dalbablog.wordpress.com

فصل هفتم

در فصل قبل برخی توزیع‌های معروف و کاربردی را برای متغیر تصادفی گسسته ارایه نمودیم و خواص و ویژگیهای هر یک را بررسی نمودیم. در این فصل به بررسی توزیعهای مبتنی بر متغیر تصادفی پیوسته می‌پردازیم.

۱.۷ توزیع یکنواخت پیوسته

ساده‌ترین توزیع پیوسته، توزیع یکنواخت می‌باشد. فرض کنید تمام نقاط در بازه (a,b) دارای امکان وقوع یکسان باشند، در این صورت متغیر تصادفی X را که برد آن مقادیر موجود در بازه (a,b) می‌باشد، متغیر تصادفی یکنواخت می‌نامند. که با نماد $X \sim U(a,b)$ نشان داده می‌شود. تابع چگالی احتمال متغیر تصادفی یکنواخت برابر است با:

$$f_X(x) = \begin{cases} \frac{۱}{b-a} & a < x < b \\ ۰ & \text{سایر مقادیر} \end{cases}$$

و به این ترتیب بدست می‌آید که چون X یکنواخت می‌باشد در نظر می‌گیریم:

$$f_X(x) = c \qquad a < x < b$$

در نتیجه:

$$\int_{-\infty}^{+\infty} f_X(x)\,dx = ۱ \quad \Rightarrow \quad \int_a^b c\,dx = c\,x\Big|_a^b = c\,(b-a) = ۱ \quad \Rightarrow \quad c = \frac{۱}{b-a}$$

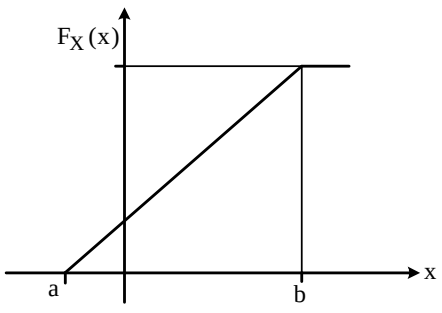
تابع توزیع متغیر تصادفی یکنواخت X برابر است با:

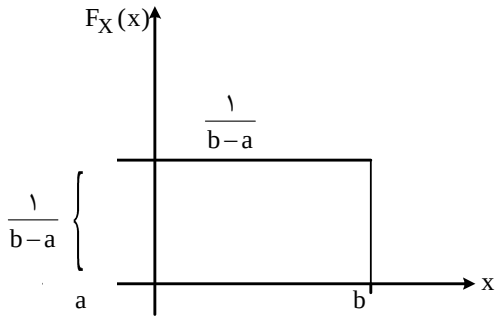
$$f_X(x) = \int_{-\infty}^{t} f_X(x) = \int_a^t \frac{۱}{b-a}\,dx = \frac{t-a}{b-a}$$

پس:

$$f_X(x) = \begin{cases} ۰ & t<a \\ \frac{t-a}{b-a} & a \le t \le b \\ ۱ & t > b \end{cases}$$

۷-۲ در شکل زیر نمودار تابع چگالی و تابع توزیع متغیر تصادفی یکنواخت نشان داده شده است:





مقادیر امید ریاضی واریانس و تابع مولد گشتاور را برای متغیر تصادفی یکنواخت X محاسبه می‌کنیم:

$$E[X] = \int_a^b x\,\frac{۱}{b-a}\,dx = \frac{۱}{b-a}\,\frac{x^۲}{۲}\Big|_a^b =$$

$$= \frac{۱}{b-a}\left(\frac{b^۲}{۲} - \frac{a^۲}{۲}\right) = \frac{(b-a)(a+b)}{۲(b-a)} = \frac{a+b}{۲}$$

dalbablog.wordpress.com

$$E[X^2] = \int_a^b x^2 \frac{1}{b-a}\, dx = \frac{1}{b-a}\left(\frac{x^3}{3}\right)\Big|_a^b$$
$$= \frac{(b-a)(b^2-ab+a^2)}{3(b-a)} = \frac{a^2+ab+b^2}{3}$$

پس:

$$var(X) = E[X^2] - E^2[X] = \frac{a^2+ab+b^2}{3} - \frac{a^2+2ab+b^2}{4} = \frac{(b-a)^2}{12}$$

$$m_X(t) = E[e^{tX}] = \int_a^b e^{tX} \frac{1}{b-a}\, dx = \frac{1}{b-a}\left(\frac{e^{tX}}{t}\right)\Big|_a^b = \frac{e^{tb}-e^{ta}}{t(b-a)}$$

با فرض $a \le c < d \le b$ مقدار احتمال $c \le x \le d$ برابر است با:

$$p(c \le x \le d) = \int_c^d \frac{1}{b-a}\, dx = \frac{d-c}{b-a}$$

ملاحظه می‌کنید که مطابق با تعریف متغیر تصادفی یکنواخت مقدار احتمال تنها به طول بازه (c,d) وابسته است و نه به مقادیر c و d.

۳-۷ مثال ۱: نمرات دانشجویان یک کلاس در درس آمار به طور یکنواخت در فاصله ۱۲ الی ۲۰ توزیع شده است مطلوبست:

الف) تابع توزیع احتمال:

ب) احتمال قرارگیری نمرات در بازه $(۲۰-۱۸)$, $(۱۵-۱۳)$.

ج) میانگین نمرات کلاس، واریانس.

حل: الف) متغیر تصادفی X دارای توزیع یکنواخت است بنابراین:

$$f_X(x) = \begin{cases} \frac{1}{۲۰-۱۲} = \frac{1}{۸} & ۱۲ \le X \le ۲۰ \\ ۰ & \text{سایر مقادیر} \end{cases}$$

ب):

$$P(۱۸ \le X \le ۲۰) = \int_{۱۸}^{۲۰} \frac{1}{۸}\, dx \; \frac{۲۰-۱۸}{۸} = \frac{۲}{۸} = \frac{1}{۴}$$
$$P(۱۳ \le X \le ۱۵) = \int_{۱۳}^{۱۵} \frac{1}{۸}\, dx \; \frac{۱۵-۱۳}{۸} = \frac{۲}{۸} = \frac{1}{۴}$$

ج):

$$E[X] = \frac{۲۰+۱۲}{۲} = ۱۶$$
$$var(X) = \frac{(۲۰-۱۲)^2}{۱۲} = \frac{۶۴}{۱۲} = \frac{۱۶}{۳}$$

۴-۷ ۲.۶ متغیر تصادفی نمایی

یک فرایند پواسون با پارامتر λ که برای یک واحد زمان نظاره می‌شود، را در نظر بگیرید اگر زمان شروع فرایند صفر $(t=۰)$ باشد و T مدت زمانی باشد که باید بگذرد تا اولین پیشامد رخ دهد در این صورت T یک متغیر تصادفی نمایی با پارامتر λ نامیده می‌شود.

تابع توزیع نمایی را با استفاده از تعریف بدست می‌آوریم:

می‌دانیم $f_X(t) = p(X \le t)$ حال اگر $t < ۰$ باشد بنا به تعریف زمان منفی معنا ندارد. بنابراین اگر $t < ۰$ داریم: $p(X \le t) = ۰$ یعنی:

$$f_X(t) = ۰ \qquad t < ۰$$

dalbablog.wordpress.com

حال فرض می‌کنیم $t \geq ۰$ باشد، احتمال اینکه هیچ پیشامدی در بازه $(۰ , t)$ رخ ندهد بنابه تابع چگالی احتمال پواسون برابر است با:

$$f_X(x=۰) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}\Big|_{x=۰} = e^{-\lambda t}$$

بنابراین $p(X > t) = e^{-\lambda t}$ اما $p(X \leq t) = ۱ - p(X > t)$ و داریم:

$$F_X(t) = p(X \leq t) = ۱ - p(X > t) = ۱ - e^{-\lambda t}$$

در نتیجه تابه توزیع متغیر تصادفی نمایی بصورت زیر بدست می‌آید:

$$F_X(t) = \begin{cases} ۱ - e^{-\lambda t} & t > ۰ \\ ۰ & t < ۰ \end{cases}$$

با مشتق‌گیری از $F_X(t)$ تابع چگالی را بدست می‌آوریم:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > ۰ \\ ۰ & x \leq ۰ \end{cases}$$

۵-۷ برای بدست آوردن مقادیر امید و واریانس از تابع مولد گشتاور استفاده می‌کنیم:

$$m_X(t) = E\left[e^{tx}\right] = \int_۰^\infty e^{tx}\ \lambda e^{-\lambda x}\ dx = \lambda \int_۰^\infty e^{tx}\ \lambda e^{-x(\lambda - t)}\ dx$$

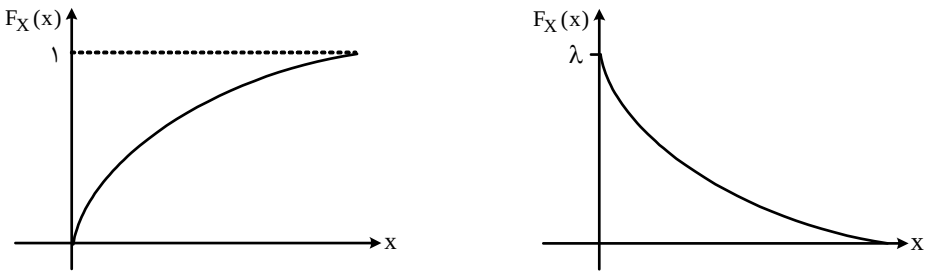
$$= \lambda\ \left(\frac{۱}{t-\lambda}\ e^{-x(\lambda-t)}\right)\Big|_۰^\infty = \frac{-\lambda}{t-\lambda} = \frac{\lambda}{\lambda - t} \quad (t < \lambda)$$

$$E[X] = m'_X(t=۰) = \frac{\lambda}{(\lambda - t)^۲}\Big|_{t=۰} = \frac{۱}{\lambda}$$

$$E[X^۲] = m''_X(t=۰) = \frac{۲\lambda}{(\lambda - t)^۳}\Big|_{t=۰} = \frac{۲}{\lambda^۲}$$

$$var(X) = E[X^۲] - E^۲[X] = \frac{۲}{\lambda^۲} - \frac{۱}{\lambda^۲} = \frac{۱}{\lambda^۲}$$

شکل زیر نمودار تابع توزیع و چگالی را برای متغیر تصادفی نمایش می‌دهد:



متغیر تصادفی نمایی با نماد $X \sim E(\lambda)$ نمایش داده می‌شود و معمولاً از آن به عنوان مدلی برای عمر قطعات و سیستم‌ها استفاده می‌شود.

۶-۷ مثال ۲: اگر لامپهای تولید شده توسط یک کارخانه به طور متوسط هر ۶ ماه یکبار بسوزند مطلوبست:

الف) تابع چگالی احتمال برای مدت زمان کارکرد لامپها.

ب) احتمال اینکه از زمان خرید، یک لامپ حداقل ۲ ماه کار کند؟

ج) احتمال اینکه یک لامپ قبل از ۴ ماه بسوزد؟

د) اگر ۱۰ عدد لامپ خریداری کنیم احتمال اینکه حداقل ۲ عدد از لامپها قبل از ۴ ماه بسوزند؟

dalbablog.wordpress.com

ه) میانگین طول عمر، واریانس و تابع مولد گشتاور را برای متغیر تصادفی X که مدت زمان کارکرد لامپها می‌باشد بدست بیاورید.

۷-۷ حل: الف) متغیر تصادفی X را طول عمر لامپهای خریداری شده در نظر می‌گیریم در این صورت X یک متغیر تصادفی نمایی با پارامتر λ می‌باشد. برای بدست آوردن پارامتر λ ابتدا هر واحد زمانی را یک ماه در نظر می‌گیریم، λ برابر است با تعداد پیشامدها در یک واحد زمانی که در اینجا چون هر ۶ ماه یک لامپ می‌سوزد معادل است با اینکه $\frac{۱}{۶}$ لامپ در هر یک ماه می‌سوزد یعنی $\lambda=\frac{۱}{۶}$ و تابع چگالی برابر است با:

$$f_X(x) = \frac{۱}{۶}\ e^{-\frac{۱}{۶}x} \qquad x > ۰$$

$$= ۰ \qquad \text{سایر مقادیر}$$

ب) احتمال اینکه حداقل ۲ ماه طول بکشد تا یک لامپ بسوزد برابر است با:

$$p(X > ۲) = \int_{۲}^{\infty} \frac{۱}{۶}\ e^{-\frac{x}{۶}}\, dx = \frac{۱}{۶}\,(-۶\, e^{-\frac{x}{۶}})\Big|_{۲}^{\infty}$$

$$= \frac{۱}{۶}\,(۰ - (-۶\, e^{-\frac{۲}{۶}})) = e^{-\frac{۱}{۳}} = ۰/۷۱۶$$

ج) احتمال اینکه یک لامپ قبل از ۴ ماه بسوزد برابر است با:

$$p(X \le ۴) = \int_{۰}^{۴} \frac{۱}{۶}\ e^{-\frac{x}{۶}}\, dx = \frac{۱}{۶}\,(-۶\, e^{-\frac{x}{۶}})\Big|_{۰}^{۴} = ۱-۶\, e^{-\frac{۴}{۶}} = ۰/۴۸۶$$

۷-۸ د) متغیر Y را بصورت زیر تعریف می‌کنیم:

$Y =$ تعداد لامپها از بین ۱۰ لامپ که در کمتر از ۴ ماه می‌سوزند.

طبق تعریف Y یک متغیر تصادفی دو جمله‌ای با پارامتر $n = ۱۰$ و $p = ۰/۴۸۶$ می‌باشد. پارامتر p از بند (ج) بدست می‌آید. احتمال اینکه هر لامپ در کمتر از ۴ ماه بسوزد برابر ۰/۴۸۶ می‌باشد، بنابراین می‌بایستی پارامتر p برابر ۰/۴۸۶ انتخاب شود. حال با احتمال اینکه حداقل ۲ عدد از لامپها قبل از ۴ ماه بسوزند برابر است با:

$$p\,(Y \ge ۲) = ۱-p(Y < ۲)\ = ۱-\left[p(Y=۱)+p(Y=۰)\right]$$

$$= ۱-\left[\binom{۱۰}{۱}\,(۰/۴۸۶)^{۱}\,(۰/۵۱۴)^{۹} + \binom{۱۰}{۰}\,(۰/۴۸۶)^{۰}\ (۰/۵۱۴)^{۱۰}\right]$$

$$= ۱-(۰/۰۱۲ + ۰/۰۰۱۲) = ۰/۹۸۶$$

پس:

$$f_X(x) = \binom{۱۰}{y}\,(۰/۴۸۶)^{y}\ (۰/۵۱۴)^{۱۰-y}$$

اگر امید ریاضی $f_Y(y)$ را محاسبه کنیم داریم: $E[X] = np = ۱۰ \times ۰/۴۸۶ = ۴/۸۶$ یعنی تقریباً از هر ۱۰ عدد لامپ خریداری شده بطور متوسط ۵ لامپ در کمتر از ۴ ماه می‌سوزند.

ه) میانگین طول عمر لامپها عبارتست از: $E[X] = \frac{۱}{\lambda} = \frac{۱}{\frac{۱}{۶}} = ۶$ که با توجه به اینکه واحد زمان را هر یک ماه در نظر گرفتیم بنابراین میانگین طول عمر هر لامپ ۶ ماه می‌باشد که از صورت مثال نیز همین انتظار می‌رفت.

dalbablog.wordpress.com

$$\text{var}(X) = \frac{۱}{\lambda^۲} = \frac{۱}{(\frac{۱}{۶})^۲} = ۳۶$$

$$m_X(t) = \frac{\lambda}{\lambda - t} = \frac{\frac{۱}{۶}}{\frac{۱}{۶} - t} = \frac{۱}{۱ - ۶t}$$

۷-۹ ۶ . ۲ . ۱ رابطه توزیع هندسی و توزیع نمایی

توزیع هندسی طبق تعریف عبارتست از تعداد آزمایشها قبل از رسیدن به اولین پیروزی توزیع نمایی نیز از جهت تعریف تشابه زیادی با توزیع هندسی دارد. توزیع نمایی هم عبارتست از مدت زمانی که در یک فرایند پواسون با پارامتر λ باید سپری شود تا اولین پیشامد رخ دهد توجه کنید که میانگین توزیع هندسی $\frac{۱}{p}$ و توزیع نمایی $\frac{۱}{\lambda}$ می‌باشد. متغیرهای تصادفی هندسی و نمایی در یک خاصیت ویژه مشترک می‌باشند که هیچ متغیر تصادفی گسسته یا پیوسته دیگری این حالت را ندارد. ویژگی فوق به بی‌حافظگی معروف است و به صورت زیر می‌باشد:

برای توزیع نمایی: $p(X > a + b \mid X > a) = p(X > b)$

برای توزیع هندسی: $p(X = a + b \mid X \geq a) = p(X = b)$

اثبات برای توزیع نمایی: فرض می‌نیم A برابر با پیشامد $X > a$ و B برابر پیشامد $X > b$ و C برابر پیشامد $X > a+b$ باشد در این صورت باید ثابت کنیم: $p(C|A) = p(B)$ می‌دانیم:

$$p(C|A) = \frac{p(C \cap A)}{p(A)}$$

از آنجا که پیشامد C زیر مجموعه‌ای از پیشامد A است $(\{X > a+b\} \subset \{X > a\})$ بنابراین $C \cap A = C$ و داریم:

$$p(C|A) = \frac{p(C \cap A)}{p(A)} = \frac{p(C)}{p(A)}$$

۷-۱۰ حال مقادیر $p(A)$ و $p(B)$ و $p(C)$ را بدست می‌آوریم:

$$p(A) = p(X > a) = \int_a^{\infty} \lambda e^{-\lambda x}\, dx = e^{-\lambda a}$$

$$p(B) = p(X > b) = \int_b^{\infty} \lambda e^{-\lambda x}\, dx = e^{-\lambda b}$$

$$p(C) = p(X > a+b) = \int_{a+b}^{\infty} \lambda e^{-\lambda x}\, dx = e^{-\lambda(a+b)}$$

$$p(C|A) = \frac{p(C)}{p(A)} = \frac{e^{-\lambda(a+b)}}{e^{-\lambda a}} = e^{-\lambda b} = p(B)$$

بی‌حافظگی به این مفهوم است که اگر قطعه‌ای یا سیستمی به مدت a واحد زمان کار کرده باشد، احتمال اینکه حداقل b واحد زمان دیگر نیز کار کند $(a+b)$ برابر است با احتمال اینکه قطعه یا سیستم از لحظه صفر بخواهد حداقل b واحد زمان کار کند.

۷-۱۱ مثال ۳: در یک کارخانه ماشین‌های تولیدی هر یک ماه نیازمند سرویس تعمیرات باشند. اگر یک ماشین تولیدی ۶ ماه بدون تعمیرات کار کرده باشد احتمال اینکه در طول ماه بعد نیازمند تعمیات باشد چقدر است؟

حل: متغیر تصادفی X را مدت زمان لازم قبل از اولین تعمیر در نظر می‌گیریم در این صورت X یک متغیر تصادفی نمایی با پارامتر $\lambda = ۱$ می‌باشد. می‌بایستی مقدار احتمال زیر را محاسبه کنیم:

$$p(X > ۶+۱ \mid X > ۶) = p(X > ۷ \mid X > ۶)$$

$$= p(X > ۱) = \int_۱^{+\infty} e^{-x}\, dx = e^{-۱} = ۰/۳۶$$

۷-۱۲ ۶ . ۳ توزیع گاما

dalbablog.wordpress.com

اگر متغیرهای تصادفی $X_1, X_2, \cdots, X_n$ دارای توزیع نمایی با پارامتر λ باشند در این صورت متغیر تصادفی Y را که برابر است با مجموع متغیرهای تصادفی X_1 تا X_n، یک متغیر تصادفی گاما می‌نامیم.

$$Y = X_1 + X_2 + \cdots + X_n$$

تابع چگالی متغیر تصادفی گاما برابر است با:

$$f_X(x) = \frac{(\lambda x)^{n-1}}{T(n)} \lambda e^{-\lambda x} \qquad x \geq 0$$

$$= 0 \qquad \text{سایر مقادیر}$$

که در آن $T(n)$ تابع گاما می‌باشد و به صورت زیر تعریف می‌شود:

$$T(n) = \int_0^\infty x^{n-1} e^{-x} \, dx$$

$T(n)$ به ازای هر $n > 0$ موجود است و روابط زیر برای تابع گاما برقرار است:

۱- $T(n+1) = n\, T(n)$

۲- $T(n) = (n-1)!$ اگر n عدد طبیعی باشد

۳- $T(\frac{1}{2}) = \sqrt{\pi}$

از آنجا که متغیر تصادفی گاما مجموع n متغیر تصادفی نمایی است پس می‌توان مقادیر امید، واریانس و تابع مولد گشتاور را با استفاده از این خاصیت بدست آورد:

$$E[X] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \frac{1}{\lambda} = \frac{n}{\lambda}$$

$$var(X) = var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} var(X_i) = \sum_{1}^{n} \frac{1}{\lambda^2} = \frac{n}{\lambda^2}$$

$$m_X(t) = E\left[e^{tX}\right] = E\left[e^{t\sum_{1}^{n} X_i}\right] = E\left[e^{tX_1}\right] E\left[e^{tX_2}\right] \cdots E\left[e^{tX_n}\right]$$

$$= \left(\frac{\lambda}{\lambda - t}\right)\left(\frac{\lambda}{\lambda - t}\right) \cdots \left(\frac{\lambda}{\lambda - t}\right) = \left(\frac{\lambda}{\lambda - t}\right)^n$$

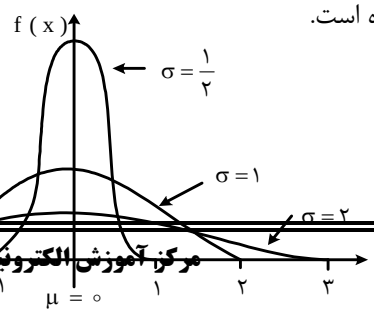
۷-۱۳ توزیع نرمال

مهمترین توزیع پیوسته که آنرا بررسی می‌کنیم توزیع نرمال می‌باشد، بسیاری از پدیده‌های طبیعی مثل قد و وزن افراد، نمرات درسی، میزان محصول در طول سال از توزیع نرمال پیروی می‌کنند.

متغیر تصادفی X دارای توزیع نرمال است و آنرا بصورت $X \sim N(\mu, \sigma^2)$ نمایش می‌دهیم اگر تابع چگالی احتمال بفرم زیر باشد:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < +\infty \,,\ \mu \in R \,,\ \sigma > 0$$

همانطور که ملاحظه می‌نید دو پارامتر توزیع نرمال همان میانگین و واریانس متغیر تصادفی X می‌باشند. مد، میانه و میانگین توزیع نرمال با یکدیگر برابر می‌باشند. در شکل زیر توزیع نرمال با مقادیر واریانس متفاوت نشان داده شده است.



مرکز آموزش الکترونیکی دانشگاه علم و صنعت ایران

dalbablog.wordpress.com

تغیر میانگین در توزیع نرمال تنها نمودار تابع را به سمت راست یا چپ منتقل می‌کند. منحنی نرمال دارای خواص زیر می‌باشد:

۱- نقاط عطف منحنی عبارتند از $x_1=\mu+\sigma$, $x_2=\mu-\sigma$.

۲- منحنی نسبت به خط $x=\mu$ متقارن است بنابراین:

$$۱-\ f_X(\mu-a)=f_X(\mu+a) \qquad (a>۰)$$

$$۲-\begin{cases} p(X>a)=p(X<-a) \\ ۱-F_X(x)=F_X(-x) \end{cases}$$

۳- مساحت سطح زیر منحنی نمودار نرمال برابر واحد می‌باشد. زیرا بوضوح داریم: $\int_{-\infty}^{+\infty} f(x)\,dx=۱$

۷-۱۴ برای بدست آوردن میانگین و واریانس از تابع مولد گشتاور استفاده می‌کنیم که بصورت زیر بدست می‌آید:

$$m_X(t)=E\left[e^{tX}\right]=\int_{-\infty}^{+\infty} e^{tx}\,\frac{۱}{\sigma\sqrt{۲\pi}}\,e^{\frac{-(x-\mu)^۲}{۲\sigma^۲}}\,dx$$

$$=\frac{۱}{\sigma\sqrt{۲\pi}}\int_{-\infty}^{+\infty} e^{tx-\frac{(x-\mu)^۲}{۲\sigma^۲}}\,dx$$

با تبدیل توان e به صورت مربع کامل داریم:

$$tx-\frac{(x-\mu)^۲}{۲\sigma^۲}=\frac{۱}{۲\sigma^۲}\left[x^۲-۲\mu x+\mu^۲-۲\sigma^۲+x\right]$$

$$=\frac{۱}{۲\sigma^۲}\left[(x-\mu x-\sigma^۲t)^۲-۲\sigma^۲t\mu-\sigma^۴t^۲\right]$$

$$=\frac{۱}{۲\sigma^۲}\left[(x-\mu-\sigma^۲t)^۲\right]+t\mu+\frac{\sigma^۲t^۲}{۲}$$

به این ترتیب:

$$m_X(t)=e^{\mu t+\frac{(\sigma t)^۲}{۲}}\int_{-\infty}^{+\infty}\underbrace{\frac{۱}{\sigma\sqrt{۲\pi}}\,e^{\frac{-(x-\mu-\sigma^۲t)^۲}{۲\sigma^۲}}}_{g_X(x)}\,dx$$

مقدار انتگرال $\int_{-\infty}^{+\infty} g(n)\,dx$ برابر یک می‌باشد زیرا $g_X(x)$ خود یک تابع نرمال با پارامترهای $N(\mu+\delta^۲t\ ,\ \delta^۲)$ می‌باشد. بنابراین مقدار تابع مولد گشتاور برابر است با:

$$m_X(t)=\ e^{\mu t+\frac{\sigma^۲t^۲}{۲}}$$

حال داریم:

$$E[X]=m'_X(t)\Big|_{t=۰}=\left(\mu+\frac{\sigma^۲}{۲}(۲t)\right)e^{\mu t+\frac{\sigma^۲t^۲}{۲}}\Big|_{t=۰}=\mu$$

dalbablog.wordpress.com

$$E[X^2] = m''_X(t)\Big|_{t=0} = \sigma^2 (e^{\mu t+\frac{\sigma^2 t^2}{2}}) \ (\mu + \frac{\sigma^2}{2}(2t))^2 \ e^{\mu t+\frac{\sigma^2 t^2}{2}} \Big|_{t=0} = \sigma^2 + \mu^2$$

$$\Rightarrow \text{var}(X) = E[X^2] - E^2[X] = \sigma^2 + \mu^2 = \sigma^2$$

۷-۱۵ تابع توزیع متغیر تصادفی نرمال X عبارتست از:

$$F_X(t) = \int_{-\infty}^{t} \frac{1}{\delta\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\delta^2}} \ dx$$

برای انتگرال فوق یک تابع ائلیه نمی‌توان بدست آورد به همین دلیل مجبور هستیم از روشهای عددی یک جدول برای مقادیر متفاوت t بدست بیاوریم. اما از آنجا که دو پارامتر μ و δ^2 نیز متغیر می‌باشند می‌بایستی روشی بدست بیاوریم که بتوان مقدار احتمال را بدون وابستگی به μ و δ^2 بدست آورد. متغیر تصادفی Z را بصورت زیر تعریف می‌کنیم:

$$Z = \frac{X-\mu}{\sigma}$$

بوضوح Z متغیر تصادفی نرمال با میانگین ۰ و واریانس ۱ می‌باشد. $Z \sim N(0,1)$ متغیر تصادفی Z را نرمال استاندارد می‌نامند. (که در فصل اول نیز برای نمونه‌های گرفته شده از یک جامعه معرفی شد).

$$E[Z] = E\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}[E[X]-\mu] = \frac{1}{\sigma}(\mu-\mu) = 0$$

$$\text{var}(Z) = \text{var}\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}\text{var}(X-\mu) = \frac{\sigma^2}{\sigma^2} = 1$$

$$m_Z(t) = e^{\frac{t^2}{2}}$$

حال با استفاده از تغیر متغیر $Z = \frac{X-\mu}{\sigma}$ می‌توان به سادگی مقادیر احتمال را از روی جدول توزیع متغیر تصادفی نرمال استاندارد بدست آورد.

تابع چگالی متغیر تصادفی نرمال استاندارد که آنرا با $\eta_Z(z)$ یا $\varphi(z)$ نمایش می‌دهیم برابر است با:

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

تابع توزیع متغیر تصادفی نرمال استاندارد برابر است با:

$$\varphi(Z) = N_Z(z) = p_z(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{Z^2}{2}} \ dx$$

جدول ضمیمه مقادیر تابع توزیع را برای $-3/5 \le Z \le 3/5$ نشان می‌دهد.

۷-۱۶ مثال ۴: با استفاده از جدول مقادیر احتمالات زیر را بدست بیاورید؟

الف) $p(Z<0) \ , \ p(Z>0)$

ب) $p(Z<1) \ , \ p(|Z|<\frac{3}{2})$

ج) $p(-1<Z<0) \ , \ p(1<Z<3)$

حل: الف)

dalbablog.wordpress.com

$$p(Z < ۰) = \varphi(۰) = \frac{۱}{۲}$$

$$p(Z > ۰) = ۱ - \varphi(۰) = ۱ - \frac{۱}{۲} = \frac{۱}{۲}$$

ب)

$$p(Z < ۱) = \varphi(۱) = ۰/۸۴۱۳$$

$$p(|Z| < \frac{۳}{۲}) = p(-\frac{۳}{۲} < Z < \frac{۳}{۲}) = \varphi(\frac{۳}{۲}) - \varphi(\frac{۳}{۲})$$

$$= ۰/۹۳۳۲ - ۰/۰۶۶۸ = ۰/۸۶۶۴$$

ج)

$$p(-۱ < Z < ۰) = \varphi(۰) - \varphi(۱) = \frac{۱}{۲} - ۰/۱۵۸۷ = ۰/۳۴۱۳$$

$$p(۱ < Z < ۳) = \varphi(۳) - \varphi(۱) = ۰/۹۹۸۷ - ۰/۸۴۱۳ = ۰/۱۵۷۴$$

با توجه به مثال فوق می‌وان خواص زیر را برای متغیر تصادفی نرمال استاندارد به دست آورد:

۱. $\varphi(-\infty) = ۰$

۲. $\varphi(+\infty) = ۱$

۳. $\varphi(a) = ۱ - \varphi(-a)$

۷-۱۷ ۱.۴.۶ محاسبه مقادیر احتمال متغیر تصادفی نرمال

فرض کنید X یک متغیر تصادفی نرمال با میانگین μ و واریانس $\sigma^۲$ باشد $(X \sim N(\mu, \sigma^۲))$ برای بدست آوردن احتمال $p(a < x < b)$ داریم:

$$p(a < x < b) = p(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma})$$

$$= P(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma})$$

$$= \varphi(\frac{b-\mu}{\sigma}) - \varphi(\frac{a-\mu}{\sigma})$$

۷-۱۸ مثال ۵: اگر $X \sim N(۵, ۱۶)$ مطلوبست محاسبه $p(۲ \le X < ۷)$.

حل: داریم $\mu = ۵$, $\sigma^۲ = ۱۶$ پس:

$$p(۲ \le X < ۷) = p(\frac{۲-۵}{۴} \le \frac{X-۵}{۴} < \frac{۷-۵}{۴})$$

$$= p(-\frac{۳}{۴} \le Z < \frac{۱}{۲})$$

$$= \varphi(\frac{۱}{۲}) - \varphi(-\frac{۳}{۴}) = ۰/۶۹۱۵ = ۰/۲۲۶۶ = ۰/۴۶۴۹$$

۷-۱۹ مثال ۶: نمرات دانشجویان یک کلاس از توزیع نرمال با میانگین ۱۵ و واریانس ۴ پیروی می‌کند اگر بدانیم تمامی نمرات از ۱۰ بیشتر هستند احتمال اینکه نمرات بین ۱۲ تا ۱۶ باشد چقدر است؟

حل: متغیر تصادفی X نرمال می‌باشد $(X \sim N(۱۵, ۴))$ می‌بایستی احتمال شرطی زیر را محاسبه کنیم:

dalbablog.wordpress.com

$$p(۱۲ < X < ۱۶) \mid X > ۱۰) = \frac{p(۱۲ < X < ۱۶ , X > ۱۰)}{p(X > ۱۰)}$$

$$= \frac{p(۱۲ < X < ۱۶)}{p(X > ۱۰)} = \frac{p(\frac{۱۲-۱۵}{۲} < \frac{X-۱۵}{۲} < \frac{۱۶-۱۵}{۲})}{p(\frac{X-۱۵}{۲} > \frac{۱۰-۱۵}{۲})}$$

$$= \frac{p(-\frac{۳}{۲} < Z < \frac{۱}{۲})}{p(Z > -\frac{۵}{۲})} = \frac{\varphi(\frac{۱}{۲}) - \varphi(-\frac{۳}{۲})}{\varphi(\frac{۵}{۲})} = \frac{۰/۶۹۱۵ - ۰/۰۶۶۸}{۰/۹۹۳۸} = ۰/۶۲۸۵$$

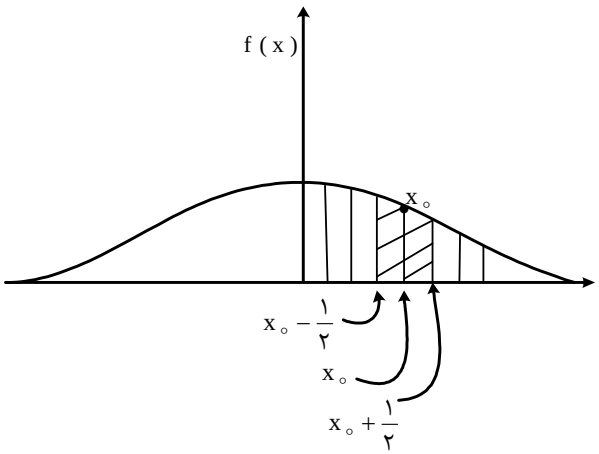
۲۰-۷ ۶ . ۴ . ۲ تقریب توزیع دو جمله‌ای به توزیع نرمال

اگر در یک توزیع دو جمله‌ای تعداد آزمایشها بسیار زیاد باشد یعنی $n \to +\infty$ و در عین حال احتمال پیروزی و شکست تقریباً برابر باشند یعنی $p \approx q \approx \frac{۱}{۲}$ در این صورت می‌توان مقادیر احتمال دو جمله‌ای را با استفاده از توزیع نرمال بدست آورد.

برای تقریب توزیع دو جمله‌ای به نرمال مقادیر امید و واریانس را برابر قرار می‌دهیم یعنی:

$$\mu = np \quad , \quad \sigma^۲ = npq$$

حال به شکل زیر توجه کنید:



برای بدست آوردن $p(X = x_۰)$

در توزیع دو جمله‌ای کافیست.

در توزیع نرمال مساحت مشخص شده در شکل را بدست بیاوریم زیرا می‌دانیم $p(X = x_۰) = f(x_۰)$. و از طرفی مساحت حاشور خورده تقریباً با مساحت یک ذوذنقه برابر است:

$$\text{مساحت ذوذنقه حاشور خورده} = \text{ارتفاع} \times \frac{\text{مجموع دو قاعده}}{۲} = (x_۰ + \frac{۱}{۲} - (x_۰ - \frac{۱}{۲}) \times \frac{f(x_۰ + \frac{۱}{۲}) + f(x_۰ - \frac{۱}{۲})}{۲}$$

$$= \frac{f(x_۰ + \frac{۱}{۲}) + f(x_۰ - \frac{۱}{۲})}{۲} \approx f(x_۰)$$

حال داریم:

$$p(X \subset x_۰) = p(x_۰ - \frac{۱}{۲} < X < x_۰ + \frac{۱}{۲})$$

$$= P \frac{x_۰ - \frac{۱}{۲} - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{x_۰ + \frac{۱}{۲} - \mu}{\sigma}$$

$$= \varphi(\frac{x_۰ + \frac{۱}{۲} - \mu}{\sigma}) - \varphi(\frac{x_۰ - \frac{۱}{۲} - \mu}{\sigma})$$

dalbablog.wordpress.com

و با جاگذاری مقادیر $\mu = np \quad , \quad \delta^2 = npq$ بدست می‌آوریم:

$$p(X=x_\circ) = \varphi\left(\frac{x_\circ + \frac{1}{2} - np}{\sqrt{npq}}\right) - \varphi\left(\frac{x_\circ - \frac{1}{2} - np}{\sqrt{npq}}\right)$$

به همین ترتیب:

$$p(X \le x_\circ) = \varphi\left(\frac{x_\circ + \frac{1}{2} - np}{\sqrt{npq}}\right)$$

$$p(a < X \le b) = \varphi\left(\frac{b + \frac{1}{2} - np}{\sqrt{npq}}\right) - \varphi\left(\frac{a - \frac{1}{2} - np}{\sqrt{npq}}\right)$$

۷-۲۱ مثال ۷: احتمال بارندگی در طول یک ماه در یک شهر ۰/۳ . مطلوبست احتمال اینکه حداقل در ۵۰ شهر از ۲۰۰ شهر کشور در طول یک ماه آینده باران ببارد؟

حل: در این مثال X یک متغیر تصادفی دو جمله‌ای با پارامترهای $X \sim B(۲۰۰ , ۰/۳)$ می‌باشد می‌خواهیم احتمال $p(X > ۵۰)$ را محاسبه کنیم:

$$p(X > ۵۰) = ۱ - p(X \le ۵۰) = ۱ - \varphi\left(\frac{۵۰ + \frac{1}{2} - (۲۰۰ \times ۰/۳}{\sqrt{۲۰۰ \times ۰/۳ \times ۰/۷}}\right)$$

$$= ۱ - \varphi\left(\frac{-۹/۵}{۶/۴}\right) = ۱ - \varphi(-۱/۴۸) = ۱ - ۰/۰۶۹۴ = ۰/۹۳۰۶$$

۷-۲۲ ۳ . ۴ . ۶ تقریب توزیع پواسون به توزیع نرمال

برای توزیعهای پواسون با λ بزرگ (بزرگتر یا مساوی ۵) می‌توان مقادیر احتمال توزیع پواسون را با استفاده از توزیع نرمال بدست آورد در این حالت روابط دقیقاً مشابه روابط تقریب دو جمله‌ای به نرمال می‌باشند با این تفاوت که به جای μ , δ^2 داریم: $\mu = \lambda \quad , \quad \sigma^2 = \lambda$

$$p(X=x_\circ) = \varphi\left(\frac{x_\circ + \frac{1}{2} - \lambda}{\lambda}\right) - \varphi\left(\frac{x_\circ - \frac{1}{2} - \lambda}{\lambda}\right)$$

$$p(X \le x_\circ) = \varphi\left(\frac{x_\circ + \frac{1}{2} - \lambda}{\lambda}\right)$$

$$p(a < X \le b) = \varphi\left(\frac{b + \frac{1}{2} - \lambda}{\lambda}\right) - \varphi\left(\frac{a - \frac{1}{2} - \lambda}{\lambda}\right)$$

۷-۲۳ مثال ۸: تلفن یک شرکت در هر ساعت تقریباً ۲۰ مرتبه زنگ می‌زند مطلوبست احتمال اینکه تلفن در طول ۲ ساعت آینده حداقل ۲۰ و حداکثر ۴۰ بار زنگ بزند؟

حل: X یک متغیر تصادفی پواسون می‌باشد واحد زمانی را هر یک ساعت در نظر می‌گیریم در این صورت $\lambda = ۱۰$ می‌باشد و می‌بایستی احتمال $p(۲۰ > X < ۴۰)$ را محاسبه کنیم:

$$p(۲۰ > X < ۴۰) = \varphi\left(\frac{۴۰ + \frac{1}{2} - ۱۰}{۱۰}\right) - \varphi\left(\frac{۲۰ - \frac{1}{2} - ۱۰}{۱۰}\right)$$

$$= \varphi(۳/۰۵) - \varphi(۰/۹۵) = ۰/۹۹۸۹ - ۰/۸۲۸۹ = ۰/۱۷$$

dalbablog.wordpress.com

فصل هشتم

توزیع‌های نمونه‌ای

در فصل‌های پیش نشان دادیم که با داشتن تابع توزیع یک متغیر تصادفی چگونه می‌توان مقدار احتمالات مورد نظر را بدست آورد. اما در عمل در بسیاری از اوقات می‌دانیم که یک متغیر تصادفی (مثلاً بارش باران در طول سال) از چه توزیع پیروی می‌کند اما مشکل اینجاست که مقدار پارامترهای توزیع را نمی‌دانیم. به عنوان مثال تعداد دفعات برنده شدن حساب بانکی یک شخص در طول ۱۰ سال یک متغیر تصادفی دو جمله‌ای با پارامترهای $X \sim B(۱۰ , p)$ می‌باشد. اما p یا همان احتمال برنده شدن حساب را نمی‌دانیم بنابراین برای بدست آوردن آن نیازمند روشهایی هستیم که بتوان پارامترهای مجهول را برآورد نمود در این فصل و فصل‌های بعدی روشهایی را برای این منظور ارایه می‌کنیم.

۸ . ۱ نمونه تصادفی

یک جمعیت آماری را با متغیر تصادفی X و توزیع احتمال $f_X(x)$ در نظر بگیرید. از این جمعیت n عضو را انتخاب می‌کنیم، این n عضو را یک نمونه nتایی گوییم. اولین عضو از نمونه را با $X_۱$ و دومین عضو را با $X_۲$ و به همین ترتیب nامین عضو انتخابی را با X_n نمایش می‌دهیم. به این ترتیب نمونه nتایی به صورت $X_۱, X_۲, \ldots, X_n$ خواهد بود که در آن هر یک از X_i یا خود یک متغیر تصادفی یا همان توزیع $f_X(x)$ می‌باشند. و از آنجا که تعداد اعضای جمعیت آماری نامتناهی یا بسیار زیاد می‌باشد $(N \to \infty)$. در این صورت انتخاب X_i یا مستقل از X_j $(i \neq j)$ می‌باشد. و این یعنی توزیع توام $X_۱$ تا X_n برابر با حاصل ضرب توزیع $X_۱$ تا X_n می‌باشد. حال با توجه به مطالب فوق تعریف نمونه تصادفی را بصورت زیر ارایه می‌کنیم:

تعریف:

گوییم $X_۱, X_۲, \ldots, X_n$ یک نمونه تصادفی از متغیر تصادفی X می‌باشد اگر و فقط اگر تابع احتمال $f(X_۱, X_۲, \ldots, X_n)$ بصورت زیر باشد:

$$f(X_۱, X_۲, \ldots, X_n) = \prod_{i=۱}^{n} f_X(x_i) \qquad \forall (x_۱, x_۲, \ldots, x_n)$$

حال در مثال‌های بعدی نحوه بکارگیری نمونه‌گیری را مشخص می‌کنیم.

