Formula Sheet

Core – Data analysis

standardised score	$z = \dfrac{x - \bar{x}}{s_x}$
lower and upper fence in a boxplot	lower $Q_1 - 1.5 \times IQR$ upper $Q_3 + 1.5 \times IQR$
least squares line of best fit	$y = a + bx$, where $b = r\dfrac{s_y}{s_x}$ and $a = \bar{y} - b\bar{x}$
residual value	residual value = actual value – predicted value
seasonal index	seasonal index = $\dfrac{\text{actual figure}}{\text{deseasonalised figure}}$

Core – Recursion and financial modelling

first-order linear recurrence relation	$u_0 = a, \quad u_{n+1} = bu_n + c$
effective rate of interest for a compound interest loan or investment	$r_{effective} = \left[\left(1 + \dfrac{r}{100n}\right)^n - 1\right] \times 100\%$

Module 1 – Matrices

determinant of a 2 × 2 matrix	$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \det A = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$
inverse of a 2 × 2 matrix	$A^{-1} = \dfrac{1}{\det A}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$, where $\det A \neq 0$
recurrence relation	$S_0 =$ initial state, $S_{n+1} = T S_n + B$

Module 4 – Graphs and relations

gradient (slope) of a straight line	$m = \dfrac{y_2 - y_1}{x_2 - x_1}$
equation of a straight line	$y = mx + c$

Significant figures vs. Decimal places

- ## Significant Figures

> 273.245 = 273.25
> (to 2 decimal places)
>
> 273.245 = 270
> (to 2 significant figures)

→ **All non- zero values are significant**

4 . 2 (2 sig figs)

→ **All zeros in between are significant**

40002 (5 sig figs)

Or, in the case of decimal values: **4 . 0002** (5 sig figs)

→ **Decimal values**

1. All **final zeros** after the decimal point are significant

   4 . 200 (4 sig figs)

2. All **leading zeros** after a decimal point are **NOT** significant

   0 . 000**422** (3 sig figs)

→ **Terminal zeros don't count UNLESS there is a decimal point at the end**

420 (2 sig figs)

420 . (3 sig figs)

420 . 0 (4 sig figs)

- ## Decimal places

→ Involves rounding values after the decimal point to however many decimal places

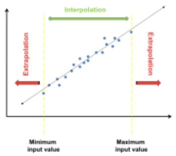
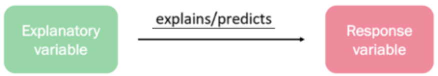
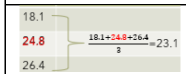
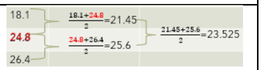
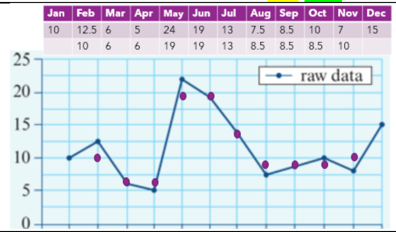
422 . 347

Round to 2 decimal places : 422 . **35**

422 . 344

Round to 2 decimal places : 422 . **34**

→ **Money must always be rounded to 2 decimal places or "to the nearest cent"**

Topics	Data Types 1A	Display/Analyse Tools	Report/Explain/Interpret/Describe
Univariate Data All in one 1.6	Categorical variables 1B: Nominal data; Ordinal data	Bar chart, Pie Chart, Frequency Table 1a, Segmented bar chart	Mode/ Modal Value Frequency types, Frequency % = $\frac{Count}{Total\ Count} \times 100\%$
	Numerical variables 1C: Discrete data; Continuous data	Boxplots 1c 1.5, Grouped Frequency Tables Stem plot 1b 1.6, dot plot 1b 1.6 Stem plot 1b 1.6, histogram 1b 1.1, log 1.2 log-histogram 1.3	Shape → Centre → Spread → Outliers → 1F **Symmetric:** Mean $\bar{x}$ 1.4 Standard Deviation S 1G 1.4 68-95-99.7% rule 1H 1.6 1I 1.7 Z-score=Z=$\frac{x-\bar{x}}{s}$ 1I 1.7 $x = \bar{x} + Z * S$ **Skewed:** 1.5 Median M or Q_2 IQR, Range 1.6 Lower Fence $= Q_1 - 1.5 * IQR$ 1.6 Upper Fence $= Q_3 + 1.5 * IQR$ 1.5 5-figure summary: Min, Q_1, Q_2, Q_3, Max IQR=$Q_3 - Q_1$, Range=Max−Min
Bivariate Data All in one 3.6	Two categorical variables 2A	Segmented bar chart, two-way frequency table 2a, parallel bar chart	Mode/ Modal Value Frequency types
	One categorical, one numerical variable 2B	Back-to-back stem plots, parallel dot plots, parallel box plots 2b	Shape → Centre → Spread → **Symmetric:** 1.4 Mean $\bar{x}$ Standard Deviation S **Skewed:** 1.5 Median M or Q_2 IQR, Range
	Two numerical variables 2C  Interpolation Extrapolation The assumptions for fitting a least squares line 1. the data is numerical 2. the association is linear 3. there are no clear outliers.	Scatterplot 2c 2.1  residual = actual data value y − predicted 3.7 value y residual = y - $\hat{y}$ Nil pattern residual 3.6 plot 2f = Linear relation Curved/ patterned residual plot ≠ linear relation	Strength → Direction → Form → 3A 3B 3C 2e LSRL y=a+bx 3.1 3.7 3.2 3.7 2e slope $b=\frac{r s_y}{s_x}$ 3.2 3.7 2e y-intercept $a = \bar{y} - b\bar{x}$ Strong/Moderate/Weak (Check r value) 2D 2.2 2.3 Positive / Negative Linear / Non-linear Reporting 2d on Coefficient of Determination r^2 3.1 Almost [r^2 in %] of [RV y] can be explained / predicted by [EV x].
Time Series 4A 4E: LSRL 4.1 4.3 4.4 4.5	Features 4A Trend Cycles Seasonality Structure change Outliers	**Moving smoothing 4.2** Moving Mean 4B 4.2 3/5 moving mean  2/4 Moving mean  Moving Median 4C 4.2 	**Seasonal Index S.I. 4D 4.6 4.7** S.I.=$\frac{Value\ for\ Season}{Yearly\ Average}$ Yearly Average=$\frac{Sum\ of\ Season\ Values}{No.of\ season\ per\ year}$ **Correct** S.I. = $(\frac{1}{S.I.} - 1) \times 100$ + means ↑, − means ↓ **Deseasonalising 4D 4.7** Deseasonalised Figure = $\frac{Actual\ Figure}{S.I.}$ $= Actual\ Figure * \frac{1}{S.I.}$ Actual figure= Deseasonalised Figure * S.I.

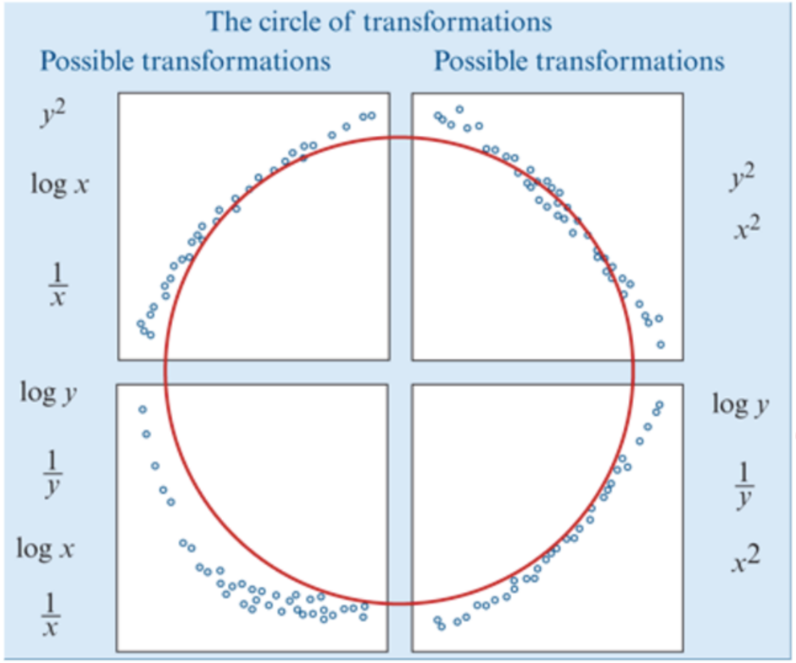
Note: Textbook Summary Notes Section #, Report Instruction Notes #. CAS Instruction Notes #

3D: Data Transformation & 3E 3.3 3.4 3.5 3.6

$Stretching\ transformation: Squared\ \&\ recipricol\ transformation$
$Compressing\ transformation:\ Logarithmic\ transformation$

$y^2 = a + bx$
$y = a + b\log x$
$y = a + \frac{b}{x}$

$\log y = a + bx$
$\frac{1}{y} = a + bx$
$y = a + b\log x$
$y = a + \frac{b}{x}$



$y^2 = a + bx$
$y = a + bx^2$

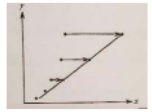
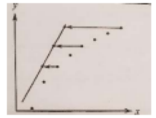
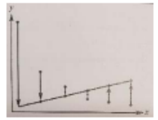
$\log y = a + bx$
$\frac{1}{y} = a + bx$
$y = a + bx^2$

- *Best transformation: strongest r/r^2 value*

Types of transformations:

- *Log: compresses the data*
- *Square: stretches the data*
- *Reciprocal: compresses values greater than 1, stretches values less than 1.*

The Effect of Each Transformation:

Type of Transformation:	Description of Effect:	One Word Description:	Graph of Transformation:
Squared Transformations (x^2 and y^2)	Spreads out the high x-values relative to the lower x-values and vice versa.	Stretching transformation ➢ x^2 stretches high x-values ➢ y^2 stretches high y-values	
Log Transformation (Log_x and Log_y)	Compresses the higher x-values relative to the lower x-values and vice versa	Compressing Transformation	
Reciprocal Transformations	Compresses larger y-values relative to smaller y-values and vice versa	Stretching and Compressing Transformation	

1D Log Scales & Graphs 1.2

Log (Base 10) Scale

Logarithms

A logarithm, or log, is a power or exponent or index of a number. That is the log of a^b is b.
For example the logs of $2^3, 5^4$, and 10^6 are 2, 3, and 6 respectively.

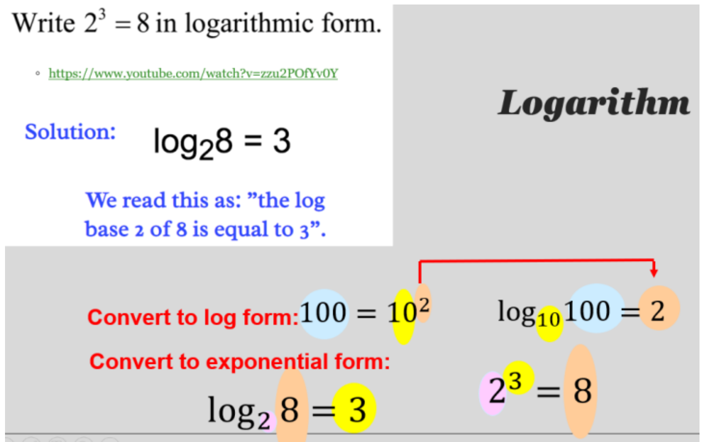
Log (Base 10) Scale

The log (base 10) scale is based of exponentials of base 10, i.e. $10, 10^2, 10^3, 10^4$.
Using the log (base 10) scale allows data ranging over several order of magnitude to be displayed.

