بسم الله الرحمن الرحیم

آمار و احتمالات مهندسی

رشته : کامپیوتر

دکتر پرویز نصیری

فصل اول

آمار توصیفی

دراین فصل مسائل زیر بررسی می شود:

- -مفاهیم اساسی
- -شاخص های گرایش مرکزی
- -شاخص های پراکندگی
- -جدول توزیع فراوانی
- -نمودارها
- -چولگی و برجستگی
- -کدگزاری
- -جامعه آماری دو بعدی

مفاهیم اساسی

۱-جامعه

۲-نمونه

۳-داده های آماری

۴-متغیر

۱–جامعه

$$X_1, X_2, \ldots, X_i, \ldots, X_N$$

که x_i عضو i ام جامعه است برای i=1,2,…,N

۲– نمونه

$$X_1, X_2, \ldots, X_i, \ldots, X_N$$

که x_i عضو i ام جامعه است برای i=1,2,…,N

۳-انواع داده های آماری

انواع داده های آماری به دو گروه، داده های دست اول (خام) و داده های دست دوم تقسیم بندی می شوند.

۴-متغیر

انواع آن:

۱-کمی

۲-کیفی

شاخص های گرایش مرکزی:

۱–میانگین
۲– میانه
۳– نما
۴– چارکها

۱- میانگین

الف–میانگین حسابی

ب- میانگین هندسی

پ–میانگین هارمونیک

ت- میانگین پیراسته

الف- میانگین حسابی

فرض کنید جامعه مورد بررسی دارای **N**عضو $X_1, X_2, ..., X_n$ باشد. میانگین جامعه از رابطه زیر بدست می آید.

$$\mu = \overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i = \frac{1}{N}(X_1 + X_2 + ... + X_N)$$

ب– میانگین هندسی

اگر Xn,...,X2,X1 یک نمونه به حجم n از جامعه مـورد بررســی باشد میانگین هندسی از رابطه زیر بدست مـی آیـد و بـا علامـت G نمایش داده می شود.

$$G = \sqrt[n]{x_1, x_2, ..., x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

پ–میانگین هارمونیک

اگر Xn,...,X2,X1 یک نمونه به حجم n از جامعه مـورد بررســی باشـد میـانگین هارمونیک از رابطه زیر بدست می آید و با علامت H نمایش داده می شود.

$$\frac{1}{H} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}$$

یا

$$H = \frac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$$

ت-میانگین پیراسته

اگر k تا از مشاهدات حذف شده باشند میانگین پیراسته از رابطـه زیـر بدست می آید . k<n

$$\bar{x}_p = \frac{1}{n-k}\sum_{i=1}^{n-k} x_i$$

۲-میانه

ویژگی ها:

الف- میانه مشاهدات را به دو بخش مساوی تقسیم می کند.

ب- منحصر به فرد است.

ج- تحت تأثیر داده های پرت قرار نمی‌گیرد.

د- محاسبه آن ساده است.

۳- نما

نمای یک مجموعه عددی است که در آن مجموعه بیش از بقیه تکرار شده باشد.

۴- چارکها

چارکهای یک مجموعه مورد بررسی عبارتست از کمیت‌ها یا مقادیری که مجموعه را به چهار قسمت مساوی تقسیم می‌کنند. محاسبه چارکها همانند میانه می‌باشد.

شاخص های پراکندگی:

۱ – دامنه
۲–واریانس
۳–انحراف معیار
۴–متغیرهای استاندارد
۵–ضریب تغییر یا تعیین
۶–انحراف چارکی
۷–گشتاورها

۱– دامنه

$R = X_{MAX} - X_{MIN}$

۲–واریانس

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

ویژگی های واریانس نمونه:

۱–واریانس عدد ثابت C برابر با صفر است.

۲–اگرمقدار ثابت a رابه مشاهدات اضافه یا ازآنها کم کنیم واریانس تغییر نمی‌کند.

۳–اگر مشاهدات در مقدار ثابت K ضرب یا برآن تقسیم شود واریانس جدید از ضرب یا تقسیم واریانس قدیم در K^2 بدست می آید

۳–انحراف معیار

انحراف معیار در نمونه جذر واریانس یا پراش می باشد.

$$\sqrt{S^2} = S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

μ= میانگین جامعه

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2$$

δ^2 = واریانس جامعه

و جذر آن انحراف معیار جامعه

۴–متغیرهای استاندارد

$$z_i = \frac{x_i - \bar{x}}{S} \qquad , \qquad i = 1,2,\ldots,n$$

ویژگی های متغیرهای استاندارد:

۱– میانگین متغیرهای استاندارد برابر صفر است.

$$\sum_{i=1}^{n} z_i = \sum_{i=1}^{n} \frac{x_i - \bar{x}}{S} = \frac{1}{S}\sum_{i=1}^{n}(x_i - \bar{x}) = \frac{1}{S}\left(\sum x_i - n\bar{x}\right) = \frac{1}{S}(n\bar{x} - n\bar{x}) = 0$$

۲–واریانس متغیرهای استاندارد برابر با ۱ است .

۳– متغیرهای استاندارد فاقد واحد اندازه گیری هستند.

۴– مقدار Z_i می تواند، منفی، صفر یا مثبت باشد.

۵- ضریب تغییر یا ضریب تعیین

$$C.V=\frac{S}{\bar{x}}=\frac{\sqrt{\frac{1}{n}\sum(x_i-\bar{x})^2}}{\frac{1}{n}\sum_{i=1}^{n}x_i}=\frac{\sqrt{n\sum(x_i-\bar{x})^2}}{\sum x_i}$$

ویژگیهای ضریب تغییر

۱- به واحد اندازه گیری بستگی ندارد.

۲- برای مقایسه دو صفت از یک جامعه با واحدهای اندازه گیری متفاوت مورد استفاده قرار می گیرد.

۳- مجموعه مشاهداتی که دارای C.V کمتری است از سازگاری و همگنی بیشتری برخوردار هستند.

۶- انحراف چارکی

$$Q.D = \frac{Q_3 - Q_1}{2}$$

ویژگیهای انحراف چارکی:

۱- این شاخص چون میزان پراکندگی در اطراف مرکز توزیع را نشان می دهد از شاخص دامنه با ثبات تر است.

۲- این شاخص چون شامل ۲۵٪ از مشاهدات کوچک و بزرگ نیست تحت تأثیر داده های پرت قرار نمی گیرد.

۳- این شاخص برای داده های کلاس بندی نیز قابل محاسبه است

۷– گشتاورها

$$m'_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^r \qquad , \qquad r = 1,2,\ldots$$

ویژگیهای گشتاورهای مرکزی:

۱– $m_1=0 \quad , \quad r=1$

۲– $r=2 \qquad m_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = S^2$

۳– تغییر در مبدأ یا اضافه و کم کردن مقدار ثابت به مشاهدات تغییری در m_r ندارد

۴–باتغییر در مقیاس یا ضرب و تقسیم کردن مقدار ثابت در مشاهدات، m_r در توان r ام مقدار ثابت ضرب یا تقسیم می شود

۵– $m_2 = m'_2 - (\bar{x} - a)^2$

جدول توزیع فراوانی

طول کلاس :

$l_i = b_i - a_i$ $\quad i = 1,2,\dots,k$

محاسبه میانگین و واریانس در جدول توزیع فراوانی :

میانگین حسابی

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i x_i$$

میانگین هندسی

$$G = \sqrt[n]{x_1^{f_1}.x_2^{f_2}.....x_k^{f_k}}$$

میانگین هارمونیک

$$H = \frac{n}{\sum_{i=1}^{k}\frac{f_i}{x_i}}$$

واریانس

$$S^2 = \frac{1}{n}\sum_{i=1}^{k} f_i (x_i - \bar{x})^2$$

محاسبه نما در جدول توزیع فراوانی

$$M = a_i + \frac{d_1}{d_1 + d_2} \times l$$

محاسبه میانه در جدول توزیع فراوانی

$$m = a_i + \frac{\frac{n}{2} - F_{i-1}}{f_i} \times l$$

محاسبه چارک ها در جدول توزیع فراوانی

$$Q_j = a_i + \frac{\frac{j \times n}{4} - F_{i-1}}{f_i} \times l$$

نمودارها:

۱–نمودار نقطه ای

۲– نمودار دایره ای

۳–نمودار میله ای

۴– نمودار مستطیلی

۵– نمودار چندضلعی فراوانی

۶– نمودار چند ضلعی تجمعی

چولگی

معیارهای محاسبه میزان چولگی عبارتند از:

۱- ضریب چولگی پیرسن

$$SK_p = \frac{\bar{x} - M}{S}$$

۲- ضریب چولگی بر اساس گشتاور مرکزی مرتبه سوم

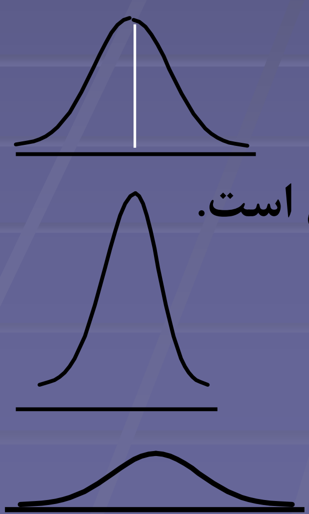
$$b = \frac{m_3}{S^3}$$

برجستگی

$$K = \frac{m_4}{S^4} - 3 = \frac{\frac{1}{n}\sum\left(x_i - \bar{x}\right)^4}{\left[\frac{1}{n}\sum\left(x_i - \bar{x}\right)^2\right]^2} - 3$$

ویژگی های برجستکی:

۱- مستقل از واحد



۲- $k=0$ میزان برجستگی صفر است و منحنی چندضلعی فراوانی بر منحنی نرمال منطبق است.

۳- $k>0$ منحنی چندضلعی فراوانی در مقایسه با منحنی نرمال دارای برجستگی است.

۴- $k<0$ منحنی چندضلعی فراوانی در مقایسه با منحنی نرمال دارای پخی است.

کدگذاری

کدگذاری مجموعه ای داده ها عبارت از عملیاتی است که طی آن از هر مشاهده عدد ثابتی را کم (اضافه) کرده و نتیجه را بر عدد ثابتی تقسیم (ضرب) می‌نمایند.

جامعه آماری دوبعدی

i	1	2	3	…	N
X_i	X_1	X_2	X_3	…	X_n
Y_i	Y_1	Y_2	Y_3	…	Y_n

بسم الله الرحمن الرحيم

فصل دوم

احتمال

دراین فصل مسائل زیر بررسی می شود:

- ۱– فضای نمونه
- ۲– پیشامد
- ۳– شمارش
- ۴– اصول شمارش
- ۵– جایگشت
- ۶– ترکیب
- ۷– احتمال
- ۸– تابع احتمال
- ۹– قوانین احتمال
- ۱۰– احتمال شرطی
- ۱۱– دو پیشامد
- ۱۲– فرمول بیز

۱– فضای نمونه:

- **مجموعه ای از همه برآمدهای ممکن یک تجربه تصادفی را فضای نمونه می‌گویند. و آن را با علامت نمایش می دهند.**
- یک سکه را آنقدر پرتاب می کنیم تا شیر ظاهر شود. فضای نمونه را بنویسید.
- که S گسسته و نامتناهی شمارا است $S=\{H, TH, TTH, \ldots\}$

۲– پیشامد:

هر زیرمجموعه ای از فضای نمونه را یک پیشامد گویند.

۲-۱ رخداد یک پیشامد

۲-۲ دو پیشامد ناسازگار

۲-۳ تفاضل پیشامد A از B

۳– شمارش

تعیین تعداد عناصر یک فضای نمونه متناهی به وسیله شمارش مستقیم، واقعاً مشکل یا لااقل خسته کننده است.

۴–اصول شمارش

فرض کنید کار X با m طریق به نامهای $X_1, X_2, \ldots, X_m$ و کار Y با N طریق به نامهای $Y_1, Y_2, \ldots, Y_n$ قابل انجام باشند. اصول شمارش عبارتند از:

۳–۱ اصل اول شمارش : اگر انجام کار Z منوط به انجام کار X یا Y باشد آنگاه کار Z را می توان به m+n طریق X1,X2,...,Xm و Y1,Y2,...,Yn با نامهای انجام داد.

۳–۲ اصل دوم شمارش : اگر انجام کار Z منوط به انجام کار X یا Y باشد آنگاه کار Z را می توان به m×n طریق زیر انجام داد:

$$(x_1, y_1), (x_1, y_2), \ldots, (x_1, y_n)$$
$$(x_2, y_1), (x_2, y_2), \ldots, (x_2, y_n)$$
$$\vdots$$
$$(x_m, y_1), (x_m, y_2), \ldots, (x_m, y_n)$$

مثال: چند عدد زوج سه رقمی از ارقام ۱، ۲، ۵، ۶ و ۹ می توان نوشت به طوریکه هر رقم فقط یک بار استفاده شود؟

از اینکه اعداد زوج باشد، برای رقم یکان فقط دو انتخاب وجود دارد پس کل طرق برابر است با ۲×۴×۳=۲۴ .

۵-جایگشت

ترتیبی از مجموعه n شیء با آرایش معین جایگشت اشیاء خوانده می شود

۱-۴ جایگشت n شی ء متمایز n(n-1)(n-2)×…×2×1-n!

۲-۴ جایگشت r تایی n شی ء متمایز n(n-1)(n-2)…(n-r+1)

۳-۴ جایگشت r تایی n شی ء متمایز با تکرار $n\,\mathrm{P}'\mathrm{r} = n\times n\times n\times \ldots \times n = n^r$

۴-۴ جایگشت با اشیاء مکرر

$$\frac{n!}{n_1!n_2!\ldots n_k!} = \binom{n}{n_1, n_2, \ldots, n_k}$$

۵-۴ جایگشت n شیء متمایز در محیط دایره

۶- ترکیب

هرگاه در جایگشت، آرایش و نظم اشیا کنار هم مورد توجه نباشد آن را ترکیب گویند.

۵-۱ ترکیب r تایی n شیء متمایز

$$_{n}C_{r}=\frac{n!}{r!(n-r)!}\longrightarrow \quad x=\frac{_{n}P_{r}}{r!}=\frac{n!}{r!(n-r)!}={}_{n}C_{r}$$

۵-۲ ترکیب r تایی n شیء با تکرار اشیاء

$$_{n+r-1}C_{r}=\frac{(n+r-1)!}{r!(n-1)!}$$

۷- احتمال

مفهوم کلاسیک:

$$P(A)=\frac{\text{تعداد حالات مساعد}}{\text{تعداد حالات کل}}=\frac{n(A)}{n(S)}=\frac{n}{A}$$

مفهوم فراوانی :احتمال یک پیشامد برابر با نسبت دفعاتی است که پیشامدهای از یک نوع در تکرار زیاد رخ خواهند داد، احتمال به مفهوم فراوانی تلقی می شود.

۸- تابع احتمال

تابعی را که به هر پیشامد عددی در بازه (۱،۰) نسبت دهد و در سه اصل زیر صدق کند تابع احتمال گویند .

اصل اول: احتمال هر پیشامد بزرگتر یا مساوی صفر است.

$$P(A) \geq 0 \quad , \qquad \forall A \subset S$$

اصل دوم: احتمال فضای نمونه S برابر با ۱ می باشد.

$$P(S) = ۱$$

اصل سوم:

$$P[A_1 \cup A_2 \cup \ldots] = \sum_{i=1} P(A_i) = P(A_1) + P(A_2) + \ldots$$

۹- قوانین احتمال

قضیه ۹-۱ اگر ϕ مجموعه تهی باشد آنگاه P()=0

برهان: می دانیم $S \cup \phi = S$ و S و ϕ دو مجموعه مجزا هستند. یعنی $S \cap \phi = \phi$ طبق اصل دوم و سوم.

$$P(S \cup \phi) = P(S) + P(\phi)$$

$$1 = 1 + P(\phi) \Rightarrow P(\phi) = 0$$

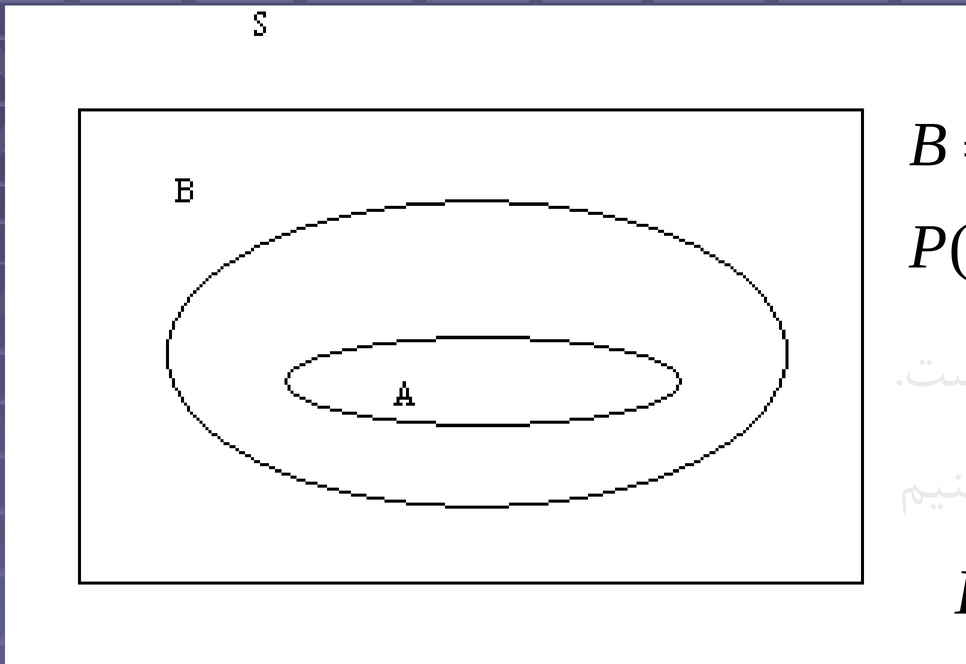
قضیه ۹-۲ اگر A^C متمم پیشامد A باشد آنگاه $P(A^C)=1-P(A)$

برهان: می دانیم $A \cap A^c = \phi$ و $A \cup A^c = S$ پس:

$$P(A \cup A^c) = P(S)\ ,\quad P(A) + P(A^c) = 1 \Rightarrow P(A) \equiv 1 - P(A^c)$$

قضیه ۹-۳ اگر $A \subset B$ باشد آنگاه $P(A) \le P(B)$

برهان: اگر $A \subset B$ باشد B را می توان به صورت دو پیشامد مجزای A و $B \cap A^c$ نوشت.



$$B = A \cup (B \cap A^c)$$

$$P(B) = P(A) + P(B \cap A^c)$$

طبق اصل اول احتمال $P(B \cap A^c) \ge 0$ است.

