Statistics

- Types of variables
- Mean, mode, median, and standard deviation
- Histograms and normal distribution
- p-value
- Statistical tests
  - Chi-squared test
  - t-test
  - ANOVA
  - Correlation test

Types of variables

Variables

- Categorical variables: discrete or qualitative

  e.g. blue/red/yellow, male/female, or musician/nonmusician

- Continuous variables: quantitative

  e.g. reaction time, IQ, or score on a test

Scientific experiments

- Independent variables: changed/controlled in scientific experiment
- Dependent variables: tested/measured in scientific experiment

  e.g. the effect of musical training (IV) on cognitive abilities (DV)

Mean, mode, median, and standard deviation

Consider these test scores:

40, 45, 53, 58, 58, 61, 70, 72, 74, 79, 80, 85, 90

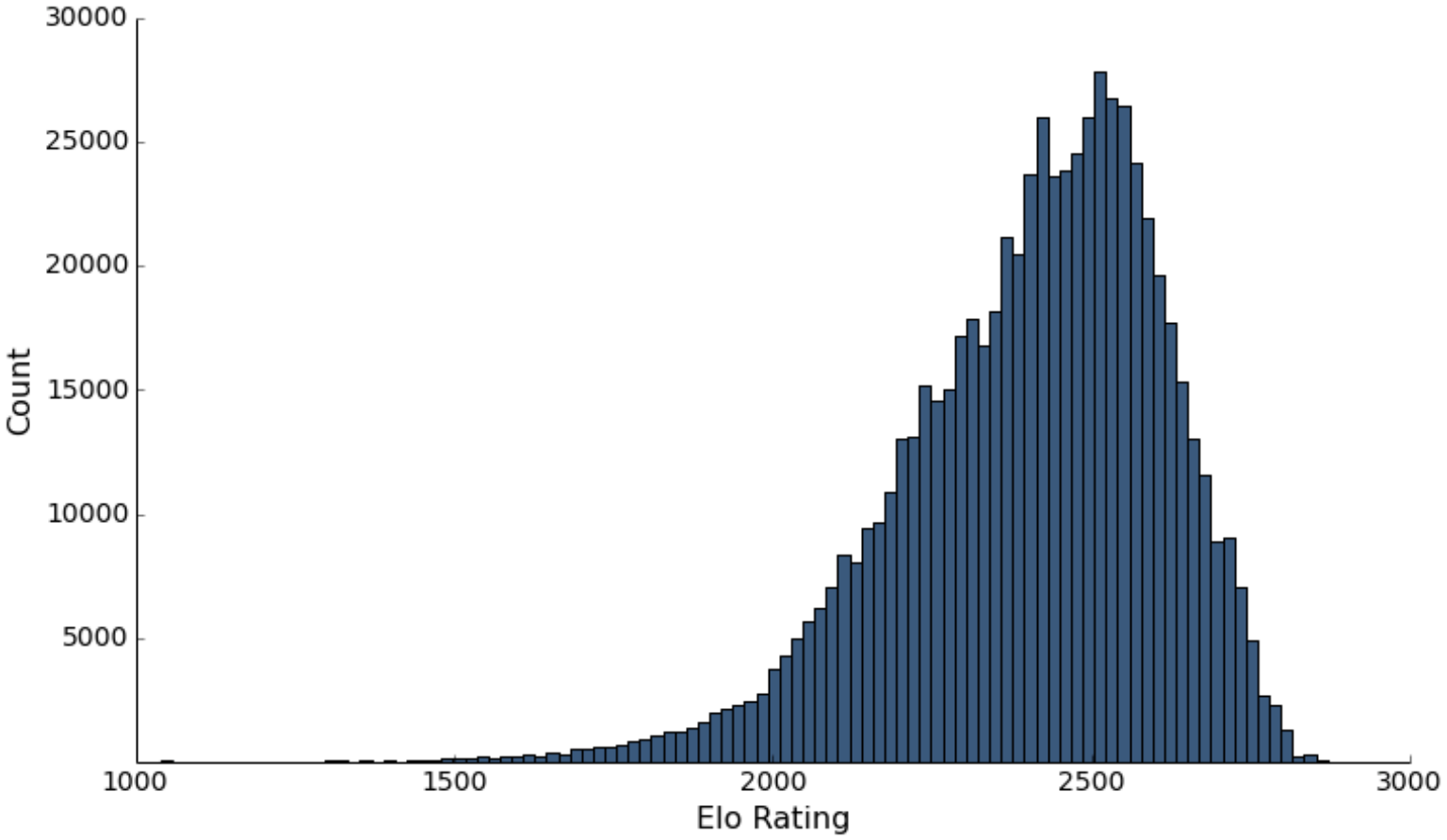
- The **mean** is the average test score (66.54)
- The **median** is the "middle" value from the list (70)
- The **mode** is the most frequent test score (58)

