Problem-06: According to a survey in 2008, the mean salary of MBA graduates in accounting was 37,000 TK per month. In a follow up study in June 2009, a sample of 48 MBA students graduating in accounting found a sample mean of 38,100 TK and a sample standard deviation of 5200 TK.

(i) Formulate the null and alternative hypothesis that can be used to determine whether the sample data support the conclusion that the MBA graduates in accounting have a mean salary greater than 37000 TK.

(ii) At 5% Level of significance what is your conclusion?

(iii) Find the p-value and state your conclusion.

(iv) Find 95% confidence interval for mean salary of MBA graduates.

Solution:

Code:

```
xbar = 38100; xbar
n = 48; n
meu = 37000; meu
sd = 5200; sd
alpha = 0.05; alpha
zcalculated = (xbar-meu)/(sd/sqrt(n)); zcalculated
ztabulated = qnorm(alpha, mean=0, sd=1, lower.tail
                =FALSE); ztabulated

if(zcalculated > ztabulated){
   print("Null Hypothesis is rejected")
}else{
   print("Null Hypothesis is accepted")
}
```

```
pvalue = pnorm(zcalculated, lower.tail=FALSE); pvalue
if(pvalue < alpha){
   print("Null Hypothesis is rejected")
}else{
   print("Null Hypothesis is accepted")
}
confidenceinterval = xbar - ztabulated*(sd/sqrt(n));
                     confidenceinterval
```

Output:

zcalculated = 1.465581
ztabulated = 1.644854 (Null hypothesis is accepted)
pvalue = 0.07138117 (Null hypothesis is accepted)
confidence interval = 36865.5

comment:

(i) The null and alternative hypothesis are given as follows:

$H_0 : \mu \leq 37000$

$H_A : \mu > 37000$

(ii) Here,

zcalculated < ztabulated

so, Null hypothesis is accepted at 5% level of significance. We can say that the MBA graduates in accounting have a mean salary which is less than or equal to 37000.

(iii) Here, p-value = 0.07138117 and $\alpha = 0.05$ where p-value > α that means null hypothesis is accepted.

(iv) Therefore, we are 95% sure that the confidence interval is 36865.5 for mean salary of MBA graduates.

Problem-07: The daily temperature (in degree celsius) of two months during summer season are shown below:

Months	Daily temperature (in degree Celsius)
1	32, 34, 31, 33, 35, 36, 34, 34, 34, 35, 32, 33, 33, 33, 32, 32, 34, 33, 32, 34, 32, 31, 33, 34, 35, 34, 33, 33, 33, 34, 34
2	34, 34, 35, 35, 35, 35, 35, 35, 36, 37, 34, 33, 34, 35, 34, 34, 36, 34, 33, 34, 32, 33, 34, 36, 35, 35, 35, 34, 35, 34

(i) Input the two sets of data using R software and save this file in CSV format in desktop.

(ii) Formulate the null hypothesis and alternative hypothesis that can be used to determine that the temperature of both months are not similar?

(iii) Calculate the value of test statistic and state your conclusion.

(iv) What is the p-value of this test? Give your conclusion based on p-value.

(v) construct box plots for these two sets of data. Do the box plots support your conclusion obtained in question (iv).

Solution:

code:

```
temp1= c(32, 34, 31, 33, 35, 36, 34, 34, 34, 35, 32, 33, 33
         33, 32, 32, 34, 33, 32, 34, 32, 31, 33, 34, 35,
         34, 33, 33, 33, 34, 34); temp1

temp2 = c(34, 34, 35, 35, 35, 35, 35, 35, 36, 37, 34,
          33, 34, 35, 34, 34, 36, 34, 33, 34, 32, 33, 34,
          36, 35, 35, 35, 34, 35, 34); temp2
```

```
d = data.frame(temp1, temp2);d
getwd()
write.csv(d, 'D:/Books/4th semested/STAT-2201/stat-2201
          /problem7.csv')
alpha = 0.05; alpha
x1bar = mean(temp1); x1bar
x2bar = mean(temp2); x2bar
n1 = length(~~n1~~ temp1); n1
n2 = length(temp2); n2
sd1 = sd(temp1); sd1
sd2 = sd(temp2); sd2
zstat = (x1bar - x2bar) / sqrt(sd1^2/n1 + sd^2/n2);
        zstat
ztab = qnorm(alpha/2, mean = 0, sd = 1); ztab
if (abs(zstat) > abs(ztab)){
   print("Null hypothesis is rejected")
}else{
   print("Null hypothesis is accepted")
}
pvalue = 2*pnorm(zstat); pvalue
if(pvalue < alpha){
   print("Null hypothesis is rejected")
}else{
   print("Null hypothesis is accepted")
}
boxplot(temp1, temp2, main = "Box Plot", xlab = "Month"
        ylab = "Temperature")
```

Problem-08: In a sample of 80 Americans, 44 wished that they were rich. In a sample of 90 Europeans, 41 wished that they were rich. Answer the following questions using R software:

(i) At $\alpha = 0.01$, is there a difference in the proportions?

