الاختبارات الاحصائية
Statistical Tests

الاستاذ المساعد

الدكتـــور احمـــد الحســـيني

1- اختبار الخطأ المعياري

الهدف من اجراء الاختبارات الاحصائية هو الكشف عن معنوية المعلمات المقدرة سواء في معادلة الانحدار الخطية المفردة (Simple Regression) او معادلة الانحدار الخطي المتعدد (Multiple Regression) , ويأخذ اختبار الخطأ المعياري الصيغة الاتية :

$$S\alpha = \sqrt{Si^2 \frac{\sum Xi^2}{n\sum xi^2}} \qquad S\beta = \sqrt{\frac{Si^2}{\sum xi^2}}$$

اما في الانحدار المتعدد فتكون صيغته :

$$S\alpha = \sqrt{Si^2 \frac{\ddot{X}1^2(\sum x2^2) + \ddot{X}2^2(\sum x1^2) - 2\ddot{X}1\ddot{X}2(\sum x1x2)}{(\sum x1^2)(\sum x2^2) - (\sum x1x2)^2}}$$

$$S\beta1 = \sqrt{\frac{Si^2(\sum x2^2)}{(\sum x1^2)(\sum x2^2) - (\sum x1x2)^2}}$$

$$S\beta2 = \sqrt{\frac{Si^2(\sum x1^2)}{(\sum x1^2)(\sum x2^2) - (\sum x1x2)^2}}$$

$Si^2 = \frac{\sum_{ei}{}^{2}}{n-k}$

n:Number of Observations

k: number of Parameters

$\sum ei^2 = \sum yi^2 \ - b\sum xiyi$ للانحدار البسيط

$\sum ei^2 = \sum Yi^2 - b1 \sum x1Y - b2 \sum x2Y$ للانحدار المتعدد

$$\sum yi^2 = \sum Yi^2 - \frac{(\sum Yi^2)}{n}$$

ولكي تكون المعلمة معنوية من الناحية الاحصائية يجب ان تكون :

