Stats 110 Homework 5

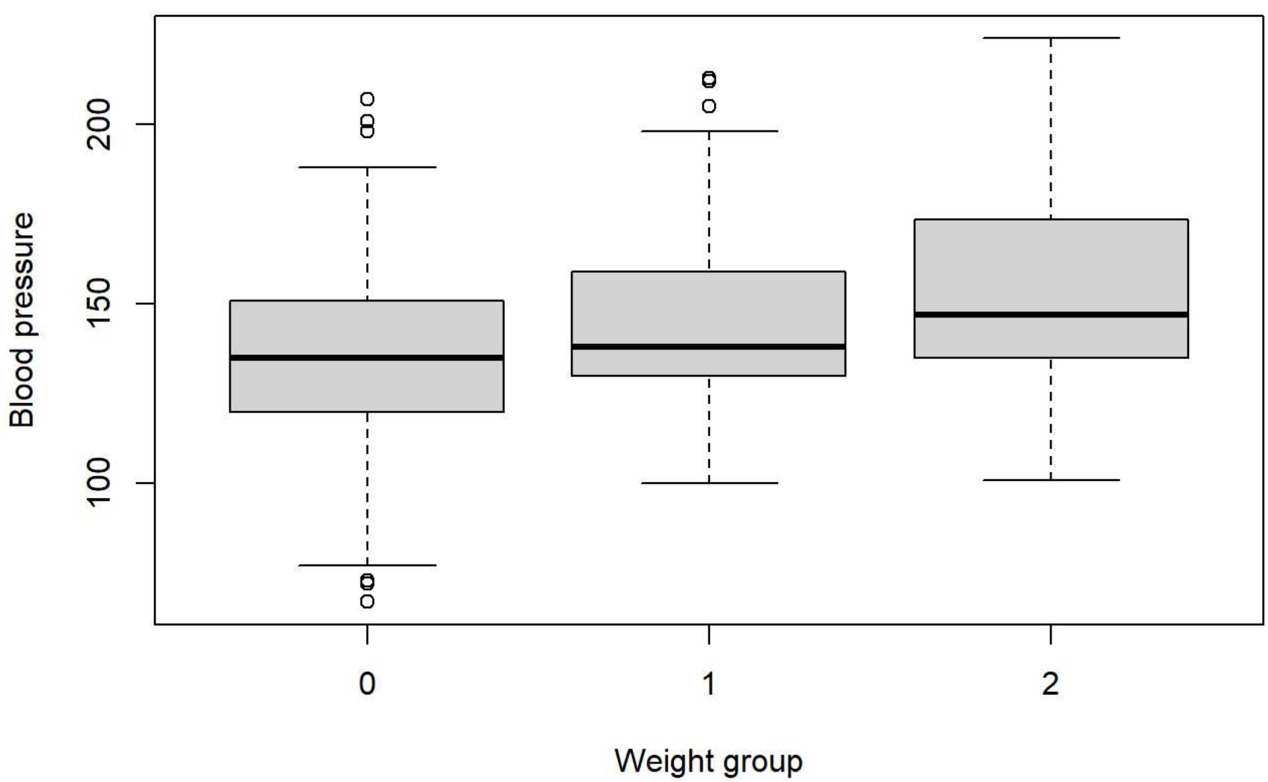
Owen Lin

11/29/2022

1. a. You can't randomly assign an ethnic group to a mother and then observe baby's weight. It can only be an observational study and thus omitted variable bias exist. For instant, it could be that mother from the same ethnic groups share a food preference shaped by the culture, which affect the birth weight of infants.
   b. We can not extend our sample's results to all US births because a random (representative) sample in North Carolina is not necessarily a representative sample of the US.

