**Week #11: Data Collection and Statistical Inference for Numeric Variables**

**Problem 1.** The Literary Digest magazine conducted a poll to predict the result of the 1936 Presidential election between Franklin Roosevelt (Democrat and incumbent) and Alf Landon (Republican). At the time, the poll was famous, because they had correctly predicted three successive elections. In 1936 they mailed questionnaires to 10 million people and asked how they planned to vote. The sampling frame was constructed from telephone directories, country club memberships, and automobile registrations. Only 2.3 million of those contacted returned their questionnaire. Based on their responses, the Literary Digest predicted that Landon would win, getting 57% of the vote. Instead, Landon got only 36% of the vote and Roosevelt won in a landslide. What happened?

1. **Undercoverage**
   - **wanted:** sample to represent all reg voters
   - **what happened:** sampling wealthier pop'n

2. **Non-response**
   - **what is the difference between those who decided to respond and those who did not**
Problem 2. Gastric freezing is a treatment for ulcers in the upper intestine. In a study published many years ago, patients with ulcers swallowed a deflated balloon with tubes attached, and then a refrigerated liquid was pumped through the balloon for an hour. The idea is that cooling the stomach will reduce its production of acid and relieve ulcers. Results from this study showed that about 1/3 of the subjects improved. What is wrong with this study? How could this experiment be improved?

No control group!

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>93 patients</td>
<td>78 patients</td>
</tr>
<tr>
<td>Trt: Gastric Freezing</td>
<td>Control: Placebo</td>
</tr>
<tr>
<td>TVt: 34% improved</td>
<td>Control: 38% improved</td>
</tr>
</tbody>
</table>

\[\text{No control group!}\]
Problem 3. The Human Toxome Project (HTP) is working to understand the scope of industrial pollution in the human body. Industrial chemicals may enter the body through pollution or as ingredients in consumer products. In October 2008, scientists at HTP tested cord blood samples for 20 newborn infants in the United States. The cord blood of the "In utero/newborn" group was tested for 430 industrial compounds, pollutants, and other chemicals, including chemicals linked to brain and nervous system toxicity, immune system toxicity, and reproductive toxicity and fertility problems. The researchers recorded the number of these targeted chemicals that were found in each infant’s cord blood. The average number of targeted chemicals found was 127.45, with a standard deviation of 25.965. Create a \( \text{99%} \) confidence interval for the average number of targeted chemicals in all infants cord blood.

a. Construct the \( \text{99%} \) confidence interval.

\[
\bar{x} \pm t^* \left( \frac{s}{\sqrt{n}} \right)
\]

\[
\bar{x} = 127.45 \quad t^* = 2.861 \quad s = 25.965 \quad n = 20
\]

\[
127.45 \pm 2.861 \left( \frac{25.965}{\sqrt{20}} \right) = 127.45 \pm 10.038
\]

\[
127.45 - 10.038 = 117.412 \\
127.45 + 10.038 = 137.488
\]

\( \text{90% CI: } (117.412, 137.488) \)

b. We are \( \text{90%} \) confident that the true avg. # of targeted chemicals found in infants’ cord blood (US, 2008) is between 117.412 chemicals and 137.488 chemicals.
Problem 4. The Human Toxome Project (HTP) is working to understand the scope of industrial pollution in the human body. Industrial chemicals may enter the body through pollution or as ingredients in consumer products. In October 2008, scientists at HTP tested cord blood samples for 20 newborn infants in the United States. The cord blood of the "In utero/newborn" group was tested for 430 industrial compounds, pollutants, and other chemicals, including chemicals linked to brain and nervous system toxicity, immune system toxicity, and reproductive toxicity and fertility problems. The researchers recorded the number of these targeted chemicals that were found in each infant’s cord blood. The average number of targeted chemicals found was 127.45, with a standard deviation of 25.965. In 2000, the average number of targeted chemicals found in infant’s cord blood was 120. Conduct a hypothesis test at the 0.01 significance level to test if the average number has changed.