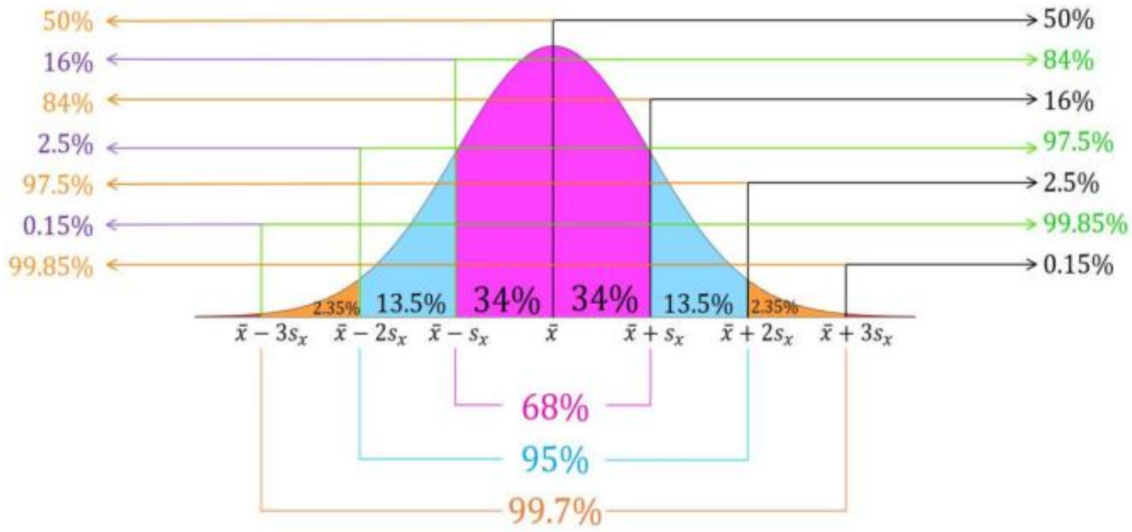
Converting Between Forms using the Log (Base 10) Scale

log value $= \log_{10}$(data value)   data value $= 10^{\text{log value}}$

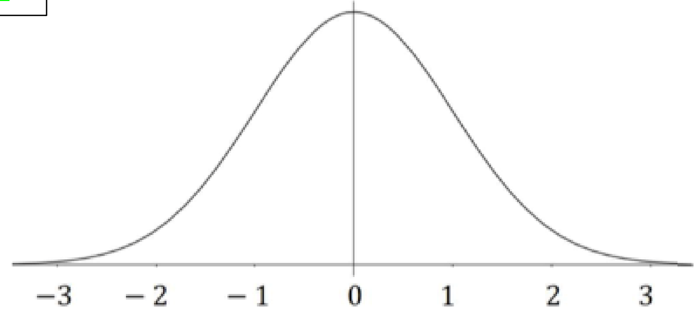
Data Value	0.001	0.01	0.1	10^n	1	10	100	1000
Log Form	$\log_{10} 0.001$	$\log_{10} 0.01$	$\log_{10} 0.1$	$\log_{10} 10^n$	$\log_{10} 1$	$\log_{10} 10$	$\log_{10} 100$	$\log_{10} 1000$
Log Value	-3	-2	-1	n	0	1	2	3
Exponent Form	10^{-3}	10^{-2}	10^{-1}	10^n	10^0	10^1	10^2	10^3

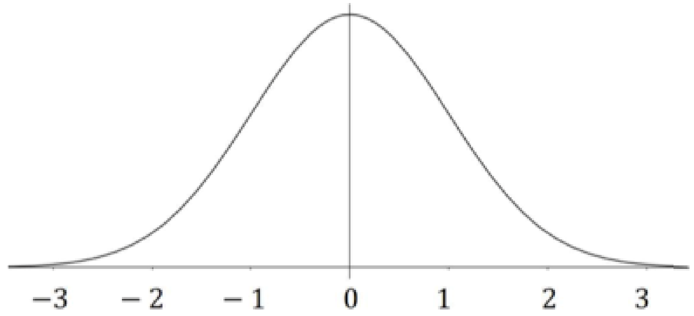


1H The Normal Distribution



1.6





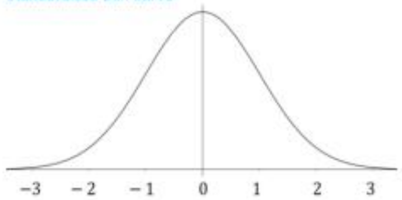
Converting to a Standard Score

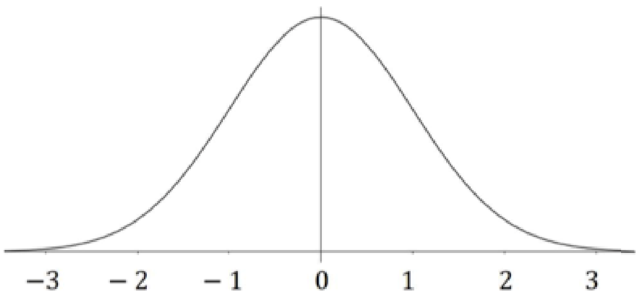
$$z = \frac{x - \bar{x}}{s}$$

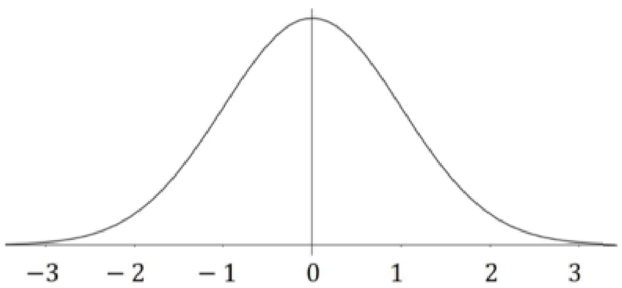
Converting to an Actual Score

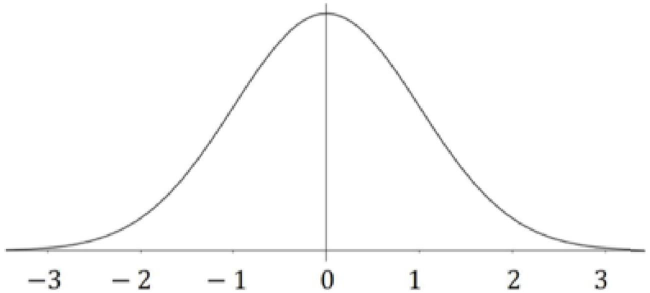
$$x = \bar{x} + z \times s$$

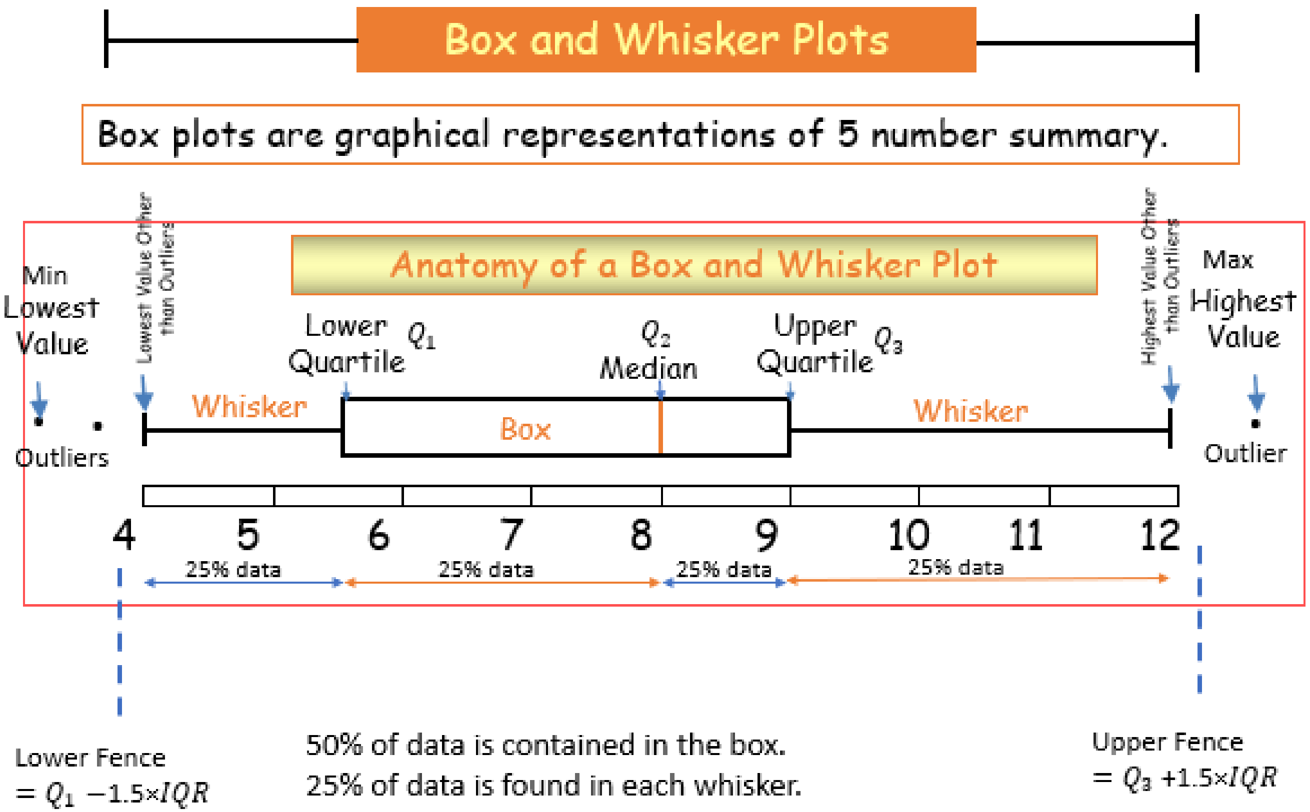
Standardised Bell Curve











$0.75 \le r \le 1$	Strong, positive, linear association
$0.5 \le r < 0.75$	Moderate, positive, linear association
$0.25 \le r < 0.5$	Weak, positive, linear association
$-0.25 < r < 0.25$	No association
$-0.5 < r \le -0.25$	Weak, negative, linear association
$-0.75 < r \le -0.5$	Moderate, negative, linear association
$-1 \le r \le -0.75$	Strong, negative, linear association

Data Summary Notes

1A: Types of data

Categorical: characteristics/qualities

- Nominal: grouped according to characteristics
- Ordinal: can be grouped and ordered

Numerical: numbers/quantities

- Discrete: whole numbers, can be counted
- Continuous: is measured

1B: displaying categorical data

Count frequency: number of times the category appears in the data

Percentage frequency: $\frac{count\ frequency}{total\ count} \times 100$

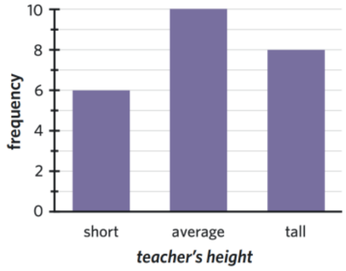
Mode: most frequently occurring value or category

Frequency Table:

teacher's height	frequency	
	number	**%**
short	6	25.0
average	10	41.7
tall	8	33.3
total	24	100.0

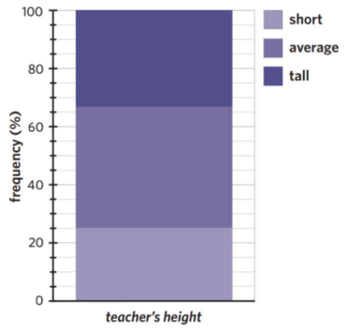
Bar chart:

- Must have gaps between bars



Segmented bar chart:

