

Grundlagen der Mathematischen Statistik

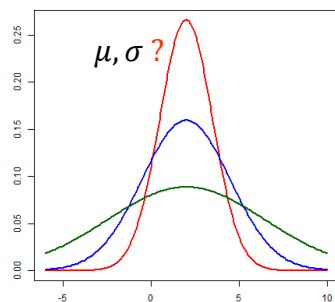
Nicht-parametrische Tests

Uwe Menzel, 2018
uwe.menzel@matstat.org
www.matstat.org

Testmethoden

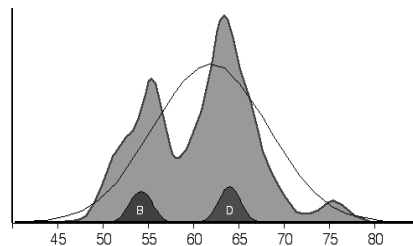
Parametrische Methoden

Die Verteilung der Population von der die Stichprobe stammt ist bekannt bis auf einen oder mehrere unbekannte Parameter, z. B. N mit unbekanntem μ und/oder σ .



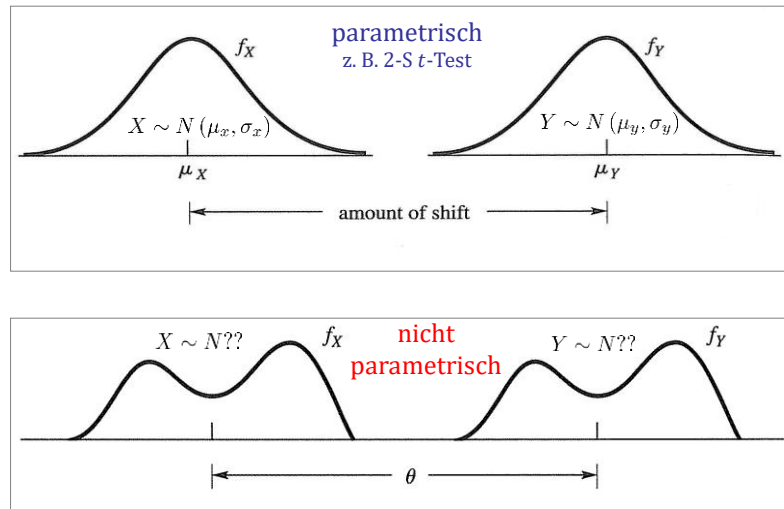
Nicht-parametrische Methoden

Es können keine sinnvollen Annahmen über die der Stichprobe zugrunde liegende Verteilung gemacht werden.



www.matstat.org

"Location shift"



www.matstat.org

Wozu nicht-parametrische Tests ?

- wenn die Voraussetzungen für einen parametrischen Test nicht vorliegen:
 - **2-Sample t-Test**: zugrunde liegende Populationen müssen normalverteilt sein
 - **One-Way ANOVA**: fordert (ungefähre) Normalverteilung für alle Gruppen, außerdem (ungefähr) homogene Varianzen
- wenn die Daten nicht oder schwer quantifiziert werden können:
 - **Ordinalskalen** (Größenverhältnisse liegen vor, aber die Differenzen sind im mathematischen Sinne bedeutungslos: z. B. Größe von Kleidungsstücken, Güteklassen: A, B, C, ...)
 - solche Daten kann man jedoch **rangordnen** (der Größe nach ordnen)

Wie gut funktionieren nicht-parametrische Tests ?

- fast so gut wie parametrische Tests wenn die Voraussetzungen wie Normalverteilung und/oder Varianzhomogenität vorliegen
 - **Trennschärfe** (power, $1 - \beta$) oft kleiner als bei parametrischen Tests
- oft besser wenn die Voraussetzungen (Normalverteilung usw.) **nicht** erfüllt sind

www.matstat.org

Welcher Test wird mit welchem "ersetzt"?

Population normalverteilt	Population nicht normalverteilt	Bemerkung
1-Sample <i>t</i> -Test (1-Sample <i>Z</i> -Test)	1-Sample Sign Test	allgemeine Verteilung R: sign.test
	1-Sample Wilcoxon Test	symmetrische Verteilung
Paired <i>t</i> -Test	Matched Pairs Sign Test	1-Sample Sign Test für Differenzen
	Wilcoxon-Signed Rank Test	1-Sample Wilcoxon für Differenzen
2-Sample <i>t</i> -Test	Mann-Whitney <i>U</i> -test	R: wilcox.test
One-Way ANOVA	Kruskal-Wallis Test	R: kruskal.test

www.matstat.org

One-Sample Sign Test

- Testet ob der **Median** einer Population einem hypothetischen Wert m_0 gleicht.
- Wenn die Nullhypothese stimmt, sollte ungefähr die Hälfte aller Werte einer Stichprobe größer als der hypothetische Median m_0 sein, während die andere Hälfte kleiner sein sollte.
- **Zufallsvariable X** : Anzahl der Stichprobenwerte, die größer als m_0 sind
- n = Stichprobengröße
- Unter H_0 sollte X **binomialverteilt mit $p = 1/2$** sein: $X \sim \text{Bin}(n, 0.5)$, denn für jeden der n Bernoulli-Versuche sollte die Wahrscheinlichkeit, dass der Wert größer als m_0 ist, 0.5 sein.

