

Name: Chandana Ramesh Galgali

Batch: B-4

Roll no: 16010422234

Tutorial Name: Tutorial 6 - Sampling using R

Date: 11 March 2024

Q.1 Test the significance of the difference between the means of two normal population with the same standard deviation from the following data.

| | Size | Mean | St. Dev |
|----------|------|------|---------|
| Sample-1 | 1000 | 25 | 5 |
| Sample-2 | 2000 | 23 | 7 |

Code on Rstudio:

```
sm1=25 # mean for sample 1
sm2=23 # mean for sample 2
sd1=5 # standard deviation of sample1
sd2=7 # standard deviation of sample2
n1=1000 # size of sample 1
n2=2000 # size of sample 2
zcal=abs((sm1-sm2)/sqrt((sd1^2/n2)+(sd2^2/n1)))
cat("Absolute value of z-calculated is =",zcal)
cat("Name & Roll No.:", "Chandana Galgali & 16010422234")
```

Output (screen shot):

The screenshot displays the RStudio interface. The script editor on the left contains the R code for testing the significance of the difference between two means. The console at the bottom shows the execution of the code, including the calculation of the z-statistic and the output of the cat() function. The Environment pane on the right shows the variables created during the execution.

```
1 sm1=25 # mean for sample 1
2 sm2=23 # mean for sample 2
3 sd1=5 # standard deviation of sample1
4 sd2=7 # standard deviation of sample2
5 n1=1000 # size of sample 1
6 n2=2000 # size of sample 2
7 zcal=abs((sm1-sm2)/sqrt((sd1^2/n2)+(sd2^2/n1)))
8 cat("Absolute value of z-calculated is =",zcal)
9 cat("Name & Roll No.:", "Chandana Galgali & 16010422234")
```

Environment

| Values | |
|--------|------------------|
| n1 | 1000 |
| n2 | 2000 |
| sd1 | 5 |
| sd2 | 7 |
| sm1 | 25 |
| sm2 | 23 |
| zcal | 8.06477838545512 |

```
> sm1=25 # mean for sample 1
> sm2=23 # mean for sample 2
> sd1=5 # standard deviation of sample1
> sd2=7 # standard deviation of sample2
> n1=1000 # size of sample 1
> n2=2000 # size of sample 2
> zcal=abs((sm1-sm2)/sqrt((sd1^2/n2)+(sd2^2/n1)))
> cat("Absolute value of z-calculated is =",zcal)
Absolute value of z-calculated is = 8.064778
> cat("Name & Roll No.:", "Chandana Galgali & 16010422234")
Name & Roll No.: Chandana Galgali & 16010422234
```

Steps of Hypothesis Testing:

1. $H_0 : \mu_1 = \mu_2$
2. $H_1 : \mu_1 \neq \mu_2$ (Nature of the test is two tailed)
3. LOS is 5%
4. Table value of Z_α is 1.96
5. Calculated value of Z : $Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}}} = 8.064778$
6. $Z_{cal} > Z_\alpha$; Therefore, H_0 is rejected
7. We can say that the difference between the population means is significant.

Q.2. The weights of eight randomly selected athletes are recorded in kilograms: 70, 75, 78, 80, 82, 85, 87, 90. The weights of twelve randomly selected basketball players are recorded in kilograms: 72, 74, 76, 78, 79, 80, 82, 83, 84, 85, 87, 88. Can it be concluded that basketball players, on average, weigh more than athletes?

Code on Rstudio:

```
x1 <- c(70, 75, 78, 80, 82, 85, 87, 90)
x2 <- c(72, 74, 76, 78, 79, 80, 82, 83, 84, 85, 87, 88)
sm1=mean(x1) # mean for sample 1
sm2=mean(x2) # mean for sample 2
sd1=sd(x1) # standard deviation of sample1
sd2=sd(x2) # standard deviation of sample2
n1=8 # size of sample 1
n2=12 # size of sample 2
tcal= abs((sm1-sm2)/sqrt(((n1*sd1^2+n2*sd2^2)/(n1+n2-2))*((1/n1)+(1/n2))))
cat("Absolute value of t-calculated is=",tcal)
cat("Name & Roll No.:", "Chandana Galgali & 16010422234")
```

