تحلیل های آماری در برنامه‌ریزی روستایی

(کارشناسی ارشد جغرافیا و برنامه‌ریزی روستایی)

دکتر مصطفی طالشی

دکتر پرویز نصیری

تابستان ۹۲

نام کتاب: تحلیل های آماری در برنامه‌ریزی روستایی (کارشناسی ارشد جغرافیا و برنامه‌ریزی روستایی)
مؤلفین:– دکتر مصطفی طالشی دکتر پرویز نصیری
ویراستار علمی:
چاپ اول:
شمارگان: نسخه
لیتوگرافی:
چاپخانه:
قیمت:
شابک:

فهرست مطالب

پیش گفتار

فصل اول

شناخت مفاهیم اساسی درآمار توصیفی

مقدمه

آمار علمی است که با جمع‌آوری، تنظیم، تجزیه و تحلیل و تفسیر مشاهدات سر و کار دارد. جمع‌آوری داده ها مرحله تهیه اندازه‌گیری یا شمارش‌ها بشمار می رود. نتایج قابل اطمینان فقط ازداده هایی که به طور صحیح جمع‌آوری شده‌اند، بدست می‌آیند. جمع‌آوری داده ها قسمت مهمی از روش آماری است که به طریق نمونه‌گیری از جامعه آماری اقتباس می‌شود. مجموعه روشها و قوانینی که نتایج را ساده‌تر کند و به تفسیر آسانتری از جمع‌آوری و طبقه‌بندی از داده‌ها یا مشاهدات منجر شود آمار توصیفی نامیده می‌شود. در بکارگیری از علم آمار این نکته حائز اهمیت است که هیچ روش آماری، به تنهایی از بروز اشتباهات، نادرستی‌ها، استدلال غلط یا نتایج غیر صحیح جلوگیری نمی‌کند. بلکه بایستی مشاهدات آماری به گونه ای انجام شود که ضریب خطای آماری را به حداقل ممکن کاهش یابد. از سوی دیگرروش‌های آماری نیز به طور صحیح به کار برده شود و نتایج را باید کسی تجزیه و تحلیل و تفسیر نماید که نه تنها با روشهای آماری بلکه با رشته‌ای که آمار در مورد آن به کار رفته است نیز آشنا باشد. در این فصل مفاهیم اساسی، روش‌های نمونه‌گیری، شاخص‌های مرکزی و غیر مرکزی، جدول توزیع فراوانی و نمودارهای آماری را مورد توجه قرار می گیرد.

۱-۱ مفاهیم اساسی

در این بخش برای آشنایی بیشتر با ادبیات آماری اصطلاحات اساسی آمار را که همواره با آنها سر و کار خواهیم داشت مورد بازشناسی قرار می گیرد.

۱-۱-۱ جامعه آماری

جامعه آماری مجموعه یا کل اندازه‌گیری‌ها از تمام اشیائی است که حداقل دارای یک صفت مشخص است. از نظر آمار یک جامعه آماری هرگز مجموعه‌ای از اشیا نیست، بلکه همیشه مجموعه‌ای از اندازه‌گیری‌ها یا شمارش‌ها می‌باشد. یک جامعه آماری ممکن است از تعداد متناهی یا نامتناهی داده تشکیل شده باشد. در حالتی که جامعه آماری متناهی و دارای N عضو باشد اعضاء آن را به صورت زیر نمایش می‌دهند.

$X_1, X_2, \ldots, X_N$

که X_1 عضو اول، X_2 عضو دوم ... و X_N عضو Nام جامعه است و هیچ ترتیبی بین آنها نیست.

مثال فرض کنید انجمن برنامه‌ریزی روستایی کشور دارای ۷ گروه عضو باشد. در این مثال $N = 7$ و اعضای آن را می‌توان جداگانه به صورت زیر با نماد $X_۱$ تا $X_۷$ نمایش داد:

$X_1 \quad X_2 \quad X_3 \quad X_4 \quad X_5 \quad X_6 \quad X_7$

۱-۱-۲ نمونه

نمونه مجموعه‌ای از اندازه‌گیری‌هاست که قسمتی یا تمامی یک جامعه را در بر می‌گیرد. اگر نمونه انتخابی قابل شمارش یا برچسب‌پذیر باشد و دارای $n \left(n \le N\right)$ عضو باشد، اعضای نمونه را به صورت زیر نمایش می‌دهند:

$X_1, X_2, \ldots, X_n$

مثال اگر سه گروه از انجمن برنامه‌ریزی روستایی برای مشارکت در یک طرح پژوهشی مشترک مشارکت داشته باشند. این گروه‌ها آن را می‌توان به صورت زیر نوشت:

$n = 3 \quad، \quad X_1 \quad X_2 \quad X_3$

۱-۱-۳ داده‌های آماری

داده‌های آماری به دو گروه، داده‌های دست اول یا خام و داده‌های دسته دوم تقسیم‌بندی می‌شود. داده‌های دست اول آن دسته از داده‌هایی هستند که روی آنها هیچگونه سازماندهی یا پردازش صورت نگرفته است. برای نمونه جمع‌آوری اطلاعات از اطلاعات پرسشنامه‌ای توسط مرکز آمار ایران در سرشماری‌ها از هزینه زندگی خانوارهای روستایی، داده‌های دست اول یا خام است و گزارشهای چاپ شده توسط مرکز آماری ایران به صورت جداول یا نمودارها از نوع داده‌های دست دوم می‌باشد.

۱-۱-۴ متغیر

متغیر علامتی است برای معرفی اعضای جامعه آماری و به ویژگی اطلاق می‌شود و تغییرات را از فرد به فردی یا از شئ به شئ دیگر نشان می‌دهد. به عنوان مثال میزان تولید یک محصول در واحد سطح و یا میزان مهاجر فرستی سکونتگاههای روستایی هر یک بعنوان متغیر تلقی می‌شود، به دلیل تغییرات این متغیر همواره بصورت کمّی یا کیفی قابل بررسی است. متغیر کمّی به متغیری گفته می‌شود که قابل اندازه‌گیری باشد و یا برای اندازه‌گیری آن مقیاسی وجود داشته باشد.

متغیر کمّی ممکن است گسسته یا پیوسته باشد. متغیر کمّی گسسته آنهایی هستند که قابل شمارش هستند. مانند میزان جمعیت روستایی، تعداد طرح‌هایی اجرایی روستایی، تعداد جمعیت بیکار، تعداد روزهای یخبندان.

متغیر کمّی پیوسته آنهایی هستند که بین هر دو مقدار دلخواه بتوانند مقداری انتخاب کنند. مانند دبی رودخانه ، سرعت باد.
اگر متغیری با x نشان داده شود، مشاهدات متوالی این متغیر به صورت:

$$x_1, x_2, \ldots, x_i, \ldots, x_n$$

نشان داده می‌شود که x_i مشابه iام و x_n عضو یا مشاهد nام و n حجم مشاهدات است. برای مثال اگر x میزان درآمد یک خانوار روستایی در یک سکونتگاه روستایی باشد و جمعیت این سکونتگاه ۲۵ خانواری به صورت زیر نمایش داده می‌شود:

$$x_1, x_2, \ldots, x_n$$

که x_1 درآمد، خانواری که میزان آن اول برآورد شد و ...، x_n درآمد خانواری است که برآوردآن آخر استحصال شده است.

برای نشان دادن جمع این n مشاهده از نماد حرف یونانی Σ استفاده می‌شود.

$$\sum_{i=1}^{n} X_i = X_1 + X_2 + \ldots + X_n$$

که اندیس i شمارنده و معمولاً عدد صحیح در نظر گرفته می‌شود.

قبل از بررسی، شاخص‌های گرایش مرکزی و پراکندگی، روشهای نمونه‌گیری که سهم ویژه ایی در تحلیل این بخش از تحلیل آماری دارد ، مورد بررسی قرار می گیرد. مهمترین روشهای نمونه‌گیری عبارتند از:

۱- نمونه‌گیری تصادفی ساده[۱]

۲- نمونه‌گیری منظم یا سیستماتیک[۲]

۳- نمونه‌گیری تصادفی یا طبقه بندی[۳]

۴- نمونه‌گیری تصادفی خوشه‌ای[۴]

---

۱. Simple Random Sampling
۲. Systematic Sampling
۳. Stratified Random Sampling
۴. Cluster Random Sampling

۱-۲ نمونه‌گیری تصادفی ساده

نمونه‌گیری تصادفی ساده یکی از روشهای متداول نمونه‌گیری است. در این روش فرض می‌شود که هر عضوی از جامعه شانس مساوی برای انتخاب شدن داشته باشد. نمونه‌گیری تصادفی ساده ممکن است با جایگذاری یا بدون جایگذاری باشد.

۱-۳ نمونه‌گیری با جایگذاری و بدون جایگذاری

در نمونه‌گیری با جایگذاری، هر مشاهده iام از جامعه پس از رؤیت یا ثبت به جامعه برگردانده می‌شود. در این روش ممکن است مشاهدات تکراری داشته باشیم. در نمونه‌گیری بدون جایگذاری پس از رؤیت یا ثبت نمونه iام، عضو مشاهده شده از جامعه به جامعه برگردانده نمی‌شود. در نمونه‌گیری با جایگذاری، شانس انتخاب هر عضو ثابت است ولی در نمونه‌گیری بدون جایگذاری ثابت نیست.

۱-۴ انتخاب نمونه تصادفی ساده

یکی از راههای انتخاب نمونه تصادفی ساده استفاده از جدول اعداد تصادفی یا استفاده از ماشین حساب و یا برنامه‌های آماده رایانه‌ای است. در اینجا طریقه استفاده از جدول اعداد تصادفی (جدول شماره ۱) شرح داده می‌شود.

فرض کنید جامعه آماری مورد بررسی شامل ۱۵ خانوار روستایی است که اطلاعات درآمد سرانه آن از طریق پژوهش میدانی جمع آوری شده است. میزان درآمد سرانه نیز بصورت هزار تومان بر حسب سرانه به صورت زیرمی باشد:

شماره خانوار :	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵
درآمد سرانه :	۱۴۳	۱۵۱	۱۵۲	۱۵۸	۱۵۱	۱۶۰	۱۵۸	۱۵۰	۱۴۳	۱۵۲	۱۵۲	۱۵۸	۱۶۱	۱۵۸	۱۵۰

می‌خواهیم نمونه‌ای تصادفی به حجم ۵ با استفاده از جدول اعداد تصادفی انتخاب کنیم. ابتدا لازم است در نقطه شروع تصادفی قرارگیریم. این مهم را به طرق

زیادی می‌توان انجام داد. یکی از راهها این است که با نوک مداد بر نقطه‌ای تصادفی در صفحه اعداد تصادفی اثر گذاریم. نقطه تصادفی نزدیکترین رقم به محلی است که مداد بر صفحه اثر می‌گذارد. فرض کنید نقطه شروع تصادفی در تقاطع سطر ششم و ستون سوم باشد، رقم مزبور در این نقطه صفر است. چون نمونه خود را باید از میان ۱۵ خانوار انتخاب کنیم، تنها می‌توان اعداد ۱ تا ۱۵ را از جدول اعداد تصادفی استخراج کرد.

بنابراین باید اعداد دورقمی را چنان انتخاب کرد که در محدوده ۱ تا ۱۵ قرارگیرند. نخستین عدد دورقمی که در نقطه شروع تصادفی اعداد جدول انتخاب می‌شود، عدد ۰۲ یا ۲ است. اعداد بعدی به ترتیب عبارتند از:

۰۲ ۹۴ ۳۹ ۰۲ ۷۷ ۵۵ ۷۳ ۲۲ ۷۰ ۹۷ ۷۹ ۰۱ ۷۱
۱۹ ۵۲ ۵۲ ۷۵ ۸۰ ۲۱ ۸۰ ۸۱ ۴۵ ۱۷ ۴۸ ۵۴ ۱۷
۸۴ ۵۶ ۱۱ ۸۰ ۹۹ ۳۳ ۷۱ ۴۳ ۰۵ ۳۳ ۵۱ ۲۹ ۶۹
۵۶ ۱۲ ۷۱

پس از انتخاب اولین عدد تصادفی، بقیه اعداد با حرکت از نقطه شروع به سمت راست انتخاب می‌شوند. می‌توان جهت حرکت را هر طور که از قبل تعیین شده رو به بالا یا به پایین ادامه داد. نمونه انتخابی متناظر با اعداد فوق عبارتند از ۲ ، ۱ ، ۱۱ ، ۵ ، ۱۲

عدد تصادفی انتخاب شده	۲	۱	۱۱	۵	۱۲
مقدار نمونه	۱۵۱	۱۴۳	۱۵۲	۱۵۱	۱۵۸

یادآوری می‌شود که ۰۲ یا عدد ۲ دو بار تکرار شده که یکی از آنها را حذف کردیم. بنابراین نمونه انتخابی برابر است با:

۱۵۱ ۱۴۳ ۱۵۲ ۱۵۱ ۱۵۸

۱-۵ نمونه‌گیری منظم یا سیستماتیک

در این روش تمام اعضای جامعه بدون هرگونه نظم و ترتیبی فهرست‌بندی می‌شوند، سپس نمونه مورد نظر با استفاده از یک نظم معین از لیست جامعه انتخاب می‌شود. اگر اعضای جامعه در لیست یا فهرست‌بندی به صورت زیر باشد:

$$X_۱, X_۲, \ldots, X_N$$

و بخواهیم یک نمونه به حجم n انتخاب کنیم، ابتدا عدد K یا تناوب را از رابطه زیر حساب می‌کنیم:

$$K = \left[\frac{N}{n}\right]$$

پس از محاسبه K، جامعه را به K قسمت تقسیم می‌کنیم و از هر دسته یک نمونه انتخاب می‌کنیم تا نمونه‌ای به حجم n به دست آید. نقطه شروع به تصادف از بین ۱ تا K انتخاب می‌شود، عضو جامعه متناظر در لیست عضو اول جامعه خواهد بود. برای مثال، فرض کنید پژوهشگری قصد دارد نگرش روستانشینان به حجم $N = 1000$ نفر را نسبت به فعالیت های اجرایی جهاد کشاورزی را مورد پژوهش قرار دهد. برای انتخاب نمونه‌ای به حجم $n = 50$، فهرستی از کلیه روستانشینان را تهیه می‌کند. پس از هر $\frac{N}{n} = \frac{1000}{50} = 20$ نفر یک نفر روستایی را به صورت تصادفی انتخاب می‌کند. برای جلوگیری از هر نوع سوگیری، یک عدد به تصادف از ۱ تا ۲۰ انتخاب می‌کنیم و متناظر با آن و اضافه کردن ۲۰ به آن نمونه‌های متناظر را از جامعه یا لیست انتخاب می‌کنیم. اگر عدد تصادفی انتخاب شده ۰۷ باشد، اعداد بعدی عبارت است از:

۰۷ ۲۷ ۴۷ ۶۷ ۸۷ ۱۰۷ ۱۲۷ ۱۴۷ ...

اعضای متناظر با اعداد فوق یعنی X_{07}، X_{27}، X_{47}، X_{67}، X_{87}، X_{107}، X_{127} X_{147} و ... نمونه‌های مورد بررسی خواهند بود.

هنگام طرح‌ریزی برای انتخاب یک نمونه از طریق یک فهرست یا امثال آن، پژوهشگر باید با دقت فهرست را بررسی کرده و مطمئن شود که هیچ الگوی چرخشی در آن وجود ندارد. به عبارت دیگر، در صورتی که دارای یک ترتیب تناوبی باشد، نمونه‌گیری منظم، عملکرد و نتیجه مناسبی نخواهد داشت.

مثال فرض کنید بخواهیم از ۲۰ سکونتگاه روستایی دهستان الف زیر ۵ سکونتگاه را به تصادف به روش سیستماتیک انتخاب کنیم.

۱- محمد آباد	۶- حمیدآباد	۱۱- شاه تقی	۱۶- فرهادگرد
۲- تقی آباد	۷- داوودی	۱۲- طاهر آباد	۱۷- قاسم آباد
۳- اسماعیل آباد	۸- کلاته علی آباد	۱۳- خلیل آباد	۱۸- کریم آباد
۴- خرو سفلی	۹- کلاته شاهرخ	۱۴- علی آباد	۱۹- نورآباد
۵- خرو علیا	۱۰- صابری	۱۵- منصوریه	۲۰- یوسف آباد

در این مثال $N = 20$، $n = 5$ در نتیجه $K = \frac{20}{5} = 4$، برای شروع انتخاب یک عدد بین ۱ و ۴ به تصادف انتخاب می‌شود. فرض کنید عدد انتخاب شده ۳ باشد. بنابراین اولین عضو نمونه شخصی است که در چارچوب شماره آن ۳ است که در این مثال اسماعیل آباد ، اولین روستایی است که انتخاب شده و می‌تواند اطلاعات جمعیتی، اشتغال و غیره را که مدنظر پژوهشگران است از این سکونتگاه پرسشگری شود، برای انتخاب سکونتگاه بعدی کافی است به شماره ۳ متوالیاً عدد ۴ اضافه کنیم تا شماره سکونتگاه بدست آید.

۳ ۷ ۱۱ ۱۵ ۱۹

با توجه به شماره، سکونتگاهها انتخاب شده عبارتند از: اسماعیل آباد، داوودی،شاه تقی، منصوریه و نورآباد.

۱-۶ نمونه‌گیری تصادفی با طبقه‌بندی[۱]

در این روش، جامعه آماری از لحاظ تشابه یا از روی یک یا چند صفت متغیر طبقه‌بندی و نمونه‌های مستقل از هر طبقه انتخاب می‌شود. در این روش، جامعه آماری به حجم N، به h طبقه یا زیرجامعه آماری همگن به حجمهای $N_۱, N_۲, \ldots, N_h$ افراز می‌شود، و از هر طبقه نمونه‌ای متناسب با حجم زیرجامعه انتخاب می‌شود.

اگر $n_۱, n_۲, \ldots, n_K$ به ترتیب نمونه‌های انتخابی از زیرجامعه‌های ۱ تا h باشد داریم:

حجم جامعه: $N_۱ + N_۳ + \ldots + N_h = N$

حجم نمونه: $n_۱ + n_۳ + \ldots + n_h = n$

انتخاب نمونه از طبقات به روشهای نمونه‌گیری تصادفی ساده یا سیستماتیک امکان‌پذیر می‌باشد.

مثال فرض کنید در بخش جوین از شهرستان سبزوار در استان خراسان رضویه سه دهستان به نام های پیراکوه ، بالا جوین و پایین جوین ۳۰ سکونتگاه روستایی وجود دارد و میزان تولید گندم آنها با توجه به موقعیت جغرافیایی به صورت زیر گزارش شده باشد. (ارقام به ده میلیون تن)

دهستان پایین جوین	۵	۶	۴	۷/۵	۸/۵	۶/۷۵	۹	۷	۸	۵/۵
دهستان بالا جوین	۴	۳	۴/۵	۳/۵	۵	۴/۲۵	۳/۷۵	۴/۲۵	۳/۲۵	۶
دهستان پیراکوه	۲/۷۵	۱/۵	۳/۵	۱	۲/۵	۳/۲۵	۳/۷۵	۳	۲	۱/۷۵

۱. Stratified Random Sampling

فرض کنید بخواهیم از این جامعه آماری یک نمونه ۹ تایی انتخاب کنیم، چون توزیع تولید در موقعیت‌های جغرافیایی یکسان نیست لذا، از هر طبقه یا دهستان یک نمونه ۳ تایی انتخاب می‌کنیم

$$N = 30 , h = 3 , n_1 = 3 , n_2 = 3 , n_3 = 3$$

برای انتخاب نمونه از داخل موقعیت‌ها می‌توان از روش نمونه‌گیری تصادفی ساده یا سیستماتیک استفاده کرد.

دهستان پایین جوین	۶	۷/۵	۵/۵
دهستان بالا جوین	۴	۳/۵	۴/۲۵
دهستان پیراکوه	۱/۵	۳/۷۵	۲/۷۵

۱-۷ نمونه‌گیری تصادفی خوشه‌ای[۱]

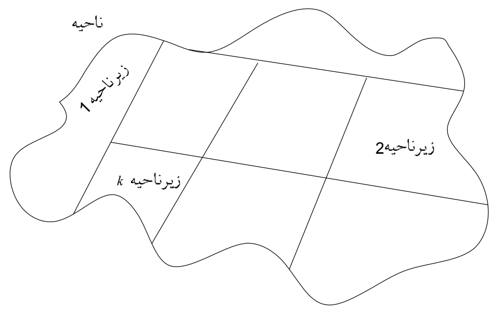
در روشهای نمونه‌گیری تصادفی ساده، سیستماتیک و طبقه‌ای، نمونه از جامعه آماری مورد بررسی انتخاب می‌شود، اما گاهی اوقات انتخاب نمونه به صورت مستقیم از اعضای جامعه ممکن نیست یا به عبارت دیگر اطلاعات کافی از جامعه آماری در دست نیست. برای مثال اگر پژوهشگر بخواهد متوسط هزینه پروژه‌های کوچک و بزرگ کشور را به دست آورد یا برآورد کند، چون فهرست کاملی در دسترس نیست، نمی‌توان بر اساس روشهای تصادفی ذکر شده، نمونه مناسبی انتخاب کرد. در چنین شرایطی از نمونه‌گیری تصادفی خوشه‌ای استفاده می‌شود. برای استفاده از این روش، جامعه آماری را به چند خوشه (مثلاً K) تقسیم می‌کنند و از بین این خوشه‌ها، با توجه به وقت بررسی و هزینه نمونه‌گیری و عوامل دیگر، چند خوشه به تصادف انتخاب می‌کنند و خوشه‌های انتخابی را سرشماری می‌کنند و نتایج به دست آمده را به جامعه آماری تعمیم می‌دهند.

---

۱. Cluster Random Sampling

مثال فرض کنید در کلان منطقه زاگرس در غرب کشور که شامل چند استان است میزان مهاجر فرستی مورد برآورد قرار گیرد.. روش برآورد یا جمع‌آوری اطلاعات را شرح دهید.

– چون ناحیه مورد بررسی فاقد اطلاعات قبلی است، گروه پژوهشی ناحیه مورد بررسی را به چند زیرناحیه (شکل) افراز می‌کنند. از بین زیرناحیه‌ها، چند ناحیه انتخاب و آنها را مورد بررسی کلی قرار می‌دهد.



شکل ۱-۱

۱-۸ شاخص‌های گرایش مرکز

در بررسی داده‌های آماری یا مشاهدات، گاهی این پرسش مطرح است که آیا می‌توان، نتیجه بررسی را در قالب یک عدد ارائه داد. برای مثال ممکن است تعدادی از سکونتگاههای روستایی مطالعات میزان درآمد سرانه بخش کشاورزی صورت گرفته است.در این ارتباط اگر هدف تحلیل متوسط درامد سرانه باشد که نیاز است با یک عدد بیان شود.

در این بخش مهمترین شاخص‌های گرایش مرکزی که عبارتند از انواع میانگین‌ها، نما، میانه و چارکها، مورد بحث قرار خواهد گرفت.

۱-۸-۱ میانگین حسابی

یکی از رایج‌ترین و مفیدترین معیار گرایش‌های مرکزی میانگین حسابی داده‌هاست. اگر مشاهدات را با $x_1, x_2, \ldots, x_n$ علامتگذاری کنیم میانگین حسابی از رابطه زیر بدست می‌آید.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

که $\bar{x}$، ایکس بار خوانده می‌شود.

مثال درمطالعات وضع موجود طرح هادی روستایی محمد آباد عمر ساختمانهایی روستایی که توسط مدیرطرح برآورد شده است. عبارت است از (بر حسب سال)

۱۸ ۱۳ ۱۷ ۱۲ ۱۰

عمر متوسط نمونه فوق را حساب کنید.

$$n = 5,\quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{5}\sum_{i=1}^{5} X_i = \frac{1}{5}[10+12+17+13+18] = 14$$ سال

ویژگی‌های میانگین حسابی[1]

الف- متغیر است، چون از یک نمونه به نمونه دیگر تغییر می‌کند.

ب- اگر به تمام مشاهدات مقدار ثابت اضافه (کم) کنیم به همان مقدار به میانگین حسابی اضافه (کم) می‌شود.

ج- اگر تمام مشاهدات را در مقدار ثابت غیرصفر ضرب (تقسیم) کنیم، همان مقدار در میانگین ضرب (تقسیم) می‌شود.

مثال اگر میانگین مشاهدات $x_1, x_2, \ldots, x_n$ برابر با ۴ باشد، میانگین مشاهدات $2x_1 + 7$، $2x_2 + 7$، …، $2x_n + 7$ برابر است با:

---

۱. Arthmatic Mean

$$2(\bar{x})+7=2(4)+7=15$$

۱-۸-۲ میانگین هندسی[1]

اگر $x_1, x_2, \ldots, x_n$ مشاهداتی به حجم n باشند میانگین هندسی از رابطه زیر بدست می‌آید و آن را با G نمایش می‌دهند.

$$G=\sqrt[n]{x_1 x_2 \ldots x_n} \quad ، \quad x_i \geq 0 \quad ، \quad i=1,2,\ldots,n$$

مثال اگر زمان صرف شده برای سه طرح هادی روستایی بر حسب سال به صورت زیر باشد، مقدار متوسط هندسی آن را بدست آورید.

