Tutorial Session 10 – Correlation and Regression

Extra Question 1.

X	Y	$\bar{x}-x$	$\bar{y}-y$	$(\bar{x}-x)\cdot(\bar{y}-y)$	$(\bar{x}-x)^2$	$(\bar{y}-y)^2$
1	100	4	-16.3	-65.3	16	266.8
2	98	3	-14.3	-43	9	205.4
3	91	2	-7.3	-14.7	4	53.8
4	86	1	-2.3	-2.3	1	5.4
5	80	0	3.7	0	0	13.4
6	77	-1	6.7	-6.7	1	44.4
7	75	-2	8.7	-17.3	4	75.1
8	76	-3	7.7	-23	9	58.8
9	70	-4	13.7	-54.7	16	186.8
$\Sigma x = 45$	$\Sigma y = 753$			$\Sigma(\bar{x}-x)\cdot(\bar{y}-y) = -227$	$\Sigma(\bar{x}-x) = 60$	$\Sigma(\bar{y}-y) = 910$

$\bar{x} = \frac{\Sigma x}{n} = \frac{45}{9} = 5$; $\bar{y} = \frac{\Sigma y}{n} = \frac{753}{9} = 83.67$

i) $r = \frac{\Sigma((x-\bar{x})(y-\bar{y}))}{\sqrt{\Sigma(\bar{x}-x)^2}\cdot\sqrt{\Sigma(\bar{y}-y)^2}} = \frac{-227}{\sqrt{60}\cdot\sqrt{910}} = \frac{-227}{(7.75)\cdot(30.17)} \Rightarrow -0.97$

There is a strong negative linear relationship between two variables.

ii) $r^2 = (-0.97)^2$
$= 0.94$, 94% of the variability in y can be explained by the model.

iii) $S_y = \sqrt{\frac{\Sigma(\bar{y}-y)^2}{(n-1)}}$; $S_x = \sqrt{\frac{\Sigma(\bar{x}-x)^2}{(n-1)}}$; $\beta_1 = r\frac{S_y}{S_x}$

$= \sqrt{\frac{910}{(9-1)}}$ $\qquad = \sqrt{\frac{60}{(9-1)}}$ $\qquad = (-0.97)\cdot\frac{10.67}{2.74}$

$\Rightarrow 10.67$ $\qquad \Rightarrow 2.74$ $\qquad = (-0.97)\cdot(3.89)$

$\Rightarrow -3.78$

$\beta_0 = \bar{y} - \beta_1\bar{x} = 83.67 - (-3.78)(5) = 83.67 + 18.9 \Rightarrow 102.57$

$\therefore$ The regression model $\Rightarrow y = \beta_0 + \beta_1 x$

$y = 102.57 - 3.78x$

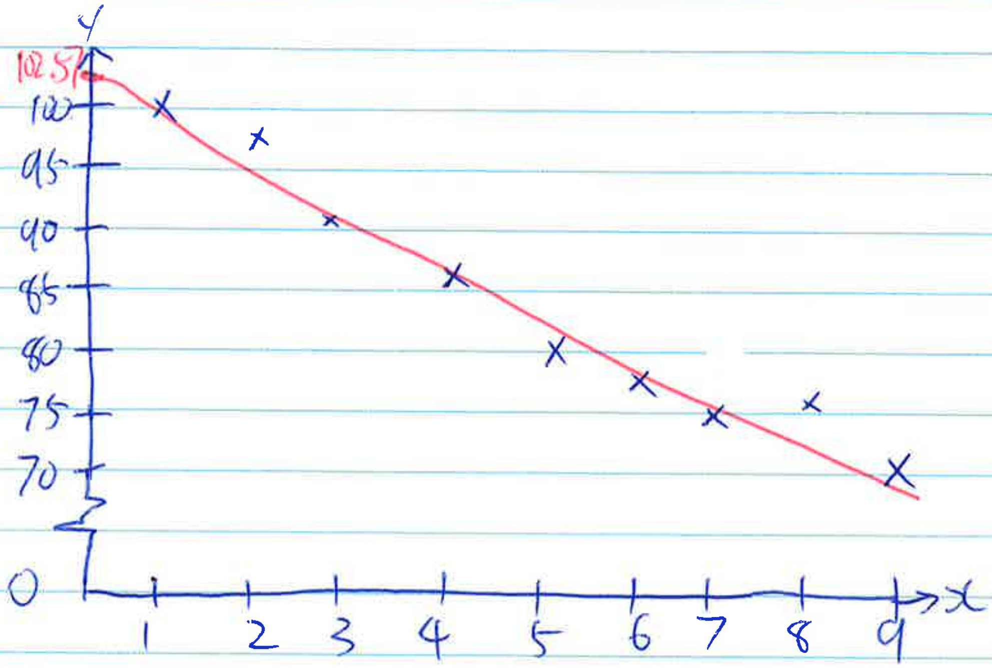
Extra Question 1 Cont'

