

analysis_regression_public_transport

```
library(tidyverse)
```

```
## Warning: Paket 'tidyverse' wurde unter R Version 4.4.2 erstellt
```

```
## Warning: Paket 'ggplot2' wurde unter R Version 4.4.3 erstellt
```

```
## Warning: Paket 'tibble' wurde unter R Version 4.4.2 erstellt
```

```
## Warning: Paket 'tidyr' wurde unter R Version 4.4.2 erstellt
```

```
## Warning: Paket 'readr' wurde unter R Version 4.4.2 erstellt
```

```
## Warning: Paket 'purrr' wurde unter R Version 4.4.3 erstellt
```

```
## Warning: Paket 'dplyr' wurde unter R Version 4.4.2 erstellt
```

```
## Warning: Paket 'stringr' wurde unter R Version 4.4.2 erstellt
```

```
## Warning: Paket 'forcats' wurde unter R Version 4.4.2 erstellt
```

```
## Warning: Paket 'lubridate' wurde unter R Version 4.4.2 erstellt
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
```

```
## v dplyr      1.1.4      v readr      2.1.5
```

```
## v forcats   1.0.0      v stringr   1.5.1
```

```
## v ggplot2    3.5.2      v tibble    3.2.1
```

```
## v lubridate  1.9.4      v tidyr     1.3.1
```

```
## v purrr      1.0.4
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2)
```

```
library(tidyr)
```

```
library(dplyr)
```

```
zaehlstellen_verkehr <- read_csv("data/dauerzaehlstellen_data.csv")
```

```

## Rows: 40418 Columns: 17
## -- Column specification -----
## Delimiter: ","
## chr   (4): ZNAME, STRTYP, RENAME, FZTYP
## dbl  (11): ZNR, STRNR, DTVMS, DTVMF, DTVMO, DTVDD, DTVFR, DTVSA, DTVSF, TVMA...
## date  (2): DATUM, TVMAXT
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

public_transport <- read_csv("data/dauerzaehlstellen_location_public_transport_1km.csv")

## Rows: 5590 Columns: 2
## -- Column specification -----
## Delimiter: ","
## dbl (2): ZNR, Public_Transport_Id
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.

znr_counts <- table(public_transport$ZNR)
znr_counts_df <- as.data.frame(znr_counts)
colnames(znr_counts_df) <- c("ZNR", "count")

znr_counts_df$ZNR <- as.character(znr_counts_df$ZNR)

gesamt <- zaehlstellen_verkehr %>% filter(RENAME == "Gesamt")

gesamt_kfz <- gesamt %>% filter(FZTYP == "Kfz")
gesamt_lkw <- gesamt %>% filter(FZTYP == "Lkw")

gesamt_kfz$ZNR <- as.integer(gesamt_kfz$ZNR)
gesamt_lkw$ZNR <- as.integer(gesamt_lkw$ZNR)

gesamt_kfz$ZNR <- as.character(gesamt_kfz$ZNR)
gesamt_lkw$ZNR <- as.character(gesamt_lkw$ZNR)

gesamt_kfz <- merge(gesamt_kfz, znr_counts_df, by = "ZNR", all.x = TRUE)
gesamt_lkw <- merge(gesamt_lkw, znr_counts_df, by = "ZNR", all.x = TRUE)

aggregated_kfz <- gesamt_kfz %>%
  group_by(ZNR) %>%
  summarise(
    DATUM = first(DATUM),
    ZNAME = first(ZNAME),
    STRTYP = first(STRTYP),
    STRNR = first(STRNR),
    RENAME = first(RENAME),
    FZTYP = first(FZTYP),
    DTVMS = sum(DTVMS, na.rm = TRUE),
    DTVMF = sum(DTVMF, na.rm = TRUE),
    DTVMO = sum(DTVMO, na.rm = TRUE),

```

```

    DTVDD = sum(DTVDD, na.rm = TRUE),
    DTVFR = sum(DTVFR, na.rm = TRUE),
    DTVSA = sum(DTVSA, na.rm = TRUE),
    DTVSF = sum(DTVSF, na.rm = TRUE),
    TVMAX = sum(TVMAX, na.rm = TRUE),
    ISTCOVID19 = first(ISTCOVID19),
    count = first(count)
  )

aggregated_lkw <- gesamnt_lkw %>%
  group_by(ZNR) %>%
  summarise(
    DATUM = first(DATUM),
    ZNAME = first(ZNAME),
    STRTYP = first(STRTYP),
    STRNR = first(STRNR),
    RENAME = first(RENAME),
    FZTYP = first(FZTYP),
    DTVMS = sum(DTVMS, na.rm = TRUE),
    DTVMF = sum(DTVMF, na.rm = TRUE),
    DTVMO = sum(DTVMO, na.rm = TRUE),
    DTVDD = sum(DTVDD, na.rm = TRUE),
    DTVFR = sum(DTVFR, na.rm = TRUE),
    DTVSA = sum(DTVSA, na.rm = TRUE),
    DTVSF = sum(DTVSF, na.rm = TRUE),
    TVMAX = sum(TVMAX, na.rm = TRUE),
    ISTCOVID19 = first(ISTCOVID19),
    count = first(count)
  )

plot_data <- aggregated_kfz %>%
  select(ZNR, count, DTVMO, DTVDD, DTVFR, DTVSA, DTVSF, TVMAX)

# In Long-Format transformieren für ggplot
plot_data_long <- pivot_longer(
  plot_data,
  cols = c(DTVMO, DTVDD, DTVFR, DTVSA, DTVSF, TVMAX),
  names_to = "Wochentag",
  values_to = "Verkehr"
)

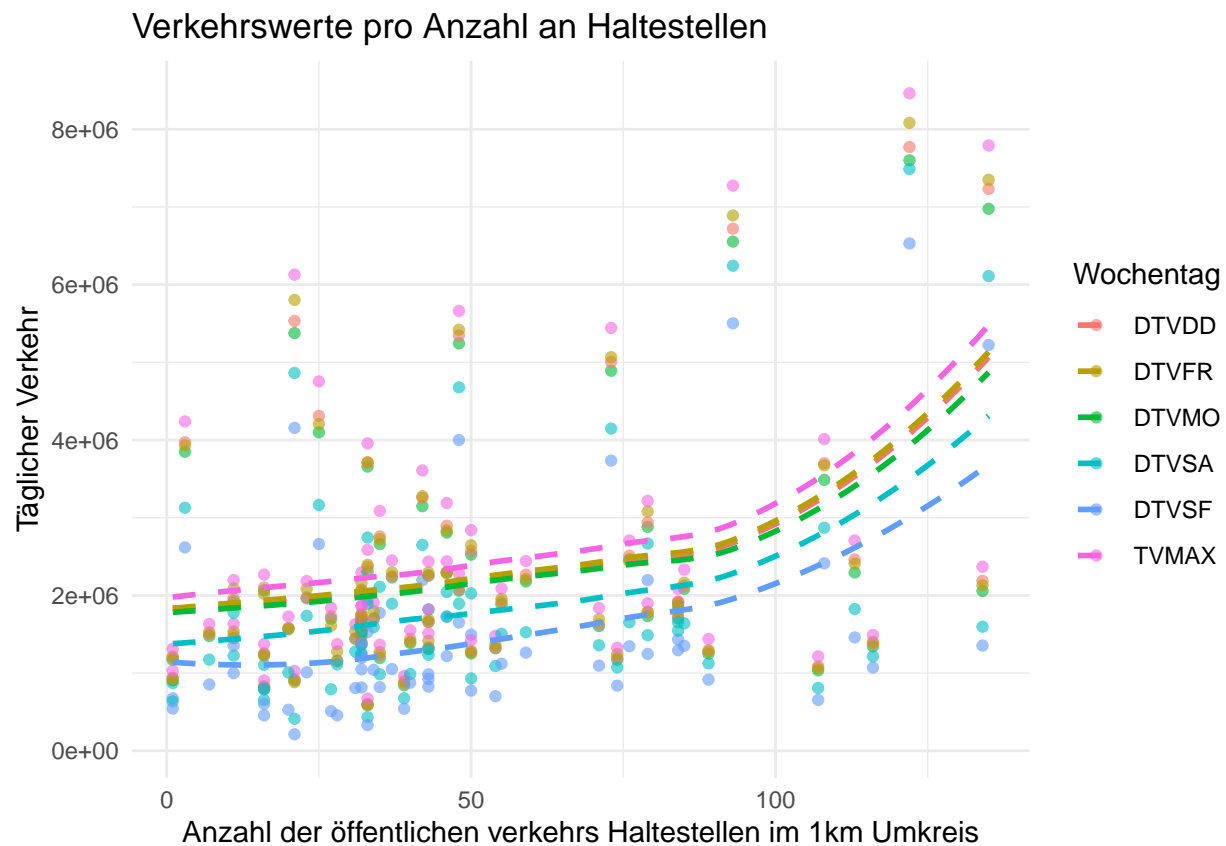
ggplot(plot_data_long, aes(x = count, y = Verkehr, color = Wochentag)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "loess", se = FALSE, linetype = "dashed") +
  labs(
    title = "Verkehrswerte pro Anzahl an Haltestellen",
    x = "Anzahl der öffentlichen verkehrs Haltestellen im 1km Umkreis",
    y = "Täglicher Verkehr",
    color = "Wochentag"
  ) +
  theme_minimal()

