Non-Parametric Analysis

Please download:
- ***relationships.sav***
- ***Field-Looks_Charis.sav***

27 Nov 2009
CPSY501
Dr. Sean Ho
Trinity Western University

Outline for today

- Non-Parametric Between-subjects Tests:
  - 2 groups: Mann-Whitney U
  - Many groups: Kruskall-Wallis H
  - Factorial variants
- Non-Parametric Within-subjects Tests:
  - 2 repetitions: Wilcoxon Signed-rank
  - Many repetitions: Friedman's ANOVA

- There exist other non-parametric options, too!

When To Use Non-Parametric

- Review: assumptions of ANOVA?
  - If met, parametric methods more power
  - ANOVA is pretty robust to non-normality
  - Try transforming data, or multi-level model
- Consider non-parametric alternatives if:
  - DV is ordinal, or
  - Violations are severe enough to affect result
- Non-parametric often use ranks, not raw scores
  - e.g., Spearman vs. Pearson correlation
  - Comparing medians rather than means

Mann-Whitney U Test

- DV: ordinal or scale, but non-parametric
- IV: dichotomous, between-subjects (2 groups)
- RQ: Is there a median difference between the two groups?
  - Non-parametric alternative to *t*-test
- What research designs might need this test?
- Data entry format:
  2 variables entered for each participant:
  - IV: group membership (0 or 1)
  - DV: outcome measure

Mann-Whitney U: SPSS

- Dataset: relationships.sav
- Analyze → Nonparametric → 2 Independent:
  - Test Variable: DV (quality of communication)
  - Grouping Variable: IV (having counselling)
  - Use "Define Groups" to choose two groups for comparison: 1= no counsel; 2 = had counsel
  - Test Type: "Mann-Whitney U"
- Optional SPSS module for Exact Tests: Computationally-expensive but more precise, especially for small / unbalanced designs

Mann-Whitney U: Output

"There was no significant effect of Having Counselling on Quality of Communication, $U = 311.00$, $p = .059$, $\text{Mdn}_H = 4.0$, $\text{Mdn}_N = 3.0$."

- Use Analyze → Descriptives → Explore to get group medians

Test Statistics[a]

	Quality of Communcation
Mann-Whitney U	**311.000**
Wilcoxon W	1014.000
Z	-1.890
Asymp. Sig. (2-tailed)	**.059**

a. Grouping Variable: Had any counselling

Mann-Whitney U: Effect Size

- Effect size must be calculated manually, using the following formula:
  - $r = Z / \sqrt{N}$
  - e.g., $r = -1.89 / \sqrt{60} \approx -0.24499$
- Use existing research or Cohen's effect size "estimates":
  - "There is a small difference between the therapy and no therapy groups, *r* = -.24"

Practise: Mann-Whitney U

- You try it!
- Is there a significant difference between spouses who report communication problems and spouses who have not (“Com_prob”), in terms of the level of conflict they experience (“Conflict”)?
- What is the size of the effect?

Kruskal-Wallis Test

- DV: ordinal or scale, but non-parametric
- IV: categorical, between-subjects (many groups)
- RQ: Is there a median difference amongst the groups?
  - Alternative to One-way ANOVA
- What research designs might need this test?
- Data entry format:
  2 variables entered for each participant:
  - IV: group membership
  - DV: outcome measure

Kruskal-Wallis: SPSS

- Analyze → Nonparametric → K Independent:
  - Test Variable: DV (level of conflict)
  - Grouping Variable: IV (type of counselling)
  - Test Type: "Kruskal-Wallis H"
  - Define Groups: specify highest and lowest group numbers to be compared
- Optional SPSS module for Exact Tests: Computationally-expensive but more precise, especially for small / unbalanced designs

Kruskal-Wallis: Output

"Type of counselling has a significant effect on participants' level of conflict,
$\chi^2(2) = 7.09$, $p = .029$."

