

CONTENTS

Unit 3 – Area of Study 1 – Data Analysis	1
Types of Data	2
Types of Variables	2
Univariate Analysis	4
Organising Data	4
Visual Displays	6
Using the TI-Nspire Calculator	12
Using the Casio Classpad Calculator	13
Using Logarithm Scales	15
Shapes of Distributions	19
Parameters in Statistics	21
Measures of Central Tendency	21
Mode	22
Median	22
Mean	22
Measures of Spread	30
The Range	30
Interquartile Range	30
Boxplots	34
The Standard Deviation	39
The Normal Model for a Bell-Shaped Distribution	41
Standard (Z) Scores	45
Samples and Populations	47
Bivariate Data	48
Two Categorical Variables	48
One Categorical and One Numerical Variable	51
Bivariate Analysis	58
Numerical and Numerical	58
The Strength of a Linear Relationship	63
Pearson’s Product Moment Correlation Coefficient / the ‘r’ Value	63
The Coefficient of Determination	65
Correlation and Causation	68
Regression – Least-Squares Regression Line	71
Calculating the Least-Squares from Summary Data	71
Calculating the Least-Squares from Raw Data	75

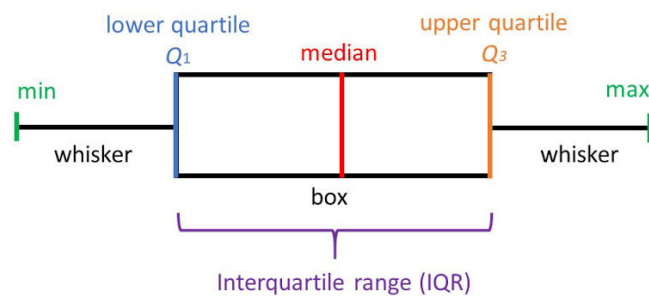
Interpolation, Extrapolation and Interpretation	79
Residuals and Residual Plots	80
Types of Residual Plots	81
Transformations	86
Transformations Using y^2 or x^2	88
Transformations Using $\frac{1}{y}$ or $\frac{1}{x}$	89
Transformations Using $\log_{10} y$ or $\log_{10} x$	90
Choosing the Correct Transformations	92
Testing Transformations	95
Displaying, Summarising and Describing Time Series Data	100
Trends	100
Cycles	101
Seasonality	101
Structural Changes	102
Outliers	102
Irregular (Random) Fluctuations	103
Smoothing	104
Deseasonalising Data	108
Predicting from Time Series	110
Exam 2 Style Questions	114
Unit 3 – Area of Study 2 – Recursion and Financial Modelling	120
Investments and Loans	121
Recursion	122
Investments	124
Simple Interest	124
Compound Interest	130
Nominal and Effective Interest Rates	135
Depreciation of Assets	137
Flat Rate Depreciation	137
Reducing Balance Depreciation	141
Unit Cost Depreciation	146
Reducing Balance Loans	148
Interest Only Loans	158
Annuity Investments	159
Annuities	164
Perpetuities	166
Additional Questions	168
Solutions	174

MEASURES OF SPREAD

THE RANGE

- The range is the difference between the smallest value and the largest value and gives an indication of the spread of results.
- $\text{Range} = X_{\max} - X_{\min}$
- For Ungrouped Data: Arrange values in order and select the lowest and highest values and then apply formula.
- For Grouped Data: Read off the smallest value in the first interval and the largest value on the last interval and subtract.
- The range is heavily influenced by outliers, because if outliers are present they will be the maximum or minimum values.

INTERQUARTILE RANGE



- One way of overcoming problems of outliers is to exclude the top and bottom quarters of the data. The range of the remaining data is the interquartile range (IQR).
- IQR = The range of values between which the middle 50% of the values lie.
- This measure of spread does not take into account all of the actual results and is not affected by outliers.
- IQR:
 - List data in order of value.
 - Locate the median. This is the 2nd quartile (the value below which half or 50% of the results lie).
 - Find the median of the lower half of the values (remember to exclude the median value in your data set if the data set is odd numbered). This median is the 1st quartile (the value below which one quarter or 25% of the results lie).
 - Find the median of the upper half of the values (remember to exclude the median value in your data set if odd numbered). This median is the 3rd quartile (the value below which three quarters or 75% of the results lie).
- $IQR = Q_3 - Q_1$

EXAMPLE 5

For the stem plot, find:

- (a) The range.
 (b) The interquartile range using the stem plot.

