Please place your name legibly at the top and the first letter of your last name in the box. Put your calculator's number on the blank below the box. Written answers should be concise and final answers should be copied into blanks provided where designated.

MULTIPLE CHOICE (3 pts each) The exam is worth 120 points

1) Which of the following tests compares the locations of two samples to determine whether the populations differ? It is often referred to as a "median test."
(A) Krustal-Wallis test.  (C) Permutation test.  (E) Sign test.
(B) Mann-Whitney U test.  (D) Principle components analysis.

2) If we have a data set of positive values; which of the following is NOT a valid data transformation?
(A) \( f(x) = x^2 + 5 \)
(B) \( f(x) = x^2 - 5 \)
(C) \( (x + 5)^2 \)
(D) \( (x - 5)^2 \)
(E) \( x + 5^2 \)

3) What value is closest to the percentage of people who test "positive" who are genuinely positive for the condition?
(A) 30%  (B) 40%  (C) 50%  (D) 60%  (E) 70%

4) What value is closest to the percentage of people who test "negative" who are actually positive for the condition?
(A) 0.1%  (B) 0.2%  (C) 0.3%  (D) 0.4%  (E) 0.5%

5) When we test the robustness of our value by taking a series of samples (with replacement) from the original data set and calculate the result again, we are doing which of the following?
(A) Bootstrapping  (C) Monte Carlo methods  (E) Spearman rank correlation
(B) Krustal-Wallis testing  (D) Numerical analysis

6) In the example involving chaos and population growth I described chaotic and stochastic systems; which of the following correctly describes these two?
(A) Chaotic systems are stochastic systems with added random factors.
(B) Chaotic systems can be described with equations, stochastic systems have some randomness.
(C) Chaotic systems have some randomness, stochastic systems can be described with equations
(D) Chaotic systems have higher variances than stochastic systems
(E) Chaotic systems have lower variances than stochastic systems

7) Which of the following would NOT cause us to consider transforming our data prior to statistical testing:
(A) The data sets in an F test are not normally distributed.
(B) The data we are plotting when doing a regression appears to be curved instead of straight.
(C) The residuals of our data values are normally distributed around the "best-fit" line.
(D) The variances of the groups we are comparing in a one-way ANOVA are very similar.
(E) The variances of the groups we are comparing in a two-way ANOVA differ.
The next 3 questions are based on the paper that you were instructed to read for this exam.

(8) Which of the following is the best description of overall result of the study?
(A) Female snails exposed to parasites laid fewer and smaller eggs than controls.
(B) Female snails exposed to parasites laid more and larger eggs than controls.
(C) Female snails exposed to parasites mated less often and with fewer males than controls.
(D) Female snails exposed to parasites mated more often and with more males than controls.
(E) Female snails exposed to parasites showed mating preferences for males that are unexposed.

(9) In the paper, a technique called "Dunnett's test" was performed. Based on your reading of the paper (and maybe a quick trip to Wikipedia), which of the following techniques discussed in class is most similar to their use of the "Dunnett's test"?
(A) A One-way ANOVA
(B) A Two-way ANOVA
(C) A series of Bonferroni corrected t tests.

(10) Which of the following best describes the parasite and host species studied?
(A) The species live in Australia and the trematode worm parasite infects snails and snakes.
(B) The species live in Canada and the trematode worm parasite infects snails and beavers.
(C) The species live in Germany and the trematode worm parasite infects snails and fish.
(D) The species live in Mexico and the trematode worm parasite infects snails and dogs.
(E) The species live in New Zealand and the trematode worm parasite infects snails and ducks.

(11) Which of the following was NOT specifically described as a design goal during the presentation of graphical displays of quantitative information?
(A) Although dark to light is an intuitive scale for quantitative values, the use of colors for indicating magnitude should often be avoided as there is no natural low to high progression of colors.
(B) Serif fonts are better than sans serif fonts because of ease of reproduction.
(C) The ratio of ink/information to space should be maximized.
(D) Thin parallel lines can create a hypnotic effect called the “Moire effect” and should be avoided.
(E) Whenever possible, and if unobtrusive, additional information should be added to figures.

