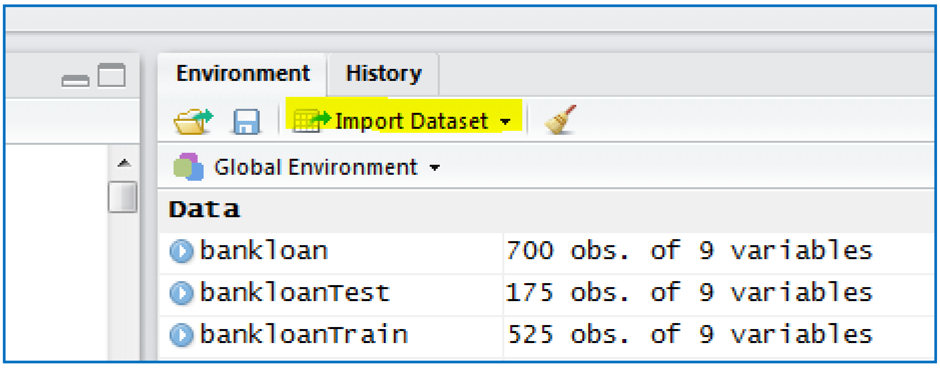
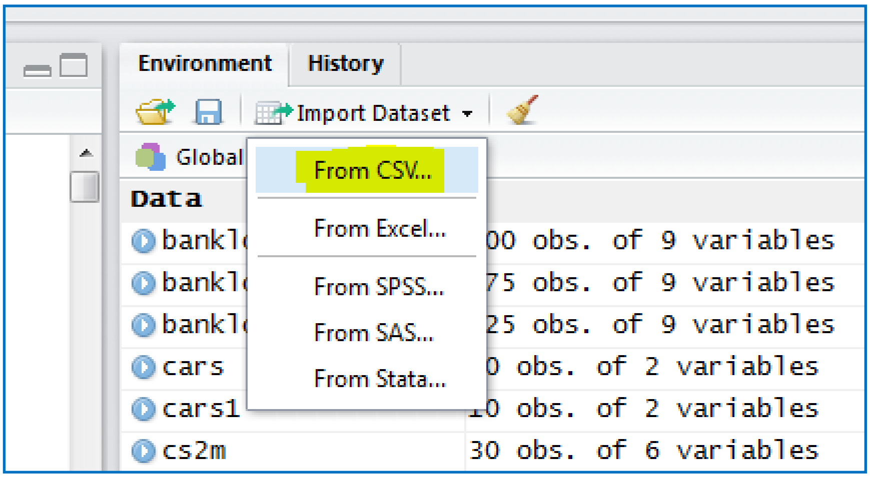
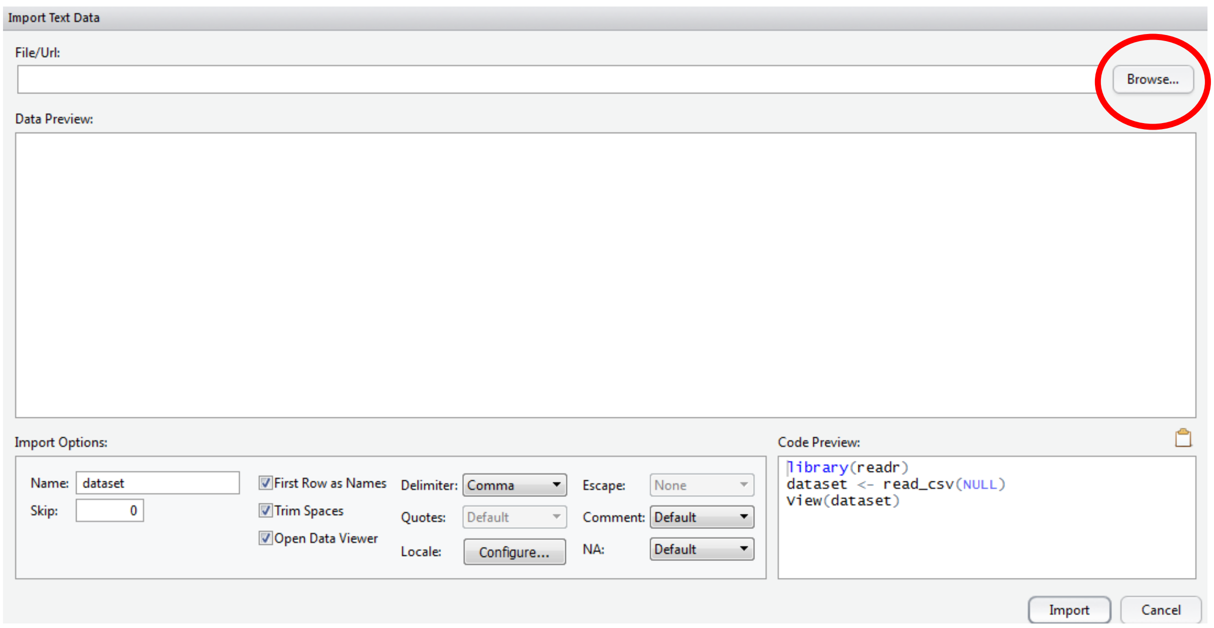
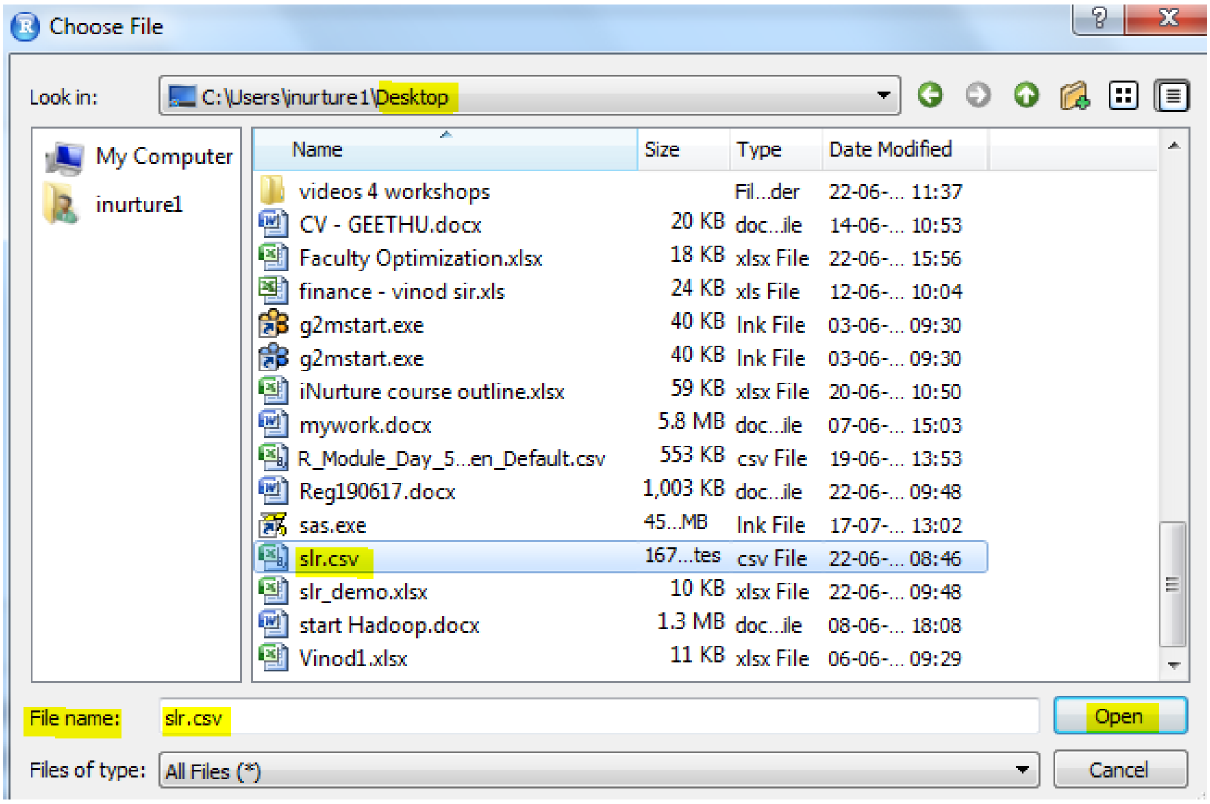
Import File in R