اگر آن را از طرف راست رابطه اخیر حذف کنیم نتیجه می شود:

$$P(A) \le P(B)$$

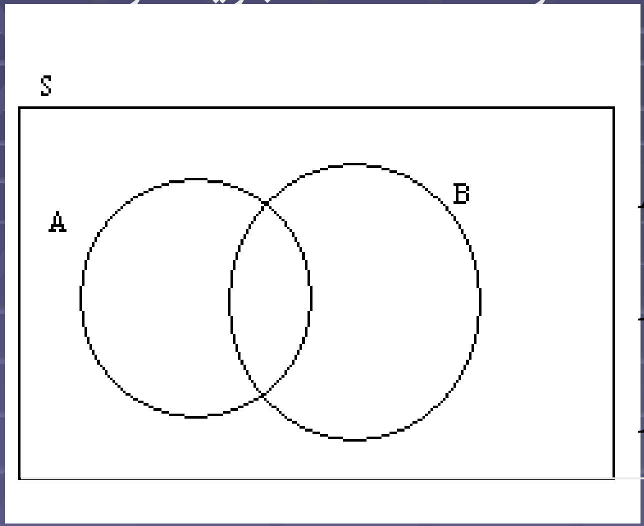
قضیه ۹-۴ اگر A یک پیشامد باشد آنگاه $0 \le P(A) \le 1$

برهان: چون $\Phi \subset A \subset S$ طبق قضیه ۹-۳ داریم:

$$P(\phi) \le P(A) \le P(S) \quad , \quad 0 \le P(A) \le 1$$

قضیه۹-۵ اگر A، B دو پیشامد دلخواه در S باشند آنگاه $P(A\cap B^c)=P(A)-P(A\cap B)$

برهان: پیشامد A را می توان به دو پیشامد مجزای $A\cap B^c$ و $A\cap B$ تجزیه کرد



$$A=(A\cap B^c)\cup(A\cap B)$$

$$P(A)=P(A\cap B^c)+P(A\cap B)$$

$$P(A\cap B^c)=P(A)-P(A\cap B)$$

قضیه۹-۶ اگر A، B دو پیشامد دلخواه در S باشند آنگاه $P(A\cup B)=P(A)+P(B)-P(A\cap B)$

برهان: پیشامد $A\cup B$ را می توان به دو پیشامد مجزای $A\cap B^c$ و B تجزیه کرد.

$$A\cup B=B\cup(A\cap B^c)\xrightarrow{\text{طبق قضیه ۹-۵}}P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

$$P(A\cup B)=P(B)+P(A\cap B^c)$$

قضیه۹–۸ اگر A ، B و C پیشامدهای دلخواه در S باشند آنگاه:

$$P(A\cup B\cup C)=P(A)+P(B)+P(C)-P(A\cap B)-P(A\cap C)$$
$$-P(B\cap C)+P(A\cap B\cap C)$$

۱۰– احتمال شرطی

احتمال شرطی پیشامد A به شرط وقوع پیشامد B به صورت زیر تعریف می‌شود

$$P(A|B)=\frac{P(A\cap B)}{P(B)} \qquad , \qquad P(B)>0$$

نکته۱۰–۱ اگر $P(A)\neq 0$ از $P(A|B)=\frac{P(A\cap B)}{P(B)}$ نتیجه می شود که:

$$P(A\cap B)=P(A|B)P(B)=P(B|A)P(A)$$

نکته ۱۰-۲اگر $P(B)=0$ باشد نتیجه می شود که $P(A\cap B)=0$ در این صورت $P(A|B)$ تعریف نشده است.

نکته ۱۰-۳ اگر پیشامدهای A_1 ، A_2، ... ، A_k دوبه دو مجزا باشند. احتمال شرطی $\bigcup_{i=1}^{k} A_i$. به شرط B برابر با:

$$P\left[\bigcup_{i=1}^{k} A_i \mid B\right]=\frac{P\left(\bigcup_{i=1}^{k} A_i \cap B\right)}{P(B)}=\frac{\sum_{i=1}^{k} P(A_i\cap B)}{P(B)}$$

۱۱– دو پیشامد مستقل

دو پیشامد A و B را مستقل گوییم، اگر رخ داد یکی تأثیری در دیگری نداشته باشد. یعنی $P(B|A)=P(B), P(A|B)=P(A)$ بنابراین A و B مستقل اند اگر :

$$P(A\cap B)=P(A\mid B)P(B)=P(B\mid A)P(A)=P(A).P(B)$$

قضیه ۱۱–۱ اگر دو پیشامد A و B مستقل باشند آنگاه A و B نیز مستقل اند.

برهان:

$$A\cup A^c=S \qquad , \qquad B=(A\cap B)\cup(A^c\cap B)$$

$$P(B)=P(A\cap B)+P(A^c\cap B)$$

$$=P(A)P(B)+P(A^c\cap B)$$

$$P(A^c\cap B)=P(B)-P(A)P(B)=P(B)(1-P(A))=P(B)P(A^c)$$

قضیه ۱۱–۲ پیشامدهای A_1 ، A_2 ، ...، A_k مستقل اند اگر و تنها اگر احتمال اشتراک هر ۲، ۳، ...، k تا از این پیشامدها مساوی حاصلضرب احتمالهای مربوطه به هر پیشامد باشد.

برای استقلال سه پیشامد A_1، A_2 و A_3 لازم است که :

$$1)\ P(A_1\cap B_2)=P(A_1)P(A_2)\ ,\ P(A_1\cap A_3)=P(A_1)P(A_3)\ ,\ P\left(A_2\cap A_3\right)=P\left(A_2\right)P\left(A_3\right)$$

$$2)\ P(A_1\cap A_2\cap A_3)=P(A_1)P(A_2)\,P(A_3)$$

قضیه ۱۱-۳ اگر احتمال وقوع پیشامد A_1 برابر P_1 و احتمال وقوع پیشامد A_2 برابر P_2 و دو پیشامد A_1 و A_2 مستقل باشند آنگاه احتمال اینکه فقط یکی از آنها اتفاق بیفتد برابر است با:

$$P_1(1-P_2)+(1-P_1)P_2$$

برهان: رخداد پیشامد A_1 برابر با رخ داد پیشامد A_1 اشتراکش با A_2 و رخداد پیشامد A_2 برابر با رخداد پیشامد A_2 اشتراکش با A^c_1 است. پس:

$$A_1 = A_1 \cap A_2^c \qquad , \qquad A_2 = A_1^c \cap A_2$$

چون A_1 و A_2 مستقل اند و مجزا هستند

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) = P(A_1 \cap A_2^c) + P(A_1^c \cap A_2)$$

$$= P(A_1)P(A_2^c) + P(A_1^c)P(A_2)$$

$$= P_1(1-P_2)+(1-P_1)P_2$$

قضیه ۱۱-۴ (قانون جمع احتمالات) فرض کنید پیشامدهای A_1 ، A_2 ، .... ، A_k پیشامدهای دو به دو مجزا از هم و اجتماع آنها S باشد و A یک پیشامد دلخواه از S باشد آنگاه:

$$P(A)=\sum_{i=1}^{k}P(A|A_1)P(A_i)$$

برهان:

$$A\cap S=A \quad , \quad \bigcup_{i=1}^{k}A_i=S$$

$$A=A\cap\left(\bigcup_{i=1}^{k}A_i\right)=A\cap(A_1\cup A_2\cup....\cup A_k)$$

$$=(A\cap A_1)\cup(A\cap A_2)\cup....\cup(A\cap A_k)$$

پیشامدهای طرف راست رابطه اخیر دوبه دو مجزا هستند. طبق اصل سوم

$$P(A)=P(A\cap A_1)+P(A\cap A_2)+....+P(A\cap A_k)$$

$$=\sum_{i=1}^{k}P(A\cap A_i)=\sum_{i=1}^{k}P(A|A_i)P(A_i)$$

۱۲ – فرمول بیز

اگر پیشامدهای A_1 ، A_2 ، ...، A_k دو به دو مجزا و $\bigcup_{i=1}^{k} A_i = S$ باشد احتمال شرطی هریک از A ها به شرط اتفاق پیشامد A از S برابر با:

$$P[A_i \mid A] = \frac{P(A_i \cap A)}{P(A)}$$

با توجه به فرمول قضیه ۱۱–۴

$$P[A_i \mid A] = \frac{P(A \mid A_i)P(A_i)}{\sum_{i=1}^{k} P(A \mid A_i)P(A_i)}$$

بسم الله الرحمن الرحيم

فصل سوم

توزیع متغیرهای تصادفی

در این فصل مسائل زیر بررسی می شود:

- ۱- متغیر تصادفی
- ۲- متغیر تصادفی گسسته
- ۳- متغیر تصادفی پیوسته
- ۴- تابع توزیع $F_{(x)}$
- ۵- تابع احتمال و تابع توزیع توام دو متغیر تصادفی
- ۶- تابع توزیع توام
- ۷- تابع چگالی احتمال و تابع توزیع حاشیه ای
- ۸- تابع چگالی احتمال و تابع توزیع شرطی
- ۹- استقلال دو متغیر تصادفی
- ۱۰- امید ریاضی
- ۱۱- گشتاورها
- ۱۲- ضریب همبستگی دو متغیر تصادفی
- ۱۳- چولگی و برجستگی در جامعه
- ۱۴- تابع مولد گشتاورها
- ۱۵- نامساوی مارکف و چبیشف

۱– متغیر تصادفی

با فرض اینکه هر تجربه تصادفی دارای فضای نمونه S باشد با تدوین یک قانون یا مجموعه ای از قوانین می‌توان اعضای فضای نمونه را به وسیله اعداد یا زوج اعداد (X_1,X_2) یا به طور کلی تر با n گانه مرتب اعداد $(X_1,X_2,\ldots,X_n)$افراز کرد.

۲– متغیر تصادفی گسسته

فرض کنید متغیر تصادفی X دارای فضای نمونه یک بعدی A باشد. به طوری که A گسسته و شمارا باشد. هرگاه بتوان تابع احتمال $P(A)$ $(A\subset A)$را برحسب تابع $f(X)$ به شکل زیر تعریف کرد:

$$P(A) = P(X \in A) = \sum_{A} f(x)$$

به طوری که f(X) در دو شرط زیر صدق کند.

۱- $$f(x) \geq 0 \quad , \quad \forall\ x \in A$$

۲- $$\sum_{A} f(x) = 1$$

Xرا متغیر تصادفی از نوع گسسته و f(X) را تابع احتمال یا پخش گسسته X گویند.

۳- متغیر تصادفی پیوسته

فرض کنید متغیر تصادفی X دارای فضای نمونه یک بعدی A باشد. به طوری که A پیوسته و بازه از اعداد حقیقی باشد. هرگاه بتوان تابع احتمال P(A)(A⊂ A)را برحسب تابع f(X) به شکل زیر تعریف کرد:

$$P(A) = P(X \in A) = \int_{A} f(x)dx$$

۴F(x)– تابع توزیع

$$F(x) = P[X \le x]$$

- تابع توزیع متغیرهای تصادفی از نوع گسسته و پیوسته به ترتیب به صورت زیر تعریف می شوند.

$$F(x) = P(X \le x) = \sum_{t \le x} f(t)$$

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$$

خواص تابع توزیع (گسسته یا پیوسته):

۱- $0 \le P(X \le x) \le 1$ یا $0 \le F(x) \le 1$

۲- F(x) یک تابع غیر نزولی است.

۳- $\lim_{x \to +\infty} F(x) = F(+\infty) = 1$ و $\lim_{x \to -\infty} F(x) = F(-\infty) = 0$

۴- F(x) در هر نقطه x از راست پیوسته است.

۵- $P(a \le X \le b) = F(b) - F(a)$

۶- $P(X = x) = F(x) - F(x^{-})$

که درآن $F(X^{-})$ حد چپ $F(X)$ در نقطه X است.

۷- $F(x^{-}) \le F(x) \le F(x^{+})$

۸- الف: در متغیر پیوسته $f(x) = \dfrac{dF(x)}{dx}$ یا $F(x) = \int_{-\infty}^{x} f(t)\,dt$

ب: در متغیر گسسته $f(x) = F(x) - F(x^{-})$

که در آن $F(x) = \sum_{x \le t} f(t)$

مثال: اگر متغیر تصادفی پیوسته X دارای تابع توزیع F(X) باشد تابع چگالی احتمال آن را بدست آورید.

$$F(x)=\begin{cases}0 & x<1\\ \dfrac{(x-1)^2}{8} & 1\le x<3\\ 1 & x\ge 3\end{cases}$$

همانطور که ملاحظه می کنید F(X) از راست پیوسته است چون:

F(1)=0

$$\lim_{x\to 1^+}F(x)=\lim_{x\to 1^+}\frac{(x-1)^2}{8}=\lim_{\varepsilon\to 0}\frac{(1+\varepsilon-1)^2}{8}=0$$

$$F(3)=1$$

$$\lim_{x\to 3^+}F(x)=\lim_{\varepsilon\to 0}F(3+\varepsilon)=1$$

$$\lim_{x\to 3^-}F(x)=\lim_{\varepsilon\to 0}F(3-\varepsilon)=\lim_{\varepsilon\to 0}\frac{(3-\varepsilon-1)^2}{8}=\frac{1}{2}$$

$$f(x)=\frac{d\ F(x)}{dx}=\begin{cases}0 & x<1\\ \dfrac{x-1}{4} & 1\le x<3\end{cases}$$

۵ – تابع احتمال و تابع توزیع توام دو متغیر تصادفی

فرض کنید متغیرهای تصادفی X و Y دارای فضای دوبعدی A باشد. بـه طـوری کـه A گسسته و شمارا باشد. هرگاه بتوان تـابع احتمـال $P(A)$ را برحسـب تـابع $f(x,y)$ بـه شکل زیر تعریف کرد

$$P(A) = P[(X,Y) \in A] = \sum_{A}\sum f(x,y)$$

به طوری که $f(x,y)$ در دو شرط زیر صدق کند.

۱– $\forall (x,y) \in A \qquad f(x,y) \geq 0$

۲– $\sum_{A}\sum f(x,y) = 1$

(X, Y) را متغیرهای تصادفی توام از نوع گسســته و $f(x,y)$ را تـابع چگـالی احتمـال یـا پخش توام گسسته گویند.

و برای حالتی که X و Y متغیرهای تصادفی از نوع پیوسته اند، می توان تابع چگالی احتمال $f(x,y)$ را روی همه صفحه تعریف کرد. تابع دومتغیره $f(x,y)$ را تابع چگالی احتمال توام متغیر X و Yگوییم اگر و تنها اگر برای هر ناحیه A از صفحهxy

$$P[(X,Y)\in A]=\int_A\int f(x,y)\,dy\,dx$$

و $f(x,y)$ همواره در دو شرط زیر صدق نماید.

۱- $$f(x,y)\geq 0 \qquad \forall\ A\in A$$

۲- $$\int_A\int f(x,y)\,dy\,dx=1$$

۶– تابع توزیع توام

اگر X و Y متغیرهای تصادفی با تابع چگالی احتمال توام $f(x,y)$ باشند تابع توزیع یا تابع توزیع تجمعی توام X و Y در حالت گسسته و پیوسته به ترتیب به صورت زیر تعریف می شود.

$$F(x,y) = P[X \leq x \, , Y \leq y] = \sum_{S \leq x} \sum_{t \leq y} f(s,t)$$

$$F(x,y) = P[X \leq x \, , Y \leq y] = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)\, dt\, ds$$

تبصره: محاسبه احتمال (X,Y) روی A به طوری که $A = \{(x,y) \mid a \leq x < b \quad , \quad c \leq y < d\}$ از فرمول زیر محاسبه می شود:

$$P[a \leq X < b \, , \; c \leq Y < d] = F(b,d) - F(a,d) - F(b,c) + F(a,c)$$

۷– تابع چگالی احتمال و تابع توزیع حاشیه ای

فرض کنید دو متغیر تصادفی X و Y دارای تابع چگالی احتمال توام $f(x,y)$ باشند و بخواهیم احتمال پیشامد $-\infty < X \le x$ را حساب کنیم. محاسبه احتمال پیشامد $-\infty < X \le x$ برای دو متغیر تصادفی X و Y با تابع چگالی احتمال توام $f(x,y)$ هم ارز است با محاسبه احتمال پیشامد ($-\infty < Y < \infty$ ، $-\infty < X \le x$) پس

$$P[-\infty < X \le x] = P[-\infty < Y < \infty \ , \ -\infty < X \le x]$$

محاسبه احتمال رابطه اخیر در حالت گسسته و پیوسته به ترتیب برابرند با:

$$P[-\infty < X \le x, \ -\infty < Y < \infty] = \sum_{S \le x} \sum_{-\infty}^{\infty} f(s,y)$$

$$P[-\infty < X \le x, \ -\infty < Y < \infty] = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(s,y)\, dy\, ds$$

اگر تعریف کنیم:

$$F_X(x) = P[X \le x] = P[-\infty < X \le x\ ,\ \ -\infty < Y < \infty] = \sum_{s\le x}\sum_{-\infty}^{\infty} f(s,y) = \sum_{s\le x} f_x(s)$$

$$F_X(x) = P[X \le x] = P[-\infty < X \le \infty,\ -\infty < Y < \infty] = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f(s,y)\,dy\,ds = \int_{-\infty}^{x} f_x(s)ds$$

تابع های $F_X(x) = \sum_{s\le x} f_x(s)$ و $F_X(x) = \int_{-\infty}^{x} f_x(s)\,ds$ را به ترتیب تابع توزیع حاشیه ای در حالت گسسته و پیوسته گویند. با معلوم بودن تابع توزیع حاشیه‌ای، تابع چگالی احتمال حاشیه ای برای متغیرهای گسسته و پیوسته به ترتیب از رابطه های زیر بدست می آیند.

