Chap. 5 Building Logistic Regression Models

- Model selection
- Model checking
- Be careful with "sparse" categorical data(infinite estimates possible)

Model Selection with Many Predictors

ex) Horseshoe crab study

$Y=$ whether female crab has satellites (1=yes, 0=no)

Explanatory variable - weight, width, color (ML, M, MD, D) (dummy variables $c_1,\ c_2,\ c_3$), Spine condition (3 categories) (dummy var. $s_1,\ s_2$).

Consider model for Crabs :

$$\mathrm{log}itP(Y=1)=\alpha+\beta_1c_1+\beta_2c_2+\beta_3c_3+\beta_4s_1+\beta_5s_2+\beta_6(weight)+\beta_7(width)$$

LR test of $H_0:\beta_1=\beta_2=\cdots=\beta_7=0$ has test stat.

$$\begin{aligned}-2(L_0-L_1)&=difference\ \ of\ \ Deviance\\&=225.8-185.2=40.6,\ \ df=7\ \ (p-vale<0.0001)\end{aligned}$$

Strong evidence at least one predictor has an effect. But, look at Wald test of individual effects(eg. weight). None of individual effects seem "significant"

Multicollinearity (Strong correlations among predictors) also plays havoc with GLMs

eg) corr(weight, width)=0.89

Individual $Z-$ test of $\frac{\hat{\beta}_i}{SE}$ affected by multicollinearity sufficient to use one of weight/width in model(we use width)

Backward elimination using color, width, spine condition as predictors

- Start with complex model, including all interactions
- Drop "least significant" (eg, largest $p-$ value) variable among highest-order terms
- Refit model
- Continue until all variables left are significant.

Note: If testing many interactions, simpler and perhaps better to test at one time as a group of terms. (like the previous example)

ex) see Table 5.2 (p141 in 2nd edition) or Table 5.1 (p127 in 3rd edition)

Model	Predictors	Deviance G^2	df	AIC	Models Compared	Deviance Difference	Corr. $r(y,\hat{\mu})$
1	$(C*S*W)$	170.44	152	212.4	-		
2	$(C*S+C*W+S*W)$	173.68	155	209.7	(2)-(1)	3.2 (df=3)	
3a	$(C*S+S*W)$	177.34	158	207.3	(3a)-(2)	3.7 (df=3)	
3b	$(C*W+S*W)$	181.56	161	205.6	(3b)-(2)	7.9 (df=6)	
3c	$(C*S+C*W)$	173.69	157	205.7	(3c)-(2)	0.0 (df=2)	
4a	$(S+C*W)$	181.64	163	201.6	(4a)-(3c)	8.0 (df=6)	
4b	$(W+C*S)$	177.61	160	203.6	(4b)-(3c)	3.9 (df=3)	
5	$(C+S+W)$	186.61	166	200.6	(5)-(4b)	9.0 (df=6)	
6a	$(C+S)$	208.83	167	220.8	(6a)-(5)	22.2 (df=1)	
6b	$(S+W)$	194.42	169	202.4	(6b)-(5)	7.8 (df=3)	
6c	$(C+W)$	187.46	168	197.5	(6c)-(5)	0.8 (df=2)	0.452
7a	(C)	212.06	169	220.1	(7a)-(6c)	24.5 (df=1)	0.285
7b	(W)	194.45	171	198.5	(7b)-(6c)	7.0 (df=3)	0.402
8	$(C=dark+W)$	187.96	170	194.0	(8)-(6c)	0.5 (df=2)	0.447
9	None	225.76	172	227.8	(9)-(8)	37.8 (df=2)	0.000

Table C=color, S=spine condition, W=width

H_0 : Model $C+S+W$ has 3 parameters for C, 2 parameters for S, and 1 parameter W.

H_a : Model $C*S*W=C+S+W+C\times S+C\times W+S\times W+C\times S\times W$.

$$LR\ stat. = difference\ in\ Deviances$$
$$= 186.6 - 170.4 = 16.2$$
$$df = 166 - 152 = 14\ (p - value = 0.30)$$

Simpler model $C + S + W$ is adequate.

At next stage, S can be dropped from model $C + S + W$

$$diff.\ in\ deviance = 187.5 - 186.6 = 0.9,\ df = 2$$

Results in model fit

$$\underset{(\alpha)}{\log it\hat{\pi}} = \underset{(\alpha)}{-12.7} + \underset{(\beta_1)}{1.3c_1} + \underset{(\beta_2)}{1.4c_2} + \underset{(\beta_3)}{1.1c_3} + \underset{(\beta_4)}{0.47(width)}$$

Setting $\beta_1 = \beta_2 = \beta_3$ gives

$$\log it\hat{\pi} = -13.0 + 1.3C + 0.48(width)$$

where $C = \begin{cases} 1 & ML, M, MD \\ 0 & D \end{cases}$

Conclude

- Given width, estimated odds of satellite for nondark crabs equal $e^{1.3} = 3.7$ times estimated odds for dark crabs.

  95% C.I. : $e^{1.3 \pm 1.96(0.525)} = (1.3,\ 10.3)$
- Given color, estimated odds of satellite multiplied by $e^{0.48 \pm 1.96(0.104)} = (1.3,\ 2.0)$ for each 1cm increase in width.