Standard deviation (SD, or σ)

- Quantity expressing by how much the members of the group differ from the mean value of the group
- In this case, *SD* = 15.42

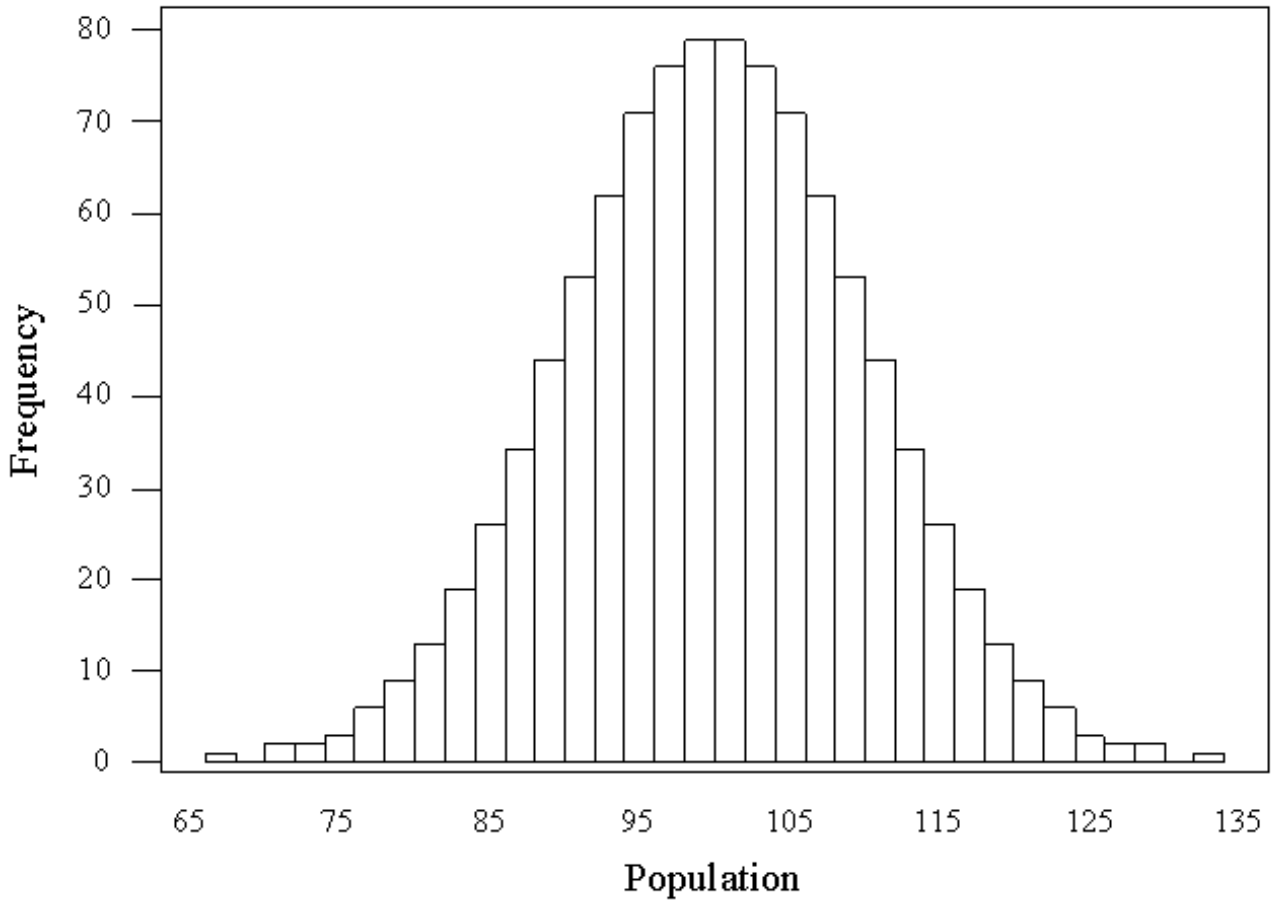
Histograms and normal distribution

