

ECON1203 Topic 1 – What is Statistics?

Key Concepts (1.1)

- **Statistics:** involves collecting, analysing & interpreting data
 - **Descriptive Statistics:** Organizing, Summarizing & Presenting data in an informative way
 - **Inferential Statistics:** A body of methods to draw conclusions or inferences about characteristics of populations based on sample data
- **Population:** The group of all items of interest (*e.g. people, animals, plants or things*)
 - **Parameter:** The descriptive measure of a population
 - Used to represent certain population characteristics
 - Usually represents the information we need (*mean # of drinks consumed at the Uni (pg5)*)
- **Sample:** A set of data drawn from the studied population
 - Use small & manageable samples to draw conclusions about the larger group (population)
 - *E.g. computing the mean # of soft drinks consumed by 500 Uni students to infer the population mean*
 - **Statistic:** is a descriptive measure of a sample, a quantity calculated from a sample data
- **Statistical Inference:** The process of making an estimate, prediction or decision about a population based on sample data
 - Confidence Level – Proportion of times the statistical inference will be correct
 - Significance Level – How frequently the conclusions will be wrong
- **Exit Poll:** A random sample of voters is asked who they voted for (statistic) & through statistical inference we estimate the results of the population.

Graphical Descriptive Techniques I – Chapter 2

Type of Data & Information (2.1)

- **Variable:** Is some **characteristic** of a population or sample (*E.g. the mark received on a statistics exam*)
 - **Values:** Possible observations of the variable
 - **Data:** The observed values of a variable
 - **Discrete Variables:** Whole numbers or categories. *E.g. cannot get 2.5 heads (coin flipping)*
 - Cannot take on all values between max/min value
 - **Continuous Variables:** Variable can be any value between max/min value incl. fractions
 - *E.g. time taken to finish an exam*
 - **Interval/Quantitative/Numerical Data** → Real numbers
 - All calculations permitted
 - **Nominal/Qualitative/Categorical** → categories
 - No Calculations permitted, only frequency or % occurrence
 - **Ordinal** → Appears to be nominal but order of their values has meaning
 - *E.g. poor - 1, fair - 2, good - 3, very good - 4, excellent – 5 (MAINTAIN # ORDER / could be 6,18,23,45,88) (pg14)*
 - Only Ranking or Ordering Date calculations e.g. Median (pg 16)
 - **Time Series:** Data referring to measurements at different points in time
 - **Cross Sectional:** Data measured at a single point in time

Nominal Data

Describing a Set of **Nominal** Data (2.2)

- **Frequency Distribution Tables**

- Frequency: How often the event occurs.
- Relative Frequency: Lists the categories/bins and the proportion with which they occur (of total)
- Cumulative Frequency: Accumulating total of each category
- **Mutually Exclusive**: Results can only be included in one category (*e.g. mode of transport – only can choose one out of car, bus, walk, cycle etc.*)
- **NB: Ordinal data should be arranged in order**

Result	Frequency	Cumulative Frequency	Relative Frequency
2	6	6	30%
3	4	10	20%
4	10	20	50%

- **Bar Charts**

- Visual representation of data presenting the frequency of events
- Typically used for **qualitative** data (hence spaces between categories/variables)

- **Pie Charts**

- Visual representation of data presenting the relative frequency
- Typically used for **qualitative** data

Describing the Relationship between Two **Nominal** Variables (2.3)

- **Bivariate**: Shows the relationship between two variables
- **Cross Tabulation/Classification Table** (Row relative frequency table) (**USED FOR NOMINAL DATA**)

Occupation	G&M	POST	STAR	SUN	Total
Blue Collar	27	18	38	37	120
White Collar	29	43	21	15	108
Professional	33	51	22	20	126
Total	89	112	81	72	354

Occupation	G&M	POST	STAR	SUN	Total
Blue Collar	.23	.15	.32	.31	1
White Collar	.27	.4	.19	.14	1
Professional	.26	.4	.17	.16	1
Total	.25	.32	.23	.2	1

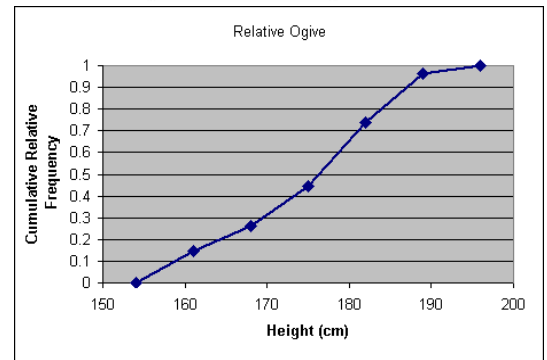
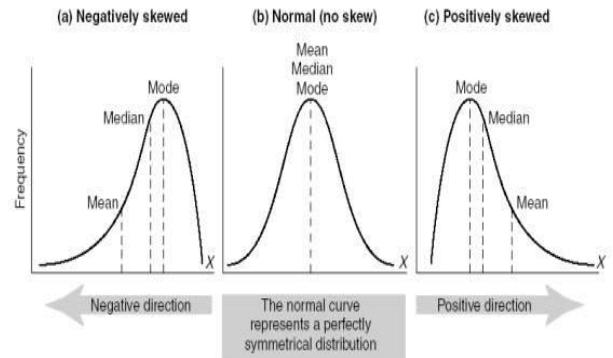
- **Graphing Techniques**

- Side By Side Bar Charts (Used for **nominal** data)

Graphical Descriptive Techniques II – Chapter 3

Graphical Techniques to Describe a Set of Interval Data (3.1)

- **Histogram (Cross Sectional)**
 - Create BINS/Classes to categorize numbers
 - too many → doesn't summarize
 - too few → not enough information
 - must be mutually exclusive & exhaustive
 - DO AN EXAMPLE
 - Sturges Formula – $1 + 3.3\ln(n) = \text{no. of bins}$
- **Key Features of Histograms**
 - **Symmetry:** If a straight line is drawn through the middle, separating 2 identical sides
 - **Skewness:** a long tail extending to the left/right
 - Which Modal Class (Uni or Bi – Modal)
 - Is it Bell Shaped
 - Outliers, Clusters Etc.
- **Ogive (Cumulative Relative Frequency Graph)**
- **Stem & Leaf**
 - Advantage over Histogram as we can see actual observations and don't lose potentially useful information



Stem	Leaf
1	80
2	40 60 70
3	10 30 60 70 90
4	10 80
5	00
6	
7	10 30
8	90

Describing Time Series Data (3.2)

- **Line Chart / Time Series Plot**
 - Plotting the variable over time

Describing the Relationship between Two Interval Variables (3.3) → Bivariate Relationships

- **Scatter Diagram/Plot**
 - Data for 2 variables, Independent (X) / Dependant (Y)
 - E.g. House size / House Price
- **Key Features of Scatter Plots**
 - **Linearity**
 - Positive/Negative Linear Relationship. **NB** Correlation is not causation
 - Non – Linear Relationship
 - No Relationship
 - Clusters
 - Outliers