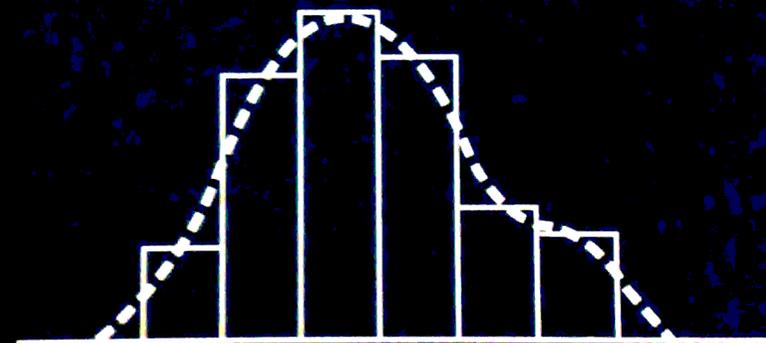




آمار و احتمال مقدماتی

چاپ سی و ششم



دکتر جواد بهبودیان

استاد دانشگاه شیراز

ویرایش ۳

در این ویرایش تنها غلطهای چاپی، محاسبات غلط و برخی تمرینها و مثالها اصلاح شده‌اند. ساختار این ویرایش با ویرایش ۲ تفاوتی چندانی ندارد.

جواد بهبودیان
دانشگاه شیراز
پاییز ۸۴

آمار و احتمال مقدماتی

تألیف دکتر جواد بهبودیان

استاد دانشگاه شیراز

| | | | |
|----|-------|--|-------|
| ۱۵ | | پیشگفتار مؤلف | |
| ۱۷ | | تهیه و تنظیم داده‌ها | فصل ۱ |
| ۱۷ | | مقدمه | ۱.۱ |
| ۱۷ | | ۱.۱.۱ جمعیت | |
| ۱۸ | | ۲.۱.۱ نمونه | |
| ۱۹ | | اندازه‌گیری و مقیاس‌سازی | ۲.۱ |
| ۱۹ | | ۱.۲.۱ مقیاسهای استیونز | |
| ۲۱ | | ۲.۲.۱ متغیرها | |
| ۲۲ | | ۳.۲.۱ داده‌ها | |
| ۲۲ | | ۴.۲.۱ سراسر کردن داده | |
| ۲۳ | | ۵.۲.۱ تمرین بخش دو | |
| ۲۴ | | جدولهای آماری | ۳.۱ |
| ۲۵ | | ۱.۳.۱ فراوانی و فراوانی نسبی | |
| ۲۵ | | ۲.۳.۱ فراوانی انباشته و فراوانی نسبی انباشته | |
| ۲۵ | | ۳.۳.۱ جدول فراوانی | |
| ۲۸ | | ۴.۳.۱ تمرین بخش سه | |
| ۳۰ | | نمودارهای آماری | ۴.۱ |
| ۳۰ | | ۱.۴.۱ هیستوگرام | |
| ۳۱ | | ۲.۴.۱ چندبهر فراوانی | |
| ۳۱ | | ۳.۴.۱ چندبهر فراوانی انباشته | |
| ۳۱ | | ۴.۴.۱ منحنی‌های فراوانی و فراوانی انباشته | |
| ۳۲ | | ۵.۴.۱ منحنی فراوانی نرمال | |
| ۳۴ | | ۶.۴.۱ تمرین بخش چهار | |
| ۳۵ | | معیارهای تمرکز | ۵.۱ |
| ۳۶ | | ۱.۵.۱ میانگین حسابی | |
| ۳۶ | | ۲.۵.۱ معدل وزنی | |
| ۳۶ | | ۳.۵.۱ میانگین هندسی | |
| ۳۶ | | ۴.۵.۱ میانگین توانفی | |

| | | |
|-----|--|-------|
| ۶۸ | مقدمه | |
| ۶۹ | آزمایش تصادفی | ۱.۱.۲ |
| ۷۰ | پشامد و فضای نمونه | ۲.۱.۲ |
| ۷۰ | رخ دادن یک پشامد | ۳.۱.۲ |
| ۷۱ | زیر پشامد | ۴.۱.۲ |
| ۷۱ | پشامدهای برابر | ۵.۱.۲ |
| ۷۱ | اعمال مجموعه‌ای روی پشامدها | ۶.۱.۲ |
| ۷۲ | فضای نمونه با پایان و بی‌پایان | ۷.۱.۲ |
| ۷۵ | تمرین بخش یک | ۸.۱.۲ |
| ۷۶ | تعبیرهای مختلف احتمال | ۲.۲ |
| ۷۷ | احتمال چیست؟ | ۱.۲.۲ |
| ۸۱ | تمرین بخش دو | ۲.۲.۲ |
| ۸۲ | مدل احتمال با فضای نمونه با پایان | ۳.۲ |
| ۸۲ | احتمال هر پشامد | ۱.۳.۲ |
| ۸۲ | مدل احتمال یکنواخت | ۲.۳.۲ |
| ۸۳ | چند فرمول برای محاسبه احتمال | ۳.۳.۲ |
| ۸۴ | روش تمییز مدل احتمال | ۴.۳.۲ |
| ۸۶ | محاسبه احتمال در مدل یکنواخت | ۵.۳.۲ |
| ۸۶ | تمرین بخش سه | ۶.۳.۲ |
| ۸۸ | قوانین شمارش | ۴.۲ |
| ۸۹ | اصول شمارش | ۱.۴.۲ |
| ۹۰ | فرمولهای شمارش | ۲.۴.۲ |
| ۹۲ | تعبیر جایگشت و ترکیب به زبان مجموعه‌ها | ۳.۴.۲ |
| ۹۳ | فرمول استرلینگ | ۴.۴.۲ |
| ۹۳ | شمارش با مدل‌های جعبه و مهره | ۵.۴.۲ |
| ۱۰۰ | تمرین بخش چهار | ۶.۴.۲ |
| ۱۰۳ | مدل احتمال با فضای نمونه دلخواه | ۵.۲ |
| ۱۰۵ | میدان سگما | ۱.۵.۲ |
| ۱۰۶ | تابع و تابع مجموعه‌ای | ۲.۵.۲ |
| ۱۰۷ | فضای احتمال | ۳.۵.۲ |
| ۱۱۱ | تمرین بخش پنج | ۴.۵.۲ |
| ۱۱۳ | مدل احتمال شرطی | ۶.۲ |

| | | |
|----|--|--------|
| ۳۷ | میانگین ریشه‌ای ربه دو | ۵.۵.۱ |
| ۳۷ | میانه | ۶.۵.۱ |
| ۳۷ | محاسبه میانه برای داده‌های گسته | ۷.۵.۱ |
| ۳۹ | محاسبه میانه برای داده‌های پیوسته | ۸.۵.۱ |
| ۴۰ | چندکها | ۹.۵.۱ |
| ۴۱ | محاسبه چندکها برای داده‌های گسته | ۱۰.۵.۱ |
| ۴۲ | محاسبه چندکها برای داده‌های پیوسته | ۱۱.۵.۱ |
| ۴۳ | نما | ۱۲.۵.۱ |
| ۴۳ | محاسبه نما برای داده‌های گسته | ۱۳.۵.۱ |
| ۴۳ | محاسبه نما برای داده‌های پیوسته | ۱۴.۵.۱ |
| ۴۶ | مقایسه میانگین، میانه و نما | ۱۵.۵.۱ |
| ۴۷ | میانگین اصلاح شده | ۱۶.۵.۱ |
| ۴۸ | تمرین بخش پنج | ۱۷.۵.۱ |
| ۴۹ | معیارهای پراکنندگی | ۶.۱ |
| ۵۰ | برد | ۱.۶.۱ |
| ۵۰ | میانگین انحرافها | ۲.۶.۱ |
| ۵۱ | واریانس و انحراف استاندارد | ۳.۶.۱ |
| ۵۲ | روش تبدیل یا روش کوتاه برای محاسبه میانگین و واریانس | ۴.۶.۱ |
| ۵۲ | داده‌های تبدیل شده | ۵.۶.۱ |
| ۵۲ | داده‌های استاندارد | ۶.۶.۱ |
| ۵۴ | ضریب تغییر | ۷.۶.۱ |
| ۵۶ | نیم برد چارکها | ۸.۶.۱ |
| ۵۶ | تمرین بخش شش | ۹.۶.۱ |
| ۵۸ | چولگی و برجستگی | ۷.۱ |
| ۵۸ | گشتاور و گشتاور مرکزی داده‌ها | ۱.۷.۱ |
| ۵۹ | چولگی | ۲.۷.۱ |
| ۵۹ | برجستگی | ۳.۷.۱ |
| ۶۰ | تمرین بخش هفت | ۴.۷.۱ |
| ۶۰ | چند نمودار جدید | ۸.۱ |
| ۶۱ | نمودارهای ساقه‌ای | ۱.۸.۱ |
| ۶۱ | نمودار جعبه‌ای | ۲.۸.۱ |
| ۶۴ | تمرین بخش هشت | ۳.۸.۱ |

| | | |
|-----|--|-------|
| ۱۵۶ | تابع چگالی متغیر تصادفی آمیخته | ۵.۳.۳ |
| ۱۵۷ | تمرین بخش سه | ۶.۳.۳ |
| ۱۵۸ | توأم دو متغیر تصادفی گسسته | ۴.۳ |
| ۱۵۹ | تعریف چگالی توأم دو متغیر تصادفی گسسته | ۱.۴.۳ |
| ۱۶۰ | تعریف دو متغیر تصادفی مستقل گسسته | ۲.۴.۳ |
| ۱۶۱ | تعریف توزیع توأم دو متغیر تصادفی | ۳.۴.۳ |
| ۱۶۲ | تمرین بخش چهار | ۴.۴.۳ |

۱۶۳ امید ریاضی یک متغیر تصادفی و کاربرد آن فصل ۴

| | | |
|-----|--|-------|
| ۱۶۳ | مفهوم امید ریاضی و ویژگیهای آن | ۱.۴ |
| ۱۶۴ | تعریف امید ریاضی | ۱.۱.۴ |
| ۱۶۶ | امید ریاضی تابعی از یک متغیر تصادفی | ۲.۱.۴ |
| ۱۶۶ | تعریف امید ریاضی $g(X)$ | ۳.۱.۴ |
| ۱۶۷ | امیا ریاضی یک چند جمله‌ای | ۴.۱.۴ |
| ۱۶۸ | امید ریاضی تابعی از دو متغیر تصادفی | ۵.۱.۴ |
| ۱۶۹ | تعریف امید ریاضی $h(X, Y)$ در حالت گسسته | ۶.۱.۴ |
| ۱۷۰ | امید ریاضی یک ترکیب خطی | ۷.۱.۴ |
| ۱۷۰ | امید ریاضی دو متغیر تصادفی هم‌توزیع | ۸.۱.۴ |
| ۱۷۲ | تمرین بخش یک | ۹.۱.۴ |
| ۱۷۲ | مفهوم واریانس و ویژگیهای آن | ۲.۴ |
| ۱۷۵ | تعریف واریانس | ۱.۲.۴ |
| ۱۷۶ | ویژگیهای واریانس | ۲.۲.۴ |
| ۱۷۷ | تعریف انحراف استاندارد | ۳.۲.۴ |
| ۱۷۷ | متغیر تصادفی استاندارد | ۴.۲.۴ |
| ۱۷۸ | تمرین بخش دو | ۵.۲.۴ |
| ۱۸۰ | مفهوم کواریانس و ویژگیهای آن | ۳.۴ |
| ۱۸۰ | تعریف کواریانس | ۱.۳.۴ |
| ۱۸۰ | محاسبه کواریانس | ۲.۳.۴ |
| ۱۸۲ | ویژگیهای کواریانس | ۳.۳.۴ |
| ۱۸۴ | تمرین بخش سه | ۴.۳.۴ |
| ۱۸۵ | مفهوم ضریب همبستگی و ویژگیهای آن | ۴.۴ |
| ۱۸۵ | تعریف ضریب همبستگی | ۱.۴.۴ |
| ۱۸۵ | ویژگیهای ضریب همبستگی | ۲.۴.۴ |
| ۱۸۹ | تمرین بخش چهار | ۳.۴.۴ |

| | | |
|-----|---------------------------------|-------|
| ۱۱۶ | قانون ضرب احتمال | ۱.۶.۲ |
| ۱۱۷ | تمرین بخش شش | ۲.۶.۲ |
| ۱۱۸ | فرمول بیز و کاربرد آن | ۷.۲ |
| ۱۱۸ | تفکیک فضای نمونه به چند پیشامد | ۱.۷.۲ |
| ۱۱۹ | تفکیک یک پیشامد به چند پیشامد | ۲.۷.۲ |
| ۱۲۰ | فرمول تفکیک احتمال | ۳.۷.۲ |
| ۱۲۰ | فرمول بیز | ۴.۷.۲ |
| ۱۲۲ | احتمال پیشین و احتمال پسین | ۵.۷.۲ |
| ۱۲۳ | تمرین بخش هفت | ۶.۷.۲ |
| ۱۲۶ | پیشامدهای مستقل | ۸.۲ |
| ۱۲۵ | دو پیشامد مستقل | ۱.۸.۲ |
| ۱۲۶ | دو پیشامد مستقل و دو پیشامد جدا | ۲.۸.۲ |
| ۱۲۷ | چند پیشامد مستقل | ۳.۸.۲ |
| ۱۲۸ | حاصلضرب دو مدل احتمال | ۴.۸.۲ |
| ۱۳۰ | تمرین بخش هشت | ۵.۸.۲ |

۱۳۲ متغیر تصادفی - تابع توزیع - تابع چگالی فصل ۳

| | | |
|-----|--------------------------------------|-------|
| ۱۳۲ | مفهوم متغیر تصادفی | ۱.۳ |
| ۱۳۲ | تعریف متغیر تصادفی | ۱.۱.۳ |
| ۱۳۵ | متغیر تصادفی گسسته و پیوسته | ۲.۱.۳ |
| ۱۳۵ | تمرین بخش یک | ۳.۱.۳ |
| ۱۳۶ | تابع توزیع | ۲.۳ |
| ۱۳۷ | تعریف تابع توزیع | ۱.۲.۳ |
| ۱۴۱ | ویژگیهای تابع توزیع | ۲.۲.۳ |
| ۱۴۲ | محاسبه احتمال از روی تابع توزیع | ۳.۲.۳ |
| ۱۴۳ | متغیر تصادفی ثابت | ۴.۲.۳ |
| ۱۴۶ | متغیرهای تصادفی هم‌توزیع | ۵.۲.۳ |
| ۱۴۷ | تمرین بخش دو | ۶.۲.۳ |
| ۱۴۸ | تابع چگالی | ۳.۳ |
| ۱۴۸ | تابع چگالی متغیر تصادفی گسسته | ۱.۳.۳ |
| ۱۴۹ | تعریف تابع چگالی متغیر تصادفی گسسته | ۲.۳.۳ |
| ۱۵۱ | تابع چگالی متغیر تصادفی پیوسته | ۳.۳.۳ |
| ۱۵۳ | تعریف تابع چگالی متغیر تصادفی پیوسته | ۴.۳.۳ |

| | | | |
|-----|---|------|--|
| ۲۱۶ | تعریف یک نمونه تصادفی | ۲.۲۶ | |
| ۲۱۶ | بافته‌های یک نمونه تصادفی | ۲.۲۶ | |
| ۲۱۷ | آماره | ۲.۲۶ | |
| ۲۱۷ | تعریف آماره | ۵.۲۶ | |
| ۲۱۷ | تعریف برآورد و برآوردیاب | ۶.۲۶ | |
| ۲۱۹ | تمرین بخش دو | ۷.۲۶ | |
| ۲۱۹ | روشهای برآوردیابی | ۲۶ | |
| ۲۱۹ | روش گشتاورها | ۱.۳۶ | |
| ۲۲۰ | قانون اعداد بزرگ | ۲.۳۶ | |
| ۲۲۳ | روش راستمنائی ماکزیمم | ۳.۳۶ | |
| ۲۲۴ | تابع راستمنائی | ۴.۳۶ | |
| ۲۲۴ | برآورد راستمنائی ماکزیمم | ۵.۳۶ | |
| ۲۲۷ | تمرین بخش سه | ۶.۳۶ | |
| ۲۲۸ | فاصله اطمینان برای یک پارامتر | ۴۶ | |
| ۲۲۹ | تعریف فاصله اطمینان | ۱.۴۶ | |
| ۲۲۹ | طرز پیدا کردن فاصله اطمینان | ۲.۴۶ | |
| ۲۳۱ | تمرین بخش چهار | ۳.۴۶ | |
| ۲۳۲ | آزمون آماری یک پارامتر | ۵۶ | |
| ۲۳۳ | ناحیه بحرانی | ۱.۵۶ | |
| ۲۳۳ | خطاها | ۲.۵۶ | |
| ۲۳۳ | توان آزمون | ۳.۵۶ | |
| ۲۳۶ | تمرین بخش پنج | ۴.۵۶ | |
| ۲۳۷ | p مقدار | ۶۶ | |
| ۲۳۹ | تعریف p مقدار | ۱.۶۶ | |
| ۲۴۱ | طرز محاسبه p مقدار | ۲.۶۶ | |
| ۲۴۳ | تمرین بخش شش | ۳.۶۶ | |
| ۲۴۴ | متغیرهای X^1 و T و F و کاربرد آن | ۷۶ | |
| ۲۴۴ | متغیر تصادفی X^2 | ۱.۷۶ | |
| ۲۴۵ | متغیر تصادفی T | ۲.۷۶ | |
| ۲۴۶ | متغیر تصادفی F | ۳.۷۶ | |
| ۲۴۶ | چند آماره مهم بر پایه توزیع نرمال | ۴.۷۶ | |
| ۲۴۸ | استقلال X^1 و S^2 | ۵.۷۶ | |
| ۲۴۸ | توزیع $\frac{(\bar{X}-\mu)}{(\frac{S}{n})}$ | ۶.۷۶ | |

چند توزیع مهم و ارتباط آنها با هم

فصل ۵

| | | |
|-----|--|------|
| ۱۹۲ | توزیع دو جمله‌ای | ۱۵ |
| ۱۹۲ | تعریف آزمایش برنولی | ۱.۱۵ |
| ۱۹۲ | تعریف متغیر تصادفی برنولی | ۲.۱۵ |
| ۱۹۳ | تعریف توزیع دو جمله‌ای | ۳.۱۵ |
| ۱۹۴ | چگالی دو جمله‌ای | ۴.۱۵ |
| ۱۹۴ | میانگین و واریانس توزیع دو جمله‌ای | ۵.۱۵ |
| ۱۹۵ | محاسبه احتمال از روی توزیع دو جمله‌ای | ۶.۱۵ |
| ۱۹۶ | تمرین بخش یک | ۷.۱۵ |
| ۱۹۷ | توزیع پواسن | ۲۵ |
| ۱۹۸ | چگالی پواسن | ۱.۲۵ |
| ۱۹۸ | میانگین و واریانس توزیع پواسن | ۲.۲۵ |
| ۲۰۰ | تمرین بخش دو | ۳.۲۵ |
| ۲۰۲ | توزیع نرمال | ۳۵ |
| ۲۰۲ | چگالی توزیع نرمال | ۱.۳۵ |
| ۲۰۲ | میانگین و واریانس توزیع نرمال | ۲.۳۵ |
| ۲۰۳ | توزیع نرمال استاندارد | ۳.۳۵ |
| ۲۰۴ | محاسبه احتمال برای متغیر نرمال غیر استاندارد | ۴.۳۵ |
| ۲۰۴ | ترکیب خطی چند متغیر نرمال | ۵.۳۵ |
| ۲۰۵ | تمرین بخش سه | ۶.۳۵ |
| ۲۰۶ | قضیه حد مرکزی و کاربرد آن | ۲۵ |
| ۲۰۷ | نمونه تصادفی | ۱.۴۵ |
| ۲۰۸ | قضیه حد مرکزی | ۲.۴۵ |
| ۲۰۹ | ارتباط میان توزیع دو جمله‌ای و نرمال | ۳.۴۵ |
| ۲۰۹ | تصحیح پیوستگی | ۴.۴۵ |
| ۲۱۲ | تمرین بخش چهار | ۵.۴۵ |

نتیجه‌گیری آماری

فصل ۶

| | | |
|-----|-------------------------|------|
| ۲۱۲ | توزیع دو جمله‌ای | ۱۶ |
| ۲۱۵ | روش منطقی احتمال و آمار | ۱.۱۶ |
| ۲۱۶ | برآورد آماری یک پارامتر | ۲.۱۶ |
| ۲۱۶ | نمونه تصادفی | ۱.۲۶ |

۷.۷.۶ واریانس آمیخته ۲۴۹

۸.۷.۶ توزیع $\frac{\chi^2}{\nu}$ ۲۵۰

۹.۷.۶ تمرین بخش هفت ۲۵۰

فصل ۷

نتیجه گیری آماری برای پارامترهای توزیع نرمال و پارامتر نسبت

۱.۷ فاصله اطمینان و آزمون آماری برای پارامترهای توزیع نرمال ۲۵۲

۱.۱.۷ فاصله اطمینان برای μ ۲۵۲

۲.۱.۷ آزمون آماری برای μ ۲۵۵

۳.۱.۷ فاصله اطمینان برای σ^2 ۲۵۷

۴.۱.۷ آزمون آماری برای σ^2 ۲۵۹

۵.۱.۷ تمرین بخش یک ۲۶۰

۲.۷ فاصله اطمینان و آزمون آماری برای پارامترهای دو توزیع نرمال مستقل ۲۶۱

۱.۲.۷ فاصله اطمینان برای $\mu_2 - \mu_1$ ۲۶۲

۲.۲.۷ آزمون آماری برای $\mu_2 - \mu_1$ ۲۶۳

۳.۲.۷ فاصله اطمینان برای $\frac{\sigma_2^2}{\sigma_1^2}$ ۲۶۴

۴.۲.۷ آزمون آماری برای $\frac{\sigma_2^2}{\sigma_1^2}$ ۲۶۵

۵.۲.۷ تمرین بخش دو ۲۶۷

۳.۷ آزمونهای گوناگون برای پارامترهای نسبت ۲۶۸

۱.۳.۷ فاصله اطمینان برای p ۲۶۹

۲.۳.۷ آزمون آماری برای p ۲۷۰

۳.۳.۷ فاصله اطمینان برای $p_2 - p_1$ ۲۷۱

۴.۳.۷ آزمون آماری برای $p_2 = p_1$ ۲۷۲

۵.۳.۷ تمرین بخش سه ۲۷۳

فصل ۸

ضرب همبستگی و خط برگشت

۲۷۵

۱.۸ برآورد ضریب همبستگی ۲۷۵

۱.۱.۸ برآورد پارامتر ρ ۲۷۶

۲.۱.۸ ویژگیهای برآورد ρ ۲۷۷

۳.۱.۸ توزیع تقریبی R ۲۷۹

۴.۱.۸ توزیع نرمال توأم ۲۸۰

۵.۱.۸ چگالی R تحت فرض نرمال ۲۸۱

۶.۱.۸ تمرین بخش یک ۲۸۲

۲.۸ خط برگشت ۲۸۳

۱.۲.۸ مدل خطی ساده ۲۸۴

۲.۲.۸ یک نمونه از متغیر مستقل و متغیر وابسته با مدل خطی ساده ۲۸۴

۳.۲.۸ روش کمترین توانهای دوم برای برآورد β و β_0 ۲۸۵

۴.۲.۸ پیش بینی متغیر وابسته به طور متوسط ۲۸۷

۵.۲.۸ برآورد خطای تصادفی ۲۸۷

۶.۲.۸ برآورد σ^2 ۲۸۷

۷.۲.۸ خط برگشت ۲۸۸

۸.۲.۸ توزیع برآورد پابها در حالت نرمال ۲۸۹

۹.۲.۸ تمرین بخش دو ۲۹۲

فصل ۹

آزمون مدل‌های احتمال به کمک توزیع χ^2

۱.۹ آزمون χ^2 برای برازندگی ۲۹۵

۱.۱.۹ تمرین بخش یک ۲۹۹

۲.۹ آزمون χ^2 برای مستقل بودن ۳۰۱

۱.۲.۹ تمرین بخش دو ۳۰۳

۳.۹ آزمون هم‌توزیمی ۳۰۴

۱.۳.۹ تمرین بخش سه ۳۰۶

پاسخ تمرینهای انتخابی

۳۰۹

واژه نامه فارسی به انگلیسی ۳۲۲

واژه نامه انگلیسی به فارسی ۳۳۱

جدول توزیع دو جمله‌ای ۳۳۸

جدول توزیع پواسن ۳۳۹

جدول توزیع نرمال ۳۴۰

جدول توزیع χ^2 ۳۴۱

جدول توزیع T ۳۴۲

جدول توزیع F ۳۴۳

گفتارهای این کتاب تقریباً طبق برنامه درس آمار و احتمال مقدماتی و درس روشهای مقدماتی آمار معرّبه شورای عالی انقلاب فرهنگی تنظیم گردیده‌اند و باید در دو نوبت، هفته‌ای ۲ ساعت تدریس و تمرین شوند. با افزودن یا کاستن چند مطلب و مثالهای مناسب می‌توان کتاب را برای درس آمار و احتمال در رشته‌های گوناگون به کار برد. به طور کلی تمام دانشجویانی که دارای دیپلم ریاضی - فیزیک یا علوم تجربی می‌باشند و می‌خواهند بر اساس ریاضی دبیرستان با مفاهیم آمار و احتمال مقدماتی آشنا شوند، می‌توانند از تمام این کتاب استفاده کنند. با این حال، در صورتی که استاد درس مصححت بدانند، می‌توان از خواندن بعضی از قسمتهای ریاضی خودداری کرد. بدون این که به فهم سایر قسمتها لطمه‌ای وارد شود.

داده‌هایی که در این کتاب آمده‌اند، پاره‌ای را دانشجویان جمع‌آوری کرده‌اند و پاره‌ای مصنوعی می‌باشند. امید است توصیه می‌شود که دانشجویان این کتاب، داده‌های مشابه را جمع‌آوری کنند تا مفاهیم آمار و احتمال بهتر روشن شوند.

مصححی درباره چاپ تازه

پس از تقریباً یک دهه که از نخستین چاپ این کتاب می‌گذرد، دگرگونی‌هایی در این چاپ تازه پدید آمده است که به شرح زیر می‌باشند:

۱. چند نمودار مانند نمودار ساقه‌ای و نمودار جمع‌بندی افزوده شده‌اند.
۲. موضوع p -مقدار (p -value)، که این روزها در نرم افزارهای آماری کاربرد فراوان دارد، در کتاب آمده است.
۳. یک فصل درباره آزمونهای آیزیدیه می‌شود.
۴. اصلاحات لازم در متن و در مثالها و تمرینها انجام شده است.
۵. پاسخ تمرینهای انتخابی در پایان کتاب افزوده شده است.

آمار هم مانند هر علم دیگر در حرکت و نوگرایی می‌باشد. با پیشرفت و متداول شدن کامپیوتر در عصر ماه این دانش تمرینش روندی نو پیدا کرده است. نویسنده امیدوار است که چاپ تازه با زمان همگام باشد و خواننده بتواند با بهره برداری از آن و انجام محاسبات کامپیوتری از مفاهیم ساده آمار و احتمال در زندگی روزانه استقبال نماید.

در پایان برخود می‌دانم از آنهایی که برای ویرایش دوم به نحوی یاری کرده‌اند صمیمانه قدردانی کنم. به ویژه از خانم حلیمه رضائی که کار ماشین‌نویسی کامپیوتری را سامان دادند و از دانشجویان آمار دانشگاه شیراز که در امر غلط‌گیری و تهیه پاسخ تمرینها همکاری کرده‌اند متشکرم. از انتشارات دانشگاه امام رضا (ع) که برای نشر این کتاب دقت و کوشش کرده‌اند بی‌اندازه سپاسگزارم.

هم‌چنین از کلیه عزیزانی که در ویرایش سوم کتاب همکاری داشته‌اند به ویژه مؤسسه نایب و حروف‌چینی جم آقای علی‌رضا شریبان و سرکار خانم الفسانه سلیمانی و کوشش انتشارات دانشگاه امام رضا (ع) متشکر و سپاسگزارم.

جواد بهبودیان

دانشگاه شیراز، دانشکده علوم، بخش آمار

پاییز ۸۲

پیشگفتار مؤلف

در عصر ما بیش از هر عصری اعداد محرک و راهمای انسانها می‌باشند. اغلب با شنیدن یا دیدن عددی که به نحوی با ما ارتباط دارد اندیشه می‌کنیم، احساساتی می‌شویم، تغییر عقیده می‌دهیم، چاره‌جویی می‌کنیم و سرانجام به کار می‌پردازیم.

اعداد اگر نماینده تقریبی واقعیتها باشند و به خوبی تجزیه و تحلیل گردند، به حق می‌توانند ما را راهنمایی کنند. اما اگر جعلی و مخلوش باشند یا به غلط تفسیر گردند، ممکن است ما را گمراه نمایند.

هرگاه بخواهیم درباره موضوعی به طریق علمی از اعداد یاری بگیریم باید گامهای زیر پیموده شوند.

- الف- کسب اطلاعات لازم وابسته به موضوع از راه پرسشنامه یا طرح آزمایش.
 - ب- اندازه‌گیری اطلاعات جمع‌آوری شده بوسیله مقیاسهای مناسب و بیان آنها به صورت اعدادی به نام داده‌ها.
 - ج- تنظیم و تجزیه و تحلیل داده‌ها طبق اصول و ضوابطی مقبول.
- دو گام اول با وجود اصول و روشهای عمومی، بستگی تام به رشته و هدف پژوهشگر دارند و ما کمتر درباره آنها صحبت خواهیم کرد. گام سوم موضوع اساسی کار ما در این نوشته‌ها می‌باشد.

داده‌ها، یعنی اطلاعاتی که به زبان اعداد ترجمه شده‌اند اغلب دستخوش عوامل شانس می‌باشند و در صورتی می‌توانند روشنگر واقعیتها گردند که:

۱. طبق قواعدی تهیه و تنظیم گردند.
۲. قوانین شانس که بر آنها حاکم می‌باشند مشخص شوند.
۳. طبق اصولی یا اطمینان کافی تجزیه و تحلیل شوند.

در واقع مراحل یاد شده موضوع آمار و احتمال می‌باشند که در این کتاب در چند فصل بیان می‌شوند.

این کتاب، برپایه تجربه و سلیقه شخصی، به عنوان کتاب درسی برای آمار و احتمال عمومی نوشته شده است. گفتارها با زبانی ساده و مثالها و تمرینهای زیاد و در ضمن با اسلوب ریاضی در حد دبیرستان ارائه شده‌اند. با خواندن و تمرین کردن این گفتارها خواننده با آمار و احتمال آشنا می‌شود و می‌تواند خود را برای گرفتن دروسهای بالاتر در این رشته آماده نماید.

مفهوم جمعیت از نظر آمار خیلی وسیع تر از مفهوم واژه‌ای آن، مانند جمعیت فلان شهر، و یا مفهوم سازمانی آن، مانند جمعیت هلال احمر، می‌باشد. برای انجام هر کار آماری روی یک جمعیت، باید آن جمعیت و ویژگی مورد مطالعه، بدون هرگونه ابهام قبلاً مشخص شوند. مثلاً جمعیت ماشینهای پیکان و قدرت ترمز.

جمعیت آماری ممکن است با پایان یا بی‌پایان باشد. مثلاً جمعیت افرادی که در دو سال گذشته به دنیا آمده‌اند با پایان، ولی جمعیت افرادی که از فروردین امسال به بعد به دنیا می‌آیند، بی‌پایان است. گاهی جمعیت مورد مطالعه به قدری اتبوه است که عملاً می‌توان آن را بی‌پایان تلقی کرد، مثل جمعیت دانه‌های گندم که در یک مزرعه برداشت می‌شود. مطالعه یک یک افراد جمعیت، به علت هزینه زیاد و کمی وقت یا نداشتن امکانات کافی، اغلب مقدور نیست. بنابراین قسمتی از جمعیت را به جای تمام آن، به نام نمونه در نظر می‌گیرند.

۲.۱.۱- نمونه

قسمتی از جمعیت را که طبق ضوابطی مقبول انتخاب می‌شود و مطالعه آن به جای مطالعه تمام جمعیت مقدور است، نمونه‌ای از جمعیت می‌نامند. معمولاً به مصداق مشت نمونه خروار است نتیجه حاصل از مطالعه نمونه را به تمام جمعیت تعمیم می‌دهند. ولی این کار احتیاط دارد. زیرا هر مشت نمی‌تواند نمونه خروار باشد و قطعاً بی‌غرضی در انتخاب مشت و اندازه مشت در این نمایندگی نقش مهمی دارد. مسئله انتخاب یک نمونه خوب به قدری مهم است که قسمت زیادی از تئوری احتمال و آمار به آن اختصاص دارد. برای حفظ بی‌طرفی و کسب حقیقت اغلب نمونه‌هایی را در نظر می‌گیرند که انتخاب آنها کاملاً شانسی باشد. اینگونه نمونه‌ها را نمونه‌های تصادفی می‌نامند و در عمل برای استخراج آنها از قرعه کشی یا جدولی به نام جدول اعداد تصادفی، و در عصر ما از کامپیوترها، کمک می‌گیرند.

برای مطالعه ویژگیهای یک جمعیت یا نمونه‌ای از آن باید دانسته‌هایی را که از راه پرسشنامه یا طرح آزمایش در آزمایشگاه به دست می‌آیند، به زبان اعداد درآورده و سپس این اعداد را طبق اصولی تجزیه و تحلیل و تفسیر کرد. به دست آوردن چنین اعدادی تا آنجا که مربوط به شمارش یا اندازه‌گیری امور کمی مانند طول، وزن، حجم، و غیره می‌باشد. امری است عادی و حتی وسائل دقیقی برای این کار ممکن است موجود باشد. ولی اندازه‌گیری امور کیفی نظیر هوش انسان یا شکنندگی چیزها، آسان نیست و مقیاسهای دقیق برای این کار موجود نمی‌باشند. با اینحال پیشرفت صنعت و علم بستگی به روشهای دقیق اندازه‌گیری امور مورد مطالعه دارد. مقیاس‌سازی و اندازه‌گیری موضوع درس جداگانه‌ای است. در اینجا تنها برای آشنائی، به مختصر اشاره‌ای در این باره می‌پردازیم.

تهیه و تنظیم داده‌ها

۱.۱- مقدمه

واژه Statistics که به فارسی آن را آمار ترجمه کرده‌اند در اغلب زبانها به دو معنی به کار می‌رود:

الف - به معنی ارقام و اعداد واقعی یا تقریبی درباره اموری از قبیل زاد و مرگ، طلاق، میزان محصولات کشاورزی و صنعتی، تصادفات رانندگی و غیره. در این رابطه معمولاً دو اثری مثلاً به نام دفترهای آمار در سازمانهای دولتی موجود است.

ب - به معنی روشهایی برای جمع‌آوری، تنظیم و تجزیه و تحلیل اطلاعات عددی درباره موضوعی. با اینکه این دو مفهوم با هم ارتباط دارند، در این کتاب ما بیشتر با مفهوم دوم سر و کار داریم. تنها در فصل اول مطالبی را تحت عنوان تهیه و تنظیم داده‌ها، که اغلب آمار توصیفی نامیده می‌شوند، شرح می‌دهیم.

ما روزانه در خیابان و بازار، در مدرسه و دانشگاه، با دیدار دوستان و خویشان و از راه رسانه‌ها در معرض داده‌های آماری می‌باشیم. برداشت و قضاوت معقول بر اساس این داده‌ها، به کمک مفهوم دوم، که جنبه علمی دارد، انجام می‌گیرد.

در آمار مطالعه و قضاوت معقول درباره موضوعهای گوناگون، بر مبنای یک جمع انجام می‌شود و قضاوت بر مبنای یک فرد خاص، اصلاً مطرح نمی‌باشد. بنابراین نخست مفهوم کلمه جمعیت را از نظر آمار شرح می‌دهیم.

۱.۱.۱- جمعیت

مجموعه افراد یا چیزهایی را که می‌خواهیم یک یا چند ویژگی درباره آنها مطالعه کنیم، یک جمعیت می‌نامیم. مثلاً جمعیت ماشینهای پیکان که در دو سال گذشته به بازار آمده‌اند از نظر قدرت ترمز، یا جمعیت نوزادانی که در سال گذشته به دنیا آمده‌اند از نظر مصرف شیر خشک.

۲.۱ اندازه‌گیری و مقیاس‌سازی

فرض کنید a یکی از افراد جمعیت π و i یک ویژگی مورد مطالعه باشد. مثلاً سیبهای یک باغ را جمعیت π را یکی از سیبها بگیریید. ویژگی i برای a ممکن است وزن سیب باشد، که امری است کمی، یا تردی سیب باشد، که امری است کیفی. منظور از اندازه‌گیری ویژگی i در a ، اطلاق یک عدد حقیقی x به a طبق ضابطه‌ای مشخص می‌باشد. به سخنی دیگر x تابعی از a به صورت $x=f(a)$ است. این تابع یا x را مقیاس و طرز تعیین عدد x را اندازه‌گیری یا مقیاس‌سازی می‌گویند. اگر i وزن سیب باشد می‌توان عدد x را با مقیاس وزن مثلاً برحسب گرم با تقریب لازم پیدا کرد. ولی اگر i تردی سیب باشد، مقیاسی شناخته شده برای اندازه‌گیری و پیدا کردن x در دست نیست.

با اینحال سیب شناسان می‌توانند منظور از تردی سیب را تعریف نمایند و مقیاسی برای تردی سیب معرفی کنند. باید توجه داشت که نوع عددی که نماینده وزن است با عددی که نماینده تردی است با هم تفاوت دارند. مثلاً می‌توان گفت که وزن یک سیب دو برابر وزن سیب دیگر است ولی این بیان در مورد تردی مفهومی ندارد.

استیونز S.S. Stevens، استاد روانشناسی دانشگاه هاروارد آمریکا، در مقاله بنیادی ۱۹۴۶ خود چهار نوع مقیاس معرفی کرده است. این مقیاسها در حدود نیم قرن است که در بعضی از کتابهای آمار بیان می‌شوند و با اینکه در عصر آمار و کامپیوتر بارها مورد انتقاد قرار گرفته‌اند، هنوز دارای اهمیت هستند. برای اطلاع بیشتر در این باره به کتابنامه رجوع کنید. اینک مقیاسهای استیونز را به کوتاهی شرح می‌دهیم.

۱.۲.۱ مقیاسهای استیونز

الف - مقیاس اسمی: هرگاه مقیاس x که معمولاً یک عدد طبیعی است، تنها برای شناسائی افراد یا چیزها یا مکانها به کار رود، آن را یک مقیاس اسمی می‌نامند. مثلاً اگر کارگران یک کارخانه از شهرهای تهران، اصفهان، شیراز و کرمان باشند و به ترتیب آنها را با اعداد ۱ و ۲ و ۳ و ۴ مشخص کنیم، این اعداد صرفاً می‌گویند که هر کارگر از کدام شهر است. کارگری که برحسب ۳ دارد از شیراز و کارگری که برحسب ۴ دارد از کرمان است. توجه کنید که این اعداد را نمی‌توان برای مقایسه با چهار عمل اصلی حساب به کار برد.

مقیاس اسمی x را با هر تابع یک به یک می‌توان به مقیاس اسمی y تبدیل کرد بی‌آنکه شناسائی تغییر نماید. مثلاً تابع یک به یک $y = x + 10$ = لایبرجسبهای ۱، ۲، ۳، ۴ را به ترتیب به برجسبهای ۱۱، ۱۲، ۱۳، ۱۴ تبدیل می‌کند.

ب - مقیاس ترتیبی: هرگاه مقیاس x که یک عدد حقیقی است، برتری را بیان کند. آن را یک مقیاس ترتیبی می‌نامند. به سخنی دیگر اگر a_2 از نظر ویژگی i بر a_1 برتری داشته باشد و ویژگی i را برای a_2 و a_1 با اعداد x_2 و x_1 نشان دهیم، باید داشته باشیم $x_2 > x_1$. مثلاً اگر مهندس یک کارخانه کارگران را از نظر مهارت با اعداد ۱، ۲، ۳، ۴ مشخص کند، کارگر شماره ۴ از کارگر شماره ۲ ماهرتر است. ولی نمی‌توان گفت که دو برابر او مهارت دارد؛ اینگونه اعداد را می‌توان تنها برای مقایسه به کار برد و نمی‌توان با آنها چهار عمل اصلی انجام داد.

مقیاس ترتیبی x را می‌توان با یک تابع صعودی به مقیاس ترتیبی y تبدیل کرد، بی‌آنکه برتری یا ترتیب مختل گردد. مثلاً تابع $x^3 = y$ اعداد ۱، ۲، ۳، ۴ را به ۱، ۸، ۲۷، ۶۴ تبدیل می‌کند و با اعداد اخیر هم کارگر چهارم از کارگر دوم ماهرتر است.

ج - مقیاس فاصله‌ای: هرگاه مقیاس x که یک عدد حقیقی است، نسبت دو تفاضل یا هر دو فاصله را حفظ کند، آن را یک مقیاس فاصله‌ای می‌نامند. به سخنی دیگر اگر با این مقیاس برای a_2 ، a_1 و ویژگی i با اعداد x_1, x_2, x_3, x_4 اندازه‌گیری شود، باید نسبت $\frac{(x_2 - x_3)}{(x_4 - x_1)}$ ثابت بماند و به واحد اندازه‌گیری بستگی نداشته باشد. مثلاً فرض کنید ویژگی i گرمای چهار جسم باشد و با مقیاس سانتیگراد داشته باشیم:

$$x_1 = 10, \quad x_2 = 15, \quad x_3 = 20, \quad x_4 = 45$$

ملاحظه می‌شود که

$$\frac{x_2 - x_3}{x_4 - x_1} = 5$$

اینک با مقیاس فارنهایت طبق فرمول $y = \frac{9x}{5} + 32$ داریم:

$$y_1 = 50, \quad y_2 = 59, \quad y_3 = 68, \quad y_4 = 113$$

بازهم ملاحظه می‌شود که

$$\frac{y_2 - y_3}{y_4 - y_1} = 5$$

در مثال بالا نکته‌های زیر قابل توجه می‌باشند:

(۱) اگر با مقیاس سانتیگراد جسم a دارای درجه حرارت صفر باشد، با مقیاس فارنهایت این جسم دارای درجه حرارت ۳۲ است. بنابراین با مقیاس فاصله‌ای صفر معنی هیچ نمی‌دهد و صرفاً جنبه قراردادی دارد.

ب - متغیر عددی: مانند شماره فرزندان یک خانواده که از راه شمارش به دست می‌آید، یا هوش یک پسر ۹ ساله و وزن یک جوان بیست ساله که از راه اندازه‌گیری با مقیاس فاصله‌ای و نسبتی به دست می‌آیند.

۳.۲.۱ داده‌ها

فرض کنید می‌خواهیم ویژگی x از یک جمعیت را، که معمولاً یک متغیر است، مطالعه کنیم. اگر این متغیر را در مورد یک یک افراد جمعیت یا نمونه‌ای از آن با مقیاس مناسب اندازه‌گیری کنیم، یک مجموعه از اعداد به دست می‌آید که آن را داده‌ها می‌نامند. داده‌ها دو نوع‌اند:

الف - داده‌های گسسته: از راه اندازه‌گیری با مقیاس‌های اسمی، ترتیبی، یا شمارشی بدست می‌آیند. داده‌های گسسته را داده‌های جدا از هم می‌نامند.

ب - داده‌های پیوسته: از راه اندازه‌گیری با مقیاس‌های فاصله‌ای یا نسبتی بدست می‌آیند.

داده‌ها اغلب بصورت انبوهی از اعداد ارائه می‌شوند و به خودی خود خام هستند. برای اینکه بتوان آنها را پخته کرد و حقایق را جويا شد باید:

الف - آنها را در جدول‌هایی تنظیم کرد.

ب - از روی جدول‌ها، نمودارهایی رسم نمود.

ج - آنها را در یک یا چند عدد مختصر کرد.

تنها بعد از طی این مراحل می‌توان، قوانین شانس حاکم بر آنها را پیدا کرده و به برداشتهای آماری و تهیه گزارش نهائی درباره ویژگی مورد مطالعه پرداخت.

۴.۲.۱ سر راست کردن داده‌های پیوسته

اندازه‌گیری با مقیاس‌های اسمی و ترتیبی، چندان مفهوم و مصداقی ندارد، زیرا داده‌های حاصل معمولاً از راه نامگذاری یا ترتیب بدست می‌آیند. اندازه‌گیری را حتماً در مورد مقیاس‌های فاصله‌ای و نسبتی باید بکار برد. داده‌های حاصل از این راه معمولاً پیوسته می‌باشند. داده پیوسته یک عدد حقیقی است که به صورت اشاری با تعدادی ارقام اعشار بیان می‌شود.

منظور از سر راست کردن یا رند کردن عدد حقیقی x تا r رقم اعشار، به دست آوردن یک عدد اشاری x' با حداکثر r رقم اعشار می‌باشد، به طوری که داشته باشیم

$$|x - x'| \leq 0.5 \times 10^{-r}$$

(۲) با مقیاس اول گرمای a_1 سه برابر گرمای a_2 است، ولی با مقیاس دوم چنین نیست. بنابراین در اندازه‌گیری با مقیاس فاصله‌ای، نسبت محفوظ نمی‌ماند.

(۳) با هر دو مقیاس گرمای a_2 از a_1 پنج برابر تفاضل گرمای a_1 از a_2 است. بنابراین با مقیاس فاصله‌ای، نسبت فاصله‌ها با تغییر مقیاس فاصله‌ای تغییر نمی‌کند.

مقیاس فاصله‌ای x را می‌توان با تابع خطی $y = ax + b$ با فرض $a > 0$ ، به مقیاس فاصله‌ای y تبدیل کرد، بی‌آنکه نسبت دو تفاضل تغییر کند.

مقیاس فاصله‌ای معمولاً در روانشناسی، تعلیم و تربیت، جامعه‌شناسی و فیزیک، به ویژه در اموری مانند هوش، حافظه، ارزیابی آزمونها، حرارت و... به کار می‌رود. این نوع مقیاس‌سازی به مهارت و تجربه فراوان در رشته مورد بحث بستگی دارد و گاهی دستخوش مجادله می‌باشد.

د - مقیاس نسبتی: هرگاه مقیاس x که یک عدد حقیقی است، نسبت را حفظ کند، آن را یک مقیاس نسبتی می‌نامند. به سختی دیگر اگر با این مقیاس برای دو جسم a_1 و a_2 ویژگی x ، مثلاً وزن، یا x_1 و x_2 اندازه‌گیری شود، باید $\frac{x_1}{x_2}$ ثابت بماند و به واحد اندازه‌گیری بستگی نداشته باشد. اگر وزن a_1 و a_2 با مقیاس گرم 6000 و 2000 باشد، با مقیاس کیلوگرم 6 و 2 است. ملاحظه می‌شود که با هر دو مقیاس وزن a_1 سه برابر وزن a_2 است.

مقیاس نسبتی برای امور کمی مانند وزن، طول، سطح، مقدار حرارت به کار می‌رود و عالیترین نوع مقیاس است. با اعداد حاصل از این مقیاس می‌توان چهار عمل اصلی حساب را انجام داد. مقیاس نسبتی x را می‌توان با یک تابع خطی $y = ax$ با فرض $a > 0$ ، به مقیاس نسبتی y تبدیل کرد، بی‌آنکه نسبت تغییر کند.

ما در این کتاب، ضمن بحث دربارهٔ مباحث آماری گوناگون، به مقیاس‌های مورد نیاز اشاره خواهیم کرد.

۲.۲.۱ متغیرها

ویژگی x ، مثلاً گروه خونی، مهارت، هوش، و وزن، در افراد مختلف جمعیت یکسان نیست و معمولاً از فردی به فرد دیگر تغییر کرده، کاهش یا افزایش می‌یابد. از این رو x را یک متغیر می‌نامیم. دو نوع متغیر داریم:

الف - متغیر گروهی: مانند گروه خونی و مهارت، که با مقیاس اسمی یا ترتیبی سنجیده می‌شوند و بر اساس آن جمعیت را گروه‌بندی می‌کنند.

ب - صفحات یک کتاب و شماره غلطها را در یک صفحه.

ج - اتومبیلهای پیکان که پارسال تولید شده‌اند و مصرف بنزین برای صد کیلومتر.

د - دانش‌آموزان یک شهر و میزان یادگیری درس ریاضی.

۳. نشان دهید که مقیاس فاصله‌ای x را می‌توان با تابع $y = ax + b$ ، $a > 0$ ، به مقیاس فاصله‌ای z تبدیل کرد.

۴. نشان دهید که مقیاس نسبتی x را می‌توان با تابع $y = ax$ ، $a > 0$ ، به مقیاس نسبتی z تبدیل کرد.

۵. اعداد زیر را تا دو رقم اعشار یا تا سه رقم اعشار سر راست کنید.

$$2/5120 \quad 2/7251 \quad 2/9252 \quad 2/2715$$

۶. اعداد ۱۹۷۲۶ و ۱۴۵۹۸ و ۱۱۷۲۳ را تا نزدیکترین ده واحد سر راست کنید؟

۷. با یکی از توابع زیر مقیاس x را به z تبدیل می‌کنیم. کدام نوع مقیاس تغییر نمی‌کند؟

$$y = \log x \quad y = x^2 + 1 \quad y = 2x + 3 \quad y = 10x$$

۸. مقیاس ریشتر (به نام چارلز ریشتر زلزله‌شناس آمریکائی متولد ۱۹۰۰ میلادی) عبارتست از لگاریتم دهمی ماکزیم دامنه‌نوسان ثبت شده روی ماشین نوار زلزله سنج برحسب میکرون. مقیاس ریشتر چه نوع مقیاسی است؟

۳.۱ جدولهای آماری

نمایشی داده‌ها را، با نظمی خاص، در چند سطر و ستون یک جدول آماری می‌گویند. یک جدول آماری باید به نحوی تنظیم شود که بتوان به آسانی پاره‌ای از دانسته‌های نهفته در داده‌ها را از روی آن خواند. شماره جدول، نام جدول، عنوانهای سطر و ستون، زیر نویس، و بالاخره مأخذ جدول در صورت لزوم باید روشن و گویا باشند. جدولی که از روی تمام داده‌ها به دست می‌آید، جدول اصلی و جدولی که از روی جدول اصلی، برای بررسی دانسته‌هایی ویژه مشتق می‌شود، جدول فرعی نامیده می‌شود. در کارهای آماری برای این که داده‌ها را خلاصه کنند، آنها را در جدولی به نام جدول فراوانی تنظیم می‌کنند.

به مثالهای زیر توجه کنید

اگر عدد ۷۳/۲۸۲۵۰۰۰ را تا یک رقم اعشار (یا نزدیکترین دهم واحد) سر راست کنیم، عدد ۷۳/۴ به دست می‌آید.

اگر عدد ۵۱/۲۹۴۳۰۰۰ را تا دو رقم اعشار (یا نزدیکترین صدم واحد) سر راست کنیم، عدد ۵۱/۲۹ به دست می‌آید.

اگر عدد ۳۵/۴۷۳۵۰۰۰ را تا سه رقم اعشار (یا نزدیکترین هزارم واحد) سر راست کنیم، عدد ۳۵/۴۷۳ به دست می‌آید.

اگر عدد ۷۸/۲۳۶۵۰۰۰۰ را تا چهار رقم اعشار (یا نزدیکترین ده هزارم واحد) سر راست کنیم، عدد ۷۸/۲۳۶۷ به دست می‌آید.

با تعریف بالا و با توجه به مثالها، قاعده کلی زیر را برای سر راست کردن یک عدد حقیقی تا r را اعشار داریم:

الف - اگر رقم اعشار $(r+1)$ ام بیشتر از ۵ (یا ۵ و رقم r زوج) باشد، عدد یک را به رقم r ام اضافه کرده بعداً تمام ارقام اعشار بعد از رقم r ام را حذف می‌کنیم.

ب - اگر رقم اعشار $(r+1)$ ام کمتر از ۵ (یا ۵ و رقم r فرد) باشد، تمام ارقام اعشار بعد از رقم r ام را حذف می‌کنیم.

به همین طریق می‌توان عددی را تا نزدیکترین واحد، تا نزدیکترین ده واحد، تا نزدیکترین صد واحد، و ... سر راست کرد. مثلاً اگر عدد ۴۲/۳ را تا نزدیکترین واحد سر راست کنیم، عدد ۴۲ به دست می‌آید. اگر عدد ۷۶ را تا نزدیکترین ده واحد سر راست کنیم عدد ۸۰ به دست می‌آید.

اگر عدد ۸۵۲۰ را تا نزدیکترین صد واحد سر راست کنیم عدد ۸۵۰۰ به دست می‌آید.

۵.۲.۱ تمرین بخش دو

۱. این ویژگیها را با چه مقیاس‌هایی می‌توان اندازه‌گیری کرد؟ کدامیک از متغیرهای مربوط گسسته و کدام پیوسته می‌باشند؟

نژاد، زمان، درآمد و شغل.

۲. منظور از جمعیت، نمونه، متغیر و مقیاس مربوط را در موارد زیر مشخص کنید:

الف - دانش‌آموزان یک مدرسه و قدرت دید.

۱.۳.۱ فراوانی و فراوانی نسبی

هرگاه n چیز از k نوع T1, T2, ..., Tk با فرض n ≥ k ≥ ۱ به ترتیب با تعدادهای f1, f2, ..., fk تشکیل شده باشند، این تعدادها را فراوانیها و f1/n, f2/n, ..., fk/n را فراوانیهای نسبی این چیزها می‌گویند. فراوانیهای نسبی را به ترتیب با r1, r2, ..., rk نشان می‌دهند و واضح است که برای i = ۱, ۲, ..., k داریم:

∑_{i=1}^k f_i = n, ۱ ≤ f_i < n, ∑_{i=1}^k r_i = ۱, ۱/n ≤ r_i < ۱

۲.۳.۱ فراوانی انباشته و فراوانی نسبی انباشته

با توجه به فراوانی و فراوانی نسبی، برای i = ۱, ۲, ..., k

s_i = ∑_{j=1}^i f_j, g_i = ∑_{j=1}^i r_j

را به ترتیب، فراوانی انباشته و فراوانی نسبی انباشته می‌گویند و واضح است که

s_k = ۱, g_k = n

مثال ۱ اگر گروهی ۱۰ نفری از ۳ شیرازی، ۵ تهرانی، و ۲ کرمانی تشکیل شده باشد، فراوانی شیرازیها، تهرانیها و کرمانیها در این گروه به ترتیب ۳، ۵ و ۲ و فراوانی نسبی آنها ۰/۳، ۰/۵ و ۰/۲ می‌باشد. فراوانی انباشته و فراوانی نسبی انباشته شیرازیها و تهرانیها به ترتیب ۰/۸ است.

۳.۳.۱ جدول فراوانی

هرگاه داده‌ها را به نحوی تقسیم کرده، آنها را برحسب فراوانیها در جدولی تنظیم کنیم، این جدول را جدول فراوانی یا جدول توزیع فراوانی می‌نامند. برای اینکه جدول فراوانی را با توجه به داده‌ها از لحاظ داده‌های گسسته و پیوسته، شرح دهیم به مثالهای زیر توجه کنید:

مثال ۲ پزشکی ۲۰ بیمار قلبی دارد که گروه خونی آنها عبارتند از:

B A O AB O A A A O O A A B B AB O AB AB O O

جدول ۱. جدول فراوانی گروه خونی ۲۰ بیمار قلبی با متغیر اسمی xi

Table with 6 columns: xi, fi, ri, gi, si. Rows: A, B, AB, O, and a total row.

می‌خواهیم از این داده‌ها یک جدول فراوانی بسازیم.

نخست چهار گروه خونی O, AB, B, A را به ترتیب با اعداد ۴، ۳، ۲، ۱ متناظر می‌کنیم تا داده‌های اسمی به دست آیند. این اعداد صرفاً برای گروه بندی و نامگذاری است و نمی‌توان روی آنها چهار عمل اصلی را انجام داد. بعد از شمارش لازم، جدول ۱، به نام جدول گروه خونی ۲۰ بیمار قلبی به دست می‌آید. داده‌های این مثال نمونه‌ای از داده‌های گروهی است که در واقع داده‌های گسسته می‌باشند.

مثال ۳ داده‌های زیر تعداد فرزندان ۲۰۰ خانواده در یکی از شهرهای ایران می‌باشند:

Large data table with 10 columns representing the number of children in 200 families.

در این مثال ویژگی مورد مطالعه، تعداد فرزندان یک خانواده است، که با اعداد ۰، ۱، ۲، ۳، ۴، ۵، ۶ مشخص شده است. این اعداد برای گروه بندی به کار می‌روند و جنبه ترتیبی دارند. داده‌های بالا

جدول ۲ جدول فراوانی تعداد فرزندان ۲۰۰ خانواده با متغیر ترتیبی x_i

| x_i | f_i | r_i | g_i | s_i |
|-------|-------|-------|-------|-------|
| ۰ | ۲۰ | ۰/۱۰ | ۲۰ | ۰/۱۰ |
| ۱ | ۳۰ | ۰/۱۵ | ۵۰ | ۰/۲۵ |
| ۲ | ۵۰ | ۰/۲۵ | ۱۰۰ | ۰/۵۰ |
| ۳ | ۴۰ | ۰/۲۰ | ۱۴۰ | ۰/۷۰ |
| ۴ | ۳۰ | ۰/۱۵ | ۱۷۰ | ۰/۸۵ |
| ۵ | ۲۰ | ۰/۱۰ | ۱۹۰ | ۰/۹۵ |
| ۶ | ۱۰ | ۰/۰۵ | ۲۰۰ | ۱/۱۰۰ |
| | ۲۰۰ | ۱/۱۰۰ | | |

نمونه‌ای از داده‌های گسسته هستند که در جدول ۲، به نام جدول فراوانی تعداد فرزندان ۲۰۰ خانواده خلاصه شده‌اند.

مثال ۴ داده‌های زیر اندازه‌های قد ۱۰۰ جوان بیست ساله در یکی از شهرهای ایران می‌باشند، که بر حسب سانتیمتر تا نزدیکترین واحد سر راست شده‌اند.

| | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| ۱۷۲ | ۱۸۴ | ۱۷۰ | ۱۵۰ | ۱۶۴ | ۱۷۱ | ۱۷۸ | ۱۷۲ | ۱۷۷ | ۱۵۱ |
| ۱۵۸ | ۱۶۵ | ۱۶۹ | ۱۷۲ | ۱۶۶ | ۱۶۴ | ۱۶۹ | ۱۵۹ | ۱۵۳ | ۱۷۰ |
| ۱۶۲ | ۱۵۴ | ۱۵۹ | ۱۸۲ | ۱۷۴ | ۱۶۲ | ۱۵۱ | ۱۶۵ | ۱۷۲ | ۱۵۶ |
| ۱۶۳ | ۱۷۰ | ۱۷۷ | ۱۸۴ | ۱۷۵ | ۱۷۱ | ۱۶۴ | ۱۵۶ | ۱۵۰ | ۱۶۵ |
| ۱۷۲ | ۱۶۹ | ۱۶۲ | ۱۷۲ | ۱۷۶ | ۱۷۵ | ۱۷۴ | ۱۶۹ | ۱۶۶ | ۱۵۴ |
| ۱۶۲ | ۱۶۷ | ۱۸۰ | ۱۶۹ | ۱۵۲ | ۱۵۹ | ۱۶۱ | ۱۶۴ | ۱۷۱ | ۱۵۶ |
| ۱۶۳ | ۱۷۰ | ۱۶۵ | ۱۷۲ | ۱۸۴ | ۱۷۰ | ۱۵۰ | ۱۶۴ | ۱۷۱ | ۱۷۰ |
| ۱۸۴ | ۱۷۷ | ۱۵۳ | ۱۷۰ | ۱۶۲ | ۱۵۴ | ۱۷۳ | ۱۷۷ | ۱۶۹ | ۱۷۲ |
| ۱۶۶ | ۱۶۴ | ۱۷۴ | ۱۶۹ | ۱۷۹ | ۱۷۷ | ۱۷۹ | ۱۷۳ | ۱۶۵ | ۱۶۵ |
| ۱۵۷ | ۱۶۳ | ۱۵۳ | ۱۵۸ | ۱۵۱ | ۱۶۲ | ۱۵۹ | ۱۶۱ | ۱۶۰ | ۱۸۳ |

| | | | | | | | | | |
|---|----|---|----|----|---|---|---|---|----|
| ۷ | ۹ | ۳ | ۱۱ | ۲ | ۵ | ۳ | ۲ | ۸ | ۳ |
| ۳ | ۲ | ۱ | ۲ | ۱۱ | ۶ | ۸ | ۹ | ۷ | ۴ |
| ۵ | ۱۱ | ۹ | ۴ | ۵ | ۲ | ۳ | ۲ | ۲ | ۷ |
| ۳ | ۵ | ۴ | ۹ | ۲ | ۲ | ۳ | ۴ | ۹ | ۱۱ |
| ۸ | ۱۱ | ۳ | ۲ | ۲ | ۶ | ۴ | ۵ | ۹ | ۸ |

چه نوع داده‌هایی داریم؟ یک جدول فراوانی کامل برای این داده‌ها پیدا کنید. چند درصد خانواده‌ها بیش از ۵ قرص در ماه مصرف می‌کنند؟

۲. در آزمونی یک پرسشنامه ۲۰ پرسشی، برای اندازه‌گیری استعداد ریاضی، به ۴۰ نفر از دانش‌آموزان کلاس پنجم ابتدائی یکی از دبستانها داده‌اند. نمره‌های این آزمون عبارتند از:

| | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|
| ۱۴ | ۱۲ | ۱۲ | ۱۳ | ۱۰ | ۷ | ۸ | ۱۱ | ۱۴ | ۱۳ |
| ۱۴ | ۱۳ | ۹ | ۱۰ | ۱۰ | ۱۴ | ۱۳ | ۱۲ | ۱۱ | ۱۰ |
| ۱۲ | ۱۷ | ۱۴ | ۹ | ۱۲ | ۱۴ | ۱۳ | ۱۳ | ۱۲ | ۱۴ |
| ۹ | ۷ | ۱۵ | ۱۵ | ۱۲ | ۱۱ | ۱۱ | ۱۲ | ۱۰ | ۹ |

داده‌ها از چه مقیاسی به دست آمده‌اند؟ یک جدول فراوانی کامل برای داده‌ها پیدا کنید. چند درصد دانش‌آموزان نمره‌های کمتر از ۱۲ دارند؟

۳. وزنه‌های ۴۰ بسته پسته تا نزدیکترین کیلو عبارتند از:

| | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| ۱۳۸ | ۱۶۴ | ۱۵۰ | ۱۳۲ | ۱۴۴ | ۱۲۵ | ۱۴۹ | ۱۵۷ | ۱۴۶ | ۱۵۸ |
| ۱۴۰ | ۱۴۷ | ۱۳۶ | ۱۴۸ | ۱۵۲ | ۱۴۴ | ۱۶۸ | ۱۲۶ | ۱۳۸ | ۱۷۶ |
| ۱۶۳ | ۱۱۹ | ۱۵۴ | ۱۶۵ | ۱۴۶ | ۱۷۳ | ۱۴۲ | ۱۴۷ | ۱۳۵ | ۱۵۳ |
| ۱۴۰ | ۱۳۵ | ۱۶۱ | ۱۳۵ | ۱۴۵ | ۱۴۲ | ۱۵۰ | ۱۵۶ | ۱۴۵ | ۱۲۸ |

چه نوع داده‌هایی داریم؟ یک جدول فراوانی کامل با ۸ رده به طولهای مساوی پیدا کنید.

۴. میزان هموگلوبین خون در ۵۰ بیمار سرطانی برحسب گرم در ۱۰۰ میلی‌لیتر عبارتند از:

| | | | |
|-----|-----|-----|-----|
| ۱۵۲ | ۱۴۰ | ۱۴۷ | ۱۵۵ |
|-----|-----|-----|-----|

| | | | | | | | | | |
|------|------|------|------|------|------|------|------|------|------|
| ۱۳/۶ | ۱۴/۸ | ۱۳/۷ | ۱۴/۲ | ۱۱/۵ | ۱۱/۹ | ۱۳/۸ | ۱۴/۶ | ۱۴/۲ | ۱۲/۷ |
| ۱۳/۴ | ۱۱/۵ | ۱۱/۹ | ۱۴/۸ | ۱۲/۷ | ۱۲/۴ | ۱۵/۳ | ۱۵/۲ | ۱۳/۵ | ۱۵/۰ |
| ۱۲/۴ | ۱۲/۰ | ۱۳/۸ | ۱۱/۷ | ۱۰/۰ | ۱۳/۲ | ۱۵/۵ | ۱۴/۰ | ۱۳/۵ | ۱۵/۰ |
| ۱۲/۷ | ۱۲/۹ | ۱۳/۷ | ۱۵/۱ | ۱۳/۵ | ۱۲/۷ | ۱۵/۷ | ۱۰/۹ | ۱۴/۰ | ۱۴/۸ |
| ۱۴/۰ | ۱۳/۸ | ۱۴/۷ | ۱۱/۹ | ۱۲/۰ | ۱۱/۴ | ۱۱/۱ | ۱۳/۷ | ۱۳/۲ | ۱۶/۲ |

با تشکیل ۷ رده به طول ۰/۹، جدول فراوانی کامل را تشکیل دهید.

۴.۱ نمودارهای آماری

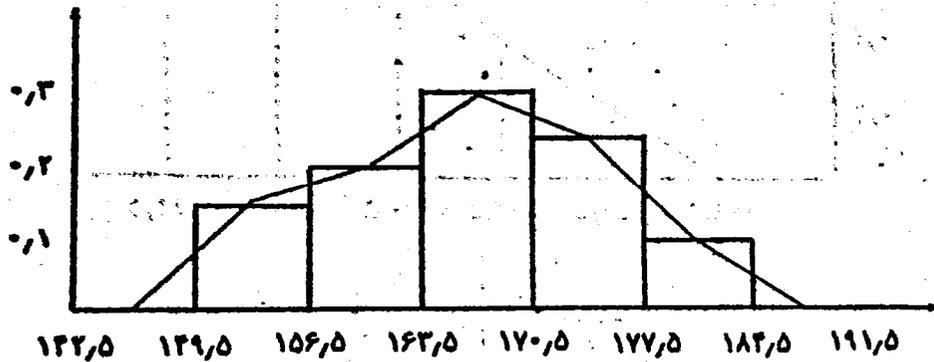
نمایش داده‌ها را، طبق قراردادهای خاص به صورت هندسی، یک نمودار آماری می‌گویند. یک نمودار آماری باید به نحوی ترسیم شود که بتوان به راحتی اطلاعات نهفته در داده‌ها را از روی آن تا حدودی با چشم بدون توضیح و تشریح اضافی دید. هر نمودار آماری، باید دارای شماره، عنوان، و در صورت لزوم زیرنویس و مأخذ باشد. مقیاسهای اندازه‌گیری روی محورهای افقی و عمودی باید مشخص باشند. نمودارهای آماری در امور اقتصادی، صنعتی، بهداشتی، و غیره به کار می‌روند و برحسب رشته مربوط آنها را به طرق مختلف ترسیم می‌کنند. در اینجا فقط چند نوع نمودار که در آمار و احتمال مورد نیاز می‌باشند شرح می‌دهیم. در پایان این گفتار هم چند نمودار جدید را معرفی می‌کنیم.

۱.۴.۱ هیستوگرام

نموداری است مرکب از چند مستطیل که از روی جدول فراوانی داده‌های پیوسته ساخته می‌شود. تعداد مستطیل‌ها برابر است با تعداد رده‌ها. قاعده هر مستطیل روی محور x جا دارد و طولش برابر است با طول واقعی رده، که هر چه باشد آن را یک واحد تلقی می‌کنیم، و مرکز آن نماینده رده است. ارتفاع هر مستطیل برابر است با فراوانی نسبی رده مربوط. هیستوگرام را برای داده‌های جدا هم، که بتوان آنها را در یک جدول فراوانی با فواصل مساوی تنظیم کرد، نیز به کار می‌برند. مساحت تمام مستطیل‌های یک هیستوگرام برابر یک واحد مربع است. اگر رده‌ها در جدول فراوانی دارای طولهای مختلف باشند، قاعده‌های تمام مستطیلها برابر نخواهد بود. در این موارد باید ارتفاع مستطیلها را طوری انتخاب کرد که مساحت تمام آنها برابر یک واحد مربع شود (تمرین ۵). در نگاره ۱ هیستوگرام جدول فراوانی ۳ داده شده است. ملاحظه می‌شود که فراوانی‌های نظیر رده‌های ۱۳۹/۵ - ۱۴۲/۵ و ۱۴۲/۵ - ۱۹۱/۵ - ۱۸۴/۵ صفر می‌باشند، ولی برای ترسیم نمودارهای بعدی آنها را منظور می‌داریم.

۲.۲.۱ چندبر فراوانی

اگر نقطه‌های وسط قاعده‌های بالائی مستطیل‌های هیستوگرام و نقطه‌های وسط رده‌های راکه بلافاصله در دو انتهای هیستوگرام بوده، و دارای فراوانی صفر هستند، به هم بپیوندیم، یک خط شکسته به دست می‌آید که آن را چندبر فراوانی می‌نامند. در نگاره ۱ چندبر فراوانی مربوط به جدول ۳ داده شده است.



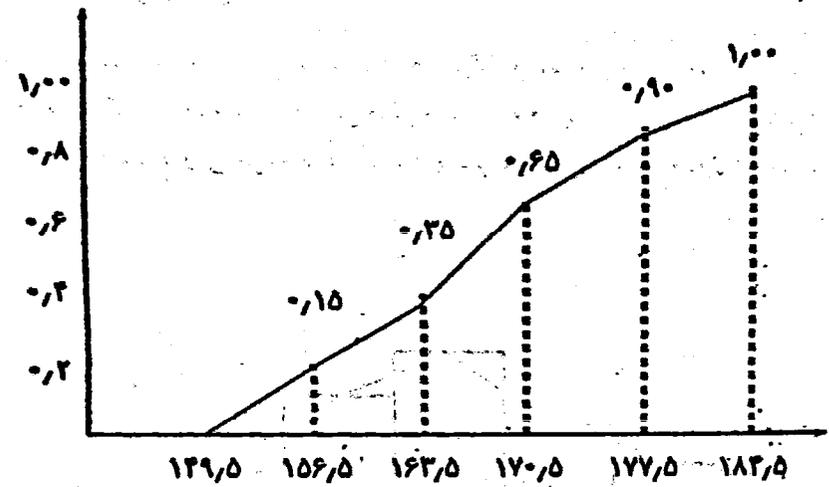
نگاره ۱ هیستوگرام و چندبر فراوانی مربوط به قدها

۲.۴.۱ چندبر فراوانی انباشته

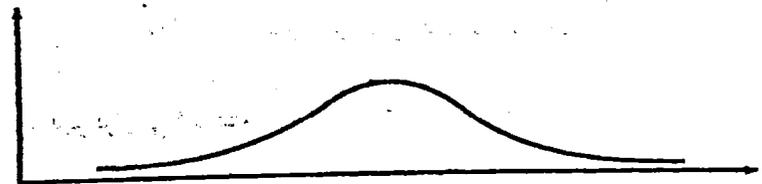
اگر نقاطی را که طول آنها مرز رده‌ها و عرض آنها فراوانی نسبی انباشته تا آن مرز باشد، به هم بپیوندیم یک خط شکسته به دست می‌آید که آن را چندبر فراوانی انباشته می‌نامند. نگاره ۲ چندبر فراوانی انباشته برای جدول ۳ است.

۲.۴.۱ منحنی‌های فراوانی و فراوانی انباشته

اگر تعداد داده‌ها زیاد و طول رده‌ها کوچک و در نتیجه رده‌ها زیاد باشد، چندبر فراوانی و چندبر فراوانی انباشته دارای اضلاع زیاد خواهند شد و می‌توان بر آنها منحنی‌هایی منطبق کرد که به ترتیب منحنی فراوانی و منحنی فراوانی انباشته نامیده می‌شوند. چون مساحت تمام مستطیل‌های هیستوگرام یک واحد مربع می‌باشد، مساحت زیرمنحنی فراوانی را هم یک واحد مربع فرض می‌کنند. منحنی فراوانی انباشته،



نگاره ۲ چندین فراوانی انباشته مربوط به قدام



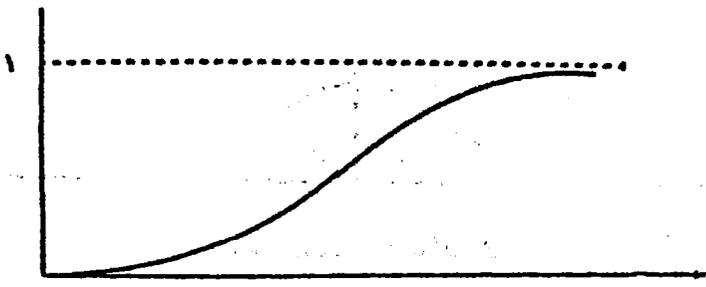
نگاره ۳ منحنی فراوانی

یک منحنی غیر نزولی است که عرضهای تقاط آن بین صفر و یک می باشند. نگاره ۳ و ۴ به ترتیب منحنی فراوانی و منحنی فراوانی انباشته را نشان می دهند.

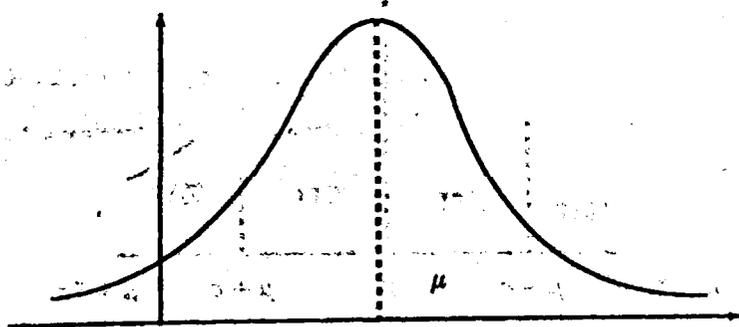
۵.۴.۱ منحنی فراوانی نرمال

اگر منحنی فراوانی دارای معادله مختصاتی

$$y = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$



نگاره ۴ منحنی فراوانی انباشته



نگاره ۵ منحنی نرمال با σ^2 کوچک

باشد، آن را منحنی فراوانی نرمال با پارامترهای μ و σ^2 و متغیر مربوط را متغیر نرمال می نامند. این منحنی نسبت به خط $x = \mu$ متقارن، دارای ماکزیمم در نقطه $(\mu, 1/\sqrt{2\pi\sigma^2})$ ، مجانب یا محور x ها، و زنگ-گونه می باشد. اگر $\mu = 0, \sigma^2 = 1$ باشد، منحنی نرمال را منحنی نرمال استاندارد می نامند. از عرض نقطه ماکزیمم روشن است که هر چه σ^2 کوچکتر باشد، منحنی کشیده تر بوده، قسمت اعظم مساحت زیر منحنی اطراف $x = \mu$ توزیع می شود. نگاره ۵ و ۶ و ۷ نمونه هایی از منحنی فراوانی نرمال هستند. نگاره ۷ منحنی نرمال استاندارد را نشان می دهد.

کلمه نرمال به معنی طبیعی و هادی است. معمولاً بعضی ویژگیهای طبیعی مانند قد و وزن، تقریباً دارای منحنی فراوانی نرمال هستند، یعنی تعداد افراد معمولی (اطراف μ) زیاد و تعداد افراد غیر طبیعی (دور از μ) کم می باشد. نگاره ۷ نحوه توزیع مساحت زیر منحنی نرمال را بر حسب مقیاس σ نسبت به

۶.۴.۱ تمرین بخش چهار

۱. برای داده‌های نرمال با میانگین μ و واریانس σ^2 چند درصد از داده‌ها بیش از $\mu + 2\sigma$ می‌باشد؟ چند درصد کمتر از $\mu - \sigma$ می‌باشد؟
۲. مختصات نقطه ماکزیمم یک منحنی نرمال را از روی معادله مختصاتی آن بدون استفاده از مشتق و با استفاده از آن بیابید؟
۳. برای داده‌های تمرین ۳ از بخش ۳، هیستوگرام و چندبهر فراوانی انباشته را رسم کنید.

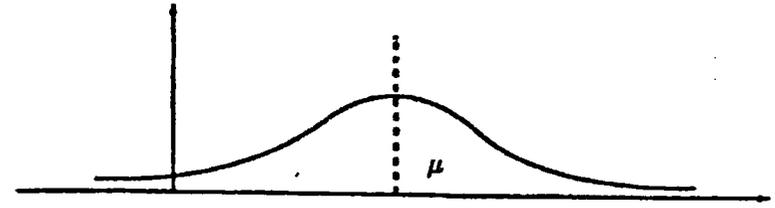
۴. برای داده‌های تمرین ۴ از بخش ۳، هیستوگرام چندبهر فراوانی، و چندبهر فراوانی انباشته را رسم کنید. یک منحنی فراوانی روی چندبهر فراوانی رسم کنید؟
۵. فراوانی رده‌های $۱۳/۵ - ۱۰/۵$ ، $۱۷/۵ - ۱۳/۵$ ، $۱۹/۵ - ۱۷/۵$ ، $۲۴/۵ - ۱۹/۵$ ، که طول آنها نابرابرند، ۵، ۷، ۲ و ۶ می‌باشد. هیستوگرام را طوری رسم کنید تا مساحت مستطیلهای مربوط برابر یک واحد مربع شود. برای این منظور از رابطه زیر استفاده کنید:

طول رده \times عرض مستطیل = فراوانی نسبی رده

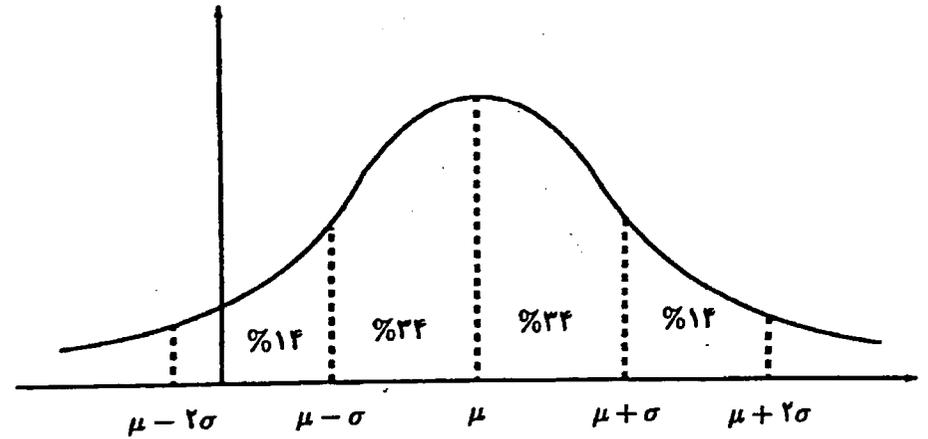
۵.۱ معیارهای تمرکز

با استفاده از جدول فراوانی و نمودارهای آماری، می‌توان تا حدودی دانسته‌های نهفته در داده‌ها را مختصر و محسوس کرد. با این حال سعی می‌شود تا این دانسته‌ها را به صورت یک یا چند عدد معقول درآورد، تا هم بتوان ایده‌ای کلی درباره ویژگی مورد مطالعه به دست آورد و هم نتیجه مطالعات را به سادگی گزارش داد. چنین اعدادی که معمولاً در حوالی مرکز منحنی فراوانی می‌باشند، معیارهای تمرکز نامیده می‌شوند. و در اینجا آنها را شرح می‌دهیم.

فرض می‌کنیم تعداد داده‌ها n و به صورت x_1, x_2, \dots, x_n با فراوانیهای f_1, f_2, \dots, f_n خلاصه شده باشند. در صورتی که داده‌ها پیوسته باشند، آنها را نماینده رده‌ها می‌گیریم.

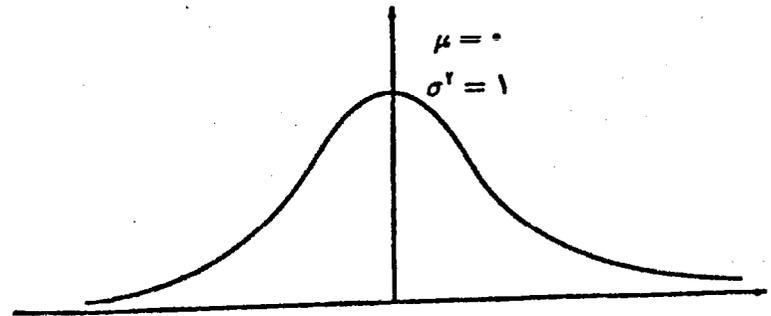


نگاره ۶ منحنی نرمال با σ^2 بزرگ



نگاره ۷ نحوه توزیع مساحت زیر منحنی نرمال

μ نشان می‌دهد.



نگاره ۸ منحنی نرمال استاندارد

مجموع داده‌ها، تقسیم بر تعداد آنها یعنی

$$\bar{x} = \frac{\sum_{i=1}^k f_i x_i}{n} \quad (1)$$

را میانگین داده‌ها می‌گویند. هرگاه تمام فراوانیها برابر یک باشند داریم $k = n$ ، و در این حال \bar{x} را در زبان معمولی معدل می‌نامند.

۲.۵.۱ معدل وزنی

اگر $0 < \omega_i < 1$ برای $i = 1, 2, \dots, k$ و $\sum_{i=1}^k \omega_i = 1$ آنگاه

$$\bar{x}_\omega = \sum_{i=1}^k \omega_i x_i \quad (2)$$

را معدل وزنی اعداد x_1, x_2, \dots, x_k با وزنهای $\omega_1, \omega_2, \dots, \omega_k$ می‌نامند. بنابراین میانگین (۱) هم یک نوع معدل وزنی می‌باشد که در آن $\omega_i = \frac{f_i}{n}$.

۳.۵.۱ میانگین هندسی

در صورتی که x_1, x_2, \dots, x_k همگی مثبت باشند، میانگین هندسی آنها می‌شود

$$G = \sqrt[k]{x_1^{f_1} x_2^{f_2} \dots x_k^{f_k}} \quad (3)$$

معمولاً هرگاه x_i ها از درصدها یا نسبت‌ها تشکیل شده باشند، مثلاً در کارهای اقتصادی یا جمعیت‌شناسی میانگین هندسی به کار می‌برند. برای محاسبه این میانگین آسانتر است که قبلاً لگاریتم آن را حساب کرد. لگاریتم این میانگین برابر است با میانگین حسابی

$$\log x_1, \log x_2, \dots, \log x_k$$

۴.۵.۱ میانگین توافقی

در صورتی که x_1, x_2, \dots, x_k همگی غیر صفر باشند، میانگین توافقی آنها که آن را با H نشان می‌دهند، فرمول

$$\frac{1}{H} = \frac{1}{n} \sum_{i=1}^k \frac{f_i}{x_i} \quad (4)$$

به دست می‌آید. این میانگین که در هینکسنجی و مطالعه شبکه‌های برق به کار می‌رود، در حقیقت برابر است با عکس میانگین حسابی $\frac{1}{x_1}, \frac{1}{x_2}, \dots, \frac{1}{x_k}$.

۵.۵.۱ میانگین ریشه‌ای رتبه دو

$$M_r = \left(\sum_{i=1}^k \frac{f_i x_i^r}{n} \right)^{\frac{1}{r}} \quad (5)$$

تعریف می‌شود، و در حقیقت برابر است با جذر میانگین حسابی

$$x_1^r, x_2^r, \dots, x_k^r$$

می‌توان ثابت کرد که برای داده‌های مثبت میان این چهار نوع میانگین حسابی، هندسی، توافقی و ریشه‌ای رتبه دو، نامساوی زیر برقرار است:

$$H \leq G \leq \bar{x} \leq M_r \quad (6)$$

برای اثبات این نامساوی باید نشان داد که میانگین ریشه‌ای رتبه r یعنی

$$M_r = \left(\sum_{i=1}^k \frac{f_i x_i^r}{n} \right)^{\frac{1}{r}}$$

تابعی صعودی از r است. سپس با قرار دادن r برابر $1, 0$ و -1 در M_r ، نامساوی (۶) به دست می‌آید.

۶.۵.۱ میانه

عدد m را میانه می‌نامند هرگاه تقریباً تعداد نصف داده‌ها از m کوچکتر باشند.

۷.۵.۱ محاسبه میانه برای داده‌های گسسته

فرض کنید n داده گسسته داشته باشیم که به طور غیر نزولی یعنی به صورت $x_1 \leq x_2 \leq \dots \leq x_n$ تنظیم شده باشند. اگر n فرد باشد داده‌ای را که در وسط قرار دارد، میانه می‌گویند. اگر n زوج باشد

جدول ۴ جدول کمکی برای محاسبه میانگینها

| x_i | f_i | $f_i x_i$ | $f_i x_i^2$ | f_i / x_i | $f_i \log x_i$ |
|-------|-------|-----------|-------------|-------------|----------------|
| ۱ | ۳ | ۳ | ۳ | ۳ | ۰/۰۰۰ |
| ۲ | ۵ | ۱۰ | ۲۰ | ۲/۵۰۰ | ۱/۵۰۵ |
| ۳ | ۲ | ۶ | ۱۸ | ۰/۶۶۷ | ۰/۹۵۴ |
| ۵ | ۳ | ۱۵ | ۷۵ | ۰/۶۰۰ | ۲/۰۹۷ |
| مجموع | ۱۳ | ۳۴ | ۱۱۶ | ۶/۷۶۷ | ۴/۵۵۶ |

$\bar{X} = 2/61 \quad G = 2/24 \quad M_2 = 2/99 \quad H = 1/92$

نصف مجموع دو داده‌ای را که در وسط قرار دارند، میانه می‌گویند. میانه را با m نشان می‌دهیم. بنابراین داریم:

برای n فرد $m = x_{\frac{(n+1)}{2}}$

برای n زوج $m = (x_{\frac{n}{2}} + x_{\frac{(n+1)}{2}}) / 2$

مثال ۵ برای پیدا کردن میانه داده‌های

- ۳۱ ۴ ۵ ۲۱ ۲۰ ۳۰ ۱۷ ۸ ۱۲

نخست آنها را به طور غیر نزولی یعنی به صورت زیر می‌نویسیم:

- ۴ ۵ ۸ ۱۲ ۱۷ ۲۰ ۲۱ ۳۰ ۳۱

چون تعداد داده‌ها، عدد فرد $n = 9$ است، میانه می‌شود $m = 17$.

مثال ۶ برای پیدا کردن میانه داده‌های ۲ ۳ ۲ ۱ ۴ ۵ ۳ ۲ صورت ۱ ۲ ۲ ۲ ۳ ۳ ۴ ۵ می‌نویسیم. چون تعداد داده‌ها عدد زوج $n = 8$ است. میانه می‌شود $m = \frac{(2+3)}{2} = 2/5$

مثال ۷ میانه برای داده‌های ۱ ۴ ۵ ۷ به ترتیب با فراوانیهای ۳ ۲ ۲ ۲ می‌شود $m = 5$ یعنی داده‌ها ششم به ترتیب غیر نزولی.

۸.۵.۱ محاسبه میانه برای داده‌های پیوسته

نخست طرز محاسبه را با یک مثال شرح می‌دهیم و سپس فرمول کلی را می‌نویسیم.

مثال ۸ برای داده‌های مربوط به قدهای ۱۰۰ جوان بیست‌ساله، که در جدول فراوانی ۳ در بخش ۳ خلاصه شده‌اند، میانه را طبق دستور زیر محاسبه می‌کنیم:

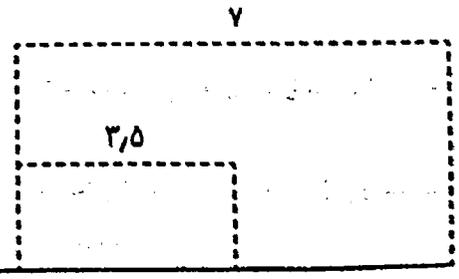
(۱) نخستین رده‌ای را در نظر می‌گیریم که فراوانی انباشته آن برابر نصف یا بیشتر از نصف داده‌ها باشد، یعنی برابر یا بیشتر از $50 = \frac{100}{2}$. این رده را که میانه در آن جا دارد رده میانه می‌نامند و می‌شود $170/5 - 163/5$.

(۲) فراوانی انباشته رده بلافاصله قبل از رده میانه یعنی ۳۵ را از نصف تعداد داده‌ها، یعنی از $50 = \frac{100}{2}$ کم می‌کنیم. داریم $15 = 50 - 35$.

(۳) فرض می‌کنیم در رده میانه، یعنی در $170/5 - 163/5$ که طولش ۷ سانتیمتر است و در آن ۳۰ داده داریم، تمام داده‌ها در فواصل مساوی توزیع شده باشند. با یک تناسب معلوم می‌شود که ۱۵ داده اول به اندازه $\frac{(15 \times 7)}{30} = 3/5$ سانتیمتر از این رده را در بر می‌گیرند.

(۴) عدد $3/5$ را با مرز پائین رده میانه، یعنی با $163/5$ جمع می‌کنیم. میانه می‌شود سانتیمتر $m = 163/5 + 3/5 = 167$.

نگاره زیر طرز محاسبه را نشان می‌دهد.



رده میانه رده بلافاصله قبل از رده میانه

با استفاده از این مثال، فرمول زیر برای محاسبه میانه داده‌های پیوسته به دست می‌آید:

$$m = L_{.15} + \left(\frac{.5n - F_{.15}}{f_{.15}} \right) \cdot h \quad (7)$$

در این فرمول m میانه، $L_{.15}$ مرز پائین رده میانه، n تعداد رده‌ها، $g_{.15}$ فراوانی انباشته رده بلافاصله قبلی از رده میانه، $f_{.15}$ فراوانی رده میانه، و w طول رده می‌باشد.

۹.۵.۱ چندکها

عدد Q_p را که در آن $0 < p < 1$ ، چندک مرتبه p می‌نامند، هرگاه تقریباً $100p$ درصد داده‌ها کوچکتر از آن باشند. مثلاً $Q_{.15}$ را چندک مرتبه 0.15 می‌گویند هرگاه تقریباً ۱۵ درصد داده‌ها کوچکتر از $Q_{.15}$ باشد.

چندکها کلی‌تر از میانه می‌باشند و در حقیقت $Q_{.5}$ همان میانه یعنی m است. به زبان هندسی با استفاده از منحنی فراوانی، اگر از نقطه Q_p خطی به موازات محور y ها رسم کنیم. مساحت زیر منحنی فراوانی که در سمت چپ این خط قرار دارد، برابر p واحد مربع می‌باشد. چندکهای معروف عبارتند از:

(۱) چارکها که به ازای $p = 0.25, 0.50, 0.75$ به دست می‌آیند و آنها را به ترتیب با Q_1, Q_2, Q_3 نشان می‌دهند؛ Q_1 و Q_3 را چارک اول و سوم و Q_2 را میانه می‌گویند. Q_1 را می‌توان میانه برای داده‌هایی پنداشت که کوچکتر یا مساوی m هستند. همچنین Q_3 را می‌توان میانه برای داده‌هایی پنداشت که بزرگتر یا مساوی m هستند.

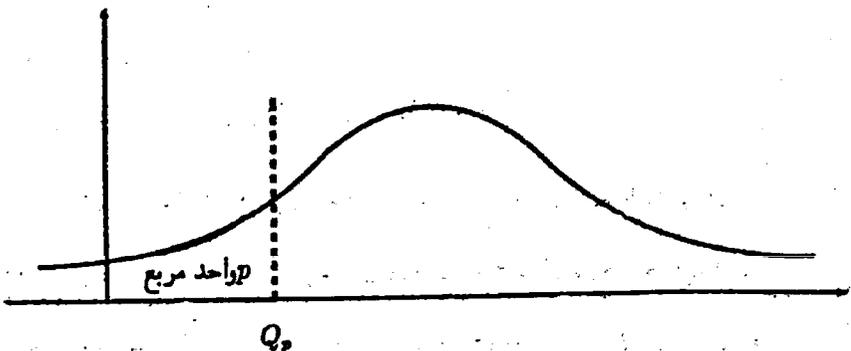
(۲) دهکها که به ازای $p = 0.1, 0.2, \dots, 0.9$ به دست می‌آیند و آنها را با D_1, D_2, \dots, D_9 نشان می‌دهند.

(۳) صدکها که به ازای $p = 0.01, 0.02, \dots, 0.99$ به دست می‌آیند و آنها را با P_1, P_2, \dots, P_{99} نشان می‌دهند.

به کمک این چارکها می‌توان میانگینهای گوناگون تعریف کرد. مثلاً میانگین وزنی زیر را میانگین سه چارکی می‌گویند.

$$i = \frac{Q_1 + 2Q_2 + Q_3}{4}$$

طرز محاسبه چندکها، مانند میانه، برای داده‌های جدا و پیوسته فرق دارند و هر یک را جداگانه شرح می‌دهیم.



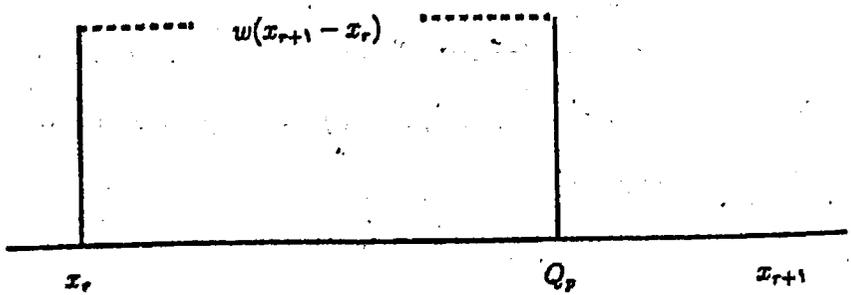
نگاره ۹ منحنی فراوانی و چندک مرتبه p

۱۰.۵.۱ محاسبه چندکها برای داده‌های گسسته

فرض کنید n داده گسسته داشته باشیم که به ترتیب غیر نزولی یعنی به صورت $x_1 \leq x_2 \leq \dots \leq x_n$ تنظیم شده باشند. برای به دست آوردن Q_p با تقریبی خوب بدین طریق عمل می‌کنیم: اگر $(n+1)p$ مساوی عدد صحیح r باشد، فرض می‌کنیم $Q_p = x_r$. در غیر این صورت بزرگترین عدد صحیح در $(n+1)p$ را مساوی r و اختلاف آنرا با r برابر w می‌گیریم. واضح است که $0 \leq w < 1$ ، حال $\omega(x_{r+1} - x_r)$ ، یعنی کسری از طول فاصله (x_r, x_{r+1}) ، را به x_r اضافه می‌کنیم. در نتیجه Q_p به صورت زیر می‌شود (به نگاره زیر توجه کنید)

$$Q_p = (1-\omega)x_r + \omega x_{r+1} \quad (۸)$$

ملاحظه می‌شود که Q_p معدل وزنی x_r و x_{r+1} است. میانه را هم می‌توانیم از این فرمول پیدا کنیم.



مثال ۹ برای داده‌های ۱۲ ۸ ۷ ۷ ۵ ۴ ۳ ۲ ۱ که به ترتیب غیر نزولی مرتب شدند $Q_{.16}$ را حساب کنید

چون $n=11$, $p=0.16$, $(n+1)p=7.12$, $r=7$, $w=0.2$

$Q_{.16} = D_6 = P_6 = 0.18x_7 + 0.18x_8 = 0.18(5) + 0.18(7) = 6.12$

مثال ۱۰ برای داده‌های خلاصه شده در جدول ۲، از بخش ۳ چارک سوم را حساب کنید

چون $n=200$, $p=0.75$, $(n+1)p=150.75$, $r=150$

و $w=0.75$ داریم

$Q_3 = 0.25x_{150} + 0.75x_{151} = 0.25(4) + 0.75(4) = 4$

۱۱.۵.۱ محاسبه چندکها برای داده‌های پیوسته

نخست داده‌ها را در یک جدول فراوانی تنظیم می‌کنیم، و سپس رده‌ای را که Q_p در آن جا دارد و رده Q_p می‌نامیم، مشخص می‌سازیم. اینک فرمول زیر را که ترکیب و نحوه پیدا کردن آن عیناً مانند فرمول (۷)، که مربوط به میانه می‌باشد، به دست می‌آوریم.

$Q_p = L_p + \left(\frac{pn - G_p}{f_p}\right)w$ (۹)

در این فرمول n تعداد داده‌ها، L_p مرز پائین رده Q_p ، یعنی نخستین رده‌ای که فراوانی انباشته آن برابر یا بیشتر از pn است، G_p فراوانی انباشته رده بلافاصله قبل از رده Q_p ، f_p فراوانی رده Q_p و w طول رده می‌باشد.

مثال ۱۱ برای داده‌های خلاصه شده در جدول ۳ از بخش ۳ دمک هفتم را حساب کنید

چون $n=100$, $p=0.7$, $pn=70$, رده D_7 می‌شود $177.5 - 170.5$. حال با استفاده از فرمول (۹) داریم

$D_7 = 170.5 + \left(\frac{70 - 65}{25}\right)7 = 171.9$

۱۲.۵.۱ نما

داده‌ای که فراوانی آن نسبت به سایر داده‌ها بیشتر باشد، نما یا مد نامیده می‌شود و آنرا با M نشان می‌دهند. کلمه مد در زبان عادی هم همین مفهوم را دارد مثلاً وقتی می‌گوئیم فلان لباس زنانه مد می‌باشد، منظور این است که اغلب خانمها آنرا می‌پوشند.

۱۳.۵.۱ محاسبه نما برای داده‌های گسسته

نخست فراوانی داده‌ها را پیدا می‌کنیم. داده‌ای را که فراوانی آن بیشتر باشد به عنوان نما اختیار می‌کنیم. اگر دو داده، دارای فراوانی‌های مساوی، بیش از سایر فراوانی‌ها، باشد هر دو را به عنوان نما اختیار می‌کنیم، و داده‌ها را دو نمائی می‌گوئیم، مشروط به اینکه این دو داده پهلوی هم نباشند. اگر پهلوی هم باشند، نصف مجموع آنها را نما می‌خوانیم. اگر تمام داده‌ها دارای فراوانی‌های مساوی باشند، می‌گوئیم داده‌ها بدون نما هستند.

