Section 7.5
Confidence Interval for the Difference of Two Means

A confidence interval for two population means is used when you have two independent random samples and you wish to make a comparison of the difference ($\mu_1 - \mu_2$).

The assumptions that need to be satisfied are:

1. Both samples must be independent SRSs from the populations of interest.
2. Both sets of data must come from normally distributed populations. If this is not the case or if we are unsure whether the population is normally distributed, the sampling distributions of $\bar{x}_1$ and $\bar{x}_2$ must be normally distributed. (Recall from section 4.4 that we can assume that the sampling distribution of $\bar{x}$ is normal for values of *n* greater than 30.)

When the population standard deviations are known, we use the formula $\left(\bar{x}_1 - \bar{x}_2\right) \pm z^* \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ and when it is unknown, we will need to find the sample standard deviations, s_1 and s_2, and use the formula $\left(\bar{x}_1 - \bar{x}_2\right) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ where t^* is the *t*-critical value based on the smaller of $n_1 - 1$ or $n_2 - 1$ degrees of freedom.

Example 1: The height (in inches) of men at UH is assumed to have a normal distribution with a standard deviation of 3.6 inches. The height (in inches) of women at UH is also assumed to have a normal distribution with a standard deviation of 2.9 inches. A random sample of 49 men and 38 women yielded respective means of 68.3 inches and 64.6 inches. Find the 90% confidence interval for the difference in the heights of men at UH and women at UH.

Let's summarize:

Men

$n_1 = 49$

$\sigma_1 = 3.6$

$\bar{x}_1 = 68.3$

Women

$n_2 = 38$

$\sigma_2 = 2.9$

$\bar{x}_2 = 64.6$

Now, do we use z^* or t^*?

Why? population s.d. is given

$z^* = \text{qnorm}\left(\frac{1+.9}{2}\right) = 1.6448$

Apply: $\left(\bar{x}_1 - \bar{x}_2\right) \pm z^* \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

$$(68.3 - 64.6) \pm 1.6448 \sqrt{\frac{3.6^2}{49} + \frac{2.9^2}{38}}$$

90% Confidence Interval: [2.55 , 4.85]

Interpretation: We are 90% confident that the true difference in the mean heights of men and women at UH is between 2.55 and 4.85 inches.

Example 2: A researcher wants to see if birds that build larger nests lay larger eggs. He selects two random samples of nests: one of small nests and the other of large nests. He weighs one egg from each nest. The data are summarized below:

	Small nests	Large nests
Sample size	60 = n_1	159 = n_2
Sample mean (g)	37.2 = $\bar{x}_1$	35.6 = $\bar{x}_2$
Sample variance	24.7 = s_1^2	39.0 = s_2^2

Find the 95% confidence interval for the difference between the average mass of eggs in small and large nests.

Finally, here we can assume normality since each sample size is greater than 30. This is from Section 4.4, the Central Limit Theorem.

Now, do we use z^* or t^*? Why? population s.d. is unknown

Also, when we have a set of samples, to find the degrees of freedom we will always use the smaller sample size. So for this problem, df = 60 – 1 = 59. There are other ways to find df, but we'll always go with the smaller sample size.

So, $t^* = qt\left(\frac{1.95}{2}, 59\right) = 2.00$

$s_1^2 = \text{Var}$

Now apply: $\left(\overline{x}_1 - \overline{x}_2\right) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Here we could do $\left(\overline{x_2} - \overline{x_1}\right)$*, as long as you indicate it.*

$$(37.2 - 35.6) \pm 2\sqrt{\frac{24.7}{60} + \frac{39}{159}}$$

95% Confidence Interval: $[-0.021, 3.221]$

We are 95% confident that the true difference between the average mass of eggs in small and large nests is between −0.021 and 3.221.