MTH107-Rcheatsheet.R

dogle

Thu Dec 15 11:44:59 2016

par(mar=c(3,3,0.5,0.5),mgp=c(1.9,0.6,0),tcl=-0.2)  
  
library(NCStats)

## Loading required package: FSA

## ## FSA v0.8.11. See citation('FSA') if used in publication.  
## ## Run fishR() for related website and fishR('IFAR') for related book.

## ## NCStats v0.4.5 by Derek H. Ogle, Northland College.

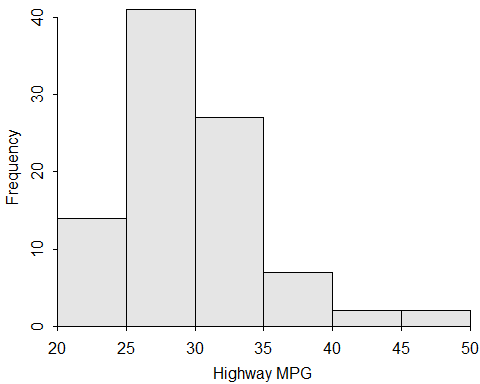
setwd("C:/aaaWork/Web/GitHub/NCMTH107")  
dfobj <- read.csv("https://raw.githubusercontent.com/droglenc/NCData/master/93cars.csv")  
str(dfobj)

## 'data.frame': 93 obs. of 26 variables:  
## $ MFG : Factor w/ 32 levels "Acura","Audi",..: 1 1 2 2 3 4 4 4 4 5 ...  
## $ Model : Factor w/ 93 levels "100","190E","240",..: 49 54 9 1 6 24 56 73 72 35 ...  
## $ Type : Factor w/ 6 levels "Compact","Large",..: 4 3 1 3 3 3 2 2 3 2 ...  
## $ MinPrice : num 12.9 29.2 25.9 30.8 23.7 14.2 19.9 22.6 26.3 33 ...  
## $ MidPrice : num 15.9 33.9 29.1 37.7 30 15.7 20.8 23.7 26.3 34.7 ...  
## $ MaxPrice : num 18.8 38.7 32.3 44.6 36.2 17.3 21.7 24.9 26.3 36.3 ...  
## $ CMPG : int 25 18 20 19 22 22 19 16 19 16 ...  
## $ HMPG : int 31 25 26 26 30 31 28 25 27 25 ...  
## $ AirBags : int 0 2 1 2 1 1 1 1 1 1 ...  
## $ DriveTrain: int 1 1 1 1 0 1 1 0 1 1 ...  
## $ Cyls : int 4 6 6 6 4 4 6 6 6 8 ...  
## $ EngSize : num 1.8 3.2 2.8 2.8 3.5 2.2 3.8 5.7 3.8 4.9 ...  
## $ HP : int 140 200 172 172 208 110 170 180 170 200 ...  
## $ RPM : int 6300 5500 5500 5500 5700 5200 4800 4000 4800 4100 ...  
## $ EngRev : int 2890 2335 2280 2535 2545 2565 1570 1320 1690 1510 ...  
## $ Manual : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 1 1 1 1 1 ...  
## $ FuelTank : num 13.2 18 16.9 21.1 21.1 16.4 18 23 18.8 18 ...  
## $ Passengers: int 5 5 5 6 4 6 6 6 5 6 ...  
## $ Length : int 177 195 180 193 186 189 200 216 198 206 ...  
## $ Wheelbase : int 102 115 102 106 109 105 111 116 108 114 ...  
## $ Width : int 68 71 67 70 69 69 74 78 73 73 ...  
## $ Uturn : int 37 38 37 37 39 41 42 45 41 43 ...  
## $ RearSeat : num 26.5 30 28 31 27 28 30.5 30.5 26.5 35 ...  
## $ Luggage : int 11 15 14 17 13 16 17 21 14 18 ...  
## $ Weight : int 2705 3560 3375 3405 3640 2880 3470 4105 3495 3620 ...  
## $ Domestic : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 2 2 2 2 2 ...

