

Grundlagen der Biostatistik und Informatik

Hypothesenprüfungen IV.

Nichtparametrische Methoden

dr. László Smeller
Semmelweis Universität
2018

Übersicht der Teste

Verteilung Stichproben	Normalverteilte Daten	Die Verteilung der Daten ist unbekannt
Eine Stichprobe	Einstichproben t-Test	Vorzeichnentest Wilcoxon Test
Zwei Stichproben	Zweistichproben t-test	Mann-Whitney U-Test
Mehrere Stichproben	ANOVA (Varianzanalyse)	Kruskal-Wallis Test

2

Nichtparametrische Methoden

Als Erinnerung: **Bedingungen der t-Teste:**

- kontinuierliches Merkmal (z.B. Körperhöhe, Körpertemperatur...)
- die Daten müssen eine Normalverteilung folgen!



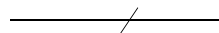
Nichtparametrische Methoden

- nur ordinale Daten (Ordinalskala)
- keine Normalverteilung (auch bei unbekannter Verteilung möglich!)

z. B. Schmerzmittel – Wie stark ist ein Schmerz? Kann nur auf einer Ordinalskala gemessen werden:

1, 2, 3, 4, 5

oder



3

Nichtparametrische Methoden

Vorteile:

- Verteilungsunabhängigkeit
- Ordinal-, Intervall-, Verhältnisskalen

Nachteile:

- Datenreduktion, Informationsverlust
- Kleinere Teststärke
- größere Wahrscheinlichkeit der Fehler 2. Art:
Nur größere Unterschiede können detektiert werden
als bei den parametrischen Teste

4

Übersicht der Teste

Verteilung Stichproben	Normalverteilte Daten	Die Verteilung der Daten ist unbekannt
Eine Stichprobe	Einstichproben t-Test	Vorzeichentest Wilcoxon Test
Zwei Stichproben	Zweistichproben t-test	Mann-Whittney U-Test
Mehrere Stichproben	ANOVA (Varianzanalyse)	Kruskal-Wallis Test

5

Eine Stichprobe: Vorzeichentest

Daten oder Datenpaaren, (Änderung oder Unterschied)

Datenreduktion:

Nur zwei Alternativen (einander ausschließende Ereignisse)

z.B.: Verbesserung oder Verschlechterung des Krankheitszustandes

Erfolg - Misserfolg

⇒ **Binomialverteilung**

Hat das Medikament eine Wirkung? D.h.: Sind signifikant mehr Fällen mit Verbesserung als mit Verschlechterung? Haben die zwei Alternativen unterschiedliche Wahrscheinlichkeit?

H_0 : Die zwei Alternativen (z.B. Verbesserung und Verschlechterung) haben dieselbe Wahrscheinlichkeit. ($p=1/2$, $q=1/2$)

Analogie: Münzenexperiment: Kopf oder Zahl

6

Vorzeichentest: Beispiel des Kopfschmerzes

Stärke von Kopfschmerzen vor und nach der Einnahme des Medikamentes werden an einer relativen Skala gegeben.

Der Kopfschmerz erniedrigt in k aus n Fällen (in $n-k$ Fällen es erhöht sich. Die „keine Änderung“ Fällen werden nicht beachtet.)

Ist die Änderung signifikant?

H_0 : Das Medikament ist unwirksam, d.h. Erniedrigung und Erhöhung des Kopfschmerzes sind gleich wahrscheinlich.

Beispiel1.: Kopfschmerzen sinkt bei 7 aus 9 Patienten.

Beispiel2.:Kopfschmerzen sinkt bei 9 aus 10 Patienten.

Bei Gültigkeit der Nullhypothese: Analogie mit dem Münzenexperiment:

Experiment: k -mal Kopf aus n Versuche.

