آمار پایه مورد استفاده در طراحی آزمایش

✐ مثال ۱–۱:

برای خرید گلوله برای آسیا یک مهندس قادر نخواهد بود که تمامی گلوله ها را مورد بررسی قرار دهد و مجبور است تعدادی از آنها را به عنوان نمونه مورد بررسی قرار دهد.

✐ مثال ۱–۲:

از یک مهندس خواسته می شود که عیار یک ماده معدنی چند میلیون تنی را بررسی کند. مشخص است که آنالیز شیمیایی همه ماده معدنی نمی تواند عملی باشد در نتیجه مهندس بایستی از طریق نمونه برداری و آنالیز، عیار ماده معدنی را مشخص کند.

در دو مثال بالا، مهندس بایستی استنتاج هایی در مورد یک جمعیت (Population) با آنالیز یک نمونه (Sample) ارائه کند.

بنابراین، آمار وسیله ای است که با آن استنتاج یا پیش گویی در مورد یک جمعیت با استفاده از داده های یک نمونه صورت می گیرد.

حدود اطمینان (Confidence Limit)

در مثال گلوله های آسیا اگر تعداد نسبتاً زیادی از آنها مورد آزمایش قرار گیرند و مشخص شود که کمترین مقاومت فشاری kg/cm^2 ۳۰۰۰ است در نتیجه می توان با اطمینان زیادی عنوان کرد که گلوله های آزمایش نشده مقاومتشان بیش از kg/cm^2 ۱۵۰۰ است. ولی بیان این که مقاومت فشاری گلوله های آزمایش نشده بیش از kg/cm^2 ۲۸۰۰ است، با اطمینان زیاد نیست.

در مورد مثال تعیین عیار ماده معدنی اگر عیار تخمینی ۱٪ باشد با اطمینان نسبتاً بالایی می توان عنوان کرد که عیار واقعی بیشتر از ۰/۵٪ می باشد ولی درجه اطمینان بیان این که عیار واقعی بیش از ۰/۷٪ می باشد به مراتب کمتر است.

مثال ۱-۳:

دو کیسه پول وجود دارد. یکی از کیسه ها فقط دارای سکه های خط دار (نوشته دار) می باشد، اگر از یکی از کیسه ها پنج سکه روی زمین بریزد و همگی آنها خط دار باشند، با چه اطمینانی می توان گفت که اینها از کیسه ای که دارای سکه های خط دار می باشد، بیرون آمده است؟

با ۹۷٪ اطمینان می توان گفت که این سکه ها از کیسه دارای سکه های خط دار بیرون آمده چون احتمال این که اینها از کیسه پول های واقعی بیرون آمده باشند ۳٪ است:

گرایش مرکزی توسط یکی از سه پارامتر زیر اندازه گیری می شود:

۱- میانگین (**Mean**): مجموع تمام مقادیر تقسیم بر تعداد آنها.

جمعیت: $\mu = \frac{\sum_{i=1}^{N} y_i}{N}$ نمونه: $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$

میانه (**Median**): مقدار مرکزی ما بین تمامی مقادیر .

نمرات درس "کنترل و مدلسازی" به ترتیب زیر بوده است، میانه نمره کلاس چه مقدار است؟

٧	١١	١٢	١٣/٥	١٤/٧	١٥/٢	١٥/٧

میانه: ١٣/٥

نما (**Mode**): مقداری که بیش از همه تکرار شده است.

برای تعیین عیار یک سنگ معدن طلا نمونه هایی برای آنالیز شیمیایی فرستاده شد و نتایج زیر بر حسب گرم بر تن بدست آمد، نما یا مد عیار سنگ معدن چه مقدار است؟

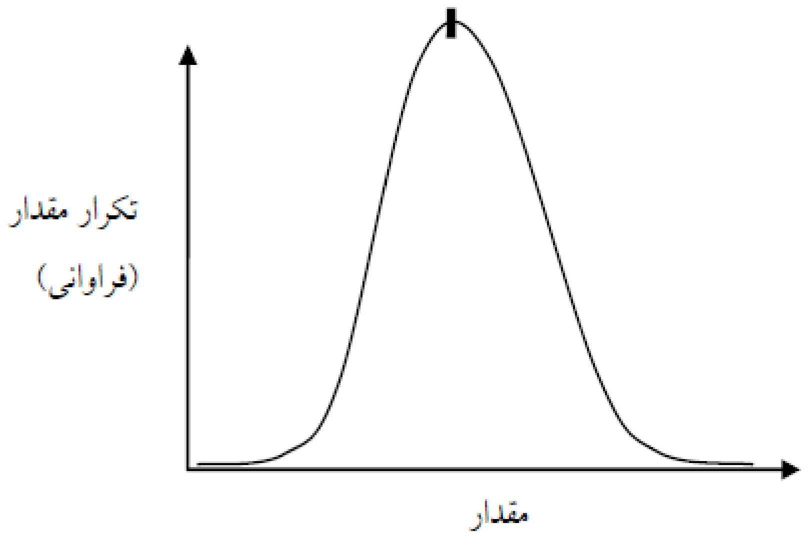
٦	٧/٥	٨/٢	٦	٧	٩/٦	٦	٨/٨

نما <u>٦</u> گرم بر تن می باشد چون ٣ بار تکرار شده است.

انواع توزیع مقادیر

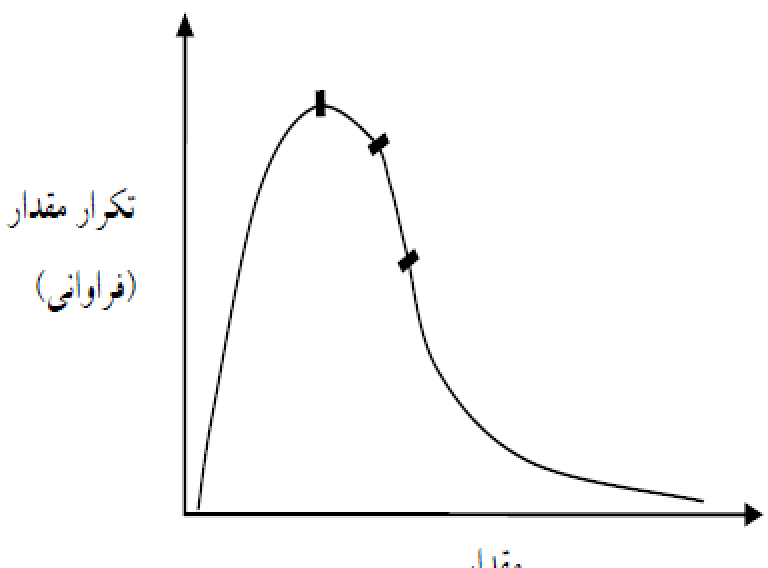
محور Y تعداد تکرار مقدار و محور X مقدار

۱- توزیع متقارن (Symmetrical Distribution)