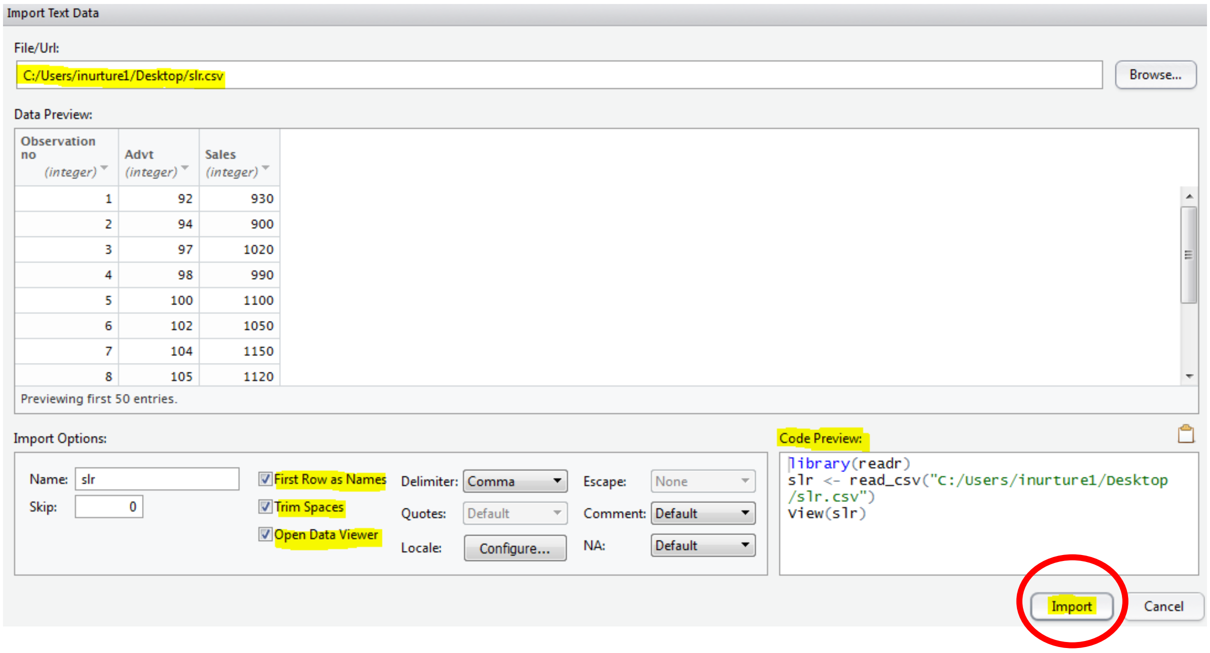
File name **slr.csv**

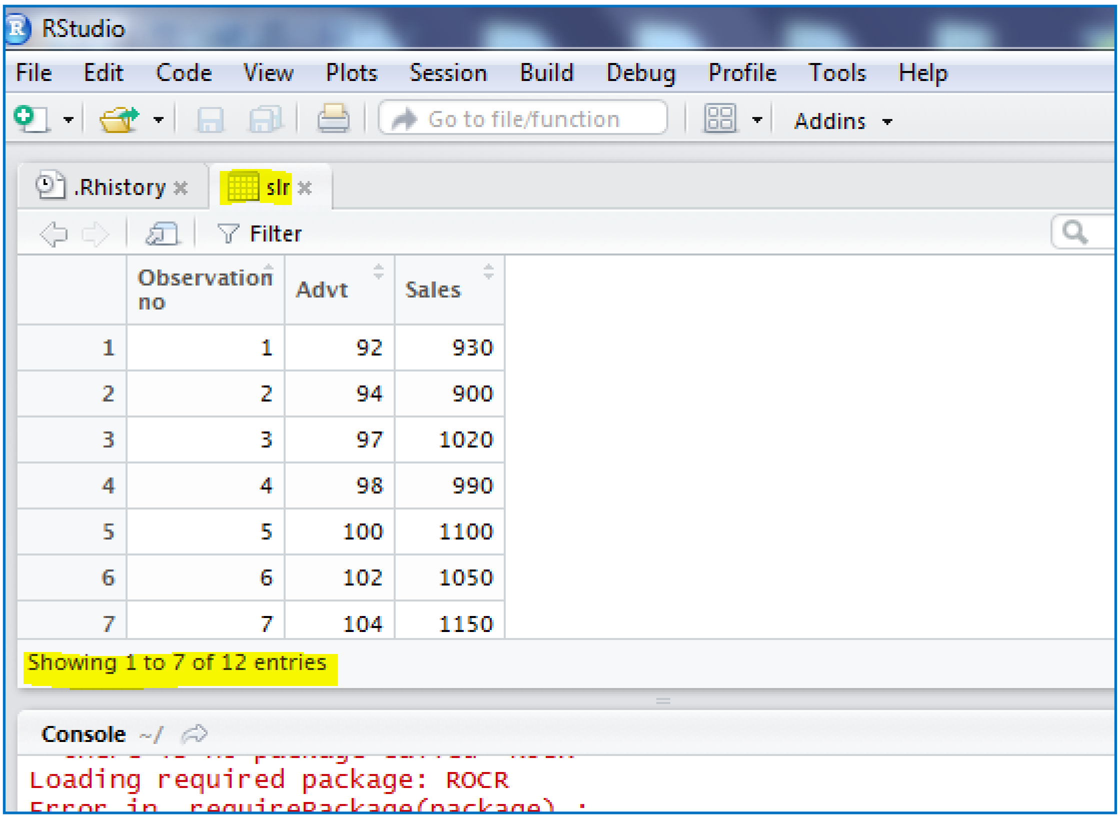




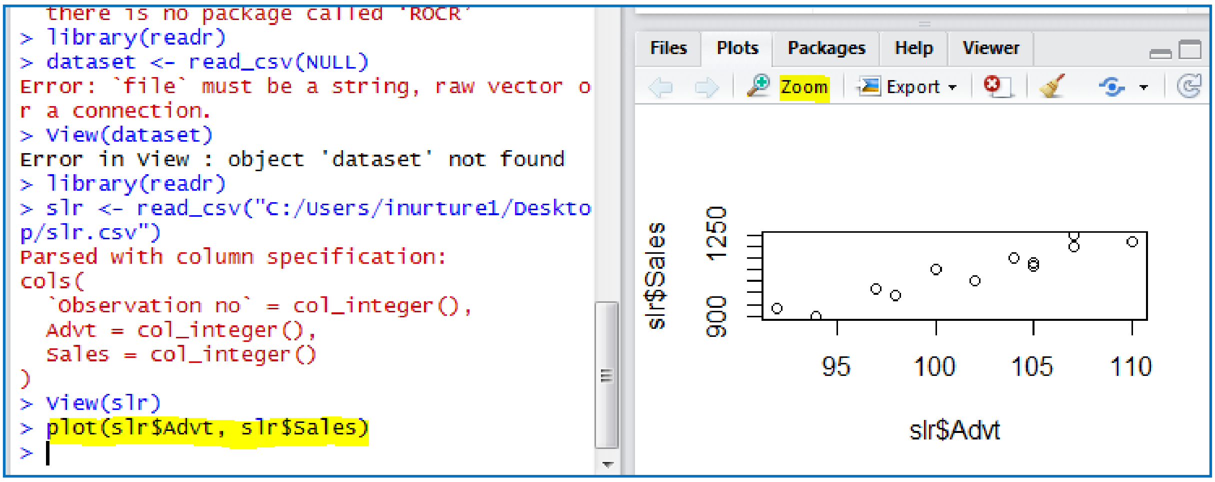


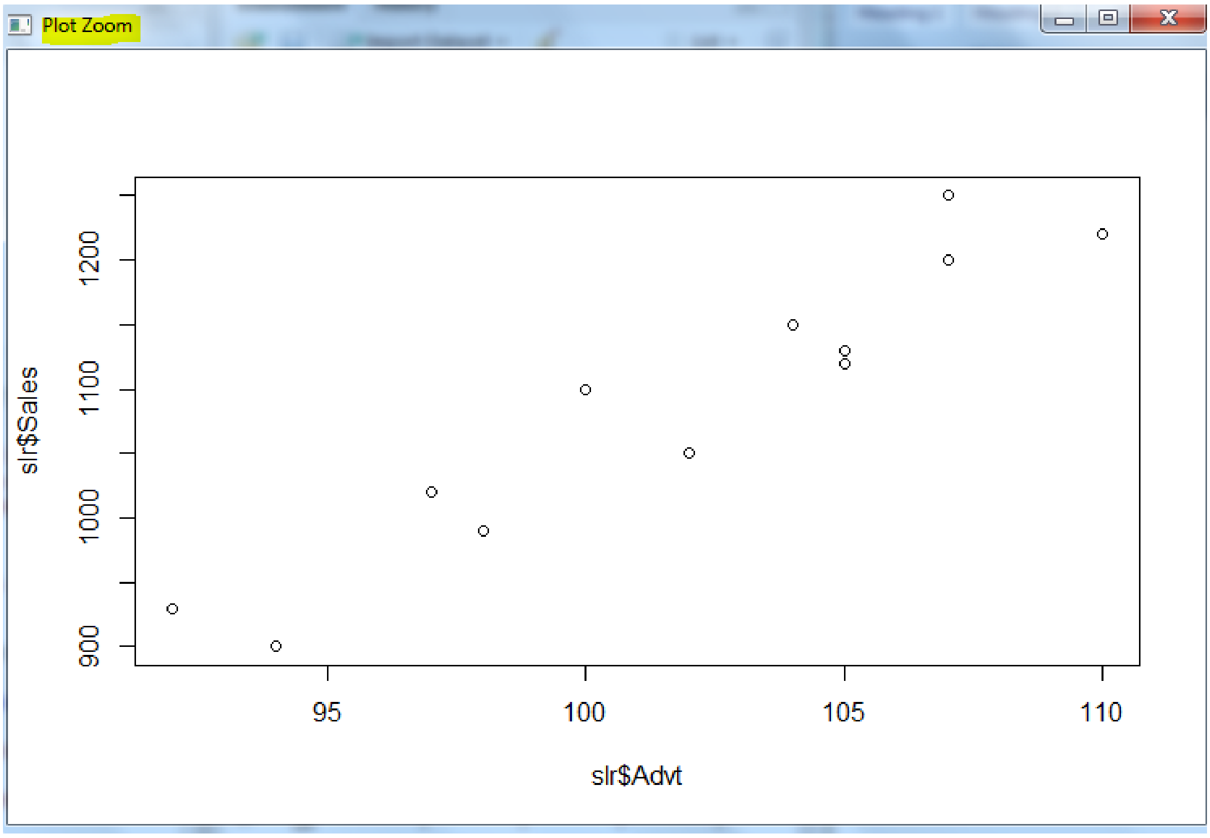

Choose File
Look in:
C:\Users\inurture1\Desktop
My Computer
inurture1
Name
Size
Type
Date Modified
videos 4 workshops Fil...der 22-06-... 11:37
CV - GEETHU.docx 20 KB doc...ile 14-06-... 10:53
Faculty Optimization.xlsx 18 KB xlsx File 22-06-... 15:56
finance - vinod sir.xls 24 KB xls File 12-06-... 10:04
g2mstart.exe 40 KB lnk File 03-06-... 09:30
g2mstart.exe 40 KB lnk File 03-06-... 09:30
iNurture course outline.xlsx 59 KB xlsx File 20-06-... 10:50
mywork.docx 5.8 MB doc...ile 07-06-... 15:03
R_Module_Day_5...en_Default.csv 553 KB csv File 19-06-... 13:53
Reg190617.docx 1,003 KB doc...ile 22-06-... 09:48
sas.exe 45...MB lnk File 17-07-... 13:02
slr.csv 167...tes csv File 22-06-... 08:46
slr_demo.xlsx 10 KB xlsx File 22-06-... 09:48
start Hadoop.docx 1.3 MB doc...ile 08-06-... 18:08
Vinod1.xlsx 11 KB xlsx File 06-06-... 09:29
File name:
slr.csv
Open
Files of type:
All Files (*)
Cancel


Import Text Data
File/Url:
C:/Users/inurture1/Desktop/slr.csv
Browse...
Data Preview:
Observation no (integer) Advt (integer) Sales (integer)
1 92 930
2 94 900
3 97 1020
4 98 990
5 100 1100
6 102 1050
7 104 1150
8 105 1120
Previewing first 50 entries.
Import Options:
Name: slr
Skip: 0
First Row as Names
Trim Spaces
Open Data Viewer
Delimiter: Comma
Quotes: Default
Locale: Configure...
Escape: None
Comment: Default
NA: Default
Code Preview:
library(readr)
slr <- read_csv("C:/Users/inurture1/Desktop/slr.csv")
View(slr)
Import
Cancel

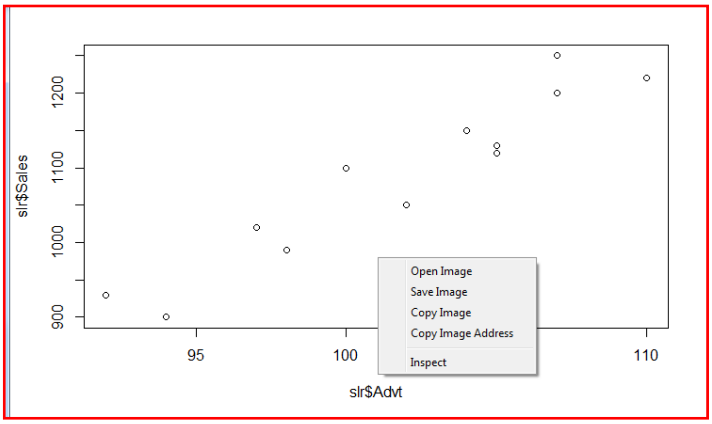


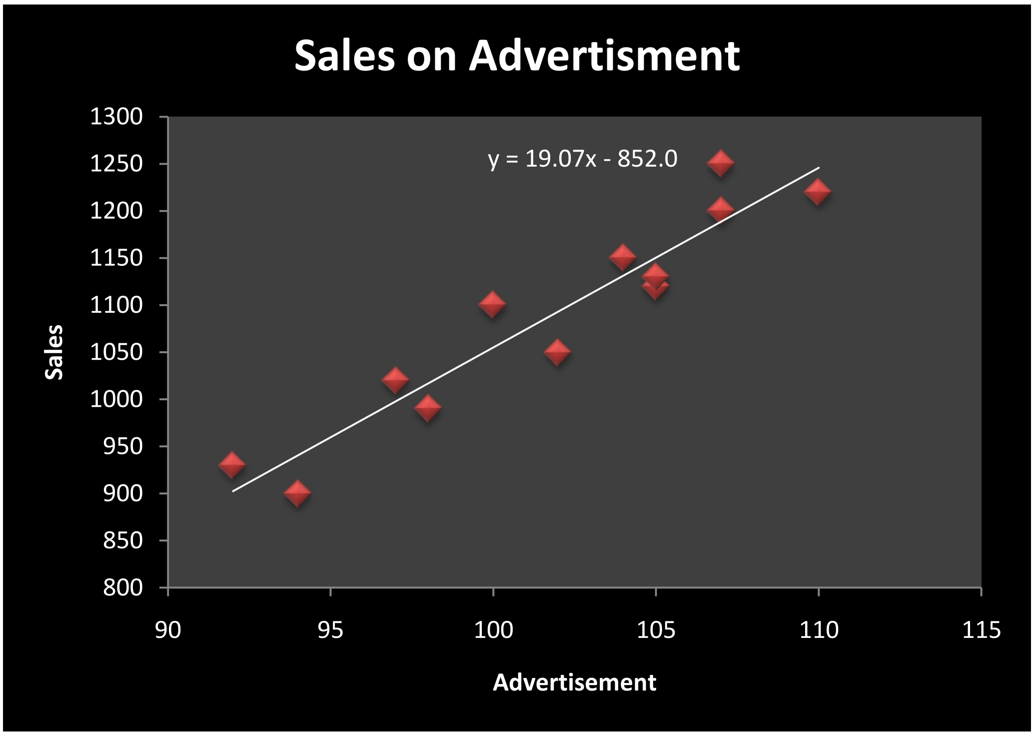
Always first plot scatter plot and see relationship





Copy plot and paste at your desired destination





Type equation here.

