# Quantitative Methoden

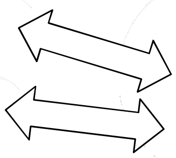
Inferenzstatistik – alea iacta est

11.04.2008

Prof. Dr. Walter Hussy und David Tobinski
UDE.EDUcation College
im Rahmen des dokFORUMs
Universität Duisburg-Essen

## Inferenzstatistik
Erläuterung

„Inferenzstatistik“ bedeutet übersetzt „schließende Statistik“. Damit ist der Schluss von den erhobenen Daten einer Stichprobe auf Werte in der Population gemeint.

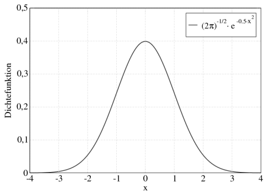


## Die Normalverteilung
Einführung

- In der Natur sind sehr viele Merkmale normalverteilt. Dies gilt beispielsweise für die Körpergröße, Intelligenz oder das Sehvermögen.
- Die Normalverteilung galt sogar zeitweise als eine Art Naturgesetz. Auch wenn diese Vorstellung mittlerweile abgelehnt wird, so zeigt sie doch die herausragende Bedeutung der Normalverteilung in der Mathematik und in den empirischen Sozialwissenschaften.
- Sehr viele statistische Verfahren, die für wissenschaftliches Arbeiten erforderlich sind, setzen voraus, dass die relevanten Merkmale normalverteilt sind.

## Die Normalverteilung
Darstellung der Funktion

Zu sehen ist eine **unimodale** und **eingipflige** Verteilung mit **glockenförmigem** Verlauf. Sie ist **symmetrisch** und nähert sich der x-Achse **asymptotisch** an. Dadurch sind die Werte für **Median**, **Modus** und **arithmetisches Mittel** identisch. Eine solche Verteilung heißt **Normalverteilung**.



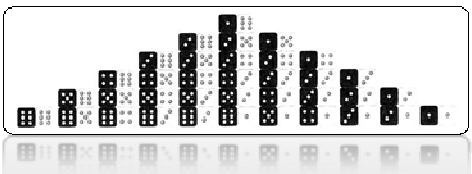

$$f(x) = \frac{1}{\sqrt{2\pi \cdot \sigma^2}} \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

## Die Normalverteilung
Wahrscheinlichkeiten unter der Normalverteilung

Aus der Symmetrie der Normalverteilung folgt, dass die Wahrscheinlichkeit, einen größeren Wert als den Mittelwert zu messen, bei p : 0,5 liegt.

Da eine Kurve aus unendlich vielen Punkten besteht, ist die Wahrscheinlichkeit für einen einzelnen Wert unendlich klein. Deshalb sind Wahrscheinlichkeitsberechnungen nur für Flächen, also Intervalle unter der Kurve möglich, niemals für einzelne Werte.
Für Normalverteilungen gilt, dass die Fläche, die von +/- einer Standardabweichung vom Mittelwert begrenzt wird, mehr als 2/3 aller (68,26%) beinhaltet. 95,44% liegen im Bereich von +/- zwei Standardabweichungen.

## Die Normalverteilung
Wahrscheinlichkeiten unter der Normalverteilung



## Mittelwertsvergleiche
Der t-Test (Verfahren für intervallskalierte Daten)

Der t-Test untersucht, ob sich zwei empirisch gefundene Mittelwerte systematisch voneinander unterscheiden. Mit Hilfe dieses Testverfahrens ist es möglich festzustellen, ob zwei betrachtete Gruppen in einem untersuchten Merkmal wirklich einen Unterschied aufweisen oder ob ein gefundener Mittelwertsunterschied rein zufällig entstanden ist.

## Der t-Test
Voraussetzungen für die Anwendung eines t-Tests

Voraussetzungen für die Anwendung eines t-Tests

Für den t-Test gibt es drei mathematische Voraussetzungen:

1. Das untersuchte Merkmal ist intervallskaliert.
2. Das untersuchte Merkmal ist in der Population normalverteilt.
3. Die Populationsvarianzen, aus denen die beiden Stichproben stammen, sind gleich (Varianzhomogenität).

## Der t-Test
Mittelwertsvergleiche

Wegen dieser Voraussetzungen gehört der t-Test zur Gruppe der parametrischen Verfahren.

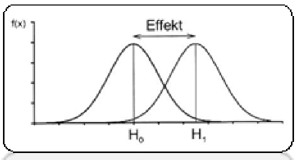
Parametrische Verfahren schätzen Populationsparameter mittels statistischer Kennwerte wie dem arithmetischen Mittel oder der Varianz, für deren Berechnung die Intervallskaliertheit der Daten Voraussetzung ist.

