Student Name ___ MARKING KEY

Eastern Goldfields College

Mathematics Applications U3&4 2017

Test 1 – Calculator Free

Total Marks: 17 marks

Time allowed: 10 minutes

No calculator or notes permitted for this section.

Answer all of the following questions. Show all working to obtain full marks.

Question 1 [3 marks]

For each of the following sets of variables, determine if there is likely to be a relationship, and if so, identify the explanatory and response variables.

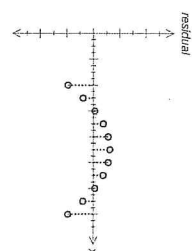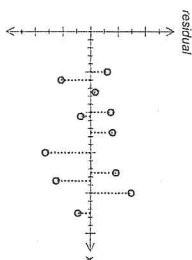|		Relationship Yes/No	Explanatory Variable	Response Variable
a)	Basketball ability and height (cm)	Y	Height	B. Ability
b)	Eye colour and Age (years)	N		
c)	Hair colour and Gender	N		
d)	Time spent cleaning and Size of house	Y	Size	Time Spent

Must have nothing here (red annotation)

Question 2 [7 marks = 3, 4]

(a) Match each of the scatter graphs with the appropriate correlation coefficient from the following list of values.

List of values: -2.3   -0.88   0.08   0.71   1.6

$r_{xy} = 0.71$ ✓

$r_{xy} = -0.88$ ✓

$r_{xy} = 0.08$ ✓
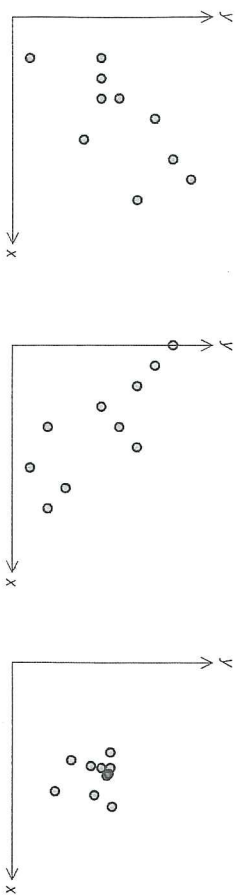
(b) Given the following **residual plots**, comment on the appropriateness of using a linear regression model for each.

• Appropriate ✓

• Residual appears random *(missing Residual 3 marks)* ✓

• No pattern in Residual ✓

• Not Appropriate ✓

• Residual has a pattern ✓

Question 3 [5 marks = 2, 1, 2]

The least squares regression line between N and t is given by $N = 0.05t + 2.51$

a) Find the average increase in N corresponding to an increase of 20 units in t

$20 \times 0.005 = 1$ ✓✓

b) Predict the value of N when $t = 50$.

$\hat{N} = 0.05(50) + 2.51$ ✓

$= 2.5 + 2.51$

$= 5.01$ ✓

c) The actual value of N when $t = 50$ is 4.95. Find the residual associated with the prediction in (b).

$Residual = 4.95 - 5.01$ ✓

$4.95 - 5.01$

$Residual = -0.06$ ✓

Question 4 [2 marks]

The variables x and y have a coefficient of linear correlation of 0.95. The least squares regression line of $y = 12.4 + 0.67x$. The residual plot associated with the least squares regression line is shown in the accompanying diagram.

Comment on the appropriateness of using the straight line model to describe the relationship between x and y.
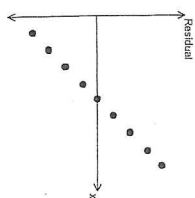
✓ • Not appropriate

✓ • pattern appears in the residual

must have both words. (red annotation)

Student Name _____

Eastern Goldfields College

Mathematics Applications 3&4 2017

Test 1 – Calculator Assumed

Total Marks: 33 marks

Time allowed: 45 minutes

One A4 page of notes IS permitted in this section.

Question 1   [10 marks – 1, 1, 2, 2, 2]

To investigate Year 12 student behaviour on Saturday afternoons, a group of Year 11 students conducted a survey in their school. They asked the Year 12 students "Which of the following activities would you most prefer to do on a Saturday afternoon?" Some of the results of the survey are given in the table below.

Percentage of students choosing the various activities

Activity	Males	Females
Watch reality television shows	4	5
Engage with social media	18	15
Play sport	36	28
Paid work	29	30
Shopping with friends	13%	22

(a) Enter the missing data value.

$100 - 28 = 72\%$ ✓

(b) What percentage of females did not prefer to play sport?

(c) There were 120 boys and 124 girls in Year 12 who responded to the survey.

(i) How many boys would prefer to play sport on Saturday afternoon?

$0.36 \times 120 = 43.2$
≈ 43 Boys. ✓

(ii) How many students would prefer to work on Saturday afternoon?

$0.29 \times 120 = 34.8$ ✓
$0.3 \times 124 = 37.2$ ✓
72 students ✓

(d) What type of graph would be best to display this data? Explain your choice of graph.

[handwritten annotations] graph, reason / explanation, eg: look stacked column graph "categorical variables", graph, "categorical" variables. ✗

(e) What can you learn from this data about the activities that students in Year 12 in **your** school would prefer to do on a Saturday afternoon? Explain your answer.

