B27
Hector G. Flores Rodriguez

Stats 110 - HW7
November 29, 2017

1. Ex. 5.15

a. Not possible.

The residuals are defined as: $\hat{\epsilon} = y - \bar{y}_k$. The average for a group, $\bar{y}_k$, cannot be be smaller than all the observed values for that group. The only way that is possible is to subtract a value less than or equal to the $\min y_{k,j}\ \forall j$.

b. Possible.

E.g. For some group i, $y_{i,1} = 5$, $y_{i,2} = 4$, $y_{i,3} = 3$, $y_{i,4} = -1$

$\bar{y_i}. = 2.75$

It is easy to see that all residuals except for $y_{i,4}$ will be positive

c. Possible.

If a higher score means a lower number.

d. Possible.

If the average for the higher score group is larger than the average for the lower score group.

2. Ex. 7.8

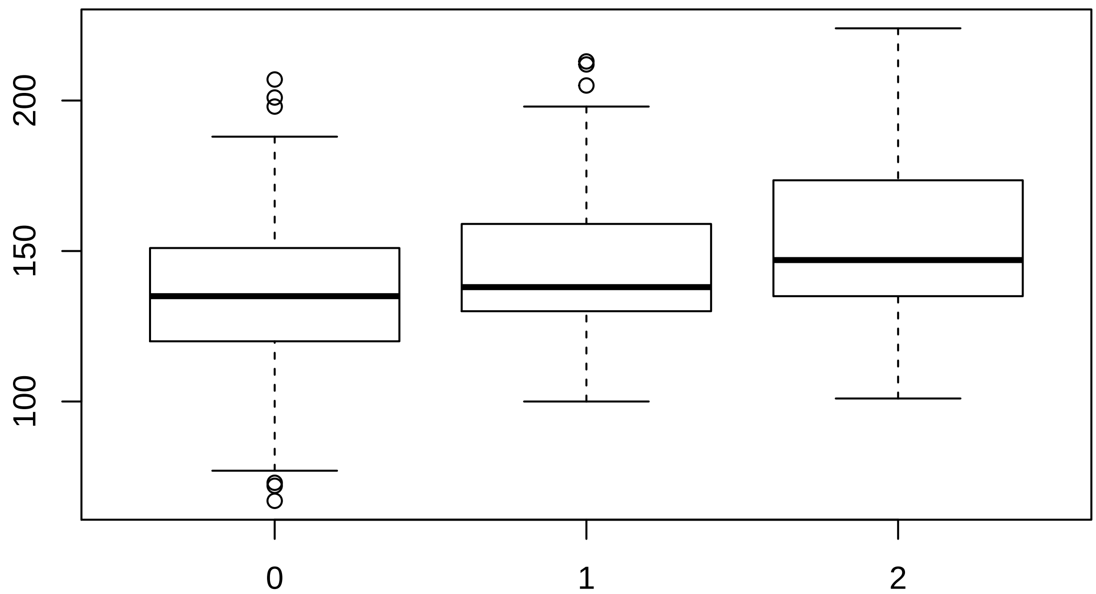
The one pice of the formula that is different between Bonferroni, Tukey, and Fisher is the critical value.

3.

Categorical variables define group membership, so there lacks a continuum on the real line between categorical variables. This means that there are no leverage points since extremes values on the the x-axis will most likely refere to one group or another.

4.

```
data = read.csv("../data/Blood1.csv")
wt = as.factor(data$Overwt)
boxplot(SystolicBP~wt, data=data)
```



5.

```r
means = tapply(data$SystolicBP, wt, mean)
sds = tapply(data$SystolicBP, wt, sd)
ns = tapply(data$SystolicBP, wt, length)

summary = rbind(means, sds, ns)
print(summary)
```

```
##                0         1         2
## means 136.31551 144.36697 153.18137
## sds    27.26852  25.07864  27.81397
## ns    187.00000 109.00000 204.00000
```

6.

Yes. The standard deviations are approximately similar however there is a shift in the mean for the obese group.

7. Write down the population model and null and alternative hypothesis for ANOVA

7-(a). Using group means. Define parameters. Specify the conditions

Model: $Y_{ij} = \mu_i + \epsilon_{ij}$

Y_{ij} := j^{th} observation under the i^{th} group

μ_i := mean of the i^{th} group

ϵ_{ij} := random error

Assumptions: $\epsilon_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$

7-(b). Factor effects version.