تابع حاشیه ای Xدر حالت گسسته

$$f_x(x) = F_X(x) - F_X(x^-)$$

تابع حاشیه ای Yدر حالت گسسته

$$f_y(y) = F_Y(y) - F_Y(y^-)$$

تابع حاشیه ای X در حالت پیوسته

$$f_x(x) = \frac{\partial F(x,y)}{\partial y}$$

تابع حاشیهای Y در حالت پیوسته

$$f_y(y)=\frac{\partial F(x,y)}{\partial x}$$

اگر تابع چگالی احتمال توام f(x,y) معلوم باشد تابع چگالی احتمال حاشیه ای X و Yبرای حالت گسسته و پیوسته به ترتیب از روابط زیر بدست می آیند.

$$f_Y(y)=\sum_{x} f(x,y) \quad , \quad f_X(x)=\sum_{y} f(x,y)$$

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

یادآوری می شود که هریک از توابع چگالی حاشیه ای X و Y به نوبه خود تابع چگالی احتمال می باشند و در تمام شرایط تابع چگالی بودن صدق می کنند.

۸ – تابع چگالی احتمال و تابع توزیع شرطی

فرض کنید دو متغیر تصادفی X و Y از نوع گسسته، دارای تابع احتمال توام $f(x,y)$ ، تابع های چگالی احتمال حاشیه ای $f_x(x)$، $f_y(y)$ و فضای نمونه A باشند. دو پیشامد A_1و A_2را به صورت زیر درنظر می گیریم.

$$A_1 \subset A = \{(x,y) | x = x_1, -\infty < y < +\infty\}$$

$$A_2 \subset A = \{(x,y) | -\infty < x < +\infty, y = y_1\}$$

می دانیم که:

$$P(A_1) = P[(x,y) \in A_1] = \sum_x \sum_y f(x,y) = \sum_x f_x(x) = f(x)$$

$$P(A_2) = P[(x,y) \in A_2] = \sum_x \sum_y f(x,y) = \sum_y f_y(y) = f(y)$$

احتمال شرطی پیشامد A_1 به شرط A_2 برابر است با:

$$P(A_1 | A_2) = \frac{P(A_1 \cap A_2)}{P(A_2)} = \frac{f(x_1, y_1)}{f(y)}$$

اگر تعریف کنیم:

$$f(x_1 | y_1) = \frac{f(x_1, y_1)}{f(y_1)}$$

آنگاه تابع احتمال شرطی x_1 به شرط y_1 برابر است با:

$$f(x_1 | y_1) = \frac{f(x_1, y_1)}{f(y_1)} \quad , \quad f(y_1) > 0$$

برای سادگی تابع احتمال X به شرط Y را به صورت زیر تعریف می کنیم.

$$f(x | y) = \frac{f(x, y)}{f(y)} \quad , \quad f(y) > 0$$

و تابع احتمال Y به شرط X را به صورت $f(y|x)$تعریف می کنیم.

$$f(y\,|\,x)=\frac{f(x,y)}{f(x)} \qquad , \qquad f(x)>0$$

در حالتی که متغیرهای تصادفی X و Y پیوسته باشند از همین نماد استفاده می‌کنیم.

$$f(x\,|\,y)=\frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)\,dx}=\frac{f(x,y)}{f(y)} \qquad , \qquad f(y)>0$$

$$f(y\,|\,x)=\frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)\,dy}=\frac{f(x,y)}{f(x)} \qquad , \qquad f(x)>0$$

یادآوری می شود که تابع چگالی احتمال شرطی نیز به نوبه خود یک تابع چگالی احتمال است و در تمام شرایط چگالی بودن صدق می کند.

۹ - استقلال دو متغیر تصادفی

فرض کنید دو متغیر تصادفی X و Y دارای تابع چگالی احتمال توام $f(x,y)$ ، توابع چگالی حاشیه ای $f_x(x)$، $f_y(y)$ ، و توابع چگالی احتمال شرطی $f(x|y)$ و$f(y|x)$ باشند.
گوییم دو متغیر تصادفی X و Y به طور احتمالی مستقل اند اگر:

$$f(x,y)=f_x(x)f_y(y)$$

یا تابع چگالی احتمال شرطی X به شرط Y مستقل از Y باشد یا تابع چگالی احتمال شرطیY به شرط X مستقل از X باشد. یا:

$$f(x,y)=f(X\mid y)f_y(y)=f(Y\mid x)f_X(x)$$

۱۰ - امید ریاضی

اگر متغیر تصادفی X دارای تابع چگالی احتمال $f(x)$ باشد. امید ریاضی در حالت گسسته و پیوسته به ترتیب به صورت زیر تعریف می شود.

$$E(X) = \sum_{x} x \, . \, f(x)$$

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

در ادبیات آماری امید ریاضی را معمولاً با μ نمایش می دهند.

ویژگیهای امید ریاضی

۱- امید ریاضی مقدار ثابت برابر با خودش است.

$$E(c) = c$$

۲-

$$E\left[\sum_{i=1}^{n} a_i X_i\right] = \sum_{i=1}^{n} a_i E(X_i)$$

۳-

$$E\left[\sum a_i X_i + \sum b_i Y_i\right] = \sum a_i E(X_i) + \sum b_i (Y_i)$$

۴- اگر X و Y مستقل از هم باشند.

$$E(XY) = E(X).E(Y)$$

۱۱ – گشتاورها

اگر متغیر تصادفی X دارای تابع چگالی احتمال $f(x)$ و a یک عدد ثابت حقیقی باشد
گشتاورهای مرتبه rام حول نقطه a در جامعه در حالت گسسته و پیوسته به ترتیب
به صورت زیر تعریف می شود.

$$E(X-a)^r = \sum_x (x-a)^r f(x)$$

$$E(X-a)^r = \int_{-\infty}^{\infty} (x-a)^r f(x)\,dx$$

در ادبیات آماری، معمولاً گشتاورهای حول نقطه میانگین جامعه (μ) را گشتاورهای
مرکزی می گویند و با نماد μ_r نمایش می دهند.

$$\mu_r = E(X-\mu)^r = \sum_x (x-\mu)^r f(x)$$

$$\mu_r = E(X-\mu)^r = \int_{-\infty}^{\infty} (x-\mu)^r f(x)\,dx$$

۱۱–۱ واریانس

اگر در فرمول گشتاورهای مرکزی r برابر با ۲ درنظر گرفته شود واریانس جامعه بدست می آید و معمولاً آن را با نمادσ^2 یا $V(X)$ نمایش می دهند.

$$V(X)=\sigma^2=E(X-\mu)^2=\sum_x (x-\mu)^2 f(x)$$

$$V(X)=\sigma^2=E(X-\mu)^2=\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx$$

باتوجه به خواص عملگر E می توان واریانس X را به صورت زیر تعریف کرد.

$$\begin{aligned} V(X)&=E(X-\mu)^2=E[X^2+\mu^2-2\mu X]=E(X^2)+\mu^2-2\mu E(X)\\ &=E(X^2)+\mu^2-2\mu^2=E(X^2)-\mu^2 \end{aligned}$$

۱۱–۲ ویژگیهای واریانس

۱- واریانس مقدار ثابت صفر است.

۲-
$$V(C)=0$$

۳-
$$V(aX)=a^2V(X)$$

$$V(aX+a)=a^2V(X)+0$$

۴-
$$V\left[\sum_{i=1}^{n} a_i X_i\right]=\sum_{i=1}^{n} a_i^2 V(X_i)$$

۵- جذر واریانس را انحراف معیار گویند.

$$\sqrt{\sigma^2}=\sqrt{V(X)}=\sigma$$

برای نمونه ویژگی ۳ را می توان به صورت زیر ثابت کرد.

$$V(aX+c)=E[aX+c-E(aX+c)]^2$$
$$=E[aX+c-aE(X)-c]^2=E[aX-aE(X)]^2$$
$$=E[a(X-E(X))]^2=a^2E(X-\mu)^2=a^2V(X)$$

۱۱–۳ کوواریانس دو متغیر تصادفی

اگر متغیرهای تصادفی X و Y دارای تابع چگالی احتمال توام $f(x,y)$ باشند کوواریانس آنها به صورت زیر تعریف می شود.

$$Cov(X,Y) = E(X-\mu_x)(Y-\mu_y)$$

که μ_x و μ_y به ترتیب امید ریاضی X و Y می باشند. رابطه بالا را می توان به صورت زیر نیز نوشت:

$$\begin{aligned}Cov(X,Y) &= E[XY-\mu_y X-\mu_x Y+\mu_x\mu_y]\\ &= E(XY)-\mu_y E(X)-\mu_x E(Y)+\mu_x\mu_y\\ &= E(XY)-\mu_x\mu_y-\mu_x\mu_y+\mu_x\mu_y = E(XY)-\mu_x\mu_y\end{aligned}$$

قضیه ۱۱–۳–۱ اگر دو متغیر X و Y مستقل باشند آنگاه:

$$Cov(X,Y)=0$$

قضیه ۱۱–۳–۲ اگر X و Y دارای تابع چگالی احتمال توام $f(x,y)$ باشند آنگاه:

$$V(X+Y)=V(X)+V(Y)+2Cov(X,Y)$$

۱۲– ضریب همبستگی دو متغیر تصادفی

ضریب همبستگی دو متغیر تصادفی X و Y را در جامعه با ρ نمایش می دهند و به صورت زیر تعریف می شود.

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X-\mu_x)^2 . E(Y-\mu_y)^2}}$$

۱۲–۱ ویژگیهای ضریب همبستگی:

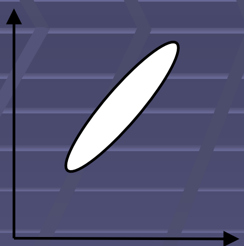
۱– همواره $-1 \le \rho \le 1$ و مستقل از واحد اندازه گیری است.

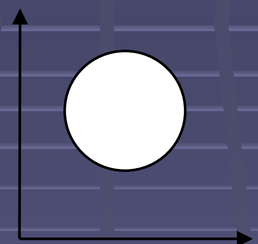
۲– هنگامی که ۱= ρاست همبستگی دو متغیر X و Y شدید و هم سو است.

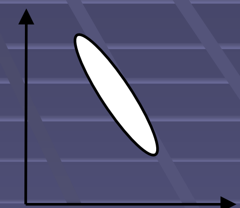
۳– هنگامی که ۱- = ρ است همبستگی دو متغیر X وY شدید و خلاف هم است.

۴– هنگامی که ρ در همسایگی صفر است همبستگی دو متغیر ضعیف است.

۵- باتوجه به مقدار ρ ، نمودار پراکنش X و Y به صورت زیر دسته بندی می‌شود


ρ=1


ρ=0


ρ=-1

۶- $$\rho_{aX+b,\, cY+d} = \rho_{X,Y}$$

یعنی اگر متغیرهای X و Y را در مقدار ثابت ضرب کنیم و مقدار ثابت به آنها اضافه کنیم تغییری در همبستگی ایجاد نمی شود.

۱۳ – چولگی و برجستگی در جامعه

در آمار توصیفی میزان چولگی و برجستگی را به ترتیب از فرمولهای زیر محاسبه کردیم.

$$b = \frac{m_3}{S^3} \qquad\qquad k = \frac{m_4}{S^4} - 3$$

میزان چولگی و برجستگی در جامعه به ترتیب از فرمولهای زیر محاسبه می شود :

$$\alpha_3 = \frac{\mu_3}{\sigma^3} = \frac{E(X-\mu)^3}{\sigma^3}$$

$$\alpha_4 = \frac{\mu_4}{\sigma^4} = \frac{E(X-\mu)^4}{\sigma^4} - 3$$

۱۴ – تابع مولد گشتاورها

اگر متغیر تصادفی X دارای تابع چگالی احتمال *f(x)* باشد تابع مولد گشتاورها در حالت گسسته و پیوسته به ترتیب به صورت زیر تعریف می شود.

$$M_X(t) = E[e^{tX}] = \sum_x e^{tx} f(x)$$

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$$

بطوری که $|t|<h$.

تابع مولد گشتاورها تعریف خاصی از امید ریاضی است. اگر $h(X)=e^{tx}$ تعریف شود $E[h(X)]$همان تعریف تابع مولد گشتاورها است و در مواقعی که محاسبه $E(X^r)$ برای بعضی از توزیع ها وقت گیر است از $M_X(t)$ استفاده می شود.

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$$

از $M_X(t)$ نسبت به t مشتق های متوالی می گیریم.

$$M'_X(t) = \frac{d}{dt} E[e^{tX}] = E(X\, e^{tX})$$

$$M''_X(t) = \frac{d^2}{dt^2} E[e^{tX}] = E(X^2\, e^{tX})$$

پس از r بار مشتق گیری

$$M^r_X(t) = \frac{d^r}{dt^r} E[e^{tX}] = E(X^r\, e^{tX})$$

برای t=0در مشتق rام

$$E(X^r) = M^r_X(0)$$

۱-۱۴ ویژگیهای تابع مولد گشتاورها

$$M_X(0) = 1$$

$$M'_X(0) = E(X) = \mu$$

$$M''_X(0) - (M'_X(0))^2 = V(X)$$

$$M_{\frac{X+a}{b}}(t) = e^{\frac{at}{b}} M_X\left(t/b\right)$$

۱۵– نامساوی مارکف و چبیشف

۱۵–۱ نامساوی مارکف

فرض کنید متغیر تصادفی X دارای فضای مفروض A باشد به طوری که اعضای A همه مثبت باشند و α یک عدد بزرگتر از صفر باشد. نامساوی مارکف را تحت قضیه زیر بیان می کنیم.

قضیه۱۵–۱ اگر متغیر تصادفی X دارای فضای A باشد و E(X) موجود باشد آنگاه همواره:

$$P[X \geq a] \leq \frac{E(X)}{a}$$

برهان: تابع اشاره I(X) را به صورت زیر تعریف می کنیم.

$$I(X) = \begin{cases} 1 & X \geq a \\ 0 & X < a \end{cases}$$

چون 0≤x است پس:

$$I(X) \leq \frac{X}{a}$$

از طرفین رابطه اخیر امید ریاضی می گیریم:

$$E[I(X)] \leq E\left[\frac{X}{a}\right]$$

$$1 \times P(X \geq a) \leq \frac{E(X)}{a} \quad , \qquad P(X \geq a) \leq \frac{E(X)}{a}$$

۲–۱۵ نامساوی چبیشف

اگر متغیر تصادفی X دارای میانگین μ و واریانس σ^2 باشد آنگاه برای هر k>0

$$P[|X-\mu| \geq k] \leq \frac{\sigma^2}{k^2}$$

اثبات:

اگرk^2را برابر α و(X- μ) را X فرض کنیم شرایط مارکف تأمین می شود و

$$P[(X-\mu)^2 \geq k^2] \leq \frac{E(X-\mu)^2}{k^2} \longrightarrow P[|X-\mu|^2 \geq k^2] \leq \frac{\sigma^2}{k^2} \longrightarrow P[|X-\mu| \geq k] \leq \frac{\sigma^2}{k^2}$$

اهمیت نامساوی مارکف و چبیشف در این است که ما را قادر می سازد با معلوم بودن میانگین و واریانس جامعه، کرانهای بالا و پایین را برای مقادیر مختلف احتمال بدست آوریم، گرچه فرم تابع چگالی احتمال معلوم نیست.

مثال: فرض کنید تعداد محصولات تولید شده در یک کارخانه در طول هفته یک متغیر تصادفی با میانگین μ=۵۰ و واریانس σ۲=۲۵ باشد. مطلوبست:

الف- احتمال اینکه تولید محصول در یک هفته معین بیش از ۷۵ باشد.

ب- احتمال اینکه محصول یک هفته معین بین ۴۰ و ۶۰ باشد.

$$P[X \geq 75] \leq \frac{E(X)}{75}, \qquad P(X \geq 75) \leq \frac{50}{75} = \frac{2}{3}$$

$$P[40 \leq X \leq 60] = P[-10 \leq X - 50 \leq 10] = P[-10 \leq X - \mu \leq 10$$

$$P[|X - \mu| \leq 10] = 1 - P[|X - \mu| > 10]$$

$$P[|X - \mu| > 10] \leq \frac{E(X-\mu)^2}{100} = \frac{25}{100} = \frac{1}{4} \longrightarrow \left\{ \begin{array}{l} P[40 \leq X \leq 60] \geq 1 - \frac{1}{4} \\ \\ P[40 \leq X \leq 60] \geq 0.75 \end{array} \right.$$

بسم الله الرحمن الرحيم

فصل چهارم

توزیع های احتمال خاص

در این فصل مسائل زیر بررسی می شود:

- ۱– توابع احتمال خاص گسسته
- ۲– توابع چگالی احتمال خاص پیوسته

۱ – توابع احتمال خاص گسسته

در این بخش توابع احتمال یکنواخت، برنولی، دوجمله ای، دو جمله ای منفی، هندسی، فوق هندسی، پواسن، سری لگاریتمی و سری لگاریتمی مارکف با ارائه الگو معرفی می شود.

