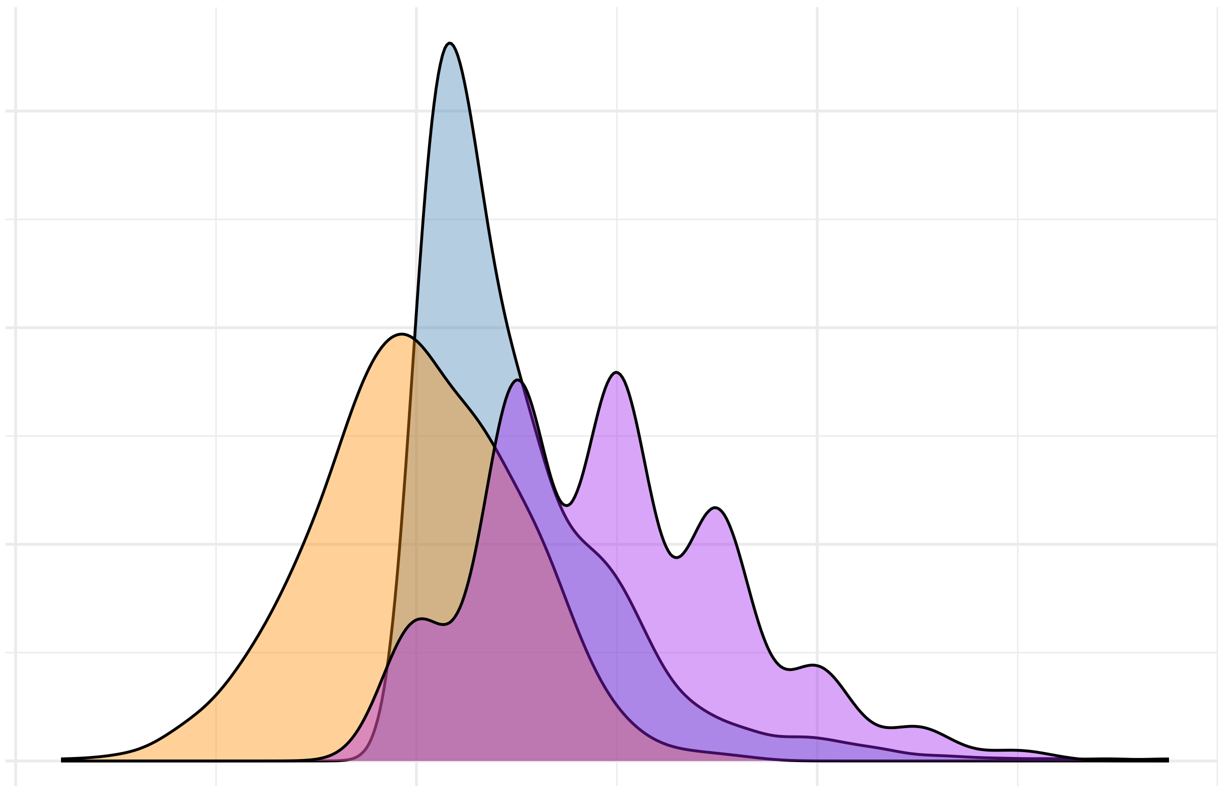
Normality Testing

What is normality testing?

Normality testing is used to determine whether your data follow a normal (bell-shaped) distribution. This matters because many common statistical tests make assumptions about the shape of the data. Some tests require normally distributed data (parametric tests) and others are designed for non-normal data (non-parametric tests). Normality testing helps you decide which type of statistical test is appropriate for your data.



Do these distributions look the same?

When to use normality testing?

Normality testing is used to determine whether your data follows a normal distribution, which is a key assumption for many statistical tests. It is particularly important when you have one continuous variable (e.g., enamel hardness, plaque scores, whitening effect) and one categorical variable with more than two groups (e.g., types of restorative materials, age groups, bleaching agents)

By performing normality testing, you can decide whether to use parametric tests (like T-test or ANOVA) for normally distributed data, or non-parametric alternatives (like Mann-Whitney U or Kruskal-Wallis) if the data is not normally distributed.

Different types of normality testing

Type of test: Shapiro-Wilk Test

When to use: A continuous variable with a sample size of anything less than around 300.

Type of test: Kolmogorov-Smirnov (K-S) Test

When to use: A continuous variable with a sample size of anything more than 300.

Type of test: Anderson-Darling Test

When to use: You want a more sensitive test to deviations in extreme values.

Type of test: Graphical methods

When to use: Histograms and Q-Q plots can be used alongside the above statistical tests to gain qualitative information.

How to implement in RStudio

```r
# Shapiro test
shapiro.test(values)

# Kolmogorov-Smirnov Test
ks.test(values, "pnorm", mean=mean(values), sd=sd(values))

# Anderson-Darling Test
ad.test(x) # from the nortest package

# Histogram
hist(values)

# Q-Q Plot
qqnorm(values)
```

Interpreting results

Interpreting normality test results involves assessing whether your continuous data follows a normal distribution, which determines if parametric tests are appropriate.

The p-value from a normality test indicates how likely it is to observe your data if it truly comes from a normal distribution. A high p-value (commonly > 0.05) suggests that the data does not significantly deviate from normality, so parametric tests are suitable. A low p-value suggests that the data significantly deviates from normality, and you may need to use non-parametric tests (e.g., Kruskal-Wallis instead of ANOVA).

Histograms and Q-Q plots provide a complementary way to assess normality. Histograms should resemble a bell-shaped curve and the data points in Q-Q plots should roughly follow a straight line and deviations at the tails indicate non-normality.

It is important to remember that small samples sizes may fail to detect non-normality, whereas large samples size with small deviations can result in a significant p-value. Therefor, always interpret p-values alongside graphical assessments to make a balanced judgment about your data's normality.