

SEN-SRN Conversion Course
Research Module



Analysing Quantitative data

Presented By:

Andrew Xuereb, SRN
PQ Dip, MSc (HSM)

Lecture Overview

- Descriptive Statistics
 - Median, Mean, Standard Deviation
 - Frequency Distribution
- Inferential Statistics
 - Testing Hypotheses, Level of significance
- Statistical Tests
 - Analysis of Variance (Anova)
 - t test
 - Chi square Test
 - Correlation



Quantitative Data - Analysis

- Dealing with figures
- Easiest way is to use computer software using:
 - Spreadsheets Program (putting them in table: rows and columns, easier to apply histogram, pie chart etc.)
 - Statistical package software, e.g. Statistical Package of Social Science (SPSS)
- Consult a statistician



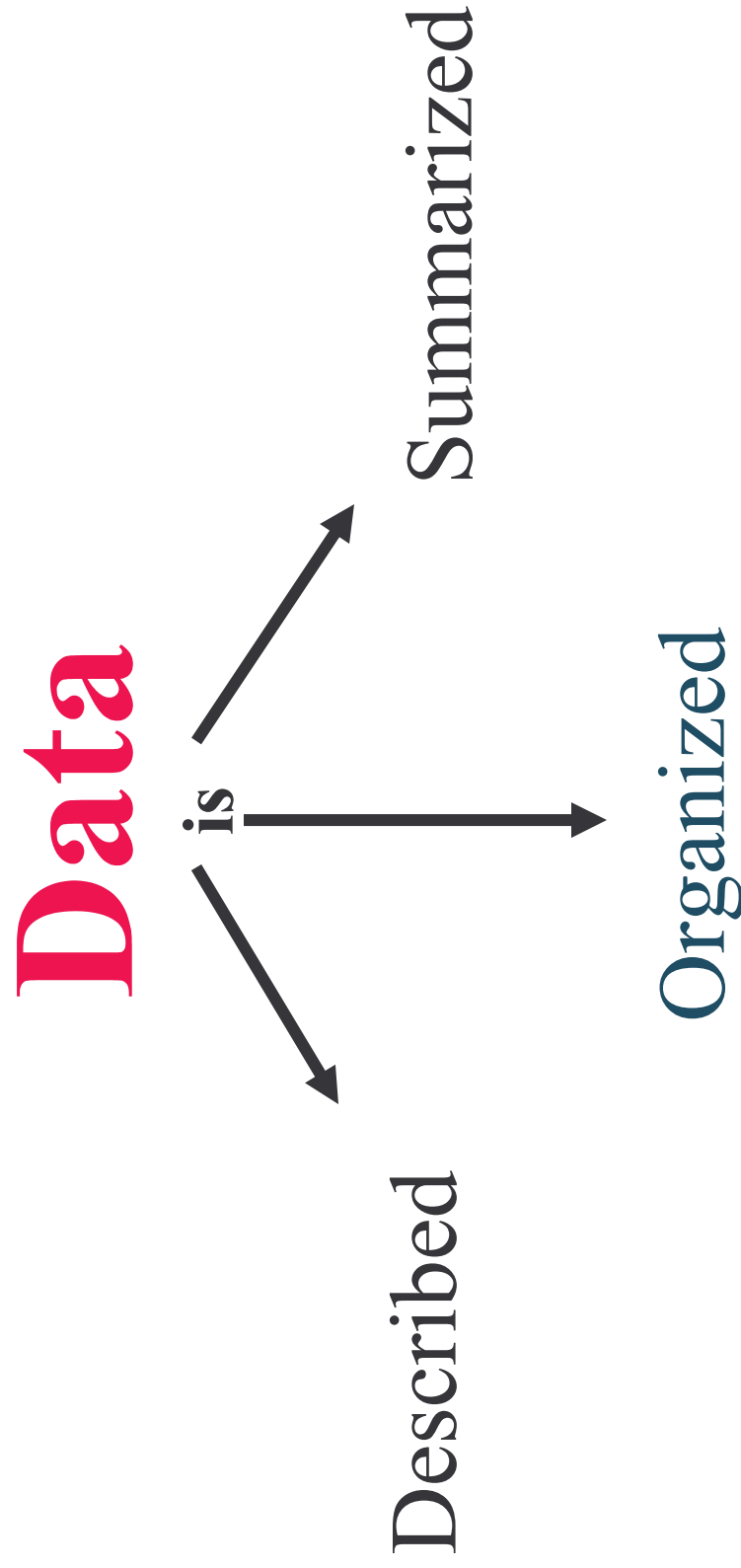
Quantitative Data – Principles of Analysis

