WEEK 6

POPULATION AND SAMPLE

SAMPLE VS POPULATION

Mostly we are not able to study a whole population of individuals.

- Time, economical considerations, practical reasons etc.

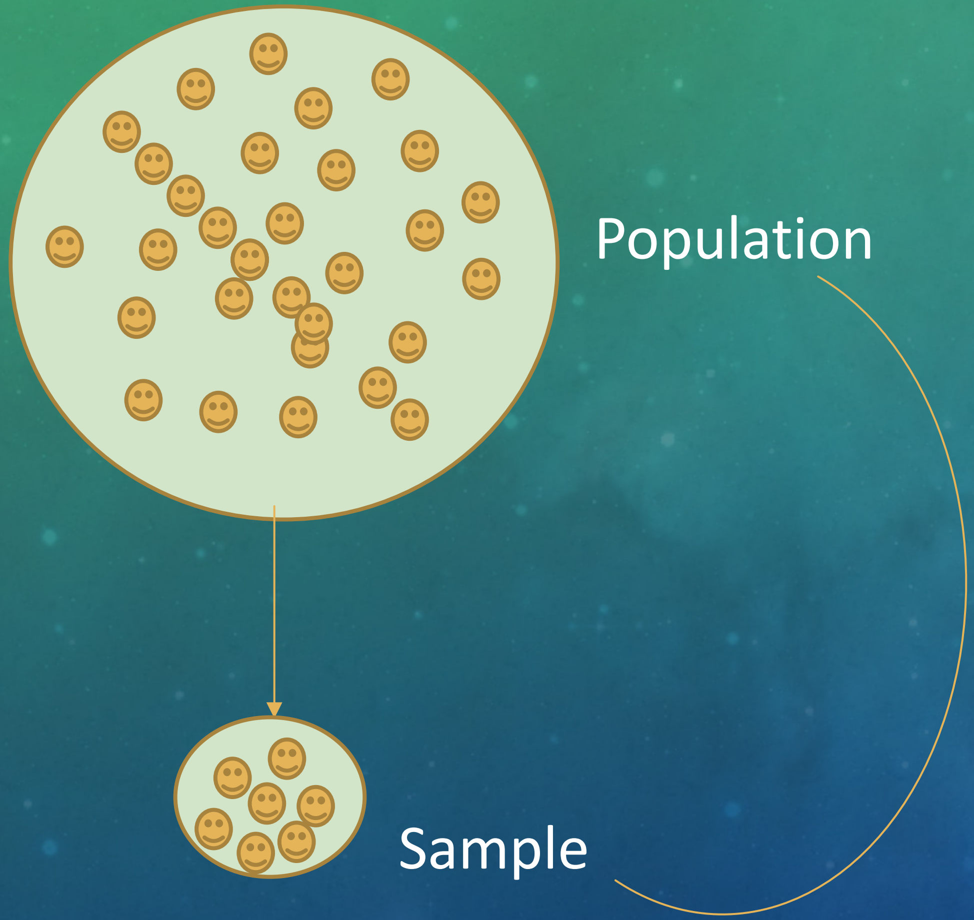
Solution?

Take a representative sample

Side effect !

There will always be a uncertainty about the population with the conclusions we made.

The goal of data analysis is to make the strongest possible conclusions from limited amouns of data.



The process of generalizing to the population from sample :

<u>Statistical inference</u>

SAMPLING ERROR

- Since we are only looking at a sample (rather than the population) there is always likely to be an error in sample estimate.

⇓

So, we need to establish the precision of the sample statistics as an estimate of population parameter

⇓

Standard error of the estimate

STANDARD ERROR OF THE ESTIMATE

A sample mean differs from the population mean depends on :

- Sample size
- Observation variability

$$(SE)S_{\bar{X}} = \frac{Standard\ deviation}{\sqrt{n}}$$

WHAT ARE “STANDARD DEVIATION AND STANDARD ERROR OF MEAN” FOR ?

- Standard deviation
  - Gives an indication of how close the observations are to their mean.
  - Average measure of deviation of each observation in dataset from mean

- Standard Error of Mean
  - Evaluates the sampling error by giving an indication of how close a sample mean is to the population mean.

CONFIDENCE INTERVAL OF MEAN

- How good is our estimate?