iv). $x = 13$, $y = 102.57 - 3.78 \times 13$

$y = 102.57 - 49.18$

$y \approx 53.4$

v).



Extra Question 2.

x	y	$\bar{x}-x$	$\bar{y}-y$	$(\bar{x}-x)\cdot(\bar{y}-y)$	$(\bar{x}-x)^2$	$(\bar{y}-y)^2$
1	60	3.4	10.5	35.5	11.3	111.2
1	65	3.4	5.5	18.7	11.3	30.8
3	70	1.4	0.5	0.7	1.9	0.3
6	73	-1.6	-2.5	4	2.7	6
10	75	-5.6	4.5	25.1	31.8	19.8
2	68	2.4	2.5	6	5.6	6.5
8	79	-3.6	-8.5	30.7	13.2	71.5
2	67	2.4	3.5	8.4	5.6	12.6
5	75	-0.6	-4.5	2.8	0.4	19.8
7	76	-2.6	-5.5	14.4	7	29.8
3	68	1.4	2.5	3.5	1.9	6.5
$\Sigma x = 48$	$\Sigma y = 776$			$\Sigma((\bar{x}-x)\cdot(\bar{y}-y)) = 149.8$	$\Sigma(\bar{x}-x)^2 = 92.5$	$\Sigma(\bar{y}-y)^2 = 314.7$

Extra Q2 Cont'

$\bar{x} = \frac{\sum x}{n} = \frac{48}{11} = 4.36$; $\bar{y} = \frac{\sum y}{n} = \frac{776}{11} = 70.55$

i) $r = \frac{\sum((\bar{x}-x)\cdot(\bar{y}-y))}{\sqrt{\sum(\bar{x}-x)^2}\cdot\sqrt{\sum(\bar{y}-y)^2}} = \frac{149.8}{\sqrt{92.5}\cdot\sqrt{314.7}} = \frac{149.8}{(9.62)(17.74)} = 0.878$

There is a strong positive linear relationship between the secretaries' experience and their typing speed.

ii) $r^2 = (0.878)^2$

$= 0.77$, 77% of the variability in typing speed can be explained by the variability of secretaries' experience.

iii) $S_y = \sqrt{\frac{\sum(\bar{y}-y)^2}{(n-1)}} = \sqrt{\frac{314.7}{(11-1)}} = 5.61$; $S_x = \sqrt{\frac{\sum(\bar{x}-x)^2}{(n-1)}} = \sqrt{\frac{92.5}{(11-1)}} = 3.04$; $\beta_1 = r\frac{(S_y)}{(S_x)}$

$= (0.878)\cdot\frac{5.61}{3.04}$

$= 0.878 \times 1.844$

$= 1.62$

$\beta_0 = \bar{y} - \beta_1\bar{x} = (70.55) - (1.62)(4.36)$

$= 70.55 - 7.06 \Rightarrow 63.48$

$\therefore$ The regression model $= y = 63.48 + 1.62x$

iv) $x = 9$, $y = 63.48 + 1.62 \times 9$

$y = 63.48 + 14.57 \Rightarrow 78.05$

v)