4. Multiple Logistic Regression

EXAMPLE Los Angeles Heart Study (200 men)

Abbreviation	Variable	Units
Ag	Age:	in years
S	Systolic Blood Pressure:	millimeters of mercury
D	Diastolic Blood Pressure:	millimeters of mercury
Ch	Cholesterol:	milligrams per DL
H	Height:	inches
W	Weight:	pounds
CNT	Coronary incident:	1 if an incident had occurred in the previous ten years; 0 otherwise

Of the 200 cases, 26 had coronary incidents.

$$y_i \sim \text{Bin}(1, p_i)$$

$$\log[p_i/(1-p_i)] = \beta_0 + \beta_1 Ag_i + \beta_2 S_i + \beta_3 D_i + \beta_4 Ch_i + \beta_5 H_i + \beta_6 W_i,$$

$i = 1, \ldots, 200.$

Variable	Estimate	Std. Error	z
Intercept	-4.5173	7.451	-0.61
Ag	0.04590	0.02344	1.96
S	0.00686	0.02013	0.34
D	-0.00694	0.03821	-0.18
Ch	0.00631	0.00362	1.74
H	-0.07400	0.1058	-0.70
W	0.02014	0.00984	2.05

$$G^2 = 134.9, \qquad df = 193$$

G^2 has no basis for comparison to a standard.

Prediction

$$\log[\hat{p}_i/(1-\hat{p}_i)] = \hat{\beta}_0 + \hat{\beta}_1 Ag_i + \hat{\beta}_2 S_i + \hat{\beta}_3 D_i + \hat{\beta}_4 Ch_i + \hat{\beta}_5 H_i + \hat{\beta}_6 W_i \,.$$

60-year-old man,
blood pressure: 140 over 90,
cholesterol: 200,
height: 69 inches,
weight: 200 pounds,

Estimated log odds of a coronary incident:

$$\begin{aligned}\log[\hat{p}/(1-\hat{p})] &= -4.5173 + .04590(60) + .00686(140) - .00694(90) \\ &\quad + .00631(200) - 0.07400(69) + 0.02014(200) = -1.2435.\end{aligned}$$

$$\hat{p} = \frac{e^{-1.2435}}{1 + e^{-1.2435}} = .224\,.$$

Plots

4.1 Informal Model Selection

Akaike's Information Criterion — AIC
Similar to C_p regression.

$$A^* = (G^2 - 134.9) - (7 - 2p)\,.$$

134.9 is G^2 for the full model
7 degrees of freedom full model
p degrees of freedom submodel
Information in $A - q$ and A^* is identical:

$$A^* = 258.1 + (A - q)$$

($258.1 =$ number of cells $- G^2$[full model] $- q$[full model] $=$
$400 - 134.9 - 7$.)

Model Variables	df	G^2	$A-q$	A^*
Ag,S,D,Ch,H,W	193	134.9	-251.1	7
Ag	198	142.7	-253.3	4.8
W	198	150.1	-245.9	12.2
H,W	197	146.8	-247.2	10.9
Ch	198	146.9	-249.1	9.0
S,D	197	147.9	-246.1	12.0
Intercept	199	154.6	-243.4	14.7

$$\log[p_i/(1-p_i)] = \gamma_0 + \gamma_1 Ag_i$$

Tests Against Full Model

Model	df	G^2
Ag	5	7.8
W	5	15.2**
H,W	4	11.9*
Ch	5	12.0*
S,D	4	13.0*
Intercept	6	19.7**

Ag and one or two other explanatory variables.

Model Variables	df	G^2	A^*
Ag,S,D,Ch,H,W	193	134.9	7.0
Ag,S,D	196	141.4	7.5
Ag,S,Ch	196	139.3	5.4
Ag,S,H	196	141.9	8.0
Ag,S,W	196	138.4	4.5
Ag,D,Ch	196	139.0	5.1
Ag,D,H	196	141.4	7.5
Ag,D,W	196	138.5	4.6
Ag,Ch,H	196	139.9	6.0
Ag,Ch,W	196	135.5	1.6*
Ag,H,W	196	138.1	4.2
Ag,S	197	141.9	6.0
Ag,D	197	141.4	5.5
Ag,Ch	197	139.9	4.0
Ag,H	197	142.7	6.8
Ag,W	197	138.8	2.9*
Ag	198	142.7	4.8

$$\log[p_i/(1-p_i)] = \gamma_0 + \gamma_1 Ag_i + \gamma_2 W_i$$

Variable	Parameter	Estimate	SE
Intercept	γ_0	-7.513	1.706
Ag	γ_1	0.06358	0.01963
W	γ_2	0.01600	0.00794

$$\log[p_i/(1-p_i)] = \eta_0 + \eta_1 Ag_i + \eta_2 W_i + \eta_3 Ch_i$$

Variable	Parameter	Estimate	SE
Intercept	η_0	-9.255	2.061
Ag	η_1	0.05300	0.02074
W	η_2	0.01754	0.003575
Ch	η_3	0.006517	0.008243

Ag and W coefficients quite stable.
Ag, W, and Ch coefficients all positive, so
small increase in age, weight, or cholesterol associated with a small increase in the odds of having a coronary incident.

Association, not causation!
Correlations between predictor variables can make interpretations of individual regression coefficients almost impossible.

Estimated full model:

Variable	Estimate	Std. Error	z
Intercept	-4.5173	7.451	-0.61
Ag	0.04590	0.02344	1.96
S	0.00686	0.02013	0.34
D	-0.00694	0.03821	-0.18
Ch	0.00631	0.00362	1.74
H	-0.07400	0.1058	-0.70
W	0.02014	0.00984	2.05

$$G^2 = 134.9, \qquad df = 193$$

Coefficient for D is $-.00694$, coefficient for S is .00686.
Use difference $S - D$ as a predictor?

$$\log[p_i/(1-p_i)] = \gamma_0 + \gamma_1 Ag_i + \gamma_2(S_i - D_i) + \gamma_3 Ch_i + \gamma_4 H_i + \gamma_5 W_i,$$

$G^2 = 134.9$ on $df = 194$.
test against full model

$$G^2 = 134.9 - 134.9 = 0$$

$df = 194 - 193 = 1$.

$H_0 : \beta_3 = -\beta_2$

$$\begin{aligned}
&\log[p_i/(1-p_i)] \\
&\quad = \beta_0 + \beta_1 Ag_i + \beta_2 S_i + (-\beta_2) D_i + \beta_4 Ch_i + \beta_5 H_i + \beta_6 W_i \\
&\quad = \beta_0 + \beta_1 Ag_i + \beta_2 (S_i - D_i) + \beta_4 Ch_i + \beta_5 H_i + \beta_6 W_i
\end{aligned}$$

4.2 Checking Lack of Fit

Tsiatis (1980), Landwehr, Pregibon, and Shoemaker (1984), and Fienberg and Gong (1984) approaches based on clustering near replicates of the regression variables

5. Logistic Regression Diagnostics

Usually: residuals for detecting outliers.

