برآورد:

تعریف: فرض کنید $X_1, \ldots, X_n$ یک نمونه تصادفی از جامعه‌ای با پارامتر مجهول θ باشد، آماره $T = T(X_1, \ldots, X_n)$ را برای θ یک برآوردگر نقطه‌ای گوییم، اگر T به هیچ پارامتر مجهولی وابسته نباشد، اما توزیع احتمال آن به θ بستگی داشته باشد.

برآورد کننده نااریب:

تعریف: برآوردگری مانند $\hat{\theta}$ را یک برآوردگر نااریب پارامتر θ گوییم اگر و تنها اگر

$$E[\hat{\theta}] = \theta$$

مثال: همانطور که قبلاً دیدیم

$$E[\bar{X}] = \mu$$

پس $\bar{X}$ یک برآوردگر نااریب برای μ (میانگین جامعه) می‌باشد.

مثال: نشان دهید که S^2، واریانس نمونه‌ای، یک برآوردگر نااریب برای σ^2 است.

جواب: با مشاهده به بخش توزیع‌های نمونه‌ای، می‌بینیم $E[S^2] = \sigma^2$ پس، S^2 یک برآوردگر نااریب برای σ^2 است.

مثال: مشابه به مثال قبل می‌توان نشان داد

$$\hat{\sigma}^2 = \frac{(n-1)S^2}{n} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n}$$

یک برآوردگر اریب برای σ^2 است، یعنی

$$E[\hat{\sigma}^2] = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

تعریف: اگر $\hat{\theta}$ یک برآوردگر از θ باشد و $\lim_{n \to \infty} E[\hat{\theta}] = \theta$، در این صورت $\hat{\theta}$ را یک برآوردگر مجانباً نااریب از پارامتر θ گوییم.

نتیجه: $\hat{\sigma}^2 = \frac{n-1}{n} S^2$ یک برآوردگر مجانباً نااریب از σ^2 است.

مثال: اگر $X_1, \ldots, X_n$ نمونه‌ای تصادفی از جامعه‌ای با تابع چگالی احتمال

$$f_X(x) = \begin{cases} e^{-(x-\delta)} & x > \delta \\ 0 & o.w \end{cases}$$

باشند، آیا $\bar{X}$ برآوردگری نااریب برای δ است؟

حل: می‌دانیم $E[\bar{X}] = \mu$، لذا ابتدا میانگین جامعه را به دست می‌آوریم:

$$\mu = \int_{\delta}^{\infty} x e^{-(x-\delta)} dx = -(x+1) e^{-(x-\delta)} \Big|_{\delta}^{\infty} = \delta + 1$$

لذا $\bar{X}$ برآوردگری اریب برای δ است. امّا $\bar{X} - 1$ نااریب است.

میانگین مربع خطا (MSE):

تعریف: اگر $\hat{\theta}$ یک برآوردگر از θ باشد، در این صورت از میانگین مربع خطا (MSE) که به صورت $MSE(\hat{\theta}) = E(\theta - \hat{\theta})^2$ تعریف می‌شود، به عنوان معیار سنجش دقت برآوردگر استفاده می‌شود.

نکته:

$$MSE(\hat{\theta}) = Var(\hat{\theta}) + bias(\hat{\theta})^2$$

که در اینجا $bias(\hat{\theta})$ به صورت زیر تعریف می‌شود:

$$bias(\hat{\theta}) = E(\hat{\theta}) - \theta$$

و میزان اریبی برآوردگر تعریف می‌شود.

اثبات:

$$\begin{aligned} MSE[\hat{\theta}] &= E[(\hat{\theta} - \theta)^2] \\ &= E[\hat{\theta}^2 - 2\theta\hat{\theta} + \theta^2] \\ &= E[\hat{\theta}^2] - 2\theta E[\hat{\theta}] + \theta^2 \\ &= E[\hat{\theta}^2] - E[\hat{\theta}]^2 + E[\hat{\theta}]^2 - 2\theta E[\hat{\theta}] + \theta^2 \\ &= Var(\hat{\theta}) + (\theta - E[\hat{\theta}])^2 = Var(\hat{\theta}) + bias^2(\hat{\theta}) \end{aligned}$$

مثال: اگر در مثال قبل، از $\bar{X}-1$ به عنوان برآوردگری استفاده کنیم، میزان MSE، $\hat{\delta}$ را بیابید.

