

# Methoden der Hydrologie: 1. Übung

Daniel Abanto

2023-11-17

## Settings

```
knitr::opts_chunk$set(warning = FALSE, message = FALSE, fig.align="center")
opts_knit$set(root.dir = '/Users/dabanto/Desktop/master/hydro_master/ezg_weiler')
opts_chunk$set(tidy.opts=list(width.cutoff=60),tidy=TRUE)
```

## Packages, die ich verwenden werde

|    |              |            |           |           |           |         |
|----|--------------|------------|-----------|-----------|-----------|---------|
| ## | gt           | patchwork  | reshape   | gridExtra | gtsummary | ggpubr  |
| ## | "0.9.0"      | "1.1.2"    | "0.8.9"   | "2.3"     | "1.7.2"   | "0.6.0" |
| ## | ggstatsplot  | smplot2    | lubridate | readr     | tidyr     | ggplot2 |
| ## | "0.11.1"     | "0.1.0"    | "1.9.2"   | "2.1.4"   | "1.3.0"   | "3.4.4" |
| ## | dplyr        | rmdformats | knitr     |           |           |         |
| ## | "1.1.3.9000" | "1.0.4"    | "1.45"    |           |           |         |

## Daten

Als erstes lade ich die Datensätze, die im .txt und .csv Format aus der LUBW- und BfG Webseite heruntergeladen wurden. Ich werde mich mit einem Pegel des Neckars sowie mit zwei verschiedenen Parametern beschäftigen, die Auskunft über die Wasserqualität geben.

```
#Neckar Abfluss
df_neckar <-
  ↪ read.table("/Users/dabanto/Downloads/2023-11-13_12-55/6335600_Q_Day.Cmd.txt", header
  ↪ = T)

#Diclofenac Konzentration
diclofenac <-
  ↪ read_csv("/Users/dabanto/Downloads/Daten_der_Chemie_Messstellen_2023_11_13.csv")

#Calcium Konzentration
calcium <-
  ↪ read_csv("/Users/dabanto/Downloads/Daten_der_Chemie_Messstellen_2023_11_13(1).csv")
```

## Datenvorverarbeitung

Die Datum Spalten werden umbenannt und als **date** Format gespeichert

```
df_neckar <- df_neckar %>%
  dplyr::rename(Datum = YYYY.MM.DD.hh.mm.) %>%
  mutate(date = ymd(Datum)) %>%
  select(-Datum)
```

```
diclofenac <- diclofenac %>%
  mutate(date = as.Date(Datum, format="%d.%m.%Y %H:%M"))

calcium <- calcium %>%
  mutate(date = as.Date(Datum, format="%d.%m.%Y %H:%M"))
```

Die Diclofenac- und Calcium-Werte wurden zwischen Januar 2010 und Dezember 2022 gemessen. Deshalb extrahiere ich die Abflussdaten, sodass sie mit den Messwerten im Einklang sind.

```
#extract data
df_neckar <- df_neckar %>%
  filter(between(date, as.Date('2010-01-01'), as.Date('2022-12-31')))
```

Die Calcium-Messwerte sind bereits als **num** gespeichert, die Diclofenac-Messwerte als **chr**.

```
diclofenac$Messwert <- gsub(",", ".", diclofenac$Messwert)

diclofenac$Messwert <- as.numeric(diclofenac$Messwert)
```

## Zusammenführung von Daten

Die Mess- und Abflusswerte werden zusammengeführt, um einen einheitlichen Datensatz zu erstellen..

```
#Abfluss- und Diclofenac-Werte werden zusammengeführt. Der Join erfolgt mit Hilfe des
↪ Datums
zsm_ds <- merge(x = df_neckar, y = diclofenac[ , c("date", "Messwert")], by = "date",
↪ all.x = TRUE)

#Die Abfluss und Messwerte-Spalte wird umbenannt, damit keine Konflikte entstehen.
zsm_ds <- zsm_ds %>%
  dplyr::rename(Messwert_diclofenac = Messwert)

zsm_ds <- zsm_ds %>%
  dplyr::rename(Abfluss = Value)

#Jetzt werden noch die Calcium-Werte gejoined.
final_ds <- merge(x = zsm_ds, y = calcium[ , c("date", "Messwert")], by = "date", all.x =
↪ TRUE)

#Hier wird auch die Messwerte-Spalte umbenannt.
final_ds <- final_ds %>%
  dplyr::rename(Messwert_calcium = Messwert)

head(final_ds, 10) %>%
  gt()
```

