

UNIVERSITÄT LEIPZIG

Medizinische Fakultät

V 10 Statistisches Testen II

Dr. rer. med. habil. Marita Ziepert /
Dr. rer. nat. Meinhard Mende,
IMISE, Universität Leipzig

imise.

Lernziele

Am Ende der Vorlesung

- kennen Sie den Unterschied zwischen parametrischen und nichtparametrischen Tests,
- können Sie den U-Test für zwei unabhängige Stichproben sowie den Wilcoxon-Test für gepaarte Stichproben rechnen und das Testergebnis interpretieren
- und wissen Sie, welcher Test in welcher Situation adäquat ist.

V10

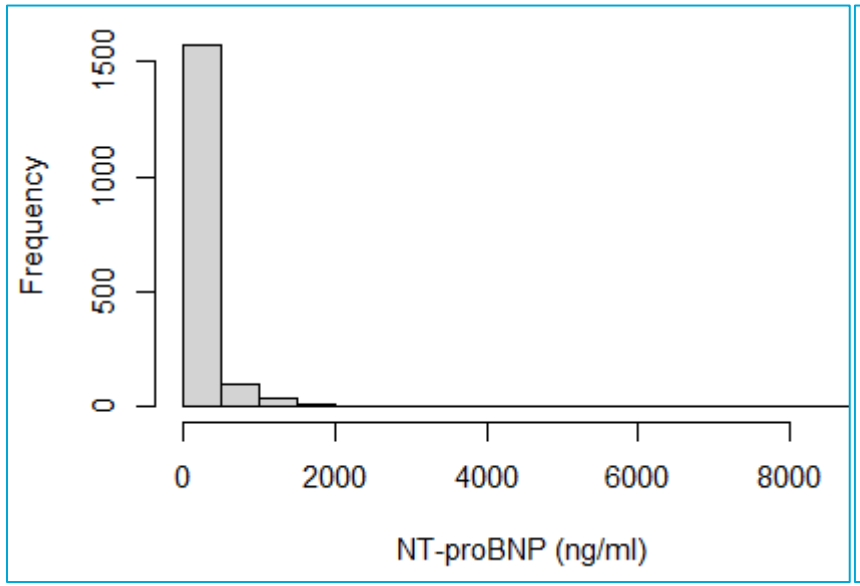
10.1 Nichtparametrische Tests zum Vergleich der Lage von Verteilungen

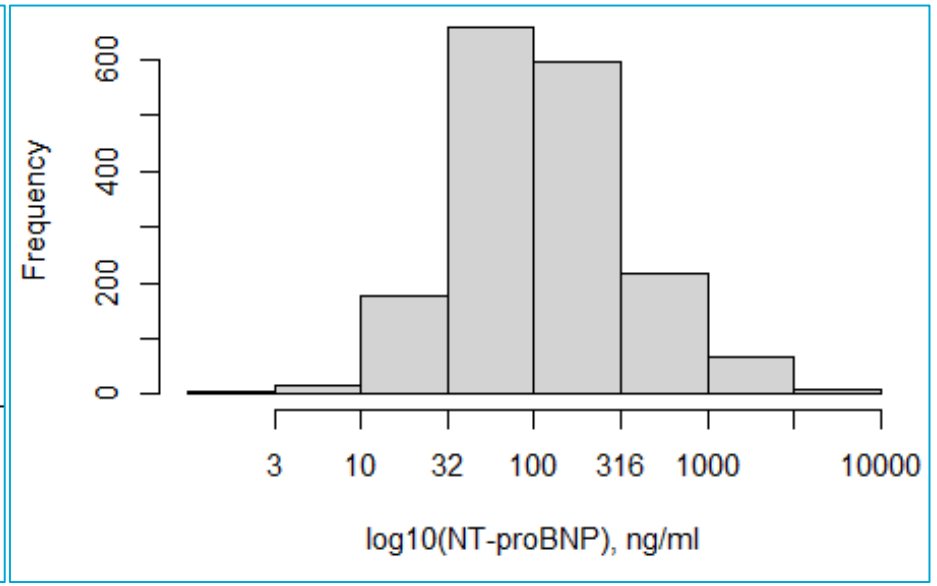
Nichtparametrische Tests - Einführung I

Wie analysiere ich Daten, die offensichtlich nicht normalverteilt (also z. B. schief verteilt / ordinal) sind oder Ausreißer haben?

1. t-Test (Welch) gilt als sehr robust (Sawilowsky & Blair 1982, Rasch et al. 2011), d.h. er ist anwendbar, wenn
   - die Verteilung nahezu symmetrisch ist oder
   - Stichprobenumfang nicht zu klein ist (n ≥ 30)
2. Transformation der Daten: z. B. Logarithmieren
   - sinnvoll z. B. bei Laborwerten
3. anderen Test anwenden, z. B. nichtparametrischen (syn.: verteilungsfreien) Test, auch Rangtest genannt

Beispiel 10. 1 Logarithmische Transformation





aus Daten der Diast-CHF-Studie

- NT-proBNP, Marker für Herzinsuffizienz bzw. Vorhofflimmern
  - mit einer (extrem) schiefen Verteilung und
  - einem Wertebereich über mehrere Zehnerpotenzen.
- Durch logarithmische Transformation wird die Verteilung ziemlich normalisiert.

