آزمون فرض آماری

تسهیل کننده آموزشی :
هاشمی نژاد

آزمون فرض آماری

آزمون فرض

هر حکمی درباره توزیع جامعه یا پارامتر جامعه را یک فرض آماری می نامند و ممکن است درست یا نا درست باشد. درست یا نادرست بودن یک فرض، باید بر مبنای اطلاعات حاصل از نمونه گیری از جامعه بررسی شود. این عمل را آزمون فرض می نامیم.

چون ادعا ممکن است درست یا نادرست باشد، بنابرای دو فرض مکمل به وجود می آید: یکی برای آن که ادعا درست باشد و دیگری برای این که ادعا درست نباشد، بنابراین شروع یک آزمون فرض، همواره شامل دو فرض آماری می باشد که در مقابل یکدیگر قرار می گیرند.

آزمون فرض آماری

فرض صفر و فرض مقابل

در بحث آزمون فرض اغلب با فرض ها یا ادعاهایی در مورد پارامترهای توزیع جوامع آماری مواجه هستیم. به این فرض یا ادعاها فرض صفر گفته می شود و آن را با H_0 نشان می دهیم . فرض آماری که در مقابل فرض صفر قرار می گیرد (ناقض فرض صفر) را فرض مقابل نامیده و آن را با H_1 نشان می دهیم.

آزمون فرض آماری

در حالت کلی آزمون های فرض یه یکی از صورت های زیر بیان می شود:

$$۱)\begin{cases} H_۰: \theta = \theta_۰ \\ H_۱: \theta = \theta_۱ \end{cases}$$

$$۲)\begin{cases} H_۰: \theta = \theta_۰ \\ H_۱: \theta \neq \theta_۰ \end{cases}$$

$$۳)\begin{cases} H_۰: \theta \geq \theta_۰ \\ H_۱: \theta < \theta_۱ \end{cases}$$

$$۴)\begin{cases} H_۰: \theta \leq \theta_۰ \\ H_۱: \theta > \theta_۱ \end{cases}$$

آزمون فرض آماری

	ادعای اصلی			
	میانگین نمران ۷۵ است	میانگین نمرات ۷۵ نیست	میانگین نمرات حداقل ۷۵ است	میانگین نمرات بالای ۷۵ است
قدم ۱: شکل قراردادی ادعای اصلی	$\mu = ۷۵$	$\mu \neq ۷۵$	$\mu \geq ۷۵$	$\mu > ۷۵$
قدم ۲: شکل قراردادی که درست است وقتی ادعای اصلی غلط است	$\mu \neq ۷۵$	$\mu =۷۵$	$\mu < ۷۵$	$\mu \leq ۷۵$
قدم ۳: فرض صفر ($H۰$) (باید شامل تساوی باشد)	$H_{۰}: \mu = ۷۵$	$H_{۰}: \mu = ۷۵$	$H_{۰}: \mu \geq ۷۵$	$H_{۰}: \mu \leq ۷۵$
قدم ۴: فرض ($H۱$) (نمی تواند شامل تساوی باشد)	$H_{۱}: \mu \leq ۷۵$	$H_{۱}: \mu \leq ۷۵$	$H_{۱}: \mu \leq ۷۵$	$H_{۱}: \mu \leq ۷۵$

مثال : ادعای داده شده را بخوانید و H۰ و H۱ را تعیین کنید.

الف) میانگین سن استادان دانشگاه بیش تر از ۳۰ سال است.

ب) میانگین IQ دانشجویان دانشگاه حداقل ۱۰۰ است.

پ) میانگین هزینه ماهانه نگهداری یک هواپیما ۳۲۷۱ دلار است.

حل:

الف)

$$\begin{cases} H_0: \pi \leq 30 \\ H_1: \pi > 30 \end{cases}$$

$\pi>30$ شکل قراردادی ادعای اصلی

$\pi\leq30$ شکل قراردادی که درست است وقتی ادعای اصلی غلط است

آزمون فرض آماری

ب)

$$\begin{cases} H_0: \pi \ge 100 & \pi \ge 100 \text{ شکل قراردادی ادعای اصلی} \\ H_1: \pi < 100 & \pi < 100 \text{ شکل قراردادی که درست است وقتی ادعای اصلی غلط است} \end{cases}$$

پ)

$$\begin{cases} H_0: \pi = 3271 & \pi = 3271 \text{ شکل قراردادی ادعای اصلی} \\ H_1: \pi \ne 3271 & \pi \ne 3271 \text{ شکل قراردادی که درست است وقتی ادعای اصلی غلط است} \end{cases}$$

آزمون فرض آماری

آماره آزمون

برای هر پارامتر مورد آزمون می توان یک آماره آزمون براساس مشاهدات نمونه ای محاسبه کرد. با توجه به مقدار به دست آمده برای این آماره در نمونه تصادفی مشاهده شده، نسبت به رد یا پذیرش فرض H_0 تصمیم گیری خواهد کرد.

آماره آزمون یا دارای توزیع کاملا شناخته شده ای است (مانند Z، t، F، ...) و یا توزیع آن را می توان به دست آورد.

