Homework #2 Solutions

***Note: You get 4 points just for turning in the assignment!***

## Question 1 (16 points total)

The world’s smallest mammal is the bumblebee bat, also known as the Kitti’s hog nosed bat. Such bats are roughly the size of a large bumblebee! Listed below are weights (in grams) from a sample of these bats. Test the claim that these bats come from the same population having a mean weight equal to 1.8g. *Beware: This data is not the same as in the lecture slides!*

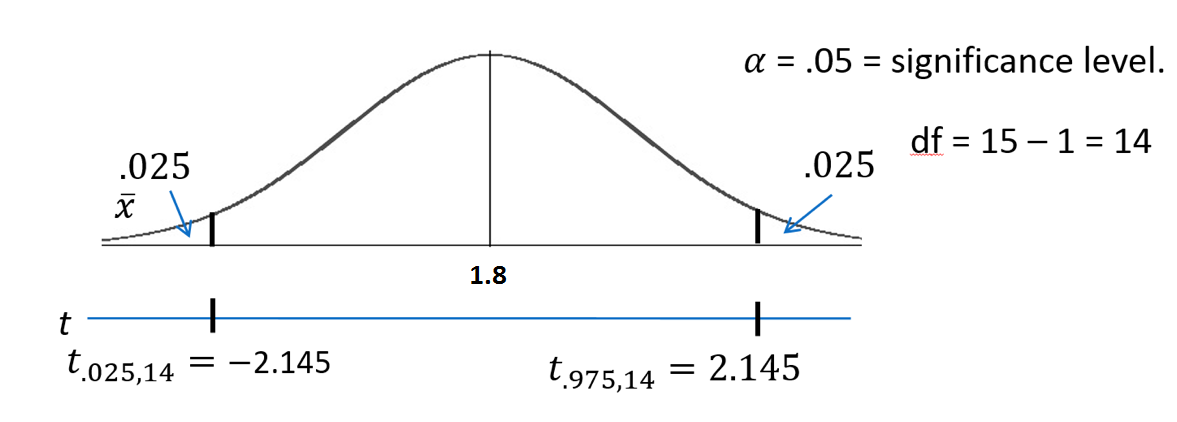
Sample: 1.7, 1.6, 1.5, 2.0, 2.3, 1.6, 1.6, 1.8, 1.5, 1.7, 1.2, 1.4, 1.6, 1.6, 1.6

### Part A (15 points total)

Perform a complete analysis using SAS. Use the six step hypothesis test with a conclusion that includes a statistical conclusion, a confidence interval, and a scope of inference (as best as can be done with the information above…there are many correct answers with the vagueness of the description of the sampling mechanism).

**Step 1 - Hypotheses (2 points):**

**Step 2 - Identification of Critical Value (1 point for drawing, 1 point for value):**



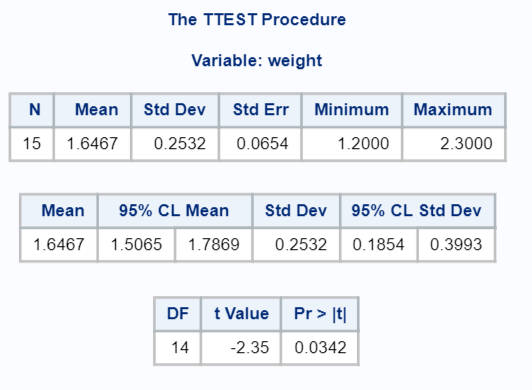
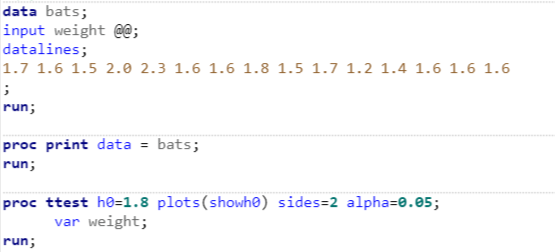
**Step 3 - Value of Test Statistic (2 points):**

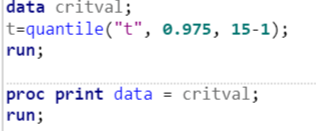
**Step 4 - Give p-value (2 points):**

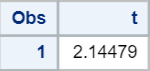
**Step 5 - Decision (2 points): Reject**  **(p=0.0342 < 0.05)**