مثال ۱۲ برای داده‌های ۱۳ ۱۳ ۱۳ ۱۳ ۱۳ ۲ ۳ ۴ ۵ ۳ ۳ ۱ نما برابر است با $M=3$ ، زیرا فراوانی داده ۳ بیش از سایر داده‌ها می‌باشد.

مثال ۱۳ برای داده‌های ۵ ۳ ۳ ۴ ۱ ۲ ۳ ۱ ۱ دو داده ۱ و ۳ که پهلوی هم نیستند و فراوانی مشترک آنها بیش از سایر فراوانیها می‌باشد، هر دو، نما اختیار می‌شوند. این داده‌ها را دو نمائی می‌گوئیم.

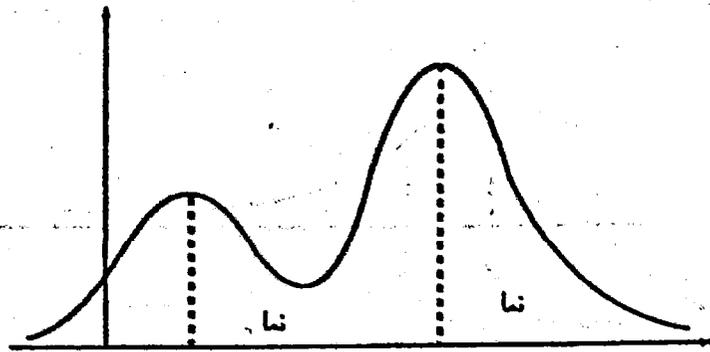
مثال ۱۴ برای داده‌های ۱ ۲ ۱ ۳ ۲ ۲ ۴ ۵ ۲ ۱ ۳ ۱ ۱ ۲ ۳ نصف دو داده ۱ و ۲ را که پهلوی هم هستند و دارای فراوانیهای ۴، یعنی بیش از سایر فراوانی‌ها، می‌باشند به عنوان نما اختیار می‌کنیم. بنابراین $M=1/5$.

مثال ۱۵ داده‌های ۲ ۱ ۳ ۲ ۵ ۳ ۱ ۳ ۵ ۱ نما ندارند، زیرا فراوانی تمام داده‌ها مساوی هستند.

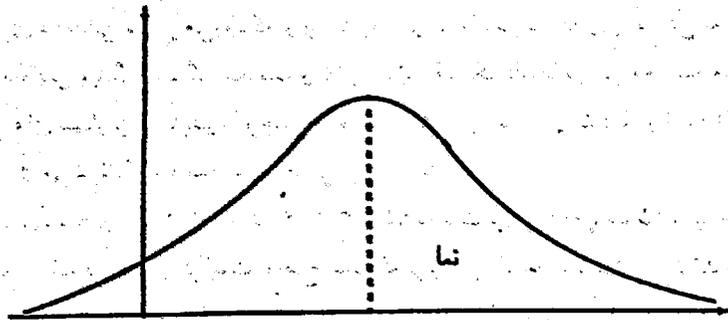
۱۴.۵.۱ محاسبه نما برای داده‌های پیوسته

نخست داده‌ها را در یک جدول فراوانی خلاصه می‌کنیم، و بعد نماینده رده‌ای را که دارای فراوانی بیشتر می‌باشد، و رده نمائی نامیده می‌شود، به عنوان نما اختیار می‌کنیم.

مثال ۱۶ برای داده‌های خلاصه شده در جدول ۳ از بخش ۳ نما را پیدا کنید.



نگاره ۱۱ منحنی فراوانی دونمائی



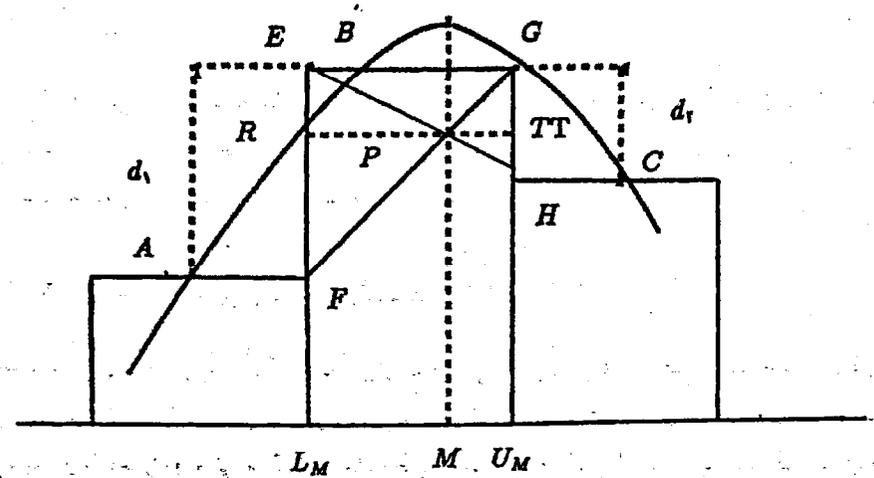
نگاره ۱۲ منحنی فراوانی یک نمائی

$L_M = ۱۶۳/۵$ ، $d_1 = ۰/۱۰$ ، $d_2 = ۰/۰۵$ و $w = ۷$. از فرمول (۱۰) داریم

$$M = ۱۶۳/۵ + \frac{۰/۱۰}{۰/۱۰ + ۰/۰۵} \times ۷$$

$$= ۱۶/۸۲$$

اگر دو رده که پهلوی هم نیستند دارای فراوانی‌های مساوی، بیش از سایر فراوانیها باشند، داده‌ها را دو نمائی می‌خوانیم. در نگاره‌های ۱۱ و ۱۲ منحنی فراوانی دو نمائی و یک نمائی را مشاهده می‌کنید.



نگاره ۱۰ طرز محاسبه نما برای داده‌های پیوسته

از روی جدول ملاحظه می‌شود که فراوانی رده $۱۷۰/۵ - ۱۶۳/۵$ ، بیش از سایر فراوانیها است. بنابراین $M = ۱۶۷$ یعنی نماینده این رده را به عنوان نما اختیار می‌کنیم.

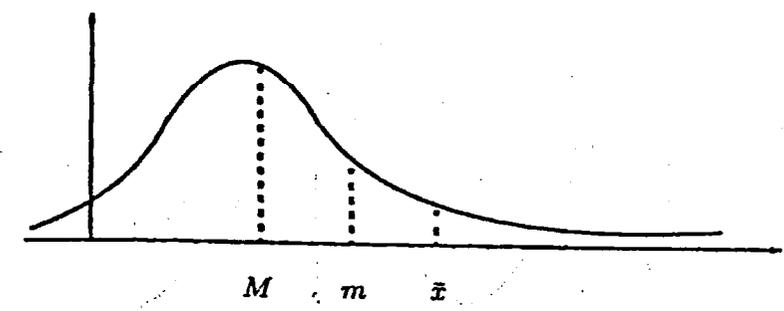
برای دقت بیشتر می‌توان نما را از فرمول

$$M = L_M + \frac{d_1}{d_1 + d_2} w \tag{۱۰}$$

به دست آورد. در این فرمول M نما، L_M مرز پائین رده نمائی، d_1 اختلاف فراوانی‌های نسبی رده نمائی و رده بلافاصله قبل از آن، d_2 اختلاف فراوانی‌های نسبی رده نمائی و رده بلافاصله بعد از آن و w طول رده می‌باشد. نگاره ۱۰ دلیل هندسی فرمول بالا را از روی هیستوگرام روشن می‌کند.

در این شکل نما را طول نقطه ماکزیمم یک سهمی فرض می‌کنیم که از وسط قاعده‌های بالائی سه مستطیل هیستوگرام می‌گذرد. می‌توان ثابت کرد که طول این نقطه ماکزیمم، با طول نقطه برخورد EH و FG یکی است. از دو مثلث متشابه EPF و GPH می‌توان اندازه RP و در نتیجه نما را به دست آورد.

مثال ۱۷ برای داده‌های خلاصه شده، در جدول ۳ از بخش ۳ نما را به کمک فرمول (۱۰) پیدا کنید. در این مثال رده $۱۶۳/۵ - ۱۷۰/۵$ رده نمائی است. فراوانی نسبی این رده $۰/۳۰$ و فراوانی‌های نسبی رده‌های بلافاصله قبل از آن و بعد از آن به ترتیب $۰/۲۰$ و $۰/۲۵$ می‌باشند. بنابراین،



نگاره ۱۳ منحنی فراوانی چوله به راست

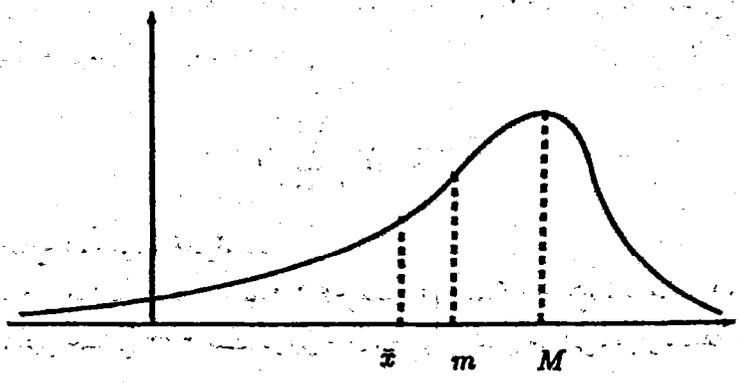
۱۵.۵.۱ مقایسه میانگین، میانه و نما

میانگین برای داده‌هایی به کار می‌رود که از راه اندازه‌گیری یا مقیاس فاصله‌ای یا نسبی به دست آمده باشند. از محاسن میانگین سادگی محاسبه و تاثیر اندازه یک یک داده‌ها در این محاسبه می‌باشد. با اینحال داده‌های بسیار بزرگ یا بسیار کوچک، به نام داده‌های دور افتاده یا پرت، بر میانگین اثر نامطلوب دارند و باعث می‌شوند که میانگین معیار خوبی برای تمرکز نباشد.

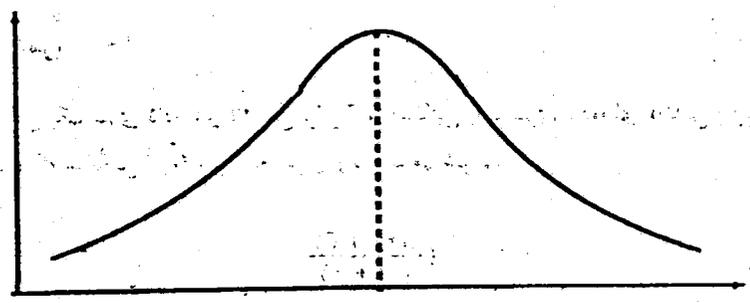
مثلاً برای داده‌های ۲ ۳ ۴ ۵ ۶ ۸ ۹ ۱۳ ۸۵ به علت بزرگی داده پرت ۸۵، میانگین می‌شود عدد ۱۵ که از بقیه داده‌ها بزرگتر است. واضح است که برای این داده‌ها عدد ۱۵ نمی‌تواند معیار تمرکز خوبی باشد، زیرا با اکثر داده‌ها تفاوت زیادی دارد. برای این داده‌ها، میانه عدد ۶ می‌باشد و معیار بهتری برای تمرکز است. ضمناً اگر داده‌ها را از راه اندازه‌گیری با مقیاس‌های اسمی یا ترتیبی به دست آمده باشند، باید به جای میانگین، میانه و نما را به کار برد. اگر منحنی فراوانی چوله به راست باشد، میانگین و میانه و نما به ترتیب روی محور طول‌ها از راست به چپ، و اگر منحنی فراوانی چوله به چپ باشد، آنها از چپ به راست قرار می‌گیرند. برای منحنی فراوانی متقارن، این سه معیار بر هم منطبق می‌شوند. به نگاره‌های ۱۳ و ۱۴ و ۱۵ توجه شود. هرگاه میزان چولگی خفیف باشد، بین میانگین و میانه و نما، رابطه تقریبی زیر برقرار می‌باشد:

$$\bar{x} - M \approx 3(\bar{x} - m) \tag{11}$$

اگر از M و m و \bar{x} ، سه خط موازی محور Y ها رسم کنیم، از نظر هندسی خطی که از M رسم می‌شود از نقطه ماکزیم منحنی فراوانی می‌گذرد، خطی که از m رسم می‌شود مساحت زیر منحنی فراوانی را نصف می‌کند و خطی که از \bar{x} رسم می‌شود محور تعادل منحنی فراوانی را مشخص می‌سازد.



نگاره ۱۴ منحنی فراوانی چوله به چپ



نگاره ۱۵ متقارن

۱۶.۵.۱ میانگین اصلاح شده

میانگین حسابی ساده‌ترین معیار تمرکز برای داده‌ها است. ولی همانطوری که گفتیم داده‌های پرت بر آن اثر نامطلوب دارند، یعنی آن را از مرکز داده‌ها دور می‌کنند. برای اصلاح این امر، داده‌های پرت را حذف می‌کنند و سپس میانگین بقیه داده‌ها را به عنوان معیار تمرکز به کار می‌برند. چنین میانگینی را میانگین اصلاح شده می‌نامند. میانگین اصلاح شده در واقع یک نوع میانگین وزنی است که بدین نحو تعریف می‌شود:

فرض کنید داده‌ها را به طریق غیر نزولی به صورت $x_1 \geq x_2 \geq \dots \geq x_n$ مرتب کرده باشیم. k تا از

کوچکترین و k تا از بزرگترین داده‌ها را حذف می‌کنیم. انتخاب $k < \frac{n}{4}$ بستگی به داده‌هایی دارد که برت به نظر می‌آیند. حال میانگین بقیه داده‌ها یعنی

$$T_k = \frac{1}{n-2k} \sum_{i=k+1}^{n-k} x_i$$

را میانگین اصلاح شده مرتبه k می‌گویند.

برای داده‌های مرتب شده ۱۵ ۷ ۴ ۱ ۶ - عدد ۴/۲۵ میانگین اصلاح شده مرتبه ۱ می‌باشد، در حالی که میانگین حسابی ۴/۲۳ است. در این مثال، ظاهراً میانگین حسابی، مرکز داده‌ها را بهتر نشان می‌دهد.

برای n های زوج میانه در حقیقت برابر میانگین اصلاح شده مرتبه $\frac{n-2}{4}$ و برای n های فرد میانه برابر میانگین اصلاح شده مرتبه $\frac{n-1}{4}$ است.

۱۷.۵.۱ تمرین بخش پنج

۱. اگر میانگین یک سری داده‌های m تایی برابر \bar{x} و میانگین یک سری داده‌های n تایی برابر \bar{y} باشد، ثابت کنید که میانگین آمیخته این دو سری از داده‌ها می‌شود.

$$\frac{(m\bar{x} + n\bar{y})}{(m + n)}$$

مثال عددی - میانگین مزد ۱۰ کارگر ۳۵۰۰ ریال و میانگین مزد ۱۵ کارگر دیگر ۴۵۰۰ ریال می‌باشد میانگین مزد این ۲۵ کارگر را حساب کنید.

۲. نشان دهید که برای داده‌های

۱۵ ۱۹ ۱۶ ۱۱ ۱۵ ۱۱ ۱۲ ۳ ۸ ۷ ۳ ۱ ۷ ۷ ۵

میانگین حسابی، هندسی، توافق، و ریشه‌ای رتبه دو، به ترتیب برابر ۹/۳، ۷/۵، ۵/۱، ۱۰/۶ می‌باشد. برای این داده‌ها میانه، چارک اول، دهک سو، و نما را حساب کنید.

۳. برای داده‌هایی که در جدول فراوانی ۳ خلاصه شده است میانگین، میانه، و نما را حساب کنید و بوسیله فرمول (۱۱) محاسبات را کنترل نمایید.

۴. برای داده‌هایی که در جدول فراوانی ۳ خلاصه شده‌اند، دهک چهارم را حساب کنید

۵. از فرمول (۸) استفاده کرده و فرمولهایی برای میانه و چارک اول پیدا کنید.

۶. از رابطه میان میانگینها را (فرمول ۶) برای داده‌های مثبت ۱، ۲، ۳، ۴، ۵، ۶، ۷، ۸، ۹، ۱۰ ثابت کنید

۷. ثابت کنید M_r تابعی صعودی از r است و سپس نامساوی (۶) را به دست آورید.

۸. میانگین اصلاح شده مرتبه ۲ را برای داده‌های زیر پیدا کنید و آن را با میانگین حسابی مقایسه نمایید.

۱۹ ۱۵ ۵ ۳ ۱ ۸ -۹

۹. میانگین سه - چارکی و میانگین اصلاح شده مرتبه یک را برای داده‌های تمرین ۲ بیابید.

۶.۱ معیارهای پراکندگی

معمولاً افراد یک جمعیت از نظر ویژگی مورد مطالعه با هم تفاوت دارند، و این خود مهمترین انگیزه برای پژوهشگری و کاربرد فن آمار می‌باشد. مشاهده تفاوت و تنوع، انسان را به کنجکاوی درباره رموز طبیعت امور اجتماعی، تربیتی و صنعتی وادار می‌کند. مثلاً نوزادانی که امروز در یک بیمارستان به دنیا می‌آیند از لحاظ گروه خونی، وزن، درجه حرارت بدن، رنگ و غیره با هم تفاوت دارند. دانش‌آموزان یک کلاس از نظر ادب و استعداد با هم فرق دارند. لامپ‌های ۶۰ وات اسرام تحت شرایط مساوی، از نظر طول عمر یکسان نیستند. میزان این تفاوتها را چگونه می‌شود سنجید، و چطور می‌توان درباره آن قضاوت کرد؟

همان طوری که قبلاً دیدیم داده‌ها را معمولاً به صورت یک عدد به نام معیار تمرکز خلاصه می‌کنند، و قسمتی از اطلاعات موجود در آنها را در این عدد منعکس می‌سازند. ولی لازم است درباره تفاوت داده‌ها با یکدیگر و میزان پراکندگی و تجمع آنها به نحوی مطالعه کرد.

مثال ۱۸ فرض کنید اعداد زیر نمره‌های دو کلاس در یک آزمون ریاضی باشند

کلاس یک: ۱۰۰ ۱۰۰ ۷۰ ۶۳ ۶۲ ۱۰ ۱۲ ۶ ۰
کلاس دو: ۲۹ ۲۱ ۲۰ ۲۳ ۶۲ ۵۱ ۵۰ ۲۲ ۲۹ ۲۸ ۴۰

با وجود اینکه میانگین برای هر دو کلاس ۲۷ می‌باشد، نحوه پراکندگی داده‌ها به دور ۲۷ و نحوه تغییرپذیری آنها نسبت به ۲۷ در این دو کلاس کاملاً متفاوت می‌باشد. در کلاس یک داده‌ها با ۲۷ تفاوت فاحش دارند، در حالی که در کلاس دو داده‌ها اطراف ۲۷ متمرکز شده‌اند به عبارت دیگر در کلاس یک میزان تغییرپذیری داده‌ها شدید و در کلاس دو ضعیف می‌باشد.

اینک چند معیار مفید را برای سنجش پراکندگی شرح می‌دهیم. فرض کنید x_1, x_2, \dots, x_k یک سری داده‌های n تایی یا فراوانی‌های k_1, k_2, \dots, k_k را میانگین \bar{x} باشد

۱.۶.۱ برد

اگر (۱) کوچکترین و (۲) بزرگترین داده باشد،

$$R = x_{(2)} - x_{(1)} \quad (12)$$

را برد داده‌ها می‌گویند.

مثال ۱۹ در مثال ۱۸ برد برای کلاسهای یک و دو به ترتیب عبارتند از:

$$R_2 = 62 - 40 = 22, \quad R_1 = 100 - 0 = 100$$

با وجودی که این معیار وسعت پراکندگی را منعکس می‌کند و طرز محاسبه آن ساده می‌باشد، بیانگر خوبی برای تغییر پذیری داده‌ها نیست. زیرا این معیار فقط به بزرگترین و کوچکترین داده بستگی دارد و اگر بقیه داده‌ها تغییر کنند، تاثیری در برد ایجاد نخواهد شد. اگر تعداد داده‌ها زیاد شود، برد بزرگ شده و یا ثابت می‌ماند.

$$R' = \frac{x_{(1)} + x_{(n)}}{2} \quad (13)$$

توجه کنید که میان برد یک نوع معیار تمرکز است. برای مثال ۱۸ میان برد نمره کلاسهای یک و دو به ترتیب ۵۰ و ۵۱ می‌باشند.

۲.۶.۱ میانگین انحرافها

قدر مطلق \bar{x} - بر انحراف از میانگین برای داده x_i و

$$d = \frac{\sum_{i=1}^k f_i |x_i - \bar{x}|}{n} \quad (14)$$

را میانگین انحرافها می‌نامند. واضح است که هر قدر داده‌ها از \bar{x} دورتر باشند، d بزرگتر خواهد بود. لازم است اضافه کنیم که $x_i - \bar{x}$ ممکن است مثبت یا منفی یا صفر باشد و با استفاده از تعریف d همواره

داریم:

$$\sum_{i=1}^k f_i (x_i - \bar{x}) = 0 \quad (15)$$

اگر میانه معیار بهتری برای تمرکز باشد، می‌توان آن را به جای میانگین در فرمول (۱۲) به کار برد.

مثال ۲۰ در مثال ۱۸ میانگین انحرافها برای کلاسهای یک و دو به ترتیب عبارتند از:

$$d_2 = 2/91, \quad d_1 = 25/56$$

اگر به جای میانگین، میانه را، که برای کلاسهای یک و دو به ترتیب $m_1 = 62$ و $m_2 = 48$ می‌باشند به کار ببریم، داریم:

$$d'_2 = 2/09, \quad d'_1 = 33/88$$

میانگین انحرافها، معیار خوبی برای میزان پراکندگی داده‌ها می‌باشد، ولی طرز محاسبه و کشف خواص ریاضی آن به علت وجود قدر مطلق، قدری پیچیده است. بنابراین به جای آن معیار دیگری به نام واریانس و یا جذر آن به نام انحراف استاندارد را به کار می‌برند.

۳.۶.۱ واریانس و انحراف استاندارد

واریانس را که دو لغت به معنی تفاوت و تغییر است، از فرمول

$$s^2 = \sum_{i=1}^k \frac{f_i (x_i - \bar{x})^2}{n} \quad (16)$$

که میانگین مجذور انحرافها می‌باشد، به دست می‌آورند. اگر تمام داده‌ها به \bar{x} نزدیک باشند، s^2 کوچک می‌شود. به ویژه اگر تمام x_i ها با \bar{x} برابر باشند s^2 صفر است. این حقایق نشان می‌دهند که s^2 معیار خوبی برای سنجش پراکندگی و تغییر پذیری داده‌ها نسبت به میانگین می‌باشد.

با استفاده از جبر مقدماتی، فرمول (۱۶) به صورت زیر که از نظر محاسبه ساده تر است، در می‌آید.

$$s^2 = \frac{1}{n} \sum_{i=1}^k f_i x_i^2 - \bar{x}^2 \quad (17)$$

اگر جمله اول فرمول بالا را که میانگین $x_1^2, x_2^2, \dots, x_k^2$ می‌باشد، با \bar{x}^2 نشان دهیم، داریم:

$$s^2 = \bar{x}^T - \bar{x}^2 \quad (18)$$

بنابراین واریانس برابر است با میانگین توان دوم داده‌ها منهای توان دوم میانگین داده‌ها. چون واریانس یک معیار دو بعدی است (زیرا توان دوم به کار می‌بریم)، از اینرو جنر مثبت فن یعنی s راه که انحراف استاندارد نامیده می‌شود به کار می‌برند. گاهی واریانس را از فرمول

$$s_u^2 = \frac{1}{n-1} \sum_{i=1}^k f_i (x_i - \bar{x})^2 \quad (19)$$

به دست می‌آورند، یعنی مجموع توان دوم انحراف را به جای n ، بر $n-1$ تقسیم می‌کنند. دلیل آماری این کار در آینده روشن خواهد شد. واضح است که میان s_u^2 و s^2 رابطه

$$s_u^2 = \frac{n}{n-1} s^2 \quad (20)$$

برقرار است. این رابطه نشان می‌دهد که برای نمونه‌های بزرگ، مثلاً نمونه‌هایی با $n > 100$ ، s_u^2 و s^2 تقریباً یا هم برابرند.

مثال ۲۱ برای داده‌های ۹ ۳ ۰ ۷ ۱۰ واریانس و انحراف استاندارد را حساب کنید.

یا استفاده از فرمولهای (۱۸) و (۲۰) داریم

$$s = \sqrt{12} = 3.46 \quad s^2 = x^2 = \bar{x}^2 = 28 - 16 = 12$$

$$s_u = \sqrt{15} = 3.87 \quad s_u^2 = \frac{5}{4} \times 12 = 15$$

۴.۶.۱ روش تبدیل یا روش کوتاه برای محاسبه میانگین و واریانس

گاهی برای محاسبه میانگین و واریانس، می‌توان داده‌ها را به داده‌های دیگری تبدیل کرده و سپس محاسبات را انجام داد. مثلاً برای محاسبه میانگین داده‌های ۱۲۰ ۱۲۵ ۱۳۰ ۱۰۵ نخست از هر یک عدد ۱۰۰ را کم می‌کنیم و سپس نتایج را بر ۵ تقسیم می‌کنیم تا داده‌های جدید ۴ ۵ ۳ ۰ به دست آیند. میانگین داده‌های اخیر می‌شود ۴، حال ۴ را در ۵ ضرب کرده، نتیجه را با ۱۰۰ جمع می‌کنیم تا ۱۲۰ یعنی میانگین داده‌های بالا پیدا شود. این روش را روش تبدیل یا روش کوتاه می‌نامند.

۵.۶.۱ داده‌های تبدیل شده

فرض کنید x_1, x_2, \dots, x_k با فرآیندهای $k, k-1, \dots, 1$ یک سری داده n تایی، و a و $b > 0$ دو عدد مناسب باشند. عدد a برای تغییر مبدأ اندازه‌گیری و عدد b را برای تغییر واحد اندازه‌گیری به کار

می‌برند. اینک

$$y_i = \frac{x_i - a}{b} \quad i = 1, 2, \dots, k \quad (21)$$

را داده تبدیل شده می‌نامند. بنابراین داده‌های تبدیل شده می‌شوند y_1, y_2, \dots, y_k با فرآیندهای $k, k-1, \dots, 1$.

اگر \bar{x} و s_x^2 میانگین و واریانس داده‌های تبدیل شده باشند، به سادگی می‌توان نشان داد که

$$\bar{x} = b\bar{y} + a \quad s_x^2 = b^2 s_y^2 \quad s_x = b s_y \quad (22)$$

در عمل نخست میانگین و واریانس را برای داده‌های تبدیل شده و سپس به کمک روابط بالا برای داده‌های اصلی محاسبه می‌کنیم.

مثال ۲۲ برای داده‌هایی که در جدول ۳ خلاصه شده‌اند، میانگین و واریانس را با روش تبدیل محاسبه کنید.

نماینده‌های رده‌ها یعنی ۱۸۱ ۱۷۴ ۱۶۷ ۱۶۰ ۱۵۳ به ترتیب با فرآیندهای ۱۰ ۲۵ ۳۰ ۲۰ ۱۵ در حکم y ها می‌باشند. برای اینکه داده‌های تبدیل شده را به دست آوریم، فرض می‌کنیم $a = 167$ و $b = 7$ باشند. حال y ها را به کمک

$$y_i = \frac{x_i - 167}{7}$$

پیدا می‌کنیم. با توجه به جدول ۵ و فرمولهای (۲۲) میانگین و انحراف استاندارد محاسبه می‌شوند.

$$\bar{y} = -\frac{5}{100} = -0.05$$

$$s_y^2 = \bar{y}^2 - y^2 = \frac{125}{100} = 0.0025 = 1/400$$

$$\bar{x} = 7(-0.05) + 167 = 166.65$$

$$s_x^2 = 7^2 (1/400) = 7/0.93$$

$$s_x = 8/22$$

جدول ۵ جدول محاسبه میانگین و انحراف استاندارد با روش گرنه

| x_i | f_i | y_i | y_i' | $f_i y_i$ | $f_i y_i'$ |
|-------|-------|-------|--------|-----------|------------|
| ۱۵۳ | ۱۵ | -۲ | ۲ | -۳۰ | ۳۰ |
| ۱۶۰ | ۲۰ | -۱ | ۱ | -۲۰ | ۲۰ |
| ۱۶۷ | ۳۰ | ۰ | ۰ | ۰ | ۰ |
| ۱۷۲ | ۲۵ | ۱ | ۱ | ۲۵ | ۲۵ |
| ۱۸۱ | ۱۰ | ۲ | ۲ | ۲۰ | ۲۰ |
| | ۱۰۰ | | | -۵ | ۱۲۵ |

اگر نمره‌های دو آزمون تقریباً دارای منحنی فراوانی نرمال باشند، تنها بعد از استاندارد کردن می‌توان آنها را با هم مقایسه کرد.

نمره استاندارد ریاضی $z' = \frac{60 - 72}{15} = -0.8$

نمره استاندارد فیزیک: $z'' = \frac{25 - 50}{20} = -0.75$

بنابراین علی در فیزیک بهتر می‌باشد، زیرا داریم $z'' > z'$.

۷.۶.۱ ضریب تغییر

نسبت انحراف استاندارد به میانگین، یعنی

$$V = \frac{S}{\bar{x}} \quad (22)$$

را که اغلب به صورت درصد بیان می‌شود، ضریب تغییر می‌نامند. این ضریب که به واحد اندازه‌گیری بستگی ندارد، در عمل برای مقایسه به کار می‌رود.

مثال ۲۴ کارخانه‌ای دو نوع لاستیک اتومبیل تولید می‌کند. برای نوع A میانگین عمر ۱۰۰۰۰ کیلومتر، با انحراف استاندارد ۲۰۰۰ کیلومتر، و برای نوع B میانگین عمر ۱۱۰۰۰ کیلومتر یا انحراف استاندارد ۱۰۰۰ کیلومتر می‌باشد. کدام نوع لاستیک بهتر است؟

ضریب تغییر برای نوع A $V_1 = \frac{2000}{10000} = 0.20 \quad V_1 = 20\%$

ضریب تغییر برای نوع B $V_2 = \frac{1000}{11000} = 0.09 \quad V_2 = 9\%$

نوع B بهتر است، زیرا هم میانگین عمر آن بیشتر است و هم ضریب تغییر آن کوچکتر. در این مثال اگر فرض کنیم که طول عمرها متغیرهای نرمال هستند، با محاسبه فاصله

$(\bar{x} - 2s, \bar{x} + 2s)$

برای هر دو نوع لاستیک، معلوم می‌شود که در حدود ۹۶٪ از لاستیک نوع A دارای طول عمری در فاصله (۱۴۰۰۰ و ۶۰۰۰)، و در حدود ۹۶٪ از لاستیک نوع B، دارای طول عمری در فاصله (۱۳۰۰۰ و ۹۰۰۰) می‌باشند. باز هم نتیجه می‌شود که لاستیک نوع B بهتر است.

۶.۶.۱ داده‌های استاندارد

فرض کنید x_1, x_2, \dots, x_k به ترتیب با فراوانیهای f_1, f_2, \dots, f_k از یک سری داده n تایی با میانگین \bar{x} و انحراف استاندارد s باشند. با تبدیل زیر

$$z_i = \frac{x_i - \bar{x}}{s} \quad i = 1, 2, \dots, k \quad (23)$$

داده‌های جدید z_1, z_2, \dots, z_k به ترتیب با فراوانیهای f_1, f_2, \dots, f_k که به نام داده‌های استاندارد به دست می‌آیند. به ویژه اگر z ها نمره‌های آزمون یک کلاس درس باشند، z ها را نمره‌های استاندارد می‌نامند. به سادگی می‌توان نشان داد که برای این داده‌ها میانگین برابر صفر و واریانس برابر یک است. از اینرو کلمه استاندارد را در مورد آنها به کار می‌برند. اگر متغیرها نرمال باشند، بعد از استاندارد شدن بهتر می‌توان آنها را با هم مقایسه و درصدهای لازم را محاسبه کرد. برای یک متغیر نرمال، z و z^2 به ترتیب تقریب‌های خوبی برای پارامترهای منحنی فراوانی یعنی μ و σ^2 می‌باشند.

برای اینکه کاربرد میانگین و انحراف استاندارد، و همچنین داده‌های استاندارد روشن شود به ذکر مثال زیر می‌پردازیم:

مثال ۲۳ نمره دانش‌آموزان یک کلاس در آزمون ریاضی، دارای میانگین ۷۲ و انحراف استاندارد ۱۵ و در آزمون فیزیک، دارای میانگین ۵۰ و انحراف استاندارد ۲۰ می‌باشند. اگر نمره علی در ریاضی ۶۰ و در فیزیک ۳۵ باشد، معلومات علی در کدام موضوع بیشتر است؟ چون این دو آزمون با مقیاس‌های مختلف به عمل آمده‌اند، مقایسه اعداد ۶۰ و ۳۵ مفهومی ندارد.

۸.۶.۱ نیم برد چارکها

نصف برد چارکهای اول و سوم، یعنی

$$Q = \frac{Q_3 - Q_1}{2} \quad (25)$$

را نیم برد چارکها می نامند. در اینجا اضافه می کنیم که

$$Q' = \frac{Q_1 + Q_3}{2} \quad (26)$$

را میان چارکی می نامند. توجه کنید که میان چارکی یک نوع معیار تمرکز است. برای توزیعهای متقارن Q' درست برابر Q_2 ، یعنی میانه می باشد. معیار Q که در حقیقت برابر فاصله Q' از Q_1 یا Q_3 است، یک نوع معیار برای سنجش پراکندگی است. به ویژه Q را در مواردی که داده های نهائی نامعلوم یا پرت، یعنی بی اندازه کوچک یا بی اندازه بزرگ باشند، به کار می برند.

مثال ۲۵ تعداد فرزندان ۱۶ خانواده به ترتیب غیر نزولی عبارتند از:

۰ ۱ ۱ ۲ ۲ ۲ ۳ ۳ ۳ ۳ ۴ ۴ ۴ ۵ ۶ ۷

برای این داده ها $Q_1 = 2$ ، $Q_2 = 3$ ، $Q_3 = 4$ و می باشد. بنابراین میان چارکی و نیم برد چارکها عبارتند از $Q' = 3$ و $Q = 1$. صرف نظر از داده های نهائی ملاحظه می شود که Q' تقریباً مرکز توزیع، و Q معیار پراکندگی می باشد.

۹.۶.۱ تمرین بخش شش

۱. میزان حرارت ظهر در یکی از شهرهای ایران در دی ماه برحسب درجه سانتیگراد عبارت است از:

۸ ۱۱ ۱۲ ۱۶ ۷ ۹ ۴ ۶ ۶ ۴ ۴ ۵ ۷ ۶ ۸

۹ ۱۰ ۱ ۱۶ ۷ ۹ ۴ ۵ ۶ ۸ ۷ ۱۲ ۱۱ ۶

برد، میان برد، انحراف استاندارد، میان چارکی، و نیم برد چارکها را حساب کنید.

۲. با استفاده از فرمول (۱۶)، فرمول (۱۸) را به دست آورید.

۳. طول عمرهای ۱۰۰ لامپ اسرام ۶۰ وات، برحسب ساعت در جدول زیر داده شده اند.

| طول عمر | فراوانی |
|---------|---------|
| ۳۰۰-۳۱۹ | ۱۰ |
| ۳۲۰-۳۳۹ | ۲۰ |
| ۳۴۰-۳۵۹ | ۳۸ |
| ۳۶۰-۳۷۹ | ۲۵ |
| ۳۸۰-۳۹۹ | ۷ |
| | ۱۰۰ |

میانگین و واریانس را با روش تبدیل حساب کنید. هیستوگرام و چند ضلعی فراوانی را رسم کنید. میانگین و میانه و نما را با هم مقایسه نمایید. درصد فراوانی را در فاصله $(\bar{x}, \bar{x} + s)$ یا فرض نرمال بودن منحنی فراوانی و از روی جدول فوق محاسبه کنید.

۴. فرمولهای (۲۲) را به دست آورید.

۵. برای داده هائی که در جدول ۲ خلاصه شده اند، میانگین و واریانس را با روش کوتاه حساب کنید.

۶. نشان دهید که برای داده های استاندارد میانگین صفر و واریانس یک می باشد.

۷. در یک کارخانه آجر فشاری میانگین مزد روزانه کارگران ۴۵۰ تومان و انحراف استاندارد ۴۵ تومان، و

در یک کارخانه روغن نباتی میانگین مزد روزانه کارگران ۶۵۰ تومان و انحراف استاندارد ۶۰ تومان

می باشد. مزد علی در کارخانه آجر فشاری ۵۰۰ تومان و مزد احمد در کارخانه روغن نباتی ۶۰۰

تومان می باشد. نسبت به کارگران کارخانه خود مزد علی بهتر است یا احمد؟ از روی ضریب تغییر

درباره مزد این دو کارخانه قضاوت کنید. اگر مردها متغیرهای نرمال باشند، مزد در کدام کارخانه بهتر

است؟

۸. نشان دهید که ضریب تغییر به واحد اندازه گیری بستگی ندارد، یعنی اگر داده ها را در عدد مثبت b

ضرب کنیم، ضریب تغییر ثابت می ماند.

۹. ده قوطی تایید دارای میانگین ۳۰۰ گرم و انحراف استاندارد ۱۰ گرم می باشند. این قوطیها از ۱۰ مغازه

با قیمت های متفاوت خریداری شده اند. میانگین قیمتها ۱۵۰ ریال و انحراف استاندارد قیمتها ۱۰

ریال است. آیا وزنها به هم نزدیکترند یا قیمتها؟

۱۰. اگر x_i ها را دو مقایسه کنیم و معدل مجموع n^2 پراتز $(x_j - x_i)^2$ را در نظر بگیریم، معیاری به صورت

$$\delta^2 = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (x_j - x_i)^2$$

برای پراکندگی به دست می آید ثابت کنید که $\delta^2 = 2s^2$.
 راهتمانی: داخل پراتز \bar{x} را اضافه و کم کنید

۱۱. در شهری مزد کارگران کارخانه پارچه بافی دارای میانگین ۵۰۰ تومان و انحراف استاندارد ۲۰ تومان و مزد کارگران کارخانه سیمان دارای میانگین ۴۵۰ تومان و انحراف استاندارد ۲۵ تومان می باشد مزد علی در کارخانه اول ۵۲۰ تومان و مزد حسن در کارخانه دوم ۴۷۰ تومان می باشد اگر مزد معیار مهارت باشد، علی ماهرتر است یا حسن؟

۱۲. ثابت کنید $\frac{s}{\bar{x}}$ با تغییر مبدأ و واحد اندازه گیری تغییر نمی کند؟

۷.۱ چولگی و برجستگی

طبیعی ترین منحنی فراوانی، منحنی فراوانی نرمال استاندارد می باشد که معادله مختصاتی آن به صورت

$$y = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (27)$$

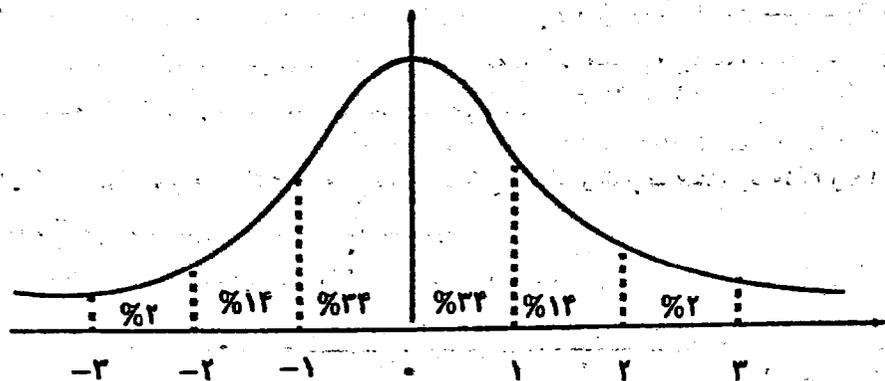
است. این منحنی زنگ - گونه و از نظر مقارن، کشیدگی، و پخی، تناسب و زیبایی خاصی دارد مساحت زیر منحنی به نحوی طبیعی طبق نگاره ۱۶ توزیع شده است.

در عمل هیچ متغیری وجود ندارد که منحنی فراوانی آن کاملاً نرمال استاندارد باشد. اغلب منحنی فراوانی داده ها نامتقارن و کشیده یا پخ می باشند و میزان نامرمانی را با دو معیار به نامهای چولگی و برجستگی می سنجند این دو معیار به میانگینهای مخصوص، به نام گشتاورها بستگی دارند.

۱.۷.۱ گشتاور و گشتاور مرکزی داده ها

فرض کنید x_1, x_2, \dots, x_n به ترتیب با فراوانیهای k_1, k_2, \dots, k_n یک سری داده n تایی باشند میانگین توان m_r x_i ها و \bar{x} ها، یعنی

$$m_r' = \frac{\sum_{i=1}^k f_i x_i^r}{n} \quad m_r = \frac{\sum_{i=1}^n f_i (x_i - \bar{x})^r}{n} \quad (28)$$



نگاره ۱۶ منحنی نرمال استاندارد

را گشتاور m_2 ، و گشتاور مرکزی m_3 داده ها می نامند. k معمولاً یک عدد طبیعی است. واضح است که m_1' برابر \bar{x} و m_2' برابر s^2 می باشد. اگر داده ها نسبت به میانگین، متقارن باشند، گشتاورهای مرکزی فرد یعنی m_3, m_5, \dots برابر صفر هستند.

۲.۷.۱ چولگی

میزان عدم تقارن منحنی فراوانی را چولگی می نامند. فرض کنید \bar{x} میانگین، m میانه، M نما، s انحراف استاندارد، و m_3 گشتاور مرکزی سوم باشند. هر کدام از فرمولهای زیر را می توان به عنوان معیار چولگی به کار برد:

$$b_1 = \frac{\bar{x} - M}{s} \quad \text{ضریب چولگی اول پیرسن}$$

$$b_2 = \frac{3(\bar{x} - m)}{s} \quad \text{ضریب چولگی دوم پیرسن} \quad (29)$$

$$g = \frac{m_3}{s^3} \quad \text{ضریب گشتاوری چولگی}$$

در فرمولهای بالا s را بدین جهت به کار برده ایم، تا این ضرائب به واحد اندازه گیری بستگی نداشته باشند. در صورتیکه داده ها نسبت به میانگین متقارن باشند، ضرائب بالا برابر صفر هستند. ولی باید متوجه بود که عکس این موضوع کاملاً صحت ندارد. برحسب این که b_1, b_2 و g مثبت یا منفی باشند، منحنی فراوانی چوله به راست (نگاره ۱۳) یا چوله به چپ (نگاره ۱۴) می باشد.

چون نما کمتر از میانگین و میانه، تحت تاثیر چولگی فرار می‌گیرد، از اینرو ضریب b_1 را به صورت بالا تعریف کرده‌اند. ولی در عمل محاسبه نما با دقت کافی مشکل می‌باشد، بنابراین b_1 را که به کمک رابطه (۱۱) از روی b_1 به دست می‌آید به کار می‌برند.

مثال ۲۶ طول عمر ۱۰۰ باتری اتومبیل دارای میانگین و میانه و انحراف استاندارد ۳/۴۸ و ۳/۵۰ و ۱/۶۵ سال می‌باشد. ضریب چولگی را محاسبه کنید.

با اطلاعات داده شده، ضریب چولگی دوم پیرسن می‌شود

$$b_2 = \frac{\sum (\bar{x} - m)}{s} = \frac{\sum (3/50 - 3/48)}{1/65} = 0.036$$

بنابراین منحنی فراوانی طول عمر باتریها کمی چوله به راست می‌باشد.

۳.۷.۱ برجستگی

میزان کشیدگی یا پخی منحنی فراوانی را نسبت به منحنی نرمال استاندارد، برجستگی آن می‌نامند. فرض کنید m_2 گشتاور مرکزی چهارم و s^4 انحراف استاندارد باشد. چون برای داده‌های نرمال $\frac{m_2}{s^4}$ به عدد ۳ نزدیک می‌باشد، معیار برجستگی را از فرمول

$$k = \frac{m_2}{s^4} - 3 \quad (30)$$

به دست می‌آورند. برحسب آنکه k مثبت یا منفی باشد، منحنی فراوانی کشیده یا پخ می‌باشد. اگر k نزدیک صفر باشد، برجستگی منحنی فراوانی طبیعی است.

مثال ۲۷ در مثال ۲۶ گشتاور مرکزی چهارم برابر است با ۹/۲۳. برجستگی را محاسبه کنید.

$$k = \frac{9/23}{(1/65)^4} - 3 = 1/25 - 3 = -1/75$$

بنابراین منحنی فراوانی طول عمر باتریها، نسبت به منحنی فراوانی نرمال استاندارد، پخ می‌باشد.

۴.۷.۱ تمرین بخش هفت

۱. ثابت کنید اگر به هر یک از داده‌ها مقدار ثابت a را اضافه کنیم، گشتاورهای مرکزی و در نتیجه معیارهای چولگی و برجستگی تغییر نمی‌کنند.

۲. نشان دهید که ضریب دوم چولگی پیرسن را می‌توان از روی ضریب اول از رابطه (۱۱) به دست آورد.

۳. گشتاور مرکزی سوم را برحسب گشتاورهای اول، دوم، و سوم به دست آورید.

۴. معیار چولگی را به کمک میانه و چارکهای اول و سوم به صورت

$$\frac{(Q_3 - m) - (m - Q_1)}{Q_3 - Q_1}$$

تعریف می‌کنند. این معیار را برای داده‌های جدول فراوانی ۳ محاسبه کنید. به کمک دهکها، چگونه می‌توان معیار چولگی را تعریف کرد؟

۵. برای جدول فراوانی ۳ چولگی و برجستگی را محاسبه کنید و درباره منحنی فراوانی قد تفسیر نمایید.

۸.۱ چند نمودار جدید

در سالهای اخیر چند نمودار مفید آماری ارائه گردیده‌اند که در مبحث آنالیز داده‌ها و آمار توصیفی به کار می‌روند. این نمودارها را می‌توان به کمک نرم افزارهای آماری یا با دست تهیه نمود. در این بخش بعضی از این نمودارها را شرح می‌دهیم.

۱.۸.۱ نمودارهای ساقه‌ای

یا هیستوگرام داده‌ها نمی‌توان یک یک داده‌ها را نمایش داد. یکی از آماردانان به نام توکی، بیست سال پیش در کتاب خود نمودار دیگری به نام نمودار ساقه‌ای یا نمودار ساقه و برگ مطرح کرد که نقش هیستوگرام را دارد و افزون بر این، تمام داده‌ها را می‌توان در آن دید. سه نوع نمودار ساقه‌ای به کار می‌رود.

نمودار ساقه‌ای ساده

این نمودار معمولاً برای داده‌هایی که به صورت اعداد کوچک هستند به کار می‌رود. فرض کنید داده‌های زیر تعداد دانش‌آموزان کلاس اول ابتدائی در ۱۵ دبستان شهر A باشند. این داده‌ها به ترتیب غیر نزولی عبارتند از

۶۰ ۵۹ ۵۴ ۴۹ ۴۷ ۴۶ ۴۶ ۴۶ ۴۱ ۴۱ ۳۵ ۳۴ ۲۵ ۲۴ ۲۲

نمودار ساقه‌ای ساده را اینگونه می‌سازند.

(۲) بزرگترین داده یعنی x_n

(۳) میانه داده‌ها یعنی m

(۴) چارک اول داده‌ها یعنی Q_1

(۵) چارک سوم داده‌ها یعنی Q_3

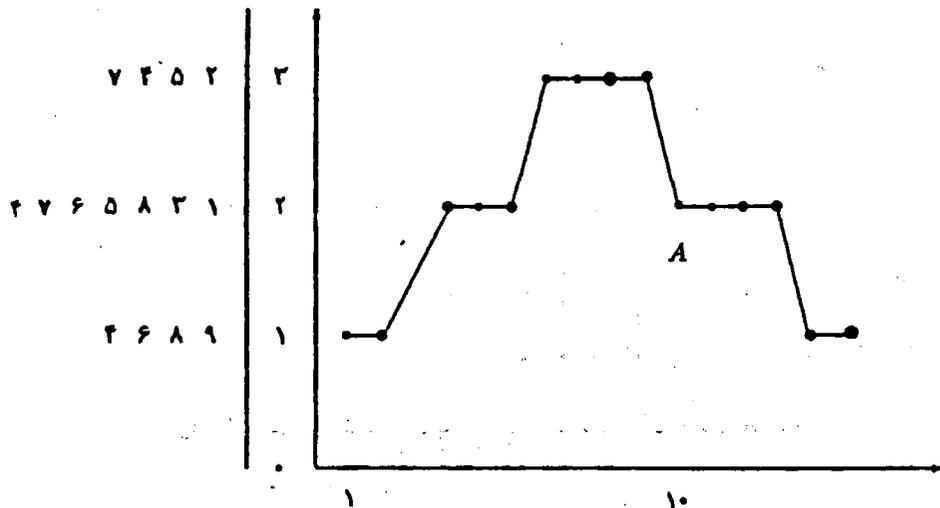
این ۵ عدد را با ۵ نقطه به طولهای x_n, Q_3, m, Q_1 و x_1 عمودی محور طولها نشان می‌دهیم حال مستطیلی می‌سازیم که دو ضلع عمودی آن از Q_1 و Q_3 بگذرند، و از m ، خطی به موازات دو ضلع عمودی رسم می‌کنیم. از Q_1 و Q_3 دو خط دیگر به نام دنباله‌ها به ترتیب تا x_1 و x_n امتداد می‌دهیم. بدین طریق نمودار جعبه‌ای یا نمودار جعبه‌ای دنباله‌دار به دست می‌آید. فاصله دو نقطه پایانی دنباله‌ها را با $R = x_n - x_1$ نشان می‌دهیم و آن را برد داده‌ها می‌نامیم. طول جعبه را با $Q = Q_3 - Q_1$ نشان می‌دهیم و آن را برد میان چارکی می‌نامیم. نمودار جعبه‌ای با یک نگاه نحوه پراکندگی داده‌ها را نشان می‌دهد. هر اندازه R و Q بزرگتر باشند پراکندگی بیشتر است. میزان تقارن یا عدم تقارن با این نمودار به خوبی دیده می‌شود. داده‌هایی که نسبت به سایر داده‌ها خیلی کوچکتر یا خیلی بزرگتر باشند، باعث می‌شوند که R بزرگ شود و در نتیجه دنباله چپ یا راست دراز گردند. معمولاً این نوع داده‌ها را که به داده‌های دور افتاده یا پرت شهرت دارند، کنار می‌گذارند و نمودار جعبه‌ای را با بقیه داده‌ها می‌سازند و داده‌های پرت را با نقاط جداگانه در پایان دنباله‌ها می‌گذارند. چنین نموداری را نمودار جعبه‌ای اصلاح شده می‌گویند. معمولاً اگر داده‌ای بیش از $1/5Q$ از Q_1 یا Q_3 دور باشد، آن را داده پرت محسوب می‌دارند.

مثال ۲۸ داده‌های زیر نمره‌های یک آزمون ۱۰۰ تستی در یک کلاس ۴۰ نفری هستند که به ترتیب غیر نزولی تهیه شده‌اند:

| | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|
| ۱۰ | ۳۸ | ۳۸ | ۳۹ | ۳۹ | ۴۰ | ۴۰ | ۴۱ | ۴۲ | ۴۵ |
| ۴۷ | ۴۸ | ۴۹ | ۴۹ | ۵۰ | ۵۱ | ۵۱ | ۵۳ | ۵۳ | ۶۱ |
| ۶۱ | ۶۳ | ۶۳ | ۶۵ | ۶۵ | ۷۰ | ۷۲ | ۷۳ | ۷۴ | ۷۵ |
| ۷۶ | ۷۶ | ۷۸ | ۷۸ | ۸۰ | ۸۰ | ۸۳ | ۸۵ | ۹۲ | ۹۸ |

می‌خواهیم نمودار جعبه‌ای این داده‌ها را رسم کنیم. چکیده پنج عددی عبارتند از

$$x_1 = 10 \quad Q_1 = 47 \quad m = 61 \quad Q_3 = 75 \quad x_n = 98$$



نگاره ۱۹ نمودار ساقه‌ای نقطه‌دار برای درجه حرارت تابستان

برای تعیین نمودار ساقه‌ای نقطه‌دار، نخست ساقه‌ها را میان دو خط عمودی و برگها را به ترتیب ثبت داده‌ها طرف چپ ساقه‌های مربوط می‌گذاریم. با استفاده از محور ساقه‌ها و محور زمان، هر داده را با یک نقطه نشان می‌دهیم. طول این نقطه زمان ثبت داده و عرضش ساقه داده است. نگاره ۱۹ نمودار ساقه‌ای نقطه‌دار داده‌های بالا را نشان می‌دهد. توجه کنید که در این نمودار برگها به ترتیب ثبت داده‌ها، نه به ترتیب صعودی، قرار می‌گیرند.

به عنوان مثال در این نگاره نقطه A می‌گوید که در دهم مرداد درجه حرارت ۲۵ است.

۲.۸.۱ نمودار جعبه‌ای

این نمودار برای نشان دادن نحوه پراکندگی داده‌ها سودمند است. برای ساختن آن روش زیر را به کار می‌بریم.

نخست داده‌ها را به ترتیب غیر نزولی به صورت $x_1 \leq x_2 \leq \dots \leq x_n$ می‌نویسیم. اینک آنها را در پنج عدد زیر به نام چکیده پنج عددی فشرده می‌کنیم:

(۱) کوچکترین داده یعنی x_1

۲. تعداد تصادفات اتومبیل در شهری در ۱۵ روز اول تابستان و ۱۲ روز اول زمستان عبارتند از:

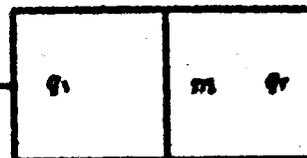
۱۲ ۱۰ ۱۵ ۲۳ ۱۲ ۲۷ ۱۶ ۲۴ ۲۱ ۲۳ ۲۲ ۱۸ ۲۵ ۳۱ ۱۹

۳۱ ۱۸ ۲۲ ۲۷ ۲۴ ۲۲ ۲۸ ۲۱ ۲۴ ۲۵ ۲۳ ۲۰

الف - نمودار ساقه‌ای پشت به پشت را بسازید.

ب - نمودار ساقه‌ای نقطه‌دار را برای تابستان و برای زمستان رسم کنید.

ج - نمودار جمع‌بندی را برای تابستان و برای زمستان رسم کنید.



نگاره ۲۰ نمودار جمع‌بندی برای نمره‌های نمره

نگاره ۲۰ نمودار جمع‌بندی دادها را نشان می‌دهد.

یادآور می‌شویم که Q_1 را میانه تمام داده‌هایی گرفتیم که از m کوچکتر یا مساوی آن هستند Q_3 را هم میانه تمام داده‌هایی گرفتیم که از m بزرگتر یا مساوی آن هستند البته می‌توان این چارکها را با روشی دقیقتر حساب کرد. ولی اختلاف چندانی قابل ملاحظه نیست.

چون $Q_3 - Q_1 = 28$ و $Q_2 = 42$ پس داده‌ای بزرگ است که از Q_1 به اندازه ۲۲ کوچکتر یا از Q_3 به اندازه ۲۲ بزرگتر باشد در این مثال چنین داده‌ای نداریم و نیازی به اصلاح نمودار جمع‌بندی نیست.

۳۸.۱ تمرین بخش هشت

۱. در آزمون ریاضی عمومی، یک کلاس ۵۰ نفری نمره‌های زیر را گرفتند:

۷۲ ۶۵ ۹۷ ۹۳ ۸۲ ۵۱ ۶۷ ۸۵ ۹۵ ۸۱
 ۸۰ ۹۲ ۸۵ ۷۱ ۵۴ ۹۱ ۹۰ ۸۳ ۶۲ ۶۳
 ۸۸ ۹۵ ۹۸ ۵۷ ۵۱ ۹۲ ۷۳ ۶۱ ۶۹ ۷۰
 ۵۲ ۷۱ ۹۳ ۶۲ ۵۵ ۸۴ ۸۱ ۶۵ ۷۲ ۶۳
 ۲۱ ۶۵ ۸۶ ۲۲ ۵۱ ۹۳ ۸۲ ۶۷ ۶۴ ۵۹

الف - نمودار ساقه‌ای را بسازید.

ب - نمودار ساقه‌ای نقطه‌دار را رسم کنید.

ج - نمودار جمع‌بندی را رسم کنید.

احتمال در علوم، مهندسی، علوم اجتماعی، علوم قضائی، پزشکی، بازاریابی، و غیره وسیله مهمی برای تجزیه و تحلیل پدیده‌ها است.

در تئوری احتمال واژه‌هایی مانند آزمایش، پیشامد، و غیره را که از زبان عادی گرفته شده‌اند، به کار می‌برند. باید نخست این واژه‌ها را دقیقاً بدون ابهام روشن نمود تا بتوان تئوری را با اسلوب ریاضی پایه‌گذاری کرد.

۱.۱.۲ آزمایش تصادفی

هر عملی را که تصادف در آن دخالت داشته باشد، در اصطلاح تئوری احتمال، آزمایش تصادفی می‌گویند. دقیقاً آزمایشی تصادفی عبارت است از عملی که تحت شرایطی ثابت و برحسب موضوعی مورد علاقه، هر بار که تکرار شود، منجر به یکی از اعضای مجموعه ناتهی S گردد. مجموعه S متشکل از تمام نتیجه‌هایی که تحت شرایط آزمایش و برحسب موضوع مورد علاقه قابل تصور است.

مثال ۱ پنج کارت یکسان، دو عدد سفید، یک عدد قرمز و دو عدد سبز را به خوبی مخلوط می‌کنیم و چشم بسته یکی را بیرون می‌آوریم. تحت این شرایط، اگر رنگ کارت بیرون آمده موضوع مورد علاقه باشد، داریم: $S = \{\text{سفید، قرمز، سبز}\}$. هر بار که آزمایش را انجام دهیم کارتی به یکی از سه رنگ یاد شده بیرون می‌آید.

اگر شرایط آزمایش و موضوع مورد علاقه را تغییر دهیم، مثلاً کارتها دارای شماره‌های ۱، ۲، ۳، ۴، ۵ باشند و موضوع مورد علاقه شماره کارت باشد آنگاه $S = \{1, 2, 3, 4, 5\}$.

مثال ۲ سکه‌ای را به تصادف می‌اندازیم و شرایط را به نحوی کنترل می‌کنیم که گم نشود و در خاک نرم روی پهلوی قرار نگیرد، بلکه بعد از سکون نقش یکی از دو روی آن که شیر و خط نامیده می‌شوند و آنها را با H و T نشان می‌دهند، از بالا دیده شود. اگر موضوع مورد علاقه نقش روی سکه بعد از سکون باشد داریم: $S = \{H, T\}$. این آزمایش را که با آن آشنائی داریم بازی شیر و خط می‌نامیم. هرگاه موضوع مورد علاقه، چیز دیگری مانند محل سکون سکه یا نوع صدای برخورد آن با زمین باشد، طبعاً S پیچیده‌تر خواهد شد.

از دو مثال بالا معلوم می‌شود که در هر آزمایش تصادفی، برای اینکه S بدون ابهام تعیین شود، باید شرایط آزمایش و موضوع مورد علاقه را قبلاً مشخص کرد.

قوانین شانس یا احتمال

۱.۲ مقدمه

انسان به علت عدم توانائی در کشف تمام رموز عالم و عدم اطمینان در برابر حوادث روزگار مجبور است مفاهیمی از قبیل تصادف و شانس را به عنوان پاره‌ای از زندگی روزانه بپذیرد. با این حال دانشمندان همواره تلاش کرده‌اند تا از راه تجربه و مشاهده به کمک علم و تکنولوژی تا آنجا که مقدور است علت پدیده‌ها را کشف نمایند و عدم اطمینان را کاهش دهند.

درباره تصادف و شانس، چه در سطح ابتدائی به صورت داستانها و چه در سطح عالی به صورت بحثهای فلسفی، هزاران کتاب و مقاله نوشته‌اند، که انگیزه‌های خوبی برای تفکر در این زمینه می‌باشند. متفکران ریاضی سعی کرده‌اند تا موضوع را لااقل در موارد ساده به کمک اعداد و قوانین ریاضی، تجزیه و تحلیل نمایند و نظام امور تصادفی را، با اینکه تصادف در ظاهر تابع هیچ نظامی نیست، پیدا کنند. می‌دانیم که قماربازان در حین تفنن و تشریش، به تصادف برد و باخت و خوش شانس و بدشانسی توجه زیاد دارند. از قضا دانش تجزیه و تحلیل امور تصادفی از این راه غیر اخلاقی آغاز شده است، این دانش را که بر پایه ریاضی علم آمار است، احتمال می‌نامند.

برای اولین بار یکی از اشراف زادگان و قماربازان معروف فرانسه، برای توضیح علت برد و باخت با تاس تخته نرد به دو ریاضیدان مشهور فرانسوی، پاسکال (۱۶۶۲-۱۶۲۳) و فرما (۱۶۶۵-۱۶۰۱)، متوسل گردید. مسائلی را که آنها حل کردند، مدت‌ها به راههای مختلف توجیه و تفسیر شد. ناگفته نماند که قبل از آنها دو دانشمند ایتالیائی، کاردان (۱۵۷۶-۱۵۰۱) و گالیله (۱۶۴۲-۱۵۶۴)، در این زمینه محاسبات گوناگون انجام داده بودند، که مورد توجه قماربازان آن زمان در ایتالیا قرار گرفت.

بنابراین از نظر تاریخی تئوری احتمال از بازیهای قمار در اوائل قرن هفدهم شروع شد و رفته رفته رشد نمود و در رشته‌های مختلف نفوذ کرد. در سال ۱۹۳۳ کولموگروف ریاضیدان روسی (۱۹۸۷-۱۹۰۳)، رساله‌ای درباره اصول احتمال منتشر کرد، و به احتمال کاملاً جنبه ریاضی داد. امروز

۲.۱.۲ پیشامد و فضای نمونه

در یک آزمایش تصادفی، هر پیشامد E عبارت است از زیر مجموعه‌ای از مجموعه تمام نتیجه‌ها. اگر مجموعه E یک عضوی به صورت $\{e\}$ باشد آن را یک پیشامد ساده می‌نامند و معمولاً تنها با e نشان می‌دهند. اگر E تهی باشد آن را پیشامد محال می‌نامند و \emptyset نشان می‌دهند و اگر E برابر با S باشد آن را پیشامد حتمی می‌نامند.

چون در آزمایشهای تصادفی، به ویژه کارهای آزمایشگاهی، نمونه‌برداری معمول است، از اینرو S را فضای نمونه می‌نامند. در حقیقت S را باید فضای پیشامدها نامید.

۳.۱.۲ رخ دادن یک پیشامد

می‌گوئیم در یک آزمایش تصادفی، پیشامد E رخ می‌دهد، هرگاه آزمایش تصادفی منجر به مشاهده یکی از اعضای E گردد.

مثال ۳ دو عدد کارت سفید به شماره‌های ۱، ۲، سه عدد کارت قرمز به شماره‌های ۳، ۴، ۵ و دو عدد کارت سبز به شماره‌های ۶، ۷ را به خوبی مخلوط کرده، چشم بسته یکی از این هفت کارت را بیرون می‌آوریم. فرض کنید شماره روی کارت موضوع مورد علاقه باشد. بنابراین فضای نمونه می‌شود $S = \{1, 2, 3, 4, 5, 6, 7\}$. پیشامد مشاهده کارت قرمز، یعنی پیشامد اینکه کارت بیرون آمده قرمز باشد، زیر مجموعه‌ای از S به صورت $\{3, 4, 5\}$ است. این پیشامد هنگامی رخ می‌دهد که یکی از کارتهای شماره ۳ یا ۴ یا ۵ بیرون آید. پیشامد مشاهده کارت مضروب ۵، یک پیشامد ساده به صورت $\{5\}$ است.

پیشامد مشاهده کارت بزرگتر از ۸، می‌شود پیشامد محال یعنی مجموعه تهی، زیرا بزرگترین شماره می‌تواند عدد ۷ باشد.

پیشامد مشاهده کارت کوچکتر از ۱۰، می‌شود پیشامد حتمی یعنی تمام S ، زیرا تمام شماره‌ها از ۱۰ کمتر هستند.

مثال ۴ در بازی شیر و خط سکه‌ای را سه بار می‌اندازیم. در این آزمایش تصادفی فضای نمونه می‌شود

$$S = \{TTT, TTH, THT, HTT, THH, HTH, HHT, HHH\}$$

به عنوان مثال، منظور از HHT این است که سکه بار اول و دوم شیر و بار سوم خط بیاید. این یک پیشامد ساده است که آن را به صورت $\{HHT\}$ و یا برای آسانی به صورت HHT نشان می‌دهیم. پیشامد بار اول و آخر شیر، می‌شود: $\{HTH, HHH\}$.

پیشامد بیش از سه شیر می‌شود پیشامد محال یعنی مجموعه تهی، زیرا حداکثر می‌توان سه شیر مشاهده کرد.

پیشامده حداکثر سه خط می‌شود پیشامد حتمی یعنی تمام S ، زیرا در این آزمایش تصادفی، مشاهده هیچ یا یک یا دو یا سه خط قطعی است.

۴.۱.۲ زیر پیشامد

اگر هر پیشامد ساده که در E است در F نیز باشد، می‌گویند E زیر پیشامد F است و می‌نویسند $E \subset F$. واضح است که رخ دادن E باعث رخ دادن F می‌شود ولی عکس این موضوع همواره درست نیست. در مثال ۳، پیشامد کارت سفید زیر پیشامد کارت کوچکتر از ۵ است، زیرا

$$\{1, 2, 3, 4\} = \text{پیشامد مشاهده کارت کوچکتر از ۵}$$

$$\{1, 2\} = \text{پیشامد مشاهده کارت سفید}$$

۵.۱.۲ پیشامدهای برابر

اگر E زیر پیشامد F و F زیر پیشامد E باشد، می‌گویند E و F برابرند و می‌نویسند: $E = F$. در مثال ۳ پیشامد کارت سفید و پیشامد کارت کوچکتر از سه برابرند، زیرا هر دو می‌شوند $\{1, 2\}$.

۶.۱.۲ اعمال مجموعه‌ای روی پیشامدها

اعمال مجموعه‌ای که روی پیشامدها می‌توان انجام داد، درست مانند اعمال روی مجموعه‌ها می‌باشند. اینک آنها را شرح می‌دهیم.

(۱) اجتماع دو پیشامد E و F

پیشامدی است که از تمام پیشامدهای ساده که در E یا F یا هر دو باشند تشکیل می‌شود و آنرا با $E \cup F$ نشان داده، پیشامد E یا F می‌نامند. این پیشامد موقعی رخ می‌دهد که لااقل یکی از دو پیشامد E و F رخ دهند.

(۲) اشتراک دو پیشامد E و F

پیشامدی که از تمام پیشامدهای ساده که هم در E و هم در F باشند تشکیل می‌شود و آنرا با $E \cap F$ نشان داده، پیشامد E و F می‌نامند. این پیشامد موقعی رخ می‌دهد که E و F هر دو با هم رخ دهند. اگر $E \cap F = \emptyset$ ، دو پیشامد E و F را جدا می‌نامند و در این حال رخ دادن هر دو با هم محال است و معمولاً می‌گویند E و F ناسازگارند.

(۳) تفاضل پیشامد E از پیشامد F

پیشامدی است که از تمام پیشامدهای ساده که در F بوده ولی در E نباشند تشکیل می‌شود و آنرا با $F - E$ نشان داده، پیشامد F نه E می‌نامند. اگر E زیر پیشامد F باشد، $F - E$ را یک تفاضل واقعی می‌نامند، زیرا تمام E را در بر دارد.

(۴) متمم پیشامد E

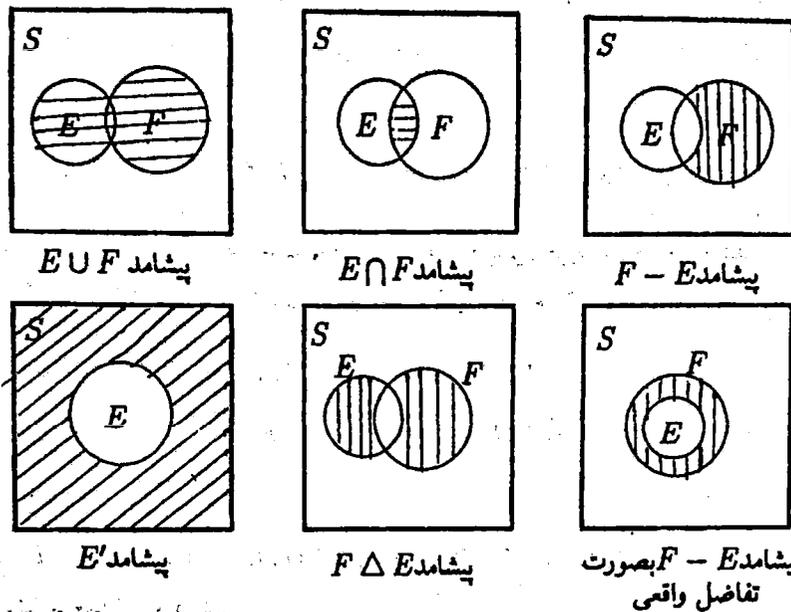
تفاضل پیشامد E از تمام فضای نمونه، یعنی $S - E$ را متمم E می‌گویند و آنرا با E' نشان داده، پیشامد نه E می‌نامند. واضح است که

$$E \cup E' = S, \quad E \cap E' = \emptyset$$

(۵) تفاضل متقارن دو پیشامد E و F

پیشامدی است که از تمام پیشامدهای ساده که یا در E یا در F ولی نه در هر دو باشند تشکیل می‌شود و آنرا با $E \Delta F$ نشان می‌دهند و پیشامد یا E یا F می‌نامند. واضح است که $E \Delta F$ و $F \Delta E$ با هم برابرند و از اینرو کلمه متقارن را به کار می‌بریم.

به کمک نگاره ۱، که نمودارهای ون نامیده می‌شوند، می‌توان اعمال مجموعه‌ای روی پیشامدها را نشان داد. در این شکل فضای نمونه را با یک مستطیل، پیشامدهای E و F را با دو دایره و پیشامد مورد نظر را به وسیله هاشور مشخص کرده‌ایم.



نگاره ۱ نمودارهای ون

مثال ۵ در مثال ۳ فرض می‌کنیم E پیشامد کارت بزرگتر از یک و F پیشامد کارت کوچکتر از پنج باشد، یعنی $E = \{۲, ۳, ۴, ۵, ۶, ۷\}$ و $F = \{۱, ۲, ۳, ۴\}$. اینک داریم:

- | | |
|--------------------------------------|-------------------------------|
| $E \cup F = \{۱, ۲, ۳, ۴, ۵, ۶, ۷\}$ | اجتماع E و F می‌شود |
| $E \cap F = \{۲, ۳, ۴\}$ | اشتراک E و F می‌شود |
| $E' = \{۱\}$ | متمم E می‌شود |
| $F - E = \{۱\}$ | تفاضل E از F می‌شود |
| $E - F = \{۵, ۶, ۷\}$ | تفاضل F از E می‌شود |
| $E \Delta F = \{۱, ۵, ۶, ۷\}$ | تفاضل متقارن E و F می‌شود |

ملاحظه کنید که در این مثال، پیشامد کارت سفید یعنی $\{۱, ۲\}$ پیشامد کارت قرمز یعنی $\{۳, ۴, ۵\}$ جدا هستند. به عبارت دیگر کارتی را که بیرون می‌آوریم نمی‌تواند هم سفید و هم قرمز باشد. ولی پیشامد کارت فرد و پیشامد کارت مضرب سه، یعنی $\{۱, ۳, ۵, ۷\}$ و $\{۲, ۴, ۶\}$ جدا نیستند و با بیرون آمدن کارت شماره سه، هر دو پیشامد با هم رخ می‌دهند.

مثال ۶ یک زن و شوهر جوان آرزو دارند که در آیند صاحب سه فرزند شوند، آنها فرض می‌کنند صاحب فرزندان دوقلو یا چند قلو نشوند. اگر پسر و دختر را به ترتیب با B و G نشان دهیم فضای نمونه می‌شود

$$S = \{BBB, BBG, BGB, GBB, BGG, GBG, GGB, GGG\}$$

به عنوان مثال منظور از BBG این است که فرزند اول و دوم پسر و فرزند سوم دختر باشد.

فرض می‌کنیم E پیشامد داشتن فرزندان پسر و دختر و F پیشامد حداقل دو پسر باشد، یعنی

$$E = \{BGG, GBG, GGB, BBG, BGB, GBB\}$$

$$F = \{BBG, GBB, BGB, BBB\}$$

اینک داریم:

اجتماع E و F می‌شود $E \cup F = \{BBB, BBG, GBB, BGB, BGG, GBG, GGB\}$

یعنی پیشامد حداکثر دو دختر.

اشتراک E و F می‌شود $E \cap F = \{BBG, GBB, BGB\}$ ، یعنی پیشامد درست دو پسر.

متمم E می‌شود $E' = \{BBB, GGG\}$ ، یعنی پیشامد فرزندان همجنس.

تفاضل E از F می‌شود $F - E = \{BBB\}$ ، یعنی پیشامد سه پسر.

تفاضل متقارن E و F می‌شود $E \Delta F = \{BBB, BGG, GBG, GGB\}$ یعنی پیشامد نداشتن

درست دو پسر.

۷.۱.۲ فضای نمونه با پایان و بی پایان

اگر تعداد اعضای فضای نمونه یک عدد مثبت باشد، آن را فضای نمونه با پایان و در غیر این صورت آن

را فضای نمونه بی پایان می‌نامند. در تمام مثالهای بالا فضای نمونه با پایان است. مثلاً در مثال ۴ فضای

نمونه دارای ۸ عضو است. برای اینکه منظور از فضای نمونه بی پایان روشن شود به مثال زیر توجه کنید:

مثال ۷ در یک آزمایش تصادفی، بازی شیر و خط را آتقدر ادامه می‌دهیم تا برای نخستین بار شیر بیاید

مثلاً اگر چهار بار اول خط و بار پنجم شیر بیاید، نتیجه آزمایش می‌شود $TTTTH$. فرض کنید اینگونه

دنباله‌ها موضوع مورد علاقه باشد. بنابراین فضای نمونه می‌شود

$$S = \{H, TH, TTH, \dots\}$$

که یک مجموعه بی پایان است. در اینجا مجموعه S را که با مجموعه اعداد طبیعی تناظر یک به یک دارد، از این پس مجموعه شمارش پذیر می‌نامیم.

در این مثال، پیشامد زوج بودن تعداد بازیها می‌شود $\{TH, TTTT, \dots\}$ و پیشامد حداکثر پنج

بازی می‌شود $\{H, TH, TTH, TTTT, TTTTH\}$ ملاحظه می‌شود که پیشامد اول یک زیر

مجموعه شمارش پذیر و پیشامد دوم یک زیر مجموعه با پایان از S می‌باشد.

۸.۱.۲ تمرین بخش یک

۱. کتابی دارای ۵۰ صفحه به رنگ سفید، ۳۰ صفحه به رنگ سبز و ۲۰ صفحه به رنگ قرمز است.

یک صفحه از کتاب را چشم بسته انتخاب می‌کنیم. فضای نمونه را برحسب موضوع مورد علاقه تعیین کنید.

۲. یک سکه ۵ ریالی و یک سکه ۱۰ ریالی را با هم به تصادف در یک سینی می‌اندازیم. برحسب

موضوع مورد علاقه، فضای نمونه را تعیین کنید. $\{HH, HT, TH, TT\}$

۳. دوستی قرار است به تصادف بین ساعت یک و دو بعد از ظهر در پارک شهر شما را ملاقات نماید.

فضای نمونه را برحسب موضوع مورد علاقه تعیین کنید. آیا این فضا با پایان است یا بی پایان؟

۴. دو تاس را با هم به تصادف می‌ریزیم. فضای نمونه را برحسب موضوع مورد علاقه تعیین کنید.

۵. در اطاقی دو صندلی به شماره‌های سفید ۱ و ۲ و سه صندلی به شماره‌های سبز ۳ و ۴ و ۵ و

پنج صندلی به شماره‌های قرمز ۶ و ۷ و ۸ و ۹ و ۱۰ قرار دارند. تمام صندلیها یکسان و هم‌رنگ

می‌باشند. شخصی به تصادف روی یکی از این صندلیها می‌نشیند. فرض کنید E پیشامد سبز

بودن، F پیشامد مضرب پنج بودن، و G پیشامد زوج بودن شماره صندلی این شخص باشد.

پیشامدهای زیر را مشخص کنید:

$$E, F, E \cup F, E \cap F, F - E, E \Delta F, E \cap G$$

$$F \cup G, E' \cup F', E' \cap F', E' \Delta F', F' - E'$$

۶. فرض کنید E ، F و G سه پیشامد در یک آزمایش باشند. پیشامد اینکه:

الف - فقط E رخ دهد

ب - حداقل دو تا از این سه پیشامد رخ دهند

ج - درست دو تا از این سه پیشامد رخ دهند

را بر حسب این سه پیشامد یا متممهای آنها بنویسید.

۷. در آزمایشی بازی شیر و خط را چهار بار تکرار می‌کنیم. فرض کنید E پیشامد شیر در بار اول و آخر و F پیشامد شیر در بار دوم و سوم باشد. پیشامدهای زیر را مشخص کنید:

$$E, F, E \cup F, F \cap F, F - E, E \Delta F$$

۸. ثابت کنید برای هر دو مجموعه E و F داریم:

$$(E \cap F)' = E' \cup F', \quad (E \cup F)' = E' \cap F'$$

این قوانین را که به قوانین دم‌رگان معروف‌اند به صورت پیشامد توجیه کنید و آنها را به بیش از دو پیشامد تعمیم دهید.

۹. ثابت کنید:

$$(E \cup F) - F = E \cap F', \quad E \Delta F = (E - F) \cup (F - E)$$

۱۰. کدامیک از روابط زیر درست است؟

$$(الف) (E \cap F) \cup G = (E \cup G) \cap (F \cup G)$$

$$(ب) (F - E) \cup G = (F \cup G) - (E \cup G)$$

$$(ج) (F - E) \cap G = (F \cap G) - (E \cap G)$$

$$(د) (F \Delta E) \cap G = (F \cap G) \Delta (E \cap G)$$

۲.۲. تعبیرهای مختلف احتمال

نوشته‌ها و زبان عادی پر از جمله‌های احتمال می‌رود، ممکن است، حدس می‌زنم، اگر شانس بزنند، خیال می‌کنم، و مانند آنها هستند. تمام این جمله‌ها همراه با عدم یقین، بی‌اطلاعی، و تردید می‌باشند. همگی متضمن یک مفهوم پردردسر به نام احتمال هستند که اگر آن را به نحوی مطلوب تعبیر کنیم و اندازه‌گیری نمائیم، تصمیم‌گیری و حقیقت‌یابی آسانتر می‌شود.

۱.۲.۲. احتمال چیست؟

در پاسخ به این پرسش گفتگوهای فلسفی سالها است که ادامه دارد و برای همیشه ادامه خواهد داشت. تنها برای یک ریاضیدان که به مفهوم فلسفی احتمال کاری ندارد و می‌خواهد در چارچوبی محدود به اصول و قضایا تحقیق کند، مشکلی ایجاد نمی‌شود و گفتگوهای فلسفی چندان در کار او تاثیر ندارند. با این حال پایه‌های فلسفی هر موضوعی می‌توانند دیدها را تغییر دهند و موجب رشد تئوریهای جدید شوند. در مورد احتمال هم این مطلب صدق می‌کند.

با وجود اینکه فلاسفه و آماردانان درباره مفهوم فلسفی احتمال وحدت نظر دارند و اغلب یکدیگر را به باد انتقاد می‌گیرند، تعبیرهای زیر در عمل مفید هستند و با اسلوب ریاضی برخوردی ندارند.

الف - تعبیر احتمال به هر طریق همشانشی

هرگاه فضای نمونه یک آزمایش از n پیشامد ساده تشکیل شده باشد که از نظر رخ دادن هیچکدام بر دیگر برتری نداشته باشد، می‌گوئیم این پیشامدها همشانش هستند. اگر کل احتمال را یک بگیریم. احتمال رخ دادن هر یک از این پیشامدهای ساده می‌شود $\frac{1}{n}$.

مثلاً در بازی شیر و خط اگر سکه کاملاً معمولی و سالم باشد، پیشامدهای H و T را می‌توان همشانش تصور کرد و احتمال رخ دادن هر کدام $\frac{1}{2}$ است. چنین سکه‌ای را از این پس سکه نالرب می‌گوئیم، و سکه‌ای را که فاقد خاصیت همشانشی باشد سکه ارب می‌نامیم.

از نظر تاریخی تعبیر احتمال به طریق همشانشی قدیمی‌ترین تعریف برای احتمال است و در بازیهای قمار به خصوص بازی با تاس تخته و با ورق، مدت‌ها نظر بازیگران و مشاوران ریاضی آنها را در قرن هفدهم به خود جلب کرده بود. لاپلاس در اوائل قرن هیجده، رساله تازیخی خود را درباره تئوری احتمال، با این تعبیر که مفهوم کلاسیک احتمال نامیده می‌شود، پایه‌گذاری کرد.

به این تعبیر احتمال، دو انتقاد عمده وارد می‌باشد. نخست اینکه مفهوم همشانشی، به فرض اینکه شرایط آن برقرار باشد، با مفهوم احتمال که مورد تعبیر می‌باشد ارتباط دارد، مثلاً وقتی در بازی شیر و خط می‌گوئیم H و T همشانش هستند، در حقیقت می‌گوئیم احتمال‌های مساوی دارند. انتقاد دیگر این است که اگر پیشامدها همشانش نباشند، با این تعبیر نمی‌توانیم به آنها احتمال نسبت دهیم. مثلاً وقتی که می‌خواهیم بچه‌دار شدن یک زن و شوهر جوان را پیش‌بینی کنیم، پیش‌آمدهای همشانش بی‌مورد است و باید تعبیر دیگری برای احتمال در نظر گرفت.

جدول ۱ فراوانی و فراوانی نسبی در ۱۰۰ بازی شیر و خط

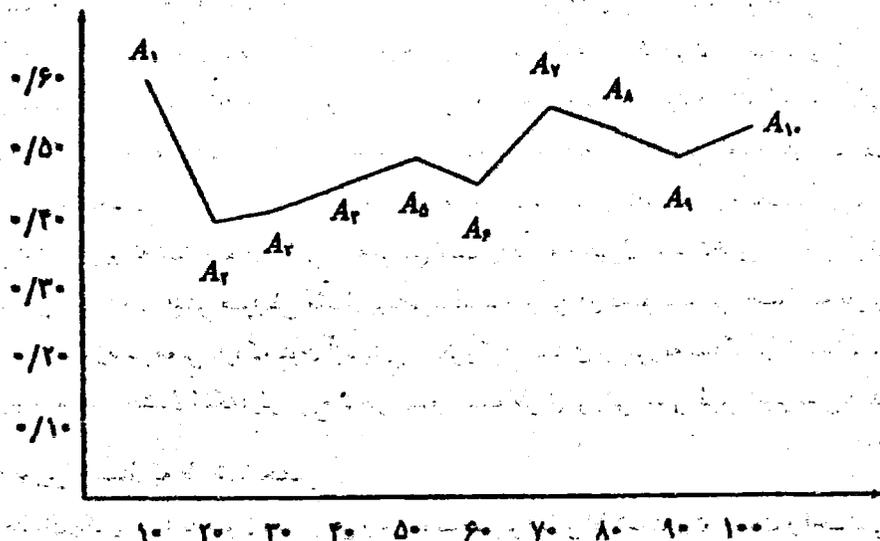
| | | | | | |
|----------------|------|------|------|------|------|
| تعداد آزمایشها | ۱۰ | ۲۰ | ۳۰ | ۴۰ | ۵۰ |
| فراوانی H | ۶ | ۸ | ۱۳ | ۱۸ | ۲۴ |
| فراوانی نسبی H | ۰/۶۰ | ۰/۴۰ | ۰/۴۳ | ۰/۴۵ | ۰/۴۸ |
| تعداد آزمایشها | ۶۰ | ۷۰ | ۸۰ | ۹۰ | ۱۰۰ |
| فراوانی H | ۲۷ | ۳۸ | ۴۳ | ۴۴ | ۵۲ |
| فراوانی نسبی H | ۰/۴۵ | ۰/۵۴ | ۰/۵۳ | ۰/۴۸ | ۰/۵۲ |

ب - تعبیر احتمال به طریق فراوانی نسبی

آزمایشی را تحت شرایط یکسان n بار تکرار می‌کنیم. فرض کنید r بار پیشامد E در این n آزمایش رخ دهد. می‌گوئیم r فراوانی و کسر $\frac{r}{n}$ فراوانی نسبی پیشامد E در این n آزمایش است. حال اگر n را بزرگ کنیم، r هم بزرگ شده یا ثابت می‌ماند و این باعث می‌گردد که $\frac{r}{n}$ دستخوش نوسان شود. یا این حال تجربه نشان می‌دهد که کسر $\frac{r}{n}$ ، با وجود همه نوسانها، سرانجام به سمت عدد ثابتی، که آن را احتمال پیشامد E به طریق فراوانی نسبی می‌نامند، گرایش پیدا می‌کند.

مثلاً اگر در بازی شیر و خط با یک سکه نازیب بخوایم احتمال پیشامد H را به طریق فراوانی نسبی تعیین کنیم، این بازی را ۱۰ بار، ۲۰ بار و بالاخره ۱۰۰ بار انجام می‌دهیم. نتیجه مشاهدات را به کمک جدول ۱ روی صفحه محورها مختصات با نقاط A_1, A_2, \dots, A_{100} نشان می‌دهیم. طول هر یک از این نقاط در نگاره ۲، نماینده تعداد آزمایشها و عرض آن نماینده فراوانی نسبی است.

البته در نگاره ۲ تعداد آزمایشها چندان زیاد نیست که بتوان وضع نهائی فراوانی نسبی را مشاهده نمود، ولی نحوه نوسان نقاط یاد شده اطراف خطی که عرضش $0/5$ است نشان می‌دهد که فراوانی نسبی با زیاد شدن تعداد آزمایشها سرانجام به $0/5$ که آن را احتمال پیشامد شیر می‌نامیم گرایش پیدا می‌کند. تعبیر احتمال به طریق فراوانی نسبی، که از لحاظ قدمت به ارسطو نسبت داده می‌شود عملاً از اواسط قرن نوزدهم متداول گردید. احتمال با این تعبیر طرفداران زیادی دارد زیرا این مفهوم در زمان عادی مردم هم رایج است. مثلاً وقتی می‌گوئیم احتمال موفقیت در این مسابقه برای فارغ التحصیلان دبیرستانها $0/80$ است، منظور این است که 80 درصد موفق می‌شوند. این تعبیر به سبب خواص عددی فراوانی نسبی، با اسلوب ریاضی که بعداً مطالعه خواهیم کرد کاملاً سازگار است. از طرفی چون در علم



نگاره ۲ توجه هندسی فراوانی نسبی شیرها در ۱۰۰ بازی شیر و خط

تجربی و اجتماعی با آزمایشهای هماتند و مکرر سروکار داریم، تعبیر احتمال به طریق فراوانی نسبی وسیله تفسیر و توجیه پدیده‌ها می‌باشد. به عنوان مثال اگر 90 درصد نوزادان در اثر خوردن یک نوع ویتامین طبی دچار کم اشتها می‌شوند، پزشکان اطفال ممکن است استعمال آن را منع نمایند. موضوع همشانسی را هم می‌توان به کمک فراوانی نسبی احساس کرد. مثلاً وقتی می‌گوئیم در بازی شیر و خط با یک سکه نازیب، پیشامد H و پیشامد T همشانس هستند منظور این است که اگر بازی را زیاد تکرار کنیم، فراوانی نسبی H و T انتظار می‌رود تقریباً با هم برابر شوند.

از نظر تاریخی کریش، ریاضیدان انگلیسی، در زندان نازیها در جنگ دوم جهانی سکه‌ای را 10000 بار پرتاب کرد و 5067 مرتبه شیر دید. از این راه نشان داد که فراوانی نسبی عدد $0/5067$ و احتمال شیرآمدن تقریباً این عدد است.

چون آزمایشهای مکرر بر هم اثر ندارند و تحت شرایط یکسان انجام می‌گیرند، بنابراین آزمایش کننده نمی‌تواند در این آزمایشها، که تعیین کننده فراوانی نسبی می‌باشد، بر اساس اطلاعات شخصی اعمال نظر یا اظهار عقیده نماید. از اینرو تعبیر احتمال به طریق فراوانی نسبی را تعبیر عینی هم می‌نامند. تعبیر احتمال به طریق فراوانی نسبی، با همه محاسنی که دارد، خالی از عیب نمی‌باشد.

نخست اینکه انجام آزمایشها تحت شرایط یکسان و تشخیص دقیق این شرایط، اگر ظاهراً مشکلی نباشد، در عمل پر از ابهام و اشکال است. مثلاً هیچکس نمی تواند ادعا کند که می تواند صدبازی شیر و خط را تحت شرایط یکسان انجام دهد، زیرا هزاران عامل ناشناخته چون فشار هوا، قدرت بازو، خشکی و غیره در کار است. حتی اگر این کار را با آخرین دستگاه های پیشرفته الکترونیکی انجام دهیم، باز هم کنترل شرایط یکسان مقدور نخواهد بود. دوم اینکه تعداد آزمایشهایی را که، باید زیاد باشد، نمی توان قطعی کرد. بنابراین نه تنها احتمال واقعی را که عددی است ایده آل و مجهول، بلکه تخمین معقولی از آن را هم نمی توان بدون داشتن ضوابطی مقبول پیدا کرد. گذشته از این دو کمبود، تعبیر احتمال به طریق فراوانی نسبی، زمانی معنی دارد که بتوان آزمایشی را تکرار کرد. بسیاری از امور به سادگی و یا اصلاً قابل تکرار نیستند. مثلاً احتمال اینکه در کره مریم موجود زنده باشد، به طریق فراوانی نسبی قابل تعبیر نمی باشد.

ج - تعبیر احتمال به طریق شخصی

اگر احتمال یک پیشامد به عنوان میزان یقین و عقیده شخص بر اساس اطلاعات و تجربیات و احساسی که نسبت به وقوع آن دارد تلقی گردد، می گویند احتمال به طریق شخصی تعبیر شده است. از اینرو تعبیر احتمال به طریق شخصی را تعبیر ذهنی، در برابر تعبیر عینی که قبلاً شرح دادیم، می نامند. وقتی پزشک خانوادگی یک بیمار می گوید احتمال این که بیمارش زنده بماند زیاد است، چنین اظهاری صرفاً حاکی از اطلاعات و عقیده این پزشک می باشد. ممکن است نظر یک فرد غیر متخصص، یا حتی پزشک دیگر، در این باره چنین نباشد.

مشکلی که با این تعبیر پیش می آید این است که احتمال اغلب جنبه کیفی و احساسی دارد. مثلاً وقتی می گوئیم احتمال می رود فردا ابری باشد، یا این جمله نمی توان میزان و مقدار احتمال را تعیین کرد. برای پایه ریزی تئوری احتمال لازم است این جنبه کیفی را به نحوی به جنبه کمی، که اندازه گیری در آن مطرح باشد، تبدیل کرد. بدین منظور عده ای احتمال شخصی را با شرط بندی پولی، که کاری است بر خلاف موازین اخلاقی، توجیه کرده اند. تنها با ذکر یک مثال به شرح این توجیه می پردازیم.

فرض کنید E پیشامد آفتابی بودن فردا و متمم آن، یعنی " E' ، پیشامد بارانی بودن فردا باشد. وقتی می گوئیم احتمال E برابر $1/6$ است، این احتمال را به طریق شخصی بدین نحو توجیه می کنیم: میزان یقین، در مورد آفتابی بودن به قدری است که حاضریم 6 به 4 ، بر سر آفتابی بودن، با شما شرط بندی کنیم. به سخی دیگر، اگر هوا بارانی باشد به شما 6 تومان می بازیم و اگر هوا آفتابی باشد از شما 4 تومان می بریم. برای حفظ تعادل، شما هم که طرف مقابل می باشید 4 به 6 ، بر سر بارانی بودن، این شرط بندی را می پذیرید. می گویند اینگونه شرط بندی منصفانه است زیرا سود متوسط هر طرف می شود

$$0 = (6/0) + (4/0) - (-6) = (6/0) + (-4) + (0/4) = 0$$

به طور کلی هرگاه در یک شرط بندی منصفانه، بر سر پیشامد E ، حاضر باشیم a تومان به b تومان با شما شرط بندی کنیم، $p = \frac{a}{a+b}$ را احتمال از نظر ما یا احتمال شخصی می نامند. اغلب می گویند امتیاز یا بخت پولی به نفع E ، در این شرط بندی a می باشد. کسر $\frac{a}{b} = \frac{p}{1-p}$ را نسبت بخت پولی به نفع E می خوانند.

در این کتاب ما به تعریف ریاضی احتمال، که دور از گفتگو است، بسنده می کنیم. البته در ضمن مثالها، جسته و گریخته، به تعبیرهای بالا اشاره خواهیم کرد.

تعریف ریاضی احتمال را نخست در مورد فضای نمونه با پایان، با زبانی ساده شرح می دهیم تا این ایده ها به قدر کافی روشن شوند. بعد از اینکه خواننده آمادگی پیدا کرد، احتمال در یک فضای نمونه دلخواه را، با روش اصل سازی، که روشی متداول در ریاضی نوین است، از نظر می گذرانیم.

۲.۲.۲ تمرین بخش دو

۱. در ظرفی چهار مهره یکسان به رنگهای سفید، سبز، قرمز، و سیاه موجود است. یکی از مهره ها را به تصادف از این ظرف بیرون می آوریم. احتمال پیشامد سفید بودن این مهره $1/4$ است. منظور از این عدد را به راه های گوناگون شرح دهید.

۲. دو تاس را با هم می ریزیم. اگر موضوع مورد علاقه خاله های هر تاس باشد، به هر پیشامد ساده چقدر احتمال نسبت می دهید؟ اگر احتمال پیشامد "جفت شدن" $1/36$ باشد منظور از این عدد را به راه های گوناگون شرح دهید.

۳. یک بسته کارت ۵۲ تایی، به شماره ۱ تا ۵۲، را به خوبی مخلوط می کنیم. سپس یک کارت به تصادف بیرون می آوریم. احتمال اینکه کارت بیرون آمده شماره ۱۰ داشته باشد چیست؟ اگر این کار را ۱۰۴۰ بار تکرار کنند انتظار دارید چندبار چنین کارتی بیرون آید؟

۴. یک آزمایش تصادفی را تحت شرایط یکسان n بار مستقلاً تکرار می کنیم. فرض کنید $R(E)$ فراوانی نسبی پیشامد دلخواه E در این آزمایش باشد. ثابت کنید که

الف - $0 \leq R(E) \leq 1$

ب - $R(E') = 1 - R(E)$

$$R(E \cup F) = R(E) + R(F)$$

ج - برای هر دو پیشامد جدای E و F

در یک مسابقه اسب دوانی شخصی می‌گوید احتمال اینکه اسب او برنده شود $\frac{1}{4}$ می‌باشد و حاضر است با شما در یک شرط‌بندی منصفانه بر سر این پیشامد شرکت کند. منظور از این شرط‌بندی منصفانه و احتمال یاد شده را، که احتمال شخصی است، شرح دهید. اگر این شخص ۱۵۰۰ تومان به ۲۰۰۰ تومان، بر سر بردن اسبش، شرط‌بندی کند، احتمال برنده شدن اسب او را پیدا کنید.

اگر احتمال "شخصی" پیشامد E برابر $0/875$ باشد، شرط‌بندی بر سر E چگونه است؟

۳.۲

مدل احتمال با فضای نمونه با پایان

یک مدل احتمال با فضای نمونه با پایان عبارت است از:

الف - مجموعه ناتهی $S = \{e_1, e_2, \dots, e_k\}$ به نام فضای نمونه یا فضای پیشامدهای ساده.

ب - عددهای مثبت p_1, p_2, \dots, p_k با مجموع یک، به طوری که هر p_i با e_i متناظر باشد.

عدد مثبت p_i را احتمال پیشامد ساده e_i می‌نامیم و می‌نویسیم $P(e_i) = p_i$.

اگر تمام احتمال را به عنوان یک واحد جرم که میان اعضای مجموعه S توزیع شده است تصور نمائیم، می‌توان p_i را به عنوان جرم e_i تعبیر کرد.

۱.۳.۲

احتمال هر پیشامد

اگر فضای نمونه با پایان باشد، هر پیشامد زیر مجموعه‌ای از S و هر زیر مجموعه S یک پیشامد می‌باشد. احتمال پیشامد E برابر است با مجموع احتمالهای پیشامدهای ساده‌ای که E را تشکیل می‌دهند. مثلاً اگر $E = \{e_1, e_2, e_5\}$ ، آنگاه $P(E) = p_1 + p_2 + p_5$ ، که می‌توان آن را به جرم E تعبیر کرد. واضح است که $P(S) = 1$ و $P(\emptyset) = 0$.

۲.۳.۲

مدل احتمال یکنواخت

اگر تمام پیشامدهای ساده دارای احتمالهای مساوی باشند، یعنی

$$p_1 = p_2 = \dots = p_k = \frac{1}{k}$$

می‌گوئیم مدل احتمال یکنواخت است، هرگاه در مدل یکنواخت پیشامد E از m پیشامد ساده و فضای نمونه از k پیشامدهای ساده تشکیل شده باشند، داریم

$$P(E) = \frac{1}{k} + \frac{1}{k} + \dots + \frac{1}{k} = \frac{m}{k} = \frac{E \text{ در } S}{S \text{ در } S}$$

به سخنی دیگر، در یک مدل احتمال یکنواخت، احتمال E برابر است با نسبت تمام حالت‌های مساعد برای رخ دادن E به تمام حالت‌های ممکن.