What is an outlier for 0 – 1 data?

0-1 data, reasonable observations can have "unusually large" residuals.

Too many 0s or 1s in situations where we would not expect them

(e.g., too many 1s that we think have a small p_i)

the problem is lack of fit.

Leverages and influential observations.

"In what sense is this observation influential?"

Observations are not "influential" in a vacuum.

Influence for: 1) fitted probabilities, 2) estimated regression parameters, 3) just about anything.

Identify important aspects of the model and examine influence measures appropriate to those aspects.

Fitted probabilities: Kullback-Leibler (K-L) divergence measure

Regression coefficients: Cook's distance
(Also an approximation to K-L and Cook and Weisberg's (1982) likelihood distance measure.)

Software dictates a focus on Cook's distance.

"Given some influential observations, what do you do about them?" worry about them!

Ignore influence or eliminate influence by deleting and refitting the model?

"Influential" depends on model being fitted.

Answers must depend on the data and the purpose of the analysis.

Fitting

$$\log[\hat{p}_i/(1-\hat{p}_i)] = \hat{\beta}_0 + \hat{\beta}_1 Ag_i + \hat{\beta}_2 Ch_i + \hat{\beta}_3 W_i,$$

Standard diagnostic quantities:

- fitted probability: $\hat{p}_i$,
- leverage: $\hat{a}_{ii}$,
- large sample standard error of $\hat{p}_i$: $\sqrt{\hat{p}_i(1-\hat{p}_i)(1-\hat{a}_{ii})}$,
- standardized residuals:

$$r_i = \frac{y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1-\hat{p}_i)(1-\hat{a}_{ii})}},$$

- Pearson residuals:

$$\tilde{r}_i = \frac{y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1-\hat{p}_i)}},$$

  the square of which is the ith component of Pearson's chi-square,

- deviance residuals:

$$\pm\sqrt{2[y_i \log(y_i/\hat{p}_i) + (1-y_i)\log((1-y_i)/(1-\hat{p}_i))]}$$

  sign is the sign of $y_i - \hat{p}_i$,

- Cook's distance: C_i

Formulae are for y_i either 0 or 1,

Remember: residuals are not very interesting.

Three predictor model: Ag, W, Ch

9 cases with the highest leverages:

Case	19	38	41	84	108	111	116	153	157
Leverage	.104	.081	.149	.079	.147	.062	.067	.090	.093

Case 41 has $Ag = 40$, $W = 169$, and $Ch = 520$.

Of 200 cases, only 9 have Ch values over 400.

Only 3 have Ch values over 428:
Case 116 with $Ch = 453$, Case 38 with $Ch = 474$, and Case 41.

Case 108 has $Ag = 51$, $W = 262$, and $Ch = 269$.

262 pounds is extremely high within the data.

Of 9 cases with high leverage, only 19, 41, and 111 had coronary incidents.

Cook's distances C_i: 32 cases with $C_i \geq .01$.

Include all 26 coronary incidents.

Other 6 cases: 4 among highest leverage cases: 2 reasonably high leverages.

Four cases with $C_i \geq .05$.

Case	C_i	Leverage
41	.112	.149
86	.078	.008
126	.079	.042
192	.064	.022

All had coronary incidents.

Case 41 easily the most influential.

Delete 41 from the data.

Fitted p_i's are pretty similar.

Variable	Estimate with Case 41	Estimate without Case 41	SE without Case 41
Intercept	−9.255	−8.813	2.058
Ag	0.05300	0.05924	0.02138
Ch	0.006517	0.004208	0.003881
W	0.01754	0.01714	0.008216

Role of cholesterol.

Case 41 deleted, estimate over standard error is
$.004208/.003881 = 1.08$.
Case 41 included, $.006517/.003575 = 1.82$.

Cholesterol questionable before; without Case 41, forget about it.

Case 41 deleted: $G^2(Ag, Ch, W) = 132.8$ on 195 df,
$G^2(Ag, W) = 134.0$ on 196 df.
$G^2 = 134.0 - 132.8 = 1.2$ with $196 - 195 = 1$ df.
No evidence for including cholesterol.

Without Case 41

Variable	Estimate	SE
GM	−7.756	1.745
Ag	0.06675	0.02013
W	0.01625	0.008042

6. Model Selection Methods

Stepwise methods or finding best subsets.

Standard programs do stepwise e.g., BMDP-LR and SAS PROC LOGISTIC.

These are similar to normal regression

Only standard program for best subset logistic regression is SAS PROC LOGISTIC?

Procedure based on score tests. (Bad!)

Do it yourself procedure:

EXAMPLE All variables in Chapman data

- Get $\hat{p}_i$'s.
- Define two variables:

  Weight:

  $$RWT_i = \hat{p}_i(1 - \hat{p}_i)$$

  Dependent variable:

  $$Y_i = \log[\hat{p}_i/(1 - \hat{p}_i)] + (y_i - \hat{p}_i)/RWT_i\,.$$

- Apply best subset regression program (e.g. BMDP-9R) to

  $$Y_i = \beta_0 + \beta_1 Ag_i + \beta_2 S_i + \beta_3 D_i + \beta_4 Ch_i + \beta_5 H_i + \beta_6 W_i + e_i,$$

  using weights RWT.

(Diagnostic statistics can be obtained from multiple regression program in same fashion.)

Based on the C_p statistic, five best-fitting models

Variables	C_p
Ag, Ch, W	1.66
Ag, W	2.88
Ag, Ch, H, W	3.13
Ag, S, Ch, W	3.49
Ag, D, Ch, W	3.59

Last three models add a worthless variable to Ag, Ch, and W model.

C_p statistics are based on one-step fits.

Variable	MLE	One-Step
Intercept	-9.2559	-9.21822
Ag	0.053004	0.0529624
Ch	0.0065179	0.00647380
W	0.017539	0.0174699

SAS PROC LOGISTIC uses score tests against intercept only model.

Similar method, but! — For example,

Score test for dropping all of Ag, S, D, Ch, H, and W,

Fit regression model as indicated — except score test $\hat{p}_i$'s are mle's from intercept only model.

$\hat{p}_i$'s: Use mle's from full model or mle's from intercept only model?

Exact score tests — or — one-step C_p statistics?

6.1 Computer Commands

SAS commands for logistic regression.

`www.stat.unm.edu/~fletcher/LLM`

Data in file 'chapman.dat'

Eight columns: case index, Ag, S, D, Ch, H, W, and Cnt.