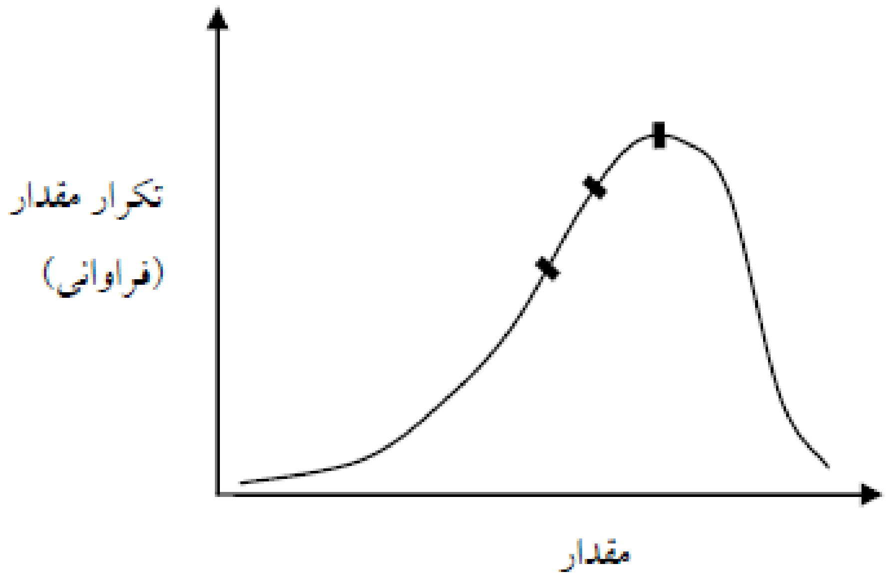
میانگین = میانه = نما

۲- توزیع به راست منحرف (Right – Skewed)



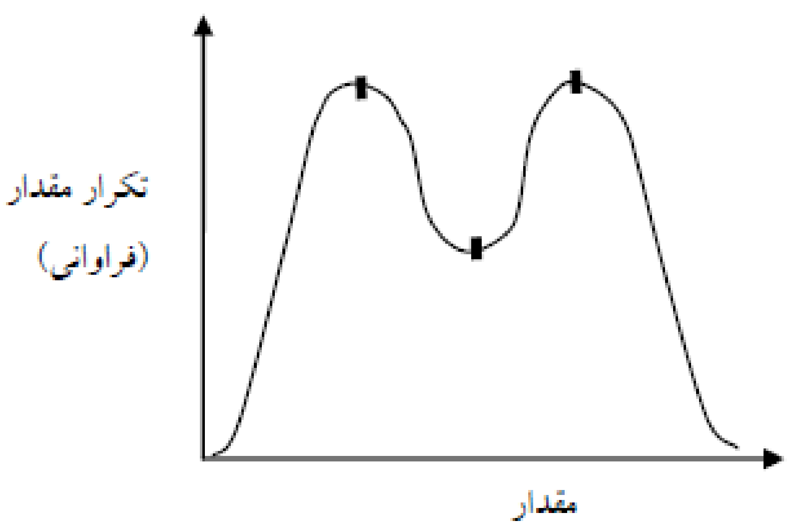
میانگین > میانه > نما

۳- توزیع به چپ منحرف (Left – Skewed)



انگین < میانه < نما

٤- توزیع دو نمایی (Bimodal)



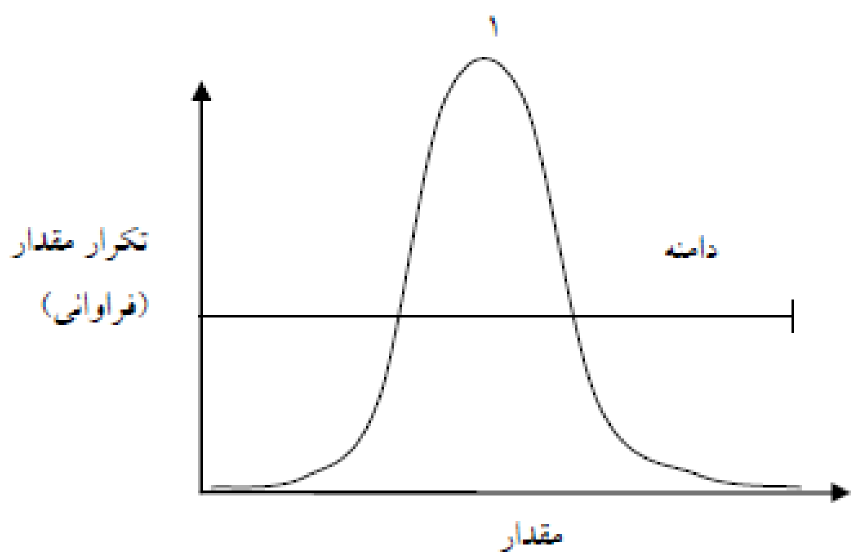
نما (۲) > میانه = میانگین > نما (۱)

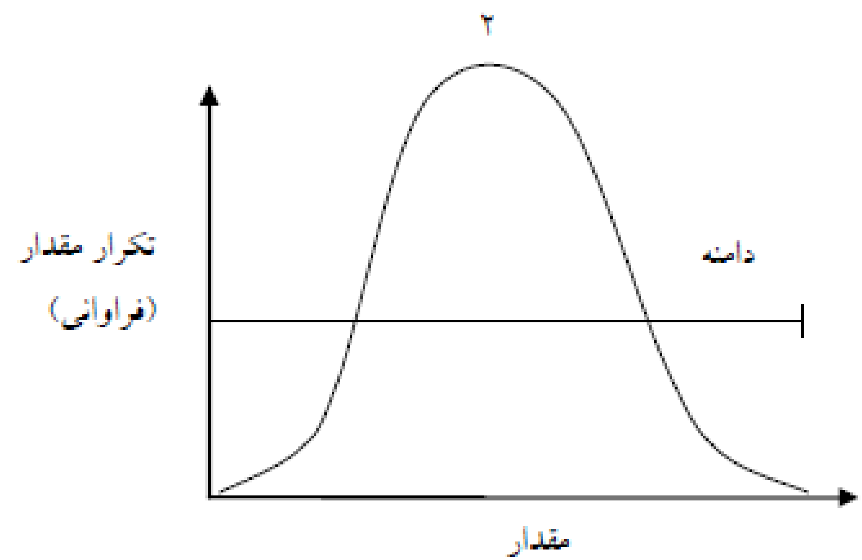
اندازه گیری تمایل مرکزی (Measuring Central Tendency)

میانگین، میانه و نما برای توصیف اطلاعات بکار می روند، ولی برای پیش بینی خیلی مفید نیستند.

مثال ۱-۹:

شکل ۱ و ۲ دارای میانگین، میانه و نمای یکسانی می باشند ولی شکل ۱ فشرده تر از شکل ۲ می باشد. در بیشتر موارد لازم است که پراکندگی داده ها نسبت به میانگین مشخص شود.





