Assignment on R

Shukrah Azeez

3/08/2022

# Dataset:Rock

## Statistical analysis

### Regression

data(rock)  
attach(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

regression<-lm(peri~area)  
summary(regression)

##   
## Call:  
## lm(formula = peri ~ area)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2306.8 -502.3 122.5 564.5 1291.9   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -471.43579 342.77487 -1.375 0.176   
## area 0.43875 0.04473 9.808 7.51e-13 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 823.1 on 46 degrees of freedom  
## Multiple R-squared: 0.6765, Adjusted R-squared: 0.6695   
## F-statistic: 96.2 on 1 and 46 DF, p-value: 7.506e-13

#### Report

##### Hypothesis

* H0: There is no significant difference between peri and area.
* H1: There is significant difference between peri and area.

##### Decision Rule

* p-value < 0.05, reject H0, otherwise, fail to reject H0.

##### Conclusion

* Since p-value(7.506e-13) < 0.05, reject H0. Therefore, There is significant difference between peri and area.
* The linear regression model for the analysis is; Y = -471.43579 + 0.43875X

### Correlation

correlation <- cor.test(peri,perm)  
correlation

##   
## Pearson's product-moment correlation  
##   
## data: peri and perm  
## t = -7.4334, df = 46, p-value = 2.048e-09  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.8454045 -0.5753425  
## sample estimates:  
## cor   
## -0.7387158

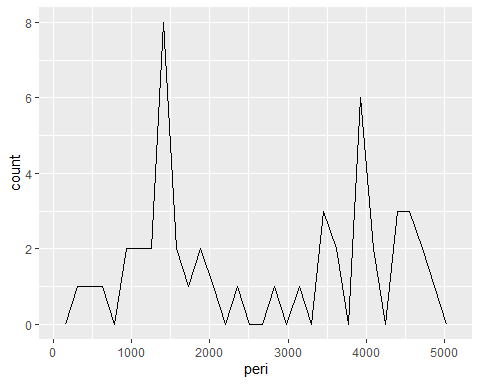
#### Report

* The analysis has a very strong negative correlation of -0.7387158.

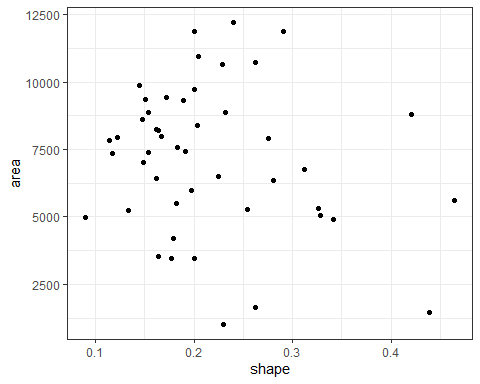
## Graphical visualisation

library(ggplot2)  
ggplot(data=rock,aes(x=peri))+  
geom\_freqpoly()

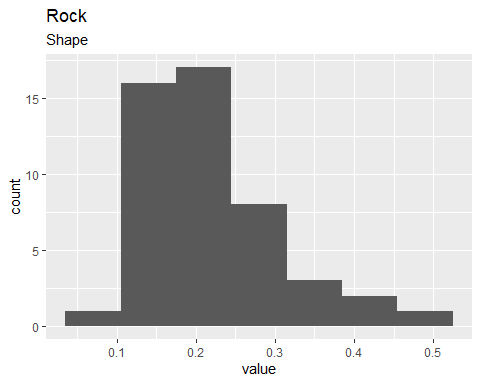
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



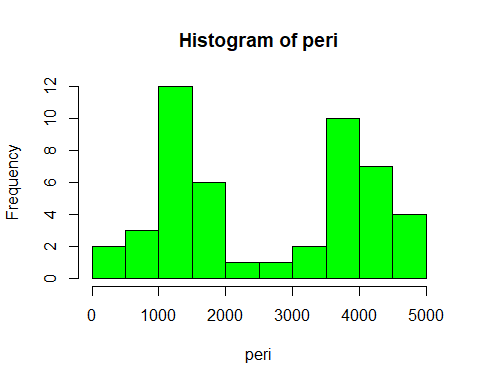
ggplot(data=rock,aes(x=shape,y=area))+  
geom\_point()+  
theme\_bw()



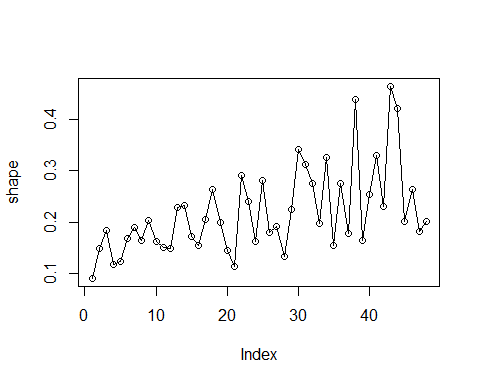
ggplot(data=rock,aes(x=shape))+  
geom\_histogram(binwidth=0.07)+  
theme\_grey()+  
labs(title="Rock", subtitle="Shape", x="value")



hist(peri,col="green")



plot(shape)  
lines(shape)



## Exploratory Data Analysis

### Mean

mean(perm)

## [1] 415.45

mean(peri)

## [1] 2682.212

mean(area)

## [1] 7187.729

mean(shape)

## [1] 0.2181104

### Variance

var(shape)

## [1] 0.006971657

var(area)

## [1] 7203045

var(perm)

## [1] 191684.8

var(peri)

## [1] 2049654

### Standard deviation

sd(area)

## [1] 2683.849

sd(peri)

## [1] 1431.661

sd(area)

## [1] 2683.849

sd(perm)

## [1] 437.8182