Name – Yash Daga Roll Number – 20BCE7323

Lab-10

Q.1) In a certain type of metal test specimen, the normal stress on a specimen is known to befunctionally related to the shear resistance. The following is a set of coded experimental data onthe two variables:

Normal Stress, x	Shear Resistance, y
26.8	26.5
25.4	27.3
28.9	24.2
23.6	27.1
27.7	23.6
23.9	25.9
24.7	26.3
28.1	22.5
26.9	21.7
27.4	21.4
22.6	25.8
25.6	24.9

•Estimate the shear resistance for a normal stress of 24.5.

•Plot the data; does it appear that a simple linear regression will be a suitable model?

Ans

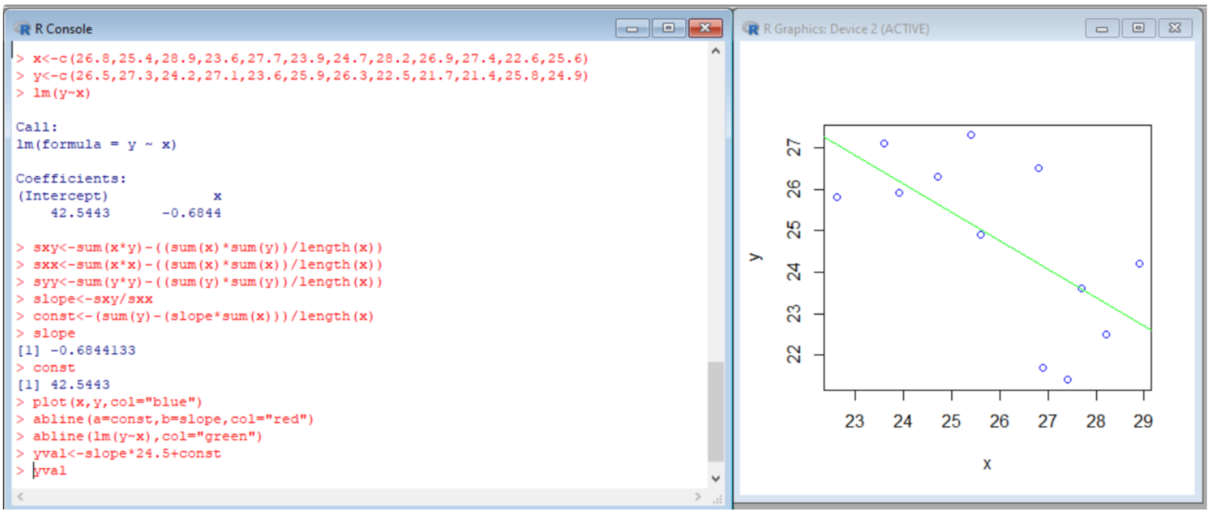
```
> x<-c(26.8,25.4,28.9,23.6,27.7,23.9,24.7,28.2,26.9,27.4,22.6,25.6)

> y<-c(26.5,27.3,24.2,27.1,23.6,25.9,26.3,22.5,21.7,21.4,25.8,24.9)

> lm(y~x)


Call:

lm(formula = y ~ x)
```

Coefficients:

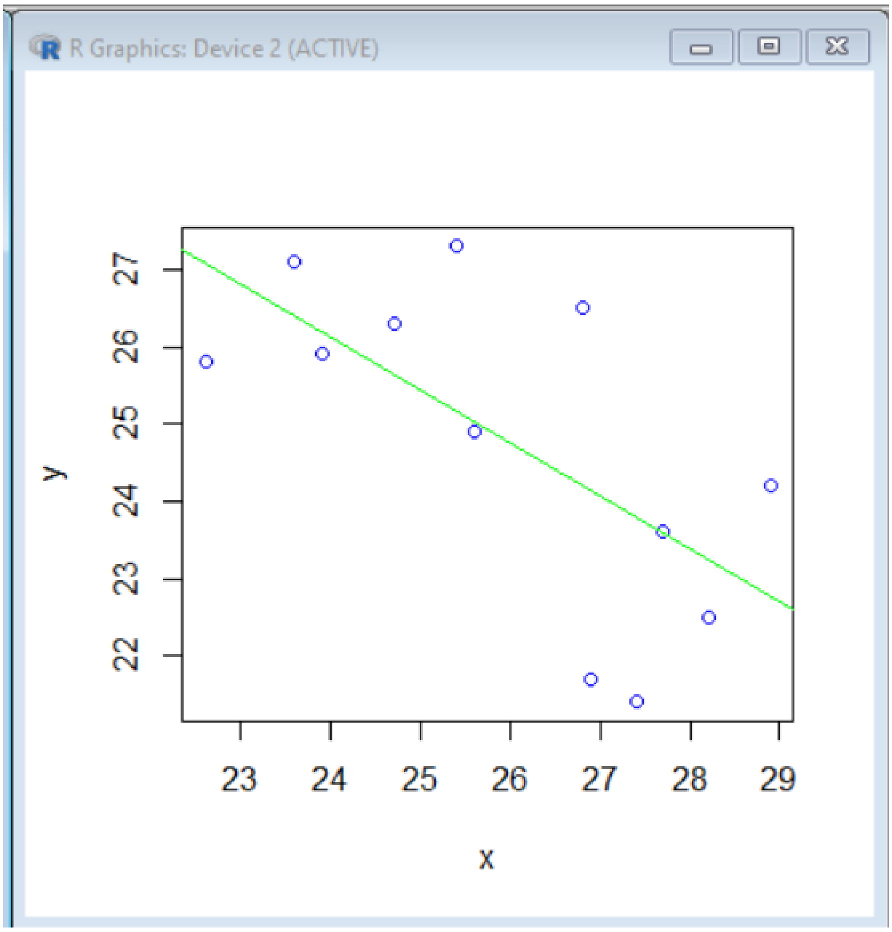
(Intercept)   x

42.5443   -0.6844

```
> sxy<-sum(x*y)-((sum(x)*sum(y))/length(x))
> sxx<-sum(x*x)-((sum(x)*sum(x))/length(x))
> syy<-sum(y*y)-((sum(y)*sum(y))/length(x))
> slope<-sxy/sxx
> const<-(sum(y)-(slope*sum(x)))/length(x)
> slope
[1] -0.6844133
> const
[1] 42.5443
> plot(x,y,col="blue")
> abline(a=const,b=slope,col="red")
> abline(lm(y~x),col="green")
> yval<-slope*24.5+const
> yval
[1] 25.77618
```



```
> x<-c(26.8,25.4,28.9,23.6,27.7,23.9,24.7,28.2,26.9,27.4,22.6,25.6)
> y<-c(26.5,27.3,24.2,27.1,23.6,25.9,26.3,22.5,21.7,21.4,25.8,24.9)
> lm(y~x)

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
    42.5443      -0.6844

> sxy<-sum(x*y)-((sum(x)*sum(y))/length(x))
> sxx<-sum(x*x)-((sum(x)*sum(x))/length(x))
> syy<-sum(y*y)-((sum(y)*sum(y))/length(x))
> slope<-sxy/sxx
> const<-(sum(y)-(slope*sum(x)))/length(x)
> slope
[1] -0.6844133
> const
[1] 42.5443
> plot(x,y,col="blue")
> abline(a=const,b=slope,col="red")
> abline(lm(y~x),col="green")
> yval<-slope*24.5+const
> yval
```