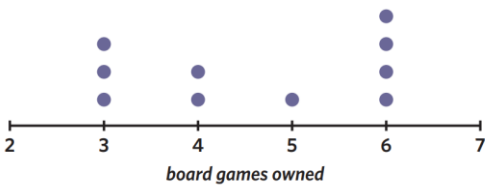
- Can be count or percentage frequency
- Must have a key



1C: Displaying Numerical data

Dot plot

- Discrete data
- Small data sets



Stem and leaf plot

- Needs a key
- Can have class intervals (splitting the stem in two if it is really large)

No intervals

Key: 1 | 2 = 1.2

1	3 3 4 6 8
2	0 4 9
3	1 1 1 4 5 8
4	2

With class intervals of 5

Key: 1 | 2 = 12

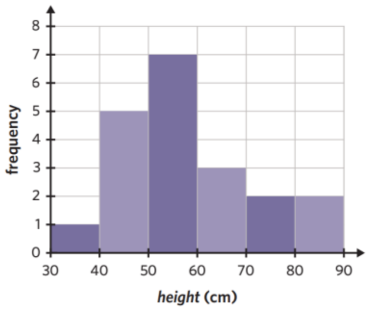
0	1 1 2 3 4
0	5 6 6 8 8 9
1	2 3 3
1	6 7 7 8 9 9

Grouped frequency tables

height (cm)	frequency	
	number	**%**
30–<40	1	5
40–<50	5	25
50–<60	7	35
60–<70	3	15
70–<80	2	10
80–<90	2	10
total	20	100

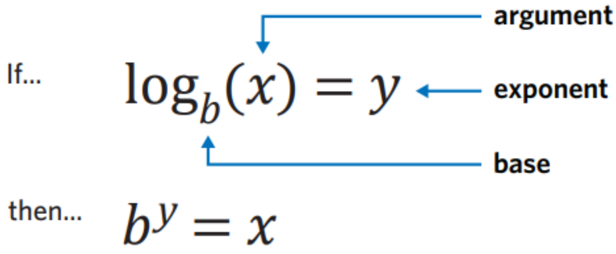
Histogram:

- Continuous data
- Intervals – no gaps between bars
- No gaps between bars
- X-axis markers are always a whole number

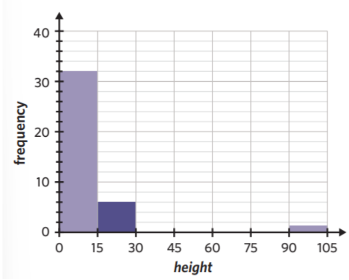


1D: Log scales and graphs

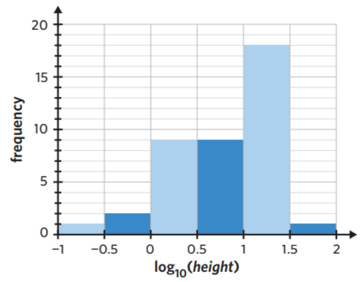
- Log scales are used to compress data that has a large range, making it more even and able to be displayed on the same set of axes.
- The base is always 10
- When undoing the log scale do ten to the power of the scale (eg. $10^{2.2} = 158.5$)



From…



To…

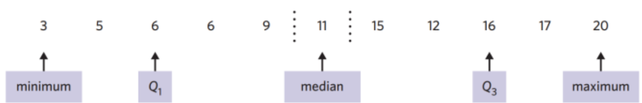


1E: the five-number summary and boxplots

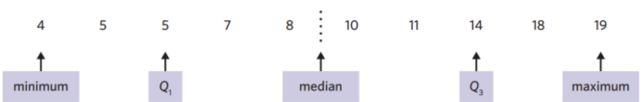
5 number summary:

- Minimum: smallest value in data set
- Q1: median of the lower half
- Median: middle value in an ordered data set
- Q3: median of the upper half
- Maximum: largest value in data set

Odd number of values:



Even number of values:



Spread: refers to how variable the data set it

Range = maximum – minimum

Interquartile range: measure of spread of the middle 50% of a data set. Accurate measure of spread when outliers are present.

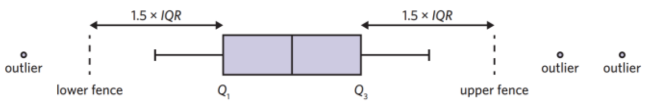
IQR = $Q_3 - Q_1$

Outliers: values which fall outside of what is 'normal'. Outliers are still the minimum and maximum value!

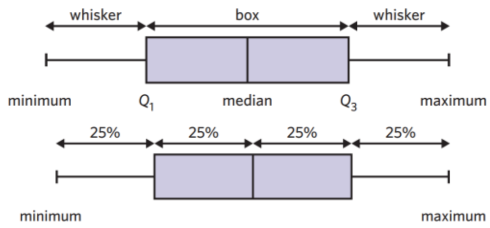
Fence: defines the boundary of what is an outlier. If a value is less than the lower fence or greater than the upper fence it is considered to be an outlier.

$$lower\ fence = Q_1 - (1.5 \times IQR)$$

$$upper\ fence = Q_3 + (1.5 \times IQR)$$

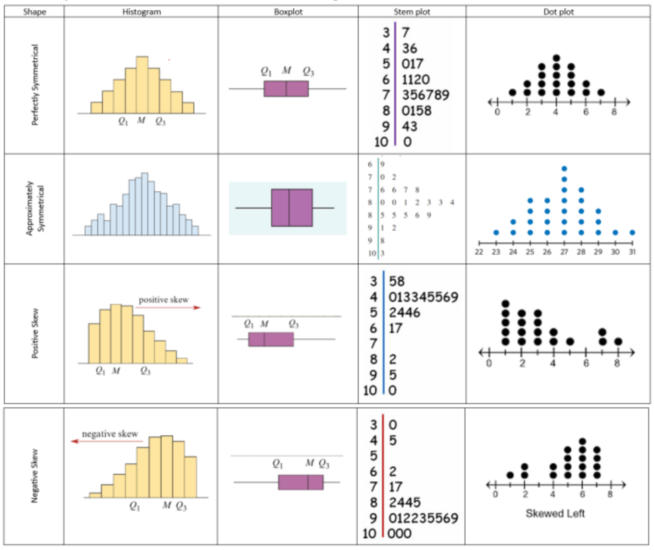


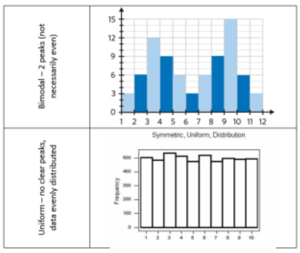
Boxplots:



1F: describing numerical data

- Shape: is the data symmetrical, skewed or have any outliers?
- Centre: What is the median value?
- Spread: What is the range and IQR?





1G: Standard deviation

Population: the entire group is used to collect data.
Sample: smaller subset of the population (this is usually what is used).

Mean: measure of centre – the AVERAGE. $\bar{x}$

- Calculated by adding all the data values together and then dividing by the number of values.

$\bar{x} = \frac{\Sigma x}{n}$, where Σx is the 'sum of all values', and n is the number of values in the data set.

Standard deviation: measure of spread based on the average deviation of each data point compared to the mean. It can be calculated by hand but please use CAS.

$$s_x = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}$$

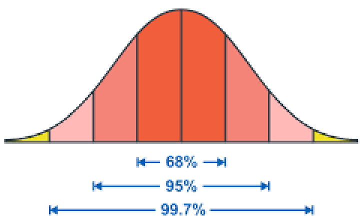
1H: The Normal Distribution

Normal Distribution: is a symmetrical (or approximately) numerical data set centred around the mean.

- Bell shaped
- Mean and median are equal

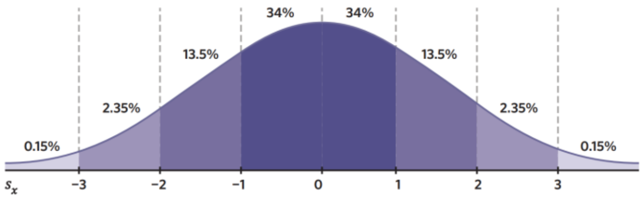
68-95-99.7% rule:

- 68% of the data lies within one standard deviation of the mean
- 95% of the data lies within two standard deviations of the mean
- 99.7% of the data lies within three standard deviations of the mean



The bell curve can be broken into each section:

- The mean lies in the centre (0)



1I: z-scores

Standardised score:

- Z-score
- Measure of the number of standard deviations between the mean and a data value
- Each data value is an 'actual score'
- Positive = above mean, negative = below mean, zero = equal to mean

$$z = \frac{x - \bar{x}}{s_x}$$

- z is the standardised score
- x is the actual score
- $\bar{x}$ is the mean
- s_x is the standard deviation

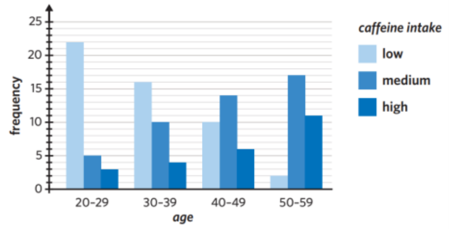
Actual score: $x = \bar{x} + (z \times s_x)$

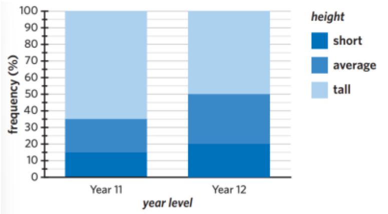
2A: association between 2 variables

Two-way frequency table:

- Columns = EV, Rows = RV
- Percentage frequency is used for greater accuracy when making comparisons if sample sizes are different

Grouped and segmented bar charts:





Describing the association between two variables:

- Whether or not an association between the two variables exists
- Appropriate percentages to support findings

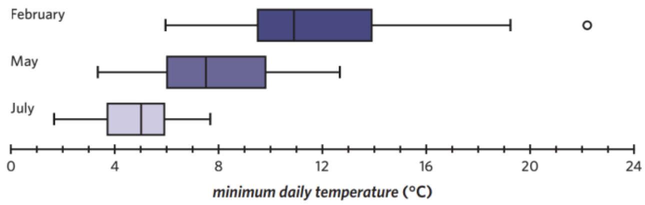
2B: association between numerical and categorical variables

- Back to back stem plot

Key: 3 | 7 = 37 points

Crocodiles		Zebras
4 0	2	
8 7 5 2	3	7
7 4 4	4	
8 2	5	2 8
3 1	6	0 1 5 9
	7	0 3 5 8
0	8	1 3 4

- Parallel boxplot



- Making comparisons: refer to 1F and compare shape, centre and spread of the two categories

2C: association between two numerical variables

Response variable: RV, may be explained or predicted by changes in the explanatory variable.

Explanatory variable: EV, used to explain or predict the changes observed in the response variable.