(ii) What is the p-value of this test? What is your conclusion compared with p-value? compare this conclusion with conclusion obtained (i).

(iii) Find the 99% confidence interval for the difference of the two proportions.

Solution:

code:

```
a1 = 44; a1
n1 = 80; n1
a2 = 41; a2
n2 = 90; n2
alpha = 0.01; alpha
p1 = a1/n1; p1
p2 = a2/n2; p2
p = (a1+a2)/(n1+n2); p
q = (1-p); q
zstat = (p1-p2)/sqrt(p*q*(1/n1+1/n2)); zstat
ztab = qnorm(alpha/2, mean=0, sd=1); ztab
if(zstat > ztab){
  print("Null hypothesis is rejected")
}else{
  print("Null hypothesis is accepted")
}
```

```
pvalue =2 * pnorm(zstat, lower.tail= FALSE);pvalue
if (Pvalue <alpha){
  print("Null hypothesis is rejected")
}else{
  print("Null hypothesis is accepted")
}
LB =(P1-P2) - abs(ztab)*sqrt (p*q*(1/n1 + 1/n2)) ; LB
UB= (P1-P2) +abs(ztab)*sqrt (p*q*(1/n1 +1/n2));UB
```

Output:

P1 = 0.55

P2 = 0.4555556

P = 0.5

q = 0.5

zstat = 1.229273

ztab = 2.575829

Null hypothesis is accepted

pvalue = 0.2189696

Null hypothesis is accepted

LB = -0.1034553

UB = 0.2923442

comment:

(i) The null and alternative hypothesis are given as follows:

$$H_0 : P_1 = P_2$$

$$H_A : P_1 \neq P_2$$

Here, zstat < ztab that means null hypothesis is accepted and we can say that there is no difference in the proportion at level of significance, $\alpha = 0.01$

(ii) Here, Pvalue = 0.2189696 and Pvalue $> \alpha$ that means null hypothesis is accepted. From (i) we can see that, null hypothesis is accepted also. So, there is no difference in the proportions.

(iii) Here, Lower bound = -0.1034553 and upper bound = 0.2923442. Therefore, we are 99% sure that the difference value in the proportion lies between -0.1034553 to 0.2923442.

Problem-09: The number of students admitted in two departments in a university in different years are as follows:

Year	Statistics	Mathematics	Year	Statistic	Mathematics
2001	40	60	2011	37	55
2002	42	64	2012	38	54
2003	45	67	2013	43	69
2004	38	55	2014	42	65
2005	40	62	2015	39	59
2006	39	66	2016	46	70
2007	46	70	2017	42	68
2008	44	65	2018	41	62
2009	43	62	2019	42	64
2010	42	56	2020	38	58

The researchers claim that the variation in admission of students in different years are not same. Answer the following question using R software:

(i) Input the data in MS Excel and save this file in csv format. Export this csv file in R.

(ii) Formulate the null and alternative hypothesis

(iii) calculate the value of appropriate test statistic and comment on your result.

(iv) Find the p-value of this test and state your conclusion.

Solution:

code:

```
data = read.csv(file.choose()); data
math = data[,3]; math
stat = data[,2]; stat
sd_math = sd(math); sd_math
sd_stat = sd(stat); sd_stat
fcal = sd_math^2 / sd_stat^2; fcal
ftab = qf(alpha, df1 = 19, df2 = 19); ftab
if (fcal > ftab){
  print("Null hypothesis is rejected")
}else{
  print("Null hypothesis is accepted")
}
pvalue = 1 - pf(fcal, df1 = 19, df2 = 19); pvalue
if(pvalue < alpha){
 print("Null hypothesis is rejected")
}else{
  print("Null hypothesis is accepted")
}
```

output:

fcal = 3.615662

ftab = 0.330321

Null hypothesis rejected

pvalue = 0.003697033

Null hypothesis rejected

comment:

(i) We export this above data from csv file in R.

