الف- اگر ۹۷/۵ درصد از کالاها کمتر از ۲۹/۸ دقیقه مونتاژ شوند مقدار σ را تعیین کنید.

ب- $P(|X-10|<5)$ را محاسبه کنید.

ج- اگر ۳ عدد از این کالا به طور مستقل و همزمان مونتاژ شوند، احتمال اینکه حداقل یکی از آنها کمتر از ۲۰ دقیقه مونتاژ شود را بیابید.

۶۸ در یک هتل متقاضیان به طور متوسط ۵ نفر در ساعت مراجعه می‌کنند. فرض کنید تا ۱۰ دقیقه قبل هنوز متقاضی نیامده باشد.

الف- احتمال اینکه متقاضی بعدی کمتر از ۲ دقیقه دیگر بیاید را بیابید.

ب- احتمال اینکه فاصله زمانی آمدن دهمین و یازدهمین متقاضی از دو دقیقه تجاوز نکند را بیابید.

۶۹ فرض کنید طول عمر لامپهای روشنایی شرکتی معین دارای توزیع نرمال با میانگین ۱۰۰۰ و انحراف معیار ۱۰۰ ساعت باشد. همچنین فرض کنید طول عمر لامپهای ساخت شرکتی دیگر نیز دارای توزیع نرمال با میانگین ۹۰۰ و انحراف معیار ۱۵۰ ساعت باشد. اگر از هر یک از این شرکتها یک لامپ خریداری شود، احتمال اینکه حداقل یکی از آنها ۹۸۰ ساعت یا بیشتر عمر کند را بیابید.

فصل ششم

توزیعهای نمونه‌ای

۱.۶ نمونه تصادفی و توزیع نمونه‌ای

همانگونه که در فصل اول اشاره شد آمار به عنوان یک موضوع علمی، شامل مفاهیم و روشهایی جهت جمع آوری، سازماندهی و خلاصه کردن داده‌های به دست آمده از قسمتی از یک جمعیت و انجام استنباط و نتیجه گیری از روی آنها در مورد کل جمعیت می‌باشد.

در هر مطالعه آماری با مجموعه افراد یا اشیاء سر و کار داریم که هدف از مطالعه کسب اطلاعات در مورد آنها می‌باشد. این مجموعه را اصطلاحاً **جمعیت** و افراد یا اشیاء تشکیل دهنده آن را اعضاء یا عناصر جمعیت گویند. در بسیاری موارد به علت پرهزینه بودن، وقت‌گیر بودن و... اطلاعات فقط از قسمتی از جمعیت جمع آوری می‌گردد و بر اساس اطلاعات حاصله از این قسمت استنباط هایی در مورد تمام جمعیت انجام می‌گیرد. انتخاب قسمتی از جمعیت را **نمونه گیری** و قسمت انتخاب شده را یک **نمونه** گویند.

در این فصل برخی از مفاهیم نمونه گیری که در فصلهای بعد مورد استفاده قرار می‌گیرند را می‌آوریم.

نمونه تصادفی

در مبحث نمونه گیری معمولاً جمعیت را با یک متغیر تصادفی X نمایش می‌دهند که این متغیر تصادفی X دارای یک توزیع احتمال بخصوص $f_X(x)$ است. این توزیع احتمال را توزیع احتمال جمعیت گویند. برای مثال اگر جمعیت لامپهای تولیدی یک کارخانه باشد و بخواهیم

مطالعه‌ای روی طول عمر لامپها داشته باشیم آنگاه X را طول عمر لامپ تولیدی کارخانه در نظر می‌گیریم و فرض می‌کنیم $X\sim E(\theta)$. همچنین اگر جمعیت نوزادان باشند و بخواهیم روی قد نوزادان مطالعه‌ای انجام دهیم، در این صورت X را طول قد نوزاد در نظر می‌گیریم و فرض می‌کنیم $X\sim N(\mu,\sigma^2)$. حال فرض کنید بخواهیم از این جمعیت یک نمونه به اندازه n انتخاب کنیم. اگر مقدار اولین عضو انتخابی را با X_1 و مقدار دومین عضو انتخابی را با X_2 و... و مقدار nامین عضو انتخابی را X_n نمایش دهیم در این صورت نمونه ما به صورت X_1، X_2، ... و X_n می‌باشد که هر کدام دارای همان توزیع احتمال جمعیت یعنی $f_X(x)$ می‌باشند و از یکدیگر مستقل هستند. این نمونه را **نمونه تصادفی** گویند. برای مثال اگر در اندازه‌گیری قد نوزادان نمونه‌ای از ۳ نوزاد انتخاب کرده و اندازه قد نوزاد iام را با X_i (i=۱,۲,۳) نشان دهیم آنگاه X_1، X_2 و X_3 نمونه تصادفی ما را تشکیل می‌دهند که هر کدام دارای همان توزیع جمعیت یعنی $N(\mu,\sigma^2)$ هستند. اگر پس از اندازه‌گیری روی ۳ نوزاد مقادیر $x_1=45$ و $x_2=50$ و $x_3=47$ سانتیمتر را به دست آوریم آنگاه x_1، x_2 و x_3 را مقادیر مشاهده شده این نمونه تصادفی گویند.

برای درک مفهوم نمونه تصادفی به مثال زیر توجه کنید.

مثال ۶.۱.۱ چهار مهره به شماره‌های ۱، ۲، ۳ و ۴ وجود دارد. می‌خواهیم یک نمونه تصادفی به اندازه n=۲ از این مهره‌ها انتخاب کنیم. اگر متغیر تصادفی X نمایانگر شماره روی یک مهره انتخابی باشد آنگاه X دارای تابع احتمال یکنواخت زیر است

x	۱	۲	۳	۴
$f_X(x)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

اگر X_1 و X_2 به ترتیب نمونه‌های انتخاب شده در مرتبه اول و دوم باشند آنگاه تعداد نمونه‌های تصادفی ممکن به فرم (X_1,X_2) برابر $16=4^2$ است که هر کدام دارای شانس یکسان $\frac{1}{16}$ برای انتخاب شدن می‌باشند. در جدول ۶.۱ نمونه‌های مختلف همراه با احتمالات آنها آورده شده است.

