MULTIPLE CHOICE (5 pts each)  

(1) This technique reduces a large number of independent variables into a smaller set of orthogonal ones in order to identify the most important influences on the dependent variable.
(A) ANOVA  
(B) Bootstrapping  
(C) Logistic regression  
(D) Multivariate regression  
(E) Principle components analysis.

(2) When we do an two-way ANOVA and get a large p value, which of the following best describes our conclusion?
(A) We have strong evidence to suggest that there is a relationship between values of factors and the means of groups with those factor values.  
(B) We lack strong evidence to suggest that there is a relationship between values of factors and the means of groups with those factor values.  
(C) We prove that there is a relationship between values of factors and the means of groups with those factor values.  
(D) We prove that there is no relationship between values of factors and the means of groups with those factor values.  
(E) One or more of the group means is different from one or more of the others.

(3) Which of the following best describes the conditions that would result in an ANOVA yielding a small p value?
(A) The variance of group means is larger than expected from the overall variance.  
(B) The variance of group means is larger than expected from the sum of the within group variances.  
(C) The variance of group means is larger than the overall variance.  
(D) The variance of group means is larger than the sum of the within group variances.  
(E) The sum of squares total is the sum of the sum of squares within and the sum of squares among.

(4) Dunett's test is a modified version of an ANOVA in which the goal is to compare all the means of a set of treatment groups to the same control. It may be used in cases where different doses of a drug are given and the goal is not so much whether they differ from one another, but rather whether any of them differ from the control treatment. If we had a control and 4 treatment groups and performed Dunnet's test and it returned a p value of 0.075 what would our conclusion be?
(A) The overall p value would be 0.3.  
(B) The mean of one or more of the treatment groups differs from at least one other.  
(C) The mean of one or more of the treatment groups differs from the control group.  
(D) The means of the treatment groups do not differ from one another.  
(E) The means of the treatment groups do not differ from the control group.

(5) Which of the following is NOT correct?
(A) We use the $X^2$ test to compare proportions.  
(B) We use the $t$ test to compare means.  
(C) We use the $F_{\text{max}}$ test to compare means.  
(D) We use the F test to compare variances.  
(E) We use the one-factor ANOVA to compare means.
(6) Prior to performing a statistical test, which of the following would NOT cause us to consider transforming our data?
(A) Non-linear relationship between two variables for which we wish to describe a relationship.
(B) Non-normal data distributions in 5 groups for which we wish to compare means.
(C) Non-normal data distributions in 5 groups for which we wish to compare variances.
(D) Unequal variances in 5 groups for which we wish to compare means.
(E) Unequal variances in 5 groups for which we wish to compare variances.

(7) Consider the plots to the right. Indicate the letter of the plot that matches the one to the upper left in terms of them being a pair of two-factor ANOVA interaction plots.

(8) Assuming that the plot to the right shows means and Tukey-Cramer comparison intervals, which of the following is the correct full set of means that a series of Bonferroni-corrected t tests would indicate are unequal?
(A) A≠B, B≠C, C≠D, E≠F.
(B) A≠C, A≠D, D≠E, D≠F.
(C) A≠D, D≠F.
(D) A≠E, A≠F, B≠E, B≠F.
(E) A≠F, B≠E.

(9) The value obtained by dividing \(SS_{\text{regression}}\) by \(SS_{\text{total}}\) is which of the following?
(A) Coefficient of determination
(B) Correlation coefficient
(C) Slope
(D) \(t\) calculated value
(E) Y-intercept

For the next two questions consider a medical test with a false positive rate of 2% and a false negative rate of 3%. The frequency of the condition in people who take this test is 4%.

(10) What value is closest to the percentage of people who test "positive" who are genuinely positive for the condition?
(A) 50%  (B) 55%  (C) 60%  (D) 65%  (E) 70%

(11) What value is closest to the percentage of people who test "negative" who are actually negative for the condition?
(A) 99.85%  (B) 99.9%  (C) 99.9%  (D) 99.95%  (E) 100%
The next 3 questions are based on the paper that you were instructed to read for this exam.