The file looks like:

```
  1 44 124  80 254 70 190 0
  2 35 110  70 240 73 216 0
  3 41 114  80 279 68 178 0
  4 31 100  80 284 68 149 0
       data continue
199 50 128  92 264 70 176 0
200 31 105  68 193 67 141 0
```

PROC GENMOD: program for generalized linear models

```
options ps=60 ls=72 nodate;
data chapman;
   infile 'chapman.dat';
   input ID Ag S D Ch H W Cnt;
   n = 1;
proc genmod data=chapman ;
   model Cnt/n = Ag Ch W / link=logit
                           dist=binomial;
run;
```

The first line controls printing.

The next four lines involve defining and reading the data and creating a variable "n" that gives the total number of possible successes for each case.

The remaining lines specify the model and that a logistic regression is to be performed.

PROC LOGISTIC: more specialized features

```
options ps=60 ls=72 nodate;
data chapman;
   infile 'chapman.dat';
   input ID Ag S D Ch H W Cnt;
proc logistic data=chapman descending;
   model Cnt=Ag Ch W / waldcl waldrl plcl
                       influence iplots lackfit rsq;
   output out=chdiag predicted=phat;
run;
proc print data=chdiag;
run;
proc logistic data=chapman descending;
   model Cnt=Ag S D Ch H W / selection = score
                             best = 3 details;
run;
```

Two calls of PROC LOGISTIC:
1) standard procedure, 2) model selection.

"proc logistic", specify data being used and "descending".
"descending" makes program model probabilities of events coded as 1 (rather than events coded as 0).
"descending" makes program model the probability of a coronary incident rather than the probability of no coronary incident.

Standard output:

- estimated regression coefficients,
- standard errors,
- values of $z^2 = [Est/\mathrm{SE}(Est)]^2$,
- $P\ values$,
- $e^{\hat{\beta}_k}$'s.

Model statement specifies dependent and predictor variables.

After / on model line, options are specified.

- "waldcl" gives intervals $\hat{\beta}_k \pm 1.96\,\mathrm{SE}(\hat{\beta}_k)$;
  call the interval (a, b).
- "waldrl" gives $e^{\hat{\beta}_k}$ and intervals (e^a, e^b).
- "plcl" gives confidence intervals for β_k's based on profile likelihoods.
- "influence" causes diagnostics to be presented,
- Cbar: $(1 - \hat{a}_{ii})C_i$.
- Index plots by specifying "iplots".
- "lackfit"
- "rsq" gives some R^2 and Adj. R^2 values,

"output" creates SAS data set of diagnostics,

"selection" option: "forward", "backwards", "stepwise", or "score".

"score" and "best = 3": three one-variable models with the best score statistics, the three best two-variable models, the three best three-variable models, etc.

7. ANOVA Type Logit Models

EXAMPLE: Muscle tension data.

Factor	Abbreviation	Levels
Change in muscle tension	T	High, Low
Weight of muscle	W	High, Low
Muscle type	M	Type 1, Type 2
Drug	D	Drug 1, Drug 2

			Drug (k)	
Tension (h)	Weight (i)	Muscle (j)	Drug 1	Drug 2
	High	Type 1	3	21
		Type 2	23	11
High				
	Low	Type 1	22	32
		Type 2	4	12
	High	Type 1	3	10
		Type 2	41	21
Low				
	Low	Type 1	45	23
		Type 2	6	22

Change in tension — response factor.
Weight, muscle type, drug — explanatory.

Saturated model

$$\begin{aligned}\log(p_{1ijk}/p_{2ijk}) &= G + W_i + M_j + D_k \\ &\quad + (WM)_{ij} + (WD)_{ik} + (MD)_{jk} \\ &\quad + (WMD)_{ijk}\,.\end{aligned}$$

$$\log(p_{1ijk}/p_{2ijk}) = (WMD)_{ijk}\,.$$

$$\log(m_{hijk}) = (\tau\omega\mu\delta)_{hijk}\,,$$

In general,

$$p_{1ijk}/p_{2ijk} = m_{1ijk}/m_{2ijk}\,.$$

Reduced model

$$\log(p_{1ijk}/p_{2ijk}) = W_i + (MD)_{jk}$$

$$\log(m_{hijk}) = (\tau\omega)_{hi} + (\tau\mu\delta)_{hjk} + (\omega\mu\delta)_{ijk}\,.$$

$(\omega\mu\delta)_{ijk}$ included to deal with sampling scheme

Table 1: Correspondence Between Some Logit and Log-Linear Models

	Logit Model	Log-Linear Model
1)	{WM}{WD}{MD}	[WMD][TWM][TWD][TMD]
2)	{WM}{WD}	[WMD][TWM][TWD]
3)	{WM}{MD}	[WMD][TWM][TMD]
4)	{WD}{MD}	[WMD][TWD][TMD]
5)	{WM}{D}	[WMD][TWM][TD]
6)	{WD}{M}	[WMD][TWD][TM]
7)	{MD}{W}	[WMD][TMD][TW]
8)	{W}{M}{D}	[WMD][TW][TM][TD]
9)	{W}{M}	[WMD][TW][TM]
10)	{W}{D}	[WMD][TW][TD]
11)	{M}{D}	[WMD][TM][TD]

Line 3 of table:

$$\log(p_{1ijk}/p_{2ijk}) = G + W_i + M_j + D_k + (WM)_{ij} + (MD)_{jk}$$

is equivalent to

$$\begin{aligned}\log(m_{hijk}) \;=\; & \gamma + \omega_i + \mu_j + \delta_k + (\omega\mu)_{ij} + (\omega\delta)_{ik} + (\mu\delta)_{jk} + (\omega\mu\delta)_{ijk} \\ & + \tau_h + (\tau\omega)_{hi} + (\tau\mu)_{hj} + (\tau\delta)_{hk} \\ & + (\tau\omega\mu)_{hij} + (\tau\omega\delta)_{hik}.\end{aligned}$$

or

$$\log(m_{hijk}) = (\omega\mu\delta)_{ijk} + (\tau\omega\mu)_{hij} + (\tau\omega\delta)_{hik}\,.$$

Logit models obtained from log-linear models by subtraction.

$$\begin{aligned}
\log(p_{1ijk}/p_{2ijk}) &= \log(m_{1ijk}/m_{2ijk}) \\
&= \log(m_{1ijk}) - \log(m_{2ijk}) \\
&= (\tau\omega)_{1i} + (\tau\mu\delta)_{1jk} + (\omega\mu\delta)_{ijk} \\
&\quad - (\tau\omega)_{2i} - (\tau\mu\delta)_{2jk} - (\omega\mu\delta)_{ijk} \\
&= [(\tau\omega)_{1i} - (\tau\omega)_{2i}] + [(\tau\mu\delta)_{1jk} - (\tau\mu\delta)_{2jk}] \\
&= W_i + (MD)_{jk}
\end{aligned}$$

$W_i \equiv [(\tau\omega)_{1i} - (\tau\omega)_{2i}] \quad (MD)_{jk} \equiv [(\tau\mu\delta)_{1jk} - (\tau\mu\delta)_{2jk}]$.