- **1st Stage:** Findings are grouped in rows and columns
- **2nd Stage:** Look for consistent trends and patterns e.g. counts, means, standard deviation (descriptive statistics)
- **3rd Stage:** Look for particular trends and patterns e.g. differences between sexes and age groups (inferential statistics)
- **4th Stage:** Presentation of findings in a series of tables and charts
- **5th Stage:** Interpretation or explanation of findings in the research report

Descriptive Statistics – What?

- **Frequency Distribution:** An analysis method that involves determining how often scores or values appear in a data set.
- **Percentages:** numerical information calculated out of every hundred
- **Mean:** is simply the sum of the values divided by the number of subjects. Also called the average.
- **Median:** It is the measure that corresponds to the middle score. It lies at the midpoint of the distribution and divides the score into halves.
- **Mode:** Identifies the most popular score measure of the central tendency. The score or value that occurs most frequently in a distribution of scores.

Descriptive Statistics – Why?



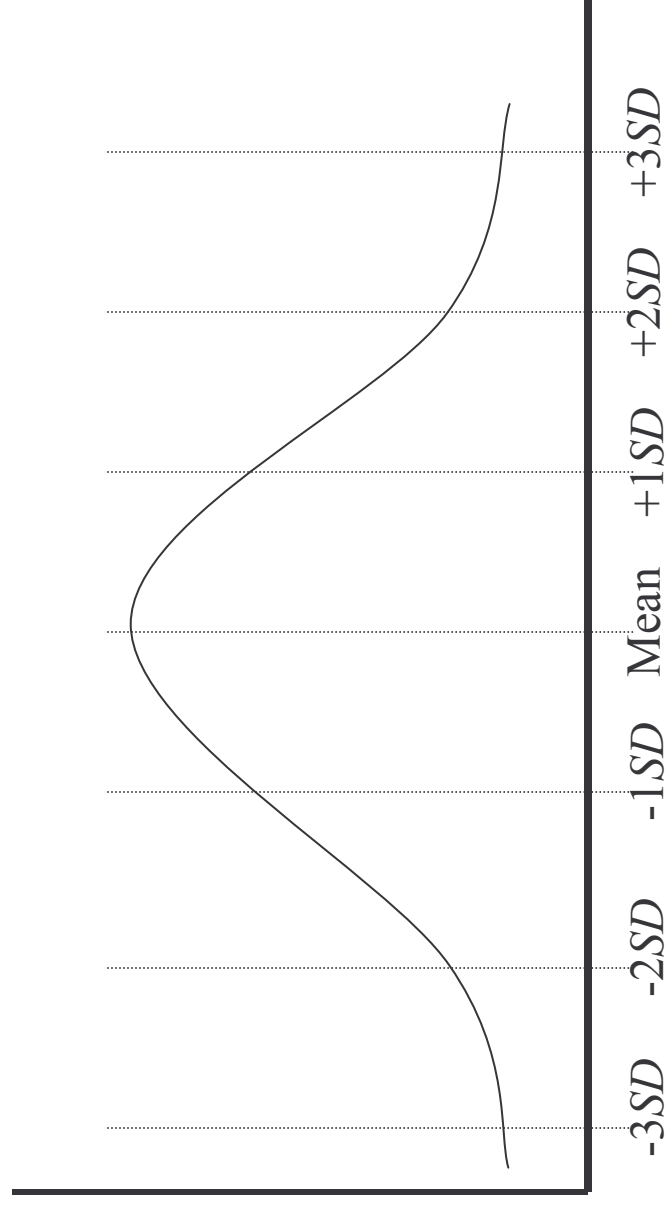
Descriptive Statistics - Table

	Total	Agree	Disagree	No Response
Respondents	100	45	39	16
Percentage	100	45	39	16

