MULTIPLE CHOICE (3 pts each)  

**The exam has 120 points total**

(1) The development and implementation of very strict informed consent processes for studies with human subjects was most strongly spurred by which of the following events?  
(A) The exposure of the Tuskegee experiment's deception of the study subjects.  
(B) The Milgram experiment exposing the degree to which people would commit crimes against study subjects.  
(C) The Talking Hans experiments’ demonstration of how pervasive the placebo effect is and how powerful its effect on people’s health can be.  
(D) The study showing how facial hair influences the deposition of bacterial cells.  
(E) The study showing how monkeys floss using human hair.

(2) There is a famous French saying, "plus ça change, plus c'est la même chose" which translates as, "the more things change, the more they stay the same." One way to interpret this concept is to realize that as things change we don’t realize it since our reference changes as well. Which of the following concepts mentioned in class most closely parallels this idea?  
(A) Cherry-picking.  
(B) Confirmation bias.  
(C) Placebo effect.  
(D) Shifting baselines.  
(E) Statistical significance.

(3) "With great power comes great responsibility" was the message most closely associated with which of the concepts or examples from class?  
(A) The Bevis effect.  
(B) The placebo effect.  
(C) Shifting baselines  
(D) The file-drawer problem.  
(E) The Milgram experiment.

(4) Why is the F test not recommended for comparing variances?  
(A) Highly different means can lead to large calculated F values.  
(B) Highly different variances can lead to large calculated F values.  
(C) It is prone to inaccurate results if the data is non-normally distributed.  
(D) It is prone to inaccurate results if the variances are large.  
(E) The test statistic is skewed when the variances differ greatly.

(5) If you do a statistical test and the p value is 0.04 what is your conclusion?  
(A) Accept H₀ and reject Hₐ.  
(B) Accept both H₀ and Hₐ.  
(C) The differences are not significant.  
(D) Reject H₀ and accept Hₐ.  
(E) Reject both H₀ and Hₐ.

(6) Which of the following is the best description of what it means when someone says that the results of their two-sample t test are "statistically significant"?  
(A) There is less than a 5% chance that the means of two populations are the same.  
(B) There is less than a 5% chance that the means of two populations are different.  
(C) There is more than a 5% chance that the means of two populations are the same.  
(D) There is more than a 5% chance that the means of two populations are different.  
(E) There is an important difference between the two population means.
ASSIGNED WEBSITE.
*The next 3 questions are based on the website assigned for this test.*

(7) Consider the following possible poll question:

*"When thinking about president Obama and his recent expensive and inefficient implementation of the Affordable Care Act ("Obamacare"), would you rate his overall performance as president as good or bad?"*

Reporting values from this would be an example of which of the following misuses of statistics?
(A) Data manipulation.  
(B) Discarding unfavorable data.  
(C) Loaded questions.

(8) The website described a misuse called "discarding unfavorable data" which I described in class with a different name. The name given in class was which of the following?
(A) Data dredging.  
(B) Data manipulation.  
(C) False causality.

(9) Which of the following misuses of statistics is possible when we make type II errors, but not if we make type I errors?
(A) Biased samples.  
(B) Data dredging.  
(C) False causality.

(10) Consider a situation in which there is a population with a mean of 30 and a second population with a mean that differs. Both populations have a standard deviation of 4. If we took two samples of size 11 from each population, which of the values below is closest to the minimum magnitude of the difference between the means of the two populations that we would be able to detect with a two-tailed homoscedastic t test?
(A) 1.6  
(B) 2.6  
(C) 3.6  
(D) 4.6  
(E) 5.6

(11) Consider a situation in which two populations differ by 20% and their standard deviations are approximately 30% of their means. If we take two identically sized samples from each population, which of the values below is closest to the minimum size of each sample we would need to take in order to correctly detect this difference with a two-tailed homoscedastic t test?
(A) 5  
(B) 10  
(C) 15  
(D) 20  
(E) 25

(12) Consider the previous question. If we collected a smaller set of sample values and failed to detect the difference that was present, what type of error have we committed?
(A) Type I.  
(B) Type II.  
(C) Inferential.