Nichtparametrische Tests - Einführung II

Parametrische Tests versus **Nichtparametrische Tests**

Voraussetzungen

- Unabhängigkeit der Beobachtungseinheiten
- Gültigkeit der Verteilungsannahme für jede Stichprobe

Effektgröße:

- gut quantifizierbar

Voraussetzungen

- Unabhängigkeit der Beobachtungseinheiten
- keine Annahme einer speziellen Verteilung

Effektgröße:

- evtl. schwierig quantifizierbar

Ein visueller Check der Verteilung mit einem Histogramm ist sinnvoll.

Nichtparametrische Tests - Einführung III

Parametrische Tests versus Nichtparametrische Tests

- Power ist bei passendem Modell höher.
- Power ist ggf. niedriger.

Bei stetigen Merkmalen liefern mit wachsender Fallzahl parametrische und nichtparametrische Verfahren immer ähnlichere Ergebnisse

- Konsequenz des zentralen Grenzwertsatzes
- Konvergenz gegen Normalverteilung

Folgerung: Für große Fallzahl parametrische Tests bevorzugen.

Entscheidungshilfe

- Checken Sie die Verteilung(en) visuell (Histogramm)
  - Halbwegs symmetrisch oder (sehr) schief?
  - Extremwerte?
- Auf welcher Skala wird gemessen
  - metrisch
  - ordinal
- Wie ist die Fallzahl?

Rangtests zum Lagevergleich von Verteilungen

Anzahl der Stichproben	Art der Stichproben	parametrischer Test	Statistisches Verfahren
1	----	ja	t-Test (Normwertvergleich)
		nein	Wilcoxon-Test (Normwertvergleich)
2	unabhängig	ja	t-Test für unabhängige Stichproben
		nein	**U-Test nach Mann-Whitney**
	abhängig	ja	t-Test für Paardifferenzen
		nein	**Wilcoxon-Test**
> 2	unabhängig	ja	univariate Varianzanalyse
		nein	H-Test nach Kruskal-Wallis
	abhängig	ja	multivariate Varianzanalyse
		nein	Friedman-Test

10.1.1 Mann-Whitney-U-Test I

Szenario:

- Lagevergleich zweier unabhängiger Gruppen
- Beobachtungswerte ordinal skaliert bzw. schief verteilt / kleine Fallzahl

- Prüfgröße berechnet sich aus den Rangsummen und der Fallzahl der beiden Gruppen.
- Wilcoxon 1945 bzw. Mann und Whitney 1947: Unabhängig voneinander äquivalente Tests.
- Wir nutzen die Bezeichnung Mann-Whitney-U-Test (in Abgrenzung vom Wilcoxon-Test für gepaarte Werte).

Mann-Whitney U-Test II

vergleicht 2 unabhängige Gruppen bzgl. ihrer Lage: $X \sim F(x), Y \sim G(x)$.

H_0: $F(x) \equiv G(x)$ bzw.

die Wahrscheinlichkeit, dass eine beliebige Beobachtung der 1. Grundgesamtheit größer ist als eine beliebig gezogene Beobachtung der 2. Grundgesamtheit ist genau ½.

- Voraussetzung: $m, n \geq 4$ $(m, n; Fallzahlen\ der\ 2\ Gruppen)$
- Prüfgröße: $U = min\ (U_1, U_2)$
- $U_1 = m \cdot n + \frac{m(m+1)}{2} - R_1$
- $U_2 = m \cdot n + \frac{n(n+1)}{2} - R_2$ (R_1, R_2…Rangsummen)

Mann-Whitney U-Test III

Algorithmus:

- Ordne die $m + n$ Beobachtungseinheiten in aufsteigender Reihenfolge,
- vergebe Ränge für alle Beobachtungeinheiten $m + n$
  - Gleichen sich mehrere Beobachtungseinheiten, ersetze Ränge durch deren arithmetisches Mittel (Bindungen, ties)
- Addiere die Ränge getrennt für beide Gruppen und berechne U_1 und U_2 (Kontrolle: $U_1 + U_2 = m \cdot n$)
- Prüfgröße: $U = min\,(U_1, U_2)$
- Testentscheidung: $U \leq U_{m;\,n;\,1-\alpha/2}$ → H_0 ablehnen

Beispiel 10.2 Behandlung von Candidiasis

- In einer randomisierten Studie wurde bei jeweils 10 Patienten mit chronischer oraler Candidiasis nach der Behandlung mit Placebo bzw. Verum (Clotrimazol) der Erfolgsscore (1 – 4) am Tag 7 nach Behandlung bestimmt (Kirkpatrick and Alling, NEJM 1978).
- Unterscheidet sich der Erfolgsscore im Mittel zwischen den mit Placebo und Verum behandelten Patienten?