A **histogram** is a graphical representation of the distribution of numerical data.



Histograms and normal distribution

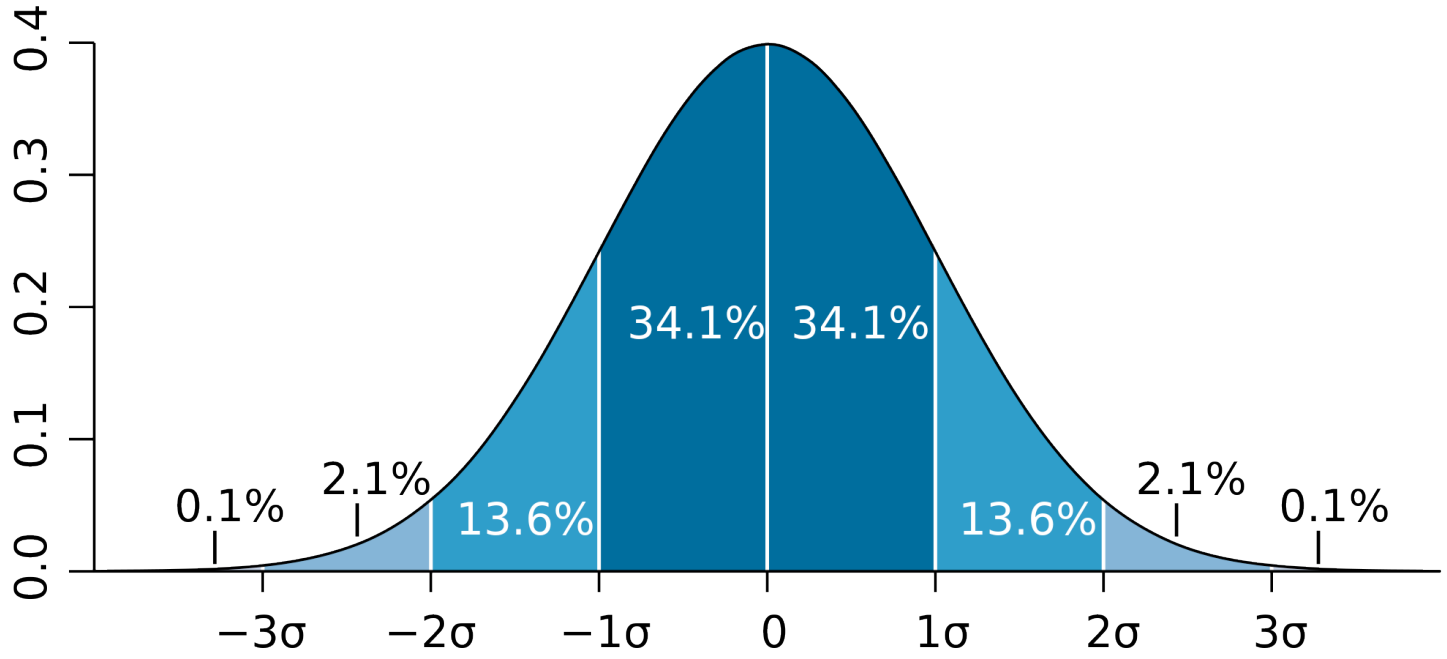
Data is **normally distributed** when the histogram creates a bell-shaped curve (i.e. normal curve).



