

Statistik

01. Inferenz

Roland Schäfer

Institut für Germanistische Sprachwissenschaft
Friedrich-Schiller-Universität Jena

stets aktuelle Fassungen: <https://github.com/rsling/VL-Deutsche-Syntax>

1

Quantitative Analyse

- Quantitative empirische Forschung
- Grundbegriffe der quantitativen Forschung

- Validität
- Ableitung des Fisher-Exakt-Tests aus ersten Prinzipien

2

Nächste Woche | Überblick

Quantitativ

- Messungen und Experimente

- Messungen und Experimente
- Hypothesen und Theorien

- Messungen und Experimente
- Hypothesen und Theorien
- Begriff der „Variablen“

- Messungen und Experimente
- Hypothesen und Theorien
- Begriff der „Variablen“
- statistische Inferenz:

- Messungen und Experimente
- Hypothesen und Theorien
- Begriff der „Variablen“
- statistische Inferenz:
 - ▶ Hypothesenpaare H_1 und H_0

- Messungen und Experimente
- Hypothesen und Theorien
- Begriff der „Variablen“
- statistische Inferenz:
 - ▶ Hypothesenpaare H_1 und H_0
 - ▶ Typ I- und Typ II-Fehler

- Messungen und Experimente
- Hypothesen und Theorien
- Begriff der „Variablen“
- statistische Inferenz:
 - ▶ Hypothesenpaare H_1 und H_0
 - ▶ Typ I- und Typ II-Fehler
 - ▶ Validität

- Messungen und Experimente
- Hypothesen und Theorien
- Begriff der „Variablen“
- statistische Inferenz:
 - ▶ Hypothesenpaare H_1 und H_0
 - ▶ Typ I- und Typ II-Fehler
 - ▶ Validität
 - ▶ exakte Hypothesentestung

- Maxwell & Delaney (2004: Kap. 1 und 2)
- Gravetter & Wallnau (2007: Kap. 1)

- beobachtbare Phänomene

- beobachtbare Phänomene
- Beobachtungen reproduzierbar

- beobachtbare Phänomene
- Beobachtungen reproduzierbar
- messbar = beobachtbar (Sinneswahrnehmung an sich irrelevant)

- beobachtbare Phänomene
- Beobachtungen reproduzierbar
- messbar = beobachtbar (Sinneswahrnehmung an sich irrelevant)
- Realismus: „wirkliche“ Phänomene und ihre Mechanismen

- beobachtbare Phänomene
- Beobachtungen reproduzierbar
- messbar = beobachtbar (Sinneswahrnehmung an sich irrelevant)
- Realismus: „wirkliche“ Phänomene und ihre Mechanismen
- Experiment

- nicht-arbiträre Genauigkeit der Messung (Wirkung und Störeinflüsse)

- nicht-arbiträre Genauigkeit der Messung (Wirkung und Störeinflüsse)
- potentiell inadäquate Messung des theoretischen Konstrukts

- nicht-arbiträre Genauigkeit der Messung (Wirkung und Störeinflüsse)
- potentiell inadäquate Messung des theoretischen Konstrukts
- \Rightarrow Vermeidung von Fehlschluss auf unechte Ursachen

- nicht-arbiträre Genauigkeit der Messung (Wirkung und Störeinflüsse)
- potentiell inadäquate Messung des theoretischen Konstrukts
- \Rightarrow Vermeidung von Fehlschluss auf unechte Ursachen
- \Rightarrow relevante Ursachen

- nicht-arbiträre Genauigkeit der Messung (Wirkung und Störeinflüsse)
- potentiell inadäquate Messung des theoretischen Konstrukts
- \Rightarrow Vermeidung von Fehlschluss auf unechte Ursachen
- \Rightarrow relevante Ursachen
- Generalisierbarkeit der Ergebnisse

- nicht-arbiträre Genauigkeit der Messung (Wirkung und Störeinflüsse)
- potentiell inadäquate Messung des theoretischen Konstrukts
- \Rightarrow Vermeidung von Fehlschluss auf unechte Ursachen
- \Rightarrow relevante Ursachen
- Generalisierbarkeit der Ergebnisse
- Überkommen der individuell gefärbten Wahrnehmung

- Gegenstand: interne (mentale) Grammatik (I-Grammatik)
universeller und individueller Teil

- Gegenstand: interne (mentale) Grammatik (I-Grammatik)
universeller und individueller Teil
- I-Grammatik bei jedem Sprecher (leicht) verschieden

- Gegenstand: interne (mentale) Grammatik (I-Grammatik)
universeller und individueller Teil
- I-Grammatik bei jedem Sprecher (leicht) verschieden
- I-Grammatik erlaubt immer binäre Grammatikalitätsentscheidung

- Gegenstand: interne (mentale) Grammatik (I-Grammatik)
universeller und individueller Teil
- I-Grammatik bei jedem Sprecher (leicht) verschieden
- I-Grammatik erlaubt immer binäre Grammatikalitätsentscheidung
- also: Linguisten können eigene Grammatik untersuchen!
(Introspektion)

- Gegenstand: interne (mentale) Grammatik (I-Grammatik)
universeller und individueller Teil
- I-Grammatik bei jedem Sprecher (leicht) verschieden
- I-Grammatik erlaubt immer binäre Grammatikalitätsentscheidung
- also: Linguisten können eigene Grammatik untersuchen!
(Introspektion)
- universalistische Aussagen auf Basis solcher Ergebnisse zulässig

- Gegenstand: interne (mentale) Grammatik (I-Grammatik)
universeller und individueller Teil
- I-Grammatik bei jedem Sprecher (leicht) verschieden
- I-Grammatik erlaubt immer binäre Grammatikalitätsentscheidung
- also: Linguisten können eigene Grammatik untersuchen!
(Introspektion)
- universalistische Aussagen auf Basis solcher Ergebnisse zulässig
- Das ist auf allen Ebenen eine Frechheit!

- Ursprung der Hypothesen: Theorien

- Ursprung der Hypothesen: Theorien
- interessante Hypothesen:

- Ursprung der Hypothesen: Theorien
- interessante Hypothesen:
 - ▶ Formulierung relevanter Kausationsbedingung (wenn, dann)

- Ursprung der Hypothesen: Theorien
- interessante Hypothesen:
 - ▶ Formulierung relevanter Kausationsbedingung (wenn, dann)
 - ▶ großer Gültigkeitsbereich (ein Sprecher vs. alle Sprecher)

- Kann die Hypothese weiter angenommen werden, oder liefert das Experiment starke Evidenz gegen sie?

