3 Regressionsmodelle für Zähldaten

Aufgabe 21

Analyse des Datensatz ’bike’

```
bike <- read.table("bike.dat", header=T)
```

(a) siehe Übungsmitschrift

(b) Berechnung der Modelle

```
# Poisson-Modell
model1a <- glm(y ~ lecture + offset(log(time)), family=poisson, data=bike)
summary(model1a)

##
## Call:
## glm(formula = y ~ lecture + offset(log(time)), family = poisson,
##     data = bike)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4637  -0.5593  -0.1254   0.3634   2.0402
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.4543     0.3162 -14.087   <2e-16 ***
## lecture       0.9187     0.3907   2.352   0.0187 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 28.079  on 23  degrees of freedom
## Residual deviance: 22.194  on 22  degrees of freedom
## AIC: 63.503
##
## Number of Fisher Scoring iterations: 5

# äquivalente Schreibweise
model1b <- glm(y ~ lecture, offset=log(time), family=poisson(link=log),
               data=bike)
summary(model1b)

##
## Call:
## glm(formula = y ~ lecture, family = poisson(link = log), data = bike,
##     offset = log(time))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
```

```
## -2.4637  -0.5593  -0.1254   0.3634   2.0402
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -4.4543     0.3162 -14.087   <2e-16 ***
## lecture       0.9187     0.3907   2.352   0.0187 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 28.079  on 23  degrees of freedom
## Residual deviance: 22.194  on 22  degrees of freedom
## AIC: 63.503
##
## Number of Fisher Scoring iterations: 5
```

```
# Quasi-Poisson-Modell
model2 <- glm(y ~ lecture, offset=log(time), family=quasipoisson(link=log),
              data=bike)
summary(model2)
```

```