شماره طرح هادی	۱	۲	۳
مدت زمان	۵	۴	۲

$$x_1=1 \quad ، \quad x_2=3 \quad ، \quad x_3=9 \quad ، \quad n=3$$
$$G=\sqrt[n]{x_1 x_2 \ldots x_n}=\sqrt[3]{x_1 \cdot x_2 \cdot x_3}=\sqrt[3]{1\times 3\times 9}=\sqrt[3]{27}=3$$

۱-۸-۳ میانگین هارمونیک[2]

اگر $x_1, x_2, \ldots, x_n$ مشاهداتی به حجم n باشند میانگین هارمونیک از رابطه زیر بدست می‌آید و آن را با H نشان می‌دهند.

$$H=\frac{n}{\sum \frac{1}{X_i}}$$

مثال مدیر طرح هادی برای انجام عملیات عمرانی در کاربریهای متفاوت روستا چهار زمان لازم برای انجام پروژه پیشنهاد می‌کند

۱. Geometric Mean
۲. Harmonic Mean

اجرایی کاربری‌های پیشنهادی	۱	۲	۳	۴
زمان لازم (برحسب ماه)	۵	۴	۲	۱

متوسط زمان برابر است با:

$$H=\frac{n}{\sum\frac{1}{X_i}}=\frac{4}{\sum_{i=1}^{4}\frac{1}{x_i}}$$

i	۱	۲	۳	۴
x_i	۵	۴	۲	۱
$\frac{1}{x_i}$	$\frac{1}{5}$	$\frac{1}{4}$	$\frac{1}{2}$	۱

$$H=\frac{4}{\left(\frac{1}{5}+\frac{1}{4}+\frac{1}{2}+1\right)}=\frac{4}{(1/95)}=۲/۰۵۱$$

ذکر این نکته ضروری است که بین میانگین حسابی، هندسی و هارمونیک رابطه مهم زیر همواره برقرار است:

$$H\leq G\leq\bar{x}$$

مثال اگر چهار مشاهده به صورت زیر باشند نشان دهید که $H\leq G\leq\bar{x}$.

۰/۲ ۰/۵ ۰/۱ ۰/۲۵

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=\frac{1}{4}(0/2+0/5+0/1+0/25)=\frac{1/05}{4}=0/2625$$

$$G=\sqrt[4]{x_1\cdot x_2\cdot x_3\cdot x_4}=\sqrt[4]{(0/2)(0/5)(0/1)(0/25)}=0/2236$$

$$H=\frac{n}{\sum\frac{1}{X_i}}=\frac{4}{\left(\frac{1}{0/2}+\frac{1}{0/5}+\frac{1}{0/1}+\frac{1}{0/25}\right)}$$

$$=\frac{4}{\frac{10}{2}+\frac{10}{5}+\frac{10}{1}+\frac{100}{25}}=\frac{4}{5+2+10+4}=\frac{4}{19}=0/2105$$

$$H = 0/2105 \leq G = 0/2236 \leq \bar{x} = 0/2625$$

۱-۸-۴ نما

نما مجموعه‌ای از مشاهدات $x_n,\ldots,x_2,x_1$ عدد یا کمیتی است که بیشتر از دیگر مشاهدات تکرار شده باشد. مشاهدات مورد بررسی ممکن است فاقد نما، یک نما یا چند نما باشند.

مثال سری داده‌های زیر به ترتیب فاقد نما، یک نما و دو نما می‌باشند.

الف- مدت زمان اجرا ۵ پروژه طرح هادی عبارتند از (بر حسب ماه)

۷ ۲ ۴ ۱ ۳

چون همه به یک اندازه تکرار شده‌اند، فاقد نما است

ب- بودجه اختصاص یافته به ۷ پروژه عبارتند از (بر حسب میلیون تومان)

۹۰۰ ۵۰۰ ۵۰ ۶۰۰ ۱۰۰ ۲۰۰ ۱۰۰

این مشاهدات دارای یک نما است که برابر با ۱۰۰ هزار تومان می‌باشد.

ج- تعداد کارکنان اجرایی یک طرح هادی روستایی به صورت زیر گزارش شده‌اند (بر حسب نفر)

۲۵ ۱۹ ۱۱ ۱۶ ۱۵ ۲۰ ۱۱ ۱۵ ۹

که ۱۱ و ۱۵ نما مشاهدات فوق می‌باشند.

۱-۸-۵ میانه

میانه مشاهدات $x_n,\ldots,x_2,x_1$ عدد یا کمیتی است که اطلاعات یا داده را به دو قسمت مساوی تقسیم می‌کند به طوری که ۵۰٪ از اطلاعات کمتر یا مساوی است و ۵۰٪ از اطلاعات بیشتر یا مساوی آن است. برای محاسبه میانه مشاهدات $x_n,\ldots,x_2,x_1$ یکی از دو حالت زیر را انجام می‌دهیم.

الف- اگر تعداد مشاهدات یا حجم نمونه n، فرد باشد. پس از مرتب کردن مشاهدات، عدد میانی میانه مشاهدات خواهد بود.

ب- اگر تعدا مشاهدات یا حجم نمونه زوج باشد، متوسط دو عدد میانی پس از مرتب کردن میانه داده‌ها خواهد بود.

۱۱ ۹ ۵ ۲ ۴ ۱- ۰ ۳ ۷ ۱

ابتدا داده‌ها را مرتب می‌کنیم

۱۱ ۹ ۷ ۵ ۴ ۳ ۲ ۱ ۰ ۱-

دو عدد میانی برابرند با: ۴ و ۳

$$\text{متوسط} = \frac{3+4}{2} = \frac{7}{2} = 3/5$$

مثال برای داده‌های زیر میانه را بدست آورید.

۱۵ ۹ ۷ ۶ ۴ ۲- ۰ ۱- ۱

ابتدا داه‌ها را مرتب می‌کنیم.

۱۵ ۹ ۷ ۶ ۴ ۱ ۰ ۱- ۲-

چون عدد میانی ۴ است، پس میانه برابر با ۴ است.

مثال سن مدیران پروژه ساختمانی شرکتی به صورت زیر گزارش شده است. میانه سن مدیران را بدست آورید.

شماره پرسنلی	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
سن (سال)	۲۷	۳۳	۵۰	۴۵	۴۰	۲۵	۳۵	۳۰	۵۵	۴۹

داده‌ها را مرتب می‌کنیم

۵۵ ۵۰ ۴۹ ۴۵ ۴۰ ۳۵ ۳۳ ۳۰ ۲۷ ۲۵

$$\text{میانه}=\frac{35+40}{2}=\frac{75}{2}=37/5$$

۱-۸-۶ چارک‌ها

چارک‌ها مشاهدات $x_1, x_2, \ldots, x_n$ اعداد یا کمیتهایی هستند که مشاهدات یا اطلاعات را به چهار قسمت مساوی تقسیم می‌کنند و به ترتیب آنها را چارک اول، دوم و سوم می‌نامند و به ترتیب با Q_1، Q_2 و Q_3 نمایش داده می‌شوند.

Q_1 عدد کمیتی است که ۲۵٪ از اطلاعات کمتر یا مساوی آن است.

Q_2 عدد یا کمیتی است که ۵۰٪ از اطلاعات کمتر یا مساوی آن است و آنرا میانه نیز می‌گویند.

Q_3 عدد یا کمیتی است که ۷۵٪ از اطلاعات کمتر یا مساوی آن است (یا ۲۵٪ اطلاعات بیشتر یا مساوی آن است).

چارک اول	$Q_1=25$ امین درصد
چارک دوم	$Q_2=50$ امین درصد
چارک سوم	$Q_3=75$ امین درصد

برای پیدا کردن موقعیت چارک‌ها از فرمول زیر استفاده می‌کنیم:

$$\text{موقعیت چارک } h\text{-ام}=n\times\frac{h}{۴}$$

مثال زمان اتمام مراحل تولید ۴۰ مزرعه کشاورزی در یک سکونتگاه روستایی به صورت زیر گزارش شده است (برحسب ماه)

۱۴	۱۳	۱۱	۱۰	۹	۸	۷	۴	۳	۳	۲	۱
۲۷	۲۵	۲۲	۲۱	۲۰/۵	۲۰	۱۹	۱۸/۵	۱۸	۱۷	۱۶	۱۵
۴۳	۴۱	۴۰	۳۹	۳۶	۳۵	۳۲	۳۱	۳۰	۲۹	۲۸	۲۷/۵
								۵۰	۴۸	۴۷	۴۵

تعداد مشاهدات ۴۰ است. چارک اول مقداری است که حداقل $40 \times \frac{1}{4} = 10$ مشاهده کوچکتر یا مساوی خودش دارد. (۱۰–ام مشاهده + ۱۱–ام مشاهده تقسیم بر۲)

$$Q_1 = \frac{11 + 13}{2} = 12$$

و چارک دوم مقداری است که حداقل $40 \times \frac{1}{2} = 20$ مشاهده کوچکتر یا مساوی با خودش دارد. (۲۰–ام مشاهده+ ۲۱–ام مشاهده تقسیم بر ۲)

$$Q_2 = \frac{20/5 + 21}{2} = 20/75$$

برای چارک سوم: $40 \times \frac{3}{4} = 30$ (۳۰–ام مشاهده+ ۳۱–ام مشاهده تقسیم بر ۲)

$$Q_3 = \frac{32 + 35}{2} = 33/5$$

۱-۹ شاخص‌های پراکندگی

گرچه شاخص‌های مرکزی یک معیار ایده‌آل تمایل به مرکز هستند، به تنهایی تصویری روشن از یک سری از مشاهدات یا توزیع آنها نمی‌دهند. برای اینکه ممکن است مشاهدات متفاوتی دارای میانگین‌های مشابه باشند. در این بخش چند معیار پراکندگی شامل دامنه، واریانس، انحراف معیار و ضریب تغییر را مورد بحث قرار می‌دهیم.

۱-۹-۱ دامنه

دامنه مشاهدات $x_n, \ldots, x_2, x_1$ از تفاوت بین کوچکترین و بزرگترین مشاهدات بدست می‌آید و اطلاعاتی در مورد پراکندگی مشاهدات حول میانگین مشاهدات نشان می‌دهد ولی از آنجا که به دو مقدار انتهایی بستگی دارد اطلاعات زیادی به دست نمی‌دهد.

مثال در یک بررسی آماری از ۱۵ خانوار روستایی درباره درآمد سرانه بخش کشاورزی، اطلاعات زیر برای ارزیابی بدست آمده است. دامنه را بدست آورید.

۱/۲ ۱/۵ ۰/۷ ۳ ۱/۷ ۱۰ ۹ ۴ ۸/۵ ۱/۷۵ ۸/۲۵ ۳/۵

۴/۹ ۵/۷۵ ۷/۷۵

دامنه = بزرگترین − کوچکترین = $9-0/7=8/3$

۱-۹-۲ واریانس نمونه

واریانس مشاهدات $x_1, x_2, \ldots, x_n$ با میانگین $\bar{x}$ از رابطه زیر بدست می‌آید و آن را با s^2 نشان می‌دهند.

مثال مدت زمان آماده سازی چهار قطعه زمین زراعی یک کشاورز برحسب ماه برابر است با:

۱ ۳ ۷ ۹

میانگین و واریانس را بدست آورید.

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i=\frac{1}{4}(1+3+7+9)=\frac{20}{4}=5$$

$$s^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})=\frac{1}{n}\sum_{i=1}^{4}(x_i-5)=\frac{1}{4}\left[(1-5)^2+(3-5)^2+(7-5)^2+(9-5)^2\right]$$

$$=\frac{1}{4}\left[(-4)^2+(-2)^2+(2)^2+(4)^2\right]=\frac{40}{4}=10$$

تبصره واریانس را می‌توان از فرمول زیر که معادل s^2 است بدست آورد.

$$s^2=\frac{1}{n}\sum_{i=1}^{n}x_i^2-\bar{x}^2$$

برای مثال اخیر

$$\sum x_i^2 = (1)^2 + (3)^2 + (7)^2 + (9)^2 = 140$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1}{4}(140) - 25 = 35 - 25 = 10$$

ویژگیهای واریانس نمونه

۱- واریانس مشاهدات $x_1, x_2, \dots, x_n$ همواره بزرگتر یا برابر صفر است

۲- اگر مقدار ثابت a را به همه مشاهدات اضافه (کم) کنیم واریانس تغییر نمی‌کند.

۳- اگر همه مشاهدات را در مقدار ثابت $K(K > ۰)$ ضرب یا تقسیم کنیم واریانس در مقدار K^2 ضرب یا تقسیم می‌شود

۳-۹-۱ انحراف معیار

واریانس یکی از پارامترهای پراکندگی است که در تجزیه و تحلیل‌های آماری زیاد به کار می‌رود، که واحد آن مجذور واحد اصلی بیان می‌شود. انحراف معیار چنین معیاری از پراکندگی است که با جذر گرفتن از واریانس به دست می‌آید بنابراین، انحراف معیار مشاهدات $x_1, x_2, \dots, x_n$ برابر است با:

$$S = \sqrt{\frac{1}{n}\sum (x_i - \bar{x})^2}$$

انحراف معیار دارای خصوصیات بسیار جالب توجهی است که در آینده اشاره خواهد شد.

مثال هزینه‌های تولید ۵ واحد صنعتی روستایی در شهرک صنعتی روستایی در شهرستان قم به صورت زیر گزارش شده است (ارقام به میلیون تومان می‌باشد)

۳ ۲ ۱- ۰ ۱

میانگین، واریانس و انحراف معیار را بدست آورید.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{5}(1 + ۰ + -1 + 2 + 3) = 1$$ (یک میلیون تومان)

به طور متوسط یک میلیون مانده حساب دارند.

$$s^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\bar{x}\right)=\frac{1}{5}\sum_{i=1}^{5}\left(x_i-1\right)^2$$
$$=\frac{1}{5}\left[(1-1)^2+(\circ-1)^2+(-1-1)^2+(2-1)^2+(3-1)^2\right]$$
$$=\frac{1}{5}\left[\circ+1+4+1+4\right]=\frac{10}{5}=2$$

واریانس برابر با ۲ میلیون تومان مربع

$$S=\sqrt{\frac{1}{n}\sum\left(x_i-\bar{x}\right)^2}=\sqrt{2}=1/4$$ میلیون تومان

۱-۹-۴ ضریب تغییر

ضریب تغییر مشاهدات $x_1,x_2,\ldots,x_n$ از رابطه زیر بدست می‌آید،

$$CV=\frac{S}{\bar{x}}=\frac{\sqrt{\frac{1}{n}\sum\left(x_i-\bar{x}\right)^2}}{\frac{1}{n}\sum_{i=1}^{n}x_i}=\frac{\sqrt{n\sum\left(x_i-\bar{x}\right)^2}}{\sum x_i}$$

که در آن $\bar{x}$ و S به ترتیب میانگین و انحراف معیار مشاهدات می‌باشند.

ضریب تغییرات را معمولاً در عدد ۱۰۰ ضرب می‌کنند و آن را به صورت زیر نشان می‌دهند.

$$\%CV=\frac{S}{\bar{x}}\times 100$$

اگر میانگین و واریانس ۵ مشاهده به ترتیب برابر با ۱۰ و ۴ باشند. ضریب تغییر را بدست آورید.

$$\%CV=\frac{S}{\bar{x}}\times 100=\frac{2}{10}\times 100=20$$

ویژگی‌های ضریب تغییر

۱- به واحد اندازه‌گیری بستگی ندارد.

۲- برای مقایسه دو صفت از یک جامعه با واحدهای اندازه‌گیری متفاوت مورد استفاده قرار می‌گیرد.

۱-۱۰ جدول توزیع فراوانی

مشاهدات آماری معمولاً از یک سری اندازه‌های بدست آمده از دسته‌ای از اشیاء تشکیل می‌شود و در مواردی که سری‌های بزرگ مشاهدات برای توصیف در دسترس باشند نیاز است در قالب خلاصه شده‌ای ارائه شود تا اطلاعات صحیحی از آنها کسب شود. برای توصیف دقیق داده‌ها، ترکیب بعضی از آنها با یکدیگر ضروری است. این نیاز با تشکیل یک جدول توزیع فراوانی برآورده می‌شود. جدول توزیع فراوانی مشاهدات را بر حسب تعداد اندازه‌هایی که داخل حدود معین قرار می‌گیرند دسته‌بندی می‌کند. صورت کلی جدول توزیع فراوانی عبارت است از:

شماره کلاس	کلاس	حد متوسط	فراوانی	درصد فراوانی	فراوانی تجمعی	درصد فراوانی تجمعی
۱	a_1-b_1	$\frac{a_1+b_1}{2}$	f_1	$\frac{f_1}{n}\times 100$	F_1	$\frac{F_1}{n}\times 100$
۲	a_2-b_2	$\frac{a_2+b_2}{2}$	f_2	$\frac{f_2}{n}\times 100$	F_2	$\frac{F_2}{n}\times 100$
. . .						
i	a_i-b_i	$\frac{a_i+b_i}{2}$	f_i	$\frac{f_i}{n}\times 100$	F_i	$\frac{F_i}{n}\times 100$
. . .						
k	a_k-b_k	$\frac{a_k+b_k}{2}$	f_k	$\frac{f_k}{n}\times 100$	$F_k=n$	$\frac{F_k}{n}\times 100$
جمع	-	-	n	۱۰۰	-	-

در این جدول توزیع فراوانی که k گروه‌بندی جمعیت است. نمادها به صورت زیر تعریف می‌شوند.

- a_1: کران پایین کلاس اول — b_1: کران بالا کلاس اول
- a_i: کران پایین کلاس i–ام — b_i: کران بالا کلاس iام
- a_k: کران پایین کلاس k–ام — b_k: کران بالا کلاس kام
- $\frac{a_i + b_i}{2}$ حد متوسط یا نمایند گروه بندی جمعیت iام برای $i = 1,2,\cdots,k$
- $b_i - a_i$ طول یا رده گروه بندی جمعیت iام برای $i = 1,2,\cdots,k$
- f_i فراوانی یا فراوانی مطلق گروه بندی جمعیت iام برای $i = 1,2,\cdots,k$
- $\frac{f_i}{n}$ فراوانی نسبی گروه بندی جمعیت iام برای $i = 1,2,\cdots,k$
- $\frac{f_i}{n} \times 100$ درصد فراوانی نسبی گروه بندی جمعیت iام برای $i = 1,2,\cdots,k$
- $F_1 = f_1$ فراوانی تجمعی گروه بندی جمعیت اول
- $F_2 = f_1 + f_2$ فراوانی تجمعی گروه بندی جمعیت دوم
- $F_2 = f_1 + f_2 + \ldots + f_i$ فراوانی تجمعی گروه بندی جمعیت iام
- $\frac{F_i}{n}$ فراوانی تجمعی نسبی گروه بندی جمعیت iام برای $i = 1,2,\cdots,k$
- $\frac{F_i}{n} \times 100$ درصد فراوانی تجمعی گروه بندی جمعیت iام برای $i = 1,2,\cdots,k$

به طوری که $F_k = \sum_{i=1}^{k} f_i = n$.

مثال جدول توزیع فراوانی، نمونه میزان درآمد سرانه روستایی سکونتگاه روستایی علی‌آباد را که از نظر صنایع روستایی و بر حسب میلیون ریال گروه‌بندی شده‌اند نشان می‌دهد. جدول توزیع فراوانی را کامل کنید.

مقاومت بتن	۱۰–۱۵	۱۵ –۲۰	۲۰–۲۵	۲۵ –۳۰	۳۰–۳۵
تعداد نمونه	۱۵	۲۰	۵۰	۱۵	۲۰

جدول توزیع فراوانی را کامل کنید.

مقاومت بتن	تعداد نمونه (فراوانی)	فراوانی نسبی	درصد فراوانی نسبی	فراوانی تجمعی	فراوانی تجمعی نسبی	درصد فراوانی تجمعی
۱۰–۱۵	۱۵	۰/۱۲۵	۱۲/۵	۱۵	۰/۱۲۵	۱۲/۵
۱۵–۲۰	۲۰	۰/۱۶۷	۱۶/۷	۳۵	۰/۲۹۲	۲۹/۲
۲۰–۲۵	۵۰	۰/۴۱۶	۴۱/۶	۸۵	۰/۷۰۸	۷۰/۸
۲۵–۳۰	۱۵	۰/۱۲۵	۱۲/۵	۱۰۰	۰/۸۳۳	۸۳/۳
۳۰–۳۵	۲۰	۰/۱۶۷	۱۶/۷	۱۲۰	۱	۱۰۰
کل	۱۲۰	۱	۱۰۰	–	–	–

در تشکیل جدول توزیع فراوانی، تعداد کلاس و طول کلاس نقش اساسی دارد. از رابطه‌ای که بین طول کلاس، تعداد کلاس و دامنه وجود دارد برای محاسبه طول کلاس استفاده می‌شود.

طول کلاس × تعداد کلاس = دامنه

$$R = k \cdot \ell$$

در تشکیل جدول توزیع فراوانی برای داده‌های خام همواره R قابل محاسبه است. اگر کلاس یا تعداد کلاس معلوم باشد دیگری از رابطه $R = k \cdot \ell$ قابل محاسبه است. اگر k و ℓ هر دو مجهول باشند. تعداد طبقات با استفاده از فرمول استرژ زیر بدست می‌آید.

$$k = 1 + 3/32 \log_{10}^{(n)}$$

که n حجم نمونه است.

مثال مقدار مصرف نهادهای تولید (سم و کود) در ۲۵ روز متوالی در یک مزرعه کشاورزی (برحسب کیلوگرم) به صورت زیر گزارش شده است.

الف– دامنه داده‌ها را بدست آورید.

ب– برای تشکیل جدول توزیع فراوانی، تعداد کلاس‌ها را بدست آورید.

ج- جدول توزیع فراوانی را تشکیل دهید و به سؤالات زیر پاسخ دهید.

۴	۳/۹	۴/۲	۳/۱	۲/۶	۱/۶	۲/۵	۱/۷۵	۱/۲	۱/۵
۵	۴/۵	۲	۱/۸	۳/۵	۴/۱	۵/۶	۵/۳	۵/۲	۴/۳
					۳/۸	۳	۲/۸	۱/۲۵	۵/۷۵

ابتدا داده را مرتب می‌کنیم

۲/۸	۲/۶	۲/۵	۲	۱/۸	۱/۷۵	۱/۶	۱/۵	۱/۲۵	۱/۲
۴/۵	۴/۳	۴/۲	۴/۱	۴	۳/۹	۳/۸	۳/۵	۳/۱	۳
					۵/۷۵	۵/۶	۵/۳	۵/۲	۵

$k = 1 + 3/32\log_{10}^{(25)} = 5/64$ ، $n = 25$ ، دامنه $= 5/75 - 1/2 = 4/55$

طول کلاس را تقریباً برابر با ۶ در نظر می‌گیریم $k = 6$

طول با توجه به رابطه $\ell = \frac{R}{k}$ برابر است با:

$$\ell = \frac{R}{K} = \frac{۴/۵۵}{۶} = ۰/۷۵۸ \sim ۰/۸$$

شماره	رده	فراوانی	فراوانی نسبی	درصد فراوانی نسبی	فراوانی تجمعی	درصد فراوانی تجمعی
۱	۱/۲-۲	۶	۰/۲۴	۲۴	۶	۲۴
۲	۲-۲/۸	۳	۰/۱۲	۱۲	۹	۳۶
۳	۲/۸-۳/۶	۴	۰/۱۶	۱۶	۱۳	۵۲
۴	۳/۶-۴/۴	۶	۰/۲۴	۲۴	۱۹	۷۶
۵	۴/۴-۵	۱	۰/۰۴	۴	۲۰	۸۰
۶	۵-۵/۸	۵	۰/۲	۲۰	۲۵	۱۰۰
		۲۵	۱	۱۰۰	—	—

۱- چند درصد از نهادهای تولید انجام شده دارای وزن بین ۲/۶ – ۲/۸ کیلوگرم است.

۲- بیشترین نهادهای تولید در چه رده وزنی قرار دارد.

۳- چند درصد نهادهای تولید دارای وزنی کمتر یا مساوی ۵ کیلوگرم است.

پاسخ

- برای پاسخ به سؤال ۱ با توجه به ستون دوم و سوم برابر با ۱۶ درصد چون ۱۶ درصد در رده ۲/۶ — ۲/۸ استفاده شده است.
- برای پاسخ به سؤال ۲، با توجه به ستون سوم یعنی فراوانی، برابر با ۲۴ درصد چون بیشترین فراوانی مربوط به کلاس اول و چهارم با ۲۴ درصد می‌باشد.
- برای پاسخ به سؤال ۳ با توجه به ستون درصد فراوانی تجمعی برابر با ۸۰ درصد می‌باشد.

یکی دیگر از استفاده‌های جدول توزیع فراوانی برای متغیرهای کیفی و گسسته با دامنه محدود است. در این جداول توزیع فراوانی با داده‌های گسسته یا کیفی هر عدد یا گروه معروف همان صورتی است که نمایش داده شده است.

مثال مدیر یک شرکت صنعتی روستایی در استان اصفهان یک مؤسسه قصد دارد در خصوص تعداد افراد تحت تکفل کارگران روستایی، مطالعه‌ای را انجام دهد. بدین منظور پرسشنامه‌ای تنظیم و اطلاعات زیر به عنوان داده‌های خام مربوط به افراد تحت تکفل هفتاد پرسنل جمع‌آوری و گزارش شده است.

۵	۰	۴	۷	۱	۲	۲	۳	۰	۵
۱	۴	۰	۳	۷	۴	۱	۱	۰	۷
۴	۵	۰	۱	۱	۷	۲	۹	۳	۱
۸	۱۱	۱	۲	۴	۵	۱	۱	۰	۴
۸	۲	۷	۶	۹	۳	۰	۱	۴	۰
۲	۰	۱	۷	۵	۴	۰	۱	۲	۷
۱	۸	۴	۷	۱	۵	۱	۳	۰	۶

جدول توزیع فراوانی را تشکیل دهید و به سؤالات زیر پاسخ دهید.

تعداد افراد تحت تکفل	فراوانی	فراوانی نسبی	درصد فراوانی نسبی	فراوانی تجمعی	درصد فراوانی تجمعی
۰	۱۱	۰/۱۵۷	۱۵/۷	۱۱	۱۵/۷
۱	۱۶	۰/۲۲۹	۲۲/۹	۲۷	۳۸/۶
۲	۷	۰/۱۰۰	۱۰/۰	۳۴	۴۸/۶
۳	۵	۰/۰۷۷	۷/۷	۳۹	۵۶/۳
۴	۹	۰/۱۲۹	۱۲/۹	۴۸	۶۹/۲
۵	۶	۰/۰۸۸	۸/۸	۵۴	۷۸
۶	۲	۰/۰۲۹	۲/۹	۵۶	۸۰/۹
۷	۸	۰/۱۱۴	۱۱/۴	۶۴	۹۲/۳
۸	۳	۰/۰۴۲	۴/۲	۶۷	۹۶/۵
۹	۲	۰/۰۲۱	۲/۱	۶۹	۹۸/۶
۱۰	۰	۰	۰	۶۹	۹۸/۶
۱۱	۱	۰/۰۱۴	۱/۴	۷۰	۱۰۰
کل	۷۰	۱	۱۰۰	–	–

الف: چند درصد از شاغلان روستایی فقط چهار نفر تحت تکفل دارند.

