

Introduction to the Generalized Linear Model

Logistic and Poisson Regression

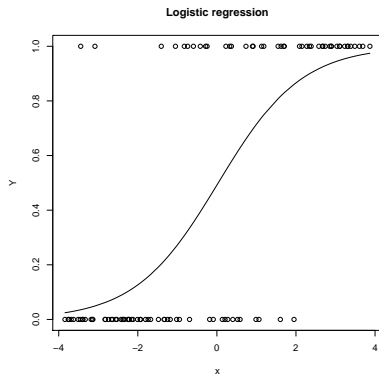
André Meichtry

School of Health Professions
Zurich University of Applied Sciences ZHAW

2020

Generalized Linear Model (GLM)

- ▶ We want to generalize the linear model to discrete or continuous outcomes
- ▶ Dichotomous event outcome, leading to **Logistic regression**
- ▶ Counts as outcome, leading to **Poisson regression**



Aspects of generalization

- ▶ Link function
- ▶ Variance function
- ▶ Distribution of the exponential family

Link function

- ▶ **Systematic part:** The expectation of the response, $\mu_i = E(Y_i)$, is transformed with a **link function**.
- ▶ This quantity is a linear function of the parameter β_j , called the **linear predictor** η_i with the link function $h(\cdot)$

$$\boxed{h(E(Y_i)) = h(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}.} \quad (1)$$

- ▶ Usual Link functions
 - ▶ Linear regression: Identity function
 - ▶ Logistic regression: logit function
 - ▶ Poisson regression: log function

Variance function

- ▶ **Random part:** The variance $\text{Var}(Y_i)$ is a function of the expectation,

$$\boxed{\text{Var}(Y_i) = \phi v(\mu_i),} \quad (2)$$

where $v(\cdot)$ is the **variance function** and ϕ is the **dispersion parameter**, which has to be estimated or not.

- ▶ Usual variance functions
 - ▶ Linear regression: $v(\mu_i) = 1$ with $\phi = \sigma^2$
 - ▶ Logistic regression: $v(\mu_i) = \mu_i(1 - \mu_i)$ and $\phi = 1$
 - ▶ Poisson regression: $v(\mu_i) = \mu_i$ and $\phi = 1$

Distributions

- ▶ Each class of a GLM follow a model with density of the **exponential family**¹. Special cases and most often used are:
- ▶ The **Normal distribution**
- ▶ The **Binomial distribution**
- ▶ The **Poisson distribution**

¹distributions of the exponential family have densities $f(y_i) = \exp\{(y_i\theta_i - A(\theta_i))/\phi + B(y_i, \phi)\}$ with θ_i as the **canonical parameter** with θ_i be expressed through μ_i . One can show that $E(Y_i) = \mu_i = A'(\theta_i)$ and $\text{Var}(Y_i) = A''(\theta_i)\phi$. The function $A(\cdot)$ fixes the exponential family, $B(\cdot)$ is a normalizing function.

Recap: Linear Model

- ▶ $Y_i \sim \mathcal{N}(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma^2)$.
- ▶ The model can be written: $Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i$.
- ▶ The expectation μ_i is

$$\mu_i = \mathbb{E}(Y_i) = \mathbf{x}_i^T \boldsymbol{\beta}. \quad (3)$$

- ▶ The link function $h(\cdot)$ is the identity and the variance function is $v(\mu_i) = 1$, the dispersion parameter is known, $\phi = \sigma^2$.
- ▶ Interpretation: β_j is the difference in expectations for two subpopulations that differ on x_j by one unit (slope).

Recap: Linear Model for Fertility

► Estimation with least squares: `lm()`

```
modlm<-lm(Fertility~.,swiss)
modlm0<-lm(Fertility~1,swiss) ##for later
summary(modlm)

##
## Call:
## lm(formula = Fertility ~ ., data = swiss)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.2743  -5.2617   0.5032   4.1198  15.3213
##
## Coefficients:
##              Estimate Std. Error t value    Pr(>|t|)
## (Intercept)    66.91518    10.70604     6.250 0.000000191
## Agriculture    -0.17211     0.07030    -2.448   0.01873
## Examination    -0.25801     0.25388    -1.016   0.31546
## Education      -0.87094     0.18303    -4.758 0.000024306
## Catholic        0.10412     0.03526     2.953   0.00519
## Infant.Mortality 1.07705     0.38172     2.822   0.00734
##
## Residual standard error: 7.165 on 41 degrees of freedom
## Multiple R-squared:  0.7067, Adjusted R-squared:  0.671
## F-statistic: 19.76 on 5 and 41 DF,  p-value: 0.000000005594
```