احتمال	$\bar{x}$	(x_1,x_2)	شماره نمونه	احتمال	$\bar{x}$	(x_1,x_2)	شماره نمونه
$\frac{1}{16}$	۲	(۳,۱)	۹	$\frac{1}{16}$	۱	(۱,۱)	۱
$\frac{1}{16}$	۲/۵	(۳,۲)	۱۰	$\frac{1}{16}$	۱/۵	(۱,۲)	۲
$\frac{1}{16}$	۳	(۳,۳)	۱۱	$\frac{1}{16}$	۲	(۱,۳)	۳
$\frac{1}{16}$	۳/۵	(۳,۴)	۱۲	$\frac{1}{16}$	۲/۵	(۱,۴)	۴
$\frac{1}{16}$	۲/۵	(۴,۱)	۱۳	$\frac{1}{16}$	۱/۵	(۲,۱)	۵
$\frac{1}{16}$	۳	(۴,۲)	۱۴	$\frac{1}{16}$	۲	(۲,۲)	۶
$\frac{1}{16}$	۳/۵	(۴,۳)	۱۵	$\frac{1}{16}$	۲/۵	(۲,۳)	۷
$\frac{1}{16}$	۴	(۴,۴)	۱۶	$\frac{1}{16}$	۳	(۲,۴)	۸

جدول ۶.۱ جدول نمونه‌های ۲ تایی ممکنه از جمعیت ۴ تایی

مشاهده می‌شود که عناصر اول و دوم نمونه تصادفی یعنی X_1 و X_2 قبل از اینکه مشاهده شوند به عنوان متغیرهای تصادفی در نظر گرفته می‌شوند و مقادیر ۱، ۲، ۳ و ۴ را با توابع احتمال زیر اختیار می‌کنند

x_i	۱	۲	۳	۴
$f_{X_i}(x_i)=P(X_i=x_i)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

i=۱,۲

که دقیقاً همان توزیع احتمال جمعیت یعنی متغیر تصادفی X می‌باشد.

تعریف ۶.۱ فرض کنید متغیر تصادفی X نمایانگر یک جمعیت با توزیع احتمال $f_X(x)$ باشد. یک نمونه تصادفی به اندازه n از این جمعیت عبارت است از جمع آوری n متغیر تصادفی مستقل X_1، X_2، ... و X_n که هر کدام دارای توزیع احتمال $f_X(x)$ هستند. مقادیر مشاهده شده این نمونه تصادفی را با x_1، x_2، ... و x_n نمایش می‌دهند.

پارامتر و آماره

هر ویژگی یک جمعیت را یک پارامتر جمعیت و یا به اختصار پارامتر گویند و ویژگی

متناظر را در نمونه آماره می‌نامند. به عنوان مثال میانگین جمعیت μ یک پارامتر جمعیت است در حالی که اگر از جمعیت نمونه‌گیری کرده و میانگین نمونه $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ را محاسبه کنیم، این میانگین نمونه یک آماره است. به عبارت دقیقتر یک آماره تابعی از نمونه تصادفی است که به پارامتر مجهول بستگی ندارد. توجه کنید که مقدار آماره از یک نمونه به نمونه دیگر تغییر می‌کند ولی مقدار پارامتر جمعیت همواره ثابت است. برای مثال در مثال ۱.۱.۶ میانگین جمعیت مقداری ثابت و برابر است با $\mu=E(X)=\sum_{x=1}^{4}xf_X(x)=\frac{1}{4}(1+2+3+4)=2/5$

در حالی که میانگین نمونه $\overline{X}$ بسته به اینکه چه نمونه‌ای استخراج شده باشد متغیر خواهد بود.

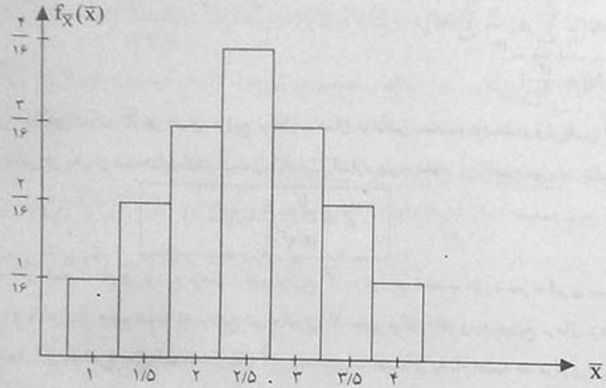
در آمار با انجام نمونه‌گیری و به دست آوردن نمونه تصادفی X_1، X_2،... و X_n آماره‌هایی مانند $\overline{X}$ و $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ را به عنوان میانگین و واریانس نمونه محاسبه کرده و از روی این آماره‌ها استنباط‌هایی در مورد پارامترهای جمعیت مانند میانگین جمعیت μ و واریانس جمعیت σ^2 انجام می‌دهیم.

توزیع نمونه‌ای

آماره که تابعی از نمونه تصادفی است خود نیز یک متغیر تصادفی است. هر چند در هر وضعیت مفروض فقط با یک نمونه و یا یک مجموعه از مشاهدات سر و کار داریم و آماره مورد نظر فقط یک مقدار متناظر را دارد، ولی با نمونه های مختلف مقدار آماره مطابق با توزیعی که از روی توزیع نمونه تصادفی معین می‌شود، تغییر می‌کند (نکته مهم این است که رفتار آماره را می‌توان بوسیله یک توزیع احتمال توصیف کرد. بنابراین هر آماره یک متغیر تصادفی است و توزیع احتمال آن را توزیع نمونه‌ای گویند).

