Hypothesis Testing

Name: Modem Praveen

Roll Number: 17BCS035

First few tests are done directly without using functions and later tests are done using function().For Z critical value and chi square test used inbuilt functions (qnorm() and chisq.test())

One sample Z-Test

Q) An automatic machine fills an aerated drink in 2000 cc bottles. A tester needs to test H0 that the average amount being filled in a bottle is at least 2000 cc. He selects a random sample of 40 bottles and records the exact content of the bottles and finds the sample mean to be 1999.6 cc. Consider the population standard deviation as 1.30 cc

Let's test the null hypothesis at the significance level of 5%.

Above population mean is known and Population standard deviation also known so it is Z-statistics

And sample size is more than 30 so we need to calculate Z-statistics.

Set null hypothesis and alternative hypothesis

H0 >= 2000 # Null Hypothesis

H1 < 2000 # Alternative Hypothesis - Lower tailed test

level of significance 0.05

alpha=0.05

n = 40 # Sample Size

Sample size id more than 30. So, need to calculate Z statistics

mu=2000

Xbar = 1999.6 # Sample mean

sigma = 1.3 # Population Std Dev

SE = sigma/sqrt(n) # Sample std deviation: 0.0422

Z_cal = (Xbar - mu)/SE   # Z score

Z_cal   # -1.96 |z|=1.96 std dev away from the mean

Z_critical=qnorm(1-alpha, lower.tail = FALSE)

Note: we can find it using Z-table also but in R we can find it with qnorm function

Z # Critical value for 95% confidence -1.64

Conclusion: Reject null hypothesis . So at 0.05 level of significance the claim of at least 2000cc of water in a bottle should be rejected.

Two sample Z-Test

Q)The amount of a certain trace element in blood is known to vary with a standard deviation of 14.1 ppm (parts per million) for male blood donors and 9.5 ppm for female donors. Random samples of 75 male and 50 female donors yield concentration means of 28 and 33 ppm, respectively. What is the likelihood that the population means of concentrations of the element are the same for men and women?

#Null hypothesis: H0: μ 1 = μ 2

#alternative hypothesis: H1 : μ 1 ≠ μ 2

Zcal= (x1_bar - x2_bar)/sqrt((sigma1)**2/n1 + (sigma2)**2/n2)

x1=28

x2=33

sigma1=14.1

sigma2=9.5

n1=75

n2=50

Z_cal=(x1-x2)/sqrt(((sigma1)**2/n1) + ((sigma2)**2/n2))

Z_cal #-2.37 |Z|=2.37

Z_critical=qnorm(1-(alpha/2)) # two tailed so alpha/2

Z_critical # 1.959964

Conclusion: Reject null hypothesis . z-score in either tail of the distribution (plus or minus) will lead to rejection of the null hypothesis of no difference.

One sample Proportion Test

Q) A researcher claims that Republican Party will win in next Senate elections especially in Florida State. A statistical data reported that 23% voted for Republican Party in last election. To test the claim a researcher surveyed 80 people and found 22 said they voted for Republican Party in last election. Is there enough evidence at $\alpha=0.05$ to support this claim?

po=0.23

n=80

p_cap=22/80

#Null Hypothesis: p= 0.23

#Alternative Hypothesis: p≠0.23

alpha=0.05

Z_cal=((p_cap-po)/sqrt(po*(1-po)/n))

Z_cal   # 0.9564

Z_critical=qnorm((1-alpha/2))

Z_critical # 1.9566,-1.9566

Conclusion: Reject Null hypothesis

Two sample Proportion test

A car manufacturer aims to improve the quality of the products by reducing the defects and also increase the customer satisfaction. Therefore, he monitors the efficiency of two assembly lines in the shop floor. In line A there are 18 defects reported out of 200 samples. While the line B shows 25 defects out of 600 cars. At α 5%, is the differences between two assembly procedures are significant?