Descriptive Statistics - Standard Deviation

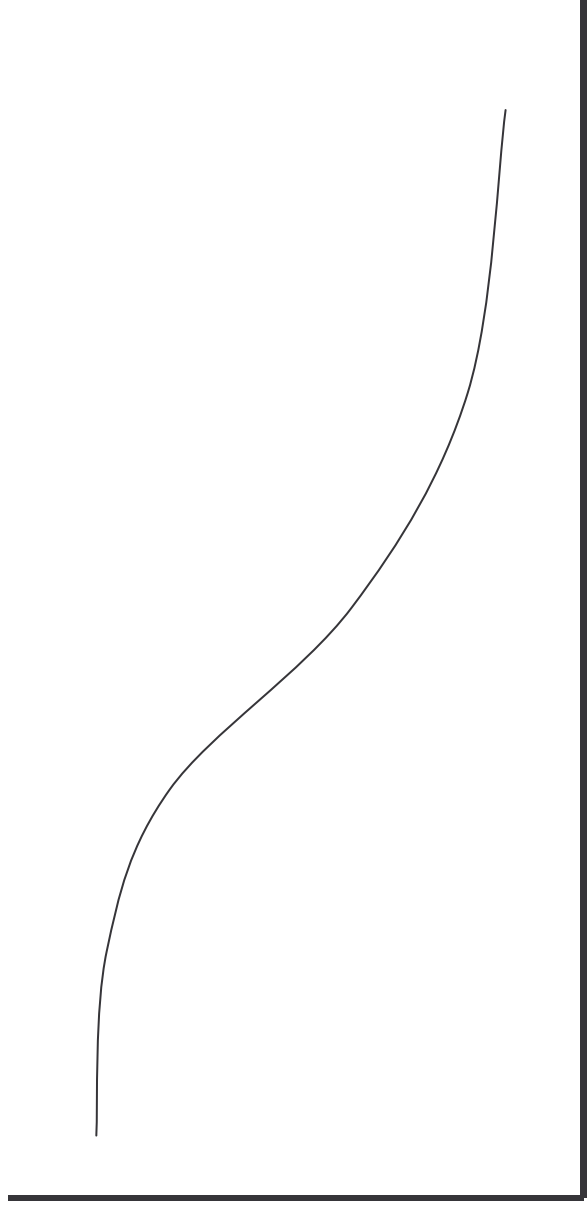
- Shows how the scores are grouped around the midpoint
- Referred to as 's' or 'SD'
- Example: 3 students with exam results:
 - **English:** 30, 60, 90 (Mean or average = 60) SD:17.32
 - **Maltese:** 59, 60, 61 (Mean or average = 60) SD: 0.57
 - Do results have the same significance although with the same mean or average for the English and Maltese result?
 - Results have significantly different SD: 17.32 **Vs** SD 0.57
- To be more descriptive: means should show SD