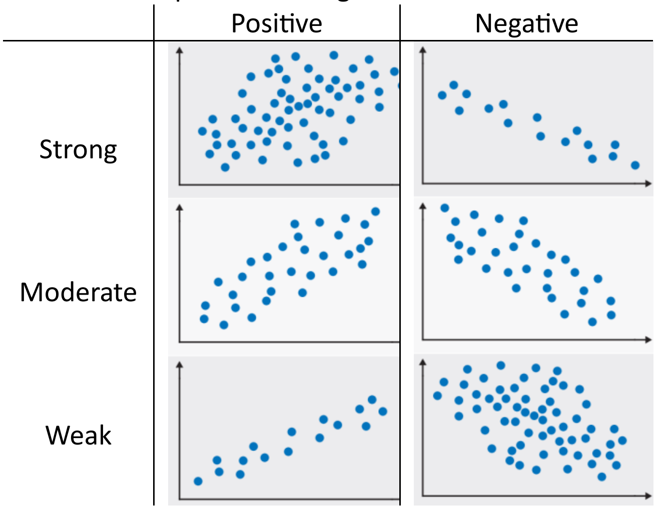
- 'EV explains the RV'

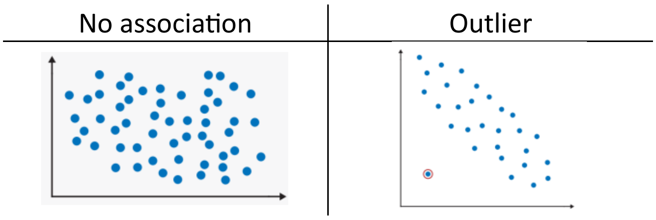
Scatterplots:

- EV = x axis, RV = y axis

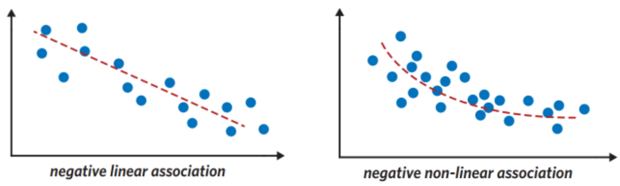
Describing relationship/analysing scatterplots:

- Strength: how close the data points are together
- Direction: positive or negative





- Form: linear (straight) or non-linear (curved)



2D: Correlation and causation

Pearson's correlation coefficient (r): numerical value that determines strength and direction between two numerical variables, assuming:

- Data is linear
- Data is numeric
- No outliers present

$0.75 \le r \le 1$	Strong, positive, linear association
$0.5 \le r < 0.75$	Moderate, positive, linear association
$0.25 \le r < 0.5$	Weak, positive, linear association
$-0.25 < r < 0.25$	No association
$-0.5 < r \le -0.25$	Weak, negative, linear association
$-0.75 < r \le -0.5$	Moderate, negative, linear association
$-1 \le r \le -0.75$	Strong, negative, linear association

Correlation and Causation: just because two variables have a high correlation, it doesn't mean that one causes the change in the other. Some explanations:

- **Common response:** a third variable that is the likely cause of correlation, acting on both variables. Eg. Number of people wearing sunscreen and feinting → the sunscreen isn't causing people to feint… the third variable would be temperature. This is common cause as temperature affects **both** variables.
- **Confounding variable:** external variable that can also produce a change to the RV. Eg. Plant height and water intake. Water intake does effect plant height (RV) but so does sun, soil quality, buys, season, temperature…
- **Coincidence:** two variables correlate but have no relation to each other. Pure chance. No logical explanation.

3A: fitting a least squares regression line

Least squares regression line (LSRL): is the line which creates the minimum sum of the squares of residuals. There are assumptions:

- Data is numerical
- The relationship between variables is linear
- There are no clear outliers present

The line is used to show the general trend in the data and is given by the equation:

$$y = a + bx$$

Intercept (a) Slope (b)

Determining LSRL from a graph: Find the intercept (a) and the slope (b).

- Intercept: read directly from the graph when the EV is 0
- Slope: choose two points on the line that you can clearly ready the coordinates. Use the rule:

$$b = \frac{rise}{run} = \frac{y_2 - y_1}{x_2 - x_1}$$

Calculating the LSRL from summary statistics:

$$b = r \times \frac{s_y}{s_x}$$

$$a = \bar{y} - b\bar{x}$$

- a is the y-intercept
- b is the gradient
- r is Pearson's correlation coefficient
- $\bar{x}$ is the mean of the explanatory variable (x)
- $\bar{y}$ is the mean of the response variable (y)
- s_x is the standard deviation of the explanatory variable (x)
- s_y is the standard deviation of the response variable (y)

Drawing the LSRL on a graph: Sub in the first value on the x-axis and the last value on the x-axis into the equation. Plot the two points, join the line using a ruler.

3B: Interpreting LSRL: use the following statements, fill in EV and RV and values of a and b.

y-intercept: when the **EV** is 0, the **RV** is **a**.

Slope: for every one-unit increase in the **EV**, the **RV increases/decreases** by **b**. (If b is positive, increases, if b is negative, decreases)

Making predictions: the LSRL can be used to predict the value of the RV from the EV.
Interpolation: predicting within the range of data.
Extrapolation: predicting outside the range of data. Less reliable.
How to predict: sub the EV value that you are predicting for into the LSRL equation to predict the RV.

3C: Performing a regression analysis:

Coefficient of determination (r^2): calculated by squaring the r value. It is turned into a percentage (×100) then interpreted. Use the statement by inputting the variable names and percentages:

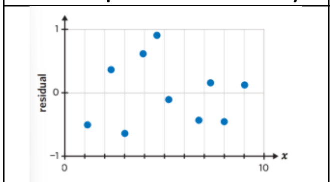
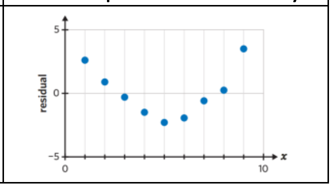
- **r^2 %** of the variation in the **RV** can be explained by the variation in the **EV**. The remaining **%** can be explained by other factors.

Residuals: residuals are the vertical distances between the data point and the LSRL.

$$residual = actual\ data\ value - predicted\ data\ value$$

- Actual value: found in the question/table of data
- Predicted value: must use the LSRL to predict the RV from the EV
- Positive residual = data point above LSRL, negative residual = data point below LSRL, zero residual = data point on the LSRL.

Residual plots:

Random scattering of points in the residual plot, supports the assumption of linearity	Clear curved pattern in the residual plot, does not support the assumption of linearity
	

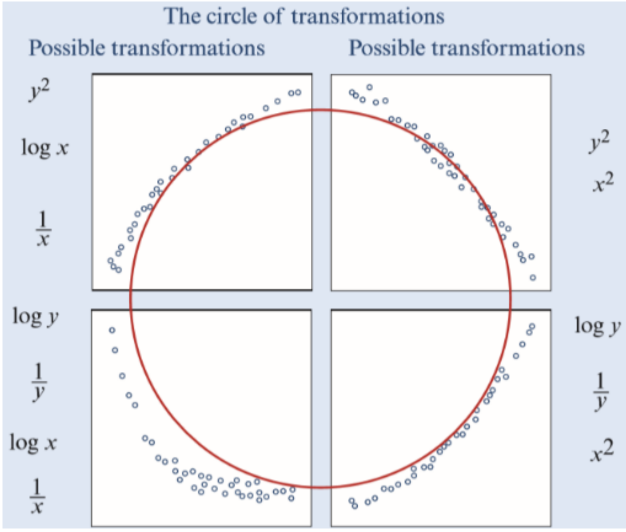
3D: Data Transformations:

You shouldn't perform a linear regression analysis for data that is nonlinear. Therefore, nonlinear data is **transformed.**

- Transformation linearise data so that regression analysis can be performed accurately.
- Match the nonlinear scatterplot with one in the diagram to help you determine the best transformation.
- **Best transformation:** strongest r/r^2 value

Types of transformations:

- Log: compresses the data
- Square: stretches the data
- Reciprocal: compresses values greater than 1, stretches values less than 1.



3E: Data transformations – applications

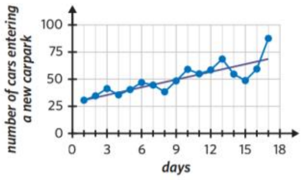
LSRL: once you have transformed your data you must create a new LSRL equation and include the transformation in the rules.
Eg. From
$y = -16.14 + 9.39x$
To
$y = -0.73 + 1.05x^2$

Making predictions: the limits of extrapolation are still present. When calculating, use solve as this will undo the transformation for you.

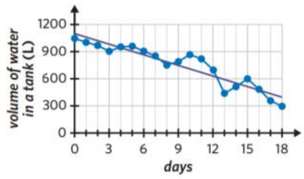
4A: Time series data and their graphs

Trends: general upwards (increasing) or downwards (decreasing) movement over time. Trend lines can be fitted directly to trends. There can be multiple trend lines.

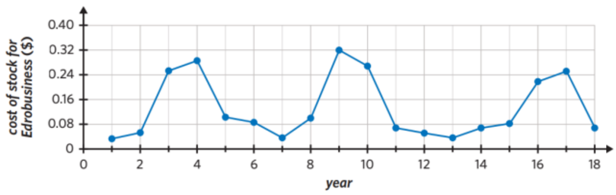
Increasing trend



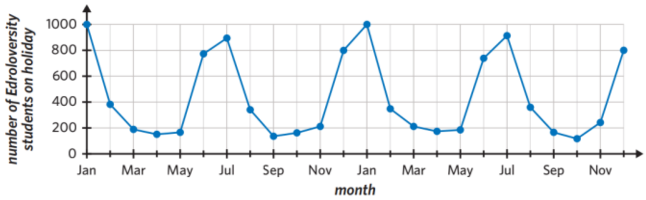
Decreasing trend



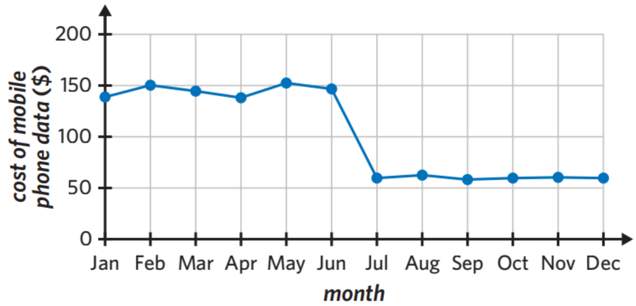
Cycles/cyclical variation: periodic movements over a period greater than 1 year. Peaks of cycles occur at approximately the same intervals, cycles can have a period which changes slightly between peaks.



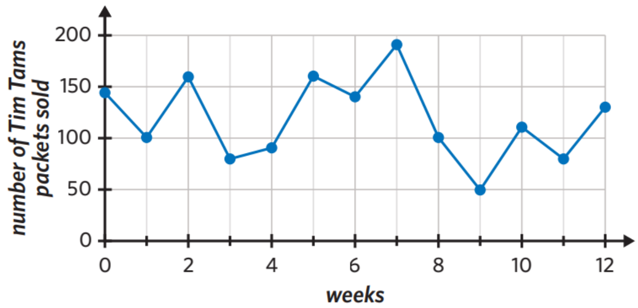
Seasonality: cyclical variation within a calendar-related period (week, month, quarter). A seasonal time series plot has regular peaks and troughs that occur at the same time each period and the length of the period must be a year or less.



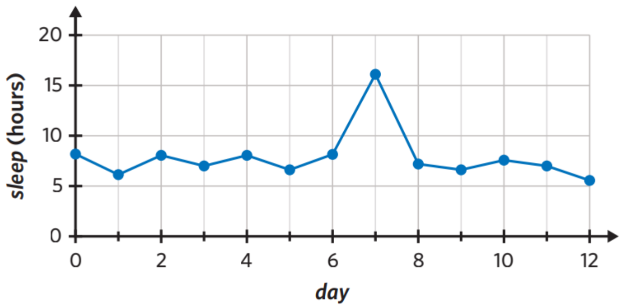
Structural Change: When an established pattern is suddenly altered. The graph then continues on the same level post structural change.



Irregular fluctuations: random variations that cannot be explained by trend, seasonality, cycles or structural change.