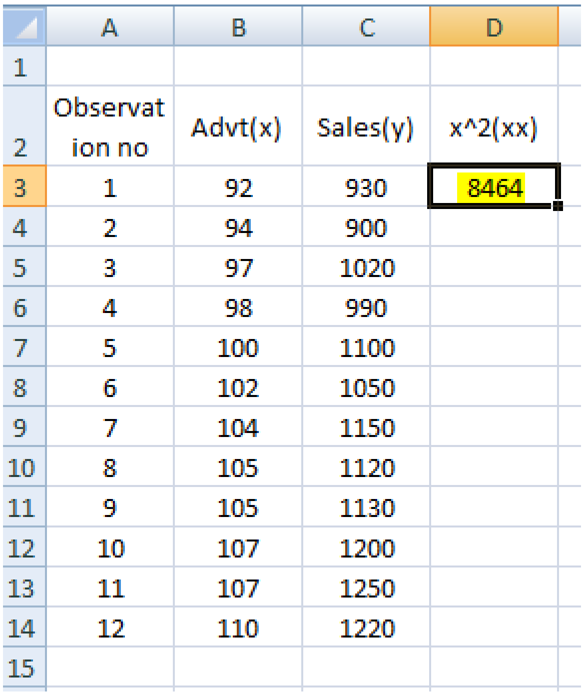
Build Linear Regression Model

```
> mod<-lm(slr$Sales~slr$Advt)
> mod

Call:
lm(formula = slr$Sales ~ slr$Advt)

Coefficients:
(Intercept)     slr$Advt
    -852.08        19.07
```

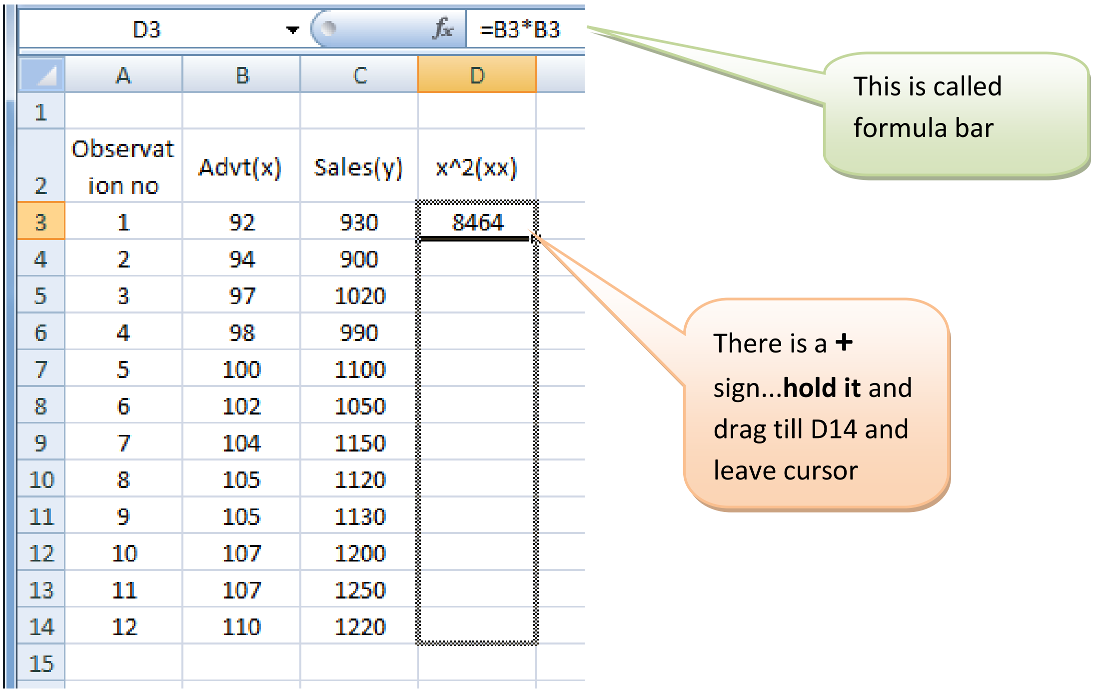
```
> summary(mod)

Call:
lm(formula = slr$Sales ~ slr$Advt)

Residuals:
    Min      1Q  Median      3Q     Max 
-43.101 -27.692  -4.383  23.589  61.547 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -852.08     203.78  -4.181  0.00188 ** 
slr$Advt       19.07       2.00   9.535 2.45e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 37.11 on 10 degrees of freedom
Multiple R-squared:  0.9009,    Adjusted R-squared:  0.891 
F-statistic: 90.93 on 1 and 10 DF,  p-value: 2.454e-06
```

	A	B	C	D
1				
2	Observation no	Advt(x)	Sales(y)	x^2(xx)
3	1	92	930	=B3*B3
4	2	94	900	
5	3	97	1020	
6	4	98	990	
7	5	100	1100	
8	6	102	1050	
9	7	104	1150	
10	8	105	1120	
11	9	105	1130	
12	10	107	1200	
13	11	107	1250	
14	12	110	1220	
15				

Leave Row 1 as blank



	A	B	C	D
1				
2	Observation no	Advt(x)	Sales(y)	x^2(xx)
3	1	92	930	8464
4	2	94	900	
5	3	97	1020	
6	4	98	990	
7	5	100	1100	
8	6	102	1050	
9	7	104	1150	
10	8	105	1120	
11	9	105	1130	
12	10	107	1200	
13	11	107	1250	
14	12	110	1220	
15				



	A	B	C	D
1				
2	Observation no	Advt(x)	Sales(y)	x^2(xx)
3	1	92	930	8464
4	2	94	900	
5	3	97	1020	
6	4	98	990	
7	5	100	1100	
8	6	102	1050	
9	7	104	1150	
10	8	105	1120	
11	9	105	1130	
12	10	107	1200	
13	11	107	1250	
14	12	110	1220	
15				

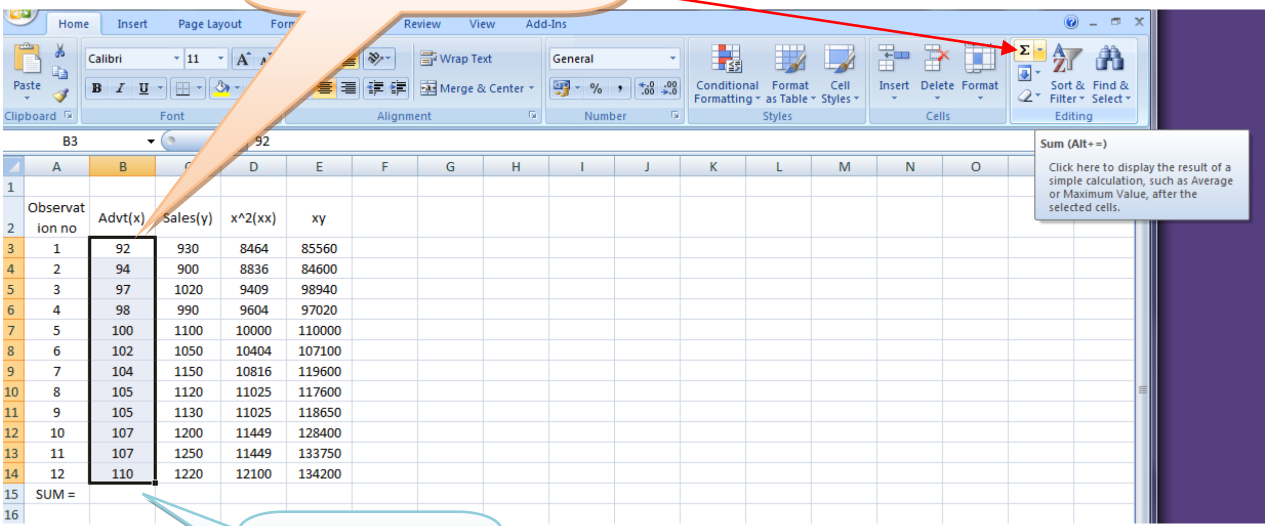
	A	B	C	D
1				
2	Observation no	Advt(x)	Sales(y)	x^2(xx)
3	1	92	930	8464
4	2	94	900	8836
5	3	97	1020	9409
6	4	98	990	9604
7	5	100	1100	10000
8	6	102	1050	10404
9	7	104	1150	10816
10	8	105	1120	11025
11	9	105	1130	11025
12	10	107	1200	11449
13	11	107	1250	11449
14	12	110	1220	12100
15				

	A	B	C	D	E
1					
2	Observation no	Advt(x)	Sales(y)	x^2(xx)	xy
3	1	92	930	8464	=B3*C3
4	2	94	900	8836	
5	3	97	1020	9409	
6	4	98	990	9604	
7	5	100	1100	10000	
8	6	102	1050	10404	
9	7	104	1150	10816	
10	8	105	1120	11025	
11	9	105	1130	11025	
12	10	107	1200	11449	
13	11	107	1250	11449	
14	12	110	1220	12100	
15					

Press ENTER →85560 will appear → There is a + sign...**hold it** and drag till E14 and leave cursor

	A	B	C	D	E
1					
2	Observation no	Advt(x)	Sales(y)	x^2(xx)	xy
3	1	92	930	8464	85560
4	2	94	900	8836	84600
5	3	97	1020	9409	98940
6	4	98	990	9604	97020
7	5	100	1100	10000	110000
8	6	102	1050	10404	107100
9	7	104	1150	10816	119600
10	8	105	1120	11025	117600
11	9	105	1130	11025	118650
12	10	107	1200	11449	128400
13	11	107	1250	11449	133750
14	12	110	1220	12100	134200
15					

Select B3 to B14 → Go to **Sum** and click it. You will have sum at B15



You will get sum here in B15

	A	B	C	D	E
1					
2	Observation no	Advt(x)	Sales(y)	x^2(xx)	xy
3	1	92	930	8464	85560
4	2	94	900	8836	84600
5	3	97	1020	9409	98940
6	4	98	990	9604	97020
7	5	100	1100	10000	110000
8	6	102	1050	10404	107100
9	7	104	1150	10816	119600
10	8	105	1120	11025	117600
11	9	105	1130	11025	118650
12	10	107	1200	11449	128400
13	11	107	1250	11449	133750
14	12	110	1220	12100	134200
15	SUM =	1221			
16					

There is a **+** sign...**hold it** and drag till E15 and leave cursor

	A	B	C	D	E
1					
2	Observation no	Advt(x)	Sales(y)	x^2(xx)	xy
3	1	92	930	8464	85560
4	2	94	900	8836	84600
5	3	97	1020	9409	98940
6	4	98	990	9604	97020
7	5	100	1100	10000	110000
8	6	102	1050	10404	107100
9	7	104	1150	10816	119600
10	8	105	1120	11025	117600
11	9	105	1130	11025	118650
12	10	107	1200	11449	128400
13	11	107	1250	11449	133750
14	12	110	1220	12100	134200
15	SUM =	1221	13060	124581	1335420
16					