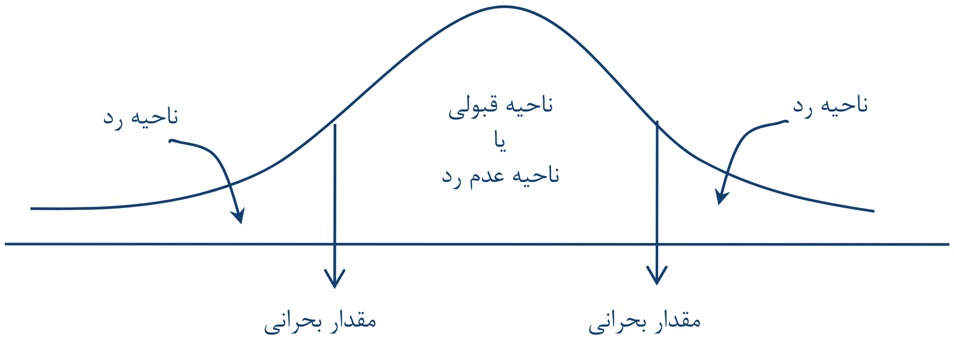
آزمون فرض آماری

جدول زیر آماره آزمون مناسب برای آزمون هر پارامتر را نشان می هد.

پارامترها	نمادآماری فرضیه	آماره	توزیع نمونه گیری	آماره آزمون
میانگین جامعه	$\pi_{٠}$	$\overline{X}$	نرمال یا t استیوانت	Z یا t
مقایسه میانگین دو جامعه	$\pi_1 - \pi_2$	$\overline{X}_1 - \overline{X}_2$	نرمال یا t استیوانت	Z یا t
نسبت موفقیت جامعه	$P_{٠}$	$\overline{P}$	نرمال	Z
مقایسه نسبت موفقیت دو جامعه	$P_{۱} - P_{۲}$	$\overline{P}_1 - \overline{P}_2$	نرمال	Z
واریانس جامعه	$\sigma^{۲}$	$\overline{S}$	مربع - کای	χ^2
مقایسه واریانس دو جامعه	$\frac{\sigma_1^2}{\sigma_2^2}$	$\frac{S_1^2}{S_2^2}$	فیشر	F

ناحیه رد و مقادیر بحرانی

ناحیه رد یک آزمون بر مبنای ساختار فرض مقابل (H_1) تعیین می گردد. توزیع آماره آزمون، مطابق شکل زیر به دو ناحیه تقسیم می شود، یکی از ناحیه رد H_0 و دیگر ناحیه قبول (عدم رد) H_0، ناحیه رد فرض H_0 را ناحیه بحرانی و مرز بین دو ناحیه رد و قبول را مقادیر بحرانی می نامیم. در واقع مقادیر بحرانی ناحیه رد را از قبول جدا می کند.



انواع خطاهای آماری

در انجام یک آزمون نهایتا تصمیمی را در مورد قبول یا رد فرض H_0 خواهیم گرفت.
قبول فرض H_0 از نظر ما معادل رد فرض H_1 و رد فرض H_0 معادل پذیرش فرض H_1 می باشد. بر مبنای تصمیمی که نسبت به رد یا قبول فرض H_0 می گیریم همواره احتمال بروز یکی از دو نوع خطای زیر وجود دارد.

خطای نوع اول(I)

رد فرض H_0 در حالی که فرض H_0 درست باشد، خطای نوع اول نامیده می شود.

احتمال متناظر با خطای نوع اول را با α نشان داده و به صورت زیرتعریف می کنیم:

$$\alpha = P(\text{خطای نوع اول}) = P(H_0 \text{ رد} \mid H_0 \text{ درست باشد})$$

خطای نوع اول α سطح معنی دار بودن یا سطح تشخیص نیز نامیده می شود.

خطای نوع دوم(II)

قبول فرض H_0 در حالی که فرض H_0 درست نباشد، خطای نوع دوم نامیده می شود.

احتمال متناظر با خطای نوع دوم را با β نشان داده و به صورت زیرتعریف می کنیم:

$$\beta = P\,(\text{خطای نوع دوم}) = P(H_0 \text{ رد} \mid H_0 \text{ درست باشد})$$

آزمون فرض آماری

نکات مهم:

۱- در هر آزمون فرض دو نوع خطا وجود دارد، خطای نوع اول و خطای نوع دوم.

۲- αوβ به یکدیگر وابسته بوده و بین آن ها رابطه معکوس وجود دارد. لذا اگر حجم نمونه انتخابی ثابت باشد، کاهش هر کدام از انواع خطاها باعث افزایش دیگری خواهد شد و بالعکس.

$$\lim_{\alpha\to 1}\beta=0 \quad , \quad \lim_{n\to 0}\beta=1$$

۳- با افزایش مقدار حجم نمونه (n) می توان هر دو خطا αوβ را به طور هم زمان کاهش داد.

$$\lim_{n\to\infty}\alpha=0 \quad , \quad \lim_{n\to\infty}\beta=0$$

آزمون فرض آماری

۴- اندازه ناحیه بحرانی، و در نتیجه احتمال ارتکاب خطای نوع اول (α) ، همیشه با تنظیم مقادیر بحرانی می تواند کاهش داده شود.