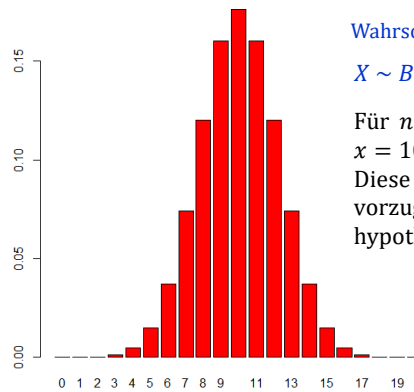
Nullhypothesen und alternative Hypothesen: (x = Observation von X)

	Fall 1	Fall 2	Fall 3
H_0	$m = m_0$	$m < m_0$	$m > m_0$
H_a	$m \neq m_0$	$m > m_0$	$m < m_0$
p -Wert	$p = 2 \cdot P(X \leq \min(x, n - x))$	$p = P(X \geq x)$	$p = P(X \leq x)$

www.matstat.org

One-Sample Sign Test

- Zufallsvariable X : Anzahl der Stichprobenwerte, die größer als m_0 sind
- Unter H_0 sollte X binomialverteilt mit $p = 1/2$ sein: $X \sim \text{Bin}(n, 0.5)$, denn für jeden der n Bernoulli-Versuche sollte die Wahrscheinlichkeit, dass der Wert größer als m_0 ist, 0.5 sein.



Wahrscheinlichkeitsfunktion

$$X \sim \text{Bin}(20, 0.5)$$

Für $n = 20$ ist die Observation $x = 10$ am wahrscheinlichsten. Diese Observation erwarten wir vorzugsweise, wenn die Nullhypothese stimmt.

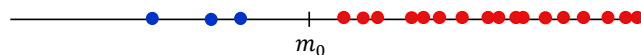
www.matstat.org

One-Sample Sign Test

p-Wert, Fall 2 (siehe Tabelle oben): H_0 wird verworfen wenn x groß ist.

$$H_a : m > m_0 \quad \Rightarrow \quad p = P(X \geq x) \quad \boxed{X = \text{Anzahl der Werte} > m_0}$$

Beispiel: Wir haben 20 Messwerte ($n = 20$), wovon 17 größer sind als der hypothetische Median, also $x = 17$.



Es scheint unwahrscheinlich, dass der hypothetische Median m_0 stimmt.

$$p = P(X \geq x) = P(X \geq 17) \quad \text{mit } X \sim \text{Bin}(20, 0.5) \quad (\text{"unter } H_0\text{"})$$

$$p = P(X \geq 17) = 1 - P(X \leq 16) = 1 - F_X(16) = 1 - 0.9987 = \underline{\underline{0.0013}}$$

Ergebnis: Der p -Wert ist kleiner als $\alpha = 0.05$, wir verwerfen daher H_0 , d. h. wir verwerfen die Annahme $p = 0.5$ und damit die Annahme, dass m_0 der Median der zugrunde liegenden Verteilung ist. Wir akzeptieren dagegen die alternative Hypothese $m > m_0$ (der wahre Wert des Medians muss größer sein als durch die Hypothese behauptet).

www.matstat.org

Verteilungsfunktion für $Bin(n, p)$

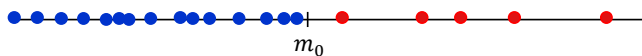
n	$k \setminus p$	0.05	0.1	0.15	0.2	0.25	0.3	0.4	0.5
20	0	0.3585	0.1216	0.0388	0.0115	0.0032	0.0008	0.0000	0.0000
	1	0.7358	0.3917	0.1756	0.0692	0.0243	0.0076	0.0005	0.0000
	2	0.9245	0.6769	0.4049	0.2061	0.0913	0.0355	0.0036	0.0002
	3	0.9841	0.8670	0.6477	0.4114	0.2252	0.1071	0.0160	0.0013
	4	0.9974	0.9568	0.8298	0.6296	0.4148	0.2375	0.0510	0.0059
	5	0.9997	0.9887	0.9327	0.8042	0.6172	0.4164	0.1256	0.0207
	6	1.0000	0.9976	0.9781	0.9133	0.7858	0.6080	0.2500	0.0577
	7	1.0000	0.9996	0.9941	0.9679	0.8982	0.7723	0.4159	0.1316
	8	1.0000	0.9999	0.9987	0.9900	0.9591	0.8867	0.5956	0.2517
	9	1.0000	1.0000	0.9998	0.9974	0.9861	0.9520	0.7553	0.4119
	10	1.0000	1.0000	1.0000	0.9994	0.9961	0.9829	0.8725	0.5881
	11	1.0000	1.0000	1.0000	0.9999	0.9991	0.9949	0.9435	0.7483
	12	1.0000	1.0000	1.0000	1.0000	0.9998	0.9987	0.9790	0.8684
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9935	0.9423
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9984	0.9793
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9941
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987	
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	
18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	

One-Sample Sign Test

p-Wert, Fall 3 (siehe Tabelle oben): H_0 wird verworfen wenn x klein ist.

$$H_a : m < m_0 \quad \Rightarrow \quad p = P(X \leq x) \quad X = \text{Anzahl Werte} > m_0$$

Beispiel: Wir haben 20 Messwerte ($n = 20$), davon 5 größer als der hypothetische Median, also $x = 5$.