- Kann die Hypothese weiter angenommen werden, oder liefert das Experiment starke Evidenz gegen sie?
- Probleme bei Prüfung:

- Kann die Hypothese weiter angenommen werden, oder liefert das Experiment starke Evidenz gegen sie?
- Probleme bei Prüfung:
 - ▶ störende Einflüsse

- Kann die Hypothese weiter angenommen werden, oder liefert das Experiment starke Evidenz gegen sie?
- Probleme bei Prüfung:
 - ▶ störende Einflüsse
 - ▶ ungenaue Operationalisierung (s. u.)

- Kann die Hypothese weiter angenommen werden, oder liefert das Experiment starke Evidenz gegen sie?
- Probleme bei Prüfung:
 - ▶ störende Einflüsse
 - ▶ ungenaue Operationalisierung (s. u.)
 - ▶ stochastische Natur des Phänomens

- Kann die Hypothese weiter angenommen werden, oder liefert das Experiment starke Evidenz gegen sie?
- Probleme bei Prüfung:
 - ▶ störende Einflüsse
 - ▶ ungenaue Operationalisierung (s. u.)
 - ▶ stochastische Natur des Phänomens
 - ▶ kleine Stichprobe (s. u.)

- Kann die Hypothese weiter angenommen werden, oder liefert das Experiment starke Evidenz gegen sie?
- Probleme bei Prüfung:
 - ▶ störende Einflüsse
 - ▶ ungenaue Operationalisierung (s. u.)
 - ▶ stochastische Natur des Phänomens
 - ▶ kleine Stichprobe (s. u.)
- falsche Positive und falsche Negative jeweils zu vermeiden

- Daten besorgen/Experiment machen

- Daten besorgen/Experiment machen
- typischerweise kleine Datenmengen

- Daten besorgen/Experiment machen
- typischerweise kleine Datenmengen
- Datenbetrachtung durch menschliche Wahrnehmung

- Daten besorgen/Experiment machen
- typischerweise kleine Datenmengen
- Datenbetrachtung durch menschliche Wahrnehmung
- Interpretation der Daten durch Nachdenken

- Daten besorgen/Experiment machen
- typischerweise kleine Datenmengen
- Datenbetrachtung durch menschliche Wahrnehmung
- Interpretation der Daten durch Nachdenken
- extrem fehleranfällig (Fehler = unzulässige Generalisierung)

- Daten besorgen/Experiment machen
- typischerweise kleine Datenmengen
- Datenbetrachtung durch menschliche Wahrnehmung
- Interpretation der Daten durch Nachdenken

- extrem fehleranfällig (Fehler = unzulässige Generalisierung)
- wichtig zur Hypothesengenerierung

- Daten besorgen/Experiment machen

- Daten besorgen/Experiment machen
- typischerweise größere Datenmengen

- Daten besorgen/Experiment machen
- typischerweise größere Datenmengen
- Datenanalyse durch Visualisierung/mathematische Datenbeschreibung

- Daten besorgen/Experiment machen
- typischerweise größere Datenmengen
- Datenanalyse durch Visualisierung/mathematische Datenbeschreibung
- Hypothesenprüfung durch Testverfahren (s. u.)

- Daten besorgen/Experiment machen
- typischerweise größere Datenmengen
- Datenanalyse durch Visualisierung/mathematische Datenbeschreibung
- Hypothesenprüfung durch Testverfahren (s. u.)
- Grundlage moderner Wissenschaft

- Operationalisierung = präzise Formulierung der Messmethode für ein theoretisches Konstrukt

- Operationalisierung = präzise Formulierung der Messmethode für ein theoretisches Konstrukt
- Bsp. Konstrukt „Satzlänge“: Wortanzahl? Phonemanzahl? Phrasenanzahl?

- Operationalisierung = präzise Formulierung der Messmethode für ein theoretisches Konstrukt
- Bsp. Konstrukt „Satzlänge“: Wortanzahl? Phonemanzahl? Phrasenanzahl?
- Bsp. Konstrukt „Satztopik“: Oha!?! (Cook & Bildhauer 2013)

- Operationalisierung = präzise Formulierung der Messmethode für ein theoretisches Konstrukt
- Bsp. Konstrukt „Satzlänge“: Wortanzahl? Phonemanzahl? Phrasenanzahl?
- Bsp. Konstrukt „Satztopik“: Oha!?! (Cook & Bildhauer 2013)
- alle genannten Beispiele: abhängig von Auxiliarrhypothesen bzw. anderen theoretischen Konstrukten (Wort, Phonem, Phrase, ...)

- von Interesse: allgemeine Gesetzmäßigkeiten

- von Interesse: allgemeine Gesetzmäßigkeiten
- also Untersuchungsgegenstand: alle X (Sprecher, Sätze, ...)

- von Interesse: allgemeine Gesetzmäßigkeiten
- also Untersuchungsgegenstand: alle X (Sprecher, Sätze, ...)
- untersuchbar: kleine Menge von X

- von Interesse: allgemeine Gesetzmäßigkeiten
- also Untersuchungsgegenstand: alle X (Sprecher, Sätze, ...)
- untersuchbar: kleine Menge von X
- wenn „alle X “ an sich sowieso sehr klein:
interessante Fragestellung nur schwer möglich

- von Interesse: allgemeine Gesetzmäßigkeiten
- also Untersuchungsgegenstand: alle X (Sprecher, Sätze, ...)
- untersuchbar: kleine Menge von X
- wenn „alle X “ an sich sowieso sehr klein:
interessante Fragestellung nur schwer möglich
- fiktives Beispiel:

- von Interesse: allgemeine Gesetzmäßigkeiten
- also Untersuchungsgegenstand: alle X (Sprecher, Sätze, ...)
- untersuchbar: kleine Menge von X
- wenn „alle X“ an sich sowieso sehr klein:
interessante Fragestellung nur schwer möglich
- fiktives Beispiel:
 - ▶ Sprecher ohne Ersatzinfinitiv (*dass ich schlafen gemusst habe*)
benutzen Modalverben öfter im Präteritum (*dass ich schlafen musste*)