ب: چند درصد از شاغلان روستایی بین ۲ تا ۶ نفر تحت تکفل دارند.

ج: چند درصد از شاغلان روستایی بیشتر یا مساوی ۳ نفر تحت تکفل دارند.

– برای پاسخ به سؤال الف با توجه به ستون‌های اول و چهارم برابر است با ۱۲/۹ درصد.

– برای پاسخ به سؤال الف با توجه به ستون‌های اول و چهارم برابر است با ۱۲/۹ درصد.

– برای پاسخ به سؤال ب با توجه به ستون‌ها اول و چهارم برابر است با

$$10/0+7/7+12/9+8/8+2/9=4/23$$ درصد

– برای پاسخ به سؤال ج با توجه به ستون فراوانی تجمعی برابر است با:

$$100-\left(15/7+22/9\right)=100-38/6=61/4$$

۱-۱۱ محاسبه میانگین و واریانس در جدول توزیع فراوانی

برای محاسبه میانگین و واریانس در جدول توزیع فراوانی از فرمول‌های زیر استفاده می‌کنیم.

میانگین: $\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i \; x_i$

واریانس: $S^2 = \frac{1}{n}\sum_{i=1}^{k} f_i \left(x_i - \bar{x}\right)^2$

که در آن، k تعداد کلاس، f_i فراوانی کلاس iام و $n = \sum_{i=1}^{k} f_i$

مثال حقوق پرداخت شده برای کارگران روستایی در سه شرکت تولید روستایی به صورت زیر گزارش شده است (ارقام به صد هزار ریال) میانگین و واریانس را محاسبه کنید.

شرکت تولید روستایی	A	B	C
میزان حقوق	۸۶	۹۲	۸۹
تعداد کارگران روستایی	۴	۱۵	۱۱

$$n = \sum_{i=1}^{k} f_i = 4 + 15 + 11 = 30$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i \; X_i = \frac{1}{30}\left[4\times 86 + 15\times 92 + 11\times 89\right] = 90/1$$

به طور متوسط هر کارگران روستایی ۹۰۱۰۰۰۰ ریال یا ۹۰۱ هزار تومان دریافت می‌کند.

$$S^2 = \frac{1}{n}\sum_{i=1}^{k} f_i \left(x_i - \bar{x}\right)^2$$

$$= \frac{1}{30}\sum_{i=1}^{k} f_i \left(x_i - \bar{x}\right)^2 = \frac{1}{30}\left[f_1\left(x_1 - \bar{x}\right)^2 + f_2\left(x_i - \bar{x}\right)^2 + f_3\left(x_i - \bar{x}\right)^2\right]$$

$$= \frac{1}{30}\left[4\left(86 - 90/1\right)^2 + 15\left(92 + 90/1\right)^2 + 11\left(89 - 90/1\right)^2\right]$$

$$= \frac{1}{30}\left[4\left(-4/1\right)^2 + 15\left(1/9\right)^2 + 11\left(-1/1\right)^2\right] = \frac{134/7}{30} = 4/49$$

مثال دستمزد ۱۰ میرآب در مزرعه کشاورزی خاصی به صورت زیر گزارش شده است (ارقام به میلیون ریال)

رده دستمزد	۱–۲	۲–۳	۳–۴	۴–۵
فراوانی	۲	۵	۲	۱

متوسط انحراف معیار دستمزدها را بدست آورید.

چون متغیر دستمزد متغیر کمی پیوسته است برای محاسبه میانگین و انحراف معیار ابتدا حد متوسط کلاس‌ها را بدست می‌آوریم. با توجه به نماینده کلاسها میانگین موزون برابر است با:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{4} f_i x_i = \frac{1}{10}\left[2\times 15 + 5\times 2/5 + 2\times 3/5 + 1\times 4/5\right]$$

$$= \frac{1}{10}(27) = 2/7$$

رده دستمزد	فراوانی (f_i)	نماینده x_i	$f_i x_i$	$(x_i - 2/7)$	$f_i\left(x_i - 2/7\right)^2$
۱–۲	۲	۱/۵	۳	–۱/۲	۲/۸۸
۲–۳	۵	۲/۵	۱۲/۵	–۰/۲	۰/۲
۳–۴	۲	۳/۵	۷	۰/۸	۱/۲۸
۴–۵	۱	۴/۵	۴/۵	۱/۸	۳/۲۴
کل	۱۰	–	۲۷	–	۷/۶

$$S^2 = \frac{1}{10}\sum_{i=1}^{4} f_i \left(x_i - 2/7\right)^2 = \frac{7/6}{10} = 0/76$$

انحراف معیار$= S = \sqrt{0/76} = 0/87$

۱-۱۲ محاسبه نما در جدول توزیع فراوانی

نمای یک مجموعه از مشاهدات مقدار یا مقادیری بود که بیشتر از بقیه تکرار شده بود. ولی در جدول توزیع فراوانی برای داده‌های پیوسته نما از فرمول زیر محاسبه می‌شود.

$$M = a_i + \frac{d_1}{d_1 + d_2} \times l$$

که:

a_i = حد پایین کلاس که نما در آن قرار دارد.

l = طول کلاس که نما درآن قرار دارد.

اگر i شماره کلاسی که نما درآن قرار دارد، باشد؛ d_1 و d_2 به صورت زیر محاسبه می‌شود.

$$d_1 = f_i - f_{i-1} \quad , \quad d_2 = f_i - f_{i+1}$$

یافتن کلاسی که نما در آن قرار دارد نقش مهمی در محاسبه نما دارد. در یک جدول توزیع فراوانی نما در کلاسی است که دارای بیشترین فراوانی است.

مثال برای جدول توزیع فراوانی زیر نما را بدست آورید.

شماره کلاس	کلاس	فراوانی	فراوانی تجمعی
۱	۲/۰–۲/۴	۵	۵
۲	۲/۴–۲/۸	۵	۱۰
۳	۲/۸–۳/۲	۹	۱۹
۴	۳/۲–۳/۶	۴	۲۳
۵	۳/۶–۴/۰	۴	۲۷
۶	۴/۰–۴/۴	۳	۳۰

چون کلاس سوم بیشترین فراوانی را دارد. پس نما در این کلاس قرار دارد وi=۳ است و نما از رابطه زیر بدست می‌آید:

$$M = a_3 + \frac{d_1}{d_1 + d_2} \times l$$

$$l = 3/2 - 2/8 = 0/4 \ , \quad d_1 = f_3 - f_2 = 4 \quad , \quad d_2 = f_3 - f_4 = 5$$
$$M = 2/8 + \frac{4}{4+5}(0/4) = 2/977$$

۱-۱۳ محاسبه میانه در جدول توزیع فراوانی

در جدول توزیع فراوانی میانه از فرمول زیر محاسبه می‌شود.

$$m = a_i + \frac{\frac{n}{2} - F_{i-1}}{f_i} \times l$$

که f_i فراوانی کلاسی که میانه در آن قرار دارد و F_{i-1} فراوانی تجمعی کلاس ما قبل کلاسی که میانه در آن قرار دارد است. برای یافتن کلاسی که میانه در آن قرار دارد اقدامات زیر را انجام می‌دهیم.

۱- $\frac{n}{2}$ را محاسبه می کنیم.

۲- در ستون فراوانی تجمعی از بالا به پایین اعداد را با $\frac{n}{2}$ مقایسه می‌کنیم. اولین عددی که بزرگتر یا مساوی $\frac{n}{2}$ باشد کلاس متناظر کلاسی است که میانه درآن قرار دارد. برای مثال قبل داریم:

$$\frac{n}{2} = \frac{30}{2} = 15 \qquad , \qquad 15 \le 19 \Rightarrow i = 3$$
$$f_i = f_3 = 9 \qquad , \qquad F_{i-1} = F_2 = 10$$
$$m = 2/8 + \frac{15-10}{9}(0/4) = 3/022$$

۱-۱۴ نمودارهای توصیفی

یکی از روشهای نمایش مشاهدات استفاده از نمودارهای آماری است. از طریق نمودارهای آماری می‌توان در حداقل زمان توزیع داده‌ها را نمایش و انتقال پیام را به

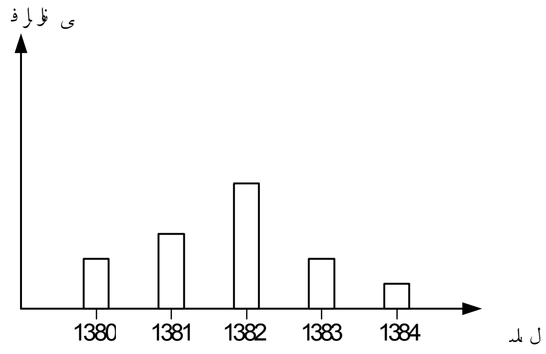
تسهیل نمود. در ادامه، نمودارهای میله‌ای، دایره‌ای، مستطیلی، چند ضلعی و فراوانی تراکمی را مورد بحث قرار می‌دهیم.

۱-۱۴-۱ نمودار میله‌ای (ستونی)[۱]

نمودار میله‌ای نموداری است که فراوانی‌ها از طریق سطوحی از میله‌ها نمایش داده می‌شوند. این نمودار خاص متغیرهای کیفی و یا متغیرهای کمی گسسته است. برای رسم آن متغیرها را در محور افقی یا محور xها و فراوانی‌ها را بر محور عمودی یا محور y ها نمایش می‌دهند و براساس تعداد فراوانی‌های مشاهده هر متغیر، ستونی با ارتفاع متناظر با فراوانی متغیر ترسیم می‌شود.

مثال تعداد طرح های روستایی اجراشده در ۵ سال گذشته در شهرستان سمنان به صورت زیر گزارش شده است نمودار میله‌ای را رسم کنید.

سال	۱۳۸۰	۱۳۸۱	۱۳۸۲	۱۳۸۳	۱۳۸۴
تعداد طرح های روستایی	۲	۳	۵	۲	۱



نمودار شماره ۱. نمودار میله‌ای (ستونی)

۱. Bar Graph

۱-۱۴-۲ نمودار دایره‌ای

نمودار دایره‌ای یکی از انواع نمودارهای قابل درک در توصیف آماری است که نمایش توزیع متغیرها به کمک آن به سادگی امکان‌پذیر است. برای این منظور کافی است که قسمتی از سطح درونی دایره را متناظر با سهم متغیر مورد مطالعه و از مجموع درجات داخلی دایره جدا و به متغیر مورد نظر تخصیص دهیم. در جدول توزیع فراوانی سطح درونی هر گروه یا رده از رابطه زیر بدست می‌آید.

$$S_i = \frac{360}{\sum_{i=1}^{k} f_i} \times f_i \quad ، \quad i = 1,2,\dots,k \quad , \quad \sum_{i=1}^{k} S_i = 360$$

در صورتی که مشاهدات ما از متغیرها برحسب درصد بیان شده باشد. می‌توان کل مشاهدات را برحسب درصد بیان نموده و برای درصد مورد نظر از متغیر مورد مطالعه، سهم درجه را معین نمود که فرمول آن عبارتست از:

$$S_i = \frac{100}{\sum_{i=1}^{k} f_i} \times f_i \quad ، \quad i = 1,2,\dots,k \quad , \quad \sum_{i=1}^{k} S_i = 100$$

مثال تعداد طرح های روستایی اجراشده در سه سال گذشته به صورت زیر گزارش شده است نمودار دایره‌ای را رسم کنید (ارقام به هزار)

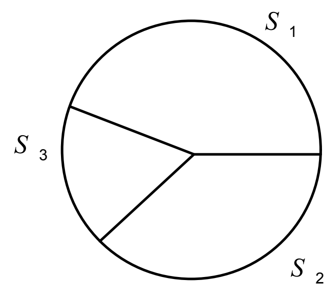
سال	۱۳۸۳	۱۳۸۴	۱۳۸۵
تعداد مجوز	۳۱۵	۸۲	۱۵۳

در مثال:

$$f_1 = 3/5 \qquad f_2 = 82 \qquad f_3 = 153$$

$$n = \sum_{i=1}^{3} f_i = 550 \quad S_i = \frac{360}{n} \times f_i \quad ، \quad i = 1,2,3$$

$$S_1 = 206/18 \qquad S_2 = 53/67 \qquad S_3 = 100/15$$



مثال درصد پیشرفت یک پروژه طرح هادی بزرگ در طول مدت پروژه به صورت زیر گزارش شده است

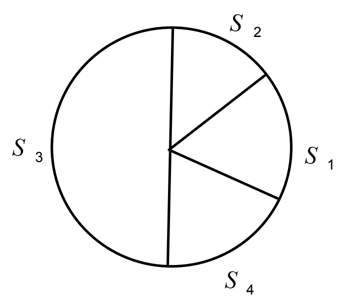
طول مدت	شش ماه اول	شش ماه دوم	شش ماه سوم	شش ماه چهارم
درصد پیشرفت	۱۰	۲۰	۵۰	۲۰

نمودار دایره‌ای را رسم کنید.

$$f_1 = 10 \quad , \quad f_2 = 20 \quad , \quad f_3 = 50 \quad , \quad f_4 = 20 \quad , \quad \sum_{i=1}^{4} f_i = 100$$

$$S_i = \frac{100}{\sum f_i} \times f_i \qquad ، \quad i = 1, 2, 3, 4$$

$$S_i = \frac{100}{100} \times 10 = 10 \qquad ، \quad S_2 = 20 \quad ، \quad S_3 = 50 \quad ، \quad S_4 = 20$$



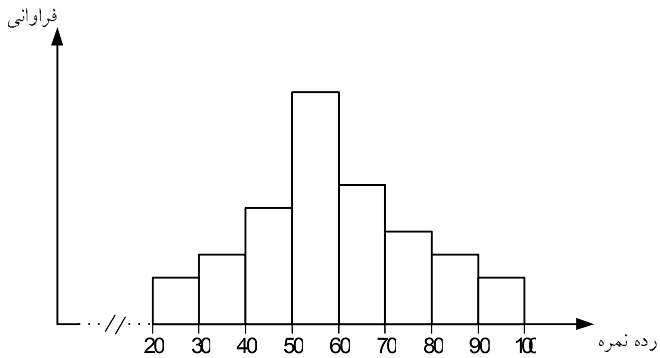
نمودار شماره ۲. نمودار دایره‌ای

۱-۱۴-۳ نمودار مستطیلی

در صورتی که متغیرهای مورد مطالعه از نوع متغیرهای پیوسته باشند از نمودار مستطیل یا هیستوگرام استفاده می‌شود و برای رسم آن، رده یا کلاس را روی محور xها و فراوانی مشاهده شده را روی محور yها مشخص می‌کنند و نمودار مربوطه را رسم می‌کنند. در مواقعی که کوچکترین مشاهده یا کران پایین کلاس اول از مبدأ یا صفر فاصله داشته باشد و یا دارای ارزش بیشتری باشد در محور xها یا افقی از برش (--//--) استفاده می‌شود.

مثال توزیع نمرات دانشجویان رشته جغرافیا و برنامه ریزی روستایی در درس آمار به صورت جدول زیر گزارش شده است. نمودار هیستوگرام را رسم کنید. (دامنه نمرات از صفر تا صد می‌باشد.)

رده نمره	۲۰-۳۰	۳۰-۴۰	۴۰-۵۰	۵۰-۶۰	۶۰-۷۰	۷۰-۸۰	۸۰-۹۰	۹۰-۱۰۰
فراوانی	۲	۳	۵	۱۰	۶	۴	۳	۲



نمودار شماره ۳. نمودار مستطیلی

۱-۱۴-۴ نمودار چند ضلعی

برای ترسیم نمودار چند ضلعی، از نقاط میانی یا نماینده رده‌های متغیرها استفاده می‌شود بنابراین نقاط میانی هر رده به عنوان نماینده آن رده در محور xها و تعداد

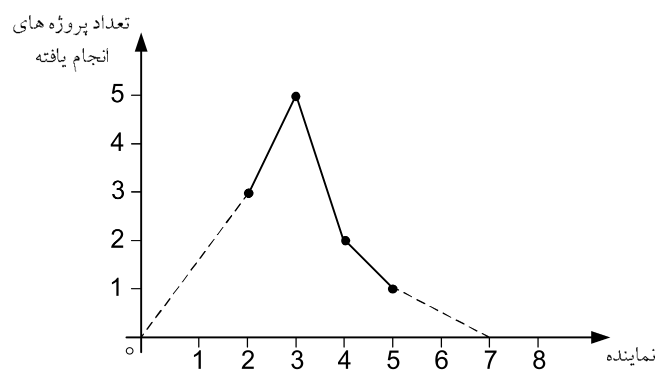
مشاهدات یا فراوانی مربوط به هر کدام از رده‌ها در محور yها مشخص می‌شود. پس از وصل نقاط بدست آمده، منحنی را از طرفین ادامه می‌دهیم تا محور xها را قطع کند.

مثال شرکت مهندس مشاوری پروژهای طرح های هادی، دوره زمانی پروژه‌های خود را پی در پی به صورت زیر گزارش داده است (برحسب سال) نمودار چند ضلعی را رسم کنید.

فاصله زمانی	۱/۵ –۲/۵	۲/۵ –۳/۵	۳/۵ –۴/۵	۴/۵ –۵/۵
تعداد پروژه‌های انجام یافته	۳	۵	۲	۱

برای رسم نمودار چند ضلعی ابتدا، متوسط یا نماینده رده‌ها را بدست می‌آوریم.

نماینده	۲	۳	۴	۵
تعداد پروژه‌های انجام یافته	۳	۵	۲	۱



نمودار شماره ۴. نمودار چندضلعی

از نمودار چند ضلعی معمولاً برای مقایسه دو گروه از متغیرها استفاده می‌شود.

مثال میزان هزینه‌های اجرایی دو پروژه طرح هادی در نوبتهای مختلف در جدول زیر گزارش شده است. نمودار چندضلعی دو پروژه را در یک نمودار رسم کنید و آن را تحلیل کنید. (ارقام به میلیارد ریال)

پروژه اول

هزینه	۰-۱/۵	۱/۵-۳	۳-۴/۵	۴/۵-۶
نوبت	۱	۲	۳	۱

پروژه دوم

هزینه	۰-۱/۵	۱/۵-۳	۳-۴/۵	۴/۵-۶
نوبت	۲	۳	۲	۱

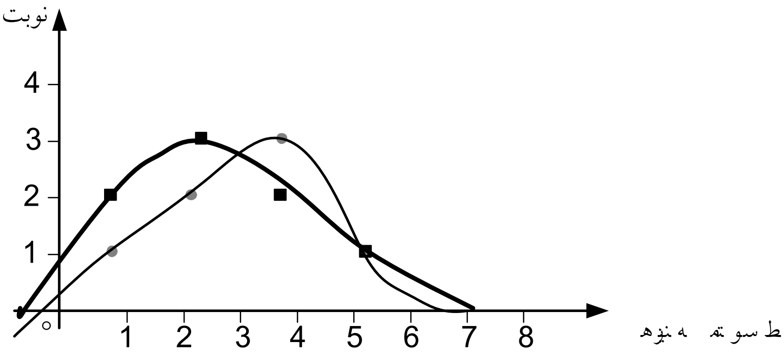
ابتدا نماینده رده‌ها را در دو پروژه بدست می‌آوریم.

پروژه اول

متوسط هزینه	۰/۷۵	۲/۲۵	۳/۷۵	۵/۲۵
نوبت	۱	۲	۳	۱

پروژه دوم

متوسط هزینه	۰/۷۵	۲/۲۵	۳/۷۵	۵/۲۵
نوبت	۲	۳	۲	۱



نمودار ۵. نمودار چند ضلعی پروژه اول و دوم

متوسط هزینه پروژه دوم برای متوسط هزینه کمتر از ۳ از تعداد نوبت‌های بیشتری نسبت به پروژه اول برخوردار است و برای متوسط هزینه بیش از ۳ پروژه اول از نظر نوبت بیشتر از پروژه دوم است.

۱-۱۴-۵ نمودار فراوانی تراکمی (اجایو)[۱]

برای رسم نمودار فراوانی تراکمی تجمعی از فراوانی تجمعی کران بالای رده‌ها استفاده می‌کنیم. به طوری که کران بالای رده‌ها روی محور xها و فراوانی تجمعی متناظر با آن را روی محور yها مشخص کرده و از وصل نقاط مذکور به هم نمودار چند ضلعی فراوانی تجمعی از درصد فراوانی تجمعی استفاده کنیم، منحنی حاصل را منحنی درصد فراوانی تجمعی گویند. در بعضی از کتب آمار از این نمودار به عنوان نمودار اجایو نیز نام می‌برند.

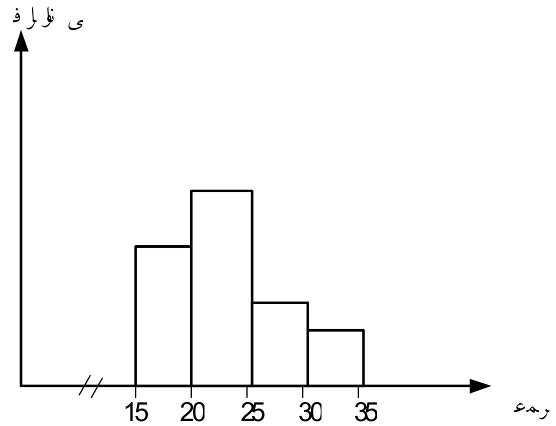
مثال یک مهندس پروژه طرح هادی روستایی در یک بررسی آماری از عمر ساختمان درسکونتگاههای روستایی دره کن اطلاعات زیر را گزارش داده است. (ارقام به ده سال)

---

۱. Ojive

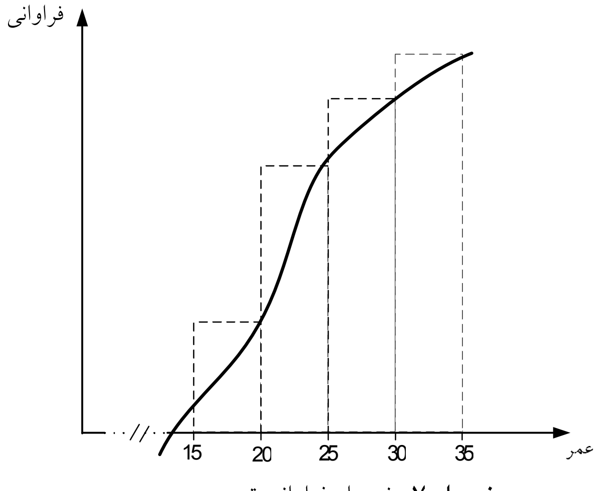
عمر	۱۵ –۲۰	۲۰–۲۵	۲۵ –۳۰	۳۰–۳۵
تعداد ساختمان در تهران	۵	۷	۳	۲

نمودار هیستوگرام و فراوانی تجمعی را رسم کنید.



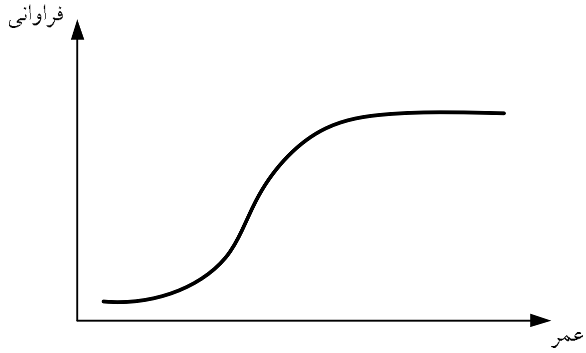
نمودار ۶. نمودار هیستوگرام

عمر	۱۵ –۲۰	۲۰–۲۵	۲۵ –۳۰	۳۰–۳۵
فراوانی تجمعی	۵	۱۲	۱۵	۱۷



نمودار ۷. نمودار فراوانی تجمعی

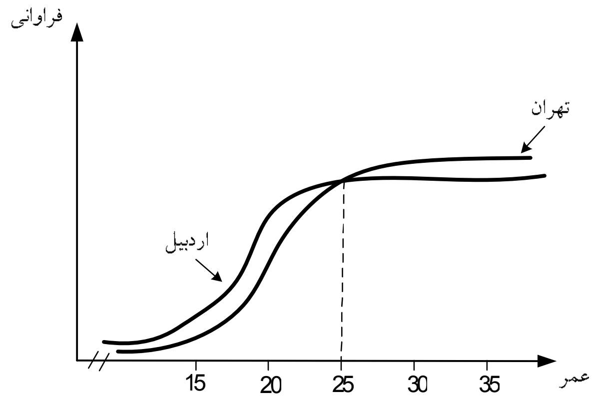
اگر نقطه چین‌ها را حذف کنیم خواهیم داشت:



مثال کاربردی

برای مقایسه عمر ساختمان‌ها در سکونتگاههای دره کن درتهران و سکونتگاههای روستایی اردبیل اطلاعات زیر بدست آمده نمودار فراوانی تجمعی را رسم کنید و آن را تحلیل کنید.

عمر	۱۵ -۲۰	۲۰–۲۵	۲۵ -۳۰	۳۰–۳۵
تعداد ساختمان روستایی در تهران	۵	۷	۳	۲
تعداد ساختمان در روستاهای اردبیل	۶	۸	۲	۱



از نمودار اجایو مشخص می‌شود تعداد ساختمانها حداکثر بیست و پنج سال عمر می‌کند در اردبیل بیشتر از تهران ولی عمر ساختمان‌های بیشتر از بیست و پنج سال در تهران بیشتر از اردبیل است.

خودآزمایی

۱- جامعه و نمونه را تعریف کرده و بررسی کنید کدام یک از موارد زیر جامعه و کدام یک نمونه است.