Logits model data as two factor table.

Smallest interesting log-linear model is independence:

$$\log(m_{hijk}) = \tau_h + (\omega\mu\delta)_{ijk}\,.$$

Looking at $\log(p_{1ijk}/p_{2ijk}) = \log(m_{1ijk}) - \log(m_{2ijk})$,

$$\log(p_{1ijk}/p_{2ijk}) = \tau_1 - \tau_2 \equiv G$$

Saturated model for two-factor table is the interaction model

$$\log(m_{hijk}) = \tau_h + (\omega\mu\delta)_{ijk} + (\tau\omega\mu\delta)_{hijk}\,,$$

$$\log(p_{1ijk}/p_{2ijk}) = G + (WMD)_{ijk}$$

- Interesting logit models correspond to modeling the interaction in this two-way table.
- More interaction than complete independence.
- Less interaction than the saturated model.

Table 2: Statistics for Logit Models

Logit Model	df	G^2	P	$A - q$
{WM}{WD}{MD}	1	0.111	.7389	−1.889
{WM}{WD}	2	2.810	.2440	−1.190
{WM}{MD}	2	0.1195	.9417	−3.8805
{WD}{MD}	2	1.059	.5948	−2.941
{WM}{D}	3	4.669	.1966	−1.331
{WD}{M}	3	3.726	.2919	−2.274
{MD}{W}	3	1.060	.7898	−4.940
{W}{M}{D}	4	5.311	.2559	−2.689
{W}{M}	5	11.35	.0443	1.35
{W}{D}	5	12.29	.0307	2.29
{M}{D}	5	7.698	.1727	−2.302

A closer look at logit model {MD}{W}.

Actual fitted log-linear models corresponding logit models.

{MD}{W} corresponds to [WMD][TMD][TW].

Logit model terms becomes interactions with response factor T interaction between all of the explanatory factors.

Table 3: Estimated Expected Cell Counts for the Log-Linear Model [WMD][TMD][TW]

			Drug (k)	
Tension (h)	Weight (i)	Muscle (j)	Drug 1	Drug 2
High	High	Type 1	2.31	20.04
		Type 2	23.75	11.90
	Low	Type 1	22.68	32.96
		Type 2	3.26	11.10
Low	High	Type 1	3.69	10.97
		Type 2	40.24	20.10
	Low	Type 1	44.32	22.03
		Type 2	6.74	22.90

Ratio of high tension change to low tension change gives estimated odds.

Table 4: Estimated Odds of High Tension Change for the Logit Model {MD}{W}

		Drug	
Weight	Muscle	Drug 1	Drug 2
High	Type 1	.625	1.827
	Type 2	.590	.592
Low	Type 1	.512	1.496
	Type 2	.483	.485

Main Effect:
High tension change odds are 1.22 times greater for high-weight muscles than for low-weight muscles.
$.625/.512 = 1.22$, $1.22 = .590/.483 = 1.827/1.495 = .592/.485$.
$\hat{m}_{11jk}\hat{m}_{22jk}/\hat{m}_{12jk}\hat{m}_{21jk} = 1.22$.

Muscle type—drug interaction:

High weights: (top Table 4).

All about .6,
except type 1, drug 2.

Estimated probability about $.6/(1 + .6) = .375$.

For type 1, drug 2, estimated odds are 1.827
estimated probability change is $1.827/(1 + 1.827) = .646$.

Similar for low-weight odds except smaller by factor of 1.22 because of main effect for weight.

Table 5: Estimated Odds for the Logit Model {WM}{MD}

		Drug	
Weight	Muscle	Drug 1	Drug 2
High	Type 1	.809	2.202
	Type 2	.569	.512
Low	Type 1	.499	1.358
	Type 2	.619	.557

Table 6: Estimated Odds for the Logit Model {WM}{MD}

		Drug	
Muscle	Weight	Drug 1	Drug 2
Type 1	High	.809	2.202
	Low	.499	1.358
Type 2	High	.569	.512
	Low	.619	.557

Other good logit model is {WM}{MD}.

Type 2 muscles: About the same regardless of weight and drug.

Contrary to our previous model, do not depend much on weight, (odds go down rather than up for higher weights).

Type 1 muscles, same dominant features as previous model.

The difference between the models {MD}{W} and {WM}{MD} is that:
In {MD}{W}, for type 2 muscles,
high weight should increase the odds,
{WM}{MD} indicates little change for high weight
(any change is a decrease)

Reason for rewriting table:
{WM}{MD} has M in both terms,
fixed level of M, the effects of W and D are additive,
(size of effects change with the level of M).

This analysis at lowest level of *technical* sophistication.
Fitted values and likelihood ratio test statistics.
Conclusions drawn without standard errors.

7.1 Computer Commands

File 'tenslr.dat' columns for number of high tension scores and low tension scores, three columns of indices for weight (high is 1), muscle type, and drug, respectively.

```
 3  3 1 1 1
21 10 1 1 2
23 41 1 2 1
11 21 1 2 2
22 45 2 1 1
32 23 2 1 2
 4  6 2 2 1
12 22 2 2 2
```

Fit {WM}{WD}{MD} using SAS PROC GENMOD.
"n" is number of individuals with level of weight, muscle type, and drug. "class" command distinguishes ANOVA type factors from regression predictors.

```
options ps=60 ls=72 nodate;
data tension;
   infile 'tenslr.dat';
   input H L W M D;
   n = H+L;
proc genmod data=tension;
   class W M D;
   model H/n = W*M W*D M*D / link=logit
                             dist=binomial;
run;
```

```
proc print data=chdiag;
run;
```

or fit log-linear model [WMD][TWM][TWD][TMD]. To fit {WM}{MD} or {WM}{D} in GENMOD, model statement uses W∗M M∗D or W∗M D, respectively.

8. Logit Models for a Multinomial Response

Gneralizations of logistic regression for more than one response category.

Abortion opinions given order: Yes, No, Undecided.

- Odds of consecutive categories (good for ordered categories)

$$\log(m_i/m_{i+1}), \qquad i = 1, \ldots, R-1\,,$$

  equivalently,

$$\log(p_i/p_{i+1}), \qquad i = 1, \ldots, R-1\,.$$

  – odds of Yes to No,
  – odds of No to Undecided.