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?
h. How would the hypothesis test have changed if we wanted to see if the number of targeted chemicals had increased, instead of seeing if it had changed?
Problem 4. The Human Toxome Project (HTP) is working to understand the scope of industrial pollution in the human body. Industrial chemicals may enter the body through pollution or as ingredients in consumer products. In October 2008, scientists at HTP tested cord blood samples for 20 newborn infants in the United States. The cord blood of the "In utero/newborn" group was tested for 430 industrial compounds, pollutants, and other chemicals, including chemicals linked to brain and nervous system toxicity, immune system toxicity, and reproductive toxicity and fertility problems. The researchers recorded the number of these targeted chemicals that were found in each infant’s cord blood. The average number of targeted chemicals found was 127.45, with a standard deviation of 25.965. In 2000, the average number of targeted chemicals found in infant’s cord blood was 120. Conduct a hypothesis test at the 0.01 significance level to test if the average number has changed.

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?
h. How would the hypothesis test have changed if we wanted to see if the number of targeted chemicals had increased, instead of seeing if it had changed?

d. TS = 1.283
   df = n-1 = 20-1 = 19
   \( \text{two} \)

\[ p-value > 0.20 > 0.10 \]

\[ p-value > 0.10 \]

\[ p-value > \alpha \]

\[ \text{Fail to Rej. } H_0 \]
Problem 4. The Human Toxome Project (HTP) is working to understand the scope of industrial pollution in the human body. Industrial chemicals may enter the body through pollution or as ingredients in consumer products. In October 2008, scientists at HTP tested cord blood samples for 20 newborn infants in the United States. The cord blood of the "In utero/newborn" group was tested for 430 industrial compounds, pollutants, and other chemicals, including chemicals linked to brain and nervous system toxicity, immune system toxicity, and reproductive toxicity and fertility problems. The researchers recorded the number of these targeted chemicals that were found in each infant’s cord blood. The average number of targeted chemicals found was 127.45, with a standard deviation of 25.965. In 2000, the average number of targeted chemicals found in infant’s cord blood was 120. Conduct a hypothesis test at the 0.01 significance level to test if the average number has changed.

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?
h. How would the hypothesis test have changed if we wanted to see if the number of targeted chemicals had increased, instead of seeing if it had changed?

f. The data does not provide statistically significant evidence that the true avg. # of targeted chemicals found in the cord blood of all US infants in 2008 is different from 120.
Problem 4. The Human Toxome Project (HTP) is working to understand the scope of industrial pollution in the human body. Industrial chemicals may enter the body through pollution or as ingredients in consumer products. In October 2008, scientists at HTP tested cord blood samples for 20 newborn infants in the United States. The cord blood of the "In utero/newborn" group was tested for 430 industrial compounds, pollutants, and other chemicals, including chemicals linked to brain and nervous system toxicity, immune system toxicity, and reproductive toxicity and fertility problems. The researchers recorded the number of these targeted chemicals that were found in each infant’s cord blood. The average number of targeted chemicals found was 127.45, with a standard deviation of 25.965. **In 2000, the average number of targeted chemicals found in infant’s cord blood was 120.** Conduct a hypothesis test at the 0.01 significance level to test if the average number has changed.

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?
h. How would the hypothesis test have changed if we wanted to see if the number of targeted chemicals had increased, instead of seeing if it had changed?

\[ H_0: \mu = 120 \]
\[ H_a: \mu \neq 120 \]

**p-value > 0.10**

Yes! 120 (\(\mu_0\)) is in the interval we failed to Rej. \(H_0\)

**One-sided**

alt. hypothesis would look different

0 p-value \(\Rightarrow\) 0.10 < \(\text{p-value}\) < 0.15
Problem 5. 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.