۵- اگر فرض صفر (H_0) نادرست باشد، در آن صورت مقدار β به حداکثر خود در صورتی خواهد رسید که مقدار واقعی پارامتر به عنوان فرض مقابل (H_1) انتخاب شده باشد.

آزمون فرض آماری

تصمیم گیری ها

قبلا گفتیم که ادعای اصلی گاهی فرض می شود و زمانی فرض مقابل. با این حال، روش، متلزم این است که ماهمواره فرض را آزمون کنیم. تصمیم گیری اصلی همواره به یکی از دو صورت زیر خواهد بود.

۱- فرض صفر $H_۰$ را رد می کنیم.

۲- فرض صفر $H_۰$ را رد نمی کنیم.

توان آزمون

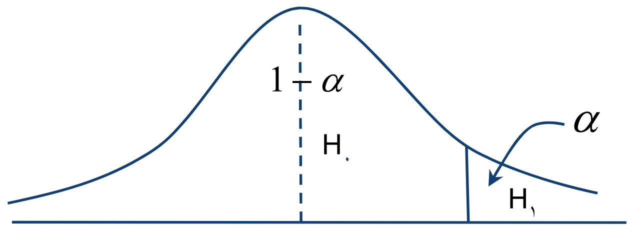
اگر βاحتمال خطای نوع دوم در یک آزمون باشد، آن گاه توان آزمون را برابر(β -۱) تعریف می کنند.

$$\text{توان آزمون} = ۱-\beta = ۱-P(H_۰ \text{ قبول} \mid H_۱ \text{ درست باشد})$$

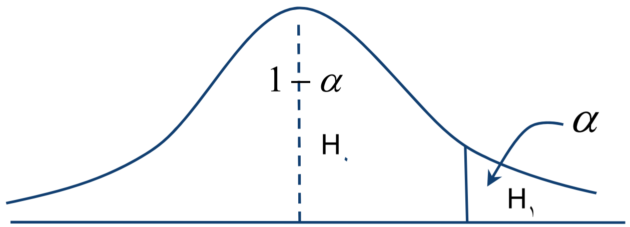
آزمونی را می توان توامند تر دانست که با توجه به مقدار مشخصی از α دارای خطای نوع دوم (β) کم تری باشد.

آزمون یک دامنه و دو دامنه

آزمون فرض را یک دامنه می نامیم اگر فرض مقابل آن (H_1) یک طرفه باشد، به عبارت دیگر آزمون به یکی از دو صورت زیر باشد:



$$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta < \theta_0 \end{cases}$$



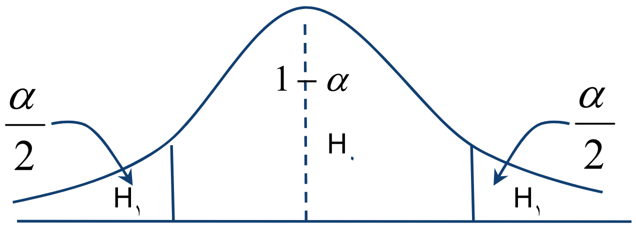
$$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$$

در چنین آزمونی، ناحیه بحرانی همواره در یک توزیع قرار دارد. ناحیه بحرانی برای فرض متقابل $\theta > \theta_0$ تماما در دنباله انتهایی سمت راست توزیع واقع است (راست دامنه) و حال آن که ناحیه بحرانی برای فرض مقابل $\theta < \theta_0$ تماما در سمت چپ توزیع قرار می گیرد (چپ دامنه)

آزمون فرض آماری

آزمون یک دامنه و دو دامنه

آزمون فرض را دو دامنه می نامیم اگر فرض مقابل دو طرفه باشد، به عبارت دیگر آزمون به صورت زیر باشد:



$$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases}$$

در این آزمون ناحیه بحرانی در دو انتهای توزیع قرار دارد. در مورد دو دامنه، α به دو قسمت مساوی بین دو دامنه تقسیم شده ناحیه بحرانی را تشکیل می دهد.

آزمون فرض آماری

برای هر یک از ادعاهای زیر، آزمون فرض را طبقه بندی کنید که راست دامنه یا چپ دامنه یا دو دامنه است؟

الف) میانگین IQ جنایتکاران بالای ۱۰۰ است.

ب) میانگین هزینه ماهانه نگهداری یک خودرو ۱۵۰۰۰ تومان است.

حل:

الف)

$$\begin{cases} H_0: \pi \le 100 \\ H_1: \pi > 100 \end{cases}$$

$\pi > 100$ شکل قراردادی ادعای اصلی

$\pi \le 100$ شکل قراردادی که درست است وقتی ادعای اصلی غلط است

آزمون فرض آماری

ب)

$$\begin{cases} H_0: \pi = 15000 \\ H_1: \pi \neq 15000 \end{cases}$$

$\pi > 15000$ شکل قراردادی ادعای اصلی

$\pi \leq 15000$ شکل قراردادی که درست است وقتی ادعای اصلی غلط است

بنابراین آزمون فرض را دو دامنه می نامیم زیرا H۱ دو طرفه است.

