Univariate Statistical Analysis

By Saba Akram

12/27/2020

# Input the churn dataset  
churn <- read.csv("D:/Data Mining and Predictive Analysis/Data sets/churn.txt",stringsAsFactors = TRUE)  
# Analyse subgroup of a data.  
subchurn<-subset(churn,churn$Int.l.Plan=="yes"&churn$VMail.Plan=="yes"&churn$Day.Mins>220)  
summary(subchurn$CustServ.Calls)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000 0.750 1.000 1.607 2.000 9.000

length(subchurn$CustServ.Calls)

## [1] 28

# One Sample T test and confidence interval for mean.  
mean.test<-t.test(x=subchurn$CustServ.Calls,mu=2.4,conf.level = .95)  
mean.test

##   
## One Sample t-test  
##   
## data: subchurn$CustServ.Calls  
## t = -2.2171, df = 27, p-value = 0.03522  
## alternative hypothesis: true mean is not equal to 2.4  
## 95 percent confidence interval:  
## 0.8733969 2.3408888  
## sample estimates:  
## mean of x   
## 1.607143

mean.test$statistic

## t   
## -2.217128

mean.test$p.value

## [1] 0.03522289

mean.test$conf.int

## [1] 0.8733969 2.3408888  
## attr(,"conf.level")  
## [1] 0.95

# One Sample proportion Test and Confidence Interval  
num.churn<-sum(churn$Churn.=="True") # Churners  
num.churn

## [1] 0

sample.size<-dim(churn)[1]  
sample.size

## [1] 3333

p<-483/sample.size # Point Estimete  
Z\_data<-(p-.15)/sqrt((.15\*(1-.15))/sample.size)  
Z\_data # Test Statistics.

## [1] -0.8222369

error<-qnorm(0.975, mean=0, sd=1)\*sqrt((p\*(1-p))/sample.size)  
error

## [1] 0.01195064

lower.bound<-p-error  
lower.bound

## [1] 0.1329639

upper.bound<-p+error  
upper.bound

## [1] 0.1568651

p.value<-2\*pnorm(Z\_data,mean = 0,sd=1)  
p.value

## [1] 0.4109421