

Vorlesung Statistik – WiSe 2016/17

Vorlesung 5 Nichtparametrische Tests 1.12.2016

Prof. Dr. Peter Martus
UKT Tübingen

Studiengänge
Molekulare Medizin FS 3(BSc)
Biochemie FS 3(BSc)
Medizininformatik FS 3 (BSc)

Vorlesung 5

- Die Rangtransformation und nichtparametrische Tests
- Der Mann-Whitney Test
- Der Wilcoxon Test für abhängige Daten
- Der Vorzeichentest
- Übersicht der Testverfahren
- Die Spearman Korrelation

Teil 1

Die Rangtransformation und nichtparametrische Tests

Die Rangtransformation

Bei der Rangtransformation werden die Messwerte der Größe nach sortiert, der kleinste Wert erhält die 1, der größte den Rang n.

Beispiel 1:

Werte	3.7	2.5	4.1	3.8	3.2	4.0
Ränge	3	1	6	4	2	5

Gleiche Werte werden auch als Bindungen bezeichnet. Bei Bindungen wird einfach der Mittelwert der zugehörigen Rangwerte verwendet.

Beispiel 2:

Werte	3.7	2.5	4.1	3.7	3.2	4.0
Ränge	3.5	1	6	3.5	2	5

Beispiel 3:

Werte	3.7	2.5	4.1	3.7	3.7	4.0
Ränge	3	1	6	3	3	5

Nichtparametrische Tests

Wenn quantitative Daten nicht normalverteilt sind oder die Skala ordinal ist, sollten keine t-Tests verwendet werden. Im allgemeinen verwendet man dann Verfahren, die auf der Rangtransformation beruhen.

Wir werden hier die beiden Tests behandeln, die als „Ersatz“ für den t-Test für unverbundene und für den t-Test für unverbundene Stichproben verwendet werden.

Es handelt sich dabei um den Test von Mann-Whitney und den Test von Wilcoxon für verbundene Stichproben.

Zum Schluss wird die Spearman-Korrelation nachgetragen, die analog die bisher behandelte Korrelation (Pearson) ersetzen kann.

Teil 2

Der Mann-Whitney Test

Der Mann-Whitney Test

Für zwei unabhängige Stichproben wird anstelle des t-Tests der Test von Mann-Whitney verwendet. Die Grundidee ist einfach:

Allen Paaren (x_A, x_B) mit Werten aus Gruppe A und Gruppe B werden zwei Zahlen zugeordnet (von denen die zweite redundant ist).

	Zahl 1	Zahl 2
$x_A < x_B$	1	0
$x_A = x_B$	0.5	0.5
$x_A > x_B$	0	1
Summe	U_A	U_B

Liegen diese Summen nahe genug bei 0 bzw. $n_A * n_B$ sind die Unterschiede zwischen beiden Gruppen signifikant.

Die Prüfgröße des Mann Whitney Tests

Nur um Rechenzeit zu sparen, werden in der Praxis andere Formeln zur Berechnung von U_A und U_B benutzt.

n_A Fallzahl Gruppe A

R_A Rangsumme Gruppe A

n_B Fallzahl Gruppe B

R_B Rangsumme Gruppe B

$$U_A = n_A * n_B + n_A * (n_A + 1) / 2 - R_A,$$

$$U_B = n_A * n_B + n_B * (n_B + 1) / 2 - R_B$$

Als Prüfgröße wird das Minimum von U_A und U_B verwendet.

$$U = \text{Min}(U_A, U_B)$$

Die Ergebnisse sind signifikant, wenn der Wert der Prüfgröße kleiner oder gleich dem kritischen Wert ist, der einer Tabelle entnommen werden kann.

Beispiel zum Mann-Whitney

Therapie	Blutdrucksenkung [mmHg]									
A	13	11	17	18	14	9	15	0	14	12
Rang	6.5	3	14	16.5	8.5	2	11	1	8.5	5
B	17	19,5	18	16	17	13	11,5	32,5	21	14,5
Rang	14	18	16.5	12	14	6.5	4	20	19	10

In beiden Stichproben ist die Schiefe nicht im Intervall $[-1,1]$, sie ist **-1.7 für Therapie A und +1.8 für Therapie B.**

Berechnungen für den Mann-Whitney Test – Variante 1

An der folgenden Tabelle können wir die Vergleiche x_A vs. x_B direkt ablesen.

