education equals

Topic: The Statistical Process and Mixed Bivariate Data Analysis

Time: 45 mins Marks: /45 marks

Calculator Assumed

Question One: [2, 2, 2: 6 marks]

a) Twenty marbles are selected randomly from a box containing marbles. These twenty marbles are intended to be a sample of the marbles in the box. Is this sample likely to be biased? Explain your answer.

b) A radio show asks people to phone in with their opinion after posing a question on a segment. 100 people phone in. Is this sample likely to be biased? Explain your answer.

c) A number of students from each year group, relative to the total number of students in the year, are randomly selected to take part in a survey. Is this group likely to be a biased representation of the students in the school? Explain your answer.

Question Two: [2, 2, 2, 1, 1, 1, 2: 11 marks]

Troye and his brother Tyde are interested in collecting and analysing data.

Troye wants to know about the exercise lifestyle of his fans. He plans to survey 25 of his fans.

a) Suggest two questions he could ask in his survey in order to gather this information.

Troye produces an online survey with yes/no questions for his fans to answer.

b) Suggest how Troye could improve the process of collecting data.

Tyde was investigating what type of exercise people aged 10 – 19 years old are involved in. He gathered the following data which came from the Bureau of Statistics.

Table 1: Participation Rate %

Sport	WA	Australia
Swimming and diving	17.2	16.5
Soccer (outdoor)	19.9	21.7
Australian Rules football	16.0	14.9
Netball	0.3	0.2
Basketball	8.5	9.2
Tennis	9.4	8.4
Martial arts	7.5	7.8
Gymnastics	1.5	1.7
Cricket (outdoor)	9.7	8.6
Rugby League	7.0	7.5
Athletics, track and field	3.0	3.2
Rugby Union	3.8	4.0
Touch football	1.9	2.1
Soccer (indoor)	4.3	2.5
Hockey	1.8	1.3
Other organised sports	10.2	11.8

c) Provide two reasonable explanations as to why neither column adds up to 100%.

d) What type of graph could be used to display this data?

e) Tyde concludes that outdoor soccer is less popular in WA than in the whole of Australia because it is too hot in WA. Using the data in the table provide a counter argument for his conclusion.

f) Which sport/s is/are more popular in WA than compared with Australia?

g) What are the most popular and least popular sports in Australia and where do these sports rank in WA?

Question Three: [3, 1, 2, 2: 8 marks]

The following table records the height and weight of nine adult females, as well as the length of their right femur bone.

	A	B	C	D	E	F	G	H	I
Height (cm)	162	160	164	165	163	180	171	190	195
Weight (kg)	65	70	73	90	76	80	78	85	100
Length of Femur (cm)	23	25	25	26	26	27	28	30	33

The least squares regression line for Height (*H*) and Length of Femur (*F)* is

$F = 0.2117H - 9.4631$ and $r_{HF} = 0.9187$

a) Calculate the least squares regression line for Weight (*W)* and Length of Femur (*F)* ` and the value for r_{WF}.

b) Use the linear regression given above to predict the length of someone's Femur given their height is 154 cm.

c) Comment on the reliability of this prediction.

d) If it is known that the 154 cm person weighs 66 kg, would it be more reliable to use this figure to predict the length of their Femur?

Question Four: [3, 2, 3, 2: 10 marks]

In Australia, girls from private schools are more likely to become injured while cheerleading than non-private school girls. The national cheerleading injury rate for private school girls is four times the national rate for all non-private school girls.

Age-specific cheerleading injury rates in Australia, 2013				
Age-group (years)	**Private schools (current students and alumni)**		**Non-private school girls (current and past)**	
	Frequency	% Frequency	Frequency	% Frequency
15-19	63	13.46	15	
20-24	134	28.63	52	
25-29	124	26.50	100	
30-34	88	18.80	125	
35-39	48	10.26	71	
40-44	11	2.35	15	

a) Complete the % Frequency column for non-private school girls.

b) Comment on the injury rates for women in Australia.

The following table shows the cheerleading injury rates for 15 – 19 year old women in three American cities. Injury rates are measured in injuries per 1 000 women.

US State	**Cheerleading injury rate**	**Total number of girls ages 15 - 19**	**Actual number of injuries for 15 – 19 year olds**
New York	63	400 217	
California	59	845 557	
South Carolina	65	93 306	

c) Using the injury rates calculate the actual number of injuries for each American city in the table above.

There are approximately 733 000 women in Australia aged 15 – 19 years old and approximately 11 000 15 – 19 year olds recorded cheerleading injuries in the last 12 month.

d) Compare the cheerleading injury rate of Australian and American teenagers.

Question Five: [4, 4, 1, 1: 10 marks]

Amanda wants to predict how much a plant will grow given the amount of water given each week. In her experiment she monitors 7 plants. The first is given 1 litre of water each week, the second 2 litres, the third 3 litres and so on. Over the next 6 weeks she measures the weekly average growth of each plant.