مثال ۲.۱.۶ در مثال ۱.۱.۶ بر اساس اطلاعات به دست آمده از انتخاب نمونه دو تایی می‌خواهیم میانگین جمعیت μ را حدس بزنیم. برای این منظور شاید بهترین حدس میانگین نمونه دوتایی یعنی $\overline{X}$ باشد. در جدول ۱.۶ میانگین نمونه‌های ممکن دوتایی و احتمالات مربوطه آورده شده است. با توجه به جدول ۱.۶ تابع احتمال و هیستوگرام زیر را برای توزیع احتمال میانگین نمونه $\overline{X}$ داریم.

$\bar{x}$	۱	۱/۵	۲	۲/۵	۳	۳/۵	۴
$P(\overline{X}=\bar{x})$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$



توجه کنید که جدول به دست آمده تمام ویژگی‌های یک توزیع احتمال را دارد. تنها اختلاف این توزیع احتمال با توزیعهای احتمال که تاکنون مورد بررسی قرار داده‌ایم در این است که نتایج به دست آمده در اثر انتخاب یک نمونه تصادفی بوده است و به همین دلیل این توزیع احتمال را توزیع نمونه‌ای گویند. در هیستوگرام فوق مشاهده می‌شود که اگر چه توزیع جمعیت مورد نظر یک توزیع یکنواخت است ولی توزیع نمونه میانگین نمونه $\overline{X}$ تغییر شکل داده و به توزیع نرمال گرایش پیدا کرده است.

در بخش‌های بعد توزیع نمونه‌ای چند آماره مهم را که در فصل بعد کاربرد دارند را مورد بررسی قرار می‌دهیم. توزیع نمونه‌ای $\bar{X}$ (نمونه‌ای به اندازه n) ۱) جامعه نرمال ⇐ توزیع نمونه‌ای $\bar{X}$ ... $\mu_{\bar{x}}=\mu$ / $\sigma^2_{\bar{x}}=\frac{\sigma^2}{n}$ (σ معلوم) ۲) جامعه غیر نرمال ⇐ اگر $n\geq30$ $\frac{\sum x_i}{n}$

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

فرض کنید که از جمعیت X با میانگین μ و واریانس σ^2 نمونه‌ای تصادفی X_1، X_2،... و X_n به اندازه n انتخاب کرده باشیم و بخواهیم توزیع نمونه‌ای میانگین نمونه $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ را به دست آوریم. برای این منظور دو حالت زیر را در نظر می‌گیریم.

الف- اگر جمعیت دارای توزیع نرمال با میانگین μ و واریانس σ^2 باشد چون $\overline{X}$ به صورت یک ترکیب خطی از متغیرهای تصادفی نرمال مستقل می‌باشد، بنابراین طبق قضیه ۱۴.۵ داریم که

$$\bar X \sim N\left(\frac{1}{n}\sum_{i=1}^{n}\mu, \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2\right)$$

و یا

$$\bar X \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

یعنی میانگین نمونه $\bar X$ نیز دارای توزیع نرمال با همان میانگین جمعیت μ است و واریانس آن از واریانس هر یک از مشاهدات کمتر است و با افزایش اندازه نمونه مقدار آن کاهش می‌یابد. بنابراین

$$\boxed{Z=\frac{\bar X-\mu}{\sigma/\sqrt{n}}\sim N(0,1)} \qquad (۱۶)$$

ب- اگر جمعیت دارای توزیع نرمال نباشد توزیع $\bar X$ به توزیع جمعیت مورد نمونه‌گیری بستگی دارد و با افزایش حجم نمونه n، توزیع نمونه‌گیری $\bar X$ تغییر شکل داده و به توزیع نرمال نزدیک می‌شود. این موضوع چنان اهمیت دارد که آن را به صورت قضیه‌ای به نام قضیه حد مرکزی در زیر بیان می‌کنیم.

قضیه ۱.۶ (قضیه حد مرکزی) فرض کنید $X_1, X_2, \ldots$ دنباله‌ای از متغیرهای تصادفی هم توزیع با میانگین μ و واریانس σ^2 متناهی باشند به قسمی که هر تعداد متناهی از این متغیرهای تصادفی از یکدیگر مستقل باشند. در این صورت توزیع حدی $Z=\frac{\bar X-\mu}{\sigma/\sqrt{n}}$ موقعی که $n\to\infty$ توزیع نرمال استاندارد است.

دارای توزیع نرمال استاندارد

به عبارت دیگر قضیه حد مرکزی بیانگر این است که وقتی اندازه نمونه n افزایش یابد توزیع میانگین نمونه $\bar X$ یک نمونه تصادفی که از هر جمعیتی گرفته شده باشد، توزیع نرمال با میانگین μ و واریانس $\frac{\sigma^2}{n}$ است.

تقریب نرمال برای توزیع نمونه‌ای $\bar X$ معمولاً موقعی که $n\geq 30$ باشد یک تقریب مناسب است و در صورتی که نمونه تصادفی از جمعیت نرمال انتخاب شده باشد آنگاه طبق قسمت (الف) برای هر مقدار n، $\bar X$ دارای توزیع نرمال است.

مثال ۱.۲.۶ یک شرکت تولیدی لاستیک اتومبیل، لاستیک‌هایی تولید می‌کند که طول عمر این لاستیک‌ها دارای توزیع نرمال با میانگین ۲۴ ماه و انحراف معیار ۲ ماه است. احتمال اینکه در یک نمونه تصادفی ۲۵تایی از لاستیک‌ها میانگین طول عمر کمتر از ۲۵ ماه باشد را بیابید.

