Ollscoil
Teicneolaíochta
an Atlantaigh

Atlantic
Technological
University

Lean Sigma BIO 08025

Lecture 5-Measure Phase-Descriptive Statistics

Dr. Stephen Daly

Six Sigma

1. Overview of the Measure Phase
2. Types of Data
3. Data Collection
4. Measures of Central Tendency
5. Measures of Variability

Purpose of Measure Phase

- To thoroughly understand the current state of the process and collect reliable data on process speed, quality, and costs that you will use to expose the underlying causes of problems

Measure Phase Deliverables

- Fully developed Value Stream Map
- Reliable data on critical inputs (Xs) and critical outputs (Ys) to be used for analysing defects, variation, process flow, and speed.
- Baseline measures of process capability, lead times, Sigma quality levels etc.
- A robust measurement system

The main steps of the Measure Phase

1. Create/Validate a value stream map to confirm current process flow.
    - Start with a basic process map and add data to generate the VSM

2. Identify outputs/inputs and process variables relevant to your project.
    - You need to collect data that relates to your project goals and targeted customers

3. Create data collection plan

4. Create data analysis plan

    Verify what types of tools can be used for the type of data you are collecting

The main steps of the Measure Phase

5. Use measurement system analysis and gage R&R
    5. Evaluation of repeatability and reproducibility

6. Collect and establish baselines

7. Update Value Stream Map with data

8. Use Littles Law to calculate Lead times (process lead time=no of things in process/avg completion rate)

The main steps of the Measure Phase

9. Perform Process Capability Evaluation

10. Make Quick Hit Improvements
    - If the data shows you can get partial benefits now then implement- however, be sure you are in a position to measure and show improvement.
    - Use a Kaizen approach
    - If solution ideas come up but the risks are high or unknown, keep track of the ideas for potential implementation down the line but continue with the DMAIC process.

11. Perform Measure Gate Review

Measurements and their importance for quality

When we use Six Sigma ideas and principles to analyse and improve a process, we are trying to answer questions such as:

- How satisfied are our customers?
- How good are our products or services?
- How good are our processes?

In many cases we are looking for both numerical and descriptive answers to such questions.

Measurements and their importance for quality

- Measurements are not just numbers – they are the outcome of a measurement process.

- Many factors can influence the value of a measurement and create variability in the value.

Measurement Variability

- Variability can come from the instrument used, the procedures followed, the actions of the people involved, the time of day, the order the measurements were taken in etc.

- The **accuracy** of a measurement is the closeness of the measurement to the true or actual value of the quantity being measured.

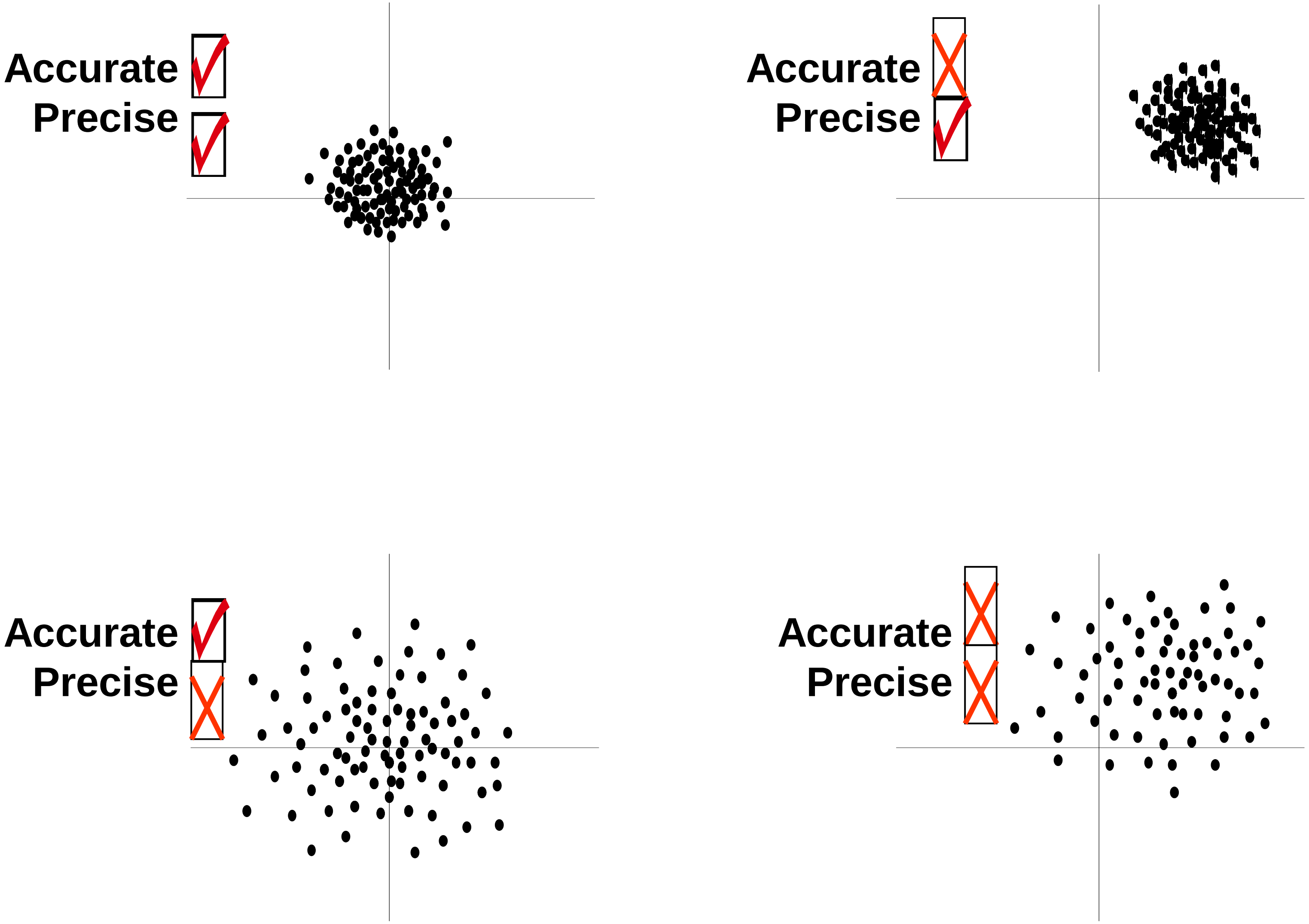
- Very few measurements are perfect – we call this variation **measurement error** or measurement noise.

Repeatability & Reproducibility

- **Repeatability** – how well repeated measurements carried out by the same person agree with themselves (this addresses one aspect of precision).

- **Reproducibility** – how well repeated measurements carried out by different people or labs agree with each other (also an aspect of precision).

The “Bull’s-Eye Analogy Accuracy & Precision



Summary

- **Accuracy (Validity)**

  Accuracy is the deviation between the actual measurement and the true value (as assessed by periodic calibration).

- **Precision (Repeatability)**

  Precision is the ability to obtain the same reading each time the measure is taken.