Criterion for selecting a model

- Use theory, other research as guide.
- Parsimony (simplicity) is good
- Can use some criterion to choose among set of models. Most popular criterion is Akaike information criterion (AIC):
  Choose model with minimum AIC

$$AIC = -2(L - no.\ model\ parameters)$$

where $L = \log\ likelihood$

- For explanatory purpose, can use automated procedure such as backward elimination.
- Ideally should have ≥ 10 outcomes of each type per predictor.

ex) $n = 1000$, $(Y = 1)$ 30 times, $(Y = 0)$ 970 times

Model should contain ≤ 3 predictors

$n = 173$ horseshoe crabs $(Y = 1) : 111$ crabs, $(Y = 0) : 62$ crabs.

use ≤ 6 predictors

Note:

- Some software (eg. PROC LOGISTIC in SAS) has options for stepwise selection procedures.
- Can further check fit with residuals for grouped data, influence measures, cross validation
- To summarize predictive power, can use correlation(Y, $\hat{\pi}$)

Predictors	Correlation
color	0.280
width	0.400
color+width	0.452
color=dark+width	0.447

input program crab(genmode and logistic)

```
data crab;
 infile 'C:Wcrabs_SAS.dat';
 input color spine width satell weight;
 if satell>0 then y=1 if satell=0 then y=0 n=1;
 weight=weight/1000 color=color-1;
 if color=4 then dark=0 if color<4 then dark=1;
run;
/* Model with color and width */
proc genmod;
 class color;
 model y/n=color width / dist=bin link=logit;
run;
/* Model with dark and width */
proc genmod;
 model y/n=dark width / dist=bin link=logit;
run;
/* Model with color and width using logistic regression */
proc logistic order=data;
 class color (ref='4')/param=ref;
 model y/n=color width;
run;
proc logistic order=data;
 class color (ref='4')/param=ref;
 model y/n=color width/selection=backward;
run;
```

Another summary : Classification table

predict $\hat{Y}=1$ if $\hat{\pi}>0.50$ and $\hat{Y}=0$ if $\hat{\pi}<0.50$

		Prediction		Total
		$\hat{Y}=1$	$\hat{Y}=0$	
Actual	$Y=1$	94	17	111
	$Y=0$	34	28	62

SAS: Get with CTABLE option in PROC LOGISTIC for various "cutpoints"

Predictive Power

- classification table

Predict $\begin{cases} \hat{Y}=1 \text{ if } \hat{\pi}>\pi_0 \\ =0 \text{ if } \hat{\pi}\le\pi_0 \end{cases}$ for some cutoff π_0 (for example, $\pi_0=0.5$)

$$Sensitivity = P(\hat{Y}=1 \mid Y=1) = \frac{94}{94+17} = 0.85$$

$$Specificity = P(\hat{Y}=0 \mid Y=0) = \frac{28}{34+28} = 0.45$$

ex) Horseshoe crab (model using width and color)

Actual	Prediction, $\pi_0=0.64$		Prediction, $\pi_0=0.50$		Total
	$\hat{Y}=1$	$\hat{Y}=0$	$\hat{Y}=1$	$\hat{Y}=0$	
$Y=1$	74	37	94	17	111
$Y=0$	20	42	34	28	62

Predictive Power

For $\pi_0=0.64$

$$Sensitivity = P(\hat{Y}=1 \mid Y=1) = \frac{74}{74+37} = 0.667$$

$$Specificity = P(\hat{Y}=0 \mid Y=0) = \frac{42}{20+42} = 0.677$$

For $\pi_0=0.50$

$$Sensitivity = P(\hat{Y}=1 \mid Y=1) = \frac{94}{94+17} = 0.85$$

$$Specificity = P(\hat{Y}=0 \mid Y=0) = \frac{28}{34+28} = 0.45$$

ROC (receiver operating characteristic) Curve

- Plot of sensitivity as a function of (1-specificity) for the possible cutoffs π_0
- Summarize predictive power for all possible π_0. When π_0 gets near 0, almost all predictions are $\hat{y}=1$ (sensitivity is near 1 and specificity is near 0)
- For a given specificity, better predictive power correspond to higher sensitivity (The better the predictive power, the higher the ROC curve)
- The area under the ROC curve - the value of a measure of predictive power (concordance index). concordance index c estimates the probability that the

predictions and the outcomes are concordant.(the observation with the large y also has the larger $\hat{\pi}$)

input program crab(ROC)

```
data crab;
 infile 'C:WWcrabs_SAS.dat';
 input color spine width satell weight;
 if satell>0 then y=1; if satell=0 then y=0; n=1;
 weight=weight/1000;
 color=color-1;
 if color=4 then dark=0; if color<4 then dark=1;
run;

proc logistic data=crab order=data;
 class color (ref='4')/param=ref;
 model y=width color / outroc=ROCData;
run;

symbol v=dot i=join;
proc gplot data=ROCData;
 plot _sensit_*_1mspec_;
run;
quit;
```

The LOGISTIC Procedure
Model Information

Data Set	WORK.CRAB
Response Variable	y
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	173
Number of Observations Used	173

Response Profile

Ordered Value	y	Total Frequency
1	1	111
2	0	62

Probability modeled is y=1.

