Technological University Dublin
Tallaght Campus
School of Enterprise Computing and Digital Transformation

Data Analysis
Pen and Paper Lab Assignment
Linear Regression

ANSWERS

Exercise 1 - Manual Linear Regression Model Fitting ANSWER

To fit a linear regresion model we must calculate:

- the mean of both x and y (independent and dependent variable): $\bar{x}, \bar{y}$
- the sum of squares of the differences of x values from the mean: $\sum(x - \bar{x})^2$
- the sum of cross-multiplication of x and y differences from the mean: $\sum(x - \bar{x})(y - \bar{y})$

Then we use these to calculate the slope (m) and the constant coefficient (c) of the linear model. This was done in a spreadsheet but in an exam the number of instances would be a lot smaller (e.g. 5), hence calculable manually in reasonable time.

	Test 1 (x)	Test 2 (y)					
	x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})^2$	$(x-\bar{x})(y-\bar{y})$	
	59	56	11.77	2.46	138.51	28.97	
	52	63	4.77	9.46	22.75	45.12	
	44	55	-3.23	1.46	10.44	-4.72	
	51	50	3.77	-3.54	14.21	-13.34	
	42	66	-5.23	12.46	27.36	-65.18	
	42	48	-5.23	-5.54	27.36	28.97	
	41	58	-6.23	4.46	38.82	-27.80	
	45	36	-2.23	-17.54	4.98	39.12	
	27	13	-20.23	-40.54	409.28	820.12	
	63	50	15.77	-3.54	248.67	-55.80	
	54	81	6.77	27.46	45.82	185.89	
	44	56	-3.23	2.46	10.44	-7.95	
	50	64	2.77	10.46	7.67	28.97	
mean: x, y	47.23	53.54					
square/crossmultiplication sums: $\Sigma(x-\bar{x})^2$, $\Sigma(x-\bar{x})(y-\bar{y})$					1006.31	1002.38	
m (slope): $\Sigma(x-\bar{x})(y-\bar{y})/\Sigma(x-\bar{x})^2$							1.00
c (constant coefficient): $\bar{y} - m\bar{x}$							6.49

Screenshot of spreadsheet `Lab4_2_Ex1_ANSWER.csv`, which can be downloaded alongisde this document.

$$m = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2} = 1.00$$

$$c = \bar{y} - m\bar{x} = 6.49$$

The linear regression model is:

y = 6.49 + x

Using the model we can predict the value of y in the case that $x = 46$:

y(x=46) = 6.49 + 46 = 52.49