##
## Call:
## glm(formula = y ~ lecture, family = quasipoisson(link = log),
##     data = bike, offset = log(time))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4637  -0.5593  -0.1254   0.3634   2.0402
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.4543     0.3302 -13.490 4.07e-12 ***
## lecture       0.9187     0.4080   2.252   0.0346 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 1.090437)
##
##     Null deviance: 28.079  on 23  degrees of freedom
## Residual deviance: 22.194  on 22  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

Modell-Vergleich:

- phi auch bei Quasi-Poisson nahezu 1; keine wesentliche Änderung bei den Signifikanzen, auch hinsichtlich des p-Wertes wenig Unterschied. → Spricht für Poisson-Modell!
- dieselben Schätzer für Koeffizienten, da phi bei Schätzgleichung rausfliegt
- geringfügig größere Standardfehler im Quasi-Poisson-Fall
- Achtung: Da im Quasi-Poisson-Modell Vtlg eigentlich nicht bekannt ist und die Devianzen und AIC nicht bestimmt werden können, gibt die Funktion im Quasi-Poisson-Modell einfach die Devianzen des Poisson-Modells aus

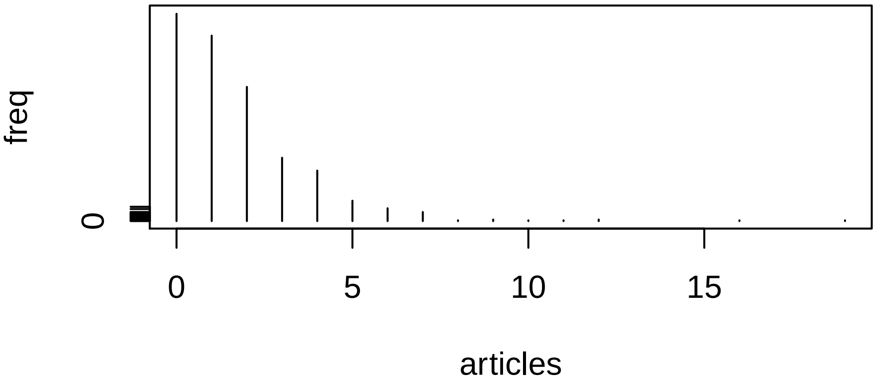
Aufgabe 22

Vorbereitungen

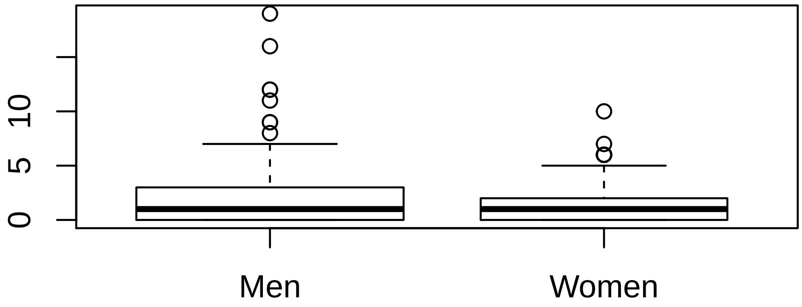
```
library(pscl)
data(bioChemists)
help(bioChemists)
attach(bioChemists)
```

(a) Deskriptive Analysen

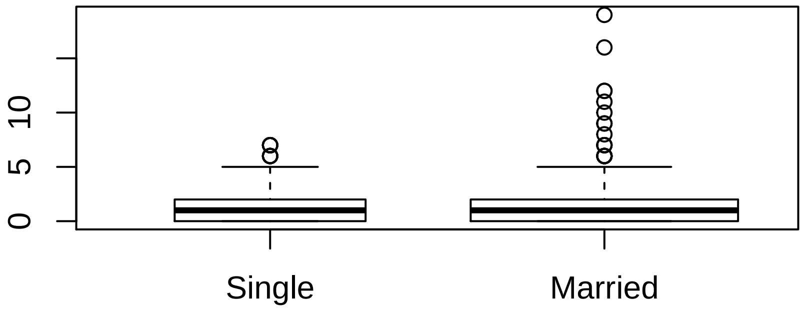
```
# Verteilung der Anzahl an Artikeln
plot(as.numeric(names(table(art))),table(art),type="h", xlab="articles",
ylab="freq")
```



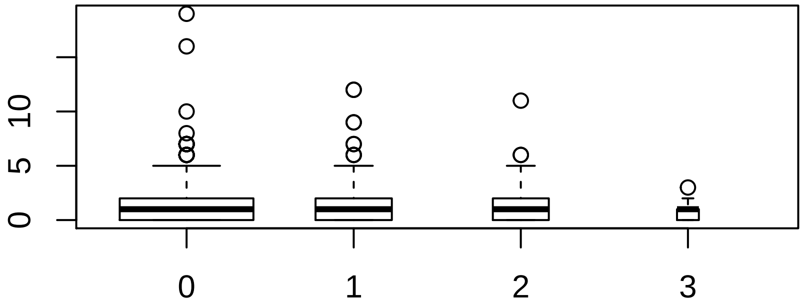
```
# nach Geschlecht
boxplot(art ~ fem, varwidth=T)
```



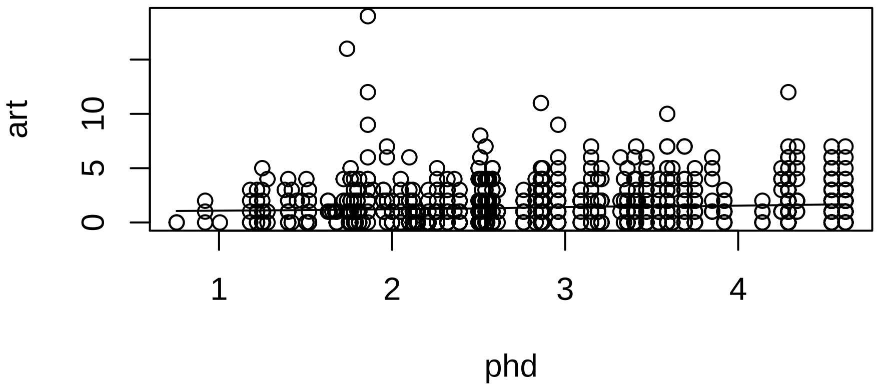
```
# nach Familienstand
boxplot(art ~ mar, varwidth=T)
```



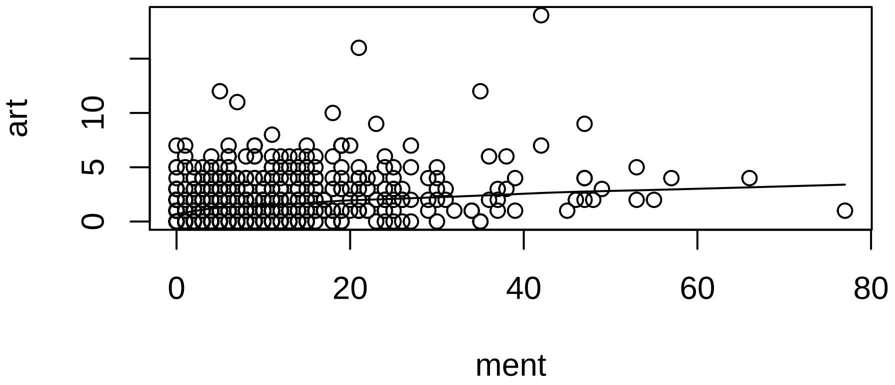
```
# nach Kinderzahl
boxplot(art ~ kid5, varwidth=T)
```



