Note # 10: Statistical Inference for Numerical Variables

Problem 1. A statistics professor at a large community college wanted to determine if there was a difference in the means of the final exam scores between students who took his statistics course online and the students who took his course face-to-face. He randomly selected 30 students from his online course. The average exam score in this group was an 85.6, with a standard deviation of 4.3. He randomly selected 30 students from his face-to-face course. The average exam score in this group was an 83.2, with a standard deviation of 2.7. Create a 98% confidence interval for the difference between the average grade of the face-to-face group versus the online group.

- **a.** Construct the 98% confidence interval.
- **b.** Interpret your confidence interval from part a.

Answer: **Unpaired**

Group 1: Face-to-Face	**Group 2: Online**	**Critical value**
$\bar{x}_1 = \bar{x}_F = 83.2$	$\bar{x}_2 = \bar{x}_O = 85.6$	$CL: 98\%$
$s_1 = s_F = 2.7$	$s_2 = s_O = 4.3$	$df = \min(n_1 - 1, n_2 - 1)$
$n_1 = n_F = 30$	$n_2 = n_O = 30$	$df = \min(30 - 1, 30 - 1)$
		$df = \min(29, 29) = 29$
		$t^* = 2.462$

a. $(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = (83.2 - 85.6) \pm (2.462) \sqrt{\frac{2.7^2}{30} + \frac{4.3^2}{30}} = -2.4 \pm (2.462)(0.927)$

$$-2.4 \pm 2.28 \quad \rightarrow \quad \mathbf{98\%}\ \boldsymbol{CI}\text{: } (\mathbf{-4.68, -0.12})$$

b. We are **98%** confident that the true difference between the average exam score for face-to-face students and the average score for online students is between -4.68 points and -0.12 points.

Problem 2. A statistics professor at a large community college wanted to determine if there was a difference in the means of the final exam scores between students who took his statistics course online and the students who took his course face-to-face. He randomly selected 30 students from his online course. The average exam score in this group was an 85.6, with a standard deviation of 4.3. He randomly selected 30 students from his face-to-face course. The average exam score in this group was an 83.2, with a standard deviation of 2.7. Conduct a hypothesis test at the 0.02 significance level to test this.

- **a.** What are the hypotheses?
- **b.** What is the significance level?
- **c.** What is the value of the test statistic?
- **d.** What is the p-value?
- **e.** What is the correct decision?
- **f.** What is the appropriate conclusion/interpretation?
- **g.** Does the hypothesis test agree with the confidence interval from question 1?

Answer:

a. $H_0: \mu_{F2F} = \mu_{Online}$ vs. $H_A: \mu_{F2F} \neq \mu_{Online}$

b. $\alpha = 0.02$

c. $TS = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{83.2 - 85.6}{\sqrt{\frac{2.7^2}{30} + \frac{4.3^2}{30}}} = \frac{-2.4}{0.927} = \mathbf{-2.589}$

$TS = -2.589\,,\quad df = \min(30-1, 30-1) = df = \min(29, 29) = 29 \rightarrow\ 0.01 < p - value < 0.02$

d. **Two-sided**: $0.01 < p - value < 0.02$

e. $p - values < 0.02 < \alpha = 0.02,\ \therefore$ **Reject H_0**

f. The data **does** provide statistically significant evidence that there is a difference between the average final exam score for the face-to-face students and the average final exam score for the online students. Based on our samples, we think the average is higher for online students.

g. 98% CI: $(\mathbf{-4.68, -0.12})$. Yes, they agree because the null value (0) is not in the interval and we rejected H_0.