Histograms and normal distribution

Data in natural and social sciences tends to be normally distributed.

In normally distributed data, about **68%** of the values are within one SD from the mean (**95%** for two SD, **99.7%** for three SD).



Histograms and normal distribution

Knowing the distribution of experimental results will inform the choice of statistical tests for the analysis of these results:

- **Parametric tests** are used for normally distributed data
- **Nonparametric tests** for data that follows another distribution

Nonparametric tests tend to rank the data, and are therefore generally less robust than parametric tests.

p-value

Most statistical tests return two values of interest:

- a value specific to the test (χ^2, t, r, F, etc.)
- a p-value

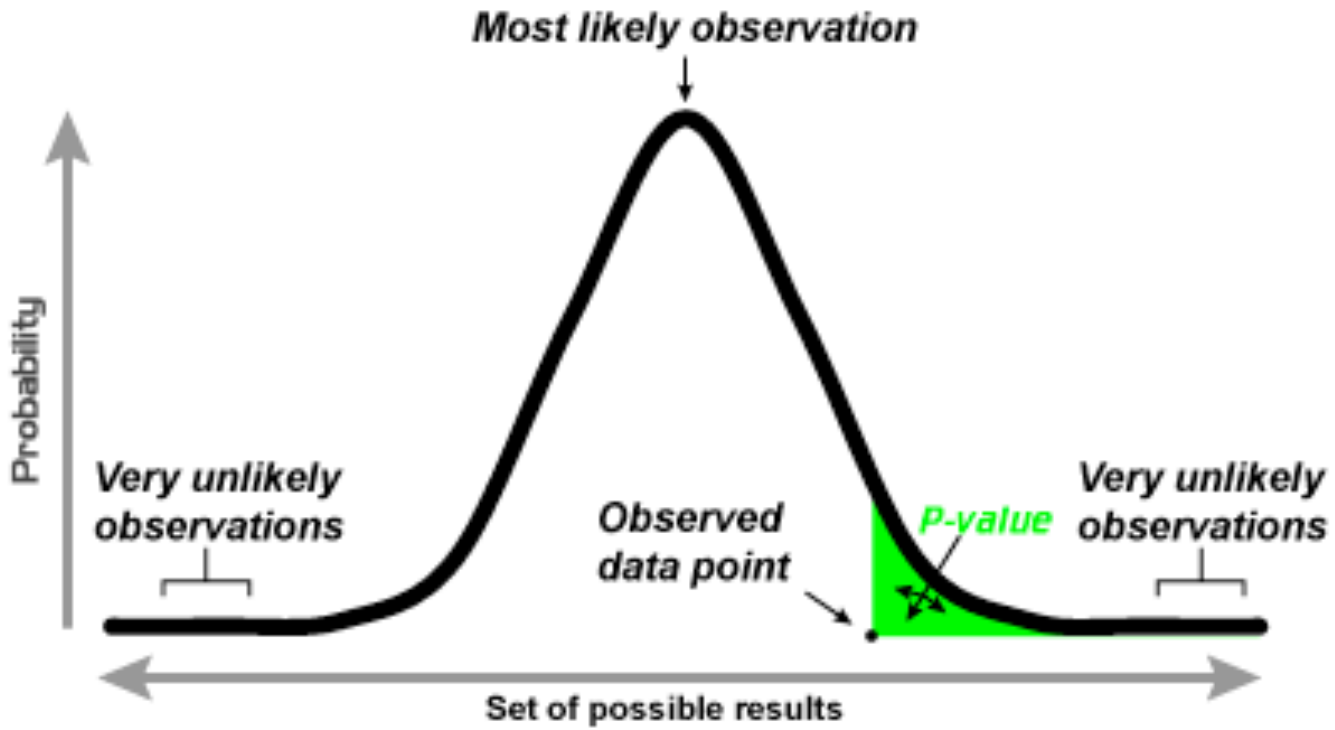
The **p-value** is the probability of obtaining a result equal or more extreme than what was actually observed, when the null hypothesis is true.