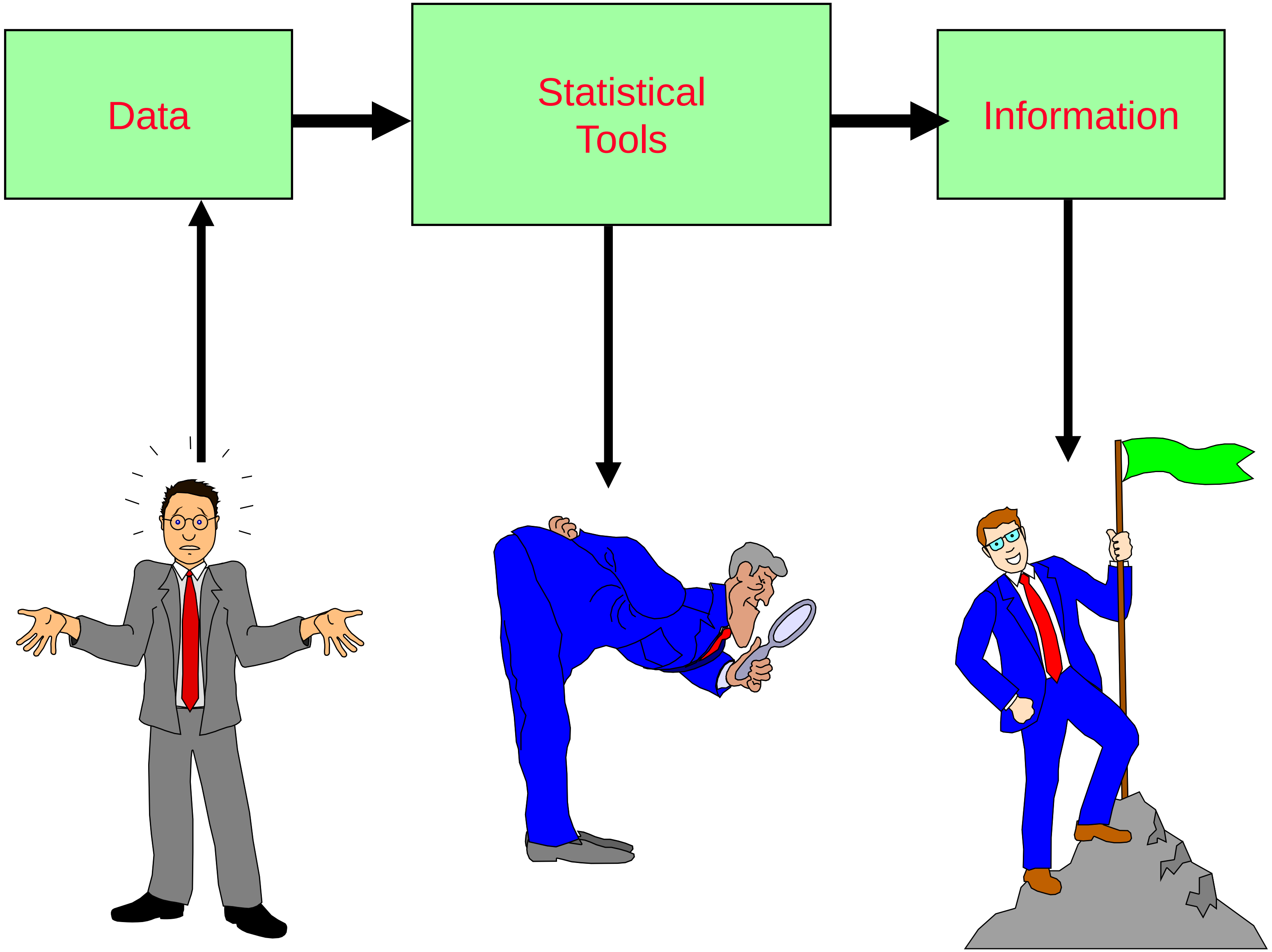
Atlantic Technological University

Types of Statistical Analysis

- Descriptive Statistics
  - Graphical Tools
  - Numerical Measures
- Inferential Statistics
  - Populations
  - Samples
- Probability
  - Linking Descriptive and Inferential Statistics

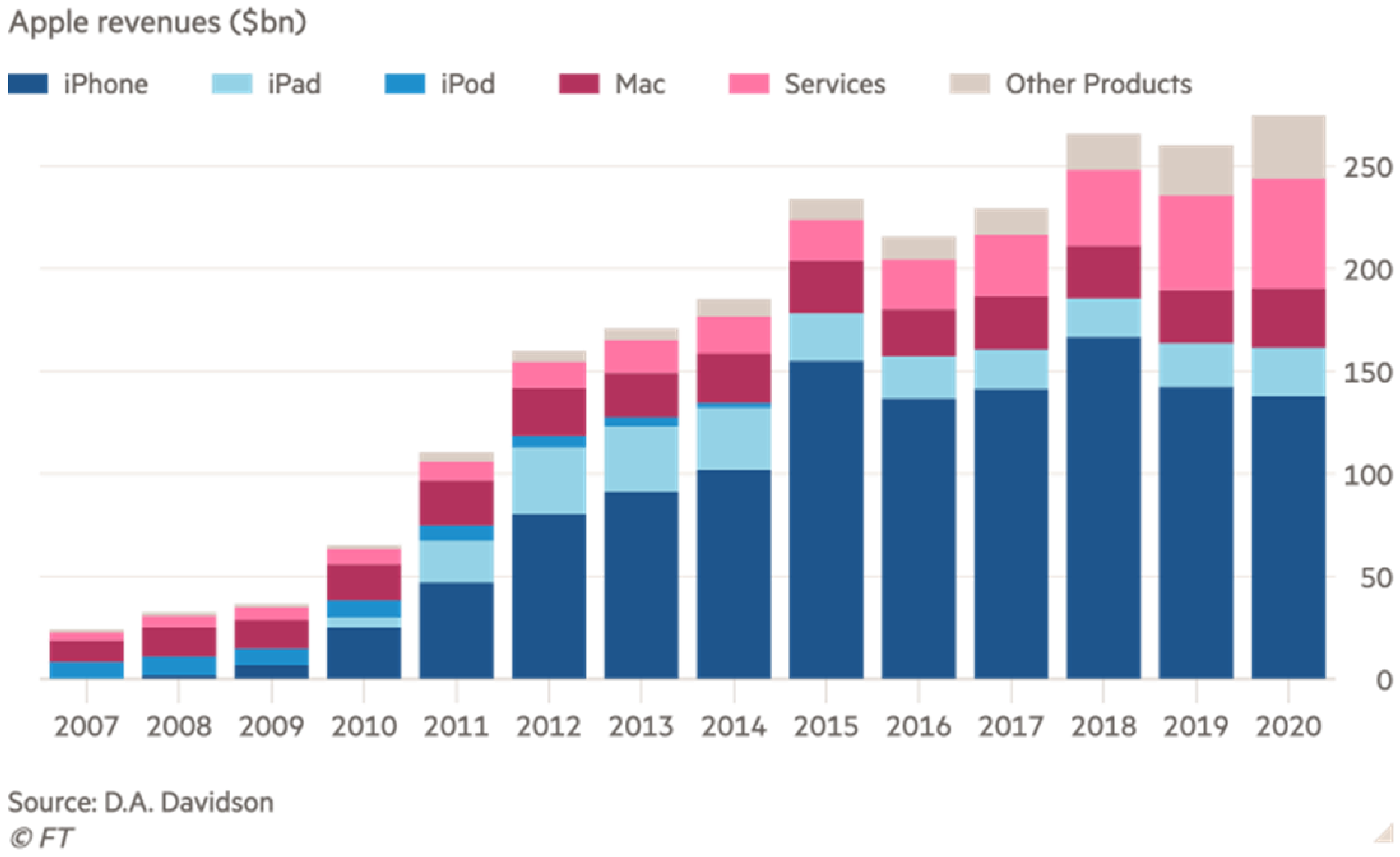
Atlantic Technological University

From Data to Information



DESCRIPTIVE STATISTICS
What does this show?

Graph
Where Apple's Profits come from?



Graphs / Charts make data visual

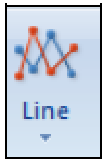
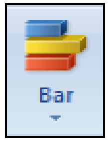
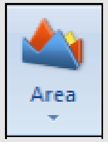
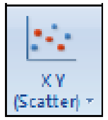
- A chart gets your point across—fast. With a chart, you can transform data to show comparisons, patterns, and trends.

- So instead of having to analyse columns of worksheet numbers in an Excel spreadsheet, you can see at a glance what the data means.

Atlantic Technological University

Excel Charts Types

Table 5-1: Chart Types

	**Column**	Column charts are used when you want to compare different values vertically side-by-side. Each value is represented in the chart by a vertical bar. If there are several series, each series is represented by a different color.
	**Line**	Line charts are used to illustrate trends over time. Each value is plotted as a point on the chart and is connected to other values by a line. Multiple items are plotted using different lines.
	**Pie**	Pie charts are useful for showing values as a percentage of a whole. The values for each item are represented by different colors.
	**Bar**	Bar charts are just like column charts, except they display information in horizontal bars rather than in vertical columns.
	**Area**	Area charts are the same as line charts, except the area beneath the lines is filled with color.
	**XY (Scatter)**	Scatter charts are used to plot clusters of values using single points. Multiple items can be plotted by using different colored points or different point symbols.
	**Other Charts**	Select from Stock, Surface, Doughnut, Bubble, or Radar-type charts. You can also make a combination chart by selecting a different type of chart for only one of the data series.