2. a. See Output

```
blood <- read.table("Blood.txt", fill = TRUE, header = TRUE)
blood$Overwt <- as.factor(blood$Overwt)
blood$Smoke <- as.factor(blood$Smoke)
boxplot(blood$SystolicBP ~ blood$Overwt, ylab = "Blood pressure", xlab = "Weight group")
```



b. Weight group 0 : mean = 136.31, sd = 27.26, size = 187
   Weight group 1 : mean = 144.36, sd = 25.07, size = 109
   Weight group 2 : mean = 153.18, sd = 27.81, size = 204

```
tapply(blood$SystolicBP, blood$Overwt, mean)
tapply(blood$SystolicBP, blood$Overwt, sd)
tapply(blood$SystolicBP, blood$Overwt, length)
```

```
##        0        1        2
## 136.3155 144.3670 153.1814
##        0        1        2
## 27.26852 25.07864 27.81397
##   0   1   2
## 187 109 204
```

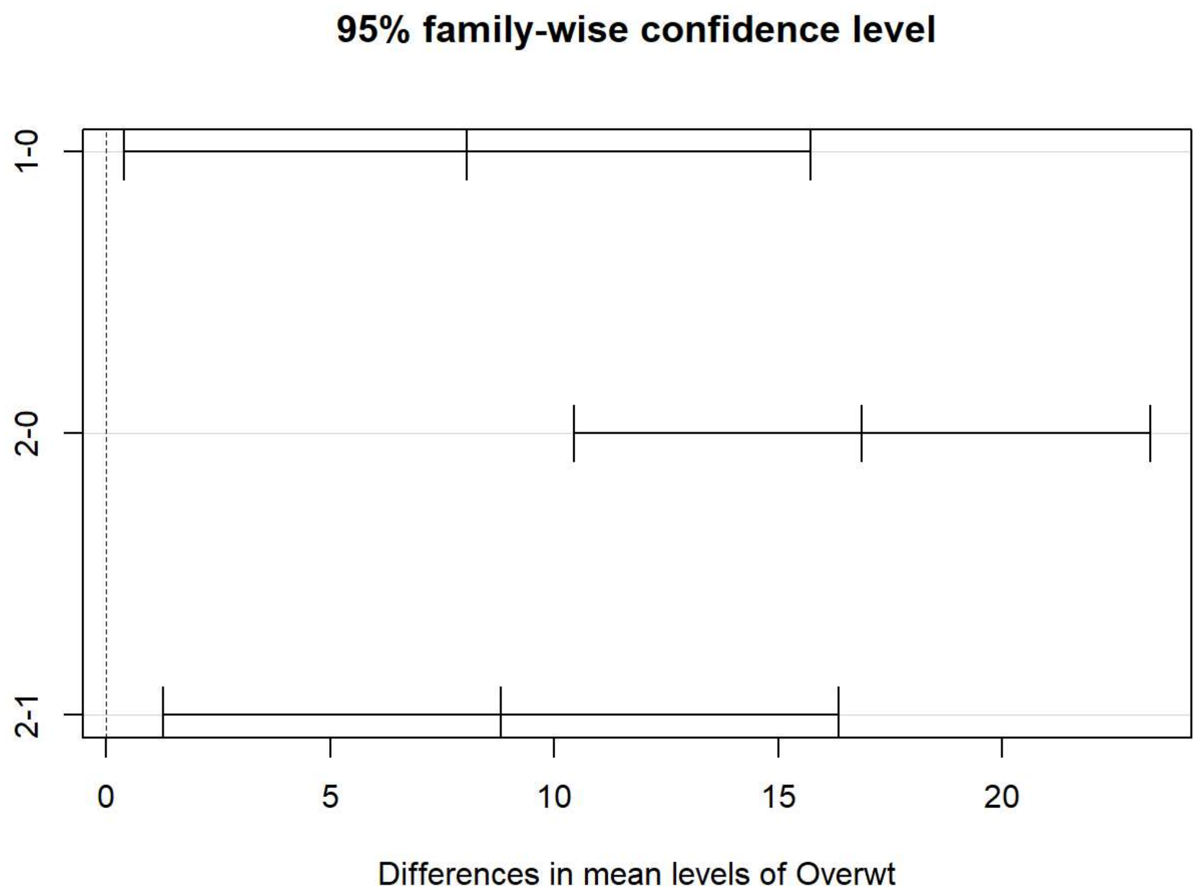
c. The sample standard deviations are roughly equal. Thus, it is appropriate to conduct an analysis of variance.
d. $SystolicBP_i = \mu + \alpha_k + \epsilon_{i,k}$, k = 0, 1, 2
   H0: $\alpha_k = 0$ for all k
   Ha: At least 1 α_k does not equal to zero.
e. H0: The difference between the mean blood pressure for three weight groups is 0.
   Ha: H0 is false.
f. Test statistic: 19.02 ~ F(2, 497) p-value: 1.1e-08 < 0.05 Conclusion: we reject the null and conclude at least one weight group

```
ano2f <- aov(SystolicBP ~ Overwt, data = blood)
summary(ano2f)
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## Overwt        2  27801   13900   19.02 1.1e-08 ***
## Residuals   497 363274     731
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

g. It seems like weight group 2 is significantly different than both 1 and 0 at 95% confidence