Problem 3. An instructor decided to create two different versions of the same exam, Version A and Version B. Prior to passing out the exams, he shuffled the exams together to ensure each student received a random version. Of the 30 students who took Version A, the average score was a 79.4, with a standard deviation of 14. Of the 27 students who took Version B, the average score was a 74.1, with a standard deviation of 20. Because he wants to ensure that the exam was fair to all students, he would like to evaluate whether the difference observed in the groups is so large that it provided convincing evidence that Version B was more difficult (on average) than Version A. Consider the students who took Version A to be Group A and the students who took Version B to be Group B. Create a 99% confidence interval for the difference between the average scores on Version A and Version B.

a. Construct the 90% confidence interval.

b. Interpret your confidence interval from part a.

Answer: Unpaired

Group 1: Version A	Group 2: Version B	Critical value
$\bar{x}_1 = \bar{x}_A = 79.4$	$\bar{x}_2 = \bar{x}_B = 74.1$	$CL: 99\%$
$s_1 = s_A = 14$	$s_2 = s_B = 20$	$df = \min(n_1 - 1, n_2 - 1)$
$n_1 = n_A = 30$	$n_2 = n_B = 27$	$df = \min(30 - 1, 27 - 1)$
		$df = \min(29, 26) = 26$
		$t^* = 2.779$

a. $(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = (79.4 - 74.1) \pm (2.779)\sqrt{\frac{14^2}{30} + \frac{20^2}{27}} = 5.3 \pm (2.779)(4.6204)$

$$5.3 \pm 12.84 \quad \rightarrow \quad \mathbf{99\%}\ \boldsymbol{CI}: \ (\mathbf{-7.54, 18.14})$$

c. We are **99%** confident that the true difference between the average score on version A and the average score on version B is between -7.54 points and 18.14 points.

Problem 4. An instructor decided to create two different versions of the same exam, Version A and Version B. Prior to passing out the exams, he shuffled the exams together to ensure each student received a random version. Of the 30 students who took Version A, the average score was a 79.4, with a standard deviation of 14. Of the 27 students who took Version B, the average score was a 74.1, with a standard deviation of 20. Because he wants to ensure that the exam was fair to all students, he would like to evaluate whether the difference observed in the groups is so large that it provided convincing evidence that Version B was more difficult (on average) than Version A. Consider the students who took Version A to be Group A and the students who took Version B to be Group B. Conduct a hypothesis test at the 0.01 significance level to test this.

- **a.** What are the hypotheses?
- **b.** What is the significance level?
- **c.** What is the value of the test statistic?
- **d.** What is the p-value?
- **e.** What is the correct decision?
- **f.** What is the appropriate conclusion/interpretation?
- **g.** Does the hypothesis test agree with the confidence interval from question 3?

Answer:

a. $H_0: \mu_A = \mu_B$ vs. $H_A: \mu_A > \mu_B$

b. $\alpha = 0.01$

c. $TS = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{79.4 - 74.1}{\sqrt{\frac{14^2}{30} + \frac{20^2}{27}}} = \frac{5.3}{4.6204} = \mathbf{1.1471}$

$TS = 1.1471\,,\quad df = \min(30-1, 27-1) = df = \min(29, 26) = 26 \rightarrow \ 0.10 < p - value < 0.15$

d. **One-sided**: $0.10 < p - value < 0.15$

e. $p - values > 0.10 > \ \alpha = 0.01,\ \therefore$ **Fail to reject H_0**

f. The data **<u>does not</u>** provide statistically significant evidence that the average score on version A is greater than the true average score on version B.

g. $99\%\ CI: (\mathbf{-7.54, 18.14})$. Yes, CI included the null value (0) and we failed to reject H_0.

Problem 5. Use the following ANOVA output to answer the questions below. Assume we are conducting this test at the 10% level.

Source	DF	Sum of Squares	Mean Squares	F Value	Prob
Groups	2	435.59259	217.7963	1.9211042	0.1569
Error (Residuals)	51	5781.8889	113.37037		
Total	53	6217.4815			

a. Based on the type of test that was run, what kind of variables are we studying?

b. How many groups were compared?

c. How many total observations were there?

d. What is the SSG? What is the SSE? What is the SST?

e. What is the MSG? What is the MSE?

f. What is the value of the test statistic?

g. What is the p-value?

h. What is the correct decision?

i. What is the appropriate conclusion/interpretation?

