

# Abgabe 1 für Computergestützte Methoden

Gruppe 99, Jakob Bohn, Nisa-Nur Civi, Darleen

02.12.2024

## 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	Datenverarbeitung . . . . .	3
2.1.1	Excel-Formel zur Abfrage von Maximalwerten . . . . .	3
2.2	Datenhaltung . . . . .	4
2.2.1	Tabellenentwurf . . . . .	4
2.2.2	SQL-Befehle zur Tabellenerstellung . . . . .	4
2.2.3	Importieren der Daten in SQLite . . . . .	5
2.2.4	SQL-Abfrage zur maximalen Temperatur . . . . .	5

# 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  $0 < \sigma^2 < \infty$  gilt. Dann konvergiert die standardisierte Summe  $Z_n$  dieser ZV für  $n \rightarrow \infty$  in Verteilung gegen eine Standardnormalverteilung:<sup>1</sup>

$$Z_n = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, 1) \quad (1)$$

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:

- Schätzen von Populationsparametern in der Statistik.
- Hypothesentests in der medizinischen Forschung.

---

<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 Datenverarbeitung

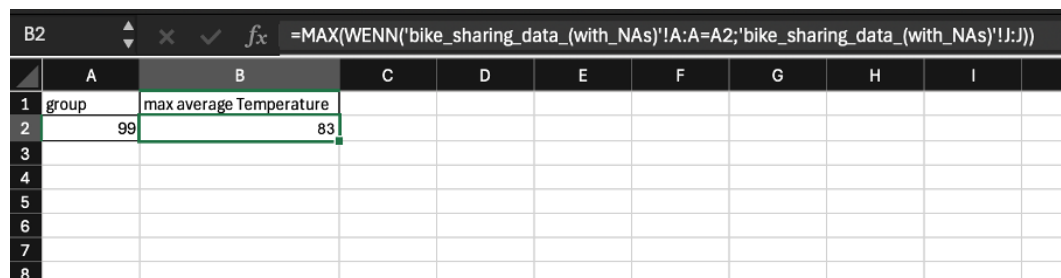
Der Datensatz für unsere Gruppe enthält verschiedene Messwerte, die an der Fahrradverleihstation „West St & Liberty St“ an jedem Tag im Jahr gemessen wurden. Dabei gibt es zu jedem Tag im Jahr Messwerte zu Temperatur, Niederschlag und Windgeschwindigkeit.

An den Daten kann man die Auslastung der Station zu verschiedenen Jahreszeiten oder zu bestimmten Wetterbedingungen vergleichen.

#### 2.1.1 Excel-Formel zur Abfrage von Maximalwerten

Die folgende Formel wird auf einem neuen Arbeitsblatt in Excel angewendet:

```
=MAX(WENN('bike_sharing_data_(with_NAs) '!A:A=A2,  
          'bike_sharing_data_(with_NAs) '!J:J))
```



	A	B	C	D	E	F	G	H	I
1	group	max average Temperature							
2	99	83							
3									
4									
5									
6									
7									
8									

Abbildung 1: Excel-Ansicht.

Diese Formel bezieht sich auf das Arbeitsblatt „bike\_sharing\_data\_(with\_NAs)“ und gibt das Maximum der Spalte J aus. Es werden nur Werte berücksichtigt, bei denen der Wert in Spalte A der entsprechenden Zeile gleich dem Wert in Zelle A2 des neuen Arbeitsblatts ist.

So können die maximalen Durchschnittstemperaturen für verschiedene Gruppen abgefragt werden. Gibt man bei A2 den Wert „99“ ein, wird geprüft: „Wenn in Spalte A der Wert 99 steht, dann wird der entsprechende Wert aus Spalte J in die MAX-Abfrage einbezogen.“

Das Maximum beträgt 83° für Gruppe 99.

Dieses Ergebnis ist in Grad Fahrenheit angegeben.

Um die Temperatur in Celsius umzuwandeln, kann folgende Formel angewendet werden:

$$\text{Celsius} = (\text{Fahrenheit} - 32) \times \frac{5}{9}$$

Nach der Umrechnung ergibt sich für die maximale Temperatur ein Wert von 28,3°C, gemessen an der Station „West St & Liberty St“.

