

Statistik

Vorlesung 0 - Organisation und Einführung

Prof. Dr. Sandra Eisenreich

18. März 2024

Hochschule Landshut

Organisation

- **Moodle Kurs:**

- Link: <https://moodle.haw-landshut.de/course/view.php?id=11040>
- Passwort: Stat_SoSe24

- **Vorlesung:** Es wird zwei Skripte geben:

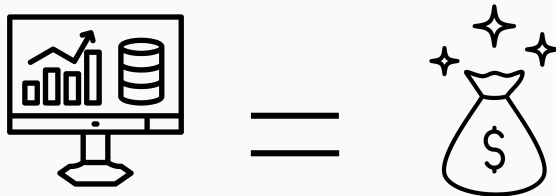
- die Vorlesungspräsentationen ohne Rechnungen, die wir uns in der Vorlesung zusammen erarbeiten
 - ein vollständiges Skript inclusive der Lösungen, damit Sie nicht mitschreiben müssen, sondern mitdenken können.
- Die handschriftlichen Vorlesungsnotizen (also die Rechnungen und Erklärungen) werden ebenfalls auf Moodle bereitgestellt.
 - Das Skript ist eine Abwandlung von vergangenen Vorlesungen von Prof. Dr. Peter Hartmann und Prof. Dr. Eduard Kromer.

- **Praktika:** 14-tägig je 1,5 h; Räume und Termine siehe Moodle
- **Übungsblätter:** 6 Übungsblätter zu den unterschiedlichen Themen
- **Erwartungshaltung:** vor der Übung *selbstständig* bearbeiten!

- **Dauer und Umfang:** 60 Minuten für 60 Punkte
- **relevanter Stoff:** das gesamte Skript
- **zulässige Hilfsmittel:**
 - nicht-programmierbarer Taschenrechner
 - Wertetabelle für die χ^2 - und Normalverteilungen
 - 1 selbst beschriebenes DIN A4 Blatt
- **Aufbau:**
 - ca 20% Multiple Choice Fragen zur Theorie aus der Vorlesung (1 Punkt für richtige Antwort, 0 Punkte für falsche oder fehlende Antwort)
 - ca. 80% Rechenaufgaben, die sich an den Übungsblättern orientieren
- **Bestehenskriterium:** mit 50% der Punkte plus die Hälfte der Punkte im Multiple Choice Teil (Erwartungswert bei zufälligem Ankreuzen) haben Sie auf jeden Fall bestanden.

Was ist Statistik?

Motivation: Warum Statistik?



Jede Datenerhebung ist eine Stichprobe. In Stichproben steckt immer der Zufall.

⇒ Statistik

“Da reicht doch gesunder Menschenverstand”? Nein!

Verteilungen

Der Zufall (und damit Stichproben) folgen Regeln. Meist haben die erhobenen Daten gewisse Strukturen, genannt "Verteilungen".

Beispiele:

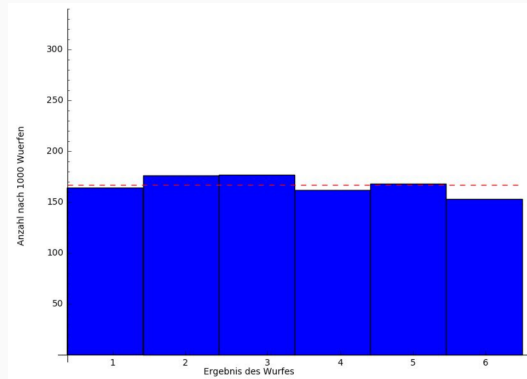


Abbildung 1: Häufigkeitsverteilung nach 1000 Würfeln mit einem Würfel.

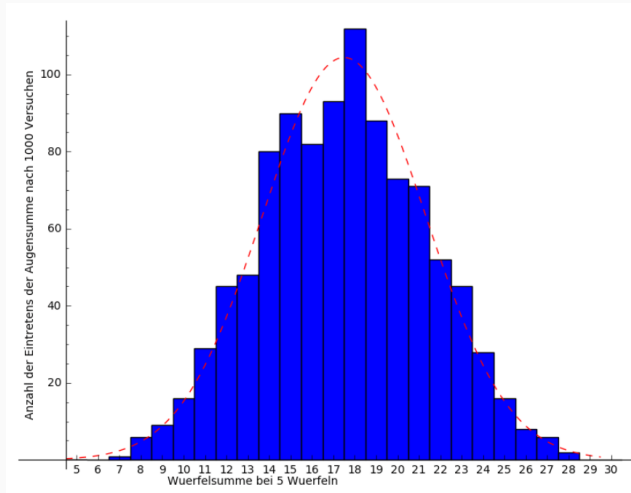


Abbildung 2: Würfeln mit 5 Würfeln. Durchschnittliche Würfelsumme nach 1000 Versuchen.

