INDEX

SN	Experiment Name
1.	Suppose we have a population of four measurements 2, 4, 6 and 8. Draw a random sample of size 2 without replacement and demonstrate that (i) The sample mean is an unbiased estimate of population mean. (ii) $v(\bar{y}) = \frac{\sigma^2}{n}\frac{N-n}{N-1}$ (iii) Verify is s^2 an unbiased estimate of σ^2? (iv) Find 95% confidence interval for population mean and population total. (v) Answer the above questions for sampling with replacement.
2.	Let X and Y denote the strength of concrete beams and cylinders. The following data are obtained X:5.9,7.2,7.3,6.3,8.1,6.8,7.0,7.6,6.8,6.5,7.0,6.3,7.9,9.0,8.2,8.7,7.8,9.7,7.4,7.7,9.7,7.8,7.7,11.6,11.3,11.8,10.7. Y:6.1,5.8,7.8,7.1,7.2,9.2,6.6,8.3,7.0,8.3,7.8,8.1, 7.4,8.5,8.9,9.8,9.7,14.1,12.6,11.2. (i) Show that $\bar{X} - \bar{Y}$ is an unbiased estimator of $\mu_1 - \mu_2$. Calculate it for the given data. (ii) Find the variance and standard deviation (standard error) of the estimator in Part(i), and then compute the estimated standard error. (iii) Calculate an estimate of the ratio σ_1/σ_2 of the two standard deviations. (iv) Suppose a single beam X and a single cylinder Y are randomly selected. Calculate an estimate of the variance of the difference X −Y.
3.	A farm grows grapes for jelly. The following data are measurements of sugar in a grapes of a sample taken from each of 30 truckloads. 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 estimates of μ and σ . (ii) Construct an approximate 90% / 95% / 80% confidence interval for μ .
4.	Draw random number of size 200 from (a) normal distribution with mean 50 and variance 26 and (b) exponential distribution with mean 60. (i) Find the estimate of the parameters by maximum likelihood method. (ii) Construct a 90% / 95% / 80% confidence interval for the parameter(s). (iii) Estimate the variance using exponential distribution
5.	The sample mean from population with pdf $f(x;\theta) = \theta e^{-\theta x}; x > 0, \theta > 0$ are given below 0.46, 0.38, 0.61, 0.82, 0.59, 0.53, 0.72, 0.44, 0.58, 0.60, 0.73, 0.55, 0.23, 0.62, 0.38, 0.27, 0.36, 0.47, 0.49, 0.71. (i) Find the estimate of θ by maximum likelihood method.

INDEX

	(ii) Construct a 90% / 95% / 80% confidence interval for θ . (iii) Estimate the variance of θ .
6.	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 5,200 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 37,000 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.
7.	The daily temperature (in degree Celsius) of two months during summer season are shown below: <table><tr><td>Months</td><td>Daily temperature (in degree Celsius)</td></tr><tr><td>1</td><td>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</td></tr><tr><td>2</td><td>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</td></tr></table>(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)
8.	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 in (i). (iii) Find the 99% confidence interval for the difference of the two proportions.

INDEX

9. The number of students admitted in two departments in a university in different years are as follows:

Year	Statistics	Mathematics	Year	Statistics	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 researcher 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.

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.

Test the significance of correlation between weight and height. Conclusion your result using p-value method

Problem-01: suppose, we have a population of four measurements 2, 4, 6 and 8. Draw a random sample of size 2 without replacement and demonstrate that

(i) The sample mean is an unbiased estimate of population mean.

(ii) $V(\bar{y}) = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$

(iii) Verify is s^2 an unbiased estimate of σ^2?

(iv) Find 95% confidence interval for population mean and population total.

(v) Answer the above questions for sampling with replacement.

Solution:

code: (without replacement)

```
data = c(2,4,6,8); data
sample.data = rbind(c(2,4), c(2,6), c(2,8), c(4,6),
                    c(4,8), c(6,8)); sample.data

population_mean = mean(data); population_mean
ybar = rowMeans(sample.data); ybar
exp_ybar = sum(ybar*(1/6)); exp_ybar

exp_ybar2 = sum((ybar^2)*(1/6)); exp_ybar2
v_ybar = exp_ybar2-(exp_ybar^2); v_ybar
N = length(data); N
n = 2; n
var = var(data); var
sigma2 = (var*(N-1))/N; sigma2
RHS = ((sigma2*(N-n))/(n*(N-1))); RHS
```

```
s2 = (sample.data[,1] -ybar)^2 +(sample.data[,2]-
        ybar)^2 ; s2
exp-s2 = sum(s2 * (1/6)); exp-s2

alpha = 0.05; alpha
ztab = qnorm(alpha/2, mean=0, sd = 1); ztab
LB = population_mean - abs(ztab) * sqrt(var/N);LB
UB = population_mean + abs(ztab)* sqrt(var/N);UB
population_total = population_mean * N; population
                                          _total

var_total = N^2 * sigma2; var_total
LB1 = population_total - abs(ztab) * sqrt(var_total
        /N); LB1
UB1 = population_total + abs(ztab) * sqrt(var_total
        /N); UB1
```