- جمعیت روستایی ایران
- انجمن فارغ‌التحصیلان برنامه ریزان روستایی
- کلیه پروژه‌های روستایی اتمام یافته
- پروژه‌های عمران روستایی با بودجه کمتر از یک میلیارد تومان
- انجمن مهندسین مشاور روستایی
- انجمن برنامه روستایی

۲- اگر داده‌های زیر درآمد سرانه خانوارهای روستایی بخش ساوجبلاغ باشند. (ارقام به صد هزار تومان است).

۶۱	۶۰	۸۵	۵۲	۴۸	۶۵	۵۴	۷۴	۴۹
۳۷	۹۰	۸۳	۴۰	۵۶	۸۱	۸۵	۵۹	۵۷

الف- $\sum_{i=1}^{18} X_i$ را بدست به آورید.

ب- $\sum X_i^2$ را بدست آورید.

۳- انواع داده‌های آماری را شرح دهید.

۴- با توجه به ویژگی سیگما نشان دهید که رابطه‌های زیر درست است.

الف- $\sum_{i=1}^{n} 2X_i = 2\sum_{i=1}^{n} X_i$

ب- $\sum_{i=1}^{n}\left(X_i + Y_i\right) = \sum_{i=1}^{n} X_i + \sum_{i=1}^{n} Y_i$

ج- $\sum_{i=1}^{n}\left(X_i + Y_i\right)^2 = \sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} Y_i^2 - 2\sum_{i=1}^{n} X_i Y_i$

۵-متوسط هزینه های تولید کشاورزی ۵۰ مزرعه نمونه در سکونتگاههای روستایی دشت قزوین برحسب میلیارد تومان به صورت زیر گزارش شده است.

۲/۷۵	۶/۲۵	۳/۵	۶/۵	۱/۵	۴	۱	۲	۳	۴
۱/۶۵	۳/۷۵	۱/۷۵	۲/۵	۲/۷۵	۲/۵	۶/۲۵	۲/۵	۴	۱
۲/۷۵	۱/۷۵	۶/۲۵	۴/۹۵	۶/۲۵	۲/۷۵	۴/۲۵	۳/۷۵	۲	۱/۵
۶/۲۵	۵/۲۵	۵/۵	۶/۲۵	۷/۲۵	۳	۲	۳/۲۵	۱	۴/۲۵
۴/۷۵	۱/۷۵	۲/۷۵	۴/۲۵	۳/۲۵	۴/۷۵	۷/۵	۶/۵	۲/۵	۳/۵

با استفاده از جدول اعدا تصادفی، نمونه‌ای به حجم ۵ انتخاب کنید.

۶- انواع روش‌های نمونه‌گیری را نام ببرید.

۷- اگر میزان مالیات برای ۲۵ پروژه انجام شده به صورت زیر باشد (ارقام به صد هزار تومان).

۱۷	**۲۰**	**۲۳/۵**	**۲۵/۹**	**۱/۵**	**۲۰/۷**	**۱۹**	**۱۶**	**۱۴/۵**	**۲۳**	**۲۲**	**۲۱/۵**	**۷/۵**
	۳۱/۵	**۳۱**	**۲**	**۲۰**	**۲۲**	**۳۰**	**۲۴**	**۲۸**	**۲۷/۱**	**۱۷**	**۱۸**	**۲۵/۳**

میانگین حسابی، هندسی و هارمونیک را حساب کنید و نشان دهید که

$$H \le G \le \overline{X}$$

۸- انواع شاخص‌های مرکزی و غیر مرکزی را نام ببرید.

۹- نشان دهید میانگین حسابی دو گروه زیر با هم برابرند.

$x_۱ \;\; x_۲ \;\; \ldots \;\; x_n$

$$\frac{(۲x_۱-۳),(۲x_۲-۳),\ldots,(۲x_n-۳)}{۲}+\frac{۳n}{۲}$$

۱۰- اگر نرخ رشد، نوسازی بافت سکونتکاههای روستایی در ۵ سال گذشته به صورت زیر باشد.

سال	**۱**	**۲**	**۳**	**۴**	**۵**
میزان رشد	**۰/۲**	**۰/۳**	**۰/۲۵**	**۰/۱۵**	**۰/۳۲**

۱۱- سه گروه از مهندسین مشاور مطالعات اجتماعی اقتصادی سکونتگاههای روستایی در دشت مغان پیشنهادهایی را برای اتمام پروژه به صورت زیر ارائه داده‌اند.

کارفرما	**۱**	**۲**	**۳**
مدت زمان (سال)	**۰/۲**	**۰/۳**	**۰/۲۵**

میانگین هارمونیک را بدست آورید.

۱۲- مهاجرت ۱۰ سکونتگاه روستایی در نیشابوربه صورت زیر گزارش شده، میانه را بدست آورید.

۱۲	۱۹	۱۶	۱۴	۱۵	۱۰	۹	۱۳	۱۱	۱۷

۱۳- میزان سرمایه گذاری اقتصادی در بخش دولتی در سکونتگاههای روستایی شهرستان ساوه در بخش مسکن روستایی به صورت زیر گزارش شده است (ارقام به میلیون تومان)

۷۸	۳۰	۶۳۲	۴۴	۱۷۰	۳۱	۲۱	۳۷	۳۹	۴۴
۳۷	۶۶	۱۸۱	۵۰	۸۲	۴۴	۴۳	۲۵۰	۶	۴۲

الف- نمای داده‌های فوق را بدست آورید.

ب- میانه داده‌های فوق را بدست آورید.

۱۴- اگر میزان تولید کشاورزی ۲۰ سکونتگاه روستایی دربجنورد بصورت تن در هکتارا به صورت زیر باشد. (ارقام به میلیون تن)

۱۹	۱۴	۱۶	۹	۵	۱۵	۱۰	۱۳	۱۱	۱۷
۱۰	۲۰	۱۶	۱۵	۱۴	۱۵	۷	۱۶	۱۸	۱۷

الف- دامنه داده‌های را بدست آورید.

ب- میانه داده‌های را بدست آورید.

ج- پس از محاسبه واریانس و انحراف معیار نمونه، ضریب تغییر را بدست آورید.

۱۵- برای جدول توزیع فراوانی زیر میانگین، میانه و واریانس را حساب کنید.

فراوانی	کلاس
۲	۹۹/۵-۱۰۴/۵
۸	۱۰۴/۵-۱۰۹/۵
۱	۱۰۹/۵-۱۱۴/۵
۱۳	۱۱۴/۵-۱۱۹/۵

۱۱۹/۵-۱۲۴/۵	۷
۱۲۴/۵-۱۲۹/۵	۱
۱۲۹/۵-۱۳۴/۵	۱

میانگین، نما، و واریانس را حساب کنید.

۱۶- تعداد کارگران ۱۰کارگاه تولید فرش دستی در بخش کلات نادری در خراسان رضوی به صورت زیر گزارش شده است.

معدل	۱	۲	۳	۴	۵	۶	۷	۸	۹	۱۰
تعداد کارگران	۱۱۰	۷۳۱	۱۰۳۱	۸۴	۲۰	۱۱۸	۱۱۶۲	۱۹۷۷	۱۰۳	۷۵۲

الف- دامنه تعداد کارگران برای ۱۰ معدل را بدست آورید.

ب- میانگین، واریانس و انحراف معیار تعداد کارگران در ۱۰ کارگاه را محاسبه کنید.

۱۷- نشان دهید که

$$S^{2} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^{2} = \frac{1}{n}\sum_{i=1}^{n} X_i^{2} - \overline{X}^{2}$$

۱۸- دو نفر از مهندسین آبیاری برای اتمام ۵ پروژه آبرسانی تحت فشار زمان‌های برحسب سال به صورت زیر ارائه داده‌اند.

مهندس	۱	۲	۳	۴	۵
I	۲	۱	۴	۵	۳
II	۱/۵	۲/۵	۳	۲	۳

ضرایب تغییر برای زمان‌های اعلام دو مهندس آبیاری را محاسبه و تحلیل کنید.

۱۹- در یک جمع آوری ازاطلاعات سن ۵۵ نفر از کشاورزان به صورت زیر گزارش شده است.

۳۲	۷۰	۴۲	۴۹	۳۷	۵۰	۳۵	۳۹	۴۰	۴۷	۵۴	۴۱
۴۲	۳۳	۴۵	۳۹	۶۹	۳۴	۳۰	۴۰	۵۰	۳۹	۵۰	۴۴
۴۵	۳۶	۳۸	۵۲	۴۲	۵۰	۳۴	۴۲	۲۷	۶۰	۶۳	۴۴

۴۲	۶۰	۳۳	۴۶	۶۳	۵۵	۲۷	۳۱	۴۶	۴۸	۴۳	۳۵
					۵۰	۵۰	۵۳	۳۴	۴۵	۴۶	۳۵

چارک‌های اول، دوم و سوم را بدست آورید و آن را تحلیل کنید.

۲۰- جدول توزیع فراوانی زیر را کامل کنید.

فراوانی	کلاس
۶	۱۱/۵-۱۲/۵
۱	۱۲/۵-۱۳/۵
۳	۱۳/۵-۱۴/۵
۶	۱۴/۵-۱۵/۵
۸	۱۵/۵-۱۶/۵
۲	۱۶/۵-۱۷/۵
۳	۱۷/۵-۱۸/۵
۱	۱۸/۵-۱۹/۵

۲۱- میزان مصرف سوخت بنزین ۲۵ تا از ماشین‌های کشاورزی شرکت سهامی زراعی نیل آباد در تربت جام به صورت زیر گزارش شده است (ارقام ۱۰ لیتر).

۳/۷۰	۳/۷۴	۳/۷۰	۳/۷۷	۳/۸۰
۳/۶۷	۳/۶۸	۳/۷۶	۳/۸۶	۳/۷۳
۳/۷۴	۳/۸۰	۳/۷۰	۳/۸۳	۳/۵۷
۳/۶۶	۳/۶۵	۳/۷۳	۳/۷۴	۳/۷۸
۳/۶۴	۳/۷۳	۳/۷۸	۳/۶۴	۳/۷۵

جدول توزیع فراوانی را تشکیل دهید.

۲۲- جدول توزیع فراوانی دستمزد ۲۰۰ نفر از کارگران روستایی شهرک صنعتی روستایی در قم به صورت زیر گزارش شده است.

رده- کلاس	۸۹/۵-۱۰۴/۵	۱۰۴/۵-۱۱۹/۵	۱۱۹/۵-۱۳۴/۵	۱۳۴/۵-۱۴۹/۵	۱۴۹/۵-۱۶۴/۵	۱۶۴/۵-۱۷۹-۵
فراوانی	۲۴	۶۲	۷۲	۲۶	۱۲	۴

- جدول توزیع فراوانی را کامل کنید.
- چند درصد از کارگزان دستمزد کمتر مساوی ۱۴۹/۵ است.
- فراوانی مربوط به کدام گروه است و چند درصد.

۲۳- برای جدول توزیع فراوانی زیر، میانگین، میانه، نما و واریانس را محاسبه کنید.

کلاس	۱-۳	۴-۶	۷-۹	۱۰-۱۲	۱۳-۱۵
فراوانی	۱	۴	۵	۱	۱

۲۴- ۴ پروژه راه سازی روستایی همراه تعداد کارگران به صورت زیر گزارش شده است.

پروژه	۱	۲	۳	۴
تعداد کارگران	۱۰۲	۱۰۰	۱۰۷	۱۰۴

- نمودار میله‌ای را رسم کنید.
- نمودار دایره‌ای را رسم کنید.

۲۶- میزان مهاجر فرستی ۲۰ سکونتگاه روستایی در شهرستان ایلام به صورت زیر داده شده است. نمودار مستطیلی را رسم کنید.

رده	۱۰-۱۲	۱۲-۱۴	۱۴-۱۶	۱۶-۱۸	۱۸-۲۰
فراوانی	۲	۳	۵	۳	۲

۲۷- برای جدول توزیع فراوانی زیر، نمودار چند ضلعی و تراکمی (اجایو) را رسم کنید.

کلاس	فراوانی
۱۲/۵-۲۷/۵	۶
۵/۲۷-۴۲/۵	۳

۴۲/۵–۵/۵۷	۵
۵/۵۷–۷۲/۵	۸
۷۲/۵–۸۷/۵	۶
۸۷/۵–۱۰۲/۵	۲

۲۸– جدول توزیع فراوانی زیر ماه‌های سپری برای اتمام ۲۳ پروژه طرح آبرسانی تخت فشاررا نشان می‌دهد.

رده– ماه	۹–۱۳	۱۴–۱۹	۲۰–۲۵	۲۶–۲۸	۲۹–۳۲
فراوانی	۱	۶	۲	۵	۹

نمودار مناسب را رسم کنید.

فصل دوم

آزمون فرض‌های آماری

مقدمه

در بسیاری از پژوهش‌های برنامه ریزی روستایی، پژوهشگر در صدد پاسخگویی به سؤال یا سؤال‌های پژوهشی است. به این منظور فرضیه یا فرضیه‌هایی را تدوین، و مورد آزمون قرار می‌دهد. در این حالت، پژوهشگر از نوع خاصی از استنباط آماری به نام «آزمون فرضیه‌های آماری» استفاده می‌کند.

به طور کلی هدف از آزمون فرضیه‌های آماری، رسیدن به این موضوع است که آیا حدس و گمانی که پژوهشگر درباره پارامتر یا پارمترهایی از جامعه زده است، مورد تأیید قرار می‌گیرند یا خیر. از آنجا که این نوع قضاوت از خطا عاری نیست، ممکن است فرضیه‌هایی که در واقع درست است، نادرست بدانیم و برعکس فرضیه‌های نادرست را درست بدانیم. فرضیه‌ای را که مبین وضعیت خاصی درباره پارامتر یا توزیع یک جامعه معین باشد، یک فرضیه آماری گویند. در واقع هر حکمی که از سوی پژوهشگر در مورد پارامترهای جامعه داده شود، یک فرضیه آماری است. آزمون فرضیه قاعده‌ای است که به وسیله آن مشخص می‌شود که آیا نمونه مورد مطالعه از نظر منطقی با فرضیه موردنظر مطابقت دارد یا خیر؟ در واقع این قاعده نوعی دستور تصمیم‌گیری است که تأیید فرضیه را برای پاره‌ای از نمونه‌ها و رد فرضیه را برای پاره‌ای دیگر بیان می‌کند.

البته باید به یاد داشته باشیم که تأیید یا رد یک فرضیه آماری با اثبات یا عدم اثبات یک گزاره ریاضی متفاوت است. یک گزاره ریاضی را اثبات می‌کنند و یا اینکه با ارائه مثال ناقض، آن را رد می‌کنند. در هر حال نتیجه‌ای که به دست می‌آید، بدون هیچ شکی برقرار است؛ اما در مقابل، نتیجهٔ حاصل از آزمون فرضیه آماری از طریق تحلیل داده‌های تجربی، حتمی و قطعی نیست.

چنان که ذکر شد، انجام آزمون فرضیه آماری از فرآیند خاصی تبعیت می‌کند که در این فصل به تفصیل به آن پرداخته می‌شود؛ اما قبل از آن لازم است با مفاهیم و اصطلاحات مرتبط با آموزن فرضیه‌های آماری آشنا شویم.

۲-۱ فرضیه صفر و فرض مقابل

از آنجا که ادعای پژوهشگر ممکن است درست یا نادرست باشد، بنابراین دو فرض مکمل در ذهن ایجاد می‌شود که تحت عنوان فرض صفر و فرض مقابل مطرح می‌شوند.

۲-۱-۱ فرض صفر

در هر آزمون آماری یک فرضیه اولیه وجود دارد که آن را فرضیه صفر می‌گویند و معمولاً با H_0 نشان می‌دهند و به عنوان فرضیه مورد آزمون در نظر گرفته می‌شود. فرض صفر عملاً برای تأیید یا رد مورد آزمون قرار می‌گیرد.

۲-۱-۲ فرض مقابل

در مقابل فرضیه صفر، گروهی از فرضیه‌های مخالف وجود دارند که با H_1 مشخص می‌گردند. بنابراین، این فرضیه عبارتی را در مورد پارامتر یا پارامترهای جامعه که مخالف با فرض صفر(H_0) باشد، در دامنهٔ مقادیر مناسب یک پارامتر ارائه می‌کند.

یعنی اگر $H_{\circ}$ پذیرفته شود، نشان می‌دهد که چه فرضی باید رد شود و برعکس. برای درک بهتر این موضوع، مثال‌های زیر ارائه می‌شود.

با توجه به نکات فوق، می‌توان چنین نتیجه گرفت که هر گاه بخواهیم یک ادعا را از طریق تأیید آن به وسیله اطلاعات حاصل از نمونه آزمون کنیم، نفی آن ادعا را فرض صفر ($H_{\circ}$) و خود آن را فرض مقابل ($H_{۱}$) در نظر می‌گیریم. به این ترتیب، فرض صفر و فرض مقابل بر اساس فرضیه‌های پژوهشی به شرح زیر خواهد بود:

$$\begin{cases} H_{\circ} : \mu \geq ۱ & \text{نقیض ادعا} \\ H_{۱} : \mu < ۱ & \text{ادعا} \end{cases}$$

در اینجا، کمیت نامعلوم، میزان مهاجر فرستی سکونتکاههای کوهستانی است. از دو حکمی که درباره این کمّیت نامعلوم وجود دارند، یکی را فرض صفر ($H_{\circ}$) و دیگری را فرض مقابل ($H_{۱}$) می‌گویند.

مثال یک گروه پژوهشی برنامه ریزی روستایی در شمال خراسان ادعا می‌کند که نسبت مهاجر فرستی سکونتکاههای کوچک کوهستانی بیش از ۷۵ درصد است. با توجه به حجم جامعه آماری، نمونه‌ای ۱۰۰ سکونتگاه روستایی از کل سکونتگاهها انتخاب و از جمعیت روستایی پرسشگری می‌شود.

در فرآیند سنجش نسبت مهاجر فرستی این سکونتگاههای روستایی، فرضیه‌های پژوهشی به شرح زیر تدوین می‌شوند:

- *نسبت مهاجرفرستی حداکثر ۷۵ درصد است.*
- *نسبت مهاجرفرستی بیش از ۷۵ درصد است.*

در چنین شرایطی، فرض صفر و فرض مقابل به شرح زیر خواهد بود:

$$\begin{cases} H_{\circ} : P \leq ۷۵ \\ H_{۱} : P > ۷۵ \end{cases}$$

چنان که مشاهده می‌شود، در مثال مذکور فرض یک یا فرض مقابل (H_1)، ادعا و فرض صفر (H_0) نقیض ادعاست. زیرا آنچه که مدنظر گروه پژوهش است، چیزی بیش از ۷۵ است.

به این ترتیب می‌توان گفت، انجام آزمون یک فرضیه آماری به وسیله نمونه، هنگامی امکان‌پذیر می‌گردد که تابعی از نمونه را به عنوان ملاک یا آماره آزمون انتخاب کنیم که حداقل دارای دو شرط زیر باشد:

۱. توزیع ملاک آزمون یا آماره آزمون انتخاب شده مشخص باشد.
۲. تأثیر غلط بودن فرضیه H_0 روی آماره آزمون معلوم باشد.

برای روشن شدن موضوع، فرض کنید از جامعه‌ای نرمال با میانگین μ و واریانس معلوم σ^2 نمونه‌ای به حجم n انتخاب کنیم و بخواهیم بزرگ بودن میانگین جامعه (μ) را از عدد مشخص μ_0 آزمون کنیم. در این صورت فرض H_0 در مقابل H_1 به صورت زیر خواهد بود.

$H_1: \mu > \mu_0$ در مقابل $H_0: \mu \leq \mu_0$

اگر $\overline{X}$، میانگین یک نمونه به حجم n را به عنوان ملاک آزمون فوق انتخاب کنیم، دو شرط مذکور مراعات گردیده است. اولاً توزیع $\overline{X}$ کاملاً مشخص و دارای توزیع نرمال با میانگین μ و واریانس $\frac{\sigma^2}{n}$ است و ثانیاً تأثیر غلط بودن فرضیه H_0 روی این ملاک (یا $\overline{X}$) در جهت افزایش آن می‌باشد. بدین ترتیب فرضیه H_0 را هنگامی غلط می‌دانیم که $\overline{X}$ بزرگ باشد. در اینجا این سؤال مطرح است که چه مرزی را برای بزرگ بودن $\overline{X}$ قائل شویم که اگر $\overline{X}$ بزرگتر از آن بود، فرضیه H_0 را رد و اگر کوچکتر از آن بود فرضیه H_0 را بپذیریم. قبل از پاسخ به این سؤال خطای نوع اول، خطای نوع دوم، سطح اطمینان و توان آزمون را تعریف می‌کنیم.

۲-۲ خطای نوع اول[۱]

رد فرض H_0 وقتی که H_0 درست است؛ به عبارت دیگر خطای نوع اول وقتی بروز می‌کند که ما به اشتباه فرضیه H_0را رد می‌کنیم؛ در صورتی که درست بوده و باید آن را می‌پذیرفتیم.

۲-۳ خطای نوع دوم[۲]

قبول فرض H_0 وقتی که H_0 نادرست است. یعنی خطای نوع دوم هنگامی رخ می‌دهد که فرضیه H_0 غلط بود و باید رد می‌شد، ولی ما آن را به اشتباه می‌پذیریم.

۲-۴ سطح اطمینان

در بخش قبلی خطای نوع اول (α) توضیح داده شد. "α" که دقت برآورد نامیده می‌شود، مقدار ثابتی است که به کمک آن حد پایین اطمینان[۳] و حد بالای اطمینان[۴] تعریف می‌کنیم. سطح دلخواه در یک توزیع آماری که μ_x در آن قرار می‌گیرد، همان فاصلهٔ اطمینان است که احتمال آن با P نشان داده می‌شود و «سطح اطمینان پژوهشگر» نامیده می‌شود. روشن است که فاصلهٔ اطمینان بر اساس سطح اطمینانی که پژوهشگر را راضی می‌کند، تعریف می‌شود. بنابراین:

اگر $\alpha = P[\text{رد فرض } H_0 \mid H_0 \text{ درست}]$؛ آنگاه:

$$\text{سطح اطمینان} = 1-\alpha = P[\text{قبول فرض } H_0 \mid H_0 \text{ درست}]$$

۲-۵ توان آزمون

در انجام آزمون فرض، مسلماً تمایل داریم از بین روش‌های مختلف آزمون، آن روشی را انتخاب کنیم که با در نظر گرفتن احتمال مشخصی برای مرتکب شدن خطای نوع

---

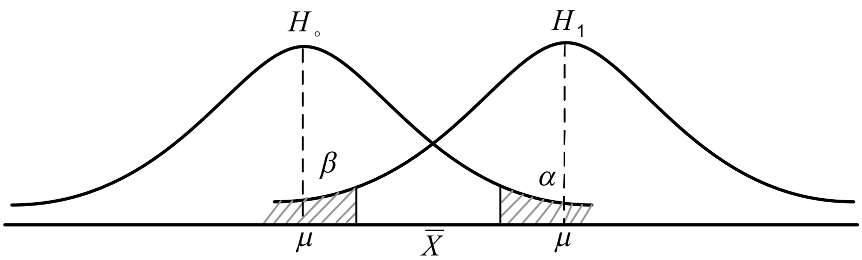
۱. Type one error
۲. Type two error
۳. Lower confidence limit
۴. Upper confidence limit

اول، احتمال ارتکاب خطای نوع دوم در آن کمتر باشد. به این ترتیب آزمونی را می‌توان برتر، توانمندتر و یا قدرتمندتر دانست که با توجه به مقدار مشخصی از α دارای خطای نوع دوم (β) کمتری باشد. توان (قدرت) یک آزمون را توسط کمیت $(1-\beta)$ سنجش می‌کنند. این کمیت نشان‌دهنده احتمال رد کردن فرضیه H_0 است، وقتی که غلط است و باید رد شود. بنابراین:

اگر $\beta = P[$قبول فرض H_0 | نادرست $H_0]$، آنگاه:

توان آزمون $= 1-\beta = P[$رد فرض H_0 | نادرست $H_0]$

در آزمون فرض $H_0: \mu \le \mu_0$ در مقابل $H_1: \mu > \mu_0$ مرز بزرگتر برای بزرگ بودن $\overline{X}$، همان خطای نوع دوم است؛ یعنی قبول H_0 در صورتی که این فرضیه غلط باشد، بیشتر می‌گردد و از طرف دیگر هر قدر این مرز کوچکتر انتخاب شود، احتمال خطای نوع اول، یعنی غلط دانستن فرضیه H_0 در صورتی که این فرضیه صحیح باشد، بیشتر می‌گردد. شکل زیر این وضعیت را نشان می‌دهد.



فرض $H_0: \mu = \mu_0$ یک فرض ساده است. تحت این فرض، توزیع جامعه مشخص است. در واقع فرض ساده، فرضی است که توزیع جامعه را مشخص می‌کند، در حالی که فرض مرکب فرضی است که تحت آن، توزیع جامعه مشخص نیست.

۲-۶ سطح معنی‌داری

معمولاً در آزمون‌های آماری ناحیه رد کردن فرضیه H_0 را بر اساس مقدار معینی که برای α در نظر می‌گیرند، تعیین می‌کنند. این مرز یا ناحیه تعیین شده که مقدار اشتباه نوع اول

است و مایلیم در تصمیم‌گیری برای H_0 تحمل کنیم، سطح معنی‌داری می‌نامند که با α نشان داده می‌شود. مقدار آن بستگی به تشخیص و نظر محقق و هم چنین موضوع مورد بررسی دارد. در بررسی‌های آماری معمولاً α را برابر ۰/۰۵ یا ۰/۰۱ انتخاب می‌کنند.

۲-۷ ناحیه بحرانی

در یک آزمون سطح زیر منحنی چگالی احتمال به دو ناحیه تقسیم می‌شود: ناحیه قبول و ناحیه رد.

ناحیه قبول: ناحیه‌ای است که H_0 رد می‌شود. این بدان معنی است که اگر مقدار آماره آزمون در این ناحیه قرار گیرد، H_0 رد می‌شود.

