Project_II

Brenda Maldonado

9/26/2021

Analysis of Variance The heartbpchol.csv data set contains continuous cholesterol (Cholesterol) and blood pressure status (BP_Status) (category: High/ Normal/ Optimal) for alive patients.

```
chol <- read.csv('heartbpchol.csv')
chol <- chol[order(chol$BP_Status),]
head(chol)
```

```
##    Cholesterol BP_Status
## 2          188      High
## 3          292      High
## 6          247      High
## 9          228      High
## 10         280      High
## 11         305      High
```

For the heartbpchol.xlsx data set, consider a one-way ANOVA model to identify differences between group cholesterol means. The normality assumption is reasonable, so you can proceed without testing normality.

a) Perform a one-way ANOVA for Cholesterol with BP_Status as the categorical predictor. Comment on statistical significance of BP_Status, the amount of variation described by the model, and whether or not the equal variance assumption can be trusted.

```
chol %>%
  group_by(BP_Status) %>%
  summarise(count = n(),
            mean = mean(Cholesterol,na.rm = TRUE),
            sd = sd(Cholesterol, na.rm=TRUE))
```

```
## # A tibble: 3 x 4
##   BP_Status count  mean    sd
##   <chr>     <int> <dbl> <dbl>
## 1 High        229  241.  44.7
## 2 Normal      245  229.  43.2
## 3 Optimal      67  222.  39.7
```

b) Comment on any significantly different cholesterol means as determined by the post-hoc test comparing all pairwise differences. Specifically explain what that tells us about differences in cholesterol levels across blood pressure status groups, like which group has the highest or lowest mean values of Cholesterol.

Exercise 2: Analysis of Variance For this problem use the bupa.csv data set. Check UCI Machine Learning Repository for more information (http://archive.ics.uci.edu/ml/datasets/Liver+Disorders). The mean corpuscular volume and alkaline phosphatase are blood tests thought to be sensitive to liver disorder related to excessive alcohol consumption. We assume that normality and independence assumptions are valid.

```
bupa <- read.csv('bupa.csv')
bupa %>% head()
```

```
##   mcv alkphos drinkgroup
## 1  85      92          1
## 2  85      64          1
## 3  86      54          1
## 4  91      78          1
## 5  87      70          1
## 6  98      55          1
```

a) Perform a one-way ANOVA for mcv as a function of drinkgroup. Comment on significance of the drinkgroup, the amount of variation described by the model, and whether or not the equal

variance assumption can be trusted

```
aov(mcv~drinkgroup,bupa)
```

```
## Call:
##    aov(formula = mcv ~ drinkgroup, data = bupa)
##
## Terms:
##                 drinkgroup Residuals
## Sum of Squares     596.355  6209.877
## Deg. of Freedom          1       343
##
## Residual standard error: 4.25495
## Estimated effects may be unbalanced
```

b) Perform a one-way ANOVA for alkphos as a function of drinkgroup. Comment on statistical significance of the drinkgroup, the amount of variation described by the model, and whether or not the equal variance assumption can be trusted.