حل چون جمعیت نرمال است پس $\bar X\sim N(24, \frac{2^2}{25})$ و در نتیجه

جامعه نرمال

$\mu=24$, $\sigma=2$, $n=25$

توزیع $\bar x$ دارای توزیع نرمال است

$p(\bar x<25)=P\left(\frac{\bar x-\mu}{\sigma/\sqrt n}<\frac{25-\mu}{\sigma/\sqrt n}\right)=P\left(Z<\frac{25-24}{2/\sqrt{25}}\right)=P(Z<\frac{5}{2})=P(Z<2.5)=\ldots$

$$P(\bar X<25)=P\left(\frac{\bar X-\mu}{\sigma/\sqrt n}<\frac{25-24}{\frac{2}{5}}\right)=P(Z<2.5)=0.9938$$

مثال ۲.۲.۶ فرض کنید مقدار سالهای تحصیل در بین افراد بالغ در کشوری معین دارای میانگین ۱۱/۱ سال و انحراف معیار ۳ سال باشد. احتمال اینکه در یک نمونه تصادفی ۱۰۰ نفری از افراد بالغ متوسط تعداد سالهای تحصیل بین ۱۱ تا ۱۲ سال باشد را بیابید.

حل چون جمعیت نرمال نیست ولی $n>30$ بنابراین تقریباً $\bar X\sim N(11.1, \frac{3^2}{100})$ و در نتیجه

$$P(11<\bar X<12)=P\left(\frac{11-11.1}{\frac{3}{10}}<\frac{\bar X-\mu}{\sigma/\sqrt n}<\frac{12-11.1}{\frac{3}{10}}\right)$$

$$\approx P(-0.33<Z<3)=P(Z<3)-P(Z\leq -0.33)$$

$$=0.9987-0.3707=0.628$$

$\mu=11.1$, $\sigma=3$, $n=100$

$P(11<\bar x<12)=P(-0.33<Z<3)=p(Z<3)-p(Z<-0.33)$

تبدیل به توزیع نرمال استاندارد

به طور توزیع نرمال

۳.۶ توزیع نمونه‌ای واریانس نمونه S^2

یک معیار پراکندگی معقول واریانس نمونه $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar X)^2$ می‌باشد که توزیع نمونه‌ای آن به توزیعی به نام توزیع مربع-کای[1] مربوط می‌شود که در زیر این توزیع را معرفی می‌کنیم.

توزیع مربع-کای

اگر $Z\sim N(0,1)$ آنگاه توزیع $Y=Z^2$ را یک توزیع مربع-کای با یک درجه آزادی گویند. همچنین اگر $Z_1, Z_2, \ldots$ و Z_n یک نمونه تصادفی nتایی از جمعیت نرمال استاندارد باشند (یعنی Z_iها از یکدیگر مستقل باشند و $Z_i\sim N(0,1)$) آنگاه توزیع $Y=Z_1^2+Z_2^2+\ldots+Z_n^2$ را یک توزیع مربع-کای با n درجه آزادی گویند و آن را با نماد $Y\sim\chi^2_{(n)}$ نمایش می‌دهند.

تابع چگالی احتمال یک متغیر تصادفی مربع-کای با n درجه آزادی عبارت است از

$$\boxed{f_Y(y)=\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}y^{\frac{n}{2}-1}e^{-\frac{y}{2}} \qquad y>0} \qquad (۲۶)$$

1- Chi-Square distribution

که در آن $\Gamma(.)$ تابع گاما می‌باشد که به صورت زیر تعریف می‌شود.

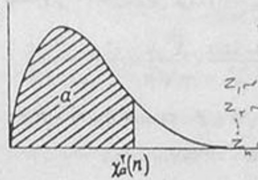
$$\Gamma(\beta) = \int_{۰}^{\infty} x^{\beta-۱} e^{-x}\,dx \qquad \beta > ۰$$

و همواره $\Gamma(\beta) = (\beta-۱)\Gamma(\beta-۱)$ و $\Gamma(\frac{۱}{۲}) = \sqrt{\pi}$ و اگر β عدد صحیحی باشد آنگاه $\Gamma(\beta) = (\beta-۱)!$.

$\Gamma(1)=1$

نمودار تابع چگالی احتمال توزیع مربع-کای در شکل ۱.۶ رسم شده است.

انحراف معیار و واریانس توزیع $\chi^۲$ دارند.



$Z \sim N(0,1)$
$Z^2 \sim \chi^2_{(1)}$
$\left.\begin{array}{l} Z_1 \sim N(0,1) \\ Z_2 \sim N(0,1) \\ \vdots \\ Z_n \sim N(0,1) \end{array}\right\} Z_1^2 + Z_2^2 + \dots + Z_n^2 \sim \chi^2_{(n)}$

شکل ۱.۶ نمودار تابع چگالی احتمال توزیع $\chi^۲_{(n)}$

توزیع مربع-کای یکی از توزیعهای مهم در توزیعهای نمونه‌ای است و کاربرد اصلی آن در مبحث استنباط آماری می‌باشد. برای توزیع مربع-کای جدول (IV) در ضمیمه ارایه گردیده است که در این جدول مقادیر $\chi^۲_\alpha(n)$ به ازای مقادیر مختلف α و n محاسبه شده است که در آن $\chi^۲_\alpha(n)$ نقطه‌ای روی محور افقی تابع چگالی احتمال توزیع $\chi^۲_{(n)}$ است که مساحت زیر منحنی قبل از این نقطه برابر α است. یعنی

$$Y \sim \chi^۲_{(n)} \Rightarrow P(Y \le \chi^۲_\alpha(n)) = \alpha$$

برای مثال از جدول (IV) داریم که $\chi^۲_{۰/۹۵}(۱۵) = ۲۵/۰$ و $\chi^۲_{۰/۸۰}(۲۵) = ۳۰/۷$.

مثال ۱.۳.۶ اگر $X \sim \chi^۲_{(۱۸)}$ مطلوب است

الف- $P(X \le ۷/۰۱)$ را محاسبه کنید.

ب- اگر $P(X > x) = ۰/۰۵$ مقدار x را به دست آورید.

