Assignment

nabiha faisal

1/19/2022

data(rock)   
head(rock)

## area peri shape perm  
## 1 4990 2791.90 0.0903296 6.3  
## 2 7002 3892.60 0.1486220 6.3  
## 3 7558 3930.66 0.1833120 6.3  
## 4 7352 3869.32 0.1170630 6.3  
## 5 7943 3948.54 0.1224170 17.1  
## 6 7979 4010.15 0.1670450 17.1

dim(rock)

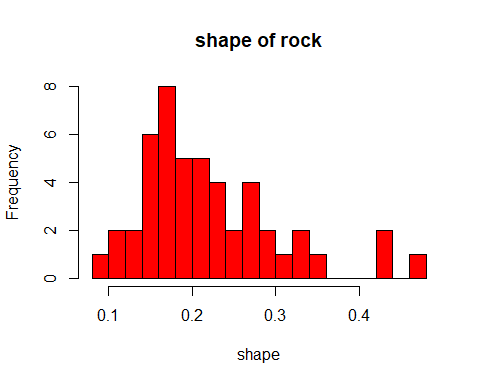
## [1] 48 4

#there are 48row and 4 col in this data  
  
print(summary(rock))

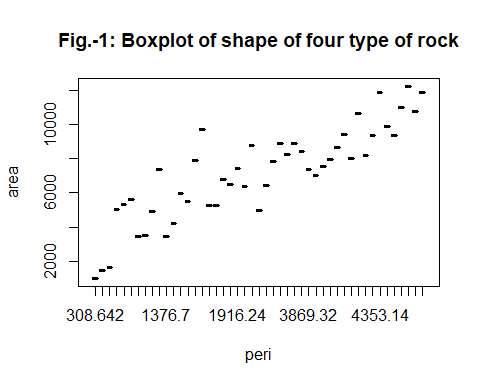
## area peri shape perm   
## Min. : 1016 Min. : 308.6 Min. :0.09033 Min. : 6.30   
## 1st Qu.: 5305 1st Qu.:1414.9 1st Qu.:0.16226 1st Qu.: 76.45   
## Median : 7487 Median :2536.2 Median :0.19886 Median : 130.50   
## Mean : 7188 Mean :2682.2 Mean :0.21811 Mean : 415.45   
## 3rd Qu.: 8870 3rd Qu.:3989.5 3rd Qu.:0.26267 3rd Qu.: 777.50   
## Max. :12212 Max. :4864.2 Max. :0.46413 Max. :1300.00

## 2. Plots

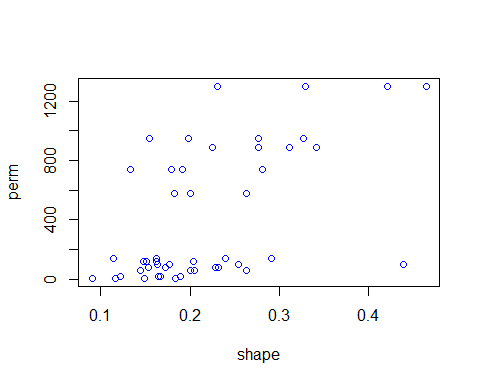
library(ggplot2)  
attach(rock)  
  
  
hist(shape,breaks= 15, col = 'red',  
 xlab= 'shape',   
 main = 'shape of rock')



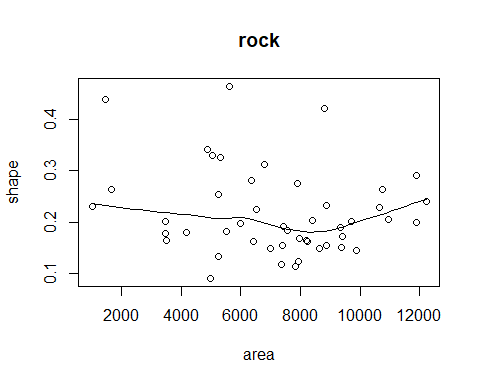
#graph shows the max observation lies between 0.1 and 0.2 and it is roughly bell shaped  
boxplot(area ~ peri, main="Fig.-1: Boxplot of shape of four type of rock", col= rainbow(3))



