

Abgabe 1 für Computergestützte Methoden

Gruppe 5

Dida Abas: 4338708

Subidya Srirangavasan: 4338455

Melissa Ergün: 4353910

01.12.2024

Inhaltsverzeichnis

1	Der zentrale Grenzwertsatz	2
1.1	Aussage	2
1.2	Erklärung der Standardisierung	2
1.3	Anwendungen	2
2	Bearbeitung Aufgabe 1	3
2.1	Thema: Datenverarbeitung	3
2.1.1	Analyse der Daten und Import in Tabellenkalkulation Excel	3
2.1.2	Nummer 1: Berechnung der höchsten mittleren Temperatur	3
2.2	Thema: Datenhaltung	4
2.2.1	Datenbank-Schema entwerfen	4
2.2.2	Umsetzung des Schemas in SQL und Import der zugeordneten Daten als CSV	5
2.2.3	Bestimmung der höchsten mittleren Temperatur mit SQL	6

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 $\mu = \mathbb{E}(X_i)$ mit der Varianz $\sigma^2 = \text{Var}(X_i)$, wobei $0 < \sigma^2 < \infty$. Dann konvergiert die standardisierte Summe Z_n dieser ZV für $n \rightarrow \infty$ in Verteilung gegen eine Standardnormalverteilung:¹

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

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

- die Durchführung von Hypothesentests
- eine Approximation von Wahrscheinlichkeiten für Summen oder Mittelwerte

¹Der zentrale Grenzwertsatz hat verschiedene Verallgemeinerungen. Eine davon ist der Lindeberg-Feller-Zentrale-Grenzwertsatz [1, S.238], der schwächere Bedingungen an die Unabhängigkeit und die identische Verteilung der ZV stellt

2 Bearbeitung Aufgabe 1

2.1 Thema: Datenverarbeitung

2.1.1 Analyse der Daten und Import in Tabellenkalkulation Excel

Bei dem gegebenen Datensatz handelt es sich um eine Csv Datei. Das heisst, dass beim Öffnen der Datei die einzelnen Werte mit einem Komma getrennt sind. Die Daten veranschaulichen einen Fahrradverleih und die an dem Tag gegebenen Wetterbedingungen. Die Sichtung der Daten ergibt:

- Spalte 1 (group) : Die Daten sind nach Gruppennummern eingeteilt, d jenes bedeutet, dass jede Gruppe einen Datensatz zugeteilt bekommen hat.
- Spalte 2 (station): Dies ist Äquivalent zu der Gruppenzuteilung der Daten. Jede Gruppe hat einen Stationsnamen/eine Station zugeteilt bekommen. In der Gruppe 5 ist der Fokus bei der Station 10 Ave & W 14 St
- Spalten 3-6 (date, day_of_year, day_of_week, month_of_year) ordnen den Daten jeweils ein Tag im Jahr zu (mit Datum, Anzahl des Tages im Jahr, Wochentag (1=Montag, 7=Sonntag), Monat) jeden Tag, chronologisch vom 01.01.2023 bis zum 31.12.2023. Dementsprechend gibt es insgesamt 365 Messungen/Daten.
- Spalten 7-11 geben die Wetterbedingungen wieder. Angegeben ist der Niederschlag (precipitation), die Windgeschwindigkeit (windspeed), die minimale, durchschnittliche und maximale Temperatur. Für die Aufgabe ist nur die durchschnittliche/mittlere Temperatur relevant.
- Die letzte Spalte teilt die Anzahl der ausgeliehenen Fahrräder zu.

Die relevanten Spalten für die Bearbeitung der Aufgaben sind *group* (hier: group 5), *station*, *date* und *average temperature*. Der Import in eine Tabellenkalkulation ist für die Aufgabe essentiell. Dazu wird die Datei gespeichert und in der Tabellen Kalkulation *Excel* eingefügt: Daten->Daten abrufen->aus Datei.

