**Week #13: Final Exam Review**

This week, we will be reviewing problems from each of the chapters we have covered this semester. This review is not all-inclusive, rather, it is meant to remind you of some important topics we have covered throughout the semester. For information about the logistics, timing, and expectation of your particular exam, please contact your instructor.

**Problem 1.** Researchers studying the effect of antibiotic treatment for acute sinusitis compared to symptomatic treatments randomly assigned 166 adults diagnosed with acute sinusitis to one of two groups: treatment or control. Study participants received either a 10-day course of amoxicillin (an antibiotic) or a placebo similar in appearance and taste. The placebo consisted of symptomatic treatments such as acetaminophen, nasal decongestants, etc. At the end of the 10-day period, patients were asked if they experienced improvements in symptoms. The distribution of responses is shown below.

<table>
<thead>
<tr>
<th></th>
<th>Showed Improvement</th>
<th>Didn’t Show Improvement</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>66</td>
<td>19</td>
<td>85</td>
</tr>
<tr>
<td>Control</td>
<td>65</td>
<td>16</td>
<td>81</td>
</tr>
<tr>
<td>Total</td>
<td>131</td>
<td>35</td>
<td>166</td>
</tr>
</tbody>
</table>

a. What are the explanatory and response variables in this study? What type of variable is each?
b. What is the population?
c. What are the parameters and their values?
d. What is the sample?
e. What are the statistics and their values?
f. In which group did a higher percentage of patients experience improvement in symptoms?
g. Your findings so far might suggest a real difference in effectiveness of antibiotic and placebo treatments for improving symptoms of sinusitis. However, this is not the only possible conclusion that can be drawn based on your finds so far. What is one other possible explanation for the observed difference between the percentages of patients in the antibiotic and placebo treatment groups that experience improvement in symptoms of sinusitis?
a. explanatory variable: treatment type (antibiotic/placebo)  
categorical, nominal

Response variable: whether or not their symptoms improved  
categorical, nominal

b. All adults diagnosed with acute sinusitis

c. ① prop. of all adult acute sinusitis suffers on antibiotic who show improvement  
value: unknown, we can est. using sample statistic

② prop. of all adult acute sinusitis suffers on placebo who show improvement  
value: unknown, we can est. using sample statistic

d. 166 adult acute sinusitis sufferers in our survey

e. ① prop. of individuals in trt. group who showed improvement  
value: \( \hat{p}_t = \frac{66}{85} = 0.77650\)

② prop. of individuals in placebo group who showed improvement, value: \( \hat{p}_p = \frac{65}{81} = 0.8025\)
f. placebo/control group

g. random chance

SAMPLING variability
Problem 2. Anna teaches a small class of 14 students. She recently gave them an exam (out of 100 points). The scores on the exam were: 0, 15, 58, 75, 78, 82, 84, 84, 89, 91, 94, 98, 100, and 100.

a. What is the five number summary of this data set?

b. Are any of the scores potentially outliers?

\[
\text{min: } 0 \\
Q_1: 75 \\
\text{Med: } 84 \\
Q_3: 94 \\
\text{max: } 100
\]

\[
\text{IQR} = Q_3 - Q_1 = 94 - 75 = 19
\]

\[
\text{b. Smaller than } Q_1 - 1.5 \times \text{IQR} = 75 - (1.5)(19) = 46.5
\]

\[
\text{bigger than } Q_3 + 1.5 \times \text{IQR} = 94 + (1.5)(19) = 122.5
\]

Yes! 0 & 15 are potentially outliers.
Problem 3. At a certain gas station, 40% of the customers request regular gas, 35% request unleaded gas, and 25% request premium gas. Of those customers requesting regular gas, only 30% fill their tanks all the way up, while the remaining 70% only fill up part of their tank. Of those customers requesting unleaded gas, 60% fill their tanks all the way up, while of those requesting premium, 50% fill their tanks all the way up.

a. Draw a tree diagram to illustrate this scenario.
b. What is the probability that a person requests regular gas?
c. What is the probability that a person doesn’t fill the tank all the way up?
d. What is the probability that a person requests regular gas or doesn’t fill the tank all the way up?
e. What is the probability that a person requests regular gas and doesn’t fill the tank all the way up?
f. If the next customer does not fill the tank all the way up (only fills it up part of the way), what is the probability that they requested regular gas?
g. If the next customer requests regular gas, what is the probability that they don’t fill the tank all the way up (only fills it up part of the way)?
**Problem 3.** At a certain gas station 40% of the customers request regular gas, 35% request unleaded gas, and 25% request premium gas. Of those customers requesting regular gas, only 30% fill their tanks all the way up, while the remaining 70% only fill up part of their tank. Of those customers requesting unleaded gas, 60% fill their tanks all the way up, while of those requesting premium, 50% fill their tanks all the way up.

