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
- Eindimensionaler Chi-Quadrat-Test
- Zweidimensionaler Chi-Quadrat-Test
- Vierfelder Chi-Quadrat-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
- 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

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'*:

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

- 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

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

- Vergleich: Bei hinreichend großen Stichproben kann dieser *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

Eindimensionaler Chi-Quadrat-Test
(z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- χ^2-Verfahren (Chi-Quadrat-Verfahren): Zurr Analyse nominalskalierter (kategorialer bzw. diskreter) Daten, d. h. der Analyse von Häufigkeiten
- Eindimensionaler χ^2-Test: Klassifikation von Versuchspersonen anhand eines Merkmals mit zwei oder mehr Stufen
- Beispiel: Überprüfung, ob signifikant mehr Frauen als Männer an einem Lernexperiment teilgenommen haben
- Vergleich zwischen beobachteten und erwarteten Häufigkeiten
- Erwartete Häufigkeiten entsprechen der Nullhypothese
- Gleichverteilungsannahme: Eine (von vielen möglichen) Nullhypothesen, bei der die erwarteten Häufigkeiten für alle Zellen gleich sind

Eindimensionaler Chi-Quadrat-Test (z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- **Beispiel:** Geschlechtsverteilung aus dem Artikel von Wang und Yeh (2013) zum Einfluss des Sexappeals pädagogischer Agenten: 121 Männer und 163 Frauen (N = 284; 57% ♀)
- **Berechnung des χ^2-Wertes:**

$$\boxed{\chi^2 = \sum_{i=1}^{k} \frac{(\mathrm{f}_{\mathrm{bi}} - \mathrm{f}_{ei})^2}{\mathrm{f}_{ei}}}$$

k	= Anzahl der Kategorien des Merkmals
f_{bi}	= Beobachtete Häufigkeit in Kategorie i
f_{ei}	= Erwartete Häufigkeit in Kategorie i
N	= Stichprobenumfang

- **Unter der Gleichverteilungsannahme gilt:**

$$\boxed{\mathrm{f}_{e1} = \mathrm{f}_{e1} = \ldots = \mathrm{f}_{ek} = \frac{N}{k}}$$

- **Beispiel:** f_{ei} = 284 : 2 = 142

$$\chi^2 = \frac{(121-142)^2}{142} + \frac{(163-142)^2}{142} = \frac{441}{142} + \frac{441}{142} \approx 6.21$$

Eindimensionaler Chi-Quadrat-Test
(z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- Bestimmung der Freiheitsgrade: df = k – 1
- Beispiel: df = 2 – 1 = 1
- Kritischer χ^2-Wert für df = 1 und α = .05 (zweiseitig): $\chi^2_{krit} \approx 3.84$
- Inferenzstatistische Entscheidung und Ergebnisdarstellung
  - Da $\chi^2_{emp} = 6.21 \geq \chi^2_{krit} = 3.84$ wird H_0 wird zugunsten der H_1 verworfen, d. h. das Ergebnis ist signifikant
  - $\chi^2(1) = 6.21$, $p = .01$, $w^2 = 0.02$
- Berechnung der Effektgröße w^2 über $w^2 = \chi^2 : N$ mit den Konventionen:

Effektgröße	Kleiner Effekt	Mittlerer Effekt	Großer Effekt
w^2	0.01	0.09	0.25

- Weitere Effektgrößen für χ^2-Verfahren: Phi φ, Cramer’s *V*, Odds-Ratio

Zweidimensionaler Chi-Quadrat-Test (z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- Zweidimensionaler χ^2-Test: Klassifikation von Versuchspersonen anhand zweier Merkmale mit jeweils zwei oder mehr Stufen
- Kontingenzanalyse: Spezieller zweidimensionaler χ^2-Test, bei dem die Unabhängigkeit der untersuchten Merkmale überprüft wird
- Beispiel: Überprüfung, ob signifikant mehr Frauen als Männer eine Lernaufgabe richtig gelöst haben
- Fiktive Rohdaten und Randhäufigkeiten zu einer Studie:

<table>
<tr><th colspan="2" rowspan="2">k x l Kreuztabelle</th><th colspan="2">Geschlecht</th><th rowspan="2">Zeilen-summe</th></tr>
<tr><th>Männer</th><th>Frauen</th></tr>
<tr><td rowspan="2">Lernaufgabe</td><td>Richtig</td><td>5</td><td>55</td><td>60</td></tr>
<tr><td>Falsch</td><td>10</td><td>30</td><td>40</td></tr>
<tr><td colspan="2">Spaltensumme</td><td>15</td><td>85</td><td>100</td></tr>
</table>

Zweidimensionaler Chi-Quadrat-Test (z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- Berechnung des χ^2-Wertes:

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(f_{bij} - f_{eij})^2}{f_{eij}}$$

k	= Anzahl der Kategorien des Merkmals A
l	= Anzahl der Kategorien des Merkmals B
f_{bij}	= Beobachtete Häufigkeit der Kombination i,j
f_{eij}	= Erwartete Häufigkeit der Kombination i,j
N	= Stichprobenumfang

- Bei der Kontingenzanalyse gilt:

$$f_{eij} = \frac{n_{i\bullet} \cdot n_{\bullet j}}{N} = \frac{\text{Zeilensumme} \cdot \text{Spaltensumme}}{N}$$

