برآورد فاصله ای و آزمون فرضیه های آماری میانگین جامعه.

توزیع های نمونه ای:

در این بخش توزیع های نمونه ای؛ برخی ویژگیها و کاربرد آنها در یافتن برآورد های فاصله ای و انجام آزمون های آماری بیان می شود.

۱-توزیع نرمال

توزیع نرمال در فصل قبل معرفی و برخی ویژگیهای مفید آن بیان شد. در این بخش کاربردهایی از آن در برآورد یابی فاصله ای و آزمون فرض های آماری معرفی می شود.

تعریف (نمونه تصادفی) یک نمونه تصادفی شمال n متغیر تصادفی مستقل و هم توزیع $X_1, X_2, ..., X_n$ است.

بدیهی است که در یک نمونه تصادفی:

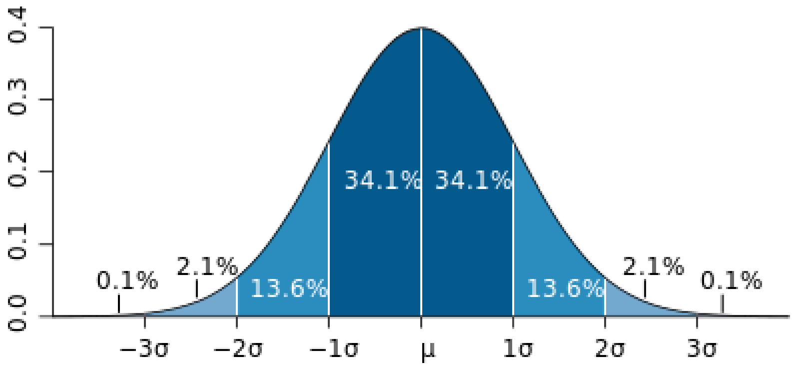
$$\begin{cases} EX_i = \mu, \quad i = 1,2,....,n \\ Var(X_i) = \sigma^2 \end{cases}$$

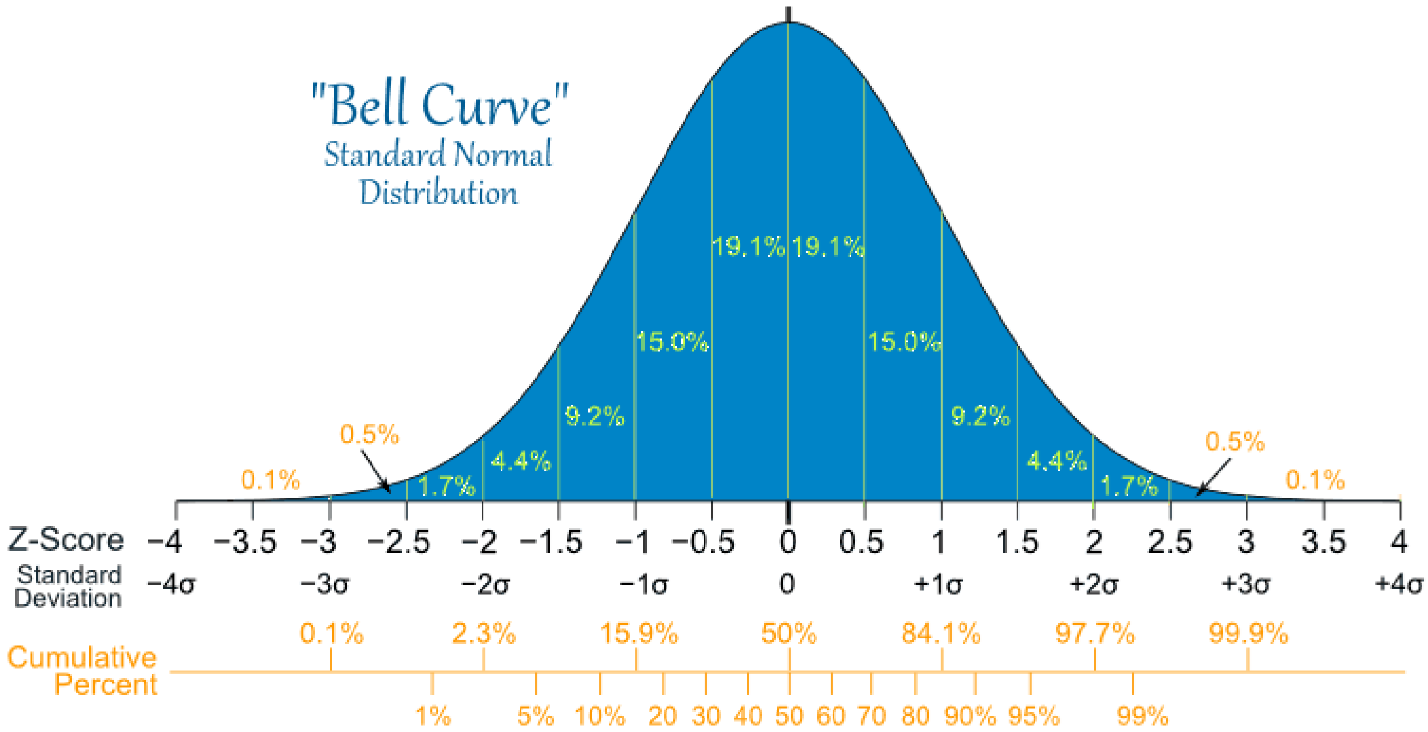
میانگین نمونه و واریانس نمونه عبارتند از:

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

قضیه حد مرکزی: اگر $X_1, X_2, ..., X_n$ یک نمونه تصادفی از جامعه ای با میانگین μ و واریانس σ^2 با شد انگاه $Z_n = \frac{\sqrt{n}(\overline{X} - \mu)}{\sigma}$ برای مقادیر به اندازه کافی بزرگ n دارای توزیع نرمان استاندارد است.

قضیه فوق بیان می کند که میانگین نمونه به طور مجانبی دارای توریع نرمال است به این دلیل در تحلیل های آماری وقتی حجم نمونه به اندازه کافی بزرگ باشد توزیع جامعه را نرمال در نظر می گیرند.

