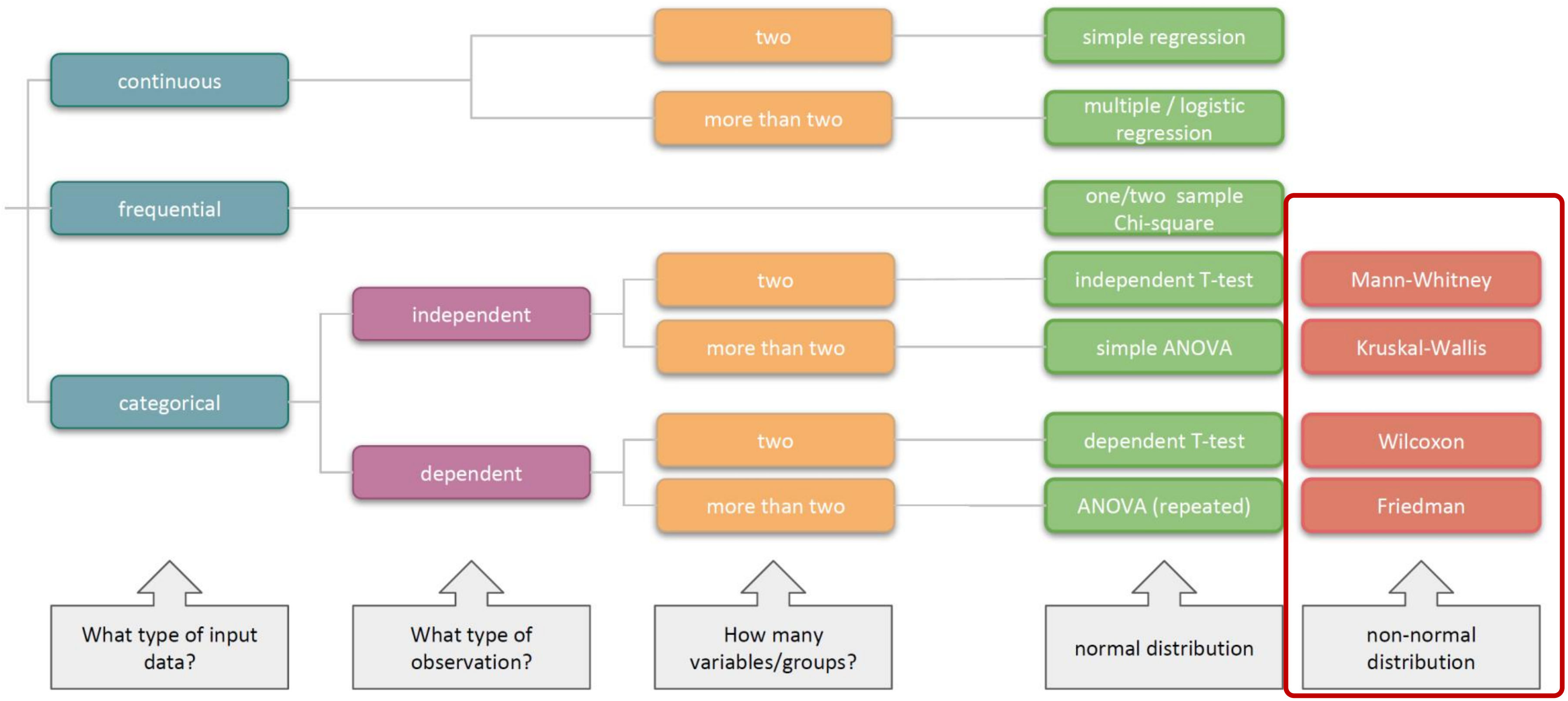
Quality and Usability Seminar: Statistics

TOPIC 10: Non-parametric tests

Agenda

1. Recap of Method Selection
2. Mann-Whitney U Test
3. Possible Research Questions
4. Task of the Week

Decision Tree



Assumption

Parametric

- Normally distributed data
- Homogeneity of variance
- Interval level
- Data independence

Non-parametric

- ~~Normally distributed data~~
- ~~Homogeneity of variance~~

☺

Mann-Whitney U Test

Usage:

two groups, **independent** → counterpart of independent t-test

Assumptions:

- Measurement at least at ordinal level, two independent variables**,** similar shape of distributions
- When reporting results: if yes, report medians, else report mean ranks

General procedure (underlying math):

1. Put the data of both group together, and calculate the rank of each item
2. For each group, calculate the sum of ranks for their items
3. Check the sum of the rank of group1, against U statistics (no matter which one is group1)

Report the effect size as well.

An example…

How depressant effects of certain recreational drugs?

10 people got ecstasy tablet to take on Saturday night, and 10 were allowed to only drink alcohol.

Level of depression is measure by BDI (an inventory) after a day.

