One population

Normality
- Visually
- Mean = median
- Skewness = 0
- Kurtosis – 3
- Jacque-Bera hypothesis test
- If population is normal, sample is normal
- n (population) ≥ 30

1. Variable is quantitative
2. Data is interval or ratio
3. Sample is random and independent
4. Sample is normally distributed

Type of data: **Numerical**

Use when:
- Mean is not an ideal measure b/c of outliers
- SD unknown (therefore cannot use z or t test)
- Sample mean not normally distributed

Parametric

Non-parametric **(use the median η)**

Mean (μ)

Known population SD

Unknown population SD, but known sample SD

Z test

t test

*assume population is normally distributed or n >30

On EViews: Satterthwaite-Welch t-test allows for unequal population variances

Sign-test

- Random sample
- At least ordinal

Binomial distribution

If n ≥ 10

Binomial approximated to normal distribution

Wilcoxon sign ranks

- More powerful, b/c takes into consideration the magnitude
- Variable quantitative and continuous
- Sample distribution is symmetric

If n (nonzero deviations) > 20 the test can be normally approximated

Type of data: **Nominal**

Categories

Two (i.e. 1,0)

Approximated to normal distribution **Z test**
- If np ≥ 5 and nq ≥ 5
- P-hat and mean have similar properties

More than two

χ^2
- Independence
- Goodness of fit

More than two populations

Numerical Data

Independent design

Normal distribution

ANOVA test

Extension of the z or t test for difference b/n two population means

Assumptions:
- Normality
- Independent random samples
- Same variance

Welch-F test (same assumptions as above, however can use when your groups have unequal variances.

Non-normal distribution

Kruskal-Wallis test

Generalisation of the Wilcoxon rank sum.

Assumptions:
- Data independent random samples
- Pops differ at most by central location
- Variable of interest = continuous
- Measurement scale at least ordinal

For k = 2, we can apply either the Wilcoxon rank-sum test or the Kruskal-Wallis test. However, while the Wilcoxon rank-sum test can be used for one-sided and two-sided alternative hypotheses alike, the Kruskal-Wallis test can only determine whether a significant difference exist between the sample means.

Randomised block design

Normal distribution

ANOVA test

Equivalent to the matched pair z or t test for the difference between two population means

Assumptions:
- Normality
- Independent random samples
- Same variance
- The block treatment effects are *additive*

Non-normal distribution

Friedman test

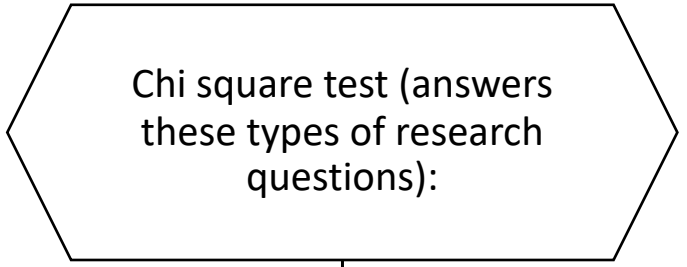
Generalisation of Wilcoxon signed rank matched pairs

Assumptions:
- Experimental design = Randomised blocks
- Sampled populations have similar spreads and shapes

Also can compare central locations of ranked data when it is: (1) quantitative, (2) not normal.

Chi square tests

- Performed on qualitative data or on quantitative data
- Measured on a nominal or ordinal scale, if:
  - The observed frequencies have been generated by a multinomial experiment:
    - There are n repeated, independent trials
    - Each trial has the same k mutually exclusive but exhaustive outcomes
    - The probability of each outcome is constant



Goodness of fit

(One population) – how well does a certain distribution fit the data?

H_0: the sample data have been drawn from a population that follows some 'known' distribution.

Independence

(One population, two variables)

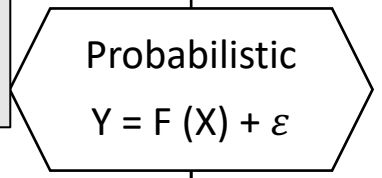
H_0: two variables are independent, so the distribution of one of them in no way depends on the distribution of the other.*This relationship can be:*

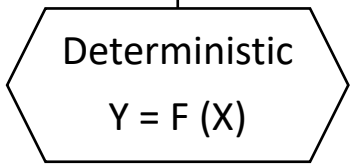
Homogeneity

(Multiple populations, one variable) – are several populations similar in terms of some characteristic?

H_0: populations are homogeneous in the sense that x follows the same distribution in these populations.

Statistics is concerned with ***probabilistic relationships*** and focuses on two questions:
- How strong is the relationship?
- What is the probable form of the relationship?





Measures of association

The simplest way to measure the strength of a linear relationship i.e. the strength of correlation, between two quantitative variables defined over the same population, is provided by the population covariance. It can be estimated by the sample covariance.

By standardising the covariance with SD we get →

Pearson (sample) correlation coefficient

Regression analysis

How to specify and estimate the relationship

Correlation analysis

How to measure the strength of the relationship

Spearman rank correlation coefficient
- Ordinal scale
- Populations are not normally distributed

Pearson correlation coefficient
- Variables quantitative
- Interval or ratio scale
- Both populations normally distributed

$|rxy| \geq 0.8$ strong linear relationship, $0.5 \leq |rxy| < 0.8$ moderately strong $0.2 \leq |rxy| < 0.5$ weak linear relationship.
Correlation between two variables, does not necessarily imply a causal relationship.