حل:

$$Var(\hat{\delta}) = Var(\bar{X}-1) = \frac{\sigma^2}{n}$$

$$E[X^2] = \int_{\delta}^{\infty} x^2 e^{-(x-\delta)} dx = -(x^2+2x+2) e^{-(x-\delta)} \Big|_{\delta}^{\infty}$$

$$= \delta^2 + 2\delta + 2$$

$$\sigma^2 = E[X^2] - E[X]^2 = (\delta^2 + 2\delta + 2) - (\delta+1)^2 = 1$$

$$\Rightarrow Var(\hat{\delta}) = Var(\bar{X}) = \frac{\sigma^2}{n} = \frac{1}{n}$$

$$MSE[\hat{\delta}] = \frac{1}{n}$$

تذکر: قبلاً دیدیم که $Var(\bar{X}) = \frac{\sigma^2}{n}$ پس اگر $\bar{X} = \hat{\mu}$ آنگاه

$$MSE[\bar{X}] = \frac{\sigma^2}{n}$$

تمرین: در مورد MSE، S^2 به عنوان برآوردگر σ^2 چه می‌توان گفت؟ آیا S_*^2 برآوردگری ناریب از σ^2 است. آیا MSE آن از S^2 کمتر است؟

کارآیی برآوردگر:

اگر بخواهیم بین دو یا چند برآوردگر ناریب یکی را انتخاب کنیم، برآوردگری را انتخاب می‌کنیم که واریانس کمتری داشته باشد، که این امر با توجه به فرمول MSE قابل توجیه است.

تعریف: اگر $\hat{\theta}_1$ و $\hat{\theta}_2$ برآوردگرهای ناریب از θ باشند، $\hat{\theta}_1$ را کاراتر از $\hat{\theta}_2$ گوییم هرگاه:

$$Var(\hat{\theta}_1) \leq Var(\hat{\theta}_2)$$

تعریف: برای برآوردگرهای $\hat{\theta}_1$ و $\hat{\theta}_2$ نسبت $\frac{Var(\hat{\theta}_1)}{Var(\hat{\theta}_2)}$ را کارایی نسبی $\hat{\theta}_2$ نسبت به $\hat{\theta}_1$ تعریف می‌کنیم. اگر $\hat{\theta}_2$ کاراتر باشد، این کسر بزرگتر از 1 است و در غیر این صورت کمتر از یک.

فاصله اطمینان :

یک فاصله اطمینان بازه‌ای است با کران بالای U و کران پایین L که می‌توان با اطمینان خاصی، احتمال $1-\alpha$، گفت که پارامتر مجهول θ از جامعه در این بازه قرار دارد؛ یعنی؛

$$P(L < \theta < U) = 1-\alpha$$

درجه اطمینان این فاصله اطمینان $1-\alpha$ است و به این فاصله یک فاصله اطمینان $100(1-\alpha)$ درصدی گفته می‌شود. معمولاً" مقدار α عددی کوچک‌تر از 0.05 اختیار می‌شود. مثلاً" اگر $\alpha=0.01$ فاصله اطمینان 99 درصدی و درجه اطمینان 0.99 خواهد بود.

فاصله اطمینان برای میانگین یک جامعه نرمال با واریانس معلوم

اگر $X_1, \ldots, X_n$ یک نمونه تصادفی از جامعه‌ای نرمال با میانگین مجهول μ و واریانس σ^2 باشد. همانطور که دیدیم؛ $\bar{X}$ از توزیع نرمال با میانگین μ و واریانس $\frac{\sigma^2}{n}$ پیروی می‌کند.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow Z = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

$$1-\alpha = P\left(-Z_{\alpha/2} < Z < Z_{\alpha/2}\right)$$

$$= P\left(-Z_{\alpha/2} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < Z_{\alpha/2}\right)$$

$$= P\left(\bar{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

لذا فاصله $\left(\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$ یک فاصله اطمینان $100(1-\alpha)$ درصدی برای μ می‌باشد.