$S\alpha$, $S\beta > \frac{1}{2}\alpha, \frac{1}{2}\beta$

وفي هذه الحالة يتم قبول فرض العدم

$Hi \neq 0$

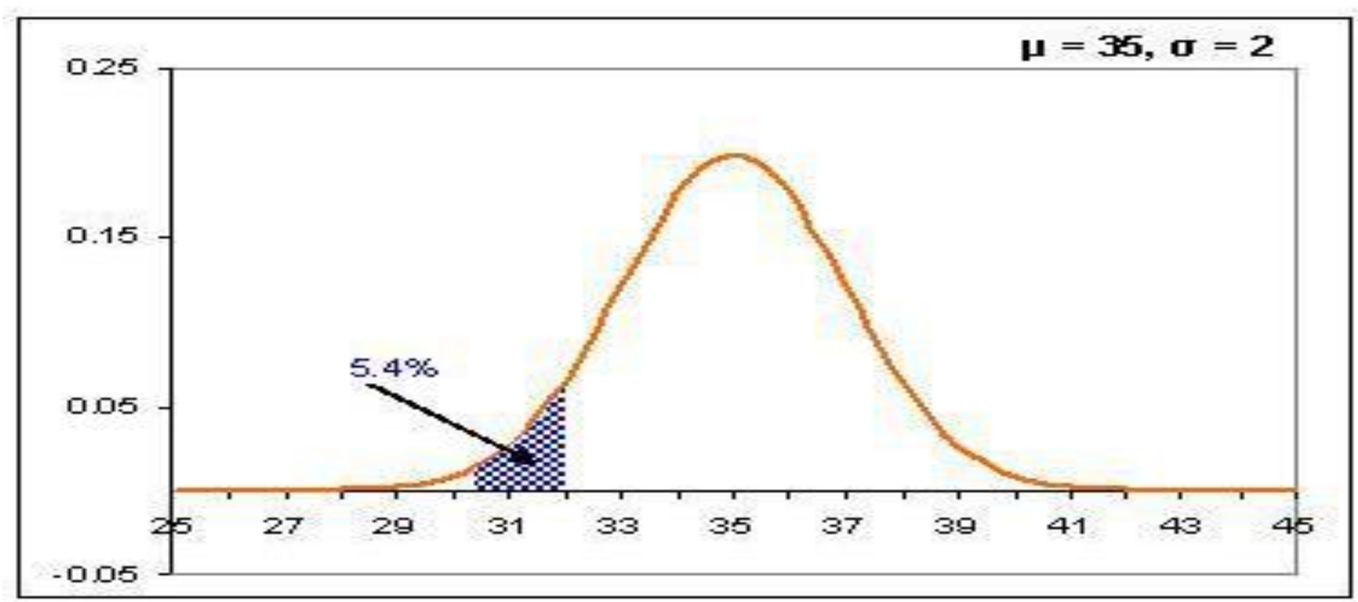
واهمال الفرض البديل

$Hi=0$

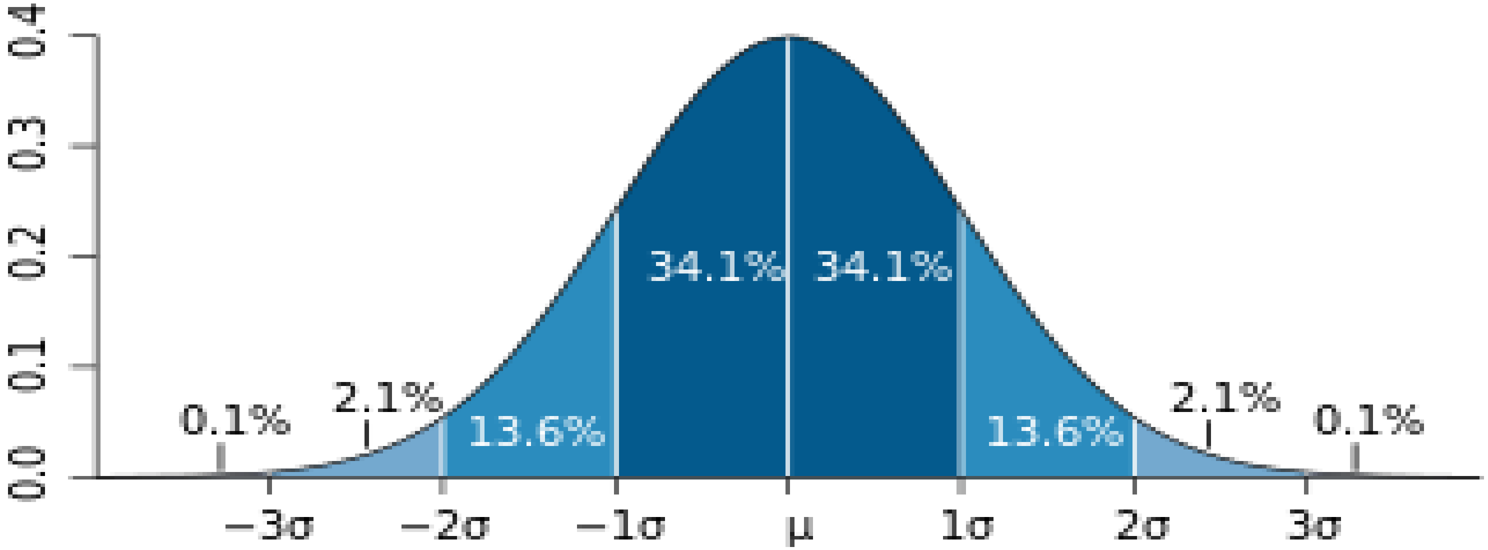
اختبار t
t- Test

ويسمى في بعض الاحيان اختبار الطالب ومن خصائص هذا الاختبار

1- انه توزيع ناقوسي جرسي



2- يقترب هذا التوزيع من التوزيع الطبيعي مع زيادة حجم العينة



وتكون صيغة الاختبار كالاتي :

$$t\alpha^* = \frac{\alpha}{S\alpha}$$

$$t\beta^* = \frac{\beta}{S\beta}$$

Simple Regression

$$t\beta 1 = \frac{\beta 1}{S\beta 1}$$

$$t\beta 2 = \frac{\beta 2}{S\beta 2}$$

Multiple Regression

فاذا كانت :

$$t\alpha^*, t\beta^* > t$$

فهذا يعني قبول الفرضية البديلة واهمال فرض العدم

تقبل $Hi=0$

تهمل $Hi \neq 0$

ملاحظة : ان اختبار t لايوضح مدى اقتراب او ابتعاد المعلمات المقدرة (α , β) عن المعلمات الحقيقية , بل يوضح المقدار الذي تختلف فيه المعلمات المقدرة عن الصفر , ولمعرفة مدى اقتراب المعلمات المقدرة عن المعلمات الحقيقية نلجأ الى مايسمى فترات الثقة (**Trust Period**)

فترة الثقة

وتشير فترات الثقة الى المدى الذي تبتعد او تقترب فيه المعلمات المقدرة عن معلمات المجتمع الاحصائي (المعلمات الحقيقية) , ويتم التعبير عن فترات الثقة بالصيغة الاتية :

$$\rho\ [\ \alpha - ts\alpha \lesssim\ a \lesssim\ \alpha + ts\alpha\ ]\ (\ 1-\ \alpha\ )$$

$$\rho\ [\beta - ts\beta \lesssim b \lesssim \beta + ts\beta]\ (\ 1-\ \alpha\ )$$

حيث ان :

P: الاحتمالية

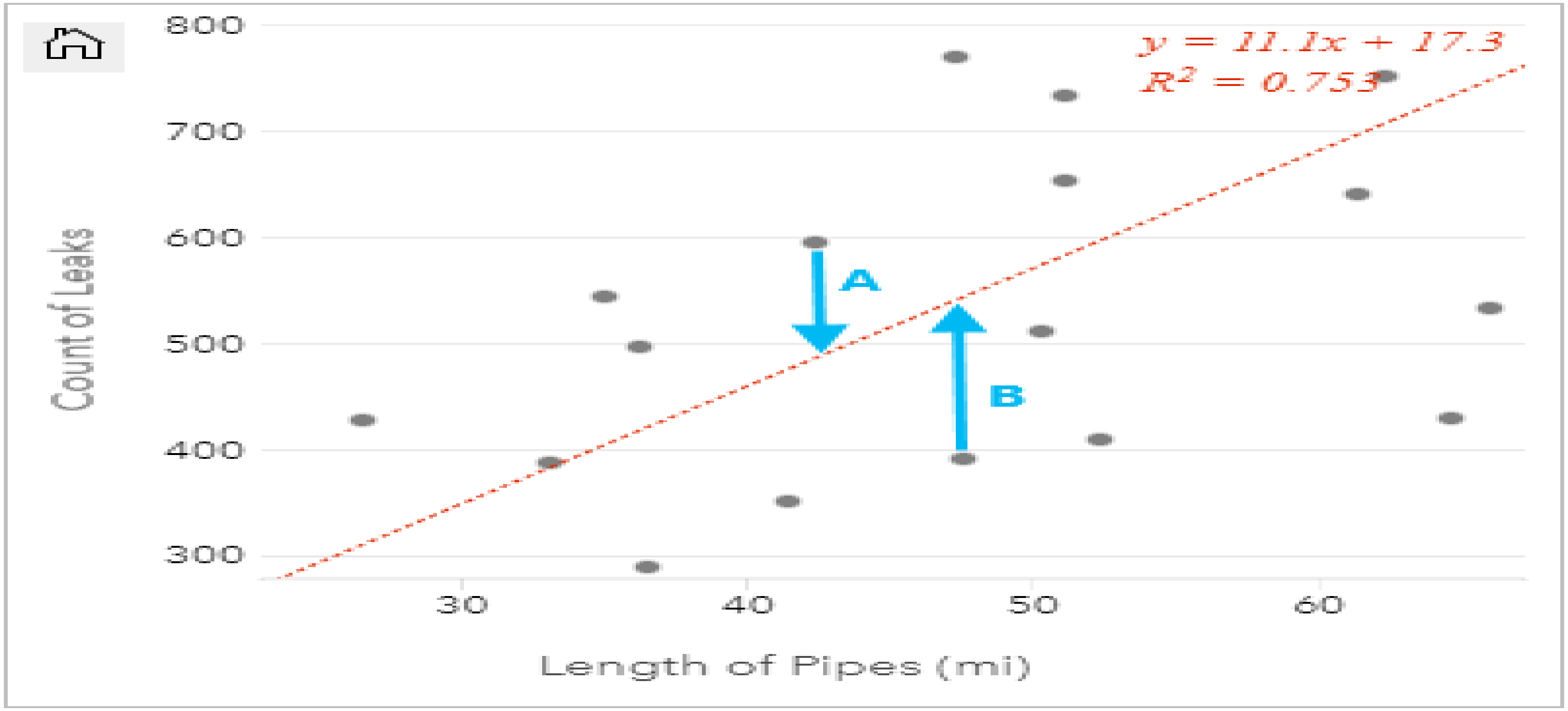
$s\alpha\ ,\ s\beta$: الخطأ المعياري

α: مستوى المعنوية

$(\ 1-\alpha\ )$: درجة الثقة

معامل التحديد (R^2)
Determination Coefficient

بعد تقدير النموذج القياسي يرغب الباحث في التعرف كم تقترب او تبتعد المشاهدات الحقيقية عن خط الانحدار المقدر , ويسمى هذا الاقتراب او الابتعاد بجودة التوفيق (Goodness of Fit)



وان مقياس هذه الجودة هو معامل التحديد (R^2) , ويسمى احيانا معامل التفسير لانه يفسر كم من المتغيرات المؤثرة في المتغير التابع يتم توضيحها عن طريق المتغيرات المستقلة ويسمى ايضا معامل التوضيح (Explain Coefficient) .

