جزوه آمار و احتمالات

(اقتصاد-مدیریت-حسابداری)

کاری از:

ساسان قاراخانی

@eghtesadsanjinovin

فصل ۱) آمار توصیفی:

۱-۱) تفاوت بین شاخه‌های آمار:

الف) آمار توصیفی: تحلیل در مورد کل داده‌های جامعه

ب) آمار توصیفی: توصیف جامعه از روی نمونه با داده‌های کمی که دارای توزیع آماری هستند.

ج) آمار ناپارامتری: دقیقاً شبیه به آمار توصیفی با این تفاوت که داده‌ها کیفی هستند و توزیع آماری ندارند.

۱-۲) انواع مقیاس اندازه‌گیری صفات:

انواع مقیاس‌ها:
- کمی
  - فاصله‌ای (صفر قراردادی): ex) درجه دماسنج
  - نسبی (صفر مطلق): ex) وزن، قد و...، که نسبت آنها فاقد واحد اندازه‌گیری می‌شود
- کیفی
  - اسمی: ex) جنسیت، نژاد و...
  - ترتیبی (رتبه‌ای): ex) تحصیلات

میانگین جدولی: درصورتی که F_i فراوانی هر دسته و f_i فراوانی نسبی هر دسته و x_i نیز وسط هر دسته باشد داریم:

$$\bar{x}=\frac{\sum F_i x_i}{\sum F_i}=\frac{\sum F_i x_i}{n}=\sum f_i x_i$$

نکته) مجموع جبری اختلاف داده‌ها از میانگین برابر صفر است: $\sum F_i(x_i-\bar{x})=0$

نکته) مجموع توان دوم اختلاف از میانگین همیشه در حداقل است یعنی داریم:

$$\forall a:\quad \sum(x_i-\bar{x})^2<\sum(x_i-a)^2$$

در جداول باز که حد پایین و بالای آن‌ها مشخص نیست بایستی از میانگین پیراسته (حذف درصدی از پایین و بالای داده‌ها) استفاده کنیم.

میانگین هندسی: در داده‌هایی که اندازه‌های آنها نسبی یا به صورت درصدی یا نرخ رشد و شاخص باشد استفاده می‌شود. توجه شود که در میانگین هندسی x_i ها بایستی به صورت نسبت یا درصد باشند.

$$\bar{X}_G=\sqrt[N]{x_1^{F_1}x_2^{F_2}\cdots x_N^{F_N}}$$

ex) درصورتی که جمعیت یک کشور در سال‌های مختلف ۴۵، ۵۰، ۵۳ و ۶۰ باشد متوسط میزان رشد جمعیت کدام است؟
بایستی اعداد را نسبی کنیم:

$$\bar{X}_G=\sqrt[3]{\frac{50}{45}\times\frac{53}{50}\times\frac{60}{53}}=\sqrt[3]{\frac{60}{45}}=1/1 \Rightarrow$$

$1/1-1=0/1$ = نرخ رشد سالیانه

ex) درصورتی که فروش در سال اول ۲ برابر سال قبل و فروش در سال دوم ۳ برابر سال اول و در سال سوم ۴ برابر سال دوم باشد به‌طور متوسط فروش چند برابر شده است؟

$$\bar{X}_G=\sqrt[3]{2\times3\times4}=\sqrt[3]{24}=2\sqrt[3]{3}$$

تذکر: درصورتی که متوسط نرخ رشد سوال می‌شد بایستی $\bar{X}_G-1$ را بدست آوریم.

ex) میزان فروش از سال ۸۸ تا ۹۲ از ۱۰۰ دلار به ۱۶۰۰ دلار رسیده است متوسط نرخ رشد سالیانه چقدر است؟

$$\bar{X}_G=\sqrt[4]{\frac{1600}{100}}=2 \Rightarrow \bar{X}_G-1=1 \rightarrow \%100$$

میانگین هارمونیک: اگر مقیاس اندازه‌گیری داده‌ها ترکیبی باشد ex) کیلومتر بر ساعت، دور در دقیقه و...، از میانگین هارمونیک استفاده می‌شود

$$\bar{X}_H=\frac{\sum F_i=N}{\frac{F_1}{x_1}+\frac{F_2}{x_2}+\cdots+\frac{F_N}{x_N}}=\frac{\sum f_i}{\frac{f_1}{x_1}+\frac{f_2}{x_2}+\cdots+\frac{f_N}{x_N}}$$

ex) ۳ ماشین کالایی را تولید می‌کنند اولی هر کالا را در ۲ دقیقه، دومی هر کالا را در ۳ دقیقه و سومی هر کالا را در ۶ دقیقه می‌سازد. اگر هر سه ماشین با هم بخواهند یک کالا را بسازند به طور متوسط در چند دقیقه خواهند ساخت؟

$$\bar{X}_H=\frac{3}{\frac{1}{2}+\frac{1}{3}+\frac{1}{6}}=3\,min$$

ex) اگر اتومبیلی یک سوم مسیر را با $60\,km/h$ و بقیه مسیر را با $80\,km/s$ و مسیر برگشت را با $100\,km/h$ حرکت کند متوسط سرعت رفت و برگشت چقدر است؟

$$\bar{X}_H=\frac{\frac{1}{3}+\frac{2}{3}+1}{\frac{\frac{1}{3}}{60}+\frac{\frac{2}{3}}{80}+\frac{1}{100}}=83/72$$

نکته) همواره رابطه‌ی زیر برقرار است:

$$\bar{X}_H<\bar{X}_G<\bar{X}$$

ex) طرز محاسبه میانه در داده‌های گسسته جدولی:

اولین دسته‌ای که فراوانی تجمعی آن از مقدار $\frac{N}{2}+\frac{1}{2}$ بیشتر است

مرکز دسته‌ها ← x_i	1	5	10	11
فراوانی F_i	8	12	10	5
فراوانی تجمعی CF_i	8	20	30	45

دسته سوم: $\frac{35}{2}+\frac{1}{2}=18 \longrightarrow x_3=10$

تذکر: درصورتی که مقدار $\frac{N}{2}+\frac{1}{2}$ مقدار صحیح نشد میانه به صورت زیر است:

$$\tilde{x}=M_d=X_L+0/5(X_{L+1}-X_L)$$

X_L: اولین مرکز دسته که فراوانی تجمعی آن از مقدار $\frac{N}{2}+\frac{1}{2}$ کوچکتر است

ex) میانه جدول روبرو چه مقداری است؟

x_i	2	8	12	15	17
F_i	20	30	10	20	20
CF_i	20	50	60	80	100

$\frac{N}{2}+\frac{1}{2}=50/5 \notin N$

$\tilde{x}=8+0/5(12-8)=10$

محاسبه میانه در داده‌های پیوسته جدولی: اولین طبقه‌ای که فراوانی تجمعی آن بیشتر از $\frac{N}{2}$ باشد دسته میانه‌دار می‌باشد

$$\tilde{x}=L+\left(\frac{\frac{N}{2}-CF_{i-1}}{F_i}\right)I=L+\left(\frac{\frac{1}{2}-Cf_{i-1}}{f_i}\right)I$$

L حد پایین دسته میانه‌دار — F_i فراوانی دسته میانه‌دار

CF_{i-1} فراوانی تجمعی دسته قبل از دسته میانه‌دار — I طول دسته.

ex) میانه جدول داده‌های پیوسته روبرو را بدست آورید؟

دسته دوم دسته میانه‌دار می‌باشد.

	1-5	5-9	9-13	13-17
F_i	5	10	3	2
CF_i	5	15	18	20

$\frac{N}{2}=10$

$\tilde{x}=5+\left(\frac{10-5}{10}\right)4=7$

نکته) همیشه مقدار $\sum|x_i-\tilde{x}|$ در حداقل است و وقتی توزیع داده‌ها نامتقارن باشد بهترین شاخص مرکزی میانه است

مد: داده‌ای که بیشترین فراوانی را دارد

در داده‌های پیوسته اگر d_1 (اختلاف فراوانی مطلق دسته مددار از دسته قبلش) و d_2 (اختلاف دسته فراوانی مطلق دسته مددار از دسته بعدش) باشد داریم:

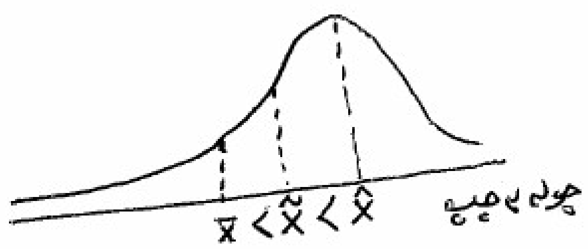
$$M_o=\hat{x}=L+\left(\frac{d_1}{d_1+d_2}\right)I$$

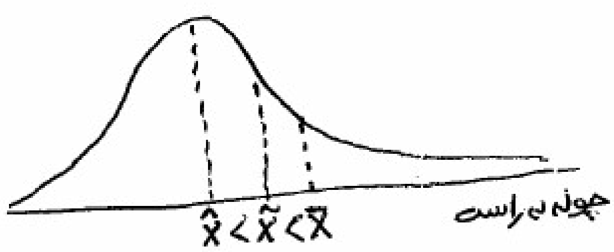
$\hat{x}=9/5+\left(\frac{4}{4+6}\right)4=9/5+1/6=11/1$

	2-5	6-9	10-13	14-17
دسته‌ها	1/5-5/5	5/5-9/5	9/5-13/5	13/5-17/5
	7	6	10 *	4

$d_1=4$ $d_2=6$

نکته) مد تنها شاخص مرکزی مورد استفاده در داده‌های کیفی است.





