8.5 Confidence intervals for two samples

In this section, we look at how to obtain confidence intervals for parameters involving two independent samples.

Two normal means (where the true variances are known)

Suppose a random sample of size n_1 from a $N(\mu_1, \sigma_2^2)$ distribution has sample mean $\bar{X}_1$. Then:

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

Similarly, an independent random sample of size n_2 from a $N(\mu_2, \sigma_2^2)$ distribution with sample mean $\bar{X}_2$ will have:

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

Recall from Chapter 6, that subtracting two independent normal random variables produces another normal random variable. Here we have:

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

This result also holds for large samples from any distribution.

Standardising this gives:

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

This is our pivotal quantity for obtaining a confidence interval for $\mu_1 - \mu_2$, the difference between the means of two distributions.

Example 8.10

The heights (in inches) of two groups of people were recorded. The results are shown in the table below:

	A	B
Sample size	25	11
Sample mean	68	71
Sample standard deviation	3	2.5

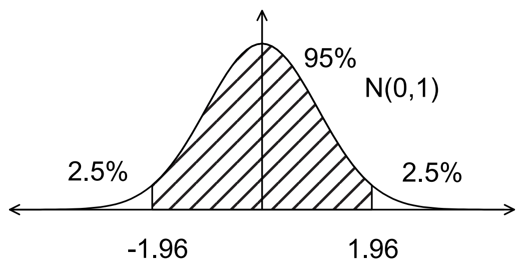
The population standard deviations of groups A and B were 2.5 and 2.1, respectively. Determine an equal-tailed 95% confidence interval for $\mu_A - \mu_B$, the difference in the means of the two populations.

Solution

Since we are calculating a confidence interval for the difference between two means, $\mu_A - \mu_B$, and we know the population variances, our pivotal quantity is:

$$\frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} \sim N(0,1)$$

For a 95% equal-tailed confidence interval, we require a probability of 2.5% on either side of the central region. From page 162 of the *Tables* we see that $P(Z > 1.9600) = 2.5\%$. Since the $N(0,1)$ distribution is symmetrical about zero our critical values are ± 1.96:



We have:

$$P\left(-1.96 < \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} < 1.96 \right) = 0.95$$

So the end points of the interval based on our particular sample can be found by rearranging:

$$\frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} = \pm 1.96$$

$$(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B) = \pm 1.96\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$$

$$(\mu_A - \mu_B) = (\bar{x}_A - \bar{x}_B) \mp 1.96\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}$$

We are told that $\sigma_A = 2.5$ and $\sigma_B = 2.1$. Our sample statistics are $\bar{x}_A = 68$, $\bar{x}_B = 71$, $n_A = 25$ and $n_B = 11$, so this gives:

$$(\mu_A - \mu_B) = (68 - 71) \mp 1.96\sqrt{\frac{2.5^2}{25} + \frac{2.1^2}{11}}$$
$$= -3 \mp 1.581$$

So our 95% confidence interval for $\mu_A - \mu_B$ to 3SF is $(-4.58, -1.42)$. ♦♦

Since both end points are negative it appears that $\mu_A - \mu_B < 0$, *ie* $\mu_A < \mu_B$.

Confidence interval for the difference in means $\mu_1 - \mu_2$ (σ^2's known)

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

This gives an equal-tailed $100(1-\alpha)\%$ confidence interval of $(\bar{x}_1 - \bar{x}_2) \mp z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$.

This result holds for samples from a normal distributions, or large samples from any distribution.

Two normal means (where the true variances are unknown)

It is more likely that we will not know the population variances (especially since we don't know the population means). So we need a pivotal quantity for $\mu_1 - \mu_2$ that involves only the sample variances S_1^2 and S_2^2.

Unfortunately, simply replacing σ_1^2 and σ_2^2 in the pivotal quantity in the previous section with the sample variances S_1^2 and S_2^2 doesn't correspond to any standard distribution. To get a standard distribution we need to make the assumption that $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Under this assumption:

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_P\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t_{n_1+n_2-2}$$

where S_P^2 is the pooled sample variance estimator:

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

Example 8.11

The heights (in inches) of two groups of people were recorded. The results are shown in the table below:

	A	B
Sample size	25	11
Sample mean	68	71
Sample standard deviation	3	2.5

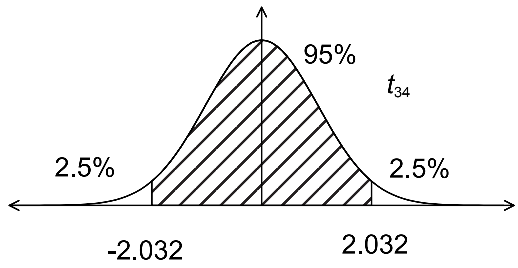
Assuming the population standard deviations of groups A and B are equal, determine a 95% confidence interval for $\mu_A - \mu_B$, the difference in the means of the two populations.

Solution

Since, we are finding a confidence interval for the difference between two means, $\mu_A - \mu_B$, and the population variances are unknown, but assumed equal, our pivotal quantity is:

$$\frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{S_P\sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} \sim t_{n_A + n_B - 2}$$

For a 95% equal-tailed confidence interval, we require a probability of 2.5% on either side of the central region. From page 163 of the *Tables* we see that $P(t_{34} > 2.032) = 2.5\%$. Since the t_{34} distribution is symmetrical about zero our critical values are ± 2.032:



We have:

$$P\left(-2.032 < \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{S_P\sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} < 2.032\right) = 0.95$$

So the end points of the interval based on our particular sample can be found by rearranging:

$$\frac{(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B)}{s_P\sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} = \pm 2.032$$

$$(\bar{x}_A - \bar{x}_B) - (\mu_A - \mu_B) = \pm 2.032 \times s_P\sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}$$

$$(\mu_A - \mu_B) = (\bar{x}_A - \bar{x}_B) \mp 2.032 \times s_P\sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}$$

Since we have $n_A = 25$, $n_B = 11$, $s_A = 3$ and $s_B = 2.5$, our pooled variance estimate is:

$$s_P^2 = \frac{1}{25 + 11 - 2}\left\{(25-1) \times 3^2 + (11-1) \times 2.5^2\right\} = 8.19118$$

Substituting this together with our sample means of $\bar{x}_A = 68$ and $\bar{x}_B = 71$ gives:

$$(\mu_A - \mu_B) = (\bar{x}_A - \bar{x}_B) \mp 2.032 \times s_P\sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}$$

$$= (68 - 71) \mp 2.032\sqrt{8.19118}\sqrt{\frac{1}{25} + \frac{1}{11}}$$

$$= -3 \mp 2.104$$