Non-Parametric Between-Participants ANOVA

PSYC5141: Experimental Methods for Psychology

Mark Hurlstone, Lancaster University

Week 4, Lecture 1

Learning Objectives

- **Introduction to non-parametric tests**
  - When and why we might use non-parametric alternatives to ANOVA
  - How rank-based tests work
- **The Kruskal–Wallis test**
  - The logic of the test
  - Calculating and interpreting the test statistic
- **Follow-up comparisons**
  - Planned comparisons and post-hoc tests
  - Correcting for multiple comparisons
- **Interpreting and consolidating results**
  - Interpreting Kruskal–Wallis and follow-up comparisons
  - Comparing Kruskal–Wallis with parametric ANOVA

Why and when do we use non-parametric tests?

When might between-participants ANOVA be inappropriate?

Single-factor between-participants ANOVA assumes:

- Independence of observations
- **Normality:** Scores are approximately normally distributed within each group
- **Homogeneity of variance:** Population variances are equal across groups

What can we do when these assumptions are not well met?

What are non-parametric tests?

- Non-parametric tests are alternatives to parametric tests such as t-tests and ANOVA
- They are often useful when data are markedly non-normal or when the dependent variable is ordinal
- Many non-parametric tests work with **ranks** rather than the original scores
- They are more flexible, but they may be less informative about the original metric of the data

- **Parametric ANOVA** → compares group means
- **Kruskal–Wallis** → compares rank distributions across groups

From observations to ranks

A simple example ...

Observation	Score	Rank
A	23	3
B	31	4
C	2	1
D	4	2

Key points:

- smallest score gets rank 1
- next smallest gets rank 2, and so on
- tied scores receive the average of their ranks
- ranks preserve **order**, but not the original distances between scores

The Kruskal–Wallis test

Same topic, different design

- In the last lecture, we used a misinformation-warning example in a **within-participants** design
- In this lecture, we are using the same topic, but we are now assuming a **between-participants** design:
  - each participant is assigned to **one** warning condition only
  - the three groups are **independent**
  - this allows us to illustrate the Kruskal–Wallis test

A Between-Participants Experiment

- A psychologist investigates whether misinformation warnings improve people's ability to distinguish between true and false information on social media
- **Independent variable:** Type of misinformation warning, with three levels:
  - A_1: No warning
  - A_2: Generic warning
  - A_3: Specific warning
- **Dependent variable:** Percentage of social media posts correctly classified as true or false
- Design: Fifteen participants, five participants per condition

Example data



A_1 No warning	A_2 Generic warning	A_3 Specific warning
20	42	72
25	47	77
30	52	82
35	57	87
85	92	97

Why might we consider a non-parametric test?

- The no-warning and generic-warning groups both exhibit positively skewed distributions
- Both groups contain an unusually high score relative to their remaining observations
- With small samples, extreme observations can have a substantial influence on group means and variance estimates
- A rank-based alternative to ANOVA may therefore be worth considering



A_1 No warning	A_2 Generic warning	A_3 Specific warning
20	42	72
25	47	77
30	52	82
35	57	87
85	92	97

The logic of Kruskal–Wallis

- If the groups do not differ, we would expect them to have **similar average ranks**
- If one condition tends to produce higher scores, that condition should also receive **higher ranks**
- So the Kruskal–Wallis test asks:

> ***“ Are the group rank totals more different than we would expect by chance?***

Step 1: Rank all of the observations

Score	Group	Rank
20	A_1	1
25	A_1	2
30	A_1	3
35	A_1	4
42	A_2	5
47	A_2	6
52	A_2	7
57	A_2	8
72	A_3	9
77	A_3	10
82	A_3	11
85	A_1	12
87	A_3	13
92	A_2	14
97	A_3	15

Step 2: Return the ranks to the original groups

Group	Scores	Ranks	Sum of ranks R_j	Mean rank
A_1	20, 25, 30, 35, 85	1, 2, 3, 4, 12	22	4.40
A_2	42, 47, 52, 57, 92	5, 6, 7, 8, 14	40	8.00
A_3	72, 77, 82, 87, 97	9, 10, 11, 13, 15	58	11.60

Checking the sum of ranks:

The sum of the ranks from 1 to N is given by:

$$\sum_{i=1}^{N} i = \frac{N(N+1)}{2}$$

Where N is the total number of observations

For our example:

$$\frac{15(15+1)}{2} = \frac{15(16)}{2} = 120$$

We can then check this against the sum of the ranks across our three groups:

$$22 + 40 + 58 = 120$$

The Kruskal–Wallis formula

“ Square each group's rank total, divide by its sample size, and sum across groups.

