Simple Linear Regression

***INTRODUCTION***

In this exercise we are provided with the dataset that consist of weight and systolic blood pressure of 26 randomly selected males in the age group 25–30.

In our dataset, weight of the males of age group 25-30 is independent variable and systolic blood pressure of male is dependent variable.

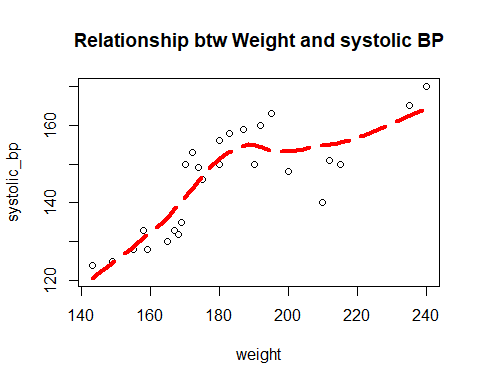
Here we are interested in

1. Obtaining a scatter plot and interpreting it.
2. Finding a regression line connecting the variables systolic pressure and weight and interpreting the plot, intercept term and regression coefficients.
3. Obtaining the fitted values and to check if sum of fitted values is equal to the sum of observed values.
4. Obtaining the residuals and interpreting it.

***ANALYSIS***

*#Assigning the weight and systolic blood pressure of 26 male of age group 25-30 to the respective variables.*  
weight=**c**(165,167,180,155,212,175,190,210,200,149,158,169,170,172,159,168,174,183,215,195,180,143,240,235,192,187)  
systolic\_bp=**c**(130,133,150,128,151,146,150,140,148,125,133,135,150,153,128,132,149,158,150,163,156,124,170,165,160,159)

*#Obtaining the scatter plot for the two variables weight and systolic blood pressure to check whether there exist a linear relationship between the variables.*  
**scatter.smooth**(weight,systolic\_bp,main="Relationship btw Weight and systolic BP", lpars = **list**(col="red", lwd = 4, lty = 5))

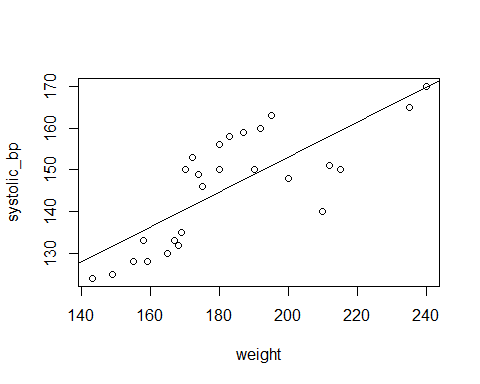


From the scatter plot, we observe that there is a linear relationship between the variables hence we try to fit a simple linear regression model to the data.  
  
*#We observe that weight of the males of age group 25-30 is independent variable and systolic blood pressure of male is dependent variable.*  
  
*#Fitting a simple linear regression model.*  
reg=**lm**(systolic\_bp**~**weight)  
reg

##   
## Call:  
## lm(formula = systolic\_bp ~ weight)  
##   
## Coefficients:  
## (Intercept) weight   
## 69.1044 0.4194

Hence, we obtain the regression coefficient as 0.4194 and the intercept of the fitted model is 69.1044.

*#Plotting the values of weight and systolic blood pressure of males of age group 25-30.*  
**plot**(weight,systolic\_bp)  
  
*#Fitting a regression line.*  
**abline**(reg)



Hence, from the above scatter plot of fitted regression equation we observe that not all the points lie exactly on the fitted regression line some of the points are above the regression line and some are below it.

*#To check whether the sum of observed value is equal to the sum of fitted values.*

*#Obtaining the fitted values of the model.*  
fit\_values=**fitted.values**(reg)  
fit\_values

## 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

Hence, the fitted values of the model are obtained.

*#Sum of observed values*  
**sum**(systolic\_bp)

## [1] 3786

Hence, the sum of the observed values is 3786.

*#Sum of fitted values*  
**sum**(fit\_values)

## [1] 3786

Hence, the sum of the fitted values is 3786.

From the above calculation we observe that sum of observed values and fitted values are equal which is equal to 3786.  
  
*#Obtaining the residuals that is residuals=observed value - fitted values.*   
d=**resid**(reg)  
d

## 1 2 3 4 5 6   
## -8.3078813 -6.1467117 5.4008907 -6.1137292 -7.0203958 3.4979667   
## 7 8 9 10 11 12   
## 1.2067387 -17.1815654 -4.9874134 -6.5972380 -2.3719749 -4.9855421   
## 13 14 15 16 17 18   
## 9.5950427 11.7562123 -7.7913901 -7.5661269 6.9173819 12.1426451   
## 19 20 21 22 23 24   
## -9.2786414 12.1096626 11.4008907 -5.0807468 0.2359785 -2.6669455   
## 25 26   
## 10.3679082 11.4649843

Hence, the residual errors are obtained.

*#Obtaining the mean of residuals.*

mean(d)

## [1] 1.712706e-16

We observe that the expected value of errors is 0 i.e. the residual errors have mean approximately equal to zero.

***CONCLUSION***

From the scatter plot we observe that there is a linear relationship between the variables that is the weight and systolic BP of males of age group 25-30 are linearly related.

Hence we proceed to fit a simple linear regression model to the data.

The fitted model is Yˆ= 69.1044+0.4194 X.

Regression coefficients represent the mean change in the response variable i.e. systolic BP for one unit of change in the predictor variable i.e. weight.

We observe that the regression coefficient is positive that is 0.4194 which indicates that as the weight (independent variable) increases the mean of the systolic BP (dependent variable) also tends to increases

We have obtained the intercept as 69.1044 which indicates the expected mean systolic BP when weight of a person is 0 but here the weight of person cannot be zero hence the the intercept do not have a meaningful interpretation

The equation shows that the coefficient for the weight to be gained is 0.4194. The coefficient indicates that for additional weight to be gained we can expect the systolic BP to increase by an average of 0.4194. That is, from the data given, we estimate that the systolic BP increases by amount is about 0.4194 for the increase in weight of the person.

From the scatter plot of fitted regression equation above, it can be seen that not all the data points fall exactly on the fitted regression line. Some of the points are above the line and some are below it; overall, the residual errors (e) have approximately mean zero.

From the above calculations we also observe that the sum of observed values is equal to the sum of fitted values.