

Ketaki Mahajan
A – 2 / 16014022050

Tutorial 6: Sampling using R (15 / 03 / 2024)

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 | Standard Deviation |
|----------|------|------|--------------------|
| Sample 1 | 1000 | 25 | 5 |
| Sample 2 | 2000 | 23 | 7 |

Code –

```
sm1 = 25 # mean for sample 1
```

```
sm2 = 23 # mean for sample 2
```

```
sd1 = 5 # standard deviation of sample 1
```

```
sd2 = 7 # standard deviation of sample 2
```

```
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 & rollno.: ", "ketaki mahajan & 16014022050")
```

R-studio Output –

The screenshot displays the R Studio interface. The top-left pane shows the source code with line numbers 1 through 14. The top-right pane, titled 'Environment', lists the objects created: n1 (numeric, 1000), n2 (numeric, 2000), sd1 (numeric, 5), sd2 (numeric, 7), sm1 (numeric, 25), sm2 (numeric, 23), and zcal (numeric, 8.064778). The bottom-left pane shows the console output, which matches the code execution results: the absolute value of the calculated z-statistic is 8.064778, followed by the name and roll number.

```
1 sm1 = 25 # mean for sample 1
2 sm2 = 23 # mean for sample 2
3
4 sd1 = 5 # standard deviation of sample 1
5 sd2 = 7 # standard deviation of sample 2
6
7 n1 = 1000 # size of sample 1
8 n2 = 2000 # size of sample 2
9
10 zcal= abs((sm1-sm2)/sqrt((sd1^2/n2)+(sd2^2/n1)))
11
12 cat("absolute value of z-calculated is ", zcal)
13 cat("name & rollno.: ", "ketaki mahajan & 16014022050")
14
```

Console Output:

```
> sm1 = 25 # mean for sample 1
> sm2 = 23 # mean for sample 2
>
> sd1 = 5 # standard deviation of sample 1
> sd2 = 7 # standard deviation of sample 2
>
> 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 & rollno.: ", "ketaki mahajan & 16014022050")
name & rollno.: ketaki mahajan & 16014022050
>
```

Steps of Hypothesis Testing –

- $H_0: \mu_1 = \mu_2$
- $H_a: \mu_1 \neq \mu_2$ (Nature of the test is two tailed)
- LOS is 5%
- Table value of Z_α is: 1.96
- Calculated value of Z: $Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}}} = 8.064778$
- Since, $Z_{cal} > Z_\alpha$, we **reject** the null hypothesis.
- Hence, to conclude, we can say that the **difference between the population means is significant.**

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 –

```
# athletes weights (sample 1)
x1 <- c(70, 75, 78, 80, 82, 85, 87, 90)

# basketball player weights (sample 2)
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 sample 1
sd2 = sd(x2) # standard deviation of sample 2

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 & rollno.: ", "ketaki mahajan & 16014022050")
```

R-Studio Output –

The screenshot shows the R-Studio interface. The top-left pane contains the R script code for a two-sample t-test. The top-right pane shows the Environment window with variables n1, n2, sd1, sd2, sm1, sm2, tcal, x1, and x2. The bottom-left pane shows the Console output, which includes the execution of the code and the calculated t-value.

```

1 # athletes weights (sample 1)
2 x1 <- c(70, 75, 78, 80, 82, 85, 87, 90)
3
4 # basketball player weights (sample 2)
5 x2 <- c(72, 74, 76, 78, 79, 80, 82, 83, 84, 85, 87, 88)
6
7 sm1 = mean(x1) # mean for sample 1
8 sm2 = mean(x2) # mean for sample 2
9
10 sd1 = sd(x1) # standard deviation of sample 1
11 sd2 = sd(x2) # standard deviation of sample 2
12
13 n1 = 8 # size of sample 1
14 n2 = 12 # size of sample 2
15
16 tcal = abs((sm1-sm2)/sqrt(((n1*sd1^2+n2*sd2^2)/(n1+n2-2))*((1/n1)+(1/n2))))
17
18 cat("absolute value of t-calculated is ", tcal)
19 cat("name & rollno.: ", "ketaki mahajan & 16014022050")
20