۳.۳.۲ چند فرمول برای محاسبه احتمال

اگر در نمودارهای ون (نگاره ۱) مساحت هر مجموعه را به عنوان احتمال پیشامد نظیر آن تلقی کنیم، با محاسبه قسمت‌های هاشوردار معلوم می‌شود که:

$$P(F - E) = P(F) - P(E \cap F)$$

$$P(E') = 1 - P(E)$$

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

به ویژه اگر $E \subset F$ ، آنگاه

$$P(F - E) = P(F) - P(E)$$

و اگر $E \cap F = \emptyset$ ، آنگاه

$$P(E \cup F) = P(E) + P(F)$$

اثبات ریاضی فرمولهای بالا را در حالت کلی در آینده خواهیم دید.

مثال ۸ در ظرفی دو مهره سفید، سه مهره سبز، و دو مهره قرمز داریم. مهره‌ها را که گذشته از رنگ کاملاً یکسان هستند، به خوبی مخلوط می‌کنیم و سپس یکی از آنها را چشم بسته بیرون می‌آوریم. فرض کنید در این آزمایش موضوع مورد علاقه رنگ مهره باشد. اگر مهره سفید و سبز و قرمز را به ترتیب با W و G و R نشان دهیم، فضای نمونه می‌شود $S = \{W, G, R\}$. با توجه به تعداد مهره‌ها و رنگ آنها در این آزمایش داریم:

$$P(R) = \frac{2}{5}, \quad P(G) = \frac{3}{5}, \quad P(W) = \frac{2}{5}$$

جدول ۲ فراوانی و فراوانی نسبی رنگ چشم در نمونه ۵۰۰۰ نفری

| رنگ چشم | جمع کل | سیاه | سبز | قهوه‌ای | آبی |
|--------------|--------|--------|--------|---------|--------|
| فراوانی | ۵۰۰۰ | ۱۵۰۰ | ۵۰۱ | ۲۰۰۰ | ۹۹۹ |
| فراوانی نسبی | ۱/۵۰۰۰ | ۰/۳۰۰۰ | ۰/۱۰۰۲ | ۰/۴۰۰۰ | ۰/۱۹۹۸ |

پاسخ به تمام این پرسشها موضوع آمار و احتمال است. برای اینکه مطلب روشن شود، اینک به مثال زیر می‌پردازیم.

مثال ۱۱ مردم یک شهر بزرگ دارای چشمان آبی، قهوه‌ای، سبز و سیاه می‌باشند. یک نفر را به تصادف در نظر گرفته فرض می‌کنیم موضوع مورد علاقه، رنگ چشم باشد. در این آزمایش فضای نمونه می‌شود

$$S = \{\text{آبی، قهوه‌ای، سبز، سیاه}\}$$

برای اینکه احتمال پیشامدهای ساده را پیدا کنیم به آمارگیری می‌پردازیم. مثلاً یک نمونه تصادفی ۵۰۰۰ تایی از مردم شهر را در نظر می‌گیریم و رنگ چشمان آنها را صورت‌برداری می‌کنیم. فرض کنید جدول ۲ را به دست آید.

اینک با استفاده از تعبیر احتمال به طریق فراوانی نسبی، از جدول بالا تقریباً داریم:

$$P(\text{سیاه}) = 0/3, P(\text{سبز}) = 0/1, P(\text{قهوه‌ای}) = 0/4, P(\text{آبی}) = 0/2$$

پس مدل احتمال تقریبی رنگ چشم مردم شهر را به کمک نمونه‌برداری به دست آوردیم. طبیعی است که هر قدر نمونه بزرگتر باشد، این مدل احتمال تقریبی، بهتر خواهد بود. به خصوص اگر از رنگ چشم تمام مردم شهر جداول فراوانی بسازیم، مدل احتمال صددرصد دقیق خواهد بود، ولی عملاً این کار مقدور نیست.

اینک با داشتن این مدل تقریبی، به این پرسش پاسخ می‌دهیم: اگر جمعیت شهر در حدود ۴۰۰۳۰۰ نفر باشد، تقریباً چند نفر دارای چشمان آبی هستند؟
احتمال پیشامد چشم آبی ۰/۲ است. بنابراین اگر احتمال را به طریق فراوانی نسبی تعبیر کنیم، پاسخ تقریبی می‌شود:

$$۸۰۰۶۰ = ۴۰۰۳۰۰ \times 0/2$$

ملاحظه می‌شود که مدل احتمال یکتواخت نیست.

حال اگر هویت مهره‌های هم‌رنگ را مشخص کنیم، یعنی دو مهره سفید را با W_1, W_2 ، سه مهره سبز را با G_1, G_2, G_3 ، و دو مهره قرمز را با R_1, R_2 نشان دهیم، آنگاه فضای نمونه با توجه به رنگ و هویت مهره‌ها، به صورت:

$$S = \{W_1, W_2, G_1, G_2, G_3, R_1, R_2\}$$

در می‌آید. اینک نظر به همشانسی پیشامدهای ساده، می‌توان به هر پیشامد ساده $\frac{1}{7}$ احتمال نسبت داد و مدل احتمال را یکتواخت تصور کرد.

با این مدل یکتواخت می‌توان احتمال را در مدل قبلی توجه کرد. مثلاً پیشامد مهره سبز در این مدل یکتواخت می‌شود $G = \{G_1, G_2, G_3\}$ با احتمال $P(G) = \frac{3}{7}$. توجه کنید که G یک پیشامد ساده در مدل اول، اما یک پیشامد مرکب در مدل دوم است.

مثال ۹ در بازی شیر و خط، سکه نالرایی را دو بار مستقلاً، یعنی بدون اینکه بار اول در بار دوم تاثیر کند، می‌اندازیم. در این آزمایش فضای نمونه می‌شود

$$S = \{TT, HT, TH, HH\}$$

که در آن مثلاً منظور از پیشامد ساده HT این است که سکه بار اول شیر و بار دوم خط بیاید. با توجه به تالیب بودن سکه و مستقل بودن بازها، پیشامدهای ساده همشانسی بوده به هر کدام احتمال $\frac{1}{4}$ تعلق می‌گیرد، یعنی، مدل احتمال یکتواخت است.

مثال ۱۰ در یک آزمایش تصادفی سکه نالرایی را یک بار می‌اندازیم، اگر شیر بیاید آزمایش را پایان می‌دهیم، و گرنه سکه را یک بار دیگر انداخته آنگاه آزمایش را پایان می‌دهیم. در این آزمایش فضای نمونه می‌شود $S = \{H, TH, TT\}$ با احتمالهای $P(H) = \frac{1}{4}$ ، $P(TH) = \frac{1}{4}$ و $P(TT) = \frac{1}{4}$. ملاحظه می‌شود که مدل احتمال یکتواخت نیست.

۲.۳.۲ روش تعیین مدل احتمال

بر خلاف مثالهای بالا که همگی ساده و مصنوعی بودند، در عمل با پدیده‌هایی روبرو می‌شویم که مدل احتمال آنها را نمی‌توان به سادگی تعیین کرد. بنابراین پرسشهای زیر معقول و منطقی می‌باشند:
چگونه می‌توان مدل احتمال یک آزمایش را پیدا کرد؟ چگونه می‌توان قضاوت کرد که فلان مدل احتمال برای فلان آزمایش الگویی برانزده است؟ آیا فایده مدل احتمال چیست؟

۵.۳.۲ محاسبه احتمال در مدل یکنواخت

همانطوری که قبلاً اشاره شد، هرگاه در مدل یکنواخت پیشامد E از m پیشامد ساده و فضای نمونه از k پیشامد ساده تشکیل شده باشند، $P(E) = \frac{m}{k}$. بنابراین محاسبه احتمال E بستگی به محاسبه m و k دارد. در بعضی موارد مستقیماً با نوشتن تمام اعضای مجموعه‌های S و E می‌توان شمارش را به سادگی انجام داد. ولی گاهی این کار مقدور نیست. برای اینکه مطلب روشن شود به مثال زیر توجه کنید.

مثال ۱۲ در ظرفی دو مهره سفید W_1 و W_2 و سه مهره سیاه B_1, B_2, B_3 داریم که همگی صرفنظر از رنگ و هویت یکسان هستند. مهره‌ها را به خوبی مخلوط کرده به تصادف دو مهره با هم از ظرف بیرون می‌آوریم. فضای نمونه با توجه به رنگ و هویت مهره‌ها می‌شود:

$$S = \{W_1W_2, B_1B_2, B_1B_3, B_2B_3, B_1W_1, B_2W_1, B_1W_2, B_2W_2, B_3W_2, B_3W_1\}$$

که دارای ده عضو است. مثلاً منظور از B_1W_1 که عین B_1B_2 می‌باشد، این است که مهره سیاه شماره دو و مهره سفید شماره یک با هم بیرون آیند.

فرض کنید که می‌خواهیم احتمال پیشامد مهره‌های هم‌رنگ را حساب کنیم. اگر این پیشامد را E بنامیم، داریم:

$$E = \{W_1W_2, B_1B_2, B_1B_3, B_2B_3\}$$

$$P(E) = \frac{4}{10} = 0.4$$

که دارای چهار عضو است. بنابراین $P(E) = 0.4$. حال اگر به جای دو مهره سفید و سه مهره سیاه دوست مهره سفید و سیصد مهره سیاه در ظرف داشته باشیم، شمارش تعداد اعضای S و E ، مستقیماً از راه نوشتن تمام اعضای این دو مجموعه عملاً محال است. بلکه باید چاره‌اندیشی کرد و بطور غیر مستقیم تعداد اعضای این دو مجموعه را تعیین نمود. خوشوقتانه قوانین شمارش، که در درسی بنام آنالیز ترکیبی مطرح می‌شوند، این مشکل را حل می‌کنند. نظر به اهمیت موضوع، بخش ۴ را به قوانین شمارش اختصاص داده‌ایم.

۶.۳.۲ تمرین بخش سه

۱. یک سکه نازیب‌راسه بار می‌اندازیم. مدل احتمال را تعیین کنید. احتمال این که شیر و خط متناوباً مشاهده شوند چقدر است؟ احتمال این که حداقل یک بار شیر مشاهده شود چقدر است؟

۲. یک تاس ارباب داریم که اگر آن را به تصادف پرتاب کنیم، شانس مشاهده هر وجه آن متناسب با تعداد خالها، در آن وجه می‌باشد. مدل احتمال را برای این تاس پیدا کنید. احتمال پیشامد خال زوج چقدر است؟

۳. یک تاس را هزار بار انداخته یافته‌ها را در جدول زیر خلاصه کرده‌ایم:

| | | | | | | |
|--------------|-----|-----|-----|-----|-----|-----|
| خال | ۱ | ۲ | ۳ | ۴ | ۵ | ۶ |
| فراوانی | ۲۱۴ | ۱۵۲ | ۱۷۸ | ۱۸۸ | ۱۶۳ | ۱۰۵ |
| فراوانی نسبی | | | | | | |

این جدول را تکمیل کرده مدل احتمال را برای این تاس به تقریب تعیین کنید. اگر این تاس را ۲۰۰۰ بار پرتاب کنیم، انتظار دارید چند بار خال مضرب سه دیده شود؟

۴. یک زن و شوهر جوان آرزو دارند در آینده سه فرزند داشته باشند. احتمال اینکه فرزند اول و آخر پسر باشند چقدر است؟

۵. در ظرفی دو مهره سفید و سه مهره سیاه داریم، که همگی گذشته از رنگ و هویت یکسان می‌باشند. مهره‌ها را به خوبی مخلوط می‌کنیم و به تصادف سه مهره با هم از ظرف بیرون می‌آوریم. مدل احتمال را با توجه به رنگ و هویت مهره‌ها تعیین کنید.

۶. در آزمایشی سکه نازیبی را دوبار می‌اندازیم. اگر هر دو بار شیر بیاید آزمایش را پایان می‌دهیم، و گرنه سکه را یک بار دیگر انداخته آنگاه آزمایش را پایان می‌دهیم. مدل احتمال را برای این آزمایش تعیین کنید.

۷. مستقیماً و به کمک نمودار ون $P(E \Delta F)$ را پیدا کنید.

۸. با فرض $P(E) = 0.6$, $P(F) = 0.7$, $S = E \cup F$ ، مطلوبست محاسبه $P(E' - F)$, $P(E \cap F)$, $P(F - E)$, $P(E \Delta F)$

۹. فرض کنید E و F جدا باشند و $P(E) = 0.3$ و $P(F) = 0.4$. مطلوبست محاسبه

$$P(E \cup F), P(E - F), P(E' - F'), P(E \Delta F)$$

۱۰. با استفاده از فرمول $P(A \cup B)$ و قوانین شرکت‌پذیری و توزیمی یک فرمول برای $P(E \cup F \cup G)$ پیدا کنید. با نمودار ون درستی فرمول را نشان دهید.

۴.۲ قوانین شمارش

ساده‌ترین نوع محاسبه، شمارش تعداد اعضای یک مجموعه با پایان کوچک است که انسان به طور طبیعی از طریق انگشتان دست، برای اولین بار به آن توجه کرده است. با این حال اگر این مجموعه با پایت خیلی پر عضو باشد، به طوری که نتوانیم یک یک اعضای آن را نشان دهیم، عمل شمارش نه تنها از عهده انسان بلکه از عهده کامپیوترهای بزرگ هم ساخته نیست. خوشبختانه به کمک اندیشه ریاضی می‌توان فرمولهائی ابداع کرد که از دشواری کار شمارش می‌کاهند فن پیدا کردن این فرمولها بر دو اصل بدیهی به نام اصول شمارش بنا شده است.

۱.۴.۲ اصول شمارش

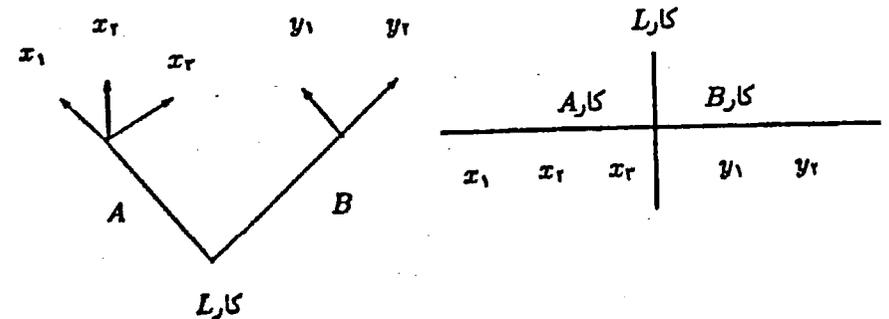
فرض کنید کار A را به m طریق یا نامهای x_1, x_2, \dots, x_m و کار B را به n طریق یا نامهای y_1, y_2, \dots, y_n بتوان انجام داد. اصول شمارش عبارتند از:

الف - اصل جمع برای شمارش

اگر انجام کار L منوط به انجام کار A یا B باشد، آنگاه کار L را می‌توان به $m + n$ طریق به نامهای $x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n$ انجام داد.

این اصل را اصل جمع می‌گویند و در آن تکیه بر روی یا است. به زبان مجموعه‌ها اگر مجموعه A دارای m عضو و مجموعه B دارای n عضو باشد، آنگاه مجموعه $A \cup B$ که اجتماع دو مجموعه جدای A و B است، دارای $m + n$ عضو است.

درستی این اصل را می‌توان به کمک جدول یا نمودار زیر برای $m = 3$ و $n = 2$ تجسم کرد



به عنوان مثال فرض کنید رفتن از دانشکده به منزل (کار L)، با تاکسی (کار A) از سه راه یا با اتوبوس (کار B) از دو راه امکان داشته باشد. بنابراین به شش طریق با تاکسی و با اتوبوس می‌توان از دانشکده به منزل رفت.

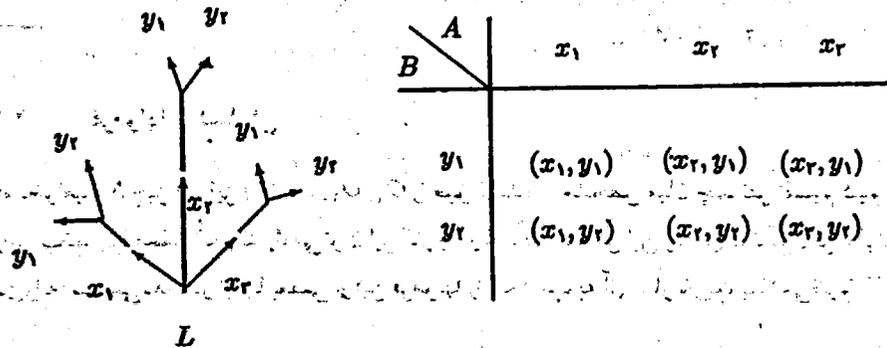
اتوبوس (کار B) از دو راه امکان داشته باشد. بنابراین به پنج طریق با تاکسی یا با اتوبوس می‌توان از دانشکده به منزل رفت.

ب - اصل ضرب برای شمارش

اگر انجام کار L منوط به انجام پیاپی کار A و B باشد، آنگاه کار L را می‌توان به mn طریق به نامهای $(x_1, y_1), (x_1, y_2), \dots, (x_m, y_n)$ انجام داد.

این اصل را اصل ضرب می‌گویند و در آن تکیه بر روی و می‌باشد. به زبان مجموعه‌ها اگر مجموعه A دارای m عضو و مجموعه B دارای n عضو باشد، آنگاه مجموعه $A \times B$ ، که حاصلضرب دکارتی دو مجموعه A و B می‌باشد، دارای mn عضو است.

درستی این اصل را می‌توان به کمک جدول یا نمودار زیر مثلاً برای $m = 3$ و $n = 2$ تجسم کرد



به عنوان مثال فرض کنید رفتن از دانشکده به منزل (کار L)، با تاکسی از دانشکده تا پارک شهر (کار A) از سه راه و با اتوبوس از پارک شهر تا منزل (کار B) از دو راه امکان داشته باشد. بنابراین به شش طریق با تاکسی و با اتوبوس می‌توان از دانشکده به منزل رفت.

این دو اصل را به جای دو کار A و B می‌توان به k کار A_1, A_2, \dots, A_k تعمیم داد.

مثال ۱۳ می‌خواهیم از پنج بازرگان، شش دبیر و چهار قاضی یک جلسه دو نفری تشکیل دهیم به طوری که اعضای جلسه دارای مشاغل مختلف باشند. جلسه را به چند طریق می‌توان تشکیل داد؟

طبق اصل دوم، این جلسه از بازرگان و دبیر به $5 \times 6 = 30$ طریق یا از بازرگان و قاضی به $5 \times 2 = 10$ طریق یا از دبیر و قاضی به $6 \times 2 = 12$ طریق تشکیل می‌شود. بنابراین طبق تعمیم اصل اول، جلسه را می‌توان به $30 + 10 + 12 = 52$ طریق تشکیل داد.

مثال ۱۴ بازی شیر و خط را ده بار مستقلاً با سکه‌ای نالریب انجام می‌دهیم. فضای نمونه از چند پیشامد ساده تشکیل می‌شود؟ احتمال پیشامد یک در میان شیر و خط چقدر است؟

این ده بازی شیر و خط متناظر با ده کار A_1, A_2, \dots, A_{10} می‌باشند که هر یک به دو طریق T و H انجام می‌پذیرد. بنابراین طبق اصل ضرب تعداد پیشامدهای ساده در فضای پیشامدها برابر است با 2^{10} . این پیشامدهای ساده همشانس می‌باشند و احتمال هر یک برابر است با 2^{-10} .

پیشامد یک در میان شیر و خط، از دو پیشامد ساده که یکی با H و دیگری با T آغاز می‌شود، تشکیل می‌گردد. بنابراین احتمال آن برابر است با $2^{-9} = 0.0019$.

مثال ۱۵ از اعضای مجموعه $\{1, 2, 3, 4, 5\}$ چند سه‌تایی مرتب (x, y, z) می‌توان ساخت؟ x, y, z را هر کدام به پنج طریق می‌توان برگزید. بنابراین تعداد سه‌تایی‌های مرتب می‌شود $5 \times 5 \times 5 = 125$. هر سه‌تایی مرتب را می‌توان به منزلهٔ مختصات قائم یک نقطه در فضای سه بعدی پنداشت.

۲.۲.۲ فرمولهای شمارش

هرگاه بخواهیم N چیز متمایز O_1, O_2, \dots, O_N را طبق قاعده‌ای مشخص میان چند نفر تقسیم کنیم یا آنها را مرتب‌نمائیم، معمولاً به چند راه می‌توان این کار را انجام داد. شمارش این راه‌های گوناگون فرمولهای مفیدی را به دست می‌دهد. در اینجا بعضی از این فرمولها را که اغلب به آنها نیاز داریم، پیدا می‌کنیم.

الف - جایگشت N چیز

ترتیبی را که می‌توان N چیز متمایز را از چپ به راست پهلوی هم گذاشت، یک جایگشت از این N چیز می‌گویند. مثلاً $O_1 O_2 O_3$ یک جایگشت از سه چیز $O_1 O_2 O_3$ است.

تعداد جایگشتهای N چیز متمایز را بدین طریق پیدا می‌کنیم:

برای ساختن یک جایگشت از N چیز متمایز باید آنها را یک به یک انتخاب کرده از چپ به راست پهلوی هم قرار دهیم. چیز اول را به N طریق (کار A_1)، چیز دوم را به $N-1$ طریق (کار A_2)، ... و چیز N ام را به یک طریق (کار A_N) می‌توان انتخاب کرد. طبق اصل ضرب، ساختن یک جایگشت معادل است با انجام پیاپی کارهای A_1, A_2, \dots, A_N که به $1 \times (N-1) \times \dots \times N$ طریق انجام می‌گیرد. این حاصلضرب را با $N!$ نشان می‌دهند و آن را N فاکتوریل می‌خوانند. بنابراین تعداد جایگشتهای N چیز

تمتایز می‌شود.

$$1 \times 2 \times \dots \times (N-1) \geq N = N! \quad (۱)$$

ب - جایگشت R تائی از N چیز

هرگاه از N چیز متمایز R تا را برگزیده، $1 \leq R \leq N$ ، به ترتیب از چپ به راست پهلوی هم قرار دهیم، آن را یک جایگشت R تائی از این N چیز می‌گویند. مثلاً $O_1 O_2 O_3$ یک جایگشت دو تائی از سه چیز O_1, O_2, O_3 می‌باشد. بنابراین $O_1 O_2$ و $O_2 O_1$ به خاطر ترتیب با هم فرق دارند.

تعداد جایگشتهای R تائی N چیز متمایز را بدین طریق پیدا می‌کنیم: برای ساختن یک جایگشت R تائی از N چیز متمایز باید R تا از این N چیز را یک به یک انتخاب کرده از چپ به راست پهلوی هم قرار دهیم. طبق اصل ضرب این کار معادل است با انجام پیاپی R کار $A_1, A_2, \dots, A_{R-1}, A_R$ که به $(N-R+1) \times (N-R+2) \times \dots \times N$ طریق انجام می‌گیرد. بنابراین تعداد جایگشتهای R تائی از N چیز متمایز می‌شود.

$$N \times (N-1) \times \dots \times (N-R+2) \times (N-R+1)$$

این حاصلضرب را با $(N)_R$ یا $P_{N,R}$ نشان می‌دهند. از گسترش $N!$ و $(N-R)!$ به صورت حاصلضرب نتیجه می‌شود که:

$$P_{N,R} = N \times (N-1) \times \dots \times (N-R+2) \times (N-R+1) = \frac{N!}{(N-R)!} \quad (۲)$$

اگر R را در این فرمول برابر N و $N-1$ بگیریم تساویهای زیر به دست می‌آیند:

$$P_{N,N} = N! = \frac{N!}{1}$$

$$P_{N,N-1} = N! = \frac{N!}{1}$$

برای اینکه تساویهای بالا معنی داشته باشند، قرارداد می‌کنیم که:

$$0! = 1! = 1$$

ج - ترکیب R تائی از N چیز

هرگاه از N چیز متمایز، یک گروه R تائی را با هم یا یک به یک بدون توجه به ترتیب برگزینیم، آن را یک ترکیب R تائی از این N چیز می گویند. مثلاً $O_1 O_2 O_3$ ، یک ترکیب دو تائی از سه چیز O_1, O_2, O_3 می باشد. بنابراین $O_1 O_2$ و $O_2 O_1$ ، به خاطر عدم توجه به ترتیب، با هم فرق ندارند.

تعداد ترکیبهای R تائی N چیز متمایز را بدین طریق پیدا می کنیم: چون در برابر تمام $R!$ جایگشتهای R تائی N چیز، تنها یک ترکیب R تائی داریم، بنابراین تعداد ترکیبهای R تائی N چیز متمایز که معمولاً آن را با $C_{N,R}$ یا $\binom{N}{R}$ نشان می دهند می شود:

$$C_{N,R} = \binom{N}{R} = \frac{P_{N,R}}{R!} = \frac{N!}{R!(N-R)!} \tag{۳}$$

۳.۴.۲ تعبیر جایگشت و ترکیب به زبان مجموعه ها

به زبان مجموعه ها، اگر N چیز متمایز را با مجموعه $\{O_1, O_2, \dots, O_N\}$ نشان دهیم، آنگاه:

الف - هر جایگشت از این N چیز، یک N تائی مرتب از اعضای این مجموعه است.

ب - هر جایگشت R تائی از این N چیز، یک R تائی مرتب از اعضای این مجموعه است.

ج - هر ترکیب R تائی از این N چیز، یک زیر مجموعه R عضوی از اعضای این مجموعه است.

مثلاً سه چیز متمایز O_1, O_2, O_3 را با مجموعه $\{O_1, O_2, O_3\}$ نشان می دهیم. سه تائی مرتب $\{O_1, O_2, O_3\}$ یک جایگشت، دو تائی مرتب $\{O_2, O_1\}$ یک جایگشت دو تائی، و زیر مجموعه دو عضوی $\{O_2, O_1\}$ یک ترکیب دو تائی از این مجموعه می باشند.

مثال ۱۶ یک خانواده پنج نفری می توانند:

الف - به $۱۲۰ = ۵!$ طریق (صف های پنج نفری تشکیل دهند. (تعداد جایگشتها)

ب - به $P_{۵,۳} = ۶۰$ طریق صف های سه نفری تشکیل دهند. (تعداد جایگشتهای سه تائی)

ج - به $C_{۵,۳} = ۱۰$ طریق گروه های سه نفری تشکیل دهند. (تعداد ترکیب های سه تائی)

توجه!

فرمول ترکیب در مورد این مثال هم مفید است: N مهره هم شکل داریم که R تا از آنها سفید (غیر متمایز از هم) و بقیه سیاه (غیر متمایز از هم) هستند. از این مهره ها می توان $C_{N,R}$ جایگشت تشکیل داد. (چرا؟)

۴.۴.۲ فرمول استرلینگ

محاسبه واقعی $N!$ که اغلب در عمل به آن نیاز داریم، برای N های بزرگ خسته کننده است. مقدار تقریبی $N!$ از فرمول زیر، به نام فرمول استرلینگ و با استفاده از لگاریتم گیری به آسانی به دست می آید:

$$N! \approx A(N) = \left(\frac{N}{e}\right)^N \sqrt{2\pi N} \tag{۴}$$

خطای نسبی این تقریب برابر است با

$$E = \frac{|N! - A(N)|}{N!}$$

این خطای نسبی حداکثر برابر است با

$$\frac{1}{(12N-1)}$$

این فرمولها در درسهای پیشرفته احتمال یا ریاضی اثبات می شوند.

مثال ۱۷ مقدار واقعی $۱۰!$ می شود ۳۶۲۸۸۰۰ و مقدار تقریبی آن به کمک فرمول استرلینگ اینگونه به دست می آید:

$$\ln A(10) = 10(\ln 10 - 1) + 0.5(\ln 2\pi + \ln 10) = 15.096082$$

بنابراین داریم

$$10! \approx A(10) = 359896$$

خطای نسبی در حدود 0.008 است.

۵.۴.۲ شمارش با مدلهای جعبه و مهره

فرض کنید تعداد N جعبه متمایز را بخواهیم با تعداد R مهره برگزیم. برحسب اینکه مهره ها متمایز یا غیر متمایز باشند و در هر جعبه گذاشتن مهره مکرر مجاز یا غیر مجاز باشد، چهار مدل داریم.

حوت و با جاگشتها
الف - مهره‌ها متمایز و مهره مکرر مجاز است

در این مدل تعداد روشهای پر کردن جعبه‌ها را با $C(N, R)$ نشان می‌دهیم و آن را بدین طریق پیدا می‌کنیم: چون R مهره متمایز داریم که مجازیم هر کدام را در هر یک از N جعبه بگذاریم، بنابراین پر کردن جعبه‌ها معادل با انجام پیاپی R کار است که هر یک را به N طریق می‌توان انجام داد. پس طبق اصل ضرب داریم:

$$C(N, R) = N^R$$

مثلاً دو مهره a و b را می‌توان به نه طریق در سه جعبه سبز و سفید و قرمز گذاشت.

ب - مهره‌ها متمایز و مهره مکرر غیر مجاز است

در این حالت باید داشته باشیم $R \leq N$ ، تعداد روشهای پر کردن جعبه‌ها را با $C(\bar{N}, R)$ نشان داده و آن را بدین طریق پیدا می‌کنیم: برای پر کردن جعبه‌ها نخست R جعبه را از بین N جعبه انتخاب کرده (کار A_1) و سپس R جعبه انتخاب شده را با R مهره متمایز یک به یک پر می‌کنیم (کار A_2). چون کار A_1 را به $\binom{N}{R}$ طریق و کار A_2 را به $R!$ طریق می‌توان انجام داد، پس طبق اصل ضرب داریم:

$$C(\bar{N}, R) = \binom{N}{R} R! = \frac{N!}{(N-R)!}$$

این درست برابر تعداد جایگشتهای R تایی N چیز متمایز است.

مثلاً دو مهره a و b را می‌توان به شش طریق در سه جعبه سبز و سفید و قرمز قرار داد، به طوری که در هر جعبه حداکثر یک مهره گذاشته شود.

ج - مهره‌ها غیر متمایز و مهره مکرر غیر مجاز است

در این مدل، که باید داشته باشیم $R \leq N$ ، تعداد روشهای پر کردن جعبه‌های را با $C(\bar{N}, \bar{R})$ نشان داده آن را بدین طریق پیدا می‌کنیم: برای پر کردن جعبه‌ها نخست R جعبه را از بین N جعبه انتخاب می‌کنیم (کار A_1) و سپس R جعبه انتخاب شده را با R مهره غیر متمایز یک به یک پر می‌کنیم (کار A_2). چون کار A_1 را به $\binom{N}{R}$ طریق و کار A_2 را تنها به یک طریق می‌توان انجام داد، پس طبق اصل ضرب داریم:

$$C(\bar{N}, \bar{R}) = \binom{N}{R} \times 1 = \frac{N!}{R!(N-R)!}$$

این درست برابر تعداد ترکیب‌های R تایی N چیز متمایز است.

مثلاً دو مهره a و a را می‌توان به سه طریق در سه جعبه سبز و سفید و قرمز قرار داد، به طوری که در هر جعبه حداکثر یک مهره گذاشته شود.

تعداد جایگشتهایی که از R چیز غیر متمایز از نوع a و $N-R$ چیز غیر متمایز از نوع b می‌توان تشکیل داد طبق این مدل می‌شود $\binom{N}{R}$ زیرا برای ساختن هر جایگشت می‌توان R جعبه از بین N جعبه متمایز را با a ها و بقیه جعبه‌ها را با b ها پر کرد.

مثلاً از a و a و b و b می‌توان تعداد $10 = \binom{4}{2}$ جایگشت تشکیل داد.

د - مهره‌ها غیر متمایز و مهره مکرر مجاز است

در این مدل تعداد روشهای پر کردن جعبه‌ها را با $C(N, \bar{R})$ نشان داده، آن را بدین طریق پیدا می‌کنیم: برای روشن شدن مطلب فرض کنید بخواهیم پنج جعبه را با سه مهره غیر متمایز پر کنیم. با توجه به شکل زیر اگر پنج جعبه را با شش خط عمودی مجسم کرده و خط اول و آخر را ثابت تصور کنیم، پر کردن این پنج جعبه با این سه مهره معادل است با ساختن یک جایگشت از چهار خط عمودی و سه ستاره، که به $35 = \binom{7}{4}$ روش امکان دارد. (به ترکیب R تایی از N رجوع کنید).



بطور کلی اگر N جعبه را با $N+1$ خط عمودی نشان دهیم و خط اول و آخر را ثابت تصور کنیم، پر کردن این N جعبه به وسیله R مهره معادل است با ساختن یک جایگشت از $N-1$ خط عمودی و R ستاره که به $\binom{N+R-1}{R}$ روش امکان دارد. بنابراین

$$C(N, \bar{R}) = \binom{N+R-1}{R}$$

مثلاً دو مهره a و a را می‌توان به شش روش در سه جعبه سبز و سفید و قرمز گذاشت.

مثال ۱۸ تعداد جوابهای درست و غیر منفی معادله دو مجهولی $X_1 + X_2 = 4$ را، که به صورت بردار (X_1, X_2) نشان داده می‌شوند، پیدا کنید.

پیدا کردن هر جواب معادل است با گذاشتن چهار مهره غیر متمایز در دو جعبه متمایز X_1 و X_2 ، که طبق مدل د جعبه و مهره به $5 = \binom{4+2-1}{2}$ طریق امکان دارد، و این جوابها عبارتند از

- $(0, 4)$ ، $(1, 3)$ ، $(2, 2)$ ، $(3, 1)$ ، $(4, 0)$

بطور کلی معادله N مجهولی $X_1 + X_2 + \dots + X_N = R$ دارای $\binom{N+R-1}{R}$ جواب درست غیر منفی است.

مثال ۱۹ از N حرف O_1, O_2, \dots, O_N چند جایگشت R تائی در صورتی که حرف مکرر مجاز باشد، می توان تشکیل داد؟

چون حرف مکرر مجاز می باشد، بنابراین هر کدام از اعضای جایگشت R تائی را می توان به N طریق انتخاب کرد. پس پاسخ این مساله، مانند مدل الف جعبه و مهره می شود N^R . مثلاً از a, b و c می توان نه جایگشت دو تائی، در صورتی که تکرار مجاز باشد، تشکیل داد که عبارتند از $aa, bb, cc, cb, bc, ca, ac, ba, ab$.

مثال ۲۰ از N حرف متمایز O_1, O_2, \dots, O_N چند ترکیب R تائی، در صورتی که حرف مکرر مجاز باشد، می توان تشکیل داد؟

فرض کنید در یک ترکیب R تائی، که در آن تکرار مجاز است، X_i تعداد دفعاتی باشد که O_i تکرار شده است. بنابراین داریم:

$$X_1 + X_2 + \dots + X_N = R$$

تعداد جوابهای درست و غیر منفی این معادله N مجهولی، یعنی $\binom{N+R-1}{R}$ درست برابر است با تعداد ترکیبهای R تائی N حرف متمایز در صورتی که تکرار مجاز باشد.

مثلاً از a, b و c می توان شش ترکیب دو تائی، در صورتی که تکرار مجاز باشد، تشکیل داد که عبارتند از aa, bb, cc, ab, ac, bc .

مثال ۲۱ فرض کنید N چیز داشته باشیم که همگی از هم متمایز نباشند و بتوان آنها را به k گروه تقسیم کرد، به طوری که N_1 تا از نوع O_1, N_2 تا از نوع O_2, \dots, N_k تا از نوع O_k باشند. فرض می کنیم چیزهای هر گروه غیر متمایز باشند. بنابراین $N_1 + N_2 + \dots + N_k = N$ می خواهیم تعداد جایگشتهائی را که از این N چیز می توان تشکیل داد پیدا کنیم.

برای اینکه یک جایگشت از این N چیز بسازیم، بدین طریق عمل می کنیم: N جعبه متمایز را در یک صف می چینیم. حال N_1 تا از این جعبه ها را انتخاب کرده آنها را با N_1 چیز از نوع O_1 که غیر متمایز می باشند یک به یک پر می کنیم. این کار را می توان به $\binom{N}{N_1}$ طریق انجام داد. سپس N_2 تا از بقیه جعبه ها را انتخاب کرده آنها را با N_2 چیز از نوع O_2 غیر متمایز می باشند یک به یک پر می کنیم.

این کار را می توان به $\binom{N-N_1}{N_2} \dots \binom{N_k}{N_k}$ طریق انجام داد. به همین طریق بقیه جعبه ها را پر می کنیم. پس طبق اصل ضرب جعبه ها را می توان به

$$\binom{N}{N_1} \binom{N-N_1}{N_2} \dots \binom{N_k}{N_k} = \frac{N!}{N_1! N_2! \dots N_k!}$$

طریق که درست برابر تعداد جایگشتهائی می باشد پر کرد. اگر N چیز از هم متمایز بودند، تعداد جایگشتهائی می شد $N!$. ولی به علت تقسیم آنها به گروه های غیر متمایز، این تعداد مطابق فرمول بالا تقلیل پیدا کرده است. اغلب این تعداد را با (N_1, N_2, \dots, N_k) نشان می دهند.

مثلاً از a, b, c, a, a و b می توان تعداد $60 = \frac{6!}{(3!2!1!)}$ جایگشت تشکیل داد.

مثال ۲۲ پنج نفر شیرازی، دو نفر اصفهانی و سه نفر تهرانی روی ده صندلی که در یک ردیف قرار دارند به تصادف می نشینند. احتمال اینکه تمام هم شهریها پهلوی هم باشند چقدر است؟

اگر فرض کنید هم شهریها غیر متمایز هستند این ده نفر به $\frac{10!}{(3!2!5!)}$ طریق همشاس می توانند پهلوی هم بنشینند. هم شهری ها، یعنی شیرازیها و اصفهانیها و تهرانیها به $6 = 3!$ طریق می توانند صندلیها را اشغال کنند. بنابراین احتمال مورد نظر می شود

$$\frac{6}{2520} = \frac{1}{420}$$

مثال ۲۳ دو جمله ای $(x+y)^n$ را بسط دهید.

اگر $(x+y)^n$ را n بار در خودش ضرب کنیم، جمله هائی به فرم $x^k y^{n-k}$ به دست می آیند که در آنها k می تواند یکی از اعداد $0, 1, \dots, n$ را اختیار کند. برای هر k ، تعداد جمله $x^k y^{n-k}$ برابر است با $\binom{n}{k}$ ، زیرا k های این جمله از k پرانتز $(x+y)$ می آیند که می توان آنها را به $\binom{n}{k}$ طریق انتخاب کرد و y ها از بقیه پرانتزها هستند. بنابراین داریم:

$$(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$$

برای اینکه استدلال بالا روشن شود $(x+y)^3$ را مستقیماً بسط دهید.

مثال ۲۴ سه جمله ای $(x+y+z)^n$ را بسط دهید.

اگر $x + y + z = n$ را n بار در خودش ضرب کنیم، جمله‌هایی به فرم $x^{n_1} y^{n_2} z^{n_3}$ به دست می‌آیند که در آنها n_1, n_2, n_3 اعداد درست و غیر منفی هستند و داریم: $n_1 + n_2 + n_3 = n$. برای هر n_1 و n_2 تعداد جمله $x^{n_1} y^{n_2} z^{n_3}$ برابر است با

$$\binom{n}{n_1} \binom{n_1 + n_2}{n_2} = \binom{n}{n_1, n_2, n_3}$$

زیرا x های این جمله از n_1 پرانتز $(x + y + z)$ به $\binom{n}{n_1}$ طریق و y ها از n_2 پرانتز دیگری به $\binom{n_1 + n_2}{n_2}$ طریق می‌آیند و z ها از بقیه پرانتزها هستند. بنابراین داریم:

$$(x + y + z)^n = \sum \binom{n}{n_1, n_2, n_3} x^{n_1} y^{n_2} z^{n_3}$$

که در آن جمع‌بندی روی تمام جوابهای درست و غیر منفی معادله $n_1 + n_2 + n_3 = n$ که به صورت (n_1, n_2, n_3) هستند، انجام می‌گیرد.

مثلاً بسط $(x + y + z)^4$ بعد از خلاصه شدن تعداد $15 = \binom{4+4-1}{4}$ جمله دارد که در آن ضرب جمله xyz^3 برابر است با $\frac{4!}{2!1!1!} = 12$ به خاطر بسطهای بالا، معمولاً $\binom{n}{n_1, n_2, n_3}$ را یک ضرب سه جمله‌ای و $\binom{n}{k}$ را که در حقیقت معادل $\binom{n}{n-k}$ می‌باشد، یک ضرب دو جمله‌ای می‌نامند.

مثال ۲۵ احتمال اینکه روزهای تولد اعضای یک خانواده ده نفری در سال تمام روزهای هفته را شامل شوند چقدر است؟ این مسئله ظاهراً هم مدل با مسئله زیر است:
ده مهر غیر متمایز را به تصادف در هفت جعبه متمایز می‌گذاریم. احتمال اینکه در هر جعبه حداقل یک مهره داشته باشیم چقدر است؟

طبق مدل د جعبه و مهره، ده مهره غیر متمایز در هفت جعبه متمایز به $\binom{16}{7}$ طریق می‌توانند قرار گیرند. از طرفی برای این که در هر جعبه لااقل یک مهره داشته باشیم، نخست یک مهره در هر یک از هفت جعبه می‌گذاریم و سپس سه مهره باقیمانده را به

$$\binom{16-7}{7} = \binom{9}{7}$$

طریق در هفت جعبه قرار می‌دهیم. چون مدل احتمال یکنواخت است، احتمال مطلوب می‌شود:

$$\frac{\binom{9}{7}}{\binom{16}{7}} = \frac{3}{286}$$

$$\frac{1}{(5 \times 4 \times 3)} = \frac{1}{60}$$

احتمال برای نمونه مرتب بی جایگذاری می‌شود

$$\frac{3!}{(5 \times 4 \times 3)} = \frac{1}{10}$$

احتمال برای نمونه نامرتب بی جایگذاری می‌شود

$$\frac{1}{\binom{5}{3}} = \frac{1}{10}$$

احتمال برای نمونه با هم می‌شود

لازم است اضافه کنیم که تعداد نمونه‌های سه تایی نامرتب با جایگذاری، درست برابر است با تعداد ترکیبهای سه تایی با تکرار مجاز، یعنی $35 = \binom{5+3-1}{3}$ (به مثال ۲۰ رجوع شود). ولی این نمونه‌ها همشانس نیستند، زیرا اگر ترتیب را رعایت کنیم نمونه ۱، ۲، ۳ به شش طریق و نمونه ۲، ۳، ۱ به سه طریق مرتب می‌شوند. بنابراین نمی‌توانیم بگوئیم احتمال برای نمونه نامرتب با جایگذاری می‌شود $\frac{1}{35}$. ملاحظه می‌شود که در دو حالت آخر احتمالها با هم مساوی می‌باشند، زیرا تعداد راه‌های بیرون آوردن ۳ کارت از ۵ کارت یک به یک بی جایگذاری و بدون توجه به ترتیب آنها درست برابر است با تعداد راه‌های بیرون آوردن هر ۳ کارت با هم. برای اطمینان، این شمارش را به هر دو روش مستقیماً انجام دهید.

۶.۴.۲ تمرین بخش چهار

۱. در آزمایشی نخست یک سکه و سپس یک تاس را می‌اندازیم. تعداد پیشامدهای ساده را به کمک یک نمودار درختی و همچنین یک جدول پیدا کنید.
۲. یک تاس نازیب را پنج بار مستقلاً می‌اندازیم. تعداد پیشامدهای ساده را به دست آورید. احتمال اینکه خال مضرب سه مشاهده نشود چقدر است؟
۳. دو دختر و سه پسر به تصادف در یک صف می‌ایستند. احتمال اینکه دختری در دو سر صف بایستد چقدر است؟ احتمال اینکه دخترها پهلوی هم بایستند چقدر است؟ احتمال اینکه پسر و دختر یک در میان بایستند چقدر است؟
۴. در شهری برای اتوموبیلها شماره‌های پنج رقمی بی صفر به کار می‌برند. چند شماره که رقم اول و آخر آنها مثل هم باشند می‌توان تهیه دید؟
۵. زده پست در اداره‌ای می‌خواهند سه پست را به علت کمی مراجعه تعطیل کنند. به چند طریق می‌توان این کار را انجام داد؟ احتمال این که پست مخصوصی تعطیل شود چقدر است؟
۶. در اداره‌ای ده پست خالی موجود است و تنها سه نفر صلاحیت دار برای اشغال این پستها پیدا می‌شوند. به چند طریق می‌توان این سه نفر را در این پستها مشغول داشت؟
۷. سه نفر شیرازی، دو نفر تهرانی و چهار نفر اصفهانی به چند طریق می‌توانند در یک صف بایستند، به فرض آن که همشهریها از هم متمایز نباشند. احتمال این که دو تهرانی در دو سر صف فرار

گیرند چقدر است؟

۸. به کمک فرمول استرلینگ ۱۵! را به تقریب محاسبه کنید و خطای نسبی این تقریب را به دست آورید.

۹. به چند طریق پنج نفر می‌توانند اطراف یک میز گرد بنشینند؟ احتمال این که جوانترین و پیرترین آنها پهلوی هم قرار گیرند چقدر است؟ (فرض کنید که این افراد هم سن نباشند).

۱۰. در پرسشنامه‌ای ده پرسش بلی - نه داده شده است. احتمال این که شخصی به تصادف آنها را با چهار بلی و شش نه پر کند چقدر است؟

۱۱. ثابت کنید که با فرض $N \geq R$ داریم:

$$\binom{N}{R} = \binom{N-1}{R} + \binom{N-1}{R-1}, \quad \binom{N}{R} = \binom{N}{N-R}$$

به کمک جعبه و مهره صحت این دو فرمول را توجیه کنید.

۱۲. ضریب $z^2 y^2 x^2$ در بسط $(2x - 3y + 4z)^4$ پیدا کنید.

۱۳. از بسط $(1 + 1)^N$ استفاده کرده ثابت کنید:

$$\binom{N}{0} + \binom{N}{1} + \dots + \binom{N}{N} = 2^N$$

حال تعداد زیر مجموعه‌های یک مجموعه N عضوی را پیدا کنید.

۱۴. ثابت کنید به $\binom{R}{N}$ طریق می‌توان R مهره غیر متمایز را در N جعبه گذاشت بطوریکه هیچیک خالی نماند ($R \geq N$). به چند طریق می‌توان این کار را انجام داد به طوری که در هر جعبه لااقل دو مهره قرار گیرد، با فرض $R \geq 2N$ ؟

۱۵. تعداد جوابهای درست مثبت معادله $x_1 + x_2 + \dots + x_N = R$ را پیدا کنید.

مثال: $x_1 + x_2 + x_3 = 5$

۱۶. در ظرفی N کارت به شماره‌های ۱ تا N داریم. از این کارتها یک نمونه R تایی بیرون می‌آوریم. با استفاده از مثال ۲۷، تعداد انواع نمونه‌ها را پیدا کنید.

۱۷. سه مهره a, b, c را به چند طریق می‌توانید در پنج جعبه متمایز قرار دهید در صورتی که حداکثر یک مهره در هر جعبه مجاز باشد؟ اگر بیش از یک مهره مجاز باشد، به چند طریق؟

۱۸. ده مهره غیر متمایز را به چند طریق می‌توانید در چهار جعبه متمایز قرار دهید در صورتی که بخواهیم در هر جعبه لااقل یک مهره قرار گیرد؟

$$\binom{9}{43}$$

۱۹. در ظرفی سه مهره سفید، چهار مهره قرمز، و پنج مهره سبز داریم. از این ظرف شش مهره با هم بیرون می‌آوریم. مطلوبست احتمال این که در این نمونه شش‌تایی، مهره‌های هم‌رنگ مساوی باشند. احتمال این که مهره‌ها از سه رنگ باشند و اکثریت با مهره‌های سبز باشد چقدر است؟

۲۰. سه نفر زن و نه نفر مرد را می‌خواهیم به گروه‌های سه نفری، چهار نفری، و پنج نفری تقسیم کنیم به طوری که در هر گروه یک زن باشد. به چند طریق می‌توان این کار را انجام داد؟

۲۱. یک خانواده ۳ نفری پنجشنبه یا جمعه متولد شده‌اند. احتمال اینکه هم پنجشنبه و هم جمعه جشن تولد داشته باشند چقدر است؟ یک بار سه نفر را غیر متمایز و یک بار متمایز بگیرید. از مدل مهره و جعبه استفاده کنید و تمام حالات را در نظر بگیرید. جواب درست کدام است؟

۲۲. ریاضیدانان ایتالیایی گاهی روی پیشامد مجموع خاله‌های سه تاس شرط‌بندی می‌کردند. آنها می‌پنداشتند که احتمال مجموع ۹ و احتمال مجموع ۱۰ برابرند، ولی عملاً مجموع ۹ را بیشتر می‌دیدند. مسئله را نزد گالیله بردند. ریاضیدانان چه اشتباهی می‌کردند؟ پاسخ درست گالیله چه بوده است؟

۲۳. ده کارگر را میان سه مهندس ایرانی، ایتالیایی و آلمانی به چند طریق می‌توان به تصادف تقسیم کرد به طوری که به ترتیب ۳ و ۴ و ۳ کارگر داشته باشند؟ اگر ۳ نفر از این کارگرها افغانی باشند، احتمال اینکه با مهندس ایرانی کار کنند چقدر است؟

۲۴. ضرائب جمله‌های بسط $(x+y)^n$ را در یک سطر بنویسید. مثلاً ضرائب $(x+y)^2$ می‌شوند ۱ ۲ ۱. اگر n ، را $0, 1, \dots, N$ بگیریم، با یک مثلث عددی فرمولهای تمرین ۱۱ را بیان دارید و از این راه طرز ساختن مثلث خیام - پاسکال را پیدا کنید.

۲۵. ثابت کنید که

$$\binom{n}{0} - \binom{n}{1} + \binom{n}{2} + \dots + (-1)^n \binom{n}{n} = 0$$

۲۶. جعبه‌ای محتوی ۲۴ لامپ است. ۴ عدد از این لامپها معیوب هستند. ۱۰ عدد از این لامپها را به تصادف بر می‌گزینیم. احتمال اینکه تمام آنها سالم باشند چقدر است؟

۵.۲ مدل احتمال با فضای نمونه دلخواه

این بخش برای دانشجویان ریاضی، آمار، فیزیک و مهندسی مفید است.

تا به حال در تمام مثالها، جز مثال ۷، فضای نمونه یک مجموعه با پایان بود. اینک مثال ۷ را که در آن فضای نمونه یک مجموعه شمارش پذیر بی پایان است، یعنی اعضایش را می توان با اعداد طبیعی متناظر کرد، بار دیگر بررسی می نمائیم. افزون بر این، مثال دیگری را مطالعه می کنیم که در آن فضای نمونه شمارش ناپذیر است.

مثال ۲۸ در یک آزمایش، بازی شیر و خط را با یک سکه نازیب مستقلاً آندرد ادامه می دهیم تا برای اولین بار شیر بیاید. چون اولین شیر ممکن است در نخستین بازی یا دومین بازی یا... بیاید، بنابراین فضای نمونه یک مجموعه شمارش پذیر به صورت $S = \{e_1, e_2, \dots\}$ است. مثلاً منظور از پیشامد e_3 این است که در سومین بازی برای اولین بار شیر بیاید، به عبارت دیگر نتیجه سه بازی به صورت TTH باشد.

در این مثال هم مانند فضای نمونه با پایان می توان برای ساختن مدل احتمال به هر پیشامد ساده e_i احتمال p_i را، به طوری که $0 < p_i < 1$ و $p_1 + p_2 + \dots = 1$ صادق باشند، نسبت داد. ضمناً هر پیشامد زیر مجموعه ای است از S و هر زیر مجموعه با پایان یا بی پایان از S را می توان یک پیشامد فرض کرد، که احتمال آن برابر است با مجموع احتمالات پیشامدهای ساده ای که آنرا تشکیل می دهند.

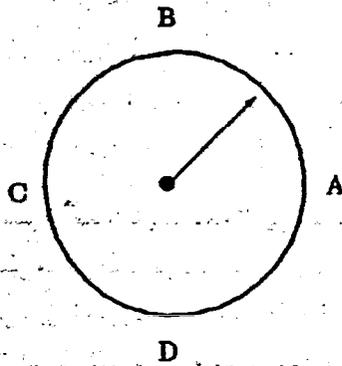
به عنوان مثال احتمال نظیر پیشامد ساده e_4 را تعیین می کنیم. چون سکه نازیب و آزمایشها مستقلاً انجام می گیرند، در چهار بازی شیر و خط 2^4 پیشامد ساده، هر کدام با احتمال $(\frac{1}{2})^4$ داریم که یکی از آنها یعنی، $TTHH$ منجر به e_4 می شود بنابراین $p_4 = (\frac{1}{2})^4$. به طور کلی برای $n = 1, 2, \dots$ احتمال نظیر پیشامد ساده e_n می شود، $p_n = (\frac{1}{2})^n$. مجموع این احتمالات، که تشکیل یک سری هندسی با قدر نسبت $\frac{1}{2}$ می دهند، برابر است با یک.

به عنوان مثال دیگر، پیشامد این که تعداد بازیها برای دیدن نخستین شیر مضرب سه باشد زیر مجموعه ای است از S به صورت $E = \{e_3, e_6, \dots\}$ با احتمال

$$P(E) = (\frac{1}{2})^3 + (\frac{1}{2})^6 + \dots = \frac{(\frac{1}{2})^3}{1 - (\frac{1}{2})^3} = \frac{1}{7}$$

در اینجا هم، مانند احتمال با فضای نمونه با پایان، اگر تمام احتمال را به عنوان یک واحد جرم که میان اعضای مجموعه شمارش پذیر $S = \{e_1, e_2, \dots\}$ توزیع شده است بپنداریم، می توان p_i را به عنوان جرم e_i تعبیر کرد.

مثال ۲۹ عقربه ای در مرکز یک قاب گرد، که روی میز جا دارد، نصب شده است. این عقربه با مختصر اشاره ای آزادانه حرکت می کند و بعد از چند دوران نوک آن به تصادف در نقطه ای از محیط قاب قرار می گیرد. فرض کنید در این آزمایش موضوع مورد علاقه محل توقف نوک عقربه باشد.



بنابراین هر پیشامد ساده نقطه ای است از دایره ای که محیط قاب را تشکیل می دهد، و فضای نمونه مجموعه ای است شمارش ناپذیر متشکل از نقاط محیط این دایره. چون نوک عقربه به تصادف در نقطه ای از محیط قاب قرار می گیرد و هیچ نقطه ای را به نقطه دیگر ترجیح نمی دهد، پس باید به تمام پیشامدهای ساده احتمالات مساوی نسبت داد. اما امکان ندارد که تمام این احتمالات مثبت و کل احتمال برابر یک باشد، زیرا فضای نمونه شمارش ناپذیر است.

با این حال می توان این پرسش را که پاسخی معقول دارد مطرح کرد: احتمال این که نوک عقربه روی کمان AB که یک چهارم محیط است قرار گیرد چقدر است؟ واضح است که، با توجه به مکانیسم دستگاه، تمام چهار ربع محیط دایره از این لحاظ همشانس می باشند، به عبارت دیگر می توان تمام فضای نمونه را به یک فضای نمونه با پایان با چهار عضو همشانس توجیه کرد. بنابراین پاسخ $\frac{1}{4}$ است. به ویژه اگر کل احتمال را به عنوان یک واحد جرم که بطور اتصالی و یکنواخت روی محیط دایره توزیع شده است، بپنداریم، می توان این احتمال را به عنوان جرم کمان AB تعبیر کرد. چنین احتمالی را که متناسب با طول کمان است، می توان با همین توجیه برای تمام کمانها و حتی مجموع چند کمان بکار برد. ضمناً ملاحظه می شود که احتمال هر پیشامد ساده، یا بطور کلی هر پیشامد شمارش پذیر، برابر صفر است. ولی اگر F یک پیشامد شمارش ناپذیر دلخواه باشد، احتمال آن را نمی توان به سادگی پیدا کرد. این مثال نشان می دهد، در حالتی که فضای نمونه شمارش ناپذیر باشد، ممکن است به هر پیشامد ساده، احتمال مثبت تعلق نگیرد، ولی در عوض بعضی از زیر مجموعه های S احتمال مثبت دارند در حقیقت انگیزه ساختن مدل احتمال به نحوی کلی تر، که فضای نمونه شمارش پذیر و شمارش ناپذیر را

در برگیره، و بدون اشکال باشد، با اینگونه مثالها شروع شد.

کلموگرف ریاضیدان و احتمالدان روسی، در سال ۱۹۳۰ میلادی برای اولین بار، با اسلوب ریاضی، احتمال را پی ریزی کرد و راه را برای پیشرفت سریع این دانش گشود.

روشی را که این دانشمند به کار برده است روش اصل سازی می باشد که در هندسه با آن آشنائی کامل داریم. در هر تئوری ریاضی این روش مبتنی بر چند مفهوم ابتدائی و تعریف نشده مانند خط و نقطه در هندسه، است که خواص آنها با چند گزاره بدون اثبات به نام اصل، مانند اصل اقلیدس در مورد خطوط موازی، بیان می شوند. با استفاده از این مقدمات و روش قیاسی منطق، خواص دیگر تحت عنوان قضیه بیان می گردند، و در طی بسط تئوری، مفاهیم تازه، که در کشف قضایای بیشتر نقش دارند تحت عنوان تعریف افزوده می شوند. لازم است اضافه کنیم که در هر تئوری، باید اصلها مستقل، تعداد آنها به اندازه کافی و سازگار باشند (یعنی یکدیگر را نقض نکنند و منجر به قضایای متناقض نشوند). ضمناً باید بر پایه مشاهدات و تجربیات انسان بنا شوند و در خارج مصداق داشته باشند تا بتوانند برای کشف حقایق و تفسیر وقایع مفید واقع گردند. اینک با زبانی ساده و سبک ریاضی، این روش را در مورد احتمال شرح می دهیم.

۱.۵.۲ میدان سیگما

مجموعه دلخواه S را در نظر می گیریم و فرض می کنیم B یک دسته نتهی از زیر مجموعه های S باشد می گوئیم B یک میدان سیگما است هرگاه، دو شرط زیر برقرار گردند:

۱. اگر زیر مجموعه E به B تعلق داشته باشد، آنگاه E^c یعنی متمم E هم به B تعلق داشته باشد.
۲. اگر جمله های دنباله E_1, E_2, \dots به B تعلق داشته باشند، آنگاه $\bigcup_{i=1}^{\infty} E_i$ ، یعنی اجتماع آنها هم به B تعلق داشته باشد.

به سختی دیگر، B را یک میدان سیگما می نامند، هرگاه روی اعضای آن اعمال متمم گیری و اجتماع شمارش پذیر انجام گیرد، اعضائی متعلق به آن به دست آیند. نام این میدان به خاطر شرط دوم می باشد. با این شرایط فوراً نتیجه می شود که مجموعه تهی و خود S هم به B تعلق دارند. ضمناً با استفاده از قانونهای دومرگان، که در نظریه مجموعه ها دیده اید، می توان نشان داد که هرگاه جمله های دنباله E_1, E_2, \dots به B تعلق داشته باشند، $\bigcap_{i=1}^{\infty} E_i$ یعنی اشتراک آنها هم به B تعلق دارد.

به طور کلی در هر مجموعه S تمام زیر مجموعه ها پر عضوترین میدان سیگما و مجموعه $\{S, \emptyset\}$ کم عضوترین میدان سیگما را می سازد. اینک به مثالهای دیگر توجه می کنیم.

مثال ۳۰ در مجموعه $S = \{a, b, c\}$ تمام هشت زیر مجموعه تشکیل یک میدان سیگما می دهند ولی یک میدان سیگمای کوچکتر را می توان از چهار زیر مجموعه $S, \{b, c\}, \{a\}$ ، و \emptyset تشکیل داد. زیر مجموعه های $S, \{a\}$ ، و \emptyset تشکیل میدان سیگما نمی دهند زیرا متمم a در میان آنها نیست.

مثال ۳۱ فرض می کنیم S مجموعه اعداد حقیقی باشد. تمام زیر مجموعه های S تشکیل یک میدان سیگما می دهند. ولی می توانیم یک میدان سیگمای کم عضوتر، که در آنالیز ریاضی مورد استفاده است و آنرا میدان برل به نام ریاضیدان فرانسوی برل (۱۸۸۸-۱۹۵۶) می نامند، معرفی کرد. میدان برل را می توان با روش زیر ساخت.

تمام زیر مجموعه های S را که به صورت فاصله های نیم باز $(-\infty, a]$ ، یا اعداد حقیقی دلخواه a می باشند، در نظر می گیریم. روی این مجموعه ها اعمال متمم گیری و اجتماع و اشتراک شمارش پذیر را انجام می دهیم. بعداً روی مجموعه های حاصل و مجموعه های داده شده، این اعمال را مرتباً بدون وقفه تکرار می کنیم. سرانجام یک دسته از زیر مجموعه های S به دست می آید که یک میدان برل نامیده می شود. با مثالی پیچیده می توان زیر مجموعه ای از S پیدا کرد که در این میدان نباشد. این میدان کوچکترین میدان سیگما است که انواع فواصل و زیر مجموعه های باپایان و شمارش پذیر S را در بر می گیرد.

۲.۵.۲ تابع و تابع مجموعه ای

به طور کلی هرگاه بتوانیم بر پایه یک رابطه به هر عضو از مجموعه ای فقط یک عضو از مجموعه ای دیگر را نسبت دهیم، این رابطه را یک تابع می گویند. مجموعه نخست را دامنه و مجموعه دوم را برد این تابع می نامند. به سختی دیگر، هر تابع رابطه ای یک به یک یا چند به یک (ولی نه یک به چند) از مجموعه ای به مجموعه ای دیگر است.

مثلاً $f(x) = x^2 + 1$ تابعی دو به یک از مجموعه اعداد حقیقی به مجموعه اعداد حقیقی مثبت است. ضابطه این تابع، هم 3 و هم -3 را به 10 می فرستد. ولی $h(x) = x^2 - 1$ تابعی یک به یک است. هر یک از این دو تابع را تابع نقطه ای می نامند.

به عنوان مثالی دیگر، رابطه ازدواج تابعی چند به یک است. دامنه این تابع زنان متاهل و برد آن مردان متاهل است. ضابطه این تابع را نمی توان مانند دو تابع نقطه ای یاد شده، با یک فرمول ریاضی بیان کرد.

اگر هر عضو از دامنه تابعی خود یک مجموعه باشد، چنین تابعی را تابع مجموعه ای می خوانند. مثلاً

فرض کنید g تعداد اعداد درستی باشد که در یک فاصله با پایان موجود است، مانند

$$g([2, 7]) = 6$$

دامنه g تمام فاصله‌های باپایان و برد آن اعداد درست غیر منفی است. به عنوان مثالی دیگر، مساحت یک مثلث یک تابع مجموعه‌ای است. دامنه این تابع تمام مثلثهای واقع در صفحه و برد آن اعداد حقیقی است. این دو تابع مجموعه‌ای، هر دو چند به یک هستند.

۳.۵.۲ فضای احتمال

فرض کنید S یک مجموعه دلخواه و B یک میدان سیگما از زیر مجموعه‌های S باشد. یک تابع مجموعه‌ای P به نام تابع احتمال را، که دامنه آن B و برد آن مجموعه اعداد حقیقی است، در نظر می‌گیریم. اگر اصلهای زیر در مورد P صادق باشند، آنگاه سه‌تایی (S, B, P) را یک فضای احتمال می‌نامند:

اصل یک - $P(S) = 1$ ، اصل دو - $P(E) \geq 0$ ، برای هر E که به B تعلق داشته باشد.

اصل سه - $P(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$ در اصل سه جمله‌های دنباله E_1, E_2, \dots به B تعلق دارند و دو به دو جدا می‌باشند یعنی

$$E_i \cap E_j = \emptyset \quad i \neq j \quad i, j = 1, 2, \dots$$

این اصل را، اصل جمع‌پذیری احتمال می‌گویند.

از این پس هر زیر مجموعه E از S را، که در میدان سیگمای B باشد یک پیشامد و $P(E)$ را احتمال آن می‌گوئیم. بنابراین هر زیر مجموعه از S را که به B تعلق نداشته باشد نمی‌توان یک پیشامد خواند و طبیعتاً احتمال ناپذیر است. خود S را فضای نمونه یا پیشامد حتمی و \emptyset را پیشامد محال می‌نامیم. اصول بالا، در حقیقت، فشرده‌ای از مشاهدات و تجربیات ما درباره آزمایش تصادفی، پیشامد و مفهوم و خواص احتمال هستند، که به نحوی کلی و انتزاعی ارائه شده‌اند. با داشتن این اصول می‌توانیم تئوری احتمال را، بدون این که به مثالهای گوناگون و حالات مختلف متوسل شویم، در چهارچوب قوانین ریاضی گسترش دهیم.

در تمام مطالب و قضایائی که دنبال می‌کنیم یک فضای احتمال ثابت (S, B, P) را در نظر داریم و در صورتی که لزوم نداشته باشد از تکرار این موضوع خودداری خواهیم کرد.

قضیه ۱ پیشامد \emptyset که آن را پیشامد محال می‌نامیم، دارای احتمال صفر است.

برهان دنباله پیشامدهای E_1, E_2, \dots را با فرض $E_1 = S$ و $E_i = \emptyset$ برای $i > 1$ در نظر می‌گیریم. چون پیشامدها در این دنباله دو به دو جدا هستند و $S = \cup_{i=1}^{\infty} E_i$ پس بنا بر اصل سه داریم

$$P(S) = P(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i) = P(S) + P(\emptyset)$$

از طرفی طبق اصل یک $P(S) = 1$ ، بنابراین بعد از حذف $P(S)$ از دو طرف داریم $P(\emptyset) = 0$.

قضیه ۲ برای پیشامدهای دو به دو جدای E_1, E_2, \dots, E_n داریم:

$$P(\cup_{i=1}^n E_i) = \sum_{i=1}^n P(E_i) \quad (5)$$

برهان دنباله پیشامدهای E_1, E_2, \dots را با فرض $E_i = \emptyset$ برای $i > n$ در نظر می‌گیریم. چون پیشامدها در این دنباله دو به دو جدا هستند و $\cup_{i=1}^n E_i = \cup_{i=1}^{\infty} E_i$ ، پس بنا بر اصل سه و قضیه یک داریم:

$$\begin{aligned} P(\cup_{i=1}^n E_i) &= P(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i) = \sum_{i=1}^n P(E_i) + \sum_{i=n+1}^{\infty} P(E_i) \\ &= \sum_{i=1}^n P(E_i) + P(\cup_{i=n+1}^{\infty} E_i) = \sum_{i=1}^n P(E_i) + P(\emptyset) = \sum_{i=1}^n P(E_i) \end{aligned}$$

بنابراین جمع‌پذیری بی‌پایان و باپایان، هر دو، در مورد تابع احتمال درست است.

قضیه ۳ اگر پیشامد E' متمم پیشامد E باشد، آنگاه

$$P(E') = 1 - P(E) \quad (6)$$

برهان چون E و E' از هم جدا هستند و $S = E \cup E'$ پس بنا بر اصل یک و قضیه ۲ داریم:

$$1 = P(S) = P(E \cup E') = P(E) + P(E')$$

$$P(E') = 1 - P(E)$$

قضیه ۴ برای هر دو پیشامد E و F داریم:

$$P(F-E) = P(F) - P(E \cap F) \quad (۷)$$

برهان به سادگی می توان نشان داد که برای هر دو مجموعه E و F داریم:

$$F = (F-E) \cup (E \cap F)$$

که در آن $E \cap F$ و $F-E$ از هم جدا هستند. پس بنابر قضیه (۲) داریم:

$$P(F) = P(F-E) + P(E \cap F)$$

$$P(F-E) = P(F) - P(E \cap F)$$

نتیجه ۱ اگر $E \subset F$ ، آنگاه $E \cap F = E$ و $P(F-E) = P(F) - P(E)$ ، یعنی در این مورد تابع

احتمال دارای ویژگی تفریق پذیری است.

نتیجه ۲ اگر $E \subset F$ ، آنگاه با استفاده از نتیجه (۱) و اصل دو داریم $P(E) \leq P(F)$. به عبارت دیگر

اگر بتوان پیشامدها را به وسیله رابطه C مرتب کرد، آنگاه تابع P حافظ ترتیب است. به ویژه برای هر

پیشامد E ، از $E \subset S$ و اصل یک و دو داریم:

$$0 \leq P(E) \leq 1$$

قضیه ۵ برای هر دو پیشامد E و F داریم:

$$P(E \cup F) = P(E) + P(F) - P(E \cap F) \quad (۸)$$

برهان به سادگی می توان نشان داد که برای دو مجموعه E و F داریم:

$$E \cup F = E \cup (F-E)$$

که در آن E و $F-E$ از هم جدا هستند. پس بنابر قضایای (۲) و (۲) داریم

$$P(E \cup F) = P(E) + P(F-E) = P(E) + P(F) - P(E \cap F)$$

مثال ۲۲ فرض می کنیم $S = \{a, b, c\}$ و میدان سیگمای B از زیر مجموعه های S ، $\{a\}$ ، $\{b, c\}$ ، \emptyset تشکیل شده باشد. با فرض $0 < p < 1$ و

$$P(\{a\}) = p, \quad P(\{b, c\}) = 1-p, \quad P(S) = 1$$

سه تایی (S, B, P) یک فضای احتمال است. در اینجا به $\{a, c\}$ نمی توان احتمال نسبت داد زیرا به میدان سیگمای B تعلق ندارد. پس این زیر مجموعه از S احتمال ناپذیر است.

مثال ۲۳ فرض می کنیم $S = \{e_1, e_2, \dots, e_n\}$ و میدان سیگمای B از تمام زیر مجموعه های S تشکیل شده باشد. اعداد مثبت p_1, p_2, \dots, p_n را که مجموع آنها برابر یک می باشد در نظر می گیریم. فرض می کنیم برای هر زیر مجموعه E داشته باشیم.

$$P(E) = \sum_{e_i \in E} p_i$$

مثلاً برای $E = \{e_2, e_5, e_7\}$ داریم $P(E) = p_2 + p_5 + p_7$.

واضح است که (S, B, P) یک فضای احتمال است. از تعریف P معلوم می شود که

$$P(\{e_i\}) = P(e_i) = p_i, \quad i = 1, 2, \dots, n$$

در حقیقت این همان مدل احتمال با فضای نمونه با پایان است که قبلاً آن را مطالعه کرده ایم.

بدیهی است که در این مثال می توان به جای مجموعه با پایان که یک مجموعه شمارش پذیر

$\{e_1, e_2, \dots\}$ را در نظر گرفت.

مثال ۳۴ فرض کنید S فاصله بسته $[0, 2]$ و میدان سیگما از تمام زیر مجموعه های آن تشکیل شده

باشد. در اینجا میدان سیگما به اندازه ای پر عضو است که نمی توان هیچ تابع مجموعه ای را مقید به

اصول احتمال کرد و یک فضای احتمال ساخت. ولی می توان یک میدان سیگمای کم عضو تر مانند

مثال (۳۱) (میدان بول)، به وسیله تکرار بی وقفه اعمال متمم گیری و اجتماع و اشتراک روی زیر

فاصله های $[a, b]$ با فرض $0 \leq a \leq b \leq 2$ به دست آورد. آنگاه می توان یک فضای احتمال ساخت. به

ویژه اگر تمام احتمال، یا به تعبیر بهتر یک واحد جرم، به طور پکنواخت و اتصالی در فاصله $[0, 2]$

توزیع شده باشد، آنگاه مقدار تابع P برای هر زیر فاصله $[a, b]$ متناسب است با طول آن یعنی $\frac{b-a}{2}$.

این مثال در حقیقت مدل احتمال آزمایشی است که در آن می خواهیم یک عدد حقیقی به تصادف

در فاصله $[0, 2]$ انتخاب کنیم. چون بعضی از اعداد حقیقی، مثلاً $\sqrt{3}$ را، باید با تقریب اعشاری بیان

۱- مضرب سه

۲- حداکثر سه

۳- حداقل سه

یا شد چقدر است؟

راهنامه‌ی: پیشامد آمدن خال شش را با E و نیامدن آن را با E' نشان دهید.

۳. از شخصی درخواست می‌کنیم تا به تصادف عددی در فاصله $[2, 7]$ در نظر گیرد. فضای احتمال را مشخص کنید. احتمال این که عدد یاد شده کوچکتر از سه باشد چقدر است؟ نشان دهید احتمال این که عدد درست در نظر گرفته شود، برابر صفر است.

۴. نقطه M را به تصادف در داخل مربع $ABCD$ که طول هر ضلع آن ۴ واحد است، انتخاب می‌کنیم. احتمال پیشامد اینکه فاصله M از قطر AC کمتر از $\sqrt{2}$ باشد، چیست؟

۵. برای مجموعه $S = \{a, b, c, d\}$ ، کوچکترین میدان سیگما را که شامل $\{a\}$ و $\{c, d\}$ به ترتیب با احتمالهای $\frac{1}{4}$ و $\frac{1}{8}$ است پیدا کنید. احتمال برای سایر اعضای این میدان سیگما را به دست آورید. یک مجموعه احتمال ناپذیر پیدا کنید.

۶. با استفاده از تمرین (۳) پیشامدهای A ، B و C را چنان بسازید که داشته باشیم:

$$P(A) = 0, \quad A \neq \emptyset$$

$$P(B) = P(C), \quad B \neq C$$

از این تمرین معلوم می‌شود که هر پیشامدی با احتمال صفر لزوماً تهی نیست. افزون بر این، تابع P همواره یک به یک نمی‌باشد.

۷. ثابت کنید که

$$P(A \cup B) \leq P(A) + P(B) \quad (\text{نامساوی بول})$$

$$P(A \cap B) \geq P(A) + P(B) - 1 \quad (\text{نامساوی بونفرونی})$$

۸. ثابت کنید که

$$P(A' \cup B') = 1 - P(A) - P(B) + P(A \cup B)$$

کرد، عملاً انجام دقیق چنین آزمایشی مقدور نیست. با این حال می‌توان احتمال بسیاری از پیشامدها را محاسبه کرد. مثلاً احتمال این که عدد انتخابی در فاصله $[0/50, 0/75]$ قرار گیرد می‌شود $\frac{0/75 - 0/50}{1} = 0/125$ ، زیرا احتمال متناسب با طول فاصله می‌باشد. به عنوان مثالی دیگر، احتمال این که عدد انتخابی برابر $1/5$ شود، صفر است. برای اثبات این موضوع عدد بسیار کوچک a را در نظر می‌گیریم. احتمال این که عدد انتخابی در فاصله $[1/5 - a, 1/5 + a]$ قرار گیرد متناسب است با طول این فاصله یعنی، $2a$. چون a دلخواه است و می‌توانیم آن را هر اندازه که بخواهیم کوچک اختیار کنیم، بنابراین احتمال این که عدد انتخابی در نزدیکیهای $1/5$ باشد، می‌تواند خیلی کوچک شود و احتمال این که درست عدد $1/5$ باشد صفر است.

مثال ۳۵ نقطه‌ای به تصادف در دایره‌ای به شعاع R و مرکز O انتخاب می‌کنیم. احتمال پیشامد این که نقطه یاد شده به مرکز دایره نزدیک باشد تا به محیط دایره چقدر است؟ فضای نمونه که مجموعه‌ای است شمارش ناپذیر، از نقاط داخل دایره (O, R) و پیشامد منظور، از نقاط داخل دایره $(O, \frac{R}{4})$ تشکیل می‌شود. اگر تمام احتمال را به یک واحد جرم که به طور یکتواخت و اتصالی در سطح داخل دایره اول توزیع شده است تعبیر کنیم، احتمال این پیشامد برابر است با نسبت مساحت دایره دوم به مساحت دایره اول، یعنی

$$\frac{\pi R^2}{4} : (\pi R^2) = \frac{1}{4}$$

مسئله‌ی مانند مثالهای (۳۴) و (۳۵) را مسائل هندسی احتمال می‌نامند حل آنها، با تعبیر کل احتمال به صورت یک واحد جرم و با فرض اینکه جرم به طور یکتواخت توزیع شده باشد، عملاً منجر به محاسبه طول یا مساحت می‌شود.

۴.۵.۲ تمرین بخش پنج

۱. نشان دهید اگر $S = \{e_1, e_2, \dots\}$ و $P(e_n) = (\frac{1}{4})^n$ برای $n = 1, 2, \dots$ ، آنگاه یک فضای احتمال داریم. هرگاه $A = \{e_1, e_2, \dots, e_{10}\}$ و $B = \{e_4, e_5, \dots, e_{10}\}$ ، احتمالهای زیر را در این فضای احتمال حساب کنید:

$$P(A - B) \quad P(A \cap B) \quad P(A \cup B)$$

۲. در آزمایشی، تاس نازیبی را آنقدر می‌ریزیم تا برای اولین بار خال شش بیاید. اگر تعداد ریزشهای لازم موضوع مورد نظر باشد، فضای احتمال را پیدا کنید. احتمال آن که تعداد ریزشهای لازم،

۹. ثابت کنید نامساوی بول و بونفرونی معادل اند، یعنی یکی را از روی دیگری می توان بدست آورد؟

۱۰. به کمک استقرای ریاضی و تمرین ۷ ثابت کنید که

$$P(U_{i=1}^n E_i) \leq \sum_{i=1}^n P(E_i) \quad (\text{نامساوی بول با } n \text{ پیشامد})$$

$$P(\cap_{i=1}^n E_i) \geq \sum_{i=1}^n P(E_i) - n + 1 \quad (\text{نامساوی بونفرونی با } n \text{ پیشامد})$$

۱۱. پیشامدهای A_1, A_2, \dots, A_n را در نظر گرفته فرض می کنیم که

$$B_1 = A_1, B_2 = A_1 \cap A_2, \dots, B_n = A_1 \cap A_2 \cap \dots \cap A_{n-1} \cap A_n$$

ثابت کنید که

$$P(U_{i=1}^n A_i) = \sum_{i=1}^n P(B_i)$$

حال نامساوی بول را برای n پیشامد و بینهایت پیشامد E_1, E_2, \dots نتیجه بگیرید.

۱۲. آیا قضیه زیر درست است؟

$$P(A-B) = P(B-A) \text{ اگر و تنها اگر } A=B$$

۱۳. فرض کنید

$$P(A) = P(B) = 0/95$$

با استفاده از نامساوی بونفرونی، حداقل مقدار $P(A \cap B)$ را بیابید.

۱۴. یک فرمول برای محاسبه $P(A \Delta B)$ بیابید.

۱۵. یک فرمول برای محاسبه $P(A \Delta B \Delta C)$ بیابید.

۶.۲ مدل احتمال شرطی

اغلب در یک آزمایش، علم به این که پیشامدی رخ داده است، ممکن است در احتمال رخ دادن پیشامدی دیگر تاثیر کند. مثلاً اگر بدانیم در هفته اول دیماه شهری بارانی بوده است، وقوع این پیشامد می تواند احتمال بارانی بودن هفته دوم دیماه را تقلیل دهد.

به طور کلی در یک آزمایش، مدل احتمال به علت پیشامدی که رخ داده است، دستخوش دگرگونی شده مدل جدیدی به نام مدل احتمال شرطی به دست می آید. احتمال شرطی دارای اهمیتی خاص می باشد و به کمک آن می توان تئوری احتمال را گسترش داد و بعضی مسائل پیچیده را حل کرد برای اینکه منظور از احتمال شرطی روشن شود، پیش از تعریف ریاضی آن به ذکر یک مثال ساده می پردازیم.

مثال ۳۶ در آزمایشی یک تاس نارنجی را مستقلاً دوبار ریخته اند. اگر بدانیم که مجموع خالها کمتر از شش است، احتمال این که خالها مساوی باشند چقدر است؟

در این آزمایش فضای نمونه متشکل است از ۳۶ پیشامد ساده هر یک با احتمال $\frac{1}{36}$. اگر فرض کنیم A پیشامد مساوی بودن خالها و B پیشامد کمتر از شش بودن مجموع خالها باشد داریم:

$$A = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6)\}$$

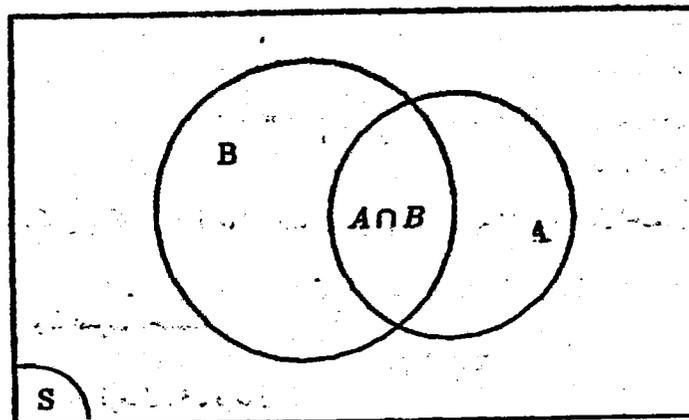
$$B = \{(1, 1), (1, 2), (1, 3), (1, 4), (2, 1), (2, 2), (2, 3), (3, 1), (3, 2), (4, 1)\}$$

$$P(A) = \frac{1}{6} \quad P(B) = \frac{5}{18}$$

چون می دانیم که B با ۱۰ پیشامد ساده رخ داده است، بر پایه علم به این موضوع یقین داریم که ۲۶ پیشامد ساده دیگر رخ نخواهند داد. پس فضای نمونه جدید منحصر به خود B می شود و احتمال پیشامد A را باید در چارچوب آن پیدا کرد.

نظر به این که آزمایش دارای مدل احتمال یکتواخت است، طبیعی است که مدل احتمال شرطی را هم یکتواخت فرض کنیم. بنابراین مدل احتمال شرطی از فضای نمونه B با ۱۰ پیشامد ساده هر یک با احتمال $\frac{1}{10}$ تشکیل می شود. حال احتمال شرطی A که آن را با $P(A|B)$ نشان می دهند و احتمال A به شرط B می خوانند، برابر است با $\frac{1}{5}$ ، زیرا این پیشامد موقعی رخ می دهد که پیشامدهای متعلق به $A \cap B = \{(1, 1), (2, 2)\}$ رخ دهد و احتمال هر یک از این پیشامدهای ساده در فضای نمونه شرطی $\frac{1}{10}$ است.

اگر مدل احتمال غیر شرطی یکتواخت نباشد، باید احتمال هر پیشامد ساده متعلق به B را متناسباً به نحوی افزایش داد که احتمال شرطی برای B ، یعنی فضای نمونه شرطی، برابر یک شود. به طور کلی، با توجه به نگاره (۳)، معقول به نظر می رسد که $P(A|B)$ را متناسب با $P(A \cap B)$ بگیریم، یعنی $P(A|B) = KP(A \cap B)$. ضریب K را باید به نحوی اختیار کرد که احتمال شرطی برای B برابر



نگاره ۳ فضای نمونه S و فضای نمونه شرطی B

یک شود. بنابراین

$$P(B|B) = P(B) = K P(B \cap B) = K P(B)$$

$$K = \frac{1}{P(B)}$$

پس با فرض $P(B) \neq 0$ داریم $K = \frac{1}{P(B)}$ اینک که منظور از احتمال شرطی روشن شد، دقیقاً آن را تعریف می‌کنیم.

تعریف - فرض کنید A و B دو پیشامد دلخواه در فضای احتمال (S, B, P) باشند به طوری که $P(B) \neq 0$. احتمال A به شرط B ، یعنی احتمال پیش آمد A با علم به اینکه پیشامد B رخ داده است برابر است با

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (9)$$

اگر $P(B) = 0$ ، آنگاه $P(A|B)$ قابل تعریف نیست.

حال فرض کنید B با $P(B) > 0$ یک پیشامد ثابت باشد و برای هر پیشامد دلخواه A احتمال شرطی $P(A|B)$ را با $P_B(A)$ نشان دهید. به سادگی می‌توان ثابت کرد که (S, B, P_B) یک فضای احتمال است. یعنی تابع مجموعه‌ای P_B از اصول احتمال پیروی می‌کند. این فضای احتمال را فضای احتمال شرطی می‌خوانیم.

مثال ۳۷ مثال (۳۶) را به کمک فرمول (۹) حل کنید.

چون $P(B) = \frac{5}{18}$ و $P(A \cap B) = \frac{1}{18}$ ، بنابراین

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{18}}{\frac{5}{18}} = \frac{1}{5}$$

ملاحظه می‌شود که با به کار بردن فرمول احتمال شرطی نیازی به تشریح مدل احتمال شرطی نیست.

۱.۶.۲ قانون ضرب احتمال

با فرض $P(B) > 0$ از فرمول (۹)، فرمول

$$P(A \cap B) = P(B)P(A|B) \quad (10)$$

که به قانون ضرب احتمال معروف است به دست می‌آید.

مثال ۳۸ جعبه‌ای محتوی ۱۷ لامپ سالم و ۴ لامپ معیوب است. از این جعبه دو لامپ را یک به یک به تصادف انتخاب کرده امتحان می‌کنیم. احتمال این که هر دو معیوب باشند چقدر است؟ فرض کنید A و B به ترتیب پیشامدهای معیوب بودن لامپ اول و دوم باشند. می‌خواهیم احتمال پیشامد $A \cap B$ را پیدا کنیم.

چون $P(A) = \frac{4}{21}$ و $P(B|A) = \frac{3}{20}$ ، بنابراین طبق قانون ضرب احتمال داریم

$$P(A \cap B) = P(A)P(B|A) = \frac{4}{21} \times \frac{3}{20} = \frac{1}{25}$$

قانون ضرب احتمال را می‌توان به k پیشامد A_1, A_2, \dots, A_{k-1} با فرض

$$P(A_1 \cap A_2 \cap \dots \cap A_{k-1}) > 0$$

تعمیم داد.

به عنوان مثال برای سه پیشامد A, B و C ، با فرض $P(A \cap B) > 0$ فرمول زیر به دست می‌آید

$$\begin{aligned} P(A \cap B \cap C) &= P[(A \cap B) \cap C] = P(A \cap B)P(C|A \cap B) \\ &= P(A)P(B|A)P(C|A \cap B) \end{aligned} \quad (11)$$

فرمول (۱۱) را می‌توان با استفاده از فرمول (۱۰) ثابت کرد.

مثال ۳۹ هفت کارت سفید، شش کارت سبز، و یک کارت قرمز را به خوبی مخلوط می‌کنیم. از میان آنها به تصادف سه کارت را یک به یک بدون جایگزینی انتخاب می‌نمایم. احتمال اینکه این سه کارت به ترتیب سفید و سبز و قرمز باشند چقدر است؟

فرض کنید A ، B و C به ترتیب پیشامدهای این باشند که کارتها سفید، سبز و قرمز باشند. از فرمول بالا داریم

$$P(A \cap B \cap C) = P(A)P(B|A)P(C|A \cap B)$$

$$= \frac{7}{52} \times \frac{6}{51} \times \frac{1}{50} = \frac{1}{520}$$

مستقیماً هم از راه شمارش می‌توانیم مساله را بدین طریق حل کنیم: جمعاً تعداد $12 \times 13 \times 14$ پیشامد ساده هم‌شانس داریم که تعداد $1 \times 6 \times 7$ از آنها منجر به پیشامد مطلوب می‌شوند. بنابراین با تقسیم تعداد حالت‌های مساعد به حالت‌های ممکن، عدد $\frac{1}{520}$ نتیجه می‌شود.

۲۶.۲ تمرین بخش شش

۱. آزمایشی را n مرتبه مستقلاً تکرار می‌کنیم. فرض کنید پیشامدهای A و B به ترتیب n_1 و n_2 مرتبه رخ دهند. صحت فرمول احتمال شرطی را با تعبیر احتمال به طریق فرولاتی نسبی نشان دهید.
۲. خانواده‌ای سه فرزند دارد. اگر فرزند اول و آخر همجنس باشند، احتمال همجنس بودن تمام فرزندان چقدر است؟ مساله را با تشریح مدل احتمال شرطی و همچنین به کمک فرمول احتمال شرطی حل کنید.
۳. یک بسته کارت ۵۲ تایی، به شماره‌های ۱ تا ۵۲، را به خوبی مخلوط می‌کنیم. سپس سه کارت یک به یک بدون جایگزینی بیرون می‌آوریم. احتمال این که شماره‌های آنها از ۱۴ کمتر باشد چیست؟ مساله را به کمک شرطی و از راه شمارش حل کنید.
۴. دو تاس را با هم می‌ریزیم. در صورتی که بدلتیم مجموع آنها عدد هفت است، احتمال اینکه مینوم خالها ۳ باشد چقدر است؟
۵. فرمول (۱۱) را با استفاده از فرمول احتمال شرطی (۱۰) و استقرای ریاضی به k پیشامد دهید.

۶. فرض کنید در یک آزمایش، B با $P(B) > 0$ یک پیشامد ثابت باشد. برای هر پیشامد دلخواه A احتمال شرطی $P(A|B)$ را با $P_B(A)$ نشان دهید. ثابت کنید تابع مجموعی P_B از اصول احتمال پیروی می‌کند. اینک فرمول‌هایی برای $P(A_1|B)$ ، $P(A_1 - A_2|B)$ ، $P(A_1 \Delta A_2|B)$ و $P(A_1 \cap A_2|B)$ بیابید.

۷. در ظرفی سه مهره سفید و پنج مهره سیاه داریم. از این ظرف دو مهره بدون جایگزینی یک به یک بیرون می‌آوریم. نشان دهید احتمال این که مهره اول سفید باشد برابر است با احتمال اینکه مهره دوم سفید باشد. (هنگام بیرون آوردن مهره دوم از رنگ مهره اول اطلاع نداریم).

۸. دو تاس را با هم می‌ریزیم. فرض کنید B هفت بودن مجموع خالها، A مضرب ۵ بودن A_1 مضرب ۴ بودن و A_2 مضرب ۳ بودن آنها باشد. احتمال‌های یاد شده در تمرین ۶ را مستقیماً و به کمک فرمول‌هایی که یافته‌اید حل کنید.

۷.۲ فرمول بیز و کاربرد آن

یکی از فرمول‌های معروف احتمال که در آن احتمال شرطی به نحوی بارز به کار می‌رود فرمول بیز می‌باشد. این فرمول را یک کشیش انگلیسی به نام توماس بیز در سال ۱۷۶۳ میلادی پیدا کرد. مقاله اصلی بیز مدتها به طور جدی مورد توجه نبود. ولی این مقاله چندی پیش برای بار دوم در یکی از مجلات علمی آمار منتشر شد. فرمول بیز در حقیقت نقطه شروع رشته‌ای از آمار به نام آمار بیز می‌باشد. آمار بیز، که از لحاظ فلسفی و تعبیر احتمال به طریق شخصی دارای اهمیت است، در چند دهه گذشته توسط آماردانان طی کتب و مقالات متعدد مورد بحث و تحقیق قرار گرفته است. این روزها درباره آمار بیز زیاد می‌توان صحبت کرد. ما تنها بعد از مقدمات لازم، فرمول بیز و کاربرد آن را با چند مثال شرح می‌دهیم.

۱.۷.۲ تفکیک فضای نمونه به چند پیشامد

فرض کنید پیشامدهای A_1, A_2, \dots, A_k که دارای احتمال مثبت هستند، دو به دو جدا از هم و اجتماع آنها برابر فضای نمونه، که باشد. یعنی

$$A_i \cap A_j = \emptyset, \quad i, j = 1, 2, \dots, k$$

$$U = A_1 \cup \dots \cup A_k = S$$

اصطلاحاً می‌گوئیم S به این پیشامدها تفکیک شده است، یا این پیشامدها یک افراز برای S هستند. گاهی این پیشامدها را دو به دو از هم جدا و فراگیر می‌نامند.

مثال ۴۰ در آزمایشی سکه‌ای را سه بار می‌اندازیم. فضای نمونه را می‌توان به پیشامدهای صخر شیر، یک شیر، دو شیر و سه شیر یعنی:

$$\{TTT\}, \{HTT, THT, TTH\}, \{HHT, HTH, THH\}, \{HHH\}$$

تفکیک کرد.

۲.۷.۲ تفکیک یک پیشامد به چند پیشامد

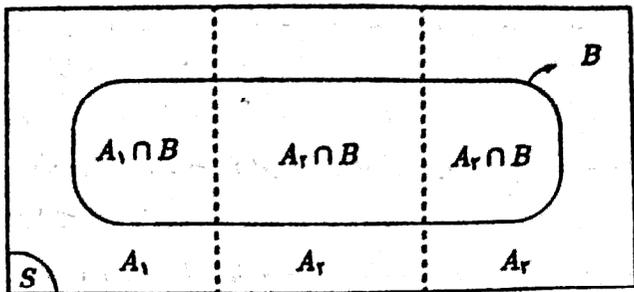
فرض کنید فضای نمونه S به پیشامدهای A_1, A_2, \dots, A_k تفکیک شده باشد. برای هر پیشامد دیگر B که دارای احتمال مثبت باشد داریم:

$$B = B \cap S = B \cap (\cup_{i=1}^k A_i) = \cup_{i=1}^k (B \cap A_i) \quad (12)$$

اصطلاحاً می‌گوئیم پیشامد B به پیشامدهای $A_1 \cap B, A_2 \cap B, \dots, A_k \cap B$ تفکیک شده است. اگر پیشامدهای A_1, A_2, \dots, A_k را به صورت علت و B را به صورت معلول تلقی کنیم، آنگاه می‌توان گفت که B معلول یکی از این علل می‌باشد. فرمول (۱۲) پیشامد B را با توجه به علل گوناگون تفکیک می‌کند. برای توجیه هندسی به نگاره (۴) که در آن S به سه پیشامد A_1, A_2, A_3 تفکیک شده است توجه کنید.

مثال ۴۱ پانزده کارت یکسان به شماره‌های یک تا پانزده داریم. یکی از کارت‌ها را به تصادف انتخاب می‌کنیم. فرض کنید فضای نمونه به کارتهای فرد و زوج تفکیک شده باشد. پیشامد کارت مضرب سه به پیشامدهای $\{۶, ۱۲\}, \{۳, ۹, ۱۵\}$ تفکیک می‌شود.

اینک از فرمول (۱۲) استفاده کرده دو فرمول مهم به نام فرمول تفکیک احتمال و فرمول بیز را به دست می‌آوریم.



نگاره ۴ تفکیک پیشامد B به سه پیشامد

۳.۷.۲ فرمول تفکیک احتمال

از فرمول (۱۲) با استفاده از قانونهای جمع و ضرب احتمال داریم:

$$P(B) = \sum_{i=1}^k P(A_i \cap B)$$

یا

$$P(B) = \sum_{i=1}^k P(B | A_i) P(A_i) \quad (13)$$

این فرمول را فرمول تفکیک احتمال یا فرمول احتمال مرکب می‌نامند و برای حل بعضی مسائل احتمال مفید است.

۴.۷.۲ فرمول بیز

از فرمول احتمال شرطی داریم

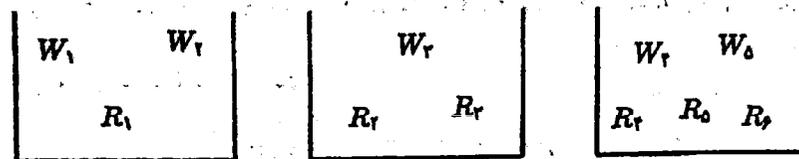
$$P(A_j | B) = \frac{P(A_j \cap B)}{P(B)} \quad j = 1, 2, \dots, k$$

حال با استفاده از فرمول تفکیک احتمال و قانون ضرب احتمال، فرمول معروف بیز به صورت

$$P(A_j | B) = \frac{P(B | A_j) P(A_j)}{\sum_{i=1}^k P(B | A_i) P(A_i)} \quad j = 1, 2, \dots, k \quad (14)$$

به دست می‌آید. بنابراین فرمول بیز، در واقع، وسیله‌ای برای محاسبه احتمال شرطی است، هنگامی که فضای نمونه به چند پیشامد تفکیک شده باشد.

مثال ۴۲. سه ظرف داریم. ظرف یک محتوی دو مهره سفید و یک مهره قرمز، ظرف دو محتوی یک مهره سفید و دو مهره قرمز، و ظرف سه محتوی دو مهره سفید و سه مهره قرمز است.



ظرف یک

ظرف دو

ظرف سه

فرض کنید ظرفها و مهره‌ها گذشته از هویت و رنگ کاملاً یکسان باشند. یکی از این سه ظرف را به تصادف انتخاب می‌کنیم و از آن به تصادف یک مهره بیرون می‌آوریم.
 الف - احتمال این که مهره سفید باشد چقدر است؟ (احتمال معلول)
 ب - اگر مهره سفید باشد، احتمال این که از ظرف سه بیرون آمده باشد چقدر است؟ (احتمال علت)
 فرض کنید B پیشامد سفید بودن مهره و A_1, A_2, A_3 به ترتیب پیشامدهای انتخاب ظروف یک، دو و سه باشند. واضح است که

$$P(A_1) = \frac{1}{3} \quad P(A_2) = \frac{1}{3} \quad P(A_3) = \frac{1}{3}$$

$$P(B|A_1) = \frac{2}{3} \quad P(B|A_2) = \frac{1}{3} \quad P(B|A_3) = \frac{2}{5}$$

الف - از فرمول تفکیک احتمال داریم:

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) \\ = \left(\frac{2}{3}\right)\left(\frac{1}{3}\right) + \left(\frac{1}{3}\right)\left(\frac{1}{3}\right) + \left(\frac{2}{5}\right)\left(\frac{1}{3}\right) = \frac{7}{15}$$

ب - از فرمول بیز داریم:

$$P(A_3|B) = \frac{P(B|A_3)P(A_3)}{P(B)} = \frac{2}{7}$$

در این مثال فضای نمونه عبارتست از

$$S = \{W_1, W_2, R_1, W_3, R_2, R_3, W_4, W_5, R_4, R_5, R_6\}$$

که در آن تمام پیشامدهای ساده همشانس نیستند، زیرا ترکیب مهره‌ها در این سه ظرف با هم فرق دارند فضای پیشامدهای ساده، به پیشامدهای

$$A_1 = \{W_1, W_2, R_1\} \quad A_2 = \{W_3, R_2, R_3\} \quad A_3 = \{W_4, W_5, R_4, R_5, R_6\}$$

تفکیک می‌شود. پیشامد B عبارت است از

$$B = \{W_1, W_2, W_3, W_4, W_5\}$$

۵.۷.۲ احتمال پیشین و احتمال پسین

هرگاه درباره رخ دادن پیشامد A عقیده خود را به صورت احتمال بیان کنیم، چنین احتمالی را که طبیعتاً باید به طریق شخصی تعبیر شود احتمال پیشین می‌نامند.
 حال اگر با کسب اطلاعات جدید، از راه رخ دادن پیشامد B ، درباره رخ دادن A تغییر عقیده دهیم، احتمال جدید A را احتمال پسین می‌نامیم. برای این که مفهوم این دو احتمال روشن شود به مثال زیر توجه کنید.