رابطه تجربی پیرسون: در چولگی‌های خفیف داریم:

$$\bar{x}-\hat{x}=3(\bar{x}-\tilde{x})$$

نکته) چارک دوم همان دهک پنجم، همان صدک پنجاهم و همان میانه است

برای محاسبه چندک‌ها از فرمول $(N+1)P$ برای داده‌های گسسته استفاده می‌شود. ابتدا باید داده‌ها را به صورت غیرنزولی sort کنیم سپس درصورتی که $(N+1)P$ عدد صحیح باشد عدد مورد نظر $X_{(N+1)P}$ خواهد بود و اگر $(N+1)P$ غیرصحیح باشد (k مقدار صحیح و r مقدار اعشاری آن باشد) خواهیم داشت

$$(1-r)X_k+rX_{k+1}$$

ex) چارک اول در داده‌های ۱۸۰، ۱۶۰، ۱۴۰، ۱۲۰، ۱۰۰، ۹۰، ۸۵، ۸۰ کدام است؟

$(N+1)P=(8+1)\frac{1}{4}=\frac{9}{4}=2/25$

$X_{2/25}=(1-0/25)X_2+0/25X_3=0/75\times85+0/25\times90=86/25$

محاسبه چندک در داده های پیوسته جدولی: مقدار NP را محاسبه کرده اولین دسته ای که فراوانی تجمعی آن بزرگتر از مقدار NP باشد دسته مورد نظر می باشد

حال چندک مورد نظر به صورت مقابل محاسبه می شود

$$\text{چندک مورد نظر} = L + \left(\frac{NP - CF_{i-1}}{F_i}\right) I$$

نکته) برای محاسبه دامنه تغییرات برای داده های پیوسته در صورتی که داده ها به اعداد صحیح گرد شده باشند به صورت $R = [X_{max} - X_{min}] + 1$ است

ولی در داده های گسسته همواره داریم $R = X_{max} - X_{min}$

واریانس جامعه و نمونه:

$$\sigma^2 = \frac{\sum F_i (x_i - \mu)^2}{n} = \frac{\sum F_i (x_i^2 - 2\mu x_i + \mu_i^2)}{n} = \frac{\sum F_i x_i^2}{n} - \mu^2 \quad \text{یا} \quad \sigma^2 = \sum f_i x_i^2 - \left(\sum f_i x_i\right)^2$$

$$S^2 = \frac{\sum F_i (x_i - \bar{x})^2}{n-1} = \frac{1}{n-1}\left(\sum F_i x_i^2 - \frac{\left(\sum F_i x_i\right)^2}{n}\right) \quad \text{یا} \quad S^2 = \frac{1}{n-1}\left(\sum x_i^2 - n\bar{x}^2\right)$$

نکته) اگر k جامعه با تعداد مشاهدات N_1، N_2 و ... و N_k و میانگین μ_1 و ... و μ_k و واریانس σ_1^2 و ... و σ_k^2 باشند در این صورت میانگین و واریانس کل جامعه ها به صورت زیر است:

$$\mu = \frac{N_1\mu_1 + \dots + N_k\mu_k}{N_1 + \dots + N_k} \quad \text{و} \quad \sigma^2 = \frac{\sum N_i \sigma_i^2}{N} + \frac{\sum N_i (\mu_i - \mu)^2}{N}$$

ضریب تغییرات: واحد ندارد و برای مقایسه دو سری داده که واحد اندازه گیری یکسان ندارند و یا میانگین آنها متفاوت است استفاده می شود

$$CV = \frac{\sigma}{\bar{x}}$$

$$m_n = \frac{\sum F_i (x_i - a)^n}{N}$$

گشتاور عمومی مرتبه n ام

گشتاور عمومی حول عدد دلخواه a:

الف) گشتاورها نسبت به مبدا صفر:

$$m_1 = \frac{\sum F_i x_i}{N} = \mu, \quad m_2 = \frac{\sum F_i x_i^2}{N} \quad \left\{ \begin{array}{l} \mu = m_1 \\ \sigma^2 = m_2 - m_1^2 \end{array} \right.$$

ب) گشتاور مرکزی (نسبت به میانگین):

$$\left.\begin{array}{l} \mu_1 = \dfrac{\sum F_i (x_i - \mu)}{N} = 0 \\ \mu_2 = \dfrac{\sum F_i (x_i - \mu)^2}{N} = \sigma^2 \end{array}\right\} \Rightarrow \left\{ \begin{array}{l} \mu_1 = 0 \\ \mu_2 = \sigma^2 \end{array} \right.$$

$$\mu_3 = \frac{\sum F_i (x_i - \mu)^3}{N} = \frac{\sum F_i (x_i^3 - 3x_i^2\mu + 3x_i\mu^2 - \mu^3)}{N} = m_3 - 3m_2m_1 + 3m_1^3 - m_1^3$$

$$SK(ax_i \pm b) = \pm SK(x_i) \quad \left\{ \begin{array}{ll} SK(x_i) & \text{if } a > 0 \\ -SK(x_i) & \text{if } a < 0 \end{array} \right.$$

ضریب چولگی پیرسون

$$SK = \frac{(\bar{x} - \hat{x})}{\sigma} = \frac{3(\bar{x} - \tilde{x})}{\sigma} \quad \text{یا} \quad SK = \frac{\mu_3}{\sigma^3}$$

$SK > 0 \Rightarrow$ چوله به راست

$SK < 0 \Rightarrow$ چوله به چپ

ضریب کشیدگی:

$$E = \frac{\mu_4}{\sigma^4} - 3$$

فصل ۲) احتمالات

نکته) اغلب در مسائلی که از کلمه (حداقل) استفاده می‌شود استفاده از راه حل مکمل راحت‌تر است.

Ex) چند عدد سه رقمی وجود دارد که حداقل یکی از ارقام آن عدد ۱ باشد (با تکرار و بدون تکرار)؟

کل حالات بدون تکرار $\boxed{9|9|8} = ۶۴۸$
کل حالات که شامل یک نباشد $\boxed{8|8|7} = ۴۴۸$
$\} ۶۴۸ - ۴۴۸ = ۲۰۰$

کل حالات با تکرار $\boxed{9|10|10} = ۹۰۰$
کل حالات بدون عدد یک $\boxed{8|9|9} = ۶۴۸$
$\} ۹۰۰ - ۶۴۸ = ۲۵۲$

جایگشت: راه های مرتب کردن n شی متمایز درکنار یکدیگر را گویند (خطی - دایره ای - یک در میان - با تکرار).

در مسائل جایگشت خطی تعداد حالات $n!$، جایگشت دایره‌ای $(n-1)!$ و در جایگشت یک در میان (اگر تعداد زوج باشد $2\times(n!)^2$ و اگر تعداد فرد باشد $n!\times m!$) خواهد بود همچنین در جایگشت ترکیبی دایره‌ای یک در میان (توجه شود در این حالت تعداد دو گروه با هم برابر است) $(n-1)!\,n!$ می باشد
و اگر تعدادی از شی ها یکسان باشند از جایگشت با تکرار استفاده می‌شود $\frac{n!}{n_1!\,n_2!\dots n_k!}$ که در آن $n_1+n_2+\dots+n_k=n$ می باشد.

ترتیب: انتخاب r شی از n شی متمایز که ترتیب انتخاب r شی مهم باشد $P_r = \frac{n!}{(n-r)!}$

ترکیب: انتخاب r شی از n شی متمایز که ترتیب انتخاب r شی مهم نباشد $C_r = \frac{n!}{r!(n-r)!} \rightarrow C\binom{n}{2} = \frac{n\times(n-1)}{2}$

نکته خیلی مهم: در سوالات از ما ساخت کلمات یا اعداد خواسته می‌شود که ۴ حالت زیر اتفاق می افتد:

۱) تکراری ندارند ← اگر تعداد کلمات یا رمز با تعداد کلمه برابر باشند (جایگشت خطی)
اگر تعداد کلمات یا رمز با تعداد کلمه برابر نباشد (ترتیب P)

۲) تکراری دارد ← اگر تعداد کلمات یا رمز با تعداد کلمه برابر باشد (جایگشت با تکرار)
اگر تعداد کلمات یا رمز با تعداد کلمه برابر نباشد (ترتیب یا ترکیب یا جایگشت با تکرار)

Ex) چند کلمه یا رمز ۳ حرفی می توان با کلمه‌ی sasan ساخت؟ ۱۳۲ تا

۱) کلماتی که حروف تکراری ندارند $P\binom{3}{3}$

۲) کلماتی که دو حرف تکراری a دارند $\binom{2}{1}\times 3$

۳) کلماتی که دو حرف تکراری s دارند $\binom{2}{1}\times 3$

Ex) با حروف continuity چند کلمه سه حرفی می توان ساخت؟

$P\binom{7}{3}+\binom{6}{1}\times3+\binom{6}{1}\times3+\binom{6}{1}\times3=۲۶۴$

$\binom{n}{0}+\binom{n}{1}+\binom{n}{2}+\dots+\binom{n}{n}=2^n \qquad \binom{n}{r}=\binom{n}{n-r}$

مثال) به چند طریق می توان از بین یک گروه ۷ نفره یک گروه حداقل ۲ نفره انتخاب کرد؟ $\binom{7}{2}+\binom{7}{3}+\dots+\binom{7}{7}=2^7-\binom{7}{0}-\binom{7}{1}=۱۲۸-۸=۱۲۰$

تقسیم ۱) تعداد حالات تقسیم n شی متمایز بین k دسته مختلف به طوری که دسته اول n_1 عضو، ... و دسته k ام n_k عضو داشته باشد:

$n_1+n_2+\dots+n_k=n \qquad \frac{n!}{n_1!\,n_2!\dots n_k!} = \binom{n}{n_1,n_2,\dots,n_k}$

Ex) به چند طریق می توان ۱۲ زندانی را در سلول ۳ نفری، ۴ نفری و ۵ نفری تقسیم کرد؟ به چند طریق می توان زندانیان را به طور مساوی بین ۳ سلول تقسیم کرد؟

$\binom{12}{3,4,5}=\frac{12!}{3!\,4!\,5!} \qquad \binom{12}{4,4,4}=\frac{12!}{4!\,4!\,4!}$

تقسیم ۲) تعداد حالات تقسیم n شی مشابه بین k فرد مختلف برابر $C\binom{n+k-1}{k-1}$ می باشد.

