APPM 4570/5570 - Take Home Exam 2 - Spring 2017

Josh Miltier

PROBLEM #1

H_0: $\sigma^2_1 = \sigma^2_2$, H_A: $\sigma^2_1 \neq \sigma^2_2$, μ=75 seconds, σ=20 seconds

Problem #1, part 1.

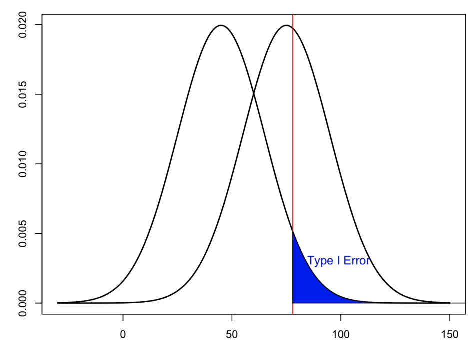
CODE:

```
var_old_data=0
var_new_data=0
var_diff=0
for(i in 1:200000){
   old_machine_data=rnorm(200,75,20)
   new_machine_data=rnorm(200,45,20)

   var_old_data[i]=var(old_machine_data)
   var_new_data[i]=var(new_machine_data)
   var_diff[i]=var_old_data[i]-var_new_data[i]
}
mean(var_old_data)-mean(var_new_data)

y = qnorm(.95, mean(var_diff), sd(var_diff))
count = 0;
for(i in 1:200000){
   if ((var_diff[i]) > y){
      count <- count + 1
   }
}

count/200000
```



OUTPUT:

> mean(var_old_data)-mean(var_new_data)
[1] -0.03022414
> qnorm(.95, mean(var_diff), sd(var_diff))
[1] 93.26884
> count/200000
[1] 0.050605 <~ Amount greater than the 5%

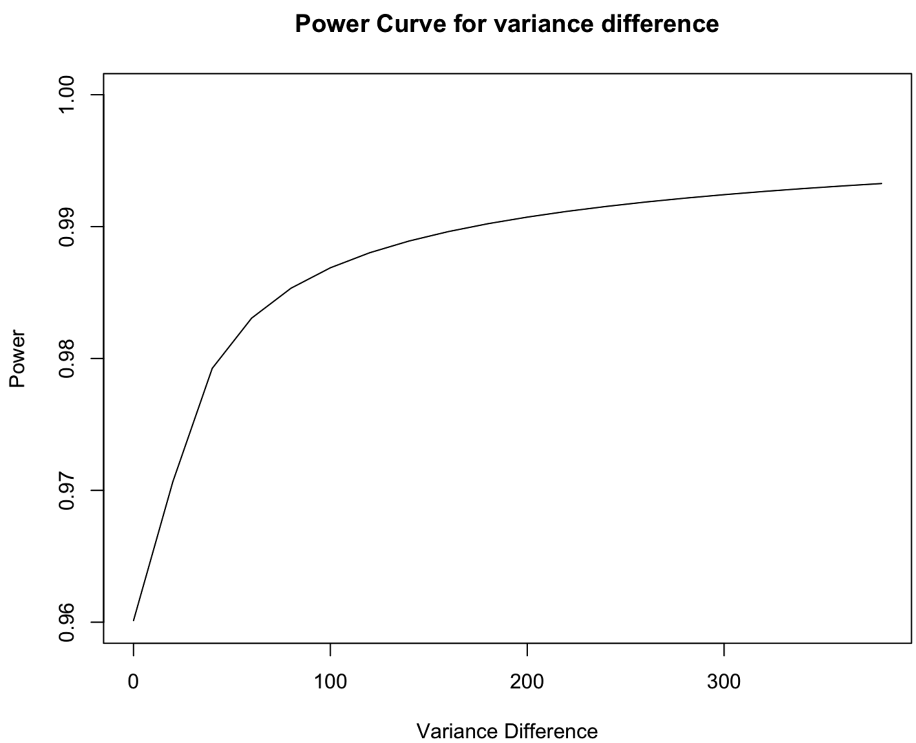
Problem #1, part 2.