مثال ۱: مدیر یک کارخانه کنسروسازی ادعا می‌ند که از هر ۱۰۰۰۰ کنسرو تولید شده تنها ۵ قوطی دارای اشکال می‌باشد، به وضوح برای اثبات ادعای وی نمی‌توانیم تمام قوطی‌ها را بررسی کنیم بنابراین از میان تمام قوطی‌ها ۱۰ قوطی انتخاب می‌کنیم (در این مثال برای سادگی تعداد کمتری نمونه انتخاب کردیم) و آنها را مورد بررسی قرار می‌دهیم. مطلوبست:

الف) تعریف متغیر تصادفی X و نمونه‌های تصادفی x_i $\quad i=۱, ۲, \ldots, ۱۰$

ب) تابع چگالی احتمال توام نمونه‌های تصادفی x_i .

حل: الف) متغیر تصادفی X را به این صورت تعریف می‌کنیم:

$$X = \begin{cases} ۰ & \text{اگر قوطی سالم باشد} \\ ۱ & \text{اگر قوطی ناسالم باشد} \end{cases}$$

به این ترتیب نمونه‌های تصادفی X_i به این صورت تعریف خواهد شد:

$$X = \begin{cases} ۰ & \text{اگر نمونه } i \text{ام سالم باشد} \\ ۱ & \text{اگر نمونه } i \text{ام ناسالم باشد} \end{cases}$$

بنابراین متغیر تصادفی X یک متغیر تصادفی برنولی با پارامتر p می‌باشد.

نمونه تصادفی عبارتست از $X_۱, X_۲, \ldots, X_n$ که هر یک نیز به نوبه خود دارای توزیع برنولی با همان پارامتر p می‌باشند. یعنی داریم:

$$f_{Xi}(xi) = p^{x_i}(۱-p)^{۱-x_i} \qquad i=۱, ۲, \ldots, ۱۰$$

$$x_i = ۰ \text{ یا } ۱$$

$$= ۰ \qquad \text{برای سایر مقادیر}$$

ب) از آنجا که نمونه‌ها از بین تعداد زیادی (۱۰/۰۰۰) کنسرو انتخاب می‌شوند بنابراین انتخاب با جایگذاری و بدون جایگذاری تقریباً معادل یکدیگر می‌باشند و می‌توان گفت که X_iها از یکدیگر مستقل می‌باشند. در این صورت تابع چگالی توام X_iها عبارتست از:

dalbablog.wordpress.com

$$f(x_1, x_2, \dots, x_{10}) = \prod_{i=1}^{10} \underset{X}{f}(x_i) = p^{\sum_{i=1}^{10} x_i} \quad (1-p)^{10-\sum_{i=1}^{10} x_i}$$

مثال ۲: می‌دانیم که طول قد دانشجویان یک دانشگاه یک متغیر تصادفی نرمال با میانگین μ و واریانس σ^2 می‌باشد. اگر به تصادف ۱۰ نفر از دانشجویان انتخاب کنیم و X_1 را طول قد نفر اول و X_2 را طول قد نفر دوم و به همین ترتیب X_{10} را طول قد نفر دهم قرار دهیم. در این صورت مطلوبست:

الف) تابع چگالی احتمال توام X_1 تا X_{10}.

ب) احتمال اینکه قد هر ۱۰ نفر از ۱/۵ متر بیشتر باشد؟

ج) احتمال اینکه هیچ کس قدی بلندتر از ۲ متر نداشته باشد؟

حل: الف) از آنجا که هر یک از X_iها نمونه‌های تصادفی می‌باشند که از یکدیگر مستقل بوده و از توزیع نرمال با میانگین μ و واریانس σ^2 پیروی می‌کند بنابراین:

$$f(X_1, X_2, \dots X_n) = \prod_{i=1}^{10} \underset{X}{f}(x_i) = \prod_{1}^{10} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma^{10}(2\pi)^5} e^{\frac{-\sum_{i=1}^{10}(x_i-\mu)^2}{2\sigma^2}}$$

ب) احتمال اینکه قد هر ۱۰ نفر از ۱/۵ متر بیشتر باشد برابر است با:

$$p(X_1 > 1/5) = 1 - p(X_1 \le 1/5) = 1 - p\left(\frac{X_1-\mu}{\sigma} < \frac{1/5-\mu}{\sigma}\right)$$

$$= 1 - N_Z\left(Z < \frac{1/5-\mu}{\sigma}\right)$$

برای X_1 تا X_{10} داریم:

$$p(X_1, \dots, X_{10} > 1/5) = \left(1 - N_Z\left(\frac{1/5-\mu}{\sigma}\right)\right)^{10}$$

ج) مطابق بند (ب) مقدار این احتمال هم برابر است با:

$$p(X_1, \dots, X_{10} < 2) = \left(N_Z\left(\frac{2-\mu}{\sigma}\right)\right)^{10}$$

همانطور که ملاحظه می‌کنید تابع چگالی توام و نتایج احتمالات به مقادیر μ و σ^2 وابسته می‌باشد. در فصل برآورد نشان می‌دهیم که با برآورد این پارامترها می‌توان مقادیر احتمالات را محاسبه نمود.

۲ . ۸ آماره‌ها

اگر با استفاده از عضوهای یک نمونه تصادفی (مثل X_1 , X_2 , ...) یک تابع بسازیم که به مقادیر مجهول پارمتر وابستگی نداشته باشد به آن تابع یک آماره می‌گوییم.

به این ترتیب اگر $X_1, X_2, \dots, X_n$ یک نمونه تصادفی باشند توابع زیر همگی آماره می‌باشند:

$$X_2 + X_4 + X_8 \quad , \quad \frac{X_n}{X_1} \quad , \quad \frac{1}{n}\sum_{i=1}^{n}(X_i)^2$$

$$\prod_{i=1}^{n} X_i \quad , \quad \sum_{i=1}^{n}(X_i - 2)^2$$

dalbablog.wordpress.com

اما توابع زیر بشرطی آماره می‌باشند که مقادیر μ و σ^2 مجهول نباشند:

$$X_n - \mu \quad ، \quad \frac{1}{n-1}\sum_{1}^{n}(X_i-\mu)^2 \quad ، \quad \frac{X_2 - X_1}{2\sigma^2} \quad ، \quad \frac{X_1-\mu}{\sigma}$$

به عنوان مثال $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ میانگین نمونه تصادفی می‌باشد که یک آماره هم می‌باشد. اما این میانگین با پارامتر μ در جامعه تفاوت می‌کند زیرا میانگین نمونه از نمونه‌ای به نمونه دیگر تغیر می‌ند اما پارامتر جامعه μ همواره ثابت است.

در استنباط آماری با گرفتن نمونه‌هایی از جامعه و بدست آوردن آماره‌هایی مثل $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ یا $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ می‌خواهیم نتیجه‌گیری‌هایی در مورد پارامترهای جامعه مثا میانگین μ و واریانس σ^2 داشته باشیم.

حال همانطور که برای یک متغیر تصادفی X مقادیری مثل میانگین، واریانس، kامین گشتاور را معرفی نمودیم برای متغیر تصادفی X نیز این مقادیر را می‌توانیم تعریف کنیم.

مقدار $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ میانگین نمونه گفته می‌شود. به همین ترتیب مقدار $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ نیز مقدار واریانس نمونه و $S=\sqrt{S^2}$ نیز مقدار انحراف از معیار نمونه نامیده می‌شود.

Kامین گشتاور نمونه عبارتست از:

$$M_K = \frac{1}{n}\sum_{i=1}^{n} X_i^k \qquad k=1,2,\ldots$$

توجه کنید که مقادیر $\overline{X}$ و S^2 یک آماره می‌باشد و در نتیجه خود یک متغیر تصادفی می‌باشد این مطلب برای مقادیر مختلف K در M_K (مثل M_1 ، M_2 ...) نیز برقرار می‌باشد.

مثال ۳: نشان دهید روابط زیر دارای S^2 برقرار است:

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

ب) $$S^2 = \frac{n}{n-1}\left(M_2 - \overline{X}^2\right)$$

حل: الف)

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

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

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

ب) بوضوح از بند الف داریم:

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

$$= \frac{n}{n-1}\left(M_2 - \overline{X}^2\right)$$

dalbablog.wordpress.com

در بخش توزیع‌های نمونه‌ای علت بکار بردن $\frac{۱}{n-۱}$ به جای $\frac{۱}{n}$ را در محاسبه $S^۲$ توضیح می‌دهیم.

مثال ۴: در مثال ۲ نتیجه، حاصل از نمونه‌گیری از قد ۱۰ نفر از دانشجویان بصورت زیر می‌باشد:

۱۷۸ , ۱۸۷ , ۱۶۱ , ۱۵۰ , ۱۷۰ , ۱۷۵ , ۱۹۲ , ۱۶۶ , ۱۷۳ , ۱۸۴

مطلوبست:

الف) میانگین و واریانس نمونه.

ب) گشتاور اول و دوم نمونه.

حل: الف)

$$\overline{X} = \frac{۱}{۱۰}\sum_{۱}^{۱۰} X_i = \frac{۱}{۱۰}(۱۷۸+۱۸۷+۱۶۱+۱۵۰+۱۷۰+۱۷۵+۱۹۲+۱۶۶+۱۷۳+۱۸۴) = ۱۷۳/۶$$

برای محاسبه واریانس نمونه از گشتاور اول و دوم استفاده می‌نیم:

$$M_۱ = M_۱ = \frac{۱}{n}\sum_{۱}^{n} X_i = \overline{X} = ۱۷۳/۶$$

$$M_۲ = \frac{۱}{n}\sum_{۱}^{n} X_i^۲ = \frac{۱}{۱۰}\sum_{۱}^{۱۰} (X_i)^۲ = \frac{۱}{۱۰}(۱۷۸^۲ + ۱۸۷^۲ + ۱۶۱^۲ + ۱۵۰^۲ + ۱۷۰^۲$$

$$+ ۱۷۵^۲ + ۱۹۲^۲ + ۱۶۶^۲ + ۱۷۳^۲ + ۱۸۴^۲ = ۳۰۲۸۰/۴$$

بنابراین $S^۲$ برابراست با:

$$S^۲ = \frac{n}{n-۱}(M_۲ - \overline{X}^۲) = \frac{۱۰}{۹}(۳۰۲۸۰/۴ - ۱۷۳/۶^۲)$$

$$= \frac{۱۰}{۹}(۱۴۳/۴۴) = ۱۵۹/۳۷$$

انحراف از معیار نمونه: $\Rightarrow S = \sqrt{۱۵۹/۳۷} = ۱۲/۶۲$

۸ . ۳ آماره مرتب، میانه و برد نمونه

نمونه تصادفی $X_۱, X_۲, \ldots, X_n$ از متغیر تصادفی X را در نظر بگیرید اگر مقادیر مشاهده شده نمونه تصادفی را از کوچک به بزرگ مرتب کنیم بطوریکه $X_{(۱)} \le X_{(۲)} \le \ldots \le X_{(n)}$ را داشته باشیم در این صورت $X_{(i)}$ را iامین آماره مرتب می‌نامیم. $(i = ۱, ۲, \ldots, n)$ همچنین به $X_{(۱)}$ حداقل مقدار نمونه و به $X_{(n)}$ حداکثر مقدار نمونه گفته می‌شود. میانه نمونه عبارتست از مقدار وسط آماره مرتب به عبارت دقیقتر $m_\circ$ را میانه نمونه در نظر می‌گیریم که بصورت زیر بدست می‌آید:

اگر n فرد باشد $m_\circ = X_{(\frac{n+۱}{۲})}$.

اگر n زوج باشد $= \frac{X(\frac{n}{۲}) + X(\frac{n}{۲}+۱)}{۲}$.

برد نمونه نیز عبارتست از تفاضل بزرگترین آماره مرتب از کوچکترین آماره مرتب که با $R_\circ$ نمایش داده می‌شود. $R_\circ = X_{(n)} - X_{(۱)}$

مثال ۵: نمونه تصادفی مشاهده شده در مثال ۴ را در نظر بگیرید، آماره مرتب، حداقل مقدار نمونه، حداکثر مقدار نمونه، میانه و برد نمونه را بدست بیاورید؟

حل: با مرتب نمودن نمونه‌های مشاهده شده داریم:

dalbablog.wordpress.com

$X_{(۱)}=۱۵۰$, $X_{(۲)}=۱۶۱$, $X_{(۳)}=۱۶۶$, $X_{(۴)}=۱۷۰$, $X_{(۵)}=۱۷۳$
$X_{(۶)}=۱۷۵$, $X_{(۷)}=۱۷۸$, $X_{(۸)}=۱۸۴$, $X_{(۹)}=۱۸۷$, $X_{(۱۰)}=۱۹۲$

$X_{(۱)}=۱۵۰$ حداقل مقدار نمونه.

$X_{(۱۰)}=۱۹۲$ حداکثر مقدار نمونه.

میانه نمونه : $\frac{X_{(\frac{۱۰}{۲})}+X_{(\frac{۱۰}{۲}+۱)}}{۲}=\frac{X_{(۵)}+X_{(۶)}}{۲}=\frac{۱۷۳+۱۷۵}{۲}=۱۷۴$

برد نمونه : $X_{(۱۰)}-X_{(۱)}=۱۹۲-۱۵۰=۴۲$

۸-۱۱

۸ . ۴ تابع توزیع نمونه‌ای

برای نمونه تصادفی $X_۱,X_۲,\dots,X_n$ و آماره مرتب $X_{(۱)}\le X_{(۲)}\le\dots\le X_{(n)}$ تابع توزیع نمونه‌ای بصورت زیر تعریف می‌شود:

$$F^*(t)=\begin{cases}۰ & t<X_{(۱)} \quad \text{برای}\\ \frac{r}{n} & X_{(r)}\le t<X_{(r+۱)} \quad (r=۱,\ ,\dots,n)\\ ۱ & t>X_{(n)}\end{cases}$$

۸-۱۲

مثال ۶: اگر نمونه‌های مشاهده شد از قد دانشجویان در مثال ۴ بصورت زیر باشد مطلوبست تابع توزیع نمونه‌ای برای نمونه تصادفی مشاهده شده؟

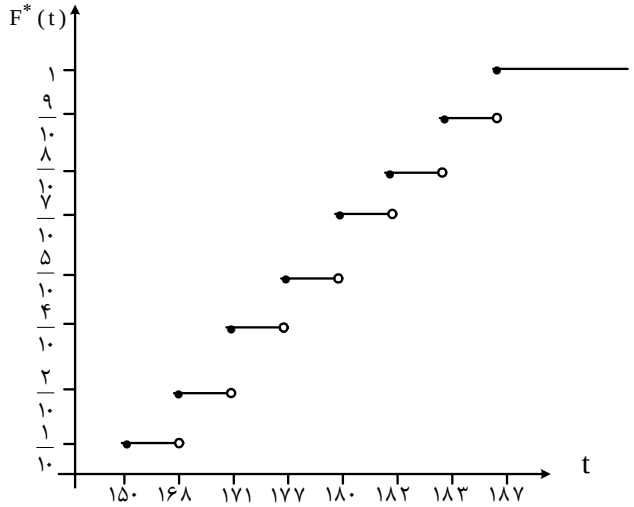
۱۵۰ , ۱۶۸ , ۱۷۱ , ۱۷۱ , ۱۷۷ , ۱۸۰ , ۱۸۰ , ۱۸۲ , ۱۸۳ , ۱۸۷

حل:

$$F^*(t)=\begin{cases}۰ & t<۱۵۰\\ \frac{۱}{۱۰} & ۱۵۰\le t<۱۶۸\\ \frac{۲}{۱۰} & ۱۶۸\le t<۱۷۱\\ \frac{۴}{۱۰} & ۱۷۱\le t<۱۷۷\\ \frac{۵}{۱۰} & ۱۷۷\le t<۱۸۰\\ \frac{۷}{۱۰} & ۱۸۰\le t<۱۸۲\\ \frac{۸}{۱۰} & ۱۸۲\le t<۱۸۳\\ \frac{۹}{۱۰} & ۱۸۳\le t<۱۸۷\\ ۱ & ۱۸۷\le t\end{cases}$$

dalbablog.wordpress.com

نمودار تابع توزیع نمونه‌ای فوق نیز بصورت زیر می‌باشد.



÷

۸-۱۳

۸ . ۵ نامساوی چبی چف

فرض کنید X یک متغیر تصادفی باشد که نحوه توزیع و تابع چگالی آنرا نمی‌دانیم در این صورت برای اظهار نظر در مورد مقادیر احتمال که متغیر تصادفی X قبول می‌کند می‌توانیم از قضیه چبی چف استفاده کنیم. این قضیه بصورت زیر می‌باشد:

اگر X یک متغیر تصادفی باشد و $g(X)$ یک تابع حقیقی غیر منفی از X باشد در این صورت:

$$p[g(X) \geq K] \leq \frac{1}{K} E[g(X)] \qquad g(X) \geq 0 \ , \ K>0$$

با انتخاب $g(X) = (X-\mu_X)^2$ و $k^2\sigma_X^2$ بجای k داریم:

$$p(X-\mu_X)^2 \geq k^2\sigma_X^2) \leq \frac{1}{k^2\sigma_X^2} E\left[(X-\mu_X)^2\right]$$

$$p(|X-\mu_X| \geq K\sigma_X) \leq \frac{1}{k^2}$$

بنابراین:

$$p(|X-\mu_X| < K\sigma_X) \geq 1-\frac{1}{k^2}$$

مقادیر احتمال که توسط قانون چبی چف بدست می‌آیند بسته به مقدار k از دقتهای مختلفی برخوردارند بطوریکه هر چه k بیشتر باشد دقت محاسبه احتمال توسط قانون چبی چف بیشتر است. این مطلب را در مثال زیر نشان می‌دهیم:

مثال ۷: اگر X یک متغیر تصادفی نرمال باشد در این صورت مطلوبست محاسبه احتمال $p(|X-\mu_X| < k\sigma_X)$ با استفاده از توزیع نرمال و قانون چبی چف به ازای مقادیر $k=1, 2, 3, 4$.

حل: ابتدا مقدار احتمال را برای کتغیر تصادفی نرمال X بدست می‌آوریم:

$$p(|X-\mu_X| < K\sigma_X) = p(-k < \frac{X-\mu_X}{\sigma_X} < k)$$

$$= N_Z(k) - N_Z(-k) = 2N_Z(k)-1$$

حال در جدول زیر مقادیر مختلف k را با استفاده از رابطه فوق و قانون چبی چف محاسبه می‌کنیم:

dalbablog.wordpress.com

K	$۲N_Z(k)-۱$	$۱-\frac{۱}{k^۲}$
۱	۰/۶۸۲۶	۰
۲	۰/۹۵۴۶	۰/۷۵
۳	۰/۹۹۷۴	۰/۸۸۸۹
۴	۱/۰۰۰	۰/۹۳۳۴

همانطور که ملاحظه می‌کنید مقادیر بدست آمده برای $k=۱,۲$ از خطای بالایی برخوردار هستند اما با افزایش k مقدار بدست آمده از قانون چبی چف به مقدار واقعی نزدیک‌تر می‌شود. می‌توان تابع چگالی احتمال را طوری تعریف نمود که به ازای هر مقدار $k \geq ۱$ مقدار واقعی احتمال با احتمال بدست آمده از رابطه چبی چف برابر باشد. این حالت را در مثال بعد نشان می‌دهیم:

مثال ۸: فرض کنید تابع چگالی احتمال متغیر تصادفی گسسته X بصورت زیر باشد:

$$f_X(x)=\frac{k^۲-۱}{k^۲} \qquad x=۰ \text{ براي } \qquad k \geq ۱$$

$$=\frac{۱}{۲k^۲} \qquad x=-k \text{ , } k \text{ برای}$$

$$=۰ \qquad \text{سایر مقادیر}$$

مطلوبست: الف) محاسبه μ_X , σ_X.

ب) محاسبه احتمال $p(|X-\mu_X|<k\sigma_X)$ به وسیله تابع احتمال و بوسیله قانون چبی چف و مقایسه آندو.

حل: الف)

$$\mu_X=\sum_{i=۱}^{n} f(xi)\,x_i = ۰\times\frac{k^۲-۱}{k^۲}+k(\frac{۱}{۲k^۲})+(-k)(\frac{۱}{۲k^۲})=۰+\frac{۱}{۲k}-\frac{۱}{۲k}=۰$$

$$E[X^۲]=\sum_{۱}^{n} x_i^۲\,f(x_i) = ۰\times\frac{k^۲-۱}{k^۲}+k^۲(\frac{۱}{۲k^۲})+(-k)^۲(\frac{۱}{۲k^۲})$$

$$=۰+\frac{۱}{۲}+\frac{۱}{۲}=۱ \quad \sigma_x^۲=E[X^۲]-E^۲[X]=\sigma_x^۲=۱-۰^۲=۱$$

ب) ابتدا احتمال دقیق را محاسبه می‌کنیم:

$$\mu_X=۰ \qquad \sigma_X=۱$$

$$\Rightarrow p(|X-\mu_X|<k\,\sigma_X)=p(|X|<K)$$

$$p(-k<X<k)=p(X=-k)+p(X=k)=\frac{۱}{۲k^۲}+\frac{۱}{۲k^۲}=\frac{۱}{k^۲}$$

اما طبق قانون چبی چف رابطه زیر برقرار می‌باشد:

$$p(|X-\mu_X|<k\,\sigma_X) \leq \frac{۱}{k^۲}$$

مقدار $\frac{۱}{k^۲}$ دقیقاً برابر با احتمال محاسبه شده می‌باشد بنابراین در این حالت مخصوص مقدار دقیق احتمال با مقدار بدست آمده از قانون چبی چف برابر می‌باشد و به عبارتی حالت مساوی در قانون چبی چف رخ می‌دهد.

۸ . ۶ قضیه حد مرکزی

dalbablog.wordpress.com

همانطور که قبلاض اشاره کردیم تابع نرمال نقش مهمی در احتمالات بازی می‌کند زیرا اولاً محاسبه احتمال نرمال با استفاده از تابع نرمال استاندارد و جدول مربوط به آن بسیار ساده می‌باشد ثانیاً بسیاری از محاسبات احتمال را همانطور که در فصلهای قبل نشان دادیم تحت شرایط خاصی می‌توان با استفاده از تابع نرمال تقریب زد. حال در اینجا نشان می‌دهیم که اگر متغیرهای تصادفی X_i $(i = ۱,۲,\dots)$ از طکدیگر مستقل باشند و از یک تابع توزیع پیروی کنند، مجموع آنها به سمت تابع نرمال میل می‌کند. در اینجا نمی‌دانیم که متغیرهای تصادفی X_i از چه توزیعی پیروی می‌کند. صورت قضیه حد مرکزی بصورت زطر می‌باشد:

فرض کنید $X_۱, X_۲, \dots, X_n$ نمونه‌ای تصادفی از جامعه نامتناهی X با میانگین μ و $\sigma^۲$ باشند در این صورت دنباله متغیرهای تصادفی $Z_۱, Z_۲, \dots, Z_n$ را به صورت زیر تعریف می‌کنیم:

$$Z_n = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad n = ۱,۲, \dots$$

که در آن $\overline{X}_n = \frac{۱}{n}\sum_{i=۱}^{n} X_i$ در این صورت اگر $F_{Z_n}(t)$ تابع توزیع متغیرهای تصادفی Z_n به ازای مقدار حقیقی t باشد داریم:

$$\lim_{n\to\infty} F_{Z_n}(t) = N_Z(t)$$

که در آن $N_Z(t)$ تابع توزیع نرمال استاندارد می‌باشد.

به عبارت ساده‌تر متغیر تصادفی $Z = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$ وقتی n به سمت بی‌نهایت میل می‌کند دارای توزیع نرمال استاندارد $N(۰, ۱)$ می‌باشد. یعنی برای nهای به اندازه کافی بزرگ داریم $F_{Z_n}(t) = N_Z(t)$ اما برای هر مقدار n داریم:

$$F_{\overline{X}_n}(t) = F_{Z_n}\left(\frac{t-\mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

در نتیجه برای nهای بزرگ:

$$F_{\overline{X}_n}(t) = N_Z\left(\frac{t-\mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

رابطه فوق به این معنی است که تابع توزیع میانگین تعداد زیادی از متغیرهای تصادفی مستقل با توزیع یکسان تقریباً مساوی است با توزیع نرمال استاندارد. به همین دلیل است که تعداد زیادی از قوانین احتمال را بوسیله نرمال تقریب می‌زنند.

۸-۱۷

به عنوان مثال به تقریب دو جمله‌ای به نرمال توجه کنید که در فصل قبل روابط آنرا بصورت زیر نشان دادیم:

$$p(X = x_۰) = N\left(\frac{x_۰ + \frac{۱}{۲} - np}{\sqrt{pq}}\right) - N\left(\frac{x_۰ - \frac{۱}{۲} - np}{\sqrt{pq}}\right)$$

می‌دانیم اگر متغیرهای تصادفی برنولی و مستقل $X_۱, X_۲, \dots, X_n$ را هر یک با پارامتر p در نظر بگیریم در این صورت $Y = \sum_{۱}^{n} X_i$ یک متغیر تصادفی دو جمله‌ای با پارامتر n و p می‌باشد حال طبق قضیه حد مرکزی میانگین X_iها به توزیع نرمال میل می‌کند یعنی تابع توزیع $\frac{Y}{n} = \frac{۱}{n}\sum_{۱}^{n} X_i = \overline{X}_n$ وقتی n بزرگ باشد برابر است باک

dalbablog.wordpress.com

$$F_{\frac{Y}{n}}(t) = N_Z(\frac{t-\mu}{\frac{\sigma}{\sqrt{n}}})$$

که در آن امید ریاضی $\frac{Y}{n}$ برابر است با $\mu = p$ و واریانس $\frac{Y}{n}$ نیز برابر با $\sigma^2 = pq$ می‌باشد حال داریم:

$$F_{\frac{Y}{n}}(t) = N_Z(\frac{t-\mu}{\frac{\sqrt{pq}}{\sqrt{n}}}) = N_Z(\frac{t-p}{\sqrt{\frac{pq}{n}}})$$

حال $F_{\frac{Y}{n}}(t)$ را به $F_Y(t)$ تبدیل می‌کنیم:

$$F_Y(t) = F_{\frac{Y}{n}}(\frac{t}{n}) = N_Z(\frac{\frac{t}{n}-p}{\sqrt{\frac{pq}{n}}}) = N_Z(\frac{t-np}{\sqrt{npq}})$$

در نتیجه $F_Y(t) = N_Z(\frac{t-np}{\sqrt{npq}})$ که با استفاده از روش ذوذنقه‌ای که در فصل قبل نشان دادیم می‌توان مقدار احتمال $p(X = x_\circ)$ را بصورت زیر بدست آورد:

$$p(X = x_\circ) = N(\frac{x_\circ + \frac{1}{2} - np}{\sqrt{npq}}) - N(\frac{x_\circ - \frac{1}{2} - np}{\sqrt{npq}})$$

dalbablog.wordpress.com

فصل نهم

توزیع‌های برخی از آماره‌ها

در فصل قبل با آماره‌ها و تعاریف و مقدمات آنها آشنا شدید، نشان دادیم که آماره‌ها خود متغیر می‌باشند و می‌توان تابع توزیع آنها را معنی نمود، حال در این فصل چندین آماره خاص را مورد بررسی قرار می‌دهیم و میانگین و واریانس آنها را محاسبه می‌کنیم.

۱.۹ توزیع نمونه‌ای میانگین نمونه $\overline{X}$

از یک جامعه با میانگین μ و واریانس δ^2 تعداد n نمونه $X_1, X_2, \ldots, X_n$ انتخاب می‌کنیم، آماره $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ را در نظر بگیرید

می‌خواهیم میانگین و واریانس $\overline{X}$ را به عنوان یک متغیر تصادفی بدست بیاوریم. طبق قضیه حد مرکزی اگر مقدار n به اندازه کافی بزرگ اختیلر شود متغیر تصادفی $Z = \frac{\overline{X}-\mu}{\frac{\delta}{\sqrt{n}}}$ یک متغیر تصادفی نرمال استاندارد می‌باشد. از طرفی داریم:

$$\overline{X} = \frac{\delta}{\sqrt{n}} Z + \mu \quad , \quad E[Z] = 0 \quad , \quad \delta_Z^2 = 1$$

بنابراین:

$$E[\overline{X}] = \frac{\delta}{\sqrt{n}} E[Z] + \mu = \frac{\delta}{\sqrt{n}} \times 0 + \mu = \mu$$

بنابراین آماره $\overline{X}$ به عنوان یک متغیر تصادفی اگر تعداد نمونه‌ها زیاد باشد $(n > 30)$ همواره دارای میانگین μ یا همان میانگین جامعه می‌باشد.

۹-۳

مثال ۱: در نقطه‌ای از یک رودخانه در هر دقیقه بطور متوسط تعداد ۲۰ ماهی عبور می‌کند بدیهی است که ماهیگیر تنها در صورتی می‌تواند در هر دقیقه این تعداد ماهی را صید کند که تقریباً تمامی طول رودخانه را با تور ماهیگیری پوشش دهد یا به عبارتی در هر دقیقه از بیشتر نقاط در طول رودخانه نمونه برداری کند که معادل است با انتخاب تعداد زیادی نمونه.

حال واریانس $\overline{X}$ را محاسبه می‌کنیم:

$$\text{var}(\overline{X}) = \text{var}\left(\frac{\delta}{\sqrt{n}} Z + \mu\right) = \text{var}\left(\frac{\delta}{\sqrt{n}} Z\right) = \frac{\delta^2}{\sqrt{n}} \text{var}(Z) = \frac{\delta^2}{\sqrt{n}}$$

بنابراین واریانس $\overline{X}$ برابر با $\frac{\delta^2}{\sqrt{n}}$ می‌باشد.

۹-۴

مثال ۲: محصول شیر تولیدی یک کارخانه بطور متوسط ۲۰ روز پس از تاریخ تولید قابل مصرف می‌باشد که با واریانس ۴ روز رخ می‌دهد. می‌خواهیم از این کارخانه ۱۰ پاکت شیر تهیه کنیم. احتمال اینکه این پاکتهای شیر بطور متوسط حداقل ۱۸ روز قابل مصرف باشند چقدر است؟

حل: برای به دست آوردن احتمال می‌بایستی $p(\overline{X} \geq 18)$ را محاسبه کنیم که در واقع از تابع توزیع متغیر تصادفی $\overline{X}$ که در اینجا نرمال $\mu \quad \frac{\delta^2}{\sqrt{n}}$ در نظر گرفته می‌شود استفاده می‌کنیم. بنابراین:

$$\mu_{\overline{X}} = 20 \quad , \quad n = 10$$

$$\delta_{\overline{X}}^2 = \frac{\delta^2}{n} = \frac{(4)^2}{10} = \frac{16}{10} \quad \rightarrow \quad \frac{\delta}{\overline{X}} = \sqrt{\frac{16}{10}} = 1/26$$

dalbablog.wordpress.com

$$p(\overline{X} \geq ۱۸) = p\left(\frac{\overline{X}-\mu_{\overline{X}}}{\delta_{\overline{X}}} \geq \frac{۱۸-\mu_{\overline{X}}}{\delta_{\overline{X}}}\right)$$

$$= p\left(\frac{\overline{X}-\mu_{\overline{X}}}{\delta_{\overline{X}}} \geq \frac{۱۸-۲۰}{۱/۲۶}\right) = p\left(\frac{\overline{X}-\mu_{\overline{X}}}{\delta_{\overline{X}}} \geq ۱/۵۸\right)$$

$$p = (Z \geq -۱/۵۸) = ۱ - p(Z \leq -۱/۵۸)$$

$$= ۱ - N_Z(-۱/۵۸) = ۱ - ۰/۰۵۷۱ = ۰/۹۴۲۹$$

۹-۵-۱

۹-۲ توزیع نمونه‌ای واریانس نمونه $S^۲$.

حال به محاسبه توزیع نمونه‌ای واریانس $S^۲$ و بدست آوردن میانگین و واریانس توزیع آن می‌پردازیم.

همانطور که از فصل‌های قبل به یاد دارید واریانس نمونه‌ها با $S^۲$ نشان داده می‌شود که بصورت زیر تعریف می‌شود:

$$S^۲ = \frac{۱}{n-۱} \sum_{i=۱}^{n} (X_i - \overline{X})^۲$$

قبل محاسبه توزیع نمونه‌ای $S^۲$ ابتدا میانگین و واریانس آنرا (از آنجا که محاسبه آن ساده می‌باشد) بدست می‌آوریم:

می‌دانیم واریانس $\overline{X}$ برابر $\frac{\delta^۲}{n}$ می‌باشد بنابراین:

$$var(\overline{X}) = \frac{\delta^۲}{n} \rightarrow E\left[(\overline{X}-\mu)^۲\right] = \frac{\delta^۲}{n}$$

$$E\left[\overline{X}^۲\right] - E^۲[X] = E\left[\overline{X}^۲\right] - \mu^۲ = \frac{\delta^۲}{n} \quad \Rightarrow \quad E\left[\overline{X}^۲\right] = \frac{\delta^۲}{n} + \mu^۲$$

حال داریم:

$$E\left[S^۲\right] = E\left[\frac{۱}{n-۱} \sum_{i=۱}^{n} (X_i - \overline{X})^۲\right] = \frac{۱}{n-۱} E\left[\sum_{i=۱}^{n} (X_i - \overline{X})^۲\right]$$

$$= \frac{۱}{n-۱} E\left[\sum_{۱}^{n} X_i^۲ - ۲\overline{X} \sum_{۱}^{n} X_i + \sum_{۱}^{n} \overline{X}^۲\right]$$

$$= \frac{n}{n-۱} E\left[\frac{۱}{n} \sum_{۱}^{n} X_i^۲ - ۲\overline{X} \frac{۱}{n} \sum_{۱}^{n} X_i + \frac{۱}{n} \sum_{۱}^{n} \overline{X}^۲\right]$$

$$= \frac{n}{n-۱} E\left[\frac{۱}{n} \sum_{۱}^{n} X_i^۲ - ۲\overline{X}^۲ + \overline{X}^۲\right]$$

$$= \frac{n}{n-۱} E\left[\frac{۱}{n} \sum_{۱}^{n} X_i^۲ - \overline{X}^۲\right] = \frac{n}{n-۱} \left(E\left[\frac{۱}{n} \sum_{۱}^{n} X_i^۲\right] - E\left[\overline{X}^۲\right]\right)$$

$$= \frac{n}{n-۱} \left(E[M_۲] - \left(\frac{\delta^۲}{n} + \mu\right)\right)$$

dalbablog.wordpress.com

۹-۵-۲

که در آن $M_۲$

گشتاور دوم نمونه‌ها می‌باشد و داریم: $E[M_۲]=m_۲$ که $m_۲$ نیز گشتاور دوم متغیر تصادفی X می‌باشد. در نهایت داریم:

$$= \frac{۱}{n-۱}\left(m_۲ - \mu^۲ - \frac{\delta^۲}{n}\right)$$

اما: $m_۲ = \delta^۲ + \mu^۲$

$$= \frac{n}{n-۱}\left(\delta^۲ + \mu^۲ - \mu^۲ - \frac{\delta^۲}{n}\right) = \frac{n}{n-۱}\left(\delta^۲ - \frac{\delta^۲}{n}\right) = \delta^۲$$

به نتیجه جالبی رسیدیم و آن اینکه میانگین واریانس نمونه‌ها برابر با واریانس جامعه می‌باشد در واقع علت اینکه در محاسبه $S^۲$ از $\frac{۱}{n-۱}$ استفاده می‌کنیم این است که در محاسبات فوق در نهایت جواب ساده شده و $\delta^۲$ به عنوان میانگین $S^۲$ $\left(E[S^۲]=\delta^۲\right)$ بدست می‌آید . در حالی که اگر $S^۲$ را بصورت زیر تعریف می‌کردیم $S^۲ = \frac{۱}{n}\sum_{i=۱}^{n}(X_i - \overline{X})^۲$ جواب نهایی $E[S^۲]$ برابر با $\left(\frac{n}{n-۱}\right)\sigma^۲$ می‌شد که بلیل وجود n در جواب نتیجه مطلوبی نیست. حال به جای بدست آوردن توزیع $S^۲$ توزیع $\left(\frac{n-۱}{\delta^۲}\right)S^۲$ را بدست می‌آوریم زیرا در محاسبه مقدار میانگین بصورت زیر ساده‌تر می‌شود.

$$E\left[\frac{n-۱}{\delta^۲} - S^۲\right] = \frac{n-۱}{\delta^۲}E[S^۲] = \frac{n-۱}{\delta^۲}\delta^۲ = n-۱$$

برای بدست آوردن توزیع $\left(\frac{n-۱}{\delta^۲}\right)S^۲$ ابتدا توزیع خی-دو یا مربع‌کای را معرفی می‌کنیم:

۹-۶

۹.۲.۱ توزیع خی دو

فرض کنید Z یک متغیر تصادفی نرمال استاندارد باشد $Z \sim N(۰, ۱)$ در این صورت متغیر تصادفی $W = Z^۲$ را یک متغیر تصادفی $\chi^۲$ (خی دو) با یک درجه آزادی می‌نامیم و به صورت $\chi^۲_۱$ نمایش می‌دهیم که تابع توزیع آن بصورت زیر بدست می‌آید:

قبلاً نشان دادیم که اگر $Y = X^۲$ باشد تابع توزیع $F_Y(t)$ برابر است با:

$$F_Y(t) = F_X(\sqrt{t}) - F_X(-\sqrt{t}) \qquad t \geq ۰$$
$$= ۰ \qquad t < ۰$$

حال داریم: $W = Z^۲$

$$F_W(t) = F_Z(\sqrt{t}) - F_Z(-\sqrt{t})$$

اما: $F_Z(\sqrt{t}) = N_Z(t)$ و $N_Z(-a) = ۱ - N_Z(a)$ بنابراین:

$$F_W(t) = N_Z(\sqrt{t}) - \left[۱ - N_Z(\sqrt{t})\right] = ۲N_Z(\sqrt{t}) - ۱$$

بنابراین با داشتن جدول مقادیر توزیع نرمال استاندارد به راحتی می‌توان مقادیر توزیع $\chi^۲_۱$ را به دست آورد. مقدار دقیق تابع چگالی $\chi^۲_۱$ بصورت زیر می‌باشد:

dalbablog.wordpress.com

$$f_X(x) = \frac{1}{\sqrt{2\pi}}\, x^{-\frac{1}{2}}\, e^{-\frac{x}{2}} \qquad x > 0$$

میانگین، واریانس و تابع مولد گشتاور آن عبارتند از:

$$E[W] = 1 \qquad var(W) = 2 \qquad m_W(t) = \left(\frac{1}{1-2t}\right)^{\frac{1}{2}}$$

۹-۷

حال فرض کنید $X_1, X_2, \ldots, X_n$ نمونه‌هایی تصادفی از یک متغیر تصادفی نرمال X با میانگین μ و واریانس δ^2 باشند در این صورت:

$$Y = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\delta^2}$$

یک متغیر تصادفی χ^2 با n درجه آزادی نامیده می‌شود و با χ^2_n نمایش داده می‌شود. به بیانی دیگر اگر $Z_1, Z_2, \ldots, Z_n$ تعداد n متغیر تصادفی نرمال استاندارد $N(0,1)$ باشند در این صورت $Y = Z_1^2, Z_2^2, \ldots, Z_n^2$ یک متغیر تصادفی χ^2 با n درجه آزادی است. تابع چگالی متغیر تصادفی χ^2_n بصورت زیر می‌باشد:

$$f_Y(y) = \frac{1}{\Gamma\left(\frac{n}{2}\right)} \left(\frac{1}{2}\right)^{\frac{n}{2}} y^{\frac{n}{2}-1}\, e^{-\frac{y}{2}} \qquad y > 0$$

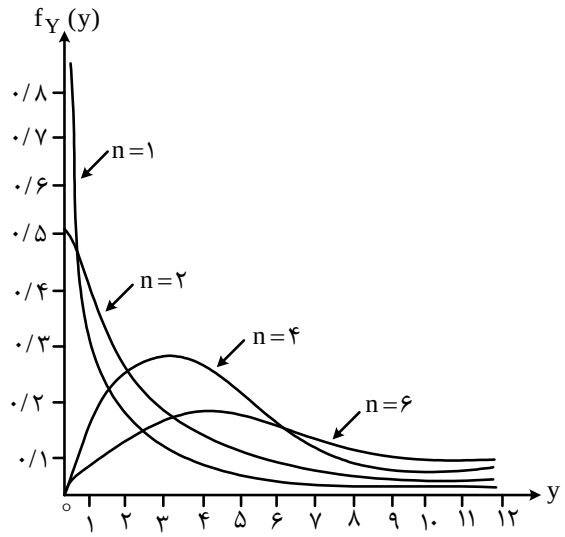
توجه کنید که متغیر تصادفی χ^2_n حالت خاصی از متغیر تصادفی گاما می‌باشد که در آن $k = \frac{n}{2}$ و $\lambda = \frac{1}{2}$ می‌باشد.

مقادیر میانگین، واریانس و تابع مولد گشتاور متغیر تصادفی χ^2_n بصورت زیر می‌باشد:

$$E[Y] = n \quad , \quad var(Y) = 2n$$

$$m_Y(t) = \left(\frac{1}{1-2t}\right)^{\frac{n}{2}}$$

نمودار منحنی متغیر تصادفی χ^2_n دارای نقطه ماکزیممی در $y = n-2$ می‌باشد و با افزایش n نمودار منحنی هر چه بیشتر به نمودار نزدیک نزدیکتر می‌شود این مطلب در شکل زیر نشان داده شده است:



dalbablog.wordpress.com

نکته: اگر $X_1, X_2, \ldots, X_n$ متغیرهای تصادفی خی دو به ترتیب با $r_1, r_2, \ldots, r_n$ درجه آزادی باشند و همچنین دو بدو مستقل از یکدیگر باشند در این صورت متغیر تصادفی $Y = \sum_{1}^{n} X_i$ دارای توزیع خی دو با $R = \sum_{i=1}^{n} r_i$ درجه آزادی می‌باشد.

۹-۸

مثال ۳: فرض کنید X_1, X_2, X_3 سه متغیر تصادفی مستقل خی دو به ترتیب با ۷ و ۳ و ۱ درجه آزادی باشند، اگر $W = X_1 + X_2 + X_3$ مطلوبست میانگین و واریانس متغیرهای تصادفی W؟

حل: متغیر تصادفی W یک متغیر تصادفی خی دو با $۱۱ = ۷+۳+۱$ درجه آزادی می‌باشد بنابراین میانگین و واریانس آن عبارتست از:

$$E[W] = ۱۱$$

$$\text{var}(W) = ۲ \times ۱۱ = ۲۲$$

حال به قضیه بسیار مهم توجه کنید:

قضیه: اگر $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از متغیر تصادفی نرمال X با میانگین μ و واریانس δ^2 باشند آنگاه:

الف) $\overline{X}$ و S^2 از یکدیگر مستقل هستند.