Analogie für Beispiel 1.: 7-mal Kopf aus 9 Versuche

Analogie für Beispiel2.: 9-mal Kopf aus 10 Versuche

7

Vorzeichentest: Analogie mit dem Münzenexperiment

Bei Münzenexperiment kann man die Wahrscheinlichkeit der unterschiedlichen Fällen ausrechnen (Binomialverteilung!):

		Anzahl von Experimenten									
		1	2	3	4	5	6	7	8	9	10
Anzahl von Kopf	0	50.0%	25.0%	12.5%	6.3%	3.1%	1.6%	0.8%	0.4%	0.2%	0.1%
	1	50.0%	50.0%	37.5%	25.0%	15.6%	9.4%	5.5%	3.1%	1.8%	1.0%
	2		25.0%	37.5%	37.5%	31.3%	23.4%	16.4%	10.9%	7.0%	4.4%
	3			12.5%	25.0%	31.3%	31.3%	27.3%	21.9%	16.4%	11.7%
	4				6.3%	15.6%	23.4%	27.3%	27.3%	24.6%	20.5%
	5					3.1%	9.4%	16.4%	21.9%	24.6%	24.6%
	6	Binomialverteilung					1.6%	5.5%	10.9%	16.4%	20.5%
	7						0.8%	3.1%	7.0%	11.7%	
	8							0.4%	1.8%	4.4%	
	9								0.2%	1.0%	
	10									0.1%	

8

Vorzeichentest: Anwendung der Binomialverteilung

Bei Gültigkeit der H_0 gibt dieselbe Tabelle die Wahrscheinlichkeit der entsprechenden Fällen:

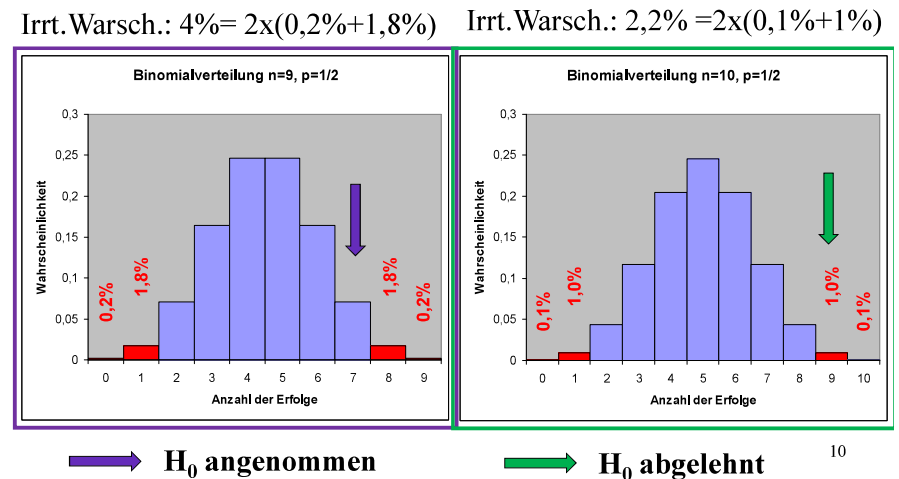
Anzahl der Verbesserungen	Anzahl von Patienten										
	1	2	3	4	5	6	7	8	9	10	
	0	50.0%	25.0%	12.5%	6.3%	3.1%	1.6%	0.8%	0.4%	0.2%	0.1%
	1	50.0%	50.0%	37.5%	25.0%	15.6%	9.4%	5.5%	3.1%	1.8%	1.0%
	2		25.0%	37.5%	37.5%	31.3%	23.4%	16.4%	10.9%	7.0%	4.4%
	3			12.5%	25.0%	31.3%	31.3%	27.3%	21.9%	16.4%	11.7%
	4				6.3%	15.6%	23.4%	27.3%	27.3%	24.6%	20.5%
	5					3.1%	9.4%	16.4%	21.9%	24.6%	24.6%
	6	Binomialverteilung					1.6%	5.5%	10.9%	16.4%	20.5%
	7							0.8%	3.1%	7.0%	11.7%
	8							0.4%	1.8%	4.4%	
	9								0.2%	1.0%	
	10									0.1%	

$$p_k = \binom{n}{k} p^k q^{n-k}$$

Irrtumswahrscheinlichkeit=5% (2.5%+2.5%)

Vorzeichentest: Anwendung der Binomialverteilung

Graphisch:



Vorzeichentest: Beispiel 3, Überlebenszeit

Überlebenszeit bei behandelten Ratten mit einem Tumor. 168, 190, 280, 221, 110, 165, 179, 250, 195, 276 (Tage)

Es ist bekannt dass die Überlebenszeit der nicht behandelten Ratten mit dieser Art der Tumoren 170 Tage beträgt (Median!)