Outliers: stands out from the general body of data. It then returns to follow the original pattern/trend



4B: Smoothing – moving means

Smoothing: evens out fluctuations to help identify any underlying trends

- Only smooth the RV
- The larger the mean smooth, the more effective (5 more effective than 3)

3 mean: use 3 values and find the mean

5 mean: use 5 values and find the mean

- Always centred around the value you are trying to smooth

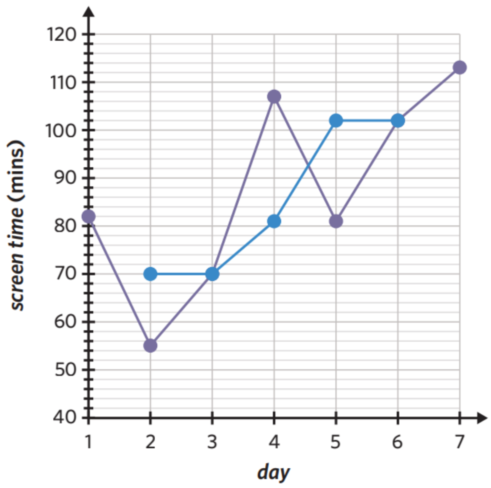
day	*temp.* (°C)	*calculation*	three-mean smoothed *temperature* (°C)
Mon	24	–	–
Tue	27	$\frac{24+27+21}{3}$	24
Wed	21	$\frac{27+21+18}{3}$	22
Thu	18	$\frac{21+18+15}{3}$	18
Fri	15	$\frac{18+15+15}{3}$	16
Sat	15	$\frac{15+15+12}{3}$	14
Sun	12	–	–

Smoothing with centring: an additional step when smoothing with an even number of points. Finding the mean of two non-centred means.

day	*temp.* (°C)	*before centring*	*after centring*
Mon	24		–
		$\frac{24+27}{2}=25.5$	
Tue	27		$\frac{25.5+24}{2}=24.75$
		$\frac{27+21}{2}=24$	
Wed	21		$\frac{24+19.5}{2}=21.75$
		$\frac{21+18}{2}=19.5$	
Thu	18		–

4C Smoothing – moving medians

- Smoothed directly on the graph
- Median smoothing only uses and odd number of points
- Smooth the RV



4D: Seasonal adjustments:

- Seasonal fluctuations exist.
- Seasonal indices (SI) are used to de-seasonalise the data to minimise the effects of seasonality. This allows trends to be more easily observed.

Rules

1. $Seasonal\ index\ (SI) = \frac{value\ for\ season}{seasonal\ average}$

2. $Seasonal\ average\ (SA) = \frac{sum\ of\ all\ seasons}{number\ of\ season}$

3. $Deseasonalised\ figure\ (DS) = \frac{value\ for\ season}{seasonal\ index}$

4. Reseasonalising data:

$$value\ for\ season = deseasonalised\ figure \times seasonal\ index$$

How to interpret a seasonal index:

$(seasonal\ index - 1) \times 100 =$ _______%

- A negative %: (season) is below the seasonal average by ___%
- A positive %: (season) is above the seasonal average by ___%

Correcting for seasonality:

$(\frac{1}{seasonal\ index} - 1) \times 100 =$ _______%

- A negative %: To correct (season) for seasonality, (unit) need to be decreased by ____%
- A positive %: To correct (season) for seasonality, (unit) need to be increased by ____%

Notes:

- The sum of the seasonal indices is equal to the number of seasons (if you are working with months of the year there are 12 seasons and therefore the seasonal indices will sum to 12)
- If there were no fluctuations, the seasonal average is 1

4E: Time series data and LSRL modelling:

Trend lines: can be fitted to time series plots if there appears to be an increasing or decreasing trend.

- The LSRL is used
- If seasonality is present, data needs to be deseasonalised first before fitting the LSRL to the deseasonalised values

Forecasting: making a prediction for the future

- You need to re-seasonalise the value if the prediction was made from a deseasonalised LSRL

1.1 Constructing a histogram from raw data

CAS 1: How to construct a histogram using the TI-Nspire CAS

Display the following set of 27 marks in the form of a histogram.

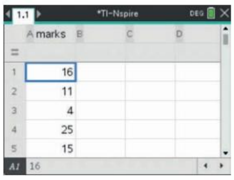
16 11 4 25 15 7 14 13 14 12 15 13 16 14
15 12 18 22 17 18 23 15 13 17 18 22 23

Steps

1 Start a new document by pressing ctrl + N (or ⌂on>**New**. If prompted to save an existing document, move the cursor to **No** and press enter.

2 Select **Add Lists & Spreadsheet**.
Enter the data into a list named *marks*.

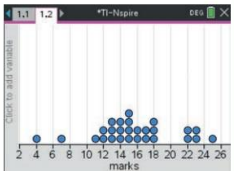
a Move the cursor to the name cell of column A and type in *marks* as the list variable.
Press enter.



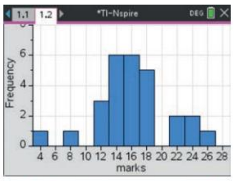
b Move the cursor down to row 1, type in the first data value and press enter. Continue until all the data have been entered. Press enter after each entry.

3 Statistical graphing is done through the **Data & Statistics** application. Press ctrl + I (or alternatively press ctrl doc▾) and select **Add Data & Statistics**.

a Press tab enter (or click on the **Click to add variable box** on the x-axis) to show the list of variables. Select *marks*.
Press enter to paste *marks* to that axis.



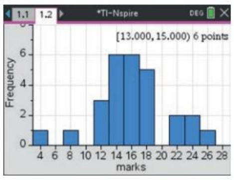
b A dot plot is displayed as the default. To change the plot to a histogram, press menu>**Plot Type> Histogram**. The histogram shown opposite has a column (or bin) width of 2, and a starting point (alignment) of 3. See Step 5 below for instructions on how to change the appearance of a histogram.



4 Data analysis

a Move the cursor over any column; a ✋ will appear and the column data will be displayed as shown opposite.

b To view other column data values, move the cursor to another column.



Note: If you click on a column, it will be selected.

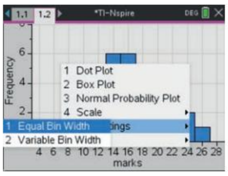
Hint: If you accidentally move a column or data point, ctrl + esc enter will undo the move.

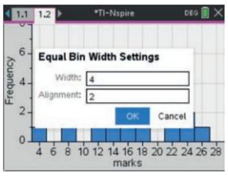
5 Change the histogram column (bin) width to **4** and the starting point to **2**.

a Press ctrl + menu to access the context menu as shown (below left).
Hint: Pressing ctrl + menu enter with the cursor on the histogram gives you a context menu that relates only to histograms. You can access the commands through menu>**Plot Properties**.

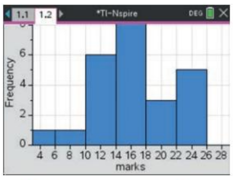
b Select **Bin Settings**>**Equal Bin Width**.

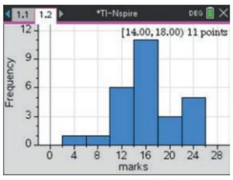
c In the settings menu (below right) change the Width to 4 and the Starting Point (Alignment) to 2 as shown. Press enter.



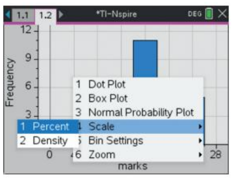


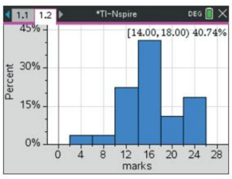
d A new histogram is displayed with column width of 4 and a starting point of 2 but it no longer fits the window (below left). To solve this problem, press ctrl + menu>**Zoom>Zoom-Data** and enter to obtain the histogram as shown below right.





6 To change the frequency axis to a percentage axis, press ctrl + menu>**Scale>Percent** and then press enter.





1.2 Working with logarithms

Example 12 Using a CAS calculator to find logs

a Find the log of 45, correct to two significant figures.

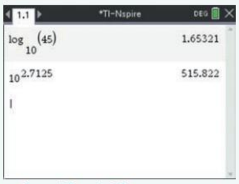
b Find the number with log equal to 2.7125, correct to the nearest whole number.

Explanation

a Open a calculator screen, type log (45) and press enter. Write down the answer correct to two significant figures.

b If the log of a number is 2.7125, then the number is $10^{2.7125}$.
Enter the expression $10^{2.7125}$ and press enter. Write down the answer correct to the nearest whole number.

Solution



a $\log 45 = 1.65\ldots$
$= 1.7$ (to 2 sig. figs)

b $10^{2.7125} = 515.82\ldots$
$= 516$ (to the nearest whole number)

1.3 Constructing a histogram with a log scale

CAS 2: Using a TI-Nspire CAS to construct a histogram with a log scale

The weights of 27 animal species (in kg) are recorded below.

1.4	470	36	28	1.0	12 000	2600	190	520
10	3.3	530	210	62	6700	9400	6.8	35
0.12	0.023	2.5	56	100	52	87 000	0.12	190

Construct a histogram to display the distribution:

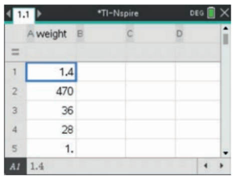
a of the body weights of these 27 animals and describe its shape

b of the log of the body weights of these animals and describe its shape.

Steps

1 a Start a new document by pressing ctrl + N.

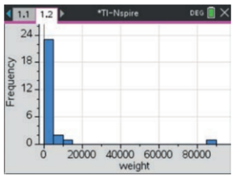
b Select **Add Lists & Spreadsheet**.
Enter the data into a column named *weight*.



2 a Press ctrl + I and select **Add Data & Statistics**.
Click on the **Click to add variable** on the x-axis and select the variable *weight*. A dot plot is displayed.

b Plot a histogram using menu>**Plot Type>Histogram**.

c Describe the shape of the distribution.

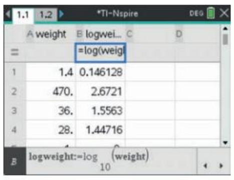


Shape: positively skewed with outliers

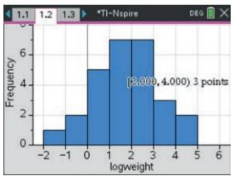
3 a Return to the **Lists & Spreadsheet** screen.

b Name another list *logweight*.

c Move the cursor to the formula cell below the *logweight* heading. Type in = **log(*weight*)**.
Press enter to calculate the values of *logweight*.



4 a Plot a histogram using a log scale. That is, plot the variable *logweight*.
Note: Use menu>**Plot Properties**>**Histogram Properties**>**Bin Settings**>**Equal Bin Width** and set the column width (bin) to 1 and alignment (start point) to −2 and use menu>**Window/Zoom**>**Zoom-Data** to rescale.



b Describe the shape of the distribution.

Shape: approximately symmetric

1.4 Calculating the standard deviation

CAS 3: How to calculate the mean and standard deviation using the TI-Nspire CAS

The following are the heights (in cm) of a group of women.

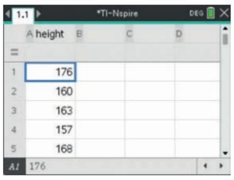
176 160 163 157 168 172 173 169

Determine the mean and standard deviation of the women's heights. Give your answers correct to two decimal places.

Steps

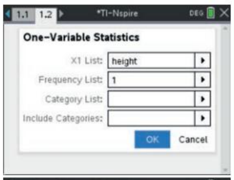
1 Start a new document by pressing ctrl + N.

2 Select **Add Lists & Spreadsheet**.
Enter the data into a list named *height*, as shown.



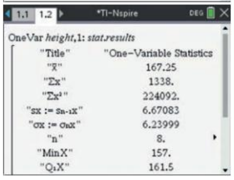
3 Statistical calculations can be done in either the **Lists & Spreadsheet** application or the **Calculator** application (used here).
Press ctrl + I and select **Add Calculator**.

a Press menu>**Statistics>Stat Calculations>One-Variable Statistics**. Press enter to accept the **Num of Lists** as 1.



b **i** To complete this screen, use the ▸ arrow and enter to paste in the list name *height*.

ii Pressing enter exits this screen and generates the results screen shown opposite.