حل الف- چون $\chi^۲_{۰/۰۱}(۱۸) = ۷/۰۱$ بنابراین $P(X \le ۷/۰۱) = ۰/۰۱$

ب-

$$P(X > x) = ۰/۰۵ \Rightarrow P(X \le x) = ۰/۹۵ \Rightarrow x = \chi^۲_{۰/۹۵}(۱۸) = ۲۸/۹$$

در قضیه زیر توزیع نمونه‌ای $S^۲$ را به دست می‌آوریم.

قضیه ۲.۶ اگر $X_۱, X_۲, \ldots, X_n$ یک نمونه تصادفی از جمعیت نرمال با میانگین μ و واریانس $\sigma^۲$ باشند آنگاه

الف- $$\frac{(n-۱)S^۲}{\sigma^۲} = \frac{\sum_{i=۱}^{n}(X_i - \overline{X})}{\sigma^۲} \sim \chi^۲_{(n-۱)}$$

ب- $\overline{X}$ و $S^۲$ از یکدیگر مستقل هستند.

مثال ۲.۳.۶ یک جمعیت نرمال دارای واریانس ۶ است. اگر نمونه تصادفی ۲۵ تایی از این جمعیت انتخاب شود احتمال اینکه واریانس نمونه بین ۳/۴۵ و ۱۰/۷۵ باشد را بیابید.

حل

$$P(۳/۴۵ < S^۲ < ۱۰/۷۵) = P\left(\frac{۲۴ \times ۳/۴۵}{۶} < \frac{(n-۱)S^۲}{\sigma^۲} < \frac{۲۴ \times ۱۰/۷۵}{۶}\right)$$
$$= P(۱۳/۸ < \chi^۲_{(۲۴)} < ۴۳)$$
$$= P(\chi^۲_{(۲۴)} < ۴۳) - P(\chi^۲_{(۲۴)} \le ۱۳/۸) = ۰/۹۹ - ۰/۰۵ = ۰/۹۴$$

واریانس نمونه‌ای

تبدیل کردیم به استاندارد $(n-1)\times$

۴.۶ توزیع نمونه‌ای $\frac{\overline{X} - \mu}{S/\sqrt{n}}$

در بخش ۲.۶ مشاهده کردیم که اگر $X_۱, X_۲, \ldots, X_n$ یک نمونه تصادفی از جمعیت نرمال با میانگین μ و واریانس $\sigma^۲$ باشند آنگاه $Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ دارای توزیع نرمال استاندارد است. در بعضی موارد مقدار $\sigma^۲$ نامعلوم است و به جای آن می‌توان از واریانس نمونه $S^۲$ به عنوان یک برآورد یا تخمین استفاده کرد. حال اگر در Z به جای σ مقدار S را قرار دهیم، توزیع نمونه‌ای $T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$ به چه صورت خواهد بود. در این بخش نشان می‌دهیم که توزیع نمونه‌ای T به توزیعی به نام توزیع t مربوط می‌شود که آن را در زیر معرفی می‌کنیم.

توزیع t[1]

فرض کنید $Z \sim N(۰,۱)$ و $Y \sim \chi^۲_{(n)}$ و X و Y از یکدیگر مستقل باشند. در این صورت توزیع متغیر تصادفی $T = \frac{Z}{\sqrt{\frac{Y}{n}}}$ را یک توزیع t با n درجه آزادی گویند و آن را با نماد $t_{(n)}$

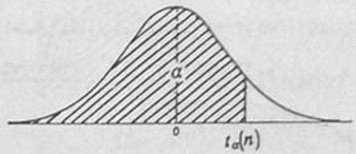
1- t-distribution

$T\sim t_{(n)}$ نمایش می‌دهند.

تابع چگالی احتمال یک متغیر تصادفی t با n درجه آزادی عبارت است از

$$f_T(t)=\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}} \qquad -\infty<t<+\infty \qquad (۳.۶)$$

نمودار تابع چگالی احتمال توزیع t در شکل ۲.۶ رسم شده است.



شکل ۲.۶ نمودار تابع چگالی احتمال توزیع $t_{(n)}$

برای توزیع t جدول (V) در ضمیمه ارایه گردیده است که در این جدول مقادیر $t_\alpha(n)$ به ازای مقادیر مختلف α و n محاسبه شده است که در آن $t_\alpha(n)$ نقطه‌ای روی محور افقی تابع چگالی احتمال توزیع $t_{(n)}$ است که مساحت زیر منحنی قبل از این نقطه برابر α است. یعنی

$$T\sim t_{(n)} \Rightarrow P(T<t_\alpha(n))=\alpha$$

برای مثال از جدول (V) داریم که $t_{0/70}(10)=0/542$ و $t_{0/95}(15)=1/75$

مثال ۱.۴.۶ اگر $T\sim t_{(16)}$ مطلوب است

الف- $P(T>1/34)$ را محاسبه کنید.

ب- اگر $P(T<t)=0/80$ باشد مقدار t را محاسبه کنید.

حل الف-
$$P(T>1/34)=1-P(T<1/34)=1-0/90=0/10$$
ب-
$$P(T<t)=0/80 \Rightarrow t=t_{0/80}(16)=0/865$$

در قضیه زیر توزیع نمونه‌ای $T=\frac{\bar{X}-\mu}{S/\sqrt{n}}$ را به دست می‌آوریم.

