AnnovaPractice

Jitesh Rawat

2/23/2021

library(palmerpenguins)  
library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3 v purrr 0.3.4  
## v tibble 3.0.4 v dplyr 1.0.2  
## v tidyr 1.1.2 v stringr 1.4.0  
## v readr 1.4.0 v forcats 0.5.0

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

library(lattice)  
library(car)

## Loading required package: carData

##   
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':  
##   
## recode

## The following object is masked from 'package:purrr':  
##   
## some

head(penguins)

## # A tibble: 6 x 8  
## species island bill\_length\_mm bill\_depth\_mm flipper\_length\_~ body\_mass\_g sex   
## <fct> <fct> <dbl> <dbl> <int> <int> <fct>  
## 1 Adelie Torge~ 39.1 18.7 181 3750 male   
## 2 Adelie Torge~ 39.5 17.4 186 3800 fema~  
## 3 Adelie Torge~ 40.3 18 195 3250 fema~  
## 4 Adelie Torge~ NA NA NA NA <NA>   
## 5 Adelie Torge~ 36.7 19.3 193 3450 fema~  
## 6 Adelie Torge~ 39.3 20.6 190 3650 male   
## # ... with 1 more variable: year <int>

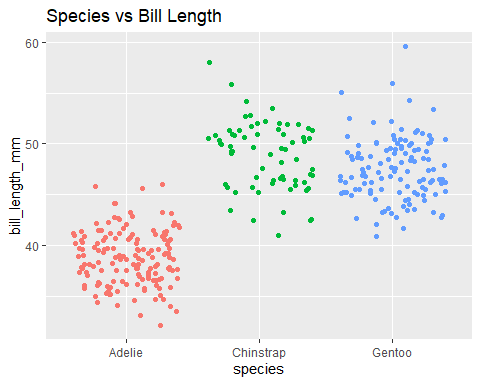
pdata <- penguins %>%   
 select(species,bill\_length\_mm)

summary(pdata)

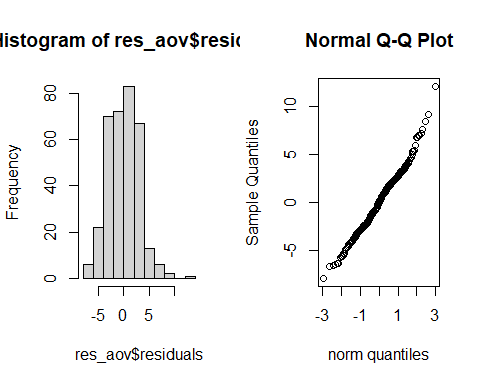
## species bill\_length\_mm   
## Adelie :152 Min. :32.10   
## Chinstrap: 68 1st Qu.:39.23   
## Gentoo :124 Median :44.45   
## Mean :43.92   
## 3rd Qu.:48.50   
## Max. :59.60   
## NA's :2

ggplot(pdata)+  
 aes(species,bill\_length\_mm,color=species)+  
 geom\_jitter()+  
 theme(legend.position="none")+  
 labs(title="Species vs Bill Length")

## Warning: Removed 2 rows containing missing values (geom\_point).



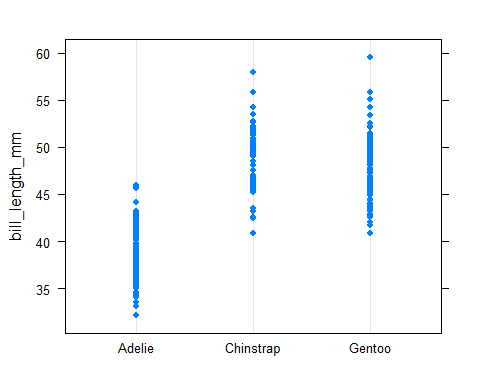
#Checking normailty visually  
  
res\_aov <- aov(bill\_length\_mm ~ species,data=pdata)  
  
par(mfrow=c(1,2)) # combine plots  
  
# histogram  
hist(res\_aov$residuals)  
  
qqnorm(res\_aov$residuals,xlab = "norm quantiles")



shapiro.test(res\_aov$residuals)

##   
## Shapiro-Wilk normality test  
##   
## data: res\_aov$residuals  
## W = 0.98903, p-value = 0.01131

#we check homogenity of variances wiht simple doptplot  
dotplot(bill\_length\_mm ~ species, data =pdata)



#Its pretty clear that the variances of the groups are similar

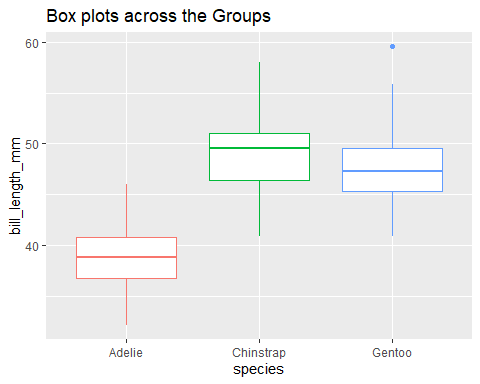
#still we apply levene test   
leveneTest(bill\_length\_mm ~ species, data=pdata)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 2 2.2425 0.1078  
## 339

#Doing this we see that p-value is .108 that is pretty good to say that the test says we accept null hypothesis that is variances are constant in groups

#We see some visulalizations before annova  
  
ggplot(pdata)+  
 aes(species,bill\_length\_mm,color=species)+  
 geom\_boxplot()+  
 theme(legend.position="none")+  
 labs(title="Box plots across the Groups")

## Warning: Removed 2 rows containing non-finite values (stat\_boxplot).



#from this plot we can see that species Adelie has smallest bills

#Lets apply one wat annova  
  
oneway.test(bill\_length\_mm ~ species,  
 data = pdata,  
 var.equal = TRUE # assuming equl variances  
 )

##   
## One-way analysis of means  
##   
## data: bill\_length\_mm and species  
## F = 410.6, num df = 2, denom df = 339, p-value < 2.2e-16

#This one way annova methods pvalue turns out to be less than 0.05 which means that the means are not equal  
#Just like how we saw from boxplot that one species has different bill length from other species

#install.packages("multcomp")  
  
library(multcomp)

## Loading required package: mvtnorm

## Loading required package: survival

## Loading required package: TH.data

## Loading required package: MASS

##   
## Attaching package: 'MASS'

## The following object is masked from 'package:dplyr':  
##   
## select

##   
## Attaching package: 'TH.data'

## The following object is masked from 'package:MASS':  
##   
## geyser

# Tukey HSD test:  
post\_test <- glht(res\_aov,  
 linfct = mcp(species = "Tukey")  
)  
summary(post\_test)

##   
## Simultaneous Tests for General Linear Hypotheses  
##   
## Multiple Comparisons of Means: Tukey Contrasts  
##   
##   
## Fit: aov(formula = bill\_length\_mm ~ species, data = pdata)  
##   
## Linear Hypotheses:  
## Estimate Std. Error t value Pr(>|t|)   
## Chinstrap - Adelie == 0 10.0424 0.4323 23.232 < 0.001 \*\*\*  
## Gentoo - Adelie == 0 8.7135 0.3595 24.237 < 0.001 \*\*\*  
## Gentoo - Chinstrap == 0 -1.3289 0.4473 -2.971 0.00883 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
## (Adjusted p values reported -- single-step method)

#by adjusted pvalues we can see that groups have same mean or not  
  
par(mar = c(3, 8, 3, 3))  
plot(post\_test)

