

Notes on Statistical Analysis for Student Projects

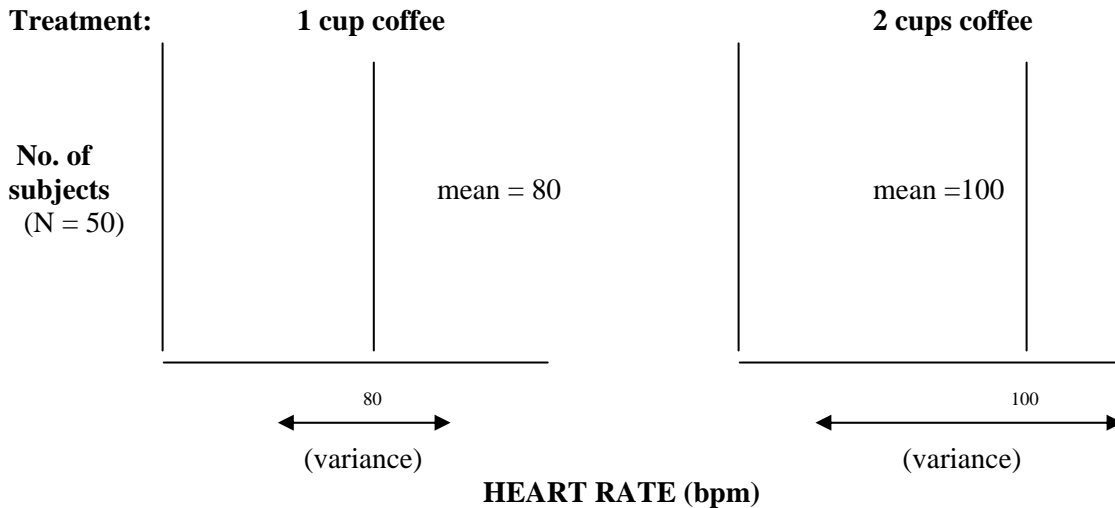
Instructions: Read and complete this exercise as a homework assignment. We will collect/discuss your answers during the next lab meeting. Please see *Pechenik or Knisely* for more information on using and reporting statistics.

Introduction

Your group research project will likely be an experiment designed to make a limited number of observations from which you would like to generalize to the whole population under study (your population of interest). Statistics are a tool used to make such inferences about a population based on a sample. A particular statistical analysis of your data (the dependent variable which you measure in your sample) is chosen before you begin to collect data, which you feel will provide the best estimate of population characteristics.

Researchers do not choose a statistical test to “wow” or prove anything (remember, you can not prove your hypothesis), but instead use statistics to define “different”. For example, you expect varying amounts of caffeine to have different effects on a subject’s heart rate. Inherent variability among test subjects will produce different measurements of heart rate, but the appropriate statistical test will determine if the results are “significantly different”. Statistical significance supports the hypothesis which states that there is a relationship between the independent and dependent variables tested in your experiment. Because your null hypothesis predicts that there will be no difference between treatment effects, statistical analysis showing no significant difference leads you to accept the null hypothesis. However, even if treatment effects are not statistically different, there may be important trends or other information which is valuable based on your results.

Using the information on mean and variance provided, draw the distributions around the mean for each of these figures (assume that the scale on the x-axis is the same).



Null hypothesis (H_0): There is no difference in heart rate as a result of 1 or 2 cups of coffee. (in other words, 1 cup = 2 cups; the trt will have no effect on heart rate).