مثال ۴۳ شخصی سکه‌ای در دست دارد که عقیده داریم با احتمال $0/1$ دو شیری است. یعنی هر دو طرف آن شیر هستند. این احتمال را که جنبه شخصی دارد، احتمال پیشین می‌گوئیم.
 برای بررسی عقیده خود در مورد دو شیری بودن سکه، از این شخص می‌خواهیم که سکه را به تصادف بیندازد. مطلوبست احتمال دو شیری بودن سکه به شرط اینکه شیر بنیاید (این احتمال را احتمال پسین می‌گوئیم). اگر سکه یک شیری باشد آن را ناریب فرض می‌کنیم.
 پیشامد دو شیری بودن سکه را با A_1 و یک شیری بودن آن را با A_2 نشان می‌دهیم. فرض کنید B پیشامد شیر آمدن سکه باشد. واضح است که

$$P(A_1) = 0/1, \quad P(A_2) = 0/9, \quad P(B|A_1) = 1, \quad P(B|A_2) = \frac{1}{4}$$

از فرمول بیز داریم

$$P(A_1|B) = \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} \\ = \frac{0/1}{0/1 + 0/45} = 0/182$$

ملاحظه می شود که احتمال پیشین $1/10$ و احتمال پسین $182/10$ است. در حقیقت چون سکه شیر آمده است، عقیده ما بر دو شیر بودن سکه افزایش یافته است. چنانچه اگر خط آمده بود یقین پیدا می کردیم که سکه یک شیر است.

مثال ۴۴ نتیجه یک آزمایش پزشکی با احتمال $8/10$ برای بیماری سرطانی مثبت، و با احتمال $9/10$ برای دیگران منفی است. فرض کنید $200/1000$ از مردم شهری سرطان داشته باشند. مطلوب است احتمال این که نتیجه مثبت حاکی از سرطان باشد.

فرض کنید A پیشامد سرطان داشتن و B پیشامد مثبت بودن نتیجه آزمایش پزشکی باشد. بنابراین

$$P(B|A) = 8/10, \quad P(B'|A') = 9/10, \quad P(A) = 200/1000$$

می خواهیم $P(A|B)$ را محاسبه کنیم. طبق فرمول بیز داریم:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A')P(A')} \\ = \frac{(8/10)(200/1000)}{(8/10)(200/1000) + (9/10)(800/1000)} = 0.016$$

ملاحظه می شود احتمال این که شخصی در این شهر سرطان داشته باشد $200/1000$ است (احتمال پیشین)، ولی اگر نتیجه آزمایش سرطان در مورد این شخص مثبت باشد، این احتمال افزایش پیدا می کند و به 0.016 می رسد (احتمال پسین).

۶.۷.۲ تمرین بخش هفت

۱. یک شرکت لامپ سازی دارای سه کارخانه است که 15% ، 35% ، 50% از لامپها را تولید می کنند. فرض کنید احتمال کم عمر بودن لامپی که توسط این کارخانه ها تولید می شود به ترتیب $2/10$ ، $1/10$ و $5/10$ باشد. اگر لامپی محصول این شرکت باشد، احتمال کم عمر بودن آن چقدر است؟

۲. در ظرفی دو مهره سفید و سه مهره سیاه داریم. از این ظرف یک مهره به تصادف بیرون آورده مهره ای با رنگ مخالف در آن می گذاریم. سپس یک مهره دیگر از ظرف بیرون می آوریم. احتمال این که مهره دوم سفید باشد چقدر است؟ اگر مهره دوم سفید باشد، احتمال سیاه بودن مهره اول چقدر است؟

۳. دو ظرف داریم. ظرف اول محتوی دو مهره سفید و سه مهره سیاه، و ظرف دوم محتوی سه مهره سفید و دو مهره سیاه است. چشم بسته یک مهره از ظرف اول بیرون آمده و در ظرف دوم

می گذاریم. سپس از ظرف دوم دو مهره بیرون می آوریم. احتمال این که این دو مهره سفید باشند چقدر است؟ اگر این دو مهره سفید باشند، احتمال سفید بودن مهره اول چقدر است؟

۴. در مثال (۴۳) اگر سکه را مستقلاً دوبار بیندازیم و هر دو بار شیر بیاید، احتمال پسین چقدر است؟ مسئله را به جای دوبار، برای n بار حل کنید و نتیجه محاسبه را تفسیر نمایید.

۵. دادگاهی با احتمال $9/10$ متهم گناهکار را محکوم و با احتمال $95/100$ متهم بیگناه را تبرئه می کند. هرگاه $50/100$ از متهمان گناهکار باشند، احتمال این که محکوم شدن ناشی از گناهکاری باشد چقدر است؟

۶. شرکتی 40% لاستیک را از کارخانه A و بقیه را از کارخانه B می خرد. معمولاً 10% لاستیک کارخانه A و 5% لاستیک کارخانه B معیوب است. از این شرکت یک حلقه لاستیک خریده اند. اگر این حلقه معیوب باشد، احتمال اینکه از کارخانه A باشد چقدر است؟

۸.۲ پیشامدهای مستقل

فرض کنید A و B دو پیشامد با احتمال مثبت در یک فضای احتمال باشند. ممکن است علم به این که B رخ داده است بهیچوجه در احتمال رخ دادن A تاثیر نکند، یعنی رابطه عددی

$$P(A|B) = P(A) \quad (*)$$

صحت داشته باشد. در این صورت می گوئیم A مستقل از B است. به سادگی می توان نشان داد که اگر A مستقل از B باشد، B هم مستقل از A است، یعنی از $(*)$ داریم:

$$P(B|A) = P(B) \quad (**)$$

برای اثبات این موضوع، با استفاده از تعریف احتمال شرطی، قانون ضرب، و رابطه $(*)$ داریم:

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A|B)P(B)}{P(A)} = P(B)$$

ضمناً رابطه زیر هم که نسبت به A و B متقارن است، به دست می آید:

$$P(A \cap B) = P(A)P(B) \quad (***)$$

این سه رابطه با هم معادل هستند و می توان هر یک را به عنوان تعریف دو پیشامد مستقل به کار برد. با این حال برای این که نشان دهیم مستقل بودن دو پیشامد امری است متقارن و پیشامدها را به داشتن احتمال مثبت مقید نسازیم، رابطه سوم را به عنوان تعریف به کار می بریم.

۱.۸.۲ دو پیشامد مستقل

در یک فضای احتمال دو پیشامد A و B را مستقل می‌نامیم هرگاه داشته باشیم

$$P(A \cap B) = P(A)P(B) \quad (15)$$

باید بیفزائیم که ویژگی مستقل بودن دو پیشامد یک رابطه عددی است نه مجموعه‌ای. برقراری این رابطه برای دو پیشامد، به تابع احتمال P بستگی دارد. مفهوم مستقل بودن دو پیشامد، نقش عملدهای در پیشبرد احتمال و آمار داشته است. این مفهوم نظریه احتمال را از نظریه آنالیز ریاضی جدا ساخته و هویت هر یک را مشخص می‌کند.

مثال ۴۵ در ظرفی دو مهره سفید و یک مهره قرمز داریم. از این ظرف دو مهره یک به یک و با جایگذاری به تصادف بیرون می‌آوریم. فرض کنید A پیشامد سفید بودن مهره اول و B پیشامد قرمز بودن مهره دوم باشد. آیا A و B مستقل هستند؟

عقل سلیم به ما حکم می‌کند که A و B مستقل هستند، زیرا با جایگذاری مهره اول در ظرف شرایط ظرف تغییر نمی‌کند و سفید بودن مهره اول در شانس قرمز بودن مهره دوم تاثیری ندارد. با این حال موضوع را طبق تعریف بررسی می‌نمائیم.

برای این که مطلب روشن شود، فضای نمونه را در جدول زیر تشریح می‌کنیم. به عنوان مثال در این جدول منظور از (R, W_1) این است که مهره اول قرمز و مهره دوم سفید باشد.

| | | I | | |
|----|-------|--------------|--------------|------------|
| | | W_1 | W_2 | R |
| II | W_1 | (W_1, W_1) | (W_1, W_2) | (W_1, R) |
| | W_2 | (W_2, W_1) | (W_2, W_2) | (W_2, R) |
| | R | (R, W_1) | (R, W_2) | (R, R) |

A →

↓ B

ملاحظه می‌شود که $P(A) = \frac{6}{9}$ ، $P(B) = \frac{2}{9}$ و $P(A \cap B) = \frac{2}{9}$. بنابراین

$$P(A \cap B) = P(A)P(B) = \frac{2}{9}$$

اگر آزمایش را بدون جایگذاری انجام دهیم، قطعاً A و B مستقل نخواهد بود زیرا رنگ مهره اول در رنگ مهره دوم تاثیر دارد در این حال فضای نمونه تنها از شش پیشامد ساده تشکیل می‌شود. یعنی پیشامدهای ساده (W_1, W_1) ، (W_1, W_2) ، (W_2, W_1) و (W_2, W_2) را باید از جدول بالا حذف کرد. محاسبه نشان می‌دهد که در آزمایش بدون جایگذاری داریم:

$$P(A \cap B) = \frac{2}{6} \quad , \quad P(B) = \frac{2}{6} \quad , \quad P(A) = \frac{4}{6}$$

بنابراین

$$P(A \cap B) \neq P(A)P(B)$$

مثال ۴۶ یک بسته کارت ۵۲ تایی، به شماره‌های ۱ تا ۵۲ تایی، را به خوبی مخلوط می‌کنیم. سپس یک کارت به تصادف بیرون می‌آوریم. فرض کنید E پیشامد "شماره کارت کمتر از ۱۴" و F پیشامد "شماره کارت بیش از ۴" و مضرب ۴ باشد. آیا E و F مستقل اند؟

$$P(E \cap F) = \frac{2}{52} \quad , \quad P(F) = \frac{12}{52} \quad , \quad P(E) = \frac{12}{52}$$

ملاحظه می‌شود که

$$P(E \cap F) = P(E)P(F) = \frac{2}{52}$$

یعنی E و F دو پیشامد مستقل هستند.

۲.۸.۲ دو پیشامد مستقل و دو پیشامد جدا

اغلب دو پیشامد جدا را به غلط به عنوان دو پیشامد مستقل می‌پندارند. اگر A و B بخواهند هم مستقل و هم جدا باشند، لازم است داشته باشیم:

$$P(A \cap B) = P(A)P(B) \quad , \quad A \cap B = \emptyset$$

بنابراین لااقل یکی از این دو پیشامد باید دارای احتمال صفر باشد. یعنی دو پیشامد مستقل با احتمالهای مثبت هرگز نمی‌توانند جدا باشند، مگر اینکه یکی از آنها دارای احتمال صفر باشد.

مثال ۴۷ تاس نازیبی را می‌اندازیم. دو پیشامد خال فرد و خال زوج جدا هستند ولی مستقل نیستند. دو پیشامد خال زوج و خال مضرب سه مستقل هستند ولی جدا نیستند. با نوشتن پیشامدها و محاسبه احتمال آنها، درستی این مطالب را می‌توان نشان داد.

۳.۸.۲ چند پیشامد مستقل

پیشامدهای A_1, A_2, \dots, A_n مستقل هستند هرگاه برای هر k تایی آنها داشته باشیم

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \dots P(A_{i_k}), \quad 2 \leq k \leq n$$

از این تعریف معلوم می‌شود که اگر n پیشامد مستقل باشند، هر تعدادی از آنها هم مستقل خواهند بود.

مثال ۴۸ برای مستقل بودن سه پیشامد A_1, A_2, A_3 باید چهار رابطه زیر همگی برقرار باشند

$$P(A_1 \cap A_2) = P(A_1)P(A_2)$$

$$P(A_1 \cap A_3) = P(A_1)P(A_3)$$

$$P(A_2 \cap A_3) = P(A_2)P(A_3)$$

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$

سه رابطه اول نشان می‌دهند که این سه پیشامد دو به دو مستقل هستند، ولی رابطه چهارم از آنها به دست نمی‌آید. به عنوان مثال، یک مدل احتمال یکتواخت با فضای نمونه

$$S = \{e_1, e_2, e_3, e_4\}$$

و سه پیشامد

$$A_1 = \{e_1, e_2\}, \quad A_2 = \{e_1, e_3\}, \quad A_3 = \{e_1, e_4\}$$

را در نظر می‌گیریم. در اینجا این سه پیشامد دو به دو مستقل هستند، ولی همگی مستقل نیستند زیرا از

$$P(A_1 \cap A_2 \cap A_3) = P(e_1) = \frac{1}{4}$$

$$P(A_1)P(A_2)P(A_3) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{8}$$

معلوم می‌شود که رابطه چهارم برقرار نیست.

ضمناً از رابطه چهارم نمی‌توان بقیه روابط را به دست آورد، یعنی این رابطه به تنهایی برای مستقل بودن این سه پیشامد کافی نیست. به عنوان مثال یک مدل احتمال یکتواخت با فضای نمونه

$$S = \{e_1, e_2, \dots, e_8\}$$

$$A_1 = \{e_1, e_2, e_3, e_4\}, \quad A_2 = \{e_2, e_5, e_6, e_8\}, \quad A_3 = \{e_5, e_6, e_7, e_8\}$$

در نظر می‌گیریم. این سه پیشامد دو به دو مستقل نیستند، ولی رابطه چهارم صحت دارد زیرا

$$P(A_1 \cap A_2 \cap A_3) = P(e_8) = \frac{1}{8}$$

$$P(A_1)P(A_2)P(A_3) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \frac{1}{8}$$

۴.۸.۲ حاصلضرب دو مدل احتمال

دو مدل احتمال زیر را در نظر می‌گیریم.

مدل اول با فضای نمونه $S_1 = \{a_1, a_2, \dots, a_m\}$ و احتمالهای p_1, p_2, \dots, p_m . مدل دوم با فضای

نمونه $S_2 = \{b_1, b_2, \dots, b_n\}$ و احتمالهای q_1, q_2, \dots, q_n . فرض کنید

$$S = S_1 \times S_2 = \{(a_i, b_j) : i = 1, 2, \dots, m, \quad j = 1, 2, \dots, n\}$$

حاصلضرب دکارتی S_1 و S_2 باشد. چون

$$\sum_{i=1}^m p_i = \sum_{j=1}^n q_j = 1$$

پس داریم:

$$\sum_{i=1}^m \sum_{j=1}^n p_i q_j = 1$$

حال اگر به هر عضو S یعنی به (a_i, b_j) احتمال $P(a_j, b_i)$ را نسبت دهیم، یک مدل احتمال جدید با فضای نمونه S ، که آن را مدل حاصلضرب می نامند، بدست می آید.
می توان نشان داد که اگر E پیشامدی متعلق به مدل اول و F پیشامدی متعلق به مدل دوم باشد حاصلضرب دکارتی $E \times F$ پیشامدی متعلق به مدل حاصلضرب است و داریم

$$P(E \times F) = P(E)P(F) \quad (16)$$

معمولاً مدل احتمال دو آزمایش که پیاپی یا با هم انجام می گیرند و روی یکدیگر تاثیر نمی کنند مانند دوبار انداختن یک سکه، یا انداختن یک تاس و یک سکه با هم، یا یک به یک بیرون آوردن دو مهره از ظرفی با جایگذاری، به صورت مدل حاصلضرب می باشد.

مثال ۴۹ یک سکه و یک تاس نالریب را پیاپی یا با هم می اندازیم. مدل احتمال برای سکه از $S_1 = \{H, T\}$ با احتمال $\frac{1}{2}$ برای هر پیشامد ساده و برای تاس از $S_2 = \{1, 2, \dots, 6\}$ با احتمال $\frac{1}{6}$ برای هر پیشامد ساده تشکیل می شود. مدل حاصلضرب از

$$S = \{(H, 1), (H, 2), \dots, (H, 6), (T, 1), (T, 2), \dots, (T, 6)\}$$

با احتمال $\frac{1}{12}$ برای هر پیشامد ساده تشکیل می شود.
فرض بر این است که در مدل حاصلضرب آزمایشهای مربوط روی هم اثر ندارند بنابراین اگر E پیشامدی متعلق به مدل اول و F پیشامدی متعلق به مدل دوم باشد، در مدل حاصلضرب دو پیشامد $A = E \times S_2$ و $B = S_1 \times F$ مستقل هستند.
برای اثبات این موضوع از فرمول (۱۶) داریم:

$$P(B) = P(S_1)P(F) = P(F), \quad P(A) = P(E)P(S_2) = P(E) \quad (17)$$

افزون بر این می توان نشان داد که:

$$A \cap B = (E \times S_2) \cap (S_1 \times F) = E \times F \quad (18)$$

حال با استفاده از (۱۶)، (۱۷) و (۱۸) داریم:

$$P(A \cap B) = P(A)P(B)$$

مثال ۵۰ در مثال ۴۹ پیشامد شیر برای سکه و پیشامد خال زوج برای تاس مستقل هستند این موضوع بدون متوسل شدن به محاسبه، محسوس است. زیرا وقتی یک سکه و تاس را با هم می اندازیم، نقش سکه و خال تاس روی هم تاثیری ندارند. ضمناً پیشامد شیر مربوط به مدل احتمال سکه و پیشامد خال زوج مربوط به مدل احتمال تاس است.

با این حال مساله را از راه محاسبه حل می کنیم. فرض کنید A پیشامد شیر و B پیشامد خال زوج باشد. بدیهی است که

$$A = \{(H, 1), (H, 2), (H, 3), (H, 4), (H, 5), (H, 6)\} = \{H\} \times S_2$$

$$B = \{(H, 2), (H, 4), (H, 6), (T, 2), (T, 4), (T, 6)\} = S_1 \times \{2, 4, 6\}$$

$$A \cap B = \{(H, 2), (H, 4), (H, 6)\} = \{H\} \times \{2, 4, 6\}$$

$$P(A) = \frac{1}{2}, \quad P(B) = \frac{1}{3}, \quad P(A \cap B) = \frac{1}{6}$$

بنابراین A و B مستقل هستند.

مدل حاصلضرب را می توان به جای دو مدل به چند مدل که دارای فضای نمونه با پایان یا بی پایان می باشند تعمیم داد. مثلاً اگر در آزمایشی سکه ای را ده بار بیندازیم، مدل احتمال از حاصلضرب ده مدل احتمال آن سکه تشکیل می شود.
به عنوان مثالی دیگر، در آزمایشی نقطه ای به تصادف در داخل مربعی که طول هر ضلع آن یک واحد است، انتخاب می کنیم. مدل احتمال این آزمایش از حاصلضرب دو مدل احتمال یکنواخت در فاصله $[0, 1]$ تشکیل می شود.

۵.۸.۲ تمرین بخش هشت

۱. یک سکه نالریب را سه بار می اندازیم. فرض کنید A پیشامد مشاهده شیر و خط با هم و B پیشامد مشاهده حداکثر یک شیر باشد. نشان دهید A و B مستقل هستند. اگر سکه را دوبار بیندازیم، نشان دهید A و B مستقل نیستند.

۲. بسته کارت ۵ تایی، به شماره های ۱ تا ۵۲ تالی، را به خوبی مخلوط می کنیم. سپس دو کارت یک به یک با جایگذاری بیرون می آوریم. احتمال اینکه هر دو کارت شماره فرد داشته باشند چقدر است؟

فصل ۳

متغیر تصادفی - تابع توزیع - تابع چگالی

۱.۳ مفهوم متغیر تصادفی

برای اینکه مفهوم متغیر تصادفی روشن شود، نخست به ذکر یک مثال می‌پردازیم.

مثال ۱ سکه نالرایی را دوبار به تصادف می‌اندازیم، شماره شیرها را در این آزمایش تصادفی با X نشان می‌دهیم. واضح است که از پیشامدی به پیشامد دیگر، مقدار X تغییر می‌کند. مثلاً اگر پیشامد HH رخ دهد X برابر ۲ و اگر پیشامد HT رخ دهد X برابر ۱ می‌شود. افزون بر این در این تغییر، عامل شانس حاکم می‌باشد. مثلاً احتمال اینکه X برابر ۲ شود، $\frac{1}{4}$ و احتمال اینکه برابر ۱ شود $\frac{1}{4}$ می‌باشد. از اینرو X را یک متغیر تصادفی می‌نامند.

در حقیقت X تابعی است که دامنه آن فضای نمونه یعنی، $S = \{TT, TH, HT, HH\}$ و برد آن زیر مجموعه‌ای از مجموعه اعداد حقیقی R به صورت $S_X = \{0, 1, 2\}$ است. این زیر مجموعه اغلب تکیه‌گاه متغیر تصادفی X نامیده می‌شود. نگاره (۱) را ببینید. ملاحظه می‌شود که

$$X(TT) = 0, \quad X(TH) = X(HT) = 1, \quad X(HH) = 2$$

این متغیر تصادفی را، با دامنه و برد آن، به صورت زیر هم نشان می‌دهند:

$$X: \{TT, TH, HT, HH\} \rightarrow \{0, 1, 2\} \subset R$$

چون قانون شانس بر تغییرات تابع X حاکم است حتماً باید آن را تابع تصادفی بنامیم ولی با اصطلاحی نارسا، همه جا به متغیر تصادفی شهرت یافته است. ضمناً کاربرد X به جای تابع $X(x)$

۳. ثابت کنید اگر A و B مستقل باشند، A' و B مستقل هستند. درباره A' و B' چه فکر می‌کنید؟

۴. اگر $P(A) = 0$ یا $P(A) = 1$ ، می‌گوئیم A یک پیشامد غیر واقعی است. ثابت کنید پیشامد غیر واقعی A از هر پیشامد دیگر B مستقل است.

۵. فرض کنید A, B, C سه پیشامد مستقل باشند. ثابت کنید

$$C, A \cup B$$

$$C, A \cap B$$

$$C, A - B$$

$$C, A \Delta B$$

$$C, B, A'$$

مستقل هستند.

۶. ثابت کنید برای یک تاس نالرایی، پیشامدهای خال زوج و خال مضرب سه مستقل هستند. آیا این موضوع برای یک تاس اریب درست است؟ با دو مثال عددی موضوع را بررسی کنید.

۷. فرمول (۱۶) را ثابت کنید.

۸. یک تاس نالرایی و یک تاس اریب، که برای آن احتمال هر خال فرد $\frac{1}{9}$ و هر خال زوج $\frac{2}{9}$ است، در دست داریم. این دو تاس را با هم می‌اندازیم. مدل احتمال را به صورت حاصلضرب نشان دهید. ثابت کنید پیشامد خال زوج برای تاس نالرایی و خال مضرب سه برای تاس اریب مستقل می‌باشند. مساله را مانند مثال (۵۰) تشریح کنید.

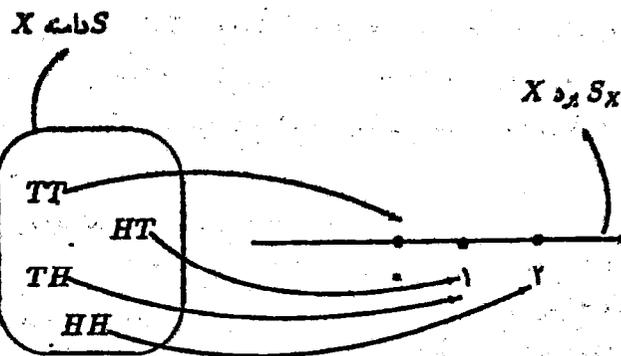
۹. سکه‌ای را سه بار می‌اندازیم. مدل احتمال را به صورت حاصلضرب تشریح کنید.

۱۰. نقطه‌ای به تصادف در مربعی که طول هر ضلع آن یک واحد است در نظر می‌گیریم. مدل احتمال را به صورت حاصلضرب تشریح کنید. احتمال اینکه طول این نقطه در فاصله $(\frac{1}{4}, 0)$ و عرض در فاصله $(0, \frac{1}{5})$ باشد چیست؟

۱۱. یک مدل احتمال یکنواخت با فضای نمونه

$$S = \{1, 2, \dots, N\}$$

در نظر می‌گیریم. ثابت کنید در این مدل دو پیشامد مستقل واقعی یافت می‌شوند، اگر و تنها اگر N عدد اول نباشد.



نگاره ۱ - متغیر تصادفی X به عنوان یک تابع

درست مانند کاربرد فربه جای تابع $f(x)$ است. در متغیر تصادفی $X(e)$ پیشامد ساده e نقش متغیر x را در تابع $f(x)$ دارد. پیشامدها را معمولاً با استفاده از X می‌توانیم نشان دهیم. مثلاً پیشامد کوچکتر بودن شماره شیرها از ۲ را به صورت زیر نشان می‌دهیم:

$$X < 2 = \{TT, TH, HT\}$$

احتمال این پیشامد را به صورت زیر می‌نویسیم:

$$P(X < 2) = P(TT, HT, TH) = \left(\frac{3}{4}\right)$$

به طور کلی پیشامد تعلق داشتن X به مجموعه عددی A به صورت زیر نوشته می‌شود:

$$X \in A = \{e \in S, X(e) \in A\}$$

در مثال ۱ داریم:

$$X \in \left[1, \frac{5}{4}\right] = \{TH, HT\}$$

$$X \in (0, 6) = \{HT, TH, HH\}$$

به ویژه پیشامد $X < 2$ را با $X \in (-\infty, 2)$ نشان می‌دهیم.

متغیر تصادفی را با حروف بزرگ لاتین، مثلاً X ، و عددی را که نتیجه انجام یک آزمایش تصادفی است، با حروف کوچک لاتین، مثلاً x ، نشان می‌دهیم. عدد x را یافته (یعنی مقدار مشاهده شده) متغیر تصادفی X می‌گوئیم. بنابراین x ، یعنی برد تابع X ، در حقیقت فضای یافته‌های X یا فضای نمونه این متغیر تصادفی است. اینک متغیر تصادفی را دقیقاً تعریف می‌کنیم.

۱.۱.۳ تعریف متغیر تصادفی

یک مدل احتمال، با فضای نمونه S ، را در نظر می‌گیریم. تابع حقیقی X را که دامنه آن S و بردش زیر مجموعه‌ای از اعداد حقیقی است، یک متغیر تصادفی روی این مدل احتمال می‌نامند. برد X را با S_x نشان می‌دهیم و آن را تکیه‌گاه X می‌نامیم. در حقیقت تابع X مجموعه S را که ممکن است عددی نباشد به یک مجموعه عددی S_x تبدیل می‌کند.

برای هر زیر مجموعه عددی A منظور از $(X \in A)$ ، پیشامد این است که یافته X در A قرار گیرد. احتمال این پیشامد را با $P(X \in A)$ نشان می‌دهیم.

حال به دو مثال دیگر که در یکی شمارش پذیر بی پایان و در دیگری شمارش ناپذیر است توجه می‌کنیم.

مثال ۲ بازی شیر و خط را با سکه‌ای نازیب، آنقدر ادامه می‌دهیم تا برای اولین بار شیر بیاید. در این آزمایش دنباله‌ای از T را که به H انتهای می‌شود، مانند $TTHH$ ، پیشامدی ساده می‌پنداریم. فضای نمونه عبارت است از:

$$S = \{H, TH, TTH, \dots\}$$

اینک فرض کنید متغیر تصادفی X شماره بازبهای لازم برای آمدن اولین شیر باشد. مثلاً اگر با سه بازی، شیر بیاید، داریم: $X(TTH) = 3$. در این مثال تکیه‌گاه X می‌شود مجموعه اعداد طبیعی، یعنی:

$$S_x = \{1, 2, \dots\}$$

احتمال پیشامد زوج بودن شماره بازبهای لازم برای مشاهده اولین شیر برابر است با:

$$P(X \in \{2, 4, \dots\}) = P(\{TH, TTH, \dots\}) \\ = \left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^4 + \dots = \frac{1}{3}$$

احتمال پیشامد بزرگتر بودن شمارهٔ بازبهای لازم از ۴، برای مشاهدهٔ اولین شیر برابر است با:

$$P(X > 4) = P(X \in \{5, 6, \dots\}) \\ = \left(\frac{1}{4}\right)^5 + \left(\frac{1}{4}\right)^6 + \dots = \frac{1}{16}$$

ملاحظه می‌شود که در مثال ۱ و ۲ مجموعهٔ یک‌شمارش‌پذیر می‌باشد. با این تفاوت که در مثال اول یک‌باپایان و در مثال دوم بی‌پایان است. در این دو مثال متغیر تصادفی، احتمال را از فضای نمونه به مجموعهٔ اعداد حقیقی می‌برد و در نقاط جداگانه جا می‌دهد. بدین جهت در این دو مثال X را متغیر تصادفی گسسته می‌نامند. حال به مثال زیر که در آن یک یک فاصله است توجه کنید.

مثال ۳ نقطه M را به تصادف در داخل دایره‌ای به شعاع ۲ واحد و مرکز صفر انتخاب می‌کنیم (دایره C). فضای نمونه S از تمام داخل این دایره تشکیل می‌شود. فرض کنید متغیر تصادفی X برای هر نقطه M ، اندازه پاره خط OM باشد، که آن را با x_M نشان می‌دهیم پس داریم: $X(M) = x_M$. تکیه‌گاه X می‌شود: $S_X = [0, 2]$. ملاحظه می‌شود که X احتمال را که بطور یکنواخت و پیوسته در داخل دایره توزیع شده است، بطور پیوسته، ولی نه یکنواخت به فاصله نیم باز $[0, 2]$ می‌برد.

اینک به عنوان مثال احتمال پیشامد اینکه نقطه‌ای انتخاب شود که فاصله آن از مرکز دایره کمتر از ۱ باشد را حساب می‌کنیم. این پیشامد می‌شود: $(X < 1)$ و از نقاط داخل دایره‌ای به شعاع ۱ و مرکز صفر تشکیل می‌شود (دایره C'). احتمال آن برابر است با مساحت دایره C' به مساحت دایره C یعنی:

$$P(X < 1) = \frac{\text{مساحت دایره } C'}{\text{مساحت دایره } C} = \frac{1}{4}$$

در این مثال تکیه‌گاه X یک فاصله عددی می‌باشد. از اینرو X را یک متغیر تصادفی پیوسته می‌نامند.

۲.۱.۳ متغیر تصادفی گسسته و پیوسته

متغیر تصادفی X را گسسته می‌گویند هرگاه برد آن یعنی یک مجموعهٔ عددی شمارش‌پذیر باشد و آن را پیوسته می‌گویند هرگاه یک فاصله عددی یا اجتماع چند فاصله عددی باشد.

۳.۱.۳ تمرین بخش یک

۱. یک سکه نازیب را سه بار به طور تصادفی می‌اندازیم. فرض کنید X شماره شیرها باشد.

الف - برای تمام پیشامدهای ساده، مقدار X را در یک جدول بنویسید

ب - احتمال پیشامدهای $(X \leq 2)$ ، $(X > 2)$ ، $(1 \leq X < 2)$ را پیدا کنید

ج - S_X را پیدا کنید. چرا X گسسته است؟

د - اگر Y شماره خطها باشد، آیا $X = Y$ ؟ آیا $S_Y = S_X$ ؟ چرا؟

۲. یک تاس نازیب را آنقدر می‌ریزیم تا برای اولین بار خال شش بیاید.

الف - اگر A پیشامد خال شش و B پیشامد خال غیر شش، در هر ریزش باشد، فضای نمونه

و احتمال هر پیشامد ساده را پیدا کنید.

ب - فرض کنید X شماره ریزشهای لازم برای مشاهدهٔ اولین شش باشد. S_X را پیدا کنید. آیا X

گسسته است؟ چرا؟

ج - احتمال اینکه X مضرب سه باشد چقدر است؟

۳. نقطه M را به تصادف در داخل مربع $ABCD$ که طول هر ضلع آن یک واحد است انتخاب می‌کنیم.

متغیر تصادفی X فاصله M از قطر AC می‌باشد.

الف - S_X را پیدا کنید. آیا X پیوسته است؟ چرا؟

ب - پیشامد $(X < \frac{1}{4})$ را پیدا کنید و احتمال آن را بدست آورید.

۴. دو تاس نازیب را به تصادف با هم می‌ریزیم. فرض کنید X مجموع خالها باشد.

الف - S_X را پیدا کنید.

ب - $P(X > 10)$ و $P(1 < X \leq 4)$ را حساب کنید.

۵. در بسته‌ای ۵ عدد لامپ داریم که دو عدد آنها معیوب می‌باشد. یک نمونه سه تایی به تصادف از این

بسته بیرون می‌آوریم. فرض کنید X شماره لامپهای معیوب در این نمونه باشد.

الف - فضای نمونه را بنویسید. پیشامدهای $(X < 2)$ و $(1 < X < 3)$ را بنویسید.

ب - تکیه‌گاه X را پیدا کنید.

۲.۳ تابع توزیع

در بخش پیش دیدیم که چگونه می‌توان به کمک متغیر تصادفی احتمال را از S به فضای نمونه که ممکن است یک مجموعهٔ غیر عددی باشد، به R ، یعنی مجموعه اعداد حقیقی منتقل کرد. اینک با

معرفی تابع توزیع نشان می‌دهیم که چگونه روی R می‌توان احتمال را به کمک یک تابع نقطه‌ای، به جای یک تابع مجموعه‌ای، محاسبه کرد.

برای هر عدد حقیقی x فاصله نیم‌بازی به صورت $A_x = (-\infty, x]$ و متغیر تصادفی X را در نظر می‌گیریم. احتمال اینکه X در این فاصله نیم باز باشد برابر است با:

$$P(X \in A_x) = P(X \leq x)$$

این احتمال یک تابع نقطه‌ای از x است که آن را با $F_X(x)$ نشان می‌دهند و تابع توزیع X می‌نامند.

۱.۲.۳ تعریف تابع توزیع

فرض کنید X یک متغیر تصادفی روی یک مدل احتمالی باشد. تابع حقیقی

$$F_X(x) = P(X \leq x)$$

را، که در آن x یک عدد حقیقی است، تابع توزیع X می‌نامیم.

هرگاه تنها با یک متغیر تصادفی X سروکار داشته باشیم، این تابع را با $F(x)$ نشان می‌دهیم. ولی اگر مثلاً دو متغیر تصادفی X و Y را داشته باشیم، تابع توزیع یکی را با $F_X(x)$ و دیگری را با $F_Y(y)$ نشان می‌دهیم.

اگر تمام احتمال را به یک واحد جرم که روی R ، یا محور طولها، به طور پیوسته یا گسته توزیع شده است تعبیر کنیم، می‌توان $F(x)$ را بدین طریق توجیه نمود: مقدار $F(x)$ در نقطه ثابت x برابر است با تمام جرم احتمالی که روی محور طولها در سمت چپ نقطه x و در خود x داریم. به عبارت دیگر $F(x)$ برابر است با احتمال انباشته شده روی محور طولها از $-\infty$ تا خود x . از اینرو تابع را تابع انباشتگی احتمال هم می‌گویند.

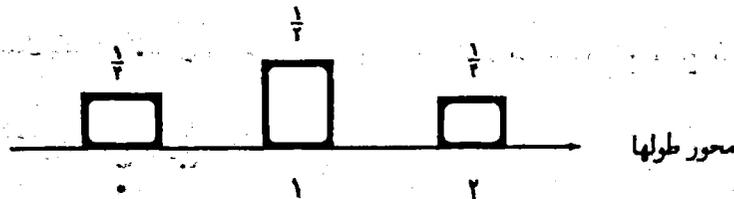
مثال ۴ (حالت گسته) سکه‌ای ناریب را به تصادف دوبار می‌اندازیم. فرض کنید متغیر تصادفی X شماره شیرها باشد. می‌خواهیم تابع توزیع X را پیدا کنیم.

همانطوری که در مثال ۱ دیدیم X یکی از اعداد ۰، ۱، ۲ را با احتمال مثبت طبق جدول زیر، که آن را جدول توزیع احتمال X می‌نامند، می‌پذیرد.

می‌توان گفت که یک واحد جرم احتمال طبق نگاره ۲ روی محور x ها در نقاط ۰ و ۱ و ۲ توزیع شده است.

جدول ۱ جدول توزیع احتمال شماره شیرها

| جمع احتمال | ۰ | ۱ | ۲ | $X=x$ |
|------------|---------------|---------------|---------------|----------|
| ۱ | $\frac{1}{4}$ | $\frac{2}{4}$ | $\frac{1}{4}$ | $P(X=x)$ |



نگاره ۲ نمایش توزیع جرم احتمال روی محور x ها

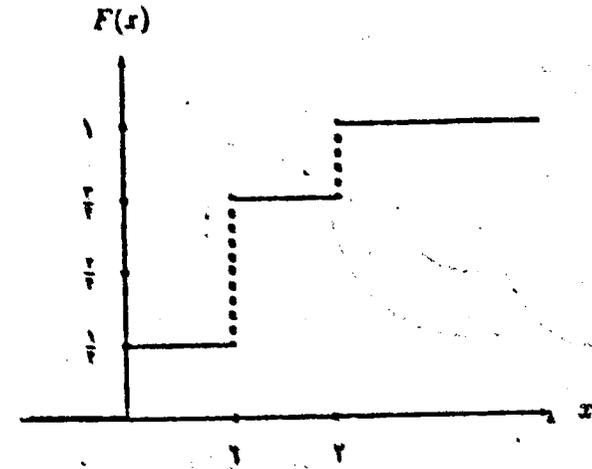
با توجهی که برای $F(x)$ کردیم، به عنوان مثال مقدار تابع توزیع X در نقطه $x = \frac{3}{4}$ برابر است با تمام احتمالی که در سمت چپ این نقطه و در خود این نقطه (در صورتیکه در این نقطه احتمال داشته باشیم) موجود است. چون احتمال تنها میان ۰ و ۱ و ۲ توزیع شده است پس:

$$F\left(\frac{3}{4}\right) = \frac{1}{4} + \frac{2}{4} = \frac{3}{4}$$

بطور کلی، طبق توجیه بالا، فرمول جبری $F(x)$ برای تمام مقادیر x به صورت زیر است:

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{4} & 0 \leq x < 1 \\ \frac{3}{4} & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

مثلاً $F\left(\frac{1}{4}\right) = \frac{1}{4}$ زیرا $-\infty$ تا نقطه‌ای به طول $\frac{1}{4}$ به اندازه $\frac{1}{4}$ احتمال (یعنی تنها احتمال در نقطه صفر) انباشته شده است. در این مثال $F(x)$ یک تابع پله‌ای است که نمودار آن در نگاره (۳) داده شده است.



نگاره ۳ نمودار F(x) در حالت گسسته

مثال ۵ (حالت پیوسته) نقطه M را به تصادف در داخل دایره‌ای به شعاع ۲ و مرکز O انتخاب می‌کنیم (دایره C). فرض کنید متغیر تصادفی X اندازه پاره خط OM باشد می‌خواهیم تابع توزیع X را پیدا کنیم.

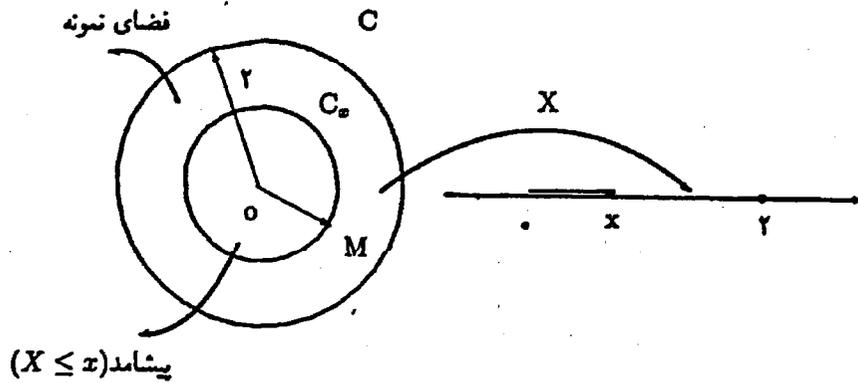
همانطوری که در مثال ۳ دیدیم X می‌تواند یکی از اعداد فاصله نیم باز [۰، ۲] را بپذیرد. بنابراین می‌توان گفت که یک واحد جرم احتمال به طور پیوسته روی محور طولها در این فاصله توزیع شده است. برای پیدا کردن F(x) پیشامد (X ≤ x) را که از تقاطع داخل و محیط دایره‌ای به مرکز O و شعاع x تشکیل می‌شود در نظر می‌گیریم (دایره C). نگاره (۴) را ببینید. طبق تعریف داریم:

$$F(x) = P(X \leq x) = \frac{\text{مساحت دایره } C_x}{\text{مساحت دایره } C} = \frac{x^2}{4}$$

با توجه به اینکه احتمال تنها در فاصله [۰، ۲] توزیع شده است فرمول جبری F(x) برای تمام مقادیر x به صورت زیر است:

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{4} & 0 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

در این مثال F(x) یک تابع پیوسته است که نمودار آن در نگاره (۵) داده شده است.



نگاره ۴ طرز تعیین توزیع F(x)

به کمک دو مثال بالا می‌توان خواص تابع توزیع را به آسانی تشریح کرد. پیش از این کار نکات زیر را درباره یک تابع حقیقی g(x) یادآور می‌شویم:

۱. مقدار g(x) را در x=a یا g(a) نشان می‌دهیم.

۲. حد g(x) را وقتی x از راست (یا چپ) به a نزدیک می‌شود، در صورتی که وجود داشته باشد، با g(a+) (یا g(a-)) نشان می‌دهیم و آن را حد راست (یا چپ) g(x) در x=a می‌نامیم.

۳. اگر g(a+) = g(a) (یا g(a-) = g(a))، آنگاه g(x) در x=a از طرف راست (یا چپ) پیوسته است.

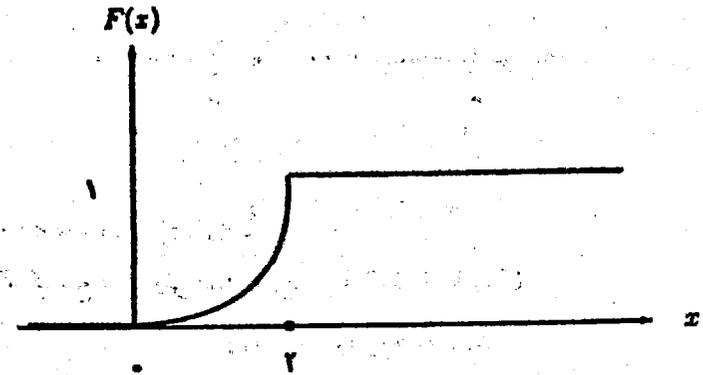
۴. اگر g(a-) = g(a+) = g(a)، آنگاه g(x) را در x=a پیوسته می‌نامیم.

۵. حد g(x) را وقتی x → +∞ (یا x → -∞)، در صورتی که وجود داشته باشد، با g(+∞) (یا g(-∞)) نشان می‌دهیم.

حال به نمودار F(x) در مثال ۴، که در نگاره (۳) داده شده است توجه می‌کنیم. ملاحظه می‌شود که مثلاً برای x=1 داریم:

$$F(1) = \frac{1}{4}, \quad F(1^+) = \frac{1}{4}, \quad F(1^-) = \frac{1}{4}$$

بنابراین F(x) در x=1 از طرف راست پیوسته است، ولی از طرف چپ ناپیوسته می‌باشد.



نگاره ۵ نمودار $F(x)$ در حالت پیوسته

با اندازه احتمال در $x=1$ می باشد. $F(1) - F(1^-) = \frac{1}{4}$ را جهش $F(x)$ در $x=1$ می نامند. ملاحظه می شود که این جهش درست برابر

در همین مثال برای $x = \frac{3}{4}$ داریم:

$$F\left(\frac{3}{4}\right) = \left(\frac{3}{4}\right)^+ = F\left(\frac{3}{4}\right)^- = \frac{3}{4}$$

بنابراین $F(x)$ در $x = \frac{3}{4}$ پیوسته است.

ضمناً ملاحظه می شود که $F(-\infty) = 0$ و $F(+\infty) = 1$.

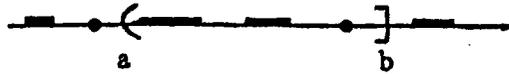
اینک به نمودار $F(x)$ در مثال ۵، که در نگاره ۵ داده شده است، توجه می کنیم. ملاحظه می شود که مثلاً برای $x=2$ داریم:

$$F(2) = F(2^+) = F(-2) = 1$$

بنابراین $F(x)$ در $x=2$ پیوسته است، ولی در این نقطه مشتق ندارد. در سایر نقاط این تابع پیوسته و دارای مشتق است.

۲.۲.۳ ویژگیهای تابع توزیع

فرض کنید $F(x)$ تابع توزیع متغیر تصادفی دلخواه (گسسته یا پیوسته) X باشد. یادآور می شویم که $F(x)$ یعنی احتمال انباشته شده روی محور طولها از $-\infty$ تا خود x . با این توجه ویژگیهای زیر را برای $F(x)$ می توان بیان کرد:



نگاره ۶ احتمال روی محور x ها

الف - $0 \leq F(x) \leq 1$ ، $-\infty < x < \infty$

ب - $F(x)$ یک تابع غیر نزولی است یعنی:

$$x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$$

ج - $F(-\infty) = 0$ ، $F(\infty) = 1$

د - برای هر عدد حقیقی $x=a$ تابع $F(x)$ لااقل از طرف راست پیوسته است یعنی:

$$F(a^+) = F(a)$$

ما از اثبات ویژگیهای بالا، که نیاز به آنالیز ریاضی دارند، خودداری می کنیم و از خواننده می خواهیم که با توجه به نمودارهای بالا درستی این ویژگیها را بررسی نماید.

برعکس می توان ثابت کرد که هر تابع حقیقی $F(x)$ با ویژگیهای بالا، تابع توزیع یک متغیر تصادفی X است.

۳.۲.۳ محاسبه احتمال از روی تابع توزیع

فرض کنید X یک متغیر تصادفی با توزیع $F(x)$ باشد. فاصله عددی دلخواه I را در نظر می گیریم. می خواهیم $P(X \in I)$ را بر حسب $F(x)$ حساب کنیم.

به عنوان مثال نخست فاصله $I = (a, b]$ را در نظر می گیریم. تمام احتمال را به عنوان یک واحد جرم که روی محور طولها به طور گسسته یا پیوسته توزیع شده است تعبیر می کنیم. $P(X \in I)$ به منزله جرمی است که در فاصله $(a, b]$ داریم. با توجه به نگاره (۶) ملاحظه می شود که:

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a)$$

با استفاده از تعریف $F(x)$ داریم:

$$P(a < X \leq b) = F(b) - F(a)$$

البته رابطه بالا را می توان دقیقاً ثابت کرد. ولی احساس درستی این رابطه با توجه بالا کافی است و عملاً بسیار مفید است. به همین طریق می توان تساویهای زیر را به آسانی پیدا کرد:

$$P(a < X < b) = F(b^-) - F(a)$$

$$P(a \leq X \leq b) = F(b) - F(a^-)$$

$$P(a \leq X < b) = F(b^-) - F(a^-)$$

$$P(X < b) = F(b^-)$$

$$P(X > a) = 1 - F(a)$$

به ویژه هرگاه فاصله I از یک نقطه x تشکیل شده باشد داریم:

$$P(X = x_0) = P(x_0 \leq X \leq x_0) = F(x_0) - F(x_0^-)$$

اگر $F(x)$ در x_0 پیوسته باشد این احتمال برابر صفر است. در غیر این صورت برابر با جهش $F(x)$ در x_0 است.

بنابراین هرگاه X یک متغیر تصادفی پیوسته باشد احتمال اینکه درست برابر با یک عدد دلخواه شود همواره صفر است، ولی برای یک متغیر تصادفی گسسته ممکن است این احتمال مثبت باشد.

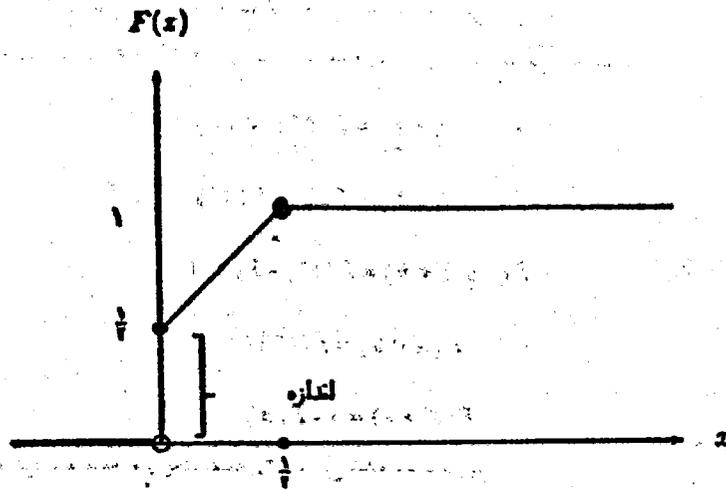
مثال ۶ تابع زیر را که نمودار آن در نگاره ۷ داده شده است در نظر می گیریم.

$$F(x) = \begin{cases} 0 & x < 0 \\ x + \frac{1}{4} & 0 \leq x < \frac{1}{4} \\ 1 & x \geq \frac{1}{4} \end{cases}$$

این تابع دارای تمام ویژگیهای یک تابع توزیع است. بنابراین $F(x)$ تابع توزیع یک متغیر تصادفی X می باشد.

ملاحظه می شود که در $x = 0$ داریم.

$$F(0^-) = 0, F(0^+) = \frac{1}{4}, F(0) = \frac{1}{4}$$



نگاره ۷ نمودار $F(x)$

بنابراین $F(x)$ در $x = 0$ ناپیوسته است. با این حال پیوسته از طرف راست است، زیرا $F(0) = F(0^+) = \frac{1}{4}$ احتمال در $x = 0$ می شود

$$P(X = 0) = F(0) - F(0^-) = \frac{1}{4}$$

مشاهده می شود که نمودار در $x = 0$ به اندازه $\frac{1}{4}$ جهش دارد در سایر نقاط $F(x)$ پیوسته است. در واقع $\frac{1}{4}$ کل احتمال در $x = 0$ متمرکز شده است و بقیه احتمال به طور پیوسته در فاصله $[0, \frac{1}{4}]$ توزیع شده است. به عبارت دیگر متغیر تصادفی X آمیخته از نوع گسسته و پیوسته می باشد بدین جهت آن را یک متغیر تصادفی آمیخته می نامند.

حال احتمال پیشامد $0 < X < \frac{1}{4}$ را حساب می کنیم

$$P(0 < X < \frac{1}{4}) = F(\frac{1}{4}^-) - F(0) = 1 - \frac{1}{4} = \frac{3}{4}$$

بطور فشرده از مثالهای بالا معلوم می شود که سه نوع تابع توزیع داریم. برحسب اینکه یک واحد جرم احتمال چگونه توزیع شده باشد:

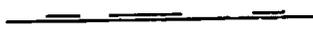
۱. هرگاه احتمال در نقاط جداگانه توزیع شده باشد، X از نوع گسسته و نمودار $F(x)$ پله ای است. $F(x)$ در این نقاط ناپیوسته و دارای جهش است.

X از نوع گسته



۲. هرگاه احتمال در یک یا چند فاصله به طور پیوسته توزیع شده باشد، X از نوع پیوسته و نمودار $F(x)$ پیوسته است.

X از نوع پیوسته



۳. هرگاه احتمال در یک یا چند فاصله به طور پیوسته و در یک یا چند نقطه به طور جداگانه توزیع شده باشد، X از نوع آمیخته است و نمودار $F(x)$ پیوسته با یک یا چند نقطه ناپیوستگی است که در آنها جهش دارد.

X از نوع آمیخته



پیش از پایان این بخش دو مطالب مهم دیگر را شرح می‌دهیم: یکی متغیر تصادفی ثابت و دیگری متغیرهای تصادفی هم‌توزیع.

۴.۲.۳ متغیر تصادفی ثابت

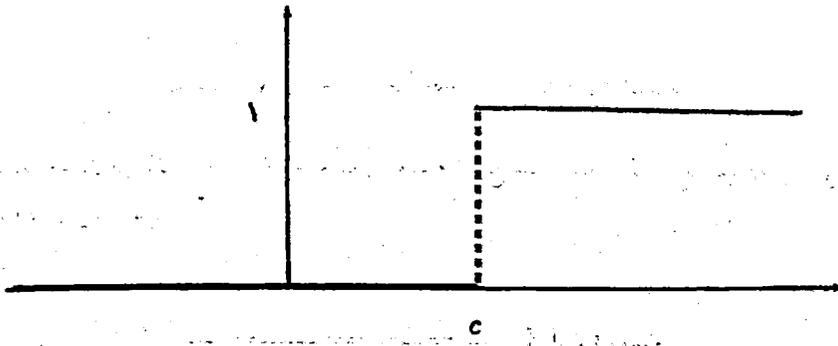
هرگاه تمام احتمال در نقطه $x=c$ متمرکز شده باشد، یعنی داشته باشیم $P(X=c)=1$ ، می‌گوئیم X یک متغیر تصادفی ثابت است. توزیع X ، که آن را یک توزیع تباہیده می‌نامند، به صورت زیر است:

$$F(x) = \begin{cases} 0 & x < c \\ 1 & x \geq c \end{cases}$$

نگاره (۸) منحنی این تابع توزیع را نشان می‌دهد.

ملاحظه می‌شود که $F(x)$ در $x=c$ به اندازه یک واحد جهش دارد.

چون تمام احتمال تنها در $x=c$ متمرکز دارد و توزیع نشده است، یقیناً X دارای مقدار ثابت c است. شاید اصطلاح توزیع تباہیده یا متغیر تصادفی تباہ شده هم به همین دلیل باشد. با به کار بردن این اصطلاح می‌توانیم به هر مقدار ثابت هم به چشم یک متغیر تصادفی نگاه کنیم.



نگاره ۸. نمودار توزیع تباہیده

۵.۲.۳ متغیرهای تصادفی هم‌توزیع

دو متغیر تصادفی X و Y را هم‌توزیع می‌نامند هرگاه برای هر عدد حقیقی z داشته باشیم:

$$P(X \leq z) = P(Y \leq z)$$

تساوی بالا، طبق تعریف تابع توزیع، معادل است با

$$F_X(z) = F_Y(z) \quad \text{برای هر عدد حقیقی } z$$

بنابراین دو متغیر تصادفی X و Y هم‌توزیع هستند، هرگاه توابع توزیع آنها برابر باشند. این نوع برابری را اغلب با $X \stackrel{D}{=} Y$ نشان می‌دهند. باید توجه داشت که این برابری مستلزم برابری بودن X و Y نیست، یعنی

$$X \stackrel{D}{=} Y \neq X = Y$$

مثال ۷ سکه نازیبی را دوبار به تصادف می‌اندازیم. فرض کنید X شماره شیرها و Y شماره خطها باشد.

در مثال ۴ توزیع X را پیدا کردیم. توزیع Y هم عیناً مثل توزیع X است. بنابراین، $X \stackrel{D}{=} Y$ ولی $X \neq Y$ زیرا مثلاً برای پیشامد ساده HH داریم.

$$X(HH) = 2 \quad Y(HH) = 0$$

یعنی X و Y نمی‌توانند دو متغیر تصادفی برابر، روی فضای نمونه $S = \{TT, HT, TH, TT\}$ باشند، در حالی که توزیع احتمال برای هر دو یکسان است.

۶.۲.۳ تمرین بخش دو

۱. سکه نازیبی را سه بار به تصادف می‌اندازیم. فرض کنید X شماره شیرها و Y شماره خطها باشد.
 - الف - جدول توزیع احتمال X را پیدا کنید. تابع توزیع X را بدست آورید و نمودار آن را رسم کنید.
 - ب - $P(X \geq 1)$ ، $P(2 \leq X < 3)$ و $P(X = 2)$ را از روی تابع توزیع و مستقیماً حساب کنید.
 - ج - نشان دهید که X و Y هم‌توزیع ولی نابرابر هستند.
 - د - فرض کنید $Z = X + Y$. تابع توزیع Z را پیدا کنید و نمودار آن را رسم نمایید.
۲. نقطه M را به تصادف در مربع $ABCD$ به ضلع یک واحد انتخاب می‌کنیم. فرض کنید متغیر تصادفی X فاصله این نقطه از قطر AC باشد.
 - الف - تابع توزیع X را پیدا کرده نمودار آن را رسم نمایید.
 - ب - $P(\frac{1}{6} < X < \frac{1}{4})$ را از روی تابع توزیع مستقیماً پیدا کنید.
 - ج - احتمال اینکه M روی قطر AC باشد چقدر است؟
 - د - اگر Y فاصله M از قطر BD باشد، نشان دهید X و Y هم‌توزیع ولی نابرابر می‌باشند.
۳. کدامیک از توابع زیر تابع توزیع می‌باشد؟ نمودار هر تابع را رسم کرده نقاط ناپیوستگی را پیدا کنید.

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{3} & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases} \quad G(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x} & x \geq 0 \end{cases}$$

$$H(x) = \begin{cases} 0 & x < 1 \\ \frac{x^2}{4} & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases} \quad T(x) = \begin{cases} 0 & x < 2 \\ x - 1 & 2 \leq x < 3 \\ 2 & x \geq 3 \end{cases}$$

۴. نشان دهید که

$$F(x) = \frac{e^x}{e^x + e^{-x}}$$

یک تابع توزیع است. نمودار این تابع را رسم کنید. m را طوری تعیین کنید تا $F(m) = \frac{1}{4}$ و $F(m)$ میانه توزیع می‌گویند.

۵. اگر X و Y هم‌توزیع باشند، ثابت کنید:

$$P(X > z) = P(Y > z)$$

۶. اگر X دارای توزیع داده شده در تمرین ۴ باشد، تابع توزیع $Y = e^X$ را بدست آورید و احتمالی:

$$F_Y(y) = P(Y \leq y) = P(e^X \leq y)$$

۷. در تمرین ۵ (بخش یک) تابع توزیع X را پیدا کرده نمودار آن را رسم نمایید.

۸. نشان دهید که

$$F(x) = 1 - \cos x \quad 0 < x < \frac{\pi}{2}$$

تابع توزیع یک متغیر تصادفی X است و آن را برای تمام اعداد حقیقی تعریف کنید. $P(X \leq \frac{\pi}{4})$ و میانه توزیع را بیابید.

۳.۳ تابع چگالی

فرض کنید X یک متغیر تصادفی با تابع توزیع $F(x)$ باشد. یادآور می‌شویم که برای هر x ثابت، $F(x)$ برابر است با احتمال انباشته شده روی محور طولها از $-\infty$ تا خود x . اینک می‌خواهیم تابعی تعریف کنیم که احتمال را در هر نقطه، برای متغیر تصادفی گسسته، و میزان فشردگی احتمال را در هر نقطه، برای متغیر تصادفی پیوسته تعیین کند: چنین تابعی را تابع چگالی احتمال یا تابع چگالی X می‌نامند و معمولاً آن را با $f_X(x)$ یا اگر تنها متغیر تصادفی X را داشته باشیم، با $f(x)$ نشان می‌دهند.

۱.۳.۳ تابع چگالی متغیر تصادفی گسسته

می‌دانیم که X را گسسته می‌گوئیم هرگاه احتمال را از فضای نمونه به محور طولها ببرد و در نقاط جداگانه جا دهد. مجموعه این نقاط را که با پایان یا شمارش پذیر می‌باشد تکیه گاه X می‌نامیم و با $\{x_1, x_2, \dots\}$ نشان می‌دهیم حال چگالی متغیر تصادفی X را به طریق زیر تعریف می‌کنیم:

۲.۳.۳ تعریف تابع چگالی متغیر تصادفی گسسته:

فرض کنید X یک متغیر تصادفی گسسته باشد. برای این متغیر، احتمال در هر نقطه روی محور طولیها تشکیل یک تابع حقیقی با ویژگیهای زیر را می‌دهد که این تابع را چگالی X می‌نامند.

$$(i) \quad f(x_i) > 0 \quad x_i \in S_X$$

$$(ii) \quad f(x_i) = 0 \quad x_i \notin S_X$$

$$(iii) \quad \sum_i f(x_i) = 1$$

برعکس هر تابع با ویژگیهای بالا، چگالی یک متغیر تصادفی گسسته است، تاکید می‌کنیم که برای یک متغیر تصادفی گسسته همواره داریم:

$$f(x_i) = P(X=x_i) \quad x_i \in S_X$$

از اینرو در بعضی از کتابها این تابع را تابع احتمال هم می‌نامند و با $P(x_i)$ نشان می‌دهند.

تابع توزیع برحسب تابع چگالی برای هر x ثابت، می‌شود:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

و تابع چگالی در هر x_i برحسب تابع توزیع می‌شود:

$$f(x_i) = P(X=x_i) = F(x_i) - F(x_{i-1})$$

منظور از $F(x_{i-1})$ ، حد چپ $F(x)$ در نقطه x_i است.

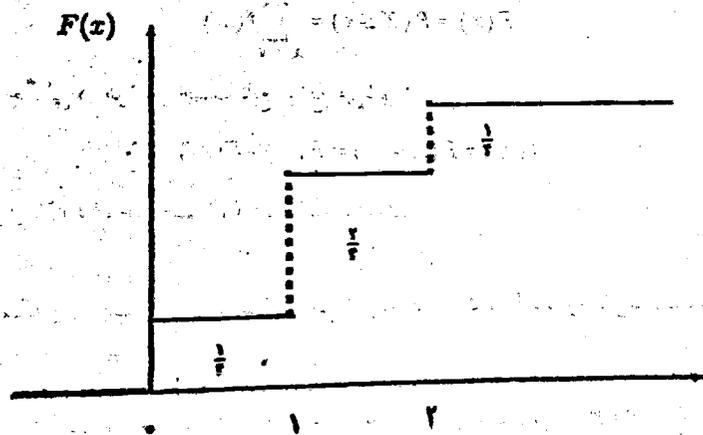
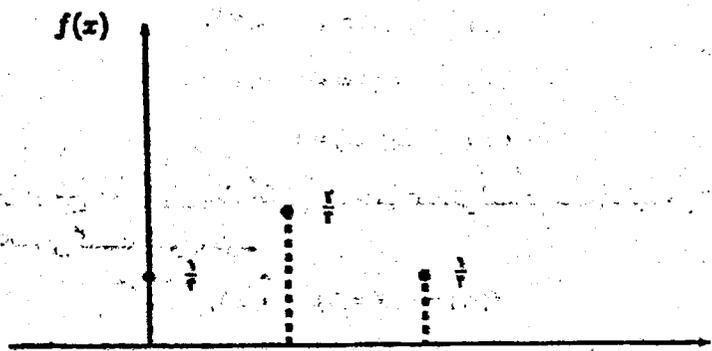
مثال ۸ سکه‌ای نارویب را دوبار به تصادف می‌اندازیم. فرض کنید X شماره شیرها باشد، در مثال ۱

دیدیم که $S_X = \{0, 1, 2\}$

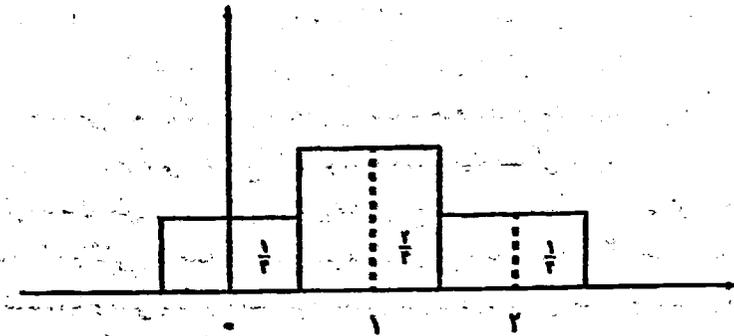
$$P(X=0) = \frac{1}{4}, \quad P(X=1) = \frac{1}{4}, \quad P(X=2) = \frac{1}{4}$$

تابع چگالی X می‌شود:

$$f(x) = \begin{cases} \frac{1}{4} & x=0 \\ \frac{1}{4} & x=1 \\ \frac{1}{4} & x=2 \\ 0 & \text{جای دیگر} \end{cases}$$



نگاره ۹ تابع توزیع و تابع چگالی شماره شیرها برای دو پرتاب سکه



نگاره ۱۰ - هیستوگرام تابع احتمال شماره شیرها در دو پرتاب سکه

نگاره (۹) ارتباط میان تابع توزیع و تابع چگالی را برای X نشان می‌دهد. به آسانی، به عنوان مثال، داریم

$$F(1) = \sum_{x \leq 1} f(x) = f(0) + f(1) = \frac{1}{4} + \frac{3}{4} = \frac{4}{4}$$

$$f(1) = F(1) - F(1^-) = \frac{4}{4} - \frac{1}{4} = \frac{3}{4}$$

اغلب برای X با مقادیر درست، نگاره (۱۰) را که هیستوگرام تابع احتمال نامیده می‌شود رسم می‌کنند. در این نگاره احتمال در هر نقطه به وسیله مساحت مستطیلی که قاعده آن یک واحد و ارتفاع آن به اندازه احتمال در آن نقطه است مشخص می‌شود. هیستوگرام تابع احتمال، با هیستوگرام داده‌ها که در گفتار نخست داشتیم شباهت دارد.

۳.۳.۳ تابع چگالی متغیر تصادفی پیوسته

می‌دانیم که X را پیوسته می‌گوئیم هرگاه احتمال را از فضای نمونه به محور طولها ببرد و در یک یا اجتماع چند فاصله به طور پیوسته جا دهد. مجموعه S_X یعنی تکیه‌گاه X ، یک یا اجتماع چند فاصله جدا است. بنابراین S_X یک مجموعه شمارش‌ناپذیر است. چون $F(x)$ ، یعنی تابع توزیع، پیوسته می‌باشد در نتیجه احتمال در هر نقطه برابر صفر است. برای تعریف تابع چگالی نخست یک مورد ساده

را تشریح می‌کنیم:

فرض کنید تمام یک واحد جرم احتمال در فاصله $S_X = [a, b]$ بطور پیوسته و یکنواخت توزیع شده باشد. به عبارت دیگر، با یک تعبیر فیزیکی، میله‌ای مستقیم و همگن که طول آن $b - a$ و جرم آن یک واحد باشد تجسم می‌کنیم. چگالی این میله برابر است با جرم یک واحد طول یعنی $\frac{1}{b-a}$. ملاحظه می‌شود که در این مثال چگالی یک تابع ثابت بوده به x نقاط میله بستگی ندارد.

حال اگر احتمال بطور پیوسته ولی نایکنواخت توزیع شده باشد، چگالی را نمی‌توان به طریق بالا پیدا کرد. کمالینکه چگالی یک میله مستقیم و غیر همگن (یعنی میله‌ای که فشردگی جرم در تمام طول آن یکنواخت نباشد) با چگالی یک میله مستقیم و همگن تفاوت دارد. در این حالت چگالی در هر نقطه x تابعی از x می‌باشد و آن را به طریق زیر پیدا می‌کنیم:

در داخل $S_X = [a, b]$ فاصله‌ای بسیار کوچک به طول γh و به مرکز ثابت x_0 اختیار می‌کنیم. اگر h به قدر کافی کوچک باشد می‌توان توزیع احتمال را در فاصله بسیار کوچک $[x_0 - h, x_0 + h]$ تقریباً یکنواخت تصور کرد. چون جرم احتمال در این فاصله برابر است با

$$P(x_0 - h \leq X \leq x_0 + h) = P(x_0 - h < X < x_0 + h) \\ = F(x_0 + h) - F(x_0 - h)$$

پس چگالی برای این فاصله کوچک، طبق تعبیر بالا، تقریباً برابر است با

$$\frac{F(x_0 + h) - F(x_0 - h)}{\gamma h}$$

حد کسر بالا را وقتی که $h \rightarrow 0$ ، در صورتی که وجود داشته باشد، با $f(x_0)$ نشان می‌دهند، و $f(x)$ را تابع چگالی متغیر تصادفی پیوسته X در نقطه‌ای دلخواه به طول x می‌نامند. اگر $F(x)$ در نقطه x_0 دارای مشتق باشد می‌توان ثابت کرد که

$$f(x_0) = \lim_{h \rightarrow 0} \frac{F(x_0 + h) - F(x_0 - h)}{\gamma h} \\ = \lim_{h \rightarrow 0} \frac{F(x_0 + h) - F(x_0)}{h} = F'(x_0)$$

چون احتمال به طور پیوسته در فاصله $[a, b]$ توزیع شده است، رابطه بالا برای تمام نقاط این فاصله، شاید بجز نقاط پایانی، درست می‌باشد. برای اینکه $f(x)$ یکنا باشد معمولاً در نقاطی که $F(x)$ مشتق ندارد $f(x)$ را برابر صفر می‌گیرند. این نتایج برای مواردی که S_X از اجتماع چند فاصله جدا تشکیل شده است نیز برقرار می‌باشند.

با این حساب $f(x)$ در تمام نقاط x بجز نقاط پایانی برابر است با مشتق $F(x)$. ضمناً در نقاط داخلی x ، تابع $f(x)$ کفو مثبت و در سایر نقاط صفر می‌باشد.
با توجه به ارتباط مشتق و انتگرال برای هر x در x روابط زیر را میان تابع چگالی و تابع توزیع داریم

$$F(x) = \int_{-\infty}^x f(t) dt$$

$$f(x) = F'(x)$$

اینک تابع چگالی را برای یک متغیر تصادفی پیوسته به طریق زیر تعریف می‌کنیم.

۴.۳.۳ تعریف تابع چگالی متغیر تصادفی پیوسته

فرض کنید X یک متغیر تصادفی پیوسته باشد میزان فشردگی احتمال در هر نقطه روی محور طولها تشکیل یک تابع حقیقی با خواص زیر می‌دهد که آن را تابع چگالی X می‌نامند:

(i) $f(x) \geq 0$ برای تمام x ها

(ii) $\int_{-\infty}^{\infty} f(x) dx = \int_{S_X} f(x) dx = 1$

بر عکس هر تابعی با خواص بالا تابع چگالی یک متغیر تصادفی پیوسته می‌باشد.

تاکید می‌کنیم که برای یک متغیر تصادفی پیوسته، برخلاف یک متغیر تصادفی گسسته، مقدار $f(x_0)$ برابر احتمال در x_0 نمی‌باشد و حتی ممکن است داشته باشیم $f(x_0) > 1$. چون $F(x)$ یک تابع پیوسته می‌باشد احتمال در هر نقطه برابر صفر است.

برای اینکه مفهوم تابع چگالی یک متغیر تصادفی پیوسته در نقطه ثابت x_0 روشن شود به توضیح زیر می‌پردازیم: فرض کنید در نقطه ثابت x_0 داشته باشیم: $f(x_0) = F'(x_0)$ با استفاده از تعریف مشتق می‌نویسیم:

$$f(x_0) = \lim_{h \rightarrow 0} \frac{F(x_0 + h) - F(x_0)}{h}$$

پس برای مقادیر بسیار کوچک h رابطه تقریبی زیر به دست می‌آید:

$$P(x_0 < X \leq x_0 + h) = F(x_0 + h) - F(x_0) = hf(x_0)$$

بنابراین احتمال در یک فاصله بسیار کوچک را می‌توان به کمک تابع چگالی از روی رابطه تقریبی بالا پیدا کرد. اگر $h \rightarrow 0$ ، آنگاه $P(X = x_0) = 0$.

نگاره (۱۱) ارتباط میان تابع توزیع و تابع چگالی را برای یک متغیر تصادفی پیوسته نشان می‌دهد توجه کنید که $P(a < X < b)$ برابر است با مساحت محصور بوسیله منحنی، تابع چگالی، محور طولها و خطوط $x = a$ و $x = b$ مساحت محصور میان منحنی چگالی و محور طولها برابر یک می‌باشد

مثال ۹ نقطه‌ای به تصادف در داخل دایره‌ای به شعاع ۲ و مرکز صفر انتخاب می‌کنیم فاصله این نقطه را تا مرکز دایره با متغیر تصادفی X نشان می‌دهیم. قبلاً دیده‌ایم که X یک متغیر تصادفی پیوسته با تابع توزیع زیر می‌باشد:

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{4} & 0 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

از این تابع نسبت به x مشتق می‌گیریم تا تابع چگالی X به دست آید:

$$f(x) = \begin{cases} \frac{x}{2} & 0 < x < 2 \\ \text{جای دیگر} & 0 \end{cases}$$

ملاحظه می‌شود که

$$\int_{-\infty}^{\infty} f(x) dx = \int_0^2 \frac{x}{2} dx = \left. \frac{x^2}{4} \right|_0^2 = 1$$

$$P\left(\frac{1}{2} < X < \frac{1}{2}\right) = \int_{\frac{1}{2}}^{\frac{1}{2}} \frac{x}{2} dx = \left. \frac{x^2}{4} \right|_{\frac{1}{2}}^{\frac{1}{2}} = \frac{1}{16} - \frac{1}{36} = \frac{5}{144}$$

مثال ۱۰ فرض کنید متغیر تصادفی X طول عمر یک لامپ بر حسب ساعت باشد نشان دهید که تابع زیر

$$f(x) = \begin{cases} \frac{1}{100} \exp\left\{-\frac{x}{100}\right\} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

می‌تواند یک تابع چگالی باشد. تابع توزیع X را پیدا کنید و $P(X > 50)$ را محاسبه نمایید

حل - تابع $f(x)$ دارای ویژگیهای یک تابع چگالی می باشد، زیرا

(i) $f(x) \geq 0$

(ii) $\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{\infty} \frac{1}{100} e^{-\frac{x}{100}} dx = 1$

پس $f(x)$ یک تابع چگالی است.
تابع توزیع را بدین طریق پیدا می کنیم:

$$F(x) = \int_{-\infty}^x f(t) dt = \int_{-\infty}^x \frac{1}{100} e^{-\frac{t}{100}} dt = 1 - e^{-\frac{x}{100}}$$

بنابراین تابع توزیع x برای تمام اعداد حقیقی چنین است:

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - \exp\left\{-\frac{x}{100}\right\} & x \geq 0 \end{cases}$$

برای محاسبه $P(X \geq 50)$ ، یعنی احتمال پیشامد اینکه طول عمر لامپ، لااقل ۵۰ ساعت باشد داریم:

$$P(X \geq 50) = 1 - P(X < 50) = 1 - P(X \leq 50) \\ = 1 - F(50) = 1 - 1 + e^{-\frac{1}{2}} = 0.607$$

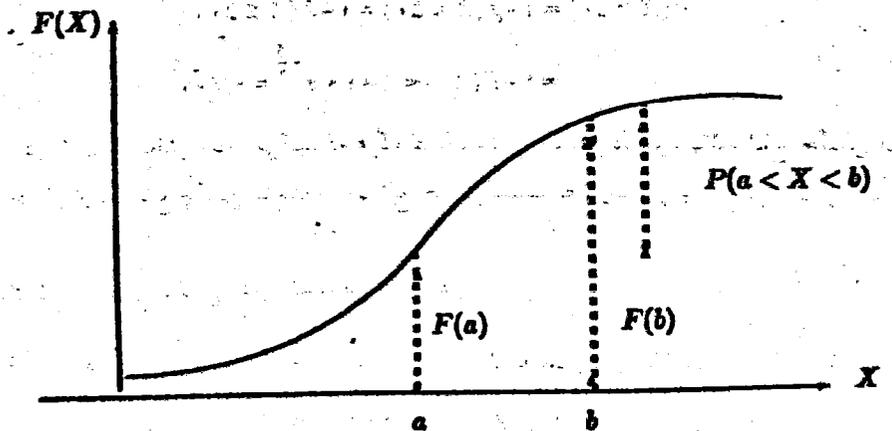
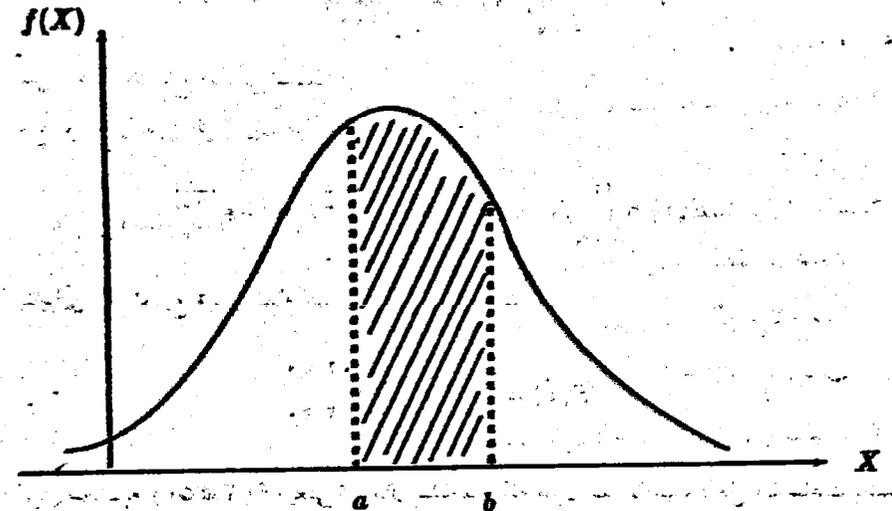
در پایان این بخش اشاره می کنیم که هرگاه X یک متغیر تصادفی دلخواه باشد تابع چگالی آن آمیخته ای از یک تابع چگالی پیوسته و یک تابع چگالی گسسته بصورت زیر می باشد:

۵.۳.۱ تابع چگالی متغیر تصادفی آمیخته

متغیر تصادفی X دارای چگالی آمیخته است اگر چگالی آن به صورت زیر باشد:

$$f(x) = pf_d(x) + (1-p)f_c(x)$$

در این رابطه $0 \leq p \leq 1$ و $f_d(x)$ و $f_c(x)$ به ترتیب دو چگالی گسسته و پیوسته می باشند. اگر $p = 1$ ، X از نوع گسسته و اگر $p = 0$ ، X از نوع پیوسته است. چون با متغیر تصادفی آمیخته از این نوع کمتر مواجه می شویم و تابع توزیع برای تشریح هر نوع متغیر تصادفی کافی است، از بحث در این باره خودداری می کنیم.



نگاره ۱۱ تابع توزیع و تابع چگالی یک متغیر تصادفی

۶.۳.۳ تمرین بخش سه

۱. سکه نازیبی را سه باره به تصادف می‌اندازیم. فرض کنید X شماره شیرها باشد تابع چگالی و تابع توزیع X را پیدا کنید. این دو تابع را رسم و با هم مقایسه نمایید. $F(2)$ و $f(2)$ را محاسبه کنید و منظور از آنها را بیان دارید.

۲. اگر در تمرین ۱ متغیر تصادفی Y شماره خطها باشد، آیا چگالی X و Y برابرند؟ چگالی $Z = X + Y$ را پیدا کنید.

۳. نقطه M را به تصادف در مربع $ABCD$ به ضلع واحد انتخاب می‌کنیم. فرض کنید متغیر تصادفی X فاصله این نقطه از قطر AC باشد.

الف - تابع چگالی X را پیدا کرده و نمودار آن را رسم کنید.

ب - $P(\frac{1}{5} < X < \frac{1}{4})$ را از روی تابع چگالی محاسبه کنید.

۴. کدامیک از توابع زیر تابع چگالی می‌باشد؟

$$f(x) = \begin{cases} x^2 & 0 < x < 1 \\ 0 & \text{جای دیگر} \end{cases}$$

$$f(x) = \begin{cases} 2e^{-2x} & x > 0 \\ 0 & \text{جای دیگر} \end{cases}$$

۵. m را طوری تعیین کنید تا

$$f(x) = \begin{cases} \frac{m}{\sqrt{x}} & 0 < x < 1 \\ 0 & \text{جای دیگر} \end{cases}$$

یک تابع چگالی شود.

۶. اگر $f_1(x)$ و $f_2(x)$ دو تابع چگالی باشند، ثابت کنید که با فرض $0 < p < 1$ ، تابع زیر هم یک تابع چگالی است:

$$f(x) = pf_1(x) + (1-p)f_2(x)$$

این چگالی را یک چگالی آمیخته از $f_1(x)$ و $f_2(x)$ می‌گویند. هر یک از دو چگالی $f_1(x)$ و $f_2(x)$ ممکن است گسسته یا پیوسته باشند.

۷. برای تابع توزیع مثال ۶، تابع چگالی را به صورت یک چگالی آمیخته بنویسید.

۸. در تمرین ۴ برای هر چگالی، تابع توزیع را بیابید و رسم کنید.

جدول ۲ فضای نمونه و مقادیر (X, Y)

| فضای نمونه | TTT | HTT | THT | TTH | HHT | HTH | THH | HHH |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|
| (X, Y) | (۰,۰) | (۱,۱) | (۱,۱) | (۱,۰) | (۲,۲) | (۲,۱) | (۲,۱) | (۳,۳) |

۴.۳ چگالی توام دو متغیر تصادفی گسسته

در یک آزمایش تصادفی ممکن است روی فضای نمونه دو متغیر تصادفی X و Y تعریف کنیم و بخواهیم آنها را به صورت یک متغیر تصادفی دو بعدی (X, Y) بررسی نماییم. در اینجا می‌توانیم قبل از هر چیز بعضی مفاهیمی را که برای یک متغیر تصادفی شرح دادیم، مانند تابع توزیع و تابع چگالی، برای (X, Y) هم تعمیم دهیم، و سپس به ارتباط این دو متغیر پردازیم. برای این منظور در حالت پیوسته آشنائی با انتگرال دوگانه ضرورت دارد. از اینرو تنها حالتی را بررسی می‌کنیم که X و Y هر دو گسسته باشند، بیشتر نتایج حاصل در حالت پیوسته هم درست هستند. به مثال زیر توجه کنید:

مثال ۱۱ سکه‌ای نازیبی را سه باره به تصادف می‌اندازیم. فرض کنید X شماره شیرها در این سه ریزش و Y شماره شیرها در دو ریزش اول باشد.

فضای نمونه از ۸ پیشامد ساده تشکیل شده است. در این مثال برای پیشامد ساده THH متغیر X برابر با ۲ و Y برابر با ۱ است. پس (X, Y) می‌شود $(2, 1)$. جدول ۲ در بالا فضای نمونه و مقادیری را که جفت (X, Y) می‌پذیرد نشان می‌دهد.

جدول (۳) را جدول توزیع احتمال توام (X, Y) می‌نامند. این جدول در عمل برای انجام محاسبات مفید است و از روی آن معلوم می‌شود که در کدام نقاط احتمال داریم.

ملاحظه می‌شود که سطر آخر جدول، توزیع احتمال X و ستون آخر جدول، توزیع احتمال Y را به ما می‌دهند.

با استفاده از جدول ۳ به تعریف زیر می‌پردازیم:

روابط زیر به آسانی نتیجه می‌شوند:

$$f(x, y) \geq 0$$

$$\sum_x \sum_y f(x, y) = 1$$

$$f_X(x) = \sum_y f(x, y)$$

$$f_Y(y) = \sum_x f(x, y)$$

در حالت پیوسته به جای علامت جمع ساده و دوگانه یعنی \sum و Σ به ترتیب علامت انتگرال ساده و دوگانه یعنی \int و \iint را به کار می‌بریم.

در مثال بالا X یکی از اعداد ۰، ۱، ۲، ۳ و Y یکی از اعداد ۰، ۱، ۲ را می‌پذیرد. البته این دو متغیر به نحوی با هم ارتباط دارند. مثلاً اگر X برابر ۳ شود Y الزاماً ۲ خواهد بود و نمی‌تواند آزادانه عدد دیگری را با احتمال مثبت بپذیرد. اصطلاحاً می‌گوئیم X و Y نامستقل هستند. اینک به مثال زیر توجه کنید:

مثال ۱۲ از یک دسته کارت چهارتایی به شماره‌های ۱، ۲، ۳، ۴ دو کارت پیاپی به تصادف با جایگذاری بیرون می‌آوریم. فرض کنید X شماره کارت اول و Y شماره کارت دوم باشد. جدول توزیع احتمال توام (X, Y) به صورت جدول (۴) است.

ملاحظه می‌شود که در این مثال X و Y هر کدام می‌توانند آزادانه یکی از اعداد ۱، ۲، ۳، ۴ را بپذیرند. بدین جهت می‌گوئیم X و Y مستقل هستند. البته اگر دو کارت را بی‌جایگذاری بیرون آوریم، آنگاه X و Y نمی‌توانند مستقل باشند. مثلاً اگر X عدد ۳ باشد؛ یعنی اول کارت شماره ۳ بیرون آید، دیگر Y نمی‌تواند ۳ شود. اینک دو متغیر تصادفی مستقل را در حالت گسسته تعریف می‌کنیم.

۲.۴.۳ تعریف دو متغیر تصادفی مستقل گسسته

فرض کنید X و Y دو متغیر تصادفی گسسته باشند، این دو متغیر را مستقل گوئیم هرگاه برای هر دو تایی (x, y) داشته باشیم:

$$P(X=x, Y=y) = P(X=x)P(Y=y)$$

جدول ۳ جدول توزیع احتمال توام (X, Y)

| $X \backslash Y$ | ۰ | ۱ | ۲ | ۳ | $f_Y(y)$ |
|------------------|---------------|---------------|---------------|---------------|---------------|
| ۰ | $\frac{1}{8}$ | $\frac{1}{8}$ | ۰ | ۰ | $\frac{2}{8}$ |
| ۱ | ۰ | $\frac{2}{8}$ | $\frac{2}{8}$ | ۰ | $\frac{4}{8}$ |
| ۲ | ۰ | ۰ | $\frac{1}{8}$ | $\frac{1}{8}$ | $\frac{2}{8}$ |
| $f_X(x)$ | $\frac{1}{8}$ | $\frac{2}{8}$ | $\frac{2}{8}$ | $\frac{1}{8}$ | ۱ |

۱.۴.۳ تعریف چگالی توام دو متغیر تصادفی گسسته

فرض کنید X و Y دو متغیر تصادفی گسسته باشند. تابع دو متغیری زیر را تابع چگالی توام X و Y می‌نامند:

$$f(x, y) = P(X=x, Y=y)$$

در هر نقطه (x, y) که احتمال داشته باشیم این تابع مثبت و در سایر نقاط صفر است. طبق جدول

بالا به عنوان مثال داریم:

$$f(2, 1) = \frac{2}{8}$$

$$f(1, 2) = 0$$

چگالی متغیر X را به تنهایی با $f_X(x)$ و چگالی متغیر Y را به تنهایی با $f_Y(y)$ نشان می‌دهیم و آنها را چگالیهای کناری می‌نامیم (زیرا در کنار پائین و کنار راست جدول بالا خودنمایی می‌کنند). ضمناً

جدول ۴ جدول توزیع احتمال توام X و Y

| $X \backslash Y$ | ۱ | ۲ | ۳ | ۴ | $f_Y(y)$ |
|------------------|----------------|----------------|----------------|----------------|---------------|
| ۱ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{4}$ |
| ۲ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{4}$ |
| ۳ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{4}$ |
| ۴ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{16}$ | $\frac{1}{4}$ |
| $f_X(x)$ | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{4}$ | ۱ |

ملاحظه می شود که این تعریف شبیه تعریف دو پیشامد مستقل است. برحسب چگالی می گوئیم X و Y مستقل هستند هرگاه برای دو تائی (x, y) داشته باشیم:

$$f(x, y) = f_X(x)f_Y(y)$$

این تعریف مستقل بودن برحسب چگالی، در حالت پیوسته هم به کار می رود.

۳.۴.۳ تعریف توزیع توام دو متغیر تصادفی

فرض کنید X و Y دو متغیر تصادفی دلخواه باشند (گسسته یا پیوسته). تابع دو متغیری زیر را تابع توزیع توام X و Y می نامند:

$$F(x, y) = P(X \leq x, Y \leq y)$$

x و y هر دو عدد حقیقی هستند و منظور از پیشامد $(Y \leq y)$ و $(X \leq x)$ اشتراک دو پیشامد $(X \leq x)$ و $(Y \leq y)$ است.

چگالی توام و توزیع توام بیش از دو متغیر تصادفی را هم می توان به طریق بالا تعریف کرد.

۴.۴.۳ تمرین بخش چهار

۱. در ظرفی ۳ مهره سفید، ۲ مهره سیاه، و ۱ مهره قرمز داریم. از این ظرف به تصادف ۵ مهره با هم بیرون می آوریم. فرض کنید X شماره مهره های سفید و Y شماره مهره های سیاه در این ۵ مهره باشد. جدول پیشامدهای ساده و مقادیر (X, Y) را برای هر پیشامد پیدا کنید. جدول توزیع احتمال (X, Y) را بدست آورید. آیا X و Y مستقل می باشند؟

۲. یک تاس ناریب را دوبار می اندازیم. فرض کنید X و Y به ترتیب شماره خالها در بار اول و دوم باشند. جدول توزیع احتمال (X, Y) را پیدا کنید. آیا X و Y مستقل هستند؟

۳. سکه ای ناریب را سه بار می اندازیم. فرض کنید X شماره بیشترین شیرهای پیاپی و Y شماره بیشترین خطهای پیاپی باشد. جدول پیشامدهای ساده و مقادیر (X, Y) را پیدا کنید. جدول توزیع احتمال (X, Y) را بدست آورید. آیا X و Y مستقل اند؟

۴. از یک دسته کارت چهارتائی به شماره های ۱، ۲، ۳، ۴ دو کارت پیاپی بی جایگذاری بیرون می آوریم. فرض کنید X شماره کارت اول و Y شماره کارت دوم باشد. جدول توزیع احتمال X و Y را پیدا کنید؟ آیا X و Y هم توزیع هستند؟

۵. دو سکه و یک تاس ناریب را با هم می ریزیم. فرض کنید X شماره شیرها و Y شماره خالها باشد. جدول توزیع احتمال (X, Y) را بیابید. آیا X و Y مستقل اند؟

$\frac{600}{400} = 1/5$ تومان باشد. بنابراین اگر ۲۰۰ بلیط به مبلغ ۶۰۰ تومان فروخته شود انتظار می‌رود که ماشین هم همین مبلغ را بپردازد.

در این مثال یک متغیر تصادفی X (یعنی شماره شیرها در ریزش سه سکه) داریم که جدول توزیع احتمال آن به صورت زیر است:

| | | | | |
|----------|---------------|---------------|---------------|---------------|
| $X=x$ | ۰ | ۱ | ۲ | ۳ |
| $P(X=x)$ | $\frac{1}{8}$ | $\frac{3}{8}$ | $\frac{3}{8}$ | $\frac{1}{8}$ |

عدد زیر را که حقیقت معدل وزنی شماره شیرها می‌باشد، امید ریاضی یا امید X می‌نامند و با $E(X)$ نشان می‌دهند (E حرف اول کلمه Expectation به معنای انتظار است)

$$E(X) = \frac{1}{8} \times 0 + \frac{3}{8} \times 1 + \frac{3}{8} \times 2 + \frac{1}{8} \times 3 = 1/5$$

تغییر فراوانی $E(X) = 1/5$ چنین است: اگر ماشین این سه سکه را بارها مثلاً ۱۰۰۰ بار بریزد، انتظار داریم معدل شیرها در هر بار بشود عدد $1/5$.

اینک مفهوم بالا را تعمیم داده آن را برای هر متغیر تصادفی گسسته یا پیوسته به صورت زیر تعریف می‌کنیم:

۱.۱.۴ تعریف امید ریاضی

فرض کنید X یک متغیر تصادفی با چگالی $f(x)$ باشد. امید ریاضی X عددی به صورت زیر است:

$$E(x) = \begin{cases} \sum x f(x) & \text{حالت گسسته} \\ \int_{-\infty}^{\infty} x f(x) dx & \text{حالت پیوسته} \end{cases}$$

امید ریاضی را با μ نشان می‌دهند و آن را میانگین X یا میانگین چگالی $f(x)$ هم می‌نامند. می‌دانیم که هرگاه x در تکیه گاه X یعنی S_X باشد، چگالی $f(x)$ مثبت و در جای دیگر صفر است. بنابراین باید مجموع یا انتگرال بالا را روی تکیه گاه X محاسبه کرد. اگر S_X بی‌پایان باشد یا برای بعضی عناصر آن $f(x)$ بی‌پایان شود، ممکن است عدد $E(X)$ وجود نداشته باشد. چون با این موارد کمتر مواجه می‌شویم، از این پس فرض می‌کنیم $E(X)$ وجود دارد. به طور کلی $E(X)$ در صورتی وجود دارد که مجموع یا انتگرال بالا به طور مطلق همگرا باشد.

امید ریاضی یک متغیر تصادفی و کاربرد آن

۱.۴ مفهوم امید ریاضی و ویژگیهای آن

در بخشهای پیش مفاهیمی از قبیل پیشامد، احتمال، و متغیر تصادفی را دقیقاً شرح دادیم. زندگی روزانه علاوه بر این مفاهیم با مفهوم دیگری به نام امید یا انتظار همراه است. در حالی که انسان همواره با امید و انتظار زندگی می‌کند، ولی نمی‌تواند بیش از حد به بعضی از رویدادها چشم امید داشته باشد. امید و انتظار هم مانند تصادف و شانس، در عین بی‌نظامی، تابع نظام و قوانینی می‌باشد که در این فصل می‌خواهیم آنها را مطالعه کنیم. از نظر تاریخی امید و انتظار مانند تصادف و احتمال از راه بازیهای قمار جنبه ریاضی پیدا کرده است. برای اینکه مطلب روشن شود به بازی شیر و خط زیر توجه نمائید:

فرض کنید با خریدن یک بلیط بتوانید در یک بازی تفریحی شرکت کنید. در این بازی یک ماشین سکه‌انداز با فشار تکمهای به تصادف سه سکه با رنگهای مختلف را با هم می‌ریزد و با فشار تکمهای دیگر به اندازه شماره شیرها، سکه ده ریالی به شما می‌پردازد.

صاحب ماشین مدعی است که اصلاً قصد قمار و سودجویی ندارد و صرفاً می‌خواهد مردم را سرگرم کند، ولی قیمت هر بلیط باید طوری باشد که بعد از چند بازی رویهم ضرر نکند. آیا منصفانه بهای هر بلیط چقدر باید باشد؟

چون سکه‌ها کاملاً به تصادف می‌ریزند، احتمال اینکه شماره شیرها ۰، ۱، ۲، ۳ شوند به ترتیب $\frac{1}{8}$ ، $\frac{3}{8}$ ، $\frac{3}{8}$ ، $\frac{1}{8}$ می‌باشد. حال اگر تعدادی زیاد بلیط، مثلاً ۴۰۰ بلیط، فروش رود انتظار می‌رود (بنابر تعبیر احتمال با فراوانی نسبی) ۵۰ مرتبه هیچ شیر، ۱۵۰ مرتبه یک شیر، ۱۵۰ مرتبه دو شیر، و ۵۰ مرتبه سه شیر مشاهده شود. بنابراین انتظار می‌رود که ماشین برای این ۴۰۰ بلیط مبلغ زیر را بپردازد:

$$400 \times \frac{1}{8} \times 0 + 400 \times \frac{3}{8} \times 1 + 400 \times \frac{3}{8} \times 2 + 400 \times \frac{1}{8} \times 3 = 600$$

تومان ۶۰۰ برای اینکه صاحب ماشین نه سود نماید نه زیان، بهای هر بلیط به طور متوسط یا منصفانه باید

مثال ۱ - (حالت گسسته) متغیر تصادفی X اعداد ۱، ۲، ۳، ۶ را به ترتیب با احتمالهای زیر

$$\frac{1}{4}, \frac{1}{8}, \frac{1}{2}, \frac{1}{8}$$

می پذیرد $E(X)$ را بیابید

حل - در این مثال $S_X = \{1, 2, 3, 6\}$. بنابراین داریم:

$$E(X) = \sum x f(x) = (1 \times \frac{1}{4}) + (2 \times \frac{1}{8}) + (3 \times \frac{1}{2}) + (6 \times \frac{1}{8}) = 2.75$$

این مثال را می توانیم به طریق زیر تعبیر کنیم:

یک دسته کارت در نظر می گیریم؛ دو کارت با شماره ۱، یک کارت با شماره ۲، چهار کارت با شماره ۳ و یک کارت با شماره ۶، کارتها را مخلوط کرده یکی را به تصادف بیرون می آوریم. شماره روی کارت بیرون آمده یک متغیر تصادفی X با امید ریاضی ۲/۷۵ است. حال اگر به تعدادی زیاد، مثلاً ۱۰۰ بار آزمایش را انجام دهیم و صد عددی را که مشاهده می کنیم جمع کرده معدل بگیریم، انتظار می رود این معدل بشود ۲/۷۵.

مثال ۲ - (حالت پیوسته) نقطه ای به تصادف در داخل دایره ای به شعاع ۲ و مرکز صفر انتخاب می کنیم. فاصله این نقطه تا مرکز را با متغیر تصادفی X نشان می دهیم. امید ریاضی X را پیدا کنید.

حل - همانطوری که در مثال ۹، فصل سوم، دیدیم X دارای تابع چگالی زیر است:

$$f(x) = \begin{cases} \frac{x}{2} & 0 < x < 2 \\ 0 & \text{جای دیگر} \end{cases}$$

فاصله $S_X = (0, 2)$ تکیه گاه X است. بنابراین داریم

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^2 x \cdot \frac{x}{2} dx = \frac{x^3}{6} \Big|_0^2 = \frac{4}{3}$$

اگر به تعدادی زیاد، مثلاً ۶۰ بار آزمایش را انجام دهیم و ۶۰ فاصله را جمع کرده معدل بگیریم، انتظار می رود این معدل بشود عدد $\frac{4}{3}$.

