Generalized Linear Models

Introduction, Motivation and Overview

Herwig Friedl

Graz University of Technology/Austria

hfriedl@tugraz.at

http://www.stat.tugraz.at/courses/glmLjubljana.html

5. May 2021

What you can expect:

We will discuss

- Ordinary Linear Models (Regression Analysis)
- Generalized Linear Models
- Maximum Likelihood Estimation & Goodness-of-Fit, Deviance
- Overdispersion
- Quasi-Likelihood Models
- Random Effects Models

Linear Models: Motivation

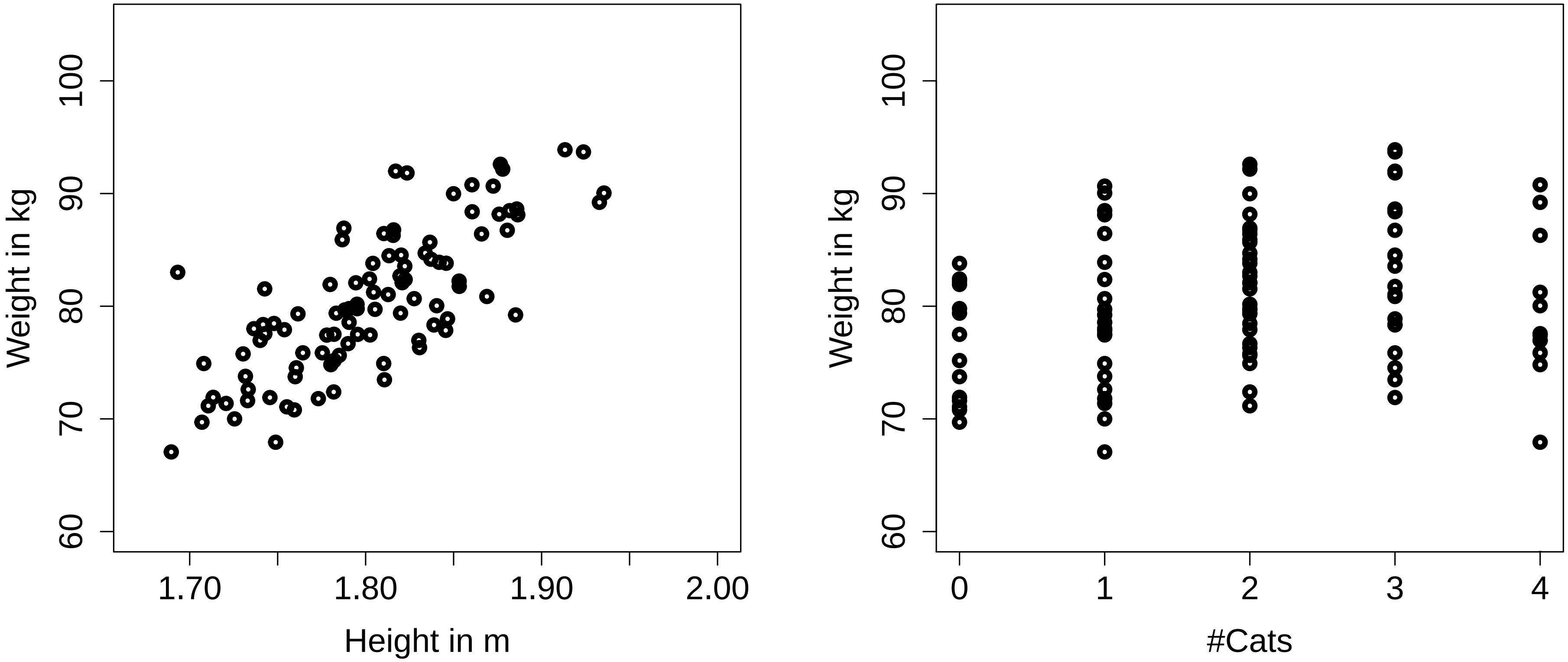
Suppose that we are interested in the average weight of male PhD students at University of Ljubljana. We put each guy's name (**population**) in a hat and randomly select 100 (**sample**).

Here they are: $y_1, y_2, \ldots, y_{100}$.

Suppose, in addition, we also measure their heights and ask for the number of cats owned by their parents.

Here they are: $h_1, h_2, \ldots, h_{100}$ and $c_1, c_2, \ldots, c_{100}$.

Linear Models: Motivation



Linear Models: Motivation

Questions:

How would you use this data to estimate the average weight of:

1. male PhD?
2. male PhD whose height is between 1.75 and 1.80 m?
3. male PhD whose parents own 3 cats?

Linear Models: Motivation

Answers:

1. $\bar{y} = \frac{1}{100}\sum_{i=1}^{100} y_i$, the sample mean
2. average the y_i's for guys whose h_is are between 1.75 and 1.80 m
3. average the y_i's for guys whose c_is are 3?
   No! Same as in 1., because the body weight certainly does not depend on the number of cats!

Intuitive description of regression:

(weight) y = variable of interest = response = dependent variable
(height) x = explanatory variable = predictor = indep. variable