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$$

$$SS_{xy} = 1335420 - \frac{1221 * 13060}{12} = 6565$$

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 124581 - \frac{(1221)^2}{12} = 344.25$$

This is called Regression Coefficient

$$\boldsymbol{b_1} = \frac{SS_{xy}}{SS_{xx}} = \frac{6565}{344.25} = \mathbf{19.07}$$

This is called Intercept

$$\boldsymbol{b_0} = \frac{\sum y}{n} - b_1 \frac{\sum x}{n} = \frac{13060}{12} - 19.07 \frac{1221}{12} = \mathbf{-852.08}$$

Regression Equation:

$$y = a + bx$$

$$y = b_0 + b_1 x$$

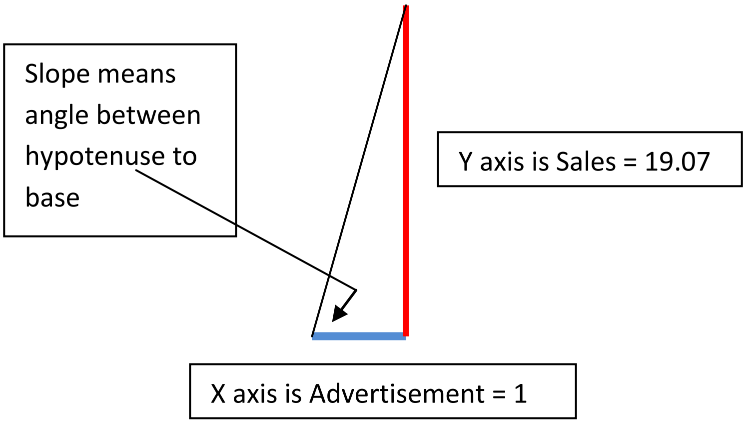
$$Sales = intercept + regression\ coefficient \times advertisement$$

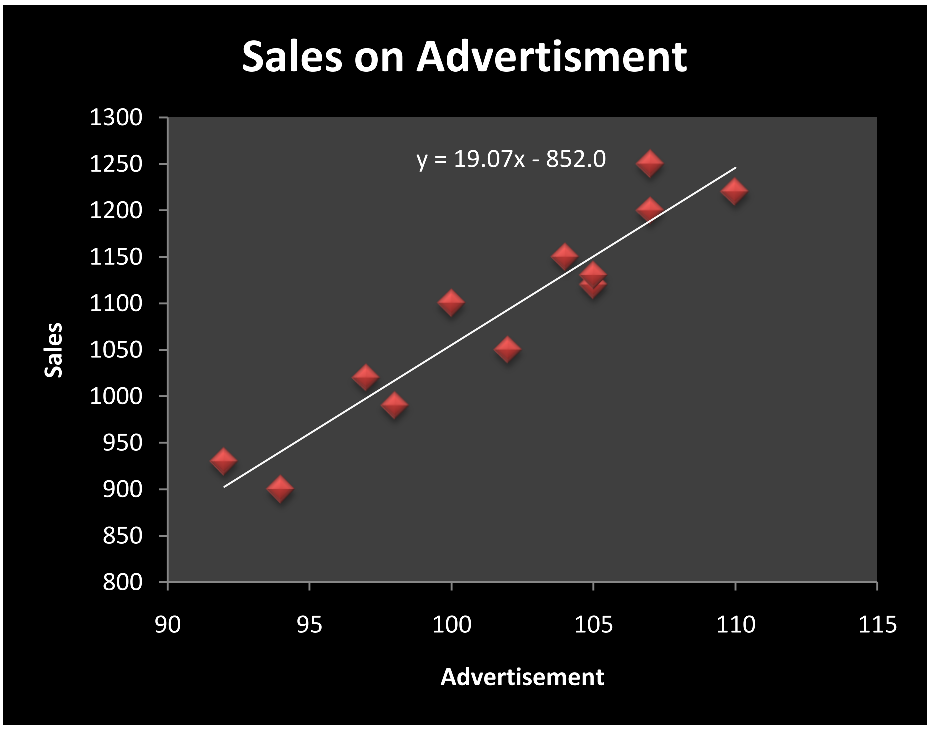
$$\boldsymbol{Sales = -852.08 + 19.07 \times advertisement}$$

Interpretation of b_1

One unit change in ***Advertisement*** will result in 19.07 times change in ***Sales***

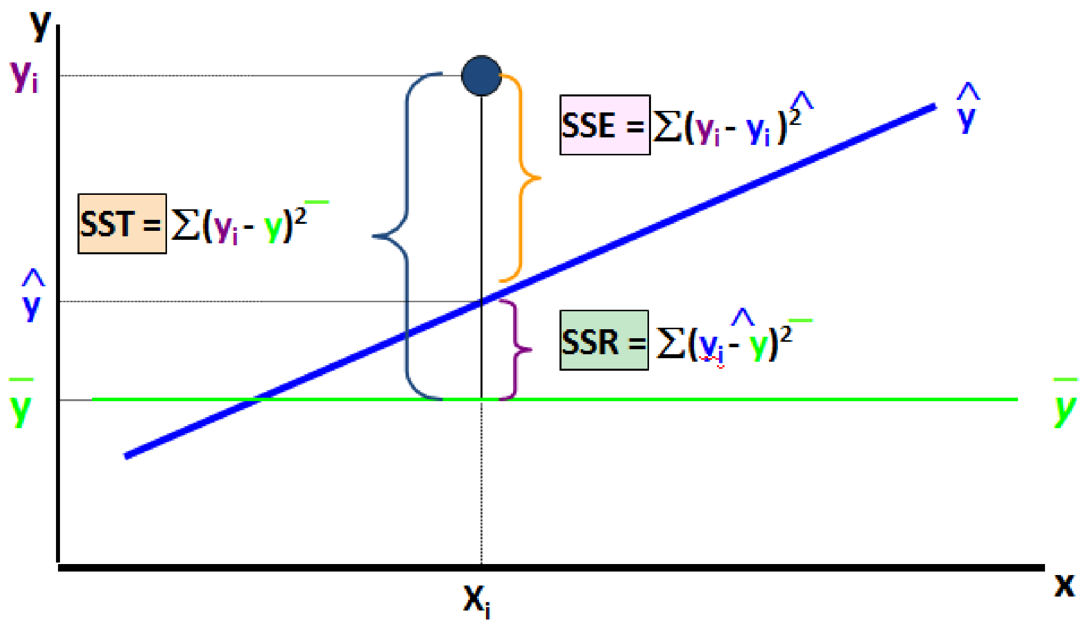
Regression coefficient is also called Slope (or Tan θ = Perpendicular/Base = Sales/Advertisement)





Standard Error of Mean

```
> summary(mod)

Call:
lm(formula = slr$Sales ~ slr$Advt)

Residuals:
    Min      1Q  Median      3Q     Max 
-43.101 -27.692  -4.383  23.589  61.547 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -852.08     203.78  -4.181  0.00188 ** 
slr$Advt       19.07       2.00   9.535 2.45e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 37.11 on 10 degrees of freedom
Multiple R-squared:  0.9009,	Adjusted R-squared:  0.891 
F-statistic: 90.93 on 1 and 10 DF,  p-value: 2.454e-06
```



ZTEST | =$E$1+$F$1*B3

	A	B	C	D	E	F
1					-852.08	19.07
2	Observation no	Advt(x)	Sales(y)	x^2(xx)	xy	Pred= yhat = a + bx = -852.08+ 19.07x
3	1	92	930	8464		=$E$1+$F$1*B3
4	2	94	900	8836	84600	
5	3	97	1020	9409	98940	
6	4	98	990	9604	97020	
7	5	100	1100	10000	110000	
8	6	102	1050	10404	107100	
9	7	104	1150	10816	119600	
10	8	105	1120	11025	117600	
11	9	105	1130	11025	118650	
12	10	107	1200	11449	128400	
13	11	107	1250	11449	133750	
14	12	110	1220	12100	134200	
15	SUM =	1221	13060	124581	1335420	
16						

$ sign is used for fixing the same calculation for all cells

Press enter → hold the **+** sign at the right bottom corner of cell F3 → drag till F15 → LEAVE the cursor → you will have calculations in all cells (F4:F15)

	A	B	C	D	E	F
1					-852.08	19.07
2	Observation no	Advt(x)	Sales(y)	x^2(xx)	xy	Pred= yhat = a + bx = -852.08+ 19.07x
3	1	92	930	8464	85560	902.36
4	2	94	900	8836	84600	940.5
5	3	97	1020	9409	98940	997.71
6	4	98	990	9604	97020	1016.78
7	5	100	1100	10000	110000	1054.92
8	6	102	1050	10404	107100	1093.06
9	7	104	1150	10816	119600	1131.2
10	8	105	1120	11025	117600	1150.27
11	9	105	1130	11025	118650	1150.27
12	10	107	1200	11449	128400	1188.41
13	11	107	1250	11449	133750	1188.41
14	12	110	1220	12100	134200	1245.62
15	SUM =	1221	13060	124581	1335420	22432.39
16						

```
> Pred<-predict(lm(slr$Sales~slr$Advt))
```

```
> Pred
        1         2         3         4         5         6         7 
 902.3965  940.5374  997.7487 1016.8192 1054.9601 1093.1009 1131.2418 
        8         9        10        11        12 
1150.3123 1150.3123 1188.4532 1188.4532 1245.6645 
```

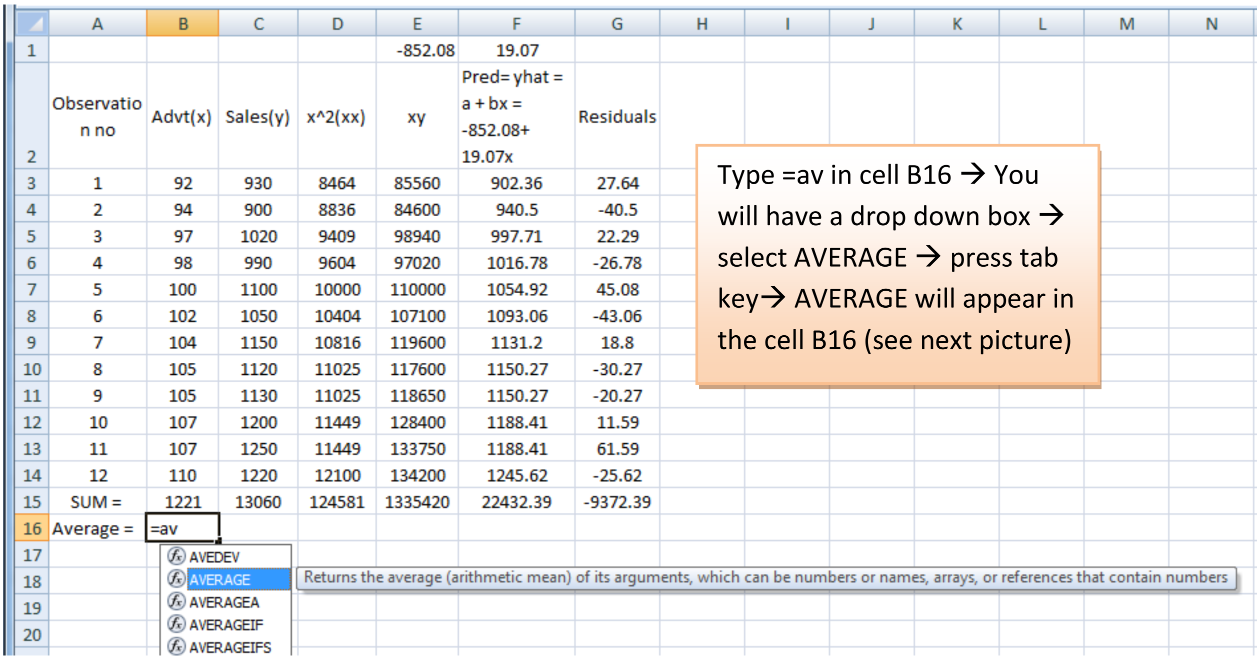
	A	B	C	D	E	F	G
1					-852.08	19.07	
2	Observation no	Advt(x)	Sales(y)	x^2(xx)	xy	Pred= yhat = a + bx = -852.08+ 19.07x	Residuals
3	1	92	930	8464	85560	902.36	=C3-F3
4	2	94	900	8836	84600	940.5	
5	3	97	1020	9409	98940	997.71	
6	4	98	990	9604	97020	1016.78	
7	5	100	1100	10000	110000	1054.92	
8	6	102	1050	10404	107100	1093.06	
9	7	104	1150	10816	119600	1131.2	
10	8	105	1120	11025	117600	1150.27	
11	9	105	1130	11025	118650	1150.27	
12	10	107	1200	11449	128400	1188.41	
13	11	107	1250	11449	133750	1188.41	
14	12	110	1220	12100	134200	1245.62	
15	SUM =	1221	13060	124581	1335420	22432.39	
16							

