

# Abgabe 1 für Computergestützte Methoden

Gruppe 43, Lea Brand, Lara Wagner

01.12.2024

## Inhaltsverzeichnis

<b>1</b>	<b>Der zentrale Grenzwertsatz</b>	<b>2</b>
1.1	Aussage . . . . .	2
1.2	Erklärung der Standardisierung . . . . .	2
1.3	Anwendungen . . . . .	2
<b>2</b>	<b>Bearbeitung zur Aufgabe 1</b>	<b>3</b>
2.1	Berechnung der höchsten mittleren Temperatur . . . . .	3
2.2	Datenbankschema entwerfen . . . . .	3
2.3	SQL und CSV . . . . .	3
2.4	Formulierung der SQL-Abfrage, um höchste mittlere Temperatur zu ermitteln . . . . .	4

# 1 Der zentrale Grenzwertsatz

Der zentrale Grenzwertsatz (ZGS) ist ein fundamentales Resultat der Wahrscheinlichkeitstheorie, das die Verteilung von Summen unabhängiger, identisch verteilter (*i.i.d.*) Zufallsvariablen (ZV) beschreibt. Er besagt, dass unter bestimmten Voraussetzungen die Summe einer großen Anzahl solcher ZV annähernd normalverteilt ist, unabhängig von der Verteilung der einzelnen ZV. Dies ist besonders nützlich, da die Normalverteilung gut untersucht und mathematisch handhabbar ist.

## 1.1 Aussage

Sei  $X_1, X_2, \dots, X_n$  eine Folge von *i.i.d.* ZV mit dem Erwartungswert  $\mu = \mathbb{E}(X_i)$  und der Varianz  $\sigma^2 = \text{Var}(X_i)$ , wobei  $0 < \sigma^2 < \infty$  gelte. Dann konvergiert die standardisierte Summe  $Z_n$  dieser ZV für  $n \rightarrow \infty$  in Verteilung gegen eine Standardnormalverteilung:<sup>1</sup>

$$Z_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \quad (1)$$

Das bedeutet, dass für große  $n$  die Summe der ZV näherungsweise normalverteilt ist mit Erwartungswert  $n\mu$  und Varianz  $n\sigma^2$ :

$$\sum_{i=1}^n X_i \sim \mathcal{N}(n\mu, n\sigma^2). \quad (2)$$

## 1.2 Erklärung der Standardisierung

Um die Summe der ZV in eine Standardnormalverteilung zu transformieren, subtrahiert man den Erwartungswert  $n\mu$  und teilt durch die Standardabweichung  $\sigma\sqrt{n}$ . Dies führt zu der obigen Formel (1). Die Darstellung (2) ist für  $n \rightarrow \infty$  nicht wohldefiniert.

## 1.3 Anwendungen

Der ZGS wird in vielen Bereichen der Statistik und der Wahrscheinlichkeitstheorie angewendet. Typische Beispiele sind:

- *Hypothesentests*
- *Stichprobenverteilung des Mittelwerts*

---

<sup>1</sup>Der zentrale Grenzwertsatz hat verschiedene Verallgemeinerungen. Eine davon ist der **Lindeberg-Feller-Zentrale-Grenzwertsatz** [[1], Seite 328], der schwächere Bedingungen an die Unabhängigkeit und die identische Verteilung der ZV stellt.

## 2 Bearbeitung zur Aufgabe 1

### 2.1 Berechnung der höchsten mittleren Temperatur

Wir haben alle Datensätze von Gruppe 43 in eine neue Tabelle eingefügt, wobei uns aufgefallen ist, dass sich alle Daten eines Datums jeweils in einer Spalte befanden. Wir haben alle Zeilen ausgewählt und durch "Teilen in Spalten" die Daten, nach Kommas, in verschiedene Spalten aufteilen können. Dabei ist uns aufgefallen, dass viele Daten nicht vorhanden (NA), falsch (Windspeed=-1) oder in der falschen Spalte waren (Windspeed=Jun 49). Wir haben dann am Ende der Spalte *average temperature* das Summenzeichen verwendet und dort *max* ausgewählt und so den Wert 28,3°C Grad herausbekommen, als höchste mittlere Temperatur.

### 2.2 Datenbankschema entwerfen

1. Normalform: Durch 2.1. enthält jede Spalte einen atomaren Wert. 2. Normalform: Aufteilung in mehrere Tabellen abhängig vom Primärschlüssel.

Wir haben in 3 Tabellen aufgeteilt, in "group", "date" und "values". In der Tabelle "group" haben wir die Spalten "group id" und "station", in "date id" haben wir die Spalten "date", "day of year", "day of week" und "month of year" und in "values" haben wir "group id", "date id", "precipitation", "windspeed", "min temperature", "average temperature", "max temperature" und "count".

### 2.3 SQL und CSV

Umsetzung des Schemas in SQL (DDL):

```
CREATE TABLE "group43" (  
  id INTEGER NOT NULL UNIQUE,  
  name TEXT NOT NULL UNIQUE,  
  PRIMARY KEY("id")  
);
```

```
CREATE TABLE "date43" (  
  id TEXT NOT NULL UNIQUE,  
  day_of_year INTEGER NOT NULL,  
  day_of_week INTEGER NOT NULL,  
  month_of_year INTEGER NOT NULL,  
  PRIMARY KEY("id")  
);
```

```
CREATE TABLE "values43" (  
  group_id TEXT NOT NULL,  
  date_id TEXT NOT NULL,  
  precipitation NUMERIC,  
  windspeed NUMERIC,
```

```

min_temperature NUMERIC,
average_temperature NUMERIC,
max_temperature NUMERIC,
count INTEGER,
PRIMARY KEY("date_id","group_id"),
FOREIGN KEY(date_id) REFERENCES date43(id),
FOREIGN KEY(group_id) REFERENCES group43(id)
);

```

Import der zugeordneten Daten als CSV:

In der Excel-Datei haben wir unsere Gruppen-Tabelle abgespeichert als CSV-Datei. Dann haben wir den DB Browser für SQLite benutzt und die CSV-Datei importiert, indem wir im Menü Importieren ausgewählt haben und unsere vorbereitete Datei importiert. Wir haben den Feld-Separator "," ausgewählt, String-Zeichen „ und die Codierung UTF-8".

## 2.4 Formulierung der SQL-Abfrage, um höchste mittlere Temperatur zu ermitteln

```

SELECT
max(average_temperature) as max_wert
FROM values43
WHERE average_temperature not in ('NA')
;

```

## Literatur

- [1] Achim Klenke. *Wahrscheinlichkeitstheorie*. Springer, 3. edition, 2013.