- Compare each level to a particular level;

$$\log(m_i/m_R), \qquad i = 1, \ldots, R-1\,.$$

  - odds of Yes to Undecided,
  - odds of No to Undecided.

  Rearrange order of levels to make more interesting.

Both sets of models equivalent to a log-linear model.

Example:

$$\log(m_{ijk}/m_{i+1\ jk}) = w_{2(j)} + w_{3(k)}, \qquad i = 1, \ldots, R-1,$$

and

$$\log(m_{ijk}/m_{Rjk}) = v_{2(j)} + v_{3(k)}, \qquad i = 1, \ldots, R-1,$$

are equivalent. (w and v parameters depend on i.)

Both are equivalent to

$$\log(m_{ijk}) = u_{23(jk)} + u_{12(ij)} + u_{13(ik)}\,.$$

- Pool response levels, compare each level to total of all others,

$$\log\left(\frac{m_i}{\Sigma_{h\neq i}\, m_h}\right), \qquad i = 1, \ldots, R\,.$$

  - odds of Yes to not Yes,
  - odds of No to not No,
  - odds of Undecided to not undecided (Decided).

- *Continuation ratios:* usually for ordered categories

$$\log\left(\frac{m_i}{\Sigma_{h=i+1}^{R} m_h}\right), \qquad i = 1, \ldots, R-1\,.$$

  – odds of Yes to not Yes,

  – odds of No to Undecided.

  Rearrange ordering: Undecided, Yes, No.

  – odds of Undecided to Decided.

  – odds of Yes to No.

- *Cumulative logits*,

$$\log\left(\frac{\Sigma_{h=1}^{i} m_h}{\Sigma_{h=i+1}^{R} m_h}\right), \qquad i = 1, \ldots, R-1\,.$$

  Abortion opinions: Undecided, Yes, No.

  – odds of Undecided to Decided.

  – odds of not opposed to opposed.

Table 7: Log-Linear Models for the Abortion Opinion Data

Model	df	G^2	$A-q$
[RSA][RSO][ROA][SOA]	10	6.12	−13.88
[RSA][RSO][ROA]	20	7.55	−32.45
[RSA][RSO][SOA]	20	13.29	−26.71
[RSA][ROA][SOA]	12	16.62	−7.38
[RSA][RSO][OA]	30	14.43	−45.57
[RSA][ROA][SO]	22	17.79	−26.21
[RSA][SOA][RO]	22	23.09	−20.91
[RSA][RO][SO][OA]	32	24.39	−39.61
[RSA][RO][SO]	42	87.54	3.54
[RSA][RO][OA]	34	34.41	−33.59
[RSA][SO][OA]	34	39.63	−28.37
[RSA][RO]	44	97.06	9.06
[RSA][SO]	44	101.9	13.9
[RSA][OA]	36	49.37	−22.63
[RSA][O]	46	111.1	19.1

EXAMPLE Abortion data.

Opinions as a response variable, so [RSA] in all models.

Table 7 has fits for all ANOVA type logit models with [RSA].

The best fitting model [RSA][RSO][OA] determines consecutive categories

$$\begin{aligned}\log(m_{hi1k}/m_{hi2k}) &= w^1_{RS(hi)} + w^1_{A(R)}\,,\\ \log(m_{hi2k}/m_{hi3k}) &= w^2_{RS(hi)} + w^2_{A(k)}\,,\end{aligned}$$

or fixed level

$$\begin{aligned}\log(m_{hi1k}/m_{hi3k}) &= v^1_{RS(hi)} + v^1_{A(k)}\\ \log(m_{hi2k}/m_{hi3k}) &= v^2_{RS(hi)} + v^2_{A(k)}\,.\end{aligned}$$

First pair: odds of support to opposing legalized abortion; odds of opposing to being undecided.

The second pair: odds of support to undecided; odds of opposing to undecided.

Choosing "undecided" as the standard level is particularly unintuitive. That undecided is the last category is no reason to chose it as the standard.

Table 8: Fitted Values for [RSA][RSO][OA]

			Age					
Race	Sex	Opinion	18-25	26-35	36-45	46-55	56-65	65+
		Support	100.1	137.2	117.5	75.62	70.58	80.10
	Male	Oppose	39.73	64.23	56.17	47.33	50.99	62.55
		Undec.	1.21	2.59	5.36	5.05	5.43	8.35
White								
		Support	138.4	172.0	152.4	101.8	101.7	110.7
	Female	Oppose	43.49	63.77	57.68	50.44	58.19	68.43
		Undec.	2.16	4.18	8.96	8.76	10.08	14.86
		Support	21.19	16.57	15.20	11.20	8.04	7.80
	Male	Oppose	8.54	7.88	7.38	7.11	5.90	6.18
		Undec.	1.27	1.54	3.42	3.69	3.06	4.02
Nonwhite								
		Support	21.40	26.20	19.98	16.38	13.64	12.40
	Female	Oppose	4.24	6.12	4.77	5.12	4.92	4.83
		Undec.	0.36	0.68	1.25	1.50	1.44	1.77

Table 9: Estimated Odds of Support versus Oppose

Legalized Abortion
(Based on the log-linear model [RSA][RSO][OA])

		Age					
Race	Sex	18-25	26-35	36-45	46-55	56-65	65+
White	Male	2.52	2.14	2.09	1.60	1.38	1.28
	Female	3.18	2.70	2.64	2.02	1.75	1.62
Nonwhite	Male	2.48	2.10	2.06	1.57	1.36	1.26
	Female	5.05	4.28	4.19	3.20	2.77	2.57

Continuation ratios with ordering: Undecided, Yes, No.

- odds of Undecided to Decided.
- odds of Yes to No.

Odds of Support versus Opposed
undecideds excluded

		Age					
Race	Sex	18-25	26-35	36-45	46-55	56-65	65+
White	Male	2.52	2.14	2.09	1.60	1.38	1.28
	Female	3.18	2.70	2.64	2.01	1.75	1.62
Nonwhite	Male	2.48	2.11	2.06	1.57	1.36	1.26
	Female	5.08	4.31	4.22	3.22	2.79	2.58

Except for nonwhite females, the odds of support are essentially identical to those obtained with undecideds included.

Table 10: Estimated Expected Cell Counts with Undecideds Eliminated

			Age					
Race	Sex	Opinion	18-25	26-35	36-45	46-55	56-65	65+
White	Male	Support	100.2	137.7	117.0	75.62	70.22	80.27
		Oppose	39.78	64.35	55.98	47.38	50.78	62.73
	Female	Support	139.2	172.2	152.3	101.6	101.7	109.9
		Oppose	43.78	63.77	57.71	50.41	58.28	68.05
Nonwhite	Male	Support	20.67	16.96	15.48	11.00	8.07	7.81
		Oppose	8.33	8.04	7.52	7.00	5.93	6.19
	Female	Support	20.84	25.17	20.21	16.78	13.98	12.97
		Oppose	4.11	5.84	4.79	5.22	5.02	5.03

G^2 without undecideds is 9.104 on 15 df.
G^2 for [RSA][RO][SO][OA] is 11.77 on 16 df.
Difference is not large, so $\log(m_{hi1k}/m_{hi2k}) = R_{(h)} + S_{(i)} + A_{(k)}$ may fit.