Output (screen shot):

The screenshot shows the RStudio interface. The script editor on the left contains the following R code:

```
1 x1 <- c(70, 75, 78, 80, 82, 85, 87, 90)
2 x2 <- c(72, 74, 76, 78, 79, 80, 82, 83, 84, 85, 87, 88)
3 sm1=mean(x1) # mean for sample 1
4 sm2=mean(x2) # mean for sample 2
5 sd1=sd(x1) # standard deviation of sample1
6 sd2=sd(x2) # standard deviation of sample2
7 n1=8 # size of sample 1
8 n2=12 # size of sample 2
9 tcal= abs((sm1-sm2)/sqrt(((n1*sd1^2+n2*sd2^2)/(n1+n2-2))*((1/n1)+(1/n2))))
10 cat("Absolute value of t-calculated is=",tcal)
11 cat("Name & Roll No.:", "Chandana Galgali & 16010422234")
```

The console on the bottom left shows the execution output:

```
> x1 <- c(70, 75, 78, 80, 82, 85, 87, 90)
> x2 <- c(72, 74, 76, 78, 79, 80, 82, 83, 84, 85, 87, 88)
> sm1=mean(x1) # mean for sample 1
> sm2=mean(x2) # mean for sample 2
> sd1=sd(x1) # standard deviation of sample1
> sd2=sd(x2) # standard deviation of sample2
> n1=8 # size of sample 1
> n2=12 # size of sample 2
> tcal= abs((sm1-sm2)/sqrt(((n1*sd1^2+n2*sd2^2)/(n1+n2-2))*((1/n1)+(1/n2))))
> cat("Absolute value of t-calculated is=",tcal)
Absolute value of t-calculated is= 0.07582603> cat("Name & Roll No.:", "Chandana Galgali & 16010422234")
Name & Roll No.: Chandana Galgali & 16010422234
>
```

The Environment pane on the right shows the following variables:

| Variable | Value |
|----------|--|
| n1 | 8 |
| n2 | 12 |
| sd1 | 6.55607678853313 |
| sd2 | 5.06921785850339 |
| sm1 | 80.875 |
| sm2 | 80.6666666666667 |
| tcal | 0.0758260300043977 |
| x1 | num [1:8] 70 75 78 80 82 85 87 90 |
| x2 | num [1:12] 72 74 76 78 79 80 82 83 84 85 ... |

Steps of Hypothesis Testing:

1. $H_0 : \mu_1 = \mu_2$
2. $H_1 : \mu_1 < \mu_2$ (Nature of the test is one tailed)
3. LOS is 5%, $df = 8 + 12 - 2 = 18$
4. Table value of t_α is 1.734
5. Calculated value of t :
$$t_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} \times \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = 0.07582603$$
6. $t_{cal} < t_\alpha$; Therefore, H_0 is accepted (left tailed test)
7. Therefore, it can not be concluded that the basketball players, on average, weigh more than the athletes.