Answer:

a. One categorical variable (defines groups) and one numerical variable (average of each group).

b. $df_G = k - 1 \rightarrow k = df_G + 1 = 2 + 1 = 3$

c. $df_T = n - 1 \rightarrow n = df_T + 1 = 53 + 1 = 54$

d. $SSG = 435.59259,$ $SSE = 5781.8889,$ $SST = 6217.4815$

e. $MSG = 217.7963,$ $MSE = 113.37037$

f. Test Statistic $F = 1.92$

g. $p - value = 0.1569$

h. $0.1569 > 0.10 \rightarrow p - value > \alpha \rightarrow \boldsymbol{Fail\ to\ reject\ H_0}$

i. The data **does not** provide statistically significant evidence that there is an association between these two variables.

Problem 6. Use the following data set, and create and compute the values for the ANOVA table.

Group 1	Group 2	Group 3
11	7	15
10	8	17
11		16
$n_1 = 3$	$n_2 = 2$	$n_3 = 3$
$\bar{x}_1 = \frac{11+10+11}{3} = \frac{32}{3} = 10.67$	$\bar{x}_1 = \frac{7+8+}{2} = \frac{15}{2} = 7.5$	$\bar{x}_1 = \frac{15+17+16}{3} = \frac{48}{3} = 16$

ANOVA Table:

Source	DF	Sum of Squares	Mean Squares	F Value	Prob
Groups	**2**	$\mathbf{93.69}$	$\mathbf{46.845}$	$\mathbf{73.89}$	**0.000193**
Error (Residuals)	**5**	$\mathbf{3.188}$	$\mathbf{0.634}$		
Total	**7**	$\mathbf{96.878}$			

Answer:

$df_G = k - 1 = 3 - 1 = \mathbf{2}$

$df_T = n - 1 = 8 - 1 = \mathbf{7}$

$df_E = df_T - df_G = 7 - 2 = \mathbf{5}$

$$\bar{\boldsymbol{x}}_{\boldsymbol{Grand}} = \frac{11+10+11+7+8+15+17+16}{8} = \frac{95}{8} = \mathbf{11.875}$$

$$SSG = \sum n_i(\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{x}}_{\boldsymbol{Grand}})^2 = 3 \times (10.67 - 11.875)^2 + 2 \times (7.5 - 11.875)^2 + 3 \times (16 - 11.875)^2$$

$$= 4.36 + 38.28 + 51.05 = \mathbf{93.69}$$

$$SST = \sum (x_i - \bar{\boldsymbol{x}}_{\boldsymbol{Grand}})^2$$

$$= (11 - 11.875)^2 + (10 - 11.875)^2 + (11 - 11.875)^2 + (7 - 11.875)^2 + (8 - 11.875)^2$$

$$+ (15 - 11.875)^2 + (17 - 11.875)^2 + (16 - 11.875)^2$$

$$= 0.766 + 3.516 + 0.766 + 23.766 + 15.016 + 9.766 + 26.266 + 17.016 = \mathbf{96.878}$$

$$SSE = SST - SSG = 96.878 - 93.69 = \mathbf{3.188}$$

$MSG = \frac{SSG}{df_G} = \frac{\mathbf{93.69}}{\mathbf{2}} = \mathbf{46.845}$ and $MSE = \frac{SSE}{df_E} = \frac{\mathbf{3.188}}{\mathbf{5}} = \mathbf{0.634}$

$F = \frac{MSG}{MSE} = \frac{46.845}{0.634} = \mathbf{73.89}$

Problem 7. A teacher wants to know if test scores are different between his classes. Test this at the 5% level.