```
# nach Prestige
plot(phd,art)
lines(lowess(phd,art))
```



```
# nach Betreuer
plot(ment,art)
lines(lowess(ment,art))
```



(b) Äquidispersionseigenschaft fraglich, falls:

- Dispersionsparameter ϕ im Quasi-Poisson-Modell deutlich größer als 1
- Signifikanzen des Poisson-Modells und Quasi-Poissonmodells sehr unterschiedlich

(c) Bekannte Modelle

```
# Poisson-Modell
poissonModell <- glm(art ~ fem + mar + kid5 + phd + ment, family=poisson,
data=bioChemists)
summary(poissonModell)

##
## Call:
## glm(formula = art ~ fem + mar + kid5 + phd + ment, family = poisson,
##     data = bioChemists)
##
## Deviance Residuals:
```

```
##     Min       1Q   Median       3Q      Max
## -3.5672  -1.5398  -0.3660   0.5722   5.4467
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.304617   0.102981   2.958   0.0031 **
## femWomen    -0.224594   0.054613  -4.112 3.92e-05 ***
## marMarried   0.155243   0.061374   2.529   0.0114 *
## kid5        -0.184883   0.040127  -4.607 4.08e-06 ***
## phd          0.012823   0.026397   0.486   0.6271
## ment         0.025543   0.002006  12.733  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 1817.4  on 914  degrees of freedom
## Residual deviance: 1634.4  on 909  degrees of freedom
## AIC: 3314.1
##
## Number of Fisher Scoring iterations: 5
```

```r
# Quasi-Poisson-Modell
qpoissonModell <- glm(art ~ fem + mar + kid5 + phd + ment, family=quasipoisson,
                      data=bioChemists)
summary(qpoissonModell)
```

```
##
## Call:
## glm(formula = art ~ fem + mar + kid5 + phd + ment, family = quasipoisson,
##     data = bioChemists)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.5672  -1.5398  -0.3660   0.5722   5.4467
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.304617   0.139273   2.187 0.028983 *
## femWomen    -0.224594   0.073860  -3.041 0.002427 **
## marMarried   0.155243   0.083003   1.870 0.061759 .
## kid5        -0.184883   0.054268  -3.407 0.000686 ***
## phd          0.012823   0.035700   0.359 0.719544
## ment         0.025543   0.002713   9.415  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 1.829006)
##
##     Null deviance: 1817.4  on 914  degrees of freedom
## Residual deviance: 1634.4  on 909  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

(d) Negativbinomial-Modell

```r
negbinomial <- glm.nb(art ~ fem + mar + kid5 + phd + ment, link=log,
data=bioChemists)
summary(negbinomial)
```

```
##
## Call:
## glm.nb(formula = art ~ fem + mar + kid5 + phd + ment, data = bioChemists,
##     link = log, init.theta = 2.264387695)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1678  -1.3617  -0.2806   0.4476   3.4524
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.256144   0.137348   1.865 0.062191 .
## femWomen    -0.216418   0.072636  -2.979 0.002887 **
## marMarried   0.150489   0.082097   1.833 0.066791 .
## kid5        -0.176415   0.052813  -3.340 0.000837 ***
## phd          0.015271   0.035873   0.426 0.670326
## ment         0.029082   0.003214   9.048  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(2.2644) family taken to be 1)
##
##     Null deviance: 1109.0  on 914  degrees of freedom
## Residual deviance: 1004.3  on 909  degrees of freedom
## AIC: 3135.9
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  2.264
##           Std. Err.:  0.271
##
##  2 x log-likelihood:  -3121.917
```