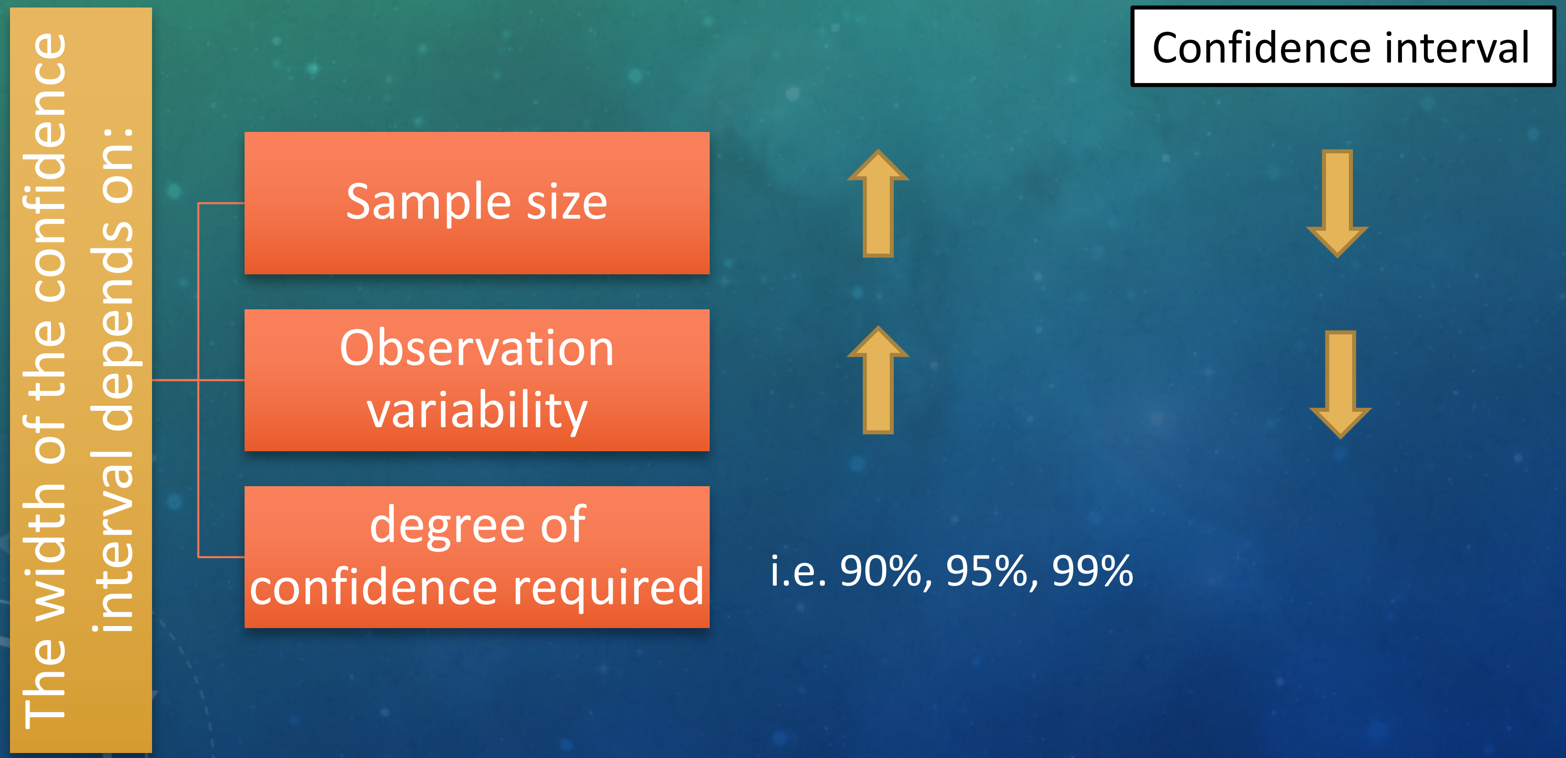
Confidence interval of mean
Upper limit
Lower limit

→ Interpreted as the range of values within which we expect the true population mean to lie with a certain probability

If the confidence interval is wide, then the sample mean is a poor estimate of the population mean.
If the confidence interval is narrow, then the sample mean is a good estimate, i.e. it is a precise estimate of the population mean.

CONFIDENCE INTERVAL OF MEAN

- If we have a 95% confidence interval for the mean, then we say that we are 95% certain that the interval contains the true population value within this interval.



HOW TO CALCULATE CONFIDENCE INTERVAL?

If we know the population std. deviation

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} = \left(\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} \right)$$

If we don't know the population std. deviation

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

$$\bar{x} \pm t_{0.05} \frac{s}{\sqrt{n}} = \left(\bar{x} - t_{0.05} \frac{s}{\sqrt{n}}, \bar{x} + t_{0.05} \frac{s}{\sqrt{n}} \right)$$