0	9	11		12	13		14	14		15		17			18				
			11.5			13			14.5		16		17	17		18	19.5	21	32.5
10	10	10		9	8.5		8	8		7		5			3.5				

Die Summe der Vergleichszahlen für Gruppe A ist 79, entsprechend die für Gruppe B $10 \cdot 10 - 79 = 21$.

Dasselbe Ergebnis sehen wir auf den folgenden Folien unter Verwendung der Formeln von Folie ...

Berechnungen für den Mann-Whitney Test – Variante 2

$$R_A = \mathbf{76}, R_B = \mathbf{134}$$

$$U_A = n_1 * n_2 + n_1 * (n_1 + 1) / 2 - R_A,$$

$$U_A = 10 * 10 + 10 * 11 / 2 - 76 = \mathbf{79},$$

$$U_B = n_1 * n_2 + n_2 * (n_2 + 1) / 2 - R_B$$

$$U_B = 10 * 10 + 10 * 11 / 2 - 134 = \mathbf{21}$$

Minimum: **21**

Interpretation

Der kritische Wert ist **23** (2 Gruppen, jeweils 10 Personen, Signifikanzniveau 0.05, zweiseitig).

Der Wert in der Stichprobe ist 21. In diesem Fall erhalten wir Signifikanz, weil es darauf ankommt, dass der Stichprobenwert kleiner oder gleich dem kritischen Wert ist.

Wir haben also gezeigt, dass ein signifikanter Unterschied zwischen beiden Gruppen existiert.

Aber was genau haben wir gezeigt?

Die Werte in Gruppe B sind größer als die in Gruppe A!

Die Mittelwerte? Die Mediane? (bei Ordinalskalen??)

Der Mann-Whitney Test

Die mathematische Hypothesenformulierung ist beim Mann-Whitney Test nicht exakt identisch zum t-Test. Strenggenommen lauten die Hypothesen:

H_0 : Die Verteilungsfunktionen der Werte in beiden Gruppen sind identisch.

H_1 : Die Verteilungsfunktion sind nicht identisch.

Wenn sich die Verteilungsfunktionen durch eine konstante Verschiebung unterscheiden (sog. „shift Alternative“), ist die folgende Hypothesenstruktur gleichwertig:

H_0 : Die Mediane in beiden Gruppen sind identisch.

H_1 : Die Mediane sind verschieden.

(Nur) für diesen Fall lautet also die Alternative der Median der zweiten Gruppe ist größer/kleiner (einseitig) oder ungleich (zweiseitig) den der zweiten Gruppe.

Dieselbe Interpretation wird aber oft gewählt, ohne dass tatsächlich eine shift Alternative vorliegt bzw. ohne dass dies überprüft wird!

Teil 3

Der Wilcoxon-Test für abhängige Stichproben

Der Wilcoxon Test für abhängige Stichproben

Wenn die Differenzen bei abhängigen Stichproben nicht normalverteilt sind, tritt der Wilcoxon Test für abhängige Stichproben an die Stelle des t-Tests.

Eine Rangtransformation ist auch Grundlage des Wilcoxon Tests. Aber diesmal sind es die Differenzen, für die Ränge vergeben werden.

Wir betrachten wieder ein Beispiel zur Blutdrucksenkung. Diesmal ist die Frage, ob die Werte nach Therapie besser sind als vor Therapie.

Die Idee ist wieder recht einfach: Wir zählen wie viele Differenzen positives, wieviele negatives Vorzeichen haben. Aber wir gewichten zusätzlich nach der absoluten Größe der Differenzen. Eine große positive Differenz kann also mehrere kleine negative „ausgleichen“.