The least squares regression line for growth of plant (G) in cm based on litres of water (L) is:

$\hat{G} = 0.023L + 1.5 \quad r = 0.62$

Caleb did the same experiment but with 4 plants and the least squares regression line calculated from the data he collected is $\hat{G} = 0.101L + 1.2 \quad r = 0.75$

a) Compare the growth predicted using each of their regression lines for 6.5 L of water and 15 L of water and comment on the reliability of the predictions.

b) The residual when predicting the growth rate using 3 L of water is 1.2 cm for Amanda's data and -0.5 cm for Caleb's data. What were the actual growth rates recorded when the plants were given 3 L of water?

c) Whose plants recorded a higher rate of growth and state this rate?

d) Give one reasonable explanation as to why the growth rates may differ.

education equals

Topic: SOLUTIONS

Time: 45 mins Marks: /45 marks

No calculator allowed

Question One: [2, 2, 2: 6 marks]

a) Twenty marbles are selected randomly from a box containing marbles. These twenty marbles are intended to be a sample of the marbles in the box. Is this sample likely to be biased? Explain your answer.

Random sample is not likely to be biased. ✓ ✓

b) A radio show asks people to phone in with their opinion after posing a question on a segment. 100 people phone in. Is this sample likely to be biased? Explain your answer.

✓ Since participation is limited to those listening to that particular radio station, the people who decide to participate may not be similar to the people who do not participate. This would likely be a biased sample. ✓

As participants are required to expend effort only those with strong opinions about the subject are likely to phone in.

c) A number of students from each year group, relative to the total number of students in the year, are randomly selected to take part in a survey. Is this group likely to be a biased representation of the students in the school? Explain your answer.

✓ Since the number of students selected from each year are being chosen randomly and the number of students is stratified to reflect the number of students in each year group, this sample is not likely to be biased. ✓

Question Two: [2, 2, 2, 1, 1, 1, 2: 11 marks]

Troye and his brother Tyde are interested in collecting and analysing data.

Troye wants to know about the exercise lifestyle of his fans. He plans to survey 25 of his fans.

a) Suggest two questions he could ask in his survey in order to gather this information.

How much time do you spend exercising? What type of exercise do you enjoy doing?

✓ ✓

Troye produces an online survey with yes/no questions for his fans to answer.

b) Suggest how Troye could improve the process of collecting data.

Provide a range of times and allow people to tick the appropriate box. Interview people face to face rather than sending an online survey.

✓ ✓

Tyde was investigating what type of exercise people aged 10 – 19 years old are involved in. He gathered the following data which came from the Bureau of Statistics.

Table 1: Participation Rate %

Sport	WA	Australia
Swimming and diving	17.2	16.5
Soccer (outdoor)	19.9	21.7
Australian Rules football	16.0	14.9
Netball	0.3	0.2
Basketball	8.5	9.2
Tennis	9.4	8.4
Martial arts	7.5	7.8
Gymnastics	1.5	1.7
Cricket (outdoor)	9.7	8.6
Rugby League	7.0	7.5
Athletics, track and field	3.0	3.2
Rugby Union	3.8	4.0
Touch football	1.9	2.1
Soccer (indoor)	4.3	2.5
Hockey	1.8	1.3
Other organised sports	10.2	11.8

c) Provide two reasonable explanation as to why neither column adds up to 100%.

People may participate in more than one sport. Some people surveyed may not participate in any sport. It could also be due to rounding. ✓ ✓

d) What type of graph could be used to display this data?

Bar or column graph ✓

e) Tyde concludes that outdoor soccer is less popular in WA than in the whole of Australia because it is too hot in WA. Using the data in the table provide a counter argument for his conclusion.

Cricket is played in summer and is played outdoors and is more popular in WA compared with Australia. ✓

f) Which sport/s is/are more popular in WA than compared with Australia?

Australian Rules Football, Tennis, Cricket, Soccer (indoor) and Hockey

✓

g) What are the most popular and least popular sports in Australia and where do these sports rank in WA?

The most popular sport in Australia is outdoor soccer and this is also the most popular sport in WA. The least popular sport in Australia is netball and this is also the least popular sport in WA.

Question Three: [3, 1, 2, 2: 8 marks]

The following table records the height and weight of nine adult females, as well as the length of their right femur bone.

	A	B	C	D	E	F	G	H	I
Height (cm)	162	160	164	165	163	180	171	190	195
Weight (kg)	65	70	73	90	76	80	78	85	100
Length of Femur (cm)	23	25	25	26	26	27	28	30	33

The least squares regression line for Height (H) and Length of Femur (F) is

$F = 0.2117H - 9.4631$ and $r_{HF} = 0.9187$

a) Calculate the least squares regression line for Weight (W) and Length of Femur (F) ` and the value for r_{WF}.