\[
\begin{align*}
\text{Group 1: Face to Face} & \quad \text{Group 2: online} \\
\bar{X}_1 = \bar{X}_F &= 83.2 & \bar{X}_2 = \bar{X}_O &= 85.6 \\
S_1 = S_F &= 2.7 & S_2 = S_O &= 4.3 \\
n_1 = n_F &= 30 & n_2 = n_O &= 30 \\
\end{align*}
\]

\[
\begin{align*}
(\bar{X}_1 - \bar{X}_2) & \pm t* \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \\
(\bar{X}_F - \bar{X}_O) & \pm t* \sqrt{\frac{S_F^2}{n_F} + \frac{S_O^2}{n_O}} \\
(83.2 - 85.6) & \pm (2.462) \sqrt{\frac{2.7^2}{30} + \frac{4.3^2}{30}} \\
-2.4 & \pm (2.462)(0.927) \\
-2.4 & \pm 2.28 \\
98\% \text{ CI: } & (-4.10, -0.12)
\end{align*}
\]

Shelby Cummings

Week #11
Problem 5. 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.

b. We are 98% confident that the true difference between the average exam score for Face to Face students and the average exam score for online students is between -4.68 points and -0.12 points.
**Problem 6.** 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.**

- What are the hypotheses?
- What is the significance level?
- What is the value of the test statistic?
- What is the p-value?
- What is the correct decision?
- What is the appropriate conclusion/interpretation?
- Does the hypothesis test agree with the confidence interval from question 5?

\[
\begin{align*}
H_0 &: M_F = M_o \\
H_A &: M_F \neq M_o \\
\alpha &= 0.02 \\
T_S &= \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \\
&= \frac{83.2 - 85.6}{\sqrt{\frac{2.7^2}{30} + \frac{4.3^2}{30}}} \\
&= \frac{-2.4}{0.927} \\
&= -2.589
\end{align*}
\]
**Problem 6.** 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 5?

\[ d. \quad T_S = \frac{-2.589}{100} \quad \text{up \, to} \quad T_S = \frac{|-2.589|}{100} = 2.589 \]

\[ df = \min(n_1-1, n_2-1) = \min(30-1, 30-1) = \min(29, 29) = 29 \]

\[ -2.589 < T_S < 2.756 \]

\[ 0.01 < p\text{-value} < 0.02 \]

\[ e. \quad p\text{-value} < 0.02 \]

\[ p\text{-value} < \alpha \]

\[ \boxed{\text{Reject } H_0} \]
Problem 6. 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 5?

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

g. **Yes! 0 (\(M_0\)) isn’t in the interval**

Rejected \(H_0\)
Problem 7. A study was conducted to investigate the effectiveness of hypnotism in reducing pain. Eight randomly selected individuals were asked to rank their pain both before and after being hypnotized. A lower score indicates less pain. The average difference (after-before) for the sample was -3.13, with a standard deviation of 2.91. **Create a 90% confidence interval for the average difference in pain levels.**

a. Construct the 90% confidence interval.

b. Interpret your confidence interval from part a.

---

*a. Paired*

\[
\bar{X}_{\text{diff}} = t^* \left( \frac{S_{\text{diff}}}{\sqrt{n_{\text{diff}}}} \right)
\]

\[
\bar{X}_{\text{diff}} = -3.13 \\
t^* = 1.895 \\
S_{\text{diff}} = 2.91 \\
n_{\text{diff}} = 8
\]

90% CL

\[
df = n - 1 = 8 - 1 = 7
\]

\[
\bar{X}_{\text{diff}} = -3.13 \pm (1.895) \left( \frac{2.91}{\sqrt{8}} \right)
\]

\[
-3.13 \pm (1.895)(1.029)
\]

\[
-3.13 \pm 1.95
\]

\[
-3.13 - 1.95 = -5.08 \\
-3.13 + 1.95 = -1.18
\]

90% CI: (-5.08, -1.18)

**b.** We are 90% confident that the true avg. diff. in pain scores (after-before hypnosis) is between -5.08 and -1.18.
Problem 8. A study was conducted to investigate the effectiveness of hypnotism in reducing pain. Eight randomly selected individuals were asked to rank their pain both before and after being hypnotized. A lower score indicates less pain. The average difference (after-before) for the sample was -3.13, with a standard deviation of 2.91. **Conduct a hypothesis test at the 0.10 significance level to test if hypnotism changes the pain 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. Does the hypothesis test agree with the confidence interval from question 7?

