**BALDIVIS SECONDARY COLLEGE**

** APPLICATIONS - Unit 3 & 4**

**2018 Test 1- Bivariate Data**

Student Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Teacher Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Time allowed for this task:** 55 minutes, in-class, test conditions.

Section 1: 15 minutes + 2 minutes reading time

Section 2: 35 minutes + 3 minutes reading time

**Materials required:** Section 1 Calculator free section (15 marks)

Standard writing equipment

SCSA Formula Sheet

Section2 Calculator assumed section (35 marks)

Calculator (to be supplied by the student)

SCSA formula Sheet

One page A4 (single sided) hand written notes

**Other materials allowed:** Drawing templates

**Marks available:** **50 marks**

**Task Weighting: 7%**

**Question 1. (3 marks)**

A linear fit is applied to a set of data.

The set of residuals is graphed on the set of axes below.

R

x

State with reasons, whether or not the linear fit was appropriate. (3)

**Question 2. (4 marks)**

Describe the association between the two variables that are graphed in the scatter diagrams below.

(a) (2)

(b) (2)

**Question 3. (8 marks)**

Researchers looking for an association between reading and numeracy levels of students compared achievement at six schools using NAPLAN test results. The graph below shows the average reading and numeracy scores for the schools and the linear regression model *y* =1.042*x* - 8.055 calculated from the data.



(a) Describe, in words, the association between the variables *x* (reading score) and *y* (numeracy score) for these schools in terms of direction and strength. (2)

(b) Consider the correlation coefficient *r* between the variables *x* and *y* .

(i) Place a cross on the scale below for your estimate of *r* . (1)



(ii) Would you expect *r* to increase or decrease if the data point (538, 570) was removed from the data set? Explain your answer. (1)

(d) One of the data points is at (562, 562). Would you predict that another school with a reading score of 562 will also have a numeracy score of 562? Justify your answer. (2)

(e) Another data point is at (634, 663). Write down a calculation to determine the residual for the linear regression model at this point, but do **not** evaluate it. (2)

**End of non-calculator Section**

 **Section 2 - Calculator Allowed**

**Total marks – 35 Working time 35 minutes**

**Question 4. (7 marks)**

Ben conjectured that many of the Year Ten girls had dyed their hair blonde. He did a survey of 100 randomly selected Year Ten students and the results are tabled below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Hair colour | | |  |
|  | Black | Brown | Blonde |  |
| Girls | 16 | 7 | 28 | 51 |
| Boys | 14 | 25 | 10 | 49 |
|  | 30 | 32 | 38 | 100 |

(a) Represent the numbers of each hair colour for girls and then for boys as row or column percentages as appropriate. (3)

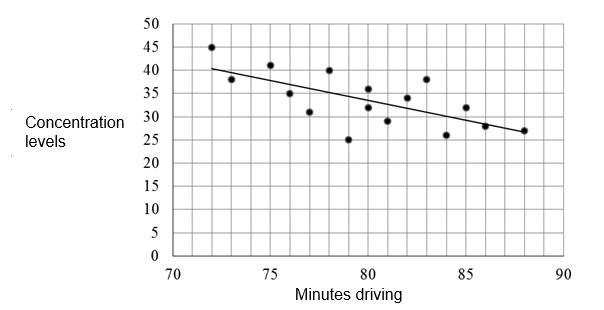
|  |  |  |  |
| --- | --- | --- | --- |
|  | Hair colour | | |
|  | Black | Brown | Blonde |
| Girls |  |  |  |
| Boys |  |  |  |

(b) Analyse the data and make a conclusion about the conjecture above. (4)

**Question 5 (12 marks)**

It has been conjectured that the longer one drives on country roads, the concentration of the driver lessens. A suitable test of concentration was devised. A study was done and the results are shown in the table and on the graph below.

|  |  |
| --- | --- |
| Minutes driving on country roads | Concentration levels |
| 72 | 45 |
| 73 | 38 |
| 75 | 41 |
| 76 | 35 |
| 77 | 31 |
| 78 | 40 |
| 79 | 25 |
| 80 | 32 |
| 80 | 36 |
| 81 | 29 |
| 82 | 34 |
| 83 | 38 |
| 84 | 26 |
| 85 | 32 |
| 86 | 28 |
| 88 | 27 |



1. Calculate the correlation coefficient (2)
2. Explain why there is 54% of unexplained variation. (2)

(c) Explain the relationship bewteen the sign of the correlation coefficient and the

gradient of the trend line shown on the graph. (2)

(d) Determine the equation of the regression line. (3)

(e) Estimate the concentration score and comment on the reliability of your estimate given the driver has driven for

(i) 74 minutes. (1)

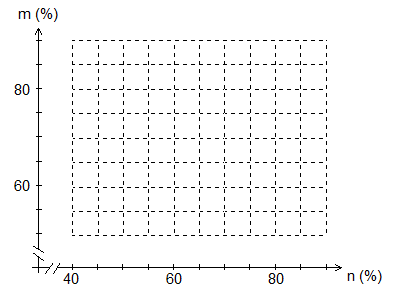
(ii) an hour and three quarters. (1)

(f) Comment on the reliability of your results in (e). (1)

**Question 6 (12 marks)**

The table and scatterplot show the percentage scores for a group of twelve candidates who took both numerical reasoning (n) and mechanical aptitude (m) tests as part of a job selection process.

|  |  |
| --- | --- |
| **n (%)** | **m (%)** |
| 47 | 61 |
| 82 | 86 |
| 67 | 75 |
| 66 | 67 |
| 71 | 69 |
| 56 | 65 |
| 68 | 72 |
| 66 | 73 |
| 52 | 66 |
| 81 | 80 |
| 77 | 82 |
| 75 | 80 |



(a) Construct the scatter plot on the graph above. (2)

(b) Determine the least squares regression line of *m* on *n*. (2)

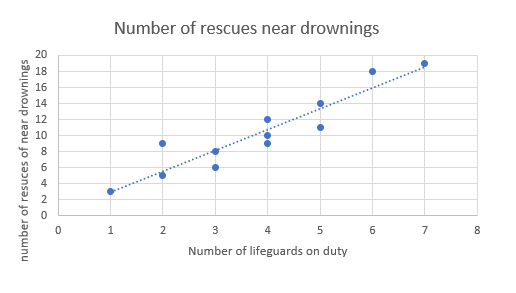
(c) Draw your line from (b) on the scatterplot above. (2)

(d) State the correlation coefficient, *rnm*, for the linear model in (b). (1)

(e) Describe two features of the scatterplot that can be determined from *rnm*  (2)

(f) A thirteenth candidate for the job scored 45 on the numerical reasoning test but due to illness was unable to complete the mechanical aptitude test. What mechanical aptitude score would you predict for this person? Explain, with reasons, how valid your prediction is. (3)

**Question 7** **(4 marks)**



1. Comment on the relationship between the number of lifeguards on duty and the number of rescues of near drownings. (2)
2. Jodie calculated the value of the correlation coefficient for the given data to be 0.96 (to 2.d.p.) and concluded that the more lifeguards were on duty, the more near drowning rescues occurred and hence it was probably better to reduce the number of lifeguards on duty. Discuss Jodie’s conclusions. (2)

**End of test**