The **null hypothesis** is what researchers are trying to disprove when conducting an experiment

e.g. musical training doesn't affect cognitive abilities

p-value

In practice, the p-value determines statistical significance. In natural and social sciences, the significance level is traditionally set at $p < .05$.



Chi-squared test

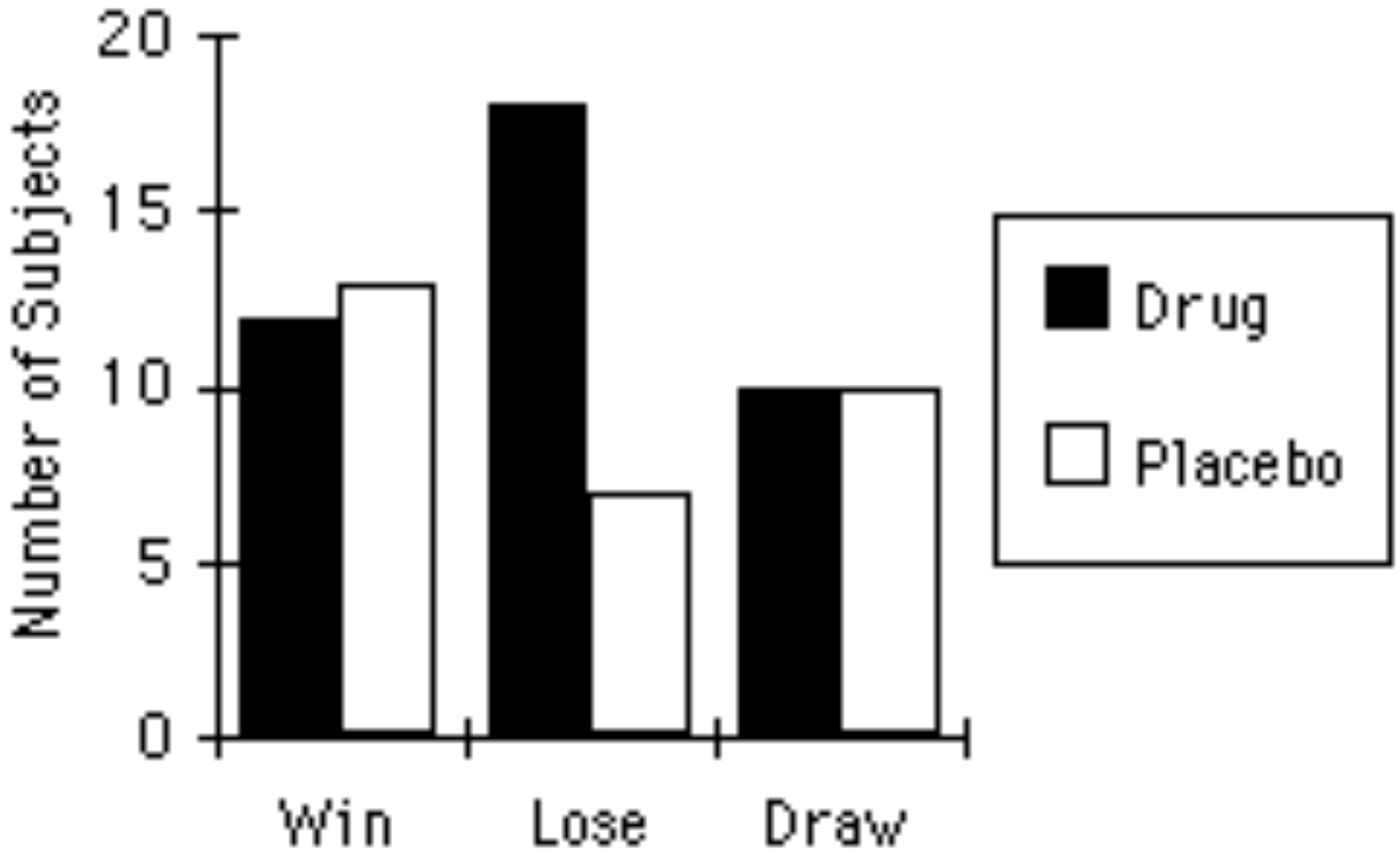
The Chi-squared test measures the association between **categorical independent variables** and **categorical dependant variables**.