(12) The proportion of variance of Y explained by variance in X is represented by the:
(A) Correlation coefficient.
(B) Correlation factor.
(C) Coefficient of determination
(D) Coefficient of interaction.
(E) Slope.

(13) A full ANOVA analysis of data that includes multiple factors and groups may typically include all of the following EXCEPT:
(A) A chi-square goodness of fit test.
(B) A comparison of inclusion intervals.
(C) A series of Bonferroni corrected t tests.
(D) An F_{max} test of the equality of variances.
(E) An F test of F=MSA/MSW.

(14) If a researcher conducts a one-way ANOVA test on an overall data set that includes 4 comparison groups and then decides to compare the groups pairwise using Bonferroni-corrected t tests, which of the following is closest to the appropriate p value to use so that the overall probability of type I error is only 0.05.
(A) 0.001
(B) 0.002
(C) 0.004
(D) 0.006
(E) 0.008
The 5 sets of interaction plots below are used for the next 3 questions.

(15) Which pair of interaction plots above most likely depicts a situation in which neither color nor humidity influence the mean, but there is an interaction?  **B**

(16) Which of the following values most closely matches the value for the blue dry sample in data set C?
(A) 8  
(B) 25  
(C) 35  
(D) 40  
(E) 50

(17) Which of the following values is NOT a possible value for the blue damp sample from the indicated data set.
(A) Set A = 20  
(B) Set B = 25  
(C) Set C = 27  
(D) Set D = 50  
(E) Set E = 18

The 5 data plots at the right are used for the next 2 questions.

(18) Which of the figures above depicts data that has the highest value of “r”?  **A**

(19) Which of the figures above depicts data that has the highest value of “R²”?  **E**
**20. CORRELATION/REGRESSION.** Consider the following data and calculations obtained from doing a regression/correlation analysis.

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>𝑥𝑖 − 𝑥̅</th>
<th>𝑦𝑖 − 𝑦̅</th>
<th>(𝑥𝑖 − 𝑥̅)²</th>
<th>(𝑦𝑖 − 𝑦̅)²</th>
<th>SP</th>
<th>𝑦̂</th>
<th>𝑦̂ − 𝑦̅</th>
<th>(𝑦̂ − 𝑦̅)²</th>
<th>𝑦𝑖 − 𝑦̂</th>
<th>(𝑦𝑖 − 𝑦̂)²</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>9</td>
<td>-9</td>
<td>-7</td>
<td>81</td>
<td>49</td>
<td>63</td>
<td>10.06</td>
<td>-5.94</td>
<td>35.23</td>
<td>-1.06</td>
<td>1.13</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>-6</td>
<td>-1</td>
<td>36</td>
<td>1</td>
<td>6</td>
<td>12.04</td>
<td>-3.96</td>
<td>15.66</td>
<td>2.96</td>
<td>8.74</td>
</tr>
<tr>
<td>8</td>
<td>11</td>
<td>-3</td>
<td>-5</td>
<td>9</td>
<td>25</td>
<td>15</td>
<td>14.02</td>
<td>-1.98</td>
<td>3.91</td>
<td>-3.02</td>
<td>9.13</td>
</tr>
<tr>
<td>12</td>
<td>22</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>36</td>
<td>6</td>
<td>16.66</td>
<td>0.66</td>
<td>0.43</td>
<td>5.34</td>
<td>28.52</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>1</td>
<td>-3</td>
<td>1</td>
<td>9</td>
<td>-3</td>
<td>16.66</td>
<td>0.66</td>
<td>0.43</td>
<td>-3.66</td>
<td>13.39</td>
</tr>
<tr>
<td>13</td>
<td>16</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>17.32</td>
<td>1.32</td>
<td>1.74</td>
<td>-1.32</td>
<td>1.74</td>
</tr>
<tr>
<td>17</td>
<td>23</td>
<td>6</td>
<td>7</td>
<td>36</td>
<td>49</td>
<td>42</td>
<td>19.96</td>
<td>3.96</td>
<td>15.66</td>
<td>3.04</td>
<td>9.26</td>
</tr>
<tr>
<td>19</td>
<td>19</td>
<td>8</td>
<td>3</td>
<td>64</td>
<td>9</td>
<td>24</td>
<td>21.28</td>
<td>5.28</td>
<td>27.83</td>
<td>-2.28</td>
<td>5.18</td>
</tr>
<tr>
<td>88</td>
<td>128</td>
<td></td>
<td>232.0</td>
<td>178.0</td>
<td>153.0</td>
<td>128.00</td>
<td>100.90</td>
<td>77.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a, 2 pts) What is the slope of the "best-fit" line?
Slope = __0.659__