Es scheint unwahrscheinlich, dass der hypothetisch Median m_0 korrekt ist.

$$p = P(X \leq x) = P(X \leq 5) \quad \text{mit } X \sim Bin(20, 0.5) \quad (\text{"unter } H_0\text{"})$$

$$p = P(X \leq 5) = F_X(5) = \underline{0.0207}$$

Ergebnis: Der p -Wert ist kleiner als $\alpha = 0.05$, wir verwerfen daher H_0 , d. h. wir verwerfen die Annahme $p = 0.5$ und damit die Annahme, dass m_0 der Median der zugrunde liegenden Verteilung ist. Wir akzeptieren dagegen die alternative Hypothese $m < m_0$ (der wahre Wert des Medians muss kleiner sein als durch die Hypothese behauptet).

Verteilungsfunktion für $Bin(n, p)$

n	$k \setminus p$	0.05	0.1	0.15	0.2	0.25	0.3	0.4	0.5
20	0	0.3585	0.1216	0.0388	0.0115	0.0032	0.0008	0.0000	0.0000
	1	0.7358	0.3917	0.1756	0.0692	0.0243	0.0076	0.0005	0.0000
	2	0.9245	0.6769	0.4049	0.2061	0.0913	0.0355	0.0036	0.0002
	3	0.9841	0.8670	0.6477	0.4114	0.2252	0.1071	0.0160	0.0013
	4	0.9974	0.9568	0.8298	0.6296	0.4148	0.2375	0.0510	0.0059
	5	0.9997	0.9887	0.9327	0.8042	0.6172	0.4164	0.1256	0.0207
	6	1.0000	0.9976	0.9781	0.9133	0.7858	0.6080	0.2500	0.0577
	7	1.0000	0.9996	0.9941	0.9679	0.8982	0.7723	0.4159	0.1316
	8	1.0000	0.9999	0.9987	0.9900	0.9591	0.8867	0.5956	0.2517
	9	1.0000	1.0000	0.9998	0.9974	0.9861	0.9520	0.7553	0.4119
	10	1.0000	1.0000	1.0000	0.9994	0.9961	0.9829	0.8725	0.5881
	11	1.0000	1.0000	1.0000	0.9999	0.9991	0.9949	0.9435	0.7483
	12	1.0000	1.0000	1.0000	1.0000	0.9998	0.9987	0.9790	0.8684
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9935	0.9423
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9984	0.9793
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9941
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

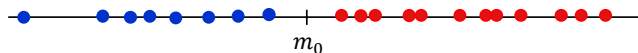
One-Sample Sign Test

p-Wert, Fall 1: H_0 wird verworfen wenn x zu klein oder zu groß ist (2-seitiger Test).

$$H_a : m \neq m_0 \Rightarrow p = P(X \geq \max(x, n - x)) + P(X \leq \min(x, n - x)) \\ = 2 \cdot P(X \leq \min(x, n - x)) \quad (\text{da } p = 0.5 \rightarrow \text{symmetr. Verteilung})$$

Beispiel 1: Wir haben 20 Messwerte ($n = 20$), davon 12 größer als der hypothetische Median, also $x = 12$.

$$X = \text{Anzahl Werte} > m_0$$



$$p = P(X \geq \max(x, n - x)) + P(X \leq \min(x, n - x)) \quad \text{mit } X \sim Bin(20, 0.5) \\ p = P(X \geq 12) + P(X \leq 8) = 1 - P(X \leq 11) + P(X \leq 8) \\ = 1 - F_X(11) + F_X(8) = 1 - 0.7483 + 0.2517 = \underline{\underline{0.5034}}$$

Ergebnis: Der p -Wert ist größer als $\alpha = 0.05$, wir verwerfen die Nullhypothese $p = 0.5$ **nicht** und damit auch nicht die Annahme, dass m_0 der Median der zugrunde liegenden Verteilung ist. Es ist also durchaus möglich, dass die Nullhypothese zutrifft.

www.matstat.org

Verteilungsfunktion für $Bin(n, p)$

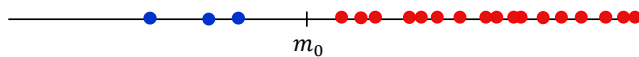
n	$k \setminus p$	0.05	0.1	0.15	0.2	0.25	0.3	0.4	0.5
20	0	0.3585	0.1216	0.0388	0.0115	0.0032	0.0008	0.0000	0.0000
	1	0.7358	0.3917	0.1756	0.0692	0.0243	0.0076	0.0005	0.0000
	2	0.9245	0.6769	0.4049	0.2061	0.0913	0.0355	0.0036	0.0002
	3	0.9841	0.8670	0.6477	0.4114	0.2252	0.1071	0.0160	0.0013
	4	0.9974	0.9568	0.8298	0.6296	0.4148	0.2375	0.0510	0.0059
	5	0.9997	0.9887	0.9327	0.8042	0.6172	0.4164	0.1256	0.0207
	6	1.0000	0.9976	0.9781	0.9133	0.7858	0.6080	0.2500	0.0577
	7	1.0000	0.9996	0.9941	0.9679	0.8982	0.7723	0.4159	0.1316
	8	1.0000	0.9999	0.9987	0.9900	0.9591	0.8867	0.5956	0.2517
	9	1.0000	1.0000	0.9998	0.9974	0.9861	0.9520	0.7553	0.4119
	10	1.0000	1.0000	1.0000	0.9994	0.9961	0.9829	0.8725	0.5881
	11	1.0000	1.0000	1.0000	0.9999	0.9991	0.9949	0.9435	0.7483
	12	1.0000	1.0000	1.0000	1.0000	0.9998	0.9987	0.9790	0.8684
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9935	0.9423
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9984	0.9793
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9941
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

One-Sample Sign Test

p-Wert, Fall 1: H_0 wird verworfen wenn x zu klein oder zu groß ist (2-seitiger Test).

