Name: Solutions. Date: _____

Baldivis Secondary College

Year 11 Applications

Test 5, 2019

Topics – Univariate Data and Comparing Data

	50

Total Time: *55 minutes*

Total Reading: *2 minutes*

Total Working: *53 minutes*

Weighting: *6% of the year.*

Equipment: *SCSA Formula Sheet; 1 page notes (A4 one side, **Unfolded**), CASIO ClassPad; Scientific Calculator*

Resource Free Section – 18 min 1 min reading time **[18 marks]**

1. [3 marks]

Circle the correct data type for the following examples:

(i) The number of languages spoken by each person at an airport

A. discrete numerical

B. nominal categorical

C. continuous numerical

D. ordinal categorical

(ii) The shoe size of all year 11s at Baldivis Secondary College

A. discrete numerical

B. nominal categorical

C. continuous numerical

D. ordinal categorical

(v) Everyone at the gym on a Saturday was asked to describe their fitness level as (low, medium or high)

A. discrete numerical

B. nominal categorical

C. continuous numerical

D. ordinal categorical

2. [3 marks]

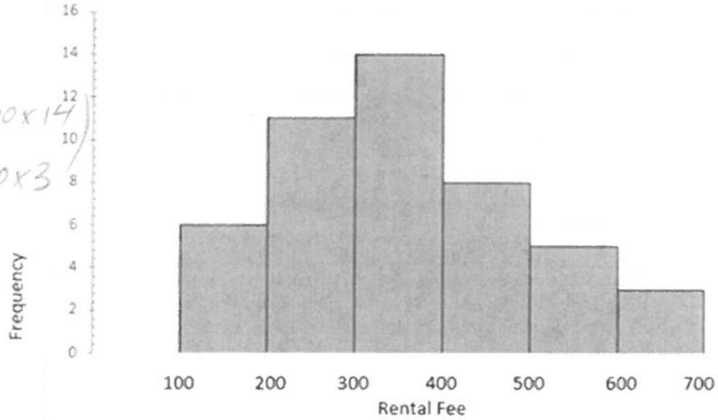
The histogram below shows the weekly rents for a number of households. Comment on the data.

mean calc

(100×6) (200×11) (300×14) (400×8) (500×5) 600×3

Total household = 47

14500 ÷ 47



- fairly symmetrical with a slight positive skew
- mean 14500 ÷ 47 = $308.51
- The mean rental fee was 308.51
- The data is unimodal with rents between $300-400
- The range is $700-$100 = $600

3. [12 marks: 6, 1, 2, 3]

The weights of 15 year old boys and girls were recorded to the nearest kilogram and listed below:

BOYS: 55, 75, 92, 48, 60, 74, 58, 84, 44, 74, 50, 56, 82, 49, 58, 69, 72, 64, 46, 62, 81, 48, 53, 78, 54, 68, 60, 52, 66, 56

GIRLS: 53, 39, 66, 40, 58, 70, 50, 38, 63, 60, 42, 36, 74, 72, 56, 66, 37, 43, 72, 64, 43, 44, 61, 49, 56, 53, 48, 53, 50, 38

a) Display the given information using a back-to-back stem and leaf plot

Boys		Girls
	3	67889
98864	4	0233489
8866543200	5	0033366
964200	6	014366
85442	7	0224
421	8	
2	9	

Key leaf stem
81 = 1 | 8

b) How many of these boys and girls were at most 66 kilograms? (not more than)

54 → need to include 66kg.

c) Which is the most appropriate measure of central tendency that should be used to compare the centres of the data sets? Explain your answer.

median should be used

* data for both boys is skewed & median (middle value) will be less affected by more extreme scores. The median score will still lie in the middle.

* the data is numerical & in order so easy to mark off & find middle value.

d) Write a report (3 sentences) comparing the distribution of weights of these groups?

Shape: The boys data set is skewed to the left slightly. The girls data set is mostly symmetrical.

location: median of the boys is 69, while the girls median was 56. The boys median was larger

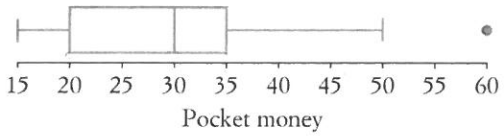
Spread: The range of the boys is 48 (92-44). The girls range is 38 (74-36). So the boys range is considerably larger.