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(N+1)$$

Symbols:

- N = total sample size
- k = number of groups
- R_j = sum of ranks for group j
- n_j = sample size for group j

For our example:

- $N = 15$
- $k = 3$
- $R_1 = 22,\ R_2 = 40,\ R_3 = 58$
- $n_1 = n_2 = n_3 = 5$

The Kruskal–Wallis formula

“ Scales the sum of the squared rank totals, accounting for the total sample size.

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(N+1)$$

Symbols:

- N = total sample size
- k = number of groups
- R_j = sum of ranks for group j
- n_j = sample size for group j

For our example:

- $N = 15$
- $k = 3$
- $R_1 = 22,\ R_2 = 40,\ R_3 = 58$
- $n_1 = n_2 = n_3 = 5$

Worked example: substituting values

For our example:

- $N = 15$
- $k = 3$
- $R_1 = 22,\ R_2 = 40,\ R_3 = 58$
- $n_1 = n_2 = n_3 = 5$

$$H = \frac{12}{15(16)}\left(\frac{22^2}{5} + \frac{40^2}{5} + \frac{58^2}{5}\right) - 3(16)$$

Worked example: substituting values

For our example:

- $N = 15$
- $k = 3$
- $R_1 = 22,\ R_2 = 40,\ R_3 = 58$
- $n_1 = n_2 = n_3 = 5$

$$H = \frac{12}{15(16)}\left(\frac{22^2}{5} + \frac{40^2}{5} + \frac{58^2}{5}\right) - 3(16)$$

Worked example: substituting values

For our example:

- $N = 15$
- $k = 3$
- $R_1 = 22,\ R_2 = 40,\ R_3 = 58$
- $n_1 = n_2 = n_3 = 5$

$$H = \frac{12}{15(16)}\left(\frac{22^2}{5} + \frac{40^2}{5} + \frac{58^2}{5}\right) - 3(16)$$

Worked example: substituting values

For our example:

- $N = 15$
- $k = 3$
- $R_1 = 22,\ R_2 = 40,\ R_3 = 58$
- $n_1 = n_2 = n_3 = 5$

$$H = \frac{12}{15(16)}\left(\frac{22^2}{5} + \frac{40^2}{5} + \frac{58^2}{5}\right) - 3(16)$$

Worked example: substituting values

For our example:

- $N = 15$
- $k = 3$
- $R_1 = 22,\ R_2 = 40,\ R_3 = 58$
- $n_1 = n_2 = n_3 = 5$

$$H = \frac{12}{15(16)}\left(\frac{22^2}{5} + \frac{40^2}{5} + \frac{58^2}{5}\right) - 3(16)$$

Worked example: substituting values

For our example:

- $N = 15$
- $k = 3$
- $R_1 = 22,\ R_2 = 40,\ R_3 = 58$
- $n_1 = n_2 = n_3 = 5$

$$H = \frac{12}{15(16)}\left(\frac{22^2}{5} + \frac{40^2}{5} + \frac{58^2}{5}\right) - 3(16)$$

$$H = \frac{12}{240}\left(96.8 + 320 + 672.8\right) - 48$$

$$H = \frac{12}{240}(1089.6) - 48$$

$$H = 54.48 - 48 = \boxed{6.48}$$

Interpreting the omnibus result

$$df = k - 1 = 3 - 1 = 2$$

$$6.48 > 5.991$$

$$H(2) = 6.48,\ p < .05$$

Chi-squared critical values			
df	α = .10	α = .05	α = .01
1	2.706	3.841	6.635
2	4.605	**5.991**	9.210
3	6.251	7.815	11.345
4	7.779	9.488	13.277
5	9.236	11.070	15.086

Upper-tail critical values. Highlighted: df = 2, α = .05.

Effect size

- We use the rank-based eta-squared estimate:

$$\eta_H^2 = \frac{H - k + 1}{N - k}$$

- Then substitute the values:

$$\eta_H^2 = \frac{6.48 - 3 + 1}{15 - 3}$$

$$\eta_H^2 = \frac{4.48}{12} = 0.37$$

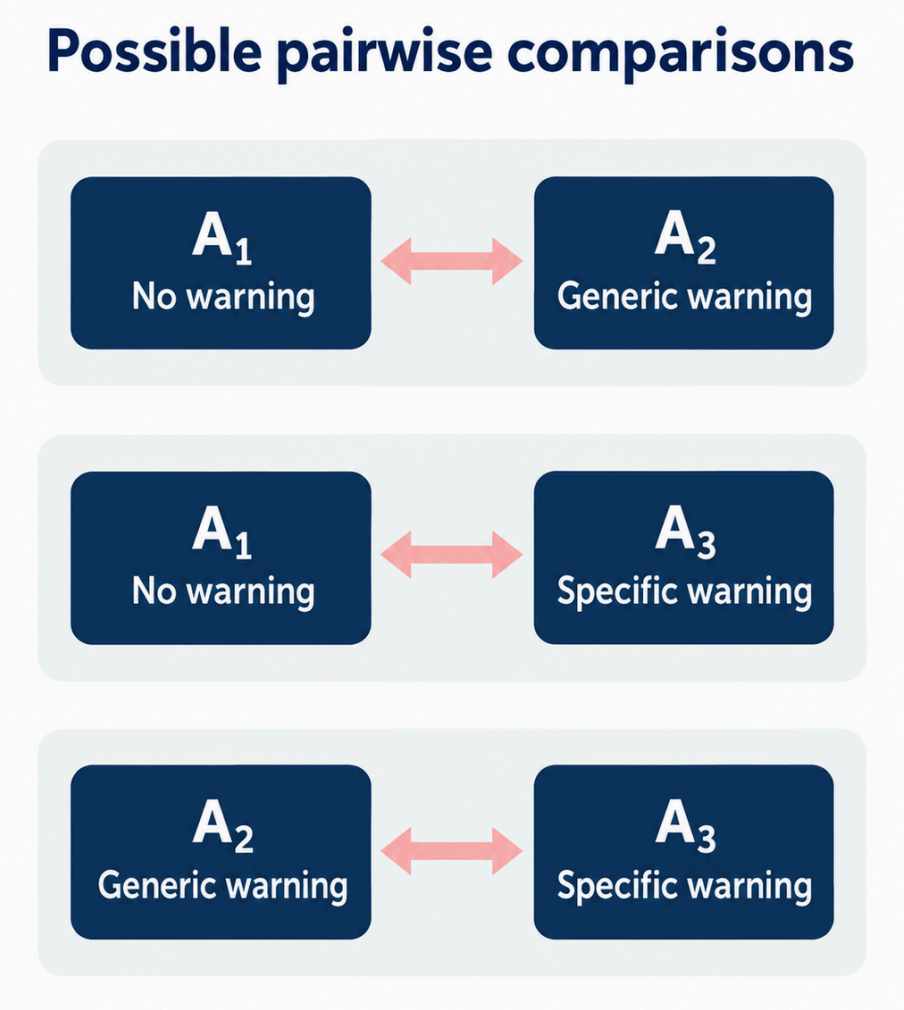
Effect size	**Interpretation**
.01 to less than .06	Small
.06 to less than .14	Moderate
.14 or greater	Large

Our result of $\eta_H^2 = .37$ would therefore be classified as a **large effect**

Follow-up comparisons

Why do we need follow-up comparisons?

- The Kruskal–Wallis test is an omnibus test: it evaluates whether there are differences among the group distributions
- However, a statistically significant result does not tell us **which groups differ**
- As with parametric ANOVA, we can investigate specific group differences using planned comparisons or post-hoc tests



Planned comparisons: The Wilcoxon rank-sum test

The Wilcoxon rank-sum test (also known as the Mann–Whitney U test) is a non-parametric alternative to the independent-samples t-test

- Compares two independent groups
- Combines the observations from both groups and ranks them from smallest to largest
- Examines whether the ranks tend to be higher in one group than the other
- Can be used to test specific hypotheses established before examining the data

Example: We might predict that participants receiving a specific warning (A_3) will achieve higher accuracy scores than those receiving no warning (A_1)