## Der t-Test
Die Fragestellung des t-Tests

**Die** zentrale Frage des t-Tests lautet: *Wie wahrscheinlich ist die empirisch gefundene oder eine größere Mittelwertsdifferenz unter allen möglichen rein theoretisch denkbaren Differenzen?*

Der wichtigste Wert für die Durchführung eines t-Tests ist die Differenz der Gruppenmittelwerte. Diese Differenz bildet den Stichprobenkennwert des t-Tests: $\bar{x}_1 - \bar{x}_2$

## Der t-Test
Die Fragestellung des t-Tests




## Der t-Test
Die Fragestellung des t-Tests

Der t-Test kann jeweils nur zwei Gruppen im Mittelwertsvergleich betrachten.

Inhaltliche Hypothesen müssen in statistische Hypothesen umformuliert werden. $\bar{x}_1 - \bar{x}_2 > 0.$

Die Bildung der Differenz wird entscheidend durch die Formulierung der statistischen Hypothese mitbestimmt: Sie legt fest, welcher Wert vom anderen abgezogen wird.

## Der t-Test
Die Nullhypothese

Für die Erklärung der Mittelwertsdifferenz gibt es neben der Annahme eines systematischen Unterschieds zwischen den beiden Gruppen eine weitere Möglichkeit: Die Differenz zwischen den Mittelwerten ist zufällig zustande gekommen und es gibt keinen echten Unterschied zwischen den beiden untersuchten Gruppen.
Die beiden Gruppen stammen im Grunde aus zwei Populationen mit demselben Mittelwert.
Die Differenz zwischen den Gruppen sollte demzufolge Null betragen.
Diese Annahme heißt deshalb Nullhypothese oder $H_0$.

## Der t-Test
Stichprobenkennwerteverteilung unter der Nullhypothese

Wird aus zwei Populationen mit identischem Populationsmittelwert jeweils eine Stichprobe gezogen, so kann die Differenz der beiden Stichprobenmittelwerte theoretisch jeden beliebigen Wert annehmen. Die zu erwartende Differenz aber ist gleich Null, denn die Stichprobenmittelwerte sind normalverteilt um ihren jeweiligen Erwartungswert, den Populationsmittelwert.

Der t-Test findet in den meisten Fällen als Nullhypothesen-signifikanztest Anwendung.

## Der t-Test
Die Bewertung des t-Werts

| | In Wirklichkeit gilt die $H_0$ | In Wirklichkeit gilt die $H_1$ |
|---|---|---|
| Entscheidung zugunsten der $H_0$ | Richtige Entscheidung | β-Fehler |
| Entscheidung zugunsten der $H_1$ | α-Fehler | Richtige Entscheidung |

## Der t-Test
Die Bewertung des t-Werts

Unter der Annahme der Nullhypothese errechnet der t-Test eine Wahrscheinlichkeit für das Auftreten der gefundenen oder einer größeren Differenz, die z.B. p = 0,03 beträgt.
Der errechnete Wert von 3% (p = 0,03) bedeutet, dass die Wahrscheinlichkeit für das Finden einer solchen oder einer größeren Differenz beim Ziehen von Stichproben aus einer identischen Population sehr gering ist. Natürlich ist diese Differenz möglich, sie ist aber sehr unwahrscheinlich. Wenn die Differenz unter Annahme der Nullhypothese sehr unwahrscheinlich ist, so trifft möglicherweise die Annahme selbst gar nicht zu.

## Der t-Test
Die Bewertung des t-Werts

Test bei unabhängigen Stichproben

| | | Levene-Test der Varianzgleichheit | | T-Test für die Mittelwertgleichheit | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | 95% Konfidenzintervall der Differenz | |
| | | F | Signifikanz | T | df | Sig. (2-seitig) | Mittlere Differenz | Standardfehler der Differenz | Untere | Obere |
| EINS | Varianzen sind gleich | ,858 | ,354 | ,288 | 74 | ,774 | ,17 | ,589 | -1,006 | 1,344 |
| | Varianzen sind nicht gleich | | | ,288 | 73,125 | ,775 | ,17 | ,589 | -1,006 | 1,344 |

Test bei unabhängigen Stichproben

| | | Levene-Test der Varianzgleichheit | | T-Test für die Mittelwertgleichheit | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | 95% Konfidenzintervall der Differenz | |
| | | F | Signifikanz | T | df | Sig. (2-seitig) | Mittlere Differenz | Standardfehler der Differenz | Untere | Obere |
| ZWEI | Varianzen sind gleich | 10,951 | ,001 | 11,957 | 74 | ,000 | 5,14 | ,430 | 4,287 | 6,002 |
| | Varianzen sind nicht gleich | | | 12,411 | 53,568 | ,000 | 5,14 | ,415 | 4,313 | 5,976 |

## Der t-Test
Die Entwicklung eines Entscheidungskriteriums

Das Signifikanzniveau muss vor der Durchführung einer Untersuchung festgelegt und begründet werden, um das Ergebnis beurteilen zu können. Per Konvention liegt es meist bei $\alpha = 0{,}05$, einige Untersuchungen verwenden 1% oder 10%. Ein auf dem 5%-Niveau signifikantes Ergebnis wird in der Literatur und in SPSS in der Regel mit einem Sternchen (*) gekennzeichnet, ein auf dem 1%-Niveau signifikantes Ergebnis mit zwei Sternchen (**).

