آمار و احتمالات

استاد:

سرکار خانم صالحی

دانشجو:

منصور نوری

www.mnouri.ir

بهار ۱۳۹۱

علم آمار

در آمار توصیفی مفهوم اولیه جامعه آماری می‌باشد. جامعه آماری دسته‌ای از اشیاء و یا افراد را گویند که لااقل در یک خاصیت مشترک باشند و ما بخواهیم موضوع یا موضوع‌هایی را در آن‌ها بررسی کنیم. جامعه آماری یا محدود است یا نامحدود و جامعه آماری محدود جوامعی هستند که اعضای آن قابل شمارش است و جامعه آماری نامحدود اعضای آن قابل شمارش نیستند. هر چه خاصیت مشترک بیشتر شود جامعه آماری محدودتر می‌شود و صفت مشترک را صفت مشخصه می‌گویند. اندازه جامعه آماری را با N نشان می‌دهند.

انواع صفت متغیر:

الف- متغیر کمی: که قابل اندازه‌گیری و قابل شمارش باشد مانند سن، قد، وزن و ...

ب – متغیر کیفی: که قابل اندازه‌گیری و شمارش نیست مانند شماره ملی، جنسیت، گروه خونی و ...

انواع متغیر کمی:

الف – متغیر کمی پیوسته: متغیری است که مجموعه مقادیر خود را از مجموعه اعداد حقیقی اختیار می‌کند. مانند وزن

ب – متغیر کمی ناپیوسته (گسسته): مجموعه مقادیر خود را از مجموعه اعداد طبیعی یا معادل هم ارز آن اختیار می‌کند.

انواع متغیر کیفی:

الف – متغیر کیفی اسمی: این متغیرها قابل شمارش نیستند مانند شماره شناسنامه که کیفی اسمی می‌باشد که فقط یک مشخصه برای فرد می‌باشد.

ب – متغیر کیفی ترتیبی: این متغیرها همگی کیفی اسمی‌اند ولی ویژگی‌های خاصی دارند که آن‌ها را در یک ترتیب قابل قبول قرار می‌دهد. مانند ایزوها، درجه نظامی، مقاطع تحصیلی و متغیری می‌تواند کیفی ترتیبی باشد که در قالب یک استاندارد باشد.

مراحل انجام یک کار آماری

۱- مشاهده یا جمع‌آوری اطلاعات
۲- پردازش یا مرتب کردن داده‌ها و تعیین شاخص‌های آماری
۳- تحلیل آماری

مشاهده:

۱- **سرشماری:** سنجش تک تک افراد جامعه آماری است. به دلایل مشکلاتی که در سرشماری وجود دارد همواره اقدام به این عمل نمی‌شود. این مشکلات شامل: هزینه بر بودن، زمان بر بودن و ... در برخی موارد موجب از بین رفتن جامعه آماری می‌شود.

۲- **نمونه‌گیری:** تعدادی از افراد جامعه را انتخاب و مورد بررسی قرار می‌دهیم. دو نکته مهم در نمونه‌گیری را می‌توان چنین گفت الف: به چه روشی انتخاب شود؟ ب: چند نمونه لازم است؟

جداول آماری

جدول توزیع فراوانی

مرتب کردن داده‌ها توسط جدول فراوانی انجام می‌شود.

حدود دسته	فراوانی مطلق
x_i - x_{i+1}	F_i
10 - 20	5
20 - 30	3
30 - 40	4
40 - 50	8
	$\sum F_i = n \rightarrow 20$

متغیر کمی پیوسته

اندازه متغیر	فراوانی مطلق	فراوانی نسبی	فراوانی تجمعی
x_i	F_i	f_i	$F(x_i)$
A	5	$f_1 = \frac{5}{20} = 0.25$	5
B	3	$f_1 = \frac{3}{20} = 0.15$	8
AB	4	$f_1 = \frac{4}{20} = 0.2$	12
O	8	$f_1 = \frac{8}{20} = 0.4$	20
	$\sum F_i = n \rightarrow 20$	$\sum_{i=1}^{4} f_i = 1$	-----

اندازه متغیر	فراوانی مطلق
x_i	F_i
A	5
B	3
AB	4
O	8
	$\sum F_i = n \rightarrow 20$

$X_{min} = 10$

$X_{max} = 50$

$K = 4$ تعداد دسته

$R =$ دامنه تغییرات (مجموع طول دسته نباید از دامنه تغییرات کوچک‌تر باشد)

$R = X_{max} - X_{min}$

$C = \frac{R}{K}$ طول دسته

فراوانی نسبی:

سهم یا نسبت صفت را در جامعه نشان می‌دهد مجموع فراوانی‌های نسبی همواره یک می‌باشد. و

فراوانی نسبی $= \frac{\text{فراوانی مطلق}}{\text{اندازه جامعه یا نمونه}} = f_i = \frac{F_i}{n}$

$$\sum_{i=1}^{n} f_i = 1 \rightarrow \sum F_i = F_1 + F_2 + \ldots + F_n = \frac{F_1}{n} + \frac{F_2}{n} + \ldots + \frac{F_n}{n} = \frac{F_1 + F_2 + \ldots + F_n}{n} = \frac{n}{n} = 1$$

فراوانی تجمعی:

$F(x_i)$ آخرین فراوانی تجمعی همان اندازه جامعه یا نمونه است و می‌گوید تا چه اندازه‌ای چقدر داریم: (تا یک طبقه چند تاست). (تفاضل دو مرکز دسته متوالی همان طول دسته است. 10=15-25)

حدود دسته	فراوانی مطلق	مرکز دسته
x_i - x_{i+1}	F_i	X'_i
10 - 20	5	$\frac{10+20}{2} = 15$
20 - 30	3	$\frac{20+30}{2} = 25$
30 - 40	4	$\frac{30+40}{2} = 35$
40 - 50	8	$\frac{40+50}{2} = 45$
	$\sum F_i = n \rightarrow 20$	

تمرین ۱:

در جدول زیر خانه‌های خالی را پر کنید؟

حدود دسته	مرکز دسته	فراوانی مطلق	فراوانی نسبی	فراوانی تجمعی
x_i - x_{i+1}	x'_i	F_i	f_i	$F(x_i)$
$\boxed{3}$-$\boxed{5}$	4	7	$f_1=\frac{7}{20}=\boxed{0.35}$	$\boxed{7}$
$\boxed{5}$-$\boxed{7}$	$\boxed{6}$	$\boxed{6}$	$f_1=\frac{6}{20}=0.3$	$\boxed{13}$
$\boxed{7}$-$\boxed{9}$	$\boxed{8}$	$\boxed{2}$	$f_1=\frac{2}{20}=\boxed{0.1}$	15
$\boxed{9}$-$\boxed{11}$	10	$\boxed{5}$	$f_1=\frac{5}{20}=\boxed{0.25}$	20
		$n = 20$	$\sum_{i=1}^{4} f_i = 1$	

$x'_1 = 4$ مرکز دسته اول

$x'_4 = 10$ مرکز دسته چهارم

$10 - 4 = 6 \quad , 4 - 1 = 3 \quad \rightarrow \frac{6}{3} = 2$ طول دسته

$F_i = f_i \times n \quad \rightarrow 0.3 \times 20 = 6$

$20 - 15 = 5 \quad , 7 + 6 = 13$

احتمالات

در آمار احتمال اندازه امکان وقوع حادثه است.

مفهوم اولیه در احتمال حادثه و پیش آمد می‌باشد. هر کاری را که راجع به وقوع یا عدم وقوع آن صحبت کنیم به آن حادثه یا پیش آمد می‌گویند. حوادث را با حروف بزرگ نمایش می‌دهند. مثلاً احتمال حادثه A برابر است با: p(A)

به دنبال هر آزمایشی یک حادثه اتفاق می‌افتد.

آزمایش: برقرار کردن مجموعه شرط‌های معین به دفعات دلخواه و تکراری است.

شروط معین: این شروط مثلاً در پرتاب یک مکعب باید دارای ویژگی‌های: در زمین سفت و سخت باشد، مولکول سازنده آن یکسان باشد و همه رویه‌های آن یک اندازه باشد.

در آمار ۳ حادثه وجود دارد:

۱- **یقین** (احتمال حوادث یقین ۱ می‌باشد)

۲- **غیرممکن** (احتمال حوادث غیرممکن صفر می‌باشد)

۳- **تصادفی** (از قبل نمی‌توان گفت چه چیزی می‌آید ولی می‌توانیم پیش‌بینی کنیم و احتمال حوادث تصادفی بین ۱ و صفر اتفاق می‌افتد)

طبق تعریف کلاسیک احتمال: احتمال هر حادثه‌ای برابر است با تعداد حالات مطلوب به تعداد حالات ممکن.

$$\text{احتمال} = \frac{\text{تعداد حالات مطلوب}}{\text{تعداد حالات ممکن}} \leftarrow P(A)=\frac{m_A}{n}$$

هیچ‌وقت m_A نمی‌تواند بزرگ‌تر از n باشد و احتمال همیشه یک عدد مثبت بین ۱ و صفر می‌باشد و احتمال هر حادثه‌ای شبیه فراوانی نسبی می‌باشد.

خواص (قوانین) احتمال

۱- احتمال هر حادثه‌ای یک عدد مثبت است. $P(A) \geq 0$

۲- احتمال هر حادثه‌ای همواره بین ۱ و صفر است. $0 \leq P(A) \leq 1$

۳- مجموع احتمال حوادث همواره یک است. $\Sigma p_i = 1$

تمام قوانینی را که در مجموعه‌ها داریم در حوادث نیز صدق می‌کند.

۴- احتمال عکس هر حادثه برابر است با یک منهای احتمال آن حادثه $P(\overline{A}) = 1- P(A)$ و منظور از $\overline{A}$ یعنی غیر از A (نه A)

۵- اگر دو حادثه ناسازگار باشند:

$$P(A.B) = P(A) \times P(B)$$
$$P(A+B) = P(A) + P(B)$$
$$P(A+B+C) = P(A) + P(B) + P(C)$$

حوادث ناسازگار: حوادثی هستند که عضو مشترک ندارند.

۶- اگر دو حادثه سازگار باشند:

$$P(A+B) = P(A) + P(B) – P(A.B)$$
$$P(A+B+C) = P(A) + P(B) + P(C) – P(AB) – P(AC) - P(BC) + P(ABC)$$
$P(A\times B) = P(A) \times P(B|A)$ احتمال شرطی
$$P(ABC) = P(A) \times P(B|A) \times P(C|AB)$$

مثال: در کیسه‌ای ۴ مهره سیاه و ۶ مهره قرمز وجود دارد. دو مهره به تصادف از آن کیسه خارج می‌کنیم مطلوب است احتمال اینکه: $(R=6)$, $(B=4)$

الف) هر دو سیاه باشد. $P(B_1 B_2)$

ب) هر دو قرمز باشد. $P(R_1 R_2)$

ج) اولی سیاه و دومی قرمز باشد. $P(BR)$

د) یکی سیاه و یکی قرمز باشد. $P(BR+RB)$

پاسخ:

الف) هر دو سیاه باشد.

$$P(B_1 B_2) = P(B_1) \times P(B_2 \mid B_1) \rightarrow \frac{4}{10} \times \frac{3}{9} = \frac{12}{90}$$

ب) هر دو قرمز باشد.

$$P(R_1 R_2) = P(R_1) \times P(R_2 \mid R_1) \rightarrow \frac{6}{10} \times \frac{5}{9} = \frac{30}{90}$$

ج) اولی سیاه و دومی قرمز باشد.

$$P(BR) = P(B) + P(R|B)\rightarrow \frac{4}{10} \times \frac{6}{9} = \frac{24}{90}$$

د) یکی سیاه و یکی قرمز باشد.

$$P(BR+RB) = P(BR) + P(RB) \rightarrow P(B) \times P(R|B)+ P(R) \times P(B|R)\rightarrow \frac{4}{10} \times \frac{6}{9} + \frac{6}{10} \times \frac{4}{9} = \frac{48}{90}$$

مثال: با توجه به جدول زیر از این جامعه ۵۰ عضوی ۲ نفر به تصادف انتخاب می‌شوند. احتمال اینکه این دو نفر **مرد یا متأهل باشند**، چیست؟

(جدول سازگار)

جنسیت	متأهل	مجرد	جمع
زن	12	8	20
مرد	18	12	30
جمع	30	20	50

$$P(\text{متأهل+مرد}) = P(\text{مرد}) + P(\text{متأهل}) - P(\text{مرد متأهل}) = \frac{30}{50} + \frac{30}{50} - \frac{18}{30} = \frac{42}{50}$$

(احتمال شرطی) $P(AB) = P(A) \times P(B|A) \rightarrow P(B|A) = \frac{P(BA)}{P(A)}$

احتمال شرطی $= \dfrac{\text{احتمال حاصل ضرب دو حادثه}}{\text{احتمال حادثه دوم}} \leftrightarrow P(B|A) = \frac{P(BA)}{P(A)}$

اگر علاوه بر شروط معین که در آزمایش وجود دارد و یک حادثه اتفاق بی افتد یک شرط دیگر را برقرار کنیم به آن احتمال شرطی می‌گویند.

مثال: یک مکعب سالم را در زمین سخت پرتاب می‌کنیم مطلوب است احتمال اینکه مجموع رویه‌ها کمتر از هفت باشد به شرط آن که رویه اول یک باشد چیست؟

$$\left\{\begin{matrix} 1,1 & 1,2 & 1,3 & 1,4 & 1,5 & 1,6 \\ 2,1 & 2,2 & 2,3 & 2,4 & 2,5 & 2,6 \\ 3,1 & 3,2 & 3,3 & 3,4 & 3,5 & 3,6 \\ 4,1 & 4,2 & 4,3 & 4,4 & 4,5 & 4,6 \\ 5,1 & 5,2 & 5,3 & 5,4 & 5,5 & 5,6 \\ 6,1 & 6,2 & 6,3 & 6,4 & 6,5 & 6,6 \end{matrix}\right\} \qquad \frac{5}{36} \qquad \frac{15}{36}$$

(احتمال شرطی) $P(A|B) = \frac{P(A|B)}{P(B)} = \frac{\frac{5}{36}}{\frac{15}{36}} = \frac{5}{15}$

قضیه احتمال متوسط و بیس (bayse)

حوادث $H_1+H_2+ \ldots..+H_n$ که دو به دو ناسازگار و بر روی هم مجموعه کامل حوادث را تشکیل می‌دهند (A) هر یک با یک احتمالی در حادثه (A) دخیلند و احتمال موضوع مورد بررسی ما در یکی از آن‌ها اندازه خاصی است که ما در نهایت بخواهیم به طور متوسط احتمال متوسط آن موضوع را در حادثه A بدانیم در واقع احتمال متوسط را خواسته‌ایم و اگر بخواهیم بدانیم این احتمال مورد نظر در رابطه با یکی از آن حوادث چه اندازه است از قضیه بیس استفاده می‌کنیم

H_i	$P(H_i)$	$P(A \mid H_i)$	
H_1	0.3	0.3	0.09
H_2	0.2	0.7	0.17
H_3	0.5	0.9	0.45
	1		0.68

H_i = حادثه چندم

PH_i .مجموعه حوادث که جمع آن همیشه برابر با یک می‌باشد

$(A \mid H_i)$ = حادثه استاندارد بودن احتمال است A

$$\sum(PH_i) \times P(A \mid H_i) = 0.09+0.14+0.45 = 0.68$$

احتمال متوسط $= P(A_i) \times P(A \mid H_i) \rightarrow P(A) = \sum(PH_i) \times P(A \mid H_i)$

بیس $= P(H_i \mid A) = \frac{P(H_i) \times P(A \mid H_i)}{\sum(PH_i) \times P(A \mid H_i)} = \frac{0.45}{0.68} = \frac{45}{68}$

مسئله (۱): کیسه‌ای که محتوی ۴ مهره سفید و ۶ مهره سیاه است، ۲ مهره بدون جای‌گذاری بیرون می‌آوریم مطلوب است احتمال اینکه: $(B=6)$, $(W=4)$

الف) هر دو مهره سفید باشد. $P(W_1 W_2)$

ب) هر دو مهره سیاه باشد. $P(B_1 B_2)$

ج) اولی سفید و دومی سیاه باشد. $P(WB)$

د) یکی سفید و دیگری سیاه باشد. $P(WB+BW)$

پاسخ مسئله (۱):

الف) هر دو سفید باشد.

$$P(W_1 W_2) = P(W_1) \times P(W_2 \mid W_1) \rightarrow \frac{4}{10} \times \frac{3}{9} = \frac{12}{90}$$

ب) هر دو سیاه باشد.

$$P(B_1 B_2) = P(B_1) \times P(B_2 \mid B_1) \rightarrow \frac{6}{10} \times \frac{5}{9} = \frac{30}{90}$$

ج) اولی سفید و دومی سیاه باشد.

$$P(WB) = P(W) + P(B|W) \rightarrow \frac{4}{10} \times \frac{6}{9} = \frac{24}{90}$$

د) یکی سفید و یکی سیاه باشد.

$$P(WB+BW) = P(WB) + P(BW) \rightarrow P(W) \times P(B|W) + P(B) \times P(W|B) \rightarrow \frac{4}{10} \times \frac{6}{9} + \frac{6}{10} \times \frac{4}{9} = \frac{48}{90}$$

مسئله (۲): جامعه‌ای به صورت جدول زیر وجود دارد از این جامعه دو نفر را تصادفی انتخاب می‌کنیم. مطلوب است احتمال این که این دو نفر **مرد یا باسواد** باشند.

