فصل نهم

توزیع‌های برخی از آماره‌ها

در فصل قبل با آماره‌ها و تعاریف و مقدمات آنها آشنا شدید، نشان دادیم که آماره‌ها خود متغیر می‌باشند و می‌توان تابع توزیع آنها را معنی نمود، حال در این فصل چندین آماره خاص را مورد بررسی قرار می‌دهیم و میانگین و واریانس آنها را محاسبه می‌کنیم.

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

از یک جامعه با میانگین μ و واریانس $\delta^۲$ تعداد n نمونه $X_۱, X_۲, \ldots, X_n$ انتخاب می‌کنیم، آماره $\overline{X} = \frac{۱}{n}\sum_{i=۱}^{n} X_i$ را در نظر بگیرید می‌خواهیم میانگین و واریانس $\overline{X}$ را به عنوان یک متغیر تصادفی بدست بیاوریم. طبق قضیه حد مرکزی اگر مقدار n به اندازه کافی بزرگ اختیلر شود متغیر تصادفی $Z = \frac{\overline{X}-\mu}{\frac{\delta}{\sqrt{n}}}$ یک متغیر تصادفی نرمال استاندارد می‌باشد. از طرفی داریم:

$$\overline{X} = \frac{\delta}{\sqrt{n}} Z + \mu \quad , \quad E[Z] = ۰ \quad , \quad \delta_Z^۲ = ۱$$

بنابراین:

$$E[\overline{X}] = \frac{\delta}{\sqrt{n}} E[Z] + \mu = \frac{\delta}{\sqrt{n}} \times ۰ + \mu = \mu$$

بنابراین آماره $\overline{X}$ به عنوان یک متغیر تصادفی اگر تعداد نمونه‌ها زیاد باشد $(n > ۳۰)$ همواره دارای میانگین μ یا همان میانگین جامعه می‌باشد.

۹-۳

مثال ۱: در نقطه‌ای از یک رودخانه در هر دقیقه بطور متوسط تعداد ۲۰ ماهی عبور می‌کند بدیهی است که ماهیگیر تنها در صورتی می‌تواند در هر دقیقه این تعداد ماهی را صید کند که تقریباً تمامی طول رودخانه را با تور ماهیگیری پوشش دهد یا به عبارتی در هر دقیقه از بیشتر نقاط در طول رودخانه نمونه برداری کند که معادل است با انتخاب تعداد زیادی نمونه.

حال واریانس $\overline{X}$ را محاسبه می‌کنیم:

$$\operatorname{var}(\overline{X}) = \operatorname{var}\left(\frac{\delta}{\sqrt{n}} Z + \mu\right) = \operatorname{var}\left(\frac{\delta}{\sqrt{n}} Z\right) = \frac{\delta^۲}{\sqrt{n}} \operatorname{var}(Z) = \frac{\delta^۲}{\sqrt{n}}$$

بنابراین واریانس $\overline{X}$ برابر با $\frac{\delta^۲}{\sqrt{n}}$ می‌باشد.

۹-۴

مثال ۲: محصول شیر تولیدی یک کارخانه بطور متوسط ۲۰ روز پس از تاریخ تولید قابل مصرف می‌باشد که با واریانس ۴ روز رخ می‌دهد. می‌خواهیم از این کارخانه ۱۰ پاکت شیر تهیه کنیم. احتمال اینکه این پاکتهای شیر بطور متوسط حداقل ۱۸ روز قابل مصرف باشند چقدر است؟

حل: برای به دست آوردن احتمال می‌بایستی $p(\overline{X} \geq ۱۸)$ را محاسبه کنیم که در واقع از تابع توزیع متغیر تصادفی $\overline{X}$ که در اینجا نرمال μ $\frac{\delta^۲}{\sqrt{n}}$ در نظر گرفته می‌شود استفاده می‌کنیم. بنابراین:

$$\mu_{\overline{X}} = ۲۰ \quad , \quad n = ۱۰$$

$$\delta_{\overline{X}}^۲ = \frac{\delta^۲}{n} = \frac{(۴)^۲}{۱۰} = \frac{۱۶}{۱۰} \quad \rightarrow \quad \frac{\delta}{\overline{X}} = \sqrt{\frac{۱۶}{۱۰}} = ۱/۲۶$$

$$p(\overline{X} \geq ۱۸) = p\left(\frac{\overline{X}-\mu_{\overline{X}}}{\delta_{\overline{X}}} \geq \frac{۱۸-\mu_{\overline{X}}}{\delta_{\overline{X}}}\right)$$

$$= p\left(\frac{\overline{X}-\mu_{\overline{X}}}{\delta_{\overline{X}}} \geq \frac{۱۸-۲۰}{۱/۲۶}\right) = p\left(\frac{\overline{X}-\mu_{\overline{X}}}{\delta_{\overline{X}}} \geq ۱/۵۸\right)$$

$$p=(Z \geq -۱/۵۸) = ۱ - p(Z \leq -۱/۵۸)$$

$$= ۱ - N_Z(-۱/۵۸) = ۱ - ۰/۰۵۷۱ = ۰/۹۴۲۹$$

۹-۵-۱

۹-۲ توزیع نمونه‌ای واریانس نمونه $S^۲$.

حال به محاسبه توزیع نمونه‌ای واریانس $S^۲$ و بدست آوردن میانگین و واریانس توزیع آن می‌پردازیم.

همانطور که از فصل‌های قبل به یاد دارید واریانس نمونه‌ها با $S^۲$ نشان داده می‌شود که بصورت زیر تعریف می‌شود:

$$S^۲ = \frac{۱}{n-۱} \sum_{i=۱}^{n} (X_i - \overline{X})^۲$$

قبل محاسبه توزیع نمونه‌ای $S^۲$ ابتدا میانگین و واریانس آنرا (از آنجا که محاسبه آن ساده می‌باشد) بدست می‌آوریم:

می‌دانیم واریانس $\overline{X}$ برابر $\frac{\delta^۲}{n}$ می‌باشد بنابراین:

$$var(\overline{X}) = \frac{\delta^۲}{n} \rightarrow E\left[(\overline{X}-\mu)^۲\right] = \frac{\delta^۲}{n}$$

$$E\left[\overline{X}^۲\right] - E^۲[X] = E\left[\overline{X}^۲\right] - \mu^۲ = \frac{\delta^۲}{n} \quad \Rightarrow \quad E\left[\overline{X}^۲\right] = \frac{\delta^۲}{n} + \mu^۲$$