4 Write down the answers to the required degree of accuracy (i.e. two decimal places).

The mean height of the women is $\bar{x}$ = 167.25 cm and the standard deviation is s = 6.67 cm.

Notes: **a** The sample standard deviation is **sx**.
b Use the ▲ ▼ arrows to scroll through the results screen to obtain values for additional statistical values.

1.5 Constructing a boxplot with outliers

CAS 4: How to construct a boxplot with outliers using the TI-Nspire CAS

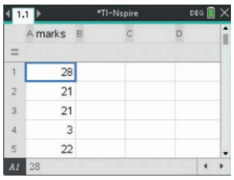
Display the following set of 19 marks in the form of a boxplot with outliers.

28 21 21 3 22 31 35 26 27 33
43 31 30 34 48 36 35 23 24

Steps

1 Start a new document by pressing ctrl+N.

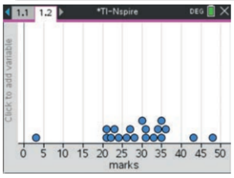
2 Select **Add Lists & Spreadsheet**. Enter the data into a list called ***marks*** as shown.



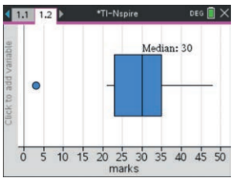
3 Statistical graphing is done through the **Data & Statistics** application.
Press ctrl + I and select **Add Data & Statistics**.

Note: A random display of dots will appear – this indicates that list data are available for plotting. Such a dot is not a statistical plot.

a Click on the **Click to add variable** on the *x*-axis and select the variable ***marks***. A dot plot is displayed by default as shown opposite.



b To change the plot to a boxplot press menu>**Plot Type**>**boxplot**. Your screen should now look like that shown opposite.



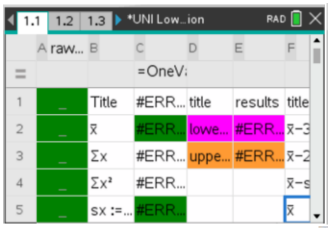
4 Data analysis
Key values can be read from the boxplot by moving the cursor over the plot or using menu>**Analyze>Graph Trace**.
Starting at the far left of the plot, we see that the:

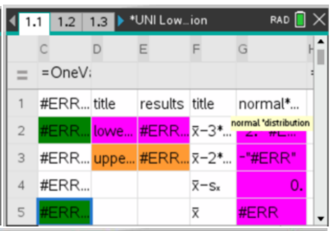
- minimum value is 3 (an outlier)
- first quartile is 23 (**Q_1 = 23**)
- median is 30 (**Median = 30**)
- third quartile is 35 (**Q_3 = 35**)
- maximum value is 48.

1.6 Using "UNI Lower Upper Fences Normal Distribution" template

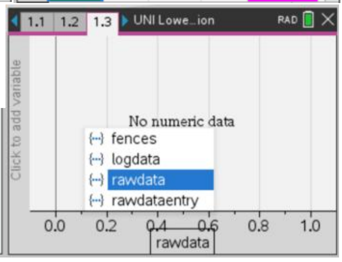
Fences Either ENTER RawData into column a of 1.1
Or
Enter Q1 into C9
Enter Q3 into C11
ENTER

Normal Distribution Either ENTER RawData into column a of 1.1
Or
Enter $\bar{x}$ into C2
Enter s_x into C5
ENTER





In following files: Click "Click to add variable" on the x–axis and select the variable "rawdata" to have following graphs. Log Histogram need to select "logdata".
1.3 BoxPlot: 1.4 DotPlot; 1.5 Histogram; 1.6 Log Histogram 1.1 Column i for Stem Plot



1.7 Using "Z SCORE" template

Enter all known data into green boxes as indicated on the column a of data title, results will come out in the coloured area

2.1 Constructing a scatterplot

CAS 1: How to construct a scatterplot using the TI-Nspire CAS

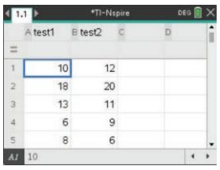
Construct a scatterplot for the set of test scores given below.
Treat *test 1* as the explanatory (i.e. *x*) variable.

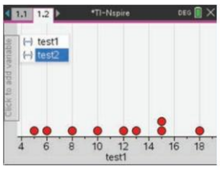
Test 1	10	18	13	6	8	5	12	15	15
Test 2	12	20	11	9	6	6	12	13	17

Steps

1 Start a new document by pressing ctrl + N.

2 Select **Add Lists & Spreadsheet**. Enter the data into lists named *test1* and *test2*.

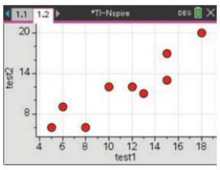




3 Press ctrl + I and select **Add Data & Statistics**.

4 **a** Click on **Click to add variable** on the *x*-axis and select the explanatory variable *test1*.

b Click on **Click to add variable** on the *y*-axis and select the response variable *test2*. A scatterplot is displayed. The plot is scaled automatically.



2.2 Calculating the correlation coefficient r using the formula

How to calculate the correlation coefficient using the formula

Use the formula to calculate the correlation coefficient, r, for the following data.

x	1	3	5	4	7
y	2	5	7	2	9

$\bar{x} = 4,\ s_x = 2.236$
$\bar{y} = 5,\ s_y = 3.082$

Give the values rounded to two decimal places.

Steps

1 Write down the values of the means, standard deviations and n.

$\bar{x} = 4 \quad s_x = 2.236$
$\bar{y} = 5 \quad s_y = 3.082 \quad n = 5$

2 Set up a table like that shown opposite to calculate $\sum(x-\bar{x})(y-\bar{y})$.

x	$(x-\bar{x})$	y	$(y-\bar{y})$	$(x-\bar{x})\times(y-\bar{y})$
1	−3	2	−3	9
3	−1	5	0	0
5	1	7	2	2
4	0	2	−3	0
7	3	9	4	12
Sum	0		0	23

$\therefore \sum(x-\bar{x})(y-\bar{y}) = 23$

3 Write down the formula for r. Substitute the appropriate values and evaluate, rounding the answer to two decimal places.

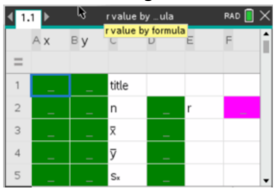
$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{(n-1)s_x s_y}$$

$$\therefore r = \frac{23}{(5-1)\times 2.236 \times 3.082}$$

$$= 0.834... = 0.83 \text{ (2 d.p.)}$$

2.3 Using "r value by formula" template to calculate the correlation coefficient r using the formula

Enter all data in green area and answer will come in pink box



3.1 Determining the equation of the least squares regression Line

CAS 1: How to determine and graph the equation of a least squares regression line using the TI-Nspire CAS

The following data give the height (in cm) and weight (in kg) of 11 people.

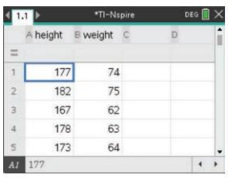
Height (x)	177	182	167	178	173	184	162	169	164	170	180
Weight (y)	74	75	62	63	64	74	57	55	56	68	72

Determine and graph the equation of the least squares regression line that will enable weight to be predicted from height. Write the intercept and slope rounded to three significant figures.

Steps

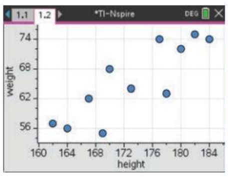
1 Start a new document by pressing ctrl + N.

2 Select **Add Lists & Spreadsheet**. Enter the data into lists named *height* and *weight*, as shown.



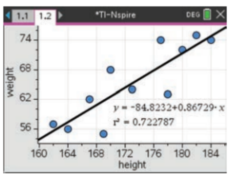
3 Identify the explanatory variable (EV) and the response variable (RV).
EV: *height*
RV: *weight*
Note: In saying that we want to predict *weight* from *height*, we are implying that *height* is the EV.

4 Press ctrl + I and select **Add Data & Statistics**. Construct a scatterplot with *height* (EV) on the horizontal (or x-) axis and *weight* (RV) on the vertical (or y-) axis.
Press menu>**Settings** and click the **Diagnostics** box. Select **OK** to activate this feature for *all* future documents. This will show the coefficient of determination (r^2) whenever a regression is performed.



5 Press menu>**Analyze**>**Regression**>**Show Linear (a + bx)** to plot the regression line on the scatterplot. Note that, simultaneously, the equation of the regression line is shown on the screen.
The equation of the regression line is:

$$weight = -84.8 + 0.867 \times height$$



The coefficient of determination is $r^2 = 0.723$, rounded to three significant figures.

3.2 Conducting a regression analysis using data

CAS 2: How to conduct a regression analysis using the TI-Nspire CAS

This analysis is concerned with investigating the association between life expectancy (in years) and birth rate (in births per 1000 people) in 10 countries.

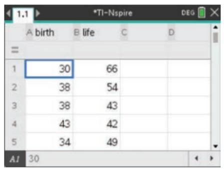
Birth rate	30	38	38	43	34	42	31	32	26	34
Life expectancy (years)	66	54	43	42	49	45	64	61	61	66

Steps

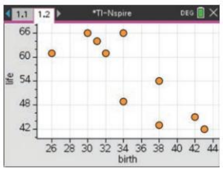
1 Write down the explanatory variable (EV) and response variable (RV). Use the variable names *birth* and *life*.

EV: *birth*
RV: *life*

2 Start a new document by pressing ctrl + N.
Select **Add Lists & Spreadsheet**.
Enter the data into the lists named *birth* and *life*, as shown.



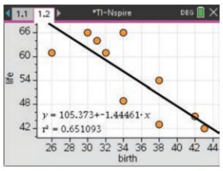
3 Construct a scatterplot to investigate the nature of the relationship between life expectancy and birth rate.



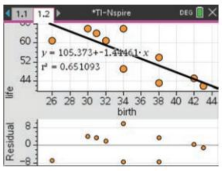
4 Describe the association shown by the scatterplot. Mention direction, form, strength and outliers.

There is a strong, negative, linear relationship between life expectancy and birth rate. There are no obvious outliers.

5 Find and plot the equation of the least squares regression line and r^2 value.
Note: Check if **Diagnostics** is activated using menu>**Settings**.



6 Generate a residual plot to test the linearity assumption.
Use ctrl + ◄ (or click on the page tab) to return to the scatterplot.
To hide the residual plot press menu>**Analyze>Residuals>Hide Residual Plot**.



7 Use the values of the intercept and slope to write the equation of the least squares regression line. Also write the values of r and the coefficient of determination.

Regression equation:
$life = 105.4 - 1.445 \times birth$
Correlation coefficient: $r = -0.8069$
Coefficient of determination: $r^2 = 0.651$

3.3 Perform a squared transformation

CAS 1: Using the TI-Nspire CAS to perform a squared transformation

The table shows the height (in m) of a base jumper for the first 10 seconds of her jump.

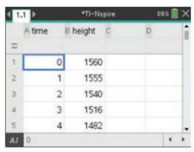
Time	0	1	2	3	4	5	6	7	8	9	10
Height	1560	1555	1540	1516	1482	1438	1383	1320	1246	1163	1070

a Construct a scatterplot displaying *height* (the RV) against *time* (the EV).
b Apply an x-squared transformation and fit a least squares line to the transformed data.
c Use the regression line to predict the height of the base jumper after 3.4 seconds.

Steps

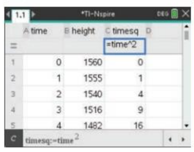
1 Start a new document by pressing ctrl + N.

