

Using Inferential statistics

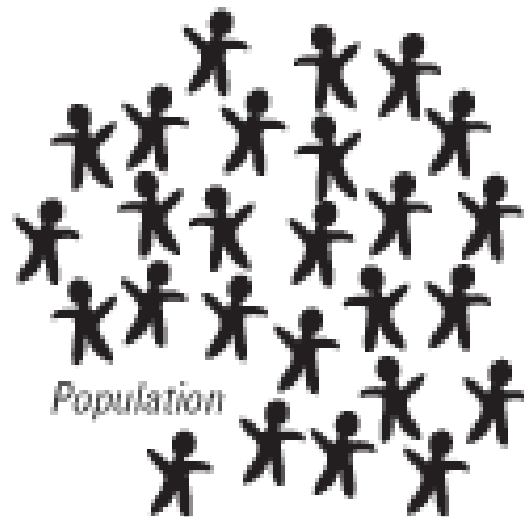
Christine Davies

Key types of data

Data type	description
'Scale', or 'Interval/ratio' data	Numerical (quantitative) data with clear, fixed intervals (eg. 1,2,3).
Ordinal data	Data which can be rank ordered, but with differences that are not clear/fixed (eg. once a day; agree strongly)
Nominal data	Data which describes categories (categorical/ qualitative data) (eg. red, green, blue).

inferential statistics: inferring
conclusions from quantitative data

We want to know about these



Population



Parameter

μ

(Population mean)

