Today

- HW 1: due February 7, 2 pm. Feb 28, Mar 21, Apr 4/11
- Intro: generalized linear models SM 10.3, Eg. 10.18, 29
- Principles of Statistics CD Chapter 2
- In the News: placebos and nocebos
- `https://www.zoology.ubc.ca/~schluter/R/`
  Bare bones introduction to R, nicely formatted
- `http://yihui.name/knitr/`
  Generating reports with Knitr and R

Binary Data: Example 10.18

- `library(SMPracticals); data(nodal)` has 53 binary observations; one per patient
- response: binary, indicating cancer has spread to lymph nodes (1) or not (0)
- covariates: `age`, `stage`, `grade`, `xray`, `acid`
- all dummy variables

```
> data(nodal)
> head(nodal)
  m r aged stage grade xray acid
1 1 1    0     1     1    1    1
2 1 1    0     1     1    1    1
3 1 1    0     1     1    1    1
4 1 1    0     1     1    1    1
5 1 1    0     1     1    1    1
6 1 0    0     1     1    1    1
> dim(nodal)
[1] 53  7
```

... example 10.18

- model

$$r_i \sim \text{Bernoulli}(p_i), \quad \log\left(\frac{p_i}{1-p_i}\right) = x_i^{\mathrm{T}}\beta$$

- likelihood function

$$L(\beta; r) \propto \prod_{i=1}^{n} p_i^{r_i}(1-p_i)^{1-r_i}$$

- log-likelihood function

$$\ell(\beta; r) = \Sigma y_i x_i^{\mathrm{T}}\beta - \log\{1 + \exp(x_i^{\mathrm{T}}\beta)\}$$

- maximum likelihood estimator $\hat{\beta}$:

$$\left.\frac{\partial \ell(\beta; r)}{\partial \beta}\right|_{\hat{\beta}} = 0$$

- asymptotic variance estimate $j^{-1}(\hat{\beta})$ $\quad j(\beta) = -\frac{\partial^2 \ell(\beta; r)}{\partial \beta \partial \beta^{\mathrm{T}}}$

... example 10.18

- model

$$r_i \sim \text{Bernoulli}(p_i), \quad \log\left(\frac{p_i}{1-p_i}\right) = x_i^{\mathrm{T}}\beta$$

- likelihood function

$$L(\beta; r) \propto \prod_{i=1}^{n} p_i^{r_i}(1-p_i)^{1-r_i}$$

- log-likelihood function

$$\ell(\beta; r) = \Sigma y_i x_i^{\mathrm{T}}\beta - \log\{1 + \exp(x_i^{\mathrm{T}}\beta)\}$$

- maximum likelihood estimator $\hat{\beta}$:

$$\left.\frac{\partial \ell(\beta; r)}{\partial \beta}\right|_{\hat{\beta}} = 0$$

- asymptotic variance estimate $j^{-1}(\hat{\beta})$ $\quad j(\beta) = -\frac{\partial^2 \ell(\beta; r)}{\partial \beta \partial \beta^{\mathrm{T}}}$

... example 10.18

- model

$$r_i \sim \text{Bernoulli}(p_i), \quad \log\left(\frac{p_i}{1-p_i}\right) = x_i^{\mathrm{T}}\beta$$

- likelihood function

$$L(\beta; r) \propto \prod_{i=1}^{n} p_i^{r_i}(1-p_i)^{1-r_i}$$

- log-likelihood function

$$\ell(\beta; r) = \Sigma y_i x_i^{\mathrm{T}}\beta - \log\{1 + \exp(x_i^{\mathrm{T}}\beta)\}$$

- maximum likelihood estimator $\hat{\beta}$:

$$\left.\frac{\partial \ell(\beta; r)}{\partial \beta}\right|_{\hat{\beta}} = 0$$

- asymptotic variance estimate $j^{-1}(\hat{\beta})$ $\quad j(\beta) = -\frac{\partial^2 \ell(\beta; r)}{\partial \beta \partial \beta^{\mathrm{T}}}$

... example 10.18

- model

$$r_i \sim \text{Bernoulli}(p_i), \quad \log\left(\frac{p_i}{1-p_i}\right) = x_i^{\mathrm{T}}\beta$$

- likelihood function

$$L(\beta; r) \propto \prod_{i=1}^{n} p_i^{r_i}(1-p_i)^{1-r_i}$$

- log-likelihood function

$$\ell(\beta; r) = \Sigma y_i x_i^{\mathrm{T}}\beta - \log\{1 + \exp(x_i^{\mathrm{T}}\beta)\}$$

- maximum likelihood estimator $\hat{\beta}$:

$$\left.\frac{\partial \ell(\beta; r)}{\partial \beta}\right|_{\hat{\beta}} = 0$$

- asymptotic variance estimate $j^{-1}(\hat{\beta})$ $\quad j(\beta) = -\frac{\partial^2 \ell(\beta; r)}{\partial \beta \partial \beta^{\mathrm{T}}}$

... example 10.18

- model

$$r_i \sim \text{Bernoulli}(p_i), \quad \log\left(\frac{p_i}{1-p_i}\right) = x_i^{\mathrm{T}}\beta$$

- likelihood function

$$L(\beta; r) \propto \prod_{i=1}^{n} p_i^{r_i}(1-p_i)^{1-r_i}$$

- log-likelihood function

$$\ell(\beta; r) = \Sigma y_i x_i^{\mathrm{T}}\beta - \log\{1 + \exp(x_i^{\mathrm{T}}\beta)\}$$

- maximum likelihood estimator $\hat{\beta}$:

$$\left.\frac{\partial \ell(\beta; r)}{\partial \beta}\right|_{\hat{\beta}} = 0$$

- asymptotic variance estimate $j^{-1}(\hat{\beta})$ $\quad j(\beta) = -\frac{\partial^2 \ell(\beta; r)}{\partial \beta \partial \beta^{\mathrm{T}}}$

... example 10.18

```
> ex1018 <- glm(r ~ . - m, data = nodal, family = binomial)
##########    r is the response, use all columns but m as covariates
> summary(ex1018)

Call:
glm(formula = r ~ . - m, family = binomial, data = nodal)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3317  -0.6653  -0.2999   0.6386   2.1502

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.0794     0.9868  -3.121   0.0018 **
aged1        -0.2917     0.7540  -0.387   0.6988
stage1        1.3729     0.7838   1.752   0.0799 .
grade1        0.8720     0.8156   1.069   0.2850
xray1         1.8008     0.8104   2.222   0.0263 *
acid1         1.6839     0.7915   2.128   0.0334 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 70.252  on 52  degrees of freedom
Residual deviance: 47.611  on 47  degrees of freedom
AIC: 59.611

Number of Fisher Scoring iterations: 5
```

RStudio

Go to file/function

Project: (None)

Console ~/

```
    Null deviance: 70.252  on 52  degrees of freedom
Residual deviance: 47.611  on 47  degrees of freedom
AIC: 59.611

Number of Fisher Scoring iterations: 5

> ?glm
> glm$vcov
Error in glm$vcov : object of type 'closure' is not subsettable
> vcov(glm)
Error in UseMethod("vcov") :
  no applicable method for 'vcov' applied to an object of class "function"
> vcov(ex1018)
            (Intercept)       aged1       stage1      grade1
(Intercept)   0.9737142 -0.33529043 -0.265570935 -0.28843059
aged1        -0.3352904  0.56852418 -0.091543659  0.13317513
stage1       -0.2655709 -0.09154366  0.614418872 -0.19964134
grade1       -0.2884306  0.13317513 -0.199641340  0.66516828
xray1        -0.2753407  0.01550127  0.084134191 -0.01627798
acid1        -0.5326605  0.10665082  0.008102781  0.14450101
                  xray1        acid1
(Intercept) -0.27534073 -0.532660509
aged1        0.01550127  0.106650818
stage1       0.08413419  0.008102781
grade1      -0.01627798  0.144501013
xray1        0.65677489  0.058370345
acid1        0.05837034  0.626431282
> diag(.Last.value)
(Intercept)       aged1      stage1      grade1       xray1       acid1
  0.9737142   0.5685242   0.6144189   0.6651683   0.6567749   0.6264313
> sqrt(.Last.value)
(Intercept)       aged1      stage1      grade1       xray1       acid1
  0.9867696   0.7540054   0.7838488   0.8155785   0.8104165   0.7914741
> coef(ex1018)
(Intercept)       aged1      stage1      grade1       xray1       acid1
 -3.0793806  -0.2917427   1.3729295   0.8719723   1.8008141   1.6839295
> coef(ex1018)/sqrt(diag(vcov(ex1018)))
(Intercept)       aged1      stage1      grade1       xray1       acid1
 -3.1206684  -0.3869239   1.7515235   1.0691457   2.2220847   2.1275863
>
```

Environment | History

Import Dataset | Clear | List

Global Environment

Data

nodal	53 obs. of 7 variables

Values

ex1018	List of 30
coefficients :	Named num [1:6] -3.079 -0.292 1.373 0.872…
..- attr(*, "names")=	chr [1:6] "(Intercept)" "aged1" "s…
residuals :	Named num [1:53] 1.07 1.07 1.07 1.07 1.07 ...
..- attr(*, "names")=	chr [1:53] "1" "2" "3" "4" ...
fitted.values :	Named num [1:53] 0.934 0.934 0.934 0.934…
..- attr(*, "names")=	chr [1:53] "1" "2" "3" "4" ...
effects :	Named num [1:53] 1.335 -0.751 -1.788 -0.704 -2…
..- attr(*, "names")=	chr [1:53] "(Intercept)" "aged1" "…

Files | Plots | Packages | Help | Viewer

R: Fitting Generalized Linear Models | Find in Topic

In addition, non-empty fits will have components `qr`, `R` and `effects` relating to the final weighted linear fit.

Objects of class `"glm"` are normally of class `c("glm", "lm")`, that is inherit from class `"lm"`, and well-designed methods for class `"lm"` will be applied to the weighted linear model at the final iteration of IWLS. However, care is needed, as extractor functions for class `"glm"` such as `residuals` and `weights` do **not** just pick out the component of the fit with the same name.

