Practical no 4

Aim: Demonstration of Hypothesis testing

Theory:

Hypothesis testing is a form of statistical inference that uses data from a sample to draw conclusions about a

population parameter or a population probability distribution.

First, a tentative assumption is made about the parameter or distribution. This assumption is called the null

hypothesis and is denoted by H0. An alternative hypothesis (denoted Ha), which is the opposite of what is

stated in the null hypothesis, is then defined. The hypothesis-testing procedure involves using sample data to

determine whether or not H0 can be rejected. If H0 is rejected, the statistical conclusion is that the alternative

hypothesis Ha is true.

For example, assume that a radio station selects the music it plays based on the assumption that the average

age of its listening audience is 30 years. To determine whether this assumption is valid, a hypothesis test could

be conducted with the null hypothesis given as H0: $\mu = 30$ and the alternative hypothesis given as Ha: $\mu \neq 30$.

Based on a sample of individuals from the listening audience, the sample mean age, $\bar{x}$, can be computed and

used to determine whether there is sufficient statistical evidence to reject H0. Conceptually, a value of the

sample mean that is “close” to 30 is consistent with the null hypothesis, while a value of the sample means that

is “not close” to 30 provides support for the alternative hypothesis. What is considered “close” and “not close”

is determined by using the sampling distribution of $\bar{x}$.

Ideally, the hypothesis-testing procedure leads to the acceptance of H0 when H0 is true and the rejection of

H0 when H0 is false. Unfortunately, since hypothesis tests are based on sample information, the possibility of

errors must be considered. A type I error corresponds to rejecting H0 when H0 is actually true, and a type II

error corresponds to accepting H0 when H0 is false. The probability of making a type I error is denoted by α,

and the probability of making a type II error is denoted by β.

Steps:

- Open Excel create a data
- Save it as .CSV(MS-DOS)
- Keep the dataset and R code in a same folder.

Dataset:

Sr no	C1	Deviation	Deviation Square
1	85.3	-12.2214	149.3633
2	86.9	-10.6214	112.8147
3	96.8	-0.72143	0.520459
4	101.1667	3.645238	-61.2767
5	106.9167	9.395238	-135.698
6	112.6667	15.14524	-210.12
7	118.4167	20.89524	-284.541
8	124.1667	26.64524	-358.962
9	129.9167	32.39524	-433.384
10	135.6667	38.14524	-507.805
11	141.4167	43.89524	-582.227
12	147.1667	49.64524	-656.648
13	152.9167	55.39524	-731.07
14	158.6667	61.14524	-805.491
15	164.4167	66.89524	-879.912
16	170.1667	72.64524	-954.334
17	175.9167	78.39524	-1028.76
18	181.6667	84.14524	-1103.18
19	187.4167	89.89524	-1177.6
20	193.1667	95.64524	-1252.02

Code:

```
datanew=read.csv("C:/Users/admin/Downloads/Desktop/Materials/COMPUTER SCIENCE/Sem 6/Data
datanew
boxplot(datanew)
m1=mean(datanew$C1)
m1
sd1=sd(datanew$C1)
sd1
plot(datanew$C1)
t.test(datanew$C1,alternative="greater",mu=100)
```

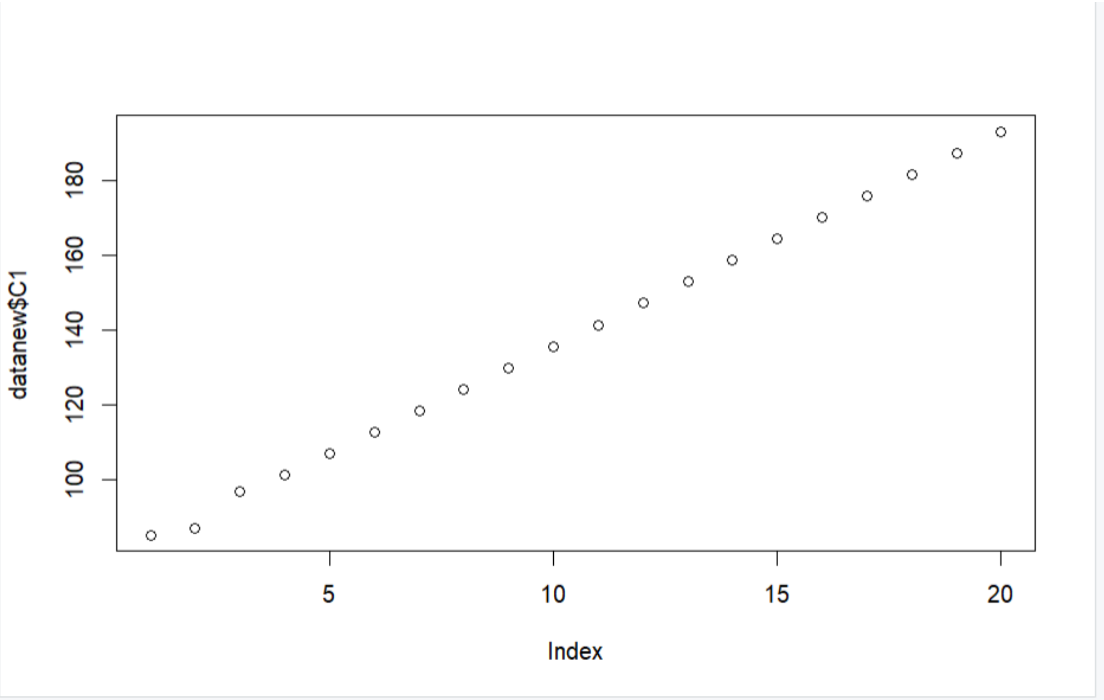
Output:

```
> datanew
   Sr.no       C1   Deviation Deviation.Square
1      1  85.3000 -12.2214286      149.3633163
2      2  86.9000 -10.6214280      112.8147449
3      3  96.8000  -0.7214285        0.5204592
4      4 101.1667   3.6452384      -61.2766837
5      5 106.9167   9.3952384     -135.6981122
6      6 112.6667  15.1452385     -210.1195408
7      7 118.4167  20.8952385     -284.5409693
8      8 124.1667  26.6452385     -358.9623979
9      9 129.9167  32.3952386     -433.3838264
10    10 135.6667  38.1452386     -507.8052550
11    11 141.4167  43.8952386     -582.2266836
12    12 147.1667  49.6452387     -656.6481121
13    13 152.9167  55.3952387     -731.0695407
14    14 158.6667  61.1452387     -805.4909692
15    15 164.4167  66.8952388     -879.9123978
16    16 170.1667  72.6452388     -954.3338264
17    17 175.9167  78.3952388    -1028.7552550
18    18 181.6667  84.1452389    -1103.1766830
19    19 187.4167  89.8952389    -1177.5981120
20    20 193.1667  95.6452389    -1252.0195410
> boxplot(datanew)
> m1=mean(datanew$C1)
> m1
[1] 138.5417
> sd1=sd(datanew$C1)
> sd1
[1] 34.02634
> plot(datanew$C1)
> t.test(datanew$C1,alternative="greater",mu=100)

        One Sample t-test

data:  datanew$C1
t = 5.0656, df = 19, p-value = 3.434e-05
alternative hypothesis: true mean is greater than 100
95 percent confidence interval:
 125.3855      Inf
sample estimates:
mean of x
 138.5417

>
```



Conclusion: Hence we have successfully learnt & performed Hypothesis testing.