جنسیت	باسواد	بی‌سواد	جمع
مرد	15	10	25
زن	12	13	25
جمع	27	23	50

P(مرد+باسواد)=P(مرد) + P(باسواد) - P(مرد و باسواد)

$$\frac{25}{50} + \frac{27}{50} - \frac{15}{50} = \frac{37}{50}$$

مسئله (۳): اگر ۳۰٪ کارکنان بخش حسابداری یک شرکت زن باشد که ۲۰٪ آن‌ها در شرف بازنشستگی هستند و در بین کارکنان مرد همین بخش ۵٪ در شرف بازنشستگی شدن هستند. یکی از کارکنان را به طور تصادفی انتخاب می‌کنیم. مطلوب است احتمال اینکه الف) فرد انتخاب شده در شرف بازنشستگی باشد و ب) مرد در شرف بازنشستگی است احتمال اینکه مرد باشد.

زن $\%30 \rightarrow$ $\%20$ شرف بازنشستگی

مرد $\%70 \rightarrow$ $\%5$ شرف بازنشستگی

H_i	$P(H_i)$	$P(A \mid H_i)$	$P(H_i) \times P(A \mid H_i)$
زن A	0.3	0.2	0.06
مرد B	0.7	0.5	0.035
	1		0.095

الف) فرد انتخاب شده در شرف بازنشستگی باشد $\sum(PH_i) \times P(A \mid H_i) = 0.06+0.035 = 0.095$

احتمال متوسط $= P(A_i) \times P(A \mid H_i) \rightarrow P(A) = \sum(PH_i) \times P(A \mid H_i)$

ب) فرد انتخاب شده مرد باشد.

بیس $= P(H_i \mid A) = \frac{P(H_i) \times P(A \mid H_i)}{\sum(PH_i) \times P(A \mid H_i)} = \frac{0.035}{0.095} = \frac{35}{95}$

مسئله (۴): چهار کیسه ظاهراً یکسان داریم در کیسه اول ۳ مهره سیاه و ۷ مهره سفید داریم. در کیسه دوم ۴ مهره سیاه و ۶ مهره سفید داریم. در کیسه سوم ۵ مهره سیاه و ۵ مهره سفید است و در کیسه چهارم ۲ مهره سیاه و ۸ مهره سفید است یکی از این کیسه‌ها را به طور تصادفی انتخاب می‌کنیم و از داخل این کیسه به طور تصادفی ۲ مهره بیرون می‌آوریم. مطلوب است :

الف) احتمال این که یکی از مهره‌ها سفید و دیگری هم سیاه باشد.

H_i	$P(H_i)$	$P(A \mid H_i)$	$P(H_i) \times P(A \mid H_i)$
B W 3 7 کیسه ۱	$\frac{1}{4}$	$\frac{C_3^1 C_7^1}{C_{10}^2} = \frac{21}{45}$ یا $(\frac{3}{10} \times \frac{7}{9} + \frac{7}{10} \times \frac{3}{9}) = \frac{21}{45}$	$\frac{1}{4} \times \frac{21}{45} = \frac{24}{180}$
4 6 کیسه ۲	$\frac{1}{4}$	$\frac{C_4^1 C_6^1}{C_{10}^2} = \frac{24}{45}$ یا $(\frac{4}{10} \times \frac{6}{9} + \frac{6}{10} \times \frac{4}{9}) = \frac{24}{45}$	$\frac{1}{4} \times \frac{22}{45} = \frac{24}{180}$
5 5 کیسه ۳	$\frac{1}{4}$	$\frac{C_5^1 C_5^1}{C_{10}^2} = \frac{25}{45}$ یا $(\frac{5}{10} \times \frac{5}{9} + \frac{5}{10} \times \frac{5}{9}) = \frac{25}{45}$	$\frac{1}{4} \times \frac{25}{45} = \frac{25}{180}$
2 8 کیسه ۴	$\frac{1}{4}$	$\frac{C_2^1 C_8^1}{C_{10}^2} = \frac{16}{45}$ یا $(\frac{2}{10} \times \frac{8}{9} + \frac{8}{10} \times \frac{2}{9}) = \frac{16}{45}$	$\frac{1}{4} \times \frac{16}{45} = \frac{16}{180}$
	1		$\sum(PH_i) \times P(A \mid H_i) = \frac{86}{180}$

محتمل‌ترین حادثه مربوط به کیسه سوم است چون بیشترین احتمال را دارد. $\frac{25}{180}$

فرمول ترکیب: $C_n^r = \frac{n!}{r!(n-1)!}$

ب) احتمال این که هر دو مهره سفید باشد.

H_i	$P(A \mid H_i)$	$P(H_i) \times P(A \mid H_i)$
B W 3 7 کیسه ۱	$\frac{C_7^2}{C_{10}^2} = \frac{21}{45}$	$\frac{1}{4} \times \frac{21}{45} = \frac{21}{180}$
4 6 کیسه ۲	$\frac{C_6^2}{C_{10}^2} = \frac{15}{45}$	$\frac{1}{4} \times \frac{15}{45} = \frac{15}{180}$
5 5 کیسه ۳	$\frac{C_5^2}{C_{10}^2} = \frac{10}{45}$	$\frac{1}{4} \times \frac{10}{45} = \frac{10}{180}$
2 8 کیسه ۴	$\frac{C_8^2}{C_{10}^2} = \frac{28}{45}$	$\frac{1}{4} \times \frac{28}{45} = \frac{28}{180}$
		$\sum(PH_i) \times P(A \mid H_i) = \frac{74}{180}$

محتمل‌ترین حادثه مربوط به کیسه چهارم است چون بیشترین احتمال را دارد. $\frac{28}{180}$

ج) احتمال اینکه دو مهره سفید و متعلق به کیسه سوم باشد.

بیس $= P(H_3 \mid A) = \frac{P(H_3) \times P(A \mid H_3)}{P(A)} = \frac{\frac{10}{180}}{\frac{74}{180}} = \frac{10}{74}$

مثال:

فضای نمونه (فضای کامل حوادث) برای دانشجو با موضوع باید درس آمار را یاد بگیرد تا آن را پاس کند چیست؟

اگر این شرط برقرار باشد که دانشجو حداکثر ۳ بار می‌تواند یک درس را بگیرد چیست؟ پاس $= P$ افتاده $= N$

فضای نمونه $= \{P, NP, NNP, \dots\}$ اگر حداکثر ۳ بار درس را گیرد پاسخ: $\{P, NP, NNP\}$

مثال:

احتمال‌های زیر را محاسبه کنید:

الف) پیشامد دقیقاً یک پسر در خانواده دو فرزندی؟

ب) پیشامد حداقل یک پسر در خانواده‌های دو فرزندی؟ پسر $= b$ دختر $= g$

ب $= \{bb, bg, gb\}$ الف $= \{bg, gb\}$

مثال:

در امتحانی ۶ سؤال داده شده است که جواب آن‌ها تعیین درستی یا نادرستی آن است اگر دانشجویی به طور تصادفی جلوی سؤال‌ها علامت بگذارد مطلوب است احتمال اینکه:

الف) در مجموع جواب‌های دانشجو حداقل یک علامت درست باشد.

ب) در مجموع جواب دانشجو حداقل ۵ علامت درست باشد.

الف) $= \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{64}$

ب) $= \frac{7}{64}$ ⟶

$\frac{1}{64}$	$\frac{1}{64}$	$\frac{1}{64}$	$\frac{1}{64}$	$\frac{1}{64}$	$\frac{1}{64}$	$\frac{1}{64}$	$= \frac{7}{64}$
NPPPPP	PNPPPP	PPNPPP	PPPNPP	PPPPNP	PPPPPN	PPPPPP	

مثال:

جدول توزیع قابلیت دانشجویان در درس آمار و روش تحقیق به شرح زیر می‌باشد. اگر دانشجویی به تصادف انتخاب شود مطلوب است محاسبه:

الف) مقدار K

ب) احتمال اینکه حداقل یک درس دانشجو ضعیف باشد

ج) احتمال این که دانشجو در درس آمار ضعیف نباشد.

د) دانشجو در درس آمار ضعیف است احتمال اینکه در درس روش تحقیق نیز ضعیف باشد.

روش تحقیق / آمار	B قوی	B متوسط	B ضعیف
A قوی	0.18	0.12	0.16
A متوسط	0.1	0.12	0.04
A ضعیف	0.15	0.04	K = 0.09
	0.43	0.28	1

الف) $\sum P(A_i) = 1 \rightarrow 0.018+0.12+0.16+\ldots+k=1 \rightarrow k = 0.09$

ب) P (روش تحقیق ضعیف + آمار ضعیف) = P(آمار ضعیف) + P(روش تحقیق ضعیف) – (آمار و روش تحقیق ضعیف)

$P(A+B)= P(A) + P(B) – P(AB) = 0.28 + 0.29 – 0.09 = 0.48$

ج) $P(\overline{A_3}) = 1- P(A_3) = 1 - 0.28 = 0.27$

د) $P(B|A) = \frac{P(AB)}{P(A)} = \frac{0.09}{0.28} = \frac{9}{28}$

متغیرهای تصادفی

متغیر تصادفی بر اساس مورد مطلوب اندازه گیری می‌شود.

پیشامد مکعب:

x_i	1	2	3	4	5	6	
p_i	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	=1

فرض کنید دو سکه را پرتاب کنیم و بگوییم پیشامد مورد نظر ما آمدن حداقل یک بار رویه خط باشد.

حل: وقتی دو سکه را پرتاب می‌کنیم پیشامدهای زیر پیش می‌آید:

HH	HT	TH	TT
0	1	1	2

احتمال این که حداقل یک بار خط بیاید:

x_i	0	1	2	
p_i	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$	=1

توزیع دوجمله‌ای

یک آزمایش را n بار انجام می‌دهیم و هر آزمایش دارای احتمال ثابت است همچنین آزمایش‌ها مستقل از هم باشند حال اگر این آزمایش را n بار انجام دهیم احتمال این که x بار حالت مورد نظر ما اتفاق بیافتد توسط فرمول زیر که فرمول توزیع دوجمله‌ای تعیین می‌شود:

$$P(X = x) = C_n^x\, p^x\, q^{n-x}$$

$q = 1 - P$ حالت غیر مورد نظر , $X =$ متغیر تصادفی , $x = 0,1,2,\dots n$

مثال:

برای خانواده‌های سه فرزندی احتمال اینکه از سه فرزند یک فرزند پسر باشد:

$n = 3 \quad , x = 1 \quad , \quad$ (فرزند پسر)$p = \frac{1}{2}$

$$q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2} \quad \rightarrow C_3^1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{3-1} = 3 \times \frac{1}{2} \times \frac{1}{4} = \frac{3}{8}$$

مثال:

اگر احتمال اصابت تیر به هدف 0.3 باشد احتمال اینکه از ۶ بار تیراندازی ۲ بار به هدف اصابت کند چقدر است:

$n = 6 \quad , x = 2 \quad , \quad p = 0.3$

$$q = 1 - p = 1 - 0.3 = 0.7 \quad \rightarrow p(X = x) = C_6^2\, (0.3)^2\, (0.7)^{6-2}$$

مثال:

۲۰٪ محصولات یک کارخانه غیر استاندارد است دو محصول را به طور تصادفی انتخاب می‌کنیم. مطلوب است:

الف: احتمال این که هر دو محصول غیر استاندارد باشد.

ب: حداقل یکی از آن‌ها غیر استاندارد باشد.

پاسخ:

الف:

$P = 20\% \rightarrow 0.2$ احتمال موجود

$n = 2$ تعداد محصول انتخابی

$x = 2$ دو محصول غیر استاندارد

$q = 1 - p$

فرمول: $P(X = x) = C_n^x\, p^x\, q^{n-x} \qquad , \qquad C_n^r = \frac{n!}{r!(n-r)!} \rightarrow C_2^2 = \frac{2!}{2!0!} = 1$

$$p(X = 2) = C_2^2\, (0.2)^2\, (1-0.2)^{2-2} \rightarrow p(X = 2) = 0.04 \times 1 = 0.04$$

ب:

$$p(X \geq 0) = 1 - p(X = 0) \rightarrow 1 - C_2^0\, (0.2)^0\, (1-0.2)^{2-0}$$
$$= 1 - (0.8)^2 = 1 - 0.64 = 0.36$$

مثال:

احتمال این که رادیوی تولید یک شرکت خراب باشد 0.4 است. فردی ۵ رادیو از این فروشگاه می‌خرد اگر کمیت تصادفی x عبارت باشد از تعداد رادیوهای خراب، مطلوب است:

الف: تابع توزیع احتمال کمیت تصادفی x.

ب: انتظار دارید بین این رادیوهای خریداری شده چند رادیو خراب باشد. (امید ریاضی)

پاسخ:

الف:

$px = x$

$P = 0.4$ احتمال موجود

$n = 5$ تعداد رادیو خریداری شده

$x = x$ احتمال کمیت تصادفی

$q = 1 - p$

فرمول: $P(X = x) = C_n^x\, p^x q^{n-x}$ (وقتی تابع توزیع را خواستند از این فرمول استفاده می کنیم)

$q = 1 - p = 1 - 0.4 = 0.6$

$p(X = x) = C_5^x\, (0.4)^x\, (1-0.4)^{5-x} \rightarrow x = 1 , 2 , 3 , 4 , 5$

ب:

امید ریاضی $\leftarrow E(X) = np$ متوسط متوسط‌هایی است که مقادیر کمیت تصادفی می‌تواند بگیرد. (مقادیر متوسط).

میانگین در احتمال $\leftarrow E(X) = \Sigma p_i x_i$ – واریانس $D(X) = npq$

امید ریاضی توزیع دوجمله ای: $E(X) = 5 \times 0.4 = 2$

واریانس توزیع دو جمله‌ای: $D(X) = 5 \times 0.4 \times 0.6 = 1.2$

توزیع فوق هندسی (هیپرژئومتریک)

در توزیع فوق هندسی احتمال ثابت نداریم از یک آزمایش به آزمایش دیگر احتمال آن فرق می‌کند. ولی در توزیع پواسن و توزیع دوجمله ای احتمال ثابت وجود دارد.

فرمول: $\frac{C_m^x C_{N-m}^{n-x}}{C_N^n}$

مثال:

۱۰ عدد ماشین وجود دارد که ۴ تای آن سالم است. ۲ ماشین به طور تصادفی انتخاب می‌کنیم احتمال اینکه:

الف: یک ماشین سالم در گروه انتخابی باشد چقدر است.

ب: احتمال اینکه حداقل یک ماشین سالم باشد چقدر است.

ج: محاسبه امید ریاضی

پاسخ:

الف:

N = 10

x = 2

$\text{فرمول}: \frac{C_m^x C_{N-m}^{n-x}}{C_N^n} \rightarrow P(X=1) = \frac{C_4^1 C_6^1}{C_{10}^2} = \frac{24}{25}$

ب:

$P(X \geq 0) = 1 - P(X=0) = 1 - \frac{C_4^0 C_6^2}{C_{10}^2} = 1 - \frac{15}{45} = \frac{30}{45}$

ج:

$E(X) = np \rightarrow E(X) = 2 \times \frac{4}{10} = 1.6$

مثال:

از یک گروه ۱۵ نفری شامل ۷ مرد و ۸ زن می‌خواهیم ۴ نفر را به عنوان نماینده انتخاب کنیم. مطلوب است احتمال اینکه در گروه انتخابی:

الف: دو زن موجود باشد.

ب: حداقل یک زن موجود باشد.

ج: حداقل یک مرد موجود باشد.

پاسخ:

الف:

N = 15

مرد: 7 , زن: 8

x = 2 , n = 4

$\text{فرمول}: \frac{C_m^x C_{N-m}^{n-x}}{C_N^n} \rightarrow P(X=2) = \frac{C_8^2 C_7^2}{C_{15}^4} = \frac{\frac{8!}{216!} \times \frac{7!}{2!5!}}{\frac{15!}{4!11!}} = \frac{28 \times 21}{3 \times 13 \times 7 \times 5}$

ب:

$P(X \geq 0) \rightarrow 1 - p(X = 0_{\text{زن}}) = 1 - \frac{C_8^0 C_7^4}{C_{15}^4} = 1 - \frac{1-35}{3 \times 13 \times 7 \times 5} = \frac{38}{39}$

ج:

$P(X \geq 0) \rightarrow 1 - p(X = 0_{\text{مرد}}) = 1 - \frac{C_7^0 C_8^4}{C_{15}^4}$

مثال:

از کارتنی محتوی ۷ توپ قرمز، ۹ توپ آبی و ۶ توپ سبز یک نمونه سه تایی انتخاب می‌کنیم. مطلوب است محاسبه احتمال اینکه از هر رنگ موجود باشد.

پاسخ:

$(7+9+6=22) \rightarrow P(X=1) = \frac{C_7^1 C_9^1 C_6^1}{C_{22}^4}$

مثال:

از گروهی متشکل از ۲۵ پزشک، ۱۵ پرستار و ۳۵ بهیار یک کمیته ۶ نفره به تصادف انتخاب می‌شود. مطلوب است محاسبه احتمال این که در این کمیته ۲ پزشک، ۳ پرستار و یک بهیار شرکت داشته باشد.

پاسخ:

$$(25+15+35=75)\ ,\ (2+3+1=6) \rightarrow P\ (X=6) = \frac{C_{25}^{2} C_{15}^{3} C_{35}^{1}}{C_{75}^{6}}$$

توزیع پواسن

برای حوادثی که احتمال آن خیلی کوچک است حوادث نادر مانند این که از توپ‌های پارچه خریداری شده از یک کارخانه ریسندگی تعداد توپ‌های پارچه که زدگی دارند خیلی کم است که در واقع برای حوادثی که احتمال وقوع آن نادر و یا خیلی کم باشد به کار می‌رود.

فرمول: $P\ (X = x) = \frac{\lambda^{x} e^{-\lambda}}{x} \qquad (\lambda = np)$

مثال:

احتمال اینکه کالایی معیوب باشد 0.0002 است. مطلوب است احتمال آن که در هر ۵۰۰۰ کالای ارسال به گمرک:

الف: همگی سالم باشد.

ب: ۲ کالا معیوب باشد.

ج: بیش از ۲ کالا معیوب باشد.

