Session 06

Generalized Linear Models 1

Nature of the generalization

- Single response variable, y
- Some candidate predictor variables $x_1, x_2, \ldots, x_p$.
- The distribution of y can only depend on the predictors through a single linear function:

$$\eta = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

- The distribution belongs to the GLM family of distributions
- There may (or may not) be an unknown scale parameter.

© CSIRO, 2000-2007

Distributions in the GLM family

- Normal - Ordinary linear models
- Binomial - Logistic regression, probit analysis
- Poisson - Log-linear models
- Gamma - Alternative to lognormal models
- Negative Binomial, &c

© CSIRO, 2000-2007

Link functions

- It is assumed that the linear predictor determines the mean of the response
- The linear predictor is unbounded, but the mean of some of these distributions (e.g. binomial) is restricted.
- The mean is assumed to be a (monotone) function of the linear predictor
- The inverse of this function is called the link function
- Choosing a link is often the first problem in constructing a GLM.

© CSIRO, 2000-2007

Examples

- Normal – Identity link
- Binomial – Logistic or Probit links
- Poisson – Log or Square-root link
- Gamma – log or inverse link

- For the binomial distribution the response is taken as the proportion of cases responding. Thus the mean lies between 0 and 1. The logistic link uses

$$\mu = \frac{\exp \eta}{1 + \exp \eta}, \qquad \eta = \log\left(\frac{\mu}{1-\mu}\right)$$

© CSIRO, 2000-2007

Why is the link function defined 'backwards'?

- Historical reasons.
- GLM theory was developed as a replacement for an older approximate theory that used transformations of the data
- The link function is defined in the same sense, but the data are never transformed. The connection is assumed between parameters.
- The newer theory produces exact MLEs, but apart from the normal/identity case, inference procedures are still somewhat approximate.

© CSIRO, 2000-2007

Practice

- Constructing GLMs in S-PLUS is almost entirely analogous to constructing LMs
- Estimation is by iteratively weighted least squares, so some care has to be taken that the iterative scheme has converged
- Some tools exist for manual and automated variable selection
- There are differences. e.g. the residuals function distinguishes four types of residual which all coincide in the case of linear models.

© CSIRO, 2000-2007

The budworm data – a toxicological experiment

Dose	Sex	Dead	Alive	Total
1	M	1	19	20
2	M	4	16	20
4	M	9	11	20
8	M	13	7	20
16	M	18	2	20
32	M	20	0	20
1	F	0	20	20
2	F	2	18	20
4	F	6	14	20
8	F	10	10	20
16	F	12	8	20
32	F	16	4	20

© CSIRO, 2000-2007

An initial example: Budworm data

- Preparing the data:

```
options(contrasts = c("contr.treatment", "contr.poly"))

ldose <- rep(0:5, 2)
numdead <- c(1, 4, 9, 13, 18, 20,
             0, 2, 6, 10, 12, 16)
sex <- factor(rep(c("M", "F"), each = 6))
SF <- cbind(numdead, numalive = 20 - numdead)
Budworms <- data.frame(ldose, sex)
Budworms$SF <- SF
rm(sex, ldose, SF)
```

© CSIRO, 2000-2007

Fitting an initial model

```
budworm.lg <- glm(SF ~ sex/ldose, family = binomial, data =
   Budworms, trace = T)
GLM    linear loop 1: deviance = 5.0137
GLM    linear loop 2: deviance = 4.9937
GLM    linear loop 3: deviance = 4.9937
GLM    linear loop 4: deviance = 4.9937

summary(budworm.lg, cor = F)

Coefficients:
                 Value Std. Error    t value
(Intercept) -2.9935418  0.5526997 -5.4162174
        sex  0.1749868  0.7783100  0.2248292
  sexFldose  0.9060364  0.1671016  5.4220677
  sexMldose  1.2589494  0.2120655  5.9366067

(Dispersion Parameter for Binomial family taken to be 1 )
```