```
TukeyHSD(ano2f, ordered = T)
plot(TukeyHSD(ano2f, ordered = T))
```




```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##     factor levels have been ordered
##
## Fit: aov(formula = SystolicBP ~ Overwt, data = blood)
##
## $Overwt
##          diff        lwr      upr     p adj
## 1-0  8.051464  0.3927115 15.71022 0.0366867
## 2-0 16.865865 10.4316024 23.30013 0.0000000
## 2-1  8.814400  1.2740746 16.35473 0.0170703
```

h. $SystolicBP_i = \mu + \alpha_k + \beta_j + \epsilon_{i,k,j}$, k = 0, 1, 2; j = 0, 1

i. See Output.

```
ano2h <- aov(SystolicBP ~ Overwt + Smoke, data = blood)
summary(ano2h)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Overwt        2  27801   13900   19.53 6.84e-09 ***
## Smoke         1  10277   10277   14.44 0.000163 ***
## Residuals   496 352997     712
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

j. H0: $\beta_j = 0$ for all j.
   Ha: H0 is false.
   Test statistic: 14.44 ~ F(1, 496)
   p-value: 0.000163 < 0.05
   Conclusion: we reject the null and conclude that smoke does
k. $SystolicBP_i = \mu + \alpha_k + \beta_j + \gamma_{kj} + \epsilon_{ikj}$, k = 0, 1, 2; j = 0, 1
l. H0: $\gamma_{kj} = 0$ for all j.
   Ha: H0 is false.
   Test statistic: 0.539 ~ F(2, 494)
   p-value: 0.583614 > 0.05
   Conclusion: we fail to reject the null and conclude that Overwt and Smoke do not have an interaction effect on the mean SystolicBP.

```
ano2l <- aov(SystolicBP ~ Overwt + Smoke + Overwt*Smoke, data = blood)
# summary(ano2L)
```

3. a. Response variable: amount eaten
   First factor: Male or Female
   Second factor: leptin or insulin injection
   b. First factor: Observational with 2 levels
   Second factor: experimental with 2 levels
   c. The results indicate that there is an interaction between the two factors in their effect on the response, because the affect of second factor is different base on gender.

4. a. Response variable: score on the set of math problems
   First factor: hyperactive or not
   Second factor: high noise and low noise
   b. First factor: Observational with 2 levels
   Second factor: experimental with 2 levels
   c. The results indicate that there is an interaction between the two factors in their effect on the response, because the performance under high noise and low noise is different base on the level of first factor.

5. a. μ = (12+17+14+16)/4 = 14.75
   b. α_1 = (12+17)/2 - 14.75 = -0.25
   α_2 = (14+16)/2 - 14.75 = 0.25
   c. β_1 = (12+14)/2 - 14.75 = -1.75
   β_2 = (17+16)/2 - 14.75 = 1.75
   d. γ_{11} = 12 - 14.75 - (-0.25 + -1.75) = -0.75
   γ_{12} = 17 - 14.75 - (-0.25 + 1.75) = 0.75
   γ_{21} = 14 - 14.75 - (0.25 + -1.75) = 0.75
   γ_{22} = 16 - 14.75 - (0.25 + 1.75) = -0.75

6. K=2, J=2, n=25 (N = 100)

```
table <- data.frame(Source = c('Face', 'Gender', 'Interaction', 'Residual', 'Total'), df = c(1,
1, 1, 96, 99), SS = c(12915, 2500, 400, 9600, 25415), MS = c(12915, 2500, 400, 100, 15915), F =
 c(129.15, 25, 4, NA, NA))

table
```

```
##        Source df    SS    MS      F
## 1        Face  1 12915 12915 129.15
## 2      Gender  1  2500  2500  25.00
## 3 Interaction  1   400   400   4.00
## 4    Residual 96  9600   100     NA
## 5       Total 99 25415 15915     NA
```