۲.۱.۴ امید ریاضی تابعی از یک متغیر تصادفی

در عمل بعضی اوقات باید امید ریاضی تابعی از متغیر تصادفی X را حساب کرد. مثلاً در ماشین سکه انداز، ممکن است صاحب ماشین بخواهد به اندازه $\frac{1}{5}$ پولی را که ماشین می ریزد مزد دریافت کند انتظار دارید هر بلیط را چقدر بفروشد؟ اگر پولی را که ماشین می ریزد با X نشان دهیم، مزد صاحب ماشین در برابر فروش هر بلیط مقدار تصادفی $\frac{X}{5}$ می شود. بنابراین امید ریاضی متغیر تصادفی

$$Y = X + \frac{X}{5} = \frac{6X}{5}$$

را که تابعی از X است حساب کرد. بطور کلی فرض کنید X یک متغیر تصادفی با چگالی $f(x)$ و $Y = g(X)$ یک تابع حقیقی باشد. امید ریاضی متغیر تصادفی $Y = g(X)$ را به صورت زیر تعریف می کنیم:

۳.۱.۴ تعریف امید ریاضی $g(X)$

امید ریاضی $g(X)$ عددی به صورت زیر است:

$$E[g(X)] = \begin{cases} \sum_x g(x) f(x) & \text{حالت گسسته} \\ \int_{-\infty}^{\infty} g(x) f(x) dx & \text{حالت پیوسته} \end{cases}$$

مثال ۳ - سکه ای را سه بار می ریزیم. فرض کنید X شماره شیرها باشد. امید ریاضی تابع $Y = X^2$ را پیدا کنید.

حل - متغیر تصادفی X یکی از اعداد ۰، ۱، ۲، ۳ را به ترتیب با احتمالهای $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$ می پذیرد. امید ریاضی Y می شود:

$$\begin{aligned} E(Y) &= E(X^2) = \sum x^2 f(x) \\ &= (0 \times \frac{1}{8}) + (1^2 \times \frac{3}{8}) + (2^2 \times \frac{3}{8}) + (3^2 \times \frac{1}{8}) \\ &= \frac{3}{8} + \frac{12}{8} + \frac{9}{8} = 3 \end{aligned}$$

مثال ۴ هرگاه X دارای چگالی زیر باشد، امید ریاضی $Y = \sqrt{X}$ را پیدا کنید.

$$f(x) = \begin{cases} \frac{x}{2} & 0 < x < 2 \\ 0 & \text{جای دیگر} \end{cases}$$

حل - چون X یک متغیر تصادفی پیوسته است، امید ریاضی X می شود:

$$E(Y) = E(\sqrt{X}) = \int_{-\infty}^{\infty} \sqrt{x} f(x) dx = \int_0^2 \sqrt{x} \cdot \frac{x}{2} dx \\ = \int_0^2 \frac{1}{2} x^{\frac{3}{2}} dx = \frac{4}{5} \sqrt{2}$$

۴.۱.۴ امید ریاضی یک چند جمله‌ای

فرض کنید X یک متغیر تصادفی با چگالی $f(x)$ و $Q(x)$ یک چند جمله‌ای درجه n بصورت $Q(x) = c_0 + c_1x + \dots + c_nx^n$ باشد با استفاده از تعریف امید ریاضی و خواص Σ یا \int داریم

$$E[Q(X)] = E(c_0 + c_1X + \dots + c_nX^n) = c_0 + c_1E(X) + \dots + c_nE(X^n)$$

برای اثبات، مثلاً در حالت پیوسته برای یک چند جمله‌ای درجه دو، داریم:

$$E(c_0 + c_1X + c_2X^2) = \int_{-\infty}^{\infty} (c_0 + c_1x + c_2x^2) f(x) dx \\ = \int_{-\infty}^{\infty} c_0 f(x) dx + \int_{-\infty}^{\infty} c_1 x f(x) dx + \int_{-\infty}^{\infty} c_2 x^2 f(x) dx \\ = c_0 \int_{-\infty}^{\infty} f(x) dx + c_1 \int_{-\infty}^{\infty} x f(x) dx + c_2 \int_{-\infty}^{\infty} x^2 f(x) dx \\ = c_0 + c_1 E(X) + c_2 E(X^2)$$

به ویژه برای تابع خطی $Y = bX + c$ داریم:

$$E(Y) = E(bX + c) = bE(X) + c$$

با فرض $b = 0$ نتیجه می‌گیریم که $E(c) = c$ یعنی امید ریاضی یک مقدار ثابت c ، می‌شود خود c .

مثال ۵ هرگاه $E(X) = 1$ و $E(X^2) = 2$ ، امید ریاضی $(X-2)(X+1)$ را پیدا کنید.

حل - چون $(X-2)(X+1) = X^2 - X - 2$ ، بنابراین داریم:

$$E[(X+1)(X-2)] = E(X^2 - X - 2) = E(X^2) - E(X) - 2 \\ = 2 - 1 - 2 = -1$$

۵.۱.۴ امید ریاضی تابعی از دو متغیر تصادفی

دو متغیر تصادفی X و Y را با چگالی توأم $f(x, y)$ در نظر می‌گیریم. فرض کنید $h(x, y)$ یک تابع دو متغیری باشد، مانند $h(x, y) = x + y$ یا $h(x, y) = xy$. می‌خواهیم امید ریاضی متغیر تصادفی $Z = h(X, Y)$ را تعریف کنیم. برای این منظور از چگالی توأم استفاده می‌کنیم. در اینجا تنها حالت گسسته را شرح می‌دهیم.

۶.۱.۴ تعریف امید ریاضی $h(X, Y)$ در حالت گسسته

امید ریاضی $h(X, Y)$ در حالتی که X و Y هر دو گسسته باشند، عددی به صورت زیر است:

$$E[h(X, Y)] = \sum_x \sum_y h(x, y) f(x, y)$$

از تعریف بالا با فرض $h(x, y) = x$ و $h(x, y) = y$ می‌توانیم $E(X)$ و $E(Y)$ را از روی چگالی توأم پیدا کنیم زیرا:

$$\sum_x \sum_y x f(x, y) = \sum_x x \sum_y f(x, y) = \sum_x x f_X(x) = E(X)$$

$$\sum_x \sum_y y f(x, y) = \sum_y y \sum_x f(x, y) = \sum_y y f_Y(y) = E(Y)$$

در حالتی که X و Y هر دو پیوسته باشند، جمع دوگانه Σ به انتگرال دوگانه تبدیل می‌شود. این حالت را در پیوسته یک شرح داده‌ایم.

مثال ۶ برای جدول توزیع احتمال توأم مثال ۱۱، فصل سوم، مقدار $E(XY)$ را پیدا کنید.

حل - احتمال نظیر هر (x, y) را در xy ضرب و سپس آنها را جمع می‌کنیم. بنابراین داریم:

$$E(XY) = (0 \times \frac{1}{8}) + (0 \times \frac{1}{8}) + (1 \times \frac{2}{8}) + (2 \times \frac{2}{8}) + (4 \times \frac{1}{8}) + (6 \times \frac{1}{8}) = 2$$

مثال ۷ دو متغیر تصادفی X و Y دارای چگالی توام $f(x,y)$ هستند. ثابت کنید که

$$E(X+Y) = E(X) + E(Y)$$

حل - تابع $h(x,y) = x+y$ را در نظر می‌گیریم. بنابر تعریف داریم:

$$\begin{aligned} E(X+Y) &= \sum_x \sum_y (x+y)f(x,y) = \sum_x \sum_y xf(x,y) + \sum_x \sum_y yf(x,y) \\ &= E(X) + E(Y) \end{aligned}$$

این نتیجه در حالت پیوسته هم درست است.

مثال ۸ دو متغیر تصادفی X و Y مستقل و به ترتیب دارای چگالیهای $f_X(x)$ و $f_Y(y)$ می‌باشند. ثابت کنید که

$$E(XY) = E(X)E(Y)$$

حل - تابع $h(x,y) = xy$ را در نظر می‌گیریم. چون X و Y مستقل هستند، چگالی توام X و Y می‌شود:

$$f_{XY}(x,y) = f_X(x)f_Y(y)$$

اینک بنابر تعریف داریم:

$$\begin{aligned} E(XY) &= \sum_x \sum_y xyf(x,y) = \sum_x \sum_y xyf_X(x)f_Y(y) \\ &= \sum_x xf_X(x) \sum_y yf_Y(y) = E(X)E(Y) \end{aligned}$$

این نتیجه در حالت پیوسته هم درست است. توجه کنید که اگر X و Y مستقل نباشند، ممکن است تساوی بالا درست نباشد. مثلاً در مثال ۱۱، فصل سوم، داریم:

$$E(X) = \frac{3}{4} \quad E(Y) = 2 \quad E(XY) = 2$$

بنابراین معلوم می‌شود که $E(XY) \neq E(X)E(Y)$

۷.۱.۴ امید ریاضی یک ترکیب خطی

متغیرهای تصادفی X_1, X_2, \dots, X_n و مقادیر ثابت a_1, a_2, \dots, a_n را در نظر می‌گیریم. متغیر تصادفی

$$X = a_1X_1 + a_2X_2 + \dots + a_nX_n$$

را یک ترکیب خطی از X_i ها می‌نامیم.

اگر چگالی توام را برای چند متغیر تصادفی و امید ریاضی را برای تابعی از آنها تعمیم دهیم به آسانی ثابت می‌شود که

$$E(X) = a_1E(X_1) + a_2E(X_2) + \dots + a_nE(X_n)$$

مثال ۹ اگر $E(X_1) = 2$ و $E(X_2) = 3$ و $E(X_3) = -4$ ، امید ریاضی

$$X = 2 + 5X_1 - 2X_2 + 2X_3$$

را پیدا کنید.

حل - اگر عدد ۳ را به عنوان یک متغیر تصادفی ثابت تلقی کنیم، بنابر فرمول بالا داریم:

$$\begin{aligned} E(X) &= 2 + 5E(X_1) - 2E(X_2) + 2E(X_3) \\ &= 2 + 10 + (-6) + (-8) = -1 \end{aligned}$$

۸.۱.۴ امید ریاضی دو متغیر تصادفی هم‌توزیع

فرض کنید $X \stackrel{D}{=} Y$ ، یعنی X و Y هم‌توزیع باشند. بدیهی است که این دو متغیر تصادفی دارای چگالی مشترک و تکیه‌گاه مشترک هستند، به سخی دیگر داریم:

$$\begin{aligned} f_X(z) &= f_Y(z) \quad z \text{ برای هر عدد حقیقی} \\ S_X &= S_Y \quad \text{برای تکیه‌گاه‌ها داریم:} \end{aligned}$$

بنابراین X و Y دارای امید ریاضی مشترک هستند، زیرا در حالت گسسته

$$E(X) = \sum_x xf_X(x) \quad x \in S_X$$

$$E(Y) = \sum_y y f_Y(y) \quad y \in S_Y$$

بدیهی است که دو مجموع بالا با هم برابرند.

در حالت پیوسته Σ به \int تبدیل می شود و همان نتیجه به دست می آید. پس به طور فشرده برای دو متغیر تصادفی هم توزیع داریم:

$$X \stackrel{D}{=} Y \Rightarrow E(X) = E(Y)$$

افزون بر این، برای هر تابع دلخواه $h(t)$ ، می توان به آسانی نشان داد که

$$X \stackrel{D}{=} Y \Rightarrow E[h(X)] = E[h(Y)]$$

البته با فرض اینکه امید ریاضی وجود داشته باشد.

مثال ۱۰ سکه ای نازیب را دو بار پرتاب می کنیم. فرض کنید X و Y به ترتیب شماره شیرها و شماره خطها باشد. نشان دهید که

$$E(X^2) = E(Y^2), \quad E(X) = E(Y)$$

حل - همانطوری که در گفتار سوم داشتیم، X و Y هم توزیع هستند و جدول احتمال مشترک آنها چنین است:

| | | | |
|------------|---------------|---------------|---------------|
| Y یا X | ۰ | ۱ | ۲ |
| احتمال | $\frac{1}{4}$ | $\frac{1}{4}$ | $\frac{1}{4}$ |

بنابراین داریم:

$$E(X) = E(Y) = (0 \times \frac{1}{4}) + (1 \times \frac{1}{4}) + (2 \times \frac{1}{4}) = 1$$

$$E(X^2) = E(Y^2) = (0^2 \times \frac{1}{4}) + (1^2 \times \frac{1}{4}) + (2^2 \times \frac{1}{4}) = 2.5$$

در تساوی دوم تابع $h(t) = t^2$ را به کار برده ایم.

۹.۱.۴ تمرین بخش یک

۱. در یک بازی با پرتاب تصادفی تاس، به اندازه شماره خالها به شما سکه ۱۰۰ ریالی می دهند. بلیط هر بازی چند ریال باید باشد تا شخص بلیط فروش به اندازه $\frac{1}{4}$ سکه ها سود ببرد؟

۲. سکه نازیبی را سه بار پرتاب می کنند. فرض کنید X شماره شیرها و Y شماره خطها باشند. نشان دهید که

$$E(X) = E(Y)$$

$$E(X^2) = E(Y^2)$$

۳. برای سکه ای ارب با $P(H) = \frac{1}{4}$ ، آیا X و Y هم توزیع هستند؟ آیا تساویهای بالا درست است؟ نشان دهید که برای هر تابع $h(t)$ داریم:

$$X \stackrel{D}{=} Y \Rightarrow E[h(X)] = E[h(Y)]$$

۴. فرض کنید X و Y مستقل باشند و داشته باشیم $E(X) = 2$ و $E(Y) = 3$. امید ریاضی $(X-1)(Y-2)$ را بیابید.

۵. فرض کنید X_1, \dots, X_n دارای امید ریاضی مشترک μ باشند. امید ریاضی \bar{X} را بیابید.

۶. نشان دهید که تابع زیر، چگالی یک متغیر تصادفی گسسته X است.

$$f(x) = \begin{cases} \frac{1}{x(x+1)} & x = 1, 2, \dots \\ 0 & \text{جای دیگر} \end{cases}$$

آیا $E(X)$ وجود دارد؟

۷. نشان دهید که تابع زیر، چگالی یک متغیر تصادفی پیوسته X است.

$$f(x) = \begin{cases} \frac{1}{x^2} & x > 1 \\ 0 & \text{جای دیگر} \end{cases}$$

آیا $E(X)$ وجود دارد؟

۸. فرض کنید A نقطه ای به طول ۲ روی محور طولها و C نقطه ای به عرض ۲ روی محور عرضها باشد. نقطه ای به تصادف داخل مربع $OABC$ اختیار می کنیم. طول و عرض این نقطه را با X و Y

نشان می‌دهیم. نشان دهید X و Y هم‌توزیع هستند. $E(X)$ و $E(Y)$ و $E(X+Y)$ و $E(X-Y)$ را بیابید.

۹. میانگین چگالیهای $f_1(x)$ و $f_2(x)$ را با μ_1 و μ_2 نشان می‌دهیم. میانگین چگالی آمیخته $f(x) = pf_1(x) + qf_2(x)$ را بیابید، $0 < p < 1$ و $q = 1 - p$.

۱۰. ثابت کنید هرگاه متغیر تصادفی X غیر منفی باشد، داریم $E(X) \geq 0$.

۱۱. چگالی توام X و Y به صورت زیر است:

$$f(x, y) = \begin{cases} \frac{xy}{8} & (1, 1), (1, 2), (2, 1), (1, 3) \\ 0 & \text{جای دیگر} \end{cases}$$

$E(X)$ و $E(Y)$ و $E(XY)$ و $E(X+Y)$ را بیابید.

راهنمایی: برای محاسبه $E(XY)$ باید مجموع دوگانه به کار برد.

۱۲. چگالی توام X و Y به صورت زیر است:

$$f(x, y) = \begin{cases} x+y & 0 < x < 1, 0 < y < 1 \\ 0 & \text{جای دیگر} \end{cases}$$

نشان دهید X و Y هم‌توزیع هستند. $E(X)$ و $E(Y)$ و $E(XY)$ را بیابید.

راهنمایی: برای محاسبه $E(XY)$ باید انتگرال دوگانه به کار برد.

۱۳. یک تاس ناریب را دوبار می‌ریزیم. فرض کنید X_1 و X_2 به ترتیب شماره خالهای بار اول و دوم باشند و $X = X_1 + X_2$ و $Y = X_1 - X_2$. نشان دهید که $E(XY) = E(X)E(Y)$.

۱۴. در تمرین ۱۳، نشان دهید X و Y مستقل نیستند. بنابراین $E(XY) = E(X)E(Y)$ شرط لازم ولی نه کافی برای مستقل بودن X و Y است.

۱۵. فرض کنید X دارای چگالی زیر باشد:

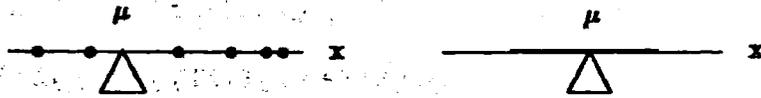
$$f_X(x) = \begin{cases} \sin x & 0 < x < \frac{\pi}{4} \\ 0 & \text{جای دیگر} \end{cases}$$

۲.۴ مفهوم واریانس و ویژگیهای آن

امید ریاضی یک متغیر تصادفی، با توجه به تعریف آن، به منزله معدل وزنی می‌باشد.

از اینرو $E(X)$ را، همانطوری که در پیش گفتیم، میانگین X یا میانگین $f_X(x)$ می‌نامند و آن را با μ و یا با μ_X نشان می‌دهند.

μ به عنوان مرکز توزیع X تلقی می‌شود و به آن پارامتر مکان می‌گویند. از نظر فیزیکی μ را بدین طریق تعبیر می‌کنیم: فرض کنید محور x ها در حکم میله‌ای همگن باشد که یک واحد جرم احتمال به طور گسسته یا پیوسته روی آن توزیع شده باشد. حال μ در سمت گراتیگاه این میله است.



دو متغیر تصادفی متفاوت ممکن است دارای میانگینهای مساوی باشند مثلاً فرض کنید X_1 اعداد $1, 0, 0, 1$ - را به ترتیب با احتمالهای $\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}$ و X_2 اعداد $5, 0, 0, 5$ - را به ترتیب با احتمالهای $\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}$ پذیرد. ملاحظه می‌شود که

$$E(X_1) = E(X_2) = 0$$



اینک می‌خواهیم معیاری بسازیم که به کمک آن بتوان درباره پراکنندگی جرم احتمال اطراف میانگین قضاوت نمود.

ظاهراً باید به امید ریاضی $X - \mu$ توجه کنیم. ولی چون همواره داریم:

$$E(X - \mu) = E(X) - \mu = 0$$

از این راه معیاری به دست نمی‌آید. امید ریاضی $|X - \mu|$ به عنوان معیار پراکنندگی معقول به نظر می‌آید، ولی به خاطر نماد قدر مطلق محاسبه آن همواره ساده نمی‌باشد. از اینرو بهتر است امید ریاضی $(X - \mu)^2$ را در نظر بگیریم.

۱.۲.۴ تعریف واریانس

میانگین $(X-\mu)^2$ را واریانس X می نامند و با $Var(X)$ یا σ_X^2 یا σ^2 نشان می دهند بنابراین

$$Var(X) = \sigma_X^2 = \sigma^2 = E((X-\mu)^2) \geq 0$$

در عمل برای محاسبه واریانس از فرمول مهم دیگری استفاده می شود که آن را به طریق زیر از فرمول بالا به دست می آوریم:

$$\begin{aligned} Var(X) &= E((X-\mu)^2) = E(X^2 + \mu^2 - 2\mu X) \\ &= E(X^2) + \mu^2 - 2\mu E(X) \end{aligned}$$

چون $E(X) = \mu$ ، پس فرمول محاسبه واریانس می شود:

$$Var(X) = E(X^2) - E^2(X)$$

از فرمول آخر نتیجه می گیریم که واریانس X برابر است با امید ریاضی X^2 منهای توان دوم امید ریاضی X .

مثال ۱۱ یک سکه نالر ب را سه بار می اندازیم فرض کنید X تعداد شیرها باشد. واریانس X را پیدا کنید:

حل - قبلاً دیدیم که $E(X) = \frac{3}{4}$ ، حال $E(X^2)$ را حساب می کنیم:

$$E(X^2) = (0 \times \frac{1}{8}) + (1^2 \times \frac{3}{8}) + (2^2 \times \frac{3}{8}) + (3^2 \times \frac{1}{8}) = 3$$

پس داریم:

$$Var(X) = E(X^2) - E^2(X) = 3 - (\frac{3}{4})^2 = \frac{3}{4}$$

مثال ۱۲ متغیر تصادفی X دارای چگالی زیر است:

$$f(x) = \begin{cases} 2x & 0 < x < 1 \\ 0 & \text{جای دیگر} \end{cases}$$

واریانس X را پیدا کنید.

حل -

$$E(X) = \int_{-\infty}^{\infty} xf(x) dx = \int_0^1 2x^2 dx = \frac{2}{3}$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^1 2x^3 dx = \frac{1}{2}$$

بنابراین داریم

$$Var(X) = E(X^2) - E^2(X) = \frac{1}{2} - (\frac{2}{3})^2 = \frac{1}{18}$$

۲.۲.۴ ویژگیهای واریانس

با استفاده از فرمول واریانس، ویژگیهای آن به دست می آیند.

الف - اگر b و c دو مقدار ثابت باشند، واریانس تابع خطی $Y = bX + c$ می شود

$$Var(Y) = Var(bX + c) = b^2 Var(X)$$

برای اثبات این ویژگی، از $\mu_Y = b\mu_X + c$ و تعریف $Var(Y)$ استفاده می کنیم.

با فرض $b = 1$ داریم:

$$Var(X+c) = Var(X)$$

به سخنی دیگر، هرگاه به X مقدار ثابت c اضافه شود، یعنی X به راست یا چپ انتقال یابد،

واریانس تغییر نمی کند، یعنی نحوه پراکندگی اطراف میانگین جدید تغییر نمی کند.

با فرض $b = 0$ داریم:

$$Var(c) = 0$$

یعنی واریانس مقدار ثابت صفر است زیرا مقدار ثابت نمی تواند پراکندگی داشته باشد. برعکس

می توان نشان داد که اگر متغیر تصادفی X دارای واریانس صفر باشد یعنی $Var(X) = 0$ ، آنگاه

$P(X = \mu_X) = 1$ ، یعنی X با احتمال یک برابر میانگین می شود. به سخنی دیگر دلری توزیع تابهده

۳.۲.۴ تعریف انحراف استاندارد

انحراف استاندارد X عبارت است از

$$\sigma_X = \sigma = \sqrt{\text{Var}(X)}$$

ملاحظه می شود که

$$\sigma_{bX}^2 = \text{Var}(bX) = b^2 \text{Var}(X)$$

$$\sigma_{bX} = |b| \sigma_X$$

در ضمن می دانیم که

$$\mu_{bX} = b\mu_X$$

بنابراین واریانس مانند سطح دارای بعد ۲ ولی انحراف استاندارد و میانگین مانند طول دارای بعد یک هستند. مثلاً اگر واحد اندازه گیری را از متر به دسیمتر تغییر دهیم واریانس صد برابر ولی میانگین و انحراف استاندارد ده برابر می شود.

۴.۲.۴ متغیر تصادفی استاندارد

اگر X دارای میانگین μ و واریانس σ^2 باشد، متغیر تصادفی زیر را

$$X^* = \frac{X - \mu}{\sigma}$$

دارای میانگین صفر و واریانس یک است. X^* را متغیر تصادفی استاندارد می گوئیم.

مثال ۱۳ وزن نوزادی متغیر تصادفی X با میانگین $3/21$ کیلوگرم و واریانس $0/64$ کیلوگرم به توان ۲ است. میانگین و انحراف استاندارد را برحسب پوند پیدا کنید.

حل - می دانیم که هر کیلوگرم تقریباً $2/2$ پوند است. پس وزن نوزاد برحسب پوند می شود متغیر تصادفی $Y = 2/2 X$ حال داریم:

$$\mu_Y = 2/2 \mu_X = (2/2)(3/21) = 7/502 \text{ پوند}$$

$$\sigma_Y^2 = (2/2)^2 \sigma_X^2 = (2/2)^2 (0/64) = 3/0.976 \text{ پوند به توان ۲}$$

$$\sigma_Y = \sqrt{3/0.976} = 1/76 \text{ پوند}$$

۵.۲.۴ تمرین بخش دو

۱. تاس ناریبی را به تصادف می ریزند. اگر خال زوج بیاید ده ریال دریافت می کنید و اگر خال فرد بیاید سه ریال می پردازید. فرض کنید X درآمد شما در این بازی باشد. میانگین و واریانس X را پیدا کنید.

۲. متغیر تصادفی X دارای چگالی زیر است:

$$f(x) = \begin{cases} \frac{1}{2} & -1 < x < 1 \\ 0 & \text{جای دیگر} \end{cases}$$

میانگین و واریانس و انحراف استاندارد X و X^2 و $|X|$ و $2X - 3$ و $(2X + 3)(X + 1)$ را پیدا کنید.

۳. اندازه قامت یک جوان بیست ساله متغیر تصادفی X با میانگین ۱۵۰ و انحراف استاندارد ۲۵ سانتیمتر است. میانگین و واریانس انحراف استاندارد X را برحسب اینج حساب کنید.

۴. متغیرهای X و Y مستقل با میانگینهای ۲ و ۳ و واریانسهای ۴ و ۵ می باشند. امید ریاضی XY و $(X + Y)(X - Y)$ را حساب کنید.

۵. ثابت کنید که اگر $E(X) = 0$ و $\text{Var}(X) = 0$ آنگاه $P(X = 0) = 1$.

۶. ثابت کنید که $\text{Var}(Y) = 0$ اگر و تنها اگر $P(Y = \mu_Y) = 1$.

راهنمایی: فرض کنید $X = Y - \mu_Y$ و از تمرین ۵ استفاده کنید.

۷. با استفاده از تعریف واریانس ثابت کنید که

$$\text{Var}(bX + c) = b^2 \text{Var}(X)$$

۸. ثابت کنید که متغیر تصادفی استاندارد $X^* = \frac{X - \mu}{\sigma}$ دارای میانگین صفر و واریانس یک است.

۹. در ظرفی N بلیط به شماره های $1, 2, \dots, N$ داریم. یکی از آنها را به تصادف بیرون می آوریم و شماره آن را X می نامیم. میانگین و واریانس X را برحسب N پیدا کنید.

۱۰. با فرض $Y > 0$ ثابت کنید که

$$E(\sqrt{Y}) \leq \sqrt{E(Y)}$$

$$E\left(\frac{1}{Y}\right) \geq \frac{1}{E(Y)}$$

چه موقع تساوی برقرار می شود؟

راهنمایی: برای اثبات نامساوی اول در فرمول محاسبه واریانس به جای X^2 بگذارید Y . برای اثبات نامساوی دوم از دو طرف نامساوی زیر امید ریاضی بگیرید.

$$(Y - \mu_Y) \left(\frac{1}{Y} - \frac{1}{\mu_Y} \right) = \frac{-(Y - \mu_Y)^2}{\mu_Y Y} \leq 0$$

۱۱. فرض کنید X یک متغیر تصادفی پیوسته غیر منفی باشد، یعنی $P(X \geq 0) = 1$. با استفاده از تساوی زیر

$$E(X) = \int_0^{\infty} xf(x) dx = \int_0^a xf(x) dx + \int_a^{\infty} xf(x) dx$$

ثابت کنید که برای هر عدد $a > 0$ داریم

$$P(X > a) \leq \frac{E(X)}{a} \quad (\text{نامساوی مارکف})$$

اگر X گسسته باشد، چه می کنید؟

۱۲. فرض کنید متغیر تصادفی Y دارای میانگین μ و واریانس σ^2 باشد. با استفاده از $X = (Y - \mu)^2$ و نامساوی تمرین ۱۱ ثابت کنید که برای $k > 0$ داریم

$$P(|Y - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad (\text{نامساوی چیبیچف})$$

اگر Y متغیر نرمال استاندارد باشد، طرف چپ نامساوی را برای $k = 1, 2, 3$ بیابید و با طرف راست مقایسه کنید.

۱۳. فرض کنید X دارای چگالی زیر باشد:

$$f(x) = \begin{cases} \frac{1}{2} & -1 < x < 1 \\ 0 & \text{جای دیگر} \end{cases}$$

کران بالایی $P(|X| > \frac{3}{4})$ را به کمک تمرین ۱۲ پیدا کنید.

۱۴. متغیر تصادفی Y اعداد -1 و 1 را با احتمالهای $\frac{1}{4}$ و $\frac{1}{4}$ می پذیرد. نشان دهید نامساوی چیبیچف به تساوی تبدیل می شود. بنابراین این نامساوی را نمی توان در حالت کلی تیزتر کرد. زیرا ممکن است تساوی هم برقرار شود.

۱۵. در تمرین ۱۳، $Var(X^2)$ را بیابید.

۳.۴ مفهوم کواریانس و ویژگیهای آن

برای اینکه کواریانس را تعریف کنیم نخست با استفاده از تعریف واریانس به محاسبه واریانس مجموع دو متغیر تصادفی X و Y با میانگینهای μ_X و μ_Y می پردازیم. چون امید $X + Y$ برابر است با $\mu_X + \mu_Y$ پس داریم

$$\begin{aligned} Var(X+Y) &= E\{[(X+Y) - (\mu_X + \mu_Y)]^2\} \\ &= E\{(X - \mu_X)^2 + (Y - \mu_Y)^2 + 2(X - \mu_X)(Y - \mu_Y)\} \\ &= Var(X) + Var(Y) + 2E\{(X - \mu_X)(Y - \mu_Y)\} \end{aligned}$$

آخرین امید ریاضی را با $Cov(X, Y)$ نشان می دهیم و آن را کواریانس X و Y می نامیم. بنابراین برای $Var(X+Y)$ فرمول زیر به نام فرمول واریانس مجموع به دست می آید:

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

۱.۳.۴ تعریف کواریانس

فرض کنید X و Y دو متغیر تصادفی با میانگینهای μ_X و μ_Y باشند کواریانس این دو متغیر را با $Cov(X, Y)$ یا σ_{XY} نشان می دهیم و به صورت زیر تعریف می کنیم:

$$Cov(X, Y) = E\{(X - \mu_X)(Y - \mu_Y)\}$$

حاصلضرب $(X - \mu_X)(Y - \mu_Y)$ ممکن است، مثبت یا منفی باشد. پس امید ریاضی آن، یعنی عدد $Cov(X, Y)$ بر حسب اینکه مثبت یا منفی باشد، بیان می دارد که X و Y بطور متوسط در یک جهت یا در دو جهت مخالف تغییر می کنند.

۲.۳.۴ محاسبه کواریانس

کواریانس را می توان با استفاده از تعریف بالا و داشتن جدول توزیع احتمال توام X و Y محاسبه کرد. در عمل از فرمول دیگری استفاده می شود که آن را به طریق زیر از روی تعریف کواریانس پیدا می کنیم:

$$\begin{aligned} Cov(X, Y) &= E\{(X - \mu_X)(Y - \mu_Y)\} \\ &= E\{XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y\} \\ &= E\{XY\} - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y \end{aligned}$$

پس فرمول محاسبه کواریانس می شود:

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

مثال ۱۴ فرض کنید جدول توزیع احتمال X و Y به صورت زیر باشد:

| | | | | |
|------------------|---------------|---------------|---------------|---------------|
| $X \backslash Y$ | ۱ | ۲ | ۳ | $f_Y(y)$ |
| ۱ | $\frac{1}{8}$ | $\frac{2}{8}$ | $\frac{3}{8}$ | $\frac{6}{8}$ |
| ۲ | $\frac{2}{8}$ | \cdot | \cdot | $\frac{2}{8}$ |
| $f_X(x)$ | $\frac{3}{8}$ | $\frac{2}{8}$ | $\frac{3}{8}$ | ۱ |

$Cov(X, Y)$ و $Var(X+Y)$ را حساب کنید.

حل - از جدول بالا داریم:

$$E(X) = 2, \quad E(X^2) = \frac{19}{4}, \quad E(Y) = \frac{5}{4}$$

$$E(Y^2) = \frac{5}{4}, \quad Var(X) = \frac{3}{4}, \quad Var(Y) = \frac{3}{16}$$

$$E(XY) = \frac{9}{4}$$

$$Cov(X, Y) = E(XY) - E(X)E(Y) = -\frac{1}{4}$$

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X, Y) = \frac{5}{16}$$

با استفاده از فرمول محاسبه کواریانس می توان به آسانی ویژگیهای زیر را ثابت کرد:

الف - $Cov(X, Y) = Cov(Y, X)$. این ویژگی نشان می دهد که جابجائی X و Y مقدار کواریانس را تغییر نمی دهد یعنی کواریانس نسبت به X و Y متقارن است.

ب - برای مقدار ثابت k داریم: $Cov(X, k) = 0$. این ویژگی نشان می دهد که هرگاه یکی از دو متغیر ثابت بماند کواریانس صفر می شود.

ج - برای هر دو مقدار ثابت b و d داریم:

$$Cov(X+b, Y+d) = Cov(X, Y)$$

این ویژگی نشان می دهد که هرگاه مبدأ اندازه گیری X و Y را تغییر دهیم کواریانس تغییر نمی کند.

د - برای دو متغیر تصادفی مستقل X و Y داریم: $Cov(X, Y) = 0$. این ویژگی به آسانی از $E(XY) = E(X)E(Y)$ و از فرمول محاسبه کواریانس نتیجه می شود و بیان می دارد که کواریانس دو متغیر تصادفی مستقل صفر است. زیرا تغییرات آنها مستقل از هم می باشد. ولی باید توجه کرد که صفر بودن کواریانس، مستلزم مستقل بودن X و Y نمی باشد (به مثال ۱۵ در زیر توجه کنید).

ه - برای هر دو مقدار ثابت a و c داریم:

$$Cov(aX, cY) = ac Cov(X, Y)$$

این ویژگی نشان می دهد که هرگاه واحد اندازه گیری را تغییر دهیم کواریانس تغییر می کند. مثلاً اگر واحد اندازه گیری X و Y را ده بار کوچک کنیم کواریانس صد برابر می شود. از ویژگیهای بالا معلوم می شود که با تغییر مبدأ و واحد اندازه گیری X و Y داریم:

$$Cov(aX+b, cX+d) = ac Cov(X, Y)$$

و - $Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$. این ویژگی نشان می دهد که کواریانس نسبت به X و Y جمع پذیر است.

ز - $Cov(X, X) = Var(X)$. این ویژگی نشان می دهد که واریانس را می توان بر حسب کواریانس بیان داشت.

۴.۳.۴ تمرین بخش سه

۱. جدول توزیع احتمال X و Y به صورت زیر است:

| | | | |
|------------------|----------------|---------------|----------------|
| $X \backslash Y$ | ۱ | ۲ | ۳ |
| ۱ | $\frac{1}{12}$ | $\frac{1}{6}$ | $\frac{1}{12}$ |
| ۲ | $\frac{1}{6}$ | ۰ | $\frac{1}{6}$ |
| ۳ | ۰ | $\frac{1}{3}$ | ۰ |

$E(X)$ و $E(Y)$ و $E(XY)$ و $Cov(X, Y)$

۲. چگالی توام X و Y به صورت زیر است:

$$f(x, y) = \begin{cases} kxy & 0 \leq x \leq 1 \\ 0 & \text{جای دیگر} \end{cases}$$



$E(X)$ و $E(Y)$ و $E(XY)$ و $Cov(X, Y)$ را بیابید.

۳. ویژگیهای کواریانس را با استفاده از فرمول محاسبه کواریانس ثابت کنید.

۴. برای $Cov(a_1X_1 + a_2X_2, b_1Y_1 + b_2Y_2)$ یک فرمول پیدا کنید و آن را تعمیم دهید.

۵. فرض کنید X و Y مستقل و هم توزیع با واریانس مشترک ۴ باشند. نشان دهید.

$$Cov(X+Y, X-Y)$$

آیا فرض مستقل و هم توزیع بودن X و Y ضرورت دارد؟

۶. در تمرین ۵،

$$Cov(X+2Y+3, X-2Y+3)$$

را بیابید.

مثال ۱۵ فرض کنید X اعداد ۱، ۰، -۱ را با احتمالهای $\frac{1}{4}$ ، $\frac{1}{4}$ ، $\frac{1}{4}$ بپذیرد. متغیر تصادفی $Y = X^2$ را در نظر می گیریم. نشان دهید که $Cov(X, Y) = 0$ ولی X و Y مستقل نیستند.

حس - ملاحظه می شود که $E(X) = 0$ ، $E(Y) = \frac{1}{4}$ و $E(XY) = 0$ پس داریم: $Cov(X, Y) = 0$. حال نشان می دهیم که X و Y مستقل نیستند. اگر X و Y بخواهند مستقل باشند باید داشته باشیم (بطور مثال):

$$P(X=1, Y=1) = P(X=1)P(Y=1)$$

ولی داریم:

$$P(X=1, Y=1) = P(X=1) = \frac{1}{4}$$

$$P(X=1)P(Y=1) = \left(\frac{1}{4}\right)\left(\frac{1}{4}\right) = \frac{1}{16}$$

مثال ۱۶ متغیرهای تصادفی X و Y دارای میانگینهای ۱ و ۲ و واریانسهای ۳ و ۴ و کواریانس -۱ می باشند. میانگینها و واریانسها و کواریانس $U = 2X - Y$ و $V = X + 2Y$ را پیدا کنید.

حل - به آسانی داریم:

$$E(U) = 2E(X) - E(Y) = 0$$

$$E(V) = E(X) + 2E(Y) = 5$$

$$Var(U) = Var(2X - Y) = 4Var(X) + Var(Y) - 2Cov(X, Y) = 20$$

$$Var(V) = Var(X + 2Y) = Var(X) + 4Var(Y) + 2Cov(X, Y) = 15$$

$$Cov(U, V) = -5 \quad (\text{چرا؟})$$

۷. برای $Var(a_1X_1 + a_2X_2)$ یک فرمول پیدا کنید و آن را تعمیم دهید.

۸. در تمرین ۵، $Var(2X - 3Y + 1)$ را بیابید.

می‌گیریم. برای حالت $\rho = 1$ متغیر تصادفی Z را به صورت زیر تعریف می‌کنیم:

$$Z = \frac{X - \mu_X}{\sigma_X} - \frac{Y - \mu_Y}{\sigma_Y}$$

بنابر فرمول واریانس مجموع و تعریف ضریب همبستگی داریم:

$$Var(Z) = 1 + 1 - 2Cov\left(\frac{X - \mu_X}{\sigma_X}, \frac{Y - \mu_Y}{\sigma_Y}\right) = 2 - 2\rho \geq 0$$

بنابراین $\rho \leq 1$ و $\rho = 1$ اگر و تنها اگر $Var(Z) = 0$ چون $E(Z) = 0$ پس $\rho = 1$ اگر و تنها اگر یا

احتمال یک داشته باشیم:

$$Z = \frac{X - \mu_X}{\sigma_X} - \frac{Y - \mu_Y}{\sigma_Y} = 0$$

به عبارت دیگر $\rho = 1$ اگر و تنها اگر میان یافته‌های X و Y رابطه خطی بالا که دارای ضریب زاویه مثبت است برقرار باشد،

به همین طریق U را به صورت زیر تعریف می‌کنیم:

$$U = \frac{X - \mu_X}{\sigma_X} + \frac{Y - \mu_Y}{\sigma_Y}$$

و نشان می‌دهیم که $\rho \geq -1$.

به طور فشرده داریم:

$$(1) \rho \text{ عددیست در فاصله بسته } [-1, 1]$$

(2) $\rho = 1$ اگر و تنها اگر یافته‌های (X, Y) روی خط زیر باشند:

$$\frac{X - \mu_X}{\sigma_X} - \frac{Y - \mu_Y}{\sigma_Y} = 0$$

(3) $\rho = -1$ اگر و تنها اگر یافته‌های (X, Y) روی خط زیر باشند:

$$\frac{X - \mu_X}{\sigma_X} + \frac{Y - \mu_Y}{\sigma_Y} = 0$$

(4) اگر ρ به 1 یا -1 نزدیک باشد، یافته‌های (X, Y) اطراف یک خط راست می‌باشند. به عبارت دیگر گرایش خطی میان X و Y زیاد است.

(5) اگر $\rho = 0$ ، X و Y نامبسته هستند. به عبارت دیگر گرایش خطی میان X و Y ناچیز است. بنابراین نامبستگی مستلزم مستقل بودن X و Y نمی‌باشد.

نگاره ۱، ویژگیهای بالا را نشان می‌دهد.

۲.۴ مفهوم ضریب همبستگی و ویژگیهای آن

ویژگیها الف تا د برای $Cov(X, Y)$ ، به عنوان معیاری برای سنجش تغییرات X و Y نسبت به هم، ویژگیهای خوبی هستند اما ویژگیها مطلوب نیست زیرا مقادیر کورلیانس را به واحد اندازه گیری متکی می‌کند. اگر X و Y را به ترتیب با σ_X و σ_Y بسنجیم یعنی به جای آنها $\frac{X}{\sigma_X}$ و $\frac{Y}{\sigma_Y}$ را به کار ببریم. ملاحظه می‌شود که دیگر این دو متغیر جدید در نتیجه $Cov\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right)$ به واحد اندازه گیری متکی نخواهد شد. این انتخاب برای تعریف ضریب همبستگی است.

۱.۴.۴ تعریف ضریب همبستگی

دو متغیر تصادفی X و Y با واریانسهای σ_X^2 و σ_Y^2 را در نظر می‌گیریم ضریب همبستگی این دو متغیر را با $\rho(X, Y)$ یا ρ نشان می‌دهیم و به صورت زیر تعریف می‌کنیم:

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

۲.۴.۴ ویژگیهای ضریب همبستگی

الف - برای مقادیر ثابت a, b, c, d با فرض $a > 0$ و $c > 0$ داریم:

$$\rho(aX + b, cY + d) = \rho(X, Y)$$

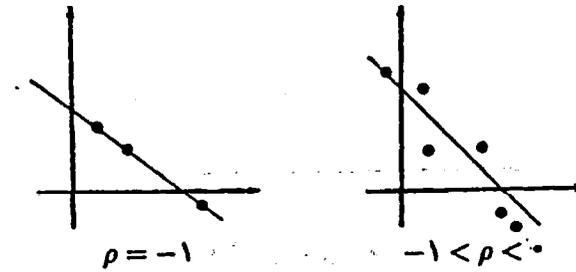
این تساوی به آسانی با استفاده از ویژگیهای واریانس و کورلیانس ثابت می‌شود. معنی این تساوی این است که ضریب همبستگی X و Y به مبدأ و واحد اندازه گیری بستگی ندارد.

ب - $-1 \leq \rho \leq 1$

$$\rho = \pm 1 \Leftrightarrow \frac{X - \mu_X}{\sigma_X} = \pm \frac{Y - \mu_Y}{\sigma_Y} \quad \text{ج}$$

برای اثبات دو ویژگی ب و ج به طریق زیر عمل می‌کنیم:

دو متغیر تصادفی استاندارد $\frac{X - \mu_X}{\sigma_X}$ و $\frac{Y - \mu_Y}{\sigma_Y}$ را که دارای میانگین صفر و واریانس یک هستند در نظر



نگاره ۱ ضریب همبستگی و ارتباط خطی میان X و Y

مثال ۱۷ در مثال ۱۴ ضریب همبستگی X و Y را پیدا کنید.

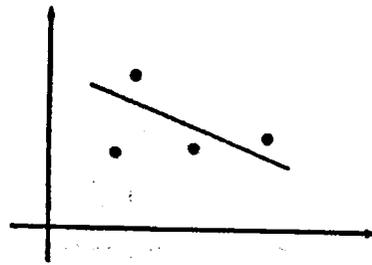
حل - در مثال ۱۴ دیدیم که

$$\sigma_X^2 = \frac{3}{4} \quad \sigma_Y^2 = \frac{3}{16} \quad \sigma_{XY} = -\frac{1}{4}$$

پس ضریب همبستگی برابر است با

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{-\frac{1}{4}}{\sqrt{\frac{3}{4}} \sqrt{\frac{3}{16}}} = \frac{-\frac{1}{4}}{\frac{3}{8}} = -\frac{2}{3}$$

یافته‌های متغیر تصادفی دو بعدی (X, Y) عبارتند از $(1, 1)$ و $(2, 1)$ و $(3, 1)$ و $(1, 2)$. نمایش هندسی این یافته‌ها چهار نقطه در صفحه محورها می‌باشد. این نقاط اطراف خطی با ضریب زاویه منفی در نگاره ۲ دیده می‌شوند.



نگاره ۲ پراکندگی نقاط اطراف یک خط

مثال ۱۸ فرض کنید میان X و Y رابطه زیر برقرار باشد:

$$(2X-1)(Y+1) = 2XY+3$$

ضریب همبستگی را پیدا کنید.

حل - رابطه بالا بعد از ساده کردن به صورت رابطه خطی زیر درمی‌آید:

$$2X - Y - 4 = 0$$

این خط دارای ضریب زاویه مثبت است. پس $\rho = 1$.

مثال ۱۹ فرض کنید که X دارای چگالی زیر باشد:

$$f(x) = \begin{cases} \frac{1}{2} & -1 < x < 1 \\ 0 & \text{جای دیگر} \end{cases}$$

با فرض $Y = X^2$ ، $\rho(X, Y)$ را پیدا کنید.

حل - به آسانی داریم

$$E(X) = 0$$

$$E(Y) = E(X^2) = \frac{1}{3}$$

$$E(XY) = E(X^3) = 0$$

$$Cov(X, Y) = E(XY) - E(X)E(Y) = 0$$

بنابراین $\rho(X, Y) = 0$. چون یافته‌های (X, Y) روی سهمی $Y = X^2$ جا دارند، صرفی بودن ضریب همبستگی دور از انتظار نمی‌باشد. در ضمن می‌توان به عنوان تمرین نشان داد که X و Y مستقل نیستند.

مثال ۲۰ فرض کنید $X+Y=1$ و $U=2X-Y$ و $V=X+2Y$. ضریب همبستگی U و V را بیابید.

حل - از حل دستگاه دو مجهولی

$$\begin{cases} 2X - Y = U \\ X + 2Y = V \end{cases}$$

یا استفاده از $X+Y=1$ داریم

$$U + 2V = 5$$

بنابراین $\rho(U, V) = -1$.

۳.۴.۴ تمرین بخش چهارم

۱. متغیرهای X و Y دارای جدول توزیع احتمال زیر می‌باشند:

| | | | |
|------------------|---------------|---------------|---------------|
| $X \backslash Y$ | ۱ | ۲ | ۳ |
| -۱ | $\frac{1}{8}$ | $\frac{2}{8}$ | $\frac{3}{8}$ |
| ۰ | ۰ | $\frac{1}{8}$ | $\frac{1}{8}$ |

$Cov(X, Y)$ و $\rho(X, Y)$ و $\rho(X-Y, Y)$ را پیدا کنید.

۲. متغیر تصادفی X دارای چگالی زیر است:

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{جای دیگر} \end{cases}$$

فرض کنید $Y = X^2$. $\rho(X, Y)$ و $\rho(X+1, 2Y-1)$ را حساب کنید.
۳. متغیرهای X و Y دارای جدول توزیع احتمال زیر می‌باشند:

| | | | |
|------------------|----------------|----------------|----------------|
| $X \backslash Y$ | -۱ | ۰ | ۱ |
| ۰ | $\frac{1}{16}$ | $\frac{2}{16}$ | $\frac{1}{16}$ |
| ۱ | $\frac{4}{16}$ | $\frac{4}{16}$ | $\frac{4}{16}$ |

نشان دهید $\rho(X, Y) = 0$ ولی X و Y نامستقل اند.

۴. متغیر تصادفی X اعداد ۲، ۰، -۲ را با احتمالات $\frac{1}{4}$ ، $\frac{1}{2}$ ، $\frac{1}{4}$ می‌پذیرد. فرض کنید $Y = X^2 + 1$ نشان دهید که X و Y نامستقل اند ولی

$$Var(X+Y) = Var(X) + Var(Y)$$

۵. در ظرفی ۳ مهره سفید و ۲ مهره سیاه و یک مهره قرمز داریم. از این ظرف چشم بسته چهار مهره با هم بیرون می‌آوریم. فرض کنید X تعداد مهره‌های سفید و Y تعداد مهره‌های سیاه در این نمونه چهارتایی باشد. جدول توزیع احتمال توام X و Y را پیدا کنید. $\rho(X, Y)$ را پیدا کنید.

۶. دو متغیر تصادفی X و Y دارای واریانس مساوی می‌باشند. ثابت کنید که مجموع و تفاضل آنها ناهمبسته هستند.

۷. فرض کنید X_1, X_2, \dots, X_n هم‌توزیع و مستقل با میانگین مشترک μ و واریانس مشترک σ^2 باشد. میانگین و واریانس \bar{X} را پیدا کنید.

۸. در تمرین ۷ کواریانس X و $X_1 - X$ را پیدا کنید.

۹. در ظرفی ۳ مهره سفید و ۴ مهره سیاه داریم. از این ظرف ۵ مهره با هم بیرون می آوریم. فرض کنید X تعداد مهره های سفید و Y تعداد مهره های سیاه در این نمونه ۵ تایی باشد. ضریب همبستگی X و Y را پیدا کنید.

۱۰. X و Y دارای میانگینهای ۱ و ۲ و واریانسهای ۳ و ۴ و کواریانس ۱- می باشند. میانگین و واریانس $Z = 2X - Y + 1$ را پیدا کنید. m را طوری تعیین کنید تا $U = mX + Y$ و $V = X - mY$ ناهمبسته باشند.

۱۱. با استفاده از ویژگیهای کواریانس ثابت کنید که $Var(bX + c) = b^2 Var(X)$

۱۲. ثابت کنید که: $\rho(aX + b, cY + d) = \pm \rho(X, Y)$

۱۳. در مثال ۱۹ نشان دهید که X و Y مستقل نیستند.

۱۴. ثابت کنید $\rho \geq -1$. چه موقع داریم $\rho = -1$ ؟

۱۵. فرض کنید میان X و Y رابطه خطی $1 = X + 2Y$ برقرار باشد و داشته باشیم $Var(X) = 4$. $Cov(X, Y)$ و $\rho(X, Y)$ را بیابید.

فصل ۵

چند توزیع مهم و ارتباط آنها با هم

در این گفتار نخست توزیمهای دو جمله ای، بواسن و نرمال را که در سرتاسر آمار و احتمال و در عمل دارای اهمیت هستند مطالعه می کنیم. سپس به کمک قضیه ای معروف و تاریخی به نام قضیه حد مرکزی درباره ارتباط این سه توزیع سخن می گوئیم.

۱.۵ توزیع دو جمله ای

پیش از اینکه منظور از توزیع دو جمله ای را بیان داریم، به تعریف آزمایش برنولی و متغیر تصادفی برنولی می پردازیم.

۱.۱.۵ تعریف آزمایش برنولی

آزمایش برنولی یک آزمایش تصادفی است که فضای نمونه آن تنها از دو پیشامد F و S ، به نام شکست و پیروزی، تشکیل شده باشد.

مثلاً بازی شیر و خط، آزمایشهای طبی که نتیجه مثبت یا منفی دارند، و پرسشهایی که پاسخ آنها بله یا نه می باشند، همگی از نوع آزمایشهای برنولی هستند. این آزمایشها به نام یکی از برادران برنولی است. آنها از ریاضیدانان معروف سوئیس در قرن هجدهم بودند.

۲.۱.۵ تعریف متغیر تصادفی برنولی

Y را روی فضای نمونه آزمایش برنولی، بدین طریق تعریف می کنیم:

$$Y(F) = 0 \quad Y(S) = 1$$

مثلاً در آزمایش با سکه اگر خط مشاهده کنیم Y را صفر و اگر شیر مشاهده کنیم Y را یک می‌گیریم. فرض کنید:

$$P(Y=0)=q; P(Y=1)=p \quad 0 < p < 1, p+q=1$$

متغیر تصادفی Y را یک متغیر تصادفی برنولی و توزیع آن را توزیع برنولی می‌نامند. این توزیع گسسته دارای چگالی زیر است:

$$f(y;p) = \begin{cases} p^y (1-p)^{1-y} & y=0,1 \\ \text{جای دیگر} & \end{cases}$$

چون با تغییر p چگالی بالا تغییر می‌کند، از اینرو p را یک پارامتر برای چگالی بالا می‌نامیم. پارامتر p احتمال پیروزی است. به آسانی داریم

$$E(Y) = p \quad \text{Var}(Y) = pq$$

۳.۱.۵ تعریف توزیع دو جمله‌ای

یک آزمایش برنولی با پارامتر p را n بار مستقلاً انجام می‌دهیم. مثلاً یک سکه، با $P(H) = p$ ، را n بار مستقلاً می‌اندازیم. فرض کنید متغیر تصادفی X تعداد پیروزیها، مثلاً تعداد شیرها، باشد، می‌گوئیم X دارای توزیع دو جمله‌ای است. چون هر آزمایش برنولی از دو پیشامد ساده F و S تشکیل می‌شود، پس تعداد پیشامدهای ساده در این n آزمایش برنولی برابر است با 2^n که هر یک به صورت رشته‌ای n تایی از F و S می‌باشد. مثلاً برای $n=3$ هشت پیشامد زیر را داریم:

$$FFF, SFF, FSF, FFS, SSF, FSS, SFS, SSS$$

پیشامد $(X=2)$ ، یعنی پیشامد اینکه تعداد پیروزیها دو باشد، عبارتست از

$$\{SSF, FSS, SFS\}$$

بطور کلی برای n آزمایش برنولی پیشامد $(X=x)$ از $\binom{n}{x}$ پیشامد ساده تشکیل می‌شود که هر یک به صورت رشته‌ای مرکب از x تا S و $n-x$ تا F است. چون آزمایشهای برنولی مستقل هستند، احتمال هر یک از این رشته‌ها می‌شود $p^x q^{n-x}$ و احتمال پیشامد $(X=x)$ می‌شود:

$$P(X=x) = \binom{n}{x} p^x q^{n-x} \quad x=0,1,2,\dots,n$$

بنابر دو جمله‌ای نیوتن داریم:

$$\sum_{x=0}^n P(X=x) = \sum_{x=0}^n \binom{n}{x} p^x q^{n-x} = (p+q)^n = 1$$

به سخنی دیگر، احتمالی که به $(X=x)$ نسبت می‌دهیم درست برابر است با یکی از جمله‌های بسط دو جمله‌ای $(p+q)^n$. از اینجاست که چگالی زیر به نام چگالی دو جمله‌ای به دست می‌آید:

۴.۱.۵ چگالی دو جمله‌ای

چگالی زیر را چگالی دو جمله‌ای با پارامترهای n و p می‌گوئیم:

$$f(x;n,p) = \begin{cases} \binom{n}{x} p^x q^{n-x} & x=0,1,2,\dots,n \\ \text{جای دیگر} & \end{cases}$$

هرگاه متغیر تصادفی X دارای این چگالی باشد، می‌گوئیم X دارای توزیع دو جمله‌ای با پارامترهای n و p است و می‌نویسیم $X \sim B(n,p)$. ملاحظه می‌شود که X یک متغیر تصادفی گسسته با تکیه‌گاه $SX = \{0,1,2,\dots,n\}$ می‌باشد.

تابع توزیع X را به صورت زیر نشان می‌دهیم:

$$B(x;n,p) = \sum_{k=0}^n \binom{n}{k} p^k q^{n-k}$$

توجه کنید که متغیر تصادفی برنولی Y هم حالتی ویژه از متغیر دو جمله‌ای با پارامترهای 1 و p است، یعنی $Y \sim B(1,p)$.

۵.۱.۵ میانگین و واریانس توزیع دو جمله‌ای

با استفاده از چگالی X می‌توان میانگین و واریانس آن را محاسبه کرد، ولی روش زیر آسانتر است. n آزمایش برنولی مستقل انجام می‌دهیم تا متغیر دو جمله‌ای X را بسازیم. فرض کنید Y_i متغیر تصادفی برنولی نظیر آزمایش i ام باشد. در نتیجه می‌نویسیم

$$X = Y_1 + Y_2 + \dots + Y_n$$

از طرفی می‌دانیم که برای متغیر تصادفی برنولی Y_i داریم:

$$E(Y_i) = p \quad \text{Var}(Y_i) = pq \quad i=1,2,\dots,n$$

با استفاده از ویژگیهای امیدریاضی و واریانس داریم:

$$E(X) = E(Y_1) + E(Y_2) + \dots + E(Y_n) = np$$

$$Var(X) = Var(Y_1) + Var(Y_2) + \dots + Var(Y_n) = npq$$

مثلاً برای $X \sim B(22, \frac{1}{4})$ داریم $E(X) = 8$ و $Var(X) = 6$.

می باشد. بارهایی که در روز با چراغ قرمز برخورد می کند متغیر تصادفی دو جمله ای $X \sim B(10, 0/6)$ است. با استفاده از جدول I داریم:

$$P(X=3) = P(X \leq 3) - P(X \leq 2) = 0/0425$$

مثال ۳ یک آزمون ریاضی، ساخته از ۱۵ تست چهار پرشی است که در هر تست تنها یک پرش درست است. شخصی بدون آمادگی لازم در این امتحان شرکت می کند و تستها را شانسی پاسخ می دهد. احتمال اینکه دست کم ۵ تست را درست پاسخ دهد چقدر است؟

حل - پاسخ دادن درست به هر تست، معادل با یک آزمایش برنولی با پارامتر $p = \frac{1}{4}$ است. بنابراین پاسخهای درست متغیر تصادفی $X \sim B(15, 0/25)$ می باشد. می خواهیم $P(X \geq 5)$ را حساب کنیم. با استفاده از جدول I داریم:

$$P(X \geq 5) = 1 - P(X \leq 4) = 0/3125$$

۶.۱.۵ محاسبه احتمال از روی توزیع دو جمله ای

محاسبه احتمال برای $1 \leq n \leq 25$ ، از روی چگالی آسان است، ولی برای $6 \leq n \leq 25$ بهتر است از جدولهایی که برای این منظور تهیه شده است استفاده کرد. جدول شماره I نمونه ای از جدول توزیع دو جمله ای می باشد که ما در این کتاب به کار می بریم. برای $n > 25$ معمولاً یک نوع روشی تقریبی به کار می رود که به زودی درباره آن صحبت خواهیم کرد. اینک به ذکر چند مثال می پردازیم:

مثال ۱ یک سکه نازیب را پنج بار مستقلاً می اندازیم. احتمال اینکه دو شیر و سه خط مشاهده شود چقدر است؟

حل - فرض کنید X تعداد شیرها باشد. واضح است که $X \sim B(5, \frac{1}{2})$ می خواهیم $P(X=2)$ را پیدا کنیم. با استفاده از چگالی X داریم:

$$P(X=2) = f(2; 5, \frac{1}{2}) = \binom{5}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{5-2} \\ = \binom{5}{2} \left(\frac{1}{2}\right)^5 = \frac{5}{16}$$

مثال ۲ در یک چهارراه چراغ سبز ۱۵ ثانیه و چراغ زرد ۵ ثانیه و چراغ قرمز ۳۰ ثانیه دوام دارد. راننده ای ده بار در روز بین ساعت ۸ تا ۹ از این چهارراه عبور می کند. احتمال اینکه درست سه بار در روز به چراغ قرمز برخورد کند چقدر است؟

حل - هر بار که این راننده از چهارراه عبور می کند، برخورد با چراغ قرمز معادل با پیروزی در یک آزمایش برنولی با پارامتر

$$p = \frac{30}{15+5+30} = 0/6$$

۷.۱.۵ تمرین بخش یک

۱. در ظرفی سه مهره سفید و دو مهره سیاه داریم که همگی گذشته از رنگ یکسان هستند. از این ظرف چهار مهره یک به یک با جایگذاری بیرون می آوریم. فرض کنید X تعداد مهره های سفید در این نمونه چهارتایی باشد. $E(X)$ و $Var(X)$ و $P(X=2)$ را محاسبه کنید. احتمال اینکه این چهار مهره یک در میان دارای رنگهای گوناگون باشند چیست؟

۲. یک آزمون زبان انگلیسی ساخته از ۲۰ تست پنج پرشی است که در هر تست تنها یک پرش درست می باشد. شخصی که اصلاً انگلیسی نمی داند در این آزمون شرکت می کند و تستها را به تصادف پاسخ می دهد. انتظار دارید چند تست را درست پاسخ دهد؟ احتمال اینکه ده تست را درست پاسخ دهد چیست؟

۳. با استفاده از چگالی دو جمله ای و تعریف امید ریاضی، $E(X)$ و $E(X(X-1))$ و سپس $E(X^2)$ و $Var(X)$ را برای متغیر تصادفی دو جمله ای X پیدا کنید.

۴. با فرض $X \sim B(n, p)$ و $U = n - X$ ، توزیع U را پیدا کنید.

۵. چهل درصد لامپهایی که در یک کارخانه تولید می شود معیوب هستند. احتمال اینکه در یک دوچین اصلاً لامپ معیوب نباشد چیست؟

۶. دو تاس ناریب را پنج بار با هم می‌ریزیم. احتمال اینکه دوبار جفت شش مشاهده شود چیست؟
 ۷. فرض کنید $X_1 \sim B(n_1, p)$ و $X_2 \sim B(n_2, p)$. اگر این دو متغیر مستقل باشند و $X = X_1 + X_2$ ثابت کنید که $X \sim B(n_1 + n_2, p)$.

۸. سکه‌ای را، با احتمال p برای شیر، آتقدر می‌ریزیم تا برای اولین بار شیر مشاهده شود. فرض کنید متغیر تصادفی X تعداد آزمایشهای لازم باشد چگالی X را پیدا کنید.

۹. سکه ناریب را، آتقدر می‌ریزیم تا دوبار شیر مشاهده شود. فرض کنید X تعداد آزمایشهای لازم باشد. $P(X=5)$ را حساب کنید.

۱۰. تمرین یک را بدون جایگذاری حل کنید.

۱۱. یک آزمایش برنولی، با احتمال p برای پیروزی، را n بار مستقلاً انجام می‌دهیم. فرض کنید X و Y به ترتیب تعداد پیروزیها و شکستها باشند. ثابت کنید احتمال حداقل m پیروزی برابر با احتمال حداکثر $n-m$ شکست است.

۱۲. سه سکه را که شانسی شیر آمدن برای آنها $\frac{1}{4}$ و $\frac{1}{3}$ و $\frac{1}{2}$ می‌باشد با هم پرتاب می‌کنیم. احتمال این که هر سه شیر بیایند چیست؟ امید ریاضی و واریانس تعداد شیرها را پیدا کنید.

۱۳. در کارخانه‌ای ۵ درصد از لامپهای تولید شده معیوب هستند. برای کنترل، هر روز ۱۰ لامپ را به تصادف انتخاب کرده آزمایش می‌کنند. انتظار می‌رود چند روز از سال بیش از ۳ لامپ معیوب مشاهده شوند؟

۱۴. فرض کنید $X \sim B(3, \frac{1}{4})$ و $Y = (X+1)^2$. $P(Y=4)$ و $E(Y)$ را حساب کنید.

۱۵. در n آزمایش برنولی مستقل، فرض کنید X شماره پیروزیها و Y شماره شکستها باشد. $Cov(X, Y)$ و $\rho(X, Y)$ را بیابید.

۲.۵ توزیع پواسن

یک توزیع گسسته که با توزیع دو جمله‌ای ارتباط نزدیک دارد، توزیع پواسن می‌باشد. برای اینکه توزیع پواسن را معرفی کنیم و این ارتباط را نشان دهیم، توزیعی دو جمله‌ای $B(n, p_n)$ را که در آن p_n (یعنی شانسی پیروزی در یک آزمایش برنولی) به n (یعنی تعداد آزمایشهای برنولی مستقل) بستگی دارد در نظر می‌گیریم. فرض کنید با بزرگشدن n احتمال p_n کوچک شود به طوری که np_n به سمت عدد $\lambda > 0$

میل نماید. می‌توان ثابت کرد که

$$\lim_{n \rightarrow \infty} \binom{n}{x} p_n^x q_n^{n-x} = e^{-\lambda} \frac{\lambda^x}{x!}$$

حال با استفاده از سری زیر

$$\sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{\lambda}$$

نشان می‌دهیم که یک چگالی روی $\{0, 1, 2, \dots\}$ می‌توان ساخت. ملاحظه می‌شود که

$$\sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$$

۱.۲.۵ چگالی پواسن

چگالی زیر را چگالی پواسن با پارامتر $\lambda > 0$ می‌گوئیم:

$$f(x; \lambda) = \begin{cases} e^{-\lambda} \frac{\lambda^x}{x!} & x = 0, 1, 2, \dots \\ 0 & \text{جای دیگر} \end{cases}$$

هرگاه متغیر تصادفی X دارای این چگالی باشد، می‌گوئیم X دارای توزیع پواسن با پارامتر λ می‌باشد و می‌نویسیم $X \sim P(\lambda)$. مشاهده می‌شود که X یک متغیر تصادفی گسسته با تکیه‌گاه $S_X = \{0, 1, 2, \dots\}$ است.

تابع توزیع X را به صورت زیر نشان می‌دهیم:

$$P(x; \lambda) = \sum_{k=0}^x e^{-\lambda} \frac{\lambda^k}{k!}$$

این توزیع به نام پواسن (۱۷۸۱-۱۸۴۰) ریاضیدان فرانسوی است.

۲.۲.۵ میانگین و واریانس توزیع پواسن

میانگین را به طریق زیر پیدا می‌کنیم:

$$E(X) = \sum_{x=0}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

$$= \lambda e^{-\lambda} \sum_{y=0}^{\infty} \frac{\lambda^y}{y!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

حل - فرض کنید X تعداد تلفن‌ها در عرض ۹ دقیقه باشد. با یک تناسب می‌توان دید که X به طور متوسط برابر است با

$$\frac{40 \times 9}{60} = 6$$

این عدد برابر با میانگین X یعنی λ است، بنابراین امید ریاضی X می‌شود

$$\lambda = E(X) = 6$$

فرض کنید این فاصله ۹ دقیقه‌ای را به فواصل زمانی بسیار کوچک T طوری بتوان تقسیم کرد تا شرایط زیر برقرار باشند:

الف - احتمال داشتن تلفن در هر T تقریباً متناسب با T باشد (شرط ایستائی)

ب - در هر T حداکثر یک تلفن بشود (شرط کمیابی)

ج - پیشامد تلفن داشتن در دو فاصله T مستقل باشند (شرط استقلال)

حال تصور کنید که تلفن شدن در فاصله زمانی T ، به منزله یک آزمایش برنولی با پیروزی باشد. طبق شرایط بالا تعداد زیاد آزمایش برنولی مستقل با p کوچک و میانگین $\lambda = 6$ انجام می‌گیرد. پس می‌توان گفت که X دارای توزیع پواسن با پارامتر $\lambda = 6$ است. با استفاده از جدول شماره II داریم:

$$P(X \geq 5) = 1 - P(X \leq 4) = 1 - \sum_{x=0}^4 \frac{e^{-6} 6^x}{x!} = 1 - 0.285 = 0.715$$

۳.۲.۵ تمرین بخش دو

۱. متغیر تصادفی X دارای توزیع پواسن با واریانس ۴ می‌باشد، $p(X > 2)$ را حساب کنید

۲. متغیر تصادفی X دارای توزیع پواسن می‌باشد و داریم:

$$3P(X=2) = P(X=1)$$

$P(X^2 - 4X = 0)$ را پیدا کنید.

۳. فرض کنید $X \sim P(\lambda)$. ثابت کنید که $E(X(X-1)) = \lambda^2$.

برای محاسبه واریانس نخست به طریق بالا ثابت می‌کنیم که

$$E(X(X-1)) = \lambda^2$$

حال با استفاده از این رابطه و $E(X) = \lambda$ داریم $E(X^2) = \lambda^2 + \lambda$ و در نتیجه

$$Var(X) = E(X^2) - E^2(X) = \lambda$$

تاکید می‌کنیم که میانگین و واریانس هر دو برابر با پارامتر λ می‌باشند.

رابطه حدی بالا به خوبی ارتباط میان توزیع دو جمله‌ای و توزیع پواسن را نشان می‌دهد. عملاً هرگاه n آزمایش برنولی مستقل داشته باشیم به طوری که n به قدر کافی بزرگ و p کوچک و np دارای مقداری معقول باشد، می‌توانیم توزیع دو جمله‌ای را با توزیع پواسن تقریب کنیم.

مثال ۴ - یک صدم ساعتهائی که در یک کارخانه تولید می‌شوند معیوب هستند. احتمال اینکه در یک بسته ۲۰۰ تائی که به تصادف پر شده است، اصلاً ساعت معیوب نباشد چقدر است؟

حل - هر ساعت که در کارخانه تولید می‌شود به منزله رخ دادن یک آزمایش برنولی با $p = 0.01$ (شانس معیوب بودن ساعت) می‌باشد. $n = 200$ آزمایش برنولی مستقل داریم. فرض کنید X تعداد ساعتهای معیوب در این بسته یا تعداد پیروزیها در این آزمایشها باشد. می‌دانیم که $X \sim B(200, 0.01)$ پس داریم:

$$P(X=0) = \binom{200}{0} (0.01)^0 (0.99)^{200} = 0.1334$$

از طرفی چون $n = 200$ بزرگ و $p = 0.01$ کوچک می‌باشد و $np = 2$ تقریباً دارای توزیع پواسن با پارامتر $\lambda = 2$ می‌باشد و داریم

$$P(X=0) \approx e^{-2} \frac{2^0}{0!} = 0.1353$$

ملاحظه می‌شود که جواب تقریبی تا دو رقم اعشار با جواب واقعی برابر است.

مثال ۵ - به یک مرکز تلفن بطور متوسط ۲۰ تلفن در فاصله ۱۰-۹ صبح می‌شود. احتمال اینکه در ۹ دقیقه حداقل ۵ تلفن به این مرکز بشود چیست؟

۴. در یک متر پارچه بطور متوسط ۲ زدگی می‌باشد. احتمال اینکه در یک قواره ۳ متری اصلاً زدگی نباشد چیست؟

۵. در یک کتاب ۴۰۰ صفحه‌ای ۴۰ غلط چاپی یافت می‌شود. احتمال اینکه در ده صفحه اول بیش از دو غلط باشد چیست؟

۶. در ظرفی ۹۹ مهره سفید و یک مهره سیاه داریم. از این ظرف ۲۰۰ بار با جایگذاری مهره‌های بیرون می‌آوریم. احتمال اینکه ۷ بار مهره سیاه دیده شود چیست؟

۷. تعداد تصادفات اتومبیل در یک چهارراه، ماهانه بطور متوسط ۴ می‌باشد. احتمال اینکه در عرض یک هفته تصادف رخ ندهد چیست؟

۸. احتمال هدف‌گیری در یک تیراندازی ۰/۰۰۰۲ می‌باشد. احتمال اینکه در ۴۰۰۰ تیراندازی، بیش از دو هدف‌گیری شود چیست؟

۹. اگر $f(x)$ چگالی توزیع پواسن باشد، ثابت کنید

$$f(x+1) = \frac{\lambda}{x+1} f(x)$$

با استفاده از این برابری ساده، با فرض $\lambda = 1, 2, 3$ برای $f(x)$ از $x=0$ تا $x=10$ یک جدول بسازید.

۱۰. اگر $X \sim B(n, p)$ و $f(x)$ چگالی X باشد، ثابت کنید که

$$f(x+1) = \frac{p(n-x)}{q(x+1)} f(x)$$

با استفاده از این برابری ساده، با فرض $n=10$ و $p=0.75, 0.5, 0.25$ برای $f(x)$ یک جدول بسازید.

۱۱. فرض کنید $X_1 \sim B(100, 0.01)$ و $X_2 \sim B(200, 0.01)$ مستقل باشند. توزیع $X_1 + X_2$ را پیدا کنید.

توزیع تقریبی $X_1 + X_2$ و X_2 را پیدا کنید.

۱۲. اگر Y_1 و Y_2 مستقل و دارای توزیع پواسن با پارامترهای ۱ و ۲ باشند، توزیع $Y_1 + Y_2$ را حدس بزنید.

دانشمندی: از تمرین ۱۱ استفاده کنید.

۳.۵ توزیع نرمال

یکی از توزیمهای معروف توزیع نرمال است که در فصل اول قدری با آن آشنا شدیم. این توزیع را، که در تمام رشته‌های آمار و احتمال نقشی ارزنده دارد، ابراهام دو موآور (۱۷۵۴-۱۶۶۷) ریاضیدان فرانسوی معرفی کرده است. در این بخش ویژگیهای این توزیع را با تفصیل بیشتر مطالعه می‌کنیم.

۱.۳.۵ چگالی توزیع نرمال

چگالی زیر را چگالی نرمال با پارامترهای μ و σ^2 می‌نامند:

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \quad -\infty < x < \infty$$

هرگاه متغیر تصادفی X دارای این چگالی باشد، می‌گوئیم X دارای توزیع نرمال با پارامترهای μ و σ^2 می‌باشد و می‌نویسیم $X \sim N(\mu, \sigma^2)$. مشاهده می‌شود که X یک متغیر تصادفی پیوسته روی R می‌باشد.

تابع توزیع X به صورت انتگرال زیر است.

$$N(x; \mu, \sigma^2) = \int_{-\infty}^x f(t; \mu, \sigma^2) dt$$

این انتگرال را تنها می‌توان به تقریب محاسبه کرد.

۲.۳.۵ میانگین و واریانس توزیع نرمال

با انتگرال‌گیری می‌توان نشان داد که

$$E(X) = \int_{-\infty}^{\infty} xf(x; \mu, \sigma^2) dx = \mu$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x; \mu, \sigma^2) dx = \mu^2 + \sigma^2$$

$$Var(X) = E(X^2) - E^2(X) = \mu^2 + \sigma^2 - \mu^2 = \sigma^2$$

بنابراین در توزیع نرمال پارامتر μ میانگین X و پارامتر σ^2 واریانس X می‌باشد.

با استفاده از جدول III برای متغیر نرمال استاندارد Z داریم:

$$P(Z \leq 2) = \Phi(2) = 0.9772$$

$$P(Z \leq 0) = \Phi(0) = 0.5000$$

$$P(Z \leq -2/2) = 1 - \Phi(2/2) = 1 - 0.9772 = 0.0228$$

$$P(Z > -\frac{1}{2}) = 1 - P(Z \leq -\frac{1}{2}) = 1 - [1 - \Phi(\frac{1}{2})] = 0.6915$$

$$P(-1 \leq Z \leq 2) = \Phi(2) - \Phi(-1) = \Phi(2) + \Phi(1) - 1 \\ = 0.9772 + 0.8413 - 1 = 0.8185$$

۴.۳.۵ محاسبه احتمال برای متغیر نرمال غیر استاندارد

هرگاه $X \sim N(\mu, \sigma^2)$ می توان ثابت کرد که تبدیل $Z = \frac{X - \mu}{\sigma}$ دارای توزیع نرمال استاندارد است. با استفاده از این تبدیل برای متغیر نرمال غیر استاندارد، احتمال را به راحتی محاسبه می کنند. مثلاً با فرض $X \sim N(2, 9)$ داریم:

$$P(X < 3) = P\left(\frac{X - \mu}{\sigma} < \frac{3 - 2}{3}\right) = P(Z < 0.33) = 0.6293$$

$$P(-5 < X < 4) = P\left(\frac{-5 - 2}{3} < \frac{X - \mu}{\sigma} < \frac{4 - 2}{3}\right) = P(-2.33 < Z < 0.67) \\ = \Phi(0.67) - \Phi(-2.33) = 0.7486 + 0.9901 - 1 \\ = 0.7387$$

۵.۳.۵ ترکیب خطی چند متغیر نرمال

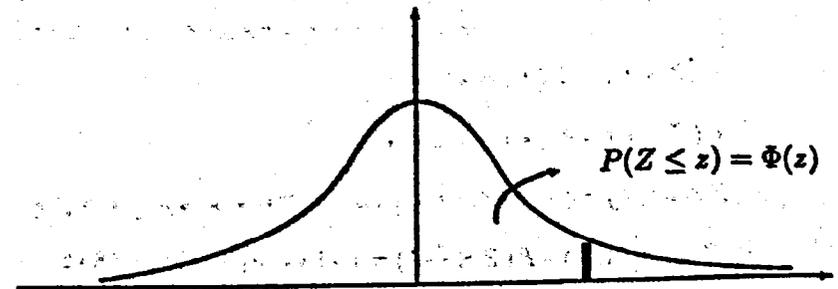
فرض کنید X_1, X_2, \dots, X_n مستقل و برای هر X_i داشته باشیم $X_i \sim N(\mu_i, \sigma_i^2)$. مقادیر ثابت $c_0, c_1, c_2, \dots, c_n$ را در نظر می گیریم. می توان ثابت کرد که

$$Y = c_0 + c_1 X_1 + c_2 X_2 + \dots + c_n X_n$$

دارای توزیع نرمال با میانگین و واریانس زیر است:

$$\mu_Y = c_0 + c_1 \mu_1 + c_2 \mu_2 + \dots + c_n \mu_n$$

$$\sigma_Y^2 = c_1^2 \sigma_1^2 + c_2^2 \sigma_2^2 + \dots + c_n^2 \sigma_n^2$$



نگاره ۱ چگالی نرمال استاندارد

۳.۳.۵ توزیع نرمال استاندارد

توزیع نرمال را با $\mu = 0$ و $\sigma^2 = 1$ ، توزیع نرمال استاندارد می نامند. معمولاً یک متغیر با این توزیع را با Z نشان می دهند. چگالی Z بصورت زیر است: چگالی توزیع نرمال استاندارد را به صورت زیر نشان می دهند:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad -\infty < z < \infty$$

منحنی متقارن این چگالی را در نگاره ۱ می بینید.

توزیع نظیر این چگالی را با $\Phi(z)$ نشان می دهند. پس داریم:

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

این انتگرال را نمی توان به صورت تابعی صریح از Z محاسبه کرد، ولی می توان آن را با بسط تابع $\phi(t)$ به صورت یک سری به تقریب محاسبه نمود. در عمل با استفاده از جدول ۳ مقدار تقریبی $\Phi(z)$ را برای بعضی از مقادیر Z محاسبه می کنند. چون منحنی $\phi(z)$ نسبت به محور عرضها متقارن است داریم

$$\Phi(-z) = 1 - \Phi(z) \quad , \quad \Phi(0) = \frac{1}{2}$$

به ویژه هرگاه X_i ها نرمال و مستقل با میانگین مشترک μ و واریانس مشترک σ^2 باشند، آنگاه معدل X_i ها یعنی

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

دارای توزیع نرمال با میانگین μ و واریانس $\frac{\sigma^2}{n}$ می باشد. این نتیجه با فرض $c_1 = c_2 = \dots = c_n = \frac{1}{n}$ و $c_0 = 0$ می باشد.

مثال ۶ درآمد روزانه شخصی از سه محل تامین می گردد. از محل اول مقدار ثابت ۱۲۰ تومان، از محل دوم مقدار تصادفی $X_1 \sim N(100, 64)$ و از محل سوم مقدار تصادفی $X_2 \sim N(80, 36)$ با فرض مستقل بودن X_1 و X_2 ، احتمال اینکه درآمد روزانه این شخص بیش از ۳۱۰ تومان بشود چقدر است؟

حل - درآمد روزانه این شخص می شود متغیر تصادفی $Y = 120 + X_1 + X_2$ ، که دارای توزیع نرمال با پارامترهای زیر است:

$$\mu = E(Y) = 120 + \mu_1 + \mu_2 = 120 + 100 + 80 = 300$$

$$\sigma^2 = Var(Y) = \sigma_1^2 + \sigma_2^2 = 64 + 36 = 100$$

حال داریم:

$$P(Y > 310) = P\left(\frac{Y - \mu}{\sigma} > \frac{310 - 300}{10}\right) = P(Z > 1) \\ = 1 - P(Z \leq 1) = 1 - \Phi(1) = 0.2420$$

۶.۳.۵ تمرین بخش سه

۱. اگر Z دارای توزیع نرمال استاندارد باشد، احتمالهای زیر را محاسبه کنید:

$$P(Z \leq a) \quad a = 0, 1, 2, 3, 4$$

$$P(-a \leq Z \leq a) \quad a = 1, 2, 3, 4$$

۲. اگر $X \sim N(2, 9)$ ، احتمالهای زیر را حساب کنید:

$$P(X < 3), P(X > -1), P(1 \leq X < 4)$$

۳. اگر Z_1 و Z_2 مستقل و هر دو نرمال استاندارد باشند، ثابت کنید که $P(Z_1 Z_2 > 0) = \frac{1}{4}$. اگر Z_1 و Z_2 مستقل و فقط مقارن باشند، آیا این رابطه درست است؟

۴. درجه حرارت در شهری دارای توزیع نرمال با میانگین ۶۸ درجه فارنهایت و انحراف معیار ۴ درجه فارنهایت می باشد. احتمال اینکه درجه حرارت بیش از ۷۰ درجه فارنهایت باشد چقدر است؟ توزیع درجه حرارت برحسب درجه سانتی گراد چیست؟

۵. دوره حاملگی دارای توزیع نرمال با میانگین ۲۷۰ روز و واریانس ۱۰۰ روز می باشد. احتمال اینکه این دوره کمتر از ۲۵۰ روز باشد چیست؟

۶. نمره دانشجویان علوم در یک آزمون ریاضی دارای توزیع $N(60, 25)$ است. استاد درس به دانشجویانی که نمره آنها بیش از $\mu + \sigma$ باشد نمره الف می دهد. در یک کلاس ۲۰۰ نفری انتظار چند الف دارند؟

۷. اگر X و Y مستقل و هر دو دارای توزیع $N(1, 4)$ باشند.

$$P(X < 2Y + 1), P(XY > 0), P(|X - Y| > 2)$$

را حساب کنید.

۸. با فرض $Z \sim N(0, 1)$ میانگین و واریانس Z را از راه انتگرال گیری پیدا کنید.

۹. با فرض $X \sim N(\mu, \sigma^2)$ میانگین و واریانس X را به کمک تمرین ۸ پیدا کنید.

۱۰. با فرض $X \sim N(2, 4)$ مطلوبیست محاسبه $P(|X| < 1)$ و $P(4X^2 > 1)$.

۴.۵ قضیه حد مرکزی و کاربرد آن

در این بخش به شرح یک قضیه معروف احتمال بنام قضیه حد مرکزی می پردازیم. از نظر تاریخی، در موآور برای اولین بار در اوائل قرن هیجدهم میلادی این قضیه را برای متغیرهای تصادفی برنولی بکار برد. بعدها دیگران، در نیمه اول قرن بیستم، قضیه را در جهت های مختلف گسترش داده آن را تکمیل نمودند.

در اینجا، تنها قضیه را برای حالتی بسیار ساده، آن هم بدون اثبات، شرح می دهیم و با ذکر چند مثال کاربرد آن را روشن می کنیم.

۱.۴.۵ نمونه تصادفی

متغیرهای تصادفی مستقل و هم‌توزیع X_1, X_2, \dots, X_n را، که دارای توزیع مشترک $F(x)$ هستند، یک نمونه تصادفی n تایی از این توزیع می‌نامند. مثلاً وقتی یک تاس نالرئب را ده بار مستقلاً می‌ریزیم، یک نمونه تصادفی ده‌تایی از این توزیع با چگالی زیر داریم:

$$f(x) = \begin{cases} \frac{1}{6} & x = 1, 2, \dots, 6 \\ 0 & \text{جای دیگر} \end{cases}$$

یافته‌های این نمونه تصادفی ده‌تایی، یعنی مقادیری که عملاً مشاهده می‌کنیم، ممکن است بشوند:

$$1, 5, 5, 4, 2, 6, 1, 3, 2, 1$$

به عنوان مثالی دیگر، فرض کنید متغیر تصادفی $X \sim N(100, 16)$ اندازه قامت پسر هفت ساله باشد اگر ۵ پسر هفت ساله را به تصادف انتخاب کنیم یک نمونه تصادفی پنج‌تایی از این توزیع نرمال داریم. یافته‌های این نمونه تصادفی پنج‌تایی تا یک سانتیمتر تقریب ممکن است بشوند:

$$92, 94, 104, 105, 98$$

حال فرض کنید T_n و \bar{X}_n به ترتیب مجموع و معدل X_i ها در یک نمونه تصادفی n تایی باشند، یعنی

$$T_n = X_1 + X_2 + \dots + X_n$$

$$\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{T_n}{n}$$

اغلب نیاز داریم درباره این دو متغیر تصادفی احتمالاتی زیر را محاسبه کنیم:

$$A < B \quad \text{برای دو عدد} \quad P(A \leq T_n \leq B)$$

$$a < b \quad \text{برای دو عدد} \quad P(a \leq \bar{X}_n \leq b)$$

در حالت کلی انجام این کار مستلزم داشتن توزیع T_n و توزیع \bar{X}_n می‌باشد که معمولاً به آنها دسترسی نداریم. البته اگر نمونه تصادفی از توزیع نرمال آمده باشد، آنگاه T_n و \bar{X}_n هم، که به صورت ترکیب

خطی از متغیرهای نرمال و مستقل X_1, X_2, \dots, X_n می‌باشند، نرمال هستند و محاسبه آسان می‌شود. آیا این نتیجه در حالت کلی هم درست است؟ باید گفت نه! ولی اگر n به اندازه کافی بزرگ باشد، می‌توان گفت که تحت شرایطی T_n و \bar{X}_n تقریباً دارای توزیع نرمال هستند. در حقیقت این بیان، عصاره قضیه حد مرکزی می‌باشد. اینک قضیه حد مرکزی را دقیقاً شرح می‌دهیم:

۲.۴.۵ قضیه حد مرکزی

فرض کنید X_1, X_2, \dots, X_n یک نمونه تصادفی از توزیع $F(x)$ با میانگین و واریانس μ و σ^2 باشد، به طوری که $\sigma^2 < \infty$. اگر n به اندازه کافی بزرگ باشد، دو نتیجه تقریبی زیر را داریم:

الف - $T_n = X_1 + X_2 + \dots + X_n$ تقریباً دارای توزیع $N(n\mu, n\sigma^2)$ است.

ب - $\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}$ تقریباً دارای توزیع $N(\mu, \frac{\sigma^2}{n})$ است.

این توزیعهای تقریبی را به صورت زیر نشان می‌دهیم:

$$T_n \approx N(n\mu, n\sigma^2) \quad \text{مجموع}$$

$$\bar{X}_n \approx N(\mu, \frac{\sigma^2}{n}) \quad \text{معدل}$$

اگر Φ توزیع نرمال استاندارد باشد، به صورت حدی برای هر z داریم:

$$\lim_{n \rightarrow \infty} P\left(\frac{T_n - n\mu}{\sqrt{n\sigma^2}} \leq z\right) = \Phi(z)$$

$$\lim_{n \rightarrow \infty} P\left(\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z\right) = \Phi(z)$$

همان طوری که قبلاً دیدیم، اگر Z نرمال استاندارد باشد (به نگاره ۱ توجه کنید)

$$\Phi(z) = P(Z \leq z)$$

و این احتمال را می‌توان از روی جدول ۳ پیدا کرد.

بزرگ بودن n به اندازه کافی شرطی مبهم می‌باشد که در تمام مسائل مورد تقریبی باعث کنجکاو می‌گردد و انسان می‌پرسد به چه بزرگی؟ در این قضیه باید گفت که اگر توزیع F چندان چوله به چپ یا راست نباشد برای $n \geq 25$ تقریب خوبی بدست می‌آید. اینک به ذکر یک مثال می‌پردازیم:

مثال ۷ ده عدد به تصادف در فاصله $(0, 1)$ انتخاب می‌کنیم. احتمال اینکه مجموع آنها بیش از شش باشد چقدر است؟

حل - این ده عدد تصادفی، یک نمونه تصادفی ده تایی $X_1, X_2, \dots, X_9, X_{10}$ را تشکیل می‌دهند. این متغیرهای تصادفی مستقل بوده و هر یک دارای توزیع یکنواخت در فاصله $(0, 1)$ ، با میانگین $\mu = \frac{1}{2}$ و واریانس $\sigma^2 = \frac{1}{12}$ می‌باشد. فرض کنید T مجموع این ده عدد تصادفی باشد. می‌خواهیم $P(T > 6)$ را محاسبه کنیم. طبق قضیه حد مرکزی توزیع تقریبی T می‌شود $N(5, \frac{5}{6})$. از اینرو احتمال مطلوب می‌شود:

$$\begin{aligned} P(X_1 + X_2 + \dots + X_{10} > 6) &= P(T > 6) = P\left(\frac{T - \mu T}{\sigma T} > \frac{6 - 5}{\sqrt{\frac{5}{6}}}\right) \\ &\approx P(Z > 1/1) = 1 - P(Z \leq 1/1) \\ &= 1 - \Phi(1/1) \approx 0.136 \end{aligned}$$

۳.۴.۵ ارتباط میان توزیع دو جمله‌ای و نرمال

فرض کنید Y_1, Y_2, \dots, Y_n یک نمونه تصادفی از توزیع برنولی با پارامتر p باشد. همان طوری که می‌دانیم متغیر تصادفی $X = \sum_{i=1}^n Y_i$ ، یعنی تعداد پیروزیها، دارای توزیع دو جمله‌ای با میانگین np و واریانس npq می‌باشد. پس بنا بر قضیه حد مرکزی، در صورتیکه n به اندازه کافی بزرگ باشد، متغیر تصادفی

$$\frac{X - np}{\sqrt{npq}}$$

تقریباً دارای توزیع نرمال استاندارد است. به سخنی دیگر $X \sim B(n, p)$ و $U \sim N(np, npq)$ تقریباً هم‌توزیع هستند. برای هر دو عدد درست a و b با فرض $0 \leq a < b \leq n$ داریم:

$$P(a \leq X \leq b) \approx P(a \leq U \leq b)$$

هر اندازه n بزرگتر و p به $\frac{1}{2}$ نزدیکتر باشد، تقریب بالا بهتر خواهد بود.

۴.۴.۵ تصحیح پیوستگی

تقریب توزیع دو جمله‌ای به وسیله توزیع نرمال در حقیقت تقریب یک توزیع گسسته به وسیله یک توزیع پیوسته می‌باشد. برای اینکه $P(a \leq X \leq b)$ با تقریب بهتری پیدا شود معمولاً آن را با

$P(a - 0.5 \leq U \leq b + 0.5)$ تقریب می‌کند. این عمل را تصحیح پیوستگی می‌نامند. دلیل افزودن یا کاستن 0.5 را به نقاط انتهایی از روی هیسئوگرام به زودی شرح می‌دهیم. می‌دانیم که برای متغیر نرمال $U \sim N(np, npq)$ بعد از استاندارد کردن داریم:

$$\begin{aligned} P(a - 0.5 \leq U \leq b + 0.5) &= P\left(\frac{a - 0.5 - np}{\sqrt{npq}} \leq Z \leq \frac{b + 0.5 - np}{\sqrt{npq}}\right) \\ &= \Phi\left(\frac{b + 0.5 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - 0.5 - np}{\sqrt{npq}}\right) \end{aligned}$$

ملاحظه می‌شود که جمله $\frac{0.5}{\sqrt{npq}}$ با بزرگ شدن n به صفر نزدیک می‌شود. پس در صورتیکه n به قدر کافی بزرگ باشد تصحیح پیوستگی عملاً بی اثر خواهد بود.

مثال ۸ فرض کنید $X \sim B(4, \frac{1}{4})$. مقدار واقعی و تقریب $P(2 \leq X \leq 4)$ را حساب کنید.

حل - می‌دانیم که

$$E(X) = 1, \quad \text{Var}(X) = 1$$

$$P(X=x) = \binom{4}{x} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{4-x} \quad x = 0, 1, 2, 3, 4$$

بنابراین جدول توزیع احتمال X به صورت زیر است

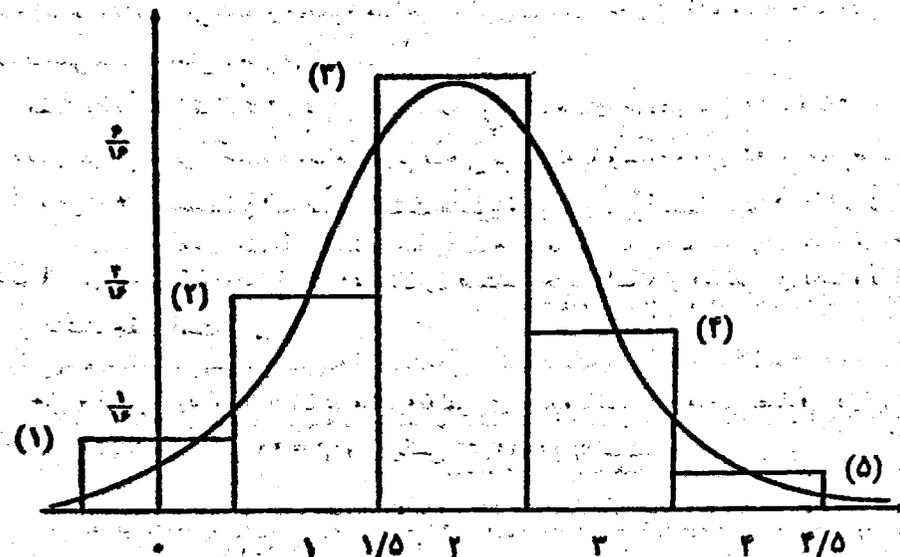
| | | | | | |
|----------|-----------------|-----------------|-----------------|-----------------|-----------------|
| $X=x$ | 0 | 1 | 2 | 3 | 4 |
| $P(X=x)$ | $\frac{1}{256}$ | $\frac{4}{256}$ | $\frac{6}{256}$ | $\frac{4}{256}$ | $\frac{1}{256}$ |

دقیقاً داریم:

$$\begin{aligned} P(2 \leq X \leq 4) &= \sum_{x=2}^4 P(X=x) \\ &= \frac{11}{16} = 0.6875 \end{aligned}$$

اگر توزیع $X \sim B(4, \frac{1}{4})$ را با توزیع $U \sim N(1, 1)$ تقریب کنیم داریم:

$$\begin{aligned} P(2 \leq X \leq 4) &= P(2 \leq U \leq 4) = P\left(\frac{2-1}{1} \leq Z \leq \frac{4-1}{1}\right) \\ &= P(1 \leq Z \leq 3) = \Phi(3) - \Phi(1) \\ &= 0.9974 - 0.2420 = 0.7554 \end{aligned}$$



نگاره ۲ منحنی نرمال و هیستوگرام

می بینید که این مقدار تقریبی با مقدار واقعی تفاوت زیاد دارد زیرا $n = 4$ بزرگ نیست. حال محاسبه را با تصحیح پیوستگی انجام می دهیم.

$$\begin{aligned} P(2 \leq X \leq 4) &\approx P(2 - 0.5 \leq U \leq 4 + 0.5) \\ &= P(1.5 \leq U \leq 4.5) \\ &= P\left(\frac{1.5 - 2}{1} \leq Z \leq \frac{4.5 - 2}{1}\right) \\ &= 0.9938 + 0.6915 - 1 = 0.6853 \end{aligned}$$

با اینکه $n = 4$ بزرگ نیست، این تقریب به مقدار واقعی نزدیک است. نگاره ۲ نشان می دهد که چرا در این مثال با استفاده از تصحیح پیوستگی تقریب بهتری داریم. ضمناً چون $p = \frac{1}{4}$ ، شکلی متقارن بدست می آید.

نخست هیستوگرام توزیع احتمال X را با استفاده از جدول توزیع احتمال بدین طریق رسم می کنیم: مستطیلهائی رسم می کنیم که دارای قاعده هائی به طول یک روی محور X می باشند و وسط قاعده ها از

نقاطی به طولهای ۰.۱، ۰.۲، ۰.۳، ۰.۴، ۰.۵ تشکیل شده باشند. مساحت یا ارتفاع هر مستطیل به اندازه احتمال نقطه وسط قاعده آن می باشد و مجموع مساحتها برابر یک است. حال از نقاط وسط قاعده هائی بتلاشی یک منحنی شبیه منحنی نرمال می گذرانیم.

مقدار واقعی $P(2 \leq X \leq 4)$ برابر است با مجموع مساحتهای مستطیلهای (۲) و (۳) و (۴) و (۵). اما مقدار تقریبی آن برابر است با مساحت زیر منحنی نرمال که به وسیله محور X ها و خطوط عمودی $x = 1/5$ و $x = 4/5$ محصور شده است (قسمت هاشور خورده) و برابر است با $P(1/5 < U < 4/5)$.

مثال ۹ در شهری ۵۶ درصد از رای دهندگان زن هستند. احتمال اینکه از ۵۰ نفر رای دهنده لاقل ۳۰ نفر زن باشند چقدر است؟

حل - فرض کنید X تعداد زن در میان ۵۰ نفر رای دهنده باشند. این متغیر تصادفی دارای توزیع $B(50, 0.56)$ با میانگین $np = 28$ و واریانس $npq = 12/32$ است.

$$\begin{aligned} P(X \geq 30) &= P(U \geq 29/5) \\ &= P\left(\frac{U - np}{\sqrt{npq}} > \frac{29/5 - 28}{\sqrt{12/32}}\right) \\ &= P(Z \geq 0.43) = 1 - \Phi(0.43) \\ &= 1 - 0.67 = 0.33 \end{aligned}$$

۵.۴.۵ تمرین بخش چهار

- در فاصله (۰، ۱) تعداد ۷۵ عدد تصادفی را انتخاب می کنیم. نشان دهید احتمال اینکه معدل این عددها در فاصله (۰/۴۵، ۰/۵۵) قرار گیرد تقریباً برابر است با ۰/۸۶۶.
- یک سکه نازیب را ۶ بار پرتاب می کنیم. اگر X تعداد شیرها باشد، احتمال $P(2 \leq X \leq 3)$ را دقیقاً و با تقریب نرمال حساب و با یک نمودار نتایج را بیان کنید.
- فرض کنید $X \sim B(n, p)$. برای عدد درست $0 \leq a \leq n$ با استفاده از تقریب نرمال $P(X = a)$ را پیدا کنید (از یک نمودار استفاده کنید).

۴. احتمال اینکه نوزاد در تاریخی که دکتر پیش بینی می کند متولد شود $\frac{1}{4}$ می باشد. احتمال اینکه از ۴۰۰ نوزاد ۱۵ نوزاد در تاریخ پیش بینی شده متولد شوند چیست؟ تقریب بواسن و نرمال را بکار برید.

۵. یک نمونه تصادفی ۷۲ تایی از چگالی زیر را در نظر می‌گیریم:

$$f(x) = \begin{cases} \frac{1}{x^2} & x > 1 \\ 0 & \text{جای دیگر} \end{cases}$$

احتمال اینکه لااقل ۵۰ عضو این نمونه از ۳ بزرگتر باشند تقریباً چقدر است؟
۶. وزارتخانه‌ای ۸ خط تلفن و ۴۰ بخش دارد. هر بخش به مدت یک ساعت در روز (۱۲ ساعت) از این خطها استفاده می‌کند. فرض کنید X تعداد خطهای اشغال شده در لحظه‌ای معین باشد.

