Review for the Final Exam (STA221)

1. (a) Review all the quiz questions, the midterm test, in class examples, homework problems;
(b) Go over the review for the midterm.

2. Given the following partial software output:

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>6</td>
<td>20.2</td>
</tr>
<tr>
<td>Within</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Total</td>
<td>88</td>
<td>110.5</td>
</tr>
</tbody>
</table>

\[ MS = \frac{S_S}{df} \]

\[ SS_F = SS_T - SS_E \]

(a) What is the value of F statistic?

\[ F = \frac{MS_C}{MS_E} = \frac{20.2/6}{90.3/82} = 3.71 \]

(b) What decision would be made regarding \( H_0: \) population means are equal?

\[ p-value = P(X \geq 3.71) = 0.00259 \]

\[ X \sim F(6, 82) \]

\[ H_0 \text{ reject} \]

3. Given the following partial information:

<table>
<thead>
<tr>
<th>Analysis of Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>AxB</td>
</tr>
<tr>
<td>Error</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

What is the interaction SS?

\[ SS_{AB} = MS_{AB} \cdot df_{AB} \]

\[ 8.61 \times 9 = 77.49 \]

\[ df_{AB} = 2.8 \]
4. If electricity power failures occur according to a Poisson distribution with an average of 3 failures every twenty weeks, calculate the probability that there will not be more than one failure during a particular week.

\[ P(X) = \frac{\lambda^x e^{-\lambda}}{x!} \]
\[ X = \# \text{ of failures} \]
\[ P(X \leq 1) = P(0) + P(1) = \frac{0.15 e^{-0.15}}{0!} + \frac{0.15^1 e^{-0.15}}{1!} = 0.06 + 0.13 = 0.99 \]

5. The correlation coefficient for the heights and weight of ten offensive backfield football players was determined to be \( r = 0.8261 \).

(a) What percentage of the variation in weight was explained by the heights of the players?

\[ r^2 = 0.8261^2 = 0.68244 \text{ or } 68.24 \%
\]

(b) Is there sufficient evidence at the \( \alpha = 0.01 \) level to claim that heights and weights are positively correlated?

\[ H_0: \rho = 0 \text{ vs } H_a: \rho > 0 \]
\[ t = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}} \sim t(n-2) \]
\[ t = \frac{0.8261 \sqrt{10-2}}{\sqrt{1-0.8261^2}} = 4.146 \]
\[ P\text{-value} = P(T \geq 4.146) \]

\[ T \sim t_8 \quad 0.001 < P\text{-value} < 0.0025 \] (use t-table)

\[ \Rightarrow \text{reject } H_0 \]

conclude that there is strong evidence that heights and weights are positively correlated.
6. **Jury Selection.** One study of grand juries in Alameda County, California, compared the demographic characteristics of jurors with the general population, to see if jury panels were representative. The results for age are shown below. The investigators wanted to know if the 66 jurors were selected at random from the population of Alameda County. (Only persons over 21 and over are considered; the county age distribution is known from Public Health Department data.) The study was published in the UCLA Law Review.

<table>
<thead>
<tr>
<th>Age</th>
<th>Count-wide % # of jurors observed</th>
<th># of jurors expected (O-E)</th>
<th>(O-E)^2/E</th>
</tr>
</thead>
<tbody>
<tr>
<td>21-40</td>
<td>42% 5</td>
<td>66.0.42 = 27.72</td>
<td>-22.72</td>
</tr>
<tr>
<td>41-50</td>
<td>23% 9</td>
<td>15.18</td>
<td>-6.18</td>
</tr>
<tr>
<td>51-60</td>
<td>16% 19</td>
<td>10.56</td>
<td>8.44</td>
</tr>
<tr>
<td>over 60</td>
<td>19% 33</td>
<td>12.54</td>
<td>20.46</td>
</tr>
<tr>
<td>Total</td>
<td>100% 66</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Do we have evidence that grand juries are selected at random for the population of Alameda County?

\[
x^2 = \frac{(-22.72)^2}{27.72} + \frac{(-6.18)^2}{15.18} + \frac{8.44^2}{10.56} + \frac{20.46^2}{12.54}
\]

\[
= 61.27
\]

\[df = 4 - 1 = 3\]

\[p < 0.0005\ \text{(almost zero)}\]

\[H_0: \text{For each age group, the proportion of jurors is consistent with the county proportion}\]

\[H_a: \text{at least one is not consistent}\]

So we reject \(H_0\).