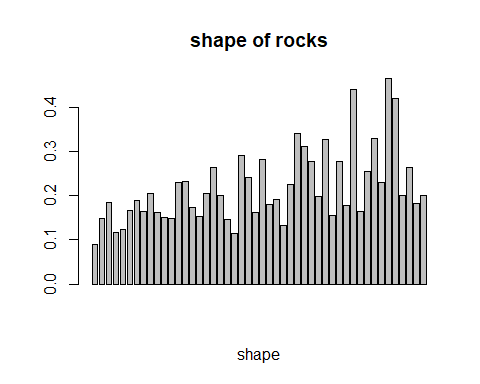
#graph shows as area in creases the perimeter in pixels also increase  
plot(shape,perm, col = "Blue")



scatter.smooth(x=area,y=shape,main="rock")



barplot(shape, main="shape of rocks",  
 xlab="shape")



#the biggest rock shape is greater than 0.4

## 3. Correlation

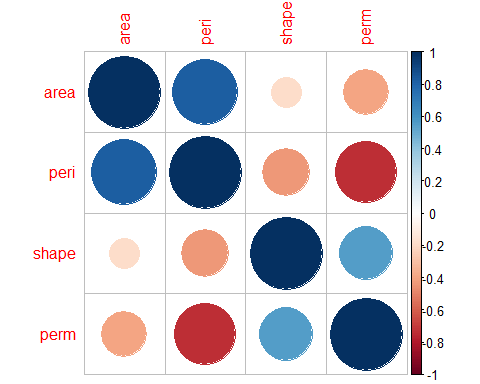
cor(rock)

## area peri shape perm  
## area 1.0000000 0.8225064 -0.1821611 -0.3966370  
## peri 0.8225064 1.0000000 -0.4331255 -0.7387158  
## shape -0.1821611 -0.4331255 1.0000000 0.5567208  
## perm -0.3966370 -0.7387158 0.5567208 1.0000000

#area is closely related to peri(perimeter in pixels of the rock)  
#shape is closely related to perm of the rock  
  
  
# -----------------------------------------------------------------------------  
library(corrplot)

## corrplot 0.92 loaded

cor.mat.rock = cor(rock)  
corrplot(cor.mat.rock)



## 3. Confidence Interval

library(Rmisc)

## Loading required package: lattice

## Loading required package: plyr

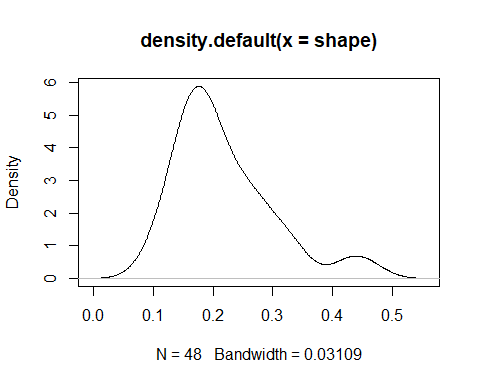
attach(rock)

## The following objects are masked from rock (pos = 7):  
##   
## area, peri, perm, shape

CI(shape, ci = 0.95) # Calculates the confidence interval of a vector of data.

## upper mean lower   
## 0.2423553 0.2181104 0.1938656

#mean lies between the confidence interval we accept the null hypothesis  
  
plot(density(shape))



shapiro.test(shape)

##   
## Shapiro-Wilk normality test  
##   
## data: shape  
## W = 0.90407, p-value = 0.0008531

# p value <0.05 so the data is not normall distributed

# 4. Hypothesis testing

# ——————-

library(stats)  
  
# One sample t-test  
# -------------------  
t.test(rock$shape, mu = 0.21) # to test: Is the mean value of mpg differ from 20 or not?