2 Select **Add Lists & Spreadsheet**.
Enter the data into lists named *time* and *height*, as shown.

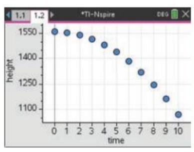


3 Name column C as *timesq* (short for 'time squared').

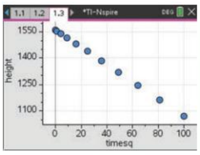
4 Move the cursor to the formula cell below *timesq*.
Enter the expression = *time*^2 by pressing =, then typing **time∧2**. Pressing enter calculates and displays the values of *timesq*.



5 Press ctrl + I and select **Add Data & Statistics**.
Construct a scatterplot of *height* against *time*. Let *time* be the explanatory variable and *height* the response variable. The plot is clearly non-linear.

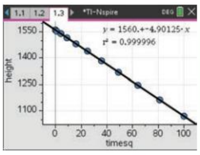


6 Press ctrl + I and select **Add Data & Statistics**. Construct a scatterplot of *height* against *time*^{2}. The plot is now linear.



7 Press menu>**Analyze>Regression>Show Linear (a + bx)** to plot the line on the scatterplot with its equation.

Note: The x in the equation on the screen corresponds to the transformed variable *time*^{2}.



8 Write down the regression equation in terms of the variables *height* and *time*^{2}.

height = 1560 – 4.90 × *time*^{2}

9 Substitute 3.4 for *time* in the equation to find the height after 3.4 seconds.

height = 1560 – 4.90 × 3.4^2 = 1503 m

3.4 Applying the log transformation

CAS 2: Using the TI-Nspire CAS to perform a log transformation

The table shows the *lifespan* (in years) and *GDP* (in dollars) of people in 12 countries. The association is non-linear.

Using the log x transformation:

- linearise the data, and fit a regression line to the transformed data (*GDP* is the EV)
- write its equation in terms of the variables *lifespan* and *GDP* rounded to three significant figures.
- use the equation of the regression line to predict the lifespan in a country with a GDP of $20 000, rounded to one decimal place.

Lifespan	*GDP*
80.4	36 032
79.8	34 484
79.2	26 664
77.4	41 890
78.8	26 893
81.5	25 592
74.9	7 454
72.0	1 713
77.9	7 073
70.3	1 192
73.0	631
68.6	1 302

Steps

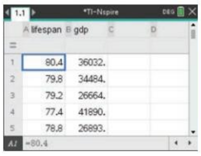
1 Start a new document by pressing ctrl + N .

2 Select **Add Lists & Spreadsheet**.

Enter the data into lists named *lifespan* and *gdp*.

3 Name column C as *lgdp* (short for log (*GDP*)).

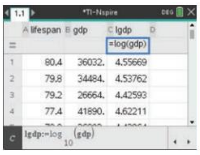
Now calculate the values of log (*GDP*) and store them in the list named *lgdp*.



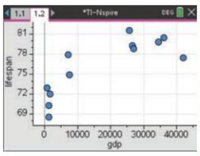
4 Move the cursor to the formula cell below the *lgdp* heading.

We need to enter the expression = **log(gdp)**.

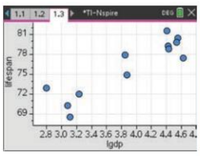
To do this, press = then type in **log(gdp)**. Pressing enter calculates and displays the values of *lgdp*.



5 Press ctrl + I and select **Add Data & Statistics**. Construct a scatterplot of *lifespan* against *GDP*. Let *GDP* be the explanatory variable and *lifespan* the response variable. The plot is clearly non-linear.

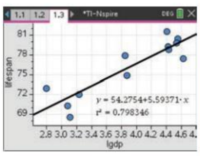


6 Press ctrl + I and select **Add Data & Statistics**. Construct a scatterplot of *lifespan* against log(*GDP*). The plot is now clearly linear.



7 Press menu>**Analyze>Regression>Show Linear (a + bx)** to plot the line on the scatterplot with its equation.

Note: The x in the equation on the screen corresponds to the transformed variable log (*GDP*).



8 Write the regression equation in terms of the variables *lifespan* and log (*GDP*).

lifespan = 54.3 + 5.59 × log (*GDP*)

9 Substitute 20 000 for *GDP* in the equation to find the lifespan of people in a country with GDP of $20 000.

lifespan = 54.3 + 5.59 × log 20 000
= 78.3 years

3.5 Applying the reciprocal transformation

CAS 3: Using the TI-Nspire CAS to perform a reciprocal transformation

The table shows the length (in cm) and width (in cm) of eight sizes of sticky labels.

Length	6.8	5.6	4.6	4.2	3.5	4.0	5.0	5.5
Width	1.8	2.0	2.5	3.0	3.5	2.6	2.0	1.9

Using the $1/y$ transformation:

- linearise the data, and fit a regression line to the transformed data (*length* is the EV)
- write its equation in terms of the variables *length* and *width*
- use the equation to predict the width of a sticky label with a length of 5 cm.

Steps

1 Start a new document by pressing ctrl + N.

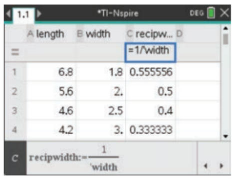
2 Select **Add Lists & Spreadsheet**.

Enter the data into lists named *length* and *width*.

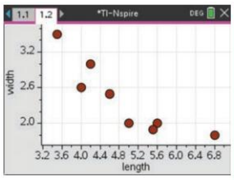
3 Name column C as *recipwidth* (short for 1/width).

Calculate the values of *recipwidth*.

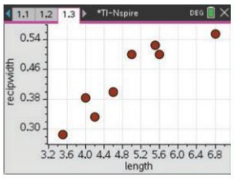
Move the cursor to the formula cell below the *recipwidth* heading. Type in **=1/width**. Press enter to calculate the values of *recipwidth*.



4 Press ctrl + I and select **Add Data & Statistics**. Construct a scatterplot of *width* against *length*. Let *length* be the explanatory variable and *width* the response variable. The plot is clearly non-linear.

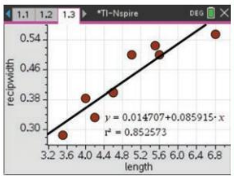


5 Press ctrl + I and select **Add Data & Statistics**. Construct a scatterplot of *recipwidth* (1/*width*) against *length*. The plot is now clearly linear.



6 Press menu>**Analyze>Regression>Show Linear (a + bx)** to plot the line on the scatterplot with its equation.

Note: The y in the equation on the screen corresponds to the transformed variable $1/width$.



7 Write down the regression equation in terms of the variables *width* and *length*.

$1/width$ = 0.015 + 0.086 × *length*

8 Substitute 5 cm for *length* in the equation.

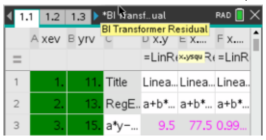
$1/width$ = 0.015 + 0.086 × 5 = 0.445

Thus width = 1/0.445 = 2.25 cm (to 2 d.p.)

3.6 Using the "BI Transformer Residual" template

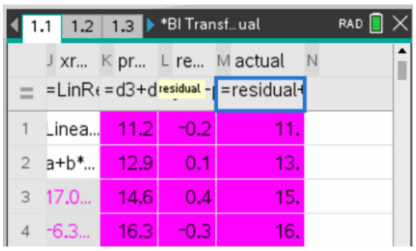
Steps

1. Enter new data into "xev", "yrv" columns (green boxes)

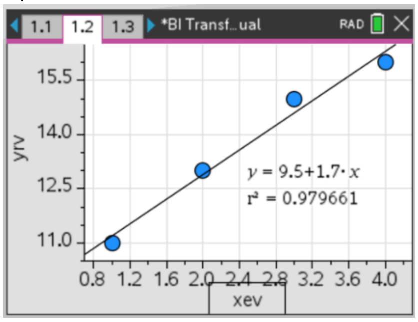


2. Transformed values will appear in columns E-J; r value and r^2 value can be found in rows 5 & 6 (blue and brown writing respectively); a & b can be found in rows 3 & 4 (pink writings) for equations.

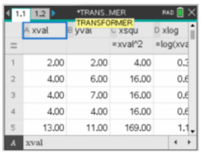
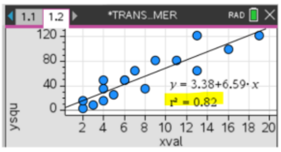
3. Bivariate data prediction Residual Actual Values can be found in Columns K L M (pink area)



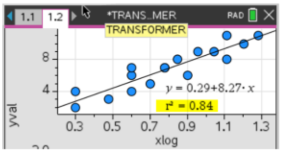
4. Transformed data plot in page 1.2. Residual plot in page 1.3. Click on "xev" & "yrv" to choose right transformation to show corresponding equations and r^2 value.

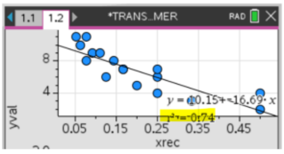


5. Example Qs

CAS Notes 3.6



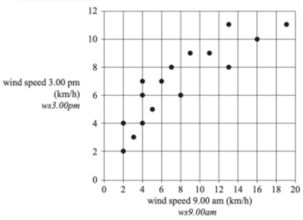


Apply the squared transformation to the variable ws3.00 pm and determine the equation of the least squares regression line that allows (ws3.00 pm)^2 to be predicted from ws9.00 am. In the boxes provided, write the coefficients for this equation, correct to 3 significant figures.

(ws3.00 pm)^2 = 3.38 + 6.59 × ws9.00 am r^2=0.82

(ws3.00 pm) = 0.29 + 8.27 × log(ws9.00 am) r^2=0.84 ✓

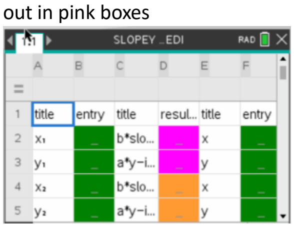
(ws3.00 pm) = 10.2 + 16.7 / ws9.00 am r^2=0.74



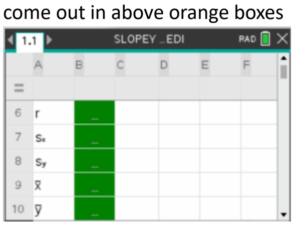
Wind speed (km/h)	
9.00 am	3.00 pm
2	2
4	6
4	7
4	4
13	11
6	7
3	3
16	10
6	7
13	8
11	9
2	4
7	8
5	5
8	6
6	7
19	11
9	9

3.7 Using "SLOPEY PREDI" template

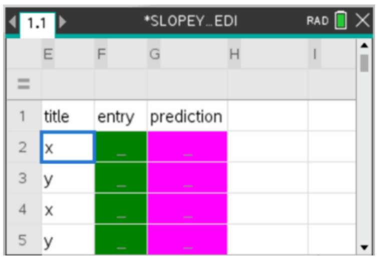
1. Two dots entering in green boxes and b slope and a y-intercept will come out in pink boxes



2. Following values entering green boxes, and b slope and a y-intercept will come out in above orange boxes



3. Given values of x or y values entering to column F, prediction will come out inside pink boxes



4.1 Constructing time series plots

CAS 1: How to construct a time series using the TI-Nspire CAS

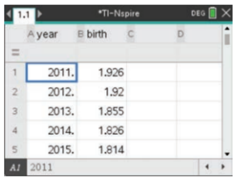
Construct a time series plot for the data presented below, which shows the birth rate in Australia (in births per woman) from 2011–2020.

year	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
birth	1.926	1.920	1.855	1.826	1.814	1.752	1.741	1.740	1.657	1.580

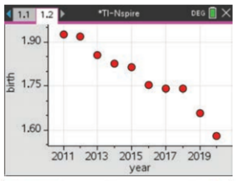
Steps

1 Start a new document by pressing ctrl + N.

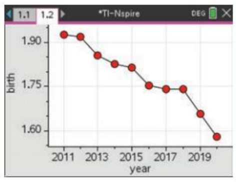
2 Select **Add Lists & Spreadsheet.** Enter the data into lists named *year* and *birth*.