الف - نشان دهید که $X \sim B(40, \frac{1}{12})$.

ب - $P(X < 8)$ را به تقریب محاسبه کنید. این احتمال را ضریب دسترسی به خط آزاد می‌نامیم.
ج - قرار است تعداد بخشها را به ۶۰ برسانند. اگر بخواهند ضریب دسترسی به خط آزاد تغییر نکند، وزارتخانه باید چند خط تلفن داشته باشد؟

۷. در ظرفی ۴۵ کارت سفید و ۵ کارت سبز داریم. ۱۰۰ بار، با جایگذاری از این ظرف به تصادف یک کارت بیرون می‌آوریم. فرض کنید X تعداد کارت‌های سبز در این نمونه ۱۰۰ تایی باشد.
 $P(X > 16)$ را به کمک توزیع دو جمله‌ای، نرمال و پواسن پیدا کنید و نتایج را با هم مقایسه نمایید.

۸. نیروی تحمل یک نوع میله به طور متوسط ۱۴ کیلوگرم با انحراف معیار ۲ کیلوگرم است. احتمال تقریبی اینکه معدل نیروی تحمل برای ۱۰۰ عدد از این میله‌ها بیش از ۱۴/۵ کیلوگرم باشد چیست؟

۹. یک نمونه تصادفی ۱۰۰ تایی از توزیع زیر را در نظر می‌گیریم:

$$F(x) = x^3 \quad 0 \leq x \leq 1$$

احتمال تقریبی اینکه معدل این نمونه بیش از ۰/۶ باشد چیست؟

۱۰. نمونه تصادفی X_1, \dots, X_n را از یک توزیع پواسن با میانگین ۲ در نظر می‌گیریم. احتمال تقریبی اینکه معدل این نمونه بیش از ۵ باشد چیست؟

نتیجه گیری آماری

در فصل نخست دیدیم که چگونه می‌توان داده‌ها، یعنی اطلاعات عددی درباره‌ی یک آزمایش تصادفی، را خلاصه کرد. در فصلهای دیگر قوانین احتمال، متغیر تصادفی، تابع توزیع، تابع چگالی، امید ریاضی و چند توزیع مهم را شرح دادیم.

در این فصل می‌خواهیم گفتار نخست را که جنبه عددی و مشاهده دارد با گفتارهای دیگر که جنبه نظری دارند و در آنها الگوهای شانس، یعنی قوانین توزیع احتمال، معرفی شده‌اند به نحوی ارتباط دهیم و به یک نوع نتیجه‌گیری به نام نتیجه‌گیری آماری یا استنباط آماری بپردازیم.

نخست منظور از آمار و فرق آن را با احتمال مطالعه می‌کنیم. سپس درباره‌ی پارامتر توزیع یک متغیر تصادفی تحت عناوین زیر سخن می‌گوئیم:

۱. برآورد یک پارامتر

۲. فاصله اطمینان برای یک پارامتر

۳. آزمون آماری یک پارامتر

در پایان این فصل به کوتاهی درباره‌ی متغیرهای X^2 و F ، که در کارهای آماری دارای کاربرد فراوان هستند سخن می‌گوئیم.

۱.۶. فرق احتمال و آمار

به زبان ساده، احتمال یعنی عددی در فاصله $[0, 1]$ که میزان یقین درباره‌ی امری را بیان می‌دارد. آمار یعنی داده‌های عددی درباره‌ی امری. مثلاً می‌گوئیم احتمال بارندگی در فروردین ماه امسال ۰/۵ است. ولی می‌گوئیم آمار نشان می‌دهد طی ۴ سال گذشته میزان باران در فروردین ماه ۰/۵۲، ۰/۶۱، ۰/۷۵، ۰/۸۰ میلی متر بوده است.

اینک که با مفهوم دقیق احتمال و مدل‌های احتمالی پیدا کرده‌ایم، و در ضمن می‌دانیم که چگونه داده‌های عددی را گروه بندی و خلاصه کنیم، منظور از علم آمار و احتمال را به صورت زیر بیان می‌داریم: آمار و احتمال عبارت است از جمع‌آوری داده‌های عددی با اصولی مقبول و پیاده کردن مدل‌های احتمالی روی آنها به منظور استخراج نتایج بنام نتایج آماری.

۱.۱.۶ روش منطقی احتمال و آمار

روش احتمال از نظر منطقی یک روش قیاسی است. در این روش باید به کبری یعنی قانون کلی (مثلاً انسان در معرض خطاست) و صغری یعنی حالت جزئی (مثلاً نویسنده این کتاب انسان است) و نتیجه یعنی نتیجه مقایسه حالت جزئی با قانون کلی (مثلاً نویسنده این کتاب در معرض خطاست) توجه کرد. روش آماری از نظر منطقی یک روش استقرائی است، یعنی روشی است که طبق آن از امور جزئی (مثلاً مشاهده ۲۸۰ قریه با آب و برق در میان ۵۰۰ قریه) به یک قانون کلی (مثلاً احتمال آب و برق داشتن در هر قریه تقریباً ۵/۱۰ می‌باشد) پی می‌بریم. برای روشن شدن مطلب به مثال زیر توجه کنید.

سکه‌ای را چندین بار مثلاً ۵۰۰ بار مستقلاً و تحت شرایط یکسان می‌اندازیم و فرض کنید ۲۸۰ بار شیر بیاید. با تعبیر احتمال به طریق فراوانی نسبی و با روش استقرائی می‌توان مقدار تقریبی p ، یعنی شانس شیر آمدن را برابر $\frac{280}{500} = 0.56$ گرفت. در این عدد تا حدودی ناباوری و تردید وجود دارد. مثلاً اگر خود شما یا دیگری از نوسکه را ۵۰۰ بار بیاندازد، ممکن است تعداد شیرها با ۲۸۰ تفاوت پیدا کنند. همچنین ممکن است شخص بخواهد با ابتکاری خاص (نه استفاده از احتمال به طریق فراوانی نسبی) از عدد ۲۸۰ استفاده کرده، مقدار تقریبی p را پیدا کند. بنابراین مقدر تقریبی p ، که آن را برآورد p می‌نامیم، بر اثر عوامل تصادفی یا روش برآورد کاملاً محرز و مشخص نیست و همواره با مقدار واقعی آن تفاوت دارد. در آمار اصول و قواعدی را مطالعه می‌کنیم تا بر اساس آنها بتوانیم بهترین مقدر تقریبی را برای پارامتر مجهول p پیدا کنیم.

اینک فرض کنید که شانس شیر آمدن برای سکه یاد شده با عدد ۰/۵۶ برآورد شود. این عدد را به طریق آماری به دست آورده‌ایم و آن را به عنوان یک قانون کلی برای شیر آمدن می‌پذیریم. حال سکه را دوبار می‌اندازیم. با استفاده از این قانون، احتمال اینکه سکه هر دو بار شیر بیاید می‌شود عدد $\frac{0.56^2}{1} = 0.3136$. در این عدد بعد از قبول قانون کلی، تردیدی وجود ندارد زیرا آن را با روش قیاسی به دست آورده‌ایم.

۲.۶ برآورد آماری یک پارامتر

متغیر تصادفی گسسته یا پیوسته X را در نظر می‌گیریم. این متغیر ممکن است روی یک جمعیت تعریف شده باشد یا نتیجه یک آزمایش تصادفی باشد. مثلاً تعداد فرزندان یک جوان ۲۵ ساله یک متغیر تصادفی گسسته و اندازه قامت او یک متغیر تصادفی پیوسته است. هرگاه سکه‌ای را پرتاب کنیم نتیجه این آزمایش تصادفی از نظر شیر و خط یک متغیر تصادفی گسسته (متغیر برنولی) و از نظر مدت زمانی که سکه در حرکت است یک متغیر تصادفی پیوسته می‌باشد.

فرض کنید چگالی X برای ما معلوم باشد ولی به یک یا چند پارامتر مجهول بستگی داشته باشد مثلاً در آزمایش برنولی، X دارای چگالی برنولی یا پارامتر مجهول p است. در مورد اندازه قامت جوانان ۲۵ ساله، از راه جمع‌آوری و خلاصه کردن داده‌ها، می‌توان حدس زد که X دارای توزیع نرمال است. فرض کنید μ پارامتر مجهول باشد. می‌خواهیم این پارامتر مجهول را به تقریب تعیین کنیم یا به زبان آماری آن را برآورد نماییم. برای این منظور نخست مفهوم نمونه تصادفی را شرح می‌دهیم.

۱.۲.۶ نمونه تصادفی

هر نوع قضاوت آماری، به کمک مشاهدات مکرر یا آزمایشهای متعدد انجام می‌گیرد. بنابراین X را مستقلاً n بار تحت شرایط یکسان به صورت X_1, X_2, \dots, X_n تصور می‌کنیم. واضح است که این متغیرهای تصادفی مستقل و هم‌توزیع می‌باشند و در عمل پس از مشاهده به صورت n عدد حقیقی x_1, x_2, \dots, x_n در می‌آیند.

۲.۲.۶ تعریف یک نمونه تصادفی

متغیرهای تصادفی X_1, X_2, \dots, X_n یک نمونه تصادفی n تایی از X هستند، هرگاه از یکدیگر مستقل و با هم‌توزیع باشند.

۳.۲.۶ یافته‌های یک نمونه تصادفی

هرگاه در عمل اعداد x_1, x_2, \dots, x_n را برای نمونه تصادفی مشاهده کنیم، آنها را یافته‌های این نمونه تصادفی می‌نامیم.

مثال ۱ یک تاس نازیب را پنج بار می‌ریزیم. نمونه تصادفی ۵ تایی X_1, X_2, \dots, X_5 به دست می‌آید.

اگر در عمل برای این نمونه تصادفی اعداد ۱، ۶، ۲، ۳، ۲، ۳ را مشاهده کنیم، این اعداد یافته‌های این نمونه تصادفی هستند.

مثال ۲ فرض کنید متغیر تصادفی X (برحسب میلی گرم) مقدار طلائی باشد که از یک کیلوگرم سنگ معدن طلا در آزمایشگاه به دست می‌آید. هرگاه این آزمایش را چهاربار مستقل انجام دهیم، یک نمونه تصادفی چهارتایی X_1, X_2, X_3, X_4 به دست می‌آید. اگر در عمل اعداد $۰/۶۱۲, ۰/۷۱۲, ۰/۵۱۳, ۰/۲۱۱$ را مشاهده کنیم، این اعداد یافته‌های این نمونه تصادفی هستند.

۴.۲.۶ آماره

فرض کنید متغیر تصادفی X دارای چگالی $f(x; \theta)$ باشد، که در آن $\theta \in A$ یک پارامتر مجهول است. مجموعه A را که پارامتر θ می‌تواند در آن تغییر کند فضای پارامتر می‌نامیم. مثلاً اگر X یک متغیر برنولی، با احتمال $0 < \theta < 1$ برای پیروزی باشد $A = (0, 1)$ فضای پارامتر است. ممکن است پارامتر θ چندبعدی به صورت $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ باشد. مثلاً در توزیع نرمال پارامتر $\theta = (\mu, \sigma^2)$ یک پارامتر دو بعدی می‌باشد و فضای پارامتر عبارت است از ناحیه بالای محور طولها در صفحه مختصات. برای اینکه مقدار تقریبی θ را پیدا کنیم، اعضای نمونه را به نحوی مناسب با هم ترکیب می‌توانیم منی تابعی مناسب از آنها می‌سازیم تا بتوانیم پارامتر مجهول را برآورد کنیم.

۵.۲.۶ تعریف آماره

تابع $U = g(X_1, X_2, \dots, X_n)$ از نمونه تصادفی X_1, X_2, \dots, X_n را که در آن پارامتر مجهول نباشد یک آماره می‌نامند. واضح است که U یک متغیر تصادفی است و توزیع آن ممکن است به پارامتر بستگی داشته یا نداشته باشد. مثلاً در توزیع $N(\mu, 1)$ ، توزیع آماره $U = \bar{X}$ یعنی $N(\mu - \frac{1}{n})$ به پارامتر بستگی دارد، ولی توزیع آماره $V = X_n - X_1$ یعنی $N(0, 2)$ به پارامتر بستگی ندارد. آماره‌ای برای برآورد کردن مفید است که توزیع آن به پارامتر مجهول بستگی داشته باشد و بتواند درباره این پارامتر ما را مطلع دارد.

۶.۲.۶ تعریف برآورد و برآورد یاب

هرگاه به جای نمونه تصادفی یافته‌های آن را در آماره $U = g(X_1, X_2, \dots, X_n)$ بگذاریم، یافته این آماره می‌شود عدد $u = g(x_1, x_2, \dots, x_n)$. اگر این عدد را به عنوان مقدار تقریبی پارامتر θ به کار ببریم،

آنگاه آماره U را یک برآورد یاب (یا برآوردگر یا برآورد کننده یا تخمین زن) و عدد u را یک برآورد (یا تخمین) برای پارامتر θ می‌گویند.

باید افزود که آماره و برآورد یاب هر دو متغیر تصادفی هستند، از اینرو آنها را با حروف بزرگ و یافته‌های آنها را با حروف کوچک نشان می‌دهیم. به سبب شباهت حروف کوچک و بزرگ در بعضی موارد (مانند s و S)، خواننده باید از متن مطلب تفاوت برآورد و برآورد یاب را تمیز دهد. توجه کنید که برآورد یاب پارامتر θ یک آماره است، ولی هر آماره برآورد یاب برای θ نمی‌باشد.

مثال ۳ فرض کنید در مثال ۲ داشته باشیم $X \sim N(\mu, 0/4)$. یک برآورد و یک برآورد یاب برای μ پیدا کنید.

حل - چون μ میانگین X است، پس انتظار داریم که اگر X را چندبار مشاهده کنیم معدل بشود μ . بنابراین یک برآورد مناسب برای μ می‌شود عدد زیر:

$$u = \bar{x} = \frac{0/211 + 0/513 + 0/712 + 0/612}{4} = 0/512$$

بطور کلی برآورد یابی که این برآورد را می‌دهد عبارتست از متغیر تصادفی زیر:

$$U = \bar{X} = \frac{X_1 + X_2 + X_3 + X_4}{4}$$

مثال ۴ عددی به تصادف در فاصله (a, b) اختیار می‌کنیم و آن را با X نشان می‌دهیم a و b را برآورد کنید. در این مثال $\theta = (a, b)$ یک پارامتر دو بعدی می‌باشد.

حل - می‌دانیم X دارای چگالی زیر است:

$$f(x; a, b) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{جای دیگر} \end{cases}$$

این چگالی را چگالی یکنواخت روی فاصله (a, b) می‌نامند. X را مستقلاً چندبار مثلاً پنج بار تصور می‌کنیم تا یک نمونه تصادفی پنج‌تایی بدست آید. فرض کنید اعداد $1/2, 1/3, 1/9, 1/4, 1/7$ یافته‌های این نمونه باشند. می‌توان کوچکترین و بزرگترین عدد یعنی $1/9$ و $1/2$ را به عنوان برآورد a و b اختیار کرد. برآورد یاب a می‌شود آماره $U_1 = \min\{X_1, X_2, \dots, X_5\}$ و برآورد یاب b می‌شود آماره $U_2 = \max\{X_1, X_2, \dots, X_5\}$.

۷.۲.۶ تمرین بخش دو

۱. یک تاس را ۵ بار می ریزیم تا نمونه تصادفی X_1, \dots, X_5 به دست آید. یافته‌های این نمونه به چند طریق می‌توانند رخ دهند؟ احتمال اینکه یافته‌ها به صورت ۱، ۳، ۵، ۱، ۲ باشند چیست؟ شانس مشاهده ۶ بار با روش استقرائی چگونه می‌یابید، هرگاه در ۱۰۰ ریزش ۲۰ بار خال ۶ مشاهده کنیم؟

۲. فرض کنید X_1, X_2, X_3 یک نمونه تصادفی از توزیع $N(\mu, 1)$ باشد. توزیع آماره‌های $U = X_1 + X_2 + X_3$ و $V = 2X_1 - X_2 - X_3$ را بیابید. کدام از این دو آماره برای برآورد کردن پارامتر μ مفید است؟ آیا $W = X_1 + X_2 - \mu$ یک آماره است؟

۳. شخصی اعداد ۸، ۶، ۴، ۷، ۵، ۳، ۹، ۲ را به تصادف در فاصله $(0, \theta)$ پیدا کرده است. پارامتر θ را چگونه برآورد می‌کند؟

۴. فرق برآورد و برآوردیاب چیست؟ فرق آماره و برآوردیاب چیست؟

۵. یک تاس اریب در دست است، که احتمال آمدن خال ۵، با آن $\frac{1}{4}$ است. این عدد را چگونه و با چه روشی یافته‌اند؟ اگر این تاس را ۳ بار بریزیم، احتمال اینکه دوبار خال ۵ مشاهده کنیم چیست؟ این احتمال را با چه روشی می‌یابیم؟

۳.۶ روشهای برآوردیابی

فرض کنید X دارای چگالی $f(x; \theta)$ باشد که در آن پارامتر θ مجهول است. برای اینکه θ را برآورد کنیم، یعنی مقدار تقریبی آن را بیابیم، نخست نمونه تصادفی X_1, X_2, \dots, X_n را از این چگالی در نظر می‌گیریم و فرض می‌کنیم اعداد x_1, x_2, \dots, x_n یافته‌های این نمونه تصادفی باشند. اینک دو روش معروف را برای برآوردیابی θ شرح می‌دهیم.

۱.۳.۶ روش گشتاورها

روش گشتاورها را برای برآورد کردن پارامترها، کارل پیرسن آماردان انگلیسی در ۱۸۹۴ به کار برده است و یکی از قدیمیترین روشها است که بر اساس گشتاورها می‌باشد. همانطوری که می‌دانیم $\mu = E(X)$ را با پایان فرض می‌کنیم) را گشتاور مرتبه نخست X می‌نامند. بدیهی است که معمولاً تابعی از

پارامتر θ است، زیرا مثلاً در حالت پیوسته داریم:

$$\mu = \int_{-\infty}^{\infty} xf(x; \theta) dx = g(\theta) \quad \text{تابعی از } \theta$$

از طرف دیگر میانگین یا معدل X_1, \dots, X_n یعنی

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

را که یک آماره می‌باشد و در آن پارامتر مجهول θ یافت نمی‌شود، گشتاور مرتبه نخست نمونه تصادفی می‌گویند. اگر به جای نمونه تصادفی یافته‌های آن را به کار ببریم آنگاه یافته آماره \bar{X} می‌شود:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

این عدد را گشتاور مرتبه نخست داده‌ها می‌گوئیم.

۲.۳.۶ قانون اعداد بزرگ

بر اساس قانونی مشهور در احتمال، به نام قانون اعداد بزرگ یا قانون معدلها، می‌توان ثابت کرد که برای هر $\epsilon > 0$ با بزرگ شدن n (اندازه نمونه تصادفی) $P(|\bar{X} - \mu| > \epsilon)$ به سمت صفر میل می‌نماید. به سختی دیگر، اگر نمونه بزرگ شود شانس نزدیک شدن معدل \bar{X} ها به μ زیاد می‌گردد.

معمولاً می‌نویسند $\mu \rightarrow \bar{X}$ و می‌گویند \bar{X} در احتمال به μ میل می‌کند، هنگامی که به n به بینهایت برود. به زبان آماری \bar{X} را برآوردیابی سازگار برای μ می‌نامند.

این قانون را برای متغیرهای تصادفی صفر و یک، یکی از برادران برونولی ارائه داد. خنشین، ریاضیدان روسی، در اوائل قرن بیستم آن را در حالتی که μ با پایان باشد ثابت کرد. اگر واریانس X با پایان باشد به کمک نامساوی چیبیف، فوراً این قانون ثابت می‌شود.

با استفاده از این قانون و با استفاده از اینکه $E(\bar{X}) = \mu$ ، می‌توان گفت که \bar{X} برآوردیابی مناسب برای μ است. اینک، با فرض اینکه θ یک بعدی باشد، از حل معادله تقریبی

$$\bar{x} = \mu = q(\theta)$$

برآوردی برای پارامتر مجهول θ به دست می‌آید که آن را برآورد گشتاوری θ می‌گویند. ما این برآورد را با $\hat{\theta}$ نشان می‌دهیم.

مثال ۵ طول عمر یک نوع لامپ متغیر تصادفی X با چگالی زیر است.

$$f(x; \theta) = \begin{cases} \theta e^{-\theta x} & \theta > 0, x > 0 \\ 0 & \text{جای دیگر} \end{cases}$$

با استفاده از یک نمونه تصادفی برآورد گشتاوری θ را بیابید. این چگالی را نمائی منفی می نامند.

حل - به کمک انتگرال جزء به جزء داریم:

$$\mu = E(X) = \int_{-\infty}^{\infty} x \theta e^{-\theta x} dx = \frac{1}{\theta} = g(\theta)$$

حال μ را به کمک \bar{x} برآورد می کنیم. برای این منظور از حل معادله تقریبی

$$\bar{x} = \frac{1}{\theta}$$

برآورد گشتاوری θ می شود عدد $\frac{1}{\bar{x}}$. اگر به جای \bar{x} متغیر تصادفی \bar{X} را بگذاریم، برآوردیاب گشتاوری

θ می شود متغیر تصادفی

$$\bar{\theta} = \frac{1}{\bar{X}}$$

توجه کنید که هم برای برآورد θ و هم برای برآوردیاب θ ، نماد $\bar{\theta}$ را به کار می بریم. از متن مطلب معلوم

می شود که $\bar{\theta}$ چه وقت عدد و چه وقت متغیر تصادفی است.

به طور کلی داریم:

گشتاور k ام متغیر تصادفی X عبارتست از

$$\mu_k = E(X^k) = \int x^k f(x; \theta) dx$$

گشتاور k ام نمونه تصادفی X_1, X_2, \dots, X_n عبارتست از

$$\bar{X}^k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

گشتاور k ام داده های x_1, x_2, \dots, x_n عبارتست از

$$\bar{x}^k = \frac{1}{n} \sum_{i=1}^n x_i^k$$

با فرض $Y = X^k$ و $Y_1 = X_1^k$ و به کمک قانون اعداد بزرگ داریم:

$$Y \xrightarrow{L} E(Y)$$

$$\bar{X}^k \xrightarrow{L} E(X^k)$$

$$\bar{X}^k \xrightarrow{L} \mu^k$$

برای $k=1$ قبلاً نشان دادیم که $\mu \xrightarrow{L} \bar{X}$ (با برابر μ می گیریم).

بنابراین μ_k یعنی گشتاور مرتبه k ام X را با \bar{X}^k برآورد می کنیم. مثالهای زیر نشان می دهند که به

کمک گشتاورهای بالاتر می توان پارامتر یک بعدی یا چند بعدی را برآورد کرد.

مثال ۶ فرض کنید $X \sim N(0, \theta)$. با استفاده از یک نمونه تصادفی برآورد و برآوردیاب گشتاوری

را بیابید.

حل - بدیهی است که

$$\mu_1 = E(X^1) = 0, \quad \mu = E(X) = 0$$

ملاحظه می شود که گشتاور نخست X به θ بستگی ندارد پس با استفاده از گشتاور دوم X برآورد و

برآوردیاب گشتاوری θ می شوند:

$$\bar{\theta} = \bar{X}^2$$

$$\bar{\theta} = \bar{X}^2$$

مثال ۷ فرض کنید $X \sim N(\mu, \sigma^2)$. با استفاده از یک نمونه تصادفی برآورد پارامترها را بیابید.

حل - می دانیم که

$$E(X) = \mu$$

$$E(X^2) = \mu^2 + \sigma^2$$

$E(X)$ و $E(X^2)$ را به ترتیب با \bar{x} و \bar{x}^2 برآورد می کنیم. از حل دو معادله تقریبی

$$\mu = \bar{x}$$

$$\mu^2 + \sigma^2 = \bar{x}^2$$

برآوردهای گشتاوری پارامترهای μ و σ^2 می شوند اعداد

$$\bar{\mu} = \bar{x}$$

$$\bar{\sigma}^2 = \overline{x^2} - \bar{x}^2 = S^2$$

برآوردهای گشتاوری این دو پارامتر عبارتند از متغیرهای تصادفی

$$\bar{\mu} = \bar{X}$$

$$\bar{\sigma}^2 = \overline{X^2} - \bar{X}^2 = S^2$$

در این مثال برای برآورد کردن پارامترها، از نرمال بودن متغیر تصادفی X استفاده نمی شود. در حقیقت نتایج بالا را می توان برای هر متغیر تصادفی X یا میانگین و ولریانس پایایان به کار برد.

۳.۳.۶ روش راستمنائی ماکزیمم

برای اینکه این روش را شرح دهیم، نخست به مثال زیر توجه نمائید:

متغیر تصادفی X دارای توزیع برنولی با پارامتر p است. این پارامتر برابر است با $\frac{1}{4}$ یا $\frac{1}{3}$ ، ولی نمی دانیم کدام. به سختی دیگر فضای پارامتر عبارت است از مجموعه دو عضوی $\{\frac{1}{3}, \frac{1}{4}\}$. نمونه تصادفی X_1, X_2, X_3, X_4 را از X در نظر می گیریم. فرض کنید یافته های این نمونه را به صورت ۱، ۰، ۱، ۱ مشاهده کرده باشیم. می دانیم که در توزیع برنولی $E(X) = p$. بنابراین برآورد گشتاوری p می شود

$$\bar{p} = \frac{1+1+0+1}{4} = \frac{3}{4}$$

عدد $\frac{3}{4}$ یا $\frac{1}{3}$ و $\frac{1}{4}$ تفاوت دارد یعنی این برآورد در فضای پارامتر نیست. پس برآورد گشتاوری در این مثال، نمی تواند مقلد p را تعیین کند. در حقیقت مقداری برای p یافتیم که در فضای پارامتر وجود ندارد. اینک روش زیر را به کار می بریم:

نخست احتمال اینکه در توزیع برنولی، یافته های بالا را مشاهده کنیم به دست می آوریم. چون X ها مستقل و هر یک دارای توزیع برنولی با پارامتر p هستند داریم:

$$P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 1) = p \cdot p \cdot (1-p) \cdot p = p^3(1-p)$$

طبیعی است که مشاهده این داده ها از توزیع برنولی زمانی صدق پیدا می کند که $p^3(1-p)$ بیشترین مقلد را روی مجموعه $(\frac{1}{3}, \frac{1}{4})$ داشته باشد. این احتمال به لری $p = \frac{1}{3}$ برابر $\frac{7}{81}$ و به

ازای $p = \frac{1}{4}$ برابر $\frac{1}{16}$ است. بنابراین $p = \frac{1}{4}$ را به عنوان برآورد پارامتر مجهول بر می گزینیم. تابع $L(p)$ را با $p^3(1-p)$ نشان داده آن را تابع راستمنائی (یا درستمنائی) می نامند. عدد $\frac{1}{4}$ ، که $L(p)$ را روی فضای پارامتر ماکزیمم می کند، برآورد راستمنائی ماکزیمم می خوانند و با $\hat{p} = \frac{1}{4}$ نشان می دهند. این روش را روش راستمنائی ماکزیمم می گویند. حال در حالت کلی روش راستمنائی ماکزیمم را شرح می دهیم.

۴.۳.۶ تابع راستمنائی

نمونه تصادفی X_1, X_2, \dots, X_n را از متغیر تصادفی X با چگالی $f(x; \theta)$ در نظر می گیریم. فرض می کنیم مقدار واقعی پارامتر θ به مجموعه A ، به نام فضای پارامتر، تعلق داشته باشد. یافته های این نمونه تصادفی را با اعداد x_1, x_2, \dots, x_n نشان می دهیم. می دانیم که X_i ها هم توزیع با X و مستقل هستند. بنابراین چگالی هر X_i برابر است با $f(x_i; \theta)$ و چگالی توام X_i ها برابر است با حاصلضرب چگالیهای آنها. این چگالی توام را با فرض اینکه x_i ها ثابت و θ متغیر باشد با $L(\theta)$ نشان می دهند و آن را تابع راستمنائی Likelihood Function برای یافته های نمونه تصادفی می نامند. به طور خلاصه داریم:

$$L(\theta) = f(x_1; \theta) f(x_2; \theta) \dots f(x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

در حالت گسسته $L(\theta)$ دقیقاً برابر است با احتمال مشاهده یافته های بالا بر حسب پارامتر مجهول θ . چه در حالت گسسته و چه در حالت پیوسته، هر اندازه $L(\theta)$ بیشتر باشد شانس مشاهده x_1, x_2, \dots, x_n از چگالی $f(x; \theta)$ بیشتر است. از اینرو مقداری از θ در A را به عنوان برآورد θ بر می گزینیم که $L(\theta)$ را ماکزیمم کند.

۵.۳.۶ برآورد راستمنائی ماکزیمم

مقداری از θ در فضای پارامتر که $L(\theta)$ را ماکزیمم می کند، برآورد راستمنائی ماکزیمم نامیده می شود و آن را با $\hat{\theta}$ نشان می دهند. معمولاً می گویند MLE Maximum Likelihood Estimate پارامتر θ می شود $\hat{\theta}$. به طور خلاصه $\hat{\theta} \in A$ برای θ می شود MLE هرگاه داشته باشیم:

$$L(\hat{\theta}) \geq L(\theta) \quad \forall \theta \in A$$

روش راستمنائی ماکزیمم را فیشر، آماردان انگلیسی، در ۱۹۱۲ ارائه داد. این روش در مقایسه با روش گشتاوری دارای ویژگیهای آماری سودمند می باشد.

مثال ۸ فرض کنید متغیر تصادفی X دارای چگالی زیر باشد:

$$f(x; p) = \begin{cases} p(1-p)^{x-1} & x = 1, 2, \dots \\ 0 & \text{جای دیگر} \end{cases}$$

این چگالی را، چگالی هندسی می نامند زیرا به صورت یک جمله از تصاعد هندسی است. در حقیقت متغیر تصادفی X تعداد آزمایشهای برنولی مستقل است که باید انجام گیرد، تا نخستین پیروزی رخ دهد.

با استفاده از یک نمونه تصادفی، MLE پارامتر $1 > p > 0$ را بیابید.

حل - فرض کنید x_1, x_2, \dots, x_n یافته های یک نمونه تصادفی از X باشند، به طوری که یکی از آنها بیش از عدد ۱ باشد. تابع راستنمایی عبارت است از

$$L(p) = \prod_{i=1}^n f(x_i; p) = p^n (1-p)^{\sum_{i=1}^n x_i - n}$$

لگاریتم $L(p)$ را با $l(p)$ نشان می دهیم و از آن نسبت به p مشتق اول و دوم می گیریم:

$$l(p) = n \ln p + (n\bar{x} - n) \ln(1-p)$$

$$l'(p) = \frac{n}{p} - \frac{n\bar{x} - n}{1-p}$$

$$l''(p) = -\frac{n}{p^2} + \frac{n\bar{x} - n}{(1-p)^2}$$

می توان نشان داد که پاسخ معادله $l'(p) = 0$ یعنی $\frac{1}{x}$ مشتق دوم را منفی می کند. بنابراین MLE پارامتر p می شود

$$\hat{p} = \frac{1}{x}$$

و برآوردیاب متناظر با آن، متغیر تصادفی

$$\hat{p} = \frac{1}{\bar{X}}$$

است. به طریق بینشی (شهودی) می توان این برآورد را به شرح زیر پیدا کرد: در $x_1 + x_2 + \dots + x_n$ آزمایش برنولی مستقل، n پیروزی رخ می دهد. پس برآورد p برابر است با فراوانی نسبی پیروزیها، یعنی عدد

$$\frac{n}{x_1 + x_2 + \dots + x_n} = \frac{1}{\bar{X}}$$

اگر تمام x_i ها برابر عدد ۱ باشند، آنگاه داریم:

$$L(p) = p^n$$

ماکزیمم این تابع در فاصله بسته $[0, 1]$ برابر است با عدد ۱ و $\hat{p} = 1$.

مثال ۹ برای چگالی مثال ۵، MLE پارامتر θ را با استفاده از یک نمونه تصادفی بیابید.

حل - فرض کنید x_1, x_2, \dots, x_n یافته های یک نمونه تصادفی از X باشد. تابع راستنمایی عبارت است از

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta) = \theta^n e^{-\theta n\bar{x}} \quad \theta > 0$$

این تابع را روی فضای پارامتر یعنی روی فاصله $(0, \infty)$ ماکزیمم می کنیم. لگاریتم $L(\theta)$ را با $l(\theta)$ نشان می دهیم و از آن نسبت به θ مشتق اول و دوم می گیریم:

$$l(\theta) = n \ln \theta - n\theta\bar{x}$$

$$l'(\theta) = \frac{n}{\theta} - n\bar{x}$$

$$l''(\theta) = -\frac{n}{\theta^2}$$

پاسخ معادله $l'(\theta) = 0$ یعنی $\frac{1}{x}$ مشتق دوم را منفی می کند. بنابراین MLE پارامتر θ می شود

$$\hat{\theta} = \frac{1}{x}$$

و برآوردیاب متناظر با آن، متغیر تصادفی

$$\hat{\theta} = \frac{1}{\bar{X}}$$

است. با توجه به مثال ۵، نتیجه می گیریم که MLE و برآورد گشتاوری برابرند. ناگفته نماند که این نتیجه همواره برقرار نیست.

مثال ۱۰ متغیر تصادفی X دارای توزیع برنولی با پارامتر $p \in \left\{ \frac{1}{4}, \frac{1}{3}, \frac{1}{2} \right\}$ است. یافته های یک نمونه تصادفی از X عبارتند از ۱، ۱، ۰، ۱، ۰، ۱، ۰، ۱.

با استفاده از این یافته ها، MLE پارامتر p را بیابید.

حل - تابع راستمنامی عبارت است از

$$L(p) = P(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 1, X_5 = 0) = p^4(1-p)^2$$

برای ماکزیمم کردن $L(p)$ از روی فضای پارامتر گسسته $(\frac{1}{4}, \frac{1}{3}, \frac{1}{2})$ نمی توان مشتق $L(p)$ را به کار برد. مستقیماً داریم

$$L(\frac{1}{4}) = \frac{1}{256}$$

$$L(\frac{1}{3}) = \frac{4}{243}$$

$$L(\frac{1}{2}) = \frac{9}{1024}$$

بنابراین MLE پارامتر p می شود عدد $\hat{p} = \frac{1}{4}$ ، زیرا $L(\frac{1}{4})$ بیشترین مقدار را روی فضای پارامتر دارد.

۶.۳.۶ تمرین بخش سه

۱. فرض کنید X یک عدد تصادفی در فاصله (a, b) باشد. برآورد گشتاوری a و b را پیدا کنید. از نمونه $1/12, 1/14, 1/15, 1/23, 1/27, 1/33, 1/37, 1/43, 1/47, 1/53, 1/57, 1/67, 1/73, 1/77, 1/83, 1/87, 1/97, 1/103, 1/107, 1/113, 1/127, 1/137, 1/147, 1/157, 1/167, 1/177, 1/187, 1/197$ استفاده کنید. آیا با روش دیگر هم می توانید a و b را برآورد کنید؟

۲. فرض کنید $X \sim N(\mu, 1)$. برآورد گشتاوری و MLE پارامتر μ را پیدا کنید.

۳. فرض کنید X دارای توزیع برنولی با پارامتر p باشد. برآورد p را با استفاده از یک نمونه تصادفی از دو راه بیابید.

۴. فرض کنید $X \sim N(0, \theta)$. MLE پارامتر θ را بیابید.

۵. برآورد یاب U را برای پارامتر θ ناربب می نامیم هرگاه $E(U) = \theta$. ثابت کنید در توزیع $N(\mu, 1)$ آماره X یک برآورد ناربب برای μ است.

۶. فرض کنید $X \sim P(\lambda)$. برآورد گشتاوری و MLE پارامتر λ را پیدا کنید.

۷. فرض کنید U_1 و U_2 دو برآورد یاب ناربب و مستقل با ولریانسهای ۲ و ۳ برای پارامتر θ باشند. ضرایب a_1 و a_2 را طوری پیدا کنید تا

$$T = a_1 U_1 + a_2 U_2$$

بتواند یک برآورد یاب ناربب با کمترین ولریانس برای θ باشد. حال ولریانسهای U_1 و U_2 و T را با هم مقایسه نمایید. کدام از این سه برآورد یاب، ناربب بهتر می باشد؟ (از تمرین ۵ استفاده کنید)

۸. متغیر تصادفی X دارای توزیع برنولی با پارامتر p است. فضای پارامتر عبارت است از

$$A = (\frac{1}{4}, \frac{1}{3}, \frac{1}{2})$$

MLE پارامتر p را بیابید

۹. اگر U_1 و U_2 دو برآورد یاب ناربب و مستقل برای پارامتر θ باشند، یک برآورد یاب ناربب برای θ^2 و برای $\theta(1-\theta)$ بر حسب U_1 و U_2 بیابید

۱۰. شانس شیر آمدن با یک ۵ ریالی p_1 و با یک ۱۰ ریالی p_2 می باشد چگونه یک برآورد یاب ناربب برای $p_2 - p_1$ پیدا می کنید؟

۴.۶ فاصله اطمینان برای یک پارامتر

در بخش پیش دیدیم که چگونه می توان برای پارامتر θ یک مقدر تقریبی به نام برآورد پیدا کرد. در این بخش می خواهیم فاصله‌های پیدا کنیم که θ بتواند با اطمینان زیاد در آن باشد. برای اینکه مفهوم چنین فاصله اطمینان را شرح دهیم، نخست به مثال زیر می پردازیم.

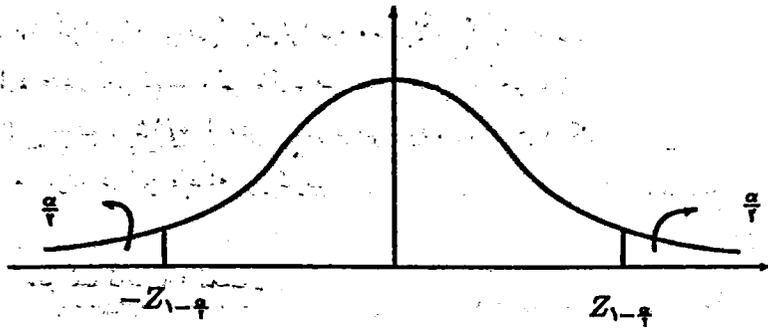
مثال ۱۱ یک نمونه تصادفی ۱۰۰ تایی از $X \sim N(\mu, 4)$ را در نظر می گیریم. احتمال اینکه پارامتر مجهول μ در فاصله تصادفی $(\bar{X} - 0.4, \bar{X} + 0.4)$ باشد چقدر است؟

حل - می دانیم که $\bar{X} \sim N(\mu, 0.4)$ و $\frac{(\bar{X} - \mu)}{0.2} \sim N(0, 1)$. پس داریم:

$$P(\bar{X} - 0.4 < \mu < \bar{X} + 0.4) = P(-2 < \frac{\bar{X} - \mu}{0.2} < 2)$$

$$= P(-2 < Z < 2) = \Phi(2) - \Phi(-2) = 0.95$$

فاصله $(\bar{X} - 0.4, \bar{X} + 0.4)$ را یک فاصله تصادفی می نامند زیرا دو سر آن متغیر تصادفی می باشند. این فاصله تصادفی پارامتر μ را با احتمال ۰/۹۵ در بر می گیرد و اصطلاحاً آن را یک فاصله اطمینان برای μ با ضریب ۰/۹۵ می گویند. معنی جمله آخر این است که تقریباً ۹۵ درصد از فواصل عددی که از این راه به دست می آیند، می توانند پارامتر μ را در برگیرند. مثلاً، اگر یافته \bar{X} بشود $4.7 = \bar{X}$ ، یافته فاصله تصادفی بالا می شود $(2.8, 6.6)$. این فاصله عددی از جمله فواصل عددی می باشد که



نگاره ۱ منحنی نرمال استاندارد

حل - می دانیم که \bar{X} برآوردیاب μ می باشد و داریم

$$Q = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1) \quad \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

پس می توانیم Q را، که توزیع آن به μ بستگی ندارد، به عنوان تابع محوری بگیریم. حال با توجه به نگاره ۱ داریم:

$$P\left(-z_{1-\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha = \gamma$$

از حل نامساوی داخل پرانتز بر حسب μ فاصله اطمینان زیر به دست می آید:

$$\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, \bar{X} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\right)$$

با داشتن α می توانیم عدد $z_{1-\frac{\alpha}{2}}$ را از جدول (۳) پیدا کنیم. جدول زیر برای پیدا کردن عدد $z_{1-\frac{\alpha}{2}}$ در دو انتهای چگالی نرمال برای بعضی مقادیر γ مفید می باشد.

| | | | | |
|--------------------------|--------|--------|--------|--------|
| γ | ۰/۸۰ | ۰/۹۰ | ۰/۹۵ | ۰/۹۹ |
| $\frac{\alpha}{2}$ | ۰/۱۰ | ۰/۰۵ | ۰/۰۲۵ | ۰/۰۰۵ |
| $z_{1-\frac{\alpha}{2}}$ | ۱/۲۸۱۶ | ۱/۶۲۲۹ | ۰/۹۶۰۰ | ۲/۵۷۵۸ |

جدول ضریب اطمینان و مقدار $z_{1-\frac{\alpha}{2}}$

تقریباً ۹۵ درصد آنها μ را در بر می گیرند. از اینرو معمولاً می گویند فاصله عددی $(2/8, 3/6)$ یک فاصله اطمینان ۹۵ درصدی می باشد. ولی تاکید می کنیم که جمله آخر به این معنی نیست که فاصله عددی $(2/8, 3/6)$ ، با احتمال $0/95$ پارامتر μ را در بر می گیرد، زیرا μ خود یک عدد می باشد و تنها با احتمال صفر یا یک می تواند در این فاصله قرار گیرد.

۱.۴.۶ تعریف فاصله اطمینان

یک نمونه تصادفی از چگالی $f(x; \theta)$ را در نظر می گیریم و فرض می کنیم آماره U برآوردیاب پارامتر θ باشد. فاصله تصادفی

$$(g_1(U), g_2(U))$$

را که دو سر آن توابعی از آماره U می باشند در نظر می گیریم. اگر داشته باشیم

$$P(g_1(U) < \theta < g_2(U)) = \gamma$$

به طوری که γ به پارامتر θ بستگی نداشته باشد و صرفاً یک عدد باشد، آنگاه می گوئیم این فاصله تصادفی یک فاصله اطمینان با ضریب γ برای پارامتر θ می باشد. هر اندازه γ به یک نزدیک تر و $E(g_1(U) - g_2(U))$ ، یعنی میانگین طول فاصله، کوچکتر باشد، فاصله یاد شده بهتر است.

اگر در فاصله تصادفی یاد شده به جای U یافته آن μ را قرار دهیم، فاصله عددی $(g_1(\mu), g_2(\mu))$ می شود یافته فاصله اطمینان بالا.

۲.۴.۶ طرز پیدا کردن فاصله اطمینان

نخست متغیر تصادفی $Q = q(U, \theta)$ را، که به صورت تابعی از آماره U و پارامتر θ می باشد، به نحوی می سازند که توزیع آن به θ بستگی نداشته باشد. Q را معمولاً تابع محوری می گویند. سپس دو عدد ثابت $a < b$ را چنان انتخاب می کنند تا داشته باشیم:

$$P(a < q(U, \theta) < b) = \gamma$$

از حل نامساوی داخل پرانتز بالا بر حسب θ یک فاصله اطمینان با ضریب γ برای θ پیدا می شوند.

مثال ۱۲ فرض کنید $X \sim N(\mu, \sigma^2)$ و σ^2 معلوم و μ پارامتر مجهول باشد. با استفاده از یک نمونه تصادفی n تایی یک فاصله اطمینان با ضریب γ برای μ بسازید.

مثال ۱۳. آهن موجود در یک کیلوگرم سنگ آهن متغیر تصادفی $X \sim N(\mu, 100)$ بر حسب میلی‌گرم می‌باشد. برای یک نمونه تصادفی ۲۵ تایی یافته \bar{X} عدد ۱۵ شده است یک فاصله اطمینان با ضریب ۰/۹۰ برای μ پیدا کنید.

حل - می‌دانیم که فاصله اطمینان به صورت زیر است:

$$\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}, \bar{X} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \right)$$

برای اینکه یافته این فاصله اطمینان را پیدا کنیم از داده‌های زیر استفاده می‌کنیم:

$\sigma^2 = 100, \sigma = 10, n = 25, \bar{x} = 15, \gamma = 0/90, \frac{\alpha}{2} = 0/05$
پس یافته فاصله اطمینان می‌شود فاصله عددی $(11/71, 18/29)$. این یکی از فواصل اطمینان ۹۰ درصدی است که ممکن است μ را در برداشته باشد. اگر در مثال بالا σ^2 مجهول باشد، باید به جای σ^2 برآورد آنرا بگذاریم. البته این کار موجب تغییر فرمول بالا می‌شود. ما روش برآورد کردن σ^2 و روش پیدا کردن فاصله اطمینان را در این حالت، در بخش ۶ از این فصل شرح می‌دهیم.
۳.۴.۶ تمرین بخش چهارم

۱. یک نمونه تصادفی ۱۰۰ تایی از $X \sim N(\mu, 4)$ را در نظر می‌گیریم. احتمال اینکه پارامتر مجهول μ در فاصله $(\bar{X} - 0/3, \bar{X} + 0/3)$ باشد چیست؟ اگر برای \bar{X} عدد ۵ را مشاهده کنیم، منظور از یافته این فاصله اطمینان را شرح دهید.

۲. طول عمر یک نوع باتری اتومبیل متغیر تصادفی $X \sim N(\mu, 25)$ بر حسب ماه است. \bar{X} طول عمر ۹ باتری عبارتند از:

۲۱, ۲۴, ۱۸, ۱۷, ۱۹, ۲۶, ۲۸, ۲۳, ۲۰

یک فاصله اطمینان ۹۰ درصدی برای پارامتر μ به دست آورید.

۳. نمره نهائی درس کلکولس در یک کلاس پر دانشجو متغیر تصادفی $X \sim N(\mu, 64)$ می‌باشد. برای یک نمونه ۳۶ تایی از این دانشجویان داریم $\bar{x} = 72$. یک فاصله اطمینان ۹۸ درصدی برای پارامتر μ پیدا کنید.

۴. فرض کنید $X \sim N(0, \sigma^2)$. احتمال اینکه σ در فاصله تصادفی $(|X|, 2|X|)$ باشد چیست؟
و b را طوری پیدا کنید تا $(a|X|, b|X|)$ یک فاصله اطمینان هشاد درصدی برای σ باشد اندازه نمونه را ۴ بگیرد.

۵. متغیر تصادفی X طول عمر یک لامپ می‌باشد. X دارای چگالی زیر است:

$$f(x; \theta) = \begin{cases} \frac{e^{-x/\theta}}{\theta} & x > 0, \theta > 0 \\ 0 & \text{جای دیگر} \end{cases}$$

۶. c را طوری پیدا کنید تا فاصله تصادفی $(0, cX)$ یک فاصله اطمینان نود درصدی برای θ باشد

۵.۶ آزمون آماری یک پارامتر

متغیر تصادفی X را که توزیع آن به پارامتر θ بستگی دارد در نظر می‌گیریم. ممکن است مقدار واقعی θ ، به علت تغییر شرایط تغییر نماید. ادعای هر نوع تغییر را باید از راه مشاهده، یعنی به کمک داده‌ها، تأیید یا انکار نمود. به زبان آماری این کار را آزمون آماری پارامتر θ می‌نامیم.

مثلاً فرض کنید $\theta = 0/8$ نسبت بیمارانی باشد که آنتی‌بیوتیک A آنها را بهبود می‌بخشد. ممکن است در طول زمان ویروس بیماری در برابر این دارو مقاوم شده باشد و بخواهیم نشان دهیم که اخیراً $\theta = 0/6$. اثبات آماری این امر، به کمک یک آزمون آماری درباره پارامتر متغیر تصادفی برنولی انجام می‌گیرد.

در این بخش نخست اصطلاحات آزمون آماری را به یاری مثال زیر شرح می‌دهیم.

سکه‌ای داریم که طبق معمول آنرا ناریب فرض می‌کنیم. این فرض را با $H_0: p = \frac{1}{4}$ نشان می‌دهیم و H_1 را فرض صفر می‌نامیم. کلمه صفر را به این دلیل به کار می‌بریم که در مورد سکه معمولی هر نوع تغییری را از نظر ناریبی ناچیز تصور می‌کنیم. شخصی که خود را درباره ضرب سکه‌ها مطلع می‌داند، می‌گوید اینگونه سکه‌ها اریب می‌باشند و باید فرض کرد که $p = \frac{1}{4}$. این فرض جدید را با $H_1: p = \frac{1}{4}$ نشان می‌دهیم و H_0 را فرض مقابل (در مقابل فرض H_0) می‌گوئیم فرضهای H_0 و H_1 را فرضهای ساده می‌نامیم. ولی فرضهائی مانند $H_0: p > \frac{1}{4}$ یا $H_1: p < \frac{1}{4}$ را فرضهای مرکب می‌گویند. در این بخش ما تنها درباره فرضهای ساده صحبت می‌کنیم. بنابراین با یک مسأله آماری به صورت زیر روبرو هستیم:

$$\begin{cases} H_0: p = \frac{1}{4} \\ H_1: p = \frac{1}{4} \end{cases}$$

خطاها را به ترتیب با α و β نشان می‌دهیم و آنها را خطای نوع اول و خطای نوع دوم می‌نامیم. بنابراین داریم:

$$\alpha = P(\text{خطای نوع اول} | H_0 \text{ درست است}) = P(H_1 | H_0 \text{ درست است})$$

$$\beta = P(\text{خطای نوع دوم} | H_1 \text{ درست است}) = P(H_0 | H_1 \text{ درست است})$$

α را اغلب میزان آزمون یا آستانه آزمون یا سطح آزمون می‌نامند و معمولاً آنرا 0.10 یا 0.05 یا 0.01 می‌گیرند. دلیل کوچک گرفتن α این است که نمی‌خواهند وضع موجود یعنی H_0 را به ناحق تغییر دهند و با خسارت زیاده‌روی شوند. در مثال بالا، متغیر تصادفی S که بر اساس آن تصمیم‌گیری می‌کنیم دارای توزیع $B(10, p)$ است. اگر H_0 درست باشد داریم $p = \frac{1}{4}$ و اگر H_1 درست باشد داریم $p = \frac{1}{3}$. پس، از جدول توزیع دو جمله‌ای داریم:

$$\alpha = P(S \in C | p = \frac{1}{4}) = P(S \leq 2 | p = \frac{1}{4})$$

$$= B(10, \frac{1}{4}; 2) = 0.0547$$

$$\beta = P(S \in \bar{C} | p = \frac{1}{3}) = P(S > 2 | p = \frac{1}{3})$$

$$= 1 - B(10, \frac{1}{3}; 2) = 0.4744$$

ملاحظه می‌شود که خطای نوع دوم زیاد است، زیرا بر پایه ناحیه بحرانی مثلاً اگر یافته S بشود ۹ باز هم H_0 را درست می‌گیریم که معقول نمی‌باشد. در این حال حتماً باید گفت که فرض $p > \frac{1}{4}$ در H_1 درست است.

۳.۵.۶ توان آزمون

احتمال رد H_0 در صورتیکه H_1 درست باشد، یعنی احتمال رد H_0 را به حق، توان آزمون می‌نامند و با π نشان می‌دهند. بنابراین توان آزمون عبارت است از:

$$\pi = P(H_1 \text{ درست است} | H_1 \text{ درست است}) = 1 - P(H_0 \text{ درست است} | H_1 \text{ درست است}) = 1 - \beta$$

ملاحظه می‌شود که جمع خطای نوع دوم و توان آزمون برابر عدد ۱ است. در مثال بالا داریم $\pi = 0.5256$.

ناحیه بحرانی

| | | | | | |
|---|---|---|---|---|----|
| ۰ | ۱ | ۲ | ۳ | ۴ | ۱۰ |
|---|---|---|---|---|----|

ناحیه رد H_0

ناحیه رد H_1

نگاره ۲ ناحیه رد H_0 و H_1

طبیعت این مسأله آماری با یک مسأله ریاضی تفاوت دارد. تنها از راه مشاهده و قرارداد می‌توان در مورد درستی یا نادرستی H_0 تصمیم گرفت. از اینرو چنین مسأله‌ای را یک آزمون آماری و نتیجه تصمیم‌گیری درباره درستی H_0 یا H_1 را انجام یک آزمون آماری می‌نامیم.

۱.۵.۶ ناحیه بحرانی

چگونه آزمون بالا را انجام دهیم؟ سکه را مستقلاً چندین بار مثلاً ۱۰ بار پرتاب می‌کنیم و متغیر تصادفی S را تعداد شیرها در این ۱۰ پرتاب می‌گیریم. در صورتی که سکه ناریب باشد، یعنی H_0 درست باشد، انتظار داریم که یافته S بشود ۵. طبیعی است که اگر S خیلی کوچک باشد، نمی‌توان H_0 را تأیید کرد. بنابراین قرارداد می‌کنیم که اگر $S \leq 2$ یعنی اگر یافته S در مجموعه $C = \{0, 1, 2\}$ بیافتد، H_0 را رد کنیم. مجموعه C را ناحیه بحرانی می‌نامیم زیرا با رد H_0 وضعی را که تا بحال مورد قبول بوده است دیگر نمی‌پذیریم. اما اگر S در مجموعه $\bar{C} = \{3, 4, \dots, 10\}$ بیافتد، دلیلی نداریم که H_0 را رد کنیم. این دو ناحیه را در نگاره ۲ نشان داده‌ایم.

آماره S را که آزمون بر پایه آن انجام می‌گیرد، آماره آزمون می‌نامند.

۲.۵.۶ خطاها

آیا قضاوت بالا بدون خطا می‌باشد؟ خیر، چون ممکن است با یک سکه ناریب ($p = \frac{1}{4}$) در ده پرتاب مثلاً دوبار شیر مشاهده کرد ولی قرارداد موجب می‌شود که H_0 را به ناحق رد کنیم. یا ممکن است با یک سکه اریب ($p = \frac{1}{3}$) در ده پرتاب مثلاً ۴ بار شیر مشاهده کرد، ولی قرارداد موجب می‌شود که H_1 را به ناحق رد کنیم. بنابراین مرتکب دو نوع خطا می‌شویم. قضاوت بالا بر اساس آن قرارداد در صورتی معقول می‌باشد که احتمال مرتکب شدن این خطاها تا آنجا که ممکن است کوچک باشد. احتمال این

مثال ۱۳- فرض کنید $X \sim N(\mu, 4)$ خطای اندازه گیری با یک دستگاه فیزیکی بر حسب میکرو باشد. برای آزمون زیر:

$$H_0: \mu = 0 \quad (S < 7) \\ H_1: \mu = 1 \quad (S \geq 7)$$

یک نمونه تصادفی ۲۵ تایی از X را در نظر می گیریم. اگر $\bar{X} \geq 7/4$ را به عنوان ناحیه بحرانی بگیریم خطاها و توان آزمون را حساب کنید.

حل - P_{H_0} و P_{H_1} را به ترتیب احتمال تحت فرض H_0 و H_1 می گیریم. حال داریم:

$$\alpha = P(H_0 \text{ درست} | H_0 \text{ وارد می کنیم}) = P_{H_0}(\bar{X} > 7/4)$$

$$= P_{H_0}\left(\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} > \frac{7/4 - 0}{\frac{2}{5}}\right) = P(Z > 1) = 1 - \Phi(1) = 0.2420$$

توان آزمون $\pi = 1 - \beta = P(H_1 \text{ درست} | H_1 \text{ وارد می کنیم}) = P_{H_1}(\bar{X} \leq 7/4)$

$$= P_{H_1}\left(\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} > \frac{7/4 - 1}{\frac{2}{5}}\right) P(Z \leq -1/5)$$

$$= \Phi(-1/5) = 1 - \Phi(1/5) = 0.668$$

توان آزمون می شود $\pi = 1 - \beta = 0.9332$

$$\pi = 1 - \beta = 0.9332$$

دلیل آنکه ناحیه بحرانی را به صورت $\bar{X} \geq 7/4$ گزیده ایم این است که \bar{X} برآورد μ می باشد پس اگر \bar{X} از صفر بیش از حد بزرگتر شود، نسبت به فرض H_0 تردید می کنیم.

مثال ۱۴- در آزمون مثال ۱۳ می خواهیم خطای نوع اول بشود $\alpha = 0.10$. ناحیه بحرانی را پیدا کنید.

حل - ناحیه بحرانی به صورت $\bar{X} > d$ است. باید d را طوری تعیین کنیم تا داشته باشیم $\alpha = 0.10$ می نویسیم:

$$\alpha = P(H_0 \text{ درست} | H_0 \text{ وارد می کنیم}) = P_{H_0}(\bar{X} > d) \\ = P_{H_0}\left(\frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} > \frac{d - 0}{\frac{2}{5}}\right) = P(Z > 2.5d) = 1 - \Phi(2.5d)$$

$$\Phi(2.5d) = 0.90$$

$$2.5d = 1.28$$

$$d = 0.512$$

۴.۵.۶ تمرین بخش پنج

۱. می خواهیم برای یک سکه آزمون زیر را انجام دهیم:

$$\begin{cases} H_0: p = \frac{1}{4} \\ H_1: p = \frac{3}{4} \end{cases}$$

سکه را ده بار پرتاب می کنیم و S را تعداد شیرها در این ده پرتاب می گیریم. اگر $S \geq 8$ فرض H_0 را در می کنیم.

الف - خطای نوع اول و دوم و توان آزمون را حساب کنید

ب - اگر $S \geq d$ را ناحیه بحرانی بگیریم، d را طوری تعیین کنید تا خطای نوع اول بیش از 0.09 نباشد.

۲. در تمرین ۲ از بخش چهار، تولید کننده باتری مدعی است که اخیراً طول عمر باتری به سبب ترکیبات جدید زیاد شده است. می خواهیم آزمون زیر را انجام دهیم:

$$\begin{cases} H_0: \mu = 20 \\ H_1: \mu = 24 \end{cases}$$

اگر $\bar{X} > 25$ ناحیه بحرانی باشد، خطای نوع اول و دوم و توان آزمون را حساب کنید بر اساس داده ها فرض H_0 وارد می کنید؟

۳. در ظرفی سر بسته ۷ مهره می باشد که ۵ از آنها سفید و ۲ بقیه سیاه هستند. می خواهیم آزمون زیر را انجام دهیم

$$\begin{cases} H_0: \theta = 2 \\ H_1: \theta = 7 \end{cases}$$

شخصی دو مهره با هم از ظرف بیرون می آورد و به ما نشان می دهد فرض H_0 را رد می کنیم اگر هر دو سفید باشند. خطای نوع اول و دوم و توان را حساب کنید

۲. در تمرین ۵ از بخش ۳ می خواهیم آزمون زیر را انجام دهیم

$$\begin{cases} H_0: \theta = 200 \\ H_1: \theta = 500 \end{cases}$$

اگر $X > 300$ مشاهده شود، فرض H_0 را رد می کنیم. خطای نوع اول و دوم را حساب کنید
۵. آموزگاری عقیده دارد که دانش آموزان به اعداد ۱۷، ۱۸، ۱۹، ۲۰ از میان ۱ تا ۲۰ بیشتر علاقه دارند. از ۲۴ دانش آموز می خواهد که یکی از اعداد ۱ تا ۲۰ را انتخاب کنند. فرض کنید X تعداد دانش آموزانی باشد که یکی از اعداد ۱۷، ۱۸، ۱۹، ۲۰ را انتخاب کردند.
الف - توزیع احتمال X را پیدا کنید، در صورتی که دانش آموزان به تصادف عدد را انتخاب کنند.
ب - متغیر تصادفی X به چه ویژگی باید باشد تا عقیده آموزگار با خطای نوع اول پنج درصد تأیید گردد؟

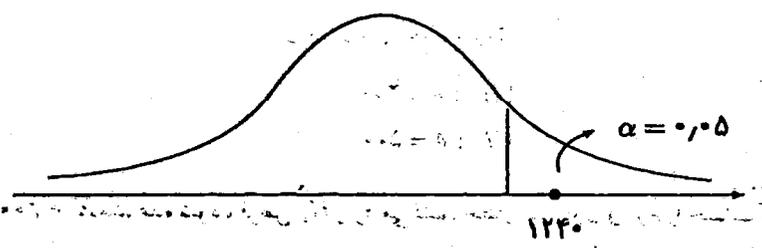
۶۶ - مقدار p

مدتی است که در بسیاری از کتابهای آماری و نرم افزارهای آماری، حتی در ماشینهای حساب، اصطلاحی به نام p -value مشاهده می شود که ما آنرا p -مقدار یا مقدار احتمال می نامیم. فیشر، آماردان انگلیسی، آن را P می نامد. لهن، آماردان دانشگاه برکلی آمریکا، به آن احتمال با معنی می گوید. اغلب مفهوم p -مقدار برای غیر آماردانان روشن نیست، با اینکه آن را به کار می برند. نخست این مفهوم را با مثال زیر شرح می دهیم.

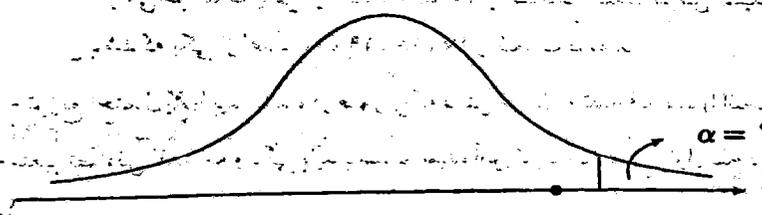
لامپهایی را که یک کارخانه تولید می کند دارای طول عمر نرمال با میانگین $\mu = 1200$ ساعت و انحراف معیار $\sigma = 200$ ساعت می باشند. صاحب کارخانه مدعی است که با ماشینهای جدید، میانگین

طول عمر افزایش یافته است. معدل طول عمر ۱۰۰ لامپ جدید ۱۲۲۰ ساعت شده است. می خواهیم با میزان $\alpha = 0.05$ این ادعا را بیازمائیم. به سخنی دیگر می خواهیم آزمون زیر را انجام دهیم:

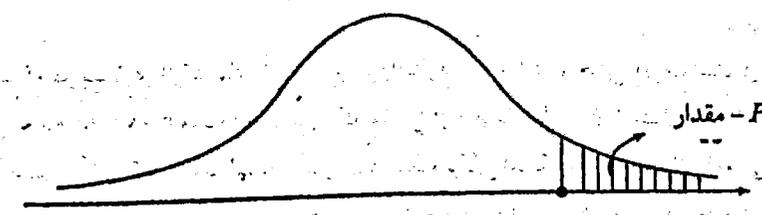
$$\begin{cases} H_0: \mu = 1200 \\ H_1: \mu > 1200 \end{cases}$$



ناحیه بحرانی با $\alpha = 0.05$



ناحیه بحرانی با $\alpha = 0.01$



P -مقدار

نگاره ۳

می دانیم که برآورده یاب پارامتر μ کماتر X است. اگر یافته این آماره یعنی \bar{X} خیلی بزرگ شود، فرض H_0 را رد می کنیم. بنابراین ناحیه بحرانی به صورت $c > \bar{X}$ است. عدد c به میزان آزمون یعنی $\alpha = 0.05$ بستگی دارد و به طریق زیر آن را می یابیم:

$$\alpha = P(H_0 \text{ رد} | H_0 \text{ درست})$$

$$0.05 = P(\bar{X} > c | H_0 \text{ درست})$$

آماره \bar{X} دارای توزیع $N(\mu, \frac{\sigma^2}{n})$ است پس تحت فرض H_0 داریم $\bar{X} \sim N(1200, 400)$ اینک می نویسیم:

$$0.05 = P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \geq \frac{c - 1200}{20}\right)$$

$$= P(Z > \frac{c - 1200}{20})$$

$$= 1 - \Phi\left(\frac{c - 1200}{20}\right)$$

سرانجام با استفاده از جدول ۳ داریم

$$\Phi\left(\frac{c - 1200}{20}\right) = 0.95$$

$$\frac{c - 1200}{20} = 1.65$$

$$c = 1233$$

بنابراین ناحیه بحرانی با میزان $\alpha = 0.05$ می شود $\bar{X} > 1233$. چون یافته \bar{X} یعنی عدد $\bar{X} = 1240$ ناحیه بحرانی می افتد، پس فرض H_0 را با میزان $\alpha = 0.05$ رد می کنیم و فرض H_1 را می پذیریم. البته اگر α را برابر 0.01 بگیریم، با محاسبه ای مانند بالا، ناحیه بحرانی به صورت $\bar{X} > 1246$ می آید. در این حالت $\bar{X} = 1240$ در ناحیه بحرانی قرار نمی گیرد و در نتیجه فرض H_0 با میزان $\alpha = 0.01$ رد نمی شود. ملاحظه می شود که رد H_0 بستگی به اندازه α دارد. معمولاً با پیروی از سنتی که فشر بنا گذارده است، α را عددی کوچک مانند 0.01 یا 0.05 یا 0.10 می گیرند تا H_0 را با کمال احتیاط رد کنند. اندازه α به نظر پژوهشگر بستگی دارد. کوچکترین مقداری که برای α می توان در نظر گرفت تا H_0 رد شود، برابر است با

مساحت هاشور خورده ای در دم چگالی X که از $\bar{X} = 1240$ شروع می گردد. این احتمال در دم را p - مقدار می گویند. در مثال بالا داریم:

$$p = P(\bar{X} \geq 1240 | H_0 \text{ درست}) = 1 - \Phi(z) = 0.0228$$

نگاره ۳ ناحیه بحران را برای $\alpha = 0.05$ و $\alpha = 0.01$ - مقدار را نشان می دهد. چنین احتمالی می گویند، با فرض درستی H_0 یعنی $\mu = 1200$ ، ممکن است فقط در ۲ بار از ۱۰۰ بار \bar{X} حداقل برابر 1240 شود. حال پژوهشگر برای خود α را آزادانه انتخاب می کند. (در صورتی که رد H_0 برای او زیانبار باشد، سعی می کند α را کوچک بگیرد.) اگر p - مقدار از این α که پژوهشگر با میل خود انتخاب کرده است کمتر باشد، فرض H_0 رد می شود.

مثلاً با $\alpha = 0.01$ فرض H_0 رد نمی شود ولی با $\alpha = 0.032$ رد می شود. با این روش، نیازی به یافتن ناحیه بحرانی نیست. اینک با استفاده از بحث بالا، p - مقدار را در حالت کلی تعریف می کنیم و روش محاسبه آن را بیان می داریم. فرض کنید متغیر تصادفی X دارای چگالی $f(x; \theta)$ باشد و بخواهیم آزمونی درباره پارامتر θ انجام دهیم. نخست نمونه تصادفی X_1, X_2, \dots, X_n را از X با یافته های x_1, x_2, \dots, x_n در نظر می گیریم. حال آماره ای مناسب مانند U ، که به صورت تابع $U = g(X_1, \dots, X_n)$ است، در نظر می گیریم. این آماره می تواند برآوردهایی برای θ باشد، مانند \bar{X} در مثال بالا که برآوردهای پارامتر μ است. به U آماره آزمون می گویند و یافته آن را با $u = g(x_1, \dots, x_n)$ نشان می دهند. افزون بر این، U را به نحوی بر می گزینیم که بتوان تابع توزیع یا تابع تقریبی آن را، تحت فرض H_0 ، پیدا کرد.

۱.۶.۶ تعریف p - مقدار

کمترین مقداری از α (یعنی میزان یا سطح آزمون) است، که یافته آماره آزمون ممکن است موجب رد فرض H_0 گردد.

۲.۶۶ طرز محاسبه p - مقدار

فرض کنید U آماره آزمون و μ یافته آن باشد معمولاً سه نوع آزمون به کار می رود:
الف - آزمون یکسویی راست به صورت

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$$

در این آزمون داریم:

$$p = P_{\theta_0}(U \geq u)$$

ب - آزمون یکسویی چپ به صورت

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$$

در این آزمون داریم:

$$p = P_{\theta_0}(U \leq u)$$

ج - آزمون دوسویی به صورت

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$$

در این آزمون داریم:

$$p = \gamma \min \{ P_{\theta_0}(U \leq u), P_{\theta_0}(U \geq u) \}$$

نماد $P_{\theta_0}(\cdot)$ یا $P(\cdot | \theta_0)$ احتمال با فرض درست بودن H_0 و نماد \min به معنی کمترین است.

مثال ۱۵ - تولید کننده باتریهای ۶ ولتی می گوید طول عمر آنها دارای توزیع $N(55, 100)$ است. معدل طول عمر ۲۵ عدد از این باتریها ۵۰ ساعت بوده است. آیا گفته تولید کننده را با میزان $\alpha = 0.05$ می پذیرید؟

حل - تولید کننده می گوید $\mu = 55$ ولی عملاً برآورده این پارامتر $\bar{x} = 50$ شده است. بنابراین آزمون زیر را برای توزیع نرمال انجام می دهیم.

$$\begin{cases} H_0 : \mu = 55 \\ H_1 : \mu < 55 \end{cases}$$

با فرض اینکه H_0 درست باشد، داریم $\bar{X} \sim N(55, 4)$. حال p - مقدار را با روش زیر می یابیم:

$$\begin{aligned} p = P(X \leq 50) &= P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{50 - 55}{2}\right) \\ &= P(Z \leq -2.5) \\ &= 0.0062 \end{aligned}$$

چون p - مقدار از 0.05 کمتر است، فرض H_0 را با میزان 0.05 حتی 0.10 رد می کنیم.

مثال ۱۶ - سکه ای را، که شانس شیر آمدن برای آن پارامتر p است، ۱۰ بار مستقلاً خنثی اندازیم و ۲ بار شیر می بینیم. آزمون زیر را با میزان $\alpha = 0.05$ انجام دهید.

$$\begin{cases} H_0 : p = \frac{1}{4} \\ H_1 : p = \frac{1}{4} \end{cases}$$

حل - فرض کنید S مجموع شیرها در ۱۰ پرتاب مستقل باشد. می دانیم که تحت فرض H_0 داریم $S \sim B(10, \frac{1}{4})$. p - مقدار را می یابیم:

$$p = P(S \leq 2) = B(10, \frac{1}{4}; 2) = 0.0547$$

چون p - مقدار از 0.05 بیشتر است، H_0 را رد نمی کنیم. به سخی دیگر با یک سکه نازیب ممکن است در ۱۰ پرتاب حداکثر دوبار به شیر به بینیم. احتمال چنین پیشامدی در مقام مقایسه با 0.05 کوچک نیست. پس دلیلی نداریم که به نازیبی سکه مشکوک گردیم.

مثال ۱۷ - رئیس یک کارخانه ساعت سازی می گوید روزانه به طور متوسط ۵ عده ساعت معیوب در میان ساعتهای تولید شده یافت می شوند. در یکی از روزها برحسب تصادف ۱۰ ساعت معیوب دیده شده اند. آیا گفته رئیس کارخانه با میزان $\alpha = 0.05$ پذیرفته می شود؟

حل - تعداد ساعتهای معیوب در روز یک متغیر تصادفی پواسن X با پارامتر λ است. آزمون دوسویی زیر را انجام می دهیم:

$$\begin{cases} H_0 : \lambda = 5 \\ H_1 : \lambda \neq 5 \end{cases}$$

خود X در حکم برآوردیاب گشتاوری λ است، زیرا $E(X) = \lambda$. بنابراین X را به عنوان آماره آزمون اختیار می‌کنیم و $-p$ مقدار را با فرض اینکه $\lambda = 5$ درست باشد می‌یابیم.

$$-p = \min \{P(X \leq 10), P(X \geq 10)\}$$

با استفاده از جدول ۲ داریم:

$$P(X \leq 10) = 0.9863$$

$$P(X \geq 10) = 1 - P(X \leq 9) = 1 - 0.9682 = 0.0318$$

پس داریم:

$$-p = \min \{0.9863, 0.0318\} = 0.0318$$

چون $-p$ مقدار از 0.05 بیشتر است، گفته رئیس کارخانه را با میزان 0.05 رد نمی‌کنیم.

۳.۶.۶ تمرین بخش شش

۱. تعداد اعضای هر خانواده در شهری بزرگ متغیر تصادفی $X \sim N(\mu, 4)$ است. در این شهر، برای یک نمونه تصادفی از ۱۰۰ خانواده، داریم $\bar{x} = 4/2$. آزمون زیر را

$$\begin{cases} H_0: \mu = 3/9 \\ H_1: \mu > 3/9 \end{cases}$$

را با میزان 0.05 و 0.10 انجام دهید. $-p$ مقدار را پیدا کنید.

۲. تمرین ۱ را برای آزمون دوسوئی زیر انجام دهید.

$$\begin{cases} H_0: \mu = 3/9 \\ H_1: \mu \neq 3/9 \end{cases}$$

۳. شانس آمدن شیر یا سکه‌ای پارامتر θ است. این سکه را ۱۰ بار انداخته و ۶ بار شیر دیده‌ایم. $-p$ مقدار را برای آزمونهای زیر پیدا کنید.

$$\begin{cases} H_0: \theta = \frac{1}{4} \\ H_1: \theta > \frac{1}{4} \end{cases} \quad \begin{cases} H_0: \theta = \frac{1}{4} \\ H_1: \theta \neq \frac{1}{4} \end{cases}$$

۴. در تمرین ۳، فرض کنید سکه را ۲۰۰ بار انداخته و ۲۴۰ بار شیر دیده باشیم. با این داده‌ها آزمونها را انجام دهید.

۵. فروشنده‌ای می‌گوید در هر بسته کاغذ به طور متوسط ۲ برگ لکه دار موجود است. در بستهای ۴ برگ لکه‌دار مشاهده کرده‌ایم. گفته فروشنده را با میزان 0.05 بیازمائید.

۶. فرض کنید داشته باشیم $X \sim N(\mu, \sigma^2)$. معدل یک نمونه تصادفی n تایی از X برابر عدد \bar{x} شده است. $-p$ مقدار را برای آزمون

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$$

با فرض معلوم بودن σ^2 به طور کلی محاسبه کنید.

۷.۶ متغیرهای χ^2 و T و F و کاربرد آن

در این بخش درباره سه متغیر تصادفی مهم و کاربرد آنها سخن می‌گوئیم. چگالی‌های این متغیرهای تصادفی پیچیده می‌باشد و ما در این درس به آنها نیاز نداریم.

فرض کنید متغیرهای تصادفی Z_1, Z_2, \dots, Z_n مستقل و هر یک دارای توزیع نرمال با میانگین صفر و واریانس یک باشند. به سخن دیگر یک نمونه تصادفی از توزیع نرمال استاندارد در نظر می‌گیریم.

۱.۷.۶ متغیر تصادفی χ^2

می‌گوئیم متغیر تصادفی مثبت

$$\chi^2 = \sum_{i=1}^n Z_i^2$$

دارای توزیع توان دوم کای Chi-square با n درجه آزادی است. عدد درست n را درجه آزادی نامیده‌اند زیرا این متغیر تصادفی تابعی از n متغیر تصادفی مستقل می‌باشد که هر کدام آزادانه می‌توانند تغییر کنند. برای اینکه درجه آزادی را مشخص کنیم این متغیر را با $\chi^2(n)$ نشان می‌دهیم. حرف یونانی χ را به توان دو می‌رسانیم تا با تعریف این متغیر تصادفی سازگار باشد. از اینرو $\chi^2(n)$ به عنوان یک نماد برای این متغیر تصادفی به کار می‌رود.

برای محاسبه بعضی از مقادیر احتمال از جدول ۴ استفاده می‌کنیم مثلاً

$$P(\chi^2(10) \leq 3.94) = 0.05$$

به عنوان تمرین می توان ثابت کرد که

$$E(\chi^2(n)) = n$$

$$Var(\chi^2(n)) = 2n$$

یکی از ویژگیهای مهم متغیر تصادفی $\chi^2(n)$ ویژگی جمع پذیری است که اینگونه بیان می شود: اگر $\chi^2(n_1)$ و $\chi^2(n_2)$ مستقل باشند، آنگاه $\chi^2(n_1) + \chi^2(n_2)$ دارای توزیع $\chi^2(n_1 + n_2)$ است. این ویژگی را می توان به یاری تعریف $\chi^2(n)$ ثابت کرد. برای n های بزرگ (مثلاً $n > 30$) با استفاده از قضیه حد مرکزی می توان نشان داد که متغیر استاندارد χ^2 یعنی

$$Z = \frac{\chi^2(n) - n}{\sqrt{2n}}$$

تقریباً دارای توزیع نرمال $N(0, 1)$ است. مثلاً با کاربرد تقریب نرمال داریم

$$P(\chi^2(200) \leq 100) \approx P(Z \leq \frac{100 - 200}{\sqrt{400}}) = P(Z \leq -5) \approx 0$$

۲.۷.۶ متغیر تصادفی T

فرض کنید $Z \sim N(0, 1)$ و متغیر تصادفی U هم توزیع با $\chi^2(n)$ باشد. اگر Z و U مستقل باشند، می گوئیم متغیر تصادفی

$$T = \frac{Z}{\sqrt{\frac{U}{n}}}$$

دارای توزیع T با n درجه آزادی است. برای اینکه درجه آزادی را مشخص کنیم این متغیر را با $T(n)$ نشان می دهیم. منحنی چگالی این متغیر تصادفی مانند منحنی چگالی نرمال استاندارد، متقارن و زنگ شکل می باشد، ولی بیش از منحنی چگالی نرمال در دو انتهای آن احتمال نهفته است.

برای محاسبه بعضی از مقادیر احتمال از جدول ۵ استفاده می کنیم. مثلاً

$$P(T(12) \leq 1/782) = 0/95$$

اگر n بزرگ باشد (مثلاً $n > 30$) می توان نشان داد که توزیع $T(n)$ تقریباً مانند توزیع متغیر تصادفی نرمال استاندارد است. مثلاً

$$P(T(50) \leq 2) = P(Z \leq 2) = 0/977$$

۲.۷.۶ متغیر تصادفی F

فرض کنید متغیر تصادفی V هم توزیع با $\chi^2(n_1)$ و متغیر تصادفی W هم توزیع با $\chi^2(n_2)$ باشد. اگر V و W مستقل باشند می گوئیم متغیر تصادفی مثبت

$$F = \frac{V/n_1}{W/n_2}$$

دارای توزیع اف با n_1 و n_2 درجه آزادی می باشد. برای اینکه درجه های آزادی را مشخص کنیم این متغیر را با $F(n_1, n_2)$ نشان می دهیم. به کمک جدول ۶ می توانیم بعضی از مقادیر احتمال را برای این متغیر تصادفی حساب کنیم. مثلاً

$$P(F(6, 3) \leq 1/94) = 0/95$$

۴.۷.۶ چند آماره مهم بر پایه توزیع نرمال

فرض کنید X_1, X_2, \dots, X_n یک نمونه تصادفی n تایی ($n > 1$) از توزیع نرمال $N(\mu, \sigma^2)$ باشد. حال نگاهی به توابع زیر از این نمونه تصادفی می کنیم و هر یک را شرح می دهیم:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

الف - آماره \bar{X} یک برآوردیاب ناریب برای پارامتر مجهول μ می باشد. این آماره دارای توزیع $N(\mu, \frac{\sigma^2}{n})$ است.

ب - با فرض معلوم بودن μ ، به آسانی می توان نشان داد که آماره S^2 یک برآوردیاب ناریب برای پارامتر σ^2 است و

$$\frac{nS^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2$$

هم توزیع با $\chi^2(n)$ می باشد. البته اگر μ مجهول باشد، S^2 آماره نیست زیرا در آن پارامتر مجهول موجود است.

ج - اگر μ مجهول باشد آماره S^2 یک برآوردیاب ناریب برای σ^2 است.

در حقیقت اگر در S^2 به جای μ برآوردیاب آن، یعنی \bar{X} را بگذاریم آماره زیر

$$S_b^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

به عنوان برآوردیاب σ^2 به دست می آید. ولی می توان ثابت کرد که

$$E(S_b^2) = \frac{n-1}{n} \sigma^2$$

چون امید ریاضی S_b^2 برابر σ^2 نمی شود. اصطلاحاً S_b^2 را یک برآوردیاب اریب برای σ^2 می گویند (b حرف اول کلمه biased یعنی اریب می باشد). واضح است که

$$E(S^2) = E\left(\frac{n}{n-1} S_b^2\right) = \sigma^2$$

بنابراین نتیجه می گیریم که S^2 یک برآوردیاب ناریب برای پارامتر σ^2 است. این نتیجه برای هر توزیعی با میانگین μ و واریانس σ^2 درست است و نرمال بودن توزیع لزومی ندارد. در بالا نشان دادیم که $\frac{nS^2}{\sigma^2}$ دارای توزیع $\chi^2(n)$ می باشد. ولی در درسهای بعد ثابت می کنند که

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma}\right)^2$$

دارای توزیع $\chi^2(n-1)$ است. چرا یک درجه آزادی کم می شود؟ چون مجموع متغیرهای $X_1 - \bar{X}, X_2 - \bar{X}, \dots, X_n - \bar{X}$ صفر است، نمی توانند آزادانه تغییر کنند. بلکه با معلوم بودن $n-1$ تا از آنها متغیر آخر معلوم می شود. به طور خلاصه برای σ^2 در توزیع نرمال یا غیر نرمال، سه برآوردیاب داریم:

| | | |
|---------|-------------------------------|-------------------------|
| S^2 | در صورتی که μ معلوم باشد: | این برآوردیاب ناریب است |
| S_b^2 | در صورتی که μ مجهول باشد: | این برآوردیاب اریب است |
| S^2 | در صورتی که μ مجهول باشد: | این برآوردیاب ناریب است |

۵.۷.۶ استقلال S^2 و \bar{X}

در درسهای بعد ثابت می کنند که \bar{X} یعنی معدل نمونه تصادفی نرمال و S^2 یعنی واریانس نمونه تصادفی نرمال، دو متغیر تصادفی مستقل می باشند. این ادعا ظاهراً بعید به نظر می آید زیرا هر دو تابع نمونه تصادفی می باشند. ولی همانطوری که قبلاً دیدیم استقلال یک رابطه احتمالی است، به این معنی که برای هر دو عدد حقیقی a و b داریم

$$P(\bar{X} < a, S^2 < b) = P(\bar{X} < a)P(S^2 < b)$$

در ضمن چون

$$S^2 = \frac{n}{n-1} S_b^2$$

پس S_b^2 و \bar{X} هم مستقل اند ولی به یاد بسپارید که S_b^2 و \bar{X} مستقل نیستند

$$۶.۷.۶ \text{ توزیع } \frac{(\bar{X} - \mu)}{\left(\frac{S}{\sqrt{n}}\right)}$$

فرض کنید X_1, X_2, \dots, X_n یک نمونه تصادفی از توزیع $N(\mu, \sigma^2)$ باشد. می دانیم که \bar{X} دارای توزیع

$N\left(\mu, \frac{\sigma^2}{n}\right)$ و $N(0, 1)$ است یا توجه به اینکه S^2 یک برآوردیاب ناریب برای

σ^2 می باشد. اگر به جای σ^2 بگذاریم S^2 ، توزیع $\frac{(\bar{X} - \mu)}{\left(\frac{S}{\sqrt{n}}\right)}$ چه خواهد شد؟ برای پاسخ به این پرسش

می نویسیم:

$$\frac{\frac{\bar{X} - \mu}{\left(\frac{S}{\sqrt{n}}\right)}}{\frac{(\bar{X} - \mu)}{\left(\frac{S}{\sqrt{n}}\right)}} = \frac{\frac{\bar{X} - \mu}{\left(\frac{S}{\sqrt{n}}\right)}}{\frac{\bar{X} - \mu}{\left(\frac{S}{\sqrt{n}}\right)}} = \frac{\sqrt{\frac{(n-1)S^2}{\sigma^2/(n-1)}}}{\frac{(\bar{X} - \mu)}{\left(\frac{S}{\sqrt{n}}\right)}}$$

از آنجائی که S^2 یک برآوردیاب ناریب برای σ^2 می باشد،

حال فرض می کنیم $Z = \frac{(\bar{X} - \mu)}{\left(\frac{\sigma}{\sqrt{n}}\right)}$ و $U = \frac{(n-1)S^2}{\sigma^2}$. می دانیم که Z دارای توزیع $N(0, 1)$ و

U دارای توزیع $\chi^2(n-1)$ می باشد. در ضمن Z و U مستقل اند، زیرا \bar{X} و S^2 مستقل اند. از اینرو، طبق

تعریف متغیر T

$$\frac{(\bar{X} - \mu)}{\left(\frac{S}{\sqrt{n}}\right)} = \frac{Z}{\sqrt{\frac{U}{(n-1)}}}$$

دارای توزیع T با $n-1$ درجه آزادی است. با این حال هرگاه n خیلی بزرگ باشد این کسر به تفریب دارای توزیع نرمال استاندارد است.

۷.۷.۶ واریانس آمیخته

فرض کنید X_1, X_2, \dots, X_m یک نمونه تصادفی از $X \sim N(\mu_1, \sigma^2)$ و Y_1, Y_2, \dots, Y_n یک نمونه تصادفی از $Y \sim N(\mu_2, \sigma^2)$ باشد، به طوری که X و Y دو نمونه مستقل بوده و $\mu_1 \neq \mu_2$. ملاحظه می شود که این دو توزیع دارای واریانس مشترک σ^2 می باشند، پارامتر σ^2 را چگونه برآورد کنیم؟ برآوردیاب S_p^2 از روی نمونه اول و نمونه دوم به ترتیب می شوند

$$S_1^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2$$

$$S_2^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2$$

حال این دو برآوردیاب را به صورت زیر با هم می آمیزیم:

$$S_p^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2}$$

برآوردیاب S_p^2 را اصطلاحاً واریانس آمیخته (pooled variance) می نامند. می توان ثابت کرد که S_p^2 هم مانند S_1^2 و S_2^2 یک برآوردیاب نااریب برای σ^2 می باشد یعنی:

$$E(S_p^2) = E(S_1^2) = E(S_2^2) = \sigma^2$$

در ضمن S_p^2 از دو برآوردیاب دیگر بهتر می باشد زیرا می توان ثابت کرد که

$$Var(S_p^2) < Var(S_1^2)$$

$$Var(S_p^2) < Var(S_2^2)$$

برای اثبات مطالب بالا باید نخست از مستقل بودن S_1^2 و S_2^2 و لزوم جمع پذیری متغیرهای χ^2 مستقل استفاده کرد و نشان داد که $\frac{(m+n-2)S_p^2}{\sigma^2}$ دارای توزیع $\chi^2(m+n-2)$ می باشد، و سپس میانگین و واریانس $\chi^2(m+n-2)$ را حساب نمود.

نسبت $\frac{S_1^2}{S_p^2}$ را می توان به عنوان برآوردیاب نسبت $\frac{\sigma_1^2}{\sigma_2^2}$ به کار برد. اگر نسبت اول به یک نزدیک باشد، شانس مساوی بودن واریانسهای X و Y زیاد است. از اینرو داشتن توزیع $\frac{S_1^2}{S_2^2}$ ضروری است.

۸.۷.۶ توزیع $\frac{S_1^2}{S_2^2}$

برای دو نمونه تصادفی نرمال و مستقل، که در بالا به آنها اشاره شد، داریم:

$$V = \frac{(m-1)S_1^2}{\sigma_1^2} \sim \chi^2(m-1)$$

$$W = \frac{(n-1)S_2^2}{\sigma_2^2} \sim \chi^2(n-1)$$

(نماد $\frac{D}{d}$ به معنی هم توزیع بودن دو متغیر تصادفی است.)

چون دو نمونه تصادفی بالا مستقل از هم می باشند، پس V و W هم مستقل اند. حال طبق تعریف

متغیر تصادفی F نتیجه می گیریم که

$$\frac{V}{W} = \frac{\frac{(m-1)S_1^2}{\sigma_1^2}}{\frac{(n-1)S_2^2}{\sigma_2^2}} = F(m-1, n-1)$$

پس می توان توزیع $\frac{S_1^2}{S_2^2}$ را از روی توزیع $F(m-1, n-1)$ با داشتن σ_1^2 و σ_2^2 محاسبه کرد. به عنوان مثال اگر از $X \sim N(\mu_1, 4)$ یک نمونه تصادفی ۱۱ تایی و از $Y \sim N(\mu_2, 1)$ یک نمونه تصادفی ۹ تایی داشته باشیم و این دو نمونه از هم مستقل باشند، آنگاه

$$P\left(\frac{S_1^2}{S_2^2} \leq 13/4\right) = P\left(\frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \leq 13/4 \times \frac{1}{4}\right) = P(F(10, 8) \leq 3/25)$$

$$= 0/95$$

۹.۷.۶ تمرین بخش هفت

۱. با استفاده از جدول χ^2 ، احتمال زیر را حساب کنید.

$$P(\chi^2(40) < 39/3)$$

۲. عدد c را طوری تعیین کنید تا داشته باشیم

$$P(\chi^2(12) > c) = 0/95$$

۳. با فرض $Z \sim N(0, 1)$ ، $E(Z^k)$ را برای $k = 1, 2, 3, 4$ حساب کنید. حال امید ریاضی و واریانس $\chi^2(n)$ را با استفاده از تعریف این متغیر پیدا کنید.

۴. با استفاده از تعریف $\chi^2(n)$ و قضیه حد مرکزی نشان دهید که $\chi^2(n)$ برای n های بزرگ تقریباً دارای توزیع $N(n, 2n)$ می باشد. حال پاسخ تمرین ۱ را به تقریب به دست آورید.

۵. با استفاده از تعریف $\chi^2(n)$ ، خاصیت جمع پذیری این متغیر را توجیه نمایید. با فرض مستقل بودن متغیرهای $\chi^2(5)$ و $\chi^2(7)$ ، احتمال زیر را حساب کنید:

$$P\left(\frac{\chi^2(5) + \chi^2(7)}{2} < 2/20\right)$$

۶. با استفاده از جدول T ، احتمال زیر را حساب کنید:

$$P(0/129 < T(10) < 0/542)$$

۷. عدد c را طوری تعیین کنید تا داشته باشیم

$$P(T(7) > c) = 0/25$$

۸. $P(T(200) > 2)$ را به تقریب حساب کنید.

۹. با استفاده از جدول F ، احتمالهای زیر را حساب کنید.

$$P(F(5, 3) > 9/01) \quad P(F(5, 5) < 11)$$

۱۰. عدد c را طوری پیدا کنید تا داشته باشیم

$$P(F(12, 5) < c) = 0/99$$

۱۱. با استفاده از تعریف S^2 ، ثابت کنید که در توزیع نرمال، $\frac{nS^2}{\sigma^2}$ دارای توزیع $\chi^2(n)$ می باشد. حال امید ریاضی و واریانس S^2 را پیدا کنید.

۱۲. فرض کنید X_1, X_2, \dots, X_n یک نمونه تصادفی از $N(1, \theta)$ باشد. یک برآوردیاب ناریب برای θ پیدا کنید. با فرض $\theta = 2$ ، احتمال اینکه این برآوردیاب کوچکتر از ۵ باشد چیست؟

۱۳. فرض کنید X_1, X_2, \dots, X_n یک نمونه تصادفی از توزیعی با میانگین μ و واریانس σ^2 باشد. تساوی زیر را ثابت کنید.

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_{i=1}^n (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2$$

حال با استفاده از تعریف S^2 ، امید ریاضی این متغیر را پیدا کنید.

۱۴. در تمرین ۱۳، اگر نمونه تصادفی از توزیع $N(\mu, \sigma^2)$ باشد، با استفاده از تعریف S^2 ، امید ریاضی و واریانس این متغیر را حساب کنید.

۱۵. فرض کنید X_1, X_2, \dots, X_{10} یک نمونه تصادفی از توزیع $N(1, 3)$ باشد. احتمال زیر را حساب کنید

$$P(\bar{X} > 2, S^2 < 1/39)$$

۱۶. برای یک نمونه تصادفی F تائی از $N(\mu_1, 2)$ و یک نمونه تصادفی F تائی از $N(\mu_2, 4)$ با فرض مستقل بودن دو نمونه، احتمال زیر را حساب کنید.

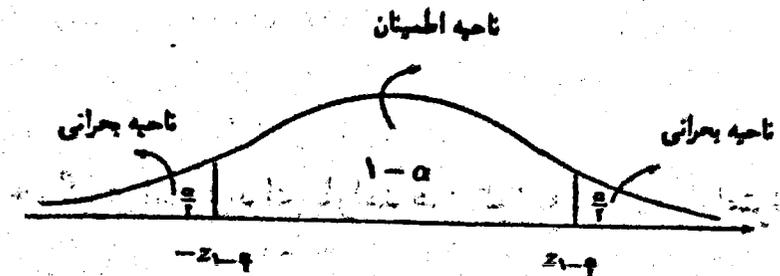
$$P\left(\frac{S_1^2}{S_2^2} > 2/252\right)$$

۱۷. امید ریاضی و واریانس S_p^2 (واریانس آمیخته) را حساب کنید.

۱۸. ثابت کنید که واریانس آمیخته ناریب است.

۱۹. $Var(S_p^2)$ را پیدا کنید و با $Var(S_1^2)$ و $Var(S_2^2)$ مقایسه نمایید.

۲۰. برای K نمونه تصادفی مستقل، واریانس آمیخته را بیابید.



نگاره ۱ چگالی تابع محوری Q

ب - σ^2 مجهول

در تابع محوری Q به جای σ^2 برآوردیاب آن S^2 را می‌کناریم تا تابع محوری زیر، که دارای توزیع تی با $n-1$ درجه آزادی است، به دست آید:

$$Q^* = \frac{\bar{X} - \mu}{S} \sqrt{n} \sim T(n-1)$$

برای تابع محوری Q^* داریم

$$P(-t_{1-\frac{\alpha}{2}}(n-1) < \frac{\bar{X} - \mu}{S} \sqrt{n} < t_{1-\frac{\alpha}{2}}(n-1)) = 1 - \alpha$$

از حل نامساوی دو سوئی داخل پرانتز بر حسب μ فاصله زیر به دست می‌آید:

$$\bar{X} - t_{1-\frac{\alpha}{2}}(n-1) \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{1-\frac{\alpha}{2}}(n-1) \frac{S}{\sqrt{n}} \quad (2)$$

فاصله (۲) مانند فاصله (۱) می‌باشد، با این تفاوت که σ به S و $z_{1-\frac{\alpha}{2}}$ (که در جدول نرمال است) به $t_{1-\frac{\alpha}{2}}(n-1)$ (که در جدول T است) تبدیل شده‌اند. این دو فاصله دارای احتمالی مساوی در دو انتها هستند (در هر دو انتها مقدار احتمال $\frac{\alpha}{2}$ می‌باشد) و می‌توان ثابت کرد که بهترین فواصل اطمینان می‌باشند. منحنی چگالی تابع محوری Q^* هم مانند نگاره ۱ می‌باشد. به ویژه اگر $n \geq 30$ باشد تقریباً $t_{1-\frac{\alpha}{2}}(n-1)$ و $z_{1-\frac{\alpha}{2}}$ برابرند.

نتیجه گیری آماری برای پارامترهای توزیع نرمال و پارامترنسبت

۱.۷ فاصله اطمینان و آزمون آماری برای پارامترهای توزیع نرمال

فرض کنید X_1, X_2, \dots, X_n یک نمونه تصادفی از $X \sim N(\mu, \sigma^2)$ باشد. در این بخش می‌خواهیم برای پارامترهای این توزیع فاصله اطمینان پیدا کنیم و در مورد آنها آزمون آماری انجام دهیم. ضریب فاصله اطمینان را با $1-\alpha$ و میزان آزمون را با α نشان می‌دهیم. در عمل α را کوچک، معمولاً 0.01 یا 0.05 یا 0.1 اختیار می‌کنند، تا فرض H_0 که تا به حال مورد قبول بوده است به ناحق رد نشود.

۱.۱.۷ فاصله اطمینان برای μ

چنین فاصله‌ای را می‌توان به آسانی پیدا کرد. دو حالت در نظر می‌گیریم:

الف - σ^2 معلوم

برای تابع محوری

$$Q = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

با توجه به نگاره ۱ داریم

$$P(-z_{1-\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < z_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

از حل نامساوی دو سوئی داخل پرانتز بر حسب μ فاصله زیر به دست می‌آید:

$$\left(\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) \quad (1)$$

۲.۱.۷ آزمون آماری برای μ

برای آزمون μ دو فرض ساده به صورت زیر در نظر می گیریم:

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$$

اگر \bar{X} یعنی برآوردیاب μ ، بیش از حد بزرگ شود H_0 را رد می کنیم و H_1 را می پذیریم، زیرا می دانیم که $\mu_1 > \mu_0$. برای تعیین ناحیه بحرانی کافی است که مقدار ثابت c را، که به بستگی دارد،

طوری پیدا کنیم تا داشته باشیم:

$$Q = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > c$$

دو حالت در نظر می گیریم:

الف - σ^2 معلوم

تحت فرض H_0 داریم $Q \sim N(0, 1)$ ، یعنی Q یا متغیر نرمال استاندارد Z هم توزیع می باشد. چون می خواهیم خطای نوع اول برابر α باشد، پس

$$\alpha = P(H_0 \text{ درست} | H_1 \text{ را رد کنیم}) = P(Z > c) = 1 - \Phi(c)$$

$$\Phi(c) = 1 - \alpha$$

$$c = z_{1-\alpha}$$

بنابراین فرض H_0 را موقعی رد می کنیم که $Q > z_{1-\alpha}$ یا

$$\bar{X} > \mu_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \quad (۳)$$

ب - σ^2 مجهول

به جای σ^2 برآوردیاب آن S^2 را می گذاریم. در نتیجه Q به

$$Q = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

تبدیل می شود که هم توزیع با $T(n-1)$ است. بنابراین فرض H_0 را موقعی رد می کنیم که $Q > t_{1-\alpha}(n-1)$ یا

$$\bar{X} > \mu_0 + t_{1-\alpha}(n-1) \frac{S}{\sqrt{n}} \quad (۴)$$

ناحیه بحرانی (۳) مانند (۴) می باشد، با این تفاوت که σ^2 به S^2 و عدد $z_{1-\alpha}$ به عدد $t_{1-\alpha}(n-1)$ تبدیل شده اند.