- Also report medians and post hoc results…

Test Statistics[a,b]

	Level of Conflict
Chi-Square	7.094
df	2
Asymp. Sig.	.029

a. Kruskal Wallis Test

b. Grouping Variable: Type of Counselling

Kruskal-Wallis: Follow-up

- K-W test is omnibus: if significant, follow-up with Mann-Whitney tests on pairs of groups
  - Must do Bonferroni correction manually, or else Type 1 error inflates: lower the level of significance to 0.05 / (# comparisons)
- Try planned contrasts based on
  - Theory / literature,
  - Your research question; or
  - Ordering groups by mean ranks, and comparing each group to next highest group

Kruskal-Wallis: Effect Size

- Overall effect size not so useful; instead,
- Calculate effect size for each pair of groups that differs significantly (in Mann-Whitney follow-up)
  - $r = Z / \sqrt{n}$
  - Where the sample size n is the total number of participants in this pair of groups

Kruskal-Wallis: Reporting

- "Type of Counselling has a significant effect on participants' level of conflict, $\chi^2(2) = 7.09$, $p = .029$."
- "Specifically, the No Counselling group had higher conflict scores, $Mdn_N = 4.0$, than did the Couples Counselling group, $Mdn_C = 3.0$, Mann-Whitney $U = 176.5$, $Z = -2.61$, $p = .009$, $r = -.37$.
- Field uses "H" for the K-W: $H(2) = 7.09$
- Note: Bonferroni correction: $\alpha = .05 / 3 \approx .017$

TRINITY WESTERN UNIVERSITY

Kruskal-Wallis: Non-signifcant

- If Kruskal-Wallis gives a non-significant result, but the research question behind the analysis is still “important”:
- The problem might be with low power, so
- Descriptive follow-up analyses can be helpful.
- See the illustration for Friedman’s ANOVA below for some clues.

Parametric vs. Non-Parametric

- There's nothing wrong with running both parametric and non-parametric tests!
- Comparison of non-parametric tests with the corresponding ANOVA may be able to lend more confidence in the overall adequacy of the patterns reported.
- Nonparametric analyses tend to have less power for well-distributed DVs, but they can be more sensitive to effects when the DV is, for instance, truly bimodal!

TRINITY WESTERN UNIVERSITY

Param vs. Non-Param: ex.

- If we do the same analysis parametrically:
  - DV: Level of Conflict
  - IV: Type of Counselling
  - One-way ANOVA, with Levene's test and Bonferroni post-hoc
- Results are similar: $F(2, 57) = 4.05$, $p = .023$
  - The No Counselling group shows more conflict than the Couples Counselling group, $M_N = 3.87$ and $M_C = 3.04$
- Confirms results with non-parametric analysis

TRINITY WESTERN UNIVERSITY

Factorial Between-Subjects

- SPSS doesn't have factorial between-subjects non-parametric analyses built-in!
- Try creating one IV that encodes all cells: e.g., 2x3 factorial → 6-level one-way, use K-W
  - (-) Follow-up gets hard with lots of groups
  - (-) Moderation/interaction analysis is harder
  - (+) Flexibility in defining groups
  - (+) Target specific cells / interactions
- Or: Run separate K-W tests for each IV
- Or: Convert to ranks and use log-linear analysis

Factorial Analysis: Example

- Research question: How do Marital Status and Type of Counselling relate to Conflict Levels?
- Check cell sizes with Analyze → … → Crosstabs:
  - The smallest cells (Individual Counselling) have 5-6 people per group, that is enough
- 2x3 factorial → convert to 1 IV with 6 levels:
  - Indiv. Counsel & Married; Indiv & Divorced; Couples Couns. & Marr; Couple & Div; etc.
  - Transform → Recode into Different, try the "If" conditions

TRINITY WESTERN UNIVERSITY

Factorial Analysis: Output

- The Kruskall-Wallis test for the combined variable is not significant.
- This suggests that the significant effect for Counselling Type is masked when combined with Marital Status.