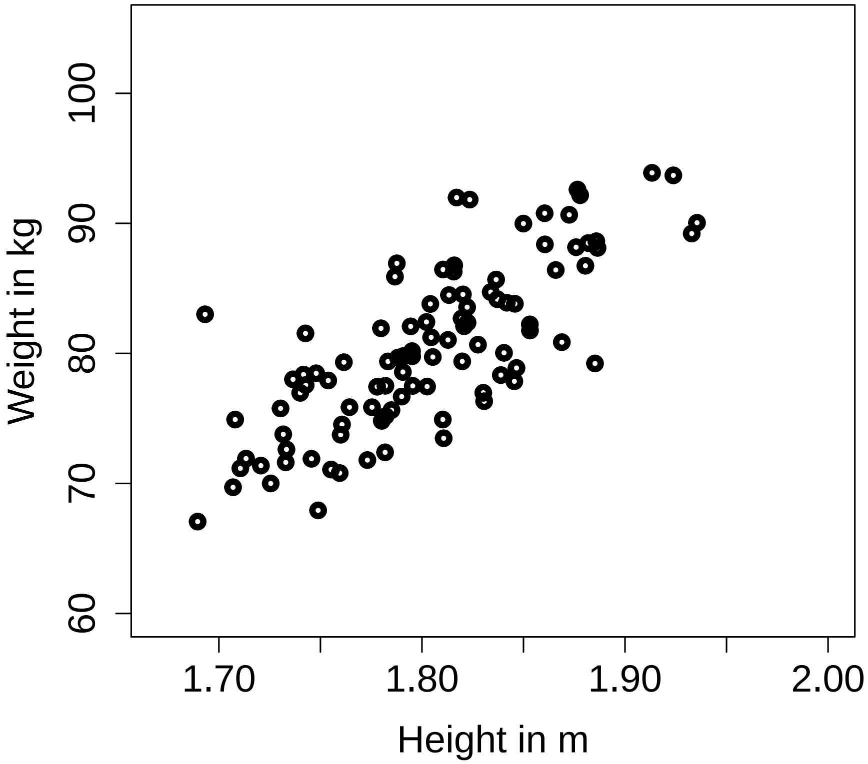
Linear Models: Motivation

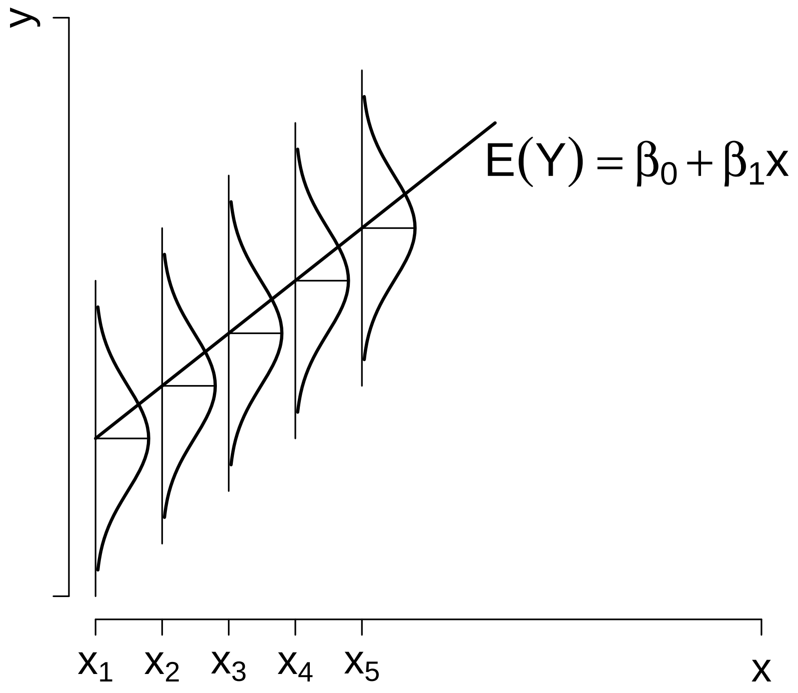
Fundamental assumption in regression

1. For each particular value of the predictor variable x, the response variable y is a random variable whose mean $\mathsf{E}(y)$ (expected value) depends on x.

2. The mean of y, $\mathsf{E}(y)$, can be written as a deterministic function of x.

Linear Models: Motivation





Linear Models, LMs

Three-Part Specification:

1. **Random Component:** $y_1, \ldots, y_n$ independent normal distributed with $\mathrm{E}(y_i) = \mu_i$, $i = 1, \ldots, n$, and constant variance $\mathrm{var}(y_i) = \sigma^2$.
2. **Systematic Component:** Fixed covariates $x_{i0}, x_{i1}, \ldots, x_{i,p-1}$ (intercept $x_{i0} = 1$) define a linear predictor

$$\eta_i = \sum_{j=0}^{p-1} x_{ij}\beta_j = \mathbf{x}_i^t \boldsymbol{\beta}$$

3. **Link Function:** between random and systematic components, here $\mu_i = \eta_i$ (identity function).

Linear Models, LMs

When is the model called *simple linear regression*?

simple: only one predictor x_i,

linear: regression function $\mathsf{E}(y) = \beta_0 + \beta_1 x$ is linear in **parameters**.

Why do we *care about* a regression model?

If a model is realistic and if we have reasonable estimates of β_0, β_1 we have:

1. the ability to predict new y_i's given a new x_i,
2. an understanding of how the mean $\mathsf{E}(y_i)$ changes with x_i.

Linear Models, LMs

Goal: Find the set of all relevant covariates (explanatory variables)

By applying statistical tests, identify those x_j's that are responsible for different means of the responses Thus, we check if a model like

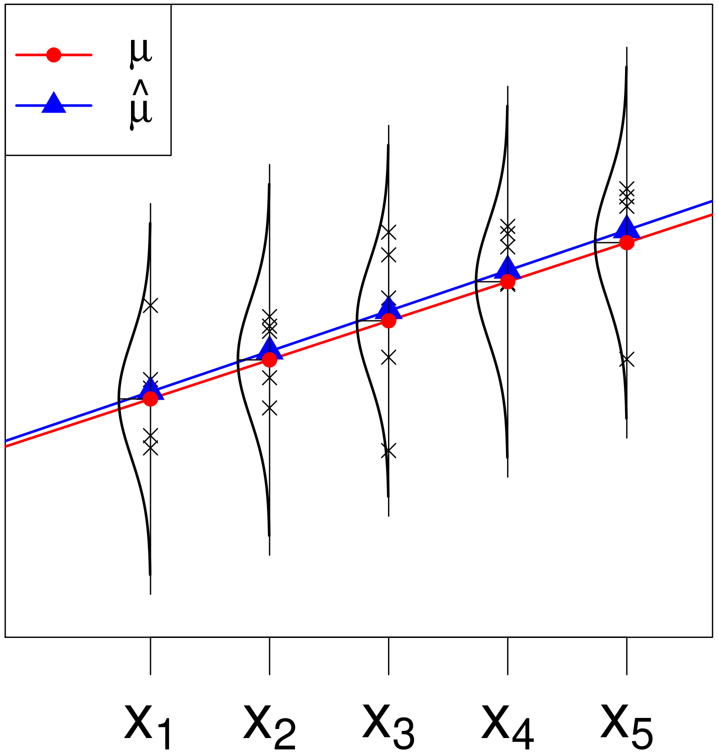
$$\mathsf{E}(y_i) = \mathbf{x}_i^t \boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_{p-1} x_{i,p-1}$$

holds, i.e. if

$$\mathsf{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}\,.$$

A covariate x_j is called **relevant**, if its associated parameter $\beta_j \neq 0$.

Typical Simple Linear Model:



Assumptions on y_i:

- normal distribution
- constant variance
- linear relationship

Goal Find a "suitable" estimate $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\mu} = \mathbf{x}^t \boldsymbol{\beta}$.

Problem Assumptions are very restrictive!

Linear Models, LMs

How to estimate β?

Minimize **Least Squares** Criterion (Sum of Squared Errors):

$$\mathsf{SSE}(\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i - \mu_i)^2 = \sum_{i=1}^{n} (y_i - \mathbf{x}_i^t \boldsymbol{\beta})^2 \,.$$

This gives the Maximum Likelihood Estimator (MLE) $\hat{\boldsymbol{\beta}}$ under a normal model. Thus,

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \mathsf{SSE}(\boldsymbol{\beta}) = \mathsf{SSE}(\hat{\boldsymbol{\beta}}) \qquad \text{and} \qquad \hat{\boldsymbol{\beta}} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y} \,.$$

Linear Models, LMs

How to estimate $\boldsymbol{\sigma^2}$?