(12) Which of the following is the best description of overall result of the study?
(A) Females are more polygamous than males.
(B) Males are more monogamous than females.
(C) Monogamy is better for females and polygamy is better for males.
(D) Monogamy is better for males and polygamy is better for females.
(E) The lack of strict monogamy or polygamy in humans may be due to differences between sexes.

(13) Which of the following describes a key line of mathematical reasoning used in the study?
(A) The different means in males and females are caused by hormone levels.
(B) The differences in variances in males and females can be explained by differences in means.
(C) The platykurtic trait distribution can be explained by the combination of two normal distributions.
(D) The skews of the traits are opposite in males and females, but result in a symmetric distribution.
(E) Transforming the SOI data into 2D:4D data provides better results.

(14) Which of the following techniques from class is most similar to one used in the study?
(A) A $X^2$ test of independence.
(B) A $X^2$ goodness of fit test.
(C) An unpaired heteroscedastic t test.
(D) A one-factor ANOVA.
(E) A two-factor ANOVA.

The next 3 questions are based on the student research projects described in class.

(15) Which of the following best describes Ashley’s fish asymmetry study?
(A) She used an ANOVA test to compare the means of FA (i.e., L-R) values before & after mortality.
(B) She used an F test to compare the variances of FA (i.e., L-R) values before & after mortality.
(C) She used an F test to compare the means of FA (i.e., L-R) values before & after mortality.
(D) She used Levene’s test to compare the variances of FA (i.e., L-R) values before & after mortality.
(E) She used Levene’s test to compare the means of FA (i.e., L-R) values before & after mortality.

(16) Which of the following best describes Beverly and Hannah’s cancer study?
(A) They examined a series of correlations and used positive and negative residuals to identify cancers that were overfunded and underfunded with respect to medical research.
(B) They examined a series of regressions and used positive and negative residuals to identify cancers that were overfunded and underfunded with respect to medical research.
(C) They examined a series of correlations and used positive and negative residuals to identify cancers that were more or less deadly.
(D) They examined a series of regressions and used positive and negative residuals to identify cancers that were more or less deadly.
(E) They compared rates of cancer in the US to rates in the UK.

(17) Natalie (the brand new graduate student) will be performing a study of evolvability which will involve which of the following techniques?
(A) Comparing morphological traits of different species to look for patterns.
(B) Comparing the DNA sequences of quickly and slowly evolving bacteria.
(C) Measuring the wings of Drosophila raised at different temperatures.
(D) Running individual-based computer simulations on a computer.
(E) Searching online databases of protein expression patterns.
The data plot to the right is used for the next 5 questions.

(18) Which of the following is closest to the correlation coefficient?
(A) 0.9
(B) 0.8
(C) 0.7
(D) 0.6
(E) 0.5

(19) This data set exhibits ____ positive and ____ negative residuals respectively.
(A) 8 , 12 (B) 9 , 11 (C) 10 , 10 (D) 11 , 9 (E) 12 , 8

(20) For this data set, which of the following pairs of values is the most likely match for the mean productivity and mean biodiversity values respectively?
(A) 1.9 , 1.9 (B) 2.0 , 2.1 (C) 2.1 , 2.0 (D) 2.2 , 1.9 (E) 2.3 , 1.8

(21) Which of the following is closest to the t value for a test of the significance of the correlation coefficient?
(A) 4.2 (B) 7.9 (C) 8.8 (D) 10.9 (E) 12.1

(22) Which of the following is the best conclusion to draw from this data?
(A) The correlation between biodiversity and productivity doesn’t mean anything.
(B) Increased biodiversity causes increased productivity.
(C) Increased productivity causes increased biodiversity.
(D) There is a significant relationship between biodiversity and productivity.
(E) There is no significant relationship between biodiversity and productivity

