* problem NO = 01

* problem Name : Suppose we have a population of four measurements 2, 4, 6, 8. Draw a random Sample of size 2 without replacement and demonstrate that,

(i) The sample mean is 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.

Source code:

(i)

```
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
pop.mean = mean(data); pop.mean
Yibar = rowMeans(Sample.data); Yibar
E.Ybar = sum(Yibar*(1/6)); E.Ybar
```

(ii)

```
E.Yibar2 = sum((Yibar^2)*(1/6)); E.Yibar2
V.Ybar = E.Yibar2 - (E.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
```

(iii)

```
s2 = (Sample.data[,1] - Yibar)^2 + (Sample.data[,2] -
                                      Yibar)^2; s2
E.s2 = sum(s2*(1/6)); E.s2
```

(iv)

```
alpha = 0.05; alpha
ztab = qnorm(alpha/2, mean=0, sd=1); ztab
LCL = pop.mean - abs(ztab)*sqrt(var/N); LCL
UCL = pop.mean + abs(ztab)*sqrt(var/N); UCL
```

```
total = pop.mean * N; total
Vartotal = N^2 * sigma2 ; Vartotal
LCLt = total - abs (ztab) * sqrt (vartotal/N) ; LCLt
UCLt = total + abs (ztab) * sqrt (vartotal/N) ; UCLt
```

Output:

(i) population mean = 5

Comment: The sample mean ($\bar{x}$) is an unbiased estimate of the population mean (μ).

(ii) $V(\bar{y}) = 1.6667$ and $\frac{\sigma^2}{n} \cdot \frac{N-n}{N-} = 1.6667$

Comment: RHS and LHS is almost equal.

(iii) population Variance = 6.6667

Comment: S^2 is an unbiased estimate of population Variance σ^2.

(iv) 95% Lower confidence level = 2.469697
95% upper confidence level = 7.530303
Total population mean = 20
Total lower confidence level = 11.23477
Total upper confidence level = 28.76523.

※ problem NO = 2

※ problem Name: 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 $\frac{\sigma_1}{\sigma_2}$ 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 $(X-Y)$.

Source code:

(i)

```
X = c(5.9,7.2,7.3,6.3,8.1,6.8,7.0,7.6,6.3,8.7,7.4,11.6,11.3,10.7); X
Y = c(6.1,5.8,7.8,7.1,7.2,9.2,6.6,8.3,7.0,8.3,8.9,14.1,12.6,11.2); Y

Xbar = mean(X) ; Xbar
Ybar = mean(Y) ; Ybar
dif = abs(Xbar-Ybar)
```

(ii)

```
Xvar = var(X) ; Xvar
Yvar = var(Y) ; Yvar
n1 = length(X) ; n1
n2 = length(Y) ; n2
SE.dif = sqrt((Xvar/n1) + (Yvar/n2)) ; SE.dif
```

(iii)

```
ratio = sqrt(Xvar)/sqrt(Yvar) ; ratio
```

(iv)

```
var.dif = Xvar + Yvar ; var.diff.
```

~~S~~ Output:

(i) $\bar{X} - \bar{Y} = 0.4342593$

$\mu_1 - \mu_2 = 0.4342593$

Comment: $\bar{X} - \bar{Y}$ is an unbiased estimator $\mu_1 - \mu_2$

(ii) Variance $= 2.754046$

Estimator standard error $= 0.5686506$

(iii) ratio $= 0.7887133$

(iv) Variance of the difference $X - Y = 7.181282$

* problem NO = 03

* problem Name: A farm grows graps for jelly. The following data are measurements of sugar in graps of a sample taken from each of 30 truckloads.

15.1, 16.0, 15.2, 12.0, 16.9, 14.4, 16.3, 15.6, 12.9, 15.3, 15.1, 16.0, 12.5, 14.3, 15.4, 15.8, 15.5, 12.5, 14.5, 14.9, 15.1, 16.0, 12.5, 14.3, 15.4, 15.8, 15.5, 12.5, 14.5, 14.9, 15.1, 15.3, 12.4, 17.2, 14.7, 14.8, 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 standard deviation σ.

(i) Find point estimates of μ and σ.

(ii) Construct an approximate 90%/95%/80% confidence interval for μ.