It can be shown that the MLE of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mathbf{x}_i^t\hat{\boldsymbol{\beta}})^2 = \frac{1}{n}\mathsf{SSE}(\hat{\boldsymbol{\beta}})\,.$$

However, since $\hat{\sigma}^2$ is biased, we will always use its unbiased version (degrees of freedom corrected)

$$S^2 = \frac{1}{n-p}\mathsf{SSE}(\hat{\boldsymbol{\beta}})\,.$$

Generalized Linear Models, GLMs

Linear Model $\Longrightarrow$ **Generalized Linear Model**

Linear Model

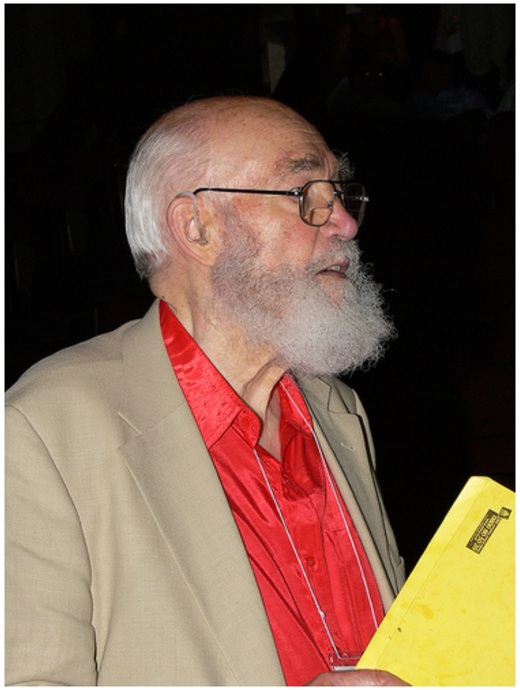
- normal distribution
- constant variance
- linear relationship between μ_i and $x_{i0}, x_{i1}, \ldots, x_{i,p-1}$

Generalized Linear Model

- any distribution from the **linear exponential family**
- variance proportional to a **function** of the mean
- linear relationship between a **function of** μ_i and $x_{i0}, x_{i1}, \ldots, x_{i,p-1}$

Generalized Linear Models: Literature

Modelling concept first introduced in 1972 by **John A. Nelder** (1924 - 2010) and **Robert W.M. Wedderburn** (1947 - 1975).



Peter McCullagh and **John A. Nelder** (1983): Generalized Linear Models, London:Chapman & Hall.

Generalized Linear Models, GLMs

1. **Random Component:** $y_1, \ldots, y_n$ independent distributed from any member of the **Linear Exponential Family (LEF)** with $\mathrm{E}(y_i) = \mu_i$ and variance $\mathrm{var}(y_i) = \phi V(\mu_i)$, $i = 1, \ldots, n$.

2. **Systematic Component:** Fixed covariates $x_{i0}, x_{i1}, \ldots, x_{i,p-1}$ (intercept $x_{i0} = 1$) define a **linear predictor**

$$\eta_i = \sum_{j=0}^{p-1} x_{ij}\beta_j = \mathbf{x}_i^t \boldsymbol{\beta}$$

3. **Link Function:** between random and systematic components

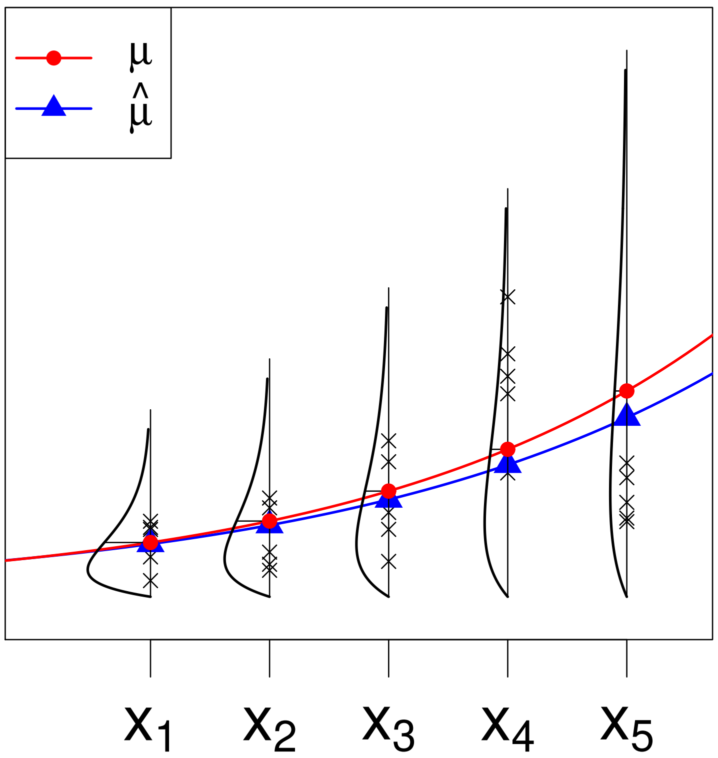
$$g(\mu_i) = \eta_i.$$

Generalized Linear Models, GLMs

Generalization of the Linear Model: Properties of the GLM

- particular choice of the **distribution** from the LEF determines the variance function $V(\mu)$
- the **dispersion parameter** ϕ allows for additional flexibility in the variance
- the LM is a **special case** of the GLM
  - constant variance function $V(\mu) = \text{constant}$
  - identical link function $g(\mu) = \mu = \eta$

Generalized Linear Models, GLMs

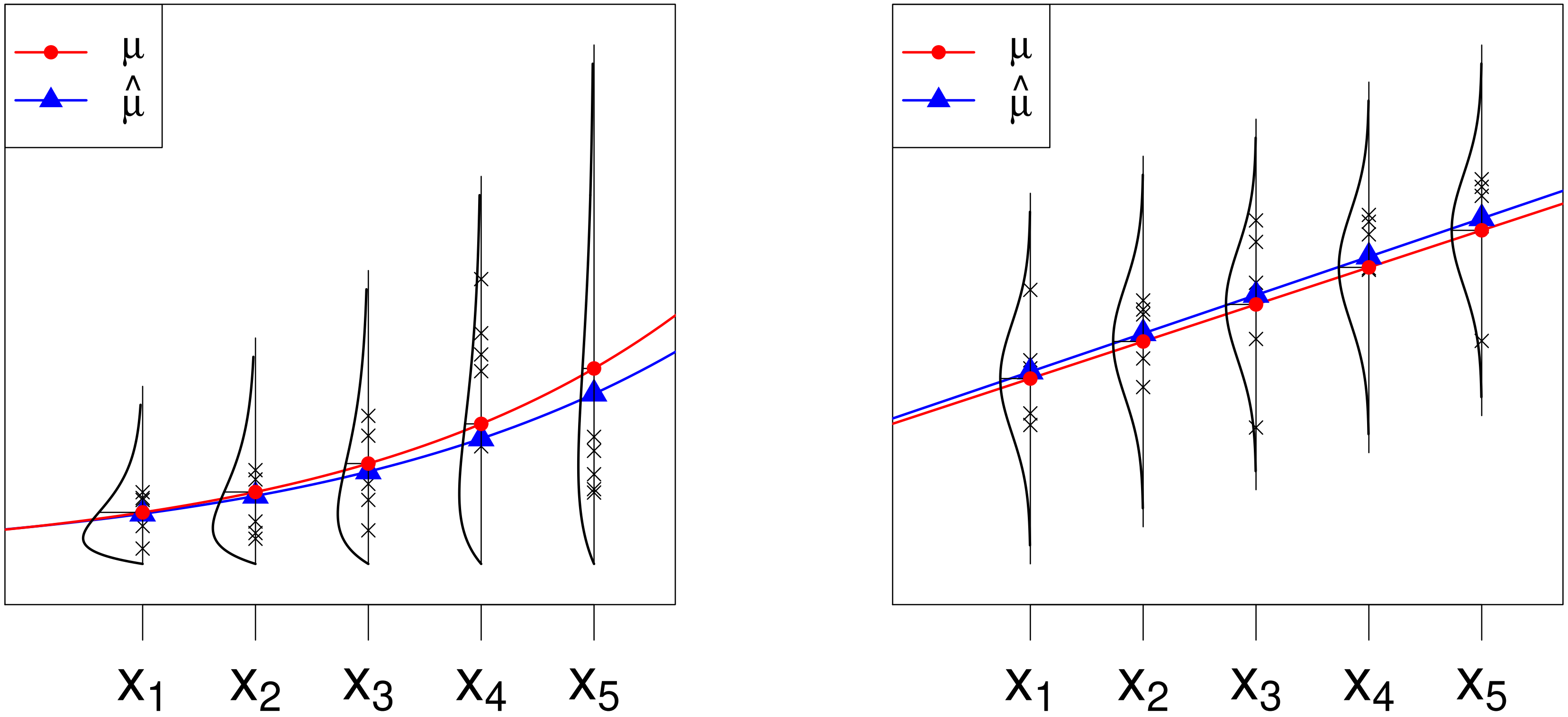


Assumptions

- distribution from the LEF
- variance as a function of μ: var$(y) = \phi V(\mu)$ with dispersion parameter ϕ
- linear relationship with the link function $g(\boldsymbol{\mu})$

Goal Find a "suitable" estimate $\hat{\boldsymbol{\beta}}$ for $g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$.

