

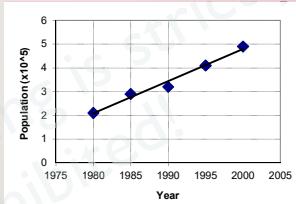
Linear Regression with Example

Year	Population
1980	2.1
1985	2.9
1990	3.2
1995	4.1
2000	4.9
2005	?

year	1980	1985	1990	1995	2000	2005
population	2.1	2.9	3.2	4.1	4.9	?

Linear Regression with Example

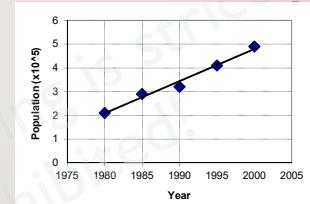
Year	Population
1980	2.1
1985	2.9
1990	3.2
1995	4.1
2000	4.9
2005	?



$$\hat{y} = \text{slope} * x + \text{intercept}$$

Linear Regression with Example

Year	Population
1980	2.1
1985	2.9
1990	3.2
1995	4.1
2000	4.9
2005	?



$$\hat{y} = \text{slope} * x + \text{intercept}$$

Manual Computation using Linear Regression Formula

$$\hat{y} = \text{slope} * x + \text{intercept}$$

$$\text{slope} = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

$$\text{intercept} = \bar{y} - \text{slope} \cdot \bar{x}$$

x	y	xy	x^2
Year	Population		
1980	2.1		
1985	2.9		
1990	3.2		
1995	4.1		
2000	4.9		
Sum			
Average			
Count (n) =			

Slope	
Intercept	

Manual Computation using Linear Regression Formula

$$\hat{y} = \text{slope} * x + \text{intercept}$$

$$\text{slope} = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

$$\text{intercept} = \bar{y} - \text{slope} \cdot \bar{x}$$

x	y	xy	x^2
Year	Population		
1980	2.1	4158	3920400
1985	2.9	5756.5	3940225
1990	3.2	6368	3960100
1995	4.1	8179.5	3980025
2000	4.9	9800	4000000
Sum	9950	17.2	19800750
Average	1990	3.44	
Count (n) =	5		

Slope	0.136
Intercept	-267.2

Linear Regression with Example

Year	Population
1980	2.1
1985	2.9
1990	3.2
1995	4.1
2000	4.9
2005	?

Verify Using Excel Function		
Slope		
Intercept		
Predict	For Year	Prediction

2005

Linear Regression with Example

Year	Population
1980	2.1
1985	2.9
1990	3.2
1995	4.1
2000	4.9
2005	?

Verify Using Excel Function		
Slope	0.136	
Intercept	-267.2	
Predict	For Year	Prediction

2005

Regression Goodness of Fit

Several indices are used to determine the goodness of fit of the model.

- R-squared, or coefficient of determination
- Adjusted R-squared
- Standard Error
- F statistics
- t statistics

R-squared (Measures of Variation)

- Total variation is made up of two parts:

$$SST = SSR + SSE$$

Total Sum of Squares

Regression Sum of Squares

Error Sum of Squares

$$SST = \sum (y_i - \bar{y})^2$$

$$SSR = \sum (\hat{y}_i - \bar{y})^2$$

$$SSE = \sum (y_i - \hat{y}_i)^2$$

where:

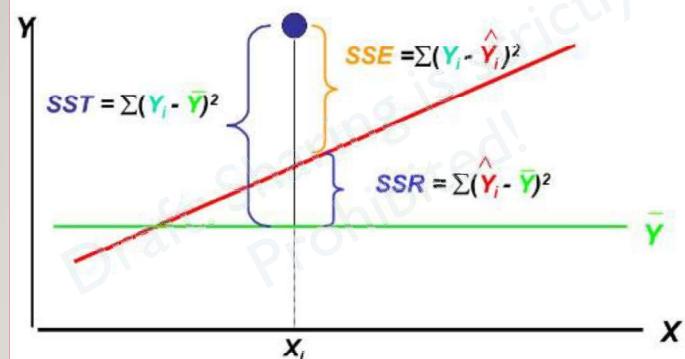
\bar{y} = Average value of the dependent variable

y_i = Observed values of the dependent variable

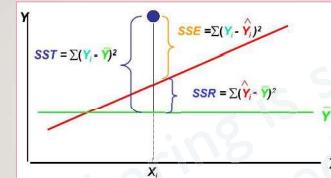
\hat{y}_i = Predicted value of y for the given x_i value

