



FAKULTÄT FÜR
WIRTSCHAFTSWISSENSCHAFT

Einführung in die statistische Datenanalyse

Bachelorabschlussseminar
Dipl.-Kfm. Daniel Cracau

Gliederung

1. Grundlagen

2. Nicht-parametrische Tests

- a. Mann-Whitney-Wilcoxon-U Test
- b. Wilcoxon-Signed-Rank Test
- c. Rank-order-correlation
- d. Chi-Square test-of-independence (Unabhängigkeitstest)
- e. Chi-Square-Goodness-of-fit test (Anpassungstest)

3. Regression

- a. Einfache Regression
- b. Multiple Regression

1. Grundlagen

Beschreibende (deskriptive) Statistik

- beschreibt Daten mit Hilfe von Kennzahlen (Mittelwert, Varianz, ...) und Grafiken (Kuchendiagramm, Balkendiagramm,...)

Schließende (induktive) Statistik

- untersucht Daten und leitet mit Hilfe von Hypothesen Eigenschaften ab

Vorgehensweise bei Hypothesentests:

- I. Formulierung der Null- und der Alternativhypothese
- II. Durchführung des Testverfahrens
- III. Annahme oder Ablehnung der Nullhypothese
- IV. Interpretation des Ergebnisses

1. Grundlagen

1. Formulierung der Null- und der Alternativhypothese

- Nullhypothesen begründen einen mathematisch korrekt formulierten Zusammenhang
- Nullhypothesen gehen immer davon aus, dass kein Zusammenhang zwischen Variablen besteht bzw. es keinen Unterschied zwischen Stichproben gibt
- Nullhypothese und Alternativhypothese schließen sich immer gegenseitig aus
- (Nullhypothese und Alternativhypothese füllen den gesamten Möglichkeitenraum aus)*

1. Grundlagen

1. Formulierung der Null- und der Alternativhypothese

- es gibt einseitige und zweiseitige Hypothesentests
 - > zweiseitige Hypothesentests testen ausschließlich auf Ungleichheit
 - > einseitige Hypothesentests testen auf einen Unterschied in eine bestimmte Richtung (größer oder kleiner)
- ***Nullhypothesen sollen immer das beinhalten, was man widerlegen möchte!***

1. Grundlagen

1. Formulierung der Null- und der Alternativhypothese

Beispiel 1

Wir wollen testen, ob Studenten genauso viel Zeit am PC verbringen wie Nicht-Studenten.

Nullhypothese: $H_0: \mu_S = \mu_{NS}$

Alternativhypothese: $H_1: \mu_S \neq \mu_{NS}$

Beispiel 2

Wir wollen testen, ob Frauen im Durchschnitt mehr für Schuhe ausgeben als Männer.

Nullhypothese*: $H_0: \mu_M = \mu_F$

Alternativhypothese: $H_1: \mu_M < \mu_F$

1. Grundlagen

II. Durchführung des Testverfahrens

- passend zur Nullhypothese wird ein statistisches Testverfahren ausgewählt

parametrische Tests

- setzen bestimmte Eigenschaften der Stichprobe voraus
- z. B. normalverteilte Daten, metrisches Skalenniveau

nicht-parametrische (verteilungsfreie) Tests

- setzen weniger Eigenschaften der Stichprobe voraus
- keine Annahmen über Verteilungen
- ordinales Skalenniveau (Rangfolge interpretierbar)

1. Grundlagen

II. Durchführung des Testverfahrens

p-value = „ α -Fehler“ = „Fehler 1. Art“

- ein Hypothesentest liefert einen **p-Wert** (*p-value*)
- gibt an, mit welcher Wahrscheinlichkeit man einen Fehler macht, wenn man die Nullhypothese ablehnt und auf die Alternativhypothese schließt

	Nullhypothese ist wahr	Nullhypothese ist nicht wahr
Nullhypothese wird nicht abgelehnt	Korrekt Schluss	β - Fehler
Nullhypothese wird abgelehnt	α - Fehler	Korrekt Schluss

Tabelle 1: korrektes und unkorrektes Schließen

1. Grundlagen

III. Annahme oder Ablehnung der Nullhypothese

- der p -Wert wird mit einem selbstgewählten Signifikanzniveau α verglichen
 - *typische* Signifikanzniveaus sind 10%, 5%, 1% und 0,1%
 - Signifikanzniveaus werden entsprechend der Stichprobengröße und der Relevanz der Untersuchung festgesetzt
- > wenn $p \leq \alpha \rightarrow$ Nullhypothese verwerfen und auf Alternativhypothese schließen
- > wenn $p > \alpha \rightarrow$ Nullhypothese nicht verwerfen

1. Grundlagen

IV. Interpretation des Ergebnisses

- je nachdem, ob die Nullhypothese abgelehnt wurde oder nicht, können jetzt Aussagen über die getesteten Eigenschaften gemacht werden

Beispiele

- Es gibt einen signifikanten Unterschied zwischen der durchschnittlichen Zeit, die Studenten am PC verbringen und der durchschnittlichen Zeit, die Nicht-Studenten am PC verbringen.
- Frauen geben durchschnittlich nicht signifikant mehr für Schuhe aus als Männer.
- ...