ولتوضيح صيغة معامل التوضيح :

$$R^2 = \frac{\text{التباين الموضح}}{\text{التباين الكلي}}$$

او يمكن التعبير عنه بالصياغات الاتية :

$$R^2 = 1 - \frac{RSS}{TSS}$$

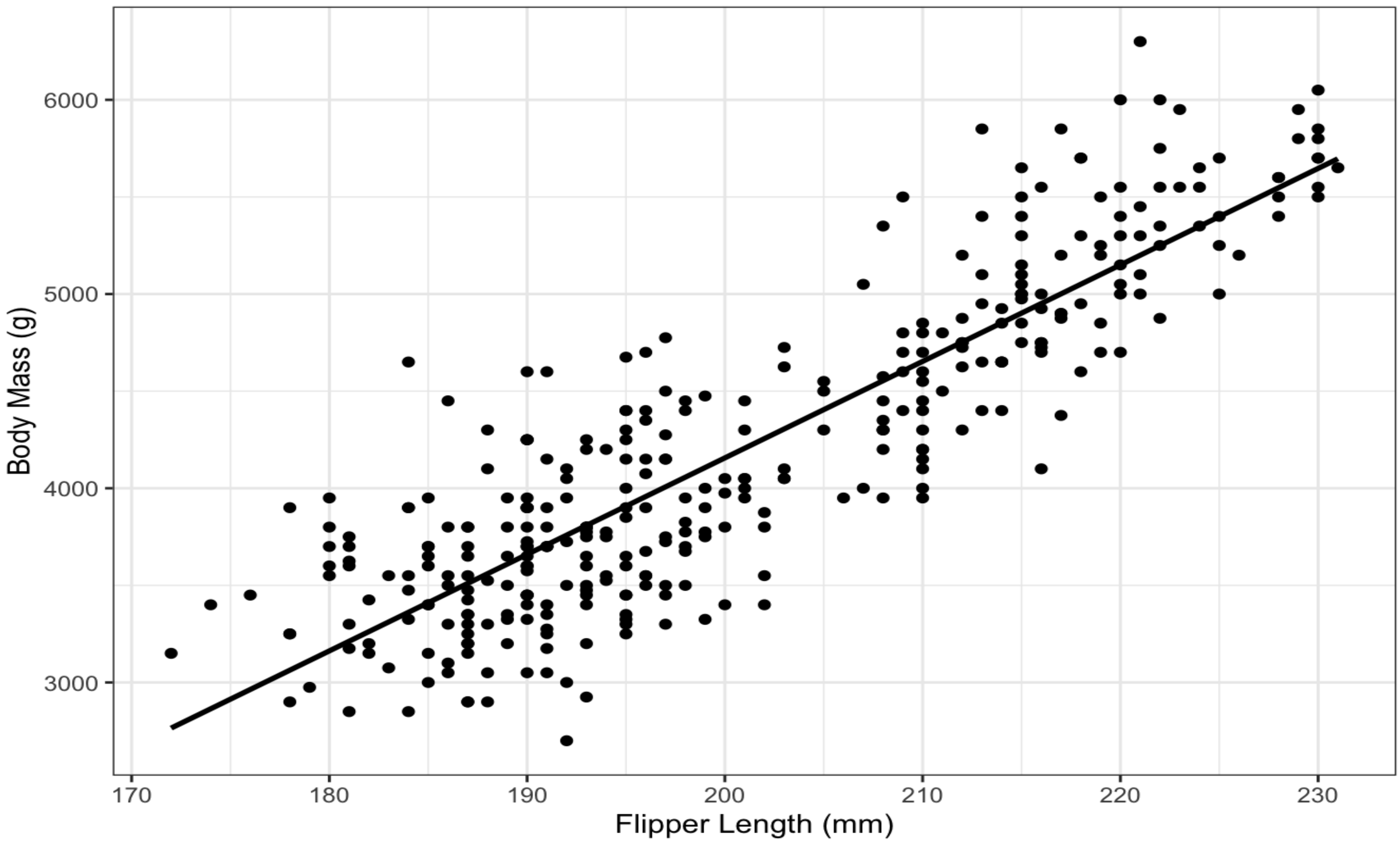
$$R^2 = \frac{\beta \sum xiyi}{\sum yi^2}$$