If a `binomial` `glm` model was specified by giving a two-column response, the weights returned by `prior.weights` are the total numbers of cases (factored by the supplied case weights) and the component `y` of the result is the proportion of successes.

Fitting functions

The argument `method` serves two purposes. One is to allow the model frame to be recreated with no fitting. The other is to allow the default fitting function `glm.fit` to be replaced by a function which takes the same arguments and uses a different fitting algorithm. If `glm.fit` is supplied as

... example 10.18

- likelihood ratio test $\beta_1 = 0$
- full model fit $\hat{\beta}$
- reduced model fit $\tilde{\beta}$: $\sup_\beta \ell(0, \beta_2, \ldots \beta_5)$ constrained MLE
-

$$2\{\ell(\hat{\beta}; r) - \ell(\tilde{\beta}; r)\} \dot{\sim} \chi^2_1$$

- likelihood ratio test $\beta_{(1)} = 0$
- full model fit $\hat{\beta}$
- reduced model fit $\tilde{\beta}$: $\sup_\beta \ell(0, \beta_{(2)})$ constrained MLE
-

$$2\{\ell(\hat{\beta}; r) - \ell(\tilde{\beta}; r)\} \dot{\sim} \chi^2_\nu$$

... example 10.18

- likelihood ratio test $\beta_1 = 0$
- full model fit $\hat{\beta}$
- reduced model fit $\tilde{\beta}$: $\sup_\beta \ell(0, \beta_2, \ldots \beta_5)$ constrained MLE
- $$2\{\ell(\hat{\beta}; r) - \ell(\tilde{\beta}; r)\} \dot{\sim} \chi^2_1$$
- likelihood ratio test $\beta_{(1)} = 0$
- full model fit $\hat{\beta}$
- reduced model fit $\tilde{\beta}$: $\sup_\beta \ell(0, \beta_{(2)})$ constrained MLE
- $$2\{\ell(\hat{\beta}; r) - \ell(\tilde{\beta}; r)\} \dot{\sim} \chi^2_\nu$$

... example 10.18

- likelihood ratio test $\beta_1 = 0$
- full model fit $\hat{\beta}$
- reduced model fit $\tilde{\beta}$: $\sup_\beta \ell(0, \beta_2, \ldots \beta_5)$ constrained MLE
- $$2\{\ell(\hat{\beta}; r) - \ell(\tilde{\beta}; r)\} \dot{\sim} \chi^2_1$$
- likelihood ratio test $\beta_{(1)} = 0$
- full model fit $\hat{\beta}$
- reduced model fit $\tilde{\beta}$: $\sup_\beta \ell(0, \beta_{(2)})$ constrained MLE
- $$2\{\ell(\hat{\beta}; r) - \ell(\tilde{\beta}; r)\} \dot{\sim} \chi^2_\nu$$

... example 10.18

- likelihood ratio test $\beta_1 = 0$
- full model fit $\hat{\beta}$
- reduced model fit $\tilde{\beta}$: $\sup_\beta \ell(0, \beta_2, \ldots \beta_5)$ constrained MLE
- $$2\{\ell(\hat{\beta}; r) - \ell(\tilde{\beta}; r)\} \dot{\sim} \chi^2_1$$
- likelihood ratio test $\beta_{(1)} = 0$
- full model fit $\hat{\beta}$
- reduced model fit $\tilde{\beta}$: $\sup_\beta \ell(0, \beta_{(2)})$ constrained MLE
- $$2\{\ell(\hat{\beta}; r) - \ell(\tilde{\beta}; r)\} \dot{\sim} \chi^2_\nu$$

... example 10.18

- likelihood ratio test $\beta_1 = 0$
- full model fit $\hat{\beta}$
- reduced model fit $\tilde{\beta}$: $\sup_\beta \ell(0, \beta_2, \ldots \beta_5)$ constrained MLE
- $$2\{\ell(\hat{\beta}; r) - \ell(\tilde{\beta}; r)\} \dot{\sim} \chi^2_1$$
- likelihood ratio test $\beta_{(1)} = 0$
- full model fit $\hat{\beta}$
- reduced model fit $\tilde{\beta}$: $\sup_\beta \ell(0, \beta_{(2)})$ constrained MLE
- $$2\{\ell(\hat{\beta}; r) - \ell(\tilde{\beta}; r)\} \dot{\sim} \chi^2_\nu$$

... example 10.18

- likelihood ratio test $\beta_1 = 0$
- full model fit $\hat{\beta}$
- reduced model fit $\tilde{\beta}$: $\sup_\beta \ell(0, \beta_2, \ldots \beta_5)$ constrained MLE
- $$2\{\ell(\hat{\beta}; r) - \ell(\tilde{\beta}; r)\} \dot{\sim} \chi^2_1$$
- likelihood ratio test $\beta_{(1)} = 0$
- full model fit $\hat{\beta}$
- reduced model fit $\tilde{\beta}$: $\sup_\beta \ell(0, \beta_{(2)})$ constrained MLE
- $$2\{\ell(\hat{\beta}; r) - \ell(\tilde{\beta}; r)\} \dot{\sim} \chi^2_\nu$$

... example 10.18

- likelihood ratio test $\beta_1 = 0$
- full model fit $\hat{\beta}$
- reduced model fit $\tilde{\beta}$: $\sup_\beta \ell(0, \beta_2, \ldots \beta_5)$ constrained MLE
- $$2\{\ell(\hat{\beta}; r) - \ell(\tilde{\beta}; r)\} \dot{\sim} \chi^2_1$$
- likelihood ratio test $\beta_{(1)} = 0$
- full model fit $\hat{\beta}$
- reduced model fit $\tilde{\beta}$: $\sup_\beta \ell(0, \beta_{(2)})$ constrained MLE
- $$2\{\ell(\hat{\beta}; r) - \ell(\tilde{\beta}; r)\} \dot{\sim} \chi^2_\nu$$

... example 10.18

- likelihood ratio test $\beta_1 = 0$
- full model fit $\hat{\beta}$
- reduced model fit $\tilde{\beta}$: $\sup_\beta \ell(0, \beta_2, \ldots \beta_5)$ constrained MLE
- $$2\{\ell(\hat{\beta}; r) - \ell(\tilde{\beta}; r)\} \dot{\sim} \chi^2_1$$
- likelihood ratio test $\beta_{(1)} = 0$
- full model fit $\hat{\beta}$
- reduced model fit $\tilde{\beta}$: $\sup_\beta \ell(0, \beta_{(2)})$ constrained MLE
- $$2\{\ell(\hat{\beta}; r) - \ell(\tilde{\beta}; r)\} \dot{\sim} \chi^2_\nu$$

... example 10.18

```
> update(ex1018, . ~ . - aged)

Degrees of Freedom: 52 Total (i.e. Null);  48 Residual
Null Deviance:     70.25
Residual Deviance: 47.76  AIC: 57.76

> 47.76 - 47.61
[1] 0.15

> update(ex1018, . ~ . - stage)

Coefficients:
(Intercept)        aged1       grade1        xray1        acid1
    -2.6866      -0.0704       1.4025       1.7479       1.7822

Degrees of Freedom: 52 Total (i.e. Null);  48 Residual
Null Deviance:     70.25
Residual Deviance: 50.81  AIC: 60.81
> 2*(50.81 - 47.61)
[1] 6.4

> pchisq(q=50.81 - 47.61, df=1, lower.tail=F)
[1] 0.07363827
```

difference between residual deviances = log-likelihood ratio

... example 10.18

```
> update(ex1018, . ~ . - age - grade)

Call:  glm(formula = r ~ aged + stage + xray + acid, family = binomial
    data = nodal)

Coefficients:
(Intercept)        aged1       stage1        xray1        acid1
    -2.7777      -0.4698       1.6634       1.8798       1.5521

Degrees of Freedom: 52 Total (i.e. Null);  48 Residual
Null Deviance:     70.25
Residual Deviance: 48.76  AIC: 58.76


> pchisq(48.76-47.61,2,lower.tail=F)
[1] 0.5627049
```

χ^2_2, because we're comparing models with, and without, both `age` and `grade`

... example 10.18

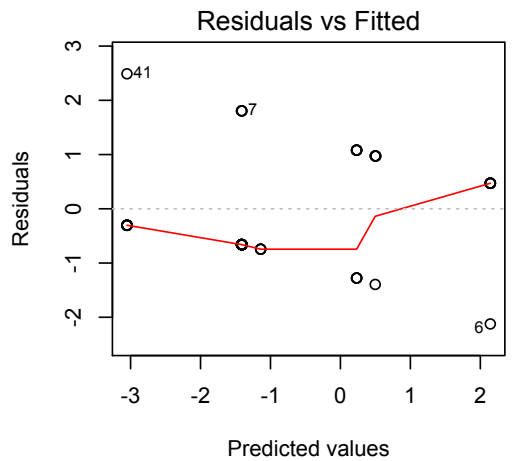
```
> step(ex1018)
Start:  AIC=59.61
cbind(r, m - r) ~ age + stage + grade + xray + acid

        Df Deviance    AIC
- age    1   47.760 57.760
- grade  1   48.760 58.760
<none>       47.611 59.611
- stage  1   50.808 60.808
- acid   1   52.660 62.660
- xray   1   52.922 62.922

Step:  AIC=57.76
cbind(r, m - r) ~ stage + grade + xray + acid

        Df Deviance    AIC
- grade  1   49.180 57.180
<none>       47.760 57.760
- stage  1   50.817 58.817
- xray   1   53.162 61.162
- acid   1   53.526 61.526

Step:  AIC=57.18
cbind(r, m - r) ~ stage + xray + acid

        Df Deviance    AIC
<none>       49.180 57.180
- acid   1   54.463 60.463
- stage  1   54.788 60.788
- xray   1   55.915 61.915
```