$$\bar{x} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

مثال: قد دانشجویان یک کلاس از توزیع نرمال با میانگین μ و واریانس 64 پیروی می‌کند. اگر ده نفر از دانشجویان کلاس را به تصادف انتخاب کنیم و میانگین قد آنها 171.3 باشد، یک فاصله اطمینان 95 و 99 درصدی برای میانگین قد دانشجویان کلاس به دست آورید.

$$\sigma^2 = 64 \qquad X_i \sim N(\mu, 64)$$

$$n = 10 \qquad \bar{x} = 171.3 \qquad \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$$

$$Z_{\alpha/2} = Z_{0.025} = 1.96$$

فاصله اطمینان 95 درصدی
$$\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \Rightarrow 171.3 \pm 1.96 \frac{8}{\sqrt{10}}$$

$$\alpha = 0.01 \Rightarrow \alpha/2 = 0.005 \Rightarrow Z_{\alpha/2} = Z_{0.005} = 2.57$$

فاصله اطمینان 99 درصدی
$$171.3 \pm 2.57 \frac{8}{\sqrt{10}}$$

$$(164.798, 177.802)$$

تذکر: اگر حجم نمونه، n، بیش از 30 باشد، طبق قضیه حد مرکزی فارغ از اینکه جامعه نرمال باشد یا نه، فاصله اطمینان بالا معتبر است.

تذکر: فاصله اطمینان منحصر به فرد نیست؛ در واقع در اینجا ما فاصله اطمینان متقارن را در نظر گرفته‌ایم؛ برای مشخص شدن بحث توجه کنید که:

$$1-\alpha = P(-Z_{\alpha/2} < Z < Z_{\alpha/2})$$
$$= P(-Z_{\frac{2\alpha}{3}} < Z < Z_{\alpha/3}) = P(-Z_{a_1\alpha} < Z < Z_{a_2\alpha})$$

که در آن $a_1 + a_2 = 1$.

فاصله اطمینان برای میانگین یک جامعه نرمال با واریانس مجهول

اگر $X_1, \ldots, X_n$ یک نمونه تصادفی از جامعه‌ای نرمال با میانگین μ و واریانس مجهول σ^2 باشند، از آماره

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

برای تعیین فاصله اطمینان استفاده می‌کنیم:

$$1-\alpha = P(-t_{\alpha/2, n-1} < T < t_{\alpha/2, n-1})$$

$$= P\left(-t_{\alpha/2, n-1} < \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} < t_{\alpha/2, n-1}\right)$$

$$= P\left(\bar{X} - t_{\alpha/2, n-1}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2, n-1}\frac{S}{\sqrt{n}}\right)$$

لذا $\bar{x} \pm t_{\alpha/2, n-1}\frac{s}{\sqrt{n}}$ یک فاصله اطمینان $100(1-\alpha)$ درصدی برای میانگین جامعه است.

$$\bar{x} \pm t_{\alpha/2, n-1}\frac{s}{\sqrt{n}}$$

تذکر مهم: اگر حجم نمونه بیش از 30 باشد، شرط نرمال بودن جامعه با توجه به قضیه حد مرکزی لازم نیست و ثانیاً توزیع t - استیودنت را می‌توان توزیع نرمال در نظر گرفت. پس فاصله اطمینان به صورت زیر خواهد بود:

$$\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$$

مثال: یک سازنده رنگ می‌خواهد متوسط زمان خشک شدن رنگ جدید دیوارهای ساختمان را بررسی کند. اگر برای 12 سطح آزمایشی با مساحت‌های برابر، میانگین زمان خشک شدن را 66.3 دقیقه و انحراف معیار را مساوی 8.4 دقیقه به دست آورد، یک فاصله اطمینان 95 درصدی برای میانگین واقعی زمان خشک شدن رنگ به دست آورید.