$$R^2 = \frac{\sum \hat{y}^2}{\sum y^2}$$

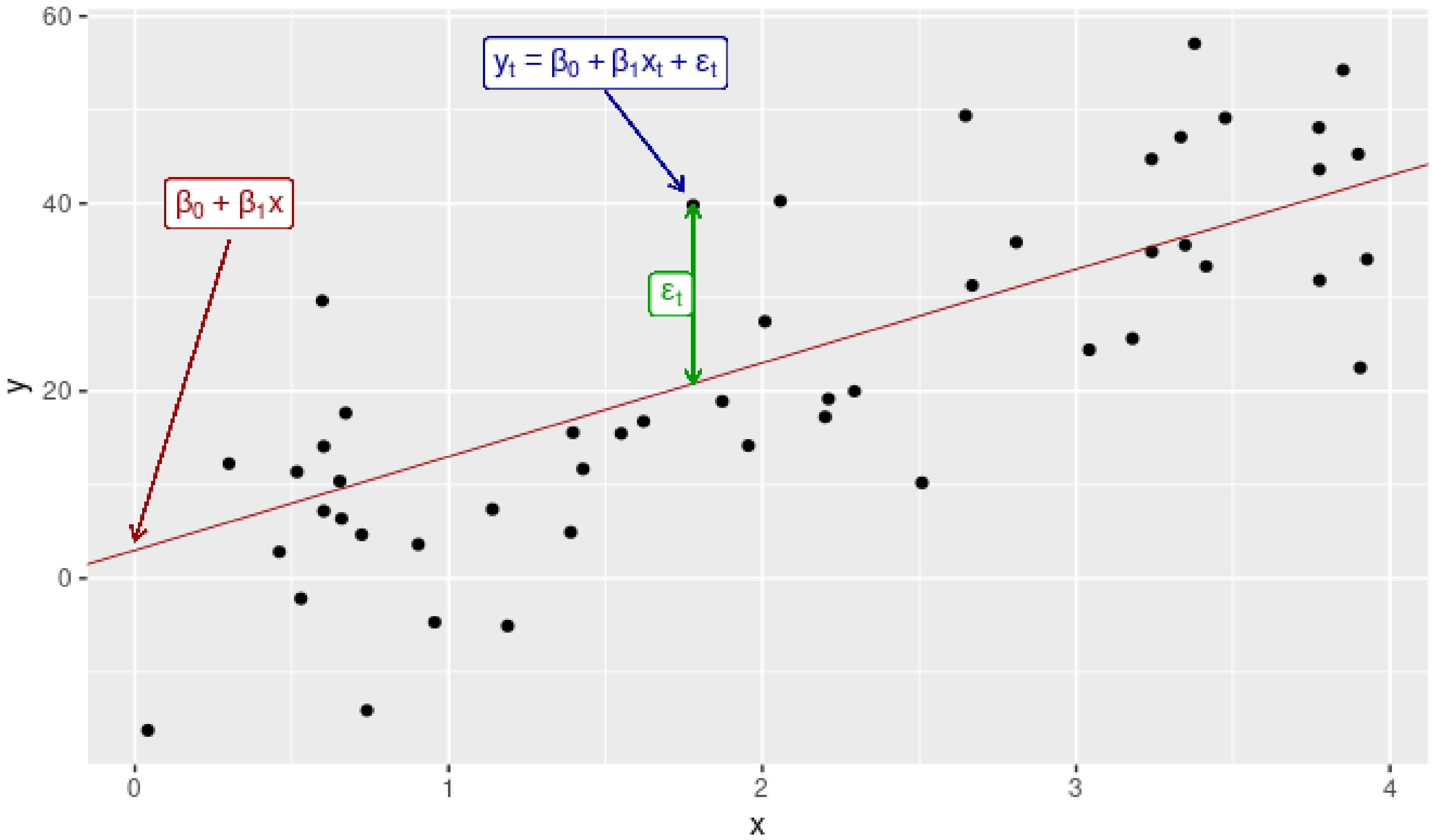
Simple Regression

$$R^2 = \frac{\beta 1 \sum x1y + \beta 2 \sum x2y}{\sum yi^2}$$

$$\sum yi^2 = \sum Yi^2 + \frac{(\sum Yi)^2}{n}$$



6000
5000
4000
3000
Body Mass (g)
170
180
190
200
210
220
230
Flipper Length (mm)



$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$
$\beta_0 + \beta_1 x$
ε_t
60
40
20
0
y
0
1
2
3
4
x

اختبار F
F- Test

F-Test Formula

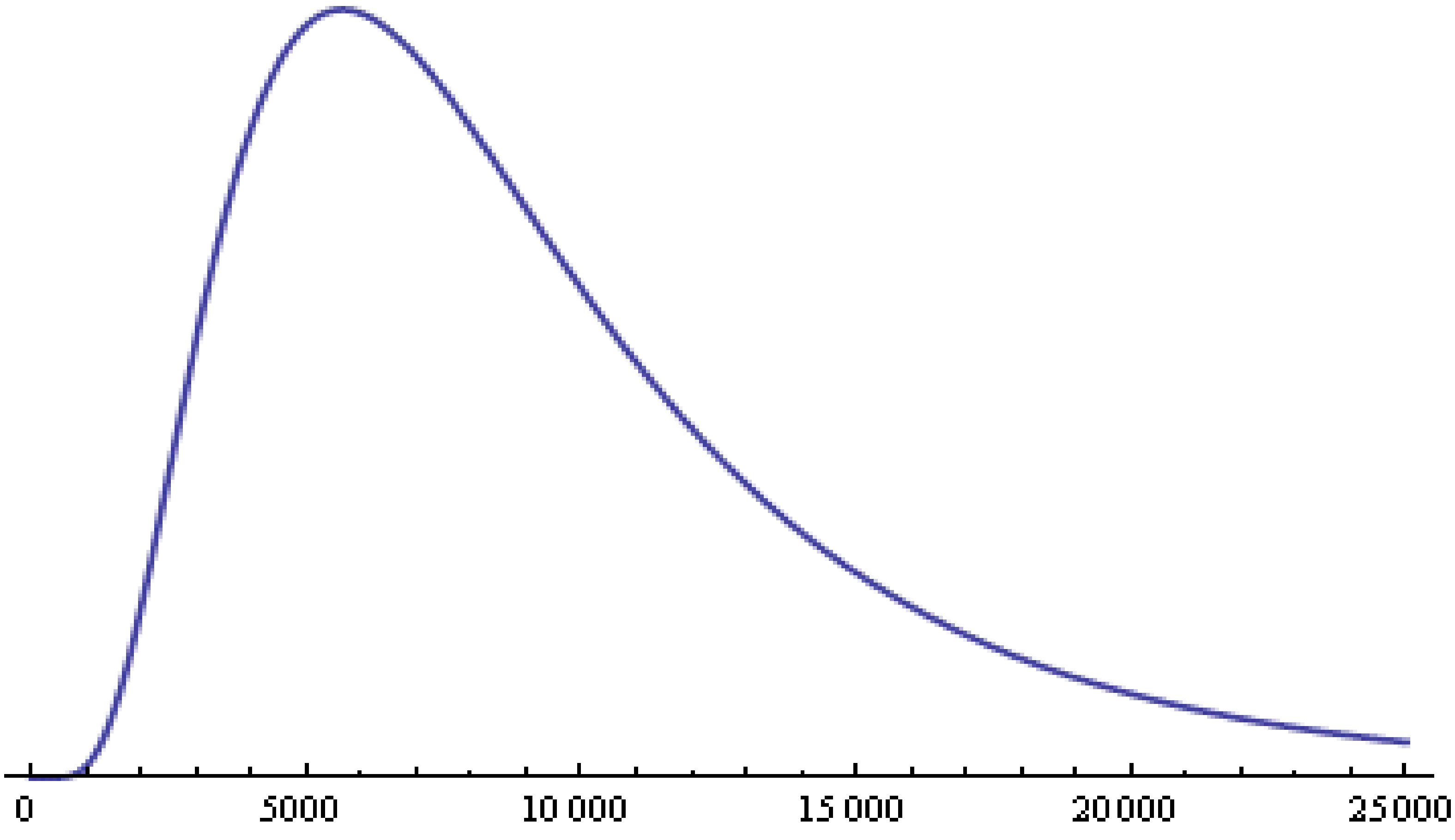
$$\text{F Value} = \frac{\text{Larger Sample Variance}}{\text{Smaller Sample Variance}} = \frac{\sigma_1^2}{\sigma_2^2}$$

ANOVA TABLE

Source of Variation	Sum of Squares (SS)	Degrees of freedom (df)	Mean Square (MS)	F
Factor (Between)	SS_{Factor}	k-1	$MS_{Factor=}$ $SS_{Factor}/k-1$	$F= MS_{Factor}/MS_{Error}$
Error (Within)	SS_{Error}	n-k	$MS_{Error=}$ $SS_{Error}/n-k$	
Total	SS_{Total}	n-1		

وبحسب الجدول السابق يتم تحليل التباين الكلي (TSS) الى مركبتين هما التباين الموضح (ESS) والتباين غير الموضح (USS) .

