Confidence Intervals for Two Means

We derive confidence intervals for the difference in two population means, $\mu_1 - \mu_2$, under three circumstances:

- when the populations are independent and normally distributed with a common variance σ^2
- when the populations are independent and normally distributed with unequal variances
- when the populations are dependent and normally distributed

Two-Sample Pooled t-Interval

- If $X_1, X_2, \ldots, X_n \sim N(\mu_X, \sigma^2)$ and $Y_1, Y_2, \ldots, Y_m \sim N(\mu_Y, \sigma^2)$ are independent random samples, then a $(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$, the difference in the population means is:

$$(\bar{X} - \bar{Y}) \pm (t_{\alpha/2, n+m-2}) S_p \sqrt{\frac{1}{n} + \frac{1}{m}}$$

where S_p^2 is the **pooled sample variance**

$$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}$$

is an unbiased estimator of the common variance σ^2.

Example

The feeding habits of two species of net-casting spiders are studied. The species, the deinopis and menneus, coexist in eastern Australia. The following data were obtained on the size, in millimeters, of the prey of random samples of the two species:

- Size of Random Pray Samples of the Deinopis Spider in Millimeters n = 10

$$12.9, 10.2, 7.4, 7.0, 10.5, 11.9, 7.1, 9.9, 14.4, 11.3$$

- Size of Random Pray Samples of the Menneus Spider in Millimeters m = 11

$$10.2, 6.9, 10.9, 11.0, 10.1, 5.3, 7.5, 10.3, 9.2, 8.8, 7.0$$

What is the difference, if any, in the mean size of the prey (of the entire populations) of the two species?

Example (cont'd)

- $n = 10$ and $m = 11$; $\sigma_x^2 = 6.32$ and $\sigma_y^2 = 3.61$;
- the pooled sample variance is given by

$$s_p^2 = \frac{(10-1)6.32 + (11-1)3.61}{10+11-2} = 4.89$$

  which leads to a pooled standard deviation of 2.21:

$$s_p = \sqrt{4.89} = 2.21;$$

- the sample means are calculated to be

$$\bar{x} = 10.26 \text{ and } \bar{y} = 8.84;$$

- then the 95% confidence interval for the difference in the population means is given by

$$(10.26 - 8.84) \pm 2.09 \times 2.21\sqrt{\frac{1}{10} + \frac{1}{11}}$$

$$(-0.59, 3.44)$$

  That is, we can be 95% confident that the actual mean difference in the size of the prey is between -0.59 mm and 3.44 mm.

Two-Sample t-Interval (different variances)

- If $X_1, X_2, \ldots, X_n \sim N(\mu_X, \sigma_X^2)$ and $Y_1, Y_2, \ldots, Y_m \sim N(\mu_Y, \sigma_Y^2)$ are independent random samples
- we are not able to get the CI because we do not know the exact distribution. It is possible only if n and m are large and we can use the normal distribution, then a $(1-\alpha)\%$ confidence interval for $\mu_X - \mu_Y$ is given by

$$(\bar{X} - \bar{Y}) \pm (z_{\alpha/2}) \sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}$$

where S_X^2 and S_Y^2 are the **sample variances** for X and Y.

Confidence intervals for the difference in two population proportions

- $X_1, X_2, \ldots, X_{n_1} \sim Be(p_1)$ and $Y_1, Y_2, \ldots, Y_{n_2} \sim Be(p_2)$ are independent random samples
- For large random samples, an (approximate) $(1-\alpha)100\%$ confidence interval for $p_1 - p_2$, the difference in two population proportions, is:

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Example

A social experiment conducted in 1962 involved three- and four-year-old children from poverty-level families in Ypsilanti, Michigan. The children were randomly assigned either to:

1. A treatment group receiving two years of preschool instruction
2. A control group receiving no preschool instruction.

The participants were followed into their adult years. Here is a summary of the data:

	Arrested for some crime	
	Yes	No
Control	32	30
Preschool	19	42

Find a 95% confidence interval for $p_1 - p_2$, the difference in the two population proportions.

Example (cont'd)

- Of the $n_1 = 62$ children serving as the control group, 32 were later arrested for some crime, yielding a sample proportion of:

$$\hat{p}_1 = 0.516 \qquad 32/62$$

- and, of the $n_2 = 61$ children receiving preschool instruction, 19 were later arrested for some crime, yielding a sample proportion of:

$$\hat{p}_2 = 0.311 \qquad 19/61$$

- A 95% confidence interval for $p_1 - p_2$ is therefore:

$$(0.516 - 0.311) \pm 1.96\sqrt{\frac{0.516 \times 0.484}{62} + \frac{0.311 \times 0.689}{61}}$$

which simplifies to:

$$0.205 \pm 0.170 = (0.035, 0.375)$$

We can be 95% confident that between 3.5% and 37.5% more children not having attended preschool were arrested for a crime by age 19 than children who had received preschool instruction.