From Data to Information
Weights of Tablets

- 3.0469
- 3.0555
- 3.0627
- 3.0412
- 3.0473
- 2.9739
- 3.0609
- 3.0023
- 2.9813
- 3.0430
- 3.0467
- 3.0465
- 3.0296
- 3.0352
- 3.0572

- 3.0476
- 3.0140
- 3.0077
- 2.9686
- 3.0354
- 2.9890
- 3.0119
- 3.0370
- 3.0111
- 3.0316
- 3.0774
- 3.0161
- 3.0188
- 3.0339
- 2.9885

- 3.0234
- 3.0749
- 3.0451
- 3.0165
- 3.0283
- 3.0432
- 3.0652
- 3.0331
- 3.0412
- 3.0218
- 3.0568
- 3.0373
- 3.0580
- 3.0918
- 3.0156

- 3.0248
- 3.0753
- 3.0023
- 3.0193
- 3.0680
- 3.0674
- 3.0113
- 3.0469
- 3.0592
- 3.0604
- 3.0258
- 3.0646
- 3.0459
- 3.0532
- 3.0347

- 3.0322
- 3.0289
- 3.0165
- 3.0494
- 3.0399
- 3.0240
- 3.0216
- 3.0480
- 3.0590
- 3.0135
- 3.0469
- 3.0558
- 3.0627
- 3.0412
- 3.0473

MEASURES OF CENTRAL TENDENCY

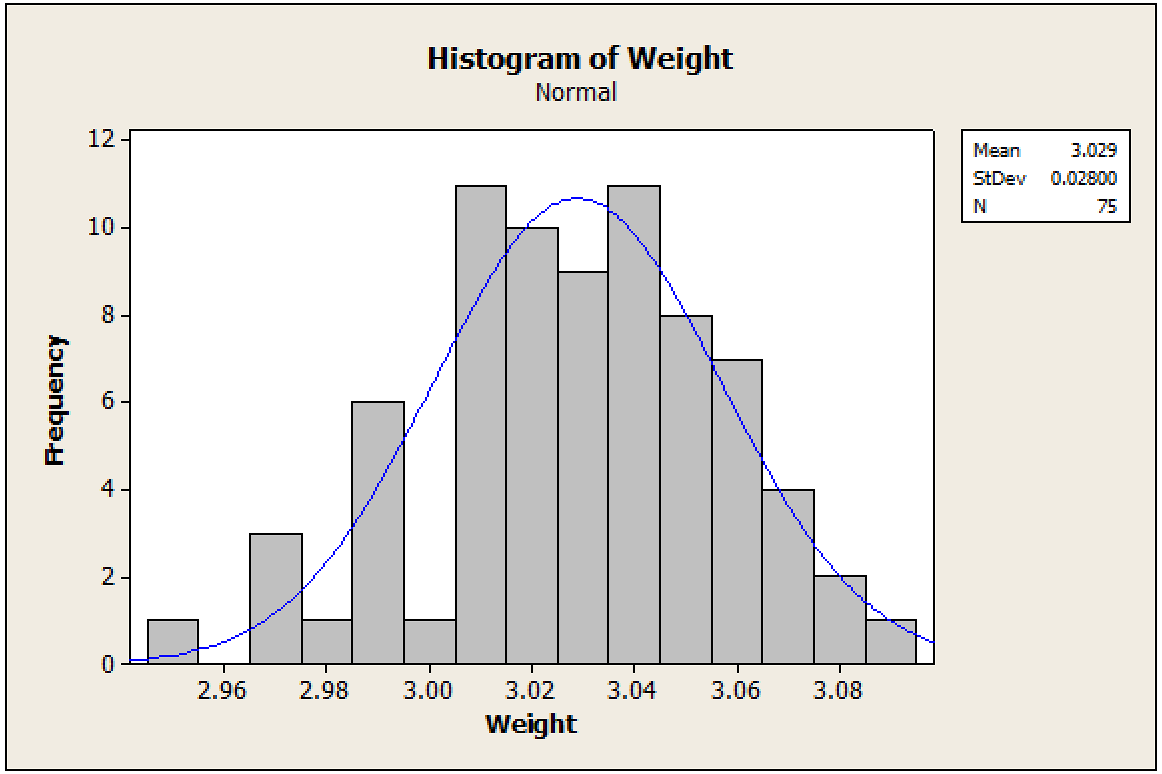
- Mean—arithmetic average
- Median—midpoint in a distribution
- Mode—most frequent score

Descriptive Statistics
Weights of Tablets
Measures of Central Tendency
Measures of Variability

- **N** = 75
- **Mean** = 3.0289 grams
- **Median** = 3.0323
- **Mode** = 3.0412
- **Std. Deviation** = 0.0280
- **Min** = 2.9459
- **Max** = 3.0918
- **Range** = 0.1459

Atlantic Technological University

From Data to Information
Histogram of Weights of Tablets



Atlantic Technological University

MEAN

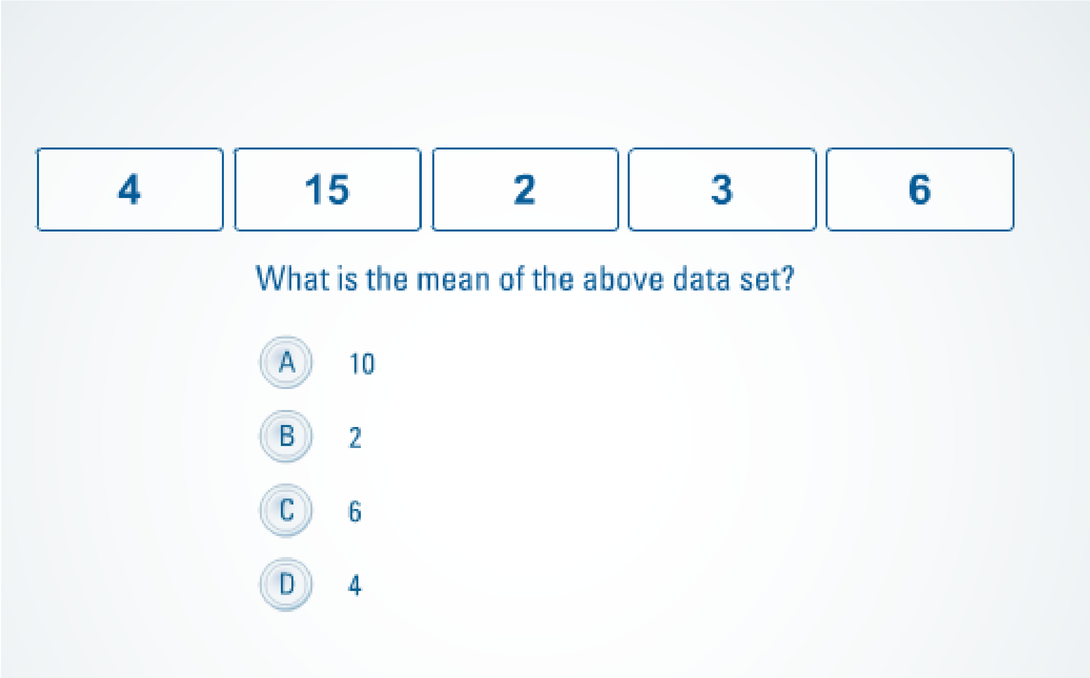
- What it is
  - Arithmetic average
  - Sum of scores/number of scores

- How to compute it
  - $\bar{X} = \frac{\Sigma X}{n}$
    - Σ = summation sign
    - X = each score
    - *n* = size of sample

1. Add up all of the scores
2. Divide the total by the number of scores

Atlantic Technological University

Using Statistics to Analyse Data (Lecture 5)
Quiz - Mean



Answer = 5.5 (6)

Atlantic Technological University

MEDIAN

- What it is
  - Midpoint of distribution
  - Half of scores above & half of scores below

- How to compute it when n is odd
  1. Order scores from lowest to highest
  2. Count number of scores
  3. Select middle score
- How to compute it when n is even
  1. Order scores from lowest to highest
  2. Count number of scores
  3. Compute X of two middle scores

Atlantic Technological University

Example

Processing Time in seconds for a customer phone call to be answered?