#Null Hypothesis: Two proportions are the same

#Alternative Hypothesis: Two proportions are not the same

alpha=0.05

p1=18/200

p2=25/600

x1=18

x2=25

n1=200

n2=600

p0=(x1+x2)/(n1+n2)

Z_cal=(p1-p2)/(sqrt(p0*(1-p0)*((1/n1)+(1/n2))))

Z_cal # 2.62

Z_critical=qnorm((1-alpha)/2)

Z_critical   # 1.9566,-1.9566

Conclusion: Reject null hypothesis

One sample t-test

Q) A professor wants to know if her introductory statistics class has a good grasp of basic math. Six students are chosen at random from the class and given a math proficiency test. The professor wants the class to be able to score above 70 on the test. The six students get scores of 62, 92, 75, 68, 83, and 95. Can the professor have 90 percent confidence that the mean score for the class on the test would be above 70?

#null hypothesis: H 0: $\mu = 70$

#alternative hypothesis: H a : $\mu > 70$

xbar=mean(c(62,92,75,68,83,95))

xbar

s=sqrt(var(c(62,92,75,68,83,95)))

s

delta=70

n=6

tcal=(xbar-delta)/(s/sqrt(n))

tcal # 1.705314

Two sample t-test

#Does right- or left-handedness affect how fast people type? Random samples of students from a typing class are given a typing speed test (words per minute), and the results are compared. Significance level for the test: 0.10. Because you are looking for a difference between the groups in either direction (right-handed faster than left, or vice versa), this is a two-tailed test.

```r
#null hypothesis: H 0: μ 1 = μ 2
#alternative hypothesis: H a : μ 1 ≠ μ 2
#   Right - n = 16 , xbar=55.8 ,s=5.7
#   left - n=9 , xbar=59.3 ,s=4.3
n1=16
n2=9
x1=55.8
x2=59.3
s1=5.7
s2=4.3
mu=0


tcal<-function(n1,n2,x1,x2,s1,s2,mu,ttab){


  "H0: there is no significant diff btw 2 samples \n"
  "H1: there is significant difference btw 2 samples \n"
  if (n1 & n2){
    tcalculated=(x1-x2-mu)/ (sqrt(((s1**2)/n1)+((s2**2)/n2)))
    cat("tcal=", tcalculated)
    if(tcalculated<ttab){
      "Accept null hypothesis"
    }else{
      "Reject null hypothesis"
    }
  }else{
    "It is not t test"
  }
}
```

tcal(16,9,55.8,59.3,5.7,4.3,0,1.714)

```
>    tcal(16,9,55.8,59.3,5.7,4.3,0,1.714)
tcal= -1.731683[1] "Accept null hypothesis"
```