**Step 6 - Conclusion (3 points for the statistical conclusion, 1 point for the confidence interval, 1 point for discussing the scope): On the basis of this test, there is enough evidence to reject the claim that the mean weight of bumblebee bats is equal to 1.8g ( from a two-sided t-test). A 95% confidence interval is [1.5065, 1.7869] grams, with a point estimate of 1.6467 g. The problem was ambiguous on the randomness of the sample; thus, we will assume that it was not a random sample, which makes inference to all bats strictly speculative.**

**SAS:**





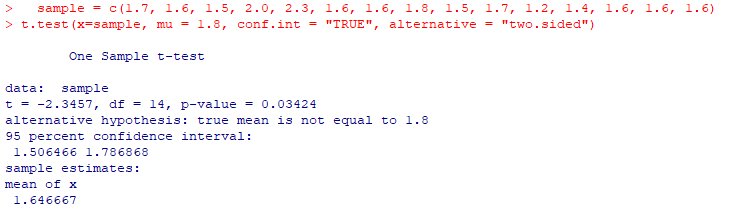


### Part B (1 point)

Inspect and run this R Code and compare the results (t statistic, p-value, and confidence interval) to those you found in SAS. To run the code simply copy and paste the below code into R.

sample = c(1.7, 1.6, 1.5, 2.0, 2.3, 1.6, 1.6, 1.8, 1.5, 1.7, 1.2, 1.4, 1.6, 1.6, 1.6)  
t.test(x=sample, mu = 1.8, conf.int = "TRUE", alternative = "two.sided")

**You should observe exactly the same results.**



## Question 2 (40 points total)

In the United States, it is illegal to discriminate against people based on various attributes. One example is age. An active lawsuit filed August 30, 2011, in the Los Angeles District Office is a case against the American Samoa Government for systematic age discrimination by preferentially firing older workers. Though the data and details are currently sealed, let’s suppose that a random sample of the ages of fired and not fired people in the American Samoa Government are listed below:

**Fired**  
> 34, 37, 37, 38, 41, 42, 43, 44, 44, 45, 45, 45, 46, 48, 49, 53, 53, 54, 54, 55, 56  
**Not Fired**  
> 27, 33, 36, 37, 38, 38, 39, 42, 42, 43, 43, 44, 44, 44, 45, 45, 45, 45, 46, 46, 47, 47, 48, 48, 49, 49, 51, 51, 52, 54

### Part A (3 points for running the test, 1 point for the p-value, and 1 point for the correct conclusion)

Perform a permutation test to test the claim that there is age discrimination. Provide the and , the p-value, and full statistical conclusion including the scope. Note: this was an example in Live Session 1. You may start from scratch or use the sample code and PowerPoint from Live Session 1. You can earn full credit for a two-sided test, but note that a one-sided test more closely answers the question of interest.

One sided:

Two sided:

fired <- c(34, 37, 37, 38, 41, 42, 43, 44, 44, 45, 45, 45, 46, 48, 49, 53, 53, 54, 54, 55, 56)  
not.fired <- c(27, 33, 36, 37, 38, 38, 39, 42, 42, 43, 43, 44, 44, 44, 45, 45, 45, 45, 46, 46, 47, 47, 48, 48, 49, 49, 51, 51, 52, 54)  
label1 <- rep('fired', 21)  
label2 <- rep('not.fired', 30)  
label.all <- as.factor(c(label1, label2))  
samoa <- data.frame(status=label.all, age=c(fired, not.fired))  
  
t.test(age ~ status, data=samoa, var.equal=TRUE)