2.1.2 Nummer 1: Berechnung der höchsten mittleren Temperatur

Um die mittlere Temperatur zu berechnen nutzt man den Befehl = *Max*([3], diesen Befehl gibt man in der Zelle unter dem letzten Wert in der Spalte J („average temperature“) (Zelle:J367), ein. Anschließend markiert man die gesamte Spalte, um alle Zahlenwerte aufzufassen und unter diesen die höchste Temperatur zu bestimmen. Man kann auch mit dem Befehl „ZelleA:ZelleB“, alle Zellen von Zelle A bis B, auffassen.

Befehl:

$$= MAX(J2 : J366)$$

Ergebnis:

$$83^{\circ}\text{F}$$

Da dieses Ergebnis in der Einheit Fahrenheit vermerkt ist, muss man dieses noch umrechnen mit Hilfe des Befehls: „=UMWANDELN(Zahl;Grösse1;Grösse2)“[4]. Dies wird erneut unter die letzte beschriebene Zelle in J eingegeben (Zelle: J368)
Befehl:

$$= \text{UMWANDELN}(J367; F; C)$$

Ergebnis:

28,33°C

2.2 Thema: Datenhaltung

1. Normalform: Es werden nur atomare Werte verwendet
2. Normalform: 1. NF erfüllt, jedes Nichtschlüssel-Attribut ist nur vom Primärschlüssel abhängig, nicht aber von einer Teilmenge desselben.

2.2.1 Datenbank-Schema entwerfen

Zunächst hat man ein Datenbank-Schema erzeugt für alle Daten in der CSV-Datei, wobei der Datensatz auf der Gruppe 5 basiert:

```
Verleih(VerleihId#, WetterId#, count)
Wetter(WetterId#, Datumid#, precipitation, windspeed, min_temperature, average_temperature, max_temperature)
Datum(DatumId#, Stationid#, day_of_year, day_of_week, month_of_year)
station(Stationid#, name)
```

Oder: Es ist möglich eine Tabelle zu erzeugen mit den für die Abgabe wichtigen Daten:

```
Verleih(VerleihId#, Gruppe, Station, Datum, Mittlere Temperatur)
```

2.2.2 Umsetzung des Schemas in SQL und Import der zugeordneten Daten als CSV

Um eine SQL-Tabelle zu erstellen wurden folgende Befehle genutzt [5] (1. Schema):

```
CREATE TABLE Stationen (  
  Id INTEGER PRIMARY KEY,  
  Name TEXT NOT NULL  
);
```

```
CREATE TABLE Datum (  
  Id INTEGER PRIMARY KEY,  
  StationId INTEGER  
  Date TEXT NOT NULL,  
  Dayofyear INTEGER,  
  Dayofweek INTEGER,  
  Month INTEGER  
  FOREIGN KEY(stationid) REFERENCES station(id)  
);
```

```
CREATE TABLE Wetter(  
  Id INTEGER PRIMARY KEY,  
  DatumId INTEGER,  
  Precipitation DECIMAL,  
  Windspeed DECIMAL,  
  MinTemperture INTEGER,  
  AverageTemperature INTEGER,  
  MaxTemperature INTEGER,  
  FOREIGN KEY (DatumId) REFERENCES Datum(id)  
);
```

```
CREATE TABLE Verleih(  
  id INTEGER PRIMARY KEY NOT NULL,  
  Wetterid INTEGER,  
  count INTEGER,  
  FOREIGN KEY (WetterId) REFERENCES Wetter (Id)  
);
```

Oder mit den gekürzten Werten:

```
CREATE TABLE Verleih (  
  ID INTEGER PRIMARY KEY,  
  (Gruppe INTEGER,-> nur bei nicht gekürzten Datensatz)  
  Station TEXT,  
  Datum TEXT,
```

Mittlere Temperatur REAL

);

Anschließend wurde die zuvor abgeänderte CSV-Datei mit den gefilterten Werten der Gruppe 5 mit dem Tool SQLiteOnline [6] in die Datenbank importiert, mithilfe des Buttons *Import*.