Pool support and oppose categories: $2 \times 2 \times 2 \times 6$ table.
[RSA][RSO][OA] fitted

Odds of Being Decided on Abortion

		Age					
Race	Sex	18-25	26-35	36-45	46-55	56-65	65+
White	Male	116.79	78.52	32.67	24.34	22.26	16.95
	Female	83.43	56.08	23.34	17.38	15.90	12.11
Nonwhite	Male	23.76	15.97	6.65	4.95	4.53	3.45
	Female	68.82	46.26	19.25	14.34	13.12	9.99

Odds decrease with age; older people are less likely to take a position.

White males most likely to state a position.

Nonwhite males least likely to state a position.

White and nonwhite females have similar odds of being decided.

The G^2 for [RSA][RSO][OA] is 5.176 on 15 df. The G^2 for the smaller model [RSA][RO][SO][OA] is 12.71 on 16 df. The difference is very large.

Model was [RSA][RSO][OA], equivalently

$$\log(m_{hi1k}/m_{hi2k}) = (RS)_{hi} + A_k$$

$G^2 = 9.104$ on 15 df, same as [RSA][RSO][OA].

Table of support versus oppose odds suggests:
(1) odds decrease as age increases (2) the odds for males are about the same.

Fit models that incorporate these.

Data suggesting models, formal tests less appropriate.

G^2's still measure quality of model fit.
Linear trend in ages. Codes ages as $k = 1, 2, \ldots, 6$. Fit

$$\log(m_{hi1k}/m_{hi2k}) = (RS)_{hi} + \gamma k\,.$$

G^2 is 10.18 on 19 df, fits very well.
$\{RS\}\{A\}$ has $G^2 = 9.104$ on 15 df,
Difference is $G^2 = 10.18 - 9.104 = 1.08$ on $19 - 15 = 4$ df.
Incorporate males having same odds of support. Recode

(h,i)	(1,1)	(1,2)	(2,1)	(2,2)
g	1	2	3	4

$$\log(m_{g1k}/m_{g2k}) = (RS)_g + A_k$$

Same model, same fit.

Recode again

	wm	wf	nm	nf
g	1	2	3	4
f	1	2	1	3
e	1	1	2	1

$$\log(m_{fe1k}/m_{fe2k}) = (RS)_{fe} + A_k$$

Same model, same fit.

Now model

$$\log(m_{fe1k}/m_{fe2k}) = (RS)_{f} + A_k \, .$$

Male groups are distinguished by e, and e does not appear.

This has $G^2 = 9.110$ on 16 df.

Compare to $\{RS\}\{A\}$: $G^2 = 9.110 - 9.104 = .006$ on $16 - 15 = 1$ df.

Incorporate both trend in ages and equality for males

$$\log(m_{fe1k}/m_{fe2k}) = (RS)_f + \gamma k\,.$$

$G^2 = 10.19$ on 20 df.

Only increased the G^2 by $10.19 - 9.10 = 1.09$.

For

$$\log(m_{fe1k}/m_{fe2k}) = \mu + (RS)_f + \gamma k\,,$$

side condition $(RS)_1 = 0$

Parameter	Estimate	SE	Est./SE
μ	1.071	.1126	9.51
$(RS)_1$	0	—	—
$(RS)_2$	.2344	.09265	2.53
$(RS)_3$	.6998	.2166	3.23
γ	$-.1410$	.02674	-5.27

All terms seem important.
With side condition, $(\widehat{RS})_2$ actually estimates $(RS)_2 - (RS)_1$.

z score 2.53 indicates white females differ from males.

$(\widehat{RS})_3$ estimates difference between nonwhite females and males.

The estimated odds of support are

	Age					
Race-Sex	18-25	26-35	36-45	46-55	56-65	65+
Male	2.535	2.201	1.912	1.661	1.442	1.253
White female	3.204	2.783	2.417	2.099	1.823	1.583
Nonwhite female	5.103	4.432	3.850	3.343	2.904	2.522

Same general characteristics as discussed earlier.

Transform into (conditional) probabilities of support.
(Easier to interpret than odds.)

Estimated probability that white female between 46 and 55 years supports legalized abortion

$$2.099/(1 + 2.099) = .677.$$

Odds about 2, probability about twice as great for support than oppose.

SOME EQUIVALENCES

$$\log(m_{fe1k}/m_{fe2k}) = (RS)_{hi} + \gamma k$$

corresponds to

$$\log(m_{hijk}) = (RSA)_{hik} + (RSO)_{hij} + \gamma_j k$$

$$\log(m_{fe1k}/m_{fe2k}) = (RS)_f + A_k$$

corresponds to

$$\log(m_{fejk}) = (RSA)_{fek} + (RSO)_{fj} + (OA)_{jk}$$

$$\log(m_{fe1k}/m_{fe2k}) = (RS)_f + \gamma k$$

corresponds to

$$\log(m_{fejk}) = (RSA)_{fek} + (RSO)_{fj} + \gamma_j k$$

Asmussen and Edwards (1983) allow fitting models that do not always include a term for the interactions among the explanatory factors. They argue that *log-linear models are appropriate for response factors as long as the model allows for collapsing over the response factors onto the explanatory factors.*

9. Logistic Discrimination and Allocation

How can you tell Swedes and Italians apart?

How can you tell different species of irises apart?

How can you identify people who are likely to have a heart attack or commit a crime?

Collect data on individuals who in each populations.

Discriminate between the populations.

To identify Swedes and Italians, data on height, hair color, eye color, and skin complexion?

To identify irises, measure petal length and width and sepal length and width.

Typically, data on several variables combined to identify the likelihood that someone belongs to a particular population.

Independent samples from each population.

Use samples to characterize the populations (discrimination).

Allocation identifies population for an individual when only variable values are known.

Typically
Discrimination data arises from a retrospective study.

Table 11: Cushing's Syndrome Data

Case	Type	TETRA	PREG	Case	Type	TETRA	PREG
1	A	3.1	11.70	12	B	15.4	3.60
2	A	3.0	1.30	13	B	7.7	1.60
3	A	1.9	0.10	14	B	6.5	0.40
4	A	3.8	0.04	15	B	5.7	0.40
5	A	4.1	1.10	16	B	13.6	1.60
6	A	1.9	0.40	17	C	10.2	6.40
7	B	8.3	1.00	18	C	9.2	7.90
8	B	3.8	0.20	19	C	9.6	3.10
9	B	3.9	0.60	20	C	53.8	2.50
10	B	7.8	1.20	21	C	15.8	7.60
11	B	9.1	0.60				

EXAMPLE 21 people with Cushing's syndrome (Aitchison and Dunsmore, 1975).