Generalized Linear Models, GLMs



Generalized Linear Models, GLMs

Limits of the GLM

To obtain estimates for the parameters in a GLM one has to choose a distribution from the one-parameter **Linear Exponential Family**

$$f(y_i|\theta_i) = \exp\left\{\frac{y_i\theta_i - c(\theta_i)}{\phi} + h(y_i,\phi)\right\} \,,$$

where θ_i is a specific function in μ_i and thus in $\boldsymbol{\beta}$.

Normal, Gamma, Binomial, Poisson, . . . are well known members.

Generalized Linear Models, GLMs

Characteristics of some common members:

Distribution	ϕ	$\mu(\theta)$	$V(\mu)$
Normal(μ, σ^2)	σ^2	θ	1
Gamma(μ, ν)	$1/\nu$	$-1/\theta$	μ^2
Poisson(μ)	1	$\exp(\theta)$	μ
Binomial$(m, \mu)/m$	$1/m \times 1$	$e^\theta/(1+e^\theta)$	$\mu(1-\mu)$

A LEF member is characterized by its variance function.

Generalized Linear Models, Estimates

How to find the MLEs?

The MLE $\hat{\mu}_i$ is defined as the **zero of the score function** (1st derivative of the log-likelihood function)

$$\frac{\partial}{\partial \mu_i} \log f(y_i|\theta_i) = \frac{y_i - \mu_i}{\phi V(\mu_i)}$$

Thus, the **MLE of μ** only depends on the **first two moments** of the assumed distribution ($\mathrm{E}(y_i) = \mu_i$, $\mathrm{var}(y_i) = \phi V(\mu_i)$).

Generalized Linear Models, Estimates

To find the MLE $\hat{\boldsymbol{\beta}}$ we apply the chain rule ($g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta}$), giving

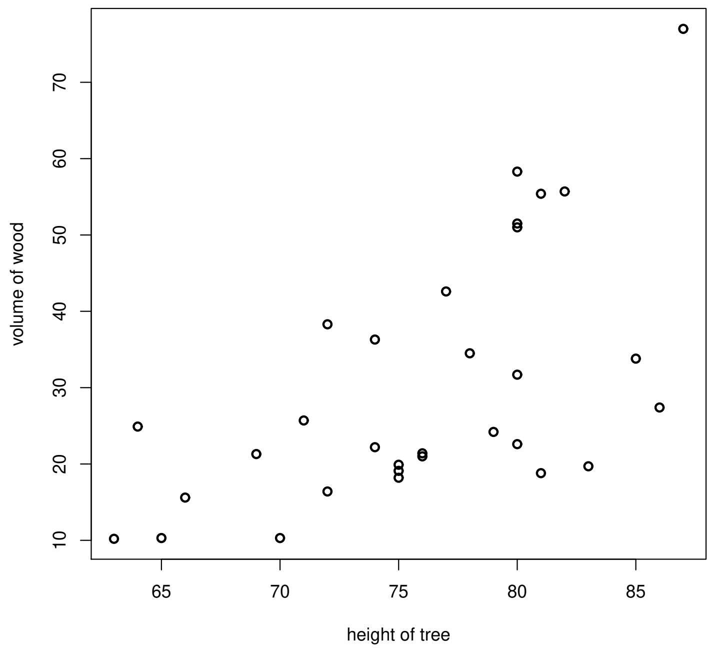
$$\frac{\partial}{\partial \boldsymbol{\beta}} \log f(\mathbf{y}|\boldsymbol{\theta}) = \sum_{i=1}^{n} \frac{y_i - \mu_i}{\phi V(\mu_i)} \cdot \frac{\partial \mu_i}{\partial \boldsymbol{\beta}},$$

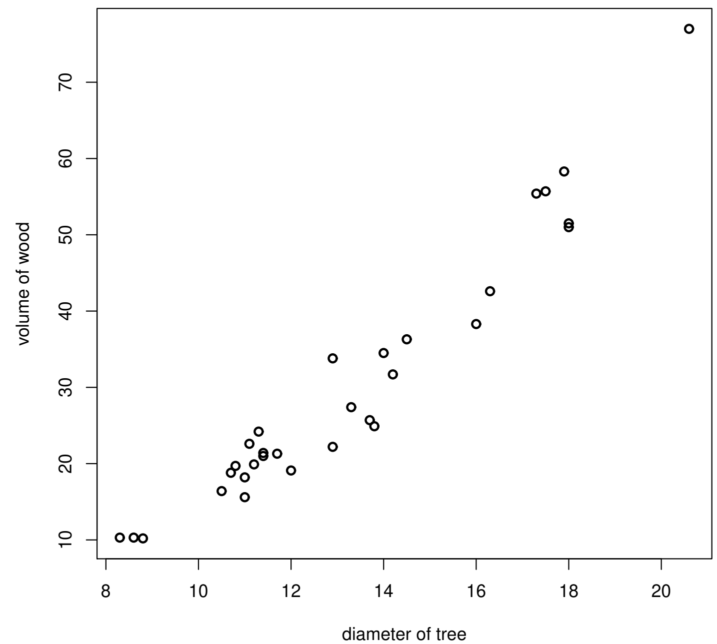
which of course also only depend on the **first two moments** of the assumed distribution and the assumed **link function**.

Notice that the score function is **highly nonlinear in $\boldsymbol{\beta}$** and therefore the zeros $\hat{\boldsymbol{\beta}}$ have to be found numerically (by iteration, IWLS).

Generalized Linear Models, Example

Example 1: $n = 31$ Black Cherry Trees, V volume of useful wood in feet^3, H height of tree in feet, D diameter of tree in inches.





