LINEAR REGRESSION & TIME SERIES

Homework - 1

Blessina Paul, MS in Data Analytics Student

Tushar Raj, MS in Data Analytics Student

Clark University

1. For the following data

X	30	28	41	18	30	45
Y	130.4	125.0	163.3	94.4	130.4	173.9

(a) Estimate the regression equation of Y on X using least squares method.
(b) What value of Y would you predict for X=40? Why this value?
(c) What value of Y would you predict for X=60? Would you have equal faith in these two predictions?
(d) Calculate the residual (The error term) for the fourth observation, e4, where X=18.
(e) Does the estimated regression line pass through mean values of x and y ?
(f) What is your opinion on the correlation between X and Y?
(g) Compute the expected (mean) value of the estimated residuals. Is the mean equal to zero?

Solution: Done in excel

a. Regression line equation:

Y	X	y	x	y*x	x*x	y^	u^	
130.4	30	-5.833333333	-2	11.66666667	4	130.33035	0.06964539	
125	28	-11.23333333	-4	44.93333333	16	124.42738	0.572624113	
163.3	41	27.06666667	9	243.6	81	162.79674	0.503262411	
94.4	18	-41.83333333	-14	585.6666667	196	94.912482	-0.51248227	
130.4	30	-5.833333333	-2	11.66666667	4	130.33035	0.06964539	
173.9	45	37.66666667	13	489.6666667	169	174.6027	-0.702695035	
136.2333333	32	0.0000000000	0	1387.2	470	136.23333	0.0000000	$= \sum \mu$

β2	2.951489362
β1	41.78567376

y at x =40 is 159.8452482

estimated regression line pass through mean values of x and y = 136.23333

$$\hat{\beta}_2 = \frac{\sum y_i x_i}{\sum x_i^2} = \frac{1387.203}{470}$$

$$= 2.95149$$

$$\hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{X}$$

$$= 136.233 - (2.95149)(32)$$

$$= 136.233 - 94.44768$$

$$= 41.7853$$

$$E(y|x_i) = \beta_1 + \beta_2 X$$

$$136.232 = 41.7853 + 2.95149$$

Estimated regression line equation

$$\hat{y}_i = 41.7853 + 2.95149x_i$$

(B) What value of Y would you predict for X = 40? Why this value?

For $x = 40$

$$E(y|x_i) = \beta_1 + \beta_2 x$$
$$= 41.7853 + 2.9549(40)$$
$$= 41.7853 + 118.0596$$
$$= 159.8449$$

The predicted value of y for $x = 40$ is : $\underline{159.8449}$

This value was calculated by the equation derived using the least square method.

When, $x_{30} \rightarrow y = 130.4$

$x_{28} \rightarrow y = 125.0$

$x_2 \rightarrow x_{30} - x_{28} = 130.4 - 125.0$
$= 5.4$

$x_1 \rightarrow x_2/2 = 5.4/2 = 2.7$

$x_{40} \rightarrow x_{41} - x_1 = 163.3 - 2.7$
$= 160.6$

Error Term:

$$\hat{u}_i = y_i - \hat{y}_i$$
$$= 159.8449 - 160.6$$
$$= -0.7551$$

(C) What value of Y would you predict for X = 60? Would you have equal faith in these two predictions?

For $x = 60$

$$E(y|x_i) = \beta_1 + \beta_2 X$$
$$= 41.7853 + 2.95149\,(60)$$
$$= 41.7853 + 177.0894$$
$$= 41.7853 + 177.0894$$
$$= 218.8747$$

To say if we have same faith in $x = 40$ & $x = 60$ prediction. We have to get the difference between the observed and predicted of y is the error, or residual.

Represents how close our prediction comes to the actual observation. The smaller the residual, the better the prediction.

