

Statistics 175 Applied Statistics

Generalized Linear Models

Jianqing Fan

Example 1 (Kyphosis data): (The data set kyphosis consists of measurements on 81 children following corrective spinal surgery. Variable collected include $Y = \text{kyphosis}$, $X_1 = \text{age in month}$, $X_2 = \text{Number} = \text{number of vertebrae in the operation}$, $X_3 = \text{start} = \text{begining of the range of vertebrae involved}$.)

```

> attach(kyphosis)
> kyphosis[1:13,]
  Kyphosis Age Number Start
1 absent  71     3     5
2 absent 158     3    14
3 present 128     4     5
4 absent   2     5     1
5 absent   1     4    15
6 absent   1     2    16
7 absent  61     2    17
8 absent  37     3    16
9 absent 113     2    16
10 present 59     6    12
11 present 82     5    14
12 absent 148     3    16
13 absent  18     5     2
> kyph.glm1 <- glm(Kyphosis ~ Age + Start + Number, family = binomial,
+ data = kyphosis)
> kyph.glm1
Call:
glm(formula = Kyphosis ~ Age + Start + Number, family = binomial, data =
  kyphosis)

Coefficients:
(Intercept)      Age      Start      Number
-2.036932  0.01093048 -0.20651  0.410601

Degrees of Freedom: 81 Total; 77 Residual
Residual Deviance: 61.37993
> summary(kyph.glm1)

Call: glm(formula = Kyphosis ~ Age + Start + Number, family = binomial, data =
  kyphosis)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.312363 -0.5484308 -0.3631876 -0.1658653  2.16133

Coefficients:
              Value Std. Error  t value
(Intercept) -2.03693225  1.44918287 -1.405573
      Age     0.01093048  0.00644419  1.696175
      Start -0.20651000  0.06768504 -3.051043
      Number  0.41060098  0.22478659  1.826626

(Dispersion Parameter for Binomial family taken to be 1 )

```

```

Null Deviance: 83.23447 on 80 degrees of freedom

Residual Deviance: 61.37993 on 77 degrees of freedom

Number of Fisher Scoring Iterations: 5

Correlation of Coefficients:
      (Intercept)      Age      Start
Age -0.4633715
Start -0.3784028 -0.2849547
Number -0.8480574  0.2321004  0.1107516
> anova(kyph.glm1, test="Chi")
Analysis of Deviance Table

Binomial model

Response: Kyphosis

Terms added sequentially (first to last)
      Df Deviance Resid. Df Resid. Dev Pr(Chi)
NULL                                80    83.2345
Age  1    1.3020          79    81.9325 0.253851
Start 1   16.6334         78    65.2991 0.000045
Number 1    3.9191         77    61.3799 0.047739
> resid <- residuals(kyph.glm1, type="deviance")
> resid[1:5]
      1          2          3          4          5
-0.7707923 -0.5111615  1.189314 -1.106713 -0.2461915
> kyph.glm2
Call:
glm(formula = Kyphosis ~ Start + Number + Age, family = binomial)

Coefficients:
(Intercept)      Start      Number      Age
-2.036932 -0.20651  0.410601  0.01093048

Degrees of Freedom: 81 Total; 77 Residual
Residual Deviance: 61.37993
> anova(kyph.glm2, test = "Chi")
Analysis of Deviance Table

Binomial model

Response: Kyphosis

Terms added sequentially (first to last)
      Df Deviance Resid. Df Resid. Dev Pr(Chi)
NULL                                80    83.2345
Start 1   15.1623         79    68.0722 0.0000987
Number 1    3.5357         78    64.5365 0.0600605
Age  1    3.1565          77    61.3799 0.0756233

```

Example 2: (Wave-soldering data) In 1988, an experiment was designed and implemented at one of AT & T's factories to investigate alternatives in the "wave-soldering" procedure for mounting electronic components on printed circuits boards. The response, measured by eye, is a count of the number of visible solder skips for a board soldered under a particular choice of levels for the experimental factors.

```
> attach(solder.balance)
> summary(solder.balance)
```

Opening	Solder	Mask	PadType	Panel	skips
S:240	Thin :360	A1.5:180	L9 : 72	1:240	Min. : 0.000
M:240	Thick:360	A3 :180	W9 : 72	2:240	1st Qu.: 0.000
L:240		B3 :180	L8 : 72	3:240	Median : 2.000
		B6 :180	L7 : 72		Mean : 4.965
			D7 : 72		3rd Qu.: 6.000
			L6 : 72		Max. :48.000
			(Other):288		

```
> paov <- glm(skips~., family=poisson, data=solder.balance)
> summary(paov)
```