Generalized Linear Models, Example

We assume that $V_i \overset{ind}{\sim} \mathsf{Normal}(\mu_i, \sigma^2)$ with a **cone like behavior** of the mean volume (after converting D from inches to feet), i.e.

$$\mathsf{E}(V_i) = \frac{\pi}{12} \cdot H_i \cdot \left(\frac{D_i}{12}\right)^2$$

This is equivalent to

$$\begin{aligned} \log(\mathsf{E}(V_i)) &= \log\frac{\pi}{12} + 1 \cdot \log(H_i) + 2 \cdot \log\frac{D_i}{12} \\ \log(\mu_i) &= \beta_0 + \beta_1 \log(H_i) + \beta_2 \log\frac{D_i}{12} \end{aligned}$$

Generalized Linear Models, Example

We use the statistic software package to do the calculation:

```
> glm(V ~ log(H) + log(D/12), family = gaussian(link=log))
 Coefficients:
             Estimate Std. Error t value Pr(>|t|)
 (Intercept) -1.57484    1.04613  -1.505 0.143422 ***
 log(H)       1.08765    0.24216   4.491 0.000111 ***
 log(D/12)    1.99692    0.08208  24.330  < 2e-16 ***
 ---
 (Dispersion parameter for gaussian family taken to be 6.41642)

     Null deviance: 8106.08  on 30  degrees of freedom
 Residual deviance:  179.66  on 28  degrees of freedom
 AIC: 150.44
 Number of Fisher Scoring iterations: 4
```

Generalized Linear Models, Example

Remember our cone model:

$$\begin{aligned}\log(\mathsf{E}(V_i)) &= \log\frac{\pi}{12} + 1\cdot\log(H_i) + 2\cdot\log\frac{D_i}{12}\\ \log(\mu_i) &= \beta_0 + \beta_1\log(H_i) + \beta_2\log\frac{D_i}{12}\end{aligned}$$

We've got $\hat\beta_0 = -1.575$, nicely comparing to $\log(\pi/12) = -1.340$.

Also $\hat\beta_1 = 1.088$ and $\hat\beta_2 = 1.997$ are both close to the respective theoretical quantities 1 and 2.

What is meant by the term **deviance**?

Generalized Linear Models, Deviance

We need a measure to assess the **goodness-of-fit** of our model.

One approach is to compare our model with the best available model. The best model allows one parameter for every single mean μ_i, thus consists of n parameters in the linear predictor. Such models are called **saturated.**

Under this setting, the MLE is $\hat{\boldsymbol{\mu}} = \mathbf{y}$.

The **scaled deviance** compares the maximum of the log-likelihood under our model with its maximum under the best model, i.e.

$$\frac{1}{\phi} D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\big(\log f(\mathbf{y}|\mathbf{y}) - \log f(\mathbf{y}|\hat{\boldsymbol{\mu}})\big) \, .$$

Generalized Linear Models, Deviance

$$\frac{1}{\phi} D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = 2\big(\log f(\mathbf{y}|\mathbf{y}) - \log f(\mathbf{y}|\hat{\boldsymbol{\mu}})\big) \,.$$

Under certain regularity conditions it can be shown that the deviance follows asymptotically a χ^2_{n-p} distribution (with $n-p$ degrees of freedom).

Since the mean of a χ^2_{n-p} variate is its degrees of freedom, we often compare the scaled deviance $D(\mathbf{y}; \hat{\boldsymbol{\mu}})/\phi$ with $n-p$.

We are **happy with the model fit**, if $D(\mathbf{y}; \hat{\boldsymbol{\mu}})/\phi \approx n-p$.

We will have to **improve the model**, if $D(\mathbf{y}; \hat{\boldsymbol{\mu}})/\phi \gg n-p$.

Generalized Linear Models, Deviance

The simplest (worst fitting) model is called the **null model** (intercept only model, i.i.d.) and assumes that all the means are the same. The respective **null deviance** is

$$\frac{1}{\phi}D(\mathbf{y};\bar{\mathbf{y}}) = 2\big(\log f(\mathbf{y}|\mathbf{y}) - \log f(\mathbf{y}|\bar{\mathbf{y}})\big)\,.$$

What to do, if the **dispersion parameter ϕ is unknown**?

Generalized Linear Models, Deviance

Consider the ratios ($i = 1, \dots, n$)

$$1 = \frac{\mathsf{E}(y_i - \mu_i)^2}{\mathsf{var}(y_i)} = \frac{\mathsf{E}(y_i - \mu_i)^2}{\phi \cdot V(\mu_i)} \;\Rightarrow\; \phi = \frac{\mathsf{E}(y_i - \mu_i)^2}{V(\mu_i)}, \; i = 1, \dots, n\,.$$

Averaging (df corrected) over its estimated versions results in the mean **Pearson statistic**

$$\hat{\phi} = \frac{1}{n-p} \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}\,.$$

Generalized Linear Models, Deviance

For Gaussian responses (constant variances) the mean **Pearson statistic** equals the **mean sum of squared errors**

$$\hat{\phi} = S^2 = \frac{1}{n-p}\,\mathsf{SSE}(\hat{\boldsymbol{\beta}})$$

and the **scaled deviance** equals the **scaled sum of squared errors**

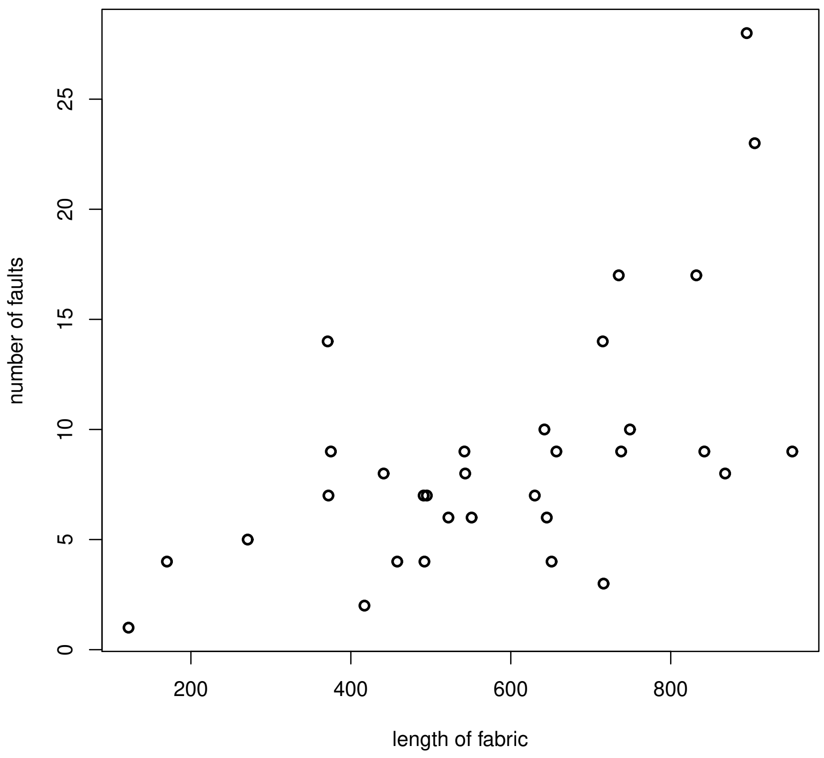
$$\frac{1}{\phi} D(\mathbf{y}; \hat{\boldsymbol{\mu}}) = \frac{1}{\sigma^2}\,\mathsf{SSE}(\hat{\boldsymbol{\beta}})\,.$$

Generalized Linear Models, Example Revisited

```
> glm(V ~ log(H) + log(D/12), family = gaussian(link=log))
 Coefficients:
             Estimate Std. Error t value Pr(>|t|)
 (Intercept) -1.57484    1.04613  -1.505 0.143422 ***
 log(H)       1.08765    0.24216   4.491 0.000111 ***
 log(D/12)    1.99692    0.08208  24.330  < 2e-16 ***
 ---
 (Dispersion parameter for gaussian family taken to be 6.41642)

     Null deviance: 8106.08  on 30  degrees of freedom
 Residual deviance:  179.66  on 28  degrees of freedom
 AIC: 150.44
 Number of Fisher Scoring iterations: 4
```

Generalized Linear Models, Example

Example 2: Fabric data. Faults f in rolls of material of length l.



