Statistics 511 Final Exam Spring 2016

Name: Solution

Section (Circle One):

3pm 4:30pm

Write your answer and all the relevant derivations on the exam paper in case for partial credit
(7.5pt) Assume the weight of a packaged product is normally distributed, and the standard deviation is 0.5.
(a) (1pt) Compute a 90% confidence interval for \( \mu \) given the average weight of 20 products is 5.45
(b) (1pt) Compute a 80% confidence interval for \( \mu \) given the average weight of 12 products is 5.75
(c) (1pt) Compute a 99% confidence interval for \( \mu \) given the average weight of 50 products is 5.25
(d) (1.5pt) How large the sample size \( n \) must be if the width of a 95% interval is to be 0.7?
(e) (1.5pt) How large the sample size \( n \) must be if the width of a 85% interval is to be 0.7?
(f) (1.5pt) How large the sample size \( n \) must be if the width of a 75% interval is to be 0.7?

\[(a) \quad 5.45 \pm \frac{2}{\sqrt{20}} (0.95) \times \frac{0.5}{\sqrt{20}} \Rightarrow 5.45 \pm 1.65 \times \frac{0.5}{\sqrt{20}} \Rightarrow (5.265, 5.634)\]

\[(b) \quad 5.75 \pm \frac{2}{\sqrt{12}} (0.90) \times \frac{0.5}{\sqrt{12}} \Rightarrow 5.75 \pm 1.28 \times \frac{0.5}{\sqrt{12}} \Rightarrow (5.565, 5.935)\]

\[(c) \quad 5.25 \pm \frac{2}{\sqrt{50}} (0.95) \times \frac{0.5}{\sqrt{50}} \Rightarrow 5.25 \pm 2.58 \times \frac{0.5}{\sqrt{50}} \Rightarrow (5.067, 5.432)\]

\[(d) \quad n \geq \left( \frac{2 \times \frac{0.5}{\sqrt{96}}}{0.7} \right)^2 = \left( \frac{2 \times 1.96 \times 0.15}{0.7} \right)^2 = 7.84 \]

\[\boxed{n \geq 8}\]

\[(e) \quad n \geq \left( \frac{2 \times \frac{0.5}{\sqrt{94}}}{0.7} \right)^2 = \left( \frac{2 \times 1.98 \times 0.15}{0.7} \right)^2 = 4.23 \]

\[\boxed{n \geq 5}\]

\[(f) \quad n \geq \left( \frac{2 \times \frac{0.5}{\sqrt{151}}}{0.7} \right)^2 = \left( \frac{2 \times 1.15 \times 0.15}{0.7} \right)^2 = 2.70 \]

\[\boxed{n \geq 3}\]
(2) (6.5pt) We know data is normally distributed and we obtained 16 samples. A 90% confidence interval for $\mu$ is $[21.6, 25.4]$.

(a) (1pt) Compute the sample mean $\bar{x}$

(b) (1pt) Compute the sample standard deviation $s$

(c) (1pt) Compute a 95% lower confidence bound for $\mu$

(d) (1pt) Compute a 80% confidence interval for $\mu$

(e) (1.5pt) Compute a 90% prediction interval for a future observation

(f) (1pt) Compute a 95% upper prediction bound for a future observation

\[ a) \quad \bar{x} = \frac{(21.6 + 25.4)}{2} = 23.5 \]

\[ b) \quad \frac{(25.4 - 21.6)}{2} = t_{15}(0.95) \times \frac{s}{\sqrt{16}} \Rightarrow s = 4.335 \]

\[ c) \quad [21.6, +\infty) \]

\[ d) \quad 23.5 \pm t_{15}(0.95) \times \frac{4.335}{\sqrt{16}} \Rightarrow 23.5 \pm 1.34 \times \frac{4.335}{\sqrt{16}} \]

\[ \approx (22.047, 24.953) \]

\[ e) \quad 23.5 \pm t_{15}(0.95) \times 4.335 \times \sqrt{1 + \frac{1}{16}} \Rightarrow 23.5 \pm 1.753 \times 4.335 \times \sqrt{1 + \frac{1}{16}} \]

\[ \approx (15.67, 31.33) \]

\[ f) \quad (-\infty, 31.33) \]
(3) (7pt) A manufacturer claims its lightbulb average lifetime is 500 hours. A consumer organization bought 11 lightbulbs with sample mean equal to 480 hours and sample variance equal to 100 hour^2. They discovered the lightbulb lifetime can be considered as normally distributed. The consumer organization was testing whether the average lightbulb lifetime was shorter than claimed.

(a) (1pt) Write down the null and the alternative hypotheses.

(b) (1pt) Compute the test statistic

(c) (1pt) Compute the rejection region at 0.10 significance level

(d) (1pt) Compute the rejection region at 0.05 significance level

(e) (1pt) Compute the p-value (i.e., the range for p-value)

(f) (1pt) At 0.05 significance level, what is your conclusion about the test?

(g) (1pt) Compute the rejection region at 0.05 level assuming we are simply testing whether the average lifetime is different than 500 hours, i.e., a different alternative hypothesis.