Overproduction of cortisol by the adrenal cortex.

A−adenoma
B−bilateral hyperplasia
C−carcinoma

Data:

TETRA – Tetrahydrocortisone

and

PREG – Pregnanetriol.

Figure 1: Cushing's Syndrome Data

Data determine 3×21 table

	Case														
Type	1	2	3	4	5	6	7	8	$\cdots$	16	17	18	19	20	21
A	1	1	1	1	1	1	0	0	$\cdots$	0	0	0	0	0	0
B	0	0	0	0	0	0	1	1	$\cdots$	1	0	0	0	0	0
C	0	0	0	0	0	0	0	0	$\cdots$	0	1	1	1	1	1

$$TL \equiv \log(\text{TETRA})$$

$$PL \equiv \log(\text{PREG})$$

Independent samples from three populations: A, B, and C.

Product-multinomial sampling.

Observations are intrinsically discrete.

Observations only within $\pm.05$ mg and $\pm.005$ mg of respective nominal values.

Categories are all *observable* combinations of TL and PL.

Huge number of possible categories, say, S.

Sparsely sampled $3 \times S$ table.

Most column totals are 0, so MLEs do not exist because

$$\text{estimated column totals} > 0$$

must equal

$$\text{observed column totals} = 0.$$

Analysis conducted on *observed* 3×21 table.

Works well even though sampling scheme is clearly wrong.

True product-multinomial sampling allows for col. totals of 0, which cannot happen in 3×21 table.

Fit

$$\log(m_{ij}) = \alpha_i + \beta_j + \gamma_{1i}(TL)_j + \gamma_{2i}(PL)_j \,,$$

$i = 1, 2, 3$, $j = 1, \ldots, 21$.

Similar to log-linear version of logit/logistic models.

β_j for each combination of variables.

$$\log(m_{1j}/m_{2j}) = (\alpha_1 - \alpha_2) + (\gamma_{11} - \gamma_{12})(TL)_j + (\gamma_{21} - \gamma_{22})(PL)_j$$

or

$$\log(m_{1j}/m_{2j}) = \alpha + \delta_1(TL)_j + \delta_2(PL)_j \,.$$

Looks like a logistic regression model but is not,

$$\log\left(\frac{m_{1j}}{m_{2j}}\right) \neq \log\left(\frac{p_{1j}}{p_{2j}}\right) \,.$$

p_{1j}/p_{2j} is NOT odds of A to B.

Probabilities are from different populations.

p_{1j}/p_{2j} is a likelihood ratio.

p_{ij} is the likelihood within population i of observing category j.

Estimate of "log likelihood ratio" is

$$\log\left(\frac{\hat{p}_{1j}}{\hat{p}_{2j}}\right) = \log\left(\frac{\hat{m}_{1j}/n_{1\cdot}}{\hat{m}_{2j}/n_{2\cdot}}\right) = \log\left(\frac{\hat{m}_{1j}}{\hat{m}_{2j}}\right) - \log\left(\frac{n_{1\cdot}}{n_{2\cdot}}\right) .$$

Asymp. variances more complicated than logistic regression.

Although odds depend on the sampling scheme, odds ratios do not.

Odds ratios handled exactly as in logistic regression.

$$\log(m_{ij}) = \alpha_i + \beta_j + \gamma_{1i}(TL)_j + \gamma_{2i}(PL)_j \,,$$

$G^2 = 12.30$ on 36 df.

G^2 cannot be compared to a χ^2, but can test reduced models.

$$\log(m_{ij}) = \alpha_i + \beta_j + \gamma_{1i}(TL)_j$$

$G^2 = 21.34$ on 38 df

$$\log(m_{ij}) = \alpha_i + \beta_j + \gamma_{2i}(PL)_j$$

$G^2 = 37.23$ on 38 df

Neither reduced model provides an adequate fit.

(χ^2 tests of model comparisons are valid.)

Use

$$\log(m_{ij}) = \alpha_i + \beta_j + \gamma_{1i}(TL)_j + \gamma_{2i}(PL)_j$$

in remainder.

Table 12: Estimated Probabilities: $\hat{p}_{ij} = \hat{m}_{ij}/n_{i\cdot}$.

	Group				Group		
Case	A	B	C	Case	A	B	C
1	.1485	.0012	.0195	12	.0000	.0295	.1411
2	.1644	.0014	.0000	13	.0000	.0966	.0068
3	.1667	.0000	.0000	14	.0001	.0999	.0000
4	.0842	.0495	.0000	15	.0009	.0995	.0000
5	.0722	.0565	.0003	16	.0000	.0907	.0185
6	.1667	.0000	.0000	17	.0000	.0102	.1797
7	.0000	.0993	.0015	18	.0000	.0060	.1879
8	.1003	.0398	.0000	19	.0000	.0634	.0733
9	.0960	.0424	.0000	20	.0000	.0131	.1738
10	.0000	.0987	.0025	21	.0000	.0026	.1948
11	.0000	.0999	.0003				

Table 12 contains estimated probabilities for the three populations.

Bayes theorem

$$\hat{\pi}(i|Data) = \frac{\hat{p}_{ij}\pi(i)}{\Sigma_{i=1}^{3}\hat{p}_{ij}\pi(i)}\,.$$

Two choices of prior probabilities

- probabilities proportional to sample sizes $\pi(i) = n_{i.}/n_{..}$
- equal probabilities $\pi(i) = \frac{1}{3}$.

Prior probabilities proportional to sample sizes are *rarely appropriate*, but relate simply to standard output,

so they get more prominence than they deserve.

Proportional probabilities are $\hat{m}_{ij}$ values.

Equal probabilities: divide $\hat{p}_{ij}$ in Table 12 by the sum of the three probabilities for each case.