| date  | Abfluss | Messwert_diclofenac | Messwert_calcium |
|-------|---------|---------------------|------------------|
| 14610 | 329     | NA                  | NA               |
| 14611 | 263     | NA                  | NA               |
| 14612 | 215     | NA                  | NA               |
| 14613 | 185     | NA                  | 80               |
| 14614 | 164     | NA                  | NA               |
| 14615 | 155     | NA                  | NA               |
| 14616 | 136     | NA                  | NA               |
| 14617 | 117     | NA                  | NA               |
| 14618 | 119     | NA                  | NA               |
| 14619 | 116     | NA                  | NA               |

## Weitere Datenvorverarbeitung

Da ich mit `facet_wrap()` arbeiten werde, muss ich die Ausrichtung des DataFrames ändern. Das bedeutet, dass ich ein langes Format brauche.

Es wird eine Kopie des Dataframes erstellt und die Abfluss-Spalte wird gelöscht, weil ich den Abfluss-Plot mit einer separaten Funktion erstellen muss.

```
melt_df <- final_ds
melt_df$Abfluss <- NULL
```

Jetzt wird die Ausrichtung des Dataframes geändert.

```
mdata <- melt(melt_df, id=c("date"))
```

## Plots

Die Messwerte und die Abflusswerte werden im nächsten Schritt geplottet.

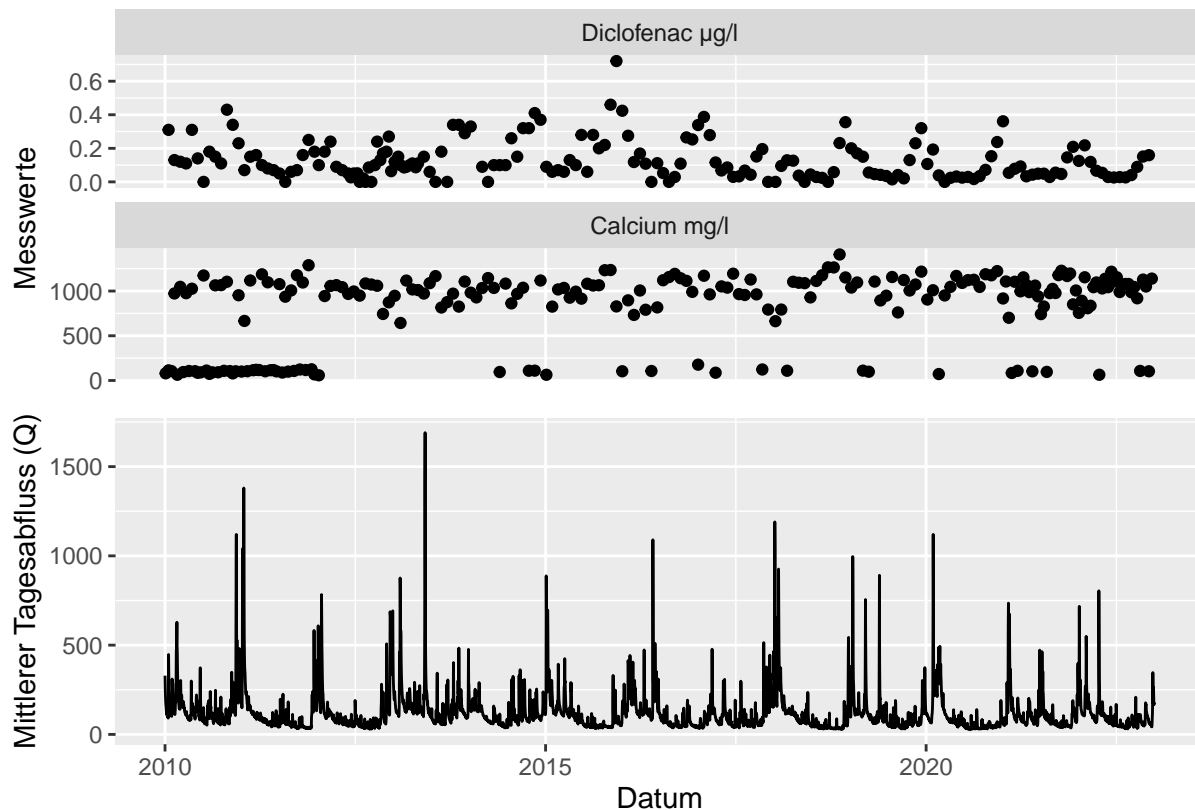
```
panel_labels <- c("Messwert_diclofenac" = "Diclofenac µg/l", "Messwert_calcium" =  
  ↪ "Calcium mg/l")  
  
p1 <- ggplot(mdata, aes(date, value)) +  
  geom_point() +  
  ylab("Messwerte") +  
  facet_wrap(~variable, scales = "free_y", ncol = 1, labeller = labeller(variable =  
    ↪ panel_labels)) +  
  theme(axis.title.x=element_blank(),  
        axis.text.x=element_blank(),  
        axis.ticks.x=element_blank())
```

