فصل دوم
آمار و احتمالات

1-2: مبانی تئوری آمار و احتمال

1-1-2: تعریف آمار

آمار به عنوان یک موضوع علمی، امروزه شامل مفاهیم و روش‌هایی است که مستلزم جمع آوری داده‌ها (به وسیلهٔ یک فرآیند آزمایش و مشاهده) و انجام استنباط و نتیجه‌گیری (به وسیلهٔ تجزیه و تحلیل این داده‌ها) هستند. آمار هنر و علم جمع آوری، تعبیر و تجزیه و تحلیل داده‌ها و استخراج تعمیم‌های منطقی در مورد پدیده‌های تحت بررسی است.

2-1-2: تعریف احتمال

بطور ساده، احتمالات به شانس وقوع یک حادثه گفته می‌شود. احتمال معمولا برای توصیف نگرش ذهن نسبت به گزاره‌هایی که ما از حقیقت آن‌ها مطمئن نیستیم، مورد استفاده قرار می‌گیرد. گزاره‌های موردنظر معمولا به فرم "آیا یک رویداد خاص رخ می-دهد؟" و نگرش ذهن ما به فرم "چقدر اطمینان داریم که این رویداد رخ خواهد داد؟" است. میزان اطمینان ما، قابل توصیف به صورت عددی می‌باشد که این عدد مقداری بین 0 و 1 را گرفته و آن را احتمال می‌نامیم. هر چه احتمال یک رویداد بیشتر باشد، ما مطمئن‌تر خواهیم بود که آن رویداد رخ خواهد داد. در واقع میزان اطمینان ما از این‌که یک واقعه (تصادفی) اتفاق خواهد افتاد را احتمال وقوع آن می‌نامند.

اگر تعداد تکرارهای یک آزمایش را n در نظر بگیریم، آن‌گاه احتمال وقوع واقعه E، به صورت زیر تعریف می‌شود:

$$P(E) = \frac{\text{number of times E occures}}{n}$$

با توجه به پیشامد P(E) داریم:

الف) $0 = (0)$ (احتمال یک پیشامد غیر ممکن برابر صفر است)

ب) $P(S) = 1$ (احتمال یک پیشامد قطعی برابر یک است)

ج) $0 \leq (\ ) \leq 1$

باید توجه داشت که در عمل، پیشامد غیرممکن وجود نداشته و پیشامدهایی که احتمال وقوع آن‌ها به صفر نزدیک باشد، به عنوان پیشامد غیرممکن درنظر گرفته می‌شود. به همین ترتیب پیشامدهایی که احتمال وقوع آن‌ها به 1 نزدیک باشند، به عنوان پیشامدهای قطعی تلقی خواهند شد.

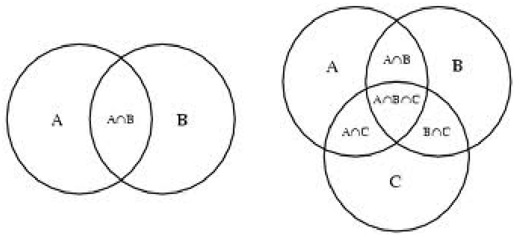
پیشامدهای A و ... ، A ، A ناسازگارند اگر احتمال وقوع هم‌زمان دو پیشامد وجود نداشته باشد. اگر پیشامدهای A و ... ، A ، A همه نتایج ممکن را در برگیرند، یعنی امکان نداشته باشد که در اثر آزمایش هیچ‌کدام از A ، A ..... یا A رخ ندهد، در اینصورت پیشامد را فرسا گوئیم.

اگر شرایط آزمایش به گونه‌ای باشد که احتمال وقوع هر یک از پیشامدهای A و ... ، A ، A برابر باشند، در این صورت پیشامد را هم‌تراز یا هم‌شانس گوئیم. اگر پیشامدهای A و ... ، A ، A هر سه خاصیت، ناسازگار، فرسا و هم‌ترازی را داشته باشند، در نظریه احتمال آن‌ها را حالت‌ها یا شانس‌ها نامیده و می‌گوئیم توسط مدل کلاسیک بیان می‌شوند. قبل از آن که وارد بحث‌های آمار و احتمال شویم، باید با مجموعه‌ها که در آینده با آن مواجه هستیم، آشنا شوید.

2-1-3: قوانین مجموعه‌ها

مجموعه به دسته‌ای از عناصر گفته می‌شود که به سبب شباهت در داشتن یک ویژگی یا خصوصیت در کنار یکدیگر قرار گرفته و مجموعه را تشکیل می‌دهند. اشتراک دو مجموعه A و B، مجموعه‌ای است که در آن اعضایی از دو مجموعه انتخاب می‌شوند که در هر دو مجموعه وجود داشته باشند. اشتراک دو مجموعه به صورت $A \cap B$ یا ندرتا AB نشان داده می‌شود. در مقابل اجتماع دو مجموعه به مجموعه‌ای گفته می‌شود که شامل همه اعضایی است که حداقل در یکی از مجموعه‌های A یا B حضور داشته باشند و با $A \cup B$ یا ندرتا $A + B$ نشان داده می‌شود.

مکمل مجموعه A، شامل عناصری است که در مجموعه A وجود ندارند. مکمل مجموعه A به صورت A نشان داده می‌شوند. نمودار ون به طور مناسب تمام این مفاهیم را به تصویر می‌کشد.



شکل 2-1: نمودار ون اجتماع و اشتراک دومجموعه

قوانین:

$$A(B \cup C) = AB \cup AC$$

$$A \cup (BC) = (A \cup B)(A \cup C)$$

قوانین دمورگان:

$$\cup \quad = \bar{} \cap \qquad\qquad =$$

$$\cap \quad = \bar{} \cup \qquad\qquad =$$

مهم است که بتوان توصیف یک مجموعه را با استفاده از علامت‌های سمبلیکی که در بالا به آن‌ها اشاره شد، بیان کرد.

2-1-4: احتمال و مجموعه‌ها

همان‌گونه که قبلا اشاره شد، پیشامدهای A و ... ، A ، A را پیشامدهای ناسازگار گوئیم اگر وقوع هم‌زمان دو پیشامد غیرممکن باشد. به عبارت دیگر

$$\cap \quad = \emptyset \qquad \neq \ ,1 \leq \ , \ \leq$$

و در نتیجه

$$\cap \qquad = 0$$

اگر A و B دو پیشامد باشند، آن‌گاه

$$(\cup) = ( ) + ( ) - (\cap)$$

$$= (\cup \cup \ldots)$$

$$= ( ) - \cap + \cap \cap$$

$$- \cdots .. (-1) (\cap \ldots \cap)$$

اگر پیشامدهای A و ... ، A ، A دو به دو ناسازگار باشند، داریم:

$$= ( )$$

اگر A و A دو پیشامد متمم باشند، آن‌گاه

$$(\bar{}) = 1 - ( )$$

1-4-1-2: احتمال شرطی

در بسیاری از مسائل، آگاهی از رخ دادن پیشامدی، در محاسبه احتمال رخ دادن پیشامدهای دیگر کمک موثری می‌کند. فرض کنید تاس سالمی را پرتاب کرده‌اید و از دوست خود می‌پرسید، احتمال آمدن 3 چقدر است، مسلما خواهد گفت 1/6. حال به او بگوئید عدد ظاهر شده فرد است در این صورت احتمال آمدن عدد 3 چقدر است، حتما خواهد گفت 1/3 .

احتمال رخ دادن پیشامد B، مشروط بر آن‌که بدانیم پیشامد A رخ داده است را احتمال شرطی گوئیم و با نماد $P(B|A)$ نشان می‌دهیم. مقدار احتمال شرطی به صورت زیر محاسبه می‌شود.

احتمال پیشامد B به شرط آن‌که پیشامد A رخ داده باشد، عبارت است از:

$$(\quad | \quad) = \frac{(\quad \cap \quad)}{(\quad)}$$

در صورتی که $0 < (\quad)$ باشد.

مثال) از جعبه‌ای که محتوی 4 مهره سفید و 3 مهره سیاه است، دو مهره بدون جایگذاری خارج می‌کنیم، مطلوبست محاسبه احتمال این‌که:

الف) هر دو مهره سفید باشد.
ب) هم رنگ نباشند.

: پیشامد این‌که توپ اول سفید باشد.

: پیشامد این‌که توپ دوم سفید باشد.

: پیشامد این‌که توپ اول سیاه باشد.

: پیشامد این‌که توپ دوم سیاه باشد.

الف) باید احتمال پیشامد $\quad \cap \quad$ را محاسبه کنیم، داریم:

$$(\quad \cap \quad) = (\quad) \; (\quad | \quad) \,, \qquad (\quad) = \frac{4}{7} \,,$$

$$(\quad | \quad) = \frac{3}{6} \,, \quad (\quad \cap \quad) = \frac{4}{7} \times \frac{3}{6}$$

ب) باید احتمال پیشامد زیر را حساب کنیم

$$(\quad \cap \quad) \cup (\quad \cap \quad)$$

چون پیشامدهای $\quad \cap \quad$ و $\quad \cap \quad$ ناسازگارند، پس

$$(\quad \cap \quad) \cup (\quad \cap \quad) = (\quad \cap \quad) + (\quad \cap \quad)$$

$$(\quad \cap \quad) = (\quad) \; (\quad | \quad) = \frac{4}{7} \times \frac{3}{6}$$

$$(\quad \cap \quad) = \ (\quad)\ (\quad | \quad)$$

در نتیجه احتمال قسمت ب برابر – + – = – است.

دو پیشامد A و B را مستقل گوئیم، اگر و تنها اگر

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

به عبارت دیگر، اگر رخ دادن پیشامد A، اثری در رخ دادن یا ندادن پیشامد B نداشته باشد، در این‌صورت دو پیشامد A و B را مستقل گوئیم.

قضیه بیز: فرض کنید مجموعه $\{ B, \ldots, B, \quad \}$ یک افراز از فضای نمونه S باشد. اگر C پیشامد دلخواه از S باشد و $P(C) \neq 0$، آن‌گاه به ازای $K=1,2,\ldots,n$ داریم:

$$P(B \quad |C) = \frac{(\quad \cap \quad)}{\Sigma \qquad (\quad \cap \quad)} = \frac{(\quad)\ (\quad | \quad)}{\Sigma \qquad (\quad)\ (\quad | \quad)}$$

مثال) کارخانه‌ای دارای سه ماشین است که به ترتیب 50%، 30%، 20% محصول را تولید می‌کنند. می‌دانیم درصد کالاهای معیوب این سه ماشین به ترتیب 3%، 4%، 5% است. مطلوبست محاسبه احتمال این‌که:

الف) اگر کالایی را به تصادف از محصولات کارخانه انتخاب کنیم، معیوب باشد.

ب) اگر کالای انتخاب شده معیوب باشد، این کالا توسط ماشین اول تولید شده باشد.

B_i : پیشامد این‌که کالای انتخاب شده مربوط به ماشین iام باشد. $i=1,2,3$

C : پیشامد این‌که کالای انتخاب شده معیوب باشد.

الف) می‌خواهیم $(\quad)$ را محاسبه کنیم، چون

$$= (\quad \cap \quad) \cup (\quad \cap \quad) \cup (\quad \cap \quad)$$

و پیشامدهای $\cap$ و $\cap$ و $\cap$ ناسازگارند، پس

$$(\quad) = (\quad \cap \quad) + (\quad \cap \quad) + (\quad \cap \quad)$$

$$= \ (\ )\ (\ |\ ) + \ (\ )\ (\ |\ ) + \ (\ )\ (\ |\ )$$

$$= \frac{}{} \times \frac{}{} + \frac{}{} \times \frac{}{} + \frac{}{} \times \frac{}{} = \frac{}{}$$

در نتیجه 3,7% کالای تولید شده معیوب هستند.

ب)

$$P(B\ |C) = \frac{(\ )\ (\ |\ )}{\Sigma\quad (\ )\ (\ |\ )} = \frac{\frac{50}{100}\times\frac{3}{100}}{\frac{37}{1000}}$$

5-1-2: انواع داده‌ها

بسیاری از داده‌هایی که از آزمایشات گوناگون بدست می‌آیند، به صورت اعداد حقیقی ارائه می‌شوند. اگر این داده‌ها از یک آزمایش به آزمایش دیگر تغییر کنند، آن‌گاه به آن‌ها متغیرهای تصادفی می‌گوییم.

1-5-1-2: متغیرهای تصادفی گسسته

در توزیع احتمال گسسته، یک متغیر تصادفی گسسته هر یک از مقادیر خود را با احتمالی معین اختیار می‌کند. اگر متغیر تصادفی X نشان‌دهنده عددی باشد که از پرتاب یک تاس نتیجه می‌شود. در این صورت X هر یک از مقادیر 1,2,3,4,5,6 را با احتمال مساوی 1/6 اختیار می‌کند که آن را به صورت جدول زیر می‌توان نوشت،

X	1	2	3	4	5	6
P(X = x)	1/6	1/6	1/6	1/6	1/6	1/6

و به صورت فرمول زیر بیان نمود:

$$(\ ) = 1/6 \quad , \qquad = 1,2,\ldots,6$$

جدول یا فرمولی که تمام مقادیر متغیر تصادفی را همراه با احتمال‌های مربوطه نشان دهد تابع احتمال می‌نامیم و آن را با ( )، ( ) و ... نشان می‌دهیم.