Class Level Information

Class	Value	Design Variables		
color	1	1	0	0
	2	0	1	0
	3	0	0	1
	4	0	0	0

Model Convergence Status
Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	227.759	197.457
SC	230.912	213.223
-2 Log L	225.759	187.457

The LOGISTIC Procedure
Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	38.3015	4	<.0001
Score	34.3384	4	<.0001
Wald	27.6788	4	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
width	1	19.6573	<.0001
color	3	6.6246	0.0849

Analysis of Maximum Likelihood Estimates

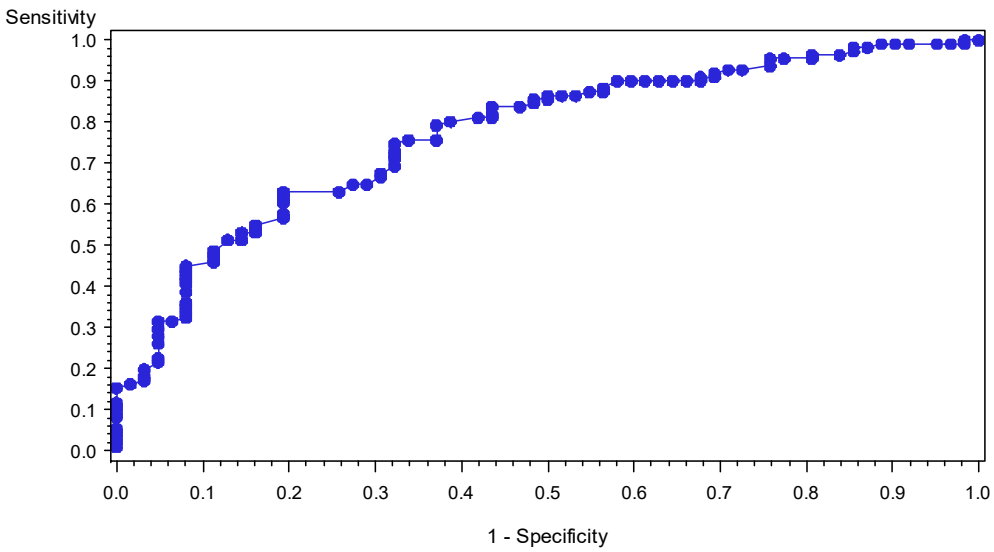
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-12.7151	2.7618	21.1965	<.0001
width		1	0.4680	0.1055	19.6573	<.0001
color	1	1	1.3299	0.8525	2.4335	0.1188
color	2	1	1.4023	0.5484	6.5380	0.0106
color	3	1	1.1061	0.5921	3.4901	0.0617

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
width	1.597	1.298	1.964
color 1 vs 4	3.781	0.711	20.102
color 2 vs 4	4.065	1.387	11.909
color 3 vs 4	3.023	0.947	9.646

Association of Predicted Probabilities and Observed Responses

Percent Concordant	76.9	Somers' D	0.543
Percent Discordant	22.6	Gamma	0.546
Percent Tied	0.5	Tau-a	0.251
Pairs	6882	c	0.771



Model Checking

Is the chosen model adequate?

● Goodness of fit test

But test using deviance G^2, X^2 limited to "non-sparse" contingency tables.

● Check whether fit improves by adding other predictors, interactions between predictors.

$LR\ stat.$ = change in deviance is useful for comparing models even when G^2 not valid as overall test of fit.

ex) Revisit Florida death penalty data

Victim's Race	Suspect's Race	Death Penalty(Y) Yes	No	n
Black	Black	4	139	143
	White	0	16	16
White	Black	11	37	48
	White	53	44	467

$\pi = P(Y = Yes)$

```
data death;
 input v $ d $ p total @@;
cards;
b b 4 143  b w  0  16
w b 11 48  w w 53 467
run;

❶ proc genmod data=death;
 class v d;
 model p/total = d v / dist=bin link=logit lrci type3;
run;

❷ proc genmod data=death;
 class v d;
 model p/total = v / dist=bin link=logit lrci;
run;
```

OUTPUT:

❶

```
The GENMOD Procedure
     Model Information

Data Set                      WORK.DEATH
Distribution                    Binomial
Link Function                      Logit
Response Variable (Events)             p
Response Variable (Trials)         total

Number of Observations Read            4
Number of Observations Used            4
Number of Events                      68
Number of Trials                     674

       Class Level Information
Class      Levels    Values
v               2    b w
d               2    b w
```

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1	0.3798	0.3798
Scaled Deviance	1	0.3798	0.3798
Pearson Chi-Square	1	0.1978	0.1978
Scaled Pearson X2	1	0.1978	0.1978
Log Likelihood		-209.4783	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	-2.0595	0.1458	-2.3565	-1.7836	199.40	<.0001
d	b	1	0.8678	0.3671	0.1140	1.5633	5.59	0.0181
d	w	0	0.0000	0.0000	0.0000	0.0000	.	.
v	b	1	-2.4044	0.6006	-3.7175	-1.3068	16.03	<.0001
v	w	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

The GENMOD Procedure
LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
d	1	5.01	0.0251
v	1	20.35	<.0001