(a) \( H_0: \mu = 500 \quad H_a: \mu < 500 \)

(b) \[ \text{test stat} = \frac{480 - 500}{\sqrt{\frac{100}{11}}} \]

(c) at 0.10, \[ \text{test stat} < -t_{10(0.10)} \]

(d) at 0.05, \[ \text{test stat} < -t_{10(0.05)} \]

(e) \[ p\text{-value} = P(T_{10} \leq -6.63) = 0.0005 \]

(f) \( p\text{-value very small, reject } H_0 \)

(g) \[ t_{10(0.05)} = 2.228 \]
large sample \( \chi^2 \) test.

(4) (8pt) In a tasting experiment, 430 out of 2000 bottles of wine were considered spoiled to certain extent. Test if more than 20\% of wine were spoiled.

(a) (1pt) Write down the null and the alternative hypotheses.
(b) (1pt) Compute the test statistic
(c) (1pt) Compute the rejection region at 0.07 significance level
(d) (1pt) Compute the rejection region at 0.03 significance level
(e) (1pt) Compute the p-value (the exact value)
(f) (1pt) At 0.01 significance level, what is your conclusion about the test?

(g) (2pt) In another experiment, 43 out of 200 bottles were spoiled. Compute the test statistic and the exact p-value given this new experiment. At 0.05 significance level, what is your conclusion about the new test?

\[
\begin{align*}
(\text{a}) & \quad H_0 : \ P = 0.2 \\
(\text{b}) & \quad \text{test stat} = \frac{430/2000 - 0.2}{\sqrt{0.2 \times 0.8 / 2000}} = 1.677 \\
(\text{c}) & \quad \text{at } 0.07, \text{ test stat } > 1.48 \quad \therefore (0.93) = 1.48 \\
(\text{d}) & \quad \text{at } 0.03, \text{ test stat } > 1.88 \quad \therefore (0.97) = 1.88 \\
(\text{e}) & \quad \text{p-value} = P(\chi^2_{10,1} > 1.677) = 0.047 \\
(\text{f}) & \quad \text{p-value} = 0.047 > 0.01, \ \text{fail to reject } H_0 \\
(\text{g}) & \quad \text{test stat} = \frac{43/200 - 0.2}{\sqrt{0.2 \times 0.8 / 200}} = 0.53 \\
(\text{h}) & \quad \text{p-value} = P(\chi^2_{10,1} > 0.53) = 0.298 \\
(\text{p-value} = 0.298 > 0.05, \ \text{fail to reject } H_0)
\end{align*}
\]
(5) (4pt) We collected samples from two normally distributed populations. One sample has 8 data points with sample mean 71 and sample variance 4.2. The other sample has 10 data points with sample mean 80 and sample variance 3.8. The sample variances are almost the same. Test if $\mu_1$ and $\mu_2$ are different.

(a) (1pt) Compute the pooled estimator of $\sigma^2$
(b) (1pt) Compute the test statistic
(c) (1pt) Write down the degrees of freedom of the test statistic if the null hypothesis is true
(d) (1pt) Compute the range of the p-value

(a) $S_p^2 = \frac{8-1}{8+10-2} \times 4.2 + \frac{10-1}{8+10-2} \times 3.8 = 3.975$

(b) $t_{test} = \frac{71-80-1}{\sqrt{3.975(\frac{1}{8} + \frac{1}{10})}} = -9.52$

(c) $df = 8+10-2 = 16$

(d) Since $H_a: \mu_1 - \mu_2 \neq 0$, $P(T_{16} < -9.52) + P(T_{16} > 9.52)$

$p$-value $\approx 0$
(6) (4pt) We have a partial 1-way ANOVA table

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Squares</th>
<th>Square</th>
<th>F</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>2</td>
<td>2.4</td>
<td>1.2</td>
<td>0.178</td>
<td>p-value &gt; 0.10</td>
</tr>
<tr>
<td>Error</td>
<td>9</td>
<td>60.70</td>
<td>6.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>11</td>
<td>63.10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) (1pt) Write down the three assumptions of ANOVA test
(b) (2pt) Compute the missing numbers and fill up the ANOVA table
(c) (1pt) List the multiple comparison confidence intervals we need to compute after we reject the ANOVA null hypothesis. (Don’t need to actually compute the intervals)

(a) D normal  (b) equal variance $\sigma^2$  (c) independent samples

(c) 3 confidence intervals

CI for $\mu_1 - \mu_2$;

CI for $\mu_1 - \mu_3$;

CI for $\mu_2 - \mu_3$.

(Since we fail to reject $H_0$, we don’t need to compute them)
(7) (3pt) Assume the fitted regression line is \( \hat{y} = 2 - 5x \).
   (a) (1pt) If a data point has \( x = 3 \), what is the predicted \( \hat{y} \)?
   (b) (1pt) If we know the actual \( y = -12.5 \) for \( x = 3 \), what is the residual?
   (c) (1pt) Could a least square regression model return the residuals as 1.5, -1.1, 0.7, -0.9, 1.3? Why or why not?

\[
\begin{align*}
\text{(a)} & \quad \hat{y} = 2 - 5 \times 3 = -13 \\
\text{(b)} & \quad -12.5 - (-13) = 0.5 \\
\text{(c)} & \quad \text{Residuals add to 0, } \sum e_i = 0; \\
& \quad \text{since } 1.5 -1.1 + 0.7 -0.9 +1.3 = 0 \\
& \quad \text{they cannot be residuals from a regression fitted model.}
\end{align*}
\]