Thus Calculating Error Term for $x = 60$

$$\hat{U}_i = y_1 - \hat{y}_1 \quad -①$$

Calculating y value when $x = 60$

$x_{30} \rightarrow y = 130.4$

$x_{45} \rightarrow y = 173.9$

$x_{15} \rightarrow x_{45} - x_{30} = 173.9 - 130.4 = 43.5$

Thus, $x_{60} \rightarrow x_{45} + x_{15} = 173.9 + 43.5 = 217.4$

Putting value of $x_{60} \rightarrow y_1 = 217$ in equation ①

$$\hat{U}_i = 217.4 - 218.8747$$
$$= -1.4747$$

The Error value of $x = 60$ is -1.4747 which is double for $x = 40$. Thus we dont have equal faith on both prediction

(D) Calculate the residual (TH Error term) for the fourth Observation, e_4, where $x = 18$

When $x = 18$, Error Term

$$\hat{U}_i = y_i - \hat{y}_i \quad \text{—— (1)}$$

Calculating $\hat{y}_i$ for $x = 18$

$$\hat{y}_i = 41.7853 + 2.95149x \quad \text{—— (2)}$$

Substituting $x = 18$ in above equation (2)

$$\hat{y}_i = 41.7853 + 2.95149 \times 18$$
$$= 41.7853 + 53.66682$$
$$= 94.912$$

Now, Substituting the value of y_i and $\hat{y}_i$ in Equation

when $x = 18$

$$\hat{U}_i = y_i - \hat{y}_i$$
$$\hat{U}_4 = 94.4 - 94.912$$
$$\hat{U}_4 = -0.512$$

The residual (The error term) for the fourth observation, e_4, where $x = 18$ is : -0.512

(E) Does the estimated regression line pass through mean values of x and y?

Estimated regression line

$$\hat{y}_i = 41.7853 + 2.95149 x_i$$

Mean value of x = 32

Mean value of y = 136.233

So by substituting in the above equation the value of x regression line is passing through the mean values.

$$\text{Since, } \hat{\beta}_1 = \hat{y} - \hat{\beta}_2 \bar{X}$$

$$\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}$$

Using the above Equation regression line is passing through the mean value.

(F) What is your opinion on the correlation between X and Y?

Correlation measures the strength of the linear association between the variables.

$$r^2 = \hat{\beta}_2^2 \left(\frac{\sum x_i^2}{\sum y_i^2} \right)$$

$$= (2.95149)^2 \times \left(\frac{470}{4095.6498} \right)$$

$$= 8.7112932 \times 0.1147559$$

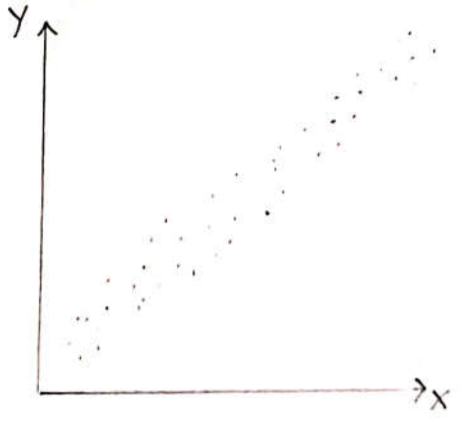
$$r^2 = 0.9996696$$

$$r = \sqrt{0.9996696}$$

$$= 0.9998347$$

Here the value of r is close to (+1) positive one and $r > 0$

Thus, Both have positive relationship and both move in Same direction.



(G) Compute the expected (mean) value of the residuals. Is the mean equal to zero?

X	Y	E(Y/x) or $\hat{Y}$	$\hat{U}_i = y_i - \hat{y}$
30	130.4	130.33	0.07
28	125	124.427	0.573
41	163.3	162.7953	0.50
18	94.4	94.912	-0.512
30	130.4	130.33	0.07
45	173.9	174.595	-0.697

