MULTIPLE CHOICE (3 pts each)

(1) If you are using a table of critical t values and you accidentally use the values for more degrees of freedom than you really have which of these is true:
   (A) The risk of making a type I error is increased and the risk of type II error is increased.
   (B) The risk of making a type I error is increased and the risk of type II error is decreased.
   (C) The risk of making a type I error is decreased and the risk of type II error is increased.
   (D) The risk of making a type I error is decreased and the risk of type II error is decreased.
   (E) The risk of making a type I error is the same and the risk of type II error is the same.

(2) Which of the following is an accurate statement?
   (A) We perform t tests to decide if two samples have different variances, whereas we perform F tests to decide if two samples have different means.
   (B) We perform t tests to decide if two populations have different variances, whereas we perform F tests to decide if two populations have different means.
   (C) We perform t tests to decide if two samples have different means, whereas we perform F tests to decide if two samples have different variances.
   (D) We perform t tests to decide if two populations have different means, whereas we perform F tests to decide if two populations have different variances.
   (E) We perform t tests to decide if two samples have different means, whereas we perform F tests to decide if two populations have different variances.

(3) The best description of what a p value represents is:
   (A) The probability that H₀ is true.
   (B) The probability of seeing the sample data if H₀ is true.
   (C) The probability of seeing the sample data if H₀ is false.
   (D) The probability of seeing the sample data if Hₐ is true.
   (E) The probability of seeing the sample data if Hₐ is false.

(4) The best description of what a p value represents is:
   (A) p is the smallest α we could choose and still reject H₀.
   (B) p is the largest α we could choose and still reject H₀.
   (C) p is the smallest α we could choose and still accept H₀.
   (D) p is the largest α we could choose and still accept H₀.
   (E) p is the best α we could choose to test H₀.

(5) Which of the following is correct?
   (A) The standard deviation turns into the standard error when the sample size is very large.
   (B) The standard error measures spread of a data set whereas the standard deviation describes the width of a region within which we believe the population mean lies.
   (C) The standard error measures spread of a data set whereas the standard deviation describes the width of a region within which we believe the sample mean lies.
   (D) The standard deviation measures spread of a data set whereas the standard error describes the width of a region within which we believe the population mean lies.
   (E) The standard deviation measures spread of a data set whereas the standard error describes the width of a region within which we believe the sample mean lies.
(6) Why is it generally better to perform two-tailed heteroscedastic t tests than one-tailed heteroscedastic t tests?
(A) We can only do the one-tailed test when the variances are equal; since that is never guaranteed we should do the two-tailed test.
(B) We can only do the one-tailed test when the variances are not equal; since that is never guaranteed we should do the two-tailed test.
(C) Doing a two-tailed t test instead of a one-tailed reduces the chance of type I error.
(D) Doing a two-tailed t test instead of a one-tailed reduces the chance of type II error.
(E) Doing a two-tailed t test is more robust to outliers in our data.

(7) Which of the following is the best description of the way the technical term "statistically significant" is most commonly used?
(A) When the null hypothesis of a statistical test can be rejected with absolute certainty
(B) When the null hypothesis of a statistical test can be rejected at the $\alpha=99\%$ level.
(C) When the null hypothesis of a statistical test can be rejected at the $\alpha=95\%$ level.
(D) When the null hypothesis of a statistical test can be rejected at the $\alpha=5\%$ level.
(E) When the null hypothesis of a statistical test can be rejected at the $\alpha=1\%$ level.

(8) A researcher is interested in whether a drug alters the pH of blood in users. She takes blood samples from a set of individuals before and after taking the experimental medication. Unfortunately a small number of the sample jar labels got switched within each time period and she cannot determine which samples are from the same individuals in a few cases (the "before" and "after" information is correct). What option below best explains her next step?
(A) Because only a few labels are incorrect she can still do a paired t test, but she should use a smaller p value threshold to decide significance.
(B) She cannot do a paired t test, but since the same people are in both samples she can analyze the data with a homoscedastic t test.
(C) She cannot do a paired t test, but she can analyze the data with a heteroscedastic t test.
(D) She cannot do a t test, but she can analyze the data with an F ratio test to answer her question.
(E) Since the labels are not correct she cannot do a t test and she needs to redo the experiment.

(9) When we look at tables of t test critical values the values typically get smaller as we move down the table. Which of the following best describes why?
(A) Lower on the table the degrees of freedom increase and the t distribution becomes more normal.
(B) Lower on the table the degrees of freedom decrease and the t distribution becomes more normal.
(C) Lower on the table the $\alpha$ values increase and the t distribution becomes more normal.
(D) Lower on the table the $\alpha$ values decrease and the t distribution becomes more normal.
(E) For larger degrees of freedom the t distribution resembles the Poisson distribution.

