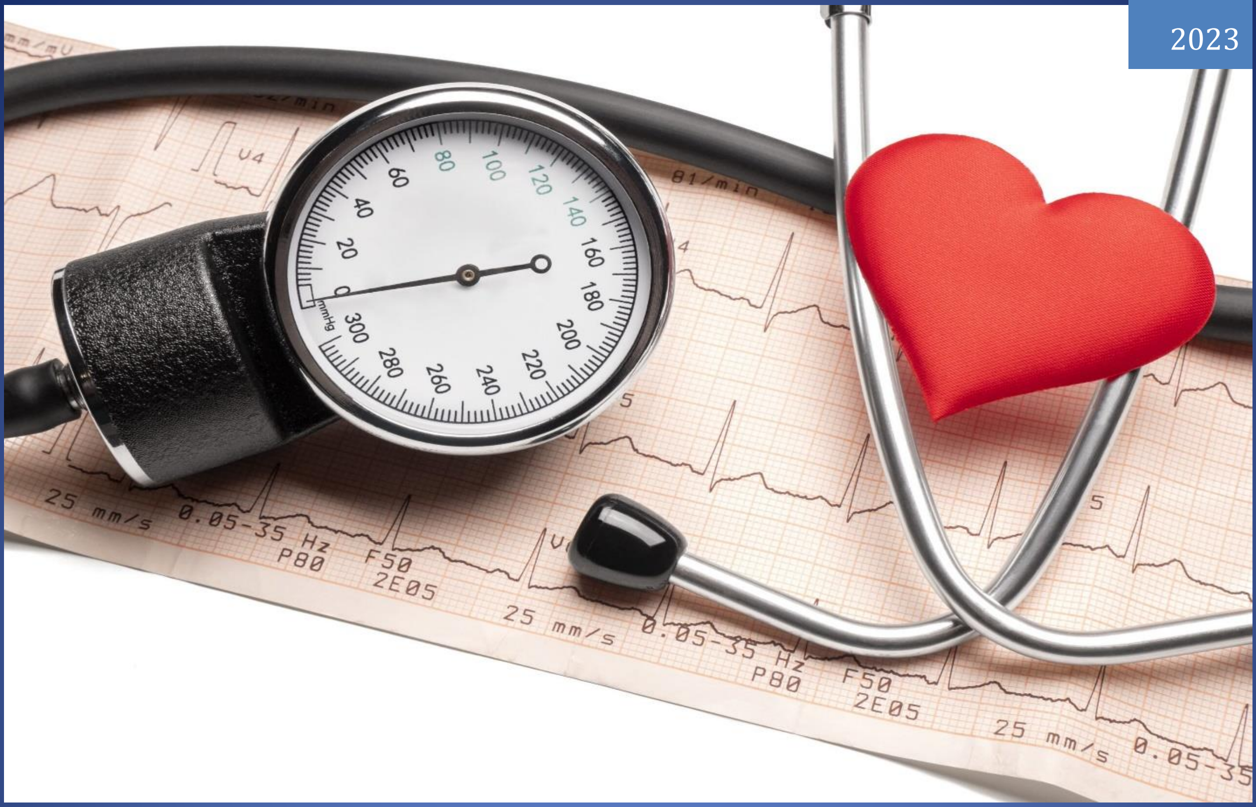
2023



Weight Vs Systolic Blood Pressure

SIMPLE LINEAR REGRESSION – LAB1

A.MAYURI(2348133)

Weight (Pounds) Vs Systolic Blood Pressure (mmHg)

A.Mayuri(2348133)

2023-11-03

Question:

The weight and systolic blood pressure of 26 randomly selected male in the age group 25–30 are shown below.

Objective:

1.Obtain the scatter plot and interpret it.

2. Find a regression line connecting the variables systolic pressure and weight. Interpret the plot, intercept term and regression coefficients. What is the inference from the sign of regression coefficient?

3. Obtain the fitted values. Does the sum of fitted values is equal to the sum of observed values ?

Based on the above questions, prepare a report with introduction, analysis and conclusions.

