Roland Schäfer. 2019. *Statistische Inferenz in der Linguistik* (Textbooks in Language Sciences 99). Berlin: Language Science Press.

This title can be downloaded at:
http://langsci-press.org/catalog
© 2019, Roland Schäfer
Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):
http://creativecommons.org/licenses/by/4.0/
ISBN: no digital ISBN
no hardcover ISBN
no softcover ISBN
ISSN: 2364-6209
no DOI

Cover and concept of design: Ulrike Harbort
Typesetting: Roland Schäfer
Fonts: Linux Libertine, Arimo, DejaVu Sans Mono
Typesetting software: XƎLaTeX

Language Science Press
Unter den Linden 6
10099 Berlin, Germany
langsci-press.org

Storage and cataloguing done by FU Berlin

no logo

Language Science Press has no responsibility for the persistence or accuracy of URLs for external or third-party Internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Dedicated to a randomly chosen individual.

Inhaltsverzeichnis

Vorwort

Danksagungen

Abbreviations and symbols

Abbreviations

ANOVA	analysis of variance
CDF	cumulative distribution function
CLT	central limit theorem
cp.	ceteris paribus (all other things being equal)
iid.	independent and identically distributed
LM	linear model
LMM	linear mixed model
GLM	linear mixed model
GLMM	generalised linear mixed model
PDF	probability density function
VCOV	variance-covariance matrix

Symbols

Symbols are overloaded ad-hoc to denote either a (possibly indexed) value such as $s_x = 1$ (for "the population mean of variable x is 1") or a function such as $s(x) = 1$ where applicable.

		Mathematische Symbole
$x \sim \mathbf{D}$		Zufallsvariable x *folgt* der Zufallsverteilung **D**
$\bar{x}$	$mean(x)$	arithmetisches Stichprobenmittel von x
$\tilde{x}$	$med(x)$	Stichprobenmedian von x
$\hat{x}$		Vorhergesagter Wert für x

	Symbole in Buchstabenform
α	Alphaniveau (Neyman-Pearson, NHST)
α_i	Intercept i
β	Betaniveau (Neyman-Pearson)
β_i	Koeffizient i der ersten Ebene (hierarchische Modelle)

df		Freiheitsgrade (*degrees of freedom*)
e		Euler-Konstante
ϵ		Fehler/Residuum auf Beobachtungsebene (Modelle)
f		Frequenz
F		F-Quotien (siehe ANOVA)
γ_i		Koeffizient i der zweiten Ebene (hierarchische Modelle)
	$cov(x, y)$	Kovarianz von x und y
H		Kruskal-Wallis-Statistik
H_0		Nullhypothese (Fisher)
H_A		Alternativhypothese (Neyman-Pearson)
M_M		Haupthypothese (Neyman-Pearson)
IQR		Interquartilsabstand
$\mathscr{L}$		Likelihood
μ		Populationsmittel
μ_i		Mittel des modellierten Effekts i (Modelle)
n		Stichprobengröße
N		Populationsgröße
O		Chance (*odds*)
p		Anteilswert (*proportion*)
P_i		i-tes Perzentil
Pr		Wahrscheinlichkeit (*probability*)
φ		Dispersionsparameter
Q_i		i-tes Quartil
r		Stichproben-Kovariations-Koeffizient
r^2		Bestimmtheitsmaß (*coefficient of determination*)
R^2		multifaktorielles Bestimmtheitsmaß
ρ		Grundgesamtheits-Kovariations-Koeffizient
s_x	$sd(x)$	Standardabweichung der Stichprobe x (*standard deviation*)
s_x^2	$var(x)$	Varianz der Stichprobe x
SE_{n_x,π_x}	$se(n_x, \pi_x)$	Standardfehler (*standard error*) für n_x und π_x
$SQ_{x,y}$	$sq(x, y)$	Quadratsumme (*sum of squares*) für Stichproben x und y
$SP_{x,y}$	$sp(x, y)$	Porduktsumme (*sum of products*) für Stichproben x und y
σ	$sd(X)$	Standardabweichung der Population
σ^2		Varianz der Population
sig		Signifikanzniveau (Fisher)
U		Mann-Whitney-Statistik
$VCOV_m$	$vcov(m)$	Varianz-Kovarianz-Matrix von m
χ^2		Chi-Quadrat-Statistik

Zuvallsverteilungen werden hier durch fettgedruckte Buchtsbaben gekennzeichnet, nicht durch den manchmal üblichen Skript-Font. Die kumulative Verteilung wird mit Strich gegeben, also **Norm**′.

Bern	$\mathscr{B}$	Bernoulliverteilung
Exp	$\mathscr{E}$	Exponentialverteilung
F	$\mathscr{F}$	F-Verteilung
Norm	$\mathscr{N}$	Normal-/Gauss-Verteilung
t	$\mathscr{T}$	t-Verteilung
Unif	$\mathscr{U}$	uniforme Verteilung
Chisq	$\mathscr{X}^2$	χ^2-Verteilung

Teil I

Wissenschaftliche Inferenz in der Linguistik

Teil II

Systeme der statistischen Inferenz

1 Ronald Fishers System der statistischen Inferenz

1.1 Randomsierung

1.2 The Lady Tasting Tea

1.3 Die Rolle der Nullhypothese

Teil III

Datenanalyse

Teil IV

Statistische Tests

2 <++>

3 Mittelwervergleiche

3.1 Beispiele in diesem Kapitel und typische Fragestellungen

Dieses Kapitel für in mögliche Inferenzen über Mittelwerte ein. Mittelwerte sind für Daten definiert, die numerisch skaliert sind, also unsere typischen Beispiele wie:

- Wortlängen,
- Satzlängen,
- numerisch gemessene Werte auf Dokumentebene,[1]
- Reaktionszeiten wie z. B. Lesezeiten,
- Scores aus Magnitude Estimation-Experimenten.