Table 13: Probabilities of Classification

Case		Group	Proportional Prior Probabilities A	B	C	Equal Prior Probabilities A	B	C
1		A	.89	.01	.10	.88	.01	.12
2		A	.99	.01	.00	.99	.01	.00
3		A	1.00	.00	.00	1.00	.00	.00
4		A	.50	.50	.00	.63	.37	.00
5	**	A	.43	.57	.00	.56	.44	.00
6		A	1.00	.00	.00	1.00	.00	.00
7		B	.00	.99	.01	.00	.99	.01
8	**	B	.60	.40	.00	.72	.28	.00
9	**	B	.58	.42	.00	.69	.31	.00
10		B	.00	.99	.01	.00	.97	.03
11		B	.00	1.00	.00	.00	1.00	.00
12	**	B	.00	.29	.71	.00	.17	.83
13		B	.00	.97	.03	.00	.93	.07
14		B	.00	1.00	.00	.00	1.00	.00
15		B	.01	.99	.00	.01	.99	.00
16		B	.00	.91	.09	.00	.83	.17
17		C	.00	.10	.90	.00	.05	.95
18		C	.00	.06	.94	.00	.03	.97
19	**	C	.00	.63	.37	.00	.46	.54
20		C	.00	.13	.87	.00	.07	.93
21		C	.00	.03	.97	.00	.01	.99

Table 14: Summary of Classifications

Allocated to Group	Proportional Prior Probabilities True Group A	B	C	Equal Prior Probabilities True Group A	B	C
A	5	2	0	6	2	0
B	1	7	1	0	7	0
C	0	1	4	0	1	5

Allocation

Model includes separate β_j for each case.

Not clear how to use model to allocate future cases.

Begin with log-linear model, develop logit models, work back to an allocation model.

$$\log(m_{ij}) = \alpha_i + \beta_j + \gamma_{1i}(TL)_j + \gamma_{2i}(PL)_j\,,$$

$i = 1, 2, 3$, $j = 1, \ldots, 21$ has 30 parameters, only 9 of interest.

Of these nine, only six are estimable.

Log probability ratio of type A relative to type B

$$\begin{aligned}
&\log(p_{1j}/p_{2j}) \\
&\quad = \log(m_{1j}/m_{2j}) - \log(n_{1\cdot}/n_{2\cdot}) \\
&\quad = (\alpha_1 - \alpha_2) + (\gamma_{11} - \gamma_{12})(TL)_j + (\gamma_{21} - \gamma_{22})(PL)_j - \log(n_{1\cdot}/n_{2\cdot})\,.
\end{aligned}$$

The log-likelihoods of A relative to C are

$$\begin{aligned}
&\log(p_{1j}/p_{3j}) \\
&\quad = \log(m_{1j}/m_{3j}) - \log(n_{1\cdot}/n_{3\cdot}) \\
&\quad = (\alpha_1 - \alpha_3) + (\gamma_{11} - \gamma_{13})(TL)_j + (\gamma_{21} - \gamma_{23})(PL)_j - \log(n_{1\cdot}/n_{3\cdot})\,.
\end{aligned}$$

These models eliminate β_j which we won't know for a new case.

Estimated parameters:

Par.	Est.	Par.	Est.	Par.	Est.
α_1	0.0	γ_{11}	-16.29	γ_{21}	-3.359
α_2	-20.06	γ_{12}	-1.865	γ_{22}	-3.604
α_3	-28.91	γ_{13}	0.0	γ_{23}	0.0

Values of 0 are side conditions.

New case with TL and PL,

$$\log(\hat{p}_1/\hat{p}_2) = 20.06+(-16.29+1.865)TL+(-3.359+3.604)PL-\log(6/10)$$

and

$$\log(\hat{p}_1/\hat{p}_3) = 28.91 - 16.29(TL) - 3.359(PL) - \log(6/5)\,.$$

TETRA $= 4.1$, PREG $= 1.10$, then $\log(\hat{p}_1/\hat{p}_2) = .24069$ and $\log(\hat{p}_1/\hat{p}_3) = 5.4226$. The likelihood ratios are

$$\begin{aligned} \hat{p}_1/\hat{p}_2 &= 1.2721 \\ \hat{p}_1/\hat{p}_3 &= 226.45 \end{aligned}$$

and by division,

$$\hat{p}_2/\hat{p}_3 = 226.45/1.2721 = 178.01\,.$$

Type A is a bit more likely than B.
Both are much more likely than type C

Estimated posterior probabilities for a new case.

Posterior odds are

$$\frac{\hat{\pi}(1|Data)}{\hat{\pi}(2|Data)} = \frac{\hat{p}_1}{\hat{p}_2}\frac{\pi(1)}{\pi(2)} \equiv \hat{O}_2$$

and

$$\frac{\hat{\pi}(1|Data)}{\hat{\pi}(3|Data)} = \frac{\hat{p}_1}{\hat{p}_3}\frac{\pi(1)}{\pi(3)} \equiv \hat{O}_3\,.$$

Also,

$$\hat{\pi}(1|Data) + \hat{\pi}(2|Data) + \hat{\pi}(3|Data) = 1.$$

Three equations in three unknowns,

solve for $\hat{\pi}(i|Data)$, $i = 1, 2, 3$.

$$\begin{aligned}
\hat{\pi}(1|Data) &= \left[1 + \frac{1}{\hat{O}_2} + \frac{1}{\hat{O}_3}\right]^{-1} = \frac{\hat{O}_2\hat{O}_3}{\hat{O}_2\hat{O}_3 + \hat{O}_3 + \hat{O}_2}\,, \\
\hat{\pi}(2|Data) &= \frac{1}{\hat{O}_2}\left[1 + \frac{1}{\hat{O}_2} + \frac{1}{\hat{O}_3}\right]^{-1} = \frac{\hat{O}_3}{\hat{O}_2\hat{O}_3 + \hat{O}_3 + \hat{O}_2}\,, \\
\hat{\pi}(3|Data) &= \frac{1}{\hat{O}_3}\left[1 + \frac{1}{\hat{O}_2} + \frac{1}{\hat{O}_3}\right]^{-1} = \frac{\hat{O}_2}{\hat{O}_2\hat{O}_3 + \hat{O}_3 + \hat{O}_2}\,.
\end{aligned}$$

TETRA = 4.10 and PREG = 1.10.
For $\pi(i) = n_{i\cdot}/n_{\cdot\cdot}$

$$\begin{aligned}\hat{\pi}(1|Data) &= .433\\ \hat{\pi}(2|Data) &= .565\\ \hat{\pi}(3|Data) &= .002\,.\end{aligned}$$

Assuming $\pi(i) = 1/3$

$$\begin{aligned}\hat{\pi}(1|Data) &= .560\\ \hat{\pi}(2|Data) &= .438\\ \hat{\pi}(3|Data) &= .002\,.\end{aligned}$$

Values of TETRA and PREG are identical to those for case 5; thus, $\hat{\pi}(i|Data)$'s are identical to those listed in Table 13 for case 5.

It is easy to just fit log-linear or logistic models to discrimination data to get $\hat{m}_{ij}$'s or $\hat{p}_{ij}$'s, respectively.

If you treat these values as estimated probabilities for being in the various populations, you are doing a Bayesian analysis with prior probabilities proportional to sample sizes.

This is rarely an appropriate methodology.