Algorithmus und Beispiel für den Wilcoxon Test

1. Berechne für jeden Patienten die Differenz Wert nach Therapie minus Wert vor Therapie
2. Streiche die Nulldifferenzen (!)
3. Bilde die Absolutwerte der Differenzen
4. Ordne diesen Absolutwerten Rangzahlen zu
5. Vergleiche die Rangsummen für positive und negative Differenzen

Pat. Nr.	1	2	3	4	5	6	7	8	9	10
vorher	119.2	156.0	145.0	137.2	124.8	160.4	141.3	144.1	151.5	130.8
nachher	107.8	156.0	137.0	142.7	112.4	125.7	135.6	125.1	143.4	136.5
Differenz	-11.4	0	-8.0	5.5	-12.4	-34.7	-5.7	-19	-7.7	5.7
Vorzeichen	-	ignoriert	-	+	-	-	-	-	-	+
Rang	6		5	1	7	9	2.5	8	4	2.5

$$\mathbf{R^-} = 6+5+7+9+2.5+8+4 = 41.5, \mathbf{R^+} = 1+2.5 = 3.5, \mathbf{Min(R^-,R^+)} = 3.5$$

Signifikanz und P-Wert:

Nach Streichung der Nulldifferenz ist die Fallzahl $n = 9$.

Von 9 Differenzen sind 7 negativ (Blutdruck wurde gesenkt), und 2 positiv (Blutdruck ist angestiegen).

Die Rangsummen sind 41.5 für die 7 negativen Differenzen (d.h. Durchschnittsrang ca. 6) und 3.5 für die 2 positiven Differenzen (Durchschnittsrang ca. 2). Insofern ist das Ergebnis durch Berücksichtigung der Rangzahlen noch deutlicher geworden.

Für den Test wird die kleinere Rangzahl $\min(R^-, R^+)$ mit dem kritischen Wert verglichen. Aus einer Tabelle können wir entnehmen, dass der kritische Wert für neun Differenzen und $\alpha = 0.05$ (zweiseitig) $\min(R^-, R^+) = 5$ beträgt. Das Ergebnis ist also signifikant.

Der P-Wert beträgt 0.024.

Teil 4

Der Vorzeichen test

Der Vorzeichentest

Wenn man die Größenordnung der Differenzen vernachlässigt, vergleicht man nur die Anzahl positiver und negativer Differenzen.

Pat. Nr.	1	2	3	4	5	6	7	8	9	10
vorher	119.2	156.0	145.0	137.2	124.8	160.4	141.3	144.1	151.5	130.8
nachher	107.8	156.0	137.0	142.7	112.4	125.7	135.6	125.1	143.4	136.5
Differenz	-11.4	0	-8.0	5.5	-12.4	-34.7	-5.7	-19	-7.7	5.7
Vorzeichen	-	ignoriert	-	+	-	-	-	-	-	+
Rang	6		5	1	7	9	2.5	8	4	2.5

Jetzt entspricht der Nullhypothese, dass die Wahrscheinlichkeiten für positive und negative Differenzen jeweils gleich, also 50% sind (0-Differenzen werden ja ignoriert).

Die Berechnung erfolgt mit Hilfe der Binomialverteilung, also mit denselben Formeln wie das Eingangsbeispiel aus VL 3.

Der Vorzeichentest

Wir berechnen also die Wahrscheinlichkeit, dass mindestens 7 von 9 Differenzen negativ sind, wenn die Wahrscheinlichkeit für eine negative Differenz 50% wäre.

$$\begin{aligned} & \binom{9}{7} \times 0.5^7 \times 0.5^2 + \binom{9}{8} \times 0.5^8 \times 0.5^1 + \binom{9}{9} \times 0.5^9 \times 0.5^0 \\ & = \\ & 0.5^9 \times (36 + 9 + 1) = \frac{1}{512} \times 46 \approx 8.8\% \end{aligned}$$

Der einseitige P-Wert ist 8.8%, aufgrund der Symmetrie der Binomialverteilung mit $p = 0.05$ ist der zweiseitige P-Wert ca. 17.6%.

Mit Hilfe des Vorzeichentests konnten wir also keinen Unterschied nachher minus vorher beweisen.