Introduction:

We are given a dataset of size 26 and this encloses adults of age 25-30. we are given information about their respective weights and systolic BP. We are developing a simple linear regression model which aims to understand the linear relationship between the dependent and the parameters. Linear regression is a wise choice to build a model as we have only two variables as our interest can most optimally have a relationship between them. However, we are proceeding with the analysis to verify and understand the same.

Variables of Interest:

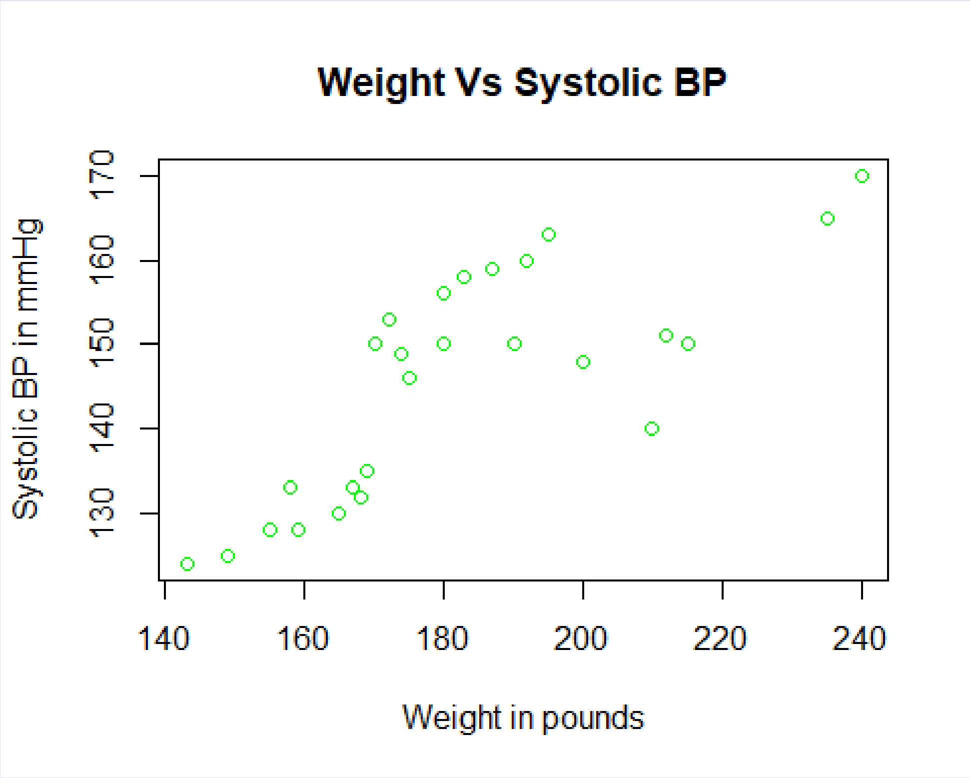
The independent (x) variable is that of the adults weight in pounds and the dependent (y) variable is their systolic BP in mmHg.

Step1: Import Dataset

```
library(readxl)
Lab11 <- read_excel("C:/Users/mayur/Desktop/Mstat/Semesters/Tri-sem2/Regression/Dataset/Lab11.xlsx")
View(Lab11)
attach(Lab11)
```

Step : Scatter Plot

```
plot(Lab11$Weight,Lab11$`Symbolic BP`,col='green',main='Weight Vs Systolic BP',xlab='Weight in pounds',ylab = 'Systolic BP in mmHg')
```



Observation:

We observe that there exist a positive linear relation between the 2 variables of interest now we will try to understand the magnitude of the same relationship.

Step3: Correlation

```
cor(Lab11$Weight,Lab11$`Symbolic BP`)

## [1] 0.7734903
```

Observation:

The correlation coefficient between the two variables is obtained as 0.7734 which is <0.8. Hence the variables have a moderately positive correlation between them.

