Pose Question & Collect Data

Come up with a specific question to answer

- Summary Question: (one variable) e.g., Find the typical height of the students in the class.
- Comparison Question: (one variable) e.g., Do boys or girls spend more time on the internet?
- Relationship Question: (two variables) e.g., Do students who study more do better in exams?

Collect Data

- What data do I need? Categorical (Qualitative): Nominal, Ordered
  Numerical (Quantitative): Discrete, Continuous
- What sampling method will I use? Simple Random, Stratified, Cluster, Quota
- How will I eliminate bias? random selection, careful questioning, who, when & where
- What will the source of data be? Primary/Secondary, questionnaire, C@S, official records

Analyse the Data

Analyse the Data – Descriptive Statistics

Statistics on the sample data

Distribution

Statistical distribution describes the number of times each possible outcome occurs in a sample.
Distribution Table / Frequency Distribution Table / Grouped Frequency Distribution Table

Choose the Appropriate Visual Representation

Nominal (male/female): Bar Chart, Line Plot (Dot Plot), Pie Chart
Ordinal (never/sometimes): Bar Chart, Line Plot, Pie Chart
Discrete (no. of cars/age in years): Bar Chart, Pie Chart, Line Plot, Stem and Leaf Plot
Continuous (height/foot length): Histogram, Stem and Leaf Plot

- Bar Charts good for comparing frequencies
- Pie Charts good for showing proportion of the total sample
- Dot plots useful for representing a small sample. Particularly good for showing central tendency, dispersion and shape.
- Stem and Leaf Plots useful for representing a sample of discrete or continuous data. Particularly good for showing central tendency, dispersion and shape.

Summary of the Data (Univariate)

- Central Tendency
  - → Mean
  - → Median
  - → Mode
- Dispersion (Spread, Variability)
  - → Range
  - → IQR: Inter Quartile Range
  - → Standard Deviation
- Shape
  - → Gaps/ Clusters
  - → Outliers
  - → Modality
  - → Symmetric
  - → Bell Shaped
  - → Skewed
  - → Normal

The Five-Number Summary

1. Maximum
2. Minimum
3. Median
4. First Quartile
5. Third Quartile

Comparison of Data (Univariate)

- All of the above summary techniques used to compare sets of data

Relationship between Variables (Bivariate)

- Scatterplots
- Correlation Coefficient
- Line of Best Fit

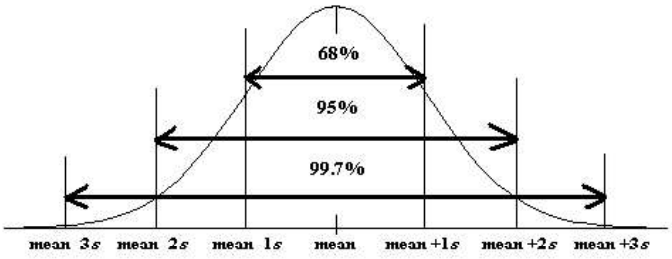
Interpret the Results

Interpretation of the Results to Answer the Question Posed

Non-Inferential Statistics

Making a generalisation about the sample data or when the sample data is the same as the population

- Interpreting the summary statistics to answer the question posed.
- Making a comparison between summary statistics: differences/similarities.
- Empirical Rule: Interpreting a Normal Distribution (for a normal distribution, almost all data will fall within three standard deviations of the mean). Otherwise known as the 68 – 95 – 99.7 rule.



- Z-scores: A z-score gives us an indication of how unusual a value is because it tells us how far it is from the mean on a Standardised Distribution Curve. If the data value sits right at the mean, it's not very far at all and its z-score is 0. A z-score of 1 tells us the data value is one standard deviation above the mean, while a z-score of -1 tells us that the value is one standard deviation below the mean.

Inferential Statistics

The data is taken a step further to make a generalisation about the population from which the sample is taken.

No deterministic statements

- We cannot make a deterministic (definite/absolute) statement about the population because the sample we took was just our best attempt to represent the population. There will be some variation.
- The vocabulary used in statements about the population must not be deterministic – use: "tends to", "estimation" , "inference"

Correlation and Association

- Is there an association between the two variables? Causation: Does one variable change because the other variable changes?
- Is there a correlation between the two variables? What does the correlation suggest about the population? E.g., One variable "tends to" increase as the other variable increases.

Margin of Error

- Since the sample is not the same size as the population there is a margin of error that accompanies any inferred statistic about the population.
- The bigger the sample, the smaller the margin of error, $\frac{1}{\sqrt{n}}$.

Hypothesis Testing using the Margin of Error

- Using the margin of error and the statistics from the sample to test if a statement about the population could be true.

Terms: *Descriptive/Non-Inferential/Inferential Statistics for teacher's information only.*