حال داریم:

$$E\left[S^۲\right] = E\left[\frac{۱}{n-۱} \sum_{i=۱}^{n} (X_i - \overline{X})^۲\right] = \frac{۱}{n-۱} E\left[\sum_{i=۱}^{n} (X_i - \overline{X})^۲\right]$$

$$= \frac{۱}{n-۱} E\left[\sum_{۱}^{n} X_i^۲ - ۲\overline{X} \sum_{۱}^{n} X_i + \sum_{۱}^{n} \overline{X}^۲\right]$$

$$= \frac{n}{n-۱} E\left[\frac{۱}{n} \sum_{۱}^{n} X_i^۲ - ۲\overline{X} \frac{۱}{n} \sum_{۱}^{n} X_i + \frac{۱}{n} \sum_{۱}^{n} \overline{X}^۲\right]$$

$$= \frac{n}{n-۱} E\left[\frac{۱}{n} \sum_{۱}^{n} X_i^۲ - ۲\overline{X}^۲ + \overline{X}^۲\right]$$

$$= \frac{n}{n-۱} E\left[\frac{۱}{n} \sum_{۱}^{n} X_i^۲ - \overline{X}^۲\right] = \frac{n}{n-۱} \left(E\left[\frac{۱}{n} \sum_{۱}^{n} X_i^۲\right] - E\left[\overline{X}^۲\right]\right)$$

$$= \frac{n}{n-۱} \left(E[M_۲] - \left(\frac{\delta^۲}{n} + \mu\right)\right)$$

۹-۵-۲

که در آن M_2

گشتاور دوم نمونه‌ها می‌باشد و داریم: $E[M_2]=m_2$ که m_2 نیز گشتاور دوم متغیر تصادفی X می‌باشد. در نهایت داریم:

$$= \frac{1}{n-1}(m_2 - \mu^2 - \frac{\delta^2}{n})$$

اما: $m_2 = \delta^2 + \mu^2$

$$= \frac{n}{n-1}(\delta^2 + \mu^2 - \mu^2 - \frac{\delta^2}{n}) = \frac{n}{n-1}(\delta^2 - \frac{\delta^2}{n}) = \delta^2$$

به نتیجه جالبی رسیدیم و آن اینکه میانگین واریانس نمونه‌ها برابر با واریانس جامعه می‌باشد در واقع علت اینکه در محاسبه S^2 از $\frac{1}{n-1}$ استفاده می‌کنیم این است که در محاسبات فوق در نهایت جواب ساده شده و δ^2 به عنوان میانگین S^2 ($E[S^2] = \delta^2$) بدست می‌آید . در حالی که اگر S^2 را بصورت زیر تعریف می‌کردیم $S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ جواب نهایی $E[S^2]$ برابر با $\sigma^2 (\frac{n}{n-1})$ می‌شد که بلیل وجود n در جواب نتیجه مطلوبی نیست. حال به جای بدست آوردن توزیع S^2 توزیع $S^2 (\frac{n-1}{\delta^2})$ را بدست می‌آوریم زیرا در محاسبه مقدار میانگین بصورت زیر ساده‌تر می‌شود.

$$E\left[\frac{n-1}{\delta^2} - S^2\right] = \frac{n-1}{\delta^2} E[S^2] = \frac{n-1}{\delta^2}\,\delta^2 = n-1$$

برای بدست آوردن توزیع $S^2 (\frac{n-1}{\delta^2})$ ابتدا توزیع خی-دو یا مربع‌کای را معرفی می‌کنیم:

۹-۶

۹.۲.۱ توزیع خی دو

فرض کنید Z یک متغیر تصادفی نرمال استاندارد باشد $Z \sim N(0, 1)$ در این صورت متغیر تصادفی $W = Z^2$ را یک متغیر تصادفی χ^2 (خی دو) با یک درجه آزادی می‌نامیم و به صورت χ_1^2 نمایش می‌دهیم که تابع توزیع آن بصورت زیر بدست می‌آید:

قبلاً نشان دادیم که اگر $Y = X^2$ باشد تابع توزیع $F_Y(t)$ برابر است با:

$$F_Y(t) = F_X(\sqrt{t}) - F_X(-\sqrt{t}) \qquad t \geq 0$$
$$= 0 \qquad t < 0$$

حال داریم: $W = Z^2$

$$F_W(t) = F_Z(\sqrt{t}) - F_Z(-\sqrt{t})$$

اما: $F_Z(\sqrt{t}) = N_Z(t)$ و $N_Z(-a) = 1 - N_Z(a)$ بنابراین:

$$F_W(t) = N_Z(\sqrt{t}) - \left[1 - N_Z(\sqrt{t})\right] = 2N_Z(\sqrt{t}) - 1$$

بنابراین با داشتن جدول مقادیر توزیع نرمال استاندارد به راحتی می‌توان مقادیر توزیع χ_1^2 را به دست آورد. مقدار دقیق تابع چگالی χ_1^2 بصورت زیر می‌باشد:

$$f_X(x) = \frac{۱}{\sqrt{۲\pi}} x^{-\frac{۱}{۲}} e^{-\frac{x}{۲}} \qquad x > ۰$$

میانگین، واریانس و تابع مولد گشتاور آن عبارتند از:

$$E[W] = ۱ \qquad var(W) = ۲ \qquad m_W(t) = (\frac{۱}{۱-۲t})^{\frac{۱}{۲}}$$

۹-۷

حال فرض کنید $X_۱, X_۲, \ldots, X_n$ نمونه‌هایی تصادفی از یک متغیر تصادفی نرمال X با میانگین μ و واریانس $\delta^۲$ باشند در این صورت:

$$Y = \sum_{i=۱}^{n} \frac{(X_i - \mu)^۲}{\delta^۲}$$

یک متغیر تصادفی $\chi^۲$ با n درجه آزادی نامیده می‌شود و با $\chi_n^۲$ نمایش داده می‌شود. به بیانی دیگر اگر $Z_۱, Z_۲, \ldots, Z_n$ تعداد n متغیر تصادفی نرمال استاندارد $N(۰, ۱)$ باشند در این صورت $Y = Z_۱^۲, Z_۲^۲, \ldots, Z_n^۲$ یک متغیر تصادفی $\chi^۲$ با n درجه آزادی است. تابع چگالی متغیر تصادفی $\chi_n^۲$ بصورت زیر می‌باشد:

$$f_Y(y) = \frac{۱}{\Gamma(\frac{n}{۲})} (\frac{۱}{۲})^{\frac{n}{۲}} y^{\frac{n}{۲}-۱} e^{-\frac{y}{۲}} \qquad y > ۰$$

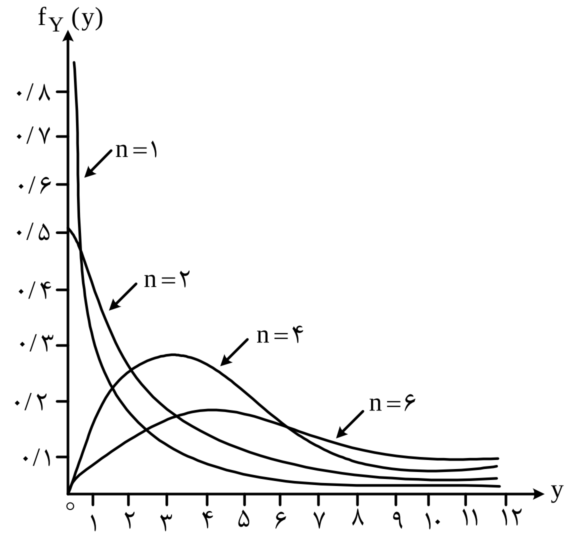
توجه کنید که متغیر تصادفی $\chi_n^۲$ حالت خاصی از متغیر تصادفی گاما می‌باشد که در آن $k = \frac{n}{۲}$ و $\lambda = \frac{۱}{۲}$ می‌باشد.

