Stat 5000Lab #3  
Fall 2024 Due Tue Sep 17th Name: Sam Olson

**Assignment**

1. Conduct the t-test for the SMS speed example in SAS and complete the following exercises:
2. Using the formula from the notes, calculate by hand a 95% confidence interval for the difference in the two treatment means. Use .

Formula:

$$\text{95% Confidence Interval} \space = (\bar{Y\_1} - \bar{Y\_2}) \pm t\_{28,0.975} S\_p \sqrt{\frac{1}{n\_1} + \frac{1}{n\_2}}$$

Where:

Calculation

# Load Data   
library(readr)  
smsspeed\_1 <- read\_csv("C:/Users/samue/OneDrive/Desktop/Iowa\_State\_PS/STAT 5000/Labs/Lab 3/smsspeed-1.csv")

## Rows: 30 Columns: 4  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): AgeGroup  
## dbl (3): Age, Own Phone, Control  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

data1 <- smsspeed\_1 %>%   
 filter(smsspeed\_1$AgeGroup == "Over30")  
  
data2 <- smsspeed\_1 %>%   
 filter(smsspeed\_1$AgeGroup == "Teens")  
  
sampleMean1 <- mean(data1$`Own Phone`)  
sampleMean2 <- mean(data2$`Own Phone`)  
difference <- sampleMean1 - sampleMean2   
difference

## [1] 44.012

#Step 2: Finding standard deviation   
s1 <- sd(data1$`Own Phone`)  
s2 <- sd(data2$`Own Phone`)  
  
#Step 3: Finding sample size   
n1 <- length(data1$`Own Phone`)  
n2 <- length(data2$`Own Phone`)  
  
numerator <- (n1-1)\*(s1^2) + (n2-1)\*(s2^2)  
denom <- n1 + n2 - 2  
pooled <- sqrt( numerator / denom )   
sqrtFactor <- sqrt(1/n1 + 1/n2)  
  
tStatDf <- 2.0484  
  
rightSide <- tStatDf\*pooled\*sqrtFactor   
  
pooled

## [1] 18.51069

rightSide

## [1] 13.84544

lb <- difference - rightSide  
ub <- difference + rightSide  
  
lb

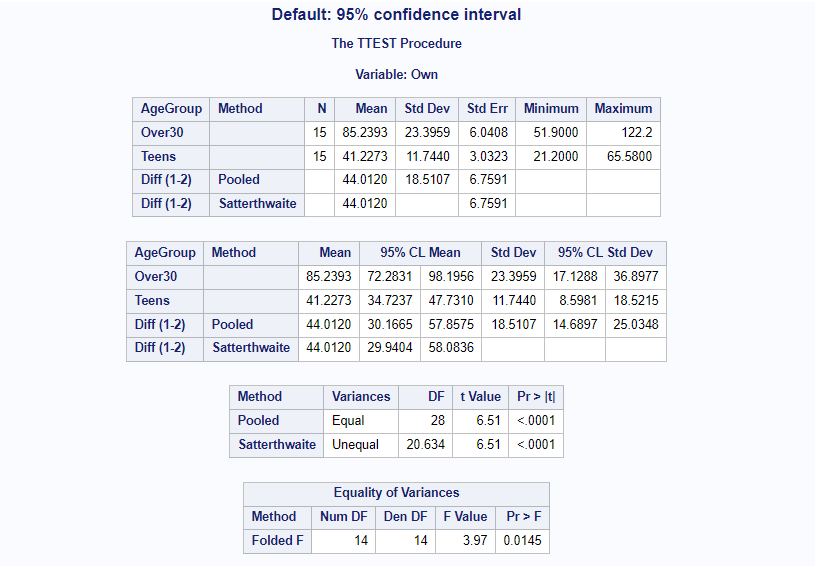
## [1] 30.16656

ub

## [1] 57.85744

This gives a 95% Confidence Interval for the Difference to be between (30.167, 57.857), where units are the difference in the amount of time it took Over 30-year olds compared to teens (I believe in seconds, but that would be wild, wouldn’t it?)