CODE:

```
var_diff = seq(0,380,20)
x_critical_.05 = (qnorm(.95) * (20/sqrt(200))) + 75

type_II_.05=.05
for (i in 2:20){
    type_II_.05[i] = ((x_critical_.05 - 75) / sd_diff[i]/sqrt(200))
}

plot(var_diff, 1-(pnorm(type_II_.05)), ylim=c(.96,1), type="l", ylab="Power", xlab="Variance Difference",
                        main="Power Curve for variance difference")
```



As the variance difference increases, the power decreases meaning that the Type II Error increases. This makes sense because the greater variance difference is changing further away from the truth of the null hypothesis.

<u>Problem #1, part 3.</u>

<u>CODE:</u>

```
power.t.test(delta=(75/45), sd=sqrt(380),power=.8,sig.level=.05)
```

<u>OUTPUT:</u>

```
> power.t.test(delta=(75/45), sd=sqrt(380),power=.8,sig.level=.05)

Two-sample t test power calculation

              n = 2148.414
          delta = 1.666667
             sd = 19.49359
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

The approximate sample size to determine the variance is less than or equal to 380 is ~2149. This sample number is realistic if nearly 2 days (for the old machine) of sampling is workable by the statistician. Otherwise, spending ~45 hours collecting data is not realistic, then the sample size to collect is also not realistic.

PROBLEM #2

Problem #2, part 1.

CODE:

```
balanceData=read.table('BalanceData.txt', header=TRUE)
balanceData

for (i in 1:120){
	elderly[i]=balanceData$BalanceMeasure[i]
}
for(i in 120:270){
	young[i-120]=balanceData$BalanceMeasure[i]
}

alpha=.01

CI_elderly = c(mean(elderly)-qnorm(1-alpha/2)*sd(elderly)/sqrt(length(elderly)), mean(elderly)+qnorm(1-alpha/
	2)*sd(elderly)/sqrt(length(elderly)))
CI_young = c(mean(young)-qnorm(1-alpha/2)*sd(young)/sqrt(length(young)), mean(young)+qnorm(1-alpha/
	2)*sd(young)/sqrt(length(young)))

CI_elderly; CI_young
```

OUTPUT:

```
> CI_elderly; CI_young
	[1] 23.44857 28.14829     #ELDERLY CONFIDENCE INTERVAL
	[1] 16.82857 18.62567     #YOUNG CONFIDENCE INTERVAL
```

With 99% confidence, the true mean is within the represented confidence intervals above, with a significance level of .01.

Problem #2, part 2.

CODE:

```
CI_diff = c((mean(elderly)-mean(young))-qnorm(1-alpha/2)*sqrt((sd(elderly)^2/length(elderly)) +
	(sd(young)^2/length(young))), mean((mean(elderly)-mean(young)))+qnorm(1-alpha/
	2)*sqrt((sd(elderly)^2/length(elderly)) + (sd(young)^2/length(young))))
CI_diff
```

OUTPUT:

```
> CI_diff
	[1]  5.555509 10.587105
```

The data is not a surprise, looking at the variation of the data as well as the confidence intervals from part 1.

Problem #2, part 3.

Claim: young men sway at least 7.5 mm (on average) less than elderly men
Opp: young men sway more than 7.5 mm (on average) than elderly men

H_0: $\mu_{Elderly} - \mu_{Young} = 7.5$ mm, H_A: $\mu_{Elderly} - \mu_{Young} < 7.5$ mm

CODE:

```
t.test(elderly, young, conf.level=.99, alternative = c("less"), mu=7.5)
```

OUTPUT:

```
> t.test(elderly, young, conf.level=.99, alternative = c("less"), mu=7.5)

        Welch Two Sample t-test

        data:  elderly and young
        t = 0.58494, df = 153.72, p-value = 0.7203
        alternative hypothesis: true difference in means is less than 7.5
        99 percent confidence interval:
          -Inf 10.36737
        sample estimates:
        mean of x mean of y
           25.79843  17.72712
```

Based on the 99% confidence interval from part 2, we Fail to Reject the null hypothesis. There is not sufficient evidence to support that this claim is is true. Supportive data based on the p-value also suggests p-value > sig. level thus FTR H_0.

Problem #2, part 4.

The Type I Error for this test = $\alpha = 0.01$.