

Further Statistics 2 Unit Test 1: Linear regression

- 1 The results of an experiment in which different masses were attached to an elastic string are shown in the table below. m is the mass attached, in grams, and x is the extension of the string, in mm.

Table 1

m	10	20	30	40	50
x	12.2	23.8	37.1	47.8	60.1

- a** Find the equation of the regression line of x on m .
Give your answer in the form $x = a + bm$ where a and b are constants to be found. **(2 marks)**
- b** Explain the significance of the value b . **(1 mark)**
- c** Use your regression line to estimate the value of x when:
- i** $m = 42.5$ grams,
 - ii** $m = 75$ grams. **(2 marks)**
- d** Explain, with a reason, which of the estimates in part **c** is most reliable. **(1 mark)**
- 2 The speed at which a sled goes down a constant incline is thought to be related to the mass placed on the sled. The table below shows the mass, m kg, and the speed v ms⁻¹.

Table 2

m	1.2	2.3	3.1	4.6	6.8
v	2.2	4.1	5.7	8.4	10.9

$$\sum m = 18 \quad \sum m^2 = 83.74 \quad \sum v = 31.3 \quad \sum v^2 = 243.51 \quad \sum mv = 142.5$$

- a** Calculate the values of S_{mm} and S_{mv} . **(2 marks)**
- b** Find the equation of the regression line of v on m in the form $v = a + bm$ where a and b are constants to be found. **(2 marks)**
- c** Explain the significance of the value a . **(1 mark)**
- d** Use your regression line to estimate that value of v when $m = 5.7$ kg. **(1 mark)**
- e** State, with a reason, whether this estimate is reliable. **(1 mark)**

- 3** A farmer plants 5 crops of tomatoes and provides each crop with a different amount of nutrient supplement. The amount of nutrient given is x grams, and the number of tomatoes produced is y .

The data is coded using $p = \frac{x - 700}{10}$ and $q = y - 65$

$$\sum p = 176 \quad \sum p^2 = 9488 \quad \sum q = 163 \quad \sum q^2 = 7883 \quad \sum pq = 8646$$

- a** Calculate the values of S_{pp} and S_{pq} . **(2 marks)**
- b** Find the equation of the regression line of y on x in the form $y = a + bx$, where a and b are constants to be found. **(5 marks)**
- c** Explain the significance of b in the model. **(1 mark)**
- 4** A particular model of van depreciates in value as it gets older. Below is a table showing the ages, x years, and the value, P £'000s, of a random sample of these vans.

Table 3

x	0.8	1.2	1.7	2.4	3.1	4.0
P	11.5	10.6	9.5	8.4	6.0	5.7

The equation of the regression line of P on x is given as $P = 12.869 - 1.9329x$.

- a** Explain the significance of the value 12.869. **(1 mark)**
- b** Calculate the residuals. **(2 marks)**
- c** Use your answer to part **b** to identify a possible outlier. **(1 mark)**
- d** State, with a reason, whether the outlier should be excluded. **(1 mark)**
- e** Ignoring the outlier, produce another model. **(2 marks)**
- f** Use the model to predict the value of a van which is six years old. **(1 mark)**
- g** State, with a reason, whether your prediction in part **f** is reliable. **(1 mark)**
- 5** A random sample of 8 pygmy possums is taken and the age, x days, and length, y cm, is recorded.

$$\sum x = 180 \quad \sum y = 45.9 \quad S_{xx} = 378 \quad S_{yy} = 2.63875 \quad S_{xy} = 31.35$$

- a** Find the equation of the regression line of y on x . Give your answer in the form $y = a + bx$ where a and b are constants to be found. **(3 marks)**
- b** Calculate the residual sum of squares (RSS). **(2 marks)**

Table 4 shows the residuals for each value of x .

Table 4

x	12	15	18	21	24	27	30	33
Residual	0.0338	-0.0149	-0.0636	0.0877	0.039	r	-0.0584	0.0929

- c** Find the value of r . **(2 marks)**
- d** By considering the signs of the residuals, explain whether or not the linear regression model is suitable for these data. **(2 marks)**
- 6** A scientist is measuring the amount of product produced by a chemical reaction when different amounts of a catalyst is added. She measures the amount of catalyst, x , in grams, and the amount of product, y , also in grams. The results are shown in the table below.

Table 5

x	2	3	4	5	6	7	8	9	10	11
y	6.1	6.4	6.8	7.2	7.5	7.9	8.0	8.5	8.9	9.2

$$S_{xx} = 82.5 \quad S_{yy} = 9.785 \quad S_{xy} = 28.35$$

- a** Show that the residual sum of squares (RSS) is 0.0429 correct to 3 significant figures. **(2 marks)**
- The equation of the y on x regression line is given as $y = 5.42 + 0.344x$.
- b** Calculate the residual values. **(3 marks)**
- c** Write down the outlier. **(1 mark)**
- d i** Comment on the validity of ignoring this outlier. **(1 mark)**
- ii** Ignoring the outlier, produce another model. **(2 marks)**
- iii** Use the new model to estimate the amount of product produced when 8 grams of catalyst is added. **(1 mark)**
- iv** Explain whether it would be sensible or not to use the model to predict the amount of product produced when 15 grams of catalyst is added. **(1 mark)**