ناحیه رد: همیشه در انتهای منحنی توزیع واقع است و ممکن است بسته به فرض مخالف در هر دو انتها و یا در یک انتهای توزیع واقع گردد، که در این صورت آزمونهای دو دنباله یا یک دنباله مطرح می‌شوند.

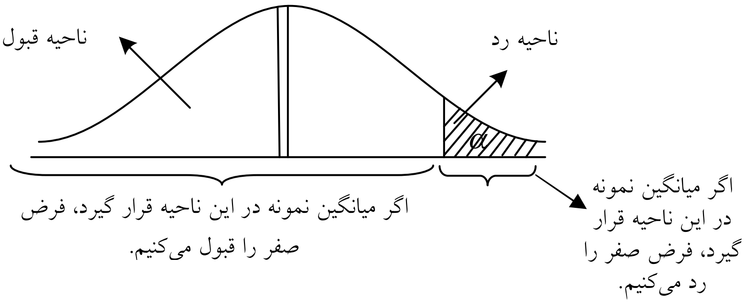
۲-۸ آزمون‌های یک دامنه[۱] و دو دامنه[۲]

یک دنباله یا دو دنباله بودن آزمون فرض آماری، به تعریف H_0 و H_1 بستگی دارد. اگر فرض مخالف (H_1) از نوع $\mu > \mu_0$، $> \mu_2$ یا $\mu_1 < 1$؛ $\sigma^2 > \sigma_0^2$؛ $1 > \sigma_2^2$ یا $1 > \sigma_1^2$ و غیره باشد، ناحیه بحرانی فقط در یک طرف منحنی توزیع احتمال قرار گرفته و آن را آزمون یک دامنه می‌نامند.

اگر H_1 از نوع $\mu > \mu_0$، $\mu_1 > \mu_2$؛ $\sigma^2 > \sigma_0^2$؛ $\sigma_1^2 > \sigma_2^2$ و غیره باشد، ناحیه بحرانی هم چنان که در شکل زیر مشاهده می‌شود، در سمت راست انتهای منحنی قرار می‌گیرد و در این صورت آن را آزمون یک دنباله راست می‌گویند.

---

۱. One tail test
۲. Two tail test



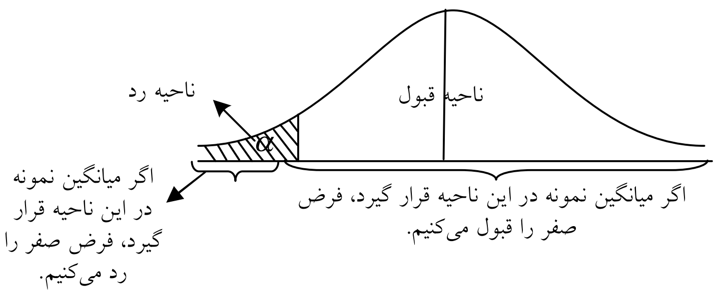
مثال فرضیه‌ای پژوهشی به صورت زیر مطرح شده است: «میانگین مهاجرپذیری "سکونتگاههای دشتی" بیش از ۷۵ است».

این فرضیه طوری به فرضیه‌های آماری تبدیل می‌شود که $H_{\circ}$ نشان‌دهنده نقیض ادعا و $H_{۱}$ نشان‌دهنده ادعای پژوهشگر است. در این حالت فرضیه آماری عبارت است از:

$$\begin{cases} H_{\circ} : \mu \leq ۷۵ & \text{نقیض ادعا} \\ H_{۱} : \mu > ۷۵ & \text{ادعا} \end{cases}$$

واضح است که $H_{۱}$ به اندازه α در دنباله راست توزیع نمونه‌گیری $\overline{x}$ تعریف می‌شود. لذا توزیع از نوع یک دنباله راست است.

در حالی که اگر $H_{۱}$ از نوع $\mu < \mu_{\circ}$، $\mu_{۱} < \mu_{۲}$؛ $\sigma^{۲} < \sigma^{۲}_{\circ}$؛ $\sigma^{۲}_{۱} < \sigma^{۲}_{۲}$ و غیره باشد، ناحیه بحرانی چنان که در شکل زیر مشاهده می‌شود، در سمت چپ انتهای منحنی قرار می‌گیرد و در این صورت آن را آزمون یک دنباله چپ می‌گویند.



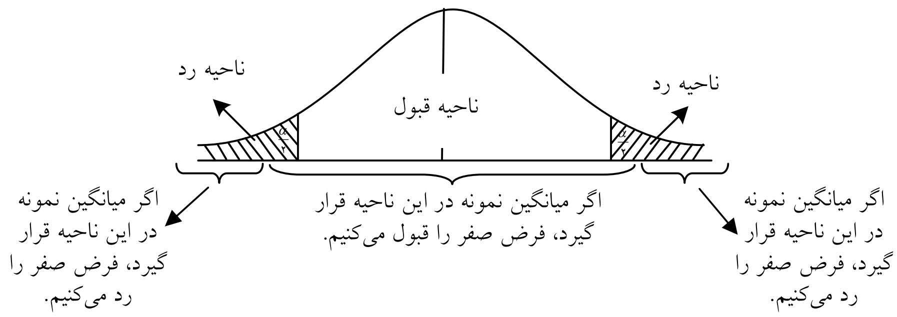
مثال اگر فرضیه‌ای پژوهشی مطرح شده قبلی به این صورت باشد: «میانگین مهاجر پذیری"سکونتگاههای دشتی" کمتر از ۷۵ است».

این فرضیه نیز طوری به فرضیه‌های آماری تبدیل می‌شود که H_0 نشان‌دهنده نقیض ادعا و H_1 نشان‌دهنده ادعا است. لذا فرضیه آماری به شرح زیر خواهد بود:

$$\begin{cases} H_0 : \mu \geq 75 & \text{نقیض ادعا} \\ H_1 : \mu < 75 & \text{ادعا} \end{cases}$$

واضح است که H_1 به اندازه α در دنبالهٔ چپ توزیع نمونه‌گیری $\overline{x}$ تعریف می‌شود. لذا توزیع از نوع یک دنبالهٔ چپ است.

اما اگر H_1 از نوع $\mu_1 \neq 0$، $\mu_1 \neq \mu_2$؛ $\sigma^2 \neq \sigma_0^2$؛ $\sigma_1^2 \neq \sigma_2^2$ و غیره باشد، ناحیه بحرانی چنان که در شکل زیر مشاهده می‌شود، در دو طرف انتهای منحنی قرار می‌گیرد. در این وضعیت، آزمون یک آزمون دو دامنه نامیده می‌شود. در آزمون دو دامنه، سطح مساوی $\frac{\alpha}{2}$ در دو سمت انتهای منحنی برای آزمون معنی‌دار بودن در سطح α قرار می‌گیرد.



مثال اگر فرضیه‌ای پژوهشی مطرح شدهٔ قبلی به این صورت باشد: «میانگین مهاجر پذیری"سکونتگاههای دشتی" برابر از ۷۵ است».

این فرضیه طوری به فرضیه‌های آماری تبدیل می‌شود که $H_\circ$ نشان‌دهنده ادعا و $H_\text{۱}$ نشان‌دهنده نقیض ادعا خواهد بود. از این‌رو فرضیه آماری به شرح زیر نوشته می‌شود:

$$\begin{cases} H_\circ : \mu = \text{۷۵} & \text{ادعا} \\ H_\text{۱} : \mu \neq \text{۷۵} & \text{نقیض ادعا} \end{cases}$$

واضح است که $H_\text{۱}$ به اندازهٔ $\frac{\alpha}{\text{۲}}$ در دنبالهٔ چپ و به اندازهٔ $\frac{\alpha}{\text{۲}}$ در دنباله راست توزیع نمونه‌گیری $\overline{x}$ تعریف می‌شود. لذا توزیع از نوع دو دنباله است.

۲-۹ مراحل آزمون فرضیه‌های آماری

برای انجام یک آزمون فرضیه می‌توان مراحل زیر را طی کرد:

۱. نوشتن فرض‌های آماری: بر اساس فرضیه‌های پژوهشی بیان شده، ابتدا باید فرض‌های آماری مناسب که بیانگر ادعا و نقیض ادعای پژوهشگر می‌باشند، نوشته شوند. بر اساس قواعد بیان شده، فرضیه‌های آماری برای میانگین جامعه، صرف‌نظر از فرضیه‌های پژوهشی یکی از تعاریف زیر را خواهد داشت:

$$\begin{cases} H_\circ : \mu \geq \mu_\circ \\ H_\text{۱} : \mu < \mu_\circ \end{cases} \qquad \begin{cases} H_\circ : \mu = \mu_\circ \\ H_\text{۱} : \mu \neq \mu_\circ \end{cases} \qquad \begin{cases} H_\circ : \mu = \mu_\circ \\ H_\text{۱} : \mu \neq \mu_\circ \end{cases}$$

۲. تعیین آماره آزمون و محاسبهٔ مقدار آن: بر اساس شرایط برآورد آماری، توزیع $\overline{x}$ که یکی از توزیع‌های Z یا t خواهد بود، تعیین می‌شود.

در حالتی که نمونه از جامعه نرمال با انحراف معیار معلوم انتخاب شود، آماره آزمون عبارت خواهد بود از:

$$Z_\circ = \frac{\overline{X} - \mu_\circ}{\frac{\sigma}{\sqrt{n}}}$$

هر گاه نمونه از جامعه‌ای نرمال با انحراف معیار نامعلوم انتخاب شود، اما حجم نمونه کوچک باشد، آماره عبارت است از:

$$t = \frac{\overline{X} - \mu_{\circ}}{\frac{S_x}{\sqrt{n}}}$$

در صورتی که حجم نمونه بزرگ باشد، توزیع بر اساس قضیه حد مرکزی از تقریب نرمال برخوردار است و آمارهٔ آزمون عبارت است از:

$$t = \frac{\overline{X} - \mu_{\circ}}{\frac{S_x}{\sqrt{n}}}$$

با انتخاب آماره آزمون مناسب، مقدار آماره آزمون تحت فرض $H_{\circ}$ و مشاهدات محاسبه می‌شود.

۳. **مقدار بحرانی محاسبه می‌شود:** در این مرحله پس از تعیین مقدار α و سطح اطمینان، آزمون از نظر یک دنباله یا دو دنباله بودن مشخص می‌شود. سپس مقدار بحرانی (یعنی مرز $H_{\circ}$ و $H_{۱}$) بر اساس α و نوع آماره آزمون تعیین می‌شود.

۴. **تصمیم‌گیری:** در مرحله تصمیم‌گیری، مقدار آماره آزمون محاسبه شده با مقدار بحرانی مقایسه می‌شود. اگر آماره آزمون در ناحیه پذیرش $H_{\circ}$ قرار گیرد، فرض $H_{\circ}$ در سطح اطمینان $(۱-\alpha)$ درصد پذیرفته می‌شود؛ در غیر این صورت داده‌های نمونه دلیل محکمی بر تأیید $H_{\circ}$ ارائه نداده و آن را رد می‌کنند.

۲-۱۰ آزمون فرض برای میانگین توزیع نرمال

از آنجا که توزیع جامعه نرمال دارای دو پارامتر به ترتیب میانگین μ و واریانس $\sigma^{۲}$ است، انواع فرض‌ها درباره μ را می‌توان به صورت‌های زیر نوشت:

$$۱)\left|\begin{array}{l} H_{\circ}: \mu = \mu_{\circ} \\ H_{۱}: \mu \neq \mu_{\circ} \end{array}\right. \qquad ۲)\left|\begin{array}{l} H_{\circ}: \mu \geq \mu_{\circ} \\ H_{۱}: \mu < \mu_{\circ} \end{array}\right. \qquad ۳)\left|\begin{array}{l} H_{\circ}: \mu \leq \mu_{\circ} \\ H_{۱}: \mu > \mu_{\circ} \end{array}\right.$$

$\mu_{\circ}$ یک عدد است که μ مورد آزمون با آن مقایسه می‌شود.

در فرض‌های آماری نوشته شده، فرض آماری ۱، یک فرض دو طرفه است. در حالی فرض‌های آماری ۲ و ۳ فرض‌های یک‌طرفه می‌باشند. برای آزمون فرض درباره μ دو حالت زیر را در نظر می‌گیریم:

الف- وقتی واریانس جامعه (σ^2) معلوم است.

فرض کنید X_1، X_2،...، X_n یک نمونه n تایی از توزیع نرمال با میانگین مجهول μ و واریانس معلوم σ^2 باشند. هدف آزمون فرض زیر درباره میانگین جامعه است، لذا داریم:

$$H_{\circ} : \mu \le \mu_{\circ} \quad \text{در مقابل} \quad H_1 : \mu > \mu_{\circ}$$

فرض $H_{\circ}$ یک فرض ساده در مقابل فرض H_1 که یک فرض مرکب یک‌طرفه راست است، قرار دارد. در بحث برآوردها، مشخص شده که میانگین نمونه $\bar{X}$ یک برآورد نقطه‌ای برای μ است؛ لذا هرگونه تصمیم‌گیری در مورد فرض فوق بر اساس $\bar{X}$ خواهد بود.

مقادیر بزرگ $\bar{X}$ پشتیبانی برای فرض H_1 یا تأییدی برای H_1 است و باعث رد فرض $H_{\circ}$ می‌شود. رد فرض $H_{\circ}$ را می‌توان به صورت‌های زیر نیز نوشت:

رد $H_{\circ}$ هم ارز است با: $\bar{X} > k_{\circ}$

رد $H_{\circ}$ هم ارز است با: $\bar{X} - \mu_{\circ} > k_1$

رد $H_{\circ}$ هم ارز است با: $\dfrac{\bar{X} - \mu_{\circ}}{\dfrac{\sigma}{\sqrt{n}}} > k_2$

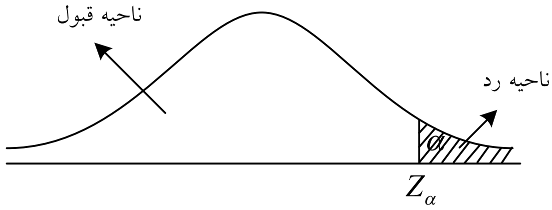
لازم به ذکر است که $k_{\circ}$، k_1 و k_2 اعداد ثابت هستند.

اما $\dfrac{\bar{X} - \mu_{\circ}}{\dfrac{\sigma}{\sqrt{n}}}$ یک کمیت محوری است و تحت فرض $H_{\circ}$ یک آماره است و از آن به عنوان آماره آزمون یاد می‌کنند.

اگر α سطح معنی‌دار آزمون باشد، با توجه به توزیع آماره آزمون داریم:

$$\alpha = P\left[\frac{\overline{X}-\mu_{\circ}}{\frac{\sigma}{\sqrt{n}}} > Z_{\alpha} \quad \text{درست} \quad H_{\circ}\right]$$

چون $\frac{\overline{X}-\mu_{\circ}}{\frac{\sigma}{\sqrt{n}}}$ تحت فرض $H_{\circ}$ دارای توزیع نرمال استاندارد است، ناحیه بحرانی و ناحیه قبول برای آزمون فرض $H_{\circ}: \mu = \mu_{\circ}$ در مقابل $H_1: \mu > \mu_{\circ}$ به صورت زیر خواهد بود:



مقدار Z_{α} با توجه به مقدار α، از جدول نرمال استاندارد محاسبه می‌شود. برای تصمیم‌گیری در مورد فرض $H_{\circ}: \mu \leq \mu_{\circ}$ در مقابل $H_1: \mu > \mu_{\circ}$، مراحل فرآیند آزمون فرض را انجام می‌دهیم.

مثال مدیریت جهاد کشاورزی شهرستان جوین در خراسان رضوی ادعا می‌کند که متوسط تولید گندم در سکونتگاههای روستایی در سال‌های گذشته بیش از ۷۵ هزار کیلو گرم است. اگر متوسط تولید ۲۵ سکونتگاه روستایی از شهرستان که به تصادف انتخاب شده‌اند برابر با $\overline{x} = ۷۶$ و انحراف معیار آن ۱۰ و نیز توزیع نمرات نرمال باشد، ادعای مدیر را با اطمینان ۹۰٪ آزمون کنید.

حل برای آزمون فرضیه این مسأله، فرآیند زیر طی می‌شود:

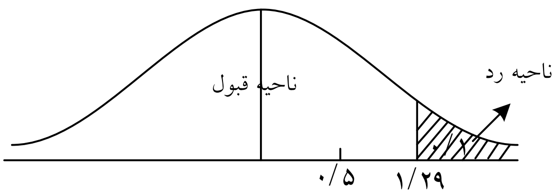
گام ۱) نوشتن فرضیه صفر و فرضیه مقابل

$H_{\circ}: \mu \leq ۷۵$ در مقابل $H_1: \mu > 75$

گام ۲) تعیین آماره آزمون

$$Z_{\circ} = \frac{\bar{x} - \mu_{\circ}}{\frac{\sigma}{\sqrt{n}}} = \frac{76 - 75}{\frac{10}{\sqrt{25}}} = 0/5$$

گام ۳) مشخص کردن مقادیر بحرانی و تقسیم توزیع به ناحیه قبول و ناحیه رد



گام ۴) تصمیم‌گیری

با توجه به مقدار $\alpha = ۰/۱$، مقدار جدول برابر با ۱/۲۹ است که در مقایسه با مقدار آماره بیشتر است. لذا فرض $H_{\circ}$ پذیرفته می‌شود.

مراحل انجام آزمون فرض $H_{\circ}: \mu \geq \mu_{\circ}$ در مقابل $H_1: \mu < \mu_{\circ}$ همانند مثال اخیر است. فقط ناحیه رد یا بحرانی به جای اینکه در سمت راست باشد در سمت چپ آن قرار می‌گیرد.

همچنین مراحل انجام آزمون فرض ساده $H_{\circ}: \mu = \mu_{\circ}$ در مقابل فرض مرکب دوطرفه $H_1: \mu \neq \mu_{\circ}$ به صورت زیر است.

در آزمون $H_{\circ}: \mu = \mu_{\circ}$ مقادیر بزرگ یا کوچک $\bar{X}$، پشتیبانی برای H_1 است و یا ممکن است باعث رد $H_{\circ}$ شود. بنابراین:

رد $H_{\circ}$ هم ارز است با : $\bar{X} < K'_{\circ}$ یا $\bar{X} > K_{\circ}$

رد $H_{\circ}$ هم ارز است با : $\bar{X} - \mu_{\circ} < K'_1$ یا $\bar{X} - \mu_{\circ} > K_1$

رد $H_{\circ}$ هم ارز است با: $\frac{\bar{X} - \mu_{\circ}}{\frac{\sigma}{\sqrt{n}}} < K'_2$ یا $\frac{\bar{X} - \mu_{\circ}}{\frac{\sigma}{\sqrt{n}}} < K_2$

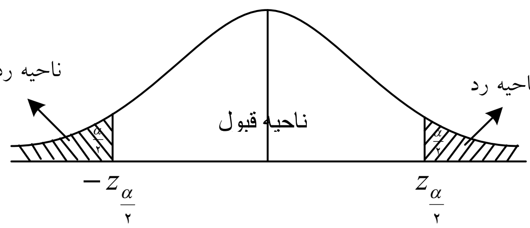
لازم به ذکر است که $K_{\circ}$، $K'_{\circ}$، K_1، K'_1، K_2 و K'_2 اعداد ثابت هستند.

$\dfrac{\bar{X}-\mu_{\circ}}{\frac{\sigma}{\sqrt{n}}}$ تحت فرض $H_{\circ}$ آماره آزمون است و برای αی معین خواهیم داشت:

$$P\left[\dfrac{\bar{X}-\mu_{\circ}}{\frac{\sigma}{\sqrt{n}}} > z_{\alpha/2} \mid H_{\circ} \quad \text{درست}\right] = \alpha/2$$

$$P\left[\dfrac{\bar{X}-\mu_{\circ}}{\frac{\sigma}{\sqrt{n}}} < -z_{\alpha/2} \mid H \quad \text{درست}\right] = \alpha/2$$

ناحیه بحرانی یا رد برای آزمون $H_{\circ}$ در مقابل H_1 به صورت زیر خواهد بود:



مقادیر $z_{\frac{\alpha}{2}}$ و $-z_{\frac{\alpha}{2}}$ که مقادیر قرینه هستند از جدول نرمال استاندارد بدست می‌آیند.

برای تصمیم‌گیری در مورد آزمون فرض $H_{\circ}$ در مقابل H_1 گام‌های چهارگانه مذکور را انجام می‌دهیم؛ با این تفاوت که در مرحله تصمیم‌گیری اگر بازه ($-z_{\frac{\alpha}{2}}$ و $z_{\frac{\alpha}{2}}$) عدد $Z_{\circ}$ را دربرداشته باشد فرض $H_{\circ}$ را می‌پذیریم در غیر اینصورت آن را رد می‌کنیم.

مثال مدیریت جهاد کشاورزی استان اردبیل ادعا می‌کند میانگین میزان عملکرد مشارکت پذیری روستاییان در جاده سازی برابر ۳۵ درصداست. این مدیریت برای بررسی فرضیه فوق یک نمونه ۱۰۰ نفرروستایی از جمعیت روستایی را انتخاب می‌کند که میانگین نمره مشارکت آنان ۲۹ و انحراف معیارشان ۳ می‌باشد. بر اساس اطلاعات حاصل از این نمونه، آزمون آماری را در سطح معنی‌دار بودن $\alpha = ۰/۱$ انجام دهید.

$$\begin{cases} H_{\circ} : \mu = ۳۵ \\ H_{۱} : \mu \neq ۳۵ \end{cases}$$

چنانکه ملاحظه می‌شود اندازه نمونه بزرگ است $(n > ۳۰)$ و با توجه به فرض‌های فوق، آزمون از نوع دو طرفه است. لذا مقدار آماره آزمون عبارت است از:

$$Z = \frac{\overline{X} - \mu_{\circ}}{\sigma_{\overline{X}}} = \frac{۲۹ - ۳۵}{\frac{S}{\sqrt{n}}} = \frac{۲۹ - ۳۵}{\frac{۳}{\sqrt{۱۰۰}}} = \frac{-۶}{۰/۳} = -۲۰$$

از مقایسه $Z = -۲۰$ با $Z_{۰/۰۵} = -۱/۶۴۵$ فرض $H_{\circ}$ رد می‌شود.

مثال کارکنان واحد توزیع نهاده های تولید در شرکت تعاونی"مرودشت" در توجیه درخواست اضافه حقوق خود ادعا می‌کنند که به طور متوسط طی ۱۳ دقیقه، نهاده های مورد نیازیک روستا را ارسال می‌کنند. مدیر این شرکت ادعای خلاف آنان دارد. وی در جهت بررسی صحت این ادعا یک نمونه ۴۰۰تایی انتخاب می‌کند که میانگین تکمیل ارسال نهادهای تولیدی یک روستا ۱۴ دقیقه با انحراف معیار ۱۰ دقیقه بوده است. با توجه به این اطلاعات آزمون مربوطه را در سطح معنی‌داری $\alpha = ۰/۰۵$ انجام دهید.

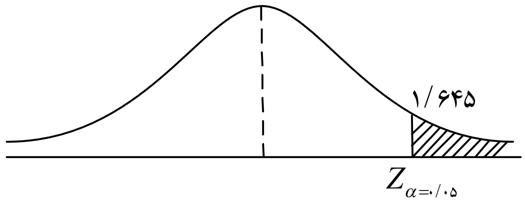
حل

$$\begin{cases} H_{\circ} : \mu \leq ۱۳ \\ H_{۱} : \mu > ۱۳ \end{cases}$$

آماره آزمون است که توزیع $\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$ دارد. اما چون حجم نمونه به اندازه کافی بزرگ است، توزیع این آماره به طرف نرمال استاندارد می‌رود با میانگین صفر و واریانس ۱ لذا مقدار آماره را محاسبه می‌کنیم.

$$\frac{۱۴-۱۳}{\frac{۱۰}{\sqrt{۴۰۰}}}=\frac{۱}{\frac{۱۰}{۲۰}}=\frac{۲۰}{۱۰}=۲$$

$$Z_{\alpha=۰/۰۵}=۱/۶۴۵$$



چنانکه مشاهده می‌شود Z محاسبه شده (یعنی آماره آزمون) بزرگتر از ۱/۶۴۵ است و در ناحیه رد آزمون قرار می‌گیرد. لذا فرض $H_۰$ رد شده و نتیجه گرفته می‌شود که میانگین واقعی اتمام ارسال یک محموله بیشتر از ۱۳ دقیقه می‌باشد.

مثال شرکت تعاونی تولیدی بسته بندی فرآورده های کشاورزی رودبار "زیتون" محصولات زیتون خود را در بسته‌های ۵۰۰ گرمی به مشتریان عرضه می‌کند. اما مؤسسه استاندارد خلاف این ادعا را مطرح می‌کند و معتقد است که وزن محصولات این شرکت کمتر است. این بازرس برای بررسی ادعای خود ۹ بسته از محصولات این شرکت را انتخاب می‌کند. میانگین بسته‌ها ۴۸۰ گرم و انحراف معیار آنها ۱۲ گرم حاصل شده است بر اساس اطلاعات بدست آمده وضعیت وزن بسته بندی محصولات این شرکت تعاونی چگونه است؟ آزمون را در سطح معنی‌دار ۰/۰۵ انجام دهید.

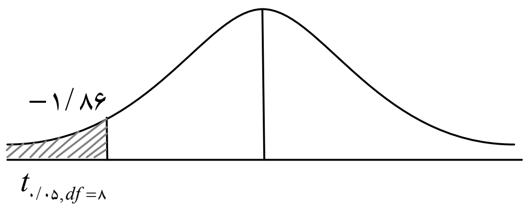
حل برای تحلیل آماری این موضوع از روش آزمون چون از نوع یک دامنه است. بنابراین خواهیم داشت:

$$\begin{cases} H_{\circ} : \mu \geq ۵۰۰ \\ H_{۱} : \mu < ۵۰۰ \end{cases}$$

آماره آزمون $\dfrac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$ است که توزیع t دارد. یعنی:

$$\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} = \frac{۴۸۰-۵۰۰}{\frac{۱۲}{\sqrt{۹}}} = \frac{-۲۰}{\frac{۱۲}{۳}} = \frac{-۲۰}{۴} = -۵$$

$$t_{\alpha=۰/۰۵, df=n} = -۱/۸۶$$



چنانکه ملاحظه می‌شود مقدار آماره آزمون در ناحیه رد قرار می‌گیرد. لذا فرض $H_{\circ}$ رد می‌شود. بنابراین براساس نمونه بدست آمده می‌توان بیان داشت وزن محصولات این شرکت تعاونی کمتر از مقدار تعیین شده است.