Ist der Median der Überlebenszeiten der behandelten Ratten unterschiedlich als 170 Tage?

H_0 : Median der Überlebenszeiten der behandelten Ratten beträgt 170 Tage.

168, 150, 280, 221, 230, 165, 179, 250, 195, 276

- - + + + - + + + +

Bei Gültigkeit der H_0 die Daten sind >170 Tage zu 50% Wahrsch.
< 170 Tage zu 50% Wahrsch.

Vorzeichentest: Anwendung der Binomialverteilung

Bei Gültigkeit der H_0 gibt diese Tabelle die Wahrscheinlichkeit der entsprechenden Fällen:

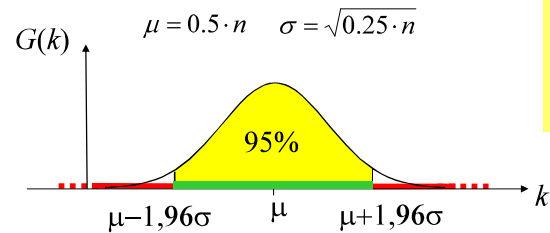
		Anzahl von Ratten									
		1	2	3	4	5	6	7	8	9	10
Anzahl der Ratten mit Überlebenszeit > 170 Tage	0	50.0%	25.0%	12.5%	6.3%	3.1%	1.6%	0.8%	0.4%	0.2%	0.1%
	1	50.0%	50.0%	37.5%	25.0%	15.6%	9.4%	5.5%	3.1%	1.8%	1.0%
	2		25.0%	37.5%	37.5%	31.3%	23.4%	16.4%	10.9%	7.0%	4.4%
	3			12.5%	25.0%	31.3%	31.3%	27.3%	21.9%	16.4%	11.7%
	4				6.3%	15.6%	23.4%	27.3%	27.3%	24.6%	20.5%
	5					3.1%	9.4%	16.4%	21.9%	24.6%	24.6%
	6	Binomialverteilung					1.6%	5.5%	10.9%	16.4%	20.5%
	7							0.8%	3.1%	7.0%	11.7%
	8								0.4%	1.8%	4.4%
	9									0.2%	1.0%
	10										0.1%

$$p_k = \binom{n}{k} p^k q^{n-k}$$

Irrtumswahrscheinlichkeit=5% (2,5%+2,5%)

Vorzeichentest: Annäherung bei $n > 20$

Annäherung bei $n > 20$ mit Normalverteilung:



Siehe: Binomialverteilung:
 $\mu = p \cdot n$
 $\sigma = \sqrt{p \cdot q \cdot n}$

z. B. 100 Patienten, 56 Verbesserungen, 34 Verschlechterungen

$H_0: ?; \mu_0 = ?; \mu = ?; \sigma = ?; \text{Entscheidung?}$

Analogie zu Einstichproben t-Test

(Lösung: $56 - 34 = 90$ $\mu = 45$ $\sigma = \sqrt{0.25 \cdot 100} = 5$ $\mu \pm 1.96 \cdot \sigma = 45 \pm 9.3 = 54.3 < 56 \Rightarrow \text{signifikant (5\% Irrtum)}$)

13

Übersicht der Teste

Verteilung	Normalverteilte Daten	Die Verteilung der Daten ist unbekannt
Stichproben		
Eine Stichprobe	Einstichproben t-Test	Vorzeichentest Wilcoxon Test
Zwei Stichproben	Zweistichproben t-test	Mann-Whitney U-Test
Mehrere Stichproben	ANOVA (Varianzanalyse)	Kruskal-Wallis Test