Two-Sample t Test

Data

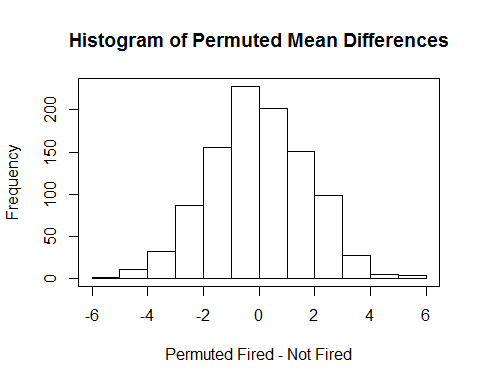
##   
## Two Sample t-test  
##   
## data: age by status  
## t = 1.0991, df = 49, p-value = 0.2771  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1.678993 5.526612  
## sample estimates:  
## mean in group fired mean in group not.fired   
## 45.85714 43.93333

number\_of\_permutations <- 1000  
xbarholder <- numeric(0)  
counter <- 0  
observed\_diff <- mean(subset(samoa, status == "fired")$age)-mean(subset(samoa, status == "not.fired")$age)  
  
set.seed(123)  
for(i in 1:number\_of\_permutations)  
{  
scramble <- sample(samoa$age, 51)  
fired.new <- scramble[1:21]  
not.fired.new <- scramble[22:51]  
diff <- mean(fired.new)-mean(not.fired.new)  
xbarholder[i] <- diff  
## if doing a two-sided t-test, the following line should be (abs(diff) > abs(observed\_diff))

Permutations Test

if(diff > observed\_diff)  
counter <- counter + 1  
}  
hist(xbarholder, xlab='Permuted Fired - Not Fired', main='Histogram of Permuted Mean Differences')  
box()

Note Comment Well



pvalue <- counter / number\_of\_permutations  
pvalue

## [1] 0.133

For a two-sided test, the p-value = 0.269.

**There is not sufficient evidence to suggest that the mean score of those who were fired is greater than (or different from) the mean age of those who were not fired (** for a one-sided and two-sided test, respectively**). Since this was a random sample of employees in Samoa, we can generalize the inference to all employed people in Samoa. Since we failed to reject the null hypothesis, we do not need to discuss whether causal conclusions can be drawn.**

*Note: Remember, your p-value will not be exactly the same. However, it should be very close (likely within 0.05).*

SAS:



### Part B (15 points total, scored exactly as Problem 1, part A)

Now run a two sample t-test appropriate for this scientific problem (use SAS) (Note: we may not have talked much about a two-sided versus a one-sided test. If you would like to read the discussion on p.44 (Statistical Sleuth), you can run a one-sided test if it seems appropriate. Otherwise, just run a two-sided test as in class. There are also examples in the Statistics Bridge Course). Be sure to include all six steps, a statistical conclusion, and scope of inference.

*Note: the solutions will look slightly different if you do a one-sided test. Either option can still receive full credit, though.*

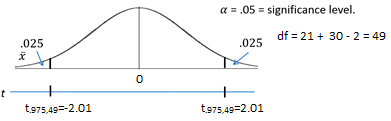
*Note: If your data is sorted in the opposite order, you will get all the same numbers but with opposite signs (for example, -1.10 instead of 1.10). This is totally fine and your answers should match up, including p-values. If your signs are opposite and your p-value doesn’t match up, it means you are reporting the wrong shaded area. In this case, if you subtract your p-value from 1 then it should match with the p-value given in the solution.*

**Step 1 - Hypotheses (2 points):**

**Or**

or

**Step 2 - Identification of Critical Value (1 point for drawing, 1 point for value) using alpha = 0.05 and 30+21-2=49 degrees of freedom: (2-sided) or (1-sided)**



Pooled d.f.

**Step 3 - Value of Test Statistic (2 points):**

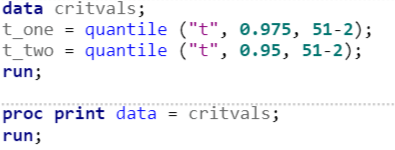
**Step 4 - Give p-value (2 points): (2-sided) or (1-sided)**

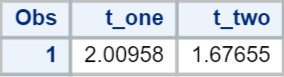
**Step 5 - Decision (2 points): Fail to Reject**

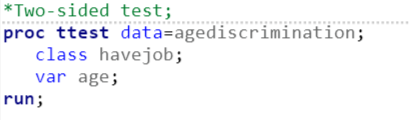
**Step 6 - Conclusion (3 points for the statistical conclusion, 1 point for the confidence interval, 1 point for discussing the scope): On the basis of this test, there is not enough evidence to suggest that the mean ages of the fired and not fired groups are different. In other words, there is not enough evidence to suggest that there is discrimination based on age ( from a two-sample t-test, from a one-sample t-test). A 95% confidence interval for this difference is years. Since the subjects in this sample were randomly sampled, inference can be generalized to the population of all employees in the American Samoa Government.**