1. Provide a screenshot of the SAS output and use it to verify your calculation.



95% Confidence Interval SAS

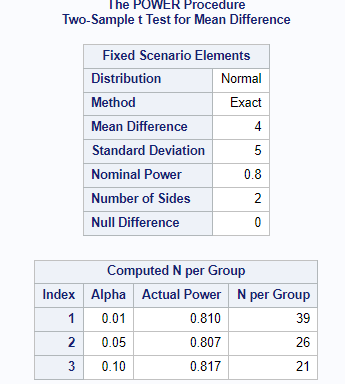
1. Interpret the confidence interval in the context of the problem.

A 95% Confidence Interval can be interpreted as a calculated range within which we can be 95% certain the true effect (true difference between two groups) lies.

Within the context of this particular question, it can be interpreted as: We are 95% Confident that the true difference between people Over 30 and Teens texting speeds is between 30.167 and 57.857 seconds; or is would take people Over 30 30.167 to 57.857 more seconds to type a specified message compared to Teens (within the context of being 95% confident).

This may also be interpreted as a commentary on the procedure of calculating the Confidence Interval: If we repeated this procedure of experimentation and calculation and constructed their respective 95% confidence intervals, these confidence intervals would contain the true difference between Over 30 and Teens texting times 95% of the time.

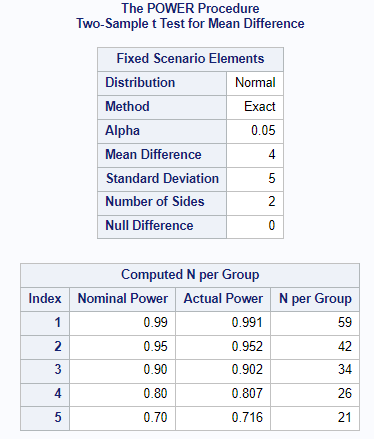
1. Use SAS to explore sample size determinations for the bone loss example using the **hypothesis testing method** and complete the following exercises:
2. Explore the effect of changing just the significance level - For , what are the resulting sample sizes? Summarize your findings in one concise sentence.



Alpha

For greater significance levels we require an increasing number of samples per Group, and the amount these sample sizes increase by is non-linear, i.e. each decrease of in (greater significant level) requires a larger number of samples to be added per Group compared to its prior significance level.

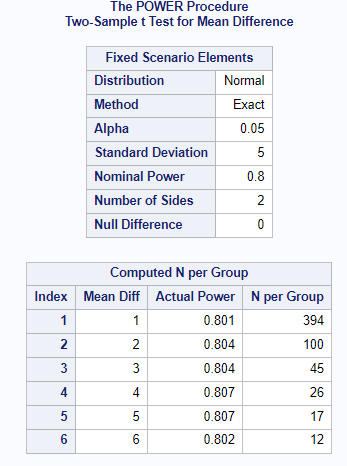
1. Explore the effect of changing just the power - For , what are the resulting sample sizes? Summarize your findings in one concise sentence.



Beta

When changing just the power in relation to the required sample size, we see that greater power (smaller ) requires larger sample sizes, and the increase in sample sizes between power levels becomes larger and larger for smaller and smaller ’s.

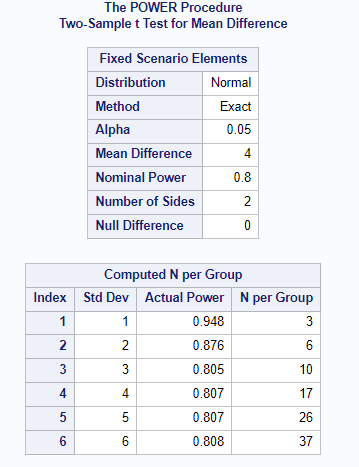
1. Explore the effect of changing just the true effect size - For , what are the resulting sample sizes? Summarize your findings in one concise sentence.



