A Simple Linear Regression and Correlation

Question Cost accountants often estiamte over head Cost based on the level of production. At the Standard Knitting Co., they have collected infromation on over head Cost in US$ (Y) and units (X) produced at different plants and want to estiamte the Simple Linear Regression Equation (Y=a+bX) to pridict future over head Cost:

A	Find the regression equation for Cost accountants.
B	Pridict the over head Cost when 50 units are produced?
C	Calculate the Standard Error of Estimate
D	Find the Coefficent of Determination
E	Find the Coefficent of Coreelation

Dr. Shabbir Ahmad

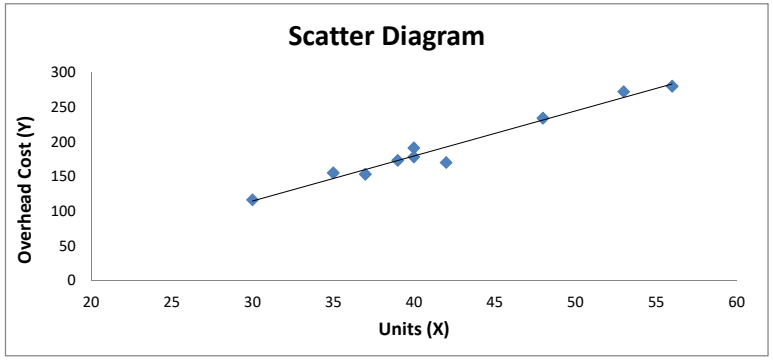
Assistant Professor
Department of Mathematics
COMSATS University Islamabad, Wah Campus

Plants	Units X	Overhead Cost (US$) Y	XY	X^2	Y^2	$\hat{Y}$	$Y-\bar{Y}$	$(Y-\hat{Y})^2$	$\hat{Y}-\bar{Y}$	$(\hat{Y}-\bar{Y})^2$	$Y-\bar{Y}$	$(Y-\bar{Y})^2$
1	40	191	7640	1600	36481	179.217	11.782993	138.8389287	-12.983	168.558	-1.2	1.44
2	42	170	7140	1764	28900	192.2	-22.2	492.84	0.000	0.000	-22.2	492.84
3	53	272	14416	2809	73984	263.6065	8.3935374	70.45147034	71.406	5098.883	79.8	6368.04
4	35	155	5425	1225	24025	146.7595	8.2404762	67.90544785	-45.440	2064.837	-37.2	1383.84
5	56	280	15680	3136	78400	283.081	-3.080952	9.492267574	90.881	8259.348	87.8	7708.84
6	39	173	6747	1521	29929	172.7255	0.2744898	0.075344648	-19.474	379.256	-19.2	368.64
7	48	234	11232	2304	54756	231.149	2.8510204	8.128317368	38.949	1517.023	41.8	1747.24
8	30	116	3480	900	13456	114.302	1.6979592	2.883065389	-77.898	6068.092	-76.2	5806.44
9	37	153	5661	1369	23409	159.7425	-6.742517	45.46153559	-32.457	1053.488	-39.2	1536.64
10	40	178	7120	1600	31684	179.217	-1.217007	1.481105558	-12.983	168.558	-14.2	201.64
	420	1922	84541	18228	395024	1922	0.00	837.557483		24778.043	0.000	25615.6
	$\sum X$	$\sum Y$	$\sum XY$	$\sum X^2$	$\sum Y^2$		$\sum(Y-\hat{Y})$	$\sum(Y-\hat{Y})^2$	$\sum(\hat{Y}-\bar{Y})$	$\sum(\hat{Y}-\bar{Y})^2$	$\sum(Y-\bar{Y})$	$\sum(Y-\bar{Y})^2$
	42	192.2										
	$\bar{X}$	$\bar{Y}$										