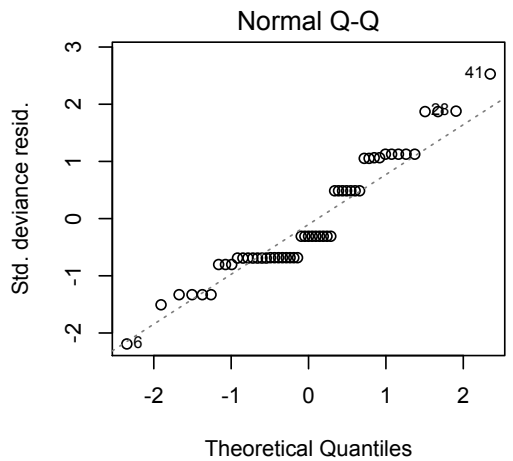
... example 10.18

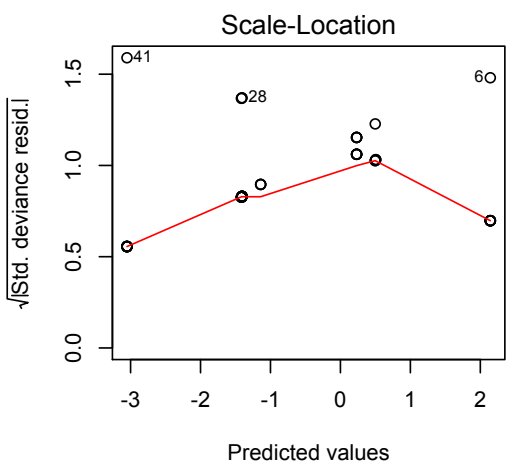
```
Call:  glm(formula = r ~ stage + xray + acid, family = binomial, data = nodal)

Coefficients:
(Intercept)       stage1        xray1        acid1
     -3.052        1.645        1.912        1.638

Degrees of Freedom: 52 Total (i.e. Null);  49 Residual
Null Deviance:     70.25
Residual Deviance: 49.18  AIC: 57.18
> ex1018.final = .Last.value
> summary(ex1018.final) # i.e. final fitted model, compare SM p.491

Call:
glm(formula = r ~ stage + xray + acid, family = binomial, data = nodal)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1231  -0.6620  -0.3039   0.4710   2.4892

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.0518     0.8420  -3.624  0.00029 ***
stage1        1.6453     0.7297   2.255  0.02414 *
xray1         1.9116     0.7771   2.460  0.01390 *
acid1         1.6378     0.7539   2.172  0.02983 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 70.252  on 52  degrees of freedom
Residual deviance: 49.180  on 49  degrees of freedom
AIC: 57.18
```

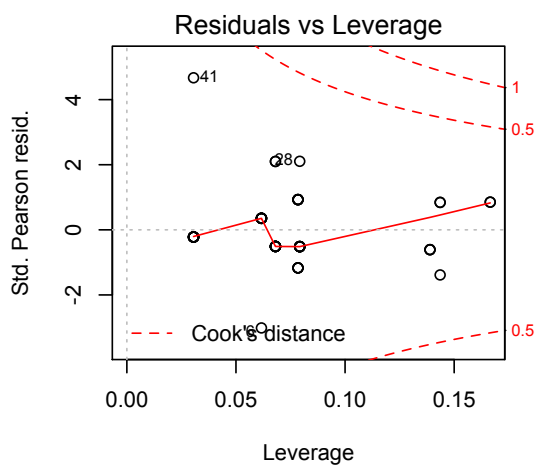
... example 10.18

```
> par(mfrow=c(2,2))
> plot(ex1018.final)
 ##  SM Figure 10.7 is better, but x-axis is p-hat, not X\beta
```









deviance residuals are default

$r_i - \hat{p}_i$ are the 'obvious'

... example 10.18

aggregated data presented in textbook

Table 10.8 Data on nodal involvement (Brown, 1980).

m	*r*	age	stage	grade	xray	acid
6	5	0	1	1	1	1
6	1	0	0	0	0	1
4	0	1	1	1	0	0
4	2	1	1	0	0	1
4	0	0	0	0	0	0
3	2	0	1	1	0	1
3	1	1	1	0	0	0
3	0	1	0	0	0	1
3	0	1	0	0	0	0
2	0	1	0	0	1	0
2	1	0	1	0	0	1
2	1	0	0	1	0	0
1	1	1	1	1	1	1
1	1	1	1	0	1	1
1	1	1	0	1	1	1
1	1	1	0	0	1	1
1	0	1	0	1	0	0
1	1	0	1	1	1	0
1	0	0	1	1	0	0
1	1	0	1	0	1	0
1	1	0	0	1	0	1
1	0	0	0	0	1	1

... example 10.18

- In data set `nodal` several patients have the same value of the covariates
- these can be added up to make a binomial observation
- ```
  > nodal2[1:4,]
    m r age stage grade xray acid
  1 6 5   0     1     1    1    1
  2 6 1   0     0     0    0    1
  3 4 0   1     1     1    0    0
  4 4 2   1     1     0    0    1
  ```
- ```
  > ex1018binom = glm(cbind(r,m-r) ~ ., data = nodal2, family = binomial)
  > summary(ex1018binom)  # stuff omitted
  Coefficients:
              Estimate Std. Error z value Pr(>|z|)
  (Intercept)  -3.0794     0.9868  -3.121  0.00180 **
  age          -0.2917     0.7540  -0.387  0.69881
  stage         1.3729     0.7838   1.752  0.07986 .
  grade         0.8720     0.8156   1.069  0.28500
  xray          1.8008     0.8104   2.222  0.02628 *
  acid          1.6839     0.7915   2.128  0.03337 *
  ---
  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

  (Dispersion parameter for binomial family taken to be 1)

      Null deviance: 40.710  on 22  degrees of freedom
  Residual deviance: 18.069  on 17  degrees of freedom
  AIC: 41.693

  Number of Fisher Scoring iterations: 5
  ```

... example 10.18

```
> step(ex1018binom)

Coefficients:
(Intercept)        stage         xray         acid
     -3.052        1.645        1.912        1.638

Degrees of Freedom: 22 Total (i.e. Null);  19 Residual
Null Deviance:      40.71
Residual Deviance: 19.64  AIC: 39.26
```

– same coefficient estimates; same estimated standard errors

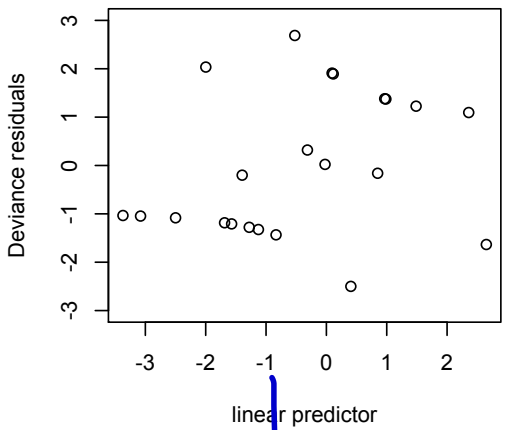
– different residual deviance and different degrees of freedom; same change in deviance

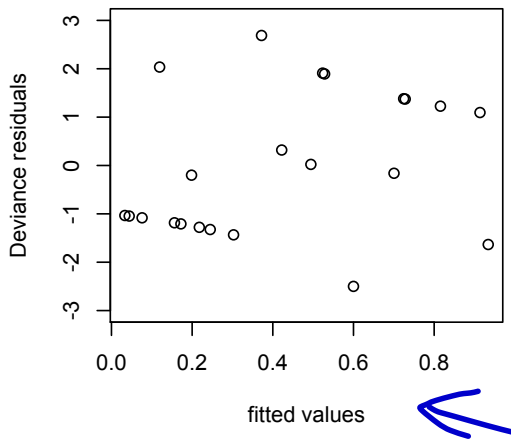
– MISTAKE in text on p. 491; residual scaled deviance is 49.180 on 49 df when fitting to all 53 observations; and cannot be used as a test of fit

– deviances in Table 10.9 are incorrect as well

`http://statwww.epfl.ch/davison/SM/` has corrected version

... example 10.18





$\hat{p}_i(\hat{\beta}) = \dfrac{e^{x_i^T\hat{\beta}}}{1+e^{x_i^T\hat{\beta}}}$

components of resid. dev.

$x_i^T\hat{\beta}$ "linear pred."

- resids ok for binom $\times$ binary

$y_i \sim \text{Bin}(m_i, p_i) \quad i = 1, \ldots, n$

$\hat{p}_i = \frac{y_i}{n_i}$ mle with no regression

$p_i(\hat{\beta}) = e^{x_i^T \hat{\beta}} / (1 + e^{x_i^T \hat{\beta}})$ mle with regression

new notation: $\tilde{\eta}_i = \text{logit}\, \hat{p}_i$

$\hat{\eta}_i = \text{logit}\, p_i(\hat{\beta})$

Residual deviance (scaled)

$$\sum_{i=1}^{n} 2\{l_i(\tilde{\eta}_i; y_i) - l_i(\hat{\eta}_i; y_i)\} \quad (10.2)$$

$l_i = \log f(y_i) \leftarrow 1$ binomial

$$r_P = \frac{y_i - n_i p_i(\hat{\beta})}{\sqrt{n_i p_i(\hat{\beta})(1 - p_i(\hat{\beta}))}}$$

$$r_{D_i} = \pm \sqrt{2\{\quad\}}$$

deviance residual

($m_i \to \infty$ as $n \to \infty$ ~ Normal)

$\sim \chi^2_{23-4}$

β_3 stage xray acid

Under model $p_i(\beta) = e^{x_i^T\beta}/(1+e^{x_i^T\beta})$,

$D \sim \chi^2_{n-p}$ — SM scaled deviance

R v deviance residual

Test of fit for the model.