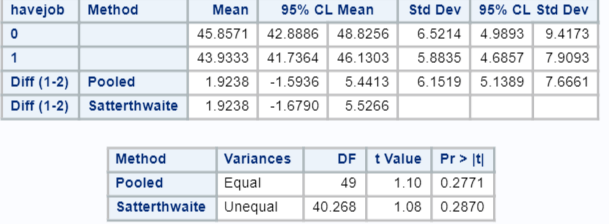
*Note: If you did a 1-sided test, the 95% CI is . A more appropriate 90% two-sided confidence interval for the difference in means is .*

**SAS:**



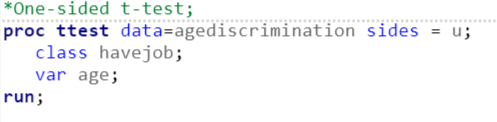
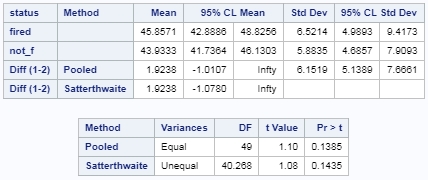




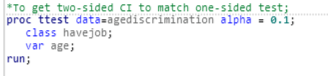


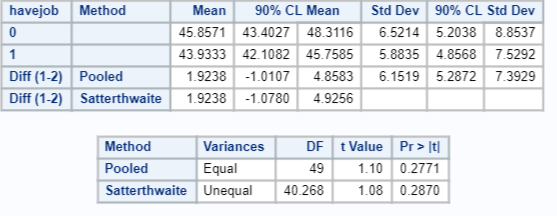
Part E: *sp*

Part D Confidence Interval

Part C: Compare to permutations test





### Part C (5 points total)

Compare this p-value to the randomized p-value found in question 2a.

**P-values will vary from test to test, but the p-values should be close because the distribution of sample means from the permutation test is approximately normal.**

### Part D (5 points total)

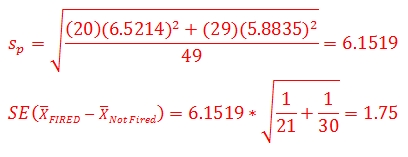
The jury wants to see a range of plausible values for the difference in means between the fired and not fired groups. Provide them with a confidence interval for the difference of means and an interpretation.

**A 95% confidence interval for this difference is years (or the 1-sided interval reported above). With 95% confidence, we expect the true mean difference in ages to fall in this interval. In particular, approximately 95% of all confidence intervals created this way should include the true difference in means. Because 0 is included in the interval and is thus a probable value, we don’t have enough evidence to conclude a difference exists.**

### Part E (5 points total)

Given the sample standard deviations from SAS, calculate by hand:

i. (2.5 points) Pooled Standard Deviation   
ii. (2.5 points) Standard Error of



### Part F (5 points total)

Inspect and run this R Code and compare the results (t statistic, p-value and confidence interval) to those you found in SAS. To run the code simply copy and paste the below code into R.

Fired = c(34, 37, 37, 38, 41, 42, 43, 44, 44, 45, 45, 45, 46, 48, 49, 53, 53, 54, 54, 55, 56)  
Not\_fired = c(27, 33, 36, 37, 38, 38, 39, 42, 42, 43, 43, 44, 44, 44, 45, 45, 45, 45, 46, 46, 47, 47, 48, 48, 49, 49, 51, 51, 52, 54)  
t.test(x = Fired, y = Not\_fired, conf.int = .95, var.equal = TRUE, alternative = "two.sided")

**You should observe exactly the same results. No real discussion is needed aside from making this observation!**

## Question 3 (25 points total)

In the last homework, it was mentioned that a Business Stats professor here at SMU polled his class and asked them how much money (cash) they had in their pocket at that very moment. The idea was that we wanted to see if there was evidence that those in charge of the vending machines should include the expensive bill/coin acceptor or if it should just have the credit card reader. However, another professor from Seattle University was asked to poll her class with the same question. Below are the results of the polls.