مراحل آزمون فرض

شروع

↓

ادعا یا فرض را که باید آزمون شود تعیین کرده آن را به شکل قراردادی بنویسید

↓

شکل قراردادی ای را بنویسید که وقتی ادعای اصلی غلط است باید درست باشد.

↓

از دو فرم قراردادی که تاکنون به دست آمده آن را که شامل شرط تساوی است H_0 و دیگری را H_1 بنامید.

↓

سطح معنی دار آزمون ، α را براساس اهمیت خطای نوع اول انتخاب کنید چنانچه رد یک H_1 درست مهم است، α را کوچک انتخاب . مقادیرمعمول عبارتند از ۰/۰۵ و ۰/۰۱

↓

چه آماره ای این آزمون مناسب است و توزیع نمونه گیری چیست؟

↓

آماره آزمون ، ناحیه آزمون و مقدار بحرانی را تعیین کنید. نموداری رسم کنید و در آن آماره آزمون، مقدار (مقادیر) بحرانی و ناحیه بحرانی را بگنجانید.

↓

اگر آماره آزمون در ناحیه بحرانی قرار بگیرد H_0 را رد کنید. اگر آماره آزمون در ناحیه بحرانی نیست، فرض صفر را رد نکنید.

↓

تصمیم قبلی را با عبارت ساده و غیر فنی بیان کنید.

↓

پایان

آزمون فرض برای میانگین یک جامعه

آزمون یکسان بودن میانگین با عدد ثابت وقتی که انحراف معیار معلوم است

۱) فرضیه آماری برای میانگین جامعه μبه صورت زیر در نظر گرفته می شود.

$$\begin{cases} H_{0}:\pi=\pi_{0} \\ H_{1}:\pi\neq\pi_{0} \end{cases} \qquad \begin{cases} H_{0}:\pi=\pi_{0} \\ H_{1}:\pi<\pi_{0} \end{cases} \qquad \begin{cases} H_{0}:\pi=\pi_{0} \\ H_{1}:\pi>\pi_{0} \end{cases}$$

۲) اگر نمونه از جامعه ای نرمال با انحراف معیار معلوم انتخاب شود. توزیع$\overline{X}$صرف نظر

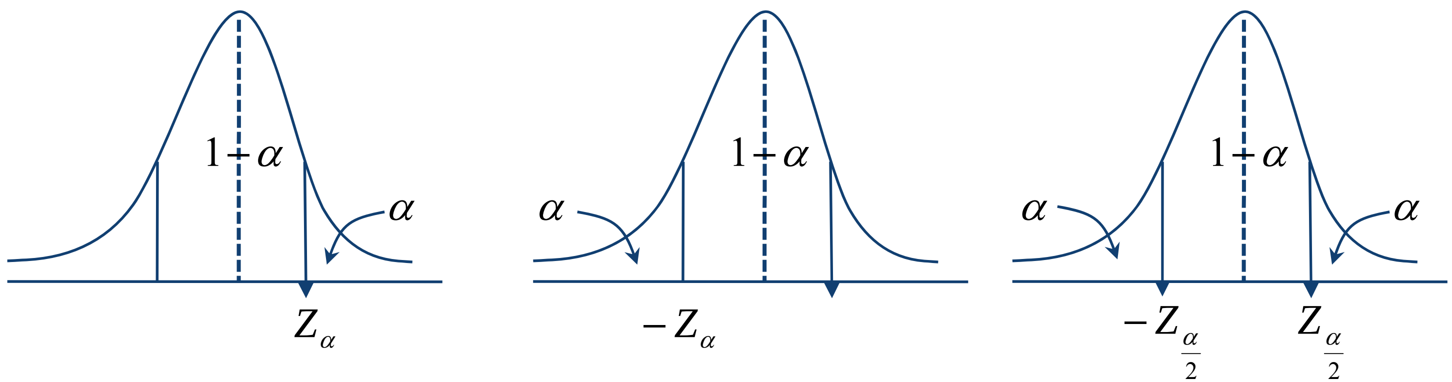
از حجم نمونه نرمال است و در نتیجه آماره آزمون Z خواهد بود.

$$Z=\frac{\overline{X}-\pi_{0}}{\frac{\sigma}{\sqrt{n}}}$$

آزمون فرض آماری

در رابطه بالا μ_0 میانگین مورد آزمون و مخرج متغیر Z همان $\sigma_{\overline{X}}$ است.

۳) ناحیه بحرانی با توجه به فرض مقابل H۱ تعیین می شود.



$H_1: \pi > \pi_0$ — ناحیه بحرانی : $Z>Z_\alpha$

$H_1: \pi > \pi_0$ — ناحیه بحرانی : $Z>-Z_\alpha$

$H_1: \pi \neq \pi_0$ — ناحیه بحرانی : $Z> Z_{\frac{\alpha}{2}}, Z<Z_\alpha$

آزمون فرض آماری

۴) اگر مقدار عددی آماره آزمون در ناحیه بحرانی قرار گیرد، فرض H_0 رد و فرض H_1 پذیرفته می شود. یعنی مقدار μ_0را در سطح تشخیص α نمی توان به عنوان میانگین جامعه پذیرفت. اگر مقدار عددی آماره آزمون در ناحیه حرانی قرار نگیرد، دلیلی برای رد کردن فرض H_0 نخواهم داشت، یعنی در نمونه دلیلی بر رد μ_0 به عنوان میانگین جامعه وجود ندارد و آن را می توان پذیرفت .