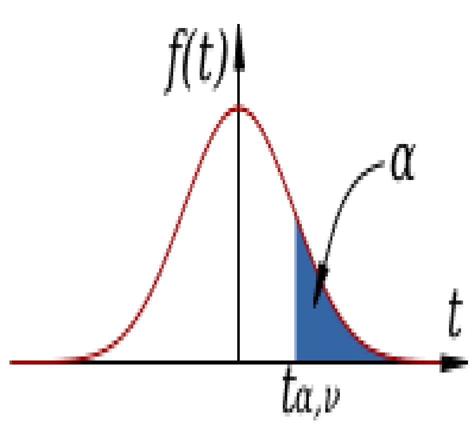


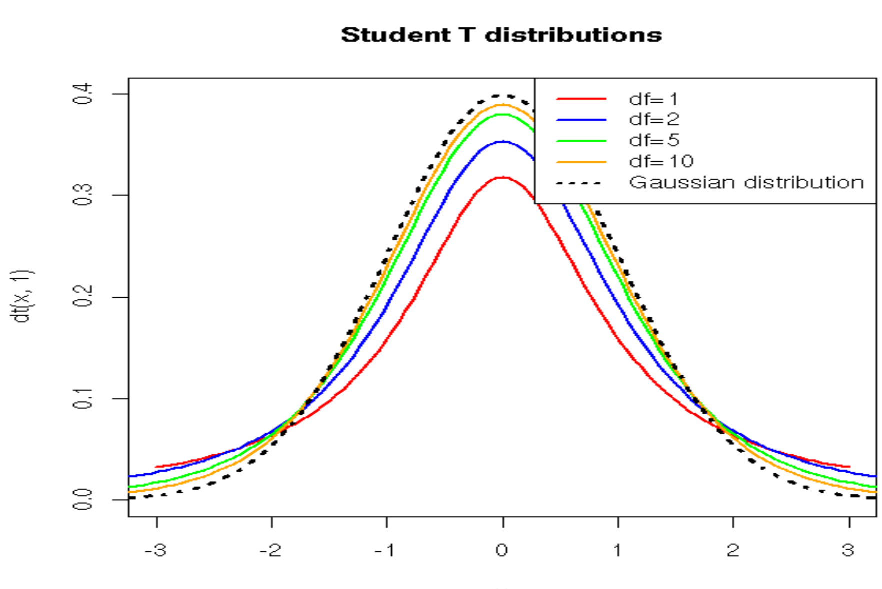


۲-توزیع تی استودنت:

اگر $X_1, X_2, ..., X_n$ یک نمونه تصادفی از جامعه نرمال با میانگین μ و واریانس $\sigma^2 < \infty$ با شد آنگاه برای هر n ؛ $T_n = \frac{\sqrt{n}(\bar{X}-\mu)}{s}$ یک متغیر تصادفی با توزیع به نام توزیع تی استودنت با درجه آزادی $n-1$ می باشد. این توزیع شبیه توزیع نرمال حول صفر متقارن بوده و برای مقادیر $n \geq 30$ تقریبا منطبق بر نر مال استاندارد می شود. در منابع از نماد زیر برای نمایش توزیع تی استودنت استفاده می شود.

$$T_n \to T(n-1)$$

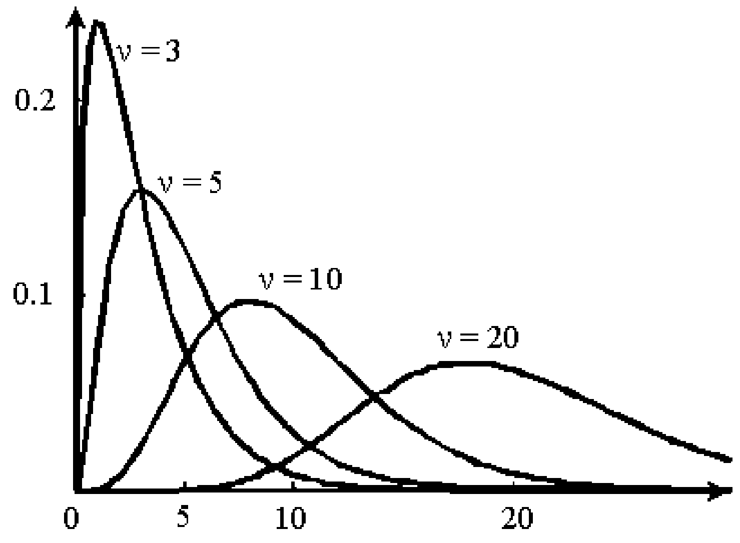


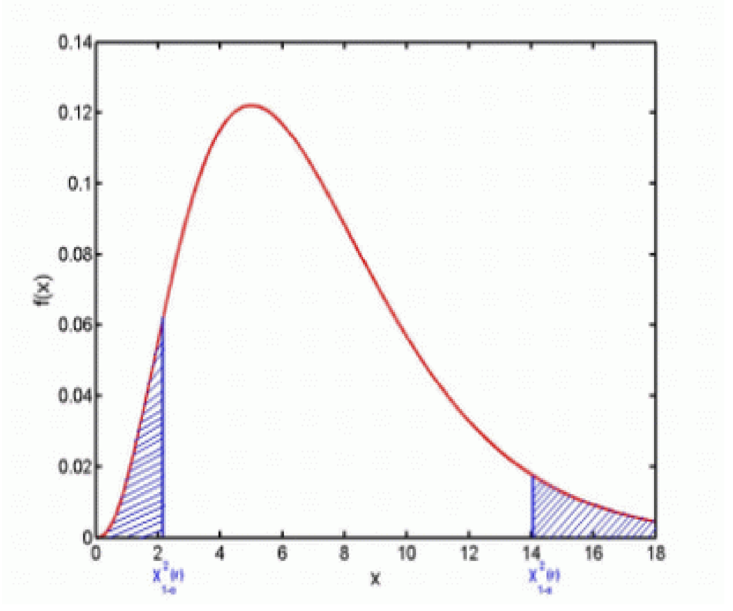


در ادامه برخی از کاربردهای توزیع نرمال و تی استودنت در تحلیل داده های آماری بیان می شود.

۳-توزیع کای دو:

اگر $X_1, X_2, ..., X_n$ یک نمونه تصادفی از جامعه نرمال با میانگین μ و واریانس $\sigma^2 < \infty$ با شد انگاه برای هر n ؛ $X^2 = \frac{\sqrt{n-1}.S^2}{\sigma^2} \to X(n-1)$ یک متغیر با توزیع به نام توزیع کای دو با درجه آزادی $n-1$ می باشد. این توزیع نا متقارن است.





نکته: یکی از کاربرد های این توزیع برای برآورد فاصله ای و آزمون فرضیه در مورد واریانس یا انحراف معیار می باشد.

۱. مبانی آمار استنباطی، مفهوم و اهداف آن
۲. روش‌های آمار استنباط (برآورد نقطه‌ای، برآورد فاصله‌ای، آزمون فرضیه‌های آماری)
۳. استنباط نقطه‌ای درباره میانگین، نسبت، واریانس و ...