Der Vorzeichentest verwendet weniger Information und hat deswegen weniger Power als t-Test und Wilcoxon-Test.

Der Vorzeichentest für binäre Daten

Verbundene Stichproben können auch in Therapiestudien für binäre Outcomes (Therapieerfolg ja/nein) zum Therapievergleich verwendet werden. Dies setzt allerdings voraus, dass derselbe Patient mit zwei verschiedenen Therapien behandelt werden kann, was vor allem bei chronischen Erkrankungen eine realistische Option darstellt.

Probanden mit chronischen Schmerzen sollen angeben, ob sie sich im vergangenen Monat insgesamt durch die Schmerzen in ihrem täglichen Leben beeinträchtigt fühlten. Der Einfachheit halber soll die Frage mit ja oder nein beantwortet werden.

Die Probanden erhalten in zwei Monaten jeweils Medikament A oder Medikament B (die Reihenfolge sollte zufällig gewählt werden). Die Probanden geben für jede der beiden Therapiephasen also die Information Beeinträchtigung ja/nein.

Beispiel zum Vorzeichentest für binäre Daten

Pat. Nr.	1	2	3	4	5	6	7	8	9	10
Therapie A	1	1	0	1	1	1	1	1	0	1
Therapie B	0	1	0	0	0	1	0	0	1	0
Differenz	1	0	0	1	1	0	1	1	-1	1
Vorzeichen	+	ignoriert	ignoriert	+	+	ignoriert	+	+	-	+

Genau wie beim vorangegangenen Beispiel können wir jetzt fragen:
Wie groß ist die Wahrscheinlichkeit, dass wir bei 7 Patienten mit unterschiedlichen Ergebnissen, 6 mal Therapie A und nur einmal Therapie B erfolgreich sehen?

Hierfür verwenden wir wieder Binomialwahrscheinlichkeiten:

$$\binom{7}{6} \times 0.5^6 \times 0.5^1 + \binom{7}{7} \times 0.5^7 \times 0.5^0 = 0.5^7 \times (7 + 1) = \frac{1}{128} \times 8 = 6.25\%$$

Offenbar ist der einseitige P-Wert 6.25%, der zweiseitige wäre 12.5%, in beiden Fällen erhalten wir kein signifikantes Ergebnis

Darstellung des Vorzeichentests über eine Kreuztabelle

Für jeden Patienten trifft genau eines der 4 Muster zu:

Therapie A	Therapie B	Anzahl
Misserfolg	Misserfolg	n_{00}
Misserfolg	Erfolg	n_{01}
Erfolg	Misserfolg	n_{10}
Erfolg	Erfolg	n_{11}

Und wir können die Ergebnisse kompakt in einer Kreuztabelle darstellen:

		Therapie B		
		Misserfolg	Erfolg	Gesamt
Therapie A	Misserfolg	$1 = n_{00}$	$1 = n_{01}$	$n_{00} + n_{01}$
	Erfolg	$6 = n_{10}$	$2 = n_{11}$	$n_{10} + n_{11}$
	Gesamt	$n_{00} + n_{10}$	$n_{01} + n_{11}$	n_{gesamt}

Vergleich mit dem Chi-Quadrat Test

Obwohl die Darstellungen der Daten für Chi-Quadrat und Vorzeichentest in der gleichen Tabellenstruktur erfolgen, sind die Analysen und Fragestellungen komplett verschieden!

		Therapie B		
		Misserfolg	Erfolg	Gesamt
Therapie A	Misserfolg	n_{00}	n_{01}	$n_{00} + n_{01}$
	Erfolg	n_{10}	n_{11}	$n_{10} + n_{11}$
	Gesamt	$n_{00} + n_{10}$	$n_{01} + n_{11}$	n_{gesamt}

		Misserfolg	Erfolg	Gesamt
Therapie	A	n_{00}	n_{01}	$n_{00} + n_{01}$
	B	n_{10}	n_{11}	$n_{10} + n_{11}$
	Gesamt	$n_{00} + n_{10}$	$n_{01} + n_{11}$	n_{gesamt}