a. Draw a tree diagram to illustrate this scenario.

b. What is the probability that a person requests regular gas?

c. What is the probability that a person doesn’t fill the tank all the way up?

d. What is the probability that a person requests regular gas or doesn’t fill the tank all the way up?

e. What is the probability that a person requests regular gas and doesn’t fill the tank all the way up?

f. If the next customer does not fill the tank all the way up (only fills it up part of the way), what is the probability that they requested regular gas?

g. If the next customer requests regular gas, what is the probability that they don’t fill the tank all the way up (only fills it up part of the way)?

\[ \text{b. } P(\text{regular}) = 0.40 \]

\[ \text{c. } P(\text{partial}) = P(\text{reg} \cap \text{part}) + P(\text{unl} \cap \text{part}) + P(\text{prem} \cap \text{part}) \]
\[ = (0.40)(0.7) + (0.35)(0.4) + (0.25)(0.5) \]
\[ = 0.28 + 0.14 + 0.125 \]
\[ = 0.545 \]
Problem 3. At a certain gas station 40% of the customers request regular gas, 35% request unleaded gas, and 25% request premium gas. Of those customers requesting regular gas, only 30% fill their tanks all the way up, while the remaining 70% only fill up part of their tank. Of those customers requesting unleaded gas, 60% fill their tanks all the way up, while of those requesting premium, 50% fill their tanks all the way up.

a. Draw a tree diagram to illustrate this scenario.

b. What is the probability that a person requests regular gas?

c. What is the probability that a person doesn’t fill the tank all the way up?

d. What is the probability that a person requests regular gas or doesn’t fill the tank all the way up?

e. What is the probability that a person requests regular gas and doesn’t fill the tank all the way up?

f. If the next customer does not fill the tank all the way up (only fills it up part of the way), what is the probability that they requested regular gas?

g. If the next customer requests regular gas, what is the probability that they don’t fill the tank all the way up (only fills it up part of the way)?

d. \( P(\text{reg } \cup \text{ part}) = P(\text{reg}) + P(\text{part}) - P(\text{reg } \cap \text{ part}) \)

\[
= 0.4 + 0.545 - (0.4)(0.7)
\]

\[
= 0.4 + 0.545 - 0.28
\]

\[
= \frac{0.665}{0.665}
\]

e. \( P(\text{reg } \cap \text{ part}) = \frac{P(\text{part}|\text{reg})P(\text{reg})}{P(\text{reg})} \)

\[
= \frac{(0.7)(0.4)}{0.28}
\]
Problem 3. At a certain gas station 40% of the customers request regular gas, 35% request unleaded gas, and 25% request premium gas. Of those customers requesting regular gas, only 30% fill their tanks all the way up, while the remaining 70% only fill up part of their tank. Of those customers requesting unleaded gas, 60% fill their tanks all the way up, while of those requesting premium, 50% fill their tanks all the way up.

a. Draw a tree diagram to illustrate this scenario.

b. What is the probability that a person requests regular gas?

c. What is the probability that a person doesn’t fill the tank all the way up?

d. What is the probability that a person requests regular gas or doesn’t fill the tank all the way up?

e. What is the probability that a person requests regular gas and doesn’t fill the tank all the way up?

f. If the next customer does not fill the tank all the way up (only fills it up part of the way), what is the probability that they requested regular gas?

g. If the next customer requests regular gas, what is the probability that they don’t fill the tank all the way up (only fills it up part of the way)?
**Problem 4.** In the United States, the ages of smartphone users follows an approximately normal distribution with a mean of 36.9 years and a standard deviation of 13.9 years.

a. What is the probability that a randomly selected smartphone user is at most 50.8 years old?

b. What is the probability that a randomly selected smartphone user is older than 50.8 years old?

c. What is the probability that a randomly selected smartphone user is between 23 years old and 50.8 years old?

d. What value is the 80th percentile of this distribution? What does that mean?

