MES – Quantitative and Qualitative Methods
Spring 2010 Quiz – due 6pm, in lab, Tuesday, Week 10, May 31.

Name: _____ Faculty: _____

This quiz is open-book, open-notes, and open computer; you have 3 hours to complete it. This is an individual effort; do NOT talk about the exam with others until after exam is due. If you use sources outside your own head, or quote directly from the texts or tutorial, cite them. You may complete the exam by hand or word processor, but if the latter, bring hard copy to lab.

(1) (40%) For each of the following scenarios, (a) state the null and alternative hypotheses, (b) indicate which statistical test you would use to analyze the data (assuming an underlying normal distribution), and (c) briefly describe the consequence of committing Type I and Type II errors.

(A) Dr. Sun takes 100 random measurements of the dissolved oxygen in the water in Budd Inlet and Dr. Star takes 100 random measurements of the dissolved oxygen levels in Case Inlet. Since dissolved oxygen levels are known to greatly stress the sea life in Puget Sound, the State has instituted a program that will give a $100,000 grant for further study of the impact of dissolved oxygen levels on the fish in the area with the lower levels. Does the water in Budd Inlet have a significantly lower dissolved oxygen level than the water in Case Inlet?

- Ho:
- Ha:
- statistical test you would use:
- consequence of committing Type I error:
- consequence of committing Type II error:

(B) Jane seeks to understand if a proposed slug education program at the local grade school will affect students' attitudes towards the slimy creatures. She administers a test about the slugs' life cycle and role in the ecosystem to a random selection of 25 students at Insect Elementary. After the students have completed the education program, Jane will again test their understanding of slugs and wants to compare the scores.

- Ho:
- Ha:
- statistical test you would use:
- consequence of committing Type I error:
- consequence of committing Type II error:

(C) John has been investigating the impact of changing climate on the health of the Western Red Cedar trees in the Pacific Northwest. He has developed a composite measure of tree health (called "HEALTH") that combines data about tree diameter, tree height, and tree growth rate (as estimated from the rings in a core sample). He also has collected data at a 100 forest locations on the site elevation, mean temperature, average rainfall, average snow-pack depth, and soil pH. He hopes to find out which set of conditions best predicts the health of the Western Red Cedar and thus assist the National Forest Service in choosing sites for new stand establishment.

- Ho:
- Ha:
- statistical test you would use:
- consequence of committing Type I error:
- consequence of committing Type II error:

(D) Dale has run analyses to determine the carbon content of leaf litter, on six specimens, using three different types of equipment. His school currently has equipment to measure photometry, but wishes to know whether the other two instruments are comparable. The outcome of this research will determine what new equipment the school will purchase.

Specimen	Photometry	Spectometry	Absorption
1	0.96	0.94	0.98
2	0.96	0.98	1.01
3	0.85	0.87	0.86
4	0.86	0.84	0.9
5	0.86	0.87	0.89
6	0.89	0.93	0.92

- Ho:
- Ha:
- statistical test you would use:
- consequence of committing Type I error:
- consequence of committing Type II error:

(2) (40%) Interpret the results of the following R data analyses.

(A) Three factors (the level of acetic acid, the amount of H2S, and the level of lactic acid) were expected to affect the taste of cheese (adapted from the Qualitative and Quantitative Data Analysis Final Exam, 2007).

- In the R output below, circle the factors that significantly contribute to taste. If you are doing this electronically, highlight those factors in yellow.
- How did you determine which were significant contributors?

- Write out the regression model (equation) for taste (using words for the variable names).

- Is the model chosen a valid one? How do you know?

Call:
lm(formula = taste$Taste ~ taste$Acetic.Acid + taste$H2S + taste$Lactic.Acid)

R Output:

Residuals:

Min	1Q	Median	3Q	Max
-17.391	-6.612	-1.009	4.908	25.449

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-28.8768	19.7354	-1.463	0.15540
taste$Acetic.Acid	0.3277	4.4598	0.073	0.94198
taste$H2S	3.9118	1.2484	3.133	0.00425 **
taste$Lactic.Acid	19.6705	8.6291	2.280	0.03108 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.13 on 26 degrees of freedom
Multiple R-squared: 0.6518, Adjusted R-squared: 0.6116
F-statistic: 16.22 on 3 and 26 DF, p-value: 3.81e-06

(B) The local utility wants to determine if there is enough biomass in adjacent forests to support a biomass gasification project. They engage two separate firms to randomly sample the forests and determine the tonnage of biomass that might be used for that project.

- Did the two firms differ significantly in their estimates of the amount of biomass? How do you know?

Call:
t.test(count$Mass~count$Firm)

R Output:

```
        Welch Two Sample t-test
             :  count$Mass by count$Firm

    t = -1.9554, df = 36.437, p-value = 0.05824
    alternative hypothesis: true difference in means is not equal to 0
    95 percent confidence interval:
     -117.419392    2.119392
    sample estimates:
    mean in group 1           mean in group 2
            401.15                    458.80
```

(C) Can you use parametric statistics to test the data depicted in Figure 1 below? Why or why not?