- von Interesse: allgemeine Gesetzmäßigkeiten
- also Untersuchungsgegenstand: alle X (Sprecher, Sätze, ...)
- untersuchbar: kleine Menge von X
- wenn „alle X“ an sich sowieso sehr klein:
interessante Fragestellung nur schwer möglich
- fiktives Beispiel:
 - ▶ Sprecher ohne Ersatzinfinitiv (*dass ich schlafen gemusst habe*)
benutzen Modalverben öfter im Präteritum (*dass ich schlafen musste*)
 - ▶ ...aber es gibt nur einen Sprecher ohne Ersatzinfinitiv

- von Interesse: allgemeine Gesetzmäßigkeiten
- also Untersuchungsgegenstand: alle X (Sprecher, Sätze, ...)
- untersuchbar: kleine Menge von X
- wenn „alle X“ an sich sowieso sehr klein:
interessante Fragestellung nur schwer möglich
- fiktives Beispiel:
 - ▶ Sprecher ohne Ersatzinfinitiv (*dass ich schlafen gemusst habe*)
benutzen Modalverben öfter im Präteritum (*dass ich schlafen musste*)
 - ▶ ...aber es gibt nur einen Sprecher ohne Ersatzinfinitiv
 - ▶ \Rightarrow dann auch kein Theorie- und Empiriebedarf

- weil Grundgesamtheit intrinsisch nicht betrachtbar:
Verwendung eingeschränkter Datensätze

- weil Grundgesamtheit intrinsisch nicht betrachtbar:
Verwendung eingeschränkter Datensätze
- ideal: uniform zufällige Stichprobe
= jedes Element der Grundgesamtheit hat die gleiche Chance beim Ziehen

- weil Grundgesamtheit intrinsisch nicht betrachtbar:
Verwendung eingeschränkter Datensätze
- ideal: uniform zufällige Stichprobe
= jedes Element der Grundgesamtheit hat die gleiche Chance beim Ziehen
- andere Möglichkeit: stratifizierte Stichprobe
= Stichprobe so zusammengesetzt, dass wichtige Eigenschaften proportional repräsentiert sind

- weil Grundgesamtheit intrinsisch nicht betrachtbar:
Verwendung eingeschränkter Datensätze
- ideal: uniform zufällige Stichprobe
= jedes Element der Grundgesamtheit hat die gleiche Chance beim Ziehen
- andere Möglichkeit: stratifizierte Stichprobe
= Stichprobe so zusammengesetzt, dass wichtige Eigenschaften proportional repräsentiert sind
- Problem bei Letzterem: haufenweise Auxiliärhypothesen

- weil Grundgesamtheit intrinsisch nicht betrachtbar:
Verwendung eingeschränkter Datensätze
- ideal: uniform zufällige Stichprobe
= jedes Element der Grundgesamtheit hat die gleiche Chance beim Ziehen
- andere Möglichkeit: stratifizierte Stichprobe
= Stichprobe so zusammengesetzt, dass wichtige Eigenschaften proportional repräsentiert sind
- Problem bei Letzterem: haufenweise Auxiliarhypothesen
- außerdem: nicht unbedingt erforderlich, s. letzten Teil der heutigen VL

- Ziel: Schluss von Stichprobe auf Grundgesamtheit

- Ziel: Schluss von Stichprobe auf Grundgesamtheit
- kritisch: Stichprobengröße

- Ziel: Schluss von Stichprobe auf Grundgesamtheit
- kritisch: Stichprobengröße
- Kriterium: Wahrscheinlichkeit, das Ergebnis per Zufall zu bekommen

- Ziel: Schluss von Stichprobe auf Grundgesamtheit
- kritisch: Stichprobengröße
- Kriterium: Wahrscheinlichkeit, das Ergebnis per Zufall zu bekommen
- starker Einfluss: Stärke des Effekts in der Stichprobe

- meist uninteressanter Typ Fragestellung:
„Wieviel Prozent X haben Eigenschaft A?“

- meist uninteressanter Typ Fragestellung:
„Wieviel Prozent X haben Eigenschaft A?“
- wegen Fehlens kausaler Zusammenhänge

- meist uninteressanter Typ Fragestellung:
„Wieviel Prozent X haben Eigenschaft A?“
- wegen Fehlens kausaler Zusammenhänge
- Bsp.: wie oft *wegen* mit Dat bzw. Gen?

- meist uninteressanter Typ Fragestellung:
„Wieviel Prozent X haben Eigenschaft A?“
- wegen Fehlens kausaler Zusammenhänge
- Bsp.: wie oft *wegen* mit Dat bzw. Gen?
- besser:
„Hängt bei X die Wahrscheinlichkeit für Eigenschaft B von A ab?“

- meist uninteressanter Typ Fragestellung:
„Wieviel Prozent X haben Eigenschaft A?“
- wegen Fehlens kausaler Zusammenhänge
- Bsp.: wie oft *wegen* mit Dat bzw. Gen?
- besser:
„Hängt bei X die Wahrscheinlichkeit für Eigenschaft B von A ab?“
- Bsp.: „Nehmen nominale Pröp eher Gen oder Dat?“

- meist uninteressanter Typ Fragestellung:
„Wieviel Prozent X haben Eigenschaft A?“
- wegen Fehlens kausaler Zusammenhänge
- Bsp.: wie oft *wegen* mit Dat bzw. Gen?
- besser:
„Hängt bei X die Wahrscheinlichkeit für Eigenschaft B von A ab?“
- Bsp.: „Nehmen denominale Präp eher Gen oder Dat?“

- meist uninteressanter Typ Fragestellung:
„Wieviel Prozent X haben Eigenschaft A?“
- wegen Fehlens kausaler Zusammenhänge
- Bsp.: wie oft *wegen* mit Dat bzw. Gen?
- besser:
„Hängt bei X die Wahrscheinlichkeit für Eigenschaft B von A ab?“
- Bsp.: „Nehmen denominale Präp eher Gen oder Dat?“

konzeptuell:

	denominale P	andere P
Dat	x_1	x_2
Gen	x_3	x_4

- Eigenschaften, quantitativ gemessen: Variablen

- Eigenschaften, quantitativ gemessen: Variablen
- im Experiment:

- Eigenschaften, quantitativ gemessen: Variablen
- im Experiment:
 - ▶ kontrolliere für Theorie irrelevante Variablen (Störvariablen)

- Eigenschaften, quantitativ gemessen: Variablen
- im Experiment:
 - ▶ kontrolliere für Theorie irrelevante Variablen (Störvariablen)
 - ▶ variiere „Ursachen-Variablen“ (unabhängige Variablen)

- Eigenschaften, quantitativ gemessen: Variablen
- im Experiment:
 - ▶ kontrolliere für Theorie irrelevante Variablen (Störvariablen)
 - ▶ variiere „Ursachen-Variablen“ (unabhängige Variablen)
 - ▶ beobachte „Wirkung-Variablen“ (abhängige Variablen)

- Problem in Astronomie, Korpuslinguistik usw.: keine Experimente möglich

- Problem in Astronomie, Korpuslinguistik usw.: keine Experimente möglich
- unabhängige Variablen nicht variierbar

- Problem in Astronomie, Korpuslinguistik usw.: keine Experimente möglich
- unabhängige Variablen nicht variierbar
- Daten liegen bereits vor bzw. fallen vom Himmel

- Problem in Astronomie, Korpuslinguistik usw.: keine Experimente möglich
- unabhängige Variablen nicht variierbar
- Daten liegen bereits vor bzw. fallen vom Himmel
- Auswahl von Datensätzen, so dass von den unabhängigen Variablen die zur Theorieprüfung nötigen Permutationen im Datensatz vorkommen

- Problem in Astronomie, Korpuslinguistik usw.: keine Experimente möglich
- unabhängige Variablen nicht variierbar
- Daten liegen bereits vor bzw. fallen vom Himmel
- Auswahl von Datensätzen, so dass von den unabhängigen Variablen die zur Theorieprüfung nötigen Permutationen im Datensatz vorkommen
- dabei Zusatzproblem bei Korpuslinguistik: Korpus meist nicht das eigene, wenig Informationen über mögliche Verzerrungen

- Problem in Astronomie, Korpuslinguistik usw.: keine Experimente möglich
- unabhängige Variablen nicht variierbar
- Daten liegen bereits vor bzw. fallen vom Himmel
- Auswahl von Datensätzen, so dass von den unabhängigen Variablen die zur Theorieprüfung nötigen Permutationen im Datensatz vorkommen
- dabei Zusatzproblem bei Korpuslinguistik: Korpus meist nicht das eigene, wenig Informationen über mögliche Verzerrungen
- Was ist die Grundgesamtheit?

- H1 (eigene Hypothese):
erwarteter Variablenzusammenhang aus Kausalrelation

Hypothesenpaare

- H_1 (eigene Hypothese):
erwarteter Variablenzusammenhang aus Kausalrelation
- H_0 (Nullhypothese): Negation der H_1

- H_1 (eigene Hypothese):
erwarteter Variablenzusammenhang aus Kausalrelation
- H_0 (Nullhypothese): Negation der H_1
- Ziel der Inferenzstatistik: Zurückweisung der H_0

- H_1 (eigene Hypothese):
erwarteter Variablenzusammenhang aus Kausalrelation
- H_0 (Nullhypothese): Negation der H_1
- Ziel der Inferenzstatistik: Zurückweisung der H_0
- Logik: *Entweder ein sehr seltenes Ereignis ist eingetreten, oder es besteht tatsächlich ein Zusammenhang.*

- H_1 (eigene Hypothese):
erwarteter Variablenzusammenhang aus Kausalrelation
- H_0 (Nullhypothese): Negation der H_1
- Ziel der Inferenzstatistik: Zurückweisung der H_0
- Logik: *Entweder ein sehr seltenes Ereignis ist eingetreten, oder es besteht tatsächlich ein Zusammenhang.*
- \Rightarrow Stärkung der H_1 , aber nicht „Beweis“!

- H_1 (eigene Hypothese):
erwarteter Variablenzusammenhang aus Kausalrelation
- H_0 (Nullhypothese): Negation der H_1
- Ziel der Inferenzstatistik: Zurückweisung der H_0
- Logik: *Entweder ein sehr seltenes Ereignis ist eingetreten, oder es besteht tatsächlich ein Zusammenhang.*
- \Rightarrow Stärkung der H_1 , aber nicht „Beweis“!

- geringe Bedeutung für sich genommen

- H_1 (eigene Hypothese):
erwarteter Variablenzusammenhang aus Kausalrelation
 - H_0 (Nullhypothese): Negation der H_1
 - Ziel der Inferenzstatistik: Zurückweisung der H_0
 - Logik: *Entweder ein sehr seltenes Ereignis ist eingetreten, oder es besteht tatsächlich ein Zusammenhang.*
 - \Rightarrow Stärkung der H_1 , aber nicht „Beweis“!
-
- geringe Bedeutung für sich genommen
 - Severity hängt von viel mehr Faktoren ab (Mayo)

- H1 (eigene Hypothese):
erwarteter Variablenzusammenhang aus Kausalrelation
 - Ho (Nullhypothese): Negation der H1
 - Ziel der Inferenzstatistik: Zurückweisung der Ho
 - Logik: *Entweder ein sehr seltenes Ereignis ist eingetreten, oder es besteht tatsächlich ein Zusammenhang.*
 - \Rightarrow Stärkung der H1, aber nicht „Beweis“!
-
- geringe Bedeutung für sich genommen
 - Severity hängt von viel mehr Faktoren ab (Mayo)
 - weitere Experimente/Replikationen

Severity (strong): We have evidence for a claim C just to the extent it survives a stringent scrutiny. If C passes a test that was highly capable of finding flaws or discrepancies from C , and yet none or few are found, then the passing result, x , is evidence for C . [Mayo 2018:14]

- übliche Hypothesen (H_1):
„Es besteht ein Zusammenhang zwischen A und B.“

- übliche Hypothesen (H_1):
„Es besteht ein Zusammenhang zwischen A und B.“
- H_0 dann: Fehlen des Zusammenhangs

- übliche Hypothesen (H_1):
„Es besteht ein Zusammenhang zwischen A und B.“
- H_0 dann: Fehlen des Zusammenhangs
- = ungerichtete Hypothese