The same model as GLM

- Estimation with maximum likelihood: `glm()`

```
modglm<-glm(Fertility~.,swiss,family=gaussian)
modglm0<-glm(Fertility~1,swiss,family=gaussian)## for later
summary(modglm)

##
## Call:
## glm(formula = Fertility ~ ., family = gaussian, data = swiss)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -15.2743  -5.2617   0.5032   4.1198  15.3213
##
## Coefficients:
##              Estimate Std. Error t value    Pr(>|t|)
## (Intercept)    66.91518   10.70604    6.250 0.000000191
## Agriculture    -0.17211    0.07030   -2.448  0.01873
## Examination   -0.25801    0.25388   -1.016  0.31546
## Education      -0.87094    0.18303   -4.758 0.000024306
## Catholic        0.10412    0.03526    2.953  0.00519
## Infant.Mortality 1.07705    0.38172    2.822  0.00734
##
## (Dispersion parameter for gaussian family taken to be 51.34251)
##
##      Null deviance: 7178  on 46  degrees of freedom
## Residual deviance: 2105  on 41  degrees of freedom
## AIC: 326.07
##
## Number of Fisher Scoring iterations: 2
```

What is different?

- ▶ “Deviance” versus Sum of Squares
- ▶ “Likelihood ratio tests” versus F -tests
- ▶ Explanations in next slides

```
anova(modglm,modglm0,test="LR")

## Analysis of Deviance Table
##
## Model 1: Fertility ~ Agriculture + Examination + Education + Catholic +
##   Infant.Mortality
## Model 2: Fertility ~ 1
##   Resid. Df Resid. Dev Df Deviance      Pr(>Chi)
## 1         41      2105
## 2         46      7178 -5  -5072.9 < 0.00000000000000022

anova(modlm,modlm0,test="F")

## Analysis of Variance Table
##
## Model 1: Fertility ~ Agriculture + Examination + Education + Catholic +
##   Infant.Mortality
## Model 2: Fertility ~ 1
##   Res.Df  RSS Df Sum of Sq    F        Pr(>F)
## 1      41 2105
## 2      46 7178 -5  -5072.9 19.761 0.0000000005594
```

Estimation and Tests

- ▶ Estimation via **Maximum Likelihood Principle**
- ▶ We look at the (logarithmic) probability of the data **as function of the parameters** and search the maximum.
- ▶ The log-likelihood l is

$$l(\beta) = \sum_{i=1}^n \log \Pr(Y_i = y_i \mid \mathbf{x}_i, \beta) \quad (4)$$

- ▶ The β that maximizes $l(\beta)$ is called the **Maximum Likelihood Estimate (MLE)** $\hat{\beta}$
- ▶ One can show that the MLE has an asymptotic normal distribution.

Estimation and Tests

- ▶ **Residual Deviance** replaces the **residual sum of squares** and is defined as

$$D = 2(l_{max} - l(\hat{\beta})) \quad (5)$$

- ▶ l_{max} is the likelihood for the “maximal”, the saturated model (one parameter for each observation i)
- ▶ **Null Deviance** replaces the **total sum of squares**

$$D = 2(l_{max} - l_0) \quad (6)$$

Estimation and Tests

- ▶ **Likelihood-Ratio-Test:** The difference in deviance

$$2(l_{Large} - l_{Small}) \quad (7)$$

- ▶ has an asymptotic chi-square distribution
 - ▶ with the difference of the number of parameters as degrees of freedom.
-
- ▶ H_0 : Modell small with p_{Small} parameters is true.
 - ▶ H_1 : Modell large with $p_{Large} > p_{Small}$ parameters is true.
 - ▶ $2(l_{Large} - l_{Small}) \overset{approx}{\sim} \chi^2_{p_{Large} - p_{Small}}$

Logistic regression

- ▶ The distribution of the Y_i is binomial,

$$Y_i \sim \text{Bin}(\mu_i = \pi_i, n = 1)$$

- ▶ The model for $\mu_i = \pi_i$ can be written

$$\text{logit}(\pi_i) = \mathbf{x}_i^T \boldsymbol{\beta}, \quad (8)$$

- ▶ The link function is $h(\pi_i) = \text{logit}(\pi_i) = \log(\pi_i/(1 - \pi_i)) = \log \text{odds}$
- ▶ The variance function $v(\pi) = \pi(1 - \pi)$ and $\phi = 1$.
- ▶ Interpretation: β_j (except for the intercept) is the difference in logits (**log odds ratio**) for two subpopulations that differ on x_j by one unit.
- ▶ $\exp(\beta_j)$ (except for the intercept) is the **odds ratio OR** for the event for two subpopulations that differ on x_j by one unit.

