INTRODUCTION TO STATISTICS
NOTES

UNIT 1

Statistics is a branch of math that deals with the collection, organization and analysis and interpretation of numerical data.

Statistic is specifically used to drawing general comclsuions about a population based on information from a sample.

A **POPULATION** is an entire group from which we want information. ALL is used to represent a population.

 The area on which the reasearch is been held

A **SAMPLE** is data collected from a SUBSET of elements in the population. They make an infrence from the population to get the sample.

A PARAMETER is <u>the value calculated from the population</u>

A STATISTIC is the <u>value calculated from a sample</u>

INFERENCE the use of sample information to make conclusions about a population.

CENSUS this is an attempt to collect information from **ALL** elements in the population

Quartiles special part of percentiles... it divides the data sept into 4 parts... 25% =Q1

50th percentile = Median =M

75TH percentile = Q3

→ interquartile range is the lower quartile less the upper quartile. (Q3 -Q1) the distance between the two. \

In trying to find these quartiles the easiest way is to find the **median first then the Q1 then Q3**

Five number summary always appear in this form... the median and quartiles are resistant measures. The 5-number summary is good for describing distributions, in particular non-symmetric ones.

Minimum values

Q1→ median for lower quartile

M→ median

Q3→ upper quartile

Maximum values

The total are under the Normal curve = 1

The larger the standard deviation the wider the spread of the variance.

Role of Mu and Sigma

The centre of the distribution is the population mean (Mu)

Spread is determined by sigma = standard deviation.

The large the standard deviation or variance median the larger the curve

EMPIRICAL RULE

If you have normally distributed with mean Mu and standard deviation.

1. The standard deviation of the mean

- Approximately 68% of the data values will be between Mu – sigma and Mu + sigma

$$P(\mu\text{-}1\sigma < X < \mu\text{+}1\sigma) = 0.68 \text{ or Mu + or - sigma}$$

To estimate the value of **the population parameter**

The hypothesis test

Test the evidence provided by the collected data that the value of the parameter supports some claims (given a certain significance level)

How to construct the confidence interval.

1. Obtain the point estimate.
2. Get critical value from statistical table
3. Calculate the standard error. This is just the standard deviation from the sampling distribution for the staticstic of interest.
4. Complete the calculation using the formula

Point estimate + or – (critical value) * (*standard deviation of the point estimate* / standard error)

The margin of error

Critical value * standard error

If p* is not given then assume it = 0.5

If you are looking for margin of error:

M= z* ASR p(1-p) / n

If you are looking for n

N= (z* / m) all squared p(1-p)

	Two tailed	**One tail upper**	**One tail lower**
H0	: p = p0	: p = p0	p = p0