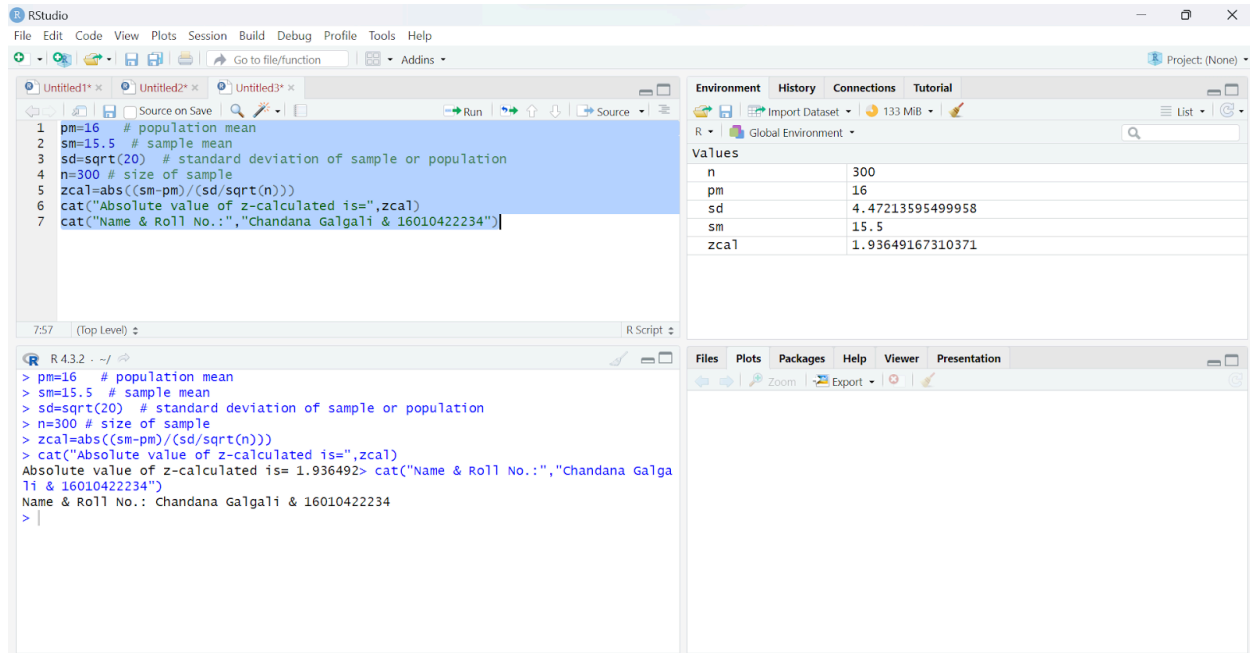
Q.3. A random sample of 300 observations has a mean of 15.5 kg. Can it be a random sample from a population whose mean is 16 kg and variance is 20 kg?

Code on Rstudio:

```
pm=16 # population mean
sm=15.5 # sample mean
sd=sqrt(20) # standard deviation of sample or population
n=300 # size of sample
zcal=abs((sm-pm)/(sd/sqrt(n)))
```

```
cat("Absolute value of z-calculated is=",zcal)
cat("Name & Roll No.:", "Chandana Galgali & 16010422234")
```

Output (screen shot):



The screenshot shows the RStudio interface. The script editor on the left contains the following R code:

```
1 pm=16 # population mean
2 sm=15.5 # sample mean
3 sd=sqrt(20) # standard deviation of sample or population
4 n=300 # size of sample
5 zcal=abs((sm-pm)/(sd/sqrt(n)))
6 cat("Absolute value of z-calculated is=",zcal)
7 cat("Name & Roll No.:", "Chandana Galgali & 16010422234")
```

The console on the bottom left shows the output of the code:

```
> pm=16 # population mean
> sm=15.5 # sample mean
> sd=sqrt(20) # standard deviation of sample or population
> n=300 # size of sample
> zcal=abs((sm-pm)/(sd/sqrt(n)))
> cat("Absolute value of z-calculated is=",zcal)
Absolute value of z-calculated is= 1.936492
> cat("Name & Roll No.:", "Chandana Galga
11 & 16010422234")
Name & Roll No.: Chandana Galgali & 16010422234
>
```

The Environment pane on the right shows the values of the variables:

| Variable | Value |
|----------|------------------|
| n | 300 |
| pm | 16 |
| sd | 4.47213595499958 |
| sm | 15.5 |
| zcal | 1.93649167310371 |

Steps of Hypothesis Testing:

1. $H_0 : \mu = 16$
2. $H_1 : \mu \neq 16$ (Nature of the test is two tailed)
3. LOS is 5%
4. Table value of Z_α is 1.96
5. Calculated value of Z : $Z_{cal} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = 1.936492$
6. $Z_{cal} < Z_\alpha$; Therefore, H_0 is accepted
7. Therefore, we can say that the sample is drawn from a population with mean 16 kg and variance 20 kg.