Press enter → hold the **+** sign at the right bottom corner of cell G3 → drag till G15 → LEAVE the cursor → you will have calculations in all cells (G4:G15)

```
> error<- residuals(lm(slr$Sales~slr$Advt))
> error
        1         2         3         4         5         6         7 
 27.60349 -40.53740  22.25127 -26.81917  45.03994 -43.10094  18.75817 
        8         9        10        11        12 
-30.31227 -20.31227  11.54684  61.54684 -25.66449 
> summary(error)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
-43.100 -27.690  -4.383   0.000  23.590  61.550 
```

	A	B	C	D	E	F	G
1					-852.08	19.07	
2	Observation no	Advt(x)	Sales(y)	x^2(xx)	xy	Pred= yhat = a + bx = -852.08+ 19.07x	Residuals
3	1	92	930	8464	85560	902.36	27.64
4	2	94	900	8836	84600	940.5	-40.5
5	3	97	1020	9409	98940	997.71	22.29
6	4	98	990	9604	97020	1016.78	-26.78
7	5	100	1100	10000	110000	1054.92	45.08
8	6	102	1050	10404	107100	1093.06	-43.06
9	7	104	1150	10816	119600	1131.2	18.8
10	8	105	1120	11025	117600	1150.27	-30.27
11	9	105	1130	11025	118650	1150.27	-20.27
12	10	107	1200	11449	128400	1188.41	11.59
13	11	107	1250	11449	133750	1188.41	61.59
14	12	110	1220	12100	134200	1245.62	-25.62
15	SUM =	1221	13060	124581	1335420	22432.39	-9372.39
16							

Find average in cell B16



	A	B	C	D	E	F	G
1					-852.08	19.07	
2	Observation no	Advt(x)	Sales(y)	x^2(xx)	xy	Pred= yhat = a + bx = -852.08+ 19.07x	Residuals
3	1	92	930	8464	85560	902.36	27.64
4	2	94	900	8836	84600	940.5	-40.5
5	3	97	1020	9409	98940	997.71	22.29
6	4	98	990	9604	97020	1016.78	-26.78
7	5	100	1100	10000	110000	1054.92	45.08
8	6	102	1050	10404	107100	1093.06	-43.06
9	7	104	1150	10816	119600	1131.2	18.8
10	8	105	1120	11025	117600	1150.27	-30.27
11	9	105	1130	11025	118650	1150.27	-20.27
12	10	107	1200	11449	128400	1188.41	11.59
13	11	107	1250	11449	133750	1188.41	61.59
14	12	110	1220	12100	134200	1245.62	-25.62
15	SUM =	1221	13060	124581	1335420	22432.39	-9372.39
16	Average =	=av					

	A	B	C	D	E
1					-852.08
2	Observatio n no	Advt(x)	Sales(y)	x^2(xx)	xy
3	1	92	930	8464	85560
4	2	94	900	8836	84600
5	3	97	1020	9409	98940
6	4	98	990	9604	97020
7	5	100	1100	10000	110000
8	6	102	1050	10404	107100
9	7	104	1150	10816	119600
10	8	105	1120	11025	117600
11	9	105	1130	11025	118650
12	10	107	1200	11449	128400
13	11	107	1250	11449	133750
14	12	110	1220	12100	134200
15	SUM =	1221	13060	124581	1335420
16	Average =	=AVERAGE(			
17		AVERAGE(**number1**, [number2], ...)			
18					

	A	B	C	D	E	F	G
1					-852.08	19.07	
2	Observation no	Advt(x)	Sales(y)	x^2(xx)	xy	Pred= yhat = a + bx = -852.08+ 19.07x	Residuals
3	1	92	930	8464	85560	902.36	27.64
4	2	94	900	8836	84600	940.5	-40.5
5	3	97	1020			97.71	22.29
6	4	98	990			016.78	-26.78
7	5	100	1100	1		054.92	45.08
8	6	102	1050	1		093.06	-43.06
9	7	104	1150	1		131.2	18.8
10	8	105	1120	11025	117600	1150.27	-30.27
11	9	105	1130	11025	118650	1150.27	-20.27
12	10	107	1200	11449	128400	1188.41	11.59
13	11	107	1250	11449	133750	1188.41	61.59
14	12	110	1220	12100	134200	1245.62	-25.62
15	SUM =	1221	13060	124581	1335420	22432.39	-9372.39
16	Average =	=AVERAGE(B3:B14					
17		AVERAGE(**number1**, [number2], ...)					
18							

Select the range B3:B14 (DONT INCLUDE B15) →PRESS enter

	A	B	C	D	E	F	G
1					-852.08	19.07	
2	Observation no	Advt(x)	Sales(y)	x^2(xx)	xy	Pred= yhat = a + bx = -852.08+ 19.07x	Residuals
3	1	92	930	8464	85560	902.36	27.64
4	2	94	900	8836	84600	940.5	-40.5
5	3	97	1020	9409	98940	997.71	22.29
6	4	98	990	9604	97020	1016.78	-26.78
7	5	100	1100	10000	110000	1054.92	45.08
8	6	102	1050	10404	107100	1093.06	-43.06
9	7	104	1150	10816	119600	1131.2	18.8
10	8	105	1120	11025	117600	1150.27	-30.27
11	9	105	1130	11025	118650	1150.27	-20.27
12	10	107	1200	11449	128400	1188.41	11.59
13	11	107	1250	11449	133750	1188.41	61.59
14	12	110	1220	12100	134200	1245.62	-25.62
15	SUM =	1221	13060	124581	1335420	22432.39	-9372.39
16	Average =	101.75					
17							

There is a **+** sign...**hold it** and drag till G16 and leave cursor

	A	B	C	D	E	F	G
1					-852.08	19.07	
2	Observation no	Advt(x)	Sales(y)	x^2(xx)	xy	Pred= yhat = a + bx = -852.08+ 19.07x	Residuals
3	1	92	930	8464	85560	902.36	27.64
4	2	94	900	8836	84600	940.5	-40.5
5	3	97	1020	9409	98940	997.71	22.29
6	4	98	990	9604	97020	1016.78	-26.78
7	5	100	1100	10000	110000	1054.92	45.08
8	6	102	1050	10404	107100	1093.06	-43.06
9	7	104	1150	10816	119600	1131.2	18.8
10	8	105	1120	11025	117600	1150.27	-30.27
11	9	105	1130	11025	118650	1150.27	-20.27
12	10	107	1200	11449	128400	1188.41	11.59
13	11	107	1250	11449	133750	1188.41	61.59
14	12	110	1220	12100	134200	1245.62	-25.62
15	SUM =	1221	13060	124581	1335420	22432.39	-9372.39
16	Average =	101.75	1088.33	10381.8	111285	1088.2925	0.040833

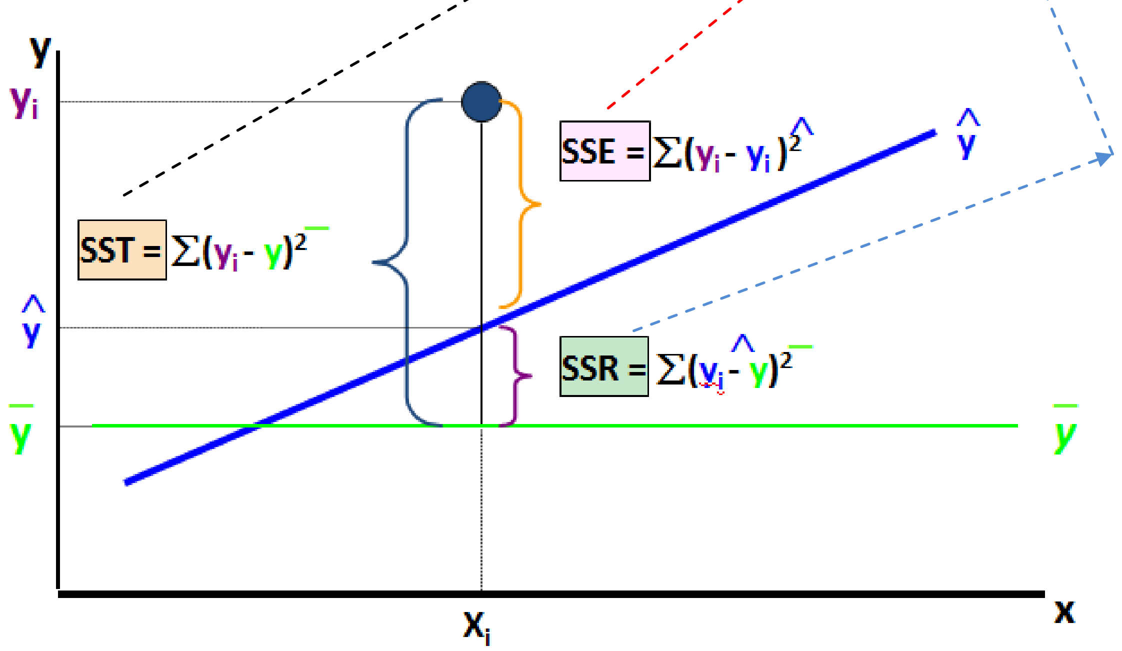
Congratulations! Good work done so far! Now we are set to crack the mystery of complicated terms (almost scary!) with the help of above numbers. We just need to calculate three more terms that is SST, SSR $ SSE

[This time am not showing each excel step and believing that you will be able to create in your excel sheet and find SUM in row 15, HOWEVER, showing that how you should use the $ sign in formulas]

	A	B	C	D	E	F	G	H	I	J
1					-852.08	19.07				
2	Observation no	Advt(x)	Sales(y)	x^2(xx)	xy	Pred= yhat = a + bx = -852.08+ 19.07x	Residuals	**SST =** (y-*AVG* y)^2	**SSR =** (yhat- *AVG* yhat)^2	**SSE =** (y-yhat)^2
3	1	92	930	8464	85560	902.36	27.64	=(C3-$C$16)^2	=(F3-$F$16)^2	=(C3-F3)^2
4	2	94	900	8836	84600	940.5	-40.5			
5	3	97	1020	9409	98940	997.71	22.29			
6	4	98	990	9604	97020	1016.78	-26.78			
7	5	100	1100	10000	110000	1054.92	45.08			
8	6	102	1050	10404	107100	1093.06	-43.06			
9	7	104	1150	10816	119600	1131.2	18.8			
10	8	105	1120	11025	117600	1150.27	-30.27			
11	9	105	1130	11025	118650	1150.27	-20.27			
12	10	107	1200	11449	128400	1188.41	11.59			
13	11	107	1250	11449	133750	1188.41	61.59			
14	12	110	1220	12100	134200	1245.62	-25.62			
15	SUM =	1221	13060	124581	1335420	22432.39	-9372.39			
16	Average =	101.75	1088.33	10381.8	111285	1088.2925	0.040833			
17										