32	23	26	26	41	36	26	38
45	36	27	39	42	39	25	33
14	39	24	34	35	23	29	31

What ‘Descriptive Statistics’ would you use to describe this data?

Sort Data

- N = 24
- Min = 14 seconds
- Mode =26 seconds
- Max = 45 seconds
- Range = 31 seconds
- Mean = 31.8 seconds
- Median = 32.5
- Q1 = 26
- Q3 = 38.75
- IQR = 12.75

1	14
2	23
3	23
4	24
5	25
6	26
7	26
8	26
9	27
10	29
11	31
12	32
13	33
14	34
15	35
16	36
17	36
18	38
19	39
20	39
21	39
22	41
23	42
24	45

Atlantic Technological University

Median , Q1 and Q3

- **Median:** (24+1)x 0.5=12.5, so between 12th and 13th values (32 and 33). The median is 32.5.

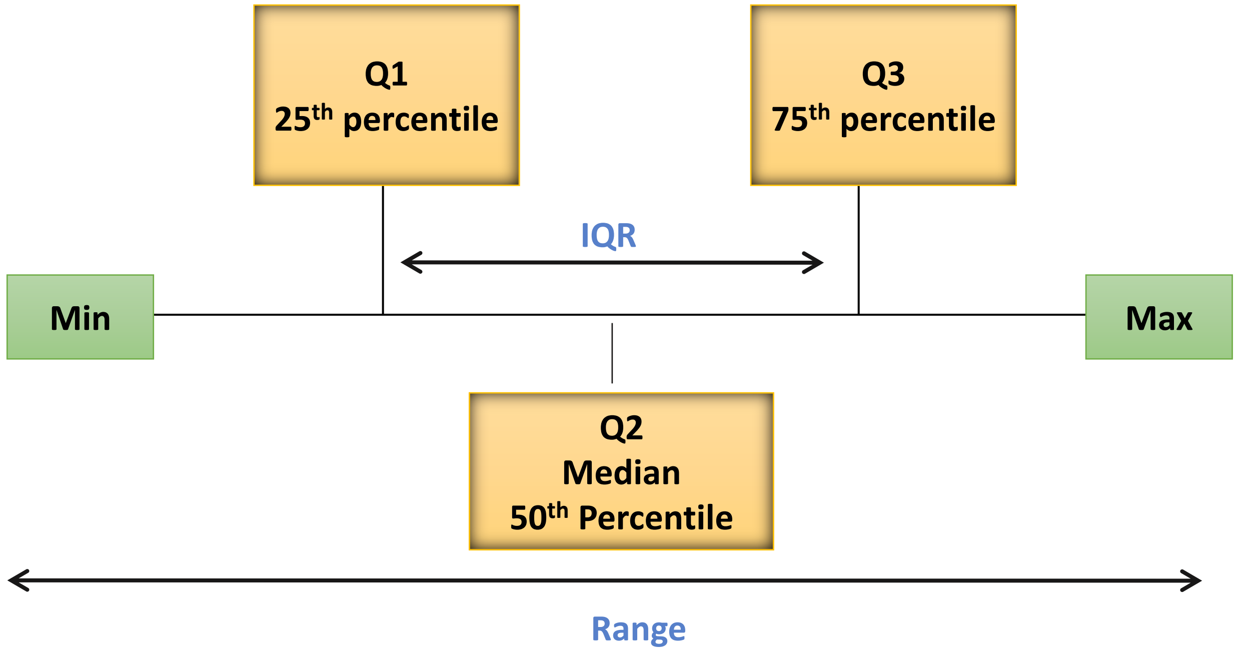
- **Q1 Lower quartile:** $(24+1)\times0.25=6.25^{th}$ observation (6th and 7th values are both 26). The lower quartile is 26.

- **Q3 Upper quartile:** $(24+1)\times0.75=18.75$ or 0.75 of the way from the 18th to the 19th value. The upper quartile is 38.75.

Atlantic Technological University

Inter Quartile Range

- Inter quartile range **(IQR)** is another range measure but this time looks at the data in terms of quarters or percentiles.
- The range of data is divided into four equal percentiles or quarters (25%).



IQR

- IQR is the range of the middle 50% of the data. It is often seen as a better measure of spread than the **range** as it is not affected by outliers i.e. because it uses the middle 50%, it is not affected by outliers or extreme values.

- Outliers – variables that are the extreme lower or upper end of the distribution. They are atypical, infrequent observations.

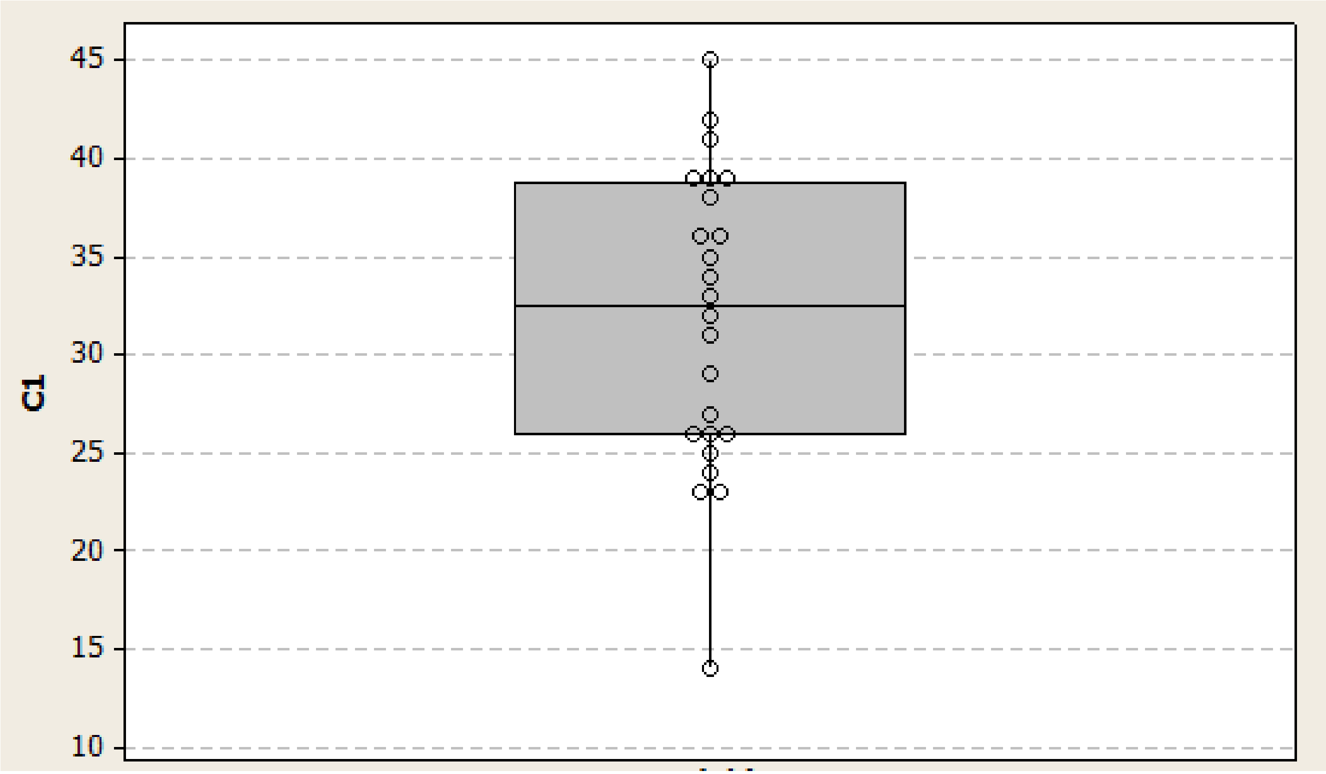
- These will influence the mean (arithmetic). Why?

