BUSINESS STATISTICS

LECTURE 1

Intro

Stats draw conclusions about large groups of individuals/items with the use of information gathered from these subjects

- Reliable forecasts
- Descriptive stats: collecting, summarizing, describing data
- Inferential stats: conclusions, making decisions concerning a population on sample data

Process of DESCRIPTIVE

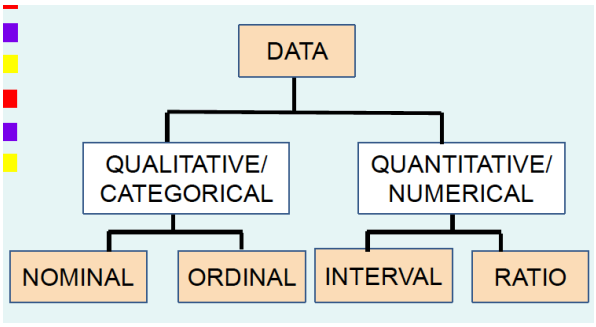
1. Collect data - survey
2. Present data – tables/graphs
3. Characterize data – sample mean

Process of INFERENTIAL

1. Estimation – using a sample mean
2. Hypothesis testing – testing a claim

Definitions

- **Population** – collection of all possible individuals, objects, or measurements of interest. Measured in parameters – UPPERCASE Greek LETTER
- **Sample** – a portion, part or subset of the population of interest. Measured in statistics 0 LOWERCASE Greek LETTER



- **Nominal** = data that is classified into non-overlapping categories and cannot be arranged in any particular order / sorted e.g. eye colour, gender

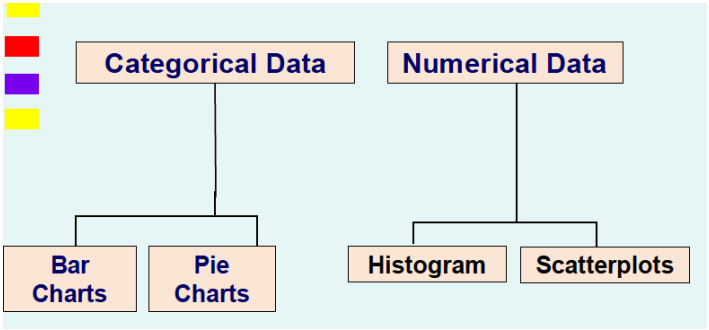
- **Ordinal** = Data that can be classified into distinct non ordering categories in which ranking is implied e.g. data can be arranged in some order/sorted. e.g. A, B, C

- **Interval =** For quantity numbers, but measurements do not have a true zero point e.g. difference between temperature on Fahrenheit scale, shoe size

- **Ratio =** The interval between inherent zero and starting point e.g. prices, speeds

Graphs for numerical data

- Scatter plot graph = A plot or graph of pairwise data from two continuous variables, to explore the relationship between them
- Pie chart = Circular display graph showing percentages in sectors of data
- Bar chart = **Each category represents a bar**, the length of the bar is the amount/frequency/percentage of the values
- Histogram = **Each bar represents a set of numbers**. Type of vertical bar where the area of each is equal to the frequency of the corresponding interval (basically bar chart but the bars

	touch)  **Data collection types** • Cross Section data = data collected at one point in time • Time series = data is collected overtime

LECTURE 2

Numerical Measures	Measures of **central tendency** – the location of the center of a set of number 1. Mode – most often 2. Median – middle 3. Mean – average (add all scores up and divided by total number of scores) **Population Mean (μ)** a. *N = number of values* b. *x = any particular value* c. *Σ = operation of adding* d. *Σx = sum of x values in the population* $\mu = \frac{\Sigma x}{N} = \frac{x_1 + x_2 + x_3 + \ldots + x_N}{N}$ **Sample Mean** a. *x = any particular value* b. *Σ = operation of adding* c. *Σx = sum of the x values in the population* d. *n = number of values or sample size* $\bar{x} = \frac{\Sigma x}{n} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$ **Median** a. Applicable for ordinal, interval and ratio data only because they're ranked and sorted b. Not effected by outliers c. E.g. property market prices d. Process: put the numbers in ascending order and then pick out the middle number **Mode** a. Applicable to all levels of measurement b. Count
Measures of dispersion/ spread	**MEASURES OF VARIRABILITY** **Range** • Largest – Smallest = range

1. POPULATION Variance σ^2 (Sigma Square)

- The average of the squared deviations from the arithmetic mean
- Explaining formula
  - μ = mean (basic formula to calculate)
  - $x - \mu$ = calculating how far each score is from the mean so do a subtracting
  - Then $x-\mu^2$ = square each number from the set
  - $\Sigma(x - \mu^2)$ = add up these numbers for SIGMA, then square root
  - Then divide by N (number of values)
  - = VARIANCE

Population Mean: $\mu = \frac{\sum x}{N}$

Population Variance: $\sigma^2 = \frac{\sum(x-\mu)^2}{N}$

2. POPULATION Standard Deviation σ (Sigma)

The standard deviation is a measure of the spread of scores within a set of data.

- Square root of the variance

Population Standard Deviation: $\sigma = \sqrt{\sigma^2}$

SAMPLE = 1. Variance and 2. Standard Deviation

Sample Mean: $xbar = \frac{\sum x}{n}$

SAMPLE Variance: $s^2 = \frac{\sum(x-xbar)^2}{n-1}$

Sample Standard Deviation: $s = \sqrt{s^2}$

Coefficient of variation (COV)

- Is defined as **the ratio** of the standard deviation to the mean, expressed as a percentage
- Measurement of relative dispersion and used to compare standard deviation/variability of datasets with different means
- *Questions will ask for interruption – so know what each formula actually means*

$$CV = \frac{s}{\bar{x}}(100)$$

s= sample standard deviation
xbar=sample mean
In the formula, multiply by 100 coverts the answer to percentage

NOTE: For population CV, replace in the formula with population parameters σ and μ

Note: so s= sigma and x = m

z score

- Negative = below average
- Positive = above average
- A z-score of 0 implies x=μ
- Normally distributed data