ادعا شده است که میانگین برق مصرفی فروردین ماه یک ناحیه تهران دست کم ۱۳۰۰ کیلوات ساعت بوده است. بدین منظور یک نمونه تصادفی به تعداد ۴۰۰ خانوار از منطقه انتخاب شده که میانگین و انحراف معیار برق مصرفی آن ها ۱۲۵۲ و ۲۵۷ کیلوات ساعت است. سطح خطای ۱٪ را در نظر گرفته و صحت ادعا را بررسی کنید.

حل:

$$\begin{cases} H_{۰}: \pi \geq ۱۳۰۰ \\ H_{۱}: \pi < ۱۳۰۰ \end{cases} , \alpha = ۰/۰۱$$

آزمون فرض آماری

چون n>۳۰ است، توزیع $\overline{X}$ از تقریب نرمال برخوردار است ، پس :

$$Z=\frac{\overline{X}-\pi_0}{\frac{\sigma}{\sqrt{n}}}=\frac{۱۲۵۲-۱۳۰۰}{\frac{۲۵۷}{\sqrt{۴۰۰}}}=-۳/۷۵$$

$$Z_\alpha=Z_{۰/۰۱}=-۲/۳۲۵$$

چون Z<Zα بنابراین مقدار عددی آماره آزمون در ناحیه بحرانی قرار می گیرد، لذا فرض H۰ رد و فرض H۱ پذیرفته می شود. یعنی میانگین برق مصرفی فروردین ماه یک ناحیه تهران کم تر از ۱۳۰۰ کیلوات ساعت است.

آزمون فرض نمونه های جفت شده (مقایسه زوج ها)

در آزمون برای میانگین دو جامعه فرض براین است که نمونه های هر گروه مستقل از یکدیگر هستند. آزمون فرضیه ای که برمبنای نمونه های غیر مستقل (وابسته) قرار دارد به آزمون نمونه های جفت شده معروف است. در چنین مواردی به طریق زیر عمل می کنیم.

هنگامی که با دو نمونه وابسته سر و کار داریم استخراج کردن و ثبت $\overline{X}_1$، S_1، n_1، $\overline{X}_2$، S_2، n_2 کار بیهوده ای است، زیرا ارتباط بین مقادیر جور شده ضایع می گردد. در عوض، اختلاف های (d) بین جفت داده ها را مطابق جدول زیر حساب می کنیم.

آزمون فرض آماری

زوج	رفتار۲ (x_i)	رفتار۲ (y_i)	تفاضل $D_i=x_i-y_i$
۱	$x_۱$	$y_۱$	$D_۱=x_۱-y_۱$
۲	$x_۲$	$y_۲$	$D_۲=x_۲-y_۲$
⋮	⋮	⋮	⋮
n	x_n	y_n	$D_n=x_n-y_n$

توزیع d_i نرمال خواهد بود. مقادیر $\bar{d}$ و S_d را برای نمونه محاسبه می کنیم.

$\bar{d}$ میانگین مقادیر d، S_d انحراف معیار مقادیر d و n تعداد داده های جفت شده .

$$\bar{d} = \frac{\sum d_i}{n} \quad ، \quad S_d^2 = \frac{\sum \left(d_i - \bar{d}\right)^2}{n-1}$$

آزمون فرض آماری

فرضیه آماری برای داده های جفت شده به صورت زیر در نظر گرفته می شود.

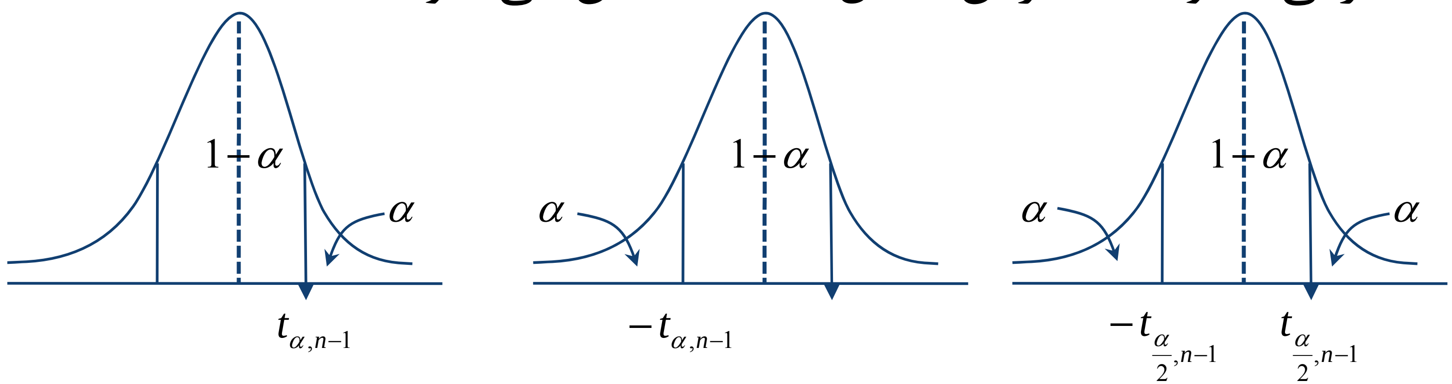
$$\begin{cases} H_0 : \pi_d = 0 \\ H_1 : \pi_d > 0 \end{cases} \qquad \begin{cases} H_0 : \pi_d = 0 \\ H_1 : \pi_d < 0 \end{cases} \qquad \begin{cases} H_0 : \pi_d = 0 \\ H_1 : \pi_d \neq 0 \end{cases}$$