(e) siehe Übungsmitschrift

(f) Zero-Inflated Poissonmodelle

```
# Poisson - Binomial(Logit-Link)
zeroinflPBl <- zeroinfl(art ~ fem + mar + kid5 + phd + ment, data=bioChemists,
dist="poisson", link="logit")
summary(zeroinflPBl)
```

```
##
## Call:
## zeroinfl(formula = art ~ fem + mar + kid5 + phd + ment, data = bioChemists,
##     dist = "poisson", link = "logit")
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -2.3253 -0.8652 -0.2826  0.5404  7.2976
##
## Count model coefficients (poisson with log link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.640838   0.121307   5.283 1.27e-07 ***
## femWomen    -0.209145   0.063405  -3.299 0.000972 ***
## marMarried   0.103751   0.071111   1.459 0.144565
## kid5        -0.143320   0.047429  -3.022 0.002513 **
## phd         -0.006166   0.031008  -0.199 0.842378
## ment         0.018098   0.002294   7.888 3.07e-15 ***
```

```
##
## Zero-inflation model coefficients (binomial with logit link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.577059   0.509386  -1.133  0.25728
## femWomen     0.109746   0.280082   0.392  0.69518
## marMarried  -0.354014   0.317611  -1.115  0.26502
## kid5         0.217097   0.196482   1.105  0.26919
## phd          0.001274   0.145263   0.009  0.99300
## ment        -0.134114   0.045243  -2.964  0.00303 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 21
## Log-likelihood: -1605 on 12 Df
```

```
# Poisson - Binomial(Probit-Link)
zeroinflPBp <- zeroinfl(art ~ fem + mar + kid5 + phd + ment, data=bioChemists,
dist="poisson", link="probit")
summary(zeroinflPBp)
```

```
##
## Call:
## zeroinfl(formula = art ~ fem + mar + kid5 + phd + ment, data = bioChemists,
##     dist = "poisson", link = "probit")
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -2.3222 -0.8686 -0.2879  0.5458  7.3162
##
## Count model coefficients (poisson with log link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.642393   0.122457   5.246 1.56e-07 ***
## femWomen    -0.207921   0.063701  -3.264  0.00110 **
## marMarried   0.105262   0.071304   1.476  0.13988
## kid5        -0.143343   0.047667  -3.007  0.00264 **
## phd         -0.007203   0.031351  -0.230  0.81829
## ment         0.018054   0.002318   7.788 6.82e-15 ***
##
## Zero-inflation model coefficients (binomial with probit link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.37233    0.29711  -1.253   0.2101
## femWomen     0.06241    0.16256   0.384   0.7011
## marMarried  -0.19094    0.18349  -1.041   0.2981
## kid5         0.12307    0.11560   1.065   0.2870
## phd         -0.00863    0.08709  -0.099   0.9211
## ment        -0.07128    0.02779  -2.565   0.0103 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 19
## Log-likelihood: -1605 on 12 Df
```

```
# Negativbinomial - Binomial(Logit-Link)
zeroinflNBl <- zeroinfl(art ~  fem + mar + kid5 + phd + ment, data=bioChemists,
dist="negbin", link="logit")
summary(zeroinflNBl)
```

```
##
## Call:
## zeroinfl(formula = art ~ fem + mar + kid5 + phd + ment, data = bioChemists,
##     dist = "negbin", link = "logit")
```

```
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.2942 -0.7601 -0.2909  0.4448  6.4155
##
## Count model coefficients (negbin with log link):
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.4167465  0.1435966   2.902  0.00371 **
## femWomen    -0.1955068  0.0755926  -2.586  0.00970 **
## marMarried   0.0975826  0.0844520   1.155  0.24789
## kid5        -0.1517325  0.0542061  -2.799  0.00512 **
## phd         -0.0007001  0.0362697  -0.019  0.98460
## ment         0.0247862  0.0034927   7.097 1.28e-12 ***
## Log(theta)   0.9763565  0.1354695   7.207 5.71e-13 ***
##
## Zero-inflation model coefficients (binomial with logit link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.19169    1.32282  -0.145  0.88478
## femWomen     0.63593    0.84892   0.749  0.45379
## marMarried  -1.49947    0.93867  -1.597  0.11017
## kid5         0.62843    0.44278   1.419  0.15582
## phd         -0.03771    0.30801  -0.122  0.90254
## ment        -0.88229    0.31623  -2.790  0.00527 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Theta = 2.6548
## Number of iterations in BFGS optimization: 43
## Log-likelihood: -1550 on 13 Df
```

```
# Negativbinomial - Binomial(Probit-Link)
zeroinflNBp <- zeroinfl(art ~  fem + mar + kid5 + phd + ment, data=bioChemists,
dist="negbin", link="probit")
summary(zeroinflNBp)
```

```
##
## Call:
## zeroinfl(formula = art ~ fem + mar + kid5 + phd + ment, data = bioChemists,
##     dist = "negbin", link = "probit")
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.2917 -0.7574 -0.2928  0.4407  6.4220
##
## Count model coefficients (negbin with log link):
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.4112993  0.1430685   2.875  0.00404 **
## femWomen    -0.1952483  0.0754646  -2.587  0.00967 **
## marMarried   0.0965461  0.0844185   1.144  0.25276
## kid5        -0.1508128  0.0541780  -2.784  0.00538 **
## phd         -0.0006429  0.0362825  -0.018  0.98586
## ment         0.0250045  0.0034921   7.160 8.05e-13 ***
## Log(theta)   0.9650591  0.1336851   7.219 5.24e-13 ***
##
## Zero-inflation model coefficients (binomial with probit link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.13866    0.79488  -0.174  0.86152
## femWomen     0.39173    0.49600   0.790  0.42966
## marMarried  -0.91748    0.58479  -1.569  0.11667
## kid5         0.39794    0.26732   1.489  0.13658
## phd         -0.02075    0.18455  -0.112  0.91049
## ment        -0.52928    0.19040  -2.780  0.00544 **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Theta = 2.6249
## Number of iterations in BFGS optimization: 28
## Log-likelihood: -1550 on 13 Df
```

```
# In den zwei Modellierungen können auch unterschiedliche Kovariablen
# verwendet werden:
zeroinflPBl2 <- zeroinfl(art ~ fem + mar + kid5 + phd + ment|fem + mar,
data=bioChemists, dist="poisson", link="logit")
summary(zeroinflPBl2)
```

```
##
## Call:
## zeroinfl(formula = art ~ fem + mar + kid5 + phd + ment | fem + mar, data = bioChemists,
##     dist = "poisson", link = "logit")
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.5654 -0.9899 -0.3074  0.5417  7.3201
##
## Count model coefficients (poisson with log link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.566465   0.115933   4.886 1.03e-06 ***
## femWomen    -0.228403   0.065131  -3.507 0.000453 ***
## marMarried   0.115022   0.071786   1.602 0.109094
## kid5        -0.171474   0.043293  -3.961 7.47e-05 ***
## phd          0.002041   0.028566   0.071 0.943027
## ment         0.021579   0.002163   9.976  < 2e-16 ***
##
## Zero-inflation model coefficients (binomial with logit link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.56508    0.29820  -5.248 1.53e-07 ***
## femWomen     0.02474    0.31461   0.079    0.937
## marMarried  -0.18758    0.31734  -0.591    0.554
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 20
## Log-likelihood: -1621 on 9 Df
```