۱–۱ تابع احتمال یکنواخت

متغیر تصادفی X دارای تابع احتمال یکنواخت با پارامتر k است اگر تابع احتمال آن به صورت زیر باشد:

$$f(x) = P(X = x) = \frac{1}{k} \qquad x = 1,2,\ldots,k$$

الگو: جعبه ای شامل صفحه کلید با شماره های ۱ تا k است. اگر هم شانس بودن را برای همه شماره ها یکسان درنظر بگیریم و تعریف کنیم

Xشماره صفحه کلید خارج شده آنگاه X دارای تابع چگالی احتمال یکنواخت گسسته است

$$E(X) = \sum_{x=1}^{k} xf(x) = \sum_{x=1}^{k} x.\frac{1}{k} = \frac{1}{k}\sum_{x=1}^{k} x = \frac{k+1}{2}$$

$$E(X^2) = \sum_{x=1}^{k} x^2 f(x) = \frac{1}{k}\sum_{x=1}^{k} x^2 = \frac{1}{k}\frac{k(k+1)(2k+1)}{6} = \frac{(k+1)(2k+1)}{6}$$

$$V(X) = E(X^2) - (E(X))^2 = \frac{(k+1)(2k+1)}{6} - \frac{(k+1)^2}{4} = \frac{(k+1)(k-1)}{12}$$

۱–۲ تابع احتمال برنولی

متغیر تصادفی X دارای تابع احتمال برنولی با پارامتر (شانس) Pاست اگر تابع چگالی احتمال آن به صورت زیر باشد.

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

الگو: جعبه ای شامل صفحه کلیدهایی از نوع دست دوم و نو با نسبت های P-۱ و P است. یک صفحه کلید به تصادف از جعبه خارج کنیم و اگر متغیر X را به صورت زیر تعریف کنیم:

$$X = \begin{cases} 0 & \text{اگر صفحه کلید خارج شده دست دوم باشد} \\ 1 & \text{اگر صفحه کلید خارج شده نو باشد} \end{cases}$$

آنگاه X دارای تابع برنولی است.

۱–۳ تابع احتمال دو جمله ای

متغیر تصادفی X دارای تابع احتمال دوجمله ای با پارامترها n و p است. اگر تابع احتمال آن به صورت زیر باشد.

$$f(x) = p(X = x) = \binom{n}{x} p^{x}(1-p)^{n-x} \qquad x = 0,1,2,\dots,n$$

الگو: جعبه ای شامل صفحه کلیدهای از نوع دست دوم و نو با نسبت های p و p–۱ است. از این جعبه در شرایط یکسان و به تصادف n صفحه کلید یکی یکی و با جایگذاری خارج می کنیم و اگر تعریف کنیم

X : تعداد صفحه کلیدهای نو خارج شده آنگاه Xدارای توزیع دوجمله ای است.

۱–۳–۱ ویژگیهای توزیع دو جمله ای

۱– $\sum_{x=0}^{n} \binom{n}{x} p^{x}(1-p)^{n-x} = [p+(1-p)]^{n} = 1$

۲– دارای نمای منحصر به فرد است.

۳–برای n=1 تابع چگالی احتمال دو جمله ای همان تابع احتمال برنولی است.

۴– دارای میانگین np و واریانس $np(1-p)$ است.

۵– $f(x) = \frac{n(n-1)(n-2)...(n-x+1)}{x!}\left(\frac{p}{1-p}\right)^{x} f(0)$

۶– برای مقادیر n و p مقدار F(x) را می توان از جدول ضمیمه ۱ محاسبه کرد مختلف

۱–۴ تابع احتمال دو جمله ای منفی (پاسکال)

متغیر تصادفی X دارای تابع احتمال دو جمله ای منفی با پارامتر r و p است اگر تابع احتمال آن به صورت زیر باشد.

$$f(x) = p(X = x) = \binom{x+r-1}{r-1} p^r (1-p)^x \qquad x = 0,1,2,\dots$$

۱–۴–۱ ویژگیهای توزیع دو جمله ای منفی

1- $$\sum_{x=0}^{\infty} \binom{x+r-1}{r-1} p^r (1-p)^x = \sum_{x=0}^{\infty} \binom{x+r-1}{x} p^r (1-p)^x = p^r p^{-r} = 1$$

۲-دارای میانگین $\frac{r(1-p)}{p}$ و واریانس $\frac{r(1-p)}{p^2}$ است.

۱–۵ تابع احتمال هندسی

متغیر تصادفی X دارای تابع احتمال هندسی با پارامتر p است اگر تابع احتمال آن به صورت زیر باشد.

$$f(x) = p(X = x) = p(1-p)^x \qquad x = 0,1,2,\ldots$$

۱–۵–۱ ویژگیهای توزیع هندسی

۱- این توزیع فاقد حافظه است یعنی $p[X \geq s+t \mid X \geq t] = p[X \geq s]$

۲- دارای میانگین $\frac{1-p}{p}$ و واریانس $\frac{1-p}{p^2}$ است.

۳- $\sum_{x=0}^{\infty} p(1-p)^x = 1$

۱–۶ تابع احتمال فوق هندسی

متغیر تصادفی X دارای تابع احتمال فوق هندسی با پارامترهای N،kوn است اگر تابع احتمال آن به صورت زیر باشد.

$$f(x) = p[X = x] = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \qquad x = 0,1,2,\ldots,n$$

۱–۶–۱ ویژگیهای توزیع فوق هندسی

۱–

$$\sum_{x=0}^{n} \binom{k}{x}\binom{N-k}{n-x} = \binom{N}{n}$$

۲– Nیک عدد صحیح مثبت، k یک عدد صحیح نامنفی (k≤N)و n یک عدد نامنفی و حداکثر برابر با N است.

۳–دارای میانگین $\frac{nk}{N}$ و واریانس $\left(\frac{nk}{N}\right)\left(\frac{N-k}{N}\right)\left(\frac{N-n}{N-1}\right)$ است.

۱–۷ تابع احتمال پواسن

متغیر تصادفی X دارای تابع احتمال پواسن با پارامتر λ است اگر تابع احتمال آن به صورت زیر باشد.

$$f(x) = p(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \qquad x = 0,1,2,\ldots$$

۱–۷–۱ ویژگیهای توزیع پواسن

۱– $\displaystyle\sum_{x=0}^{\infty}\frac{e^{-\lambda}\lambda^{x}}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^{x}}{x!} = 1$

۲– دارای نمای منحصر به فرد است.

۳– دارای میانگین λ و واریانس λ است.

۴– برای مقادیر مختلف λ می توان از جدول ضمیمه (۲) مقدار F(x) را محاسبه کرد

۱–۸ تابع احتمال سری لگاریتمی

متغیر تصادفی X دارای تابع احتمال سری لگاریتمی با پارامتر α است اگر تابع احتمال آن به صورت زیر باشد.

$$f(x) = P(X = x) = \frac{-1}{\ln(1-\alpha)} \; \frac{\alpha^x}{x} \qquad x = 1,2,\ldots$$

۱–۹ تابع احتمال سری لگاریتمی مارکف

متغیر تصادفی X دارای تابع احتمال سری لگاریتمی مارکف با پارامترهای α و β است اگر تابع احتمال آن به صورت زیر باشد.

$$f(x) = P(X = x) = \frac{-1}{\ln(1-\alpha)} \cdot \frac{(1-\beta)^x - \left[1 - \frac{\beta}{1-\alpha}\right]^x}{x} \qquad x = 1,2,\ldots$$

۱–۹–۱ ویژگیهای سری لگاریتمی مارکف

۱- برای $\beta=1-\alpha$ توزیع سری لگاریتمی مارکف به توزیع سری لگاریتمی تبدیل می شود.

۲- دارای میانگین $\dfrac{-\alpha}{\beta \ln(1-\alpha)}$ است.

مثال: طول نوبت بارندگی دارای توزیع سری لگاریتمی مارکف با $\alpha=0/63$ و $\beta=0/3$ است مطلوبست:

الف- احتمال اینکه طول نوبت بارندگی برابر با ۱ باشد.

ب- احتمال اینکه طول نوبت بارندگی حداکثر ۲ باشد.

$$f(x)=P(X=x)=\frac{-1}{\ln(1-0.63)}\cdot\frac{(1-0.3)^x-\left[1-\dfrac{0.3}{1-0.63}\right]^x}{x} \longrightarrow f(1)=0.514$$

$$f(x\le 2)=f(1)+f(2)=0.514+0.228=0.742$$

۲ – توابع چگالی احتمال خاص پیوسته

در این بخش توابع چگالی احتمال یکنواخت، نرمال، نرمال استاندارد، نمایی، گاما، کی‌دو، بتا، استودنت و فیشر ارائه می شود.

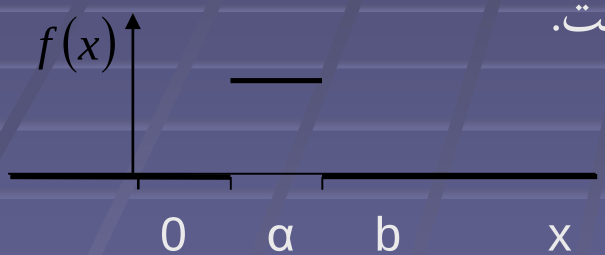
۲–۱ تابع چگالی احتمال یکنواخت (مستطیلی)

متغیر تصادفی X دارای تابع چگالی احتمال یکنواخت با پارامترهای α و b است اگر تابع چگالی آن به صورت زیر باشد.

$$f(x)=\frac{1}{b-a} \qquad a<x<b$$

۲–۱–۱ ویژگیهای تابع چگالی احتمال یکنواخت

۱- نمودار f(X) برای $-\infty<a<b<\infty$ به صورت زیر است.



۲- تابع توزیع F(X) برابر است با:

$$F(x)=P(X\le x)=\int_{-\infty}^{x} f(t)\,dt=\begin{cases} 0 & x<a \\ \dfrac{x-a}{b-a} & a\le x<b \\ 1 & x\ge b \end{cases}$$

۳- دارای میانگین $\frac{a+b}{2}$ و $\frac{(b-a)^2}{12}$ واریانس است.

۴- برای α=0 ، b=۱ تابع $f(x)$ را روی بازه (0،۱) گویند و آن را به صورت زیر تعریف می کنند.

$$f(u)=1 \qquad 0<u<1$$

$$f(u)=I_{(0,1)}^{(u)}$$

که $I_{(0,1)}^{(u)}$ را تابع نشانگر گویند.

ذکر این نکته ضروری است که تابع توزیع هر متغیر تصادفی همانند 0<u<1 و F(u) عمل می کند چون $F(+\infty)=1$ ، $F(-\infty)=0$ پس u=F(x) است. از این خاصیت در آمار برای شبیه سازی متغیرهای تصادفی استفاده می‌کنند.

۲–۲ تابع چگالی احتمال نرمال

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

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

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

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

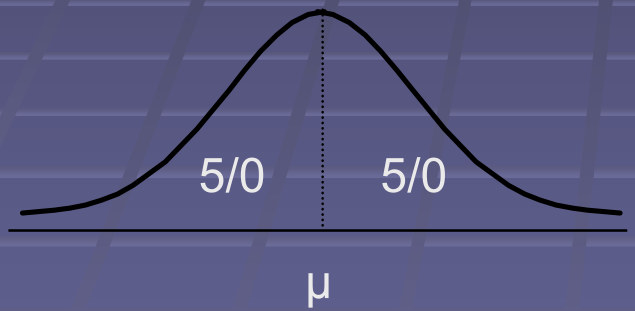
$$\sigma > 0$$

μ و σ2 پارامترهای توزیع نرمال هستند.

۲–۲–۱ ویژگیهای توزیع نرمال

۱- این توزیع نسبت به محور y=μ دارای تقارن است.

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



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

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

۲–۳ توزیع نرمال استاندارد

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

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

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

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

مقادیر مختلف F(x)را می توان با توجه به ویژگی Z از جدول ضمیمه (۳) بدست آورد که

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

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

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

۲–۴ تابع چگالی احتمال نمایی

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

$$f(x) = \frac{1}{\theta} e^{-x/\theta} \qquad x > 0 \ , \quad \theta > 0$$

توزیع نمایی کاربردهای مهمی دارد. از جمله در مدلهای صف بندی، می توان نشان داد که زمان انتظار مابین ورودی های متوالی از توزیع نمایی پیروی می کند

۲–۴–۱ ویژگیهای توزیع نمایی

۱–
$$F(x) = P(X < x) = \int_{-\infty}^{x} f(t)\,dt = 1 - e^{-x/\theta}$$

۲– فاقد حافظه است.
$$P[X > s+t \mid X > t] = P(X > s)$$

۳– دارای میانگین θ و واریانس θ^2 است.

۴– اگر u دارای توزیع یکنواخت روی (0,۱) باشد آنگاه –ln(u) دارای توزیع نمایی با θ=۱ است.

۲–۵ تابع چگالی احتمال گاما

متغیر تصادفی X دارای تابع چگالی احتمال گاما با پارامترهای α و β است اگر تابع چگالی احتمال آن به صورت زیر باشد.

$$f(x) = \frac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} \qquad x > 0 \ , \quad \alpha > 0 \ , \quad \beta > 0$$

حالت خاص: برای $\alpha=1$ ، $\beta=\theta$ توزیع گاما به توزیع نمایی تبدیل می شود. تابع چگالی احتمال گاما باتوجه به ویژگی تابع گاما تعریف می شود. چون:

$$\int_0^{\infty} f(x)dx = \int_0^{\infty} \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)} dx = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}\int_0^{\infty} (\beta u)^{\alpha-1} e^{-u} \beta \, du$$

$$= \frac{1}{\Gamma(\alpha)}\int_0^{\infty} e^{-u} u^{\alpha-1} du = 1$$

$$\Gamma(\alpha) = \int_0^{\infty} u^{\alpha-1} e^{-u} du$$

۲-۵-۱ ویژگیهای توزیع گاما

۱- $F(x)=P(X\le x)=\int_{-\infty}^{x}\frac{1}{\beta^{\alpha}\Gamma(\alpha)}x^{\alpha-1}e^{-x/\beta}dx=1-\sum_{i=0}^{\alpha-1}\frac{1}{i!}\left(\frac{x}{\beta}\right)^{i}e^{-x/\beta}$

۲- دارای میانگین $\alpha\beta$ و واریانس $\alpha\beta^{2}$ است.

مثال : در یک شهر مصرف برق روزانه دارای توزیع گاما با α=۳ و β=۲ است. اگر ظرفیت روزانه ۱۲ میلیون کیلووات ساعت باشد. احتمال اینکه برق موجود برای یک روز کافی باشد چقدر است؟

$$P(X\le 12)=F(12)=1-\sum_{i=0}^{3}\frac{1}{i!}\left(\frac{12}{2}\right)^{i}e^{-12/2}$$

$$=1-e^{-6}[1+6+18+36]=0.849$$

۲–۶ تابع چگالی احتمال کی‌دو

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

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

توزیع کی دو حالت خاص توزیع گاما است

$$\left(\alpha = \frac{r}{2} \ , \qquad \beta = 2\right)$$

۲–۶–۱ ویژگیهای توزیع کی دو

۱– r را درجه آزادی توزیع گویند.

۲– دارای میانگین r و واریانس 2r است.

۳–

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

۴– مقادیر مختلف F(x) را می توان برای مقادیر مختلف r از جدول ضمیمه (۵) بدست آورد.

۲–۷ تابع چگالی احتمال بتا

متغیر تصادفی X دارای تابع چگالی احتمال بتا با پارامترهای α و β است اگر تابع چگالی احتمال آن به صورت زیر باشد.

$$f(x) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\ x^{\alpha-1}\ (1-x)^{\beta-1} \qquad 0 < x < 1\ , \quad \alpha > 0\ , \quad \beta > 0$$

۲–۷–۱ ویژگیهای توزیع بتا

۱– $$\int_0^1 x^{\alpha-1}\ (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

۲– برای $\alpha=1$ و $\beta=1$ توزیع بتا به توزیع یکنواخت پیوسته تبدیل می شود.

۳– دارای میانگین $\dfrac{\alpha}{\alpha+\beta}$ و واریانس $\dfrac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)}$ است.

۲–۸ تابع چگالی احتمال استودنت (توزیع t)

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

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

که **r** را درجه آزادی توزیع **t** گویند.

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

۲–۸–۱ ویژگیهای توزیع t:

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

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

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

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

۲–۹ تابع چگالی احتمال فیشر

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

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

که r_1و r_2 به ترتیب درجه آزادی صورت و مخرج خوانده میشود برای مقادیر مختلف r_1و r_2 مقادیر مختلف F(x) از جدول ضمیمه (۶) قابل محاسبه است.

بسم الله الرحمن الرحيم

فصل پنجم

توزیع های نمونه گیری

در این فصل مسائل زیر بررسی می شود:

۱– برآوردگر
۲– توزیع مشترک
۳– توابع خطی از متغیرهای تصادفی مستقل
۴– توزیع میانگین
۵– قضیه حد مرکزی
۶– تقریب نرمال برای توزیع دو جمله ای
۷– توزیع واریانس نمونه
۸– توزیع t
۹– توزیع نسبت واریانس دو نمونه

۱- برآوردگر

هر تابعی از نمونه را که به پارامتر یا پارامترهای جامعه بستگی نداشته باشد برآوردگر یا آماره گویند. چون مقدار برآوردگر از نمونه ای به نمونه دیگر تغییر می‌کند متغیری است تصادفی. مقدار عددی آماره یا برآوردگر را برآورد گویند.