5

Source code:

(i)

```
x = c(16.0, 15.2, 12.0, 16.9, 14.4, 16.3, 15.6, 12.9, 15.3,
      15.8, 15.5, 12.5, 14.5, 15.1, 12.5, 14.3, 15.4, 15.4,
      13.0, 12.6, 14.9, 15.1, 15.3, 17.2, 14.7, 14.8); x

n = length(x); n

muhat = sum(x)/n ; muhat
sigmahat = sqrt(sum(x^2) - n * muhat^2)/n) ; sigmahat
```

(ii)

```
alpha = 0.10 ; alpha
ztab = qnorm(alpha/2; mean=0, sd=1); ztab
LCL = muhat - abs(ztab)* sqrt(sigmahat/n) ; LCL
UCL = muhat + abs(ztab)* sqrt(sigmahat/n); UCL
```

output:

(i) $\mu = 14.72$

$\sigma = \cancel{14.72}\ 1.357547$

(ii) 90% lower confidence level ~~of~~ = 14.3701.

90% upper confidence level = 15.0699.

** problem No = 4

** problem Name: 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 esmitimate 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.

** Source code:

```
norm = rnorm(200,50,26); norm
exp = rexp(60); exp
```

(i)

```
n = length(norm); n
muhat = sum(norm)/n; muhat
Sigmahat = sqrt((sum(norm^2)-n*muhat^2)/n)); sigmahat
thetahat = 1/mean(exp); Thetahat
```

(ii)

```
alph = 0.10; alpha
Ztab = qnorm(alpha/2, mean = 0, sd = 1); Ztab
LCL = muhat - abs(Ztab)*sqrt(sigma hat/n); LCL
UCL = muhat + abs(Ztab)*sqrt(sigma hat/n); UCL
```

(iii)

```
Var = 1/(n/sum(exp^2)); Var
```

Output: (i) $\mu = 49.06334$

$\alpha = 25.08312$

$\theta = 0.9423804$

(ii) 90% lower confidence level = 48.48083

90% upper confidence level = 49.64585

(iii) Variance = 0.6147883.

Problem No = 5

Problem Name: 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.

Output:

~~(i) Theta = 1.897533~~

Source code:

(i)

```
x = c(0.46, 0.38, 0.61, 0.82, 0.59, 0.53, 0.72, 0.58,
      0.60, 0.73, 0.55)
n = length(x); n
theta = n/(sum(x)); theta.
```

(ii)

```
muhat = sum(x)/n ; muhat
sigmahat = sqrt((sum(x^2) - n * muhat^2)/n); sigmahat
alpha = 0.010 ; alpha
ztab = qnorm(alpha/2, mean=0, sd=1) ; ztab
LCL = muhat - abs(ztab) * sqrt(sigmahat/n); LCL
```

(iii) Vartheta = 1/(n * (1/theta)); vartheta

Output :

(i) $\theta = 1.897533$

(ii) 90% lower confidence level = 0.3831464
90% upper confidence level = 0.6708536

(iii) Variance of θ = 0.09487666 .

problem No = 6

problem Name: According to a survey in 2008 the mean salary of MBA gradabtes in accounting was 37000 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 38100 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 concalusion 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.

Source code:

```
xbar = 38100 ; xbar
n = 48 ; n
sd = 5200 ; sd
mu = 37000 ; mu
alpha = 0.05 ; alpha
Zstat = (xbar - mu) / (sd/sqrt(n)) ; Zstat
Ztab = qnorm(alpha, mean=0, sd=1, lower.tail=False) ; Ztab
if (Zstat > Ztab) {
  print("Null hypothesis is rejected") }
 else { print("Null hypothesis is accepted") }
pval = pnorm(Zstat, lower.tail=FALSE) ; pval
if (pval < alpha) {
  print("Null hypothesis is rejected") }
 else { print("Null hypothesis is accepted") }
cl = xbar - Ztab * sd/sqrt(n) ; cl
```

Output:

(i) sd = 5200
mu = 37000

(ii) Ztab = 1.644854
Zstat = 1.465581

∴ Ztab > Zstat

So Null hypothesis is accepted.

Conclusion: MBA graduates in according have a mean Salary is not greater than 37000 tK.

(iii) pval = 0.07138117

∴ pval > alpha

So Null hypothesis is accepted.

(iv) 95% confidence interval is 36865.45.

* problem NO = 7

* problem Name: 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).

