

Test 1-Take Home

Jarod Wright

9/27/2021

4

```
library(stats)
library(dplyr)
library(ggplot2)
data(Orange)
attach(Orange)

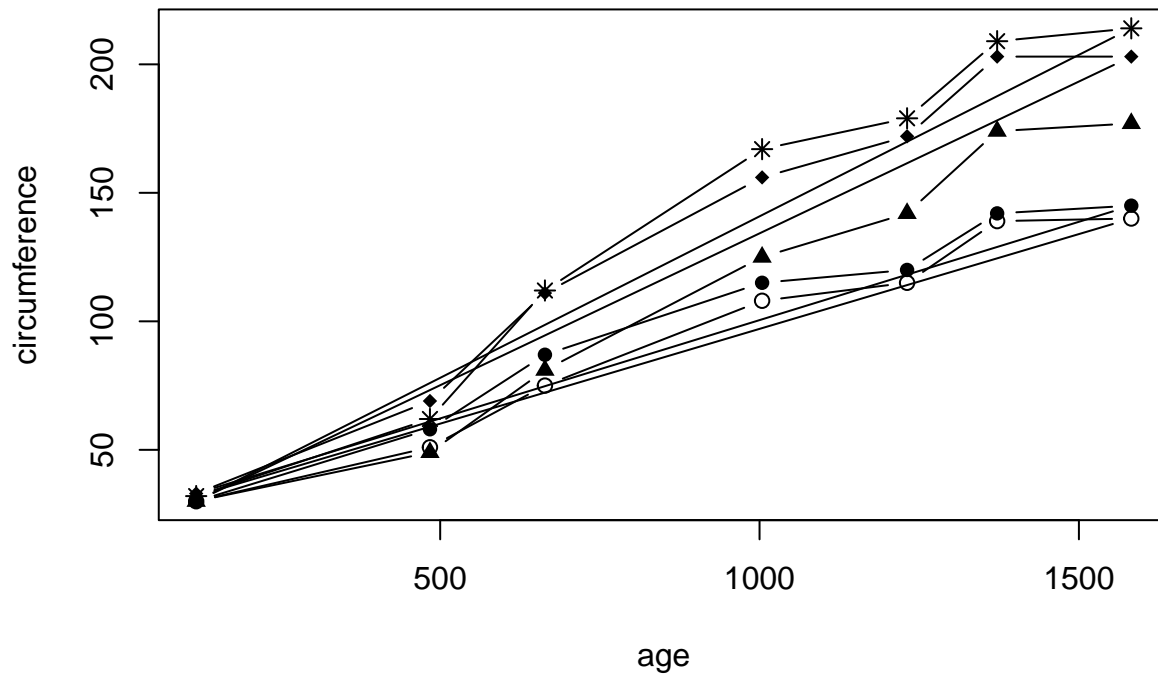
set.seed(123)

dim(Orange)

## [1] 35  3
Orange$Tree = as.factor(Orange$Tree)
unique(Orange$Tree) # 5 tree types

## [1] 1 2 3 4 5
## Levels: 3 < 1 < 5 < 2 < 4
plot(circumference~age,type = "b",pch = c(1,16,17,18,8)[Orange$Tree], main = "Circumference with Age")
```

Circumference with Age



```
cor.test(age,circumference)
```

```
##
##  Pearson's product-moment correlation
##
## data:  age and circumference
## t = 12.9, df = 33, p-value = 1.931e-14
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.8342364 0.9557955
## sample estimates:
##      cor
## 0.9135189
```

#correlation is significant

#4) There is a linear relationship between circumference and age.

5

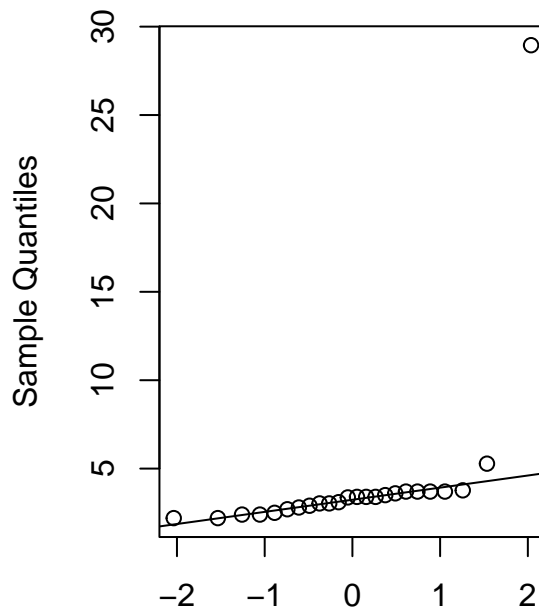
```
library(MASS)
data(chem)
View(chem)

par(mfrow=c(1,2))
qqnorm(chem)
qqline(chem)
```