e.g. favourite colour of boys/girls

The test returns the χ^2 value, which indicates a difference in **proportions between groups** if the test is significant.

Chi-squared test

Results are typically displayed in a bar chart.



t-test

The t-test measures the association between a **categorical independent variable with two levels** and a **continuous dependant variable**.

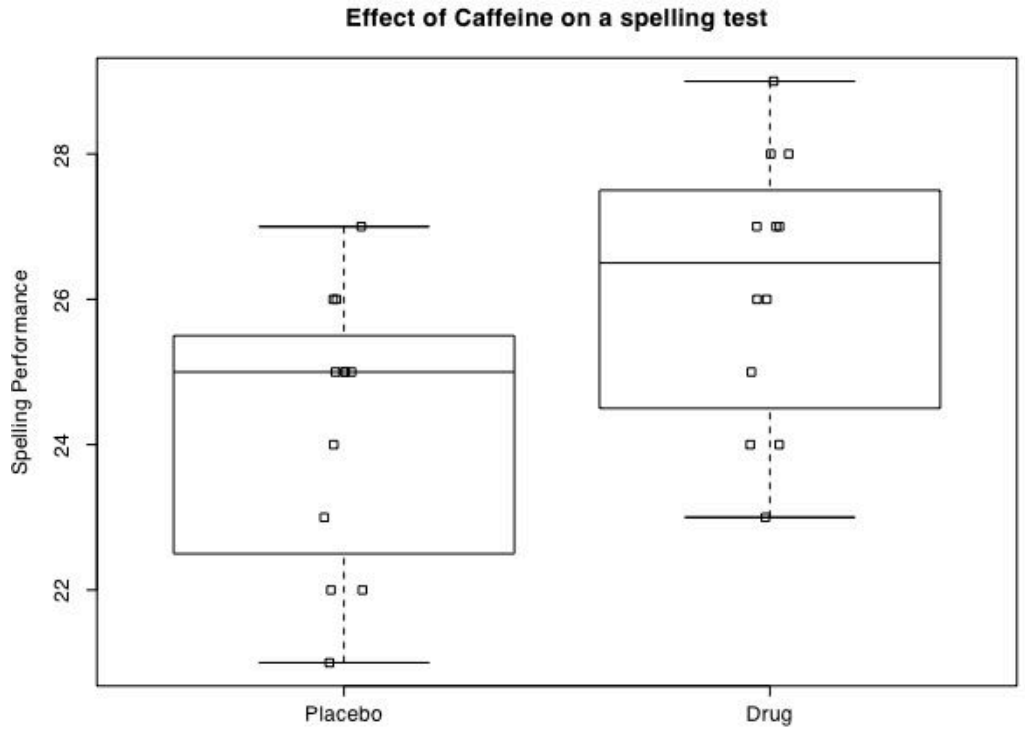
e.g. test scores of musicians/nonmusicians

The test returns the *t* value, which indicates a difference in **mean values between groups** if the test is significant.

t-test

Results are typically displayed in a box plot.

- The line is the middle of each box represents the median value
- The top and bottom of the box represent quartiles 1 and 3
- The whiskers typically represent one standard deviation from the mean



ANOVA

The analysis of variance (ANOVA) measures the association between a **categorical independent variable with more than two levels** and a **continuous dependant variable**.