R-squared (Measures of Variation)

$$R^2 = \frac{SSR}{SST} = \frac{(SST - SSE)}{SST} = 1 - \frac{SSE}{SST}$$



R-squared (Measures of Variation)



$$\begin{aligned} R^2 &= \frac{SSR}{SST} \\ &= \frac{(SST - SSE)}{SST} \\ &= 1 - \frac{SSE}{SST} \end{aligned}$$

R-squared (Measures of Variation)

$$R^2 = 1 - \frac{SSE}{SST}$$

Zero Regression Error

$$R^2 = 1 - \frac{0}{SS_{Total}} \rightarrow 1.0$$

R-squared: Calculate SST, SSE & SSR

	x	y				
	Year	Population	Prediction	Error	Square Error	Sq. Mean Difference
1980	2.1	2.08	-0.02	0.0004	1.80	
1985	2.9	2.76	-0.14	0.0196	0.29	
1990	3.2	3.44	0.24	0.0576	0.06	
1995	4.1	4.12	0.02	0.0004	0.44	
2000	4.9	4.8	-0.1	0.01	2.13	
Sum				0.088	4.71	
Mean	avg(x)	avg(y)			MSE	MST
Count (n)					df	df

	x	y				
	Year	Population	Prediction	Error	Square Error	Sq. Mean Difference
1980	2.1	2.08	-0.02	0.0004	1.80	
1985	2.9	2.76	-0.14	0.0196	0.29	
1990	3.2	3.44	0.24	0.0576	0.06	
1995	4.1	4.12	0.02	0.0004	0.44	
2000	4.9	4.8	-0.1	0.01	2.13	
Sum				0.088	4.71	
Mean	avg(x)	avg(y)			MSE	MST
Count (n)					df	df

	x	y				
	Year	Population	Prediction	Error	Square Error	Sq. Mean Difference
1980	2.1	2.08	-0.02	0.0004	1.80	
1985	2.9	2.76	-0.14	0.0196	0.29	
1990	3.2	3.44	0.24	0.0576	0.06	
1995	4.1	4.12	0.02	0.0004	0.44	
2000	4.9	4.8	-0.1	0.01	2.13	
Sum				0.088	4.71	
Mean	avg(x)	avg(y)			MSE	MST
Count (n)					df	df

R-squared: Calculate SST, SSE & SSR

	Degree of freedom	Sum of square	Mean square	F
Regression	$q-1$	$SST - SSE = \sum_i (\hat{y}_i - \bar{y})^2$	$MSR = \frac{SST - SSE}{q-1}$	$F = \frac{MSR}{MSE}$
Residual (Error)	$n-q$	$SSE = \sum_i (y_i - \hat{y}_i)^2$	$MSE = \frac{SSE}{n-q}$	
Total	$n-1$	$SST = \sum_i (y_i - \bar{y})^2$		

Linear Regression Equation: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_n X_n$

Simple Linear Regression: $Y = \beta_0 + \beta_1 X_1$

= Intercept + X1 * Slope

No. of Coefficients q = 2 (β_0 = Intercept, β_1 = Slope)

R-squared: Calculate SST, SSE & SSR

	x	y				
	Year	Population	Prediction	Error	Square Error	Sq. Mean Difference
1980	2.1	2.08	-0.02	0.0004	1.80	
1985	2.9	2.76	-0.14	0.0196	0.29	
1990	3.2	3.44	0.24	0.0576	0.06	
1995	4.1	4.12	0.02	0.0004	0.44	
2000	4.9	4.8	-0.1	0.01	2.13	
Sum						
Mean	avg(x)	avg(y)			MSE	MST
Count (n)					df	df

Slope	0.136
Intercept	-267.2

Number of Coefficients	R Square
	MSE
	MST
	Adjusted R Square
	Standard Error

R-squared: Calculate SST, SSE & SSR

$$R^2 = 1 - \frac{SSE}{SST}$$

$$R^2_{adj} = 1 - \frac{MSE}{MST}$$

$$MSE = SSE / (n-q)$$

$$Std Error = \sqrt{MSE} = \sqrt{\frac{SSE}{n-q}}$$

$$MST = SST / (n-1)$$

$$F = \frac{MSR}{MSE}$$

$$MSR = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{q-1} = \frac{SST - SSE}{q-1}$$