ب) $\frac{(n-1)}{\delta^2} S^2$ یک متغیر تصادفی χ^2 با $n-1$ درجه آزادی می‌باشد توجه کنید که $(\frac{n-1}{\delta^2}) S^2$ برابر است با:

$$(\frac{n-1}{\delta^2}) S^2 = \frac{n-1}{\delta^2} \frac{1}{n-1} \sum_{1}^{n} (X_i - \overline{X})^2 = \sum_{1}^{N} \frac{(X_i - \overline{X})^2}{\delta^2}$$

و این با آنچه در تعریف متغیر تصادفی χ^2_n متفاوت می‌باشد زیرا در تعریف از میانگین جامعه μ استفاده نمودیم: $\sum_{1}^{n} \frac{(X_i - \mu)^2}{\delta^2}$.

۹-۹

حال نشان می‌دهیم که چرا با جایگزین نمودن $\overline{X}$ به جای μ یک درجه آزادی از n کم می‌شود و درجه آزادی $\frac{(n-1)}{\delta^2} S^2$ برابر با $n-1$ می‌شود.

$$\sum_{i=1}^{n} (X_i - \mu)^2 = \sum_{1}^{n} \left[(X_i - \overline{X}) + (\overline{X} - \mu) \right]^2$$

$$= \sum_{1}^{n} (X_i - \overline{X})^2 + \sum_{1}^{n} (\overline{X} - \mu)^2 + ۲ (\overline{X} - \mu) \sum_{1}^{n} (X_i - \overline{X})$$

$$= \sum_{1}^{n} (X_i - \overline{X})^2 + n (\overline{X} - \mu)^2$$

در نتیجه بدست می‌آید:

$$\frac{\sum (X_i - \mu)^2}{\delta^2} = \frac{\sum (X_i - \mu)^2}{\delta^2} + \frac{(\overline{X} - \mu)^2}{\frac{\delta^2}{n}}$$

dalbablog.wordpress.com

حال می‌دانیم که عبارت سمت چپ تساوی فوق یک متغیر تصادفی خی دو با n درجه آزادی است و عبارت $\frac{(\overline{X}-\mu)^2}{\frac{\delta^2}{n}}$ نیز یک متغیر تصادفی X^2 با ۱ درجه آزادی است بنابراین جمله $\frac{\sum (X_i-\mu)^2}{\delta^2}$ می‌بایستی یک متغیر تصادفی χ^2 با $n-1$ درجه آزادی باشد.

مثال ۴: از یک جامعه نرمال با میانگین μ و واریانس δ^2 یک نمونه تصادفی ۱۱تایی انتخاب می‌کنیم. متغیر تصادفی $Y = \frac{10\, S^2}{\delta^2}$ را در نظر بگیرید مطلوبست:

الف) محاسبه $p(3/94 < Y < 18/3)$.

ب) با استفاده از احتمال فوق رابطه بین S^2 و δ^2 را در جامعه معین کنید.

حل: الف) ابتدا احتمال را بصورت زیر ساده می‌کنیم:

$$p(3/94 < Y < 18/3) = p(Y < 18/3) - p(Y < 3/94)$$

حال از روی جدول مقادیر احتمال برای متغیر تصادفی خی دو و با توجه به اینکه $Y = \frac{10\, S^2}{\delta^2}$ یک متغیر تصادفی خی دو با ۱۰ درجه آزادی است مقادیر احتمال بصورت زیر بدست می‌آیند:

$$p(Y < 18/3) - p(Y < 3/94) = 0/95 / 0/05 = 0/9$$

ب) می‌دانیم $p(3/94 < Y < 18/3) = 0/9$ بنابراین:

$$p(3/94 < Y = \frac{10\, S^2}{\delta^2} < 18/3) = p(\frac{3/94\, \delta^2}{10} < S^2 < \frac{18/3\, \delta^2}{10})$$

$$\Rightarrow p(0/394\ \delta^2 < S^2 < 1/83\ \delta^2) = 0/9$$

حال می‌بینیم که S^2 در بازه‌ای نزدیک δ^2 قرار دارد یعنی اگر از یک جامعه نرمال با میانگین μ و واریانس δ^2 تعداد ۱۱ نمونه انتخاب کنیم به احتمال بسیار زیاد $(0/9)$ مقدار واریانس نمونه‌ها نزدیک به واریانس جامعه می‌باشد.

۹-۱۱-۱

۹ . ۳ تابع توزیع آمارۀ مرتب

قبلاً آمارۀ مرتب را معرفی نمودیم، می‌دانیم آماره‌های مرتب خود یک متغیر تصادفی می‌باشند. بنابراین می‌توانیم تابع توزیع آنها را بدست بیاوریم.

فرض کنید $X_1, X_2, \ldots, X_n$ تعداد n نمونه تصادفی از متغیر تصادفی X باشند در این صورت تابع توزیع برای آمارۀ مرتب $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ عبارتست از:

$$F_{X(j)}(t) = \sum_{k=j}^{n} \binom{n}{k} \left(F_X(t)\right)^k \left(1-F_X(t)\right)^{n-k} \qquad t \in R$$

$F_X(t)$ تابع توزیع متغیر تصادفی X می‌باشد.

$F_{X(j)}(t)$ به صورت زیر بدست می‌آید:

اگر حداقل j متغیر تصادفی نمونه، کمتر یا مساوی با t باشند در این صورت $X_{(j)}$ یا jامین آمارۀ مرتب کمتر یا مساوی با t می‌باشد. نمام متغیرهای تصادفی نمونه مستقل می‌باشند و برای تمام متغیرهای تصادفی نمونه $F_X(t)$ برابر است با:

dalbablog.wordpress.com

احتمال کمتر یا مساوی با t بودن. بنابراین تعداد متغیرهای تصادفی نمونه که کمتر یا مساوی با t می‌باشند یک متغیر تصادفی دو جمله‌ای با پارامترهای n و $p = F_X(t)$ می‌باشد، در نتیجه:

$$f_{X(j)}(t) = \sum_{k=j}^{n} \binom{n}{k} \left(F_X(t)\right)^{j-1} \left(1-F_X(t)\right)^{n-j} f_X(t)$$

با قرار دادن $j = 1$ و $j = n$ تابع توزیع و چگالی برای $\min(X_j)$ و $\text{Max}(X_j)$ بصورت زیر بدست می‌آید:

$$\text{Min}\underset{j}{(X)} \text{ تابع توزیع} : F_{X(1)}(t) = \sum_{k=1}^{n} \binom{n}{k} \left(F_X(t)\right)^{k} \left(1-F_X(t)\right)^{n-k} = 1-\left(1-F_{X(1)}(t)\right)^{n}$$

$$\min\underset{j}{(X)} \text{ تابع چگالی} : f_{X(1)}(t) = \binom{n}{1} \left(f_X(t)\right)^{1-1} \left(1-F_X(t)\right)^{n-1} f_X(t) = n\left(1-F_X(t)\right)^{n-1} f_X(t)$$

همانطور که ملاحظه می‌کنید رابطه $f_X(t) = \dfrac{d F_{X(1)}(t)}{d t}$ نیز بین تابع توزیع و تابع چگالی برقرار است.

$$\text{Max}(X_j) \text{ تابع توزیع} : F_{X(n)}(t) = \sum_{k=n}^{n} \binom{n}{k} \left(F_X(t)\right)^{k} \left(1-F_X(t)\right)^{n-k} = \left(F_X(t)\right)^{n}$$

$$\text{Max}(X_j) \text{ تابع چگالی} : f_{X(n)}(t) = n\binom{n}{n} \left(f_X(t)\right)^{n-1} \left(1-F_X(t)\right)^{n-n} f_X(t) = n\left(F_X(t)\right)^{n-1} f_X(t)$$

مثال ۵: متغیر تصادفی X را مدت زمان بین هر دو تماس تلفنی در یک شرکت در نظر می‌گیریم که بطور میانگین هر ۱۰ دقیقه یکبار رخ می‌دهد اگر بطور تصادفی ۴ نمونه از متغیر تصادفی X اختیار کنیم مطلوبست:

الف) تابع توزیع آمارهٔ مرتب نمونه تصادفی.

ب) تابع توزیع و چگالی ماکزیمم و مینیمم نمونه‌ها.

ج) احتمال اینکه در نمونه‌های مشاهده شده کوچکترین بازه زمانی بین دو تماس بیشتر از ۲ دقیقه باشد چقدر است؟

حل: الف) ابتدا توجه کنید که X یک متغیر تصادفی نمایی با پارامتر $\lambda = \dfrac{1}{10}$ می‌باشد. بنابراین تابع توزیع X عبارتست از:

$$F_X(t) = 1 - e^{-\lambda t} = 1 - e^{-\frac{1}{10}t} \qquad t > 0$$

بنابراین تابع توزیع آمارهٔ مرتب برای ۴ نمونه مشاهده شده عبارتست از:

$$F_{X(j)}(t) = \sum_{k=j}^{4} \binom{n}{k} \left(1-e^{\frac{-1}{10}t}\right)^{k} \left(e^{-\frac{1}{10}t}\right)^{n-k}$$

۴-۱۲-۲

ب) ماکزیمم آماره‌های مرتب در این مثال $X_{(4)}$ می‌باشد. که تابع توزیع و چگالی آن بدست می‌آید:

$$F_{X(4)}(t) = \left(F_X(t)\right)^{4} = \left(1-e^{-\frac{1}{10}t}\right)^{4} \qquad t > 0$$

$$F_{X(4)}(t) = 4\left(1-e^{-\frac{1}{10}t}\right)^{3} \frac{1}{10} e^{-\frac{1}{10}t}$$

مینیمم آماره‌های مرتب $X_{(1)}$ می‌باشد که تابع توزیع و چگالی آن نیز بصورت زیر می‌باشد:

dalbablog.wordpress.com

$$F_{X(1)}(t) = 1 - (1 - F_X(t))^4 = 1 - (1 - 1 + e^{-\frac{1}{10}t})^4 = 1 - e^{-\frac{4}{10}t}$$

$$F_{X(1)}(t) = 4(1 - 1 + e^{-\frac{1}{10}t})^3 \frac{1}{10} e^{-\frac{1}{10}t} = \frac{4}{10} e^{-\frac{3}{10}t} e^{-\frac{1}{10}t} = \frac{4}{10} e^{-\frac{4}{10}t}$$

توجه کنید که توزیع کوچکترین آماره مرتب در این مثال خود یک متغیر تصادفی نمایی با پارامتر $\lambda = \frac{4}{10}$ می‌باشد.

ج) احتمال اینکه کوچکترین بازه زمانی بین دو تماس در نمونه‌ها بیشتر از ۲ دقیقه باشد برابر است با: $p(X_{(1)} > 2)$ و داریم:

$$p(X_{(1)} > 2) = 1 - p(X_{(1)} < 2) = 1 - F_{X(1)}(t=2)$$

$$= 1 - (1 - e^{-\frac{4}{10} \times 2}) = e^{-0/8} \approx 0/449$$

۹-۱۳

مثال ۶: برای آماره مرتب $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ از یک متغیر تصادفی پیوسته X مقادیر مورد انتظار برای احتمال بزرگترین عضو نمونه بدست بیاورید:

حل: می‌بایستی $E\left[F_X(X_{(n)})\right]$ را بدست بیاوریم داریم:

$$f_{X(n)}(t) = n(F_X(t))^{n-1} f_X(t)$$

$$E\left[f_x(x_{(n)})\right] = \int_{-\infty}^{+\infty} f_X(t)\, n\, (F_X(t))^{n-1} f_X(t)$$

$$= \int_{-\infty}^{+\infty} (f_X(t))^{n+1} \int_{-\infty}^{+\infty} = \frac{n}{n+1} - 0 = \frac{n}{n+1}$$

۹-۱۴

۹.۴ توزیع نمونه‌ای $\frac{\overline{X} - \mu}{\frac{\delta}{\sqrt{n}}}$.

در این فصل نشان دادیم که توزیع $\frac{\overline{X} - \mu}{\frac{\delta}{\sqrt{n}}}$ نرمال استاندارد است، در صورتی که $X_1, X_2, \ldots, X_n$ از جامعه‌ای نرمال با میانگین μ و واریانس δ^2 انتخاب شده باشد. همینطور نشان دادیم که در صورتی که تعداد نمونه‌ها زیاد باشد $(n > 30)$ توزیع $\frac{\overline{X} - \mu}{\frac{\delta}{\sqrt{n}}}$ مستقل از توزیع جامعه دارای توزیع نرمال استاندارد می‌باشد.

حال اگر واریانس یک جامعه مجهول باشد بناچار می‌بایستی از واریانس نمونه S^2 در محاسبه $\frac{\overline{X} - \mu}{\frac{\delta}{\sqrt{n}}}$ استفاده کنیم. به این ترتیب می‌بایستی ابتدا تابع توزیع $\frac{\overline{X} - \mu}{\frac{\delta}{\sqrt{n}}}$ را بدست بیاوریم.

dalbablog.wordpress.com

هر گاه Z یک متغیر تصادفی نرمال استاندارد باشد $(Z \sim N(0,1))$ و Y نیز یک متغیر تصادفی خی دو با n درجه آزادی باشد $(Y \sim \chi^2_n)$ و Z و Y از یکدیگر مستقل باشند در این صورت متغیر تصادفی $T = \frac{Z}{\sqrt{\frac{Y}{n}}}$ را یک متغیر تصادفی t با n درجه آزادی می‌نامیم و بصورت $T \sim t_{(n)}$ نمایش می‌دهیم.

۹-۱۵

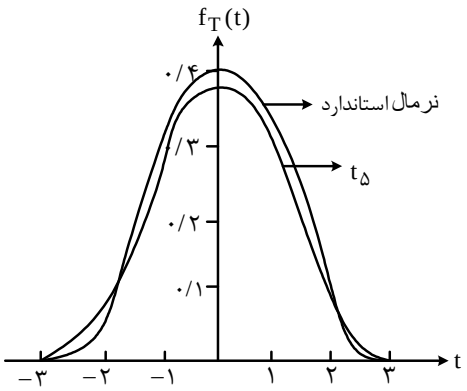
تابع چگالی متغیر تصادفی T بصورت زیر می‌باشد:

$$f_T(t) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}} \left[1+\frac{t^2}{n}\right]^{-\frac{n+1}{2}} \qquad -\infty < t < +\infty$$

میانگین و واریانس متغیر تصادفی t_n عبارتند از:

$$E[t_n] = 0 \quad , \quad var(t_n) = \frac{n}{n-2} \qquad n > 2$$

نمودار منحنی متغیر تصادفی t_n بسیار مشابه متغیر تصادفی نرمال استاندارد می‌باشد به شکل زیر که متغیر تصادفی t_5 را با نرمال استاندارد مقایسه می‌کند توجه کنید.



در واقع با افزایش درجه آزادی متغیر تصادفی t_n نمودار منحنی هر چه بیشتر به نمودار منحنی نرمال استاندارد نزدیک می‌شود.

۹-۱۶

توجه: اگر $X \sim t_1$

توزیع t_1 بصورت زیر می‌باشد:

$$f_X(x) = \frac{1}{\pi(1+x^2)} \qquad -\infty < x < +\infty$$

حال توجه کنید که معکوس توزیع X یعنی $Y = \frac{1}{X}$ برابر است با:

$$f_Y(y) = \frac{1}{\pi(1+y^2)}$$

به عبارت دقیق‌تر عکس یک متغیر t_1 بازهم t_1 می‌باشد.

dalbablog.wordpress.com

حال به محاسبه توزیع $\frac{\overline{X}-\mu}{\frac{\delta}{\sqrt{n}}}$ می‌پردازیم. طبق تعریف می‌دانیم اگر $Z \sim N(0,1)$ و $Y \sim \chi^2_n$ متغیر تصادفی $T = \frac{Z}{\sqrt{\frac{Y}{n}}}$ یک متغیر تصادفی t_n می‌باشد حال در نظر بگیرید: $Z = \frac{\overline{X}-\mu}{\frac{\delta}{\sqrt{n}}}$ و $Y = \frac{(n-1)S^2}{\delta^2}$ قبلاً نشان دادیم که Z و Y به ترتیب متغیرهای تصادفی نرمال استاندارد و χ^2_{n-1} می‌باشند در نتیجه متغیر تصادفی $T = \frac{Z}{\sqrt{\frac{Y}{n-1}}}$ می‌بایستی یک متغیر تصادفی t با $n-1$ درجه آزادی می‌باشد. با ساده نمودن عبارت T داریم:

$$T = \frac{\frac{\overline{X}-\mu}{\frac{\delta}{\sqrt{n}}}}{\sqrt{\frac{(n-1)S^2}{\delta^2(n-1)}}} = \frac{\frac{\overline{X}-\mu}{\frac{\delta}{\sqrt{n}}}}{\sqrt{\frac{S^2}{\delta^2}}} = \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$$

در نتیجه متغیر تصادفی $T = \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$ یک متغیر تصادفی t با $n-1$ درجه آزادی می‌باشد.

توجه کنید در صورتی که نمونه‌های تصادفی $X_1, X_2, \ldots, X_n$ از جامعه‌ای نرمال باشند رابطه $T = \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$ یک متغیر تصادفی t_{n-1} می‌باشد.

اما اگر جامعه نرمال نباشد یا توزیع آنرا ندانیم یا n به اندازه کافی بزرگ $(n > 30)$ باشد آنگاه باز هم $T = \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$ یک متغیر تصادفی t_{n-1} می‌باشد حتی در این حالت می‌توان با توجه به تشابه نمودار منحنی‌های t_{n-1} و نرمال استاندارد می‌توان گفت که توزیع $T = \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$ برای $n > 30$ همان نرمال استاندارد می‌باشد.

۹-۱۷

محاسبه مقادیر احتمال t_n.

برای بدست آوردن مقادیر مختلف احتمال t_n از جدول احتمالات $F_t(x)$ استفاده می‌کنیم بطوریکه $F_t(x)$ برابر است با:

$$F_t(x) = \int_{-\infty}^{x} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}} (1+x^2)^{\frac{-(n+1)}{2}} dx = 1-\alpha$$

$$F_t(x) = p(T < x)$$

مقادیر هر یک از ردیف یا معادل با درجه آزادی و مقادیر ستون‌ها معادل با مقدار احتمال $F_t(x)$ می‌باشد. جدول زیر بخشی از جدول تابع توزیع t را نشان می‌دهد:

$1-\alpha$ / n	۰/۶	۰/۷۵	۰/۹۰	۰/۹۵	۰/۹۷۵
۱	۰/۳۲۵	۱/۰۰	۳/۰۷۸	۶/۳۱۴	۱۲/۷۰۶
۲	۰/۲۸۹	۰/۸۱۶	۱/۸۸۶	۲/۹۲	۴/۳۰۳
۳	۰/۲۷۷	۰/۷۶۵	۱/۶۳۸	۲/۳۵۳	۳/۱۸۲

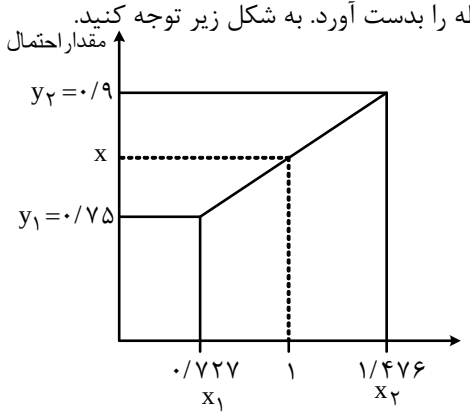
مرکز آموزش الکترونیکی دانشگاه علم و صنعت ایران

dalbablog.wordpress.com

حال با توجه به جدول به عنوان مثال $p(t_{۲} < ۲/۹۲)$ برابر با ۰/۹۵ می‌باشد توجه کنید که مقادیر احتمال از عناوین هر یک از ستون‌ها بدست می‌آیند. برای بدست آوردن سایر مقادیر احتمال از تقریب زدن استفاده می‌کنیم.

۹-۱۸

به عنوان مثال برای بدست آوردن مقدار احتمال $p(t_{۵} < ۱)$ از آنجا که در جدول عدد ۱ در ردیف پنجم ما بین دو مقدار ۱/۴۷۶ و ۰/۷۲۷ می‌باشد می‌بایستی با استفاده از روش خطی مقدار این احتمال را تقریب بزنیم در این روش ما بین دو نقطه $(۰/۷۲۷ , ۰/۷۵)$ و $(۱/۴۷۶ , ۰/۹)$ یک خط در نظر می‌گیریم، با نوشتن معادله خط به سادگی می‌توان مقادیر احتمال ما بین این دو نقطه را بدست آورد. به شکل زیر توجه کنید.



معادله خط عبارتست از: $y = \frac{y_{۲} - y_{۱}}{x_{۲} - x_{۱}} (x - x_{۱}) + y_{۱}$

بنابراین:

$$y = \frac{۰/۹ - ۰/۷۵}{۱/۴۷۶ - ۰/۷۲۷} (x - ۰/۷۲۷) + ۰/۷۵$$

حال برای بدست آوردن احتمال هر یک از مقادیر ما بین $x_{۱}, x_{۲}$ کافیست آن را به جای x در معادله بالا قرار دهیم تا y که مقدار احتمال تقریبی می‌باشد بدست آید.

برای احتمال $p(t_{۵} < ۱)$ داریم:

$$p(t_{۵} < ۱) = y = \frac{۰/۹ - ۰/۷۵}{۱/۴۷۶ - ۰/۷۲۷} (۱ - ۰/۷۲۷) + ۰/۷۵ = ۰/۸$$

مثال: تعداد مراجعه کنندگان به یک فروشگاه در طول روز یک متغیر تصادفی نرمال می‌باشد. مدیر فروشگاه می‌داند که بطور متوسط روزانه ۱۲۰ نفر به فروشگاه مراجعه می‌کنند. تعداد مراجعه کنندگان به فروشگاه در طول یک هفته بصورت زیر بدست آمده است:

۹۰ , ۱۱۰ , ۷۵ , ۱۳۰ , ۱۵۰ , ۱۲۰ , ۱۰۰

مطلوبست:

الف) احتمال اینکه میانگین تعداد مراجعه کنندگان حداقل ۱۲۵ نفر در روز باشد؟

ب) احتمال اینکه تفاوت میانگین نمونه‌ها با میانگین جامعه حداکثر ۲ واحد باشد؟

dalbablog.wordpress.com

حل: الف) ابتدا از روی نمونه‌های بدست آمده مقدار $\overline{X}$ و S^2 را بدست می‌آوریم. داریم:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{7}(90+110+75+130+150+120+100) = 110/7$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{X})^2 = \frac{1}{6}[(90-110/7)^2+(75-110/7)^2+(130-110/7)^2+(150-110/7)^2$$

$$(120-110/7)^2+(100-110/7)^2] = 636/9 \qquad \Rightarrow \qquad S = 25/2$$

حال می‌بایستی مقدار احتمال $p(\overline{X} > 125)$ را بدست بیاوریم: (با توجه به اینکه واریانس جامعه را نمی‌دانیم بنابراین می‌بایستی از توزیع t استفاده کنیم.)

$$p(\overline{X} > 110) = p(\overline{X}-\mu > 125-120) = p\left(\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} > \frac{125-120}{\frac{25/2}{\sqrt{7}}}\right)$$

$$= p\left(\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} > \frac{5}{9/52}\right) = p(t_6 > 0/525) = 1-p(t_6 < 0/525) = 1-0/68 = 0/32$$

ب) در این حالت می‌بایستی احتمال $p(\overline{X}-\mu < 2)$ را بدست بیاوریم:

$$p(\overline{X}-\mu < 2) = p\left(\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} < \frac{2}{9/52}\right) = p(t_6 < 0/21) \approx 0/58$$

۹-۲۰

توزیع نمونه‌ای نسبت واریانسهای نمونه

اگر از دو جامعه نرمال دو نمونه مستقل به اندازه‌های n_1 و n_2 بگیریم برای بدست آوردن نسبت واریانسهای دو نمونه $\left(\frac{S_1}{S_2}\right.$ یا $\left.\frac{S_2}{S_1}\right)$ از توزیعی به نام F (فیشر) استفاده می‌کنیم.

تعریف: اگر U یک متغیر تصادفی خی دو با m درجه آزادی $(U \sim \chi^2_m)$ و V یک متغیر تصادفی خی دو با n درجه آزادی $(V \sim \chi^2_m)$ باشند.

در این صورت توزیع متغیر تصادفی $F = \frac{\frac{U}{m}}{\frac{V}{n}}$ را یک متغیر F با m و n درجه آزادی می‌نامند و با نماد $F \sim F_{m,n}$ نمایش می‌دهند. تابع چگالی احتمال توزیع $F_{m,n}$ بصورت زیر می‌باشد:

$$f_F(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)\left(\frac{m}{n}\right)^{\frac{m}{2}}}{\Gamma\left(\frac{m}{n}\right)\Gamma\left(\frac{n}{2}\right)}\, x^{\frac{m}{2}-1}\left(1+\frac{m}{n}x\right)^{-\frac{m+n}{2}} \qquad x > 0$$

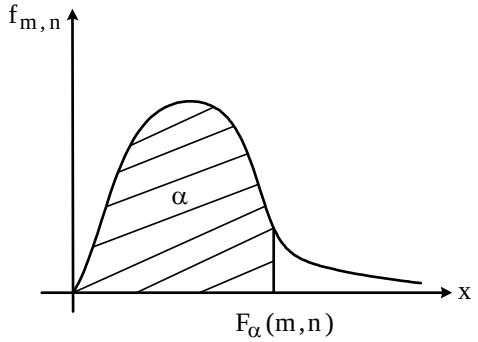
۹-۲۱

مقادیر میانگین و واریانس توزیع $F_{m,n}$ عبارتند از:

$$E[X] = \frac{n}{n-2} \qquad n > 2 \quad , \quad var(X) = 2\left(\frac{1}{n}+\frac{1}{m}\right)$$

dalbablog.wordpress.com

شکل نمودار تابع توزیع $F_{m,n}$ بصورت زیر می‌باشد:



حال برای بدست آوردن توزیع نسبت $\frac{S_1^2}{S_2^2}$ بصورت زیر عمل می‌کنیم:

Y و X را دو جمعیت نرمال با واریانسهای δ_1^2 و δ_2^2 در نظر می‌گیریم

فرض: $X \sim N(\mu_1 , \delta_1^2)$; $Y \sim N(\mu_2 , \delta_2^2)$

می‌دانیم:

$$V = \frac{(n_1 - 1) S_1^2}{\delta_1^2} \sim \chi^2_{(n_1 - 1)}$$

$$V = \frac{(n_2 - 1) S_2^2}{\delta_2^2} \sim \chi^2_{(n_2 - 1)}$$

حال با توجه به تعریف تابع توزیع F داریم:

$$F = \frac{\frac{U}{n-1}}{\frac{V}{n_2 - 1}} = \frac{\frac{(n_1 - 1) S_1^2}{\delta_1^2 (n_1 - 1)}}{\frac{(n_2 - 1) S_2^2}{\delta_2^2 (n_2 - 1)}} = \frac{\frac{S_1^2}{\delta_1^2}}{\frac{S_2^2}{\delta_2^2}} = \frac{S_1^2 \delta_2^2}{S_2^2 \delta_1^2} \sim F_{n_1 - 1 , n_2 - 1}$$

۹-۲۲

مثال: اگر X یک متغیر تصادفی t با n درجه آزادی باشد مربع متغیر تصادفی X دارای چه توزیعی می‌باشد؟

حل: فرض می‌کنیم: $Z \sim N(0 , 1)$ و $Y = \chi^2_n$ در این صورت داریم:

$$X = t_n \sim \frac{Z}{\sqrt{\frac{Y}{n}}}$$

$$\Rightarrow X^2 = \frac{Z^2}{\frac{Y}{n}}$$

از آنجا که مربع یک متغیر تصادفی نرمال استاندارد یک متغیر تصادفی خی دو با یک درجه آزادی است. بنابراین:

$$X^2 = \frac{\chi_1^2}{\frac{Y}{n}} = \frac{\frac{\chi_1^2}{1}}{\frac{\chi_n^2}{n}} \sim F_{1,n}$$

dalbablog.wordpress.com

بنابراین مربع یک متغیر تصادفی t با n درجه آزادی یک متغیر تصادفی F با $n,1$ درجه آزادی می‌باشد.

۹-۲۳

مثال: تعداد تصادفات در دو شهر A و B متغیر تصادفی نرمال به ترتیب با واریانس ۲۰ و ۲۲ تصادف در روز می‌باشد. اگر از دو جامعه فوق دو نمونه به ترتیب با حجم ۲۱ و ۳۱ نمونه انتخاب کرده باشیم و S_A^2 , S_B^2 به ترتیب واریانسهای نمونه‌های گرفته شده از شهر A و B باشد محاسبه احتمال اینکه S_A^2 حداقل ۱/۵ برابر S_B^2 باشد؟

حل: داریم:

$$\sigma_A^2 = ۲۰ \quad ; \quad \sigma_B^2 = ۲۲$$

می‌بایستی احتمال $p(S_A^2 \geq ۱/۵\ S_B^2)$ را محاسبه کنیم داریم:

$$p(S_A^2 \geq ۱/۵\ S_B^2) = p\left(\frac{S_A^2}{S_B^2} \geq ۱/۵\right)$$

$$p\left(\frac{\frac{S_A^2}{\sigma_A^2}}{\frac{S_B^2}{\sigma_B^2}} \geq ۱/۵ \ \left(\frac{۲۲}{۲۰}\right)\right) = p\left(\frac{S_A^2\,\sigma_B^2}{S_B^2\,\sigma_A^2} \geq ۱/۶۵\right)$$

$$= ۱ - p(F_{۲۰,۳۰} < ۱/۶۵) \approx ۱ - ۰/۹ = ۰/۱$$

۹-۲۴

تقریب توزیع F به توزیع خی دو.

در صورتی که در یک متغیر تصادفی $F_{m,n}$ شرط $n \to \infty$ برقرار باشد می‌توانیم برای محاسبه مقادیر احتمال mF از متغیر تصادفی χ_m^2 استفاده نمایم. به عبارت دیگر $mF_{m,n}$ با شرط $n \to \infty$ دارای توزیع χ_m^2 می‌باشد.

توزیع نمونه‌ای اختلاف میانگین‌ها

با استفاده از توزیع F احتمالات مربوط به نسبت واریانسهای دو نمونه را بدست آوریم حال می‌خواهیم برای دو نمونه گرفته شده از دو جامعه نرمال توزیع اختلاف میانگین‌ها را بدست آوریم.

از دو جامعه به ترتیب با میانگین‌های $\mu_۱, \mu_۲$ و واریانسهای $\sigma_۱^2, \sigma_۲^2$ دو نمونه به اندازه $n_۱, n_۲$ انتخاب می‌کنیم. در این صورت برای میانگین نمونه‌ها داریم:

$$\overline{X}_۱ \sim N\left(\mu_۱ , \frac{\sigma_۱^2}{n_۱}\right)$$

$$\overline{X}_۲ \sim N\left(\mu_۲ , \frac{\sigma_۲^2}{n_۲}\right.$$

از آنجا که $\overline{X}_۱, \overline{X}_۲$ از یکدیگر مستقل می‌باشد و $\overline{X}_۱ - \overline{X}_۲$ نیز یک ترکیب خطی‌از دو متغیر تصادفی نرمال می‌باشد بنابراین $\overline{X}_۱ - \overline{X}_۲$ نیز نرمال است و توزیع آن عبارتست از:

$$\overline{X}_۱ - \overline{X}_۲ \sim N\left(\mu_۱ - \mu_۲ , \frac{\sigma_۱^2}{n_۱} + \frac{\sigma_۲^2}{n_۲}\right)$$

با تبدیل آن به متغیر تصادفی نرمال استاندارد داریم:

dalbablog.wordpress.com

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

حتی اگر دو جمعیت نرمال نباشند با استفاده از قضیه حد مرکزی می‌دانیم که اگر داشته باشیم $n_1, n_2 \geq 30$ باز هم هر یک از $\overline{X}_2, \overline{X}_1$ دارای توزیع تقریبی نرمال می‌باشند. و در نتیجه $\overline{X}_1 - \overline{X}_2$ نیز دارای توزیع نرمال خواهد بود. و روابط فوق برقرار خواهند بود.

۹-۲۵

مثال: در مثال قبل اگر متوسط تعدد تصادفات در شهر A ۱۰۰ و در شهر B ۱۲۰ تصادف در روز باشند. اگر دو نمونه به حجم ۳۰ انتخاب کنیم در این صورت مطلوبست احتمال اینکه میانگین تعداد تصادفات در شهر B حداقل ۱۸ تصادف از شهر A بیشتر باشد

حل: با توجه به مفروضات مساله داریم:

$$\mu_A = 100 \qquad \sigma_A^2 = 20 \qquad n_A = 30$$

$$\mu_B = 120 \qquad \sigma_B^2 = 22 \qquad n_B = 30$$

$$\overline{X}_A - \overline{X}_B \sim N\left(100 - 120, \frac{20}{30} + \frac{22}{30}\right) \sim N(-20, 1/4)$$

حال می‌بایستی احتمال $p(\overline{X}_B \geq \overline{X}_A + 18)$ را محاسبه کنیم:

$$p(\overline{X}_B \geq \overline{X}_A + 18) = p(\overline{X}_A - \overline{X}_B \leq -18)$$

$$= p(\overline{X}_A - \overline{X}_B - (\mu_A - \mu_B)) \leq -18 - (100 - 120))$$

$$= p\left(\frac{\overline{X}_A - \overline{X}_B - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} \leq \frac{2}{1/18}\right)$$

$$= p(Z \leq 1/69) = 0/975$$

۹-۲۶

اگر در محاسبه توزیع اختلاف میانگین نمونه‌ها واریانس دو جامعه نامعلوم اما مساوی باشد $(\sigma_1^2 = \sigma_2^2 = \sigma^2)$ می‌توانیم از واریانس نمونه‌ها $(S_2^2, S_1^2, \ldots)$ به عنوان تخمین σ^2 استفاده کنیم. در این حالت از میانگین وزنی این دو واریانس برای برآورد σ^2 استفاده می‌کنیم:

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

در این حالت اگر دو جامعه نرمال باشند داریم:

$$\frac{(n_1 + n_2 - 2) S_p^2}{\sigma^2} = \frac{(n_1 - 1) S_1^2}{\sigma^2} + \frac{(n_2 - 1) S_2^2}{\sigma^2} = \chi^2_{(n_1 + n_2 - 2)}$$

حال با توجه به توزیع t داریم:

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{(n_1 + n_2 - 2)}$$

بنابراین در محاسبه احتمال مربوط به نمونه‌های بدست آمده از دو جامعه با واریانس نامعلوم اما مساوی می‌بایستی از توزیع t با $n_1 + n_2 - 2$ درجه آزادی استفاده کنیم.

dalbablog.wordpress.com

فصل دهم

S-۱

۱- برآورد

در فصل قبل با نمونه‌گیری از یک جامعه با معلوم بودن تابع چگالی و پارامترهای جامعه مثل میانگین و واریانس آشنا شدید. همچنین نحوه بدست آوردن احتمالات مربوط به آماره‌های خاص مثل میانگین، نسبت واریانسها و ..... را نشان دادیم.

در عمل بسیاری ار اوقات می‌دانیم که یک جامعه مثلاً نرمال می‌باشد اما مقدار دقیق پارامترهای جامعه را که μ و $\sigma^{۲}$ می‌باشند نمی‌دانیم، در این حالت با نمونه‌گیری از جامعه و با استنباط آماری روی نمونه‌ها می‌توانیم مقادیر پارامترهای مجهول جامعه را با دقت زیادی بدست بیاوریم. بنابراین در این فصل به بحث پیرامون روشهای برآوردیابی می‌پردازیم.

۱۰۱۰۲ برآورد نقطه‌ای

با نمونه‌گیری از جامعه دو پارامتر میانگین نمونه‌ها و واریانس نمونه‌ها را بدست می‌آوریم. این دو پارامتر نقش اصلی را در برآورد میانگین و واریانس جامعه بازی می‌کنند. بطور کلی با استفاده از میانگین و واریانس نمونه‌ها به دو طریق می‌توان پارامترهای مجهول جامعه را برآورد نمود. در روش اول که به برآورد نقطه‌ای معروف است با برابر قرار دادن میانگین و واریانس نمونه‌ها با میانگین جامعه، این دو پارامتر مجهول را بدست می‌آوریم. اما در روش دوم که به برآورد فاصله‌ای معروف است برای پارامتر مجهول جامعه مثل میانگین یک بازه با استفاده از پارامترهای معلوم نمونه بدست می‌آوریم و نشان می‌دهیم که با احتمال زیاد پارامتر مجهول جامعه در این بازه قرار دارد.

در برآورد نقطه‌ای از دو روش زیر استفاده می‌کنیم که در ادامه با آنها آشنا می‌شوید:

۱- برآورد به روش گشتاورها

۲- برآورد به روش حداکثر احتمال (در ..... ماکزیمم).

۱۰۱۰۱۰۳ برآورد به روش گشتاورها

قبلاً kامین گشتاور متغیر تصادفی X را بصورت زیر معرفی نمودیم:

$$m_k = E\left[X^K\right]$$

به همین ترتیب می‌توانیم kامین گشتاور نمونه‌ها را بصورت زیر معرفی کنیم:

$$m'_k = \frac{۱}{n} \sum_{i=۱}^{n} X_i{}^k$$

n: تعداد نمونه‌ها

با توجه به اینکه kامین گشتاور متغیر تصادفی X به تمام مقادیر احتمال متغیر X وابسته است و همچنین kامین گشتاور نمونه‌ها نیز به هر یک از مقادیر نمونه‌ها وابسته است بنابراین این طور بنظر می‌رسد که با استفاده از گشتاورها بتوان بنحوی پارامترهای مجهول جامعه را تقریب زد.

در این روش به این صورت عمل می‌کنیم که اگر p عدد پارامتر مجهول داشته باشیم به ترتیب گشتاور اول نمونه را با گشتاور اول متغیر تصادفی x، گشتاور دوم نمونه را با گشتاور دوم متغیر تصادفی و گشتاور pام نمونه را با گشتاور pام متغیر تصافی X برابر قرار می‌دهیم به این ترتیب دستگاه معادلات زیر بدست می‌آید:

$$\begin{cases} m_۱ = m'_۱ \\ m_۲ = m'_۲ \\ \quad \vdots \\ m_k = m'_k \end{cases} \qquad k = ۱, ۲, \ldots, p$$

اگر $\theta_۱, \theta_۲, \ldots, \theta_p$ مقادیر مجهول پارامترها باشد با حل دستگاه فوق هر یک از مقادیر θ_i بدست می‌آید.

حال به مثال زیر توجه کنید:

S-۲

dalbablog.wordpress.com

۴ مثال ۱: می‌دانیم تعداد مراجعه کنندگان به یک پمپ بنزین در طول روز یک متغیر تصادفی پواسون با پارامتر مجهول λ می‌باشد. اگر در طول ۱۰ روز مشاهدات زیر را برای تعداد مراجعات به پمپ بنزین بدست آورده باشیم مطاوبست تعیین پارامتر مجهول λ؟

۱۲۰ , ۹۰ , ۸۵ , ۱۱۰ , ۱۳۵ , ۹۵ , ۱۰۰ , ۱۱۱ , ۱۱۵ , ۹۸

حل: برای متغیر تصادفی پواسون داریم:

$$m_1 = E[X'] = E[X] = \lambda$$

از آنجا که تنها یک پارامتر مجهول λ داریم استفاده از اولین گشتاور کفایت می‌کند. حال اولین گشتاور نمونه را بدست می‌آوریم:

$$m'_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} = \frac{1}{10}(120 + 90 + 85 + 110 + 135 + 95 + 100 + 111 + 115 + 98) = 105/9$$

حال با برابر قرار دادن اولین گشتاور نمونه با اولین گشتاور متغیر تصادفی X مقدار λ بدست می‌آید:

$$m_1 = m'_1 = 105/9 \quad \Rightarrow \quad \lambda = 105/9$$

۵ مثال ۲: در یک سری مسابقات تیراندازی n مرتبه به هدف شلیک می‌ند که احتمال اصابت گلوله به هدف p می‌باشد. نتایج حاصل از ۱۰ مرتبه شرکت تیرانداز در مسابقات بصورت زیر می‌باشد: (نتایج بر حسب تعداد دفعات اصابت گلوله به هدف می‌باشند)

۷ , ۵ , ۶ , ۸ , ۸ , ۶ , ۷ , ۴ , ۹ , ۸

مطلوبست برآورد پارامترهای مجهول n و p؟

ابتدا توجه می‌نیم که در این مثال تعداد پیروزی‌ها در n مرتبه انجام آزمایشهای مستقل برنولی با احتمال پیروزی p مورد نظر است. بنابراین متغیر تصادفی X دو جمله‌ای با پارامترهای n و p می‌باشد.

در این مثال دو پارامتر مجهول داریم بنابراین از گشتاور مرتبه اول و دوم استفاده می‌کنیم:

$$m_1 = E[X] = np$$

$$m_2 = E[X^2] = \sigma^2 + \mu^2 = npq + n^2p^2 = n(n-1)p + np = np(np + 1 - p)$$

توجه کنید که مقادیر اول و دوم گشتاور متغیر تصادفی X همواره با استفاده از میانگین و واریانس متغیر تصادفی X بدست می‌آیند. به عبارتی در برآورد به روش گشتاورها مستقل از اینکه توزیع متغیر تصادفی X چه باشد همواره داریم:

$$\begin{cases} m_1 = \mu \\ m'_1 = \overline{X} \end{cases} \quad \Rightarrow \quad \mu = \overline{X}$$

$$\begin{cases} m_2 = \sigma^2 + \mu^2 \\ m'_2 = \frac{1}{n}\sum_{i=1}^{n} \overline{X}_i^2 \end{cases} \quad \Rightarrow \quad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} (\overline{X}_i - \overline{X})^2 = \frac{1}{n}\sum_{i=1}^{n} \overline{X}_i^2 - \overline{X}^2$$

یعنی برآورد گشتاوری میانگین و واریانس هر توزیعی برابر است با میانگین و واریانس نمونه‌ها. به همین دلیل $\overline{X}$, $\frac{1}{n}\sum (X_i - \overline{X})^2$ برآوردهای توزیع – آزاد نامیده می‌شوند، یعنی برآوردهایی که مستقل از توزیع تصادفی X می‌باشند.