❷

The GENMOD Procedure
Model Information

Data Set	WORK.DEATH
Distribution	Binomial
Link Function	Logit
Response Variable (Events)	p
Response Variable (Trials)	total
Number of Observations Read	4
Number of Observations Used	4
Number of Events	68
Number of Trials	674

Class Level Information

Class	Levels	Values
v	2	b w
d	2	b w

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	2	5.3940	2.6970
Scaled Deviance	2	5.3940	2.6970
Pearson Chi-Square	2	5.8109	2.9054
Scaled Pearson X2	2	5.8109	2.9054
Log Likelihood		-211.9854	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter		DF	Estimate	Standard Error	Likelihood Ratio 95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	-1.9526	0.1336	-2.2234	-1.6989	213.68	<.0001
v	b	1	-1.7045	0.5237	-2.9072	-0.7995	10.59	0.0011
v	w	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Goodness of fit

For model fit with $d=1$ $(Black\ suspect)$ or $d=0$ $(White)$ and $\nu=1$ $(Black\ victim)$ or $\nu=0$ $(White)$

$$logit\,\hat{\pi}=-2.06+0.87d-2.40\nu$$

$$\hat{\pi}=\frac{e^{-2.06+0.87d-2.40\nu}}{1+e^{-2.06+0.87d-2.40\nu}}$$

eg) for 467 cases with white suspect, victim, $d=\nu=0$

$$\hat{\pi}=\frac{e^{-2.06}}{1+e^{-2.06}}=0.113$$

Fitted count "Yes"=467 (0.113)=5.28 (53: observed "Yes")
Fitted count "No"=467 (0.887)=414.2 (414: observed "No")

Summarizing fit over 8 cells,

$$X^2=\sum\frac{(obs-fit)^2}{fit}=0.20$$

$$G^2=2\sum obs.\log\left(\frac{obs}{fit}\right)=0.38=deviance\ \text{for model}$$

$df=4-3$ (no. binomial observation - no. model parameters)

For $G^2=0.38$, $p-value=0.54$ for H_0 : model holds (No evidence of lack of fit)

Note:

- Model assumes lack of interaction between d and ν in effects on Y. So goodness of fit test in this example is a test of H_0 : no interaction.
- For continuous predictors or many predictors with small $\hat{\mu}_i$, G^2 and X^2 are not approx. χ^2. For better approximation, can group data befor applying X^2, G^2. Hosmer-Lemershow test groups using ranges of $\hat{\pi}$ values(available in PROC LOGISTIC). Or, can try to group predictor value(if only 1 or 2 predictors)

Residuals for Logistic Regression

At setting i or explanatory variables, let

y_i = no. successes

n_i = no. trials (preferably 'large')

$\hat{\pi}$ = estimated prob. of success, besed on ML model fit.

For a binomial GLM, Pearson residuals are

$$e_i = \frac{y_i - n_i\hat{\pi}_i}{\sqrt{n_i\hat{\pi}_i(1-\hat{\pi}_i)}} \quad (X^2 = \sum_i e_i^2)$$

e_i (called Reschi in SAS GENMOD) is approx. $N(0,\ \nu)$, when model holds, but $\nu < 1$

Standardized Pearson residual (adjusted residual in some books, SPSS)

- $$r_i = \frac{y_i - n_i\hat{\pi}_i}{SE} = \frac{y_i - n_i\hat{\pi}_i}{\sqrt{n_i\hat{\pi}_i(1-\hat{\pi}_i)(1-h_i)}}$$

  where h_i is called "leverage" (r_i labelled stReschi in SAS)

  $\Rightarrow\ 0 < h_i < 1$ High values of h_i (close to 1) correspond to extreme points in the design space
- r_i is approx. $N(0,\ 1)$ when model holds
- $|\ r_i\ | > 2$ or 3 (approx.) suggests lack of fit

ex) $Y =$ admitted into graduate school at Berkely ($1 = Yes,\ 0 = No$)

Data on textbook

$G =$ gender ($g = 0\ female,\ g = 1\ male$)

$D = department\ (A,\ B,\ C,\ D,\ E,\ F)$

$$d_1 = \begin{cases} 1 & dept.\ A \\ 0 & otherwise \end{cases},\ \cdots,\ d_5 = \begin{cases} 1 & dept.\ E \\ 0 & otherwise \end{cases}$$

For dept. F, $d_1 = d_2 = \cdots = d_5 = 0$

- Model

$$logit P(Y=1) = \alpha + \beta_1 d_1 + \cdots + \beta_5 d_5 + \beta_6 g \qquad (1)$$

  Seems to fit poorly (deviance $G^2 = 20.2,\ df = 5,\ p - value = 0.01$)
- Simpler models fit poorly. eg, model with $\beta_6 = 0$ assumes Y indep. of G controlling for D, has

$$G^2 = 21.7,\ df = 6,\ p - value = 0.001$$

  It seems to fit poorly
- Standardized Pearson residuals for model (1) show model fits well except in Dept. A ($|\ r_i\ | = r_2 = 4.03$), where fewer males accepted than model predicts.

In oter dept.s, model with no gender effect is adequate.