h. How would the hypothesis test change if we wanted to determine if hypnotism reduced pain (instead of if it changes pain levels)?

\[ H_0 : \mu_{\text{diff}} = \emptyset \]

\[ H_A : \mu_{\text{diff}} \neq \emptyset \]

\[ \alpha = 0.10 \]

\[ T_S = \frac{\bar{X}_{\text{diff}} - \mu_0}{S_{\text{diff}}/\sqrt{n_{\text{diff}}}} \]

\[ \bar{X}_{\text{diff}} = -3.13 \]

\[ \mu_0 = \emptyset \]

\[ S_{\text{diff}} = 2.91 \]

\[ n_{\text{diff}} = 8 \]

\[ = \frac{-3.13 - \emptyset}{2.91/\sqrt{8}} = \frac{-3.13}{1.029} = -3.042 \]
Problem 8. A study was conducted to investigate the effectiveness of hypnotism in reducing pain. Eight randomly selected individuals were asked to rank their pain both before and after being hypnotized. A lower score indicates less pain. The average difference (after-before) for the sample was -3.13, with a standard deviation of 2.91. **Conduct a hypothesis test at the 0.10 significance level to test if hypnotism changes the pain 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. Does the hypothesis test agree with the confidence interval from question 7?
h. How would the hypothesis test change if we wanted to determine if hypnotism reduced pain (instead of if it changes pain levels)?

d. \( T_S = -3.042 \rightarrow |T_S| = |-3.042| = 3.042 \)
   \( \text{df} = n-1 = 8-1 = 7 \)
   \( 2.998 < |T_S| < 3.499 \)
   \( 0.01 < p-value < 0.02 \)

eg. \( p-value < 0.02 \)
   \( p-value < \alpha \)
   \( \text{Reject } H_0 \)
Problem 8. A study was conducted to investigate the effectiveness of hypnotism in reducing pain. Eight randomly selected individuals were asked to rank their pain both before and after being hypnotized. A lower score indicates less pain. The average difference (after-before) for the sample was -3.13, with a standard deviation of 2.91. Conduct a hypothesis test at the 0.10 significance level to test if hypnotism changes the pain 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. Does the hypothesis test agree with the confidence interval from question 7?
h. How would the hypothesis test change if we wanted to determine if hypnotism reduced pain (instead of if it changes pain levels)?

f. The data does provide stat. sign. evidence of a true avg. difference between the pain scores before and after hypnosis. Based on our sample, we think the scores decrease after hypnosis (on avg.).

g. Yes! $\bar{x}$ (Mo) is not in the interval, Rej. Ho
Problem 8. A study was conducted to investigate the effectiveness of hypnotism in reducing pain. Eight randomly selected individuals were asked to rank their pain both before and after being hypnotized. A lower score indicates less pain. The average difference (after-before) for the sample was -3.13, with a standard deviation of 2.91. **Conduct a hypothesis test at the 0.10 significance level to test if hypnotism changes the pain 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. Does the hypothesis test agree with the confidence interval from question 7?
h. How would the hypothesis test change if we wanted to determine if hypnotism reduced pain (instead of if it changes pain levels)?

\[ H_0: \mu = 0 \quad \text{versus} \quad H_A: \mu < 0 \]

\[ t = \frac{-3.13}{2.91 / \sqrt{8}} = \frac{-3.13}{2.91 \times 0.3535} = \frac{-3.13}{1.029} = -2.997 \]

\[ p-value < 0.01 \]

The conclusion is to reject the null hypothesis and conclude that hypnotism reduces pain.