مثال) تقسیم ۵ مداد مشابه بین ۳ نفر: $\binom{5+3-1}{3-1} = \binom{7}{2} = \frac{7\times6}{2} = ۲۱$

تقسیم ۳) تعداد حالات مختلف تقسیم n شی مختلف بین k فرد مختلف k^n می باشد

قضیه انطباق: تعداد حالاتی که توپ های متفاوت 1 تا n در ظروف 1 تا n قرار گیرند به طوری که شماره توپ با شماره هیچ ظرفی یکسان نباشد

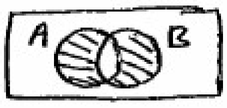
$$n!\left(\sum_{k=0}^{n}\frac{(-1)^k}{k!}\right)=n!\left(1-\frac{1}{1!}+\frac{1}{2!}-\dots+\frac{(-1)^n}{n!}\right)$$

ex) ۴ نامه برای ۴ نفر فرستاده می شود به چند حالت نامه ی هیچ کس به خودش نمی رسد؟

$$4!\left(1-\frac{1}{1!}+\frac{1}{2!}-\frac{1}{3!}+\frac{1}{4!}\right)=24\left(\frac{1}{2}-\frac{1}{6}+\frac{1}{24}\right)=12-4+1=9$$

پیشامد ناسازگار: اگر دو پیشامد هرگز به صورت همزمان رخ ندهند $A\cap B=\phi$

تفاضل متقارن: $A\Delta B=(A\cup B)-(A\cap B)=(A-B)\cup(B-A)$



پیشامد مستقل: اتفاق یکی بر رخداد دیگری هیچ اثری نداشته باشد $P(A\cap B)=P(A)\times P(B)$

درصورتی که A و B مستقل از هم باشند (A و B') و (A' و B) و (A' و B') نیز از هم مستقل اند

ex) احتمال زنده ماندن زن و شوهری تا ۲۰ سال دیگر به ترتیب $\frac{3}{5}$ و $\frac{1}{2}$ می باشد احتمال این که حداقل یکی از آنها تا ۲۰ سال دیگر زنده بماند چقدر است؟

$$P(A)[1-P(B)]+P(B)[1-P(A)]+P(A)P(B)=\frac{3}{5}\left(1-\frac{1}{2}\right)+\frac{1}{2}\left(1-\frac{3}{5}\right)+\frac{1}{2}\times\frac{3}{5}=\frac{3}{10}+\frac{2}{10}+\frac{3}{10}=\frac{7}{10}$$

قوانین دمورگان:

$$A'\cup B'=(A\cap B)'\Longrightarrow P(A'\cup B')=1-P(A\cap B)\qquad P(A\cap B')=P(A-B)$$

$$A'\cap B'=(A\cup B)'\Longrightarrow P(A'\cap B')=1-P(A\cup B)$$

احتمال شرطی: درصورت سوال با کلماتی چون می دانیم، متوجه می شویم، ... یک جمله خبری داده می شود که فضای نمونه را کوچک تر می کند

$$P(A|B)=\frac{P(A\cap B)}{P(B)}$$

ex) در ظرفی ۳ مهره سفید، ۴ قرمز و ۲ آبی است درصورتی یک مهره انتخاب شود و متوجه شویم قرمز نیست احتمال آن که آبی باشد چقدر است؟ $n(S)=9-4=5$ $\quad n(A)=2\Rightarrow P=\frac{2}{5}$

ex) اگر $P(A)=0{,}2$ و $P(B)=0{,}3$ باشد احتمال آنکه فقط A یا فقط B رخ دهد کدام است؟

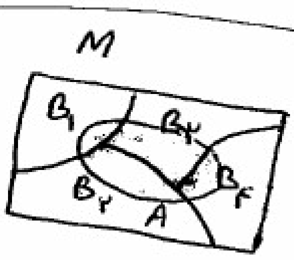
$$P(A)[1-P(B)]+P(B)[1-P(A)]=0{,}2\times0{,}7+0{,}3\times0{,}8=0{,}14+0{,}24=0{,}38$$

نکته) احتمال های شرطی به صورت زیر می باشد توجه شود که در مکمل گیری طرف راست شرط همواره بدون تغییر می ماند:

$$P(A'|B)=1-P(A|B)$$

$$P(A'|B')=1-P(A|B')$$

قضیه بیز: درصورتی که B_i ها مجموعه های دو به دو ناسازگار باشند و $\cup B_i=M$ باشد



حال اگر A رخ دهد احتمال رخداد B_j طبق قضیه بیز برابر است با:

$$P(A)=P(A\cap B_1)+\dots+P(A\cap B_4)=P(A|B_1)\cdot P(B_1)+\dots+P(A|B_4)\cdot P(B_4)=\sum P(A|B_i)P(B_i)$$

$$P(B_j|A)=\frac{P(A\cap B_j)}{P(A)}=\frac{P(A|B_j)P(B_j)}{\sum P(A|B_i)P(B_i)}$$

فصل ۳) متغیر تصادفی

درصورتی که $f(x)$ تابع احتمال باشد $\sum f(x)=1$ است.

نکته) تابع توزیع تجمعی $F(x)$ غیرنزولی و همیشه از راست پیوسته است یعنی

$$\lim_{x\to a^+} F(x)=F(a) \qquad F(a)=P(X\leqslant a)$$

$$P(a\leqslant X<b)=F(b^-)-F(a^-) \qquad P(X=b)=F(b)-F(b^-)$$

$$P(a\leqslant X\leqslant b)=F(b)-F(a^-) \qquad P(X>a)=1-P(X\leqslant a)=1-F(a)$$

$$P(a<X<b)=F(b^-)-F(a) \qquad P(a<X\leqslant b)=P(X\leqslant b)-P(X\leqslant a)=F(b)-F(a)$$

✓ برای تابع احتمال پیوسته احتمال در یک نقطه معنا ندارد $P(X=a)=0$ می باشد

با کمک تابع توزیع (تجمعی) می‌توان میانه، مد و چندک ها را محاسبه نمود:

۱) محاسبه مد از روی تابع توزیع: مد مقداری است که در تابع احتمال بیشترین احتمال را دارد یعنی جایی که $f(x)$ ماکزیمم می‌شود:

$$F'(x)=f(x) \Rightarrow F''(x)=f'(x)=0 \Rightarrow x=?$$ مقدار مد

۲) محاسبه میانه و دهک و صدک از روی تابع توزیع:

$F(b)=\frac{a}{100}$ ← b مقدار صدک a ام

$F(b)=\frac{1}{2}$ ← b مقدار میانه

$F(b)=\frac{a}{4}$ ← b مقدار چارک a ام

$F(b)=\frac{a}{10}$ ← b مقدار دهک a ام

توزیع شرطی متغیرهای تصادفی توام گسسته:

$$P(X=x|Y=y)=\frac{f(x,y)}{f_Y(y)}$$

$f_Y(y) \longrightarrow$ توزیع حاشیه ای y

$x \backslash y$	۰	۱
-1	۰/۲	۰/۱
۲	۰/۳	۰/۴

$$P(X=2|Y=0)=\frac{f(2,0)}{f_Y(0)}=\frac{0.3}{0.2+0.3}=\frac{3}{5}$$

$$P(Y=1|X=-1)=\frac{f(-1,1)}{f_X(-1)}=\frac{0.1}{0.2+0.1}=\frac{1}{3}$$

* X و Y را مستقل از هم گویند هرگاه

$$\forall x,y:\ f(x,y)=f(x)\cdot f(y)$$

یعنی برای هر مقدار x و y ضرب احتمال حاشیه ای ها برابر احتمال توام باشد.

$X \backslash Y$	۰	۱	f_X
-1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
۱	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
f_Y	$\frac{1}{2}$	$\frac{1}{2}$	۱

Ex) درصورتی که $0<x,y<1$ ، $f(x,y)=6x^2y$ مقدار $P(0<X<\frac{3}{4}, \frac{1}{3}<Y<1)$ کدام است؟

$$P=\int_{\frac{1}{3}}^{1}\int_{0}^{\frac{3}{4}} 6x^2y\,dx\,dy=\int_{\frac{1}{3}}^{1}\left[2x^3\Big|_0^{\frac{3}{4}}\right]y\,dy=\int_{\frac{1}{3}}^{1}\frac{27}{32}y\,dy=\frac{27}{64}\left(1^2-\frac{1}{9}\right)=\frac{8}{9}\times\frac{27}{64}=\frac{3}{8}$$

توزیع احتمال های حاشیه ای:

$$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\,dy \qquad f_Y(y)=\int_{-\infty}^{+\infty} f(x,y)\,dx$$

Ex) در تابع احتمال توام روبرو مقدار c چقدر است؟

$$f(x,y)=cxy^2,\quad 0\leqslant x\leqslant y\leqslant 1$$

$$\int_0^1\int_x^1 cxy^2\,dy\,dx=\int_0^1 c\left[\frac{1}{3}-\frac{1}{3}x^3\right]x\,dx=c\left[\frac{x^2}{6}-\frac{1}{15}x^5\right]=c\left[\frac{1}{6}-\frac{1}{15}\right]=1\Rightarrow c=10$$