ويستند اختبار F الى توزيع F وهذا التوزيع هو توزيع منحرف يتراوح بين (0,α)


0
5000
10 000
15 000
20 000
25 000

ويتم مقارنة قيمة *F المحتسبة مع F الجدولية وفق مستوى معنوية معين ودرجة حرية (df) مقدارها (K-1) للبسط و (n-K) للمقام فاذا كانت :

$F^* > F$

يتم قبول الفرض البديل

$Hi \neq 0$

ويهمل فرض العدم

$Hi = 0$

Exercise

Daily electricity demand for Victoria, Australia, during 2014 is contained in Elec daily. The data for the first 20 days can be obtained as follows.

A- Plot the data and find the regression model for Demand with temperature as an explanatory variable. Why is there a positive relationship?

B- Produce a residual plot. Is the model adequate?
C- Are there any outliers or influential observations?

PROBLEM 2: Simple Linear Regression.

Background: Work through Matlab Exercise 2 (posted with this assignment). Steps #8 and #9 pertain to this problem.

Consider the following sample of ordered pairs (x, y) where
y = the purity of oxygen produced in a chemical distillation process, and
x = the percentage of hydrocarbons that are present in the main condenser of the distillation unit.

Observation Number	Hydrocarbon Level; x (in %)	Purity y (in %)	Observation Number	Hydrocarbon Level; x (in %)	Purity y (in %)
1	0.99	90.01	11	1.19	93.65
2	1.02	89.05	12	1.15	92.52
3	1.15	91.43	13	0.98	90.56
4	1.29	93.74	14	1.01	89.54
5	1.46	96.73	15	1.11	89.85
6	1.36	94.45	16	1.2	90.39
7	0.87	87.59	17	1.26	93.25
8	1.23	91.77	18	1.32	93.41
9	1.55	99.42	19	1.43	94.98
10	1.4	93.65	20	0.95	87.33

0. Enter above data either as a single 20x2 matrix or as two 20x1 vectors. Be sure to preserve the ordering of the pairs.

a. Create a scatterplot of this data. Label both the x and y axis' and title the graph "scatterplot" Labels can be added on the graph itself by choosing "Insert" drop down menu.

b. Determine the correlation coefficient.
The corrcoef(x,y) command produces a 2x2 matrix with 1's down the principle diagonal Cell 1,1 provides the correlation of x with itself and cell 2,2 is the correlation of y with itself (both of which are 1). Cell 2,1 provides the correlation between x and y, cell 1,2 provides the correlation between y and x (both, of course would be the same)

c. Find the equation of the least square regression line, $\hat{y} = b_0 + b_1 x$ and determine the p-value of the slope term b_1. To generate the line and obtain the **p-value** you should use the fitlm command (refer to Matlab exercise 2). Using the 5% Significance Level, if the p-value < .05, then the linear model is statistically significant.

d. You can also display the regression equation on your scatterplot by selecting "Tools" drop down menu, select "basic fitting", linear.

e. If your regression model is statistically significant then use it to estimate Oxygen purity (in %) when the hydrocarbon level in the main condenser of the distillation unit is 1%.

Multiple Linear Regression

Multiple linear regression is a method we can use to quantify the relationship between two or more predictor variables and a response variable.

Multiple regression formulas analyze the relationship between dependent and multiple independent variables. For example, the equation Y represents the formula is equal to a plus bX1 plus cX2 plus dX3 plus E where Y is the dependent variable, and X1, X2, and X3 are independent variables. A is the intercept, b, c, and d are the slopes, and E is the residual value.

Multiple Regression Formula

$$Y = mx1 + mx2 + mx3 + b$$

Where,

Y= the dependent variable of the regression

M= slope of the regression

X1=first independent variable of the regression

The X2=second independent variable of the regression

The X3=third independent variable of the regression

B= constant

Example: Multiple Linear Regression

Suppose we have the following dataset with one response variable y and two predictor variables X1 and X2:

y	X_1	X_2
140	60	22
155	62	25
159	67	24
179	70	20
192	71	15
200	72	14
212	75	14
215	78	11

Use the following steps to fit a multiple linear regression model to this dataset.

Step 1: Calculate $X1^2$, $X2^2$, X1Y, X2Y and X1X2.

	y	X_1	X_2
	140	60	22
	155	62	25
	159	67	24
	179	70	20
	192	71	15
	200	72	14
	212	75	14
	215	78	11
Mean	181.5	69.375	18.125
Sum	1452	555	145

	X_1^2	X_2^2	X_1y	X_2y	X_1X_2
	3600	484	8400	3080	1320
	3844	625	9610	3875	1550
	4489	576	10653	3816	1608
	4900	400	12530	3580	1400
	5041	225	13632	2880	1065
	5184	196	14400	2800	1008
	5625	196	15900	2968	1050
	6084	121	16770	2365	858
Sum	38767	2823	101895	25364	9859

Step 2: Calculate Regression Sums.
Next, make the following regression sum calculations:

$\Sigma X1^2 = \Sigma X1^2 - (\Sigma X1)^2 / n = 38{,}767 - (555)^2 / 8 = 263.875$

$\Sigma X2^2 = \Sigma X2^2 - (\Sigma X2)^2 / n = 2{,}823 - (145)^2 / 8 = 194.875$

$\Sigma X1Y = \Sigma X1Y - (\Sigma X1 \Sigma Y) / n = 101{,}895 - (555 * 1{,}452) / 8 = 1{,}162.5$

$\Sigma X2Y = \Sigma X2Y - (\Sigma X2 \Sigma Y) / n = 25{,}364 - (145 * 1{,}452) / 8 = -953.5$

$\Sigma X1X2 = \Sigma X1X2 - (\Sigma X1 \Sigma X2) / n = 9{,}859 - (555 * 145) / 8 = -200.375$

	y	X_1	X_2
	140	60	22
	155	62	25
	159	67	24
	179	70	20
	192	71	15
	200	72	14
	212	75	14
	215	78	11
Mean	181.5	69.375	18.125
Sum	1452	555	145

	X_1^2	X_2^2	X_1y	X_2y	X_1X_2
	3600	484	8400	3080	1320
	3844	625	9610	3875	1550
	4489	576	10653	3816	1608
	4900	400	12530	3580	1400
	5041	225	13632	2880	1065
	5184	196	14400	2800	1008
	5625	196	15900	2968	1050
	6084	121	16770	2365	858
Sum	38767	2823	101895	25364	9859
Reg Sums	263.875	194.875	1162.5	-953.5	-200.375

Step 3: Calculate b0, b1, and b2.

