

Semesterbegleitende Abgabe WiSe2026

Statistik im Bachelor Wirtschaftsinformatik

Dozent_innen: Martin Spott, Michael Heimann, Anja Hirsch

Stand: 28.10.2025

Gruppe:

Gruppe_12 1

Autor_innen:

Eter Rami Porath Daniel Kiedacz David Julian

Einleitung

In diesem Semester beschäftigen wir uns mit Kraftstoffpreisen in Berlin. Dazu benutzen wir Daten von der Organisation Tankerkönig [1], die jede Preisänderung angeschlossener Tankstellenbetreiber_innen mit Zeitstempel protokolliert. In der Datei *tankerkoenigWiSe2025-26.csv* finden Sie den Ausschnitt der Daten, den Sie für die Aufgaben benutzen sollen. Dabei haben wir für jede Tankstelle und jede Stunde den Median der von Tankerkönig registrierten Preisänderungen der Kraftstoffe Diesel und E5 ermittelt.

Regeln

Arbeiten Sie in Ihrer Moodlegruppe von bis zu vier Studierenden, die alle in derselben Übungsgruppe sein und die eine zu wählende Abgabe auch in den zugehörigen Übungen vorstellen müssen (Terminkalender folgt). Jede/r von Ihnen hat in Moodle nur Zugriff auf die Gruppeneinteilung Ihrer in LSF festgelegten Übungsgruppe, wo Sie sich in einer Gruppe eintragen müssen, um Punkte zu bekommen.

Bearbeiten Sie die Aufgaben fortlaufend in diesem RMarkdown-Dokument und geben es bei jeder Teilabgabe komplett als RMarkdown- und als mit Knit übersetztes HTML-Dokument ab. Stellen Sie sicher, dass das RMarkdown-Dokument lauffähig ist. Gehen Sie dazu davon aus, dass sich auf unseren Rechnern die Daten im selben Verzeichnis befinden werden wie Ihr RMarkdown-Dokument.

Benutzen Sie nur Verfahren und Techniken, die wir in der Vorlesung besprochen haben. Sie können aber gerne komplexere Visualisierungen einsetzen, wenn dies sinnvoll erscheint.

Erklären Sie kurz Ihr Vorgehen vor jedem Codeabschnitt und interpretieren die Ergebnisse danach. Schreiben Sie Erklärungen/Interpretationen in Textform zwischen die Codeblöcke, nicht als Kommentare in die Codeblöcke. Sie können den Code aber gerne zusätzlich knapp kommentieren.

Für jede weitere Woche bekommen Sie eine Aufgabenstellung in RMarkdown, die Sie bitte in diesem Dokument fortlaufend anhängen und dort bearbeiten. Geben Sie immer das gesamte Dokument einschließlich der vorangehenden Aufgaben ab, da die Aufgaben teilweise aufeinander aufbauen und Sie vorangehende Lösungen später benötigen.

Woche 1 – 28.10.2025

Beschreiben Sie zwischen den Codeblöcken in Textform, was Sie warum tun und was Sie herausfinden.