فرقی نمی‌کند ولی باید یکی از بازه ها رعایت شود:

$$\int_0^1\int_y^1 cxy^2\,dx\,dy=\int_0^1 c\left[\frac{1}{2}-\frac{y^2}{2}\right]y^2\,dy=\frac{c}{2}\left[\frac{y^3}{3}-\frac{y^5}{5}\right]_0^1=\frac{c}{2}\left[\frac{1}{3}-\frac{1}{5}\right]=\frac{c}{10}=1\Rightarrow c=10$$

Ex) اگر $0\leqslant x\leqslant 1$ ، $f(x)=a+bx^2$ و $E(X)=\frac{3}{5}$ باشد $a+b$ چقدر است؟

$$\int_0^1 f(x)\,dx=1\Rightarrow a+\frac{b}{3}=1$$

$$\int_0^1 x f(x)=\frac{3}{5}\Rightarrow \frac{a}{2}+\frac{b}{4}=\frac{3}{5}$$

$$\Rightarrow\begin{cases}3a+b=3\\ 2a+b=\frac{12}{5}\end{cases}$$

$$a=\frac{3}{5},\ b=\frac{6}{5}\Rightarrow a+b=\frac{9}{5}$$

خواص امید ریاضی:

$E(X)=\int x f(x)\,dx$

$E(aX+b)=aE(X)+b$

$E(g(X))=\int g(x)f(x)\,dx$

$var(X)=E(X^2)-(E(X))^2$

$E(E(X))=E(X)$

$E(a)=a$

$E(XY)=E(X)E(Y)$ X و Y مستقل باشند

کوواریانس $Cov(X,Y)$: معیاری است که نشان دهنده شدت رابطه‌ی خطی بین دو متغیر است

$$cov(X,Y)=E(XY)-E(X)E(Y)=E((X-E(X))(Y-E(Y)))$$

$cov(X,X)=var(X)$

$cov(X,Y)=cov(Y,X)$

$cov(X,a)=0$

$cov(aX\pm b, cY\pm d)=ac\,cov(X,Y)$

$cov(aX_1+bX_2, Y)=a\,cov(X_1,Y)+b\,cov(X_2,Y)$

$var(aX\pm bY\pm c)=a^2 var(X)+b^2 var(Y)-2ab\,cov(X,Y)$

اگر X و Y مستقل باشند $cov(X,Y)=0$

نکته) $E(X|Y=y)=\sum x P(X|Y=y)$

ex) فرض کنید X و Y دارای توزیع توام باشند $E(X|Y=0)$ کدام است؟

$Y \backslash X$	0	1	P_Y
0	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{4}{8}$
1	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{3}{6}$
P_X	$\frac{11}{24}$	$\frac{13}{24}$	1

X	0	1
$P(X\|Y=0)$	$\frac{P(0,0)}{P_Y(0)}=\frac{\frac{1}{8}}{\frac{4}{8}}=\frac{1}{4}$	$\frac{P(1,0)}{P_Y(0)}=\frac{\frac{3}{8}}{\frac{4}{8}}=\frac{3}{4}$

$$E(X|Y=0)=0\times\frac{1}{4}+1\times\frac{3}{4}=\frac{3}{4}$$

تابع مولد گشتاور $M_X(t)=E(e^{xt})$

خواص تابع مولد گشتاور:

اگر r بار از تابع مولد گشتاور $M_X(t)$ نسبت به t مشتق بگیریم وسپس t را مساوی صفر قرار دهیم گشتاور r ام بدست خواهد آمد:

$$E(X^r)=\frac{d^r M_X(t)}{dt^r}\Big|_{t=0}$$

ex) اگر X متغیر تصادفی باشد و

اگر X متغیر تصادفی باشد و تابع احتمال آن $x\in R$ $\quad f(x)=e^{-|x|}$ باشد تابع مولد گشتاور آن کدام است؟

$$M_X(t)=\int_{-\infty}^{\infty} e^{xt}e^{-|x|}\,dx=\int_{-\infty}^{0} e^{(t+1)x}\,dx+\int_{0}^{\infty} e^{-(1-t)x}\,dx=\frac{1}{t+1}+\frac{1}{1-t}=\frac{1}{1-t}$$

ex) اگر برای متغیر X تابع مولد گشتاور $M_X(t)$ باشد برای $aX+b$ تابع مولد گشتاور کدام است؟

$$E(e^{(ax+b)t})=E(e^{bt}e^{axt})=e^{bt}E(e^{atX})=e^{bt}M_X(at)$$

توزیع توابعی از متغیر تصادفی: برای بدست آوردن توزیع احتمال $Y=g(X)$ ابتدا بایستی تابع تجمعی Y را بدست آورده سپس از آن نسبت به y مشتق بگیریم تا تابع احتمال Y بدست آید.

ex) در صورتی که $f(x)=x$ ، $0\leqslant x\leqslant 1$ باشد و $Y=X^3$ باشد تابع احتمال Y کدام است؟

$$F_Y(y)=P(Y\leqslant y)=P(X^3\leqslant y)=P(X\leqslant y^{\frac{1}{3}})=\int_0^{y^{\frac{1}{3}}} \underbrace{x}_{f(x)}\,dx=\frac{1}{2}y^{\frac{2}{3}}\Rightarrow f_Y(y)=\frac{1}{3}y^{-\frac{1}{3}}$$

ضریب همبستگی خطی: میزان ارتباط خطی میان دو متغیر را نشان می دهد

$$\rho=\frac{cov(X,Y)}{\sqrt{var(X)}\sqrt{var(Y)}} \qquad -1\leqslant\rho\leqslant 1$$

جمع، تفریق، ضرب و تقسیم تاثیری در ضریب همبستگی خطی ندارد

فصل ۴) توزیع‌های آماری

الف) توزیع های گسسته:

۱) توزیع یکنواخت: هرگاه مقادیر مختلف X احتمال یکسان را به ما دهند

$$f(x)=P(X=x)=\frac{1}{n}$$

$$F(x)\begin{cases}0 & x<x_1\\ \frac{i}{n} & x_i\leqslant x<x_{i+1}\\ 1 & x>x_n\end{cases}$$

$$E(X)=\sum_{1}^{n}x f(x)=\frac{1}{n}\sum x=\frac{1}{n}\left(\frac{n(n+1)}{2}\right)=\frac{n+1}{2}$$

$$var(X)=E(X^2)-(E(X))^2=\sum\frac{x^2}{n}-\left(\frac{n+1}{2}\right)^2=\frac{1}{n}\times\frac{n(n+1)(2n+1)}{6}-\left(\frac{n+1}{2}\right)^2=\frac{n^2-1}{12}$$

$$M_X(t)=E(e^{xt})=\sum e^{xt}\cdot\frac{1}{n}=\frac{1}{n}\left(e^t+e^{2t}+\dots+e^{nt}\right)=\frac{1}{n}\left(\frac{e^t(1-e^{nt})}{(1-e^t)}\right)$$

۲) توزیع برنولی: آزمایشی که دو نتیجه شکست و پیروزی را در پی دارد

$$f(x)=p^x(1-p)^{1-x}\qquad x=0,1$$

$$E(X)=p\qquad var(X)=pq\qquad M_X(t)=q+pe^t$$

۳) توزیع دوجمله‌ای: اگر آزمایش برنولی را به صورت مستقل (تکرار با جایگذاری) n بار تکرار کنیم و X تعداد موفقیت ها باشد داریم:

$$f(x)=\binom{n}{x}p^x q^{n-x}=\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

$$E(X)=np\qquad var(X)=npq\qquad M_X(t)=(q+pe^t)^n$$

حالت خاصی از توزیع دوجمله‌ای که در آن $q=p=\frac{1}{2}$ می‌باشد به صورت زیر است:

$$f(x)=\frac{\binom{n}{x}}{2^n}$$

نکات توزیع دوجمله ای:

۱) درصورتی که $p<\frac{1}{2}$ باشد توزیع دوجمله ای چوله به راست و اگر $p>\frac{1}{2}$ باشد چوله به چپ است.

۲) محتمل‌ترین پیشامد (مد توزیع) در توزیع دوجمله ای $p(n+1)$ برای وقوع X خواهد بود

۴) توزیع چندجمله‌ای: همان توزیع دوجمله ای است اما نتایج آن بیش از دو نتیجه است

$$f(x_1,x_2,\dots,x_n)=P(X_1=x_1,\dots,X_n=x_n)=\binom{n}{x_1,\dots,x_n}p_1^{x_1}p_2^{x_2}\dots p_n^{x_n}=\frac{n!}{x_1!\,x_2!\dots x_n!}p_1^{x_1}p_2^{x_2}\dots p_n^{x_n}$$

۵) توزیع فوق هندسی: اگر در آزمایش های تصادفی دو حالتی که تعداد n بار آزمایش غیرمستقل (بدون جایگذاری) انجام شود

$$f(x)=\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}\qquad E(X)=\frac{nM}{N}\qquad var(X)=\frac{N-n}{N-1}\left(\frac{nM}{N}\right)\left(1-\frac{M}{N}\right)$$

نکته) درصورتی که n نسبت به N بسیار کوچک باشد ($\frac{n}{N}\leqslant 0/05$) می‌توان از توزیع دوجمله ای ($p=\frac{M}{N}$) برای تقریب توزیع فوق هندسی استفاده کرد (در توزیع دوجمله‌ای احتمال موفقیت (p) به دلیل با جایگذاری همواره ثابت است اما در توزیع فوق هندسی احتمال موفقیت به علت این که بدون جایگذاری است ثابت نیست ولی اگر $n<<N$ باشد می‌توان آن را ثابت فرض کرد و ($\frac{N-n}{N-1}\to 1$) را نادیده گرفت.