ب- وقتی واریانس جامعه مجهول است

در مفروضات و مثال‌های اخیر واریانس جامعه، معلوم فرض شده بود. در حالتی که واریانس جامعه نرمال، مجهول باشد، فرض‌های آماری زیر را چگونه باید مورد آزمون قرار دهیم؟

برای آزمون فرض‌های فوق یک نمونه تصادفی از توزیع نرمال با میانگین μ و واریانس مجهول $\sigma^{۲}$ می‌گیریم. چون طبیعت جامعه نرمال است و در جامعه نرمال

واریانس نمونه یک برآورد نااریب برای واریانس جامعه است، لذا واریانس نمونه یک برآورد برای σ^2 است، یعنی:

$$\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

در این حالت آماره آزمون برابر است با $T = \dfrac{\overline{X}-\mu}{\dfrac{S}{\sqrt{n}}}$ که از توزیع استودنت با $n-1$ درجه آزادی پیروی می‌کند.

بنابراین برای آزمون فرض‌ها مراحل زیر را (که در حالت قبل نیز بیان شد) انجام می‌دهیم:

۱- تعریف فرض‌های آماری که یکی از فرض‌های سه‌گانه زیر خواهد بود:

$$1)\left|\begin{array}{l} H_\circ : \mu = \mu_\circ \\ H_1 : \mu > \mu_\circ \end{array}\right. \qquad 2)\left|\begin{array}{l} H_\circ : \mu = \mu_\circ \\ H_1 : \mu < \mu_\circ \end{array}\right. \qquad 3)\left|\begin{array}{l} H_\circ : \mu = \mu_\circ \\ H_1 : \mu \neq \mu_\circ \end{array}\right.$$

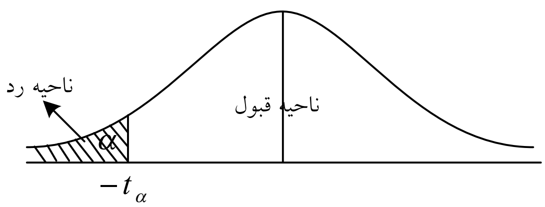
۲- مقدار آماره آزمون را تحت فرض $H_\circ$ و مشاهدات محاسبه می‌کنیم.

$$T_\circ = \frac{\bar{x}-\mu_\circ}{\dfrac{S}{\sqrt{n}}}$$

۳- مقدار T_α را از جدول استودنت با $n-1$ درجه آزادی بدست می‌آوریم.

$$P\left[T < t_\alpha\right] = \alpha$$

۴- تصمیم‌گیری می‌کنیم. یعنی اگر مقدار $T_\circ$ از t_α کوچکتر باشد، فرض $H_\circ$ را رد می‌کنیم در غیر اینصورت آن را می‌پذیریم.



مثال آزمون کارکردهای آبیاری۶ مزرعه کشاورزی درناحیه شبکه آبیاری تحت فشار بیله سوار عملکرد آنها به صورت ۲۴، ۲۸، ۲۱، ۲۳، ۳۲، ۲۲ دقیقه با مصرف یک گالن از نوع معینی سوخت بوده است. آیا در سطح خطای ۵ درصد می‌توان گفت که متوسط کارکرد این نوع موتور حداکثر برای هر گالن با این نوع سوخت ۲۹ دقیقه می‌باشد؟

حل برای حل این مسأله مراحل چهارگانه مذکور را طی می‌کنیم.

۱- فرض‌های آماری را تعریف می‌کنیم.

$$\begin{cases} H_{\circ} : \mu \leq ۲۹ \\ H_{۱} : \mu > ۲۹ \end{cases}$$

۲- آماره آزمون را محاسبه می‌کنیم. به این منظور ابتدا $\bar{x}$ و S^2 را محاسبه می‌کنیم.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 25 \qquad S^2 = \frac{1}{n-1}\sum\left(x_i - \bar{x}\right)^2 = 17/6$$

لذا خواهیم داشت:

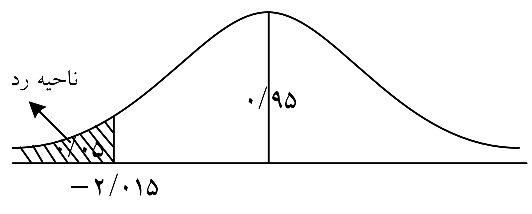
$$T_{\circ} = \frac{\bar{x} - \mu_{\circ}}{\frac{S}{\sqrt{n}}} = \frac{25-29}{\frac{4/19}{\sqrt{6}}} = -2/34$$

۳- مقدار $\alpha = ۰/۰۵$ را با $d.f = ۶ - ۱ = ۵$ بدست می‌اوریم. یعنی:

$$P\left[T < t_{\alpha}\right] = \alpha$$

$$P\left[T < t_{\alpha}\right] = ۰/۰۵ \Rightarrow t_{\alpha} = -۲/۰۱۵$$

۴- تصمیم‌گیری. چون $T_{\circ} < t_{\alpha}$ است، لذا فرض $H_{\circ}$ رد می‌شود.



مثال میزان اندازه گیری پنج حلقه چاه عمیق در دشت جوین ، منجر به ۱۴/۵، ۱۴/۲، ۱۴/۴، ۱۴/۳ و ۱۴/۶ لیتر برپانیه در هر حلقه شده است. فرض $H_\circ : \mu = 14$ را در مقابل $H_1 : \mu \neq 14$ آزمون کنید، اگر $\alpha = 0/05$ باشد.

حل

۱- فرض های آماری این مسأله داده شده است. یعنی:

$$n = 5 \quad , \quad \bar{x} = 14/4 \quad , \quad S^2 = 0/025 \quad , \quad \alpha = 0/05 \quad , \quad \frac{\alpha}{2} = 0/025$$

$$\begin{cases} H_\circ : \mu = ۱۴ \\ H_۱ : \mu \neq ۱۴ \end{cases}$$

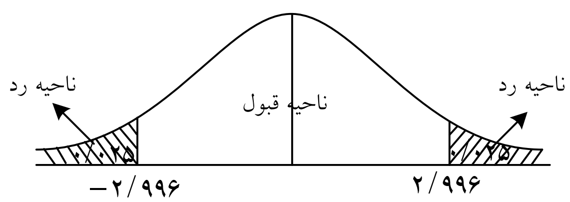
۲- مقدار آماره آزمون را محاسبه می‌کنیم.

$$T_\circ = \frac{\bar{x} - \mu_\circ}{\frac{S}{\sqrt{n}}} = \frac{14/4 - 14}{\frac{0/158}{\sqrt{5}}} = 5/66$$

۳- مقدار $d.f = n - ۱$ و $t_{\frac{\alpha}{۲}}$ را از جدول t استخراج می‌کنیم.

$$P\left[T < -t_{\frac{\alpha}{۲}}\right] = ۰/۰۲۵ \Rightarrow -t_{\frac{\alpha}{۲}} = -۲/۹۹۶$$

$$P\left[T > t_{\frac{\alpha}{۲}}\right] = ۰/۰۲۵ \Rightarrow t_{\frac{\alpha}{۲}} = ۲/۹۹۶$$



۴- تصمیم‌گیری می‌کنیم. چون $\left(-t_{\frac{\alpha}{2}}, t_{\frac{\alpha}{2}}\right)$ عدد $T_{\circ} = 5/66$ را در بر ندارد فرض $H_{\circ}$ رد می‌شود.

۲-۱۱ آزمون فرض‌های دو طرفه با استفاده از فاصله اطمینان

در مواقع آزمون فرض ساده در مقابل فرض مرکب دو طرفه، می‌توان با استفاده از فاصله اطمینان بدست آمده، برای پارامتر مفروض جامعه نسبت به قبول یا رد فرض $H_{\circ}$ اقدام کرد. روش آزمون در حالت خاص برای مثال اخیر ارائه می‌شود.

در مثال اخیر، نواحی رد یا بحرانی از دو رابطه زیر بدست می‌آید.

$$P\left[T < -t_{\frac{\alpha}{2}}\right] = \frac{\alpha}{2} \qquad\qquad P\left[T > t_{\frac{\alpha}{2}}\right] = \frac{\alpha}{2}$$

با استفاده از دو رابطه اخیر، احتمال اینکه متغیر T تحت $H_{\circ}$ در ناحیه قبول قرار گیرد برابر است با:

$$P\left[-t_{\alpha/2} < T < t_{\alpha/2}\right] = 1-\alpha$$

اما

$$T = \frac{\bar{X} - \mu_{\circ}}{\frac{S}{\sqrt{n}}}$$

پس

$$P\left[-t_{\alpha/2} < \frac{\bar{X} - \mu_{\circ}}{\frac{s}{\sqrt{n}}} < t_{\alpha/2}\right] = 1-\alpha$$

اگر پیشامد احتمال اخیر را نسبت به $\mu_{\circ}$ حل کنیم، داریم:

$$\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu_{\circ} < \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}$$

رابطه اخیر ناحیه اطمینان دو طرفه برای $\mu_{\circ}$ تحت فرض $H_{\circ}$ است. برای آزمون فرض $H_{\circ}: \mu = \mu_{\circ}$ در مقابل $H_1: \mu \neq \mu_{\circ}$ با استفاده از فاصله اطمینان، ابتدا فاصله اطمینان را برای $\mu_{\circ}$ بدست می‌آوریم. اگر $\mu = \mu_{\circ}$ تحت فرض $H_{\circ}$ در این فاصله قرار گرفت، فرض $H_{\circ}$ پذیرفته می‌شود؛ در غیر اینصورت $H_{\circ}$ رد می‌شود. در مثال اخیر، فاصله اطمینان برای $\mu_{\circ}$ برابر است با:

$$\bar{x} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{S}{\sqrt{n}}$$

$$14/4 - 2/776\frac{0/158}{\sqrt{5}} < \mu < 14/4 + 2/776\frac{0/158}{\sqrt{5}}$$

$$14/204 < \mu < 14/596$$

چون تحت فرض $H_{\circ}$، $\mu = \mu_{\circ} = 14$ در این فاصله قرار دارد، لذا فرض $H_{\circ}$ پذیرفته می‌شود.

۲-۱۲ آزمون فرض برای تفاضل میانگین دو جامعه نرمال مستقل

در تحقیقات کاربردی، مسائل متعددی موجودند که در آنها فرض‌هایی درباره تفاضل بین میانگین‌های دو جامعه مورد توجه است. برای مثال ممکن است بخواهیم متوسط درآمد سرانه یا میزان متوسط تولید سرانه کشاورزی را که توسط دو تولید کننده روستایی تولید می‌شود، با هم مقایسه کنیم. آزمون فرض برابری میانگین دو جامعه را در دو حالت زیر مورد بررسی قرار می‌دهیم.

الف- واریانس‌های دوجامعه معلوم‌اند

گاهی اوقات ممکن است یک مدیر توسعه روستایی بخواهد متوسط نمرات کارایی توسعه کشاورزی را در سکونتگاههای روستایی، متوسط مواد مصرفی نهاده های تولید از نوع A و B، متوسط راندمان تولید و متوسط زمان بهره برداری تولید از A و B را با هم مقایسه کند. برای انجام این مقایسه، اگر طبیعت دو جامعه نرمال و مستقل از

هم باشند، و به ترتیب دارای میانگین‌های μ_1 و μ_2 و واریانس σ_1^2 و σ_2^2 باشند، انواع فرض‌ها در مورد مقایسه میانگین‌های دو جامعه عبارتند از:

$$1)\left|\begin{array}{l} H_\circ : \mu_1 = \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{array}\right. \qquad 2)\left|\begin{array}{l} H_\circ : \mu_1 = \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{array}\right. \qquad 3)\left|\begin{array}{l} H_\circ : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{array}\right.$$

یا

$$1)\left|\begin{array}{l} H_\circ : \mu_1 - \mu_2 = \circ \\ H_1 : \mu_1 - \mu_2 > \circ \end{array}\right. \qquad 2)\left|\begin{array}{l} H_\circ : \mu_1 - \mu_2 = \circ \\ H_1 : \mu_1 - \mu_2 < \circ \end{array}\right. \qquad 3)\left|\begin{array}{l} H_\circ : \mu_1 - \mu_2 = \circ \\ H_1 : \mu_1 - \mu_2 \neq \circ \end{array}\right.$$

آماره آزمون برای مقایسه تفاضل $\mu_1 - \mu_2$ بر اساس نمونه تصادفی از جامعه اول و دوم خواهد بود. اگر $\bar{X}_2$ $\bar{X}_1$, به ترتیب میانگین نمونه اول و دوم باشند، کمیت محوری همان کمیت محوری بحث شده در فصل قبل خواهد بود.

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$$

که تحت $H_\circ$ دارای توزیع نرمال استاندارد است. $Z_\circ$ مشاهده شده بر اساس مشاهده نمونه از دو جامعه برابر است با:

$$Z_\circ = \frac{\bar{x}_1 - \bar{x}_2 - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$$

روش آزمون فرض $H_\circ$ در مقابل H_1، همانند آزمون برای میانگین توزیع نرمال با واریانس معلوم است. برای آزمون فرض $H_\circ : \mu_1 = \mu_2$ در مقابل $H_1 : \mu_1 > \mu_2$ مراحل زیر را انجام می‌دهیم.

۱- آماره آزمون را تحت فرض $H_\circ$ و مشاهدات محاسبه می‌کنیم ($Z_\circ$).

۲- مقدار Z_α را از رابطه $P(Z < Z_\alpha) = \alpha$ بدست می‌آوریم.

۳- اگر $Z_\circ > Z_\alpha$ باشد فرض $H_\circ$ را رد می‌کنیم.

مثال مدیر توسعه روستایی محلی (برای مثال دهیار روستایی) ادعا می‌کند تفاضل متوسط راندمان تولید از نوع A و B کمتر از ۰/۰۵ است. برای بررسی این ادعا، از هر نوع متوسط راندمان تولیدی ۳۲ نمونه انتخاب و متوسط نمونه‌ها به ترتیب $\bar{x}_1 = 0/136$ و $\bar{x}_2 = 0/083$ باشند، در سطح ۵ بدست آمده است. اگر $\sigma_1 = 0/004$ و $\sigma_2 = 0/005$ باشند، در سطح ۵ درصد فرض $H_\circ: \mu_1 - \mu_2 = 0/05$ را در مقابل $H_1: \mu_1 - \mu_2 > 0/05$ آزمون کنید.

حل پس از تعریف فرض‌های آماری،آماره آزمون را محاسبه می‌کنیم.

$$m = n = ۳۲ \quad , \quad \bar{x}_۱ = ۰/۱۳۶ \quad , \quad \bar{x}_۲ = ۰/۰۸۳ \quad , \quad \sigma_۱ = ۰/۰۰۴ \quad , \quad \sigma_۲ = ۰/۰۰۵$$

$$\alpha = 0/05 \quad , \quad \mu_1 - \mu_2 = 0/05$$

$$Z_\circ = \frac{(\bar{x}_۱ - \bar{x}_۲) - (\mu_۱ - \mu_۲)}{\sqrt{\frac{\sigma_۱^۲}{m} + \frac{\sigma_۲^۲}{n}}} = \frac{(۰/۱۳۶ - ۰/۰۸۳) - ۰/۰۵}{\sqrt{\frac{۰/۰۰۰۰۱۶}{۳۲} + \frac{۰/۰۰۰۰۲۵}{۳۲}}} = ۲/۶۵$$

حال مقدارها را از جدول مربوطه به دست می‌آوریم.

$$P[Z < Z_\alpha] = 0/05 \Rightarrow Z_\alpha = -1/645$$

چون $Z_\circ > Z_\alpha$، پس فرض $H_\circ$ پذیرفته می‌شود.

ب- واریانس‌های دو جامعه مجهول‌اند

اگر در آزمون فرض برابری میانگین دو جامعه نرمال، واریانس‌ها مجهول و برابر هم یا نزدیک به هم باشند و حجم نمونه‌ها کوچک باشند، آماره آزمون تحت فرض $H_\circ$ دارای توزیع استودنت با $m+n-2$ درجه آزادی خواهد بود.

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{S_P^2(\frac{1}{m} + \frac{1}{n})}}$$

که

$$S_P^2 = \frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2}$$

لازم به ذکر است که S_P^2 را در منابع آماری، واریانس آمیخته یا ادغام شده می‌گویند.

مثال برای مقایسه درآمد استحصالی زنان و مردان در یک واحد تولیدی صنایع روستایی در شهرک صنعتی قم بررسی آماری از ۷ مرد و ۵ زن اطلاعات زیر بدست آمده است.

متوسط حقوق دریافتی ۷ مرد $= \bar{x}_۱ = ۳۱/۷۱ \quad , \quad S_۱^۲ = ۲/۹۰$

متوسط حقوق دریافتی ۵ زن $= \bar{x}_۲ = ۳۵/۲ \quad , \quad S_۲^۲ = ۲/۲۰$

آیا با اطمینان ۹۵٪ می‌توان گفت که متوسط حقوق دریافتی مردان و زنان در جامعه برابر است؟ (ارقام به ۱۰۰/۰۰۰ ریال است)

حل ابتدا فرض‌های آماری را می‌نویسیم.

$$\begin{cases} H_\circ : \mu_۱ = \mu_۲ \\ H_۱ : \mu_۱ \neq \mu_۲ \end{cases}$$

$$m = ۷ \quad , \quad \bar{x}_۱ = ۳۱/۷۱ \quad , \quad S_۱^۲ = ۲/۹۰ \quad , \quad \alpha = ۰/۰۵$$

$$n = ۵ \quad , \quad \bar{x}_۲ = ۳۵/۲ \quad , \quad S_۲^۲ = ۲/۲۰ \quad , \quad \frac{\alpha}{۲} = ۰/۰۲۵$$

درجه آزادی $= m + n - ۲ = ۱۲ - ۲ = ۱۰$

سپس بر اساس اطلاعات دو جامعه آماره آزمون را محاسبه می‌کنیم. به این منظور ابتدا S_P^2 را حساب می‌کنیم. یعنی:

$$S_P^۲ = \frac{(m-۱)S_۱^۲ + (n-۱)S_۲^۲}{m+n-۲} = \frac{۶ \times ۲/۹۰ + ۴ \times ۲/۲۰}{۷+۵-۲} = ۲/۶۲$$

حال مقدار آماره عبارت خواهد بود از:

$$T = \frac{\bar{x}_۱ - \bar{x}_۲ - (\mu_۱ - \mu_۲)}{S_p \sqrt{\frac{۱}{m} + \frac{۱}{n}}} = \frac{۳۱/۷۱ - ۳۵/۲ - ۰}{۷+۵-۲} = -۳/۶۸$$

مقدار T جدول با توجه به درجه آزادی و مقدار α عبارت است از:

$$P\left[T<-t_{\frac{\alpha}{۲}}\right]=۰/۰۲۵\Rightarrow -t_{\frac{\alpha}{۲}}=-۲/۲۲۸$$

$$P\left[T>t_{\frac{\alpha}{۲}}\right]=۰/۰۲۵\Rightarrow t_{\frac{\alpha}{۲}}=۲/۲۲۸$$

در نهایت، چون $T=-۳/۶۸<-۲/۲۲۸$ است، فرض $H_{۰}$ رد می‌شود.

مثال مدیریت شرکت تعاونی روستایی نیل آباد در تربت جام که تولید کننده دو فرآورده لبنی برای مثال پنیر است برای تضمین بازار فروش مدیر عامل این شرکت مایل است از بین دو سری پنیر تولیدی که قیمت‌های یکسانی‌دارند، پنیری را که دارای بازار فروش بیشتری است، انتخاب کند .بهمین منظور از هر گروه نمونه‌ای تصادفی به حجم ۱۰۰ انتخاب می‌کند. نتیجه بررسی‌ها نشان می‌دهد که نمونه دریافتی از گروه A دارای میانگین ۱۱۸۰ کارتن در روز فروش و انحراف معیار ۱۲۰ ساعت و نمونه دریافتی از گروه B دارای میانگین ۱۱۶۰ کارتن در روز فروش و انحراف معیار ۴۰ بود. با توجه به اطلاعات فوق چه تصمیمی باید گرفت با فرض اینکه واریانس‌های دو گروه مساوی $\alpha=۰/۰۵$ باشد.

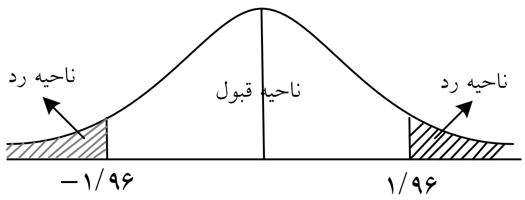
حل فرض‌های آماری به شرح زیر خواهند بود:

$$\begin{cases} H_{۰}:\mu_{۱}=\mu_{۲} \\ H_{۱}:\mu_{۱}\neq\mu_{۲} \end{cases}$$

با توجه به حجم نمونه‌ها که $n>۳۰$ می‌باشند، می‌توانیم برای محاسبه آماره آزمون از توزیع نرمال استاندارد استفاده کنیم. لذا خواهیم داشت:

$$Z=\frac{\overline{X}_{۱}-\overline{X}_{۲}}{\sqrt{\frac{S_{۱}^{۲}}{n_{۱}}+\frac{S_{۲}^{۲}}{n_{۲}}}}=\frac{۱۱۸۰-۱۱۶۰}{\sqrt{\frac{۱۴۴۰۰}{۱۰۰}+\frac{۱۶۰۰}{۱۰۰}}}=\frac{۲۰}{\sqrt{۱۴۴+۱۶}}=\frac{۲۰}{\sqrt{۱۶۰}}=۱/۵۸$$

این آزمون دو دامنه بوده و مناطق بحرانی آن در دو دامنه راست و چپ قرار دارد.



با توجه به اینکه مقدار محاسبه شده در منطقه مجاز واقع شده، فرض H_0 را نمی‌توانیم رد کنیم و نتیجه می‌گیریم که متوسط فروش دو گروه در سطح معنی‌دار ۰/۰۵ باهم برابرند.

۲-۱۳ آزمون فرض آماری با استفاده از P – مقدار[۱]

چنان که در قبلاً ذکر شد، تحلیل‌گر آمار همیشه با داده‌های از نوع کمّی سر و کار ندارد. بخشی از فرضیه‌های پژوهشی تدوین شده در رشتهٔ جغرافیا و برنامه ریزی روستایی می تواند از نوع نسبت‌پذیرباشند. از این‌رو فرضیه‌های مربوط به مقیاس‌های کیفی از نسبت موفقیت مورد ادعا آزمون می‌شوند. بر این اساس یکی از روش‌های آزمون فرض آماری، استفاده از P –مقدار یا مقدار احتمال است. اکنون که نرم‌افزارهای آماری ملاک آزمون فرض آماری را برحسب P –مقدار ارائه می‌کنند، لازم دیدیم که این بحث را در بخش جداگانه مورد بحث قرار دهیم.

فرض کنید $\overline{X}$ میانگین یک نمونه تصادفی از توزیع نرمال با میانگین μ و واریانس معلوم σ^2 باشد و بخواهیم فرض $H_0 : \mu = \mu_0$ را در مقابل فرض $H_1 : \mu < \mu_0$، آزمون کنیم. در این آزمون مقادیر کوچک $\overline{X}$ باعث رد فرض H_0 می‌شود. اگر α سطح معنی‌دار آزمون باشد فرض H_0 رد می‌شود اگر احتمال پیشامد $\overline{X} \le \overline{x}$ تحت فرض H_0 کمتر یا مساوی α باشد یعنی P –مقدار برابر است با:

---

۱. P-Value

$$\text{مقدار}-P = P\left[\bar{X} \le \bar{x} \mid H_{\circ}\right] \le \alpha$$

یا

$$\text{مقدار}-P = P\left[\frac{\bar{X}-\mu_{\circ}}{\frac{\sigma}{\sqrt{n}}} < \frac{\bar{x}-\mu_{\circ}}{\frac{\sigma}{\sqrt{n}}}\right] = \left[Z < \frac{\bar{x}-\mu_{\circ}}{\frac{\sigma}{\sqrt{n}}}\right] \le \alpha$$

در استفاده از روش P –مقدار اگر P –مقدار کمتر یا مساوی α تعیین شده باشند، فرض $H_{\circ}$ رد می‌شود.

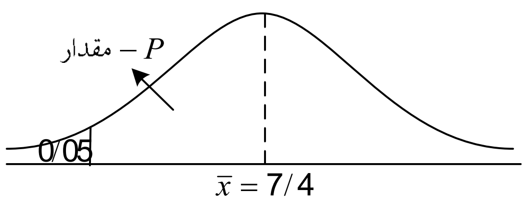
مثال یک نمونه ۳۶ تایی از یک کارگاه تولید کننده قالی دست باف در کاشان به طور متوسط دارای ۷/۴ ابریشم از مواد اصلی است. اگر توزیع جامعه مورد بررسی نرمال با انحراف معیار $\sigma = 0/48$ باشد، فرض $H_{\circ}: \mu = 7/5$ را در مقابل $\mu < 7/5$ در سطح ۵ درصد آزمون کنید.

