Statistics Problem Set 2 (45 pt) – To be done individually
Hypothesis tests for comparisons of two samples: Two-sample t-test and Mann-Whitney U test
(Reserve Reading: Townend 2002; Ch 2, 6, & 7)

CONTEXT: This problem concerns a previous Bio 270 study of the white pine (Pinus strobus) population at Range Ponds State Park (RPSP) in Poland, Maine. A primary focus of that study was to characterize the age and size structure, and spacing of the pine population.

In this exercise, the goal is first to summarize two sample data sets (calculate the descriptive statistics) and then to compare the two samples for differences in the means using an inferential statistical test. The data sets were collected from a non-harvested area (Not cut = control) and one taken in a close-by, harvested area (Cut = treatment) to determine what effect the harvest had on the population characteristics. Each student will be assigned one of four variables to do: age, stem diameter at breast height (DBH), stem height, or stem spacing as average nearest neighbor distance (NND) based on your last name first letter.

Data: The data can be accessed as Excel files via a link on the Bio 270 webpage:
http://abacus.bates.edu/~ganderso/biology/bio270/270supplements.html

1. Open Prism, and create a new table by selecting Column, and choose a graph type that you think will be appropriate for showing the descriptive statistics.
2. Click on the link to your assigned data set and copy/paste the data into Prism.
3. Save the Prism file using the filename format yourlastnamehere-stats2.pzf, where you put in your last name as indicated.

Analysis: Assigned variable: ________________________

1. Summarize the data: Calculate the descriptive statistics for each sample using the Analyze: Column Analyses; Column Statistics command. Select the D’Agostino and Pearson omnibus test for normality, too, to test each group of data for fit with a normal distribution.

2. Record the appropriate values below that describe the central tendency and variability in each sample. (5 pt)

<table>
<thead>
<tr>
<th>Plot</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>SEM</th>
<th>25th &amp; 75th percentiles</th>
<th>n</th>
<th>Range (max-min)</th>
<th>min and max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-cut</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cut</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Examine the descriptive statistics for obvious patterns. Do the samples the two plots appear to be different? Why or why not? (3 pt)

_____________________________________________________________________________________

4. Graph the descriptive statistics to visualize differences: Prism made an initial graph of the data for you when you entered the data. Look at that graph now. Based on the graphical presentation of the descriptive statistics, do you think there may be a significant difference in the means between the two plots? Why or why not? (2 pt)

_____________________________________________________________________________________

_____________________________________________________________________________________

Stats Problem Set 2: 2016
5. Since we are interested in determining whether or not the population characteristic has changed with respect to this variable, write down the Null and Alternate hypotheses that we could test: (4 pts)

NULL (H₀):

ALTERNATE (H₁):

6. Use the flow chart provided to determine which hypothesis test you should use to test your Null hypothesis. Which test will you use, and why? (2 pt)

7. List the three assumptions about the data that must be met to appropriately use a parametric test and indicate whether the assumption has been satisfied (circle Yes or No). (6 pt – 1 pt each)

   a. __________________________________________________________________________ Yes No
   b. __________________________________________________________________________ Yes No
   c. __________________________________________________________________________ Yes No

8. If any of these assumptions (7) about the data for using the parametric test are not met, what kind of data manipulation could you perform to alter the characteristics of the data distribution to satisfy the assumptions and be able to use the parametric test? (We’re looking for a general action here, not a specific.) (4 pt – 2 pt each)

   a. __________________________________________________________________________
   b. If the action you suggest in “8.a.” fails to resolve the issue, what is the fallback option for doing a hypothesis test?

9. Based on #7 and #8, perform the hypothesis test you deem most appropriate and summarize the outcome below. (5 pt)

   Name of test used: ____________________________ (1 pt)

<table>
<thead>
<tr>
<th>Test statistic (1/2 pt)</th>
<th>Degrees of freedom (1/2 pt)</th>
<th>P value (1/2 pt)</th>
<th>Reject or fail to reject the Null (1/2 pt)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Briefly define in words what the p value tells us in the context of the Null hypothesis (NOTE: not just significant. vs. not significant.): (2 pt)

“The p-value is the probability that ____________________________ ”

10. Write a brief results statement (as it might appear in a Results section) which includes the biological result, the parenthetical stat summary, and reference to a figure in which the result is shown (call it Fig. 1). (4 pt)

11. In Prism, enter the data into a Column format table and make a bar graph comparing the two sample means (include SD or SEM error bars), write an appropriate figure legend (call it Figure 1), print it, and then STAPLE to this homework. NOTE: If your data required use of a non-parametric test, change the graph type to a dot plot (see image at right) showing the median with interquartile ranges. (10 pt)
Flow Chart for Selecting Commonly Used Statistical Tests

**Parametric Assumptions:**
1. Independent, unbiased samples
2. Data normally distributed
3. Equal variances

**Type of data?**
- Continuous
- Discrete, categorical

**Type of question**
- Relationships
- Differences

**Do you have a true independent variable?**
- Yes: Regression Analyses
- No: Correlation Analysis

**Correlation Analysis**
- Parametric: Pearson’s r
- Nonparametric: Spearman’s Rank Correlation

**Differences between what?**
- Means

**Tests for Equal Variances**
- $F_{max}$ test, Brown and Smythe’s test, Bartlett’s tests

**How many treatment groups?**
- Two groups
- More than two groups

**Parametric assumptions satisfied?**
- Yes: Student’s unpaired t-test, Paired t-test
- No: Mann-Whitney U or Wilcoxon Rank sums test

**If significant, do a post hoc test, e.g., Tukey’s or Bonferroni’s**

**Data transform worked?**
- Yes: ANOVA
- No: Nonparametric

**Parametric**
- ANOVA
- Kruskal-Wallis Test

**If significant, do a Dunn’s Test**