Works for binomial data, but NOT for binary data. Special feature of binomial.

$$r_i = \pm\sqrt{2\{\ell_i(\tilde{\eta}_i; y_i) - \ell_i(\hat{\eta}_i, y_i)\}}$$

deviance residual

Parameter interpretation

- $$\log \frac{\Pr(Y = 1 \mid x)}{\Pr(Y = 0 \mid x)} = x^T \beta$$
- $$p(x) = \frac{\exp(x^T \beta)}{1 + \exp(x^T \beta)}$$
- odds of 'success' increase by a factor of e^{β_j} for every 1-unit increase in x_j
- thus for Ex 10.8, odds of nodal involvement increase by $e^{1.91}$ when `acid =1`, relative to `acid = 0`
- all other variables held fixed
- "fitted odds when all explanatory variables take their lower levels are $e^{-3.05} = 0.047$"
- corresponds to $\Pr(Y = 1 \mid 0, 0, 0) = 0.045$ ("no such cases in the data" is incorrect)

Link function

- responses $y_i, i = 1, \ldots, n$
- covariates $x_i = 1, \ldots, n$ ($1 \times p$ vectors, rows of X matrix)
- model: systematic component

$$g\{\mathrm{E}(y_i)\} = g(\mu_i) = x_i^{\mathrm{T}}\beta = \eta_i$$

- model: random component

$$y_i \mid x_i \sim f(\cdot; \theta_i)$$

- θ_i is a function of μ_i, is a function of $\beta_1, \ldots, \beta_p$
  - example: $\log\left(\frac{p}{1-p}\right) = x_i^{\mathrm{T}}\beta$ — logit link
  - example: $\Phi^{-1}(p) = x_i^{\mathrm{T}}\beta$ — probit link
  - example: $p = x_i^{\mathrm{T}}\beta$ — identity link
- parameter interpretation depends on the link function, i.e. on the model parameterization

Link function

- responses $y_i, i = 1, \ldots, n$
- covariates $x_i = 1, \ldots, n$ ($1 \times p$ vectors, rows of X matrix)
- model: systematic component

$$g\{\mathrm{E}(y_i)\} = g(\mu_i) = x_i^{\mathrm{T}}\beta = \eta_i$$

- model: random component

$$y_i \mid x_i \sim f(\cdot; \theta_i)$$

- θ_i is a function of μ_i, is a function of $\beta_1, \ldots, \beta_p$
  - example: $\log\left(\frac{p}{1-p}\right) = x_i^{\mathrm{T}}\beta$ – logit link
  - example: $\Phi^{-1}(p) = x_i^{\mathrm{T}}\beta$ – probit link
  - example: $p = x_i^{\mathrm{T}}\beta$ – identity link
- parameter interpretation depends on the link function, i.e. on the model parameterization

Link function

- responses $y_i, i = 1, \ldots, n$
- covariates $x_i = 1, \ldots, n$ ($1 \times p$ vectors, rows of X matrix)

- model: systematic component

$$g\{\mathrm{E}(y_i)\} = g(\mu_i) = x_i^{\mathrm{T}}\beta = \eta_i$$

- model: random component

$$y_i \mid x_i \sim f(\cdot; \theta_i)$$

- θ_i is a function of μ_i, is a function of $\beta_1, \ldots, \beta_p$
  - example: $\log\left(\frac{p}{1-p}\right) = x_i^{\mathrm{T}}\beta$ – logit link
  - example: $\Phi^{-1}(p) = x_i^{\mathrm{T}}\beta$ – probit link
  - example: $p = x_i^{\mathrm{T}}\beta$ – identity link
- parameter interpretation depends on the link function, i.e. on the model parameterization

Link function

- responses $y_i, i = 1, \ldots, n$
- covariates $x_i = 1, \ldots, n$ ($1 \times p$ vectors, rows of X matrix)

- model: systematic component

$$g\{\mathrm{E}(y_i)\} = g(\mu_i) = x_i^{\mathrm{T}}\beta = \eta_i$$

- model: random component

$$y_i \mid x_i \sim f(\cdot; \theta_i)$$

- θ_i is a function of μ_i, is a function of $\beta_1, \ldots, \beta_p$
  - example: $\log \frac{p_i}{1-p_i} = x_i'\beta$ logit link
  - example: $\Phi^{-1}(p_i) = x_i'\beta$ probit link
  - example: $p_i = x_i'\beta$ identity link
- parameter interpretation depends on the link function, i.e. on the model parameterization

Link function

- responses $y_i, i = 1, \ldots, n$
- covariates $x_i = 1, \ldots, n$ ($1 \times p$ vectors, rows of X matrix)

- model: systematic component

$$g\{\mathrm{E}(y_i)\} = g(\mu_i) = x_i^{\mathrm{T}}\beta = \eta_i$$

- model: random component

$$y_i \mid x_i \sim f(\cdot; \theta_i)$$

- θ_i is a function of μ_i, is a function of $\beta_1, \ldots, \beta_p$
  - example: $\log \frac{p_i}{1-p_i} = x_i^T\beta$ logit link
  - example: $\Phi^{-1}(p_i) = x_i^T\beta$ probit link
  - example: $p_i = x_i^T\beta$ identity link
- parameter interpretation depends on the link function, i.e. on the model parameterization

Link function

- responses $y_i, i = 1, \ldots, n$
- covariates $x_i = 1, \ldots, n$ ($1 \times p$ vectors, rows of X matrix)

- model: systematic component

$$g\{\mathrm{E}(y_i)\} = g(\mu_i) = x_i^{\mathrm{T}}\beta = \eta_i$$

- model: random component

$$y_i \mid x_i \sim f(\cdot; \theta_i)$$

- θ_i is a function of μ_i, is a function of $\beta_1, \ldots, \beta_p$
  - example: $\log \frac{p_i}{1-p_i} = x_i^{\mathrm{T}}\beta$ logit link
  - example: $\Phi^{-1}(p_i) = x_i^{\mathrm{T}}\beta$ probit link
  - example: $p_i = x_i^{\mathrm{T}}\beta$ identity link
- parameter interpretation depends on the link function, i.e. on the model parameterization

Bin $y_i = \frac{r_i}{m_i}$

$E(y_i) = p_i$

Link function

- responses $y_i, i = 1, \ldots, n$
- covariates $x_i = 1, \ldots, n$ ($1 \times p$ vectors, rows of X matrix)

- model: systematic component

$$g\{\mathrm{E}(y_i)\} = g(\mu_i) = x_i^{\mathrm{T}}\beta = \eta_i$$

- model: random component

$$y_i \mid x_i \sim f(\cdot; \theta_i)$$

- θ_i is a function of μ_i, is a function of $\beta_1, \ldots, \beta_p$
  - example: $\log \frac{p_i}{1-p_i} = x_i^{\mathrm{T}}\beta$ logit link
  - example: $\Phi^{-1}(p_i) = x_i^{\mathrm{T}}\beta$ probit link
  - example: $p_i = x_i^{\mathrm{T}}\beta$ identity link
- parameter interpretation depends on the link function, i.e. on the model parameterization

Link function

- responses $y_i, i = 1, \ldots, n$
- covariates $x_i = 1, \ldots, n$ ($1 \times p$ vectors, rows of X matrix)

- model: systematic component

$$g\{\mathrm{E}(y_i)\} = g(\mu_i) = x_i^{\mathrm{T}}\beta = \eta_i$$

- model: random component

$$y_i \mid x_i \sim f(\cdot; \theta_i)$$

- θ_i is a function of μ_i, is a function of $\beta_1, \ldots, \beta_p$
  - example: $\log \frac{p_i}{1-p_i} = x_i^{\mathrm{T}}\beta$ logit link
  - example: $\Phi^{-1}(p_i) = x_i^{\mathrm{T}}\beta$ probit link
  - example: $p_i = x_i^{\mathrm{T}}\beta$ identity link
- parameter interpretation depends on the link function, i.e. on the model parameterization

Link function

- responses $y_i, i = 1, \ldots, n$
- covariates $x_i = 1, \ldots, n$ ($1 \times p$ vectors, rows of X matrix)

- model: systematic component

$$g\{\mathrm{E}(y_i)\} = g(\mu_i) = x_i^{\mathrm{T}}\beta = \eta_i$$

- model: random component

$$y_i \mid x_i \sim f(\cdot; \theta_i)$$

- θ_i is a function of μ_i, is a function of $\beta_1, \ldots, \beta_p$
  - example: $\log \frac{p_i}{1-p_i} = x_i^{\mathrm{T}}\beta$ logit link
  - example: $\Phi^{-1}(p_i) = x_i^{\mathrm{T}}\beta$ probit link
  - example: $p_i = x_i^{\mathrm{T}}\beta$ identity link
- parameter interpretation depends on the link function, i.e. on the model parameterization

family {stats} R Documentation

Family Objects for Models

Description

Family objects provide a convenient way to specify the details of the models used by functions such as glm. See the documentation for glm for the details on how such model fitting takes place.

Usage

§6.3 §10.2

```
family(object, ...)

binomial(link = "logit")
gaussian(link = "identity")
Gamma(link = "inverse")
inverse.gaussian(link = "1/mu^2")
poisson(link = "log")
quasi(link = "identity", variance = "constant")
quasibinomial(link = "logit")
quasipoisson(link = "log")
```

$g(EY_i) = \eta_i = x_i^T \beta$

↑ nonlinear

Arguments

link	a specification for the model link function. This can be a name/expression, a literal character string, a length-one character vector or an object of class "link-glm" (such as generated by make.link) provided it is not specified *via* one of the standard names given next. The gaussian family accepts the links (as names) identity, log and inverse; the binomial family the links logit, probit, cauchit, (corresponding to logistic, normal and Cauchy CDFs respectively) log and cloglog (complementary log-log); the Gamma family the links inverse, identity and log; the poisson family the links log, identity, and sqrt and the inverse.gaussian family the links 1/mu^2, inverse, identity and log. The quasi family accepts the links logit, probit, cloglog, identity, inverse, log, 1/mu^2 and sqrt, and the function power can be used to create a power link function.
variance	for all families other than quasi, the variance function is determined by the family. The quasi family will accept the literal character string (or unquoted as a name/expression) specifications "constant", "mu(1-mu)", "mu", "mu^2" and "mu^3", a length-one character vector taking one of those values, or a list containing components varfun, validmu, dev.resids, initialize and name.
object	the function family accesses the family objects which are stored within objects created by modelling functions (e.g., glm).