میانگین مربعات فاصله، واریانس (Variance) نام دارد و مهمترین پارامتر توصیف پراکندگی می باشد.

معادله ریاضی واریانس:

واریانس جمعیت $\sigma^2 = \frac{\sum_{i=1}^{N}(y_i - \mu)^2}{N}$

واریانس نمونه $s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{(n-1)}$

در محاسبه واریانس جمعیت، مجموع مربعات تقسیم بر N می شود چون میانگین به طور مستقل از داده ها (y_i) در دسترس می باشد ولی در مورد واریانس نمونه، میانگین ($\bar{y}$) از داده ها (y_i) محاسبه می شود و در نتیجه یک درجه آزادی کمتر دارد و تقسیم بر $n-1$ می شود.

معمولاً واریانس بدست آمده از نمونه ها کوچکتر از مقدار واقعی آن (σ^2) می باشد، در نتیجه بر $n-1$ تا حدی این مساله را جبران می کند.

مفهوم درجه آزادی (Degree of Freedom)

مثال ۱-۱۱:

اگر از شما خواسته شود که ۵ عدد فرد انتخاب کنید شما ۵ انتخاب آزاد دارید (درجه آزادی = ۵). ولی اگر از شما خواسته شود ۵ انتخاب کنید که میانگین آنها ۱۰ باشد، ٤ عدد را می توانید به دلخواه انتخاب کنید ولی عدد پنجم را نمی توانید به دلخواه انتخاب کنید به عبارت دیگر، درجه آزادی شما ٤ می باشد.

واحد واریانس توان دوم واحد اصلی است و برای بیان اندازه تمایل مرکزی نامناسب است.

رابطه ریاضی انحراف معیار

$$(\text{نمونه})\ s = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{(n-1)}}$$

$$(\text{جمعیت})\ \sigma = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \mu)^2}{N}}$$

انحراف معیار نسبی ($\frac{s}{\bar{y}}$ یا $\frac{\sigma}{\mu}$) و واریانس نسبی ($\frac{s^2}{\bar{y}^2}$ یا $\frac{\sigma^2}{\mu^2}$) که بی بعد هستند نیز بکار می رود.

اهمیت عملی انحراف معیار

ریاضی دان روسی (Chebysheff) در مورد کاربرد انحراف معیار قضیه زیر را ارائه کرد:

$$A = 1 - \frac{1}{K^2}$$

A بخشی از اندازه گیری ها است که در محدوده **K** برابر انحراف معیار از میانگین قرار خواهند گرفت.

بر اساس تئوری (Chebysheff)، چه بخشی از اندازه گیری ها در محدوده یک انحراف معیار از میانگین خواهند بود؟ چه بخشی در محدوده ۲ و ۳ برابر انحراف معیار خواهند بود؟

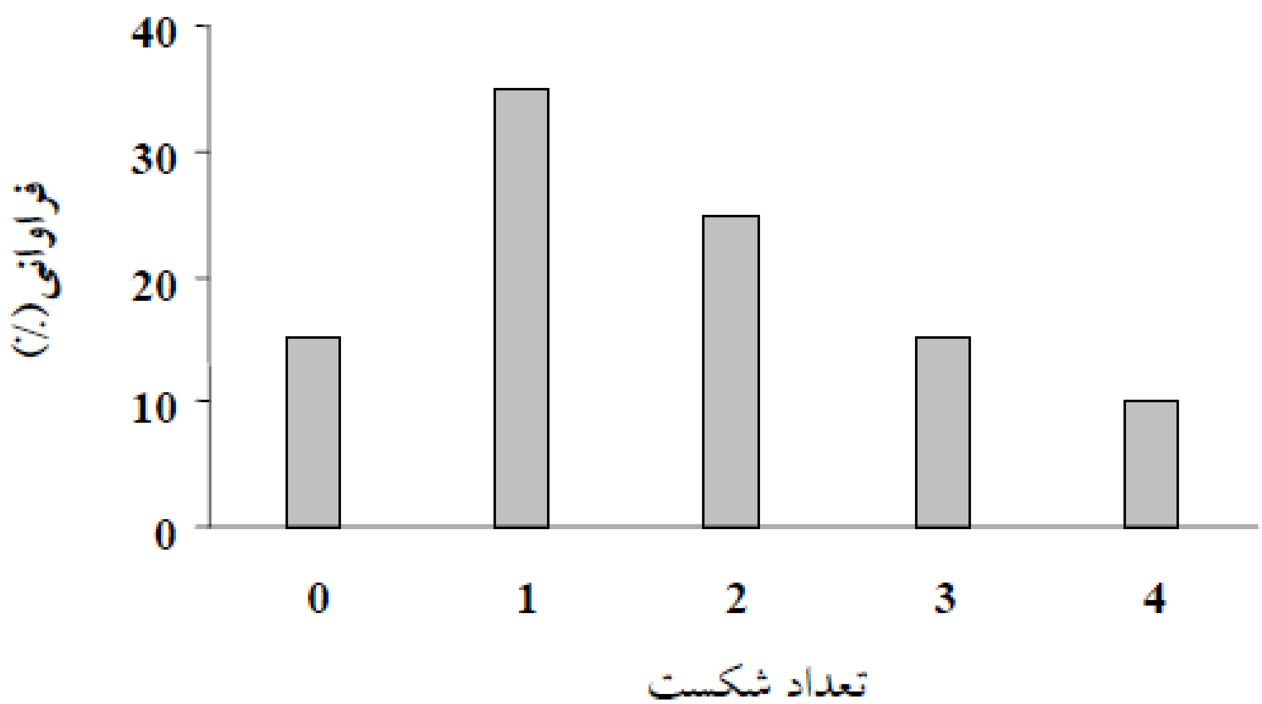
از نظر تئوری ممکن است هیچ اندازه ای در محدوده یک انحراف معیار قرار نگیرد. $\Rightarrow$ $\left(1 - \frac{1}{1}\right) = 0$

٪۷۵ اندازه گیری ها در ۲ انحراف معیار $\left(1 - \frac{1}{2^2}\right) = \frac{3}{4}$

روشهای تصویری (Pictorial Methods)