Test Statistics(a,b)

	Level of Conflict
Chi-Square	8.753
df	5
Asymp. Sig.	.119

a Kruskal Wallis Test
b Grouping Variable: Counselling Type & Marital Status

Factorial Analysis: Main effect

- The idea of a “masking effect” of Marital Status shows as well when we test that main effect alone.

Test Statistics(a)

	Level of Conflict
Mann-Whitney U	337.000
Wilcoxon W	802.000
Z	-1.752
Asymp. Sig. (2-tailed)	.080

a Grouping Variable: Marital Status

Theory-Guided Interaction Test

- Divorced & No counselling group:
  - We might assume this group should have high conflict levels!
  - Compare with some of the other 5 groups using Mann-Whitney U tests
- This can be a “theoretically guided” replacement for interaction tests in non-parametric factorial analysis.
- The choice depends on conceptual relations between the IVs.

Kruskall-Wallis: Practise

- Does number of children (range 0-3) have a significant effect on quality of marital communication?

TRINITY WESTERN UNIVERSITY

Wilcoxon Signed-rank Test

- DV: ordinal or scale, but non-parametric
- IV: within-subjects, with 2 repetitions
- RQ: Is there a median change between times?
  - Parallel to paired-samples *t*-test
- What research designs might need this test?
- Data entry format:
  2 variables entered for each participant:
  - DV at first measurement
  - DV at second measurement

Wilcoxon Signed-rank: SPSS

- Analyze → Nonparametric → 2 Related Samples:
  - Test Pairs: DV before and DV after
  - Test Type: Wilcoxon
- Practise: does level of conflict decrease from pre-therapy (Pre-conf) to post-therapy (Conflict)?
- Multiple pairs may be specified in one analysis
  - But no built-in Bonferroni correction
- Exact tests also available with add-on module

Wilcoxon Signed-rank: Output

- There was a significant reduction in level of conflict after therapy, $T = 4.5$, $p = .002$.
  - Or: $Z = -3.09$, $p = .002$ (& calc. effect size)

Ranks

		N	Mean Rank	Sum of Ranks
Pre-therapy level of Conflict - Level of Conflict	Negative Ranks	1[a]	4.50	4.50
	Positive Ranks	13[b]	7.73	100.50
	Ties	46[c]		
	Total	60		

a. Pre-therapy level of Conflict < Level of Conflict

b. Pre-therapy level of Conflict > Level of Conflict

c. Pre-therapy level of Conflict = Level of Conflict

	Pre-therapy level of Conflict - Level of Conflict
Z	-3.094[a]
Asymp. Sig. (2-tailed)	.002

a. Based on negative ranks.

b. Wilcoxon Signed Ranks Test

TRINITY WESTERN UNIVERSITY

Wilcoxon Signed: Effect Size

- Effect size must be calculated manually, using the following formula:
  - $r = Z / \sqrt{N}$
  - e.g., $r = -3.09 / \sqrt{120} \approx -0.28$
- N is total number of observations, not participants! (60 participants) * (2 observation)
- Use existing research or Cohen's effect size "estimates":
  - "The reduction in level of conflict after therapy was significant but small, *r* = -.28"

Wilcoxon Signed-rank: Practise

- Do levels of conflict change significantly between pre-therapy (Pre_conf) and 1 year after therapy (Follow_conf)?
- If so, calculate the size of the effect.
  - Participant attrition at time 3 (Follow_conf) affects the total number of observations!
- Example reporting: “There was a significant reduction in level of conflict after therapy, $T = 4.5$ [or: $Z = -3.09$], $p = .002$, $r = -.28$.”

TRINITY WESTERN UNIVERSITY

Friedman's ANOVA

- DV: ordinal or scale, but non-parametric
- IV: within-subjects, several repetitions
- RQ: Is there a median change over time?
  - Parallel to repeated-measures ANOVA
- What research designs might need this test?
- Data entry format:
  one variable for each repetition of the measure
  - DV at first measurement
  - DV at second measurement
  - DV at third measurement, etc.