(g) siehe Übungsmitschrift

(h) Hurdlemodelle

```
# Poisson - Poisson
hurdlePP <- hurdle(art ~ fem + mar + kid5 + phd + ment, data=bioChemists,
dist="poisson", zero.dist="poisson")
summary(hurdlePP)
```

```
##
## Call:
## hurdle(formula = art ~ fem + mar + kid5 + phd + ment, data = bioChemists,
##     dist = "poisson", zero.dist = "poisson")
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -2.4271 -0.9098 -0.2885  0.5515  7.0065
##
## Count model coefficients (truncated poisson with log link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.67114    0.12246   5.481 4.24e-08 ***
## femWomen    -0.22858    0.06522  -3.505 0.000457 ***
## marMarried   0.09649    0.07283   1.325 0.185209
## kid5        -0.14219    0.04845  -2.934 0.003341 **
## phd         -0.01273    0.03130  -0.407 0.684343
## ment         0.01875    0.00228   8.222  < 2e-16 ***
## Zero hurdle model coefficients (censored poisson with log link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.177848   0.175006  -1.016  0.30951
## femWomen    -0.131307   0.092409  -1.421  0.15534
## marMarried   0.195368   0.104390   1.872  0.06127 .
## kid5        -0.171740   0.066553  -2.581  0.00987 **
## phd          0.027041   0.046084   0.587  0.55735
## ment         0.036926   0.005631   6.558 5.46e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 22
## Log-likelihood: -1607 on 12 Df
```

```r
# Test, ob Koeffizientenvektoren unterhalb und oberhalb der Hürde gleich sind:
# H_0: \beta_1 = \beta_2 wird abgelehnt
hurdletest(hurdlePP)
```

```
## Wald test for hurdle models
##
## Restrictions:
## count_((Intercept) - zero_(Intercept) = 0
## count_femWomen - zero_femWomen = 0
## count_marMarried - zero_marMarried = 0
## count_kid5 - zero_kid5 = 0
## count_phd - zero_phd = 0
## count_ment - zero_ment = 0
##
## Model 1: restricted model
## Model 2: art ~ fem + mar + kid5 + phd + ment
##
##   Res.Df Df  Chisq Pr(>Chisq)
## 1    909
## 2    903  6 86.103  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
# Negativbinomial - Poisson
hurdleNP <- hurdle(art ~ fem + mar + kid5 + phd + ment, data=bioChemists,
dist="negbin", zero.dist="poisson")
summary(hurdleNP)
```

```
##
## Call:
## hurdle(formula = art ~ fem + mar + kid5 + phd + ment, data = bioChemists,
##     dist = "negbin", zero.dist = "poisson")
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.2694 -0.8247 -0.2552  0.4798  6.2988
##
```

```
## Count model coefficients (truncated negbin with log link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.355125   0.196832   1.804  0.07120 .
## femWomen    -0.244672   0.097218  -2.517  0.01184 *
## marMarried   0.103417   0.109430   0.945  0.34463
## kid5        -0.153260   0.072229  -2.122  0.03385 *
## phd         -0.002933   0.048067  -0.061  0.95134
## ment         0.023738   0.004287   5.537 3.07e-08 ***
## Log(theta)   0.603472   0.224995   2.682  0.00731 **
## Zero hurdle model coefficients (censored poisson with log link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.177848   0.175006  -1.016  0.30951
## femWomen    -0.131307   0.092409  -1.421  0.15534
## marMarried   0.195368   0.104390   1.872  0.06127 .
## kid5        -0.171740   0.066553  -2.581  0.00987 **
## phd          0.027041   0.046084   0.587  0.55735
## ment         0.036926   0.005631   6.558 5.46e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Theta: count = 1.8285
## Number of iterations in BFGS optimization: 25
## Log-likelihood: -1554 on 13 Df
```

```
# Poisson - Binomial
hurdlePB <- hurdle(art ~ fem + mar + kid5 + phd + ment, data=bioChemists,
dist="poisson", zero.dist="binomial", link="logit")
summary(hurdlePB)
```

```
##
## Call:
## hurdle(formula = art ~ fem + mar + kid5 + phd + ment, data = bioChemists,
##     dist = "poisson", zero.dist = "binomial", link = "logit")
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -2.4105 -0.8913 -0.2817  0.5530  7.0324