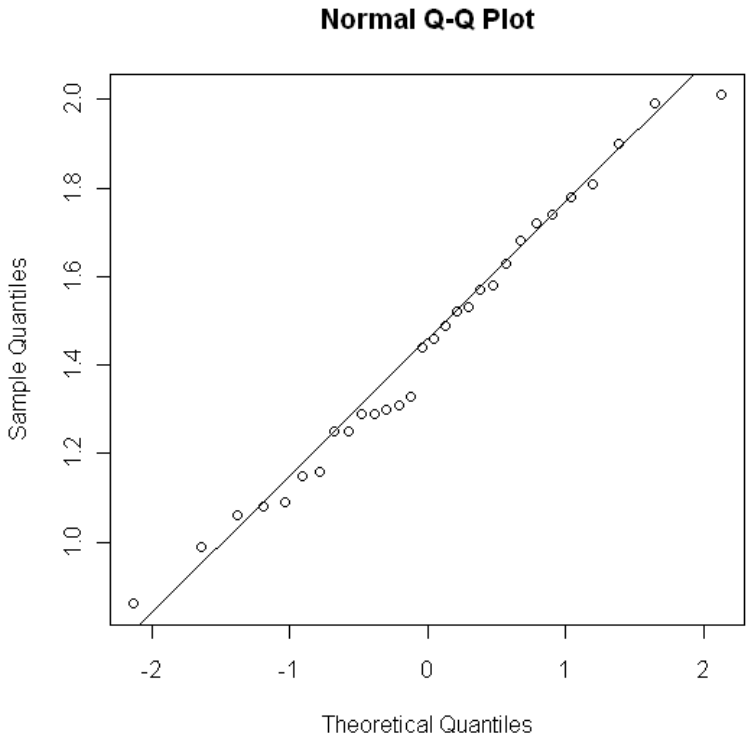

Figure 1: Sample Data Distribution

(D) Examine carefully Figure 2 below.

- What does the figure tell you about the spread of the two categories of data?

- Are there any outliers in the data?

- Would you expect to see a statistically significant difference between 1 & 2? Why or why not?

- What statistical test would you run to determine if your hunch is correct?

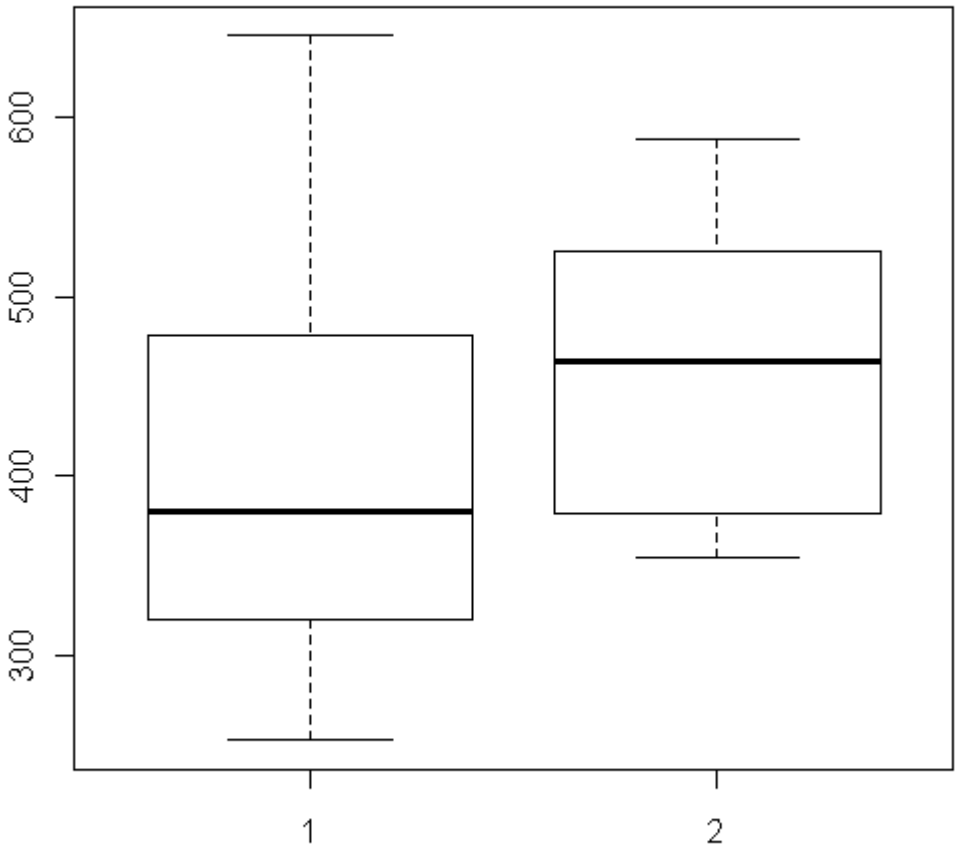

Figure 2: Boxplot

(E) Michael prepared the plot below (Figure 3) as a precursor to his data analysis of CO2 emissions of the 50 United States.