```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

```
## Warning: Removed 6 rows containing non-finite outside the scale range
## ('stat_smooth()').
```

```
## Warning: Removed 6 rows containing missing values or values outside the scale range
## ('geom_point()').
```



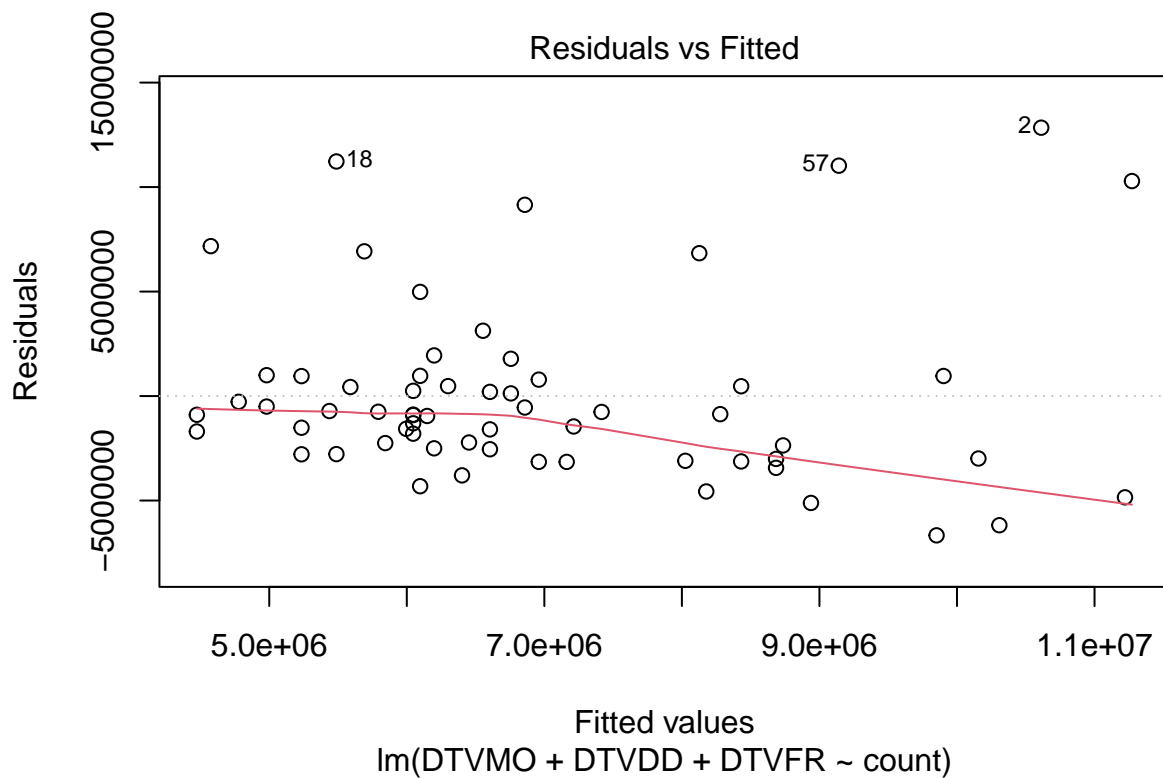
Regressionen KFZ

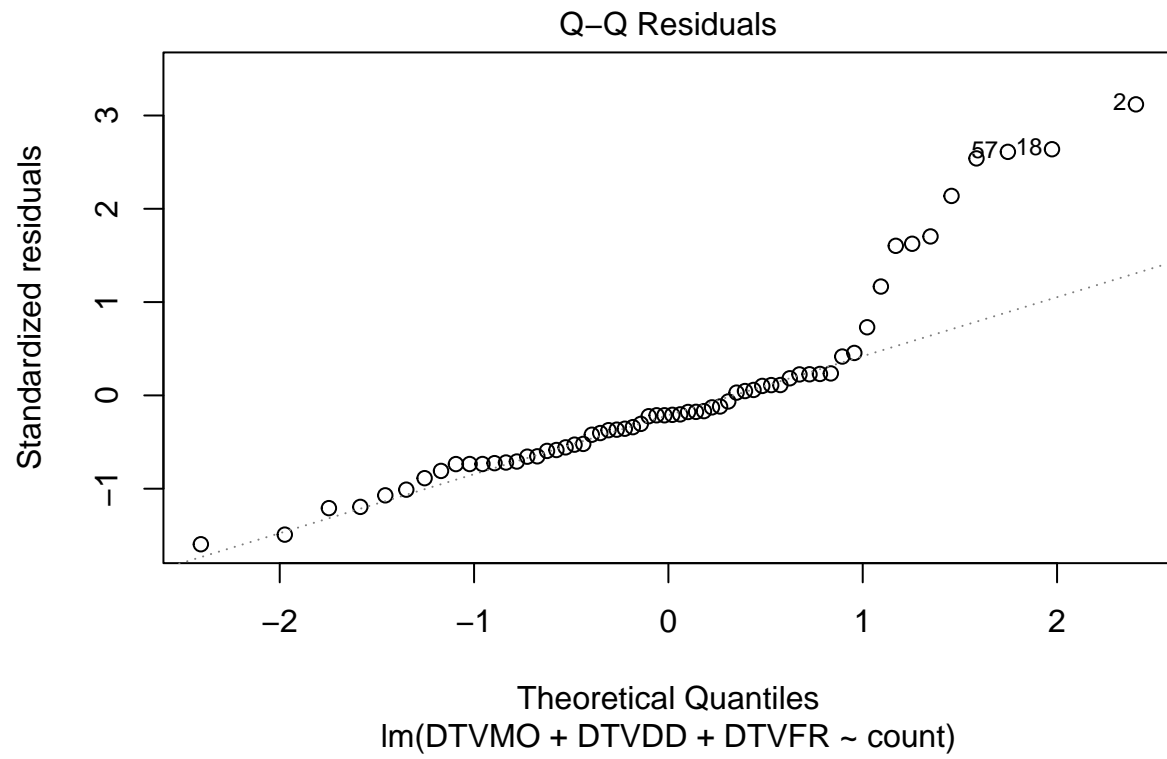
```
# Regression Woche (MO-FR)
regression_kfz_woche <- lm(DTVMO + DTVDD + DTVFR ~ count, data = aggregated_kfz)
summary(regression_kfz_woche)
```

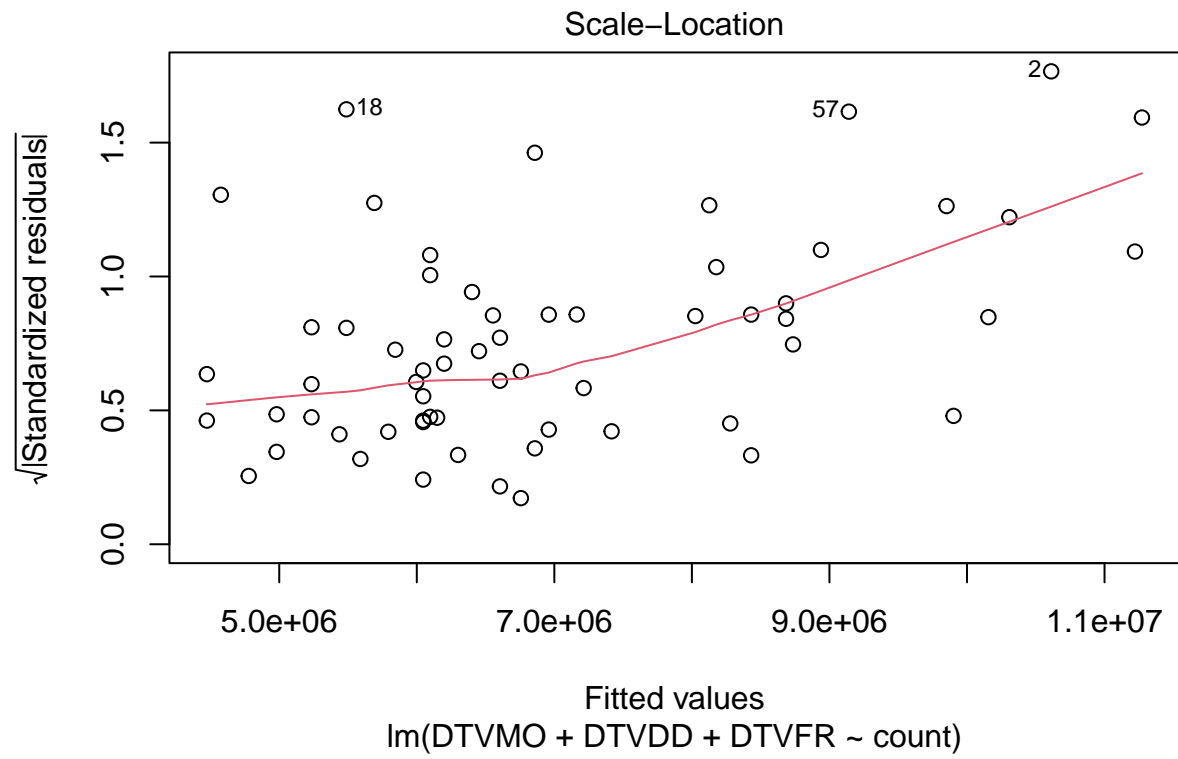
```
##
## Call:
## lm(formula = DTVMO + DTVDD + DTVFR ~ count, data = aggregated_kfz)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6666535 -2720378  -893034   912198 12842076
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4422767    984719   4.491 3.28e-05 ***
```