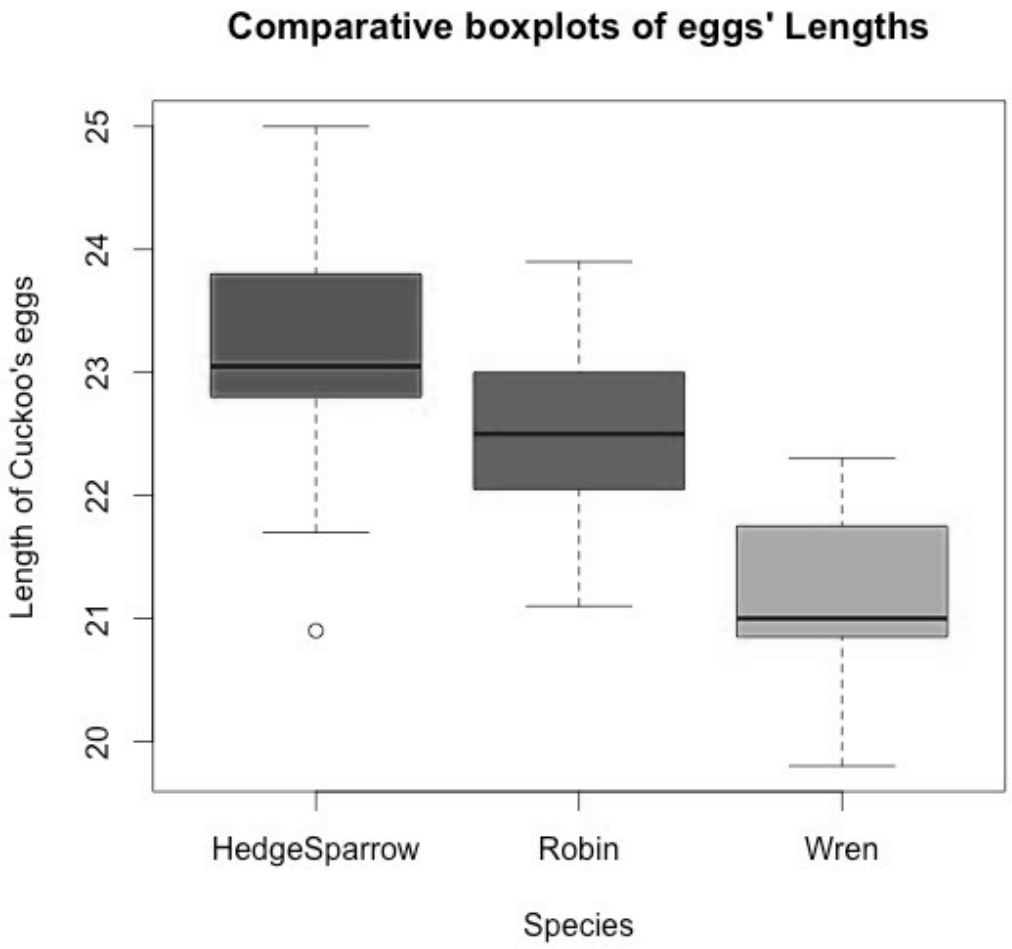
e.g. test scores of professional musicians/amateurs/nonmusicians

The test returns the *F* value, which indicates a difference in **mean values between groups** if the test is significant.

Post hoc tests are generally conducted if the ANOVA returns a significant result, to assess the differences between individual groups.

ANOVA

Results are typically displayed in a box plot.