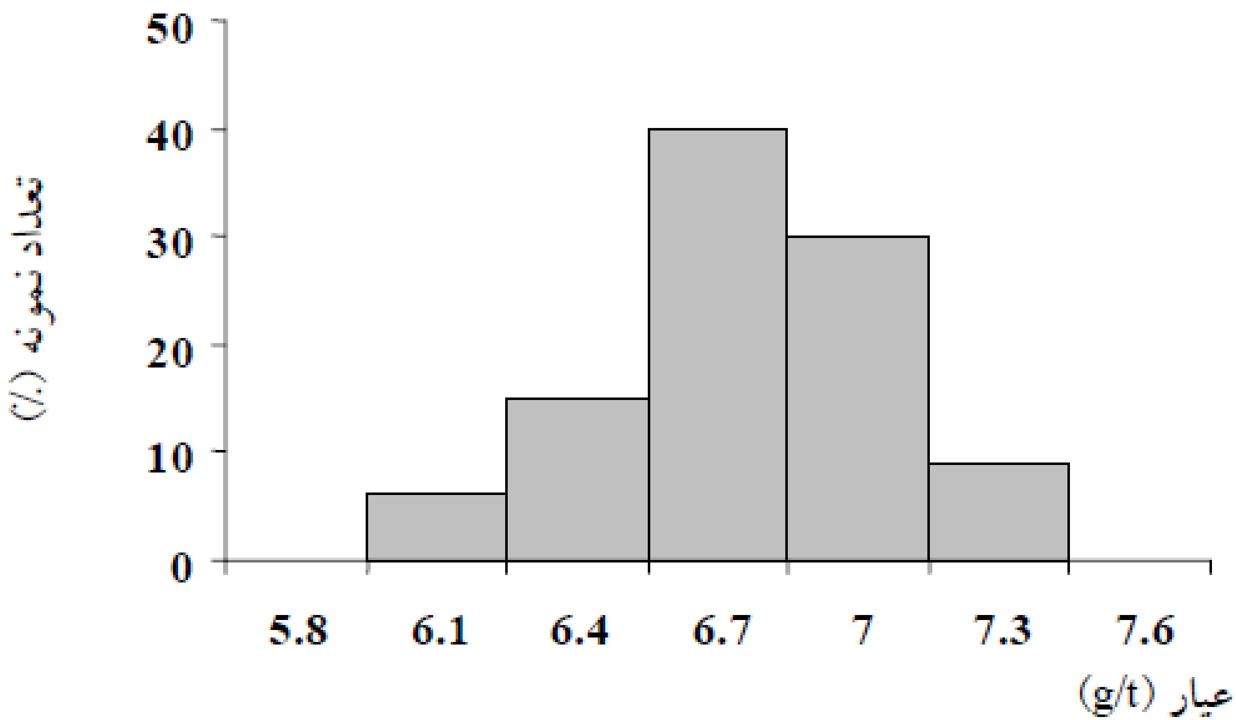
۱- نمودار میله ای (Bar chart)

این نوع نمودار برای داده های گسسته بکار می رود.



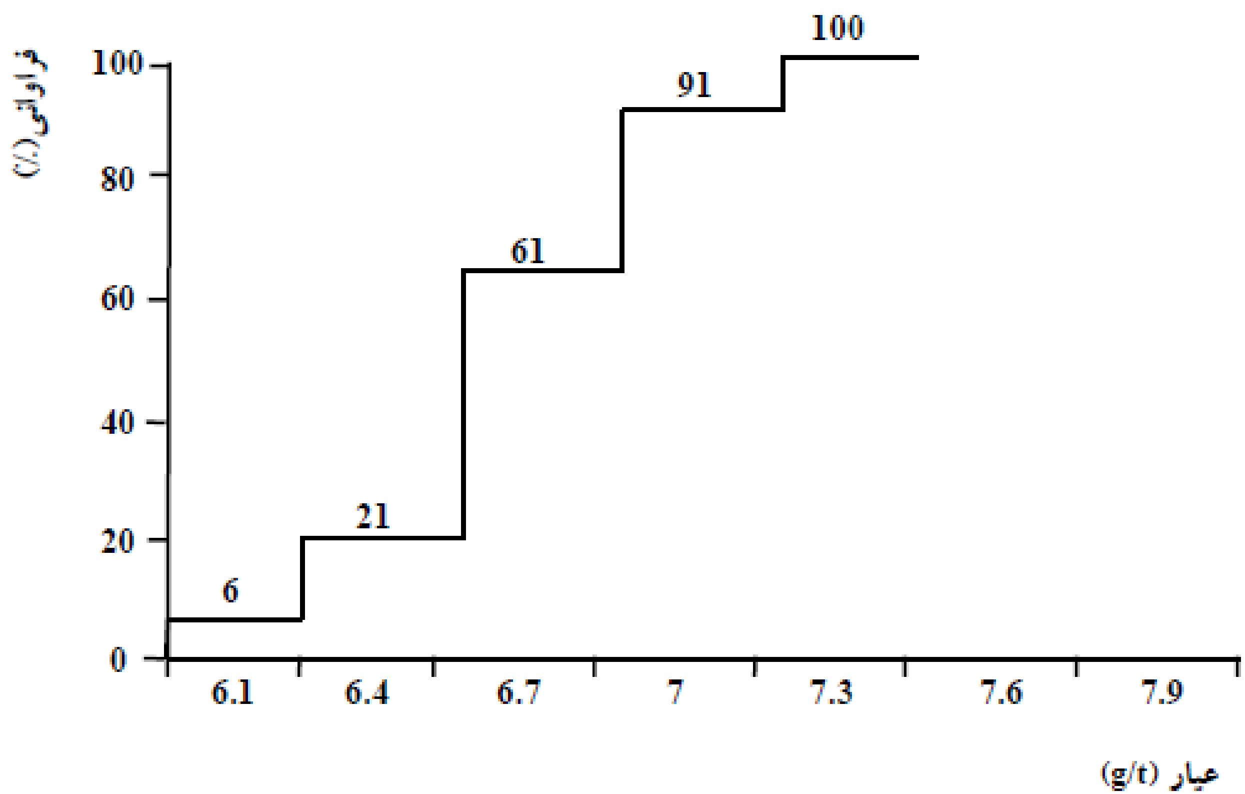
۲– نمودار ستونی (Histogram)

این نوع نمودار برای داده های پیوسته بکار می رود.