Oft interessiert uns, ob sich die Zentraltendenzen (als Abhängige) in verschiedenen Gruppen unterschiedlich sind, wobei die Gruppen durch die Unabhänigge gegeben ist. Beispiele wären z. B. Unterschieden der Mittel von:

- Wortlängen von zwei verschiedenen Autoren,
- Nomen-Verb-Quotienten in Texten aus verschiedenen Genres,
- Lesezeiten unter verschiedenen syntaktischen Extraktionsbedingungen,
- ME-Bewertungen von Standard- und Ninctstandard-Varianten.

Eventuell interessiert auch der Vergleich solcher Messungen mit einer quasi als bekannt angenommenen Population. Liegt zum Beispiel ein sehr großes Korpus des Gegenwartsdeutschen vor, aus dem die mittlere Wortlänge für ein Genre mit hoher Konfidenz auf Populationsebene geschätzt werden kann, wollen wir evtl. die mittlere Wortlänge aus einer kleineren Sammlung von Texten mit dem als bekannt angenommenen Populationsmittel vergleichen. Diese Texte können auch durchaus in einem leidlichen gut kontrollierten Szenario entstanden sein, z. B. als Aufsätze einer schulischen Lerner*innen-Gruppe.

[1] Wenn diese Werte aber als Anteilwerte bzw. Prozentwerte erfasst wurden, sind die hier vorgestellten Tests nicht anwendbar. Warum, wird in diesem und den folgenden Kapiteln deutlich.

In den beiden genannten Fällen sollte nun, wenn wir eine statistischen Inferenz anstreben und einen Test durchführen wollen, eine substantielle theoretische Hypothese vorliegen, gemäß derer es im Mittel einen tatsächlichen Unterschied zwischen mindestens zwei Gruppen oder der Grundgesamtheit und einer Gruppe gibt. Die Gruppen werden definiert durch eine kategoriale Unabhängige. Nehmen wir die Lesezeiten als Beispiel, dann wäre die substantielle Hypothese wahrscheinlich die aus einem bestimmten Processing-Modell abgeleitete psycholinguistische Vermutung, dass ein bestimmter Extraktionstyp mit mentalen Operationen verbunden ist, die zu einer Längung der Lesezeit führen (z. B. Surprisal).

TODO Erkläre, dass es eigentlich immer um GG-Mittel geht!

Die substantielle Hypothese mündet stets die inferenzlogische Form wie in Gleichung 3.1. $\mathscr{H}$ ist definiert als: „Es gibt einen Unterschied zwischen den Mitteln der zwei Populationen X_1 und X_2."

$$\mathscr{H} := \bar{X_1} \neq \bar{X_2} \tag{3.1}$$

Hypothese $\mathscr{H}$ in Gleichung 3.1 ist grundlegend verschieden von $\mathscr{H}*$ in Gleichung 3.2. $\mathscr{H}*$ wird hier definiert als: „Es gibt einen Mittelwertunterschied zwischen Stichprobe x_1 und Stichprobe x_2."

$$\mathscr{H}* := \bar{x_1} \neq \bar{x_2} \tag{3.2}$$

Diese Hypothese ist, wie in Teil I ausführlich argumentiert wurde, trivialerweise fast immer richtig. Außerdem ist sie *beweisbar*, indem wir die Mittelwerte der Stichproben bilden und vergleichen. Das macht sie wissenschaftlich völlig *uninteressant*. Die angestrebte Inferenz zielt immer auf einen grundlegenden datengenerierenden Prozess, nicht auf einzelne Stichproben.

Die korrekte inferenzlogische Form der substantiellen Hypothese in Gleichung 3.1 ist aber nicht mit der eigentlichen Hypothese zu verwechseln. Die (in der Regel in Worten oder in einem Modellformalismus) zu formulierende Theorie $\mathscr{T}$ *bettet* $\mathscr{H}$ *kausal ein*, beschreibt also einen Mechanismus, der dazu führt, dass die Reaktionszeiten oder die Wortlängen usw. in den beobachteten Gruppen verschieden sind. Es muss also gelten, dass $\mathscr{T} \hookleftarrow \mathscr{H}$.

3.2 Verteilungen von Stichprobenmitteln und die zentralen Grenzwertsätze

3.3 Stichproben- und Populationsmittel

3.4 Stichproben- und Stichprobenmittel

3.5 Messwiederholungen

3.6 Simulationen zu t-Tests

3.7 Voraussetzungen für t-Tests testen

3.8 Zur Absurdität von t-Tests in der Korpuslinguistik

3.9 Power von t-Tests

3.10 Interpretation von t-Tests in N-P

Teil V

Modelle

Fisher-Yates Test|see Fisher Exact Test