Note:

● Simpson's paradox holds

$Marginal\ \hat{\theta} = \frac{1198 \times 1278}{1493 \times 557} = 1.84$ between Y and g

● Controlling for dept.,

$\hat{\theta}_{Yg(dept)} = e^{-0.0999} = 0.90$ for logit model

● In dept. A, $\hat{\theta} = 0.35$ (odds of admission lower for males)

● If delete dept. A, $G^2 = 2.6,\ df = 4$ (Good fit).

```
options ls=100 ps=70
data berkeley;
 input dept $ gender yes no @@;
   n=yes+no;
 cards
 a 0 512 313  a 1  89  19
 b 0 353 207  b 1  17   8
 c 0 120 205  c 1 202 391
 d 0 138 279  d 1 131 244
 e 0  53 138  e 1  94 299
 f 0  22 351  f 1  24 317
 ;
run;

proc genmod data=berkeley;
  class dept gender;
  model yes/n=gender dept / dist=bin link=logit residuals type3;
run;
proc genmod data=berkeley;
  class dept gender;
  model yes/n=dept / dist=bin link=logit residuals;
run;
```

Enter data as 12 binomials
(1=female, 0=male)

The GENMOD Procedure

Model Information

Data Set	WORK.BERKELEY
Distribution	Binomial
Link Function	Logit
Response Variable (Events)	yes
Response Variable (Trials)	n
Number of Observations Read	12
Number of Observations Used	12
Number of Events	1755
Number of Trials	4526

Class Level Information

Class	Levels	Values
dept	6	a b c d e f
gender	2	0 1

Response Profile

Ordered Value	Binary Outcome	Total Frequency
1	Event	1755
2	Nonevent	2771

Parameter Information

Parameter	Effect	dept	gender
Prm1	Intercept		
Prm2	gender		0
Prm3	gender		1
Prm4	dept	a	
Prm5	dept	b	
Prm6	dept	c	
Prm7	dept	d	
Prm8	dept	e	
Prm9	dept	f	

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	5	20.2043	4.0409
Scaled Deviance	5	20.2043	4.0409
Pearson Chi-Square	5	18.8243	3.7649
Scaled Pearson X2	5	18.8243	3.7649
Log Likelihood		-2593.7442	
Full Log Likelihood		-44.5720	
AIC (smaller is better)		103.1440	

→ $p-value = 0.001$

AICC (smaller is better)	131.1440
BIC (smaller is better)	106.5383

Algorithm converged.

lower admissions rate males

The GENMOD Procedure
Analysis Of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-2.6246	0.1577	-2.9337	-2.3154	276.88	<.0001
gender	0	1	-0.0999	0.0808	-0.2583	0.0586	1.53	0.2167
gender	1	0	0.0000	0.0000	0.0000	0.0000	.	.
dept	a	1	3.3065	0.1700	2.9733	3.6396	378.38	<.0001
dept	b	1	3.2631	0.1788	2.9127	3.6135	333.12	<.0001
dept	c	1	2.0439	0.1679	1.7149	2.3729	148.24	<.0001
dept	d	1	2.0119	0.1699	1.6788	2.3449	140.18	<.0001
dept	e	1	1.5672	0.1804	1.2135	1.9208	75.44	<.0001
dept	f	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

LR Statistics For Type 3 Analysis

Source	DF	Chi-Square	Pr > ChiSq
gender	1	1.53	0.2159
dept	5	763.40	<.0001

Observation Statistics

Observation	Raw Residual	Pearson Residual	Deviance Residual	Std Deviance Residual	Std Pearson Residual	Likelihood Residual
1	-17.26992	-1.253808	-1.248674	-4.010799	-4.027288	-4.025693
2	17.269919	3.5186675	3.7189203	4.2564872	4.027288	4.2033791
3	-0.639509	-0.056021	-0.056009	-0.279662	-0.279722	-0.27972
4	0.6395092	0.2689516	0.270608	0.281445	0.2797222	0.2813152
5	10.754724	1.2628723	1.2533375	1.8666311	1.8808316	1.8744428
6	-10.75472	-0.920778	-0.92434	-1.888107	-1.880832	-1.882578
7	0.7926099	0.0826077	0.0825674	0.1411928	0.1412619	0.1412383
8	-0.79261	-0.085732	-0.085771	-0.141327	-0.141262	-0.141286
9	7.3191902	1.2415132	1.2205137	1.6058628	1.6334923	1.6175896
10	-7.31919	-0.844033	-0.850933	-1.646846	-1.633492	-1.637068
11	-0.957096	-0.206201	-0.207564	-0.304644	-0.302644	-0.303574
12	0.9570957	0.2064808	0.2051779	0.3007342	0.3026438	0.3017565

The GENMOD Procedure

Model Information

Data Set	WORK.BERKELEY
Distribution	Binomial
Link Function	Logit
Response Variable (Events)	yes
Response Variable (Trials)	n

Number of Observations Read	12
Number of Observations Used	12
Number of Events	1755
Number of Trials	4526

Class Level Information

Class	Levels	Values
dept	6	a b c d e f
gender	2	0 1

Response Profile

Ordered Value	Binary Outcome	Total Frequency
1	Event	1755
2	Nonevent	2771

Parameter Information

Parameter	Effect	dept
Prm1	Intercept	
Prm2	dept	a
Prm3	dept	b
Prm4	dept	c
Prm5	dept	d
Prm6	dept	e
Prm7	dept	f

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	6	21.7355	3.6226
Scaled Deviance	6	21.7355	3.6226
Pearson Chi-Square	6	19.9384	3.3231
Scaled Pearson X2	6	19.9384	3.3231
Log Likelihood		-2594.5099	
Full Log Likelihood		-45.3376	
AIC (smaller is better)		102.6752	
AICC (smaller is better)		119.4752	
BIC (smaller is better)		105.5846	

Algorithm converged.