(13) Which of the following is the best conceptual description of what a 95% confidence interval represents?
(A) A region where we think the population's mean value is.  
(B) A region where we think the sample's mean value is.  
(C) A region containing most of the population data values.  
(D) A region containing most of the sample data values.  
(E) A region where we think the calculated t value probably is.
(14) The best description of what a p value represents is:
(A) The probability that $H_0$ is true.
(B) The probability of seeing the sample data if $H_0$ is true.
(C) The probability of seeing the sample data if $H_0$ is false.
(D) The probability of seeing the sample data if $H_A$ is true.
(E) The probability of seeing the sample data if $H_A$ is false.

(15) Which of the following tends to result in lower p values when doing a t test comparing two samples from two different populations?
(A) Smaller calculated t values.
(B) Smaller differences between means.
(C) Smaller sample means.
(D) Smaller sample sizes.
(E) Smaller sample variances.

(16) Why is it generally better to perform two-tailed heteroscedastic t tests than one-tailed heteroscedastic t tests?
(A) Doing a one-tailed t test is more robust to outliers in our data.
(B) Doing a two-tailed t test instead of a one-tailed reduces the chance of type I error.
(C) Doing a two-tailed t test instead of a one-tailed reduces the chance of type II error.
(D) One-tailed tests are only appropriate when the variances are equal; since that is never guaranteed we should do the two-tailed test.
(E) One-tailed tests are only appropriate when the variances are not equal; since that is never guaranteed we should do the two-tailed test.

(17) Why do values in the t table decrease as you move downwards?
(A) The sample size is getting smaller.
(B) The t distribution narrows to resemble the normal distribution.
(C) The t distribution widens to resemble the normal distribution.
(D) The standard deviation is getting smaller.
(E) The standard deviation is getting larger.

(18) FDA approval requires that to be a "drug" a substance must:
(A) Be safe and 100% harmless.
(B) Be effective and 100% harmless.
(C) Be artificial and 100% effective.
(D) Be safe and relatively harmless.
(E) Be effective and relatively harmless.

(19) Which of the following was a piece of advice I gave during the discussion of making good graphs and figures?
(A) Do not use background images behind scatterplots.
(B) Use sans serif fonts whenever possible for clarity.
(C) Use the color red to illustrate larger values and the color blue to represent smaller values.
(D) Use solid symbols instead of open ones in figures.
(E) When using 3D pie charts, rotate the values you want to minimize to the front.

(20) The figure from class showing that California appears to be a deathtrap due to hunting from serial killers had several misleading aspects; which of the following was NOT one of them?
(A) The use of black and white instead of color.
(B) The incorporation of historical behaviors that are irrelevant today.
(C) The lack of accounting for the historical population size of the states.
(D) The lack of accounting for the modern population size of the states.
(E) The sudden change from stippling or parallel lines to a much more intense solid black as the magnitude changed.
(21) Consider a situation in which hospital researchers are trying to determine the relative effectiveness of two different treatments for a fictional condition called "Spectrox Toxaemia". The first treatment (medical) involves providing the patient with a drug approved for the condition. The second treatment (surgical) involves performing an operation on the patient. They conduct a non-blinded, randomized, trial (not blinded because everyone know which procedure is taken for each patient and randomized since patients will be assigned into two groups randomly) and record the change in CBT (a fictional health statistic where higher values are better) for each patient. The data for the two sets of patients in their trial is given below along with the mean, variance, and standard deviation of the values in each set.

Changes in CBT:

<table>
<thead>
<tr>
<th>Medical</th>
<th>Surgical</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>13</td>
</tr>
<tr>
<td>17</td>
<td>16</td>
</tr>
<tr>
<td>24</td>
<td>20</td>
</tr>
<tr>
<td>22</td>
<td>19</td>
</tr>
<tr>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>21</td>
<td>14</td>
</tr>
<tr>
<td>21</td>
<td>17</td>
</tr>
<tr>
<td>24</td>
<td>19</td>
</tr>
<tr>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>17</td>
<td>14</td>
</tr>
</tbody>
</table>

Mean: 20 17
Var: 8.6 8.2
SD: 2.9326 2.8636

(a, 2 pts each) Calculate the 95% confidence interval for the improvement in the CBT scores for the medical intervention and provide the lower and upper values of this interval to the right.

