MATH 1342

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 $\overline{x}_1$ and $\overline{x}_2$ must be normally distributed. (Recall from section 4.4 that we can assume that the sampling distribution of $\overline{x}$ is normal for values of n greater than 30.)

On your quizzes, group 1 will be the group that you were provided information about first.

Calculations:

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.

If the t-test is used, the smaller of your two sample sizes will be used to calculate your degree of freedom.

Confidence Intervals

General formula: *statistic* ± *margin of error*

One-sample z-test: $\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$

Two-proportion z-test: $(\hat{p}_1 - \hat{p}_2) \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

One-sample t-test: $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$

One-proportion z-test: $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

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

Two-sample t-test: $(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Example:

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.

Group 1: Men
Mean: 68.3
(pop) stnd dev: 3.6
n = 49

CL : 90%

z-test (since we have known population standard deviations)

Group 2: Women
Mean: 64.6
(pop) stnd dev: 2.9
n = 38

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

```
> (68.3-64.6)-qnorm(1.90/2)*sqrt(3.6^2/49+2.9^2/38)
[1] 2.553541
> (68.3-64.6)+qnorm(1.90/2)*sqrt(3.6^2/49+2.9^2/38)
[1] 4.846459
```

[2.55, 4.85]
We are 90% confident that the difference in heights between men and women and UH is between 2.55 and 4.85 inches.

Example:

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:

Use a 95% confidence level

	Small nests	Large nests
Sample size	60	159
Sample mean (g)	37.2	35.6
Sample variance	24.7	39.0

s^2 →

Group 1 (small nests):
mean: 37.2
s^2 : 24.7
n = 60

Group 2 (large nests):
mean: 35.6
s^2: 39.0
n = 159

Since we are provided variances from our samples (rather than our populations) we must use a t-test (with 59 degrees of freedom…smaller sample size minus one)

Workspace: $(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

```
> (37.2-35.6)-qt(1.95/2,59)*sqrt(24.7/60+39.0/159)
[1] -0.02185543
> (37.2-35.6)+qt(1.95/2,59)*sqrt(24.7/60+39.0/159)
[1] 3.221855
>
```

(smaller sample size) - 1

[-0.02, 3.22]

We are 95% confident that the difference in weight between eggs from small nests and large nests is between -0.02g and 3.22g.

Popper 26:

A study was conducted to see if males or females purchased more expensive cars. A simple random sample of 25 males was surveyed with a mean car cost of $28,000 with a standard deviation of $73. A simple random sample of 20 women was surveyed with a mean car cost of $26,500 with a standard deviation of $120. Determine, with a confidence level of 90%, the interval of the differences between their car costs.

1. What is the difference of mean car cost for the two groups?

   a. 28000 b. 26500 c. 47 d. 1500

2. How many degrees of freedom should be used here? $n_1 = 25$, $n_2 = 20$

   a. 20 b. 19 c. 25 d. 24

3. What is the t* value needed?

   a. 1.7291 b. 1.7247 c. 1.3277 d. 1.2816

4. What is the margin of error for the difference in car cost?

   a. 11.538 b. 38.928 c. 5.164 d. 52. 821

5. What is the confidence interval of the difference between these two groups?

   a. [1447.18, 1552.82] b. [1095.73, 1551.7] c. [26500,28000] d. [1481,1519]

```
> 28000-26500
[1] 1500
> qt(1.90/2,19)
[1] 1.729133
> qt(1.90/2,19)*sqrt(73^2/25+120^2/20)
[1] 52.82097
> 1500-qt(1.90/2,19)*sqrt(73^2/25+120^2/20)
[1] 1447.179
> 1500+qt(1.90/2,19)*sqrt(73^2/25+120^2/20)
[1] 1552.821
```