DATA ANALYSIS WITH R

BRUNO BELLISARIO, PHD

SESSION 5 STATISTICAL INFERENCE#3

NON-PARAMETRIC STATISTICS

→A large portion of the field of statistics and statistical methods is dedicated to data where the distribution is known.

→Samples of data where we already know or can easily identify the distribution of are called parametric data. Often, parametric is used to refer to data that was drawn from a Gaussian distribution in common usage. Data in which the distribution is unknown or cannot be easily identified is called nonparametric.

→In the case where you are working with nonparametric data, specialized nonparametric statistical methods can be used that discard all information about the distribution. As such, these methods are often referred to as distribution-free methods.

NON-PARAMETRIC STATISTICS

→The traditional nonparametric tests are primarily rank-based tests.

→Instead of using the numeric values of the dependent variable, the dependent variable is converted into relative ranks.

→The advantage of using these rank-based tests is that they don't make many assumptions about the distribution of the data. Instead, their conclusions are based on the relative ranks of values in the groups being tested.

ADVANTAGES OF NON-PARAMETRIC TESTS

→Most of traditional nonparametric tests are relatively common.

→They are appropriate for interval/ratio or ordinal dependent variables.

→ Their nonparametric nature makes them appropriate for data that don't meet the assumptions of parametric analyses. These include data that are skewed, non-normal, contain outliers, or possibly are censored.

DISADVANTAGES OF NON-PARAMETRIC TESTS

→Tests are typically named after their authors, with names like Mann–Whitney, Kruskal–Wallis, and Wilcoxon signed-rank. It may be difficult to remember these names, or to remember which test is used in which situation.

→Most of the traditional nonparametric tests are limited by the types of experimental designs they can address. They are typically limited to a one-way comparison of independent groups (e.g., Kruskal–Wallis), or to unreplicated complete block design for paired samples (e.g., Friedman). The aligned ranks transformation approach, however, allows for more complicated designs.

→Readers are likely to find a lot of contradictory information in different sources about the hypotheses and assumptions of these tests.

→Authors will often treat the hypotheses of some tests as corresponding to tests of medians, and then list the assumptions of the test as corresponding to these hypotheses. However, if this is not explicitly explained, the result is that different sources list different assumptions that data must meet for the test to be valid. This creates unnecessary confusion in the mind of students trying to correctly employ these tests.

INTERPRETATION OF NON-PARAMETRIC TESTS

→In general, these tests determine if there is a systematic difference among groups. This may be due to a difference in location (e.g., median) or in the shape or spread of the distribution of the data.

→With the Mann–Whitney and Kruskal–Wallis tests, the difference among groups that is of interest is the probability of an observation from one group being larger than an observation from another group. If this probability is 0.50, this is termed “stochastic equality”, and when this probability is far from 0.50, it is sometimes called “stochastic dominance”.

→Without additional assumptions about the distribution of the data, the Mann–Whitney and Kruskal–Wallis tests do not test hypotheses about the group medians.

EFFECT SIZE STATISTICS

→Effect size statistics for traditional nonparametric tests include Cliff's delta and Vargha and Delaney's A for Mann–Whitney, and epsilon-squared and Freeman's "coefficient of determination" (Freeman's theta) for Kruskal–Wallis. There is also an r statistic for Mann–Whitney and the paired signed-rank test. Kendall's W can be used for Friedman's test.

→Some effect size statistics determine the degree to which one group has data with higher ranks than other groups. They tend to vary from 0 (groups have data that are stochastically equal) to 1 (one group stochastically dominates). They are related to the probability that a value from one group will be greater than a value from another group.

→As rank-based measures, these effect size statistics do not indicate the difference in absolute values between groups.

APPROPRIATE USE OF TRADITIONAL NON-PARAMETRIC TESTS

→Some authors caution against using traditional nonparametric tests with ordinal dependent variables, since many of them were developed for use with continuous (interval/ratio) data.

→Other authors argue that, since these tests rank-transform data before analysis and have adjustments for tied ranks, that they are appropriate for ordinal data.

→Nonparametric tests are commonly used for interval/ratio data when the data fail to meet the assumptions of parametric analysis.

→Given these considerations and the fact that parametric statistics are often relatively robust to minor deviations in their assumptions, some authors argue that it is often better to stick with parametric analyses for interval/ratio data if it's possible to make them work.

ONE-SAMPLE WILCOXON SIGNED-RANK TEST

The one-sample Wilcoxon test is a rank-based test that begins with calculating the difference between the observed values and the default value. Because of this subtraction operation in the calculations, the data are assumed to be interval.

Appropriate data

→One-sample data

→Data are interval or ratio

Hypotheses

→Null hypothesis (simplified): The population from which the data were sampled is symmetric about the default value.

→Alternative hypothesis (simplified, two-sided): The population from which the data were sampled is not symmetric about the default value.

TWO-SAMPLE MANN–WHITNEY U TEST

→The two-sample Mann–Whitney U test is a rank-based test that compares values for two groups. A significant result suggests that the values for the two groups are different. It is equivalent to a two-sample Wilcoxon rank-sum test.

→Without further assumptions about the distribution of the data, the Mann–Whitney test does not address hypotheses about the medians of the groups. Instead, the test addresses if it is likely that an observation in one group is greater than an observation in the other. This is sometimes stated as testing if one sample has stochastic dominance compared with the other.

→The test assumes that the observations are independent. That is, it is not appropriate for paired observations or repeated measures data.

TWO-SAMPLE MANN–WHITNEY U TEST

Appropriate data

→Two-sample data. That is, one-way data with two groups only

→Dependent variable is ordinal, interval, or ratio

→Independent variable is a factor with two levels. That is, two groups

→Observations between groups are independent. That is, not paired or repeated measures data

→In order to be a test of medians, the distributions of values for each group need to be of similar shape and spread. Otherwise, the test is typically a test of stochastic equality.