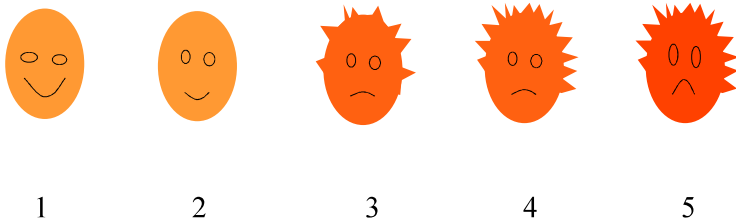
Rang-Teste

14

Prinzip der Rang Teste

Rang: Position eines Wertes innerhalb einer nach der Größe sortierten Wertereihe

z.B. Kopfschmerzen:



1

2

3

4

5

Mit Hilfe der Ränge führt man eine Gleichverteilung ein!

15

Verbundene Ränge

Wenn zwei oder mehrere ursprüngliche Daten gleich sind:

originale Daten	3; 16; 7; 1; 13; 16; 12
geordnete Daten	1; 3; 7; 12; 13; 16; 16
Ränge	1; 2; 3; 4; 5; 6,5; 6,5

Verbundene Ränge:

die bekommen den Durchschnittsrang

16

Durchschnitt der Ränge und Median der Daten

Median der Daten

originale Daten	3; 16; 7; 1; 13; 16; 12
geordnete Daten	1; 3; 7; 12; 13; 16; 16
Ränge	1; 2; 3; 4; 5; 6,5; 6,5

Durchschnittlicher Rang

17

Durchschnitt der Ränge

In steigende Reihe

geordnete Daten: $x_1, x_2, \dots, x_{(n-1)/2}, x_{(n+1)/2}, \dots, x_{n-1}, x_n$

Ränge: $1, 2, \dots, (n-1)/2, (n+1)/2, \dots, n-1, n$

(n ist ungerade)

$$\text{Durchschnitt der Ränge: } \bar{R} = \frac{1}{n} \sum_{i=1}^n i = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}$$

Durchschnittlicher Rang = Rang des Medians

Wenn n ist gerade:

$$\text{Median} = (x_{n/2} + x_{n/2+1})/2$$

$$\text{Durchschnittlicher Rang} = (n+1)/2$$

Rangteste testen den Median!

18

Eine Stichprobe: Wilcoxon-Vorzeichen Rangtest

Eine Stichprobe (Gepaarte Test)

Ordinale Daten (Symmetrische Verteilung)

Ist der Median der Datenreihe gleich Null?

(oder ein bestimmter Wert)?

H_0 : Der Median der Daten ist Null (oder ein bestimmter Wert: M_0).

Durchführung:

Die Daten bekommen Range nach ihren Betrag (Abstand von M_0)

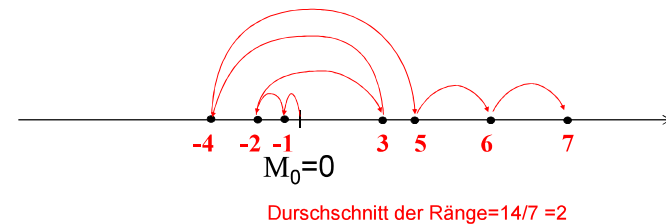
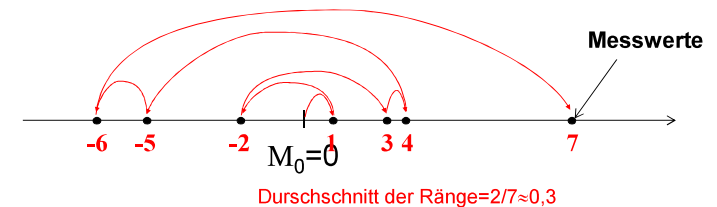
Die Ränge bekommen Vorzeichen.

Der Durchschnitt der Ränge wird geprüft.

