Lines of Regression

Q.1. Given two regression lines $4x + 3y + 7 = 0$ and $3x + 4y + 8 = 0$, determine :

i. The regression line of y on x.
ii. The regression line of x on y.
iii. The coefficient of correlation

Solution : 1

We have the lines : $4x + 3y + 7 = 0$ and $3x + 4y + 8 = 0$

Let $4x + 3y + 7 = 0$ be of x on y and $3x + 4y + 8 = 0$ be the line of y on x.

Therefore, from $4x + 3y + 7 = 0$

We have $x = -3/4y - 7/4$ ------------------ (i)

And $bxy = -3/4$

And from $3x + 4y + 8 = 0$

We have $y = -3/4x - 8/4$ ----------------- (ii)

And $byx = -3/4$

Therefore, $r = \sqrt{(byxbxy)} = \sqrt{\{(-3/4)\times(-3/4)\}} = -3/4 = -0.75$.

i. $3x + 4y + 8 = 0$ of y on x.
ii. $4x + 3y + 7 = 0$ of x on y.
iii. correlation coefficient $= -0.75$.

Q.2. There are two series of index numbers : P for price index and S for stock of a commodity. The means and standard deviation of P are 100 and 8 and of S are 103 and 4 respectively. The correlation coefficient between the two series is 0.4. With these data, obtain the regression lines of P on S and S on P.

Solution : 2

	Commodity P(x)	Commodity S(y)
Mean	$M_x = 100$	$M_y = 103$
S.D.	$\sigma_x = 8$	$\sigma_y = 4$

Correlation coefficient (r) = 0.4.

Therefore, $b_{yx} = (r\sigma y)/\sigma x = (0.4\times4)/8 = 0.2$, And $b_{yx} = (r\sigma x)/\sigma y = (0.4\times8)/4 = 0.8$.

Therefore, regression line of P on S is,

$(x - Mx) = b_{yx}(y - M_y)$

Or, $(x - 100) = 0.8(y - 103)$

Or, $x - 100 = 0.8y - 82.4$

Or, $x - 0.8y = 17.6$

And regression line of S on P is

$(y - My) = b_{yx}(x - M_x)$

Or, $(y - 103) = 0.2(x - 100)$

Or, $y - 103 = 0.2x - 20$

Or, $0.2x - y + 83 = 0$

Q.3. If the two regression lines of a bivariate distribution are $4x - 5y + 33 = 0$ and $20x - 9y - 107 = 0$,

i. calculate the arithmetic means of x and y respectively.
ii. estimate the value of x when y = 7.
iii. find the variance of y when $\sigma_x = 3$.

Solution : 3

We have, $4x - 5y + 33 = 0 \Rightarrow y = 4x/5 + 33/5$ --------------- (i)

And $20x - 9y - 107 = 0 \Rightarrow x = 9y/20 + 107/20$ ------------- (ii)

i. Solving (i) and (ii) we get, mean of x = 13 and mean of y = 17.

ii. Second line is line of x on y $x = (9/20) \times 7 + (107/20) = 170/20 = 8.5$

iii. $byx = r(\sigma y/\sigma x) \Rightarrow 4/5 = 0.6 \times \sigma_y/3$ [$r = \sqrt{(b_{yx}.b_{xy})} = \sqrt{\{(4/5)(9/20)\}} = 0.6$
$\Rightarrow \sigma y = (4/5)(3/0.6) = 4$

Q.4. The date for marks in Physics and History obtained by ten students are given below:

Marks in Physics	15	12	8	8	7	7	7	6	5	3
Marks in History	10	25	17	11	13	17	20	13	9	15

Using this data:

(a) Compute Karl Pearson's coefficient of correlation between the marks in Physics and History obtained by the ten students.

(b) (i) Find the line of regression in which Physics is taken as the independent variable.
(ii) A candidate had scored 10 marks in Physics but was absent from the History test. Estimate his probable score for the latter test.