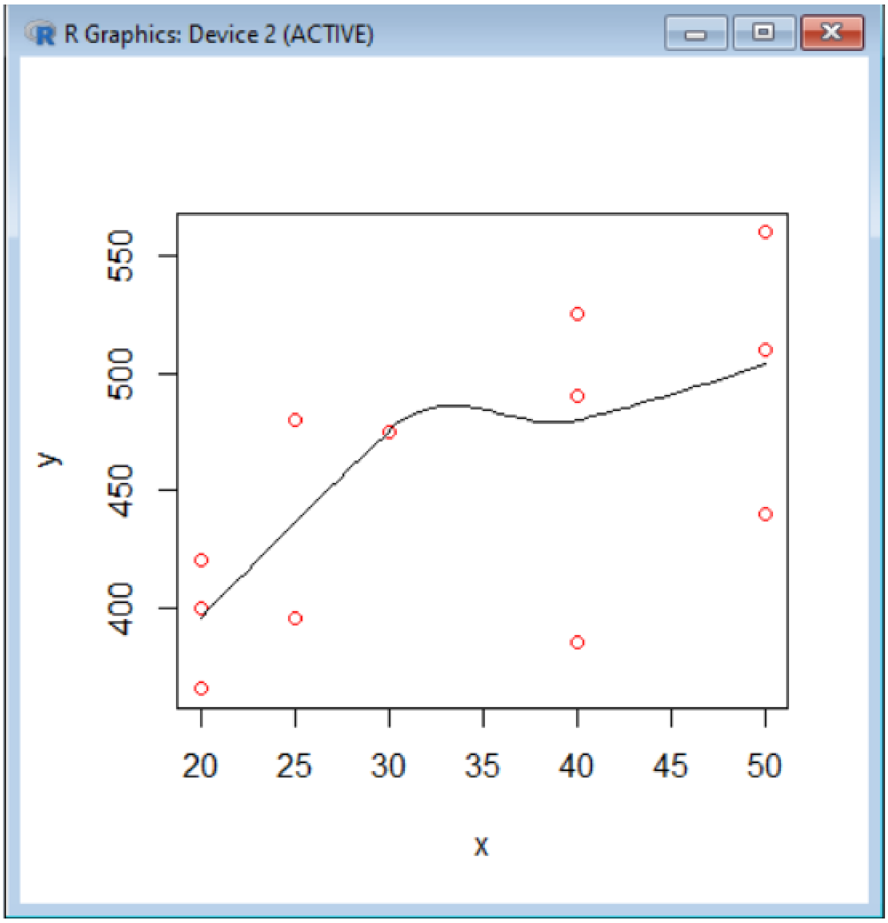


Q.2) A study was made by a retail merchant to determine the relation between weekly advertisingexpenditures and sales.(a)Plot a scatter diagram.(b)Find the equation of the regression line to predict weekly sales from advertising expenditures.

Ans:

Advertising Costs ($)	Sales ($)
40	385
20	400
25	395
20	365
30	475
50	440
40	490
20	420
50	560
40	525
25	480
50	510

```
> scatter.smooth(x,y,col="red")
> x<- c(40,20,25,20,30,50,40,20,50,40,25,50)
> y<- c(385,400,395,365,475,440,490,420,560,525,480,510)
> sx<- sum(x)
> sy<-sum(y)
> sxy<- sum(x*y)
> sx2<-sum(x^2)
> lx<-length(x)
> ly<-length(y)
> lxy<-length(x*y)
> b<-(lx*sxy-sx*sy)/(lx*sx2-sx^2)
> b
[1] 3.220812
> a<-(sy-b*sx)/lx
> a
[1] 343.7056
> scatter.smooth(x,y,col="red")
> 
```



Code:

```
> x<- c(40,20,25,20,30,50,40,20,50,40,25,50)

> y<- c(385,400,395,365,475,440,490,420,560,525,480,510)

> sx<- sum(x)

> sy<-sum(y)

> sxy<- sum(x*y)

> sx2<-sum(x^2)

> lx<-length(x)

> ly<-length(y)

> lxy<-length(x*y)

> b<-(lx*sxy-sx*sy)/(lx*sx2-sx^2)

> b

[1] 3.220812
```

```
> a<-(sy-b*sx)/lx
> a
[1] 343.7056
> scatter.smooth(x,y,col="red")
```