$n=12$ ، $\bar{x} = 66.3$ ، $S = 8.4$ ، $\alpha = 0.05 \Rightarrow \alpha/2 = 0.025$

$$t_{\alpha/2, n-1} = t_{0.025, 11} = 2.201$$

$$\Rightarrow \bar{x} \pm t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} = 66.3 \pm 2.201 \frac{8.4}{\sqrt{12}}$$

$\Rightarrow (61, 71.6)$ فاصله اطمینان 95 درصدی

فاصله اطمینان برای اختلاف میانگین دو جامعه با واریانس های معلوم :

فرض کنید X_1 ، X_2 ، ... ، X_{n_1} نمونه ای به حجم n_1 از جامعه ای نرمال با میانگین μ_1 و واریانس σ_1^2 و Y_1 ، ... ، Y_{n_2} نمونه ای به حجم n_2 از جامعه ای نرمال با میانگین μ_2 و واریانس σ_2^2 باشد و این دو جامعه از هم مستقل باشند. همانطور که قبلاً دیدیم :

$$X_i \sim N(\mu_1, \sigma_1^2) \Rightarrow \bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{n_1})$$

$$Y_i \sim N(\mu_2, \sigma_2^2) \Rightarrow \bar{Y} \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$$

لذا

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

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

پس فاصله اطمینان $100(1-\alpha)$ درصدی برای اختلاف میانگین دو جامعه به صورت زیر خواهد بود :

$$(\bar{x} - \bar{y}) \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

نکته 1 : اگر حجم دو نمونه از 30 بیشتر باشد ، فرمول فوق برای جامعه های غیر نرمال نیز درست است .

نکته 2 : اگر حجم دو نمونه بیش از 30 باشد و واریانس های دو جامعه مجهول باشند کافیست در فرمول فوق از واریانس نمونه ای به جای واریانس جامعه استفاده کنیم .

$$(\bar{x}-\bar{y}) \pm Z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

مثال : انحراف معیار طول عمر لامپ های نوع A برابر ۱۰ و لامپ های نوع B برابر 15 است. یک نمونه تصادفی 35 تایی از لامپ های نوع A در نظر گرفته‌ایم و متوسط طول عمر آنها برابر 512 بوده است. یک نمونه تصادفی ۴۰ تایی از لامپ های نوع B را نیز در نظر گرفته و متوسط طول عمر آنها را 525 به دست آورده‌ایم. یک فاصله اطمینان 95 درصدی برای تفاضل میانگین طول عمر این دو نوع لامپ به دست آورید.

حل :

$\sigma_A = 10 \,,\ \sigma_B = 15 \quad n_A = 35 \quad \bar{x}_A = 512 \,,\ n_B = 40 \quad \bar{x}_B = 525$

$$(\bar{x}_B-\bar{x}_A) \pm Z_{\alpha/2}\sqrt{\frac{\sigma_A^2}{n_A}+\frac{\sigma_B^2}{n_B}}$$

$$\alpha = 0.05 \Rightarrow \alpha/2 = 0.025 \Rightarrow Z_{\alpha/2} = Z_{0.025} = 1.96$$

$$(525-512) \pm 1.96\sqrt{\frac{100}{35}+\frac{225}{40}} \Rightarrow (7.291 \,,\ 18.703)$$

فاصله اطمینان برای اختلاف میانگین دو جامعه نرمال با واریانس های مجهول :

دو نمونه به حجم n_1 و n_2 از دو جامعه نرمال مستقل از هم با میانگین های μ_1 و μ_2 و واریانس های مجهول σ_1^2 و σ_2^2 در نظر بگیرید. می خواهیم یک فاصله اطمینان $100(1-\alpha)$ درصدی برای اختلاف میانگین دو جامعه یعنی $\mu_1-\mu_2$ بیابیم. در اینجا حتما باید واریانس دو جامعه برابر باشند : $\sigma_1^2=\sigma_2^2=\sigma^2$

حال به تعریف واریانس آمیخته S_p^2 می پردازیم :

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

از طرفی

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

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

به همین ترتیب داریم:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

پس با توجه به تعریف توزیع tـ استیودنت داریم:

$$T = \frac{Z}{\sqrt{\frac{Y}{n_1+n_2-2}}} \sim t(n_1+n_2-2)$$

از طرفی

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$$

بنابراین مشابه قبل فاصله اطمینان به صورت زیر خواهد بود:

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2,\, n_1+n_2-2}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

مثال (کتاب دکتر فرنوش): میزان برق مصرفی دو شهر A و B متغیرهای تصادفی نرمال با $\sigma_A^2=150$ و $\sigma_B^2=100$ هستند. برای مقایسه متوسط مصرف در دو شهر یک نمونه 15 تایی از دو شهر انتخاب می‌کنیم و نتایج زیر را به دست می‌آوریم:

$\bar{X}_A = 2590$, $\bar{X}_B = 2560$

الف) با فاصله اطمینان 99 درصدی چه نظری می‌توان درباره تفاضل میانگین مصرف در دو شهر داشت.

ب) اگر برای دو نمونه داشته باشیم $S_A^2=165$ و $S_B^2=120$ یک فاصله اطمینان ۹۹ درصدی برای حالتی که واریانسها نامعلوم ولی مساوی باشند بسازید.

حل:

الف)

$\bar{x}_A = 2590$, $n_A = 15$, $\sigma_A^2 = 100$

$\bar{x}_B = 2560$, $n_B = 15$, $\sigma_B^2 = 150$

$\alpha = 0.01 \Rightarrow Z_{\alpha/2} = Z_{0.005} = 2.57$

$$(\bar{x}_A - \bar{x}_B) \pm Z_{\alpha/2}\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$$

$$(2590 - 2560) \pm 2.57\sqrt{\frac{100}{15} + \frac{150}{15}}$$

$(19.48, 40.51)$ فاصله اطمینان ۹۹ درصدی $\mu_A - \mu_B$

لذا با اطمینان ۹۹ درصد چون کل بازه بزرگتر از صفر است، می توان گفت که متوسط مصرف در شهر A از شهر B بیشتر است.

ب)

$$S_p^2 = \frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2} = \frac{14\times 120 + 14\times 165}{15+15-2}$$

$$= 142.5 \Rightarrow S_p = \sqrt{S_p^2} = 11.93$$

$$\alpha = 0.01 \Rightarrow \alpha/2 = 0.005 \Rightarrow t_{\alpha/2, n_A+n_B-2} = t_{0.005, 28} = 2.76$$

$$(\bar{x}_A - \bar{x}_B) \pm t_{\alpha/2, n_A+n_B-2}\, S_p\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}$$

$$(2590 - 2560) \pm 2.76 \times (11.93) \times \sqrt{\frac{1}{15} + \frac{1}{15}}$$

$(17.97, 42)$

فاصله اطمینان برای نسبت p :

فرض کنید $X_1, \ldots, X_n$ یک نمونه تصادفی به حجم n، که در آن $n > 30$ از یک جامعه برنولی با نسبت p باشد. در این صورت اگر $\hat{p}$ به صورت

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$$

تعریف گردد با توجه به قضیه حد مرکزی داریم:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} \simeq \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}\hat{q}}{n}}}$$

که در اینجا $\hat{q} = 1 - \hat{p}$. لذا $\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$ یک فاصله اطمینان $100(1-\alpha)$ درصدی برای $\hat{p}$ خواهد بود.

مثال: برای بررسی شیوع بیماری فشارخون در افراد بیش از 30 سال یک نمونه 100 تایی تصادفی در نظر گرفته ایم و مشاهده شد که 23 نفر دارای این بیماری هستند. یک فاصله اطمینان 99 درصدی برای نسبت p شیوع این بیماری در افراد بالای به از 30 سال سن بیابید.

حل:

$$\hat{p} = 0.23 \quad n = 100 \quad \hat{q} = 0.77 = 1 - 0.23$$

$$\alpha = 0.01 \Rightarrow \alpha/2 = 0.005 \Rightarrow Z_{\alpha/2} = Z_{0.005} = 2.57$$

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

$$0.23 \pm 2.57\sqrt{\frac{0.23 \times 0.77}{100}}$$

نکته: اگر بخواهیم با احتمال $1-\alpha$ خطای برآورد p با $\hat{p}$ کمتر از e باشد، با توجه به فرمول فاصله اطمینان برای تعیین n داریم،

$$e = Z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} \Rightarrow e^2 = Z_{\alpha/2}^2 \frac{\hat{p}\hat{q}}{n} \Rightarrow n = Z_{\alpha}^2 \frac{\hat{p}\hat{q}}{e^2}$$

امّا چون p و q هر دو اعدادی بین صفر و یک هستند، لذا $\max \hat{p}\hat{q} = 1/4$ لذا مقدار n را

برابر با

$$n = \frac{Z_{\alpha/2}^2}{4e^2}$$

در نظر می گیریم.