قضیه ۳.۶ اگر $\bar{X}$ و S^2 به ترتیب میانگین و واریانس یک نمونه تصادفی به اندازه n از یک

جامعه غیرنرمال + σ معلوم ← $\bar{X}$ دارای توزیع نرمال و نرمال استاندارد

جمعیت نرمال با میانگین μ و واریانس σ^2 باشند آنگاه

$$\boxed{T=\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t_{(n-1)}} \qquad (۴.۶)$$

دارای توزیع t با درجه آزادی (n-1)

اثبات با توجه به فرمول (۱.۶) و قضیه ۲.۶ داریم که

$$\boxed{Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)} \quad , \quad Y=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2_{(n-1)}$$

و Z و Y از یکدیگر مستقل هستند. پس $T=\frac{Z}{\sqrt{\frac{Y}{n-1}}}\sim t_{(n-1)}$. از طرفی

نرمال استاندارد

$$t_{(n)}=T=\frac{Z}{\sqrt{\frac{Y}{n-1}}}=\frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}}=\frac{\bar{X}-\mu}{S/\sqrt{n}}$$

و بنابراین
$$T=\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t_{(n-1)}$$

توجه کنید که اگر $n\to\infty$ آنگاه می‌توان نشان داد که توزیع حدی $T=\frac{\bar{X}-\mu}{S/\sqrt{n}}$ یک توزیع نرمال استاندارد است. بنابراین برای $n\geq 30$ توزیع T تقریباً نرمال استاندارد است و به همین علت در جدول (V) مقادیر درجه آزادی بزرگتر از ۳۰ با ∞ نشان داده شده است و مقادیر این ردیف از جدول، با جدول توزیع نرمال یکی است.

مثال ۲.۴.۶ نمرات یک کلاس از دانشجویان دارای توزیع نرمال با میانگین ۱۵ است. اگر از این کلاس یک نمونه ۲۰تایی انتخاب کنیم و مشاهده کنیم که انحراف استاندارد نمرات آنها ۴/۲۸ است، احتمال اینکه میانگین نمرات این افراد از ۱۷ بیشتر باشد را بیابید.

σ نامعلوم
درجه آزادی ۱۹ = (n-1)

حل در اینجا $\mu=15$، $S=4/28$ و $n=20$ است بنابراین

$$P(\bar{X}>17)=P\left(\frac{\bar{X}-\mu}{S/\sqrt{n}}>\frac{17-15}{\frac{4/28}{\sqrt{20}}}\right)=P(t_{(19)}>2/09)$$

$$= 1 - P(t_{(۱۹)} \le ۲/۰۹) = 1 - ۰/۹۷۵ = ۰/۰۲۵$$

۵.۶ توزیع نمونه‌ای اختلاف میانگین‌ها $\bar{X}_1 - \bar{X}_2$:

فرض کنید که دو جمعیت داشته باشیم که جمعیت اول دارای میانگین μ_1 و واریانس σ_1^2 و جمعیت دوم دارای میانگین μ_2 و واریانس σ_2^2 باشد. یک نمونه تصادفی n_1 تایی از جمعیت اول انتخاب کرده و میانگین نمونه آن را با $\bar{X}_1$ و واریانس نمونه آن را با S_1^2 نمایش می‌دهیم و یک نمونه تصادفی n_2 تایی از جمعیت دوم انتخاب کرده و میانگین نمونه آن را با $\bar{X}_2$ و واریانس نمونه آن را با S_2^2 نمایش می‌دهیم. همچنین فرض کنید که نمونه‌گیری از دو جمعیت مستقل از یکدیگر باشند. در این بخش می‌خواهیم توزیع نمونه‌ای اختلاف میانگین‌ها یعنی $\bar{X}_1 - \bar{X}_2$ را به دست آوریم. برای این منظور ۲ حالت زیر را در نظر می‌گیریم.

حالت اول: واریانس دو جمعیت σ_1^2 و σ_2^2 معلوم باشند

$\bar{X} \to t, z$

در این حالت بستگی به نرمال بودن جمعیتها دو حالت زیر پیش می‌آید

الف - اگر دو جمعیت نرمال باشند آنگاه طبق قضیه ۱۴.۵ داریم که

$\frac{(n-1)S^2}{\sigma^2} \to \chi^2$

$$\bar{X}_1 \sim N(\mu_1, \frac{\sigma_1^2}{n_1}) \quad , \quad \bar{X}_2 \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$$

و $\bar{X}_1$ و $\bar{X}_2$ از یکدیگر مستقل هستند. بنابراین $\bar{X}_1 - \bar{X}_2$، که یک ترکیب خطی از متغیرهای تصادفی نرمال است، نیز دارای توزیع نرمال است یعنی

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

هر دو جامعه نرمال

هر دو جامعه غیرنرمال

و یا

$$\boxed{Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)} \quad (۵.۶)$$

دارای توزیع نرمال استاندارد

ب - اگر دو جمعیت نرمال نباشند آنگاه طبق قضیه حد مرکزی برای اندازه نمونه $n_1 \ge ۳۰$ و $n_2 \ge ۳۰$ هر کدام از $\bar{X}_1$ و $\bar{X}_2$ دارای توزیع تقریبی نرمال بوده و چون از یکدیگر مستقل هستند بنابراین $\bar{X}_1 - \bar{X}_2$ نیز تقریباً دارای توزیع نرمال است و نتیجه قسمت (الف) نیز در این حالت

برقرار می‌باشد.

مثال ۱.۵.۶ دو کارخانه تولید کابل A و B وجود دارند. کابل‌هایی که توسط کارخانه A تولید می‌شود، به طور متوسط تحمل ۴۰۰۰ پوند نیروی کششی و انحراف معیار ۳۰۰ پوند را دارند و کابل‌هایی که توسط کارخانه B تولید می‌شود، به طور متوسط تحمل ۴۵۰۰ پوند نیرو با انحراف معیار ۲۰۰ پوند را دارند. اگر ۱۰۰ کابل نوع A و ۵۰ کابل نوع B مورد آزمایش قرار گیرند، احتمال اینکه متوسط تحمل نیروی کششی B حداقل ۶۰۰ پوند بیش از نیروی کششی A باشد را بیابید.