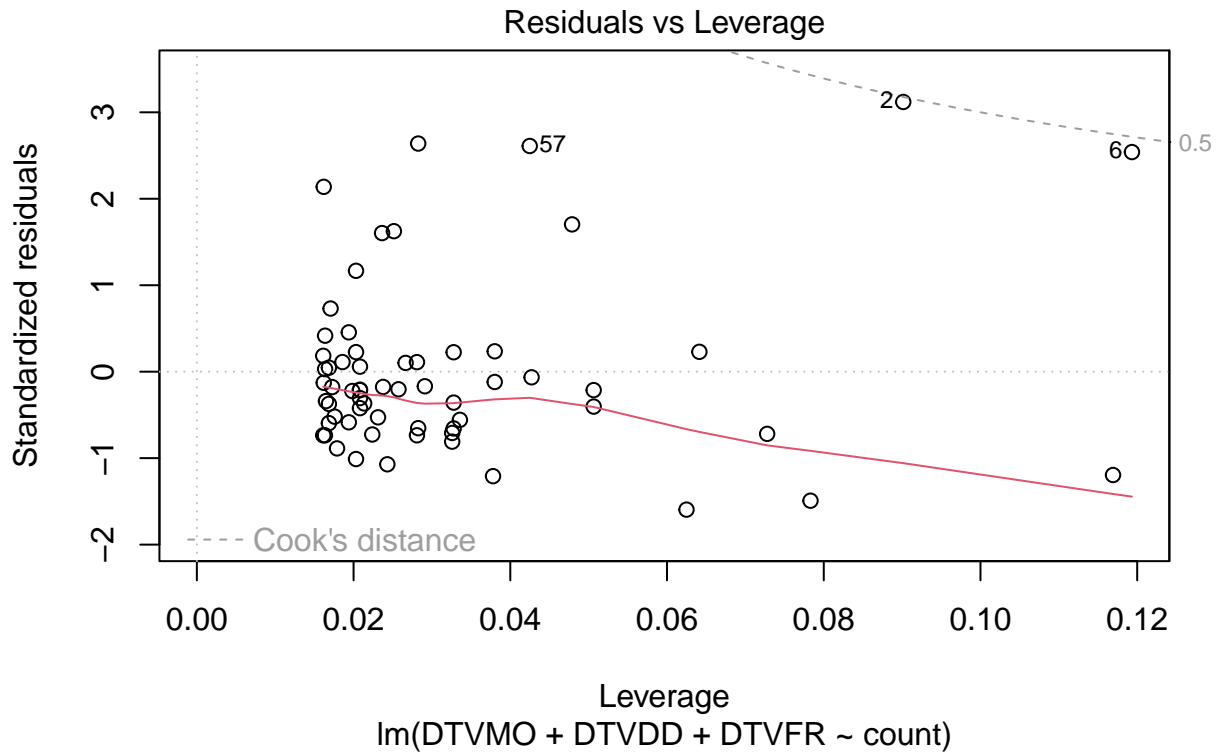
```
## count          50736      16330   3.107  0.00289 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4316000 on 60 degrees of freedom
## (1 Beobachtung als fehlend gelöscht)
## Multiple R-squared:  0.1386, Adjusted R-squared:  0.1242
## F-statistic: 9.653 on 1 and 60 DF,  p-value: 0.002887

# Traffic count ist abhängig von den Haltestellen
plot(regression_kfz_woche)
```









```
# Regression Wochenende und Feiertage
```

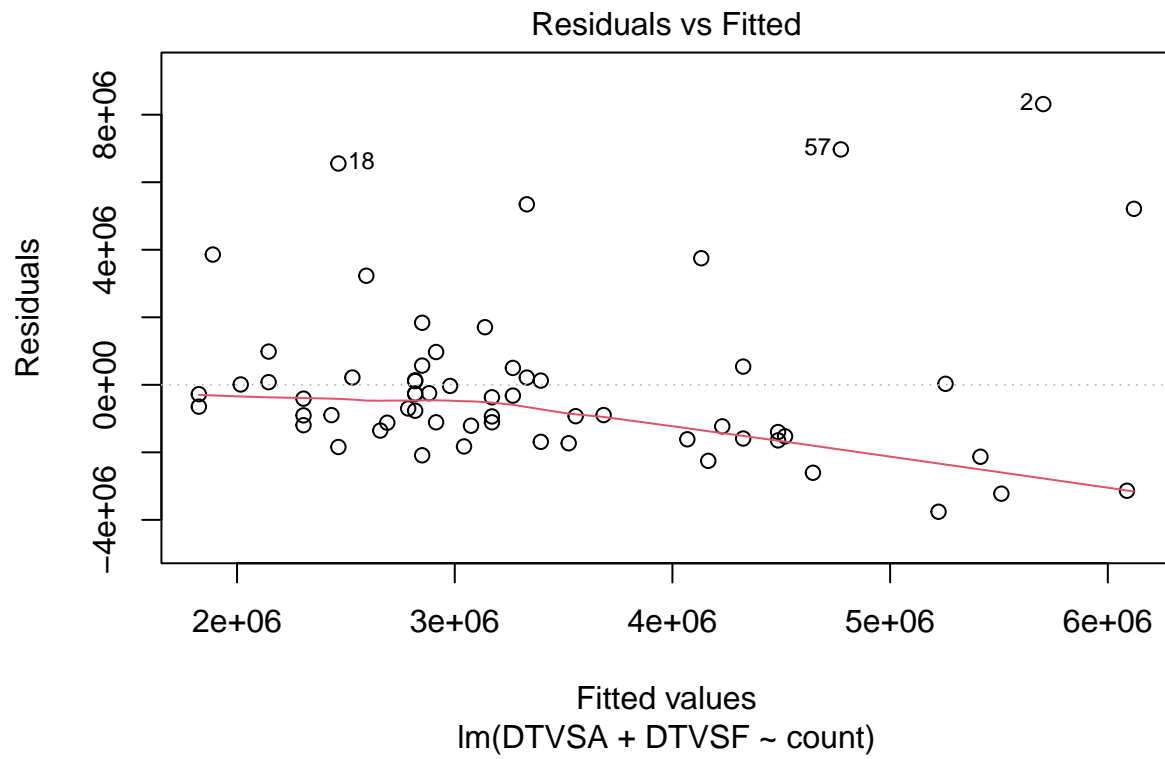
```
regression_kfz_wochenende <- lm(DTVSA + DTVSF ~ count, data = aggregated_kfz)
```

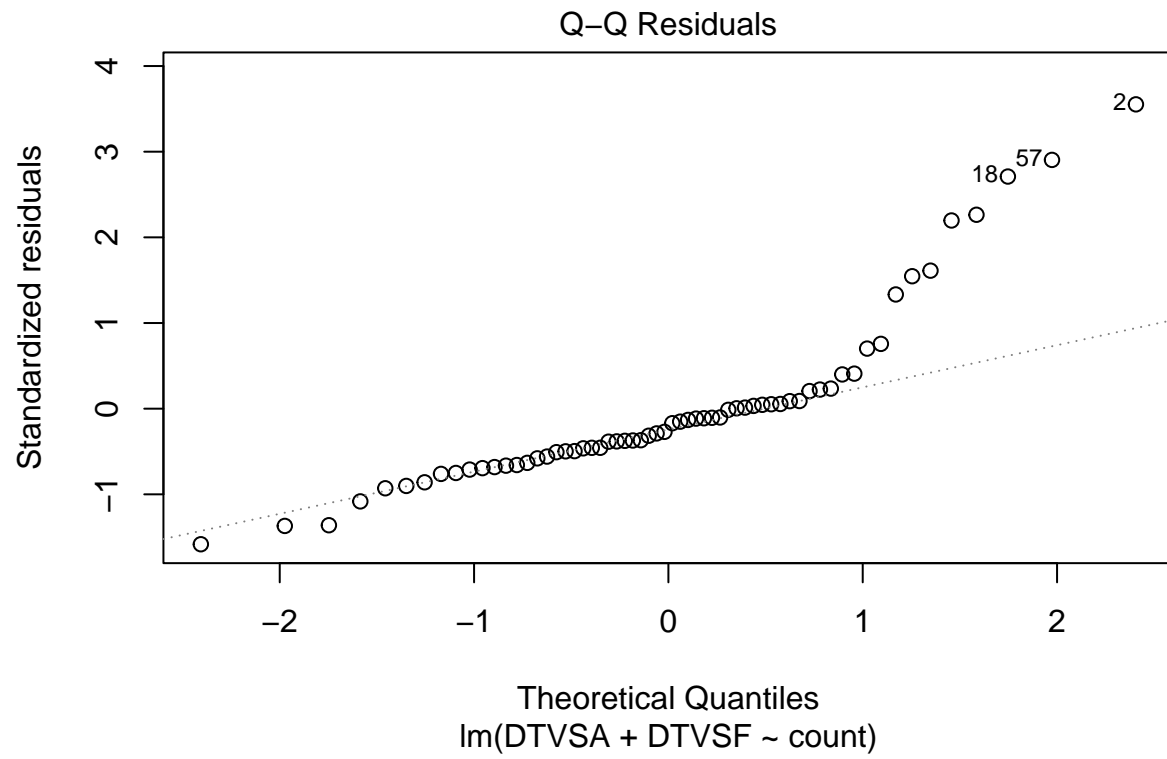
```
summary(regression_kfz_wochenende)
```

