 Demonstrating the empirical rule (teacher guide)

This guide will model how staff can analyse collected data to verify the empirical rule. The specific dot point is:

* Use collected data to illustrate that, for normally distributed random variables, approximately 68% of data will have z-scores between -1 and 1, approximately 95% of data will have z-scores between -2 and 2 and approximately 99.7% of data will have z-scores between -3 and 3 (known as the Empirical Rule)

Part 1: Collect the data

The data used in this guide were sourced from New Zealand’s census at school dataset limited to Year 12 and 13.

* + demonstrating-the-empirical-rule.XLSX, this is a completed MS Excel workbook with the raw data, formulas, pivot table and graphs described in this guide. This is intended to be used as a reference workbook.

demonstrating-the-empirical-rule-template.XLSX, this is a MS Excel workbook containing the raw data and statistics worksheet to be completed.

For this guide, we will examine the heights of the students. These are located in column R of the raw data worksheet in demonstrating-the-empirical-rule.XLSX.

Part 2: Check if the data is visually approximately normal

* Create a frequency distribution table for the data by:
	+ Selecting the data: In this case the R is left clicked.



* + From the toolbar, select insert then PivotTable



* + Check the PivotTable is table being inserted into a new worksheet then click ok.



* + On the right hand side of your screen you will have a PivotTable field menu. Left click and drag and drog the variable into the rows and values fields.



Note: If it doesn’t default to Count of…, this can be set by selecting the drop down arrow then value field settings.

* Create a column graph of the data to see if its approximately normal.
	+ Click somewhere into the pivot table:



* + Insert a column graph by selecting insert then the column chart symbol.



* + Check the distribution is approximately normal.
* Note: If you wish to construct a xy scatter plot, the data would first need to be copied and then pasted as values only.



Part 3: Statistics, z-score ranges and percentage of scores

The following calculations and screenshots are from the Statistics worksheet of the MS Excel file.

* Calculate the mean height of the data set: =AVERAGE('Raw Data'!R:R)

For your worksheet, this can be done by

* + typing =average(
	+ selecting the worksheet where your data is
	+ selecting the column letter where the data is.
	+ Pressing the enter key
* Calculate the mean height of the data set: =STDEV.P('Raw Data'!R:R)

Follow the method above after typing =STDEV.P(

* Calculate the number of scores in the data set: =COUNT('Raw Data'!R:R)

Follow the method above after typing =COUNT(

* Calculate z-score ranges by adding or subtracting appropriate multiples of the standard deviation from the mean. Sample formulas:



* Notes:
	+ Despite having 1000 scores, only 992 are shown as there were blank responses.
	+ The cells have been reformatted to display 2 decimal places.



* Calculate the number of scores in the ranges. This can be done by subtracting the number of scores less than the lower bound from the number of scores less than the upper bound.

Sample formula for -1 to 1 (Cell D7): =COUNTIF('Raw Data'!R:R,"<"&Statistics!C7)-COUNTIF('Raw Data'!R:R,"<"&Statistics!B7)

* Calculate the percentage of scores in the ranges.

Sample formula for -1 to 1 (Cell E7): =D7/$B$4