- übliche Hypothesen (H_1):
„Es besteht ein Zusammenhang zwischen A und B.“
- H_0 dann: Fehlen des Zusammenhangs
- = ungerichtete Hypothese
- gerichtete Version: Benennung des genauen numerischen Zusammenhangs

- übliche Hypothesen (H_1):
„Es besteht ein Zusammenhang zwischen A und B.“
- H_0 dann: Fehlen des Zusammenhangs
- = ungerichtete Hypothese
- gerichtete Version: Benennung des genauen numerischen Zusammenhangs
- gerichtete Hypothesen: stärkere Aussage, schwerer zu zeigen, immer noch kein „Beweis“

- je nach

- je nach
 - ▶ Stärke des Effekts

- je nach
 - ▶ Stärke des Effekts
 - ▶ Homogenität der Grundgesamtheit/Stichprobe

- je nach
 - ▶ Stärke des Effekts
 - ▶ Homogenität der Grundgesamtheit/Stichprobe
 - ▶ Stichprobengröße

- je nach
 - ▶ Stärke des Effekts
 - ▶ Homogenität der Grundgesamtheit/Stichprobe
 - ▶ Stichprobengröße
- Gefahr zweier möglicher Inferenzfehler:

- je nach
 - ▶ Stärke des Effekts
 - ▶ Homogenität der Grundgesamtheit/Stichprobe
 - ▶ Stichprobengröße
- Gefahr zweier möglicher Inferenzfehler:
 - ▶ Typ I (α): H_0 wird fälschlicherweise zurückgewiesen

- je nach
 - ▶ Stärke des Effekts
 - ▶ Homogenität der Grundgesamtheit/Stichprobe
 - ▶ Stichprobengröße
- Gefahr zweier möglicher Inferenzfehler:
 - ▶ Typ I (α): H_0 wird fälschlicherweise zurückgewiesen
 - ▶ Typ II (β): H_0 fälschlicherweise nicht zurückgewiesen

- je nach
 - ▶ Stärke des Effekts
 - ▶ Homogenität der Grundgesamtheit/Stichprobe
 - ▶ Stichprobengröße
- Gefahr zweier möglicher Inferenzfehler:
 - ▶ Typ I (α): H_0 wird fälschlicherweise zurückgewiesen
 - ▶ Typ II (β): H_0 fälschlicherweise nicht zurückgewiesen
- Typ II-Fehler mehr gefürchtet, weil Artikel dann nicht angenommen wird

- je nach
 - ▶ Stärke des Effekts
 - ▶ Homogenität der Grundgesamtheit/Stichprobe
 - ▶ Stichprobengröße
- Gefahr zweier möglicher Inferenzfehler:
 - ▶ Typ I (α): H_0 wird fälschlicherweise zurückgewiesen
 - ▶ Typ II (β): H_0 fälschlicherweise nicht zurückgewiesen
- Typ II-Fehler mehr gefürchtet, weil Artikel dann nicht angenommen wird
- inhaltlich beide gleich schwerwiegend

- Logik der Hypothesenprüfung (genauer):
Das Testergebnis ist so unwahrscheinlich per Zufall (H_0) zu erzielen, dass H_1 plausibel ist.
- die akzeptierte Schranke: α -Niveau
- Linguistik/Sozialwissenschaften: $\alpha = 0.05$
- wenn erreicht, gilt: Wenn in der GG der Zusammenhang nicht besteht, würde man nur in einer von 20 Stichproben die beobachtete Verteilung erwarten.
- *If I tell you, you have a 5% chance of being shot when you walk through the door, you go through the window.*
- α =Häufigkeit, mit der man sich langfristig positiv irrt
(H_0 korrekt, aber zurückgewiesen)
- vs. Teststärke: β =Häufigkeit, mit der man sich langfristig negativ irrt
(H_0 falsch, aber nicht zurückgewiesen)

Gefahren für Typ I/II-Fehler

- math. Bedingungen für Test nicht erfüllt (Typ I)

Gefahren für Typ I/II-Fehler

- math. Bedingungen für Test nicht erfüllt (Typ I)
- kumulierter α -Fehler (Typ I, s. nächste Folie)

Gefahren für Typ I/II-Fehler

- math. Bedingungen für Test nicht erfüllt (Typ I)
- kumulierter α -Fehler (Typ I, s. nächste Folie)
- kleine Stichprobe (Typ II)

Gefahren für Typ I/II-Fehler

- math. Bedingungen für Test nicht erfüllt (Typ I)
- kumulierter α -Fehler (Typ I, s. nächste Folie)
- kleine Stichprobe (Typ II)
- zu große Variation in der GG (Typ II)

Gefahren für Typ I/II-Fehler

- math. Bedingungen für Test nicht erfüllt (Typ I)
- kumulierter α -Fehler (Typ I, s. nächste Folie)
- kleine Stichprobe (Typ II)
- zu große Variation in der GG (Typ II)

- in Korpora: schlechte Zusammensetzung des Korpus
⇒ Phänomen mit mangelhafter Dispersion (Typ I)

- Irrtum beim Herstellen des Kausalzusammenhangs

- Irrtum beim Herstellen des Kausalzusammenhangs
- Grund (Korpuslinguistik): verzernte Stichprobenzusammensetzung

- Irrtum beim Herstellen des Kausalzusammenhangs
- Grund (Korpuslinguistik): verzernte Stichprobenzusammensetzung
- Bsp.:

- Irrtum beim Herstellen des Kausalzusammenhangs
- Grund (Korpuslinguistik): verzerrte Stichprobenzusammensetzung
- Bsp.:
 - ▶ H1: Im DECOW2012 kommt öfter das Pronomen *son* vor als im DWDS Kernkorpus, weil es erst nach 2000 zum eigenständigen Pronomen wurde.

- Irrtum beim Herstellen des Kausalzusammenhangs
- Grund (Korpuslinguistik): verzerrte Stichprobenzusammensetzung
- Bsp.:
 - ▶ H1: Im DECOW2012 kommt öfter das Pronomen *son* vor als im DWDS Kernkorpus, weil es erst nach 2000 zum eigenständigen Pronomen wurde.
 - ▶ Ho wird auf Basis zweier Stichproben (DECOW2012, DWDS) zurückgewiesen.

