

# Replikations-Skript zur Vorlesung 11: Logit- und Probit-Modelle

Claudius Gräbner

KW 4 2021

## Contents

1	Motivation	1
2	Schätzung der Modelle	3
3	Interpretation der Ergebnisse	4

In diesem Dokument werden alle Abbildungen und Tabellen aus der siebten Vorlesung repliziert. Dabei gebe ich der Info wegen *allen* R Code. Entsprechend sind bestimmt auch einige Befehle dabei, die Sie jetzt noch nicht kennen.

Zudem nehme ich an, dass im Arbeitsverzeichnis der Ordner `data/T11/` existiert und in diesem der Datensatz `nonlinmodels_schweizer-arbeit.csv` enthalten ist (dieser ist über die Repository zur Vorlesung verfügbar).

Folgende Pakete werden zudem in diesem Skript verwendet:

```
library(tidyverse)
library(data.table)
library(ggpubr)
library(latex2exp)
library(here)
library(icaeDesign)
```

Beachten Sie, dass das Paket `icaeDesign` nicht über die zentrale Paketverwaltung verfügbar ist. Es muss folgendermaßen installiert werden:

```
devtools::install_github("graebner/icaeDesign")
```

```
schweiz_al <- fread(here("data/T11/nonlinmodels_schweizer-arbeit.csv"),
  colClasses = c("double", rep("double", 5), "factor"))
```

## 1 Motivation

```
ols_nonlin <- ggplot(
  data = schweiz_al,
  mapping = aes(x=Einkommen_log, y=Arbeitslos, group=1)) +
  scale_x_continuous(limits = c(7, 14)) +
  labs(
    title = "Einkommen und Arbeitslosigkeit",
    y = "Arbeitslosigkeit",
    x = "Arbeitsunabh. Einkommen (log)" # ,
    # caption = "Daten: Kleiber & Zeilis (2008)."
  ) +
```

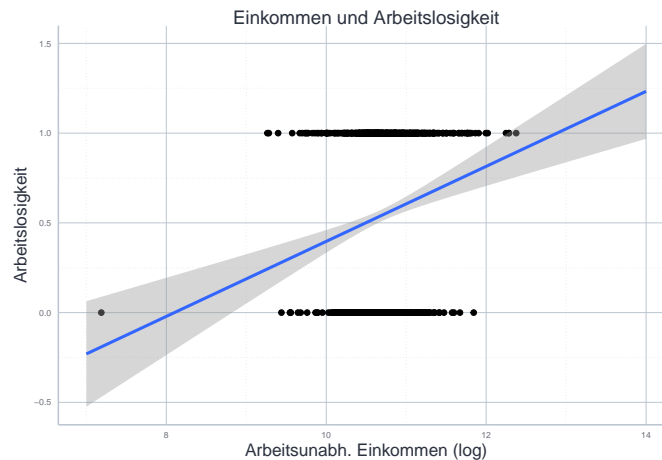
```

geom_point() + geom_smooth(method = "lm", fullrange=TRUE) +
theme_icae() +
theme(axis.title = element_text(size=12),
      plot.title = element_text(size=13)
)

ggsave(plot = ols_nonlin,
      filename = here("figures/T11/01_ols-nonlin.pdf"),
      width = 4, height = 3)

ols_nonlin

```



```

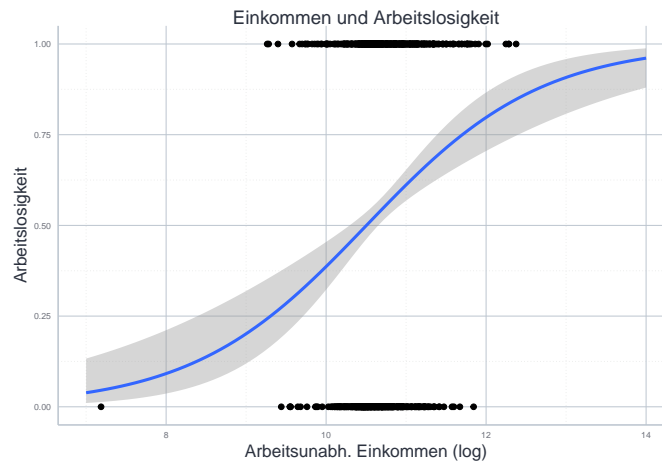
logit_expl <- ggplot(
  data = schweiz_al,
  mapping = aes(x=Einkommen_log, y=Arbeitslos, group=1)) +
  scale_x_continuous(limits = c(7, 14)) +
  labs(
    title = "Einkommen und Arbeitslosigkeit",
    y = "Arbeitslosigkeit",
    x = "Arbeitsunabh. Einkommen (log)" # ,
    # caption = "Daten: Kleiber & Zeilis (2008)."
  ) +
  geom_point() +
  geom_smooth(aes(y=Arbeitslos), method = "glm",
              method.args = list(family = "binomial"),
              fullrange=TRUE, se = TRUE) +

  theme_icae() +
  theme(axis.title = element_text(size=12),
        plot.title = element_text(size=13)
  )

ggsave(plot = logit_expl,
      filename = here("figures/T11/02_logit-expl.pdf"),
      width = 4, height = 3)

logit_expl

```



## 2 Schätzung der Modelle

Das *logit*-Modell:

```
arbeitslogit_test <- glm(
  Arbeitslos ~ Einkommen_log + Alter,
  family = binomial(link = "logit"),
  data = schweiz_al)
```