© CSIRO, 2000-2007

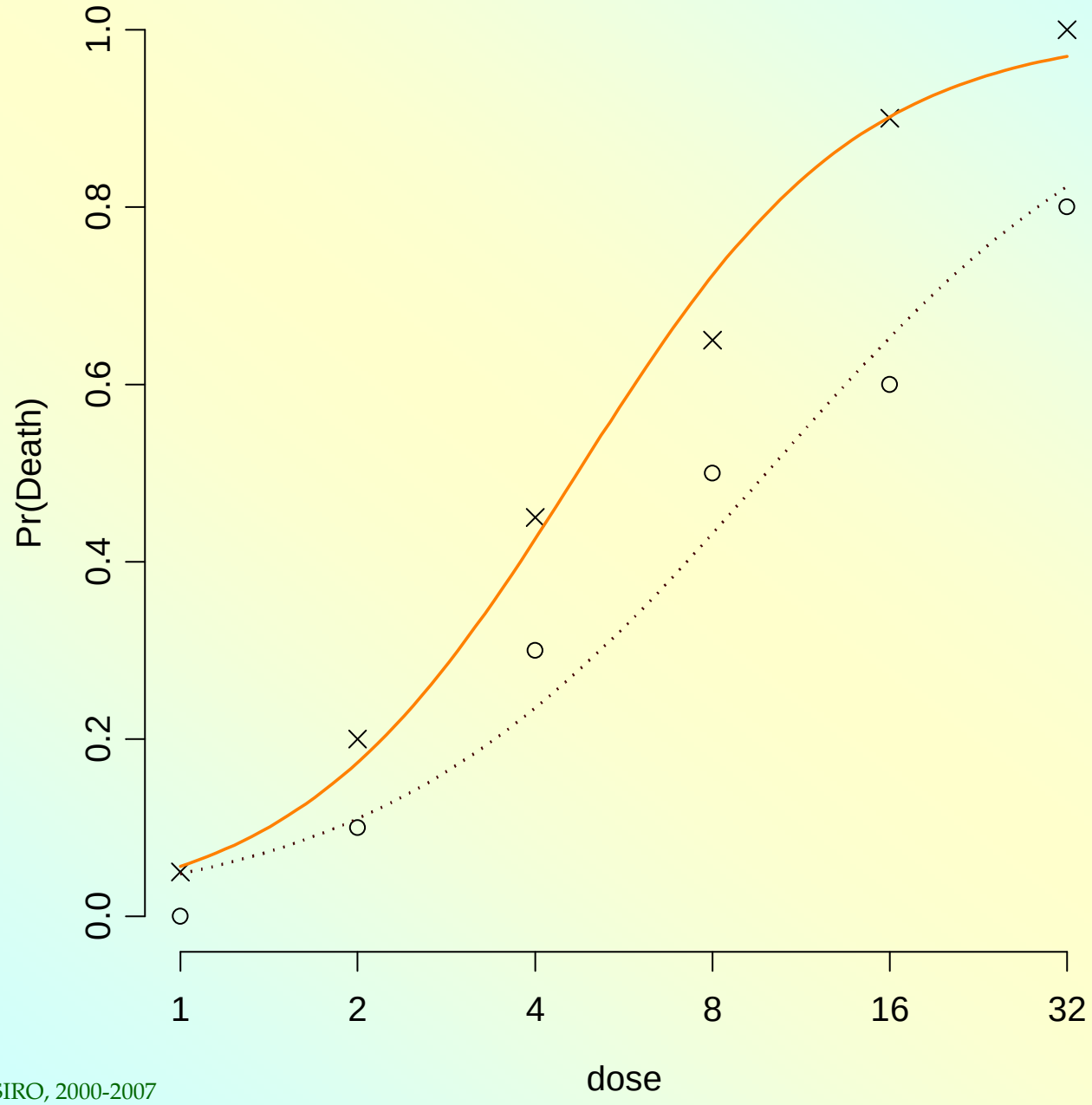
Displaying the fit

```
attach(Budworms)
plot(c(1,32), c(0,1), type = "n", xlab = "dose",
   log = "x", axes = F,ylab = "Pr(Death)")
axis(1, at = 2^(0:5))
axis(2)

points(2^ldose[1:6], numdead[1:6]/20, pch = 4)
points(2^ldose[7:12], numdead[7:12]/20, pch = 1)

ld <- seq(0, 5, length = 100)
lines(2^ld, predict(budworm.lg, data.frame(ldose = ld,
  sex = factor(rep("M", length(ld)), levels = levels(sex))),
  type = "response"), col = 3, lwd = 2)
lines(2^ld, predict(budworm.lg, data.frame(ldose = ld,
  sex = factor(rep("F", length(ld)), levels = levels(sex))),
  type = "response"), lty = 2, col = 2, lwd = 2)
detach()
```

© CSIRO, 2000-2007



© CSIRO, 2000-2007

Is sex significant?