(23) The procedure that involves scrambling our data over and over and recomputing the observed statistic many times in order to compare it to the original statistic is the:
(A) Monte Cristo test (C) Permutation test (E) Bootstrapping
(B) Monte Carlo test (D) Parameterization test

(24) An ecologist believes that pollutants are disrupting the development of the fish in a series of lakes downstream from a chemical plant. She does not predict that the overall size of the fish would be affected; rather that the polluted populations would have more smaller-than-usual fish and more larger-than-usual fish compared to unpolluted lakes. If she collects a large number of fish from each of several lakes, which of the following is the most appropriate test for her to perform to test her hypothesis?
(A) Correlation (D) One-way ANOVA (E) Fmax test
(B) Regression (E) Two-way ANOVA
**On this page and the next two, report all calculated non-integer values to the nearest 0.001**

25. **CORRELATION/REGRESSION.** Consider the following data and calculations obtained from doing a regression/correlation analysis. **NOTE**: three columns have been left blank for you to complete in order to answer the questions below.

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>Y</th>
<th>(x_i - \bar{x})</th>
<th>(y_i - \bar{y})</th>
<th>((x_i - \bar{x})^2)</th>
<th>((y_i - \bar{y})^2)</th>
<th>SPM</th>
<th>(\hat{y})</th>
<th>(\hat{y} - \bar{y})</th>
<th>((\hat{y} - \bar{y})^2)</th>
<th>(y_i - \hat{y})</th>
<th>((y_i - \hat{y})^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>10</td>
<td>-9</td>
<td>81</td>
<td></td>
<td></td>
<td></td>
<td>11.46</td>
<td>-5.54</td>
<td>30.67</td>
<td>-1.46</td>
<td>2.14</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>17</td>
<td>-6</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td>13.31</td>
<td>-3.69</td>
<td>13.63</td>
<td>3.69</td>
<td>13.63</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>-3</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td>15.15</td>
<td>-1.85</td>
<td>3.41</td>
<td>-3.15</td>
<td>9.95</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td>17.00</td>
<td>0.00</td>
<td>0.00</td>
<td>2.00</td>
<td>4.00</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>16</td>
<td>3</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td>18.85</td>
<td>1.85</td>
<td>3.41</td>
<td>-2.85</td>
<td>8.10</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>22</td>
<td>5</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td>20.08</td>
<td>3.08</td>
<td>9.47</td>
<td>1.92</td>
<td>3.70</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>23</td>
<td>10</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td>23.15</td>
<td>6.15</td>
<td>37.87</td>
<td>-0.15</td>
<td>0.02</td>
<td></td>
</tr>
</tbody>
</table>

(a, 4 pts) What is the slope of the "best-fit" line?  
Slope = __________

(b, 4 pts) What is the Y-intercept of the "best-fit" line?  
Y-int. = __________

(c, 4 pts) What is \(R^2\) for the "best-fit" line?  
\(R^2 = \) __________

(d, 4 pts) Calculate F for an ANOVA test of the slope of this data.  
\(F = \) __________

(e, 4 pts) What is the most precise range of p values you can provide for an ANOVA analysis of the significance of the slope?  
(\(\_\_\_ < p < \_\_\_\_)  
(Remember that 0 and 1.0 may be valid values for this question)
26. ANOVA. Although pine trees don’t shed their needles, maybe the chlorophyll levels in their needles change over the course of the year. Consider the hypothetical data below showing the levels of chlorophyll in three needles taken from plants grown in a greenhouse under standardized conditions in which only the soils is varied (control plots and fertilizer plots). The samples are taken over the course of a year (Spring, Summer, Fall, Winter). On this page you will shows the results of an F_{\text{max}} test and provide the one-factor and two-factor ANOVA tables. On the next page you will provide the interaction plots and your interpretations of your results.

Note: these values (with some descriptive statistics) are reproduced on the otherwise blank last page.