-End Section 1-

Resource Section – 35 min 1 min reading time NAME: _____ [32marks]

4. [8 marks, 1,1,3,3]

The following boxplot shows the amount of pocket money ($) 20 children were paid per month.



a) What is the median pocket money paid?

$30

b) Complete the sentence:

'Twenty-five per cent of children had less than $20 of pocket money per month.'

c) How many children were given pocket money of $35 or less?

0.75 x 20
= 15 students

d) Prove why the data value of $60 is an outlier.

1.5 x IQR + Q3
= 1.5 x 15 + 35
= 57.5
Any data greater than 57.5 is an outlier

IQR = Q3 - Q1
= 35 - 20
= 15

5. [12 marks: 10, 2]

The lengths of words in typical definitions in two biology books were counted and found as follows:

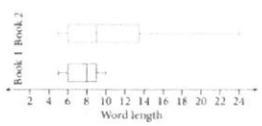
Book 1: 5, 8, 10, 6, 9, 6, 10, 8, 5, 6, 8, 10, 8, 7, 9, 7, 8, 10

Book 2: 12, 8, 5, 24, 7, 9, 5, 5, 22, 5, 13, 9, 10, 9, 5, 17, 14, 8, 14, 7

a) Construct parallel box plots to represent the data on the additional graph paper.

b) Compare, using appropriate summary statistics, the difficulty of the text, assuming that longer words make a text harder to read.

a

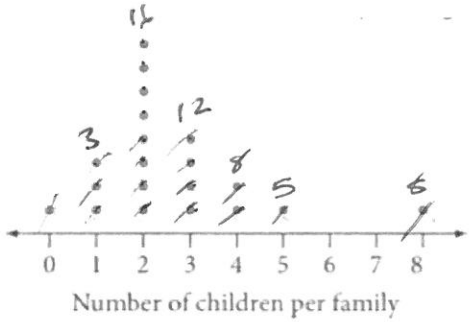


b Book 2 is harder as it has longer words in it, indicated by its higher median (9 compared to 8 and mean (10.4 compared to 7.8).

c Book 1 has a more consistent word length, indicated by its lower standard deviation (1.7 compared to 5.5).

6. [8 marks: 1, 2, 1, 1, 1, 2]

This dot plot shows the number of children in each family living on Thredbo Drive.



a) How many families live on Thredbo Drive?

20

b) Calculate the mean number of children per family.

0, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4, 5, 8

mean = 2.6

c) What is the outlier?

The family that has 8 students.

d) If the outlier is removed from the data set, how is the mean affected?

The mean with outlier in is 2.6 children.
When outlier removed the mean drops to 2.3 child.
∴ the mean reduces.

e) Give a reason why the mean is not a good measure of centre for this data?

The mean is not a good measure as the data is skewed. Even when the outlier is removed the median still stays the same at 2 children.

f) An extra family moved into Thredbo Drive, increasing the mean number of children to 3. How many children did the new family have?

Total children now = 52

For new mean = 21 x 3
= 63

So the new family has added 63 - 52 = 11 children to the street

7 a 20 b 2.6
c 8 d decreases
e An outlier is present and 2.6 children is not realistic, as data is discrete.
f 11

7. **[4 marks: 2, 2]**

A student went for a run every morning during the month of June along a variety of routes. The time taken for each run was recorded to the nearest minute.

Run time (minutes)	Number of days
16-18	2
19-21	4
22-24	5
25-27	10
28-30	4
31-33	3
34-36	2

The mean of these 30 times is 25.8 minutes and the standard deviation is 4.6 minutes.

The following month the student continued their morning runs. The 31 times for July are shown in the table below.

Run time (minutes)	Number of days
13-15	4
16-18	4
19-21	3
22-24	7
25-27	8
28-30	3
31-33	2

a) Find the mean and standard devitaion fpr the month of July

$\bar{x} = 22.7$ minutes

$\sigma_X = 5.2$ minutes

b) In which month were the run times more consistent? Justify your answer with reference to suitable statistics

The June times were more consistent.

This can be justified by noting that the standard deviation for June is smaller than that of July (4.6 < 5.2).

-End of Section 2-