Step4: Model

```
model=lm(Lab11$`Symbolic BP`~Lab11$Weight)
summary(model)
```

```
##
## Call:
## lm(formula = Lab11$`Symbolic BP` ~ Lab11$Weight)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -17.182  -6.485  -2.519   8.926  12.143
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  69.10437   12.91013   5.353 1.71e-05 ***
## Lab11$Weight  0.41942    0.07015   5.979 3.59e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.681 on 24 degrees of freedom
## Multiple R-squared:  0.5983, Adjusted R-squared:  0.5815
## F-statistic: 35.74 on 1 and 24 DF,  p-value: 3.591e-06
```

Interpretation:

if Y=B1+B2(x) is the model, then from the above table we observe that 69.10437 is the intercept and the B2 value(slope value) is 0.41942. ie, for a unit change in weight there will be 0.41942 mmHg change in BP. Also the positive B2 (Cov(x,y)/Sd(x)*Sd(y)) value shows that there is a positive relationship between the independent variable and the dependent variable.Since the only way for B2 to be positive is when cov(x,y) is positive.

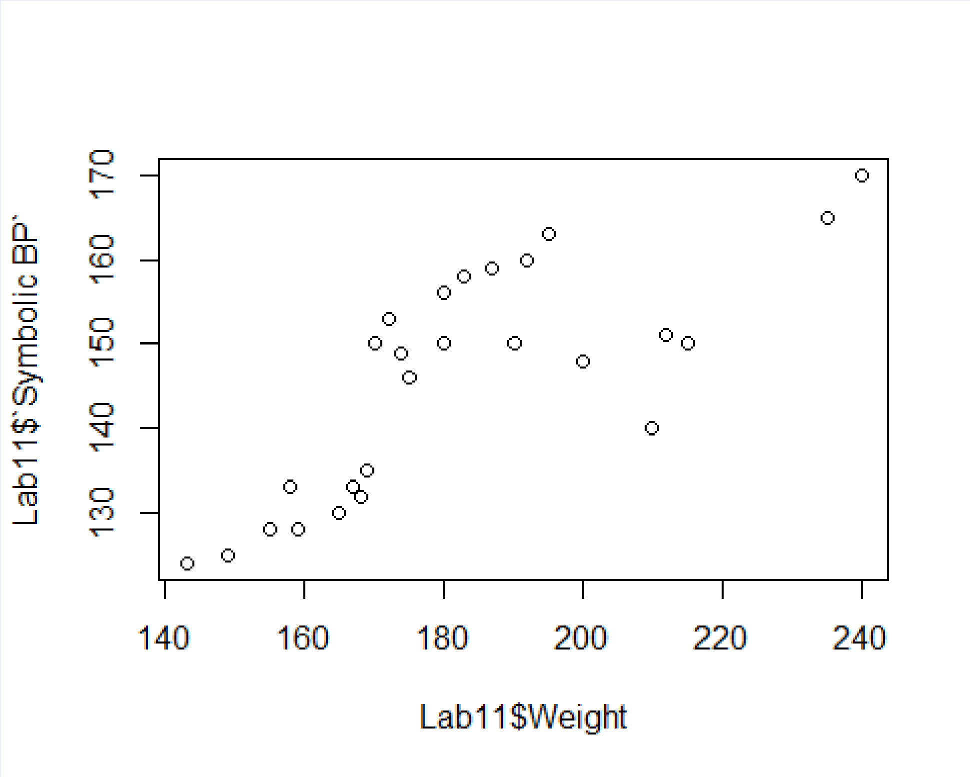
Step5: Fitting Model

```
Fit=fitted.values(model)
Fit
```