Upper value = __21.970__
Lower value = __18.030__

(b, 3 pts each) The pharmaceutical company that manufactures the drug reports that use of the drug results in a mean improvement of CBT value of 22.2. Perform a one-sample t test comparing the changes seen in this experiment to those expected. Fill in the blanks below with the calculated t value and the critical value you would use to reject the null hypothesis at an overall \( \alpha \) value of 5%.

\[ t_{\text{calc}} = -2.488 \quad t_{\text{crit}} (\alpha=0.05) = 2.228 \]

(c, 2 pts each) Calculate the 99% confidence interval for the improvement in the CBT scores for the medical intervention and provide the lower and upper values of this interval to the right.

Upper value = __22.802__
Lower value = __17.198__

(d, 3 pts each) The researchers conducted a literature search and found a paper describing the results of a separate study of the surgery which reported a mean improvement of CBT value of 18.3. Perform a one-sample t test comparing the changes seen in this experiment to those expected. Fill in the blanks below with the calculated t value and the critical value you would use to reject the null hypothesis at an overall \( \alpha \) value of 5%.

\[ t_{\text{calc}} = -1.506 \quad t_{\text{crit}} (\alpha=0.05) = 2.228 \]
To compare the two treatments to one another the researchers would need to do an unpaired t test.

(e, 2 pts each) Fill in the blanks below to correctly describe the next step in their analysis.

To know whether they could do either of the two tests they must perform an F variance ratio test. If they accept the F test null hypothesis they can do the __________ t test, whereas if they reject the null hypothesis they can only do the __________ t test.

(f, 3 pts each) Conduct an F variance ratio test and fill in the blanks with the values indicated. Report all calculated values to the nearest 0.01. Tables of F critical values are on the last page.

\[ F_{\text{calc}} = 1.049 \]
\[ F_{\text{crit}} (\alpha = 0.05) = 3.72 \]

Now conduct a two-sample t test comparing the two treatments. You may use the df appropriate for the t test with equal variances.

First, you will provide a technical description of your result
Then you will interpret the results of your test in plain English.

(g, 6 pts) State the technical conclusion of your t test in the box below. Your answer must follow the grammar described in class and NOT have extraneous wording or terms. You must correctly use either the phrase “significantly higher” or “no significant difference” and provide the most precise range of p values that your test results allow.

Note: no credit will be given for ANY text outside the box.

**The mean change in CBT value for the medical treatment is significantly higher than the mean change for the surgery treatment (0.02 < p < 0.04)**

(h, 4 pts) What is the conclusion the hospital researchers will take from their study? I.e., what recommendation would they give to the hospital administration or their patients regarding treatment of Spectrox Toxaemia? State the plain language version of your conclusion in the box below. Do not use statistical or extraneous terms.

Note: no credit will be given for ANY text outside the box.

**Patients with Spectrox Toxaemia should be administered the medical treatment instead of the surgical one.**
Callitrichines are a subfamily of New World monkeys which includes marmosets and tamarins (pictured). These species often live in very social groups and cooperatively raise their offspring which may be one reason why twins are very common. Behavioral experiments on primates are both potentially very informative (they are very close biological relatives) and ethically challenging (due to their relatively high intellect and capacity for suffering we require strict ethical guidelines as enforced by a review board and set of regulations termed the IACUC procedure). Since this is a biostats course we will see twinning in tamarins this as a useful feature that would allow us to do paired t tests on captive-born tamarins if the two individuals from each set of parents are placed into different treatments.

Imagine that we provide each twin from 11 sets of parents with one of two different training regimes; one is provided with an hour of free play with toys in isolation each day while the other works with a researcher to learn specific tasks and tricks for an hour each day. After several months of these different treatments, the individuals are introduced into a room by themselves with a novel puzzle which contains food. The puzzle can be opened by the monkeys and the time taken from their entrance into the room until they obtain the food is recorded.

Consider the two sets of time values below:

<table>
<thead>
<tr>
<th>Twin pair:</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>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solo play</td>
<td>29</td>
<td>26</td>
<td>34</td>
<td>30</td>
<td>27</td>
<td>29</td>
<td>31</td>
<td>25</td>
<td>31</td>
<td>32</td>
<td>25</td>
</tr>
<tr>
<td>Training sessions</td>
<td>24</td>
<td>24</td>
<td>29</td>
<td>28</td>
<td>29</td>
<td>29</td>
<td>31</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>25</td>
</tr>
</tbody>
</table>

Conduct a paired t test comparing the two treatments.