۶) توزیع پواسن: اگر X نشان دهنده تعداد موفقیت ها در یک فاصله زمانی یا مکانی مشخص باشد

λ (پارامتر توزیع): متوسط تعداد اتفاقات در بازه زمانی یا مکانی مشخص است

$$f(x)=\frac{e^{-\lambda}\lambda^x}{x!}\qquad E(X)=var(X)=\lambda$$

تقریب توزیع دوجمله‌ای به وسیله توزیع پواسن:

اگر در توزیع دوجمله‌ای n بزرگ و p کوچک باشد می‌توان توزیع دوجمله‌ای را با پواسن به صورت $\lambda=np$ تقریب زد: ($np\approx 5$).

$$\binom{n}{x}p^x(1-p)^{n-x}=\frac{e^{-np}(np)^x}{x!}$$

۷) توزیع هندسی: تکرار آزمایش های برنولی برای رسیدن به اولین موفقیت:

$f(x)=q^{x-1}p$

$E(X)=\frac{1}{p}$ $var(X)=\frac{q}{p^2}$ $M_X(t)=\frac{pe^t}{1-qe^t}$

ex) یک تاس سالم را آنقدر پرتاب می کنیم تا برای اولین بار عدد ۶ ظاهر شود. احتمال این که پرتاب های لازم برای رسیدن به اولین ۶ عدد فرد باشد چقدر است؟

$P(X=x)=\frac{1}{6}+(\frac{5}{6})^2(\frac{1}{6})+(\frac{5}{6})^4(\frac{1}{6})+\dots=\frac{\frac{1}{6}}{1-\frac{25}{36}}=\frac{6}{11}$

۸) توزیع دوجمله ای منفی: تکرار آزمایش های برنولی تا رسیدن به r امین موفقیت

$f(r)=\binom{n-1}{r-1}p^r q^{n-r}$

$E(X)=\frac{r}{p}$ $var(X)=\frac{rq}{p^2}$ $M_X(t)=(\frac{pe^t}{1-qe^t})^r$

ب) توزیع های پیوسته:

۱) یکنواخت: $X\sim U(a,b)$

$f(x)=\frac{1}{b-a}, a\leqslant x\leqslant b$ $F(x)=\frac{x-a}{b-a}$ و $a\leqslant x\leqslant b$

$E(X)=\frac{a+b}{2}$ $var(X)=\frac{(b-a)^2}{12}$ $M_X(t)=\frac{e^{tb}-e^{ta}}{t(b-a)}$

۲) توزیع گاما: اگر در توزیع پواسن X مدت زمان لازم برای رخ دادن α امین اتفاق باشد X دارای توزیع گاما است

β تعداد متوسط در واحد زمان

$f(x)=\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$ $x>0$ $\Gamma(\alpha)\to(\alpha-1)!$

$E(X)=\frac{\alpha}{\beta}$ $Var(X)=\frac{\alpha}{\beta^2}$ $M_X(t)=(\frac{\beta}{\beta-t})^\alpha$

۳) توزیع نمایی: در توزیع گاما اگر $\alpha=1$ شود (X زمان لازم برای رخ دادن اولین اتفاق) $X\sim E(\beta)$

توزیع بدون حافظه

$f(x)=\beta e^{-\beta x}$ $F(X)=1-e^{-\beta x}$ $E(X)=\frac{1}{\beta}$ $Var(X)=\frac{1}{\beta^2}$ $M_X(t)=\frac{\beta}{\beta-t}$

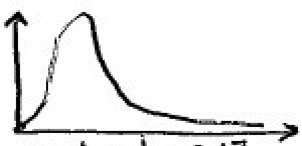
تذکر مهم) واحد β تعداد بر واحد زمان است درصورتی که درصورت سؤال متوسط طول عمر (واحد زمان) ذکر شود منظور $E(X)$ می باشد نه β! پس در این حالت $\beta=\frac{1}{E(X)}$ می باشد

توزیع خی دو (χ^2): در توزیع گاما اگر $\alpha=\frac{n}{2}$ و $\beta=\frac{1}{2}$ باشد

هرچه درجه آزادی (n) افزایش یابد توزیع متقارن تر می شود

$E(X)=n$ و $Var(X)=2n$

$\chi^2_{(n)}=Z_1^2+Z_2^2+\dots+Z_n^2$



توزیع چوله به راست

توزیع نرمال: $X\sim N(\mu,\sigma^2)$

۱) $\mu=\hat{X}=\tilde{X}$ و نقاط عطف در $\mu\pm\sigma$ می باشد

$M_X(t)=e^{\mu t+\frac{1}{2}\sigma^2t^2}$

$P(\mu-\sigma<X<\mu+\sigma)=0.68$

$P(\mu-2\sigma<X<\mu+2\sigma)=0.95$

$P(\mu-3\sigma<X<\mu+3\sigma)=0.997$

$\Phi(-a)=1-\Phi(a)$

$P(-a\leqslant X\leqslant a)=\Phi(a)-\Phi(-a)=1-2\Phi(-a)=2\Phi(a)-1$

تقریب پواسن با نرمال: درصورتی که $\lambda>10$ باشد با فرض $\sigma^2=\mu=\lambda$ می توان پواسن را با نرمال تقریب زد ولی چون یک توزیع گسسته با یک توزیع پیوسته تقریب زده می شود (درصورتی که در سوال ذکر شود) بایستی تصحیح پیوستگی انجام شود

$P(X=a)=P(a-\frac{1}{2}<X<a+\frac{1}{2})$

$P(a\leqslant X\leqslant b)=P(a-\frac{1}{2}<X<b+\frac{1}{2})$

$P(X>a)=P(X\geqslant a+1)=P(X>a+\frac{1}{2})$

$P(X<b)=P(X\leqslant b-1)=P(X<b-\frac{1}{2})$

تقریب دوجمله ای با نرمال: درصورتی که $p\to\frac{1}{2}$ میل کند و $np>5$ باشد می توان $np=\mu$ و $npq=\sigma^2$ در نظر گرفت

توزیع t استودنت:

$t=\frac{Z}{\sqrt{\frac{\chi^2_n}{n}}}$

$E(t)=0$ و $Var(t)=\frac{n}{n-2}$

$F(n_1,n_2,\frac{\alpha}{2})=\frac{1}{F(n_2,n_1,1-\frac{\alpha}{2})}=\frac{\chi^2_{n_1}/n_1}{\chi^2_{n_2}/n_2}$

$F(1,n)=\frac{Z^2}{\chi^2_n/n}=t^2_{(n)}$

فصل ۵) نمونه گیری

آماره: هر تابعی از نمونه تصادفی که شامل پارامترهای مجهول جامعه نباشد را گویند

* انواع نمونه گیری:

۱) نمونه گیری ساده: انتخاب هر عضو جامعه دارای شانس برابر است (با جایگذاری و بدون جایگذاری)

۲) نمونه گیری سیستماتیک: نمونه گیری که در آن ترتیب اعضای جامعه مهم است و بر طبق یکی از اعضا با رعایت ترتیب بقیه اعضا انتخاب می شود

۳) نمونه گیری گروهی: جامعه به چند گروه تقسیم می شود که اعضای هر گروه همگن هستند ولی تعداد اعضا برابر نیستند (اختلاف بین گروه هاست)

۴) نمونه گیری خوشه ای: جامعه به چند خوشه با اعضای ناهمگن و با تعداد اعضای برابر تقسیم می شود و سپس یک خوشه به عنوان نمونه انتخاب می شود

۵) نمونه گیری چند مرحله ای: ابتدا جامعه را به چند قسمت تقسیم کرده و از هر قسمت در هر بار تعدادی فیلتر می شوند تا به نمونه مورد نظر برسیم.

توزیع نمونه ای:

برای هر پارامتر جامعه یک آماره وجود دارد که خود آماره یک متغیر تصادفی است و دارای یک تابع احتمال است که به تابع احتمال آن، توزیع نمونه ای گویند.

$E(\bar{X})=\mu$

$$var(\bar{X})=\begin{cases}\dfrac{\sigma^2}{n} & \text{اگر جامعه نامتناهی باشد}\\[2mm] \dfrac{N-n}{N-1}\,\dfrac{\sigma^2}{n} & \text{اگر جامعه متناهی باشد}\end{cases}$$

قضیه حد مرکزی:

۱) در صورتی که از یک جامعه غیرنرمال به تعداد n نمونه انتخاب شود در صورتی که n به اندازه کافی بزرگ باشد توزیع نمونه ای میانگین نمونه $(\bar{X})$ تقریبا نرمال خواهد بود

نامساوی مارکف: برای هر $c>0$ و $g(x)$ مثبت خواهیم داشت:

$$P(g(x)\geqslant c)\leqslant \frac{E(g(x))}{c}$$: میزان حداکثر احتمال

نامساوی چبی شف: در صورتی توزیع احتمال نامشخص با میانگین μ و انحراف معیار σ داشته باشیم خواهیم داشت:

حداقل احتمال: $$P(|x-\mu|\leqslant k\sigma)\geqslant 1-\frac{1}{k^2} \Longrightarrow P(\mu-k\sigma\leqslant x\leqslant \mu+k\sigma)\geqslant 1-\frac{1}{k^2}$$

حداکثر احتمال: $$P(|x-\mu|\geqslant k\sigma)\leqslant \frac{1}{k^2} \Longrightarrow P(x\geqslant \mu+k\sigma \text{ یا } x\leqslant \mu-k\sigma)\leqslant \frac{1}{k^2}$$

قانون اعداد بزرگ:

احتمال آنکه قدر مطلق میانگین متغیرهای تصادفی x_1 و ... و x_n از میانگین امید ریاضی آنها کوچکتر از هر مقدار مثبت ε باشد با افزایش نامحدود n برابر ۱ می باشد:

$$\lim_{n\to\infty} P\left(\left|\frac{x_1+\dots+x_n}{n}-\frac{E(x_1)+\dots+E(x_n)}{n}\right|<\varepsilon\right)=1$$

$$\lim_{n\to\infty} P(|\bar{X}-\mu|<\varepsilon)=1$$

قانون قوی اعداد بزرگ:

نتیجه گیری از قانون اعداد بزرگ

۱) اگر از جامعه ای به حجم N به تعداد n نمونه گیری شود
الف) N نامحدود باشد: $E(\bar{x})=\mu_x$, $var(\bar{x})=\dfrac{\sigma_x^2}{n}$
ب) N محدود باشد: $E(\bar{x})=\mu_x$, $var(\bar{x})=\dfrac{N-n}{N-1}\,\dfrac{\sigma_x^2}{n}$

۲) اگر جامعه نرمال با واریانس σ^2 معلوم باشد هر ترکیب خطی از x ها باز دارای توزیع نرمال است $\quad \bar{x}\sim N(\mu,\dfrac{\sigma_x^2}{n})$

۳) اگر جامعه نرمال با واریانس σ^2 مجهول باشد
الف) $n>30$: $\bar{x}\sim N(\mu,\dfrac{S_x^2}{n}) \longrightarrow \dfrac{\bar{x}-\mu}{S/\sqrt{n}}\sim Z(0,1)$
ب) $n<30$: $\bar{x}\sim N(\mu,\dfrac{S_x^2}{n}) \longrightarrow \dfrac{\bar{x}-\mu}{S/\sqrt{n}}\sim t_{(n-1)}$

۴) اگر توزیع جامعه نامعلوم ولی با واریانس σ^2 معلوم باشد طبق قضیه حد مرکزی داریم:

$\bar{x}\sim N(\mu,\dfrac{\sigma^2}{n}) \qquad n>30$

$n<30 \longrightarrow$ استفاده از نامساوی چبی شف

۱) توزیع نمونه‌ای مجموع میانگین دو نمونه: درصورتی که دو جامعه نرمال باشند و یا n_1 و $n_2 > 30$ باشد و σ_1^2 و σ_2^2 ها معلوم باشند:

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

ب) درصورتی که واریانس دو جامعه نامعلوم ولی برابر باشند ($\sigma_2^2 = \sigma_1^2$) داریم:

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \Rightarrow T = \frac{(\bar{X}_1 \pm \bar{X}_2) - (\mu_1 \pm \mu_2)}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{(n_1+n_2-2)}$$

تذکر: درصورتی که هدف مقایسه میانگین واقعی دو جامعه باشد بایستی مقدار $\mu_1 \pm \mu_2$ صفر در نظر گرفته شود (دو جامعه اختلافی ندارند)

---

۲) توزیع نمونه‌ای واریانس (S^2):

الف) جامعه نرمال با میانگین مجهول μ و واریانس معلوم σ^2:

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

$$E(S^2) = \quad , \; var(S^2) = \frac{2\sigma^4}{n-1} \quad , \; S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \Rightarrow \frac{\sum (x-\bar{x})^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

ب) جامعه نرمال با میانگین معلوم μ و واریانس معلوم σ^2:

$$\frac{nS^2}{\sigma^2} \sim \chi^2_{(n)} \quad , \; S^2 = \frac{\sum (x_i - \bar{x})^2}{n} \; , \; E(S^2) = \sigma^2 \; , \; var(S^2) = \frac{2\sigma^4}{n}$$

ج) جامعه نرمال با میانگین مجهول μ و حجم نمونه $n > 30$:

$$\frac{nS^2}{\sigma^2} \sim \chi^2_{(n-1)} \quad , \; S^2 = \frac{\sum (x_i - \bar{x})^2}{n} \; , \; E(S^2) = \frac{n-1}{n}\sigma^2 \; , \; var(S^2) = \frac{2(n-1)}{n}\sigma^4$$

۳) توزیع نمونه‌ای نسبت واریانس های نمونه‌ای:

$$F(n_1-1, n_2-1) = \frac{\chi^2_{(n_1-1)}/n_1-1}{\chi^2_{(n_2-1)}/n_2-1} = \frac{\frac{\cancel{(n_1-1)}S_1^2}{\sigma_1^2} \Big/ \cancel{n_1-1}}{\frac{\cancel{(n_2-1)}S_2^2}{\sigma_2^2} \Big/ \cancel{n_2-1}} = \frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \sim F(n_1-1, n_2-1) \qquad ex) \; P\left(\frac{S_1^2}{S_2^2} > 5\right) = ?$$

---

۴) توزیع نمونه‌ای نسبت نمونه‌ای $(\bar{p})$:

درصورتی که $\bar{p}$ نسبت موفقیت نمونه و p نسبت موفقیت جامعه باشد داریم:

$$\bar{P} \sim N\left(P, \frac{pq}{n}\right)$$

$$\frac{\bar{P} - P}{\sqrt{\frac{P(1-P)}{n}}} \sim Z(0,1)$$

---

۵) توزیع نمونه‌ای مجموع یا تفاضل نسبت دو نمونه: n_1 و n_2 بزرگ هستند:

$$\frac{(\bar{P}_1 \pm \bar{P}_2) - (P_1 \pm P_2)}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}} \sim Z(0,1)$$

فصل ۶) برآورد

هر ویژگی یک جامعه را پارامتر آن جامعه گویند و هر ویژگی متناظر با آن در نمونه را یک آماره می نامند.

نکته) برآورد نقطه‌ای برای پارامترهای σ^2، μ و P به ترتیب آماره‌های S^2، $\bar{X}$ و $\bar{P}$ می باشند.

۱) برآوردگرهای نقطه‌ای با روش حداکثر درستنمایی (MLE):
تابع احتمال را n بار در هم ضرب کرده و از آن Ln می گیریم سپس از آن مشتق گرفته و برابر صفر قرار می دهیم تا برآوردی از پارامتر بدست آید.

$$f(x_1,\dots,x_n) = f(x_1)\cdot f(x_2)\dots f(x_n)$$

میزان اریبی $= E(\hat\theta-\theta) = E(\hat\theta)-\theta$

کارایی نسبی $\hat\theta_1$ به $\hat\theta_2$ برابر $\frac{var(\hat\theta_2)}{var(\hat\theta_1)}$ می باشد

تذکر: در بین برآوردگرها $\bar X$ همواره برآوردگری نااریب و با کمترین واریانس (کارا) می باشد.

نکته) $var(\bar X)=\frac{\sigma^2}{n}$ و $var(Md)=\frac{\sigma^2}{n}\times\frac{\pi}{2}$ می باشد

۲) MSE (میانگین توان دوم خطاها):

$$MSE(\hat\theta)=E((\hat\theta-\theta)^2)=var(\hat\theta)+(E(\hat\theta-\theta))^2=var(\hat\theta)+(\text{اریبی})^2$$

در بین برآورد کننده‌های دلخواه (چه اریب و چه نااریب) برآوردکننده‌ای بهتر (کاراتر) است که MSE کمتری داشته باشد.

مثال) اگر $X_1,\dots,X_n$ یک نمونه تصادفی از توزیع یکنواخت بر فاصله $(\theta,\theta+1)$ باشد مقدار میانگین توان دوم خطاها برای $\hat\theta=\bar X-\frac12$ کدام است؟

$E(\hat\theta)=E(\bar X-\frac12)=E(\bar X)-\frac12=\frac{\theta+1+\theta}{2}-\frac12=\theta$ نااریب

$MSE(\hat\theta)=var(\hat\theta)+0^2=var(\bar X-\frac12)=\frac{1}{12n}\Rightarrow\lim_{n\to\infty}MSE(\hat\theta)=0$ ← $\hat\theta$ برآوردگر سازگاری می باشد

نکته) درصورتی که یک برآوردگر سازگار باشد با افزایش n، آماره $\hat\theta$ به پارامتر θ نزدیکتر می شود و $\lim_{n\to\infty}MSE(\hat\theta)=0$ برابر صفر می شود

۳) فاصله اطمینان:
با احتمال $(1-\alpha)\%$ (α سطح خطا و $1-\alpha$ سطح اطمینان) پارامتر جامعه (θ) در فاصله اطمینان $(\hat\theta-e,\hat\theta+e)$ (e خطای برآورد و $\hat\theta$ برآورد نمونه‌ای) قرار می‌گیرد

تذکر: هرچه سطح اطمینان (فاصله اطمینان) کوچکتر باشد برآورد دقیق‌تر است چون خطای برآورد (e) کاهش می‌یابد.