newdf1 <- filterD(dfobj,Type=="Sporty")  
newdf2 <- filterD(dfobj,HMPG>30)  
newdf3 <- filterD(dfobj,Domestic!="Yes")  
newdf4 <- filterD(dfobj,Type %in% c("Sporty","Small"))  
  
  
Summarize(~HMPG,data=dfobj,digits=1)

## n mean sd min Q1 median Q3 max   
## 93.0 29.1 5.3 20.0 26.0 28.0 31.0 50.0

hist(~HMPG,data=dfobj,xlab="Highway MPG")



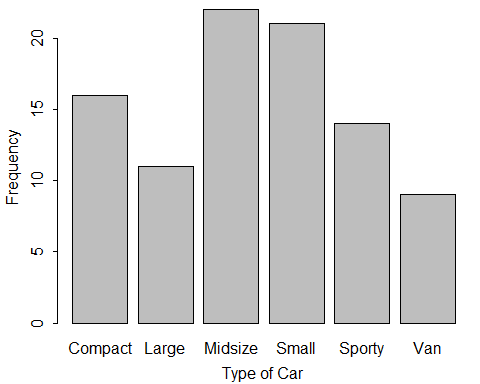
( freq1 <- xtabs(~Type,data=dfobj) )

## Type  
## Compact Large Midsize Small Sporty Van   
## 16 11 22 21 14 9

percTable(freq1,digits=1)

## Type  
## Compact Large Midsize Small Sporty Van Sum   
## 17.2 11.8 23.7 22.6 15.1 9.7 100.1

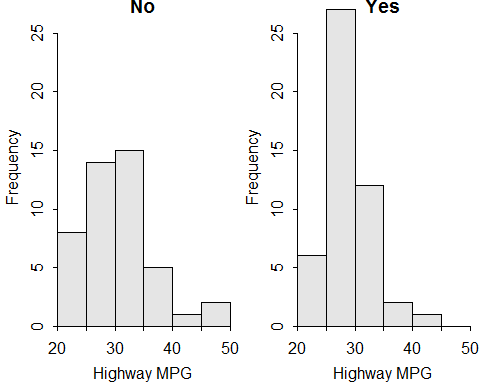
barplot(freq1,xlab="Type of Car",ylab="Frequency")



Summarize(HMPG~Domestic,data=dfobj,digits=1)

## Domestic n mean sd min Q1 median Q3 max  
## 1 No 45 30.1 6.2 21 25 30 33 50  
## 2 Yes 48 28.1 4.2 20 26 28 30 41

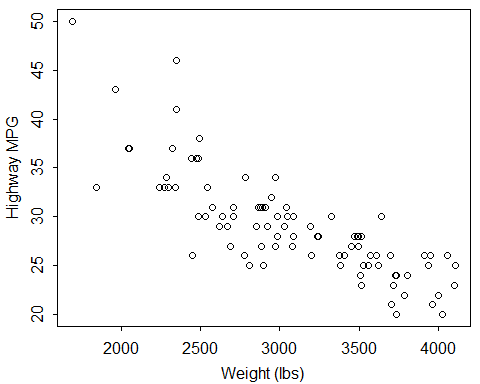
OP <- par(mfrow=c(1,2))  
hist(HMPG~Domestic,data=dfobj,xlab="Highway MPG")



par(OP)  
  
corr(HMPG~Weight,data=dfobj)

## [1] -0.8106581

plot(HMPG~Weight,data=dfobj,xlab="Weight (lbs)",ylab="Highway MPG")



( freq2 <- xtabs(~Domestic+Manual,data=dfobj) )

## Manual  
## Domestic No Yes  
## No 6 39  
## Yes 26 22

percTable(freq2,digits=1)

## Manual  
## Domestic No Yes Sum  
## No 6.5 41.9 48.4  
## Yes 28.0 23.7 51.7  
## Sum 34.5 65.6 100.1

percTable(freq2,margin=1,digits=1)

## Manual  
## Domestic No Yes Sum  
## No 13.3 86.7 100.0  
## Yes 54.2 45.8 100.0

percTable(freq2,margin=2,digits=1)