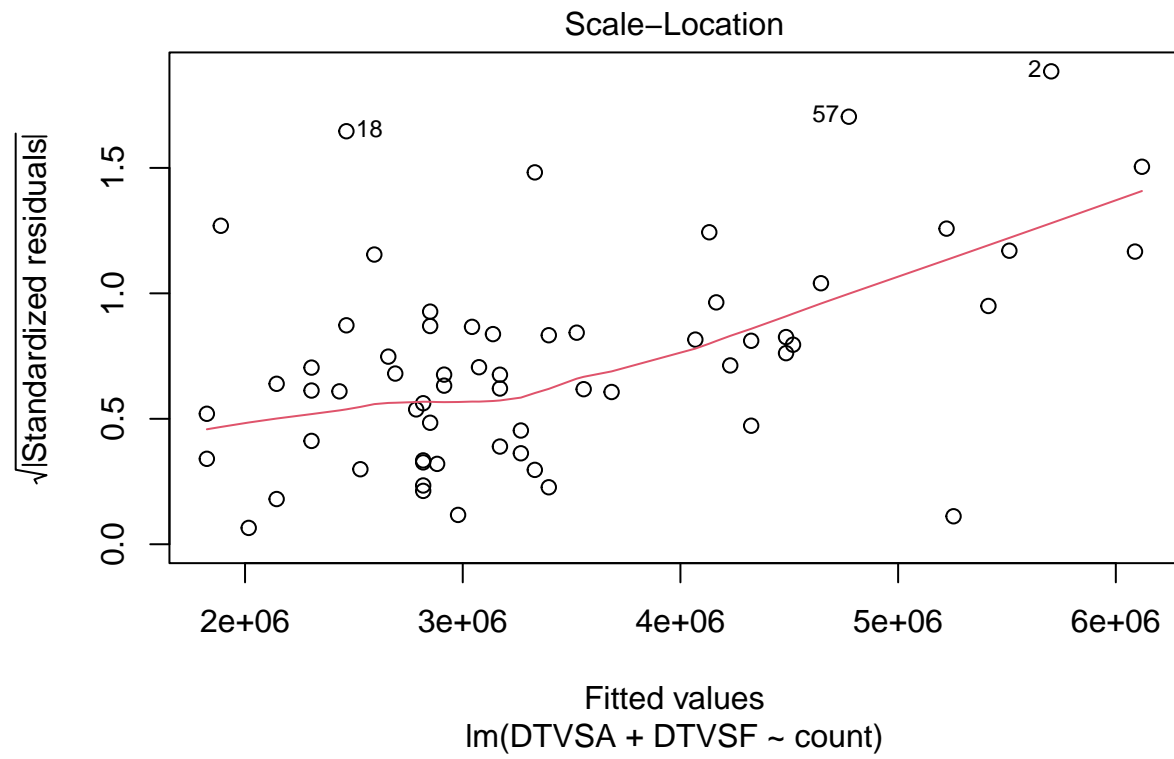
```
##
## Call:
## lm(formula = DTVSA + DTVSF ~ count, data = aggregated_kfz)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3759810 -1388765 -527938  215706  8315174
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1792480    559849   3.202  0.00219 **
## count         32056      9284   3.453  0.00102 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2454000 on 60 degrees of freedom
## (1 Beobachtung als fehlend gelöscht)
## Multiple R-squared:  0.1658, Adjusted R-squared:  0.1519
## F-statistic: 11.92 on 1 and 60 DF, p-value: 0.001023
```

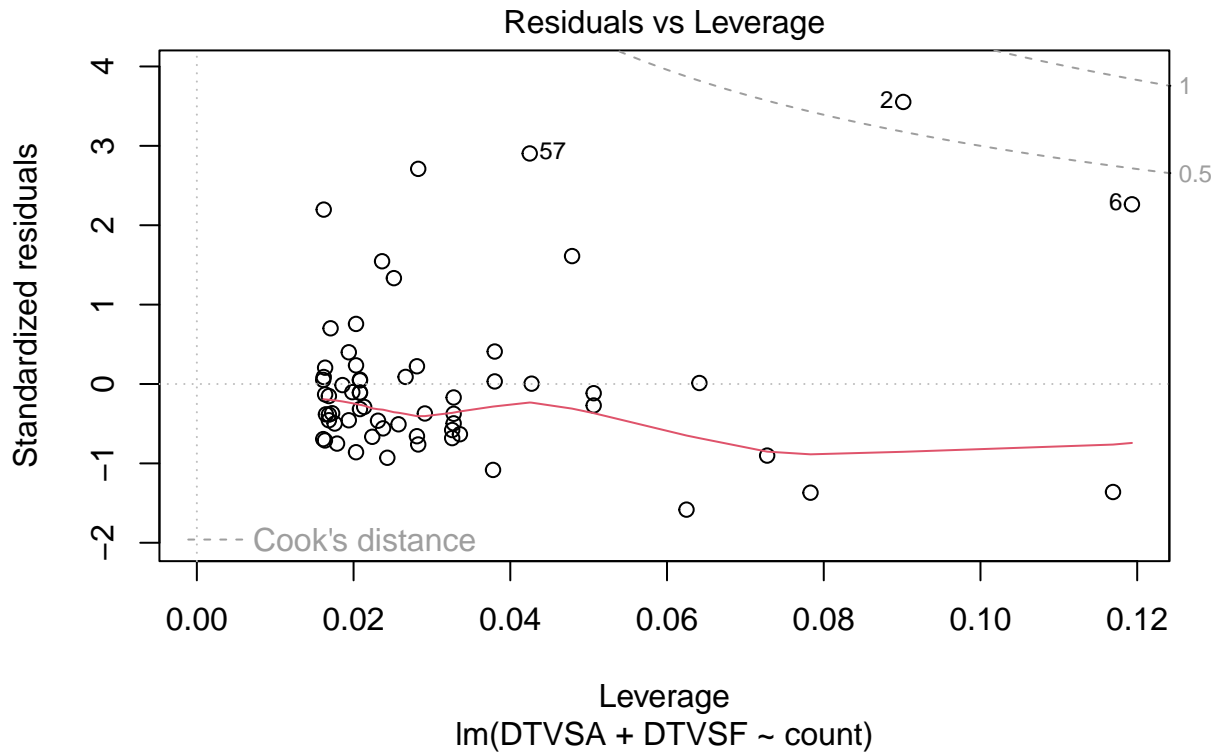


```
# Traffic count ist abhängig von den Haltestellen  
plot(regression_kfz_wochenende)
```