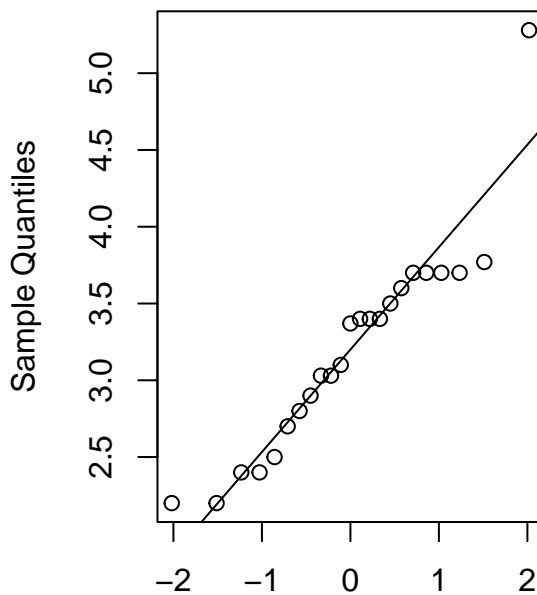
```
shapiro.test(chem)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data: chem  
## W = 0.31147, p-value = 1.283e-09  
  
chem <- chem[-17]  
qqnorm(chem)  
qqline(chem)
```

Normal Q-Q Plot



Normal Q-Q Plot



```
shapiro.test(chem)
```

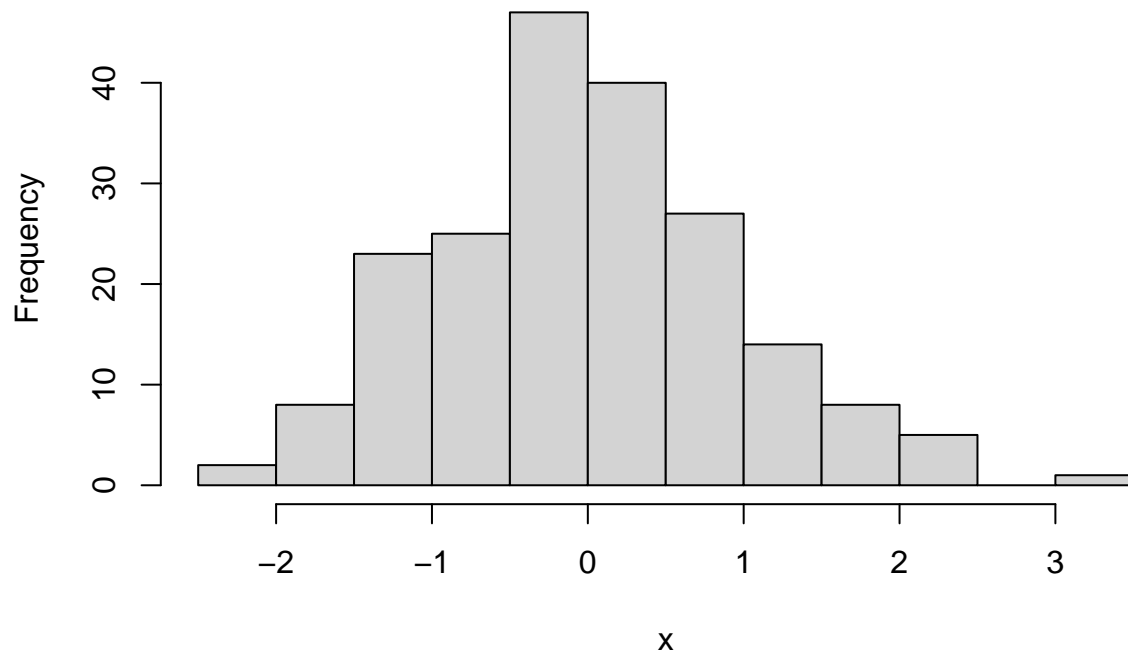
```
##  
## Shapiro-Wilk normality test  
##  
## data: chem  
## W = 0.9041, p-value = 0.03074
```

Null Hyp. is the sample came from normal dist. The Shapiro-Wilk test suggest chem does not come from a normal distribution. Assuming significance of 0.05, both p-values < 0.05 suggest evidence of non-normality. Both QQ-plots do not strongly suggest non-normality as they look like straight lines.