Einordnung:

- ordinale Skala
- (sehr) kleine Fallzahl: n = 10

⇨ nichtparametrischer Test

Beispiel 10.2 Daten

Gruppe 1 (Verum)		Gruppe 2 (Placebo)	
Nr.	Score	Nr.	Score
1	1	1	1
2	1	2	4
3	1	3	4
4	1	4	4
5	1	5	4
6	1	6	4
7	2	7	4
8	2	8	4
9	2	9	4
10	3	10	4

Beispiel 10.2 II

1. Zielgröße: Erfolgsscore (ordinal)
2. Stichproben: 2 unabhängige
3. Hypothesen: für beliebige Paare X_1, X_2:

$$H_0: p(X_1 > X_2) = \frac{1}{2} \qquad \text{vs.} \qquad H_A: p(X_1 > X_2) \neq \frac{1}{2}$$

oder

- H_0: Die Verteilungen der Erfolgsscores unterscheiden sich zwischen der mit Verum und Placebo behandelten Gruppe nicht.
- H_A: Die Verteilungen der Erfolgsscores unterscheiden sich zwischen der mit Verum und Placebo behandelten Gruppe.

Beispiel 10.2 III

4. Signifikanzniveau: $\alpha = 0{,}05; zweiseitig$
5. Prüfgröße: Stichprobe,

   daraus Rangsummen R_1, R_2

$$U_1 = m \cdot n + \frac{m(m+1)}{2} - R_1$$

$$U_2 = m \cdot n + \frac{n(n+1)}{2} - R_2$$

$$U = \min(U_1, U_2)$$

6. Testvoraussetzungen: $m, n > 4$

Beispiel 10.2 IV

7. Ablehnbereich: $m = 10, n = 10$

   Kritischer Wert für U-Test: $U_{n\,=\,10;\,m\,=\,10;\,1\,-\,0{,}05/2} = 23$

$$zweiseitig \quad \alpha = 0{,}05; \quad einseitig \quad \alpha/2 = 0{,}025$$

m \ n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	-																			
2	-	-																		
3	-	-	-																	
4	-	-	-	0																
5	-	-	0	1	2															
6	-	-	1	2	3	5														
7	-	-	1	3	5	6	8													
8	-	0	2	4	6	8	10	13												
9	-	0	2	4	7	10	12	15	17											
→ 10	-	0	3	5	8	11	14	17	20	23										
11	-	0	3	6	9	13	16	19	23	26	30									
12	-	1	4	7	11	14	18	22	26	29	33	37								
13	-	1	4	8	12	16	20	24	28	33	37	41	45							
14	-	1	5	9	13	17	22	26	31	36	40	45	50	55						
15	-	1	5	10	14	19	24	29	34	39	44	49	54	59	64					
16	-	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75				
17	-	2	6	11	17	22	28	34	39	45	51	57	63	69	75	81	87			

Beispiel 10.2 V

8. Berechnung der Prüfgröße:

$$\frac{12+13+14+15+16+17+18+19+20}{9} = 16$$

$$\frac{1+2+3+4+5+6+7}{7} = 4$$

$$(8+9+10) : 3 = 9$$

Gruppe 1 (Verum)				Gruppe 2 (Placebo)			
Nr.	Score	Prov. Rang	Rang R_1	Nr.	Score	Prov. Rang	Rang R_2
1	1	1	4	1	1	7	4
2	1	2	4	2	4	12	16
3	1	3	4	3	4	13	16
4	1	4	4	4	4	14	16
5	1	5	4	5	4	15	16
6	1	6	4	6	4	16	16
7	2	8	9	7	4	17	16
8	2	9	9	8	4	18	16
9	2	10	9	9	4	19	16
10	3	11	11	10	4	20	16
Summe			$R_1 = 62$				$R_2 = 148$

Beispiel 10.2 VI

- $U_1 = m \cdot n + \frac{m(m+1)}{2} - R_1 = 10 \cdot 10 + \frac{10 \cdot 11}{2} - 62 = \mathbf{93} = U_V$
- $U_2 = m \cdot n + \frac{n(n+1)}{2} - R_2 = 10 \cdot 10 + \frac{10 \cdot 11}{2} - 148 = \mathbf{7} = U_P$
- $U = \min(U_1; U_2) = \min(93; 7) = \mathbf{7}$