(b, 2 pts) What is the Y-intercept of the "best-fit" line?
Y-int. = __8.746__

(c, 2 pts) What is the correlation coefficient and coefficient?
r = __0.753__

(d, 2 pts) What is the coefficient of determination?
R² = __0.567__

(e, 2 pts) Calculate F for an ANOVA test of the slope of this data.
F = __7.852__

(f, 2 pts) What is the most precise range of p values you can provide for an ANOVA analysis of the significance of the slope? (Remember that 0 and 1.0 may be valid values for this range)
_0.025_ < p < _0.05_

(g, 2 pts) Is there a significant relationship between the X and Y values?
Circle one: Yes No.
21. ANOVA. The sizes of trees can vary based on the latitude and the amount of rainfall. Consider the hypothetical data below showing the diameters of 4 randomly chosen trees in 9 regions. The latitude varies from equatorial to mid-latitude to artic. The environment varies from xeric (dry) to mesic to hydric (wet). On this page you will perform a one-way ANOVA on this data and on the next page you will conduct a two-way ANOVA. Note: these values are reproduced on the otherwise blank last page.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Environment</td>
<td>Hydric</td>
<td>Mesic</td>
<td>Xeric</td>
<td>Hydric</td>
<td>Mesic</td>
<td>Xeric</td>
<td>Hydric</td>
<td>Mesic</td>
<td>Xeric</td>
</tr>
<tr>
<td>15</td>
<td>10</td>
<td>8</td>
<td>12</td>
<td>15</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>10</td>
<td>12</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>13</td>
<td>9</td>
<td>11</td>
<td>11</td>
<td>16</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
<td>10</td>
<td>12</td>
<td>6</td>
<td>10</td>
<td>13</td>
<td>6</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

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

$F_{\text{max}} = 13.000$  
Critical $F_{\text{max}}$ to use for $\alpha=0.05 = 93.9$

(b, 2 pts) State the statistical conclusion of your $F_{\text{max}}$ test in the box below. Use the grammar presented in lecture and be sure to 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.

None of the variances are significantly different from one another ( $p > 0.05$ )

(c, 2 pts) Based on the results of your $F_{\text{max}}$ test, should you transform this data in order to perform an ANOVA analysis?

Circle one: Yes  No.

(d, 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 (1 pt. ea):

<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>8</td>
<td>160</td>
<td>20</td>
<td>3.971</td>
<td>$p &lt; 0.025$</td>
</tr>
<tr>
<td>Within</td>
<td>27</td>
<td>136</td>
<td>5.037</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td>296</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(e, 2 pts) State the conclusions of the ANOVA in the box below. Use the grammar presented in lecture and be sure to 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.