Als nächstes wird der Abfluss dargestellt. Dafür verwenden wir den Dataframe **final\_ds**.

```
p2 <- ggplot(final_ds, aes(date, Abfluss)) +  
  xlab("Datum") + ylab("Mittlerer Tagesabfluss (Q)") +  
  geom_line()
```

Beide Plots werden in einer Grafik mit einer gemeinsamen X-Achse dargestellt.

```
combined_plot <- p1 / p2  
combined_plot
```



## Analyse der Beziehung

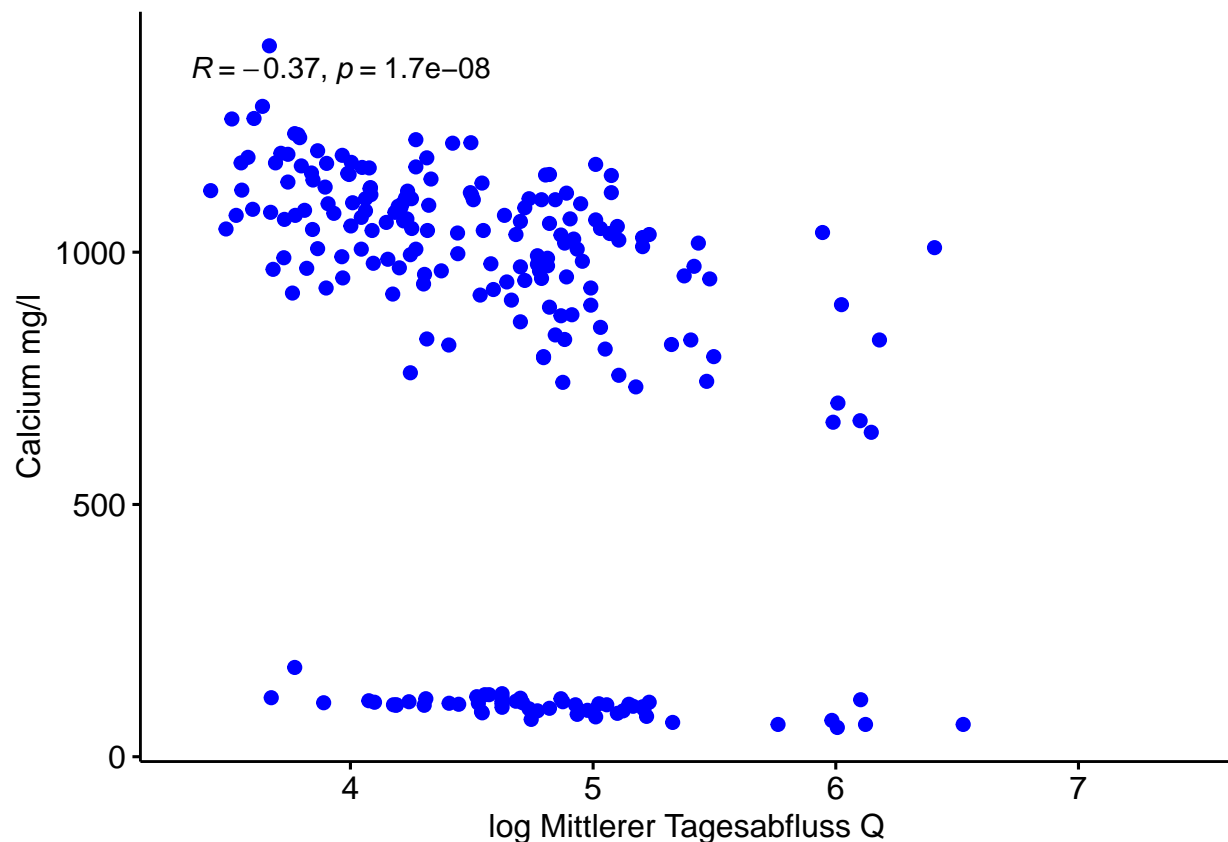
Als nächstes wird die Beziehung zwischen Abfluss und Messwerte untersucht.