	A	B	C	D	E	F	G	H	I	J
1					-852.08	19.07				
2	Observation no	Advt(x)	Sales(y)	x^2(xx)	xy	Pred= yhat = a + bx = -852.08+ 19.07x	Residuals	SST = (y-*AVG*y)^2	SSR = (yhat-*AVG*yhat)^2	SSE = (y-yhat)^2
3	1	92	930	8464	85560	902.36	27.64	25069.44444	34570.89456	763.9696
4	2	94	900	8836	84600	940.5	-40.5	35469.44444	21842.62306	1640.25
5	3	97	1020	9409	98940	997.71	22.29	4669.444444	8205.189306	496.8441
6	4	98	990	9604	97020	1016.78	-26.78	9669.444444	5114.037656	717.1684
7	5	100	1100	10000	110000	1054.92	45.08	136.1111111	1113.723756	2032.2064
8	6	102	1050	10404	107100	1093.06	-43.06	1469.444444	22.72905625	1854.1636
9	7	104	1150	10816	119600	1131.2	18.8	3802.777778	1841.053556	353.44
10	8	105	1120	11025	117600	1150.27	-30.27	1002.777778	3841.210506	916.2729
11	9	105	1130	11025	118650	1150.27	-20.27	1736.111111	3841.210506	410.8729
12	10	107	1200	11449	128400	1188.41	11.59	12469.44444	10023.51381	134.3281
13	11	107	1250	11449	133750	1188.41	61.59	26136.11111	10023.51381	3793.3281
14	12	110	1220	12100	134200	1245.62	-25.62	17336.11111	24751.94226	656.3844
15	SUM =	1221	13060	124581	1335420	22432.39	-9372.39	138966.6667	125191.6418	13769.2285
16	Average =	101.75	1088.33	10381.8	111285	1088.2925	0.040833			
17										

Explained and Unexplained Variation



Now the moment has to come to understand the calculation Residual Standard Error (37.11)

```
> summary(mod)

Call:
lm(formula = slr$Sales ~ slr$Advt)

Residuals:
    Min      1Q  Median      3Q     Max 
-43.101 -27.692  -4.383  23.589  61.547 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -852.08     203.78  -4.181  0.00188 ** 
slr$Advt       19.07       2.00   9.535 2.45e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 37.11 on 10 degrees of freedom
Multiple R-squared:  0.9009,    Adjusted R-squared:  0.891 
F-statistic: 90.93 on 1 and 10 DF,  p-value: 2.454e-06
```

$$Standard\ Error_{residuals} = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{\frac{13769.2285}{12-1-1}} = \mathbf{37.11}$$

What is the practical use of 37.11 (Standard Error Residual)?

Say, you want to predict sales for a given advertisement budget of 95.

$$Sales_{@95} = b_0 + b_1 x = -852.08 + 19.07 \times 95 = 959.57$$

It is never a good idea to express prediction in POINT ESTIMATE....rather; we should present CONFIDENCE INTERVALS as shown below:

$$Upper\ Limit_{sales} = \hat{y} + Critical\ value\ of\ Z\ at\ 95\%\ CL \times SE_{residuals}$$

$$Upper\ Limit_{sales} = 959.57 + 1.96 \times 37.11 = \mathbf{1032.306}$$

$$Lower\ Limit_{sales} = \hat{y} - Critical\ value\ of\ Z\ at\ 95\%\ CL \times SE_{residuals}$$

$$Lower\ Limit_{sales} = 959.57 - 1.96 \times 37.11 = \mathbf{886.83}$$

So, you will say that you are 95% confident that the sales correspond to 95 budget of advertisement will lie between 886.83 and 1032.306

Now let's understand t test for slope

```
> summary(mod)

Call:
lm(formula = slr$Sales ~ slr$Advt)

Residuals:
    Min      1Q  Median      3Q     Max 
-43.101 -27.692  -4.383  23.589  61.547 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -852.08     203.78  -4.181  0.00188 ** 
slr$Advt       19.07       2.00   9.535 2.45e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 37.11 on 10 degrees of freedom
Multiple R-squared:  0.9009,	Adjusted R-squared:  0.891 
F-statistic: 90.93 on 1 and 10 DF,  p-value: 2.454e-06
```

These 3 terms, *t*-statistics (9.535), Std Error of regression coefficient (2.00) and Probability (of committing Type I Error) is 2.45e-06 [2.45/1000000 = 0.00000245 = almost 0] are associated with a t test which tests following Null Hypothesis:

Ho: The slope of advertisement with sales is not significant

Ha: The slope of advertisement with sales is significant

In mathematical symbols:

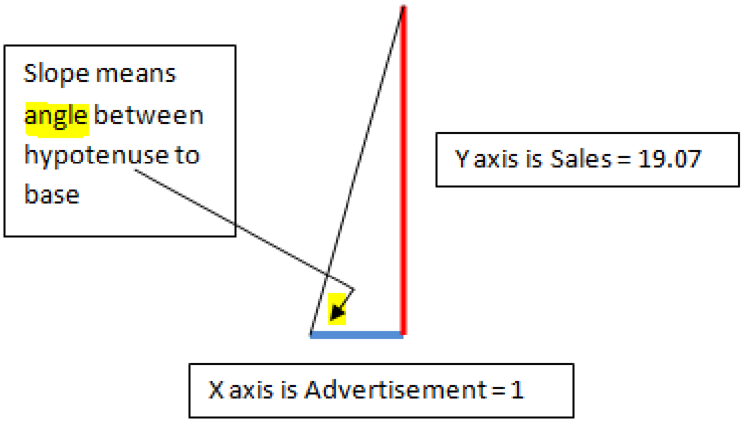
$$H_0: \beta_1 = 0(\textit{there is } \textbf{no} \textit{ linear relationship})$$

$$H_a: \beta_1 \neq 0(\textit{there is linear relationship})$$

Interpretation of b_1

One unit change in Advertisement will result in 19.07 times change in Sales

Regression coefficient is also called Slope (or Tan θ = Perpendicular/Base = Sales/Advertisement)



NO SLOPE will look like this:

See the angle....as a matter of fact there is NO angle....or

$$\beta_1 = 0(\textit{there is } \textbf{no} \textit{ linear relationship})$$

So, first we will calculate the test statistics.

$$t = \frac{b_1 - \beta_1}{S_b}$$

$$S_b = \frac{\textit{Standard Error of residuals}}{\sqrt{SS_{xx}}}$$

$$S_b = \frac{Standard\ Error\ of\ residuals}{\sqrt{SS_{xx}}}$$

$$S_b = \frac{37.11}{\sqrt{344.25}} = 2.00$$

S_b is called Standard Error of Regression Coefficient

$$Standard\ Error_{residuals} = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{\frac{13769.2285}{12-1-1}} = \mathbf{37.11}$$

	A	B	C	D	E
1					
2	Observation no	Advt(x)	Sales(y)	x^2(xx)	xy
3	1	92	930	8464	85560
4	2	94	900	8836	84600
5	3	97	1020	9409	98940
6	4	98	990	9604	97020
7	5	100	1100	10000	110000
8	6	102	1050	10404	107100
9	7	104	1150	10816	119600
10	8	105	1120	11025	117600
11	9	105	1130	11025	118650
12	10	107	1200	11449	128400
13	11	107	1250	11449	133750
14	12	110	1220	12100	134200
15	SUM =	1221	13060	124581	1335420
16					

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n}$$

$$SS_{xy} = 1335420 - \frac{1221 * 13060}{12} = 6565$$

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 124581 - \frac{(1221)^2}{12} = 344.25$$

```
> summary(mod)

Call:
lm(formula = slr$Sales ~ slr$Advt)

Residuals:
    Min      1Q  Median      3Q     Max 
-43.101 -27.692  -4.383  23.589  61.547 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -852.08     203.78  -4.181  0.00188 ** 
slr$Advt       19.07       2.00   9.535 2.45e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 37.11 on 10 degrees of freedom
Multiple R-squared:  0.9009,	Adjusted R-squared:  0.891 
F-statistic: 90.93 on 1 and 10 DF,  p-value: 2.454e-06
```

S_b is called Standard Error of Regression Coefficient

```
> summary(mod)

Call:
lm(formula = slr$Sales ~ slr$Advt)

Residuals:
    Min      1Q  Median      3Q     Max 
-43.101 -27.692  -4.383  23.589  61.547 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -852.08     203.78  -4.181  0.00188 ** 
slr$Advt       19.07       2.00   9.535 2.45e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 37.11 on 10 degrees of freedom
Multiple R-squared:  0.9009,	Adjusted R-squared:  0.891 
F-statistic: 90.93 on 1 and 10 DF,  p-value: 2.454e-06
```

$$t = \frac{b_1 - \beta_1}{S_b}$$

$$t = \frac{19.07 - 0}{2} = \mathbf{9.535}$$

Now let's discuss about Pr(>|t|) or Probability value (Significance value)

Here the rule goes like this:

If, p value is <,= 0.05 (for 5% Level of Significance) → REJECT Ho

If, p value is > 0.05 (for 5% Level of Significance) → ACCEPT Ho

As p value is 0.000 which is less than 0.05, REJECT the Ho (and ACCEPT Ha) and conclude that "Slope is significant"

```
> summary(mod)

Call:
lm(formula = slr$Sales ~ slr$Advt)

Residuals:
    Min      1Q  Median      3Q     Max 
-43.101 -27.692  -4.383  23.589  61.547 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -852.08     203.78  -4.181  0.00188 ** 
slr$Advt       19.07       2.00   9.535 2.45e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 37.11 on 10 degrees of freedom
Multiple R-squared:  0.9009,    Adjusted R-squared:  0.891 
F-statistic: 90.93 on 1 and 10 DF,  p-value: 2.454e-06
```

Verify residuals with following R commands

```
> error<- residuals(lm(slr$Sales~slr$Advt))
> error
        1         2         3         4         5         6         7 
 27.60349 -40.53740  22.25127 -26.81917  45.03994 -43.10094  18.75817 
        8         9        10        11        12 
-30.31227 -20.31227  11.54684  61.54684 -25.66449 
> summary(error)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
-43.100 -27.690  -4.383   0.000  23.590  61.550 
```

Multiple R- squared: 0.9009

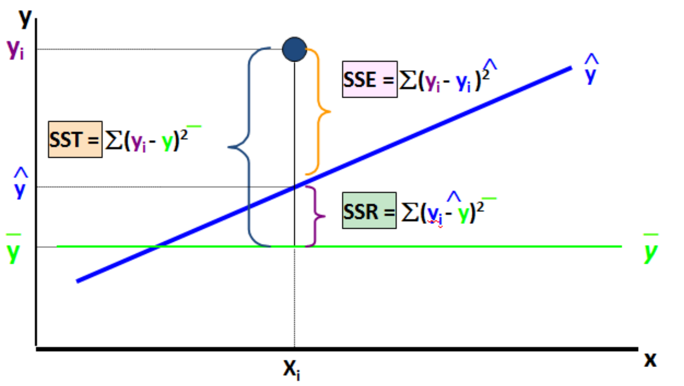
90% of the variance in Sales can be explained by Advertisement [Remaining 10% is unexplained variance....due to factors outside the model]