- This is a marginal term and so its meaning has to be interpreted carefully.
- Watch what happens if ldose is re-centred

```
budworm.lgA <- update(budworm.lg, . ~ sex/I(ldose - 3))
GLM    linear loop 1: deviance = 5.0137
GLM    linear loop 2: deviance = 4.9937
GLM    linear loop 3: deviance = 4.9937
GLM    linear loop 4: deviance = 4.9937

summary(budworm.lgA, cor = F)$coefficients
                      Value Std. Error   t value
     (Intercept) -0.2754324  0.2305173 -1.194845
             sex  1.2337258  0.3769761  3.272689
sexFI(ldose - 3)  0.9060364  0.1671016  5.422068
sexMI(ldose - 3)  1.2589494  0.2120655  5.936607
```

© CSIRO, 2000-2007

Checking for curvature

```
anova(update(budworm.lgA, . ~ . + sex/I((ldose - 3)^2)), test =
   "Chisq")
GLM    linear loop 1: deviance = 3.1802
GLM    linear loop 2: deviance = 3.1716
GLM    linear loop 3: deviance = 3.1716
GLM    linear loop 4: deviance = 3.1716
GLM    linear loop 1: deviance = 5.0137
GLM    linear loop 2: deviance = 4.9937
GLM    linear loop 3: deviance = 4.9937
GLM    linear loop 4: deviance = 4.9937
GLM    linear loop 1: deviance = 121.8135
GLM    linear loop 2: deviance = 118.7995
GLM    linear loop 3: deviance = 118.7986
GLM    linear loop 4: deviance = 118.7986
Analysis of Deviance Table

Terms added sequentially (first to last)
                          Df Deviance Resid. Df Resid. Dev   Pr(Chi)
                    NULL                    11   124.8756
                     sex  1   6.0770         10   118.7986 0.0136955
    I(ldose - 3) %in% sex  2 113.8049          8     4.9937 0.0000000
I((ldose - 3)^2) %in% sex  2   1.8221          6     3.1716 0.4021031
```

© CSIRO, 2000-2007

A final model: parallelism

```
budworm.lg0 <- glm(SF ~ sex + ldose - 1, family =
   binomial, Budworms, trace = T)
GLM    linear loop 1: deviance = 6.8074
GLM    linear loop 2: deviance = 6.7571
GLM    linear loop 3: deviance = 6.7571
GLM    linear loop 4: deviance = 6.7571

anova(budworm.lg0, budworm.lgA, test = "Chisq")
Analysis of Deviance Table

                          Terms Resid. Df Resid. Dev
1             sex + ldose - 1          9   6.757064
2 sex + I(ldose - 3) %in% sex          8   4.993727

   Test Df Deviance   Pr(Chi)
1 vs. 2  1 1.763337 0.1842088
```

© CSIRO, 2000-2007

Effective dosages (V&R p193)

```
> dose.p
function(obj, cf = 1:2, p = 0.5)
{
  eta <- family(obj)$link(p)
  b <- coef(obj)[cf]
  x.p <- (eta - b[1])/b[2]
  names(x.p) <- paste("p = ", format(p), ":", sep = "")
  pd <-  - cbind(1, x.p)/b[2]
  SE <- sqrt(((pd %*% vcov(obj)[cf, cf]) * pd) %*% c(1, 1))
  res <- structure(x.p, SE = SE, p = p)
  oldClass(res) <- "glm.dose"
  res
}
> print.glm.dose
function(x, ...)
{
  M <- cbind(x, attr(x, "SE"))
  dimnames(M) <- list(names(x), c("Dose", "SE"))
  x <- M
  NextMethod("print")
}
```

© CSIRO, 2000-2007

Example

```
> dose.p(budworm.lg0, cf = c(1, 3), p = 1:3/4)
                Dose        SE
p = 0.25: 2.231265 0.2499089
p = 0.50: 3.263587 0.2297539
p = 0.75: 4.295910 0.2746874
```