Stem	Leaf
0	8
1	8 9
2	2 7 8
3	0 1 3 6 9
4	2 4 5 5 6 7 9
5	4 7 8 8 9 9
6	0 0 0

5 | 4 means 54 kg

Solution

1. Calculate range from lowest and highest scores.

$$\begin{aligned}
 \text{Range} &= \text{Highest score} - \text{Lowest score} \\
 &= x_H - x_L \\
 &= 60 - 8 \\
 &= 52
 \end{aligned}$$

2. Count the number of scores, n .

$$n = 27$$

3. Find the median by using the rule

$$\text{Median score} = \frac{n+1}{2} \text{th score.}$$

$$\text{Median score} = \frac{n+1}{2} \text{th score}$$

$$= \frac{27+1}{2} = 14 \text{th score}$$

Median is 14th scores

Median is 45 kg

4. Find the lower quartile, Q_1 (1st quartile) from the 13 scores in the lower half of the data set.

Q_1 is 7th score

$Q_1 = 30$ kg

5. Find the upper quartile, Q_3 (3rd quartile) from the 13 scores in the upper half of the data set.

Q_3 is 7th last score

$Q_3 = 58$ kg

6. Calculate the interquartile range from the difference between the 1st and 3rd quartiles.

$$\begin{aligned}
 IQR &= Q_3 - Q_1 \\
 &= 58 - 30 \\
 &= 28
 \end{aligned}$$

QUESTION 18

The number of hours worked in the past week by all 24 part-time casual workers for a department store are summarised in the stem & leaf plot.

Number of hours worked in a week by 24 part-time casual workers.

Stem	leaf
0	
0 ⁵	5 5 8 8 9 9 9
1 ⁰	2 2 3 4 4 4
1 ⁵	5 5 5 5 7
2 ⁰	0 4
2 ⁵	5 8
3 ⁰	2 3

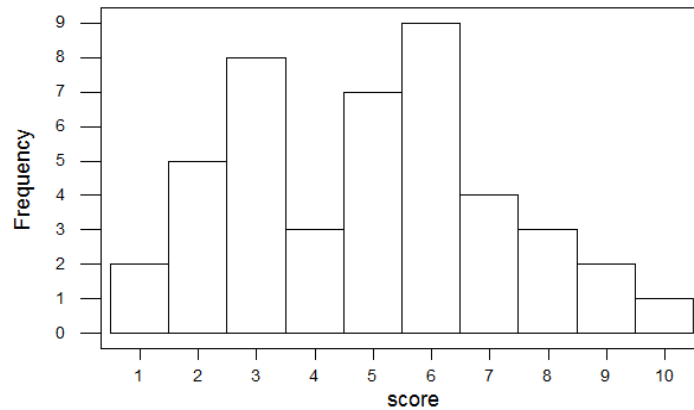
2⁰|4 means 24 hours

- (a) State the type of data and the shape of distribution.

- (b) Choose the most appropriate measure of central tendency and spread. Give a reason for your choice.

- (c) Calculate the centre and spread chosen in (b).

The display below relates to Questions 19 and 20.



QUESTION 19

For the display shown above, the IQR would be calculated using:

- A $7 - 2 = 5$
- B $7 - 3 = 4$
- C $7 - 4 = 3$
- D $6 - 3 = 3$
- E $6 - 4 = 2$

QUESTION 20

The mean score for this data set would be closest to:

- A 4.8
- B 4.9
- C 5.0
- D 5.1
- E 5.2

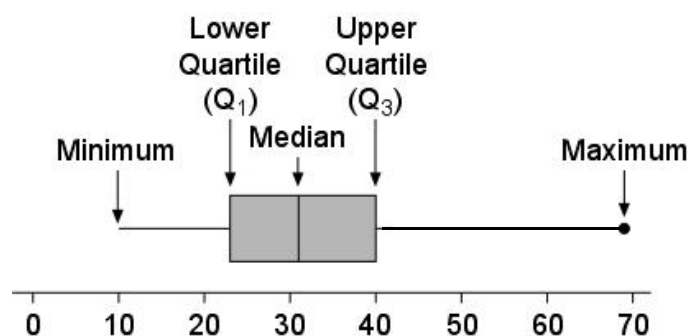
QUESTION 21

A survey was conducted of 35 customers in a restaurant and from the sample the average age of diners was 41 years, with a range of 37 years. A new customer walks into the restaurant who is aged 70. If the youngest person in the restaurant was 22 years of age, what would be the effect on the mean and range?

- A The range will remain the same but the average will increase.
- B Both the range and the average age will increase.
- C The average will increase but the range will remain the same.
- D Both the range and average will decrease.
- E Neither the range nor the average will be affected.

BOXPLOTS

A **boxplot** is a graphical display of five summary statistics: x_{\min} , Q_1 , Q_2 , Q_3 and x_{\max} .