6

```
X = matrix(nrow = 100, ncol = 2)  
x = rnorm(200)  
hist(x) #looks like a bell
```

Histogram of x



```
for (i in 1:100){
  x1 = sample(x,size = 100, replace = F)
  x1.m = mean(x1)
  x1.sd = sd(x1)

  X[i,] = c(x1.m,x1.sd)
}

X <- data.frame(X)
colnames(X) <- c("MEAN","SD")
X.mean <- mean(X$MEAN)
X.sd <- mean(X$SD)

f1 <- function(N,n){
  X1 = matrix(nrow = n,ncol = 2)
  d=rnorm(N)
  hist(d)

  for (i in 1:n){
    x2 = sample(x,size = n, replace = F)
    x2.m = mean(x1)
    x2.sd = sd(x1)

    X1[i,] = c(x1.m,x1.sd)
  }

  X1 <- data.frame(X1)
  colnames(X1) <- c("MEAN","SD")
  X1.mean <- mean(X1$MEAN)
  X1.sd <- mean(X1$SD)
```

```

}

#7
library(readxl)

covid.age <- read_excel("/Users/jarodw/Downloads/Public-Dataset-Age.XLSX")

covid <- read_excel("/Users/jarodw/Downloads/Public-Dataset-Daily-Case-Info.XLSX")
View(covid)

covid <- covid %>% arrange(AGE_RANGE)
covid <- covid[-1,] # remove 1960 date
View(covid)

par(mfrow=c(2,2))
plot(covid$TOTAL_CASES~covid$DATE, main="Total Cases v. Date", ylab = "Total Cases", col= "orange", xlab = "Date")

hist(covid$TOTAL_CASES, main = "Total Case Count Frequency", xlab = "Total Cases", col = "blue")

covid.age <- covid.age %>% filter(AGE_RANGE != "Pending")
young <- covid.age %>% filter(AGE_RANGE=="11-20 years" | AGE_RANGE=="0-10 years")
old <- setdiff(covid.age,young)

total = sum(covid.age$AR_NEWCASES)
total.cases.young = sum(young$AR_NEWCASES)
total.cases.old = sum(old$AR_NEWCASES)

case.count.proportion = total.cases.young/total

population.est = 6829174
population.est.young = .221*population.est
population.est.old = (1-0.221)*population.est

prop.test(x = c(total.cases.young,total.cases.old), n = c(population.est.young, population.est.old),
          alternative = "less")

##
## 2-sample test for equality of proportions with continuity correction
##
## data:  c(total.cases.young, total.cases.old) out of c(population.est.young, population.est.old)
## X-squared = 94.392, df = 1, p-value = 1
## alternative hypothesis: less
## 95 percent confidence interval:
## -1.000000000 0.004001125
## sample estimates:
##      prop 1      prop 2
## 0.1795531 0.1761336

prop.test(x = c(total.cases.young,total.cases.old), n = c(population.est.young, population.est.old),
          alternative = "greater")

##
## 2-sample test for equality of proportions with continuity correction
##

```

```
## data: c(total.cases.young, total.cases.old) out of c(population.est.young, population.est.old)
## X-squared = 94.392, df = 1, p-value < 2.2e-16
## alternative hypothesis: greater
## 95 percent confidence interval:
## 0.002837726 1.0000000000
## sample estimates:
## prop 1 prop 2
## 0.1795531 0.1761336
```

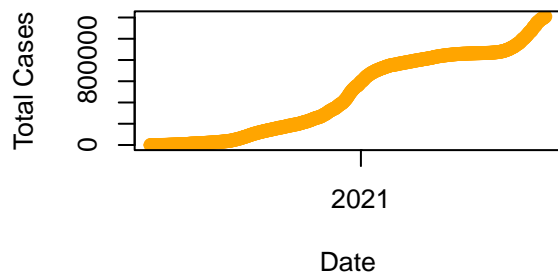
*#We can reject the Null Hypothesis that
#the proportion of cases for older people is higher.*

```
before.vacc <- covid[88:285,]
after.vacc <- covid[286:483,]
```

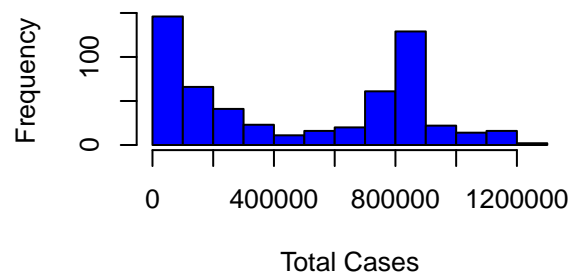
```
t.test(before.vacc$NEW_CASES,after.vacc$NEW_CASES)
```

```
##
## Welch Two Sample t-test
##
## data: before.vacc$NEW_CASES and after.vacc$NEW_CASES
## t = 1.3609, df = 381.79, p-value = 0.1744
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -126.5024 695.3003
## sample estimates:
## mean of x mean of y
## 2271.929 1987.530
```

Total Cases v. Date



Total Case Count Frequency



The true difference in means is not equal to zero. The mean number of new cases decreases after the vaccine date.

*Census data reports 22.1% of TN population is under age 18 data for those 19,20 years old will not be accounted for