Generalized Linear Models, Example

We consider a **Poisson model for counts**

$$f_i \overset{ind}{\sim} \text{Poisson}(\mu_i = \exp(\beta_0 + \beta_1 \log l_i)),$$

i.e. $\mu_i > 0$ and

$$\log \mu_i = \beta_0 + \beta_1 \log l_i$$

Question: Mean number of faults proportional to length ($\beta_1 = 1$)?

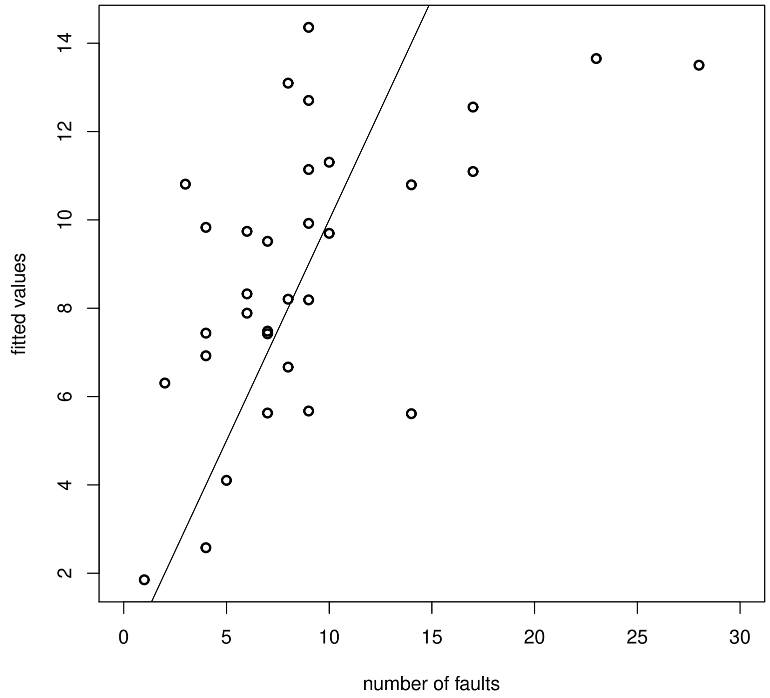
$$\mu_i = \exp(\beta_0) \cdot l_i^{\beta_1}$$

Generalized Linear Models, Example

```
> glm(f ~ log(l), family=poisson(link=log))
 Coefficients:
             Estimate Std. Error z value Pr(>|z|)
 (Intercept)  -4.1730     1.1352  -3.676 0.000237 ***
 log(l)        0.9969     0.1759   5.668 1.45e-08 ***
 ---
 (Dispersion parameter for poisson family taken to be 1)

     Null deviance: 103.714  on 31  degrees of freedom
 Residual deviance:  64.537  on 30  degrees of freedom
 AIC: 191.84
 Number of Fisher Scoring iterations: 4
```

$\hat{\beta}_1 \approx 1$ but **Deviance is more than twice the degrees of freedom!**

Generalized Linear Models, Example



Model fit is not really bad but **variance** seems to be **larger than assumed** under the Poisson model!

We say that there is some **overdispersion** w.r.t. the Poisson variance.

What now?

Generalized Linear Models, Example

Limits of the GLM

To obtain estimates for the parameters in a GLM one has to choose a distribution from the **exponential family**.

Normal, Gamma, Binomial, Poisson, . . . are well known members.

New Approach

Choose variance function $V(\mu)$ which does not necessarily belong to a distribution from the exponential family $\Rightarrow$ **quasi-likelihood approach**.

Quasi-Likelihood Estimation

Remember: the MLE $\hat{\boldsymbol{\mu}}$ is defined as the **zero of the score function**

$$\frac{\partial}{\partial \mu_i} \log f(y_i|\theta_i) = \frac{y_i - \mu_i}{\phi V(\mu_i)} \, .$$

Thus, the **MLE $\hat{\mu}$** only depends on the **first two moments**.

Instead of an exponential family distribution we now only assume $\mathsf{E}(y) = \mu$ and an arbitrary variance model $\mathsf{var}(y) = \phi V(\mu)$.

Thus, the above function is no longer a score function from a likelihood model! However, we still use it to define $\hat{\boldsymbol{\mu}}$.

Quasi-Likelihood Estimation

Define the **Maximum Quasi-Likelihood Estimator** (MQLE) $\hat{\boldsymbol{\mu}}$ as the zero of

$$\frac{\partial}{\partial \mu_i} \log q(y_i|\mu_i) = \frac{y_i - \mu_i}{\phi V(\mu_i)}\,.$$

This **quasi-score** function has many properties in common with a log-likelihood derivative (Wedderburn, 1974, 1976). Therefore, the integral

$$\log q(y|\mu) = \int^{\mu} \frac{y-t}{\phi V(t)} dt$$

should behave like a log-likelihood of y for μ. Wedderburn showed the equivalence of $f(\cdot)$ and $q(\cdot)$ for linear, one-parameter exponential families.

Quasi-Likelihood Estimation

We refer to $\log q(\mathbf{y}|\boldsymbol{\mu})$ as the (log) **quasi-likelihood** which is only based on a **mean-variance relation**.

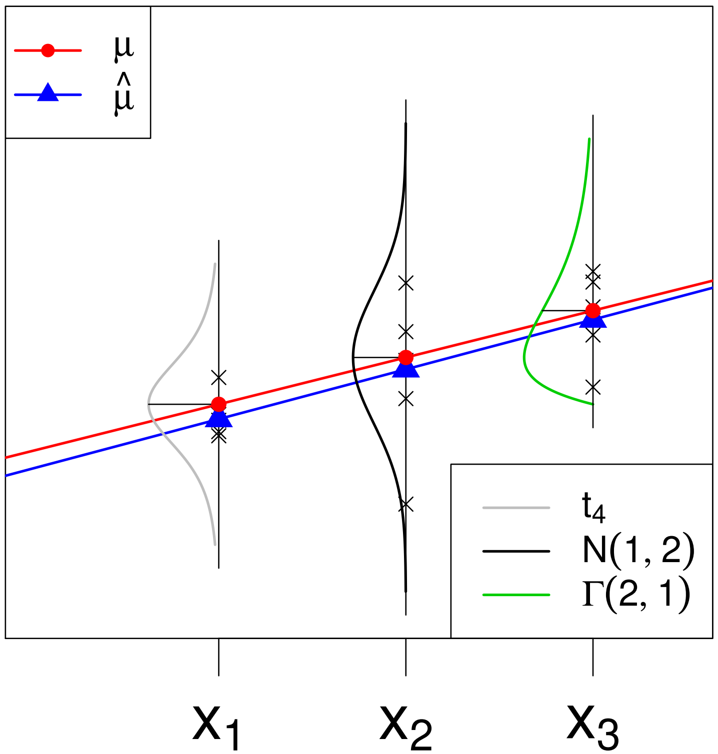
For the entire sample, the **quasi-deviance** is defined as

$$\begin{aligned} D(\mathbf{y};\hat{\boldsymbol{\mu}}) &= 2\phi\big(\log q(\mathbf{y}|\mathbf{y}) - \log q(\mathbf{y}|\hat{\boldsymbol{\mu}})\big) \\ &= 2\sum_{i=1}^{n}\int_{\hat{\mu}_i}^{y_i}\frac{y-t}{V(t)}dt\,. \end{aligned}$$