13

Source code:

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

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

data = cbind(temp1, temp2); data
getwd()
write.csv(data, 'c:/users/HP/Desktop/data.csv')
alpha = 0.05; alpha
n1 = length(temp1); n1
n2 = length(temp2); n2
xbar1 = mean(temp1); xbar1
xbar2 = mean(temp2); xbar2
sd1 = sd(temp1); sd1
sd2 = sd(temp2); sd2
zstat = (xbar1-xbar2)/sqrt(sd1^2/n1 + sd2^2/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") }

pval = 2*pnorm(~~ztab~~); pval
if (pval<alpha) { print("Null hypothesis is rejected") }
else { print(" Null hypothesis is accepted") }
```

```
boxplot (temp1, temp2, main="boxplot", xlab="month", ylab="Temperature")
LCL = (xbar1-xbar2) - abs(ztab) * sqrt(sd1^2/n1 + sd2^2/n2) ; LCL
UCL = (xbar1-xbar2) + abs(ztab) * sqrt(sd1^2/n1 + sd2^2/n2) ; UCL
```

Output: (i) Saved file in desktop in csv format

(ii) $H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

(iii) ztab = -1.959964

zstat = -4.341794

abs(zstat) > abs(ztab)

So Null hypothesis is rejected.

Conclusion: The temperature of both months are not similar.

(iii) pval = 0.9999929

Null hypothesis is ~~accepted~~ Rejected.

(iv) Lower confidence level = -1.779158

upper confidence level = -0.6724553

Problem No = 8:

Problem Name: 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 the 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 two proportions.

Source code :

```
alpha = 0.01 ; alpha
a1 = 44 ; a1
n1 = 80 ; n1
a2 = 41 ; a2
n2 = 90 ; n2
p1 = a1/n1 ; p1
p2 = a1/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, lower.tail=FALSE); ztab
if (zstat > ztab) { print("Null hypothesis is rejected") }
else { print("Null hypothesis is accepted") }
pval = 2*pnorm(zstat, lower.tail=FALSE) ; pval
if (pval<alph) { print("Null hypothesis is rejected") }
else { print("Null hypothesis is accepted") }
LCL = (p1-p2) - abs(ztab)* sqrt(p*q*(1/n1+1/n2) ; LCL
UCL = (p1-p2) + abs(ztab)* sqrt(p*q*(1/n1+1/n2) ; UCL
```

Output :

(i) Difference in the proportions 0.094445

Null hypothesis is accepted.

(ii) pval = 0.2189696

pval > alpha so Null hypothesis is accepted

Conclusion : Question ① and Question (ii) are the same Null hypothesis is accepted.

(iii) 99% Lower confidence level = -0.1034553

99% upper confidence level = 0.2923442.

Problem No = 09

Problem Name: 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	66	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 questions using R software.

(i) Input the data in ms excel and save the 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.

Source 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
alpha = 0.05; alpha
Fcal = sd.math^2/sd-stat^2; Fcal
Ftab = qf(alpha, df1=19, df2=19, lower.tail="FALSE"); Ftab
if (Fcal > Ftab){ print(" Null hypothesis is rejected")}
else{ print(" Null hypothesis is accepted")}
pval = 1-pf(Fcal, df1=19, df2=19, lower.tail="FALSE"); pval
if (pval<alpha){ print(" Null hypothesis is rejected")}
else{ print(" Null hypothesis is accepted")}
```

output:

(i) saved file in csv format.

(ii) sd-math = 4.738729

(iii) sd-stat = 0.01

(iii) Null hypothesis is rejected.

conclusion: The researcher claim that the variation in admission students different year are not same.

(iv) Null hypothesis is rejected.

** problem No = 10

** problem Name: The following are the hights (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 correlation between weight and height. conclusion your result using p-value method.

Source code:

```
X=c(160,165,159,164,168,155,158,154,153,152,154); X
Y=c(70,72,64,63,72,62,56,56,60,58,58,56,60); Y

n=length(X) ; n
data=cbind(X,Y); data
m= data.frame(X,Y) ; m
write.csv(data,'c:/users/Hp/desktop/test2/data6.csv')
alph=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(pha/2,n-2); ttab
if (abs(tcal) > abs(ttab)){ print("Null hypothesis is rejected")}
else{print("Null hypothesis is accepted")}

pval = 2*pt(tcal,n-2, lower.tail=FALSE); pval
if (pval<alpha){ print("Null hypothesis is rejected")}
else{ print("Null hypothesis is accepted")}
```

~~5~~ output:

(i) Saved this file in csv format.

(ii) Null hypothesis rejected.

conclusion: The weight of animals significantly increased due to the increase in height.

(iii) Null hypothesis is rejected.