Normal Distribution Curve

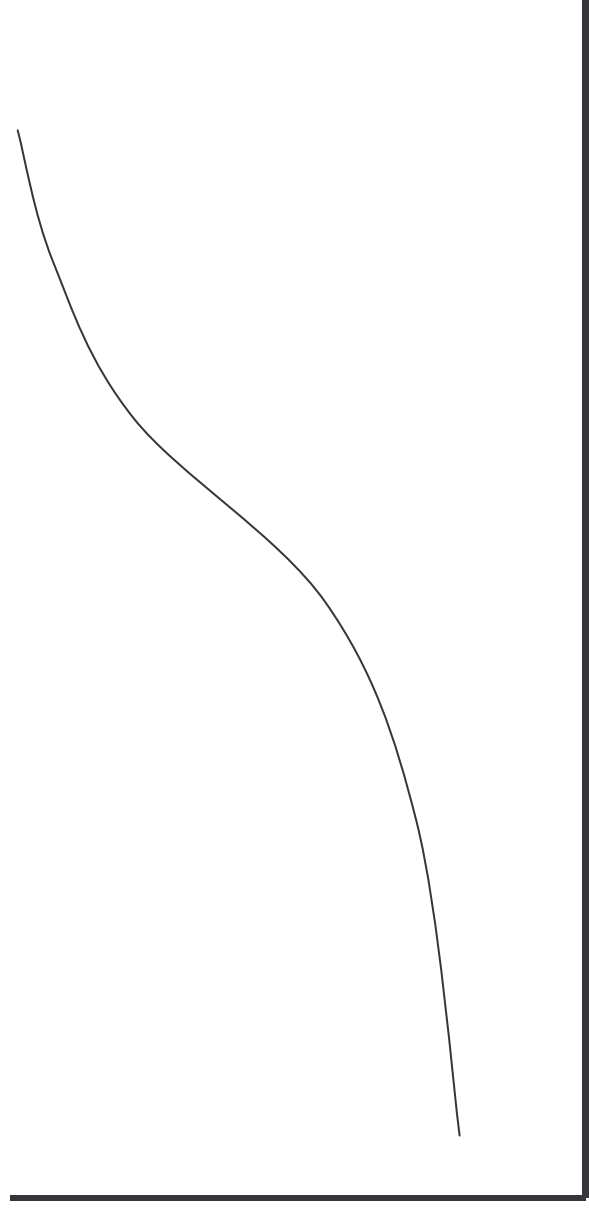


- Graph shows that 68% of population lie within 1 SD on either side of the mean

Positively Skewed



Negatively Skewed



Ordinal Vs Nominal Data

- **Ordinal level data:** refers to numbers, scores along some sort of dimension e.g. stress scale
- **Nominal level data:** assignment of characteristics into categories, e.g. male (category 1) and females (category 2)
- Important to distinguish for the identification of the best statistical tool fitting the type of data
- Consult a statistician if necessary

Level of Significance

- Two most frequently used level of significance (α or alpha) are 0.5 and 0.1. Often referred to p value
- ≤ 0.5 : Equal or Less less than 1 in 20 chance that it could happen by chance alone
- ≤ 0.1 : Equal or Less less than 1 in 100 chance that it could happen by chance alone
- **Type 1 error (*alpha error*)**: An error created by rejecting the null hypothesis when it is true (i.e. the researcher concludes that a relationship exists when in fact it does not
- **Type 2 error (*beta error*)**: Occurs when a researcher concludes that differences exists between groups were due to chance when in fact they were due to the effects of the Independent variable



Statistical Tests

- **t test:** Statistical test used to determine whether means of two groups are significantly different
- **ANOVA (analysis of Variance):** used to compare the means of two or more groups

Chi-square (χ^2)

- Used in nominally scaled data. Assess whether a relationship exists between two or more categories. Values in each cell cannot be less than 5.

Table 4.25 Levels of Short-term Absence by Levels of Stress

Level of Short Term Absence	Not stressed vs stressed		Total
	Not stressed	Stressed	
Low	19	22	41
Moderate	9	36	45
High	10	50	60
Total	38	108	146

Pearson Chi-squared Test: $\chi^2=12.368$, d.f. 2, $p=0.002$

Correlation

- **Correlation Coefficient:** Seeks to assess the extent of association between two variables. Results vary from +1.0 (a perfect positive or direct association) to zero (no correlation) to -1.0 (a perfect negative or indirect association).
- Positive association occurs when a variable's score tends to increase or decrease as the other increases or decreases in the same direction (negative in the opposite direction)

Absence Measures Vs Self-estimation of Stress for Total Population

<i>Absence Measures</i>	Total Population (n=143)	
	<i>r</i>	<i>p</i>
Attitudinal Index	0.169	0.021
Total Short-term	0.208	0.006
Total Spells	0.203	0.007
Total Days	0.114	NS

Statistical Test: Spearman's Correlation Coefficient