در نمونه گیری تصادفی از دو جامعه نرمال و وابسته که در آن ها μ_d میانگین داده های جفت شده است، آماره آزمون زیر دارای توزیع t استیوانت با n-۱ درجه آزادی می باشد.

$$t = \frac{\bar{d}}{\frac{S_d}{\sqrt{n}}}$$

آزمون فرض آماری

۳) ناحیه بحرانی با توجه به فرض مقابل (H۱) تعیین می شود.



$H_1:\pi_d>0$ $H_1:\pi_d<0$ $H_1:\pi_d\neq 0$

ناحیه بحرانی : $t>t_{\alpha\ ,n-1}$

ناحیه بحرانی : $t<-t_{\alpha\ ,n-1}$

ناحیه بحرانی : $t> t_{\frac{\alpha}{2},n-1}$, $t<t_{\alpha}$

۴) اگر مقدار عددی آماره آزمون در ناحیه بحرانی قرار گیرد، فرض H۰ رد و فرض H۱ پذیرفته می شود یعنی در سطح تشخیص α اختلاف بین مقادیر x و y قبل و بعد از دوره معنی داری می باشد و در غیر این صورت مقادیر x و y قبل و بعد از دوره تغییری نکرده است.

آزمون فرض آماری

در تئوری های مدیریت بیان شده که مهمترین انگیزه « پول» است. یکی از دانشجویان مدیریت میزان رضایت شغلی ۷ کارمند را قبل از افزایش حقوق ماهانه و بعد از آن اندازه گیری و در این جدول خلاصه کرده است:

کارمند	۱	۲	۳	۴	۵	۶	۷
قبل از افزایش	۵۰	۷۰	۸۰	۶۰	۵۵	۵۰	۶۸
بعد از افزایش	۶۰	۷۵	۸۰	۷۰	۶۵	۶۲	۷۰

فرض نرمال بودن میزان انگیزه را پذیرفته، صحت ادعا را در سطح خطای یک در صد آزمون کنید.

حل:

$$\begin{cases} H_{۰}: \pi d \le ۰ \\ \qquad\qquad\qquad , \alpha = ۰/۰۱ \\ H_{۱}: \pi d > ۰ \end{cases}$$

زوج	۱	۲	۳	۴	۵	۶	۷
x_i	۵۰	۷۰	۸۰	۶۰	۵۵	۵۰	۶۸
y_i	۶۰	۷۵	۸۰	۷۰	۶۵	۶۲	۷۰
$d_i=y_i-x_i$	۱۰	۵	۰	۱۰	۱۰	۱۲	۲

$$\bar{d} = \frac{\sum d_i}{n} = \frac{۱۰ + ۵ + \ldots + ۲}{۷} = ۷$$

$$S_d = \sqrt{\frac{\sum \left(di - \bar{d}\right)^{۲}}{n-۱}} = \sqrt{\frac{\left(۱۰ - ۷\right)^{۲} + \left(۵ - ۷\right)^{۲} + \ldots \left(۲ - ۷\right)^{۲}}{۷ - ۱}}$$

$$t = \frac{\bar{d}}{\frac{S_d}{\sqrt{n}}} = \frac{۷}{\frac{۴/۶۵۵}{\sqrt{۷}}} = ۳/۹۸$$

$$t_{\alpha, n-1} = t_{۰/۰۱, ۶} = ۳/۱۴۳$$

آزمون فرض آماری

چون ۳/۹۸>۳/۱۴۳ بنابراین مقدار عددی آماره آزمون در ناحیه بحرانی قرار می گیرد، لذا فرض H_0 رد و فرض H_1 پذیرفته می شود. یعنی افزایش حقوق باعث رضایت شغلی کارمندان می شود.

آزمون فرض آماری

آزمون فرض نسبت موفقیت در یک جامعه

$$\begin{cases} H_0: P = P_0 \\ H_1: P \neq P_0 \end{cases} \qquad \begin{cases} H_0: P \geq P_0 \\ H_1: P < P_0 \end{cases} \qquad \begin{cases} H_0: P \leq P_0 \\ H_1: P > P_0 \end{cases}$$