پاسخ:

الف: همگی سالم باشد (❈ این اعداد از جدول پواسن استخراج گردیده است)

$P = 0.0002\ \ ,\ n = 5000$

$\lambda = np \Longrightarrow 5000 \times 0.0002 = 1$

$p\ (X=0) = 0/3679$❈

ب: ۲ کالا معیوب باشد (❈ این اعداد از جدول پواسن استخراج گردیده است)

$p\ (X=2) = (x= 1\ ,\ \lambda = 1) = 0/7358$❈

$(x= 2\ ,\ \lambda = 1) = 0/6767$❈

$p(X=2) = 0.7358 - 0.6767 = 0.0591$

ج: بیش از ۲ کالا معیوب باشد (❈ این اعداد از جدول پواسن استخراج گردیده است)

$p\ (X>2) = 1 - p(X<2) = 1 - 0.6767$❈ $= 0/3233$

مثال:

از هر ۱۰۰۰ نفر یک نفر به نوعی بیماری مبتلا می‌شوند اگر ۱۵۰۰ نفر به طور تصادفی انتخاب کنیم احتمال اینکه ۴ نفر مبتلا به این نوع بیماری باشد چقدر است؟

پاسخ (❈ این اعداد از جدول پواسن استخراج گردیده است)

$P = 0.0001\ \ ,\ n = 1500$

$\lambda = np \Longrightarrow 1500 \times 0.0001 = 1.5$

$p\ (X=2) = (x= 4\ ,\ \lambda = 1.5) = 0/9814$❈

$(x= 3 , \lambda = 1.5) = 0/9344$*

$p(X=2) = 0.9814 – 0.9344 = 0.0470$

$p (X=4) = 0/0470$*

مثال:

احتمال تصادف منجر به مرگ در یک شهر ۳۰۰۰۰ نفری در هر روز 0.0001 است مطلوب است محاسبه احتمال این که در یک روز معین:

الف: تعداد تصادف منجر به مرگ صفر باشد.

ب: تعداد تصادف منجر به مرگ بیش از ۲ نفر باشد.

پاسخ:

الف: تعداد تصادف منجر به مرگ صفر باشد. (* این اعداد از جدول پواسن استخراج گردیده است)

$P = 0.0001 \ , n = 30000$

$\lambda = np \Longrightarrow 30000 \times 0.0001 = 3$

$p (X=0) = 0/0498$*

ب: تعداد تصادف منجر به مرگ بیش از ۲ نفر باشد (* این اعداد از جدول پواسن استخراج گردیده است)

$p (X>2) = 1- p(X<2) = 1- 0.4232$* $= 0/5768$

مثال:

تعداد تصادف‌های اتومبیل در یک چهارراه ماهانه به طور متوسط ۴ تصادف می‌باشد مطلوب است احتمال اینکه:

الف: در عرض یک ماه تصادف نشود.

ب: حداقل یک بار تصادف شود.

ج: سه بار تصادف شود.

پاسخ:

الف: در عرض یک ماه تصادف نشود. (* این اعداد از جدول پواسن استخراج گردیده است)

$\lambda = 4$

$p (X=0) = 0/018316$*

ب: حداقل یک بار تصادف شود (* این اعداد از جدول پواسن استخراج گردیده است)

$p (X\geq1) = 1- p(X=0) = 1- 0.018316$* $= 0/981684$

ج: سه بار تصادف شود. (* این اعداد از جدول پواسن استخراج گردیده است)

$\lambda = 4$

$p (X=3) = 0/4335$* $- 0.2381$* $= 0.1954$

روش استفاده از جدول پواسن:

	$\lambda = 3.8$	$\lambda = 3.9$	$\lambda = 4$	$\lambda = 4.1$	$\lambda = 4.2$
0			0.018316		
1			0.0916		
2			0.2381		
3			0.4335		

مثال:

در یک کتاب ۴۰۰ صفحه‌ای ۴۰ غلط چاپی وجود دارد. مطلوب است احتمال اینکه در ۱۰ صفحه:

الف: یک غلط موجود باشد.

ب: بیش از ۲ غلط وجود داشته باشد.

ج: کمتر از یک غلط وجود داشته باشد.

د: احتمال اینکه بین ۲ تا ۴ غلط وجود داشته باشد.

	λ = 0.8	λ = 0.9	λ = 1	λ = 1.1	λ = 1.2
0			0.3679		
1			0.7358		
2			0.9197		
3			0.9810		
4			0.9963		

پاسخ:

الف: یک غلط موجود باشد. (❉ این اعداد از جدول پواسن استخراج گردیده است)

$\frac{400}{10} = \frac{40}{x} \rightarrow x = 1 \quad \rightarrow \lambda = 1$

$p(X < 1) = p(X = 0) = 0.7358^{❉} - 0.3679^{❉} = 0/3679$

ب: بیش از ۲ غلط وجود داشته باشد (❉ این اعداد از جدول پواسن استخراج گردیده است)

$p(X > 2) = 1 - p(X < 2) = 1 - 0.9197^{❉} = 0/0903$

ج: کمتر از یک غلط وجود داشته باشد. (❉ این اعداد از جدول پواسن استخراج گردیده است)

$p(X < 1) = 0.7358^{❉}$

د: احتمال اینکه بین ۲ تا ۴ غلط وجود داشته باشد د (❉ این اعداد از جدول پواسن استخراج گردیده است)

$p(2 \leq X \leq 4) = 0.9963^{❉} - 0/9197^{❉}$

مثال:

به طور متوسط بین ساعت ۸ تا ۹ صبح ۴ تلفن به یک دفتر زده می‌شود. مطلوب است احتمال اینکه:

الف: در این مورد هیچ تماس تلفنی گرفته نشود.

ب: کمتر از ۳ تماس گرفته شود.

پاسخ:

الف: در این مورد هیچ تماس تلفنی گرفته نشود. (❉ این اعداد از جدول پواسن استخراج گردیده است)

$\lambda = 4$

$p(X = 0) = 0.0183^{❉}$

ب: کمتر از ۳ تماس گرفته شود (❉ این اعداد از جدول پواسن استخراج گردیده است)

$p(X < 3) = 0.4335^{❉}$

مثال (نمونه سؤال امتحانی):

لوازم مورد نیاز یک موسسه توسط ۳ کارخانه ساخته می‌شود به طوری که کارخانه اول ۳۰٪، کارخانه دوم ۴۰٪ و بقیه را کارخانه سوم تأمین می‌نماید. بر اساس آمارهای گذشته مشخص گردید که ۲٪ محصولات کارخانه اول، ۳٪ کارخانه دوم و ۴٪ محصولات کارخانه سوم معیوب است از کالای خریداری شده یک محصول را به طور تصادفی انتخاب می‌کنیم مطلوب است احتمال اینکه:

الف: احتمال این که محصول انتخاب شده معیوب باشد.

ب: احتمال اینکه محصول انتخاب شده سالم باشد.

ج: محصول انتخاب شده معیوب است احتمال این که محصول توسط کارخانه دوم تولید شده باشد.

پاسخ:

الف: 0.03

H_i	$P(H_i)$	$P(A \mid H_i)$	
H_1	0.3	0.02	0.006
H_2	0.4	0.03	0.012
H_3	0.3	0.04	0.012
	1		0.03

$$\sum(PH_i) \times P(A \mid H_i) = 0.006+0.012+0.012 = 0.03$$

ب:

$$P(\bar{A}) = 1- P(A) = 1 - 0.03 = 0.97$$

ج:

$$\text{بیس} = P(H_2 \mid A)= \frac{P(H_i)\times P(A \mid H_i)}{\sum(PH_i) \times P(A \mid H_i)} = \frac{0.012}{0.03} = \frac{12}{30}$$

مثال (نمونه سؤال امتحانی):

از ۱۵ عدد لامپ که ۵ عدد آن معیوب است به طور تصادفی ۳ لامپ را انتخاب می‌کنیم. مطلوب است:

الف: احتمال این که هر سه لامپ سالم باشد.

ب: احتمال اینکه حداقل یک لامپ سالم باشد...

پاسخ:

الف: احتمال این که هر سه لامپ سالم باشد

$$P\ (\text{معیوب})\ \frac{5}{15} = \frac{1}{3}$$

$$P\ (\text{سالم})\ 1 - \left(\text{احتمال معیوب}\right) = 1 - \frac{1}{3} = \frac{2}{3}$$

$$P\ (X = x) = C_n^x\ p^x\ q^{n-x}$$

$$p(X = 3) = C_3^3\ (\tfrac{2}{3})^3\ (\tfrac{1}{3})^{3-3} = 1 \times \frac{8}{27} \times 1 = \frac{8}{27} \quad \text{احتمال ۳ لامپ سالم}$$

ب: احتمال اینکه حداقل یک لامپ سالم باشد.

$$P\ (\text{حداقل یک سالم})\ 1 - \left(\text{سالم}\ X = 0\right)$$

$$p(X = 0) = C_3^0\ (\tfrac{2}{3})^0\ (\tfrac{1}{3})^{3-0} = 1- 1\times 1 \times \frac{1}{27} = 1 - \frac{1}{27} = \frac{26}{27} \quad \text{احتمال حداقل یک لامپ سالم}$$

مثال (نمونه سؤال امتحانی):

احتمال اینکه محصول ضمن تولید معیوب گردد 0.015 است از محصولات تولید شده تعداد ۲۰۰ واحد محصول را به طور تصادفی انتخاب می‌کنیم. مطلوب است:

الف: احتمال این که ۴ واحد محصول معیوب باشد.

ب: احتمال اینکه حداکثر ۴ واحد محصول معیوب باشد.

پاسخ:

الف: احتمال این که ۴ واحد محصول معیوب باشد. (❋ این اعداد از جدول پواسن استخراج گردیده است)

$P = 0.015 \quad , n = 200$

$\lambda = np \Longrightarrow 200 \times 0.015 = 3$

$p\ (X=4) = 0/8153❋ - 0/6472❋ = 0.1681$

ب: احتمال اینکه حداکثر ۴ واحد محصول معیوب باشد (❋ این اعداد از جدول پواسن استخراج گردیده است)

$p(X<4) = \ 0.8153❋$

مثال (نمونه سؤال امتحانی):

از یک مجموعه از محصولات تولید شده در یک کارخانه به حجم ۵۰۰ واحد که تعداد ۴۰ واحد آن غیر استاندارد می‌باشد به طور تصادفی ۲۰ واحد محصول را انتخاب می‌کنیم. کمیت تصادفی X عبارت است از تعداد محصولات غیر استاندارد بین ۲۰ محصول انتخاب شده. مطلوب است:

الف: قانون توزیع کمیت تصادفی X.

ب: امید ریاضی کمیت تصادفی X.

پاسخ:

الف:

$N = 500$

$n = 40$

$P\ (X = x) = \frac{C_m^x C_{N-m}^{n-x}}{C_N^n} \rightarrow\ P\ (X = x) = \frac{C_{20}^x C_{500-x}^{40-x}}{C_{500}^{40}} \ = x = 0,1,2,3,\ldots 20$

ب:

$E(X) = np \rightarrow\ p = \frac{40}{500} = 0.08 \quad \rightarrow 20 \times 0.08 = 1.6$

مثال (نمونه سؤال امتحانی):

یک قنادی خامه مصرفی خود را از ۳ فروشگاه تأمین می‌کند و روزانه معمولاً ۱۱۰ کیسه خامه از این سه فروشگاه به ترتیب ۳۰ کیلو از فروشگاه اول، ۴۵ کیلو از فروشگاه دوم و بقیه را از فروشگاه سوم خریداری می‌کند و آن‌ها را بدون مخلوط کردن در پخت شیرینی‌ها استفاده می‌کند. احتمال فاسد بودن خامه از این سه فروشگاه به ترتیب ۲٪، ۳٪ و ۵٪ است. اگر در یک روز خاص یک قطعه شیرینی خامه‌ای را جهت مصرف به تصادف انتخاب کنیم:

الف: چقدر احتمال دارد که شیرینی انتخابی فاسد باشد.

ب: چقدر احتمال دارد شیرینی انتخابی سالم باشد.

ج: اگر شیرینی فاسد باشد چقدر احتمال دارد که خامه آن از فروشگاه سوم تأمین شده باشد.

مثال (نمونه سؤال امتحانی):

در جعبه‌ای ۴ مهره سیاه و ۶ مهره سفید قرار دارد از این جعبه ۴ مهره به تصادف یکی پس از دیگری انتخاب می‌کنیم. مطلوب است احتمال اینکه:

الف: هر دو مهره سیاه باشد.

ب: یکی سیاه و دیگری سفید باشد.

مثال (نمونه سؤال امتحانی):

یکی از چهارراه‌های شهر تهران روزانه ۱۰۰۰ اتومبیل رفت و آمد می‌کند بر اساس آمارهای گذشته احتمال اینکه حادثه‌ای خطرناک در این چهارراه اتفاق بی افتد 0.003 است مطلوب است احتمال اینکه:

الف: حداکثر ۳ تصادف خطرناک در این چهارراه اتفاق بی افتد.

ب: فقط ۲ تصادف خطرناک در این چهارراه رخ دهد.

استنتاج آماری

استنتاج آماری فرآیند ذهنی است که در نتیجه گیری یا قضاوت نسبت به خصوصیات تمامی اعضای یک جامعه آماری بر اساس مطالعه نتایج به دست آمده از مشاهده بر روی تعداد معدودی از اعضای آن (نمونه تصادفی که زیر جامعه‌ای از جامعه اصلی است) با رعایت شرط تصادفی بودن انتخاب اعضا صورت می‌گیرد.

انتخاب تصادفی، انتخابی را گویند که تمامی اعضای جامعه امکان معین و یکسانی را برای انتخاب شدن داشته باشند. انتخاب تصادفی به دو روش صورت می‌گیرد: یکی بدون جایگزینی (بدون بازگردان) و دیگر، با جایگزینی، چون استقلال آزمایش‌ها یکی از ویژگی‌های مهم در انتخاب تصادفی به طریقه بازگردان است، به همین دلیل این روش در نظریه نمونه نقش بسیار مهمی دارد. اگر حجم جامعه اصلی نسبت به حجم نمونه خیلی بزرگ باشد. به طوری که N<۰/۰۵ باشد نمونه‌گیری بدون جایگزینی تفاوتی با روش بازگردان نخواهد داشت.

استنتاج آماری از دو بخش تشکیل شده است:

۱. تخمین زدن

۲. آزمون فرضیه

در بسیاری از زمینه‌های کاربردی محققان در صدد تعیین پارامترها مورد نظر در جامعه آماری هستند که دسترسی به آن‌ها به طور مستقیم و با روش سرشماری از جامعه آماری امکان پذیر نیست. در چنین موقعیت‌هایی محققان ناچارند به نمونه‌هایی از جوامع آماری برای دستیابی به پارامترهای مورد نظر اکتفا کنند. در واقع در علم آمار چگونگی قضاوت و نتیجه‌گیری نسبت به قانون توزیع متغیر تصادفی با استفاده از نتایج به دست آمده از نمونه تصادفی، استنباط آماری نامیده می‌شود.

نمونه گیری و توزیع‌های نمونه گیری

همان طور که قبلاً اشاره شد، در بسیاری از موارد به جای اینکه اطلاعات مورد نظر خود را از جامعه اصلی به دست آوریم از یک نمونه تصادفی n تایی استفاده می‌کنیم که تعدادی از اعضای جامعه آماری به حجم N می‌باشند که به این عمل نمونه گیری می‌گویند و دلایل نمونه گیری عبارتند از:

۱. دسترسی به همه اعضای جامعه آماری همواره ممکن نیست.
۲. سرشماری هزینه‌ی زیادی در بردارد.
۳. سرشماری نیاز به زمان زیادی دارد.
۴. در برخی موارد سرشماری موجب دگرگونی (تغییر) و یا حتی از بین رفتن جامعه آماری می‌گردد.

نمونه‌گیری به روش‌های مختلفی انجام می‌پذیرد.

۱. نمونه‌گیری تصادفی ساده
۲. نمونه‌گیری سیستماتیک
۳. نمونه‌گیری طبقه بندی شده یا گروه‌بندی شده
۴. نمونه‌گیری خوشه‌ای

با توجه به اهداف تحقیق به یکی از روش‌های فوق نمونه مورد نظر را به حجم n انتخاب می‌کنیم. البته در درس‌های بعد چگونگی تعیین n را نیز می‌آموزیم.

اگر شاخص به دست آمده از جامعه آماری با روش سرشماری را پارامتر بنامیم، در نمونه‌گیری، شاخص به دست آماده از یک نمونه n تایی از جامعه‌ای که N عضو دارد را آماره یا تابع نمونه‌ای می‌نامیم.

یکی از جنبه‌های مهم آمار استنباطی تعمیم آماره به پارامتر است. اما از آنجا که اندازه‌های نمونه، از نمونه‌ای به نمونه دیگر در تغییر است، لذا مقدار یک آماره از نمونه‌ای به نمونه دیگر تغییر خواهد کرد؛ و برای دستیابی به استنباط درست و پایا از نمونه بر اساس آماره باید به تعیین توزیع آن آماره قادر باشیم یعنی بتوان تابع احتمالی که از نمونه‌گیری مکرر حاصل می‌شود و در شکل کامل خود آن را توزیع نمونه گیری می‌گوییم را تعیین کنیم. هم‌چنین باید بدانیم تعمیم نتایج بدست آمده از نمونه به جامعه همواره با مقداری خطا همراه است که به هر صورت این خطا اجتناب ناپذیر است ولی هرچه حجم نمونه به حجم جامعه نزدیک‌تر باشد خطای ممکن به حداقل می‌رسد و نتیجه تعمیم نیز معتبرتر است.