The formula to calculate b1 is:

$[(\Sigma X2^2)(\Sigma X1Y) - (\Sigma X1X2)(\Sigma X2Y)] / [(\Sigma X1^2)(\Sigma X2^2) - (\Sigma X1X2)^2]$

Thus, b1 = [(194.875)(1162.5) – (-200.375)(-953.5)] / [(263.875) (194.875) – (-200.375)2] = 3.148

The formula to calculate β2 is: $[(\Sigma X1^2)(\Sigma X2Y) - (\Sigma X1X2)(\Sigma X1Y)] / [(\Sigma X1^2)(\Sigma X2^2) - (\Sigma X1X2)^2]$

Thus, β2 = [(263.875)(-953.5) – (-200.375)(1152.5)] / [(263.875) (194.875) – (-200.375)2] = -1.656

The formula to calculate α is: $\bar{Y} - \beta 1\overline{X1} - \beta 2\overline{X2}$
Thus, α= 181.5 – 3.148(69.375) – (-1.656)(18.125) = -6.867

Step 5: Placeα, β1, and β2 in the estimated linear regression equation.

The estimated linear regression equation is: $\hat{y} = \alpha + \beta 1 * X1 + \beta 2 * X2$

In our example, it is $\hat{y} = -6.867 + 3.148X1 - 1.656X2$

How to Interpret a Multiple Linear Regression Equation?

Here is how to interpret this estimated linear regression equation:

$$\hat{y} = -6.867 + 3.148X1 - 1.656X2$$

$\alpha = -6.867$. When both predictor variables are equal to zero, the mean value for y is -6.867.

$\beta 1 = 3.148$. A one unit increase in X1 is associated with a 3.148 unit increase in y, on average, assuming X2 is held constant.

$\beta 2 = -1.656$. A one unit increase in X2 is associated with a 1.656 unit decrease in y, on average, assuming X1 is held constant.

Example

Let us try and understand the concept of multiple regression analysis with the help of another example. Now, let us find out the relation between the salary of a group of employees in an organization, the number of years of experience, and the age of the employees.

Income ($)	Age	Experience (In Years)
26315	18	5
39493	20	7
37209	22	8
24380	23	6
25751	23	7
44629	25	5
38616	2	8
33305	28	6
36848	29	5
42551	32	7
25700	37	9
37303	41	6
24659	46	7
32617	49	8
35771	53	6

	C	D	E	F	G	H	I	J	K
18	**SUMMARY OUTPUT**								
19									
20	Regression Statistics								
21	Multiple R	0.209324041							
22	R Square	0.043816554							
23	Adjusted R Square	-0.115547354							
24	Standard Error	7178.186331							
25	Observations	15							
26									
27	ANOVA								
28		df	SS	MS	F	Significance F			
29	Regression	2	28333987.64	14166994	0.2749465	0.7642709			
30	Residual	12	618316308.1	51526359					
31	Total	14	646650295.7						
32									
33		Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
34	Intercept	41308.23971	11234.93105	3.676768	0.0031675	16829.42783	65787.0516	16829.42783	65787.0516
35	X Variable 1	-71.43027042	142.5070357	-0.50124	0.6252701	-381.926428	239.065887	-381.926428	239.065887
36	X Variable 2	-824.7583445	1554.894782	-0.53043	0.6054928	-4212.58304	2563.06635	-4212.58304	2563.06635
37									
38		-37019974.13							

The regression equation for the above example will be

$y = \alpha + \beta 1X1 + \beta 2X2$

$y = 41308.23 - 71.41308X1 - 824.758X2$

In this particular example, we will see which variable is the dependent variable and which variable is the independent variable. The dependent variable in this regression equation is the salary, and the independent variables are the experience and age of the employees.

Ŕ²

$$Adjusted\ R^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

Where

R^2 Sample R-Squared

N Total Sample Size

p Number of independent variable

Adjusted r-squared formulas

There seem to exist several formulas to calculate Adjusted R-squared.

- Wherry's formula: $1-(1-R^2)\frac{(n-1)}{(n-v)}$
- McNemar's formula: $1-(1-R^2)\frac{(n-1)}{(n-v-1)}$
- Lord's formula: $1-(1-R^2)\frac{(n+v-1)}{(n-v-1)}$
- Stein's formula: $1-[\frac{(n-1)}{(n-k-1)}\frac{(n-2)}{(n-k-2)}\frac{(n+1)}{n}](1-R^2)$

Why use Adjusted R-Squared and not R-Squared?

Assume, you have a random variable having a casual relationship with the dependent variable. The addition of such a random variable to the model will still improve the model's R-squared statistic. However, the Adjusted R Squared statistic will decrease and penalize the model if the explanatory variable does not contribute to the model. It is evident from the Adjusted R-Squared formula.

Example

Let's say we have two data sets, X & Y, each containing 20 random data points. First, calculate the Adjusted R Squared for the data set X & Y.

	A	B	H
3			
4	X	Y	
5	60	43	
6	67	17	
7	75	14	
8	88	97	
9	99	93	
10	29	63	
11	14	21	
12	92	100	
13	43	63	
14	94	86	
15	19	69	
16	30	9	
17	6	34	
18	3	5	
19	18	74	
20	70	96	
21	27	49	
22	5	89	
23	92	19	
24	52	34	
25			

Mean is calculated as:

	A	B	C
4	X	Y	
5	60	43	
6	67	17	
7	75	14	
8	88	97	
9	99	93	
10	29	63	
11	14	21	
12	92	100	
13	43	63	
14	94	86	
15	19	69	
16	30	9	
17	6	34	
18	3	5	
19	18	74	
20	70	96	
21	27	49	
22	5	89	
23	92	19	
24	52	34	

Mean is calculated as

		X_m	Y_m
29	Formula	=AVERAGE(A5:A24)	=AVERAGE(B5:B24)
30	Mean	49.2	53.8

Mean of Data Set X = 49.2

Mean of Data Set Y = 53.8

We need to calculate the difference between the data points and the mean value.

C5 | =A5-B29

	A	B	C	H
3				
4	X	Y	$X - X_m$	
5	60	43	10.9	
6	67	17		
7	75	14		
8	88	97		
9	99	93		
10	29	63		
11	14	21		
12	92	100		
13	43	63		
14	94	86		
15	19	69		
16	30	9		
17	6	34		
18	3	5		
19	18	74		
20	70	96		
21	27	49		
22	5	89		
23	92	19		
24	52	34		
25				

Similarly, calculate for all the data set of X.