Paired Difference t-test

#A farmer decides to try out a new fertilizer on a test plot containing 10 stalks of corn. Before applying the fertilizer, he measures the height of each stalk. Two weeks later, he measures the stalks again, being careful to match each stalk's new height to its previous one. The stalks would have grown an average of 6 inches during that time even without the fertilizer. Did the fertilizer help? Use a significance level of 0.05.
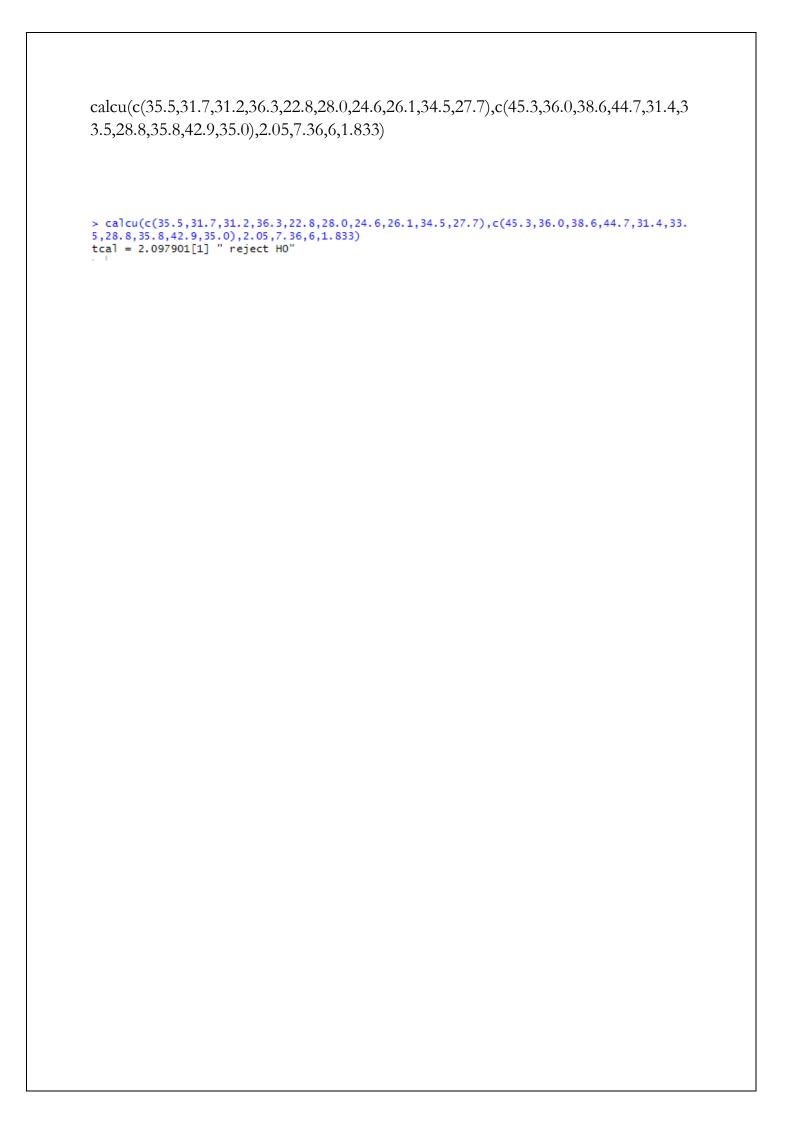
#null hypothesis: H 0: $\mu = 6$

alternative hypothesis: H a : $\mu > 6$

data is given in word document

```
calcu <-function(x1,x2,s,xbar,mu,ttab){
  n=length(x1)
  tcal=(xbar-mu)/(s/sqrt(n))
  cat("tcal =",tcal)

  if(tcal < ttab){
    "Accept H0"
  }else{
    " reject"
  }

}
```

calcu(c(35.5,31.7,31.2,36.3,22.8,28.0,24.6,26.1,34.5,27.7),c(45.3,36.0,38.6,44.7,31.4,33.5,28.8,35.8,42.9,35.0),2.05,7.36,6,1.833)

```
> calcu(c(35.5,31.7,31.2,36.3,22.8,28.0,24.6,26.1,34.5,27.7),c(45.3,36.0,38.6,44.7,31.4,33.
5,28.8,35.8,42.9,35.0),2.05,7.36,6,1.833)
tcal = 2.097901[1] " reject H0"
```

Chi square test

\# gender and preference of taking pet v1=Dog , v2=cat

data=matrix(c(207,231,282,242),nrow=2)

```
chi<-function(data,critical_value){
  "H0:There is no significant difference between two categorical variables"
  "H1:There is significant difference between two categorical variables"
  cal=chisq.test(data)
  cal
  if(cal$p.value<critical_value){
    "Reject Hypothesis"
  }else{
    "Accept Hypothesis"
  }
}
chi(data,0.05)


cal=chisq.test(data)
```

```
> chi(data,0.05)
[1] "Reject Hypothesis"
> cal=chisq.test(data)
> cal

        Pearson's Chi-squared test with Yates' continuity correction

data:  data
X-squared = 3.8454, df = 1, p-value = 0.04988
```