Logik: Wenn die Nullhypothese gültig ist, es gibt gleich viele und gleich große positive und negative Ränge, Durchschnitt der Ränge ist Null!

19

Beispiele:



20

Wilcoxon-Vorzeichen Rangtest: Einführung mit einem Beispiel

Überlebenszeit der Ratten:

168, 150, 280, 221, 230, 165, 179, 250, 195, 276

Ist der Median der Überlebenszeiten unterschiedlich von 170 Tage?

H_0 : Der Median der Überlebenszeiten beträgt 170 Tage.

Überlebenszeitenunterschiede der Ratten im Vergleich zur 170 Tage:

-2, -20, +110, +51, +60, -5, +9, +80, +25, +106

Geordnet nach Betrag der Änderung:

-2, -5, +9, -20, +25, +51, +60, +80, +106, +110,

Ränge (nach Betrag der Änderung):

1, 2, 3, 4, 5, 6, 7, 8, 9, 10

Ränge mit Vorzeichen:

-1, -2, +3, -4, +5, +6, +7, +8, +9, +10

Durchschnitt: 4.10
Standardabw.: 4.91

21

Wilcoxon Vorzeichen Rangtest: Beispiel der Überlebenszeiten der Ratten

Der Durchschnitt folgt einer Normalverteilung, wenn genug viele Daten sind (Zentraler Grenzwertsatz)

Anwendung der t-Verteilung (Annäherung!):

$$t_{n-1} = \frac{\bar{R}}{\frac{s}{\sqrt{n}}}$$

← Durchschnitt der Ränge
 ← Standardabweichung der Ränge
 ← Anzahl der Daten

Freiheitsgrad

Entscheidung: wie beim Einstichproben t-Test

Ränge mit Vorzeichen:

-1, -2, +3, -4, +5, +6, +7, +8, +9, +10 → Durchschnitt: 4.10
Standardabw.: 4.91

$$t_9 = \frac{4,10}{4,91/\sqrt{10}} = 2,64$$

⇒ $t_9 > t_{9,5\%}$ ⇒ H_0 wird abgelehnt

$t_{9,5\%} = 2,26$ (aus der Tabelle) $p < 5\%$ (mit Excel)

22

Teststärke bei Tests für eine Stichprobe (Güte, Trennschärfe, Macht)

Je kleiner ist die Datenreduktion, desto besser ist die Teststärke.

z.B.:

Datenreduktion

Vorzeichentest > Wilcoxon Vorz.R.T. > t-Test

Teststärke:

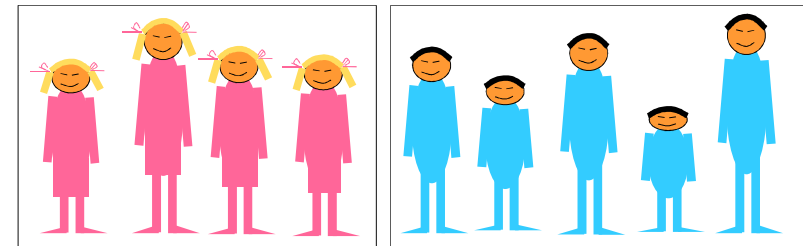
Vorzeichentest < Wilcoxon Vorz.R.T. < t-Test

↑ Für Daten mit
beliebiger symmetrischer
Verteilung (Ordinale Daten) ↑

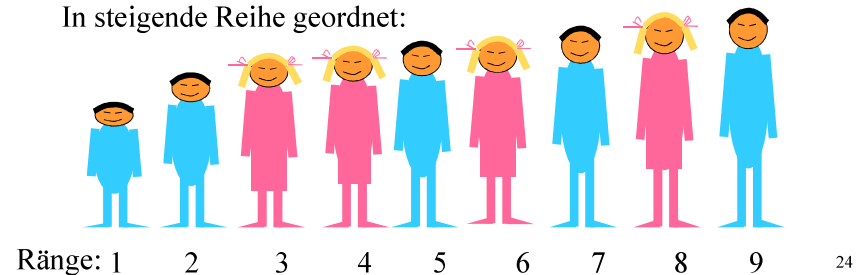
Nur für normalverteilte
Daten (Kontinuierliches
Merkmal)