Hypotheses

→Null hypothesis: The two groups are sampled from populations with identical distributions. Typically, that the sampled populations exhibit stochastic equality.

→Alternative hypothesis (two-sided): The two groups are sampled from populations with different distributions. Typically, that one sampled population exhibits stochastic dominance.

TWO-SAMPLE MANN—WHITNEY U TEST

Effect size

Statistics of effect size for the Mann–Whitney test report the degree to which one group has data with higher ranks than the other group. They are related to the probability that a value from one group will be greater than a value from the other group. Unlike p-values, they are not affected by sample size.

→Vargha and Delaney's A

→Cliff's delta

→Kendall's tau-b

→Freeman's theta and epsilon-square

	small	medium	large
r	0.10 – < 0.30	0.30 – < 0.50	≥ 0.50
tau-b	0.10 – < 0.30	0.30 – < 0.50	≥ 0.50
Cliff's delta	0.11 – < 0.28	0.28 – < 0.43	≥ 0.43
Vargha and Delaney's A	0.56 – < 0.64	0.64 – < 0.71	≥ 0.71
	>0.34 – 0.44	> 0.29 – 0.34	≤ 0.29
Freeman's theta	0.11 – < 0.34	0.34 – < 0.58	≥ 0.58
epsilon-squared	0.01 – < 0.08	0.08 – < 0.26	≥ 0.26

KRUSKAL-WALLIS TEST

→The Kruskal–Wallis test is a rank-based test that is like the Mann–Whitney U test but can be applied to one-way data with more than two groups.

→Without further assumptions about the distribution of the data, the Kruskal–Wallis test does not address hypotheses about the medians of the groups. Instead, the test addresses if it is likely that an observation in one group is greater than an observation in the other.

→The test assumes that the observations are independent. That is, it is not appropriate for paired observations or repeated measures data.

KRUSKAL-WALLIS TEST

Post-hoc tests

To determine which groups are different from others, post-hoc testing can be conducted. The most common post-hoc test for the Kruskal–Wallis test is the Dunn test.

Appropriate data

→One-way data

→Dependent variable is ordinal, interval, or ratio

→Independent variable is a factor with two or more levels. That is, two or more groups

→Observations between groups are independent. That is, not paired or repeated measures data

→In order to be a test of medians, the distributions of values for each group need to be of similar shape and spread

Hypotheses

→Null hypothesis: The groups are sampled from populations with identical distributions. Typically, that the sampled populations exhibit stochastic equality.

→Alternative hypothesis (two-sided): The groups are sampled from populations with different distributions. Typically, that one sampled population exhibits stochastic dominance.

KRUSKAL-WALLIS TEST

Effect size

Statistics of effect size for the Kruskal–Wallis test provide the degree to which one group has data with higher ranks than another group. They are related to the probability that a value from one group will be greater than a value from another group. Unlike p-values, they are not affected by sample size.

→Freeman's theta and epsilon-squared

→Cliff's delta

→Vargha and Delaney's A

	small	medium	large
epsilon-squared	0.01 – < 0.08	0.08 – < 0.26	≥ 0.26
Freeman's theta, k = 2	0.11 – < 0.34	0.34 – < 0.58	≥ 0.58
Freeman's theta, k = 3	0.05 – < 0.26	0.26 – < 0.46	≥ 0.46
Freeman's theta, k = 5	0.05 – < 0.21	0.21 – < 0.40	≥ 0.40
Freeman's theta, k = 7	0.05 – < 0.20	0.20 – < 0.38	≥ 0.38
Freeman's theta, k = 7	0.05 – < 0.20	0.20 – < 0.38	≥ 0.38
Maximum Cliff'sdelta	0.11 – < 0.28	0.28 – < 0.43	≥ 0.43
Maximum Vargha and Delaney's A	0.56 – < 0.64	0.64 – < 0.71	≥ 0.71
	> 0.34 – 0.44	> 0.29 – 0.34	≤ 0.29

FRIEDMAN TEST

→The Friedman test determines if there are differences among groups for two-way data structured in a specific way, namely in an unreplicated complete block design. In this design, one variable serves as the treatment or group variable, and another variable serves as the blocking variable.

→It is the differences among treatments or groups that we are interested in. We aren't necessarily interested in differences among blocks, but we want our statistics to take into account differences in the blocks. In the unreplicated complete block design, each block has one and only one observation of each treatment.

FRIEDMAN TEST

Appropriate data

→Two-way data arranged in an unreplicated complete block design

→Dependent variable is ordinal, interval, or ratio

→Treatment or group independent variable is a factor with two or more levels. That is, two or more groups

→Blocking variable is a factor with two or more levels

→Blocks are independent of each other and have no interaction with treatments

Hypotheses

→Null hypothesis: The medians of values for each group are equal.

→Alternative hypothesis (two-sided): The medians of values for each group are not equal.

ALIGNED-RANKS TRANSFORMATION ANOVA

→Aligned ranks transformation ANOVA (ART anova) is a nonparametric approach that allows for multiple independent variables, interactions, and repeated measures.

→Since the aligning process requires subtracting values, the dependent variable needs to be interval in nature. That is, strictly ordinal data would be treated as numeric in the process.

ALIGNED-RANKS TRANSFORMATION ANOVA

→All independent variables must be nominal

→All interactions of fixed independent variables need to be included in the model

→Post-hoc comparisons can be conducted for main effects

→For interactions effects, post-hoc comparisons can be conducted for two-way models. For more complex models, difference-in-difference post-hoc comparisons can be conducted.

→For fixed-effects models, eta-squared can be calculated as an effect size