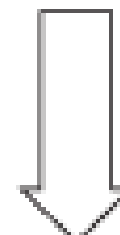
*Random
selection*



We have these to work with



Sample



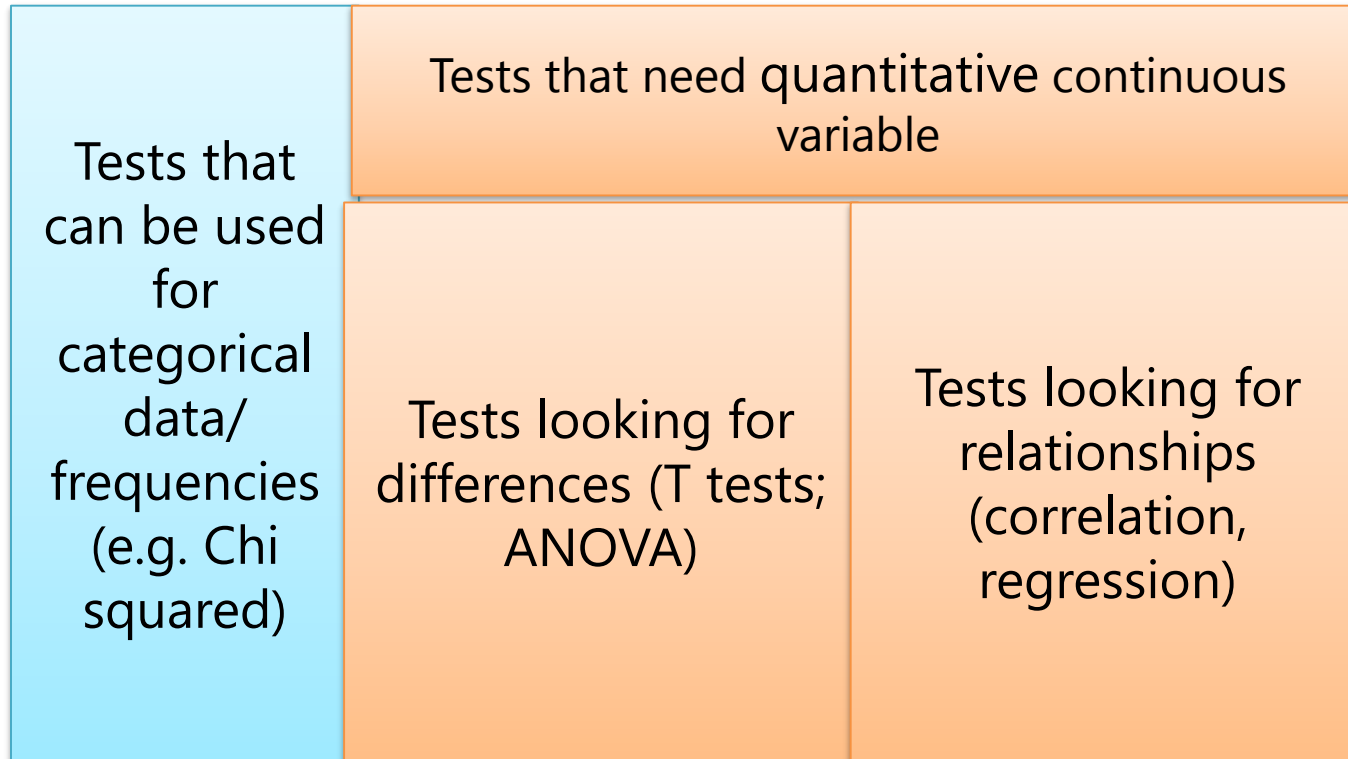
Inference

$\bar{x}$

Statistic

(Sample mean)

Key Types of Statistical test



The process of carrying out a statistical test:

Use sample
to calculate
a 'test
statistic'

Find p value*

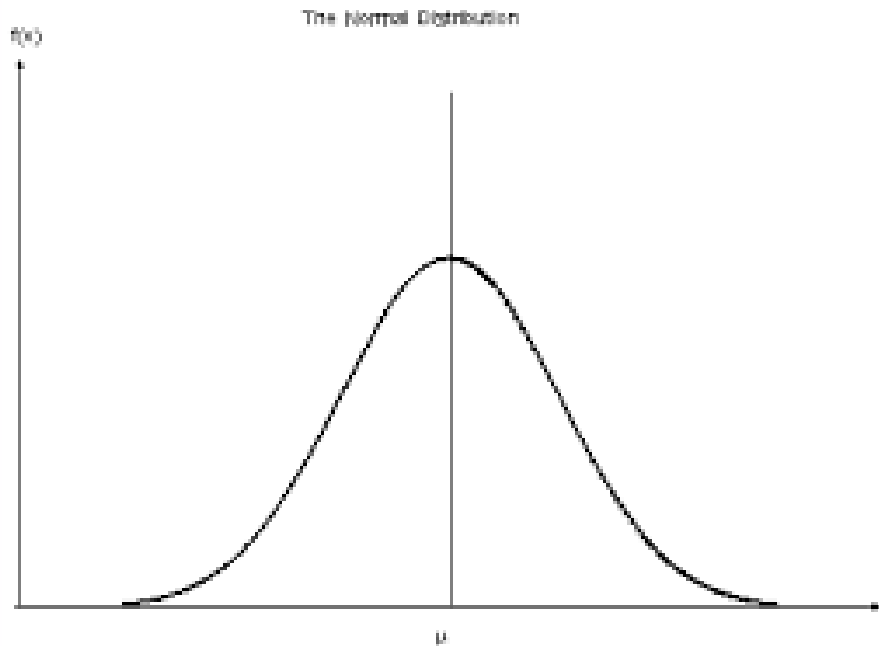
Decide
significance
level (α)