10 people record their height: *160, 162, 164, 166, 168, 170, 172, 174, 176 and 200 cm tall. With those values the mean is 171cm. 200cm is the outlier – take it out and the median is 168cm.*

Atlantic Technological University

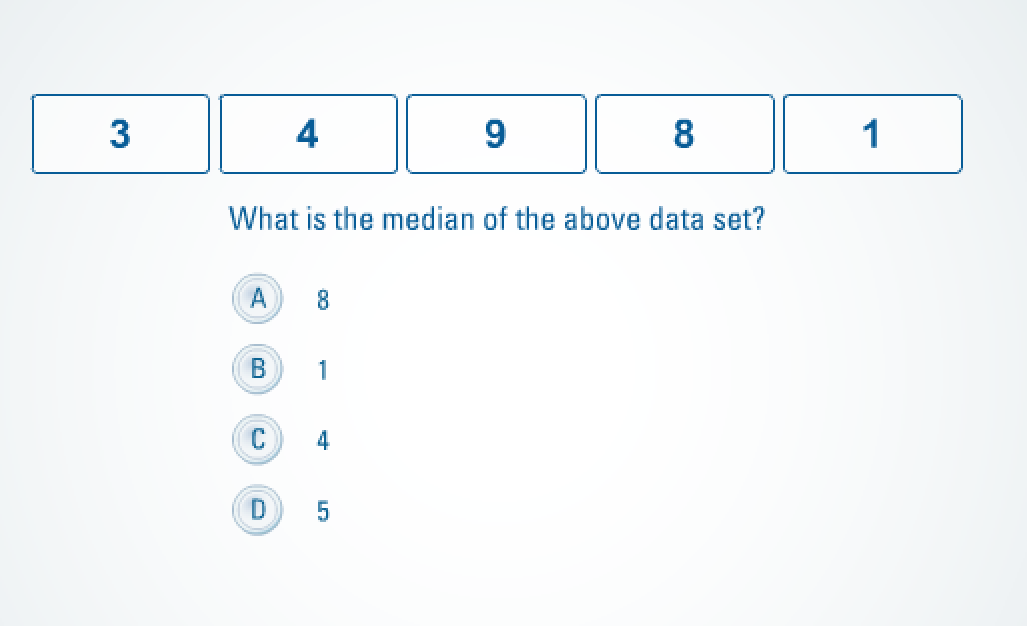
Box Plot of Call Centre Answering Times

- Median = 32.5
- Q1 = 26
- Q3 = 38.75
- IQR = 12.75



Atlantic Technological University

Using Statistics to Analyse Data
Quiz - Median



Answer = 4 i.e. the middle value

MODE

- What it is
  - Most frequently occurring score

- What it is not!
  - How often the most frequent score occurs

Atlantic Technological University

WHEN TO USE WHICH MEASURE

Measure of Central Tendency	Level of Measurement	Use When	Examples
Mode	Nominal	Data are categorical	Eye color, party affiliation
Median	Ordinal	Data include extreme scores / outliers	Salary levels in a company, Age
Mean	Interval and ratio	Quantitative Data	Time, Weight, Length

Measures of Variability / Dispersion

- Measures of dispersion: statistical measures that summarise the amount of spread or variation in the distribution of values in a variable.

- So, how values are spread within a distribution.

	Description
Range	Difference between the highest (maximum) and lowest (minimum) value in the distribution of values
Variance	The measure of the spread.
Standard deviation	Shows the relation that a set of data has to the mean of the sample data.

Range

- Range is simply the difference between the highest and lowest value in the distribution of values.

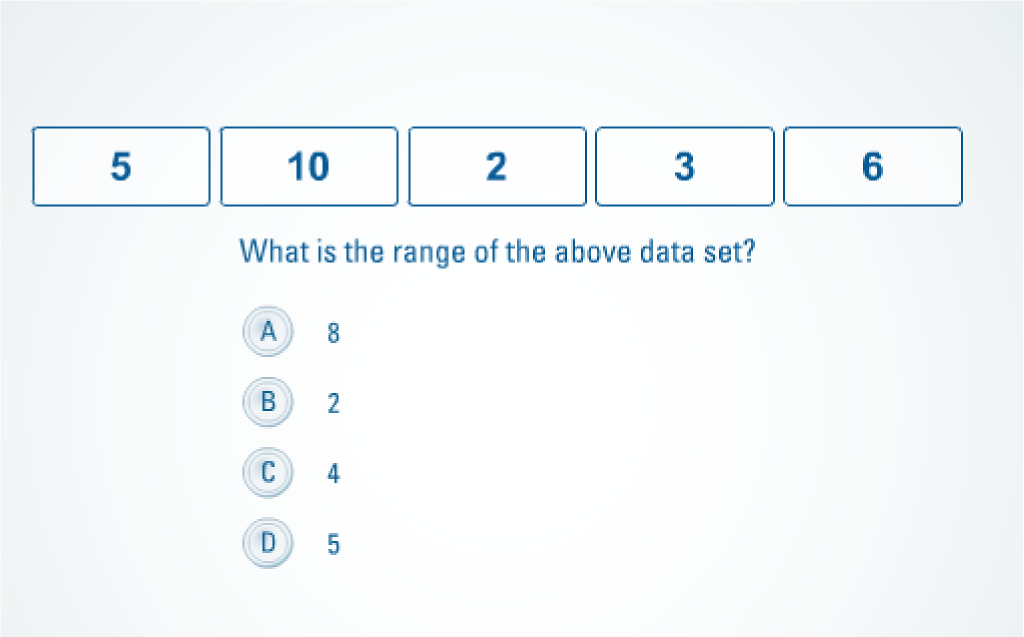
Example:

Weekly income of 10 people:

€180	€220	€280	€320	€280	€180	€350	€280	€330	€220

Range is maximum income minus minimum income: 350-180 = €170.

Using Statistics to Analyse Data
Quiz - Range



Variance

- Where the mean is a measure of the centre of a group of numbers, the variance is the measure of the spread.
- It involves measuring the distance between each of the values and the mean.
- To calculate the variance :
  1. calculate the mean
  2. for each value in the distribution subtract the mean and then square the result (the squared difference)
  3. calculate the average of those squared differences.

Note: Why do we Square the result? If you just subtracted the mean from each value in the distribution you would end up with positive numbers for those values above the mean and negative numbers for those below the mean. Add these results together and they would cancel each other out giving you a variance of zero. As multiplying 2 negative numbers together gives a positive, we square each result. By squaring them, you make all the deviations positive and they can add up.
In addition, squaring gives much larger weight to large numbers (positive or negative) than to numbers close to zero: 50^2=2,500 is bigger than 25^2=625.

Variance

$$s^2 = \frac{\sum\left(X_i - \bar{X}\right)^2}{n-1}$$

= Sum of (observed value – mean score) ^{2}

Total number of scores -1

- The larger the variance value the further the observed values of the data set are dispersed from the mean.
- A variance value of zero means all observed values are the same as the mean.

Atlantic Technological University

Standard deviation

- Standard deviation: how far on average each value is from the mean.
- Problem with Variance: because the differences are squared, the units of Variance are not the same as the units of the data.
- This can make interpretation of the results problematic.
- If the Variance is square rooted, the units of Variance then correspond to those of the data set.
- This square rooting of the Variance is reported as the Standard Deviation.

Standard deviation

- So, in most disciplines, standard deviation is used more frequently than variance.

- Standard deviation scores are used to generate standardised or z scores.

Atlantic Technological University

Standard deviation

- Standard deviation = The square root of the variance.

$$s = \sqrt{\frac{\sum\left(X - \bar{X}\right)^2}{n-1}}$$

- Σ = summation sign
- X = each score
- X = mean
- *n* = size of sample

- As it is square rooted the results correspond to the original data units. E.g. if the variable is height recorded in cm then the standard deviation can be interpreted as cm.

