Case study 1: analysis of the daily average wind speed in New York in 1973

Foundations for inference using R

# Case study 1: The wind speed in the airquality dataset

## Exploratory analysis of the daily average of the wind speed

The airquality dataset is a R object gives information about 153 daily air quality measurements () in New York, May to September 1973.

dim(airquality)

## [1] 153 6

head(airquality)

## Ozone Solar.R Wind Temp Month Day  
## 1 41 190 7.4 67 5 1  
## 2 36 118 8.0 72 5 2  
## 3 12 149 12.6 74 5 3  
## 4 18 313 11.5 62 5 4  
## 5 NA NA 14.3 56 5 5  
## 6 28 NA 14.9 66 5 6

The variable of primary interest, Wind, is the average wind speed in miles per hour at 0700 and 1000 hours at LaGuardia Airport. We use the R package ggplot2 to explore the data. Figure 1 shows histogram of wind speed.

ggplot(airquality, aes(x = Wind)) +  
 geom\_histogram(fill = "skyblue", color = "black")+  
 ylab("Frequency")

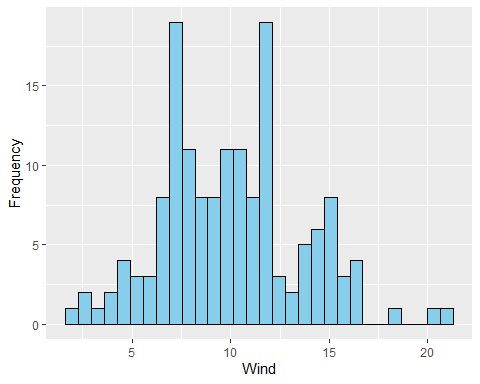


Figure 1: Histogram of wind speed.

The sample mean () and standard deviation () of the wind speed are 9.957516 and 3.523001, respectivly.

wind<-airquality$Wind  
M.wind<-mean(wind)  
SD.wind<-sqrt(var(wind))  
M.wind

## [1] 9.957516

SD.wind

## [1] 3.523001

## A confidence interval for the population mean

We constrct a confidence interval for the population mean using the R fucntion z.test. We assume that and the function uses the to select the critical value . For a confidence interval, . A confidence interval for the wind speed is [9.399284; 10.515749]. This means that we are confident that the true average wind speed lies within this range.

## Test of hypothesis about the population mean

library(TeachingDemos)  
wind=na.omit(airquality$Wind)  
z.test(wind,sd=SD.wind)

##   
## One Sample z-test  
##   
## data: wind  
## z = 34.961, n = 153.00000, Std. Dev. = 3.52300, Std. Dev. of the sample  
## mean = 0.28482, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 9.399284 10.515749  
## sample estimates:  
## mean of wind   
## 9.957516

Testing the hypotheses whether the wind speed is equal to 9 versus a two-sided alternative hypothesis at the significant level of 0.05 can be formulated by:

We use the z.test() function and specify mu=9$.

z.test(wind, SD.wind, mu=9)

##   
## One Sample z-test  
##   
## data: wind  
## z = 3.3619, n = 153.00000, Std. Dev. = 3.52300, Std. Dev. of the sample  
## mean = 0.28482, p-value = 0.0007742  
## alternative hypothesis: true mean is not equal to 9  
## 95 percent confidence interval:  
## 9.399284 10.515749  
## sample estimates:  
## mean of wind   
## 9.957516

Since p-value = 0.0007742 which is much smaller than , there is sufficient evidence to say that the mean of the wind speed is not equal to 9.