ANOVA Tutorial Sheet Solutions

## QUESTION 1.

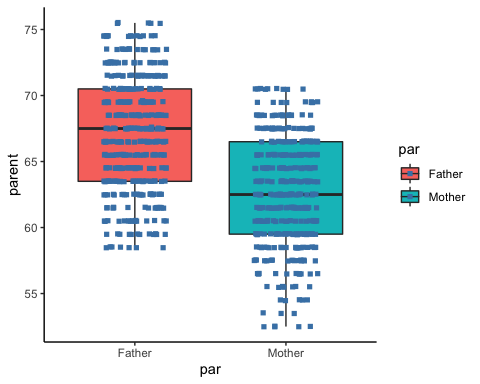
Read in data and ;

MYPEARSON<-read.csv("PearsonLeeSimple.csv")  
Pearson\_child\_parent<-lm(parent~par,data=MYPEARSON)  
summary(Pearson\_child\_parent)

##   
## Call:  
## lm(formula = parent ~ par, data = MYPEARSON)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10.2607 -3.2607 0.3761 3.3761 8.3761   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 67.1239 0.2151 312.03 <2e-16 \*\*\*  
## parMother -4.3632 0.3092 -14.11 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.221 on 744 degrees of freedom  
## Multiple R-squared: 0.2111, Adjusted R-squared: 0.21   
## F-statistic: 199.1 on 1 and 744 DF, p-value: < 2.2e-16

* ii Child’s height vs father’s height;

library(ggplot2)  
ggplot(MYPEARSON, aes(x = par, y = parent, fill = par)) +  
 geom\_boxplot() +  
 geom\_jitter(shape = 15,  
 color = "steelblue",  
 position = position\_jitter(0.21)) +  
 theme\_classic()

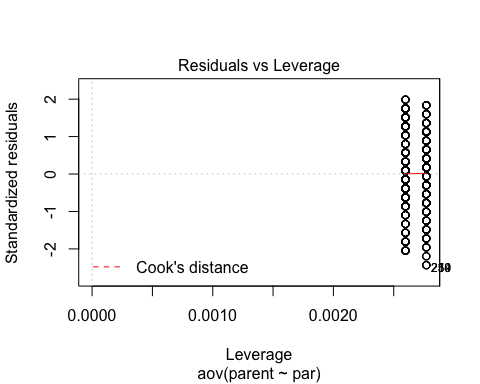
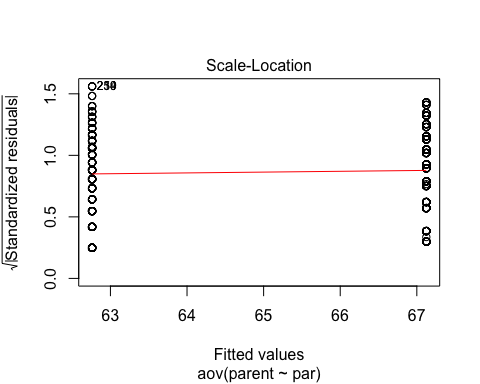
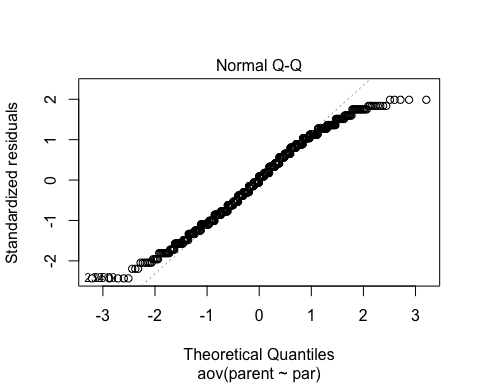
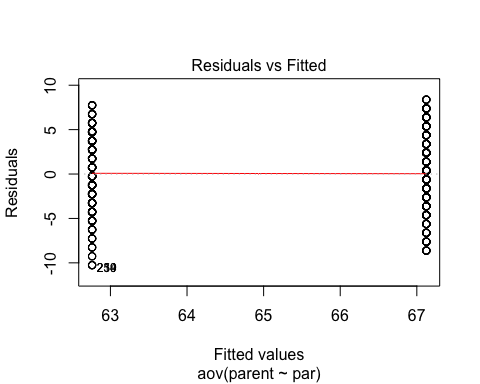


* iii Child’s height vs mother’s height;

anova\_one\_way <- aov(parent~par, data = MYPEARSON)  
summary(anova\_one\_way)

## Df Sum Sq Mean Sq F value Pr(>F)   
## par 1 3547 3547 199.1 <2e-16 \*\*\*  
## Residuals 744 13255 18   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

plot(anova\_one\_way)



## Tukey Follow Up

TukeyHSD(anova\_one\_way)

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = parent ~ par, data = MYPEARSON)  
##   
## $par  
## diff lwr upr p adj  
## Mother-Father -4.363231 -4.970319 -3.756144 0

## t-test

T\_test <- t.test(parent~par, data = MYPEARSON,paired = FALSE)  
T\_test

##   
## Welch Two Sample t-test  
##   
## data: parent by par  
## t = 14.124, df = 743.16, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 3.756744 4.969719  
## sample estimates:  
## mean in group Father mean in group Mother   
## 67.12390 62.76066

## BIGGER ANOVA

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✓ tibble 3.1.2 ✓ dplyr 1.0.7  
## ✓ tidyr 1.1.3 ✓ stringr 1.4.0  
## ✓ readr 1.4.0 ✓ forcats 0.5.1  
## ✓ purrr 0.3.4

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(reshape2)

##   
## Attaching package: 'reshape2'

## The following object is masked from 'package:tidyr':  
##   
## smiths

#TEST <- gather(MYPEARSON, child,parent -c(par,chl))  
  
TEST <-melt(MYPEARSON, id.vars = c("par","chl"), variable.name = "type")

## Warning: attributes are not identical across measure variables; they will be  
## dropped

summary(TEST)

## par chl type value   
## Father:1540 Daughter:1564 X :746 Length:2984   
## Mother:1444 Son :1420 child :746 Class :character   
## parent:746 Mode :character   
## gp :746