	A	B	C
3			
4	X	Y	$X - X_m$
5	60	43	10.9
6	67	17	17.9
7	75	14	25.9
8	88	97	38.9
9	99	93	49.9
10	29	63	-20.2
11	14	21	-35.2
12	92	100	42.9
13	43	63	-6.2
14	94	86	44.9
15	19	69	-30.2
16	30	9	-19.2
17	6	34	-43.2
18	3	5	-46.2
19	18	74	-31.2
20	70	96	20.9
21	27	49	-22.2
22	5	89	-44.2
23	92	19	42.9
24	52	34	2.9
25			

Similarly, calculate it for data set Y also.

X	Y	$X - X_m$	$Y - Y_m$
60	43	10.9	-10.8
67	17	17.9	-36.8
75	14	25.9	-39.8
88	97	38.9	43.3
99	93	49.9	39.3
29	63	-20.2	9.3
14	21	-35.2	-32.8
92	100	42.9	46.3
43	63	-6.2	9.3
94	86	44.9	32.3
19	69	-30.2	15.3
30	9	-19.2	-44.8
6	34	-43.2	-19.8
3	5	-46.2	-48.8
18	74	-31.2	20.3
70	96	20.9	42.3
27	49	-22.2	-4.8
5	89	-44.2	35.3
92	19	42.9	-34.8
52	34	2.9	-19.8

Calculate the square of the difference for both the data sets X and Y.

X	Y	$X - X_m$	$Y - Y_m$	$(X - X_m)^2$	$(Y - Y_m)^2$
60	43	10.9	-10.8	117.7	115.6
67	17	17.9	-36.8	318.6	1350.6
75	14	25.9	-39.8	668.2	1580.1
88	97	38.9	43.3	1509.3	1870.6
99	93	49.9	39.3	2485.0	1540.6
29	63	-20.2	9.3	406.0	85.6
14	21	-35.2	-32.8	1235.5	1072.6
92	100	42.9	46.3	1836.1	2139.1
43	63	-6.2	9.3	37.8	85.6
94	86	44.9	32.3	2011.5	1040.1
19	69	-30.2	15.3	909.0	232.6
30	9	-19.2	-44.8	366.7	2002.6
6	34	-43.2	-19.8	1861.9	390.1
3	5	-46.2	-48.8	2129.8	2376.6
18	74	-31.2	20.3	970.3	410.1
70	96	20.9	42.3	434.7	1785.1
27	49	-22.2	-4.8	490.6	22.6
5	89	-44.2	35.3	1949.2	1242.6
92	19	42.9	-34.8	1836.1	1207.6
52	34	2.9	-19.8	8.1	390.1

Multiply the difference in X with Y.

	A	B	C	D	E	F	G
3							
4	X	Y	$X - X_m$	$Y - Y_m$	$(X - X_m)^2$	$(Y - Y_m)^2$	$(X - X_m) * (Y - Y_m)$
5	60	43	10.9	-10.8	117.7	115.6	-116.6
6	67	17	17.9	-36.8	318.6	1350.6	-656.0
7	75	14	25.9	-39.8	668.2	1580.1	-1027.5
8	88	97	38.9	43.3	1509.3	1870.6	1680.3
9	99	93	49.9	39.3	2485.0	1540.6	1956.6
10	29	63	-20.2	9.3	406.0	85.6	-186.4
11	14	21	-35.2	-32.8	1235.5	1072.6	1151.2
12	92	100	42.9	46.3	1836.1	2139.1	1981.8
13	43	63	-6.2	9.3	37.8	85.6	-56.9
14	94	86	44.9	32.3	2011.5	1040.1	1446.4
15	19	69	-30.2	15.3	909.0	232.6	-459.8
16	30	9	-19.2	-44.8	366.7	2002.6	857.0
17	6	34	-43.2	-19.8	1861.9	390.1	852.2
18	3	5	-46.2	-48.8	2129.8	2376.6	2249.8
19	18	74	-31.2	20.3	970.3	410.1	-630.8
20	70	96	20.9	42.3	434.7	1785.1	880.9
21	27	49	-22.2	-4.8	490.6	22.6	105.2
22	5	89	-44.2	35.3	1949.2	1242.6	-1556.3
23	92	19	42.9	-34.8	1836.1	1207.6	-1489.0
24	52	34	2.9	-19.8	8.1	390.1	-56.3
25							

Correlation Coefficient is calculated using the formula given below

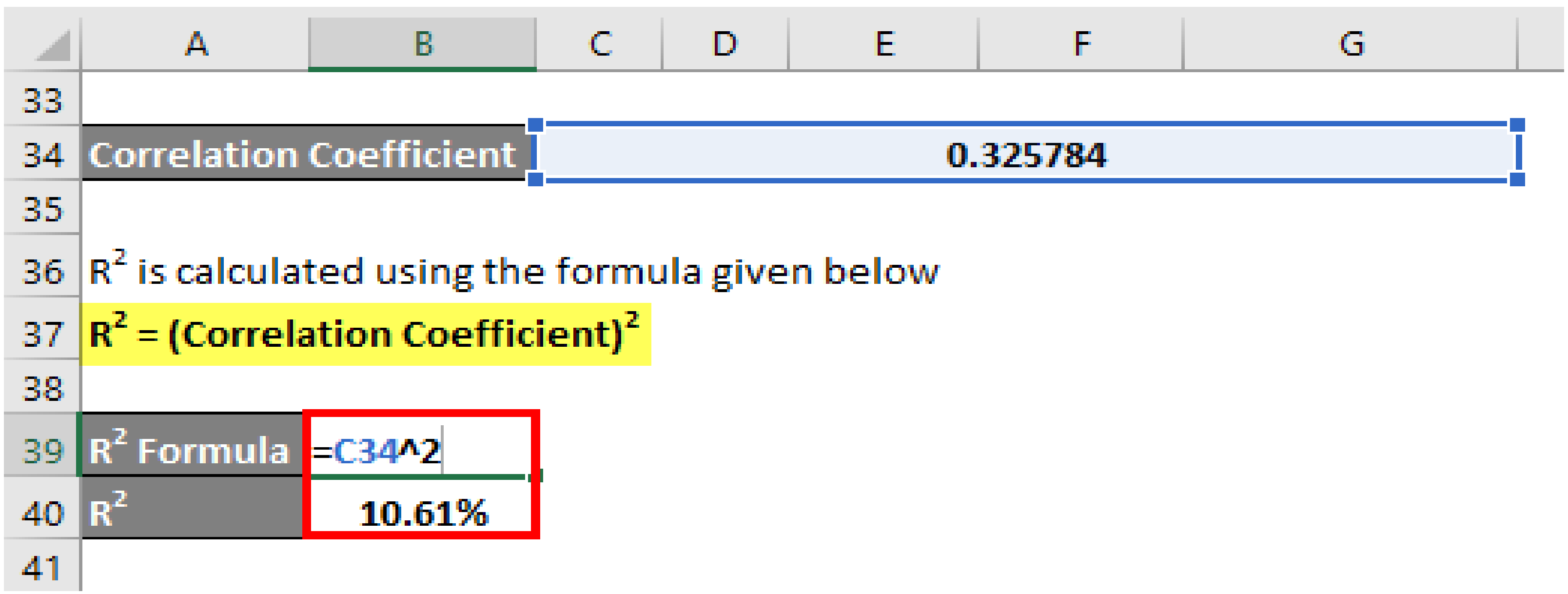
Correlation Coefficient = $\Sigma [(X - Xm) * (Y - Ym)] / \sqrt{[\Sigma (X - Xm)^2 * \Sigma (Y - Ym)^2]}$