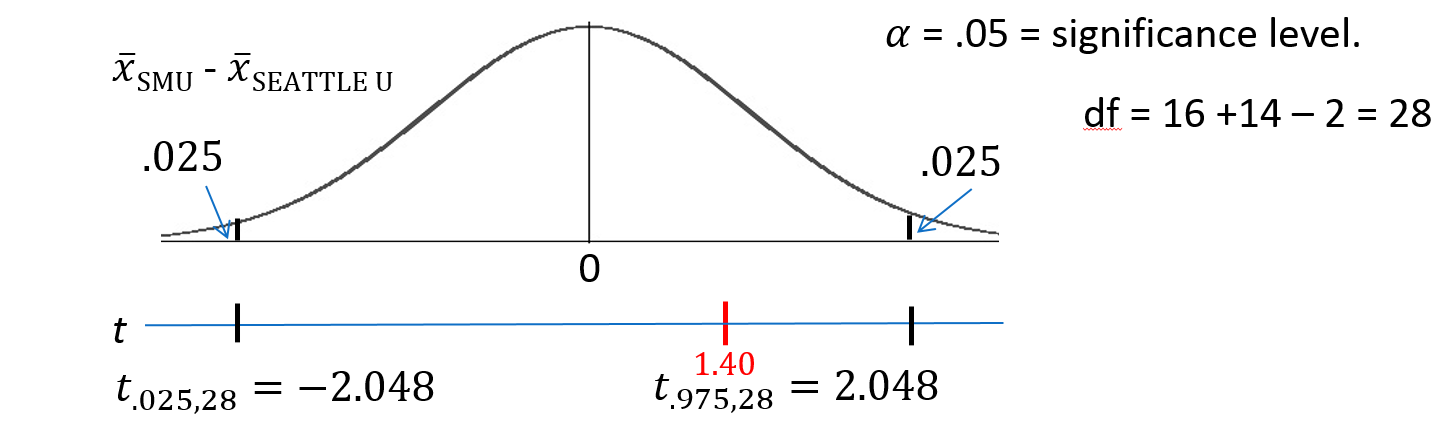
**SMU**  
> 34, 1200, 23, 50, 60, 50, 0, 0, 30, 89, 0, 300, 400, 20, 10, 0  
**Seattle U**  
> 20, 10, 5, 0, 30, 50, 0, 100, 110, 0, 40, 10, 3, 0

### Part A (15 points, score like the other hypothesis tests)

Run a two-sample t-test to test if the mean amount of pocket cash from students at SMU is different than that of students from Seattle University. Write up a complete analysis: all 6 steps, including a statistical conclusion and scope of inference (like the one from the PowerPoint). This should include identifying the and as well as the p-value. Also, include the appropriate confidence interval. FUTURE DATA SCIENTIST’S CHOICE!: YOU MAY USE SAS OR R TO DO THIS PROBLEM!

**Step 1 - Hypotheses (2 points):**

**Step 2 - Identification of Critical Value (1 point for drawing, 1 point for value):**



**Step 3 - Value of Test Statistic (2 points):**

**Step 4 - Give p-value (2 points):**

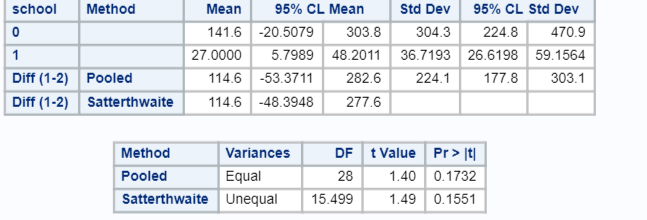
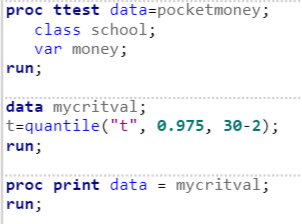
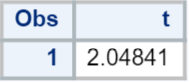
**Step 5 - Decision (2 points): Fail to Reject**

**Step 6 - Conclusion (3 points for the statistical conclusion, 1 point for the confidence interval, 1 point for discussing the scope): Based on this test, there is not enough evidence to suggest that the mean amount of pocket cash of the SMU students is different than that of the students from Seattle U ( from a two-sided t-test). A 95% confidence interval for this difference is [-$53, $282]. Since the subjects in this sample were not randomly sampled, the results only generalize to the subjects in the study (no need to discuss causal conclusions for a non-significant result).**