e. Jenna is younger than 62% of smartphone users. How old is Jenna?
Problem 4. In the United States, the ages of smartphone users follow an approximately normal distribution with a mean of 36.9 years and a standard deviation of 13.9 years.

a. What is the probability that a randomly selected smartphone user is at most 50.8 years old?

b. What is the probability that a randomly selected smartphone user is older than 50.8 years old?

c. What is the probability that a randomly selected smartphone user is between 23 years old and 50.8 years old?

d. What value is the 80th percentile of this distribution? What does that mean?

e. Jenna is younger than 62% of smartphone users. How old is Jenna?
Problem 4. In the United States, the ages of smartphone users follows an approximately normal distribution with a mean of 36.9 years and a standard deviation of 13.9 years.

a. What is the probability that a randomly selected smartphone user is at most 50.8 years old?

b. What is the probability that a randomly selected smartphone user is older than 50.8 years old?

c. What is the probability that a randomly selected smartphone user is between 23 years old and 50.8 years old?

d. What value is the 80th percentile of this distribution? What does that mean?

e. Jenna is younger than 62% of smartphone users. How old is Jenna?

\[
\begin{align*}
P(23 < X < 50.8) &= P(-1.00 < Z < 1.00) \\
&= P(Z < 1.00) - P(Z < -1.00) \\
&= 0.8413 - 0.1587 = 0.6826
\end{align*}
\]
Problem 4. In the United States, the ages of smartphone users follows an approximately normal distribution with a mean of 36.9 years and a standard deviation of 13.9 years.

a. What is the probability that a randomly selected smartphone user is at most 50.8 years old?
b. What is the probability that a randomly selected smartphone user is older than 50.8 years old?
c. What is the probability that a randomly selected smartphone user is between 23 years old and 50.8 years old?
d. What value is the 80th percentile of this distribution? What does that mean?
e. Jenna is younger than 62% of smartphone users. How old is Jenna?
Problem 4. In the United States, the ages of smartphone users follows an approximately normal distribution with a mean of 36.9 years and a standard deviation of 13.9 years.

a. What is the probability that a randomly selected smartphone user is at most 50.8 years old?

b. What is the probability that a randomly selected smartphone user is older than 50.8 years old?

c. What is the probability that a randomly selected smartphone user is between 23 years old and 50.8 years old?

d. What value is the 80th percentile of this distribution? What does that mean?

e. Jenna is younger than 62% of smartphone users. How old is Jenna?
Problem 5. The mean number of minutes for app engagement by a tablet user is 8.2 minutes with a standard deviation of 1 minute. Suppose you take a sample of 60 tablet users and measure how long an app keeps them engaged.

a. Describe the sampling distribution.
b. Find the 90th percentile for the sample mean time for app engagement for a tablet user. What does this mean?
c. What is the probability that a specific user has an engagement time between 8 minutes and 8.5 minutes?
d. What is the probability that the sample mean is between 8 minutes and 8.5 minutes?

a. **Sampling Dist'n (of the sample mean)**
   \[ \text{mean} = \mu = 8.2 \]
   \[ \text{Stdev} = \sigma / \sqrt{n} = \frac{1}{\sqrt{60}} \approx 0.1291 \]
   \[ \text{shape} = \text{approx. normal} \]

b. **90th percentile**
   \[ z = 1.28 \]
   \[ 1.28 = \frac{X - 8.2}{0.1291} \]
   \[ X = (1.28)(0.1291) + 8.2 \]
   \[ X = 8.365 \text{ minutes} \]
   90% of samples (n = 60) will have an avg. engagement time less than 8.365 minutes.
Problem 5. The mean number of minutes for app engagement by a tablet user is 8.2 minutes with a standard deviation of 1 minute. Suppose you take a sample of 60 tablet users and measure how long an app keeps them engaged.

a. Describe the sampling distribution.
b. Find the 90th percentile for the sample mean time for app engagement for a tablet user. What does this mean?
c. What is the probability that a specific user has an engagement time between 8 minutes and 8.5 minutes?
d. What is the probability that the sample mean is between 8 minutes and 8.5 minutes?