شناخت توزیع نمونه‌گیری باعث شناخت و تعیین بهترین تخمین پارامتر از میان آماره‌های مختلف می‌گردد.
صرف نظر از اینکه استفاده از چه روش آمار استنباطی مورد نظر است، قدرت آن به روش به کار گرفته شده برای انتخاب نمونه بستگی دارد. در صورتی که نمونه نماینده واقعی جامعه نباشد (نمونه دارای اریب باشد) پیش بینی صحیح و دقیق درباره‌ی پارامترهای جامعه امکان‌پذیر نیست.

اریب نمونه‌گیری را می‌توان با به کار بردن روش‌های نمونه‌گیری صحیح و مناسب و در نظر گرفته مشخصات عناصر جامعه کاهش داد.

نمونه تصادفی n متغیر مشاهده شده از یک کمیت تصادفی است که در n آزمایش مستقل از هم به دست آمده است. ما به مفهوم نمونه باید با دو دید بنگریم. اگر بخواهیم از نتایج به دست آمده فقط خود زیر جامعه را توصیف کنیم به نمونه با عینک مجموعه اعداد نگاه می‌کنیم. ولی هرگاه بخواهیم از این اعداد نتیجه گیری کنیم و به جامعه اصلی تعمیم دهیم باید به جای اعداد به آن‌ها با عینک متغیر نگاه کنیم.

قبلاً گفتیم یکی از روش‌های نمونه‌گیری، نمونه‌گیری تصادفی ساده است. در این روش شانس انتخاب هر یک از عناصر جامعه آماری در نمونه nتایی یکسان است.

برای انتخاب یک نمونه n تایی از یک جامعه N عضوی $C_{\mathrm{N}}^{\mathrm{n}}$ نمونه ممکن وجود دارد؛ و با این توصیف شانس انتخاب هر یک از نمونه‌های nتایی عبارت است از: $\frac{1}{C_{\mathrm{N}}^{\mathrm{n}}}$ برای درک بهتر موضوع از یک مثال ساده استفاده می‌کنیم.

فرض کنید جامعه آماری مورد بررسی پنج عضو به قرار زیر داشته باشد.

X_1: 6, 9, 12, 15, 18

در این جامعه پنج عضوی میانگین، میانه و واریانس را محاسبه می‌کنیم.

$$\text{میانگین } \mu = \frac{\sum x_i}{N} = \frac{6+9+12+15+18}{5} = 12 \text{ و میانه } M_e=12$$

$$6^2 = \frac{\sum (X_i-\mu)^2}{N} = \frac{(6-12)^2+(9-12)^2+(12-12)^2+(15-12)^2+(18-12)^2}{5} = \frac{90}{5}$$

$$6^2 = 18$$

حال نمونه‌هایی از این جامعه که ۳ عضو داشته باشند انتخاب می‌کنیم. تعداد نمونه‌های ۳ عضوی ممکن عبارت است از:

$$C5^3 = \frac{5!}{3!2!} = 10$$

۱۰ نمونه ممکن را به ترتیب می‌نویسیم و برای هر یک مقادیر $\bar{X}$ (میانگین) و Me (میانه) و S^2 (تخمین واریانس) را محاسبه می‌کنیم.

شماره نمونه	مقادیر نمونه	$\bar{X} = \frac{\sum X_i}{n}$	Me	$S^2 = \frac{\sum (X_i-\bar{X})^2}{n-1}$
۱	۱۲و۹و۶	۹	۹	۹
۲	۱۵و۹و۶	۱۰	۹	۲۱
۳	۱۸و۹و۶	۱۱	۹	۳۹
۴	۱۵و۱۲و۶	۱۱	۱۲	۲۱
۵	۱۸و۱۲و۶	۱۲	۱۲	۳۶
۶	۱۸و۱۵و۱۲	۱۳	۱۵	۳۹
۷	۱۵و۱۲و۹	۱۲	۱۲	۹
۸	۱۸و۱۲و۹	۱۳	۱۲	۲۱
۹	۱۸و۱۵و۹	۱۴	۱۵	۲۱
۱۰	۱۸و۱۵و۱۲	۱۵	۱۵	۹

تابع احتمال $\bar{X}$ و Me را تعیین می‌کنیم و سپس امید ریاضی آن‌ها را محاسبه می‌نماییم.

	۹	۱۰	۱۱	۱۲	۱۳	۱۴	۱۵	
$P(\bar{X})$	۰/۱	۰/۱	۰/۲	۰/۲	۰/۲	۰/۱	۰/۱	$\Sigma P(\bar{X})=1$
$P(\bar{X})\times(\bar{X})$	۰/۹	۱	۲/۲	۲/۴	۲/۶	۱/۴	۱/۵	$\Sigma P(\bar{X})\times\bar{X} = E(\bar{X})=12$

Me	۹	۱۲	۱۵	
P (Me)	۰/۳	۰/۴	۰/۳	$\Sigma P(\text{Me})=1$
$P(\text{Me})\times(\text{Me})$	۲/۷	۴/۸	۴/۵	$\Sigma P(\text{Me})\times\text{Me} = E(\bar{X})=12$

همان طور که ملاحظه می‌کنید امید ریاضی میانگین و امید ریاضی میانه با میانگین اصلی جامعه برابر شد. اما اگر به انتخاب یکی از آماره‌های $\bar{X}$ و Me برای تخمین پارامتر μ یا Me علاقه‌مند باشیم کدام یک از آن‌ها را بهتر است انتخاب کنیم.

یکی از روش‌ها این است که واریانس آن‌ها را محاسبه کنیم. زیرا از دو تخمین زن، تخمین زنی که واریانس کمتری دارد با ارزش‌تر است. بنابراین $D(Me), D(\bar{X})$ را محاسبه می‌کنیم:

$D(\bar{X}) = P(\bar{X}) \times \bar{X}^2 - [E(\bar{X})]^2$

$D(\bar{X}) = (9^2 \times 0/1 + 10^2 \times 0/1 + 11^2 \times 0/2 + 13^2 \times 0/2 + 14^2 \times 0/1 + 15^2 \times) - 12^2$

$D(\bar{X}) = 147 - 144 = 3$

$D(\text{Me}) = P(\text{Me}) \times \text{Me}^2 - [E(\text{Me})]^2$

$D(\text{Me}) = (9^2 \times 0/3 + 12^2 \times 0/4 + 15^2 \times 0/3) - 12^2$

$D(\text{Me}) = 149/4 - 144 = 5/4$

$$\left.\begin{array}{l} D(\bar{X}) = 3 \\ D(\text{Me}) = 5/4 \end{array}\right\} D(\bar{X}) = < D(\text{Me})$$

یعنی پراکندگی توزیع میانگین $(\bar{X})$ کم تر از پراکندگی توزیع میانه (Me) است بنابراین آماره $(\bar{X})$ دارای ویژگی مطلوب‌تری نسبت به (Me) است.

آیا S^2 خواص مطلوب را به عنوان یک تخمین زن برای واریانس (6^2) دارد؟

برای جواب به این سؤال، جدول توزیع نمونه گیری S^2 را می‌نویسیم؛ و امید ریاضی S^2 $(E\ S^2)$ را به دست می‌آوریم)

S^2	9	21	36	39	
$P(S^2)$	0/3	0/4	0/1	0/2	$\Sigma P(S^2) = 1$
$P(S^2) \times S^2$	2/7	8/4	3/6	7/8	$\Sigma S^2 \times P(S^2) = E(S^2) = 22/5$

ملاحظه می‌شود که امید ریاضی S^2 برابر ۲۲/۵ گردید در حالی که واریانس جامعه ۱۸ بود یعنی برابر نیستند. این اختلاف ناشی از نمونه گیری از جامعه محدود است که باید آن را با ضریب $\frac{N}{N-1}$ اصلاح کرد یعنی:

$$E(S^2) = \frac{N}{N-1}6^2$$

$$\frac{22}{5} = \frac{5}{5-1} \times 18$$

در واقع در جامعه‌های آماری بزرگ یا نامحدود $\left(\frac{n}{N} < 0/50\right)$ مقدار N/N-1 به سمت یک میل می‌کند و در نتیجه $E(S^2)=6^2$ خواهد بود.

تئوری ارزیابی یا تخمین

تخمین زدن یک نوع روش قضاوت است نسبت به مقادیر پارامترهای جامعه اصلی بر اساس یک نمونه تصادفی n تایی.

پارامترهای یک جامعه را به دو روش می‌توان ارزیابی کرد یا تخمین زد: یکی از تخمین نقطه‌ای و دیگری تخمین فاصله‌ای (برآورد فاصله‌ای).

در تخمین نقطه‌ای تابعی مناسب از متغیرهای نمونه انتخاب می‌گردد و مقدار عددی آن به عنوان نزدیک‌ترین مقدار به پارامتر جامعه قبول می‌شود ولی در تخمین فاصله‌ای بر اساس متغیرهای نمونه. دو حد از یک فاصله تشکیل داده می‌شود که این فاصله با احتمال معینی مقدار پارامتر را در بر می‌گیرد.

پارامترهای قانون توزیع را با θ نمایش می‌دهند. مثلاً در توزیع نرمال دو پارامتر وجود دارد که عبارتند از: $\theta_1 = \mu$ و $\theta_2 = 6^2$

تنها منبعی که می‌تواند اطلاعاتی راجع به چگونگی قانون توزیع به ما بدهد یک نمونه تصادفی با n متغیر است. هر تابع نمونه‌ای (آماره) که به منظور تخمین پارامترهای قانون توزیع تعریف می‌گردد. «تخمین زن» نام دارد و آن را با $\tilde{\theta}$ نمایش می‌دهند. به عبارت دیگر تخمین نقطه‌ای θ همان $\tilde{\theta}$ است. مثلاً $(\bar{X})$ تخمین نقطه‌ای μ و S^2 تخمین نقطه‌ای 6^2 و $\hat{P}$ تخمین نقطه‌ای نسبت P است.

هرچه عضوهای نمونه (حجم نمونه) بیشتر باشد دقت تخمین زن بیشتر است. ولی حتی با نمونه‌ای به حجم دو (۲) نیز می‌توان میانگین جامعه را تخمین زد که مسلماً با خطای زیاد همراه است.

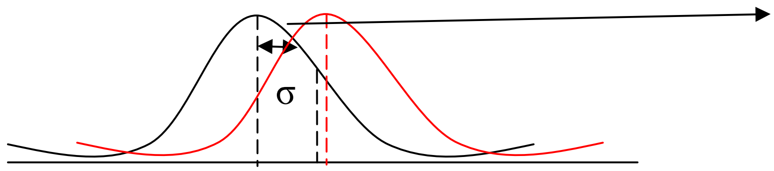
یک تخمین زن خوب باید دارای خواص زیر باشد:

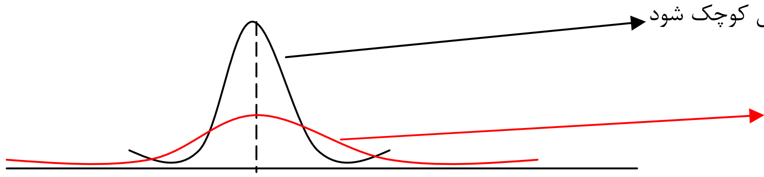
۱. خاصیت سازگاری یا پایداری از تخمین زن در مرحله اول انتظار می‌رود که با زیاد شدن حجم نمونه عددی حاصل شود که فاصله تخمین زن رفته رفته با پارامتر جامعه کم‌تر شود. چنین تخمین زنی را تخمین زن پایدار یا سازگار گویند.
۲. خاصیت ناتور (نااریب) بودن هرگاه امید ریاضی تخمین زن با خود پارامتر جامعه برابر باشد، این تخمین زن را ناتور (بدون تورش) می‌گویند. اگر چند تخمین زن ناتور داشته باشیم، آن که دارای واریانس کمتری باشد، با احتمال بیشتری به مقدار واقعی پارامتر نزدیک تر است.

اصل عدم:

حوادث با احتمال‌های کوچک که عملاً در یک آزمایش وقوع نمی‌کنند.

$$\bar{X} - \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}} \le \mu \le \bar{X} + \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}$$





$$\varepsilon(x) = \frac{1}{\sigma\sqrt{2n}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \rightarrow -\infty < u > +\infty$$

کمیت استاندارد = $u = \frac{x - E(\alpha)}{\sigma} \quad \rightarrow \varepsilon(u) = \frac{1}{\sqrt{2n}} e^{-\frac{u^2}{2}}$

مثال:

یکی از کارخانه‌های خودرو سازی می‌خواهد مدل جدید را به بازار عرضه کند برای این که یک فاصله اعتماد ٪۹۵ متوسط مصرف بنزین برای خودروهای جدید بدست آید مصرف ۳۶ خودرو را در یک فاصله مشخص اندازه گیری کرد و بر اساس آن میانگین ۱۳/۹ لیتر در ۱۰۰ کیلومتر با انحراف معیار ۲/۳ لیتر محاسبه شد. فاصله اعتماد مورد نظر را بدست آورید.

الف: وقتی واریانس جامعه معلوم است.

$$\bar{X} - \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}} \le \mu \le \bar{X} + \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}$$

ب: وقتی واریانس جامعه مجهول است.

$$\bar{X} - \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2},n-1} \le \mu \le \bar{X} + \frac{s}{\sqrt{n}} t_{1-\frac{\alpha}{2},n-1}$$

پاسخ:

$$1 - \alpha = \%95 \rightarrow \alpha = 0 = 0.05 \rightarrow \frac{\alpha}{2} = \frac{0.05}{\alpha} = 0.025$$

$$1 - \frac{\alpha}{2} = 1 - 0.025 = 0.975$$

$u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$ طبق جدول تابع توزیع نرمال

$$13/9 - \frac{2.3}{\sqrt{36}} \times 1/96 \le \mu \le 13/9 + \frac{2.3}{\sqrt{36}} \times 1/96$$

$$13/9 - 0.75 \le \mu \le 13/9 + 0.75$$

$$13/15 \le \mu \le 14.65$$

توزیع T

با زیاد شدن درجه آزادی شکل به توزیع نرمال واقعی نزدیک‌تر می‌شود. درجه آزادی تعداد متغیرهای مستقل یک تابع نمونه‌ای است.

فرمول: $D(T) = \frac{\nu}{\nu-2}$ $\qquad E(T) = 0 \qquad T = \sqrt{\frac{\chi^2_\nu}{\nu}}$

توزیع χ^2 (کای‌دو)

اگر کمیت X که دارای توزیع نرمال است به توان ۲ برسد آن کمیت دیگر دارای توزیع نرمال نیست بلکه توزیع کای‌دو دارد:

$$\Sigma U_i^2 = u_1^2 + u_2^2 + \ldots + u_n^2 \sim \chi^2_{p\,;\,n-1}$$

$$\chi^2_{p\,;\,n-1} = \chi^2_{0.975;12} = 23.3$$

$$E(\chi^2) = \nu \qquad D(\chi^2) = 2\nu$$

مثال: (شماره ۵ جزوه)

از یک جامعه نرمال نمونه‌ای به حجم N=16 انتخاب گردید و بر اساس آن $\overline{X} = 72$ و $S^2 = 4$ به دست آمده یک فاصله اطمینان برای میانگین جامعه با احتمال اعتماد ۹۹٪ بدست آورید.

(اگر بر اساس نمونه باید از فرمول زیر استفاده گردد.)

$n = 16$

$\overline{X} = 72$

$S^2 = 4 \quad (s = 2)$ تخمین واریانس

$1 - \alpha = 0.99$

$\rightarrow \underline{L}\,,\,\overline{L}_{(\mu)}$

$1 - \alpha = 0.99 \Longrightarrow \alpha = 0.01 \Longrightarrow \frac{\alpha}{2} = \frac{0.01}{2} = 0.005$

$n - 1 \Longrightarrow 16 - 1 = 15$

$1 - \frac{\alpha}{2} = 1 - 0.005 = 0.0995 \Longrightarrow t_{0.995;15} = 2.95$ $\quad \leftarrow$ عدد طبق جدول t

$72 - \frac{2}{\sqrt{16}} \times 2.95 \leq \mu \leq 72 + \frac{2}{4} \times 2.95$

$70.525 \leq \mu \leq 73.475$

مثال: (شماره ۶ جزوه)

دستمزد کارگران در یک صنعت بر طبق قانون نرمال توزیع شده است. می‌خواهیم یک فاصله اعتماد با احتمال ۹۵٪ برای میانگین دستمزد تعیین کنیم به این منظور نمونه‌ای به حجم N=25 نفر از کارگران به تصادف انتخاب شده‌اند و $\overline{X} = 5000$ و $S = 1000$ محاسبه شده است فاصله اعتماد را تعیین کنید.

$n = 25$

$\overline{X} = 5000$

$S = 1000$ تخمین انحراف معیار

$S^2 = ? \rightarrow \underline{L}\ , \overline{L}_{(\mu)} = ?$

$1 - \alpha = 0.95$

$1 - \alpha = 0.95 \Rightarrow 1 - \frac{\alpha}{2} = 0.975 \Rightarrow t_{1-\frac{\alpha}{2},n-1} \quad \Rightarrow t_{10.975} = 2.06 \leftarrow$ عدد طبق جدول t

$n - 1 \Rightarrow 25 - 1 = 24$

$5000 - \frac{1000}{\sqrt{25}} \times 2.06 \leq \mu \leq 5000 + \frac{1000}{5} \times 2.06$

$4588 \leq \mu \leq 5412$

مثال: (شماره ۷ جزوه)

فرض کنید طول عمر یک نوع باطری دارای قانون توزیع نرمال با واریانس σ=100 و میانگین نامعلوم μ است چه تعداد نمونه برای تخمین μ لازم است اگر بدانیم حداکثر خطا ۲/۵ و احتمال ۹۵٪ باشد.