مقادیر میانگین، واریانس و تابع مولد گشتاور متغیر تصادفی $\chi_n^۲$ بصورت زیر می‌باشد:

$$E[Y] = n \quad , \quad var(Y) = ۲n$$

$$m_Y(t) = (\frac{۱}{۱-۲t})^{\frac{n}{۲}}$$

نمودار منحنی متغیر تصادفی $\chi_n^۲$ دارای نقطه ماکزیممی در $y = n-۲$ می‌باشد و با افزایش n نمودار منحنی هر چه بیشتر به نمودار نزدیک نزدیکتر می‌شود این مطلب در شکل زیر نشان داده شده است:



نکته: اگر $X_1, X_2, \ldots, X_n$ متغیرهای تصادفی خی دو به ترتیب با $r_1, r_2, \ldots, r_n$ درجه آزادی باشند و همچنین دو بدو مستقل از یکدیگر باشند در این صورت متغیر تصادفی $Y=\sum_1^n X_i$ دارای توزیع خی دو با $R=\sum_{i=1}^n r_i$ درجه آزادی می‌باشد.

۹-۸

مثال ۳: فرض کنید X_1, X_2, X_3 سه متغیر تصادفی مستقل خی دو به ترتیب با ۷ و ۳ و ۱ درجه آزادی باشند، اگر $W = X_1 + X_2 + X_3$ مطلوبست میانگین و واریانس متغیرهای تصادفی W؟

حل: متغیر تصادفی W یک متغیر تصادفی خی دو با ۱۱ = ۷+۳+۱ درجه آزادی می‌باشد بنابراین میانگین و واریانس آن عبارتست از:

$$E[W] = 11$$

$$var(W) = 2 \times 11 = 22$$

حال به قضیه بسیار مهم توجه کنید:

قضیه: اگر $X_1, X_2, \ldots, X_n$ یک نمونه تصادفی از متغیر تصادفی نرمال X با میانگین μ و واریانس δ^2 باشند آنگاه:

الف) $\overline{X}$ و S^2 از یکدیگر مستقل هستند.

ب) $\frac{(n-1)}{\delta^2} S^2$ یک متغیر تصادفی χ^2 با $n-1$ درجه آزادی می‌باشد توجه کنید که $(\frac{n-1}{\delta^2}) S^2$ برابر است با:

$$(\frac{n-1}{\delta^2}) S^2 = \frac{n-1}{\delta^2} \frac{1}{n-1} \sum_1^n (X_i - \overline{X})^2 = \sum_1^N \frac{(X_i - \overline{X})^2}{\delta^2}$$

و این با آنچه در تعریف متغیر تصادفی χ_n^2 متفاوت می‌باشد زیرا در تعریف از میانگین جامعه μ استفاده نمودیم: $\sum_1^n \frac{(X_i - \mu)^2}{\delta^2}$.

۹-۹

حال نشان می‌دهیم که چرا با جایگزین نمودن $\overline{X}$ به جای μ یک درجه آزادی از n کم می‌شود و درجه آزادی $\frac{(n-1)}{\delta^2} S^2$ برابر با $n-1$ می‌شود.

$$\sum_{i=1}^n (X_i - \mu)^2 = \sum_1^n \left[(X_i - \overline{X}) + (\overline{X} - \mu)\right]^2$$

$$= \sum_1^n (X_i - \overline{X})^2 + \sum_1^n (\overline{X} - \mu)^2 + 2(\overline{X} - \mu) \sum_1^n (X_i - \overline{X})$$

$$= \sum_1^n (X_i - \overline{X})^2 + n(\overline{X} - \mu)^2$$

در نتیجه بدست می‌آید:

$$\frac{\sum (X_i - \mu)^2}{\delta^2} = \frac{\sum (X_i - \mu)^2}{\delta^2} + \frac{(\overline{X} - \mu)^2}{\frac{\delta^2}{n}}$$

حال می‌دانیم که عبارت سمت چپ تساوی فوق یک متغیر تصادفی خی دو با n درجه آزادی است و عبارت $\frac{(\overline{X}-\mu)^۲}{\frac{\delta^۲}{n}}$ نیز یک متغیر تصادفی $X^۲$ با ۱ درجه آزادی است بنابراین جمله $\frac{\sum (X_i-\mu)^۲}{\delta^۲}$ می‌بایستی یک متغیر تصادفی $\chi^۲$ با $n-۱$ درجه آزادی باشد.

مثال ۴: از یک جامعه نرمال با میانگین μ و واریانس $\delta^۲$ یک نمونه تصادفی ۱۱تایی انتخاب می‌کنیم. متغیر تصادفی $Y = \frac{۱۰ S^۲}{\delta^۲}$ را در نظر بگیرید مطلوبست:

الف) محاسبه $p(۳/۹۴ < Y < ۱۸/۳)$.

ب) با استفاده از احتمال فوق رابطه بین $S^۲$ و $\delta^۲$ را در جامعه معین کنید.

حل: الف) ابتدا احتمال را بصورت زیر ساده می‌کنیم:

$$p(۳/۹۴ < Y < ۱۸/۳) = p(Y < ۱۸/۳) - p(Y < ۳/۹۴)$$

حال از روی جدول مقادیر احتمال برای متغیر تصادفی خی دو و با توجه به اینکه $Y = \frac{۱۰ S^۲}{\delta^۲}$ یک متغیر تصادفی خی دو با ۱۰ درجه آزادی است مقادیر احتمال بصورت زیر بدست می‌آیند:

$$p(Y < ۱۸/۳) - p(Y < ۳/۹۴) = ۰/۹۵ / ۰/۰۵ = ۰/۹$$

ب) می‌دانیم $p(۳/۹۴ < Y < ۱۸/۳) = ۰/۹$ بنابراین:

$$p(۳/۹۴ < Y = \frac{۱۰ S^۲}{\delta^۲} < ۱۸/۳) = p(\frac{۳/۹۴\ \delta^۲}{۱۰} < S^۲ < \frac{۱۸/۳\ \delta^۲}{۱۰})$$

$$\Rightarrow p(۰/۳۹۴\ \delta^۲ < S^۲ < ۱/۸۳\ \delta^۲) = ۰/۹$$

حال می‌بینیم که $S^۲$ در بازه‌ای نزدیک $\delta^۲$ قرار دارد یعنی اگر از یک جامعه نرمال با میانگین μ و واریانس $\delta^۲$ تعداد ۱۱ نمونه انتخاب کنیم به احتمال بسیار زیاد $(۰/۹)$ مقدار واریانس نمونه‌ها نزدیک به واریانس جامعه می‌باشد.