<table>
<thead>
<tr>
<th></th>
<th>Sp</th>
<th>Su</th>
<th>Fa</th>
<th>Wi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cntl.</td>
<td>18</td>
<td>12</td>
<td>9</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>15</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>18</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>Fert.</td>
<td>17</td>
<td>7</td>
<td>3</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>14</td>
<td>9</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>18</td>
<td>3</td>
<td>20</td>
</tr>
</tbody>
</table>

(a, 2 pts ea) Conduct an F_{\text{max}} test and provide the values requested below:

F_{\text{max}} = ______  Critical F_{\text{max}} to use for \alpha=0.05 = ______

No matter what these values may or may not indicate, you will NOT transform your data to continue for the purposes of this exam.

(b, 1 pt ea) Regardless of the result of your F_{\text{max}} test above, conduct a one-factor ANOVA analysis on the raw data and complete the table below:

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(c, 1 pt ea) Regardless of the result of your F_{\text{max}} test on the previous page, conduct a two-factor ANOVA analysis on the raw data and complete the table below.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Season</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soil type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interact.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(d, 5 pts ea) Complete the pair of interaction plots below with the symbols requested.
In the left plot you MUST USE the following symbols: spring (O), summer (Δ), fall (□), winter (◊)
In the right plot you MUST USE the following symbols: control (O), fertilizer (Δ)

(e, 6 pts) State the conclusions of the one-factor ANOVA test in the box below. Use the grammar presented in lecture, state whether any results are "significant" or not, and include a statement about your degree of confidence with the most precise range of p values you can.
Note: no credit will be given for ANY text outside the box or hard to read answers and a lack of precision or unnecessary information and filler text will result in a loss of points.

(f, 6 pts) State the full conclusions of the two-factor ANOVA in the box below. Use the grammar presented in lecture, state whether each factor or interaction is "significant" or not, and include a statement about your degree of confidence with the most precise range of p values you can.
Note: no credit will be given for ANY text outside the box or hard to read answers and a lack of precision or unnecessary information and filler text will result in a loss of points.

(g, 4 pts) Based on the description of the experiment, do you consider this data sufficient to demonstrate that the season, the soil type, or an interaction between the two caused a change in the chlorophyll levels in these plants? If so, why? If not, why not?
### Critical F values, $\alpha=0.05$

<table>
<thead>
<tr>
<th>df for denominator</th>
<th>df for numerator</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>10.13</td>
<td>9.55</td>
<td>9.28</td>
<td>9.32</td>
<td>9.01</td>
<td>8.94</td>
<td>8.89</td>
<td>8.85</td>
<td>8.81</td>
<td>8.79</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>7.71</td>
<td>6.94</td>
<td>6.59</td>
<td>6.39</td>
<td>6.26</td>
<td>6.16</td>
<td>6.09</td>
<td>6.04</td>
<td>6.00</td>
<td>5.96</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6.61</td>
<td>5.79</td>
<td>5.41</td>
<td>5.19</td>
<td>5.05</td>
<td>4.95</td>
<td>4.89</td>
<td>4.82</td>
<td>4.74</td>
<td>4.77</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5.99</td>
<td>5.44</td>
<td>4.76</td>
<td>4.53</td>
<td>4.39</td>
<td>4.28</td>
<td>4.21</td>
<td>4.15</td>
<td>4.10</td>
<td>4.06</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>5.59</td>
<td>4.74</td>
<td>4.35</td>
<td>4.12</td>
<td>3.97</td>
<td>3.87</td>
<td>3.79</td>
<td>3.73</td>
<td>3.68</td>
<td>3.64</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5.32</td>
<td>4.46</td>
<td>4.07</td>
<td>3.84</td>
<td>3.69</td>
<td>3.59</td>
<td>3.50</td>
<td>3.44</td>
<td>3.39</td>
<td>3.35</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>5.12</td>
<td>4.26</td>
<td>3.86</td>
<td>3.63</td>
<td>3.48</td>
<td>3.37</td>
<td>3.29</td>
<td>3.23</td>
<td>3.18</td>
<td>3.14</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>4.96</td>
<td>4.03</td>
<td>3.71</td>
<td>3.48</td>
<td>3.33</td>
<td>3.22</td>
<td>3.14</td>
<td>3.07</td>
<td>3.02</td>
<td>2.98</td>
<td></td>
</tr>
</tbody>
</table>