- Irrtum beim Herstellen des Kausalzusammenhangs
- Grund (Korpuslinguistik): verzernte Stichprobenzusammensetzung
- Bsp.:
 - ▶ H1: Im DECOW2012 kommt öfter das Pronomen *son* vor als im DWDS Kernkorpus, weil es erst nach 2000 zum eigenständigen Pronomen wurde.
 - ▶ Ho wird auf Basis zweier Stichproben (DECOW2012, DWDS) zurückgewiesen.
 - ▶ wirkliche Ursache: Registerunterschiede

- Korrektheit des theoretischen Konstrukts

- Korrektheit des theoretischen Konstrukts
- eigentlich aus der Psychologie

- Korrektheit des theoretischen Konstrukts
- eigentlich aus der Psychologie
- aber riesiges mißachtetes Problem in der Linguistik

- Korrektheit des theoretischen Konstrukts
- eigentlich aus der Psychologie
- aber riesiges mißachtetes Problem in der Linguistik
- Bsp.:

- Korrektheit des theoretischen Konstrukts
- eigentlich aus der Psychologie
- aber riesiges mißachtetes Problem in der Linguistik
- Bsp.:
 - ▶ Beobachtung: Das Deutsche bewahrt genus-typische Pluralflexion am Substantiv.

- Korrektheit des theoretischen Konstrukts
- eigentlich aus der Psychologie
- aber riesiges mißachtetes Problem in der Linguistik
- Bsp.:
 - ▶ Beobachtung: Das Deutsche bewahrt genus-typische Pluralflexion am Substantiv.
 - ▶ Konstrukt: Nominalklammer/Klammerprinzip (NP-Kongruenzklammer Artikel-Substantiv, Ronneberger-Sibold 2010)

- Korrektheit des theoretischen Konstrukts
- eigentlich aus der Psychologie
- aber riesiges mißachtetes Problem in der Linguistik
- Bsp.:
 - ▶ Beobachtung: Das Deutsche bewahrt genus-typische Pluralflexion am Substantiv.
 - ▶ Konstrukt: Nominalklammer/Klammerprinzip (NP-Kongruenzklammer Artikel-Substantiv, Ronneberger-Sibold 2010)
 - ▶ Hypothese zu Beobachtung: Flexionserhalt stärkt Klammerprinzip

- Korrektheit des theoretischen Konstrukts
- eigentlich aus der Psychologie
- aber riesiges mißachtetes Problem in der Linguistik
- Bsp.:
 - ▶ Beobachtung: Das Deutsche bewahrt genus-typische Pluralflexion am Substantiv.
 - ▶ Konstrukt: Nominalklammer/Klammerprinzip (NP-Kongruenzklammer Artikel-Substantiv, Ronneberger-Sibold 2010)
 - ▶ Hypothese zu Beobachtung: Flexionserhalt stärkt Klammerprinzip
 - ▶ Das Konstrukt ist hochgradig beliebig und unterdefiniert.

- Korrektheit des theoretischen Konstrukts
- eigentlich aus der Psychologie
- aber riesiges mißachtetes Problem in der Linguistik
- Bsp.:
 - ▶ Beobachtung: Das Deutsche bewahrt genus-typische Pluralflexion am Substantiv.
 - ▶ Konstrukt: Nominalklammer/Klammerprinzip (NP-Kongruenzklammer Artikel-Substantiv, Ronneberger-Sibold 2010)
 - ▶ Hypothese zu Beobachtung: Flexionserhalt stärkt Klammerprinzip
 - ▶ Das Konstrukt ist hochgradig beliebig und unterdefiniert.
 - ▶ Abhilfe: nur Konstrukte/Hypothesen, die starke Vorhersagen generieren (s. Junggrammatiker)

- Generalisierbarkeit der Ergebnisse (über Raum, Zeit usw.)

- Generalisierbarkeit der Ergebnisse (über Raum, Zeit usw.)
- Problem: zu große Homogenität der Stichprobe
(was für statistische Validität wiederum gut ist)

- Generalisierbarkeit der Ergebnisse (über Raum, Zeit usw.)
- Problem: zu große Homogenität der Stichprobe (was für statistische Validität wiederum gut ist)
- Bezug auf Korpora:

- Generalisierbarkeit der Ergebnisse (über Raum, Zeit usw.)
- Problem: zu große Homogenität der Stichprobe (was für statistische Validität wiederum gut ist)
- Bezug auf Korpora:
 - ▶ zu spezifische Stratifikation (DeReKo)

- Generalisierbarkeit der Ergebnisse (über Raum, Zeit usw.)
- Problem: zu große Homogenität der Stichprobe (was für statistische Validität wiederum gut ist)
- Bezug auf Korpora:
 - ▶ zu spezifische Stratifikation (DeReKo)
 - ▶ verzerrte Stichprobe (evtl. traditionelle Webkorpora)

- Statistik als Teil der rationalen wissenschaftlichen Argumentation, der Interpretation von Experimenten

- Statistik als Teil der rationalen wissenschaftlichen Argumentation, der Interpretation von Experimenten
- daher: möglichst kein Mathematik-Jargon

- Statistik als Teil der rationalen wissenschaftlichen Argumentation, der Interpretation von Experimenten
- daher: möglichst kein Mathematik-Jargon
- eingeschränkte Induktion als theoriegeleitete Dateninterpretation

- Statistik als Teil der rationalen wissenschaftlichen Argumentation, der Interpretation von Experimenten
- daher: möglichst kein Mathematik-Jargon
- eingeschränkte Induktion als theoriegeleitete Dateninterpretation
- Kontrolle aller unabhängigen Variablen

- Statistik als Teil der rationalen wissenschaftlichen Argumentation, der Interpretation von Experimenten
- daher: möglichst kein Mathematik-Jargon
- eingeschränkte Induktion als theoriegeleitete Dateninterpretation
- Kontrolle aller unabhängigen Variablen
- alle anderen (Stör-)Variablen konzeptuell zufallsgebunden

- Behauptung: Dame X kann am Geschmack erkennen, ob der Tee oder die Milch zuerst in die Tasse gegossen wurde.