۹-۱۱-۱

۹ . ۳ تابع توزیع آمارۀ مرتب

قبلاً آمارۀ مرتب را معرفی نمودیم، می‌دانیم آماره‌های مرتب خود یک متغیر تصادفی می‌باشند. بنابراین می‌توانیم تابع توزیع آنها را بدست بیاوریم.

فرض کنید $X_۱, X_۲, \ldots, X_n$ تعداد n نمونه تصادفی از متغیر تصادفی X باشند در این صورت تابع توزیع برای آمارۀ مرتب $X_{(۱)}, X_{(۲)}, \ldots, X_{(n)}$ عبارتست از:

$$F_{X(j)}(t) = \sum_{k=j}^{n} \binom{n}{k} (F_X(t))^k (۱-F_X(t))^{n-k} \qquad t \in R$$

$F_X(t)$ تابع توزیع متغیر تصادفی X می‌باشد.

$F_{X(j)}(t)$ به صورت زیر بدست می‌آید:

اگر حداقل j متغیر تصادفی نمونه، کمتر یا مساوی با t باشند در این صورت $X_{(j)}$ یا jامین آمارۀ مرتب کمتر یا مساوی با t می‌باشد. نمام متغیرهای تصادفی نمونه مستقل می‌باشند و برای تمام متغیرهای تصادفی نمونه $F_X(t)$ برابر است با:

احتمال کمتر یا مساوی با t بودن. بنابراین تعداد متغیرهای تصادفی نمونه که کمتر یا مساوی با t می‌باشند یک متغیر تصادفی دو جمله‌ای با پارامترهای n و $p = F_X(t)$ می‌باشد، در نتیجه:

$$f_{X(j)}(t) = \sum_{k=j}^{n} \binom{n}{k} \left(F_X(t)\right)^{j-1}\left(1-F_X(t)\right)^{n-j} f_X(t)$$

با قرار دادن $j = 1$ و $j = n$ تابع توزیع و چگالی برای $\min(X_j)$ و $\mathrm{Max}(X_j)$ بصورت زیر بدست می‌آید:

$$\mathrm{Min}(X_j) \text{ تابع توزیع} : F_{X(1)}(t) = \sum_{k=1}^{n} \binom{n}{k} \left(F_X(t)\right)^k \left(1-F_X(t)\right)^{n-k} = 1-(1-F_{X(1)}(t))^n$$

$$\min(X_j) \text{ تابع چگالی} : f_{X(1)}(t) = \binom{n}{1} \left(f_X(t)\right)^{1-1} \left(1-F_X(t)\right)^{n-1} f_X(t) = n\left(1-F_X(t)\right)^{n-1} f_X(t)$$

همانطور که ملاحظه می‌کنید رابطه $f_X(t) = \dfrac{dF_{X(1)}(t)}{dt}$ نیز بین تابع توزیع و تابع چگالی برقرار است.

$$\mathrm{Max}(X_j) \text{ تابع توزیع} : F_{X(n)}(t) = \sum_{k=n}^{n} \binom{n}{k} \left(F_X(t)\right)^k \left(1-F_X(t)\right)^{n-k} = \left(F_X(t)\right)^n$$

$$\mathrm{Max}(X_j) \text{ تابع چگالی} : f_{X(n)}(t) = n\binom{n}{n} \left(f_X(t)\right)^{n-1} \left(1-F_X(t)\right)^{n-n} f_X(t) = n\left(F_X(t)\right)^{n-1} f_X(t)$$

مثال ۵: متغیر تصادفی X را مدت زمان بین هر دو تماس تلفنی در یک شرکت در نظر می‌گیریم که بطور میانگین هر ۱۰ دقیقه یکبار رخ می‌دهد اگر بطور تصادفی ۴ نمونه از متغیر تصادفی X اختیار کنیم مطلوبست:

الف) تابع توزیع آمارهٔ مرتب نمونه تصادفی.

ب) تابع توزیع و چگالی ماکزیمم و مینیمم نمونه‌ها.

ج) احتمال اینکه در نمونه‌های مشاهده شده کوچکترین بازه زمانی بین دو تماس بیشتر از ۲ دقیقه باشد چقدر است؟

حل: الف) ابتدا توجه کنید که X یک متغیر تصادفی نمایی با پارامتر $\lambda = \frac{1}{10}$ می‌باشد. بنابراین تابع توزیع X عبارتست از:

$$F_X(t) = 1 - e^{-\lambda t} = 1 - e^{-\frac{1}{10}t} \qquad t > 0$$

بنابراین تابع توزیع آمارهٔ مرتب برای ۴ نمونه مشاهده شده عبارتست از:

$$F_{X(j)}(t) = \sum_{k=j}^{4} \binom{n}{k} \left(1-e^{\frac{-1}{10}t}\right)^k \left(e^{-\frac{1}{10}t}\right)^{n-k}$$

۴-۱۲-۲

ب) ماکزیمم آماره‌های مرتب در این مثال $X_{(4)}$ می‌باشد. که تابع توزیع و چگالی آن بدست می‌آید:

$$F_{X(4)}(t) = \left(F_X(t)\right)^4 = \left(1-e^{-\frac{1}{10}t}\right)^4 \qquad t > 0$$

$$F_{X(4)}(t) = 4\left(1-e^{-\frac{1}{10}t}\right)^3 \frac{1}{10} e^{-\frac{1}{10}t}$$

مینیمم آماره‌های مرتب $X_{(1)}$ می‌باشد که تابع توزیع و چگالی آن نیز بصورت زیر می‌باشد:

$$F_{X(۱)}(t) = ۱-(۱-F_X(t))^۴ = ۱-(۱-۱+e^{\frac{۱}{۱۰}t})^۴ = ۱-e^{-\frac{۴}{۱۰}t}$$

$$F_{X(۱)}(t) = ۴(۱-۱+e^{-\frac{۱}{۱۰}t})^۳ \frac{۱}{۱۰} e^{-\frac{۱}{۱۰}t} = \frac{۴}{۱۰} e^{-\frac{۳}{۱۰}t} e^{-\frac{۱}{۱۰}t} = \frac{۴}{۱۰} e^{-\frac{۴}{۱۰}t}$$

توجه کنید که توزیع کوچکترین آماره مرتب در این مثال خود یک متغیر تصادفی نمایی با پارامتر $\lambda = \frac{۴}{۱۰}$ می‌باشد.