(with replacement)

```
data = c(2,4,6,8); data
sample.data = rbind (c(2,2),c(2,4),c(2,6), c(2,8),
                     c(4,2),c(4,4)c(4,6)c(4,8), c(6,2),
                     c(6,4)c(6,6)c(6,8), c(8,2),c(8,4),c(8,6),
                     c(8,8)); sample.data

pop.mean = mean(data); pop.mean
ybar = rowMeans(sample-data); ybar
exp-ybar = sum(ybar * (1/6)); exp-ybar
exp-ybar2 = sum ((ybar^2) * (1/6)); exp-ybar2
v-ybar = exp_ybar2 - (exp-ybar^2) ;v-ybar
N = length (data); N
n = 2; n
```

```
var = var(data); var
sigma2 = (var * (N-1)) / N; sigma2
RHS = ((sigma2 * (N-n)) / (n * (N-1))); RHS
s2 = (sample.data[,1] - ybar)^2 + (sample.data
      [,2] - ybar)^2 ; s2
exp_s2 = sum(s2 * (1/6)); exp_s2
alpha = 0.05; alpha
ztab = qnorm(alpha/2, mean = 0, sd = 1); ztab
LB = pop.mean - abs(ztab) * sqrt(var/N) ; LB
UB = pop.mean + abs(ztab) * sqrt(var/N); UB
population_total = pop.mean * N; population_total
var_total = N^2 * sigma2; var_total
LB1 = population_total - abs(ztab) * sqrt(var_
      total/N); LB1
UB1 = population_total + abs(ztab) * sqrt(var_
      total/N); UB1
```

Output:

For without replacement,

```
population_mean = 5
exp_ybar = 5
v_ybar = 1.6666667
RHS = 1.6666667
sigma2 = 5
exp_s2 = 6.6666667
LB = 2.469697
UB = 7.530303
LB1 = 11.23477
UB1 = 28.76523
```

for with replacement,

pop. mean = 5

exp-ybar = 13.33333

v-ybar = 104.444

RHS = 1.6666667

sigma2 = 5

exp-s2 = 13.3333

LB = 2.469697

UB = 7.530303

LB1 = 11.23477

UB1 = 28.76523

comment:

For without replacement,

(i) Here, Population mean $(\mu) = 5$ and expected value of sample mean $(E(\bar{y})) = 5$. Therefore both are equal and we can say that the sample mean is an unbiased estimate of population mean.

(ii) Here, $v(\bar{y}) = 1.666667$ and $\frac{\sigma^2}{n} \frac{N-n}{N-1} = 1.666667$ therefore both are equal.

(iii) Here, $\sigma^2 = 5$ and expected value of sample variance$(s^2) = 6.6666667$. Therefore both are not equal.

(iv) For population mean,

Lower bound, LB = 2.469697

Upper bound, UB = 7.530303

Therefore we are 95% confident that population mean is between 2.46 and 7.530303

For population total,

Lower bound, LB1 = 11.23477

Upper bound, UB1 = 28.76523

therefore, we are 95% confident that the population total is between 11.23477 and 28.76523.

For with replacement,

(i) Here, population mean = 5 and expected value of sample mean ($E(\bar{y})$) = 13.3333. Therefore, both are not equal and we can say that the sample mean is not an unbiased estimate of population mean.

(ii) Here, $V(\bar{y}) = 104.444$ and $\frac{\sigma^2}{n} \frac{N-n}{N-1} = 1.666667$ therefore, both are not equal.

(iii) Here, $\sigma^2 = 5$ and expected value of sample variance ($E(s^2)$) = 13.3333 therefore, both are not equal. ~~and~~

(iv) For population mean,

Lower bound, LB = 2.469697

Upper bound, UB = 7.530303

therefore, we are 95% confident that population mean is between 2.469697 and 7.530303.

For Population total,

Lower bound, LB1 = ~~2.469697~~ 11.23477

Upper bound, UB1 = ~~7~~ 28.76523

therefore, we are 95% confident that population total is between 11.23477 and 28.76523

Problem-02: let x and Y denote the strength of concrete beams and cylinders. The following data are obtained

X: 5.9, 7.2, 7.3, 6.3, 8.1, 6.8, 7.0, 7.6, 6.8, 6.5, 7.0, 6.3, 7.9, 9.0, 8.2, 8.7, 7.8, 9.7, 7.4, 7.7, 9.7, 7.8, 7.7, 11.6, 11.3, 11.8, 10.7.