3 Press ctrl + I and select **Add Data & Statistics**. Construct a scatterplot of *birth* against *year*. As is the case for a time series plot, *year* is the explanatory variable and *birth* the response variable.



4 To display as a connected time series plot, move the cursor to the main graph area and press ctrl + menu >**Connect Data Points**. Press enter.



4.2 Using the "SMOOTHIE" template

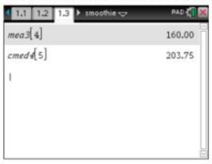
Using the "smoothie" template

If all formulas are correct, the following results should be visible for the dataset provided. Check that your template has generated the same smoothed values. Then save the file again (ctrl S)

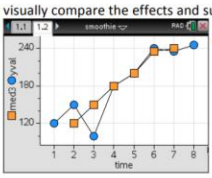
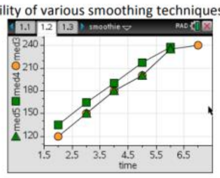
	A time	B yval	C mea3	D mea5	E med3	F med5	G mea4	H cmea4	I med4	J cmed4
=										
1	1.00	120.00	_	_	_	_	_	_	_	_
2	2.00	150.00	123.33	_	120.00	_	137.50	_	135.00	_
3	3.00	100.00	143.33	150.00	150.00	150.00	157.50	147.50	165.00	150.00
4	4.00	180.00	160.00	174.00	180.00	180.00	180.00	168.75	190.00	177.50
5	5.00	200.00	206.67	191.00	200.00	200.00	213.75	196.88	217.50	203.75
6	6.00	240.00	225.00	220.00	235.00	235.00	230.00	221.88	237.50	227.50
7	7.00	235.00	240.00		240.00					
8	8.00	245.00								

The template can be used to:

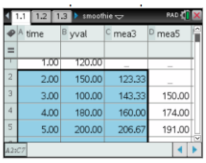
- find the correct smoothed values based on a given time period, or range of time periods

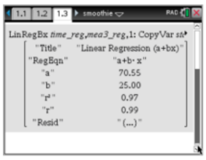


- visually compare the effects and suitability of various smoothing techniques

- create smoothed bivariate data sets upon which regression methods may be applied (would need to be copied and pasted to new lists to ensure that the two variables had same number of values, and lined up with time period correctly





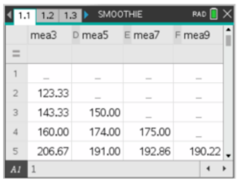
VIP Note!: To use the template with a new time series, select the **time** column and press ▲ until the entire column is selected. Then press menu **> Data > Clear Data** to clear the data (this does not delete the variable name, which is important as these names are used in the formulas in other columns).
Repeat for the **yval** column, and then enter the data for the new time series.

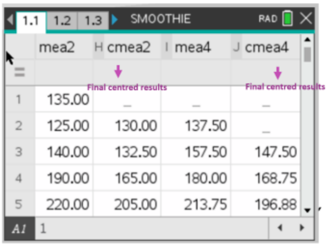
Using CAS SMOOTHIE to do 2/4-mean smoothing

CAS Notes 4.2

3, 5, 7, 9 mean smoothing

2, 4, 6, 8 mean smoothing



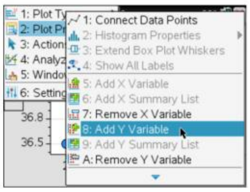


4.3 Comparing two or three time series graphs

To get both *temperature* and *threepointsmooth* on the *y*-axis, follow these steps.

To add a second *y*-variable, press:

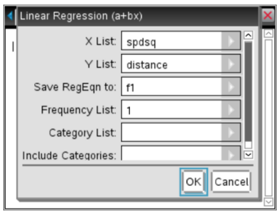
- MENU menu
- 2: Plot Properties 2
- 6: Add Y Variable 6
- Select threepointsmooth
- Enter enter



4.4 A full statistic display of LSRL

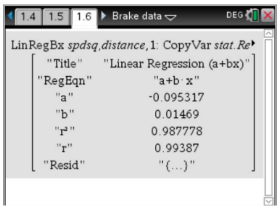
A full statistical display can be shown in a **Calculator** page using menu**>Statistics>Stat Calculations>Linear Regression (a+bx)**.

*Note: if you have performed a linear regression in the **Data & Statistics** page you can access the statistics in the **Calculator** page by pressing* var ***>Stat Results***.



The regression equation y=-0.0953+0.0147x in this case is:

distance= -0.0953 + 0.0147 *speed squared*

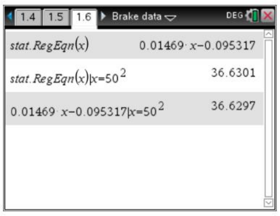


4.5 Using regression equation to predict

Use the regression equation to predict the stopping distance (to the nearest metre) of a car travelling at 50 km/h

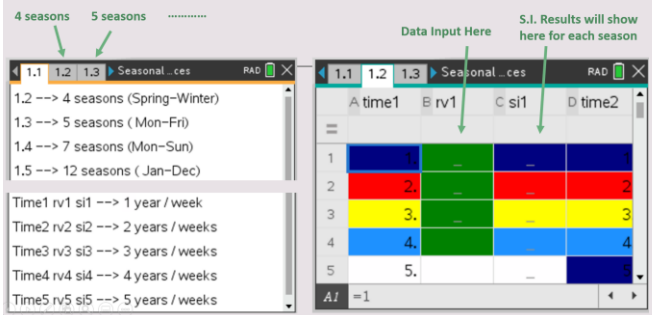
In the **Calculator** page recall the regression equation by pressing var and select **stat.regeqn** to paste to work area.

You need to add (x) after pasting as shown. When substituting in the speed value remember that the x now represents *speed squared* so enter as 50^2.



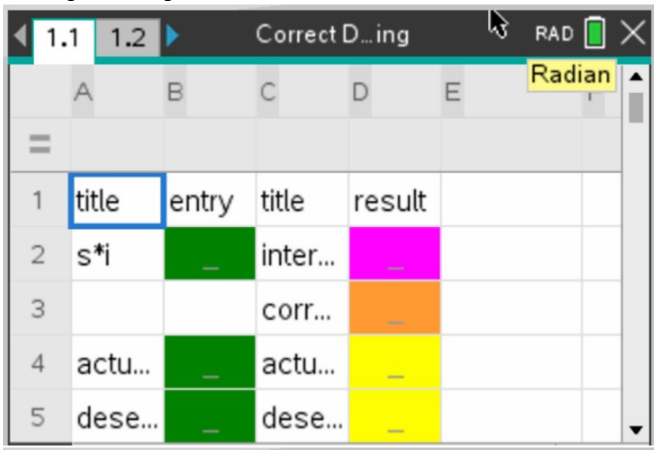
Answer: 37 m (to the nearest metre)

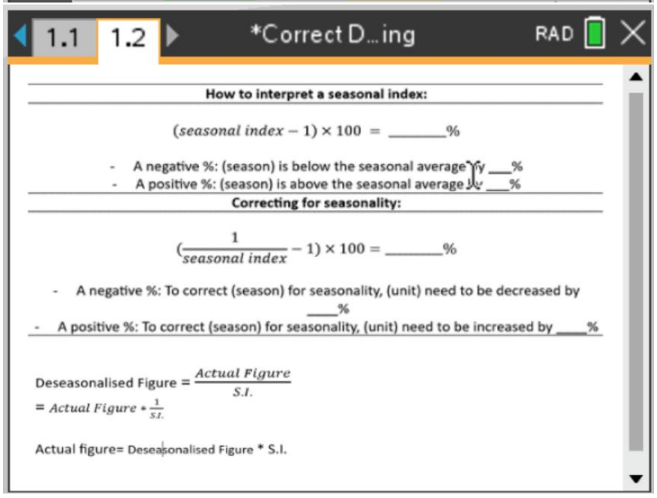
4.6 Using "SEASONING" template



4.7 Using "Correct De Seasoning" template

Entering data into green area, answers come out in coloured area





1a. Univariate Categorical Data: Frequency Table

The **[types of categories]** of **[total frequency] [frequency type]** were classified as **[list of categories].**

Modal Category

The majority of **[frequency type], [modal percentage],** were found to be **[modal category].**
Of the remaining **[frequency types], [percentage X]** were found to be **[category X],** while **[percentage Y]** were found to be **[category Y],** and while **[etc.].**

Equal Categories

The **[frequency types]** all had roughly the same percentages where **[category X]** had **[percentage X], [category Y]** had **[percentage Y], [etc.].**

1b. Univariate Numerical Data: Histogram, Dot Plot, Stem Plot

The shape of the distribution is **[symmetric/positively skewed/negatively skewed]**
The distribution has a **[standard dev./range/IQR]** of **[value]**
The distribution has a **[mean/median/mode]** of **[value]**
The distribution **[has #/has no]** outliers.

Refer to 1F: Describing numerical data

1c. Univariate Numerical Data: Box Plot

The distribution is **[positively skewed/negatively skewed]** with **[outliers/no outliers]**. The distribution is centered at **[value],** the median value. The spread of the distribution, is measured by the IQR, is **[value]** and, as measured by the range **[value].** If outliers present: There are **[value]** many outliers: **[list of outliers]**

2a. Bivariate Data (Both Categorical): Two-way Frequency Table

Worked example: Is there an association between interest in sports and age group?
Yes, the percentage of males with a high level of interest in sport steadily decreases with age group from 56.5 % for the 'under 18 years' age group, to 35.0% for the '36-50 years' age group.

Interest in sport	*Age group (%)* Under 18 years	19–25 years	26–35 years	36–50 years
High	56.5	50.2	40.7	35.0
Medium	30.1	34.4	36.8	45.8
Low	13.4	13.4	22.5	20.3
Total	100.0	100.0	100.0	100.0

2b. Bivariate Data (one categorical, One Numerical): Comparing two boxplots:

The distributions at **[variable name]** are **[symmetric/positively skewed/negatively skewed]** for both **[boxplot variables]**. There **[are/are no]** outliers. The median **[variable name]** is higher for **[boxplot 1], (M= value),** than **[boxplot 2], (M= value).** The IQR is also greater for **[boxplot 1], (IQR= value),** than **[boxplot 2], (IQR= value).** The range of **[variable name]** is also greater for **[boxplot 1], (R= value),** than **[boxplot 2], (R= value).**

2c. Bivariate Data (Both Numerical): Scatter Plot

There is a **[strong/moderate/weak], [positive/negative],[linear/non-linear]** relationship between **[response variable y]** and **[explanatory variable x]**. There **[are/are no]** clear outliers.

2d. The coefficient of determination (r^2):

The coefficient of determination indicates that **[r^2 x 100]** % of the variation in **[response variable]** is explained by the variation in **[explanatory variable]** and **[remaining %]** is explained by other factors.

2e. Least squares line:

The equation of the regression line is: **[response variable]** = **[a]** + **[b]** x **[explanatory variable]**

Slope (b):

On average, **[response variable] [increases/decreases]** by **[b units]** for every one **[unit]** increase in **[explanatory variable].**

y- intercept (a):

When **[explanatory variable]** is 0, **[response variable]** is predicted to be **[a units].**

2f. Residual Plot

The residual plot shows a **[random scatter/ curved pattern]** indicating there is a **[linear/non- linear]** relationship between **[response variable]** and **[explanatory variable].**