ج) احتمال اینکه کوچکترین بازه زمانی بین دو تماس در نمونه‌ها بیشتر از ۲ دقیقه باشد برابر است با: $p(X_{(۱)} > ۲)$ و داریم:

$$p(X_{(۱)} > ۲) = ۱-p(X_{(۱)} < ۲) = ۱-F_{X(۱)}(t=۲)$$

$$= ۱-(۱-e^{-\frac{۴}{۱۰}\times ۲}) = e^{-۰/۸} \approx ۰/۴۴۹$$

۹-۱۳

مثال ۶: برای آماره مرتب $X_{(۱)}, X_{(۲)}, \ldots, X_{(n)}$ از یک متغیر تصادفی پیوسته X مقادیر مورد انتظار برای احتمال بزرگترین عضو نمونه بدست بیاورید:

حل: می‌بایستی $E\left[F_X(X_{(n)})\right]$ را بدست بیاوریم داریم:

$$f_{X(n)}(t) = n(F_X(t))^{n-۱} f_X(t)$$

$$E\left[f_x(x_{(n)})\right] = \int_{-\infty}^{+\infty} f_X(t)\, n\, (F_X(t))^{n-۱} f_X(t)$$

$$= \int_{-\infty}^{+\infty} (f_X(t))^{n+۱} \int_{-\infty}^{+\infty} = \frac{n}{n+۱} - ۰ = \frac{n}{n+۱}$$

۹-۱۴

۴ . ۹ توزیع نمونه‌ای $\frac{\overline{X}-\mu}{\frac{\delta}{\sqrt{n}}}$.

در این فصل نشان دادیم که توزیع $\frac{\overline{X}-\mu}{\frac{\delta}{\sqrt{n}}}$ نرمال استاندارد است، در صورتی که $X_۱, X_۲, \ldots, X_n$ از جامعه‌ای نرمال با میانگین μ و واریانس $\delta^۲$ انتخاب شده باشد. همینطور نشان دادیم که در صورتی که تعداد نمونه‌ها زیاد باشد $(n > ۳۰)$ توزیع $\frac{\overline{X}-\mu}{\frac{\delta}{\sqrt{n}}}$ مستقل از توزیع جامعه دارای توزیع نرمال استاندارد می‌باشد.

حال اگر واریانس یک جامعه مجهول باشد بناچار می‌بایستی از واریانس نمونه $S^۲$ در محاسبه $\frac{\overline{X}-\mu}{\frac{\delta}{\sqrt{n}}}$ استفاده کنیم. به این ترتیب می‌بایستی ابتدا تابع توزیع $\frac{\overline{X}-\mu}{\frac{\delta}{\sqrt{n}}}$ را بدست بیاوریم.

هر گاه Z یک متغیر تصادفی نرمال استاندارد باشد $(Z \sim N(۰\,,۱))$ و Y نیز یک متغیر تصادفی خی دو با n درجه آزادی باشد $(Y \sim \chi^۲_n)$ و Z و Y از یکدیگر مستقل باشند در این صورت متغیر تصادفی $T = \dfrac{Z}{\sqrt{\dfrac{Y}{n}}}$ را یک متغیر تصادفی t با n درجه آزادی می‌نامیم و بصورت $T \sim t_{(n)}$ نمایش می‌دهیم.

۹-۱۵

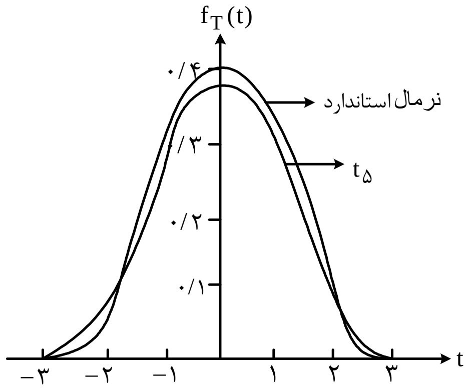
تابع چگالی متغیر تصادفی T بصورت زیر می‌باشد:

$$f_T(t) = \frac{T\left(\frac{n+۱}{۲}\right)}{T\left(\frac{n}{۲}\right)\sqrt{n\pi}} \left[۱+\frac{t^۲}{n}\right]^{-\frac{n+۱}{۲}} \qquad -\infty < t < +\infty$$

میانگین و واریانس متغیر تصادفی t_n عبارتند از:

$$E[t_n] = ۰ \qquad , \qquad var(t_n) = \frac{n}{n-۲} \qquad n > ۲$$

نمودار منحنی متغیر تصادفی t_n بسیار مشابه متغیر تصادفی نرمال استاندارد می‌باشد به شکل زیر که متغیر تصادفی $t_۵$ را با نرمال استاندارد مقایسه می‌کند توجه کنید.



در واقع با افزایش درجه آزادی متغیر تصادفی t_n نمودار منحنی هر چه بیشتر به نمودار منحنی نرمال استاندارد نزدیک می‌شود.

۹-۱۶

توجه: اگر $X \sim t_۱$

توزیع $t_۱$ بصورت زیر می‌باشد:

$$f_X(x) = \frac{۱}{\pi\,(۱+x^۲)} \qquad -\infty < x +\infty$$

حال توجه کنید که معکوس توزیع X یعنی $Y = \dfrac{۱}{X}$ برابر است با:

$$f_Y(y) = \frac{۱}{\pi\,(۱+y^۲)}$$

به عبارت دقیق‌تر عکس یک متغیر $t_۱$ بازهم $t_۱$ می‌باشد.

حال به محاسبه توزیع $\frac{\bar{X}-\mu}{\frac{\delta}{\sqrt{n}}}$ می‌پردازیم. طبق تعریف می‌دانیم اگر $Z \sim N(0,1)$ و $Y \sim \chi_n^2$ متغیر تصادفی $T=\frac{Z}{\sqrt{\frac{Y}{n}}}$ یک متغیر تصادفی t_n می‌باشد حال در نظر بگیرید: $Z=\frac{\bar{X}-\mu}{\frac{\delta}{\sqrt{n}}}$ و $Y=\frac{(n-1)S^2}{\delta^2}$ قبلاً نشان دادیم که Z و Y به ترتیب متغیرهای تصادفی نرمال استاندارد و χ_{n-1}^2 می‌باشند در نتیجه متغیر تصادفی $T=\frac{Z}{\sqrt{\frac{Y}{n-1}}}$ می‌بایستی یک متغیر تصادفی t با $n-1$ درجه آزادی می‌باشد. با ساده نمودن عبارت T داریم:

$$T=\frac{\frac{\bar{X}-\mu}{\frac{\delta}{\sqrt{n}}}}{\sqrt{\frac{(n-1)S^2}{\delta^2(n-1)}}}=\frac{\frac{\bar{X}-\mu}{\frac{\delta}{\sqrt{n}}}}{\sqrt{\frac{S^2}{\delta^2}}}=\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$$

در نتیجه متغیر تصادفی $T=\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ یک متغیر تصادفی t با $n-1$ درجه آزادی می‌باشد.