روش‌های آماری برای تجزیه و تحلیل داده‌ها به دو دسته‌ی کلی روش‌های آمار توصیفی و روش‌های آمار استنباطی دسته‌بندی می‌شوند. روش های آمار توصیفی شامل روش‌های جمع آوری داده‌ها، پردازش اولیه داده‌ها، دسته‌بندی و خلاًصه کردن و همچنین رسم نمودارها و جداول مختلف آماری به همراه محاسبه شاخص‌های آماری تمایل مرکزی و پراکندگی است. در فصل قبلی با مفاهیم اولیه آمار و همچنین مسائل مرتبط با آمار توصیفی آشنا شده‌اید. در این فصل به تشریح گزیده‌ای از روش‌های آمار استنباطی و کاربرد آن‌ها در مسایل مختلف علوم زمین می پردازیم.

اصولاً آماراستنباط مبتنی بر تجزیه و تحلیل داده‌های جمع آوری شده از روش‌های نمونه‌گیری است. با استفاده از روش‌های مختلف استنباط آماری می‌توان روابط مختلف بین متغیرهای وابسته و مستقل را بررسی نموده و از روی این روابط و یا برازش مدل‌های مختلف به داده‌ها برای آینده نتیجه‌گیری نمود. همچنین با استفاده از این روش‌ها و با در نظر گرفتن سطح اطمینان معین، می‌توان نتایج بدست آمده از نمونه تصادفی را به جامعه تعمیم داد.

روش‌های آمار استنباطی عمدتاً شامل آزمون فرضیه‌های آماری و برازش مدل به داده‌ها است. گاهی اوقات نیز می‌توان روش برآورد فاصله‌ای را به عنوان یکی از روش‌های آمار استنباطی به شمار آورد.

وقتی براساس نمونه تصادفی تقریبی از میانگین، انحراف استاندارد یا واریانس و دیگر شاخص‌های آماری را بدست می‌آوریم، در واقع برآورد نقطه‌ای از آن شاخص‌ها بر اساس نمونه تصادفی را محاسبه می‌کنیم. واضح است که اگر نمونه تصادفی تغییر یابد، این برآوردها نیز عوض خواهند شد. لذا اگر بخواهیم با داشتن فقط یک نمونه تصادفی در مورد جامعه آماری و شاخص‌های آن استنباطی داشته باشیم، می‌توان از روش برآورد دیگری با نام برآورد فاصله‌ای استفاده نمود. در برآورد فاصله‌ای در واقع با درنظر گرفتن یک سطح اطمینان معین، فاصله‌ای را به عنوان حد بالا و پایین برای مقدار شاخص آماری در جامعه ارایه می‌کنیم.

برآورد فاصله ای:

اگر $X_1, X_2,....,X_n$ یک نمونه تصادفی از جامعه ای با پارامتر مجهول θ باشد. یک برآورد کننده نقطه ای از θ تابعی از نمونه تصادفی است که هر مقدار آن یک تخمین برای θ باشد. به دلیل اینکه میزان خطای برآورد نقطه ای قابل مقایسه نمی باشد؛ فاصله ای به مرکزیت برآورد نقطه ای طوری تعیین می شود که با احتمال زیاد مقدار واقعی θ را در بر داشته باشد در این صورت طول بازه به عنوان یک معیار دقت برآورد به کار برده می شود. یعنی اگر $T(X_1, X_2,....X_n)$ یک برآورد گر نقطه ای برای پارامتر θ با شد آنگاه $(T(.)-E, T(.)+E)$ یک برآورد فاصله ای برای θ می باشد که در آن E به عنوان خطا در نظر گرفته میشود. بدیهی است که هر چه طول بازه کوچکتر باشد برآورد کننده از دقت بیشتری برخوردار است. اگر توزیع برآورد کننده معلوم باشد برآورد فاصله ای تعیین می شود.

مثال ۱: در یک بررسی آماری بزرگی ۳۲ زمین لرزه ی اتفاق افتاده در یک منطقه به صورت زیر ثبت شده است .

۴,۷ ,۴,۶ ,۵,۳ ,۴,۸ ,۴,۴ ,۴,۵ ,۴,۷ ,۴,۷ ,۴,۱ ,۴,۴ ,۳,۱ , ۴,۶ ,۴,۵ ,۵ ,۵,۳ ,۵,۵ ,۶,۸ ,۴,۵

-۳,۸ ,۴,۲ ,۳,۶ ,۳,۸ ,۴,۶ ,۵,۵ ,۴ ,۴,۶ ,۴,۲ ,۵ ,۵,۳ ,۴,۲ ,۴,۹ ,۴,۳ ,

مطلوبست الف: برآورد فاصله ای ۹۵ درصد برای متوسط بزرگی زمین لرزه ها در این منطقه را بدست آورید.

ب: آیا در سطح ۵ درصد می توان گفت میانگین بزرگی زمین لرز ه های اتفاق افتاده در این منطقه برابر ۵ ریشتر است.

برآورد فاصله ای میانگین: Confidence interval

الف- اگر $\sigma=\sigma_0$ آنگاه

$$(\bar{X}-Z_{\frac{\alpha}{2}}\frac{\sigma_0}{\sqrt{n}},\ \ \bar{X}+Z_{\frac{\alpha}{2}}\frac{\sigma_0}{\sqrt{n}})$$

یک برآورد فاصله ای با ضریب اطمینان $1-\alpha$ برای میانگین جامعه است.

در این حالت حجم نمونه لازم برای براورد میانگین جامعه با میزان دقت حداکثر d از رابطه زیر تعیین می شود.

$$n \ge \frac{4Z^2_{\frac{\alpha}{2}}.\sigma_0^2}{d^2}$$

ب- اگر واریانس مجهول باشد در این صورت از اینکه $T_n=\frac{\sqrt{n}(\bar{X}-\mu)}{s}\to T(n-1),$ برآورد فاصله ای با ضریب $1-\alpha$ به صورت زیر به دست می آید و همچنین حجم نمونه لازم برای برآورد میانگین جامعه با حداکثر خطای d از رابطه زیر محاسبه می شود.

$$(\bar{X}-t_{\frac{\alpha}{2}}(n-1)\frac{\sigma_0}{\sqrt{n}},\ \ \bar{X}+t_{\frac{\alpha}{2}}(n-1)\frac{\sigma_0}{\sqrt{n}})$$

$n=\max\{n_0,\ n_1\}$ **که در آن** n_0 **اندازه یک نمونه اولیه و** $n_1=\left[\frac{4S_0^2.t^2_{\frac{\alpha}{2}}(n_0-1)}{d^2}\right]$

نکته: اگر توزیع جامعه مورد مطالعه نرمال نباشد ولی $n\ge 30$ نتایج فوق معتبر هستند.

آزمون فرضیه های آماری: Statistical Hypothesis Test

هدف: تصمیم گیری درمورد یک ادعا بر اساس اطلاعات حاصل از نمونه.