Model: $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$

μ := overall mean level

α_i := differential effect of group i

7-(c). Null and alternative hypothesis using means then factor effects

Means:

H_0: $\mu_1 = \mu_2 = \cdots = \mu_I$

H_1: At least two population means are different

Factor effect:

H_0: $\alpha_1 = \cdots = \alpha_I = 0$

H_1: At least one $\alpha_i \neq 0$

8. Conduct one-way ANOVA to compare mean blood pressure

8-(a). State null and alternative hypothesis

H_0: $\mu_0 = \mu_1 = \mu_2$, where μ_0 is mean normal group; μ_1 is mean overweight group; μ_0 is mean obese group

H_1: At least two population means are different

8-(b). Show anova table

```r
res = aov(SystolicBP ~ wt, data=data)
summary(res)
```

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

8-(c). Test statistic and p-value

```r
oneway.test(SystolicBP ~ wt, data=data, var.equal=TRUE)
```

```
## 
##  One-way analysis of means
## 
## data:  SystolicBP and wt
## F = 19.017, num df = 2, denom df = 497, p-value = 1.101e-08
```

8-(d). Conclusion

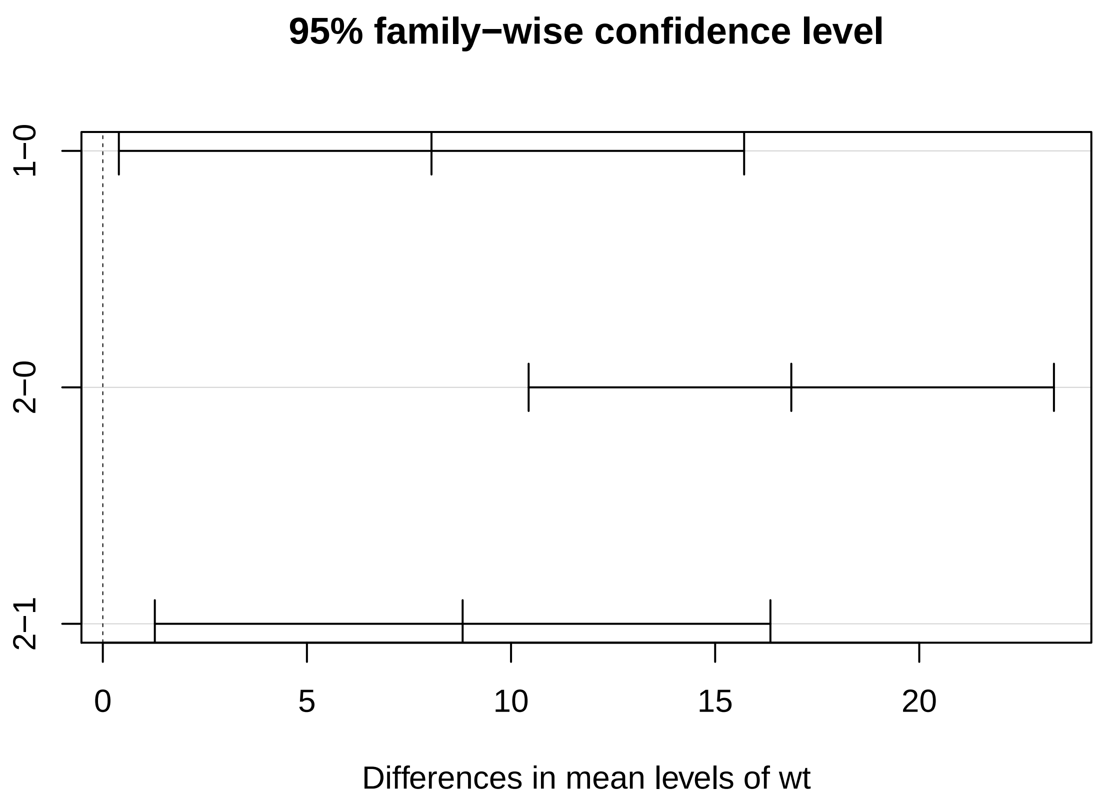
Since the p-value is small, we reject the null hypothesis in favor of the alternative. Hence, the mean levels of systolic blood pressure are different across the three weight groups (normal, overweight, obese).

9. Use Tukey multiple comparison procedure to determine which population means differ.

```r
tukey_test = TukeyHSD(res)
print(tukey_test)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = SystolicBP ~ wt, data = data)
##
## $wt
##           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
```

```r
plot(tukey_test)
```



10. Summary.

Here we compared the systolic blood pressure among three different weight groups: 0=normal; 1=overweight; 2=obese. With 95% confidence, we have found that the mean levels of systolic blood pressure are different across the three weight groups. Furthermore, with 95% confidence, we conclude that there is a significant difference in the mean levels of systolic blood pressure between: normal and overweight; normal and obese; overweight and obese.