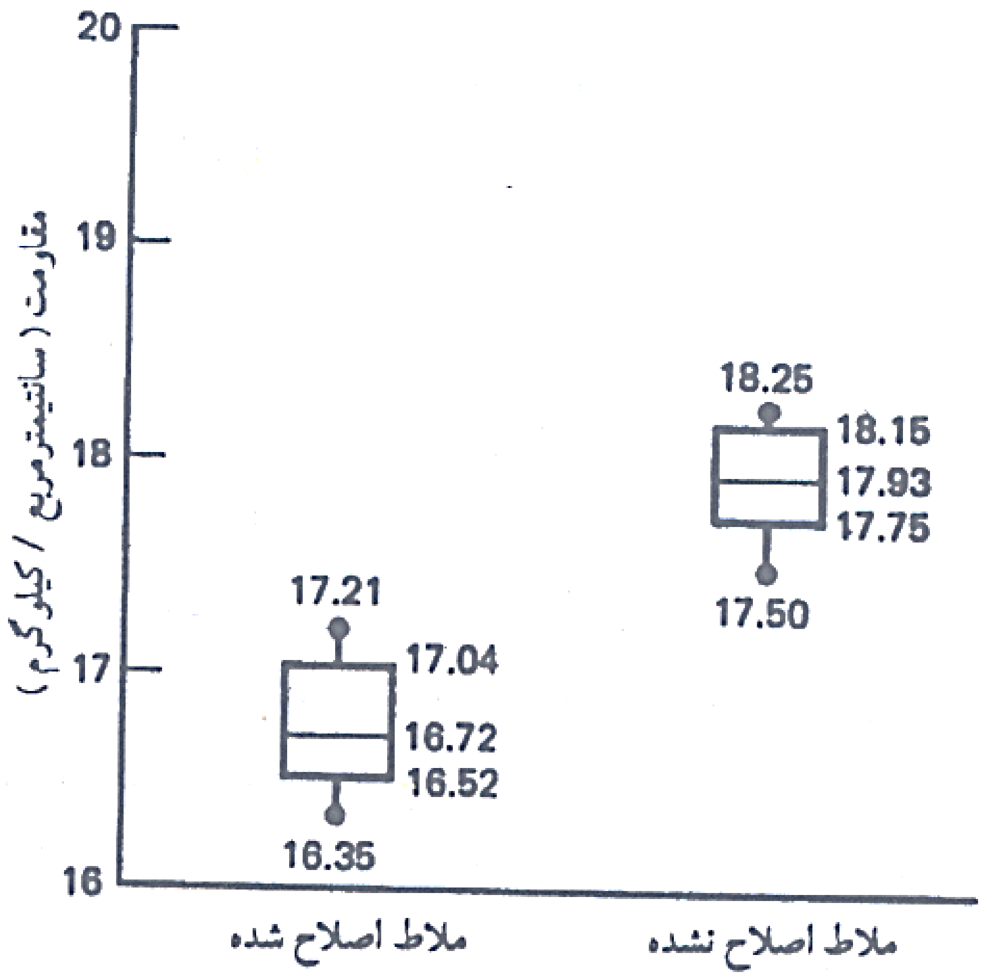
۳- نمودار تجمعی فراوانی (The Cumulative Frequency)

در این نمودار فراوانی تجمعی نسبی اندازه های مختلف رسم می شود.



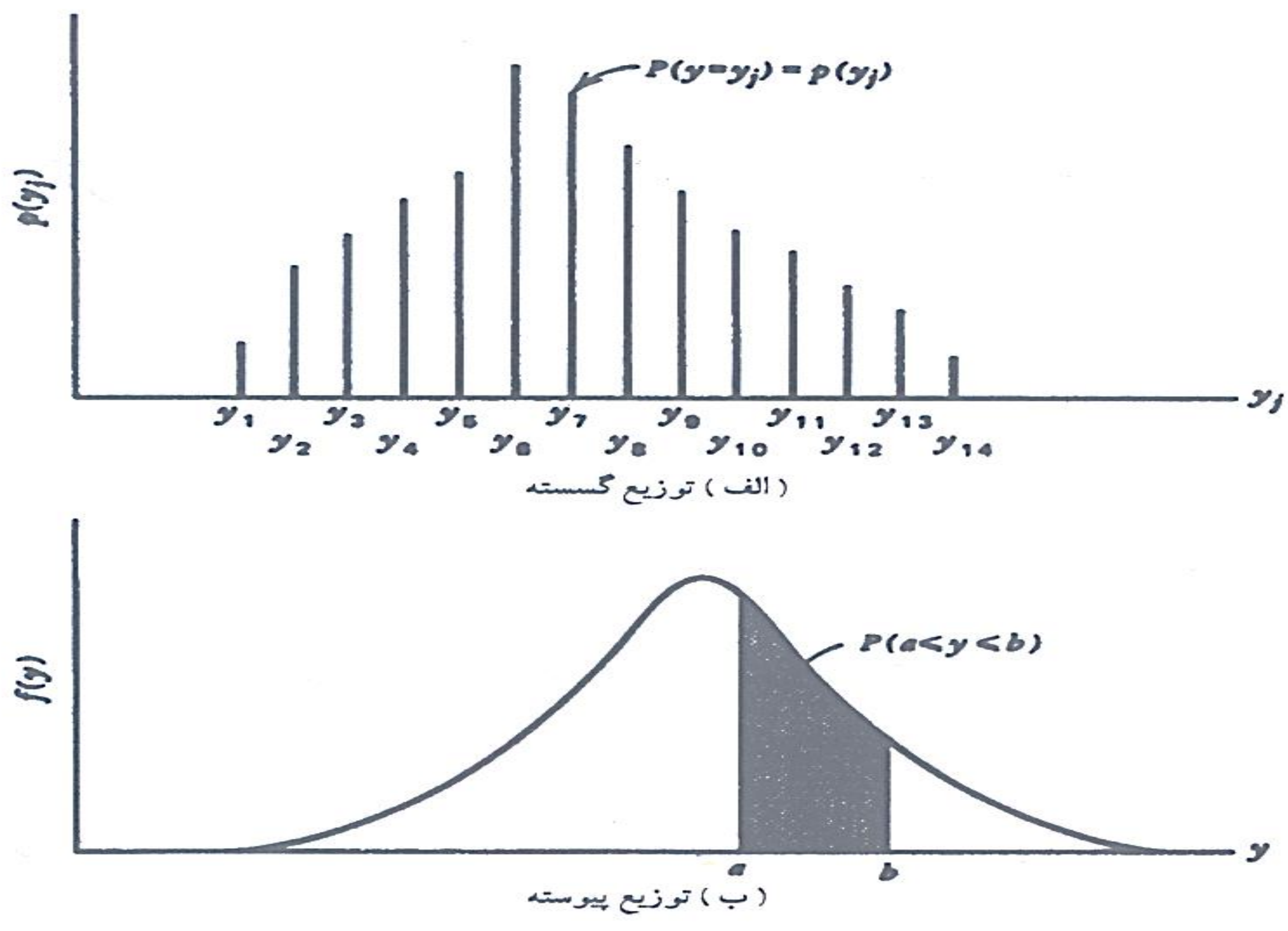
نمایش داده ها:

(جعبه ای)



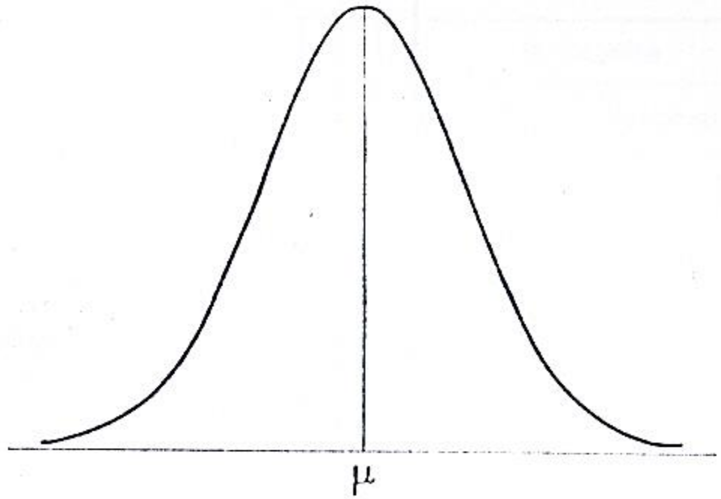
نمایش داده ها:

(پیوسته و گسسته)

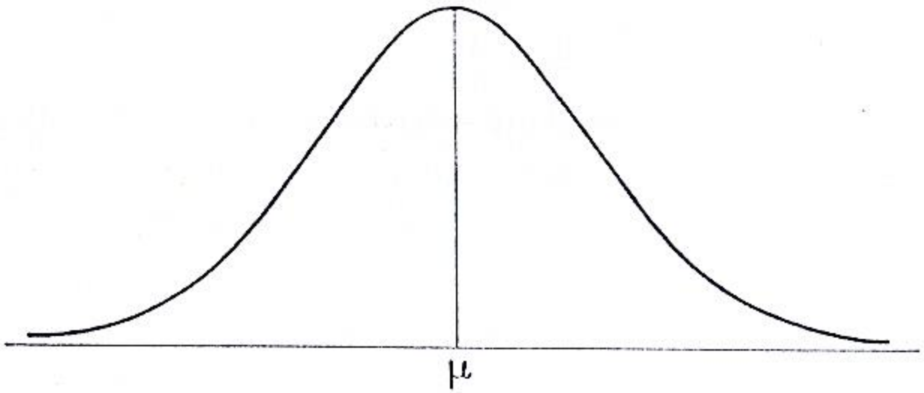


نمایش داده ها:

انحراف معیار



Curve A



Curve B

تابع چگالی احتمال نرمال

متغیر تصادفی نرمال یکی از توزیع های مهم آماری در حالت پیوسته است. متغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس σ^2است اگر تابع چگالی احتمال آن به صورت زیر باشد.

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

$$-\infty < x < \infty$$

$$-\infty < \mu < \infty$$

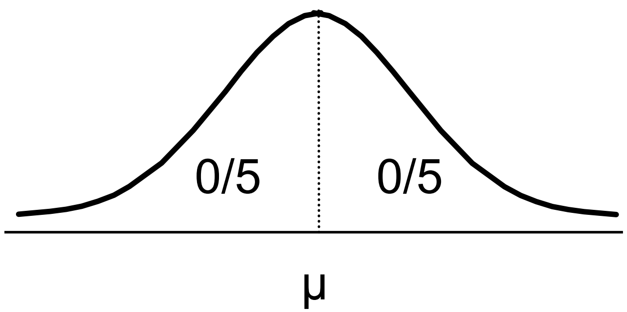
$$\sigma > 0$$

μ و σ2 پارامترهای توزیع نرمال هستند.

ویژگیهای توزیع نرمال

۱- این توزیع نسبت به محور $y=\mu$ دارای تقارن است.

۲- $\int_{-\infty}^{\infty} f(x)\,dx = 1$



۳- $P[X \le \mu] = P[X \ge \mu] = 0/5$

۴- برای $\mu=0$ و $\sigma^2=1$، توزیع نرمال را توزیع نرمال استاندارد گویند.

توزیع نرمال استاندارد

متغیر تصادفی Z دارای توزیع نرمال استاندارد است اگر تابع چگالی احتمال آن به صورت زیر باشد.

$$f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2} \qquad -\infty < Z < \infty$$

همانطور که ملاحظه می کنید این توزیع فاقد پارامتر است و برای راحتی متغیر نرمال استاندارد را با Z نمایش می دهند. در حقیقت Z همان متغیر X است با میانگین صفر و واریانس یک.

$$f(x) = f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2} \qquad -\infty < z < \infty$$

مقادیر مختلف F(X) را می توان با توجه به ویژگی Z از جدول ضمیمه منابع آماری بدست آورد.

$$F(z) = P(Z < z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}t^2} dt$$

متغیر تصادفی نرمال استاندارد نسبت به محور دارای تقارن است. یعنی:

$$\int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2} dz = \int_{0}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}z^2} dz = \frac{1}{2}$$

تابع چگالی احتمال کی‌دو

متغیر تصادفی X دارای تابع چگالی احتمال کی‌دو با پارامتر r است اگر تابع چگالی احتمال آن به صورت زیر باشد.

$$f(x) = \frac{1}{2^{\frac{r}{2}}\Gamma(\frac{r}{2})} x^{\frac{r}{2}-1} e^{-x/2} \qquad x > 0 \ , \quad r = 1,2,3,\ldots$$

توزیع کی دو حالت خاص توزیع گاما است $\left(\alpha = \frac{r}{2} , \quad \beta = 2\right)$

ویژگیهای توزیع کی دو

۱- rرا درجه آزادی توزیع گویند.

۲- دارای میانگین r و واریانس 2r است.

۳- $F(x) = P(X \le x) = 1 - \sum_{i=0}^{r-1} \frac{1}{i!}\left(\frac{x}{2}\right)^{i} e^{-x/2}$

۴- مقادیر مختلف F(x) را می توان برای مقادیر مختلف r از جدول ضمیمه منابع آماری بدست آورد.

تابع چگالی احتمال استودنت (توزیع t)

متغیر تصادفی X دارای توزیع t با پارامتر r است اگر تابع چگالی احتمال آن به صورت زیر باشد.

$$f(x)=\frac{\Gamma\left[\frac{(r+1)}{2}\right]}{\Gamma\left(\frac{r}{2}\right)}\cdot\frac{1}{\sqrt{r\pi}}\cdot\frac{1}{\left(1+\frac{x^2}{r}\right)^{\left(\frac{r+1}{2}\right)}}$$

که r را درجه آزادی توزیع t گویند.

$$-\infty<x<\infty \quad , \qquad r>0$$

ویژگیهای توزیع t:

۱- $\int_{-\infty}^{\infty} f(x)\,dx=1$

۲- برای r >۱ دارای میانگین صفر و برای r >۲ دارای واریانس $\frac{r}{r-2}$ است.

۳- در توزیع استودنت اگر درجه آزادی r از حد تصور بزرگتر باشد توزیع، بر توزیع نرمال استاندارد منطبق می شود.

۴- مقادیر مختلف F(x) برای مقادیر مختلف درجه آزادی r از جدول ضمیمه منابع آماری قابل محاسبه است.

تابع چگالی احتمال فیشر

متغیر تصادفی X دارای توزیع فیشر با پارامترهای r_1و r_2 است، اگر تابع چگالی آن به صورت زیر باشد

$$f(x)=\frac{\Gamma\left[\frac{(r_1+r_2)}{2}\right]}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)}.\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}}\ \frac{\frac{r_1}{x^2}-1}{\left(1+\frac{r_1}{r_2}.x\right)^{\frac{r_1+r_2}{2}}} \qquad x>0$$

که r_1و r_2 به ترتیب درجه آزادی صورت و مخرج خوانده میشود برای مقادیر مختلفr_1و r_2 مقادیر مختلف F(x) از جدول ضمیمه منابع آماری قابل محاسبه است.