## Mittelwertsvergleiche
Verfahren für Rangdaten

Durch eine Zuordnung von Rangplätzen wird eine künstliche Äquidistanz zwischen den Werten erzeugt, die viele mathematische Operationen wie z.B. die Mittelwertsbildung erst ermöglicht. Die nichtparametrischen Verfahren für Rangdaten arbeiten nicht mit Populationsparametern und –verteilungen, sondern legen ihre eigene Verteilung (die der Rangplätze) zu Grunde.

## Der Mann-Whitney U-Test
Verfahren für Rangdaten

Der U-Test für unabhängige Stichproben oder auch Mann-Whitney U-Test ist ein Verfahren zur Auswertung eines Zwei-Gruppen-Experiments, dessen Bedingungen sich in einer unabhängigen Variable unterscheiden.
Ähnlich wie der t-Test für unabhängige Stichproben prüft der U-Test, ob die Unterschiede in den zwei Gruppen bezüglich einer abhängigen Variable zufälligen oder systematischen Einflüssen unterliegen. Anders als der t-Test aber analysiert der U-Test die Messwerte nicht direkt, sondern die ihnen zugeordneten Rangplätze. Der U-Test stellt primär ein Instrument für ordinalskalierte Daten dar.

## Der Mann-Whitney U-Test
Verfahren für Rangdaten

Er ist aufgrund seiner größeren Vorraussetzungsfreiheit in folgenden Fällen dem t-Test für unabhängige Stichproben vorzuziehen:
- Die Intervallskalenqualität der abhängigen Variable ist zweifelhaft.
- Das Merkmal folgt in der Population keiner Normalverteilung.
- Die Annahme der Varianzhomogenität ist verletzt, so dass der t-Test für unabhängige Stichproben keine zuverlässigen Ergebnisse mehr liefert.

## Der Mann-Whitney U-Test
### Zuordnung der Rangplätze

Für die Berechnung des U-Werts ist es zuerst notwendig, alle Messwerte in eine gemeinsame Rangreihe zu bringen. Danach wird für jede Gruppe die Summe der Rangplätze sowie der durchschnittliche Rangplatz der Gruppe berechnet.

$$T_i = \sum_{m=1}^{n_i} R_{mi}$$

i : Gruppe

$n_i$ : Anzahl der Versuchspersonen in Gruppe i

$R_{mi}$ : Rangplatz der m-ten Versuchsperson in der Gruppe i

## Der Mann-Whitney U-Test
### Rangplatzüber- und Rangplatzunterschreitungen

Der statistische Kennwert U wird gebildet, indem für jeden Rangplatz einer Person der einen Gruppe die Anzahl an Personen aus der anderen Gruppe gezählt wird, die diesen Rangplatz überschreitet.

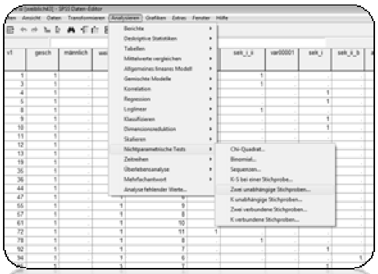
Der statistische Kennwert U` wird gebildet, indem für jeden Rangplatz einer Person der einen Gruppe die Anzahl an Personen aus der anderen Gruppe gezählt wird, die diesen Rangplatz unterschreitet.

## Der Mann-Whitney U-Test
### Rangplatzüber- und Rangplatzunterschreitungen

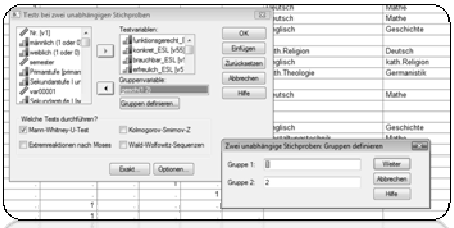
Ein größerer Unterschied zwischen den Gruppen führt in jedem Fall zu einem größeren Unterschied zwischen U und U'. Besteht kein Unterschied zwischen den beiden Gruppen, gibt es genauso viele Rangunter- wie Rangüberschreitungen.