در حالتی که متغیر تصادفی گسسته باشد، تابع احتمال را توزیع احتمال می‌نامیم.

$$f(x) = P(X = x) = P(x(s) = x) \quad \text{و} \quad s \in S$$

مثال) سکه‌ای را سه‌بار پرتاب می‌کنیم، اگر متغیر تصادفی X نشان‌دهنده تعداد شیر در این آزمایش باشد، توزیع احتمال X را به صورت جدول و فرمول بیان کنید.

فضای نمونه S عبارت است از:

$$S = \{TTT, TTH, THT, HTT, THH, HTH, HHT, HHH\}$$

متغیر تصادفی X مقادیر 0,1,2,3 را با احتمال زیر قبول می‌کند

$$f(0) = P(X = 0) = 1/8 \quad , f(1) = P(X = 1) = 3/8$$

$$f(2) = P(X = 2) = 3/8 \quad , f(3) = P(X = 3) = 1/8$$

X	0	1	2	3
$f(x) = P(X = 2)$	1/8	3/8	3/8	1/8

فرمول زیر جدول بالا را ارائه می‌دهد:

$$(\ ) = \frac{\ }{8} \quad , \quad = 0,1,2,3$$

فضای نمونه دارای 8 نمونه هم‌تراز می‌باشد. از طرفی تعداد حالت‌های نتیجه آزمایش x شیر و $3 - x$ خط است، که در صورت آمده است.

مثال) سکه‌ای طوری اریب شده که احتمال آمدن شیر 2 برابر خط است. به عبارت دیگر احتمال خط آمدن 1/3 است. سکه را آن‌قدر پرتاب می‌کنیم تا خط بیاید، فرض کنید متغیر تصادفی X تعداد دفعات پرتاب باشد. توزیع احتمال X را به صورت جدول و فرمول بیان کنید.

فضای نمونه این آزمایش از پیشامد ساده‌ای تشکیل شده است که هر یک از تعدادی H (شیر) و به دنبال آن‌ها فقط یک T (خط) می‌باشد.

$$= \{\ ,\ \ ,\ \ ,\ \ , \ldots\}$$

توزیع احتمال تعداد آزمایش‌های لازم برای آمدن یک خط:

x	1	2	3	4	.	...
f(x)	1/3	$(2/3)^{1}1/3$	$(2/3)^{2}1/3$	$(2/3)^{3}1/3$	-	-

همان‌طور که ملاحظه می‌کنید، نوشتن کامل جدول غیرممکن است ولی می‌توان آن را با فرمول بیان نمود.

$$(\ ) = (2/3) \quad \times \frac{1}{3}$$

با توجه به مثال‌های بیان شده مشاهده می‌کنید که نوشتن توزیع احتمال به صورت هم جدول و هم فرمول برای یک متغیر تصادفی ممکن است به سادگی امکان‌پذیر نباشد و باید به یکی از این دو اکتفا نمود. همان‌طور که در مثال‌های قبل ملاحظه می‌کنید، همواره $\sum f(x) = 1$ است. تابع f(x) را یک تابع توزیع احتمال برای متغیر تصادفی گسسته X گوئیم (X مقادیر x، x، ... را اختیار می‌کند) اگر

الف) $(\ )0 \geq 0 , \forall$

ب) $\sum f(x\ ) = 1$

پ) $(\ = \ ) = f(x\ )$

در مورد مثال بالا داریم:

$$\sum \quad - (-) \quad = - \sum \quad (-) \quad = - \times \frac{}{/} = - \times 3 = 1$$

2-5-1-2: داده‌های پیوسته

در مورد داده‌های پیوسته، تابع چگالی احتمال به این صورت است که احتمال این‌که یک متغیر تصادفی پیوسته دقیقا یکی از مقادیر خود را اختیار کند برابر صفر می‌باشد، به عبارت دیگر اگر X یک متغیر تصادفی پیوسته باشد، آن‌گاه

$(\ = \ ) = 0$ (یکی از مقادیر متغیر می‌باشد)

این موضوع عجیب به نظر می‌رسد. برای روشن شدن مطلب به بررسی یک مثال ساده می‌پردازیم، فرض کنید متغیر تصادفی X نشان‌دهنده طول قد افراد 21سال به بالا باشد،

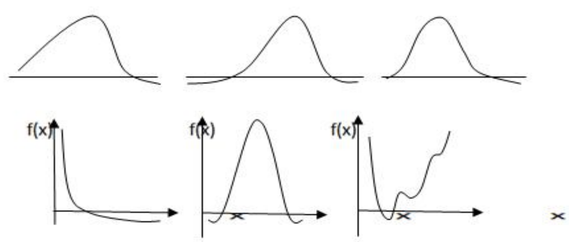
بین هر دو طول قد، مانند 167,5 و 168,5 سانتی‌متر بی‌نهایت طول قد وجود دارد که احتمال این‌که یکی از آن طول قدها دقیقا برابر 168 سانتی‌متر باشد صفر است، پس

$$(\quad <= \quad <= \quad) = (\quad = \quad) + (\quad < \quad < \quad) + (\quad = \quad)$$

$$= \quad (\quad < \quad <= \quad) = (\quad <= \quad < \quad) = (\quad < \quad < \quad)$$

به عبارت دیگر اگر متغیر تصادفی X پیوسته باشد، احتمال در فاصله مطرح خواهد شد و مهم نیست که نقاط انتهایی فاصله منظور نشده باشند.

با توجه به این‌که احتمال در هر نقطه صفر است، لذا امکان نوشتن تابع احتمال به صورت جدول وجود ندارد و فقط با فرمول می‌توان تابع احتمال را مشخص نمود. چون متغیر X روی یک فضای نمونه پیوسته تعریف می‌شود، تابع احتمال ( ) پیوسته بوده و به عنوان مثال به یکی از صورت‌های زیر می‌باشد.



2-2: تابع احتمال f(x) پیوسته

در حالتی‌که متغیر تصادفی پیوسته باشد، تابع احتمال را تابع چگالی احتمال می‌نامیم.

همانطور که اشاره شد، برای متغیر تصادفی پیوسته، احتمال در فاصله معنی دارد و احتمال وقتی که تابع چگالی احتمال مشخص باشد برابر است با مساحت زیر منحنی ناحیه‌ای که واقع در فاصله مورد نظر، بین تابع و محور xها وجود دارد. به عبارت دیگر اگر ( ) تابع احتمال متغیر تصادفی X باشد، آن‌گاه

$$(\quad < \quad < \quad) = \qquad (\;)$$

تابع ( ) را یک تابع چگالی احتمال برای متغیر تصادفی X روی تمام اعداد حقیقی می‌نامیم، اگر داشته باشیم:

$\forall x \epsilon R$, $f(x) \geq 0$

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

$P(a<X<b)=\int f(x)\, dx$

اگر تابع f پیوسته و یا قطعه‌ای باشد، $\int f(x)\, dx$ موجود است.

مثال) متغیر تصادفی X دارای تابع چگالی احتمال زیر می‌باشد.

$$( ) = \begin{cases} & 1 \leq \quad \leq 2 \\ & 2 < \quad < 3 \\ 0 & h \end{cases}$$

مطلوب است محاسبه:

الف) مقدار a

ب) $(\quad > 2)$

پ) $(1/2 < \quad < 3/2)$

با توجه به شرط‌های ارائه شده، داریم:

الف) $\int \quad dx + \int ax\, dx = 1 \quad \Longrightarrow \quad a= 6/29$

ب) $(\quad > 2) = \int \quad ( ) \quad = \int \frac{}{29} \quad = \frac{}{}$

پ) $(1/2 < \quad < 3/2) = \int_{/}^{/} \quad ( ) \quad = \int^{/} \frac{}{} \quad = \frac{}{}$

اگر X یک متغیر تصادفی پیوسته با تابع چگالی احتمال $( )$ باشد، تابع توزیع تجمعی را با $( )$ نشان داده و به صورت زیر تعریف می‌کنیم.

$$( ) = (\quad \leq \quad) = \quad ( )$$

با توجه به تعریف تابع توزیع تجمعی داریم:

$$(\quad \leq \quad \leq \quad) = ( ) - ( )$$

تابع چگالی احتمال را می‌توان از تابع توزیع تجمعی بدست آورد. به عبارت دیگر اگر f تابعی پیوسته باشد، آن‌گاه

$$(\ ) = \ (\ )$$

و اگر f همه جا پیوسته نباشد، فرمول بالا به ازای xهایی که f در آن‌ها پیوسته است، برقرار می‌باشد.

مثال) اگر متغیر تصادفی دارای تابع چگالی احتمال ( ) باشد.

الف) تابع توزیع تجمعی آن را بنویسید.

ب) $(2 \leq \ > 1)$ را حساب کنید.

$$(\ ) = \begin{cases} -x & 0 < \ < 3 \\ 0 & o.w \end{cases}$$

الف) اگر $0 >$ آن‌گاه $(\ ) = 0$

اگر $3 > \ \leq 0$ آن‌گاه

$$(\ ) = \int f(t)\ dt = \int -t\ dt = —$$

اگر $x \geq 3$ آن‌گاه

$$(\ ) = \ (\ ) + \ (\ ) + \ (\ )$$

در نتیجه

$$(\ ) = \begin{cases} 0 & x < 0 \\ — x & 0 \leq x < 3 \\ 1 & x \geq 3 \end{cases}$$

ب) $(1 < \ \leq 2) = \ (\ \leq 2) - \ (\ \leq 1)$

$$= \ (2) - \ (1) = — \ - — = —$$

6-1-2: جمع‌آوری داده‌های احتمالاتی و نمونه‌گیری تصادفی

اگر از جامعه‌ای به حجم N واحد نمونه‌ای، به تعداد ($>$) واحد را به صورت تصادفی به طریقی انتخاب کنیم که هر یک از ترکیبات n واحد نمونه‌ای که ممکن است انتخاب شود، دارای احتمال انتخاب مساوی باشند، گوئیم از جامعه مورد نظر یک نمونه تصادفی به حجم n انتخاب کرده‌ایم. نباید نمونه‌های ورودی بایاس باشند. یعنی طوری انتخاب شوند که نتیجه مطلوب نزدیک باشد. هم‌چنین به گونه‌ای انتخاب شوند که نشان‌دهنده مشت نمونه خروار باشند. اغلب در بحث نظریه نمونه‌گیری، واحدهای نمونه‌ای یک جامعه را با ($= 1,2,3,\ldots,$) و واحدهای نمونه‌ای واقع در یک نمونه به حجم n را با ($= 1,2,3,\ldots,$) نشان می‌دهند.

نمونه‌گیری تصادفی همیشه به دو صورت انجام می‌شود: نمونه‌گیری تصادفی بدون جایگذاری واحدهای نمونه‌ای و نمونه‌گیری تصادفی با جایگذاری واحدهای نمونه‌ای.

تفاوت این دو روش در این است که در اولی هیچ‌یک از نمونه‌های انتخابی تکراری نیست و یا به عبارت دیگر در یک نمونه به حجم n که بدون جایگذاری از جامعه‌ای با حجم N انتخاب شده، تعداد n واحد مجزا وجود دارد. ولی در نمونه‌گیری تصادفی با جایگذاری، احتمال تکراری بودن واحدهای نمونه صفر نیست، به این معنی‌که ممکن است با این روش یک واحد نمونه‌ای دوبار، سه بار و یا بیشتر (به تعداد حجم نمونه) انتخاب شود.

در نمونه‌گیری تصادفی با جایگذاری، احتمال این‌که در یک استخراج، هر واحد نمونه‌ای از جامعه جز نمونه قرار گیرد، مساوی است با − (N حجم جامعه است). زیرا حجم جامعه به واسطه جایگذاری مجدد انتخاب شده در جامعه، ثابت می‌ماند.

تعداد دفعاتی که می‌توان از جامعه‌ای با حجم N واحد نمونه‌ای، نمونه‌ای به حجم n با روش جایگذاری انتخاب کرد، مساوی است با N، زیرا این‌کار به منزله ساختن n تایی‌های مرتب از مجموعه‌ای با حجم N می‌باشد و تعداد این n تایی‌های مرتب عبارت است از N.

مثال) فرض کنید جامعه‌ای از 6 واحد نمونه‌ای تصادفی تشکیل شده باشد و بخواهیم از این جامعه یک نمونه به حجم $n = 3$ بدون جایگذاری انتخاب کنیم، واحدهای این جامعه را با قرعه‌های $, \ldots, \quad ,$ مشخص می‌کنیم و همه را در یک جعبه قرار می‌دهیم، در اولین استخراج احتمال اینکه هر یک از 6 قرعه تهیه شده انتخاب شود و جز نمونه قرار گیرد

مساوی است با 1/6 . پس از استخراج اولین واحد نمونه‌ای، تعداد 5 قرعه در جعبه باقی می‌ماند و احتمال این‌که در دومین استخراج هریک از آن‌ها انتخاب شود مساوی است با 1/5 . پس از انتخاب دومین واحد نمونه، چهار قرعه در جعبه باقی خواهد ماند و احتمال این‌که در سومین استخراج هریک از این چهار قرعه انتخاب گردد نیز مساوی است با ¼ و الی آخر.

برای محاسبه واریانس و میانگین نمونه به صورت زیر عمل می‌کنیم.

