UNIVERSITY OF BATH

Data Science and Statistics in Research: unlocking the power of your data

Session 2.6:

Non-parametric statistics

OUTLINE

Non-parametric Testing

Parametric Tests

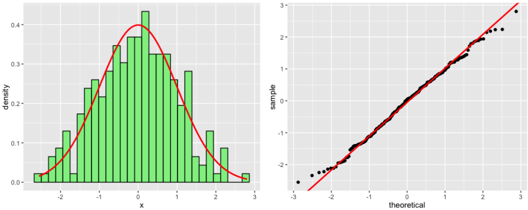
- The statistical tests you have seen so far require that the data can be assumed to follow a particular distribution, often the Normal distribution.
- This type of testing is called parametric.
- What happens when we have data that
  - may be clearly non-Normal?
  - might be Normal, but there is not enough data to establish this?
- Parametric methods are usually fine to use with reasonably-sized samples as long as the data are unimodal and roughly symmetric about the mean.
- If data are severely non-Normal and sample sizes are small, these methods may be unreliable.

Non-parametric Tests

- Non-parametric tests compare medians rather than means, and use a rank order of observations.
- They make no assumptions about the underlying distributions of the data.
- They may also be suitable for comparing nominal (categorical) and ordinal (ordered categorical) data.
- In a similar way to parametric tests, we
  - construct two alternative hypotheses
  - calculate a test statistic
  - compare to critical values to give a p-value
  - accept one hypothesis.
- Is there way of checking if I can use parametric statistics with my data?

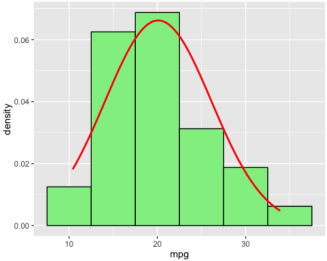
Checking for Normality

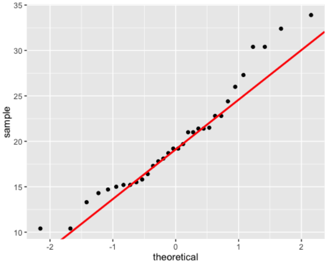
- **Histograms**: Use these to look at the distribution of your data. If the histogram is symmetric around a value then your data is normal.
- **Q-Q Plots**: If your data is normally distributed, the dots should follow the line.



EXAMPLE: MOTOR TREND CAR ROAD TESTS

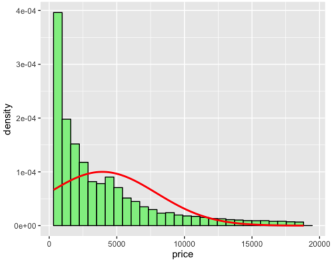
- The `mtcars` dataset in `R` contains fuel consumption and 10 other aspects of 32 cars from 1973-74.
- Is it reasonable to assume fuel consumption is normally distributed?
- Let's check this, by creating a histogram and Q-Q plot of fuel consumption.

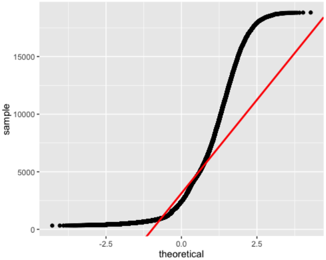




Example: Diamonds Dataset

- The `diamonds` dataset in `R` that records prices and other attributes of 53,940 round cut diamonds.
- Can we assume that prices of the diamonds are normally distributed?
- Let's check this, by creating a histogram and Q-Q plot of fuel consumption.





One sample tests

Example: Yields from a Barley Field

- The `immer` dataset in `R` give information about a farmer who grows five varieties of barley in six locations in 1931 and 1932 and measures the yearly yield.
- The farmer thinks the true median yield of barley in 1931 is 117.
- We set up the following hypothesis
  - null: the true median barley yield in 1931 is 117
  - alternative: the true median barley yield in 1931 is not 117.
- How do we test this hypothesis?
- There are not enough observations to deduce whether the data is normally distributed, so we will use a non-parametric test.

One-sample Wilcoxon Signed Rank Test

What is it for?

- A one-sample Wilcoxon Signed Rank Test is used to determine whether the median of a sample significantly differs from a specified value when there is evidence of non-normality.

What does it do?

- It tells you whether the true population median is significantly different to a known specified value, using a sample of that population.

What is the output?

- A p-value which indicates the probability that the data are consistent with the null hypothesis, that there is no difference between the true median and the known specified value.

One-sample Wilcoxon Signed Rank Test

How do you interpret the output?

- If the p-value is small, typically <0.05, then there is enough evidence to reject the null hypothesis.

What restrictions are there on its use?

- Only used when the data are severely non-normal and sample sizes are small. If this is not the case then a one-sample t-test is likely to be more powerful.

Example: Yields from a Barley Field

- We want to test the following hypotheses
  - null: the true median barley yield in 1931 is 117
  - alternative: the true median barley yield in 1931 is not 117.
- We choose a significance level of 0.05 for our test and construct the statistical decision rule
  - **IF** the p-value is less than 0.05
  - **THEN** we have enough evidence to reject the null hypothesis
  - **OTHERWISE** there is not enough evidence to reject the null hypothesis.