COMPUTING THE STANDARD DEVIATION

X
13
14
15
12
13
14
13
16
15
9
$\overline{X} = 13.4$

1. List scores and compute mean

COMPUTING THE STANDARD DEVIATION

X	$(X-\bar{X})$
13	-0.4
14	0.6
15	1.6
12	-1.4
13	-0.4
14	0.6
13	-0.4
16	2.6
15	1.6
9	-4.4
$\bar{X} = 13.4$	$\sum X = 0$

1. List scores and compute mean
2. Subtract mean from each score

Atlantic Technological University

COMPUTING THE STANDARD DEVIATION

X	$(X - \overline{X})$	$(X - \overline{X})^2$
13	-0.4	0.16
14	0.6	0.36
15	1.6	2.56
12	-1.4	1.96
13	-0.4	0.16
14	0.6	0.36
13	-0.4	0.16
16	2.6	6.76
15	1.6	2.56
9	-4.4	19.36
$\overline{X}$ =13.4	$\sum X = 0$	

1. List scores and compute mean
2. Subtract mean from each score
3. Square each deviation

Atlantic Technological University

COMPUTING THE STANDARD DEVIATION

X	$(X - \overline{X})$	$(X - \overline{X})^2$
13	-0.4	0.16
14	0.6	0.36
15	1.6	2.56
12	-1.4	1.96
13	-0.4	0.16
14	0.6	0.36
13	-0.4	0.16
16	2.6	6.76
15	1.6	2.56
9	-4.4	19.36
$\overline{X} = 13.4$	$\sum X = 0$	$\sum X^2 = 34.4$

1. List scores and compute mean
2. Subtract mean from each score
3. Square each deviation
4. Sum squared deviations

Atlantic Technological University

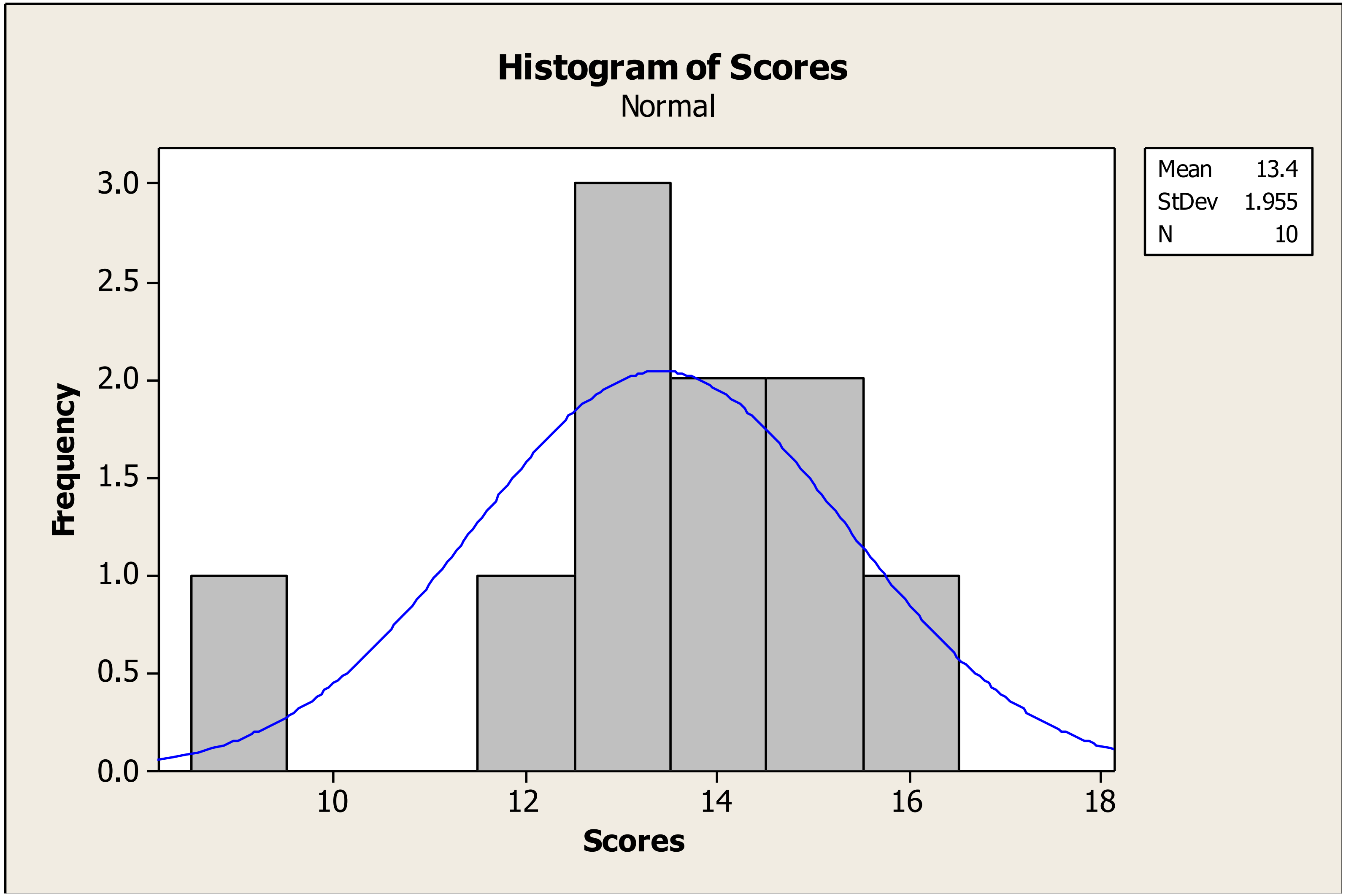
COMPUTING THE STANDARD DEVIATION

X	$(X - \bar{X})$	$(X - \bar{X})^2$
13	-0.4	0.16
14	0.6	0.36
15	1.6	2.56
12	-1.4	1.96
13	-0.4	0.16
14	0.6	0.36
13	-0.4	0.16
16	2.6	6.76
15	1.6	2.56
9	-4.4	19.36
$\bar{X}$ =13.4	$\sum X = 0$	$\sum X^2 = 34.4$

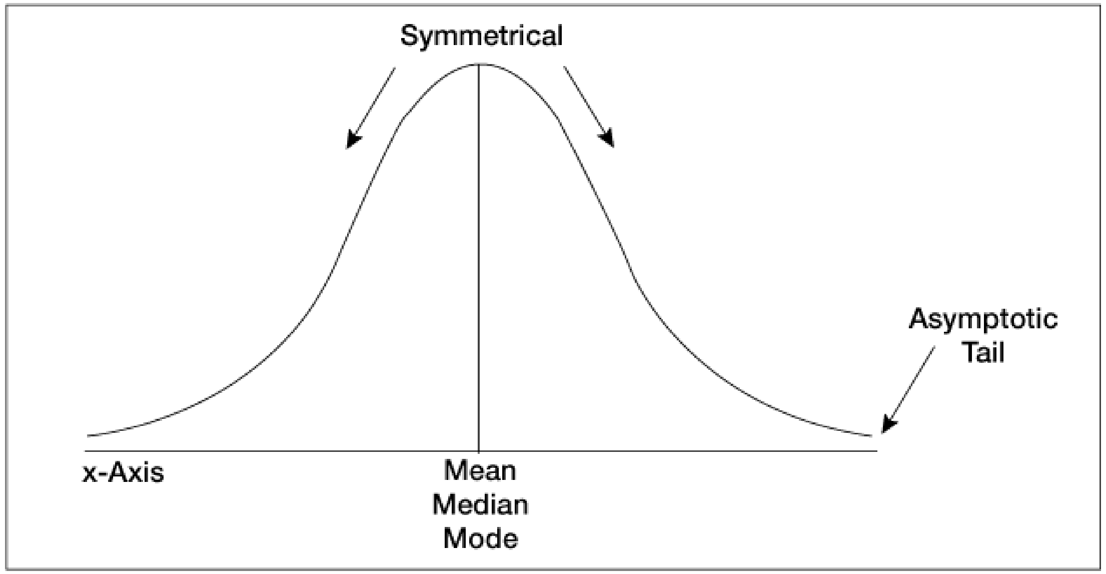
1. List scores and compute mean
2. Subtract mean from each score
3. Square each deviation
4. Sum squared deviations
5. Divide sum of squared deviation by $n - 1$
   - 34.4/9 = 3.82 (= s^2)
6. Compute square root of step 5
   - $\sqrt{3.82}$ = 1.955

Atlantic Technological University

Histogram of Data



THE NORMAL (BELL SHAPED) CURVE

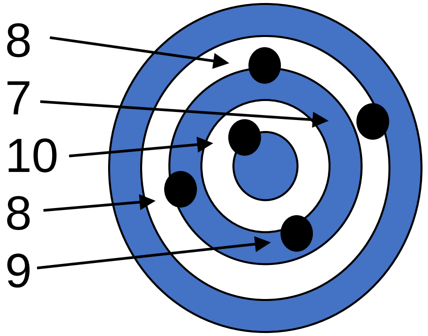


- Mean = median = mode
- Symmetrical about midpoint
- Tails approach X axis, but do not touch