۱-۱ ویژگیهای برآوردگر کارا

۱- ناریب باشد.

۲- دارای کمترین واریانس باشد.

۲ - توزیع مشترک

فرض کنید متغیر تصادفی گسسته X دارای تابع احتمال $f(x)=p(X=x)$ باشد و $X_n,\ldots,X_2,X_1$ نمونه های تصادفی مستقل از هم باشند. اگر $x_n,\ldots,x_2,x_1$ مقادیر متناظر مشاهده برای $X_n,\ldots,X_2,X_1$ باشند. پیشامدهای $X_n=x_n,\ldots,X_2=x_2,X_1=x_1$ به طور مجزا از هم مستقل اند و احتمال توام آنها برابر است با:

$$P[X_1=x_1,X_2=x_2,\ldots,X_n=x_n]=P(X_1=x_1)P(X_2.x_2)\ldots P(X_n=x_n)$$

اگر از نماد $f(x_1,x_2,\ldots,x_n)$ به جای $P[X1=x1,X2=x2,\ldots,Xn=xn]$ استفاده کنیم:

$$f(x_1,x_2,\ldots,x_n)=f(x_1)\,f(x_2)\,\ldots\,f(x_n)=\prod_{i=1}^{n} f\left(x_i\right)$$

برای $n=۲$، $f(x_1,x_2)=f(x_1)\,f(x_2)$ تابع چگالی احتمال توام X_1 و X_2 می‌باشد. در حالتی که متغیر تصادفی X از نوع پیوسته است تابع چگالی مشترک یا توام را با $f(x_1,x_2,\ldots,x_n)$ نیز نمایش می دهند.

۳ - توابع خطی از متغیرهای تصادفی مستقل

فرض کنید $X_1,X_2,\ldots,X_n$ نمونه تصادفی مستقل با توزیع مشترک $f(x_1,x_2,\ldots,x_n)$ باشند. یک تابع خطی یا آماره را می توان در حالت کلی به صورت زیر تعریف کرد.

$$\sum_{i=1}^{n} a_i X_i = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n$$

که α_i ها مقادیر ثابت هستند.

اگر X_1 و X_2به ترتیب دارای میانگین های μ_1 و μ_2 و واریانسهای σ_1^2، σ_1^2 باشند. میانگین و واریانس Y به طریق زیر محاسبه میشود:

$$\mu_y = E(a_1X_1 + a_2X_2) = \sum_{x_1}\sum_{x_2}(a_1x_1 + a_2x_2)\, f_1(x_1)\, f_2(x_2)$$

$$= a_1\left[\sum_{x_1} x_1 f_1(x_1)\right]\left[\sum_{x_2} f_2(x_2)\right] + a_2\left[\sum_{x_1} f_1(x_1)\right]\left[\sum_{x_2} x_2 f_2(x_2)\right]$$

$$= a_1E(X_1) + a_2E(X_2) = a_1\mu_1 + a_2\mu_2$$

$$\sigma_y^2 = V(Y) = E\left[Y - \mu_y^2\right] = E\left[(a_1X_1 + a_2X_2) - a_1\mu_1 - a_2\mu_2\right]^2$$

$$= E\left[(a_1(X_1 - \mu_1) + a_2(X_2 - \mu_2)\right]^2$$

$$= E\left[(a_1^2(X_1 - \mu_1)^2 + a_2^2(X_2 - \mu_2)^2 + 2a_1a_2(X_1 - \mu_1)(X_2 - \mu_2)\right]$$

$$= a_1^2V(X_1) + a_2V(X_2) + 0$$

$$= a_1^2\sigma_1^2 + a_2^2\sigma_2^2$$

---

۴ - توزیع میانگین

در آمار توصیفی، میانگین نمونه تصادفی به صورت $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ تعریف شده بود. در این بخش، توزیع $\overline{X}$ را باتوجه به توزیع جامعه ای که نمونه از آن گرفته شده بدست می آوریم.

قضیه ۴-۱ اگر $X_1, X_2, \ldots, X_n$ نمونه های مستقل و هم توزیع از جامعه ای با میانگین μ و واریانس σ^2 باشند آنگاه میانگین نمونه $\overline{X}$ دارای میانگین μ و واریانس $\frac{\sigma^2}{n}$ است.

برهان: چون $\overline{X}$ یک ترکیب خطی از X_i هاست، پس:

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

$$V(\overline{X}) = V\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

$\overline{X}$ تابعی از X_i هاست و به پارامترهای جامعه μ و σ^2 بستگی ندارد. یک آماره نااریب و دارای کمترین واریانس است.

قضیه ۴-۲ اگر $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی n تایی از جامعه نرمال با میانگین μ و واریانس σ^2 باشند آنگاه $\overline{X}$ دارای توزیع نرمال با میانگین μ و واریانس $\frac{\sigma^2}{n}$ است.

برهان: $X \sim N(\mu, \sigma^2)$ است. تابع مولد گشتاورهای آن برابر است با:

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

$$M_{\overline{X}}(t) = E\left[e^{t\overline{X}}\right] = E\left[e^{t.\frac{1}{n}\sum X_i}\right] = E\left[e^{\frac{t}{n}(X_1 + X_2 + \ldots + X_n)}\right] = E\left[e^{\frac{tX_1}{n} + \frac{tX_2}{n} + \ldots + \frac{tX_n}{n}}\right]$$

$$= \left(E\left[e^{\frac{tX}{n}}\right]\right)^n$$

چون X_i ها مستقل و هم توزیع اند.

$$=\left[M_x\left(\frac{t}{n}\right)\right]^n=\left[e^{\mu\frac{t}{n}+\frac{1}{2}\frac{\sigma^2t^2}{n^2}}\right]^n=e^{\mu t+\frac{\sigma^2t^2}{2n}}$$

$e^{\mu t+\frac{\sigma^2t^2}{2n}}$ تابع مولد گشتاورهای متغیر تصادفی نرمال با میانگین و واریانس است.

از اینکه یک ترکیب خطی از ها و ها از هم مستقل‌اند، امید ریاضی و واریانس مستقیماً به صورت زیر نیز محاسبه می شود.

لم ۴–۳ اگر شرایط قضیه ۴-۲ برقرار باشد متغیر $Z=\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ دارای توزیع نرمال استاندارد است.

$$E(Z)=E\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)=\frac{\sqrt{n}}{\sigma}E(\overline{X}-\mu)=\frac{\sqrt{n}}{\sigma}(E\overline{X}-\mu)=0$$

$$V(\overline{X})=V\left(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)=\frac{n}{\sigma^2}V(\overline{X}-\mu)=\frac{n}{\sigma^2}.\frac{\sigma^2}{n}=1$$

۵ – قضیه حد مرکزی

اگر $\overline{X}$ میانگین نمونه تصادفی $X_1, X_2, \ldots, X_n$ از توزیعی (جامعه ای) با میانگین μ و واریانس متناهی $\sigma^2 < \infty$ باشند آنگاه توزیع متغیر تصادفی $Z = \dfrac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ میل می کند به توزیع نرمال استاندارد اگر $n \to \infty$

این قضیه با استفاده از تابع مولد گشتاورها به راحتی اثبات می شود.

۶- تقریب نرمال برای توزیع دوجمله ای

در توزیع دوجمله ای با پارامترهای n و p برای nهای بزرگ محاسبه احتمال گاهی اوقات با استفاده از جدول ضمیمه (۱) خسته کننده و گاهی ممکن است جدولی با چنین nای در دسترس نباشد.

می دانیم اگر Y دارای توزیع دوجمله ای باشد، می توان Yرا به صورت جمعی از متغیرهای برنولی یعنی $Y=\sum_{i=1}^{n}X_i$ نوشت که X_iها متغیرهای برنولی با میانگین p و واریانس p(1-p)- می باشند و مقادیری که Yاختیار می کند اعداد صحیح 0، 1، 2، ...،n است.

متغیر Y را که از نوع گسسته است می توان با توجه به نتیجه قضیه حد مرکزی به وسیله متغیر نرمال استاندارد تقریب زد. احتمال پیشامد Y=k را می توان به صورت زیر تقریب زد.

$$p[Y=k]=p\left[k-\frac{1}{2}<Y<k+\frac{1}{2}\right]\approx\int_{k-\frac{1}{2}}^{k+\frac{1}{2}}f(y)\,dy$$

$$P\left[Y=k\right]\approx P\left[\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}}<\frac{Y-np}{\sqrt{np(1-p)}}<\frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right]$$

$$P[Y=k] \approx P\left[\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}} < Z < \frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right]$$

$$=\phi\left[\frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right]-\phi\left[\frac{k-\frac{1}{2}-np}{\sqrt{np(1-p)}}\right]$$

که تابع $\phi(t)$ برابر است با:

$$\phi(t)=p[Z<t]=\int_{-\infty}^{t}\frac{1}{\sqrt{2\pi}}\,e^{-\frac{1}{2}z^2}\,dz$$

۷ – توزیع واریانس نمونه

واریانس نمونه nتایی در آمار توصیفی به صورت $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ تعریف شده

بود. اکنون برای ناریب بودن، آن را به صورت $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ تعریف می‌کنیم.

قضیه ۷–۱ اگر متغیر Z دارای توزیع نرمال استاندارد باشد آنگاه Z^2 دارای توزیع کی‌دو با یک درجه آزادی است.

برهان: با استفاده از تابع مولد گشتاورها

$$M_z(t) = E[e^{tz}] = e^{\frac{1}{2}t^2}$$

برای متغیر Z

برای متغیر Z^2

$$M_{z^2}(t) = E[e^{tz^2}] = \int_{-\infty}^{\infty} e^{tz^2}\frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}z^2}dz = \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\ e^{-\frac{1}{2}[1-2t]z^2}dz$$

با فرض $u = \sqrt{1-2t}\, z$

$$M_{z^2}(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \frac{du}{\sqrt{1-2t}} = (1-2t)^{-\frac{1}{2}}$$

قضیه ۷–۲ اگر متغیرهای مستقل $Z_1, Z_2, \dots, Z_n$ دارای توزیع نرمال استاندارد باشند

آنگاه $\sum_{i=1}^{n} Z_i^2$ دارای توزیع کی‌دو با n درجه آزادی است.

اثبات این قضیه با استفاده از تابع مولد گشتاورها آسان است که در اینجا بدون اثبات می پذیریم. از این قضیه استنتاج می شود که اگر دو متغیر مستقل دارای توزیع کی‌دو باشند جمع آنها نیز توزیع کی‌دو است. در مورد تفاضل هم در شرایط خاص درست است. یعنی اگر دو متغیر مستقل دارای توزیع کی‌دو باشند تفاضل آنها نیز دارای توزیع کی‌دو است با تفاضل درجه آزادی دو متغیر.

قضیه ۷-۳ اگر $\overline{X}$ و S^2 به ترتیب میانگین و واریانس نمونه $X_1, X_2, \ldots, X_n$ از جامعه نرمال با میانگین μ و واریانس σ^2 باشد آنگاه

الف- $\overline{X}$ و S^2 از هم مستقل اند.

ب- متغیر $\frac{(n-1)S^2}{\sigma^2}$ دارای توزیع کی‌دو با $n-1$ درجه آزادی است.

۸- توزیع t

فرض کنید که $X_1, X_2, \ldots, X_n$ یک نمونه nتایی از توزیع نرمال با میانگین μ و واریانس σ^2 باشد. می دانیم متغیر $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ دارای توزیع نرمال استاندارد و متغیر $\frac{(n-1)S^2}{\sigma^2}$ دارای توزیع کی‌دو با $n-1$ درجه آزادی است. متغیر T را که تابعی از دو متغیر است به صورت زیر تعریف می کنیم.

$$T = \frac{\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\frac{S}{\sigma}} = \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$$

۹– توزیع نسبت واریانس دو نمونه

فرض کنید از جامعه اول که دارای توزیع نرمال با واریانس σ_1^2 است نمونه $X_1، X_2، ....، X_{n_1}$ به حجم n_1 و از جامعه دوم که دارای توزیع نرمال با واریانس σ_2^2 است نمونه $Y_1، Y_2، ....، Y_{n_2}$ به حجم n_2 موجود باشند. اگر $S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \overline{X})^2$ و $S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \overline{Y})^2$ به ترتیب واریانس های نمونه اول و دوم باشند متغیرهای $\frac{(n_1 - 1)S_1^2}{\sigma_1^2}$ و $\frac{(n_2 - 1)S_2^2}{\sigma_2^2}$ طبق قضیه 5-7-4 دارای توزیع کی‌دو با درجه آزادی n_1-1 و n_2-2 هستند.

قضیه **1-9** اگر متغیرهای تصادفی مستقل X و Y به ترتیب دارای توزیع کی‌دو با m و n درجه آزادی باشند آنگاه متغیر $F = \dfrac{X/m}{Y/n}$ دارای توزیع فیشر با m و n درجه آزادی است. معمولاً m را درجه آزادی صورت و n را درجه آزادی مخرج می‌گویند و آن را با علامت $F(m,n)$ نمایش می‌دهند.

برای بدست آوردن توزیع نسبت واریانس دو نمونه، متغیر F را به صورت زیر تعریف می‌کنیم.

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

متغیر F شرایط قضیه 5-9-1 را دارا می‌باشد و دارای توزیع فیشر با n_1-1 و n_2-1 درجه آزادی می‌باشد.

تبصره 5–9–2 اگر متغیر تصادفی F دارای توزیع فیشر با m و n درجه آزادی باشد آنگاه متغیر $\frac{1}{F}$ دارای توزیع فیشر با n و m درجه آزادی است. یعنی:

$$F_{(\alpha,m,n)} = \frac{1}{F_{(1-\alpha,n,m)}}$$

این نتیجه به ما اجازه می‌دهد توزیع F را تنها برای دنباله سمت راست در جدول ضمیمه (6) درج کنیم.

مثال 5–9–3 با توجه به جدول ضمیمه (6) مقدار C را برای احتمالات زیر محاسبه کنید.

$$P[F_{(9,10)} > C] = 0.05 \Rightarrow C = 3.02$$

$$P[F_{(15,15)} > C] = 0.01 \Rightarrow C = 3.52$$

$$P[F_{(8,10)} < C] = 0.95 \Rightarrow P[F_{(8,10)} > C] = 0.05 \Rightarrow C = 3.07$$

پایان فصل ۵

فصل ۶

برآورد نقطه ای و فاصله ای پارامتر

1-6 برآورد نقطه‌ای

فرض کنید متغیر تصادفی X دارای تابع چگالی احتمال $f_\theta(x)$ باشد به طوری که $\theta \in \Theta$. اگر x_1، x_2،....، x_n مشاهدات نمونه باشند برآورد نقطه‌ای برای θ ، مقدار عددی برآوردگر بر اساس نمونه مشاهده شده خواهد بود. اگر فرض کنیم $\bar{x}$ مقدار عددی برآوردگر $\bar{X}$ برای θ باشد احتمال اینکه $\bar{x}$ دقیقاً برابر با θ باشد صفر است. به همین خاطر از $\bar{x}$ به عنوان یک برآورد برای θ یاد می‌کنند.

برای اینکه فرقی بین θ و برآورد آن قایل شویم، برآورد θ را به $\hat{\theta}$ نمایش می‌دهیم. اختلاف $|\hat{\theta}-\theta|$ را خطای برآورد نقطه‌ای گویند. در ادامه این بخش روش برآورد گشتاورها و روش درستنمایی ماکزیمم ارائه می‌شود.

2-6 روش گشتاورها

برآورد پارامتر به روش گشتاورها یکی از روشهای قدیمی است که در سال 1894 توسط کارل پیرسون پیشنهاد شد و اکنون هم در برآورد پارامتر بیشتر توزیعها قابل استفاده است.

فرض کنید متغیر تصادفی X دارای تابع چگالی احتمال $f_\theta(x)$ و $x_1, x_2, \ldots, x_n$ مشاهدات نمونه تصادفی $X_1, X_2, \ldots, X_n$ باشند. برآورد پارامتر θ به روش گشتاورها از برابری گشتاورهای نمونه و جامعه به صورت زیر بدست می‌آید.

گشتاور مرتبه r ام جامعه: $\mu_r = E(X^r)$

گشتاور غیرمرکزی مرتبه r ام نمونه حول نقطه صفر: $m'_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r$

3-6 روش درستنمایی ماکزیمم

روش درستنمایی ماکزیمم در سال 1912 توسط فیشر ارائه شد و در مواردی کـه بـرآورد به روش گشتاورها و روش درستنمایی ماکزیمم یکسان نیستند، برآورد به روش درستنمایی ماکزیمم را به روش گشتاورها ترجیح می‌دهند.

اگر x_1،x_2،....،x_n مشاهدات متناظر نمونـه تـصادفی X_1،X_2،....،X_n از تـابع چگـالی احتمال $f_\theta(x)$، $\theta \in \Theta$ باشند تابع چگالی توام x_iها برابر است با:

$$f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_\theta(x_i)$$

در رابطه اخیر تنها متغیر θ است و می توان رابطه را فقط تابعی از θ نوشت.