Damit die Beziehung besser analysiert werden kann, werden die Abflusswerte logarithmiert.

```
final_ds$log_q <- log(final_ds$Abfluss)
```

Der lineare Zusammenhang zwischen den Variablen wird mit Hilfe einer Pearson-Korrelation untersucht.

```
ggscatter(final_ds, x = "log_q", y = "Messwert_calcium",  
          color = "blue", cor.coef = TRUE,  
          cor.method = "pearson",  
          xlab = "log Mittlerer Tagesabfluss Q", ylab = "Calcium mg/l")
```



Als nächstes wird eine lineare Regression mit `lm()` erstellt, um den  $R^2$  Wert sowie das Konfidenzintervall zu ermitteln

```
#lm modell  
model <- lm(log_q~Messwert_calcium, data=final_ds)  
  
#Erstellen der Tabelle  
x <- tbl_regression(  
  model,  
  pvalue_fun = function(x) {  
    if_else(  
      is.na(x),  
      NA_character_,  
      if_else(x < 0.001, format(x, digits = 3, scientific = TRUE), format(round(x, 3),  
        ↪ scientific = T))  
    )  
  },
```

```

estimate_fun = function(x) {
  if_else(
    is.na(x),
    NA_character_,
    if_else(x < 0.001, format(x, digits = 3, scientific = TRUE), format(round(x, 3),
      ↪ scientific = T)),
    )
  }
)

add_glance_ex1 <-
  x %>%
  add_glance_table(include = c(nobs, r.squared))

add_glance_ex1

```

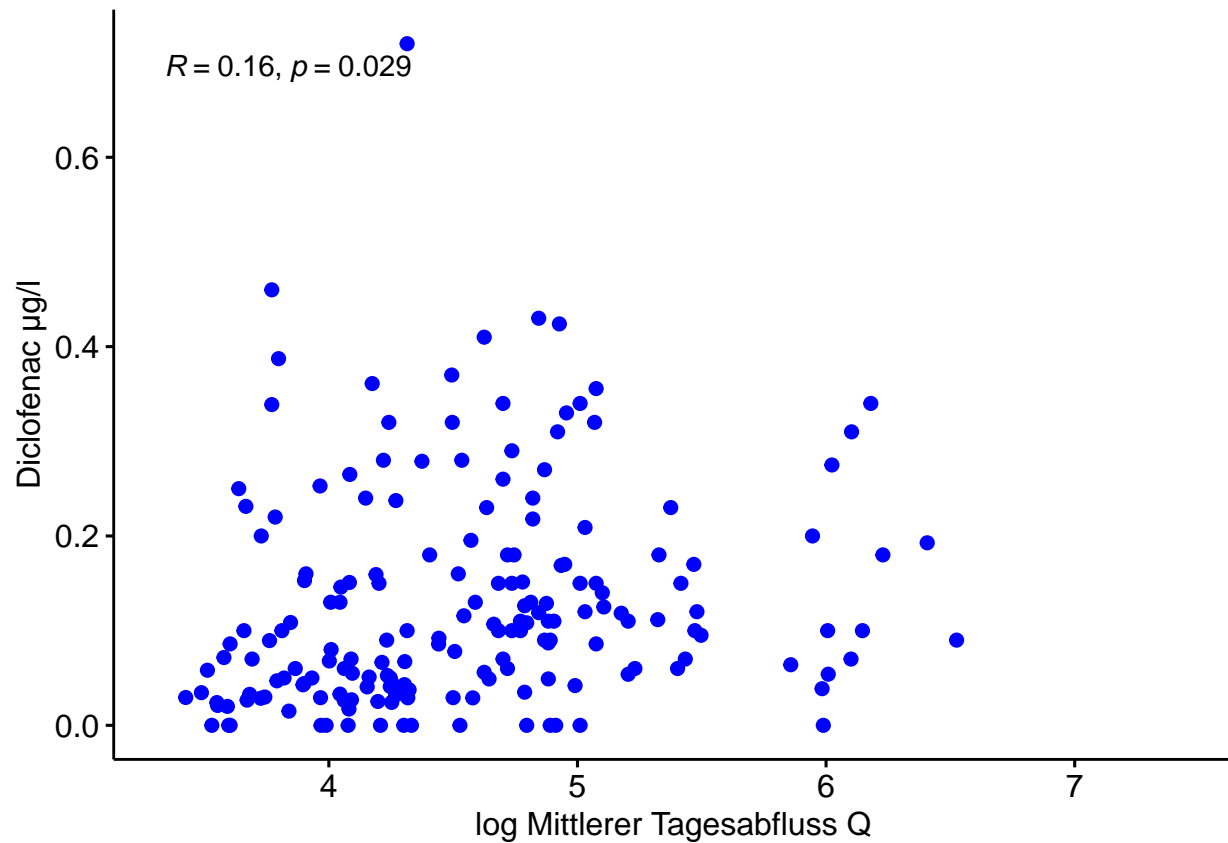
| Characteristic   | Beta      | 95% CI               | p-value |
|------------------|-----------|----------------------|---------|
| Messwert_calcium | -5.62e-04 | -7.52e-04, -3.73e-04 | 1.7e-08 |
| No. Obs.         | 221       |                      |         |
| R <sup>2</sup>   | 0.135     |                      |         |