R-squared: Calculate SST, SSE & SSR

$$R^2 = 1 - \frac{SSE}{SST}$$

$$R^2_{adj} = 1 - \frac{MSE}{MST}$$

$$MSE = SSE / (n-q)$$

$$Std Error = \sqrt{MSE} = \sqrt{\frac{SSE}{n-q}}$$

$$MST = SST / (n-1)$$

$$F = \frac{MSR}{MSE}$$

$$MSR = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{q-1} = \frac{SST - SSE}{q-1}$$

	x	y				
	Year	Population	Prediction	Error	Square Error	Sq. Mean Difference
1980	2.1	2.08	-0.02	0.0004	1.80	
1985	2.9	2.76	-0.14	0.0196	0.29	
1990	3.2	3.44	0.24	0.0576	0.06	
1995	4.1	4.12	0.02	0.0004	0.44	
2000	4.9	4.8	-0.1	0.01	2.13	
Sum						
Mean	avg(x)	avg(y)			MSE	MST
Count (n)					df	df

Number of Coefficients	R Square
	MSE
	MST
	Adjusted R Square
	Standard Error

R-squared: Calculate SST, SSE & SSR

$$R^2 = 1 - \frac{SSE}{SST}$$

$$R^2_{adj} = 1 - \frac{MSE}{MST}$$

$$MSE = SSE / (n - q)$$

$$Std. Error = \sqrt{\frac{SSE}{n - q}}$$

$$MST = SST / (n - 1)$$

$$F = \frac{MSR}{MSE}$$

$$MSR = \frac{\sum (y_i - \bar{y})}{q-1} = \frac{SST - SSE}{q-1}$$

	x	y				
	Year	Population	Prediction	Error	Square Error	Sq. Mean Difference
	1980	2.1	2.08	-0.02	0.0004	1.80
	1985	2.9	2.76	-0.14	0.0196	0.29
	1990	3.2	3.44	0.24	0.0576	0.06
	1995	4.1	4.12	0.02	0.0004	0.44
	2000	4.9	4.8	-0.1	0.01	2.13
	Sum				0.088	4.71
	Mean	avg(x)	avg(y)		0.029	1.178
	Count (n)	5			df = 3	df = 4

Number of Coefficients	2
R Square	0.981
MSE	0.029
MST	1.178
Adjusted R Square	0.975
Standard Error	0.1712698

Calculate Statistics (Using Tool)

Regression Statistics						
Multiple R	0.991					
R Square	0.981					
Adjusted R Square	0.975					
Standard Error	0.171					
Observations	5					

ANOVA						
	df	SS	MS	F	Significance F	
Regression	1	4.624	4.624	157.636	0.001	
Residual	3	0.088	0.029			
Total	4	4.712				

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-267.2	21.556	-12.396	0.001	-335.801	-198.599	-335.801	-198.599
X Variable 1	0.136	0.011	12.555	0.001	0.102	0.170	0.102	0.170

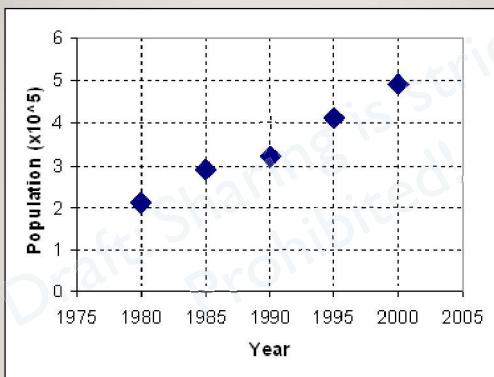
Calculate Statistics (Using Tool)

A	B	C	D	E	F	G	H	I
1 SUMMARY OUTPUT								
2								
3 Regression Statistics								
4 Multiple R	0.991							
5 R Square	0.981							
6 Adjusted R Square	0.975							
7 Standard Error	0.171							
8 Observations	5							
9								
10 ANOVA								
11	df	SS	MS	F	Significance F			
12 Regression	1	4.624	4.624	157.636	0.001			
13 Residual	3	0.088	0.029					
14 Total	4	4.712						
15								
16	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
17 Intercept	-267.2	21.556	-12.396	0.001	-335.801	-198.599	-335.801	-198.599
18 X Variable 1	0.136	0.011	12.555	0.001	0.102	0.170	0.102	0.170
19								
Here are slope and intercept of regression line								
3. Absolute value of t statistics must be larger than 1.645								