خطای برآورد میانگین: $e=Z_{\frac\alpha2}\frac{\sigma}{\sqrt n}$

* به α سطح خطا، e خطای برآورد و $Z_{\frac\alpha2}$ ضریب اطمینان گویند

نکته) هرچه α کاهش یابد میزان خطای برآورد e، فاصله اطمینان و سطح اطمینان $(1-\alpha)$ افزایش خواهد یافت

مفهوم فاصله اطمینان با خطای ۵٪ ⇐ اگر نمونه گیری را ۱۰۰ بار تکرار کنیم ۹۵ بار فاصله اطمینانی بدست می‌آوریم که پارامتر جامعه θ را در بر می‌گیرد

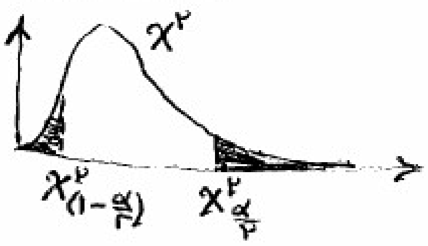
$$\boxed{Z_{0.1}=1.28,\quad Z_{0.05}=1.645,\quad Z_{0.025}=1.96,\quad Z_{0.01}=2.32,\quad Z_{0.005}=2.58}$$

الف) فاصله اطمینان میانگین جامعه:

واریانس جامعه معلوم: $(\bar X - Z_{\frac\alpha2}\frac{\sigma}{\sqrt n},\ \bar X + Z_{\frac\alpha2}\frac{\sigma}{\sqrt n})$

واریانس جامعه نامعلوم: $(n<30)$ $(\bar X - t_{(n-1)}\frac{S}{\sqrt n},\ \bar X + t_{(n-1)}\frac{S}{\sqrt n})$

ب) فاصله اطمینان واریانس جامعه با توزیع نرمال:



میانگین جامعه نامعلوم: $\frac{(n-1)S^2}{\chi^2_{(\frac\alpha2,n-1)}}\le\sigma^2\le\frac{(n-1)S^2}{\chi^2_{(1-\frac\alpha2,n-1)}}$

میانگین جامعه معلوم: $\frac{nS^2}{\chi^2_{(\frac\alpha2,n-1)}}\le\sigma^2\le\frac{nS^2}{\chi^2_{(1-\frac\alpha2,n-1)}}$

ج) فاصله اطمینان نسبت واریانس دو جامعه نرمال:

$$\frac{\frac{S_1^2}{S_2^2}}{F(n_1-1, n_2-1, \frac{\alpha}{2})} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{\frac{S_1^2}{S_2^2}}{F(n_1-1, n_2-1, 1-\frac{\alpha}{2})}$$

د) فاصله اطمینان برای تفاضل یا مجموع میانگین دو جامعه نرمال:

واریانس هر دو جامعه معلوم:

$$(\bar{X}_1 \pm \bar{X}_2) - Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq \mu_1 \pm \mu_2 \leq (\bar{X}_1 \pm \bar{X}_2) + Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

واریانس ها نامعلوم ولی برابر:

$$\begin{cases} \left((\bar{X}_1 \pm \bar{X}_2) - S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\, t_{(n_1+n_2-2)} \;,\; (\bar{X}_1 \pm \bar{X}_2) + t_{(n_1+n_2-2)} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right) \\ S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \end{cases}$$

واریانس ها نامعلوم و نابرابر:

$$\begin{cases} \left((\bar{X}_1 \pm \bar{X}_2) - t_{df, \frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_1}} \;,\; (\bar{X}_1 \pm \bar{X}_2) + t_{df, \frac{\alpha}{2}} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right) \\ df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2-1}} \end{cases}$$

هـ) فاصله اطمینان برای نسبت جامعه

$$\bar{p} - Z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}\bar{q}}{n}} \leq p \leq \bar{p} + Z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}\bar{q}}{n}}$$

$$(\bar{p}_1 \pm \bar{p}_2) - Z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}_1\bar{q}_1}{n_1} + \frac{\bar{p}_2\bar{q}_2}{n_2}} \leq p_1 \pm p_2 \leq (\bar{p}_1 \pm \bar{p}_2) + Z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}_1\bar{q}_1}{n_1} + \frac{\bar{p}_2\bar{q}_2}{n_2}}$$

---

۴) تعیین تعداد نمونه

الف) برای میانگین جامعه:

$$e = Z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}} \Rightarrow \sqrt{n} \geq Z_{\frac{\alpha}{2}} \frac{\sigma}{e} \Rightarrow n \geq \left(\frac{Z_{\frac{\alpha}{2}}\sigma}{e}\right)^2$$

ب) برای واریانس جامعه:

$$e = t_{(n-1)} \frac{S}{\sqrt{n}} \Rightarrow n_1 = \left(\frac{t_{(n-1)} S}{e}\right)^2$$

۱) ابتدا نمونه‌ای کوچک به حجم n_0 گرفته و n_1 را محاسبه می‌کنیم:

۲) در صورتی که $n_1 \geq n_0$ باشد n_0 به عنوان تعداد نمونه انتخاب می‌شود و در غیر این صورت n_2 را با جایگذاری $n = n_1 + 1$ محاسبه می‌کنیم

۳) گام دوم را آنقدر بررسی می‌کنیم تا به جواب مطلوب برسیم.

ج) برای نسبت جامعه:

$$e = Z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}\bar{q}}{n}} \Rightarrow \sqrt{n} \geq \frac{Z_{\frac{\alpha}{2}} \sqrt{\bar{p}\bar{q}}}{e} \Rightarrow n \geq (\bar{p}\bar{q})\left(\frac{Z_{\frac{\alpha}{2}}}{e}\right)^2$$

در صورتی که $\bar{p}$ (نسبت نمونه) مشخص نباشد بهترین و محافظه‌کارانه‌ترین تعداد نمونه برای $\bar{p} = \bar{q} = \frac{1}{2}$ بدست خواهد آمد

فصل ۷) آزمون فرضیه ها :

ادعا همیشه در فرض H_1 بیان می‌شود مگر اینکه شامل تساوی باشد که جای فرض H_0 و H_1 عوض می‌شود.

نکته) فرض H_0 همواره شامل مساوی خواهد بود و به دنبال آن هستیم که فرض H_0 را بپذیریم مگر آنکه خلاف آن ثابت شود

خطای نوع اول (α) و خطای نوع دوم (β):

رد H_0 به اشتباه ← پذیرش H_0 به اشتباه ←

$\alpha = P(H_0$ رد $|$ H_0 صحیح باشد$) \longrightarrow$ اعدام بی گناه

$\beta = P(H_0$ پذیرش $|$ H_0 صحیح نباشد$) \longrightarrow$ آزادی قاتل

$\beta^* = P(H_0$ رد $|$ H_1 صحیح باشد$) = 1-\beta \longrightarrow$ تشخیص درست قاتل

توان آزمون = احتمال رد H_0 به درستی؛ ← مکمل خطای نوع دوم

۱) خطای α و β با هم رابطه‌ی عکس دارند با افزایش یکی دیگری کاهش می‌یابد

۲) با افزایش حجم نمونه α و β هر دو کاهش می‌یابند

۳) همواره $\alpha+\beta \leq 1$ می‌باشد

۴) α با توان آزمون ($\beta^* = 1-\beta$) رابطه مستقیم و با خطای برآورد (ε) ← دقت برآورد و ضریب اطمینان $Z_{\frac{\alpha}{2}}$ رابطه عکس دارد

۵) هرچه بخواهیم اطمینان افزایش یابد بایستی فاصله اطمینان بزرگتر و α کوچکتر شود و هرچه بخواهیم دقت بیشتر شود بایستی فاصله اطمینان کوچکتر شود و β کاهش یا α افزایش یابد

قسمت‌های هاشور خورده : ناحیه بحرانی

در آزمون فرض برای میانگین آماره $Z=\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ را با قرار دادن مقدار μ از فرض H_0 محاسبه می‌کنیم سپس براساس α که در مسئله ذکر می‌شود، مقدار Z محاسباتی را با فضای پذیرش H_0 یا ناحیه بحرانی تحلیل می‌کنیم.

نکته) درصورتی‌که میانگین جامعه نرمال ما مشخص نباشد آماره مربوط به آزمون فرضیه‌ها به صورت :

$$\begin{cases} \text{if } n \geq 30 \longrightarrow Z_0 = \dfrac{\bar{X}-\mu}{S/\sqrt{n}} \\ \text{if } n < 30 \longrightarrow t_0 = \dfrac{\bar{X}-\mu}{S/\sqrt{n}} \end{cases}$$

آماره آزمون مقایسه نسبت دو جامعه :

$$\bar{P}_1 = \frac{X_1}{n_1}, \quad \bar{P}_2 = \frac{X_2}{n_2}$$

$$\bar{P} = \frac{X_1+X_2}{n_1+n_2}$$

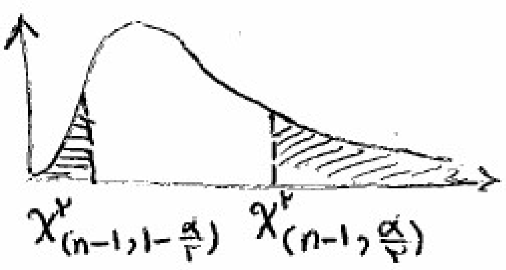
$$Z = \frac{\bar{P}_1 - \bar{P}_2}{\sqrt{\bar{p}\bar{q}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

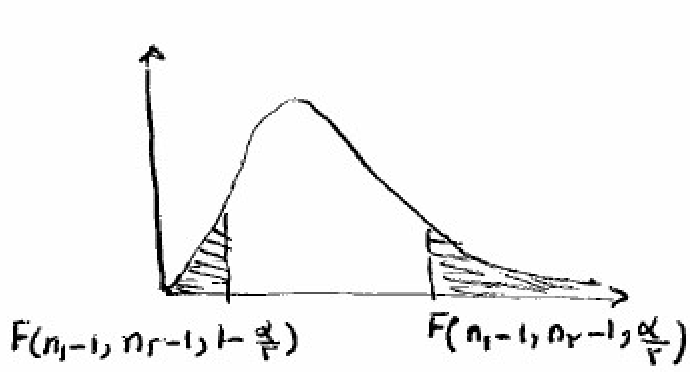
آماره آزمون مربوط به واریانس یک جامعه :

$$\chi_0^2 = \frac{(n-1)S^2}{\sigma^2} \qquad \begin{cases} H_0: \sigma^2 = \sigma_0^2 \\ H_1: \sigma^2 \neq \sigma_0^2 \end{cases}$$

آماره آزمون مربوط به نسبت (مقایسه) واریانس دو جامعه :

$$F_0 = \frac{S_1^2}{S_2^2} \qquad \begin{cases} H_0: \sigma_1^2 = \sigma_2^2 \\ H_1: \sigma_1^2 \neq \sigma_2^2 \end{cases}$$





نکته) درصورتی که فاصله اطمینان شامل عدد صفر نیز باشد فرضیه H_0 را نمی‌توان رد نمود.