- Beispiel: $f_{e11} = (15 \cdot 60) : 100 = 9$; $f_{e12} = 51$; $f_{e21} = 6$; $f_{e22} = 34$

$$\chi^2 = \frac{(5-9)^2}{9} + \frac{(55-51)^2}{51} + \frac{(10-6)^2}{6} + \frac{(30-34)^2}{34} \approx 5.23$$

Zweidimensionaler Chi-Quadrat-Test
(z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- Bestimmung der Freiheitsgrade: $df = (k - 1) \cdot (l - 1)$
- Beispiel: $df = (2 - 1) \cdot (2 - 1) = 1$
- Kritischer χ^2-Wert für df = 1 und $\alpha = .05$ (zweiseitig): $\chi^2_{krit} \approx 3.84$
- Inferenzstatistische Entscheidung
  - Da $\chi^2_{emp} = 5.23 \geq \chi^2_{krit} = 3.84$ wird H_0 zugunsten der H_1 verworfen, d. h. das Ergebnis ist signifikant
- Ergebnisdarstellung:
  - $\chi^2(1) = 5.23$, $p = .02$, $w^2 = 0.05$

Vierfelder Chi-Quadrat-Test (z. B. Rasch, Friese, Hofmann & Naumann, 2021)

- Vierfelder χ^2-Test: Spezialfall des zweidimensionalen χ^2-Tests auf Unabhängigkeit mit zwei dichotomen Merkmalen, d. h. zwei Merkmalen mit jeweils zwei Stufen
- 2 x 2 Kontingenztabelle als Versuchsplan:

2 x 2 Kreuztabelle		Merkmal B	
		B_1	B_2
Merkmal A	A_1	a	b
	A_2	c	d

- Berechnung des χ^2-Wertes:

$$\chi^2 = \frac{N \cdot (a \cdot d - b \cdot c)^2}{(a+b) \cdot (c+d) \cdot (a+c) \cdot (b+d)}$$

Vierfelder Chi-Quadrat-Test

Berechnen Sie den χ^2-Wert für den nachfolgenden Datensatz mit der Formel zum Vierfelder χ^2-Test!

- ☐ A: $\chi^2 = 5.23$
- ☐ B: $\chi^2 = 3.84$
- ☐ C: $\chi^2 = 0.64$

2 x 2 Kreuztabelle		Geschlecht: Männer	Geschlecht: Frauen
Lernaufgabe	Richtig	5	55
Lernaufgabe	Falsch	10	30

Beispiele für nonparametrische Verfahren in Fachzeitschriften

which showed that homoscedasticity was met. For these variables, we therefore conducted an analysis of variance (ANOVA), with cueing condition (Cueing vs. No-cueing) as the between-participants factor. As a Levene's test on the transfer and nonsignalled information retention scores revealed that homoscedasticity was not met, $F(1,30) = 5.71$, $p < .05$, and $F(1,30) = 12.42$, $p < .05$, these data were analysed using the nonparametric Mann–Whitney *U* test.

For the nonsignalled information retention task, the Mann–Whitney test indicated that there was no significant difference between the cueing group ($M = 3.68$, $SD = 2.1$) and the no-cueing group ($M = 3.41$, $SD = .66$), Mann–Whitney $U = 94$, $p = .21$.

Quelle: Jamet (2014)

tion effect, $F(1,32) = 3.06$, $MSE = 497.06$, $p = .09$. Because performance data were not normally distributed in all conditions, we also conducted non-parametric Mann–Whitney *U* tests, which showed the same results: a significant effect of 'Worked Examples' $Z = 4.13$, $p < .001$, $r = .688$, but no significant effect of 'Process Steps' $Z = .127$, $p = .899$.

Quelle: Nievelstein, van Gog, van Dijck und Boshuizen (2013)

The announcement about participating in this experiment was published at two German universities (TU Dortmund and University of Regensburg), therefore most students were studing at these universities: Sixty of them studied at TU Dortmund, 35 at University of Regensburg, and 15 at other universities. Most students studied psychology or pedagogical subjects. Experimental groups did not differ in aspects of university membership either (Chi-square test with 4 groups x 3 universities; frequencies of Dortmund/ Regensburg/ other = 18/7/4 and 14/12/2 for the system paced visual and auditory group, respectively; frequencies of Dortmund/ Regensburg/ other = 12/9/4 and 15/8/5 for the learner paced visual and auditory group, respectively; Chi-square (6) = 3.70, ns).

Quelle: Stiller, Freitag, Zinnbauer und Freitag (2009)

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
- Chi-Quadrat-Verfahren zur Analyse nominalskalierter Daten, sprich der Analyse von Häufigkeiten

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)
  - Verfahren für Nominaldaten (S. 125–145)

Weiterführende Literatur I

- Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler* (7. Aufl.). Berlin: Springer.
  - Nicht-parametrische Tests (S. 129–136)
  - Analyse von Häufigkeiten (S. 137–152)
- 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)
  - Vergleich von Häufigkeitsverteilungen zwischen zwei unabhängigen Stichproben (S. 354–360)
  - Der Zweistichproben-χ^2-Test (S. 360–365)
  - 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 nominalskalierter Daten: Chi-Quadrat (χ^2-)Tests (S. 543–570)
  - Verfahren zur Analyse ordinalskalierter Daten (S. 571–586)