AIC

- as terms are added to the model, deviance always decreases
- because log-likelihood function always increases
- similar to residual sum of squares
- Akaike Information Criterion penalizes models with more parameters
- $$AIC = 2\{-\ell(\hat{\beta}; y) + p\}$$

  SM (4.57)
- comparison of two model fits by difference in AIC

AIC

- as terms are added to the model, deviance always decreases
- because log-likelihood function always increases
- similar to residual sum of squares
- Akaike Information Criterion penalizes models with more parameters
- $$AIC = 2\{-\ell(\hat{\beta}; y) + p\}$$

  SM (4.57)
- comparison of two model fits by difference in AIC

AIC

- as terms are added to the model, deviance always decreases
- because log-likelihood function always increases
- similar to residual sum of squares

- Akaike Information Criterion penalizes models with more parameters
- $$AIC = 2\{-\ell(\hat{\beta}; y) + p\}$$

  SM (4.57)
- comparison of two model fits by difference in AIC

AIC

- as terms are added to the model, deviance always decreases
- because log-likelihood function always increases
- similar to residual sum of squares

- Akaike Information Criterion penalizes models with more parameters
- $$AIC = 2\{-\ell(\hat{\beta}; y) + p\}$$

  SM (4.57)
- comparison of two model fits by difference in AIC

AIC

- as terms are added to the model, deviance always decreases
- because log-likelihood function always increases
- similar to residual sum of squares

- Akaike Information Criterion penalizes models with more parameters
- $$AIC = 2\{-\ell(\hat{\beta}; y) + p\}$$

SM (4.57)

- comparison of two model fits by difference in *AIC*

AIC

- as terms are added to the model, deviance always decreases
- because log-likelihood function always increases
- similar to residual sum of squares

- Akaike Information Criterion penalizes models with more parameters
- $$AIC = 2\{-\ell(\hat{\beta}; y) + p\}$$

  SM (4.57)
- comparison of two model fits by difference in *AIC*

... AIC

```
> step(ex1018)
Start:  AIC=59.61
cbind(r, m - r) ~ age + stage + grade + xray + acid

        Df Deviance    AIC
- age    1   47.760 57.760
- grade  1   48.760 58.760
<none>       47.611 59.611
- stage  1   50.808 60.808
- acid   1   52.660 62.660
- xray   1   52.922 62.922

Step:  AIC=57.76
cbind(r, m - r) ~ stage + grade + xray + acid

        Df Deviance    AIC
- grade  1   49.180 57.180
<none>       47.760 57.760
- stage  1   50.817 58.817
- xray   1   53.162 61.162
- acid   1   53.526 61.526

Step:  AIC=57.18
cbind(r, m - r) ~ stage + xray + acid

        Df Deviance    AIC
<none>       49.180 57.180
- acid   1   54.463 60.463
- stage  1   54.788 60.788
- xray   1   55.915 61.915
```

Generalized linear models: theory

- $$f(y_j; \mu_j, \phi_j) = \exp\{\frac{y_j\theta_j - b(\theta_j)}{\phi_j} + c(y_j; \phi_j)\}$$
- $\mathrm{E}(y_j \mid x_j) = b'(\theta_j) = \mu_j$ defines μ_j as a function of θ_j
- $g(\mu_j) = x_j^T\beta = \eta_j$ links the n observations together via covariates
- $g(\cdot)$ is the link function; η_j is the linear predictor
- $\mathrm{Var}(y_j \mid x_j) = \phi b''(\theta_j) = \phi V(\mu_j)$
- $V(\cdot)$ is the variance function

Generalized linear models: theory

- $$f(y_j; \mu_j, \phi_j) = \exp\{\frac{y_j\theta_j - b(\theta_j)}{\phi_j} + c(y_j; \phi_j)\}$$
- $\mathrm{E}(y_j \mid x_j) = b'(\theta_j) = \mu_j$ defines μ_j as a function of θ_j
- $g(\mu_j) = x_j^T\beta = \eta_j$ links the n observations together via covariates
- $g(\cdot)$ is the link function; η_j is the linear predictor
- $\mathrm{Var}(y_j \mid x_j) = \phi b''(\theta_j) = \phi V(\mu_j)$
- $V(\cdot)$ is the variance function

Generalized linear models: theory

- $$f(y_j; \mu_j, \phi_j) = \exp\{\frac{y_j\theta_j - b(\theta_j)}{\phi_j} + c(y_j; \phi_j)\}$$
- $\mathrm{E}(y_j \mid x_j) = b'(\theta_j) = \mu_j$ defines μ_j as a function of θ_j
- $g(\mu_j) = x_j^T\beta = \eta_j$ links the n observations together via covariates
- $g(\cdot)$ is the link function; η_j is the linear predictor
- $\mathrm{Var}(y_j \mid x_j) = \phi b''(\theta_j) = \phi V(\mu_j)$
- $V(\cdot)$ is the variance function

Generalized linear models: theory

- $$f(y_j; \mu_j, \phi_j) = \exp\{\frac{y_j\theta_j - b(\theta_j)}{\phi_j} + c(y_j; \phi_j)\}$$
- $\mathrm{E}(y_j \mid x_j) = b'(\theta_j) = \mu_j$ defines μ_j as a function of θ_j
- $g(\mu_j) = x_j^T\beta = \eta_j$ links the n observations together via covariates
- $g(\cdot)$ is the link function; η_j is the linear predictor
- $\mathrm{Var}(y_j \mid x_j) = \phi b''(\theta_j) = \phi V(\mu_j)$
- $V(\cdot)$ is the variance function

Generalized linear models: theory

- $$f(y_j; \mu_j, \phi_j) = \exp\{\frac{y_j\theta_j - b(\theta_j)}{\phi_j} + c(y_j; \phi_j)\}$$
- $\mathrm{E}(y_j \mid x_j) = b'(\theta_j) = \mu_j$ defines μ_j as a function of θ_j
- $g(\mu_j) = x_j^T\beta = \eta_j$ links the n observations together via covariates
- $g(\cdot)$ is the link function; η_j is the linear predictor
- $\mathrm{Var}(y_j \mid x_j) = \phi b''(\theta_j) = \phi V(\mu_j)$
- $V(\cdot)$ is the variance function

Generalized linear models: theory

$$f(y_j; \mu_j, \phi_j) = \exp\{\frac{y_j\theta_j - b(\theta_j)}{\phi_j} + c(y_j; \phi_j)\}$$

- $E(y_j \mid x_j) = b'(\theta_j) = \mu_j$ defines μ_j as a function of θ_j
- $g(\mu_j) = x_j^T\beta = \eta_j$ links the n observations together via covariates
- $g(\cdot)$ is the link function; η_j is the linear predictor
- $\text{Var}(y_j \mid x_j) = \phi b''(\theta_j) = \phi V(\mu_j)$
- $V(\cdot)$ is the variance function

ϕ Scale parameter

but usually $\phi_j = \phi a_j$, a_j known

Examples

- Normal
- Binomial
- Poisson
- Gamma/Exponential
- Inverse Gaussian

family {stats} R Documentation

Family Objects for Models

Description

Family objects provide a convenient way to specify the details of the models used by functions such as glm. See the documentation for glm for the details on how such model fitting takes place.

Usage

```
family(object, ...)

binomial(link = "logit")
gaussian(link = "identity")
Gamma(link = "inverse")
inverse.gaussian(link = "1/mu^2")
poisson(link = "log")
quasi(link = "identity", variance = "constant")
quasibinomial(link = "logit")
quasipoisson(link = "log")
```

Scale parameter ϕ_j

- in most cases, either ϕ_j is known, or $\phi_j = \phi a_j$, where a_j is known
- Normal distribution, $\phi =$
- Binomial distribution $\phi_j =$
- Gamma distribution, $\phi =$
- maximum likelihood estimate of ϕ may be poor (by analogy with normal theory linear model)
- $$\hat{\phi} = \frac{1}{n-p} \sum_{j=1}^{n} \frac{(y_j - \hat{\mu}_j)^2}{a_j V(\hat{\mu}_j)}$$

Scale parameter ϕ_j

- in most cases, either ϕ_j is known, or $\phi_j = \phi a_j$, where a_j is known
- Normal distribution, $\phi =$
- Binomial distribution $\phi_j =$
- Gamma distribution, $\phi =$
- maximum likelihood estimate of ϕ may be poor (by analogy with normal theory linear model)
- $$\hat{\phi} = \frac{1}{n-p} \sum_{j=1}^{n} \frac{(y_j - \hat{\mu}_j)^2}{a_j V(\hat{\mu}_j)}$$

Scale parameter ϕ_j

- in most cases, either ϕ_j is known, or $\phi_j = \phi a_j$, where a_j is known
- Normal distribution, $\phi =$
- Binomial distribution $\phi_j =$
- Gamma distribution, $\phi =$
- maximum likelihood estimate of ϕ may be poor (by analogy with normal theory linear model)
- $$\hat{\phi} = \frac{1}{n-p} \sum_{j=1}^{n} \frac{(y_j - \hat{\mu}_j)^2}{a_j V(\hat{\mu}_j)}$$

Scale parameter ϕ_j

- in most cases, either ϕ_j is known, or $\phi_j = \phi a_j$, where a_j is known
- Normal distribution, $\phi =$
- Binomial distribution $\phi_j =$
- Gamma distribution, $\phi =$
- maximum likelihood estimate of ϕ may be poor (by analogy with normal theory linear model)
- $$\hat{\phi} = \frac{1}{n-p} \sum_{j=1}^{n} \frac{(y_j - \hat{\mu}_j)^2}{a_j V(\hat{\mu}_j)}$$