Definitions

1) Random Sampling and Probability: Random sampling, or probability sampling, is a sampling method that allows for the randomization of sample selection, i.e., each sample has the same probability as other samples to be selected to serve as a representation of an entire population.
2) Binomial distribution: A binomial distribution can be thought of as simply the probability of a SUCCESS or FAILURE outcome in an experiment or survey that is repeated multiple times.
3) Poisson Distribution: Poisson distribution is a probability distribution that can be used to show how many times an event is likely to occur within a specified period of time.
4) Normal Distribution: probability distribution that is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graph form, normal distribution will appear as a bell curve.
5) Test of Hypothesis-z test: Z-test is a statistical test to determine whether two population means are different when the variances are known and the sample size is large. Z-test is a hypothesis test in which the **z**-statistic follows a normal distribution. A **z**-statistic, or **z**-score, is a number representing the result from the z-test.
6) Test of Hypothesis - T test: A **t**-test is a statistical test that compares the means of two samples. It is used in hypothesis testing, with a null hypothesis that the difference in group means is zero and an alternate hypothesis that the difference in group means is different from zero.
7) Correlation and Regression: Correlation quantifies the strength of the linear relationship between a pair of variables, whereas regression expresses the relationship in the form of an equation.

Application Oriented Problem

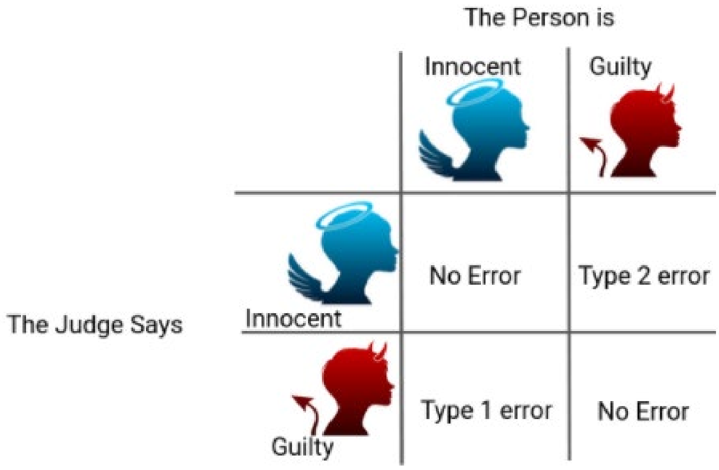
Applications of z-test and t-test problems

- **Fundamentals of Hypothesis Testing**
  - Basic Concepts – Null Hypothesis, Alternative Hypothesis, Type 1 Error, Type 2 Error, and Significance Level
  - Steps to Perform Hypothesis Testing
  - Directional Hypothesis
  - Non-Directional Hypothesis Test
- **What is the Z Test?**
  - One-Sample Z Test
  - Two Sample Z Test
- **What is the t-Test?**
  - One-Sample t-Test
  - Two-Sample t-Test
- **Deciding between the Z Test and t-Test**

Concept behind Hypothesis:

Example: A person is on trial for a criminal offense and the judge needs to provide a verdict on his case. Now, there are four possible combinations in such a case:

- First Case: The person is innocent and the judge identifies the person as innocent
- Second Case: The person is innocent and the judge identifies the person as guilty
- Third Case: The person is guilty and the judge identifies the person as innocent
- Fourth Case: The person is guilty and the judge identifies the person as guilty



Its is just an act where the analysts test their judgement.

Sample for z-test

If girls on average score higher than 600 in the exam. We have the information that the standard deviation for girls' scores is 100. So, we collect the data of 20 girls by using random samples and record their marks. Finally, we also set our α value (significance level) to be 0.05.