The GENMOD Procedure

Analysis Of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-2.6756	0.1524	-2.9744	-2.3769	308.10	<.0001
dept	a	1	3.2691	0.1671	2.9417	3.5966	382.88	<.0001
dept	b	1	3.2185	0.1749	2.8757	3.5613	338.63	<.0001
dept	c	1	2.0600	0.1674	1.7319	2.3880	151.45	<.0001
dept	d	1	2.0108	0.1699	1.6778	2.3438	140.07	<.0001
dept	e	1	1.5861	0.1798	1.2337	1.9385	77.82	<.0001
dept	f	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

Observation Statistics

Observation	Raw Residual	Pearson Residual	Deviance Residual	Std Deviance Residual	Std Pearson Residual	Likelihood Residual
1	-19.43087	-1.412995	-1.40638	-4.133629	-4.153073	-4.150827
2	19.430868	3.9053116	4.1323246	4.3944879	4.1530728	4.3672258
3	-1.188034	-0.104129	-0.104087	-0.503506	-0.503708	-0.503699
4	1.1880342	0.4928272	0.497804	0.5087944	0.5037077	0.5085781
5	6.0021786	0.6976841	0.6949825	0.8647048	0.8680662	0.8658963
6	-6.002179	-0.516503	-0.517671	-0.870029	-0.868066	-0.868762
7	-3.632576	-0.375617	-0.376411	-0.547028	-0.545873	-0.54642
8	3.6325758	0.3960931	0.3951975	0.5446391	0.5458732	0.5452238
9	4.9229452	0.8207706	0.8119543	0.989787	1.0005342	0.9933147
10	-4.922945	-0.572193	-0.575427	-1.00619	-1.000534	-1.002388
11	-2.030812	-0.428298	-0.434127	-0.628187	-0.619753	-0.623795
12	2.0308123	0.4479439	0.4417581	0.6111942	0.6197526	0.6152965

Sparse Data

Caution: Parameter estimates in logistic regression can be infinite.

ex1)

	S	F
1	8	2
0	10	0

Model

$$\log\left(\frac{P(S)}{P(F)}\right)=\alpha+\beta x$$

$e^{\hat{\beta}}=odds\ \ ratio=\frac{8\times 0}{10\times 2}=0$

$\hat{\beta}=\log\ \ odds\ \ ratio=-\infty$

ex2) Multi-center clinical trial (5centers, each with 2×2 table)

ex3) $y=\begin{cases}1 & \text{for } x<50\\ 0 & \text{for } x>50\end{cases}$

$$\log itP(Y=1)=\alpha+\beta x$$

has $\hat{\beta}=\infty$. software may not realize this!

❶ PROC GENMOD : $\hat{\beta}=3.84,\ SE=15601054$

❷ PROC LOGISTIC gives warning

❸ SPSS : $\hat{\beta}=1.83,\ SE=674.8$

Infinite estimates exists when we can separate x values where $y=1$ from $x-$ values where $y=0$.(perfect discrimination)

Inference about Conditional Associations in $2\times2\times K$ tables(X, Y, Z)

Consider H_0 : X and Y independent, given Z

1. Using logistic regression

$logit[P(Y=1 \mid X=i, Z=k)] = \alpha + \beta x_i + \beta_k^Z,\ i=1,2,\ k=1,\cdots,K$

with $x_1 = 1,\ x_2 = 0$ (dummy variable), $\{\beta_k^Z\}$ K parameters for effects of Z with constraint such as $\beta_K^Z = 0$ or $\beta_1^Z = 0$

$H_0 : \beta = 0$. Test using Wald or LR test.

2. Cochran-Mental-Haenszel Test (CMH): a non-model-base test

Cochran → binomial case given k (1950)

Mental-Haenszel → Hypergeometric dist. given k (1959)

K 2×2 tables of each

n_{11k}	n_{12k}	n_{1+k}
n_{21k}	n_{22k}	n_{2+k}
n_{+1k}	n_{+2k}	n_{++k}

Under H_0, Conditional on $\{n_{1+k},\ n_{2+k},\ n_{+1k},\ n_{+2k}\}$ for each partial table, n_{11k} has hypergeometric dist.

$$\mu_{11k} = E_{H_0}(n_{11k}) = \frac{n_{1+k}n_{+1k}}{n_{++k}}$$

$$Var(n_{11k}) = \frac{n_{1+k}n_{2+k}n_{+1k}n_{+2k}}{n_{++k}^2(n_{++k}-1)}$$

Mentel and Haenszel(1959) proposed

$$M^2 = \frac{(\sum_k n_{11k} - \sum_k \mu_{11k})^2}{\sum_k Var(n_{11k})}$$

$\sum_k (n_{11k} - \mu_{11k})$

Under H_0, $M^2 \xrightarrow{d} \chi_1^2$

($n_{11k} - \mu_{11k}$ is fluctuated around 0 (+ or -). It is bad)