0.967302836

a = -80.44285714

b = 6.491496599



A Equation of Straight Line

$$Y = a + bX$$

$$b = \frac{n\sum XY - \sum X \sum Y}{n\sum X^2 - (\sum X)^2} = \frac{10(84541) - (420)(1922)}{10(18228) - (420)^2} = 6.4915$$

$$a = \bar{Y} - b\bar{X} = 192.2 - 6.4915 * 42 = -80.38$$

So finally we obtain the regression equation as:

$$Y = -80.38 + 6.4915X$$

B Overhead at 50 Units

AS Y= - 80.38 + 6.4915 X

X=50 Y= - 80.38 + 6.4915 (50)

$\hat{Y}$ = 244.2

C Statndard Error of Estimate

The Variance of Regression is

$$S_{y.x}^2 = \frac{\sum(Y-\hat{Y})^2}{n-2} = \frac{837.5972}{10-2} = 104.6996$$

OR

$$S_{y.x}^2 = \frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2} = \frac{395024 - (-80.4429)*1922 - 6.491497*84541}{10-2} = 104.6996$$

The Standard deviation of estimate is

$$S_{y.x} = \sqrt{\frac{\sum(Y-\hat{Y})^2}{n-2}} \text{ OR } S_{y.x} = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{n-2}} = \sqrt{104.6996} = 10.2323$$

D Coefficent of Determination

The proportion of total variation in Y(Dep Variable), that is explained due to X (Indep Var).

Total SS = RSS + ESS

$$\sum(Y-\bar{Y})^2 = \sum(\hat{Y}-\bar{Y})^2 + \sum(Y-\hat{Y})^2$$

$$R^2 = \frac{RSS}{TSS} = \frac{\sum(\hat{Y}-\bar{Y})^2}{\sum(Y-\bar{Y})^2} \quad \text{OR} \quad R^2 = \frac{a\sum Y + b\sum YX - \frac{(\sum Y)^2}{n}}{\sum Y^2 - \frac{(\sum Y)^2}{n}} = 0.97203$$

$$R^2 = 1 - \frac{ESS}{TSS} = 1 - \frac{\sum(Y-\hat{Y})^2}{\sum(Y-\bar{Y})^2} \quad \text{OR} \quad R^2 = 1 - \frac{\sum Y^2 - a\sum Y - b\sum YX}{\sum Y^2 - \frac{(\sum Y)^2}{n}} = 0.97203$$

Dr. Shabbir Ahmad
Assistant Professor
Department of Mathematics
COMSATS University Islamabad, Wah Campus

E Coefficent of Correlation

$$r = r_{xy} = r_{yx} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2] \times [n\sum Y^2 - (\sum Y)^2]}}$$

$$r_{yx} = \frac{10 * 84541 - 420 * 1922}{\sqrt{[10 * 18228 - (420)^2][10 * 395024 - (1922)^2]}} = 0.983516$$

Y	X	U = (X-10)/5	V = (Y-20)/10
191	40	6	17.1
170	42	6.4	15
272	53	8.6	25.2
155	35	5	13.5
280	56	9.2	26
173	39	5.8	15.3

Properties of Correlation Coeff

The degree of inter dependence between two variables

1. The Geomatric Mean of regression Coefficients is Correltaion Coefficient.
2. $-1 \le r_{yx} \le 1$ — $r_{yx} = -1$ Perfect Negative Corr; $r_{yx} = 0$ No Corr; $r_{yx} = +1$ Perfect Psitive Corr — Ideal Situations
3. $r_{yx} = r_{xy}$
4. $r_{yx} = r_{uv}$ *where* $U = \frac{X - C_1}{h_1}$ *and* $V = \frac{Y - C_2}{h_2}$

$r_{yx} = 0.97396$ and $r_{uv} = 0.97396$

ryx =	0.973959374
ruv =	0.973959374

Dr. Shabbir Ahmad
Assistant Professor
Department of Mathematics
COMSATS University Islamabad, Wah Campus