Example with one continuous predictor

► Simulate some data:

```
set.seed(10)
##Sample Size
N<-30
##predictor
x<-sort(runif(N,-5,5))
##parameters
alpha<-0
beta<-1
##linear predictor
eta<-alpha+x*beta
##linear predictor logit E(Y) = eta
##Inverse function of the logit function is the logistic function,
##logistic(eta)=E(Y)=mu (here mu is a probability, often written pi)
pi<-exp(eta)/(1+exp(eta))
##Draw N from binomial distribution with parameters pi and 1
Y<-rbinom(N,size=1,prob=pi)
```

Example with one continuous predictor

```
headTail(cbind(x,Y))
```

```
##      x      Y
## 1 -4.48      0
## 2 -4.15      0
## 3 -3.86      0
## 4 -2.75      0
## ...   ...   ...
## 27  2.75      1
## 28  3.36      1
## 29  3.38      1
## 30  3.65      1
```

```
mod0<-glm(Y~1,family="binomial")
mod<-glm(Y~x,family="binomial")
summary(mod)
```

```
##
## Call:
## glm(formula = Y ~ x, family = "binomial")
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6720  -0.5881  -0.1558   0.4501   2.2860
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -0.1250     0.5440  -0.230  0.81826
## x              1.0696     0.3597   2.974  0.00294
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 41.054  on 29  degrees of freedom
## Residual deviance: 21.906  on 28  degrees of freedom
## AIC: 25.906
##
## Number of Fisher Scoring iterations: 5
```


Wald-tests and LRT-Tests

- ▶ Tests of individual coefficients based on approximative normality are called **Wald**-tests.
- ▶ Crude assumption about the shape of the likelihood.
- ▶ The LRT takes the likelihood values as they are.
- ▶ Therefore LR-tests are usually superior to Wald-tests
- ▶ They are asymptotically equivalent.
- ▶ `confint()` constructs LR confidence intervals if a `glm`-object is given as argument.

```
anova(mod,mod0,test="LR")

## Analysis of Deviance Table
##
## Model 1: Y ~ x
## Model 2: Y ~ 1
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1         28      21.906
## 2         29      41.054 -1   -19.148 0.0000121

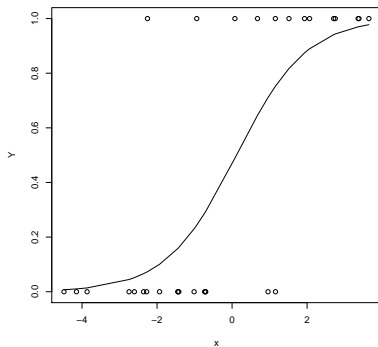
confint(mod)

## Waiting for profiling to be done...

##           2.5 %    97.5 %
## (Intercept) -1.2351381 0.9959719
## x           0.4936061 1.9548316
```

Model and data

```
pred<-predict(mod,type="response")  
plot(x,Y)  
lines(x,pred)
```



Residual analysis

What residuals are is not unambiguous:

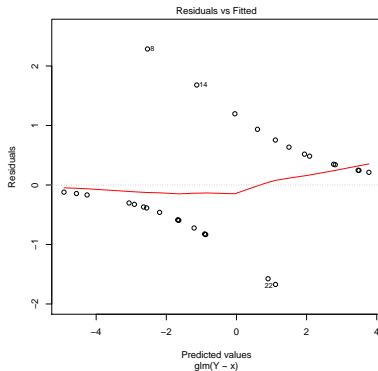
- ▶ Raw residuals (Response residuals) $R_i = Y_i - \hat{\pi}_i$
- ▶ Working residuals (transformed on the space of the linear predictor)
- ▶ Deviance residuals: $\text{sign}(Y_i - \hat{\pi}_i) \cdot \sqrt{d_i}$ with d_i as the contribution i to the deviance².
- ▶ Pearson residuals (Raw residuals divided by the standard deviation)

²would be equal to the square root of a squared residual in normal distribution.