Das *probit*-Modell:

```
arbeitsprobit_test <- glm(
  Arbeitslos ~ Einkommen_log + Alter,
  family = binomial(link = "probit"),
  data = schweiz_al)
```

```
summary(arbeitslogit_test)
```

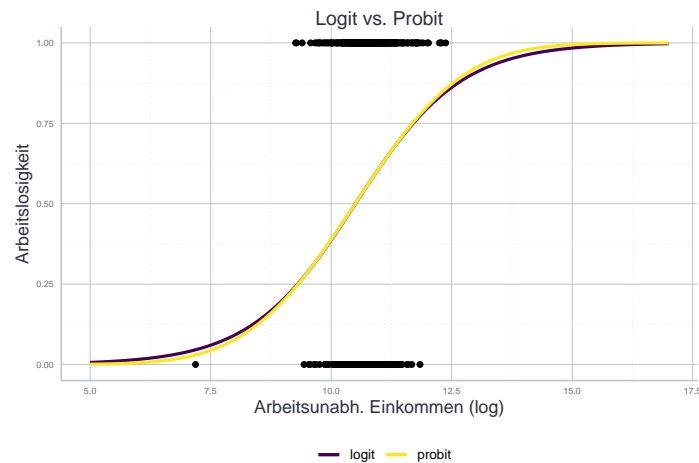
```
##
## Call:
## glm(formula = Arbeitslos ~ Einkommen_log + Alter, family = binomial(link = "logit"),
##      data = schweiz_al)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7448  -1.1855   0.8128   1.1017   1.8279
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -10.381739   2.003223  -5.183 2.19e-07 ***
## Einkommen_log  0.920045   0.185414   4.962 6.97e-07 ***
## Alter         0.018013   0.006612   2.724 0.00645 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1203.2  on 871  degrees of freedom
## Residual deviance: 1168.5  on 869  degrees of freedom
## AIC: 1174.5
```

```
##
## Number of Fisher Scoring iterations: 4

logit_probit <- ggplot(
  data = schweiz_al,
  mapping = aes(x=Einkommen_log, y=Arbeitslos, group=1)) +
  ggtitle("Logit vs. Probit") +
  ylab("Arbeitslosigkeit") + xlab("Arbeitsunabh. Einkommen (log)") +
  geom_point() + scale_x_continuous(limits = c(5, 17)) +
  scale_color_viridis_d() +
  geom_smooth(
    aes(y=Arbeitslos, color="logit"), method = "glm",
    method.args = list(family = binomial(link = "logit")),
    fullrange=TRUE, se = FALSE, alpha=0.5) +
  geom_smooth(
    aes(y=Arbeitslos, color="probit"), method = "glm",
    method.args = list(family = binomial(link = "probit")),
    fullrange=TRUE, se = FALSE, alpha=0.5) +
  theme_icae() +
  theme(axis.title = element_text(size=12),
        plot.title = element_text(size=13)
        )

ggsave(plot = logit_probit,
       filename = here("figures/T11/03_logit-probit.pdf"),
       width = 4, height = 3)

logit_probit
```



### 3 Interpretation der Ergebnisse

```
arbeitslogit <- glm(
  Arbeitslos ~ Einkommen_log + Alter + Ausbildung_Jahre + Kinder_jung +
  Kinder_alt + Auslaender,
  family = binomial(link = "logit"),
  data = schweiz_al)
summary(arbeitslogit)
```

```
##
```

```
## Call:
## glm(formula = Arbeitslos ~ Einkommen_log + Alter + Ausbildung_Jahre +
##      Kinder_jung + Kinder_alt + Auslaender, family = binomial(link = "logit"),
##      data = schweiz_al)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2681  -1.0675   0.5383   0.9727   1.9384
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -10.374346    2.166852  -4.788 1.69e-06 ***
## Einkommen_log    0.815041    0.205501   3.966 7.31e-05 ***
## Alter          0.051033    0.009052   5.638 1.72e-08 ***
## Ausbildung_Jahre -0.031728    0.029036  -1.093  0.275
## Kinder_jung     1.330724    0.180170   7.386 1.51e-13 ***
## Kinder_alt      0.021986    0.073766   0.298  0.766
## Auslaender1    -1.310405    0.199758  -6.560 5.38e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1203.2  on 871  degrees of freedom
## Residual deviance: 1052.8  on 865  degrees of freedom
## AIC: 1066.8
##
## Number of Fisher Scoring iterations: 4
```

```
predicted_probs <- predict(object = arbeitslogit,
  newdata = data.frame(
    "Einkommen_log" = c(10, 10),
    "Alter"=c(30, 30),
    "Ausbildung_Jahre" = c(5, 5),
    "Kinder_alt" = c(0, 0),
    "Kinder_jung"= c(1, 2),
    "Auslaender" = factor(c(0, 0))
  ),
  type = "response")
predicted_probs
```

```
##          1          2
## 0.6175431 0.8593445
```

```
diff(predicted_probs)
```

```
##          2
## 0.2418014
```

```
diff(
  predict(object = arbeitslogit,
    newdata = data.frame(
      "Einkommen_log" = c(10, 10),
      "Alter"=c(30, 30),
      "Ausbildung_Jahre" = c(5, 5),
      "Kinder_alt" = c(0, 0),
```

```
      "Kinder_jung"= c(1, 2),  
      "Auslaender" = factor(c(1, 1))  
    ),  
    type = "response")  
)
```

```
##           2  
## 0.3189543
```