Testing for conditional independence of X and Y given Z

In fact, this is the score test of $H_0 : \beta = 0$ in logit model

$$logit[P(Y=1)] = \alpha + \beta x_i + \beta_k^Z$$

(Day and Byar, 1979) (tink about suff. stat.)

ex)

Center	Group Treatment	Response S	Response F
1	Treatment	0	5
	Placebo	0	9
2	T	1	12
	F	0	10
3	T	0	7
	F	0	5
4	T	6	3
	F	2	6
5	T	5	9
	F	2	12

0	5	5
0	9	9
0	14	14

$n_{111} = 0$

$\hat{\mu}_{111} = 0$

$\widehat{Var}(n_{111}) = 0$

No information for treatment effect exists in center 1, 3

CMH test statistic → [CMH test does not give any direction(higher or lower). Just give significance. But Wald, LR test give it]

$$M^2 = 5.02,\ df = 1,\ p-value = 0.025$$

Model

$$logit[P(Y=1)] = \alpha + \beta x + \beta_k^Z$$

has $\hat{\beta} = 1.546,\ s.e. = 0.702$ ($e^{1.546} = 4.69$ estimated odds for each partial table)

$Wald\ stat. = 4.85$ for $H_0 : \beta = 0$ ($p-value = 0.028$)

LR stat. comparing this model to model with $\beta = 0$ is 5.49 ($p-value = 0.019$)

```
data trial;
input center $ group $ response $ count @@;
cards;
1 treatment success 0 1 treatment failure 5
1 placebo   success 0 1 placebo   failure 9
2 treatment success 1 2 treatment failure 12
2 placebo   success 0 2 placebo   failure 10
3 treatment success 0 3 treatment failure 7
3 placebo   success 0 3 placebo   failure 5
4 treatment success 6 4 treatment failure 3
4 placebo   success 2 4 placebo   failure 6
5 treatment success 5 5 treatment failure 9
5 placebo   success 2 5 placebo   failure 12
;
run;

proc freq data=trial;
weight count;
table center*group*response/cmh;
run;
```

Summary Statistics for group * response
Controlling for center

Cochran-Mantel-Haenszel Statistics (Based on table Scores)(CMH)

Statistic	Alternative Hypothesis	DF	Value	Prob
1	Nonzero Correlation	1	5.0170	0.0251
2	Row Mean Scores Differ	1	5.0170	0.0251
3	General Association	1	5.0170	0.0251

Estimates of the Common Relative Risk (Row1/Row2)

MH odds ratio estimate

Types of Study	Method	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	Mantel-Haenszel	4.7151	1.1840	18.7768
	Logit **	3.9677	1.0978	14.3395
Cohort (Col1 Risk)	Mantel-Haenszel	1.2143	1.0231	1.4413
	Logit	1.1252	0.9719	1.3026
Cohort (Col2 Risk)	Mantel-Haenszel	0.3575	0.1350	0.9466
	Logit **	0.3890	0.1547	0.9782

** These logit estimators use a correlation of 0.5 in every cell of those tables that contain a zero. Tables with a zero row or a zero column are not included in computing the logit estimators.

Breslow-Day Test for
Homogeneity of the odds Ratios

Chi-Square	0.3611
DF	2
Pr > ChiSq	0.8348

Total Sample Size = 94

```
/* SAS (genmod) showing effects of sampling zeroes(sandoz data) */
data sparse;
input center $ treat success n @@;
cards;
1 1 0 5 1 0 0 9
2 1 1 13 2 0 0 10
3 1 0 7 3 0 0 5
4 1 6 9 4 0 2 8
5 1 5 14 5 0 2 14
;
run;

proc genmod data=sparse;
class center;
model success/n=treat center/dist=bin link=logit;
run;
```

The GENMOD Procedure

Model Information

Data Set	WORK.SPARSE
Distribution	Binomial
Link Function	Logit
Response Variable (Events)	success
Response Variable (Trials)	n
Number of Observations Read	10
Number of Observations Used	10
Number of Events	16
Number of Trials	94

Class Level Information

Class	Levels	Values
center	5	1 2 3 4 5

Response Profile

Ordered Value	Binary Outcome	Total Frequency
1	Event	16
2	Nonevent	78

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	4	0.5021	0.1255
Scaled Deviance	4	0.5021	0.1255
Pearson Chi-Square	4	0.3602	0.0900
Scaled Pearson X2	4	0.3602	0.0900
Log Likelihood		-28.8701	
Full Log Likelihood		-6.4293	
AIC (smaller is better)		24.8587	
AICC (smaller is better)		52.8587	
BIC (smaller is better)		26.6742	

Algorithm converged.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept		1	-2.0223	0.6700	-3.3354	-0.7092	9.11	0.0025
treat		1	1.5460	0.7017	0.1708	2.9212	4.85	0.0276
center	1	1	-25.9997	213410.4	-418303	418250.7	0.00	0.9999
center	2	1	-2.1802	1.1327	-4.4003	0.0399	3.70	0.0543
center	3	1	-25.9070	188688.5	-369849	369796.7	0.00	0.9999
center	4	1	1.0631	0.7011	-0.3110	2.4373	2.30	0.1294
center	5	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

NOTE: The scale parameter was held fixed.

too large in absolute values.