Quasi-Likelihood Estimation

related to	$V(\mu)$	$\log q(y\|\mu)$
Normal	1	$-\frac{1}{2}(y-\mu)^2$
Poisson	μ	$y \log \mu - \mu$
Gamma	μ^2	$-y/\mu - \log \mu$
—	μ^ξ	$\mu^{-\xi}\left(\frac{\mu y}{1-\xi} - \frac{\mu^2}{2-\xi}\right)$
Binomial	$\mu(1-\mu)$	$y \log \frac{\mu}{1-\mu} + \log(1-\mu)$
—	$\mu^2(1-\mu)^2$	$(2y-1)\log \frac{\mu}{1-\mu} - \frac{y}{\mu} - \frac{1-y}{1-\mu}$
NegBin	$\mu + \mu^2/k$	$y \log \frac{\mu}{k+\mu} + k \log \frac{k}{k+\mu}$

Quasi-Likelihood Approach:



Assumptions for y_i:

- the distribution is not specified **explicitly**
- variance as a function of μ: var$(y) = \phi \cdot V(\mu)$
- linear relationship between the explanatory variables and the link function $g(\mu)$

Advantage Distribution need **not** to be specified completely, the knowledge of $V(\mu)$ suffices.

Quasi-Likelihood Estimation, Example 2 Revisited

Example 2: Fabric data. Faults f in rolls of material of length l.

```
> glm(f ~ log(l), family=poisson(link=log))
 Coefficients:
             Estimate Std. Error z value Pr(>|z|)
 (Intercept)  -4.1730     1.1352  -3.676 0.000237 ***
 log(l)        0.9969     0.1759   5.668 1.45e-08 ***
 ---
 (Dispersion parameter for poisson family taken to be 1)

     Null deviance: 103.714  on 31  degrees of freedom
 Residual deviance:  64.537  on 30  degrees of freedom
```

Overdispersion: Try a **quasi-Poisson variance** model and assume var$(y_i) = \phi \cdot V(\mu_i)$, $\phi > 0$. (For the Poisson variance $\phi = 1$.)

Quasi-Likelihood Estimation, Example 2 Revisited

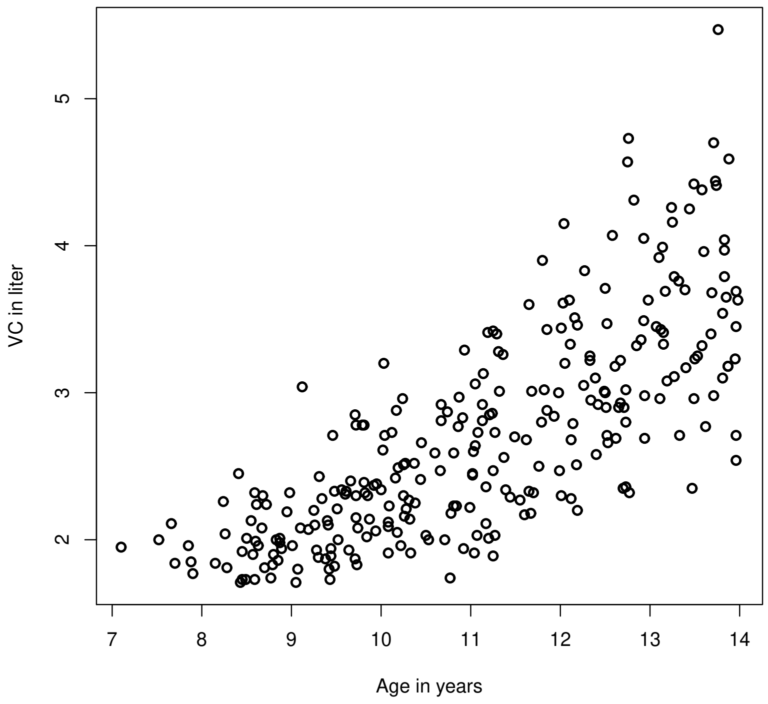
Can be easily fitted in :

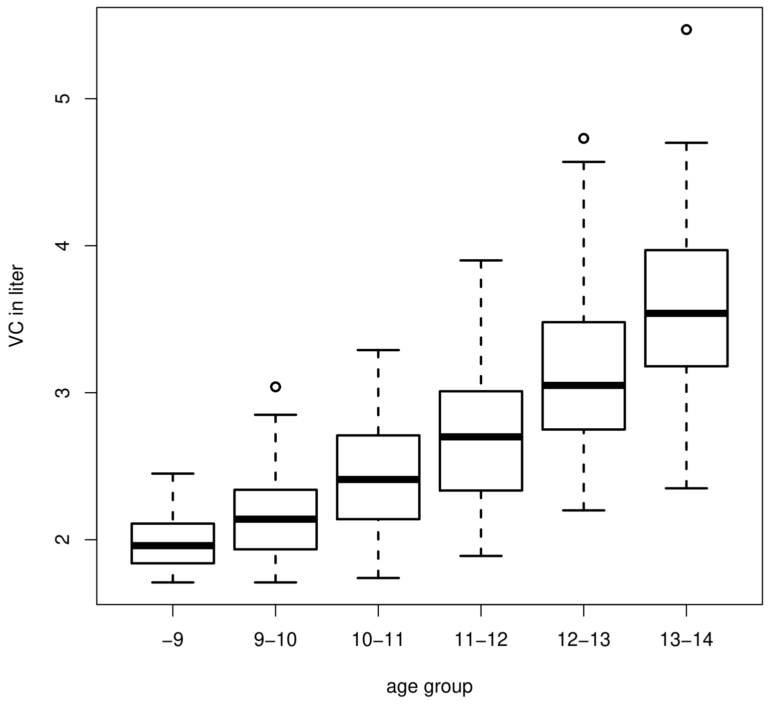
```
> glm(f ~ log(l), family=quasipoisson(link=log)))

 Coefficients:
             Estimate Std. Error t value Pr(>|t|)
 (Intercept)  -4.1730     1.7094  -2.441 0.020752 *
 log(l)        0.9969     0.2649   3.764 0.000727 ***
 ---
 (Dispersion parameter for quasipoisson family taken to be 2.267506)

     Null deviance: 103.714  on 31  degrees of freedom
 Residual deviance:  64.537  on 30  degrees of freedom
```

Now, $D(\mathbf{y};\hat{\boldsymbol{\mu}})/\hat{\phi} = 64.537/2.267 = 28.468$ is close to its df of 30.

Quasi-Likelihood Estimation, Example

Example 3: Vital capacity (lung volume) in liter of $n = 277$ girls aged from 7 to 14 years. Denote the observed VC values by y_i.





Quasi-Likelihood Estimation, Example

Assumption: The VC-mean growths **exponentially in age**, i.e.