فرضیه آماری:

مسأله آزمون فرضیه‌ها در اکثر مباحث آماری به عنوان یک مسأله اساسی مورد استفاده واقع می‌شود. در بیشتر مسایل، یک تناظر یک به یک بین آزمون فرضیه و برآورد فاصله‌ای پارامتر متناظر در جامعه وجود دارد.

معمولاً در آزمون فرضیه‌های آماری به بررسی ادعای محقق در مورد مقدار یا مقادیر پارامتر جامعه می‌پردازیم. لذا در آزمون فرضیه‌ها دو فرضیه به نام‌های فرضیه صفر یا اولیه که معمولاً با H_0 نشان داده می‌شود و فرضیه مقابل یا مخالف که با نماد H_1 نشان داده می‌شود، وجود دارد. فرضیه آماری هرگونه ادعایی در خصوص جامعه است که هیچ نوع رابطه یا تفاوتی را بیان نمی‌کند. فرضیه مقابل هر گونه فرضیه‌ای است که نفی H_0 باشد. بنابراین در اکثر موارد ادعای محقق تحت عنوان فرضیه مقابل درنظر گرفته می‌شود.
برای بررسی این فرضیه‌ها، با توجه به‌این که پارامتر مربوطه در جامعه نامعلوم است، باید نمونه‌ای تصادفی از جامعه برگزیده و با توجه به آماره های معین که بر اساس مشاهدات نمونه‌ای مقدار آن به دست می‌آید در مورد رد کردن یا رد نکردن فرضیه آماری تصمیم‌گیری نمود. به طور کلی یک فرضیه گزاره‌ای در مورد پارامتر مجهول جامعه است.
به جهت درک بیشتر تناظر بین آزمون فرضیه‌ها و برآورد فاصله‌ای به مثال زیر توجه کنید:
فرض کنید می‌خواهیم برای پارامتر میانگین جامعه یک برآورد فاصله‌ای بدست آوریم. در این صورت حد وسط فاصله اطمینان، برآورد نقطه‌ای پارامتر میانگین جامعه است. حال اگر آزمون فرضیه به صورتی باشد که تحت فرضیه اولیه، پارمتر میانگین برابر مقداری معلوم باشد و فرضیه مقابل، نقیض آن باشد، با توجه به نمونه تصادفی از جامعه می‌توان به بررسی فرضیه‌ها پرداخت. اگر براساس نمونه، میانگین داده‌ها مقداری باشد که در حدود اطمینان برای پارمتر مجهول جامعه واقع گردد، فرضیه اولیه رد نمی‌شود. در غیر این صورت باید فرضیه اولیه را رد کنیم.

انواع خطاها

در آزمون فرضیه‌های آماری در دو حالت ممکن است که مرتکب خطا گردیم. اولین خطا که به خطای نوع اول مشهور است، وقتی رخ می دهد که فرضیه اولیه‌ای که واقعاً درست است را به اشتباه با توجه به نتایج نمونه رد کنیم. احتمال ارتکاب به خطای نوع اول را با نماد آلفا (α) نشان می‌دهیم. خطای بعدی که به عنوان خطای نوع دوم شناخته می‌شود، وقتی رخ می‌دهد که فرضیه مقابل که واقعاً درست است را به اشتباه رد کنیم. احتمال ارتکاب به‌این خطا را با نماد بتا (β) نشان می‌دهیم. بنابراین داریم:

α= احتمال ارتکاب خطای نوع اول= **P(** رد فرض $\mathbf{H_0}$ به اشتباه **)**= $P[RH_0|H_0\ is\ true\ ]$

β= احتمال ارتکاب خطای نوع دوم= **P(** رد فرض مقابل به اشتباه **)**= $P[H_1|H_1 is\ true]$

تذکر مهم

ارتکاب خطای نوع اول در مسایل آزمون فرضیه به مراتب بدتر از خطای نوع دوم است. به عنوان مثال فرض کنید، فردی متهم به یک نوع خلاًف است. اگر در دادگاه، قاضی با توجه به تحقیقات به عمل آمده و مستندات پرونده، رای به مجرمیت این فرد بدهد در حالی که واقعا فرد گناهکار نیست، این مسأله اشتباه خیلی بدتری است نسبت به حالتی که فرد واقعاً گنهکار است، اما قاضی با توجه به شواهد موجود او را تبرئه نماید.

آماره آزمون

حال که مسأله آزمون فرضیه تشریح شد، می‌خواهیم بدانیم که چگونه از روی نمونه تصادفی در مورد فرضیه‌ها تصمیم‌گیری نمائیم. یکی از مشخصه‌هایی که به ما برای تصمیم‌گیری کمک می‌کند، آماره (آزمون) است. آماره تابعی از مشاهدات است که به پارامتر مجهول بستگی ندارد. بنابراین با داشتن یک نمونه تصادفی می‌توان مقدار آماره را تعیین نمود و از روی آن برای رد کردن یا عدم رد فرضیه اولیه در مقابل فرضیه مقابل تصمیم‌گیری کرد.

ناحیه بحرانی

ناحیه بحرانی آزمون مجموعه‌ای است که به ازای تمامی‌مقادیر آن، فرضیه اولیه رد می‌شود. بنابراین ناحیه بحرانی مقادیری از آماره آزمون است که به ازای آن فرضیه H_0 رد می‌شود.

سطح معنی داری آزمون

احتمال رد فرضیه اولیه وقتی در واقع آن فرضیه درست است، را سطح معنی‌داری آزمون نیز می‌نامند. معمولاً در تحقیقات تجربی یا شبه تجربی که احتمال ارتکاب خطای نوع اول به صورت اختیاری در دست محقق می‌باشد، مقدار سطح معنی داری آزمون برابر ۵ یا یک درصد در نظر گرفته می‌شود.

آزمون‌های یک طرفه و دو طرفه

با توجه به ناحیه رد و نوع فرضیه مقابل، آزمون فرضیه‌های آماری به دو دسته آزمون یک طرفه و دو طرفه تقسیم بندی می‌شود. آزمون‌های یک طرفه آزمون‌هایی هستند که ناحیه بحرانی آن‌ها فقط در یک طرف (راست یا چپ) توزیع احتمال آماره آزمون واقع می‌شود. آزمون‌هایی که ناحیه رد آن‌ها در دو طرف راست و چپ قرار می‌گیرد، آزمون‌های دو طرفه نامیده می‌شوند. آزمون‌های یک طرفه نیز به دو دسته آزمون‌های یک طرفه راست و یک طرفه چپ دسته بندی می‌شوند. اگر ناحیه رد یک آزمون یک طرفه در سمت راست باشد، به آن آزمون یک طرفه راست یا بالا دنباله‌ای گفته می‌شود و در غیر این صورت آن را یک طرفه چپ یا پائین دنباله‌ای می‌نامند.
اگر محقق در ادعای خود تأکید بر نوعی برتری روش جدید نسبت به وضع موجود داشته باشد، همانند این که روش جدید آموزشی نسبت به روش فعلی کاراتر است، آن‌گاه آزمون فرضیه مربوطه یک طرفه خواهد بود. به عنوان مثال فرض کنید می‌خواهیم $H_0: \mu \ge 50$ را در مقابل $H_1: \mu < 50$ آزمون کنیم. چون تنها به ازای مقادیر کوچک میانگین جامعه، فرضیه اولیه رد می‌شود و ناحیه رد آزمون در طرف چپ (از لحاظ محور اعداد) واقع می‌شود، این آزمون، یک آزمون فرضیه یک طرفه چپ است. اما اگر $H_0: \mu \le 50$ باشد، باید فرضیه مقابل را به صورت $H_1: \mu > 50$ در نظر گیریم. در این صورت چون ناحیه بحرانی در سمت راست μ قرار می‌گیرد، آزمون فوق از نوع یک طرفه راست است.

