**Week 2 – Notes 1/3 (Tuesday)**

**Statistics - Terminology**

**Population** = All data information

**Sample** = A sample of the population (a subset from the population)]

**Qualitative =** Attributes, non-numerical

**Quantitative =** Numericals

**Quantification =** Turning quantitative to numerical

**Descriptive =** Describing a dataset. “A picture of the data”

**Central tendency =** Mean, Median, Mode

Mean = Average / Median = Central number / Median = Most common number

**Variability =** Range (max-min), Variance (Standard Deviation), Kurtosis

Standard Deviation = How far a group of numbers is from mean (squared root of variance) / **Variance** = Average of the square differences from the mean.

**Inferential =** Inferring (Drawing) conclusions from a dataset

Parametric = Distribution is known / Non-parametic: no assumption

**Kinds of data:** # Important. We need to know which Data are dealing with to define the problem

**Numerical =** Made of numbers

Continuous = Infinite options > Age, weight, blood, pressure

Discrete =Finite options > Shore size, number of children

**Categorical =** Made of words

Ordinal = The Data has a hierarchy> Pain severity, satisfaction rating

Nominal = The Data has no hierarchy> Eye colour, dog breed, blood type

Probability distribution = Mathematical function that describes the probability of different possibles values of a variable.

Graphical user interface

Description automatically generated

Graphical user interface, text

Description automatically generated

Continuous variables

Text

Description automatically generated with medium confidence

Graphical user interface, application

Description automatically generated

Graphical user interface, text

Description automatically generated

Chart

Description automatically generated

Text

Description automatically generated

Covariance and correlation important.

Reverse correlation

Graphical user interface, website

Description automatically generated