Friedman's ANOVA: SPSS

- Analyze → Nonparametric → K Related Samples:
  - Test Variables: each repetition of DV
  - Test Type: "Friedman"
- No easy way in SPSS to do non-parametric mixed-design analysis
- Optional SPSS module for Exact Tests: Computationally-expensive but more precise, especially for small / unbalanced designs

Friedman's ANOVA: Output

- "Levels of conflict changed significantly over time, $\chi^2(2, N = 57) = 9.07$, $p = .011$."
- "Specifically…" *[report post hoc results…]*

Test Statistics[a]

N	57
Chi-Square	9.065
df	2
Asymp. Sig.	.011

a. Friedman Test

Friedman's ANOVA: Follow-up

- If Friedman’s is significant, follow-up with a series of Wilcoxon Signed-ranks tests:
- Either do all pairs of levels of the RM (post-hoc):
  - Remember manual Bonferroni correction: α = .05 / (# comparisons)
  - But Bonferroni is often too conservative
- Or target specific planned comparisons (better):
  - e.g., (time1 vs. time2), (time2 vs. time3), … k-levels of IV → only k-1 comparisons
  - Useful if power is low, e.g. many cells: control Type II error

Friedman's ANOVA: Example

- IV: Time (3 cells)
- Targeted follow-up analyses:
  - Run 3 Wilcoxon’s signed-rank tests
  - Bonferroni correction: $\alpha = .05 / 3 \approx .017$
- Pre vs. Post: $Z = -3.09$, $p = .002$, $r = -.28$
- Pre vs. 1yr: $Z = -2.44$, $p = .015$, $r = -.22$
- Post vs. 1yr: not significant
- Conclusion: improvement after therapy is maintained at the follow-up assessment.

Friedman's ANOVA: Reporting

- "Levels of conflict significantly changed over time, χ^2 (2, N = 57) = 9.07, p = .011.
- "Specifically, conflict decreased from pre-therapy levels at post-therapy observations, Z = -3.09, p = .002, r = -.28,
- "and levels remained below pre-therapy conflict levels one year later, Z = -2.44, p = .015, r = -.22."

Friedman's: Non-significance

- If Friedman’s is not significant:
  - Is it because of low power (Type II error)?
- Run a series of Wilcoxon Signed-ranks tests, but
  - Focus on effect sizes, not significance levels
- If the effect sizes are “moderate”, say > .25, then the results could be worth reporting.
- Enough detail should be reported to be useful for future meta-analyses.
- This could be true of any analysis;
  we just use Friedman's to illustrate

Friedman's ANOVA: Practise

- Dataset: Field-Looks_Charis.sav
  - “Looks or Personality” (Field text)
  - Two RM IVs in the dataset
- RQ: Is there a significant difference between participants’ judgements of people who are of average physical appearance,
  but present as dull (“ave_none”);
  somewhat charismatic (“ave_some”),
  or as having high charisma (“ave_high”)?
- If so, conduct follow-up tests to identify where the specific differences lie.

Summary: Between-Subjects

Parametric	Non-Parametric
t-test between groups (1 IV, dichotomous)	Mann-Whitney *U* test
One-way ANOVA (1 IV, many levels) • Post-hoc: *t*-tests	Kruskal-Wallis *H* test • Post-hoc: M-W
Factorial ANOVA (several IVs) • Post-hoc: *t*-tests	K-W with encoding of cells in one IV • Post-hoc: M-W

Summary: Within-Subjects

Parametric

- Paired/related *t*-test (1 RM IV w/2 levels)
- Repeated-ms ANOVA (1 RM IV with many levels)
  - Post-hoc: paired *t*-tests between levels of RM

Non-Parametric

- Wilcoxon Signed-Rank
- Friedman's ANOVA
  - Post-hoc: Wilcoxon Signed-rank between levels of the RM