##
## Count model coefficients (truncated poisson with log link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.67114    0.12246   5.481 4.24e-08 ***
## femWomen    -0.22858    0.06522  -3.505 0.000457 ***
## marMarried   0.09649    0.07283   1.325 0.185209
## kid5        -0.14219    0.04845  -2.934 0.003341 **
## phd         -0.01273    0.03130  -0.407 0.684343
## ment         0.01875    0.00228   8.222  < 2e-16 ***
## Zero hurdle model coefficients (binomial with logit link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.23680    0.29552   0.801   0.4230
## femWomen    -0.25115    0.15911  -1.579   0.1144
## marMarried   0.32623    0.18082   1.804   0.0712 .
## kid5        -0.28525    0.11113  -2.567   0.0103 *
## phd          0.02222    0.07956   0.279   0.7800
## ment         0.08012    0.01302   6.155 7.52e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 12
## Log-likelihood: -1605 on 12 Df
```

```
# Negativbinomial - Binomial
hurdleNB <- hurdle(art ~ fem + mar + kid5 + phd + ment, data=bioChemists,
```

```
dist="negbin", zero.dist="binomial", link="logit")
summary(hurdleNB)

##
## Call:
## hurdle(formula = art ~ fem + mar + kid5 + phd + ment, data = bioChemists,
##     dist = "negbin", zero.dist = "binomial", link = "logit")
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.2581 -0.8036 -0.2497  0.4745  6.2753
##
## Count model coefficients (truncated negbin with log link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.355125   0.196832   1.804  0.07120 .
## femWomen    -0.244672   0.097218  -2.517  0.01184 *
## marMarried   0.103417   0.109430   0.945  0.34463
## kid5        -0.153260   0.072229  -2.122  0.03385 *
## phd         -0.002933   0.048067  -0.061  0.95134
## ment         0.023738   0.004287   5.537 3.07e-08 ***
## Log(theta)   0.603472   0.224995   2.682  0.00731 **
## Zero hurdle model coefficients (binomial with logit link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.23680    0.29552   0.801   0.4230
## femWomen    -0.25115    0.15911  -1.579   0.1144
## marMarried   0.32623    0.18082   1.804   0.0712 .
## kid5        -0.28525    0.11113  -2.567   0.0103 *
## phd          0.02222    0.07956   0.279   0.7800
## ment         0.08012    0.01302   6.155 7.52e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Theta: count = 1.8285
## Number of iterations in BFGS optimization: 15
## Log-likelihood: -1553 on 13 Df
```

```
# Ober- und unterhalb der Schranke können auch unterschiedliche Kovariablen
# verwendet werden:
HurdlePP2 <- hurdle(art ~ fem + mar + kid5 + phd + ment|fem + mar,
data=bioChemists, dist="poisson", zero.dist="poisson")
summary(HurdlePP2)

##
## Call:
## hurdle(formula = art ~ fem + mar + kid5 + phd + ment | fem + mar, data = bioChemists,
##     dist = "poisson", zero.dist = "poisson")
##
## Pearson residuals:
##     Min      1Q  Median      3Q     Max
## -1.1758 -1.0208 -0.3143  0.5019  6.9170
##
## Count model coefficients (truncated poisson with log link):
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.67114    0.12246   5.481 4.24e-08 ***
## femWomen    -0.22858    0.06522  -3.505 0.000457 ***
## marMarried   0.09649    0.07283   1.325 0.185209
## kid5        -0.14219    0.04845  -2.934 0.003341 **
## phd         -0.01273    0.03130  -0.407 0.684343
## ment         0.01875    0.00228   8.222  < 2e-16 ***
## Zero hurdle model coefficients (censored poisson with log link):
```

```
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.22013    0.09136   2.410    0.016 *
## femWomen    -0.14093    0.08737  -1.613    0.107
## marMarried   0.04443    0.09227   0.482    0.630
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Number of iterations in BFGS optimization: 21
## Log-likelihood: -1638 on 9 Df
```