Teil 5

Überblick Statistische

Tests

Übersicht der bisher behandelten Testverfahren

	Art der Stichprobe		
Skala/Verteilung	Eine	Zwei abhängige	Zwei unabhängige
Nominal	Nicht behandelt (Chi-Quadrat Test)	Nicht Behandelt (Bowker Test)	Chi-Quadrat Test
Binär	Binomialtest (VL 3)	Vorzeichentest	Chi-Quadrat Test
Ordinal	Analog zum Wilcoxon Test	Wilcoxon Test für abhängige Stichproben	Mann-Whitney Test
Quantitativ nicht Normalverteilt	Analog zum Wilcoxon Test	Wilcoxon Test für abhängige Stichproben	Mann-Whitney Test
Quantitativ Normalverteilt	t-Test für eine Stichprobe	t-Test für abhängige Stichproben	t-Test für unabhängige Stichproben

Man kann für jede Skala auch die Tests der darüber stehenden Skalen verwenden (außer diejenigen der Nominalskala). Ggf. muss man die Daten in Kategorien aufteilen.

Alternative Bezeichnungen für statistische Tests

Stichproben

Unabhängige = Unverbundene = Ungepaarte Stichproben

Abhängige = Verbundene = Paarige Stichproben

Tests für binäre Daten

Vorzeichentest für binäre Daten = Mc Nemar Test

Nichtparametrische Tests

Mann-Whitney-Test = Rangsummentest = Wilcoxon-Test für unabhängige Stichproben

Wilcoxon-Test für abhängige Stichproben = Vorzeichenrang-Test

Teil 6

Die Spearman Korrelation

Die Korrelation nach Spearman

Auch für den Pearson-Korrelationskoeffizienten r (VL 2) existiert ein „Ersatzverfahren“ für nicht normalverteilte Daten, der Spearman-Korrelationskoeffizient r_s . Basis sind jeweils die Ränge getrennt für die x - und die y -Variable.

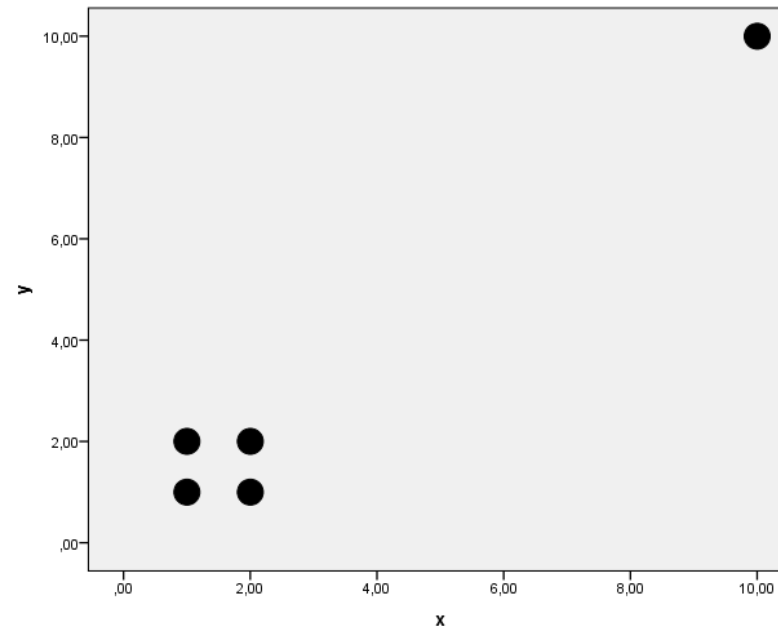
Variante 1: Man berechnet r für die Ränge.

Variante 2: Man berechnet für jedes Wertepaar die Rangdifferenz d .
Dann verwendet die Formel

$$r_s = 1 - \frac{6 \times \sum_{i=1}^n d_i^2}{n \times (n^2 - 1)}$$

Beispiel zur Korrelation nach Spearman

X	1	1	2	2	10
Y	1	2	1	2	10



$$r = 0.983 \quad r_s = 0.556$$