2.2.3 Bestimmung der höchsten mittleren Temperatur mit SQL

Für die SQL-Abfrage zur höchsten mittleren Temperatur wurde der SQL-Befehl SELECT MAX[7] genutzt:

1.Schema

```
SELECT MAX (AverageTemperature) AS hoechstemittlereTemperatur  
FROM Wetter
```

2.Schema

```
SELECT MAX((average_temperature-32)*5/9) AS hoechstemittlereTempera-  
tur  
FROM Verleih  
(WHERE group = 5 -> für ungekürzten Datensatz)
```

Ergebnis:

28,33°C

Literatur

- [1] Achim Klenke. *Wahrscheinlichkeitstheorie*. Springer, 3 edition, 2013.
- [2] StudySmarter. Der zentrale Grenzwertsatz. <https://www.studysmarter.de/studium/mathematik-studium/statistik-studium/der-zentrale-grenzwertsatz/>. letzter Zugriff: 01.12.2024.
- [3] Tim Aschermann. Excel den höchsten Wert in einer Zahlenreihe ermitteln. <https://praxistipps.chip.de/excel-den-hoechsten-wert-einer-zahlenreihe-ermitteln-10036>. letzter Zugriff: 28.11.2024.
- [4] Microsoft. UMWANDELN (Funktion). <https://support.microsoft.com/de-de/office/umwandeln-funktion-d785bef1-808e-4aac-bdcd-666c810f9af2>. letzter Zugriff: 28.11.2024.
- [5] Maik Teßmer. Computergestützte Methoden, Block 1, Vorlesung 3/3. Veröffentlichung im Moodle Lernraum, letzter Zugriff: 01.12.2024.
- [6] SQLite. SQLite. <https://sqliteonline.com>. letzter Zugriff: 01.12.2024.
- [7] IONOS Redaktion. LaTeX: Fett schreiben - so geht's. <https://www.ionos.de/digitalguide/server/konfiguration/sql-min-and-max/>. letzter Zugriff: 01.12.2024.
- [8] Lennart Oelschläger. Computergestützte Methoden, Block 2, Vorlesung 1/2. Veröffentlichung im Moodle Lernraum, letzter Zugriff: 26.11.2024.
- [9] Matheretter. Liste von Mathematischen Zeichen. <https://www.matheretter.de/wiki/mathe-zeichen>. letzter Zugriff: 26.11.2024.
- [10] Latex-antworten. Mit Bibtex zitieren. <https://latex-fragen.de/mit-bibtex-eine-webseite-zitieren/>. Letzter Zugriff: 26.11.2024.
- [11] Sabrina Dankers. Wert ins Literaturverzeichnis ohne cite. <https://www.mrunix.de/forums/showthread.php?59320-Werk-ins-Literaturverzeichnis-ohne-cite>. letzter Zugriff: 27.11.2024.
- [12] Jonas Gunreben. Latex: Zahlen hochstellen und Tiefstellen - so geht's. <https://praxistipps.chip.de/latex-zahlen-hochstellen-und-tiefstellen-so-gehts>. letzter Zugriff: 25.11.2024.
- [13] nilis. Reihenfolge der Quellenverweise. <https://www.mrunix.de/forums/archive/index.php/t-60284.html>. letzter Zugriff: 27.11.2024.
- [14] localghost. Das Grad Zeichen in Latex. <https://www.mrunix.de/forums/archive/index.php/t-49360.html>. letzter Zugriff: 28.11.2024.

- [15] Michael Mierke. Latex: Fett schreiben - so geht's. <https://www.heise.de/tipps-tricks/LaTeX-Fett-schreiben-so-geht-s-4979128.html>. letzter Zugriff: 28.11.2024.
- [16] Datenbanken verstehen. SQL MAX() Funktion. <https://www.studysmarter.de/studium/mathematik-studium/statistik-studium/der-zentrale-grenzwertsatz/>. letzter Zugriff: 01.12.2024.
- [17] Micheal Mierke. Latex Unterstrich einfügen. <https://www.heise.de/tipps-tricks/LaTeX-Unterstrich-einfuegen-so-klappt-s-4425609.html>. letzter Zugriff: 01.12.2024.
- [18] SQL Lernen. CREATE TABLE. <http://www.sql-lernen.de/create-table.php>. letzter Zugriff: 01.12.2024.