$$H_a : m \neq m_0 \Rightarrow p = P(X \geq \max(x, n - x)) + P(X \leq \min(x, n - x)) \\ = 2 \cdot P(X \leq \min(x, n - x)) \quad (\text{da } p = 0.5 \rightarrow \text{symmetr. Verteilung})$$

Beispiel 2: Wir haben 20 Messwerte ($n = 20$), davon 17 größer als der hypothetische Median, also $x = 17$.



$$p = P(X \geq \max(x, n - x)) + P(X \leq \min(x, n - x)) \quad \text{mit } X \sim Bin(20, 0.5)$$

$$p = P(X \geq 17) + P(X \leq 3) = 1 - P(X \leq 16) + P(X \leq 3) \\ = 1 - F_X(16) + F_X(3) = 1 - 0.9987 + 0.0013 = \underline{\underline{0.0026}}$$

Ergebnis: Der p-Wert ist kleiner als $\alpha = 0.05$, wir verwerfen daher H_0 , d. h. wir verwerfen die Annahme $p = 0.5$ und damit die Annahme, dass m_0 der Median der zugrunde liegenden Verteilung ist. Wir akzeptieren dagegen die alternative Hypothese $m \neq m_0$ (der wahre Wert des Medians muss sich von dem hypothetischen Wert unterscheiden)

Verteilungsfunktion für $Bin(n, p)$

n	$k \setminus p$	0.05	0.1	0.15	0.2	0.25	0.3	0.4	0.5
20	0	0.3585	0.1216	0.0388	0.0115	0.0032	0.0008	0.0000	0.0000
	1	0.7358	0.3917	0.1756	0.0692	0.0243	0.0076	0.0005	0.0000
	2	0.9245	0.6769	0.4049	0.2061	0.0913	0.0355	0.0036	0.0002
	3	0.9841	0.8670	0.6477	0.4114	0.2252	0.1071	0.0160	0.0013
	4	0.9974	0.9568	0.8298	0.6296	0.4148	0.2375	0.0510	0.0059
	5	0.9997	0.9887	0.9327	0.8042	0.6172	0.4164	0.1256	0.0207
	6	1.0000	0.9976	0.9781	0.9133	0.7858	0.6080	0.2500	0.0577
	7	1.0000	0.9996	0.9941	0.9679	0.8982	0.7723	0.4159	0.1316
	8	1.0000	0.9999	0.9987	0.9900	0.9591	0.8867	0.5956	0.2517
	9	1.0000	1.0000	0.9998	0.9974	0.9861	0.9520	0.7553	0.4119
	10	1.0000	1.0000	1.0000	0.9994	0.9961	0.9829	0.8725	0.5881
	11	1.0000	1.0000	1.0000	0.9999	0.9991	0.9949	0.9435	0.7483
	12	1.0000	1.0000	1.0000	1.0000	0.9998	0.9987	0.9790	0.8684
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9935	0.9423
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9984	0.9793
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9941
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

One-Sample Sign Test

- **Frage:** Wollen Mäuse eine eigenen Spiegel?
- 16 Mäuse mit eigenem Käfig haben jeweils einen Raum **mit** Spiegel und einen Raum **ohne** Spiegel
- **Messung:** Wo hielten sich die Mäuse meist auf?:



Maus	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
mit		x					x				x					
ohne	x		x	x	x	x		x	x	x		x	x	x	x	x

Sherwin, C.M. 2004. Mirrors as potential environmental enrichment for individually housed laboratory mice. *Appl. Anim. Behav. Sci.* 87: 95-103.

$H_0: p = 0.5$; $H_a: p \neq 0.5$ (zweiseitiger Test) $x = 3$ (Observation)

$X \sim Bin(16, 0.5)$ ("unter H_0 ")

$$p = P(X \leq 3) + P(X \geq 13) = P(X \leq 3) + 1 - P(X \leq 12) = F_X(3) + 1 - F_X(12) \\ = 0.0106 + 1 - 0.9894 = \underline{0.0212}$$

Fazit: Die Nullhypothese ($p = 0.5$) wird verworfen. Es ist den Mäusen also nicht egal, wo sie sich aufhalten. Aus den Messwerten ist ersichtlich, dass die Mäuse es vorziehen, **keinen** Spiegel zu haben.