۶ حال مقادیر گشتاور اول و دوم نمونه را بدست می‌آوریم:

$$m'_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{10}(7+5+6+8+8+6+7+4+9+8) = 6/8$$

$$m'_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \frac{1}{10}(49+25+36+64+64+36+49+16+81+64) = 48/4$$

به این ترتیب دستگاه معادلات زیر بدست می‌آید:

dalbablog.wordpress.com

$$\begin{cases} m_1 = m'_1 & \Rightarrow \quad np = 6/8 \\ m_2 = m'_2 & \Rightarrow \quad np = (np + 1 - p) = 48/4 \end{cases}$$

با حل دستگاه بدست می‌آوریم:

$$6/8\,(6/8 + 1 + p) = 48/4 \quad \Rightarrow \quad 53/04 - 6/8\,p = 48/4 \quad \Rightarrow \quad p = 0/68$$

$$np = 6/8 \;\rightarrow\; n\,(0/68) = 6/8 \Rightarrow n = 10$$

بنابراین تیرانداز فوق در هر مسابقه می‌بایستی حدوداً ۱۰ مرتبه به هدف شلیک کرده باشد که احتمال موفقیت وی در هر شلیک ۰/۶۸ می‌باشد.
برآورد به روش گشتاورها همواره نتایج دقیق و رضایت بخشی نمی‌دهد به مثال بعد در این زمینه توجه کنید:

۷ مثال ۳: فرض کنید متغیر تصادفی X در بازه $[\,0, a\,]$ بصورت یکنواخت توزیع شده باشد در این صورت مقدار a را به ازای نمونه‌های بدست آمده زیر برآورد کنید:
الف) ۵ و ۴ و ۳ و ۱ و ۸ و ۵ و ۴ و ۲ و ۰ و ۱
ب) ۵ و ۱۰ و ۴۰

حل: برای متغیر تصادفی پیوسته یکنواخت داریم:

$$m_1 = E[X] = \int_0^a x \,\frac{1}{a}\, dx = \frac{1}{a}\,\left(\frac{x^2}{2}\right)\Big|_0^a = \frac{a}{2}$$

همینطور برای نمونه‌ها داریم:

$$m'_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{10}\,(1+0+2+4+5+8+1+3+4+5) = 3/3$$

بنابراین در حالت کلی:

$$a = 2\overline{X} \quad \Leftarrow \quad m'_1 = \overline{X} = \frac{a}{2}$$

یعنی برآورد a عبارتست از دو برابر میانگین نمونه‌های بدست آمده که در اینجا با توجه به نمونه‌های الف a برابر می‌شود با: $a=2\overline{X}=2\times 3/3=6/6$.
حال توجه کنید که $a = 6/6$ یعنی نمونه‌ها در اصل از بازه $[\,0\,,\,6/6\,]$ انتخاب شده‌اند در حالی که در بین نمونه‌ها عدد ۸ موجود می‌باشد که داخل بازه فوق نمی‌باشد و این یعنی برآورد به روش گشتاورها دارای کمبودهایی می‌باشد. همین حالت برای نمونه‌های (ب) نیز صادق است:
ب)

$$\overline{X} = \frac{1}{3}\,(40+10+5) = 18/3 \quad \Rightarrow \quad a = 2\overline{x} = 36/6$$

بنابراین برآورد به روش گشتاورها در بعضی موارد نتایج دلخواه را بدست نمی‌دهد در نتیجه می‌بایستی از معیاری استفاده کنیم که میزان کارایی یک روش برآورد را نشان دهد تا بتوان میان روشهای مختلف برآورد، بهترین روش را برای مسایل مختلف انتخاب نمود.

S-۳

۲.۱.۱۰ برآورد به روش حداکثر احتمال

متغیر تصادفی X با تابع چگالی احتمال $f_X(x)$ و پارامترهای مجهول $\theta_1, \theta_2, \ldots, \theta_p$ را در نظر بگیرید. از این متغیر تعداد n نمونه استخراج می‌کنیم که عبارتند از $x_1, x_2, \ldots, x_n$ حال $X_1, X_2, \ldots, X_n$ را متغیرهای تصادفی متناظر با نمونه‌ها در نظر می‌گیریم.
می‌دانیم تابع چگالی احتمال توام متغیرهای تصادفی $X_1, \ldots, X_n$ مقادیر احتمال وقوع نمونه‌ها را بدست می‌دهد. مقادیری از θ_iها که تابع چگالی احتمال توام X_iها را ماکزیمم می‌کند برآورد پارامترهای مجهول θ_i می‌باشد. زیرا به ازای ماکزیمم شدن تابع چگالی احتمال توام X_iها در واقع احتمال وقوع نمونه‌ها به بیشترین مقدار خود می‌رسد. برای بدست آوردن برآورد θ_iها فرض می‌کنیم $L(\theta_1, \theta_2, \ldots, \theta_p)$ تابع چگالی احتمال توام متغیرهای تصادفی X_i باشد در این صورت با توجه به اینکه X_iها از یکدیگر مستقل باشند تابع چگالی احتمال توام آنها عبارتست از:

$$L(\theta_1, \theta_2, \ldots, \theta_p) = f_{X_1}(x_1)\,.\,f_{X_2}(x_2)\cdots f_{X_n}(x_n)$$

dalbablog.wordpress.com

حال با مشتق‌گیری از $L(\theta_1, \ldots, \theta_p)$ نسبت به θ_iها و برابر صفر قرار دادن این مشتق‌ها دستگاهی از معادلات بدست می‌آید که با حل این دستگاه مقادیر هر یک از θ_iها بدست می‌آید. معمولاً در محاسبات برای سادگی از $Ln(L)$ مشتق می‌گیریم. به مثال زیر توجه کنید:

۹ مثال ۴: نمونه‌های $X_1, X_2, \ldots, X_n$ را از متغیر تصادفی برنولی X داریم مطلوبست: برآورد پارامتر p به روش حداکثر احتمال (MLE)؟

حل: تابع چگالی متغیر تصادفی برنولی X به صورت زیر می‌باشد:

$$f_X(x) = p^x (1-p)^{1-x} \qquad x = 0, 1$$
$$= 0 \qquad \text{سایر مقادیر}$$

به این ترتیب تابع چگالی احتمال توام نمونه‌ها عبارتست از:

$$L(p) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdots f_{X_n}(x_n) = \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}$$

$$p^{\sum_{i=1}^{n} x_i} (1-p)^{\sum_{i=1}^{n} (1-x_i)} \quad \Rightarrow \quad L(p) = p^{\sum_{i=1}^{n} x_i} (1-p)^{n-\sum_{i=1}^{n} x_i}$$

حال برای اینکه محاسبات ساده‌تر شوند از طرفین رابطه بالا Ln می‌گیریم:

$$Ln(L(p)) = (n\,(p^{\sum x_i} (1-p)^{n-\sum x_i})$$
$$= \sum x_i \; Ln(p) + (n - \sum x_i)\,(n(1-p)$$

فرض می‌کنیم $H(p) = Ln(L(p))$ در این صورت می‌بایستی معادله $\frac{dH}{dP} = 0$ را حل کنیم تا مقدار مجهول p بدست آید. بنابراین:

$$\frac{dH}{dP} = 0 \quad \Rightarrow \quad \sum x_i \,(\frac{1}{p}) - (n - \sum x_i)\,(\frac{1}{1-p}) = 0$$

همانطور که ملاحظه می‌کنید با بکار بردن Ln در محاسبات، مشتق‌گیری بسیار ساده‌تر می‌شود. حال داریم:

$$\sum x_i \,(\frac{1}{p}) = (n - \sum x_i)\,(\frac{1}{1-p})$$
$$\Rightarrow \frac{1-p}{p} = \frac{n - \sum x_i}{\sum x_i} \quad \Rightarrow \quad \frac{1}{p} - 1 = \frac{n}{\sum x_i} - 1$$
$$\Rightarrow \frac{1}{p} = \frac{n}{\sum x_i} \quad \Rightarrow \quad \hat{p} = \frac{1}{n} \sum x_i$$

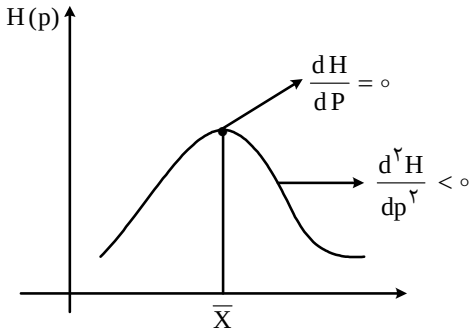
($\hat{p}$ برآورد پارامتر مجهول p می‌باشد).

بنابراین برآورد پارامتر مجهول p برابر است با $\overline{X} = \frac{1}{n} \sum x_i$ توجه کنید مقدار $\hat{p} = \overline{X}$ بدست آمده تنها در صورتی جواب درستی می‌باشد که نقطه $\overline{X}$ واقعاً مقدار ماکزیمم تابع $H(P)$ باشد. مثلاً ممکن است این نقطه، نقطه مینیمم تابع $H(P)$ باشد. برای اطمینان از اینکه $\hat{p} = \overline{X}$ مقدار ماکزیمم $H(P)$ می‌باشد باید شرط زیر برقرار باشد:

$$\frac{d^2 H}{dp^2} < 0$$

یعنی تقعر تابع $H(P)$ حول $\hat{p} = \overline{X}$ روبه پایین باشد به این ترتیب $\hat{p} = \overline{X}$ نقطه ماکزیمم تابع می‌باشد. در این رابطه به شکل زیر توجه کنید:

dalbablog.wordpress.com



حال داریم:

$$\frac{d^{۲}H}{dp^{۲}} = \sum x_i \left(\frac{-۱}{p^{۲}}\right) - \left(n - \sum x_i\right)\left(\frac{۱}{(۱-p)^{۲}}\right)$$

که به ازای هر مقدار p منفی است. بنابراین $\overline{X}$ به عنوان برآورد p در متغیر تصادفی برنولی می‌باشد.

۱۱ حال برآورد p را با استفاده از روش گشتاورها بدست می‌آوریم و نتیجه را با روش حداکثر احتمال مقایسه می‌کنیم:
مطابق روش گشتاورها داریم:

$$m_۱ = p$$

$$\Rightarrow \quad m_۱ = m'_۱ \Rightarrow \quad p = \overline{X}$$

$$m'_۱ = \frac{۱}{n}\sum x_i$$

همانطور که ملاحظه می‌کنید نتیجه برآورد از هر دو روش یک نتیجه را بدست می‌دهد. بطور کلی از آنجا که روش حداکثر معمولاً برآوردهای بهتری می‌دهد، هرگاه نتیجه دو روش برآورد با یکدیگر برابر نبود، نتیجه روش برآورد حداکثر احتمال را ملاک قرار می‌دهیم.در برآورد به روش گشتاورها نشان دادیم که این روش برای مواردی مثل مثال ۳ برآورد مطلوبی ارایه نمی‌کند بنابراین در مثال بعدی مثال ۳ را با روش برآورد حداکثر احتمال حل می‌کنیم:

S ۴

۱۲ مثال ۵: متغیر X یک متغیر تصادفی یکنواخت در بازه $(۰ , a)$ می‌باشد. مطلوبست برآورد پارامتر مجهول a به روش حداکثر احتمال؟

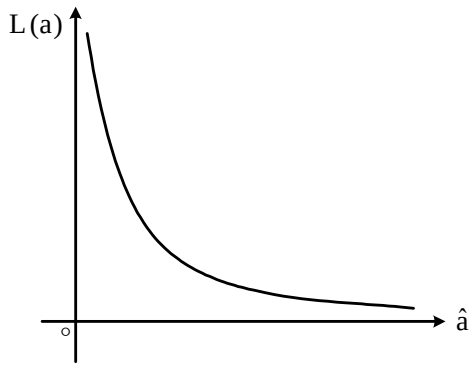
حل: تابع چگالی احتمال X عبارتست از:

$$f_X(x) = \frac{۱}{a} \qquad \text{برای } ۰ < x < a$$

با در نظر گرفتن یک نمونه تصادفی به اندازه n تابع چگالی احتمال توام بصورت زیر بدست می‌آید:

$$L(a) = \frac{۱}{a} \cdot \frac{۱}{a} \cdots \frac{۱}{a} = \frac{۱}{a^n}$$

منحنی نمایش $L(a)$ را در شکل زیر مشاهده می‌کنید.



dalbablog.wordpress.com

از روی منحنی پیداست که شیب منحنی در هیچ نقطه‌ای صفر نمی‌شود بنابراین مقدار مشتق $L(a)$ در هیچ نقطه‌ای صفر نمی‌شود. اما با توجه به نمودار می‌بینیم که هر چقدر a به صفر نزدیک می‌شود مقدار $L(a)$ بیشتر می‌شود، و به عبارتی $L(a)$ به ازای کوچکترین مقدار a ماکزیمم می‌شود، از طرفی کوچکترین مقدار a نباید از بزرگترین نمونه بدست آمده کمتر باشد بنابراین برآورد a به روش حداکثر احتمال در این مثال برابر است با بزرگترین نمونه بدست آمده یعنی nامین آماره مرتب $\hat{a} = X_{(n)}$.

بوضوح این برآورد کاراتر از برآورد به روش گشتاورها می‌باشد.

۱۳ مثال ۶: اگر متغیر تصادفی X یک متغیر تصادفی پواسون با پارامتر λ باشد مطلوبست برآورد λ به روش حداکثر احتمال و به روش گشتاورها؟

حل: تابع چگالی متغیر تصادفی پواسون عبارتست از:

$$f_X(x) = \frac{e^{-\lambda}\lambda^x}{x_i} \qquad x = ۰,۱,۲,\dots$$

$$= ۰ \qquad \text{سایر مقادیر}$$

$$L(\lambda) = \frac{e^{-\lambda}\lambda^{x_۱}}{x_۱!}.\frac{e^{-\lambda}\lambda^{x_۲}}{x_۲!}\cdots\frac{e^{-\lambda}\lambda^{x_n}}{x_n!} = \frac{e^{-n\lambda}\ \lambda^{\sum x_i}}{x_۱!\,x_۲!\cdots x_n!}$$

$$H(\lambda) = Ln(L(\lambda)) = -n\lambda + \sum x_i\ Ln(\lambda) - Ln(x_۱!\,x_۲!\cdots x_n!)$$

$$\frac{dH}{d\lambda} = -n + \sum x_i\,(\frac{۱}{\lambda}) = ۰ \quad \Rightarrow \quad \hat{\lambda} = \frac{۱}{n}\sum x_i = \overline{X}$$

حال این مساله را به روش گشتاورها حل می‌کنیم:

$$m_۱ = \mu = \lambda$$

$$\Rightarrow \quad m_۱ = m'_۱ \ \Rightarrow \ \hat{\lambda} = \overline{X}$$

$$m'_۱ = \frac{۱}{n}\sum x_i = \overline{X}$$

ملاحظه می‌کنید که نتایج هر دو روش در این حالت نیز یکسان می‌باشند.

۱۴ مثال ۷: اگر متغیر تصادفی X یک متغیر تصادفی نرمال با میانگین μ و واریانس $\sigma^۲$ باشد مطلوبست برآورد پارامترهای μ و $\sigma^۲$ به روشهای حداکثر احتمال و گشتاورها؟

حل: تابع چگالی احتمال متغیر تصادفی X عبارتست از:

$$f_X(x) = \frac{۱}{\sigma\sqrt{۲\pi}}\ e^{-\frac{(x-\mu)^۲}{۲\sigma^۲}}$$

$$L(\mu,\sigma^۲) = \prod_{i=۱}^{n}\frac{۱}{\sigma\sqrt{۲\pi}}\ e^{-\frac{(x_i-\mu)^۲}{۲\sigma^۲}} = \frac{۱}{(۲\pi\sigma^۲)^{\frac{n}{۲}}}\ e^{-\frac{۱}{۲}\sum_{i=۱}^{n}\frac{(x_i-\mu)^۲}{\sigma^۲}}$$

$$H(\mu,\sigma^۲) = Ln\,(L((\mu,\sigma^۲)) = Ln\,(\frac{۱}{(۲\pi\sigma^۲)^{\frac{n}{۲}}}) + Ln\ (e^{-\frac{۱}{۲}\sum_{i=۱}^{n}\frac{(x_i-\mu)^۲}{\sigma^۲}})$$

$$= -\frac{n}{۲}\,(n(۲\pi\sigma^۲) - \frac{۱}{۲}\sum_{i=۱}^{n}\frac{(x_i-\mu)^۲}{\sigma^۲}$$

dalbablog.wordpress.com

$$\Rightarrow H(\mu,\sigma^2) = -\frac{n}{2}(n(2\pi) - \frac{n}{2}(n\sigma^2 - \frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}$$

$$\frac{\partial H}{\partial \mu} = -\frac{1}{2}(-2) = \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial H}{\partial \sigma^2} = -\frac{n}{2}(\frac{1}{\sigma^2}) - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\,(\frac{-1}{(\sigma^2)^2})$$

$$= \frac{-n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

حال می‌بایستی معادلات زیر را حل کنیم:

$$\begin{cases}\frac{1}{\hat{\sigma}^2}\sum(x_i-\hat{\mu}) = 0\\ -\frac{n}{2\hat{\sigma}^2}+\frac{1}{2(\hat{\sigma}^2)^2}\sum(x_i-\mu)^2 = 0\end{cases}$$

۱۵ جواب معادله اول عبارتست از:

$$\sum(x_i-\hat{\mu}) = 0 \quad \Rightarrow \quad \sum x_i - \sum\hat{\mu} = 0 \quad \rightarrow \quad \sum x_i - n\hat{\mu} = 0$$

$$\Rightarrow \hat{\mu} = \frac{1}{n}\sum x_i = \overline{X}$$

با قرار دادن $\hat{\mu} = \overline{X}$ در معادله دوم داریم:

$$-\frac{n}{2\hat{\sigma}^2} + \frac{1}{2(\hat{\sigma}^2)^2}\sum(x_i-\overline{X})^2 = 0$$

$$n2\hat{\sigma}^2 + \sum(x_i-\overline{X})^2 = 0 \quad \Rightarrow \quad \hat{\sigma}^2 = \frac{1}{n}\sum(x_i-\overline{X})^2$$

بنابراین برآورد پارامترهای مجهول μ و σ^2 به روش حداکثر احتمال عبارتند از:

$$\hat{\mu} = \overline{X} \quad , \quad \hat{\sigma}^2 = \frac{n-1}{n}S^2$$

۱۶ حال برآورد را به روش گشتاورها بدست می‌آوریم:

$$m_1 = \mu$$

$$m_2 = \sigma^2 + \mu^2$$

$$m_1' = \frac{1}{n}\sum_{i=1}^{n}x_i = \overline{X}$$

$$m_2' = \frac{1}{n}\sum_{i=1}^{n}x_i^2$$

$$\begin{cases}m_1 = m_1'\\ m_2 = m_2'\end{cases} \quad \Rightarrow \quad \begin{cases}\mu = \overline{X}\\ \frac{1}{n}\sum_{i=1}^{n}x_i^2 = \sigma^2+\mu^2\end{cases}$$

از معادله اول داریم $\hat{\mu} = \overline{X}$ با جاگذاری در معادله دوم بدست می‌آید:

dalbablog.wordpress.com

$$\frac{1}{n}\sum_{i=1}^{n} x_i^2 = \hat{\sigma}^2 + \overline{X}^2 \qquad \Rightarrow \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{X}^2$$

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

در نتیجه برآورد پارامترهای μ و σ^2 به روش گشتاورها نیز همان نتایج برآورد به روش حداکثر را بدست می‌دهد.

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

S-۵

۱۶ مثال ۸: اگر متغیر تصادفی X بصورت یکنواخت در بازه $[-b\ ,\ b]$ توزیع شده باشد مطلوبست برآورد b به روش گشتاورها و حداکثر احتمال؟

حل: میانگین متغیر تصادفی X عبارتست از:

$$E[X] = \int_{-b}^{b} dx = \frac{1}{2b}\left(\frac{x^2}{2}\right)\Big|_{-b}^{b} = 0$$

حال طبق روش گشتاورها می‌بایستی از حل معادله $m_1 = m'_1$ پارامتر b بدست آید اما می‌بینیم که:

$$m_1 = E[X] = 0$$

$$m'_1 = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{X}$$

$$m_1 = m'_1 \ \Rightarrow \ \overline{X} = 0$$

همانطور که ملاحظه می‌کنید $\overline{X} = 0$ کمکی در یافتن پارامتر مجهول به ما نمی‌کند. در واقع این مثال حالت خاصی است که در آن جهت تعین پارامتر مجهول می‌بایستی از گشتاور دوم نمونه و متغیر تصادفی X استفاده کنیم:

$$m_2 = E\left[X^2\right] = \int_{-b}^{b} x^2 \frac{1}{2b}\,dx = \frac{1}{2b}\left(\frac{x^3}{3}\right)\Big|_{-b}^{b}$$

$$= \frac{1}{2b}\left(\frac{b^3}{3} + \frac{b^3}{3}\right) = \frac{b^2}{3}$$

به همین ترتیب داریم:

$$m'_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \quad \Rightarrow m_2 = m'_2 \ \Rightarrow \ \frac{b^2}{3} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

$$\Rightarrow b^2 = 3\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) \ \rightarrow \ \hat{b} = \sqrt{3\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)} = \sqrt{3m'_2}$$

بنابراین برآورد پارامتر مجهول b در این مثال برابر $\sqrt{3m'_2}$ می‌باشد.

۱۷ روش برآورد پارامتر b به روش حداکثر احتمال کاملاً مشابه مثال ۵ می‌باشد به این صورت که:

$$f_X(x) = \frac{1}{2b} \qquad -b < x < b$$

$$L(b) = \frac{1}{2b}\cdot\frac{1}{2b}\cdots\frac{1}{2b} = \frac{1}{(2b)^n}$$

مجدداً از آنجا که $L(b)$ با نزدیک شدن به صفر ماکزیمم می‌شود بنابراین کوچکترین مقداری که b می‌تواند بخود بگیرد برآورد آن می‌باشد و از آنجا که b نمی‌تواند از بزرگترین مقدار نمونه کمتر باشد بنابراین در این مثال نیز برآورد b عبارتست از $\max|X_i|$ یا همان nامین آماره مرتب $X_{(n)}$.

dalbablog.wordpress.com

۲.۱.۱۸ خواص برآورد کننده‌ها

در بخش روشهای گشتاورها و حداکثر احتمال را برای برآورد پارامترهای مجهول جامعه معرفی نمودیم و نشان دادیم که این دو روش همواره برآورد کننده‌های مشابه بدست نمی‌دهند، بنابراین نیازمند یک معیاری برای سنجش میزان کارایی یک برآوردگر نسبت به برآوردگر دیگری می‌باشیم، در این بخش به معرفی خواص برآورد کننده‌ها و روشهایی برای سنجش میزان کارایی آنها می‌پردازیم.

۱.۲.۱.۱۹ متوسط مربع خطا (MSE)

فرض کنید $\hat{\theta}$ برآورد پارامتر θ باشد در این صورت می‌دانیم که $\left|\hat{\theta}-\theta\right|$ مقدار خطای برآورد می‌باشد. معمولاً برای تخمین میزان خطا از مربع $(\hat{\theta}-\theta)$ استفاده می‌شود و ار آنجا که در اینجا n نمونه استخراج می‌شود برای هر بار استخراج این n نمونه مقدار $(\hat{\theta}-\theta)^۲$ محاسبه می‌شود و میانگین آن ملاک خواهد بود. به این ترتیب معیار متوسط مربع خطا بصورت زیر بدست می‌آید:

$$MSE\ (\hat{\theta}) = E\left[(\hat{\theta}-\theta)^۲\right]$$

توجه کنید که همواره در عمل با توجه به استخراج n نمونه و برآورد پارامتر θ مقداری خطا در برآورد پارامتر θ وجود دارد، بنابراین برای اینکه بهترین برآورد را داشته باشیم می‌بایستی خطای برآورد را حداقل کنیم. بوضوح اگر $\hat{\theta}=\theta$ باشد خطای برآورد صفر می‌باشد و بهترین حالت بدست آمده است.

برای اینکه بتوان خطای برآورد را حداقل نمود ابتدا مقدار متوسط مربع خطا را ساده می‌کنیم:

$$MSE = E\left[(\hat{\theta}-\theta)^۲\right] = E\left[(\hat{\theta}^۲ -۲\theta\hat{\theta}-\theta)^۲\right]$$
$$= E\left[\hat{\theta}^۲\right] - ۲E\left[\theta\hat{\theta}\right] + E\left[\theta^۲\right]$$

حال یک $E^۲\left[\hat{\theta}\right]$ به سمت راست عبارت بالا اضافه و کم می‌کنیم:

$$\Rightarrow MSE = E\left[\hat{\theta}^۲\right] - E^۲\left[\hat{\theta}\right] + E^۲\left[\hat{\theta}\right] - ۲E\left[\theta\hat{\theta}\right] + E\left[\theta^۲\right]$$

توجه کنید که عبارت A همان واریانس $\hat{\theta}$ می‌باشد. می‌دانیم $\hat{\theta}$ خود یک آماره می‌باشد و از آنجا که آماره‌ها خود یک متغیر تصادفی می‌باشند، بنابراین واریانس $\hat{\theta}$ معنی‌دار می‌باشد.

$$\Rightarrow MSE = var(\hat{\theta}) + (E\left[\hat{\theta}\right]-\theta)^۲$$

بنابراین متوسط مربع خطا بصورت فوق خلاصه می‌شود که در آن عبارت $E\left[\hat{\theta}\right]-\theta$ در واقع اندازه اختلاف مرکز توزیع $\hat{\theta}$ از θ را نشان می‌دهد.

۲۰ برای اینکه MSE حداقل شود می‌بایستی مقادیر $(E\left[\hat{\theta}\right]-\theta)^۲, var(\hat{\theta})$ حداقل شوند. از طرفی می‌دانیم $(E\left[\hat{\theta}\right]-\theta)^۲, var(\hat{\theta})$ هر دو عبارتی همواره مثبت می‌باشند بنابراین MSE در صورتی حداقل می‌شود که این دو عبارت حداقل شوند. حال کمترین مقداری را که این دو عبارت بخود می‌پذیرد را محاسبه می‌کنیم:

$$var(\hat{\theta}) = E\left[\hat{\theta}^۲\right] - E^۲\left[\hat{\theta}\right]$$

اگر $E^۲\left[\hat{\theta}\right]=E\left[\hat{\theta}^۲\right]$ واریانس $\hat{\theta}$ برابر صفر می‌شود که کمترین مقداری است که یک عبارت مثبت بخود می‌پذیرد. بنابراین $\min(var(\hat{\theta})) = ۰$

همینطور حداقل $(E\left[\hat{\theta}\right]-\theta)^۲$ با توجه به مثبت بودن عبارت زمانی رخ می‌دهد که کل عبارت صفر شود یعنی:

$$\min\left[(E\left[\hat{\theta}\right]-\theta)^۲\right] = ۰ \quad \Rightarrow \quad (E\left[\hat{\theta}\right]-\theta)^۲ = ۰ \quad \Rightarrow \quad E\left[\hat{\theta}\right] = \theta$$

می‌دانیم در عمل $var(\hat{\theta})$ هرگز صفر نمی‌شود (مگر اینکه تمام مقادیر نمونه‌ها برابر باشند که آن هم بی‌معنی است). اما رخ دادن تساوی $E\left[\hat{\theta}\right] = \theta$ امکان‌پذیر است بنابراین از میان برآوردگرهای مختلف برای θ برآوردگری مناسب‌تر است که شرایط زیر را داشته باشد.

۱- واریانس برآوردگر $(\hat{\theta})$ کمترین مقدار ممکن را داشته باشد.

۲- $E\left[\hat{\theta}\right] = \theta$ باشد.

شرط دوم به برآورد کننده ناااریب معروف است که به صورت زیر تعریف می‌شود:

dalbablog.wordpress.com

برآورد کننده ناریب: برآورد کننده $\hat{\theta}$ برای θ ناریب نامیده می‌شود اگر $E\left[\hat{\theta}\right] = \theta$.

برابر تساوی $E\left[\hat{\theta}\right] = \theta$ به این معنی است که اگر مثلاً kبار از جامعه تعداد n نمونه استخراج کنیم و به ازای هر بار استخراج n نمونه مقدار $\hat{\theta}$ را محاسبه کنیم و در نهایت میانگین $\hat{\theta}_i$ها ($E\left[\hat{\theta}\right] = \frac{1}{k}\sum_{i=1}^{k}\hat{\theta}_i$) می‌بایستی برابر با پارامتر مجهول θ شود. برای روشن شدن مطلب به مثال زیر توجه کنید.

۲۱ مثال ۹: آیا برآوردگر بدست آمده برای توزیع نرمال، برنولی، پواسون ناریب می‌باشد؟

حل: قبلاً نشان دادیم که برآوردگر μ و σ^2 در توزیع نرمال با استفاده از روشهای گشتاورها و حداکثر احتمال عبارتست از: $\hat{\mu} = \overline{X}$

$\hat{\sigma}^2 = \frac{n-1}{n} S^2$

برای $\hat{\mu}$ داریم:

توجه کنید که قبلاً نشان دادیم برای متغیر تصادفی X با هر توزیعی $E\left[\overline{X}\right]$ برابر μ می‌باشد. $E\left[\hat{\mu}\right] = E\left[\overline{X}\right] = \mu$

بنابراین $\hat{\mu} = \overline{X}$ در توزیع نرمال یک برآورد ناریب برای μ می‌باشد.

حال برای $\hat{\sigma}^2$ داریم:

$$E\left[\hat{\sigma}^2\right] = E\left[\frac{n-1}{n} S^2\right] = \frac{n-1}{n} E\left[S^2\right]$$

۲۲ در فصل قبل نشان دادیم که اگر $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})$ آنگاه داریم: $E\left[S^2\right] = \sigma^2$

بنابراین:

$$E\left[\hat{\sigma}^2\right] = \frac{n-1}{n}\sigma^2$$

ملاحظه می‌کنید که برای $\hat{\sigma}^2$ مقدار $E\left[\hat{\sigma}^2\right]$ برابر با σ^2 نمی‌باشد بنابراین برآوردگر $\hat{\sigma}^2 - \frac{n-1}{n} S^2$ برآوردگر ناریبی برای σ^2 نمی‌باشد. در این حالت می‌گوییم برآوردگر $\hat{\sigma}^2$ یک برآوردگر اریب برای σ^2 می‌باشد.

توجه کنید که $\hat{\sigma}^2$ تنها به دلیل وجود ضریب $(\frac{n-1}{n})$ اریب می‌باشد، به عبارتی تعداد نمونه‌ها عامل اصلی اریب بودن این برآوردگر می‌باشد بنابراین اگر تعداد نمونه‌ها زیاد باشد $(n \to \infty)$ داریم: $\lim_{n\to\infty} (\frac{n-1}{n})\sigma^2 = \sigma^2$

یعنی با شرط $n \to \infty$ برآوردگر $\hat{\sigma}^2$ یک برآوردگر ناریب برای σ^2 می‌باشد. به این نوع برآوردگرها، برآوردگر ناریب مجانبی می‌گویند که بصورت زیر تعریف می‌شود:

برآوردگر ناریب مجانبی: اگر $\hat{\theta}$ یک برآوردگر اریب برای θ بر اساس نمونه تصادفی nتایی باشد، می‌گوییم $\hat{\theta}$ یک برآوردگر ناریب مجانبی برای θ است هرگاه داشته باشیم:

$$\lim_{n\to\infty} E\left[\hat{\theta}\right] = \theta$$

۲۳ برای توزیع برنولی قبلاً بدست آوردیم که:

$$\hat{p} = \overline{X} \quad \Rightarrow \quad E\left[\hat{p}\right] = E\left[\overline{X}\right] = \mu = p$$

همینطور برای توزیع پواسون داریم:

$$\hat{\lambda} = \overline{X}$$
$$E\left[\hat{\lambda}\right] = E\left[\overline{X}\right] = \mu = \lambda$$

یک نتیجه کلی که از برآوردگر ناریب بدست می‌آوریم این است که هرگاه پارامتر مجهول توزیع برابر با میانگین توزیع متغیر تصادفی X باشد

dalbablog.wordpress.com

(یعنی $\mu = E[X] = \theta$) همینطور از آنجا که برای هر توزیعی مستقل از نوع توزیع داریم $E[\overline{X}] = \mu$ در این صورت $\overline{X}$ همواره به عنوان یک برآوردگر نااریب این گونه از توزیعها، منظور می‌شود.

خاصیت نااریبی اگر چه یک معیار برای محک بهتر بودن یک برآوردگر می‌باشد اما از آنجا که برای یک پارامتر می‌توان چندین برآوردگر نااریب بدست آورد، بنابراین می‌بایستی معیارهای دیگری وجود داشته باشند تا بتوان با کمک آنها از میان برآوردگرهای نااریب بهترین را انتخاب نمود. برای این منظور معیار کارایی یا موثرتر بودن را معرفی می‌کنیم:

تعریف: اگر $\hat{\theta}_1, \hat{\theta}_2$ هر دو برآوردگرهایی نااریب برای پارامتر θ باشند در این صورت می‌گوییم $\hat{\theta}_1$ کاراتر یا موثرتر از $\hat{\theta}_2$ می‌باشد هرگاه:

$$var(\hat{\theta}_1) < var(\hat{\theta}_2)$$

علت استفاده از این معیار با توجه به تعریف متوسط مربع خطا (MSE) کاملاً قابل توجیه است.

۲۴ مثال ۱۰: از یک جامعه نرمال با میانگین μ و واریانس σ^2 یک نمونه به حجم n انتخاب شده است. در این صورت کدامیک از برآوردگرهای نااریب زیر معیار مناسب‌تری برای برآورد پارامتر μ می‌باشد؟

الف) $\hat{\mu}_1 = \overline{X}$

ب) $\hat{\mu}_2 = \frac{1}{3}(X_1 + X_3 + X_5)$

حل: با توجه به اینکه هر دو برآوردگر نااریب می‌باشند بنابراین واریانس آنها را محاسبه نموده و با یکدیگر مقایسه می‌کنیم:

$$var(\hat{\mu}_1) = var(\overline{X}) = \frac{\sigma^2}{n}$$

$$var(\hat{\mu}_2) = var\left(\frac{1}{3}(X_1 + X_3 + X_5)\right) = \frac{1}{9}(var(X_1) + var(X_3) + var(X_5))$$

$$= \frac{1}{9}(3\sigma^2) = \frac{\sigma^2}{3}$$

با مقایسه دو واریانس بدست آمده داریم:

$$\frac{\sigma^2}{n} < \frac{\sigma^2}{3} \qquad \text{برای } n > 3$$

بنابراین اگر تعداد نمونه‌های استخراج شده از متغیر تصادفی X بیستر از ۳ باشد خواهیم داشت:

$$var(\hat{\mu}_1) < var(\hat{\mu}_2)$$

و در نتیجه با توجه به معیار کارایی می‌توان نتیجه گرفت که برآورد کننده $\overline{X}$ بهتر از $\frac{1}{3}(X_1 + X_3 + X_5)$ در این مثال می‌باشد.

تعریف: برای دو برآوردگر $\hat{\theta}_1, \hat{\theta}_2$ نسبت $\frac{var(\hat{\theta}_1)}{var(\hat{\theta}_2)}$ را کارایی نسبی برآوردگر $\hat{\theta}_2$ نسبت به برآوردگر $\hat{\theta}_1$ می‌نامند.

بوضوح اگر مقدار کسر از یک بیشتر باشد می‌گوییم برآوردگر $\hat{\theta}_2$ از برآوردگر $\hat{\theta}_1$ کاراتر یا موثرتر است.

S-۷

۲۵ مثال ۱۱: از یک جامعه نمونه‌ای به حجم ۲۰ با میانگین $\overline{X}_1 = ۱۵$ و واریانس $\sigma_1^2 = ۴$ استخراج کرده‌ایم. همچنین نمونه دیگری به حجم ۳ با میانگین $\overline{X}_2 = ۱۸۰$ و واریانس $\sigma_2^2 = ۲$ استخراج کرده‌ایم. اگر $\overline{X}$ را به عنوان برآوردگر میانگین جامعه انتخاب کنیم کارایی نسبی $\overline{X}_1$ را محاسبه کنید؟

حل: ابتدا واریانس $\overline{X}_1$ و $\overline{X}_2$ را محاسبه می‌کنیم:

dalbablog.wordpress.com

$n_۱ = ۲۰ \ , \quad \overline{X}_۱ = ۱۵ \quad , \quad \sigma_۱^۲ = ۴$

$n_۲ = ۳۰ \ , \quad \overline{X}_۲ = ۱۸ \quad , \quad \sigma_۱^۲ = ۲$

$$var(\overline{X}_۱) = \frac{\sigma_۱^۲}{n_۱} = \frac{۴}{۲۰} = \frac{۱}{۵}$$

$$var(\overline{X}_۲) = \frac{\sigma_۲^۲}{n_۲} = \frac{۲}{۳۰} = \frac{۱}{۱۵}$$

$$\overline{X}_۲ \text{ نسبت به } \overline{X}_۱ \text{ نسبی کارایی} = \frac{var(\overline{X}_۲)}{var(\overline{X}_۱)} = \frac{\frac{۱}{۱۵}}{\frac{۱}{۵}} = \frac{۵}{۱۵} = \frac{۱}{۳}$$

بنابراین استفاده از نمونه متوسط استخراج شده دوم برای برآورد میانگین جامعه کارایی بیشتری دارد و موثرتر می‌باشد. همچنین توجه کنید که در محاسبه کارایی از متوسط دو نمونه استخراج شده استفاده‌ای نمی‌شود.

۲۶ تا بحال نشان دادیم که یک برآوردگر به شرطی قابل قبول است که در درجه اول ناریب باشد و همچنین کمترین واریانس را در میان سایر برآوردگرها داشته باشد بنابراین تعریف زیر را داریم:

تعریف: اگر $\hat{\theta}$ یک برآورد کننده ناریب برای θ باشد. در این صورت $\hat{\theta}$ بهترین برآورد کننده ناریب خطی برای θ می‌باشد. هرگاه:

۱- $\hat{\theta}$ تابعی خطی از $X_۱, X_۲, \ldots, X_n$ باشد.

۲- برای هر θ $\quad E\left[\hat{\theta}\right] = \theta$.

۳- از میان تمام برآورد کننده‌های $\hat{\theta}_i$ که در شرایط ۱ و ۲ صدق می‌کند داشته باشیم:

$$var\ (\hat{\theta}) \ \leq \ var\ (\hat{\theta}_i)$$

با توجه به استفاده از روش متوسط مربع خطا (MSE) تعریف بالا و شرایط آن کاملاً موجه می‌باشد. همچنین توجه کنید که اگر $\hat{\theta}$ در شرایط فوق صدق کند در این صورت $\hat{\theta}$ موثرترین برآورد کننده ناریب خطی نیز می‌باشد.

حال به مثال زیر توجه کنید:

۲۷ مثال ۱۲: فرض کنید X یک متغیر تصادفی با میانگین μ و واریانس $\sigma^۲$ باشد. هرگاه $X_۱, X_۲, \ldots, X_n$ یک نمونه تصادفی از X باشد نشان دهید توزیع متغیر تصادفی X هر چه باشد، برآورد کننده $\overline{X}$ بهترین برآورد کننده ناریب خطی برای μ می‌باشد؟

حل: ابتدا فرض می‌کنیم $\hat{\theta}$ یک برآورد کننده ناریب خطی برای μ باشد در این صورت $\hat{\theta}$ می‌بایستی به فرم زیر باشد:

$$\hat{\theta} = \sum_{i=۱}^{n} (a_i \ x_i) + b$$

حال متغیرهای a_i و b را طوری بدست می‌آوریم که $\hat{\theta}$ بهترین برآورد کننده برای μ باشد. برای این منظور مطابق تعریف $\hat{\theta}$ می‌بایستی در شرایط تعریف بهترین برآورد کننده ناریب خطی صدق کند. یعنی: $E\left[\hat{\theta}\right] = \mu$ بنابراین:

dalbablog.wordpress.com

$$E\left[\hat{\theta}\right]= E\left[\sum_{i=1}^{n}(a_i\ x_i)+b\right]\ =\ E\left[\sum_{i=1}^{n}a_i\ x_i\right]+b$$

$$=E\sum_{i=1}^{n}(a_i\,E[X_i])+b=\sum_{i=1}^{n}(a_i\ m_1)+b$$

$$=m_1\sum_{i=1}^{n}a_i+b=E[X]\sum_{i=1}^{n}a_i+b \quad \Rightarrow E\left[\hat{\theta}\right]=\mu\sum_{i=1}^{n}a_i+b$$

حال چون بایستی داشته باشیم $E\left[\hat{\theta}\right]=\nu$ بنابراین:

$$\mu\sum_{i=1}^{n}a_i+b=\mu \quad \Rightarrow \quad \begin{cases}\sum_{i=1}^{n}a_i=1\\ b=0\end{cases}$$

۲۸ حال می‌بایستی واریانس برآورد کننده $\hat{\theta}$ را بدست آورده و سپس آنرا حداقل کنیم:

$$var(\hat{\theta})=var(\sum_{i=1}^{n}a_i\,X_i\ +b)=var\ (\sum a_i\,X_i)$$

$$=\sum var(a_i\,X_i)\ =\ \sum a_i^{2}\ var(X_i)=\ \sum a_i^{2}\ \sigma^{2}=\sigma^{2}\sum_{i=1}^{n}a_i^{2}$$

بنابراین برای حداقل نمودن واریانس $\sum a_i^{2}$ می‌بایستی حداقل شود به عبارتی باید مقادیر $a_1,a_2,\ldots,a_n$ را طوری پیدا نمود که در دو شرط زیر صدق کنند:

۱- $\sum_{i=1}^{n}a_i=1$

۲- $\sum a_i^{2}$ مینیمم شود.