Scale parameter ϕ_j

- in most cases, either ϕ_j is known, or $\phi_j = \phi a_j$, where a_j is known
- Normal distribution, $\phi =$
- Binomial distribution $\phi_j =$
- Gamma distribution, $\phi =$
- maximum likelihood estimate of ϕ may be poor (by analogy with normal theory linear model)
- $$\hat{\phi} = \frac{1}{n-p} \sum_{j=1}^{n} \frac{(y_j - \hat{\mu}_j)^2}{a_j V(\hat{\mu}_j)}$$

Scale parameter ϕ_j

- in most cases, either ϕ_j is known, or $\phi_j = \phi a_j$, where a_j is known
- Normal distribution, $\phi =$
- Binomial distribution $\phi_j =$
- Gamma distribution, $\phi =$
- maximum likelihood estimate of ϕ may be poor (by analogy with normal theory linear model)
- $$\hat{\phi} = \frac{1}{n-p} \sum_{j=1}^{n} \frac{(y_j - \hat{\mu}_j)^2}{a_j V(\hat{\mu}_j)}$$

Example 10.29

City	Rain	r/m	City	Rain	r/m	City	Rain	r/m	City	Rain	r/m
1	1735	2/4	11	2050	7/24	21	1756	2/12	31	1780	8/13
2	1936	3/10	12	1830	0/1	22	1650	0/1	32	1900	3/10
3	2000	1/5	13	1650	15/30	23	2250	8/11	33	1976	1/6
4	1973	3/10	14	2200	4/22	24	1796	41/77	34	2292	23/37
5	1750	2/2	15	2000	0/1	25	1890	24/51			
6	1800	3/5	16	1770	6/11	26	1871	7/16			
7	1750	2/8	17	1920	0/1	27	2063	46/82			
8	2077	7/19	18	1770	33/54	28	2100	9/13			
9	1920	3/6	19	2240	4/9	29	1918	23/43			
10	1800	8/10	20	1620	5/18	30	1834	53/75			

Table 10.19 Toxoplamosis data: rainfall (mm) and the numbers of people testing positive for toxoplasmosis, r, our of m people tested, for 34 cities in El Salvador (Efron, 1986).

Terms	df	Deviance
Constant	33	74.21
Linear	32	74.09
Quadratic	31	74.09
Cubic	30	62.63

Table 10.20 Analysis of deviance for polynomial logistic models fitted to the toxoplasmosis data.

- incidence of toxoplasmosis as a function of rainfall
- residual deviances approximately twice the degrees of freedom

Example 10.29

City	Rain	r/m	City	Rain	r/m	City	Rain	r/m	City	Rain	r/m
1	1735	2/4	11	2050	7/24	21	1756	2/12	31	1780	8/13
2	1936	3/10	12	1830	0/1	22	1650	0/1	32	1900	3/10
3	2000	1/5	13	1650	15/30	23	2250	8/11	33	1976	1/6
4	1973	3/10	14	2200	4/22	24	1796	41/77	34	2292	23/37
5	1750	2/2	15	2000	0/1	25	1890	24/51			
6	1800	3/5	16	1770	6/11	26	1871	7/16			
7	1750	2/8	17	1920	0/1	27	2063	46/82			
8	2077	7/19	18	1770	33/54	28	2100	9/13			
9	1920	3/6	19	2240	4/9	29	1918	23/43			
10	1800	8/10	20	1620	5/18	30	1834	53/75			

Table 10.19 Toxoplamosis data: rainfall (mm) and the numbers of people testing positive for toxoplasmosis, r, our of m people tested, for 34 cities in El Salvador (Efron, 1986).

Terms	df	Deviance
Constant	33	74.21
Linear	32	74.09
Quadratic	31	74.09
Cubic	30	62.63

Table 10.20 Analysis of deviance for polynomial logistic models fitted to the toxoplasmosis data.

- incidence of toxoplasmosis as a function of rainfall
- residual deviances approximately twice the degrees of freedom

... example 10.29

```
> data(toxo)
  rain  m  r
1 1620 18  5
2 1650 30 15
3 1650  1  0
4 1735  4  2
> toxo.glm0 = glm(cbind(r,m-r) ~ rain + I(rain^2) + I(rain^3), data = toxo,
family = binomial)

> anova(toxo.glm0)
 ...
          Df Deviance Resid. Df Resid. Dev
NULL                         33     74.212
rain       1   0.1244        32     74.087
I(rain^2)  1   0.0000        31     74.087
I(rain^3)  1  11.4529        30     62.635
> toxo.glm1 = glm(cbind(r,m-r) ~ poly(rain,3), data = toxo, family = binomial)

> summary(toxo.glm1)
...
Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)              0.02427    0.07693   0.315 0.752401
poly(rain, degree = 3)1 -0.08606    0.45870  -0.188 0.851172
poly(rain, degree = 3)2 -0.19269    0.46739  -0.412 0.680141
poly(rain, degree = 3)3  1.37875    0.41150   3.351 0.000806 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 74.212  on 33  degrees of freedom
Residual deviance: 62.635  on 30  degrees of freedom
```

Design of Studies

CD, Ch.2

- common objectives
- to avoid systematic error, that is distortion in the conclusions arising from sources that do not cancel out in the long run
- to reduce the non-systematic (random) error to a reasonable level by replication and other techniques
- to estimate realistically the likely uncertainty in the final conclusions
- to ensure that the scale of effort is appropriate

Design of Studies

CD, Ch.2

- common objectives
- to avoid systematic error, that is distortion in the conclusions arising from sources that do not cancel out in the long run
- to reduce the non-systematic (random) error to a reasonable level by replication and other techniques
- to estimate realistically the likely uncertainty in the final conclusions
- to ensure that the scale of effort is appropriate

Design of Studies

CD, Ch.2

- common objectives
- to avoid systematic error, that is distortion in the conclusions arising from sources that do not cancel out in the long run
- to reduce the non-systematic (random) error to a reasonable level by replication and other techniques
- to estimate realistically the likely uncertainty in the final conclusions
- to ensure that the scale of effort is appropriate

Design of Studies

CD, Ch.2

- common objectives
- to avoid systematic error, that is distortion in the conclusions arising from sources that do not cancel out in the long run
- to reduce the non-systematic (random) error to a reasonable level by replication and other techniques
- to estimate realistically the likely uncertainty in the final conclusions
- to ensure that the scale of effort is appropriate

Design of Studies

CD, Ch.2

- common objectives
- to avoid systematic error, that is distortion in the conclusions arising from sources that do not cancel out in the long run
- to reduce the non-systematic (random) error to a reasonable level by replication and other techniques
- to estimate realistically the likely uncertainty in the final conclusions
- to ensure that the scale of effort is appropriate

... design of studies

- we concentrate largely on the careful analysis of individual studies
- in most situations synthesis of information from different investigations is needed
- but even there the quality of individual studies remains important
- examples include overviews (such as the Cochrane reviews)
- in some areas new investigations can be set up and completed relatively quickly; design of individual studies may then be less important

... design of studies

- we concentrate largely on the careful analysis of individual studies
- in most situations synthesis of information from different investigations is needed
- but even there the quality of individual studies remains important
- examples include overviews (such as the Cochrane reviews)
- in some areas new investigations can be set up and completed relatively quickly; design of individual studies may then be less important

... design of studies

- we concentrate largely on the careful analysis of individual studies
- in most situations synthesis of information from different investigations is needed
- but even there the quality of individual studies remains important
- examples include overviews (such as the Cochrane reviews)
- in some areas new investigations can be set up and completed relatively quickly; design of individual studies may then be less important

... design of studies

- we concentrate largely on the careful analysis of individual studies
- in most situations synthesis of information from different investigations is needed
- but even there the quality of individual studies remains important
- examples include overviews (such as the Cochrane reviews)
- in some areas new investigations can be set up and completed relatively quickly; design of individual studies may then be less important

... design of studies

- we concentrate largely on the careful analysis of individual studies
- in most situations synthesis of information from different investigations is needed
- but even there the quality of individual studies remains important
- examples include overviews (such as the Cochrane reviews)
- in some areas new investigations can be set up and completed relatively quickly; design of individual studies may then be less important

... design of studies

- formulation of a plan of analysis
- establish and document that proposed data are capable of addressing the research questions of concern
- main configurations of answers likely to be obtained should be set out
- level of detail depends on the context
- even if pre-specified methods must be used, it is crucial not to limit analysis
- planned analysis may be technically inappropriate
- more controversially, data may suggest new research questions or replacement of objectives
- latter will require confirmatory studies

... design of studies

- formulation of a plan of analysis
- establish and document that proposed data are capable of addressing the research questions of concern
- main configurations of answers likely to be obtained should be set out
- level of detail depends on the context
- even if pre-specified methods must be used, it is crucial not to limit analysis
- planned analysis may be technically inappropriate
- more controversially, data may suggest new research questions or replacement of objectives
- latter will require confirmatory studies

... design of studies

- formulation of a plan of analysis
- establish and document that proposed data are capable of addressing the research questions of concern
- main configurations of answers likely to be obtained should be set out
- level of detail depends on the context
- even if pre-specified methods must be used, it is crucial not to limit analysis
- planned analysis may be technically inappropriate
- more controversially, data may suggest new research questions or replacement of objectives
- latter will require confirmatory studies

... design of studies

- formulation of a plan of analysis
- establish and document that proposed data are capable of addressing the research questions of concern
- main configurations of answers likely to be obtained should be set out
- level of detail depends on the context
- even if pre-specified methods must be used, it is crucial not to limit analysis
- planned analysis may be technically inappropriate
- more controversially, data may suggest new research questions or replacement of objectives
- latter will require confirmatory studies

... design of studies

- formulation of a plan of analysis
- establish and document that proposed data are capable of addressing the research questions of concern
- main configurations of answers likely to be obtained should be set out
- level of detail depends on the context
- even if pre-specified methods must be used, it is crucial not to limit analysis
- planned analysis may be technically inappropriate
- more controversially, data may suggest new research questions or replacement of objectives
- latter will require confirmatory studies