### Critical F values, $\alpha=0.025$

<table>
<thead>
<tr>
<th>df for denominator</th>
<th>df for numerator</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>19.23</td>
<td>16.23</td>
<td>14.78</td>
<td>13.80</td>
<td>13.01</td>
<td>12.33</td>
<td>11.76</td>
<td>11.29</td>
<td>10.91</td>
<td>10.59</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>14.15</td>
<td>12.03</td>
<td>10.70</td>
<td>9.89</td>
<td>9.22</td>
<td>8.65</td>
<td>8.19</td>
<td>7.78</td>
<td>7.41</td>
<td>7.10</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>12.19</td>
<td>10.36</td>
<td>9.17</td>
<td>8.39</td>
<td>7.74</td>
<td>7.19</td>
<td>6.77</td>
<td>6.40</td>
<td>6.06</td>
<td>5.77</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>10.68</td>
<td>9.16</td>
<td>7.88</td>
<td>7.13</td>
<td>6.47</td>
<td>5.93</td>
<td>5.48</td>
<td>5.10</td>
<td>4.76</td>
<td>4.48</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>9.27</td>
<td>7.79</td>
<td>6.56</td>
<td>5.86</td>
<td>5.25</td>
<td>4.73</td>
<td>4.30</td>
<td>3.93</td>
<td>3.60</td>
<td>3.32</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>7.99</td>
<td>6.64</td>
<td>5.47</td>
<td>4.84</td>
<td>4.27</td>
<td>3.77</td>
<td>3.38</td>
<td>3.03</td>
<td>2.71</td>
<td>2.44</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>6.86</td>
<td>5.58</td>
<td>4.44</td>
<td>3.86</td>
<td>3.33</td>
<td>2.85</td>
<td>2.47</td>
<td>2.13</td>
<td>1.82</td>
<td>1.53</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>5.83</td>
<td>4.62</td>
<td>3.54</td>
<td>2.99</td>
<td>2.50</td>
<td>2.05</td>
<td>1.70</td>
<td>1.38</td>
<td>1.10</td>
<td>0.85</td>
<td></td>
</tr>
</tbody>
</table>

### Critical Fmax values.

Upper values are $\alpha=0.05$. Lower values are $\alpha=0.025$. 

**Critical Fmax values.**

Upper values are $\alpha=0.05$. Lower values are $\alpha=0.025$.
This is the data for the ANOVA analyses.

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>Var</th>
<th>Mean</th>
<th>Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cntl-Sp</td>
<td>15</td>
<td>7</td>
<td>21</td>
<td>9</td>
</tr>
<tr>
<td>Cntl-Su</td>
<td>13</td>
<td>31</td>
<td>15</td>
<td>9</td>
</tr>
<tr>
<td>Cntl-Fa</td>
<td>5</td>
<td>12</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Cntl-Wi</td>
<td>23</td>
<td>9</td>
<td>17</td>
<td>13</td>
</tr>
<tr>
<td>Fert-Sp</td>
<td>21</td>
<td>9</td>
<td>15</td>
<td>7</td>
</tr>
<tr>
<td>Fert-Su</td>
<td>15</td>
<td>9</td>
<td>13</td>
<td>31</td>
</tr>
<tr>
<td>Fert-Fa</td>
<td>11</td>
<td>12</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>Fert-Wi</td>
<td>17</td>
<td>13</td>
<td>23</td>
<td>9</td>
</tr>
<tr>
<td>Overall</td>
<td>15</td>
<td>38.08696</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>