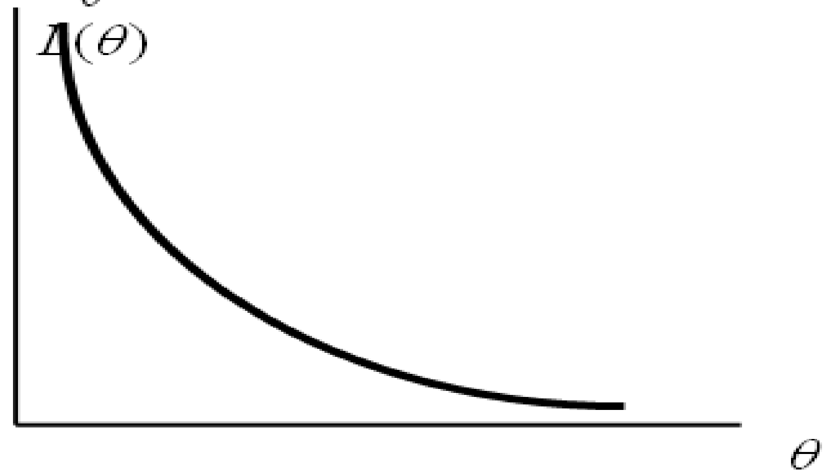
$$L(\theta)=\prod_{i=1}^{n} f_{\theta}(x_i) \quad , \qquad \theta \in \Theta$$

در ادبیات آماری $L(\theta)$ را تابع درستنمایی گویند. هدف روش درستنمایی ماکزیمم این است که θ را طوری پیدا کند که $L(\theta)$ ماکزیمم شود. از آنجا که $L(\theta)$ و $\ln L(\theta)$ در مقدار یکسانی از θ ماکزیمم می‌شود برای راحتی θ را طوری پیدا می‌کنیم که $\ln L(\theta)$ را ماکزیمم کند.

مثال 6–3–4 فرض کنید X دارای تابع چگالی احتمال یکنواخت روی بازه $(0,\theta)$ باشد θ را براساس یک نمونه n تایی برآورد کنید.

$$f_\theta(x)=\frac{1}{\theta} \quad , \quad 0<x<\theta \quad , \quad \Theta=(0,\infty)$$

$$L(\theta)=\prod_{i=1}^{n} f_\theta(x_i)=\frac{1}{\theta^n}$$



باتوجه به شیب $L(\theta)$ ، $L(\theta)$ در هیچ جا صفر نمی شود لذا احتیاجی به مشتق گرفتن از $L(\theta)$ و مساوی قراردادن آن با صفر نیست.

4–6 برآورد فاصله‌ای

در بخش برآورد نقطه‌ای اشاره به این حقیقت شد که برآورد نقطه‌ای نمی‌تواند برابر با مقدار واقعی پارامتر باشد. به عنوان مثال فرض کنید مدیر کارخانه‌ای ادعا مـی‌کنـد کـه لامپهـای تولیدی این کارخانه دارای عمر متوسط بین 1500 ± 10 ساعت است. اگر براسـاس یـک نمونه n تایی برآورد نقطه‌ای برای ادعای مدیر داشته باشیم مسلماً برآورد ما یـک نقطـه از بازه $(1500+10$ و $1500-10)$ خواهد بود که به خودی خود متضمن اطلاعـاتی دربـاره میزان احتمال اینکه برآوردگر مقداری نزدیک به مقدار واقعی مجهول قبول کند، نیـست.

تعریـف 6–4–1 اگـر X_1، X_2،....، X_n یـک نمونـه تـصادفی از $f_\theta(x)$ باشـد و $T_1(X_1,\ldots,X_n)$ و $T_2(X_1,\ldots,X_n)$ دو آماره باشند به طـوری کـه $T_1<T_2$ بـازه (T_1,T_2) را یک بازه اطمینـان با ضریب اطمینان $100(1-\alpha)\%$ بـرای θ گـوییم. اگـر احتمال اینکه دو آماره T_1، T_2، θ در برداشته باشند مستقل از θ باشد.

$$P\left[T_1(X_1,\ldots,X_n) < \theta < T_2(X_1,\ldots,X_n)\right] = 1-\alpha$$

T_1 و T_2 را به ترتیب حدود اطمینان پایینی و بالایی θ و $T_2 - T_1$ را طول بازه اطمینان گویند.

برای بدست آوردن بازه اطمینان برای θ یا تابعی از θ، $\tau(\theta)$ لازم است کمیت محوری تعریف کنیم.

تعریف 6–4–2 هر تابعی از نمونه و پارامتر را که توزیع آن مستقل از پارامتر باشد کمیت محوری گویند. به عنوان مثال $\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ برای نمونه n تایی از جامعه نرمال با میانگین μ و واریانس معلوم σ^2 یک کمیت محوری است. چون $\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ دارای توزیع نرمال استاندارد

است. از آنجا که برای یافتن فاصله اطمینان نیاز بـه داشـتن آمـاره اسـت، دانـستن توزیـع جامعه‌ای که نمونه ازآن استخراج شده است ضروری است. فرض کنید متغیـر تـصادفی Z دارای توزیع نرمال استاندارد باشد. احتمال اینکه دو عدد $1/96$ و $-1/96$ متغیر تـصادفی Z را در برداشته باشند برابر با $0/95$ است.

$$P[-1/96 < Z < 1/96] = 0/95$$

عکس آن نیز درست است. اگر دو عدد a و b متغیر تصادفی Z را با احتمال $0/95$ در برداشته باشند a و b به ترتیب برابرند با $-1/96$ و $1/96$. در حالت کلی بـرای متغیـر تصادفی Z رابطه زیر همواره برقرار است.

$$P\left[-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right] = (1-\alpha)$$

که برای $\alpha = 0/05$ رابطه اخیر برابر است با:

$$P[-1/96 < Z < 1/96] = 0/95$$

5–6 فاصله اطمینان برای میانگین توزیع نرمال

حالت اول: اگرمتغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریـانس معلـوم σ^2 باشد براساس یک نمونه nتایی یک فاصله اطمینان، بـا ضـریب اطمینـان (α -1)100٪ برای μ به صورت زیر بدست می‌آید.

در این حالت می‌دانیم که متغیر $\dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ یک کمیت محوری و دارای توزیع نرمال استاندارد است. لذا:

$$P=\left[-z_{\frac{\alpha}{2}}<\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}<z_{\frac{\alpha}{2}}\right]=1-\alpha$$

یعنی احتمال رخداد پیشامد $-z_{\frac{\alpha}{2}} < \dfrac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_{\frac{\alpha}{2}}$ برابر با $(1-\alpha)$ است. اگر پیشامد فـوق را نسبت به μ حل کنیم داریم:

$$\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

برای یک نمونه n تایی بازه اطمینان برای μ برابر است با:

$$\overline{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

برای $\alpha = 0/05$

$$\overline{X} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96 \frac{\sigma}{\sqrt{n}}$$

6-6 فاصله اطمینان برای P در توزیع دو جمله‌ای

اگر متغیر تصادفی X دارای توزیع دو جمله‌ای با پارامتر مجهول P و n معلوم باشد $E(X)=nP$ و $V(X)=nP(1-P)$ است.

از رابطه $E(X)=nP$ داریم:

$$E\left(\frac{X}{n}\right)=P \quad , \qquad V\left(\frac{X}{n}\right)=\frac{P(1-P)}{n}$$

متغیر تصادفی X مجموع چند متغیر برنولی است، $\frac{X}{n}$ یک برآوردگر برای P جامعه است لذا:

$$\hat{P}=\frac{x}{n} \quad , \qquad E(\hat{P})=P \qquad , \qquad V(\hat{P})=\frac{P(1-P)}{n}$$

باتوجه به روابط اخیر اگر حجم نمونه از حدی بزرگتر باشد کمیت محوری زیر دارای توزیع تقریبی نرمال استاندارد خواهد بود.

$$z = \frac{\frac{X}{n} - P}{\sqrt{\frac{P(1-P)}{n}}}$$

$$P\left[-z_{\frac{\alpha}{2}} < \frac{\frac{X}{n} - P}{\sqrt{\frac{P(1-P)}{n}}} < z_{\frac{\alpha}{2}}\right] \approx 1-\alpha$$

از رابطه اخیر:

$$\frac{X}{n} - z_{\frac{\alpha}{2}}\sqrt{\frac{P(1-P)}{n}} < P < \frac{X}{n} + z_{\frac{\alpha}{2}}\sqrt{\frac{P(1-P)}{n}}$$

اگر x مقدار مشاهده X باشد فاصله اطمینان اخیر قابل ارزیابی نیست چون واریانس $\frac{X}{n}$ شامل پارامتر P است. چون $\hat{P}=\frac{x}{n}$ یک برآورد نقطه‌ای برای P است. برآورد نقطه‌ای واریانس $\frac{X}{P}$ برابر است با:

$$\frac{\hat{P}(1-\hat{P})}{n}=\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}$$

فاصله اطمینان تقریبی برای P برابر است با:

$$\frac{x}{n}-z_{\frac{\alpha}{2}}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}<P<\frac{x}{n}+z_{\frac{\alpha}{2}}\sqrt{\frac{\frac{x}{n}\left(1-\frac{x}{n}\right)}{n}}$$

7-6 فاصله اطمینان برای تفاضل میانگین دو جامعه

فرض کنید X_1، X_2، ....، X_n یک نمونه m تایی از جامعه ای نرمال با میانگین μ_1 و واریانس معلوم σ_1^2 و Y_1، Y_2، ....، Y_n یک نمونه n تایی از جامعه دیگر نرمال با میانگین μ_2 و واریانس معلوم σ_2^2 باشند. می دانیم که $\overline{X}$ و $\overline{Y}$ به ترتیب برآوردگرهای μ_1 و μ_2 می باشند. لذا هرگونه استنباط روی μ_1 و μ_2 براساس $\overline{X}$ و $\overline{Y}$ خواهد بود. در این بخش هدف، برآورد فاصله اطمینان برای $\mu_2 - \mu_1$ است. باتوجه به خواص برآوردگرها، $\overline{Y} - \overline{X}$ یک برآوردگر برای $\mu_2 - \mu_1$ است. چون $\overline{X}$ و $\overline{Y}$ تک تک دارای توزیع نرمال هستند تفاضل آنها نیز نرمال است که میانگین و واریانس آن به صورت زیر محاسبه می‌شود.

$$E(\overline{Y} - \overline{X}) = E(\overline{Y}) - E(\overline{X}) = \mu_2 - \mu_1$$

$$v(\overline{Y} - \overline{X}) = v(\overline{Y}) + v(\overline{X}) = \frac{\sigma_2^2}{n} + \frac{\sigma_1^2}{m}$$

پس متغیر $\bar{Y}-\bar{X}$ دارای نرمال با میانگین $\mu_2-\mu_1$ و واریـانس $\frac{\sigma_2^2}{n}+\frac{\sigma_1^2}{m}$ اسـت. اگـر متغیر Z متغیر استاندارد شده $\bar{Y}-\bar{X}$ باشد، Z برابر خواهد بود با:

$$Z=\frac{\bar{Y}-\bar{X}-(\mu_2-\mu_1)}{\sqrt{\frac{\sigma_2^2}{n}+\frac{\sigma_1^2}{m}}}$$

چون Z نرمال استاندارد است لذا داریم:

$$P\left[-z_{\frac{\alpha}{2}}<\frac{\bar{Y}-\bar{X}-(\mu_2-\mu_1)}{\sqrt{\frac{\sigma_2^2}{n}+\frac{\sigma_1^2}{m}}}<z_{\frac{\alpha}{2}}\right]=1-\alpha$$

از رابطه اخیر داریم:

$$(\bar{Y}-\bar{X})-z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}<\mu_2-\mu_1<(\bar{Y}-\bar{X})+z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}$$

این رابطه، یک فاصله اطمینان $100(1-\alpha)\%$ برای $\mu_2-\mu_1$ است. اگـر دو جامعـه دارای واریانس مشترک σ^2 و حجم نمونه ها مساوی n باشند فاصله اطمینان بـرای $\mu_2-\mu_1$ برابر است با:

$$(\bar{Y}-\bar{X})-z_{\frac{\alpha}{2}}\sqrt{\frac{2\sigma^2}{n}}<\mu_2-\mu_1<(\bar{Y}-\bar{X})+z_{\frac{\alpha}{2}}\sqrt{\frac{2\sigma^2}{n}}$$

8–6 فاصله اطمینان برای واریانس جامعه

فرض کنید که متغیر تصادفی X دارای توزیع نرمال با میـانگین مجهـول μ و واریـانس مجهول σ^2 باشد. اگر X_1، X_2،....، X_n یک نمونه از این جامعه باشد طبـق قـضیه 5-4-7، $\frac{(n-1)S^2}{\sigma^2}$ دارای توزیع کی دو با $n-1$ درجه آزادی است. بـرای یـافتن فاصـله اطمینان می‌توان از $\frac{(n-1)S^2}{\sigma^2}$ به عنوان کمیت محوری استفاده کرد.

$$P\left[\chi^2_{[1-\frac{\alpha}{2},\ n-1]} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{[\frac{\alpha}{2},\ n-1]}\right] = 1-\alpha$$

که اعداد $\chi^2_{[\frac{\alpha}{2},\ n-1]}$ و $\chi^2_{[1-\frac{\alpha}{2},\ n-1]}$ از جدول کی دو قابل محاسبه است. از پیشامد احتمال اخیر می توان σ^2 را برحسب S^2 به صورت زیر بدست آورد.

$$\frac{(n-1)S^2}{\chi^2_{[\frac{\alpha}{2},\ n-1]}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{[1-\frac{\alpha}{2},\ n-1]}}$$

رابطه اخیریک فاصله اطمینان(α -1)100٪ برای σ^2 است.فاصله اطمینان(α -1)100٪ برای σ یا انحراف معیار برابر خواهد بود با:

$$\sqrt{\frac{(n-1)S^2}{\chi^2_{[\frac{\alpha}{2},\ n-1]}}} \le \sigma \le \sqrt{\frac{(n-1)S^2}{\chi^2_{[1-\frac{\alpha}{2},\ n-1]}}}$$

9-6 فاصله اطمینان برای نسبت دو واریانس

اگر X_1، X_2،...، X_m یک نمونه تصادفی از جامعه نرمال با میانگین μ_1 و واریانس σ_1^2 و Y_1، Y_2،...، Y_n یک نمونه تصادفی از جامعه نرمال دیگر با میانگین μ_2 و واریانس σ_2^2 باشند در بخش 5-9 فصل پنجم نشان دادیم که متغیر $F=\left(\frac{\sigma_2}{\sigma_1}\right)^2\frac{S_1^2}{S_2^2}$ دارای توزیع فیشر با $m-1$ و $n-1$ درجه آزادی است. متغیر F یک کمیت محوری است چون توزیع آن مستقل از σ_1^2 و σ_2^2 است لذا،

$$P\left[f_{(1-\frac{\alpha}{2},\, m-1,\, n-1)} < F < f_{(\frac{\alpha}{2},\, m-1,\, n-1)}\right]=1-\alpha$$

$$P\left[f_{(1-\frac{\alpha}{2},\, m-1,\, n-1)} < \left(\frac{\sigma_2}{\sigma_1}\right)^2\frac{S_1^2}{S_2^2} < f_{(\frac{\alpha}{2},\, m-1,\, n-1)}\right]=1-\alpha$$

که $f_{(\frac{\alpha}{2}, m-1, n-1)}$ و $f_{(1-\frac{\alpha}{2}, m-1, n-1)}$ از جدول توزیع فیشر بدست می‌آیند. اگر $S_1 = \frac{1}{m-1}\sum (x_i - \bar{x})^2$ و $S_2 = \frac{1}{n-1}\sum (y_i - \bar{y})^2$ به ترتیب واریانس نمونه اول و دوم باشند فاصله اطمینان برای $\frac{\sigma_1^2}{\sigma_2^2}$ از رابطه زیر بدست می‌آید.

$$\frac{1}{f_{(\frac{\alpha}{2},m-1,n-1)}} \; \frac{S_1^2}{S_2^2} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{1}{f_{(1-\frac{\alpha}{2},m-1,n-1)}} \frac{S_1^2}{S_2^2}$$

مثال 6-9-1 نمرات زیر نمونه‌ای از نمرات برنامه نویس در دو گروه 1 و 2 می‌باشد. اگر فرض نرمال بودن نمرات در دو گروه پذیرفته شود یک فاصله اطمینان 90% برای نسبت واریانس دو جامعه بدست آورید.

گروه اول	12	10	14	13	11		
گروه دوم	17	15	14	16	17	17	16

$m = 5 \quad , \qquad n = 7$

$$S_1^2 = \frac{1}{m-1}\sum (x_i - \bar{x})^2 = 2.5$$

$$S_2^2 = \frac{1}{n-1}\sum (y_i - \bar{y})^2 = 1.33$$

$$\alpha = 0.1 \quad , \quad \frac{\alpha}{2} = 0.05 \quad , \quad 1 - \frac{\alpha}{2} = 0.95$$

$$f_{(\frac{\alpha}{2},\, m-1,\, n-1)} = f_{(0.05,4,6)} = 4.53$$

$$f_{(1-\frac{\alpha}{2},\, m-1,\, n-1)} = f_{(0.95,4,6)} = \frac{1}{f_{(0.95,6,4)}} = \frac{1}{6.16} = 0.162$$

$$\frac{1}{f_{(\frac{\alpha}{2},m-1,n-1)}}\frac{S_1^2}{S_2^2} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{1}{f_{(1-\frac{\alpha}{2},m-1,n-1)}}\frac{S_1^2}{S_2^2}$$

$$\frac{1}{4.53}\frac{2.5}{1.33} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{1}{0.162}\frac{2.5}{1.33}$$

$$0.415 \le \frac{\sigma_1^2}{\sigma_2^2} \le 11.603$$

پایان فصل ۶

آزمون فرض های آماری

فصل ۷

7–1 مفاهیم اولیه

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

تعریف 7–1–1 هر فرض آماری که توزیع جامعه را کاملاً مـشخص کنـد فـرض سـاده و فرضی که توزیع جامعه را مشخص نکند فرض مرکب گویند.