قبلا قراردادهای زیر را می‌پذیریم:

: تعداد واحدهای نمونه‌ای در جامعه

y : مقدار صفت تحت مطالعه برای t امین واحد نمونه ($= 1,2,3,\ldots,$)

y: مقدار میانگین واقعی صفت در جامعه:

$$y = - \sum y$$

کمیت S که با فرمول زیر بیان شده است، متوسط توان دوم انحرافات از میانگین در جامعه محدود نامیده می‌شود.

$$= \frac{1}{-1} \quad (-)$$

$$= \frac{}{} \sum (y - Ny)$$

و واریانس مقادیر نمونه‌ها در جامعه :

$$( ) = \quad = \frac{1}{} \quad (-) = \frac{-1}{}$$

اگر حجم نمونه تصادفی را با   نمایش دهیم، داریم :

$$y = \frac{1}{}$$

$$= \frac{1}{\quad - 1} \quad (\quad - \quad) = \frac{1}{\quad - 1} \quad (\quad - \quad)$$

که به ترتیب برآوردکننده میانگین واقعی صفت و میانگین مربعات انحرافات نمونه نامیده می‌شوند. برای محاسبه واریانس این برآوردکننده یعنی ( ) دو حالت نمونه‌گیری با جایگذاری را به شرح زیر درنظر می‌گیریم:

2-1-6-1 : نمونه‌گیری تصادفی با جایگذاری

مطابق روش محاسبه واریانس می‌توان نوشت:

$$(\quad) \quad = \quad [\quad - \quad (\quad)]$$

اندیس w برای مشخص‌کردن روش نمونه‌گیری تصادفی با جایگذاری به کار رفته است. با قرار دادن $= -\Sigma \quad y$ در رابطه فوق داریم:

$$(\quad) = - \mathrm{var}\,(y_i) + \text{——} \mathrm{cov}\,(y_i, y_j)$$

2-1-6-2 : نمونه‌گیری تصادفی بدون جایگذاری

به طوری‌که دیده می‌شود:

$$(\quad) = \quad , \quad (\quad) =$$

محاسبه ( ) (اندیس برای مشخص کردن نمونه‌گیری تصادفی با روش بدون جایگذاری به کار رفته است)، مانند محاسبه ( ) در حالت نمونه‌گیری تصادفی با جایگذاری انجام می‌پذیرد تا رابطه زیر به دست آید.

$$(\quad) = \frac{1}{\quad} \quad (\quad) + \frac{\quad - 1}{\quad} \quad ,$$

ولی برخلاف آن حالت، , به علت وابستگی و مساوی صفر نمی باشد.

$$, \quad = - \frac{\quad}{\quad - 1}$$

بنابراین مقدار متوسط اشتباه نمونه‌گیری از تساوی زیر بدست می‌آید:

$$\sigma \quad = \frac{}{\sqrt{}} \quad \sqrt{1 - \frac{}{}}$$

چون ضریب ($1 - \frac{}{}$) از عدد یک کوچکتر است، از مقایسه واریانس‌های به دست آمده در حالت‌های با جایگذاری و بدون جایگذاری، نتیجه می‌شود که متوسط اشتباه نمونه‌گیری بدون جایگذاری بیشتر از متوسط اشتباه نمونه‌گیری با جایگذاری است. هم‌چنین در نمونه‌گیری بدون جایگذاری، با زیادشدن حجم نمونه، اشتباه نمونه‌گیری کم می‌شود و در صورتی که $N = n$، اشتباه نمونه‌گیری به طورکلی از بین می‌رود ولی این خاصیت در نمونه‌گیری تصادفی با جایگذاری وجود ندارد. در صورتی‌که حجم جامعه در مقایسه با حجم نمونه خیلی زیاد باشد، نمونه‌گیری تصادفی با جایگذاری و بدون جایگذاری را می‌توان یکسان فرض کرد. در واقع دقت نمونه‌گیری تصادفی از جامعه‌هایی که حجم آن‌ها در مقایسه با حجم نمونه خیلی زیاد باشد، ویا جامعه‌های نامحدود، در دوحالت با جایگذاری و بدون جایگذاری یکسان است. بنابراین داریم:

$$(\ ) \quad = (\frac{1}{} - \frac{1}{}) \quad = (1 - \frac{}{}) \frac{}{}$$

3-6-1-2 : اندازه حجم نمونه برای تعیین دقت معین

دقت نمونه‌گیری معمولا در ارتباط با اشتباه مورد قبول در برآورد پارامتر جامعه تعیین می‌گردد. اغلب کوشش می‌شود که حجم نمونه با معلوم بودن دقت خواسته شده برای نمونه‌گیری و ضریب اعتماد معین، تعیین شود. چنانچه اشتباه مورد قبول را با $y \in$ ($\in$
y مقدار اشتباه میانگین نمونه در نمونه‌گیری با حجم n یا خطای حدی می‌باشد) و ضریب اعتماد خواسته شده را با $(1 - \alpha)$ نشان دهیم، اندازه n را طوری تعیین می‌کنیم که رابطه زیر برقرار باشد:

$$[| \quad - \quad | \geq \quad] =$$

بنابراین معلوم می‌شود که n می‌تواند در ارتباط با حدود اعتماد میانگین جامعه مشخص گردد. به طور کلی تجربه نشان می‌دهد که به طور متوسط اگر از یک جامعه به دفعات مکرر نمونه انتخاب کنیم و y دارای توزیع نرمال باشد، در 68% دفعات استخراج نمونه، در صورتی‌که نمونه‌گیری بدون جایگذاری باشد، داریم:

$$|y - y| \le \sqrt{\frac{N-n}{nN}}\,S$$

و در 95% از دفعات استخراج نمونه، داریم:

$$|y - y| \le 2\sqrt{\frac{N-n}{nN}}\,S$$

در حالت کلی انتظار می‌رود که به طور متوسط در $(1-\alpha)\%$ از دفعات استخراج داشته باشیم:

$$-t_{\alpha,}\sqrt{\frac{N-n}{nN}}S \le y - y \le t_{\alpha,}\sqrt{\frac{N-n}{nN}}S$$

و با قرار دادن $\in y = y - y$ در نهایت داریم:

$$n = \frac{Nt_{,}\ S}{N \in y + t_{,}\ S}$$

اگر N به اندازه کافی بزرگ باشد، داریم:

$$N = t_{,}\ \frac{s}{(\in y)}$$

2-2: توزیع‌های احتمالاتی

2-2-1: توزیع‌های گسسته

2-2-1-1: توزیع برنولی

آزمایشی تصادفی را در نظر می‌گیریم که دارای دو نتیجه موفقیت با احتمال p و شکست با احتمال q باشد. چنین آزمایشی را برنولی می‌گویند. اگر متغیر تصادفی x به صورت زیر تعریف شود:

$$\begin{cases} X=1 & \text{اگر نتیجه آزمایش موفقیت باشد} \\ X=0 & \text{اگر نتیجه آزمایش شکست باشد} \end{cases}$$

در این‌صورت x را متغیر تصادفی برنولی گوییم و آن را با نماد X~ B(1,p) نشان می‌دهیم. ویژگی‌های این توزیع به صورت زیر می‌باشند:

تابع احتمال برنولی $x=0,1$ $\quad$ $(\ ) = P(X = x) = p\ (1 - p)$

$$(\ ) = \quad , \quad (\ ) =$$

مثال) سکه‌ای طوری اریب شده است که احتمال آمدن شیر، دو برابر احتمال آمدن خط است، اگر X نشان‌دهنده تعداد شیر در یک آزمایش باشد، میانگین و واریانس متغیر تصادفی X را حساب کنید.

$$(\text{شیر}) = 2 \quad (\text{خط})$$

$$(\text{خط}) + 2 \quad (\text{خط}) = 1 \Rightarrow \quad (\text{خط}) = 1/3,\ p(\text{شیر})=2/3$$

$$(\ ) = \frac{2}{3}, \qquad (\ ) = 2/3 * 1/3 = 2/9$$

2-2-1-2: توزیع دوجمله‌ای

احتمال‌هایی که با توزیع دو جمله‌ای داده می‌شوند، می‌توانند از راه‌های زیر پیش آیند: (1) هنگام نمونه‌گیری با جایگذاری از یک جامعه متناهی (2) هنگام نمونه‌گیری از یک جامعه نامتناهی (که اغلب به جامعه‌ای بی‌نهایت بزرگ اطلاق می‌شود) و یا بدون جایگذاری.

فرض کنید که بخواهیم از یک توده N شی که N عدد از آن‌ها معیوب و ($N_($ عدد سالم اند، یک نمونه n تایی با جایگذاری خارج کنیم. به عبارت دیگر، به تصادف یک عضو از توده را خارج می‌کنیم، آن را آزمایش و نتیجه را ثبت کرده و در توده قرار می‌دهیم، و آن را به طور کامل مخلوط می‌کنیم . فرایند را $n - 1$ باردیگر تکرار می‌کنیم.

X : تعداد موفقیت‌ها در n بار آزمایش برنولی

آن را با نماد X ~B(n, p) نمایش می‌دهیم. ویژگی‌های آن به صورت زیر است:

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

$$E(X) = np \quad , \qquad Var(X) = npq$$

مثال) دو تاس را 100 بار می‌ریزیم و تعداد نه‌ها را یادداشت می‌کنیم.

الف) احتمال رخ دادن x بار نه چقدر است؟

ب) احتمال این‌که دست‌کم سه بار نه رخ دهد چقدر است؟

آشکار است که هر بار ریختن دو تاس را برای پیشامد نه یا غیر نه آزمایش می‌کنیم، احتمال به دست آوردن یک نه در هر بار ریختن دو تاس برابر p=1/9 است. از این رو

$$P(x) = \binom{100}{x} \left(\frac{1}{9}\right) \left(\frac{8}{9}\right) \qquad , x = 0,1,2,\dots,100$$

ب)

$$P(x \geq 3) = 1 - p(x < 3) = 1 - p(x \leq 2)$$

$$= 1 - \left[\quad \binom{100}{x} \left(\frac{1}{9}\right) \left(\frac{8}{9}\right) \quad \right]$$

$$= 1 - (0.000008 + 0.000097 + 0.000603)$$

$$= 0.9993$$

3-1-2-2: توزیع فوق هندسی

فرض کنید جامعه‌ای به دو قسمت جمعیت موفق و جمعیت شکست تقسیم می‌شود که یکی M تایی و دیگری N – M تایی است (N کل جمعیت است). می‌خواهیم n عضو از این جمعیت N تایی را بدون جایگذاری انتخاب کنیم. چنین آزمایشی را فوق هندسی گویند.
وقتی نمونه‌ای از n شی به تصادف از جامعه‌ای متناهی با N شی استخراج می‌شود که در آن نمونه‌گیری بدون جایگذاری انجام شده و هر عنصر جامعه می‌تواند به روشی ساده به صورت تعلق به دو رده مجزا، دو حالتی شود.

تعداد موفقیت ها در یک آزمایش فوق هندسی :

$$(\quad = \quad) = \frac{\qquad}{(\quad)}$$

$$(\quad) = \frac{\quad}{\quad}$$

$$\text{تصحیح ضریب} = \frac{-}{-1} \qquad (\ ) = \frac{-}{-1} \cdot \frac{}{} \cdot (1 - -)$$

مثال) از جعبه‌ای شامل 4توپ سفید و 1توپ سیاه یک نمونه تصادفی 2تایی خارج می‌کنیم.   الف- با جایگذاری   ب- بدون جایگذاری

امید ریاضی متغیر تصادفی X که نشان‌دهنده تعداد توپ در نمونه است را در هر حالت محاسبه نمائید.

الف-

X	0	1	2
f(x)	1/25	8/25	16/25

$$E(X) = 0 \times \frac{1}{25} + 1 \times \frac{8}{25} + 2 \times \frac{16}{25} = \frac{8}{5}$$

ب-

$$P(X = x) = \frac{}{}$$

X	1	2
$P(X = x)$	$\frac{2}{5}$	$\frac{3}{5}$

$$E(X) = 1 \times \frac{2}{5} + 2 \times \frac{3}{5} = \frac{8}{5}$$

همان‌طور که ملاحظه می‌کنید، نمونه‌گیری چه با جایگذاری و چه بدون جایگذاری باشد، مقدار امید ریاضی فرق نمی‌کند.

4-1-2-2: توزیع پوآسن

اینک یک توزیع احتمال مهم، یعنی توزیع پوآسون را درنظر می‌گیریم که از شکل حدی تابع احتمال توزیع دو جمله‌ای وقتی $n \to \infty, p \to 0$ و np ثابت بماند، بدست می‌آید، که $np = \lambda$ در نظر گرفته می‌شود.

اگر متغیر تصادفی X نشان‌دهنده تعداد موفقیت‌ها در یک فاصله زمانی یا در یک مکان مشخص باشد آن‌گاه X را یک متغیر تصادفی پوآسون گویند و آن را با نماد $X \sim P(\lambda)$ نشان می‌دهند. مشخصات این توزیع به صورت زیر است:

$$(\ \ ) = P(X = x) = \frac{e \quad .\lambda}{x!}$$

مثال) دو درصد پیچهایی که با یک ماشین ساخته می‌شوند معیوب‌اند و تولید معیوب‌ها در جریان ساخت به طور تصادفی رخ می‌دهد. اگر پیچ‌ها در جعبه‌های 100 تایی بسته‌بندی شوند، احتمال این‌که جعبه‌ای معین دارای x معیوب باشد، چقدر است؟

احتمال این‌که جعبه دارای x معیوب باشد با توزیع دوجمله‌ای تعیین می‌شود

$$p(x) = \binom{100}{x} \left(\frac{2}{100}\right) \left(1 - \frac{2}{100}\right) \qquad , \quad x = 0,1,\dots,100$$

چون n=100 و p=0.02 و np=2، تقریب پوآسون برای $p(x)$ به صورت $P(x) = \frac{\quad}{!}$ نشان داده می‌شود.