اینک چند نکته مهم را شرح می دهیم:

یک - ناحیه بحرانی (۳) و (۴)، خواه σ^2 معلوم یا مجهول باشد، به μ_1 بستگی ندارند و تنها از $\mu_1 > \mu_0$ استفاده شده است. از اینرو می توان این دو ناحیه بحرانی را برای آزمون زیر هم که به آزمون یکسوئی شهرت دارد به کار برد.

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$$

در این آزمون، فرض H_0 را، که تحت آن توزیع X مشخص و قابل محاسبه می باشد، یک فرض ساده می نامند. ولی فرض H_1 را که تحت آن توزیع قابل محاسبه نمی باشد (زیرا نمی دانیم که μ چقدر است، تنها می دانیم که از μ_0 بزرگتر است) یک فرض مرکب می گویند.

دو - ناحیه بحرانی برای آزمون یکسوئی

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$$

بر اساس کوچک بودن \bar{X} به دست می آید. اگر مانند بالا عمل کنیم دو ناحیه بحرانی زیر به دست می آیند:

$$\bar{X} < \mu_0 - z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \quad (۵)$$

$$\bar{X} < \mu_0 - t_{1-\alpha}(n-1) \frac{S}{\sqrt{n}} \quad (۶)$$

سه - ناحیه بحرانی برای آزمون دو سوئی

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

بر اساس کوچک بودن یا بزرگ بودن \bar{X} به دست می آید. اگر احتمال را در هر سو $\frac{\alpha}{2}$ بگیریم. ناحیه بحرانی زیر به دست می آید:

الف) σ^2 معلوم

$$\bar{X} > \mu_0 + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad \text{یا} \quad \bar{X} < \mu_0 - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \quad (۷)$$

ب) σ^2 مجهول

$$\bar{X} > \mu_0 + t_{1-\frac{\alpha}{2}}(n-1) \frac{S}{\sqrt{n}} \quad \text{یا} \quad \bar{X} < \mu_0 - t_{1-\frac{\alpha}{2}}(n-1) \frac{S}{\sqrt{n}} \quad (۸)$$

جمله $t_{1-\frac{\alpha}{2}}(n-1) \frac{S}{\sqrt{n}}$ را جمله خطا می گویند. کوچکی آن به α و n بستگی دارد. چهار - به عنوان مثال، ملاحظه می شود که فاصله اطمینان (۱) و ناحیه بحرانی دو سوئی (۷) با هم ارتباط نزدیک دارند و برای تعیین آنها به طریق زیر عمل می کنیم:

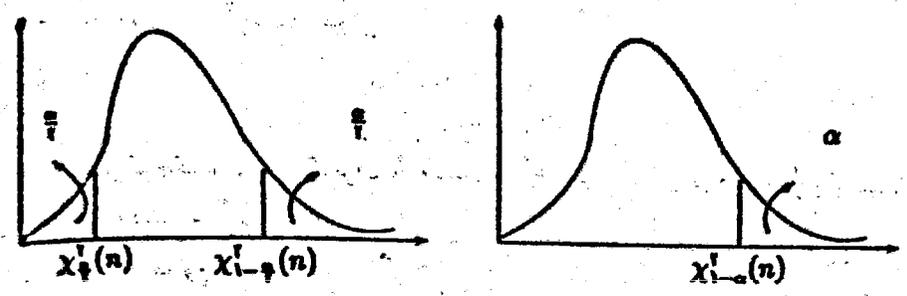
تابع محوری Q را که تابعی از پارامتر μ و آماره \bar{X} می باشد پیدا می کنیم که با فرض $Q \sim N(\mu, \sigma^2)$ دارای توزیع $N(0, 1)$ است. روی محور طولها دو نقطه $-z_{1-\frac{\alpha}{2}}$ و $z_{1-\frac{\alpha}{2}}$ را اختیار می نماییم. از نامساوی $|Q| < z_{1-\frac{\alpha}{2}}$ فاصله اطمینان و از نامساوی $|Q| > z_{1-\frac{\alpha}{2}}$ ناحیه بحرانی به دست می آید. به عبارت دیگر فاصله $(-z_{1-\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}})$ روی محور طولها، مخصوص فاصله اطمینان و مکمل آن مخصوص ناحیه بحرانی است (دو نقطه انتهائی را نادیده می گیریم، چون دارای احتمال صفر هستند) (به نگاره ۱ نگاه کنید).

به همین طریق فاصله اطمینان (۲) و ناحیه بحرانی دو سوئی (۸) با هم ارتباط دارند. افزون بر این، این مشاهده کلی بوده، از این پس آن را در موارد دیگر هم به کار می بریم.

پنج - بنا بر قضیه حد مرکزی نتایج بالا برای نمونه های بزرگ ($n \geq 30$)، که از توزیع نرمال نباشند، نیز به طور تقریبی درست اند.

۳.۱.۷ فاصله اطمینان برای σ^2

دو حالت در نظر می گیریم:



نگاره ۲ فاصله اطمینان آزمون σ^2

الف - μ معلوم

آماره S^2 یک برآوردیاب ناربیب برای σ^2 است و

$$Q_1 = n \frac{S^2}{\sigma^2}$$

دارای توزیع $\chi^2(n)$ می باشد. پس Q_1 یک تابع محوری است. روی محور طولها فاصله $(x_{1-\frac{\alpha}{2}}^*(n), x_{\frac{\alpha}{2}}^*(n))$ را اختیار می کنیم. دو سر این فاصله را به یاری جدول $\chi^2(n)$ با داشتن n و α می یابیم. (به نگاره ۲ نگاه کنید). برای این تابع محوری داریم:

$$P\left(\frac{nS^2}{\sigma^2} < x_{1-\frac{\alpha}{2}}^*(n) < x_{\frac{\alpha}{2}}^*(n)\right) = 1 - \alpha$$

با حل نامساوی دو سوئی داخل پرانتز بر حسب σ^2 ، یک فاصله اطمینان با احتمالات مساوی در دو انتها می شود:

$$\left(\frac{nS^2}{x_{1-\frac{\alpha}{2}}^*(n)}, \frac{nS^2}{x_{\frac{\alpha}{2}}^*(n)} \right) \quad (۹)$$

ب - μ مجهول

آماره S^2 یک برآوردیاب ناربیب برای σ^2 است و

$$Q_1^* = \frac{(n-1)S^2}{\sigma^2}$$

دارای توزیع $\chi^2(n-1)$ می باشد، پس اگر در فاصله (۹)، S^2 را به S^2 و n را به $n-1$ تبدیل کنیم. فاصله اطمینان در این حالت به دست می آید.

۲.۱.۷ آزمون آماری برای σ^2

برای آزمون

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{cases}$$

دو حالت در نظر می گیریم:

الف - μ معلوم

چون $\frac{nS^2}{\sigma_0^2}$ تحت فرض H_0 دارای توزیع $\chi^2(n)$ است، پس ناحیه بحرانی می شود:

$$\frac{nS^2}{\sigma_0^2} > \chi^2_{1-\alpha}(n) \quad (10)$$

ب - μ مجهول

چون $\frac{(n-1)S^2}{\sigma_0^2}$ تحت فرض H_0 دارای توزیع $\chi^2(n-1)$ است، پس ناحیه بحرانی مانند ناحیه بحرانی (۱۰) می باشد، با این تفاوت که باید S^2 را به s^2 و n را به $n-1$ تبدیل کرد. به همین طریق می توان به عنوان تمرین ناحیه بحرانی آزمون یکسرنی

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 \end{cases}$$

و ناحیه بحرانی آزمون دوسرنی

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases}$$

را با تغییرات لازم به دست آورید.

مثال ۱ مقدار محصولی را که یک کارخانه شیمیایی در روز تولید می کند متغیر تصادفی X برحسب تن، با توزیع نرمال می باشد. برای ۵۰ روز $\bar{x} = ۸۷۱$ و $s = ۲۱$ مشاهده شده اند. یک فاصله اطمینان نود درصدی برای میانگین μ به دست آورید.

حل - از فاصله اطمینان (۲) استفاده می کنیم. داده ها عبارتند از

$$n = 50, \bar{x} = 871, s = 21, z_{0.95}(29) = z_{0.95} = 1/625$$

دو انتهای فاصله می شوند

$$871 \pm 1/625 \left(\frac{21}{\sqrt{50}} \right)$$

بنابراین فاصله (۸۶۶/۱۱، ۸۷۵/۸۹) به دست می آید.

مثال ۲ مدت آرام بخشی یک نوع مسکن، متغیر تصادفی نرمال X برحسب دقیقه می باشد. آمار نشان می دهد که برای ۱۲ مصرف کننده $\bar{x} = ۴۲$ و $s = ۱۱/۹$. آزمون زیر را با میزان پنج درصد انجام دهید

$$\begin{cases} H_0 : \sigma^2 = 100 \\ H_1 : \sigma^2 > 100 \end{cases}$$

حل - ناحیه بحرانی آزمون از نامساوی زیر به دست می آید:

$$(n-1) \frac{s^2}{\sigma_0^2} > \chi^2_{1-\alpha}(n-1)$$

$$n = 12, \sigma^2 = 100, s^2 = (11/9)^2, \alpha = 0.05$$

چون $\chi^2_{0.95}(11) = 19/67$ و $(n-1)s^2/\sigma_0^2 = 15/58$ ، پس H_0 را رد نمی کنیم.

۵.۱.۷ تمرین بخش یک

۱. در مثال ۱ یک فاصله اطمینان نود درصدی برای σ^2 به دست آورید.

۲. در مثال ۱ آزمون

$$\begin{cases} H_0 : \sigma^2 = 400 \\ H_1 : \sigma^2 > 400 \end{cases}$$

را با میزان ده درصد انجام دهید.

۳. در مثال ۲ یک فاصله اطمینان نود درصدی برای μ به دست آورید.

۲.۷ فاصله اطمینان و آزمون آماری برای پارامترهای دو توزیع نرمال مستقل

فرض کنید X_1, X_2, \dots, X_n یک نمونه تصادفی از $X \sim N(\mu_1, \sigma_1^2)$ و Y_1, Y_2, \dots, Y_m یک نمونه تصادفی $Y \sim N(\mu_2, \sigma_2^2)$ باشد. این دو متغیر تصادفی و در نتیجه دو نمونه را مستقل از هم می‌گیریم. مثلاً نمونه اول را اندازه‌های قد بیست دانش آموز هفت ساله در شیراز و نمونه دوم را اندازه‌های قد پانزده دانش آموز هفت ساله در تبریز تصور کنید.

در این بخش می‌خواهیم برای پارامترهایی مانند $\mu_2 - \mu_1$ و $\frac{\sigma_2^2}{\sigma_1^2}$ فاصله اطمینان پیدا کنیم و در مورد آنها آزمون آماری انجام دهیم.

چون این دو نمونه نرمال از هم مستقل می‌باشند، پس $\bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{n})$ و $\bar{Y} \sim N(\mu_2, \frac{\sigma_2^2}{m})$ هم مستقل هستند و $\bar{Y} - \bar{X} \sim N(\mu_2 - \mu_1, \frac{\sigma_2^2}{m} + \frac{\sigma_1^2}{n})$ یک برآوردیاب نااریب برای $\mu_2 - \mu_1$ است. بدیهی است که

$$\sigma_{\bar{Y} - \bar{X}}^2 = \text{var}(\bar{Y} - \bar{X}) = \frac{\sigma_2^2}{m} + \frac{\sigma_1^2}{n}$$

افزون بر این، دو آماره

$$S_1^2 = \frac{1}{n-1} \sum_{j=1}^n (Y_j - \bar{Y})^2 \quad \text{و} \quad S_2^2 = \frac{1}{m-1} \sum_{i=1}^m (X_i - \bar{X})^2$$

مستقل بوده و به ترتیب برآوردیابهای نااریب برای σ_1^2 و σ_2^2 می‌باشند. در فصل پیش دیدیم که

$$\frac{(n-1)S_1^2}{\sigma_1^2} \sim \chi^2(n-1) \quad \text{و} \quad \frac{(m-1)S_2^2}{\sigma_2^2} \sim \chi^2(m-1)$$

۱.۲.۷ فاصله اطمینان برای $\mu_2 - \mu_1$

برای پیدا کردن این فاصله $\bar{Y} - \bar{X}$ را استاندارد می‌کنیم. تابع محوری

$$Q = \frac{\bar{Y} - \bar{X} - (\mu_2 - \mu_1)}{\sigma_{\bar{Y} - \bar{X}}} \sim Z \sim N(0, 1)$$

بدهست می‌آید. دو حالت در نظر می‌گیریم و ضریب اطمینان را $1 - \alpha$ فرض می‌کنیم.

الف - σ_1^2 و σ_2^2 معلوم

در این حالت $\sigma_{\bar{Y} - \bar{X}}$ معلوم می‌باشد و داریم

$$P(-z_{1-\frac{\alpha}{2}} < Q < z_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

۴. در مثال ۲، اگر $\mu = 35$ ، یک فاصله اطمینان نود درصدی برای σ^2 به دست آورید.

۵. فرض کنید X دارای توزیع $N(\mu, \sigma^2)$ باشد. اگر μ و σ^2 هر دو مجهول باشند، با استفاده از نمونه تصادفی X_1, X_2, \dots, X_n فرم ناحیه بحرانی را برای آزمونهای زیر پیدا کنید (خطای نوع اول را α بگیرید)

$$\begin{cases} H_0: \sigma^2 = \sigma_0^2 \\ H_1: \sigma^2 \neq \sigma_0^2 \end{cases} \quad \begin{cases} H_0: \sigma^2 = \sigma_0^2 \\ H_1: \sigma^2 < \sigma_0^2 \end{cases}$$

۶. در مثال ۲، اگر $\sigma^2 = 100$ آزمون

$$\begin{cases} H_0: \mu = 35 \\ H_1: \mu > 35 \end{cases}$$

را با میزان ده درصد انجام دهید، و توان آزمون را برای $\mu_1 = 40$ حساب کنید.

۷. در مثال ۱، آزمون زیر را از راه p -مقدار با $\alpha = 0.10$ انجام دهید.

$$\begin{cases} H_0: \mu = 70 \\ H_1: \mu \neq 70 \end{cases}$$

۸. در مثال ۲، آزمون زیر را از راه p -مقدار با $\alpha = 0.05$ انجام دهید.

$$\begin{cases} H_0: \mu = 8 \\ H_1: \mu < 8 \end{cases}$$

۹. مطالعه آماری در گذشته نشان داده است که در عرض سال به طور متوسط کارمندان ۱۵ روز در اثر بیماری غیبت دارند. پژوهشگری برای ۲۵ کارمند در سال گذشته شماره روزهای غیبت را به شرح زیر ثبت کرده است:

| | | | | | | | | | |
|----|----|----|----|----|----|----|----|----|----|
| ۵ | ۲۵ | ۱۰ | ۰ | ۳ | ۵۰ | ۱۲ | ۱۴ | ۴۰ | ۱۲ |
| ۳۲ | ۸ | ۴ | ۲۷ | ۲۰ | ۱۴ | ۱۸ | ۱۶ | ۱۰ | ۱ |
| ۲۲ | ۵۸ | ۵ | ۲۳ | ۹ | | | | | |

با میزان ۵ درصد بیازمائید که در سال گذشته روزهای غیبت بیش از ۱۵ روز است. از p -مقدار استفاده کنید.

۱۰. در تمرین ۹، با میزان ده درصد بیازمائید که در سال گذشته روزهای غیبت با ۱۵ اختلاف چندانی ندارند.

از حل نامساوی دوسوئی داخل پراکنش بالا بر حسب $\mu_2 - \mu_1$ فاصله زیر به دست می آید.

$$\bar{Y} - \bar{X} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} < \mu_2 - \mu_1 < \bar{Y} - \bar{X} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} \quad (11)$$

ب - σ_1^2 و σ_2^2 مجهول

در این حالت در صورتی که داشته باشیم $\sigma_1^2 = \sigma_2^2 = \sigma^2$ را به وسیله واریانس آمیخته یعنی

$$S_p^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2}$$

برآورد می کنیم. در نتیجه یک برآوردیاب ناریب برای $\sigma_{\bar{Y}-\bar{X}}^2$ می شود

$$\widehat{\sigma_{\bar{Y}-\bar{X}}^2} = S_p^2 \left(\frac{1}{n} + \frac{1}{m} \right)$$

حال از تابع محوری

$$Q^* = \frac{\bar{Y} - \bar{X} - (\mu_2 - \mu_1)}{\sqrt{\widehat{\sigma_{\bar{Y}-\bar{X}}^2}}} \stackrel{D}{=} T(m+n-2)$$

استفاده می کنیم. بنابراین فاصله اطمینان برای $\mu_2 - \mu_1$ مانند فاصله (۱۱) می شود، با این تفاوت که

باید به جای $\sigma_{\bar{Y}-\bar{X}}^2$ برآوردیاب آن یعنی $\widehat{\sigma_{\bar{Y}-\bar{X}}^2}$ و به جای عدد $z_{1-\frac{\alpha}{2}}(m+n-2)$ را $t_{1-\frac{\alpha}{2}}$ قرار داد. در نتیجه فاصله زیر به دست می آید:

$$\bar{Y} - \bar{X} - t_{1-\frac{\alpha}{2}}(m+n-2) S_p \sqrt{\frac{1}{m} + \frac{1}{n}} < \mu_2 - \mu_1 < \bar{Y} - \bar{X} + t_{1-\frac{\alpha}{2}}(m+n-2) S_p \sqrt{\frac{1}{m} + \frac{1}{n}}$$

۲.۲.۷ آزمون آماری برای $\mu_2 - \mu_1$

برای انجام آزمون زیر

$$\begin{cases} H_0 : \mu_2 - \mu_1 = a \\ H_1 : \mu_2 - \mu_1 > a \end{cases}$$

از آماره $\bar{Y} - \bar{X}$ استفاده می کنیم. اگر این آماره بیش از حد بزرگ شود H_0 را رد می کنیم. دو حالت در

نظر می گیریم:

الف - σ_1^2 و σ_2^2 معلوم

با استفاده از تابع محوری Q ، که در بالا به آن اشاره شده، ناحیه بحرانی می شود $Q > z_{1-\frac{\alpha}{2}}$ یا

$$\bar{Y} - \bar{X} > \mu_2 - \mu_1 + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} \quad (13)$$

ب - σ_1^2 و σ_2^2 مجهول

در صورتی که داشته باشیم $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ، با استفاده از تابع محوری Q^* ، که در بالا به آن اشاره شده،

ناحیه بحرانی می شود $Q^* > t_{1-\alpha}(m+n-2)$ یا

$$\bar{Y} - \bar{X} > \mu_2 - \mu_1 + t_{1-\alpha}(m+n-2) S_p \sqrt{\frac{1}{m} + \frac{1}{n}} \quad (14)$$

به همین طریق می توان برای آزمونهای زیر

$$\begin{cases} H_0 : \mu_2 - \mu_1 = a \\ H_1 : \mu_2 - \mu_1 < a \end{cases} \quad \begin{cases} H_0 : \mu_2 - \mu_1 = a \\ H_1 : \mu_2 - \mu_1 \neq a \end{cases}$$

ناحیه بحرانی پیدا کرد.

در صورتی که σ_1^2 و σ_2^2 مجهول و نابرابر باشند، به دست آوردن فاصله اطمینان و انجام آزمون

آماره برای $\mu_2 - \mu_1$ ، مستلزم به کار بردن فرمولهای پیچیده می باشد. ما از ذکر اینگونه فرمولها در اینجا

خودداری می کنیم، تنها یادآور می شویم که اگر اندازه نمونهها یعنی m و n بزرگ باشند می توان σ_1^2 و σ_2^2

را با S_1^2 و S_2^2 تقریب کرده آنها را دانسته تصور نمود و سپس فرمولهای (۹) و (۱۱) را به کار برد.

۳.۲.۷ فاصله اطمینان برای $\frac{\sigma_1^2}{\sigma_2^2}$

در بخش ۵ نشان دادیم که

$$Q_1 = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} = \frac{S_1^2}{S_2^2} \frac{\sigma_2^2}{\sigma_1^2} \stackrel{D}{=} F(m-1, n-1)$$

بنابراین Q_1 یک تابع محوری است و داریم

$$P(F_{\frac{\alpha}{2}}(m-1, n-1) < Q_1 < F_{1-\frac{\alpha}{2}}(m-1, n-1)) = 1 - \alpha$$

از حل نامساوی دو سوئی داخل پرانتز بالا بر حسب σ_1^2/σ_2^2 فاصله زیر به دست می آید:

$$\frac{S_1^2}{S_2^2} F_{\frac{\alpha}{2}}(m-1, n-1) < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} F_{1-\frac{\alpha}{2}}(m-1, n-1) \quad (15)$$

در اینجا لازم است درباره استفاده از جدول F نکته ای مهم را یادآور شویم. به عنوان مثال اگر $\alpha = 0.10$ و درجه های آزادی صورت و مخرج به ترتیب $m_1 = 8$ و $m_2 = 6$ باشند، از جدول F داریم $F_{0.10}(8, 6) = 4/15$ اغلب جدولها احتمال طرف راست را منظور می دارند و به جای $F_{0.10}(8, 6)$ می نویسند $F_{0.05}(8, 6)$ ولی ما برای اینکه با تعریف توزیع احتمال هماهنگی داشته باشیم، $F_{0.10}(8, 6)$ را نقطه ای می دانیم که در طرف چپ آن به اندازه 0.10 احتمال داریم. عدد $F_{0.10}(8, 6)$ یعنی نقطه ای که طرف چپ آن 0.05 احتمال است، در جدول یافت نمی شود، زیرا جدول تنها تقاطی را می دهد که در طرف چپ آنها 0.10 یا 0.05 یا 0.99 احتمال موجود است. یا اینحال می توان به عنوان تمرین رابطه عددی

$$F_{\alpha}(m_1, m_2) = \frac{1}{F_{1-\alpha}(m_2, m_1)} \quad 0 < \alpha < 1$$

را ثابت کرد. با استفاده از این رابطه داریم

$$F_{0.05}(8, 6) = \frac{1}{F_{0.95}(6, 8)} = \frac{1}{2/58} = 0.28$$

۴.۲.۷ آزمون آماری برای $\frac{\sigma_1^2}{\sigma_2^2}$

اغلب می خواهیم بیازمائیم که دو توزیع نرمال مستقل دارای واریانسهای برابر می باشند، تا بتوانیم درباره $\mu_1 - \mu_2$ آزمون انجام دهیم. از اینرو آزمون زیر را انجام می دهیم:

$$\begin{cases} H_0: \sigma_1^2 = \sigma_2^2 \\ H_1: \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

چون این آزمون دوسوئی می باشد، مکمل فاصله عددی

$$(F_{\frac{\alpha}{2}}(m-1, n-1), F_{1-\frac{\alpha}{2}}(m-1, n-1))$$

ناحیه بحرانی را مشخص می کند. بنابراین تحت فرض H_0 ، یعنی با فرض $\frac{\sigma_1^2}{\sigma_2^2} = 1$ ، ناحیه بحرانی به صورت زیر است

$$\frac{S_1^2}{S_2^2} < F_{\frac{\alpha}{2}}(m-1, n-1) \quad \text{یا} \quad \frac{S_1^2}{S_2^2} > F_{1-\frac{\alpha}{2}}(m-1, n-1) \quad (16)$$

این ناحیه بحرانی را می توان بدین طریق توجیه کرد: با فرض درست بودن H_0 داریم

$$\frac{S_1^2}{S_2^2} \stackrel{D}{=} F(m-1, n-1)$$

چون $\frac{S_1^2}{S_2^2}$ را می توان به عنوان برآوردیاب $\frac{\sigma_1^2}{\sigma_2^2}$ به کار برد، پس اگر $\frac{S_1^2}{S_2^2}$ بیش از حد بزرگ یا کوچک شود باید فرض H_0 را رد کرد.

مثال ۵ درس زبان را دو یک کلاس ۹ نفری با معلم زن و در یک کلاس ۷ نفری با معلم مرد تدریس کرده اند. در پایان یک امتحان استاندارد به هر دو کلاس داده و برای نمره ها در این دو کلاس، نتایج زیر را به دست آورده اند.

$$\bar{x} = 85 \quad s_1 = 4 \quad \bar{y} = 81 \quad s_2 = 5$$

با فرض نرمال بودن و مستقل بودن نمره ها، آزمون زیر را با میزان ده درصد انجام دهید

$$\begin{cases} H_0: \sigma_1^2 = \sigma_2^2 \\ H_1: \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

حل - از ناحیه بحرانی (۱۶) استفاده می کنیم. ملاحظه می شود که

$$\frac{S_1^2}{S_2^2} = \frac{16}{25} = 0.64 \quad F_{0.05}(8, 6) = 0.28 \quad F_{0.95}(8, 6) = 4/15$$

چون

$$F_{0.05}(8, 6) < S_1^2/S_2^2 < F_{0.95}(8, 6)$$

پس دلیلی نداریم که H_0 را رد کنیم.

مثال ۶ با فرض $\sigma_1^2 = \sigma_2^2 = \sigma^2$ در مثال ۵، آزمون زیر را با میزان ده درصد انجام دهید

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

حل - در حقیقت می خواهیم بیازمائیم که نمره ها در هر دو کلاس برابرند. نخست ولریانس مشترک یعنی $\sigma_1^2 = \sigma_2^2 = \sigma^2$ را به وسیله S^2 برآورد می کنیم.

$$S_p^2 = \frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2} = 19/86$$

می توان نشان داد که ناحیه بحرانی آزمون با فرض درست بودن H_0 ، یعنی با فرض $\mu_2 = \mu_1$ ، یعنی $\mu_2 - \mu_1 = 0$ می شود:

$$X - Y < -t_{1-\alpha/2}(m+n-2)S_p\sqrt{\frac{1}{m} + \frac{1}{n}} \quad \text{یا} \quad X - Y > t_{1-\alpha/2}(m+n-2)S_p\sqrt{\frac{1}{m} + \frac{1}{n}}$$

ملاحظه می گردد که

$$\bar{x} - \bar{y} = 85 - 81 = 4 \quad t_{0.95}(14) = 1.76$$

$$t_{0.95}(14)S_p\sqrt{\frac{1}{9} + \frac{1}{9}} = 3.95$$

پس داده ها در نامساوی اول ناحیه بحرانی صدق می کنند. بنابراین فرض H_0 را با میزان ده درصد رد می کنیم.

۵.۲.۷ تمرین بخش دو

۱. در مثال ۵، یک فاصله اطمینان نود درصدی برای $\mu_2 - \mu_1$ پیدا کنید.

۲. در مثال ۵، یک فاصله اطمینان نود درصدی برای σ_2^2 / σ_1^2 پیدا کنید.

۳. ناحیه بحرانی آزمون زیر را

$$\begin{cases} H_0: \sigma^2 = \sigma_1^2 \\ H_1: \sigma^2 > \sigma_1^2 \end{cases}$$

با میزان α برای دو توزیع نرمال مستقل پیدا کنید.

۴. ناحیه بحرانی آزمون زیر را

$$\begin{cases} H_0: \mu_2 = \mu_1 \\ H_1: \mu_2 > \mu_1 \end{cases}$$

با میزان α برای دو توزیع نرمال مستقل پیدا کنید.

۵. یک فاصله اطمینان با ضریب $1-\alpha$ برای $2\mu_1 + 3\mu_2$ در دو توزیع نرمال مستقل پیدا کنید.

۶. در یک آزمون حساب سال پنجم ابتدائی، نمره های ۸ دانش آموز پسر عبارتند از

۱۰، ۱۷، ۱۴، ۱۲، ۱۱، ۱۶، ۱۸، ۱۹

و نمره های ۶ دانش آموز دختر عبارتند از

۱۶، ۱۸، ۱۷، ۱۳، ۱۴، ۱۱

با فرض نرمال بودن نمره ها، با میزان ده درصد بیازمائید که بطور متوسط دانش آموزان پسر و دانش آموزان دختر در درس حساب هم قوه می باشند.

۷. فرض کنید در دو توزیع نرمال مستقل μ_2 و μ_1 معلوم باشند. یک فاصله اطمینان با ضریب $1-\alpha$ برای σ_2^2 / σ_1^2 پیدا کنید.

۸. در تمرین ۶، با میزان ده درصد بیازمائید که واریانسها برابر می باشند.

۹. رابطه $F_{\alpha}(m_1, m_2) = \frac{1}{F_{1-\alpha}(m_2, m_1)}$ را ثابت کنید.

۱۰. طول عمر ۹ باتری اتومبیل ساخت کارخانه A به طور متوسط ۱۳۰۵ ساعت با انحراف معیار ۴۲۰

ساعت، و طول عمر ۱۶ باتری اتومبیل ساخت کارخانه B به طور متوسط ۱۲۱۰ ساعت با انحراف معیار ۳۸۰ ساعت بوده است. تحت فرض نرمال

الف - با میزان ۱۰ درصد بیازمائید که واریانسها برابرند؛

ب - با میزان ۱۰ درصد بیازمائید که میانگین طول عمر برای باتریهای دو کارخانه یکسان است.

ج - یک فاصله اطمینان با ضریب نود درصد برای تفاضل میانگینها پیدا کنید.

۳.۷ آزمونهای گوناگون برای پارامترهای نسبت

پارامتر نسبت و نتیجه گیری آماری برای آن در عمل کاربرد فراوان دارد. مثلاً در یک کارخانه لازم است که مرتب نسبت فرآورده های معیوب را به تمام فرآورده ها گزارش دهند. برای پزشک اهمیت دارد که بدانند چند درصد از بیمارانش با دارویی که تازه به بازار آمده است معالجه شده اند. این پارامتر را می توان به عنوان پارامتر توزیع برنولی یعنی p (شانس پیروزی) تعبیر کرد. اغلب ممکن است لازم شود که دو پارامتر نسبت را با هم مقایسه نمائیم و درباره تفاضل آنها نتیجه گیری آماری کنیم.

در این بخش می خواهیم برای یک پارامتر نسبت با فواصل دو پارامتر نسبت، فاصله اطمینان پیدا کنیم و برای آزمون آنها آزمون آماری انجام دهیم. مطابق معمول ضریب اطمینان را $1 - \alpha$ و میزان آزمون را α می گیریم.

فرض کنید X_1, X_2, \dots, X_n یک نمونه تصادفی از متغیر برنولی X با پارامتر $0 < p < 1$ باشد می دانیم که میانگین و واریانس X عبارتند از p و pq . از طرفی $\sum_{i=1}^n X_i = n - X$ یعنی تعداد پیروزیها در n آزمایش برنولی مستقل (با شانس p برای پیروزی و شانس $q = 1 - p$ برای شکست) دارای توزیع دو جمله ای با میانگین np و واریانس npq می باشد برآوردیاب گشتاوری و MLE پارامتر p می شود

$$\hat{p} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X}$$

افزون بر این بنا بر قضیه حد مرکزی، در صورتی که n به اندازه کافی بزرگ باشد (مثلاً $n > 20$) متغیر تصادفی

$$Q = \frac{\sum_{i=1}^n X_i - np}{\sqrt{npq}} = \frac{\bar{X} - p}{\sqrt{\frac{pq}{n}}} = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

دارای توزیع مجانبی نرمال استاندارد، یعنی تقریباً دارای توزیع $N(0, 1)$ است. بنابراین برای n های بزرگ Q یک تابع محوری است.

۱.۳.۷ فاصله اطمینان برای p

چون تقریباً $Q \sim N(0, 1)$ پس

$$P(-z_{1-\frac{\alpha}{2}} < Q < z_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

از حل نامساوی دو سویی

$$-z_{1-\frac{\alpha}{2}} < \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} < z_{1-\frac{\alpha}{2}}$$

می توان یک فاصله اطمینان برای p به دست آورد ولی چون در این نامساوی p از درجه دو می شود، فاصله ای ساده به دست نمی آید. اگر در مخرج Q به جای p و q برآوردهای آنها یعنی \hat{p} و $\hat{q} = 1 - \hat{p}$ را بگذاریم، می توان ثابت کرد

$$Q^* = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}\hat{q}}{n}}}$$

هم برای n بزرگ تقریباً دارای توزیع $N(0, 1)$ است. از اینرو برای به دست آوردن فاصله اطمینان به جای Q متغیر تصادفی Q^* را به کار می بریم، به آسانی می توان نشان داد که فاصله اطمینان تقریبی می شود

$$\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} \quad (17)$$

فاصله (۱۷) مانند فاصله (۱) برای پارامتر μ در توزیع نرمال می باشد، با این تفاوت که \hat{p} را به جای \bar{X} و $\hat{p}\hat{q}$ (یعنی برآوردیاب واریانس متغیر برنولی) را به جای σ^2 گذارده ایم. چون متغیر برنولی گسسته می باشد، محاسبه دقیق فاصله اطمینان برای n های کوچک آسان نمی باشد و باید از روی جدول توزیع دو جمله ای یا آزمایش و خطا این کار را انجام داد.

۲.۳.۷ آزمون آماری برای p

برای انجام آزمون آماری

$$\begin{cases} H_0: p = p_0 \\ H_1: p < p_0 \end{cases}$$

فرض H_0 را موقعی رد می کنیم که \hat{p} یعنی برآوردیاب p بیش از حد کوچک شود. بنابراین ناحیه بحرانی به وسیله نامساوی $\hat{p} < c$ تعیین می شود. مقدار ثابت c به α بستگی دارد و نمی توان آن را دقیقاً پیدا کرد، زیرا \hat{p} یک متغیر تصادفی گسسته است. از اینرو برای n های بزرگ از تابع محوری $Q^* = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}\hat{q}}{n}}}$

$$Q^* = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}\hat{q}}{n}}}$$

که تحت فرض H_0 به تقریب دارای توزیع $N(0, 1)$ است استفاده می کنیم. اگر Q^* بیش از حد کوچک شود، یعنی $Q^* < k$ ، فرض H_0 را رد می کنیم. برای تعیین k ، با فرض درست بودن H_0 ، از

$$\alpha = P(Q^* < k) = P\left(\frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}\hat{q}}{n}}} < k\right)$$

استفاده می کنیم. بدیهی است که تقریباً $k = -z_{1-\alpha}$. پس ناحیه بحرانی تقریبی می شود

$$\hat{p} < p_0 - z_{1-\alpha} \sqrt{\frac{\hat{p}\hat{q}}{n}} \quad (18)$$

ملاحظه می شود که این ناحیه بحرانی مانند ناحیه بحرانی (۵) برای آزمون

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$$

در مورد توزیع نرمال (σ^2 معلوم) است، با این تفاوت که به جای μ_0 و \bar{X} و σ^2 به ترتیب p_0 و \hat{p} و $\frac{\hat{p}\hat{q}}{n}$ را گذارده ایم.

به همین طریق می توان برای آزمونهای زیر ناحیه بحرانی را پیدا کرد:

$$\begin{cases} H_0: p = p_0 \\ H_1: p > p_0 \end{cases} \quad \begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0 \end{cases}$$

۳.۳.۷ فاصله اطمینان برای $p_2 - p_1$

نمونه تصادفی X_1, X_2, \dots, X_m را از توزیع $B(1, p_1)$ و نمونه تصادفی Y_1, Y_2, \dots, Y_n را از توزیع $B(1, p_2)$ در نظر می گیریم. فرض می کنیم این دو توزیع برنولی و همچنین این دو نمونه از هم مستقل باشند.

بدیهی است که برای m و n بزرگ تقریباً داریم:

$$\hat{p}_1 \sim N(p_1, \frac{p_1 q_1}{m}) \quad \hat{p}_2 \sim N(p_2, \frac{p_2 q_2}{n})$$

چون این دو نمونه از هم مستقل می باشند، پس تقریباً داریم

$$\hat{p}_2 - \hat{p}_1 \sim N(p_2 - p_1, \frac{p_1 q_1}{m} + \frac{p_2 q_2}{n})$$

با استفاده از تابع محوری

$$Q = \frac{\hat{p}_2 - \hat{p}_1 - (p_2 - p_1)}{\sqrt{\frac{p_1 q_1}{m} + \frac{p_2 q_2}{n}}} \stackrel{D}{=} Z \sim N(0, 1)$$

می توان یک فاصله اطمینان برای $p_2 - p_1$ به دست آورد. برای این منظور نامساوی

$$-z_{1-\frac{\alpha}{2}} < Q < z_{1-\frac{\alpha}{2}}$$

را برحسب $p_2 - p_1$ حل می کنیم. در نتیجه داریم

$$\hat{p}_2 - \hat{p}_1 - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{m} + \frac{\hat{p}_2 \hat{q}_2}{n}} < p_2 - p_1 < \hat{p}_2 - \hat{p}_1 + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_1 \hat{q}_1}{m} + \frac{\hat{p}_2 \hat{q}_2}{n}}$$

فاصله اطمینان بالا مانند فاصله (۱۱) برای $\mu_2 - \mu_1$ در مورد دو توزیع نرمال مستقل می باشد.

۳.۳.۷ آزمون آماری $p_2 = p_1$

برای انجام آزمون

$$\begin{cases} H_0: p_1 = p_2 = p \\ H_1: p_1 < p_2 \end{cases}$$

از تابع محوری Q در بالا استفاده می کنیم. تحت فرض H_0 ، $\hat{p}_2 - \hat{p}_1$ دارای واریانس $p(\frac{1}{m} + \frac{1}{n})$ می باشد. به جای این واریانس برآوردیاب آن را به کار می بریم.

بدیهی است که $\hat{p} = (m\hat{p}_1 + n\hat{p}_2)/(m+n)$ در نتیجه به

$$Q^* = \frac{\hat{p}_2 - \hat{p}_1}{\sqrt{\hat{p}\hat{q}(\frac{1}{m} + \frac{1}{n})}}$$

تبدیل می شود. اگر $\hat{p}_2 - \hat{p}_1$ یا Q^* بیش از حد بزرگ شود، فرض H_0 را رد می کنیم. پس ناحیه بحرانی می شود

$$\hat{p}_2 - \hat{p}_1 > z_{1-\alpha} \sqrt{\hat{p}\hat{q}(\frac{1}{m} + \frac{1}{n})} \quad (۲۰)$$

به همین طریق می توان برای آزمونهای زیر ناحیه بحرانی پیدا کرد

$$\begin{cases} H_0: p_1 = p_2 = p \\ H_1: p_1 > p_2 \end{cases} \quad \begin{cases} H_0: p_1 = p_2 = p \\ H_1: p_1 \neq p_2 \end{cases}$$

مثال ۷ در شهری بزرگ پیش از رای گیری برای انتخاب شخصی به عنوان نماینده، یک آمارگیری ساده انجام داده اند. در یک نمونه تصادفی ۱۰۰ نفری ۵۹ رای موافق مشاهده شده است. یک فاصله اطمینان برای نسبتی از رای دهندگان موافق در این شهر پیدا کنید.

حل - فرض کنید p نسبت مورد نظر باشد. بنابراین احتمال اینکه هر یک از رای دهندگان رای موافق دهند p می باشد. در یک نمونه تصادفی از توزیع برنولی، با شانس p برای پیروزی، ۵۹ پیروزی داریم. از اینرو $\hat{p} = ۰/۵۹$. اگر ضریب اطمینان را $۰/۹۵$ بگیریم و از فاصله (۱۷) استفاده کنیم، دو سر این فاصله عبارتند از

$$\hat{p} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}\hat{q}}{n}} = ۰/۵۹ \pm ۱/۹۶ (۰/۰۲۹)$$

بنابراین یک فاصله اطمینان ۹۵ درصدی برای p می شود $(۰/۴۹۶, ۰/۶۸۶)$.

مثال ۸ در یک نمونه تصادفی ۲۰۰ نفری از شهر A شماره با سوادان ۱۲۰ نفر و در یک نمونه تصادفی ۵۰۰ نفری از شهر B این شماره ۲۴۰ نفر است. با میزان پنج درصد بیازمائید که هر دو شهر به کمک نسبت با سواد دارند.

حل - نسبت با سوادان در این دو شهر به ترتیب عبارتند از

$$\hat{p}_1 = \frac{120}{200} = 0/60$$

$$\hat{p}_2 = \frac{240}{500} = 0/48$$

بر اساس این دو برآورد به نظر می آید که نسبت با سوادان در شهر A بیشتر می باشد بنابراین آزمون زیر را انجام می دهیم:

$$\begin{cases} H_0: p_1 = p_2 = p \\ H_1: p_1 > p_2 \end{cases}$$

ناحیه بحرانی را می توان مانند ناحیه بحرانی (۲۰) پیدا کرد. فرض H_0 را موقعی رد می کنیم که

$$\hat{p}_2 - \hat{p}_1 < -z_{1-\alpha} \sqrt{\hat{p}\hat{q} \left(\frac{1}{m} + \frac{1}{n} \right)}$$

بدیهی است که برآورد p می شود

$$\hat{p} = \frac{120 + 240}{200 + 500} = 0/51$$

زیرا تحت فرض H_0 نسبت با سوادان در هر دو شهر یکسان است و عملاً در یک نمونه تصادفی ۷۰۰ نفری ۳۶۰ نفر با سواد مشاهده کرده ایم. طرف چپ نامساوی بالا می شود ۰/۱۲ - و طرف راست آن ۰/۶۸۶ - . بنابراین نامساوی برقرار است و فرض H_0 با میزان پنج درصد رد می شود.

۵.۳.۷ تمرین بخش سه

۱. در مثال ۷ با میزان ده درصد بیازمائید که پنجاه درصد از رای دهندگان رای موفق می دهند در این

مثال P - مقدار را بیابید.

۲. در مثال ۸ یک فاصله اطمینان نود درصدی برای $p_1 - p_2$ پیدا کنید.

۳. ناحیه بحرانی آزمون دوسوئی زیر را برای پارامتر توزیع برنولی با میزان α چگونه انجام می دهید؟

$$\begin{cases} H_0: p = p_0 \\ H_1: p \neq p_0 \end{cases}$$

۴. پزشکی مدعی است که هفتاد درصد از بیماران قلبی با داروئی که اخیراً کشف شده است بهبود می یابند. اگر ۸۰ بیمار از ۱۵۰ بیمار با این دارو بهبود یابند، با میزان ده درصد ادعای پزشک را بیازمائید. P - مقدار را بیابید

۵. در یک نمونه تصادفی از ۵۰۰ دانشجوی پسر ۳۰۰ نفر طرفدار کلاسهای بعد از ظهر، و در یک نمونه تصادفی از ۱۰۰ دانشجوی دختر ۶۲ نفر طرفدار این موضوع هستند. با میزان ده درصد بیازمائید که پسر و دختر هم رای می باشند.

۶ تولیدکنندهای، کاغذ را کارتتهای ۱۰۰۰ برگی تحویل می دهد. در ۲۵ کارت، ۸ کارت کمتر از ۱۰۰۰ برگ دارند. فرض کنید پارامتر p درصد کارتتهای با کمتر از ۱۰۰۰ برگ باشد. آزمون زیر را با میزان $\alpha = 0/10$ انجام دهید

$$\begin{cases} H_0: p = 0/25 \\ H_1: p > 0/25 \end{cases}$$

۷. پروندههای یک بیمارستان نشان می دهد که از ۱۰۰۰۰ مرد ۵۲ نفر و از ۱۰۰۰۰ زن ۲۳ نفر بیماری قلبی دارند. آیا بیماران قلبی مرد بیش از بیماران قلبی زن می باشند؟ از P - مقدار استفاده کنید و α را ده درصد بگیرید

۸. در تمرین ۷ یک فاصله اطمینان با ضریب نود درصدی برای تفاضل نسبتهای بیماران قلبی مرد و زن بیابید

۹. تولیدکنندهای می خواهد دو نوع حشره کش A و B را با هم مقایسه نماید. دو اطاق هم اندازه هر یک با ۱۰۰۰ حشره در نظر گرفته است و در یک اطاق بستهای از حشره کش A و در اطاق دیگر بستهای از حشره کش B مصرف می کنند. تعداد ۸۲۵، ۷۶۰ حشره به ترتیب با حشره کش A و B نابود می شوند. یک فاصله اطمینان نود درصدی برای تفاضل نسبتها پیدا کنید

۱۰. در تمرین ۹ با $\alpha = 0/10$ بیازمائید که حشره کش A موثرتر از حشره کش B می باشد

و - اگر X و Y مستقل باشند، $\rho = 0$ ولی اگر $\rho = 0$ ممکن است مستقل نباشند. در این بخش می‌خواهیم پارامتر ρ را با روش گشتاوری برآورد کنیم.

۱.۱.۸ برآورد پارامتر ρ

نمونه تصادفی $(X_1, Y_1), (X_2, Y_2), (Y_3, Y_3), \dots, (X_n, Y_n)$ را، بسا یافته‌های $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ از جفت (X, Y) در نظر می‌گیریم. مثلاً اگر X و Y به ترتیب اندازه قد و اندازه وزن یک کودک پنج ماهه باشند، با انتخاب ده کودک به تصادف یک نمونه تصادفی ده تایی به دست می‌آید که یافته‌های آن بعد از اندازه‌گیری مشخص می‌شوند. برای اینکه ρ را با روش گشتاوری برآورد کنیم. نخست σ_X^2 و σ_Y^2 و σ_{XY} را به کمک آماره‌های زیر برآورد می‌کنیم:

$$\hat{\sigma}_X^2 = S_X^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \bar{X}^2 - \bar{X}^2$$

$$\hat{\sigma}_Y^2 = S_Y^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2 = \bar{Y}^2 - \bar{Y}^2$$

$$\hat{\sigma}_{XY} = S_{XY} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) = \bar{XY} - \bar{X}\bar{Y}$$

اینک پارامتر ρ را به یاری

$$R = \frac{S_{XY}}{S_X S_Y}$$

برآورد می‌کنیم. یافته‌های آماره‌های بالا را با حروف کوچک به صورت زیر نشان می‌دهیم:

$$s_x^2 = \bar{x}^2 - \bar{x}^2$$

$$s_y^2 = \bar{y}^2 - \bar{y}^2$$

$$s_{xy} = \bar{xy} - \bar{x}\bar{y}$$

$$r = \frac{s_{xy}}{s_x s_y}$$

لازم است که تقارن میان ρ و R را شرح دهیم:

فصل ۸

ضریب همبستگی و خط برگشت

در این فصل ضریب همبستگی و خط برگشت را که با هم ارتباط نزدیک دارند با زبانی ساده شرح می‌دهیم. این مفاهیم در رشته‌های علوم، مهندسی، و علوم تربیتی و اجتماعی دارای کاربرد فراوان می‌باشند. بنابراین می‌کوشیم تا این واژه‌های آماری را با ریاضی مقدماتی روشن کنیم. در بخش ۱، ضریب همبستگی و طرز محاسبه آن را از نظر احتمال و از نظر آمار بیان می‌داریم. در بخش ۲ خط برگشت یا خط رگرسیون را تشریح می‌کنیم.

۱.۸ برآورد ضریب همبستگی

در بخش ۶ فصل چهارم، مفهوم ضریب همبستگی دو متغیر تصادفی X و Y را، که دارای چگالی توام $f(x, y)$ هستند، شرح دادیم. این ضریب را که در حقیقت یکی از ویژگی‌های چگالی توام است با پارامتر ρ نشان دادیم و به صورت زیر تعریف کردیم:

$$\rho = \rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

به ویژه نشان دادیم که:

الف - $-1 \leq \rho \leq 1$ یا $|\rho| \leq 1$

ب - برای $\rho > 0$ داریم:

$$\rho(aX + b, cY + d) = \rho(X, Y)$$

به عبارت دیگر ضریب همبستگی با تغییر مبدأ اندازه‌گیری و تغییر واحد اندازه‌گیری تغییر نمی‌کند.

ج - برای $\rho = \pm 1$ رابطه خطی زیر میان X و Y برقرار است:

$$\frac{X - \mu_X}{\sigma_X} = \pm \frac{Y - \mu_Y}{\sigma_Y}$$

۱. ρ یک پارامتر می باشد و تنها با داشتن فرمول چگالی توام (X, Y) می توان آن را محاسبه کرد ولی چون معمولاً چگالی توام را نمی دانیم، ρ در دست نیست و آن را برآورد می کنیم. این پارامتر را اغلب ضرب همبستگی جمعیت می نامند.

۲. R یک متغیر تصادفی است که به عنوان برآوردیاب ρ به کار می رود. توزیع این برآوردیاب را نمی توان در حالت کلی پیدا کرد در حالتی خاص هم این توزیع چندان ساده نمی باشد.

۳. r ، یعنی یافته R ، یک عدد است و در حقیقت یک برآورد برای پارامتر بر اساس یافته های نمونه تصادفی می باشد. این عدد را ضرب همبستگی نمونه می نامند.

۲.۱.۸ ویژگیهای برآورد r

این برآورد مانند ρ دارای ویژگیهای زیر است:

الف - با تغییر مبدا اندازه گیری و تغییر واحد اندازه گیری، مقدار r تغییر نمی کند. برای اثبات، تغییر مبدا و واحد را با تبدیل زیر نشان می دهیم:

$$x'_i = ax_i + b \quad a > 0$$

$$y'_i = cy_i + d \quad c > 0$$

این تبدیل $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ را به $(x'_1, y'_1), (x'_2, y'_2), \dots, (x'_n, y'_n)$ تبدیل می کند. با یافته های تبدیل شده داریم

$$r' = \frac{s_{x'} s_{y'}}{s_{x'} s_{y'}} = \frac{ac s_{xy}}{a s_x c s_y} = \frac{s_{xy}}{s_x s_y} = r$$

مثلاً فرض کنید اندازه قد یک کودک پنج ماهه بر حسب اینچ و اندازه وزن او بر حسب پوند باشد حال اگر اندازه قد بر حسب سانتیمتر با مبدا بیست سانتیمتر و او را اندازه وزن بر حسب کیلوگرم با مبدا نیم کیلوگرم فرض می کنیم. تبدیل زیر به دست می آید:

$$x' = 2/45x - 20$$

$$y' = 0/45y - 0/5$$

با این تبدیل مقدار r تغییر نمی کند.

ب - همواره داریم: $-1 \leq r \leq 1$

برای اثبات $r \leq 1$ نامساوی زیر را به کار می بریم:

$$\sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s_x} - \frac{y_i - \bar{y}}{s_y} \right)^2 \geq 0$$

طرف چپ را بسط می دهیم و از تعریف \bar{x} و \bar{y} و s_x استفاده می کنیم. در نتیجه داریم:

$$n - 2nr + n \geq 0$$

یا

$$r \leq 1$$

برای اثبات $r \geq -1$ در نامساوی بالا - را به + تبدیل می کنیم.

ج - با فرض $r = \pm 1$ نقاط $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ روی یک خط راست قرار می گیرند.

برای اثبات، ملاحظه می شود که اگر $r = 1$ و تنها اگر نامساوی بالا به تساوی تبدیل شود یا اگر و تنها اگر داشته باشیم

$$\frac{x_i - \bar{x}}{s_x} - \frac{y_i - \bar{y}}{s_y} = 0 \quad i = 1, 2, \dots, n$$

بنابراین اگر $r = 1$ ، نقاط (x_i, y_i) روی خط

$$\frac{x - \bar{x}}{s_x} - \frac{y - \bar{y}}{s_y} = 0$$

که دارای ضرب زاویه مثبت است قرار می گیرند.

به همین طریق ثابت می کنیم که اگر $r = -1$ نقاط (x_i, y_i) روی خط

$$\frac{x - \bar{x}}{s_x} + \frac{y - \bar{y}}{s_y} = 0$$

که دارای ضرب زاویه منفی است قرار می گیرند.

ویژگیهای بالا نشان می دهند که r شدت همبستگی خط میان دو متغیر را بیان می دارد، و تغییر مبدا

و واحد اندازه گیری r را تغییر نمی دهد.

$$\sigma_W^2 = \frac{1}{n-3}$$

می باشد. از اینرو به طور تقریبی داریم

$$Z = \frac{W - \mu_W}{\sigma_W} \sim N(0, 1)$$

با استفاده از این توزیع تقریبی می توان برای ρ فاصله اطمینان پیدا کرد و درباره آن آزمون آماری انجام داد

مثال ۲ برای یک نمونه ۱۰۳ تایی از (X, Y) داریم $\rho = 0/5$. آزمون زیر را با میزان $\alpha = 0/05$ انجام دهید:

$$\begin{cases} H_0: \rho = 0/6 \\ H_1: \rho \neq 0/6 \end{cases}$$

حل - با استفاده از تبدیل فیشر داریم

$$W = \frac{1}{2} \ln \frac{1+0/5}{1-0/5} = 0/55$$

$$\mu_W = \frac{1}{2} \ln \frac{1+0/6}{1-0/6} = 0/69$$

$$\sigma_W^2 = \frac{1}{n-3} = \frac{1}{100}$$

$$z = \frac{W - \mu_W}{\sigma_W} = \frac{0/55 - 0/69}{0/1} = -1/4$$

از روی جدول نرمال داریم $z_{1-\alpha/2} = z_{0/975} = 1/96$. بنابراین ناحیه بحرانی به صورت $|z| > 1/96$ است. چون $|z| = 1/4$ در ناحیه بحرانی نمی باشد، پس فرض H_0 را با میزان پنج درصد رد نمی کنیم.

۴.۱.۸ توزیع نرمال توام

جفت (Z_1, Z_2) دارای توزیع نرمال توام استاندارد است. هرگاه چگالی توام آن به صورت زیر باشد:

$$f(z_1, z_2; \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(z_1^2 - 2\rho z_1 z_2 + z_2^2)} \quad |\rho| \leq 1$$

مثال ۱ فرض کنید X نمره درس ریاضی و Y نمره درس فارسی یک دانش آموز اول راهنمایی باشد جدول زیر نمره های شش دانش آموز را نشان می دهد.

| | | | | | | |
|------------------|----|----|----|----|----|----|
| نمره ریاضی X_i | ۱۲ | ۱۰ | ۱۴ | ۱۱ | ۱۲ | ۹ |
| نمره فارسی Y_i | ۱۸ | ۱۷ | ۲۳ | ۱۹ | ۲۰ | ۱۵ |

برآورد ضریب همبستگی را حساب کنید.

حل - یافته های نمونه تصادفی عبارتند از:

$$(12, 18), (10, 17), (14, 23), (11, 19), (12, 20), (9, 15)$$

بعد از انجام محاسبات لازم داریم:

$$\bar{x} = 11/333, \quad \bar{y} = 18/666, \quad \bar{x}^2 = 131, \quad \bar{y}^2 = 352/667, \quad \bar{xy} = 215/333$$

$$s_x^2 = 2/555, \quad s_y^2 = 6/222, \quad s_{xy} = 3/777$$

$$r = \frac{s_{xy}}{s_x s_y} = 0/947$$

چون این ضریب به یک نزدیک می باشد، پس رابطه خطی میان نمره و فارسی شدید است، و به طور متوسط این دو نمره در جهت یکدیگر افزایش یا کاهش می یابند.

۴.۱.۸ توزیع تقریبی R

فیشر، آماردان انگلیسی، تبدیل زیر را که به تبدیل فیشر شهرت دارد به کار برده است:

$$W = \frac{1}{2} \ln \frac{1+R}{1-R}$$

او ثابت کرده است که برای نمونه های بزرگ (مثلاً $n \geq 25$) تقریباً دارای توزیع نرمال با میانگین و واریانس تقریبی

$$\mu_w = \frac{1}{2} \ln \frac{1+\rho}{1-\rho}$$

می توان ثابت کرد که Z_1 و Z_2 هر کدام دارای توزیع نرمال استاندارد هستند و ضرب همبستگی آنها برابر است با پارامتر ρ . به ویژه Z_1 و Z_2 با چگالی توام بالا، مستقل اند اگر و تنها اگر $\rho = 0$. حال تبدیل زیر را در نظر می گیریم:

$$\begin{cases} X = \sigma_1 Z_1 + \mu_1 \\ Y = \sigma_2 Z_2 + \mu_2 \end{cases}$$

واضح است که $X \sim N(\mu_1, \sigma_1^2)$ و $Y \sim N(\mu_2, \sigma_2^2)$ در ضمن ضرب همبستگی X و Y برابر است با ضرب همبستگی Z_1 و Z_2 یعنی ρ . می گوئیم جفت (X, Y) دارای توزیع نرمال توام یا توزیع نرمال دو بعدی است. با فرض $\mu_1 = \mu_2 = 0$ و $\sigma_1^2 = \sigma_2^2 = 1$ ، جفت (X, Y) هم دارای توزیع نرمال توام استاندارد می باشد. جفت نرمال توام دارای ویژگی مهم زیر است:

X و Y ، با توزیع نرمال توام، مستقل اند، اگر و تنها اگر $\rho = 0$.

به عنوان مثال اندازه قد یک جوان را یا X و اندازه وزن او را با Y نشان می دهیم. معمولاً جفت (X, Y) دارای توزیع نرمال توام است و X و Y مستقل نیستند.

۵.۱.۸ چگالی R تحت فرض نرمال

در حالت خاصی که X و Y هر دو نرمال و مستقل باشند، می توان ثابت کرد که آماره

$$V = \frac{R\sqrt{n-2}}{\sqrt{1-R^2}}$$

دقیقاً دارای توزیع $T(n-2)$ است. با استفاده از آماره V و ویژگی جفت نرمال توام (X, Y) می توان مستقل بودن X و Y را آزمون کرد. این آزمون معادل با آزمون $\rho = 0: H_0$ می باشد.

مثال ۳ فرض کنید X نمره زبان و Y نمره ریاضی یک دانشجوی سال اول باشد و بدتیم که جفت (X, Y) دارای توزیع نرمال توام است. برای یک نمونه تصادفی ده تایی از این دانشجویان داریم $\rho = 0.1$. با میزان پنج درصد مستقل بودن نمره های این دو درس را بیازمائید.

حل - آزمون زیر را انجام می دهیم، زیرا پانته R خیلی کوچک است.

$$\begin{cases} H_0: \rho = 0 \\ H_1: \rho \neq 0 \end{cases}$$

یافته آماره V برابر است با

$$v = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.1\sqrt{10-2}}{\sqrt{1-0.01}} = 0.282$$

آماره V با فرض نرمال بودن و مستقل بودن X و Y دارای توزیع $T(8)$ است. از روی جدول T داریم $t_{0.05}(8) = 1.96$. چون یافته V در ناحیه بحرانی نمی باشد، فرض مستقل بودن X و Y را با میزان پنج درصد رد نمی کنیم.

۶.۱.۸ تمرین بخش یک

۱. جدول زیر نمره های ۷ دانشجو در امتحان ریاضی و زبان می باشد.

| | | | | | | | |
|------------|----|----|----|----|----|----|----|
| نمره ریاضی | ۱۲ | ۱۳ | ۱۰ | ۱۳ | ۱۵ | ۱۱ | ۱۲ |
| نمره زبان | ۱۲ | ۱۴ | ۱۳ | ۱۵ | ۱۱ | ۱۴ | ۱۱ |

۲. را حساب کنید. به هر نمره ریاضی ۲ اضافه می کنیم و هر نمره زبان را در $1/25$ ضرب می نمائیم.

برای نمره های تبدیل شده ضرب همبستگی را حساب کنید و با ۲ مقایسه نمائید.

۲. برای داده های $(75, 50)$ و $(50, 25)$ و $(100, 125)$ و $(125, 50)$ یک تبدیل مناسب به کار برید و سپس ضرب همبستگی را حساب کنید.

۳. فرض کنید در تمرین ۱، نمره ریاضی و نمره زبان دارای توزیع نرمال توام باشند. مستقل بودن این دو نمره را با میزان ده درصد بیازمائید.

۴. ثابت کنید: $|s_{xy}| \leq s_x s_y$

۵. آیا $\rho = 0$ مستلزم $r = 0$ می باشد؟ چرا؟

۶. برای یک جفت تصادفی از (X, Y) ، با فرض $\mu_x = \mu_y = 1$ و $\sigma_x^2 = \sigma_y^2 = 2$ ، داریم $\rho = -1$. چه رابطه ای میان X و Y برقرار است؟

۷. ثابت کنید اگر $r = -1$ ، آنگاه (x_i, y_i) ها روی یک خط راست جا دارند. ضرب همبستگی را برای داده های $(2, 7)$ و $(0, 3)$ و $(-1, 1)$ و $(1, 3)$ و $(-2, -1)$ بیابید. آیا این نقاط روی یک خط راست قرار دارند؟

۸. ضریب همبستگی میان قد و وزن ۱۰۰ نوزاد برابر است با $\rho = 0.3$ آزمون زیر را با میزان ده درصد انجام دهید.

$$\begin{cases} H_0: \rho = 0.25 \\ H_1: \rho > 0.25 \end{cases}$$

آزمون را با پیدا کردن ناحیه بحرانی و با استفاده از p - مقدار انجام دهید.

۹. در مثال ۲ یک فاصله اطمینان با ضریب نود درصد برای ρ بیابید.

۱۰. اگر جفت (X, Y) دارای توزیع نرمال توأم باشد، $X \pm Y$ دارای توزیع نرمال است. میانگین و واریانس $X \pm Y$ را پیدا کنید و ضریب همبستگی $X + Y$ و $X - Y$ را بیابید.

۱۱. برای مثال ۳، با پیدا کردن ناحیه بحرانی و همچنین محاسبه p - مقدار آزمون زیر را انجام دهید.

$$\begin{cases} H_0: \rho = 0 \\ H_1: \rho > 0 \end{cases}$$

اگر داده‌ها از توزیع نرمال دو بعدی باشند، آزمون را چگونه انجام می‌دهید؟

۲.۸ خط پرگشت

نمره کنکور دانشجویی که در مهرماه ثبت نام کرده است ۸۳ می‌باشد. آیا می‌توان درباره معدل سالانه او پیش‌بینی کرد؟ تحت شرائطی، با استفاده از تجربیات گذشته، ممکن است این کار را انجام داد. ولی معدل سالانه او، که در پایان سال تحصیلی گزارش خواهد شد، فعلاً یک متغیر تصادفی است که به عوامل بیشمار از جمله نمره کنکور بستگی دارد.

در این مثال دو متغیر داریم: نخست، نمره کنکور که از آغاز آن را دقیقاً می‌دانیم، این متغیر غیر تصادفی را با x نشان می‌دهیم و آن را متغیر مستقل یا قابل کنترل می‌نامیم. دوم، معدل سالانه که یک متغیر تصادفی، وابسته به نمره کنکور، می‌باشد، این متغیر تصادفی شرطی را با $Y|x$ نشان می‌دهیم و آن را متغیر وابسته می‌نامیم.

برای سادگی، فرض می‌کنیم $Y|x$ یک تابع خطی از x به صورت $a + bx$ و واریانس آن σ^2 باشد. البته این میانگین و واریانس ممکن است علاوه بر x به متغیرهای غیر تصادفی دیگر هم بستگی داشته باشند یا به صورت خطی نباشند. فعلاً در این درس، ما همین حالت ساده را مطالعه می‌کنیم.

در این حالت ساده ما انتظار داریم که معدل سالانه برای افرادی که نمره کنکور آنها x است، به طور متوسط بشود $a + bx$. در اثر عوامل تصادفی به اندازه $Y|x - (a + bx)$ اختلاف تصادفی به وجود می‌آید. این اختلاف تصادفی را با متغیر تصادفی E ، نشان می‌دهیم و آن را خطای تصادفی می‌نامیم.

۱.۲.۸ مدل خطی ساده

$$Y|x = a + bx + E$$

را یک مدل خطی ساده می‌گویند. در این مدل، متغیر غیر تصادفی x متغیر مستقل است و از آغاز آن را دقیقاً می‌دانیم. متغیر تصادفی $Y|x$ متغیر وابسته به x است و دارای میانگین $a + bx$ و واریانس σ^2 می‌باشد. متغیر تصادفی E خطای تصادفی است و دارای میانگین صفر و واریانس σ^2 می‌باشد. پارامترهای این مدل عبارتند از a و b و σ^2 .

۲.۲.۸ یک نمونه از متغیر مستقل و متغیر وابسته یا مدل خطی ساده

در مدل خطی بالا برای متغیر مستقل x مقادیر ثابت x_1, x_2, \dots, x_n را در نظر می‌گیریم. فرض می‌کنیم این مقادیر با متغیرهای تصادفی شرطی $Y|x_1, Y|x_2, \dots, Y|x_n$ متناظر باشند. ما این متغیرهای تصادفی شرطی را برای راحتی با Y_1, Y_2, \dots, Y_n و یافته‌های آنها را با y_1, y_2, \dots, y_n نشان می‌دهیم. برای هر $i = 1, 2, \dots, n$ داریم:

$$Y_i = a + bx_i + E_i$$

$$E(Y_i) = a + bx_i, \quad Var(Y_i) = \sigma^2$$

$$E(E_i) = 0, \quad Var(E_i) = \sigma^2$$

به طور خلاصه یک نمونه از متغیرهای مستقل و وابسته به صورت زیر است:

$$(x_1, Y_1), (x_2, Y_2), \dots, (x_n, Y_n)$$

در این نمونه x_i ها از آغاز معلوم هستند ولی هر Y_i متغیر تصادفی است. بعد از مشاهده، یافته‌های زیر برای نمونه بالا به دست می‌آیند:

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

۳.۲.۸ روش کمترین توانهای دوم برای برآورد a و b

فرض می‌کنیم

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

یافته‌های یک نمونه در مدل خطی باشد. مثلاً آنها را نمرهای کنکور n دانشجو می‌گیریم. این نمرها را هنگام ثبت نام در دست داریم. بعد از پایان سال تحصیلی، معدل سالانه آنها به صورت y_i (یعنی یافته‌های متغیرهای تصادفی Y_i ها) گزارش گردیده‌اند. خطای تصادفی e_i را با استفاده از y_i به جای Y_i ، با

$$e_i = y_i - (a + bx_i)$$

نشان می‌دهیم. در روش کمترین توانهای دوم، برای اینکه مدل خطی برای نمونه بالا مصداق پیدا کند و برازنده آن باشد، مجموع توانهای دوم خطاها، یعنی عبارت زیر که تابعی از a و b می‌باشد، باید می‌نیمم گردد:

$$A(a, b) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n [y_i - (a + bx_i)]^2$$

برای می‌نیمم کردن تابع $A(a, b)$ می‌توان از مشتق این تابع نسبت به a و b استفاده کرد. ولی ما با استفاده از جبر مقدماتی، می‌نیمم این تابع را به روش زیر پیدا می‌کنیم. نخست نمادهای زیر را تعریف می‌نمائیم:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad s_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$$

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$r = \frac{s_{xy}}{s_x s_y}$$

اینک با افزودن و کاستن \bar{x} و همچنین \bar{y} می‌نویسیم:

$$\begin{aligned} A(a, b) &= \sum_{i=1}^n [y_i - \bar{y} - b(x_i - \bar{x}) + \bar{y} - a - b\bar{x}]^2 \\ &= ns_y^2 + nb^2 s_x^2 + n(\bar{y} - a - b\bar{x})^2 - 2bns_{xy} \\ &= ns_x^2 \left(b - \frac{s_{xy}}{s_x}\right)^2 + n(\bar{y} - a - b\bar{x})^2 + ns_y^2 - n\frac{s_{xy}^2}{s_x^2} \\ &\geq n\left(s_y^2 - \frac{s_{xy}^2}{s_x^2}\right) = A \end{aligned}$$

بدیهی است که می‌نیمم مقدار $A(a, b)$ موقعی بدست می‌آید که a و b در روابط زیر صدق کنند:

$$\begin{cases} b - \frac{s_{xy}}{s_x} = 0 \\ \bar{y} - a - b\bar{x} = 0 \end{cases}$$

از حل دستگاه بالا برحسب a و b ، برآوردهای این دو پارامتر می‌شوند:

$$\begin{cases} \hat{b} = \frac{s_{xy}}{s_x^2} \\ \hat{a} = \bar{y} - \hat{b}\bar{x} \end{cases}$$

به کمک تساویهای زیر معلوم می‌شود که \hat{b} ترکیبی خطی از y_i ها است.

$$\hat{b} = \frac{s_{xy}}{s_x^2} = \sum_{i=1}^n \frac{(x_i - \bar{x})(y_i - \bar{y})}{ns_x^2} = \sum_{i=1}^n \frac{x_i - \bar{x}}{ns_x^2} y_i$$

بنابراین \hat{a} و \hat{b} هر دو ترکیبهای خطی از y_1, y_2, \dots, y_n هستند.

هرگاه در \hat{a} و \hat{b} به جای y_i بگذاریم Y_i ، مطابق معمول برآوردهای a و b ، که دو متغیر تصادفی هستند، به دست می‌آیند. ما برای برآوردهای \hat{a} و \hat{b} را به کار می‌بریم. از متن مطلب می‌فهمیم که کجا برآورد و کجا برآوردیاب داریم.

به آسانی می‌توان نشان داد که برای دو متغیر تصادفی \hat{a} و \hat{b} داریم:

$$E(\hat{a}) = a \quad E(\hat{b}) = b$$

یعنی این دو برآوردیاب ناریب می‌باشند.

۴.۲.۸ پیش‌بینی متغیر وابسته به طور متوسط

مرگاه در

$$E(Y|x) = a + bx$$

به جای a و b برآوردهای آنها را بگذاریم، برآورد این میانگین به دست می‌آید ما این برآورد را با $\hat{Y} = \hat{a} + \hat{b}x$ نشان می‌دهیم. با استفاده از این رابطه می‌توانیم با داشتن x (مثلاً نمره کنکور)، مقدار متغیر وابسته را (مثلاً معدل سالانه را) به طور متوسط به وسیله \hat{Y} پیش‌بینی کنیم.

۵.۲.۸ برآورد خطای تصادفی

فرض کنید x_i مقدار متغیر مستقل و y_i یافته متغیر وابسته باشد. حال اگر متغیر وابسته را با داشتن x_i به وسیله $\hat{y}_i = \hat{a} + \hat{b}x_i$ پیش‌بینی کنیم، برآورد خطای تصادفی می‌شود

$$e_i = y_i - \hat{y}_i = y_i - \hat{a} - \hat{b}x_i$$

عدد e_i را پس مانده Residual وگرسون می‌گویند. اگر در این رابطه \hat{a} و \hat{b} دو برآورد کننده باشند و Y_i به \hat{Y}_i تبدیل کنیم، آنگاه E_i می‌شود یک متغیر تصادفی. امید ریاضی E_i برابر صفر است.

معمولاً $\sum_{i=1}^n E_i^2 = SSE$ یعنی مجموع توانهای دوم برآورد خطاها را با SSE که علامت اختصاری Sum of the Squares of Errors است نشان می‌دهند. بدیهی است که یافته SSE برابر است با می‌نیم تابع $A(a, b)$ یعنی

$$A(\hat{a}, \hat{b}) = SSE = n(s_y^2 - \frac{s_{xy}^2}{s_x^2})$$

۶.۲.۸ برآورد σ^2

می‌دانیم که

$$E(E_i^2) = Var(E_i) + E^2(E_i) = \sigma^2$$

چون E_i برآورد کننده E_i می‌باشد، پس می‌توانیم σ^2 را به وسیله

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n E_i^2 = \frac{1}{n} SSE$$

برآورد کنیم. ملاحظه می‌شود که $\hat{\sigma}^2$ درست برابر است با $\frac{1}{n}$ می‌نیم مجموع توانهای دوم خطاها.

پس داریم:

$$\hat{\sigma}^2 = s_y^2 - \frac{s_{xy}^2}{s_x^2} = s_y^2 - \hat{b}^2 s_x^2$$

مرگاه در $\hat{\sigma}^2$ به جای y_i بگذاریم Y_i ، برآوردیاب σ^2 به دست می‌آید می‌توان ثابت کرد که:

$$E(\hat{\sigma}^2) = \frac{n-2}{n} \sigma^2$$

بنابراین یک برآورد نااریب برای σ^2 می‌شود:

$$\hat{\sigma}^2 = \frac{n}{n-2} \hat{\sigma}^2 = \frac{n}{n-2} (s_y^2 - \hat{b}^2 s_x^2)$$

۷.۲.۸ خط برگشت

در بالا دیدیم که با داشتن x ، می‌توانیم میانگین متغیر وابسته را به کمک خط $\hat{y} = \hat{a} + \hat{b}x$ پیش‌بینی کنیم. با تغییر x نقاط (x, \hat{y}) روی این خط، که آنرا خط برگشت یا خط وگرسون می‌نامیم، قرار می‌گیرند. با جایگذاری \hat{a} و \hat{b} معادله این خط را می‌توان به طریق دیگر هم نوشت:

$$\hat{y} = (\bar{y} - \hat{b}\bar{x}) + \hat{b}x$$

$$\hat{y} - \bar{y} = \hat{b}(x - \bar{x})$$

چون

$$\hat{b} = \frac{s_{xy}}{s_x^2} = \frac{s_{xy}}{s_x s_y} \frac{s_y}{s_x} = r \frac{s_y}{s_x}$$

پس خط برگشت به صورت زیر در می‌آید:

$$\hat{y} - \bar{y} = r \frac{s_y}{s_x} (x - \bar{x})$$

ضریب زاویه این خط برابر است با $r \frac{s_y}{s_x} = \hat{b}$ و خط از نقطه (\bar{x}, \bar{y}) می‌گذرد. اگر $r = \pm 1$ باشد این

خط بر خط

$$\frac{y - \bar{y}}{s_y} = \pm \frac{x - \bar{x}}{s_x} = r$$

که در بخش پیش دیدیم منطبق می‌شود. در این حالت می‌توان گفت که میان متغیر مستقل و متغیر وابسته یک رابطه خطی برقرار است و پیش‌بینی بدون خطا انجام می‌گیرد.

۸.۲.۸ توزیع برآورده یابها در حالت نرمال

فرض کنید Y_i ها مستقل و دارای توزیع نرمال باشند، یعنی

$$Y_i \sim N(a + bx_i, \sigma^2) \quad i = 1, 2, \dots, n$$

چون \hat{a} و \hat{b} هر کدام بصورت یک ترکیب خطی از متغیرهای نرمال مستقل Y_1, Y_2, \dots, Y_n هستند، پس هر کدام دارای توزیع نرمال می باشند. از طرفی داریم

$$E_i \sim N(0, \sigma^2)$$

$$\frac{E_i^2}{\sigma^2} \stackrel{D}{=} \chi^2(1)$$

$$\sum_{i=1}^n \frac{\hat{E}_i^2}{\sigma^2} \stackrel{D}{=} \chi^2(n)$$

حال اگر به جای E_i برآورد یاب آن یعنی $\hat{E}_i = Y_i - \hat{a} - \hat{b}x_i$ را قرار دهیم دو درجه آزادی از دست می رود (زیرا به جای دو پارامتر a و b برآورده یابهای آنها را گذردیم) و داریم

$$\sum_{i=1}^n \frac{\hat{E}_i^2}{\sigma^2} \stackrel{D}{=} \chi^2(n-2)$$

خلاصه، با استفاده از مطالب بالا و انجام محاسباتی که ما از ذکر آنها در اینجا خودداری می کنیم، نتایج زیر به دست می آیند:

$$\hat{a} \sim N\left(a, \frac{\sigma^2}{n} \left(1 + \frac{\bar{x}^2}{s_x^2}\right)\right)$$

$$\hat{b} \sim N\left(b, \frac{\sigma^2}{ns_x^2}\right)$$

$$\frac{(n-2)\hat{\sigma}^2}{\sigma^2} \stackrel{D}{=} \chi^2(n-2)$$

افزون بر این می توان ثابت کرد که \hat{a} و \hat{b} و همچنین $\hat{\sigma}^2$ و $\hat{\sigma}$ مستقلند. حال با استفاده از تعریف توزیع T که در فصل ۵ به آن اشاره شد، سه تابع محوری مهم زیر برای نتیجه گیریهای گسائی درباره پارامترها

دست می آیند:

$$Q_1 = \frac{(n-2)\hat{\sigma}^2}{\sigma^2} \stackrel{D}{=} \chi^2(n-2)$$

$$Q_2 = \frac{\sqrt{n}(\hat{a}-a)}{\sqrt{\hat{\sigma}^2 \left(1 + \frac{\bar{x}^2}{s_x^2}\right)}} \stackrel{D}{=} T(n-2)$$

$$Q_3 = \frac{\sqrt{n}(\hat{b}-b)}{\sqrt{\hat{\sigma}^2}} \stackrel{D}{=} T(n-2)$$

به کمک این متغیرهای محوری، همانطوری که در فصلهای ۶ و ۷ دیدیم، می توانیم برای پارامترهای a و b و σ^2 فاصله اطمینان پیدا کنیم و درباره آنها آزمون آماری انجام دهیم.

مثال ۴ جدول زیر نمره کنکور و معدل سالانه ده دانشجوی علوم را نشان می دهد. این نمره ها به معیار صد استاندارد شده اند. یک مدل خطی ساده را در نظر گرفته و پارامترها را برآورد کنید.

| | | | | | | | | | | |
|-------|----|----|----|----|----|----|----|----|----|----|
| x_i | ۵۲ | ۷۵ | ۳۴ | ۲۷ | ۵۷ | ۲۸ | ۳۹ | ۲۱ | ۴۳ | ۶۴ |
| y_i | ۷۵ | ۹۸ | ۵۶ | ۸۹ | ۹۲ | ۷۳ | ۶۵ | ۵۲ | ۷۸ | ۸۲ |

جدول نمره کنکور و معدل سالانه

حل - فرض کنید \hat{x} نمره کنکور و \hat{y} معدل سالانه برای دانشجوی i ام باشد. با استفاده از جدول بالا داریم

$$\bar{x} = 46 \quad \bar{x}^2 = 2116 \quad s_x^2 = 247/4$$

$$\bar{y} = 76 \quad \bar{y}^2 = 5776 \quad s_y^2 = 205/6$$

$$\bar{xy} = 3685/4 \quad \bar{x}\bar{y} = 3496 \quad s_{xy} = 189/4$$

$$r = \frac{s_{xy}}{s_x s_y} = 0.82$$

برآورد ضرب همبستگی این دو نمره با استفاده از جدول بالا برابر 0.82 است. بنابراین تا حدودی همبستگی خطی میان نمره کنکور و معدل سالانه برقرار می باشد. اگر هر (x_i, y_i) را با یک نقطه در صفحه محور مختصات نشان دهیم، ملاحظه می شود که این نقاط تمایل دارند که اطراف یک خط

راست باشد. (به نگاره ۱ توجه کنید). پس فرض مدل خطی ساده دور از انتظار نمی باشد. برآورد پارامترها می شوند

$$\hat{b} = \frac{s_{xy}}{s_x^2} \approx 0.177$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} \approx 40.178$$

$$\hat{\sigma}^2 = \frac{n}{n-2} (s_y^2 - \hat{b}^2 s_x^2) \approx 75/75$$

مثال ۵: در مثال ۴ خط برگشت را پیدا کنید و برای هر x_i متغیر وابسته را پیش بینی نمایید.

حل: خط برگشت عبارتست از

$$\hat{y} = \hat{a} + \hat{b}x$$

یا

$$\hat{y} = 40.178 + 0.177x$$

به عنوان مثال برای $x_1 = 52$ ، متغیر وابسته به طور متوسط با مقدار

$$\hat{y}_1 = 40.178 + 0.177(52) = 80.182$$

پیش بینی می شود. آنچه را که برای متغیر وابسته مشاهده کردیم $y_1 = 75$ می باشد. پس برآورد خطای تصادفی یا پس مانده رگرسیون می شود:

$$\hat{e}_1 = y_1 - \hat{y}_1 = 75 - 80.182 = -5.182$$

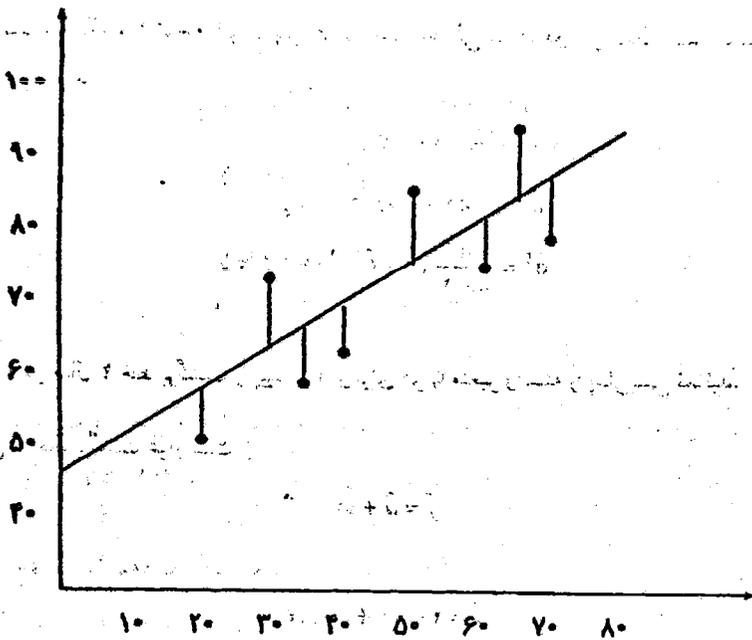
در شکل زیر خط برگشت و نقاط (x_i, y_i) و (x_i, \hat{y}_i) و \hat{e}_i را مشاهده می کنید.

مثال ۶: در مثال ۴، آزمون زیر را با میزان $\alpha = 0.05$ انجام دهید:

$$\begin{cases} H_0: b = 0 \\ H_1: b \neq 0 \end{cases}$$

حل: از تابع محوری

$$Q_T = \frac{\sqrt{n}(\hat{b} - b)}{\sqrt{\hat{\sigma}^2 / s_x^2}}$$



نگاره ۱: خط برگشت و برآورد خطاهای تصادفی

که دارای توزیع T با $n-2$ درجه آزادی است استفاده می کنیم. تحت فرض H_0 یافته این تابع می شود:

$$Q_T = \frac{\sqrt{10} \cdot (0.177 - 0)}{\sqrt{\frac{75/75}{227/2}}} = 4/40$$

از طرفی ناحیه بحرانی با میزان پنج درصد می شود $|t| > 2.306$. بنابراین فرض H_0 را رد می کنیم. این موضوع به طور شهودی هم معقول به نظر می آید، زیرا (x_i, y_i) ها اطراف خطی با ضریب زاویه غیر صفر جا دارند.

۹.۲.۸ تمرین بخش دو

۱. در مثال ۴، یک فاصله اطمینان با ضریب نود درصد برای b پیدا کنید.

۲. در مثال ۴، یک فاصله اطمینان با ضریب نود درصد برای σ^2 پیدا کنید.

۳. در مثال ۴، اگر نمره کنکور دانشجویی ۸۳ باشد، معدل سالانه او را پیش‌بینی کنید.
۴. در جدول زیر، متغیر مستقل عبارتست از نیروی کشش به یک میله فولادی و متغیر وابسته افزایش طول میله است.

| | | | | | |
|--------------|---|---|---|---|---|
| متغیر مستقل | ۱ | ۲ | ۳ | ۴ | ۵ |
| متغیر وابسته | ۲ | ۴ | ۵ | ۶ | ۸ |

برای این دو متغیر یک مدل خطی ساده فرض می‌کنیم، برآورد پارامترهای این مدل خطی را پیدا کنید.

۵. در تمرین ۴، خط رگرسیون را پیدا کرده رسم کنید.
۶. در تمرین ۴، اگر متغیر مستقل $3/5$ باشد، متغیر وابسته را پیش‌بینی کنید.
۷. در تمرین ۴، پس مانده‌ها و SSE و برآورد σ^2 را بیابید.
۸. در تمرین ۴، تحت شرایط نرمال یک فاصله اطمینان با ضریب نود درصد برای σ^2 را پیدا کنید.
۹. یک نوع ماده شیمیایی داریم که در x درجه حرارت لاگرم آن تجزیه می‌شود در پنج آزمایش داده‌های زیر به دست آمده‌اند.

| | | | | | |
|-----|----|----|---|---|---|
| x | -۲ | -۱ | ۰ | ۱ | ۲ |
| y | ۵ | ۴ | ۳ | ۲ | ۱ |

- یک مدل خطی ساده برای این دو متغیر فرض کنید. خط بازگشت را پیدا کرده رسم نمایید.
۱۰. فرض کنید در تمرین ۹، درجه حرارت ۳ باشد. مقدار ماده‌ای را که تجزیه می‌شود پیش‌بینی کنید.
۱۱. ثابت کنید در یک مدل خطی ساده داریم:

$$\sum_{i=1}^n (y_i - a - bx_i)^2 = SSE + n(a - \hat{a})^2 + \sum_{i=1}^n (x_i - \bar{x})^2 (b - \hat{b})^2$$

۱۲. با فرض نرمال و مستقل بودن Y_i ها توزیع \hat{a} و \hat{b} را کاملاً مشخص نمایید.
۱۳. ثابت کنید که امید ریاضی $E\hat{\beta}$ صفر است.
۱۴. در مدل خطی

$$Y = mx + E$$

برآورد m را به کمک $(x_1, Y_1), \dots, (x_n, Y_n)$ پیدا کنید.

مشاهده نتیجه آزمایش در C_i برابر است با p_i ، پس انتظار داریم که در این n آزمایش به طور متوسط np_i بار نتیجه آزمایش در C_i مشاهده کنیم. عدد np_i را (که ممکن است عدد درست نباشد) با e_i نشان می‌دهند و آن را فراوانی مورد انتظار Expected Frequency می‌گویند. بدیهی است که

$$\sum_{i=1}^k O_i = \sum_{i=1}^k e_i = n$$

فرض H_0 را چنین بیان می‌کنیم: آزمایش تصادفی بالا دارای مدل احتمال زیر است:

| | | | | |
|--------|-------|-------|-----|-------|
| مجموعه | C_1 | C_2 | ... | C_k |
| احتمال | p_1 | p_2 | ... | p_k |

چگونه این آزمون را انجام دهیم؟ بدیهی است که هرگاه فراوانیهای مشاهده شده و فراوانیهای مورد انتظار تحت مدل بالا بیش از حد تفاوت داشته باشند، نسبت به فرض H_0 بدگمان می‌شویم. به سخی دیگر هرگاه آماره

$$\sum_{i=1}^k (O_i - e_i)^2$$

بیش از حد بزرگ باشد، فرض H_0 رد می‌شود. ولی توزیع این آماره را نمی‌توان پیدا کرد. پیرسن به جای این آماره، آماره زیر را به کار برد

$$\sum_{i=1}^k \frac{(O_i - e_i)^2}{e_i}$$

و نشان داد که اگر n بزرگ باشد، این آماره، تحت فرض H_0 ، تقریباً دارای توزیع χ^2 با $k-1$ درجه آزادی است. رابطه میان فراوانیها سبب می‌شود که درجه آزادی به جای k بشود $k-1$. این آماره را V_k نشان می‌دهیم و آن را آماره χ^2 پیرسن می‌خوانیم. اگر با استفاده از جدول توزیع χ^2 داشته باشیم:

$$V_k > \chi_{1-\alpha}^2(k-1)$$

آنگاه فرض H_0 (یعنی برازندگی مدل) را با میزان α رد می‌کنیم. معمولاً تعداد رده‌ها، یعنی k ، را به نحوی انتخاب می‌کنند تا فراوانی داده‌ها در هر رده دست کم ۵ باشد. رده‌هایی را که دارای فراوانی کمتر از ۵ باشند با رده‌های دیگر ادغام می‌کنند.

مثال ۱ یک تاس تخته را ۱۲۰ بار مستقلاً می‌اندازیم. جدول فراوانی خالهای مشاهده شده در جدول

آزمون مدل‌های احتمال به کمک توزیع χ^2

تا به حال در تمام آزمونها فرض می‌کردیم که نوع توزیع متغیر تصادفی X یعنی مدل احتمال را می‌دانیم و آزمون را برای پارامتر آن توزیع انجام می‌دادیم، مثلاً برای پارامتر μ در توزیع نرمال. اینک می‌خواهیم درباره توزیع X آزمون کنیم. مثلاً فرض کنید X شماره تصادفهای رانندگی روزانه در یکی از چهارراهها در فاصله زمانی ساعت ۸ تا ۹ باشد. با داشتن آمار تصادفها برای ۲۰ روز می‌خواهیم بیازمائیم که X دارای توزیع پواسن است. به عنوان مثالی دیگر، فرض کنید X وزن یک نوزاد باشد، با داشتن وزن ۱۰۰ نوزاد می‌خواهیم بیازمائیم که X دارای توزیع نرمال است.

توزیعی را که گمان می‌بریم داده‌ها از آن آمده باشند، مثلاً پواسن یا نرمال، توزیع برازنده بر داده‌ها و آزمونی را که باید برای بررسی درستی این گمان انجام داد آزمون برازندگی Goodness Of Fit Test می‌نامند. مشهورترین آزمون برازندگی، آزمون χ^2 است. این آزمون به وسیله آماره‌ای انجام می‌گیرد که برای نمونه‌های بزرگ تقریباً دارای توزیع χ^2 است. چنین آزمونی را پیرسن، آماردان انگلیسی، در ۱۹۰۰ میلادی پیشنهاد کرد. آزمون χ^2 را برای مستقل بودن دو متغیر تصادفی، مثلاً مستقل بودن قد و وزن نوزاد، و برای هم‌توزیع بودن دو متغیر تصادفی، مثلاً هم‌توزیع بودن وزن نوزاد ایرانی و وزن نوزاد چینی، نیز به کار می‌برند. در این فصل با مثالهای ساده، اینگونه آزمونها را شرح می‌دهیم.

۱.۹ آزمون χ^2 برای برازندگی

فرض کنید یک آزمایش تصادفی انجام گیرد و نتیجه آن را تنها در یکی از رده‌های دو به دو جدای C_1, C_2, \dots, C_k مشاهده کنیم. احتمال مشاهده نتیجه این آزمایش تصادفی را در C_i با عدد $0 < p_i < 1$ نشان می‌دهیم، به طوری که $p_1 + p_2 + \dots + p_k = 1$.

اینک آزمایش تصادفی را n بار مستقلاً تکرار می‌کنیم. فرض کنید O_i بار نتیجه را در C_i مشاهده نمائیم. متغیر تصادفی O_i را فراوانی مشاهده شده Observed Frequency می‌نامند. چون احتمال

بر داده شده است:

| | | | | | | |
|---------|----|----|----|----|----|----|
| خال i | ۱ | ۲ | ۳ | ۴ | ۵ | ۶ |
| فراوانی | ۱۸ | ۲۳ | ۱۶ | ۲۱ | ۱۸ | ۲۴ |

با میزان ۵ درصد بیازمائید که این تاس نازیب است.

حل - باید بیازمائیم که این تاس دارای مدل احتمال زیر است

| | | | | | | |
|--------|---------------|---------------|---------------|---------------|---------------|---------------|
| خال | ۱ | ۲ | ۳ | ۴ | ۵ | ۶ |
| احتمال | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ |

در این مثال رده C_i عبارتست از $\{i\}$ یعنی خود خال i .

فراوانی مورد انتظار برای خال i در صورتی که مدل احتمال یعنی فرض H_0 درست باشد می‌شود:

$$e_i = n \times p_i = 120 \times \frac{1}{6} = 20$$

اینک جدول زیر را تنظیم می‌کنیم:

| | | | | | | | |
|-----------------------------|----------------|----------------|-----------------|----------------|----------------|-----------------|-----|
| خال i | ۱ | ۲ | ۳ | ۴ | ۵ | ۶ | جمع |
| O_i | ۱۸ | ۲۳ | ۱۶ | ۲۱ | ۱۸ | ۲۴ | ۱۲۰ |
| e_i | ۲۰ | ۲۰ | ۲۰ | ۲۰ | ۲۰ | ۲۰ | ۱۲۰ |
| $\frac{(O_i - e_i)^2}{e_i}$ | $\frac{4}{20}$ | $\frac{9}{20}$ | $\frac{16}{20}$ | $\frac{1}{20}$ | $\frac{4}{20}$ | $\frac{16}{20}$ | ۲/۵ |

چون $k = 6$ ، پس درجه آزادی V_k می‌شود $k - 1 = 5$. اینک داریم:

$$v_k = 2/5 \quad \chi^2_{/95}(5) = 11/1 \quad v_k < \chi^2_{/95}(5)$$

بنابراین فرض H_0 را رد نمی‌کنیم، یعنی نازیبی تاس را با میزان ۵ درصد می‌پذیریم.

مثال ۲ فرض کنید X طول عمر یک نوع لامپ باشد، بر حسب ماه. طول عمر ۲۰۰ لامپ از این نوع در جدول زیر داده شده است:

| | | | | |
|-------|------------|----------------|----------------|------------|
| C_i | $X \leq 3$ | $3 < X \leq 6$ | $6 < X \leq 9$ | $X \geq 9$ |
| O_i | ۵۳ | ۴۲ | ۳۵ | ۷۰ |

آزمون زیر را انجام دهید:

$$H_0: X \sim F(x) = 1 - e^{-\frac{x}{12}}, x > 0$$

حل - فراوانیهای مورد انتظار را تحت فرض H_0 می‌یابیم. مثلاً برای محاسبه e_1 می‌نویسیم:

$$p_1 = P(C_1) = P(X \leq 3) = F(3) = 0/221$$

$$e_1 = np_1 = 200 \times 0/221 = 24/2$$

نتیجه محاسبات در جدول زیر دیده می‌شود:

| | | | | |
|-------|-------|-------|-------|-------|
| C_i | C_1 | C_2 | C_3 | C_4 |
| O_i | ۵۳ | ۴۲ | ۳۵ | ۷۰ |
| e_i | ۴۴/۲ | ۳۴/۶ | ۲۶/۸ | ۹۴/۴ |

$$v_2 = 12/22 > \chi^2_{/9}(2) = 6/25$$

بنابراین با میزان ۵ درصد فرض H_0 رد می‌شود.

در صورتی که در مدل احتمال یک یا چند پارامتر مجهول باشد، باید نخست آنها را با استفاده از داده‌ها برآورد کرد. با برآورد هر پارامتر از درجه آزادی V_k یک درجه کاسته می‌شود. برای اینکه موضوع روشن شود به مثال ۳ توجه کنید.

مثال ۳ جدول زیر شماره تصادفها را برای ۵۰ هفته نشان می‌دهد.

| | | | | |
|-----------------------|----|----|---|----------|
| شماره تصادفها در هفته | ۰ | ۱ | ۲ | ≥ 3 |
| فراوانی هفته‌ها | ۳۲ | ۱۲ | ۶ | ۰ |

فرض کنید Y شماره تصادفها در هر هفته باشد. با میزان ۵ درصد بیازمائید که Y دارای توزیع پواسن است.

۲. در یک مغازه آجیل فروشی، ماشینی تخمه و پسته و بادام و فندق را با هم می‌آمیزد. در یک بسته ۵۰۰ دانه‌ای، ۲۶۹ تخمه و ۱۱۲ پسته و ۶۲ بادام و ۲۵ فندق دیده‌ایم. با میزان ۵ درصد بیازمائید که آمیختگی ماشین متناسب با ۱، ۲، ۲، ۵ است.

۳. در درس آمار و احتمال جدول فراوانی نمره‌ها چنین است:

| | | | | | |
|---------|-----|----|----|----|----|
| نمره | الف | ب | ج | د | هـ |
| فراوانی | ۱۴ | ۱۸ | ۲۲ | ۲۰ | ۱۶ |

با میزان ۱۰ درصد بیازمائید که توزیع نمره‌ها یکنواخت است.

۴. رابطه زیر را ثابت کنید:

$$\sum_{i=1}^k \frac{(O_i - e_i)^2}{e_i} = \sum_{i=1}^k \frac{O_i^2}{e_i} - n$$

۵. تجربه نشان داده است که نمره درس ریاضی عمومی هر دانشجو متغیر تصادفی X با توزیع $N(\mu, \sigma^2)$ است. استادی مدعی است که نمره‌های امتحان را روی منحنی برده بنابر دستور زیر نمره داده است:

| | |
|------------|---|
| نمره (الف) | $X > \mu + \frac{3}{4}\sigma$ |
| نمره (ب) | $\mu + \frac{1}{4}\sigma < X < \mu + \frac{3}{4}\sigma$ |
| نمره (ج) | $\mu - \frac{1}{4}\sigma < X < \mu + \frac{1}{4}\sigma$ |
| نمره (د) | $\mu - \frac{3}{4}\sigma < X < \mu - \frac{1}{4}\sigma$ |
| نمره (هـ) | $X < \mu - \frac{3}{4}\sigma$ |

جدول زیر نمره‌های کلاس ۵۰۰ نفری این استاد را نشان می‌دهد.

| | | | | | |
|---------|-----|-----|-----|----|----|
| نمره | الف | ب | ج | د | هـ |
| فراوانی | ۵۰ | ۱۱۵ | ۲۲۰ | ۹۰ | ۲۵ |

ادعای این استاد را با میزان پنج درصد بیازمائید.

۶. جدول زیر فراوانی تصادفها را در یک چهارراه برای ۷۰ روز غیر تعطیل، در فاصله ۷ تا ۹ صبح، نشان می‌دهد.

| | | | | | | | | |
|------------------------|-----|----|----|----|----|---|---|---|
| فراوانی روزها | ۲ | ۱۰ | ۱۵ | ۱۲ | ۱۲ | ۶ | ۶ | ۵ |
| بیش از ۷ شماره تصادفها | ۱۴۰ | ۲ | ۳ | ۴ | ۵ | ۶ | ۷ | ۷ |

حل - فرض H_0 می‌گوید که Y دارای چگالی زیر است:

$$H_0: f(y) = P(Y=y) = \frac{e^{-\lambda} \lambda^y}{y!} \quad y = 0, 1, \dots$$

با استفاده از داده‌های بالا برآورد پارامتر λ می‌شود:

$$\hat{\lambda} = \frac{(0 \times 22) + (1 \times 12) + (2 \times 6)}{50} = 0.48$$

در این مثال رده سوم و چهارم را با هم ادغام می‌کنیم. بنابراین تنها سه رده $C_1 = (Y=0)$ و $C_2 = (Y=1)$ و $C_3 = (Y \geq 2)$ را در نظر می‌گیریم و برآورد احتمال هر رده را تحت فرض H_0 می‌یابیم:

$$\hat{p}_1 = P(Y=0) = e^{-0.48} = 0.619$$

$$\hat{p}_2 = P(Y=1) = e^{-0.48} \times 0.48 = 0.297$$

$$\hat{p}_3 = P(Y \geq 2) = 1 - \hat{p}_1 - \hat{p}_2 = 0.084$$

حال فراوانیهای مورد انتظار عبارتند از:

$$\hat{e}_1 = n\hat{p}_1 = 50 \times 0.619 = 30.95$$

$$\hat{e}_2 = n\hat{p}_2 = 50 \times 0.297 = 14.85$$

$$\hat{e}_3 = n\hat{p}_3 = 50 \times 0.084 = 4.20$$

چون یک پارامتر را برآورد کرده‌ایم، آماره V_3 تقریباً دارای توزیع χ^2 با یک درجه آزادی است. به آسانی داریم:

$$v_3 = 1/354 \quad \chi^2_{0.95}(1) = 3.84$$

پس فرض H_0 را با میزان ۵ درصد رد نمی‌کنیم.