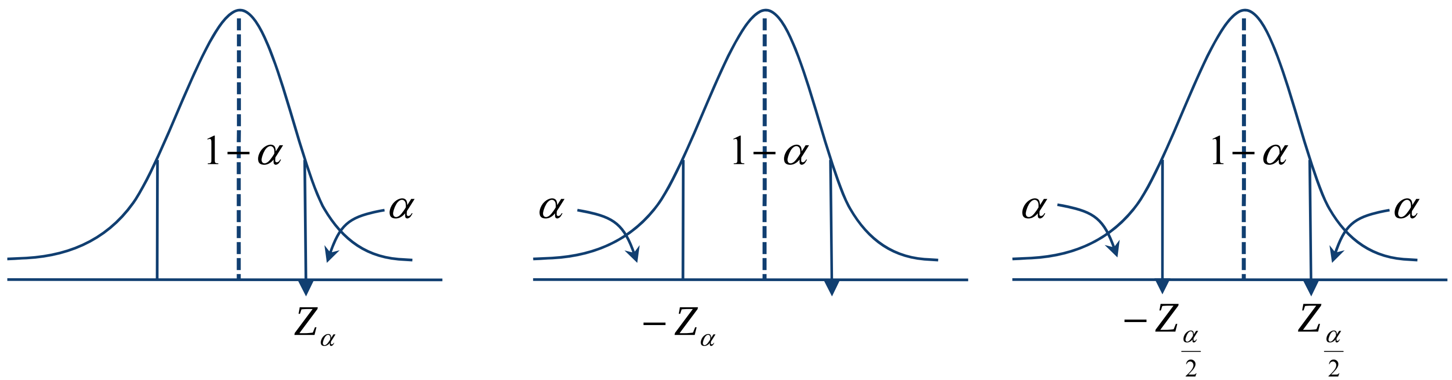
توزیع نمونه گیری $\bar{d}$ ،نرمال است و آماره آزمون Z است.

$$Z = \frac{\bar{P} - P_0}{\sqrt{\frac{P_0(1 - P_0)}{n}}}$$

در رابطه بالا P_0، نسبت مورد آزمون و مخرج متغیر Z همان $\sigma_{\bar{P}}$ است.

آزمون فرض آماری

۳) ناحیه بحرانی با توجه به فرض مقابل (H۱) تعیین می شود.



$H_1 : P > P_0$ ناحیه بحرانی : $Z>Z_\alpha$

$H_1 : P < P_0$ ناحیه بحرانی : $Z>-Z_\alpha$

$H_1 : P \neq P_0$ ناحیه بحرانی : $Z> Z_{\frac{\alpha}{2}}, Z<Z_\alpha$

۴) اگر مقدار عددی آماره آزمون در ناحیه بحرانی قرار گیرد، فرض H۰ رد و فرض H۱ پذیرفته می شود یعنی عدد P۰ را در سطح تشخیص α نمی توان به عنوان نسبت جامعه قبول کرد. اگر مقداری عددی آماره آزمون در ناحیه بحرانی قرار نگیرد، فرض H۰ رد نمی شود، یعنی در نمونه دلیلی بر رد عدد H۰ به عنوان نسبت جامعه وجود ندارد.

ادعا شده است که وضع بهداشتی خوابگاه دانشجویی نامناسب است. بدین منظور یک نمونه تصادفی ۱۵۰ نفره از بین دانشجویان به طور تصادفی انتخاب شده است که ۵۰ نفر از آن ها از وضع بهداشت خوابگاه ها شکایت داشته اند. در سطح خطای ۲ در صد محاسبه کنید آیا می توان گفت ادعای فوق درست است؟

حل:

$$n = ۱۵۰, X = ۵۰, \overline{\mathrm{P}} = \frac{X}{n} = \frac{۵۰}{۱۵۰} = ۰/۳۳۳$$

$$\begin{cases} H_۰ : P \leq ۰ \\ H_۱ : P > ۰ \end{cases}, \alpha = ۰/۰۲$$

آزمون فرض آماری

$$Z = \frac{\overline{P} - P_{\cdot}}{\sqrt{\frac{P_{\cdot}\left(1 - P_{\cdot}\right)}{n}}} = \frac{۰/۳۳۳ - ۰}{\sqrt{\frac{۰\left(۱ - ۰\right)}{۱۵۰}}}$$

$$Z_{\alpha} = Z_{۰/۰۲} = ۲/۰۵۵$$

بنابراین مقدار عدد آماره آزمون در ناحیه بحرانی قرار می گیرد، لذا فرض H۰ رد و فرض H۱ پذیرفته می شود.یعنی وضع بهداشتی خوابگاه مناسب است.

آزمون نیکویی برازش – آزمون χ^2 ساده

هدف ما آزمون معنی دار بودن اختلاف بین فراوانی های مشاهده شده و فراوانی هایی است که از نظر تئوری انتظار داریم. چون آزمون می کنیم که تا چند اندازه توزیع فراوانی مشاهده شده (فراوانی تجربی) بر توزیع فراوانی مورد انتظار (فراوانی نظری) منطبق می شود (یا برازنده است) این روش اغلب به عنوان آزمون نیکویی برازش معروف است.

۱) فرضیه آماری به صورت زیر است:

H_0: فراوانی های تجربی و نظری یکسان هستند

H_1 : فراوانی های تجربی و نظری یکسان نیستند

آزمون فرض آماری

۲) آماره آزمون نیکویی برازش به صورت زیر است:

$$\chi^۲ = \sum_{i=۱}^{k} \frac{(O_i - e_i)^۲}{e_i}$$

که در آن Oi و eiبه ترتیب فراوانی های مشاهده شده (فراوانی تجربی و مورد فراوانی های مورد انتظار (فراوانی نظری) می باشند و دارای توزیع مربع کای با (k-m-۱) درجه آزادی می باشد، که در آن k تعداد جملات در فرمول و m تعداد پارامترهای مستقلی است که بر مبنای داده های نمونه ای برآورد می شود.

