

Datenmanagement & -analyse

Übung 8 – Statistische Inferenz und lineare Regressionen

Dr. Nikolai Stein

Lehrstuhl für WI & BA

Julius-Maximilians-Universität Würzburg

Sommersemester 2021

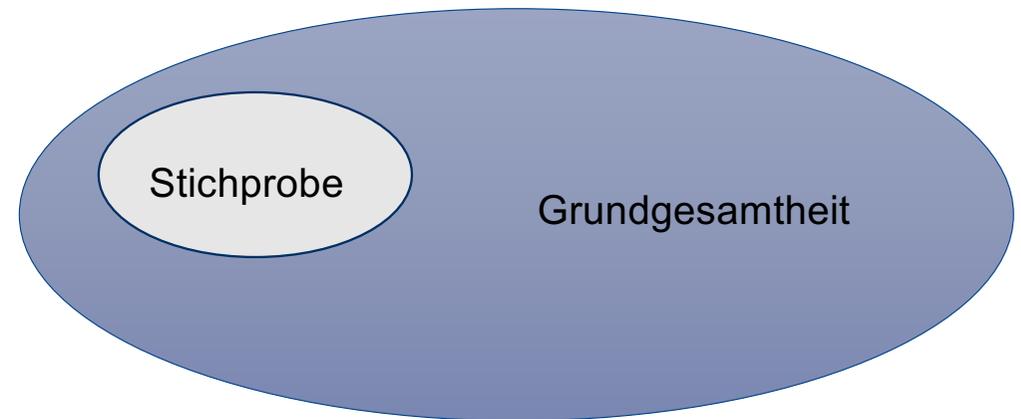


- Wann immer wir es mit Zufallsgrößen zu tun haben, müssen wir eine Stichprobe ziehen, um Rückschlüsse auf die zugrunde liegende Verteilung zu ziehen
- Die große Frage: Wie erhält man ein Beurteilungssicherheitsniveau bezüglich dieser Größen?
- Anwendungsbeispiele:
 - Wie können wir nach einer medizinischen Studie die Ergebnisse interpretieren, um die Wirksamkeit einer Behandlung zu erforschen?
 - Wie kann man die Ergebnisse von zwei aufeinanderfolgenden Vorlesungsevaluationen vergleichen?

Statistische Inferenz / Hypothesentest

Warum Hypothesentests?

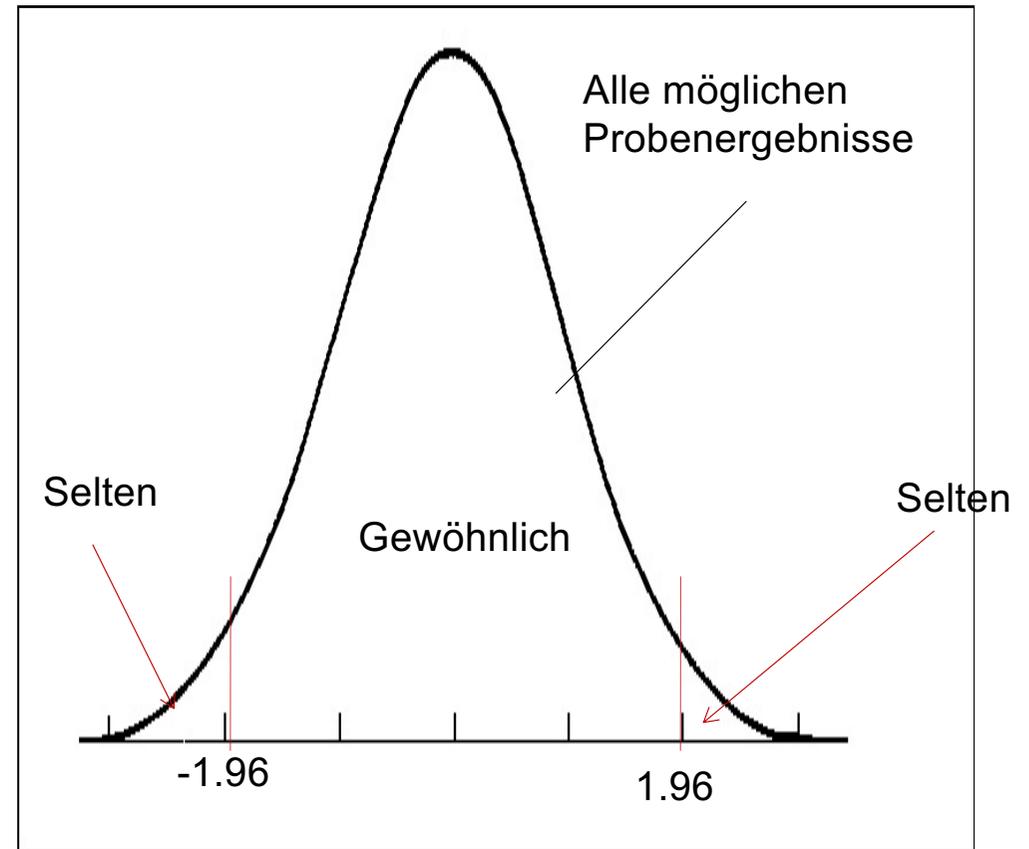
- Der erste Schritt in jeder Studie ist der Test gegen den Zufall.
- Wir können keine Schlussfolgerungen über unsere Ergebnisse ziehen, ohne sicherzustellen, dass unsere Ergebnisse nicht zufällig sind.
- Wir wissen nie genau, wie die wahre Situation ist, aber wir versuchen, die Möglichkeit eines Fehlers zu minimieren.