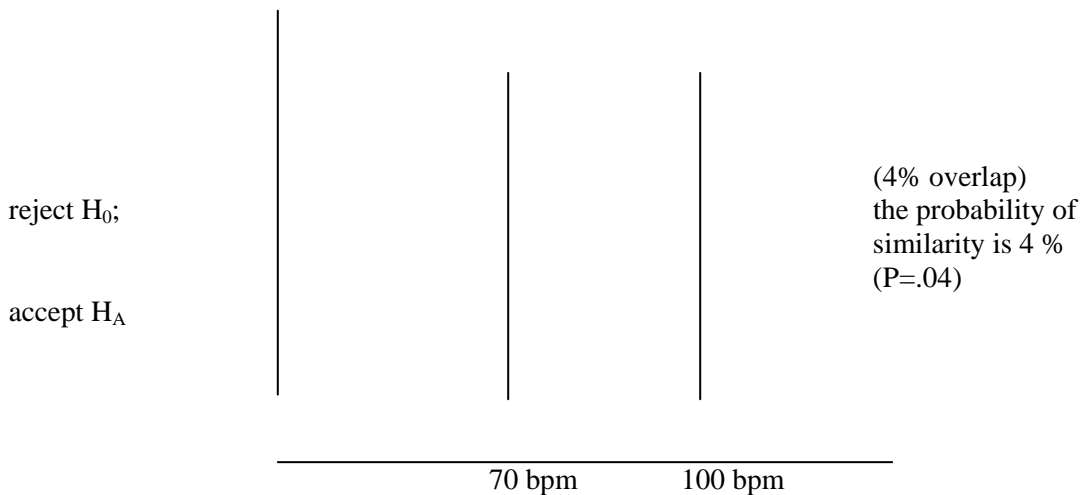
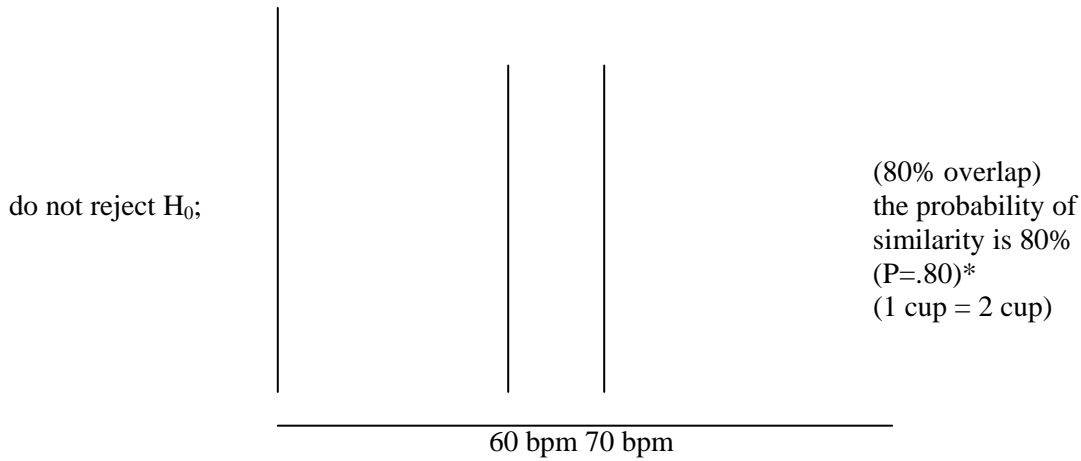
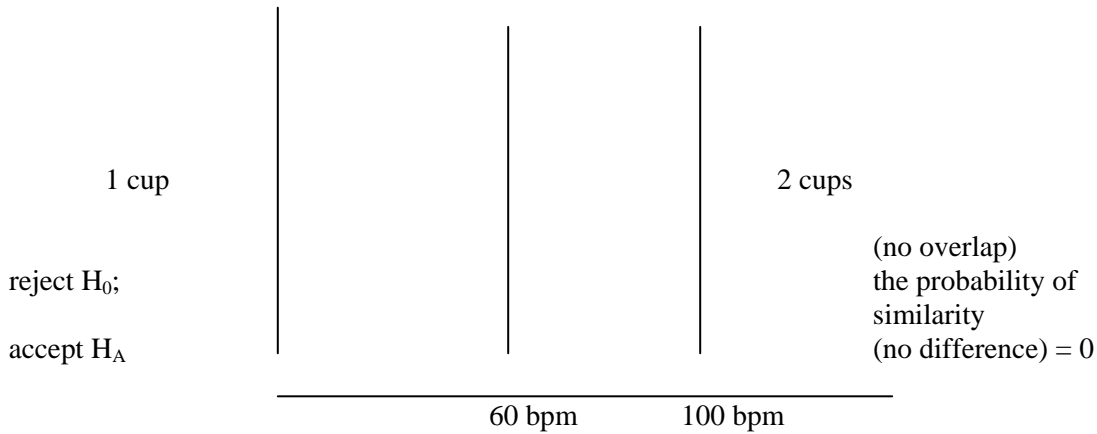
Research hypothesis (H_A): There is a difference in heart rate as a result of treatment differences.

Two potential questions may be answered by statistical analyses:

Is the difference in mean (average) heart rate between the two treatments significant?