(10) Which of the following does NOT accurately described one of the biases I described in medical drug testing is the US?
(A) More drugs are tested on adults that on children.
(B) More drugs are tested on males than on females.
(C) More drugs are tested on the young than the elderly.
(D) More drugs are tested on Caucasians than on minorities.
(E) More drugs are tested on prisoners than on other types of people.

(11) If we do a two sample t test which of the following would result in the lowest p value?
(A) The means are similar and the standard errors are low.
(B) The means are similar and the standard errors are high.
(C) The means are not similar and the standard errors are low.
(D) The means are not similar and the standard errors are high.
(E) The means and standard errors are similar to each other.
ASSIGNED WEBSITE.
The next 3 questions are based on the website assigned for this test.

(12) In class I described a question I put on last year’s exam in which I asked students to estimate my age under two different conditions (“Dr. Carter” vs “Ashley Carter”). That little test was an exercise in deliberately trying to observe the effects of which of the following?
   (A) Data dredging.    (D) Loaded questions.
   (B) Discarding unfavorable data.  (E) Overgeneralization.
   (C) False causality.

(13) Many evolutionary psychology studies are performed on undergraduate students in the US who tend to be white and from families with incomes above the median for the US. When these results are used to make statements about the evolutionary history of all humans these studies may be guilty of which misuse of statistics?
   (A) Data dredging.    (D) Loaded questions.
   (B) Discarding unfavorable data.  (E) Overgeneralization.
   (C) False causality.

(14) We now have the technology to study hundreds of thousands of genetically variable points in people's genomes (termed SNPs) quickly and affordably. A GWAS study is one in which we look at the frequencies of all these SNP genetic variants in two groups, one with a disorder known to be at least partially genetically influenced and one set of healthy controls. These studies are often successful in identifying places in the genome associated with the genetic disorder, but the results must be interpreted with caution because this method is prone to which of the following misuses of statistics?
   (A) Data dredging.    (D) Loaded questions.
   (B) Discarding unfavorable data.  (E) Overgeneralization.
   (C) False causality.

(15) The following information is true, more detail is available at:
   wikipedia.org/wiki/Kano_Trovafloxacin_trial_litigation & wikipedia.org/wiki/Trovafloxacin
   The Kano trovafloxacin legal trial arose from a clinical trial conducted by the pharmaceutical company Pfizer in Nigeria during an epidemic of meningococcal meningitis. In the clinical trial Pfizer gave 100 children its new antibiotic trovafloxacin while another 100 received the previously approved and highly effective treatment ceftriaxone, but at a substantially reduced dose. Five children given trovafloxacin and six given the reduced dose of ceftriaxone died. The participants and their families were not told that they were part of an experimental medical trial or that the charity "Doctors Without Borders" was offering the standard treatment at a full strength dose in another part of the same building.
   The clinical trial involving 200 children was done in 1996, Pfizer settled the legal case with the city of Kano, Nigeria out of court in 2009. Trovafloxacin has since been essentially pulled from use in developed countries due to the high risk of severe liver damage.
   Which of the following studies described in class is this the most similar to?
   (A) Beavis experiment.   (D) Milgram experiment.
   (B) Bradley experiment.   (E) Tuskegee experiment.
   (C) Huff experiment.

(16) If you do a statistical test and the p value is 0.06 what is your conclusion?
   (A) Reject H₀ and accept Hₐ.   (D) Accept both H₀ and Hₐ.
   (B) Accept H₀ and reject Hₐ.   (E) Reject both H₀ and Hₐ.
   (C) The difference is significant.
This is based on a true story. In my lab I am considering an experiment testing whether very dilute concentrations of alcohol dissolve cholesterol faster than pure water in which cholesterol is completely insoluble. As part of the preliminary data I will expose small crystals of cholesterol to a dilute solution of 0.05% alcohol for 48 hours. I will measure the mass of the crystals before and after; imagine that I do this and obtain the data below.

<table>
<thead>
<tr>
<th>Crystal</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>29</td>
<td>25</td>
</tr>
<tr>
<td>2</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>3</td>
<td>34</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>31</td>
<td>29</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>6</td>
<td>31</td>
<td>27</td>
</tr>
<tr>
<td>7</td>
<td>32</td>
<td>28</td>
</tr>
<tr>
<td>8</td>
<td>25</td>
<td>26</td>
</tr>
<tr>
<td>9</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>10</td>
<td>29</td>
<td>25</td>
</tr>
</tbody>
</table>

Answer the following questions and report all calculated values to the nearest 0.01.

