LungCapData

Amira Ibrahim

October 2, 2019
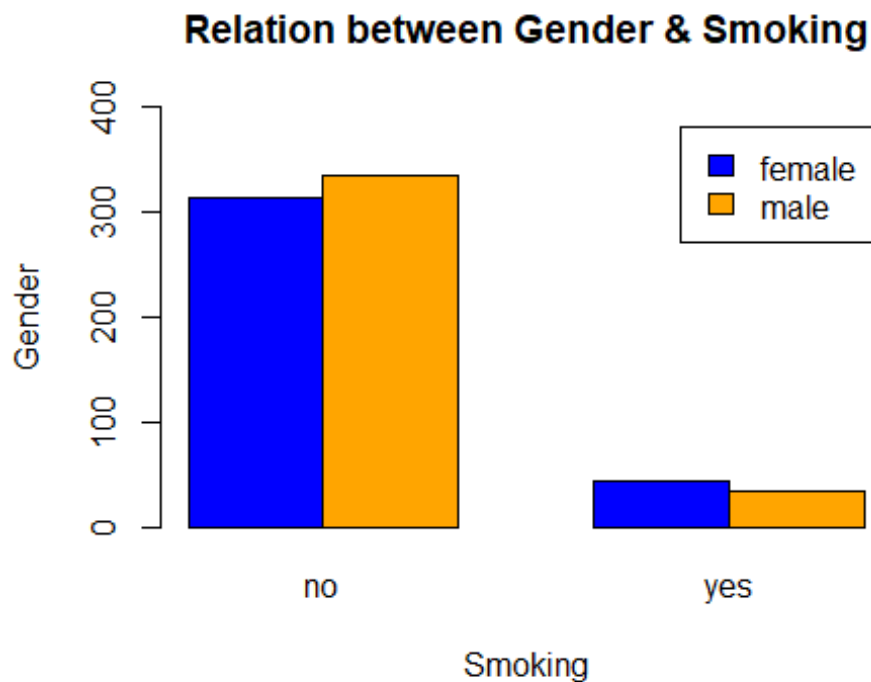
```
LungCapData <- read.delim(file.choose(),header = TRUE)
attach(LungCapData)
```

relation between Gender and Smoke :

```
Table1 <- table(Gender ,Smoke)
Table1

##         Smoke
## Gender     no yes
##    female 314  44
##    male   334  33

barplot(Table1 , beside = TRUE , legend=TRUE ,xlab = "Smoking" , ylab =
"Gender" ,
        main = "Relation between Gender & Smoking" ,ylim = c(0,400),col =
c("blue" , "orange") )
```

categorical variables by chisq test :

H0 : No relation between smoking frequency and gender

```
chisq.test(Table1 , correct = TRUE)

##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  Table1
## X-squared = 1.7443, df = 1, p-value = 0.1866
```

p-value > 0.05 , Fail to reject H0

calculate OR , RR :

```
library(epiR)

## Loading required package: survival

## Warning: package 'survival' was built under R version 3.6.1

## Package epiR 1.0-2 is loaded

## Type help(epi.about) for summary information

##

epi.2by2(Table1 , method = "cohort.count" , conf.level = 0.95)

##              Outcome +    Outcome -      Total       Inc risk *
## Exposed +         314           44        358             87.7
## Exposed -         334           33        367             91.0
## Total             648           77        725             89.4
##               Odds
## Exposed +     7.14
## Exposed -    10.12
## Total         8.42
##
## Point estimates and 95% CIs:
## -------------------------------------------------------------------
## Inc risk ratio                                0.96 (0.92, 1.01)
## Odds ratio                                    0.71 (0.44, 1.14)
## Attrib risk *                                 -3.30 (-7.79, 1.19)
## Attrib risk in population *                   -1.63 (-5.32, 2.06)
## Attrib fraction in exposed (%)                -3.76 (-9.12, 1.34)
## Attrib fraction in population (%)             -1.82 (-4.34, 0.64)
## -------------------------------------------------------------------
##  Test that odds ratio = 1: chi2(1) = 2.077 Pr>chi2 = 0.15
##  Wald confidence limits
```

```
##  CI: confidence interval
##  * Outcomes per 100 population units
```

Odds of Females not smoking are 0.71 times odds of males not smoking

```
1/0.71
```

```
## [1] 1.408451
```