Score
650
730
510
670
480
800
690
530
590
620
710
670
640
780
650
490
800
600
510
700

- Mean Score for Girls is 641
- The size of the sample is 20
- The population mean is 600
- Standard Deviation for Population is 100

z score $= \dfrac{\overline{x} - \mu}{\sigma / \sqrt{n}}$

$= \dfrac{641 - 600}{100/\sqrt{20}}$

$= 1.8336$

p value $= .033357.$

Critical Value = 1.645

Z score > Critical Value

P value < 0.05

```
> y=c(650,730,510,670,480,800,690,530,590,620,710,670,640,780,650,490,800,600,5
> z.test(y,sigma.y=1)
```

```
        One Sample  -test

data:  y
t = 24.17, df = 19, p-value = 4.965e-16
alternative hypothesis: true mean is greater than 100
95 percent confidence interval:
 602.2968      Inf
sample estimates:
mean of x
      641

>
```

H1: mean> 600 is true

Reject Null-hypothesis cause p-value is less than 0.05

Sample for t-test

Let's say we want to determine if on average girls score more than 600 in the exam. We do not have the information related to variance (or standard deviation) for girls' scores. To a perform t-test, we randomly collect the data of 10 girls with their marks and choose our α value (significance level) to be 0.05 for Hypothesis Testing

Girls_Score
587
602
627
610
619
622
605
608
596
592

- Mean Score for Girls is 606.8
- The size of the sample is 10
- The population mean is 600
- Standard Deviation for the sample is 13.14

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

$$= \frac{606.8 - 600}{13.14/\sqrt{10}}$$

$$= 1.64$$

Critical Value = 1.833

t score < Critical Value

P value = 0.0678

P value > 0.05

$H_0 : \mu \leq 600$

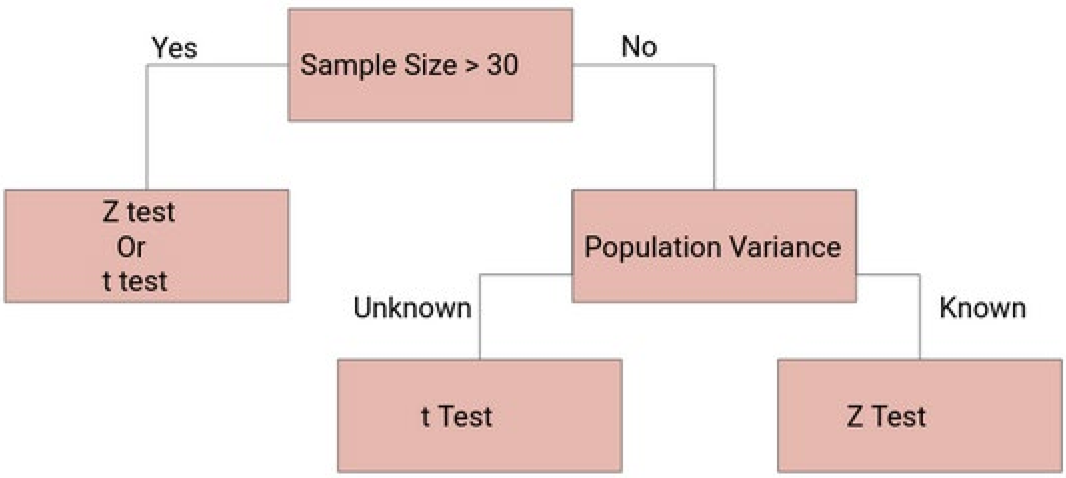
$\boxed{H_1 : \mu > 600}$

P-value is greater than 0.05 thus fail to reject the null hypothesis

```
> x=c(587,602,627,610,619,622,605,608,596,592)
> t.test(x,alternative="greater",mu=13.4)

        One Sample t-test

data:  x
t = 142.82, df = 9, p-value < 2.2e-16
alternative hypothesis: true mean is greater than 13.4
95 percent confidence interval:
 599.1838      Inf
sample estimates:
mean of x
    606.8

> |
```

Deciding between Z Test and T-Test

So when we should perform the Z test and when we should perform t-Test? It's a key question we need to answer if we want to master statistics.



If the sample size is large enough, then the Z test and t-Test will conclude with the same results. For a **large sample size**, **Sample Variance will be a better estimate** of Population variance so even if population variance is unknown, we can **use the Z test using sample variance.**

Similarly, for a **Large Sample**, we have a high degree of freedom. And since **t-distribution approaches the normal distribution**, **t**he difference between the z score and t score is negligible.

Link :[Application of my Problem]