- Rechenkontrolle: $U_1 + U_2 = 93 + 7 = m \cdot n = 10 \cdot 10 = 100$

Beispiel 10.2 VII

9. Testentscheidung: $U = 7 \leq U_{10;\,10;\,1-0{,}05/2} = 23$

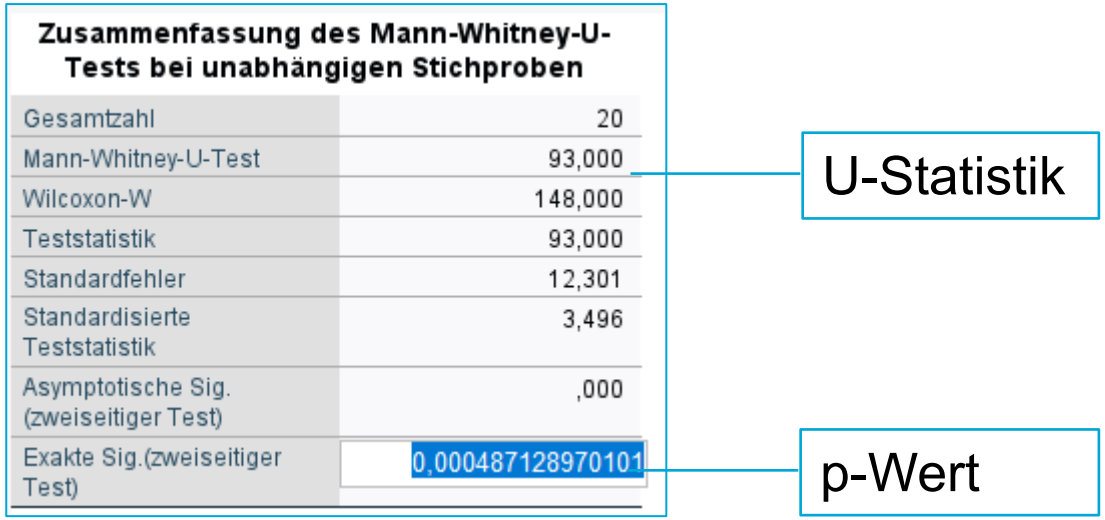
⇨ H_0 wird abgelehnt

10. Interpretation:

Die Verteilungen der Erfolgsscores unterscheiden sich bei Patienten in der Verum- und Placebogruppe signifikant zum Niveau 5%.
Der mediane Score ist in der Verumgruppe 1, in der Placebogruppe dagegen 4.

B 10.2 Mit Software SPSS

- SPSS-Ausgabe:



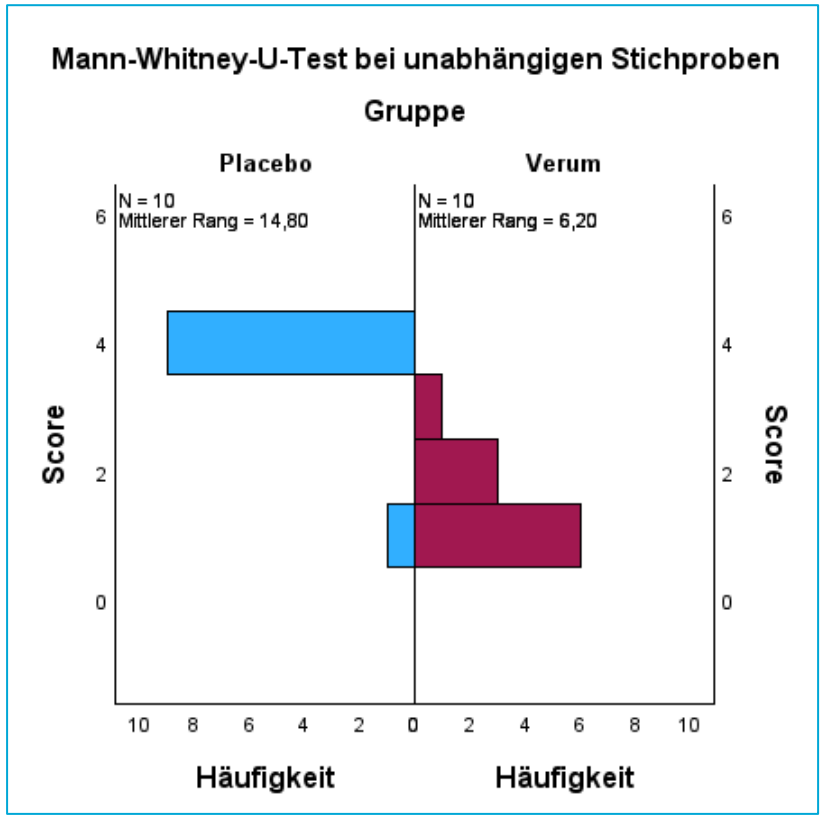
- Wegen $n \leq 30$ den exakten Test auswerten:
- $p = 0{,}00049 \leq \alpha = 0{,}05$

⇨ H_0 wird abgelehnt

B 10.2 Mit Software

Diagramm:

- instruktives Doppel-Histogramm
- mittlere Ränge zeigen Richtung des Effekts
- Konfidenzintervall nach Hodges & Lehmann
  (Ann Math Stat 1963;34(2):598-611)



Konfidenzintervallübersicht

Konfidenzintervalltyp	Parameter	Schätzer	95,0%-Konfidenzintervall Unterer	Oberer
Hodges-Lehman-Mediandifferenz bei unabhängigen Stichproben	Differenz zwischen Medianen von Score in Kategorien von Gruppe.	-3,000	-3,000	-2,000

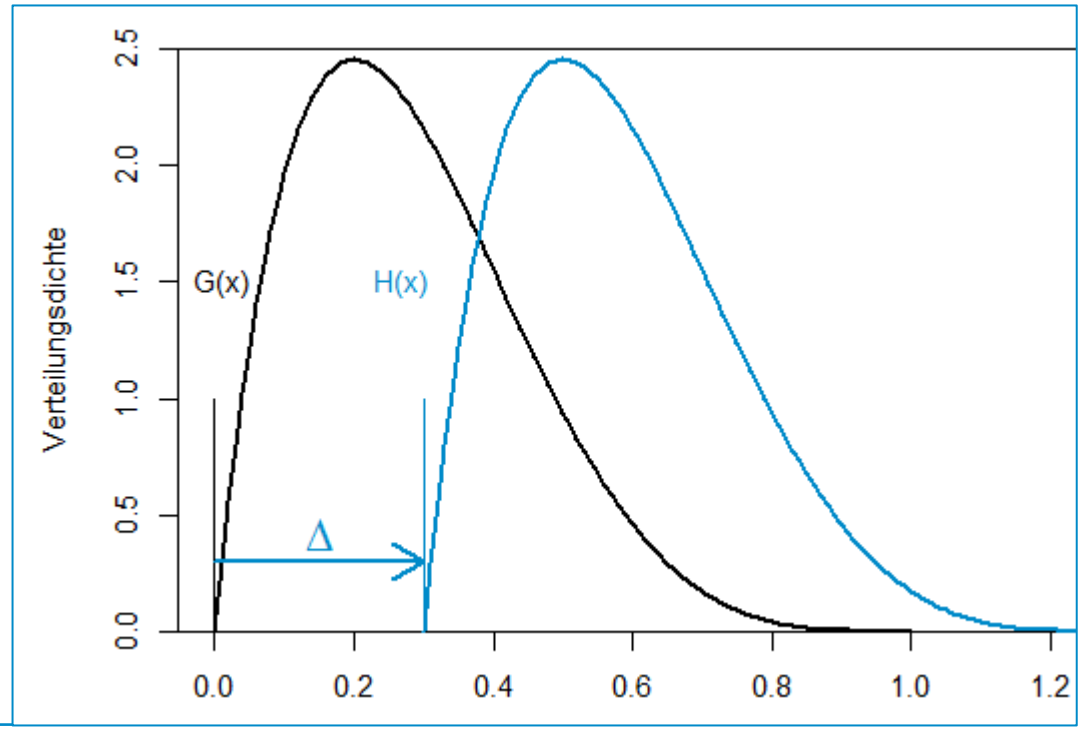
Anmerkungen



1. Entgegen anders lautenden Gerüchten prüft der Mann-Whitney U-Test nicht die Gleichheit der Mediane; das erledigt z. B. der Mediantest.
2. Der M-W U-Test ist optimal für die Shift-Hypothese als Alternativhypothese.
   Es gibt andere Alternativhypothesen.
3. Wegen der Robustheit des Welch-Tests bleiben für den M-W U-Test zwei Einsatzgebiete bei metrischen Größen:

– ordinale Zielvariablen

– kleine Fallzahlen mit Ausreißern



Rangtests zum Lagevergleich von Verteilungen

Anzahl der Stichproben	Art der Stichproben	parametrischer Test	Statistisches Verfahren
1	----	ja	t-Test (Normwertvergleich)
		nein	Wilcoxon-Test (Normwertvergleich)
2	unabhängig	ja	t-Test für unabhängige Stichproben
		nein	**U-Test nach Mann-Whitney** ✓
	abhängig	ja	t-Test für Paardifferenzen
		nein	**Wilcoxon-Test**
> 2	unabhängig	ja	univariate Varianzanalyse
		nein	H-Test nach Kruskal-Wallis
	abhängig	ja	multivariate Varianzanalyse
		nein	Friedman-Test

10.1.2 Wilcoxon-Test

Beispiel 10.3

In einer Stichprobe von 8 Sportlern einer Handballmannschaft wurde die Abfluggeschwindigkeit des Balles vor und nach dem Training bestimmt.

- Unterscheidet sich die *mittlere* Abfluggeschwindigkeit vor und nach dem Training?

