TECHNISCHE UNIVERSITÄT
IN DER KULTURHAUPTSTADT EUROPAS
CHEMNITZ

Professur Psychologie digitaler Lernmedien

Institut für Medienforschung Philosophische Fakultät

Einführung in die Statistik

Nonparametrische Verfahren

Gattaca (1997). Columbia TriStar Film.

Prof. Dr. Günter Daniel Rey

Überblick

- Einführung
- Mann-Whitney *U*-Test
- Wilcoxon-Test
- Kruskal-Wallis *H*-Test
- Friedman-Test

Einführung
(z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- Parametrische Verfahren (z. B. *t*-Test und Varianzanalyse) vs. nonparametrische Verfahren
- Vorteile parametrischer Verfahren
  - Einbezug von mehr Informationen bei der Auswertung
  - Höhere Teststärke (in den meisten Fällen)
- Nachteile parametrischer Verfahren
  - Intervallskalenniveau und Normalverteilung (u. a.) als Voraussetzung

Einführung
(z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- Empfehlungen
  - Verfahren primär nach inhaltlichen Gesichtspunkten auswählen
  - Einsatz parametrischer Verfahren als Regelfall (bei intervallskalierten Daten und hinreichend großen Stichproben)
  - Einsatz nonparametrischer Verfahren bei groben Verletzungen der Annahmevoraussetzungen
- Verfahren für Rangdaten vs. Verfahren für nominalskalierte Daten
- Nachfolgend vorgestellte Verfahren
  - Mann-Whitney *U*-Test
  - Wilcoxon-Test
  - Kruskal-Wallis *H*-Test
  - Friedman-Test

Mann-Whitney *U*-Test
(z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- Mann-Whitney *U*-Test (bzw. *U*-Test für unabhängige Stichproben): Nonparametrisches Verfahren zur Auswertung einer Studie mit zwei unabhängigen Gruppen (vgl. *t*-Test)
- Analyse der Rangplätze beim *U*-Test, der speziell für ordinalskalierte Messwerte geeignet ist
- Empfehlung: *U*-Test anstelle des *t*-Tests bei…
  - zweifelhaftem Intervallskalenniveau oder
  - nicht vorliegender Normalverteilung oder
  - verletzter Varianzhomogenität (sowie weiteren Verletzungen der Annahmevoraussetzungen)

Beispiel zur Berechnung eines *U*-Tests

- Beispiel: Fiktive Rohdaten zu einer Studie zum seductive detail Effekt

Mit seductive details

VPN	Transfer	Rang
1	4.0	8
2	4.5	7
3	7.0	4
4	3.0	9
5	2.0	10

Ohne seductive details

VPN	Transfer	Rang
6	7.5	3
7	6.5	5
8	8.5	1
9	5.0	6
10	8.0	2

- Hinweis: SPSS vergibt im Gegensatz hierzu für den niedrigsten Zahlenwert den niedrigsten Rangplatz (d. h. den „ersten Platz“)

Mann-Whitney *U*-Test (z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- Formel zur Überprüfung der Rangplatzunterschiede mittels *U*-Test:

$$U = n_1 \cdot n_2 + \frac{n_1 \cdot (n_1 + 1)}{2} - T_1$$

n_1 = Anzahl an Versuchspersonen in Gruppe 1
n_2 = Anzahl an Versuchspersonen in Gruppe 2
T_1 = Rangsumme für Gruppe 1

- Für das Beispiel gilt:
  - Anzahl an Versuchspersonen in Gruppe 1 und 2: Jeweils 5
  - Rangsumme für Gruppe 1: 8 + 7 + 4 + 9 + 10 = 38
- Berechnung:

$$U = 5 \cdot 5 + \frac{5 \cdot (5+1)}{2} - 38 = 25 + 15 - 38 = 2$$

Mann-Whitney *U*-Test

Welcher *U*-Wert resultiert für den nachfolgenden Datensatz?