SMU = c(34, 1200, 23, 50, 60, 50, 0, 0, 30, 89, 0, 300, 400, 20, 10, 0)  
Seattle = c(20, 10, 5, 0, 30, 50, 0, 100, 110, 0, 40, 10, 3, 0)  
school1 <- rep('SMU', 16)  
school2 <- rep('Seattle', 14)  
school <- as.factor(c(school1, school2))  
all.money <- data.frame(name=school, money=c(SMU, Seattle))  
  
t.test(money ~ name, data=all.money, var.equal=T, alternative='two.sided')

##   
## Two Sample t-test  
##   
## data: money by name  
## t = -1.3976, df = 28, p-value = 0.1732  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -282.62112 53.37112  
## sample estimates:  
## mean in group Seattle mean in group SMU   
## 27.000 141.625

SAS:

Part B: Compare to permutations test

### Part B (10 points)

Compare the p-value from this test with the one you found from the permutation test from last week. Provide a short 2 to 3 sentence discussion on your thoughts as to why they are the same or different.

**The p-value from the two-sample t-test was 0.1732, while the p-value from the permutation test last week was somewhere near 0.15 (answers will vary due to randomness). Recall from last week that the distribution of the permuted differences of sample means was very non-normal (if you need to refresh your memory, refer to the histogram in the Unit 1 solutions). This shows that the normal distribution (or one that is approximated by a t distribution) is not always a good approximation of the distribution of a statistic. Although the normal distribution is obviously a poor fit for this distribution, it just so happens that it does a good job of approximating the tails of the bimodal permuted distribution. The main point here is that the outlier in this data had a drastic effect on the shape of the distribution of sample means, thus disqualifying the normal distribution from being a good approximation of the distribution to approximate p-values. The p-values for the t-test and permutation test are close here, but this cannot be depended on to be consistent.**

*Note: It should be noted that the p-values are close but pay careful attention to the discussion above. The critical piece to mention is that the outlier has a drastic effect on the outcomes.* ***If you do not discuss the outlier in any capacity, 2 points are taken off.***

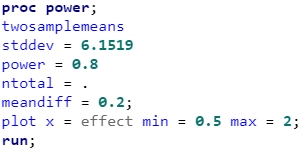
Think about the CLT. Does this makes sense?

