

## HW 8 FOR STAT601, Spring, 2011

1) For the data on heights of men and women verify (by hand) the values of the pooled two sample t-statistic and the unpooled two sample t-statistic, and verify the p-values (as accurately as possible) from the t-table.

2) We want to analyze the “Bluebonnets and Air Pollution” data that is given at:

<http://stat.tamu.edu/~sherman/classes/STAT601.spring10.html>

Two measures of health of bluebonnets are the height of the plant and the number of red petals. The main question is whether pollution causes some changes in the plants. A simple proxy for pollution is whether the plant is “near” or “far” from a highway. The SAS program: a) gives the data, b) “runs” PROC univariate, and c) two sample t-tests comparing the heights and red petal counts of the near and far plants.

a) Set up the hypotheses in symbols and in words.

b) Use the Univariate results to explore the distribution of the observations. (separately for near and far). Is the t-test (approximate normality of the observations) reasonable here?

c) State the null and alternative hypotheses to test for significant differences in heights between the near and far plants. Is the assumption of equal variances in the two groups reasonable? Test. Give the proper p-value and make your decision in terms of the null hypotheses, and in words that anyone would understand.

d) State the null and alternative hypotheses to test for significant differences in red petal counts between the near and far plants. Give the correct p-value and make your decision in terms of the null hypotheses, and in words that anyone would understand.

3) We want to analyze the “Comparing the Costs of Two Software Packages” data. Read the data description and answer the following questions.

a) Perform the first analysis using the differences between costs. State the null and alternatives, compute the p-value, and give your decision. The PROC means gives you the mean (xbar) and the standard deviation of the data. Use these to form the “t-statistic” and get an approximate p-value from the t-table.

b) Perform the second analysis that compares the two packages using the two sample t-test. State the null and alternatives, state the p-value, and give your decision.

c) Explain which of a) and b) is correct and why. Which of the analyses was more able to detect difference between the packages?

4) Verify that in the Salk Polio Vaccine trial the children who did not get any treatment at all (nonconsenters) had a significantly SMALLER chance of getting polio than the children who received the placebo. [This dramatically points out the need to “keep all things equal” when comparing two treatments].

5) A new regimen argues that it is a partial cure for breast cancer. The researchers believe that their regimen, if strictly adhered to, will reduce the risk of breast cancer in women aged 30-50 from 6.8% (the current rate) to 4.5% (a 34% reduction). This latter figure was arrived at from a small (nonrandomized) pilot study of 1000 women who tried the regimen. Approximately how many women do they need to recruit in a clinical trial to have a 80% chance of rejecting the null

hypothesis of no benefit when testing at  $\alpha = .05$ , assuming the pilot values are correct.