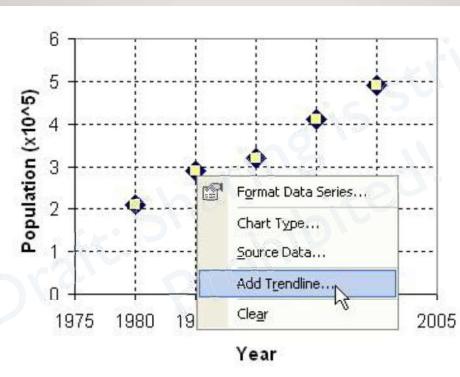
Regression model needs to pass all the criteria below:

1. The R square must be bigger than 0.80
2. The significant F (from ANOVA) must be smaller than 0.05
3. The absolute value of t-statistics must be larger than 1.96 for $\alpha=0.05$ and must larger than 1.645 for $\alpha=0.10$

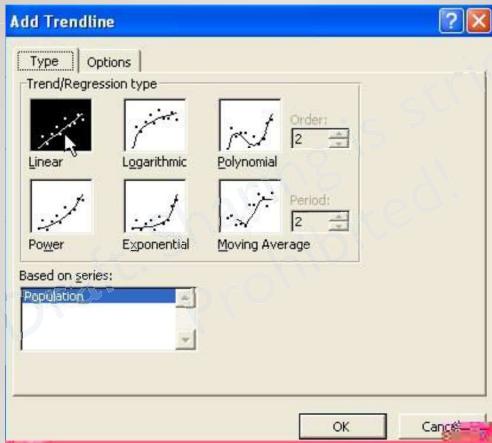
Regression analysis using chart (Scatter Plot)



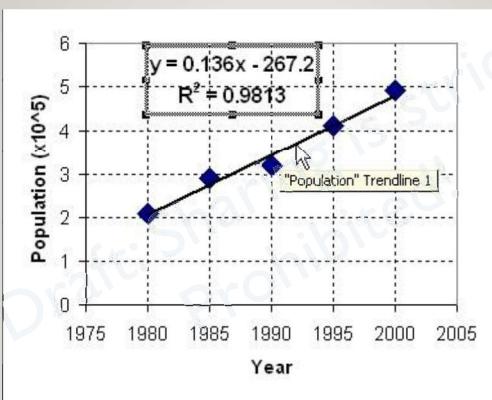
Regression analysis using chart



Regression analysis using chart



Regression analysis using chart



Manual Computation we did

$$\hat{y} = \text{slope} * x + \text{intercept}$$

$$\text{slope} = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

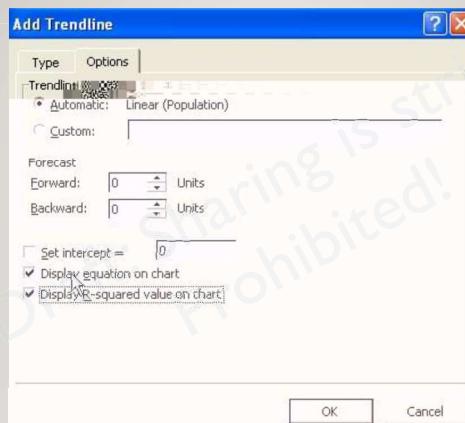
$$\text{intercept} = \bar{y} - \text{slope} \cdot \bar{x}$$

	x	y	xy	x^2
Year	Population			
1980	2.1	4158	3920400	
1985	2.9	5756.5	3940225	
1990	3.2	6368	3960100	
1995	4.1	8179.5	3980025	
2000	4.9	9800	4000000	
Sum	9950	17.2	34262	19800750
Average	1990	3.44		
Count (n) =	5			

Slope 0.136

Intercept -267.2

Regression analysis using chart



Manual Computation we did

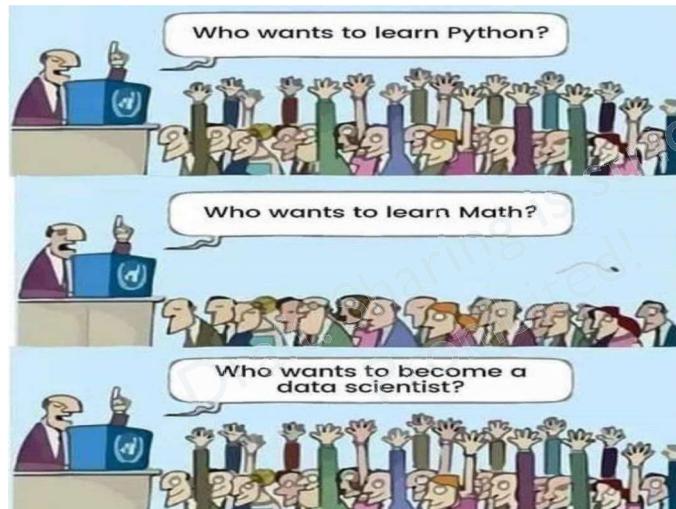
$$\hat{y} = \text{slope} * x + \text{intercept}$$