Residual analysis

- ▶ Working residuals against linear predictor
- ▶ Response residuals against fitted values

```
plot(mod, which=1)
```



Example 2: HIV

```
d.hiv<-read.csv("https://raw.githubusercontent.com/mcdr65/PhDCareRehab/master/Data/HIV.csv")
#str(d.hiv)
f<-1:11
d.hiv[,f]<-lapply(d.hiv[,f],as.factor)
```

```
head(d.hiv)
```

##	id	age3	gender	race3	educ4	employment	disability	dep	anxpoms8	paindic	aids
## 1	201	1	1	4	3	0	0	0	<NA>	1	1
## 2	202	1	1	2	4	0	1	0	0	1	0
## 3	204	2	1	5	4	1	0	1	1	1	0
## 4	205	1	1	5	3	1	0	0	1	0	0
## 5	206	3	1	5	1	0	1	1	1	1	1
## 6	207	1	2	4	4	0	1	1	0	1	1

```
summary(d.hiv)
```

##	id	age3	gender	race3	educ4	employment	disability	dep	anxpoms8	paindic	aids
## 201	: 1	1: 86	1:216	2:122	1: 47	0:267	0 : 80	0:159	0 :163	0 :141	0:151
## 202	: 1	2:133	2: 77	4:129	2: 94	1: 49	1 :235	1:157	1 :145	1 :174	1:165
## 204	: 1	3: 97	3: 23	5: 65	3:107		NA's: 1		NA's: 8	NA's: 1	
## 205	: 1				4: 68						
## 206	: 1										
## 207	: 1										
## (Other)	:310										

Example 2: Marginal Wald tests

```
fit <- glm(aids~age3+gender+race3+educ4+employment+disability+dep+paindic, family="binomial",data=d.hiv)
summary(fit)

##
## Call:
## glm(formula = aids ~ age3 + gender + race3 + educ4 + employment +
##      disability + dep + paindic, family = "binomial", data = d.hiv)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.816  -1.095   0.652   1.008   2.063
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.960436   0.570402  -1.684   0.0922
## age32        0.191440   0.299411   0.639   0.5226
## age33        0.004589   0.325009   0.014   0.9887
## gender2     -0.253102   0.306745  -0.825   0.4093
## gender3     -0.603355   0.500796  -1.205   0.2283
## race34       0.560104   0.296623   1.888   0.0590
## race35       0.380728   0.338857   1.124   0.2612
## educ42      -0.127337   0.400751  -0.318   0.7507
## educ43       0.515187   0.393228   1.310   0.1901
## educ44      -0.010437   0.436651  -0.024   0.9809
## employment1 -0.210049   0.501418  -0.419   0.6753
## disability1  0.909227   0.406726   2.235   0.0254
## dep1        -0.451316   0.256693  -1.758   0.0787
## paindic1     0.411642   0.256803   1.603   0.1089
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 434.48  on 313  degrees of freedom
## Residual deviance: 399.67  on 300  degrees of freedom
## (2 observations deleted due to missingness)
## AIC: 427.67
##
## Number of Fisher Scoring iterations: 4
```

Example 2: Sequential LR tests

```
anova(fit,test="LR")

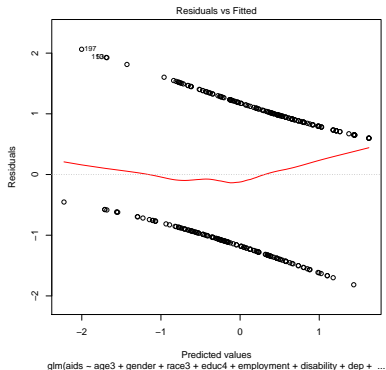
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: aids
##
## Terms added sequentially (first to last)
##
##
```

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
## NULL			313	434.48	
## age3	2	0.9382	311	433.54	0.625560
## gender	2	4.5941	309	428.95	0.100553
## race3	2	3.3381	307	425.61	0.188425
## educ4	3	7.0244	304	418.59	0.071125
## employment	1	9.6518	303	408.93	0.001892
## disability	1	4.7482	302	404.19	0.029330
## dep	1	1.9338	301	402.25	0.164346
## paidindc	1	2.5849	300	399.67	0.107885

- ▶ One could proceed with different model comparisons.

Example 2: Tukey-Anscombe Plot

```
plot(fit,which=1)
```



Poisson regression

- ▶ Model for counts Y_i
- ▶ The distribution of the Y_i is Poisson, $Y_i \sim \text{Pois}(\mu_i)$ with expectation μ_i .
- ▶ The model is

$$\log(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta} \quad (9)$$

- ▶ The link function is $h(\mu_i) = \log(\mu_i)$
- ▶ The variance function is $v(\mu_i) = \mu_i$ und $\phi = 1$.
- ▶ β_j is the difference in the logs of expected counts
- ▶ $\exp(\beta_j)$ is the (risk, rate, count) ratio for two subpopulations that differ on x_j by on unit (except for the intercept.)