☐ A: U = 18 ☐ B: U = 19 ☐ C: U = 38 ☐ D: U = 39

EG

VPN	Rang
1	1
2	8
3	3
4	6
5	12
6	9

KG

VPN	Rang
7	7
8	2
9	4
10	5
11	10
12	11

Mann-Whitney *U*-Test
(z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- Statistischer Kennwert *U*: Summe der Rangplatzüberschreitungen
- Statistischer Kennwert *U*‘: Summe der Rangplatzunterschreitungen

$$\boxed{U' = n_1 \cdot n_2 + \frac{n_2 \cdot (n_2 + 1)}{2} - T_2}$$

n_1 = Anzahl an Personen in Gruppe 1
n_2 = Anzahl an Personen in Gruppe 2
T_2 = Rangsumme für Gruppe 2

- Zusammenhang zwischen *U* und *U*‘:

$\boxed{U = n_1 \cdot n_2 - U'}$ bzw. $\boxed{U' = n_1 \cdot n_2 - U}$

- Berechnung für das Beispiel:

$$U' = 5 \cdot 5 - 2 = 23$$

Mann-Whitney *U*-Test
(z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- Statistische Nullhypothese des *U*-Tests: *U* = *U'*
- Erwarteter *U*-Wert unter der Nullhypothese:

$$\boxed{\mu_U = \frac{n_1 \cdot n_2}{2}}$$

- Berechnung für das Beispiel:

$$\mu_U = \frac{5 \cdot 5}{2} = 12.5$$

- Je stärker sich der empirische *U*-Wert (bzw. *U'*) vom erwarteten *U*-Wert unterscheidet, desto eher ist das Ergebnis signifikant
- Beispiel: Ist der Unterschied zwischen 2 und 12.5 signifikant?

Mann-Whitney *U*-Test

- Zwei Möglichkeiten bei der inferenzstatistischen Entscheidung
  - Nicht signifikant ($U_{emp} > U_{krit}$): H_0 wird vorläufig beibehalten
  - Signifikant ($U_{emp} \leq U_{krit}$): H_0 wird zugunsten der H_1 verworfen
- Wichtig: Je kleiner U_{emp} ist, desto eher ist der Test signifikant!
- Beispiel: Da $U_{emp} = 2 \leq U_{krit} = 4$ (laut Tabelle für $\alpha = .05$ (einseitig)) wird H_0 zugunsten der H_1 verworfen, d. h. das Ergebnis ist signifikant
- Inferenzstatistische Ergebnisdarstellung
  - Signifikantes Ergebnis: Angabe des *U*-Wertes (alternativ des *z*-Wertes) und des *p*-Wertes sowie der Effektstärke (über *z*-Wert berechenbar); Beispiel: $U = 2$, $p < .05$, $r = .69$
  - Nicht signifikantes Ergebnis: Zusätzlich Angabe der Teststärke

Wilcoxon-Test
(z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- Wilcoxon-Test (bzw. *W*-Test): Nonparametrisches Verfahren zur Auswertung einer Studie mit zwei abhängigen Stichproben (vgl. *t*-Test für abhängige Stichproben)
- Beispiel: Studien mit Messwiederholung
- Analyse der Rangplätze (vgl. *U*-Test) in vier Schritten:
  1. Differenzbildung der Messwertpaare
  2. Ermittlung der Betragswerte zu jeder Paardifferenz
  3. Rangreihenbildung zu den Absolutbeträgen der Differenzen
  4. Negatives Vorzeichen für alle absoluten Rangplätze, die zu einer negativen Paardifferenz gehören
- „Gerichtete Ränge“ (engl. signed ranks): Ergebnis der Analyse
- Wichtig: Paardifferenzen mit dem Wert 0 <u>nicht</u> berücksichtigen

Wilcoxon-Test

- Beispiel: Fiktive Rohdaten zu einer Studie (vgl. vorherige Sitzung)

IQ

VPN	Leistung
Sheldon	9.5
Leonard	6.5
Howard	4.5
Rajesh	8.5
Penny	2.0

Basketball

VPN	Leistung
Sheldon	1.5
Leonard	2.0
Howard	2.5
Rajesh	7.0
Penny	6.0

Gerichtete Ränge

VPN	Betrag	Rang
Sheldon	8.0	5
Leonard	4.5	4
Howard	2.0	2
Rajesh	1.5	1
Penny	4.0	-3

- Summe positiver gerichteter Ränge ($R_{positiv}$): 5 + 4 + 2 + 1 = 12
- Summe negativer gerichteter Ränge ($R_{negativ}$): -3

Wilcoxon-Test
(z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- Betragsmäßig kleinerer Wert entspricht dem Testwert *W*:

$$\boxed{W = \left|\min\left(\sum R_{positiv}, \sum R_{negativ}\right)\right|}$$

- Beispiel: $W = 3$
- Kritischer *W*-Wert für $\alpha = .05$ (einseitig) laut Tabelle: $W_{krit} = 0$
- Zwei Möglichkeiten bei der inferenzstatistischen Entscheidung
    - Nicht signifikant ($W_{emp} > W_{krit}$): H_0 wird vorläufig beibehalten
    - Signifikant ($W_{emp} \leq W_{krit}$): H_0 wird zugunsten der H_1 verworfen
- Wichtig: Je kleiner W_{emp} ist, desto eher ist der Test signifikant!
- Beispiel: Da $W_{emp} = 3 > W_{krit} = 0$ wird H_0 vorläufig beibehalten, d. h. das Ergebnis ist nicht signifikant

Wilcoxon-Test

Welcher *W*-Wert resultiert für den nachfolgenden Datensatz?