آزمون نیکویی برازش (χ^2 ساده) : برای مشخص کردن نوع توزیع مجهول جامعه می باشد

اگر O_i فراوانی مشاهده شده و e_i فراوانی قابل انتظار (فراوانی تجربی) باشد و k تعداد طبقات و m تعداد پارامترها است

$$\chi^2=\sum_{i=1}^{k}\frac{(O_i-e_i)^2}{e_i}\sim\chi^2_{k-m-1}$$

$X \backslash Y$	Y_1	Y_2	...	Y_t	
X_1	O_{11}	O_{12}		O_{1t}	$O_{1.}$
X_2					
$\vdots$					$\vdots$
X_s	O_{s1}	O_{s2}		O_{st}	$O_{s.}$
$O_{.j}$	$O_{.1}$	$O_{.2}$	......	$O_{.t}$	$O_{..}$

آزمون استقلال دو متغیر تصادفی

$$\chi^2=\frac{\sum_{j=1}^{t}\sum_{i=1}^{s}(O_{ij}-e_{ij})^2}{e_{ij}}\sim\chi^2_{(s-1)(t-1)}$$

$$e_{ij}=\frac{O_{i.}O_{.j}}{n}$$

قدر وابستگی $C=\sqrt{\frac{\chi^2}{\chi^2+N}}$ ضریب توافقی گویند

		درجه آزادی
بین گروه‌ها (تیمار)	SSR	$k-1$
داخل گروه‌ها (خطاها)	SSE	$k(n-1)$
کل	SST	$nk-1$

آنالیز واریانس : $H_0: \mu_1=\mu_2=\dots=\mu_M$

در جامعه k تیمار داریم که از هر کدام به اندازه n نمونه اختیار می‌کنیم

تیمار	1	2		n	
1	x_{11}	x_{12}		x_{1n}	$\bar{x}_1$
2	x_{21}	x_{22}		x_{2n}	$\bar{x}_2$
k	x_{k1}	x_{k2}		x_{kn}	$\bar{x}_k$
					$\bar{\bar{x}}$

$$SSR=n\sum_{i=1}^{k}(\bar{\bar{x}}-\bar{x}_i)^2$$

$$SSE=\sum_{j=1}^{n}\sum_{i=1}^{k}(x_{ij}-\bar{x}_i)^2$$

$$SST=\sum_{j=1}^{n}\sum_{i=1}^{k}(x_{ij}-\bar{\bar{x}})^2$$

$$F(k-1,k(n-1))=\frac{\frac{SSR}{k-1}}{\frac{SSE}{k(n-1)}}$$

P-Value : حداقل میزان α جهت رد فرض H_0 می باشد

فصل ۸) رگرسیون:

بهترین تابع پیش‌بینی $E(Y|X)$ می‌باشد

$$E(Y|X=2)=\sum y f(y|X=2)=\frac{\sum y f(y, x=2)}{f_X(2)}$$

در معادله رگرسیون دو متغیره ($\bar{Y}$ و $\bar{X}$) صدق می‌کند بنابراین داریم:

$$\hat{\alpha}=\bar{Y}-\hat{\beta}\bar{X} \quad , \quad \hat{\beta}=\frac{\sum(X_i-\bar{X})(Y_i-\bar{Y})}{\sum(X_i-\bar{X})^2}=\frac{\sum x_i y_i}{\sum x_i^2}=\frac{\sum X_iY_i-n\bar{X}\bar{Y}}{\sum X_i^2-n\bar{X}^2}=\frac{S_{xy}}{S_{xx}}$$

نکته) اگر فرض کنیم $\hat{\alpha}=\bar{Y}$ باشد معادله رگرسیون $\hat{Y}=\hat{\alpha}+\hat{\beta}X_i$ را می‌توان به صورت $\hat{Y}=\hat{\alpha}+\hat{\beta}(X_i-\bar{X})$ نوشت.

$$\text{if } \hat{\alpha}=\bar{Y} \Rightarrow (\hat{Y}_i-\bar{Y})=\hat{\beta}(X-\bar{X}) \qquad y_i=\hat{\beta}x_i$$

توجه شود دیگر در این حالت نبایستی $\hat{\alpha}$ را به صورت $\bar{Y}-\hat{\beta}\bar{X}$ محاسبه کرد و جواب $\bar{Y}$ می‌باشد

حالت‌های خاص:

۱) درصورتی که رگرسیون عرض از مبدأ نداشته باشد ($\alpha=0$)

$$\hat{\beta}=\frac{\sum X_iY_i}{\sum X_i^2}$$

۲) درصورتی که رگرسیون شیب نداشته باشد ($\beta=0$)

$$\hat{\alpha}=\bar{Y}$$

خطای معیار برآورد (Se)

$$Se=\sqrt{\frac{\sum(Y_i-\hat{Y}_i)^2}{n-2}}=\sqrt{\frac{\sum e_i^2}{n-2}}=\sqrt{\frac{\sum(Y_i-\bar{Y})^2-\hat{\beta}^2\sum(X_i-\bar{X})^2}{n-2}}=\sqrt{\frac{S_{yy}-\hat{\beta}^2S_{xx}}{n-2}}=\sqrt{\frac{S_{yy}-\hat{\beta}S_{xy}}{n-2}}$$

$$\hat{\alpha}\sim N\left(\alpha, \frac{\sigma^2\sum X_i^2}{n\sum x_i^2}\right)$$

$$\frac{\hat{\alpha}-\alpha}{S_a}\sim t_{(n-2)} \qquad S_a=\frac{Se}{\sqrt{\frac{1}{n}+\frac{\bar{X}^2}{\sum X_i^2-n\bar{X}^2}}}$$

$$2)\ \hat{\beta}\sim N\left(\beta, \frac{\sigma^2}{\sum x_i^2}\right)$$

$$\frac{\hat{\beta}-\beta}{\frac{Se}{\sqrt{\sum x_i^2}}}\sim t_{(n-2)}$$

جدول آنالیز واریانس:

$$\sum(Y_i-\bar{Y})^2=\sum(\hat{Y}_i-\bar{Y})^2+\sum(Y_i-\hat{Y}_i)^2 \Rightarrow SST=SSReg+SSE$$

	SS	df	MS
Reg	$SSReg=\hat{\beta}S_{xx}$	$K-1$	$\frac{SSReg}{K-1}$
خطا	$SSE=SST-\hat{\beta}S_{xx}$	$n-k$	$\frac{SSE}{n-k}$
کل	$SST=\sum Y_i^2-n\bar{Y}^2$	$n-1$	

$$F(K-1, n-K)=\frac{\frac{SSReg}{K-1}}{\frac{ESS}{n-k}}=\frac{\frac{\sum XY-n\bar{X}\bar{Y}}{K-1}}{\frac{\sum Y_i^2-n\bar{Y}^2-\sum XY+n\bar{X}\bar{Y}}{n-k}}$$

ضریب همبستگی خطی:

$$\rho=r=\frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}=\frac{Cov(X, Y)}{\sqrt{\sigma_x^2}\sqrt{\sigma_y^2}}$$

$$\hat{\beta}=\frac{S_{xy}}{S_{xx}}=\frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}\times\frac{\sqrt{S_{yy}}}{\sqrt{S_{xx}}}=r\sqrt{\frac{S_{yy}}{S_{xx}}}$$

در رگرسیون دو متغیره آماره آزمون $\begin{cases}\rho=0\\ \rho\neq 0\end{cases} \Rightarrow$

$$F(1, n-2)=t_{(n-2)}^2 \Rightarrow t=\sqrt{F}=\sqrt{\frac{R^2/1}{(1-R^2)/n-2}}=\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

ضریب تعیین (r^2): درصدی از تغییرات Y که توسط متغیر X توضیح داده می‌شود

درجه همبستگی بین دو متغیر را با یستی با ضریب تعیین بررسی کنیم نه ضریب همبستگی!

$$R^2 = r^2 = \frac{SSReg}{SSReg + SSE}$$

$$\text{if} \begin{cases} \hat{y} = \hat{\alpha}_1 + \hat{\beta}_1 x_1 \\ \hat{x}_1 = \hat{\alpha}_2 + \hat{\beta}_2 y \end{cases} \Longrightarrow \boxed{r^2 = \hat{\beta}_1 \hat{\beta}_2}$$

در رگرسیون چندگانه $W = \frac{\sum e_i^2}{\sigma^2} = \frac{SSE}{\sigma^2}$ دارای توزیع $\chi^2_{(n-k-1)}$ می‌باشد

$$\bar{R}^2 = 1 - \left(\frac{n-1}{n-k}\right)(1 - R^2)$$

$$\text{آماره } F = \frac{\frac{R_u^2 - R_r^2}{k-1}}{\frac{1 - R_u^2}{n-k}} = \frac{\frac{RSS_r - RSS_u}{m}}{\frac{1 - RSS_r}{n-k}}$$

$$DW = \frac{\sum_{i=2}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

---

k منطقه و n نمونه از هر منطقه

$$F(k-1, k(n-1)) = \frac{\overbrace{SSReg/k-1}^{\text{مجموع مربعات بین گروهی (تیمار)}}}{\underbrace{ESS/k(n-1)}_{\text{مجموع مربعات درون گروهی (خطاها)}}}$$

نکته) توزیع t چون تر و کوتاه‌تر از توزیع z می‌باشد یعنی مقادیر بحرانی آن بزرگتر از z می‌باشد

تذکر: در آزمون فرضیه $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu = \mu_1 \end{cases}$ هرچه $\Delta\mu = \mu_1 - \mu_0$ بیشتر باشد توان آزمون بیشتر است و هرچه $|\mu - \mu_0|$ کمتر باشد توان آزمون بیشتر است

جزوه آمار و احتمالات

(اقتصاد-مدیریت-حسابداری)

کاری از:

ساسان قاراخانی

@eghtesadsanjinovin