(a, 3 pts ea) What is the 95% confidence interval of the mean size of my starting crystals?

The 95% CI is from ______ to ________

I am concerned about whether different shaped crystals act differently so I am concerned about whether the variances of the masses are the same before and after the procedure.

(b, 2 pts ea) If I were to conduct an F ratio variance test, what values would I get/use?

F_{calc} = _______  df(num) = _______  df(den) = _______

Note that I can do a one-tailed t test on my data because there is no way the crystals can get heavier, they can only be dissolved and get lighter. Conduct the appropriate test on this data using the information provided above.

(c, 3 pts ea) Fill in the blanks below. Provide the tcrit. value for determining statistical significance. Round the df value to the appropriate integer.

df = _______  t_{calc} = _______  t_{crit} = _______

(d, 3 pts) Using the data above and the results of the test, make a concise statistical statement about the results of the study in the empty box below. Include a statement about your degree of confidence using the most precise p value range you can and whether the conclusions are significant or not significant. Include reference to the masses of the crystals and what that means. Think before you write, no credit will be given for any text outside the box FOR ANY REASON.
This is based on a true story. In the fall one of my students will begin a project measuring the distance between two wing vein intersections on the wings of *Drosophila* subjected to different types of artificial selection. Some flies will be selected to have larger distances (Up line) and some will be selected to have smaller distances (Down line). After a few generations she will compare them to see if they have different mean distances (in either direction just in case).

Consider the data below for 7 individuals chosen from each line:

<table>
<thead>
<tr>
<th>indiv.</th>
<th>Up line</th>
<th>Down Line</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>33</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
<td>29</td>
</tr>
<tr>
<td>3</td>
<td>35</td>
<td>33</td>
</tr>
<tr>
<td>4</td>
<td>35</td>
<td>31</td>
</tr>
<tr>
<td>5</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>6</td>
<td>34</td>
<td>32</td>
</tr>
<tr>
<td>7</td>
<td>32</td>
<td>28</td>
</tr>
</tbody>
</table>

Conduct the appropriate test on this data to determine whether the flies in the two lines exhibit different mean wing vein intersection distances. Report all requested values to the nearest 0.01.

(a, 3 pts) In the box below describe which test you decided to do and why, support your answer. Think before you write, no credit will be given for any text outside the box FOR ANY REASON.

(b, 3 pts ea) Indicate the the degrees of freedom your test will use, the t value you calculate and the relevant $t_{\text{crit}}$ value for deciding significance:

$$ df = ________ $$
$$ t_{\text{calc}} = ________ $$
$$ t_{\text{crit}} = ________ $$

(c, 3 pts) Using the data and the results of your test, make a concise biological statement about the results of your test in the empty box below. Include a statement about your degree of confidence using the most precise p value range you can and whether the conclusions are significant or not significant and in what manner. Think before you write, no credit will be given for any text outside the box FOR ANY REASON.
(19) Students often ask for their exams to be "curved" which would technically mean distributing the grades as a normal distribution centered around 75% so that there are equal numbers of "B" and "D" grades and "A" and "F" grades. If we centered a normal distribution around a score of 75% and gave "C" grades to the middle region of width one standard deviation (i.e., from $Z = -0.5$ to $Z = 0.5$) and the other letter grades to the similar regions of each side (i.e., $Z < -1.5$ for "F", $Z = -1.5$ to 0.5 for "D", $Z = 0.5$ to 1.5 for "B", and $> 1.5$ for "F") we would expect to give approximately 7% "A" and "F" grades, 24% "D" and "B" grades and 38% "C" grades.

In the spring of 2013, the 136 final student grades in Bio 260 were: 6 F 15 D 42 C 50 B 23 A

Perform a $X^2$ analysis to determine whether the grades I assigned in the spring appear to fit the "curved" distribution or not. This will be a "Goodness of Fit" test comparing the observed values (the ones I gave) to the number of values predicted from the description of scaling by using the normal distribution above.

(a, 2 pts) How many degrees of freedom will your test have? df = ______

(b, 1 pt ea) How many grades of each type would you expect to see if "curving" was done as described above? Report all these values to the nearest 0.01

A: ______  B: ______  C: ______  D: ______  F: ______

(c, 5 pts) What is your $X^2$ calculated value $X^2 = ______$

(d, 3 pts) Using the results of your test, make a concise statement about the results of your analysis in the empty box below. Include a statement about your degree of confidence using the most precise p value range you can and whether the conclusions are significant or not significant. Think before you write, no credit will be given for any text outside the box FOR ANY REASON.
TABLE OF $t_\alpha$ VALUES

Table shows the $t$ values corresponding to the indicated critical $\alpha$ value.