۱.۱.۹ تمرین بخش یک

۱. سکه‌ای را ۱۰۰ بار به تصادف پرتاب کرده ۶۳ بار شیر و ۳۷ بار خط دیده‌ایم. با میزان پنج درصد بیازمائید که این سکه ناریب است.

با میزان پنج درصد بیازمائید که شماره تصادفها در روز، دارای توزیع پواسن است. راهنمایی - نخست پارامتر میانگین را برآورد کنید و سپس $P(X=x)$ را برای شماره تصادفها بیابید.

۲.۹ آزمون χ^2 برای مستقل بودن

دو ویژگی A و B از جمعیتی را در نظر می‌گیریم، مثلاً قد و وزن، یا استعداد ریاضی و استعداد ادبیات، یا درآمد و درجه تحصیلی. فرض می‌کنیم که بتوانیم این دو ویژگی را به نحوی با دو متغیر تصادفی پیوسته یا گسسته X و Y نشان دهیم. اینک می‌خواهیم بیازمائیم که این دو ویژگی یا X و Y مستقل اند (فرض H_0). فرض کنید که ویژگی A در r رده A_1, \dots, A_r و ویژگی B در s رده B_1, \dots, B_s بررسی گردد. بنابراین هر دو ویژگی با هم در rs رده C_{11}, \dots, C_{rs} بررسی می‌شوند. منظور از C_{ij} عبارتست از داشتن ویژگی A_i در رده A و ویژگی B_j در رده B_j ، مثلاً درآمد متوسط و درجه لیسانس. اگر دو ویژگی A و B بخواهند مستقل باشند احتمال توأم زیر برای رده C_{ij} به دست می‌آید:

$$p_{ij} = P(C_{ij}) = P(A_i)P(B_j)$$

برای اینکه این احتمال را تحت فرض H_0 برآورد کنیم یک نمونه تصادفی n تائی در نظر می‌گیریم و می‌نویسیم

$$\hat{p}_{ij} = \frac{\text{فراوانی } B_j}{n} \cdot \frac{\text{فراوانی } A_i}{n}$$

حال فراوانی مورد انتظار برای C_{ij} ، تحت فرض H_0 ، می‌شود:

$$e_{ij} = n \cdot \hat{p}_{ij}$$

اگر O_{ij} فراوانی مشاهده شده برای C_{ij} باشد، آماره χ^2 می‌شود

$$V = \sum_{i=1}^r \sum_{j=1}^s \frac{(O_{ij} - e_{ij})^2}{e_{ij}}$$

درجه آزادی V برابر است با $d = (r-1)(s-1)$ ، زیرا باید از $rs-1$ تعداد پارامترهائی را که برآورد می‌کنیم کم نماییم. بنابراین داریم:

$$d = rs - 1 - (r-1) - (s-1) = (r-1)(s-1)$$

اگر مقداری را که برای V مشاهده می‌کنیم از $\chi^2_{\alpha}(d)$ بزرگتر شود فرض H_0 را با میزان α رد می‌کنیم.

مثال ۴ در شهری بزرگ ۱۰۰۰ خانواده را به تصادف انتخاب می‌کنیم و از هر خانواده دربارهٔ تعداد فرزندان و سواد خانواده پرسش می‌کنیم. می‌خواهیم بیازمائیم که این دو ویژگی مستقل اند. فرض می‌کنیم تعداد فرزندان در سه رده A_1 (هیچ یا یک فرزند) و A_2 (دو یا سه فرزند) و A_3 (بیش از سه فرزند) و سواد خانواده در دو رده B_1 (با سواد) و B_2 (بی‌سواد) تنظیم شده باشند. جدول فراوانی زیر دیده شده است:

در این جدول فراوانیهای مشاهده شده و مورد انتظار را می‌بینید. به عنوان مثال مشاهده کردیم که در این نمونه ۱۰۰۰ تائی، ۱۸۲ خانواده بی‌سواد دارای هیچ یا یک فرزند هستند. فراوانی مورد انتظار یعنی $201/6$ را چنین می‌یابیم:

| تعداد فرزندان \ سواد خانواده | A_1 | A_2 | A_3 | جمع |
|------------------------------|-------------|-------------|-------------|------|
| B_1 | ۱۸۲ (۲۰۱/۶) | ۲۱۵ (۲۱۰/۶) | ۲۰۳ (۱۸۷/۸) | ۶۰۰ |
| B_2 | ۱۵۲ (۱۳۲/۴) | ۱۳۶ (۱۴۰/۴) | ۱۱۰ (۱۲۵/۸) | ۴۰۰ |
| جمع | ۳۳۶ | ۳۵۱ | ۳۱۳ | ۱۰۰۰ |

$$\hat{p}_{11} = \frac{\text{فراوانی } B_1}{n} \cdot \frac{\text{فراوانی } A_1}{n}$$

$$= 0/336 \times 0/600 = 0/2016$$

بنابراین فراوانی مورد انتظار می‌شود

$$e_{11} = n \times \hat{p}_{11} = 1000 \times 0/2016 = 201/6$$

درجه آزادی V برابر است با $d = (3-1)(2-1) = 2$ حال داریم:

$$V = \frac{(182 - 201/6)^2}{201/6} + \dots + \frac{(110 - 125/8)^2}{125/8} = 8/069$$

چون $\chi^2_{0.95}(2) = 5.991$ است پس با میزان پنج درصد فرض مستقل بودن تعداد فرزندان و سواد خانواده را رد می‌کنیم.

۱.۲.۹ تمرین بخش دو

۱. یک سکه نالرئب و یک تاس چهاروجهی نالرئب را ۸۰ بار با هم انداخته‌اند. با استفاده از جدول فراوانی زیر، با میزان ده درصد بیازمائید که فرایند سکه (H یا T) و فرایند تاس (۱ یا ۲ یا ۳ یا ۴) از هم مستقل اند. معمولاً مستقل بودن این دو فرایند را به طور شهودی می‌پذیرند.

| | | | | |
|-----------|----|----|----|----|
| تاس \ سکه | ۱ | ۲ | ۳ | ۴ |
| H | ۸ | ۶ | ۱۲ | ۱۰ |
| T | ۱۲ | ۱۰ | ۸ | ۱۴ |

۲. در تمرین بالا آزمون را در حالت‌های زیر چگونه انجام می‌دهید؟

الف - اگر سکه ارب و تاس نالرئب باشد.

ب - اگر سکه نالرئب و تاس ارب باشد.

ج - اگر سکه ارب و تاس ارب باشد.

دانهائی - در هر مورد نخست پارامترهای مجهول را بیابید و سپس فراوانیهای مورد انتظار را پیدا کنید.

۳. با استفاده از جدول فراوانی زیر، مستقل بودن مصرف سیگار و عصبی بودن را با میزان پنج درصد بیازمائید.

| | | |
|----------|--------|------------|
| | سیگاری | غیر سیگاری |
| عصبی | ۲۱ | ۳۶ |
| غیر عصبی | ۴۸ | ۲۵ |

۴. یک سکه ۵۰ ریالی و یک سکه ۱۰۰ ریالی را ۵۰ بار با هم بریزید.

جدول فراوانی را برای HH و HT و TH و TT مشاهده کنید. حال مستقل بودن فرایندهای این دو سکه را با میزان ده درصد بیازمائید.

۳.۹ آزمون هم‌توزیمی

یک ویژگی را در دو جمعیت در نظر می‌گیریم، مثلاً با سوادی در دو شهر بزرگ. فرض می‌کنیم که بتوزیم این ویژگی را به نحوی در جمعیت اول با متغیر تصادفی X و در جمعیت دوم با Y نشان دهیم.

اینک می‌خواهیم بیازمائیم که X و Y هم‌توزیع هستند (فرض H_0). ویژگی مورد مطالعه را در رده‌های C_1, \dots, C_k تنظیم می‌نمائیم. فرض کنید احتمال اینکه یکی از اعضای جمعیت اول در C_i باشد برابر p_i و احتمال اینکه یکی از اعضای جمعیت دوم در C_i باشد برابر q_i شود، به طوری که

$$\sum_{i=1}^k p_i = \sum_{i=1}^k q_i = 1$$

اگر هر دو جمعیت در رابطه با این ویژگی هم‌توزیع باشند، باید داشته باشیم:

$$p_i = q_i = r_i \quad i = 1, \dots, k$$

در صورتی که H_0 درست باشد، r_i ، یعنی احتمال مشترک، احتمال این است که یکی از اعضای دو جمعیت در C_i باشد.

برای اینکه آزمون را انجام دهیم، یک نمونه تصادفی m تائی از جمعیت اول و یک نمونه تصادفی n تائی از جمعیت دوم در نظر می‌گیریم. فرض کنید این دو نمونه از هم مستقل باشند. فراوانی مشاهده شده C_i را برای نمونه اول با M_i و برای نمونه دوم با N_i نشان می‌دهیم. فراوانیهای مورد انتظار تحت فرض H_0 می‌شوند.

$$e_i = mp_i = mr_i$$

$$f_i = nq_i = nr_i$$

دو آماره

$$U = \sum_{i=1}^k \frac{(M_i - e_i)^2}{e_i}$$

$$V = \sum_{i=1}^k \frac{(N_i - f_i)^2}{f_i}$$

مستقل بوده و هر یک به تقریب (با فرض بزرگ بودن m و n) با $\chi^2(k-1)$ هم‌توزیع می‌باشند. بنابراین تحت فرض H_0 به تقریب داریم:

$$W = U + V \stackrel{D}{=} \chi^2(2k-2)$$

| | فرآوانی | ابتدائی | راهنمائی | دبیرستانی | جمع |
|----------|---------|---------|----------|-----------|------|
| شهره دوم | O_j | ۵۰۰ | ۲۰۰ | ۱۰۰ | ۱۰۰۰ |
| | E_j | ۲۰۰ | ۲۰۰ | ۲۰۰ | |
| شهره اول | O_j | ۵۰۰ | ۶۰۰ | ۲۰۰ | ۱۰۰۰ |
| | E_j | ۶۰۰ | ۶۰۰ | ۲۰۰ | |
| | جمع | ۱۰۰۰ | ۱۰۰۰ | ۵۰۰ | ۲۵۰۰ |

اینک یافته‌های آماره‌های U و V و W عبارتند از

$$u = \frac{(500-200)^2}{200} + \frac{(400-200)^2}{200} + \frac{(100-200)^2}{200} = 75$$

$$v = \frac{(500-600)^2}{600} + \frac{(1600-600)^2}{600} + \frac{(400-300)^2}{300} = 50$$

$$w = u + v = 75 + 50 = 125$$

$$d = 6 - 2 - 2 = 2 \text{ درجه آزادی}$$

$$w > \chi^2_{0.99}(2) = 9.21$$

بنابراین فرض H_0 را با میزان یک درصد رد می‌کنیم.

۱.۳.۹ تمرین بخش سه

۱. یک سکه طلا را ۱۰۰ بار انداخته‌ایم، ۴۰ بار شیر و ۶۰ خط آمده است. در حالی که یک سکه نقره در ۸۰ پرتاب، ۳۵ بار شیر و ۴۵ بار خط آمده است با میزان ده درصد بیازماتید که این دو سکه از نظر شیر و خط آمدن هم‌توزیع هستند.

۲. تمرین ۱ را با استفاده از تقریب نرمال و آزمون نرمال را انجام دهید.

از آنجائی که پارامتر τ_i مجهول می‌باشد، به جای آن برآوردش یعنی

$$\hat{\tau}_i = \frac{m_i + n_i}{m + n}$$

را به کار می‌بریم. چون باید $\tau_1, \tau_2, \dots, \tau_{k-1}$ یعنی $k-1$ پارامتر، را برآورد کنیم، پس درجه آزادی می‌شود

$$d = 2k - 2 - (k-1) = k-1$$

اگر برای یافته آماره W داشته باشیم

$$W > \chi^2_{1-\alpha}(k-1)$$

فرض H_0 را با میزان α رد می‌کنیم.

مثال ۵. در شهری از ۱۰۰۰ دانش آموز که به تصادف انتخاب شده‌اند ۵۰۰ نفر در دوره ابتدائی و ۴۰۰ نفر در دوره راهنمائی و ۱۰۰ نفر در دوره دبیرستان می‌باشند. در شهری دیگر از ۱۵۰۰ دانش آموز، اینگونه افراد عبارتند از ۵۰۰ و ۶۰۰ و ۴۰۰. با میزان یک درصد بیازماتید که این دو شهر از نظر ترکیب دانش‌آموزان این سه دوره هم‌توزیع می‌باشند.

در این دو شهر رویم از ۲۵۰۰ دانش آموز، ۱۰۰۰ نفر در دوره ابتدائی و ۱۰۰۰ نفر در دوره راهنمائی و ۵۰۰ نفر در دوره دبیرستان می‌باشند. اگر فرض H_0 یعنی هم‌توزیعی درست باشد، داریم

$$\hat{\tau}_1 = \frac{1000}{2500} = 0.4, \quad \hat{\tau}_2 = \frac{1000}{2500} = 0.4, \quad \hat{\tau}_3 = \frac{500}{2500} = 0.2$$

جدول فرآوانیهای مشاهده شده و مورد انتظار را در زیر می‌بینید:

۳. تخم گل هلندی از ۱۰۰ گل، ۲۴ گل سفید و ۳۰ گل صورتی و ۴۶ گل قرمز داده است. در حالی که نوع بلژیکی همین تخم از ۱۰۰ گل، ۲۰ گل سفید و ۴۰ گل صورتی و ۴۰ گل قرمز داده است. با میزان ده درصد بیازمائیید که هر دو نوع تخم گل، از نظر رنگ گل هم توزیع هستند.

۴. دو تاس را ۵۰ بار به تصادف بریزید و جدول فراوانی خالها را به بینید. حال با میزان پنج درصد نشان دهید که هر دو تاس از نظر توزیع خالها یکسان هستند.

پاسخ تمرینهای انتخابی

فصل اول

بخش دو صفحه ۲۳

۱. نژاد: اسمی، جدا، گسسته

۲. الف) جمعیت: دانش آموزان یک مدرسه، نمونه: بچه‌هایی که اسم آنها با م شروع می‌شود، متغیر: قدرت دید.

۴. مستقل از a

۵. $2/471$ $2/47$ $2/4715$

۷. $y = 10x$: مقیاس‌های اسمی، ترتیبی، فاصله‌ای و نسبتی تغییر نمی‌کنند.

بخش سه صفحه ۲۸

۱. داده‌ها از نوع جدا هستند، ۴۰٪ از خانواده‌ها بیش از ۵ قرض مصرف می‌کنند.

۲. داده‌ها با مقیاس ترتیبی اندازه‌گیری شده‌اند، ۶۲/۵٪ از دانش آموزان نمره کمتر از ۱۲ دارند.

۳. داده‌ها از نوع پیوسته هستند.

بخش چهار صفحه ۳۴

۱. ۲٪ بیش از $2\sigma + \mu$ و ۱۶٪ کمتر از $\mu - \sigma$

بخش پنج صفحه ۲۸

۱. $m = 8, Q_1 = 5, D_r = 6/6, M = 7$

۲. $\bar{x} = 166/65, m = 167, M = 168/2$

۳. $Q_2 = 162/67$

۴. میانگین اصلاح شده $\bar{x} = 2/712, z = 0$

بخش شش صفحه ۵۶

۱. $Q' = 1/275, Q = 2/62, R = 12, R' = 10, S = 12/21$

۲. $\bar{x} = 229/3, S_x^2 = 451/96, m = 250/526$

محل دوم

بخش یک صفحه ۷۵

۱. موضوع: رنگ صفحه.

$S = \{\text{قرمز, سفید, سبز}\}$

۲. فضا نمونه ناپایدار است.

الف) $E \cap F \cap G'$

ب) $(E \cap F) \cap (G \cap F) \cup (E \cap G) \cup (E \cap F \cap G)$

ج) $(E \cap F \cap G') \cup (E \cap G \cap F') \cup (E' \cap G \cap F)$

۱۰. درست، درست، درست، درست

بخش دو صفحه ۸۱

۱. $20, \frac{1}{52}$

۲. $\frac{2}{7}$

۳. $\frac{1}{7}$

بخش سه صفحه ۸۶

۱. $\frac{5}{8}, \frac{2}{8}$

۲. $\frac{2}{5}$

۳. ۶۰۰

۴. $\frac{1}{4}$

۵. $\frac{1}{10}$

۸. $0/30, 0/40, 0/70, 0/3$

۹. $0/70, 0/30, 0/40, 0/7$

بخش چهار صفحه ۱۰۰

۱. $(\frac{4}{6})^5, 6^5$

۲. $\frac{1}{10}, \frac{4}{10}, \frac{1}{10}$

۳. 9^2

۴. $\frac{2}{10}, 120$

۵. ۷۲۰

۶. $0/28, 1260$

۷. $0/006, 1/3 \times 10^{12}$

۸. $\frac{1}{7}, 24$

۹. $0/205$

۱۰. ۷۲۵۷۶۰

۱۱. 3^N

۱۲. $\binom{R-N-1}{R-N}$

۱۳. $6 \cdot \binom{R-1}{R-N}$

$$0.0825$$

بخش هشت صفحه ۱۳۰

$$\frac{1}{4} \cdot 2$$

فصل سوم

بخش یک صفحه ۱۳۵

$$1. (ب) \frac{7}{8}, \text{ صفر}, \frac{2}{8}$$

(د) خیر، بلی

$$2. (ج) 2750$$

$$3. (ب) 9$$

$$4. \frac{1}{6}, \frac{1}{12}$$

بخش دو صفحه ۱۴۷

$$1. (ب) \frac{2}{8}, \frac{7}{8}, \frac{2}{8}$$

$$2. (الف) \sqrt{2} - x^2, 0 \leq x < \frac{\sqrt{2}}{2}$$

$$(ب) 96$$

(ج) صفر

۳. $T(x)$ ، تابع توزیع نیست ولی بقیه تابع توزیع می باشد.

$$m = 0.2$$

$$F_Y(y) = \frac{y^2}{y^2 + 1}, y \geq 0$$

$$m = \frac{\pi}{2} \cdot 1 - \frac{\sqrt{2}}{2}$$

$$\binom{N}{R} \cdot \frac{N!}{(N-R)!R!} \cdot \frac{N!}{(N-R)!} \cdot \frac{N^R}{R!} \cdot N^R \cdot 16$$

$$17. 250,60$$

$$18. 84$$

$$19. 390, 252$$

$$20. 7560$$

$$23. 0.0825, 4200$$

$$26. 0.94$$

بخش پنج صفحه ۱۱۱

$$2. \frac{25}{36}, \frac{91}{216}, \frac{25}{91}$$

$$3. \frac{1}{5}$$

$$4. \frac{2}{3}$$

۱۲. خیر

بخش شش صفحه ۱۱۷

$$2. \frac{1}{2}$$

$$3. 0.13$$

$$7. \frac{2}{8}$$

بخش هفت صفحه ۱۲۳

$$1. 0.255$$

$$2. \frac{9}{11}, \frac{11}{25}$$

$$3. \frac{4}{7}, \frac{7}{25}$$

$$4. \frac{2^n}{2^n + 9}, 0.38$$

$E(X-Y) = \dots$ $E(X+Y) = 2$ $E(X^2) = E(Y^2) = \frac{7}{3}$ $E(X) = E(Y) = 1$ $E(X+Y)(X-Y) = \dots$

$p\mu_1 + q\mu_2 = 1$

$E(X+Y) = \frac{17}{3}$ $E(XY) = \frac{9}{3}$ $E(Y) = 2$ $E(X) = \frac{5}{3}$ \dots

$E(XY) = \frac{1}{3}$ $E(X) = E(Y) = \frac{7}{12}$ \dots

$E(\cos^2 X) = \frac{1}{3}$ $E(\cos X) = \frac{1}{3}$ \dots

بخش دو صفحه ۱۷۸

$22/25, 2/5, 1$

$2/95, 8/69, \frac{11}{3}, 1/15, \frac{2}{3}, -2, 1/29, \frac{1}{12}, \frac{1}{2}, 1/3, \frac{2}{52}, \frac{1}{3}, 1/58, \frac{1}{3}$

$10, 100, 60, 3$

$-6, 6, 4$

$\frac{N^2-1}{12}, \frac{N+1}{2}$

$\frac{2}{3}, 13$

$\frac{2}{45}, 15$

بخش سه صفحه ۱۸۴

$0, 4, 2, 2, 1$

$\frac{2}{225}, \frac{2}{9}, \frac{2}{5}, \frac{8}{15}$

صفر ۵

$52, 8$

بخش چهار صفحه ۱۸۹

$0/287, 0/1, \frac{1}{22}, 1$

بخش سه صفحه ۱۵۷

$\frac{2}{8}, \frac{7}{8}, 1$

بلی ۲

$f_X(x) = 2\sqrt{x} - 2x, 0 < x < \frac{\sqrt{2}}{2}$ الف ۳

ب ۰/۰۹۶

$\frac{1}{2}, 5$

$1 - e^{-2x}, x > 0$ ۸

بخش چهار صفحه ۱۶۲

۱. خیر

۲. بلی

۳. خیر

۴. خیر

فصل چهارم

بخش یک صفحه ۱۷۲

$227/5, 1$

$EX^2 = EY^2$ ولی $E(X) \neq E(Y)$ و X و Y هم توزیع نیستند

۱. ۲

۴. ۵

۶. $E(X)$ وجود ندارد

۷. $E(X)$ وجود ندارد

بخش دو صفحه ۲۰۰

۰/۷۶۲.۱

۰/۱۳۶.۲

۰/۰۰۲۵.۴

۰/۰۸۰۳.۵

۰/۰۰۳۴.۶

۰/۳۹۳۲.۷

۰/۰۴۷۴.۸

۱۱. $P(۳), P(۲), P(۱)$

۱۲. $P(۳)$

بخش سه صفحه ۲۰۵

۰/۶۲۹۳.۰, ۰/۸۴۱۳.۰, ۰/۳۷۴۷.۲

$N(۲۰, ۴/۹۴), ۰/۳=۸۵.۴$

۰/۰۲۲۸.۵

۳۲.۶

۰/۶۷۳۶.۰, ۰/۵۷۳۳.۰, ۰/۴۸۴۰.۷

۰/۲۴۱۷.۰, ۰/۶۱۷۰.۱۰

بخش چهار صفحه ۲۱۲

۰/۵۴۷۹.۰, ۰/۵۴۶۸.۲

۰/۰۳۵۷.۰, ۰/۰۳۴۷.۴

۵ صفر

۲. $\frac{\sqrt{13}}{۲}, \frac{\sqrt{13}}{۲}$

۵. $\frac{-\sqrt{۲}}{۲}$

۷. $\frac{\sigma^2}{n}$

۸ صفر

-۱.۹

۱۰. $\frac{۱ \pm \sqrt{5}}{۲}$

فصل پنجم

بخش یک صفحه ۱۹۶

۱. $۰/۲۵, \frac{۲۴}{۲۵}, \frac{۱۲}{۵}$

۲. $۰/۰۲۰۵$

۴. $B(n, q)$

۵. $(۰/۶)^{۱۲}$

۶. $۰/۰۰۷۶$

$f(x) = pq^{x-1} \quad x = ۱, ۲, ۳, \dots, n$

۹. $۰/۱۲۵$

۱۰. $\frac{۱}{۵}, \frac{۶}{۲۵}, \frac{۱۲}{۵}$

۱۲. $\frac{۹۵}{۱۲۲}, \frac{۱۳}{۱۲}, \frac{۱}{۱۲}$

۱۳. $۰/۳۷۵$

۱۴. $۲/۶۲۵, ۰/۴۲$

۱۵. $-۱, -npq$

۳. (۶۸/۸۹، ۷۵/۱۱)

۹/۵.۵

بخش پنج صفحه ۲۳۶

۱. الف) ۰/۵۲۵۶، ۰/۴۷۴۴، ۰/۰۵۲۷

ب) ۸

۲. ۰/۲۷۴۳، ۰/۷۲۵۷، ۰/۰۰۱۳

۳. $\frac{2}{7}, \frac{5}{7}, \frac{1}{21}$

۴. ۰/۴۵، ۰/۲۲۳۱

۵. الف) $X \sim B(22, \frac{1}{5})$

بخش شش صفحه ۲۴۳

۱. p - مقدار = ۰/۰۶۶۸

۲. p - مقدار = ۰/۱۳۳۶

۳. ۰/۲۵۷۸، ۰/۵۱۵۶

۴. p - مقدار صفر می شود، (برای هر دو آزمون)

بخش هفت صفحه ۲۵۰

۱. ۰/۵۰

۲. ۶/۵۷

۳. صفر، ۱، صفر، ۳، $2n, n$

۴. ۰/۴۷

۵. ۰/۰۲۵

۶. ۰/۱۵

۶. ب) ۰/۹۹۸۲

۷. ب) ۰/۰۳۶۶

فصل ششم

بخش دو صفحه ۲۱۹

۱. $\frac{1}{5}, 65$

۲. $U, N(0, 6), N(24, 3)$ ، خیر

۳. بزرگترین مقدار را بعنوان برآورد θ اختیار می کنیم

بخش سه صفحه ۲۲۷

۱. بله، کوچکترین و بزرگترین مشاهده

۲. \bar{X}, \bar{X}

۳. \bar{X}, \bar{X}

۴. \bar{X}^2

۶. \bar{X}, \bar{X}

۷. $T, \frac{T}{5}, \frac{T}{5}$ بهتر است

۸. $\frac{1}{4}, \frac{1}{4}$

۹. $U_1(1-U_1), U_1, U_1$

۱۰. $\bar{X}_2 - \bar{X}_1$

بخش چهار صفحه ۲۳۱

۱. ۰/۸۶۶۲

۲. (۱۹/۰۳، ۲۲/۵۲)

بخش دو صفحه ۲۶۷

۱. $(-۷/۹۵, -۰/۰۵)$

۲. $(۰/۱۵, ۲/۲۹)$

۵. اگر σ_1 و σ_2 مجهول باشند:

$$\sqrt{\frac{2}{m} + \frac{9}{n}} < 2\mu_1 + 2\mu_2 < \sqrt{\frac{2}{m} + \frac{9}{n}} + t_{1-\frac{\alpha}{2}} - \frac{\alpha}{2} (m+n-2) S_p \sqrt{\frac{2}{m} + \frac{9}{n}}$$

۶. نمی توانیم H_0 را رد کنیم.

۸. نمی توانیم H_0 را رد کنیم.

۱۰. الف) نمی توانیم H_0 را رد کنیم.

ب) نمی توانیم H_0 را رد کنیم.

ج) $(-۳۷۵/۹۹$ و $۱۸۵/۹۹)$

بخش سه صفحه ۲۷۳

۱. H_0 رد می شود.

۲. $(۰/۰۵۲۲$ و $۰/۱۸۷۸)$

۴. H_0 رد می شود

۵. نمی توانیم H_0 را رد کنیم.

فصل هشتم

بخش یک صفحه ۲۸۲

۱. $-۰/۲۴۵$

۲. $۰/۲۴۷$

۵. خیر

۰/۸۰۲۷

۰/۰۰۱۳۸

۰/۰۵، ۰/۹۹

۹/۸۸، ۱۰

$\frac{2\sigma^2}{n}$ ، σ^2 ، ۱۱

۰/۸، $\bar{X} - 2\bar{X} + 1$ ، ۱۲

$\frac{n-1}{n} \sigma^2$ ، ۱۳

$\frac{2\sigma^2}{n-1}$ ، σ^2 ، ۱۴

۱۵. صفر

۰/۰۶۲۱، ۱۶

$\frac{2\sigma^2}{m+n-2}$ ، σ^2 ، ۱۷

فصل نهم

بخش یک صفحه ۲۶۰

۱. $(۳۲۰/۱۳، ۶۲۰/۹۵)$

۲. H_0 را رد نمی کنیم.

۴. $(۱۰۲/۱۸، ۴۱۰/۲۷)$

۶. H_0 را رد می کنیم.

۰/۸۵، ۹

۰/۳۱، ۱۰

$$۲. V_k = ۱۶/۳۲۴ > \chi_{.۹۵}^2(۳) = ۷/۸۱$$

$$۳. V_k = ۱۰ > \chi_{.۹}^2(۴) = ۷/۷۸$$

$$۵. V_k = ۲۲/۷۵ > ۹/۴۹ = \chi_{.۹۵}^2(۴)$$

$$۶. V_k = ۱/۸۹ < ۱۲/۶ = \chi_{.۹۵}^2(۶)$$

بخش دو صفحه ۳۰۳

$$۱. V_k = ۴/۸ < \chi_{.۹}^2(۷)$$

$$۲. الف) $V_k = ۴/۰۴ < \chi_{.۹}^2(۶) = ۱۰/۶$$$

$$ب) $V_k = ۳/۲۷ < \chi_{.۹}^2(۴) = ۷/۷۸$$$

$$ج) $V_k = ۲/۴۹ < \chi_{.۹}^2(۳) = ۶/۲۵$$$

$$۳. $V_k = ۱۰/۷۳ > \chi_{.۹۵}^2(۱) = ۳/۸۴$$$

بخش سه صفحه ۳۰۶

$$۱. $\omega = u + v = ۰/۲۶۴ = \chi_{.۹}^2(۱) = ۲/۷۱$$$

$$۲. $Z = -۰/۵۴ > z_{.۰۵} = -۱/۶۴۵$$$

$$۳. $\omega = u + v = ۲/۲۱۰ < \chi_{.۹}^2(۲) = ۴/۶۱$$$

$$X + Y = ۲.۶$$

$$۰.۹۵۹۴.۷$$

$$۰.۲۹۸۱.۸ - p$$

$$۰.۳۶۷.۰.۶۱۴.۹$$

$$\sigma_1^2 + \sigma_2^2 \pm 2\rho\sigma_1\sigma_2, \mu_1 \pm \mu_2.۱۰$$

بخش دو صفحه ۲۹۲

$$۰.۴۴۵.۱/۰۹۵.۱$$

$$۳۹/۰۹۷, ۲۲۱/۹۷۸.۲$$

$$۱۰۴/۶۹.۳$$

$$\hat{a}^2 = ۰/۱۳۳, \hat{a} = ۰/۸, \hat{b} = ۱/۴.۴$$

$$\hat{y} = ۰/۸ + ۱/۴x.۵$$

$$۵/۷.۶$$

$$SSE = ۰/۴, \hat{a}_T = ۰/۱۳۳.۷$$

$$۰/۰۵۱.۱/۱۳۶.۸$$

$$\hat{y} = ۳ - x.۹$$

۱۰. صفر

$$\hat{m} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}.۱۴$$

فصل نهم

بخش یک صفحه ۲۹۲

$$۱. $V_k = ۶/۶۷ > \chi_{.۹۵}^2(۱) = ۳/۸۴$$$

پ

| | |
|---------------|--|
| Finite | با پایان، ۷۴ |
| Significance | با معنایی، ۲۳۷ |
| Odds | بخت پولی، ۸۱ |
| Estimate | برآورد، تخمین، ۲۱۷ |
| MLE | برآورد راستمنامی ماکزیمم، ۲۲۴ |
| Estimator | برآوردیاب، برآوردگر، برآورد کننده، ۲۱۷ |
| Estimation | برآوردیابی، ۲۱۷ |
| Kurtosis | برجستگی، ۵۸ |
| Range | برد، ۵۰ |
| Borel | برل (۱۸۸۸-۱۹۵۶) ریاضیدان فرانسوی، ۱۰۶ |
| Regression | برگشت، رگرسیون، ۲۷۵، ۲۹۴ |
| Bernoulli, J. | برنولی، (۱۶۵۴-۱۷۰۵) ریاضیدان سوئیسی، ۱۹۲ |
| Boole, G. | بول (۱۸۱۵-۱۸۶۴) ریاضیدان انگلیسی، ۱۱۲ |
| Bonferoni, | بونفرونی، احتمال دان، ۱۱۲ |
| At random | به تصادف، ۱۳۵ |
| Infinite | بی پایان، ۷۴ |
| Bayes, T | بیز (۱۷۰۱-۱۷۶۱) کشیش و فیلسوف انگلیسی، ۱۱۸ |

پ

| | |
|---------------|--|
| Pascal | پاسکال، (۱۶۲۳-۱۶۶۲) ریاضیدان فرانسوی، ۶۸ |
| Poisson, S.D. | پواسن (۱۷۸۱-۱۸۴۰) ریاضیدان فرانسوی، ۱۹۷ |
| Pearson, K. | پیرسن (۱۸۷۵-۱۹۳۶) آماردان انگلیسی، ۵۹، ۲۹۴ |
| Event | پیشامد، ۷۰ |

ت

| | |
|-----------------------|--------------------|
| Distribution Function | تابع توزیع، ۱۳۶ |
| Degenerate | تباهیده، ثابت، ۱۳۵ |
| Combination | ترکیب، ۹۲ |

واژه نامه فارسی به انگلیسی

آ

| | |
|-------------------------|--|
| Experiment | آزمایش، ۶۹ |
| Random Experiment | آزمایش تصادفی، ۶۹ |
| Test | آزمون، ۲۳۲ |
| Goodness of Fit Test | آزمون برازندگی، ۲۹۵ |
| Statistics | آمار، ۱۷ |
| Descriptive Statistics | آمار توصیفی، ۱۷ |
| Statistic | آماره، ۲۱۷ |
| Union | اجتماع، ۷۱ |
| Probability | احتمال، ۶۸ |
| Posterior Probability | احتمال پسین، ۱۲۲ |
| Prior Probability | احتمال پیشین، ۱۲۲ |
| Conditional Probability | احتمال شرطی، ۱۱۳ |
| Stirling, J. | استرلینگ (۱۶۹۲-۱۷۷۰) ریاضیدان انگلیسی، ۹۳ |
| Induction | استقرا، ۲۱۵ |
| Stevens, S.S | استونز استاد روانشناسی دانشگاه هاروارد، ۱۹ |
| Intersection | اشتراک، ۷۲ |
| Expectation | امید، انتظار، ۱۶۳ |
| Deviation | انحراف، ۵۰ |
| Standard Deviation | انحراف معیار ۳۸، ۵۱، ۱۷۷ |
| Measuring | اندازه گیری، ۱۹ |

| | |
|--------------------------|--|
| Continuous Data | داده‌های پیوسته، ۲۲ |
| Coded Data | داده‌های تبدیل شده، ۵۲ |
| Discrete Data | داده‌ها گسسته جدا، ۲۲ |
| Crab | دانشمند ایتالیایی (۱۵۷۶-۱۵۰۱)، ۶۸ |
| Degrees of Freedom | درجه آزادی، ۲۴۴ |
| Binomial | دو جمله‌ای، ۱۹۳ |
| De Moire, A. | دو موآور (۱۷۵۴-۱۶۶۷) ریاضیدان فرانسوی، ۲۰۶ |
| Bimodal | دو نمائی، ۴۵ |
| Decile | دهک، ۴۰ |
| Likelihood | راستمنائی، درستمنائی، ۲۲۴ |
| Maximum Likelihood | راستمنای ماکزیمم، ۲۲۴ |
| Occur | رخ دادن، ۷۵ |
| Moment Method | روش گشتاوری، ۲۱۹ |
| Rounding | مستقیم کردن، رند کردن، ۲۲ |
| Chance | شانس، اقبال، ۶۸ |
| Personal | شخصی، ۸۰ |
| Counting | شمارش، ۸۸ |
| Percentile | صدک، ۴۰ |
| Confidence Coefficient | ضریب اطمینان، ۲۳۰ |
| Coefficient of Variation | ضریب تغییر، ۵۵ |

| | |
|-----------------------|---------------------------------------|
| Linear Combination | ترکیب خطی، ۱۷۰ |
| Continuity Correction | تصحیح پیوستگی، ۲۰۹ |
| Difference | تفاضل، ۷۲ |
| Symmetric Difference | تفاضل متقارن، ۷۲ |
| Power | توان، ۲۳۴ |
| Chi-square | توان دوم کی، ۲۴۴ |
| Mixed Distribution | توزیع آمیخته، ۱۴۵، ۱۵۶ |
| Tuky, J.W. | توکی، آماردان آمریکائی، ۶۱ |
| Permutation | جایگشت، ۹۰ |
| Disjoint | جدا، ۱۲۶ |
| Table | جدول، ۲۴ |
| Quartile | چارک، ۴۰ |
| Chebychev | چیچف، (۱۸۹۱-۱۸۲۱) ریاضیدان روسی، ۱۷۹ |
| Density | چگالی، ۱۴۸ |
| Joint Density | چگالی توام، ۱۵۸ |
| Quantile | چندک، ۴۰ |
| Skewness | چولگی، ۵۸ |
| Skewed | چوله، ۴۶ |
| Errors | خطاها، ۲۳۳ |
| Random Error | خطای تصادفی، ۲۸۷ |
| Khinchine, A. | خینشین (۱۹۵۹-۱۸۹۲) ریاضیدان روسی، ۲۲۰ |
| Datum | داده، ۲۲ |
| Data | داده‌ها، ۲۲ |
| Standard Data | داده‌های استاندارد، ۵۴ |

| | |
|----------------------|-------------------------|
| Standard Variable | ۱۷۷، دره |
| Continuous Variable | ۱۳۵، سنه |
| Random Variable | ۱۳۲، زدن |
| Categorical Variable | ۲۱، رسم |
| Discrete Variable | سنه، جدا، ۱۳۵ |
| Complement | ۷۱ |
| Independent | ۱۳۲، |
| Observed Value | ۲۱۶، یافته |
| Average | ۳۶، |
| Harmonic Mean | توانقی، ۳۶ |
| Weighted Average | وزنی، با ضریب، ۳۶ |
| Dispersion Measure | پراکنندگی، ۳۹ |
| Central Measure | تمرکز، مرکزی، میانی، ۳۵ |
| T-value | ۳۳۷، |
| Nominal Scale | ن اسمی، ۱۹ |
| Ordered Scale | ن ترتیبی، ۲۰ |
| Scaling | ن سازی، ۱۹ |
| Interval Scale | ن فاصله‌ای، ۲۰ |
| Ratio Scale | ن نسبتی، ۲۱ |
| Curve | نی، ۳۱ |
| Frequency Curve | نی فراوانی، ۳۱ |
| Normal Curve | نی نرمال، ۳۲ |
| Mean | گین، ۱۶۲، ۳۶ |
| Trimmed Mean | گین اصلاح شده، ۲۷ |
| Arithmetic Mean | گین حسابی، ۳۶ |
| Square Root Mean | گین ریشه‌ای، ۳۷ |
| Trigartile Mean | گین سه چارگی، ۲۰ |
| Geometric Mean | گین هندسی، ۳۶ |
| Median | انه، ۳۷ |

| | |
|-------------------------------|---|
| | ف |
| Confidence Interval | فاصله اطمینان، ۲۲۸ |
| Frequency | فراوانی، ۲۵ |
| Cumulative Frequency | فراوانی انباشته، ۲۵ |
| Observed Frequency | فراوانی مشاهده شده، ۲۹۵ |
| Expected Frequency | فراوانی مورد انتظار، ۲۹۶ |
| Relative Frequency | فراوانی نسبی، ۲۵ |
| Ferma, | فرما، (۱۶۶۵-۱۶۰۱) ریاضیدان فرانسوی، ۶۸ |
| Sample Space | فضای نمونه، ۷۰ |
| Fisher, R. A. | فیشر (۱۸۹۰-۱۹۶۲) آماردان انگلیسی، ۶۷، ۲۷۹ |
| | ق |
| Central Limit Theorem | قضیه حد مرکزی، ۲۰۶ |
| Deduction | قیاس، ۲۱۵ |
| | ک |
| Kerrich, J | کریش ریاضیدان انگلیسی، ۷۹ |
| Least Sum of the Square (LSS) | کمترین توانهای دوم، ۲۸۵ |
| Covariance | کواریانس، همپراش، ۱۸۰ |
| Kolomogrov | کولموگروف، (۱۹۰۳-۱۹۸۷) ریاضیدان روسی، ۶۸ |
| | گ |
| Galileo | گالیله، (۱۶۴۲-۱۵۶۴) دانشمند ایتالیایی، ۶۸ |
| Moment | گشتاور، ۵۸ |
| Central Moment | گشتاور مرکزی، ۵۸ |
| | ل |
| Lehmann | لهمن آماردان آمریکایی، ۲۳۷ |
| | م |
| Markov, A. | مارکوف (۱۸۵۶-۱۹۲۲) ریاضیدان روسی، ۱۷۹ |

واژه نامه انگلیسی به فارسی

| | |
|-----------------------|--|
| A | |
| Arithmetic mean | میانگین حسابی، ۳۶ |
| At random | به تصادف، ۱۳۵ |
| Average | معدل، ۳۶ |
| B | |
| Bayes, T | بیز (۱۷۰۱-۱۷۶۱) فیلسوف و کشیش انگلیسی، ۱۱۸ |
| Bernoulli, J | برنولی (۱۶۵۴-۱۷۰۵)، ۱۹۲ |
| Bimodal | دو نمایی، ۴۵ |
| Binomial | دو جمله‌ای، ۱۹۳ |
| Bivariate Normal | نرمال توام، ۲۸۳ |
| Bonferroni | بونفرونی احتمال دان، ۱۱۲ |
| Boole, G | بول (۱۸۱۵-۱۸۶۴) ریاضیدان انگلیسی، ۱۱۲ |
| Borel | برل (۱۸۷۱-۱۹۵۶) ریاضیدان فرانسوی، ۱۰۶ |
| Box Plot | نمودار جعبه‌ای، ۶۴ |
| C | |
| Cardan | دانشمند ایتالیایی (۱۵۷۶-۱۵۰۱)، ۶۸ |
| Categorical Variable | متغیر گروهی، ۲۱ |
| Central Limit Theorem | قضیه حد مرکزی، ۲۰۶ |
| Central Measure | معیار تمرکز، مرکزی، میانی، ۳۵ |
| Central Moment | گشتاور مرکزی، ۵۸ |

| | |
|-------------------------|----------------------------------|
| Sigma Field | میدان سیگما، ۱۰۵ |
| ن | |
| Critical Region | ناحیه بحرانی، ۲۳۳ |
| Inference | نتیجه‌گیری، استنباط، برداشت، ۲۱۴ |
| Standard Normal | نرمال استاندارد، ۳۴ |
| Bivariate Normal | نرمال توام، ۲۸۳ |
| Proportion | نسبت، ۲۶۸ |
| Mode | نما، ۴۳ |
| Graph | نمودار، ۳۰ |
| Box Plot | نمودار جعبه‌ای، ۶۴ |
| Stem Graph | نمودار ساقه‌ای، ۶۱ |
| Random Sample | نمونه تصادفی، ۲۱۶ |
| و | |
| Variance | واریانس، پراش، ۱۷۵، ۵۱ |
| Pooled Variance | واریانس آمیخته، ۲۴۹ |
| ه | |
| Hunter | هانتر، آماردان آمریکائی، ۶۳ |
| Correlation | همبستگی، ۱۸۵ |
| Histogram | هستوگرام، همبند، ۳۰ |
| Equally Likelihood | همشانس، ۷۷ |
| Identically Distributed | همتوزیع، ۱۴۶ |
| ی | |
| Uniform | یکنواخت، ۸۲ |

| | |
|------------------------|---|
| Descriptive Statistics | آمار توصیفی، ۱۷ |
| Deviation | انحراف، ۵۰ |
| Difference | تفاضل، ۷۲ |
| Discrete Data | داده‌های گسسته، جدا، ۲۲ |
| Discrete Variable | متغیر گسسته، جدا، ۱۳۵ |
| Disjoint | جدا، ۱۲۶ |
| Dispersion Measure | معیار پراکنندگی، ۴۹ |
| Distribution Function | تابع توزیع، ۱۳۶ |
| E | |
| Equally Likelihood | همشانسی، ۷۷ |
| Errors | خطاها، ۲۳۳ |
| Estimate | برآورد، تخمین، ۲۱۷ |
| Estimation | برآوردیابی، ۲۱۷ |
| Estimator | برآوردیاب، برآوردگر، برآورد کننده، ۲۱۷ |
| Event | پیشامد، ۷۰ |
| Expectation | امید، انتظار، ۱۶۳ |
| Expected Frequency | فراوانی مورد انتظار، ۲۹۶ |
| Experiment | آزمایش، ۶۹ |
| F | |
| Ferma | فرما، ریاضیدان فرانسوی (۱۶۶۵-۱۶۰۱)، ۶۸ |
| Finite | باپایان، ۷۴ |
| Fisher, R.A. | فیشر (۱۹۶۲-۱۸۹۰) آماردان انگلیسی، ۶۷، ۲۷۹ |
| Frequency | فراوانی، ۲۵ |
| Frequency Curve | منحنی فراوانی، ۳۱ |
| G | |
| Galileo | گالیله، دانشمند ایتالیایی (۱۶۴۲-۱۵۶۴)، ۶۸ |
| Geometric Mean | میانگین هندسی، ۳۶ |

| | |
|--------------------------|--|
| Chance | شانس، اقبال، ۶۸ |
| Chebyshev | چیچف (۱۸۹۱-۱۸۲۱) ریاضیدان روسی، ۱۷۹ |
| Chi-square | توان دوم کی، ۲۲۴ |
| Coded Data | داده‌های تبدیل شده، ۵۲ |
| Coefficient Of Variation | ضریب تغییر، ۵۵ |
| Combination | ترکیب، ۹۲ |
| Complement | متمم، ۷۲ |
| Conditional Probability | احتمال شرطی، ۱۱۳ |
| Confidence Coefficient | ضریب اطمینان، ۲۳۰ |
| Confidence Interval | فاصله اطمینان، ۲۲۸ |
| Continuity Correction | تصحیح پیوستگی، ۲۰۹ |
| Continuous Data | داده‌های پیوسته، ۲۲ |
| Continuous Variable | متغیر پیوسته، ۱۳۵ |
| Correlation | همبستگی، ۱۸۵ |
| Counting | شمارش، ۸۸ |
| Covariance | کوارینانس، همپراش، ۱۸۰ |
| Critical Region | ناحیه بحرانی، ۲۳۳ |
| Cumulative Frequency | فراوانی انباشته، ۲۵ |
| Curve | منحنی، ۳۱ |
| D | |
| Data | داده‌ها، ۲۲ |
| Datum | داده، ۲۲ |
| Decile | دهک، ۴۰ |
| Deduction | قیاس، ۲۱۵ |
| Degenerate | تباهمیده، لابت، ۱۴۵ |
| Degrees Of Freedom | درجه آزادی، ۲۲۴ |
| De Moivre, A. | دو موآور (۱۶۶۷-۱۷۵۴) ریاضیدان فرانسوی، ۲۰۶ |
| Density | چگالی، ۱۴۸ |

| | |
|--------------------|--|
| M | |
| Markov, A. | مارکف (۱۸۵۶-۱۹۹۲) ریاضیدان روسی، ۱۷۹ |
| Maximum Likelihood | راستنمایی ماکزیمم، ۲۲۴ |
| Mean | میانگین، ۱۶۴، ۳۶ |
| Measuring | اندازه‌گیری، ۱۹ |
| Median | میانه، ۲۷ |
| Mixed Distribution | توزیع آمیخته، ۱۴۵، ۱۵۶ |
| MLE | برآورد درستنمایی ماکزیمم، ۲۲۴ |
| Mode | نما، ۴۳ |
| Moment | گشتاور، ۵۸ |
| Moment Method | روش گشتاوری، ۲۱۹ |
| N | |
| Nominal Scale | مقیاس اسمی، ۱۹ |
| Normal Curve | منحنی نرمال، ۳۲ |
| O | |
| Observed Frequency | فراوانی مشاهده شده، ۲۹۵ |
| Observed Value | یافته، مشاهده، ۲۱۶ |
| Occur | رخ دادن، ۷۰ |
| Odds | بخت پولی، ۸۱ |
| Ordinal Scale | مقیاس ترتیبی، ۲۰ |
| P | |
| Pascal | پاسکال، ریاضیدان فرانسوی (۱۶۴۲-۱۶۲۳)، ۶۸ |
| Pearson, K. | پیرسن (۱۸۵۷-۱۹۳۶) آماردان انگلیسی، ۵۹، ۲۹۴ |
| Percentile | صدک، ۴۰ |
| Permutation | جایگشت، ۹۰ |
| Personal | شخصی، ۸۰ |
| Post Probability | احتمال پیشین، ۱۲۲ |

| | |
|-------------------------------|---|
| Goodness of Fit Test | آزمون برازندگی، ۲۹۵ |
| Graph | نمودار، ۳۰ |
| H | |
| Harmonic Mean | معدل توافقی، ۳۶ |
| Histogram | هیستوگرام، همبند، ۳۰ |
| Hunter | مانتر، آماردان آمریکایی، ۶۳ |
| I | |
| Identically Distributed | همتوزیع، ۱۴۶ |
| Independent | مستقل، ۱۲۴ |
| Induction | استقرا، ۲۱۵ |
| Inference | نتیجه‌گیری، استنباط، برداشت، ۲۱۴ |
| Infinite | بی‌پایان، ۷۴ |
| Intersection | براک، ۷۲ |
| Interval Scale | طیاس فاصله‌ای، ۲۰ |
| J | |
| Joint Density | چگالی تزام، ۱۵۸ |
| K | |
| Kerrich, J | کریش، ریاضیدان انگلیسی، ۷۹ |
| Khinchine, A. | خینشین (۱۸۹۲-۱۹۵۹) ریاضیدان روسی، ۲۲۰ |
| Kolmogorov | کولموگوروف، ریاضیدان روسی (۱۹۰۳-۱۹۸۷)، ۶۸ |
| Kurtosis | برجستگی، ۵۸ |
| L | |
| Least Sum Of The Square (LSS) | کمترین توانهای دوم، ۲۸۵ |
| Lehmann | لهمن آماردان آمریکایی، ۲۳۷ |
| Likelihood | راستنمایی، درستنمایی، ۲۲۴ |
| Linear Combination | ترکیب خطی، ۱۷۰ |

| | |
|----------------------|--|
| Square Root Mean | میانگین ریشه‌ای، ۳۷ |
| Standard Data | داده‌های استاندارد، ۵۲ |
| Standard Deviation | انحراف معیار، ۱۷۷، ۵۱ |
| Standard Normal | نرمال استاندارد، ۳۴ |
| Standard Variable | متغیر استاندارد، ۱۷۷ |
| Statistic | آمار، ۲۱۷ |
| Statistics | آماره، ۱۷ |
| Stem Graph | نمودار ساقه‌ای، ۶۱ |
| Stevens, S.S. | استیونز، استاد روانشناسی دانشگاه هاروارد، ۱۹ |
| Stirling | استرلینگ (۱۶۹۲-۱۷۷۰) ریاضیدان انگلیسی، ۹۳ |
| Symmetric Difference | تفاضل متقارن، ۷۲ |
| T | |
| Table | جدول، ۲۴ |
| Test | آزمون، ۲۳۲ |
| Trimmed Mean | میانگین اصلاح شده، ۴۷ |
| Triquartile Mean | میانگین سه چارکی، ۴۰ |
| Tuky, J.W. | توکی، آماردان آمریکایی، ۶۱ |
| U | |
| Uniform | یکنواخت، ۸۲ |
| Union | اجتماع، ۷۱ |
| V | |
| Variance | واریانس، پراش، ۱۷۵، ۵۱ |
| W | |
| Weighted Average | معدل وزنی، با ضریب، ۳۶ |

| | |
|-----------------------|---|
| Probability | احتمال، ۶۸ |
| Proportion | نسبت، ۲۶۸ |
| Poisson, S.D. | پواسن (۱۸۲۰-۱۷۸۱) ریاضیدان فرانسوی، ۱۹۷ |
| Pooled Variance | ولریانس آمیخته، ۲۲۹ |
| Posterior Probability | احتمال پسین، ۱۲۲ |
| Power | توان، ۲۳۲ |
| P-Value | مقدار P - مقدار احتمال، ۲۳۷ |
| Q | |
| Quantile | چندک، ۴۰ |
| Quartile | چارک، ۴۰ |
| R | |
| Random Error | خطای تصادفی، ۲۸۷ |
| Random Experiment | آزمایش تصادفی، ۶۹ |
| Random Sample | نمونه تصادفی، ۲۱۶ |
| Random Variable | متغیر تصادفی، ۱۳۲ |
| Range | برده، ۵۰ |
| Ratio Scale | مقیاس نسبتی، ۲۱ |
| Regression | برگشت، رگرسیون، ۲۷۵، ۲۹۲ |
| Relative Frequency | فراوانی نسبی، ۲۵ |
| Rounding | سُر راست کردن، رند کردن، ۲۲ |
| S | |
| Sample | فضای نمونه، ۷۰ |
| Scaling | مقیاس سازی، ۱۹ |
| Sigma Field | میدان سیگما، ۱۰۵ |
| Significance | با معنایی، ۲۳۷ |
| Skewed | چوله، ۲۶ |
| Skewness | چولگی، ۵۸ |

| r | μ | | | | | | | | |
|---|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
| 0 | 0.9048 | 0.8187 | 0.7408 | 0.6730 | 0.6065 | 0.5488 | 0.4966 | 0.4493 | 0.4066 |
| 1 | 0.9953 | 0.9825 | 0.9631 | 0.9384 | 0.9098 | 0.8781 | 0.8442 | 0.8088 | 0.7725 |
| 2 | 0.9998 | 0.9989 | 0.9964 | 0.9921 | 0.9856 | 0.9769 | 0.9659 | 0.9526 | 0.9371 |
| 3 | 1.0000 | 0.9999 | 0.9997 | 0.9992 | 0.9982 | 0.9966 | 0.9942 | 0.9909 | 0.9865 |
| 4 | | 1.0000 | 1.0000 | 0.9999 | 0.9998 | 0.9996 | 0.9992 | 0.9986 | 0.9977 |
| 5 | | | | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9998 | 0.9997 |
| 6 | | | | | | | 1.0000 | 1.0000 | 1.0000 |

| r | μ | | | | | | | | |
|----|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 1.0 | 1.5 | 2.0 | 2.5 | 3.0 | 3.5 | 4.0 | 4.5 | 5.0 |
| 0 | 0.3679 | 0.2231 | 0.1353 | 0.0821 | 0.0498 | 0.0302 | 0.0183 | 0.0111 | 0.0067 |
| 1 | 0.7358 | 0.5578 | 0.4060 | 0.2873 | 0.1991 | 0.1359 | 0.0916 | 0.0611 | 0.0404 |
| 2 | 0.9197 | 0.8088 | 0.6767 | 0.5438 | 0.4232 | 0.3208 | 0.2381 | 0.1736 | 0.1247 |
| 3 | 0.9810 | 0.9344 | 0.8571 | 0.7576 | 0.6472 | 0.5366 | 0.4335 | 0.3423 | 0.2650 |
| 4 | 0.9963 | 0.9814 | 0.9473 | 0.8912 | 0.8153 | 0.7254 | 0.6288 | 0.5321 | 0.4405 |
| 5 | 0.9994 | 0.9955 | 0.9834 | 0.9580 | 0.9161 | 0.8576 | 0.7851 | 0.7029 | 0.6160 |
| 6 | 0.9999 | 0.9991 | 0.9955 | 0.9858 | 0.9665 | 0.9347 | 0.8893 | 0.8311 | 0.7622 |
| 7 | 1.0000 | 0.9998 | 0.9989 | 0.9958 | 0.9881 | 0.9733 | 0.9489 | 0.9134 | 0.8666 |
| 8 | | 1.0000 | 0.9998 | 0.9989 | 0.9962 | 0.9901 | 0.9786 | 0.9577 | 0.9319 |
| 9 | | | 1.0000 | 0.9997 | 0.9989 | 0.9967 | 0.9919 | 0.9829 | 0.9682 |
| 10 | | | | 0.9999 | 0.9997 | 0.9990 | 0.9972 | 0.9933 | 0.9863 |
| 11 | | | | 1.0000 | 0.9999 | 0.9997 | 0.9991 | 0.9976 | 0.9945 |
| 12 | | | | | 1.0000 | 0.9999 | 0.9997 | 0.9992 | 0.9980 |
| 13 | | | | | | 1.0000 | 0.9999 | 0.9997 | 0.9993 |
| 14 | | | | | | | 1.0000 | 0.9999 | 0.9998 |
| 15 | | | | | | | | 1.0000 | 0.9999 |
| 16 | | | | | | | | | 1.0000 |

| r | μ | | | | | | | | |
|----|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | 5.5 | 6.0 | 6.5 | 7.0 | 7.5 | 8.0 | 8.5 | 9.0 | 9.5 |
| 0 | 0.0041 | 0.0025 | 0.0015 | 0.0009 | 0.0006 | 0.0003 | 0.0002 | 0.0001 | 0.0001 |
| 1 | 0.0266 | 0.0174 | 0.0113 | 0.0073 | 0.0047 | 0.0030 | 0.0019 | 0.0012 | 0.0008 |
| 2 | 0.0684 | 0.0620 | 0.0430 | 0.0296 | 0.0203 | 0.0138 | 0.0093 | 0.0062 | 0.0042 |
| 3 | 0.2017 | 0.1512 | 0.1118 | 0.0818 | 0.0591 | 0.0424 | 0.0301 | 0.0212 | 0.0149 |
| 4 | 0.3575 | 0.2851 | 0.2237 | 0.1730 | 0.1321 | 0.0996 | 0.0744 | 0.0550 | 0.0403 |
| 5 | 0.5289 | 0.4457 | 0.3690 | 0.3007 | 0.2414 | 0.1912 | 0.1496 | 0.1157 | 0.0885 |
| 6 | 0.6860 | 0.6063 | 0.5265 | 0.4497 | 0.3782 | 0.3134 | 0.2562 | 0.2068 | 0.1649 |
| 7 | 0.8095 | 0.7440 | 0.6728 | 0.5987 | 0.5246 | 0.4530 | 0.3856 | 0.3239 | 0.2687 |
| 8 | 0.8944 | 0.8472 | 0.7916 | 0.7291 | 0.6620 | 0.5925 | 0.5231 | 0.4557 | 0.3918 |
| 9 | 0.9462 | 0.9161 | 0.8774 | 0.8305 | 0.7764 | 0.7166 | 0.6530 | 0.5874 | 0.5218 |
| 10 | 0.9747 | 0.9574 | 0.9332 | 0.9015 | 0.8622 | 0.8159 | 0.7634 | 0.7060 | 0.6453 |
| 11 | 0.9890 | 0.9799 | 0.9661 | 0.9466 | 0.9208 | 0.8881 | 0.8487 | 0.8030 | 0.7520 |
| 12 | 0.9955 | 0.9912 | 0.9840 | 0.9730 | 0.9573 | 0.9362 | 0.9091 | 0.8758 | 0.8364 |
| 13 | 0.9983 | 0.9964 | 0.9929 | 0.9872 | 0.9784 | 0.9658 | 0.9486 | 0.9261 | 0.8981 |
| 14 | 0.9994 | 0.9986 | 0.9970 | 0.9943 | 0.9897 | 0.9827 | 0.9726 | 0.9585 | 0.9400 |
| 15 | 0.9998 | 0.9995 | 0.9988 | 0.9976 | 0.9954 | 0.9918 | 0.9862 | 0.9780 | 0.9665 |
| 16 | 0.9999 | 0.9998 | 0.9996 | 0.9990 | 0.9980 | 0.9963 | 0.9934 | 0.9889 | 0.9823 |
| 17 | 1.0000 | 0.9999 | 0.9998 | 0.9996 | 0.9992 | 0.9984 | 0.9970 | 0.9947 | 0.9911 |
| 18 | | 1.0000 | 0.9999 | 0.9999 | 0.9997 | 0.9994 | 0.9987 | 0.9976 | 0.9957 |
| 19 | | | 1.0000 | 1.0000 | 0.9999 | 0.9997 | 0.9995 | 0.9989 | 0.9980 |
| 20 | | | | | 1.0000 | 0.9999 | 0.9998 | 0.9996 | 0.9991 |
| 21 | | | | | | 1.0000 | 0.9999 | 0.9998 | 0.9996 |
| 22 | | | | | | | 1.0000 | 0.9999 | 0.9999 |
| 23 | | | | | | | | 1.0000 | 0.9999 |
| 24 | | | | | | | | | 1.0000 |

| r | p | P | | | | | | | | | |
|----|----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | | 0.10 | 0.20 | 0.25 | 0.30 | 0.40 | 0.50 | 0.60 | 0.70 | 0.80 | 0.90 |
| 5 | 0 | 0.5905 | 0.3277 | 0.2373 | 0.1681 | 0.0778 | 0.0312 | 0.0102 | 0.0024 | 0.0003 | 0.0000 |
| | 1 | 0.9185 | 0.7373 | 0.6328 | 0.5282 | 0.3370 | 0.1875 | 0.0870 | 0.0308 | 0.0067 | 0.0005 |
| | 2 | 0.9914 | 0.9421 | 0.8965 | 0.8369 | 0.6826 | 0.5000 | 0.3174 | 0.1631 | 0.0579 | 0.0086 |
| | 3 | 0.9995 | 0.9933 | 0.9844 | 0.9692 | 0.9130 | 0.8125 | 0.6630 | 0.4718 | 0.2627 | 0.0815 |
| | 4 | 1.0000 | 0.9997 | 0.9990 | 0.9976 | 0.9898 | 0.9688 | 0.9222 | 0.8319 | 0.6723 | 0.4095 |
| | 5 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 10 | 0 | 0.3487 | 0.1074 | 0.0563 | 0.0282 | 0.0060 | 0.0010 | 0.0001 | 0.0000 | 0.0000 | 0.0000 |
| | 1 | 0.7361 | 0.3758 | 0.2440 | 0.1493 | 0.0464 | 0.0107 | 0.0017 | 0.0001 | 0.0000 | 0.0000 |
| | 2 | 0.9298 | 0.6778 | 0.5256 | 0.3828 | 0.1673 | 0.0547 | 0.0123 | 0.0016 | 0.0001 | 0.0000 |
| | 3 | 0.9872 | 0.8791 | 0.7759 | 0.6496 | 0.3823 | 0.1719 | 0.0548 | 0.0106 | 0.0009 | 0.0000 |
| | 4 | 0.9984 | 0.9672 | 0.9219 | 0.8497 | 0.6331 | 0.3770 | 0.1662 | 0.0474 | 0.0064 | 0.0002 |
| | 5 | 0.9999 | 0.9936 | 0.9803 | 0.9527 | 0.8338 | 0.6230 | 0.3669 | 0.1503 | 0.0328 | 0.0016 |
| | 6 | 1.0000 | 0.9991 | 0.9965 | 0.9894 | 0.9452 | 0.8281 | 0.6177 | 0.3504 | 0.1209 | 0.0128 |
| | 7 | 1.0000 | 0.9999 | 0.9996 | 0.9984 | 0.9877 | 0.9453 | 0.8327 | 0.6172 | 0.3222 | 0.0702 |
| | 8 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9983 | 0.9893 | 0.9536 | 0.8507 | 0.6242 | 0.2639 |
| | 9 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9990 | 0.9940 | 0.9718 | 0.8526 | 0.6513 |
| | 10 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 15 | 0 | 0.2059 | 0.0352 | 0.0134 | 0.0047 | 0.0005 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 1 | 0.5490 | 0.1671 | 0.0802 | 0.0353 | 0.0052 | 0.0005 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 2 | 0.8159 | 0.3980 | 0.2161 | 0.1268 | 0.0271 | 0.0037 | 0.0003 | 0.0000 | 0.0000 | 0.0000 |
| | 3 | 0.9444 | 0.6482 | 0.4613 | 0.2969 | 0.0905 | 0.0176 | 0.0019 | 0.0001 | 0.0000 | 0.0000 |
| | 4 | 0.9873 | 0.8358 | 0.6865 | 0.5155 | 0.2173 | 0.0592 | 0.0094 | 0.0007 | 0.0000 | 0.0000 |
| | 5 | 0.9978 | 0.9389 | 0.8516 | 0.7216 | 0.4032 | 0.1509 | 0.0338 | 0.0037 | 0.0001 | 0.0000 |
| | 6 | 0.9997 | 0.9819 | 0.9434 | 0.8689 | 0.6098 | 0.3036 | 0.0951 | 0.0152 | 0.0008 | 0.0000 |
| | 7 | 1.0000 | 0.9958 | 0.9827 | 0.9500 | 0.7869 | 0.5000 | 0.2131 | 0.0500 | 0.0042 | 0.0000 |
| | 8 | 1.0000 | 0.9992 | 0.9958 | 0.9848 | 0.9050 | 0.6964 | 0.3902 | 0.1311 | 0.0181 | 0.0003 |
| | 9 | 1.0000 | 0.9999 | 0.9992 | 0.9963 | 0.9662 | 0.8491 | 0.5968 | 0.2784 | 0.0611 | 0.0023 |
| | 10 | 1.0000 | 1.0000 | 0.9999 | 0.9993 | 0.9907 | 0.9408 | 0.7827 | 0.4845 | 0.1642 | 0.0127 |
| | 11 | 1.0000 | 1.0000 | 1.0000 | 0.9999 | 0.9999 | 0.9981 | 0.9824 | 0.9095 | 0.7031 | 0.3518 |
| | 12 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9997 | 0.9963 | 0.9729 | 0.8732 | 0.6020 |
| | 13 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9995 | 0.9948 | 0.9647 | 0.8329 |
| | 14 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9995 | 0.9953 | 0.9648 |
| | 15 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 20 | 0 | 0.1216 | 0.0115 | 0.0032 | 0.0008 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 1 | 0.3917 | 0.0692 | 0.0243 | 0.0076 | 0.0005 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 2 | 0.6769 | 0.2061 | 0.0913 | 0.0355 | 0.0016 | 0.0002 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 3 | 0.8670 | 0.4114 | 0.2252 | 0.1071 | 0.0160 | 0.0013 | 0.0001 | 0.0000 | 0.0000 | 0.0000 |
| | 4 | 0.9568 | 0.6296 | 0.4148 | 0.2375 | 0.1256 | 0.0207 | 0.0016 | 0.0000 | 0.0000 | 0.0000 |
| | 5 | 0.9887 | 0.8042 | 0.6172 | 0.4164 | 0.2500 | 0.0577 | 0.0065 | 0.0003 | 0.0000 | 0.0000 |
| | 6 | 0.9976 | 0.9133 | 0.7858 | 0.6080 | 0.4159 | 0.1316 | 0.0210 | 0.0013 | 0.0000 | 0.0000 |
| | 7 | 0.9996 | 0.9679 | 0.8982 | 0.7723 | 0.5956 | 0.2517 | 0.0565 | 0.0051 | 0.0001 | 0.0000 |
| | 8 | 0.9999 | 0.9900 | 0.9391 | 0.8867 | 0.7553 | 0.4119 | 0.1275 | 0.0171 | 0.0006 | 0.0000 |
| | 9 | 1.0000 | 0.9974 | 0.9861 | 0.9520 | 0.8725 | 0.5881 | 0.2447 | 0.0480 | 0.0026 | 0.0000 |
| | 10 | 1.0000 | 0.9994 | 0.9961 | 0.9829 | 0.9435 | 0.7483 | 0.4044 | 0.1133 | 0.0100 | 0.0001 |
| | 11 | 1.0000 | 0.9999 | 0.9991 | 0.9949 | 0.9435 | 0.7483 | 0.4044 | 0.1133 | 0.0100 | 0.0001 |
| | 12 | 1.0000 | 1.0000 | 0.9998 | 0.9987 | 0.9790 | 0.8684 | 0.5841 | 0.2277 | 0.0321 | 0.0004 |
| | 13 | 1.0000 | 1.0000 | 1.0000 | 0.9997 | 0.9935 | 0.9423 | 0.7500 | 0.3920 | 0.0867 | 0.0024 |
| | 14 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9984 | 0.9793 | 0.8744 | 0.5836 | 0.1958 | 0.0113 |
| | 15 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9997 | 0.9941 | 0.9490 | 0.7625 | 0.3704 | 0.0432 |
| | 16 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9997 | 0.9840 | 0.8929 | 0.5886 | 0.1330 |
| | 17 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9995 | 0.9945 | 0.9645 | 0.7939 |
| | 18 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9998 | 0.9954 | 0.9645 | 0.9308 |
| | 19 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9992 | 0.9885 | 0.8784 |
| | 20 | | | | | | | | | | |

$\alpha = .05$

| $m_1 \backslash m_2$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|----------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1 | 161.45 | 199.50 | 215.71 | 224.58 | 230.16 | 233.99 | 236.77 | 238.88 | 240.54 |
| 2 | 18.513 | 19.000 | 19.164 | 19.247 | 19.296 | 19.330 | 19.353 | 19.371 | 19.385 |
| 3 | 10.128 | 9.5521 | 9.2766 | 9.1172 | 9.0135 | 8.9406 | 8.8868 | 8.8452 | 8.8123 |
| 4 | 7.7086 | 6.9443 | 6.5914 | 6.3883 | 6.2560 | 6.1631 | 6.0942 | 6.0410 | 5.9988 |
| 5 | 6.6079 | 5.7861 | 5.4095 | 5.1922 | 5.0503 | 4.9503 | 4.8759 | 4.8183 | 4.7725 |
| 6 | 5.9874 | 5.1433 | 4.7571 | 4.5337 | 4.3874 | 4.2839 | 4.2066 | 4.1468 | 4.0990 |
| 7 | 5.5914 | 4.7374 | 4.3468 | 4.1203 | 3.9715 | 3.8660 | 3.7870 | 3.7257 | 3.6767 |
| 8 | 5.3177 | 4.4590 | 4.0662 | 3.8378 | 3.6875 | 3.5806 | 3.5005 | 3.4381 | 3.3881 |
| 9 | 5.1174 | 4.2565 | 3.8626 | 3.6331 | 3.4817 | 3.3738 | 3.2927 | 3.2296 | 3.1789 |
| 10 | 4.9646 | 4.1028 | 3.7083 | 3.4780 | 3.3258 | 3.2172 | 3.1355 | 3.0717 | 3.0204 |
| 11 | 4.8443 | 3.9823 | 3.5874 | 3.3567 | 3.2039 | 3.0946 | 3.0123 | 2.9480 | 2.8962 |
| 12 | 4.7472 | 3.8853 | 3.4903 | 3.2592 | 3.1059 | 2.9961 | 2.9134 | 2.8486 | 2.7964 |
| 13 | 4.6672 | 3.8056 | 3.4105 | 3.1791 | 3.0254 | 2.9153 | 2.8321 | 2.7669 | 2.7144 |
| 14 | 4.6001 | 3.7389 | 3.3439 | 3.1122 | 2.9582 | 2.8477 | 2.7642 | 2.6987 | 2.6458 |
| 15 | 4.5431 | 3.6823 | 3.2874 | 3.0556 | 2.9013 | 2.7905 | 2.7066 | 2.6408 | 2.5876 |
| 16 | 4.4940 | 3.6337 | 3.2389 | 3.0069 | 2.8524 | 2.7413 | 2.6572 | 2.5911 | 2.5377 |
| 17 | 4.4513 | 3.5915 | 3.1968 | 2.9647 | 2.8100 | 2.6987 | 2.6143 | 2.5480 | 2.4943 |
| 18 | 4.4139 | 3.5546 | 3.1599 | 2.9277 | 2.7729 | 2.6613 | 2.5767 | 2.5102 | 2.4563 |
| 19 | 4.3808 | 3.5219 | 3.1274 | 2.8951 | 2.7401 | 2.6283 | 2.5435 | 2.4768 | 2.4227 |
| 20 | 4.3513 | 3.4928 | 3.0984 | 2.8661 | 2.7109 | 2.5990 | 2.5140 | 2.4471 | 2.3928 |
| 21 | 4.3248 | 3.4668 | 3.0725 | 2.8401 | 2.6848 | 2.5727 | 2.4876 | 2.4205 | 2.3661 |
| 22 | 4.3009 | 3.4434 | 3.0491 | 2.8167 | 2.6613 | 2.5491 | 2.4638 | 2.3965 | 2.3419 |
| 23 | 4.2793 | 3.4221 | 3.0280 | 2.7955 | 2.6400 | 2.5277 | 2.4422 | 2.3748 | 2.3201 |
| 24 | 4.2597 | 3.4028 | 3.0088 | 2.7763 | 2.6207 | 2.5082 | 2.4226 | 2.3551 | 2.3002 |
| 25 | 4.2417 | 3.3852 | 2.9912 | 2.7587 | 2.6030 | 2.4904 | 2.4047 | 2.3371 | 2.2821 |
| 26 | 4.2252 | 3.3690 | 2.9751 | 2.7426 | 2.5868 | 2.4741 | 2.3883 | 2.3205 | 2.2655 |
| 27 | 4.2100 | 3.3541 | 2.9604 | 2.7278 | 2.5719 | 2.4591 | 2.3732 | 2.3053 | 2.2501 |
| 28 | 4.1960 | 3.3404 | 2.9467 | 2.7141 | 2.5581 | 2.4453 | 2.3593 | 2.2913 | 2.2360 |
| 29 | 4.1830 | 3.3277 | 2.9340 | 2.7014 | 2.5454 | 2.4324 | 2.3463 | 2.2782 | 2.2229 |
| 30 | 4.1709 | 3.3158 | 2.9223 | 2.6896 | 2.5336 | 2.4205 | 2.3343 | 2.2662 | 2.2107 |
| 40 | 4.0848 | 3.2317 | 2.8387 | 2.6060 | 2.4493 | 2.3359 | 2.2490 | 2.1802 | 2.1240 |
| 60 | 4.0012 | 3.1504 | 2.7581 | 2.5252 | 2.3683 | 2.2540 | 2.1665 | 2.0970 | 2.0401 |
| 120 | 3.9201 | 3.0718 | 2.6802 | 2.4472 | 2.2900 | 2.1750 | 2.0867 | 2.0164 | 1.9588 |
| ∞ | 3.8415 | 2.9957 | 2.6049 | 2.3719 | 2.2141 | 2.0986 | 2.0096 | 1.9384 | 1.8799 |

$\alpha = .10$

| | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|----------|
| 60.195 | 60.705 | 61.220 | 61.740 | 62.002 | 62.265 | 62.529 | 62.794 | 63.061 | 63.328 | |
| 9.3916 | 9.4081 | 9.4247 | 9.4413 | 9.4496 | 9.4579 | 9.4663 | 9.4746 | 9.4829 | 9.4913 | |
| 5.2304 | 5.2156 | 5.2003 | 5.1845 | 5.1764 | 5.1681 | 5.1597 | 5.1512 | 5.1425 | 5.1337 | |
| 3.9199 | 3.8955 | 3.8689 | 3.8443 | 3.8310 | 3.8174 | 3.8036 | 3.7896 | 3.7753 | 3.7607 | |
| 3.2974 | 3.2682 | 3.2380 | 3.2067 | 3.1905 | 3.1741 | 3.1573 | 3.1402 | 3.1228 | 3.1050 | |
| 2.9369 | 2.9047 | 2.8712 | 2.8363 | 2.8183 | 2.8000 | 2.7812 | 2.7620 | 2.7423 | 2.7222 | |
| 2.7025 | 2.6681 | 2.6322 | 2.5947 | 2.5753 | 2.5555 | 2.5351 | 2.5142 | 2.4928 | 2.4708 | |
| 2.5380 | 2.5020 | 2.4642 | 2.4246 | 2.4041 | 2.3830 | 2.3614 | 2.3391 | 2.3162 | 2.2926 | |
| 2.4163 | 2.3789 | 2.3396 | 2.2983 | 2.2768 | 2.2547 | 2.2320 | 2.2085 | 2.1843 | 2.1592 | |
| 2.3226 | 2.2841 | 2.2435 | 2.2007 | 2.1784 | 2.1554 | 2.1317 | 2.1072 | 2.0818 | 2.0554 | |
| 2.2482 | 2.2087 | 2.1671 | 2.1230 | 2.1000 | 2.0762 | 2.0516 | 2.0261 | 1.9997 | 1.9721 | |
| 2.1878 | 2.1474 | 2.1049 | 2.0597 | 2.0360 | 2.0115 | 1.9861 | 1.9597 | 1.9323 | 1.9036 | |
| 2.1376 | 2.0966 | 2.0532 | 2.0070 | 1.9827 | 1.9576 | 1.9315 | 1.9043 | 1.8759 | 1.8462 | |
| 2.0954 | 2.0537 | 2.0095 | 1.9625 | 1.9377 | 1.9119 | 1.8852 | 1.8572 | 1.8280 | 1.7973 | |
| 2.0593 | 2.0171 | 1.9722 | 1.9243 | 1.8990 | 1.8728 | 1.8454 | 1.8168 | 1.7867 | 1.7551 | |
| 2.0281 | 1.9854 | 1.9399 | 1.8913 | 1.8656 | 1.8388 | 1.8108 | 1.7816 | 1.7507 | 1.7182 | |
| 2.0009 | 1.9577 | 1.9117 | 1.8624 | 1.8362 | 1.8090 | 1.7805 | 1.7506 | 1.7191 | 1.6856 | |
| 1.9770 | 1.9333 | 1.8868 | 1.8368 | 1.8103 | 1.7827 | 1.7537 | 1.7232 | 1.6910 | 1.6567 | |
| 1.9557 | 1.9117 | 1.8647 | 1.8142 | 1.7873 | 1.7592 | 1.7298 | 1.6988 | 1.6659 | 1.6308 | |
| 1.9367 | 1.8924 | 1.8449 | 1.7938 | 1.7667 | 1.7382 | 1.7083 | 1.6768 | 1.6433 | 1.6074 | |
| 1.9197 | 1.8750 | 1.8272 | 1.7756 | 1.7481 | 1.7193 | 1.6890 | 1.6569 | 1.6228 | 1.5862 | |
| 1.9043 | 1.8593 | 1.8111 | 1.7590 | 1.7312 | 1.7021 | 1.6714 | 1.6389 | 1.6042 | 1.5668 | |
| 1.8903 | 1.8450 | 1.7964 | 1.7439 | 1.7159 | 1.6864 | 1.6554 | 1.6224 | 1.5871 | 1.5490 | |
| 1.8775 | 1.8319 | 1.7831 | 1.7302 | 1.7019 | 1.6721 | 1.6407 | 1.6073 | 1.5715 | 1.5327 | |
| 1.8658 | 1.8200 | 1.7708 | 1.7175 | 1.6890 | 1.6589 | 1.6272 | 1.5934 | 1.5570 | 1.5176 | |
| 1.8550 | 1.8090 | 1.7596 | 1.7059 | 1.6771 | 1.6468 | 1.6147 | 1.5805 | 1.5437 | 1.5036 | |
| 1.8451 | 1.7989 | 1.7492 | 1.6951 | 1.6662 | 1.6356 | 1.6032 | 1.5686 | 1.5313 | 1.4906 | |
| 1.8359 | 1.7895 | 1.7395 | 1.6852 | 1.6560 | 1.6252 | 1.5925 | 1.5575 | 1.5198 | 1.4784 | |
| 1.8274 | 1.7808 | 1.7306 | 1.6759 | 1.6465 | 1.6155 | 1.5825 | 1.5472 | 1.5090 | 1.4670 | |
| 1.8195 | 1.7727 | 1.7223 | 1.6673 | 1.6377 | 1.6065 | 1.5732 | 1.5376 | 1.4989 | 1.4564 | |
| 1.7627 | 1.7146 | 1.6624 | 1.6052 | 1.5741 | 1.5411 | 1.5056 | 1.4672 | 1.4248 | 1.3769 | |
| 1.7070 | 1.6574 | 1.6034 | 1.5435 | 1.5107 | 1.4755 | 1.4373 | 1.3952 | 1.3476 | 1.2915 | |
| 1.6524 | 1.6012 | 1.5450 | 1.4821 | 1.4472 | 1.4094 | 1.3676 | 1.3203 | 1.2646 | 1.1926 | |
| 1.5987 | 1.5458 | 1.4871 | 1.4206 | 1.3832 | 1.3419 | 1.2951 | 1.2400 | 1.1686 | 1.0000 | |

$\alpha = .01$

| F_{α} | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|--------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1 | 4052.2 | 4999.5 | 5403.3 | 5624.6 | 5763.7 | 5859.0 | 5928.3 | 5981.6 | 6022.5 |
| 2 | 98.503 | 99.000 | 99.166 | 99.249 | 99.299 | 99.332 | 99.356 | 99.374 | 99.388 |
| 3 | 34.116 | 30.817 | 29.457 | 28.710 | 28.237 | 27.911 | 27.672 | 27.489 | 27.345 |
| 4 | 21.198 | 18.000 | 16.694 | 15.977 | 15.522 | 15.207 | 14.976 | 14.799 | 14.659 |
| 5 | 16.258 | 13.274 | 12.060 | 11.392 | 10.967 | 10.672 | 10.456 | 10.289 | 10.158 |
| 6 | 13.745 | 10.925 | 9.7795 | 9.1483 | 8.7459 | 8.4661 | 8.2600 | 8.1016 | 7.9761 |
| 7 | 12.246 | 9.5466 | 8.4513 | 7.8467 | 7.4604 | 7.1914 | 6.9928 | 6.8401 | 6.7188 |
| 8 | 11.259 | 8.6491 | 7.5910 | 7.0060 | 6.6318 | 6.3707 | 6.1776 | 6.0289 | 5.9106 |
| 9 | 10.561 | 8.0215 | 6.9919 | 6.4221 | 6.0569 | 5.8018 | 5.6129 | 5.4671 | 5.3511 |
| 10 | 10.044 | 7.5594 | 6.5523 | 5.9943 | 5.6363 | 5.3858 | 5.2001 | 5.0567 | 4.9424 |
| 11 | 9.6460 | 7.2057 | 6.2167 | 5.6683 | 5.3160 | 5.0692 | 4.8861 | 4.7445 | 4.6315 |
| 12 | 9.3302 | 6.9266 | 5.9526 | 5.4119 | 5.0643 | 4.8206 | 4.6395 | 4.4994 | 4.3875 |
| 13 | 9.0738 | 6.7010 | 5.7394 | 5.2053 | 4.8616 | 4.6204 | 4.4410 | 4.3021 | 4.1911 |
| 14 | 8.8616 | 6.5149 | 5.5639 | 5.0354 | 4.6950 | 4.4558 | 4.2779 | 4.1399 | 4.0297 |
| 15 | 8.6831 | 6.3589 | 5.4170 | 4.8932 | 4.5556 | 4.3183 | 4.1415 | 4.0045 | 3.8948 |
| 16 | 8.5310 | 6.2262 | 5.2922 | 4.7726 | 4.4374 | 4.2016 | 4.0259 | 3.8896 | 3.7804 |
| 17 | 8.3997 | 6.1121 | 5.1850 | 4.6690 | 4.3359 | 4.1015 | 3.9267 | 3.7910 | 3.6822 |
| 18 | 8.2854 | 6.0129 | 5.0919 | 4.5790 | 4.2479 | 4.0146 | 3.8406 | 3.7054 | 3.5971 |
| 19 | 8.1850 | 5.9259 | 5.0103 | 4.5003 | 4.1708 | 3.9386 | 3.7653 | 3.6305 | 3.5225 |
| 20 | 8.0960 | 5.8489 | 4.9382 | 4.4307 | 4.1027 | 3.8714 | 3.6987 | 3.5644 | 3.4567 |
| 21 | 8.0166 | 5.7804 | 4.8740 | 4.3688 | 4.0421 | 3.8117 | 3.6396 | 3.5056 | 3.3981 |
| 22 | 7.9454 | 5.7190 | 4.8166 | 4.3134 | 3.9880 | 3.7583 | 3.5867 | 3.4530 | 3.3458 |
| 23 | 7.8811 | 5.6637 | 4.7649 | 4.2635 | 3.9392 | 3.7102 | 3.5390 | 3.4057 | 3.2986 |
| 24 | 7.8229 | 5.6136 | 4.7181 | 4.2184 | 3.8951 | 3.6667 | 3.4959 | 3.3629 | 3.2560 |
| 25 | 7.7698 | 5.5680 | 4.6755 | 4.1774 | 3.8550 | 3.6272 | 3.4568 | 3.3239 | 3.2172 |
| 26 | 7.7213 | 5.5263 | 4.6366 | 4.1400 | 3.8183 | 3.5911 | 3.4210 | 3.2884 | 3.1818 |
| 27 | 7.6767 | 5.4881 | 4.6009 | 4.1056 | 3.7848 | 3.5580 | 3.3882 | 3.2558 | 3.1494 |
| 28 | 7.6356 | 5.4529 | 4.5681 | 4.0740 | 3.7539 | 3.5276 | 3.3581 | 3.2259 | 3.1195 |
| 29 | 7.5976 | 5.4205 | 4.5378 | 4.0449 | 3.7254 | 3.4995 | 3.3302 | 3.1982 | 3.0920 |
| 30 | 7.5625 | 5.3904 | 4.5097 | 4.0179 | 3.6990 | 3.4735 | 3.3045 | 3.1726 | 3.0665 |
| 40 | 7.3141 | 5.1785 | 4.3126 | 3.8283 | 3.5138 | 3.2910 | 3.1238 | 2.9930 | 2.8876 |
| 60 | 7.0771 | 4.9774 | 4.1259 | 3.6491 | 3.3389 | 3.1187 | 2.9530 | 2.8233 | 2.7185 |
| 120 | 6.8510 | 4.7865 | 3.9493 | 3.4796 | 3.1735 | 2.9559 | 2.7918 | 2.6629 | 2.5586 |
| ∞ | 6.6349 | 4.6052 | 3.7816 | 3.3192 | 3.0173 | 2.8020 | 2.6393 | 2.5113 | 2.4073 |

$\alpha = .05$

| F_{α} | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|--------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|----------|
| 1 | 241.88 | 243.91 | 245.95 | 248.01 | 249.05 | 250.09 | 251.14 | 252.20 | 253.25 | 254.32 |
| 2 | 19.396 | 19.413 | 19.429 | 19.446 | 19.454 | 19.462 | 19.471 | 19.479 | 19.487 | 19.496 |
| 3 | 8.7855 | 8.7446 | 8.7029 | 8.6602 | 8.6385 | 8.6166 | 8.5944 | 8.5720 | 8.5494 | 8.5265 |
| 4 | 5.9644 | 5.9117 | 5.8578 | 5.8025 | 5.7744 | 5.7459 | 5.7170 | 5.6878 | 5.6581 | 5.6281 |
| 5 | 4.7351 | 4.6777 | 4.6188 | 4.5581 | 4.5272 | 4.4957 | 4.4638 | 4.4314 | 4.3984 | 4.3650 |
| 6 | 4.0600 | 3.9999 | 3.9381 | 3.8742 | 3.8415 | 3.8082 | 3.7743 | 3.7398 | 3.7047 | 3.6688 |
| 7 | 3.6365 | 3.5747 | 3.5108 | 3.4445 | 3.4105 | 3.3758 | 3.3404 | 3.3043 | 3.2674 | 3.2298 |
| 8 | 3.3472 | 3.2840 | 3.2184 | 3.1503 | 3.1152 | 3.0794 | 3.0428 | 3.0053 | 2.9669 | 2.9276 |
| 9 | 3.1373 | 3.0729 | 3.0061 | 2.9365 | 2.9005 | 2.8637 | 2.8259 | 2.7872 | 2.7475 | 2.7067 |
| 10 | 2.9782 | 2.9130 | 2.8450 | 2.7740 | 2.7372 | 2.6996 | 2.6609 | 2.6211 | 2.5801 | 2.5379 |
| 11 | 2.8536 | 2.7876 | 2.7186 | 2.6464 | 2.6090 | 2.5705 | 2.5309 | 2.4901 | 2.4480 | 2.4045 |
| 12 | 2.7534 | 2.6866 | 2.6169 | 2.5436 | 2.5055 | 2.4663 | 2.4259 | 2.3842 | 2.3410 | 2.2962 |
| 13 | 2.6710 | 2.6037 | 2.5331 | 2.4589 | 2.4202 | 2.3803 | 2.3392 | 2.2966 | 2.2524 | 2.2064 |
| 14 | 2.6021 | 2.5342 | 2.4630 | 2.3879 | 2.3487 | 2.3082 | 2.2664 | 2.2230 | 2.1778 | 2.1307 |
| 15 | 2.5437 | 2.4753 | 2.4035 | 2.3275 | 2.2878 | 2.2468 | 2.2043 | 2.1601 | 2.1141 | 2.0658 |
| 16 | 2.4935 | 2.4247 | 2.3522 | 2.2756 | 2.2354 | 2.1938 | 2.1507 | 2.1058 | 2.0589 | 2.0096 |
| 17 | 2.4499 | 2.3807 | 2.3077 | 2.2304 | 2.1898 | 2.1477 | 2.1040 | 2.0584 | 2.0107 | 1.9604 |
| 18 | 2.4117 | 2.3421 | 2.2686 | 2.1906 | 2.1497 | 2.1071 | 2.0629 | 2.0166 | 1.9681 | 1.9168 |
| 19 | 2.3779 | 2.3080 | 2.2341 | 2.1555 | 2.1141 | 2.0712 | 2.0264 | 1.9796 | 1.9302 | 1.8780 |
| 20 | 2.3479 | 2.2776 | 2.2033 | 2.1242 | 2.0825 | 2.0391 | 1.9938 | 1.9464 | 1.8963 | 1.8432 |
| 21 | 2.3210 | 2.2504 | 2.1757 | 2.0960 | 2.0540 | 2.0102 | 1.9645 | 1.9165 | 1.8657 | 1.8117 |
| 22 | 2.2967 | 2.2258 | 2.1508 | 2.0707 | 2.0283 | 1.9842 | 1.9380 | 1.8895 | 1.8380 | 1.7831 |
| 23 | 2.2747 | 2.2036 | 2.1282 | 2.0476 | 2.0050 | 1.9605 | 1.9139 | 1.8649 | 1.8128 | 1.7570 |
| 24 | 2.2547 | 2.1834 | 2.1077 | 2.0267 | 1.9838 | 1.9390 | 1.8920 | 1.8424 | 1.7897 | 1.7331 |
| 25 | 2.2365 | 2.1649 | 2.0889 | 2.0075 | 1.9643 | 1.9192 | 1.8718 | 1.8217 | 1.7684 | 1.7110 |
| 26 | 2.2197 | 2.1479 | 2.0716 | 1.9898 | 1.9464 | 1.9010 | 1.8533 | 1.8027 | 1.7488 | 1.6906 |
| 27 | 2.2043 | 2.1323 | 2.0558 | 1.9736 | 1.9299 | 1.8842 | 1.8361 | 1.7851 | 1.7307 | 1.6717 |
| 28 | 2.1900 | 2.1179 | 2.0411 | 1.9586 | 1.9147 | 1.8687 | 1.8203 | 1.7689 | 1.7138 | 1.6541 |
| 29 | 2.1768 | 2.1045 | 2.0275 | 1.9446 | 1.9005 | 1.8543 | 1.8055 | 1.7537 | 1.6981 | 1.6377 |
| 30 | 2.1646 | 2.0921 | 2.0148 | 1.9317 | 1.8874 | 1.8409 | 1.7918 | 1.7396 | 1.6835 | 1.6223 |
| 40 | 2.0772 | 2.0035 | 1.9245 | 1.8389 | 1.7929 | 1.7444 | 1.6928 | 1.6373 | 1.5766 | 1.5089 |
| 60 | 1.9926 | 1.9174 | 1.8364 | 1.7480 | 1.7001 | 1.6491 | 1.5943 | 1.5343 | 1.4673 | 1.3893 |
| 120 | 1.9105 | 1.8337 | 1.7505 | 1.6587 | 1.6084 | 1.5543 | 1.4952 | 1.4290 | 1.3519 | 1.2539 |
| ∞ | 1.8307 | 1.7522 | 1.6664 | 1.5705 | 1.5173 | 1.4591 | 1.3940 | 1.3180 | 1.2214 | 1.0000 |

| $\alpha = .01$ | | | | | | | | | |
|----------------|--------|--------|--------|--------|--------|--------|--------|--------|----------|
| 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
| 6055.8 | 6106.3 | 6157.3 | 6208.7 | 6234.6 | 6260.7 | 6286.8 | 6313.0 | 6339.4 | 6366.0 |
| 99.399 | 99.416 | 99.432 | 99.449 | 99.458 | 99.466 | 99.474 | 99.483 | 99.491 | 99.501 |
| 27.229 | 27.052 | 26.872 | 26.690 | 26.598 | 26.505 | 26.411 | 26.316 | 26.221 | 26.125 |
| 14.546 | 14.374 | 14.198 | 14.020 | 13.929 | 13.838 | 13.745 | 13.652 | 13.558 | 13.463 |
| 10.051 | 9.8883 | 9.7222 | 9.5527 | 9.4665 | 9.3793 | 9.2912 | 9.2020 | 9.1118 | 9.0204 |
| 7.8741 | 7.7183 | 7.5590 | 7.3958 | 7.3127 | 7.2285 | 7.1432 | 7.0568 | 6.9690 | 6.8801 |
| 6.6201 | 6.4691 | 6.3143 | 6.1554 | 6.0743 | 5.9921 | 5.9084 | 5.8236 | 5.7372 | 5.6495 |
| 5.8143 | 5.6668 | 5.5151 | 5.3591 | 5.2793 | 5.1981 | 5.1156 | 5.0316 | 4.9460 | 4.8588 |
| 5.2565 | 5.1114 | 4.9621 | 4.8080 | 4.7290 | 4.6486 | 4.5667 | 4.4831 | 4.3978 | 4.3105 |
| 4.8492 | 4.7059 | 4.5582 | 4.4054 | 4.3269 | 4.2469 | 4.1653 | 4.0819 | 3.9965 | 3.9090 |
| 4.5393 | 4.3974 | 4.2509 | 4.0990 | 4.0209 | 3.9411 | 3.8596 | 3.7761 | 3.6904 | 3.6025 |
| 4.2961 | 4.1553 | 4.0096 | 3.8584 | 3.7805 | 3.7008 | 3.6192 | 3.5355 | 3.4494 | 3.3608 |
| 4.1003 | 3.9603 | 3.8154 | 3.6646 | 3.5868 | 3.5070 | 3.4253 | 3.3413 | 3.2548 | 3.1654 |
| 3.9394 | 3.8001 | 3.6557 | 3.5052 | 3.4274 | 3.3476 | 3.2656 | 3.1813 | 3.0942 | 3.0040 |
| 3.8049 | 3.6662 | 3.5222 | 3.3719 | 3.2940 | 3.2141 | 3.1319 | 3.0471 | 2.9595 | 2.8684 |
| 3.6909 | 3.5527 | 3.4089 | 3.2588 | 3.1808 | 3.1007 | 3.0182 | 2.9330 | 2.8447 | 2.7528 |
| 3.5931 | 3.4552 | 3.3117 | 3.1615 | 3.0835 | 3.0032 | 2.9205 | 2.8348 | 2.7459 | 2.6530 |
| 3.5082 | 3.3706 | 3.2273 | 3.0771 | 2.9990 | 2.9185 | 2.8354 | 2.7493 | 2.6597 | 2.5660 |
| 3.4338 | 3.2965 | 3.1533 | 3.0031 | 2.9249 | 2.8442 | 2.7608 | 2.6742 | 2.5839 | 2.4893 |
| 3.3682 | 3.2311 | 3.0880 | 2.9377 | 2.8594 | 2.7785 | 2.6947 | 2.6077 | 2.5168 | 2.4212 |
| 3.3098 | 3.1729 | 3.0299 | 2.8796 | 2.8011 | 2.7200 | 2.6359 | 2.5484 | 2.4568 | 2.3603 |
| 3.2576 | 3.1209 | 2.9780 | 2.8274 | 2.7488 | 2.6675 | 2.5831 | 2.4951 | 2.4029 | 2.3055 |
| 3.2106 | 3.0740 | 2.9311 | 2.7805 | 2.7017 | 2.6202 | 2.5355 | 2.4471 | 2.3542 | 2.2559 |
| 3.1681 | 3.0316 | 2.8887 | 2.7380 | 2.6591 | 2.5773 | 2.4923 | 2.4035 | 2.3099 | 2.2107 |
| 3.1294 | 2.9931 | 2.8502 | 2.6993 | 2.6203 | 2.5383 | 2.4530 | 2.3637 | 2.2695 | 2.1694 |
| 3.0941 | 2.9579 | 2.8150 | 2.6640 | 2.5848 | 2.5026 | 2.4170 | 2.3273 | 2.2325 | 2.1315 |
| 3.0618 | 2.9256 | 2.7827 | 2.6316 | 2.5522 | 2.4699 | 2.3840 | 2.2938 | 2.1984 | 2.0965 |
| 3.0320 | 2.8959 | 2.7530 | 2.6017 | 2.5223 | 2.4397 | 2.3535 | 2.2629 | 2.1670 | 2.0642 |
| 3.0045 | 2.8685 | 2.7256 | 2.5742 | 2.4946 | 2.4118 | 2.3253 | 2.2344 | 2.1378 | 2.0342 |
| 2.9791 | 2.8431 | 2.7002 | 2.5487 | 2.4689 | 2.3860 | 2.2992 | 2.2079 | 2.1107 | 2.0062 |
| 2.8005 | 2.6648 | 2.5216 | 2.3689 | 2.2880 | 2.2034 | 2.1142 | 2.0194 | 1.9172 | 1.8047 |
| 2.6318 | 2.4961 | 2.3523 | 2.1978 | 2.1154 | 2.0285 | 1.9360 | 1.8363 | 1.7263 | 1.6006 |
| 2.4721 | 2.3363 | 2.1915 | 2.0346 | 1.9500 | 1.8600 | 1.7628 | 1.6557 | 1.5330 | 1.3805 |
| 2.3209 | 2.1848 | 2.0385 | 1.8783 | 1.7908 | 1.6964 | 1.5923 | 1.4730 | 1.3246 | 1.0000 |