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

2–7 آزمون فرض برای میانگین توزیع نرمال

فرض کنید X_1، X_2،....، X_n یک نمونه nتایی از توزیع نرمال با میانگین مجهول μ و واریانس معلوم σ^2 باشند. هدف آزمون فرض زیر در مورد میانگین جامعه است.

$H_0 : \mu = \mu_0$ در مقابل $H_1 : \mu > \mu_0$

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

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

رد $H_\circ$ هم ارز است با : $\bar{X} >$ عدد

رد $H_\circ$ هم ارز است با : $\bar{X} - \mu_\circ >$ عدد

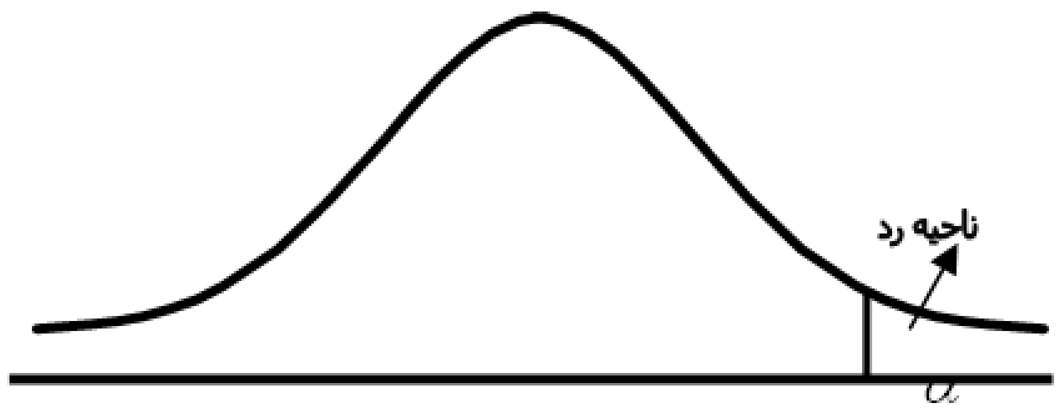
رد $H_\circ$ هم ارز است با: $\dfrac{\bar{X} - \mu_\circ}{\frac{\sigma}{\sqrt{n}}} >$ عدد

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

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

$$\alpha = P\left[\frac{\bar{X} - \mu_\circ}{\frac{\sigma}{\sqrt{n}}} > Z_\alpha \,\middle|\, H_\circ \text{ درست}\right]$$

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



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

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

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

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

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

3-7 آزمون فرضهای دو طرفه با استفاده از فاصله اطمینان

در مواقع آزمون فرض ساده در مقابل فرض مرکب دو طرفه می‌توان با استفاده از فاصله اطمینان بدست آمده برای پارامتر مفروض جامعه نسبت به قبول یا رد فرض $H_\circ$ اقدام کرد.

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

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

با استفاده از دو رابطه اخیر احتمال اینکه متغیر T تحت $H_{\circ}$ در ناحیه قبول قرار گیرد برابر است با

$$P\left[-t_{\alpha/2} < T < t_{\alpha/2}\right] = 1-\alpha$$

اما

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

پس

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

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

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

رابطه اخیر ناحیه اطمینان دو طرفه برای $\mu_\circ$ تحت فرض $H_\circ$ است. برای آزمون فرض $H_\circ : \mu = \mu_\circ$ در مقابل $H_1 : \mu \neq \mu_\circ$ با استفاده از فاصله اطمینان، ابتدا فاصله اطمینان را برای $\mu_\circ$ بدست می‌آوریم. اگر $\mu = \mu_\circ$ تحت فرض $H_\circ$ در این فاصله قرار گرفت فرض $H_\circ$ پذیرفته می‌شود در غیراینصورت رد می‌شود.

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

یکی از روشهای آزمون فرض آماری استفاده از P –مقدار یا مقدار احتمال است. اکنون که نرم افزارهای آماری ملاک آزمون فرض آماری را برحسب P –مقدار ارائه می‌کنند، لازم دیدیم که این بحث را در بخش جداگانه مورد بحث قرار دهیم.

فرض کنید $\overline{X}$ میانگین یک نمونه تصادفی از توزیع نرمال با میانگین μ، واریانس معلوم σ^2 باشد و بخواهیم فرض $H_0 : \mu = \mu_0$ را در مقابل فرض $H_1 : \mu < \mu_0$ آزمون کنیم.

در این آزمون مقادیر کوچک $\overline{X}$ باعث رد فرض H_0 می‌شود. اگر α سطح معنی‌دار آزمون باشد فرض H_0 رد می‌شود اگر احتمال پیشامد $\overline{X} \le \bar{x}$ تحت فرض H_0 کمتر یا مساوی α باشد یعنی P –مقدار برابر است با

$$\text{مقدار} - P = p\left[\overline{X} \le \bar{x} \middle| \quad H_0\right] \le \alpha$$

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

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

5–7 آزمون فرض برای پارامتر توزیع دوجمله‌ای

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

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

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

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

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

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

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

آماره آزمون

$$Z_\circ = \frac{\frac{x}{n} - P_\circ}{\sqrt{\frac{x}{n}(1-\frac{x}{n})}} = \frac{\frac{8}{50} - 0.2}{\sqrt{\frac{8}{50}(1-\frac{8}{50})}} = -0.109$$

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

6–7 آزمون فرض برای تفاضل میانگین دو جامعه

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

فرض کنید دو جامعه مورد بررسی مستقل و دارای توزیع نرمال وبه ترتیب دارای میانگین‌های μ_1 و μ_2 و واریانس σ_1^2 و σ_2^2 باشند. انواع فرضها در مورد مقایسه میانگین‌های دو جامعه عبارتند از:

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

$$1)\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 > 0 \end{cases} \qquad 2)\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 < 0 \end{cases} \qquad 3)\begin{cases} H_0 : \mu_1 - \mu_2 = 0 \\ H_1 : \mu_1 - \mu_2 \neq 0 \end{cases}$$

آماره آزمون برای مقایسه تفاضل $\mu_1 - \mu_2$ براساس نمونه تصادفی از جامعه اول و دوم خواهد بود. اگر $\bar{X}_2, \bar{X}_1$ به ترتیب میانگین نمونه اول و دوم باشند کمیت محوری، همان کمیت محوری بحث شده در فصل 6 خواهد بود.

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$$

که Z تحت فرض H_0 دارای توزیع نرمال استاندارد خواهد بود Z_0 مشاهده شده براساس مشاهده نمونه از دو جامعه برابر است با

$$Z_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}$$

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

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

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

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

7-7 آزمون فرض برای واریانس جامعه

یکی از راههای بررسی تغییرپذیری جامعه، بررسی یا آزمون فرض درباره واریانس جامعه است. به عنوان مثال مهندس کنترل کیفیت باید مراقبت نماید که تغییرپذیری اندازه‌ها از حد معینی بیشتر نشود یا یک داروساز ممکن بخواهد بداند که آیا میزان تغییر در اثر بخشی یک دارو در حدود قابل قبول است یا نه.

انواع فرضهایی که می‌توان درباره واریانس داشت عبارتند از:

$$1)\left|\begin{array}{l} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{array}\right. \qquad 2)\left|\begin{array}{l} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 \end{array}\right. \qquad 3)\left|\begin{array}{l} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{array}\right.$$

اگر جامعه مورد بررسی دارای توزیع نرمال با میانگین μ مجهول و واریانس σ^2 باشد درفصل 6 نشان داده شد که کمیت $\frac{(n-1)s^2}{\sigma^2}$ یک کمیت محوری است و تحت

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

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

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

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

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

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

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

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

برای آزمون فـرض $H_0 = \sigma^2 = \sigma_0^2$ در مقابـل $H_1 : \sigma^2 > \sigma_0^2$ کـافی اسـت آمـاره $\chi_\circ^2 = \dfrac{(n-1)s^2}{\sigma_\circ^2}$ تحت فرض $H_\circ$ و مـشاهدات محاسـبه و بـا مقـدار جـدول $\chi^2_{(n-1,\alpha)}$ مقایسه نمود. اگر $\chi_\circ^2 > \chi^2_{(n-1,\alpha)}$ باشد فرض $H_\circ$ رد می‌شود.

8–7 آزمون فرض برای نسبت دو واریانس

در این بخش به دنبال شیوه‌ای می‌گردیم که پاسخ مناسب آماری را در مورد صحت فرض اختلاف و یا عدم اختلاف واقعی بین واریانسهای دو جامعه ارائه دهد. تصمیم‌گیریهایی که به سازگاری یا عدم سازگاری پراش(واریانس) دو جامعه مربوط می‌شود معمولاً براساس آزمون نسبت واریانسها قرار دارد.

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

اگر متغیرتصادفی X دارای توزیع نرمال با میانگین μ_1 و واریانس σ_1^2 و متغیر تصادفی Y دارای توزیع نرمال با میانگین μ_2 و واریانس σ_2^2 و متغیرهای X و Y از هم مستقل باشند آنگاه فرضهایی می‌توان برای مقایسه واریانس دو جامعه به صورت زیر نوشت:

$$1)\begin{cases} H_\circ : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 > \sigma_2^2 \end{cases} \qquad 2)\begin{cases} H_\circ : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 < \sigma_2^2 \end{cases} \qquad 3)\begin{cases} H_\circ : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

اگر ${S_1}^2$ واریانس یک نمونه m تایی از جامعه اول و ${S_2}^2$ واریانس یک نمونه n تایی از جامعه دوم باشد در فصل 6 دیدیم که آماره $F = \left(\frac{\sigma_2}{\sigma_1}\right)^2 \frac{S_1^2}{S_2^2}$ دارای توزیع فیشر با $m-1$ و $n-1$ درجه آزادی است. برای آزمون فرضهای اخیر از آماره F استفاده می‌کنیم.

در مقایسه واریانس‌های دو جامعه برای سهولت این قرارداد را همیشه درنظر داریم که واریانس نمونه بزرگتر را در صورت قرار می‌دهیم به قسمی که نسبت واریانسهای نمونه همیشه بزرگتر یا مساوی 1 باشد.

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

1- مقدار آماره را تحت H_0 و نمونه از رابطه زیر بدست می‌آوریم.

$$F_0 = \left(\frac{\sigma_2}{\sigma_1}\right)^2 \frac{S_1^2}{S_2^2}$$

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

3- اگر $F_0 > F_{(m-1, n-1, \alpha)}$ باشد فرض H_0 را رد می‌کنیم در غیر اینصورت آن را می‌پذیریم.

پایان فصل ۷

بسم الله الرحمن الرحيم

آمار و احتمالات مهندسی

رشته : کامپیوتر

دکتر پرویز نصیری

فصل ۸

همبستگی و رگرسیون

در این فصل مطالب ذیل ارائه می شود:

- ضریب همبستگی
- خط رگرسیون
- پیش بینی
- آزمون فرض برای β
- آزمون فرض برای α

1-8 ضریب همبستگی

در فصل سوم ضریب همبستگی دو متغیر تصادفی X و Y را به صورت زیر تعریف کردیم:

$$\rho = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X-\mu_x)^2 . E(Y-\mu_y)^2}}$$

بررسی نمودار پراکنش و محاسبه ضریب همبستگی بین دو متغیر تصادفی X و Y، و تجزیه و تحلیل مقدار ضریب همبستگی در نقاط $\rho = 1$ یا $\rho = -1$ این نکته را در ذهن متصور می‌کند که بین دو متغیر X و Y یک رابطه خطی وجود دارد. اگر ρ یکی از دو مقدار 1 یا 1− را اختیار نکند آیا در صفحه $X\,Y$ خطی وجود دارد که احتمالات برای X و Y در نواری در اطراف این خط متمرکز باشد؟ تحت برخی شرایط جواب این سئوال مثبت است. تحت این شرایط می‌توان ρ را به عنوان اندازه‌ای برای شدت تمرکز احتمالات X و Y در اطراف این خط دانست.

بررسی همبستگی دو متغیر تصادفی X و Y در جامعه مستلزم دانستن توزیع توأم آنها یا بررسی تمام اعضاء جامع دو بعدی است که همراه با هزینه و صرف وقت زیاد است. لذا بـه جای بررسی رابطه بین X و Y در جامعه به بررسی در نمونـه اکتفـا مـی‌کنـیم. یکـی از روشهایی که بوسیله آن می‌توان همبستگی بین دو متغیر را نـشان داد نمـودار پـراکنش یـا پراکندگی است که بوسیله آن نوع ارتباط دو متغیر مشخص می‌شود.

8–1–2 خصوصیات ضریب همبستگی نمونه‌ای پیرسن

1– همواره $-1 \le r \le 1$ و مستقل از واحد اندازه‌گیری است.

2– هنگامی که $r = 1$ است، همبستگی دو متغیر X و Y شدید وهمسو است افزایش یکی باعث افزایش دیگری می‌شود.

3- هنگامی که $r=-1$ است، همبستگی دو متغیر X و Y شدید وخلاف هم و افزایش یکی باعث کاهش دیگری می‌شود.

4- هنگامی که r در همسایگی صفر است، همبستگی دو متغیر ضعیف است.

2-8 خط رگرسیون

فرض کنید $f(x,y)$ تابع چگالی احتمال توام دو متغیر X و Y و $f(x)$ تابع چگالی احتمال حاشیه‌ای X باشد. تابع چگالی احتمالی شرطی Y به شرط $X=x$ برابر است با:

$$f(y \mid X=x)=\frac{f(x,y)}{f(x)}$$

و میانگین شرطی Y به شرط $X=x$ برابر است با:

$$E(Y \mid X=x)=\int_{-\infty}^{\infty} y f(y|x) dy=\frac{\int_{-\infty}^{\infty} y f(x, y) dy}{f(x)}$$

میانگین Y به شرط $X=x$ فقط تابعی از x خواهد بود که این تابع تحت شرایطی خطی است و به صورت زیر تعریف می‌شود:

$$E(Y|X)=\alpha+\beta X$$

برای راحتی $E(Y|X)$ را به Y نمایش می‌دهند و معادله خط رگرسیون را در جامعه به صورت زیر می‌نویسند:

$$Y=\alpha+\beta X$$

که α و β به ترتیب عرض از مبداء و شیب خط رگرسیون و X را متغیر مستقل و Y را متغیر وابسته می‌نامند.

معادله خط رگرسیون در جامعه را می‌توان برای تک تک زوج‌ها به صورت زیر نوشت:

$$Y_i = \alpha + \beta X_i \qquad i = 1,2,\ldots,N$$

اگر همه X_iها و Y_iها را دقیقاً مشاهده یا اندازه بگیریم رابطه اخیر همواره برقرار است و می‌توان با مشاهده همه زوج‌های جامعه پارامترهای α و β را محاسبه کرد. اما اغلب در اندازه‌گیریها یا محاسبه X_iها و Y_iها دچار اشتباه و خطا هستیم و ممکن است مقدار X_iها یا Y_iها را کمتر یا بیشتر اندازه بگیریم. در این حالت دچار خطایی به اندازه ε_i خواهیم شد. لذا معادله خط رگرسیون را به صورت زیر در نظر می‌گیریم.

$$Y_i = \alpha + \beta x_i + \varepsilon_i$$

اگر در اندازه‌گیریها همه ε_iها صفر باشند معادله اخیر همان معادله $Y_i = \alpha + \beta X_i$ خواهد بود. فرض می‌شود ε_i یک متغیر تصادفی نرمال با میانگین صفر و واریانس σ^2 است. چون Y_i تابع خطی از ε_i است. Y_iها نیز متغیر تصادفی نرمال با میانگین $\alpha + \beta X_i$ و واریانس σ^2 می‌باشند. یادآوری می‌شود که X_iها متغیرهای تصادفی نیستند.