تحت برخی شرایط، توزیع دو جمله‌ای تقریبی برای توزیع فوق هندسی فراهم می‌سازد و تحت شرط‌هایی توزیع پوآسون، تقریبی برای توزیع دو جمله‌ای فراهم می‌کند. آشکار است که تحت این دو مجموعه از شرط‌ها، توزیع پوآسون، تقریبی برای توزیع فوق هندسی فراهم می‌آورد.

مثال) می‌دانیم که در نوعی فرایند عایق‌کاری تعداد شکست‌های عایق‌کاری در هر یارد سیم 0,07 است. احتمال یافتن x تا از این نوع شکستگی‌ها در قطعه‌ای به طول 16 یارد از سیم چقدر است؟

$\lambda=0.07$, $\mu=\lambda L= (0.07)(16)=1.12$

$(\ \ )=\frac{(\ .\ )}{!}$

2-2-1-5: توزیع دو جمله‌ای منفی

اگر آزمایش برنولی را آنقدر ادامه دهیم تا r موفقیت بدست آوریم و پس از رخ دادن r امین موفقیت آزمایش را متوقف کنیم، چنین آزمایشی را آزمایش دو جمله‌ای منفی می‌نامیم.

متغیر تصادفی X را که نشان‌دهنده تعداد دفعات در آزمایش دو جمله‌ای منفی است، متغیر تصادفی دو جمله‌ای منفی می‌نامیم.

اگر متغیر تصادفی X به صورت زیر تعریف شود:

X: تعداد آزمایشات برای رسیدن به r امین موفقیت

با یک متغیر تصادفی دو جمله‌ای سروکار داریم. مشخصات این توزیع به صورت زیر می‌باشد:

$$P(X = x) = \binom{x-1}{r-1} p\ \ (q)^{\ -} \qquad x = r, r+1, \ldots$$

$$E(X) = \frac{r}{p} \quad \text{و} \quad Var(X) = \frac{rq}{p}$$

مثال) احتمال اینکه در پرتاب متوالی یک سکه، سومین شیر در هفتمین آزمایش بدست آید چقدر است؟

$$X = 7\ , p = 1/2\ , r = 3$$

$$P(X = 7) = \begin{matrix} 7-1 \\ 3-1 \end{matrix} \ \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \quad = \frac{15}{128} = 0.1172$$

2-2-1-6: توزیع هندسی

اگر در توزیع دو جمله‌ای منفی r=1 باشد، آن‌گاه با یک متغیر تصادفی هندسی روبرو هستیم:

X: تعداد آزمایشات مستقل برنولی تا رسیدن به اولین موفقیت

متغیر هندسی را با نماد x~Ge(p) نشان می‌دهند. مشخصات آن به صورت زیر می‌باشد:

$$(\ \ ) = P(X = x) = \qquad\qquad x = 1,2,3, \ldots$$

$$E(x) = \frac{1}{p} \quad , \quad Var(X) = \frac{q}{p}$$

2-2-1-7: توزیع یکنواخت گسسته

اگر متغیر تصادفی X مقادیر یکسان اختیار کند با توزیع یکنواخت گسسته روبرو هستیم. این توزیع را با نماد X~DU(k) نشان می‌دهند. مشخصات آن به صورت روبرو می‌باشد:

$$(\ \ ) = P(X = x) = \frac{1}{k} \ x = x1, x2, x3, \dots xk$$

$$? = E(X) = \frac{1}{k} \quad x \quad , Var(x) = \frac{1}{k} \quad (x - ?)$$

مثال) کلاسی دارای 20 دانش آموز می‌باشد، می‌خواهیم به تصادف یک نفر را انتخاب و کتابی را به او هدیه دهیم، توزیع احتمال آن را مشخص کنید.

$$f(x) = \frac{1}{20} \quad x = 1,2, \dots, 20$$

2-2-2: توزیع‌های پیوسته

2-2-2-1: توزیع یکنواخت پیوسته

اگر X دارای تابع چگالی احتمال به صورت زیر باشد گوئیم X دارای توزیع یکنواخت پیوسته در بازه (a, b) است و آن را با نماد X~U(a, b) نشان می‌دهیم. مشخصات این توزیع به صورت زیر می‌باشد:

$$fx\,(x) = \frac{1}{-} \qquad a < \quad <$$

$$0 \qquad\qquad h$$

$$F_X(x) = \begin{cases} 0 & < \\ \dfrac{x-a}{b-a} & a \le x < \\ 1 & b \le a \end{cases}$$

$$E(X) = \frac{a+b}{2} \quad Var(X) = \frac{(b-a)}{12}$$

مثال) از یک ایستگاه اتوبوس مشخص، اتوبوس‌ها به فاصله 20 دقیقه از یکدیگر حرکت می‌کنند (مثلا اگر اولی ساعت چهار حرکت کند، بعدی ساعت چهار و بیست دقیقه و ...) مسافری بین ساعت پنج تا پنج و چهل دقیقه از این ایستگاه حرکت کرده است.

الف) حساب کنید احتمال آن را که مسافر کمتر از 10 دقیقه منتظر اتوبوس مانده باشد.
ب) بیشتر از 10دقیقه منتظر مانده باشد.

$$f(x) = \frac{1}{5{:}40 - 5} = \frac{1}{40}$$

الف:

$$P(10 < \quad < 20) + \quad (30 < \quad < 40) = \quad \frac{1}{40}dx + \quad \frac{1}{40}dx = \frac{1}{2}$$

ب:

$$P(0 < \quad < 10) + \quad (20 < \quad < 30) = \quad \frac{1}{40}dx + \quad \frac{1}{40}dx = \frac{1}{2}$$

2-2-2-2: توزیع گاما

نام این توزیع از تابع گاما گرفته شده است که به صورت زیر تعریف می‌شود:

$$\Gamma(z) = \quad t \quad e \quad dt$$

$$\Gamma(z) = (z-1)!$$

$$\Gamma(z + 1) = z.\Gamma(\ ) = \ .(\ - 1)!$$

متغیر تصادفی گاما به صورت زیر تعریف می‌شود:

X: مدت زمان لازم جهت اتفاق افتادن

آن را با نماد $X \sim \Gamma(\alpha, -)$ نشان می‌دهند که در اینجا − میانگین تعداد اتفاقات براساس یک فرایند پوآسون می‌باشد. مشخصات این توزیع به صورت زیر است:

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

$$E(X) = \alpha\beta \ ; \quad Var(X) = \alpha\beta$$

اگر در آزمایش پوآسون با میانگین n، x زمان αامین موفقیت باشد، توزیع X یک توزیع گاما با پارامتر α و − است و همانطور که ملاحظه می‌کنید توزیع نمایی حالت خاصی از توزیع گاما با $\alpha = 1$ می‌باشد.

مثال) فرض کنید در هر ساعت به طور متوسط 30 اتومبیل وارد یک پارکینگ می‌شوند، احتمال اینکه مامور پارکینگ حداقل 5 دقیقه منتظر بماند تا دومین اتومبیل وارد پارکینگ شود، چقدر است؟

متوسط تعداد اتومبیل‌هایی که در دقیقه وارد پارکینگ می‌شوند برابر است با $- = —$. اگر متغیر تصادفی X زمان انتظار برحسب دقیقه باشد، این متغیر تصادفی دارای توزیع گاما با پارامتر $\alpha = 2$ و $\beta = 2$ می‌باشد. بنابراین

$$P(X \geq 5) = \quad \frac{x \quad e^{-}}{2 \ \Gamma(2)}\,dx$$

$$= \quad \frac{1}{4}\,x\,e^{-}\,dx = \frac{7}{2}\,e^{-}$$

2-2-2-3: توزیع نمایی

اگر آزمایش برنولی را آنقدر ادامه دهیم تا اولین پیروزی بدست آید به توزیع هندسی می‌رسیم. می‌خواهیم بحث مشابهی را در مورد توزیع پوآسون انجام دهیم تا به توزیع نمایی برسیم. در توزیع پوآسون، Y یک متغیر تصادفی جداست که نمایانگر تعداد پیروزی در یک ناحیه پیوسته یا یک فاصله زمانی است، حال متغیر تصادفی X را با شرایط و مشخصات توزیع پوآسون درنظر می‌گیریم ولی فرض می‌کنیم X یک متغیر تصادفی پیوسته است. فرض کنید n میانگین تعداد پیروزی در واحد زمان باشد. پس توزیع پوآسون متناظر دارای میانگین nx در مدت زمان x خواهد بود. توزیع پوآسون در رابطه با متغیر تصادفی Y عبارت است از :

$$P(y) = \frac{(nx)\quad e}{y!} \qquad , y = 0,1,2,3,\ldots$$

حال تابع توزیع تجمعی X را می‌نویسیم $F\ (x) = P(X \leq x)$

متغیر تصادفی X به صورت زیر تعریف می‌شود:

X: زمان انتظار بین دو پیشامد متوالی در فرایند پوآسون

یا

X: زمان انتظار تا مشاهده اولین پیشامد در فرایند پوآسون

آن را با نماد $X \sim Exp(-)$ نشان می‌دهند و مشخصات این توزیع به صورت زیر می‌باشد:

$$f(x) = \begin{cases} \frac{1}{\beta}e^{-} & X > 0 \\ 0 & ow \end{cases}$$

$$E(X) = \beta \qquad Var(X) = \beta$$

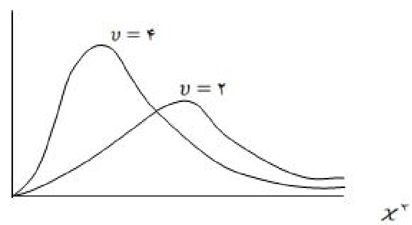
2-2-2-4: توزیع خی- دو

در توزیع گاما اگر $\alpha = -$ و $\beta = 2$ باشد، چگالی خی- دو تعریف می‌شود. مربع کای- دو را با نماد $X \sim \chi\ (\nu)$ نشان می‌دهند. مشخصات این توزیع به صورت زیر است:

$$f(x,k) = \begin{cases} \frac{1}{\Gamma(k/2)2^{-}} x^{-} \quad e^{-} & \text{for } X > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

$$E(X) = \upsilon \qquad Var(X) = 2\upsilon$$

مقدار χ منفی نمی‌باشد و منحنی توزیع آن حول $\chi = 0$ متقارن نیست، در زیر منحنی‌های خی- دو برای درجه آزادی $\upsilon = 4$ و $\upsilon = 7$ رسم شده است.



2-3: منحنی‌های خی- دو برای درجه آزادی $\upsilon = 4$ و $\upsilon = 7$

مثال) اگر X درای توزیع خی- دو با درجه آزادی 10 باشد، مطلوبست محاسبه
P(3/5 <   < 20/5)

$$(3/5 < \quad < 20/5) = \quad (\quad < 20/5) - \quad (\quad < 3/5)$$

$$= 0.975 - 0.025 = 0.95$$

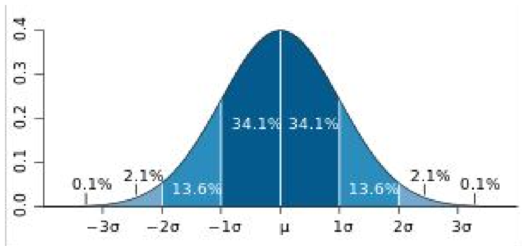
2-2-2-5: توزیع نرمال

کاربردی‌ترین توزیع پیوسته در آمار، توزیع نرمال است. متغیر X دارای توزیع نرمال با میانگین   و واریانس σ است. هرگاه تابع چگالی احتمال آن به صورت زیر باشد:

$$x, ?, \sigma \quad = \quad \frac{1}{\overline{2\pi\sigma}} \; e^{\frac{-(x-?)}{2\sigma}} \qquad x \in \mathbb{R}$$

$$E(X) = ? \qquad Var(X) = \sigma$$

منحنی تابع چگالی احتمال به صورت زنگی شکل است.

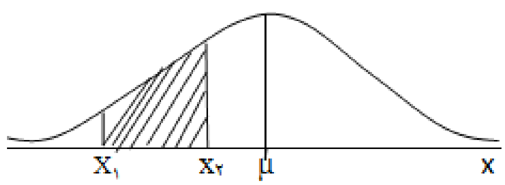


4-2: منحنی تابع چگالی احتمال

متغیر تصادفی نرمال با میانگین 0 = و واریانس 1 = σ را متغیر تصادفی نرمال استاندارد می‌نامیم و آن را با Z نشان می‌دهیم.