حل

$$n = ۳۶ \quad , \quad \bar{x} = ۷/۴ \quad , \quad \sigma = ۰/۴۸ \quad , \quad \alpha = ۰/۰۵$$

$$\text{مقدار}-P = P\left[\bar{X} \le \bar{x} \mid H_{\circ}\right] = P\left[\frac{\bar{X}-\mu_{\circ}}{\frac{\sigma}{\sqrt{n}}} < \frac{\bar{x}-\mu_{\circ}}{\frac{\sigma}{\sqrt{n}}}\right] = P\left[Z < \frac{7/407-7/5}{\frac{0/48}{\sqrt{36}}}\right]$$

$$= p\left[Z < -1/25\right] = 0/1056$$



چون P - مقدار کمتر از ۰/۰۵ نیست، فرض $H_{\circ}$ پذیرفته می‌شود.

مثال یک کارخانه تولید کننده لامپ چراغ تراکتور در شهرک صنعتی روستایی در اصفهان ادعا می‌کند که طول عمر لامپ تولیدی دارای توزیع نرمال با میانگین $\mu = 800$ و انحراف معیار ۴۰ ساعت است. اگر یک نمونه ۳۰ تایی دارای متوسط ۷۸۸ ساعت باشد، فرض $H_{\circ} : \mu = 800$ را در مقابل $H_1 : \mu \neq 800$ در سطح ۵ درصد آزمون کنید.

حل

$$n = ۳۰ \quad , \quad \bar{x} = ۷۸۸ \quad , \quad \alpha = ۰/۰۵ \quad , \quad \frac{\alpha}{۲} = ۰/۰۲۵$$

در این مثال، مقدار بزرگ یا کوچک $\overline{X}$ باعث رد $H_{\circ}$ می‌شود.

P -مقدار در این مثال از دو رابطه زیر بدست می‌آید:

$$\text{مقدار}-P_1 = P\left[\overline{X} > \bar{x} \,\middle|\, \text{درست} \quad H_{\circ}\right] = P\left[Z > \frac{\bar{x} - \mu_{\circ}}{\frac{\sigma}{\sqrt{n}}}\right]$$

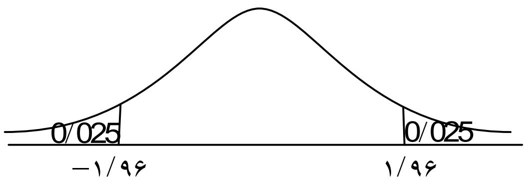
$$\text{مقدار}-P_2 = P\left[\overline{X} < \bar{x} \,\middle|\, \text{درست} \quad H_{\circ}\right] = P\left[Z < \frac{\bar{x} - \mu_{\circ}}{\frac{\sigma}{\sqrt{n}}}\right]$$

$$\text{مقدار}-P_1 = P\left[Z > \frac{788 - 800}{\frac{40}{\sqrt{30}}}\right] = P\left[Z > -1/643\right] = 0/9495$$

$$\text{مقدار}-P_2 = P\left[Z < \frac{788 - 800}{\frac{40}{\sqrt{30}}}\right] = P\left[Z < -1/643\right] = 0/0505$$

P_1 –مقدار و P_2 –مقدار را با $\frac{\alpha}{۲}=۰/۰۲۵$ مقایسه می‌کنیم.

اگر مقدار P_1 –مقدار یا P_2 –مقدار کمتر از ۰/۰۲۵ باشد، فرض H_0 رد می‌شود.



چون $P_1 > 0/025$ –مقدار یا $P_2 > 0/025$ –مقدار است پس فرض H_0 پذیرفته می‌شود.

مثال برای بررسی متوسط درآمد سرانه در طرح ساماندهی اقتصادی دهستان گوده شهرستان بستک در بوشهر، ۹ نفر انتخاب شده‌اند. اگر متوسط درآمد و انحراف معیار به ترتیب $\bar{x} = 32/8$ و ۴/۵۱ باشد، فرض $H_0 : \mu = ۳۰$ را در مقابل $H_1 : \mu > 30$ با اطمینان ۹۰٪ آزمون کنید. (ارقام به صدهزار ریال می‌باشد)

$n = 9$, $\alpha = 0/1$, $\bar{x} = 32/8$, $s = 4/51$, درجه آزادی =۸

$$\text{مقدار–}P = P\left[\bar{X} > \bar{x} \mid \text{درست } H_0\right] = P\left[\frac{\bar{X}-\mu_0}{\frac{s}{\sqrt{n}}} > \frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}\right]$$

$$= P\left[T > \frac{32/8 - 30}{\frac{4/51}{\sqrt{9}}}\right] = P\left[T > 1/86\right]$$

از اینکه آماره T دارای توزیع استودنت با $n-۱$ درجه آزادی است از جدول استودنت P –مقدار بدست می‌آید.

$\text{مقدار–}P = P[T > ۱/۸۶] = ۰/۰۵$

چون P –مقدار کمتر از ۰/۰۱ است، فرض H_0 رد می‌شود.

۲-۱۴ آزمون فرض برای پارامتر توزیع دوجمله‌ای

می‌دانیم که اگر متغیر تصادفی X دارای توزیع دوجمله‌ای با پارامتر مجهول n و P باشد، تابع چگالی احتمال آن برابر است با

$$f(x)=P(X=x)=\binom{n}{x}P^{x}(1-P)^{n-x} \quad , \quad x=1,2,\cdots,n$$

با فرض اینکه n معلوم است، انواع فرض‌ها درباره پارامتر P عبارتند از:

$$1)\left| \begin{array}{l} H_0 : P = P_0 \\ H_1 : P > P_0 \end{array} \right. \qquad 2)\left| \begin{array}{l} H_0 : P = P_0 \\ H_1 : P < P_0 \end{array} \right. \qquad 3)\left| \begin{array}{l} H_0 : P = P_0 \\ H_1 : P \neq P_0 \end{array} \right.$$

برآوردگر P برابر با $\frac{X}{n}$ است که X مقدار پیروزی در n آزمایش است.

برای انجام آزمون فرض‌های فوق از کمیت محوری $Z=\dfrac{\frac{X}{n}-P}{\sqrt{\frac{P(1-P)}{n}}}$ که برای n بزرگ تحت فرض H_0 دارای توزیع نرمال استاندارد است، استفاده می‌شود.

چون Z دارای توزیع نرمال استاندارد است، نواحی قبول یا رد آن نیز همانند آزمون میانگین توزیع نرمال است.

مثال محموله‌ای شامل ۵۰ رایانه پیله وران روستایی در بازارچه مرزی **باز**ارچه مرزی باشماق مریوان است، اگر ۸ رایانه در این محموله معیوب باشند، آیا در سطح ۵ درصد می‌توان گفت که نسبت رایانه‌های معیوب در جامعه کمتر از ۲۰ درصد است؟

$$n=50 \; , \quad x=8 \; , \quad \alpha=0/05 \; , \quad P=0/2$$

$$\left| \begin{array}{l} H_0 : P = 0/2 \\ H_1 : P < 0/2 \end{array} \right.$$

آماره آزمون

$$Z_{\circ} = \frac{\frac{x}{n} - P_{\circ}}{\sqrt{\frac{x}{n}(1-\frac{x}{n})}} = \frac{\frac{8}{50} - 0/2}{\sqrt{\frac{8}{50}(1-\frac{8}{50})}} = -0/109$$

مقدار جدول برای $\alpha = ۰/۰۵$ برابر است با $z_{\alpha} = -۱/۶۴$؛ چون $Z_{\circ} > -Z_{\alpha}$ است، فرض $H_{\circ}$ پذیرفته می‌شود.

۲-۱۵ آزمون فرض مقایسه نسبت موفقیت در دو جامعه آماری

در بسیاری از پژوهش‌ها بالاخص در حوزه برنامه ریزی روستایی و مدیریت توسعه روستایی، نسبت‌های موفقیت دو جامعه مورد مقایسه قرار می‌گیرند که نوع بررسی‌ها را اصطلاحاً مطالعات تطبیقی می‌گویند. اگر نسبت موفقیت‌ها در جامعه P باشد، بر اساس آنچه که قبلاً در مورد نسبت موفقیت یک جامعهٔ آماری گفته شد، توزیع نسبت موفقیت‌ها در نمونه ($\overline{P}$) دارای توزیع نرمال است، به شرط آن که $n\overline{p}$ و $n\overline{q}$ هر دو بزرگتر یا مساوی ۵ باشند. در عین حال میانگین و انحراف معیار متغیر $\overline{P}$ برابر بود با:

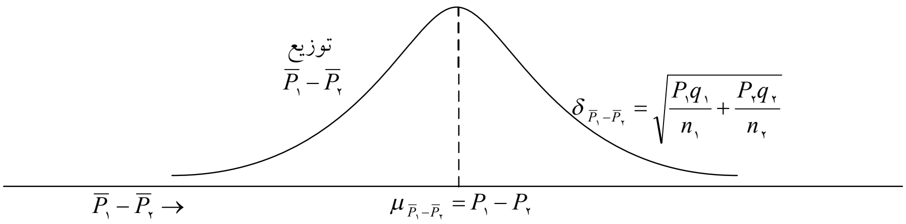
$$\mu_{\overline{p}} = P \qquad \sigma_{\overline{p}} = \sqrt{\frac{pq}{n}}$$

حال تصور کنید می‌خواهیم نسبت خاصی را در دو جامعه با یکدیگر مقایسه کنیم. فرض کنید $P_۱$ نشان‌دهنده نسبت مورد نظر در جامعه اول و $P_۲$ نسبت مورد نظر در جامعه دوم باشد. به منظور مقایسه این دو نسبت، اگر $\overline{P}_۱$، $\overline{P}_۲$ به ترتیب نسبت‌های نمونه‌ای به حجم $n_۱$ و $n_۲$ از دو جامعه باشند،متغیر $\overline{P}_۱ - \overline{P}_۲$ دارای توزیع نرمال با میانگین $\mu_{\overline{P}_۱ - \overline{P}_۲}$ و انحراف معیار $\sigma_{\overline{P}_۱ - \overline{P}_۲}$ است. لذا خواهیم داشت:

$$\mu_{\overline{P}_۱ - \overline{P}_۲} = P_۱ - P_۲$$

$$\sigma_{\overline{P}_۱ - \overline{P}_۲} = \sqrt{\frac{P_۱ - q_۱}{n_۱} + \frac{P_۲ - q_۲}{n_۲}}$$

با در اختیار داشتن میانگین و انحراف معیار متغیر تصادفی $\overline{P}_1 - \overline{P}_2$، می‌توان توزیع آن را به صورتی که در شکل زیر مشاهده می‌شود، رسم کرد.



در آزمون فرضیه برای نسبت یک جامعه که قبلاً به آن پرداخته شد، آزمون‌هایی مانند $H_0 : P = ۰/۲۵$ مطرح بود. در این آزمون‌ها مشکلی برای محاسبه انحراف معیار $\overline{P}$ که در گرو دانستن P است وجود نداشت، زیرا کمیت P را با فرض درست بودن H_0، برابر مقدار عنوان شده آن توسط H_0 (مثلاً ۰/۲۵) در فرمول قرار می‌دادیم.

$$\sigma_{\overline{P}} = \sqrt{\frac{pq}{n}}$$

اما آزمون فرض برابری دو نسبت، یعنی $H_1 : P_1 - P_2 = 0$، کمیتی را برای P_1 و P_2 ارائه نمی‌کند و تنها عنوان می‌کند که با هم مساوی هستند. در نتیجه عملاً نمی‌توان انحراف معیار متغیر تصادفی $\overline{P}_1 - \overline{P}_2$ را به صورت زیر محاسبه کرد:

$$\sigma_{\overline{P}_1 - \overline{P}_2} = \sqrt{\frac{P_1 - q_1}{n_1} + \frac{P_2 - q_2}{n_2}}$$

و لذا بر اساس آن نمی‌توان آماره آزمون زیر را تشکیل داد:

$$Z = \frac{\left(\overline{P}_1 - \overline{P}_2\right) - \left(P_1 - P_2\right)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}}$$

اما از آنجا که فرض بر صحت $H_۰$ است، مگر خلاف آن ثابت شود، می‌توان تصور کرد که $P_۱$ و $P_۲$ دارای کمیتی مساوی و برابر P هستند، یعنی:

$$p_۱ = p_۲ = P$$

اکنون اگر این مقدار را در فرمول انحراف معیار متغیر تصادفی $\overline{P}_۱ - \overline{P}_۲$ قرار دهیم، خواهیم داشت:

$$\sigma_{\overline{P}_۱ - \overline{P}_۲} = \sqrt{\frac{P_۱ - q_۱}{n_۱} + \frac{P_۲ - q_۲}{n_۲}} = \sqrt{Pq\left(\frac{۱}{n_۱} + \frac{۱}{n_۲}\right)}$$

از آنجا که کمیت P برای ما مجهول است، باید مقدار آن را از نمونه، برآورد کنیم. سؤالی که اکنون مطرح می‌شود آن است که چگونه می‌توان برآوردی از نسبت مشترک در دو جامعه (P) که ما آن را $(\overline{P})$ نشان می‌دهیم، به دست آورد. چون فرض بر صحت $H_۰$ است، مگر خلاف آن ثابت شود، می‌توان چنین تصور کرد که نسبت‌ها با هم مساوی هستند و در نتیجه چنین انگاشت که دو نمونه تصادفی گرفته شده، هر دو متعلق به یک جامعه هستند، بنابراین می‌توان تعداد موفقیت‌ها در دو نمونه را شمارش کرده به جمع دو نمونه تقسیم کرد تا برآوردی از $(\overline{P})$ به دست آید.

$$\overline{P} = \frac{X_۱ + X_۲}{n_۱ + n_۲}$$

با داشتن $\overline{P}$ می‌توان $\overline{q}$ را به صورت $\overline{q} = ۱ - \overline{P}$ محاسبه کرد. به این ترتیب می‌توان برآوردی از انحراف معیار متغیر تصادفی $\overline{P}_۱ - \overline{P}_۲$ را به صورت زیر به دست آورد.

$$S_{\overline{P}_۱ - \overline{P}_۲} = \sqrt{Pq\left(\frac{۱}{n_۱} + \frac{۱}{n_۲}\right)}$$

برای تشخیص اینکه آیا حجم نمونه‌های $n_۱$ و $n_۲$ به اندازه کافی بزرگ هستند تا بتوان اطمینان حاصل کرد که متغیرهای تصادفی $\overline{P}_۱$ و $\overline{P}_۲$ دارای توزیع نرمال هستند یا نه، چاره‌ای نیست، جز آنکه حدس بزنیم از بین $P_۱$ و $P_۲$ کدام یک از همه کوچکتر است و مقدار آن چیست. سپس حجم نمونه $n_۱$ و $n_۲$ را در آن ضرب کنیم و اطمینان

حاصل کنیم که هر دو حاصل‌ضرب بزرگتر یا مساوی ۵ است. به عنوان مثال فرض کنید حدس بزنیم که $P_۲$ کوچکتر از همه و برابر ۰/۰۴ است. در این صورت $n_۱$ و $n_۲$ هر کدام باید حداقل برابر ۱۲۵ باشند تا شرط ۵=۰/۰۴× ۱۲۵ برقرار شود.
اکنون می‌توان متغیر تصادفی $\overline{P}_۱-\overline{P}_۲$ را به صورت زیر استاندارد کرد:

$$Z=\frac{\left(\overline{P}_۱-\overline{P}_۲\right)-\left(P_۱-P_۲\right)}{\sqrt{\overline{P}\,\overline{q}\left(\frac{۱}{n_۱}+\frac{۱}{n_۲}\right)}}$$

با در اختیار داشتن آماره آزمون که دارای توزیع نرمال استاندارد است، به سادگی می‌تواند آزمونی برای $\overline{P}_۱-\overline{P}_۲$ انجام شود.

مثال یک شرکت تولید کنندهٔ لاستیک خودرو به صورت دو شیفت صبح و بعد از ظهر مشغول به فعالیت است. اطلاعات مربوط به شیفت صبح و بعد از ظهر این شرکت به شرح زیر است:

شیفت صبح (۱)	شیفت صبح (۲)
$n_۱$ =۱۰۰	$n_۲$ =۱۰۰
۵ حلقه لاستیک زیر ۱۰,۰۰۰ کیلومتر کار می‌کنند	۱۰ حلقه لاستیک زیر ۱۰,۰۰۰ کیلومتر کار می‌کنند

می‌خواهیم بدانیم که آیا در دو شیفت صبح و بعد از ظهر نسبت لاستیک‌هایی که زیر ۱۰,۰۰۰ کیلومتر کار می‌کنند مساوی است؟

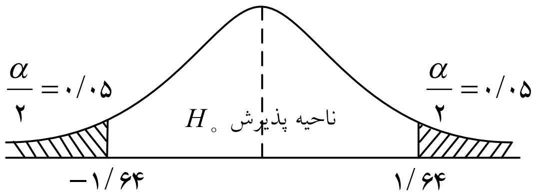
حل ابتدا فرض صفر و فرضیه مقابل را به صورت زیر می‌نویسیم:

$$H_۰: P_۱=P_۲ \Rightarrow P_۱-P_۲=۰$$
$$H_۱: P_۱\neq P_۲ \Rightarrow P_۱-P_۲\neq ۰$$

اگر سطح معنی‌دار بودن را ۰/۱ در نظر بگیریم، خواهیم داشت:

$$\alpha = ۰/۱ \Rightarrow \frac{\alpha}{۲} = ۰/۰۵ \Rightarrow Z_{\frac{\alpha}{۲}} = ۱/۶۴$$

لذا مقادیر بحرانی ۱/۶۴- و ۱/۶۴+ مربوطه، ناحیه رد را از عدم رد به گونه‌ای که در شکل زیر مشهود است، جدا می‌کنند.



برای محاسبه کمیت آماره آزمون، با توجه به اطلاعات مسأله داریم:

نسبت لاستیک‌هایی که در شیفت صبح کار نمی‌کنند $\quad \overline{P}_۱ = \frac{X_۱}{n_۱} = \frac{۵}{۱۰۰} = ۰/۰۵$

نسبت لاستیک‌هایی که در شیفت بعد از ظهر کار نمی‌کنند $\quad \overline{P}_۲ = \frac{X_۲}{n_۲} = \frac{۱۰}{۱۰۰} = ۰/۱$

به این ترتیب نسبت لاستیک‌های کار نکرده عبارت است از:

$$\overline{P} = \frac{X_۱ + X_۲}{n_۱ + n_۲} = \frac{۵+۱۰}{۱۰۰+۱۰۰} = ۰/۰۷۵$$

کمیت آماره آزمون برابر است با:

$$Z = \frac{\left(\overline{P}_۱ - \overline{P}_۲\right) - \left(P_۱ - P_۲\right)}{\sqrt{\hat{P}\hat{q}\left(\frac{۱}{n_۱} + \frac{۱}{n_۲}\right)}} = \frac{\left(۰/۰۵ - ۰/۱\right) - ۰}{\sqrt{\left(۰/۰۷۵\right)\left(۰/۹۲۵\right)\left(\frac{۱}{۱۰۰} + \frac{۱}{۱۰۰}\right)}} = -۱/۳۴$$

از آنجا که کمیت آماره آزمون در ناحیه عدم رد قرار می‌گیرد، فرضیه صفر را نمی‌توان رد کرد. بنابراین نتیجه می‌گیریم که با توجه به اطلاعات موجود، دلیلی وجود ندارد که نسبت لاستیک‌های کار نکرده در زیر ۱۰,۰۰۰ کیلومتر در دو شیفت صبح و بعد از ظهر متفاوت باشند.

۲-۱۶ آزمون فرض برای واریانس جامعه

یکی از راههای بررسی تغییرپذیری جامعه، بررسی یا آزمون فرض درباره واریانس جامعه است. به عنوان مثال مدیر روستایی (دهیار) آیا میزان تغییر تولید روستایی دریک محصول عمده در اثربخشی قیمت در حدود قابل قبول است یا نه.

انواع فرض‌هایی که می‌توان درباره واریانس داشت، عبارتند از:

$$1)\left|\begin{array}{l} H_\circ: \sigma^2=\sigma_\circ^2 \\ H_1: \sigma^2>\sigma_\circ^2 \end{array}\right. \qquad 2)\left|\begin{array}{l} H_\circ: \sigma^2=\sigma_\circ^2 \\ H_1: \sigma^2<\sigma_\circ^2 \end{array}\right. \qquad 3)\left|\begin{array}{l} H_\circ: \sigma^2=\sigma_\circ^2 \\ H_1: \sigma^2 \neq \sigma_\circ^2 \end{array}\right.$$

اگر جامعه مورد بررسی دارای توزیع نرمال با میانگین μ مجهول و واریانس σ^2 باشد کمیت $\frac{(n-1)s^2}{\sigma^2}$ یک کمیت محوری است و تحت فرض $H_\circ$ دارای توزیع کای‌مربع یا کی‌دو با $(n-1)$ درجه آزادی است. لذا برای انجام آزمون فرض‌ها از آماره فوق استفاده می‌کنیم. برای آزمون فرض واریانس جامعه مراحل زیر را دنبال می‌کنیم:

۱- تعریف فرض آماری که یکی از فرض‌های سه‌گانه فوق خواهد بود. اگر $H_\circ=\sigma^2=\sigma_\circ^2$ را در مقابل $H_1=\sigma^2>\sigma_\circ^2$ در نظر بگیریم، چون s^2 یک برآورد نقطه‌ای برای σ^2 است، لذا مقادیر بزرگ s^2 باعث رد $H_\circ$ می‌شود. همچنین خواهیم داشت:

رد فرض $H_\circ$ هم ارز است با $\quad s^2>k_\circ$

رد فرض $H_\circ$ هم ارز است با $\quad (n-1)s^2>k_1$

رد فرض $H_\circ$ هم ارز است با $\quad \frac{(n-1)s^2}{\sigma^2}>k_2$

$k_\circ$، k_1 و k_2 اعداد ثابت هستند.

۲- محاسبه آماره آزمون.

آماره $\frac{(n-1)s^2}{\sigma_\circ^2}$ تحت فرض $H_\circ$ دارای توزیع کی‌دو با $n-1$ درجه آزادی است؛ اگر α سطح معنی‌دار آزمون باشد.

$$\alpha = P\left[\frac{(n-1)s^2}{\sigma^2_\circ} > \chi^2_{(n-1,\alpha)}\right]$$

۳- تعیین مقدار $\chi^2_{(n-1,\alpha)}$ از جدول توزیع کی‌دو با $n-1$ درجه آزادی.

۴- تصمیم‌گیری. برای آزمون فرض $H_\circ = \sigma^2 = \sigma^2_\circ$ در مقابل $H_1 = \sigma^2 > \sigma^2_\circ$ کافی است آماره $\chi^2_\circ = \frac{(n-1)s^2}{\sigma^2_\circ}$ تحت فرض $H_\circ$ و مشاهدات، محاسبه و با مقدار جدول $\chi^2_{(n-1,\alpha)}$ مقایسه نمود. اگر $\chi^2_\circ > \chi^2_{(n-1,\alpha)}$ باشد، فرض $H_\circ$ رد می‌شود.

مثال بسته‌های ۵۰۰ گرمی چای در کارخانه چای خشک کنی روستایی در دهستان رودبنه به وسیله یک دستگاه اتوماتیک پر می‌شوند. اگر وزن بسته‌ها از توزیع نرمال با میانگین μ و واریانس σ^2 پیروی کنند، برای یک نمونه ۵تایی مشاهده شده، ۴۹۹، ۵۰۲، ۵۰۱، ۴۹۸ و ۵۰۰ فرض $H_\circ : \sigma^2 = 2/45$ را در مقابل $H_1 : \sigma^2 > 2/45$ در سطح ۵ درصد آزمون کنید.

$$n = 5 \quad , \quad \alpha = 0/05 \quad , \quad s^2 = \frac{1}{n-1}\sum\left(x_i - \bar{x}\right)^2 = 2/5$$

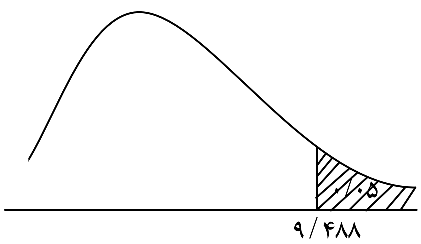
حل

آماره آزمون

$$\chi^2_\circ = \frac{(n-1)s^2}{\sigma^2_\circ} = \frac{4 \times 2/5}{2/45} = 4/082$$

مقدار جدول

$$P\left[\chi^2 > \chi^2_{(n-1,\alpha)}\right] = 0/05 \Rightarrow \chi^2_{(4,\alpha)} = 9/488$$



چون $\chi^2_{\circ} < \chi^2_{(4,\alpha)}$ است فرض $H_{\circ}$ پذیرفته می‌شود.

مثال مسئول آزمایشگاه کنترل کیفیت میگوشرکت "صیادان روستایی بندرعباس" معتقد است که واریانس اندازه‌هایی که در طول آزمایش ثبت می‌کند، کوچکتر از ۲ است. در یک آزمایش او اندازه‌های ۴/۱، ۵/۲ و ۱۰/۲ را ثبت کرده است. اگر اندازه‌ها دارای توزیع نرمال باشند، آیا می‌توان ادعای مسئول آزمایشگاه شرکت " صیادان روستایی بندرعباس " را در سطح $\alpha = ۰/۰۱$ پذیرفت؟

۱) آزمون فرض

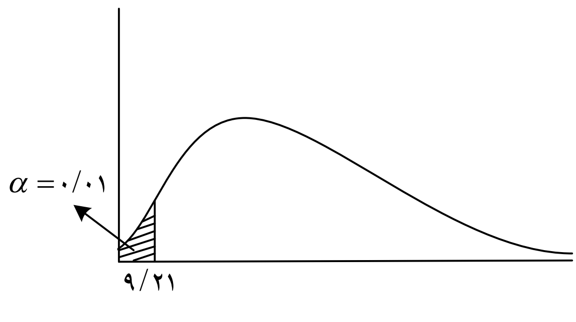
$$\begin{cases} H_{\circ} : \sigma^{۲} \geq ۲ \\ H_{۱} : \sigma^{۲} < ۲ \end{cases}$$

۲) محاسبه آماره آزمون

$$\chi^{۲}_{\circ} = \frac{(n-۱)S^{۲}}{\sigma^{۲}_{\circ}} = \frac{(۲)(۱۰/۵۷)}{۲} = ۱۰/۵۷$$

۳) مقدار بحرانی

$$\chi^{۲}_{(\alpha=۰/۰۱,\ \beta=۲)} = ۹/۲۱$$



چنانکه مشاهده می‌شود مقدار آماره آزمون در ناحیه اطمینان قرار می‌گیرد، پس فرض $H_{\circ}$ مورد تأیید قرار می‌گیرد و فرض مسئول آزمایشگاه رد می‌شود.

