20181003a Phenotypes

cheungngo

October 3, 2018

### 

## Package and data

library(BiocInstaller)  
biocLite('COPDSexualDimorphism.data')

library(COPDSexualDimorphism.data)  
data(lgrc.expr.meta)  
# The variable pkyrs in the expr.meta data.frame represents pack years smoked  
# Other variables include gender (self-explanatory) and diagmaj (disease status)

names(expr.meta)

## [1] "tissueid" "sample\_name" "newid" "GENDER" "age"   
## [6] "cigever" "pkyrs" "diagmaj" "gender"

### 

## What is the number of female participants in this study?:

unique(expr.meta$gender)

## [1] 2-Female 1-Male   
## Levels: 1-Male 2-Female

sum(expr.meta$gender=='2-Female')

## [1] 110

### 

## What is the median of the distribution of pack years smoked in this cohort (women and men)?

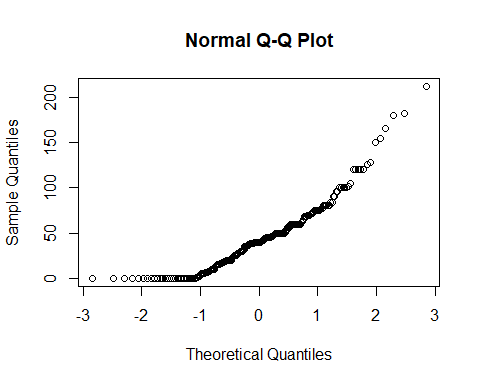
median(expr.meta$pkyrs)

## [1] 40

### 

## True or False: The distribution of pack-years smoked is well-approximated by a Gaussian (Normal) probability distribution.

qqnorm(expr.meta$pkyrs)

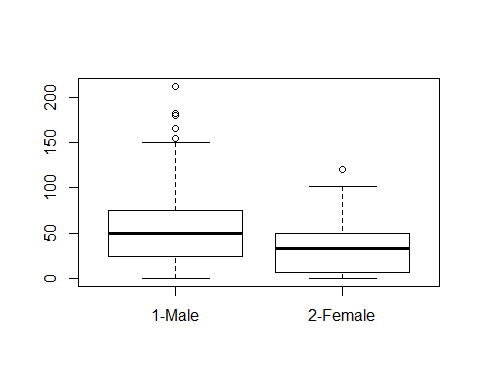


# FALSE

### 

## Which of the following is an aspect of the display that would suggest caution in using the t test in comparing males and females with respect to pack years smoked?

boxplot(pkyrs~gender, data=expr.meta)



# plot 01  
# Distributions appear quite asymmetric, with long tails skewed towards high values.

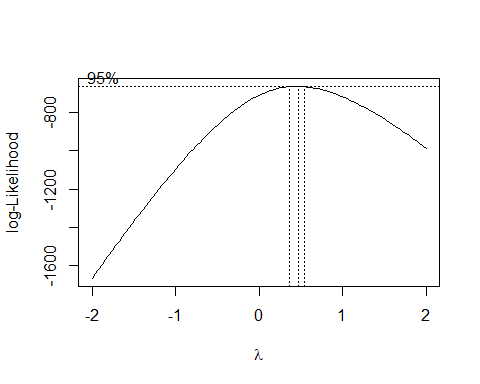
### 

## Variable transformation using boxcox

# transform the pkyrs into a positive var for analysis  
expr.meta$pyp1 = expr.meta$pkyrs+1

library(MASS)

lm1 = lm(pyp1~gender, data=expr.meta)  
bc1 = boxcox(lm1)

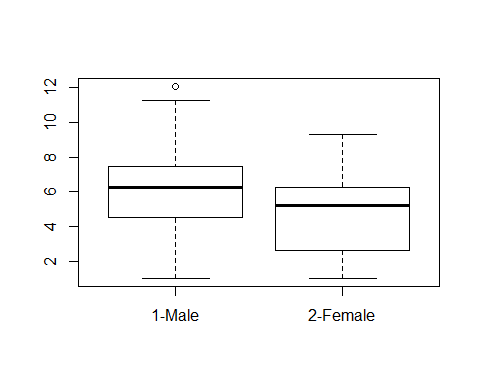


lambda = bc1$x[which(rank(-bc1$y)==1)]  
lambda

## [1] 0.4646465

# if lambda is 0.5, we use sqrt(pyp1)

boxplot(I(pyp1^lambda)~gender, data=expr.meta)



# plot 02  
# the skewness seems disappeared