مثال ۳: در یک بررسی آماری از یک توده سنگ ماسه این سوال مورد توجه است. آیا متوسط میزان تخلخل در این نمونه سنگها بیش از ۲۰ درصد است. در این مثال فرضیه های صفر و مقابل به صورت زیر بیان می شوند.

فرضیه صفر: متوسط میزان تخلخل حداکثر برابر ۲۰ درصد است

فرضیه مقابل: متوسط میزان تخلخل بیشتر از ۲۰ درصد است. به زبان ریاضی این فرضیه را به صورت زیر میتوان نمایش داد.

$$\begin{vmatrix} H_0: & \mu \le 20 \\ H_1: & \mu > 20 \end{vmatrix}$$

انواع خطا:

خطای نوع ۱ : رد فرض صفر به اشتباه که معادل است با پذیرش فرض مقابل به اشتباه. در این صورت

$$\alpha = P[error1] = P[RH_0 \mid H_0 \; is \; true]$$

را احتمال خطای نوع اول یا سطح آزمون نامند.

خطای نوع ۲ : رد فرض مقابل به اشتباه که معادل است با قبول فرض صفر به اشتباه. در این صورت

$$\beta = P[error\,2] = P[RH_1 \mid H_1 \, is \, true] = P[AH_0 \mid H_0 \; is \; fals]$$

را خطای احتمال خطای نوع دوم نامند.

بدیهی است در یک آزمون ایده آل احتمال خطا های نوع ۱ و ۲ حداقل مقدار دارند.

آزمون آماری: قاعده یا معیاری است که با خلاصه کردن مشاهدات حاصل از نمونه حکم به تایید یا رد فرض H۰ می دهد.

نکته: مقدار α را سطح آزمون نامیده و با معلوم بودن α و توزیع برآورد کننده نقطه ای پارامتر مجهول آزمون کاملا مشخص می شود.

مقدار احتمال:(p-value)

مقداری است در فاصله (۰،۱) و سطح معنی داری آزمون را نشان می دهد یعنی اینکه مقدار سطح ازمون تا چه حد قابل کاهش می باشد تا نتایج نمونه رای به رد فرضیه صفر دهد. و معمولا برای فرضیه های یک طرفه برای آزمون فرضیه صفر در مقابل فرضیه دیگر از قاعده زیر استفاده می شودو در تمام تحلیل های آماری مقدار احتمال توسط نرم افزار محاسبه می شود که تفسیر مقدار آن به عهده پژوهشگر می باشد.

$$RH_0 \Leftrightarrow p-value<\alpha$$

آزمون میانگین جامعه نرمال:

از اینکه $\bar{X}$ یک برآوردگر مناسب برای پارامتر میانگین مجهول جامعه (μ) مورد نظر، می باشد. برای آزمون فرضیه های مربوط به میانگین جامعه از برآورد کننده $\bar{X}$ استفاده می شود.

فرضیه های دوطرفه:

$$\begin{vmatrix} H_0: & \mu=\mu_0 \\ H_1: & \mu\neq\mu_0 \end{vmatrix}$$

۱- *واریانس جامعه (σ) معلوم:* جهت تعیین فاصله اطمینان برای μ و نواحی رد یا قبول فرض .H از توزیع نرمال استاندارد (Z) استفاده می شود. در صورتی که n > ۳۰ بدون توجه به توزیع جامعه، می توان از توزیع نرمال جهت محاسبه پارامترهای مجهول استفاده نمود.

آزمون: $$RH_0 \Leftrightarrow Z_0=\frac{\sqrt{n}\left|\bar{X}-\mu_0\right|}{\sigma_0}>Z_{\frac{\alpha}{2}}$$

که در آن $$Z_0=\frac{\sqrt{n}\left(\bar{X}-\mu_0\right)}{\sigma_0}$$

آزمون بر اساس مقدار احتمال: $$RH_. \Leftrightarrow p-value<\alpha$$

۲- *واریانس جامعه مجهول:* جهت تعیین فاصله اطمینان برای μ و نواحی رد یا قبول فرض .H از توزیع t با درجه آزادی k=n-۱ استفاده می شود. در صورتی که n > ۳۰ می توان از توزیع نرمال جهت محاسبه پارامترهای مجهول استفاده نمود.

آزمون: $$RH_0 \Leftrightarrow \frac{\sqrt{n}\left|\bar{X}-\mu_0\right|}{S}>t_{\frac{\alpha}{2}}(n-1)$$

که در آن: $$T=\frac{\sqrt{n}\left(\bar{X}-\mu_0\right)}{S}$$

آزمون بر اساس مقدار احتمال: $$RH_. \Leftrightarrow p-value<\alpha$$

فرضیه های یک طرفه:

$$1)\quad \begin{vmatrix} H_0:\mu=\mu_0(\mu\leq\mu_0) \\ H_1:\mu>\mu_0 \end{vmatrix}$$

۱- *واریانس جامعه (σ) معلوم:*

آزمون : $$RH_0 \Leftrightarrow Z_0=\frac{\sqrt{n}\left(\bar{X}-\mu_0\right)}{\sigma_0}>Z_{\alpha}$$

آزمون بر اساس مقدار احتمال: $$RH_0 \Leftrightarrow p-value<\alpha$$

۲- *واریانس جامعه مجهول:*

آزمون : $$RH_0 \Leftrightarrow \frac{\sqrt{n}\left(\bar{X}-\mu_0\right)}{S}>t_{\alpha}(n-1)$$

آزمون بر اساس مقدار احتمال: $RH_0 \Leftrightarrow p-value < \alpha$

$$2)\quad \begin{cases} H_0 : \mu = \mu_0 (\mu \geq \mu_0) \\ H_1 : \mu < \mu_0 \end{cases}$$

۱- واریانس جامعه (σ) معلوم:

آزمون: $RH_0 \Leftrightarrow Z_0 = \frac{\sqrt{n}\left(\overline{X}-\mu_0\right)}{\sigma_0} < -Z_\alpha$ *آزمون بر اساس مقدار احتمال:* $RH_0 \Leftrightarrow p-value < \alpha$

۲- واریانس جامعه مجهول:

آزمون: $RH_0 \Leftrightarrow \frac{\sqrt{n}\left(\overline{X}-\mu_0\right)}{S} < -t_{\alpha}(n-1)$ *آزمون بر اساس مقدار احتمال:* $RH_0 \Leftrightarrow p-value < \alpha$

Interval Estimation for mean:
Step ۱. Type your data in one column .