Odds of males not smoking are 1.4 times odds of Females not smoking

check normality

```
library(moments)
skewness(LungCap)
```
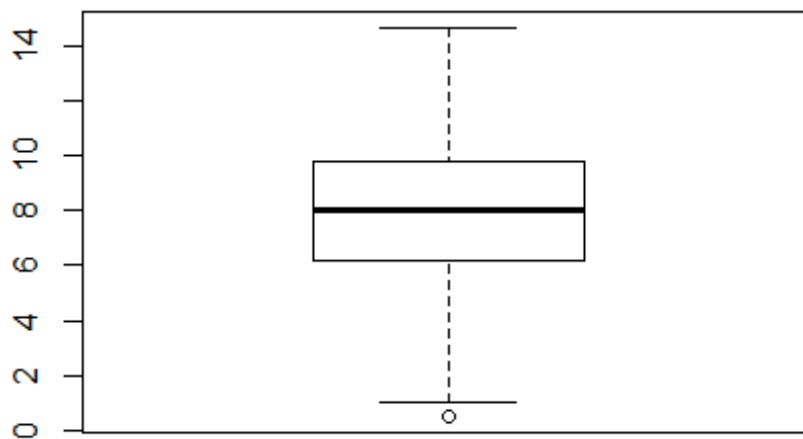
```
## [1] -0.2274017
```

accepted level from -1 to +1

```
kurtosis(LungCap)
```

```
## [1] 2.68148
```

accepted level from -2 to +2 may to +3

```
boxplot(LungCap)
```


visually ,data is normally distributed

One-sample t-test for lung Capacity :

Test H0 =8 , conf.interval = 0.95 :

```
t.test(LungCap , mu=8 , alternative = "two.sided" , conf.level = 0.95)

##
##  One Sample t-test
##
## data:  LungCap
## t = -1.3842, df = 724, p-value = 0.1667
## alternative hypothesis: true mean is not equal to 8
## 95 percent confidence interval:
##  7.669052 8.057243
## sample estimates:
## mean of x
##  7.863148
```
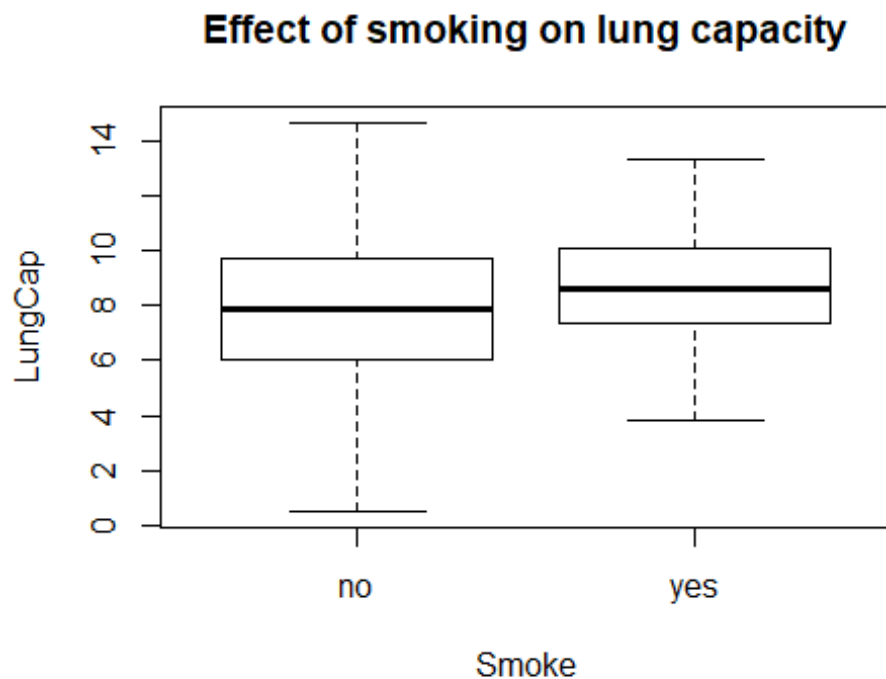
p-value >0.05 , fail to reject H0

Relation between Smoke & lung Capacity :

H0 : mean of smokers = mean of non smokers :

```
boxplot(LungCap~Smoke , main = "Effect of smoking on lung capacity")
```

check variance :

```
var(LungCap[Smoke == "yes"])

## [1] 3.545292

var(LungCap[Smoke == "no"])

## [1] 7.431694
```

so variance not equal

```
t.test(LungCap~Smoke , mu=0 , alternative = "two.sided" , var.eq = F,
conf.level =  0.95)

##
##   Welch Two Sample t-test
##
## data:  LungCap by Smoke
## t = -3.6498, df = 117.72, p-value = 0.0003927
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   -1.3501778 -0.4003548
## sample estimates:
##   mean in group no mean in group yes
##          7.770188          8.645455
```

p-value < 0.05 , reject H0 , Smoking has a significant effect on lung capacity

H0 : Median of lung capacity of smokers = Median of lung capacity of non smokers

```
wilcox.test(LungCap~Smoke , mu=0 , alternative = "two.sided" ,
            conf.int=T,  conf.level =  0.95 , paired=F , exact=F,correct=F)

##
##   Wilcoxon rank sum test
##
## data:  LungCap by Smoke
## W = 20128, p-value = 0.005533
## alternative hypothesis: true location shift is not equal to 0
## 95 percent confidence interval:
##   -1.399989 -0.249989
## sample estimates:
## difference in location
##            -0.8000564
```

p-value < 0.05 , reject H0 , Smoking has a significant effect on lung capacity