اما به دلیلی که قبلاً گفتیم، بررسی کل جامعه مقرون به صرفه نیست. لذا به جای بررسی جامعه به بررسی یک نمونه nتایی می‌پردازیم. دراین حالت خط رگرسیون در نمونه را به صورت زیر تعریف می‌کنیم.

$$y_i = a + bx_i + e_i \quad , \quad i = 1,2,\ldots,n$$

که e_i اندازه خطا در مشاهده iام است. مقادیر بدست آمده برای a و b براساس نمونه nتایی، به ترتیب تخمینی برای α و β خواهد بود. در ادامه بحث فرض می‌شود. e_iها از هم مستقل و دارای توزیع نرمال با میانگین صفر و واریانس ${\sigma_e}^2$ می‌باشند. از رابطه زیر می‌توان e_iها را به صورت زیر بدست آورد:

$e_i = y_i - a - bx_i$

می‌توان با می‌نیمم کردن $\sum_{i=1}^{n} e_i$ یا $\sum_{i=1}^{n} e_i^2$ ، مقادیر a و b را بدست آورد. ولی بهتر اســت با می‌نیمم کردن $\sum_{i=1}^{n} e_i^2$، a و b را بدست آورد (چرا؟)

لذا:

$$e_i = y_i - a - bx_i$$

$$e_i^2 = (y_i - a - bx_i)^2$$

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

با فرض برابری این رابطه

$$\varphi(a,b) = \sum_{i=1}^{n} e_i^2$$

تابع $\varphi(a,b)$ از تـابع b و a برای بدسـت آوردن a و b خواهد بود. بـرای بدسـت آوردن a و b از تـابع $\varphi(a,b)$ یکبار نسبت به a و یکبار نسبت به b مشتق می‌گیریم و برابر صفر قرار می‌دهیم.

$$\frac{\partial \varphi(a,b)}{\partial a} = \sum_{i=1}^{n} (y_i - a - bx_i) = 0$$

$$\frac{\partial \varphi(a,b)}{\partial b} = \sum_{i=1}^{n} x_i (y_i - a - bx_i) = 0$$

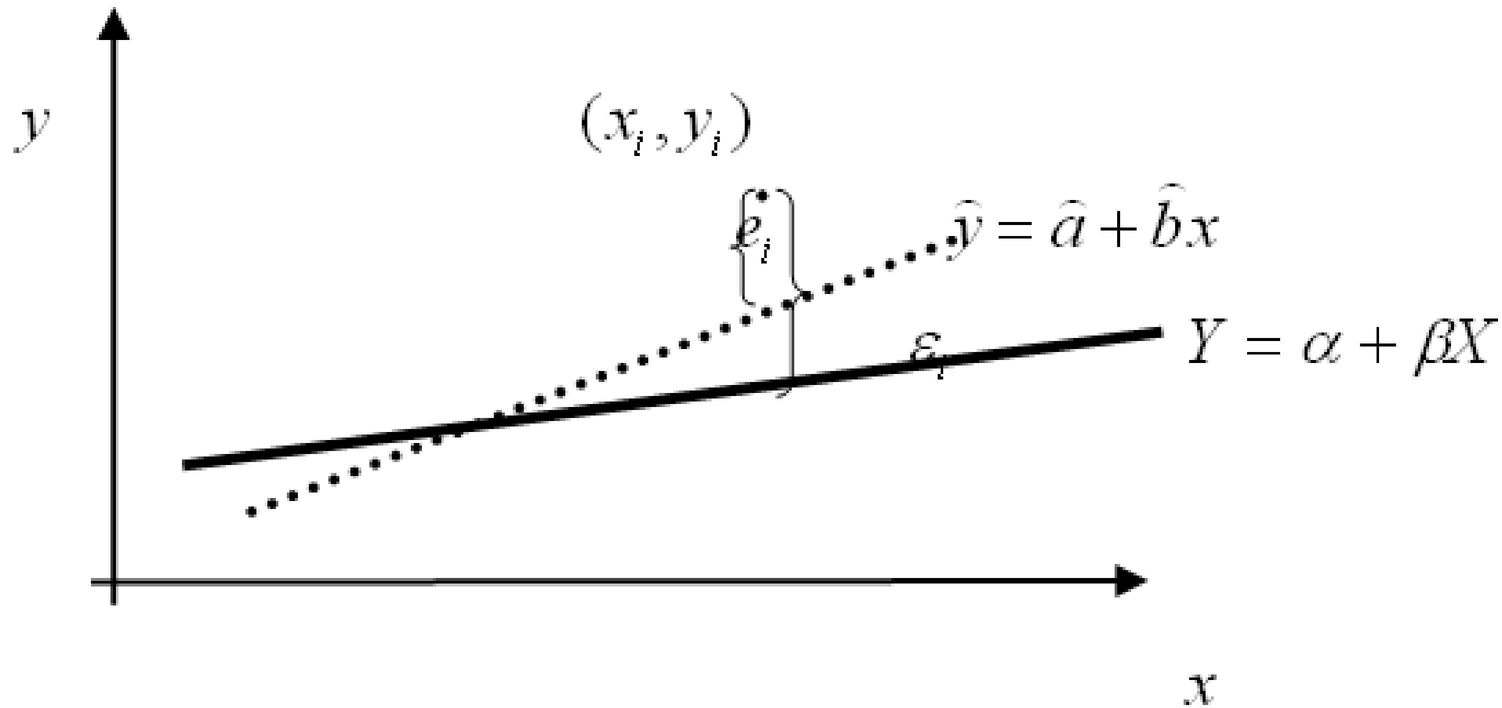
از حل دو معادله اخیر a و b عبارتند از

$$\hat{b} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\ \bar{y}}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \quad , \quad \hat{a} = \bar{y} - \hat{b}\bar{x}$$

که $\hat{a}$ و $\hat{b}$ به ترتیب عرض از مبداء و شیب خط رگرسیون در نمونه می‌باشـند. و معادلـه خط برازش شده عبارتست از

$$\hat{y}_i = \hat{a} + \hat{b}x_i \quad , \quad i = 1,2,\ldots,n$$

در نمودار زیر مقادیر ε_i و e_i برای نقطه‌ای مشخصی با هم مقایسه می‌شوند.



3-8 پیش بینی

با محاسبه مقادیر $\hat{a}$ ، $\hat{b}$ خط رگرسیون نمونه $\hat{y}=\hat{a}+\hat{b}x$ کاملاً مشخص مـی‌شـود. بـا معلوم بودن خط رگرسیون برازش شده در نمونه، مقدار $\hat{y}_\circ$ برای مقدار مشاهده شــده $x_\circ$ را، مقدار پیش بینی گویند.

مثـال 8–3–1 در مثال اخیر مقادیر پیش بینی را برای مقادیر $x_\circ = 70/33$ ، $x_\circ = 76/33$ و $x_\circ = 80/12$ بدست آورید.

خط برازش داده شده: $\hat{y} = 194/824+3/557\,x$

برای $x_\circ = 70/33$: $\hat{y}_\circ = 194/824+3/557(70/33)=444/9878$

برای $x_\circ = 76/33$: $\hat{y}_\circ = 194/824+3/557(76/33)=466/3298$

برای $x_\circ = 80/12$: $\hat{y}_\circ = 194/824+3/557(80/12)=479/8108$

4-8 آزمون فرض برای β

در بخش 8-2 متذکر شدیم که مقدار بدست آمده برای b براساس یک نمونه تصادفی n تایی یک برآورد گر برای β است به طوری که

$$\hat{\beta} = \hat{b} = \frac{\sum X_i Y_i - n\overline{X}\ \overline{Y}}{\sum (X_i - \overline{X})^2} = \frac{\sum (X_i - \overline{X}) Y_i}{\sum (X_i - \overline{X})^2} = \sum_{i=1}^{n} c_i Y_i$$

$$c_i = \frac{X_i - \overline{X}}{\sum (X_i - \overline{X})^2}$$

از معادله اخیر نتیجه می‌شود که $\hat{\beta}$ یک تابع خطی از Y_iهاست. و چون Y_iها دارای توزیع نرمال هستند $\hat{\beta}$ نیز دارای توزیع نرمال با میانگین و واریانس زیر است:

$$E(\hat{\beta}) = E\left[\sum_{i=1}^{n} c_i Y_i\right] = \sum_{i=1}^{n} c_i E(Y_i) = \sum_{i=1}^{n} c_i (\alpha + \beta X_i)$$

$$= \alpha \sum_{i=1}^{n} c_i + \beta \sum_{i=1}^{n} c_i X_i = \beta$$

مقادیر مثبت یا منفی β در معادله خط رگرسیون $Y=\alpha+\beta X$ بیانگر این واقعیت است که بین X و Y رابطه خطی مستقیم وجود دارد. در مواقعی که β نامعلوم است. تنها می‌توان از داده‌ها نمونه استنباطی روی β انجام داد. با آزمون فرض روی β می‌توان پی به رابطه خطی X و Y برد. رد فرض $H_0:\beta=0$ در مقابل $H_1:\beta\neq 0$ دلیل کافی برای وجود رابطه خطی مستقیم بین X و Y می‌باشد. برای آزمون فرض $H_0:\beta=0$ در مقابل $H_1:\beta\neq 0$ از آماره آزمون زیر تحت فرض H_0 استفاده می‌کنیم

$$Z=\frac{\hat{b}-\beta_{\circ}}{\sqrt{\dfrac{\sigma^2}{\sum(X_i-\overline{X})^2}}}$$

با فرض معلوم بودن σ^2 ، متغیر Z دارای توزیع نرمال استاندارد است.

در مواقعی که σ^2 نامعلوم است از برآوردگر نمونه‌ای آن استفاده می‌کنیم.

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{n-2}\sum_{i=2}^{n}\left[y_i - \hat{a} - \hat{b}x_i\right]^2$$

در این حالت آماره آزمون تحت فرض H_0 عبارتست از:

$$T = \frac{\hat{b} - \beta}{\sqrt{\frac{\hat{\sigma}^2}{\sum (X_i - \overline{X})^2}}}$$

که دارای توزیع استودنت با $n-2$ درجه آزادی است. برای آزمون فـرض مراحـل زیـر را انجام می‌دهیم.

1- مقدار $T_0 = \frac{\hat{b} - \beta}{\sqrt{\frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2}}}$ را تحت فرض H_0 و نمونه محاسبه می‌کنیم.

2– مقدار $t(n-2, \alpha/2)$ را از جدول استودنت بدست می‌آوریم.

3– اگر $T_\circ > t(n-2, \alpha/2)$ یا $T_\circ < -t(n-2, \alpha/2)$ باشد فرض H_0 را رد می‌کنیم.

5-8 آزمون فرض برای α

در بخش 8-2 از $\hat{a}$ به عنوان یک برآوردگر نقطه‌ای برای α یاد کردیم به طوری که:

$$\hat{\alpha} = \hat{a} = \bar{y} - \hat{b}\,\bar{x} = \frac{1}{n}\sum_{i=1}^{n} Y_i - \bar{X}_i \sum_{i=1}^{n} c_i Y_i = \sum_{i=1}^{n} \left[\frac{1}{n} - \bar{X} c_i\right] Y_i$$

$$E(\hat{\alpha}) = \sum_{i=1}^{n} \left[\frac{1}{n} - \bar{X} c_i\right] E(Y_i) = \sum_{i=1}^{n} \left[\frac{1}{n} - \bar{X} c_i\right] (\alpha + \beta X_i)$$

$$= \sum_{i=1}^{n} \left[\frac{\alpha}{n} + \frac{\beta X_i}{n} - \alpha \bar{X} c_i - \beta \bar{X} c_i X_i\right] = \alpha$$

$$V(\hat{\alpha}) = \sum_{i=1}^{n} \left[\frac{1}{n} - \overline{X}c_i\right]^2 V(Y_i) = \sigma^2 \sum_{i=1}^{n} \left[\frac{1}{n} - \overline{X}c_i\right]^2$$

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

همانند $\hat{\beta}$ ، $\hat{\alpha}$ هم دارای توزیع نرمال با میانگین α و واریـانس $\sigma^2 \left(\frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2}\right)$ است.

در خط رگرسیون، از عرض از مبدأ α می‌توان به عنوان یک مؤلفه ثابت در رابطه بین X و Y یاد کرد. مثلاً وقتی که Y هزینه یک نوبت تولید و X تعداد واحـدها در هـر نوبـت باشد، α هزینه ثابت شروع به کار نوبت تولید و β هزینه متغیر به ازای هر واحـد تولیـد می‌باشد. اگر $\alpha = 0$ باشد خط رگرسیون از مبدأ می‌گذرد و مدل رگرسیونی به مدل ساده $Y = \beta X$ تبدیل می‌شود. آزمون فرض روی α همانند آزمون فرض روی β است.

برای آزمون فرض $H_0 : \alpha = 0$ در مقابل $H_1 : \alpha \neq 0$ آماره آزمون تحت فرض H_0 برابر است با:

$$Z = \frac{\hat{a} - \alpha}{\sqrt{\sigma^2 \left[\frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2} \right]}}$$

که دارای توزیع نرمال استاندارد است.

اگر σ^2 مجهول باشد از برآوردگر آن $\sigma_e^2 = \frac{1}{n-2} \sum_{i=1}^{n} e_i^2$ استفاده می‌کنیم و آماره آزمون تحت فرض H_0 در این حالت برابر است با:

$$T = \frac{\hat{a} - \alpha}{\sqrt{\hat{\sigma}_e^2 \left[\frac{1}{n} + \frac{\overline{X}^2}{\sum (X_i - \overline{X})^2} \right]}}$$

و دارای توزیع استودنت با $n-2$ درجه آزادی است. برای آزمون فرض مراحل زیر را انجام می‌دهیم:

1- مقدار $T_{\circ} = \dfrac{\hat{a} - \alpha}{\sqrt{\hat{\sigma}_e^2 \left[\dfrac{1}{n} + \dfrac{\overline{X}^2}{\sum (X_i - \overline{X})^2} \right]}}$ راتحت فرض H_0 و نمونه محاسبه می‌کنیم.

2- مقدار $t(n-2, \alpha/2)$ را از جدول استودنت بدست می‌آوریم.

3- اگر $T_{\circ} > t(n-2, \alpha/2)$ یا $T_{\circ} < -t(n-2, \alpha/2)$ باشد فرض H_0 را رد می‌کنیم.

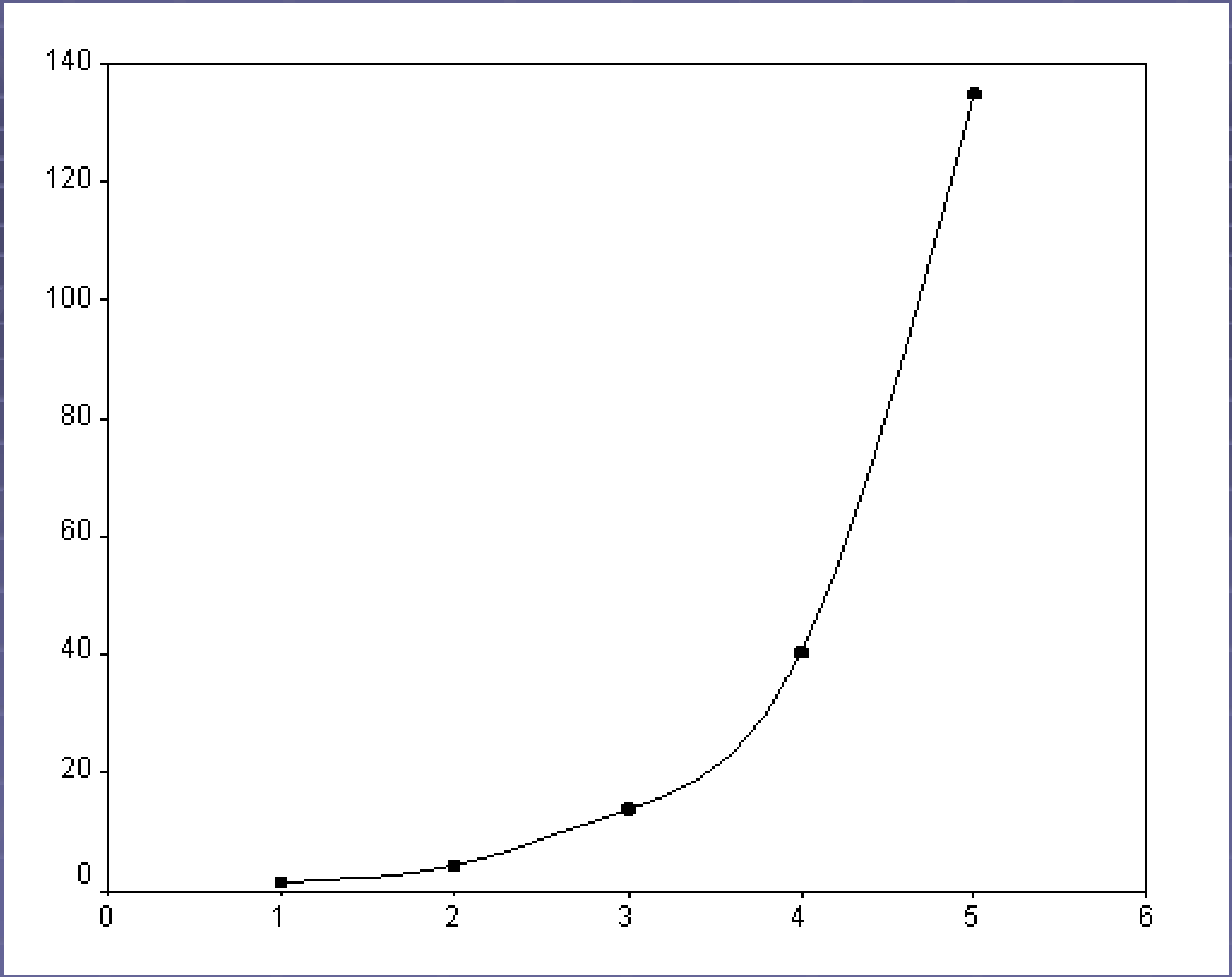
مثال 8-5-2 جدول زیر را در نظر بگیرید

x_1	1	2	3	4	5
y_1	1/6	4/5	13/8	40/2	135

الف- نمودار پراکنش را رسم کنید.

ب- اگر منحنی حاصل از نمودار پراکنش دارای معادله هندسی $y_1 = cx_1^d$ باشد مقـادیر c و d را بدست آورید.

ج- برای x_1=2/5 مقدار y_1 را پیش بینی کنید.


140
120
100
80
60
40
20
0
0
1
2
3
4
5
6

معادله $y_1 = cx_1^d$ با گرفتن لگاریتم

$$\log(y_1) = \log(c) + d\log(x_1)$$

به معادله $y = a + bx$ تبدیل می‌شود کـه در آن $y = \log(y_1)$ ، $a = \log(c)$، $b = d$ و $x = \log(x_1)$ است. با توجه به تبدیلات

x	0	0/30	0/48	0/60	0/69
y	0/20	0/65	1/14	1/60	2/13

ضرایب a و b با توجه به تبدیلات محاسبه می‌شود.

$$\hat{b} = \frac{\sum x_i y_i - n\bar{x}\ \bar{y}}{\sum (x_i - \bar{x})^2} = 2/684$$

$\hat{a} = \bar{y} - b\bar{x} = 0/033$

معادله رگرسیون تبدیل یافته:

$$\hat{y} = 0/033 + 2/684\, x$$

معادله رگرسیون قبل از تبدیل:

$$\hat{y}_1 = 1/079 + x_1^{2/684}$$

چون

$$C = (10)^{\hat{a}} = (10)^{0.033} = 1.079$$

مقدار پیش بینی برای $x_1 = 2/5$ برابر است با

$$\hat{y}_1 = 1/079 + (2/5)^{2/684} = 12/78$$

پایان فصل ۸