Die gleichen Schritte werde ich für die Diclofenac-Messwerte anwenden.

```

ggscatter(final_ds, x = "log_q", y = "Messwert_diclofenac",
  color = "blue", cor.coef = TRUE,
  cor.method = "pearson",
  xlab = "log Mittlerer Tagesabfluss Q", ylab = "Diclofenac µg/l")

```



$R^2$  Wert und Konfidenzintervall

```
#lm modell
model <- lm(log_q~Messwert_diclofenac, data=final_ds)

#Erstellen der Tabelle
x <- tbl_regression(
  model,
  pvalue_fun = function(x) {
    if_else(
      is.na(x),
      NA_character_,
      if_else(x < 0.001, format(x, digits = 3, scientific = TRUE), format(round(x, 3),
        ↪ scientific = T))
    )
  },
  estimate_fun = function(x) {
    if_else(
      is.na(x),
      NA_character_,
      if_else(x < 0.001, format(x, digits = 3, scientific = TRUE), format(round(x, 3),
        ↪ scientific = T)),
    )
  }
)

add_glance_ex1 <-
```

```
x %>%
  add_glance_table(include = c(nobs, r.squared))

add_glance_ex1
```

| Characteristic      | Beta    | 95% CI             | p-value |
|---------------------|---------|--------------------|---------|
| Messwert_diclofenac | 9.6e-01 | 9.9e-02, 1.821e+00 | 2.9e-02 |
| No. Obs.            | 180     |                    |         |
| R <sup>2</sup>      | 0.026   |                    |         |

## Diskussion

Zwischen der Konzentration von Kalzium und dem mittleren Tagesabfluss besteht ein negativer Zusammenhang, wie auf der Abbildung ersichtlich ist. Allerdings beträgt das Bestimmtheitsmaß lediglich 0,135, was auf einen geringen linearen Zusammenhang hinweist.

Zwischen der Konzentration von Diclofenac und dem mittleren Tagesabfluss besteht ein positiver Zusammenhang. Das Bestimmtheitsmaß ist in diesem Fall sogar noch geringer (0,026), sodass man interpretieren kann, dass kein linearer Zusammenhang vorliegt.

Eine mögliche Erklärung für die geringen Werte ist das Fehlen von Daten. Für die Calcium-Messwerte gibt es 221 Messungen, während es für die Diclofenac-Messwerte nur 180 gibt. Im Gegensatz dazu sind tägliche Abflussdaten vorhanden.