$n = ؟$

$\xi = 2.5$

$1 - \alpha = 0.95$

$1 - \frac{\alpha}{2} = 1.96 \Rightarrow \Sigma = \frac{\delta}{\sqrt{n}} u_{1-\frac{\alpha}{2}} \quad \Rightarrow \Sigma^t = \frac{\delta^2}{n} {u^t}_{1-\frac{\alpha}{2}} \quad \Rightarrow n = \frac{\delta^2 \, {u^t}_{1-\frac{\alpha}{2}}}{\xi^t}$

مثال:

فرض کنید که طول عمر یک باطری دارای توزیع نرمال با امید ریاضی (میانگین) مجهول با S = 100 در یک نمونه N = 9 و $\overline{X} = 10$ برآورد شده است در سطح احتمال ۹۵٪ میانگین واقعی جامعه را به صورت فاصله‌ای تخمین بزنید.

$n = 9$

$\overline{X} = 10$

$S = 100$

$1 - \alpha = 0.95$

$1 - \alpha = 0.95 \Rightarrow 1 - \frac{\alpha}{2} = 0.975 \Rightarrow u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96 \Rightarrow$ عدد طبق جدول

$10 - \frac{100}{\sqrt{9}} \times 1.96 \leq \mu \leq 10 + \frac{100}{9} \times 2.06$

$4588 \leq \mu \leq 5412$

مثال: (شماره ۱۰ جزوه)

متغیر تصادفی در جامعه اصلی بر طبق قانون نرمال با امید ریاضی و واریانس مجهول توزیع شده است به منظور تخمین زدن امید ریاضی نمونه‌ای به حجم ۹ انتخاب می‌شود مطلوب است برآورد فاصله‌ای (تخمین فاصله‌ای یا فاصله اعتماد) برای امید ریاضی جامعه فوق با احتمال اعتماد $\gamma = 98\%$ به فرض اینکه نتایج مشاهدات در نمونه به قرار زیر به دست آمده باشد. X: 8, 9, 10, 6, 5. 7, 9, 8 ,6

با استفاده از فرمول: $S^2 = \frac{\Sigma(x_i-\bar{x})^2}{n-1}$

تخمین فاصله‌ای با فاصله اطمینان برای مقایسه دو میانگین یا تفاضل دو میانگین $(\mu_1-\mu_2)$ در دو جامعه

الف: با واریانس معلوم:

$$(\bar{X}_1-\bar{X}_2) - \sqrt{\frac{\delta_1^2}{n_1}+\frac{\delta_2^2}{n_2}}\,u_{1-\frac{\alpha}{2}} \leq \mu_1-\mu_2 \leq (\bar{X}_1-\bar{X}_2) + \sqrt{\frac{\delta_1^2}{n_1}+\frac{\delta_2^2}{n_2}}\,u_{1-\frac{\alpha}{2}}$$

ب: با واریانس مجهول:

$$(\bar{X}_1-\bar{X}_2) - \sqrt{\bar{S}^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\,t_{1-\frac{\alpha}{2};v_1+v_2-2} \leq \mu_1-\mu_2 \leq (\bar{X}_1-\bar{X}_2) + \sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\,t_{1-\frac{\alpha}{2};v_1+v_2-2}$$

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

مثال: (شماره ۸ جزوه)

سازمان چای ایران جهت اطلاع از قیمت خرده فروشی یک نوع چای تولیدی این سازمان دو نمونه تصادفی از شمال و جنوب تهران از فروشگاه‌های مختلف انتخاب نمود و قیمت فروش هر بسته چای ۱۰۰ گرمی را از ان‌ها پرسیده است نتایج به صورت زیر بوده است:

شمال: $\bar{X}_1 = 5800\ ,\ S_1^2 = 100\ ,\ n_1 = 30$

جنوب: $\bar{X}_1 = 5400\ ,\ S_2^2 = 90\ ,\ n_2 = 30$

فرض کنید قیمت هر بسته چای از توزیع نرمال پیروی کند مطلوب است یک فاصله اطمینان ۹۰٪ برای اختلاف میانگین قیمت فروش این دو نوع چای در کلیه فروشگاه‌های شمال و جنوب تهران با یکدیگر.

$$S^2 = \frac{29\times100+29\times90}{30+30-2} = \frac{29\times100+90}{58} = 95$$

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

$$1-\frac{\alpha}{2} = 1-0.05 = 0.95 \implies t_{1-\frac{\alpha}{2};v_1+v_2-2} = t_{0.095;58} = 1.645$$

$$(5800-5400) - \sqrt{95\left(\frac{1}{30}\times\frac{1}{30}\right.}\times 1.645 \leq \mu_1-\mu_2 \leq 400 + \sqrt{\frac{95}{15}}\times 1.645$$

$$395.87 \leq \mu_1-\mu_2 \leq 404.13 \implies \underline{L}\ ,\ \bar{L}_{(\mu_1-\mu_2)} = 395.87\ ,\ 404.13$$

مثال: (شماره ۹ جزوه)

از کمیت تصادفی x که بر طبق قانون نرمال با امید ریاضی و واریانس نامعلوم توزیع شده نمونه‌ای به حجم n=25 انتخاب می‌شود مطلوب است فاصله اعتماد برای امید ریاضی کمیت تصادفی فوق با احتمال اعتماد $\gamma = 0.98$ به فرض اینکه تخمین پارامترها در نمونه انتخاب شده به صورت $\overline{x} = 120$ و S2 = 16 به دست آمده باشد.

(هر وقت داده‌ها بر اساس نمونه باشد)

$x \sim N(\mu = ?, \sigma^2 = ?)$

$n = 25$

$\underline{L}, \overline{L}_{(\mu)} = ?$

$\overline{X} = 120$

$S^2 = 16$

$1 - \alpha = 0.98 \Rightarrow \alpha = 0.02 \Rightarrow \frac{\alpha}{2} = \frac{0.02}{2} = 0.01$

$1 - \frac{\alpha}{2} = 1 - 0.001 = 0.99 \Rightarrow t_{1-\frac{\alpha}{2};n-1} = t_{0.99;24} = 2.49$

$\overline{X} - \frac{\sigma}{\sqrt{n}} t_{1-\frac{\alpha}{2},n-1} \leq \mu \leq \overline{X} + \frac{\sigma}{\sqrt{n}} t_{1-\frac{\alpha}{2},n-1}$

$120 - \frac{4}{\sqrt{25}} \times 2.49 \leq \mu \leq 120 + \frac{4}{9} \times 2.49$

$118.008 \leq \mu \leq 121.992 \Rightarrow \underline{L}, \overline{L}_{(\mu_1 - \mu_2)} = 118.008, 121.992$

مثال: (شماره ۲۳ جزوه)

بررسی‌ها نشان می‌دهد که توزیع وزن محصولات تولید شده یک کارخانه بزرگ توزیع نرمال و انحراف معیار آن ۲۱ تن است از آنجا که اندازه گیری وزن محصولات به طور روزانه امکان پذیر نیست یک نمونه ۵۰ روزه از تولیدات انتخاب شده است که میانگین وزن آن ۸۷۱ تن است در سطح اطمینان ۹۰٪ تخمین فاصله‌ای میانگین واقعی وزن محصولات تولید شده طی یک روز را محاسبه کنید.

$x \sim N(\mu = ?, S = ?)$

$n = 50$

$\underline{L}, \overline{L}_{(\mu)} = ?$

$\overline{X} = 871$

$1 - \alpha = 0.90 \Rightarrow \alpha = 0.1 \Rightarrow \frac{\alpha}{2} = \frac{0.1}{2} = 0.05$

$1 - \frac{\alpha}{2} = 1 - 0.05 = 0.95 \Rightarrow u_{1-\frac{\alpha}{2}} = u_{0.95} = \frac{1.64+1.65}{2} = 1.645$

$\overline{X} - \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}} \leq \mu \leq \overline{X} + \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}$

$871 - \frac{21}{\sqrt{50}} \times 1.645 \leq \mu \leq 871 + \frac{21}{\sqrt{50}} \times 1.645$

$866.005 \leq \mu \leq 875.935 \Rightarrow \underline{L}, \overline{L}_{(\mu_1 - \mu_2)} = 866.005, 875.935$

مثال: (شماره ۲۴ جزوه)

بازاریابی در صدد بررسی و برآورد قدرت خرید ساکنان یک محله تهران است او مجبور است یک نمونه تصادفی ۱۰ تایی از بین خریداران انتخاب و قدرت خرید را برای هر یک از آن‌ها اندازه گیری کند قدرت خرید نمونه فوق بر حسب ۱۰.۰۰۰ تومان چنین است: $x_i = 8, 7, 5, 4. 12, 15, 10, 13, 14, 12$ قدرت خرید ساکنان محله از توزیع نرمال برخوردار است در سطح اطمینان ۹۵٪ میانگین قدرت خرید را برآورد کنید. (تخمین بزنید)

$x \sim N$

$x_i = 8, 7, 5, 4. 12, 15, 10, 13, 14, 12$

$n= 10$

$1 - \alpha = 0.90$

$\underline{L} , \overline{L}_{(\mu)} = ?$

$$\overline{X} = \frac{\Sigma x_i}{n} = \frac{100}{10} = 10 \quad \rightarrow \quad S^2 = \frac{\Sigma (x_i - \overline{x})^2}{n} = \frac{4+9+25+36+4+25+9+16+4}{9} = \frac{132}{9} = 14.7$$

$$1 - \alpha = 0.90 \implies 1- \frac{\alpha}{2} = 0.975$$

$$t_{1-\frac{\alpha}{2};n-1} = t_{0.975;9} = 2.26 \rightarrow \overline{X} - \frac{\sigma}{\sqrt{n}} t_{1-\frac{\alpha}{2},n-1} \le \mu \le \overline{X} + \frac{\sigma}{\sqrt{n}} t_{1-\frac{\alpha}{2},n-1}$$

$$10 - \frac{\sqrt{14.7}}{\sqrt{10}} \times 2.26 \le \mu \le 10 + \frac{\sqrt{14.7}}{\sqrt{10}} \times 2.26 \rightarrow 7.228 \le \mu \le 12.712$$

مثال: (شماره ۲۸ جزوه)

بازاریابی در صدد بررسی و برآورد قدرت خرید ساکنان یک محله در تهران است به این منظور یک نمونه تصادفی ۶ تایی از بین خریداران انتخاب و قدرت خرید را برای هر یک از آن‌ها اندازه گیری کرد که بر حسب ۱۰.۰۰۰ تومان به صورت چنین است: $x_i = 8, 7, 15, 12. 10, 7$ مطلوب است یک فاصله اطمینان ۹۹٪ برای میانگین قدرت خرید ساکنان این محله.

$x \sim N$

$x_i = 8, 7, 15, 12. 10, 7$

$n= 6$

$1 - \alpha = 0.99$

$\underline{L} , \overline{L}_{(\mu)} = ?$

x_i	F_i	$F_i x_i$	$F_i x_i^2$
7	2	14	98
8	1	18	64
10	1	10	100
12	1	12	144
15	1	15	225
	$n = 6$	$\Sigma F_i x_i = 59$	$\Sigma F_i x_i^2 = 631$

$$\overline{X} = \frac{\Sigma F_i x_i}{n} = \frac{59}{6} = 9.83$$

$$S^2 = \frac{1}{n-1} [\Sigma F_i x_i^2 - \frac{1}{n} (\Sigma F_i x_i)^2]$$

$$S^2 = \frac{1}{5} [631 - \frac{1}{6} (59)^2]$$

$$S^2 = \frac{50.84}{5} = S^2 = 10.168$$

$$1 - \alpha = 0.99 \implies 1- \frac{\alpha}{2} = 0.995$$

$$t_{1-\frac{\alpha}{2};n-1} = t_{0.995;5} = 4.03 \rightarrow \overline{X} - \frac{\sigma}{\sqrt{n}} t_{1-\frac{\alpha}{2},n-1} \le \mu \le \overline{X} + \frac{\sigma}{\sqrt{n}} t_{1-\frac{\alpha}{2},n-1}$$

$$9.83 - \frac{\sqrt{10.168}}{\sqrt{6}} \times 4.03 \le \mu \le 9.83 + \frac{\sqrt{10.168}}{\sqrt{6}} \times 4.03$$

تخمین فاصله‌ای (برآورد فاصله‌ای – فاصله اطمینان) برای پارامتر P

برای این منظور از فرمول زیر استفاده می‌گردد. $\widehat{P}$ برابر است با تخمین نقطه‌ای p

$$\widehat{P} - \sqrt{\frac{\widehat{P}\,\widehat{q}}{n}}u_{1-\frac{\alpha}{2}} \le p \le \widehat{P} + \sqrt{\frac{\widehat{P}\,\widehat{q}}{n}}u_{1-\frac{\alpha}{2}}$$

مثال: (شماره ۱۳ جزوه)

از جامعه‌ای با توزیع دو نقطه‌ای نمونه‌ای به حجم $n = 40$ انتخاب می‌شود مطلوب است فاصله اعتماد با احتمال $\gamma = 0.95$ برای پارامتر P به فرض اینکه در نمونه انتخاب شده $m_A = 22$ به دست آمده باشد.

$n = 40$

$1- \alpha = 0.95$

$m_A = 22$

$\widehat{P} = \frac{m_A}{n} = \frac{22}{40} = 0.55$

$\widehat{q} = 1 - \widehat{P} = 1 – 0.55 = 0.45$

$1- \alpha = 0.95 \rightarrow 1 - \frac{\alpha}{2} = 0.95 \rightarrow u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$

$$\widehat{P} - \sqrt{\frac{\widehat{P}\,\widehat{q}}{n}}u_{1-\frac{\alpha}{2}} \le p \le \widehat{P} + \sqrt{\frac{\widehat{P}\,\widehat{q}}{n}}u_{1-\frac{\alpha}{2}}$$

$$0.55 - \sqrt{\frac{0.55\times0.45}{40}} \times 1.96 \le p \le 0.55 + \sqrt{\frac{0.55\times0.45}{40}} \times 1.96$$

$$0.39579 \le p \le 0.70421$$

مثال: (شماره ۱۴ جزوه)

به منظور تخمین زدن نسبت محصولات درجه یک بین واحدهای محصولات تولید شده در یک موسسه تولیدی از بین محصول تولید شده نمونه‌ای به حجم $n = 25$ واحد انتخاب شد و توسط بازرس فنی به درجات ۱، ۲ و ۳ طبقه‌بندی گردید و معلوم شد که تعداد $m = 5$ واحد نمونه جزء محصول درجه یک می‌باشد. مطلوب است فاصله اعتماد برای نسبت واحدهای محصول درجه یک با احتمال اعتماد $\gamma = 0.95$

$n = 25$

$1- \alpha = 0.95$

$m_A = 5$

$\widehat{P} = \frac{m_A}{n} = \frac{5}{25} = 0.2$

$\widehat{q} = 1 - \widehat{P} = 1 – 0.2 = 0.8$

$1- \alpha = 0.95 \rightarrow 1 - \frac{\alpha}{2} = 0.95 \rightarrow u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$

$$\widehat{P} - \sqrt{\frac{\widehat{P}\,\widehat{q}}{n}}u_{1-\frac{\alpha}{2}} \le p \le \widehat{P} + \sqrt{\frac{\widehat{P}\,\widehat{q}}{n}}u_{1-\frac{\alpha}{2}}$$

$$0.2 - \sqrt{\frac{0.2\times0.8}{25}} \times 1.96 \le p \le 0.2 + \sqrt{\frac{0.2\times0.8}{25}} \times 1.96$$

$$0.0432 \le p \le 0.3568$$

تخمین فاصله‌ای (برآورد فاصله‌ای – فاصله اطمینان) برای پارامتر واریانس (σ^2)

در این توزیع از تخمین χ^2 و از فرمول زیر استفاده می‌گردد.

$$\frac{(n-1)s^2}{\chi^2_{1-\alpha;n-1}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{\alpha;n-1}}$$

مثال: (شماره ۲۱ جزوه)

از یک جامعه نرمال با واریانس نامعلوم یک نمونه تصادفی ۱۶ تایی انتخاب می‌کنیم اگر واریانس نمونه $S^2 = 4$ باشد یک فاصله اعتماد ۹۵٪ برای σ^2 پیدا کنید.

$n = 16$
$s^2 = 4$
$1 - \alpha = 0.95$
$\underline{L}\ ,\ \overline{L}_{(\mu)} = ?$

$$\chi^2_{0.05;15} = 7.26 \quad , \quad \chi^2_{0.095;15} = 25$$

$$\frac{15 \times 4}{25} \leq \sigma^2 \leq \frac{15 \times 4}{7.26} \quad \rightarrow \quad 2.04 \leq \sigma^2 \leq 8.26$$

آزمون فرضیه‌های آماری

در انجام تحقیقات علمی اغلب به مسائلی برخورد می‌شود که پاسخ آن به درستی بر محقق معلوم نیست. در چنین شرایطی اقدام به تدوین فرضیه می‌گردد.

فرضیه عبارت از ادعایی است که توسط محققین، مدیران، مسئولین کنترل کیفیت و … در مورد مسائل اجتماعی، اقتصادی، سیاسی پیش کشیده می‌شود.

فرضیه آماری گزاره‌ای راجع به چگونگی توزیع صفت متغیر (کمیت تصادفی) است که درستی یا نادرستی آن به طور مطلق معلوم نیست.

این فرضیه‌ها را بر اساس نتایج مشاهدات در نمونه می‌توان آزمون کرد.