2. Nicht-parametrische Tests

a. Mann-Whitney-Wilcoxon-U Test

Was wird getestet?

- Mittelwerte zweier unabhängiger Stichproben

Testhypothese:

Die Mittelwerte der beiden Stichproben sind identisch. $\rightarrow H_0: \mu_1 = \mu_2$

Beispiele:

- Zahlungsbereitschaft für Schuhe: Mann vs. Frau
- Zeit am PC pro Tag: jung vs. alt
- Bewertung eines Kinofilms: Deutschland vs. USA

2. Nicht-parametrische Tests

b. Wilcoxon-Signed-Rank Test

Was wird getestet?

- Mittelwerte zweier abhängiger Stichproben (paarweise Beobachtungen)

Testhypothese:

Die Mittelwerte der beiden Stichproben sind identisch. $\rightarrow H_0: \mu_1 = \mu_2$

Beispiele:

- Zahlungsbereitschaft Schuhe vs. Hose (1 Person)
- Zeit am PC pro Tag: früher vs. heute (1 Person)
- Bewertung zweier Kinofilme (1 Person)

2. Nicht-parametrische Tests

c. Rank-order correlation

Was wird getestet?

- monotoner Zusammenhang zwischen zwei Variablen
- Rangkorrelationskoeffizient im Intervall $[-1; 1]$

Testhypothese:

Der Rangkorrelationskoeffizient ist null.

→ $H_0: r_s = 0$

Beispiele:

- Zusammenhang Lernaufwand und Endnote
- Zusammenhang PC-Nutzung und Altersgruppe
- Zusammenhang Filmlänge und Kinocharts

2. Nicht-parametrische Tests

d. Chi-Square test of independence (Unabhängigkeitstest)

Was wird getestet?

- stochastische Unabhängigkeit zweier Variablen

Testhypothese:

Die Variablen (X und Y) sind stochastisch unabhängig.

$$\rightarrow H_0: P(X) * P(Y) = P(X \cap Y)$$

Beispiele:

- Fußballfan (ja/nein) und Geschlecht
- Pendler (ja/nein) und Student/Professor
- Altersgruppe und Anzahl der Flirtkontakte

2. Nicht-parametrische Tests

e. Chi-Square Goodness of fit test (Anpassungstest)

Was wird getestet?

- Wahrscheinlichkeitsverteilung einer Variable

Testhypothese:

Die Variable (X) besitzt die angenommene Wahrscheinlichkeitsverteilung. $\rightarrow H_0: F(X) = F_0(X)$

Beispiele:

- Internetnutzer und Weltbevölkerung
- Klausurnoten und Normalverteilung
- aktueller Zensus und alter Zensus

3. Regressionen

a. Einfache Regression

Was wird getestet?

- linearer Zusammenhang zwischen abhängiger Variable (Y) und einer unabhängigen Variable (X)

Regressionsmodel:

Die abhängige Variable ist eine lineare Funktion der unabhängigen Variable.

$$\rightarrow Y(X) = a + b * X + \epsilon$$

mit **a**...Konstante, **b**...Regressionskoeffizient und **ϵ** ...Störterm mit Erwartungswert null ($E[\epsilon]=0$)

Beispiel:

- Semesterzahl in Abhängigkeit vom Alter

3. Regressionen

b. Multiple Regression

Was wird getestet?

- linearer Zusammenhang zwischen abhängiger Variable (Y) und mehrerer unabhängiger Variablen (X_j)

Regressionsmodel:

Die abhängige Variable ist eine (mehrdimensional) lineare Funktion der unabhängigen Variablen.

$$\rightarrow Y(X_j) = a + \sum b_j * X_j + \epsilon$$

$$Y(X_j) = a + b_1 * X_1 + b_2 * X_2 \dots + b_n * X_n + \epsilon$$

mit b_j ...Regressionskoeffizienten der Variablen

Beispiel:

- Gehalt in Abhängigkeit von Alter, Beruf, Studium,...

Vielen Dank für Ihre Aufmerksamkeit!

www.ovgu.de