Math 181B: Homework 6

Merrick Qiu

Exercise 1

$$\begin{aligned}
E[SSTOT] &= E[SSTR + SSE] \\
&= E[SSTR] + E[SSE] \\
&= E[SSTR] + E\left[\sum_{j=1}^{k}(n_j - 1)S_j^2\right] \\
&= E[SSTR] + \sum_{j=1}^{k}(n_j - 1)\sigma^2 \\
&= E[SSTR] + (n-k)\sigma^2 \\
&= (k-1)\sigma^2 + \sum_{j=1}^{k} n_j(\mu_j - \mu)^2 + (n-k)\sigma^2
\end{aligned}$$

Exercise 2

```
setwd("C:/Users/merri/Documents/MATH-31H/MATH 181B/Homework 6")
teen = read.csv("Adolescent.csv")$x
adult = read.csv("Adult.csv")$x
middle = read.csv("MiddleAge.csv")$x

People = data.frame(times = c(teen, adult, middle),
                    Age= c(rep('Teen', length(teen)), rep('Adult', length(adult)),
                    rep('Middle', length(middle))))
summary(aov(times~Age, People))

# > summary(aov(times~Age, People))
#              Df Sum Sq Mean Sq F value    Pr(>F)
# Age          `2  173.6   86.80   22.46 3.08e-09 ***
# Residuals 147  568.1    3.86
# ---

# H0: mu_teen = mu_adult = mu_middle, H1: At least one mu is different
# Assume that Yij are independent, normal, and
# have the same variation for each j
# Since the p-value of 3.08e-09 is <0.05, we reject the null hypothesis and
# the mean screen times of all the age groups are not the same.
```

No, you cannot conclude that average adolescents spend more time in front of the screen than the other two groups because the conclusion only says that at least one group has a different mean. For the same reason, you cannot conclude that adults spend more time than middle aged adults.

```
t.test(teen, adult, var.equal = T) # p-value = 1.667e-05
t.test(teen, middle, var.equal = T) # p-value = 2.448e-09
t.test(adult, middle, var.equal = T) # p-value = 0.04485

# Using the Bonferroni correction we use alpha/n = 0.01667
# Since mean of teen is greater than adult and middle and
# the p-value is <alpha/n, we conclude that teens have more
# screen time on average than adults and middle.
# Since 0.04485 > alpha/n, we cannot conclude that adults
# watch more tv than middle.
```

Exercise 3

```
reject <- function(p) {
  first = rnorm(50, 1, 1);
  second = rnorm(50, 1, 1);
  third = rnorm(50, 1, 1);
  p1 = t.test(first, second, var.equal = T)$p.value;
  p2 = t.test(second, third, var.equal = T)$p.value;
  p3 = t.test(first, third, var.equal = T)$p.value;
  return (p1<p | p2<p | p3<p);
}

regular_errors = rep(0, 1000);
bonferroni_errors = rep(0, 1000);
for(i in 1:1000) {
  regular_errors[i] = reject(0.05)
  bonferroni_errors[i] = reject(0.05/3)
}
sum(regular_errors)/1000
sum(bonferroni_errors)/1000
# > sum(regular_errors)/1000
# [1] 0.118
#
# > sum(bonferroni_errors)/1000
# [1] 0.042

# The proportion of type 1 errors is indeed <0.05 for the bonferroni correction,
# but it is more than double 0.05 without the correction.
```