تمام عملیاتی که به منظور روشن ساختن درستی یا نادرستی یک فرضیه پیش کشیده می‌شود آزمون فرضیه نام دارد. در واقع هدف از آزمون فرضیه آماری تعیین این موضوع است که با توجه به اطلاعات به دست آمده از داده‌های نمونه‌ی مشاهده شده، ادعای مطرح شده (فرضیه) درباره‌ی خصوصیتی از جامعه تایید می‌شود یا خیر؟

تعاریف و مفاهیم اساسی تئوری عمومی آزمون‌ها

به طور کلی فرضیه‌های آماری بر دو نوع اند: یکی فرضیه راجع به قانون توزیع (نا پارامتری) و دیگری راجع به پارامترهای قانون توزیع (پارامتری)

اگر فرضیه آماری قانون توزیع یا پارامتر جامعه را کاملاً مشخص کند آن را فرضیه آماری ساده گویند و در غیر این صورت آن را فرضیه آماری مرکب نامند. به عبارت دیگر فرضیه ساده فرضیه‌ای را گویند که با بیان آن قانون توزیع کمیت کاملاً مشخص گردد مثلاً در مورد بارش می‌توان گفت:

$X \sim N\,(\mu = 350\,mm\,,\ \ \delta^2 = 25m^2$

در اینجا هیچ ابهامی از نظر شناسایی قانون توزیع باقی مانده است چون توزیع نرمال دو پارامتر دارد یکی میانگین و دیگر واریانس که هر دو در اینجا مشخص شده است. ولی فرضیه‌هایی که با بیان ان قانون توزیع کمیت تصادفی کاملاً معلوم نباشد، فرضیه مرکب نام دارد. مثلاً در مورد بارش:

$X \sim N\,(\mu = 350\,mm\ ,\ \ \delta^2 = ?$

چون بی نهایت توزیع نرمال وجود دارد که دارای متوسط 350 mm بارش است، تا وقتی که واریانس (δ^2) داده نشود فرضیه مرکب است.

یک آزمون فرضیه دارای مراحل منطقی است که در ذیل به آن‌ها اشاره می‌گردد.

مرحله اول: در هر آزمون فرضیه دو فرض آماری در مقابل یک دیگر ساخته می‌شود.

یکی فرض آماری عدم تفاوت (H_0) و دیگری فرض مقابل که با (H_1) نشان داده می‌شود.

ماهیت آزمون کردن یک فرضیه آماری در این است که مشخص می‌کند آیا نتایج مشاهدات با فرضیه بیان شده سازگاری دارد یا نه؟

مرحله دوم: در این مرحله سطح احتمال انتخاب می‌شود که متناظر با حوادثی است که در شرایط مورد بررسی وقوع آن غیر ممکن فرض شده است و آن را با $\propto$ (آلفا) می‌تواند ۰/۰۵، ۰/۰۱ و... باشد.

سطح احتمال تحت شرایط مورد بررسی ممکن است در طرفین فاصله اطمینان و یا در سمت راست و یا در سمت چپ فاصله اطمینان تعیین شود.

مرحله سوم: در این مرحله ملاک آزمون فرضیه آماری ساخته می‌شود. ملاک آزمون مفهومی است که بر اساس نمونه و به کمک نتایج مشاهدات نسبت به سازگاری یا ناسازگاری فرضیه تصمیم‌گیری می‌کند.

برای ساختن ملاک آزمون از یک اصل مهم در آمار به نام اصل عدم استفاده شده است. (اصل عدم می‌گویند، حوادث با احتمال‌های کوچک در یک آزمایش امکان وقوع ندارند) بر اساس اصل عدم غیرممکن است که در یک مشاهده خارج از فاصله $\mu - 3\delta$ و $\mu + 3\delta$ عددی داشته باشیم.

ملاک آزمون را با K نمایش می‌دهند که چند نمونه از آن در ذیل معرفی می‌گردد:

$$K = U = \frac{\bar{X} - \mu_0}{\frac{\delta}{\sqrt{n}}}$$

$$K = T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$$

$$K = \chi^2 = \frac{(n-1)s^2}{\delta_0^2}$$

مرحله چهارم: در مرحله چهارم ناحیه بحرانی تشکیل می‌شود. مقادیری از تابع آزمون که با احتمال آلفا یا کوچک‌تر از α می‌تواند وقوع کند ناحیه بحرانی تابع آزمون را تشکیل می‌دهد و آن را با نماد W نشان می‌دهند.

ناحیه بحرانی برای ملاک آزمون ناحیه‌ای است که احتمال قرار گرفتن ملاک آزمون در آن ناحیه برابر با سطح احتمال $\propto$ باشد. چون سطح احتمال $\propto$ سطحی است که عملاً وقوع حادثه در آن غیرممکن است؛ لذا ناحیه بحرانی ناحیه‌ای است که ملاک آزمون با فرض درست بودن H_0 نمی‌توان مقادیر آن ناحیه را اختیار کند.

مکمل ناحیه بحرانی را ناحیه پذیرش یا ناحیه قبل و یا ناحیه مقادیر مجاز نامند و آن ناحیه‌ای است که اگر مقدار ملاک آزمون در آن ناحیه قرار گیرد، دلیلی برای رد فرضیه H_0 وجود ندارد.

مرحله پنجم: بعد از ساختن ناحیه بحرانی در مرحله پنجم نمونه‌ای تصادفی با حجم مورد نیاز (در فصل قبل راجع به اندازه حجم نمونه صحبت شد) انتخاب می‌شود و بر اساس آن مقدار عددی ملاک آزمون محاسبه می‌شود.

مرحله ششم: این مرحله، مرحله تصمیم گیری می‌باشد. در واقع بر اساس مقایسه مقدار عددی ملاک آزمون و ناحیه بحرانی تصمیم گیری انجام می‌شود.

هرگاه مقدار عددی ملاک آزمون در ناحیه بحرانی قرار گیرد $K\epsilon W$ فرضیه H_0 رد و فرضیه مقابل پذیرفته می‌شود.

در واقع فرضیه H_0 را تا زمانی که خلاف آن اثبات نشده به عنوان فرضیه درست حفظ می‌کنیم.

در قضاوت به کمک ملاک آزمون نسبت به سازگاری یا ناسازگاری فرضیه احتمال خطا وجود دارد. به طوری که اگر یک فرضیه درست را به علت انتخاب نمونه نامناسب فرضیه ناسازگار (نادرست) اعلام کنیم چنین خطایی را خطای نوع اول گویند که احتمال آن برابر آلفا است.

$$P\,(\mathrm{I}) = P\,(K\in W \mid H_0) = \alpha$$

گاهی نیز ممکن است یک فرضیه نادرست را فقط به دلیل اینکه تابع آزمون در ناحیه بحرانی قرار نگرفته است رد نکرده و به عنوان فرضیه درست قبول گردد که این خطا را خطای نوع دوم گویند و مقدار آن برابر β (بتا) است.

$$P\,(\mathrm{II}) = P\,(K\in W \mid \overline{H}_0) = \beta$$

اگر خطای نوع اول (α) را خیلی کوچک کنیم باعث افزایش خطای نوع دوم می‌گردد. به طوری که اگر $\propto$ به سمت صفر میل کند در حل احتمال خطای نوع دوم همیشه یک (حتمی) است.

در واقع وقتی حجم نمونه (n) را خیلی بزرگ کنیم به طوری که برابر حجم جامعه گردد آن گاه $\propto = 0$ و $\beta = 0$ خواهد شد.

شمای کل آزمون فرضیه‌ها

مرحله اول

H_0: $\theta = \theta_0$

H_1: $\theta \neq \theta_0$ یا $\theta > \theta_0$ یا $\theta < \theta_0$

سطح معنادار بودن احتمال

مرحله دوم → α = ...

مرحله سوم

$K = f\,(X_1, X_2, \ldots\ldots, X_n)$

یک تابع نمونه‌ای را به عنوان تابع آزمون (ملاک آزمون) انتخاب می‌کنیم.

به طوری که با فرض درست بودن فرضیه H_0 قانون توزیع آن کاملاً مشخص شده باشد.

مرحله چهارم ← W (   )

ناحیه بحرانی بر اساس سطح احتمال آلفا ساخته می‌شود.

مرحله پنجم ← انتخاب نمونه (test critrum)

مرحله ششم ← تصمیم گیری

قاعده تصمیم:

هرگاه مقدار عددی ملاک آزمون که بر اساس نتایج مشاهدات محاسبه می‌شود، در ناحیه بحرانی قرار گیرد فرضیه H_0 را تا زمانی که خلاف آن اثبات نشده به عنوان فرضیه درست حفظ می‌کنیم.

تذکـر:

اگر فرضیه $H_1: \theta + \theta_0$ باشد آزمون دو طرفه است که در این صورت ناحیه بحرانی در دو طرف فاصله اطمینان تشکیل می‌گردد. ولی اگر فرضیه $H_1: \theta > \theta_0$ یا $H_1: \theta < \theta_0$ باشد آزمون فرضیه یک طرفه است؛ و در این صورت به ترتیب ناحیه بحرانی در طرف راست فاصله اطمینان و طرف چپ فاصله اطمینان تشکیل می‌گردد.

مثال: (شماره ۱۵ جزوه)

در یک نمونه تصادفی ۴۰۰ تایی از مردم یک شهر مشخص شد که ۸۰ نفر بیکار هستند یک فاصله اطمینان ٪۹۵ برای نسبت واقعی بیکاران این شهر پیدا کنید.

$n = 400$

$1 - \alpha = 0.95$

$m = 80$

$\underline{L}\ ,\ \overline{L}_{(p)} = ?$

$\widehat{P} = \frac{m}{n} = \frac{80}{400} = 0.2$

$\hat{q} = 1 - \widehat{P} = 1 - 0.2 = 0.8$

$1 - \alpha = 0.95 \rightarrow 1 - \frac{\alpha}{2} = 0.95 \rightarrow u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$

$$\widehat{P} - \sqrt{\frac{\widehat{P}\,\hat{q}}{n}}\, u_{1-\frac{\alpha}{2}} \le p \le \widehat{P} + \sqrt{\frac{\widehat{P}\,\hat{q}}{n}}\, u_{1-\frac{\alpha}{2}}$$

$$0.2 - \sqrt{\frac{0.2\times0.8}{400}} \times 1.96 \le p \le 0.2 + \sqrt{\frac{0.2\times0.8}{400}} \times 1.96$$

$$0.1608 \le p \le 0.2392$$

مثال: (شماره ۱۶ جزوه)

از بین خانواده‌های کارمندان دولت ۴۰۰۰ خانوار به صورت تصادفی انتخاب شده است که آیا دارای تلویزیون می‌باشند یا نه؟ از مشاهدات به دست آمده معلوم گردید ۳۲۰۰ خانوار دارای تلویزیون می‌باشند. مطلوب است ارزیابی نسبت خانوارهای دارای تلویزیون بین تمام خانواده‌های کارمندان دولت.

$n = 4000$

$1-\alpha = 0.95$

$m = 3200$

$\underline{L}\ ,\ \overline{L}_{(p)} = ?$

$$\widehat{P} = \frac{m}{n} = \frac{3200}{4000} = 0.8$$

$$\hat{q} = 1 - \widehat{P} = 1 - 0.8 = 0.2$$

$$1-\alpha = 0.95 \rightarrow 1 - \frac{\alpha}{2} = 0.95 \rightarrow u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$$

$$\widehat{P} - \sqrt{\frac{\widehat{P}\,\hat{q}}{n}}\,u_{1-\frac{\alpha}{2}} \leq p \leq \widehat{P} + \sqrt{\frac{\widehat{P}\,\hat{q}}{n}}\,u_{1-\frac{\alpha}{2}}$$

$$0.8 - \sqrt{\frac{0.2\times0.8}{400}} \times 1.96 \leq p \leq 0.8 + \sqrt{\frac{0.2\times0.8}{400}} \times 1.96$$

مثال: (شماره ۱۷ جزوه)

به منظور تخمین زدن نسبت اشیاء نقص‌دار (p) در پارتی محصول تولید شده در نظر است نمونه‌ای به حجم n انتخاب شود از بازرسی‌های پارتی‌های اشیاء تولید شده واریانس نسبت اشیاء نقص‌دار برابر 0.01 به دست آمده است، مطلوب است حجم نمونه مورد نیاز برای تخمین زدن نسبت اشیاء نقص‌دار به شرط اینکه خطای حدی این تخمین از ۲٪ تجاوز نکنند و احتمال اعتماد کمتر از ۹۵٪ نباشد.

$1-\alpha = 0.95$

$p.q = 0.01$

$\xi = 0.02$

$\underline{L}\ ,\ \overline{L}_{(p)} = ?$

$$n = \frac{p.q \times u^2_{1-\frac{\alpha}{2}}}{\xi^2} = \frac{0.01 \times (1.96)^2}{(0.02)^2} = \frac{0.038446}{0.0004} = \frac{384.16}{4} = 96.04 \sim 97$$

مثال: (شماره ۱۸ جزوه)

به منظور تخمین زدن خانوارهای کم درآمد در یک ناحیه روستایی که دارای ۱۰۰۰ خانوار می‌باشد ۲۰۰ خانوار به عنوان نمونه انتخاب شده و وضع درآمد خانوار نمونه مطالعه گردید. در نمونه انتخاب شده ۸۰ خانوار کم درآمد تشخیص داده شد. مطلوب است تخمین نسبت خانوارهای کم درآمد در ناحیه روستایی فوق از طریق تعیین یک فاصله اطمینان ۹۵٪.

$N = 1000$
$n = 200$
$1- \alpha = 0.95$
$m = 80$
$\underline{L}\ , \overline{L}_{(p)} = ?$

$\widehat{P} = \frac{m}{n} = \frac{80}{200} = 0.4$
$\hat{q} = 1 - \widehat{P} = 1 - 0.4 = 0.6$
$1- \alpha = 0.95 \rightarrow 1 - \frac{\alpha}{2} = 0.95 \rightarrow u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$

$$\widehat{P} - \sqrt{\frac{\widehat{P}\,\hat{q}}{n}}\,u_{1-\frac{\alpha}{2}} \leq p \leq \widehat{P} + \sqrt{\frac{\widehat{P}\,\hat{q}}{n}}\,u_{1-\frac{\alpha}{2}}$$

$$0.4 - \sqrt{\frac{0.4\times 0.6}{200}} \times 1.96 \leq p \leq 0.4 + \sqrt{\frac{0.4\times 0.6}{200}} \times 1.96$$

مثال: (شماره ۱۹ جزوه)

مدیر بزرگ شرکت A در نظر دارد شرکت B را که سماور برقی تولید می‌کند بخرد قبلاً شرکت B ۳۰٪ از کل بازار سماور را در اختیار داشته است. اگر بخواهیم نسبت سهام شرکت B را با اطمینان ۹۰٪ با خطای 0.025 تخمین بزنیم به چه تعداد نمونه از بازار سماور احتیاج داریم؟

$p = 0.3$
$1- \alpha = 0.90$
$\xi = 0.025$
$n = ?$

$\hat{q} = 1 - \widehat{P} = 1 - 0.3 = 0.7$
$1- \alpha = 0.90 \rightarrow 1 - \frac{\alpha}{2} = 0.95 \rightarrow \frac{1.64+1.65}{2} = 1.645$

$$n = \frac{p.q \times u^2_{1-\frac{\alpha}{2}}}{\xi^2} = \frac{0.03 \times 0.7 \times (1.645)^2}{(0.025)^2} = \frac{0.21\times 2.766025}{625} \sim 910$$

مثال: (شماره ۲۶ جزوه)

به منظور ارزیابی میانگین وزن قوطی‌های کنسرو در یک واحد تولیدی تعداد $n = 6$ قوطی را به صورت تصادفی از خط تولید انتخاب و وزن نموده‌ایم. نتایج وزن قوطی‌های کنسرو به صورت زیر بود:

$x_i = 195, 199, 200, 201. 199, 196$

فرض کنید وزن قوطی‌های کنسرو در این کارخانه از توزیع نرمال پیروی می‌کند. مطلوب است محاسبه:

الف: فاصله اطمینان ۹۵٪ برای میانگین وزن قوطی‌های کنسرو در این کارخانه.

ب: فاصله اطمینان ۹۵٪ برای انحراف معیار قوطی‌های کنسرو در این کارخانه.