	A	B	C	D	E	F	G	H	I	J	K	L
1					-852.08	19.07						
2	Observation no	Advt(x)	Sales(y)	x^2(xx)	xy	Pred= yhat = a + bx = -852.08+ 19.07x	Residuals	SST = (y-*AVG*y)^2	SSR = (yhat-*AVG*yhat)^2	SSE = (y-yhat)^2		
3	1	92	930	8464	85560	902.36	27.64	25069.44444	34570.89456	763.9696		
4	2	94	900	8836	84600	940.5	-40.5	35469.44444	21842.62306	1640.25		
5	3	97	1020	9409	98940	997.71	22.29	4669.444444	8205.189306	496.8441		
6	4	98	990	9604	97020	1016.78	-26.78	9669.444444	5114.037656	717.1684		
7	5	100	1100	10000	110000	1054.92	45.08	136.1111111	1113.723756	2032.2064		
8	6	102	1050	10404	107100	1093.06	-43.06	1469.444444	22.72905625	1854.1636		
9	7	104	1150	10816	119600	1131.2	18.8	3802.777778	1841.053556	353.44		
10	8	105	1120	11025	117600	1150.27	-30.27	1002.777778	3841.210506	916.2729		
11	9	105	1130	11025	118650	1150.27	-20.27	1736.111111	3841.210506	410.8729		
12	10	107	1200	11449	128400	1188.41	11.59	12469.44444	10023.51381	134.3281		
13	11	107	1250	11449	133750	1188.41	61.59	26136.11111	10023.51381	3793.3281		
14	12	110	1220	12100	134200	1245.62	-25.62	17336.11111	24751.94226	656.3844		
15	SUM =	1221	13060	124581	1335420	22432.39	-9372.39	138966.6667	125191.6418	[illegible]		
16	Average =	101.75	1088.33	10381.8	111285	1088.2925	0.040833		0.900875331			
17								R Square = SSR/SST				
18												
19												

Dr. Vinod:
R Sqaure = SSR/SST = 125191.6418/138966.6667

```
> summary(mod)

Call:
lm(formula = slr$Sales ~ slr$Advt)

Residuals:
    Min      1Q  Median      3Q     Max 
-43.101 -27.692  -4.383  23.589  61.547 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -852.08     203.78  -4.181  0.00188 ** 
slr$Advt       19.07       2.00   9.535 2.45e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 37.11 on 10 degrees of freedom
Multiple R-squared:  0.9009,	Adjusted R-squared:  0.891 
F-statistic: 90.93 on 1 and 10 DF,  p-value: 2.454e-06
```

```
> cor(slr$Advt,slr$Sales)
[1] 0.9491666
> 0.949*0.949
[1] 0.900601
> 0.949^2
[1] 0.900601
```



Adjusted R – squared: 0.891

$$Adj\ R\ Square = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)}$$

$$n = nos\ of\ observations = 12, \qquad k = nos\ of\ predictors = 1$$

H	I	J
SST = $(y\text{-}AVGy)^2$	SSR = $(yhat\text{-}AVGyhat)^2$	SSE = $(y\text{-}yhat)^2$
25069.44444	34570.89456	763.9696
35469.44444	21842.62306	1640.25
4669.444444	8205.189306	496.8441
9669.444444	5114.037656	717.1684
136.1111111	1113.723756	2032.2064
1469.444444	22.72905625	1854.1636
3802.777778	1841.053556	353.44
1002.777778	3841.210506	916.2729
1736.111111	3841.210506	410.8729
12469.44444	10023.51381	134.3281
26136.11111	10023.51381	3793.3281
17336.11111	24751.94226	656.3844
138966.6667	125191.6418	13769.2285

$$Adj\ R\ Square = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)}$$

$$Adj\ R\ Square = 1 - \frac{13769.2285/(12-1-1)}{138966.6667/(12-1)}$$

```
> 1-(13769.2285/10)/(138966.6667/11)
[1] 0.8910087
```

Now we need to understand the last line: F-statistics 90.93 on 1 and 10 DF, p-value: 2.45e-06

```
> summary(mod)

Call:
lm(formula = slr$Sales ~ slr$Advt)

Residuals:
    Min      1Q  Median      3Q     Max 
-43.101 -27.692  -4.383  23.589  61.547 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -852.08     203.78  -4.181  0.00188 ** 
slr$Advt       19.07       2.00   9.535 2.45e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 37.11 on 10 degrees of freedom
Multiple R-squared:  0.9009,	Adjusted R-squared:  0.891 
F-statistic: 90.93 on 1 and 10 DF,  p-value: 2.454e-06
```

Degrees of freedom for regression = No of predictors, k=**1**

Degrees of freedom for residuals = n-k-1 = 12-1-1=**10**

MS (Regression) = SS (Regression)/df =125197.46/1 = **125197.46**

ANOVA

	df	SS	MS	F	Significance F
Regression	1	125197.46	125197.46	90.93	0.00
Residual	10	13769.21	1376.92		
Total	11	138966.67			

F = MS (Regression)/MS (Residual) =125197.46/13769.21 = **90.93**

H	I	J
SST = (y-*AVG*y)^2	**SSR =** (yhat-*AVG*yhat)^2	**SSE =** (y-yhat)^2
25069.44444	34570.89456	763.9696
35469.44444	21842.62306	1640.25
4669.444444	8205.189306	496.8441
9669.444444	5114.037656	717.1684
136.1111111	1113.723756	2032.2064
1469.444444	22.72905625	1854.1636
3802.777778	1841.053556	353.44
1002.777778	3841.210506	916.2729
1736.111111	3841.210506	410.8729
12469.44444	10023.51381	134.3281
26136.11111	10023.51381	3793.3281
17336.11111	24751.94226	656.3844
138966.6667	125191.6418	13769.2285

MS (Residual) = SS (Residual)/df =13769.21/10 = **1376.92**

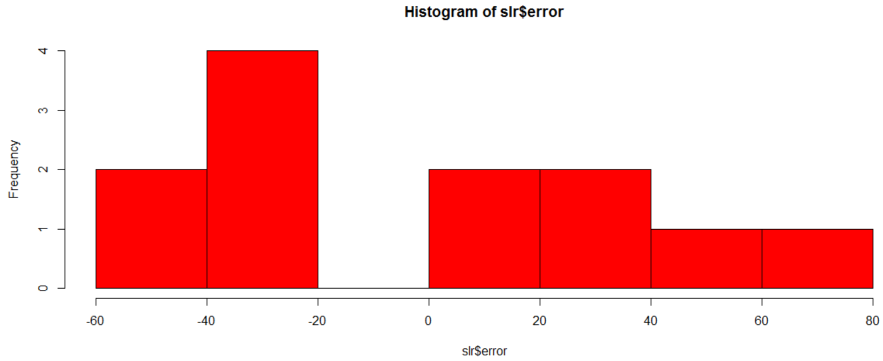
Ho*: *Model is not good
Ha*: *Model is good

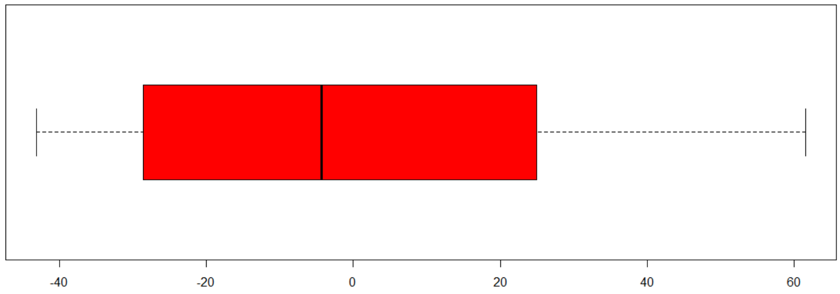
As p-value (Significance value) is 2.454e-06 which is nothing but 0, we reject the null hypothesis and conclude that the model is statistically a good fit model.

Now we have understood all involved calculations and their interpretations! Kudos to all!

Assumption 1: Normality of error

```
> hist(slr$error, col = "red")
> boxplot(slr$error, col = "red", horizontal = T)
```

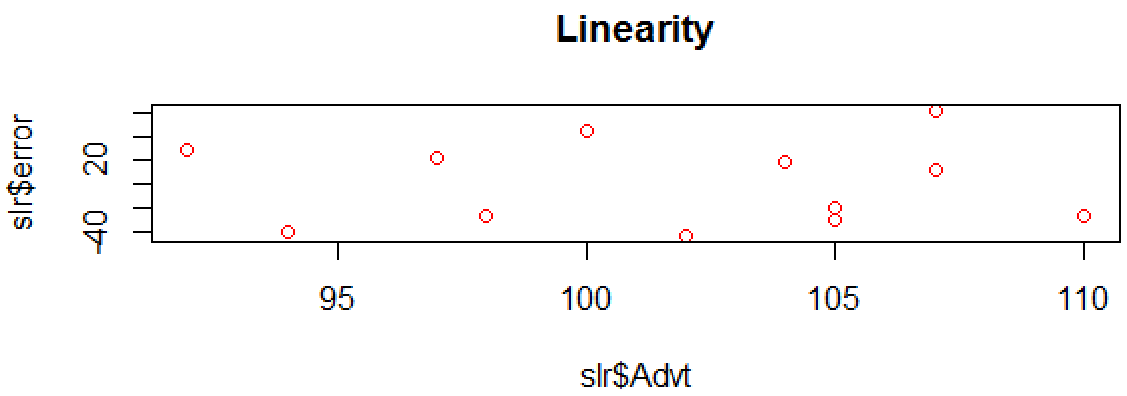




Assumption 2: Linearity

X axis: advertisement; Y axis: error

```
> plot(slr$Advt, slr$error, main = "Linearity", col = "red")
```