- Behauptung: Dame X kann am Geschmack erkennen, ob der Tee oder die Milch zuerst in die Tasse gegossen wurde.
- prä-fishersches Konzept: alle Störvariablen kontrollieren und gleich machen, sonst keine valide Inferenz möglich

- Behauptung: Dame X kann am Geschmack erkennen, ob der Tee oder die Milch zuerst in die Tasse gegossen wurde.
- prä-fishersches Konzept: alle Störvariablen kontrollieren und gleich machen, sonst keine valide Inferenz möglich
- Fisher: Das ist prinzipiell unmöglich, umständlich, teuer, unnötig!

- Behauptung: Dame X kann am Geschmack erkennen, ob der Tee oder die Milch zuerst in die Tasse gegossen wurde.
- prä-fishersches Konzept: alle Störvariablen kontrollieren und gleich machen, sonst keine valide Inferenz möglich
- Fisher: Das ist prinzipiell unmöglich, umständlich, teuer, unnötig!
- wenn alle irrelevanten Stör-Faktoren zufällig, dann:

- Behauptung: Dame X kann am Geschmack erkennen, ob der Tee oder die Milch zuerst in die Tasse gegossen wurde.
- prä-fishersches Konzept: alle Störvariablen kontrollieren und gleich machen, sonst keine valide Inferenz möglich
- Fisher: Das ist prinzipiell unmöglich, umständlich, teuer, unnötig!
- wenn alle irrelevanten Stör-Faktoren zufällig, dann:
 - ▶ Variiere die relevante unabhängige Variable.

- Behauptung: Dame X kann am Geschmack erkennen, ob der Tee oder die Milch zuerst in die Tasse gegossen wurde.
- prä-fishersches Konzept: alle Störvariablen kontrollieren und gleich machen, sonst keine valide Inferenz möglich
- Fisher: Das ist prinzipiell unmöglich, umständlich, teuer, unnötig!
- wenn alle irrelevanten Stör-Faktoren zufällig, dann:
 - ▶ Variiere die relevante unabhängige Variable.
 - ▶ Vergleiche das Ergebnis mit zufällig erwartbaren Ergebnissen.

- Behauptung: Dame X kann am Geschmack erkennen, ob der Tee oder die Milch zuerst in die Tasse gegossen wurde.
- prä-fishersches Konzept: alle Störvariablen kontrollieren und gleich machen, sonst keine valide Inferenz möglich
- Fisher: Das ist prinzipiell unmöglich, umständlich, teuer, unnötig!
- wenn alle irrelevanten Stör-Faktoren zufällig, dann:
 - ▶ Variiere die relevante unabhängige Variable.
 - ▶ Vergleiche das Ergebnis mit zufällig erwartbaren Ergebnissen.
 - ▶ Wie unwahrscheinlich ist das erzielte Ergebnis unter der Zufallsannahme?

- acht Tassen (zwei Milch zuerst, zwei Tee zuerst)

Wahrscheinlichkeiten, richtig zu raten

- acht Tassen (zwei Milch zuerst, zwei Tee zuerst)
- Mit wie vielen richtigen Treffern wären Sie zufrieden?

Wahrscheinlichkeiten, richtig zu raten

- acht Tassen (zwei Milch zuerst, zwei Tee zuerst)
- Mit wie vielen richtigen Treffern wären Sie zufrieden?

Wahrscheinlichkeiten, richtig zu raten

- acht Tassen (zwei Milch zuerst, zwei Tee zuerst)
- Mit wie vielen richtigen Treffern wären Sie zufrieden?

- acht Tassen (zwei Milch zuerst, zwei Tee zuerst)
- Mit wie vielen richtigen Treffern wären Sie zufrieden?
- Es muss die Wahrscheinlichkeit errechnet werden, eine, zwei, drei oder vier Tassen richtig zu raten.

- acht Tassen (zwei Milch zuerst, zwei Tee zuerst)
- Mit wie vielen richtigen Treffern wären Sie zufrieden?
- Es muss die Wahrscheinlichkeit errechnet werden, eine, zwei, drei oder vier Tassen richtig zu raten.
- Typischerweise schätzen Menschen solche Kombinatorikprobleme intuitiv falsch ein.

$$P(\text{richtig per Zufall}) = \frac{\text{Anzahl richtiger Zuweisungen}}{\text{Anzahl aller potentiellen Zuweisungen}} \quad (1)$$

$$P(\text{richtig per Zufall}) = \frac{\text{Anzahl richtiger Zuweisungen}}{\text{Anzahl aller potentiellen Zuweisungen}} \quad (1)$$

- Anzahl richtiger Zuweisungen: 1

$$P(\text{richtig per Zufall}) = \frac{\text{Anzahl richtiger Zuweisungen}}{\text{Anzahl aller potentiellen Zuweisungen}} \quad (1)$$

- Anzahl richtiger Zuweisungen: 1
- mögliche Zuweisungen: einfaches kombinatorisches Problem

mögliche Zuweisung von acht Tassen zu Milch/Tee zuerst

- erste MZ-Tasse: eine von 8

mögliche Zuweisung von acht Tassen zu Milch/Tee zuerst

- erste MZ-Tasse: eine von 8
- zweite MZ-Tasse: eine von 7

mögliche Zuweisung von acht Tassen zu Milch/Tee zuerst

- erste MZ-Tasse: eine von 8
- zweite MZ-Tasse: eine von 7
- dritte MZ-Tasse: eine von 6

mögliche Zuweisung von acht Tassen zu Milch/Tee zuerst

- erste MZ-Tasse: eine von 8
- zweite MZ-Tasse: eine von 7
- dritte MZ-Tasse: eine von 6
- vierte MZ-Tasse: eine von 5

mögliche Zuweisung von acht Tassen zu Milch/Tee zuerst

- erste MZ-Tasse: eine von 8
- zweite MZ-Tasse: eine von 7
- dritte MZ-Tasse: eine von 6
- vierte MZ-Tasse: eine von 5
- fünfte MZ-Tasse: STOPP (automatisch TZ)

mögliche Zuweisung von acht Tassen zu Milch/Tee zuerst

- erste MZ-Tasse: eine von 8
- zweite MZ-Tasse: eine von 7
- dritte MZ-Tasse: eine von 6
- vierte MZ-Tasse: eine von 5
- fünfte MZ-Tasse: STOPP (automatisch TZ)
- Möglichkeiten, 4 Tassen aus 8 auszuwählen:
 $8 \cdot 7 \cdot 6 \cdot 5 = 1680$