Step ۲. Choose Analyze------Descriptive stat--------Explore

Step ۳ Select names your data in Explore dialog box and transfer into Dependent list, then select Statistics for Display-----Ok.

Test involving a single population Mean:
Step ۱. Type your data in one column

Step ۲. Choose Analyze -------Compare means------One sample T- Test , then transfer name of your data into Test variables: select Test values -----Ok

مثال ۲. ده نمونه از یک توده ماسه سنگ از جامعه ای که تخلخلی بیش از ۱۸ درصد دارد استخراج کرده ایم. آیا داده ها در سطح ۵ درصد تایید میکنند که میزان تخلخل در این جامعه واقعا بیش از ۱۸ درصد است. ۱۳ ۱۷ ۱۵ ۲۳ ۲۷ ۲۹ ۱۸ ۲۷ ۲۰ ۲۴ .

مقایسه دو جامعه مستقل نرمال از طریق میانگین ها (μ۱ - μ۲)

در این بخش هدف کلی بررسیهای مقایسه ای؛ تعیین شباهتها یا کشف و اندازه گیری اختلاف پارامترها در دو جامعه آماری می باشد. در بررسیهای مقایسه ای؛ اغلب این موضوع مطرح می شود تا بین دو طرح مقابل به صورت نمونه های مستقل یا وابسته؛ قضاوت گردد؛ یعنی ماهیت بررسی ایجاب می کند که گاهی نمونه های مستقل و گاهی نمونه های وابسته مطالعه شوند. به مثال زیر توجه کنید:

مثال۳-از دو منطقه خلج و کشکان، تعداد ۳۰ نمونه ماسه سنگ از قسمت میانی لایه ها برای انجام آزمایش XRF گرفته شده و فراوانی اکسید منیزیم در ماسه سنگ های دو منطقه تعیین گردیده است. حال سوال این است که آیا متوسط میزان اکسید منیزیم در نمونه های مربوط به برش خلج و برش کشکان مساوی هستند یا خیر؟

برش کشکان	برش خلج
۱,۲۹	۳,۱۳
۸,۶۷	۵,۵۰
۲,۹۵	۲,۵۹
۲۶,۳۹	۱۲,۳۳
۱۵,۴۸	۲,۶۶
۲۳,۷۵	۲,۷۳

۲۲,۲۵	۱۳. ۳۵
۱۵,۷۹	۱۲,۶۳
۲۰,۸۱	۱۴,۰۲
۲۳,۰۹	۱۳
۲۰,۳۶	۱۴,۲۱
۲۸,۶۸	۱۵,۹۷
۲۸,۸۴	۸,۵۸
۲۷,۹۰	۱۲,۸۱
۲۷,۴۲	۹,۱۲
۳۴,۵۱	۸,۶۲
۲۷,۲۳	۱۵,۳۴
۲۵,۴۷	۱۶,۷۹
۳۱,۴۵	۱۵,۷۶
۳۳,۵۴	۱۷,۸۴
۳۱,۰۳	۱۷,۵۲
۲۶,۵۶	۱۵,۱۵
۲۹,۳۱	۱۸,۰۶
۲۶,۶۸	۱۴,۱۹
۳۲,۴۶	۱۷,۴۳
۴۹,۴۲	۳,۱۹
۴۳,۴۶	۷,۱۸
۲۶	۱۹,۱۴
۴۷	۲,۴۵
۳۱,۹۳	۱۶,۲۴

برای تحلیل این داده ابتدا بایستی میانگن دو نمونه را که مستقل از هم بدست آمده اند محاسبه کرده با مقایسه آنها با استفاده از روشهای مناسب آماری که بیان شد به سوال مطرح شده پاسخ منطقی داده می شود. ابتدا انواع فرضیه های آماری که در این گونه تحلیل با آن مواجه خواهیم شد را معرفی می نماییم سپس به تحلیل این مثال می پردازیم.

الف: فرضیه دوطرفه

1) $\begin{vmatrix} H_0: & \mu_1 = \mu_2 \\ H_1: & \mu_1 \neq \mu_2 \end{vmatrix}$

۱ - واریانس جامعه معلوم: **جهت تعیین فاصله اطمینان برای ($\mu_1 - \mu_2$) و نواحی رد یا قبول فرض .H از توزیع نرمال استاندارد (Z) استفاده می شود.**

آزمون یا قاعده تصمیم گیری: $RH_0 \Leftrightarrow |Z_0| = \dfrac{|\overline{X}-\overline{Y}|}{\sqrt{\dfrac{S_1^{\,2}}{n}+\dfrac{S_2 2}{m}}} > Z_{\frac{\alpha}{2}}$ **آزمون بر اساس مقدار احتمال:** $RH_0 \Leftrightarrow p-value < \alpha$

۲- *واریانس های دو جامعه مجهول:* جهت تعیین فاصله اطمینان برای (μ۱ - μ۲) و نواحی رد یا قبول فرض .H از توزیع t با درجه آزادی k استفاده می شود.

آزمون: $RH_0 \Leftrightarrow |t_0| = \dfrac{|\overline{X}-\overline{Y}|}{S_p\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} > t_{\frac{\alpha}{2}}(k)$ که در آن $S_p = \dfrac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}$, $k = n+m-2$ و ***آزمون بر اساس مقدار احتمال*** *به صورت* $RH_0 \Leftrightarrow p-value < \alpha$ **می باشد.**

ب: فرضیه یک طرفه

$$2)\quad \begin{vmatrix} H_0: & \mu_1 \le \mu_2 \\ H_1: & \mu_1 > \mu_2 \end{vmatrix}$$

۱- واریانس جامعه معلوم:

آزمون: $RH_0 \Leftrightarrow Z_0 = \dfrac{\overline{X}-\overline{Y}}{\sqrt{\dfrac{S_1^{\,2}}{n}+\dfrac{S_2 2}{m}}} > Z_\alpha$ و ***بر اساس مقدار احتمال:*** $RH_0 \Leftrightarrow p-value < \alpha$

۲- واریانس های دو جامعه مجهول:

