THE UNIVERSITY of EDINBURGH | usher institute

Non-parametric Methods

Introduction

Niall Anderson

Supported by

THE UNIVERSITY of EDINBURGH | IDI Data-Driven Innovation

Outline

- Identifying a problem
- Possible solution?
- Example – 2 sample problem
- Non-parametric analogues of standard methods

Making Assumptions…

- Standard methods for continuous data e.g. Student's t-test assume data sampled from Normal distribution
- Need to estimate parameters of assumed Normal distribution
- Reality is not always Normal!
- T-tests etc can be quite ***robust*** (insensitive to departure from Normality)
- More extreme cases may require an alternative…

Non-parametric Methods

AKA ***rank-based*** or ***distribution-free*** methods

Avoid Normality assumption by replacing data with ranks: e.g.

Data:	7	4	9	17	11	6	21	14
Rank:	3	1	4	7	5	2	8	6

We analyse the ranks, instead of the original data…

Example – 2 sample problem

Assume the data on the previous slide comprise measurements from 2 groups, A and B:

Data:	7	4	9	17	11	6	21	14
Rank:	3	1	4	7	5	2	8	6
Group:	A	A	A	A	B	B	B	B

If A and B are sampled from similar distributions, then ranks should be randomly distributed between A and B.

Thus, group-specific sets of ranks would be similar: alternatively think of sum of ranks:

Group A sum = 15, Group B sum = 21

Close, but not same…

Can use a probability argument to work out how likely (or not) those sums of ranks are to be found, assuming same (unspecified) distribution of measurements for A and B.

Known as ***Wilcoxon Rank Sum*** test or ***Mann-Whitney*** test.

Here, p = 0.484 – A and B appear to come from same distribution.

Why “Non-parametric”?

- No need to estimate the parameters (mean & SD) of the assumed Normal distribution
- “Rank-based” = as it sounds!
- “Distribution-free” = no requirement for a specific probability distribution to generate the observed data.

Essentially, all equivalent terminology.

Analogs…

Parametric Test	Non-parametric Analog
1 sample t-test	Wilcoxon Signed Rank test
Paired t-test	Wilcoxon Matched Pairs test
2 sample t-test	Wilcoxon Rank Sum/ Mann-Whitney test
One-way ANOVA	Kruskall-Wallis
Pearson's Correlation Coefficient	Spearman Rank Correlation Coefficient

A number of other tests may be encountered in the literature.

Worth remembering: these methods are only really important for ***continuous*** data (because otherwise a Normality assumption may be made) – not encountered with categorical data.