Call: glm(formula = skips ~ Opening + Solder + Mask + PadType + Panel, family = poisson, data = solder.balance)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.661532	-1.086761	-0.4406795	0.6114746	3.942936

	Value	Std. Error	t value
(Intercept)	0.735679734	0.029480508	24.954785
Opening.L	-1.338897707	0.037898452	-35.328560
Opening.Q	0.561940334	0.042005350	13.377828
Solder	-0.777627385	0.027309904	-28.474190
Mask1	0.214096793	0.037719290	5.676056
Mask2	0.329383406	0.016528258	19.928501
Mask3	0.330750657	0.008946418	36.970177
PadType1	0.055000439	0.033193067	1.656986
PadType2	0.105788229	0.017332685	6.103395
PadType3	-0.104859727	0.015162522	-6.915718
PadType4	-0.122876529	0.013604968	-9.031740
PadType5	0.013084728	0.008852899	1.478016
PadType6	-0.046620368	0.008837701	-5.275169
PadType7	-0.007583584	0.006976023	-1.087093
PadType8	-0.135502138	0.010597577	-12.786144
PadType9	-0.028288193	0.006563991	-4.309603
Panel1	0.166761164	0.021027817	7.930503
Panel2	0.029213741	0.011743659	2.487618

Number of Fisher Scoring Iterations: 4

Correlation of Coefficients:

	(Intercept)	Opening.L	Opening.Q	Solder	Mask1	Mask2
Opening.L	0.4472341					
.....						

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

```

+ "contr.treatment","contr.treatment"))
# The effect remains for duration of the session. You could also change
# part of them to be "contr.sum", "contr.poly", and your own contrasts.
> paov1 <- glm(skips~., family=poisson, data=solder.balance)
> summary(paov1)

Call: glm(formula = skips ~ Opening + Solder + Mask + PadType + Panel, family =
      poisson, data = solder.balance)
Deviance Residuals:
      Min       1Q   Median       3Q      Max
-3.661532 -1.086761 -0.4406795  0.6114746  3.942936

Coefficients:
              Value Std. Error    t value
(Intercept)  1.66335154 0.08019714  20.7407847
  OpeningM   -1.63497719 0.04800876 -34.0558131
  OpeningL   -1.89348730 0.05359650 -35.3285595
    Solder   -1.09973119 0.03862204 -28.4741896
   MaskA3    0.42819359 0.07543858   5.6760557
   MaskB3    1.20224701 0.06694657  17.9583069
   MaskB6    1.86648283 0.06308027  29.5890108
 PadTypeD4   0.11000088 0.06638613   1.6569858
 PadTypeL4   0.37236513 0.06266263   5.9423793
 PadTypeD6  -0.25865024 0.07304340  -3.5410486
 PadTypeL6  -0.55845370 0.07992636  -6.9871030
 PadTypeD7   0.01156078 0.06799047   0.1700353
 PadTypeL7  -0.38020544 0.07566018  -5.0251723
 PadTypeL8  -0.16115190 0.07110446  -2.2664106
 PadTypeW9  -1.32758606 0.10528746 -12.6091565
 PadTypeL9  -0.52645088 0.07911660  -6.6541140
   Panel2    0.33352233 0.04205563   7.9305028
   Panel3    0.25440239 0.04277219   5.9478453

(Dispersion Parameter for Poisson family taken to be 1 )

Null Deviance: 6855.69 on 719 degrees of freedom

Residual Deviance: 1130.48 on 702 degrees of freedom

Number of Fisher Scoring Iterations: 4

Correlation of Coefficients:
.....

> anova(paov1, test = "Chi")

Analysis of Deviance Table

Poisson model

Response: skips

Terms added sequentially (first to last)

```

	Df	Deviance	Resid. Df	Resid. Dev	Pr(Chi)
NULL			719	6855.690	
Opening	2	2524.563	717	4331.128	0.000000e+00
Solder	1	936.955	716	3394.173	0.000000e+00
Mask	3	1653.092	713	1741.080	0.000000e+00
PadType	9	542.463	704	1198.617	0.000000e+00
Panel	2	68.137	702	1130.480	1.554312e-15

Example 3 (Ship damage data): Try the following commands:

```

ship.dat <- read.table("ship.dat",header=T)
attach(ship.dat)
options(contrasts=c("contr.treatment", "contr.treatment", "contr.treatment"))

ship.glm <- glm(Damage ~log(Survice + 1) + Type+Year+Period, family=poisson)
summary(ship.glm)

```