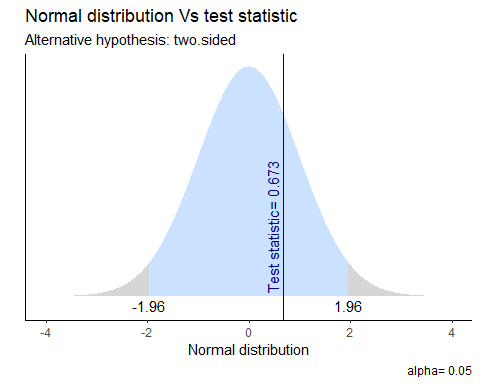
##   
## One Sample t-test  
##   
## data: rock$shape  
## t = 0.67297, df = 47, p-value = 0.5043  
## alternative hypothesis: true mean is not equal to 0.21  
## 95 percent confidence interval:  
## 0.1938656 0.2423553  
## sample estimates:  
## mean of x   
## 0.2181104

#alternative hypoyhesis not rejected  
t.test(rock$shape, mu =0.21 , alternative = 'greater') # if you want to test on sided alternative.

##   
## One Sample t-test  
##   
## data: rock$shape  
## t = 0.67297, df = 47, p-value = 0.2521  
## alternative hypothesis: true mean is greater than 0.21  
## 95 percent confidence interval:  
## 0.1978886 Inf  
## sample estimates:  
## mean of x   
## 0.2181104

library(gginference)  
ggttest(t.test(rock$shape, mu = 0.21))

## Warning: geom\_vline(): Ignoring `data` because `xintercept` was provided.



#   
# ----------------------------------------------

**5. Chi-square test**

attach(rock)

## The following objects are masked from rock (pos = 4):  
##   
## area, peri, perm, shape

## The following objects are masked from rock (pos = 9):  
##   
## area, peri, perm, shape

chisq.test(perm,shape)

## Warning in chisq.test(perm, shape): Chi-squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: perm and shape  
## X-squared = 516, df = 506, p-value = 0.3695

#there is no association between the shape and permiablity of rock  
chisq.test(area,shape)

## Warning in chisq.test(area, shape): Chi-squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: area and shape  
## X-squared = 2208, df = 2116, p-value = 0.08009

#there is slight association between the area and shape  
# --------------------

**6. Analysis of Variance**

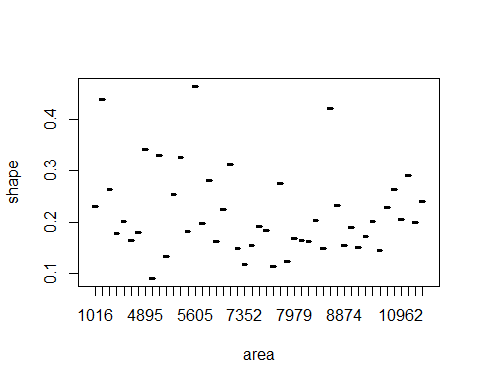
attach(rock)

## The following objects are masked from rock (pos = 3):  
##   
## area, peri, perm, shape

## The following objects are masked from rock (pos = 5):  
##   
## area, peri, perm, shape

## The following objects are masked from rock (pos = 10):  
##   
## area, peri, perm, shape

# ONE WAY ANOVA   
  
boxplot(shape ~area, col= rainbow(5))



model1 <- aov(shape ~area)  
summary(model1)

## Df Sum Sq Mean Sq F value Pr(>F)  
## area 1 0.0109 0.010873 1.579 0.215  
## Residuals 46 0.3168 0.006887

#F value is greater than P value so we will reject the null hypothesisand thus the shape per area is not equal.  
#null hypothesis=shape per area is equal  
#alternative hypothesis= shape per area not equal.