☐ A: W = 11 ☐ B: W = 10 ☐ C: W = 1 ☐ D: W = 5 ☐ E: W = 6

Messung 1

VPN	Rang
1	1
2	8
3	7
4	6
5	12
6	4

Messung 2

VPN	Rang
1	7
2	3
3	4
4	5
5	10
6	11

Gerichtete Ränge

VPN	Betrag	Rang
1		
2		
3		
4		
5		
6		

Kruskal-Wallis *H*-Test
(z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- Kruskal-Wallis *H*-Test (bzw. Rangvarianzanalyse): Nonparametrisches Verfahren zur Auswertung einer Studie mit mehr als zwei unabhängigen Gruppen (vgl. Varianzanalyse)
- Analyse der Rangplätze: Analog zum *U*-Test über die Rangsummen T_i
- Formel zur Berechnung des Testwertes *H*:

$$H = \left[\frac{12}{N \cdot (N+1)} \right] \cdot \left[\sum_{i=1}^{p} \frac{T_i^2}{n_i} \right] - 3 \cdot (N+1)$$

N = Stichprobenumfang
T_i = Rangsumme für Gruppe i
n_i = Anzahl an Personen in Gruppe i

Kruskal-Wallis *H*-Test
(z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- **Vergleich:** Bei hinreichend großen Stichproben kann der *H*-Wert mit einem kritischen Chi-Quadrat-Wert (χ^2) mit df = $p - 1$ verglichen werden
- **Wenn $H \geq \chi^2_{krit}$,** dann ist das Ergebnis signifikant
- **Post-hoc-Analysen** u. a. mit *U*-Tests und Bonferroni(-Holm)-Korrektur

Kruskal-Wallis *H*-Test

Welcher *H*-Wert resultiert für den nachfolgenden Datensatz?

☐ A: H = 12.0 ☐ B: H = 12.5 ☐ C: H = 13.0

Gruppe 1

VPN	Transfer
1	1.0
2	1.5
3	2.0
4	2.5
5	3.0

Gruppe 2

VPN	Transfer
6	3.5
7	4.0
8	4.5
9	5.0
10	5.5

Gruppe 3

VPN	Transfer
11	6.0
12	6.5
13	7.0
14	7.5
15	8.0

Friedman-Test
(z. B. Bühner & Ziegler, 2009)

- Friedman-Test (bzw. Rangvarianzanalyse): Nonparametrisches Verfahren zur Auswertung einer Studie mit mehr als zwei abhängigen Gruppen (z. B. mehr als zwei Messzeitpunkte)
- Analyse der Rangplätze: Bildung der Rangplätze zunächst innerhalb des jeweiligen Falls, anschließende Summierung der Ränge pro Messung/Treatment (R_i)
- Formel zur Berechnung des Testwertes F_r (oder T)

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

N = Stichprobenumfang
k = Anzahl der Messungen/Treatments
R_i = Rangsumme der Messung i

Friedman-Test
(z. B. Bühner & Ziegler, 2009)

- Vergleich: Bei hinreichend großen Stichproben ist die Teststatistik F_r annähernd χ^2-verteilt mit $df = k - 1$
- Wenn $F_r \geq \chi^2_{krit}$, dann ist das Ergebnis signifikant
- Post-hoc-Analysen u. a. mit *Wilcoxon*-Tests und Bonferroni (-Holm)-Korrektur

Friedman-Test

Beispielrechnung

VPN	MZP 1 Transfer	MZP2 Transfer	MZP3 Transfer	MZP 1 Rang	MZP 2 Rang	MZP 3 Rang
1	1.0	3.5	6.0	1	2	3
2	1.5	4.0	6.5	1	2	3
3	2.0	4.5	7.0	1	2	3
4	2.5	5.0	7.5	1	2	3
5	3.0	5.5	8.0	1	2	3