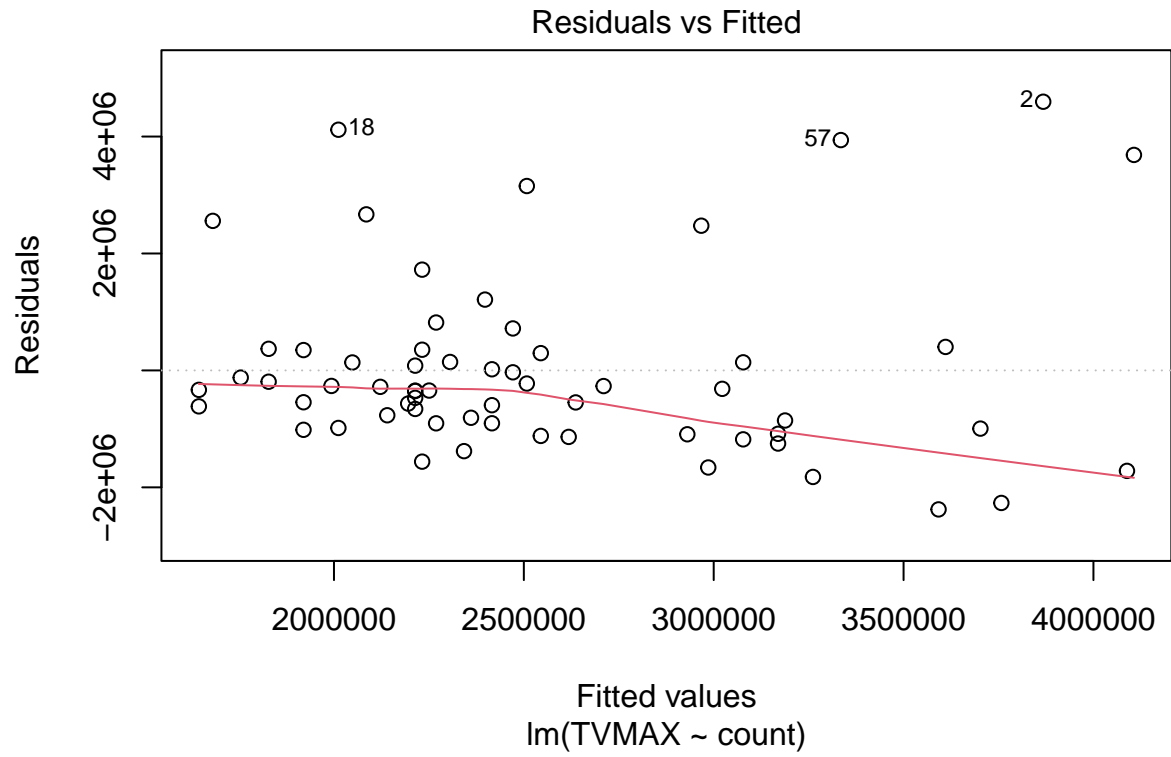


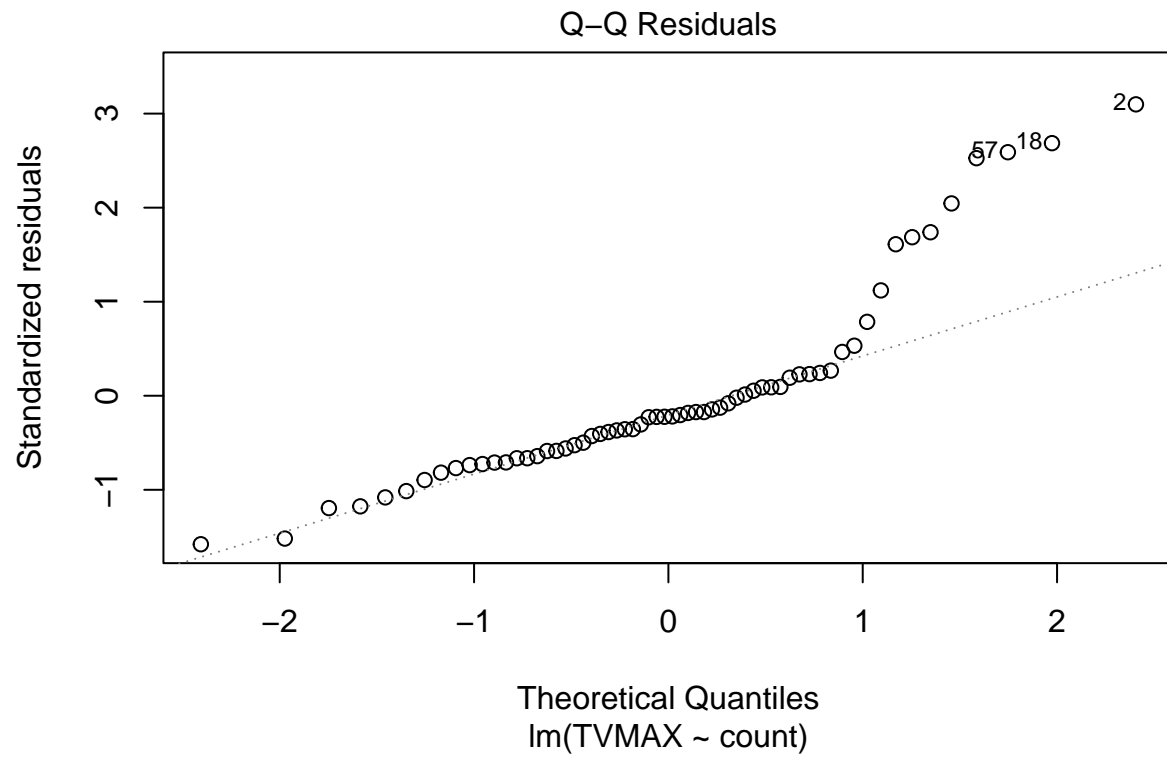


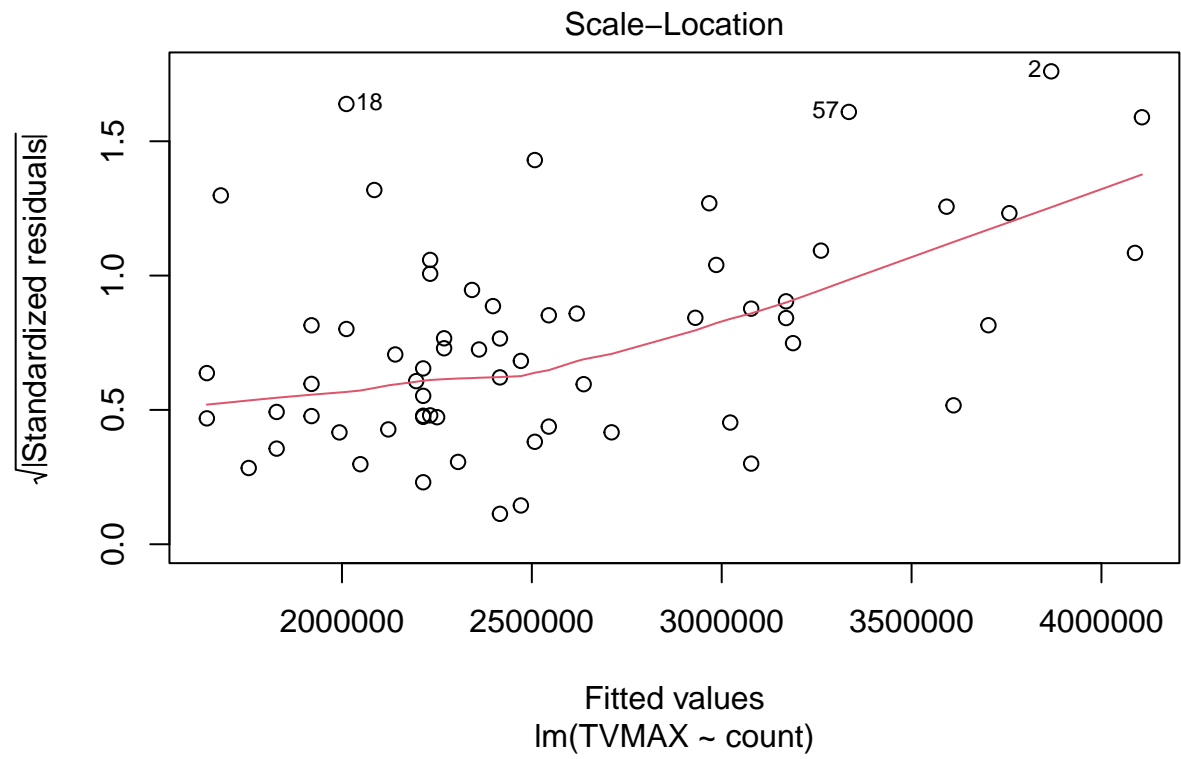
```
# Regression maximaler Wert
regression_kfz_max <- lm(TVMAX ~ count, data = aggregated_kfz)
summary(regression_kfz_max)
```

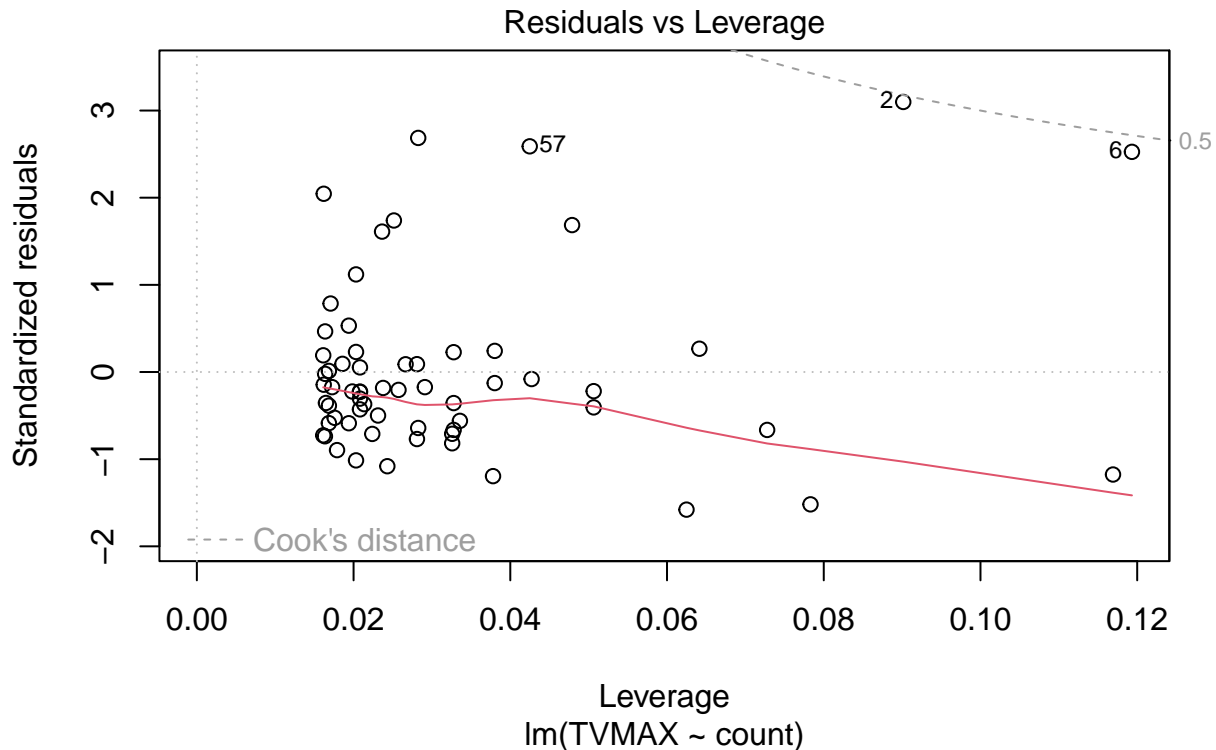
```
##
## Call:
## lm(formula = TVMAX ~ count, data = aggregated_kfz)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2376963 -964248 -338433  334884  4595548
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1625808     354780   4.583 2.38e-05 ***
## count         18377       5883    3.124  0.00275 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1555000 on 60 degrees of freedom
## (1 Beobachtung als fehlend gelöscht)
## Multiple R-squared:  0.1399, Adjusted R-squared:  0.1255
## F-statistic: 9.756 on 1 and 60 DF, p-value: 0.002751
```

```
plot(regression_kfz_max)
```