Y: 6.1, 5.8, 7.8, 7.1, 7.2, 9.2, 6.6, 8.3, 7.8, 8.1, 7.4, 8.5, 8.9, 9.8, 9.7, 14.1, 12.6, 11.2

(i) show that $\bar{x}-\bar{Y}$ is an unbiased estimator of $\mu_1-\mu_2$. Calculate it for the given data.

(ii) Find the variance and standard deviation (standard error) of the estimator in Part (i) and then compute the estimated standard error.

(iii) Calculate an estimate of the ratio $\frac{\sigma_1}{\sigma_2}$ of the two standard deviations.

(iv) Suppose a single beam x and a single cylinder Y are randomly selected. Calculate an estimate of the variance of the difference X - Y

Solution:

code:

```
x = c(5.9, 7.2, 7.3, 6.3, 8.1, 6.8, 7.0, 7.6, 6.8, 6.5, 7.0,
      6.3, 7.9, 9.0, 8.2, 8.7, 7.8, 9.7, 7.4, 7.7, 9.7, 7.8, 7.7,
      11.6, 11.3, 11.8, 10.7); x

Y = c(6.1, 5.8, 7.8, 7.1, 7.2, 9.2, 6.6, 8.3, 7.8, 8.1, 7.4, 8.5,
      8.9, 9.8, 9.7, 14.1, 12.6, 11.2); Y

ln_x = length(x); ln_x
ln_Y = length(Y); ln_Y
xbar = mean(x); xbar
Ybar = mean(Y); Ybar
dif = abs(xbar-Ybar); dif

var_x = var(x); var_x
var_Y = var(Y); var_Y
```

n1 = length(x); n1
n2 = length(Y); n2
std_error = sqrt((var_x/n1) + (var_Y/n2)); std_error
ratio = sqrt(var_x)/sqrt(var_Y); ratio
var_diff = (var_x + var_Y); var_diff

Output:

xbar = 8.140741
Ybar = 8.575
dif = 0.4342593
var_x = 2.754046
var_Y = 4.427237
n1 = 27
n2 = 20
std_error = 0.5686506
ratio = 0.7887133
var_diff = 7.181282

comment:

(i) We know, $E(\bar{X}-\bar{Y}) = E(\bar{X}) - E(\bar{Y}) = \mu_1 - \mu_2$
Hence, the unbiased ~~data~~ estimate based on the given data is $dif(\bar{x}-\bar{Y}) = 0.4342593$

(ii) Here, variance of x (var_x) = 2.754046 and variance of Y (var_Y) = 4.427237 and standard error = 0.5686506

(iii) The ratio $\left(\frac{\sigma_1}{\sigma_2}\right) = 0.7887133$

(iv) var(x-Y) = 7.181282; it the variance of the difference x-Y.

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-220
            /problem7.csv')
alpha = 0.05; alpha
x1bar = mean(temp1); x1bar
x2bar = mean(temp2); x2bar
n1 = length(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")
```

Output:

Zstat = -4.341794

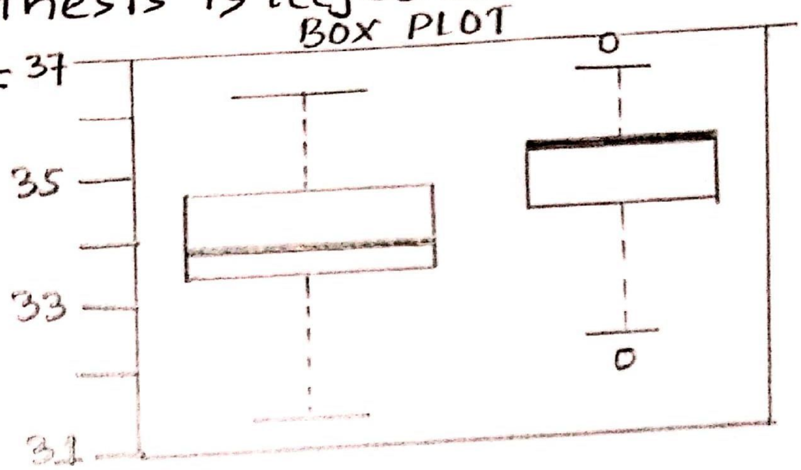
Ztab = -1.959964

Null hypothesis rejected

pvalue = 1.41324e-05

Null hypothesis is rejected

Box Plot =



comment:

(i) we input the two sets of data using R software and save this fine csv format in desktop.

(ii) Let X_1 and X_2 be the temparatuerce of month-1 and month-2 respectively. Assume that $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$. we need to test the hypothesis

$$H_0: \mu_1 = \mu_2$$

$$H_A: \mu_1 \neq \mu_2$$

(iii) Here, we use z statistic because n_1 and n_2 both are greater than 30 and the test statistic, Zcalculated = -4.341794 and tabulated value as Ztabulated = -1.959964. Since Zcalculated > Ztabulated, so null hypothesis is rejected at 5% level of significance and we can say that the temperature of both months are not similar.

(iv) Here, p value = $1.41324e^{-05}$ which is less than $\alpha = 0.05$ and the null hypothesis is rejected. So, we can say that the temparature of both months are not similar.

(v) Yes, the box plots support my conclusion obtained in question (iv).

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(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 > α 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 researcher 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
pNull 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 data set 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 = corr(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 statistics 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 deriviation 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.

Problem-05: The sample mean from population with pdf $f(x;\theta) = \theta e^{-\theta x}$; $x > 0, \theta > 0$ are given below - 0.46, 0.38, 0.61, 0.82, 0.59, 0.53, 0.72, 0.44, 0.58, 0.60, 0.73, 0.55, 0.23, 0.62, 0.38, 0.27, 0.36, 0.47, 0.49, 0.71.

(i) Find the estimate of θ by maximum likelihood method.

(ii) Construct a 90% / 95% / 80% confidence interval for θ.

(iii) Estimate the variance of θ.

Solution:

Code:

```
data=c(0.46,0.38,0.61,0.82,0.59,0.53,0.72,0.44,0.58,0.60,0.73,0.55,0.23,0.62,0.38,0.27,
       0.36,0.47,0.49,0.71); data
n = length(data);n
xbar = sum(data)/n;xbar
theta = 1/xbar;theta
alpha = 0.05;alpha
ztab1 = qnorm(1-(alpha/2)); ztab1
LB1 = theta-abs(ztab1)*(theta/sqrt(n)); LB1
UB1 = theta+abs(ztab1)*(theta/sqrt(n)); UB1
var_theta = 1/theta^2; var_theta
```

Output:

```
xbar = 0.527
theta = 1.897533
alpha = 0.05
LB1 = 1.065918
UB1 = 2.729149
var_theta = 0.277729
```

Problem-04: Draw random number of size 200 from (a) normal distribution with mean 50 and variance 26 and (b) exponential distribution with mean 60.

(i) Find the estimate of the parameters by maximum likelihood method.

(ii) construct a 90% / 95% / 80% confidence interval for the parameter(s).

(iii) Estimate the variance using exponential distribution.

Solution:

code:

```
norm = rnorm(200, 50, 26); norm
exp = rexp(200, 1/60); exp
n = length(norm); n
n1 = length(exp); n1
muhat = sum(norm)/n; muhat
sigmahut = sqrt(sum(norm^2)-(muhat^2))/n); sigmahut
lemda = 1/mean(exp); lemda

alpha = 0.05; alpha
ztab = qnorm(alpha/2, mean=0, sd =1); ztab
LB = 50 - abs(ztab) * sqrt(26/n); LB
UB = 50 + abs(ztab) * sqrt (26/n); UB
ztab1 = qnorm(1-(alpha/2)); ztab1
LB1 = lemda - abs(ztab1) *(lemda/sqrt(n1)); LB1
UB1 = lemda + abs(ztab1)*(lemda/sqrt(n1)); UB1
var_exp = 1/(lemda^2); var_exp
var_exp1 = var(exp); var_exp1
```

Output:

muhat = 48.90645

sigma hut = 28.64749

lemda = 0.0166925

LB = 49.29332

UB = 50.70668

LB1 = 0.01437947

UB1 = 0.01900643

var_exp1 = ~~3588.672~~ 3251.737

comment:

(i) By maximum likelihood method, the
parameters of the normal distribution are
$\hat{\mu}$ and $\hat{\sigma}$ where, $\hat{\mu} = 48.90645$ and $\hat{\sigma} = 28.64749$
and the parameters of the exponential distribution
are $\hat{\lambda}$ where, $\hat{\lambda} = 0.0166925$

(ii) For 95% confidence interval, $\alpha = 0.05$.
The parameter of normal distribution is mean
where lower bound = 49.29332 and upper bound
= 50.70668 that means we are 95% sure that
the mean lies between 49.29332 and 50.70668.
Again, the parameter of exponential distribution
is λ where lower bound = 0.01437947 and upper
bound = 0.01900643 that means we are 95% sure
that the mean lies between 0.01437947 and
0.01900643.

(iii) The variance using exponential distribution
is 3251.737

comment:

(i) By maximum liklihood method, the estimate of $\theta = 1.897533$

(ii) For 95% confidence interval, alpha = 0.05
Here, lower bound = 1.065918 and upper bound = 2.729149. so, we are 95% sure that θ is lies between 1.065918 and 2.729149.

(iii) The variance of $\theta = 0.277729$