مثال: در مثال قبل حجم نمونه چقدر باشد که خطای $\hat{p}$ با احتمال 95 درصد کمتر از 0.02 باشد.

$$e = 0.02 \qquad \alpha = 0.05 \Rightarrow \alpha/2 = 0.025 \Rightarrow Z_{\alpha/2} = Z_{0.025} = 1.96$$

$$n = \frac{(1.96)^2}{4 \times 4 \times 10^{-4}} = 2401$$

فاصله اطمینان برای اختلاف نسبت دو جامعه $P_1 - P_2$:

اگر $X_1, \ldots, X_{n_1}$ یک نمونه تصادفی به حجم n_1 از یک جامعه برنولی با نسبت P_1 و $Y_1, \ldots, Y_{n_2}$ یک نمونه تصادفی به حجم n_2 از جامعه برنولی دیگری با نسبت P_2 باشد و همچنین دو جامعه از یکدیگر مستقل باشند و $n_1, n_2 \geq 30$ با توجه به قضیه حد مرکزی داریم:

$$\hat{P}_1 \sim N\left(P_1, \frac{P_1 q_1}{n_1}\right)$$

$$\hat{P}_2 \sim N\left(P_2, \frac{P_2 q_2}{n_2}\right)$$

$$\hat{P}_1 - \hat{P}_2 \sim N\left(P_1 - P_2, \frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}\right)$$

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (P_1 - P_2)}{\sqrt{\frac{P_1 q_1}{n_1} + \frac{P_2 q_2}{n_2}}}$$

لذا $(\hat{P}_1 - \hat{P}_2) \pm Z_{\alpha/2}\sqrt{\frac{\hat{P}_1 \hat{q}_1}{n_1} + \frac{\hat{P}_2 \hat{q}_2}{n_2}}$ یک فاصله اطمینان $100(1-\alpha)$ درصدی برای $P_1 - P_2$ است.

مثال: برای بررسی نسبت شیوع بیماری فشارخون در بین زنان و مردان یک نمونه صد نفره از هر کدام در نظر گرفته ایم. مشاهده شده است که 32 مرد و 25 زن مبتلا به فشار خون بالا هستند. اولاً یک فاصله اطمینان 80 درصدی برای اختلاف نسبت شیوع این بیماری بین زنان و مردان بدست آورید و سپس آن را تحلیل نمایید.

$$\hat{P}_1 = 0.32 \,,\ \hat{P}_2 = 0.25 \quad \alpha = 0.2 \Rightarrow \alpha/2 = 0.1 ,\ Z_{\alpha/2} = Z_{0.1} = 1.282$$

$$\hat{q}_1 = 0.68 ,\ \hat{q}_2 = 0.75$$

$$(\hat{P}_1 - \hat{P}_2) \pm Z_{\alpha/2} \sqrt{\frac{\hat{P}_1 \hat{q}_1}{n_1} + \frac{\hat{P}_2 \hat{q}_2}{n_2}}$$

$$\Rightarrow (0.32 - 0.25) \pm 1.282 \sqrt{\frac{(0.68)(0.32)}{100} + \frac{(0.25)(0.75)}{100}}$$

فاصله اطمینان برای مشاهدات زوجی:

فرض کنید می خواهیم تأثیر یک رژیم غذایی خاص را بر روی افراد بررسی کنیم. برای این منظور یک نمونه تصادفی به حجم n در نظر گرفته و وزن آنها را ثبت می کنیم و سپس پس از یک دوره 6 ماهه از مصرف این رژیم غذایی وزن این افراد را دوباره ثبت می کنیم. واضح است که وزن هر فرد قبل و بعد از رژیم غذایی به هم وابسته است. بنابراین در اینجا مشاهدات زوجی مانند (X_i , Y_i) داریم. برای رفع مشکل استقلال متغیر D_i را به صورت $D_i = X_i - Y_i$ تعریف می کنیم. اگر D_i دارای توزیع نرمال با میانگین μ_D باشد، در این صورت

$$T = \frac{\bar{D} - \mu_D}{\frac{S_D}{\sqrt{n}}} \sim t_{(n-1)}$$

پس فاصله زیر یک فاصله اطمینان $100(1-\alpha)$ درصدی برای $\mu_D = \mu_X - \mu_Y$ خواهد بود:

$$\bar{D} \pm t_{\alpha/2, n-1} \frac{S_D}{\sqrt{n}}$$

مثال: می‌خواهیم میزان تأثیر یک رژیم غذایی خاص را در کاهش وزن افراد بررسی کنیم. برای بررسی ده نفر به طور تصادفی انتخاب شده و وزن آنها قبل و یک ماه بعد از مصرف رژیم غذایی اندازه‌گیری شده و اطلاعات آن در جدول زیر آمده است. بر اساس اطلاعات این جدول یک فاصله اطمینان 98 درصدی برای اختلاف میانگین وزن این افراد قبل و بعد از مصرف رژیم غذایی به دست آورید.

	وزن بعد	وزن قبل	D
1	76	81	-5
2	60	52	8
3	85	87	-2
4	58	70	-12
5	91	86	5
6	75	77	-2
7	82	90	-8
8	64	63	1
9	79	85	-6
10	88	83	5

$\bar{D} = -1/6 \quad S_D = 6.38 \quad \alpha = 0.02 \Rightarrow \alpha/2 = 0.01$

$t_{0.01,9} = 2.82 \Rightarrow \bar{D} \pm t_{\alpha/2} \frac{s_d}{\sqrt{n}}$

$-1.6 \pm 2.82 \frac{6.38}{\sqrt{10}} \qquad (-7.29, 4.09)$

فاصله اطمینان برای واریانس یک جامعه نرمال :

اگر $x_1, \dots, x_n$ یک نمونه تصادفی از جامعه ای نرمال با میانگین μ و واریانس σ^2 باشد در این صورت

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

دارای توزیع خی دو با $n-1$ درجه آزادی است. لذا

$$1-\alpha = P\left(\chi^2_{1-\alpha/2,n-1} < \chi^2_{n-1} < \chi^2_{\alpha/2,n-1}\right)$$

$$= P\left(\chi^2_{1-\alpha/2,n-1} < \frac{(n-1)s^2}{\sigma^2} < \chi^2_{\alpha/2,n-1}\right)$$

$$= P\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}\right)$$

بنابراین فاصله $\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}\right)$ یک فاصله اطمینان $100(1-\alpha)$ درصدی برای σ^2 است.

مثال : وزن جعبه هایی که توسط یک شرکت تولید می شود از توزیع نرمال پیروی می کند. اگر یک نمونه تصادفی ۹ تایی از جعبه ها را در نظر گرفته و وزن آن ها برابر با ۱۰۰، 103، 105، 102، 108، 107، 106، 101، 103 باشد، یک فاصله اطمینان ۹۰ درصدی برای واریانس وزن جعبه ها بیابید.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 103.88 \qquad s^2 = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n-1} = 7.61$$

$$s = 2.7588 \qquad \alpha = 0.1 \Rightarrow \alpha/2 = 0.05$$

$$\chi^2_{0.05,8} = 15.507 \quad , \quad \chi^2_{0.95,8} = 2.73$$

$$\left(\frac{8 \times 7.61}{15.50}, \frac{8 \times 7.61}{2.73}\right)$$

نکته: اگر میانگین جامعه معلوم باشد می توان از آماره

$$\chi^2 = \frac{n S_\mu^2}{\sigma^2} \sim \chi^2(n)$$

استفاده کرد. در این صورت فاصله اطمینان به صورت زیر خواهد بود:

$$\left(\frac{n S_\mu^2}{\chi^2_{\alpha/2, n}} , \frac{n S_\mu^2}{\chi^2_{1-\alpha/2, n}} \right)$$