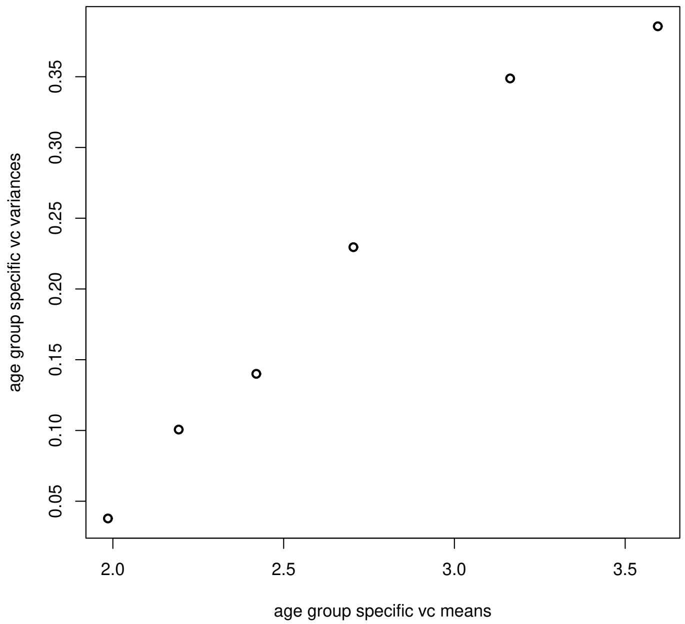
$$\mathsf{E}(y_i) = \mu_i = \exp(\beta_0 + \beta_1 \mathsf{age}_i)$$

Age-group specific means and variances:

Age	–9	9–10	10–11	11–12	12–13	13–
n	38	47	45	47	51	49
$\bar{y}$	1.99	2.19	2.42	2.70	3.16	3.60
S^2	0.04	0.10	0.14	0.23	0.35	0.39

Quasi-Likelihood Estimation, Example

How to find a suitable variance model?



Evidence:

Age-group specific **variances** seem to increase **linearly in the means** but with a shift to the right.

This implies

var$(y_i) = \phi \cdot (\mu_i + \alpha)$.

Quasi-Likelihood Estimation, Example

Since $\mu + \alpha = \mathsf{E}(y + \alpha)$ we estimate α by $-\min(y_i) = -1.7$ and define

$$y_i^* = y_i - 1.7$$

for which $\mathsf{E}(y_i^*) = \mu_i^* = \mu_i - 1.7$ and $\mathsf{var}(y_i^*) = \phi \cdot \mu_i^*$.

Age	–9	9–10	10–11	11–12	12–13	13–
$\overline{y}^*$	0.29	0.49	0.72	1.00	1.46	1.90
S^{*2}	0.04	0.10	0.14	0.23	0.35	0.39
$S^{*2}/\overline{y}^*$	0.13	0.20	0.19	0.23	0.24	0.20

The average of all $S^{*2}/\overline{y}^*$ terms is 0.198, which roughly estimates the dispersion ϕ.

Quasi-Likelihood Estimation, Example

Thus, we fit a **loglinear quasi-Poisson model** for the shifted responses y_i^*, i.e. we use the link

$$g(\mu_i^*) = \log(\mu_i^*) = \beta_0 + \beta_1 \text{age}_i \,.$$

```
> glm(I(vc-1.7) ~ age, family=quasipoisson(link=log))

 Coefficients:
             Estimate Std. Error t value Pr(>|t|)
 (Intercept)  -3.9053     0.2089  -18.70   <2e-16 ***
 age           0.3382     0.0172   19.66   <2e-16 ***
 ---
 (Dispersion parameter for quasipoisson family taken to be 0.201)
```

Quasi-Likelihood Estimation, Example

```
> glm(I(vc-1.7) ~ age, family=quasipoisson(link=log))

 Coefficients:
             Estimate Std. Error t value Pr(>|t|)
 (Intercept)  -3.9053     0.2089  -18.70   <2e-16 ***
 age           0.3382     0.0172   19.66   <2e-16 ***
 ---
 (Dispersion parameter for quasipoisson family taken to be 0.201)

    Null deviance: 142.818  on 276  degrees of freedom
Residual deviance:  56.889  on 275  degrees of freedom
```

The scaled deviance is now $56.889/0.201 = 282.71$ (comp. with df$= 275$), and the mean Pearson statistic 0.201 estimates the dispersion well (comp. with mean deviance $56.889/275 = 0.207$).

Summary I

Linear Model

- normal distribution
- constant variance
- mean is a linear combination of some explanatory variables

Generalized Linear Model

- distribution from the exponential family
- variance is a function of the mean
- additional link function

Summary II

Generalized Linear Model

- distribution from the exponential family
- variance is a function of the mean
- additional link function

QL Approach

- define **only** $V(\mu)$
- complete specification of the distribution is **not** necessary

What about modelling dependent responses?

Random Effect Models for Overdispersion

General problem of **overdispersion** in Poisson and binomial models: Deviance from the model is much larger than the residual df.

Interpret this situation as evidence that there are other factors varying which are not accounted for in the model, but which are associated with the response:

Random Effect Models for Overdispersion

A simple way of representing the extra variation is by including a **random effect** in the **linear predictor**:

$$g(\mu_i) = \mathbf{x}_i^t \boldsymbol{\beta} + z_i\,,$$

where the random effects z_i are an (iid) random sample from some distribution $G(z)$.

Here μ_i denotes the **conditional mean** given the random effect.

Random Effect Models for Overdispersion

Example 2: Fabric data reconsidered.

$$y_i \overset{ind}{\sim} \text{Poisson}(\exp(\beta_0 + \beta_1 \log l_i)),$$

i.e.

$$\log \mu_i = \beta_1 + \beta_2 \log l_i$$

Revise model:

$$y_i | z_i \overset{ind}{\sim} \text{Poisson}(\exp(\beta_1 + \beta_2 \log l_i + z_i)), \quad z_i \overset{iid}{\sim} G(z).$$

What now ?

Random Effect Models for Overdispersion

Let us first assume that the random effects z_i are iid unit mean gamma variables with shape α (conjugate distribution).

The counts are then marginally **negative binomial** variables with $\mathsf{E}(y_i) = \mu_i$ and $\mathsf{var}(y) = \mu + \mu^2/\alpha$.

Here, α quantifies the amount of overdispersion.

The special case $\alpha = \infty$ corresponds to no overdispersion (Poisson).

Random Effect Models for Overdispersion

Again, R offers a function to estimate this model:

```
> glm.nb(f ~ log(l))

 Coefficients:
             Estimate Std. Error z value Pr(>|z|)
 (Intercept)  -3.7951     1.4577  -2.603  0.00923 **
 log(l)        0.9378     0.2280   4.114 3.89e-05 ***
 ---
 (Dispersion parameter for NegBin(8.667) family taken to be 1)
     Null deviance: 50.28  on 31  degrees of freedom
 Residual deviance: 30.67  on 30  degrees of freedom
               Theta:  8.67
           Std. Err.:  4.17
```

Random Effect Models for Overdispersion

Could we also handle models like

$$y_i | z_i \stackrel{ind}{\sim} \text{Poisson}(\mu_i), \quad z_i \stackrel{iid}{\sim} N(0, 1)$$

$$\log \mu_i = \beta_1 + \beta_2 \log l_i + \sigma_z z_i$$

Yes!

The **EM algorithm** has to be applied to get the MLEs.

Now what? Interested in details?

Generalized Linear Models:

(A series of blocked lectures with some practicals)