- Should he proceed and build a linear model of CO2 as described by the area of a state? Why or why not?

- If so, what would a linear model contribute to his understanding of the data analysis? If not, can you suggest another tact he might take (if there is none, say why not)?

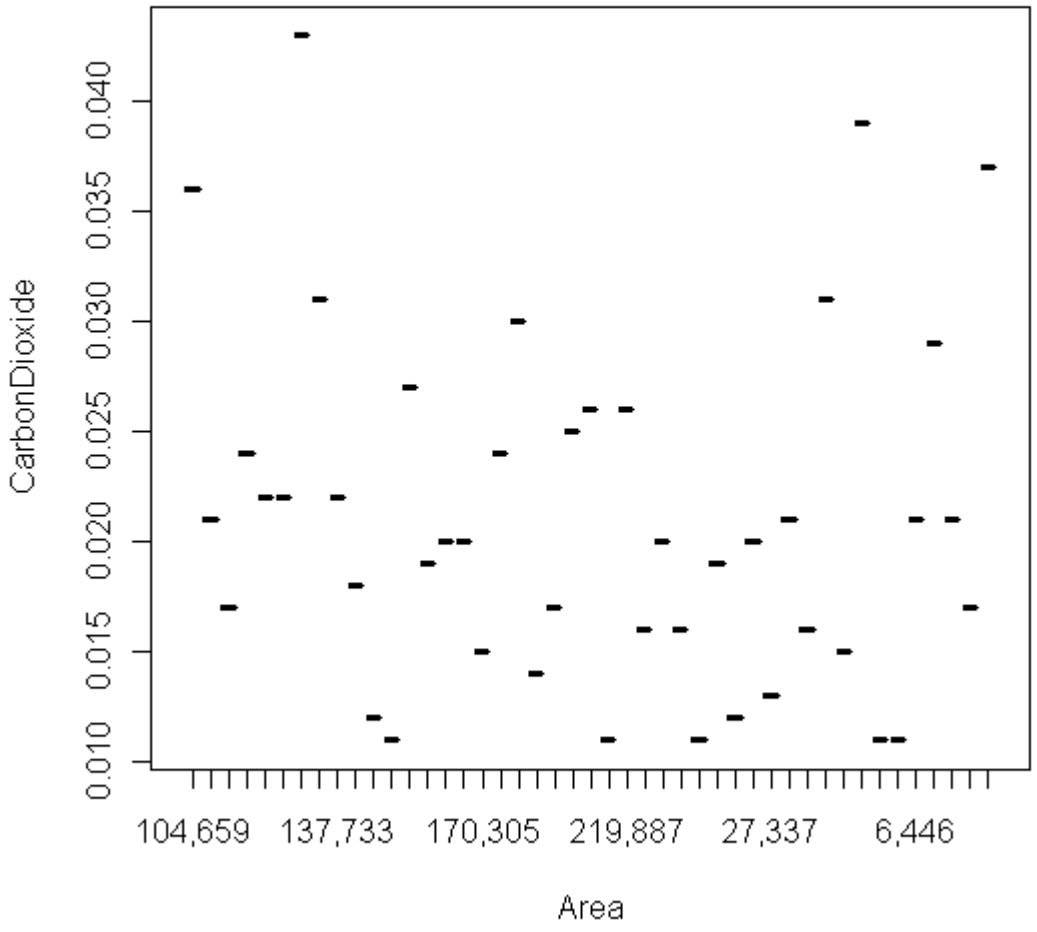

Figure 3: Carbon Dioxide and State Area

(3) (20%) This quarter you worked with one or two other students to analyze one or more datasets that had been prepared and published by scientists at the H. J. Andrews Experimental Forest. Each question is worth ~3 points, except (E), which is worth 5.

(A) Articulate as 5 major activities (in order) what you did to complete the analysis. What were these five activities? One of the steps should be "perform and interpret statistical analyses"; mention one statistical test you used.

(B) What was the most difficult aspect of the analysis for you?

(C) Hopefully you noticed and used the metadata that accompanies your dataset(s). If you could now advise the researcher(s) who published your dataset(s) on how they might improve the metadata so the data could be more easily re-used, what would you tell them?

(D) What did you learn from doing the data analysis?

(E) If you knew when you started the project what you know now, what would you have done differently?

(F) (5 pts) How might qualitative techniques be used to supplement the study that generated the data you analyzed? (Think creatively here!) Be specific to say what data might be collected, how it would be collected, and what you might learn from this supplementary study.