© CSIRO, 2000-2007

A second example: low birth weight

```
options(contrasts = c("contr.treatment", "contr.poly"))
attach(birthwt)
race <- factor(race, labels = c("white", "black", "other"))
table(ptl)
  0  1 2 3
159 24 5 1
ptd <- factor(ptl > 0)
table(ftv)
  0  1  2 3 4 6
100 47 30 7 4 1
ftv <- factor(ftv)
levels(ftv)[ - (1:2)] <- "2+"
table(ftv)
  0  1 2+
100 47 42
bwt <- data.frame(low = factor(low), age, lwt, race, smoke =
  (smoke > 0), ptd, ht = (ht > 0), ui = (ui > 0), ftv)
detach()
rm(race, ptd, ftv)
```

© CSIRO, 2000-2007

Initial Model

```
birthwt.glm <- glm(low ~ ., family = binomial, data = bwt)
dropterm(birthwt.glm, test = "Chisq")

Single term deletions

Model:
low ~ age + lwt + race + smoke + ptd + ht + ui + ftv
       Df Deviance      AIC      LRT   Pr(Chi)
<none>    195.4755 217.4755
   age  1 196.4174 216.4174 0.941876 0.3317959
   lwt  1 200.9495 220.9495 5.473941 0.0193020
  race  2 201.2268 219.2268 5.751273 0.0563802
 smoke  1 198.6738 218.6738 3.198241 0.0737175
   ptd  1 203.5841 223.5841 8.108539 0.0044057
    ht  1 202.9339 222.9339 7.458402 0.0063141
    ui  1 197.5855 217.5855 2.109974 0.1463418
   ftv  2 196.8337 214.8337 1.358185 0.5070769
```

© CSIRO, 2000-2007

Interactions?

- The previous model output suggests removing fitv and age. This is confirmed by successive deletion.
- What happens, though, when we check for interactions between factors and curvatures in the numeric predictors?

© CSIRO, 2000-2007

```
birthwt.step2 <- stepAIC(birthwt.glm,  ~ .^2 +
   I(scale(age)^2) + I(scale(lwt)^2), trace = F)
birthwt.step2$anova

Stepwise Model Path
Analysis of Deviance Table

Initial Model:
low ~ age + lwt + race + smoke + ptd + ht + ui + ftv

Final Model:
low ~ age + lwt + smoke + ptd + ht + ui + ftv + age:ftv
   + smoke:ui


        Step Df Deviance Resid. Df Resid. Dev      AIC
1                                 178   195.4755 217.4755
2  + age:ftv  2 12.47490          176   183.0006 209.0006
3 + smoke:ui  1  3.05681          175   179.9438 207.9438
4     - race  2  3.12959          177   183.0734 207.0734
```

© CSIRO, 2000-2007

Final model: two important interactions

```
dropterm(birthwt.step2, test = "Chisq")
Single term deletions

Model:
low ~ age + lwt + smoke + ptd + ht + ui + ftv +
  age:ftv + smoke:ui
         Df Deviance      AIC      LRT    Pr(Chi)
<none>    183.0734 207.0734
   lwt  1 191.5590 213.5590  8.48561 0.00357967
   ptd  1 193.5880 215.5880 10.51462 0.00118434
    ht  1 191.2108 213.2108  8.13743 0.00433607
age:ftv  2 199.0029 219.0029 15.92950 0.00034750
smoke:ui  1 186.9861 208.9861  3.91270 0.04792245
```

© CSIRO, 2000-2007

Checking for linearity on age within ftv

- An alternative to the method given in MASS
- The idea is to fit separate spline terms within the levels of **`ftv`**, but keeping all other important terms (including interactions).
- It is important that spline terms be chosen with enough knots to allow non-linear behaviour to become apparent, but not so much that the fit becomes nearly indeterminate.