آزمون: $RH_0 \Leftrightarrow t_0 = \dfrac{\overline{X}-\overline{Y}}{S_p\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} > t_\alpha(k)$ و ***بر اساس مقدار احتمال:*** $RH_0 \Leftrightarrow p-value < \alpha$

$$3)\quad \begin{vmatrix} H_0: & \mu_1 \le \mu_2 \\ H_1: & \mu_1 > \mu_2 \end{vmatrix}$$

۱- واریانس جامعه معلوم:

آزمون : $RH_0 \Leftrightarrow Z_0 = \dfrac{\overline{X}-\overline{Y}}{\sqrt{\dfrac{S_1^{\,2}}{n}+\dfrac{S_2 2}{m}}} < -Z_\alpha$ و ***بر اساس مقدار احتمال:*** $RH_0 \Leftrightarrow p-value < \alpha$

۲- واریانس های دو جامعه مجهول:

آزمون : $RH_0 \Leftrightarrow t_0 = \dfrac{\overline{X}-\overline{Y}}{S_p\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}} < -t_\alpha(k)$ و ***بر اساس مقدار احتمال:*** $RH_0 \Leftrightarrow p-value < \alpha$

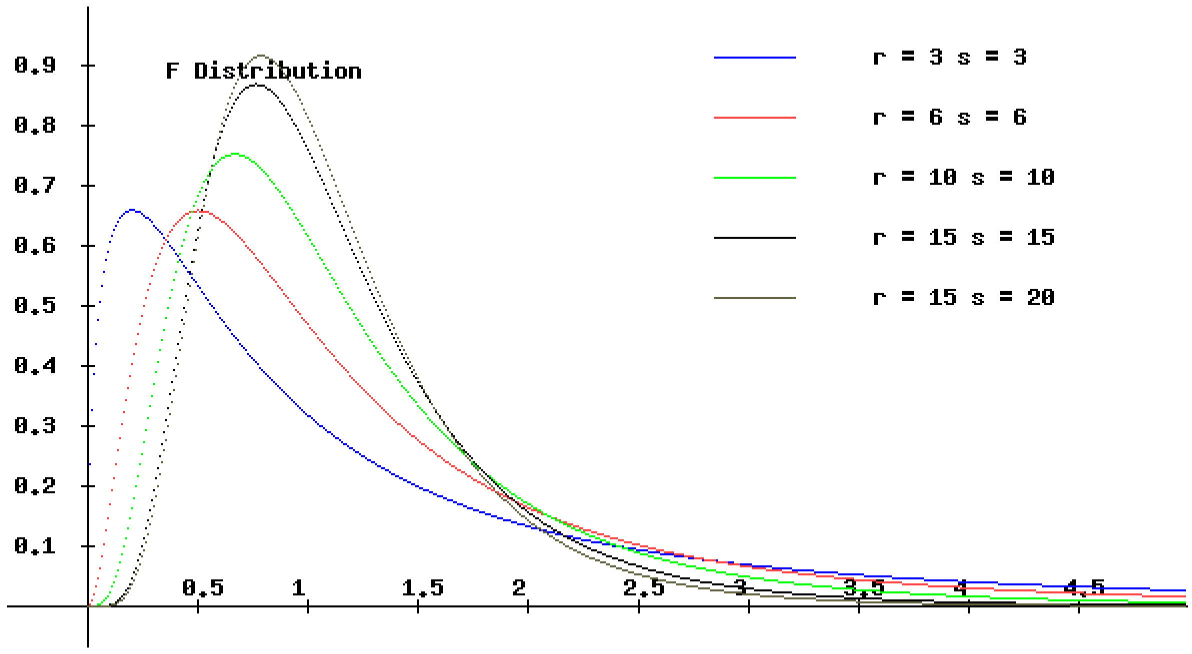
توزیع فیشر(F):

برای بررسی تساوی واریانس های دو نمونه مستقل از دو جامعه نرمال (σ۱/σ۲) از توزیع فیشر(F) با درجه آزادی k۱ , k۲ استفاده می شود.

$$\frac{S_1^2}{S_2^2}\times\frac{\sigma_2^2}{\sigma_1^2} \approx F(k_1,k_2) \qquad \begin{aligned} k_1 &= n-1 \\ k_2 &= m-1 \end{aligned}$$

به شرط اینکه واریانسها برابر باشند :

$$\sigma_1=\sigma_2=\sigma \Rightarrow \left\{\begin{aligned} \frac{S_1^{\,2}}{S_2^{\,2}} &\approx F(k_1,k_2) \\ \frac{S_2^{\,2}}{S_1^{\,2}} &\approx F(k_2,k_1) \end{aligned}\right\}$$



فرضیه های آماری: $\begin{cases} H_0: \sigma_1 = \sigma_2 \\ H_1: \sigma_1 \neq \sigma_2 \end{cases}$

آزمون: $AH_0 \Leftrightarrow p-value > \alpha$

Comparing two population means with independent samples.

Step۱ . Type your data in two columns. Variable ۱ contain data and variable ۲ contain cods "۱ for group۱" and "۲ for group۲".

Step۲. Choose **Analyze------Compare means-----Independent sample T Test.**

Step۳. In **independent samples T Test** select **file of data** and **transfer into Test variables**. Then select **cod** and **transfer into Grouping Variable**, then click **Define groups** then **enter ۱ for group۱ and ۲ for group۲------continue------OK**

آزمون t با نمونه‌های جفت

آزمون t با نمونه‌های جفت (Paired-Sample T Test) برای تجزیه و تحلیل آزمون‌هایی به کار می‌رود که هر دو نمونه در دو وضعیت متفاوت مورد مشاهده قرار می‌گیرد. در این آزمایش، اغلب اندازه متغیر در دو وضعیت(قبل و بعد) مورد بررسی قرار می‌گیرد.

فرض صفر در طرح داده‌های جفت این است که اختلافی بین میانگین‌های دو نمونه وجود ندارد و فرض مقابل اینست که بین میانگین‌های دو نمونه اختلاف وجود دارد. $\begin{cases} H_0: \mu_D = \mu_2 - \mu_1 = 0 \\ H_1: \mu_D = \mu_2 - \mu_1 \neq 0 \end{cases}$

آماره آزمون برای بررسی فرض فوق به صورت زیر است: $t = \dfrac{\sqrt{n}\overline{D}}{s_D}$ و $\overline{D} = \dfrac{1}{n}\sum_{i=1}^{n} D_i$

که در آن $D_i = X_{2i} - X_{1i}, i = 1,2,\dots,n$

آزمون بر اساس مقدار احتمال: $RH_0 \Leftrightarrow p-value < \alpha$

بر اسا س اماره تی فرض صفر تایید نمی شود اگر و تنها اگر $|t| > t_{\frac{\alpha}{2}}(n-1)$

مثال۴- برای جلوگیری از نشست ساختمان استاندارد پرواز فرودگاه مهرآباد تهران، عملیات حفاری و تزریق سیمان در زیر پی ساختمان صورت گرفته است. مقاومت فشاری تک محوری قبل از انجام عملیات تزریق سیمان و بعد از انجام عملیات تزریق سیمان بر روی نمونه های اخذ شده از محل زیر پی ساختمان انجام گرفته است. آیا نتایج بدست آمده تاثیر عملیات تزریق را در بهبود شرایط ژئوتکنیکی و مقاومتی پی ساختمان تایید می کند؟

اهمیت بررسی موضوع: از آن جا که انجام عملیات حفاری و تزریق دوغاب چه با پایه سیمانی و چه با پایه شیمیایی به صورت فازبندی و مرحله ای صورت می گیرد این آزمون به ما کمک می کند که انجام عملیات تزریق را تا مرحله ای (فاز) ادامه دهیم که به نتایج دلخواه برسیم که این عمل را با این آزمون به خوبی ارزیابی کرد.

