Two samples confidence intervals

Confidence intervals for $\mu_1-\mu_2$, σ_1^2, σ_2^2. Known

Let $Y_{11} \dots Y_{1,n_1}$ denote a random of size n_1 from $N(\mu_1, \sigma_1^2)$, σ_1^2 Known

Let $Y_{21} \dots Y_{2,n_2}$ denote a random of size n_2 from $N(\mu_2, \sigma_2^2)$, σ_2^2 Known.

Y_1's and Y_2's are independent r.v's.

Problem to find a confidence interval for $\mu_1-\mu_2$

Let $\bar{Y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} Y_{1i}$, $\bar{Y}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_{2i}$.

$$E[\bar{Y}_1] = \mu_1 \qquad Var(\bar{Y}_1) = \frac{\sigma_1^2}{n_1}$$

$$E[\bar{Y}_2] = \mu_2 \qquad Var(\bar{Y}_2) = \frac{\sigma_2^2}{n_2}$$

$$E[\bar{Y}_1-\bar{Y}_2] = \mu_1-\mu_2 \qquad Var(\bar{Y}_1-\bar{Y}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\frac{\bar{Y}_1-\bar{Y}_2-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

A $1-\alpha$ confidence interval for $\mu_1-\mu_2$ is

$$\bar{Y}_1-\bar{Y}_2 \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

Exercise

The length of life of brand Y_1 light bulbs is assumed to be $N(\mu_1, 784)$. The length of life of brand Y_2 light bulbs is assumed to be $N(\mu_2, 627)$ and independent of Y_1. If a random sample of $n_1 = 56$ brand Y_1 light bulbs yielded a mean $\bar{Y}_1 = 937.4$ hours and a random sample of $n_2 = 57$ brand Y_2 light bulbs yielded a mean of $\bar{Y}_2 = 988.9$ hours, find a 90% confidence interval for $\mu_1 - \mu_2$

$$\bar{Y}_1 - \bar{Y}_2 \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = 937.4 - 988.9 \pm (2.575) \sqrt{\frac{784}{56} + \frac{627}{57}}$$

$$= -51.5 \pm 12.875 = [-64.375, -38.625]$$

Confidence intervals for $\mu_1-\mu_2$, σ_1^2, σ_2 unknown, but equal

Let $Y_{11}, \dots, Y_{1n_1}$ denote a random sample of size n_1 from $N(\mu_1, \sigma^2)$

Let $Y_{21} \dots Y_{2,n_2}$ denote a random sample of size n_2 from $N(\mu_2, \sigma^2)$

The r.v's Y_{1i}'s and Y_{2j}'s are independent

A $1-\alpha$ confidence interval for $\mu_1-\mu_2$ is

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

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \text{pooled estimator of } \sigma^2$$

Note that $\bar{Y}_1 - \bar{Y}_2 \sim N\left(\mu_1-\mu_2, \frac{\sigma^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$

So $$\frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim N(0,1)$$

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

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

Since $\frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ and $\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{\sigma^2}$ are independent r.v's

$$\frac{\dfrac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}}{\sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1+n_2-2)\sigma^2}}} = \frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-1)$$

8.101 A mathematics test is given to a class of 50 students randomly selected from high school 1 and also to a class of 4 students randomly selected from high school 2. For the class at high school 1, the sample mean is 75 points and the sample standard deviation is 10 points. For the class at high school 2, the sample mean is 72 points and the sample standard deviation is 8 points. Construct a 95% confidence interval for the difference in the mean scores. What assumptions are necessary?

The two samples confidence interval for $\mu_1 - \mu_2$ is

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

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = \frac{(49)10^2 + 49(8)^2}{50+50-2} = 82$$

$$t_{\frac{\alpha}{2}} = t_{0.025} = 1.96$$

$$75 - 72 \pm 1.96\sqrt{82}\sqrt{\frac{1}{50}+\frac{1}{50}} = 3 \pm 3.55$$

$(-0.55, 6.55$

8.71 Two new drugs were given to patients with heart disease. The first drug lowered the blood pressure of 16 patients an average of 11 points, with a standard deviation of 6 points The second drug lowered the blood presure of 20 patients an average of 12 points. Determine a 95% confidence interval for the difference in the mean reductions in blood pressure, assuming that the measurements are normally distributed with equal variances

