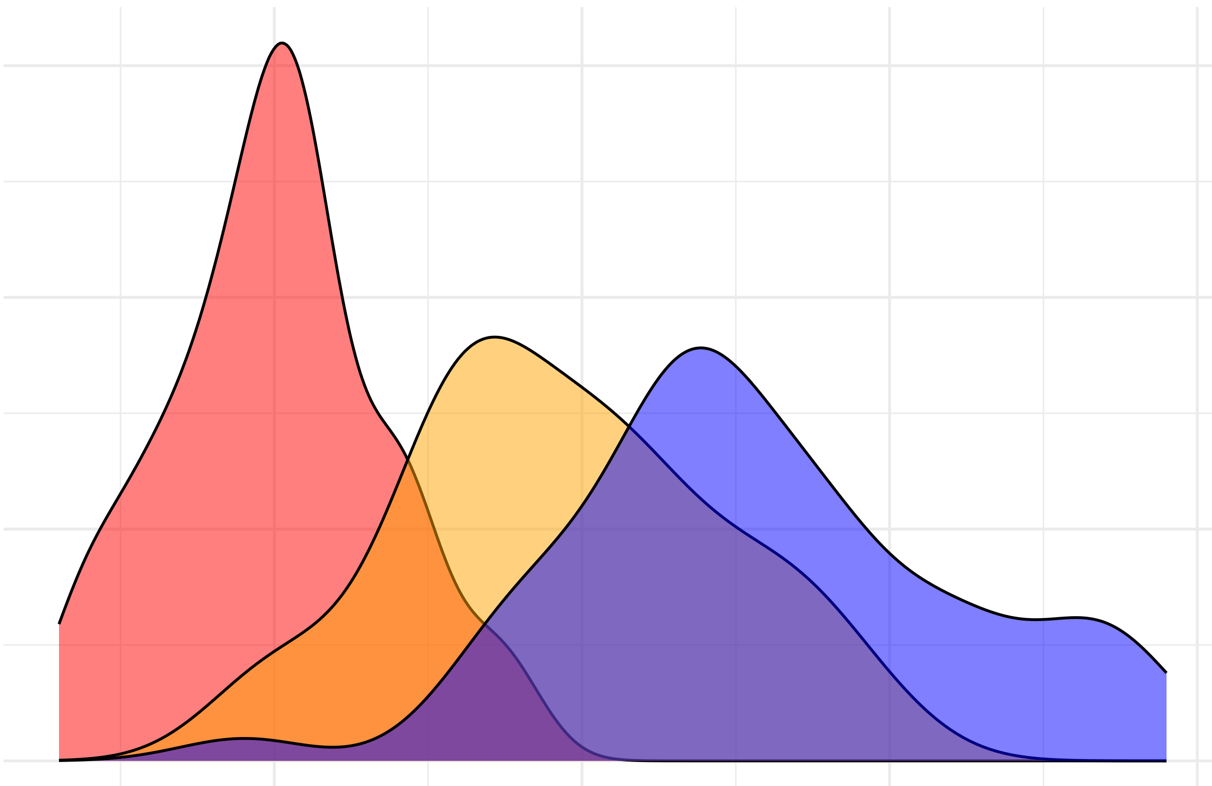
ANOVA

What is ANOVA?

ANOVA (Analysis of Variance) is a statistical test used to determine whether there are significant differences between the means of **three** or more groups. It is very useful in answering the question - "Do at least one of these groups differ from the others?".



Are these populations different?

When to use ANOVA?

When you are in a situation where you have one **continuous** variable and then a **categorical** variable with more than two groups.

Examples could include; comparing enamel hardness between multiple restorative materials, comparing plaque scores across three age groups or comparing whitening effect of different bleaching agents.

Assumptions of ANOVA?

The assumptions of an ANOVA that must be met are as follows:

1) Continuous outcome
2) Normally distributed data (Close to a bell shaped curve)
3) Homogeneity of variance (Similar variation between group A and B)
4) Independence of observations

Different types of ANOVA

Type of test: One-way ANOVA

Example: Comparing the orthodontic wire force across three different types of wire.

Type of test: Two-way ANOVA

Example: Investigating the effect of different drugs and dosages on enamel wear.

Type of test: Repeated measures ANOVA

Example: Investigation of pain perception at different time points after treatment (1 week, 6 months, 12 months)

In cases where the second assumption is not met (normal distribution of data) then a non-parametric Kruskal-Wallis test can be used.

Post-hoc analysis

When running ANOVA you are not able to tell which groups are significantly different from the others. If you were comparing A,B and C then you may find there is a significant difference but you need to know where that difference comes from. If an ANOVA result is considered not significant then there is no need to do this further analysis.

One example of a post-hoc test is called the Tukey test (Not Turkey), which compares A vs B, A vs C and B vs C. This will then give you information as to which groups are different from each other so you can infer information from your data.

An example of this is you may be comparing the effectiveness of a new toothpaste against two currently available products. Your ANOVA may be significant so then when you run the post-hoc test you may find that the new toothpaste is much better than the other two products but those products are very similar in quality.

How to implement in RStudio

```r
# Fit the ANOVA model
anova_model <- aov(continous_variable ~ categorical_variable, data = df)

# View the results
summary(anova_model)
```

```r
# Run the post-hoc
TukeyHSD(anova_model)
```

Interpreting results

Interpreting ANOVA results involves two main stages: first, understanding the primary ANOVA table to see if any group differences exist overall, and second, using post-hoc tests to pinpoint exactly which groups differ if the overall test is significant.

The p-value tells you how likely it is to observe your results if there is actually no real difference between the groups. So a low p-value means that it is very unlikely to see your data when there is no real difference. A threshold of 0.05 is commonly used in research and this means there is a 5% chance the results you observe are due to chance and not another reason.

In the post-hoc results there is an adjusted p-value instead, which is done to account for the effects of multiple testing. This is important because if you ran 100 statistical tests on something that isn't different, based on a threshold of 0.05 then you would expect 5 of them to be wrongly reported. The adjustment of p-values is an attempt to avoid false positives.