آزمون فرض های آماری

7–1 مفاهیم اولیه

در فرض‌هایی که روی پارامتر جامعه اتخاذ می‌شود ممکن است فرض ساده یا مرکب باشد و برای آزمون آنها ممکن است مرتکب خطای نوع اول یا نوع دوم یا تواماً نوع اول و دوم شویم که در این فصل فقط خطای نوع اول و دوم را مورد بحث قرار می‌دهیم. برای روشن شدن مطلب، فرض ساده، فرض مرکب، خطای نوع اول، دوم و توان آزمون را تعریف می‌کنیم.

تعریف 7–1–1 هر فرض آماری که توزیع جامعه را کاملاً مشخص کند فرض ساده و فرضی که توزیع جامعه را مشخص نکند فرض مرکب گویند.

تعریف 7–1–2 در انجام آزمون فرض، اگر فرض آماری درست باشد و ما آن را رد کنیم مرتکب خطای نوع اول و اگر فرض آماری را که پذیرفتیم نادرست باشد مرتکب خطای نوع دوم می‌شویم.

2–7 آزمون فرض برای میانگین توزیع نرمال

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 در مورد میانگین جامعه است.

$$H_0 : \mu = \mu_0 \quad \text{در مقابل} \quad H_1 : \mu > \mu_0$$

فرض H_0 یک فرض ساده در مقابل فرض H_1 که یک فرض مرکب یک طرفه است قرار دارد. در بحث برآوردها، مشخص شده که میانگین نمونه $\overline{X}$ یک برآورد نقطه‌ای برای μ است لذا هرگونه تصمیم‌گیری در مورد فرض فوق براساس $\overline{X}$ خواهد بود.

مقادیر بزرگ $\overline{X}$ پشتیبانی برای فرض H_1 یا تأییدی H_1 است و باعث رد فرض H_0 می‌شود. رد فرض H_0 را می‌توان به صورتهای زیر نیز نوشت:

رد H_0 هم ارز است با : $\quad \bar{X} >$ عدد

رد H_0 هم ارز است با : $\quad \bar{X} - \mu_0 >$ عدد

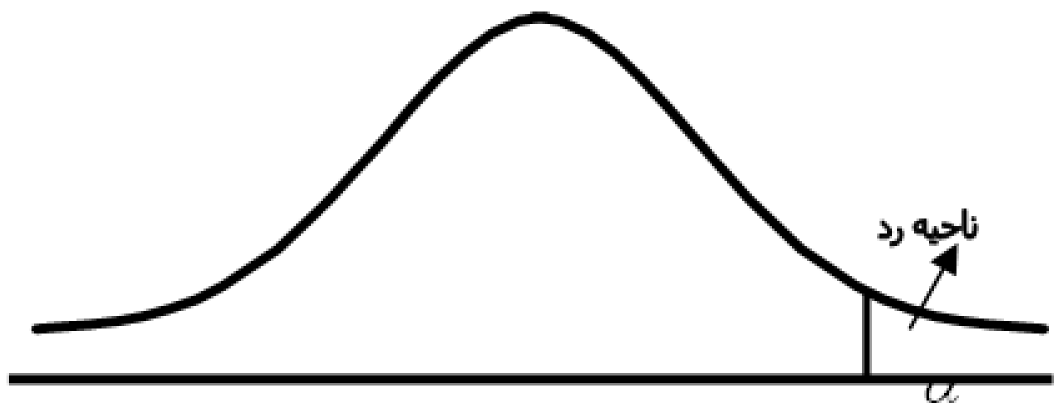
رد H_0 هم ارز است با: $\quad \dfrac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} >$ عدد

اما $\dfrac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$ یک کمیت محوری است و تحت فرض H_0 یک آماره است و از آن به عنوان آماره آزمون یاد می‌کنند.

اگر α سطح معنی دار آزمون باشد با توجه به توزیع آماره آزمون داریم:

$$\alpha = P\left[\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > Z_\alpha \;\middle|\; \text{درست} \quad H_0\right]$$

چون $\frac{\bar{X}-\mu_\circ}{\frac{\sigma}{\sqrt{n}}} > Z_\alpha$ تحت فرض $H_\circ$ دارای توزیع نرمال استاندارد است ناحیه بحرانی و ناحیه قبول برای آزمون فرض فوق به صورت زیر خواهد بود.



مقدار Z_α با توجه به مقدار α از جدول نرمال استاندارد محاسبه می‌شود. برای تصمیم-گیری در مورد فرض $H_\circ : \mu = \mu_\circ$ در مقابل $H_1 : \mu > \mu_\circ$ مراحل زیر را انجام می‌دهیم.

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

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

2- مقدار Z_α را از جدول نرمال استاندارد بدست می‌آوریم.

3- اگر مقدار $Z_\circ$ از مقدار Z_α بزرگتر باشد فرض $H_\circ$ رد می‌شود در غیراینصورت پذیرفته می‌شود.

3-7 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.

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

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

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

پس

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

اگر پیشامد احتمال اخیر را نسبت به $\mu_\circ$ حل کنیم داریم.

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

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 غیراینصورت رد می‌شود.

4–7 آزمون فرض آماری با استفاده از P –مقدار

یکی از روشهای آزمون فرض آماری استفاده از P –مقدار یا مقدار احتمال است. اکنون که نرم افزارهای 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، واریانس معلوم σ^2 باشد و بخواهیم فرض $H_\circ : \mu = \mu_\circ$ را در مقابل فرض $H_1 : \mu < \mu_\circ$ آزمون کنیم.

در این آزمون مقادیر کوچک $\overline{X}$ باعث رد فرض $H_\circ$ می‌شود. اگر α سطح معنی‌دار آزمون باشد فرض $H_\circ$ رد می‌شود اگر احتمال پیشامد $\overline{X} \le \bar{x}$ تحت فرض $H_\circ$ کمتر یا مساوی α باشد یعنی P –مقدار برابر است با

$$\text{مقدار} - P = p\left[\overline{X} \le \bar{x} \middle| \quad H_\circ\right] \le \alpha$$

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

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

5–7 آزمون فرض برای پارامتر توزیع دوجمله‌ای

فرض کنید متغیر تصادفی X دارای توزیع دوجمله‌ای با پارامتر مجهول P و n معلوم باشد. انواع فرضها در مورد پارامتر P عبارتند از

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

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

چون Z دارای توزیع نرمال استاندارد است نواحی قبول یا رد همانند آزمون برای میانگین توزیع نرمال است.