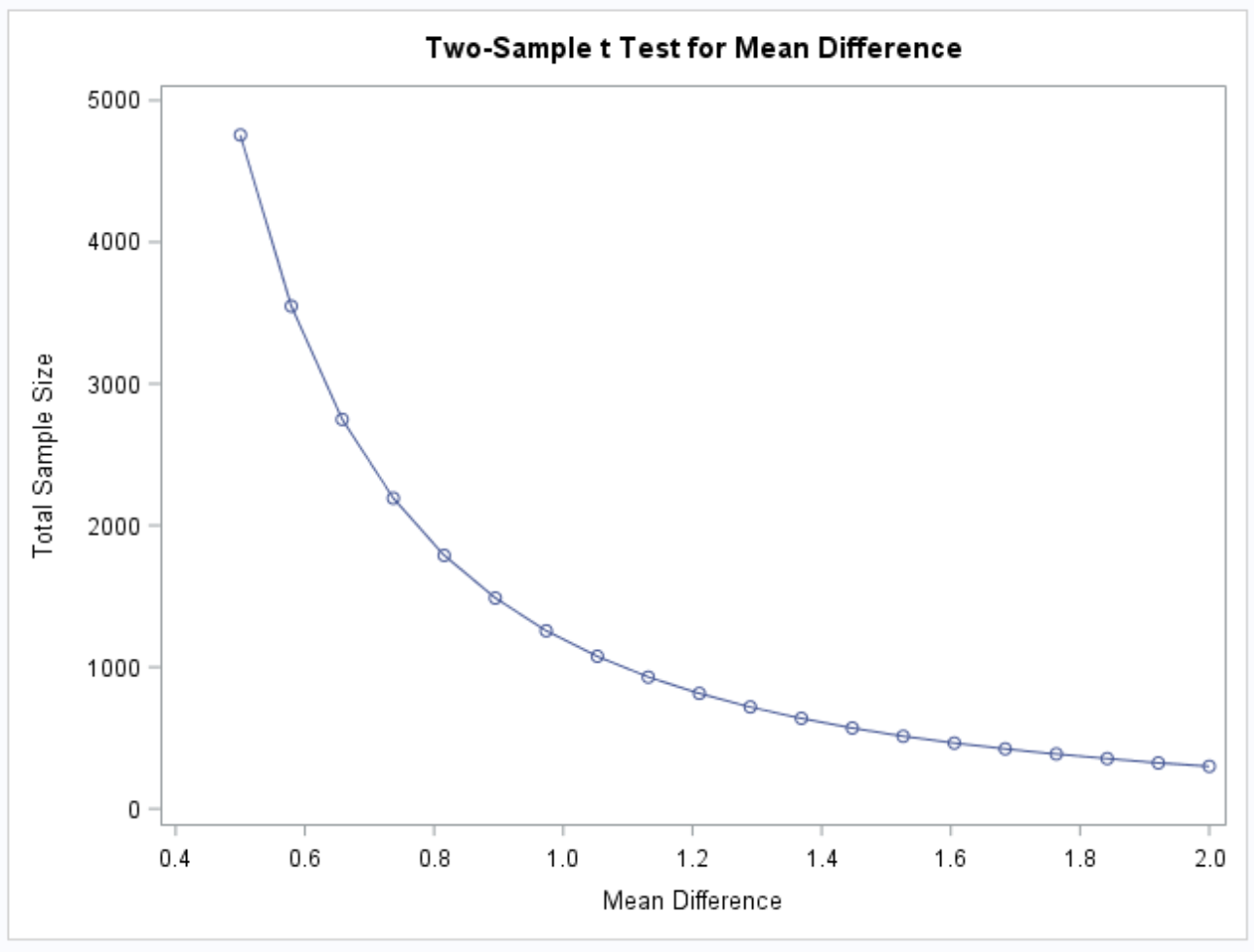
I never learned this. Remember, these are not large samples.

## Question 4 (15 points total)

### Part A (5 points)

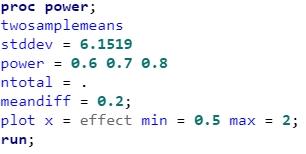
Consider the Samoan discrimination problem data (see data from question 2).Calculate the estimate of the pooled standard deviation from the Samoan discrimination problem (see data from question 2). Use this estimate to build a power curve. Use a meandiff = 0.2. Assume we would like to be able to detect effect sizes between 0.5 and 2 and we would like to calculate the sample size required to have a test that has a power of .8. Simply cut and paste your power curve and SAS code. HINT: USE THE CODE FROM DR. McGEE’s lecture. Instead of using groupstddevs, use stddev since we are using the pooled estimate.