Solution : 4

(a) See Chapter on Correlation coefficient.

(b) Here, byx = $[\Sigma xy - (\Sigma x \Sigma y)/n]/[\Sigma x^2 - (\Sigma x)^2/n]$

= [1192 – (78×150)/10]/[714 – (78×78)/10]

= 22/105.6= 5/24.

And $\bar{x}$ = Σx/n = 7.8 ; $\bar{y}$ = Σy/n = 15.

(i) Line of y on x is given by $y - \bar{y}$ = byx $(x - \bar{x})$

Or, y – 15 = 5/24(x – 7.8) = 5/24(10 – 7.8)

Or, y – 15 = 5/24(2.2) = 11/24

Or, y = 11/24 + 15

= 371/24 = 15.4 (Approx.)

Q.5. Evaluate : $\int[\cos x/\{(1 + \sin x)(2 + \sin x)\}].dx$.

Solution : 5

Let I = $\int[\cos x/\{(1 + \sin x)(2 + \sin x)\}].dx$

[Put sin x = t then cos x dx = dt]

Then I = $\int dt/\{(1 + t)(2 + t)\}$

= $\int[1/\{(1 + t)(2 - 1)\} + 1/\{(1 - 2)(2 + t)\}].dt$

= $\int[1/(1 + t) - 1/(2 + t)].dt$

= $\log|1 + t| - \log|2 + t|$

= $\log|(1 + t)/(2 + t)| + c$

= $\log|(1 + \sin x)/(2 + \sin x)| + c$.

Q.6. Evaluate : $\int dx/(\sin x + \sin 2x)$.

Solution : 6

Let $I = \int dx/(\sin x + \sin 2x)$

$= \int dx/(\sin x + 2 \sin x \cos x)$

$= \int dx/\{\sin x (1 + 2\cos x)\}$

$= \int[\sin x/\{\sin^2 x(1 + 2\cos x)\}].dx$

$= \int[\sin x/\{(1 - \cos^2 x)(1 + 2\cos x)\}].dx$

[Put $\cos x = t$ then $- \sin x\, dx = dt$]

Then $I = - \int dt/\{(1 - t^2)(1 + 2t)\}$

$= - \int dt/\{(1 + t)(1 - t)(1 + 2t)\}$;

Now $1/\{(1 + t)(1 - t)(1 + 2t)\}$

$= 1/[(1 + t)\{1- (-1)\}\{1 + 2(- 1)] + 1/[(1 + 1)(1 - t)\{1 + 2(1)\}] + 1/[(1 - 1/2)\{1 - (- 1/2\} (1 + 2t)]$

$= (-1/2)\{1/(1 + t)\}+ (1/6)\{1/(1 - t)\}+ (4/3)\{1/(1 + 2t)\}$

Therefore, $I = (1/2)\int dt/(1 + t) - (1/6)\int dt/(1 - t) - (4/3)\int dt/(1 + 2t)$

$= (1/2)\log |1 + t|+ (1/6) \log|1 - t|- (4/3)(1/2) \log|1 + 2t|+ c$

$= (1/2) \log|1 + \cos x|+ (1/6) \log|1 - \cos x|- (2/3) \log|1 + 2 \cos x|+ c.$

Q.7. Evaluate : $\int dx/[x(x^5 + 1)]$.

Solution : 7

Let $I = \int dx/[x(x^5 + 1)]$

$= \int[x^4/\{x^5 (x^5 + 1)\}].dx$

[Multiplying Nr & Dr by x4] [Put $x^5 = t$

then $x^4 dx = (1/5)dt$]

Therefore, $I = (1/5)\int dt/[t(t + 1)]$

$= (1/5)\int[1/t - 1/(t + 1)].dt$

$= (1/5) [\log|t| - \log|t + 1|] + c$

$= (1/5) \log|t/(t + 1)| + c$

$= (1/5) \log |x^5/(x^5 + 1)| + c.$