Decide to
accept or
reject null
hypothesis

*** SPSS/Excel will do this for you!**

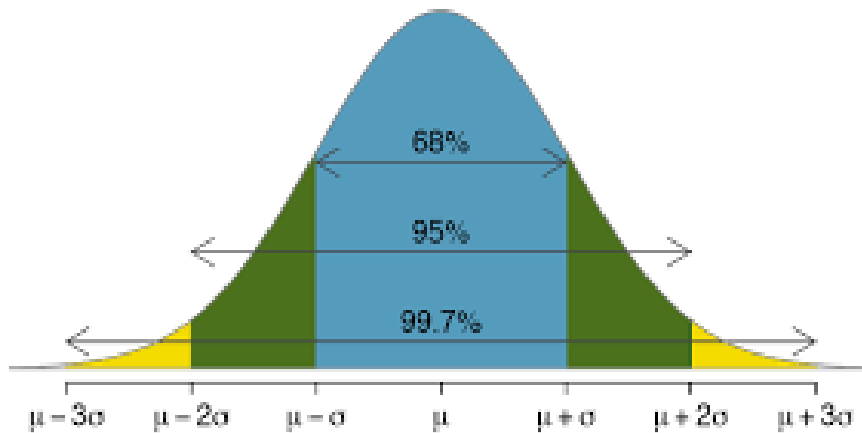
The normal distribution, and how it leads to 'p values'

The normal distribution



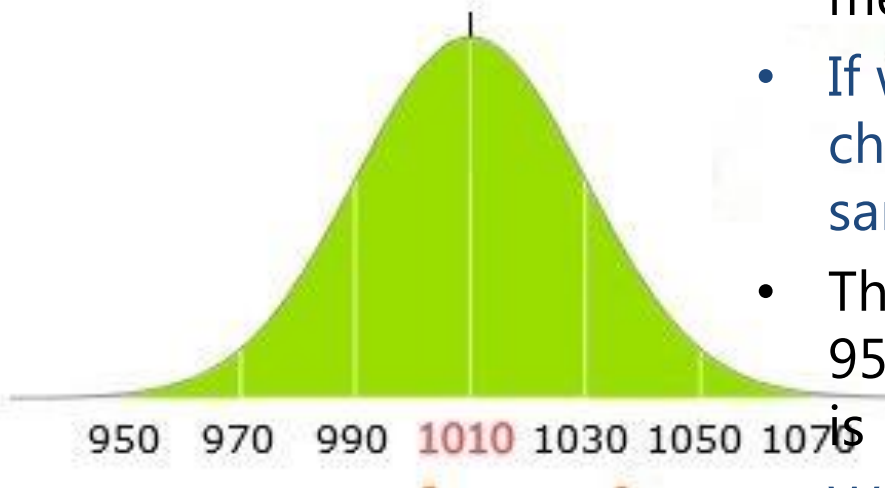
- This 'bell-shaped' curve shows the population of a particular value (here, termed 'x'), eg. height of children in a school; weight of patients in a hospital
- Most values (of x) are clustered around the central value
- The central value is the population mean, μ

Normal distribution, & standard deviation



95% of all values are found within the central part of the distribution with approx ± 2 standard deviations

Example:



- This represents a population with a normal distribution (eg. height of children aged 6-10, in mm, in a particular town)
- The population mean, $\mu = 1010$
- Most heights are clustered around the mean
- If we took many random samples of children's heights, we would find 95% of sample means between 970 and 1050
- The probability of a very low value (eg. 950) belonging to this population is low: it is less than 5%
- We might conclude that a child with height 950mm (or sample with this mean) is significantly different from this population

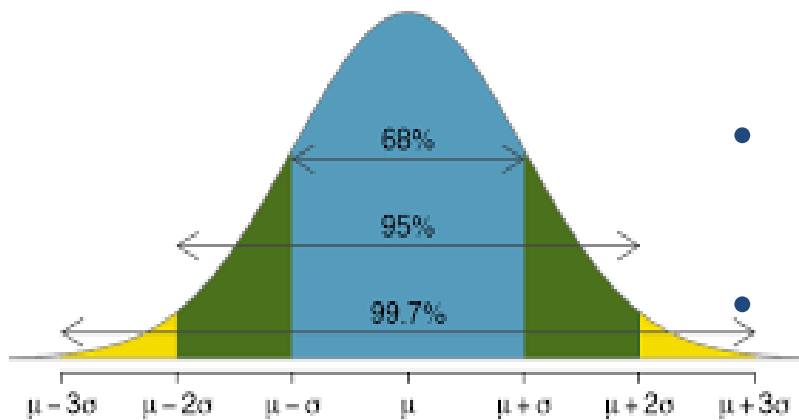
We can use this idea to infer whether samples are likely to come from particular populations

EG:



- You have bought a pack of Welsh-cakes that looks extremely 'spotty', ie. with a high raisin content 😊
- You form a hypothesis that it has a significantly different raisin content compared to a typical pack
- How do you prove or disprove your hypothesis?