$\hat{F} = 0.2407W + 7.821\ (4dp)\ \ r_{WF} = 0.8596$

✓ ✓ ✓

b) Use the linear regression given above to predict the length of someone's Femur given their height is 154 cm.

Femur length ~23.14 cm (2 dp) ✓

c) Comment on the reliability of this prediction.

Not reliable since it is an extrapolation.

✓ ✓

d) If it is known that the 154 cm person weighs 66 kg, would it be more reliable to use this figure to predict the length of their Femur?

Yes this would be more reliable as it would be an interpolation and the correlation coefficient is also high for weight vs Femur.

✓ ✓

Question Four: [3, 2, 3, 2: 10 marks]

In Australia, girls from private schools are more likely to become injured while cheerleading than non-private school girls. The national cheerleading injury rate for private school girls is four times the national rate for all non-private school girls.

Age-specific cheerleading injury rates in Australia, 2013				
Age-group (years)	**Private schools (current students and alumni)**		**Non-private school girls (current students and alumni)**	
	Frequency	% Frequency	Frequency	% Frequency
15-19	63	13.46	15	3.97
20-24	134	28.63	52	13.76 ✓
25-29	124	26.50	100	26.46 ✓
30-34	88	18.80	125	33.07
35-39	48	10.26	71	18.78
40-44	11	2.35	15	3.97

a) Complete the % Frequency column for all women.

Total frequency = 378 ✓

b) Comment on the injury rates for women in Australia.

The most number of injuries for all women was recorded for women aged between 25 and 34 years of age. More private (former private) school girls appear to be getting injured younger with most injuries occurring between age 20 and 29. ✓ ✓

The following table shows the cheerleading injury rates for 15 – 19 year old women in three American cities. Injury rates are measured in injuries per 1 000 women.

US State	**Cheerleading injury rate**	**Total number of girls ages 15 - 19**	**Actual number of injuries for 15 – 19 year olds**
New York	32	400 217	$400217 \div 1000 \times 32 = 12\,800$ ✓
California	19	845 557	16 100 ✓
South Carolina	13	93 306	1 200 ✓

c) Using the injury rates calculate the actual number of injuries for each American city in the table above.

There are approximately 733 000 women in Australia aged 15 – 19 years old and approximately 11 000 15 – 19 year olds recorded cheerleading injuries in the last 12 month.

d) Compare the cheerleading injury rate of Australian and American teenagers.

$$11\,000 \div 733\,000 \times 1000 \approx 15$$

The cheerleading injury rate for Australian teenagers is 15 compared to three American cities, 32, 19 and 13. The cheerleading injury rate for Australian teenagers is lower than the average rate of these three cities but without knowing more statistics about American teenagers we cannot make any conclusive statements about all American teenagers. ✓✓

Question Five: [4, 4, 1, 1: 10 marks]

Amanda wants to predict how much a plant will grow given the amount of water given each week. In her experiment she monitors 7 plants. The first is given 1 litre of water each week, the second 2 litres, the third 3 litres and so on. Over the next 6 weeks she measures the weekly average growth of each plant.

The least squares regression line for growth of plant (G) in cm based on litres of water (L) is:

$\hat{G} = 0.023L + 1.5 \quad r = 0.62$

Caleb did the same experiment but with 4 plants and the least squares regression line calculated from the data he collected is $\hat{G} = 0.101L + 1.2 \quad r = 0.75$

a) Compare the growth predicted using each of their regression lines for 6.5 L of water and 15 L of water and comment on the reliability of the predictions.

Amanda: $6.5\ L \rightarrow \hat{G} = 1.6495$ cm $\quad 15\ L \rightarrow \hat{G} = 1.845\ cm$ ✓

Caleb: $6.5\ L \rightarrow \hat{G} = 1.8565$ cm $\quad 15\ L \rightarrow \hat{G} = 2.715\ cm$ ✓

Even though Caleb's data has a higher correlation coefficient the prediction based on 6.5 L of water is an extrapolation from Caleb's data but is an interpolation according to Amanda's data. This makes Amanda's prediction more reliable when compared to Caleb's but her prediction would not be very reliable itself because the correlation is low. Neither prediction would be reliable for 15 L because it is extrapolating from both sets of data. ✓ ✓

b) The residual when predicting the growth rate using 3 L of water is 1.2 cm for Amanda's data and -0.5 cm for Caleb's data. What were the actual growth rates recorded when the plants were given 3 L of water?

Amanda: $\hat{G} = 1.569$ ✓ $\quad G = 1.569 + 1.2 = 2.769$ cm ✓

Caleb: $\hat{G} = 1.503$ ✓ $\quad G = 1.503 - 0.5 = 1.003$ cm ✓

c) Whose plant has the higher rate of growth and state this rate?

Caleb's $\quad$ 0.101 cm per litre ✓

d) Give one reasonable explanation as to why the growth rates may differ.

Plants need sunlight to grow, perhaps Caleb's plants had more sunlight. ✓