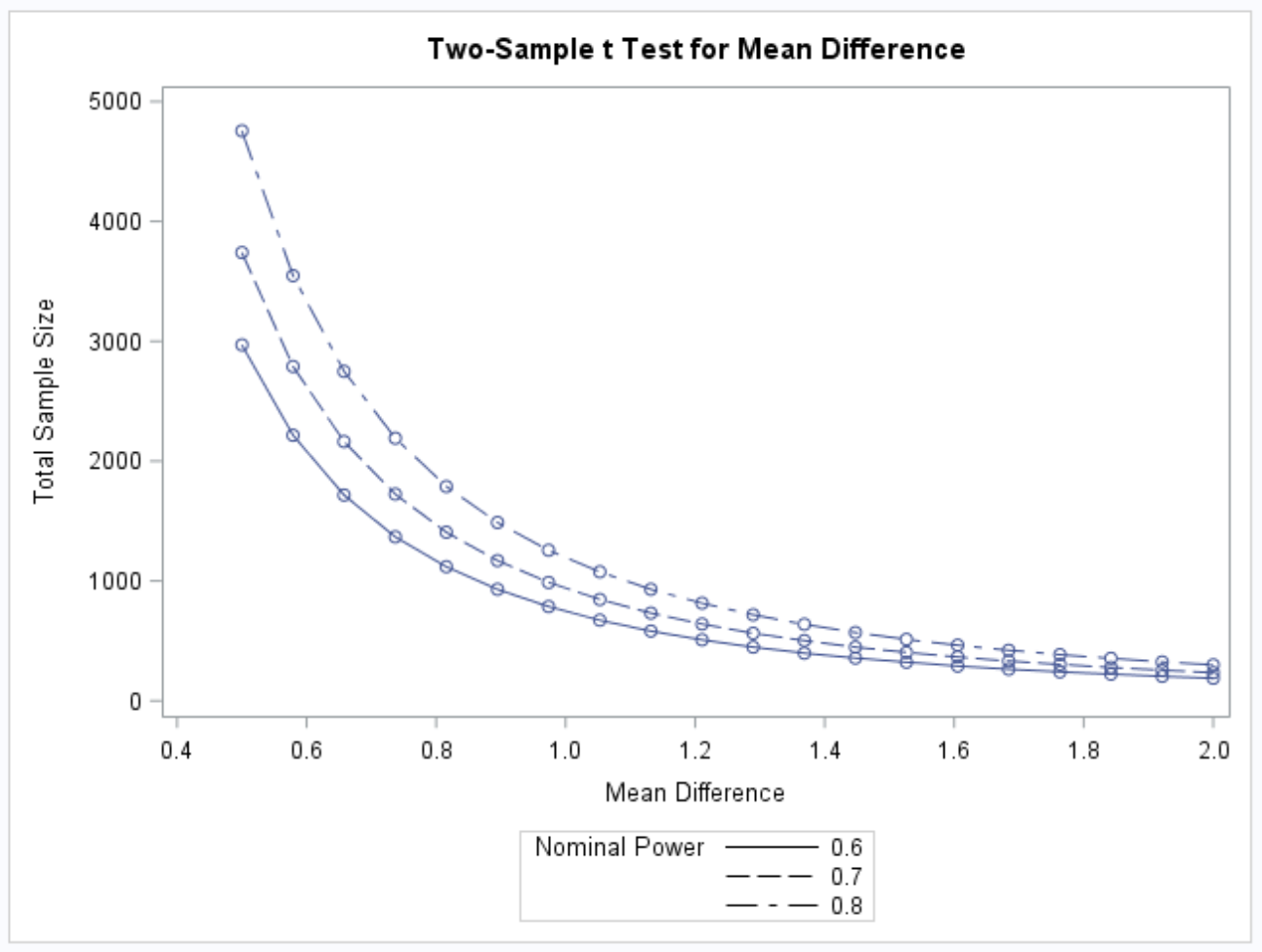
**SAS Code**  


**SAS Output**  


### Part B (5 points)

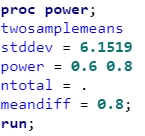
Now let’s say that we decided that we may be able to live with slightly less power if it means savings in sample size. Provide the same plot as above but this time calculate curves of sample size (yaxis) vs effect size (.5 to 2) (x axis) for power = 0.8, 0.7 and 0.6. There should be three plots on your final plot. Simply cut and paste your power curve and SAS code. HINT: USE THE CODE FROM DR. McGEE’s lecture… instead of groupstddevs, use stddev since we are using the pooled estimate. The effect size here refers to a difference in means, though there are many effect size metrics, such a Cohen’s D.

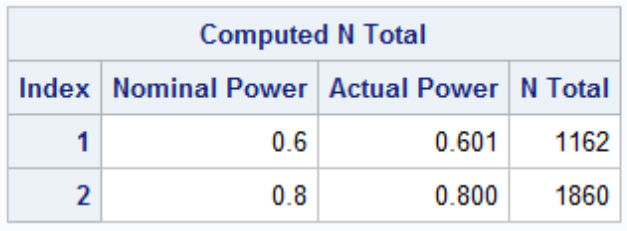
**SAS Code**  


**SAS Output**  


### Part C (5 points)

Using your last plot, estimate the savings in sample size from a test aimed at detecting an effect size of 0.8 (meandiff) with a power of 80% versus a power of 60%.

**SAS Code**  


**SAS Output**  


**Savings: 1860 - 1162 = 698**

*Note: If you only reported the output and not the savings, you lose one point.*