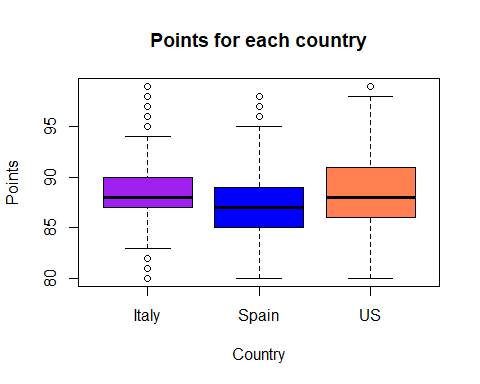
L4

Sokolova V

wine<- read.csv(file = "C:\\Users\\User\\Desktop\\3курс\\mmoi\\4\\winemag-data-130k-v2.csv", header = TRUE)  
countr <- subset(wine, (country == "US" | country == "Italy" | country == "Spain") & points<100 , select = c(country,points))  
countr$country <- factor(countr$country)

#значення по кожній з градацій   
boxplot(points ~ country, data=countr,   
 xlab = "Country" , ylab = "Points",  
 main = "Points for each country", col = c("purple", "blue", "coral"))



#значення середніх   
aggregate(x = countr$points, by = list(countr$country), FUN = mean)

## Group.1 x  
## 1 Italy 88.55989  
## 2 Spain 87.28834  
## 3 US 88.56288

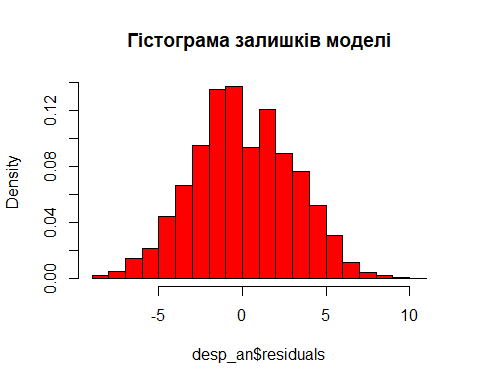
#Провести дисперсійний аналіз для цих даних  
desp\_an<-aov(points ~ country, data = countr)  
lm\_an <- lm(points ~ country, data = countr)  
summary(desp\_an)

## Df Sum Sq Mean Sq F value Pr(>F)   
## country 2 9893 4947 548.6 <2e-16 \*\*\*  
## Residuals 80678 727409 9   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

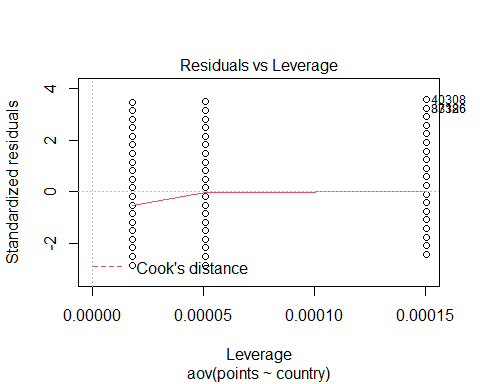
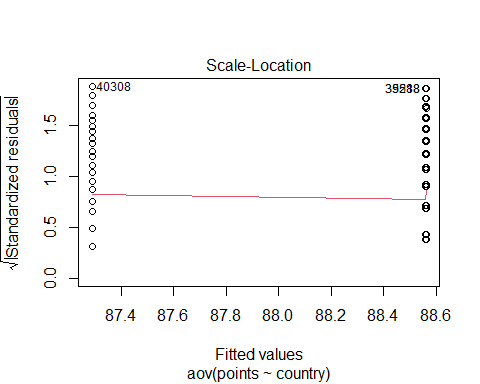
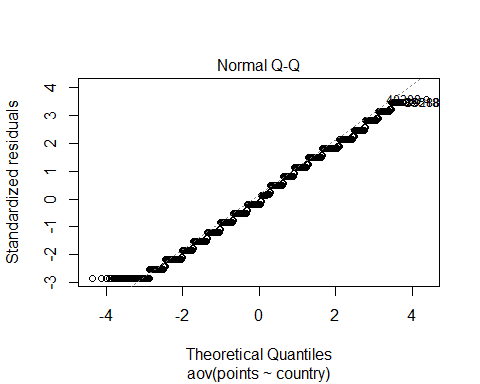
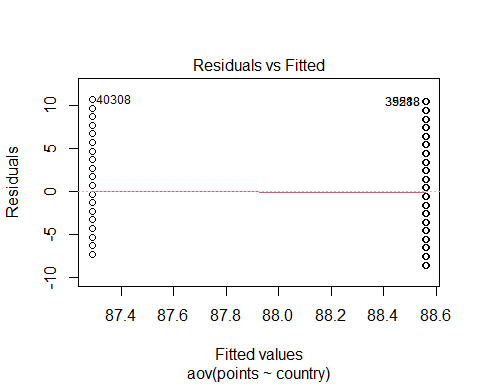
summary(lm\_an)

##   
## Call:  
## lm(formula = points ~ country, data = countr)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8.5629 -1.5629 -0.5599 2.4371 10.7117   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 88.559889 0.021483 4122.333 <2e-16 \*\*\*  
## countrySpain -1.271552 0.042642 -29.819 <2e-16 \*\*\*  
## countryUS 0.002991 0.025039 0.119 0.905   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.003 on 80678 degrees of freedom  
## Multiple R-squared: 0.01342, Adjusted R-squared: 0.01339   
## F-statistic: 548.6 on 2 and 80678 DF, p-value: < 2.2e-16

hist(desp\_an$residuals, main = "Гістограма залишків моделі", freq = F, col = "red")



plot(desp\_an)



kruskal.test(points ~ country, data = countr)

##   
## Kruskal-Wallis rank sum test  
##   
## data: points by country  
## Kruskal-Wallis chi-squared = 997.05, df = 2, p-value < 2.2e-16

bartlett.test(points ~ country, data = countr)

##   
## Bartlett test of homogeneity of variances  
##   
## data: points by country  
## Bartlett's K-squared = 731.05, df = 2, p-value < 2.2e-16

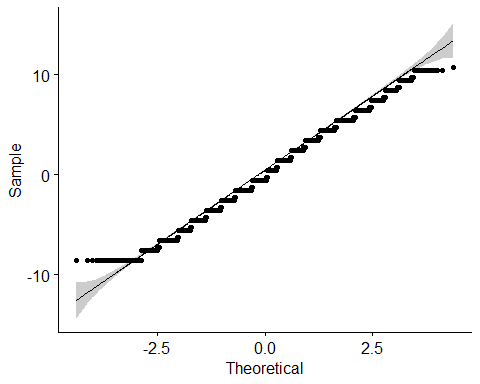
#Провести аналіз контрастів  
contrasts(countr$country)

## Spain US  
## Italy 0 0  
## Spain 1 0  
## US 0 1

library(ggplot2)  
library(ggpubr)

## Warning: package 'ggpubr' was built under R version 4.1.2

ggqqplot(desp\_an$residuals)



shapiro.test(desp\_an$residuals[0:5000])

##   
## Shapiro-Wilk normality test  
##   
## data: desp\_an$residuals[0:5000]  
## W = 0.98737, p-value < 2.2e-16