توجه کنید در صورتی که نمونه‌های تصادفی $X_1, X_2, \ldots, X_n$ از جامعه‌ای نرمال باشند رابطه $T=\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ یک متغیر تصادفی t_{n-1} می‌باشد.

اما اگر جامعه نرمال نباشد یا توزیع آنرا ندانیم یا n به اندازه کافی بزرگ $(n>30)$ باشد آنگاه باز هم $T=\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ یک متغیر تصادفی t_{n-1} می‌باشد حتی در این حالت می‌توان با توجه به تشابه نمودار منحنی‌های t_{n-1} و نرمال استاندارد می‌توان گفت که توزیع $T=\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$ برای $n>30$ همان نرمال استاندارد می‌باشد.

۹-۱۷

محاسبه مقادیر احتمال t_n.

برای بدست آوردن مقادیر مختلف احتمال t_n از جدول احتمالات $F_t(x)$ استفاده می‌کنیم بطوریکه $F_t(x)$ برابر است با:

$$F_t(x)=\int_{-\infty}^{x}\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}}(1+x^2)^{\frac{-(n+1)}{2}}dx=1-\alpha$$

$$F_t(x)=p(T<x)$$

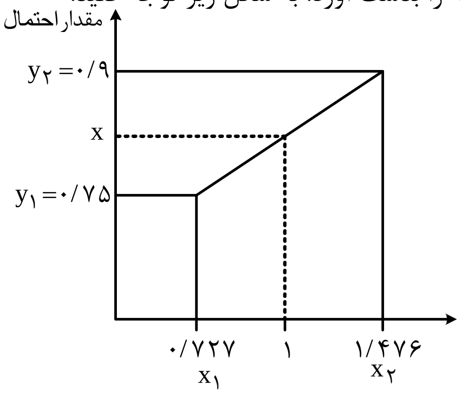
مقادیر هر یک از ردیف یا معادل با درجه آزادی و مقادیر ستون‌ها معادل با مقدار احتمال $F_t(x)$ می‌باشد. جدول زیر بخشی از جدول تابع توزیع t را نشان می‌دهد:

$n \backslash 1-\alpha$	۰/۶	۰/۷۵	۰/۹۰	۰/۹۵	۰/۹۷۵
۱	۰/۳۲۵	۱/۰۰	۳/۰۷۸	۶/۳۱۴	۱۲/۷۰۶
۲	۰/۲۸۹	۰/۸۱۶	۱/۸۸۶	۲/۹۲	۴/۳۰۳
۳	۰/۲۷۷	۰/۷۶۵	۱/۶۳۸	۲/۳۵۳	۳/۱۸۲

حال با توجه به جدول به عنوان مثال $p(t_{۲} < ۲/۹۲)$ برابر با ۰/۹۵ می‌باشد توجه کنید که مقادیر احتمال از عناوین هر یک از ستون‌ها بدست می‌آیند. برای بدست آوردن سایر مقادیر احتمال از تقریب زدن استفاده می‌کنیم.

۹-۱۸

به عنوان مثال برای بدست آوردن مقدار احتمال $p(t_{۵} < ۱)$ از آنجا که در جدول عدد ۱ در ردیف پنجم ما بین دو مقدار ۱/۴۷۶ و ۰/۷۲۷ می‌باشد می‌بایستی با استفاده از روش خطی مقدار این احتمال را تقریب بزنیم در این روش ما بین دو نقطه $(۰/۷۵ , ۰/۷۲۷)$ و $(۰/۹ , ۱/۴۷۶)$ یک خط در نظر می‌گیریم، با نوشتن معادله خط به سادگی می‌توان مقادیر احتمال ما بین این دو نقطه را بدست آورد. به شکل زیر توجه کنید.



معادله خط عبارتست از: $y = \frac{y_{۲} - y_{۱}}{x_{۲} - x_{۱}}(x - x_{۱}) + y_{۱}$

بنابراین:

$$y = \frac{۰/۹ - ۰/۷۵}{۱/۴۷۶ - ۰/۷۲۷}\ (x - ۰/۷۲۷) + ۰/۷۵$$

حال برای بدست آوردن احتمال هر یک از مقادیر ما بین $x_{۱}, x_{۲}$ کافیست آن را به جای x در معادله بالا قرار دهیم تا y که مقدار احتمال تقریبی می‌باشد بدست آید.

برای احتمال $p(t_{۵} < ۱)$ داریم:

$$p(t_{۵} < ۱) = y = \frac{۰/۹ - ۰/۷۵}{۱/۴۷۶ - ۰/۷۲۷}\ (۱ - ۰/۷۲۷) + ۰/۷۵ = ۰/۸$$

مثال: تعداد مراجعه کنندگان به یک فروشگاه در طول روز یک متغیر تصادفی نرمال می‌باشد. مدیر فروشگاه می‌داند که بطور متوسط روزانه ۱۲۰ نفر به فروشگاه مراجعه می‌کنند. تعداد مراجعه کنندگان به فروشگاه در طول یک هفته بصورت زیر بدست آمده است:

۹۰ , ۱۱۰ , ۷۵ , ۱۳۰ , ۱۵۰ , ۱۲۰ , ۱۰۰

مطلوبست:

الف) احتمال اینکه میانگین تعداد مراجعه کنندگان حداقل ۱۲۵ نفر در روز باشد؟

ب) احتمال اینکه تفاوت میانگین نمونه‌ها با میانگین جامعه حداکثر ۲ واحد باشد؟

حل: الف) ابتدا از روی نمونه‌های بدست آمده مقدار $\overline{X}$ و $S^۲$ را بدست می‌آوریم. داریم:

$$\overline{X}=\frac{۱}{n}\sum_{i=۱}^{n} x_i=\frac{۱}{۷}(۹۰+۱۱۰+۷۵+۱۳۰+۱۵۰+۱۲۰+۱۰۰)=۱۱۰/۷$$

$$S^۲=\frac{۱}{n-۱}\sum_{i=۱}^{n}(x_i-\overline{X})^۲=\frac{۱}{۶}[(۹۰-۱۱۰/۷)^۲+(۷۵-۱۱۰/۷)^۲+(۱۳۰-۱۱۰/۷)^۲+(۱۵۰-۱۱۰/۷)^۲$$

$$(۱۲۰-۱۱۰/۷)^۲+(۱۰۰-۱۱۰/۷)^۲]=۶۳۶/۹ \qquad \Rightarrow \qquad S=۲۵/۲$$

حال می‌بایستی مقدار احتمال $p(\overline{X}>۱۲۵)$ را بدست بیاوریم: (با توجه به اینکه واریانس جامعه را نمی‌دانیم بنابراین می‌بایستی از توزیع t استفاده کنیم.)