برای حداقل نمودن $\sum a_i^{2}$ ابتدا مقدار a_1 را از شرط اول محاسبه نموده و حاصل را در شرط دوم $(\sum a_i^{2})$ قرار می‌دهیم. به صورت زیر:

$$\sum_{i=1}^{n}a_i=1 \qquad \Rightarrow \qquad a_1+\sum_{i=2}^{n}a_i=1 \qquad \Rightarrow \qquad a_1=1-\sum_{i=2}^{n}a_i$$

فرض می‌کنیم $Q=\sum_{i=1}^{n}a_i^{2}$ در این صورت باید Q مینیمم شود پس:

$$Q=\sum_{i=1}^{n}a_i^{2}=a_1^{2}+\sum_{i=2}^{n}a_i=(1-\sum_{i=2}^{n}a_i)^{2}+\sum_{i=2}^{n}a_i^{2}$$

$$\Rightarrow Q=(1-\sum_{i=2}^{n}a_i)^{2}+\sum_{i=2}^{n}a_i^{2}$$

حال برای حداقل نمودن Q می‌بایستی مشتقات پاره‌ای Q را نسبت به a_iها محاسبه نموده و برابر صفر قرار دهیم تا مقادیری بشکل $a_1^{*},a_2^{*},\ldots,a_n^{*}$ که Q را مینیمم می‌کنند بدست آید:

$$\frac{\partial Q}{\partial a_j}=2(-1)\ (1-\sum_{i=1}^{n}a_i)+2a_j \qquad j=2,3,\cdots,n$$

dalbablog.wordpress.com

۲۹ توجه کنید که مقدار a_1^* از روی سایر a_i^*ها $(i=2,3,\cdots,n)$ بشکل زیر بدست می‌آید:

$$a_1^* = 1 - \sum_{i=2}^{n} a_i^*$$

رابطه ۱-۱۰

$$\frac{\partial Q}{\partial a_j} = 0 \quad \Rightarrow \quad \begin{cases} a_j^* = (1 - \sum_{i=2}^{n} a_i^*) = a_1^* \\ a_1^* = 1 - \sum_{i=2}^{n} a_i^* \end{cases}$$

حال با حل دستگاه فوق مقادیر a_j^*ها بدست می‌آید، برای حل ابتدا کل معادلات را جداگانه نوشته و طرفین معادلات را جمع می‌کنیم. به این ترتیب داریم:

$$a_1^* = 1 - \sum_{i=2}^{n} a_i^*$$

$$a_2^* = 1 - \sum_{i=2}^{n} a_i^*$$

$$a_3^* = 1 - \sum_{i=2}^{n} a_i^*$$

$$\vdots$$

$$a_n^* = 1 - \sum_{i=2}^{n} a_i^*$$

$$\text{حاصل جمع معادلات} = \sum_{i=1}^{n} a_i^* = n - n \sum_{i=2}^{n} a_i^* = n \underbrace{(1 - \sum_{i=2}^{n} a_i^*)}_{a_i^*} = n a_1^*$$

بنابراین بدست می‌آید:

$$\begin{cases} \sum_{i=1}^{n} a_i^* = n a_1^* \\ \sum_{i=1}^{n} a_i^* = 1 \end{cases} \quad \Rightarrow \quad n a_1^* = 1 \Rightarrow a_1^* = \frac{1}{n}$$

و با توجه به رابطه ۱-۱۰ بدست می‌آید:

$$(j=1,2,\cdots,n) \quad a_j^* = \frac{1}{n}$$

بنابراین $\sum_{i=1}^{n} a_i^2$ و با شرط $\sum_{i=1}^{n} a_i = 1$ به ازای $a_j^* = \frac{1}{n}$ حداقل می‌شود. بنابراین برآوردگر $\hat{\theta}$ برابر می‌شود با: $\hat{\theta} = \sum_{i=1}^{n} \frac{1}{n} X_i = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}$

در نتیجه $\overline{X}$ بهترین برآوردگر ناریب خطی برای μ می‌باشد و این مساله مستقل از توزیع متغیر تصادفی X می‌باشد. به همین دلیل این مطلب از اهمیت فراوانی برخوردار است از طرفی می‌توان نشان داد که اگر برآوردگر $\hat{\theta}$ تابعی خطی از X_iها هم نباشد باز هم $\overline{X}$ دارای کوچکترین واریانس می‌باشد. حال می‌توان گفت که $\overline{X}$ برای پارامتر μ در تابع نرمال و برای پارامتر p در توزیع برنولی و برای پارامتر λ در توزیع پواسون بهترین برآوردگر ناریب می‌باشد.

dalbablog.wordpress.com

۳۰ . ۱۰ ۳ برآورد کننده سازگار

یکی دیگر از خواصی که برای برآورد کننده‌ها می‌توان بدست آورد خاصیتی است که به سازگاری معروف است. اگر داشته باشیم:

$$\lim_{n\to\infty} \quad MSE\ (\hat{\theta}) = \lim E\left[(\hat{\theta} - \theta)^{۲}\right] = ۰$$

یعنی به ازای تعداد زیادی نمونه از متغیر تصادفی X مقدار متوسط مربع خطا برابر صفر شود، از طرفی طبق نامساوی مارکف (رجوع کنید به ثضیه چبی شف) داریم:

$$p(g(x) \geq k) \leq \frac{E[g(x)]}{k} \Rightarrow ۱-p(g(x) \leq k) \leq \frac{E[g(x)]}{k}$$

$$\Rightarrow p(g(x) \geq k) \geq ۱ - \frac{E[g(x)]}{k}$$

حال اگر $k = \varepsilon^{۲}\ ,\ (\varepsilon > ۰)$ $g(x) = (\hat{\theta}-\theta)^{۲}$ باشد در این صورت نامساوی بصورت زیر خواهد بود:

$$p((\hat{\theta}-\theta)^{۲} \leq \varepsilon^{۲}) \geq ۱-\frac{E\left[(\hat{\theta}-\theta)^{۲}\right]}{\varepsilon^{۲}}$$

$$\Rightarrow p(|\hat{\theta}-\theta| < \varepsilon) \geq ۱-\frac{E\left[(\hat{\theta}-\theta)^{۲}\right]}{\varepsilon^{۲}}$$

حال اگر شرط $\lim_{n\to\infty} MSE = ۰$ برقرار باشد می‌بایستی $E\left[(\hat{\theta}-\theta)^{۲}\right]$ برابر صفر باشد در این صورت نامساوی بصورت زیر خلاصه می‌شود:

$$p(\left|\hat{\theta}-\theta\right| < \varepsilon) \geq ۱ - \frac{۰}{\varepsilon^{۲}} \Rightarrow p(\left|\hat{\theta}-\theta\right| < \varepsilon) \geq ۱$$

و از آنجا که احتمال حداکثر برابر ۱ واحد می‌باشد بنابراین:

$$p(\left|\hat{\theta}-\theta\right| < \varepsilon) = ۱ \qquad \forall \varepsilon > ۰$$

یا به بیان دیگر:

$$p(\left|\hat{\theta}-\theta\right| \geq \varepsilon) = ۰ \qquad \forall \varepsilon > ۰$$

در صورتی که احتمالات فوق برای یک برآورد کننده $\hat{\theta}$ برقرار باشند می‌گوییم برآورد کننده دارای خاصیت سازگاری می‌باشد.
بوضوح خاصیت سازگاری زمانی که تعداد نمونه‌ها به اندازه کافی بزرگ باشد نشان دهنده برتری یک برآورد کننده نسبت به دیگری می‌باشد. اما در حالتی که تعداد نمونه‌ها کم باشد، سازگار بودن یک برآورد کننده معیاری برای بهتر بودن آن نمی‌باشد.
قبل از ارایه یک مثال به قضیه زیر نیز توجه کنید:

۳۱ قضیه: اگر داشته باشیم: $\lim_{n\to\infty} E\left[\hat{\theta}\right] = \theta$ و $\lim_{n\to\infty} var\left[\hat{\theta}\right] = ۰$ آنگاه $\hat{\theta}$ یک برآورد کننده سازگار برای θ می‌باشد. این قضیه از آنجا نتیجه می‌شود که در تعریف سازگاری داشتیم:

$$\lim_{n\to\infty} MSE = ۰ \quad \Rightarrow \quad \lim_{n\to\infty} E[(\hat{\theta}-\theta)^{۲}] = ۰$$

$$\Rightarrow \lim_{n\to\infty} (var(\hat{\theta}) + (E[\hat{\theta}] - \theta)^{۲}) = ۰ \quad \Rightarrow \quad \begin{cases} \lim_{n\to\infty} var(\hat{\theta}) = ۰ \\ \lim_{n\to\infty} E[\hat{\theta}] = \theta \end{cases}$$

۳۲ مثال ۱۳: اگر X یک متغیر تصادفی با میانگین μ و واریانس $\sigma^{۲}$ باشد نشان دهید توزیع X هر چه باشد $\overline{X}$ یک برآورد کننده سازگار از روی نمونه‌های $X_۱, X_۲, \ldots, X_n$ برای μ می‌باشد؟

dalbablog.wordpress.com

حل: می‌دانیم:

$$E[\overline{X}] = \mu$$

$$var(\overline{X}) = \frac{\sigma^2}{n}$$

به این ترتیب:

$$\lim_{n\to\infty} E[X] = \lim_{n\to\infty} \mu = \mu$$

$$\lim_{n\to\infty} var(\overline{X}) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0$$

در نتیجه $\overline{X}$ یک برآورد کننده سازگار برای μ می‌باشد.

۳۳ مثال: اگر نمونه‌های $X_1, X_2, \ldots, X_n$ را از یک جامعه نرمال با میانگین μ و واریانس σ^2 اسخراج کنیم در این صورت نشان دهید که S^2 یک برآورد کننده سازگار برای σ^2 می‌باشد؟

حل: قبلاً نشان دادیم که $\frac{(n-1)S^2}{\sigma^2}$ یک متغیر تصادفی خی دو با $n-1$ درجه آزادی است (χ^2_{n-1}) و بنابراین میانگین آن $n-1$ و واریانس آن $2(n-1)$ می‌باشد:

$$E\left[\frac{(n-1)S^2}{\sigma^2}\right] = n-1$$

$$var\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1)$$

حال با ساده نمودن روابط فوق داریم:

$$\frac{n-1}{\sigma^2} E[S^2] = n-1 \quad \Rightarrow \quad E[S^2] = \frac{n-1}{n-1}\sigma^2 \quad \Rightarrow \quad E[S^2] = \sigma^2$$

$$var\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1) \quad \Rightarrow \quad \frac{(n-1)^2}{\sigma^4} var(S^2) = 2(n-1)$$

$$var(S^2) = \frac{2\sigma^4}{n-1} \quad \Rightarrow \quad \lim_{n\to\infty} var(S^2) = \lim_{n\to\infty} \frac{2\sigma^4}{n-1} = 0$$

بنابراین S^2 یک برآورد کننده سازگار برای σ^2 می‌باشد.

S-۹

۳۴ .۱۰ .۴ آماره کافی

اگر بتوان یک آماره یافت که تمام اطلاعات را در بازه یک پارامتر مجهول خلاصه کند آنگاه آنرا یک آماره کافی برای پارامتر مجهول می‌نامیم. آماره‌های کافی به صورت زیر تعریف می‌شود:

تعریف: اگر نمونه تصادفی $X_1, X_2, \ldots, X_n$ را از یک متغیر تصادفی با پارامتر مجهول θ داشته باشیم در این صورت آمارهٔ $y=g(X_1, X_2, .., X_n)$ را یک آماره کافی برای θ می‌نامیم اگر و فقط اگر برای هر مقدار $Y=y$ توزیع مشروط هر آماره بفرم $h(X_1, X_2, .., X_n)$ و با شرط $Y=y$ به θ وابسته نباشد.

به این معنی که اگر مقدار آماره کافی معلوم باشد $(Y=y)$ آنگاه به خود مقادیر نمونه نیازی نیست و مقادیر نمونه اطلاعات بیشتری در بازه پارامتر مجهول θ نسبت به آماره کافی بدست آمده، نمی‌دهند.

dalbablog.wordpress.com

۳۵ مثال ۱۴: فرض می‌کنیم $X_۱, X_۲$ یک نمونه از متغیر تصادفی برنولی با پارامتر مجهول p باشند در این صورت نشان دهید که $Y=X_۱+X_۲$ یک آماره کافی برای p می‌باشد؟

حل: تابع احتمال متغیر تصادفی برنولی بفرم زیر است:

$$p_X(x) = p^x (۱-p)^{۱-x} \qquad x=۰,۱$$
$$= ۰ \qquad \text{سایر مقادیر}$$

حال تعریف می‌کنیم: $Y=X_۱+X_۲=g(X_۱,X_۲)$ باید نشان دهیم اگر $V=h(X_۱,X_۲)$ یک آماری دیگر باشد در این صورت $P_{V,Y}(V|Y=y)$ به θ وابسته نمی‌باشد. برای این منظور ابتدا لازم است توزیع توام V و Y را محاسبه کنیم که عبارتست از:

$$P_{V,Y}(V=h(۰,۰), Y=۰) = (۱-p)(۱-p) = (۱-p)^۲$$
$$P_{V,Y}(V=h(۰,۱), Y=۱) = p(۱-p)$$
$$P_{V,Y}(V=h(۱,۰), Y=۰) = (۱-p)p$$
$$P_{V,Y}(V=h(۱,۱), Y=۲) = p.p = p^۲$$

حال توجه کنید که توزیع $Y=X_۱+X_۲$ یک توزیع دو جمله‌ای با پارامتر $n=۲$ و p می‌باشد بنابراین: $P_Y(y) = \binom{۲}{y} p^y (۱-p)^{۲-y}$.

مقدار توزیع مشروط برای هر یک از مقادیر $Y=y$ عبارتست از:

$$P_{V,Y}(V=h(۰,۰) \mid Y=۰) = \frac{(۱-p)^۲}{\binom{۲}{۰} p^۰ (۱-p)^{۲-۰}} = ۱$$

$$P_{V,Y}(V=h(۰,۱) \text{ یا } h(۱,۰) \mid Y=۱) = \frac{p(۱-p)}{\binom{۲}{۱} p(۱-p)} = \frac{۱}{۲}$$

$$P_{V,Y}(V=h(۱,۱) \mid Y=۲) = \frac{p^۲}{\binom{۲}{۲} p^۲ (۱-p)^۰} = ۱$$

بنابراین اگر V هر آماره دلخواهی باشد، تابع احتمال مشروط آن به p وابسته نمی‌باشد در نتیجه $Y=X_۱+X_۲$ یک آماره کافی برای پارامتر مجهول p می‌باشد.

۳۶ محاسبه کافی بودن یک آماره با توجه به تعریف کار بسیار مشکلی است، اما با استفاده از روش تجزیه فیشر- نیمن یک راه نسبتاً ساده برای اثبات کافی بودن یک آماره وجود دارد:

قضیه: $\hat{\theta}$ را یک برآورگر کافی θ می‌نامیم اگر و فقط اگر تابع چگالی احتمال توام نونه‌ها را بتوان بفرم زیر تجزیه کرد:

$$f(x_۱, x_۲, \ldots, x_n . \theta) = \prod_{i=۱}^{n} f_X(x_i) = g(\hat{\theta}, \theta) . h(x_۱, x_۲, \ldots, x_n)$$

مثال ۱۵: اگر یک نمونه nتایی از یک متغیر تصادفی برنولی داشته باشیم نشان دهید $Y = \sum_{i=۱}^{n} X_i$ یک آماره کافی برای p می‌باشد؟

حل:

dalbablog.wordpress.com

$$\prod_{i=1}^{n} P_X(x_i) = \prod_{i=1}^{n} \left[p^{x_i} (1-p)^{1-x_i} \right] = p^{\sum x_i} (1-p)^{n-\sum x_i}$$

حال داریم:

$$\begin{cases} g(\hat{\theta}, \theta) = p^{\sum x_i} (1-p)^{1-x_i} \\ h(x_1, x_2, \ldots, x_n) = 1 \end{cases}$$

بنابراین $Y = \sum_{i=1}^{n} X_i$ یک آماره کافی برای p می‌باشد.

۳۷ توجه کنید هرگاه $\hat{\theta}$ یک برآورد کننده نااریب برای θ بر حسب یک آماره کافی باشد آنگاه می‌توان نشان داد که $\hat{\theta}$ دارای کمترین واریانس در میان سایر برآورد کننده‌های نااریب برای θ می‌باشد. به همین دلیل است که برآورد کننده‌ها را بر پایه آماره‌های کافی بدست می‌آورند.

قبلاً نشان دادیم که $(\frac{n-1}{n}) S^2$ یک برآورد کننده نااریب برای σ^2 متغیر تصادفی نرمال می‌باشد، حال با استفاده از برآورد کننده کافی یک برآورد کننده نااریب و کافی برای متغیر تصادفی نرمال بدست می‌آوریم.

برای متغیر تصادفی نرمال با میانگین μ و واریانس σ^2 داریم:

$$\prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \left(\frac{1}{\sigma\sqrt{2\pi}}\right) e^{\frac{-(x_i-\mu)^2}{2\sigma^2}} = \left(\frac{1}{\sigma}\right)^n \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} e^{\frac{-\sum(x_i-\mu)^2}{2\sigma^2}}$$

$$= g(\hat{\theta}, \theta) \cdot h(x_1, x_2, \cdots, x_n)$$

که در آن:

$$h(x_1, x_2, \cdots, x_n) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}}$$

از طرفی به جای μ نیز می‌توان از برآورد کافی آن $\overline{X}$ استفاده نمود. در نتیجه $\sum (x_i - \overline{X})^2$ یک آماره کافی برای σ^2 می‌باشد. از طرفی قبلاً نشان دادیم که:

$$E\left[\frac{1}{n} \sum (x_i - \overline{X})^2 \right] = \frac{n-1}{n} \sigma^2$$

$$\Rightarrow E\left[\frac{n-1}{n} \left(\underbrace{\frac{1}{n-1} \sum (x_i - \overline{X})^2}_{S^2} \right) \right] = E\left[\frac{n-1}{n} S^2 \right] = \frac{n-1}{n} \sigma^2$$

$$\Rightarrow \frac{n-1}{n} E[S^2] = \frac{n-1}{n} \sigma^2 \quad \Rightarrow \quad E[S^2] = \sigma^2$$

بنابراین خود S^2 یک برآورد کننده نااریب و کافی برای σ^2 می‌باشد. و دارای کوچکترین واریانس در میان چنین برآورد کننده‌هایی می‌باشد.

۳۸ . ۱۰ ۵ خاصیت پایایی برآورد کننده‌ها

یکی از نتایج کلی که از برآورد کننده‌های حداکثر احتمال بدست می‌آید خاصیت پایایی برآورد کننده حداکثر احتمال است. اگر $\hat{\theta}$ برآورد کننده حداکثر احتمال θ باشد و بخواهیم تابعی از θ مثل $\gamma = h(\theta)$ را برآورد کنیم کافیست به جای θ مقدار برآورد شده آنرا که $\hat{\theta}$ می‌باشد جایگزین کنیم در این صورت برآورد γ عبارتست از: $\hat{\gamma} = h(\hat{\theta})$

این مطلب نشان می‌دهد که برای محاسبه برآورد تابعی از پارامتر مجهول نیازی به محاسبه مجدد برآورد کننده نمی‌باشد بلکه کافیست یکبار برآورد برای پارامتر θ بدست آید. توجه کنید که این مطلب فقط برای برآورد کننده‌هایی که به روش حداکثر احتمال بدست می‌آیند صادق می‌باشد.

dalbablog.wordpress.com

مثال ۱۶: اگر X یک متغیر تصادفی هندسی نوع اول باشد مطلوبست:

الف) برآورد پارامتر p به روش حداکثر احتمال.

ب) برآورد تابع $h(p) = p\,(X \le ۲)$ با استفاده از خاصیت پایایی.

حل: تابع احتمال توزیع هندسی نوع اول عبارتست از:

$$P_X(x) = p\,(۱-p)^x \qquad x = ۰,۱,۲,\cdots$$
$$= ۰ \qquad \text{سایر مقادیر}$$

$$L(p) = \prod_{i=۱}^{n} p(۱-p)^{x_i} = p^n\,(۱-p)^{\sum x_i}$$

$$H(p) = Ln(L(p)) = n\,Ln(p) + \sum x_i\,Ln(۱-p)$$

$$\frac{dH}{dP} = \frac{n}{p} - \frac{\sum x_i}{۱-\hat{p}} \quad \Rightarrow \quad \frac{۱-\hat{p}}{\hat{p}} = \frac{۱}{n}\sum x_i = \overline{X} \quad \Rightarrow \quad \frac{۱}{\hat{p}} - ۱ = \overline{X} \quad \Rightarrow \quad \hat{p} = \frac{۱}{۱+\overline{X}}$$

بنابراین برآورد p به روش حداکثر احتمال برابر $\frac{۱}{۱+\overline{X}}$ می‌باشد.

ب)

$$p(X \le ۲) = p(X=۰) + p(X=۱) + p(X=۲)$$
$$= p + p(۱-p) + p(۱-p)^۲$$
$$\gamma = h(p) = p(۱ + (۱-p) + (۱-p)^۲)$$

حال داریم:

بنابراین خاصیت پایایی $\hat{\gamma} = h(\hat{p})$ بنابراین:

$$\hat{\gamma} = h(\hat{p}) = \frac{۱}{۱+\overline{X}}\left(۱ + (۱ - \frac{۱}{۱+\overline{X}}) + (۱ - \frac{۱}{۱+\overline{X}})^۲\right)$$

S-۱۰

۴۰ ۱۰ . ۶ برآورد فاصله‌ای

در برآورد نقطه‌ای با توجه به نمونه‌های بدست آمده نشان دادیم برای پارامترهای مجهول جامعه می‌توان عددی بدست آورد که با خطای کمی بیانگر مقدار واقعی پارامتر مجهول باشد. اما واقعاً نمی‌دانیم با چه مقدار خطایی پارامتر مجهول را برآورد کرده‌ایم برآورد فاصله‌ای روشی است که در آن برای پارامتر مجهول یک بازه در نظر می‌گیریم و نشان می‌دهیم که با احتمالی معین پارامتر مجهول در بازه مورد نظر قرار دارد.

فرض کنید نمونه‌های $X_۱, X_۲, \ldots, X_n$ را از یک متغیر تصادفی X بدست آورده باشیم در این صورت فاصله $(L_۱, L_۲)$ را طوری می‌سازیم که با احتمال بسیار زیاد مثلاً ۹۵٪ شامل پارامتر مجهول جامعه (مثلاً μ) باشد. به فاصله $(L_۱, L_۲)$ یک فاصله اطمینان می‌گویند زیرا مطمئن هستیم که به احتمال زیاد (مثلاً ۹۵٪) این فاصله شامل پارامتر مجهول می‌باشد. مقادیر $L_۱, L_۲$ هر کدام آماره‌ای هستند که بر اساس مقادیر معلوم نمونه‌ها بدست می‌آیند. حال به تعریف دقیق فاصله اطمینان توجه کنید:

هرگاه مقادیر $X_۱, X_۲, \ldots, X_n$ تعداد n نمونه تصادفی بدست آمده از یک متغیر تصادفی X با پارامتر مجهول θ باشند. در این صورت می‌گوییم دو آماره $L_۱, L_۲$ یک $(۱-\alpha)$ ۱۰۰ درصد فاصله اطمینان برای θ تشکیل می‌دهند. هرگاه:

$$p(L_۱ \le \theta \le L_۲) \ge ۱-\alpha$$

توجه کنید که $۱-\alpha$ مقدار احتمال می‌باشد و بنابراین در بازه ۰ تا ۱ قرار دارد، در نتیجه برای بیان آن بصورت درصد آنرا درصد ضرب کرده‌ایم و به صورت $(۱-\alpha)$ ۱۰۰٪ می‌نویسیم.

در برآورد فاصله‌ای از آماره‌های بدست آمده از فصل قبل استفاده می‌کنیم. اینکه کدام پارامتر جامعه مجهول باشد، تعین می‌کند که از کدام آماری در ساختن فاصله اطمینان استفاده کنیم. مثال بعد روش ساختن فاصله اطمینان برای جامعه نرمال را نشان می‌دهد.

dalbablog.wordpress.com

۴۱ مثال ۱۷: در شهر A بارش باران یک متغیر تصادفی نرمال با میانگین μ سانتی‌متر و واریانس ۴ می‌باشد اگر میزان بارش را در طول ۱۶ روز اندازه بگیریم و میانگین ۱۰ سانتی‌متر را بدست بیاوریم، در این صورت با احتمال ۹۵٪ میانگین بارش باران در شهر A در چه فاصله اطمینانی قرار می‌گیرد؟

حل: معلومات مساله عبارتند از:

$n = ۱۶$

$\overline{X} = ۱۰$

$\mu =$ مجهول

$\sigma^۲ = ۴ \rightarrow \sigma = ۲$

ابتدا توجه کنید که با توجه به جدول متغیر تصادفی نرمال استاندارد رابطه زیر برقرار می‌باشد:

$p(-۱/۹۶ \le Z \le ۱/۹۶) = ۰/۹۵$ (رابطه ۱۰-۲)

زیرا داریم:

$N_Z(۱/۹۶) = p(Z \le ۱/۹۶) = ۰/۹۷۵$

$N_Z(-۱/۹۶) = ۱ - N_Z(۱/۹۶) = ۰/۰۲۵$

$\Rightarrow p(-۱/۹۶ \le Z \le ۱/۹۶) = p(Z \le ۱/۹۶) - p(Z \le ۱/۹۶) = ۰/۹۷۵ - ۰/۰۲۵ = ۰/۹۵$

۴۲ در فصل نمونه‌گیری نشان دادیم که اگر واریانس جامعه معلوم باشد آمارهٔ $Z = \frac{X-\mu}{\frac{\sigma}{\sqrt{n}}}$ یک متغیر تصادفی نرمال استاندارد می‌باشد و از طرفی با توجه به رابطه ۱۰ – ۲ چون Z نیز یک متغیر تصادفی نرمال استاندارد می‌باشد می‌توان نوشت:

$$P(-۱/۹۶ \le Z \le ۱/۹۶) = P(-۱/۹۶ \le \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \le ۱/۹۶)$$

با ساده نمودن نامساوی $-۱/۹۶ \le \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \le ۱/۹۶$ بدست می‌آوریم:

$$\overline{X} - ۱/۹۶ (\frac{\sigma}{\sqrt{n}}) \le \mu \le \overline{X} + ۱/۹۶ (\frac{\sigma}{\sqrt{n}})$$

بنابراین رابطه ۱۰ – ۲ بصورت زیر بدست می‌آید:

$$p(\overline{X} - ۱/۹۶ (\frac{\sigma}{\sqrt{n}}) \le \mu \le \overline{X} + ۱/۹۶ (\frac{\sigma}{\sqrt{n}})) = ۰/۹۵$$

همانطور که ملاحظه می‌کنید برای μ یک فاصله اطمینان ۹۵ درصدی به شکل $(L_۱, L_۲)$ بدست آوردیم که در آن $L_۱, L_۲$ عبارتند از:

$$L_۱ = \overline{X} - ۱/۹۶ (\frac{\sigma}{\sqrt{n}})$$

$$L_۲ = \overline{X} + ۱/۹۶ (\frac{\sigma}{\sqrt{n}})$$

حال با جایگذاری $\overline{X}$ و σ و n در رابطه بدست می‌آوریم:

$$L_۱ = ۱۰ - ۱/۹۶ (\frac{۲}{\sqrt{۱۶}}) = ۹/۰۲$$

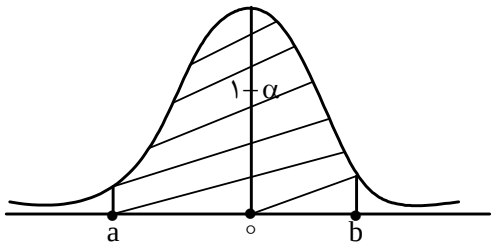
$$L_۲ = ۱۰ + ۱/۹۶ (\frac{۲}{\sqrt{۱۶}}) = ۱۰/۹۸$$

در نتیجه: $p(۹/۰۲ \le \mu\ ۱۰۹۸) = ۰/۹۵$

dalbablog.wordpress.com

یعنی ۹۵ درصد اطمینان داریم که میانگین جامعه در بازه (۹/۰۲ , ۱۰/۹۸) قرار دارد حال برای اینکه بتوانیم یک فاطله اطمینان $(1-\alpha)100$ درصدی در حالت کلی برای جامعه نرمال با میانگین مجهول و واریانس معلوم بسازیم ابتدا مفهوم $Z_{1-\frac{\alpha}{2}}$ را معرفی می‌کنیم.

۴۳ نمودار زیر منحنی نمایش توزیع نرمال استاندارد می‌باشد ملاحظه می‌کنید که مساحت مشخص شده در شکل برابر $1-\alpha$ می‌باشد:



برای پیدا کردن فاصله اطمینان می‌بایستی اعداد a و b را طوری پیدا کرد که:

$$p(a \le Z \le b) = p(a \le \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \le b) = 1-\alpha$$

با فرض اینکه فاصله متقارن باشد مساحت زیر منحنی بعد از نقطه b و قبل از نقطه a برابر $\frac{\alpha}{2}$ می‌باشد. در نتیجه مساحت کل زیر منحنی قبل از نقطه b برابر است با $1-\frac{\alpha}{2} = 1-\alpha + \frac{\alpha}{2}$ که آنرا با $Z_{1-\frac{\alpha}{2}}$ نمایش می‌دهیم. به این ترتیب $Z_{1-\frac{\alpha}{2}}$ نقطه‌ای روی محور افقی است که مساحت زیر منحنی قبل از آن برابر $1-\frac{\alpha}{2}$ می‌باشد. این مطلب با توجه به تعریف تابع توزیع نرمال استاندارد بصورت زیر نیز قابل بیان است:

$$p(Z \le Z_{1-\frac{\alpha}{2}}) = 1-\frac{\alpha}{2}$$

حال می‌توان به جای b و a در $p(a \le Z \le b)$ مقادیر $Z_{1-\frac{\alpha}{2}}$ و $-Z_{1-\frac{\alpha}{2}}$ را جایگزین نمود و با ساده نمودن نامساوی خواهیم داشت:

$$p(-Z_{1-\frac{\alpha}{2}} \le Z \le Z_{1-\frac{\alpha}{2}}) = p(-Z_{1-\frac{\alpha}{2}} \le \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \le Z_{1-\frac{\alpha}{2}})$$

$$p=(-Z_{1-\frac{\alpha}{2}}(\frac{\sigma}{\sqrt{n}}) \le \overline{X}-\mu \le Z_{1-\frac{\alpha}{2}}(\frac{\sigma}{\sqrt{n}}))$$

$$p=(\overline{X}-Z_{1-\frac{\alpha}{2}}(\frac{\sigma}{\sqrt{n}}) \le \mu \le \overline{X}+Z_{1-\frac{\alpha}{2}}(\frac{\sigma}{\sqrt{n}}))=1-\alpha$$

بنابراین با توجه به آخرین رابطه یک فاصله $(1-\alpha)100$ درصد برای میانگین (μ) جامعه نرمال با معلوم بودن σ^2 بدست آوردیم.

۴۴ مثال ۱۸: میزان فروش ماهانه یک فروشگاه یک متغیر تصادفی نرمال با میانگین مجهول μ و انحراف معیار ۵۰ هزار تومان در ماه می‌باشد. نتایج حاصل از ۲۵ ماه فروش نشان می‌دهند که میانگین فروش در طول این ۲۵ ماه عبارتست از ۵۵۰ هزار تومان. اگر ۹۰ درصد اطمینان داشته باشیم میانگین (μ) در فاصله (a,b) می‌باشد مقادیر a و b را مشخص کنید؟

حل: معلومات مثال عبارتند از:

dalbablog.wordpress.com

$n = ۲۵$

$\overline{X} = ۵۵۰$ $\quad \mu =$ مجهول $\quad \sigma = ۵۰$

می‌دانیم فاصله اطمینان از رابطه زیر بدست می‌آید:

$$p\left(\overline{X} - Z_{۱-\frac{\alpha}{۲}}\left(\frac{\sigma}{\sqrt{n}}\right) \le \mu \le \overline{X} + Z_{۱-\frac{\alpha}{۲}}\left(\frac{\sigma}{\sqrt{n}}\right)\right) = ۱-\alpha$$

ابتدا می‌بایستی مقدار α را بدست بیاوریم و سپس از روی آن $Z_{۱-\frac{\alpha}{۲}}$ بدست می‌آید.

$$\%۱۰۰(۱-\alpha)\ \%۹۰ \quad \Rightarrow \quad ۱-\alpha = ۰/۹ \quad \Rightarrow \quad \alpha = ۰/۱$$

$$Z_{۱-\frac{\alpha}{۲}} \overset{\alpha=۰/۱}{=} Z_{۱-\frac{۰/۱}{۲}} = Z_{۰/۹۵}$$

۴۵ با توجه به جدول توزیع نرمال استاندارد $Z_{۰/۹۵}$ عبارتست از:

$$N_Z(Z_{۰/۹۵}) = ۰/۹۵ \quad \Rightarrow \quad Z_{۰/۹۵} = ۱/۶۴۵$$

بنابراین بدست می‌آوریم:

$$p\left(۵۵۰ - ۱/۶۴۵\left(\frac{۵۰}{\sqrt{۲۵}}\right) \le \mu \le ۵۵۰ + ۱/۶۴۵\left(\frac{۵۰}{\sqrt{۲۵}}\right)\right) = ۰/۹$$

$$\Rightarrow p\left(۵۳۳/۵۵ \le \mu \le ۵۶۶/۴۵\right) = ۰/۹$$

بنابراین با احتمال ۰/۹ متوسط درآمد فروشگاه در بازه $(۵۳۳/۵۵\ ,\ ۵۶۶/۴۵)$ قرار دارد. توجه کنید که طول بازه فاصله اطمینان در حالت کلی برابر است با:

$$L_۲ - L_۱ = \left(\overline{X} + Z_{۱-\frac{\alpha}{۲}}\frac{\sigma}{\sqrt{n}}\right) - \left(\overline{X} - Z_{۱-\frac{\alpha}{۲}}\left(\frac{\sigma}{\sqrt{n}}\right)\right) \quad \Rightarrow L_۲ - L_۱ = ۲Z_{۱-\frac{\alpha}{۲}}\left(\frac{\sigma}{\sqrt{n}}\right)$$

ملاحظه می‌کنید که طول بازه مستقل از $\overline{X}$ می‌باشد و تنها به تعداد نمونه‌ها وابسته می‌باشد. یعنی می‌توان تعداد نمونه‌ها را طوری انتخاب نمود که هر فاصله اطمینان دلخواهی برای μ بدست آید.

به عنوان مثال در مثال قبل با ۲۵ نمونه طول بازه عبارتست از:

$$L_۲ - L_۱ = ۲(۱/۶۴۵)\left(\frac{۵۰}{\sqrt{۱۰۰}}\right) = ۱۶/۴۵$$

پیداست که با ۱۰۰ نمونه طول بازه نصف شده است.

۱۰ . ۶ . ۱ فاصله اطمینان برای جامعه نرمال با میانگین مجهول μ و واریانس مجهول $\sigma^۲$. حال فاصله اطمینان را برای حالتی بدست می‌آوریم که علاوه بر میانگین جامعه، واریانس نیز مجهول باشد. از فصل نمونه‌گیری می‌دانیم متغیر تصادفی $T = \dfrac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}}$ یک متغیر تصادفی t با $n-۱$ درجه آزادی می‌باشد بنابراین مجدداً مشابه روشی که قبلاً برای بدست آوردن فاصله اطمینان ارایه کردیم داریم:

$$p\left(-t_{۱-\frac{\alpha}{۲}} \le t_{(n-۱)} \le t_{۱-\frac{\alpha}{۲}}\right) = p\left(-t_{۱-\frac{\alpha}{۲}} \le \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \le t_{۱-\frac{\alpha}{۲}}\right)$$

$$= p\left(\overline{X} - t_{۱-\frac{\alpha}{۲}}\left(\frac{S}{\sqrt{n}}\right) \le \mu \le \overline{X} + t_{۱-\frac{\alpha}{۲}}\left(\frac{S}{\sqrt{n}}\right)\right) = ۱-\alpha$$

dalbablog.wordpress.com

که در آن $t_{۱-\frac{\alpha}{۲}}$ عبارتست از: $t(n-۱)\ (T \le t_{۱-\frac{\alpha}{۲}}) = ۱-\frac{\alpha}{۲}$.

بنابراین فاصله اطمینان در این حالت عبارتست از:

$$L_۱ = \overline{X} - t_{۱-\frac{\alpha}{۲}} (\frac{S}{\sqrt{n}})$$

$$L_۲ = \overline{X} + t_{۱-\frac{\alpha}{۲}} (\frac{S}{\sqrt{n}})$$

۴۷ مثال ۱۹: میزان مصرف بنزین در هر ماه یک متغیر تصادفی نرمال با کیانگین مجهول μ و واریانس $\sigma^۲$ می‌باشد اگر اطلاعات زیر از یک نمونه در طول ۶ ماه بدست آمده باشد مطلوبست یک فاصله اطمینان ۹۰ درصدی برای متوسط میزان مصرف بنزین؟

$$\sum_{i=۱}^{۶} x_i = ۲۴۶ \qquad \sum_{i=۱}^{۶} x_i^۲ = ۱۱۲۰۰$$

حل:

$$\overline{X} = \frac{۱}{n} \sum_{۱}^{n} x_i = \frac{۱}{۶} (۱۴۶) = ۴۱$$

$$S^۲ = \frac{۱}{n-۱} (\sum_{۱}^{n} (x_i - \overline{X})^۲) = \frac{n}{n-۱} (\frac{۱}{n} \sum_{۱}^{n} (x_i - \overline{X})^۲)$$

$$= \frac{n}{n-۱} (\frac{۱}{n} \sum_{i=۱}^{n} x_i^۲ - \overline{X}^۲) = \frac{۱}{n-۱} (\sum_{i=۱}^{n} x_i^۲ - n\overline{X}^۲)$$

$$\Rightarrow S^۲ = \frac{۱}{۵} (۱۱۲۰۰ - ۶(۴۱)^۲) = \frac{۱}{۵} (۱۱۱۴) = ۲۲۲/۸$$

$$\Rightarrow S = ۱۴/۹۲$$

حال برای یک فاصله اطمینان ۹۰ درصدی داریم:

$$۱-\alpha = ۰/۹ \Rightarrow \alpha = ۰/۱$$

$$t_{۱-\frac{\alpha}{۲}} = t_{./۹۵}$$

۴۸ با توجه به جدول t با ۵ درجه آزادی مقدار $t_{./۹۵}$ برابر ۲/۰۱۵ بدست می‌آید.

در نتیجه:

$$L_۱ = \overline{X} - t_{./۹۵} (\frac{S}{\sqrt{n}}) = ۴۱ - ۲/۰۱۵ (\frac{۱۴/۹۲}{\sqrt{۶}}) = ۲۸/۷۲$$

$$L_۲ = \overline{X} + t_{./۹۵} (\frac{S}{\sqrt{n}}) = ۴۱ + ۲/۰۱۵ (\frac{۱۴/۹۲}{\sqrt{۶}}) = ۵۳/۲۷$$

بنابراین با ۹۰ درصد اطمینان می‌توانیم بگوییم که متوسط مصرف بنزین ماهانه مابین ۲۸/۷۲ و ۵۳/۲۷ میلیون لیتر خواهد بود.

برای حالتی که $\sigma^۲$ مجهول باشد طول بازهٔ فاصله اطمینان عبارتست از:

$$L_۲ - L_۱ = ۲t_{۱-\frac{\alpha}{۲}} (\frac{S}{\sqrt{n}})$$

اما توجه می‌کنید که در این حالت طول بازه به متغیر S نیز وابسته می‌باشد و در نتیجه به هیچ عنوان نمی‌توان اندازه نمونه را طوری انتخاب نمود که دقیقاً یک بازه با طول معین بدست آید.

dalbablog.wordpress.com

نکته: قبلاً نشان دادیم که اگر تعداد نمونه‌ها زیاد باشد $(n > ۳۰)$ $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ و $\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$ هر دو به توزیع نرمال استاندارد میل می‌کنند، در نتیجه تمام مطالب گفته شده در حالتی که $n > ۳۰$ باشد برای متغیر تصادفی X با هر توزیعی برقرار می‌باشد.

۴۹ . ۱۰ . ۶ . ۲ فاصله اطمینان برای واریانس جامعه نرمال با میانگین مجهول μ و واریانس $\sigma^۲$.

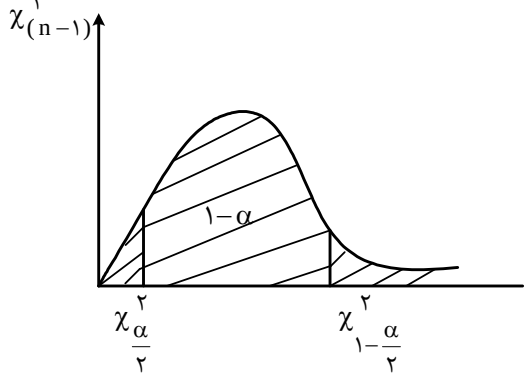
روش بدست آوردن فاصله اطمینان برای واریانس یک جامعه نرمال با میانگین مجهول μ و واریانس $\sigma^۲$ کاملاً مشابه روشی است که برای میانگین جامعه استفاده شد. در این حالت از آماره $\frac{(n-۱)S^۲}{\sigma^۲}$ استفاده می‌کنیم. می‌دانیم که $\frac{(n-۱)S^۲}{\sigma^۲}$ یک متغیر تصادفی خی دو با $n-۱$ درجه آزادی است.

می‌دانیم:

$$p(\chi^۲_{\frac{\alpha}{۲}} \le \chi^۲_{(n-۱)} \le \chi^۲_{۱-\frac{\alpha}{۲}}) = ۱-\alpha$$

$$\Rightarrow p(\chi^۲_{\frac{\alpha}{۲}} \le \frac{(n-۱)S^۲}{\sigma^۲} \le \chi^۲_{۱-\frac{\alpha}{۲}}) = ۱-\alpha \qquad (۳-۱۰)$$

توجه کنید که مقدار ابتدایی بازه $(L_۱)$ برابر $\chi^۲_{\frac{\alpha}{۲}}$ می‌باشد زیرا مطابق شکل منحنی توزیع $\chi^۲_{(n-۱)}$ مساحت زیر منحنی قبل از $\chi^۲_{\frac{\alpha}{۲}}$ برابر $\frac{\alpha}{۲}$ می‌باشد.



به عبارتی $p(\chi^۲_{(n-۱)} \le \chi^۲_{\frac{\alpha}{۲}}) = \frac{\alpha}{۲}$.

حال با ساده نمودن رابطه (۱۰-۳) داریم:

$$p(\frac{(n-۱)S^۲}{\chi^۲_{۱-\frac{\alpha}{۲}}} \le \sigma^۲ \le \frac{(n-۱)S^۲}{\chi^۲_{\frac{\alpha}{۲}}}) = ۱-\alpha$$

بنابراین بازه زیر یک فاصله اطمینان $(۱-\alpha)۱۰۰$ درصدی برای واریانس بدست می‌دهد:

$$L_۱ = \frac{(n-۱)S^۲}{\chi^۲_{۱-\frac{\alpha}{۲}}} \quad , \quad L_۲ = \frac{(n-۱)S^۲}{\chi^۲_{\frac{\alpha}{۲}}}$$

مثال ۲۰-۵: کارخانه‌ای ظروف پلاستیکی تولید می‌کند، ضایعات تولید شده توسط ماشین‌های تولیدی یک متغیر تصادفی نرمال با میانگین μ و واریانس $\sigma^۲$ می‌باشد، اگر پراکندگی ضایعات تولید شده در روز زیاد شود (واریانس تعداد ضایعات زیاد شود). می‌بایستی ماشین‌های تولیدی را تعمیر

dalbablog.wordpress.com

نمود. برای تخمین واریانس یک نمونه ۲۵ تایی از تعداد ضایعات تولید شده در هر روز می‌گیریم و نتایج $\sum x_i = ۱۷۰$, $\sum x_i^۲ = ۲۷۵۰$ را بدست می‌آوریم.