اگر X یک متغیر تصادفی نرمال با میانگین و واریانس σ باشد، برای محاسبه احتمال، از متغیر —— = Z استفاده نموده و آن را به نرمال استاندارد تبدیل می‌کنیم و احتمال مربوط به نرمال استاندارد را از جدول بدست می‌آوریم.

$$P(x1 < \quad < \quad 2) = P(z1 < \quad < \quad 2)$$



5-2: تابع چگالی احتمال در بازه $x1 < \quad < \quad 2$

مثال) معدل نمره 300 دانشجوی یک دانشکده تقریبا دارای توزیع نرمال با 2/1 = ? و σ = 1/2 است. در صورتی‌که معدل‌ها تا یک دهم تقریب محاسبه شده باشند.

الف) احتمال این‌که معدل یک دانشجو که به تصادف انتخاب شود، تقریبا 2/5 باشد را محاسبه کنید.

ب) احتمال این‌که معدل یک دانشجو که به تصادف انتخاب شود، دقیقا 2/5 باشد را محاسبه کنید.

چون معدل‌ها تا یک دهم تقریب محاسبه شده‌اند و دو نقطه انتهایی فاصله نیز مورد نظر است، باید سطح محصور واقع بین نقاط $3/55 = 2$ و $2/45 = 1$ را محاسبه کنیم.

$$(2/45 < \quad < 3/55) = \left(\frac{2/45 - 2/1}{1/2} < \quad < \frac{3/55 - 2/1}{1/2}\right)$$
$$= (0.29 < \quad < 1/21)$$
$$= (\quad < 1/21) - (\quad < 0.29)$$
$$= 0.8861 - 0.6141 = 0.2728$$

بنابراین معدل 27,28% دانشجویان در فاصله [2/5 , 3/5] قرار دارد و داریم

100	27/28	
300	x	, x=81/84

تقریبا 82 نفر دارای معدلی در فاصله [2/5 , 3/5] می‌باشند.

ب) باید سطح محصور واقع بین $x_1=2/45$ و $x_2=2/55$ را محاسبه کنیم

$$(2/45 < \quad < 2/55) = \left(\frac{2/45 - 2/1}{1/2} < \quad < \frac{2/55 - 2/1}{1/2}\right)$$
$$= (\quad < 0.375) - (\quad < 0.29)$$
$$= 0.6461 - 0.6141 = 0.032$$

و یا می‌توان گفت که تقریبا 10 نفر دارای معدل 2/5 می‌باشند.

اگر X یک متغیر تصادفی دو جمله ای با میانگین $=$ و واریانس $\sigma =$ باشد، شکل حدی توزیع متغیر تصادفی

$$= \frac{}{\sqrt{\quad}}$$

هنگامی که n به سمت بی‌نهایت میل کند، توزیع نرمال استاندارد می‌باشد.

توجه: هنگامی که n بزرگ و p به $-$ نزدیک باشد، تقریب بسیار خوب است.

اگر X یک متغیر تصادفی نرمال با میانگین ? و واریانس σ باشد، $(\sigma > 0)$، آن‌گاه متغیر تصادفی دارای توزیع خی- دو با درجه آزادی یک است.

$$= \left(\frac{-}{\quad}\right)$$

اگر ، ... ، ، n متغیر تصادفی مستقل نرمال با میانگین‌های ?، ? ، ...، ? و واریانس‌های σ ، σ، ...، σ باشند، آن‌گاه

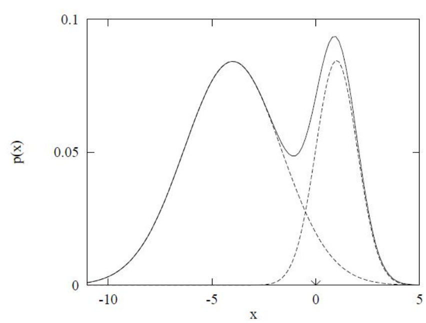
$$= \sum \quad (\text{——})$$

دارای توزیع خی- دو با درجه آزادی n است.

3-2: چگالی مخلوط

در روش‌های مبتنی بر مدل برای آن‌که بتوان رفتار داده‌ها را مورد بررسی قرار داد ابتدا یک مدل که نزدیکترین مدل به رفتار داده‌ها است را در نظر می‌گیریم و تحلیل را بر روی آن مدل انجام می‌دهیم. چگالی که بیشتر مورد استفاده قرار می‌گیرد، چگالی نرمال است. به دلیل آن‌که پراکندگی داده‌ها در حالت طبیعی به صورت نرمال است.

اما یک مشکل وجود دارد و آن زمانی است که داده‌هایی با پراکندگی مختلف داشته باشیم، به عنوان مثال اگر دریکی از کلاس‌های نمونه پراکندگی در مرکز و در دیگری در اطراف آن باشد، میانگین این داده‌ها یکسان می‌شود اما پراکندگی آن‌ها مختلف است. بنابراین به جای استفاده از چگالی واحد[1] از چگالی مخلوط[2] استفاده می‌کنیم که بهترین مدل برای این‌گونه داده‌ها است.



6-2: نمودار چگالی مخلوط

چگالی مخلوط چگالی مولفه‌ها را به یک پارامتر خاص کلاس محدود می‌کند. با فرض این‌که برای داده‌های موجود مناسب و یا از لحاظ محاسباتی مناسب باشند. برای تعیین تعداد چگالی‌ها می‌توان ازدسته‌بندی اتوماتیک مانند LBG استفاده کرد.

---

[1] Single density
[2] Mixture densities

این چگالی با $P(x;\theta$) ارائه می‌شود. پارامتر θ ، sامین پارامتر مولفه‌های چگالی مخلوط است. از π برای اشاره به فاکتور وزن sامین مولفه چگالی مخلوط استفاده می‌کنیم. وزن‌ها باید این دو ویژگی را داشته باشند: (i) باید غیر منفی باشد $\pi > 0$ و (ii) مجموع آن‌ها یک شود $\sum \pi = 1$ وزن‌های π که نسبت هر چگالی در چگالی مخلوط یا وزن آن چگالی است را می‌توان به عنوان احتمال p(s) فرض کرد که از یک نمونه داده از مولفه s، چگالی مخلوط استخراج شده است.

چگالی مخلوط، k مولفه به صورت زیر تعریف می‌شود:

( ) = (;)

مجموعه پارامترهای چگالی مخلوط را به صورت زیر بیان می‌کنیم.

={ , ,..., , , , ... , }

با فرض اینکه در اینجا تمام داده‌ها دارای توزیع منحصربه‌فرد و غیروابسته[3] درنظر گرفته شده اند. به عبارت دیگر، احتمال بردار مجموعه داده‌ها، حاصلضرب احتمال‌های آن‌ها است.

می‌توان چگالی مخلوط را به عنوان مدل‌کننده یک پردازش درنظر گرفت. زمانی که اولین منبع s بر طبق توزیع چندگانه[4]، { , ... , , } انتخاب می‌شود و سپس یک نمونه در آن بر طبق مولفه چگالی $p(x;\theta$) استخراج می‌شود. بنابراین، احتمال منبع انتخاب شده s و داده x به صورت $p(x;\theta$) است. احتمال حاشیه‌ای داده‌های انتخاب شده با استفاده از چگالی مخلوط محاسبه می‌شود.

منبعی که بردار داده x را تولید کرده می‌توان به عنوان اطلاعات از دست رفته در نظر گرفت. به عبارت دیگر، پردازش بر روی بردار x صورت می‌گیرد و منبعی که داده را تولید کرده است دیگر نمی شناسیم.

یک کمیت استخراج شده مهم احتمال ثانویه[5] است که بر روی مولفه‌های چگالی مخلوط با توجه به بردار داده‌ها تعریف شده است. می‌توان آن را به عنوان توزیعی که یک بردار داده خاص را از روی مولفه‌های مخلوط استخراج می‌کند، در نظر گرفت. به عنوان مثال، "بردار

---

[3] Identically and independency distributed (IID)
[4] Multinomial
[5] Posterior probability

داده از روی کدام چگالی مولفه استخراج شده است؟" و یا "بردار داده به کدام دسته تعلق دارد؟" احتمال ثانویه برای مولفه‌های مخلوط با استفاده از قانون بیز تعریف شده است.

الگوریتم ماکزیمم انتظار، پارامترهای مدل مخلوط را از روی داده‌هایی که باعث ضرورت استفاده از احتمال ثانویه شده‌اند، تخمین می‌زند. مدلسازی بر اساس چگالی[6] مخلوط به عنوان تخمین احتمال نیمه- پارامتری شناخته می‌شود و می‌تواند میان جریان تخمین چگالی پارامتری و غیر پارامتری قرار گیرد. تخمین احتمال پارامتریک، فرض می‌کند که داده‌ها از یک کلاس پارامتریک همانند داده‌های کلاس نرمال، استخراج شده‌اند. مساله تخمین سپس جستجوی پارامترهای نرمالی را که بهترین داده‌های مناسب را پیدا می‌کند، کاهش می‌دهد. فرضی که بر پارامترهای تخمین چگالی استوار باشد، اغلب موارد غیرواقعی است، اما پارامترهای بسیار کارائی را استخراج می‌کند. در متد تخمین غیر پارامتریک، یک چگالی خاص برای استخراج داده‌ها درنظر گرفته نمی‌شود.

تخمین غیر پارامتریک، یک فرم چگالی مخلوط با یک مولفه مخلوط برای هر نقطه داده در مجموعه داده‌ها را تخمین می‌زند. هسته‌ها به عنوان مولفه‌ها انتخاب می‌شوند.

یک تخمین‌زننده غیر پارامتریک معروف تخمین‌زننده پارتزن[7] است که از مولفه‌های نرمال با با میانگین‌های یکسان برای تطبیق نقاط داده و کوواریانس‌های مشابه استفاده می‌کند.

تخمین غیرپارامتریک می‌تواند چگالی یک کلاس بزرگ را پیاده‌سازی کند. هزینه‌ای که ما باید برای محاسبه نقطه جدید تخمین زده شده بدهیم این است که ما باید تمام هسته‌ها را ارزیابی کنیم که این محاسبه برای مجموعه داده‌های بزرگ، پیچیدگی محاسباتی زیادی دارد.

مدلسازی بر اساس چگالی میان دو جریان تعادل برقرار می‌کند: چگالی یک کلاس بزرگ را می‌تواند پیاده‌سازی کندو چگالی آن را به صورت کارائی ارزیابی کند، به گونه‌ای که تنها تعداد نسبتا کمی از تابع چگالی‌ها ارزیابی شود.

[6] Mixture modelling
[7] Parzen estimstor

4-2: تخمین پارامترهای یک مدل احتمالاتی

در آمار استنباطی از روی نمونه بدست آمده و نتایج آن در مورد پارامترهای جامعه، نتیجه گیری و استنباط می‌کنیم. این به دو بخش تئوری تخمین (برآورد) و آزمون فرضیه تقسیم‌بندی می‌شود. خود برآوردگر شامل دو بخش برآورد نقطه‌ای و فاصله اطمینان می‌باشد.

1-4-2: برآورد نقطه‌ای

این برآوردگر برای یک پارامتر به سادگی انجام می‌شود، کافیست یک نمونه تصادفی را انتخاب کرده و مقدار یک برآوردگر خوب را برای پارامتر مورد نظر بدست آورد، مثلا اگر پارامتر مورد نظر μ باشد برآوردگر را در نظر گرفته و مقدار X (میانگین نمونه) را برای یک نمونه تصادفی منتخب محاسبه می‌کنیم. این مقدار، برآورد نقطه‌ای خوبی برای μ است. و یا اگر پارامتر موردنظر σ باشد، کافیست برآوردگر را درنظرگرفته و به کمک یک نمونه تصادفی منتخب مقدار (واریانس نمونه) را بدست آوریم، این مقدار، برآورد نقطه‌ای خوبی برای σ خواهد بود. اما برآورد نقطه‌ای دارای خطای زیاد می‌باشد، زیرا مقدار آن با تغییر نمونه (حجم نمونه ثابت باشد) تغییر خواهد کرد. بنابراین سعی می‌کنیم تا بجای برآورد نقطه‌ای، یک برآورد فاصله‌ای تعیین کنیم.

2-4-2: برآوردگر نقطه‌ای (میانگین جامعه) و واریانس جامعه

اگر توزیع جامعه‌ای ناشناخته باشد، آن گاه ? و σ ناشناخته‌اند. بنابراین آن‌ها را از روی نمونه تصادفی برآورد می‌کنیم.

برآوردگر θ را برآوردگر نااریب گوئیم، اگر داشته باشیم

$$? = E(\hat{\theta}) = \theta$$

می دانیم ? = - ? پس برای نمونه‌ای با حجم دلخواه n، همواره X یک برآورد نااریب برای ? است و نیز S برآوردگر نااریب برای σ است.

مثال) اگر متغیر تصادفی X نشان‌دهنده تعداد موفقیت در یک جامعه دو جمله‌ای با نسبت p باشد، نشان دهید که برآورد — = یک برآوردگر نااریب برای نسبت p است.

باید نشان دهیم که $E(P) = p$

$$E(P) = E(\frac{X}{n}) = \frac{1}{n} E(X) = \frac{1}{n} np = p$$

توجه: برآورد نااریب یک پارامتر، منحصر به فرد نیست. چون برآوردگرهای میانه و میانگین ( ) هر دو برآوردگرهای نااریب برای ? هستند و ما می‌خواهیم از میان برآوردگرهای نااریب یک پارامتر، بهترین برآوردگر را تعیین کنیم. بنابراین از میان برآوردگرهای نااریب پارامتر θ، برآوردگری که کمترین واریانس را داشته باشد، کاراترین برآوردگر نامیده می‌شود.