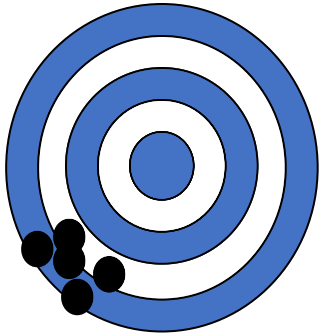
Example
Variance & Standard Deviation

Variability - Emmett

- Who is better – Emmett or Jake?
- Deviation = distance between observations and the mean (or average)



Emmett



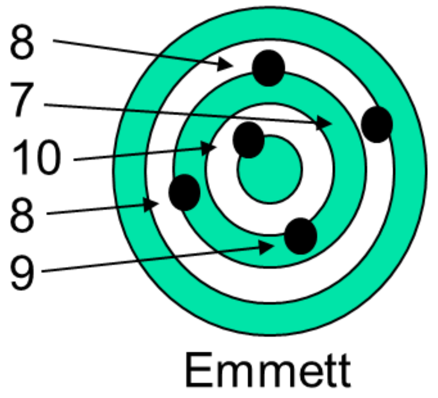
Jake

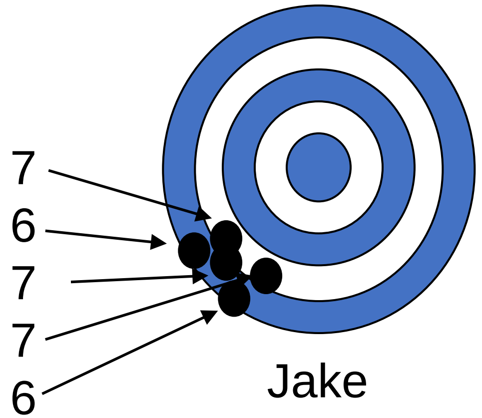
Emmett	Observations	Deviations
	10	10 - 8.4 = 1.6
	9	9 – 8.4 = 0.6
	8	8 – 8.4 = -0.4
	8	8 – 8.4 = -0.4
	7	7 – 8.4 = -1.4
averages	**8.4**	**0.0**

Atlantic Technological University

Variability - Jake

- Who is better – Emmett or Jake?
- Deviation = distance between observations and the mean (or average)



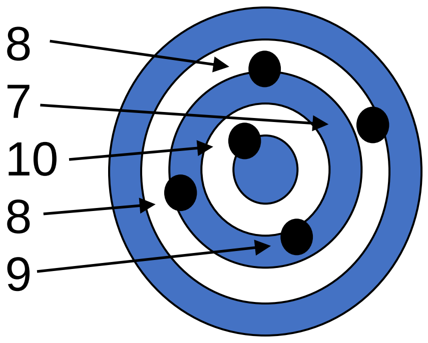


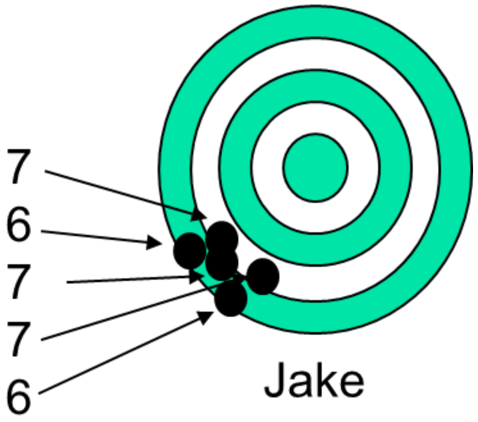
Jake	Observations	Deviations
	7	7 – 6.6 = 0.4
	7	7 – 6.6 = 0.4
	7	7 – 6.6 = 0.4
	6	6 – 6.6 = -0.6
	6	6 – 6.6 = -0.6
averages	**6.6**	**0.0**

Atlantic Technological University

Variance - Emmett

- Variance = average distance between observations and the mean squared



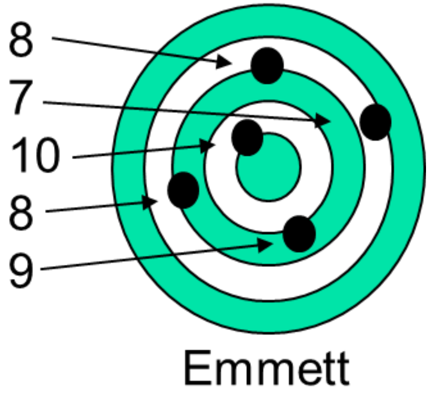


Emmett	Observations	Deviations	Squared Deviations
	10	10 - 8.4 = 1.6	2.56
	9	9 – 8.4 = 0.6	0.36
	8	8 – 8.4 = -0.4	0.16
	8	8 – 8.4 = -0.4	0.16
	7	7 – 8.4 = -1.4	1.96
averages	**8.4**	**0.0**	**1.04**

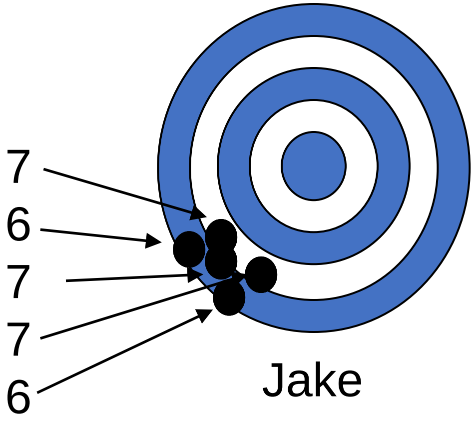
Variance

Variance - Jake

- Variance = average distance between observations and the mean squared



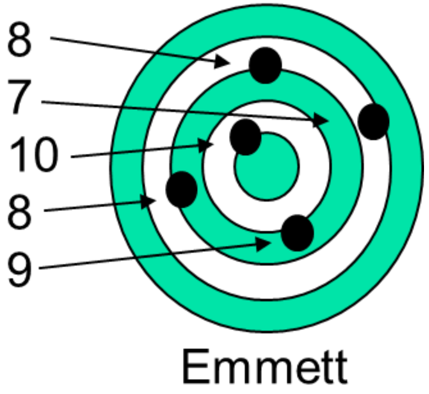
Jake	Observations	Deviations	Squared Deviations
	7		
	7		
	7		
	6		
	6		
averages			

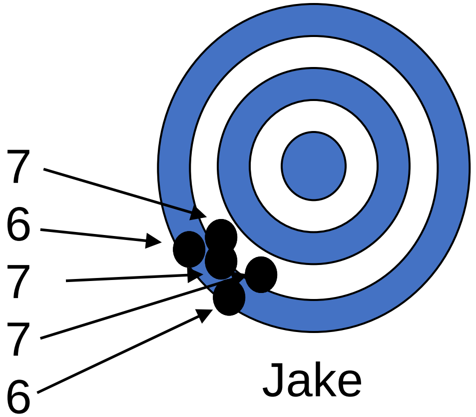


Atlantic Technological University

Variance - Jake

- Variance = average distance between observations and the mean squared



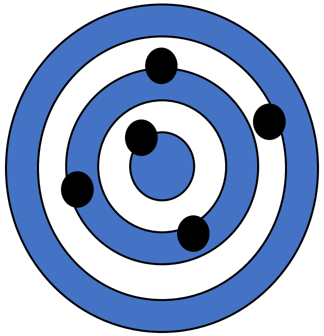


	Observations	Deviations	Squared Deviations
	7	7 - 6.6 = 0.4	0.16
	7	7 - 6.6 = 0.4	0.16
	7	7 - 6.6 = 0.4	0.16
	6	6 – 6.6 = -0.6	0.36
	6	6 – 6.6 = -0.6	0.36
averages	**6.6**	**0.0**	**0.24**

Variance

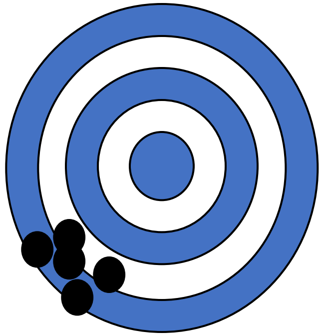
Standard Deviation

- Standard deviation = square root of variance



Emmett

	Variance	Standard Deviation
Emmett	1.04	1.02
Jake	0.24	0.4898979



Jake

Note: Assume above data is the whole Population so Divide by n for Std. Deviation.