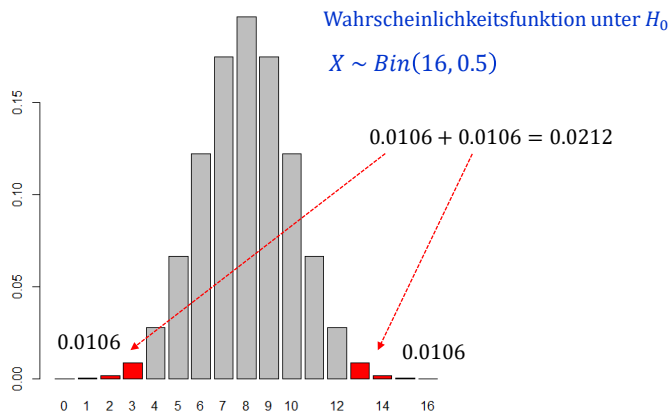
www.matstat.org

Verteilungsfunktion für $Bin(n, p)$

n	$k \setminus p$	0.05	0.1	0.15	0.2	0.25	0.3	0.4	0.5
16	0	0.4401	0.1853	0.0743	0.0281	0.0100	0.0033	0.0003	0.0000
	1	0.8108	0.5147	0.2839	0.1407	0.0635	0.0261	0.0033	0.0003
	2	0.9571	0.7892	0.5614	0.3518	0.1971	0.0994	0.0183	0.0021
	3	0.9930	0.9316	0.7899	0.5981	0.4050	0.2459	0.0651	0.0106
	4	0.9991	0.9830	0.9209	0.7982	0.6302	0.4499	0.1666	0.0384
	5	0.9999	0.9967	0.9765	0.9183	0.8103	0.6598	0.3288	0.1051
	6	1.0000	0.9995	0.9944	0.9733	0.9204	0.8247	0.5272	0.2272
	7	1.0000	0.9999	0.9989	0.9930	0.9729	0.9256	0.7161	0.4018
	8	1.0000	1.0000	0.9998	0.9985	0.9925	0.9743	0.8577	0.5982
	9	1.0000	1.0000	1.0000	0.9998	0.9984	0.9929	0.9417	0.7728
	10	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9809	0.8949
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9951	0.9616
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9991	0.9894
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9979
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

One-Sample Sign Test

Der p -Wert ist die Wahrscheinlichkeit für mindestens so extreme Testresultate wie das aktuell beobachtete, berechnet unter der Annahme dass die Nullhypothese wahr ist.



Observation $x = 3$: unter H_0 ist es sehr unwahrscheinlich, dass eine solche Observation zustande kommt. H_0 wird daher verworfen.

One-Sample Sign Test

Große Stichprobe ($n > 25$)

$$X \sim \text{Bin}(n, p)$$

$$X \sim N\left(n \cdot p, \sqrt{n \cdot p \cdot (1-p)}\right) \quad \text{wenn } V(X) = n \cdot p \cdot (1-p) > 5$$

$$X \sim \text{Bin}(n, 0.5) \quad \text{unter } H_0 \quad (\text{siehe Vorlesung F7})$$

$$X \sim N\left(n/2, 1/2\sqrt{n}\right) \quad \text{Normalapproximation, unter } H_0$$

$$Z = \frac{X - n/2}{1/2\sqrt{n}} \sim N(0, 1) \quad \text{Wenn die Nullhypothese wahr ist, sollte } Z \text{ standard-normalverteilt sein}$$

Z ist standard-normalverteilt unter H_0 . Eine Observation von Z sollte also nicht zu weit vom Zentrum der Dichtefunktion für $N(0, 1)$ liegen. Wenn die Observation dahingegen sehr weit in der Flanke der Dichtefunktion liegt, verwerfen wir die Nullhypothese, denn die Observation ist unwahrscheinlich unter H_0 . Die **kritischen Regionen** sind also (vgl. Vorlesung F12):

$$\Omega_{krit} = \{|z| \geq \lambda_{\alpha/2}\} \quad \text{zweiseitiger Test } (H_a: p \neq 0.5)$$

$$\Omega_{krit} = \{z \geq \lambda_{\alpha}\} \quad \text{einseitiger Test mit } H_a: p > 0.5$$

$$\Omega_{krit} = \{z \leq -\lambda_{\alpha}\} \quad \text{einseitiger Test mit } H_a: p < 0.5$$

www.matstat.org

One-Sample Wilcoxon Test

- testet ob der **Median** für eine Population einem hypothetischen Wert m_0 gleicht (wie Sign-Test)
- Unterschied zum Sign-Test: **die Verteilung der zugrunde liegenden Population muss symmetrisch sein.**

Beispiel: siehe Stichprobe $\{x_i\}$ in der Tabelle:

Frage: Ist der Median gleich Null? ($m_0 = 0$?)

Teststatistik:

- man **rangordnet** die $|x_i|$ ($R_i = \text{Rang}$, Tabelle)
 - (Rang 1 = kleinster Wert)
- wenn $m_0 = 0$ getestet wird, berechnet man:
 - die Rangsumme der positiven x_i (**rot** in Tabelle)
 - die Rangsumme der negativen x_i (**blau** in Tabelle)

$$W^+ = \sum_i R_i^+ \quad W^- = \sum_i R_i^-$$

Rangsumme für positive x_i Rangsumme für negative x_i

$$W^+ = 3 + 8 + 5 = 16$$

$$W^- = 6 + 9 + 4 + 2 + 10 + 11 + 12 + 7 + 1 + 13 = 75$$

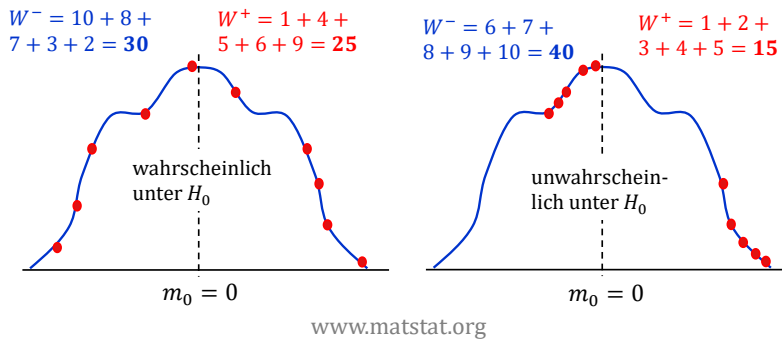
x_i	$ x_i $	R_i
-3,1	3,1	6
-6,3	6,3	9
1,2	1,2	3
-2,0	2,0	4
-1,0	1,0	2
-7,2	7,2	10
5,6	5,6	8
2,2	2,2	5
-12,0	12,0	11
-12,3	12,3	12
-5,3	5,3	7
-0,1	0,1	1
-23,4	23,4	13

www.matstat.org

One-Sample Wilcoxon Test

Wenn der Median einer **symmetrischen** Verteilung 0 ist

- ... sollten W^+ und W^- ungefähr gleich groß sein
(siehe Bild links: $W^- = 25$; $W^+ = 30$)
- ... ist es demgegenüber unwahrscheinlich, dass z. B. alle negativen Observierungen große Absolutwerte haben und alle positiven Observierungen kleine Absolutwerte haben
(siehe Bild rechts: $W^- = 15$; $W^+ = 40$)