# 7.Linear and Multiple Regression Models:

# Linear Regression line formula:  
# -------------------------------  
attach(rock)

## The following objects are masked from rock (pos = 3):  
##   
## area, peri, perm, shape

## The following objects are masked from rock (pos = 4):  
##   
## area, peri, perm, shape

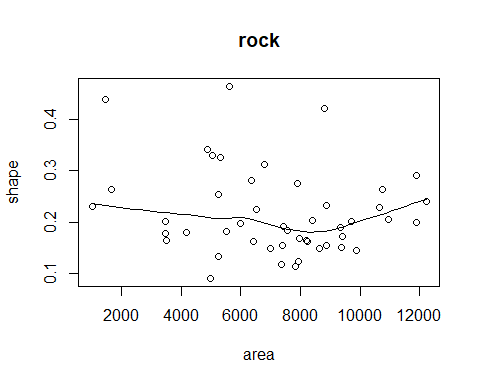
## The following objects are masked from rock (pos = 6):  
##   
## area, peri, perm, shape

## The following objects are masked from rock (pos = 11):  
##   
## area, peri, perm, shape

fit.LR <- lm(area ~ shape, data = rock)  
summary(fit.LR)

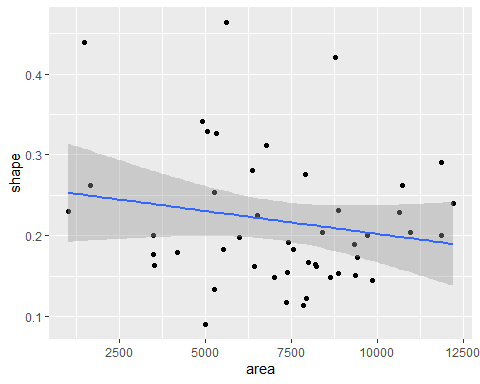
##   
## Call:  
## lm(formula = area ~ shape, data = rock)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6101.6 -1512.3 104.6 1765.3 5152.9   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8465 1087 7.788 6.08e-10 \*\*\*  
## shape -5855 4660 -1.256 0.215   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2667 on 46 degrees of freedom  
## Multiple R-squared: 0.03318, Adjusted R-squared: 0.01216   
## F-statistic: 1.579 on 1 and 46 DF, p-value: 0.2153

#equation:  
#area=8465+(-5855)shape  
#comments:  
#intercept is significant for this segregation  
#regulation model in not good overall   
#corroletion between area and shape is 18%  
#p value shows that with decrease in shape will result in increase in area  
  
scatter.smooth(x=area,y=shape,main="rock")



require(ggplot2)  
ggplot(rock,aes(y=shape,x=area))+geom\_point()+geom\_smooth(method = "lm")

## `geom\_smooth()` using formula 'y ~ x'



#plot shows us the linearity and independence of observation   
  
  
  
  
  
# Multiple Regression line formula:  
# -------------------------------  
  
  
fit.MR <- lm(area ~ shape + peri + perm, data = rock)  
summary(fit.MR)

##   
## Call:  
## lm(formula = area ~ shape + peri + perm, data = rock)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3256.3 -781.9 -121.3 652.8 4788.3   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -407.0690 919.8868 -0.443 0.660280   
## shape 2992.3143 2735.3722 1.094 0.279937   
## peri 2.1934 0.1966 11.156 2.06e-14 \*\*\*  
## perm 2.5492 0.6976 3.654 0.000684 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1300 on 44 degrees of freedom  
## Multiple R-squared: 0.7805, Adjusted R-squared: 0.7655   
## F-statistic: 52.14 on 3 and 44 DF, p-value: 1.561e-14

#equation:  
#area=-407.06+2992shape+2.19peri+2.54perm  
#intercept is not significant for this segregation  
#change in shape is not significant with change in area whereas change in perimeter and permiability are highly significant with change in area  
#corroletion between area and these three variables is 88%  
  
   
  
#t test for intercept   
#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
  
#t\_{b\_{0}}=\frac{b\_{0}-\beta\_{0}}{s\_{b\_{0}}}  
  
#t test for slope  
#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
  
#t\_{b\_{1}}=\frac{b\_{1}-\beta\_{1}}{s\_{b\_{1}}}   
   
   
  
  
# ---------------------------------------  
anova(fit.LR)

## Analysis of Variance Table  
##   
## Response: area  
## Df Sum Sq Mean Sq F value Pr(>F)  
## shape 1 11233766 11233766 1.5788 0.2153  
## Residuals 46 327309336 7115420

#f value is grater than p value so we will reject null hypothesis  
# The F-statistic value tells overall result is not good.