Hypothesis Testing Concept

Author : Kurnia Indah Putri Emzar

INTRODUCTION

Hypothesis testing concept in R programming is a process of testing the hypothesis made by the researcher or to validate the hypothesis. To perform hypothesis testing, a random sample of data from the population is taken and testing is performed. Based on the results of testing, the hypothesis is either selected or rejected. This concept is known as **Statistical Inference**.

First, we start by making dataset. The data on the weight of 10 rat samples were given.

In :

```
##1. Make a dataset
set.seed(1234)
my_data <- data.frame( name = paste0(rep("M_", 10), 1:10), weight = round(rnorm(10, 20, 2), 1) )
```

Output :

	name	weight
1	M_1	17.6
2	M_2	20.6
3	M_3	22.2
4	M_4	15.3
5	M_5	20.9
6	M_6	21.0
7	M_7	18.9
8	M_8	18.9
9	M_9	18.9
10	M_10	18.2

STEP 1 : FIND THE SUMMARY STATISTIC OF DATA

After we make a dataset, the next step we find the summary statistic on the weight of 10 rat samples.

In :

```
##2. Find summary statistik
summary(my_data$weight)
```

Output :

We have the Output from data :

```
> summary(my_data$weight)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  15.30   18.38   18.90   19.25   20.82   22.20 
```

We got from summary :

MEAN FROM DATA = 19.25 kg

MEDIAN = 18.90 kg

STEP 2 : BOXPLOT VISUALIZATION

Before we check the hypothesis, we need to visualize data into boxplot.

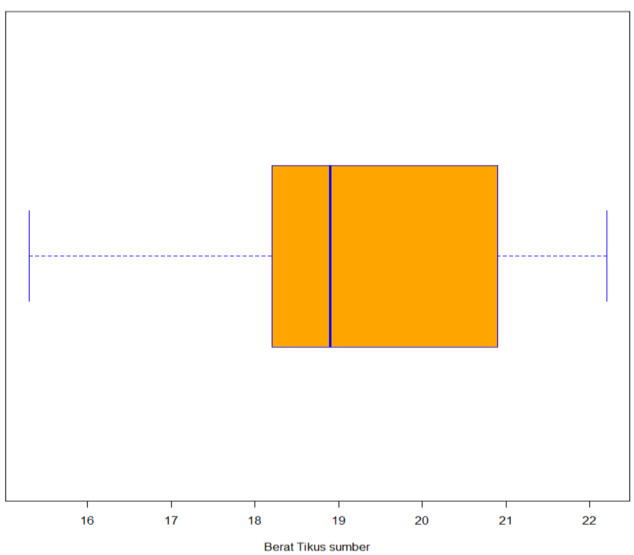
In :

```
hist(my_data$weight,
        main = "Berat tikus percobaan


                     ",
        xlab = "Berat Tikus sumber",
        col = "orange",
        border = "blue",
        freq = FALSE
)
```

Output :

We got the Output :



The test result shows the value difference between mean and median is not too far away, there are no outliers and also from the histogram and boxplot graphs that the data distribution is close to symmetrical skew.

STEP 3 : ONE SAMPLE T-TEST

A one-sample t-test checks whether a sample mean differs from the population mean. Let's create some dummy age data for the population of voters in the entire country and a sample of voters in Minnesota and test the whether the average age of voters Minnesota differs from the population.

In :

```
test <- t.test(my_data$weight, mu = 25)
test

p_value <- test$p.value
print(p_value < 0.05)
```

or we can use another way :

```
t <- test$statistic
t_table_1 <- qt(p=0.025, df=11)
t_table_2 <- qt(p=0.975, df=11)

print((t< t_table_1) | (t> t_table_2))
```

Output :

1.

Name	Type	Value
test	list [10] (S3: htest)	List of length 10
statistic	double [1]	-9.078319
parameter	double [1]	9
p.value	double [1]	7.953383e-06
conf.int	double [2]	17.8 20.7
estimate	double [1]	19.25
null.value	double [1]	25
stderr	double [1]	0.6333772
alternative	character [1]	'two.sided'
method	character [1]	'One Sample t-test'
data.name	character [1]	'my_data$weight'

2.

```
        One Sample t-test

data:  my_data$weight
t = -9.0783, df = 9, p-value = 7.953e-06
alternative hypothesis: true mean is not equal to 25
95 percent confidence interval:
 17.8172 20.6828
sample estimates:
mean of x
    19.25
```

STEP 4 : INTERPRET RESULTS

The test result shows the test statistic "t" is equal to -9.0783. This test statistic tells us how much the sample mean deviates from the null hypothesis. If the t-statistic lies outside the quantiles of the t-distribution corresponding to our confidence level and degrees of freedom, we reject the null hypothesis.
We can see from the output, p.value < 0.05, so we can **reject the-null hypothesis**. So, mean of sample from experiment is 25 kg and from the t-test we get 19.25 kg. Mean value in t-test is not equal with mean value from experiment.