Is the difference in variance (spread around the mean) significantly different?

Assume that the number of subjects in each treatment group is equal. Using the information provided, draw the distributions around the means for each of these figures.



* P = the probability that the treatments are the same

Choose a decision rule (alpha)

After you have decided which is the most appropriate statistical test for your experiment, but before you collect data, you must choose a decision rule, or how much overlap of the populations is acceptable and still decide to reject your null hypothesis. In other words, the alpha level is the

percent of instances you can accept that you are wrong when the alternative hypothesis is accepted. Generally, the alpha is set at 5% (.05), and therefore calculated values of P less than 0.05 allow rejection of the null hypothesis (a <5% possibility that treatment effects are the same). P values greater than or equal to 0.05% force you to accept the null hypothesis.

Population of Interest

Because it is impossible to test every individual on earth, research in human physiology relies on testing a sample of the whole population. Choices with respect to a sample depend on your hypothesis and experimental methods. You must choose an appropriate sample (who should be included/excluded), have a reasonably large sample to account for inherent variability, and insure that all treatment groups contain the same number of subjects. All of these decisions need to be made in order to be confident that the results based on a sample reflect the larger population.

t - test

A t-test to compare means is often chosen in experiments consisting of two levels of the independent variable. (for example, in which there is one control treatment and one experimental treatment, or one pre-treatment measurement and one post-treatment measurement). A P value will be calculated and compared to the alpha to make the decision to accept or reject H_0 . This is a less powerful statistical test than others, such as the ANOVA.

ANOVA (Analysis of variance – refer to the figures on the first page of this handout)

This tests attempts to answer the question: Does the inherent variability among test subjects account for the apparent difference between treatments, or are these differences due to the treatment itself? The ANOVA is used to compare pairs of treatments or treatment levels, and is a much stronger statistical test than the t-test. One big advantage is that several treatment levels (more than 2) can all be compared pair-wise. Like the t-test, a P value is the end-product of this test, and is compared to the alpha to make a decision.

Linear Regression

Linear regression analysis may help to establish a cause and effect relationship between the independent and dependent variables, and can predict the effects of treatment levels not tested, but within the same range. Draw the best fit line through the “data cloud” in the following figures to illustrate two different possible outcomes of linear regression analysis applied to the results of a coffee experiment.

Figure A:

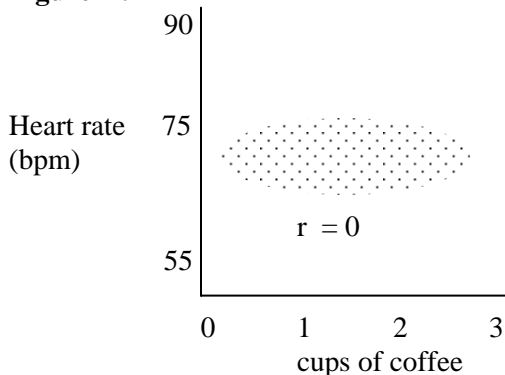
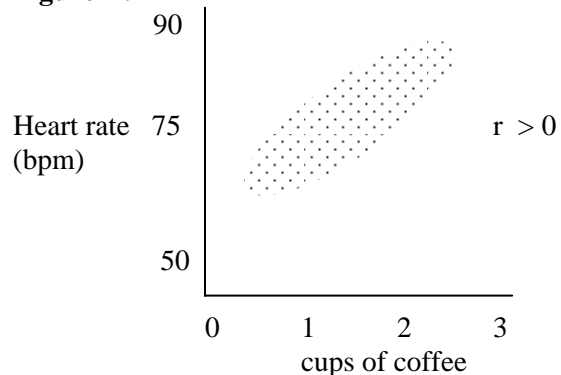


Figure B:



H_0 : There is no relationship between the number of cups of coffee and heart rate.

H_A : There is a direct linear relationship between the number of cups of coffee and heart rate.

The best fit line is drawn through the “data cloud”, and the slope of the line analyzed by linear regression. The slope (r) may be anywhere in the range between -1 and 1.

If the slope is 0 (as in A), with a probability of error greater than or equal to 0.05 (P), then accept H_0 (there is no difference between treatments).

If the slope is significantly different from 0 (as in B) ($P < 0.05$), then reject H_0 .

What would be the predicted effect of 1.5 cups of coffee on heart rate based on the line drawn in Figure B?