... design of studies

- formulation of a plan of analysis
- establish and document that proposed data are capable of addressing the research questions of concern
- main configurations of answers likely to be obtained should be set out
- level of detail depends on the context
- even if pre-specified methods must be used, it is crucial not to limit analysis
- planned analysis may be technically inappropriate
- more controversially, data may suggest new research questions or replacement of objectives
- latter will require confirmatory studies

... design of studies

- formulation of a plan of analysis
- establish and document that proposed data are capable of addressing the research questions of concern
- main configurations of answers likely to be obtained should be set out
- level of detail depends on the context
- even if pre-specified methods must be used, it is crucial not to limit analysis
- planned analysis may be technically inappropriate
- more controversially, data may suggest new research questions or replacement of objectives
- latter will require confirmatory studies

... design of studies

- formulation of a plan of analysis
- establish and document that proposed data are capable of addressing the research questions of concern
- main configurations of answers likely to be obtained should be set out
- level of detail depends on the context
- even if pre-specified methods must be used, it is crucial not to limit analysis
- planned analysis may be technically inappropriate
- more controversially, data may suggest new research questions or replacement of objectives
- latter will require confirmatory studies

Unit of study and analysis

- smallest subdivision of experimental material that may be assigned to a treatment
- Example: RCT – unit may be a patient, or a patient-month (in crossover trial)
- Example: public health intervention – unit is often a community/school/...
- split plot experiments have two classes of units of study and analysis
- in investigations that are not randomized, it may be helpful to consider what the primary unit of analysis would have been, had a randomized experiment been feasible
- the unit of analysis may not be the unit of interpretation – ecological bias
- on the whole, limited detail is needed in examining the variation within the unit of study

Unit of study and analysis

- smallest subdivision of experimental material that may be assigned to a treatment
- Example: RCT – unit may be a patient, or a patient-month (in crossover trial)
- Example: public health intervention – unit is often a community/school/...
- split plot experiments have two classes of units of study and analysis
- in investigations that are not randomized, it may be helpful to consider what the primary unit of analysis would have been, had a randomized experiment been feasible
- the unit of analysis may not be the unit of interpretation – ecological bias
- on the whole, limited detail is needed in examining the variation within the unit of study

Unit of study and analysis

- smallest subdivision of experimental material that may be assigned to a treatment
- Example: RCT – unit may be a patient, or a patient-month (in crossover trial)
- Example: public health intervention – unit is often a community/school/...
- split plot experiments have two classes of units of study and analysis
- in investigations that are not randomized, it may be helpful to consider what the primary unit of analysis would have been, had a randomized experiment been feasible
- the unit of analysis may not be the unit of interpretation – ecological bias
- on the whole, limited detail is needed in examining the variation within the unit of study

Unit of study and analysis

- smallest subdivision of experimental material that may be assigned to a treatment
- Example: RCT – unit may be a patient, or a patient-month (in crossover trial)
- Example: public health intervention – unit is often a community/school/...
- split plot experiments have two classes of units of study and analysis
- in investigations that are not randomized, it may be helpful to consider what the primary unit of analysis would have been, had a randomized experiment been feasible
- the unit of analysis may not be the unit of interpretation – ecological bias
- on the whole, limited detail is needed in examining the variation within the unit of study

Unit of study and analysis

- smallest subdivision of experimental material that may be assigned to a treatment
- Example: RCT – unit may be a patient, or a patient-month (in crossover trial)
- Example: public health intervention – unit is often a community/school/...
- split plot experiments have two classes of units of study and analysis
- in investigations that are not randomized, it may be helpful to consider what the primary unit of analysis would have been, had a randomized experiment been feasible
- the unit of analysis may not be the unit of interpretation – ecological bias
- on the whole, limited detail is needed in examining the variation within the unit of study

Unit of study and analysis

- smallest subdivision of experimental material that may be assigned to a treatment
- Example: RCT – unit may be a patient, or a patient-month (in crossover trial)
- Example: public health intervention – unit is often a community/school/...
- split plot experiments have two classes of units of study and analysis
- in investigations that are not randomized, it may be helpful to consider what the primary unit of analysis would have been, had a randomized experiment been feasible
- the unit of analysis may not be the unit of interpretation – ecological bias
- on the whole, limited detail is needed in examining the variation within the unit of study

Unit of study and analysis

- smallest subdivision of experimental material that may be assigned to a treatment
- Example: RCT – unit may be a patient, or a patient-month (in crossover trial)
- Example: public health intervention – unit is often a community/school/...
- split plot experiments have two classes of units of study and analysis
- in investigations that are not randomized, it may be helpful to consider what the primary unit of analysis would have been, had a randomized experiment been feasible
- the unit of analysis may not be the unit of interpretation – ecological bias
- on the whole, limited detail is needed in examining the variation within the unit of study

Types of observational studies

- secondary analysis of data collected for another purpose
- estimation of a some feature of a defined population (could in principle be found exactly)
- tracking across time of such features
- study of a relationship between features, where individuals may be examined
  - at a single time point
  - at several time points for different individuals
  - at different time points for the same individual
- experiment: investigator has complete control over treatment assignment
- census
- meta-analysis: statistical assessment of a collection of studies on the same topic

Types of observational studies

- secondary analysis of data collected for another purpose
- estimation of a some feature of a defined population (could in principle be found exactly)
- tracking across time of such features
- study of a relationship between features, where individuals may be examined
  - at a single time point
  - at several time points for different individuals
  - at different time points for the same individual
- experiment: investigator has complete control over treatment assignment
- census
- meta-analysis: statistical assessment of a collection of studies on the same topic

Types of observational studies

- secondary analysis of data collected for another purpose
- estimation of a some feature of a defined population (could in principle be found exactly)
- tracking across time of such features
- study of a relationship between features, where individuals may be examined
  - at a single time point
  - at several time points for different individuals
  - at different time points for the same individual
- experiment: investigator has complete control over treatment assignment
- census
- meta-analysis: statistical assessment of a collection of studies on the same topic

Types of observational studies

- secondary analysis of data collected for another purpose
- estimation of a some feature of a defined population (could in principle be found exactly)
- tracking across time of such features
- study of a relationship between features, where individuals may be examined
  - at a single time point
  - at several time points for different individuals
  - at different time points for the same individual
- experiment: investigator has complete control over treatment assignment
- census
- meta-analysis: statistical assessment of a collection of studies on the same topic

Types of observational studies

- secondary analysis of data collected for another purpose
- estimation of a some feature of a defined population (could in principle be found exactly)
- tracking across time of such features
- study of a relationship between features, where individuals may be examined
  - at a single time point
  - at several time points for different individuals
  - at different time points for the same individual
- experiment: investigator has complete control over treatment assignment
- census
- meta-analysis: statistical assessment of a collection of studies on the same topic

Types of observational studies

- secondary analysis of data collected for another purpose
- estimation of a some feature of a defined population (could in principle be found exactly)
- tracking across time of such features
- study of a relationship between features, where individuals may be examined
  - at a single time point
  - at several time points for different individuals
  - at different time points for the same individual
- experiment: investigator has complete control over treatment assignment
- census
- meta-analysis: statistical assessment of a collection of studies on the same topic

Types of observational studies

- secondary analysis of data collected for another purpose
- estimation of a some feature of a defined population (could in principle be found exactly)
- tracking across time of such features
- study of a relationship between features, where individuals may be examined
  - at a single time point
  - at several time points for different individuals
  - at different time points for the same individual
- experiment: investigator has complete control over treatment assignment
- census
- meta-analysis: statistical assessment of a collection of studies on the same topic

Types of observational studies

- secondary analysis of data collected for another purpose
- estimation of a some feature of a defined population (could in principle be found exactly)
- tracking across time of such features
- study of a relationship between features, where individuals may be examined
  - at a single time point
  - at several time points for different individuals
  - at different time points for the same individual
- experiment: investigator has complete control over treatment assignment
- census
- meta-analysis: statistical assessment of a collection of studies on the same topic

Types of observational studies

- secondary analysis of data collected for another purpose
- estimation of a some feature of a defined population (could in principle be found exactly)
- tracking across time of such features
- study of a relationship between features, where individuals may be examined
  - at a single time point
  - at several time points for different individuals
  - at different time points for the same individual
- experiment: investigator has complete control over treatment assignment
- census
- meta-analysis: statistical assessment of a collection of studies on the same topic

Types of observational studies

- secondary analysis of data collected for another purpose
- estimation of a some feature of a defined population (could in principle be found exactly)
- tracking across time of such features
- study of a relationship between features, where individuals may be examined
  - at a single time point
  - at several time points for different individuals
  - at different time points for the same individual
- experiment: investigator has complete control over treatment assignment
- census
- meta-analysis: statistical assessment of a collection of studies on the same topic

Avoidance of systematic error

CD §2.4

- "distortion in the conclusions arising from irrelevant sources that do not cancel out in the long run"
- can arise through systematic aspects of, for example, a measuring process, or the spatial or temporal arrangement of units
- this can often be avoided by design, or adjustment in analysis
- can arise by the entry of personal judgement into some aspect of the data collection process
- this can often be avoided by randomization and blinding

Avoidance of systematic error

CD §2.4

- "distortion in the conclusions arising from irrelevant sources that do not cancel out in the long run"
- can arise through systematic aspects of, for example, a measuring process, or the spatial or temporal arrangement of units
- this can often be avoided by design, or adjustment in analysis
- can arise by the entry of personal judgement into some aspect of the data collection process
- this can often be avoided by randomization and blinding

Avoidance of systematic error

CD §2.4

- "distortion in the conclusions arising from irrelevant sources that do not cancel out in the long run"
- can arise through systematic aspects of, for example, a measuring process, or the spatial or temporal arrangement of units
- this can often be avoided by design, or adjustment in analysis
- can arise by the entry of personal judgement into some aspect of the data collection process
- this can often be avoided by randomization and blinding

Avoidance of systematic error

CD §2.4

- "distortion in the conclusions arising from irrelevant sources that do not cancel out in the long run"
- can arise through systematic aspects of, for example, a measuring process, or the spatial or temporal arrangement of units
- this can often be avoided by design, or adjustment in analysis
- can arise by the entry of personal judgement into some aspect of the data collection process
- this can often be avoided by randomization and blinding

Avoidance of systematic error

CD §2.4

- "distortion in the conclusions arising from irrelevant sources that do not cancel out in the long run"
- can arise through systematic aspects of, for example, a measuring process, or the spatial or temporal arrangement of units
- this can often be avoided by design, or adjustment in analysis
- can arise by the entry of personal judgement into some aspect of the data collection process
- this can often be avoided by randomization and blinding

In the News

THE GLOBE AND MAIL

January 12, 2014

The placebo effect: A new study underscores its remarkable power

By ADRIANA BARTON

Research suggests a placebo's therapeutic impact can be so strong that even patients who know they're taking a sugar pill may start feeling better

In the not too distant future, your family doctor's first line of treatment for minor illnesses such as migraine and irritable bowel syndrome may well be snake oil. Prescribing a placebo, or sugar pill, is a stealthy way to raise a patient's expectations of getting better.