The confidence interval is

$$\bar{Y}_1 - \bar{Y}_2 \pm t_{\frac{\alpha}{2}} S_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$\nu = n_1+n_2-2 = 16+20-2 = 34$ $\qquad t_{0.025}(34) = 1.96$

$$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2} = \frac{(15)6^2+19\cdot 8^2}{34} = 51.6470$$

$$11-12 \pm 1.96\sqrt{51.6470}\sqrt{\frac{1}{16}+\frac{1}{20}} = -1 \pm 4.7245$$

$$= (-5.7245, 3.7245)$$

Confidence intervals for $\frac{\sigma_1^2}{\sigma_2}$

Let $Y_{11}, \ldots, Y_{1n_1}$ denote a random sample of size n_1 from $N(\mu_1, \sigma_1^2)$

Let $Y_{21} \ldots Y_{2,n_2}$ denote a random sample of size n_2 from $N(\mu_2, \sigma_2^2)$

Assume that the r.v's Y_{1i}'s and Y_{2j}'s are independent.

Then, $\left[\frac{S_1^2}{S_2} \frac{1}{F_{\frac{\alpha}{2}}(n_1-1, n_2-1)}, \frac{S_1^2}{S_2^2} F_{\frac{\alpha}{2}}(n_2-1, n_1-1)\right]$

is a $1-\alpha$ confidence interval for $\frac{\sigma_1^2}{\sigma_2^2}$

Since $\dfrac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \sim F(n_1-1, n_2-1)$

$$1-\alpha = P\left(F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1) \le \dfrac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \le F_{\frac{\alpha}{2}}(n_1-1, n_2-1)\right)$$

$$= P\left(\frac{S_1^2}{S_2^2} \frac{1}{F_{\frac{\alpha}{2}}(n_1-1, n_2-1)} \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{S_1^2}{S_2^2} \frac{1}{F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1)}\right)$$

So, $\left[\frac{S_1^2}{S_2^2} \frac{1}{F_{\frac{\alpha}{2}}(n_1-1, n_2-1)}, \frac{S_1^2}{S_2^2} \frac{1}{F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1)}\right]$ is $1-\alpha$ confidence interval for $\frac{\sigma_1^2}{\sigma_2^2}$

Now $\frac{1}{F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1)} = F_{\frac{\alpha}{2}}(n_1-1, n_2-1)$

because

$$P\left(F(n_1-1, n_2-1) \ge \frac{1}{F_{1-\frac{\alpha}{2}}(n_2-1, n_1-1)}\right) = P\left(F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1) \ge F(n_2-1, n_1-1)\right) = \frac{\alpha}{2}$$

8.108 A pharmaceutical manufacturer purchases raw material from two different suppliers. The mean level of impurities is approximately the same for both suppliers, but the manufacturer is concerned about the variability in the amount of impurities from supplier to supplier to supplier. To compare the variation in percentage impurities for the two suppliers the manufacturer selects ten shipments from each supplier and measures the percentage of impurities in each shipment. The sample variances were $s_1^2 = 0.273$ and $s_2^2 = 0.094$, respectively. Form a 95% confidence interval for the ratio of the true population variances.

$$\frac{\frac{s_1^2}{\sigma_1^2}}{\frac{s_2^2}{\sigma_2^2}} \sim F(n-1, m-1) \sim F(9,9)$$ Find a and b such that.

$0.95 = P(a \le F(9,9) \le b)$

$0.025 = P(b \le F(9,9))$ $\quad b = 4.03$

$0.975 = P(a \le F(9,9)) = P(F(9,9) \le \frac{1}{a})$

$0.025 = P(F(9,9) \ge \frac{1}{a})$ $\quad \frac{1}{a} = 4.03$ $\quad a = \frac{1}{4.03}$

The 95% confidence interval is $$\frac{1}{4.03} \le \frac{\frac{s_1^2}{\sigma_1^2}}{\frac{s_2^2}{\sigma_2^2}} \le 4.03$$

$$\frac{1}{4.03}\frac{s_1^2}{s_2^2} \le \frac{\sigma_1^2}{\sigma_2^2} \le 4.03\,\frac{s_1^2}{s_2^2}$$

$\Vert$

$\frac{1}{4.03}\,\frac{0.273}{0.094}$

$\Vert$

$0.7206 \le \frac{\sigma_1^2}{\sigma_2^2} \le 11.7041$