**One sided test**

- **Alt. Hypothesis would change**
- **p-value would change**

\[ 0.005 < p-value < 0.001 \]

Conclusion: Stay the same
Problem 9. A researcher visiting the STAT department at TAMU wanted to test the theory, "For STAT 302 students, there is an association between smoking and GPR.” To test this theory, the researcher used data from an anonymous survey of 233 STAT 302 students in the Fall 2002 semester. Assume that this data set is equivalent to a random sample. Each student in the survey was asked which category they fell in: never smoked, former smoker, or current smoker. They were also asked to report their GPR. Conduct an ANOVA test to see if there is a relationship between smoking and GPR.

a. What are the hypotheses?
b. What is the significance level?
c. The ANOVA table below is partially filled in. Complete the missing spaces.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking (Groups)</td>
<td>2</td>
<td>12.31</td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Error (Residuals)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>65.84</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

d. What is the value of the test statistic?
e. What is the p-value?
f. What is the correct decision?
g. What is the appropriate conclusion/interpretation?

\[ \begin{align*} 
\text{a. } & H_0: \mu_{FS} = \mu_{NS} = \mu_{CS} \\
& H_A: \text{At least one } \mu \text{ is different} \\
\text{b. } & \alpha = 0.05 
\end{align*} \]
Problem 9. A researcher visiting the STAT department at TAMU wanted to test the theory, "For STAT 302 students, there is an association between smoking and GPR." To test this theory, the researcher used data from an anonymous survey of 233 STAT 302 students in the Fall 2002 semester. Assume that this data set is equivalent to a random sample. Each student in the survey was asked which category they fell in: never smoked, former smoker, or current smoker. They were also asked to report their GPR. Conduct an ANOVA test to see if there is a relationship between smoking and GPR.

a. What are the hypotheses?
b. What is the significance level?
c. The ANOVA table below is partially filled in. Complete the missing spaces.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking (Groups)</td>
<td>2</td>
<td>12.31</td>
<td>6.155</td>
<td>210.410</td>
<td>0.000</td>
</tr>
<tr>
<td>Error (Residuals)</td>
<td>230</td>
<td>53.53</td>
<td>0.233</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>232</td>
<td>65.84</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

d. What is the value of the test statistic?
e. What is the p-value?
f. What is the correct decision?
g. What is the appropriate conclusion/interpretation?

c. \( \text{df}_G = K - 1 = 3 - 1 = 2 \)
\( \text{df}_T = n - 1 = 233 - 1 = 232 \)
\( \text{df}_E = \text{df}_T - \text{df}_G = 232 - 2 = 230 \)

\( \text{SSE} = \text{SST} - \text{SSG} = (65.84 - 12.31) = 53.53 \)
\( \text{MSG} = \frac{\text{SSG}}{\text{df}_G} = \frac{12.31}{2} = 6.155 \)
\( \text{MSE} = \frac{\text{SSE}}{\text{df}_E} = \frac{53.53}{230} = 0.233 \)

\( F = \frac{\text{MSG}}{\text{MSE}} = \frac{6.155}{0.233} = 26.410 \)
Problem 9. A researcher visiting the STAT department at TAMU wanted to test the theory, "For STAT 302 students, there is an association between smoking and GPR." To test this theory, the researcher used data from an anonymous survey of 233 STAT 302 students in the Fall 2002 semester. Assume that this data set is equivalent to a random sample. Each student in the survey was asked which category they fell in: never smoked, former smoker, or current smoker. They were also asked to report their GPR. Conduct an ANOVA test to see if there is a relationship between smoking and GPR.

a. What are the hypotheses?
b. What is the significance level?
c. The ANOVA table below is partially filled in. Complete the missing spaces.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking (Groups)</td>
<td>2</td>
<td>12.31</td>
<td>6.155</td>
<td>26.416</td>
<td>0</td>
</tr>
<tr>
<td>Error (Residuals)</td>
<td>230</td>
<td>53.33</td>
<td>0.233</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>232</td>
<td>65.84</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

d. What is the value of the test statistic?
e. What is the p-value?
f. What is the correct decision?
g. What is the appropriate conclusion/interpretation?

d. Test statistic = F-value = 26.416

e. p-value = \( \phi \)
f. \( 0 < 0.05 \)

\[ p\text{-value} < \alpha \]  

Reject \( H_0 \)

9. The data does provide statistically significant evidence that at least one avg. GPRs is different (is an assoc. b/t smoking & GPR for STAT 302 students).