Szenario:

- metrische Zielvariable
- zwei Messreihen von denselben Probanden
- (sehr) kleine Fallzahl, Verteilung unbekannt

Wilcoxon-Test I

Idee:

- Prüfgröße berechnet sich aus den Rangsummen der Differenzen der Beobachtungswerte
- H_0: Beide Stichproben stammen aus derselben Grundgesamtheit, haben also dieselbe Verteilung.
- H_A: Eine Stichprobe hat in der Tendenz größere Werte (stochastic superiority)

s. Fay & Proschan. Stat Surv 2010;4:1-39

Wilcoxon-Test II

Algorithmus:

- Sortiere die Absolutbeträge der Differenzen |d $_i$| aufsteigend
- Vergebe Ränge für alle Beträge $|d_i| > 0,\ i = 1, \ldots n^*$
  ($n^* \leq n$ ist die Anzahl von Differenzen $\neq 0$)
  - betragsmäßig gleiche Differenzen: Ränge durch deren arithmetisches Mittel ersetzen
- Addiere die Ränge getrennt für positive (R_+) und negative (R_-) Differenzen

  (Kontrolle: $R_+ + R_- = \frac{n^*(n^*+1)}{2}$)
- Prüfgröße: $R = min(R_+, R_-)$
- Testentscheidung: Ist $R \leq R_{n*,1-\alpha/2}$ ⇨ H_0 ablehnen

Beispiel 10.3 II

Spieler	Abfluggeschwindigkeit des Balles [km/h] vor dem Training	Abfluggeschwindigkeit des Balles [km/h] nach dem Training	Differenz
1	41	49	-8
2	44	41	3
3	48	56	-8
4	55	51	4
5	57	72	-15
6	58	52	6
7	60	66	-6
8	68	80	-12

– Was vermuten Sie nach dem ersten Blick?

Beispiel 10.3 III

1. Zielgröße: Abfluggeschwindigkeit (metrisch)
2. Stichproben: 2 abhängige
3. Hypothesen: H_0: $\tilde{\mu}_d = 0$ vs. H_A: $\tilde{\mu}_d \neq 0$

bzw.

- H_0: Die mediane Abfluggeschwindigkeit vor und nach dem Training unterscheidet sich nicht.
- H_A: Die mediane Abfluggeschwindigkeit vor und nach dem Training unterscheidet sich.

Beispiel 10.3 IV

4. Signifikanzniveau: $\alpha = 0{,}05; \text{zweiseitig}$

– Stichprobe: $\mathrm{R}_{+}, \mathrm{R}_{-}$

5. Prüfgröße:

$$R = \min(R+, R-)$$

6. Testvoraussetzungen: $n \geq 6$ $(n = n_1 = n_2)$ Fallzahlen beider Gruppen

Beispiel 10.3 V

7. Ablehnbereich: $n^* = 8$,

kritischer Wert:

- $R_{n*=8;0,95;zweiseitig} = 3$
- Ablehnbereich also [0, 3]

Kritische Werte für Wilcoxon-Test

Test	zweiseitig			einseitig		Test	zweiseitig			einseitig	
n	5%	1%	0,1%	5%	1%	n	5%	1%	0,1%	5%	1%
6	0			2		54	514	445	368	550	473
7	2			3	0	55	536	465	385	573	493
8	3	0		5	1	56	557	484	402	595	514
9	5	1		8	3	57	579	504	420	618	535
10	8	3		10	5	58	602	525	438	642	556
11	10	5	0	13	7	59	625	546	457	666	578
12	13	7	1	17	9	60	648	567	476	690	600
13	17	9	2	21	12	61	672	589	495	715	623
14	21	12	4	25	15	62	697	611	515	741	646
15	25	15	6	30	19	63	721	634	535	767	669
16	29	19	8	35	23	64	747	657	556	793	693
17	34	23	11	41	27	65	772	681	577	820	718
18	40	27	14	47	32	66	798	705	599	847	742
19	46	32	18	53	37	67	825	729	621	875	768
20	52	37	21	60	43						

Beispiel 10.3 VI

8. Berechnung der Prüfgröße:

- $R = \min \begin{Bmatrix} R_+ = 6{,}5, \\ R_- = 29{,}5 \end{Bmatrix}$
  $= 6{,}5$
- Kontrolle:
- $R_+ + R_- = 29{,}5 + 6{,}5$
  $= 36 = n^* \cdot \frac{n^*+1}{2}$
  $= 8 \cdot \frac{9}{2} = 36$

Spieler i	Abfluggeschwindigkeit des Balls		$d_i=x_i-y_i$	$\lvert d_i \rvert$	Rang von $\lvert d_i \rvert$	Rang R_- $d_i<0$	Rang R_+ $d_i>0$
	vor Training x_i	nach Training y_i					
1	41	49	-8	8	5,5	5,5	
2	44	41	3	3	1		1
3	48	56	-8	8	5,5	5,5	
4	55	51	4	4	2		2
5	57	72	-15	15	8	8	
6	58	52	6	6	3,5		3,5
7	60	66	-6	6	3,5	3,5	
8	68	80	-12	12	7	7	
Summe						$R_-=29{,}5$	$R_+=6{,}5$

Beispiel 10.3 VII

9. Testentscheidung: $R = 6{,}5 > R_{n*=8;0{,}95;zweiseitig} = 3$

⇨ d. h., H_0 wird beibehalten.