C3 $\updownarrow$ $\times$ $\checkmark$ $f_x$ $= (B3 - 32) * 5 / 9$										
	A	B	C	D	E	F	G	H	I	
1		max average Temperature								
2	group	°Fahrenheit	°Celsius							
3	99	83,0 °F	28,3 °C							
4										
5										
6										
7										
8										

Abbildung 2: Excel-Ansicht 2.

## 2.2 Datenhaltung

### 2.2.1 Tabellenentwurf

Schemata für die Tabellen *Stationen*, *Verleih* und *Wetter*:

- **Stationen:** (Id#, Name)
- **Verleih:** (Id#, StationId#, Datum, Verleih\_anzahl)
- **Wetter:** (Id#, Datum, temperatur\_mittel(in °C), niederschlag, windgeschwindigkeit)

Die Spalte *StationId#* verknüpft die Tabellen *Stationen* und *Verleih* als Fremdschlüssel.

### 2.2.2 SQL-Befehle zur Tabellenerstellung

Die Tabellen können mit den folgenden SQL-Befehlen erstellt werden:

```
CREATE TABLE Stationen (
    id INTEGER PRIMARY KEY,
    name TEXT NOT NULL
);
```

```
CREATE TABLE Verleih (
    id INTEGER PRIMARY KEY,
    station_id INTEGER,
    datum TEXT NOT NULL,
    verleih_anzahl INTEGER NOT NULL,
    FOREIGN KEY (station_id) REFERENCES Stationen(id)
);
```

```
CREATE TABLE Wetter (
    id INTEGER PRIMARY KEY,
    datum TEXT NOT NULL,
    temperatur_mittel FLOAT,
    niederschlag FLOAT,
```

```
        windgeschwindigkeit FLOAT
    );
```

### 2.2.3 Importieren der Daten in SQLite

Um die Daten aus vorbereiteten CSV-Dateien in SQLite zu importieren, können folgende Befehle verwendet werden:

```
.mode csv

.import "/Users/jakobbohn/stationen.csv" Stationen
.import "/Users/jakobbohn/stationen.csv" Verleih
.import "/Users/jakobbohn/stationen.csv" Wetter
```

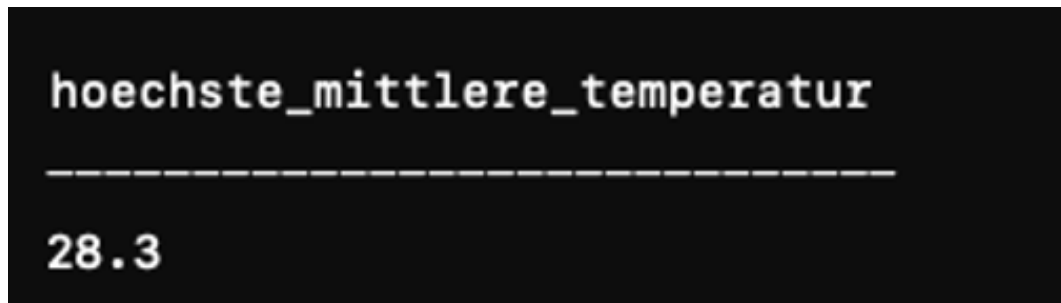
Der Dateipfad muss entsprechend angepasst werden. Je nach verwendetem CSV-Trennzeichen (Komma oder Semikolon) ist dieses ggf. zu ändern.

### 2.2.4 SQL-Abfrage zur maximalen Temperatur

Um die maximale mittlere Temperatur aus der Tabelle *Wetter* abzufragen, kann folgende SQL-Abfrage verwendet werden:

```
SELECT MAX(temperatur_mittel) AS hoechste_mittlere_temperatur
FROM Wetter;
```

Das Ergebnis gibt die höchste mittlere Temperatur im Datensatz aus.



```
hoechste_mittlere_temperatur
-----
28.3
```

Abbildung 3: Ergebnisse der Abfrage.

## Literatur

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