- The IQR is the interquartile range or middle 50% of the data represented as a box.
- The lines drawn from the box to the extreme values are the whiskers.
- The total length of the plot from the end of one whisker to the end of the other represents the range. The longer the whiskers, the greater the range.
- Outliers are not included in the whiskers – they are represented as crosses, but outliers are **included** in the range.
- A value is classified as an outlier if:
 - **Graphically** – its distance from the nearer quartile is greater than 1.5 times the IQR.
 - **Numerically** – an outlier lies below $Q_1 - 1.5 \times \text{IQR}$ or above $Q_3 + 1.5 \times \text{IQR}$.
- The median is marked in the box as a vertical line.
- If the data has a symmetrical distribution, the median is drawn in the middle of the box.
- If the data is positively skewed, the line is drawn on the left-hand side of the box. (Median will be closer in value to the minimum than maximum.)
- If the data is negatively skewed, the line is drawn on the right-hand side of the box. (Median will be closer in value to the maximum than the minimum.)


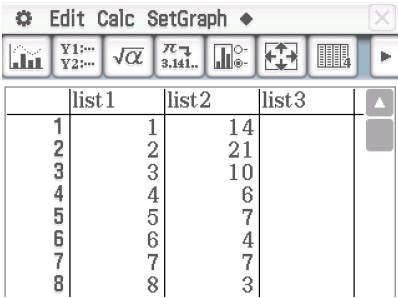


EXAMPLE 6

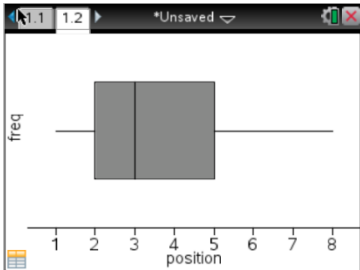

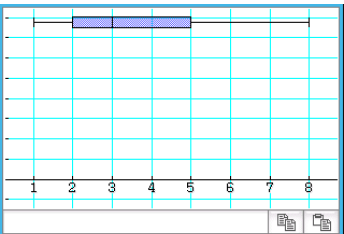
The final position on the ladder for a local footy team during its 72 year history was summarised in a frequency table.

Use the data from the frequency table to construct a boxplot and comment on the type of distribution.

Position	Frequency
1	14
2	21
3	10
4	6
5	7
6	4
7	7
8	3

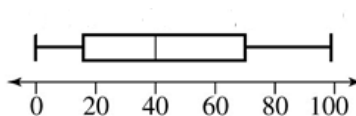
Solution

TI-Nspire	Casio ClassPad
<p>Control → I → Menu → 4. Add Lists & Spreadsheet Control → I → Menu → 5. Add Data & Statistics Enter data into lists as shown and name lists</p> 	<p>Enter data in Statistics menu</p> 
<p>On Data and Statistics page add variable position → Menu → 2. Plot properties → 9. Add Y summary list → select freq</p> 	<p>SetGraph → Setting → set as shown for a frequency table. Tick show outliers → set</p> 

TI-Nspire	Casio ClassPad
<p>Menu → 1.Plot type → 2. Boxplot</p>  <p>The image shows a TI-Nspire calculator screen displaying a boxplot. The x-axis is labeled 'position' and ranges from 1 to 8. The y-axis is labeled 'freq'. The boxplot has a minimum at 1, a first quartile at 2, a median at 3, a third quartile at 5, and a maximum at 8.</p>	<p> will draw boxplot in lower screen</p>  <p>The image shows a Casio ClassPad screen with a grid. A boxplot is drawn in the lower portion of the screen. The x-axis is labeled 1 to 8. The boxplot has a minimum at 1, a first quartile at 2, a median at 3, a third quartile at 5, and a maximum at 8.</p>
<p>Menu → 4.Analyse → A.Graph trace will allow you to trace the important features</p>	<p>Analysis → trace→ use 4 way key to trace the important features</p>
<p>Upper whisker and half of the box have a large range of value for top 50% of the data. This is positively skewed.</p>	

QUESTION 22

The range and the IQR, respectively, for the boxplot below are:



- A 100 and 40
- B 99 and 54
- C 99 and 40
- D 99 and 65
- E 90 and 54

QUESTION 23

Use a CAS calculator to display the following data sets as boxplot. Describe the distribution displayed in the boxplot in terms of shape, centre, spread and outliers.

Score, x	Frequency, f
4	3
5	7
6	11
7	6
8	3

QUESTION 24

An ordered data set of 11 heights in cm is given below:

160, 161, 165, 166, 166, 167, 167, 168, 170, 176, 205

A boxplot of the data set would be best drawn with its upper whisker ending at the value:

- A 205
- B 176
- C 170
- D 177.5
- E 175

QUESTION 25

Demonstrate that there are no outliers in the data set below:

21, 23, 34, 36, 37, 41, 44, 45, 46, 50, 56, 68, 75

Solution

QUESTION 26

In a data set, the value 90 would not be an outlier, but the value 91 would be an outlier. If all data values are whole numbers and the upper quartile is 60, then the lower quartile:

- A Must be less than 40.
- B Must be exactly 40.
- C Must be greater than 40.
- D Must be exactly 30.
- E Must be greater than 30.