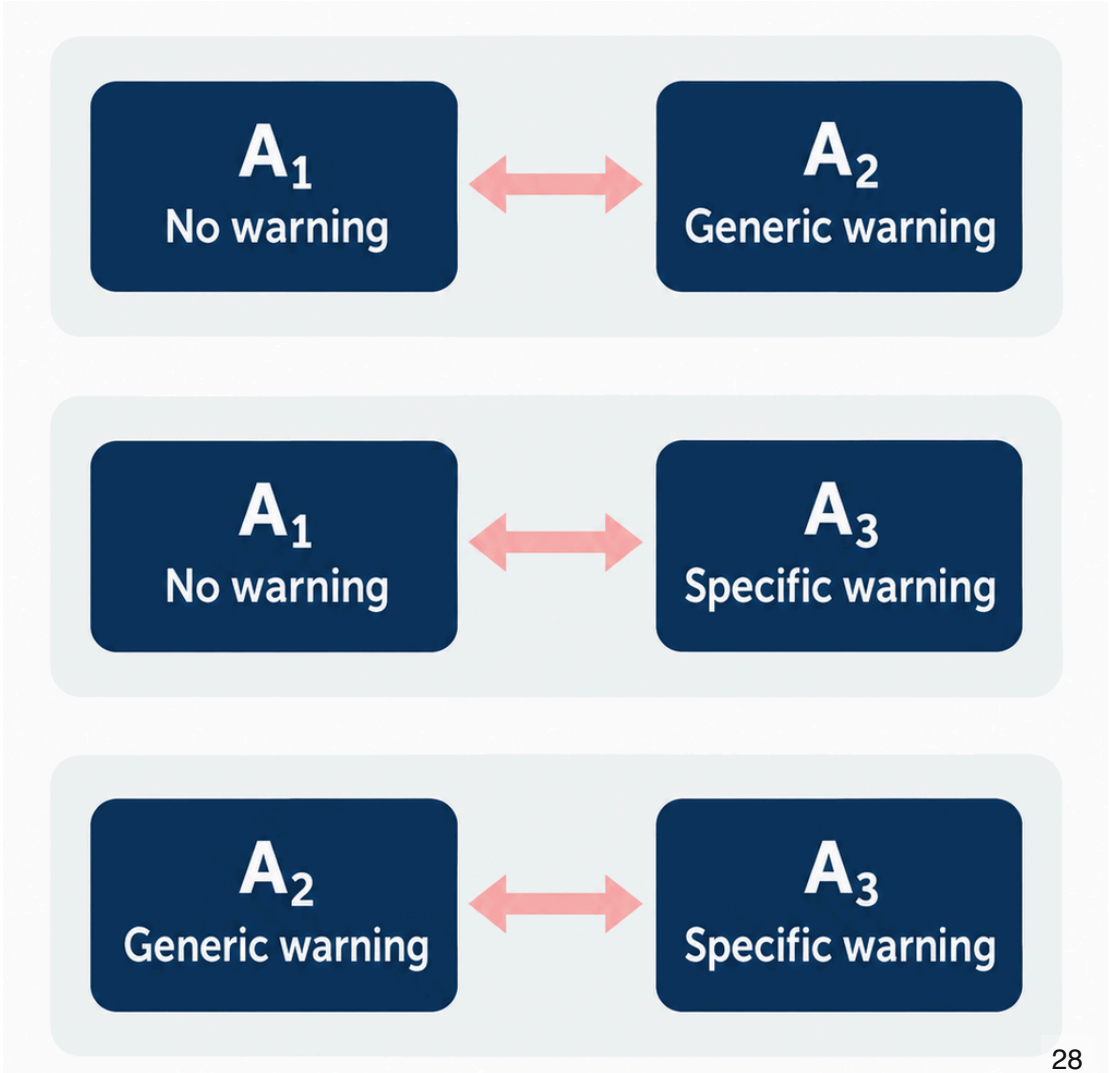
- We could therefore specify A_1 versus A_3 as a planned comparison

Post-hoc comparisons: Dunn's test

When we want to investigate differences between multiple pairs of groups, we can use Dunn's test

- A rank-based procedure commonly used for pairwise comparisons following Kruskal–Wallis
- Compares the mean ranks of two groups
- Uses the ranks calculated across all groups, rather than ranking each pair separately
- Produces a test statistic and *p*-value for each comparison

For our example, we could compare (see right):



Correcting for multiple comparisons

Conducting multiple statistical tests increases the probability of making at least one Type I error

As with parametric ANOVA, we can control the familywise error rate by adjusting for multiple comparisons

Two common methods are:

- **Bonferroni correction:** Controls the familywise error rate by dividing the significance level by the number of comparisons (or equivalently adjusting *p*-values)
- **Holm correction:** Applies a sequential adjustment to *p*-values, generally providing greater statistical power than Bonferroni while still controlling the familywise error rate