© CSIRO, 2000-2007

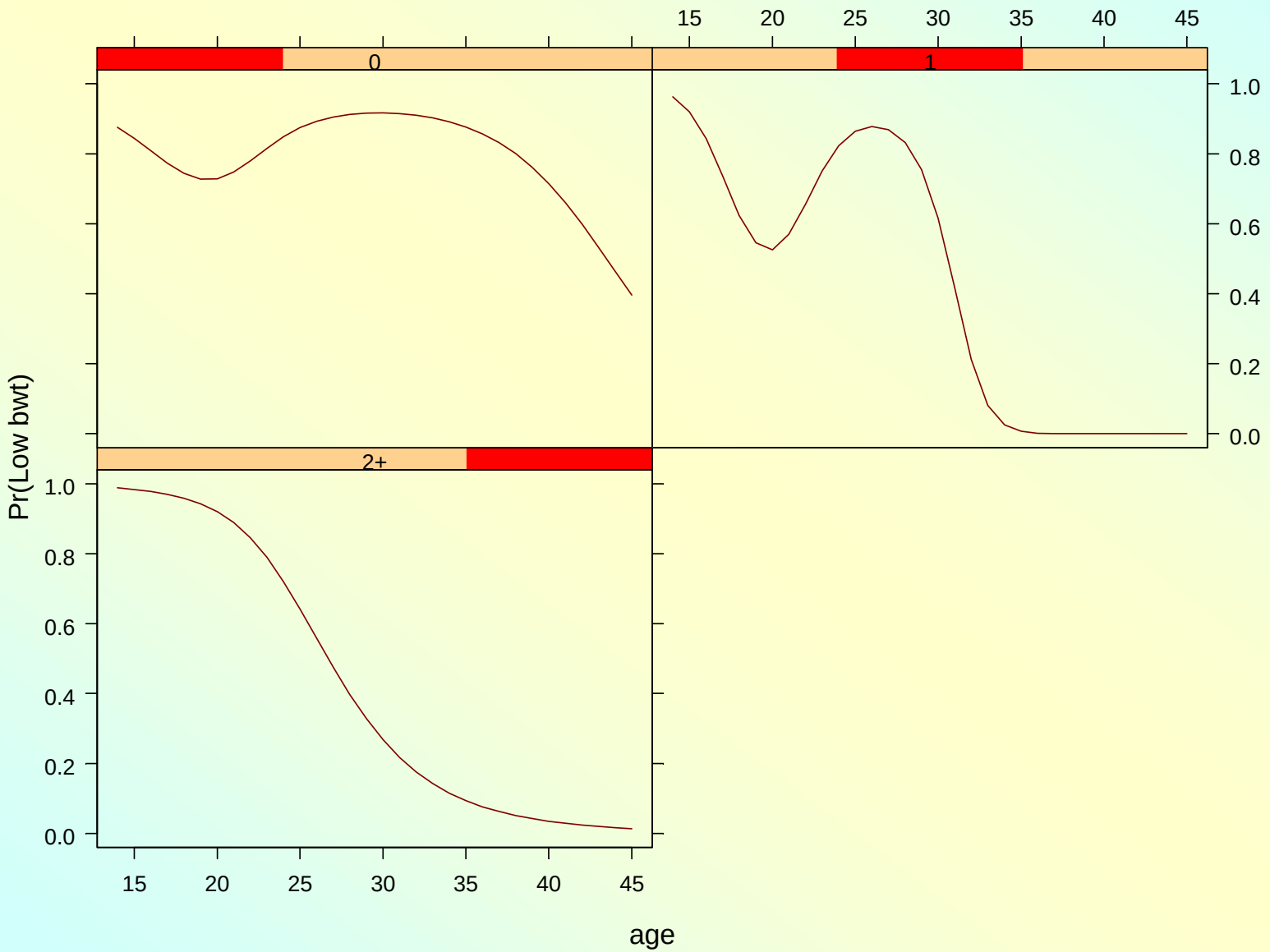
```
attach(bwt)
BWT <- expand.grid(age=14:45, lwt = mean(lwt),
   race = factor("white", levels = levels(race)),
   smoke = c(T,F),
   ptd = c(T,F),
   ht = c(T,F),
   ui = c(T,F),
   ftv = levels(ftv))
detach()

nsAge <- function(x)
   ns(x, knots = quantile(bwt$age, 1:2/3),
   Boundary.knots = range(bwt$age))

birthwt.glm2 <- glm(low ~ lwt + ptd + ht + smoke * ui +
   ftv/nsAge(age), binomial, bwt, trace = F)

prob <- predict(birthwt.glm2, BWT, type = "resp")
xyplot(prob ~ age | ftv, BWT, type = "l",
   subset = smoke == F & ptd == F & ht == F & ui == F,
   as.table = T, ylim = c(0, 1), ylab = "Pr(Low bwt)")
```

© CSIRO, 2000-2007



© CSIRO, 2000-2007