(ii) The null and alternative hypothesis are given as follows:

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_A : \sigma_1^2 \neq \sigma_2^2$$

(iii) Here, $f_{cal} > f_{tab}$ that means null hypothesis is rejected and we can say that the students in different years are not same.

(iv) Here, Pvalue = 0.003697033 whic is less than $\alpha = 0.05$ that means null hypothesis is rejected and the students in different years are not same.

Problem-10: The following are the heights (X in cm) and weights (in kg) of 15 persons.

X	160	165	159	164	168	155	158	155	152	159	158	154	153	152	154
Y	70	72	64	63	72	65	62	56	56	60	58	58	55	56	60

(i) Input the dataset using R software and save this file in csv format.

(ii) Test the hypothesis that the weight of animals significantly increased due to the increase in height? conclusion your result using p-value method.

(iii) Test the significance of correllation between weight and height. conclusion your result using p-value method.

Solution:

Code:

```
x=c(160,165,159,164,168,155,158,155,152,159,
    158,154,153,152,154);x

Y=c(70,72,64,63,72,65,62,56,56,60,58,58,55,
    56,60);Y

n=length(x);n
m=data.frame(x,Y);m
getwd()
Write.csv(m,'D:/Books/4th semester/STAT-2201/
          stat-2202/problem 10.csv')

alpha=0.05;alpha
reg=lm(m$Y~m$X,m);reg
summary(reg);
```

```
r = cor(x, y); r
tcal = r * sqrt(n-2)/(1 - r^2); tcal
ttab = qt(alpha/2, n-2); ttab
if (tcal > ttab){
  print("Null hypothesis is rejected")
}else{
  print("Null hypothesis is accepted")
}
```

Output:

Intercept = -93.2493

m$x = 0.9830

pvalue = 0.000136

r = 0.828

comment:

(i) We input this dataset using R software and save this file in csv formate.

(ii) We know, $Y = \alpha + \beta X + \epsilon$ where Y is dependent variable, X is independent variable, α is intercept and β is slope or coefficient of X and ϵ is random error term.

Here $\alpha = -93.2493$ and $\beta = 0.9830$. Pvalue = 0.000136 which is less that 0.05. So the null and alternative hypothesis are given as follows,

$H_0 : \beta = 0$

$H_A : \beta \neq 0$

$\therefore$ Pvalue = 0.000136 < 0.05 that means null hypothesis is rejected and the weight of animals don't significantly increased due to the increase in height.

(iii) Here, correlation, $r = 0.8284113$. This statistic follows student's t distribution with $(n-2)$ degrees of freedom.

Here, $t_{calculated} > t_{tabulated}$ that means the null hypothesis is rejected and the weight of animals significantly don't increased due to the increase in height.

Problem-03: A farm grows grapes for jelly. The following data are measurements of sugar in a grapes of a sample taken from each of 30 truck-loads.

16.0, 15.2, 12.0, 16.9, 14.4, 16.3, 15.6, 12.9, 15.3, 15.1, 15.8, 15.5, 12.5, 14.5, 14.9, 15.1, 16.0, 12.5, 14.3, 15.4, 15.4, 13.0, 12.6, 14.9, 15.1, 15.3, 12.4, 17.2, 14.7, 14.8

Assume that these observations of a random variable X that has mean μ and the standard deviation σ

(i) Find point estimation of μ and σ

(ii) construct an approximate 90% confidence interval for μ.

Solution:

code:

```
x=c(16.0,15.2,12.0,16.9,14.4,16.3,15.6,12.9,15.3,15.1,
    15.8,15.5,12.5,14.5,14.9,15.1,16.0,12.5,14.3,15.4,
    15.4,13.0,12.6,14.9,15.1,15.3,12.4,17.2,14.7,14.8
n=length(x);n
meu=sum(x)/n; meu
sigma=sqrt((sum(x^2)-n*(meu*2))/n);sigma
alpha=0.10;alpha
ztab=qnorm((alpha/2), mean=0, sd=1);ztab
LB=meu-abs(ztab)*sqrt(sigma/n);LB
UB=meu+abs(ztab)*sqrt(sigma/n);UB
```

Output:

mean = 14.72

sigma = 1.357547

alpha = 0.1

Ztab = -1.644854

LB = 14.3701

UB = 15.0699

comment:

(i) Here, Point estimates of $\mu = 14.72$ and $\sigma = 1.357547$

(ii) Therefore, we are 90% sure that the μ lies between 14.3701 and 15.0699.