در صورتی که یک فاصله اطمینان ۹۰ درصدی برای واریانس بدست بیاوریم و طول فاصله اطمینان بیشتر از ۵۰ شود می‌بایستی ماشینهای تولیدی را تعمیر کنیم، آیا نیازی به تعمیر می‌باشد؟

حل: معلومات مثال عبارتند از:

$$n = ۲۵$$

$$\overline{X} = \frac{۱}{۲۵}(۱۷۰) = ۶/۸$$

$$S^۲ = \frac{۱}{n-۱}(\sum x_i^{\,۲} - n\overline{X}^۲) = \frac{۱}{۲۴}(۲۷۵۰ - ۲۵(۶/۸)^۲) = ۶۶/۴۱$$

$$\Rightarrow S = ۸/۱۴$$

$$۱-\alpha = ۰/۹ \quad \rightarrow \quad \alpha = ۰/۱ \quad \rightarrow \quad \chi^۲_{\frac{\alpha}{۲}}(n-۱) = \chi^۲_{۰/۰۵}(۲۴) = ۱۳/۸$$

$$\chi^۲_{۱-\frac{\alpha}{۲}}(n-۱) = \chi^۲_{۰/۹۵}(۲۴) = ۳۶/۴$$

بنابراین:

$$L_۱ = \frac{(n-۱)S^۲}{\chi^۲_{۱-\frac{\alpha}{۲}}} = \frac{(۲۴)(۶۶/۴۱)}{۳۶/۴} = ۴۳/۷۸$$

$$L_۲ = \frac{(n-۱)S^۲}{\chi^۲_{۱-\frac{\alpha}{۲}}} = \frac{۲۴(۶۶/۴۱)}{۱۳/۸} = ۱۱۵/۴۹$$

$$L_۲ - L_۱ = ۱۱۵/۴۹ - ۴۳/۷۸ = ۷۱/۷$$

با توجه به اینکه طول بازه فاصله اطمینان بیشتر از ۵۰ واحد شده است بنابراین می‌بایستی ماشینهای تولیدی تعمیر شوند.

۵۱ ۱۰ . ۶ . ۳ فاصله اطمینان برای متغیر تصادفی نمایی با پارامتر مجهول

می‌دانیم متغیر تصادفی نمایی عموماً مدلی برای طول عمر وسایل الکتریکی، تجهیزات و ... می‌باشد. بنابراین اگر بخواهیم اطمینان داشته باشیم که متوسط طول عمر یک متغیر تصادفی نمایی حداقل برابر تعداد ساعات معین است در این صورت می‌بایستی $\frac{۱}{\lambda}$ از عدد معین بیشتر باشد بنابراین برای متغیر تصادفی نمایی تنها یک فاصله اطمینان یک طرفه می‌سازیم.

هرگاه $X_۱, X_۲, \ldots, X_n$ یک نمونه تصادفی از یک متغیر تصادفی نمایی با پارامتر λ باشد آنگاه:

$$p\left(\lambda \le \frac{\chi^۲_{۱-\alpha}}{۲n\overline{X}}\right) = ۱-\alpha$$

یک فاصله اطمینان $(۱-\alpha)$ ۱۰۰ درصدی برای λ تشکیل می‌دهد که در آن $\chi^۲_{۱-\alpha}$ مقداری از توزیع $\chi^۲$ با $۲n$ درجه آزادی است که مساحت زیر منحنی نمایش آن برابر $۱-\alpha$ باشد.

dalbablog.wordpress.com

۵۲ مثال ۲۱: طول عمر یک مدار الکتریکی یک متغیر تصادفی نمایی با پارامتر λ می‌باشد تعداد ۲۰ عدد از این مدارها را آزمایش می‌کنیم تا زمانی که از کار بیفتند. و ملاحظه می‌کنیم که مجموع طول آنها برابر ۲۳۶۰ ساعت می‌شود مطلوبست: تخمین متوسط طول عمر مدار مربوط با فاصله اطمینان ۹۰ درصدی:

حل: معلومات مساله عبارتند از:

$$n = ۲۰$$

$$\sum_{i=۱}^{۲۰} x_i = ۲۳۶۰ \quad \rightarrow \quad \overline{X} = \frac{۱}{n} \sum_{۱}^{n} x_i = \frac{۱}{۲۰} (۲۳۶۰) = ۱۱۸$$

$$۱-\alpha = ۰/۹ \quad \rightarrow \quad \chi^{۲}_{۱-\alpha}(۲n) = \chi^{۲}_{۰/۹}(۴۰) = ۵۱/۸$$

بنابراین یک فاصله اطمینان یک طرفه ۹۰ درصدی برای λ عبارتست از:

$$p\left(\lambda \le \frac{\chi^{۲}_{۱-\alpha}}{۲n\overline{X}}\right) = ۱-\alpha \quad \Rightarrow \quad L_۲ = \frac{۵۱/۸}{۲(۲۰)(۱۱۸)} = ۰/۰۱۰۹$$

توجه کنید که مقدار متوسط طول عمر برای توزیع نمایی $E[X] = \frac{۱}{\lambda}$ می‌باشد بنابراین حداقل متوسط طول عمر مدار عبارتست از:

$$\frac{۱}{۰/۰۱۰۹} = ۹۱/۱۱$$

به عبارتی ۹۰ درصد اطمینان داریم که هر مدار حداقل ۹۱/۱۱ ساعت کار می‌کند.

S-۱۳

۵۳ ۱۰ . ۶ . ۴ فاصله اطمینان برای نسبت p با تعداد نمونه‌های زیاد در توزیع برنولی

نمونه‌های $X_۱, X_۲, \ldots, X_n$ را از یک متغیر تصادفی برنولی با پارامتر مجهول p استخراج می‌کنیم اگر تعداد نمونه‌ها زیاد باشد در این صورت می‌دانیم متغیر تصادفی $\overline{X}$ تقریباً بصورت نرمال با میانگین p و واریانس $\frac{p(۱-p)}{n}$ توزیع می‌شود. به این ترتیب می‌توان نشان داد که یک فاصله اطمینان $۱۰۰(۱-\alpha)$ درصدی برای p عبارتست از:

$$p\left(\overline{X} - Z_{۱-\frac{\alpha}{۲}} \sqrt{\frac{\overline{X}(۱-\overline{X})}{n}} \le p \le \overline{X} + Z_{۱-\frac{\alpha}{۲}} \sqrt{\frac{\overline{X}(۱-\overline{X})}{n}}\right) = ۱ - \alpha$$

توجه کنید که در محاسبه مقدار $\overline{X}$ برابر است با نسبت تعداد موفقیتها به کل تعداد نمونه‌ها به عنوان مثال اگر از ۲۰ مرتبه تیراندازی به هدف ۱۵ بار هدف مورد اصابت قرار گیرد در این صورت $\overline{X} = \hat{p} = \frac{۱۵}{۲۰}$ خواهد بود.

۵۴ مثال ۲۲: احتمال اینکه یک بازیکن فوتبال یک ضربه پنالتی را گل کند یک متغیر تصادفی برنولی با پارامتر p می‌باشد. بازیکن مربوطه در تمرینات از تعداد ۲۰۰ ضربه پنالتی ۱۶۰ مرتبه را موفق به زدن گل شده است. اگر بخواهیم در مسابقات بازیکن مربوطه را مامور زدن ضربه پنالتی کنیم با چه احتمالی می‌توانیم ۹۵ درصد اطمینان داشته باشیم که وی گل می‌زند؟

حل: معلومات مساله عبارتند از:

dalbablog.wordpress.com

$n = ۲۰۰$

$\overline{X} = \hat{p} = \frac{۱۶۰}{۲۰۰} = \frac{۱۶}{۲۰} = ۰/۸$

$۱ - \alpha = ۰/۹۵ \quad \rightarrow \quad \alpha = ۰/۰۵ \quad \rightarrow \quad Z_{۱-\frac{\alpha}{۲}} = Z_{۰/۹۷۵} = ۱/۹۶$

بنابراین یک فاصله ۹۵ درصدی برای p عبارتست از:

$$L_۱ = \overline{X} - Z_{۱-\frac{\alpha}{۲}} \left(\sqrt{\frac{\overline{X}(۱-\overline{X})}{n}} \right) = ۰/۸ - ۱/۹۶ \sqrt{\frac{۰/۸\ (۰/۲)}{۲۰۰}} = ۰/۷۴$$

$$L_۲ = \overline{X} + Z_{۱-\frac{\alpha}{۲}} \left(\sqrt{\frac{\overline{X}(۱-\overline{X})}{n}} \right) = ۰/۸ + ۱/۹۶ \sqrt{\frac{۰/۸\ (۰/۲)}{۲۰۰}} = ۰/۸۵$$

بنابراین با ۹۵ درصد اطمینان می‌توانیم بگوییم که احتمال گل شدن ضربه پنالتی توسط بازیکن مربوطه در بازه $(۰/۷۴ , ۰/۸۵)$ قرار دارد.

۵۵ ۱۰ . ۶ . ۵ فاصله اطمینان برای تفاضل نسبت دو جامعه با تعداد نمونه‌های زیاد

اگر دو نمونه با حجم‌های $n_۱, n_۲$ از دو متغیر تصادفی برنولی در اختیار داشته باشیم، در این صورت برای مقایسه پارامترهای $p_۱, p_۲$ متغیرهای تصادفی از تفاضل این دو نسبت استفاده می‌کنیم به این ترتیب می‌توان نشان داد که یک فاصله اطمینان $(۱-\alpha)۱۰۰$ درصدی برای تفاضل دو نسبت از دو متغیر تصادفی برنولی عبارتست از:

$$p\left(\overline{X}_۱ - \overline{X}_۲ - Z_{۱-\frac{\alpha}{۲}} \sqrt{\frac{\overline{X}_۱(۱-\overline{X}_۱)}{n_۱} + \frac{\overline{X}_۲(۱-\overline{X}_۲)}{n_۲}} \right) \le p_۱ - p_۲ \le \overline{X}_۱ - \overline{X}_۲ + Z_{۱-\frac{\alpha}{۲}}$$

$$\sqrt{\frac{\overline{X}_۱(۱-\overline{X}_۱)}{n_۱} + \frac{\overline{X}_۲(۱-\overline{X}_۲)}{n_۲}}) = ۱-\alpha$$

که در آن $\overline{X}_۱ = \hat{P}_۱ \ , \ \overline{X}_۲ = \hat{P}_۲$

۵۶ مثال ۲۳: در دو شهر A و B احتمال متولد شدن نوزادان پسر را با متغیرهای تصادفی برنولی $X_۱, X_۲$ نشان می‌دهیم. در شهر A از ۳۰۰ نوزاد متولد شده ۱۸۰ نوزاد پسر می‌باشند و در شهر B از ۴۰۰ نوزاد متولد شده ۱۹۰ نوزاد پسر متولد شده‌اند. می‌خواهیم احتمال متولد شدن نزاد پسر را در دو شهر A و B مقایسه کنیم. برای این منظور از یک فاصله اطمینان ۹۵ درصدی استفاده می‌کنیم مطلوبست حدود فاصله اطمینان؟
حل: معلومات مساله عبارتند از:

dalbablog.wordpress.com

$n_1 = ۳۰۰$

$\overline{X}_1 = \hat{p}_1 = \frac{۱۸۰}{۳۰۰} = \frac{۱۸}{۳۰} = ۰/۶$

$n_2 = ۴۰۰$

$\overline{X}_2 = \hat{p}_2 = \frac{۱۹۰}{۴۰۰} = ۰/۴۷۵$

$۱-\alpha = ۰/۹۵ \rightarrow \alpha = ۰/۰۵ \rightarrow Z_{۱-\frac{\alpha}{۲}} = Z_{۰/۹۷۵} = ۱/۹۶$

$$L_1 = \overline{X}_1 - \overline{X}_2 - Z_{۱-\frac{\alpha}{۲}} \sqrt{\frac{(۱-\overline{X}_1)\,\overline{X}_1}{n_1} + \frac{\overline{X}_2(۱-\overline{X}_2)}{n_2}}$$

$$= ۰/۶ - ۰/۴۷۵ - ۱/۹۶ \sqrt{\frac{(۰/۶)۰/۴}{۳۰۰} + \frac{(۰/۴۷۵)۰/۵۲۵}{۴۰۰}} = ۰/۰۵$$

$$L_2 = \overline{X}_1 - \overline{X}_2 + Z_{۱-\frac{\alpha}{۲}} \sqrt{\frac{\overline{X}_1(۱-\overline{X}_1)}{n_1} + \frac{\overline{X}_2(۱-\overline{X}_2)}{n_2}}$$

$$= ۰/۶ - ۰/۴۷۵ + ۱/۹۶ \sqrt{\frac{(۰/۶)۰/۴}{۳۰۰} + \frac{(۰/۴۷۵)۰/۵۲۵}{۴۰۰}} = ۰/۱۹۸$$

فاصله اطمینان ۹۵٪ $p(۰/۰۵ \le p_1 - p_2 \le ۰/۱۹۸) = ۰/۹۵$

همچنین می‌توان نوشت: $\Rightarrow p(p_2 + ۰/۰۵ \le p_1 - p_2 + ۰/۱۹۸) = ۰/۹۵$

بر اساس مساوی فوق احتمال متولد شدن نوزاد پسر در شهر A با اطمینان ۹۵ درصد در بازه $(p_2 + ۰/۰۵ \ , \ p_2 + ۰/۱۹۸)$ قرار دارد.

S-۱۴

۵۷ . ۱۰ . ۶ . ۶ فاصله اطمینان برای نسبت واریانسهای دو متغیر تصادفی نرمال

دو نمونه به حجم n_1 و n_2 را با واریانس نمونه‌ای S_1^2 و S_2^2 از دو جامعه نرمال با واریانس σ_1^2 و σ_2^2 استخراج می‌کنیم. برای مقایسه واریانسهای دو جامعه از نسبت آنها $(\frac{S_1^2}{S_2^2})$ استفاده می‌کنیم. قبلاً نشان دادیم که متغیر تصادفی $\frac{S_1^2 \sigma_2^2}{S_2^2 \sigma_1^2}$ دارای توزیع F با $n_1 - ۱ \ , \ n_2 - ۱$ درجه آزادی است به این ترتیب می‌توان فاصله اطمینان $(۱-\alpha)$ ۱۰۰ درصدی را برای نسبت $\frac{\sigma_1^2}{\sigma_2^2}$ بصورت زیر بدست آورد:

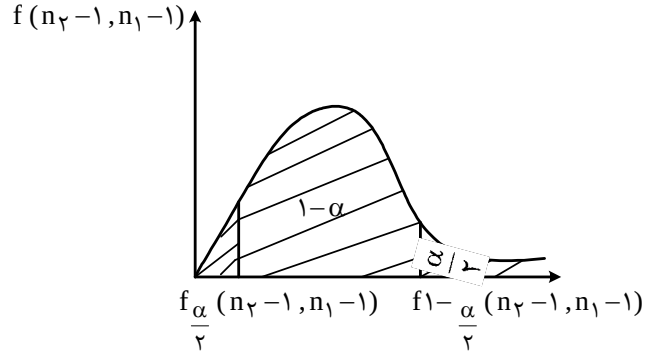
$$p\left(\frac{S_1^2}{S_2^2} \ \frac{۱}{f_{۱-\frac{\alpha}{۲}}(n_1-۱, n_2-۱)} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{S_1^2}{S_2^2} \ \frac{۱}{f_{۱-\frac{\alpha}{۲}}(n_1-۱, n_2-۱)}\right) = ۱-\alpha$$

با توجه به اینکه $f_{\frac{\alpha}{۲}}(n_1, n_2) = \frac{۱}{f_{۱-\frac{\alpha}{۲}}(n_2, n_1)}$ بنابراین رابطه فوق به صورت زیر ساده می‌شود:

$$p\left(\frac{S_1^2}{S_2^2} \ \frac{۱}{f_{۱-\frac{\alpha}{۲}}(n_1-۱, n_2-۱)} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{S_1^2}{S_2^2} \ f_{۱-\frac{\alpha}{۲}}(n_1-۱, n_2-۱)\right) = ۱-\alpha$$

dalbablog.wordpress.com

که در آن $f_{\frac{\alpha}{2}}(n_1-1, n_2-1)$ مقداری از متغیر تصادفی F با n_1-1, n_2-1 درجه آزادی است بطوریکه سطح زیر منحنی به ازای مقادیر کوچکتر از آن برابر $\frac{\alpha}{2}$ باشد. به شکل زیر توجه کنید:



توجه کنید در تمام فواصل اطمینان که در روابط آنها از S^2 استفاده شده است اگر میانگین جامعه (μ) معلوم باشد به جای استفاده از $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{X})^2$ از S^2 استفاده می‌شود که به جای $\overline{X}$ مقدار μ جایگزین شده است یعنی : $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)^2$.

۵۸ مثال ۲۴: در یک کارخانه از دو ماشین A و B برای برش میله‌های فولادی استفاده می‌شود در صورتی که واریانس طول میله‌های بریده شده توسط هر یک از ماشینها سه برابر دیگری شود تصمیم به تعمیر ماشین مربوطه می‌گیریم. در یک نمونه ۲۵ تایی از میله‌های بریده شده توسط ماشین A مقدار S_1^2 برابر ۱۷۰ می‌باشد، همچنین در یک نمونه ۳۱ تایی از میله‌های بریده شده توسط B مقدار S_2^2 برابر ۲۵ می‌باشد. با اطمینان ۹۵٪ برای نسبت $\frac{\sigma_1^2}{\sigma_2^2}$ چه تصمیمی می‌توان در مورد ماشین A و B گرفت؟

حل: داریم:

$n_1 = ۲۵ \qquad n_2 = ۳۱$

$S_1^2 = ۱۷۰ \qquad S_2^2 = ۲۵$

$۱-\alpha = ۰/۹ \rightarrow \alpha = ۰/۱ \rightarrow f_{۱-\frac{\alpha}{۲}}(n_1-۱ , n_2-۲) = f_{۰/۹۵}(۲۴ , ۳۰) = ۱/۹۴$

$f_{۱-\frac{\alpha}{۲}}(n_1-۱ , n_2-۲) = f_{۰/۹۵}(۳۰ , ۲۴) = ۱/۸۹$

به این ترتیب یک فاصله اطمینان ۹۵ درصدی عبارتست از:

$$p\left(\frac{۱۷۰}{۲۵}\frac{۱}{۱/۹۴} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{۱۷۰}{۲۵}۱/۸۹\right) = ۰/۹۵$$

$$\Rightarrow p\left(۳/۵ \le \frac{\sigma_1^2}{\sigma_2^2} \le ۱۲/۸۵\right) = ۰/۹۵$$

بنابراین نسبت واریانسها با ۹۵ درصد اطمینان در بازه $(۳/۵ , ۱۲/۸۵)$ قرار دارد.

و در نتیجه: $p(۳/۵ , ۱۲/۸۵\,\sigma_2^2) = ۰/۹۵$

یعنی با ۹۵ درصد اطمینان می‌توان گفت که واریانس ماشین A سه برابر واریانس ماشین B می‌باشد بنابراین ماشین A بایستی تعمیر شود.

۵۹ ۱۰ . ۶ . ۷ فاصله اطمینان برای تفاضل میانگین‌های دو جامعه نرمال

dalbablog.wordpress.com

از دو جامعه نرمال با واریانسهای σ_1^2 و σ_2^2 دو نمونه به حجم n_1 و n_2 می‌گیریم. قبلاً نشان دادیم که $Z=\dfrac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}$ یک متغیر تصادفی نرمال استاندارد می‌باشد. بنابراین می‌توانیم یک فاصله اطمینان $100(1-\alpha)$ درصد برای تفاضل میانگین‌های دو جامعه نرمال با واریانس معلوم بصورت زیر بدست آوریم:

$$p\left(\overline{X}_1-\overline{X}_2-Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \le \mu_1-\mu_2 \le \overline{X}_1-\overline{X}_2+Z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)=1-\alpha$$

و در صورتی که واریانسهای دو جامعه نامعلوم اما مساوی باشند ابتدا واریانس مشترک دو نمونه را بصورت زیر برآورد می‌کنیم:

$$\sigma_1^2=\sigma_1^2=\sigma^2$$

$$\hat{\sigma}^2=S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

فاصله اطمینان در این حالت از رابطه زیر بدست می‌آید:

$$p\left(\overline{X}_1-\overline{X}_2-t_{1-\frac{\alpha}{2}}(n_1+n_2-2)\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \le \mu_1-\mu_2 \le \overline{X}_1-\overline{X}_2+t_{1-\frac{\alpha}{2}}(n_1+n_2-2)S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)=1-\alpha$$

۶۰ مثال ۲۵: میزان برق مصرفی دو شهر A و B یک متغیر تصادفی نرمال می‌باشد برای مقایسه متوسط مصرف یک نمونه ۱۵ تایی از دو شهر انتخاب می‌کنیم و نتایج زیر را بدست می‌آوریم:

$\overline{X}_A=۲۵۹۰ \qquad \sigma_A^2=۱۵۰$

$\overline{X}_B=۲۵۶۰ \qquad \sigma_B^2=۱۰۰$

مطلوبست:

الف) با فاصله اطمینان ۹۹ درصدی چه نظری می‌توان درباره تفاضل متوسط مصرف در دو شهر داشت؟

ب) اگر برای دو نمونه داشته باشیم $S_A^2=۱۶۵$, $S_B^2=۱۲۰$ یک فاصله اطمینان ۹۹ درصدی برای حالتی که واریانسها نامعلوم باشند و مساوی بسازید؟

حل:الف)

$$1-\alpha=۰/۹۹ \quad \rightarrow \quad \alpha=۰/۰۱ \quad \rightarrow \quad Z_{1-\frac{\alpha}{2}}=Z_{۰/۹۹}=۲/۵۷۵$$

$$\rightarrow \quad L_1=۲۵۹۰-۲۵۶۰-۲/۵۷۵\sqrt{\frac{۱۵۰}{۱۵}+\frac{۱۰۰}{۱۵}}=۱۹/۴۸$$

$$L_2=۲۵۹۰-۲۵۶۰-۲/۵۷۵\sqrt{\frac{۱۵۰}{۱۵}+\frac{۱۰۰}{۱۵}}=۴۰/۵۱$$

بنابراین با ۹۹ درصد اطمینان می‌توانیم بگوییم که $۱۹/۴۸ \le \mu_1-\mu_2 \le ۴۰/۵۱$ و چون $\mu_1-\mu_2$ در یک بازه مثبت قرار گرفته‌اند می‌توانیم با ۹۹ درصد اطمینان بگوییم که متوسط مصرف شهر A از شهر B بیشتر است.

۶۱ ب)

$$S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}=\frac{۱۴(۱۲۰+۱۶۵)}{۱۵+۱۵-۲}=۱۴۲/۵ \quad \Rightarrow \quad S_p=۱۱/۹۳$$

dalbablog.wordpress.com

$$۱-\alpha = ۰/۹۹ \quad \rightarrow \quad \alpha = ۰/۰۱ \quad \rightarrow \quad t_{۱-\frac{\alpha}{۲}}(n_۱+n_۲-۲) = t_{۰/۹۹۵}(۲۸) = ۲/۷۶$$

$$L_۱ = \overline{X}_۱ - \overline{X}_۲ - t_{۱-\frac{\alpha}{۲}}(n_۱+n_۲-۲)\, S_p \sqrt{\frac{۱}{n_۱}+\frac{۱}{n_۲}} = ۲۵۹۰ - ۲۵۶۰ - ۲/۷۶\,(۱۱/۹۳)\sqrt{\frac{۱}{۱۵}+\frac{۱}{۱۵}} = ۱۷/۹۷$$

$$L_۲ = \overline{X}_۱ - \overline{X}_۲ - t_{۱-\frac{\alpha}{۲}}(n_۱+n_۲-۲)\, S_p \sqrt{\frac{۱}{n_۱}+\frac{۱}{n_۲}} = ۲۵۹۰ - ۲۵۶۰ + ۲/۷۶\,(۱۱/۹۳)\sqrt{\frac{۱}{۱۵}+\frac{۱}{۱۵}} = ۴۲$$

ملاحظه می‌کنید که در حالت دوم که واریانس مشترک را بصورت تقریبی محاسبه نمودیم باز هم فاصله اطمینان مقداری نزدیک به حالت الف شده است.

dalbablog.wordpress.com

فصل یازدهم

۱-۱۱ آزمون فرض

در فصل قبل با روشهای برآورد پارامترهای مجهول جامعه آشنا شدید، به عبارتی نشان دادیم که چگونه می‌توان با استفاده از نتایج حاصل از نمونه‌گیری پارامترهای مجهول یک جامعه آماری با توزیع معلوم را بدست آورد. حال در این فصل نشان می‌دهیم که چگونه می‌توان در مورد صحت یک فرضیه بر اساس یک جامعه آماری قضاوت نمود. به عنوان مثال فرض می‌کنیم که یک داروی بخصوص بر روی یک بیماری اثر مثبتی دارد به این ترتیب می‌بایستی برای اثبات این فرضیه، عده‌ای از بیماران را انتخاب کرده و دارو را روی آنها آزمایش می‌کنیم و با توجه به نتایج بدست آمده در مورد صحت یا عدم صحت فرضیه مورد نظر تصمیم می‌گیریم. روشی که در آن فرض مربوطه را پذیرفته یا رد می‌کنیم به آزمون فرض معروف می‌باشد، در این فصل نحوه آزمون فرضهای مختلف را معرفی می‌کنیم.

۱۱-۲

۱۱ . ۱ فرضها و آزمونها

تعریف دقیق فرض و آزمون فرض عبارتند از:

فرض: یک فرض عبارتست از گزاره‌ای درباره قانون احتمال یک متغیر تصادفی (که قابل نمونه‌گیری باشد)

آزمون فرض: نمونه‌گیری از متغیر تصادفی مربوطه و تصمیم‌گیری در مورد پذیرفتن یا رد فرض مورد نظر بر اساس نمونه بدست آمده.

بطور کلی نتیجه یک آزمون فرض یکی از چهار مورد زیر می‌باشد:

۱- فرض را رد می‌کنیم هنگامی که واقعاً صحیح باشد.
۲- فرض را رد می‌کنیم هنگامی که واقعاً صحیح نباشد.
۳- فرض را رد می‌پذیریم هنگامی که واقعاً صحیح باشد.
۴- فرض را رد می‌پذیریم هنگامی که واقعاً صحیح نباشد.

برای وشن شدن مطلب به مثال زیر توجه کنید:

۱۱-۳ مثال ۱: گروهی از پزشکان معتقدند یک دارو نیروزا می‌باشد. برای تصمیم‌گیری در این مورد آنرا روی ۲۰ نفر از ورزشکاران آزمایش می‌کنند. در صورتی که دارو در بیشتر از ۷۰ درصد مواقع مثبتی بر عملکرد عادی ورزشکاران داشته باشد، پزشکان آنرا جزء داروهای نیروزا طبقه بندی می‌کنند. متغیر تصادفی X را برابر تعداد ورزشکارانی در بین ۲۰ نفر در نظر می‌گیریم که دارو تاثیر مثبتی بر عملکرد آنها داشته است. در این صورت X یک متغیر تصادفی دو جمله‌ای با پارامترهای $X \sim B(p\ ,\ ۲۰)$ می‌باشد. که در آن p درصد موثر بودن دارو می‌باشد و مقداری نامعلوم می‌باشد. نیروزا بودن دارو به این معنی است که $p > ۰/۷۰$ به عبارتی اگر از روی نمونه‌ها به این نتیجه برسیم که $p > ۰/۷۰$ می‌باشد در این صورت می‌توان گفت که دارو نیروزا بوده است. نتایج حاصل از نمونهٔا یکی از دو حالت (فرض) زیر را نتیجه می‌دهد:

۱- داروی مورد نظر نیروزا می‌باشد : $p > ۰/۷۰$
۲- داروی مورد نظر نیروزا نمی‌باشد : $p > ۰/۷۰$

۱۱-۴ در روش آزمون یکی از فرضها را فرض خنثی یا فرض صفر در نظر می‌گیریم و آنرا با $H_۰$ نمایش می‌دهیم به عبارتی $H_۰$ در این مثال همان فرض است که مورد ادعای پزشکان می‌باشد یعنی فرض داروی مورد نظر نیروزا می‌باشد $(p > ۰/۷۰)$ فرض $H_۰$ خواهد بود.
در مقابل هر فرضی که فرض $H_۰$ را رد کند فرض مخالف یا مکمل نامیده می‌شود و با $H_۱$ نمایش داده می‌شود. که در این مثال فرض داروی مورد نظر نیروزا نمی‌باشد $(p \leq ۰/۷۰)$ فرض $H_۱$ می‌باشد. بنابراین:

دارو نیروزاست $H_۰ = p > ۰/۷۰$

dalbablog.wordpress.com

دارو نیروزا نیست $H_1 = p \leq ۰/۷۰$

توجه کنید که فرض مخالف صورتهای متفاوتی می‌تواند داشته باشد مثلاً اگر فرض $H_۰$ برابر $p = ۰/۷$ می‌بود در این صورت فرض H_1 هر یک از حالات زیر می‌توانست باشد:

$H_1 : p > ۰/۷$

$H_1 : p \neq ۰/۷$

$H_1 : p = ۰/۵$

تصمیم‌گیری در مورد اینکه کدام فرض را فرض مخالف در نظر بگیریم اهمیت زیادی دارد و معمولاً به صورت مساله بستگی دارد.

در ایم مثال پزشکان برای اثبات ادعای خود ۲۰ ورزشکار را انتخاب نموده و با توجه به متغیر تصادفی X این احتمال وجود دارد که نتایج طوری بدست آیند که پرشکان دارو را نیروزا معرفی کنند در حالی که واقعاً دارو نیروزا بوده است.

۵-۱۱ به عنوان مثالی دیگر فرض کنید یک سکه را ۲۰ مرتبه پرتاب کنیم و ۱۶ شیر مشاهده کنیم بوضوح مشاهدهٔ ۱۶ شیر تنها با ۲۰ مرتبه پرتاب سکه دلیلی بر ناسالم بودن سکه نمی‌باشد در مثال داروی نیروزا هم این مطلب صدق می‌کند. بنابراین در حالت کلی چهار حالت زیر بدست می‌آید:

۱- $H_۰$ را می‌پذیریم هنگامی که واقعاً صحیح باشد.

۲- $H_۰$ را رد می‌کنیم هنگامی که واقعاً صحیح باشد.

۳- $H_۰$ را می‌پذیریم هنگامی که واقعاً صحیح نباشد.

۴- $H_۰$ را رد می‌کنیم هنگامی که واقعاً صحیح نباشد.

مشاهده می‌کنید که حالت دوم و سوم نشان دهنده خطا در تصمیم‌گیری می‌باشند حالت دوم را خطای نوع اول (ریسک فروشنده می‌نامیم و حالت سوم را خطای نوع دوم (ریسک مشتری) می‌نامیم. خطای نوع اول را با α نمایش می‌دهند و آنرا سطح معنی‌دار یا سطح تشخیص آزمون می‌نامند و خطای نوع دوم را با B نمایش می‌دهند.

۶-۱۱ ۱۱ . ۲ ناحیه بحرانی و آماره آزمون

برای بیان ناحیه بحرانی و آماره آزمون مجدداً به مثال ۱ توجه کنید. در مثال ۱ نشان دادیم که:

$H_۰ : p > ۰/۷۰$

$H_1 : p \leq ۰/۷۰$

چگونگی انجام آزمون به این ترتیب است که متغیر تصادفی X را برابر تعداد ورزشکاران که دارو بروی عملکرد آنها تاثیر مثبتی داشته در نظر می‌گیریم در این صورت اگر مقدار مشاهده شده X مثلاً برابر ۱۵ باشد در نتیجه $p = \frac{۱۵}{۲۰} = \frac{۳}{۴} = ۰/۷۵$ و فرض $H_۰$ را قبول می‌کنیم زیرا $p > ۰/۷۰$ اما اگر مقدار مشاهده شده X مثلاً برابر ۵ باشد بوضوح فرض $H_۰$ را رد می‌کنیم و فرض H_1 را می‌پذیریم زیرا $p = \frac{۵}{۲۰} = ۰/۲۵$ و $p \leq ۰/۷۰$.

در حالت کلی زمانی فرض $H_۰$ را می‌پذیریم که مقدار مشاهده شدهٔ X از عدد ۱۴ بیشتر باشد زیرا: $p = \frac{۱۴}{۲۰} = ۰/۷۰$

و اگر مقدار مشاهده شده X کوچکتر یا مساوی ۱۴ باشد فرض $H_۰$ را رد می‌کنیم یعنی اگر مقادیر مشاهده شده X متعلق به مجموعه $C = \{x \mid x \leq ۱۴\}$ باشد آنگاه $H_۰$ را رد می‌کنیم و در غیر این صورت $H_۰$ را می‌پذیریم. به آماره X که بر اساس مقادیر مشاهده شده آن $H_۰$ را رد می‌کنیم یا می‌پذیریم آماره آزمون گویند و به ناحیه C که به ازای مقادیر آن $H_۰$ رد می‌شود ناحیه بحرانی می‌گوییم. به تعریف زیر توجه کنید:

dalbablog.wordpress.com

۱۱-۷ تعریف: به آماره $T = T(X_1, X_2, \cdots, X_n)$ که بر اساس مقادیر مشاهده شده آن یک فرض را می‌پذیریم یا رد می‌کنیم، آماره آزمون گویند. و به ازای مجموعه مقادیری از این آماره که بر اساس آنها فرض $H_\circ$ رد می‌شود ناحیه بحرانی آزمون می‌گوییم و با نماد C نمایش می‌دهیم و به متمم ناحیه بحرانی که با C' نمایش می‌دهیم ناحیه پذیرش آزمون می‌گوییم.

با توجه به تعریف در آزمون یک فرض به این صورت عمل می‌کنیم که ابتدامقادیر نمونه‌های $X_1, X_2, \cdots, X_n$ را جمع آوری می‌کنیم و بر اساس آنها آماره آزمون را که بشکل $t = T(x_1, x_2, \cdots, x_n)$ می‌باشد تشکیل می‌دهیم حال اگر $t \in C$ باشد فرض $H_\circ$ را رد می‌کنیم و در غیر این صورت آن را می‌پذیریم.

۳ . ۱۱ ۸ خطای آزمون

اشاره کردیم که رد تصمیم‌گیری ممکن است دو نوع خطای نوع اول (α) و نوع دوم (β) رخ دهد، که در هر دو حالت نتیجه مطلوب نمی‌باشد و در عمل می‌خواهیم تا جای ممکن مقدار خطای α و β را کاهش دهیم. با استفاده از احتمالات شرطی خطای نوع اول و دوم بصورت زیر بدست می‌آیند:

$\alpha = p$ ($H_\circ$ رد شود | فرض $H_\circ$ درست باشد) $= p(T(X_1, \cdots X_n) \in C$ | $H_\circ$ درست باشد)

$\beta = p$ ($H_\circ$ پذیرفته شود | فرض $H_\circ$ درست باشد) $= p(T(X_1, \cdots X_n) \notin C$ | H_1 درست باشد)

اگر در آزمون فرض به این نتیجه برسیم که $H_\circ$ باید رد شود و در حالی که H_1 واقعاً درست باشد یعنی مقدار احتمال زیر را داشته باشیم:

p ($H_\circ$ رد شود | H_1 درست باشد)

$= p$ ($H_\circ$ رد شود | $H_\circ$ واقعاً نادرست باشد)

آشکار است که هر چه مقدار این احتمال بیشتر باشد آزمون نتیجه دقیق‌تری بدست داده است به همین دلیل به این احتمال توان آزمون گفته می‌شود. و با علامت β^* نمایش داده می‌شود.

توجه کنید که مقدار β^* برابر با $1-\beta$ می‌باشد:

$\beta^* = p$ ($H_\circ$ رد شود | H_1 درست باشد) $= 1-p$ ($H_\circ$ پذیرفته شود | H_1 درست باشد)

$= 1-\beta$

۹ مثال ۲: شخصی ادعا می‌کند که سکه استفاده شده برای تعین زمین بازی در یک مسابقه فوتبال یکنواخت نبوده و با احتمال ۰/۶ شیر می‌آمده است. برای تحقیق صحت ادعای وی سکه را ۲۰ مرتبه پرتاب می‌کنیم مطلوبست:

الف) تعین ناحیه بحرانی آزمون و فرضهای آزمون.

ب) محاسبه مقدار احتمال خطای نوع اول و خطای نوع دوم.

ج) محاسبه توان آزمون.

حل: الف) X را تعداد شیرهای بدست آمده در ۲۰ مرتبه پرتاب سکه در نظر می‌گیریم در این صورت X یک متغیر تصادفی دو جمله‌ای با پارامتر p و ۲۰ می‌باشد. $X \sim \beta(۲۰, p)$.

طبق تعریف فرض صفر خلاف ادعای مطرح شده می‌باشد بنابراین:

$$\begin{cases} H_\circ : \ p = \frac{۱}{۲} \\ H_۱ : \ p = \frac{۶}{۱۰} \end{cases}$$

با توجه به اینکه در متغیر تصادفی برنولی $\mu = np$ می‌باشد بنابراین اگر در ۲۰ مرتبه پرتاب سکه حداقل تعداد $۲۰ \times \frac{۶}{۱۰} = ۱۲$ شیر مشاهده کنیم می‌بایستی فرض $H_\circ$ را رد کنیم به این ترتیب ناحیه بحرانی عبارتست از:

$$C = \{x \mid x \geq ۱۲\}$$

ب) محاسبه خطای نوع اول:

dalbablog.wordpress.com

$$\alpha = p\,(X \in C \mid \text{درست باشد}) = p(X \geq ۱۲ \mid p = \frac{۱}{۲})$$

$$۱ - p\,(X \leq ۱۱ \mid p = \frac{۱}{۲}) = ۱ - ۰/۷۴۸۳ = ۰/۲۵۱۷$$

محاسبه خطای نوع دوم:

$$\beta = p\,(X \notin C \mid \text{درست } H_۱) = p(X < ۱۲ \mid p = ۰/۶)$$
$$= p(X \leq ۱۱ \mid p = ۰/۶) = ۰/۴۰۴۴$$

ج) محاسبه توان آزمون:

$$\beta^* = ۱ - \beta = ۰/۵۹۵۶$$

ملاحظه می‌کنید که احتمال خطای نوع اول کم و احتمال خطای نوع دوم زیاد می‌باشد.

۱۰-۱۱ مثال ۳: در مثال قبل اگر ناحیه بحرانی بصورت $C = \{x \mid x \geq ۱۳\}$ باشد احتمال خطای نوع اول و دوم و توان آزمون را محاسبه کنید؟

حل:

$$\alpha = p(X \geq ۱۳ \mid p = \frac{۱}{۲}) = ۱ - p\,(X \leq ۱۲ \mid p = \frac{۱}{۲}) = ۱ - ۰/۸۶۸۴ = ۰/۱۳۱۶$$

$$\beta = p(X < ۱۳ \mid p = ۰/۶) = p\,(X \leq ۱۲ \mid p = ۰/۶) = ۰/۵۸۴۱$$

$$\beta^* = ۱ - \beta = ۰/۴۱۵۹$$

ملاحظه می‌کنید که در حالت دوم با متغیر ناحیه بحرانی مقدار احتمال نوع اول کاهش یافت اما این مساله موجب افزایش احتمال خطای نوع دوم شده است.

بنابراین می‌بایستی ناحیه بحرانی طوری انتخاب شود که همزمان با در نظر گرفتن یک مقدار حداکثر برای خطای نوع اول باعث حداقل نمودن نوع دوم شود. به این ترتیب توان آزمون نیز حداکثر می‌شود.

در مثال بعد با قرار دادن یک مقدار معین برای احتمال خطای نوع اول (α) تلاش می‌نیم احتمال خطای نوع دوم را تا جای ممکن کاهش دهیم و در نتیجه توان آزمون را افزایش دهیم.

۱-۱۱-۱۱ مثال ۴: X یک متغیر تصادفی نرمال با میانگین مجهول μ و واریانس ۹ می‌باشد. آزمون فرض زیر را در نظر بگیرید:

$$\begin{cases} H_۰ : \mu = ۰ \\ H_۱ : \mu = ۱ \end{cases}$$

اگر از X یک نمونه ۲۵ تایی بگیریم و ناحیه بحرانی بصورت $C = \{(X_۱, \cdots, X_{۲۵}) \mid \overline{X} \geq C\}$ باشد مقدار C را طوری بدست بیاورید که $\alpha = ۰/۱$ باشد و سپس احتمال خطای نوع دوم و توان آزمون را محاسبه کنید؟

حل: ابتدا توجه کنید که $\overline{X}$ دارای میانگین μ و واریانس $\frac{\sigma^۲}{n}$ می‌باشد. یعنی $\overline{X} \sim (\mu, \frac{۹}{۲۵})$ به این ترتیب:

$$\alpha = ۰/۱ = p(\overline{X} > C \mid \mu = ۰) = p\,(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} > \frac{C - \mu}{\frac{\sigma}{\sqrt{n}}} \mid \mu = ۰)$$

$$= p(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} > \frac{C - ۰}{\frac{۳}{۵}}) = p\,(Z > ۱/۶۶\,C) = ۰/۱$$

$$\Rightarrow ۱ - p\,(Z \leq ۱/۶۶\,C) = ۰/۱ \quad \Rightarrow p(Z \leq ۱/۶۶\,C) = ۰/۹$$

رابطه اخیر معادل است با اینکه $Z_{۰/۹} = ۱/۶۶\,C$ از جدول مقدار $Z_{۰/۹}$ برابر است با ۱/۲۸ بنابراین:

dalbablog.wordpress.com

$$Z_{./۹} = ۱/۲۸ = ۱/۶۶\ C \quad \Rightarrow \quad C = ۰/۷۷$$

به این ترتیب ناحیه بحرانی بصورت زیر بدست می‌آید:

$$C = \{(X_۱, \cdots, X_{۲۵}) \mid \overline{X} > ۰/۷۷\}$$

ملاحظه می‌کنید مه در این مثال برای اینکه بتوانیم مقدار خطای نوع اول را به دلخواه کاهش دهیم به ناچار می‌بایستی بازه ناحیه بحرانی را متغیر فرض می‌کردیم، حال مقدار خطای نوع دوم را بدست می‌آوریم:

$$\beta = p\,(\overline{X} \le C \mid \mu = ۱) = p\,(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{۰/۷۷-\mu}{\frac{\sigma}{\sqrt{n}}} \mid \mu = ۱)$$

$$= p(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \le \frac{۰/۷۷-۱}{\frac{۳}{۵}}) = p\,(Z \le -۰/۳۸) = N_Z(-۰/۳۸) = ۰/۳۵۲۰$$

توان آزمون: $\beta^* = ۱ - \beta = ۱ - ۰/۳۵۲۰ = ۰/۶۴۸$

۱۲-۱۱ ۱۱ . ۴ انواع فرضها

فرض $H_۰ : \mu = ۲$ را در نظر بگیرید، این فرض با بیان اینکه میانگین جامعه برابر ۲ می‌باشد توزیع جامعه را معلوم می‌کند و نشان می‌دهد که پارامتر مجهول جامعه میانگین برابر عدد ۲ می‌باشد، به این نوع فرضها که توزیع جامعه را کاملاً مشخص می‌سازند فرضهای ساده می‌کوییم.