≈ 0

$$\text{Mean of residual} = \frac{\sum \hat{U}_i}{n}$$

$$= \frac{0}{6} = 0$$

$$\boxed{\text{Mean of Residual} = 0}$$

2. Here are some data from the OECD from the British consumer price inflation and for the unemployment rate.

Year	Unemployment rate % (*ur*)	Inflation rate % (p)
1991	8.6	
1992	9.7	4.2
1993	10.2	2.5
1994	9.3	2.0
1995	8.5	2.7
1996	7.9	2.5
1997	6.8	1.8
1998	6.1	1.6
1999	5.9	1.3
2000	5.4	0.8
2001	5.0	1.2
2002	5.1	1.3
2003	4.9	1.4
2004	4.7	1.3
2005	4.7	2.0
2006		2.2

Consider a Phillips curve:

$\pi_t = \alpha + \beta\, ur_{t-1} + e_t$, t=1992 to 2006

(a) Estimate the Phillips curve by OLS, retaining the final observation for prediction (i.e. don't include it in calculating the regression line).

(b) What are the estimates of α and β?

(c) Predict the 2006 level of inflation by inserting the 2005 unemployment rate into your equation. How good is the prediction?

Solution:

	UR	P						
year	Unemployment rate % (ur)	Inflation rate % (p)	ur	p	ur*p	ur^2	p^	et
1991	8.6		1.592857143					
1992	9.7	4.2	2.692857143	2.3	3.6635714	2.5371939	2.3547471	1.845252918
1993	10.2	2.5	3.192857143	0.6	1.6157143	7.2514796	2.6687877	-0.168787668
1994	9.3	2	2.292857143	0.1	0.3192857	10.194337	2.8115334	-0.811533389
1995	8.5	2.7	1.492857143	0.8	1.8342857	5.2571939	2.5545911	0.145408908
1996	7.9	2.5	0.892857143	0.6	0.8957143	2.2286224	2.3261979	0.173802062
1997	6.8	1.8	-0.207142857	-0.1	-0.089286	0.7971939	2.1549031	-0.354903073
1998	6.1	1.6	-0.907142857	-0.3	0.0621429	0.0429082	1.8408625	-0.240862487
1999	5.9	1.3	-1.107142857	-0.6	0.5442857	0.8229082	1.6410185	-0.341018478
2000	5.4	0.8	-1.607142857	-1.1	1.2178571	1.2257653	1.5839202	-0.783920189
2001	5	1.2	-2.007142857	-0.7	1.125	2.5829082	1.4411745	-0.241174469
2002	5.1	1.3	-1.907142857	-0.6	1.2042857	4.0286224	1.3269779	-0.026977892
2003	4.9	1.4	-2.107142857	-0.5	0.9535714	3.6371939	1.355527	0.044472964
2004	4.7	1.3	-2.307142857	-0.6	1.2642857	4.440051	1.2984287	0.001571252
2005		2		0.1	-0.230714	5.3229082	1.2413305	0.758669541
2006								
	7.007142857	1.9	0.0000	0.00000	14.38	50.369286	1.9	0.00000

β= 0.285491442

α= -0.100479317

(a) Estimate the philips curve by OLS, retaining the final observation for prediction (i.e. don't Include don't Include it in Calculating the regression line).

$\pi_t = \alpha + \beta ur_{t-1} + e_t$

$\pi_t = -0.100479317 + 0.28549144 2 ur_{t-1} + e_t$

The value of α and β is subsituted in the equation which is calculated in the Excel spread sheet attached.

(B) What is the estimated of α and β?

$\alpha = -0.100149317$

$\beta = 0.28549144 2$

Please refer the attached Excel spred sheet for the calculation step of α and β.

(C) Predict the 2006 level of Inflation by Inserting the 2005 unemployment rate into your equation. How good is the prediction?

$\pi_t = \alpha + \beta ur_{t-1} + e_t \quad -(1)$

From Excel spred sheet:

$\alpha = -0.100149317$

$\beta = 0.28549144 2$

$e_t = 0.758669541$

$ur_{t-1} = 4.7$

Substituting in Equation (1)

$\pi_t = -0.100149317 + 0.28549144 2 (4.7)$

$= 1.2416$, which is close to the given value of 2006, so prediction is good.