Analysis of Quantitative data

Introduction

Anne Segonds-Pichon
v2020-08
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:
  • Linear: Correlation
  • Non-linear: Curve fitting
  • Model diagnostics: Goodness-of-fit
• Non-parametric tests
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>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$H_0$ True (no effect)</td>
</tr>
</tbody>
</table>
| Reject $H_0$          | Type I error $\alpha$  
                         | False Positive          | Correct          |
|                       |                         |                         | True Positive    |
| Do not reject $H_0$   | Correct                  |                          |
|                       | True Negative           | Type II error $\beta$   |
|                       |                         | False Negative          |

• Traditionally, a test or a difference is said to be “significant” if the probability of type I error is: $\alpha \leq 0.05$

• High specificity = low False Positives = low Type I error
• High sensitivity = low False Negatives = low Type II error
Statistical inference

Sample → Difference → Meaningful? → Yes → Real? → Statistical test
Big enough? → Statistic e.g. t, F ...
= Difference + Noise + Sample
Stats are all about understanding and controlling variation.

If the noise is low then the signal is detectable ... = statistical significance

... but if the noise (i.e. interindividual variation) is large then the same signal will not be detected = no statistical significance

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

  ![Histogram of Raven egg lengths](image)

  • 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

  ![Skewness Diagram]

  - Skewness < 0: Negatively skewed
  - Skewness = 0: Normal (no skew)
  - Skewness > 0: Positively skewed

• **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.

  ![Kurtosis Diagram]

  - More peaked distribution: kurtosis > 0
  - Flatter distribution: kurtosis < 0
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) and Curve fitting