Stat/Math 415 Homework 4

Due Friday Oct 4; Joseph Sepich (jps6444)

1 Problem 7.2-3

We have the following results

- n = 12
- $\overline{x} = 65.7$ inches
- $s_x = 4$ inches
- m = 15
- $\overline{y} = 68.2$ inches
- $s_y = 3$ inches

Find approximate confidence interval for 98% confidence for the difference of means $\mu_x - \mu_y$. Assume population variances are equivalent.

The problem descrption gives us an $\alpha = 0.02$ and calls for using formula 2 with sample pooled variance.

$$S_p = \sqrt{\frac{(n-1)s_x^2}{n+m-2} + \frac{(m-1)s_y^2}{n+m-2}}$$

$$S_p = \sqrt{\frac{(12-1)4^2}{12+15-2} + \frac{(15-1)3^2}{12+15-2}}$$

$$S_p = \sqrt{\frac{11 * 16 + 14 * 9}{25}} \approx 3.4756$$

The formula we want to use is:

$$(\overline{x} - \overline{y}) + -t_{\alpha/2} s_p \sqrt{\frac{1}{n} + \frac{1}{m}}$$

So we look in the table and see that for a t distribution with n + m -2 = 25 degrees of freedom for 0.02 / 2 = 0.01 as our alpha / 2 value we get $t_{\alpha/2} = 2.485$.

Plugging our values into our formula we have:

$$(65.7 - 68.2) + -2.485 * 3.4756 \sqrt{\frac{1}{12} + \frac{1}{15}}$$

This gives us the bounds **[-2.5 - 3.3450, -2.5 + 3.3450] = [-5.8450, 0.8450]** This would be the 98% confidence interval.

2 Problem 7.2-6

2.1 Part a

In order to finda 95% confidence interval for the difference in means, we must first calculate our summary statistics for both sample mean and sample variance.

$$\overline{x} = \frac{3.26 + 2.26 + 2.62 + 2.62 + 2.36 + 3.00 + 2.62 + 2.40 + 2.30 + 2.40}{10} = 2.584$$
$$\overline{y} = \frac{1.80 + 1.46 + 1.54 + 1.42 + 1.32 + 1.56 + 1.36 + 1.64 + 2.00 + 1.54}{10} = 1.564$$

To see if we can use t distribution instead of approximation with normal let's check for similar sample varaince:

$$s_x^2 = \frac{(3.26 - 2.584)^2 + (2.26 - 2.584)^2 + ... + (2.30 - 2.584)^2 + (2.40 - 2.584)^2}{9} \approx 0.10414$$
$$x_y^2 = \frac{(1.80 - 1.564)^2 + (1.46 - 1.564)^2 + ... + (2.00 - 1.564)^2 + (1.54 - 1.564)^2}{9} \approx 0.04281$$

Since both variances are similar let's use the following formula:

$$(\overline{x} - \overline{y}) + -t_{\alpha/2} s_p \sqrt{\frac{1}{n} + \frac{1}{m}}$$

$$S_p = \sqrt{\frac{(n-1)s_x^2}{n+m-2} + \frac{(m-1)s_y^2}{n+m-2}}$$
$$S_p = \sqrt{\frac{(10-1)0.10414}{10+10-2} + \frac{(10-1)0.04281}{10+10-2}} \approx 0.2712$$

Looking at the t distribution table for 10 + 10 - 2 = 18 degrees of freedom for $\alpha/2 = 0.025$ we get t = 2.101.

Plugging our values into our formula we have:

$$(2.584 - 1.564) + -2.101 * 0.2712 \sqrt{\frac{1}{10} + \frac{1}{10}}$$

This gives us the bounds [**1.02 - 0.2548, 1.02 + 0.2548**] = [**0.7652, 1.2748**] This would be the 95% confidence interval.

2.2 Part c

The wedge is necessary to do its job. If the wedge was not necessary then the force required to remove the seal with and without the seal would be the same, so the difference in means would be 0. Our 95% confidence interval for the difference in means does not contain 0, so we can say with 95% confidence the the wedge is necessary and is doing its job.