(a, 5 pts each) Fill in the blanks below with the calculated t value and the critical value you would use to reject the null hypothesis at an overall $\alpha$ value of 5%.

$$t_{\text{calc}} = 2.544 \quad t_{\text{crit}} (\alpha=0.05) = 2.228$$

(b, 5 pts) State the technical conclusion of your t test in the box below. Your answer must follow the grammar described in class and NOT have extraneous wording or terms. You must correctly use either the phrase "significantly more" or "no significant difference" and provide the most precise range of p values that your test results allow. Note: no credit will be given for ANY text outside the box.

**The monkeys in the solo play treatment took significantly more minutes to solve the puzzle than the ones that had the training sessions (0.02 < p < 0.04).**

(c, 5 pts) What is the conclusion that this data would cause us to make? I.e., what is the effect of free play versus intensive training on the ability to solve a novel problem in tamarins? State the plain language version of your conclusion in the box below. Do not use statistical or extraneous terms. Note: no credit will be given for ANY text outside the box.

**Training sessions seem to improve problem solving abilities more than solo play.**
**TABLES OF \( t_\alpha \) and \( F_\alpha \) 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_{0.01} )</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>63.60</td>
<td>127.30</td>
</tr>
<tr>
<td>3</td>
<td>1.638</td>
<td>2.353</td>
<td>3.182</td>
<td>3.482</td>
<td>4.541</td>
<td>5.841</td>
<td>7.453</td>
</tr>
<tr>
<td>4</td>
<td>1.533</td>
<td>2.132</td>
<td>2.776</td>
<td>2.999</td>
<td>3.747</td>
<td>4.604</td>
<td>5.598</td>
</tr>
<tr>
<td>5</td>
<td>1.476</td>
<td>2.015</td>
<td>2.571</td>
<td>2.757</td>
<td>3.365</td>
<td>4.032</td>
<td>4.773</td>
</tr>
<tr>
<td>6</td>
<td>1.440</td>
<td>1.943</td>
<td>2.447</td>
<td>2.612</td>
<td>3.143</td>
<td>3.707</td>
<td>4.317</td>
</tr>
<tr>
<td>7</td>
<td>1.415</td>
<td>1.895</td>
<td>2.365</td>
<td>2.517</td>
<td>2.998</td>
<td>3.499</td>
<td>4.029</td>
</tr>
<tr>
<td>9</td>
<td>1.383</td>
<td>1.833</td>
<td>2.262</td>
<td>2.398</td>
<td>2.821</td>
<td>3.250</td>
<td>3.690</td>
</tr>
<tr>
<td>10</td>
<td>1.372</td>
<td>1.812</td>
<td>2.228</td>
<td>2.359</td>
<td>2.764</td>
<td>3.169</td>
<td>3.581</td>
</tr>
<tr>
<td>11</td>
<td>1.363</td>
<td>1.796</td>
<td>2.201</td>
<td>2.328</td>
<td>2.718</td>
<td>3.106</td>
<td>3.497</td>
</tr>
<tr>
<td>12</td>
<td>1.356</td>
<td>1.782</td>
<td>2.179</td>
<td>2.303</td>
<td>2.681</td>
<td>3.055</td>
<td>3.428</td>
</tr>
<tr>
<td>14</td>
<td>1.345</td>
<td>1.761</td>
<td>2.145</td>
<td>2.264</td>
<td>2.624</td>
<td>2.977</td>
<td>3.326</td>
</tr>
<tr>
<td>15</td>
<td>1.341</td>
<td>1.753</td>
<td>2.131</td>
<td>2.249</td>
<td>2.602</td>
<td>2.947</td>
<td>3.286</td>