مقاومت بعد از تزریق	مقاومت قبل از تزریق
۳۲	۵۳
۸۹	۷۰
۳۲	۴۶
۱۱۵	۸۳
۳۸	۷۱
۹۳	۵۵
۱۰۳	۴۸
۸۱	۵۸
۷۷	۲۵
۲۸	۶۸
۱۰۰	۵۳
۵۷	۹۲
۸۸	۱۰۰
۱۱۷	۵۹
۶۷	۷۸
۶۴	۸۳
۹۴	۴۲
۱۰۸	۴۱
۶۳	۳۷
۸۶	۲۷
۵۹	۲۱
۶۲	۲۳
۳۸	۵۷
۲۸	۳۰
۱۰۷	۶۰
۳۸	۹۴
۶۶	۳۰

۵۸	۶۲
۸۴	۳۶
۸۶	۴۵
۹۵	۷۰
۳۰	۷۶

Step ۱. Type your data in two variable "before for Variable ۱ " and "after for variable۲".

Step۲. Choose Analyze------Compare means-----Paired sample T Test.

Step۳. In Paired sample T-Test select the variables name "before" and "after" so that under current selections box we have variable۱:before variable۲.after.

Then transfer them into paired variables. As before—after and then OK

مثال۵ . داده های در صد کوارتز در ۲۰ نمونه مقطع نازک از یک سنگ آزرین به صورت زیر بدست آمده است.

۲۳,۵ ۱۶,۵ ۲۵,۴ ۱۹,۱ ۱۹,۳ ۲۲,۴ ۲۰,۹ ۲۴,۹ ۲۳,۳ ۱۹,۹ ۲۲,۵ ۲۳,۶ ۲۱,۸ ۱۹,۲ ۱۸,۹ ۲۱,۱ ۲۰,۹ ۱۹,۷ ۲۲,۴ ۲۳,۱

یک بازه اطمینان ۹۵ درصد برای متوسط واقعی و یک بازه اطمینان ۹۵ درصد برای انحراف معیار درصد کوارتز بسازید.

مثال۶: داده های زیر طول فسیلها را نشان می دهد که از دو افق A و B به تصادف برداشته شده اند.

A. ۳,۲ ۳,۱ ۳,۱ ۳,۳ ۲,۹ ۲,۹ ۳,۵ ۳ B. ۳,۱ ۳,۱ ۲,۸ ۳,۱ ۳ ۲,۶ ۳ ۳,۱ ۲,۸ ۲,۹۶

الف- آیا تفاوت معنی داری بین متوسط واقعی طول فسیل دو افق وجود دارد.

ب- آیا شاهدی وجود دارد که نشان دهد فسیلهای افق A طویلتر از افق B است.

مثال ۷-مقایسه دوجامعه مستقل نرمال ازطریق میانگینها ($\mu 2$ - $\mu 1$)

متوسط بارندگی ماهیانه ۲۰ ایستگاه باران سنجی وزارت نیرو در حوالی شهر گالیکش در دو ماه مهر و فروردین بر حسب میلیمتر در جدول زیر آورده شده است. آیا تفاوت معناداری میان میانگین بارندگی در ماه مهر و فروردین این ۲۰ ایستگاه وجود دارد. یکبازهاطمینان ۹۵ درصدی برای μ_1-μ_2 بسازید.

شماره ایستگاه	مهر	فروردین
۱	۵۰	۸۶
۲	۴۸,۱	۹۳,۱
۳	۱۶,۱	۲۷
۴	۵۱,۷	۱۰۰,۵
۵	۴۴,۵	۹۲,۳
۶	۴۴	۸۶,۱
۷	۵۵	۷۸
۸	۲۴,۲	۸۱,۴
۹	۵۹	۵۸
۱۰	۵۰	۶۹,۷
۱۱	۲۰,۳	۹۵,۱
۱۲	۳۳,۵	۹۰
۱۳	۴۷	۴۵,۳

۱۴	۳۹	۷۴
۱۵	۵۱,۷	۸۷,۶
۱۶	۲۹,۹	۵۹,۷
۱۷	۳۶,۴	۶۶
۱۸	۵۱,۸	۹۳,۶
۱۹	۲۲	۸۹,۲
۲۰	۴۰	۹۵

مثال ۸- آیا تفاوت معنی داری بین مقادیر جذب سیمان بر حسب **Kg/m** در مقاطع مختلف دو گمانه واقع در دو تکیه گاه سد شیروان وجود دارد (تکیه گاه راست کد ۱ و تکیه گاه چپ کد ۲)؟

اهمیت بررسی موضوع: با توجه به میانگین جذب سیمان و مقایسه آن با مقادیر نفوذپذیری و مشخصات دسته درزه ها، در مورد این که پارامترها و طراحی عملیات تزریق در جاهای مختلف ساختگاه سد باید یکسان یا متفاوت صورت پذیرد می توان قضاوت کرد.

مقادیر جذب سیمان (Kg/m) در تکیه گاه راست	۲۹	۱۲	۱۰	۳۱	۳۳	۱۰	۱۳	۱۴	۲۱	۱۱	۱۰	۱۱	۲۶	۱۵	۱۲۲	۲۰
۷	۱۲	۱۰	۱۴	۱۱	۲۰	۱۷	۱۲	۱۱	۱۳	۱۲	۶	۲۱	۱۳	۱۱	۲۴	۱۳

مقادیر جذب سیمان (Kg/m) در تکیه گاه چپ	۱۵	۱۴	۷	۴	۴	۱۱	۱۱	۱۲	۱۷	۶	۱۷	۸	۱۴	۱۷	۱۰	۱۰
۳	۷	۶	۱۱	۷	۸	۷	۱۴	۲۱	۱۲	۱۰	۱۷	۸	۴	۸	۴	۵