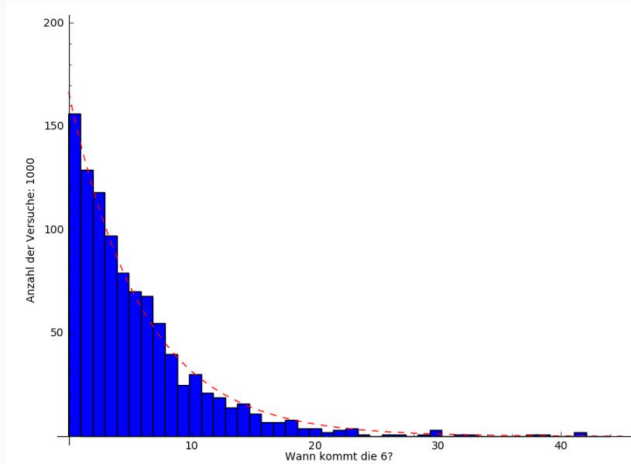
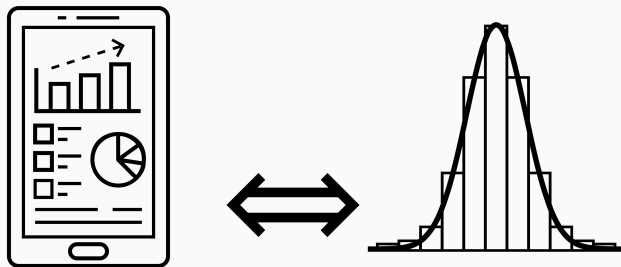


Abbildung 3: Mensch-Ärgere-dich-nicht: Warten auf den Sechser



Daten analysieren = Verstehen welche statistische Verteilung hinter den Daten steckt.

Beispiel: Wie generiert ChatGPT Text?

“Die Maus frisst den ... ???”

ChatGPT lernt aus viel Text, wie **wahrscheinlich** jedes Wort nach einem gewissen Kontext kommt, also die **Wahrscheinlichkeitsverteilung** von Sprache.

⇒ Künstliche Intelligenz \simeq angewandte Statistik (Lernt eine Verteilung aus Daten)

Was kann man mit vorhandenen Daten machen?

- Vorhersagen treffen (Welches Produkt wird am besten ankommen?)
- Konfidenz einer Vorhersage einschätzen (wie sicher kann ich mir da sein, bevor ich alles Geld in die Entwicklung von obigem Produkt stecke?)
- Werte abschätzen, die man nicht exakt bestimmen kann (entstehende Kosten durch fehlerhafte Produkte)
- Daten interpretieren (Coronatest positiv - wie wahrscheinlich bin ich krank?)

- Marktforschung
- Zensus / demographische Erhebung
- Finanzanalysen
- Vorhersagen (z.B. mit KI)
- betriebswirtschaftliche Kennzahlen
- Prozess- und Qualitätskontrollen
- Naturwissenschaft (Statistiken zur Überprüfung von Gesetzmäßigkeiten)
- ...

Beispiel für die Vorlesung: Qualitätskontrolle in der Produktion

- bei der industriellen Fertigung eines Teiles können nicht alle Teile überprüft werden (zu teuer oder zerstörende Prüfung)
- in regelmäßigen Abständen werden Stichproben genommen
- die Anzahl der defekten Teile wird festgehalten
- übersteigt die Zahl einen Schwellenwert, muss in die Produktion eingegriffen werden

Man will zu einer Aussage wie der Folgenden:

"Mit 99% Wahrscheinlichkeit sind weniger als 0.01% der Teile fehlerhaft."

Was ist Statistik?

Duden 2022:

1. Wissenschaft von der zahlenmäßigen Erfassung, Untersuchung und Auswertung von Massenerscheinungen
2. schriftlich fixierte Zusammenstellung, Aufstellung der Ergebnisse von Massenuntersuchungen, meist in Form von Tabellen oder grafischen Darstellungen

Wikipedia 2022:

Statistik „ist die Lehre von Methoden zum Umgang mit quantitativen Informationen“ (Daten). Sie ist eine Möglichkeit, „eine systematische Verbindung zwischen Erfahrung (Empirie) und Theorie herzustellen“. Unter Statistik versteht man die Zusammenfassung bestimmter Methoden zur Analyse empirischer Daten. [...]

Der Begriff **Statistik** wird synonym

- im Sinne einer Datensammlung
- im Sinne einer Kennzahl
- als Aktivität der Datenerhebung
- als wissenschaftliche Disziplin

verwendet.

- Hartmann, Peter; Mathematik für Informatiker, Springer-Vieweg; 7. Auflage; 2019
- Georgii, Hans-Otto; Stochastik; de Gruyter, 5. Auflage; 2015
- Krengel, Ulrich; Einführung in die Wahrscheinlichkeitstheorie und Statistik; Springer-Vieweg; 8. Auflage; 2005
- Henze, Norbert; Stochastik für Einsteiger; Springer; 10. Auflage; 2013
- Meintrup, David; Schäfer, Stefan; Stochastik; Springer; 1. Auflage; 2005
- Behrends, Ehrhard; Elementare Stochastik; Springer-Vieweg; 2013