$x \sim N(\mu = ?\ , S^2 = ?)$

$x_i = 195, 199, 200, 201. 199, 196$

$n= 6$

$1 - \alpha = 0.95$

$\underline{L}\ , \overline{L}_{(\mu)} = ?$

$$\overline{X} - \frac{\sigma}{\sqrt{n}} t_{1-\frac{\alpha}{2},n-1} \le \mu \le \overline{X} + \frac{\sigma}{\sqrt{n}} t_{1-\frac{\alpha}{2},n-1}$$

$$\overline{X} = \frac{\Sigma x_i}{n} = \frac{188}{6} = 198 \quad \rightarrow \quad S^2 = \frac{\Sigma (x_i - \overline{x})^2}{n-1} = \frac{9+0+4+0+4}{5} = \frac{26}{5} = 5.2$$

$$1 - \alpha = 0.95 \implies 1 - \frac{\alpha}{2} = 0.975$$

$$t_{1-\frac{\alpha}{2};n-1} = t_{0.975;5} = 2.57 \rightarrow$$

$$198 - \frac{2.28}{\sqrt{6}} \times 2.57 \le \mu \le 198 + \frac{2.28}{\sqrt{6}} \times 2.57$$

آزمون میانگین در یک جامعه با واریانس معلوم

1) $H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$ یا $\mu \ge \mu_0 \rightarrow w\,(u \ge u_{1-\frac{\alpha}{2}})$ یا $\mu \le \mu_0 \rightarrow w\,(u \le u_{1-\frac{\alpha}{2}})$

2) $\alpha = 0.05$

3) $k = u = \dfrac{\bar{X} - \mu_0}{\frac{\delta}{\sqrt{n}}}$

4) $w\,(u \le -u_{1-\frac{\alpha}{2}} \quad , \quad u \ge u_{1-\frac{\alpha}{2}})$

5) انتخاب نمونه

6) تصمیم گیری

$\rightarrow k \in w \rightarrow$ فرضیه موجود رد و فرضیه مقابل پذیرفته می‌شود

$\rightarrow k \notin w \rightarrow$ فرضیه موجود را می‌پذیریم

آزمون میانگین در یک جامعه با واریانس مجهول

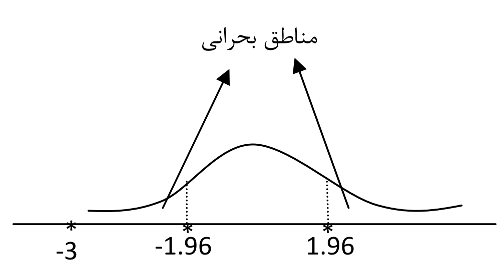
1) $H_0: \mu = \mu_0$
   $H_1: \mu \neq \mu_0$
2) $\alpha = 0.05$
3) $k = T = \dfrac{\bar{X} - \mu_0}{\frac{\delta}{\sqrt{n}}}$
4) $w\ (T \leq -t_{1-\frac{\alpha}{2},n-1} \quad , \quad T \geq t_{1-\frac{\alpha}{2},n-1})$
5) انتخاب نمونه
6) تصمیم گیری
$\rightarrow k \in w \rightarrow$ فرضیه موجود رد و فرضیه مقابل پذیرفته می‌شود
$\rightarrow k \notin w \rightarrow$ فرضیه موجود را می‌پذیریم

مثال: (شماره ۴۸ جزوه)

متغیر تصادفی x در جامعه اصلی بر طبق قانون نرمال با امید ریاضی مجهول و انحراف معیار ۲ توزیع شده است. نسبت به امید ریاضی مجهول فرضیه‌ای به صورت $H_0 = \mu = 10$ پیش کشیده است. به منظور آزمون فرضیه فوق در نظر است نمونه‌ای به حجم ۴ انتخاب شود. مطلوب است آزمون فرضیه فوق به فرض اینکه مقادیر مشاهده شده در نمونه انتخاب شده به قرار زیر باشد: $x_i = 5, 7, 7, 9$

$x \sim N\ (\mu = ?\ , S = ?)$
$H_0 = \mu = 10$
$x_i = 5, 7, 7, 9$
$n = 4$
$\bar{X} = ?$

1) $H_0: \mu = 10$
   $H_1: \mu \neq 10$
2) $\alpha = 0.05$
3) $k = u = \dfrac{\bar{X} - \mu_0}{\frac{\delta}{\sqrt{n}}}$
4) $w\ (u \leq -u_{1-\frac{\alpha}{2}} \quad , \quad u \geq u_{1-\frac{\alpha}{2}})$
   $w\ (u \leq -u_{0.975} \quad , \quad u \geq u_{0.975})$
   $w\ (u \leq -1.96 \quad , \quad u \geq 1.96)$
5) انتخاب نمونه
$\bar{X} = \dfrac{\Sigma x_i}{n} = \dfrac{5+7+7+9}{4} = \dfrac{28}{4} = 7$
$k = u = \dfrac{7-10}{\frac{2}{\sqrt{4}}} = \dfrac{-3}{\frac{3}{2}} = -3$
6) تصمیم گیری
$-3 < -1.96 \Rightarrow k \in w \rightarrow$



چون مقدار عددی ملاک آزمون در ناحیه بحرانی قرار گرفته است فرضیه H_0 رد و فرضیه مقابل پذیرفته می‌شود.

مثال: (شماره ۴۹ جزوه)

یک کارشناس کشاورزی معتقد است که رشد نوعی بذر بر طبق قانون رمان با ارتفاع متوسط 25cm و انحراف معیار 2cm توزیع شده است. نمونه مرکب از ۱۰۰ بذر از همین نوع را تحت شرایط جدیدی می‌کاریم اگر معلوم شود که این نمونه دارای میانگین 22cm است آیا نظر کارشناس را در سطح احتمال $\alpha = 0.05$ می‌توانیم رد کنیم.

$x \sim N\,(\mu = 25\ ,\ S = 2)$

$H_0 = \mu = 25$

$n = 100$

$\bar{X} = 22$

1) H_0: $\mu = 25$

H_1: $\mu \neq 25$

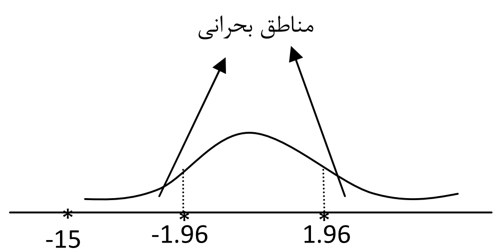
2) $\alpha = 0.05$

3) $k = u = \dfrac{\bar{X} - \mu_0}{\frac{\delta}{\sqrt{n}}}$

4) $w\ (u \leq u_{1-\frac{\alpha}{2}}\quad ,\quad u \geq u_{1-\frac{\alpha}{2}}\ )$

$w\ (u \leq u_{0.975}\quad ,\quad u \geq u_{0.975}\ )$

$w\ (u \leq -1.96\quad ,\quad u \geq 1.96\ )$



5) انتخاب نمونه

$k = u = \dfrac{22-25}{100} = \dfrac{-3}{\frac{2}{10}} = \dfrac{-30}{2} = -15$

6) تصمیم گیری

$-15 < -1.96 \Longrightarrow k \in w$

چون مقدار عددی ملاک آزمون در ناحیه بحرانی قرار گرفته است فرضیه H_0 رد و فرضیه مقابل پذیرفته می‌شود.

مثال: (شماره ۵۰ جزوه)

یک دستگاه اتو مات استوانه‌هایی با قطر a = 10cm تولید می‌کند. این اندازه به توسط طرح فنی تعیین شده است و همواره کنترل می‌گردد اندازه قطر استوانه‌ها (x) بر طبق قانون نرمال با امید ریاضی مجهول و انحراف معیار 0.5 توزیع می‌شود.

واحد بازرسی فنی در تمام مراحل فرآیند تولید نمونه‌ای به حجم ۲۵ از خط تولید انتخاب می‌کند و می‌خواهد مطمئن باشد آیا در حین تولید به دلیل دخالت برخی عوامل اندازه تعیین شده برای این دستگاه تغییر کرده است یا نه؟ در صورتی که بدانیم در ۲۵ نمونه اندازه گیری شده $\bar{X} = 9.9$ محاسبه گردد.

$x \sim N\,(\mu = ?\ ,\ S = 0.5)$

$H_0 = \mu = 10$

$n = 25$

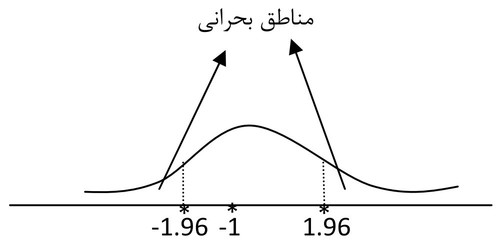
$\bar{X} = 9.9$

1) H_0: $\mu = 10$

H_1: $\mu \neq 10$

2) $\alpha = 0.05$

3) $k = u = \frac{\bar{X} - \mu_0}{\frac{\delta}{\sqrt{n}}}$



4) $w\ (u \leq u_{1-\frac{\alpha}{2}} \quad , \quad u \geq u_{1-\frac{\alpha}{2}})$

$w\ (u \leq u_{0.975} \quad , \quad u \geq u_{0.975})$

$w\ (u \leq -1.96 \quad , \quad u \geq 1.96)$

5) انتخاب نمونه

$k = u = \frac{9.9-10}{\frac{0.5}{\sqrt{25}}} = \frac{-0.1}{\frac{0.5}{5}} = -1$

6) تصمیم گیری

$-1 \nless -1.96 \Rightarrow k \notin w$

چون مقدار عددی ملاک آزمون در ناحیه بحرانی قرار نگرفته است فرضیه H_0 را تا زمانی که خلاف آن اثبات نشده به عنوان فرضیه درست حفظ می‌کنیم.

مثال: (شماره ۵۱ جزوه)

به فرض اینکه بدهی‌های سررسیده یک فروشگاه بزرگ بر طبق قانون نرمال توزیع شده باشد برای ارزیابی بدهی‌های سررسیده فروشگاه نمونه‌ای به حجم ۱۰ فقره حساب انتخاب و میانگین بدهی‌های سررسیده $\bar{X} = 276000$ و $S = 120000$ به دست آمده است اگر معاون فروشگاه با حسابرسی کامل از تمام بدهی‌های سررسیده میانگین بدهی‌های سررسیده را ۲۹۱۰۰۰ تومان به دست آورده باشد در این مورد چه قضاوتی خواهید کرد؟

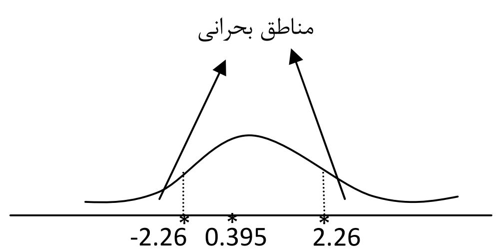
$x \sim N\ (\mu = ?\ ,\ S^2 = ?)$

$H_0 = \mu = 291000$

$n = 10$

$\bar{X} = 276000$

$S = 120000$



1) $H_0: \mu = 291000$

$H_1: \mu \neq 291000$

2) $\alpha = 0.05$

3) $k = T = \frac{\bar{X} - \mu_0}{\frac{\delta}{\sqrt{n}}}$

4) $w\ (T \leq -t_{1-\frac{\alpha}{2}, n-1} \quad , \quad T \geq t_{1-\frac{\alpha}{2}, n-1})$

$w\ (T \leq -t_{0.975;9} \quad , \quad T \geq t_{0.975;9})$

$w\ (t \leq -2.26 \quad , \quad t \geq 2.26)$

5) انتخاب نمونه

$k = T = \frac{276000-291000}{\frac{120000}{\sqrt{10}}} = \frac{-15}{\frac{120}{3.1}} = \frac{-15\times 3.1}{120} = \frac{-3.1}{8} = -0.395$

6) تصمیم گیری

$-0.395 \nless -2.26 \Rightarrow k \notin w$

چون مقدار عددی ملاک آزمون در ناحیه بحرانی قرار نگرفته است فرضیه H_0 را تا زمانی که خلاف آن اثبات نشده به عنوان فرضیه درست حفظ می‌کنیم.

مثال: (شماره ۵۲ جزوه)

در یک موسسه تولیدی طی فرآیند تولید یک نوع محصول راندمان کار یک نفر کارگر در یک ساعت ۱۲ تعیین شده است به منظور بالا بردن راندمان تولید دفتر فنی موسسه تغییراتی در خط تولید داده است. بر طبق ادعایی دفتر فنی موسسه تغییرات داده شده باید موثر واقع گردد و راندمان کارگران کارگران افزایش یابد. به این منظور کارایی تغییرات داده شده تصمیم گرفته شد ۴ کارگر را از بین کارگرانی که دز این خط تولید کارمی کنند بعنوان نمونه انتخاب کنند و راندمان کار آن‌ها را تعیین نمایند تا مطمئن شوند که راندمان کار آن‌ها بالا رفته است و یا اینکه این فقط یک دعا بوده است؟ به فرض اینکه در نمونه انتخاب شده راندمان کار در یک ساعت برای یک فرد به قرار زیر به دست آمده باشد این فرضیه را آزمون کنید. $X_i = 14, 15, 13, 14$

$X_i = 14, 15, 13, 14$

$H_0 = \mu = 12$

$x \sim N(\mu = ?\ , S^2 = ?)$

1) $H_0: \mu = 12$

$H_1: \mu > 12$

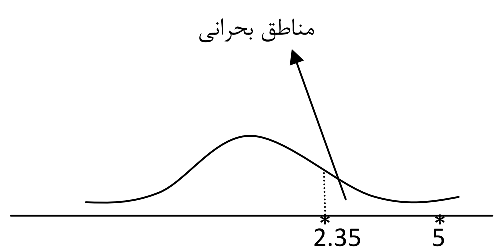
2) $\alpha = 0.05$

3) $k = T = \dfrac{\bar{X} - \mu_0}{\frac{\delta}{\sqrt{n}}}$

4) $w\ (T \geq t_{1-\frac{\alpha}{2}, n-1})$

$w\ (T \geq t_{0.975;3})$

$w\ (t \geq 2.35)$



5) انتخاب نمونه

$\overline{X} = \dfrac{\Sigma x_i}{n} = \dfrac{14+15+13+14}{4} = 14$

$S^2 = \dfrac{\Sigma\,(x_i - \bar{x})^2}{n-1} = \dfrac{0+1+1+0}{3} = \dfrac{2}{3} = 0.66$

$k = T = \dfrac{14-12}{\frac{\sqrt{0.66}}{\sqrt{4}}} = \dfrac{2}{\frac{0.8}{2}} = \dfrac{4}{0.8} = 5$

6) تصمیم گیری

$5 > 2.35 \Longrightarrow k \in w$

چون مقدار عددی ملاک آزمون در ناحیه بحرانی قرار گرفته است فرضیه H_0 رد و فرضیه مقابل پذیرفته می‌شود.

مثال: (شماره ۵۳ جزوه)

سطح زیر کشت حبوبات در استانی متغیر تصادفی با توزیع نرمال است. ۵ مزرعه از این استان را به صورت تصادفی انتخاب نموده‌ایم که سطح زیر کشت حبوبات آن‌ها در سال قبل به صورت زیر بوده است:

$X_i = 1, 1, 0.5, 2, 0.5$

مسئولین وزارت کشاورزی معتقدند میانگین سطح زیر کشت حبوبات در این استان نسبت به سال قبل کمتر از ۱/۵ هکتار بوده است. آیا با نظر آن‌ها با توجه به داده‌های فوق موافقید؟

$X_i = 1, 1, 0.5, 2, 0.5$

$H_0 = \mu = 1.5$

$x \sim N(\mu = ?, S^2 = ?)$

$n = 5$

$\alpha = 0.05$

1) $H_0: \mu = 1.5$

$H_1: \mu < 1.5$

2) $\alpha = 0.05$

3) $k = T = \frac{\bar{X} - \mu_0}{\frac{\delta}{\sqrt{n}}}$

4) $w\ (T <- t_{1-\alpha, n-1})$

$w\ (T <- t_{0.95\,;\,4})$

$w\ (t \le -2.13)$

5) انتخاب نمونه

$\overline{X} = \frac{\Sigma x_i}{n} = \frac{5}{5} = 1$

$S^2 = \frac{\Sigma (x_i - \overline{x})^2}{n-1} = \frac{0+0+0.25+1+0.25}{4} = \frac{1.5}{4} = 0.375$

$k = T = \frac{1-1.5}{\frac{0.6}{\sqrt{5}}} = \frac{-0.5}{\frac{0.6}{2.2}} = \frac{-1.1}{0.6} = -1.83$

6) تصمیم گیری

$-1.83 \nless -2.13 \Longrightarrow k \in w$

چون مقدار عددی ملاک آزمون در ناحیه بحرانی قرار گرفته است فرضیه H_0 رد و فرضیه مقابل پذیرفته می‌شود.

مثال: (شماره ۵۴ جزوه)

تاکنون میانگین مدت زمان لازم جهت ثبت نام یک دانشجو در دانشکده‌ای ۳۰ دقیقه بوده است. اخیراً دانشکده به روش جدیدی ثبت نام می‌نماید به صورتی که در یک نمونه تصادفی ۲۰ تایی از دانشجویان که به روش جدید ثبت نام کرده‌اند نتایج به صورت زیر بدست آمده است. $\bar{X} = 25$ دقیقه و $S = 8$ دقیقه اگر مدت زمانی ثبت نام از توزیع نرمال پیروی کند آیا می‌توان گفت که روش جدید بهتر از روش قدیم است. ($\alpha = 0.05$)

$H_0 = \mu = 30$

$x \sim N(\quad)$

$n = 20$

$\bar{X} = 25$

$S = 8$

$\alpha = 0.05$

1) $H_0: \mu = 30$

$H_1: \mu < 30$

2) $\alpha = 0.05$

3) $k = T = \frac{\bar{X} - \mu_0}{\frac{\delta}{\sqrt{n}}}$

4) $w\ (T <- t_{1-\alpha, n-1})$

$w\ (T <- t_{0.95\,;\,19})$

$w\ (t \leq -1.72)$

5) انتخاب نمونه

$$k = T = \frac{25-30}{\frac{8}{\sqrt{20}}} = \frac{-5}{\frac{8}{4.47}} = \frac{4.47(-5)}{8} = -2.79$$

6) تصمیم گیری

$-2.79 < -2.13 \Longrightarrow k \in w$

چون مقدار عددی ملاک آزمون در ناحیه بحرانی قرار گرفته است فرضیه H_0 رد و فرضیه مقابل پذیرفته می‌شود.

مثال: (شماره ۵۸ جزوه)

مدیر تولید یکی از برنامه‌های خاص تلویزیونی مدعی است که نسبت بینندگان علاقه‌مند به این برنامه ۷۰ درصد می‌باشد. به منظور آزمون این ادعا نمونه‌ای به حجم $n = 400$ خانوار به طور تصادفی انتخاب شد و نظر آن‌ها را نسبت به برنامه خاص پرسش نمودیم معلوم گردید که ۲۶۰ خانوار این برنامه را تماشا می‌کنند. مطلوب است آزمون ادعای فوق در سطح احتمال $\alpha = 0.05$

$H_0 = p = 0.7$

$x \sim N(\quad)$

$n = 400$

$m = 260$

$\alpha = 0.05$

1) $H_0: p = 0.7$

$H_1: p \neq 0.7$

2) $\alpha = 0.05$

3) $k = u = \frac{\hat{p} - p_0}{\sqrt{\frac{p.q}{n}}}$

$\widehat{P} = \frac{m}{n} = \frac{260}{400} = 0.65$

$\hat{q} = 1 - \widehat{P} = 1 - 0.65 = 0.35$

4) $w\,(u \leq -u_{1-\frac{\alpha}{2}}\ ,\ u \geq u_{1-\frac{\alpha}{2}})$

$w\,(u \leq -1.96\ ,\ u \geq 1.96)$

5) انتخاب نمونه

$k = u = \frac{0.65-0.7}{\sqrt{\frac{0.7\times0.3}{400}}} = \frac{-0.05}{\sqrt{\frac{0.41}{20}}} = -2.2$

6) تصمیم گیری

$-2.2 < -1.96 \Rightarrow k \in w$

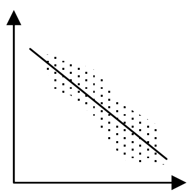
چون مقدار عددی ملاک آزمون در ناحیه بحرانی قرار گرفته است فرضیه H_0 رد و فرضیه مقابل پذیرفته می‌شود.