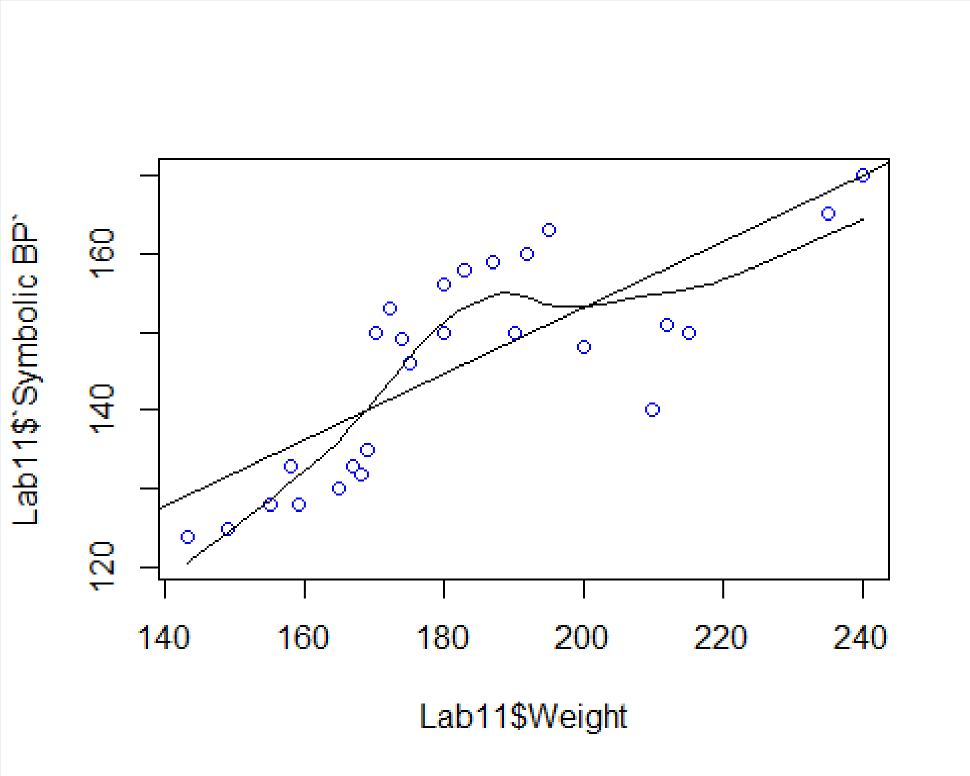
```
##        1        2        3        4        5        6        7        8
## 138.3079 139.1467 144.5991 134.1137 158.0204 142.5020 148.7933 157.1816
##        9       10       11       12       13       14       15       16
## 152.9874 131.5972 135.3720 139.9855 140.4050 141.2438 135.7914 139.5661
##       17       18       19       20       21       22       23       24
## 142.0826 145.8574 159.2786 150.8903 144.5991 129.0807 169.7640 167.6669
##       25       26
## 149.6321 147.5350
```

Step6 : Fitted Values VS Observed Values

```
plot(Lab11$Weight,Lab11$`Symbolic BP`)
```




```
scatter.smooth(Lab11$Weight,Lab11$`Symbolic BP`,col='blue')
#using the abline command
abline(lm(Lab11$`Symbolic BP`~Lab11$Weight))
```



Interpretation:

The graph above is the best possible fit for the given data set. It has minimized the sum of square of errors for the observed data set.ie the independent variable - the systolic Blood pressure.

Step7: To Verify if sum of observed value equals Fitted Values.

```
M=data.frame(Fit)
M

##         Fit
## 1  138.3079
## 2  139.1467
## 3  144.5991
## 4  134.1137
## 5  158.0204
## 6  142.5020
## 7  148.7933
## 8  157.1816
## 9  152.9874
## 10 131.5972
## 11 135.3720
## 12 139.9855
```

```
## 13 140.4050
## 14 141.2438
## 15 135.7914
## 16 139.5661
## 17 142.0826
## 18 145.8574
## 19 159.2786
## 20 150.8903
## 21 144.5991
## 22 129.0807
## 23 169.7640
## 24 167.6669
## 25 149.6321
## 26 147.5350
```

```
sum(M)
```

```
## [1] 3786
```

```
sum(Lab11$`Symbolic BP`)
```

```
## [1] 3786
```

Observation:

Here we observe that the sum of systolic pressure under fitted values is same as the observed/recorded values. Hence verified.

Conclusion:

From the above exercise we can conclude the following there is a moderate linear positive relationship between weight and systolic BP.

With a correlation coefficient of 0.77.

The model is of the form: y=B1+B2*X +E where B1=69.10437 B2=0.41942

Ie, Y=69.10437+0.41942*X+E

Also the estimates are significant from the above table .

The positive sign of B2 in an indication that the variables have a positive relationship.

The sum of fitted values is equal to sum of observed values.

ie, an increase in weight can lead to Blood pressure issues. On the contrary there are times where chronic increase in blood pressure due to lack of iodine/salt can push up the weight of a person significantly. This can also be verified by clinical analysis along with regression analysis.