One or more of the group means is significantly different from one or more of the others ( $p < 0.025$ ).
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>Environment</td>
<td>2</td>
<td>72</td>
<td>36</td>
<td>7.147</td>
<td>$p &lt; 0.025$</td>
</tr>
<tr>
<td>Latitude</td>
<td>2</td>
<td>24</td>
<td>12</td>
<td>2.382</td>
<td>$p &gt; 0.05$</td>
</tr>
<tr>
<td>Interact.</td>
<td>4</td>
<td>64</td>
<td>16</td>
<td>3.176</td>
<td>$0.025 &lt; p &lt; 0.05$</td>
</tr>
<tr>
<td>Within</td>
<td>27</td>
<td>136</td>
<td></td>
<td>5.037</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>35</td>
<td>296</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The environment has a significant effect on tree diameter ($p < 0.025$).

The latitude does not have a significant effect on tree diameter ($p > 0.05$).

There is a significant interaction between environment and latitude ($0.025 < p < 0.05$).

Complete the pair of interaction plots below. Clearly indicate which symbols you use to represent each latitude and environment in each figure. (Note that the order of the labels is not the same as in the data table.)
Critical F values, \( \alpha=0.05 \)

<table>
<thead>
<tr>
<th>df for numerator</th>
<th>df for denominator</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, 9.55, 9.28, 9.32, 9.01, 8.94, 8.89, 8.85, 8.81, 8.79</td>
<td>6.61, 6.57, 6.53, 6.49, 6.45, 6.41, 6.38, 6.35, 6.31, 6.28</td>
<td>6.05, 6.02, 6.00, 5.97, 5.94, 5.91, 5.88, 5.85, 5.82, 5.79</td>
<td>5.61, 5.55, 5.49, 5.43, 5.37, 5.31, 5.25, 5.19, 5.13, 5.08</td>
<td>5.00, 4.93, 4.87, 4.81, 4.75, 4.69, 4.63, 4.57, 4.50, 4.44</td>
<td>5.00, 4.93, 4.87, 4.81, 4.75, 4.69, 4.63, 4.57, 4.50, 4.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5.99, 5.94, 5.87, 5.81, 5.75, 5.70, 5.65, 5.60, 5.55, 5.50</td>
<td>5.52, 5.46, 5.40, 5.34, 5.28, 5.23, 5.18, 5.13, 5.08, 5.03</td>
<td>5.00, 4.93, 4.87, 4.81, 4.75, 4.69, 4.63, 4.57, 4.50, 4.44</td>
<td>5.00, 4.93, 4.87, 4.81, 4.75, 4.69, 4.63, 4.57, 4.50, 4.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5.59, 5.53, 5.47, 5.41, 5.36, 5.30, 5.25, 5.20, 5.15, 5.10</td>
<td>5.20, 5.13, 5.07, 5.01, 4.95, 4.90, 4.84, 4.79, 4.74, 4.69</td>
<td>5.00, 4.93, 4.87, 4.81, 4.75, 4.69, 4.63, 4.57, 4.50, 4.44</td>
<td>5.00, 4.93, 4.87, 4.81, 4.75, 4.69, 4.63, 4.57, 4.50, 4.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>5.26, 5.19, 5.13, 5.07, 5.01, 4.95, 4.89, 4.84, 4.79, 4.74</td>
<td>5.00, 4.93, 4.87, 4.81, 4.75, 4.69, 4.63, 4.57, 4.50, 4.44</td>
<td>5.00, 4.93, 4.87, 4.81, 4.75, 4.69, 4.63, 4.57, 4.50, 4.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5.00, 4.93, 4.87, 4.81, 4.75, 4.69, 4.63, 4.57, 4.50, 4.44</td>
<td>5.00, 4.93, 4.87, 4.81, 4.75, 4.69, 4.63, 4.57, 4.50, 4.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Critical F values, \( \alpha=0.025 \)

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>9</td>
<td>7</td>
<td>15</td>
<td>6</td>
<td>12</td>
<td>11</td>
<td>8</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>10</td>
<td>12</td>
<td>9</td>
<td>7</td>
<td>12</td>
<td>7</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>13</td>
<td>9</td>
<td>11</td>
<td>11</td>
<td>16</td>
<td>7</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>12</td>
<td>10</td>
<td>12</td>
<td>6</td>
<td>10</td>
<td>13</td>
<td>6</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>