10. Interpretation:

Die mediane Abfluggeschwindigkeit vor dem Training lag bei 56 km/h, nach dem Training bei 54 km/h.

Diese Differenz ist zum Signifikanzniveau 5% nicht signifikant.

Statistik und Interpretation

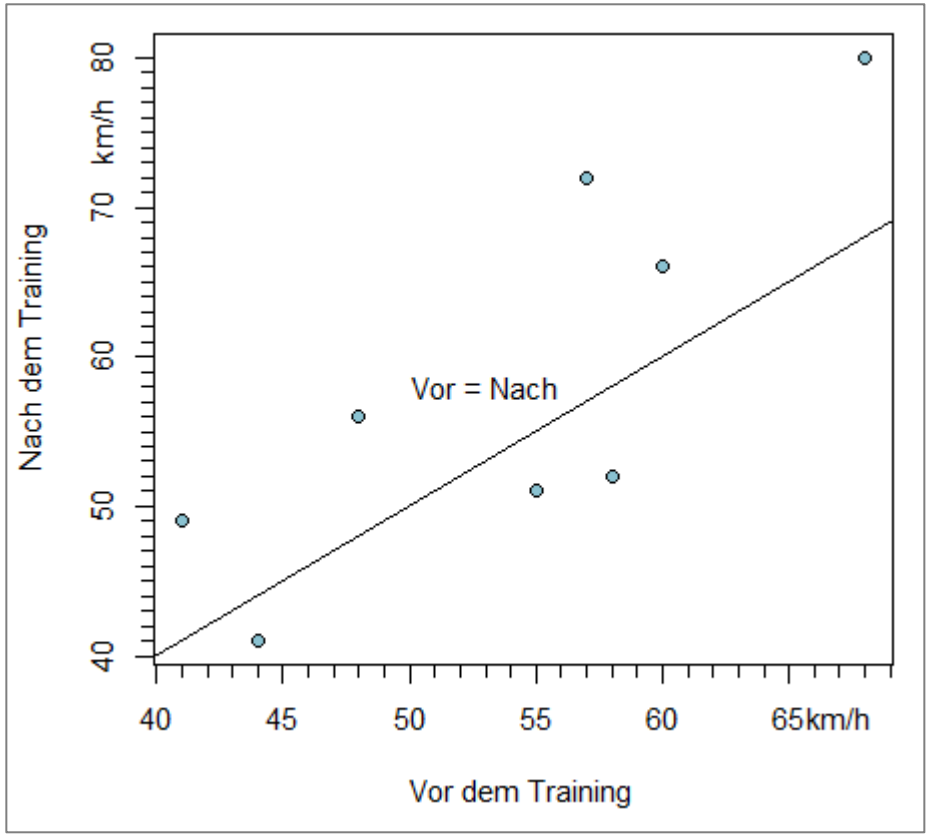
Kritisches Nachdenken ist wesentlich:

- 5 Spieler steigerten ihre Abwurfgeschwindigkeit,
- 3 Spieler verschlechterten sich deutlich.
- Welche Differenz gilt als relevant?
- Der Test liefert keine statistische Signifikanz, aber dennoch ein für den Trainer wichtiges (significant) Ergebnis.
- Die Kausalität hinter den Zahlen bleibt spekulativ (z. B. Ermüdungserscheinungen nach Training).

Differenz
-8
3
-8
4
-15
6
-6
-12

Ergänzende Grafik

– Der Scatterplot Vor vs. Nach zeigt deutlich die Entwicklung der Geschwindigkeiten.



Anmerkungen: Rangtests zum Lagevergleich

- Für n > 30 sind die Prüfgrößen approximativ standardnormalverteilt ~ *N(0,1);*
- Die Statistiksoftware SPSS wendet standardmäßig für n > 30 die Normal-approximation an. (Dann könnte man meist auch den Welch-Test anwenden.)
- Die Berechnung der asymptotischen Prüfgrößen erfolgt nach folgenden Formeln:

U-Test: $z = \frac{U - \frac{m \cdot n}{2}}{\sqrt{\frac{m \cdot n \cdot (m+n+1)}{12}}}$

Wilcoxon-Test: $z = \frac{R - \frac{n^* \cdot (n^*+1)}{4}}{\sqrt{\frac{n^* \cdot (n^*+1) \cdot (2 \cdot n^*+1)}{24}}}$



- Falls n ≤ 30, sollte man darauf achten, dass der exakte Test angewandt wird.