به روش مشابه، واریانس نمونه‌ای S را می‌توان به عنوان برآوردکننده نقطه‌ای واریانس جامعه یعنی σ به کار برد.

به طورکلی داریم، اگر برای نمونه‌ای تصادفی به حجم n از جامعه‌ای با میانگین ناشناخته μ و واریانس ناشناخته σ مقادیر X و S معین شده باشند، آن‌گاه X یک برآوردکننده نقطه‌ای نااریب برای μ با واریانس $\sigma = \sigma / n$ است. به علاوه S یک برآوردکننده نقطه‌ای نااریب برای است.

3-4-2: برآورد فاصله (فاصله اطمینان)

یک برآورد فاصله‌ای برای پارامتر ، فاصله‌ای است به صورت > > که در آن a و b یک برآورد نقطه‌ای پارامتر مانند θ است که از نمونه تصادفی بدست می‌آید و به نوع توزیع برآوردگر θ مربوط می‌باشند. چون نمونه‌های تصادفی مختلف (با حجم ثابت n) مقادیر متفاوت را برای برآورد نقطه‌ای θ بدست می‌دهند، بنابراین فاصله‌های متفاوت برای پارامتر تشکیل می‌شود که بعضی از آن‌ها پارامتر را شامل و بعضی پارامتر را شامل نخواهند شد. چون فاصله‌های اطمینان را به کمک توزیع برآوردگر θ تعیین خواهیم کرد، بنابراین این توزیع ما را قادر می‌سازد که a و b را طوری تعیین کنیم که درصدی از فاصله‌ها، شامل پارامتر باشند، به عنوان مثال اگر a و b طوری محاسبه شوند که 95% از فاصله‌ها شامل پارامتر باشند، این فاصله را 5% فاصله اطمینان گوئیم و به طورکلی اگر a و b طوری محاسبه شوند که 100(-1) % از فاصله‌ها شامل باشند، آنرا 100(-1) % فاصله اطمینان و ضریب (-1) را ضریب اطمینان گوئیم. $(1 > \alpha > 0)$.

2-4-3-1: برآورد فاصله اطمینان میانگین جامعه   برای جامعه به اندازه کافی بزرگ

با کمک برآورد نقطه‌ای و توزیع ، فاصله اطمینان برای ? را بدست می‌آوریم. می‌دانیم که اگر نمونه‌های تصادفی با حجم n، (n به اندازه کافی بزرگ) از جامعه‌ای با میانگین ? و واریانس معلوم σ انتخاب شود، آن‌گاه متغیر تصادفی $\frac{}{/\sqrt{}}$ = بنا به قضیه حد مرکزی دارای توزیع نرمال استاندارد خواهد بود. از طرفی / را مقداری از توزیع نرمال درنظر می‌گیریم که مساحت 1 − /2 در سمت چپ آن قرار دارد یعنی

$$P(Z \le z_{\ /}) = 1 - \alpha/2$$

با توجه به شکل منحنی توزیع نرمال استاندارد و خاصیت تقارن این توزیع می‌توان نوشت:

$$P(-z_{\ /} < \ < z_{\ /}) = 1 - \alpha$$

که اگر به ازای $\frac{}{/\sqrt{}}$ = ، Z را قرار دهیم، در انتها داریم:

$$P(X - \frac{\sigma}{\sqrt{n}} z_{\ /} < \ < X + \frac{\sigma}{\sqrt{n}} z_{\ /})$$

بنابراین می‌توان گفت، اگر یک نمونه تصادفی به حجم n، (n به اندازه کافی بزرگ) از یک جامعه با واریانس معلوم   انتخاب شود و میانگین نمونه   محاسبه شود،آنگاه

100%(− 1) فاصله اطمینان برای جامعه به صورت زیر است:

$$- \frac{}{\sqrt{}} \ _{/} < \ < \ + \frac{}{\sqrt{}} \ _{/}$$

به عبارت دیگر می‌توان گفت، 100%(− 1) فاصله اطمینان برای ? وقتی که نمونه تصادفی با حجم n از جامعه‌ای با واریانس معلوم   و /   مقداری از توزیع نرمال استاندارد است که مساحت $\alpha/2$ − 1 در سمت چپ آن قرار دارد.

اگر ضریب − 1 را ثابت فرض کنیم و بخواهیم فاصله اطمینان خوبی را ارائه دهیم کافی است حجم نمونه را زیاد انتخاب کنیم، زیرا فاصله اطمینان خوب هنگامی بدست می آید که ضریب اطمینان بالا و هم‌زمان طول فاصله کم باشد. مثلا ترجیح می‌دهیم که میانگین طول

عمر یک نوع لامپ با 95% اطمینان در فاصله 8 تا 9 سال قرار گیرد تا آنکه با 99% اطمینان در فاصله 3 تا 10 سال قرار گیرد.

2-4-3-2: برآورد فاصله اطمینان میانگین جامعه و واریانس معلوم برای جامعه n≥30

در این جامعه بدون توجه به نوع توزیع در رابطه زیر برای میانگین معلوم ? و واریانس معلوم ، همان فاصله اطمینان ارائه شده در جامعه‌های به اندازه کافی بزرگ را داریم و این فاصله اطمینان بدون توجه به نوع توزیع قابل محاسبه است و از دقت فاصله اطمینان خوبی برخوردار است.

$$-\frac{}{\sqrt{}} \quad / < \; < \; + \frac{}{\sqrt{}} \quad /$$

2-4-3-3: برآورد فاصله اطمینان میانگین جامعه با واریانس نامعلوم برای جامعه n≥30

در این جامعه به جای σ از s (انحراف نمونه) استفاده می شود. در این صورت باز هم می‌توان از فاصله اطمینان ارائه شده در جامعه به اندازه کافی بزرگ استفاده کرد.

$$-\frac{}{\sqrt{}} \quad / < \; < \; + \frac{}{\sqrt{}} \quad /$$

2-4-3-4: برآورد فاصله اطمینان توزیع جامعه تقریبا نرمال و واریانس معلوم برای جامعه n<30

در صورت معلوم بودن توزیع جامعه تقریبا نرمال و واریانس جامعه، می‌توان از فاصله اطمینان ارائه شده در جامعه به اندازه کافی بزرگ استفاده کرد.

2-4-3-5: برآورد میانگین جامعه‌ای با واریانس نامعلوم و جامعه‌ای با حجم n<30

یکی از عواملی که امکان انتخاب نمونه با حجم 30 ≥ n را به ما نمی دهد، ارزش کالا است یا مثلا می‌خواهیم نوعی دارو را آزمایش کنیم که امکان عارضه قلبی دارد و نمی‌توانیم این دارو را بر روی افراد زیاد آزمایش کنیم. در حالت کلی، اگر واریانس جامعه نامعلوم بوده و نتوانیم نمونه‌ای با حجم 30 ≥ n را انتخاب کنیم و توزیع جامعه نیز مشخص نباشد،

نمی‌توانیم یک برآورد فاصله‌ای برای میانگین جامعه بدست آوریم، اما اگر توزیع جامعه تقریبا نرمال باشد و اگر از به عنوان برآوردگر ? استفاده شود، می‌دانیم متغیر تصادفی زیر دارای توزیع t با درجه آزادی $v = n - 1$ می‌باشد.

$$T = \frac{X - ?}{S/\sqrt{n}}$$

از طرفی داریم:

$$P(-t \quad / \ < \frac{X - ?}{S/\sqrt{n}} < \ t \quad / \ ) \ = \ 1 - \alpha$$

بنابراین برای یک نمونه با حجم n $(n < 30)$، که میانگین و انحراف معیار s از آن محاسبه شده باشد، $(\ - 1)$ 100% فاصله اطمینان عبارت است از :

$$^{-} - \frac{}{\sqrt{}} \quad / < \quad < ^{-} + \frac{}{\sqrt{}} \quad /$$

مثال) وزن 10 روغن برحسب کیلوگرم به صورت زیر گزارش شده است

9.9,9.8,10.1,10.3,10.1,9.7,10.2, 9.8,10.8, 10.4

یک فاصله اطمینان 99% برای میانگین تمام روغن‌های مشابه تعیین کنید، در صورتی که بدانیم توزیع روغن‌ها تقریبا نرمال است.

حل: با استفاده از نمونه داده شده داریم:

$$^{-} = 10.06\ , s = 0.245$$

$$1 - \quad = \ 0.99 \Rightarrow \quad = 0.01 \ \Rightarrow 1 - \quad /2 \ = \ 0.995$$

با استفاده از جدول فیشر با درجه آزادی 9 داریم $t_{0.995} = 3.25$ و با استفاده از فاصله اطمینان برای این حالت داریم:

$$^{-} - \frac{}{\sqrt{}} \quad / < \quad < ^{-} + \frac{}{\sqrt{}} \quad /$$

$$-10.06 - \frac{\cdot}{\sqrt{}} \times 3.25 < \quad < \ +10.06 - \frac{\cdot}{\sqrt{}} \times 3.25$$

$$9.8 \ < \ ? \ < \ 10.31$$

مثال) یک نمونه تصادفی 10تایی از پیچ ساخته شده توسط کارخانه A انتخاب و قطر آن‌ها را اندازه می‌گیریم، میانگین $= 4.38$ سانتی متر و انحراف معیار $s = 0.06$ بدست آمده

است. اگر توزیع قطر پیچها تقریبا نرمال باشند، یک فاصله اطمینان 95% برای میانگین قطر تمام پیچ‌های ساخته شده مشابه، تعیین کنید.

حل: ابتدا $t_{0.975}$ با درجه آزادی 9 را از جدول پیدا می‌کنیم $t_{0.975} = 2.26$ با توجه به فرمول زیر، مقدار آن را محاسبه می‌کنیم.

$$^{-} - \frac{}{\sqrt{}} \quad / \quad < \quad < {}^{-} + \frac{}{\sqrt{}} \quad /$$

$$3.38 - \frac{\cdot}{\sqrt{}} \times 2.26 < \quad < 4.38 + \frac{\cdot}{\sqrt{}} \times 2.26$$

$$4.438 < \quad < 4.422$$

6-3-4-2: برآورد واریانس یک جامعه (n<30)

یک برآوردگر خوب برای ، برآوردگر S^2 است، بنابراین یک مقدار از این برآوردگر S^2، که از یک نمونه تصادفی با حجم n محاسبه شود، برآورد نقطه‌ای خوبی برای است. حال سعی می‌کنیم به کمک توزیع این برآوردگر، یک فاصله اطمینان برای تعیین کنیم. می‌دانیم که اگر نمونه‌ها از یک جامعه نرمال انتخاب شوند. آماره

$$X2 = \frac{(n-1)\ S}{\sigma}$$

دارای توزیع خی-دو با درجه آزادی $v = \quad - 1$ است، از طرفی می‌دانیم

$$P(\chi\_ < X < \chi \quad \_) = 1 - \alpha$$

بنابراین داریم:

$$P[\frac{(n-1)S}{\chi \quad \_} < \sigma < \frac{(n-1)S}{\chi\_}] = 1 - \alpha$$

بنابراین فاصله اطمینان به صورت زیر است:

$$\frac{(\quad)}{\_} < \sigma < \frac{(\quad)}{\_}$$

S واریانس محاسبه شده از نمونه می‌باشد.

مثال) یک نمونه تصادفی با حجم n=20 از یک جامعه نرمال، دارای میانگین x=32.8 انحراف معیار s=4.51 است.یک فاصله اطمینان 95% برای و از این جامعه بدست آورید.

حل: با توجه به جدول توزیع خی-دو با درجه آزادی 19 داریم.

$$, \chi_{.} \quad = 8.91 \quad \chi^2_{1-\frac{\alpha}{2}} = \chi^2_{0.975} = 32.9$$

$$\frac{\times(\ .\ )}{.} < \quad < \frac{\times(\ .\ )}{.} \quad , \quad 11.75 < \quad < 43.37$$

$$3.42 < \sigma < 6.58$$

5-2: آزمایش فرضیه‌ها

هر ادعا یا حدسی در مورد پارامتر یک جامعه که ممکن است درست یا نادرست باشد را یک فرض آماری می‌گویند.

در بسیاری از موارد بررسی الگوها، برآورد کردن مقدار عددی پارامتر جامعه، مورد استفاده مستقیم ما نیست و اگر هم برآورد مقدار واقعی پارامتر را بخواهیم باید از طریق سرشماری، این پارامتر را بدست آوریم که در عمل ممکن است این امر امکان پذیر نباشد. لذا با استفاده از روش‌های نمونه‌گیری به بررسی حدس و گمان‌هایی که در مورد آن‌ها زده می‌شود می‌پردازیم که به این عمل آزمون فرض گفته می‌شود.

1-5-2: خطای آزمون‌ها

فرض صفر و فرض متقابل: در آزمون فرض همواره با دو فرضیه مواجه هستیم، که یکی را فرض صفر ( ) و دیگری را فرض مقابل ( ) می‌نامیم.

الف- خطای نوع اول: رد کردن فرض در حالی که این فرض درست باشد، خطای نوع اول گفته می‌شود. به احتمال خطای نوع اول یا سطح معنی دار گفته می‌شود.

α =P(رد | درست نباشد)

ب- خطای نوع دوم: پذیرش فرض در حالی که فرض درست نباشد را خطای نوع دوم گویند، به احتمال خطای نوع دوم β گفته می‌شود.