Die Nullhypothese des U-Tests lautet deshalb:
U = U'

## Der Mann-Whitney U-Test
SPSS



## Der Mann-Whitney U-Test
SPSS



## Der Mann-Whitney U-Test
SPSS

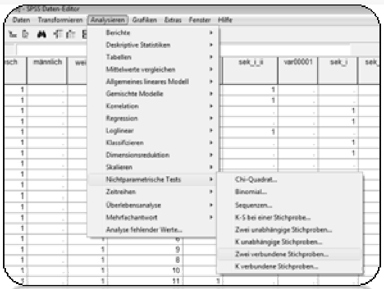
| | funktionsgerecht_ESL | konkret_ESL | brauchbar_ESL | erfreulich_ESL | wichtig_ESL | wissenschaftlich_ESL |
|---|---|---|---|---|---|---|
| Mann-Whitney-U | 724,500 | 584,000 | 737,000 | 628,000 | 558,500 | 836,500 |
| Wilcoxon-W | 1627,500 | 1487,000 | 1640,000 | 1531,000 | 1461,500 | 1697,500 |
| Z | -1,599 | -2,704 | -1,317 | -2,312 | -2,946 | -,230 |
| Asymptotische Signifikanz (2-seitig) | ,110 | ,007 | ,188 | ,021 | ,003 | ,818 |

a. Gruppenvariable: weiblich = 1 männlich = 2

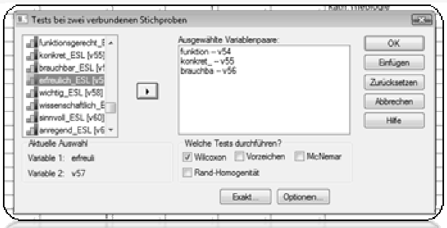
## Der Wilcoxon-Test
### Verfahren für Rangdaten

Der Wilcoxon-Test oder W-Test ist das nichtparametrische Pendant zum t-Test für abhängige Stichproben. Er ist immer in solchen Fällen das nichtparametrische Verfahren der Wahl, wenn die Messwerte zweier Stichproben in irgendeiner Weise voneinander abhängig sind. Der W-Test für abhängige Stichproben arbeitet ebenfalls mit der Analyse einer Rangreihe.

## Der Wilcoxon-Test
### Verfahren für Rangdaten



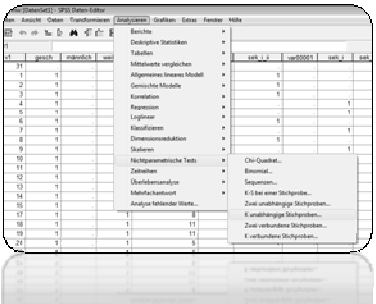

## Der Wilcoxon-Test
### Verfahren für Rangdaten



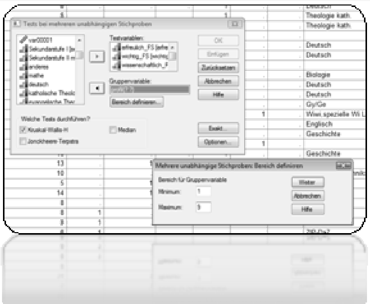
## Der Kruskal-Wallis H-Test
Verfahren für Rangdaten

Der Kruskal-Wallis H-Test ist ein Verfahren für die statistische Auswertung ordinalskalierter Daten von mehr als zwei unabhängigen Gruppen. Er bietet eine Alternative für die einfaktorielle Varianzanalyse ohne Messwiederholung, wenn deren mathematische Vorraussetzungen nicht erfüllt sind. Deshalb heißt er auch Rangvarianzanalyse.

## Der Kruskal-Wallis H-Test
Verfahren für Rangdaten



## Der Kruskal-Wallis H-Test
Verfahren für Rangdaten



## Der Kruskal-Wallis H-Test

Verfahren für Rangdaten

**Statistik für Test[a,b]**

| | konkret_FS | brauchbar_FS | erfreulich_FS | wichtig_FS | wissenschaftlich_FS | sinnvoll_FS | anregend_FS | praxisnah_FS |
|---|---|---|---|---|---|---|---|---|
| Chi-Quadrat | 9,079 | 4,904 | 3,155 | 7,028 | 7,852 | 4,993 | 2,954 | 12,831 |
| df | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| Asymptotische Signifikanz | ,336 | ,768 | ,924 | ,534 | ,448 | ,758 | ,937 | ,118 |

a. Kruskal-Wallis-Test

b. Gruppenvariable: 1=Mathe 2=Deutsch 3=Theologie 4=Gesellschaftswissenschaft 5=Sprache 6=Naturwissenschaft 7=Ing 8=Musik/Kunst 9=Sport