حال فرض $H_۱ : \mu \ge ۲$ را در نظر بگیرید، واضح است که اگر این فرض درست باشد با بیان یک بازه برای پارامتر مجهول جانعه، توزیع را بصورت دقیق مشخص نمی‌کند به این فرضها که با درست بودنشان پارامتر مجهول جامعه و در نتیجه توزیع جامعه بصورت دقیق مشخص نمی‌شود فرضهای مرکب می‌گوییم.

به عنوان مثال در مثال ۴ فرضهای $H_۰ : \mu = ۰$ و $H_۱ : \mu = ۱$ فرضهای ساده می‌باشند اما در مثال ۱ فرضهای $H_۰ : p \ge ۰/۶$ و $H_۱ : p < ۰/۶$ فرضهای مرکب می‌باشند.

۱۳-۱۱ ۱۱ . ۵ انواع آزمونها

اگر θ یک پارامتر نامعلوم جامعه باشد و بخواهیم آزمونهایی در مورد این پارامتر انجام دهیم، آزمون هر فرض آماری که در آن فرضیه مقابل $(H_۱)$ یک طرفه باشد آزمون یک طرفه نامیده می‌شود. به عنوان مثال برای مقدار ثابت $\theta_۰$ از θ آزمونهای زیر همگی یک طرفه می‌باشند:

$$\begin{cases} H_۰ : \theta = \theta_۰ \\ H_۱ : \theta \ge \theta_۰ \end{cases} \qquad \begin{cases} H_۰ : \theta \ge \theta_۰ \\ H_۱ : \theta < \theta_۰ \end{cases}$$

اگر در یک آزمون فرضیه مقابل $(H_۱)$ دو طرفه باشد آن آزمون را آزمون دو طرفه می‌نامیم. مانند آزمون زیر:

$$\begin{cases} H_۰ : \theta = \theta_۰ \\ H_۱ : \theta \ne \theta_۰ \end{cases}$$

۱۴-۱۱ ۱۱ . ۶ مراحل انجام یک آزمون فرض

برای انجام یک آزمون فرض می‌بایستی مراحل زیر را بصورت گام به گام طی نمود:

۱- تعین فرضهای صفر $H_۰$ و فرض مقابل $H_۱$.

۲- تعین یک طسح معنی‌دار α که معمولاً یکی از اعداد ۰/۰۱ ، ۰/۰۵ یا ۰/۱ در نظر گرفته می‌شود.

۳- تعین آماره آزمون $T = T(X_۱, \cdots, X_n)$ که عموماً بر اساس برآوردگر نقطه‌ای پارامتر مجهول θ بدست می‌آید.

۴- تعین ناحیه بحرانی آزمون که از روی آماره آزمون، فرض مقابل آزمون و سطح معنی‌دار α بدست می‌آید.

۵- محاسبه مقدار آماره آزمون که از روی نمونه‌های تصادفی $X_۱, X_۲, \cdots, X_n$ بدست می‌آید.

dalbablog.wordpress.com

۶- نتیجه‌گیری – اگر مقدار محاسبه شده آماره آزمون درون ناحیه بحران باشد فرض $H_\circ$ را رد و در غیر این صورت فرض $H_\circ$ را می‌پذیریم.
توجه کنید که عموماً در حل مسایل آزمون فرض، با در نظر گرفتن انواع آزمون فرضها و تعین آماره آزمون و ناحیه بحرانی برای هر یک از آزمون فرض‌ها محاسبه مراحل ۳ و ۴ بسیار ساده‌تر می‌شود.

۱۱-۱۵ . ۷ آزمون فرضهای آماری برای پارامترهای جامعه

برای سادگی انجام آزمونهای فرض بهترین راه محاسبه آماره آزمون و ناحیه بحرانی برای فرضهای متفاوت می‌باشد. بنابراین در این بخش ابتدا آزمون فرض را روی میانگین یک جمعیت را زمانی که واریانس معلوم می‌باشد انجام می‌دهیم:

فرض می‌کنیم X یک متغیر تصادفی با میانگین مجهول μ و واریانس معلوم σ^2 باشد. نمونه‌های تصادفی $X_1, X_2, \cdots, X_n$ را انتخاب می‌کنیم، می‌خواهیم آزمونهایی را روی میانگین μ انجام دهیم. در این صورت سه حالت کلی زیر را در نظر می‌گیریم:

۱- آزمون فرض $\begin{cases} H_\circ : \mu = \mu_\circ \\ H_1 : \mu > \mu_\circ \end{cases}$ که در آن $H_\circ$ رد می‌شود اگر $\overline{X} > C$

۲- آزمون فرض $\begin{cases} H_\circ : \mu = \mu_\circ \\ H_1 : \mu < \mu_\circ \end{cases}$ که در آن $H_\circ$ رد می‌شود اگر $\overline{X} > C$

۳- آزمون فرض $\begin{cases} H_\circ : \mu = \mu_\circ \\ H_1 : \mu \neq \mu_\circ \end{cases}$ که در آن $H_\circ$ رد می‌شود اگر $\overline{X} < C_1$ یا $\overline{X} > C_2$

توجه کنید که در هر حالت مقدار C که بیانگر ناحیه بحرانی می‌باشد با توجه به سطح معنی‌دار α بدست می‌آید. حال برای هر حالت آماره آزمون و مقدار C را محاسبه می‌کنیم:

۱۱-۱۶ حالت اول: $H_\circ$ رد می‌شود اگر $\overline{X} > C$ بنابراین خطای نوع اول عبارتست از: $\alpha = p(\overline{X} > C \mid \mu = \mu_\circ)$

اگر X یک متغطر تصادفی نرمال باشد یا $n \geq ۳۰$ باشد قبلاً نشان دادیم که $Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ یک متغیر تصادفی نرمال استاندارد می‌باشد بنابراین:

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(\circ, ۱)$$

$$\alpha = p(\overline{X} > C \mid \mu = \mu_\circ) = p\left(\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} > \frac{C - \mu_\circ}{\frac{\sigma}{\sqrt{n}}}\right) = p\left(Z > \frac{C - \mu_\circ}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$\Rightarrow ۱ - p\left(Z > \frac{C - \mu_\circ}{\frac{\sigma}{\sqrt{n}}}\right) = \alpha \quad \Rightarrow \quad p\left(Z > \frac{C - \mu_\circ}{\frac{\sigma}{\sqrt{n}}}\right) = ۱ - \alpha$$

$$\Rightarrow \frac{C - \mu_\circ}{\frac{\sigma}{\sqrt{n}}} = Z_{۱-\alpha} \quad \Rightarrow \quad C = \frac{\sigma}{\sqrt{n}}(Z_{۱-\alpha}) + \mu_\circ$$

حال با توجه به مقدار C ناحیه بحرانی ازمون بصورت زیر بدست می‌آید:

$$\overline{X} > C \quad \rightarrow \quad \overline{X} > \mu_\circ + \frac{\sigma}{\sqrt{n}} Z_{۱-\alpha} \quad \text{یا} \quad \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} > Z_{۱-\alpha}$$

نتیجه می‌گیریم که:

dalbablog.wordpress.com

در آزمون $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$ فرض H_0 رد می‌شود اگر و فقط اگر:

$$Z = \frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > Z_{1-\alpha}$$

توجه کنید که اگر فرض H_0 بصورت $H_0 : \mu \le \mu_0$ باشد می‌توان نشان داد که ناحیه بحرانی باز هم به صورت رابطه بالا می‌باشد.

۱۱-۱۷ حالت دوم: آزمون فرض $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$ H_0 را رد می‌کنیم اگر $\overline{X} < C$ مشابه حالت قبل اگر C را محاسبه کنیم مقدار ناحیه بحرانی بصورت زیر بدست می‌آید:

$$C = \mu_0 - Z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \quad \Rightarrow \quad \overline{X} < C \quad \rightarrow \quad \overline{X} < \mu_0 - \frac{\sigma}{\sqrt{n}} Z_{1-\alpha}$$

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < -Z_{1-\alpha}$$

بنابراین در حالت کلی در آزمون $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$ فرض H_0 رد می‌شود اگر و فقط اگر $Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < -Z_{1-\alpha}$ در این حالت هم اگر فرض H_0 بصورت $H_0 : \mu \ge \mu_0$ می‌بود باز هم ناحیه بحرانی بصورت رابطه فوق بدست می‌آمد.

۱۱-۱۸ حالت سوم: آزمون فرض $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$ که در آن H_0 رد می‌شود اگر $\overline{X} > C_2$ یا $\overline{X} < C_1$ باشد بنابراین خطای نوع اول عبارتست از:

$$\alpha = p(\overline{X} < C_1 \text{ یا } \overline{X} > C_2 \mid \mu = \mu_0) \rightarrow 1-\alpha = p(C_1 < \overline{x} < C_2 \mid \mu = \mu_0)$$

$$p = \left(\frac{C_1 - \mu_0}{\frac{\sigma}{\sqrt{n}}} < \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{C_2 - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$= p\left(\frac{C_1 - \mu_0}{\frac{\sigma}{\sqrt{n}}} < Z < \frac{C_2 - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right) = 1-\alpha$$

که در آن:

$$\frac{C_2 - \mu_0}{\frac{\sigma}{\sqrt{n}}} = -\frac{C_1 - \mu_0}{\frac{\sigma}{\sqrt{n}}} < Z_{1-\alpha}$$

و در نتیجه بدست می‌آوریم:

$$C_1 = \mu_0 - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$C_2 = \mu_0 + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

و ناحیه بحرانی به صورت زیر بدست می‌آید:

$$\overline{X} > \mu_0 + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \qquad \text{یا} \qquad \overline{X} < \mu_0 - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

dalbablog.wordpress.com

که معادل است با:

$$\left|\frac{X'-\mu_{\circ}}{\frac{\sigma}{\sqrt{n}}}\right| > Z_{1-\frac{\alpha}{2}}$$

بنابراین در حالت کلی در آزمون فرض $\begin{cases} H_{\circ}: \mu = \mu_{\circ} \\ H_{1}: \mu \neq \mu_{\circ} \end{cases}$ فرض $H_{\circ}$ رد می‌شود اگر وفقط اگر:

$$|Z| = \left|\frac{X'-\mu_{\circ}}{\frac{\sigma}{\sqrt{n}}}\right| > Z_{1-\frac{\alpha}{2}}$$

در مثالهای بعدی هر حالت را مورد بررسی قرار می‌دهیم.

۱۹-۱۱ مثال ۵: توان شکنندگی (مقدار نیروی لازم برای شکستن) کابلهایی که توسط یک شرکت تولید می‌شوند دارای میانگین ۱۸۰۰ پوند و انحراف معیار ۱۰۰ می‌باشند. محققان شرکت با اعمال تکنیک جدیدی که در مراحل ساخت اعمال کرده‌اند ادعا کرده‌اند که توان شکنندگی افزایش یافته است. برای آزمودن این ادعا یک نمونه ۵۰ تایی از کابلها تحت آزمون قرار می‌گیرند و میانگین توان شکنندگی ۱۸۵۰ پوند بدست می‌آید. آیا در سطح معنی‌دار ۰/۰۱ این ادعا پذیرفته است؟

حل: ابتدا هر یک ار فرضهای صفر و مقابل را تعریف می‌کنیم:

$H_{\circ}$: $\mu = ۱۸۰۰$ هیچ تغییری در توان شکنندگی رخ نداده است.

H_{1} : $\mu > ۱۸۰۰$ توان شکنندگی افزایش یافته است.

ملاحظه می‌کنید که حالت اول رخ داده است و با توجه به مطالب ارایه شده آماره آزمون و ناحیه بحرانی بصورت زیر خواهند بود:

$$Z = \frac{X'-\mu_{\circ}}{\frac{\sigma}{\sqrt{n}}} > Z_{1-\frac{\alpha}{2}}$$

که در آن:

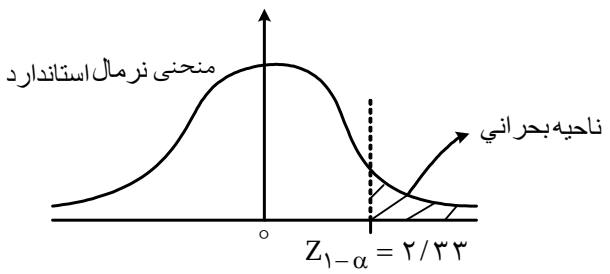
$\alpha = ۰/۰۱$

$\Rightarrow ۱-\alpha = ۱-۰/۰۱ = ۰/۹۹ \quad \Rightarrow \quad Z_{1-\alpha} = Z_{۰/۹۹} = ۲/۳۳$

یعنی ناحیه بحرانی بصورت $Z > ۲/۳۳$ می‌باشد و اگر مقدار آماره Z بزرگتر از ۲/۳۳ باشد فرض $H_{\circ}$ را رد می‌کنیم و به عبارتی ادعای محققان را می‌پذیریم. مقدار آماره Z برابر است با:

$$Z = \frac{X'-\mu_{\circ}}{\frac{\sigma}{\sqrt{n}}} >= \frac{۱۸۵۰-۱۸۰۰}{\frac{۱۰۰}{\sqrt{۵۰}}} = ۳/۵۵$$

بنابراین در سطح معنی‌دار $\alpha = ۰/۰۱$ نتایج نشان می‌دهند که $Z > ۲/۳۳$ می‌باشد و در نتیجه ادعای محققان را در مورد افزایش توان شکنندگی کابلها می‌پذیریم. در نمودار زیر ناحیه رد فرض $H_{\circ}$ را ملاحظه می‌کنید:



dalbablog.wordpress.com

۱۱-۲۰ مثال ۶: عمر متوسط ۱۰۰ عدد از لامپهای مهتابی تولید شده توسط یک کارخانه برابر ۱۵۷۰ ساعت با انحراف معیار ۱۲۰ ساعت بدست آمده است. کارخانه تولید کننده ادعا می‌کند که عمر متوسط لامپها برابر ۱۶۰۰ ساعت می‌باشد در حالی که مصرف کنندگان این ادعا را قبول ندارند. صحت ادعای کارخانه سازنده را در سطح $\alpha = ۰/۰۵$ و $\alpha = ۰/۰۱$ بررسی کنید؟

حل: در این حالت دو فرض زیر را پیش رد داریم:

$$\begin{cases} H_۰ : \mu = ۱۶۰۰ \\ H_۱ : \mu \neq ۱۶۰۰ \end{cases}$$

ملاحظه می‌کنید که در این حالت یک آزمون دو طرفه انجام می‌گیرد که در آن آماره آزمون و ناحیه بحرانی بصورت زیر می‌باشد:

$$|Z| = \left| \frac{X' - \mu_۰}{\frac{\sigma}{\sqrt{n}}} \right| > Z_{۱-\frac{\alpha}{۲}}$$

الف) ابتدا آزمون فرض را در سطح معنی‌دار $\alpha = ۰/۰۵$ انجام می‌دهیم:

$$\alpha = ۰/۰۵ \quad \rightarrow \quad ۱-\frac{\alpha}{۲} = ۰/۹۷۵$$

$$\rightarrow \quad Z_{۱-\frac{\alpha}{۲}} = Z_{./۹۷۵} = ۱/۹۶$$

بنابراین فرض $H_۰$ را رد می‌کنیم اگر $|Z| > ۱/۹۶$ باشد یا به عبارتی:

$$|Z| > ۱/۹۶ \quad \text{یا} \quad Z < -۱/۹۶$$

۱۱-۲۱ حال مقدار آماره آزمون را بدست می‌آوریم:

ابتدا توجه کنید که مقدار واقعی انحراف معیار طول عمر لامپها را نداریم بنابراین از واریانس یا انحراف معیار $\overline{X}$ برای تخمین واریانس واقعی طول عمر لامپها استفاده می‌کنیم:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{۱۲۰}{\sqrt{۱۰۰}} = ۱۲$$

به این ترتیب آماره آزمون بصورت زیر بدست می‌آید:

$$Z = \frac{X' - \mu_۰}{\frac{\sigma}{\sqrt{n}}} > \frac{۱۵۷۰ - ۱۶۰۰}{۱۲} = -۲/۵$$

ملاحظه می‌کنید که ۲/۵- خارج از بازه $(۱/۹۶ \, , \, -۱/۹۶)$ قرار دارد بنابراین $H_۰$ را در سطح معنی‌دار ۰/۰۵ رد می‌کنیم یعنی ادعای مصرف کنندگان در مورد عدم صحت میانگین طول عمر مطرح شده توسط کارخانه سازنده، صحیح می‌باشد.

ب) حال آزمون فرض را در سطح معنی‌دار ۰/۰۱ انجام می‌دهیم:

$$\alpha = ۰/۰۱ \quad \rightarrow \quad ۱-\frac{\alpha}{۲} = ۰/۹۹۵$$

$$\rightarrow \quad Z_{۱-\frac{\alpha}{۲}} = Z_{./۹۹۵} = ۲/۵۸$$

در این حالت اگر آماره آزمون در خارج از بازه $(۲/۵۸ \, , \, -۲/۵۸)$ قرار داشته باشد $H_۰$ را رد می‌کنیم. مقدار آماره آزمون در این حالت برابر همان مقدار بدست آمده در بند الف مثال می‌باشد:

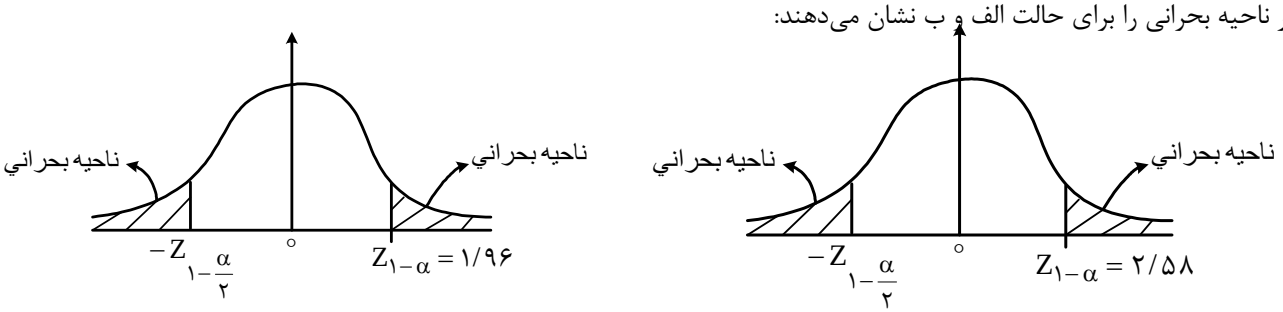
$$Z = -۲/۵ \quad \Rightarrow \quad |Z| > ۲/۵۸$$

۱۱-۲۲ از آنجا که مقدار Z برابر ۲/۵ می‌باشد و در بازه $(۲/۵۸ \, , \, -۲/۵۸)$ قرار دارد بنابراین در سطح معنی‌دار $\alpha = ۰/۰۱$ و فرض $H_۰$ را می‌پذیریم.

dalbablog.wordpress.com

ملاحظه می‌کنید که با تغییر سطح معنی‌داری و به دنبال آن تغییر ناحیه بحرانی، این احتمال وجود دارد که تصمیم‌گیری در مورد رد یا پذیرش فرض $H_\circ$ کاملاً عوض شود. توجه کنید که در حالت دوم $(\alpha = ۰/۰۱)$ فرض $H_\circ$ را می‌پذیریم اما این به معنی رد فرض $H_۱$ نمی‌باشد بلکه در این حالت می‌گوییم نمی‌توان در مورد رد فرض $H_۱$ نظری داد یا به بعارت دیگر هیچ تصمیمی نمی‌گیریم.

دو نمودار زیر ناحیه بحرانی را برای حالت الف و ب نشان می‌دهند:



۱۱-۲۳ آزمون فرض برای میانگین نرمال با واریانس نامعلوم

در صورتی که واریانس جامعه نامعلوم باشد نشان دادیم که آماره $T = \dfrac{\overline{X} - \mu_\circ}{\frac{S}{\sqrt{n}}}$ دارای توزیع t با $n-۱$ درجه آزادی است، بنابراین در این حالت و با توجه به روشهای ارایه شده برای بدست آوردن ناحیه بحرانی در بخش قبل، می‌توان نشان داد که آماره آزمون و ناحیه بحرانی برای هرحالت بصورت زیر بدست می‌آیند:

۱- در آزمون فرض $\begin{cases} H_\circ : \mu = \mu_\circ \text{ یا } \mu \le \mu_\circ \\ H_۱ : \mu > \mu_\circ \end{cases}$ $H_\circ$ رد می‌شود اگر و فقط اگر: $T = \dfrac{\overline{X} - \mu_\circ}{\frac{S}{\sqrt{n}}} > t_{۱-\alpha}(n-۱)$

۲- در آزمون فرض $\begin{cases} H_\circ : \mu = \mu_\circ \text{ یا } \mu \ge \mu_\circ \\ H_۱ : \mu < \mu_\circ \end{cases}$ $H_\circ$ رد می‌وشد اگر و فقط اگر: $T = \dfrac{\overline{X} - \mu_\circ}{\frac{S}{\sqrt{n}}} < -t_{۱-\alpha}(n-۱)$

۳- در آزمون فرض $\begin{cases} H_\circ : \mu = \mu_\circ \\ H_۱ : \mu \ne \mu_\circ \end{cases}$ $H_\circ$ رد می‌شود اگر و فقط اگر: $|T| = \left| \dfrac{\overline{X} - \mu_\circ}{\frac{S}{\sqrt{n}}} \right| > t_{۱-\frac{\alpha}{۲}}(n-۱)$

۱۱-۲۴ مثال ۷: تعداد نامه‌های رسیده به یک شرکت در طول ۱۰ روز عبارتست از:

۱ , ۳ , ۲ , ۵ , ۶ , ۴ , ۵ , ۸ , ۹ , ۱

اگر توزیع تعداد نامه‌های رسیده از متغیر تصادفی نرمال پیروی کند، آیا در سطح معنی‌دار $\alpha = ۰/۰۱$ می‌توان ادعا نمود که بطور متوسط روزانه حداثل ۴ نامه به شرکت پست ارسال می‌شود؟

حل: فرضهای زیر را در نظر می‌گیریم: $\begin{cases} H_\circ : \mu = ۴ \\ H_۱ : \mu > ۴ \end{cases}$

واریانس جامعه نامعلوم است بنابراین برای محاسبه از آماره $T = \dfrac{\overline{X} - \mu_\circ}{\frac{S}{\sqrt{n}}}$ استفاده می‌کنیم که در آن $\overline{X}$ و S برابر هستند با:

dalbablog.wordpress.com

$$\overline{X} = \frac{۱}{۱۰}(۱+۳+۲+۵+۶+۴+۵+۸+۹+۱) = ۴/۴$$

$$S^۲ = \frac{۱}{n-۱}\left(\sum_{i=۱}^{n} X_i^۲ - \frac{\overline{X}^۲}{n}\right) = \frac{۱}{۹}\left\{(۱+۹+۴+۲۵+۳۶+۱۶+۲۵+۶۴+۸۱+۱) - \frac{(۴/۴)^۲}{۱۰}\right\}$$

$$= \frac{۱}{۹}(۲۶۲ - \frac{(۴/۴)^۲}{۱۰}) = ۲۸/۸ \quad \Rightarrow \quad S = \sqrt{۲۹} = ۵/۳۶$$

فرض $H_۰$ در صورتی رد می‌شود که:

$$T > t_{۱-\alpha}(n-۱)$$

$$\alpha = ۰/۰۱ \quad \rightarrow \quad ۱-\alpha = ۰/۹۹$$

$$t_{۱-\alpha} = t_{۰/۹۹}(n-۱) = t_{۰/۹۹}(۹) = ۲/۸۲$$

مقدار آماره آزمون عبارتست از:

$$T = \frac{\overline{X} - \mu_۰}{\frac{S}{\sqrt{n}}} = \frac{۴/۴ - ۴}{\frac{۵/۳۶}{\sqrt{۱۰}}} = ۰/۲۳$$

از آنجا که مقدار آماره آزمون کمتر از ۲/۸۲ می‌باشد بنابراین فرض $H_۰$ را می‌پذیریم.

۲۵-۱۱ ۱۱ . ۹ آزمون فرض برای واریانس یک جامعه نرمال

در صورتی که بخواهیم ادعاهایی را در مورد واریانس یک جامعه نرمال بررسی کنیم می‌دانیم که آماره $\chi^۲ = \frac{(n-۱)S^۲}{\sigma^۲}$ دارای توزیع خی دو با $n-۱$ درجه آزادی است بنابراین می‌توان حالتهای زیر را برای هر حالت بدست آورد:

۱- در آزمون $\begin{cases} H_۰ : \sigma^۲ = \sigma_۰^۲ \text{ یا } \sigma^۲ \le \sigma_۰^۲ \\ H_۱ : \sigma^۲ > \sigma_۰^۲ \end{cases}$ $H_۰$ رد می‌شود اگر و فقط اگر: $\chi^۲ = \frac{(n-۱)S^۲}{\sigma_۰^۲} > \chi^۲_{۱-\alpha}(n-۱)$

۲- در آزمون فرض $\begin{cases} H_۰ : \sigma^۲ = \sigma_۰^۲ \text{ یا } \sigma^۲ \ge \sigma_۰^۲ \\ H_۱ : \sigma^۲ < \sigma_۰^۲ \end{cases}$ $H_۰$ رد می‌شود اگر و فقط اگر: $\chi^۲ = \frac{(n-۱)S^۲}{\sigma_۰^۲} < \chi^۲_{\alpha}(n-۱)$

۳- در آزمون فرض $\begin{cases} H_۰ : \sigma^۲ = \sigma_۰^۲ \\ H_۱ : \sigma^۲ \ne \sigma_۰^۲ \end{cases}$ $H_۰$ رد می‌شود اگر و فقط اگر

:

$$\chi^۲ = \frac{(n-۱)S^۲}{\sigma_۰^۲} > \chi^۲_{۱-\frac{\alpha}{۲}}(n-۱) \quad \text{یا} \quad = \chi^۲ = \frac{(n-۱)S^۲}{\sigma_۰^۲} < \chi^۲_{\frac{\alpha}{۲}}(n-۱)$$

۲۶-۱۱ مثال ۹: یک کارخانه تولید نوشابه ماشین آلات جدیدی برای پرنمودن شیشه‌ها خریداری نموده است که ادعا شده است این ماشین‌ها با انحراف معیار ۱۰ میلی لیتر شیشه‌ها را پر می‌کنند. برای بررسی صحت ادعا یک نمونه تصادفی ۱۰ تایی از شیشه‌های پر شده را انتخاب می‌کنیم و نتایج $\overline{X} = ۲۵۵$ و $S^۲ = ۱۸۰$ بدست امده است. در سطح معنی‌دار $\alpha = ۰/۰۵$ آیا ادعای مطرح شده صحیح است؟

حل: فرضها عبارتند از:

dalbablog.wordpress.com

$$\begin{cases} H_{\circ} : \sigma^{۲} = ۱۰۰ \\ H_{۱} : \sigma^{۲} \neq ۱۰۰ \end{cases}$$

با توجه به آزمونها آماره آزمون و ناحیه بحرانی بصورت زیر خواهند بود:

$$\chi^{۲}_{۱-\frac{\alpha}{۲}}(n-۱) < \chi^{۲} = \frac{(n-۱)S^{۲}}{\sigma^{۲}_{\circ}} \quad \text{یا} \quad \chi^{۲} = \frac{(n-۱)S^{۲}}{\sigma^{۲}_{\circ}} < \chi^{۲}_{\frac{\alpha}{۲}}(n-۱)$$

در سطح معنی‌دار ۰/۰۵ داریم:

$$\alpha = ۰/۰۵ \qquad \rightarrow \qquad ۱-\frac{\alpha}{۲} = ۰/۹۷۵ \qquad \frac{\alpha}{۲} = ۰/۰۲۵$$

پس:

$$\chi^{۲}_{۱-\frac{\alpha}{۲}}(n-۱) = \chi^{۲}_{۰/۹۷۵}(۹) = ۱۹$$

$$\chi^{۲}_{\frac{\alpha}{۲}}(n-۱) = \chi^{۲}_{۰/۰۲۵}(۹) = ۲/۷$$

$$\chi^{۲} = \frac{(n-۱)S^{۲}}{\sigma^{۲}_{\circ}} = \frac{۹(۱۸۰)}{۱۰۰} = ۱۶/۲$$

از آنجا که آماره $\chi^{۲} = ۱۶/۲$ درون بازه $(۱۹ \ , \ ۲/۷)$ قرار دارد بنابراین ادعای $H_{\circ}$ رد نمی‌شود و در نتیجه ادعا مطرح شده در سطح معنی‌دار ۰/۰۵ قابل قبول است.

۲۷-۱۱ ۱۱ . ۹ آزمون فرض برای تفاضل میانگین‌ها

یک نمونه تصادفی $n_{۱}$ تایی از جامعه نرمال با میانگین $\mu_{۱}$ و واریانس $\sigma^{۲}_{۱}$ انتخاب می‌کنیم، همچنین یک نمونه $n_{۲}$ تایی از جامعه نرمال دیگری با میانگین $\mu_{۲}$ و واریانس $\sigma^{۲}$ انتخاب می‌کنیم بطوریکه دو نمونه از یکدیگر مستقل باشند در این حالت برای آزمون تفاضل میانگین‌های دو جامعه از آزمون‌های زیر می‌توانیم استفاده کنیم که در سه حالت اول واریانس دو جامعه معلوم و در سه حالت بعدی واریانس مجهول فرض شده است:

۱- در آزمون فرض $\begin{cases} H_{\circ} : \mu_{۱}-\mu_{۲} = d_{\circ} \ \text{یا} \ \mu_{۱}-\mu_{۲} \leq d_{\circ} \\ H_{۱} : \mu_{۱}-\mu_{۲} > d_{\circ} \end{cases}$ فرض $H_{\circ}$ رد می‌شود اگر و فقط اگر: $Z = \dfrac{(\overline{X}_{۱}-\overline{X}_{۲}) - d_{\circ}}{\sqrt{\dfrac{\sigma^{۲}_{۱}}{n_{۱}} + \dfrac{\sigma^{۲}_{۲}}{n_{۲}}}} > Z_{۱-\alpha}$

۲- در آزمون فرض $\begin{cases} H_{\circ} : \mu_{۱}-\mu_{۲} = d_{\circ} \ \text{یا} \ \mu_{۱}-\mu_{۲} \geq d_{\circ} \\ H_{۱} : \mu_{۱}-\mu_{۲} < d_{\circ} \end{cases}$ فرض $H_{\circ}$ رد می‌شود اگر و فقط اگر: $Z = \dfrac{(\overline{X}_{۱}-\overline{X}_{۲}) - d_{\circ}}{\sqrt{\dfrac{\sigma^{۲}_{۱}}{n_{۱}} + \dfrac{\sigma^{۲}_{۲}}{n_{۲}}}} < Z_{۱-\alpha}$

۳- در آزمون فرض $\begin{cases} H_{\circ} : \mu_{۱}-\mu_{۲} = d_{\circ} \\ H_{۱} : \mu_{۱}-\mu_{۲} \neq d_{\circ} \end{cases}$ فرض $H_{\circ}$ رد می‌شود اگر و فقط اگر: $|Z| = \left| \dfrac{(\overline{X}_{۱}-\overline{X}_{۲}) - d_{\circ}}{\sqrt{\dfrac{\sigma^{۲}_{۱}}{n_{۱}} + \dfrac{\sigma^{۲}_{۲}}{n_{۲}}}} \right| > Z_{۱-\alpha}$

۲۸-۱۱ در حالتی که واریانسهای دو جامعه نامعلوم اما برابر باشند $(\sigma^{۲}_{۱} = \sigma^{۲}_{۲} = \sigma^{۲})$ روابط آزمونهای فرض بصورت زیر خواهند بود:

dalbablog.wordpress.com

۱- در آزمون فرض $\begin{cases} H_۰ : \mu_۱ - \mu_۲ = d_۰ \text{ یا } \mu_۱ - \mu_۲ \le d_۰ \\ H_۱ : \mu_۱ - \mu_۲ > d_۰ \end{cases}$ فرض $H_۰$ رد می‌شود اگر و فقط اگر:

$$T = \frac{(\overline{X}_۱ - \overline{X}_۲) - d_۰}{S_P \sqrt{\frac{۱}{n_۱} + \frac{۱}{n_۲}}} > t_{۱-\alpha}(n_۱ + n_۲ - ۲)$$

که در آن $S_P^۲ = \frac{(n_۱ - ۱) S_۱^۲ + (n_۲ - ۱) S_۲^۲}{n_۱ + n_۲ - ۲}$ می‌باشد.

۲- در آزمون فرض $\begin{cases} H_۰ : \mu_۱ - \mu_۲ = d_۰ \text{ یا } \mu_۱ - \mu_۲ \ge d_۰ \\ H_۱ : \mu_۱ - \mu_۲ < d_۰ \end{cases}$ فرض $H_۰$ رد می‌شود اگر و فقط اگر:

$$T = \frac{(\overline{X}_۱ - \overline{X}_۲) - d_۰}{S_P \sqrt{\frac{۱}{n_۱} + \frac{۱}{n_۲}}} < t_{۱-\alpha}(n_۱ + n_۲ - ۲)$$

۳- در آزمون فرض $\begin{cases} H_۰ : \mu_۱ - \mu_۲ = d_۰ \\ H_۱ : \mu_۱ - \mu_۲ \ne d_۰ \end{cases}$ فرض $H_۰$ رد می‌شود اگر و فقط اگر:

$$|T| = \left| \frac{(\overline{X}_۱ - \overline{X}_۲) - d_۰}{S_P \sqrt{\frac{۱}{n_۱} + \frac{۱}{n_۲}}} \right| > t_{۱-\frac{\alpha}{۲}}(n_۱ + n_۲ - ۲)$$

۲۹-۱۱ مثال ۱۰: یک امتحان از دو کلاس A و B گرفته شده است. کلاس A شامل ۴۰ دانشجو و کلاس B شامل ۵۰ دانشجو می‌باشد. میانگین نمرات در کلاس A ۷۴ با انحراف معیار ۸ و در کلاس B ۷۸ با انحراف معیار ۷ می‌باشد آیا نمرات دانشجویان این دو کلاس متفاوت معنی‌داری در سطوح ۰/۰۵ و ۰/۰۱ با هم دارند؟

حل: هر یک از دو کلاس را دو جامعه با میانگین‌های $\mu_۱$ و $\mu_۲$ در نظر می‌گیریم داریم:

$\overline{X}_۱ = ۷۴ \qquad \sigma_۱ = ۸$

$\overline{X}_۲ = ۷۸ \qquad \sigma_۲ = ۷$

$n_۱ = ۴۰$

$n_۲ = ۵۰$

فرض‌های آزمون عبارتند از:

$H_۰$: $\mu_۱ = \mu_۲$ اختلاف تنها ناشی از شانس است.

$H_۱$: $\mu_۱ \ne \mu_۲$ بین دو کلاس اختلاف معنی‌داری وجود دارد.

در این حالت آماره آزمون و ناحیه بحرانی عبارتند از:

$$|Z| = \left| \frac{(\overline{X}_۱ - \overline{X}_۲) - d_۰}{\sqrt{\frac{\sigma_۱^۲}{n_۱} + \frac{\sigma_۲^۲}{n_۲}}} \right| > Z_{۱-\alpha}$$

که در این مثال مقدار $d_۰ = ۰$ می‌باشد.

dalbablog.wordpress.com

۳۰-۱۱ الف) در سطح معنی‌دار ۰/۰۵ داریم:

$$\alpha = ۰/۰۵ \quad \rightarrow \quad ۱-\frac{\alpha}{۲} = ۰/۹۷۵$$

$$Z_{۱-\frac{\alpha}{۲}} = Z_{۰/۹۷۵} = ۱/۹۶$$

در این حالت فرض $H_۰$ رد می‌شود اگر $|Z| > ۱/۹۶$ حال داریم:

$$Z = \frac{(\overline{X}_۱ - \overline{X}_۲) - d_۰}{\sqrt{\frac{\sigma_۱^۲}{n_۱} + \frac{\sigma_۲^۲}{n_۲}}} = \frac{۷۴ - ۷۸}{\sqrt{\frac{۸۲}{۴۰} + \frac{۷۲}{۵۰}}} = \frac{-۴}{۱/۶۰۶} = -۲/۴۹$$

از آنجا که $Z = -۲/۴۹$ در بازه $(۱/۹۶ , -۱/۹۶)$ قرار ندارد بنابراین در سطح معنی‌دار ۰/۰۵ اختلاف معنی‌داری بین دو کلاس A و B وجود دارد و فرض $H_۰$ رد می‌شود. به این ترتیب به احتمال بیشتر کلاس دوم دارای نتایج بهتری می‌باشد.

ب) برای سطح معنی‌دار ۰/۰۱ داریم:

$$\alpha = ۰/۰۱ \quad \rightarrow \quad ۱-\frac{\alpha}{۲} = ۰/۹۹۵$$

$$Z_{۱-\frac{\alpha}{۲}} = Z_{۰/۹۹۵} = ۲/۵۸$$

در این حالت مقدار آماره آزمون برابر با مقدار بدست آمده از بند الف می‌باشد که برابر $-۲/۴۹$ می‌باشد. و از آنجا که $-۲/۴۹$ در بازه $(۲/۵۸ , -۲/۵۸)$ قرار دارد بنابراین در سطح معنی‌دار ۰/۰۱ اختلاف معنی‌داری بین دو کلاس وجود ندارد. توجه کنید که در حالت کلی نتایج حاصل از سطح معنی‌دار ۰/۰۵ را ملاک تصمیم‌گیری قرار می‌دهند. زیرا آماردانان نشان داده‌اند که تقریباً بهترین سطح برای ملاک بودن در تصمیم‌گیری‌ها، سطح معنی‌دار ۰/۰۵ می‌باشد.

۳۱-۱۱ مثال ۱۱: در کشاورز A و B در مزرعه‌های خود گندم می‌کارند، کشاورز A از نوعی ضد آفت جدید استفاده می‌کند. میانگین برداشت محصول در هر کیلومتر مربع از ۱۲ کیلومتر مربع از زمین کشاورز A برابر ۱۳۹ کیلو با انحراف معیار ۱۰ کیلو می‌باشد و در زمین کشاورز B برابر ۱۳۱ کیلو با انحراف معیار ۱۱ کیلو می‌باشد. آیا می‌توان ادعا کرد که در سطح معنی‌دار ۰/۰۵ و ۰/۰۱.

حل: فرض آزمون عبارتند از:

$H_۰$: $\mu_۱ = \mu_۲$ اختلاف در تولید تنها ناشی از شانس است.

$H_۱$: $\mu_۱ > \mu_۲$ ضد آفت در افزایش تولید موثر بوده است.

$\mu_۱$ میانگین تولید محصول کشاورز A و $\mu_۲$ میانگین تولید محصول کشاورز B می‌باشد.

آماره آزمون و ناحیه بحرانی در این حالت عبارتند از:

$$T = \frac{(\overline{X}_۱ - \overline{X}_۲) - d_۰}{S_P\sqrt{\frac{۱}{n_۱} + \frac{۱}{n_۲}}} > t_{۱-\alpha}(n_۱ + n_۲ - ۲)$$

الف) در سطح‌معنی‌دار ۰/۰۵ داریم:

$$\alpha = ۰/۰۵$$

$$۱-\alpha = ۰/۹۵ \qquad t_{۱-\alpha}(n_۱ + n_۲ - ۲) = t_{۰/۹۵}\ (۱۲+۱۲-۲)$$

$$= t_{۰/۹۵}(۲۲) = ۱/۷۲$$

$$\overline{X}_۱ = ۱۳۹ \qquad S_۱ = ۱۰ \qquad n_۱ = ۱۲$$

$$\overline{X}_۲ = ۱۳۱ \qquad S_۲ = ۱۱ \qquad n_۲ = ۱۲$$

dalbablog.wordpress.com

مقدار S_p برابر است با:

$$S_p = \sqrt{\frac{(n_1-1)\,S_1^2 + (n_2-1)\,S_2^2}{n_1+n_2-2}} = \sqrt{\frac{11(10)^2+11(11)^2}{12+12-2}} = 10/51$$

۳۲-۱۱ حال مقدار آماره آزمون برابر می‌شود با:

$$T = \frac{\overline{X}_1 - \overline{X}_2}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{139-131}{10/51\sqrt{\frac{1}{12}+\frac{1}{12}}} = 1/85$$

با توجه به مقدار آماره آزمون $T = 1/85$ که از $t_{./95}(22) = 1/72$ بیشتر است می‌توان نتیجه گرفت که فرض $H_\circ$ در سطح معنی‌دار ۰/۰۵ رد می‌شود و به این ترتیب ضدآفت در تولید محصول بیشتر موثر بوده است.

ب) در سطح معنی‌دار ۰/۰۱ داریم:

$$\alpha = 0/01 \quad \rightarrow \quad 1-\alpha = 0/99$$

$$t_{1-\alpha}(n_1+n_2-2) = t_{./99}(22) = 2/51$$

از آنجا که $T = 1/85 \not> 2/51$ بنابراین فرض $H_\circ$ رد نمی‌شود و تفاوت معنی‌داری در سطح ۰/۰۱ در تولید محصول وجود ندارد. اما از آنجا که ۰/۰۵ را ملاک تصمیم‌گیری می‌گیریم بنابراین در حالت کلی ضد آفت در تولید محصول بیشتر موثر بوده است.

