quiz2stat631

Jessica Grover

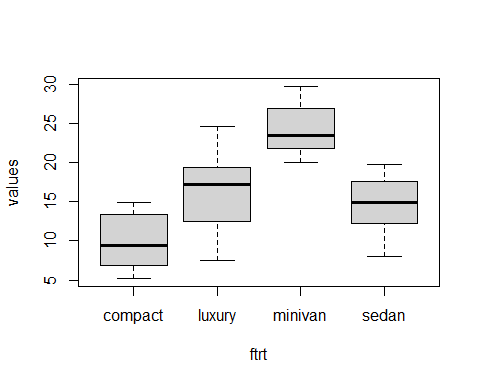
3/24/2022

#a.Construct the side-by-side boxplot to compare lengths of time among different types of cars and write down your observations.

d1<-read.csv("CarRental.csv",header = TRUE)  
library(emmeans)  
head(d1)

## X values ind  
## 1 1 8.28 compact  
## 2 2 14.84 compact  
## 3 3 5.24 compact  
## 4 4 6.18 compact  
## 5 5 6.96 compact  
## 6 6 12.99 compact

d1$ftrt<- as.factor(d1$ind)  
boxplot(values~ftrt, data = d1)

 The box plot above confirms that the time taken by the compact is lowest and the time taken by the minivan is the largest.

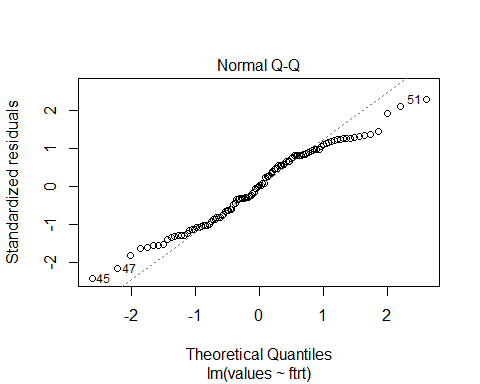
#b. Is there any statistical evidence that different types of cars are rented for varying lengths of time? Perform the corresponding test and report your conclusions.

model1<-lm(values~ftrt,data=d1)  
anova(model1)

## Analysis of Variance Table  
##   
## Response: values  
## Df Sum Sq Mean Sq F value Pr(>F)   
## ftrt 3 2295.1 765.04 54.482 < 2.2e-16 \*\*\*  
## Residuals 108 1516.5 14.04   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The p-value is less than 0.05, so we are rejecting the null hypothesis. We are 95% confident that there is atleast one difference in the rental car’s length of time.

plot(model1,which=2)

 The above plot concludes that assumptions seems to be reasonable.

#C.Apply Tukey’s simultaneous confidence interval method with α = 0.05 to find the type of cars that has the shortest rental time and the type of cars that has the largest rental time.

library(emmeans)  
lsm1<- emmeans(model1, ~ftrt)  
tk<- summary(contrast(lsm1,method="pairwise",adjust="tukey"),infer=c(T,T),level=0.95,side="two-sided")  
tk

## contrast estimate SE df lower.CL upper.CL t.ratio p.value  
## compact - luxury -6.28 0.955 108 -8.778 -3.79 -6.578 <.0001  
## compact - minivan -14.52 1.164 108 -17.556 -11.48 -12.478 <.0001  
## compact - sedan -4.68 0.881 108 -6.983 -2.38 -5.315 <.0001  
## luxury - minivan -8.23 1.251 108 -11.499 -4.97 -6.583 <.0001  
## luxury - sedan 1.60 0.994 108 -0.991 4.19 1.612 0.3764  
## minivan - sedan 9.84 1.195 108 6.717 12.95 8.229 <.0001  
##   
## Confidence level used: 0.95   
## Conf-level adjustment: tukey method for comparing a family of 4 estimates   
## P value adjustment: tukey method for comparing a family of 4 estimates

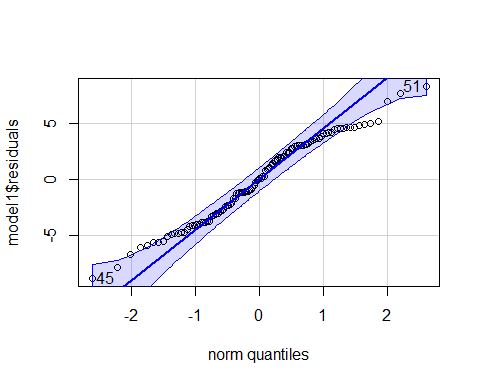
It is seen that luxury cars has shortest rental time. Minivan has the highest rental time because zero not included and it means that population parameters are highly significant

#d. Check the assumptions of your model. Are the conclusions from your model valid?

library(car)

## Loading required package: carData

#summary(mod)  
qqPlot(model1$residuals)

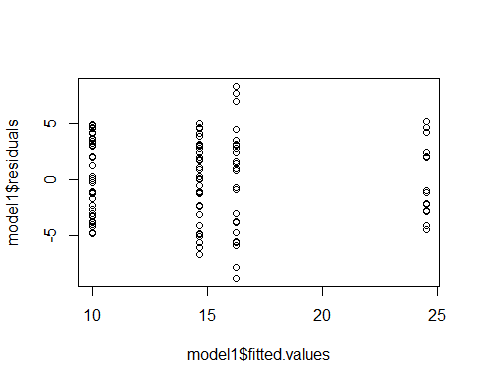


## [1] 45 51

shapiro.test(model1$residuals)

##   
## Shapiro-Wilk normality test  
##   
## data: model1$residuals  
## W = 0.97465, p-value = 0.0314

plot(model1$fitted.values,model1$residuals)

 There is constant variance so it does not violates the assumption of variance.