	A	B	C	D	E	F	G
3							
4	X	Y	$X - X_m$	$Y - Y_m$	$(X - X_m)^2$	$(Y - Y_m)^2$	$(X - X_m) * (Y - Y_m)$
5	60	43	10.9	-10.8	117.7	115.6	-116.6
6	67	17	17.9	-36.8	318.6	1350.6	-656.0
7	75	14	25.9	-39.8	668.2	1580.1	-1027.5
8	88	97	38.9	43.3	1509.3	1870.6	1680.3
9	99	93	49.9	39.3	2485.0	1540.6	1956.6
10	29	63	-20.2	9.3	406.0	85.6	-186.4
11	14	21	-35.2	-32.8	1235.5	1072.6	1151.2
12	92	100	42.9	46.3	1836.1	2139.1	1981.8
13	43	63	-6.2	9.3	37.8	85.6	-56.9
14	94	86	44.9	32.3	2011.5	1040.1	1446.4
15	19	69	-30.2	15.3	909.0	232.6	-459.8
16	30	9	-19.2	-44.8	366.7	2002.6	857.0
17	6	34	-43.2	-19.8	1861.9	390.1	852.2
18	3	5	-46.2	-48.8	2129.8	2376.6	2249.8
19	18	74	-31.2	20.3	970.3	410.1	-630.8
20	70	96	20.9	42.3	434.7	1785.1	880.9
21	27	49	-22.2	-4.8	490.6	22.6	105.2
22	5	89	-44.2	35.3	1949.2	1242.6	-1556.3
23	92	19	42.9	-34.8	1836.1	1207.6	-1489.0
24	52	34	2.9	-19.8	8.1	390.1	-56.3
30							
31	Correlation Coefficient is calculated using the formula given below						
32	Correlation Coefficient = $\Sigma [(X - X_m) * (Y - Y_m)] / \sqrt{[\Sigma (X - X_m)^2 * \Sigma (Y - Y_m)^2]}$						
33							
34	Correlation Coefficient Formula	=SUM(G5:G24)/SQRT((SUM(E5:E24)*SUM(F5:F24)))					
35	Correlation Coefficient	0.325784					
36							

Correlation Coefficient = 0.325784

R^2 is calculated using the formula given below

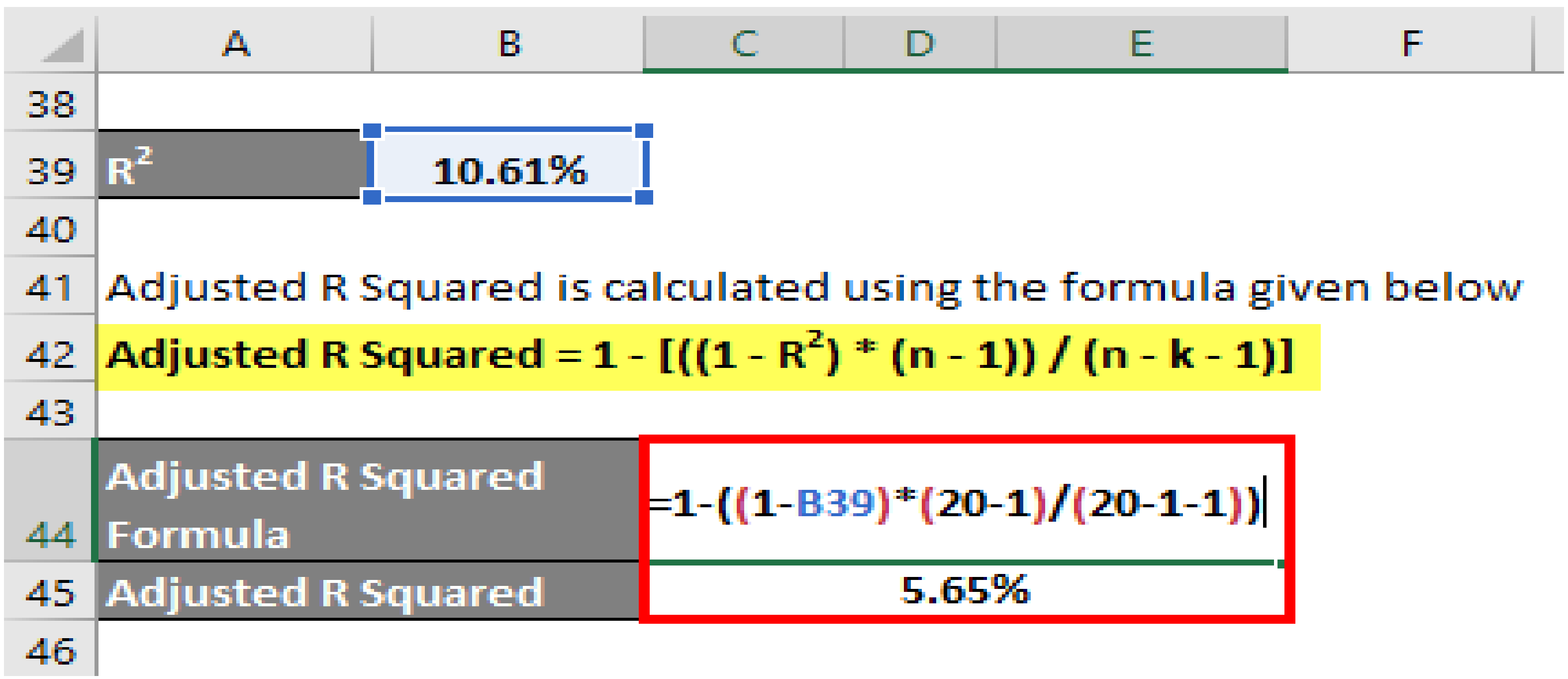
R^2 = (Correlation Coefficient)^2



R^2 = 10.61%

Adjusted R Squared is calculated using the formula given below

Adjusted R Squared = $1 – [((1 – R^2) * (n – 1)) / (n – k – 1)]$



- Adjusted R Squared = 1 – ((1 – 10.61%) * (20 – 1)/(20 – 1 – 1))

- Adjusted R Squared = 5.65%