Example: Yields from a Barley Field

- There are 30 samples in our dataset.
- Sample median and mean yield of barley in 1931 is 102.95.
- Performing a one-sample Wilcoxon rank test gives us a p-value of 0.0732.
- This is greater than the chosen significance level.
- There is not enough evidence to reject the null hypothesis.

Mann Whitney U Test

Example: Prices of round cut diamonds

- The `diamonds` dataset in `R` records prices and other attributes of 53,940 round cut diamonds.
- This dataset provides information on the cut of the diamonds
  - graded either ‘Fair’ ‘Good’, ‘Very Good’, ‘Premium’ or ‘Ideal’.
- The jeweller wants to test whether the prices of diamonds with a ‘Fair’ cut are different from those with a ‘Good’ cut.

Example: Prices of round cut diamonds

- We set up the following hypothesis
  - null: the distribution of prices between diamonds with a 'Fair' cut is the same from those with a 'Good' cut
  - alternative: the distribution of prices between diamonds with a 'Fair' cut is different from those with a 'Good' cut.
- How do we test this hypothesis?
- The prices are severely right skewed so we use non-parametric tests.

Mann Whitney U Test

What is it for?

- A Mann Whitney U test is used to compare the distribution of a numeric variable between two independent groups when there is evidence of non-Normality.

What does it do?

- It tells you whether the distribution of the variable is significantly different between the two groups.

What is the output?

- A p-value which indicates the probability that the data are consistent with the null hypothesis of no difference between the groups in terms of the ranks of the observations.

Mann Whitney U Test

How do you interpret the output?

- If the p-value is small, typically 0.05, then there is sufficient evidence to reject the null hypothesis.

What restrictions are there on its use?

- Only used when the data are severely non-normal and sample sizes are small. If this is not the case then an independent t-test is likely to be more powerful.

Example: Prices of round cut diamonds

- We want to test the following hypotheses
  - null: the distribution of prices between diamonds with a 'Fair' cut is the same from those with a 'Good' cut
  - alternative: the distribution of prices between diamonds with a 'Fair' cut is different from those with a 'Good' cut.
- We choose a significance level of 0.05 for our test and construct the statistical decision rule
  - **IF** the p-value is less than 0.05
  - **THEN** we have enough evidence to reject the null hypothesis
  - **OTHERWISE** there is not enough evidence to reject the null hypothesis.

Example: Prices of round cut diamonds

- There are 1610 'Fair' cut diamonds and 4906 'Good' cut diamonds in our dataset.
- Sample median price of 'Fair' cut diamonds is $3282.
- Sample median price of 'Good' cut diamonds is $3050.5.
- Performing a Mann-Whitney U test gives us a p-value of <0.0001.
- This is less than the chosen significance level.
- We have enough evidence to reject the null hypothesis.
- We conclude that the true median price between diamonds with a 'Fair' cut are different from those with a 'Good' cut.

Wilcoxon Signed Rank Test

Example: Yields from a Barley Field Trial

- The `immer` dataset in `R` give information about a farmer who grows five varieties of barley in six locations in 1931 and 1932 and measures the yearly yield.
- The farmer wants to test if the amount yields in 1931 and 1932 are different.
- We set up the following hypothesis
  - null: there is no difference in the distributions of yield between 1931 and 1932
  - alternative: there is a difference in the distributions of yield between 1931 and 1932.
- How do we test this hypothesis?
- There are not enough observations to deduce whether the data is normally distributed so we use non-parametric tests.

Wilcoxon Signed Rank Test

What is it for?

- A Wilcoxon Signed Rank Test is used to compare the distribution of a numeric variable of two groups when there is evidence of non-Normality.
- The two groups must be of equal size and subjects in one sample are paired with one in the other.

What does it do?

- It tells you whether the distribution of the variable is significantly different between the two groups.

What is the output?

- A p-value which indicates the probability that the data are consistent with the null hypothesis of no difference between the groups in terms of the ranks of the observations.

Wilcoxon Signed Rank Test

How do you interpret the output?

- If the p-value is small, typically <0.05, then there is enough evidence to reject the null hypothesis.

What restrictions are there on its use?

- Only used when the data are severely non-normal and sample sizes are small. If this is not the case then a paired t-test is likely to be more powerful.

Example: Yields from a Barley Field

- We want to test the following hypotheses
  - null: there is no difference in the distributions between yields between 1931 and 1932
  - alternative: there is a difference in the distribution between yields between 1931 and 1932.
- We choose a significance level of 0.05 for our test and construct the statistical decision rule
  - **IF** the p-value is less than 0.05
  - **THEN** we have enough evidence to reject the null hypothesis
  - **OTHERWISE** there is not enough evidence to reject the null hypothesis.

Example: Yields from a Barley Field

- There are 30 samples in our dataset.
- Sample median yield of barley in 1931 is 102.95.
- Sample median yield of barley in 1932 is 92.95.
- Performing an Wilcoxon rank test gives us a p-value of 0.0053.
- This is less than the chosen significance level.
- We therefore have enough evidence to reject the null hypothesis.
- We conclude that there is a difference in the true median yield between 1931 and 1932.

Any Questions?