β=P(پذیرش | درست باشد)

توان آزمون: احتمال رد کردن فرض در صورتی که درست باشد، یعنی احتمال رد کردن فرض به درستی را توان آزمون گویند. توان آزمون مکمل خطای نوع دوم است.

توان آزمون = *= = (فرض رد می‌شود | درست باشد)

-1=(درست باشد |پذیرش) -1= =1-

مثال) فرض کنید یک نوع کپسول سرماخوردگی (A) چند سال است که در کشور تولید می‌شود و ثابت شده است که پس از 5 روز، 70% موثر می‌باشد، کارخانه‌ای مدعی است که کپسول مشابه A با نام B تولید کرده است که اثر آن پس از 5 روز، بیشتر از 70% است. ملاحظه می‌کنید که در اینجا دو فرض آماری هستند که در مقابل یکدیگر قرار می‌گیرند.

$$\begin{cases} : p = 0.7 \text{ (کپسول A پس از 5روز،70\% موثر است)} : \\ : P > 0.7 \text{ (کپسول B پس از 5روز، بیشتر از 70\%موثر است)} : \end{cases}$$

حال می‌خواهیم بررسی کنیم با توجه به هزینه‌ای که برای ساخت کپسول B باید مصرف شود، آیا استفاده از آن فایده‌ای دارد. واضح است که کپسول B زمانی استفاده خواهد شد که از کیفیت بالاتری نسبت به A برخوردار باشد. برای بررسی این مطلب دو فرض آماری مانند A را در مقابل یکدیگر قرار می‌دهیم، و به کمک نمونه‌گیری یکی را رد و دیگری را قبول می‌کنیم. فرض کنید یک نمونه تصادفی 20 تایی از افراد را انتخاب و کپسول B را به آن‌ها داده‌ایم، اگر در این آزمایش پس از 5 روز، تاثیر دارو بر روی کمتر از 16 نفر موثر واقع شده باشد، ادعای کارخانه B را بی‌مورد می‌دانیم. چون تصمیم‌گیری به کمک نمونه انجام می‌شود، امکان دو نوع خطا اجتناب ناپذیراست، یکی آنکه را رد کنیم در صورتی که درست باشد یا به عبارت دیگر نمونه نشان دهد که دارو بر روی 16 نفر یا بیشتر موثر بوده است و دیگری آنکه فرض را بپذیریم در صورتی که درست نباشد. به عبارت دیگر، نمونه نشان دهد که دارو بر روی کمتر از 16 نفر موثر است. این دو نوع خطا را که عامل اصلی آن نمونه می‌باشد را خطای نوع اول و دوم می‌نامیم.

در این مثال آماری که براساس آن تصمیم گیری صورت می‌گیرد، تعداد افرادی است که اثر کپسول B در آنها موثر بوده است نقش تعیین کننده‌ای دارد. در حقیقت صفر تا 20 را به دو دسته تقسیم می‌کنیم، یکی اعداد کمتر از 16 و دیگری اعداد بزرگتر یا مساوی 16، اگر نتیجه آزمایش در دسته اول قرار گیرد، تصمیم می‌گیریم که کپسول B تفاوت چندانی با

کپسول A ندارد و در نتیجه را قبول می‌کنیم اما اگر نتیجه آزمایش در دسته دوم قرار گیرد، تصمیم می‌گیریم که کپسول B از A بهتر است و در نتیجه فرض را رد می‌کنیم. در واقع 0 تا 20 را به دو ناحیه تقسیم می‌کنیم، یکی ناحیه رد و دیگری ناحیه قبول، در زیر به تعریف این دو ناحیه می‌پردازیم.

ناحیه‌ای که باعث رد فرض صفر می‌شود را ناحیه بحرانی یا ناحیه رد و ناحیه‌ای که باعث قبول فرض صفر می‌شود را ناحیه قبولی و مرز بین این دو ناحیه را نقطه بحرانی می‌نامیم.

= P(خطای نوع اول) = P(رد | درست باشد)

$$= P(X > 15.5 | \quad = 0.7) = \qquad b(x;\ 20,0.7) = 0.4375$$

β = P(خطای نوع دوم) = P(H پذیرش | H باشد درست)

$$= P(X < 15.5 | \quad > 0.7)$$

ملاحظه می‌کنید که β قابل محاسبه نمی‌باشد، زیرا فرض مقابل به طور کامل مشخص نشده است، اما اگر فرض مقابل را تغییر داده و به صورت 0.8 = : در نظر بگیریم، β قابل محاسبه شده و داریم.

$$= \quad (\quad < 15.5 \mid p = 0.8) = \sum \qquad (x; 20,0.8) = 0.3704$$

در یک آزمون خوب باید هر دو خطای گفته شده در آن کم باشد، به عبارت دیگر نه فرضی را بی مورد رد و نه فرضی را بی مورد قبول کنیم، در واقع شرایط ایده‌آل آنست که α و β به طور هم‌زمان کوچک باشند.

با تعدیل ناحیه بحرانی می‌توان فهمید که، α و β به هم وابسته بوده و کاهش یکی باعث افزایش دیگری می‌شود، پس با تغییر ناحیه بحرانی می‌توان یکی ازخطاها را کم و دیگری را زیاد نمود.

اکنون می‌خواهیم ببینیم که چگونه می‌توان این خطاها را هم‌زمان با هم کاهش داد. در مثال مورد بررسی، یک نمونه 100 تایی انتخاب و نقطه بحرانی را به دلخواه عدد 73,5 اختیار می‌کنیم، مقادیر α و β به صورت زیر محاسبه می‌شود:

$$\alpha = P(X > 73.5 \mid p = 0.7)$$

چون حجم نمونه بزرگ است، برای محاسبه از توزیع نرمال با میانگین و واریانس زیر استفاده می‌کنیم.

$$? = np = 100 \times 0.7 = 70 \quad , \quad = npq = 21$$

$$\simeq P(Z > \frac{}{}) = 1 - P(Z < 0.76) = 0.2236$$

و به همین ترتیب برای محاسبه β نیز همانند بالا از توزیع نرمال با میانگین و واریانس زیر استفاده می‌کنیم.

$$? = np = 100 \times 0.8 = 80 \quad , \quad = npq = 16$$
$$\beta = P(X < 73.5 \mid p = 0.8) \simeq P(Z < \frac{}{})$$
$$= P(Z < -1.625) = 0.0521$$

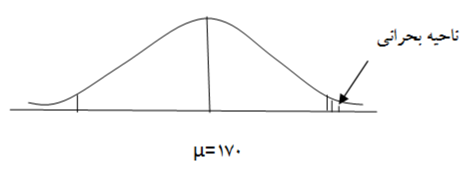
بنابراین با افزایش حجم نمونه دو خطا به طور همزمان کاهش می‌یابد.

با توجه به این مثال می‌توان نتیجه گرفت که اگر آزمون فرض به یک توزیع گسسته مربوط می‌شود، برای بررسی آن باید نقطه بحرانی را به طور دلخواه انتخاب کرد و به کمک آن احتمال خطاهای نوع اول و دوم (در صورت امکان) را بدست آورده و اگر این مقادیر بزرگ نباشند ناحیه بحرانی را مشخص و چنانچه نتیجه آزمایش در ناحیه بحرانی قرار گرفت، فرض صفر را رد و در غیر این صورت فرض صفر پذیرفته می‌شود. اما اگر احتمال خطاهای نوع اول و دوم زیاد باشد، باید نقطه بحرانی را تغییر دهیم (حجم نمونه ثابت) تا مقادیر α و β به مقادیر مورد نظر برسند ولی اگر با تغییر نقطه بحرانی مقادیر α و β به مقادیر مطلوب نرسند، در صورت امکان حجم نمونه را افزایش می‌دهیم.

مثال) بررسی توزیع آماره پیوسته

فرض کنید برآورد شده است که میانگین قد دانشجویان یک دانشگاه 171سانتیمتر و انحراف معیار 9 سانتیمتر باشد. برای بررسی درستی یا نادرستی این گزارش (فرض آماری) آزمونی به صورت زیر انجام می‌دهیم: (آزمون فقط برای میانگین انجام می‌شود.)

$$\begin{cases} : ? = 170 & (\ = 9) \\ : ? \neq 170 & (\ = 9) \end{cases}$$



7-2: بررسی فرض آماری

برای بررسی این فرض آماری، یک نمونه تصادفی n= 36 تایی از دانشجویان را انتخاب و طول قد آن‌ها را اندازه می‌گیریم، می‌دانیم آماره‌ای که برای بررسی این گزارش استفاده می‌شود آماره X است. ابتدا ناحیه‌ای بحرانی را درنظر گرفته و به کمک آن α و β را (در صورت امکان) محاسبه می‌کنیم و به کمک آن ناحیه قبول را می‌پذیریم:

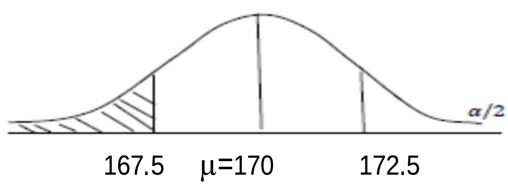
فرض کنید ناحیه بحرانی به طور دلخواه و به صورت زیر انتخاب شود.

ناحیه بحرانی $X < 167.5$ یا $X > 172.5$

ناحیه قبول $167.5 < \quad < 172.5$

به عبارت دیگر اگر میانگین نمونه، نمونه تصادفی 36تایی از 172,5 بیشتر و از 167,5 کمتر باشد می‌پذیریم که میانگین قد دانشجویان 170 سانتیمتر نبوده و در قد دانشجویان تغییر به وجود آمده‌است. اما اگر میانگین نمونه 36تایی بین 167,5 و 172,5 قرار گیرد، می‌پذیریم که قد تغییر نکرده و میانگین همان 170 سانتیمتر است و با استفاده از قضیه حد مرکزی، X دارای توزیع نرمال با میانگین 170 = ? سانتیمتر و انحراف معیار

$= \frac{}{\sqrt{}} = \frac{}{\sqrt{}} = 1.5$ است.



8-2: ناحیه بحرانی و ناحیه قبول

اکنون برای آن‌که بتوانیم درستی فرض مشخص کنیم α و β را محاسبه می‌کنیم.

$\alpha = P(\bar{X} > 172.5 \mid ? = 170) + P(X < 167.5 \mid ? = 170)$

$= P(Z > \frac{\cdot}{\cdot}) + P(Z < \frac{\cdot}{\cdot})$

$= 1 - P(Z < -1.66) = 2 * 0.0485 = 0.0970$

با توجه به مقدار α می‌توان گفت که 9,7% از تمام نمونه‌های تصادفی که با حجم n=36 انتخاب می‌شوند، باعث رد فرض صفر شده و یا به عبارت دیگر باعث خواهد شد که بپذیریم میانگین قد،170سانتیمتر نیست، در حالی که میانگین واقعی قد همان 170سانتیمتر می‌باشد. اگر بخواهیم این مقدار خطا را کاهش دهیم می‌توان حجم نمونه را بزرگ‌تر اختیار

کرد، به عنوان مثال اگر به جای یک نمونه با حجم n = 36، نمونه با حجم n = 64 از دانشجویان را انتخاب کرده و طول قد آن‌ها را اندازه بگیریم، با خطای کمتری فرض صفر را رد خواهیم کرد. البته در این حالت هیچگونه تغییری در ناحیه بحرانی رخ نداده است و فقط حجم نمونه را به n = 64 تغییر می‌دهیم. با محاسبه α می‌بینیم که مقدار خطای نوع اول با افزایش حجم نمونه، کاهش یافته است.

برای اینکه آزمون خوبی داشته باشیم باید خطای نوع دوم محاسبه شود، اما در حالی که فرض مقابل به صورت ‎170 ≠ ? :‎ باشد، محاسبه β امکان‌پذیر نیست ولی فرض صفر بدست می‌آید. با انتخاب مقادیر مختلف برای μ می‌توان β را محاسبه نمود تا تضمین کافی در مورد رد فرض صفر وجود داشته باشد. به عنوان مثال اگر خطای 2,6% در مثال قبل مورد قبول باشد می‌توان برای مقادیر مختلف ‎172.5 ≥ ?‎ و یا ‎167.5 ≤ ?‎ (چند مقدار دلخواه) خطای نوع دوم را محاسبه کرده و سپس تصمیم گیری صورت می‌پذیرد، مثلا ‎175 = ? :‎ و یا ‎165 = ? :‎ اختیار شده و مقدار β را برای این دو فرض محاسبه می‌نمائیم. هنگامی که فرض مقابل ‎175 = ? :‎ و ‎64 =‎ باشد، داریم:

β = (167.5 < < 172.5 | ? = 175) = (——— < < ———)

= (−6.66 < < −2.22) = (< −2.22)− (< −6.66)

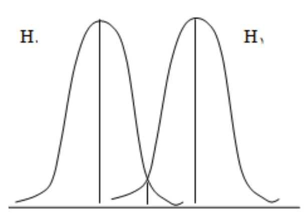
= 0.0132

و اگر ‎165 = ? :‎ باشد، آنگاه داریم.