Source	DF	Sum of Squares	Mean Squares	F Value	Prob
Class (Groups)	3	3.798	**1.266**	**2.487**	0.066358
Error (Residuals)	**81**	41.228	**0.509**		
Total	84	**45.026**			

- **a.** What kind of test should we conduct?
- **b.** What are the hypotheses?
- **c.** What is the significance level?
- **d.** The ANOVA table above is partially filled in. Complete the missing spaces.
- **e.** How many different classes was the teacher comparing?
- **f.** How many total students were in those classes?
- **g.** What is the value of the test statistic?
- **h.** What is the p-value?
- **i.** What is the correct decision?
- **j.** What is the appropriate conclusion/interpretation?

Answer:

- **a.** ANOVA.
- **b.** $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ vs. $H_A: At\ least\ one\ mean\ is\ different.$
- **c.** $\alpha = 0.05$
- **d.** To complete the ANOVA table, we need to calculate:

  $df_E = df_T - df_G = 84 - 3 = \mathbf{81}$

  $SST = SSG + SSE = 3.798 + 41.228 = \mathbf{45.026}$

  $MSG = \frac{SSG}{df_G} = \frac{3.798}{3} = \mathbf{1.266}$ and $MSE = \frac{SSE}{df_E} = \frac{41.228}{81} = \mathbf{0.509}$

  $F = \frac{MSG}{MSE} = \frac{1.266}{0.509} = \mathbf{2.487}$
- **e.** $df_G = k - 1 \rightarrow k = df_G + 1 = 3 + 1 = \mathbf{4}$
- **f.** $df_T = n - 1 \rightarrow n = df_T + 1 = 84 + 1 = \mathbf{85}$
- **g.** Test Statistic $F = \mathbf{2.487}$
- **h.** $p - value = \mathbf{0.066358}$
- **i.** $0.066358 > 0.05 \rightarrow p - value > \alpha \rightarrow \boldsymbol{Fail\ to\ reject\ H_0}$
- **j.** The data **does not** provide statistically significant evidence that there is an association between class and test score.

Problem 8. A commuter thinks that the time it takes them to commute is different based on what day of the week it is. For a few weeks, they record their commute time each day, Monday through Friday. Test their claim at the 5% level.

Source	DF	Sum of Squares	Mean Squares	F Value	Prob
Day (Groups)	**4**	14.28	**3.57**	**3.37**	0.037122
Error (Residuals)	**15**	**15.92**	**1.06**		
Total	19	30.20			

a. What kind of test should we conduct?

b. What are the hypotheses?

c. What is the significance level?

d. The ANOVA table above is partially filled in. Complete the missing spaces.

e. What is the value of the test statistic?

f. What is the p-value?

g. What is the correct decision?

h. What is the appropriate conclusion/interpretation?

Answer:

a. ANOVA.

b. $H_0: \mu_M = \mu_{Tu} = \mu_W = \mu_{Th} = \mu_F$ vs. $H_A: At\ least\ one\ mean\ is\ different.$

c. $\alpha = 0.05$

d. To complete the ANOVA table, we need to calculate:

$df_G = k - 1 = 5 - 1 = \mathbf{4}$

$df_E = df_T - df_G = 19 - 4 = \mathbf{15}$

$SSE = SST - SSG = 30.2 - 14.28 = \mathbf{15.92}$

$MSG = \frac{SSG}{df_G} = \frac{\mathbf{14.28}}{\mathbf{4}} = \mathbf{3.57}$ and $MSE = \frac{SSE}{df_E} = \frac{\mathbf{15.92}}{\mathbf{15}} = \mathbf{1.06}$

$$F = \frac{MSG}{MSE} = \frac{3.57}{1.06} = \mathbf{3.37}$$

e. $F = \mathbf{3.37}$

f. $p - value = \mathbf{0.037122}$

g. $0.037122 < 0.05 \quad \rightarrow \boldsymbol{Reject\ H_0}$

h. The **data does** provide statistically significant evidence that commute time and day of the week are associated.