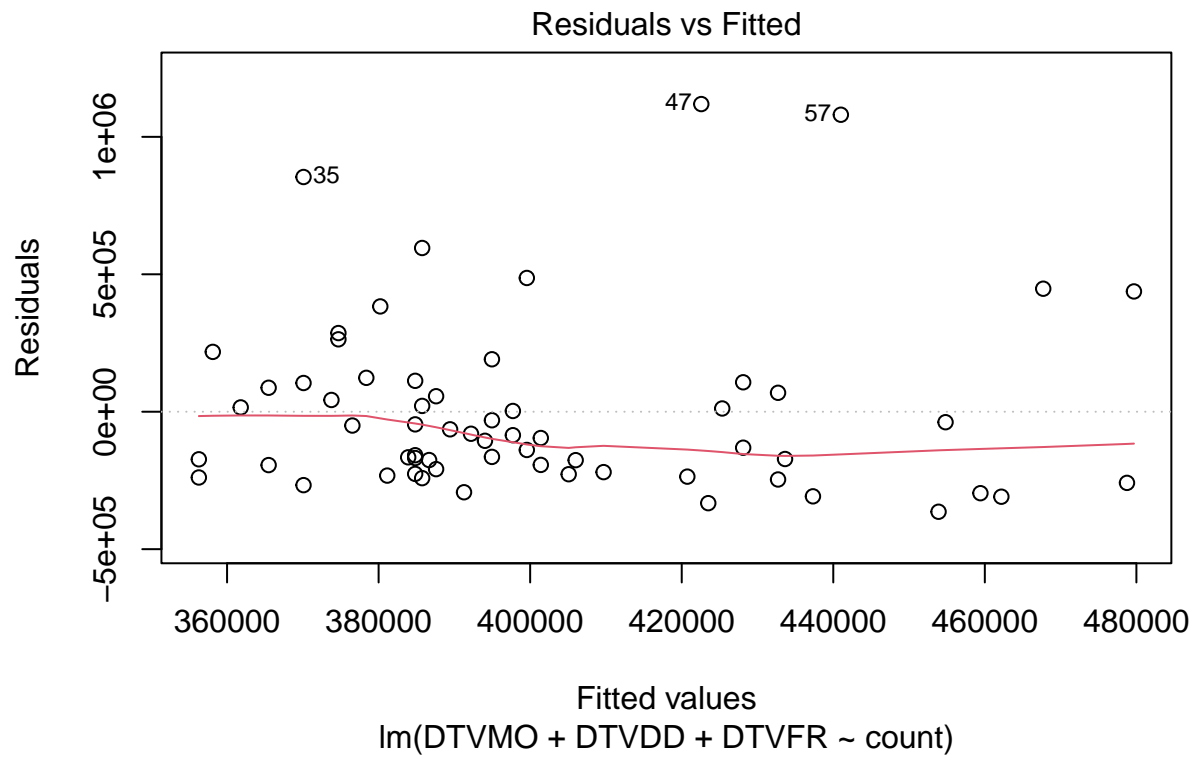
Regressionen Lkw

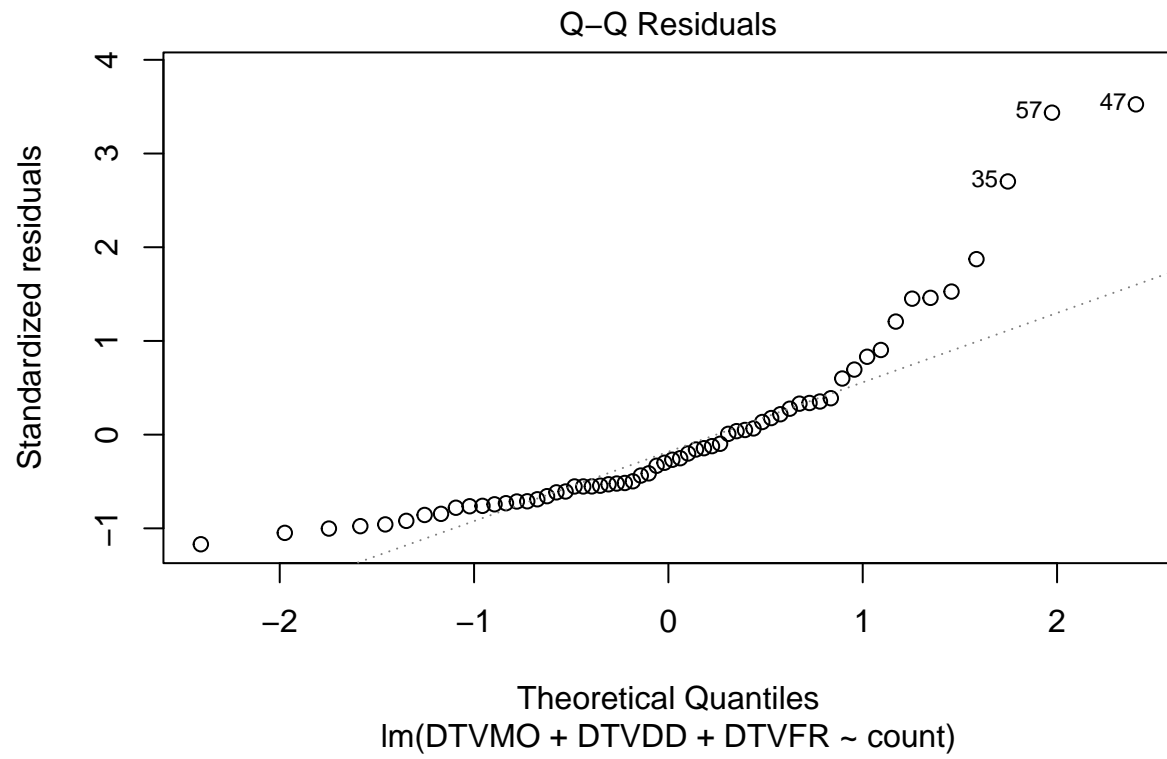
```
# Regression Lkw Woche (MO-FR)
regression_lkw_woche <- lm(DTVM0 + DTVDD + DTVFR ~ count, data = aggregated_lkw)
summary(regression_lkw_woche)
```