Example



- You find the mean raisin content of your Welsh-cakes (total number, divided by no. of cakes)
- You compare your mean with a typical raisin distribution
- If it falls within the 'extreme' ends (ie. 5% at the edges), we would say that the probability of belonging to this population is <0.05 (ie. Probability, **$p < 0.05$**)
- This means that it is **significantly different**: in our example, the raisin content is significantly higher

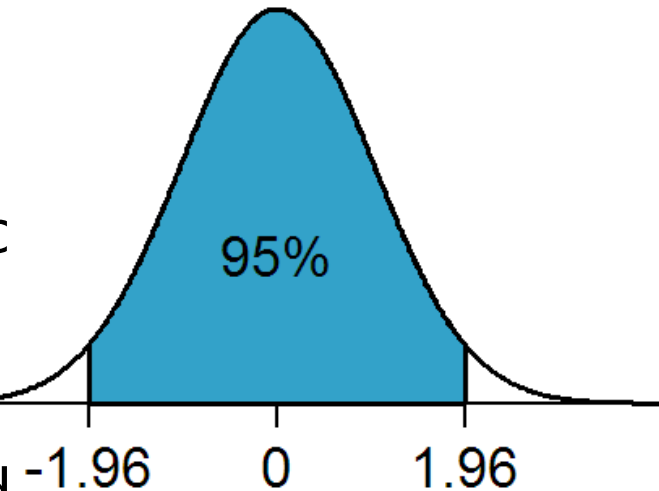
Setting out Hypotheses

Typically, we use the following convention:

- **Null hypothesis** (H_0): there is no significant difference between individuals/groups, and any observed difference is due to chance
- **Alternate hypothesis** (H_1): there is a significant difference

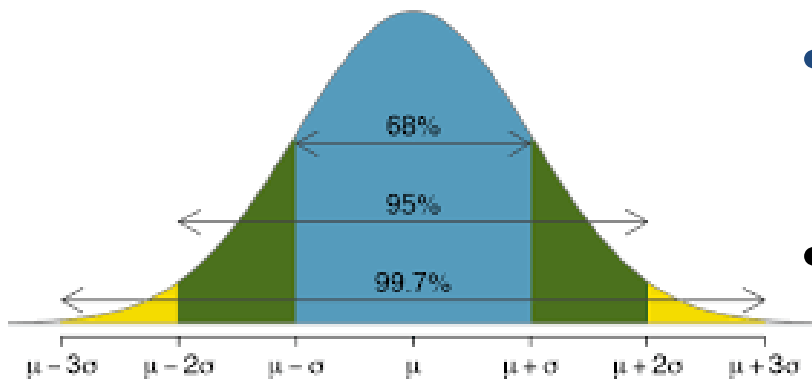
Statistical tests and p values

- Statistical tests do not give definite answers, just probabilities, or '**p values**'
- We decide if these probabilities are significant
- If we find a p value that is *smaller than* 0.05, this implies that the sample statistic is **significant** at this level
- We would then reject H_0 and accept H_1
- Eg. in the Welsh-cake example, we would deduce that our pack ('sample') is significantly different from typical Welsh-cakes ('population')



What is a significance level (α)?

- Typically, this is set at 5% (ie. $p < 0.05$; outside the middle 95%)
- This means that we accept H_1 if $p < 0.05$
- If we want to be really sure of a significant difference, we use a 1% significance level, where $p < 0.01$ (ie. outside the middle 99%)



Re-cap video:

<https://www.youtube.com/watch?v=eyknGvncKLw>

Key Types of Statistical test