مثال مدیر عامل بازار بورس محصولات کشاورزی ادعا می‌کند که یک (انحراف معیار) بازده سهام شرکت‌های عرضه کننده سهام در بازار بورس کمتر از ۵ تومان است. بدین منظور یکی از شرکت‌های کارگزاری، ۲۵ شرکت از بین شرکت‌های عرضه کننده سهام در بازار بورس را به طور تصادفی انتخاب کرده که میانگین بازده آنها ۴ انحراف معیارشان ۴ تومان است. اگر بازده شرکت‌های بازار بورس از توزیع نرمال برخوردار باشد، ادعا را در سطح خطای ۵ درصد آزمون کنید.

۱) آزمون فرض

$$\begin{cases} H_۰ : \sigma_X^۲ \geq ۲۵ \\ H_۱ : \sigma_X^۲ < ۲۵ \end{cases}$$

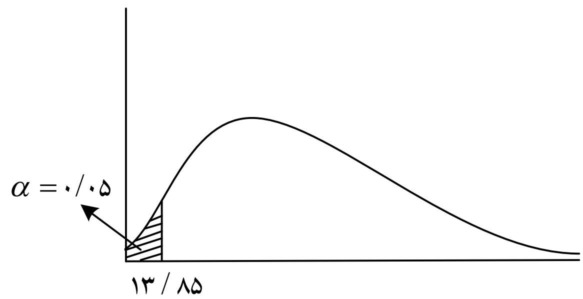
۲) محاسبه آماره آزمون

$$\chi^۲ = \frac{(n-۱)S_X^۲}{\sigma_۰^۲} = \frac{(۲۵-۱)(۴)^۲}{۲۵} = ۱۵/۳۶$$

۳) مقدار بحرانی

$$\chi_{(۰/۰۵\,,\,۲۴)} = ۱۳/۸۵$$

۴) تصمیم‌گیری



چنانکه ملاحظه می‌شود آماره آزمون در مقایسه با مقدار بحرانی در ناحیه اطمینان قرار می‌گیرد. بنابراین با ۹۵ درصد اطمینان می‌توان گفت که $H_۰$ پذیرفته می‌شود.

۲-۱۷ آزمون فرض برای نسبت دو واریانس

همچنان که دیدیم، وقتی که تفاوت بین میانگین‌های دو جامعه تحت مطالعه باشد و استفاده از توزیع نرمال یا استودنت در تعیین فاصله اطمینان و آزمون فرض مورد نظر باشد، می‌توان فرض تساوی واریانس دو جامعه مزبور را در نظر داشت. حال سؤال این است که آیا تفاوت بین واریانس‌های دو نمونه بر تفاوت واقعی میان واریانس‌های دو جامعه دلالت می‌کند یا اگر تفاوت بین واریانس‌های نمونه وجود داشته باشد، آیا تفاوت بر اثر شانس و برحسب تصادف است یا واقعاً واریانس‌های دو جامعه با هم اختلاف دارند؟

در این بخش به دنبال شیوه‌ای هستیم که پاسخ مناسب آماری را در مورد صحت فرض اختلاف و یا عدم اختلاف واقعی بین واریانس‌های دو جامعه ارائه دهد. تصمیم‌گیری‌هایی که به سازگاری یا عدم سازگاری پراش (واریانس) دو جامعه مربوط می‌شود، معمولاً بر اساس آزمون نسبت واریانس‌ها قرار دارد.

در آزمون فرض، این فرض را آزمون می‌کنیم که نسبت واریانس‌های دو جامعه برابر یک است.

اگر متغیر تصادفی X دارای توزیع نرمال با میانگین μ_1 و واریانس σ_1^2 و متغیر تصادفی Y دارای توزیع نرمال با میانگین μ_2 و واریانس σ_2^2 و متغیرهای X و Y از هم مستقل باشند، آنگاه فرض‌هایی می‌توان برای مقایسه واریانس دو جامعه به صورت زیر نوشت:

$$1)\left|\begin{array}{l} H_\circ : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 > \sigma_2^2 \end{array}\right. \qquad 2)\left|\begin{array}{l} H_\circ : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 < \sigma_2^2 \end{array}\right. \qquad 3)\left|\begin{array}{l} H_\circ : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{array}\right.$$

اگر S_1^2 واریانس یک نمونه m تایی از جامعه اول و S_2^2 واریانس یک نمونه n تایی از جامعه دوم باشد آماره $F = \left(\frac{\sigma_2}{\sigma_1}\right)^2 \frac{S_1^2}{S_2^2}$ دارای توزیع فیشر با $m-1$ و $n-1$ درجه آزادی است. برای آزمون فرض‌های اخیر از آماره F استفاده می‌کنیم.

در مقایسه واریانس‌های دو جامعه، برای سهولت این قرارداد را همیشه در نظر داریم که واریانس نمونه بزرگتر را در صورت قرار می‌دهیم به قسمی که نسبت واریانس‌های نمونه همیشه بزرگتر یا مساوی ۱ باشد.

برای آزمون فرض برای نسبت دو واریانس، مراحل چهارگانه مذکور در قسمت‌های قبل را مورد توجه قرار می‌دهیم.

برای نمونه در آزمون فرض $H_\circ = \sigma_1^2 = \sigma_2^2$ در مقابل $H_1 = \sigma_1^2 > \sigma_2^2$ مراحل زیر را انجام می‌دهیم:

۱- تعریف آماره آزمون

۲- مقدار آماره را تحت $H_\circ$ و نمونه، از رابطه زیر بدست می‌آوریم.

$$F_\circ = \left(\frac{\sigma_2}{\sigma_1}\right)^2 \frac{S_1^2}{S_2^2}$$

۳- مقدار $F_{(m-1,n-1,\alpha)}$ را از جدول فیشر بدست می‌آوریم.

۴- تصمیم‌گیری می‌کنیم. یعنی اگر $F_\circ > F_{(m-1,n-1,\alpha)}$ باشد، فرض $H_\circ$ را رد می‌کنیم در غیر اینصورت آن را می‌پذیریم.

مثال در آزمونی که برای مقایسه پراکندگی نمرات مدیران میانی و سطح بالا در ارزیابی کارآیی شرکت تعاون روستایی مغان انجام شده، اطلاعات زیر به دست آمده است.

نمرات کارایی ۱۰ مدیر میانی	۱۲	۱۳	۱۴	۱۷	۱۰	۱۸	۱۱	۱۹	۱۵	۱۶
نمرات کارایی ۸ مدیر سطح بالا	۱۴	۱۷	۱۵	۱۹	۱۲	۱۵	۲۰	۱۸		

آیا با اطمینان ۹۵٪ می‌توان گفت که واریانس یا پراکندگی نمرات مدیران میانی بیشتر از مدیران سطح بالا است؟

حل ابتدا فرض آماری را می‌نویسیم.

$n_1 = 10 \quad \overline{X}_1 = 14/5 \quad S_1^2 = 9/17 \quad H_\circ = \sigma_1^2 = \sigma_2^2$

$n_2 = 8 \quad \overline{X}_2 = 16/25 \quad S_2^2 = 7/35 \quad H_1 = \sigma_1^2 > \sigma_2^2$

مقدار آماره بر اساس اطلاعات مسأله برابر است با:

$$F_\circ = \left(\frac{\sigma_2^2}{\sigma_1^2}\right)^2 \times \frac{S_1^2}{S_2^2} = 1 \times \frac{9/17}{7/35} = 1/2476$$

استخراج مقادیر بحرانی در درجه آزادی‌های تعیین شده از جدول مربوطه

$$F_{(m-1, n-1, 5\%)} = F_{(9,7,5\%)} = 3/676$$

در نهایت چون $F_\circ = 1/2476 < F_{(9,7,5\%)} = 3/676$ لذا فرض $H_\circ$ پذیرفته می‌شود.

مثال یک نمونه تصادفی ۱۵تایی از کارکنان مرد و یک نمونه تصادفی ۲۰تایی از کارکنان زن یک شهرک صنعت روستایی قزوین مورد آزمون قرار گرفتند. متوسط نمره کارکنان مرد این سازمان ۸۴ و انحراف معیار آن ۶ بود، در حالی که متوسط نمره زنان ۷۸ و انحراف معیار آنان ۶ محاسبه شد. فرضیه $\sigma_1^2 \geq \sigma_2^2$ را در برابر $\sigma_1^2 < \sigma_2^2$ در سطح معنی‌داری $\alpha = 0/05$ آزمون کنید.

۱) فرض‌های آماری

$$\begin{cases} H_\circ : \sigma_1^2 \geq \sigma_2^2 \\ H_1 : \sigma_1^2 < \sigma_2^2 \end{cases}$$

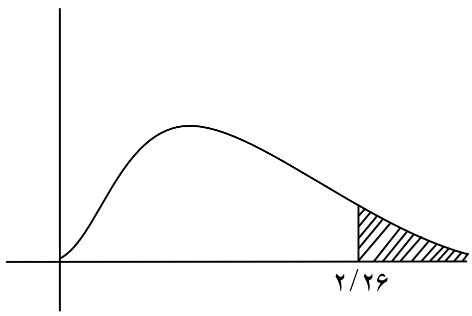
۲) محاسبه آماره آزمون

$$F = \frac{S_1^2}{S_2^2} = \frac{84}{36} = \frac{16}{9} = 1/78$$

۳) مقادیر بحرانی

$$F_{19, 24, \alpha = 0/05} = 2/26$$

۴) تصمیم‌گیری



با توجه به این که ۲/۲۶ > ۱/۷۸ است، لذا مقدار F محاسبه شده در منطقه مجاز واقع شده و فرض H_0 را نمی‌توان رد کرد. بنابراین می‌توان گفت که واریانس نمره کارکنان مرد و کارکنان زن این سازمان با هم برابرند.

مثال مدیریت کشاورزی یزد کنترل کیفیت شرکتی، یک گزارش مدیریتی تهیه کرده است که در آن آمده است: "پراکندگی محصولات کشاورزی تولیدی به وسیله سکونتگاههای"دشتی" بیشتر از سکونتگاههای "کوهپایه ایی" است. مدیر کشاورزی به منظور بررسی گزارش فوق، از سکونتگاههای " دشتی " و " کوهپایه ایی " نمونه‌هایی انتخاب کرده که حاصل اطلاعات نمونه‌ها به شرح جدول زیر است. توزیع وزن محصولاتی که به وسیله ماشین "الف" و "ب" تولید می‌شوند، نرمال است. صحت جمله را در سطح خطای ۵ درصد آزمون کنید.

سکونتگاههای دشتی	سکونتگاههای کوهپایه ایی
$n_A = ۱۶$	$n_B = ۸$
$\overline{X}_A = ۱۵\ kg$	$\overline{X}_B = ۱۵/۵\ kg$
$S_A = ۴/۵\ kg$	$S_B = ۲/۲۵\ kg$

۱) فرض‌های آماری

$$\begin{cases} H_{0}: \sigma_A^2 \le \sigma_B^2 \\ H_{1}: \sigma_A^2 > \sigma_B^2 \end{cases}$$

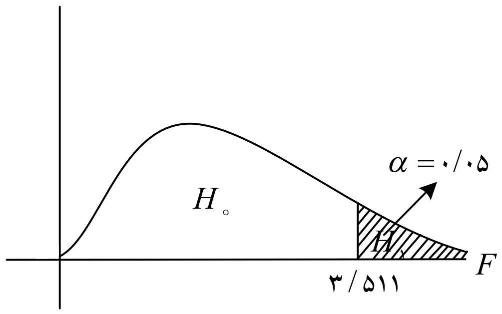
۲) محاسبه آماره آزمون

$$F = \frac{S_A^2}{S_B^2} = \frac{(4/5)^2}{(2/25)^2} = 4$$

۳) مقادیر بحرانی

$$F_{0/05,15,7} = 3/511$$

$$\begin{cases} \alpha = 0/05 \\ df_1 = n_A - 1 = 16 - 1 = 15 \\ df_2 = n_B - 1 = 8 - 1 = 7 \end{cases}$$



۴) تصمیم‌گیری

مقدار آماره آزمون در ناحیه H_1 قرار می‌گیرد یعنی داده‌ها دلالت کافی بر تأیید H_0 در سطح خطای ۵% ندارند. به علاوه چون H_1 بیان کننده ادعاست، پس می‌توان گفت ادعای مسئول کنترل کیفت در سطح اطمینان ۹۵% صحت دارد.

خودآزمایی

۱– الف: فرق بین فرض ساده و مرکب چیست؟

ب: خطای نوع اول و دوم را تعریف کنید.

۲– از دو گروه سکونتگاههای روستایی مختلف برای تولید برنج به ترتیب ۵ و ۷ سکونتگاه به طور تصادفی انتخاب شده و تولید آنها در این محصول پرسشگری شده است. این تولیدات به صورت زیر بوده است:

گروه اول	۸	۱۰	۱۴	۱۲	۱۱		
گروه دوم	۹	۱۰	۱۲	۱۴	۱۵	۱۳	۱۱

با فرض اینکه توزیع نمرات در دو گروه نرمال بوده است، آیا می‌توان گفت که متوسط تولید برنج در دو گروه مساوی است؟

۳– کارکنان بخش توزیع کالاهای مصرفی در یک شرکت حمل و نقل روستایی در توجیه درخواست اضافه حقوق خود ادعا کرده‌اند که آنان به طور متوسط طی ۱۳ دقیقه کار ارسال یک محموله را انجام می‌دهند. به عنوان مدیر این بخش شما چه تصمیمی می‌گیرید، اگر از یک نمونه ۴۰۰تایی میانگین تکیمل ارسال یک محموله ۱۴ دقیقه با انحراف معیار ۱۰ دقیقه محاسبه شود؟ (در سطح معنی‌دار $\alpha =$ ۰/۰۵ آزمون کنید).

۴– یک حلقه چاه نیمه عمیق به طور متوسط در هر دقیقه ۴۵ لیتر از مزرعه ایی را مشروب می‌سازد. برای افزایش بازدهی راندمان آبیاری یک حلقه چاه کمکی به آن اضافه شده و سپس در سه نوبت نمونه‌گیری انجام شد. نتایج نمونه‌گیری‌ها ۴۶، ۴۷ و ۴۸ در دقیقه بود. آیا این افزایش از نظر آماری معنی‌دار است یا صرفاً تصادفی بوده است؟ در سطح معنی‌دار ۰/۰۵ آزمون کنید(فرض کنید بازدهی ماشین توزیع نرمال دارد).

۵– اگر متغیر تصادفی X دارای توزیع دوجمله‌ای با پارامترهای P و $n=$ ۵ باشد کدام یک از فرض‌های زیر در مورد P، ساده و کدام مرکب است؟

$$\begin{cases} H_{\circ}: p = \frac{۱}{۲} & , \\ H_{۱}: p \neq \frac{۱}{۲} & , \end{cases} \quad \begin{cases} H_{\circ}: p = ۰/۹ & , \\ H_{۱}: p > \frac{۱}{۵} & , \end{cases} \quad \begin{cases} H_{\circ}: p \leq \frac{۱}{۴} \\ H_{۱}: \circ \leq p \leq ۰/۲ \end{cases}$$

۶- یک کارگاه قالیبافی به گونه‌ای تنظیم شده است که طول بافتی که انجام می شود ۲۰ سانتی متر باشد. میانگین یک نمونه ۱۰تایی از تولیدات این کارگاه ۲۰/۳ سانتی متر و انحراف معیار ۰/۲ سانتی متر محاسبه شد. آیا نتیجه به دست آمده حاکی از این است که کارگاه از تنظیم خارج شده است؟ در سطح معنی‌داری ۰/۰۵ آزمون کنید (فرض کنید طول قطعات ماشین توزیع نرمال دارد).

۷- جهت کنترل و ارزیابی مصرف نهاده های تولید در یک مزرعه تولید گندم در طی ۷ روز میزان مصرف مواد اولیه اندازه‌گیری شده و نتایج مشاهدات به صورت زیر به دست آمد:

(بر حسب تن) X : ۱۳ ،۸ ،۸ ،۸ ،۱۲ ،۶ ،۵ ،۱۱

مدیر این واحد تولیدی ادعا می‌کند که متوسط مصرف نهاده های تولید در این مزرعه روزانه ۱۰ تن می‌باشد. آیا بر اساس شواهد به دست آمده، می‌توان ادعای این مدیر را پذیرفت (خطای نوع اول را $\alpha =۰/۰۵$ بگیرید).

۸- فرض کنید متغیر تصادفی X دارای توزیع پواسن با پارامتر λ باشد. برای آزمون فرض $H_{\circ}: \lambda = ۲$ در مقابل $\lambda = 1/5$، اگر ناحیه بحرانی به صورت $C = \{x \mid x \leq 1\}$ باشد α، β و توان آزمون را حساب کنید.

۹- در یک نظرخواهی از ۱۵۰۰ روستایی از ساکنان یک منطقه در مورد کیفیت فعالیت اجرایی دستگاههای اجرایی روستایی ، ۵۵۰ نفر از آنان فعالیت اجرایی مورد نظر را با کیفیت دانسته‌اند. در منطقه دیگر از ۱۴۰۰ نفر که مورد سؤال واقع شده بودند، ۴۲۰ نفر کالا را با کیفیت دانسته‌اند. آیا بر اساس اطلاعات بدست آمده می‌توانیم بگوییم که نظر ساکنان دو منطقه یکسان است؟ (خطای نوع اول $\alpha =۰/۰۵$ را در نظر بگیرید).

۱۰- دو نوع باطری تراکتور مورد آزمایش قرار گرفته‌اند تا متوسط عمر آنها تعیین گردد. به طور تصادفی ۲۵ باطری نوع «الف» و ۲۵ باطری نوع «ب» انتخاب شده و مورد آزمایش قرار گرفته‌اند. نتیجه در جدول زیر خلاصه شده است.

باطری نوع «الف»	باطری نوع «ب»
$\overline{X}_1 =$ ۱۱۰/۶ ساعت	$\overline{X}_2 =$ ۱۰۳/۸ ساعت
$S_1 =$ ۱۰ ساعت	$S_2 =$ ۱۲ ساعت
۱۶ باطری بیش از ۱۰۰ ساعت عمر کرده‌اند.	۱۴ باطری بیش از ۱۰۰ ساعت عمر کرده‌اند.

با استفاده از آزمون فرض در خصوص موارد زیر نتیجه‌گیری کنید: ($\alpha =$ ۰/۱)

الف) آیا نسبت باطری‌های نوع «الف» که بیش از ۱۰۰ ساعت عمر می‌کنند، ۰/۹۵ است؟

ب) آیا دلیلی وجود دارد که باطری‌های نوع «الف» به طور متوسط بیشتر از ۹۵ ساعت عمر می‌کنند؟

ج) آیا اختلافی بین پراکندگی عمردو نوع باطری وجود دارد؟

د) آیا اختلافی بین متوسط عمر باطری نوع «الف» و «ب» وجود دارد؟

۱۱- فرض کنید X در فاصله θ تا $-\theta$ به طور یکنواخت توزیع شود. یک مقدار x را مشاهده می‌کنیم و می‌خواهیم فرض $H_0 : \theta =$ ۱ را در مقابل $H_1 : \theta =$ ۱/۵ آزمون کنیم. تصمیم می‌گیریم که اگر مقدار نمونه از ۰/۹۹ تجاوز کند H_0 را رد کنیم، α ، β و $1-\beta$ را حساب کنید.

۱۲- فرض کنید $\overline{X}$ میانگین یک نمونه تصادفی با حجم $n = 36$ از جامعه نرمال با میانگین μ و واریانس $\sigma^2 = 9$ باشد. برای آزمون فرض $H_0 : \mu =$ ۵۰ در مقابل $H_1 : \mu >$ ۵۰ اگر $\overline{X} \geq 50/8$ باشد فرض H_0 رد می‌شود. احتمال ارتکاب خطای نوع را حساب کنید.

۱۳- دو نمونه تصادفی مستقل به اندازه‌های $n_1 =$ ۳۰ و $n_2 =$ ۵۰ از دو جامعه با توزیع نرمال با میانگین‌های μ_1 و μ_2 و واریانس‌های $\sigma_1^2 =$ ۱۵۰ و $\sigma_2^2 =$ ۲۰۰ گرفتیم،

میانگین‌های این نمونه‌ها $\overline{X}_{1} = ۱۵/۸$ و $\overline{X}_{2} = ۱۱$ محاسبه شدند. آیا در سطح معنی‌دار ۰/۰۵ می توانیم بگوییم که میانگین جامعه اول حداقل ۳ واحد بیشتر از میانگین جامعه دوم است؟

۱۴- به منظور مقایسه وضع ساکنان دو ناحیه «الف» و «ب»، در نظر است از ناحیه «الف» و «ب» نمونه‌هایی انتخاب شده و وضعیت سواد روستانشینان آنها بررسی شود. با این هدف از ناحیه «الف» نمونه‌ای به حجم $n_{1} = ۱۰۰$ و از ناحیه «ب» نمونه‌ای به حجم $n_{2} = ۳۰۰$ انتخاب شده است. در نمونه اول ۱۱ نفر و در نمونه دوم ۶۹ نفر با سواد بودند. با توجه به این اطلاعات آیا می‌توانیم بگوییم که نسبت با سوادان در دو ناحیه یکسان است؟ (α=۰/۰۵).

۱۵- اگر مقدار راندمان تولید گندم در واحد سطح یک شرکت تعاونی زراعی در قطعات متفاوت به شرح ذیل باشد، دارای توزیع نرمال با میانگین μ و واریانس σ^2 باشد با استفاده از نمونه زیر فرض $H_{\circ} : \mu = ۳/۳$ را درمقابل $H_1 : \mu < ۳/۳$ در سطح ۵ درصد آزمون کنید.

۲/۴۵	۴/۲۲	۴/۷۵	۴/۴۷	۲/۵۸	۳/۴۰	۱/۱۲	۱/۴۳	۳/۶۵

۱۶- اگر $\overline{X}$ میانگین یک نمونه ۱۶تایی از توزیع نرمال با میانگین μ و انحراف معیار $\sigma = 8$ باشد، برای آزمون $H_{\circ} : \mu = 35$ در مقابل $H_{\circ} : \mu > ۳۵$، فرض $H_{\circ}$ را رد می‌کنیم، اگر $\overline{X} > 36/5$ باشد. مقدار P –مقدار را حساب کنید.

۱۷- دستمزد ماهانه کارگران بخش خدمات شرکت تعاونی زراعی بر طبق قانون نرمال توزیع شده با انحراف معیار σ=۱۰,۰۰۰ ریال و امید ریاضی آن باید μ=۵۲,۸۰۰ ریال باشد. جهت بررسی وضعیت حقوقی این کارگران به طور تصادفی ۱۶ کارگر از این گروه انتخاب شده و میانگین دستمزد آنها $\overline{X}$=۴۸۸۰۰ ریال محاسبه شد. آیا بر اساس اطلاعات بدست آمده می‌توانیم بگوییم که مدیر کارخانه به آنها ظلم کرده است؟ (خطای نوع اول α=۰/۰۵).

۱۸- مدیر جهاد کشاورزی ادعا می‌کند که ۶۰٪ جمعیت روستایی در فعالیت عمرانی مشارکت می‌کنند. برای اثبات ادعای خود ۱۰۰ خانوار روستایی انتخاب شده است و ۵۵ خانوار مشارکت فعال داشته اند. با توجه به نتیجه به دست آمده در مورد ادعای او چه قضاوتی می‌توان کرد؟ (خطای نوع اول را α=۰/۰۵ بگیرید).

۱۹- در یک دشت زراعی ، تجربه نشان داده است که ۱۵ درصد از چاه های موجود به علت وجود افت منابع آب زیرزمینی پس از ۲۰ سال فاقد آبدهی می‌شوند. مدیریت سازمان آب جدید با ایجاد شرایط ادعا می‌کند که از ۱۴۰ حلقه چاه حداکثر ۱۹ حلقه چاه دچار افت آبدهی می‌شوند. آیا در سطح ۵ درصد ادعای مدیریت جدید پذیرفته می‌شود؟

۲۰- اگر $\overline{X}$ و $\overline{Y}$ به ترتیب میانگین‌های نمونه‌های تصادفی به حجم $n_1=14$ و $n_2=18$ از جامعه‌های نرمال با میانگین μ_1، واریانس $\sigma_1^2=26$ و میانگین μ_2 و واریانس $\sigma_2^2=21$ باشند، چگونگی آزمون $H_{\circ}:\mu_1-\mu_2=c$ را در مقابل $H_1:\mu_1-\mu_2<c$ در سطح ۰/۰۲۵ برای $c=\circ$ و $c=1$ شرح دهید.

۲۱- دو تولیدکننده زراعی ادعا می‌کنند که بر اساس فرآیند جدید در تولید، میزان تولیدی نسبت به گذشته افزایش یافته است. در یک بررسی نمونه‌ای از این دو تولیدکننده، اطلاعات زیر بدست آمده است. آیا در سطح ۵ درصد می‌توان فرض $H_{\circ}:\mu_1=\mu_2$ را در مقابل $H_1:\mu_1\neq\mu_2$ پذیرفت؟

	تولید کننده اول	**تولید کننده دوم**
حجم نمونه	۱۰۰	۱۰۰
میانگین	۷۹۸	۸۲۶
واریانس	۷۹۸۲	۹۰۰۱

۲۲- در یک نمونه تصادفی از راندمان آبیاری، نمرات ۱۰ کشاورز عبارتند از:

۵۰ ۶۵ ۵۸ ۸۲ ۸۰ ۶۰ ۵۲ ۴۰ ۸۲ ۹۵

در سطح ۵ درصد فرض $H_{\circ}:\sigma=۱۷$ را در مقابل $H_1:\sigma>۱۷$ آزمون کنید.