3 Problem 7.2-9

3.1 Part a

To find the 90% confidence interval for the mean in difference of percentages of males we must used a paired difference in means approach. To do this we must make the proper calculations for $D = x_1 - x_2$ where x_2 is post program for males and x_1 is pre program for males.

1. 11.10 - 9.97 = 1.13
2. 19.50 - 15.80 = 3.70
3. 14.00 - 13.02 = 0.98
4. 8.30 - 9.28 = -0.98
5. 12.40 - 11.51 = 0.89
6. 7.89 - 7.40 = 0.49
7. 12.10 - 10.70 = 1.40
8. 8.30 - 10.40 = -2.10
9. 12.31 - 11.40 = 0.91
10. 10.00 - 11.95 = -1.95

From here we can then use the following equation:

$$\overline{D} + -t_{\alpha/2}\frac{s_D}{\sqrt{n}}$$

So let's calculate the sample mean and sample standard deviation:

$$\overline{D} = \frac{1.13 + 3.7 + ... + 0.91 - 1.95}{10} \approx 0.447$$

$$s_D = \sqrt{\frac{(1.13 - 0.447)^2 + (3.7 - 0.447)^2 + ... + (0.91 - 0.447)^2 + (-1.95 - 0.447)^2}{9}} \approx 1.7296$$

We also can find from the distribution chart for alpha of 0.05 to use the value t = 1.833 with 9 degrees of freedom. Plugging our values in we get.

$$0.447 + -1.833\frac{1.7296}{\sqrt{10}}$$

This gives us the bounds **[0.447 - 1.0026, 0.447 + 1.0026] = [-0.5556, 1.4496]** This would be the 90% confidence interval.

3.2 Part b

To find the 90% confidence interval for the mean in difference of percentages of females we must used a paired difference in means approach. To do this we must make the proper calculations for $D = x_1 - x_2$ where x_2 is post program for females and x_1 is pre program for females.

1. 22.90 - 22.89 = 0.01
2. 31.60 - 33.47 = -1.87
3. 27.70 - 25.75 = 1.95
4. 21.70 - 19.80 = 1.90
5. 19.36 - 18.00 = 1.36

6. 25.03 - 22.33 = 2.7
7. 26.90 - 25.26 = 1.64
8. 25.75 - 24.90 = 0.85
9. 23.63 - 21.80 = 1.83
10. 25.06 - 24.28 = 0.78

From here we can then use the following equation:

$$\overline{D} + -t_{\alpha/2}\frac{s_D}{\sqrt{n}}$$

So let's calculate the sample mean and sample standard deviation:

$$\overline{D} = \frac{0.01 - 1.87 + ... + 1.83 + 0.78}{10} \approx 1.115$$

$$s_D = \sqrt{\frac{(0.01 - 1.115)^2 + (-1.87 - 1.115)^2 + ... + (1.83 - 1.115)^2 + (0.78 - 1.115)^2}{9}} \approx 1.2904$$

We also can find from the distribution chart for alpha of 0.05 to use the value t = 1.833 with 9 degrees of freedom. Plugging our values in we get.

$$1.115 + -1.833\frac{1.2904}{\sqrt{10}}$$

This gives us the bounds **[1.115 - 1.0026, 1.115 + 1.0026] = [0.3670, 1.8630]** This would be the 90% confidence interval.

3.3 Part c

According to the data we **cannot** say with 90% confidence that the percentages of body fat in males has decreased; however since 0 is not contained in our confidence interval for differences in females we can say with 90% confidence that there has been a decrease in percentages of body fat for them.