- a. Importieren Sie die Daten.

```
data <- read.csv("tankerkoenigWiSe2025-26.csv")
head(data)
```

```
##   X                                id      date      time      name
## 1 1 005056a9-779e-1ee9-81a5-0ae97d8d86da 2023-01-01 06:00:00 ORLEN Tankstelle
## 2 2 005056a9-779e-1ee9-81a5-0ae97d8d86da 2023-01-01 07:00:00 ORLEN Tankstelle
## 3 3 005056a9-779e-1ee9-81a5-0ae97d8d86da 2023-01-01 08:00:00 ORLEN Tankstelle
## 4 4 005056a9-779e-1ee9-81a5-0ae97d8d86da 2023-01-01 09:00:00 ORLEN Tankstelle
## 5 5 005056a9-779e-1ee9-81a5-0ae97d8d86da 2023-01-01 10:00:00 ORLEN Tankstelle
## 6 6 005056a9-779e-1ee9-81a5-0ae97d8d86da 2023-01-01 11:00:00 ORLEN Tankstelle
##   brand      street house_number post_code opening_time closing_time
## 1 ORLEN Mariendorfer Damm      306.0    12107      06:00:00      22:00:00
## 2 ORLEN Mariendorfer Damm      306.0    12107      06:00:00      22:00:00
## 3 ORLEN Mariendorfer Damm      306.0    12107      06:00:00      22:00:00
## 4 ORLEN Mariendorfer Damm      306.0    12107      06:00:00      22:00:00
## 5 ORLEN Mariendorfer Damm      306.0    12107      06:00:00      22:00:00
## 6 ORLEN Mariendorfer Damm      306.0    12107      06:00:00      22:00:00
##   e5_median diesel_median
## 1      1849          1949
## 2      1839          1929
## 3      1879          1959
## 4      1814          1879
## 5      1769          1829
## 6      1799          1849
```

```
str(data)
```

```
## 'data.frame':    242877 obs. of  13 variables:
## $ X              : int  1 2 3 4 5 6 7 8 9 10 ...
## $ id             : chr  "005056a9-779e-1ee9-81a5-0ae97d8d86da" "005056a9-779e-1ee9-81a5-0ae97d8d86da" ...
## $ date           : chr  "2023-01-01" "2023-01-01" "2023-01-01" "2023-01-01" ...
## $ time           : chr  "06:00:00" "07:00:00" "08:00:00" "09:00:00" ...
## $ name           : chr  "ORLEN Tankstelle" "ORLEN Tankstelle" "ORLEN Tankstelle" "ORLEN Tankstelle" ...
## $ brand          : chr  "ORLEN" "ORLEN" "ORLEN" "ORLEN" ...
## $ street         : chr  "Mariendorfer Damm" "Mariendorfer Damm" "Mariendorfer Damm" "Mariendorfer Damm" ...
## $ house_number   : chr  "306.0" "306.0" "306.0" "306.0" ...
## $ post_code      : int  12107 12107 12107 12107 12107 12107 12107 12107 12107 12107 ...
## $ opening_time   : chr  "06:00:00" "06:00:00" "06:00:00" "06:00:00" ...
## $ closing_time   : chr  "22:00:00" "22:00:00" "22:00:00" "22:00:00" ...
## $ e5_median      : int  1849 1839 1879 1814 1769 1799 1779 1759 1789 1749 ...
## $ diesel_median : int  1949 1929 1959 1879 1829 1849 1839 1814 1849 1809 ...
```

b. Beschreiben Sie die Daten. Was sind die Merkmale, was die Merkmalsträger?

Die Merkmale sind -> x, id, date, time, name, brand, street, house_number, post_code, opening_time, closing_time, e5_median, diesel_median. Die Merkmalsträger sind -> die einzelnen Tankstellenbeobachtungen.

c. Geben Sie für alle Merkmale an, ob sie qualitativ/quantitativ, diskret/stetig und auf welchen Skalen sie definiert sind.

x = quantitativ diskret metrisch

id = qualitativ – nominal

date = qualitativ diskret ordinal

time = quantitativ stetig metrisch

name = qualitativ – nominal
 brand = qualitativ – nominal
 streetm = qualitativ – nominal
 house_number = quant diskret metrisch
 post_code = qualitativ – nominal
 opening_time = quan stetig metrisch
 closing_time = quant stetig ,etrisch
 e5_median = quantitativ stetig metrisch
 dieseln_median = quantitativ stetig metrisch

d. Berechnen Sie für das Merkmal **brand** (Marke) die Tabellen mit absoluter und relativer Häufigkeit.

```
abs_brand <- table(data$brand)
```

```
rel_brand <- prop.table(table(data$brand))
```