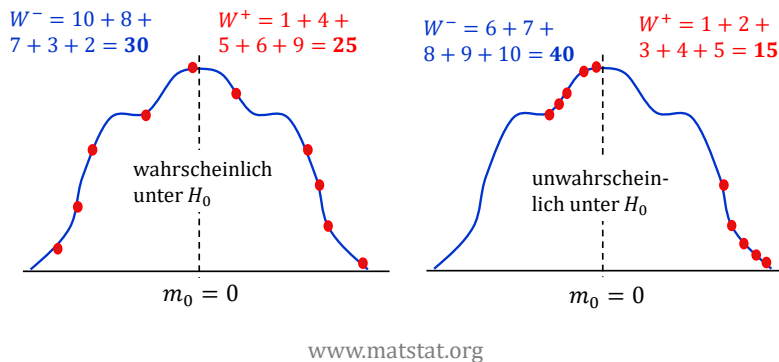


1-S Wilcoxon Test

- H_0 wird verworfen wenn der Unterschied zwischen W^+ und W^- zu groß wird.
- Dies ist gleichbedeutend damit, dass eine der beiden Rangsummen klein ist.

⇒ Teststatistik: $W = \min(W^+, W^-)$

Kritische Region: $\Omega_{krit} = \{W < W_0\}$ H_0 wird für kleine Werte der Teststatistik verworfen



One-Sample Wilcoxon Test

Tabelle der W_0 (kritische Werte)

Critical values		Wilcoxon					
One-sided	Two-sided	$n = 5$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$n = 10$
$P = .05$	$P = .10$	1	2	4	6	8	11
$P = .025$	$P = .05$		1	2	4	6	8
$P = .01$	$P = .02$			0	2	3	5
$P = .005$	$P = .01$				0	2	3
One-sided	Two-sided	$n = 11$	$n = 12$	$n = 13$	$n = 14$	$n = 15$	$n = 16$
$P = .05$	$P = .10$	14	17	21	26	30	36
$P = .025$	$P = .05$	11	14	17	21	25	30
$P = .01$	$P = .02$	7	10	13	16	20	24
$P = .005$	$P = .01$	5	7	10	13	16	19
One-sided	Two-sided	$n = 17$	$n = 18$	$n = 19$	$n = 20$	$n = 21$	$n = 22$
$P = .05$	$P = .10$	41	47	54	60	68	76
$P = .025$	$P = .05$	35	40	46	52	59	66
$P = .01$	$P = .02$	28	33	38	43	50	56
$P = .005$	$P = .01$	23	28	32	37	43	49
One-sided	Two-sided	$n = 23$	$n = 24$	$n = 25$	$n = 26$	$n = 27$	$n = 28$
$P = .05$	$P = .10$	83	92	101	110	120	130
$P = .025$	$P = .05$	73	81	90	98	107	117
$P = .01$	$P = .02$	62	69	77	85	93	102
$P = .005$	$P = .01$	55	68	76	84	92	100

wird mit Hilfe von Kombinatorik berechnet

One-Sample Wilcoxon Test

Große Stichprobe ($n > 25$)

$$Z = \frac{W^+ - \mu}{\sigma} \sim N(0, 1) \quad \text{standard-normal}$$

$$\mu = \frac{1}{2} \cdot \frac{n \cdot (n + 1)}{2} \quad \text{Erwartungswert für Teststatistik } W^+ \text{ (unter } H_0: \text{ Hälfte der totalen Rangsumme)}$$

$$\sigma^2 = \frac{n \cdot (n + 1) \cdot (2n + 1)}{24} \quad \text{Varianz für Teststatistik } W^+$$

$$\Omega_{krit} = \left\{ |z| > \lambda_{\alpha/2} \right\} \quad \left(\text{für zweiseitigen Test} \right)$$

"größer als" - denn wir testen W^+

Wilcoxon-Signed Rank Test für gepaarte Stichproben

- zwei gepaarte Stichproben (z. B. "vorher/nachher")
- gibt es eine systematische Verschiebung ("location shift") zwischen beiden Verteilungen oder sind sie identisch (H_0) ?
- Beispiel: Aluminiumgehalt von Bäumen in einen kontaminierten Areal im August und im November →

Laureysens, I., R. Blust, L. De Temmerman, C. Lemmens and R. Ceulemans. 2004. Clonal variation in heavy metal accumulation and biomass production in a poplar coppice culture. I. Seasonal variation in leaf, wood and bark concentrations. Environ. Pollution 131: 485-494.