Rangtests zum Lagevergleich von Verteilungen

Anzahl der Stichproben	Art der Stichproben	parametrischer Test	Statistisches Verfahren
1	----	ja	t-Test (Normwertvergleich)
		nein	Wilcoxon-Test (Normwertvergleich)
2	unabhängig	ja	t-Test für unabhängige Stichproben
		nein	**U-Test nach Mann-Whitney** ✓
	abhängig	ja	t-Test für Paardifferenzen
		nein	**Wilcoxon-Test** ✓
> 2	unabhängig	ja	univariate Varianzanalyse
		nein	H-Test nach Kruskal-Wallis
	abhängig	ja	multivariate Varianzanalyse
		nein	Friedman-Test

V 10

10.2 Vergleich der Lage von Verteilungen aus mehr als 2 Stichproben

Lagevergleich für mehr als 2 Stichproben I

Idee:

- mehrere (paarweise) 2-Gruppen-Vergleiche durchführen

- CAVE: In jedem Test kann es einen Fehler I. oder II. Art geben. Wiederholtes Testen kumuliert die Wahrscheinlichkeit für fehlerhafte Entscheidungen!



Jotquadrat: Grafitto in Bonn. commons. wikipedia.org

Wiederholung: Fehler 1. und 2. Art

– Mögliche Fehlentscheidungen:

Testentscheidung aufgrund Stichprobe	*wahre* Situation (Biologie)	
	H_0 richtig	H_A richtig
H_0 beibehalten	richtig $1 - \alpha$	falsch negativ β Fehler 2. Art
H_0 verwerfen und H_A annehmen	falsch positiv α Fehler 1. Art	richtig $1 - \beta$ (*Power*)

– Fehler 1. Art: Fälschliches Verwerfen der Null-Hypothese

– Fehler 2. Art: Fälschliches Beibehalten der Null-Hypothese

Multiples Testen

Problem: Gesamt – α erhöht sich mit der Anzahl der durchgeführten Tests:

- Der Test ist so konstruiert, dass der Fehler I. Art durch das Signifikanzniveau α begrenzt wird.
- k … Anzahl der durchgeführten Tests
- Bei jedem Test kann ein Fehler I. Art auftreten. Bei k Test ergibt sich ein
- $\alpha_{gesamt} = 1 - (1 - \alpha)^k$

zum Beispiel: 4 Tests zum Signifikanzniveau 5%

- $\alpha_{gesamt} = 1 - (1 - 0{,}05)^4 = 0{,}185$ (18%!)

Risiko beim multiplen Testen ☺

- Dr. Sorglos klettert mit 20 miteinander verknüpften Seilen, von denen jedes mit je 5% Wahrscheinlichkeit reißt.
- Wie hoch ist die Wsk., dass das Seil hält?

- $P(\text{Seil hält}) = 0{,}95^{20} = 0{,}358$

aus: Dubben / Beck-Bornholdt:
Der Hund, der Eier legt, S. 66



Lagevergleich für mehr als 2 Stichproben II

Lösung des Problems: (einfaktorielle) Varianzanalyse

– vermag, Mittelwerte von $k \geq 2$ Gruppen zu vergleichen:
  H_0: $\mu_1 = \mu_2 = \ ... = \mu_k$ vs. H_A: $\mu_i \neq \mu_j$ für ein Paar (i, j)

1. Zunächst beantwortet ein Globaltest die Frage
   Gibt es überhaupt einen Unterschied?
2. Wird das mit Ja beantwortet, prüft ein Post-hoc-Test die Frage:
   Die Mittelwerte welcher Gruppen unterscheiden sich?

– Dabei wird eine Korrektur für multiples Testen vorgenommen
  (z. B. Bonferroni-Korrektur).

Zusammenfassung

- Nichtparametrische Tests werden ohne konkrete Verteilungs-Annahmen abgeleitet.
  - Zu jedem t-Test-Szenario gibt es analoge Rangtests.
  - Diese passen für ordinale bzw. metrische Größen bei kleinen Stichproben oder Ausreißern.
- Beim Mehr-Gruppen-Vergleich erfolgen paarweise Tests nur, wenn der Globaltest signifikant war.
- Ziel ist eine hohe Power von Tests. Sie hängt ab
  - von der Effektstärke (z. B. Mittelwertunterschied) und
  - von der Variation (z. B. Standardabweichung)
  - Bei gegebenem Szenario kann sie nur durch eine höhere Fallzahl erhöht werden.

UNIVERSITÄT LEIPZIG

Medizinische Fakultät

VIELEN DANK!

Dr. rer. nat. Meinhard Mende

Senior-Wissenschaftler, Institut für Medizinische Informatik, Statistik und Epidemiologie und Zentrum für Klinische Studien

Härtelstraße 16-18, 04107 Leipzig

Meinhard.mende@zks.uni-leipzig.de