N = 5

k = 3

$R_1 = 5$ $R_1 = 10$ $R_1 = 15$

$$F_r = \left[\frac{12}{Nk(N+1)}\sum_{i=1}^{k} R_i^2\right] - 3N(k+1) = \left[\frac{12}{5 \cdot 3(5+1)} \cdot (5^2 + 10^2 + 15^2)\right] - 3 \cdot 5(3+1) = \mathbf{10}$$

$df = k - 1 = 3 - 1 = 2$ → kritischer $\chi^2_{krit} \approx 5.991$ → $\mathbf{10 > 5.991}$

Nullhypothese (zentrale Tendenzen der Stichproben sind gleich) wird abgelehnt

Beispiele für nonparametrische Verfahren in Fachzeitschriften

individual characteristics affected the dropout. Consequently, we took a closer look at potential variables that might have affected a dropout. For this, we conducted Mann-Whitney-U-Tests between participants who withdrew from the study in the first learning phase and the other participants. We neither found significant differences for personal interest in educational psychology ($U = 1537.000$, $Z = -0.002$, $p = .998$), nor academic self-concept ($U = 1711.000$, $Z = 0.912$, $p = .362$), nor self-rated prior knowledge ($U = 1772.000$, $Z = 1.210$, $p = .226$). During

Quelle: Krebs, Braschoß und Eitel (2024)

Table 1 presents descriptive statistics across groups for the key variables measured in the experiment. First, we tested whether the groups differed in their prior knowledge test score. Performance on the prior knowledge test was generally low (1.55, $SD = 1.46$) and highly skewed, so we used the nonparametric Kruskal-Wallis rank test to compare groups on prior knowledge. Results indicated the groups did not significantly differ in prior knowledge, $\chi^2(3) = 1.60$, $p = 0.660$.

Quelle: Zhang und Fiorella (2024)

We first examine SPE across treatments. Averages can be inferred from the left panel of Fig. 4. The upper panel of Fig. 5 depicts the distribution of SPE for both treatments. In the VIRT treatment, supervisors rated worker performance significantly higher (1.24, $SD = 1.60$) than in the REAL treatment (0.53, $SD = 1.39$, Cohen's $d = 0.47$; $p < 0.01$, Wilcoxon rank sum test, two-sided). We detected a moderate effect size for the main effect. When we observe the distributions of SPE across treatments, we find that supervisors' performance evaluations were significantly differently distributed across treatments ($p = 0.016$, two-sample Kolmogorov–Smirnov test for equality of distribution

Quelle: Reiter, Mohnen und Walkowitz (2024)

Zusammenfassung

- Nonparametrische Verfahren: Einsatz bei groben Verletzungen der Annahmevoraussetzungen parametrischer Verfahren
- Mann-Whitney *U*-Test für Studien mit zwei unabhängigen Gruppen
- Wilcoxon-Test für Studien mit zwei abhängigen Stichproben
- Kruskal-Wallis *H*-Test für Studien mit mehr als zwei unabhängigen Gruppen
- Friedman-Test für Studien mit mehr als zwei abhängigen Stichproben

Prüfungsliteratur

- Rasch, B., Friese, M., Hofmann, W., & Naumann, E. (2021). *Quantitative Methoden 2: Einführung in die Statistik für Psychologie, Sozial- & Erziehungswissenschaften* (5. Aufl.). Heidelberg: Springer.
    - Verfahren für Rangdaten (S. 105–122)

Weiterführende Literatur I

- Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Berlin: Springer.
  - Nicht-parametrische Tests (S. 129–136)
- Bühner, M., & Ziegler, M. (2019). *Statistik für Psychologen und Sozialwissenschaftler*. Pearson.
  - Friedman-Test (S. 470–482)
- Eid, M., Gollwitzer, M., & Schmitt, M. (2017). *Statistik und Forschungsmethoden* (5. Aufl.). Weinheim: Beltz.
  - Vergleich zweier Stichprobenmediane (Wilcoxon-Rangsummen-Test bzw. *U*-Test) (S. 343–349)
  - Test auf Gruppenunterschiede für Rangdaten (Kruskal-Wallis-Test) (S. 454–456)

Weiterführende Literatur II

- Leonhart, R. (2022). *Lehrbuch Statistik. Einstieg und Vertiefung* (5. Auflage). Bern: Huber.
  - Nicht-parametrische Testverfahren (S. 229–256)
- Sedlmeier, P., & Renkewitz, F. (2018). *Forschungsmethoden und Statistik: Ein Lehrbuch für Psychologen und Sozialwissenschaftler* (3. Aufl.). München: Pearson.
  - Verfahren zur Analyse ordinalskalierter Daten (S. 571–586)