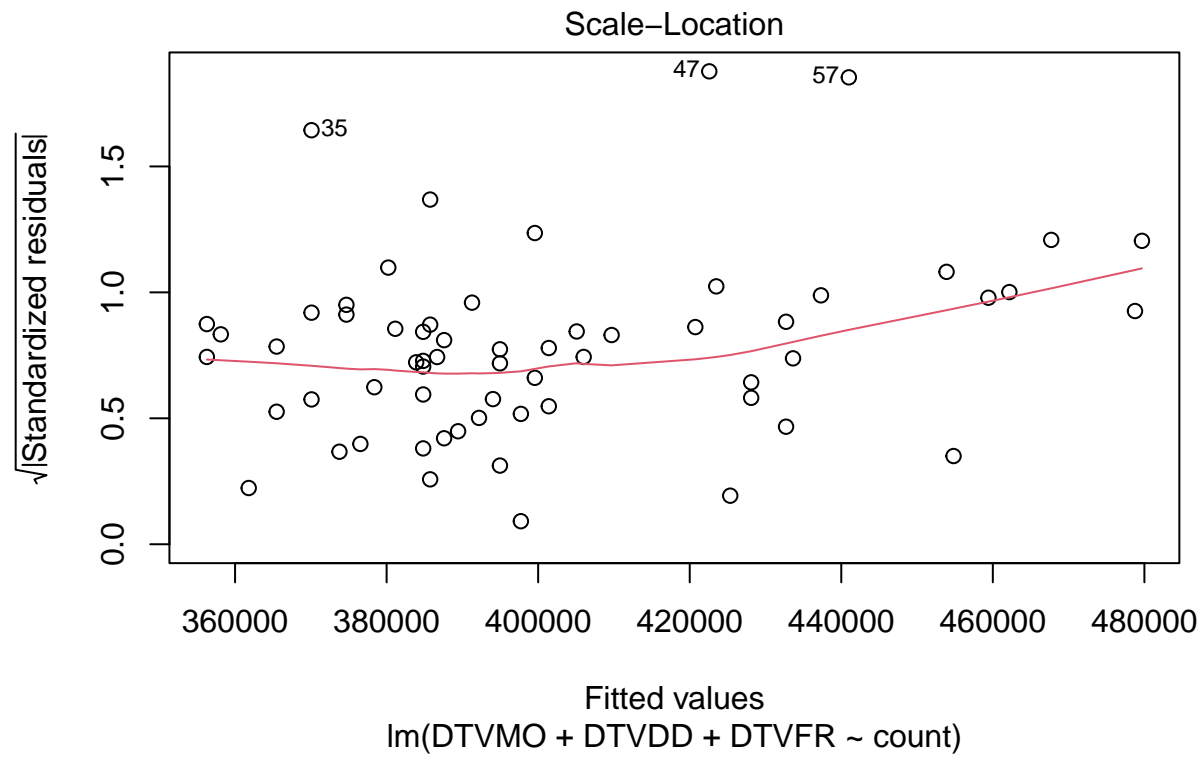
```
##
## Call:
## lm(formula = DTVM0 + DTVDD + DTVFR ~ count, data = aggregated_lkw)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -363879 -217054  -90495  100154 1119168
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 355355.8     73314.4   4.847 9.23e-06 ***
## count         920.9       1215.8   0.757  0.452
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 321300 on 60 degrees of freedom
## (1 Beobachtung als fehlend gelöscht)
## Multiple R-squared:  0.009472, Adjusted R-squared: -0.007037
## F-statistic: 0.5737 on 1 and 60 DF, p-value: 0.4517
```

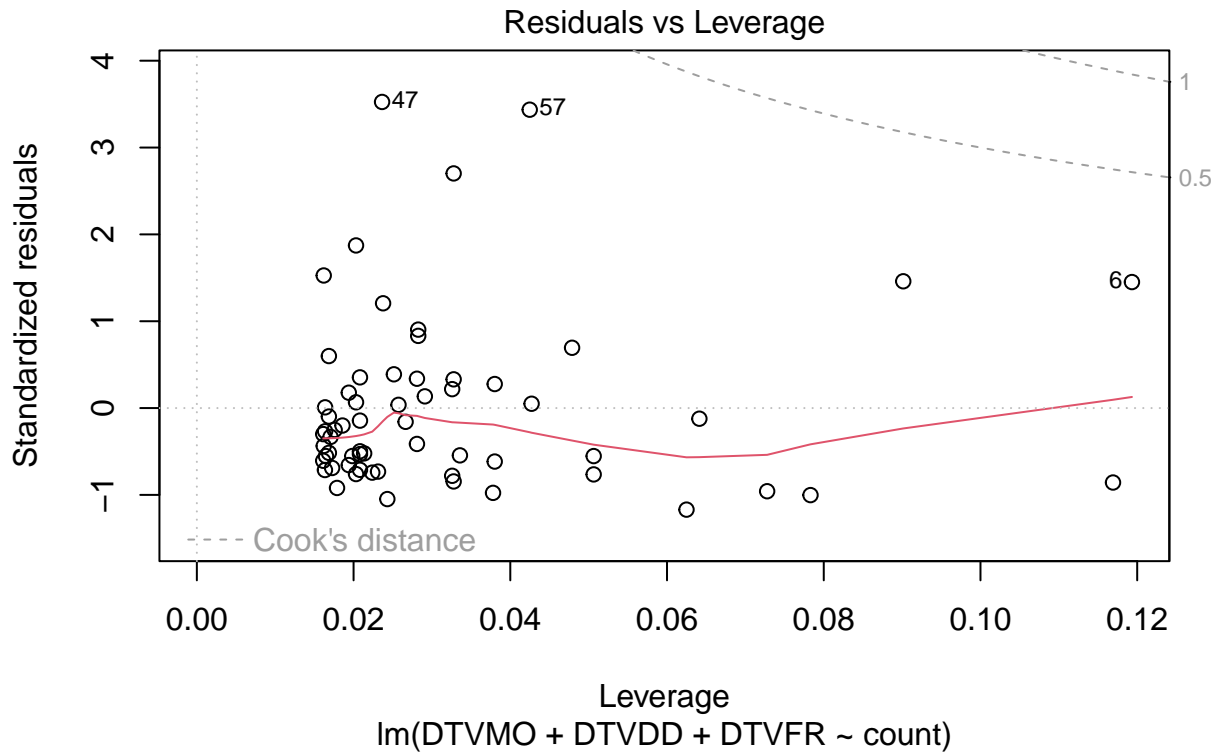


```
# Traffic count ist nicht abhängig von den Haltestellen  
plot(regression_lkw_woche)
```