But according to new research, the therapeutic effects of a placebo are so powerful that an inert pill has a good chance of reducing symptoms – even if patients know they are taking a dummy pill.

Harvard researcher Dr. Ted Kaptchuk made this counterintuitive conclusion in a study published last week in Science Translational Medicine. Kaptchuk and colleagues found that the placebo effect greatly enhanced pain relief in migraine sufferers who had the expectation they were getting an effective drug, compared to when they took the active drug with the incorrect label "placebo." More surprising, however, is that the patients reported significant pain relief, compared to an untreated migraine attack, even when they knew they had swallowed nothing more than a sugar pill.

... in the News

Kam-Hansen, et al., Science Translational Medicine

Altered Placebo and Drug Labeling Changes the Outcome of Migraine Attacks

Slavenka Kam-Hansen *et al.*
Sci Transl Med **6**, 218ra5 (2014);
DOI: 10.1126/scitranslmed.3006175

Editor's Summary

Placebo and Medication Effects in Episodic Migraine

Placebo and medication effects are intimately related in clinical practice and drug development. Kam-Hansen *et al.* investigated how information—ranging from "negative" to "neutral" to "positive"—patients, who received either active drug or placebo, modified their headache pain as measured by pa pain scores. In a randomized order over six consecutive attacks, 66 patients with episodic migraine re placebo or Maxalt (10-mg rizatriptan) under three information conditions (told placebo, told Maxalt or p Maxalt). Each participant also reported on an initial no-treatment attack, yielding a total of 459 docume attacks. Maxalt was superior to placebo for pain relief. Increasing information from negative to neutral progressively enhanced the effects of both placebo and Maxalt. The efficacy of open-label placebo wa that of no treatment. Relative to no treatment, the placebo, under each information condition, accounte 50% of the drug effect. The benefits of placebo persisted even when the placebo was honestly describ treatment involves medication or placebo, the information provided to patients and the ritual of pill taki components of medical care.

... in the News

Kam-Hansen, et al., Science Translational Medicine

www.sciencetranslationalmedicine.org/cgi/content/full/6/218/218ra5/DC1

Supplementary Materials for

Altered Placebo and Drug Labeling Changes the Outcome of Episodic Migraine Attacks

Slavenka Kam-Hansen, Moshe Jakubowski, John M. Kelley, Irving Kirsch, David C. Hoaglin, Ted J. Kaptchuk, Rami Burstein*

*Corresponding author. E-mail: rburstei@bidmc.harvard.edu

Published 8 January 2014, *Sci. Transl. Med.* **6**, 218ra5 (2014)
DOI: 10.1126/scitranslmed.3006175

The PDF file includes:

Methods
Scripted Information Read to Participants
Randomization of Treatment and Treatment Labeling

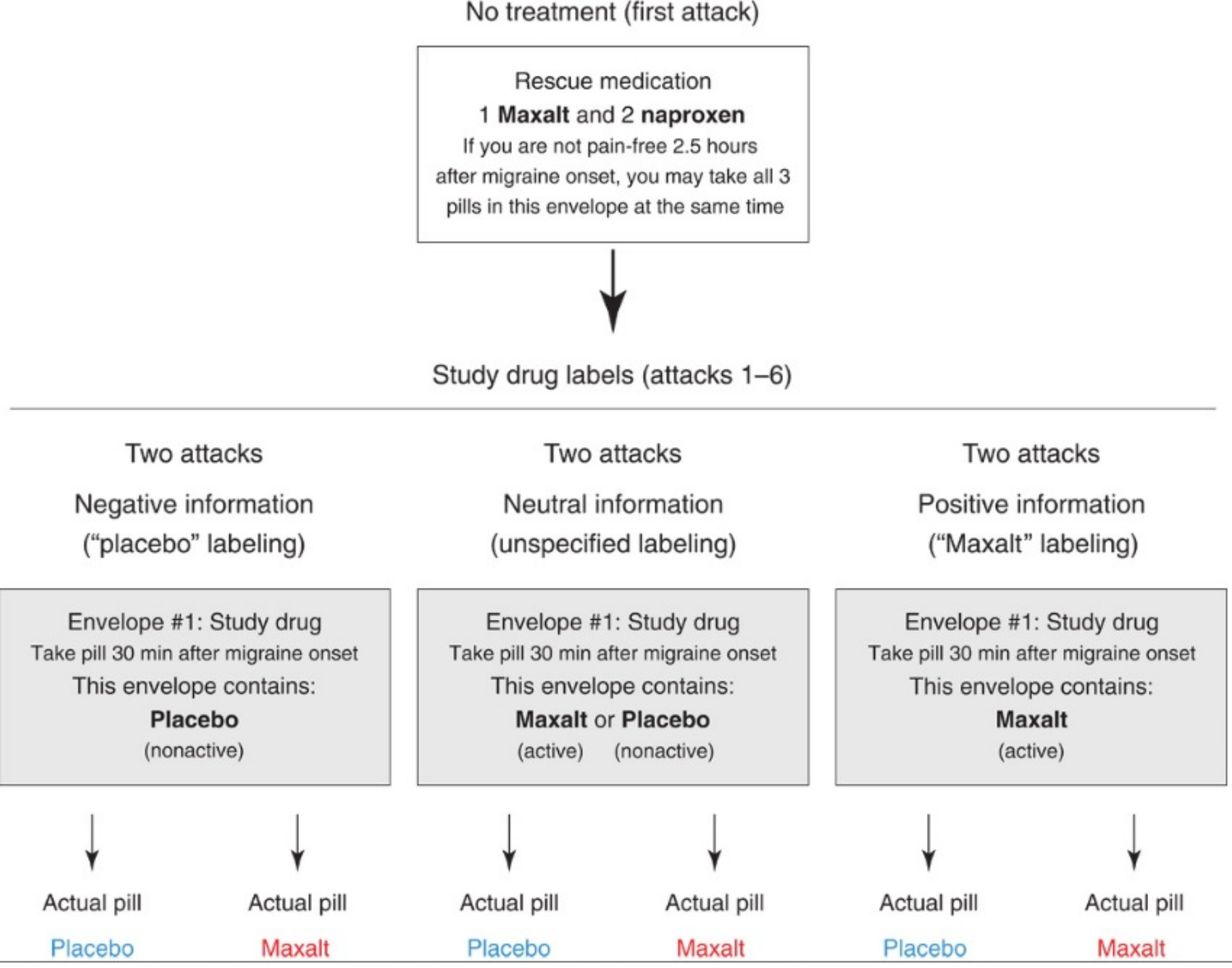

No treatment (first attack)
Rescue medication
1 Maxalt and 2 naproxen
If you are not pain-free 2.5 hours after migraine onset, you may take all 3 pills in this envelope at the same time
Study drug labels (attacks 1–6)
Two attacks
Negative information
("placebo" labeling)
Envelope #1: Study drug
Take pill 30 min after migraine onset
This envelope contains:
Placebo
(nonactive)
Two attacks
Neutral information
(unspecified labeling)
Envelope #1: Study drug
Take pill 30 min after migraine onset
This envelope contains:
Maxalt or Placebo
(active) (nonactive)
Two attacks
Positive information
("Maxalt" labeling)
Envelope #1: Study drug
Take pill 30 min after migraine onset
This envelope contains:
Maxalt
(active)
Actual pill
Placebo
Actual pill
Maxalt
Actual pill
Placebo
Actual pill
Maxalt
Actual pill
Placebo
Actual pill
Maxalt

Table S5. Structure of the eight treatment sequences and assignment of subjects to treatment sequences

Treatment sequence[a]	Treatment conditions						Number of subjects		
	Attack 1	Attack 2	Attack 3	Attack 4	Attack 5	Attack 6	Recruited	Dropped out	Analyzed
5	*M* – M	*M* – P	*P* – M	*P* – P	*U* – M	*U* – P	10	1	9
7	*P* – M	*P* – P	*M* – M	*M* – P	*U* – M	*U* – P	9	2	7
1	*U* – M	*U* – P	*M* – M	*M* – P	*P* – M	*P* – P	9	2	7
3	*U* – M	*U* – P	*P* – M	*P* – P	*M* – M	*M* – P	10	0	10
2	*U* – P	*U* – M	*M* – P	*M* – M	*P* – P	*P* – M	9	2	7
4	*U* – P	*U* – M	*P* – P	*P* – M	*M* – P	*M* – M	9	2	7
6	*M* – P	*M* – M	*P* – P	*P* – M	*U* – P	*U* – M	10	1	9
8	*P* – P	*P* – M	*M* – P	*M* – M	*U* – P	*U* – M	10	0	10
						Totals	**76**	**10**	**66**

The 6 pill/label combinations are abbreviated as follows: the first letter (in *italic*) denotes the label (*M* for 'Maxalt', *P* for 'Placebo', *U* for the unspecified 'Maxalt *or* Placebo'); the second letter (in color) denotes the actual pill (M for maxalt, P for placebo). [a]Sequence numbers correspond to the order they were entered in the GLMM analyses (cf. table S6).