abs_brand

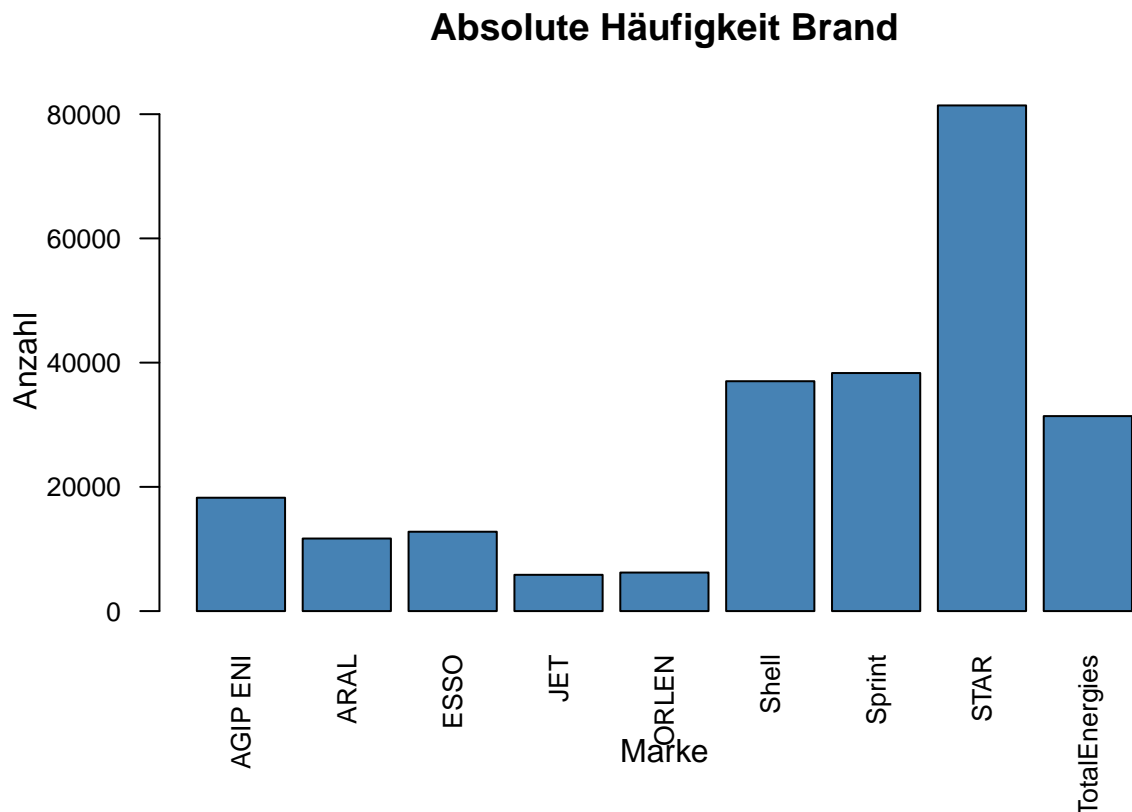
```
##
##      AGIP ENI      ARAL      ESSO      JET      ORLEN
##      18250      11680      12775      5840      6205
##      Shell      Sprint      STAR TotalEnergies
##      37005      38325      81407      31390
```

rel_brand

```
##
##      AGIP ENI      ARAL      ESSO      JET      ORLEN
##      0.07514091  0.04809019  0.05259864  0.02404509  0.02554791
##      Shell      Sprint      STAR TotalEnergies
##      0.15236107  0.15779592  0.33517789  0.12924237
```

e. Stellen Sie die Tabelle absoluter Häufigkeit von **brand** als Balkendiagramm dar. Beschriften Sie alle Axen und geben dem Diagramm eine Überschrift.

```
barplot(abs_brand, xlab = "Marke", ylab = "Anzahl", main = "Absolute Häufigkeit Brand", las = 2, col =
```



f. Was bedeuten die absoluten Häufigkeiten im Kontext unserer Daten? Warum gibt es Unterschiede zwischen den Werten der verschiedenen Marken?

Die absoluten Häufigkeiten geben an, wie oft jede Marke im Datensatz vorkommt, also wie viele Beobachtungen (Tankstellen) einer bestimmten Marke enthalten sind.

Beispiel:

Wenn Aral = 200, Shell = 150, dann bedeutet das: Es wurden 200 Tankstellen von Aral und 150 von Shell erfasst.

Unterschiedliche Marktanteile der Marken (Aral, Shell, Esso, etc.)

Unterschiedliche Anzahl an Tankstellen je Region

Eventuell nicht überall die gleichen Beobachtungszeitpunkte im Datensatz

Quellen

[1] Tankerkönig UG, URL: <https://tankerkoenig.de>, letzter Aufruf: 20.10.2025