Effect Size

We require larger sample sizes to detect smaller effect sizes, i.e. the larger the effect size, the smaller the calculated sample size required, holding all else equal.

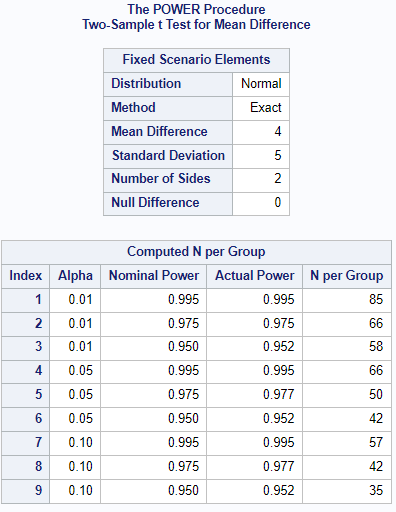
1. Explore the effect of changing just the estimated population variance - For , what are the resulting sample sizes? Summarize your findings in one concise sentence.



95% Confidence Interval SAS

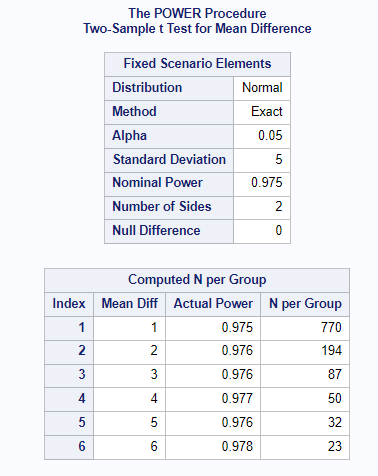
As the estimated population variance increases, we need increasing larger sample sizes, and the rate at which these sample sizes increase is increasing, e.g. the difference (change in sample size) between 5 and 6 Std Dev is larger than the difference between 1 and 2 Std Dev.

1. Use SAS to explore sample size determinations for the bone loss example using the **confidence interval method** and complete the following exercises:
2. Explore the effect of changing just the significance level - For , what are the resulting sample sizes? Summarize your findings in one concise sentence.

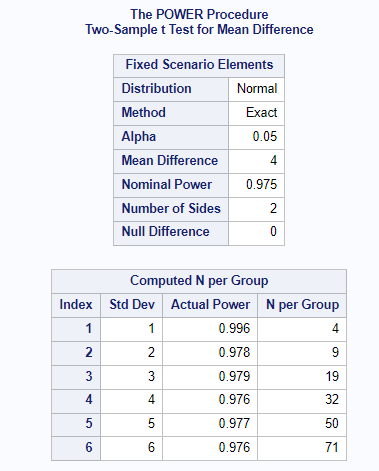


Consistent with the findings of Q2, albeit via a different method: Higher significance levels (lower ) require significantly larger sample sizes per group.

1. Explore the effect of changing just the true effect size - For , what are the resulting sample sizes? Summarize your findings in one concise sentence.

 Consistent with the findings of Q2, 3. albeit for a different method: We require much larger sample sizes per Group for smaller differences between groups, and smaller sample sizes per Group for larger differences between groups.

1. Explore the effect of changing just the estimated population variance - For , what are the resulting sample sizes? Summarize your findings in one concise sentence.



Consistent with the findings of Q2, 4. albeit for a different method: Increasing estimated population variance results in much larger required sample sizes, and the rate of increase for these samples grows larger as the variance increases.

1. Think about how the sample size determination using the confidence interval method relates to the standard error method. Summarize your findings in one concise sentence.

These two methods are related in that both methods require and take as input the **pooled estimate of the population variance** to estimate the required sample size.

**Total:** 50 points **# correct:** **%:**