= (167.5 < ¯ < 172.5 | ? = 165) = (——— < < ———)

= (2.22 < < 6.66) = (< 6.66) − (< 2.22)

= 1 − 0.09868 = 0.0132



2-9: ناحیه بحرانی

بنابراین با توجه به مقدار β در این دو حالت، می‌توان گفت که اگر $165 > ?$ و یا $175 < ?$ باشد، مقدار β از 0,0132 نیز کمتر خواهد شد، هنگامی که $n = 64$ اختیار شود در نتیجه شانس پذیرفتن هنگامی که نادرست باشد، بسیار کم است و چنین آزمونی، آزمون خوبی است.

اگر آماره‌ای که تصمیم گیری براساس آن صورت می‌گیرد، دارای توزیع پیوسته باشد می‌توان ابتدا خطای نوع اول را اختیار کرد و به کمک آن ناحیه بحرانی را تعیین نمود و نیازی نیست که مانند گسسته‌ها ابتدا ناحیه بحرانی را انتخاب و α را محاسبه کنیم و در صورتی که قابل قبول نباشد، ناحیه بحرانی را تغییر دهیم.

با توجه به دو مثال که بررسی شد، نتایج زیر را برای آزمون فرض به طور کلی می‌توان بیان نمود.

1- در هر آزمون فرض دو نوع خطا وجود دارد، خطای نوع اول و خطای نوع دوم.
2- خطای نوع اول و دوم به یکدیگر وابسته بوده و افزایش یکی باعث کاهش دیگری می‌شود.
3- با تغییر ناحیه بحرانی (ناحیه‌های بحرانی) می‌توان احتمال خطای نوع اول α و یا سطح تشخیص را تغییر داد.
4- با افزایش مقدار حجم نمونه n، می‌توان هر دو خطا را به طور همزمان کاهش داد.
5- اگر فرض صفر نادرست باشد، مقدار β ماکزیمم خواهد شد، هنگامی که مقدار واقعی پارامتر موردنظر به عنوان فرض مقابل انتخاب شده باشد.

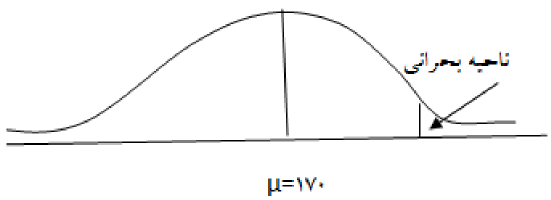
2-5-2: آزمون یک‌طرفه

آزمون فرض را یک‌طرفه می‌نامیم اگر فرض مقابل آن یک‌طرفه باشد، به عبارت دیگر آزمون به یکی از دو صورت زیر باشد.

الف $\begin{cases} :\theta = \theta \\ :\theta > \theta \end{cases}$ ب $\begin{cases} :\theta = \theta \\ :\theta < \theta \end{cases}$

در چنین آزمونی، ناحیه بحرانی همواره در یک سمت توزیع فراوانی قرار دارد. در حالتی که آزمون الف مورد نظر باشد، ناحیه بحرانی درانتهای سمت راست توزیع قرار دارد. مانند آزمون طول قد دانشجویان یک دانشکده اگر به صورت زیر باشد.

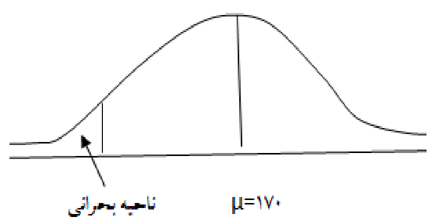
: ? = 170
: ? > 170



2-10: ناحیه بحرانی درانتهای سمت راست در آزمون یک‌طرفه

و اگر آزمون ب مورد نظر باشد، ناحیه بحرانی در انتهای سمت چپ توزیع قرار دارد. مانند آزمون قد دانشجویان اگر به صورت زیر درنظر گرفته شود.

: ? = 170
: ? < 170



2-11: ناحیه بحرانی درانتهای سمت چپ در آزمون یک‌طرفه

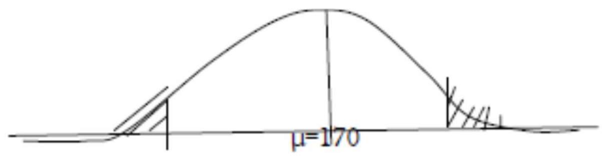
2-5-3: آزمون دوطرفه

آزمون فرض را دو طرفه گوئیم اگر فرض مقابل دوطرفه باشد، به عبارت دیگر آزمون به صورت زیر باشد.

$: \theta = \theta \quad , \quad : \theta \neq \theta$

در این آزمون ناحیه بحرانی در دو انتهای توزیع قرار دارد، مثلا اگر آزمون قد دانشجویان به صورت زیر باشد.

: ? = 170
: ? ≠ 170



12-2: ناحیه بحرانی در آزمون دوطرفه

چگونگی تشکیل یک آزمون یک‌طرفه یا دو طرفه بستگی به نتیجه خواسته شده از آزمون فرض دارد.

به عنوان مثال اگر بخواهیم ادعای کارخانه‌ای را بررسی کنیم که مدعی است نوعی لامپ تولید کرده که از لامپ‌های مشابه موجود در بازار مرغوب‌تر است، ابتدا فرض می‌کنیم لامپ جدید مشابه لامپ موجود است و آزمون یک‌طرفه‌ای به صورت زیر تشکیل می‌دهیم

(p_0 نسبت تعداد لامپ‌هایی که طول عمر آنها از 100 ساعت بیشتر است)

$$\begin{cases} : p = p_0 \\ : p > {}_0 \end{cases}$$

و اگر بخواهیم بررسی کنیم که لامپ جدید مرغوب‌تر از لامپ موجود نیست، آزمون یک‌طرفه به صورت زیر تشکیل می‌دهیم.

$: p = p_0 \quad , \quad : p < {}_0$

اما اگر بخواهیم ببینیم که لامپ جدید از نظر مرغوبیت نسبت به لامپ موجود چه وضعی دارد، آزمونی به صورت زیر درنظر می‌گیریم

$: p = p_0 \quad , \quad : p \neq p_0$

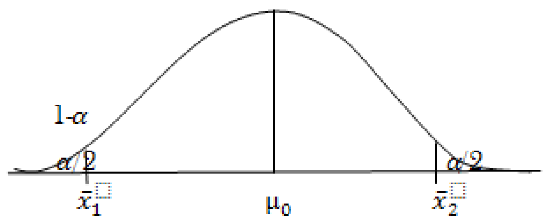
4-5-2: آزمون میانگین و واریانس

فرض کنید می‌خواهیم μ، میانگین جامعه‌ای را در سطح تشخیص α آزمون کنیم، در صورتی که بدانیم واریانس جامعه مقدار معلوم σ است. برای انجام آزمونی به صورت زیر

: ? = ?₀ , : ? ≠ ?₀

ابتدا آماره مناسبی برای μ انتخاب می‌کنیم، می‌دانیم آماره مناسب برای این منظور آماره است. از طرفی می‌دانیم که برای n نسبتا بزرگ ($n >= 30$) توزیع این آماره، توزیع نرمال با میانگین μ و واریانس — می‌باشد. حال با توجه به اینکه سطح تشخیص و

آزمون دو طرفه است. می‌توان دو نقطه بحرانی ¯ و ¯ را بدست آورد و به کمک آن ناحیه بحرانی را مشخص نمود.



13-2: ناحیه بحرانی در آزمون دوطرفه با سطح تشخیص

ناحیه بحرانی ¯ > X, ¯ <X   ناحیه قبول ¯ > X < ¯

پس از تشکیل ناحیه بحرانی، مقدار میانگین نمونه را با استفاده از نمونه منتخب محاسبه کرده و در صورتی که این مقدار در ناحیه بحرانی قرار گیرد، فرض صفر را رد می‌کنیم.

باید توجه داشت، در مواردی که توزیع آماره مناسب، نرمال است می‌توان کلیه مراحل را با استفاده از توزیع نرمال استاندارد انجام داد. بدین ترتیب که با استفاده از متغیر $\frac{}{/\sqrt{}}$ آماره مناسب را به نرمال استاندارد تبدیل کرد و ناحیه بحرانی و نقاط بحرانی را بر اساس این توزیع بنویسیم، در این صورت ناحیه بحرانی عبارت است از

> _ , < _

و ناحیه قبول _ > Z < _ - است. پس از تشکیل ناحیه بحرانی به کمک مقدار میانگین نمونه ¯، مقدار $\frac{}{/\sqrt{}}$ را محاسبه می‌کنیم، اگر این مقدار در ناحیه بحرانی قرار گرفت، فرض صفر را رد می‌کنیم.

تمام مراحلی که در بالا انجام شد. به طور کلی برای هر آزمونی که آماره مناسب آن دارای یکی از توزیع‌های Z، T، X^2 و F باشد، قابل استفاده می‌باشد. با توجه به مطالب بالا می‌توان یک صورت کلی برای آزمون فرض‌هایی ارائه کرد که توزیع آماره آن‌ها به یکی از چهار صورت فوق باشد. برای چنین آزمون‌هایی مراحل زیر به کار برده می‌شود.

1- تشکیل فرض صفر   : =

2- تشکیل فرض مقابل

: < or > or ≠

3- انتخاب میزان معنی‌دار بودن (سطح تشخیص)

4- انتخاب آماره مناسب برای پارامتر θ، تشخیص نوع توزیع آن و سپس تشکیل ناحیه بحرانی.

5- محاسبات: به کمک نمونه تصادفی انتخاب شده از جامعه، مقدار آماره مناسب محاسبه می‌شود.

6- نتیجه‌گیری: اگر مقدار آماره در ناحیه بحرانی قرار گیرد فرض رد می‌شود و در غیر این صورت فرض پذیرفته می‌شود.

مثال) فرض کنید دارای توزیع نرمال میانگین ? و واریانس 36 است. در آزمونی به صورت

: ? = 50 , : ? = 55

ناحیه بحرانی و n را طوری تعیین کنید که به ازای آن‌ها خطای نوع اول 0,05 و خطای دوم 0,1 باشد.

فرض کنید نقطه بحرانی باشد. با توجه به تعاریف α و β داریم

= P(> c | ? = 50) = 0.05

$0.05 = P(Z > \frac{}{/\sqrt{}}) \Rightarrow P(Z < \frac{}{/\sqrt{}}) = 0.95$

$c - 50 = 1.645 \times \frac{}{\sqrt{}}$ (1)

$0.1 = P(< c \mid ? = 55) = P(Z < \frac{}{/\sqrt{}})$

$c - 55 = -1.282 \times \frac{}{\sqrt{}}$ (2)

از (1) و (2) نتیجه می‌گیریم، n = 13 و c = 52.8

مثال) یک نمونه 5 تایی از جامعه‌ای نرمال انتخاب کرده ایم، نتایج به صورت زیر است

1,9، 2,4، 3,5، 4,2، 3

آیا می‌توان گفت، میانگین جامعه کمتر از 3,5 است.

1. : ? = 3.5
2. : ? < 3.5
3. $\alpha = 0.05$

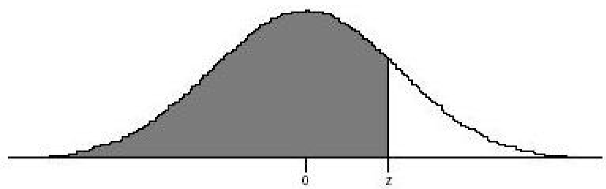
4. $T = \frac{}{/\sqrt{}}$ و ناحیه بحرانی $T < -t_{0.95} = -2.132$

5. $t = \frac{}{.\ \sqrt{}} = -1.238$ $s = 0.9$ و $\bar{} = 3$

6. چون -1.238 بزرگتر از -2.132 است، پس رد نمی‌شود

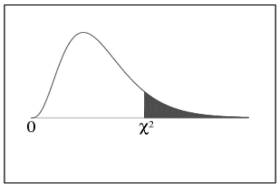
در این فصل ضمن تعریف قوانین مجموعه‌ها و طرح مسائل مختلفی از احتمالات، به تعریف توزیع‌های احتمالاتی گسسته و پیوسته وکاربرد آن در نظریه آمار واحتمال پرداخته شده است. با توجه به اینکه در روش‌های مبتنی بر مدل، ابتدا یک مدل که نزدیکترین مدل به رفتار داده‌ها است را در نظر می‌گیریم و تحلیل را بر روی آن مدل انجام می‌دهیم، در ادامه، نحوه انتخاب مدل مناسب، پارامترهای لازم و نتیجه گیری و استنباط بر اساس نتایج به‌دست‌آمده، مورد بحث و بررسی قرار گرفته است.

پیوست:



Normal Deviate z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-4.0	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.9	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.8	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.7	.0001	.0001	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379

Chi-Square Distribution Table



The shaded area is equal to α for $\chi^2 = \chi^2_\alpha$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

مراجع:

1. بزرگ نیا. ابوالقاسم، صادقی. حسن، توردالنیوس، "مبانی بررسی‌هایی نمونه‌ای" ، انتشارات آستان قدس مشهد،1370
2. شیروانی. پرویز، "نظریه نمونه‌گیری"
3. Irwin Guttman, S.S.Wilks, J.stuart Hunter , “Introductory Engineering Statistics “
4. Richard L.Scheaffer, “Introduction to Probability and its applications”
5. Jakob Verbeek, Lecture’s notes “MIXTURE DENSITY ESTIMATION”
6. Murry r Spiegel, “theory and problems of statistics”