4 Problem 7.2-11

In order to find the 95% confidence interval for the difference in means we should use a normal distribution approximation, since n appears to be sufficiently large and the sample variations are clearly not equal. We would approximate X-Y to be a normal distribution with the mean $\mu_x - \mu_y$ and a variance of $\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}$, so our confidence interval equation would be:

$$\overline{x} - \overline{y} - z_\alpha\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}$$

The value of z in the normal distributoin table for an alpha of 0.05 is 1.645, so let's plug in our values:

$$671 - 480 - 1.645\sqrt{\frac{129^2}{60} + \frac{93^2}{60}}$$

$$191 - 1.645\sqrt{421.5}$$

This makes our lower bound interval **[191 - 33.77262, infinity] = [157.2274, infinity]**. This would be the 95% confidence interval.

5 Problem 7.3-3

5.1 Part a

Our sample proportion p hat is $\overline{x} = 167/330 = 0.5061$, the sum of yes answers over the sample size.

5.2 Part b

To create the 90% confidence interval we can use the variables $\alpha = 0.1, \overline{x} = 0.5061, n = 330$. The value for z of 0.05 is what we used in the last homework problem of 1.645. Then we can plug this int our equation:

$$\overline{x} + -z_{\alpha/2}\sqrt{\frac{\overline{x}(1-\overline{x})}{n}}$$

$$0.5061 + -1.645\sqrt{\frac{0.5061(1-0.5061)}{330}}$$

This makes our 90% confidence interval for p hat [**0.4608, 0.5514**]

6 Problem 7.3-5

6.1 Part a

The point estimate of p would be $\overline{x} = \frac{9}{50} = 0.18$, the sum of yes answers over the sample size.

6.2 Part b

95% confidence interval calls for an alpha of 0.05, so z of half alpha would be 1.96.

$$\frac{y}{n} + -z_{\alpha/2}\sqrt{\frac{(y/n)(1-y/n)}{n}}$$

$$0.18 + -1.96\sqrt{\frac{0.18(1-0.18)}{50}}$$

This makes our 95% confidence interval for p hat [**0.0735, 0.2865**]

6.3 Part c

$$\frac{\overline{x} + z_{\alpha/2}^2/(2n) + -z_{\alpha/2}\sqrt{\overline{x}(1-\overline{x}/n + z_{\alpha/2}^2/(4n^2))}}{\overline{x} + z_{\alpha/2}^2/n}$$

$$\frac{0.18 + 1.96^2/(100) + -1.96\sqrt{0.18(1-0.18)/50 + 1.96^2/(4*50^2))}}{0.18 + 1.96^2/50}$$

This makes our 95% confidence interval for p hat [**0.0648, 0.2751**]

6.4 Part d

$$p_0 = \frac{y+2}{n+4} = \frac{11}{54} \approx 0.2037$$

$$p_0 + -z_{\alpha/2}\sqrt{\frac{p_0(1-p_0)}{n+4}}$$

$$0.2037 + -1.96\sqrt{\frac{0.2037(1-0.2037)}{54}}$$

This makes our 95% confidence interval for p [**0.0963, 0.3111**]

7 Problem 7.3-9

7.1 Part a

We know the following information:

- $x_1 = 1009$
- $n_1 = 1230$
- $y_2 = 207$
- $n_2 = 340$

The point estimate for the first proportion would be $1009/1230 \approx 0.8203$ and the point estimate for the second proportion would be $207/340 = 0.6088$.

This makes the point estimate for $p_1 - p_2$ be $0.8203 - 0.6088 = 0.2115$.

7.2 Part b

We are again finding a 95% confidence interval so alpha over 2 is 0.025 making the value of z 1.96. The equation to estimate the interval would be:

$$(\overline{x}_1 - \overline{x}_2) + -1.96\sqrt{\frac{\overline{x}_1(1-\overline{x}_1)}{n_1} + \frac{\overline{x}_2(1-\overline{x}_2)}{n_2}}$$

$$(0.8203 - 0.6088) + -1.96\sqrt{\frac{0.8203(1-0.8203)}{1230} + \frac{0.6088(1-0.6088)}{340}}$$

This makes our 95% confidence interval for $p_1 - p_2$ to be [**0.1554, 0.2676**]