For our example, we will use Holm-adjusted *p*-values

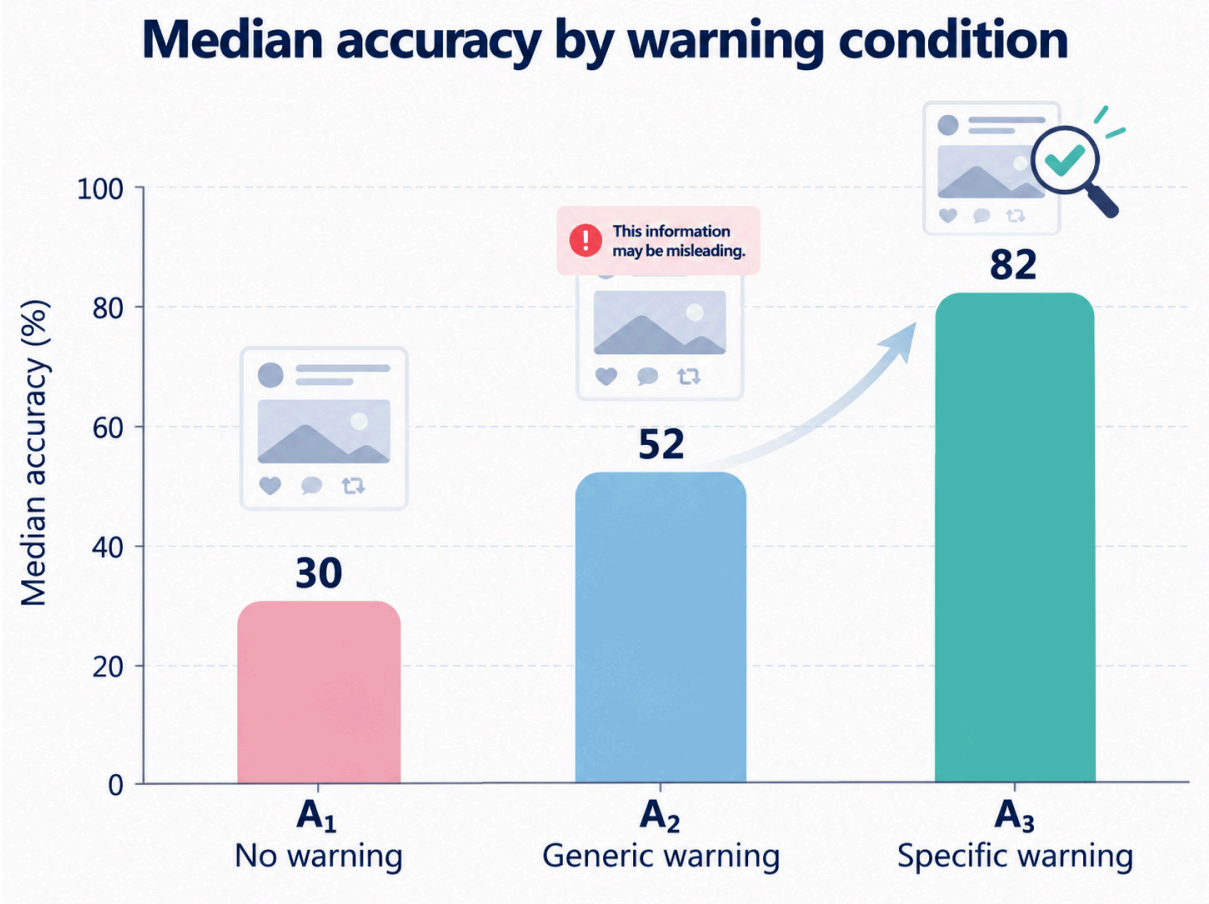
Interpreting follow-up comparisons

A significant Kruskal–Wallis test indicated differences among the warning conditions:

$$H(2) = 6.48, \quad p = .039$$

We conducted pairwise comparisons using Dunn's test with Holm correction

Comparison	Adjusted p	Significant?
A_1 vs A_2	.406	No
A_1 vs A_3	.033	Yes
A_2 vs A_3	.406	No



Between-Participants ANOVA versus Kruskal–Wallis

Feature	Between-Participants ANOVA	Kruskal–Wallis
Design	Independent groups	Independent groups
Data analysed	Original scores	Ranks of scores
Comparison	Group means	Group rank distributions
Normality	Assumes normally distributed scores within groups	Does not require normality
Homogeneity of variance	Assumes equal population variances	Does not require equal variances, although differences in distributional shape or spread affect interpretation
Test statistic	F	H
Reference distribution	F distribution	Chi-squared distribution (approximate)
Planned comparisons	Planned contrasts or t-tests	Wilcoxon rank-sum tests
Post-hoc comparisons	Tukey's HSD	Dunn's test

Summary and Looking Ahead

- The Kruskal–Wallis test is a non-parametric alternative to single-factor between-participants ANOVA
- Rather than comparing group means, it uses ranks to evaluate differences between independent groups
- The H statistic measures differences in the groups' rank distributions and is evaluated using a chi-squared reference distribution
- Planned comparisons and post-hoc tests allow us to investigate specific group differences

In the next lecture, we will consider the non-parametric alternative to single-factor within-participants ANOVA: the Friedman test