</tr>
<tr>
<td>16</td>
<td>1.337</td>
<td>1.746</td>
<td>2.120</td>
<td>2.235</td>
<td>2.583</td>
<td>2.921</td>
<td>3.252</td>
</tr>
<tr>
<td>17</td>
<td>1.333</td>
<td>1.740</td>
<td>2.110</td>
<td>2.224</td>
<td>2.567</td>
<td>2.898</td>
<td>3.222</td>
</tr>
<tr>
<td>18</td>
<td>1.330</td>
<td>1.734</td>
<td>2.101</td>
<td>2.214</td>
<td>2.552</td>
<td>2.878</td>
<td>3.197</td>
</tr>
<tr>
<td>19</td>
<td>1.328</td>
<td>1.729</td>
<td>2.093</td>
<td>2.205</td>
<td>2.539</td>
<td>2.861</td>
<td>3.174</td>
</tr>
<tr>
<td>20</td>
<td>1.325</td>
<td>1.725</td>
<td>2.086</td>
<td>2.197</td>
<td>2.528</td>
<td>2.845</td>
<td>3.153</td>
</tr>
<tr>
<td>21</td>
<td>1.323</td>
<td>1.721</td>
<td>2.080</td>
<td>2.189</td>
<td>2.518</td>
<td>2.831</td>
<td>3.135</td>
</tr>
<tr>
<td>22</td>
<td>1.321</td>
<td>1.717</td>
<td>2.074</td>
<td>2.183</td>
<td>2.508</td>
<td>2.819</td>
<td>3.119</td>
</tr>
<tr>
<td>23</td>
<td>1.319</td>
<td>1.714</td>
<td>2.069</td>
<td>2.177</td>
<td>2.500</td>
<td>2.807</td>
<td>3.104</td>
</tr>
<tr>
<td>24</td>
<td>1.318</td>
<td>1.711</td>
<td>2.064</td>
<td>2.172</td>
<td>2.492</td>
<td>2.797</td>
<td>3.091</td>
</tr>
<tr>
<td>25</td>
<td>1.316</td>
<td>1.708</td>
<td>2.060</td>
<td>2.167</td>
<td>2.485</td>
<td>2.787</td>
<td>3.078</td>
</tr>
<tr>
<td>26</td>
<td>1.315</td>
<td>1.706</td>
<td>2.056</td>
<td>2.162</td>
<td>2.479</td>
<td>2.779</td>
<td>3.067</td>
</tr>
<tr>
<td>27</td>
<td>1.314</td>
<td>1.703</td>
<td>2.052</td>
<td>2.150</td>
<td>2.473</td>
<td>2.771</td>
<td>3.057</td>
</tr>
<tr>
<td>28</td>
<td>1.313</td>
<td>1.701</td>
<td>2.048</td>
<td>2.154</td>
<td>2.467</td>
<td>2.763</td>
<td>3.047</td>
</tr>
<tr>
<td>29</td>
<td>1.311</td>
<td>1.699</td>
<td>2.045</td>
<td>2.150</td>
<td>2.462</td>
<td>2.756</td>
<td>3.038</td>
</tr>
<tr>
<td>30</td>
<td>1.310</td>
<td>1.697</td>
<td>2.042</td>
<td>2.147</td>
<td>2.457</td>
<td>2.750</td>
<td>3.030</td>
</tr>
<tr>
<td>40</td>
<td>1.303</td>
<td>1.684</td>
<td>2.021</td>
<td>2.123</td>
<td>2.423</td>
<td>2.704</td>
<td>2.971</td>
</tr>
<tr>
<td>50</td>
<td>1.295</td>
<td>1.676</td>
<td>2.009</td>
<td>2.109</td>
<td>2.403</td>
<td>2.678</td>
<td>2.937</td>
</tr>
<tr>
<td>60</td>
<td>1.296</td>
<td>1.671</td>
<td>2.000</td>
<td>2.099</td>
<td>2.390</td>
<td>2.660</td>
<td>2.915</td>
</tr>
<tr>
<td>80</td>
<td>1.292</td>
<td>1.664</td>
<td>1.990</td>
<td>2.088</td>
<td>2.374</td>
<td>2.639</td>
<td>2.887</td>
</tr>
<tr>
<td>100</td>
<td>1.290</td>
<td>1.660</td>
<td>1.984</td>
<td>2.081</td>
<td>2.364</td>
<td>2.626</td>
<td>2.871</td>
</tr>
<tr>
<td>inf.</td>
<td>1.282</td>
<td>1.646</td>
<td>1.962</td>
<td>2.056</td>
<td>2.330</td>
<td>2.581</td>
<td>2.813</td>
</tr>
<tr>
<td></td>
<td>1.282</td>
<td>1.640</td>
<td>1.960</td>
<td>2.054</td>
<td>2.326</td>
<td>2.576</td>
<td>2.807</td>
</tr>
</tbody>
</table>