HW 3

Adon Rosen

Date: 2019-10-27

Question 1

Question 1a

```r
mean.se.vals <- summarySE(data = in.data, measurevar = 'salary', groupvars = 'gender')
## Declare our variables we will need to calculate our t statistic
m.m <- mean.se.vals$salary[2]
m.sd <- mean.se.vals$sd[2]
m.size <- mean.se.vals$N[2]
f.m <- mean.se.vals$salary[1]
f.sd <- mean.se.vals$sd[1]
f.size <- mean.se.vals$N[1]

## Now calc the t stat
pooled.standard <- sqrt(((m.size-1)*m.sd^2+(f.size-1)*f.sd^2)/(m.size+f.size-2))
t.stat.by.hand <- (m.m-f.m)/(pooled.standard*sqrt(1/m.size+1/f.size))

## Now calulate the p value for the t statistic
t.stat.by.hand.p.value <- pt(t.stat.by.hand, df=298, lower.tail = F)

print(t.stat.by.hand)
```

```
[1] 16.82665
```

```r
print(t.stat.by.hand.p.value)
```

```
[1] 1.990191e-45
```

Question 1b

```r
out.val <- t.test(salary ~ gender, data=in.data, var.equal=T)
to.report.stat <- out.val$statistic
to.report.prob <- out.val$p.value
print(to.report.stat)
```

```
        t
-16.82665
```

```r
print(to.report.prob)
```

```
[1] 3.980382e-45
```

Utiliizng a t test to explore differences in means between the two genders, we conclude the salarys do display significant differences across genders. The resultant models statistic was -16.83 with a p value of 0. This value reflects the probability of observing this difference in means given undera true null hypothesis.

Question 1c

Utilizing a variance test to explore differences in group variance, we fail to reject the null hypothesis that the genders display equivalent variance. The resultant models statisc was 0.79 with a p value of 0.19792.

Question 1d

Differences in mean salary were tested between male () and female () professors.First, assumptions for a parametric t test were checked by measuring for differences in sample variance using an f-test, here we observed a nonsignificant difference in variance between the two populations (F(99,199)=.79, p=.19). This suggests the data are suitable for a parametric t-test. Significant differences in means were observed using a t test (t(219)=17, p<.0005). This value suggested that male colleagues have greater mean salaries than their female counter parts.

Question 2

Question 2a

Report my hypotheses here:

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_a : \mu_i \neq \mu_j \text{ For some i and j}$$

```
age.group <- c(rep("Young Adult", 4), rep("Adult", 4), rep("Older Adult", 4))
wais.vals <- c(17,21,20,18,16,12,14,14,9,11,8,8)
in.data <- data.frame(age.group, wais.vals)
in.data.sum <- summarySE(data=in.data, measurevar = 'wais.vals', groupvars = 'age.group')
in.data$SumSquaresW <- NA
in.data$SumSquaresW[1:4] <- (in.data$wais.vals[1:4] - in.data.sum$wais.vals[which(in.data.sum$age.group=
in.data$SumSquaresW[5:8] <- (in.data$wais.vals[5:8] - in.data.sum$wais.vals[which(in.data.sum$age.group=
in.data$SumSquaresW[9:12] <- (in.data$wais.vals[9:12] - in.data.sum$wais.vals[which(in.data.sum$age.grou
in.data$SumSquaresB <- (in.data$wais.vals - mean(in.data$wais.vals))^2
sum.of.squares.between <- sum(in.data$SumSquaresB) - sum(in.data$SumSquaresW)
sum.of.squares.within <- sum(in.data$SumSquaresW)
mean.squares.between <- sum.of.squares.between / (3 - 1)
mean.squares.within <- sum.of.squares.within / (12 - 3)
f.stat <- mean.squares.between / mean.squares.within
f.stat.p.val <- pf(q=f.stat, df1=2, df2=9, lower.tail = F)
```

The null hypothesis will be rejected with these data given the observed F statistic of 37.5 which has a p value of 4×10^{-5}. This in turn suggests these populations do not have equivalent means.

Question 2b

```
out.mod <- aov(wais.vals ~ age.group, data=in.data)
to.report <- summary(out.mod)
to.report.stat <- as.matrix(to.report)[[1]][1,4]
to.report.prob <- as.matrix(to.report)[[1]][1,5]
```

Question 3

```r
diff.vec <- c(4,4,1,2,-3,5,3,2,-4,2,1,-1,2,7,0,4,6,3,4,-1)
mean.diff <- mean(diff.vec)
sd.diff <- sd(diff.vec)
se.diff <- sd.diff / sqrt(length(diff.vec))
t.stat <- mean.diff / se.diff
t.stat.p.val <- pt(t.stat, df=19, lower.tail = F)
```