\[
\begin{align*}
Z_A &= \frac{8.0 - 8.2}{1} = -0.2 = -0.20 \\
CP_A : P(X < 8.0) = P(Z < -0.20) = 0.4207 \\
Z_B &= \frac{8.5 - 8.2}{1} = 0.3 = 0.30 \\
CP_B : P(X < 8.5) = P(Z < 0.30) = 0.6179
\end{align*}
\]

\[
P(8 < X < 8.5) = P(-0.20 < Z < 0.30)
= P(Z < 0.30) - P(Z < -0.20)
= 0.6179 - 0.4207
= 0.1972
\]
**Problem 5.** The mean number of minutes for app engagement by a tablet user is 8.2 minutes with a standard deviation of 1 minute. Suppose you take a sample of 60 tablet users and measure how long an app keeps them engaged.

a. Describe the sampling distribution.
b. Find the 90th percentile for the sample mean time for app engagement for a tablet user. What does this mean?
c. What is the probability that a specific user has an engagement time between 8 minutes and 8.5 minutes?
d. What is the probability that the sample mean is between 8 minutes and 8.5 minutes?

\[
\begin{align*}
Z_A &= \frac{8.0 - 8.2}{0.1291} = \frac{-0.2}{0.1291} = -1.55 \\
CP_A : P(\bar{X} < 8.0) &= P(z < -1.55) = 0.0606 \\
Z_B &= \frac{8.5 - 8.2}{0.1291} = \frac{0.3}{0.1291} = 2.32 \\
CP_B : P(\bar{X} < 8.5) &= P(z < 2.32) = 0.9898 \\
\end{align*}
\]

\[
P(8 < \bar{X} < 8.5) = P(-1.55 < z < 2.32) \\
= P(z < 2.32) - P(z < -1.55) \\
= 0.9898 - 0.0606 \\
= 0.9292
\]
Problem 6. A student polls their school to see if students in the school district are for or against the new legislation regarding school uniforms. They survey 600 students and find that 480 are against the new legislation.

a. What is the 90% confidence interval?
b. Would you expect the 97% confidence interval to be wider or narrower?
c. What is the correct interpretation of your confidence interval in part a?
d. Are the assumptions met? Explain.

\[
\hat{p} = \frac{480}{600} = 0.80
\]

\[
z^* = 1.645
\]

\[
n = 600
\]

\[
0.80 \pm 1.645 \times \sqrt{\frac{0.80(1-0.80)}{600}}
\]

\[
0.80 \pm 1.645 \times \sqrt{\frac{0.80(0.20)}{600}}
\]

\[
0.80 \pm 0.02686
\]

\[
0.80 - 0.02686 = 0.77314
\]

\[
0.80 + 0.02686 = 0.82686
\]

90% CI: (0.77314, 0.82686)

b. As CL increases, \( z^* \) also increases.

If \( z^* \) increases, margin of error increases.

As margin of error increases, our interval gets wider.

We would expect a 97% CI to be wider than a 90% confidence interval.
Problem 6. A student polls their school to see if students in the school district are for or against the new legislation regarding school uniforms. They survey 600 students and find that 480 are against the new legislation.

a. What is the 90% confidence interval? 
   \[ 90\% CI: (0.77314, 0.82686) \]

b. Would you expect the 97% confidence interval to be wider or narrower?

c. What is the correct interpretation of your confidence interval in part a?

   We are 90% confident that the true proportion of all students in this school district who are against the new school uniform legislation is between 0.77314 and 0.82686.

   C. 1) Confidence level 2) Describe the population parameter 3) Confidence interval

   We are 90% confident that the true proportion of all students in this school district who are against the new school uniform legislation is between 0.77314 and 0.82686.

d. Are the assumptions met? Explain.

   d. 1) Independence
      - Random sample
      - \( n \) < 10% of pop'n

   2) Sample size
      - \( n \hat{p} \geq 10 \) \( \Rightarrow (600)(0.8) = 480 \geq 10 \)
      - \( n(1-\hat{p}) \geq 10 \) \( \Rightarrow (600)(0.2) = 120 \geq 10 \)
Problem 7. Jeffrey, as an eight-year old, has an established mean time of 16.43 seconds for swimming the 25-yard freestyle. His dad, Frank, thought that Jeffrey could swim the 25-yard freestyle faster using goggles. Frank bought Jeffrey a new pair of expensive goggles and time Jeffrey swimming the 25-yard freestyle 15 different times (with breaks in between). For these 15 swims, Jeffrey’s mean time was 16.0 seconds, with a standard deviation of 0.8 seconds. Frank thought that the goggles helped Jeffrey to swim faster than the 16.43 seconds. Conduct a hypothesis test to see if the data supports this claim at the 5% level. Assume that the swim times for the 25-yard freestyle are normally distributed.