[handwritten] ✓ Nothing. Different school / demographic etc.

Question 2   [3 marks – 1, 2]

A survey into the number of 'likes' 👍 on a post and the number of friends in Facebook showed a correlation coefficient, r, equal to 0.69.

a) According to this result what percentage of the variation in number of likes could be accounted for by the variation in the number of friends?

$r^2 = (0.69)^2 = 0.4761$
$\approx 47.61\%$ ✓ variation in # of likes could be accounted for by variation in Facebook

b) Does this mean that the number of friends in Facebook influences the number of 'likes' on a post? Discuss and identify any reasons to support your answer.

[handwritten] No, it explain the relationship ✓

[red annotations] No, 47% ✓, OK lurking variable

Question 4   [10 marks – 2, 1, 1, 3, 2, 2]

Follow the activities suggested to investigate the claim that the "regression lines should only be used to for predictions when the data contain no obvious outliers."

The table below shows the marks for 11 students in the two different parts of a test.

Student	A	B	C	D	E	F	G	H	I	J	K
Part 1	10	12	12	9	10	12	14	7	12	14	12
Part 2	9	10	8	8	7	9	12	6	11	11	3

(f) Plot the data on the axes below.

Outlier
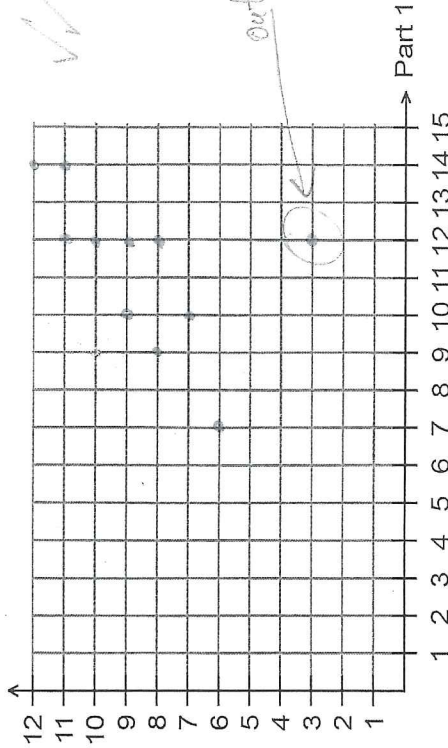
Part 1

Part 2

The regression line for the table data is
Part 2 mark = 1.31 + 0.64 × *Part 1 mark*

The coefficient of correlation, r, is 0.5224.

(f) Use this equation to predict a Part 2 mark for a student who scored 13 in Part 1.

$$Part\ 2 = 1.31 + 0.64 (13)$$
$$= 9.63$$

(f) Circle the point which represents the outlier.

Question 3   [9 marks – 3, 1, 2, 3]

The 437 students who completed a particular business management course came from two different colleges, Stack and Bosun. Information about the number of students from each college and their grades achieved is as follows:

- Every one of the 437 students received a grade of either A, B, C, D or E.
- 241 of the students were from Stack and 77 of these received an A grade.
- Altogether 145 B grades were awarded which was 44 more than the total number of A grades awarded.
- The number of Stack college students receiving a B grade was 16 more than the number from Stack receiving an A.
- Of the 158 students receiving a C grade, 60 were from Stack.
- 9 of the Stack students received a D.
- The number of Bosun students receiving an E was 4 times the number of Stack students receiving an E.

a) Copy and complete the following two way table showing the number of students from each college achieving each grade.

	A	B	C	D	E	Totals
Stack College	77	77+16 / 93	60	9	2	241
Bosun College	24	52	98	14	8 (2×4=)	196
Totals	145−44 / 101	145	158	23	10	437

b) Which is the explanatory variable and which is the response variable?

Exp = College
Resp = Grade

c) Complete the table below, showing either row or column percentages as appropriate.

	A	B	C	D	E	Totals
Stack College	31.9%	38.6	24.9	3.7	0.8	99.9
Bosun College	12	27	50	7	4	100
Totals						

d) Comment on whether there seems to be an association between the variables, explaining your reasoning and describing the association.

(d) Remove the outlier for the data and calculate the new

(i) correlation coefficient

$r = 0.866029\underset{\smile}{4} = 0.866 \ (3dp)$

(ii) equation for the regression line

Part 2 = 0.75 (Part 1) + 0.67 (2dp)

✓ notation ✓ m, c.

(e) Complete the following table by using the new equation to predict a Part 2 mark for each of the students listed. Using 'I' or 'E', indicate in the table whether the prediction is an interpolation or an extrapolation.

Part 1	13	5
Part 2	✓10.42 ~~10.45~~	4.42 ← wrong coeff 4.43 ✗ used stat's sech...
Interpolation or Extrapolation?	✓ I	E

Accept 10.4.

(f) Which of the two predictions, for the Part 1 score of 13, is the more reliable (answer for part (b) or (e)? Justify your answer.

(e)
outlier removed. ✓

~ **END OF TEST** ~