($\because$) center 1 and center 3 : $n_{+1k} = 0,\ k = 1,3$

estimated odds ratio=$e^{1.546} = 4.69$

Note: M^2 is a test sta. for testing conditional independence of X and Y given Z

● Mentel and Haenszel (1959) suggested

$$\hat{\theta}_{MH} = \frac{\sum_k n_{11k} n_{22k} / n_{++k}}{\sum_k n_{12k} n_{21k} / n_{++k}}$$ => the strength of association (simillar to odds ratio)

for estimating an assumed common odds ratio for the K partial tables.

*[It is preferred over the ML estimator when K is large and the data are sparse (The ML estimator $\hat{\beta}$ of the log odds ratio then tends to be too large in absolute value)]

Robins et al. (1986) derived a std. error for $\log\hat{\theta}_{MH}$

ex) $\hat{\theta}_{MH} = 4.72$

95% C.I. (1.18, 18.78) for θ_{MH}

Note:

● For $2\times2\times K$ tables, Breslow and Day (1980) gave large-sample Chi-squared test for equality of K odds ratios ($df = K-1$)

ex) B-D stat.=0.36, df=2 (not 4=$K-1$ [2 centers do not have successes])

or, could use goodness-of-fit test(deviance pearson) of model,

$$logit(\pi) = \alpha + \beta x + \beta_k^Z$$

which is equivalent to residual odds ratios (e^{β} in each stratum), with $df = K-1$

⇒ There is also a small-sample test (Zelon, 1972, in statXact)

● Small-sample exact test of conditional indep. of X and Y given Z(i.e., $H_0 : \beta = 0$) uses dist. of $\sum_k n_{11k}$, conditional on $\{n_{i+k}\}$ and $\{n_{+jk}\}$ (StatXact)

(SAS V8.2 Proc logistic has this test)

● CMH test generalized to test conditional indep. in $I\times J\times K$ table, for ordered and unordered rows and columns (Landis, et al., 1978) see 7.5

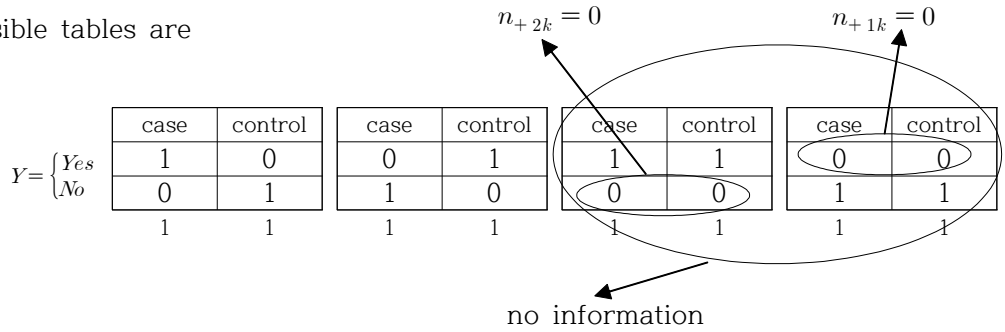
PROC FREQ in SAS

Stat.	X	Y	df
Nonzero corr.	Ordinal	Ordinal	1
Row means diff.	Nominal	Ordinal	$I-1$
General Assoc.	Nominal	Nominal	$(I-1)(J-1)$

ex) K case-control pair (matched pairs with $n_{1+k} = n_{2+k} + 1$ for k)

Pair	Exposure (Y: response)	case (cancer)	Control (no cancer)
	Yes		
1	N0		
		1	1
	Yes		
2	N0		
		1	1
⋮	⋮		
	Yes		
K	N0		
		1	1

Possible tables are



(In general ML theory, as $n \to \infty$, the parameter space is fixed)

LR test of $H_0 : \beta = 0$ has asymptotic bias on $n \to \infty$ with fixed no. parameter

Here $n = 2K$. As $n \uparrow$, so does K and no. parameter.

In fact, $\hat{\beta} \xrightarrow{p} 2\beta$ (Andersen, 1980)

Note: tables

case	control
1	1
0	0
1	1

and

case	control
0	0
1	1
1	1

make no contribution to CMH M^2, but for other tables, $\sum_k n_{11k}$ is a sum of i.i.d, Bernoulli r.v's and $M \rightarrow$ normality as no. of such tables $\uparrow$ ($M^2 \xrightarrow{d} \chi_1^2$)

In fact, $\hat{\beta}_{MH} \xrightarrow{p} \theta(e^{\beta})$

In $2 \times 2 \times K$ tables, Model

$$logitP(Y=1) = \alpha + \beta x + \beta_k^Z$$

Note:

- An alternative inference approach is conditional logistic regression.
  eliminating $\{\beta_k^Z\}$ by conditioning on their sufficient statistics (see 6.7) (SAS V8.2)
  * Frequentist: $k \rightarrow \infty$, ML consistency is broken but conditional MLE given Z is consistent.
  * Bayesian β_k^Z : random effect is considered.

- For multi-center clinical trial data, alternative approach treats center effects as random.
  eg) $\beta_k^Z = u_k \sim N(\mu, \sigma^2)$
  (Agresti & Hartzel, Stat. in Med., 2000)