<table>
<thead>
<tr>
<th>df</th>
<th>$t_{0.1}$</th>
<th>$t_{0.05}$</th>
<th>$t_{0.025}$</th>
<th>$t_{0.02}$</th>
<th>$t_{0.01}$</th>
<th>$t_{0.005}$</th>
<th>$t_{0.0025}$</th>
<th>$t_{0.001}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.078</td>
<td>6.314</td>
<td>12.710</td>
<td>15.890</td>
<td>31.820</td>
<td>63.660</td>
<td>127.300</td>
<td>318.300</td>
</tr>
<tr>
<td>3</td>
<td>1.638</td>
<td>2.353</td>
<td>3.182</td>
<td>4.482</td>
<td>5.414</td>
<td>7.841</td>
<td>10.210</td>
<td>12.702</td>
</tr>
<tr>
<td>4</td>
<td>1.533</td>
<td>2.132</td>
<td>2.776</td>
<td>3.747</td>
<td>4.604</td>
<td>5.959</td>
<td>7.173</td>
<td>8.417</td>
</tr>
<tr>
<td>5</td>
<td>1.476</td>
<td>2.015</td>
<td>2.571</td>
<td>3.365</td>
<td>4.032</td>
<td>4.773</td>
<td>5.893</td>
<td>7.042</td>
</tr>
<tr>
<td>7</td>
<td>1.415</td>
<td>1.895</td>
<td>2.365</td>
<td>2.998</td>
<td>3.499</td>
<td>4.029</td>
<td>4.785</td>
<td>5.409</td>
</tr>
<tr>
<td>8</td>
<td>1.397</td>
<td>1.860</td>
<td>2.306</td>
<td>2.896</td>
<td>3.355</td>
<td>3.833</td>
<td>4.501</td>
<td>5.041</td>
</tr>
<tr>
<td>9</td>
<td>1.383</td>
<td>1.833</td>
<td>2.262</td>
<td>2.821</td>
<td>3.250</td>
<td>3.690</td>
<td>4.297</td>
<td>4.780</td>
</tr>
<tr>
<td>10</td>
<td>1.372</td>
<td>1.812</td>
<td>2.228</td>
<td>2.764</td>
<td>3.189</td>
<td>3.625</td>
<td>4.144</td>
<td>4.502</td>
</tr>
</tbody>
</table>

Table of critical $X^2$ values

<table>
<thead>
<tr>
<th>df</th>
<th>0.05</th>
<th>0.01</th>
<th>0.001</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.84</td>
<td>6.64</td>
<td>10.83</td>
</tr>
<tr>
<td>2</td>
<td>5.99</td>
<td>9.21</td>
<td>13.82</td>
</tr>
<tr>
<td>3</td>
<td>7.82</td>
<td>11.35</td>
<td>16.27</td>
</tr>
<tr>
<td>4</td>
<td>9.49</td>
<td>13.28</td>
<td>18.47</td>
</tr>
<tr>
<td>5</td>
<td>11.07</td>
<td>15.09</td>
<td>20.52</td>
</tr>
<tr>
<td>6</td>
<td>12.59</td>
<td>16.81</td>
<td>22.46</td>
</tr>
<tr>
<td>7</td>
<td>14.07</td>
<td>18.48</td>
<td>24.32</td>
</tr>
<tr>
<td>8</td>
<td>15.51</td>
<td>20.09</td>
<td>26.13</td>
</tr>
<tr>
<td>9</td>
<td>16.92</td>
<td>21.67</td>
<td>27.88</td>
</tr>
<tr>
<td>10</td>
<td>18.31</td>
<td>23.21</td>
<td>29.59</td>
</tr>
<tr>
<td>11</td>
<td>19.68</td>
<td>24.73</td>
<td>32.12</td>
</tr>
<tr>
<td>12</td>
<td>21.03</td>
<td>26.22</td>
<td>32.91</td>
</tr>
<tr>
<td>13</td>
<td>22.36</td>
<td>27.69</td>
<td>34.53</td>
</tr>
<tr>
<td>14</td>
<td>23.69</td>
<td>29.14</td>
<td>36.12</td>
</tr>
<tr>
<td>15</td>
<td>25.00</td>
<td>30.58</td>
<td>37.70</td>
</tr>
</tbody>
</table>

$\alpha$ is the critical value for the $t$ distribution.

$t_\alpha$ is the $t$ value corresponding to the indicated critical $\alpha$ value.