حل در اینجا چون $n_A = ۱۰۰$ و $n_B = ۵۰$ پس تقریباً

$$\bar{X}_A - \bar{X}_B \sim N(۴۰۰۰ - ۴۵۰۰, \frac{۳۰۰^2}{۱۰۰} + \frac{۲۰۰^2}{۵۰})$$

و یا تقریباً $\bar{X}_A - \bar{X}_B \sim N(-۵۰۰, ۱۷۰۰)$ بنابراین

$$P(\bar{X}_B \ge \bar{X}_A + ۶۰۰) = P(\underbrace{\bar{X}_A - \bar{X}_B}_{\text{دارای توزیع نرمال}} \le -۶۰۰)$$

$$= P(\frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} \le \frac{-۶۰۰ + ۵۰۰}{\sqrt{۱۷۰۰}})$$

$$\approx P(Z \le -۲/۴۳) = ۰/۰۰۷۵$$

حالت دوم: واریانس دو جمعیت نامعلوم اما مساوی باشند

در این حالت $\sigma_1^2 = \sigma_2^2 = \sigma^2$ می‌باشد و σ^2 که واریانس مشترک دو جمعیت می‌باشد نامعلوم است. در جمعیت اول می‌توان از S_1^2 و در جمعیت دوم می‌توان از S_2^2 به عنوان یک برآورد یا تخمین برای σ^2 استفاده کرد. اما بهتر است که از یک میانگین وزنی این دو کمیت برای برآورد σ^2 استفاده کنیم. بنابراین از واریانس مشترک دو نمونه یعنی

$$\boxed{S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}} \quad (۶.۶)$$

به عنوان یک برآورد یا تخمین σ^2 استفاده می‌کنیم. حال اگر دو جمعیت نرمال باشند آنگاه با استفاده از قضیه ۲.۶ داریم که

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

و بنابراین با استفاده از تعریف توزیع t به راحتی می‌توان نشان داد که

$$\boxed{T=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{S_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t_{(n_1+n_2-2)}} \qquad (۷.۶)$$

۶.۶ توزیع نمونه‌ای نسبت واریانسهای نمونه

دو جمعیت مفروض در بخش ۵.۶ را در نظر بگیرید. در این بخش می‌خواهیم توزیع نمونه‌ای نسبت واریانسهای نمونه یعنی $\frac{S_1^2}{S_2^2}$ را به دست آوریم. توزیع نمونه‌ای این نسبت به توزیعی بنام توزیع F مربوط می‌شود که آن را در زیر معرفی می‌کنیم.

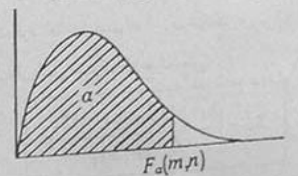
توزیع F[1]

اگر $U\sim\chi^2_{(m)}$ و $V\sim\chi^2_{(n)}$ و متغیرهای تصادفی U و V از یکدیگر مستقل باشند آنگاه توزیع متغیر تصادفی $F=\frac{U/m}{V/n}$ را یک توزیع F با m و n درجه آزادی گویند و آن را با نماد $F\sim F_{(m,n)}$ نمایش می‌دهند.

تابع چگالی احتمال یک توزیع F با درجات آزادی m و n عبارت است از

$$\boxed{f_F(x)=\frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}\left(\frac{m}{n}\right)^{\frac{m}{2}}\frac{x^{\frac{m}{2}-1}}{(1+\frac{m}{n}x)^{\frac{m+n}{2}}} \qquad x>0} \qquad (۸.۶)$$

نمودار تابع چگالی احتمال توزیع F در شکل ۳.۶ رسم شده است.



شکل ۳.۶ نمودار تابع چگالی احتمال توزیع $F_{(m,n)}$

1- F-distribution

برای توزیع F جدول (VI) در ضمیمه ارایه گردیده است که در این جدول $F_\alpha(m,n)$ به ازای مقادیر مختلف m و n و $\alpha=0/90, 0/95, 0/99$ محاسبه شده است. که در آن $F_\alpha(m,n)$ نقطه‌ای روی محور افقی تابع چگالی احتمال توزیع $F_{(m,n)}$ است که مساحت زیر منحنی قبل از این نقطه برابر α است (برای مثال از جدول (VI) داریم که $F_{./90}(6,14)=2/24$ و $F_{./95}(20,10)=2/77$).

مثال ۱.۶.۶ اگر $F\sim F_{(10,12)}$ مطلوب است

الف- $P(F<2/75)$ را محاسبه کنید.

ب- اگر $P(F>x)=0/01$ باشد مقدار x را محاسبه کنید.

حل الف- چون $F_{./95}(10,12)=2/75$ پس $P(F<2/75)=0/95$

ب- $P(F>x)=0/01 \Rightarrow P(F\le x)=0/99 \Rightarrow x=F_{./99}(10,12)=4/30$

در قضیه زیر توزیع نمونه‌ای $\frac{S_1^2}{S_2^2}$ را به دست می‌آوریم.

قضیه ۴.۶ اگر S_1^2 و S_2^2 به ترتیب واریانسهای نمونه‌های تصادفی مستقل به اندازه‌های n_1 و n_2 از جمعیتهای نرمال با واریانسهای σ_1^2 و σ_2^2 باشند آنگاه

$$\boxed{F=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F_{(n_1-1,n_2-1)}} \qquad (۹.۶)$$

اثبات از قضیه ۲.۶ نتیجه می‌شود که

$$U=\frac{(n_1-1)S_1^2}{\sigma_1^2}\sim\chi^2_{(n_1-1)} \quad , \quad V=\frac{(n_2-1)S_2^2}{\sigma_2^2}\sim\chi^2_{(n_2-1)}$$

و U و V از یکدیگر مستقل هستند. بنابراین از تعریف توزیع F داریم که

$$F=\frac{U/(n_1-1)}{V/(n_2-1)}\sim F_{(n_1-1,n_2-1)}$$

از طرفی

$$F=\frac{(n_1-1)S_1^2/[\sigma_1^2(n_1-1)]}{(n_2-1)S_2^2/[\sigma_2^2(n_2-1)]}=\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

بنابراین نتیجه قضیه حاصل می‌شود.