www.matstat.org

Wilcoxon-Signed Rank Test für gepaarte Stichproben

- man rangordnet die absoluten Differenzen ($|\Delta_i|$) → R_i (Rang 1 = kleinster Wert)
- berechne wiederum W^+ und W^- :

$$W^+ = \sum_i R_i^+ \quad W^- = \sum_i R_i^-$$

$$W^+ = 3 + 8 + 5 = 16$$

$$W^- = 6 + 9 + 4 + 2 + 10 + 11 + 12 + 7 + 1 + 13 = 75$$



www.matstat.org

Aug	Nov	Δ_i	$ \Delta_i $	R_i
8,1	11,2	-3,1	3,1	6
10,0	16,3	-6,3	6,3	9
16,5	15,3	1,2	1,2	3
13,6	15,6	-2,0	2,0	4
9,5	10,5	-1,0	1,0	2
8,3	15,5	-7,2	7,2	10
18,3	12,7	5,6	5,6	8
13,3	11,1	2,2	2,2	5
7,9	19,9	-12,0	12,0	11
8,1	20,4	-12,3	12,3	12
8,9	14,2	-5,3	5,3	7
12,6	12,7	-0,1	0,1	1
13,4	36,8	-23,4	23,4	13

Wilcoxon-Signed Rank Test für gepaarte Stichproben

- Wenn beide Verteilungen identisch wären (H_0) sollte ungefähr die Hälfte der paarweisen Differenzen positiv sein, während die andere Hälfte negativ sein sollte.
- Außerdem sollten positive und negative Differenzen mit demselben Absolutwert gleich wahrscheinlich sein.
- Unter H_0 erwarten wir also, dass die Rangsumme der positiven Differenzen ungefähr so groß ist wie die Rangsumme der negativen Differenzen.
- Wir führen also einen 1-S Wilcoxon Test für die paarweisen Differenzen durch:

$$W^+ = \sum_i R_i^+ \quad W^- = \sum_i R_i^-$$

Rangsumme für positive Δ_i Rangsumme für negative Δ_i

$$W^+ = 3 + 8 + 5 = 16$$

$$W^- = 6 + 9 + 4 + 2 + 10 + 11 + 12 + 7 + 1 + 13 = 75$$

} siehe Tabelle oben

Teststatistik: $W = \min(W^+, W^-)$ H_0 wird für kleine Werte der Teststatistik verworfen

Kritische Region: $\Omega_{krit} = \{W < W_0\}$

www.matstat.org

Mann-Whitney U-Test

- H_0 : zwei unabhängige (nicht-gepaarte) Stichproben kommen von identisch verteilten Populationen
- H_a : es besteht eine Verschiebung zwischen den Verteilungen ("location shift")
- entspricht dem Studentschen t -Test für nicht-normalverteilte Populationen

Beispiel: zwei Stichproben A und B:

- $x_A = \{25, 26, 27, 31\}$ und $x_B = \{28, 29, 32, 35\}$ (für die Praxis viel zu klein)

Teststatistik (U):

- ordne alle Werte der Größe nach (kleinster Wert zuerst)
- zähle: wie viele Werte von A kommen vor jedem B-Wert? $\rightarrow U_B$
- zähle: wie viele Werte von B kommen vor jedem A-Wert? $\rightarrow U_A$

Wert, Rangordnung	25	26	27	28	29	31	32	35
Von Stichprobe:	A	A	A	B	B	A	B	B
Beitrag zu U_B				3	3		4	4

3 Werte von Stichprobe A kommen vor diesen B

4 Werte von Stichprobe A kommen vor diesen B

www.matstat.org

Teststatistik, Mann-Whitney U -test

- zähle: wie viele Werte von A kommen vor jedem B-Wert? $\rightarrow U_B$
- zähle: wie viele Werte von B kommen vor jedem A-Wert? $\rightarrow U_A$

Wert, Rangordnung	25	26	27	28	29	31	32	35
Von Stichprobe:	A	A	A	B	B	A	B	B
Beitrag zu U_B				3	3		4	4

3 Werte von Stichprobe A
kommen vor diesen B

4 Werte von Stichprobe A
kommen vor diesen B

$$U_B = 3 + 3 + 4 + 4 = 14$$

Wert, Rangordnung	25	26	27	28	29	31	32	35
Von Stichprobe:	A	A	A	B	B	A	B	B
Beitrag zu U_A	0	0	0			2		

0 Werte von Stichprobe B
kommen vor diesen A

2 Werte von Stichprobe B
kommen vor diesem A

$$U_A = 0 + 0 + 0 + 2 = 2$$

Mann-Whitney U -test

Teststatistik (U):

$$U = \min(U_A, U_B)$$

Kritische Region:

- sehr kleine Werte der Teststatistik U deuten auf einen Unterschied zwischen den Verteilungen der Populationen A und B hin
- die kritischen Werte (U_0) sind tabelliert

$$\Omega_{krit} = \{U < U_0\}$$

Voraussetzungen; Trennschärfe (Power):

- unabhängige zufällige Stichproben
- (etwas) kleinere Power im Vergleich zum Studentischen t -Test wenn die Populationen normalverteilt sind
- größere Power als Studentischer t -Test für viele andere Verteilungen

Mann-Whitney U -test

U hängt mit W (Wilcoxon) zusammen: n_1, n_2 Stichprobengrößen

Wert	25	26	27	28	29	31	32	35
Stichprobe	A	A	A	B	B	A	B	B
Beitrag zu U_B				3	3		4	4
Beitrag zu U_A	0	0	0			2		
Rang (Wilcoxon)	1	2	3	4	5	6	7	8

$$\left. \begin{aligned} W_A &= 1 + 2 + 3 + 6 = 12 \\ W_B &= 4 + 5 + 7 + 8 = 24 \end{aligned} \right\} \text{Wilcoxon's Rangsumme}$$

$$\left. \begin{aligned} U_A &= n_1 \cdot n_2 + \frac{n_1 \cdot (n_1 + 1)}{2} - W_A = 16 + \frac{20}{2} - 12 = 14 \\ U_B &= n_1 \cdot n_2 + \frac{n_2 \cdot (n_2 + 1)}{2} - W_B = 16 + \frac{20}{2} - 24 = 2 \end{aligned} \right\} \text{Zusammenhang zwischen } U \text{ und } W$$

$$U_A + U_B = n_1 \cdot n_2 = 16$$

Diese Formeln werden oft in der Praxis verwendet (Berechnung von W einfacher).