$$\text{slope} = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

$$\text{intercept} = \bar{y} - \text{slope} \cdot \bar{x}$$

	x	y	xy	x^2
Year	Population			
1980	2.1			
1985	2.9			
1990	3.2			
1995	4.1			
2000	4.9			
Sum				
Average				
Count (n) =				

Slope	
Intercept	



Linear Regression with Linear Algebra

$$y = m \cdot x + c$$

$$\begin{aligned}m \cdot x_1 + c &= y_1 \\m \cdot x_2 + c &= y_2 \\\vdots \\m \cdot x_n + c &= y_n\end{aligned}$$

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \cdot \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\mathbf{A} = [x \ 1] \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} m \\ c \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

Draft Share Prohibited!

Linear Regression with Linear Algebra

$$y = m \cdot x + c$$

$$\begin{aligned}m \cdot x_1 + c &= y_1 \\m \cdot x_2 + c &= y_2 \\\vdots \\m \cdot x_n + c &= y_n\end{aligned}$$

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \cdot \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\mathbf{A} \cdot \mathbf{b} = \mathbf{y}$$

$$(\mathbf{A}^T \cdot \mathbf{A}) \cdot \mathbf{b} = \mathbf{A}^T \cdot \mathbf{y}$$

$$(\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot (\mathbf{A}^T \cdot \mathbf{A}) \cdot \mathbf{b} = (\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^T \cdot \mathbf{y}$$

$$\mathbf{b} = (\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^T \cdot \mathbf{y}$$

$$\mathbf{A} = [x \ 1] \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} m \\ c \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$\mathbf{A}^T \cdot \mathbf{A} =$$

$$\mathbf{A}^T \cdot \mathbf{y} =$$

Linear Regression with Linear Algebra

$$y = m \cdot x + c$$

$$\begin{aligned}m \cdot x_1 + c &= y_1 \\m \cdot x_2 + c &= y_2 \\\vdots \\m \cdot x_n + c &= y_n\end{aligned}$$

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \cdot \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\mathbf{A} \cdot \mathbf{b} = \mathbf{y}$$

$$(\mathbf{A}^T \cdot \mathbf{A}) \cdot \mathbf{b} = \mathbf{A}^T \cdot \mathbf{y}$$

$$(\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot (\mathbf{A}^T \cdot \mathbf{A}) \cdot \mathbf{b} = (\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^T \cdot \mathbf{y}$$

$$\mathbf{b} = (\mathbf{A}^T \cdot \mathbf{A})^{-1} \cdot \mathbf{A}^T \cdot \mathbf{y}$$

$$\mathbf{A} = [x \ 1] \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} m \\ c \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$\mathbf{A}^T \cdot \mathbf{A} = \begin{bmatrix} \sum x^2 & \sum x \\ \sum x & n \end{bmatrix}$$

$$\mathbf{A}^T \cdot \mathbf{y} = \begin{bmatrix} \sum xy \\ \sum y \end{bmatrix}$$

Linear Regression with Linear Algebra

	1980	1985	1990	1995	2000
$\mathbf{A} \cdot \mathbf{b}$	1	1	1	1	1
$\mathbf{A}' \cdot \mathbf{A}$	19800750	9950			
	9950	5			
Inverse of $\mathbf{A}' \cdot \mathbf{A}$					
	0.004	-7.96			
	-7.96	15840.6			
$\mathbf{L} = \text{Inv}(\mathbf{A}' \cdot \mathbf{A}) \cdot \mathbf{A}'$	-0.04	-0.02	8.88E-16	0.02	0.04
	79.8	40	0.2	-39.6	-79.4
$\mathbf{b} = \text{Inv}(\mathbf{A}' \cdot \mathbf{A}) \cdot \mathbf{A}' \cdot \mathbf{y} = \mathbf{L} \cdot \mathbf{y}$	0.136				
	-267.2				

THANK YOU!