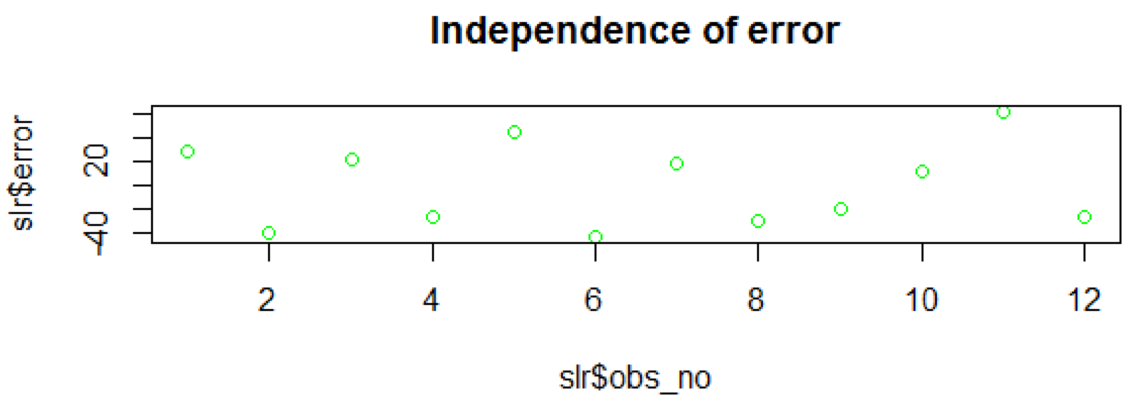


Assumption 3: Independence of error

X axis: observation number; Y axis: error

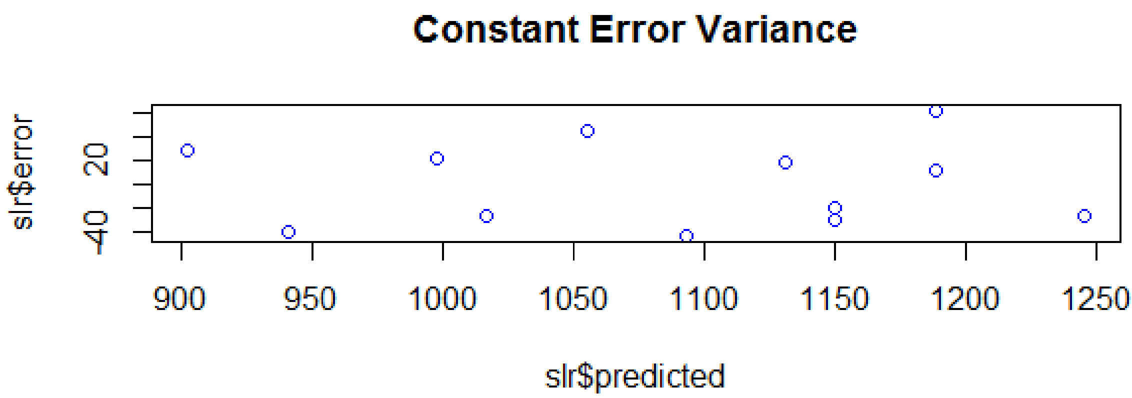
```
> obs_no<- c(1:12)
> slr$obs_no<- NULL
> slr$obs_no<- obs_no
> slr
   Advt Sales predicted     error obs_no
1    92   930  902.3965  27.60349      1
2    94   900  940.5374 -40.53740      2
3    97  1020  997.7487  22.25127      3
4    98   990 1016.8192 -26.81917      4
5   100  1100 1054.9601  45.03994      5
6   102  1050 1093.1009 -43.10094      6
7   104  1150 1131.2418  18.75817      7
8   105  1120 1150.3123 -30.31227      8
9   105  1130 1150.3123 -20.31227      9
10  107  1200 1188.4532  11.54684     10
11  107  1250 1188.4532  61.54684     11
12  110  1220 1245.6645 -25.66449     12
```

```
> plot(slr$obs_no, slr$error, main = "Independence of error", col = "green")
```



Assumption 4: Independence of error

X axis: predicted values; Y axis: error



Multiple Regression

File: **mtcars.csv**

```
> fit<- lm(mpg~ disp+hp+wt+drat, data= mtcars)
> fit

Call:
lm(formula = mpg ~ disp + hp + wt + drat, data = mtcars)

Coefficients:
(Intercept)         disp           hp           wt         drat
  29.148738     0.003815    -0.034784    -3.479668     1.768049

> ki<- lm(mtcars$mpg~ mtcars$disp+mtcars$hp+mtcars$wt+mtcars$drat)
> ki

Call:
lm(formula = mtcars$mpg ~ mtcars$disp + mtcars$hp + mtcars$wt + 
    mtcars$drat)

Coefficients:
(Intercept)  mtcars$disp    mtcars$hp    mtcars$wt  mtcars$drat
  29.148738     0.003815    -0.034784    -3.479668     1.768049
```

```
> summary(fit)

Call:
lm(formula = mpg ~ disp + hp + wt + drat, data = mtcars)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.5077 -1.9052 -0.5057  0.9821  5.6883 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 29.148738   6.293588   4.631  8.2e-05 ***
disp         0.003815   0.010805   0.353  0.72675    
hp          -0.034784   0.011597  -2.999  0.00576 ** 
wt          -3.479668   1.078371  -3.227  0.00327 ** 
drat         1.768049   1.319779   1.340  0.19153    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.602 on 27 degrees of freedom
Multiple R-squared:  0.8376,	Adjusted R-squared:  0.8136 
F-statistic: 34.82 on 4 and 27 DF,  p-value: 2.704e-10
```

Multicollinearity

```
> vif(fit)
    disp       hp       wt     drat 
8.209402 2.894373 5.096601 2.279547 
> vif(fit)>5
 disp    hp    wt  drat 
 TRUE FALSE  TRUE FALSE 
```

$$VIF_i = \frac{1}{1 - {R_i}^2}$$

Experiments

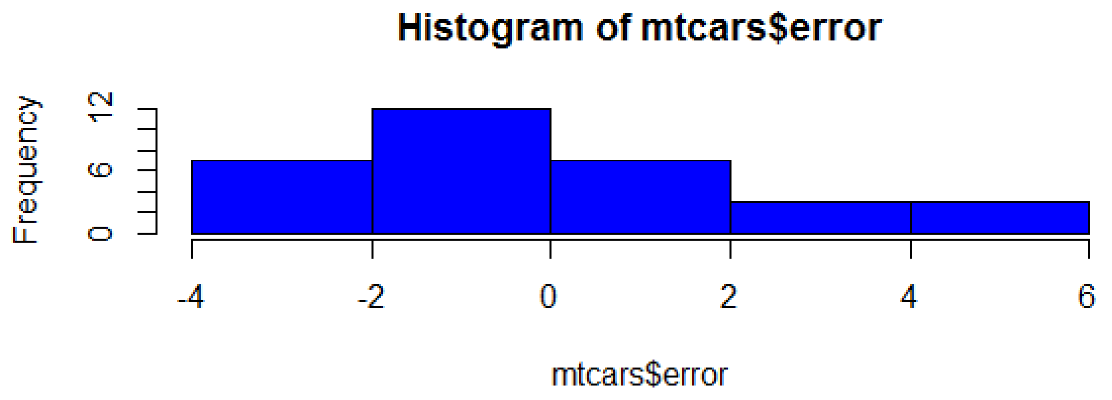
```
> summary(ti)

Call:
lm(formula = mtcars$mpg ~ mtcars$hp + mtcars$wt)

Residuals:
   Min     1Q Median     3Q    Max 
-3.941 -1.600 -0.182  1.050  5.854 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 37.22727    1.59879  23.285  < 2e-16 ***
mtcars$hp   -0.03177    0.00903  -3.519  0.00145 ** 
mtcars$wt   -3.87783    0.63273  -6.129 1.12e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.593 on 29 degrees of freedom
Multiple R-squared:  0.8268,	Adjusted R-squared:  0.8148 
F-statistic: 69.21 on 2 and 29 DF,  p-value: 9.109e-12
```

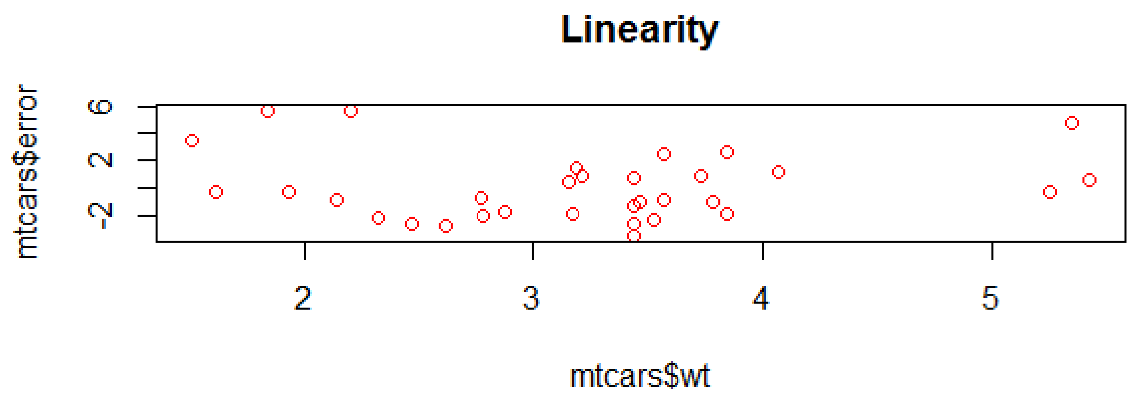
```
> summary(hi)

Call:
lm(formula = mtcars$mpg ~ mtcars$hp)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.7121 -2.1122 -0.8854  1.5819  8.2360 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 30.09886    1.63392  18.421  < 2e-16 ***
mtcars$hp   -0.06823    0.01012  -6.742 1.79e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.863 on 30 degrees of freedom
Multiple R-squared:  0.6024,    Adjusted R-squared:  0.5892 
F-statistic: 45.46 on 1 and 30 DF,  p-value: 1.788e-07
```

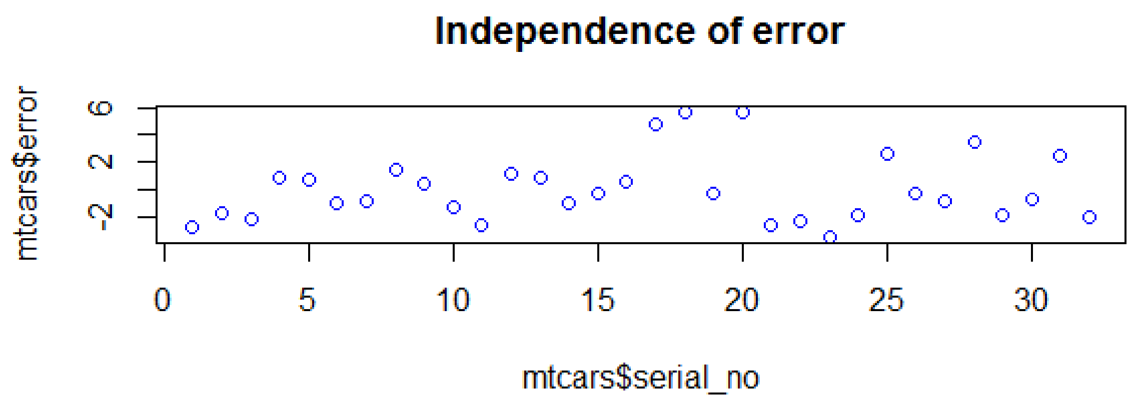
```
> summary(fi)

Call:
lm(formula = mtcars$mpg ~ mtcars$wt)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.5432 -2.3647 -0.1252  1.4096  6.8727 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  37.2851     1.8776  19.858  < 2e-16 ***
mtcars$wt    -5.3445     0.5591  -9.559 1.29e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.046 on 30 degrees of freedom
Multiple R-squared:  0.7528,    Adjusted R-squared:  0.7446 
F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
```

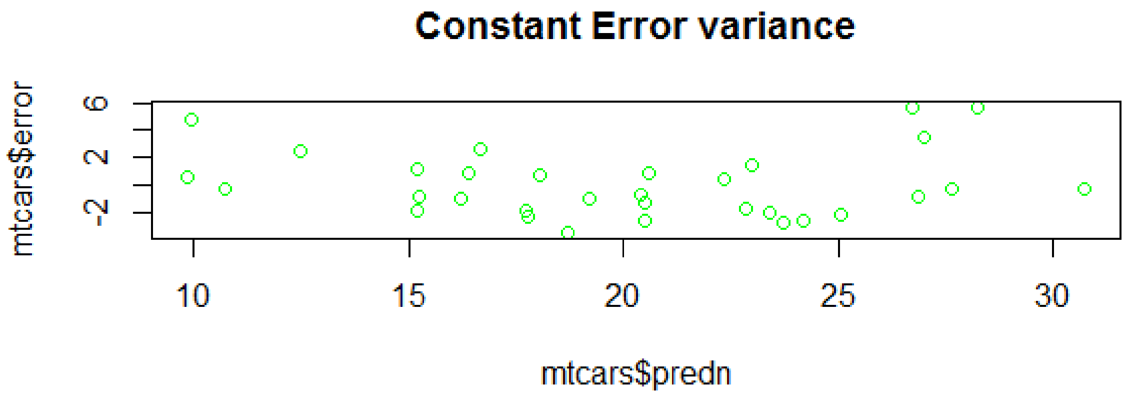
Model	Predictors	R Square
1. fit	disp, hp, wt, drat	0.8376
2. ti	hp, wt	0.8268
3. hi	hp	0.6024
4. fi	wt	0.7528

Assumptions of model: fit








Constant Error variance
mtcars$error
6
2
-2
10
15
20
25
30
mtcars$predn