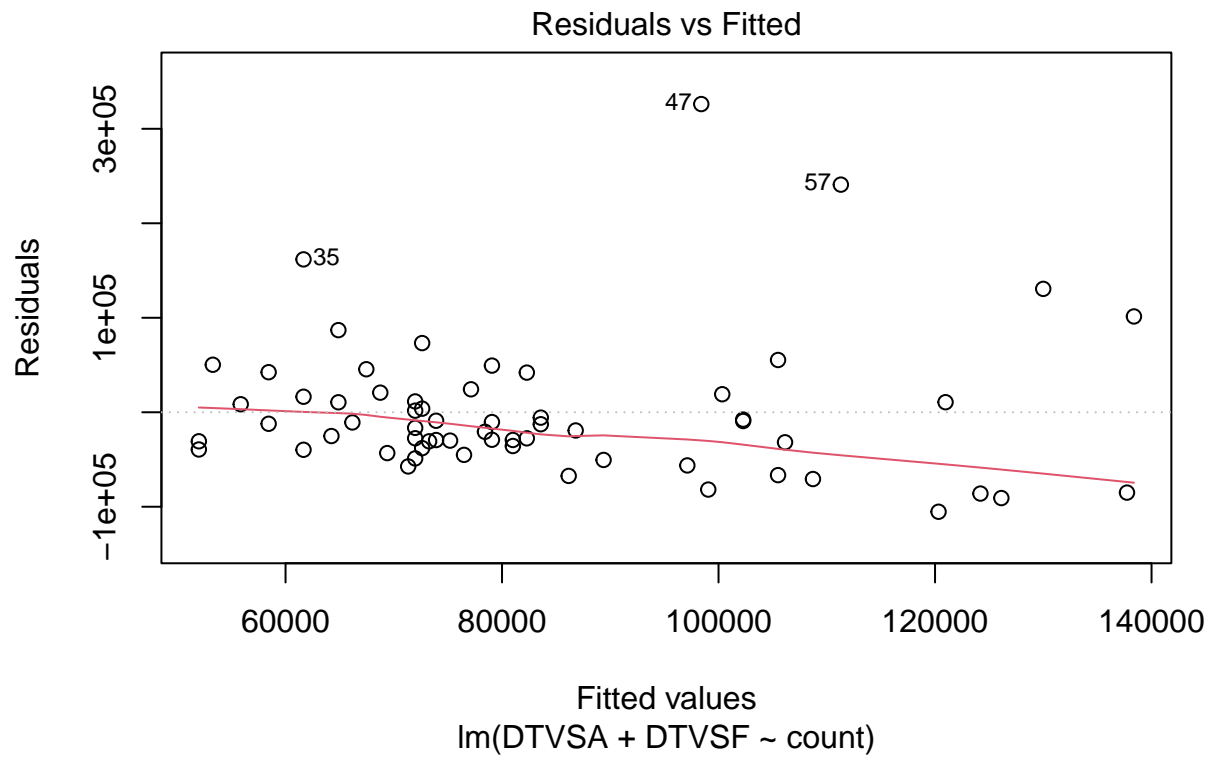


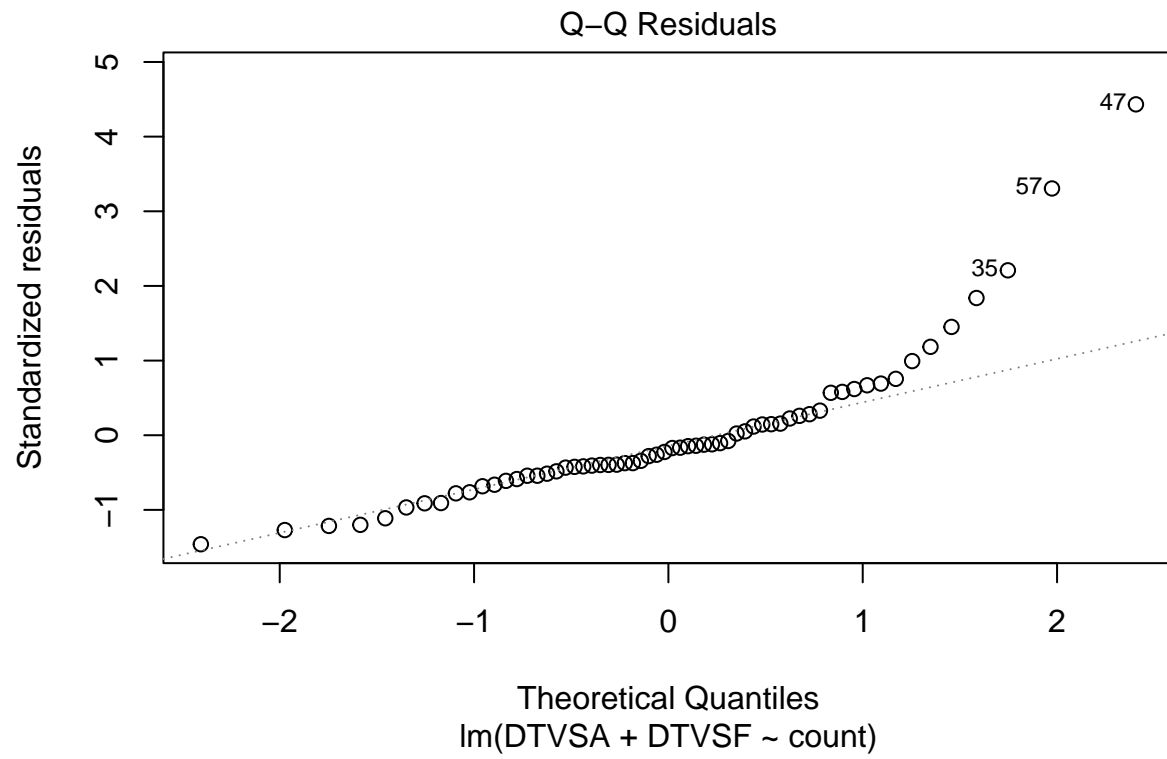
```
# Regression Lkw Wochenende
```

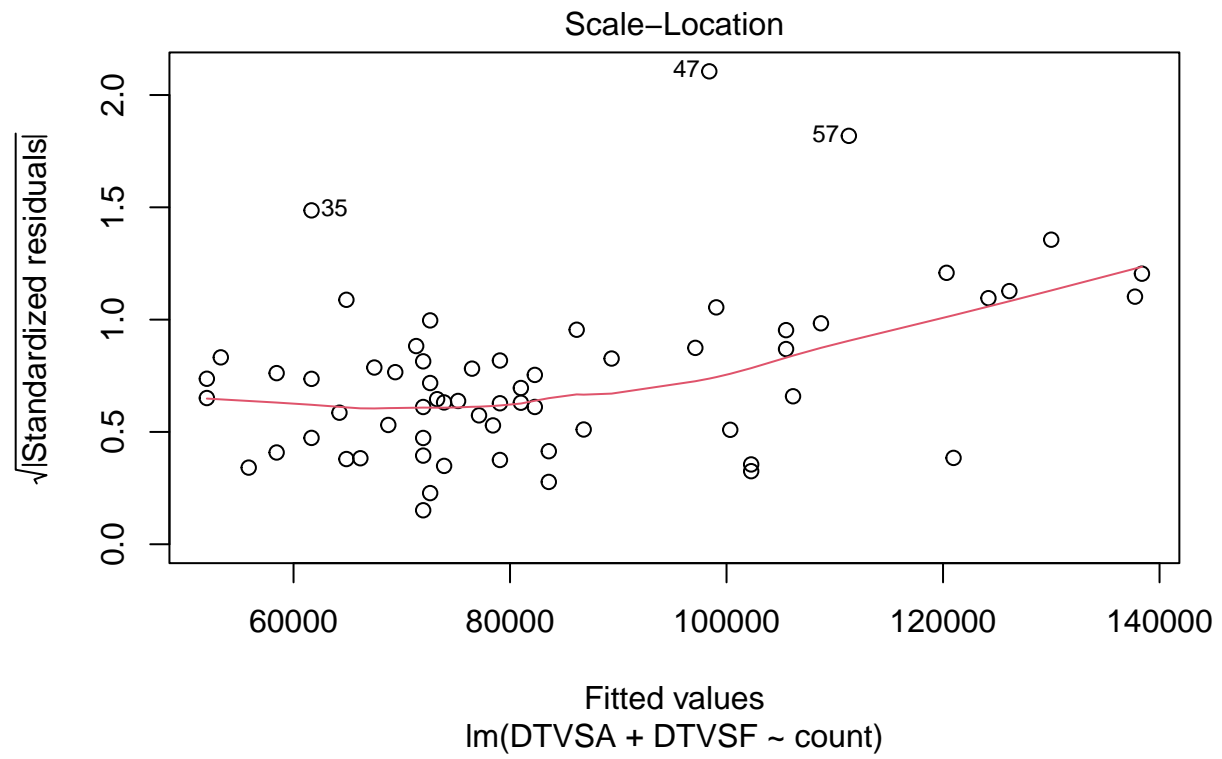
```
regression_lkw_wochenende <- lm(DTVSA + DTVSF ~ count, data = aggregated_lkw)
summary(regression_lkw_wochenende)
```

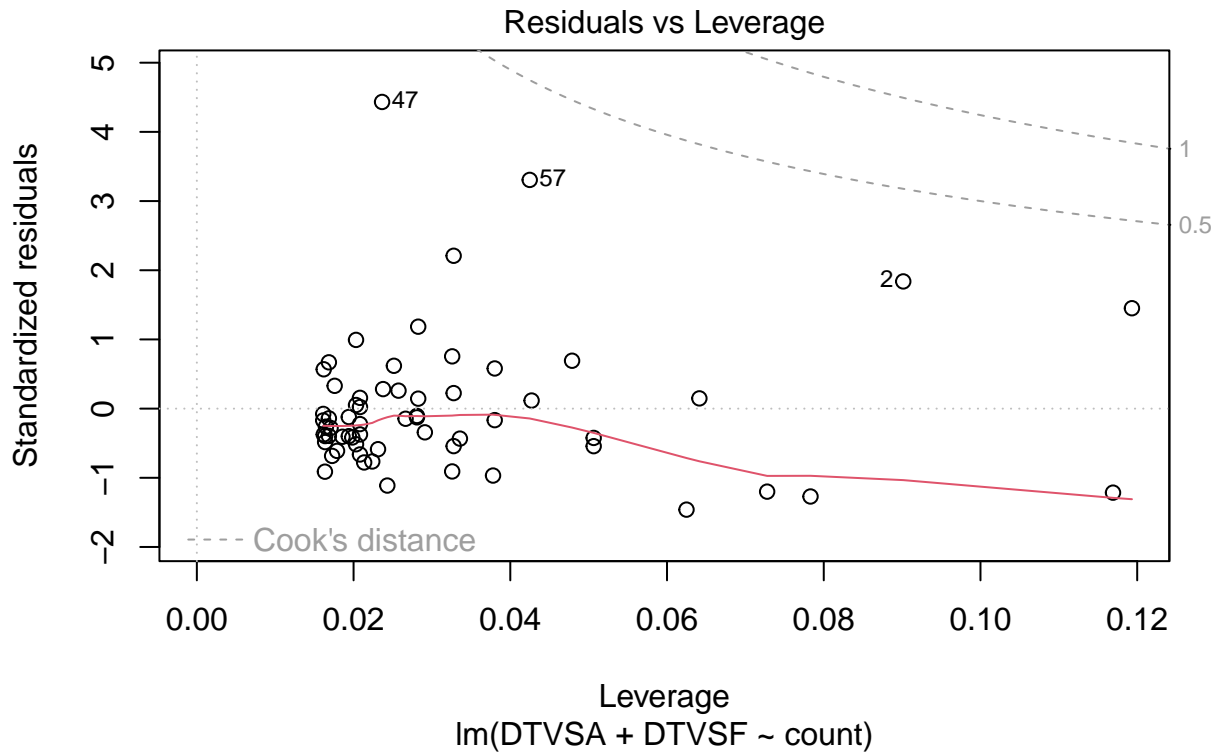
```
##
## Call:
## lm(formula = DTVSA + DTVSF ~ count, data = aggregated_lkw)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -105286  -39032  -14581   18418   326154
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   51346.5    16990.9     3.022  0.00369 **
## count          644.7       281.8     2.288  0.02568 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 74470 on 60 degrees of freedom
## (1 Beobachtung als fehlend gelöscht)
## Multiple R-squared:  0.08024,    Adjusted R-squared:  0.06492
## F-statistic: 5.235 on 1 and 60 DF,  p-value: 0.02568
```

```
# Traffic count ist abhängig von den Haltestellen  
plot(regression_lkw_wochenende)
```









```
# Regression maximaler Wert
regression_lkw_max <- lm(TVMAX ~ count, data = aggregated_lkw)
summary(regression_lkw_max) # nicht signifikant
```

```
##
## Call:
## lm(formula = TVMAX ~ count, data = aggregated_lkw)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -135129  -81311  -36080   37298  426372
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 138291.2    27669.8     4.998 5.33e-06 ***
## count         356.5       458.9     0.777   0.44
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 121300 on 60 degrees of freedom
## (1 Beobachtung als fehlend gelöscht)
## Multiple R-squared:  0.009959, Adjusted R-squared:  -0.006542
## F-statistic: 0.6035 on 1 and 60 DF, p-value: 0.4403
```



```
plot(regression_lkw_max)
```