## Manual  
## Domestic No Yes  
## No 18.8 63.9  
## Yes 81.2 36.1  
## Sum 100.0 100.0

( bfl <- lm(HMPG~Weight,data=dfobj) )

## Coefficients:  
## (Intercept) Weight   
## 51.601365 -0.007327

fitPlot(bfl,xlab="Weight (lbs)",ylab="Highway MPG")  
rSquared(bfl)

## [1] 0.6571665

predict(bfl,data.frame(Weight=3000))

## 1   
## 29.62019

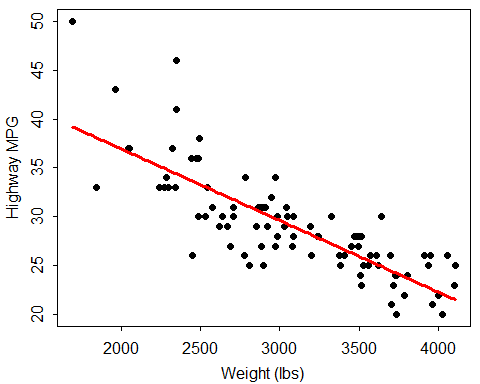
z.test(dfobj$HMPG,mu=26,alt="greater",conf.level=0.95,sd=6)

## One Sample z-test with dfobj$HMPG   
## z = 4.9601, n = 93.000, Std. Dev. = 6.000, Std. Dev. of the sample  
## mean = 0.622, p-value = 3.523e-07  
## alternative hypothesis: true mean is greater than 26   
## 95 percent confidence interval:  
## 28.06264 Inf   
## sample estimates:  
## mean of dfobj$HMPG   
## 29.08602

t.test(dfobj$HMPG,mu=26,alt="two.sided",conf.level=0.99)

## One Sample t-test with dfobj$HMPG   
## t = 5.5818, df = 92, p-value = 2.387e-07  
## alternative hypothesis: true mean is not equal to 26   
## 99 percent confidence interval:  
## 27.63178 30.54026   
## sample estimates:  
## mean of x   
## 29.08602

levenesTest(HMPG~Domestic,data=dfobj)



## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)   
## group 1 5.3595 0.02286 \*  
## 91   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

t.test(HMPG~Manual,data=dfobj,alt="less",conf.level=0.99,var.equal=TRUE)

## Two Sample t-test with HMPG by Manual   
## t = -4.2183, df = 91, p-value = 2.904e-05  
## alternative hypothesis: true difference in means is less than 0   
## 99 percent confidence interval:  
## -Inf -1.980103   
## sample estimates:  
## mean in group No mean in group Yes   
## 26.12500 30.63934

exp <- c(1,1,1,1,1,1)/6  
( gof <- chisq.test(freq1,p=exp,rescale.p=TRUE,correct=FALSE) )

## Chi-squared test for given probabilities with freq1   
## X-squared = 8.871, df = 5, p-value = 0.1143

gof$expected

## Compact Large Midsize Small Sporty Van   
## 15.5 15.5 15.5 15.5 15.5 15.5

gof$residuals

## Type  
## Compact Large Midsize Small Sporty Van   
## 0.1270001 -1.1430011 1.6510017 1.3970014 -0.3810004 -1.6510017

gofCI(gof,digits=3)

## p.obs p.LCI p.UCI p.exp  
## Compact 0.172 0.109 0.261 0.167  
## Large 0.118 0.067 0.199 0.167  
## Midsize 0.237 0.162 0.332 0.167  
## Small 0.226 0.153 0.321 0.167  
## Sporty 0.151 0.092 0.237 0.167  
## Van 0.097 0.052 0.174 0.167

( chi <- chisq.test(freq2,correct=FALSE) )

## Pearson's Chi-squared test with freq2   
## X-squared = 17.1588, df = 1, p-value = 3.438e-05

chi$expected

## Manual  
## Domestic No Yes  
## No 15.48387 29.51613  
## Yes 16.51613 31.48387

chi$residuals

## Manual  
## Domestic No Yes  
## No -2.410160 1.745645  
## Yes 2.333627 -1.690214