مثال ۲.۶.۶ اگر S_1^2 و S_2^2 نمایانگر واریانسهای نمونه‌های تصادفی مستقل به اندازه‌های ۲۵ و

۳۱ تایی به ترتیب از جمعیت‌های نرمال با واریانسهای ۱۰ و ۱۵ باشند، احتمال اینکه S_1^2 حداقل ۱/۲۶ برابر S_2^2 باشد چقدر است؟

حل

$$P\left(\frac{S_1^2}{S_2^2}\geq 1/26\right)=P\left(\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\geq 1/26\left(\frac{15}{10}\right)\right)=P(F_{(22,30)}\geq 1/89)$$

$$= 1-P(F_{(22,30)}< 1/89) = 1-0/95=0/05$$

۷.۶ تمرینات

۱ جعبه‌ای حاوی ۵ مهره است که از ۱ تا ۵ شماره گذاری شده‌اند. از این جعبه نمونه‌ای مرکب از ۲ مهره به تصادف و با جایگذاری استخراج می‌شود.

الف- با تشکیل جدول توزیع احتمال توأم X_1 و X_2، توزیع $\overline{X}=\frac{X_1+X_2}{2}$ را محاسبه کنید.

ب- میانگین و واریانس $\overline{X}$ را از روی توزیع آن محاسبه کنید.

ج- میانگین و واریانس جمعیت را پیدا کنید.

د- نمودار توزیع جمعیت (برای یک مشاهده X) و نمودار توزیع نمونه‌ای $\overline{X}$ را رسم کنید.

۲ جعبه‌ای پر از مهره‌های هم شکل است که یک سوم آنها با شماره ۲ و یک سوم با شماره ۴ و یک سوم با شماره ۶ مشخص شده‌اند.

الف- یک مهره استخراج می‌گردد اگر شماره آن X باشد، میانگین μ و انحراف معیار σ جمعیت را بیابید.

ب- نمونه‌ای مرکب از دو مهره با جایگذاری استخراج می‌شود. اگر $\overline{X}$ میانگین نمونه باشد، جدول توزیع احتمال $\overline{X}$ را به دست آورید و از روی آن میانگین و واریانس $\overline{X}$ را حساب کنید.

ج- قسمت (ب) را برای نمونه‌ای متشکل از ۳ مهره حل کنید.

۳ اندازه‌گیری قطر گام نوعی پیچ دارای توزیع نرمال با میانگین ۰/۴۰۰۸ اینچ و انحراف معیار ۰/۰۰۰۳ اینچ است. حدود مشخصات طراحی به صورت ۰/۰۰۱±۰/۴ اینچ ارایه شده است. نمونه

های ۴ تایی از قطرهای گام گرفته می‌شود. احتمال اینکه میانگین نمونه در داخل حدود مشخصات قرار گیرد را بیابید. احتمال تجاوز میانگین از ۰/۴۰۰۵ را بیابید.

۴ عرض یک شکاف که بر یک قطعه از آلیاژ آلومینیوم که با ریخته‌گری تولید می‌شود، توزیع نرمال با میانگین ۰/۹ اینچ و انحراف معیار ۰/۰۲ اینچ دارد. حدود مشخصات طراحی عبارت از ۰/۰۰۵±۰/۹ اینچ است. هر ساعت نمونه‌هایی ۵تایی از آلیاژ ریخته‌گری گرفته شده و میانگین آن محاسبه می‌شود. چند درصد از این میانگین‌های نمونه در خارج از حدود مشخصات طراحی قرار می‌گیرند؟ حدود را طوری تعیین کنید که درصد میانگین‌های نمونه که خارج از حدود قرار می‌گیرند معادل ۰/۲۷ درصد باشد.

۵ مقدار پولی که مردم شهر معینی در کیف خود دارند بطور متوسط ۹۰۰۰ تومان با انحراف معیار ۲۵۰۰ تومان است. احتمال آنکه گروهی مرکب از ۲۲۵ نفر به طور متوسط بیش از ۹۱۰۰ تومان همراه داشته باشند را بیابید.

۶ فرض کنید طول دانه‌های زنجیر دوچرخه‌ای به طور متوسط ۰/۵۰ اینچ با انحراف معیار ۰/۰۴ اینچ باشد. استانداردهای سازنده دوچرخه اقتضا می‌کند که زنجیر بین ۴۹ و ۵۰ اینچ طول داشته باشد. اگر زنجیرها از ۱۰۰ دانه درست شده باشند، چه نسبتی از آنها مطابق استاندارد می‌باشند؟

۷ اگر اندازه نمونه‌ای ۳۶ و خطای استاندارد آن ۲ باشد، اندازه نمونه چقدر باید باشد تا اینکه خطای استاندارد آن کاهش پیدا نموده و به ۱/۲ برسد؟

۸ در فاصله (۰,۱) تعداد ۷۵ عدد را به تصادف انتخاب می‌کنیم. نشان دهید احتمال اینکه معدل این عددها در فاصله (۰/۵۵,۰/۴۵) قرار گیرد تقریباً برابر ۰/۸۶۶ است.

۹ تعداد مکالمات تلفنی که از طریق مرکزی انجام می‌شود به طور متوسط ۴ مکالمه در دقیقه است. احتمال تقریبی اینکه طی یک ساعت آینده حداقل ۲۵۰ مکالمه از طریق این مرکز انجام شود را بیابید.

۱۰ وزن یک گلوله فلزی دارای میانگین ۵/۰۲ اونس و انحراف معیار ۰/۳۰ اونس است. احتمال اینکه مجموع وزن یک نمونه تصادفی ۱۰۰تایی انتخاب شده از این نوع گلوله‌های فلزی (الف) بین ۴۹۶ تا ۵۰۰ اونس (ب) بیش از ۵۱۰ اونس باشد را بیابید.

۱۱ پژوهشگری می‌خواهد میانگین جمعیتی را با استفاده از نمونه‌ای برآورد کند. وی مایل است