**STAT 40001/MA 59800 Statistical Computing Fall 2017**

**Lab-13**

1. United Airlines Flight 179 is a daily flight from Boston to San Francisco. Flight 180 goes in the other direction (SF to Boston). The data Flight 179 in *Lock5withR*  package provides the airborn flying times for each flight on the three dates each month (5th, 15th and 25th) in 2010. Test the hypothesis whether there is enough evidence that Flight179 takes 50 minutes longer than Flight 180 to reach to the destination.

install.packages("Lock5withR")

library(Lock5withR)

data(package="Lock5withR")

attach(Flight179)

t.test(Flight179,Flight180,alt="greater",mu=50)

Welch Two Sample t-test

data: Flight179 and Flight180

t = 1.5584, df = 62.523, p-value = 0.06209

alternative hypothesis: true difference in means is greater than 50

95 percent confidence interval:

49.5441 Inf

sample estimates:

mean of x mean of y

357.8611 301.4722

p-value > 0.05, do not have enough evidence to reject null hypothesis.

1. The ‘*bnlearn*’ package in R includes the data set coronary. Please access the data and print all the variables. Test the hypothesis whether proportion of the people with blood pressure level higher than 140 differ by smoking status.

install.packages("bnlearn")

library(bnlearn)

data(coronary)

head(coronary)  
 Smoking M. Work P. Work Pressure Proteins Family

1 no no no <140 <3 neg

2 no no no <140 <3 neg

3 no no no <140 <3 neg

4 no no no <140 <3 neg

5 no no no <140 <3 neg

6 no no no <140 <3 neg

names(coronary)

[1] "Smoking" "M. Work" "P. Work" "Pressure" "Proteins" "Family"

attach(coronary)

xtabs(~Smoking+Pressure)

Pressure

Smoking <140 >140

no 515 446

yes 539 341

prop.test(x=c(446,341),n=c(961,880))

2-sample test for equality of proportions with continuity correction

data: c(446, 341) out of c(961, 880)

X-squared = 10.702, df = 1, p-value = 0.00107

alternative hypothesis: two.sided

95 percent confidence interval:

0.03045299 0.12274680

sample estimates:

prop 1 prop 2

0.4640999 0.3875000

p-value < 0.05, reject null hypothesis. The proportion of the people with blood pressure level higher than 140 differ by smoking status.

1. The estimation of the average shrinkage percentage of plastic clay should have an error bound of 0.2 with 98% confidence. A pilot sample of 50 gave standard deviation of 1.2. Determine the sample size that should be used  
   library(BSDA)

nsize(b=0.2,sigma=1.2,conf.level = 0.98,type='mu')

The required sample size (n) to estimate the population   
mean with a 0.98 confidence interval so that the margin   
of error is no more than 0.2 is 195 .

1. A food processing company, considering the marketing of a new product, is interested in the proportion p of consumers that would try the new product. In a pilot sample of 40 randomly chosen consumers, 9 said that they would purchase the new product and give it a try. What sample size is needed for the 90% CI for p to have length 0.1.

nsize(b=0.05,p=9/40,conf.level = 0.9,type='pi')

The required sample size (n) to estimate the population proportion of successes with a 0.9 confidence interval so that the margin of error is no more than 0.05 is 189 .

1. Suppose that you are determining the required sample size for a two-sided independent samples t-test with 80% power and significant level of 0.05.  
   Generate a table showing the required sample size for each of the following effect sizes:

d = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5

Create a table showing the effect size versus sample size. Plot the graph for effect size versus sample size.

install.packages('pwr')

library(pwr)

samplesize=cbind(NULL,NULL)

for (i in c(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5))

{

power=pwr.t.test(d=i, power=0.8,sig.level=0.05,type="two.sample",alt="two.sided")

samplesize=rbind(samplesize,cbind(power$d,power$n))

}

samplesize

[,1] [,2]

[1,] 0.1 1570.733048

[2,] 0.2 393.405696

[3,] 0.3 175.384667

[4,] 0.4 99.080325

[5,] 0.5 63.765610

[6,] 0.6 44.585766

[7,] 0.7 33.024567

[8,] 0.8 25.524582

[9,] 0.9 20.386308

[10,] 1.0 16.714722

[11,] 1.1 14.001873

[12,] 1.2 11.942258

[13,] 1.3 10.343014

[14,] 1.4 9.077679

[15,] 1.5 8.060321

1. Suppose that you are determining the power of the test for a given sample size for a two-sided independent samples t-test with significant level of 0.05 and effect size d=0.7.

Generate a table showing the power of the test for following sample size:

n=5,10,15,20,25,30,35,40,45,50,55,60,65,70,75,80,85,90,100

Generate a table showing the power for each of the sample size.

> library(pwr)

> power=cbind(NULL,NULL)

> for(i in seq(5,100,5)){

+ p1=power.t.test(d=0.7,n=i,sig.level=0.05,alt="two.sided",type="two.sample")

+ power=rbind(power,cbind(p1$n,p1$power))}

> power

[,1] [,2]

[1,] 5 0.1631800

[2,] 10 0.3163866

[3,] 15 0.4566869

[4,] 20 0.5782714

[5,] 25 0.6790886

[6,] 30 0.7599031

[7,] 35 0.8229728

[8,] 40 0.8711328

[9,] 45 0.9072448

[10,] 50 0.9339067

[11,] 55 0.9533297

[12,] 60 0.9673141

[13,] 65 0.9772788

[14,] 70 0.9843134

[15,] 75 0.9892382

[16,] 80 0.9926597

[17,] 85 0.9950205

[18,] 90 0.9966389

[19,] 95 0.9977420

[20,] 100 0.9984898

> plot(power,xlab="Sample Size", ylab="Power", main="Sample Size Vs. Power ",type="b",col=2)

>