مثال 7–5–1 محموله‌ای شامل 50 رایانه است، اگر 8 رایانه در این محموله معیوب باشد آیا در سطح 5 درصد می‌توان گفت نسبت معیوب در جامعه کمتر از 20 درصد است؟

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

$$\begin{cases} H_\circ : P = 0.2 \\ H_1 : P < 0.2 \end{cases}$$

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 $P[Z < -Z_\alpha] = 0/05 \Rightarrow -Z_\alpha = -1/64$ چون $Z_\circ > -Z_\alpha$ پس فرض $H_\circ$ پذیرفته می‌شود.

6–7 آزمون فرض برای تفاضل میانگین دو جامعه

در تحقیقات کاربردی، مسائل متعددی موجودند که در آنها فرضهایی درباره تفاضل بین دو میانگین دو جامعه مورد توجه است. برای مثال ممکن است بخواهیم متوسط سرعت یا عمر متوسط رایانه‌هایی را که توسط دو سازنده تولید می‌شود باهم مقایسه کنیم. اگر توزیع دو جامعه مورد بررسی معلوم باشد، آزمون فرض برابری میانگین دو جامعه امکان پذیر است.

فرض کنید دو جامعه مورد بررسی مستقل و دارای توزیع نرمال وبه ترتیب دارای میانگین‌های μ_1 و μ_2 و واریانس ${\sigma_1}^2$ و ${\sigma_2}^2$ باشند. انواع فرضها در مورد مقایسه میانگین‌های دو جامعه عبارتند از:

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

$$1)\left|\begin{array}{l} H_\circ : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 > 0 \end{array}\right. \qquad 2)\left|\begin{array}{l} H_\circ : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 < 0 \end{array}\right. \qquad 3)\left|\begin{array}{l} H_\circ : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{array}\right.$$

آماره آزمون برای مقایسه تفاضل $\mu_1 - \mu_2$ براساس نمونه تصادفی از جامعه اول و دوم خواهد بود. اگر $\bar{X}_1, \bar{X}_2$ به ترتیب میانگین نمونه اول و دوم باشند کمیت محوری، همان کمیت محوری بحث شده در فصل 6 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 Z تحت فرض $H_\circ$ دارای توزیع نرمال استاندارد خواهد بود $Z_\circ$ مشاهده شده براساس مشاهده نمونه از دو جامعه برابر است با

$$Z_\circ = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$$

روش آزمون فرض H_0 در مقابل H_1 همانند آزمون برای میانگین توزیع نرمال با واریانس معلوم است. برای آزمون فرض $H_0 : \mu_1 = \mu_2$ در مقابل $H_0 : \mu_1 > \mu_2$ مراحل زیر را انجام می‌دهیم.

1- آماره آزمون را تحت فرض H_0 و مشاهدات محاسبه می‌کنیم (Z_0).

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

3- اگر $Z_0 > Z_\alpha$ باشد فرض H_0 را رد می‌کنیم.

7–7 آزمون فرض برای واریانس جامعه

یکی از راههای بررسی تغییرپذیری جامعه، بررسی یا آزمون فرض درباره واریانس جامعه است. به عنوان مثال مهندس کنترل کیفیت باید مراقبت نماید که تغییرپذیری اندازه‌ها از حد معینی بیشتر نشود یا یک داروساز ممکن بخواهد بداند که آیا میزان تغییر در اثر بخشی یک دارو 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 درفصل 6 نشان داده شد که کمیت $\frac{(n-1)s^2}{\sigma^2}$ یک کمیت محوری است و تحت

فرض $H_\circ$ دارای توزیع کی دو با $(n-1)$ درجه آزادی است. لذا برای انجام آزمون فرضها از آماره فوق استفاده می‌کنیم. برای آزمون فرض $H_0 = \sigma^2 = \sigma_0^2$ در مقابل $H_1 = \sigma^2 > \sigma_0^2$ مراحل زیر را انجام می‌دهیم.

چون s^2 یک برآورد نقطه‌ای برای σ^2 است لذا مقادیر بزرگ s^2 باعث رد $H_\circ$ می‌شود.

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

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

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

آماره $\frac{(n-1)s^2}{\sigma_0^2}$ تحت فرض $H_\circ$ دارای توزیع کی دو با $n-1$ درجه آزادی است. اگر α سطح معنی دار آزمون باشد.

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

مقدار $\chi^2_{(n-1,\alpha)}$ از جدول توزیع کی دو بدست می‌آید.

برای آزمون فرض $H_0 = \sigma^2 - \sigma_0{}^2$ در مقابل $H_1 - \sigma^2 > \sigma_0{}^2$ کافی است آماره $\chi^2_\circ = \dfrac{(n-1)s^2}{\sigma_\circ^2}$ 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.

8–7 آزمون فرض برای نسبت دو واریانس

در این بخش به دنبال شیوه‌ای می‌گردیم 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 براساس آزمون نسبت واریانسها قرار دارد.

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

اگر متغیرتصادفی 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 در فصل 6 دیدیم که 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 درنظر داریم که واریانس نمونه بزرگتر را در صورت قرار می‌دهیم به قسمی که نسبت واریانسهای نمونه همیشه بزرگتر یا مساوی 1 باشد.

برای نمونه برای آزمون فرض $H_0 = \sigma_1^2 = \sigma_2^2$ در مقابل $H_1 = \sigma_1^2 > \sigma_2^2$ مراحل زیر را انجام می‌دهیم.

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

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

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

3- اگر $F_0 > F_{(m-1, n-1, \alpha)}$ باشد فرض $H_\circ$ را رد می‌کنیم در غیر اینصورت آن را می‌پذیریم.

معنی دار بودن از نظر آماری:

مثال: داده های مربوط به آزمایش مقاومت کششی ملاط سیمان پورتلند

ملاط اصلاح شده (Y1)	16.85	16.40	17.21	16.35	16.52	17.04	16.96	17.15	16.59	16.57
ملاط اصلاح نشده (Y2)	16.62	16.75	17.37	17.12	16.98	16.87	17.34	17.02	17.08	17.27

معنی دار بودن از نظر آماری:

مثال: داده های مربوط به آزمایش مقاومت کششی ملاط سیمان پورتلند

ملاط اصلاح شده (Y1)	16.85	16.40	17.21	16.35	16.52	17.04	16.96	17.15	16.59	16.57	N1= 10	Mean1= 16.76	S.D1= 0.316
ملاط اصلاح نشده (Y2)	16.62	16.75	17.37	17.12	16.98	16.87	17.34	17.02	17.08	17.27	N2= 10	Mean2= 17.04	S.D2= 0.248

معنی دار بودن از نظر آماری:

آزمون دونقطه ای t:

$$Z_0 = \frac{\overline{y_1} - \overline{y_2}}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$t_0 = \frac{\overline{y_1} - \overline{y_2}}{S_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \Longrightarrow \quad t_0 = 2.20$$

$$S_P^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 0.081$$

مقدار $t_{0.025،18}$ =۲/۱۰۱

از آنجاییکه $t_0 > t_{0.025،18}$

بنابراین نتیجه می شود که میانگین مقاومت کششی این دو فرمولاسیون با یکدیگر تفاوت دارند.