Was können wir aus einer Stichprobe über die Grundgesamtheit schließen?

Häufige vs. seltene Ergebnisse

- Wir beginnen immer mit der Annahme, dass die Nullpopulation wahr ist. In diesem Fall fragen wir:
 - Wie wahrscheinlich wäre unser aktuelles Ergebnis, wenn die Nullpopulation die wahre Grundgesamtheit wäre?
- Ein Ergebnis in der Mitte der Normalkurve ist sehr wahrscheinlich (**gewöhnlich**).
- Ein Ergebnis in den Schwänzen der Normalkurve ist viel unwahrscheinlicher (**selten**).



Wir verwenden $\pm 1,96$ (2 Standardabweichungen) als Grenze zwischen häufig und selten.

Wir laufen immer Gefahr, eine falsche Schlussfolgerung zu ziehen:

1. H_0 ist wahr und der Test verwirft H_0 korrekterweise nicht
2. H_0 ist falsch und der Test verwirft H_0 korrekterweise
3. H_0 ist wahr aber Test verwirft H_0 fälschlicherweise
4. H_0 ist falsch aber der Test verwirft H_0 fälschlicherweise nicht

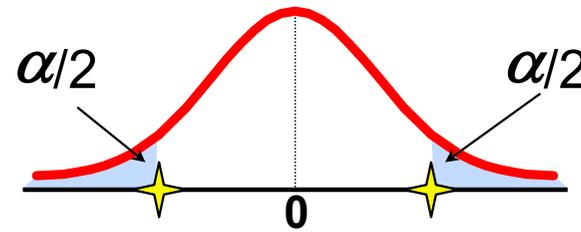
- Das Ergebnis 3 wird als Fehler vom Typ I bezeichnet (α).
- Das Ergebnis 4 wird als Fehler vom Typ II bezeichnet (β).
- Typischerweise sind wir am meisten über Fehler vom Typ I besorgt:
 - Unschuldige verurteilte Person
 - Unwirksame Behandlung zugelassen
 - Kranke Person gilt als gesund

Ablehnungsbereich zum Signifikanzniveau α

Zweiseitiger Test

$$H_0: \mu = 3$$

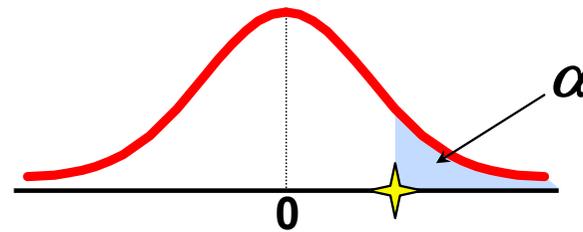
$$H_1: \mu \neq 3$$



Einseitiger Test

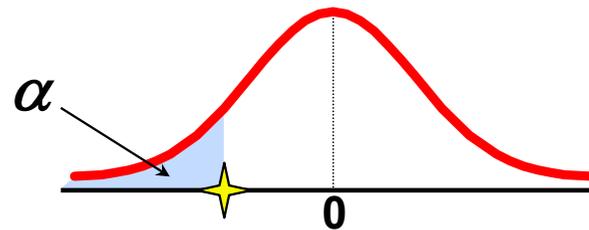
$$H_0: \mu \leq 3$$

$$H_1: \mu > 3$$



$$H_0: \mu \geq 3$$

$$H_1: \mu < 3$$



★ Grenzwert

Ablehnungsbereich

Schritte des Hypothesentestverfahrens

1. Identifizieren Sie den Populationsparameter und formulieren Sie die zu prüfenden Hypothesen.
 2. Wählen Sie ein *Signifikanzniveau* (Risiko, eine falsche Schlussfolgerung zu ziehen).
 3. Bestimmen Sie eine Entscheidungsregel, auf die Sie eine Schlussfolgerung stützen können.
-
4. Daten sammeln und Teststatistik berechnen.
 5. Wenden Sie die Entscheidungsregel an und ziehen Sie eine Schlussfolgerung.

Beispiel für eine einfache lineare Regression

- Beispiel: Vorhersage des Verkaufspreises eines Hauses (abhängige Variable y) basierend auf der Grundstücksgröße (unabhängige Variable x)
 - Verkaufspreis und Größe von 21613 kürzlich verkauften Häusern sind verfügbar
- Identifizieren Sie die erklärendste (in Bezug auf die Beispieldaten) Funktion der Form
- Die anfängliche Annahme eines linearen Zusammenhangs kann mit Hilfe eines Streudiagramms überprüft werden

