Analysis of Quantitative data

Introduction

Anne Segonds-Pichon
v2020-12
Outline of this section

• Assumptions for parametric data

• Comparing two means: Student’s t-test

• Comparing more than 2 means
  • One factor: One-way ANOVA
  • Two factors: Two-way ANOVA

• Relationship between 2 continuous variables: Correlation
Introduction

• Key concepts to always keep in mind
  – Null hypothesis and error types
  – Statistics inference
  – Signal-to-noise ratio
The null hypothesis and the error types

- The null hypothesis \((H_0)\): \(H_0 = \text{no effect}\)
  - e.g. no difference between 2 genotypes
- The aim of a statistical test is to reject or not \(H_0\).

<table>
<thead>
<tr>
<th>Statistical decision</th>
<th>True state of (H_0)</th>
<th>(H_0) True (no effect)</th>
<th>(H_0) False (effect)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject (H_0)</td>
<td>Type I error (\alpha)</td>
<td>Correct</td>
<td>Correct</td>
</tr>
<tr>
<td></td>
<td>False Positive</td>
<td>(\text{True Positive})</td>
<td>(\text{Type II error} (\beta))</td>
</tr>
<tr>
<td>Do not reject (H_0)</td>
<td>Correct</td>
<td>(\text{True Negative})</td>
<td>False Negative</td>
</tr>
</tbody>
</table>

- Traditionally, a test or a difference is said to be “significant” if the probability of type I error is: \(\alpha =< 0.05\)
- **High specificity** = low **False Positives** = low **Type I error**
- **High sensitivity** = low **False Negatives** = low **Type II error**
Statistical inference

Difference

Meaningful?

Yes

Real?

Statistical test

Statistic
e.g. t, F ...

Big enough?

Sample

Difference

+ Noise +

Population
• Stats are all about understanding and controlling variation.

If the noise is low then the signal is detectable ...

... but if the noise (i.e. interindividual variation) is large then the same signal will not be detected

• In a statistical test, the ratio of signal to noise determines the significance.
Analysis of Quantitative Data

• Choose the correct statistical test to answer your question:

  – They are 2 types of statistical tests:

    • **Parametric tests** with 4 assumptions to be met by the data,

    • **Non-parametric tests** with no or few assumptions (e.g. Mann-Whitney test) and/or for qualitative data (e.g. Fisher’s exact and $\chi^2$ tests).
Assumptions of Parametric Data

• All parametric tests have 4 basic assumptions that must be met for the test to be accurate.

First assumption: Normally distributed data

– Normal shape, bell shape, Gaussian shape

• Transformations can be made to make data suitable for parametric analysis.
Assumptions of Parametric Data

- **Frequent departures from normality:**
  - **Skewness:** lack of symmetry of a distribution

- **Kurtosis:** measure of the degree of ‘peakedness’ in the distribution
  - The two distributions below have the same variance approximately the same skew, but differ markedly in kurtosis.
Assumptions of Parametric Data

**Second assumption: Homoscedasticity (Homogeneity in variance)**

- The variance should not change systematically throughout the data.

**Third assumption: Interval data (linearity)**

- The distance between points of the scale should be equal at all parts along the scale.

**Fourth assumption: Independence**

- Data from different subjects are independent
  - Values corresponding to one subject do not influence the values corresponding to another subject.
  - Important in repeated measures experiments
Analysis of Quantitative Data

• Is there a difference between my groups regarding the variable I am measuring?
  – e.g. are the mice in the group A heavier than those in group B?
    • Tests with 2 groups:
      – Parametric: Student’s t-test
      – Non parametric: Mann-Whitney/Wilcoxon rank sum test
    • Tests with more than 2 groups:
      – Parametric: Analysis of variance (one-way and two-way ANOVA)
      – Non parametric: Kruskal Wallis (one-way ANOVA equivalent)

• Is there a relationship between my 2 (continuous) variables?
  – e.g. is there a relationship between the daily intake in calories and an increase in body weight?
    • Test: Correlation (parametric or non-parametric)