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?

\[ \text{null value: } M_0 = 16.43 \]

\[ \begin{align*}
\text{a. } H_0: \mu &= 16.43 \\
\text{HA: } \mu &< 16.43 \\
\text{b. } \alpha &= 0.05 \\
\text{c. } TS &= \frac{\bar{x} - M_0}{S/\sqrt{n}} = \frac{16.0 - 16.43}{0.8/\sqrt{15}} = \frac{-0.43}{0.2006} = -2.08 \\
\end{align*} \]
Problem 7. Jeffrey, as an eight-year old, has an established mean time of 16.43 seconds for swimming the 25-yard freestyle. His dad, Frank, thought that Jeffrey could swim the 25-yard freestyle faster using goggles. Frank bought Jeffrey a new pair of expensive goggles and timed Jeffrey swimming the 25-yard freestyle 15 different times (with breaks in between). For these 15 swims, Jeffrey’s mean time was 16.0 seconds, with a standard deviation of 0.8 seconds. Frank thought that the goggles helped Jeffrey to swim faster than the 16.43 seconds. Conduct a hypothesis test to see if the data supports this claim at the 5% level. Assume that the swim times for the 25-yard freestyle are normally distributed.

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?

\[
\text{d. } T_S = -2.08 \quad \text{look up } |T_S| = |-2.08| = 2.08
\]
\[
df = n-1 = 15-1 = 14
\]
\[
0.025 < \text{p-value} < 0.05
\]

\[
\text{E. } \text{p-value} < 0.05
\]
\[
\text{p-value} < \alpha \quad \boxed{\text{Reject } H_0}
\]

f. The data does provide statistically significant evidence that Jeffrey’s true avg. swim time with the goggles is less than 16.43 seconds.
Problem 8. In a study of 400 type 2 diabetic patients, an investigator used a simple linear regression model to study the association between systolic blood pressure (response variable) and blood sugar (predictor variable). The estimated slope of the line was 2.5, with a standard error of 0.5. Does this study provide statistically significant evidence of a positive association at the 5% level?

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. What is the 98% confidence interval for the slope parameter?

\[
\begin{align*}
\text{a. } H_0 & : \beta_1 = 0 & \text{[null value = } 0] \\
\text{H}_A & : \beta_1 > 0
\end{align*}
\]

\[
\begin{align*}
\text{b. } \alpha = 0.05
\end{align*}
\]

\[
\begin{align*}
\text{c. } TS = \frac{\text{point estimate - null value}}{\text{std. error}} = \frac{2.5 - 0}{0.5} = \frac{2.5}{0.5} = 5.00
\end{align*}
\]

\[
\begin{align*}
\text{d. } TS = 5.00 \\
\text{df} = n-2 = 400-2 = 398 \text{ (use df = 100)} \\
\text{one-tailed } >
\end{align*}
\]

\[
\begin{align*}
\text{TS} > 3.390 \\
p-value < 0.0005
\end{align*}
\]
Problem 8. In a study of 400 type 2 diabetic patients, an investigator used a simple linear regression model to study the association between systolic blood pressure (response variable) and blood sugar (predictor variable). The estimated slope of the line was 2.5, with a standard error of 0.5. Does this study provide statistically significant evidence of a positive association at the 5% level?

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. What is the 98% confidence interval for the slope parameter?

\[ p\text{-value} < 0.0005 < 0.05 \]
\[ p\text{-value} < 0.05 \]
\[ p\text{-value} < \alpha \]
\[ \text{Reject } H_0 \]

f. The data does provide statistically significant evidence of a positive association between systolic blood pressure and blood sugar.
Problem 8. In a study of 400 type 2 diabetic patients, an investigator used a simple linear regression model to study the association between systolic blood pressure (response variable) and blood sugar (predictor variable). The estimated slope of the line was 2.5, with a standard error of 0.5. Does this study provide statistically significant evidence of a positive association at the 5% level?

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. What is the 98% confidence interval for the slope parameter?

\[
\text{point estimate} = b_1 = 2.5 \\
\text{std. error} = 0.5 \\
t^* = 2.364 \\
df = n-2 = 400-2 = 398 \\
\text{(use df=100)}
\]

\[
2.5 \pm (2.364)(0.5) \\
2.5 \pm 1.182 \\
2.5 - 1.182 = 1.318 \\
2.5 + 1.182 = 3.682 \\
98\% \text{ CI: (1.318, 3.682)}
\]