$$p(\overline{X}>۱۱۰)=p(\overline{X}-\mu>۱۲۵-۱۲۰)=p(\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}>\frac{۱۲۵-۱۲۰}{\frac{۲۵/۲}{\sqrt{۷}}})$$

$$=p(\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}>\frac{۵}{۹/۵۲})=p(t_۶>۰/۵۲۵)=۱-p(t_۶<۰/۵۲۵)=۱-۰/۶۸=۰/۳۲$$

ب) در این حالت می‌بایستی احتمال $p(\overline{X}-\mu<۲)$ را بدست بیاوریم:

$$p(\overline{X}-\mu<۲)=p(\frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}<\frac{۲}{۹/۵۲})=p(t_۶<۰/۲۱)\approx ۰/۵۸$$

۹-۲۰

توزیع نمونه‌ای نسبت واریانسهای نمونه

اگر از دو جامعه نرمال دو نمونه مستقل به اندازه‌های $n_۱$ و $n_۲$ بگیریم برای بدست آوردن نسبت واریانسهای دو نمونه ($\frac{S_۱}{S_۲}$ یا $\frac{S_۲}{S_۱}$) از توزیعی به نام F (فیشر) استفاده می‌کنیم.

تعریف: اگر U یک متغیر تصادفی خی دو با m درجه آزادی ($U\sim\chi_m^۲$) و V یک متغیر تصادفی خی دو با n درجه آزادی ($V\sim\chi_m^۲$) باشند.

در این صورت توزیع متغیر تصادفی $F=\frac{\frac{U}{m}}{\frac{V}{n}}$ را یک متغیر F با m و n درجه آزادی می‌نامند و با نماد $F\sim F_{m,n}$ نمایش می‌دهند. تابع چگالی احتمال توزیع $F_{m,n}$ بصورت زیر می‌باشد:

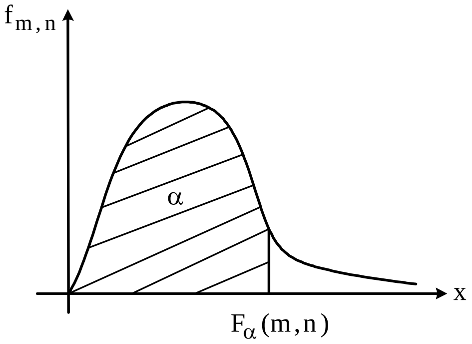
$$f_F(x)=\frac{\Gamma(\frac{m+n}{۲})(\frac{m}{n})^{\frac{m}{۲}}}{\Gamma(\frac{m}{n})\ \Gamma(\frac{n}{۲})}\ x^{\frac{m}{۲}-۱}\ (۱+\frac{m}{n}x)^{-\frac{m+n}{۲}} \qquad x>۰$$

۹-۲۱

مقادیر میانگین و واریانس توزیع $F_{m,n}$ عبارتند از:

$$E[X]=\frac{n}{n-۲} \quad n>۲ \quad , \quad var(X)=۲(\frac{۱}{n}+\frac{۱}{m})$$

شکل نمودار تابع توزیع $F_{m,n}$ بصورت زیر می‌باشد:



حال برای بدست آوردن توزیع نسبت $\frac{S_1^2}{S_2^2}$ بصورت زیر عمل می‌کنیم:

Y و X را دو جمعیت نرمال با واریانسهای δ_1^2 و δ_2^2 در نظر می‌گیریم

فرض: $X \sim N(\mu_1 , \delta_1^2) \qquad ; \qquad Y \sim N(\mu_2 , \delta_2^2)$

می‌دانیم:

$$V = \frac{(n_1-1)\, S_1^2}{\delta_1^2} \sim \chi^2_{(n_1-1)}$$

$$V = \frac{(n_2-1)\, S_2^2}{\delta_2^2} \sim \chi^2_{(n_2-1)}$$

حال با توجه به تعریف تابع توزیع F داریم:

$$F = \frac{\frac{U}{n-1}}{\frac{V}{n_2-1}} = \frac{\frac{(n_1-1)\, S_1^2}{\delta_1^2\ (n_1-1)}}{\frac{(n_2-1)\, S_2^2}{\delta_2^2\ (n_2-1)}} = \frac{\frac{S_1^2}{\delta_1^2}}{\frac{S_2^2}{\delta_2^2}} = \frac{S_1^2\, \delta_2^2}{S_2^2\, \delta_1^2} \sim F_{n_1-1\ ,\ n_2-1}$$

۹-۲۲

مثال: اگر X یک متغیر تصادفی t با n درجه آزادی باشد مربع متغیر تصادفی X دارای چه توزیعی می‌باشد؟

حل: فرض می‌کنیم: $Z \sim N(0 , 1)$ و $Y = \chi^2_n$ در این صورت داریم:

$$X = t_n \sim \frac{Z}{\sqrt{\frac{Y}{n}}}$$

$$\Rightarrow \quad X^2 = \frac{Z^2}{\frac{Y}{n}}$$

از آنجا که مربع یک متغیر تصادفی

نرمال استاندارد یک متغیر تصادفی خی دو با یک درجه آزادی است. بنابراین:

$$X^2 = \frac{\chi_1^2}{\frac{Y}{n}} = \frac{\frac{\chi_1^2}{1}}{\frac{\chi_n^2}{n}} \sim F_{1,n}$$

بنابراین مربع یک متغیر تصادفی t با n درجه آزادی یک متغیر تصادفی F با $n,1$ درجه آزادی می‌باشد.