Correlation test

The correlation test measures the association between a **continuous independent variable** and a **continuous dependant variable**.

e.g. test scores depending on length of musical training

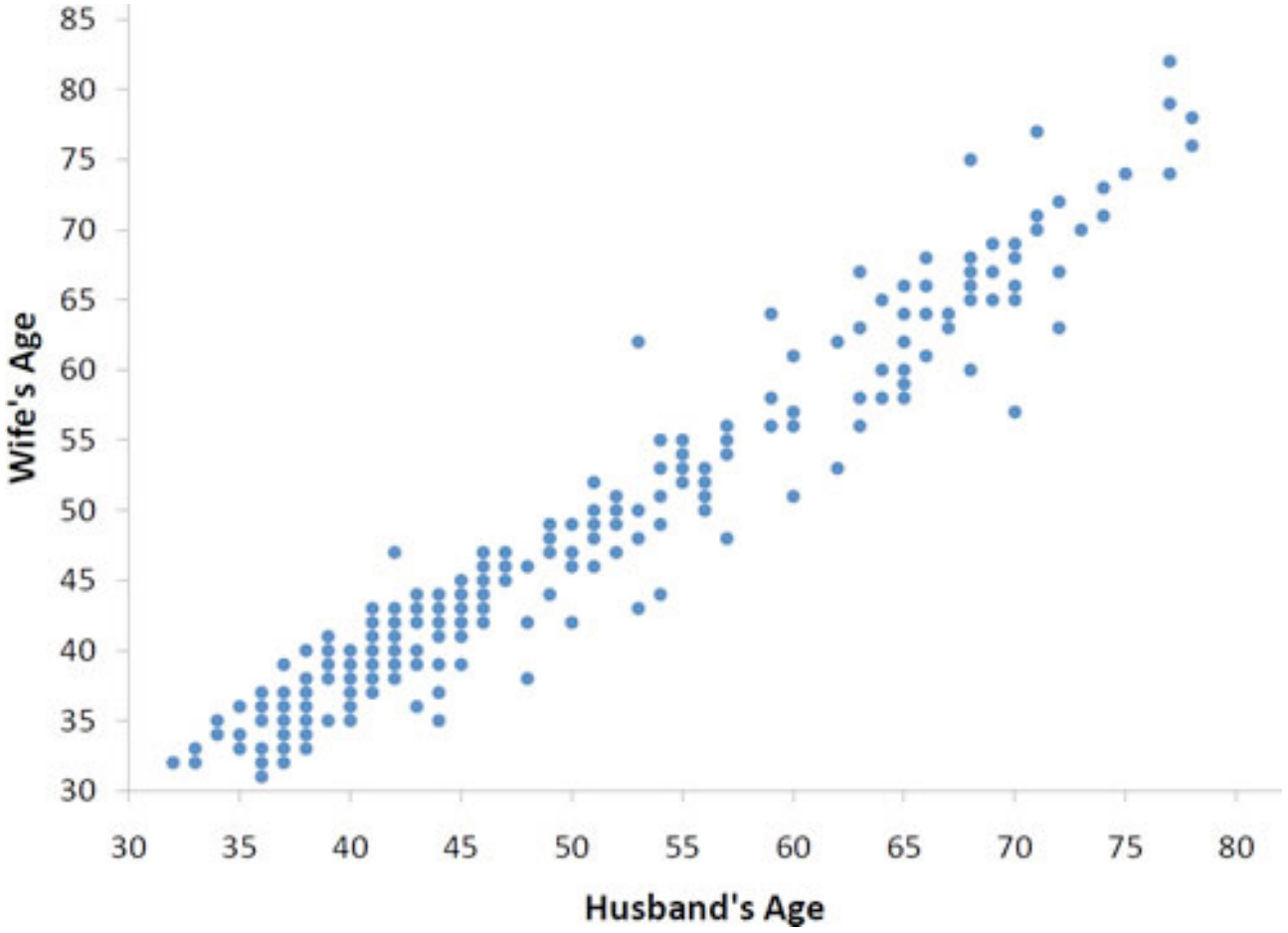
The test returns the *r* value, which indicates the magnitude of the correlation between these two variables.

Note: r^2 indicates the proportion of the variance in the DV that is predictable from the IV.

e.g. if r = .50, it means that the length of musical training accounts for 25% of the variance in test scores

Correlation test

Results are typically displayed in a scatter plot.



Correlation test

The magnitude of the correlation is easily visualised on the scatter plot.

The numbers below each plot represent the correlation coefficient *r*.