Note: Man-Whitney test and **Wilcoxon rank-sum** test are very similar, Here we use the Wilcoxon rank-sum test

TABLE 15.1 Data for drug experiment

Participant	Drug	BDI (Sunday)
1	Ecstasy	15
2	Ecstasy	35
3	Ecstasy	16
4	Ecstasy	18
5	Ecstasy	19
6	Ecstasy	17
7	Ecstasy	27
8	Ecstasy	16
9	Ecstasy	13
10	Ecstasy	20
11	Alcohol	16
12	Alcohol	15
13	Alcohol	20
14	Alcohol	15
15	Alcohol	16
16	Alcohol	13
17	Alcohol	14
18	Alcohol	19
19	Alcohol	18
20	Alcohol	18

From Andy Field’s book

Sunday Data

Score	13	13	14	15	15	15	16	16	16	16	17	18	18	18	19	19	20	20	27	35
Potential Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Actual Rank	1.5	1.5	3	5	5	5	8.5	8.5	8.5	8.5	11	13	13	13	15.5	15.5	17.5	17.5	19	20
Group	A	E	A	A	A	E	A	A	E	E	E	E	A	A	E	A	E	A	E	E

Sum of Ranks for Alcohol (A) = 90.5 Sum of Ranks for Ecstasy (E) = 119.5

Test statistic: the lowest number: Ws = 90.5

How to determine whether this test statistic is significant?

1. Calc mean (W) and std. error (SE) of test:

$$\overline{W}_s = \frac{n_1(n_1+n_2+1)}{2} \qquad \overline{W}_s = \frac{10(10+10+1)}{2} = 105$$

$$SE_{\overline{W}_s} = \sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}} \qquad SE_{\overline{W}_s} = \sqrt{\frac{(10\times 10)(10+10+1)}{12}} = 13.23$$

From Andy Field's book

2. Convert the test statistic to z-score

$$z = \frac{X - \overline{X}}{s} = \frac{W_s - \overline{W}_s}{SE_{\overline{W}_s}}$$

$$z_{\text{Sunday}} = \frac{W_s - \overline{W}_s}{SE_{\overline{W}_s}} = \frac{90.5 - 105}{13.23} = -1.10$$

3. If the absolute value is bigger than 1.96, the test is significant at $p < .05$

→ Not significant

Wilcoxon signed-rank test

Usage:

two groups, **dependent** → counterpart of dependent t-test

Assumptions:

Measurement at least at ordinal level, two groups**,** Delta (g1-g2) has a symmetrical shape (if not; transfer data)

General procedure (underlying math):

1. Calculate the difference *d*, and rank |*d*| (neglect 0s)
2. Sum the ranks for positive ones (d>0) and negative ones (d<0) → S+ and S-
3. The test statistic is the min(S+,S-)

Report the effect size as well.

Kruskal-Wallis Test

Usage:

several groups, **independent** → simple ANOVA

Assumptions:

Measurement at least at ordinal level, +two independent variable**,** independence of observations, same shape? Yes: compare median, no: compare mean ranks

General procedure (underlying math):

1. Similar to Mann-whitney, calc the sum of the ranks for each group
2. Calculate the test statistics from formula (H)
3. Compare with chi-square distribution (degree of freedom: n_groups-1)

$$H = \frac{12}{n(n+1)} \sum_h \frac{S_h^2}{n_h} - 3(n+1)$$

S_h Sum of the rank for group h
n: total sample size
n_h number of sample in group h

Quality & Usability Lab

Kruskal-Wallis Test

Post hoc test:

compare each two pair with Mann-whitney U test

BUT lost of tests leads to inflate the Type I error

To avoid that:

- Use Bonferroni adjustment: i.e. For N comparisons $\alpha = \frac{0.05}{N}$
- Do not make many comparisons, e.g. just all groups against control?

Friedman Test

Usage:

several groups, **dependent** → one-way repeated measure ANNOVA

Assumptions:

One group is measured +3 times, at least ordinal level,

$$\left[\frac{12}{nk(k+1)}\sum_{i=1}^{k} R_i^2\right] - 3n(k+1)$$

Guides

Kruskal Wallis Test:
https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php
https://sixsigmastudyguide.com/kruskal-wallis-non-parametric-hypothesis-test/
https://www.youtube.com/watch?v=YB5Eza6j-Lo

Friedman Test:
https://statistics.laerd.com/spss-tutorials/friedman-test-using-spss-statistics.php

Non-parametric tests in general:
https://www.youtube.com/watch?v=pWEWHKnwg_0

4. Task of the week

Task of the week (delivery: 11.02.2020)

Provide a script and html file which calulates the suitable non-parametric test to answer the following research questions (RQ). Please also report the results as a text conclusion including the test statistic value with degree of freedom, significance value as well as pairwise comparisions.

DB01

1.RQ: Does changing the ***bitrate* (independent variable: 2000, 4000, 6000, 50000 kbps)** have a significant effect on the *video quality* (VQ) ratings (dependent variable). Please consider **all ratings at a resolution of 1080p and framerate of 60 fps for the FIRST game**. Use the ratings provided in the gaming video quality dataset.

DB04

2.RQ: Does the *testStimulus* (independent variable) have a significant influence on speech quality ratings (dependent variable). Please assume that each of the six files was assessed by a different set of test participants **and only use the ratings of the first repetition.** Use the quality ratings provided in the dataset “speech_quality_repetition_dataset”.

Webpage (delivery: 11.02.2020)

Two pages: Kruskal-Wallis H and Friedman Test

For each

- Requirements to use the test (assumptions)
- Difference to ANOVA (procedure)
- When to use the test: examples
- How to report the results (test statistic, significance values, effect size, pair-wise comparisons