(i) Wie kann man allgemein Modelle für Zähldaten vergleichen?

- Residuenanalyse
- mit Devianz nicht, da keine hierarchische Struktur
- AIC (problematisch wenn Likelihood nicht bekannt (Quasi-Poisson))
- am besten um Prognosegüte der Modelle zu überprüfen: Kreuzvalidierung nur wie vergleicht man dann Prognose und wahre Werte?
  - Trefferrate: nicht ideal, da auch wichtig, wie weit der Fit abweicht
  - quadratische Abweichung: nicht ideal, da Zähldaten ganzzahlig und Fit reellwertig
  - Kullback-Leibler-Distanz zwischen wahrer Vtlg (Einpunktvtlg) und geschätzter Vtlg
  - kumulierter quadrierter Score

```
# hier am leichtesten zu bestimmen: AIC
AIC(poissonModell,negbinomial,hurdlePP,hurdleNP,hurdlePB,hurdleNB,
zeroinflPBl,zeroinflPBp,zeroinflNBp,zeroinflPBl2)
```

```
##                df      AIC
## poissonModell  6 3314.113
## negbinomial    7 3135.917
## hurdlePP      12 3238.408
## hurdleNP      13 3134.978
## hurdlePB      12 3234.623
## hurdleNB      13 3131.193
## zeroinflPBl   12 3233.546
## zeroinflPBp   12 3234.944
## zeroinflNBp   13 3125.782
## zeroinflPBl2   9 3259.191
```