23

Vergleich von zwei Stichproben: Mann-Whitney Test

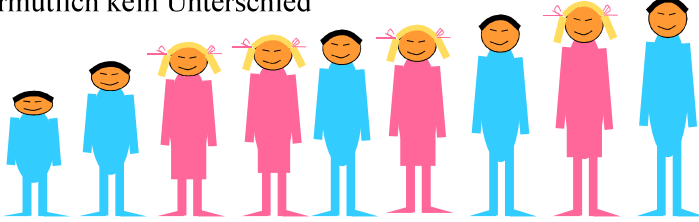


In steigende Reihe geordnet:

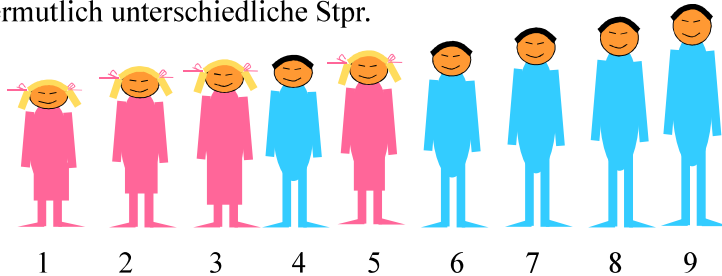


24

vermutlich kein Unterschied



vermutlich unterschiedliche Stpr.



25

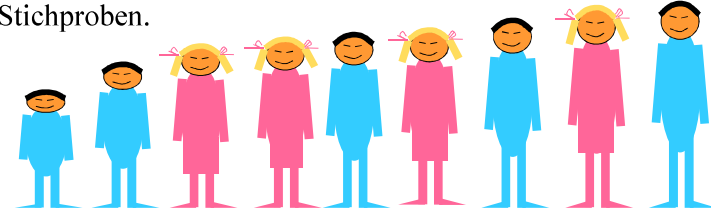
Mann – Whitney U Test (Annäherung)

(Auch als Wilcoxon Rank Summe Test genannt)

Vergleich von zwei Stichproben (n_1, n_2)

H_0 : Die zwei Stichproben stammen aus der selben Grundgesamtheit

1. Zuordnung der Ränge in den zwei zusammengeordneten Stichproben.



Ränge: 1 2 3 4 5 6 7 8 9

2. Bestimmung die Summen der Ränge in eine Gruppe: T_1 .

$$T_1 = 1+2+5+7+9=24$$

26

Mann – Whitney U Test: Annäherung

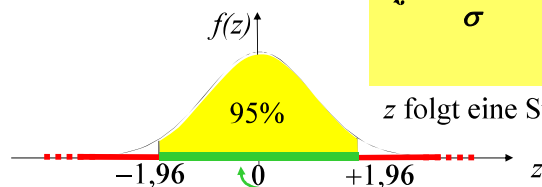
Bei Gültigkeit der Nullhypothese folgen die Daten der Gruppe 1 eine Gleichverteilung, mit möglichen werten von $1 \dots n_1+n_2$)

Erwartungswert und die theoretische Streuung von T_1 können berechnet werden:

$$\mu = \frac{n_1(n_1 + n_2 + 1)}{2}$$

$$\sigma = \sqrt{n_1 n_2 (n_1 + n_2 + 1) / 12}$$

$$z = \frac{T_1 - \mu}{\sigma} = \frac{T_1 - \frac{n_1(n_1 + n_2 + 1)}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}$$



z folgt eine Standard-Normalverteilung (wenn H_0 gültig ist)

z.B. $T_1=24, n_1=5, n_2=4 \Rightarrow z = -0,245 \Rightarrow H_0$ wird angenommen

27

Kruskal – Wallis Test

- Vergleich von mehreren Stichproben
- Mit unbekannter Verteilung der Daten
- Rangtestverfahren.

28