- bisher: jedes Set von 4 MZ-Tassen ist in verschiedenen Reihenfolgen in der Menge der Möglichkeiten

- bisher: jedes Set von 4 MZ-Tassen ist in verschiedenen Reihenfolgen in der Menge der Möglichkeiten
- Möglichkeiten, vier Tassen zu ordnen:
 $4 \cdot 3 \cdot 2 \cdot 1 = 24$

- bisher: jedes Set von 4 MZ-Tassen ist in verschiedenen Reihenfolgen in der Menge der Möglichkeiten
- Möglichkeiten, vier Tassen zu ordnen:
 $4 \cdot 3 \cdot 2 \cdot 1 = 24$
- Reihenfolge hier egal, also:

- bisher: jedes Set von 4 MZ-Tassen ist in verschiedenen Reihenfolgen in der Menge der Möglichkeiten
- Möglichkeiten, vier Tassen zu ordnen:
 $4 \cdot 3 \cdot 2 \cdot 1 = 24$
- Reihenfolge hier egal, also:

- bisher: jedes Set von 4 MZ-Tassen ist in verschiedenen Reihenfolgen in der Menge der Möglichkeiten
- Möglichkeiten, vier Tassen zu ordnen:
 $4 \cdot 3 \cdot 2 \cdot 1 = 24$
- Reihenfolge hier egal, also:

$$\text{Anzahl aller potentiellen Zuweisungen} = \frac{1680}{24} = 70 \quad (2)$$

Wenn Dame X also genau richtig liegt

- per Zufall genau richtig:
in einem von 70 Fällen, $p = 0.014$

Wenn Dame X also genau richtig liegt

- per Zufall genau richtig:
in einem von 70 Fällen, $p = 0.014$
- α -Niveau von 0.05 also erreicht

- eigentlich Binomialkoeffizient

- eigentlich Binomialkoeffizient
- „Lotto-Kombinationen“: k aus n
ohne Zurücklegen und ohne Beachtung der Reihenfolge

- eigentlich Binomialkoeffizient
- „Lotto-Kombinationen“: k aus n
ohne Zurücklegen und ohne Beachtung der Reihenfolge

- eigentlich Binomialkoeffizient
- „Lotto-Kombinationen“: k aus n
ohne Zurücklegen und ohne Beachtung der Reihenfolge

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (3)$$

- die drei richtigen: $\binom{4}{3}$

- die drei richtigen: $\binom{4}{3}$
- die eine falsche aus vier TZ: $\binom{4}{1}$

- die drei richtigen: $\binom{4}{3}$
- die eine falsche aus vier TZ: $\binom{4}{1}$
- Anzahl der Möglichkeiten drei richtige aus vier MZ
und dann eine falsche aus vier TZ zu ziehen: $\binom{4}{3} \cdot \binom{4}{1} = 4 \cdot 4 = 16$

- die drei richtigen: $\binom{4}{3}$
- die eine falsche aus vier TZ: $\binom{4}{1}$
- Anzahl der Möglichkeiten drei richtige aus vier MZ
und dann eine falsche aus vier TZ zu ziehen: $\binom{4}{3} \cdot \binom{4}{1} = 4 \cdot 4 = 16$

- die drei richtigen: $\binom{4}{3}$
- die eine falsche aus vier TZ: $\binom{4}{1}$
- Anzahl der Möglichkeiten drei richtige aus vier MZ und dann eine falsche aus vier TZ zu ziehen: $\binom{4}{3} \cdot \binom{4}{1} = 4 \cdot 4 = 16$

$$P(\text{drei richtig per Zufall}) = \frac{16}{70} = 0.229 \quad (4)$$

- Bei $\alpha = 0.05$ reichen also drei richtige nicht im Ansatz!

- Bei $\alpha = 0.05$ reichen also drei richtige nicht im Ansatz!
- alle schlechteren Ergebnisse: folglich auch nicht ausreichend

- Bei $\alpha = 0.05$ reichen also drei richtige nicht im Ansatz!
- alle schlechteren Ergebnisse: folglich auch nicht ausreichend
- Und bei 30 von 40 richtigen (also insgesamt 80 Tassen)?

Nächste Woche | Überblick

- 1 Statistik, Inferenz und probabilistische Grammatik
- 2 Deskriptive Statistik
- 3 Nichtparametrische Verfahren
- 4 z-Test und t-Test
- 5 ANOVA
- 6 Freiheitsgrade und Effektstärken
- 7 Power
- 8 Lineare Modelle
- 9 Generalisierte Lineare Modelle
- 10 Gemischte Modelle

- Cook, Philippa & Felix Bildhauer. 2013. Identifying “aboutness topics”: two annotation experiments. *Dialogue and Discourse* 4(2), 118–141.
- Gravetter, Frederick J. & Larry B. Wallnau. 2007. *Statistics for the Behavioral Sciences*. 7. Aufl. Belmont: Thomson.
- Maxwell, Scott E. & Harold D. Delaney. 2004. *Designing experiments and analyzing data: a model comparison perspective*. Mahwa, New Jersey, London: Taylor & Francis.
- Ronneberger-Sibold, Elke. 2010. Der Numerus – das Genus – die Klammer : die Entstehung der deutschen Nominalklammer im innergermanischen Vergleich. In Antje Dammel, Sebastian Kürschner & Damaris Nübling (Hrsg.), *Kontrastive Germanistische Linguistik. Teilband 2*, Bd. 206/209 (Germanistische Linguistik), 719–748. Hildesheim: Olms.

Kontakt

Prof. Dr. Roland Schäfer
Institut für Germanistische Sprachwissenschaft
Friedrich-Schiller-Universität Jena
Fürstengraben 30
07743 Jena

<https://rolandschaefer.net>
roland.schaefer@uni-jena.de

Creative Commons BY-SA-3.0-DE

Dieses Werk ist unter einer Creative Commons Lizenz vom Typ *Namensnennung - Weitergabe unter gleichen Bedingungen 3.0 Deutschland* zugänglich. Um eine Kopie dieser Lizenz einzusehen, konsultieren Sie

<http://creativecommons.org/licenses/by-sa/3.0/de/> oder wenden Sie sich brieflich an Creative Commons, Postfach 1866, Mountain View, California, 94042, USA.