Importance of Measurement

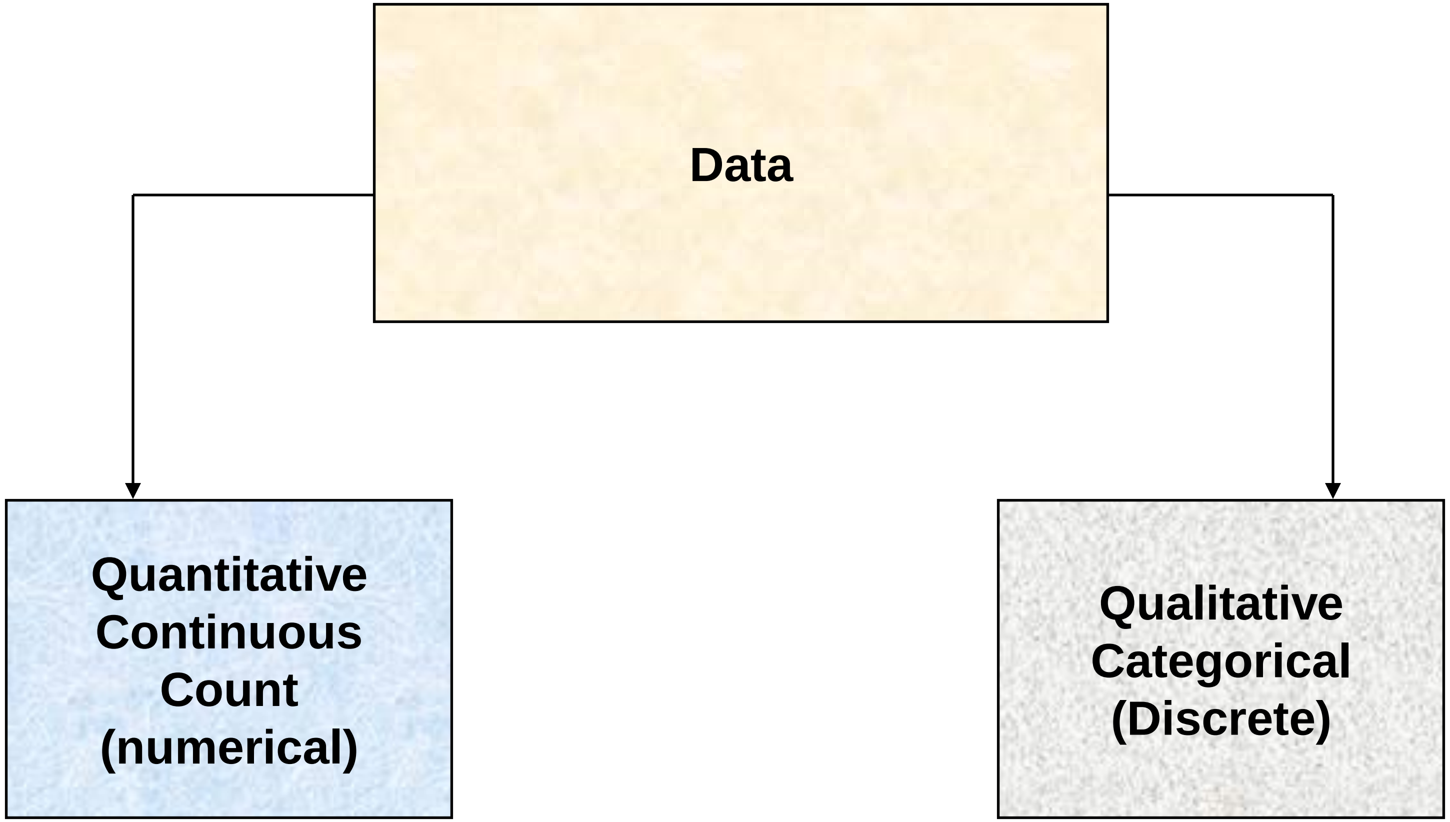
As you begin to understand your processes in terms of how the customer sees them, you need to measure the characteristics that help “predict” customer satisfaction.

- *You cannot define as a process what you do not understand....*
- *You cannot measure what you cannot define as a process.....*
- *You cannot manage what you cannot measure.....*

Be Data Driven !

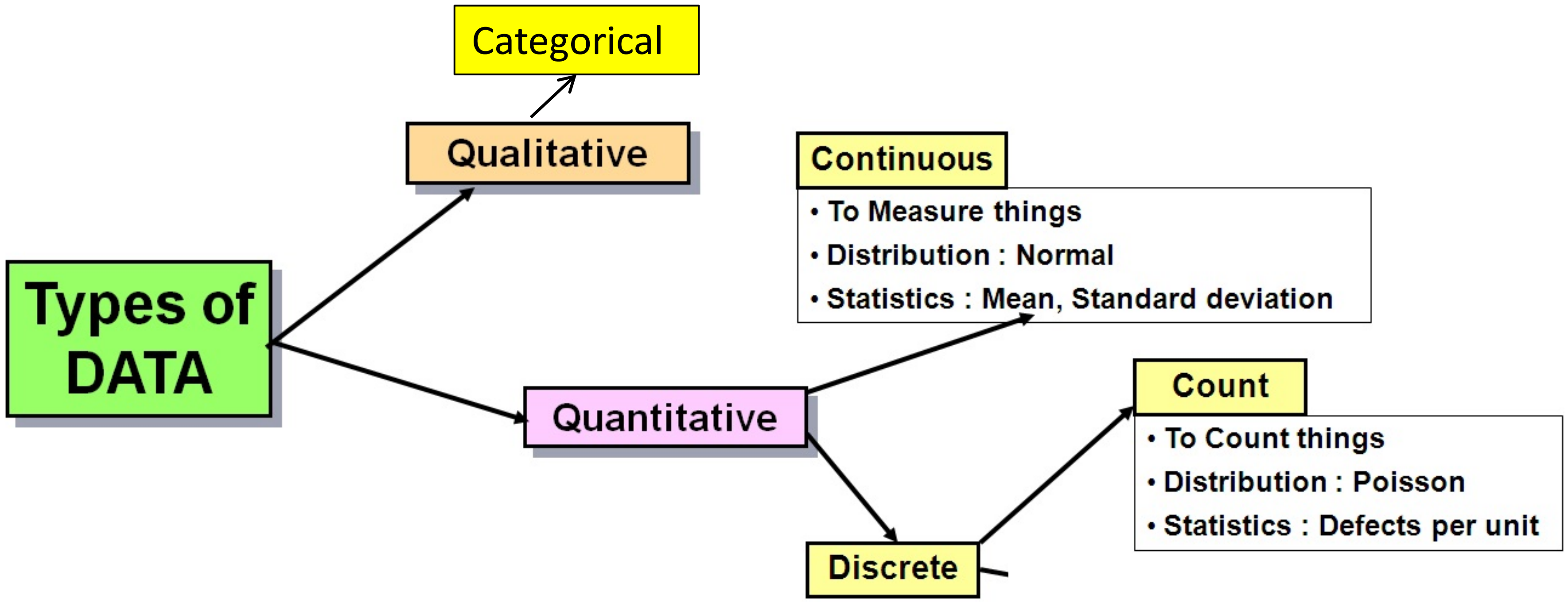
"If you don't have data as backup, you are just another person with an opinion."

Data Categories



Atlantic Technological University

Types of Data



Data Types - Quantitative

- **Continuous measurement** – sometimes called **quantitative** or **variable data** – these are the answer to the question “how much?” and are a quantity (number).

- Continuous variables can usually take on any number value, although this may be restricted to certain intervals.

- Some examples are the exact weight of a quarter pound hamburger, the length of a part, the time taken to complete a task etc.

Data Types - Qualitative

- **Categorical measurements** – sometimes called **qualitative** or **attribute data** – these are variables associated with some category or quality of the object being measured.
- Some examples are:
  - male/female,
  - defective (coded as 1) or good (0),
  - response scales (where 1 might be total dissatisfaction and 5 total satisfaction),
  - year at university (first year, second year etc)
  - income bracket (0 to less than €10000, €10000 to less than €20000 etc).

End of Lecture

Atlantic Technological University