```

Environment Window:

| Name | Type | Length | Size | Value |
|------|---------|--------|-------|---------------------------------------|
| n1 | numeric | 1 | 56 B | 8 |
| n2 | numeric | 1 | 56 B | 12 |
| sd1 | numeric | 1 | 56 B | 6.55607678853313 |
| sd2 | numeric | 1 | 56 B | 5.06921785850339 |
| sm1 | numeric | 1 | 56 B | 80.875 |
| sm2 | numeric | 1 | 56 B | 80.6666666666667 |
| tcal | numeric | 1 | 56 B | 0.0758260300043977 |
| x1 | numeric | 8 | 112 B | num [1:8] 70 75 78 80 82 85 87 90 |
| x2 | numeric | 12 | 176 B | num [1:12] 72 74 76 78 79 80 82 83... |

Console Output:

```

> # athletes weights (sample 1)
> x1 <- c(70, 75, 78, 80, 82, 85, 87, 90)
>
> # basketball player weights (sample 2)
> 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 sample 1
> sd2 = sd(x2) # standard deviation of sample 2
>
> 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 & rollno.: ", "ketaki mahajan
& 16014022050")
name & rollno.: ketaki mahajan & 16014022050

```

Steps of Hypothesis Testing –

- $H_0: \mu_1 = \mu_2$
- $H_a: \mu_1 < \mu_2$ (Nature of the test is one tailed)
- LOS is assumed as 5%
- $DOF = 8 + 12 - 2 = 18$
- Table value of t_α is: 1.7341
- Calculated value of t: $t_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} \times \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = 0.07582603$

- Since, $t_{cal} < t_\alpha$, we **accept** the null hypothesis.
- Hence, to conclude, there is **no evidence** that says basketball players, on average, weigh more than athletes.

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 are 20 kg?

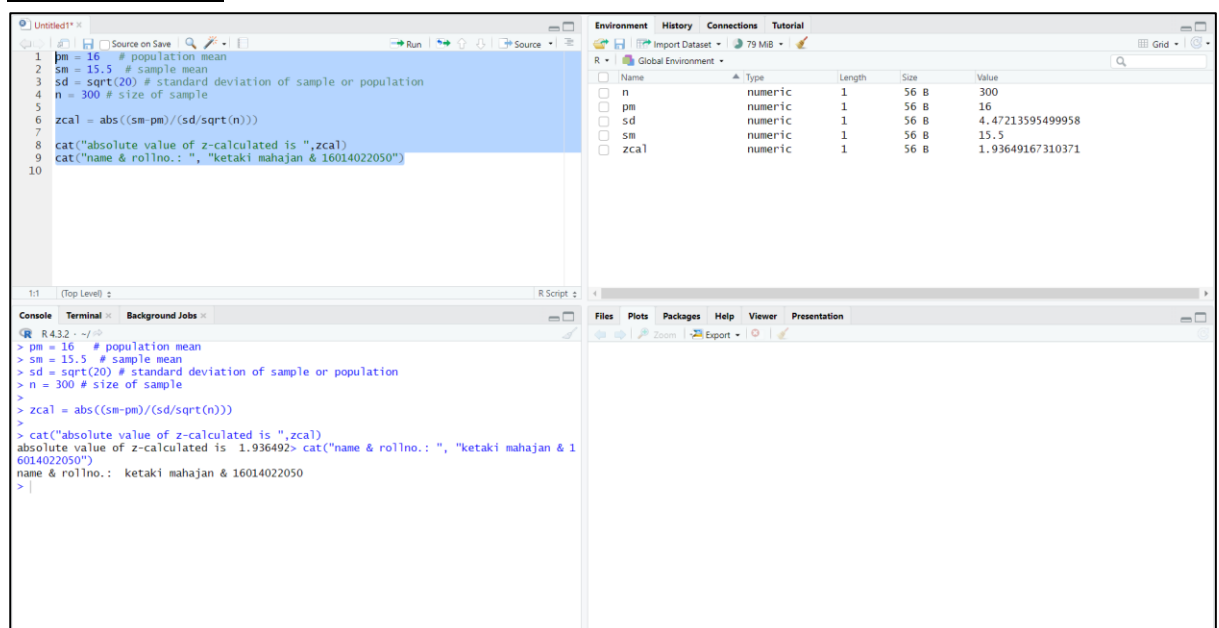
Code –

```
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)
cat("name & rollno.: ", "ketaki mahajan & 16014022050")
```

R-Studio Output –



Steps of Hypothesis Testing –

- $H_0: \mu = 16$
- $H_a: \mu \neq 16$ (Nature of the test is two tailed)
- LOS is 5%
- Table value of Z_α is: 1.96
- Calculated value of Z: $Z_{cal} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = 1.936492$
- Since $Z_{cal} < Z_\alpha$, we **accept** the null hypothesis.
- Hence, to conclude, we can say that the **sample is drawn from a population** with mean 16.