Homework 2 Due: See Canvas

1. In a study to test the efficacy of a diet in reducing obesity, 20 obese participants (initial BMI ≥ 30) were enrolled. One participant dropped out after few weeks. The participants' BMI values were observed before the program started and after following the diet plan for one month. The data in *bmi.txt* gives the change in BMI of these 19 participants. Suppose μ denote the average change in BMI due to the diet. The null hypothesis is: $H_0 : \mu = 0$.

   (a) Compute the t-ratio for testing the null hypothesis.

   (b) Suppose you'd like to test the null hypothesis against the *one-sided* alternative $H_A : \mu < 0$. Now only large negative values of the t-ratio are considered "extreme". Compute the one-sided p-value and provide a conclusion in the context of the problem.
   Note: To get the one-sided p-value you use: In **SAS** you use `proc ttest side=L`
   In **R** you have to pass the argument `alternative = "less"` to `t.test`.
   In **JMP**, this one-sided p-value is reported as `Prob < t`.

   (c) Now test the null hypothesis against the *two-sided* alternative: $H_A : \mu \neq 0$. Find the p-value and provide a conclusion in the context of the problem.

   (d) Between the one-sided and the two-sided alternatives which one do you think is more objective and why?

2. This problem is based on a study of genetic contributions to Downs syndrome. The data consists of weights, in grams, of two groups of mice: non-transgenic (`NONTRANS`) and transgenic (`TRANS`) mice with a extra 285E6 DNA fragment. Assignment of mice to these groups is not random; it depends on whether the gene injection process "works" or not. Summaries of the two data sets are provided in Figure 1.

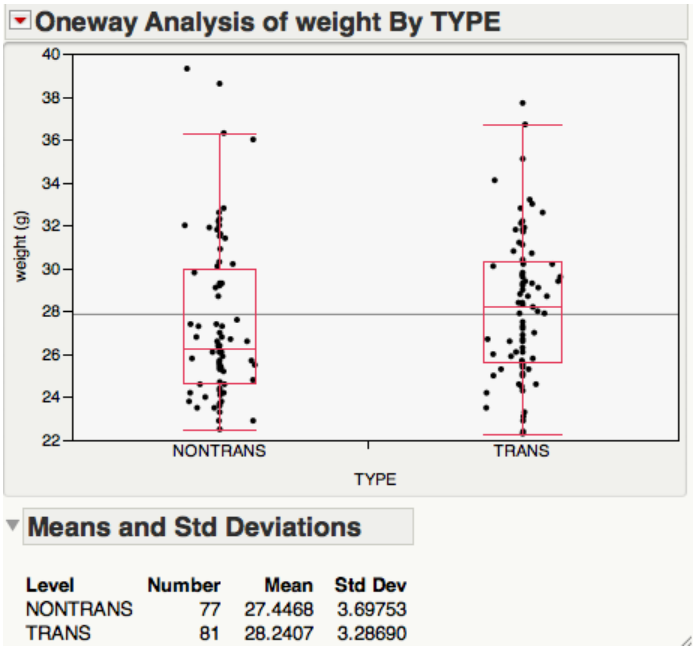

Figure 1: Summaries for the MICE data set. Points were jittered to avoid overplotting.

   (a) Calculate the pooled estimate of the standard deviation for the two independent samples.

   (b) Calculate the standard error of the difference between the two averages.

   (c) Conduct a pooled t-test to test whether the means are the same. Report here the test statistic, degrees of freedom, p-value, and a short interpretation. Show all your work.

(d) Is it appropriate to make a causal conclusion (e.g. presence of the 285E6 fragment increases the weight of mice) from a study like this? Explain (briefly) why or why not.

3. Many studies of blood pressure have tried to identify environmental and life style characteristics that influence blood pressure. Some studies are randomized experiments; many are not. One study looked at men in a particular city. The study compared the blood pressure of men who consistently ate oatmeal 3 times a week to that of those who had never eaten oatmeal. The data are summarized in the table below.

Group	n	average	s.d.
Oatmeal haters	49	155	14
Oatmeal eaters	16	135	12

(a) The variability among observations is larger in which group of men? Briefly explain how you determined your answer.

(b) Compute the pooled estimate of the standard deviation.

(c) Compute a 99% confidence interval for the difference in mean blood pressure between the groups. Show all your work. Interpret this interval within the context of the study.

(d) Results from this study are reported as "Eating oatmeal reduces your blood pressure". Is this claim appropriate? Why or why not?

4. **Chapter 2: 18 (3rd edition of the book), not in 2nd edition** (Galapagos Finches.) One of the classic data sets in evolutionary ecology is the long term data collected by Peter and Rosemary Grant on finches in the Galapagos. The data in `ex0218.txt` describe the size of the beak of finches on Daphne Major. If you have the 3rd edition, this study is case study 2.1. It is not in the 2nd edition, but you don't need any more than this description to do the problem. The data set describes the beak size of all 751 finches that were measured in 1976 and all 89 finches that were measured in 1978. They measured all these species of finch on the island each year. 1977 was a severe drought year; the only food source was a large, tough seed that was ignored under normal conditions. However, to survive, a bird must have eaten those large, tough seeds.

(a) Draw side-by-side box plots of beak size for the 1976 and 1978 birds.

(b) Use a two-sample (pooled) t-test to test the hypothesis of no change in beak size. Contrary to the book's question, please report a two-sided p-value, and interpret the results within the context of the problem. Show any work and make sure you specify the distribution and degrees of freedom used.

(c) Estimate the increase in mean beak size between 1976 and 1978. Also report the standard error of this estimate and a 95% confidence interval, calculated using standard methods discussed in class.

(d) Typical life span for these finches is 5-20 years. Does this raise any concerns about the assumption of independence between the two groups?

5. **Chapter 3: 27 (3rd)/28 (2nd)** (Pollen Removal) Ignore all the book's questions. Answer these new questions. The data are provided in `bee.txt`. The first column is the proportion of pollen (ignore for this problem), the second column is the duration of the visit, the third column indicates worker (W) or queen (Q). Consider the data on duration of visit (2nd column). You want to find out whether workers spend more time in flowers than do queens. It is unclear whether the mean or the median is more appropriate to describe typical visit length.

(a) Report the means and the standard deviations of the groups.

(b) Use a pooled t-test to test whether the mean visit lengths are equal. Report the t-ratio, df, p-value, and write a conclusion within the context of the data.

(c) Compute a 90% confidence interval for the difference in mean visit lengths, and give an interpretation within the context of these data.

(d) Run diagnostic tests – provide the side-by-side boxplots and the QQ-plots. Comment on your findings.