۳۳-۱۱ . ۱۰ آزمون فرض برای واریانسهای دو جامعه

اگر دو جامعه نرمال داشته باشیم و n_1 نمونه از جامعه اول با انحراف معیار S_1 و n_2 نمونه با انحراف معیار S_2 از جامعه دوم انتخاب کنیم در این صورت برای انجام آزمونهایی در مورد واریانسهای دو جامعه می‌توانیم از آماره و ناحیه بحرانی زیر استفاده کنیم:

۱- در آزمون فرض $\begin{cases} H_\circ : \sigma_1^2 = \sigma_2^2 \text{ یا } \sigma_1^2 \le \sigma_2^2 \\ H_1 : \sigma_1^2 > \sigma_2^2 \end{cases}$ فرض $H_\circ$ رد می‌شود اگر و فقط اگر: $F = \dfrac{S_1^2}{S_2^2} > F_{1-\alpha}(n_1-1,\, n_2-1)$

۲- در آزمون فرض $\begin{cases} H_\circ : \sigma_1^2 = \sigma_2^2 \text{ یا } \sigma_1^2 \ge \sigma_2^2 \\ H_1 : \sigma_1^2 < \sigma_2^2 \end{cases}$ فرض $H_\circ$ رد می‌شود اگر و فقط اگر: $F = \dfrac{S_1^2}{S_2^2} < F_{\alpha}(n_1-1,\, n_2-1)$

۳- در آزمون فرض $\begin{cases} H_\circ : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \ne \sigma_2^2 \end{cases}$ فرض $H_\circ$ رد می‌شود اگر و فقط اگر:

$$F = \frac{S_1^2}{S_2^2} < F_{\frac{\alpha}{2}}(n_1-1,\, n_2-1) \text{ یا } F = \frac{S_1^2}{S_2^2} > F_{1-\frac{\alpha}{2}}(n_1-1,\, n_2-1)$$

۳۴-۱۱ مثال ۱۲: استادی یک درس را در دو کلاس A و B تدریس می‌کند. کلاس A تعداد ۱۶ دانشجو و کلاس B تعداد ۲۵ دانشجو دارد. استاد از هر دو کلاس یک امتحان را می‌گیرد. میانگین نمرات دانشجویان در هر کلاس تقریباً برابر است اما واریانس کلاس A برابر ۸۶/۴ و واریانس کلاس B برابر ۱۵۰ می‌باشد. (نمرات از ۱۰۰ واحد می‌باشند) آیا در دو سطح معنی‌دار ۰/۰۵ و ۰/۰۱ می‌توان نتیجه گرفت که واریانس کلاس B از واریانس کلاس A بیشتر می‌باشد؟

حل: فرضهای آزمون عبارتند از:

$H_\circ$: $\sigma_1 = \sigma_2$ اختلاف واریانس‌ها ناشی از شانس می‌باشد.

dalbablog.wordpress.com

$H_۱$: $\sigma_۱ > \sigma_۲$ واریانس کلاس B از واریانس کلاس A بیشتر است.

داریم:

$n_۱ = ۱۶ \qquad S_۱^۲ = ۸۶/۴$

$n_۲ = ۲۵ \qquad S_۲^۲ = ۱۵۰$

آماره آزمون و ناحیه بحرانی عبارتند از:

$$F = \frac{S_۱^۲}{S_۲^۲} < F_\alpha (n_۱ - ۱ , n_۲ - ۱)$$

الف) در سطح معنی‌دار ۰/۰۵ داریم:

$\alpha = ۰/۰۵$

$$F_\alpha (n_۱ - ۱ , n_۲ - ۱) = \frac{۱}{F_{۱-\alpha} (n_۲ - ۱ , n_۱ - ۱)}$$

$$\Rightarrow F_{./۰۵} (۱۵ , ۲۴) = \frac{۱}{F_{./۹۵} (۲۴ , ۱۵)} = \frac{۱}{۲/۱۱} = ۰/۴۷۳$$

۳۵-۱۱ حال آماره آزمون را بدست می‌آوریم:

$$F = \frac{S_۱^۲}{S_۲^۲} = \frac{۸۶/۴}{۱۵۰} = ۰/۵۷۶$$

از آنجا که $f = ۰/۵۷۶ \not< ۰/۴۷۳$ بنابراین نمی‌توانیم فرض $H_۰$ را رد کنیم و در نتیجه می‌توان گفت که واریانس دو کلاس در سطح معنی‌دار ۰/۰۵ تفاوت معنی‌داری با یکدیگر ندارند.

ب) در سطح معنی‌دار ۰/۰۱ داریم:

$\alpha = ۰/۰۱$

$$F_\alpha (n_۱ - ۱ , n_۲ - ۱) = \frac{۱}{F_{۱-\alpha} (n_۲ - ۱ , n_۱ - ۱)}$$

$$\Rightarrow F_{./۰۱} (۱۵ , ۲۴) = \frac{۱}{F_{./۹۹} (۲۴ , ۱۵)} = \frac{۱}{۲/۸۹} = ۰/۳۴۶$$

باز هم $f = ۰/۵۷۶ \not< ۰/۳۴۶$ بنابراین در سطح معنی‌دار ۰/۰۱ هم می‌توانیم $H_۰$ را رد کنیم یعنی در این حالت هم تفاوت معنی‌داری میان واریانسهای دو کلاس وجود ندارد.

dalbablog.wordpress.com

فصل دوازدهم

S-۱

۱- دگرسیون-۱

دگرسیون

در فصلهای قبلی بیشتر تجزیه و تحلیل‌های آماری روی یک صفت از جامعه یا متغیر تصادفی متمرکز بود در این فصل قصد داریم بصورت همزمان دو صفت از جامعه یا دو متغیر تصادفی را مورد بررسی قرار دهیم. این بررسی شامل پیش بینی مقادیر یکی از متغیرها از روی مقادیر متغیر دیگراست که به مساله برگشت یا دگرسیون معروف می‌باشد.

به عنوان مثال فرض کنید بخواهیم تاثیر میزان مصرف شیر را در افزایش قد بدست بیاوریم و یا بخواهیم میزان وزن فرزند را از روی وزن پدرش پیش بینی کنیم. ملاحظه می‌کنید که در این گونه مسایل دو متغیر تصادفی مورد مطالعه به نوعی به یکدیگر وابسته می‌باشند. به عبارت دقیقتر یک متغیر تصادفی مثل X را مستقل و متغیر تصادفی Y را وابسته به آن در نظر می‌گیریم و یا برعکس Y را مستقل و X را وابسته به آن در نظر می‌گیریم. آشکار است که انتخاب هر یک از دو حالت به نوع مساله بستگی دارد.

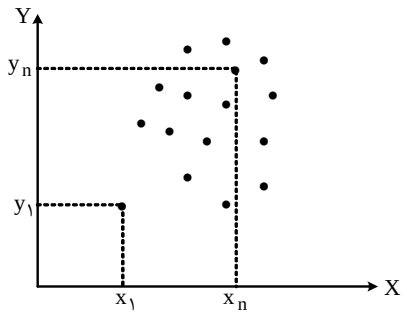
در مسایل دگرسیون برای یافتن رابطه بین متغیر تصادفی مستقل X و متغیر وابسته Y ابتدا یک نمونه nتایی از متغیر تصادفی x جمع آوری می‌کنیم که نتایج آن بصورت $x_1, x_2, \cdots, x_n$ می‌باشند. سپس مقادیر متناظر با هر یک از نمونه‌های بدست آمده (x_iها) را که همان مقادیر معادل متغیر تصادفی وابسته Y می‌باشند بدست می‌آوریم. به این ترتیب برای x_iها مقادیر متناظر $y_1, y_2, \cdots, y_n$ بدست می‌آیند. که می‌توانیم نتیجه را بصورت زوج‌های مرتب زیر نشان دهیم.

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \cdots, (x_n, y_n)$$

۲- دگرسیون-۲ به عنوان مثال مقادیر x_iها می‌توانند میزان طول قد افراد یک جامعه و مقادیر y_iها میزان مصرف شیر هر یک از نمونه‌ها باشند. به این ترتیب زوج مرتب $(۱۸۰, ۲)$ بیانگر این است که در نمونه‌گیری یکی از افراد جامعه دارای صول قد $۱۸۰\,cm$ بوده است و و وی روزانه دو لیوان شیر مصرف کرده است.

پس از بدست آمدن زوج‌های مرتب (x_i, y_i) ملاحظه می‌نید که هر زوج مرتب می‌تواند معادل یک نقطه در صفحه باشد. با رسم نقاط مورد نظر در صفحه یک تصویر کلی از رابطه X و Y بدست می‌آوریم. به شکلهای زیر که برای سه نمونه جداگانه می‌باشند توجه کنید:

نمودار ۱

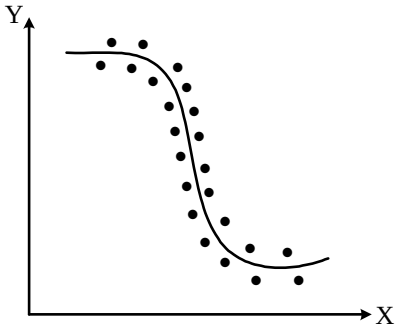


در این نمودار ملاحظه می‌کنید که زوج‌های مرتب (x_i, y_i) بصورت کاملاً پراکنده توزیع شده‌اند و به این ترتیب نتیجه می‌گیریم که رابطه‌ای بین متغیرهای تصادفی X و Y وجود ندارد.

ملاحظه می‌کنید که در این نمونه یک رابطه خطی بین مقادیر x_i و y_i وجود دارد و می‌توان نوشت $y_i = a\,x_i + b$.

نمودار ۲

dalbablog.wordpress.com



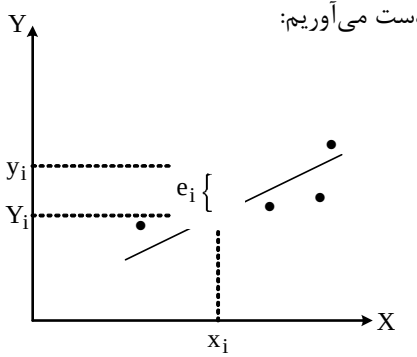
نمودار ۳

در این نمودار X و Y به یکدیگر وابسته می‌باشند اما این وابستگی از نوع غیر خطی می‌باشد.

با توجه به دو نمودار ۲ و ۳ این طور به نظر می‌رسد که در حالت کلی دو متغیر تصادفی X و Y اگر از یکدیگر مستقل نباشند یا بصورت خطی و یا بصورت غیر خطی به یکدیگر وابسته می‌باشند.
قبلاً نشان دادیم که برای اندازه‌گیری میزان وابستگی دو متغیر X و Y می‌توان از ضریب همبستگی استفاده نمود. اما در مسایل دگرسیون پیش بینی متغیر Y از روی X و یا بالعکس از اهمیت ویژه‌ای برخوردار است. بنابراین نیازمند روشی هستیم که بتوان در صورت نیاز با ثابت در نظر گرفتن یکی از مقادیر X یا Y مقدار دیگری را بدست بیاوریم. برای این منظور مفهوم بردازش منحنی را مطرح می‌کنیم.

۱- دگرسیون خطی – ۳ ۱۲ . ۱ دگرسیون خطی

اگر بین دو متغیر X و Y یک رابطه خطی وجود داشته باشد می‌توانیم یک خط را طوری رسم کنیم که نقاط (x_i , y_i) کمترین فاصله را با خط مورد نظر داشته باشند. به این عمل بردازش منحنی می‌گویند. معادله خط را بصورت $Y = a X + b$ در نظر می‌گیریم که در آن a و b مقادیر مجهول می‌باشند و در این حالت X متغیر مستقل و Y متغیر وابسته به آن در نظر گرفته می‌شود. مقادیر aو b می‌بایستی طوری محاسبه شوند که مجموع فاصله نقاط (x_i , y_i) از خط $Y = a X + b$ حداقل شود. در این حالت به $Y = a X + b$ معادله دگرسیون Y می‌گویند.
برای حداقل نمودن فاصله نقاط (x_i , y_i) از خط دگرسیون مقدار خطای e_i را مطابق نمودار زیر بدست می‌آوریم:



با توجه به نمودار Y_i مقدار پیش بینی شده توسط خط دگرسیون می‌باشد که با مقدار واقعی y_i به اندازه $e_i = | y_i - Y_i |$ فاصله دارد که این فاصله همان خطای پیش بینی می‌باشد.
برای بدست آوردن بهترین نتیجه، مجموع مربعات خطا را حداقل می‌کنیم که عبارتست از:

$$SSE = E = \sum_{i=1}^{n} (y_i - Y_i)^2$$

که در آن:

$$Y = a X + b \;\Rightarrow\; Y_i = a X_i + b \quad\Rightarrow\quad E = \sum_{i=1}^{n} (a X_i + b - Y_i)^2$$

۲- دگرسیون خطی-۴ برای مینیمم نمودن E مشتقات پاره‌ای آنرا نسبت به a و b بدست آورده و برابر صفر قرار می‌دهیم:

dalbablog.wordpress.com

$$\frac{\partial E}{\partial a} = \sum_{i=1}^{n} 2x_i (a x_i + b - y_i) = 0$$

$$\frac{\partial E}{\partial b} = \sum_{i=1}^{n} 2 (a x_i + b - y_i) = 0$$

به این ترتیب دستگاه معادلات زیر بدست می‌آید:

$$\begin{cases} \sum_{i=1}^{n} 2x_i (a x_i + b - y_i) = 0 \\ \sum_{i=1}^{n} 2 (a x_i + b - y_i) = 0 \end{cases}$$

با حل این دستگاه مقادیر مجهول a و b بدست می‌آیند. که عبارتند از:

$$a = \frac{S_{xy}}{S_{xx}} \quad , \quad b = \bar{y} - a\bar{X}$$

که در آن:

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$$

به معادله $y = a x + b$ خط دگرسیون Y روی x می‌گویند.یعنی Y از روی x بدست آمده است.

۱-مثال۱-۵ مثال ۱: در یک بررسی میزان قد پدران و فرزندان آنها بصورت جدول زیر بدست آمده است:

قد پدران X	۱۶۵	۱۶۰	۱۷۰	۱۶۳	۱۷۳	۱۵۷	۱۷۸	۱۶۸	۱۷۳	۱۷۰	۱۷۵	۱۸۰
قد فرزندان Y	۱۷۳	۱۶۸	۱۷۳	۱۶۵	۱۷۵	۱۶۸	۱۷۳	۱۶۵	۱۸۰	۱۷۰	۱۷۳	۱۷۸

مطلوبست:

الف) بردازش خطی داده‌ها

ب) رسم نمودار داده‌ها و خط دگرسیون.

ج) با توجه به معادله دگرسیون پیش بینی کنید که اگر قد پدری ۱۸۵cm باشد قد فرزند او چقدر خواهد بود؟

حل: برای بدست آوردن a و b در معادله $y = a x + b$ جدول زیر را تشکیل می‌دهیم:

x	y	x^2	xy	y^2
۱۶۵	۱۷۳	۲۷۲۲۵	۲۸۵۴۵	۲۹۹۲۹
۱۶۰	۱۶۸	۲۵۶۰۰	۲۶۸۸۰	۲۸۲۲۴
۱۷۰	۱۷۳	۲۸۹۰۰	۲۹۴۱۰	۲۹۹۲۹
۱۶۳	۱۶۵	۲۶۵۶۹	۲۶۸۹۵	۲۷۲۲۵
۱۷۳	۱۷۵	۲۹۹۲۹	۳۰۲۷۵	۳۰۶۲۵
۱۵۸	۱۶۸	۲۴۹۶۴	۲۶۵۴۴	۲۸۲۲۴
۱۷۸	۱۷۳	۳۱۶۸۴	۳۰۷۹۴	۲۹۹۲۹
۱۶۸	۱۶۵	۲۸۲۲۴	۲۷۷۲۰	۲۷۲۲۵
۱۷۳	۱۸۰	۲۹۹۲۹	۳۱۱۴۰	۳۲۴۰۰
۱۷۰	۱۷۰	۲۸۹۰۰	۲۸۹۰۰	۲۸۹۰۰
۱۷۵	۱۷۳	۳۰۶۲۵	۳۰۲۷۵	۲۹۹۲۹
۱۸۰	۱۷۸	۳۲۴۰۰	۳۲۰۴۰	۳۱۶۸۴
$\sum$	$\sum$	$\sum$	$\sum$	$\sum$

dalbablog.wordpress.com

۲- مثال ۱-۶

$$\overline{X}=\frac{۱}{n}\sum x_i=\frac{۱}{۱۲}\,۲۰۳۳=۱۶۹/۴۱$$

$$\overline{Y}=\frac{۱}{n}\sum y_i=\frac{۱}{۱۲}\,۲۰۶۱=۱۷۱/۷۵$$

$$S_{xy}=\sum_{i=۱}^{n} x_i\,y_i-n\,\overline{x}\,\overline{y}=۳۴۹۴۱۸-۱۲(۱۶۹/۴۱)\,(۱۷۱/۷۵)=۲۵۰/۲۵$$

$$S_{xx}=\sum x_i^{۲}-n\,\overline{x}^{۲}=۳۴۴۹۴۹-۱۲\,(۱۶۹/۴۱)^{۲}=۵۲۴/۹۱$$

$$a=\frac{S_{xy}}{S_{xx}}=\frac{۲۵۰/۲۵}{۵۲۴/۹۱}=۰/۴۷۷$$

$$b=\overline{Y}-a\,\overline{X}=۱۷۱/۷۵-۰/۴۷۷\,(۱۶۹/۴۱)=۹۰/۹$$

بنابراین معادله خط دگرسیون عبارتست از:

$$y=۰/۴۷۷\,x+۹۰/۹$$

ب) نمودار داده‌ها و خط دگرسیون عبارتست از:

ج) با توجه به خط دگرسیون می‌توان قد فرزند یک پدر با قد ۱۸۵ را به فرم زیر بدست آورد.

$$y=۰/۴۷۷\times۱۸۵+۹۰/۹=۱۷۹/۱۴۵$$

S-۳ ۱۲. ۲ ضریب همبستگی و دگرسیون

۷-۱

قبلاً ضریب همبستگی دو متغیر X و Y را بصورت زیر تعریف نمودیم:

$$\rho_{XY}=\frac{\sigma_{XY}}{\sigma_X\,\sigma_Y}$$

می‌توان با ساده نمودن معادله دگرسیون مقدار ضریب همبستگی را وارد معادله نمود:

dalbablog.wordpress.com

$$y = a x + b \qquad a = \frac{S_{xy}}{S_{xx}} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \left(\frac{\sigma_y}{\sigma_x}\right) = \rho_{xy} \frac{\sigma_y}{\sigma_x}$$

توجه کنید که در اینجا در محاسبه ρ_{xy} و σ_x و σ_y از مقادیر نمونه‌های مشاهده شده استفاده می‌شود.

به این ترتیب معادله خط دگرسیون بصورت زیر بدست می‌آید:

$$y - \bar{y} = \rho_{xy} \frac{\sigma_y}{\sigma_x} (x - \bar{x})$$

توجه کنید که همواره برای داده‌های مشاهده شده (x_i , y_i) دو معادله دگرسیون وجود دارد یک معادله بر حسب y نسبت به x می‌باشد و معادله دیگر بر حسب x نسبت به y می‌باشد معادله خط دگرسیون x روی y را می‌توان بصورت زیر بدست آورد:

$$x = \alpha y + \beta$$

$$\alpha = \frac{S_{xy}}{S_{yy}} \qquad \beta = \bar{x} - \alpha \bar{y}$$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i - n \bar{y}^2$$

۲-۸ با نوشتن معادله دگرسیون x روی y و استفاده از ضریب همبستگی بدست می‌آوریم:

$$x - \bar{x} = \rho_{xy} \frac{\sigma_x}{\sigma_y} (y - \bar{y})$$

بنابراین در حالت کلی دو خط دگرسیون y روی x و x روی y خواهیم داشت که عبارتند از:

$$y - \bar{y} = \rho_{xy} \frac{\sigma_y}{\sigma_x} (x - \bar{x})$$

$$x - \bar{x} = \rho_{xy} \frac{\sigma_x}{\sigma_y} (y - \bar{y})$$

با برابر قرار دادن $x = y$ در معادلات فوق محل تلاقی دو خط دگرسیون نقطه $\begin{cases} x = \bar{x} \\ y = \bar{y} \end{cases}$ بدست می‌آید. همچنین توجه کنید که:

$$a = \rho_{xy} \frac{\sigma_y}{\sigma_x}$$

$$\Rightarrow a\alpha = \rho_{xy} \frac{\sigma_y}{\sigma_x} \rho_{xy} \frac{\sigma_x}{\sigma_y} = \rho_{xy}^2$$

$$\alpha = \rho_{xy} \frac{\sigma_x}{\sigma_y}$$

بنابراین $a\alpha = \rho_{xy}^2$ و می‌بایستی مقداری در بازه $[0 , 1]$ داشته باشد.

۲-۹ مثال ۲: در مثال ۱ مطلوبست:

الف) محاسبه ضریب همبستگی نمونه با توجه به مقادیر محاسبه شده a و b.

ب) معادله دگرسیون x روی y از روی ضریب همبستگی.

ج) رسم نمودار داده‌ها و دو خط دگرسیون y روی x و x روی y.

د) پیش بینی قد پدر اگر قد فرزند وی ۱۷۹cm باشد.

حل: الف) ضریب همبستگی با توجه به روابط ارایه شده عبارتست از:

dalbablog.wordpress.com

$$a = \rho_{xy} \frac{\sigma_y}{\sigma_x} = ۰/۴۷۷$$

$$S_{yy} = \sum y_y^۲ - n\bar{y}^۲ = ۳۵۴۲۲۳ - ۱۲(۱۷۱/۷۵)^۲ = ۲۴۶/۲۵$$

$$\sigma_y = \sqrt{S_{yy}} = \sqrt{۲۴۶/۲۵} = ۱۵/۶۹$$

$$\sigma_x = \sqrt{S_{xx}} = \sqrt{۵۲۴/۹۱} = ۲۲/۹۱$$

$$\Rightarrow a = ۰/۴۷۷ = \rho_{xy} \frac{۱۵/۶۹}{۲۲/۹۱} \quad \Rightarrow \quad \rho_{xy} = ۰/۶۹۶$$

ب) معادله دگرسیون x روی y با توجه به ρ_{xy} برابر است با:

$$x = \alpha y + \beta = \rho_{xy} \frac{\sigma_x}{\sigma_y} (y - \bar{y}) + \bar{x} = ۰/۶۹۶ (\frac{۲۲/۹۱}{۱۵/۶۹}) (y - ۱۷۱/۷۵) + ۱۶۹/۴۱$$

$$\Rightarrow \quad x = ۱/۰۱۶ y - ۵/۱۲$$

ج) نمودار داده‌ها و دو خط دگرسیون و روی x و x روی y عبارتست از:

د) با استفاده از خط دگرسیون x روی y مقدار x را برای $y = ۱۷۹$ پیش بینی می‌کنیم:

$$x = ۱/۰۱۶ y - ۵/۱۲ = ۱/۰۱۶ (۱۷۹) - ۵/۱۲ = ۱۷۶/۷۴۴$$

۱۲. ۳ دگرسیون منحنی‌های چند جمله‌ای

با تعمیم روشی که برای برازش داده‌ها بصورت خطی ارایه شد به سادگی می‌توان به داده‌ها یک منحنی چند جمله‌ای بفرم $y = a_n x^n + a_{n-۱} x^{n-۱} + \cdots + a_۱ x + a_۰$ برازش داد. این کار را برای یک منحنی درجه دوم نشان می‌دهیم:

$$Y = a x^۲ + b x + c$$

$$E = \sum_{i=۱}^{n} (y_i - Y_i)^۲ = \sum_{i=۱}^{n} (a x_i^۲ + b x_i + c - y_i)^۲$$

حال مجدداً با مشتق‌گرفتن نسبت به مقادیر مجهول a و b و c و برابر صفر قرار دادن معادلات یک دستگاه بدست می‌آوریم که با حل آن a و b و c بدست می‌آیند:

dalbablog.wordpress.com

$$\frac{\partial E}{\partial a} = \sum_{i=1}^{n} 2x_i^2 (a x_i^2 + b x_i + c - y_i) = 0$$

$$\frac{\partial E}{\partial b} = \sum_{i=1}^{n} 2x_i (a x_i^2 + b x_i + c - y_i) = 0$$

$$\frac{\partial E}{\partial c} = \sum 2 (a x_i^2 + b x_i + c - y_i) = 0$$

با داشتن مقادیر نمونه‌های (x_i, y_i) به سادگی می‌توان دستگاه فوق را حل نموده و مقادیر مجهول a و b و c را بدست آورد.

۱۱ - estenbate amari

استنباط آماری بر روی ضرایب دگرسیونی

در بخش قبل بر اسا نمونه (x_1, y_1) , (x_2, y_2), ⋯ , (x_n, y_n) مدل ساده دگرسیونی را معرفی نمودیم و به کمک روش کمترین مربعات برآوردگرهای $\hat{a} = \bar{y} - \hat{b}\bar{x}$, $\hat{b} = \frac{S_{xy}}{S_{xx}}$ را معرفی نمودیم. برای $y_i = a + b x_i + E_i$

در این بخش می‌خواهیم مدل احتمالی دگرسیون را معرفی کرده و بر اساس آن برخی از فرمهای آماری را بر روی a و b آزمون کنیم.
فرض کنید در مدل ساده دگرسیونی خطا یعنی E_i متغیر تصادفی با توزیع نرمال به فرم زیر باشد:

$$E_i \sim N(0, \sigma^2)$$

همچنین متغیرهای تصادفی E_i به ازای $i = 1, 2, \cdots, n$ را مستقل از هم در نظر می‌گیریم در این صورت چون Y_i یک ترکیب خطی از متغیر تصادفی E_i می‌باشد بنابراین:

$$Y_i \sim N(a + b x_i, \sigma^2)$$

۱۲ - estenbate amari

محاسبه برآوردگرهای a و b و σ^2

به ازای هر مقدار ثابت X_i تابع چگالی متغیر تصادفی Y_i برابر سات با:

$$f_{Y_i|X_i}(y_i|x_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_i - (a + b x_i)}{\sigma}\right)^2} \quad -\infty < y_i < \infty$$

می‌توان به کمک روش ماکزیمم درستنمایی که روش دیگری است ۶ برآوردگرهای a و b و σ بذای بدست آوردن برآوردگرهای ماکزیمم درستنمایی پارامترهای a و b و σ به کمک نمونه تصادفی (X_1, Y_1) , (X_2, Y_2), ⋯ , (X_n, Y_n) ابتدا از تابع درستنمایی (یا لگاریتم آن که که ساده‌تر است) نسبت به a و b و σ مشتق می‌گیریم.
عبارتهای حاصل را برابر صفر قرار می‌دهیم، و سپس دستگاه معادلات حاصل را حل می‌کنیم.
بنابراین با مشتق گیری جزیی از:

$$Ln(L) = -n Ln(G) - \frac{n}{2} Ln(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} [Y_i - (a + b x_i)]^2$$

نسبت به a و b و σ و برابر گذاشتن عبارت‌های حاصل ۶ به دست می‌آوریم:

dalbablog.wordpress.com

$$\frac{\partial Ln(L)}{\partial a} = \frac{1}{\sigma^2} + \sum_{i=1}^{n} \left[Y_i - (a + b x_i) \right] = 0$$

$$\frac{\partial Ln(L)}{\partial b} = \frac{1}{\sigma^2} + \sum_{i=1}^{n} X_i \left[Y_i - (a + b x_i) \right] = 0$$

$$\frac{\partial Ln(L)}{\partial \sigma} = \frac{-n}{\sigma} + \frac{1}{\sigma^3} + \sum_{i=1}^{n} \left[Y_i - (a + b x_i) \right]^2 = 0$$

از دو معادله اول $\hat{a}$ و $\hat{b}$ مشابه با برآوردگرهای به روش کمترین مربعات به صورت زیر به دست می‌آید.

$$\hat{b} = \frac{S_{xy}}{S_{xx}}$$

$$\hat{a} = \overline{Y} - \hat{b}\,\overline{X}$$

با قرار دادن $\hat{a}$ و $\hat{b}$ در معادله سوم برآوردگر $\hat{\sigma}^2$ برابر می‌شود با:

$$\hat{\sigma}^2 = \frac{S_{yy} - \hat{b}\, S_{xy}}{n} = \frac{SSE}{n}$$

اگر در براوردگر ماکزیمم درستنمایی $\hat{\sigma}^2$، n را به $n-2$ تبدیل کنیم آنگاه برآوردگر $S^2 = \frac{SSE}{n-2}$ ناریب خواهد شد یعنی در این حالت:

$E\left[S^2\right] = E\left[\frac{SSE}{n-3}\right] = \sigma^2$ می‌شود.

۱۳ mesal

مثال ۱۳: فرض کنیم یک کمپانی می‌خواهد تأثیر تبلیغات را در فروش کالاهای تولید شده بررسی کند بدین منظور داده‌های زیر را بعد از ۱۰ ماه بدست می‌آورد. این داده‌ها در جدول زیر مرتب شده است.

به کمک داده‌های بدست آمده برآورد a و b و σ^2 را بدست اورید.

حل:

dalbablog.wordpress.com

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n} = 9/28 - \frac{(9/4)^2}{10} = 0/444$$

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n} = 924/8 - \frac{(9/4)(959)}{10} = 23/34$$

$$\overline{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{959}{10} = 95/9 \quad , \quad \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{9/4}{10} = 0/94$$

پس:

$$\hat{b} = \frac{S_{xy}}{S_{xx}} = \frac{23/34}{0/444} = 52/5676 \approx 52/57$$

$$\hat{a} = \overline{Y} - \hat{b}\overline{X} = 95/9 - (52/5676)(0/94) \approx 46/49$$

بنابراین معادله خط دگرسیون به صورت: $y = 46/49 + 52/57\ x$ می‌شود.

۱۴Mesal

برای محاسبه برآوردگر نااریب σ^2 یعنی S^2 به صورت زیر عمل می‌کنیم:

$$S_{yy} = \sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n} = 93/569 - \frac{(959)^2}{10} = 1600/9$$

با جایگذاری S_{yy} و $\hat{b}$ و S_{xy} در فرمول SSE خواهیم داشت:

$$SSE = S_{xy} - \hat{b}\ S_{xy} = 1600/9 - (52/5676)(23/34) = 3700$$

پس:

$$S^2 = \frac{SSE}{n-2} = \frac{373/97}{8} = 46/750$$

حال از برآوردگرهای ماکزیمم ددرستنمایی ضرایب دگرسیون ساده در آزمون فرضهایی درباره a و b و در ساختن فاصله‌های اطمینان برای این پارامترها استفاده می‌کنیم. بدین منظور نتایج زیر را بدون اثبات می‌پذیریم:

$$\hat{a} \sim N\left(a\ ,\ \frac{\sigma^2 \sum_{i=1}^{n} X_i^2}{n\ S_{xx}}\right) \qquad (1)$$

$$\hat{b} \sim N\left(b\ ,\ \frac{\sigma^2}{S_{xx}}\right) \qquad (2)$$

(۳) اگر $S^2 = \frac{SSE}{n-2}$ آنگاه $\frac{SSE}{\sigma^2} = \frac{(n-2)S^2}{\sigma^2} \sim X^2_{(n-2)}$

(۴) $\hat{a}, S^2$ مستقل از هم هستند.

(۵) $\hat{b}, S^2$ مستقل از هم هستند.

dalbablog.wordpress.com

۱۵ – estenbate amari

استباط آماری بر روی b

رابطه خطی بین x و y اولین موضوعی است که مورد بررسی قرار می‌گیرد. بنابراین علاقمندیم بدانیم که آیا داده‌ها به اندازه کافی اطلاعات در رابطه با ارتباط خطی بودن y با x می‌دهد یا نه؟ آزمون فرضهای آماری که بر روی پارامتر b ساخته می‌شود ارتباط بین y و x را روشن می‌کند بعنوان مثال اگر y با افزایش یا کاهش x تغییر می‌کند می‌توان فرض $b = ۰$ را در مقابل فرض $b \neq ۰$ ازمون کرد. با توجه به اینکه برآوردگر $\hat{b}$ دارای توزیع $\hat{b} \sim N\left(b , \frac{\sigma^۲}{S_{xx}}\right)$ می‌باشد می‌توان از تست آماری $T = \frac{\hat{b} - b_۰}{\frac{\sigma}{\sqrt{S_{xx}}}} \sim t_{(n-۲)}$.

برای آزمون فرض‌های

(۱) $H_۰: b = b_۰$ ، $H_۱: b \neq b_۰$

(۲) $H_۰: b = b_۰$ ، $H_۱: b > b_۰$

(۳) $H_۰: b = b_۰$ ، $H_۱: b < b_۰$

استفاده نمود.

نواحی بحرانی برای آزمون فرض‌های بالا در سطح معنی‌دار α را می‌توان در جدول زیر خلاصه نمود.

هم‌چنین می‌توان با استفاده از تست آماری T با ضریب اطمینان $۱ - \alpha$ فاصله اطمینان زیر را برای پارامتر b معرفی نمود.

$$P\left(\hat{b} - t_{(۱-\frac{\alpha}{۲})}^{(n-۲)} \frac{S}{\sqrt{S_{xx}}} < b < \hat{b} + t_{(۱-\frac{\alpha}{۲})}^{(n-۲)} \frac{S}{\sqrt{S_{xx}}}\right) = ۱-\alpha$$

۱۶–مثال ۱-۴

مثال ۴: با توجه به داده‌های مثال ۱ تعیین کنید که آیا پارامتر b اختلاف معنی‌داری از مقدار ۰ دارد یا نه (بوسیله استفاده از یک مدل خطی بین فروش ماهانه و مقدار تبلیغات)

حل: می‌خواهیم آزمون فرض زیر را انجام دهیم:

$H_۰: b = b_۰$

$H_۱: b \neq b_۰$

از تست آماری:

$$T = \frac{\hat{b} - ۰}{\frac{S}{\sqrt{S_{xx}}}} \sim t_{(۱-\frac{\alpha}{۲})}^{(n-۲)} = t_{(۱-\frac{\alpha}{۲})}^{(۸-۲)} = t_{(۱-\frac{\alpha}{۲})}^{(۶)}$$

استفاده می‌کنیم. با استفاده از جدول توزیع T به ازای $\alpha = ۰/۵$ $t_{./۹۷۰}^{(۶)} = ۲/۳۰۶$ بنابراین چون:

dalbablog.wordpress.com

$$T = \frac{\hat{b}}{\frac{S}{\sqrt{S_{xx}}}} = \frac{۵۲/۵۷}{\frac{۶/۸۴}{\sqrt{۰/۴۴۴}}} = ۵/۱۲ > ۲/۳۰۶$$

باشد پس فرض $H_۰$ را رد می‌کنیم. بنابراین بر اساس این مشاهدات می‌توان گفت که هزینه‌های مربوط به تبلیغات تأثیر در پیش بینی فروش ماهانه کالا دارد. با توجه به اطلاعات داده شده می‌توان یک فاصله اطمینان با ضریب اطمینان $۱-\alpha=۰/۹۵$ نیز برای یک پارامتر b بدست آورد.

جایگذاری مقادیر $\hat{b}$ و S و S_{xx} و $t^{(n-۲)}_{(۱-\frac{\alpha}{۲})}$ در $\hat{b} \pm t^{(n-۲)}_{(۱-\frac{\alpha}{۲})} \times \frac{S}{\sqrt{S_{xx}}}$ اطمینان بدست می‌آید.

$$۵۲/۵۷ \pm ۲/۳۰۶ \frac{۶/۸۴}{\sqrt{۰/۴۴۴}} \quad \Rightarrow \quad ۵۲/۵۷ \pm ۲۳/۶۷ \quad \Rightarrow \quad [۲۸ , ۹۰ , ۷۶]$$

این فاصله نشان می‌دهد که با احتمال ۰/۹۵ مقدار پارامتر b که بوسیله برآوردگر $\hat{b}$ برآورد می‌شود را دارا می‌باشد.

۱۷-مثال ۲-۴

بطور مشابه می‌توان برای پارامتر a فاصله اطمینان و آزمون فرض ساخت. با توجه به اینکه برآوردگر $\hat{a}$ دارای توزیع $\hat{a} \sim N\left(a, \frac{\sigma^۲ \sum x_i^۲}{n S_{xx}}\right)$ می‌باشد می‌توان از تست آماری $T = \frac{\hat{a}-a}{S\sqrt{\frac{\sum x_i^۲}{n S_{xx}}}} \sim t(n-۲)$ استفاده نمود و خواص بحرانی را برای فرض‌های آماری مختلف بر روی پارامتر b در سطح معنی‌دار α در جدول زیر ساخت. همچنین یک فاصله اطمینان با ضریب اطمینان $(۱-\alpha)$ ۱۰۰% برای a بصورت:

$$P\left(\hat{a} - t^{(n-۲)}_{(۱-\frac{\alpha}{۲})} S\sqrt{\frac{\sum x_i^۲}{n S_{xx}}} < a < \hat{a} + t^{(n-۲)}_{(۱-\frac{\alpha}{۲})} S\sqrt{\frac{\sum x_i^۲}{n S_{xx}}}\right) = ۱-\alpha = ۰$$

برآورد $E[Y|X]$ مقدار میانگین Y به شرط مقدار داده شده X.

برآورد میانگین Y به شرط مقدار داده شده X یکی از مسائل مهم است که می‌بایست مورد بررسی قرار گیرد. بعنوان مثال ممکن است مسئول حفاظت جان کارگران کارخانه علاقمند باشد که متوسط حوادثی که برای کارگران اتفاق می‌افتد، به ازای ساعات خاصی از آموزش که به کارگران داده می‌شود را برآورد کند.

فرض کنید که x و y یک رابطه خطی بر اساس مدل احتمالی دگرسیونی داشته باشد بطوریکه:

$$E[Y|X] = a + b\,x$$

دیدیم که تابع چگالی شرطی متغیر تصادفی Y به ازای مقدار داده شده X برابر است با:

$$f_{Y|X}(y|x) = \frac{۱}{\sigma\sqrt{۲\pi}}\, e^{-\frac{۱}{۲\sigma^۲}(y-(a+b\,x))^۲} \qquad -\infty < y < \infty$$

۱۸-مثال ۳-۴

حال می‌خواهیم با فرض خطی بودن رابطه بین x و y فاصله اطمینان و آزمون فرض برای پارامتر $E[Y|X] = a + b\,x$ به ازای مقدار داده شده x بدست آوریم. همانطور که قبلاً اشاره شد $a + b\,x_p$ میانگین شرطی Y به شرط $X = x_p$ است و از $\hat{Y} = \hat{a} + \hat{b}\,x_p$ به عنوان یک برآورد برای $a + b\,x_p$ یاد کردیم. حال می‌خواهیم فرض $H_۰ : E\left[Y|X=x_p\right) = E_۰$ را در مقابل فرض $H_۱ : E\left[Y|X=x_p\right) \neq E_۰$ ازمون می‌کنیم می‌توانیم با توجه به تابع توزیع $\hat{Y}$ تست آماری زیر را برای آزمون فرض فوق معرفی نماییم.

dalbablog.wordpress.com

$$T = \frac{\hat{Y} - E_{\circ}}{S_{\hat{Y}}} = \frac{\hat{Y} - E_{\circ}}{\sqrt{S^{2}\left[\frac{1}{n} + \frac{(x_p - \bar{x})^{2}}{S_{xx}}\right]}} \sim t(n-2)$$

متغیر تصادفی T دارای توزیع t با $(n-2)$ درجه آزادی می‌باشد. بنابراین در سطح معنی‌دار α فرض $H_{\circ}$ زمانی رو می‌شود که $|T| > t^{(n-2)}_{(1-\frac{\alpha}{2})}$ باشد. بطور مشابه می‌توان فرض $H_{\circ}$ داده شده را در مقابل فرض‌های $H_1: E[Y|X=x_p] > E_{\circ}$ یا $H_1: E[Y|X=x_p] < E_{\circ}$ آزمون نمود که در این صورت فرض $H_{\circ}$ به ترتیب زمانی رد خواهد شد که داسته باشیم:

$$T > -t^{(n-2)}_{(1-\frac{\alpha}{2})} \quad , \quad T < -t^{(n-2)}_{(1-\frac{\alpha}{2})}$$

از تست آماری فرض شده می‌توان برای پیدا کردن فاصله اطمینان برای $E[Y|X=x_p]$ نیز استفاده نمود.

$$P(\hat{Y} - t^{(n-2)}_{(1-\frac{\alpha}{2})} \sqrt{S^{2}\left[\frac{1}{n} + \frac{(x_p - \bar{x})^{2}}{S_{xx}}\right]} < a + b x_p (\hat{Y} + t^{(n-2)}_{(1-\frac{\alpha}{2})}) \sqrt{S^{2}\left[\frac{1}{n} + \frac{(x_p - \bar{x})^{2}}{S_{xx}}\right]}$$

۱۹-مثال

مثال: از داده‌های مثال ۲ استفاده کرده و یک فاصله اطمینان با ضریب اطمینان ٪۹۵ برای $X=۱/۰$ پیدا کنید.

حل:

ابتدا $\hat{Y}$ را در نقطه $X_p=۱/۰$ تخمین می‌زنیم:

$$\hat{Y} = \hat{a} + \hat{b} x_p$$

در مثالهای قبل $\hat{a}=۴۶/۴۹$ و $\hat{b}=۵۲/۵۷$ را بدست آوردیم.

$$\hat{Y}=۴۶/۴۹ + (۵۲/۵۷)\ (۱/۰) = ۹۹/۰۶$$

پس:

فرمول مناسب برای فاصله اطمینان برابر بود با:

$$\hat{Y} \pm t^{(n-2)}_{(1-\frac{\alpha}{2})} \sqrt{S^{2}\left[\frac{1}{n} + \frac{(x_p - \bar{x})^{2}}{S_{xx}}\right]}$$

با جایگذاری مقادیر مناسب در عبارت فوق که قبلاض محاسبه شده است خواهیم داشت:

$$۹۹/۰۶ \pm (۲/۳۰۶) \sqrt{۴۶/۷۵\left[\frac{(۱/۰ - ۰/۹۴)}{۰/۴۴}\right]}$$

بعد از محاسبات انجام شده حاصل می‌شود:

$$۹۹/۰۶ \pm ۵/۱۸ \quad \Rightarrow \quad [۹۳/۸۸\ ,\ ۱۰/۲۴]$$

می‌توان از محاسبات فوق نتیجه گرفت که برای هر واحد هزینه تبلیغات ($۱۰/۰۰۰) متوسط فروش ماهانه با احتمال ٪۹۵ در فاصله $۹۳۸۸۰۰ و $۱۰۴۲۴۰۰ خواهد بود. در شکل زیر برای $E[Y|X]$ رسم شده است.

dalbablog.wordpress.com

Zarebe -۲۰

ضریب همبستگی

در بسیاری از اوقات نیاز به شاخصی داریم که چگونگی ارتباط بین دو متغیر x و y را اندازه بگیرد. این شاخص ضریب همبستگی خطی بین دو متغیر x و y نامیده می‌شود.

ضریب همبستگی خطی نمونه‌ای برآورد نامناسب برای ضریب همبستگی خطی تعریف شده در فصل‌های قبل است این معیار میزان قوت ارتباط خطی بین متغیرهای x و y را اندازه می‌گیرد و اولین بار دانشمند معروفی انگلیسی به نام کارل پیرسون آنرا معرفی نمود از این جهت به آن ضریب همبستگی خطی پیرسون نیز می‌گویطم و به فرم زیر تعریف می‌شود:

$$r = \frac{S_{xy}}{\sqrt{S_{xx}\, S_{yy}}}$$