**Conclusion:** At least one is not consistent. or jurors are not selected at random.
7. *Decay of polyester fabrics in landfills.* How quickly do synthetics such as polyester decay in landfills? A researcher buried polyester strips in the soil for different lengths of time, then dug up the strips and measured the force required to break them. Breaking strength is easy to measure and is a good indicator of decay. Lower strength means the fabric has decayed. Part of the study involved burying 10 polyester strips in well-drained soil in the summer. Five of the strips, chosen at random, were dug up after 2 weeks; the other 5 were dug up after 16 weeks. Here are the breaking strengths in pounds:

<table>
<thead>
<tr>
<th></th>
<th>118</th>
<th>126</th>
<th>128</th>
<th>120</th>
<th>129</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 weeks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16 weeks</td>
<td>124</td>
<td>98</td>
<td>110</td>
<td>140</td>
<td>109</td>
</tr>
</tbody>
</table>

Is there evidence that breaking strengths are lower for strips buried longer? (Use a rank test)

\[
\begin{align*}
W &= 1 + 2 + 3 + 6 + 10 = 22 \\
\mu_w &= \frac{n_1(N+1)}{2} = \frac{5(10+1)}{2} = 22.5 \\
6_w &= \sqrt{\frac{n_1n_2(N+1)}{12}} = \sqrt{\frac{5 \cdot 5 \cdot 11}{12}} = 4.8 \\
W \text{ is not far from } \mu_w \\
\Rightarrow & \text{ fail to reject } H_0: \text{ no difference}
\end{align*}
\]

P-value = \[P(W \geq 22) = P(W \geq 22.5)\]

\[
\begin{align*}
&= P(2 \geq \frac{22.5 - 22.5}{4.8}) = P(Z \geq 0.00) \\
&= 0.8944 \\
\Rightarrow & \text{ fail to reject } H_0
\end{align*}
\]
8. Below are scores for 24 students who took the same final examination, but who are from the
groups in which were used different teaching techniques

<table>
<thead>
<tr>
<th>Group I</th>
<th>Group II</th>
<th>Group III</th>
<th>Group IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>81</td>
<td>55</td>
<td>63</td>
<td>51</td>
</tr>
<tr>
<td>90</td>
<td>60</td>
<td>53</td>
<td>67</td>
</tr>
<tr>
<td>69</td>
<td>61</td>
<td>64</td>
<td>80</td>
</tr>
<tr>
<td>66</td>
<td>72</td>
<td>67</td>
<td>70</td>
</tr>
<tr>
<td>81</td>
<td>39</td>
<td>56</td>
<td>68</td>
</tr>
<tr>
<td>75</td>
<td>85</td>
<td>70</td>
<td>67</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Total</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>462</td>
<td>372</td>
<td>373</td>
<td>403</td>
</tr>
</tbody>
</table>

Suppose $SS(\text{total}) = 700$, and $SS(\text{between groups}) = 200$.

(a) What is the observed value of the statistic one computes to test $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$
against $H_1: \text{not all 4 means are equal}$?

(b) If $\alpha = .01$, what is the critical value of the statistic?

(c) Suppose that the observed value of the statistic is 4.1 while the critical value is 5.3. What is
your conclusion?

\begin{align*}
(a) \quad F &= \frac{MS_C}{MS_E} = \frac{SS_C/df_C}{SS_E/df_E} = \frac{200}{500/20} \\
&= \frac{2}{3} \cdot \frac{20}{5} = \frac{3}{2} = 1.5
\end{align*}

(b) Use $F$-table with $df_s: 3 \text{ and } 20$

\begin{align*}
F^*_{2:0.01} &= 4.94
\end{align*}

(c) $F = 4.1$

\begin{align*}
F^* &= 5.3
\end{align*}

fail to reject $H_0$