۳) اگر فراوانی های مشاهده شده نزدیک به مقادیر انتظاری مربوطه باشند، مقدار $\chi^۲$کوچک بوده و آن خود نشانه خوبی انطباق (یا برازش) است. اگر مقادیر مشاهده

آزمون فرض آماری

به مقدار قابل ملاحظه ای از مقادیر انتظاری دور باشند، مقدار χ^2 بزرگ بوده و انطباق (یا برازش) ضعیف خواهد بود. خوبی برازش (مقدار χ^2 کوچک) منجر به رد نکردن، H_0 شده و حال آن که ضعیف بودن برازش (مقدار χ^2 بزرگ) منتهی به رد H_0 خواهد شد. در نتیجه ناحیه بحرانی در سمت راست توزیع مربع کای واقع خواهد شد(راست دامنه است)

آزمون فرض آماری

مثال: تاسی را ۱۲۰ با پرتاب می کنیم، نتایج زیر مشاهده شده است.

روی تاس	۱	۲	۳	۴	۵	۶
مشاهده شده	۲۰	۲۲	۱۷	۱۸	۱۹	۲۴

آیا نتایج مشاهده با مقادیر مورد انتظار با احتمال ۹۵٪ مطابقت دارد یا خیر؟ به عبارت دیگر با احتمال ۹۵٪ تاس را می توان سالم فرض کردیا خیر؟

حل:

$\alpha = ۰/۰۵$,

$H_۰$: فراوانی های تجربی و نظری یکسان هستند (تاس سالم می باشد)

$H_۱$: فراوانی های تجربی و نظری یکسان نیستند (تاس سالم نیست)

حل: از نظر تنوری اگر تاس سالم باشد انتظار خواهیم داشت که هر کدام از اعداد ۱ تا ۶ را ۲۰ بار نشان دهد $\left(۱۲۰ \times \frac{۱}{۶} = ۲۰\right)$.

فراواني تجربي Oi	۲۴	۱۹	۱۸	۱۷	۲۲	۲۰
فراواني تجربي ei	۲۰	۲۰	۲۰	۲۰	۲۰	۲۰
ei-Oi	۴	-۱	-۲	-۳	۲	۰

$$\chi^{۲} = \sum \frac{(Oi - ei)^{۲}}{ei} = \frac{۰ + ۴ + ۹ + ۴ + ۱ + ۱۶}{۲۰} = ۱/۷$$

$$= k - m - ۱ = ۶ - ۱ = ۵$$

$$\chi^{۲}_{۰/۰۵,۵} = ۱۱/۷$$

چون ۱/۷<۱۱/۰۵بنابراین $\chi^{۲} < \chi^{۲}_{۰/۰۵,۵}$ یعنی مقدار عددی آماره آزمون در ناحیه بحرانی قرار نمی گیرد. لذا فرض $H_۰$ رد نمی شود و نتیجه گیری می کنیم فراوانی های تجربی و نظری یکسان هستند (توزیع یکنواخت است) به عبارتی تاس سالم است.

آزمون استقلال – آزمون χ^2 مضاعف

در قسمت طبقه بندی فراوانی تجربی و نظری با یک متغیر مطرح گردید. در مواردی لازم است طبقه بندی با دو متغیر مورد بررسی قرار بگیرد که به آن χ^2 مضاعف می گویند. در آزمون χ^2 مضاعف با جداول توافقی در تماس هستیم. جدول توافقی به عنوان رابطه مفیدی برای تحلیل وابستگی یک متغیر با متغیر دیگر به کار می رود. این فقط یک وابستگی آماری است که نمی تواند به عنوان رابطع علت و معلولی ذاتی استفاده شود.

در حال کلی یک جدول توافقی $r \times c$ دارای r سطر که معرف حالت های مختلف x_1، x_2، ... xr، از یک صفت متغیر و دارای c ستون که معرف حالت های مختلف y_1،y_2، ... y_n از یک صفت متغیر دیگر است.

H_0 :دو متغیر تصادفی مورد سئوال مستقل هستند

H_1 :دو متغیر تصادفی مورد سئوال مستقل نیستند

آزمون فرض آماری

۲)آماره آزمون مستقل به صورت زیر است:

$$\chi^{۲} = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(o_{ij}-e_{ij}\right)^2}{e_{ij}}$$

که در آن O_{ij} فراوانی خانه مشاهده شده (فراوانی تجربی) و e_{ij} فراوانی مورد انتظار (فراوانی نظری) برای خانه ای که در محل تلاقی سطر iام و ستون jام قرار دارد.

برای هر خانه در جدول، فراوانی مورد انتظار e_{ij} می تواند به طریق زیر تعیین گردد.

$$e_{ij} = \frac{\text{(جمع سطر jام)(جمع سطر iام)}}{\text{جمع کل}}$$

در یک جدول توافقی با rسطر و c ستون، درج آزادی به صورت زیر است:

درجه آزادی =(r−۱)(c−۱)