www.matstat.org

Mann-Whitney U -test

Große Stichprobe ($n > 25$)

$$Z = \frac{U - \mu_U}{\sigma_U} \sim N(0, 1) \quad \dots \text{ ist standard-normal verteilt}$$

$$\mu_U = \frac{n_1 \cdot n_2}{2} \quad \text{Erwartungswert für Teststatistik } U$$

$$\sigma_U = \sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}} \quad \text{Standardabweichung für Teststatistik } U$$

$$\Omega_{krit} = \left\{ |z| > \lambda_{\alpha/2} \right\} \quad \text{kritische Region (für zweiseitigen Test)}$$

Alternative Namen für Mann-Whitney U -Test:

- Mann-Whitney-Wilcoxon Test
- Wilcoxon rank-sum Test
- Wilcoxon-Mann-Whitney Test

www.matstat.org

Kruskal-Wallis Test

- wird an Stelle von One-Way ANOVA verwendet, wenn die den Stichproben zugrunde liegenden Populationen nicht normalverteilt sind
- k : Anzahl der Populationen ("Gruppen", "treatments")
- H_0 : alle k Populationen haben die gleiche Verteilung
- H_a : mindestens eine Population hat eine verschobene Verteilung ("location shift")

Teststatistik:

- alle $n = n_1 + n_2 + \dots + n_k$ Observationen werden dem Rang nach geordnet
 - n_i = Anzahl der Stichproben-Werte für Gruppe i

R_i : Summe aller Ränge für Stichprobe i

$$\bar{R}_i = \frac{R_i}{n_i} \quad \text{Mittelwert aller Ränge für Stichprobe } i$$

$$\bar{R} = \frac{1}{n} \cdot \frac{n \cdot (n+1)}{2} = \frac{n+1}{2} \quad \text{Mittelwert aller Ränge (Mittelwert der ersten } n \text{ natürlichen Zahlen)}$$

$$\frac{n \cdot (n+1)}{2} = \text{Summe aller Ränge} = \text{Summe der ersten } n \text{ natürlichen Zahlen}$$

www.matstat.org

Kruskal-Wallis Test

Definition:
$$V = \sum_{i=1}^k n_i \cdot (\bar{R}_i - \bar{R})^2$$

V entspricht der Quadratsumme SS_{tr} bei der ANOVA (siehe Vorlesung **F13**):

- wenn H_0 wahr ist sollten alle \bar{R}_i (und \bar{R}) ungefähr gleich sein
- $\rightarrow V$ muss unter H_0 klein sein

Teststatistik:
$$H = \frac{12 \cdot V}{n \cdot (n+1)} = \frac{12}{n \cdot (n+1)} \cdot \sum_{i=1}^k \frac{R_i^2}{n_i} - 3 \cdot (n+1)$$

(klein unter H_0)

Kritische Region:

- Wenn H_0 gilt und außerdem alle n_i ausreichend groß sind ($n_i \geq 5$), ist die Teststatistik H ungefähr χ^2 -verteilt.
- H_0 wird verworfen wenn H (also auch V) groß ist – dies bedeutet ja, dass die Gruppennittelwerte \bar{R}_i viel vom gemeinsamen Mittelwert \bar{R} abweichen.

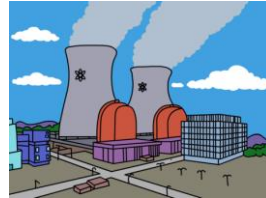
$$\Omega_{krit} = \{ H > \chi_{\alpha}^2(f) \} \quad f = k - 1$$

www.matstat.org

Kruskal-Wallis Test

Experiment: Wir wollen wissen, ob die Temperaturerhöhung des Meerwassers in der Nähe eines Kernkraftwerkes einen Einfluss auf das Gewicht der Fische hat:

Gewicht der Fische			
38°F	42°F	46°F	50°F
22	15	14	17
24	21	28	18
16	26	21	13
18	16	19	20
19	25	24	21
	17	23	



- 4 Gruppen
- $n_i = (5, 6, 6, 5)$
- $n = 22$

www.matstat.org

Kruskal-Wallis Test



kruskal_wallis.R (siehe Webseite)

```
x_38 = c(22, 24, 16, 18, 19)
x_42 = c(15, 21, 26, 16, 25, 17)
x_46 = c(14, 28, 21, 19, 24, 23)
x_50 = c(17, 18, 13, 20, 21)
weight = c(x_38, x_42, x_46, x_50)
temp = c(rep(38, length(x_38)), rep(42, length(x_42)),
         rep(46, length(x_46)), rep(50, length(x_50)))
fish = data.frame(weight = weight, temp = temp)
boxplot(weight ~ temp, data = fish)
kruskal.test(weight ~ temp, data = fish) # Kruskal-Wallis rank sum Test
# Kruskal-Wallis chi-squared = 2.0404, df = 3, p-value = 0.5641
```



www.matstat.org