۹-۲۳

مثال: تعداد تصادفات در دو شهر A و B متغیر تصادفی نرمال به ترتیب با واریانس ۲۰ و ۲۲ تصادف در روز می‌باشد. اگر از دو جامعه فوق دو نمونه به ترتیب با حجم ۲۱ و ۳۱ نمونه انتخاب کرده باشیم و S_A^2 , S_B^2 به ترتیب واریانسهای نمونه‌های گرفته شده از شهر A و B باشد محاسبه احتمال اینکه S_A^2 حداقل ۱/۵ برابر S_B^2 باشد؟

حل: داریم:

$$\sigma_A^2 = 20 \qquad ; \qquad \sigma_B^2 = 22$$

می‌بایستی احتمال $p(S_A^2 \geq 1/5\ S_B^2)$ را محاسبه کنیم داریم:

$$p(S_A^2 \geq 1/5\ S_B^2) = p\left(\frac{S_A^2}{S_B^2} \geq 1/5\right)$$

$$p\left(\frac{\frac{S_A^2}{\sigma_A^2}}{\frac{S_B^2}{\sigma_B^2}} \geq 1/5\ \left(\frac{22}{20}\right)\right) = p\left(\frac{S_A^2\,\sigma_B^2}{S_B^2\,\sigma_A^2} \geq 1/65\right)$$

$$= 1 - p(F_{20,30} < 1/65) \approx 1 - 0/9 = 0/1$$

۹-۲۴

تقریب توزیع F به توزیع خی دو.

در صورتی که در یک متغیر تصادفی $F_{m,n}$ شرط $n \to \infty$ برقرار باشد می‌توانیم برای محاسبه مقادیر احتمال mF از متغیر تصادفی χ_m^2 استفاده نمایم. به عبارت دیگر $mF_{m,n}$ با شرط $n \to \infty$ دارای توزیع χ_m^2 می‌باشد.

توزیع نمونه‌ای اختلاف میانگین‌ها

با استفاده از توزیع F احتمالات مربوط به نسبت واریانسهای دو نمونه را بدست آوریم حال می‌خواهیم برای دو نمونه گرفته شده از دو جامعه نرمال توزیع اختلاف میانگین‌ها را بدست آوریم.

از دو جامعه به ترتیب با میانگین‌های μ_1, μ_2 و واریانسهای σ_1^2, σ_2^2 دو نمونه به اندازه n_1, n_2 انتخاب می‌کنیم. در این صورت برای میانگین نمونه‌ها داریم:

$$\overline{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$$

$$\overline{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

از آنجا که $\overline{X}_1, \overline{X}_2$ از یکدیگر مستقل می‌باشد و $\overline{X}_1 - \overline{X}_2$ نیز یک ترکیب خطی‌از دو متغیر تصادفی نرمال می‌باشد بنابراین $\overline{X}_1 - \overline{X}_2$ نیز نرمال است و توزیع آن عبارتست از:

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

با تبدیل آن به متغیر تصادفی نرمال استاندارد داریم:

$$Z = \frac{(\overline{X}_۱ - \overline{X}_۲) - (\mu_۱ - \mu_۲)}{\sqrt{\frac{\sigma_۱^۲}{n_۱} + \frac{\sigma_۲^۲}{n_۲}}} \sim N(۰,۱)$$

حتی اگر دو جمعیت نرمال نباشند با استفاده از قضیه حد مرکزی می‌دانیم که اگر داشته باشیم $n_۱, n_۲ \geq ۳۰$ باز هم هر یک از $\overline{X}_۱, \overline{X}_۲$ دارای توزیع تقریبی نرمال می‌باشند. و در نتیجه $\overline{X}_۱ - \overline{X}_۲$ نیز دارای توزیع نرمال خواهد بود. و روابط فوق برقرار خواهند بود.

۹-۲۵

مثال: در مثال قبل اگر متوسط تعدد تصادفات در شهر A ۱۰۰ و در شهر B ۱۲۰ تصادف در روز باشند. اگر دو نمونه به حجم ۳۰ انتخاب کنیم در این صورت مطلوبست احتمال اینکه میانگین تعداد تصادفات در شهر B حداقل ۱۸ تصادف از شهر A بیشتر باشد

حل: با توجه به مفروضات مساله داریم:

$$\mu_A = ۱۰۰ \qquad \sigma_A^۲ = ۲۰ \qquad n_A = ۳۰$$

$$\mu_B = ۱۲۰ \qquad \sigma_B^۲ = ۲۲ \qquad n_B = ۳۰$$

$$\overline{X}_A - \overline{X}_B \sim N(۱۰۰ - ۱۲۰, \frac{۲۰}{۳۰} + \frac{۲۲}{۳۰}) \sim N(-۲۰, ۱/۴)$$

حال می‌بایستی احتمال $p(\overline{X}_B \geq \overline{X}_A + ۱۸)$ را محاسبه کنیم:

$$p(\overline{X}_B \geq \overline{X}_A + ۱۸) = p(\overline{X}_A - \overline{X}_B \leq -۱۸)$$

$$= p(\overline{X}_A - \overline{X}_B - (\mu_A - \mu_B)) \leq -۱۸ - (۱۰۰ - ۱۲۰))$$

$$= p\left(\frac{\overline{X}_A - \overline{X}_B - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^۲}{n_A} + \frac{\sigma_B^۲}{n_B}}} \leq \frac{۲}{۱/۱۸}\right)$$

$$= p(Z \leq ۱/۶۹) = ۰/۹۷۵$$

۹-۲۶

اگر در محاسبه توزیع اختلاف میانگین نمونه‌ها واریانس دو جامعه نامعلوم اما مساوی باشد $(\sigma_۱^۲ = \sigma_۲^۲ = \sigma^۲)$ می‌توانیم از واریانس نمونه‌ها $(S_۲^۲, S_۱^۲, ...)$ به عنوان تخمین $\sigma^۲$ استفاده کنیم. در این حالت از میانگین وزنی این دو واریانس برای برآورد $\sigma^۲$ استفاده می‌کنیم:

$$S_p^۲ = \frac{(n_۱ - ۱) S_۱^۲ + (n_۲ - ۱) S_۲^۲}{(n_۱ + n_۲ - ۲)}$$

در این حالت اگر دو جامعه نرمال باشند داریم:

$$\frac{(n_۱ + n_۲ - ۲) S_p^۲}{\sigma^۲} = \frac{(n_۱ - ۱) S_۱^۲}{\sigma^۲} + \frac{(n_۲ - ۱) S_۲^۲}{\sigma^۲} = \chi^۲_{(n_۱ + n_۲ - ۲)}$$

حال با توجه به توزیع t داریم:

$$T = \frac{(\overline{X}_۱ - \overline{X}_۲) - (\mu_۱ - \mu_۲)}{S_p \sqrt{\frac{۱}{n_۱} + \frac{۱}{n_۲}}} \sim t_{(n_۱ + n_۲ - ۲)}$$

بنابراین در محاسبه احتمال مربوط به نمونه‌های بدست آمده از دو جامعه با واریانس نامعلوم اما مساوی می‌بایستی از توزیع t با $n_۱ + n_۲ - ۲$ درجه آزادی استفاده کنیم.