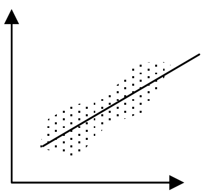
رگرسیون

در رگرسیون ما می‌خواهیم ارتباط را بسازیم نقطه‌هایی که روی یک خط نیستند بلکه حول و حوش خط مورد نظر ما هستند که ما می‌خواهیم آن خط را پیدا کنیم.

معادلات رگرسیون

- معادله رگرسیون y بر حسب x ← $y = a + bx$ (b ← شیب خط و a ← عرض از مبدأ)
- معادله رگرسیون x بر حسب y ← $x = a' + b'y$





کوواریانس y و x جهت تغییرات را نشان می‌دهد و مشخص می‌کند که ارتباط مستقیم (مثبت) یا ارتباط معکوس (منفی) است

$$\text{Cov } x,y = \frac{\Sigma (x_i-\bar{x})(y_i-\bar{y})}{n} \quad \text{یا} \quad = \frac{\Sigma xy - \frac{1}{n}\Sigma x \, \Sigma y}{n}$$

$$b = \frac{\text{Cov } x,y}{\text{واریانس صفت } x} = \frac{\Sigma (x_i-\bar{x})(y_i-\bar{y})}{\Sigma (x_i-\bar{x})^2} \quad \text{یا} \quad = \frac{\Sigma xy - \frac{1}{n}\Sigma x \, \Sigma y}{\Sigma x^2 - \frac{1}{n}(\Sigma x)^2}$$

$$a = \bar{y} - b\bar{x}$$

$$b' = \frac{\text{Cov } x,y}{\text{واریانس صفت } y} = \frac{\Sigma (x_i-\bar{x})(y_i-\bar{y})}{\Sigma (y_i-\bar{y})^2} \quad \text{یا} \quad = \frac{\Sigma xy - \frac{1}{n}\Sigma x \, \Sigma y}{\Sigma y^2 - \frac{1}{n}(\Sigma y)^2}$$

$$a' = \bar{x} - b'\bar{y}$$

روش محاسبه ضریب همبستگی (r)

فرمول ضریب همبستگی: $r = \sqrt{b \times b'}$

اگر r برابر با صفر باشد همبستگی وجود ندارد.

اگر r برابر با +۱ باشد همبستگی (ارتباط) کامل و مستقیم است.

اگر r برابر با -۱ باشد همبستگی (ارتباط) کامل و معکوس است.

اگر r برابر با بین صفر و یک باشد همبستگی (ارتباط) ناقص و مستقیم است. $0 < r < 1$

اگر r برابر با بین -۱ و صفر باشد همبستگی (ارتباط) ناقص و معکوس است. $-1 < r < 0$

مثال: (شماره ۹۰ جزوه)

میزان سرمایه گذاری (x) در بخش تحقیق و توسعه و میزان سودآوری سالانه (y) حاصل از آن طی یک نمونه تصادفی ۶ ساله به صورت جدول زیر بوده است:

فرض کنید x و y از توزیع نرمال دو متغیره پیروی می‌کند. مطلوب است تخمین معادله خط رگرسیون پیش بینی سود سالانه بر حسب سرمایه گذاری در این بخش.

x	y
5	31
11	40
4	30
5	34
3	25
2	20

x	y	xy	x^2	y^2
5	31	155	25	961
11	40	440	121	1600
4	30	120	16	900
5	34	170	25	1156
3	25	75	9	625
2	20	40	4	400
Σx	Σy	Σxy	Σx^2	Σy^2
30	180	1000	200	5642

$$b = \frac{\Sigma xy - \frac{1}{n}\Sigma x\,\Sigma y}{\Sigma x^2 - \frac{1}{n}(\Sigma x)^2} = \frac{1000 - \frac{1}{6}\times 30 \times 180}{200 - \frac{1}{6}\times(30)^2} = \frac{100}{50} = 2$$

$$a = \bar{y} - b\bar{x} = \frac{\Sigma y_i}{n} - \frac{\Sigma x_i}{n} = \frac{180}{6} - 2\left(\frac{30}{6}\right) = 30 - 10 = 20$$

$y = a + bx \Longrightarrow \boxed{y = 20 + 2x} \leftarrow$ معادله رگرسیون بر حسب x

$$b` = \frac{\Sigma xy - \frac{1}{n}\Sigma x\,\Sigma y}{\Sigma y^2 - \frac{1}{n}(\Sigma y)^2} = \frac{1000 - \frac{1}{6}\times 30 \times 180}{5642 - \frac{1}{6}\times(180)^2} = \frac{100}{242} = 0.41$$

$$a` = \bar{x} - b`\bar{y} \Longrightarrow 5 - 0.41(30) \longrightarrow a` = 5 - 12.3 = -7.3$$

$x = a` + b`y \Longrightarrow \boxed{x = -7.3 + 0.41y} \leftarrow$ معادله رگرسیون بر حسب y

مثال: (شماره ۹۱ جزوه)

به منظور بررسی اثر درآمد (x) به میزان خرید کالای لوکس (y) نمونه‌ای به حجم ۵ از یک منطقه شهری انتخاب گردید که نتایج آن به صورت جدول زیر بدست آمده است:

فرض کنید x و y از قانون نرمال دو متغیره پیروی می‌کند. مطلوب است تخمین معادله رگرسیون y بر حسب x و x بر حسب y. و آیا بین درآمد و میزان خرید کالای لوکس ارتباط خطی برقرار است؟

x	y
2	1
4	3
5	4
7	6
9	8

x	y	xy	x^2	y^2
2	1	2	4	1
4	3	12	16	9
5	4	20	25	16
7	6	42	49	36
9	8	72	81	64
Σx 27	Σy 22	Σxy 148	Σx^2 175	Σy^2 126

$$b = \frac{\Sigma xy - \frac{1}{n}\Sigma x\,\Sigma y}{\Sigma x^2 - \frac{1}{n}(\Sigma x)^2} = \frac{148 - \frac{1}{5}\times 27\times 22}{175 - \frac{1}{5}\times(27)^2} = \frac{29.2}{29.2} = 1$$

$$a = \bar{y} - b\bar{x} = \frac{\Sigma y_i}{n} - \frac{\Sigma x_i}{n} = \frac{25}{5} - 1\left(\frac{27}{5}\right) = 4.4 - 5.4 = -1$$

$y = a + bx \Longrightarrow \boxed{y = -1 + x}$ ← معادله رگرسیون بر حسب x

$$b' = \frac{\Sigma xy - \frac{1}{n}\Sigma x\,\Sigma y}{\Sigma y^2 - \frac{1}{n}(\Sigma y)^2} = \frac{1000 - \frac{1}{6}\times 30\times 180}{126 - \frac{1}{6}\times(22)^2} = \frac{29.2}{29.2} = 1$$

$$a' = \bar{x} - b'\bar{y} \Longrightarrow 5.4 - 1(4.4) \longrightarrow a' = 1$$

$x = a' + b'y \Longrightarrow \boxed{x = 1 + y}$ ← معادله رگرسیون بر حسب y

$r = \sqrt{b \times b'} = \sqrt{1 \times 1} \longrightarrow \boxed{r = 1}$ ⟵ همبستگی (ارتباط) کامل و مستقیم است

مثال: (شماره ۹۸ جزوه)

در یک رشته از فعالیت اقتصادی جهت ارتباط بین مصرف انرژی (x) و میزان تولید (y) از n = 15 واحد تولیدی که به صورت تصادفی انتخاب شدند اطلاعات زیر بدست آمده:

$\Sigma x = 67.5$

$\Sigma y = 142.5$

$\Sigma (x_i - \bar{x})^2 = 312$

$\Sigma (y_i - \bar{y})^2 = 60$

$\Sigma (x_i - \bar{x})(y_i - \bar{y}) = 42$

فرض کنید (x, y) از توزیع نرمال دو متغیره پیروی می‌کنند. مطلوب است مطالعه رگرسیون خطی y بر حسب x

$\Sigma x = 67.5$

$\Sigma y = 142.5$

$\Sigma (x_i - \bar{x})^2 = 312$

$\Sigma (y_i - \bar{y})^2 = 60$

$\Sigma (x_i - \bar{x})(y_i - \bar{y}) = 42$

$$b = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{42}{312} = 0.134$$

$$a = \bar{y} - b\bar{x} = \frac{\Sigma y_i}{n} - \frac{\Sigma x_i}{n} = \frac{142.5}{15} - 0.134\left(\frac{67.5}{15}\right)$$

مثال: (نمونه سؤالات امتحانی)

یک کارخانه‌ی مواد شیمیایی به طوری طراحی شده است روزانه به طور متوسط ۸۰۰ تن محصول داشته باشد. در ۵ روز متوالی محصول این کارخانه بر حسب تن به شرح زیر بوده است: ۸۰۲، ۷۹۳، ۷۹۰، ۸۰۵، ۷۸۵. با فرض نرمال بودن میزان محصول آیا در سطح معنا داری ۵٪ این داده نشان دهنده‌ی کاهش در مقدار متوسط میزان محصول روزانه کارخانه است؟ (آزمون فرض)

$x \sim N(\mu = ?\,,\, S^2 = ?)$
$H_0 = \mu = 800$
$n = 5$

1) $H_0: \mu = 800$
$H_1: \mu < 800$
2) $\alpha = 0.05$
3) $k = T = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$
4) $w\,(T \leq -t_{1-\frac{\alpha}{2},n-1})$
$w\,(T \leq -t_{0.95;4})$
$w\,(t \leq -2.13)$
5) انتخاب نمونه

$$\bar{X} = \frac{\Sigma x_i}{n} = \frac{802+793+790+805+785}{5} = \frac{3975}{5} = 795$$

$$S^2 = \frac{\Sigma\,(x_i - \bar{x})^2}{n-1} = \frac{49+4+25+100+100}{4} = \frac{278}{4} = 69.5$$

$$S = \sqrt{s^2} = \sqrt{69.5} = 8.3$$

$$k = T = \frac{795-800}{\frac{8.3}{\sqrt{5}}} = \frac{-0.5}{\frac{8.3}{2.2}} = \frac{-11}{8.3} = \frac{-115}{83} - 1.3$$

6) تصمیم گیری

$-1.3 \nless -2.13 \Longrightarrow k \notin w$

چون مقدار عددی ملاک آزمون در ناحیه بحرانی قرار نگرفته است فرضیه H_0 را تا زمانی که خلاف آن اثبات نشده به عنوان فرضیه درست حفظ می‌کنیم.

مثال: (نمونه سؤالات امتحانی)

در یک نظرخواهی در مورد یک طرح جدید آموزشی از ۴۵۰ نفری که مورد سؤال قرار گرفته ۳۷۸ نفر موافق اجرای طرح بودند. یک فاصله اطمینان ۹۵٪ برای نسبت واقعی افرادی که موافق اجرای این طرح جدید هستند را بدست آورید. (تخمین فاصله‌ای نسبت p)

$n = 450$
$\alpha = 0.95$
$m = 378$
$\underline{L}\,,\, \overline{L}_{(p)} = ?$

$\widehat{P} = \frac{m}{n} = \frac{378}{450} = 0.84$

$\widehat{q} = 1 - \widehat{P} = 1 - 0.84 = 0.16$

$\alpha = 0.95 \rightarrow 1- \alpha = 0.005 \rightarrow \frac{\alpha}{2} = \frac{0.05}{2} = 0.025 \rightarrow 1 - \frac{\alpha}{2} = 1- 0.025 = 0.975$

$u_{1-\frac{\alpha}{2}} = u_{0.975} = 1.96$

$\widehat{P} - \sqrt{\frac{\widehat{P}\,\widehat{q}}{n}}\, u_{1-\frac{\alpha}{2}} \leq p \leq \widehat{P} + \sqrt{\frac{\widehat{P}\,\widehat{q}}{n}}\, u_{1-\frac{\alpha}{2}}$

$0.84 - \sqrt{\frac{0.84\times0.16}{450}} \times 1.96 \leq p \leq 0.84 + \sqrt{\frac{0.84\times0.16}{450}} \times 1.96$

مثال: (نمونه سؤالات امتحانی)

۲۰٪ دانشجویان دانشکده مدیریت در رشته مدیریت بازرگانی ۵٪ درصد در رشته مدیریت دولتی و بقیه در رشته مدیریت صنعتی تحصیل می‌کنند. ۴۰٪ از دانشجویان رشته مدیریت صنعتی‌ترمی هستند. دانشجویی را به طور تصادفی از این دانشکده به طور تصادفی انتخاب می‌کنیم. اگر بدانیم که او بومی است احتمال اینکه رشته‌اش مدیریت صنعتی باشد چقدر است؟

H_i	$P(H_i)$	$P(A\|H_i)$	$P(H_i)\,P(A\|H_i)$
بازرگانی H_1	0.2	0.4	0.08
دولتی H_2	0.5	0.25	0.125
صنعتی H_3	0.3	0.3	0.09
	1		$\Sigma P(H_i)\,P(A\|H_i)= 0.295$

$P(H_3|A) = \frac{0.09}{0.295} = \frac{90}{295}$

مثال: (نمونه سؤالات امتحانی)

دو پیشامد A و B در یک فضای نمونه بر اساس آن‌ها احتمال‌های زی بدست آمده است:

$P(A\cup B) = \frac{7}{8}$. $P(A\cap B) = \frac{1}{8}$, $P(\overline{A}) = \frac{5}{8}$

مطلوب است محاسبه احتمال هر یک از پیشامدهای زیر:

الف: $P(A)$

ب: $P(B)$

ج: $P(B|A)$

الف) $P(\overline{A}) = 1 - P(A) \rightarrow P(A) = 1 - \frac{5}{8} = \frac{3}{8}$

ب) $P(B|A)= P(A) + P(B) - P(A\cap B) \rightarrow \frac{7}{8} = \frac{3}{8} + P(B) - \frac{1}{4}$

$P(B) = \frac{7}{8} - \frac{1}{8} = \frac{6}{8}$

ب) $P(B|A)= \frac{\frac{1}{4}}{\frac{3}{8}} = \frac{2}{3}$

مثال: (نمونه سؤالات امتحانی)

از جعبه‌ای که ۲۰ ساعت در ان وجود دارد و ۵ تای آن‌ها خراب است ۴ ساعت را بصورت تصادفی بیرون می‌آوریم. متغیر تصادفی x نشانگر تعداد ساعت‌های خراب در نمونه چهارتایی است. مطلوب است:

الف: قانون توزیع متغیر تصادفی.

ب: میانگین و انحراف معیار.

ج: احتمال اینکه ساعت خرابی در نمونه دیده نشود.

د: احتمال اینکه ساعت سالمی در نمونه دیده نشود.

$N = 20$

$n = 4$

$m = 5$

الف) $p(X=x) = \frac{C_m^x C_{N-m}^{n-x}}{C_N^n} \rightarrow P(X = x) = \frac{C_5^x C_{16}^{4-x}}{C_{20}^4}$

ب) احتمال خراب بودن $= P(A) = \frac{5}{20} = \frac{1}{4}$

$E(x) = np = 4 \times \frac{1}{4} = 1$

$D(x) = npq = 1 \times \frac{2}{3} = \frac{3}{4} \rightarrow q = \frac{3}{4} \rightarrow \sqrt{D(x)} = \sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{2}$

ج) $P(x = 0) = \frac{C_5^0 C_{15}^4}{C_{20}^4}$ و د) $P(1 - x = 0) = \frac{C_5^4 C_5^0}{C_{20}^4}$

مثال: (نمونه سؤالات امتحانی)

یک شرکت تولیدی قصد دارد تعداد سفارش‌های رسیده (y) را با توجه به هزینه تبلیغات (x) محصولش را بررسی کند. جدول زیر نتایج حاصل از مطالعات یک نمونه تصادفی ۵تایی را نشان می‌دهد. با فرض این که x و y از توزیع نرمال دو متغیره پیروی می‌کند مطلب وب است:

الف: معادله خط رگرسیون تعداد سفارش بر حسب هزینه تبلیغات. (y بر حسب x)

ب: ضریب همبستگی بین دو متغیر

x	y	xy	x^2	y^2
17	35	595	289	1225
21	42	882	441	1765
25	60	1500	625	3600
18	62	1116	324	3844
19	52	988	361	2704
Σx 100	Σy 251	Σxy 5081	Σx^2 2040	Σy^2 13137

$b = \frac{\Sigma xy - \frac{1}{n}\Sigma x \Sigma y}{\Sigma x^2 - \frac{1}{n}(\Sigma x)^2} = \frac{5081 - \frac{1}{5} \times 100 \times 251}{2040 - \frac{1}{5} \times (100)^2} = \frac{61}{40} = 1.525$

$b` = \frac{\Sigma xy - \frac{1}{n}\Sigma x \Sigma y}{\Sigma y^2 - \frac{1}{n}(\Sigma y)^2} = \frac{5081 - \frac{1}{5} \times 100 \times 251}{13137 - \frac{1}{5} \times (251)^2} = \frac{61}{536} = 1.113$

$a = \bar{y} - b\bar{x} = \frac{\Sigma y_i}{n} - \frac{\Sigma x_i}{n} = \frac{251}{5} - 1.525\left(\frac{100}{5}\right) = 19.7 \rightarrow y = a + bx \Longrightarrow \boxed{y = 19.7 + 1.525x}$

$r = \sqrt{b \times b`} = \sqrt{1.525 \times 1.113} \longrightarrow \boxed{r = 0.415}$ ⟵ همبستگی (ارتباط) ناقص و مستقیم است

پایان جزوه

با آرزوی توفیق الهی