Multilevel Models for Clustered Dichotomous Outcomes

Don Hedeker
University of Illinois at Chicago

www.uic.edu/∼hedeker/ml.html

Hedeker, D. (2005). Generalized linear mixed models. In B. Everitt & D. Howell (Eds.), Encyclopedia of Statistics in Behavioral Science. Wiley.

Multilevel models for categorical outcomes

- dichotomous outcomes
  - mixed-effects logistic regression

- ordinal outcomes
  - mixed-effects ordinal logistic regression
    - proportional odds model
    - partial or non-proportional odds model

- discrete or grouped time-to-event data
  - mixed-effects dichotomous or ordinal regression
  - replace logistic link with complementary log-log link to yield proportional (and non-proportional) hazards models

SMOKING
NON-
SMOKING

©1999 Humerus Cartoons Syndicate, humerusjkb@aol.com

Logistic Regression Model

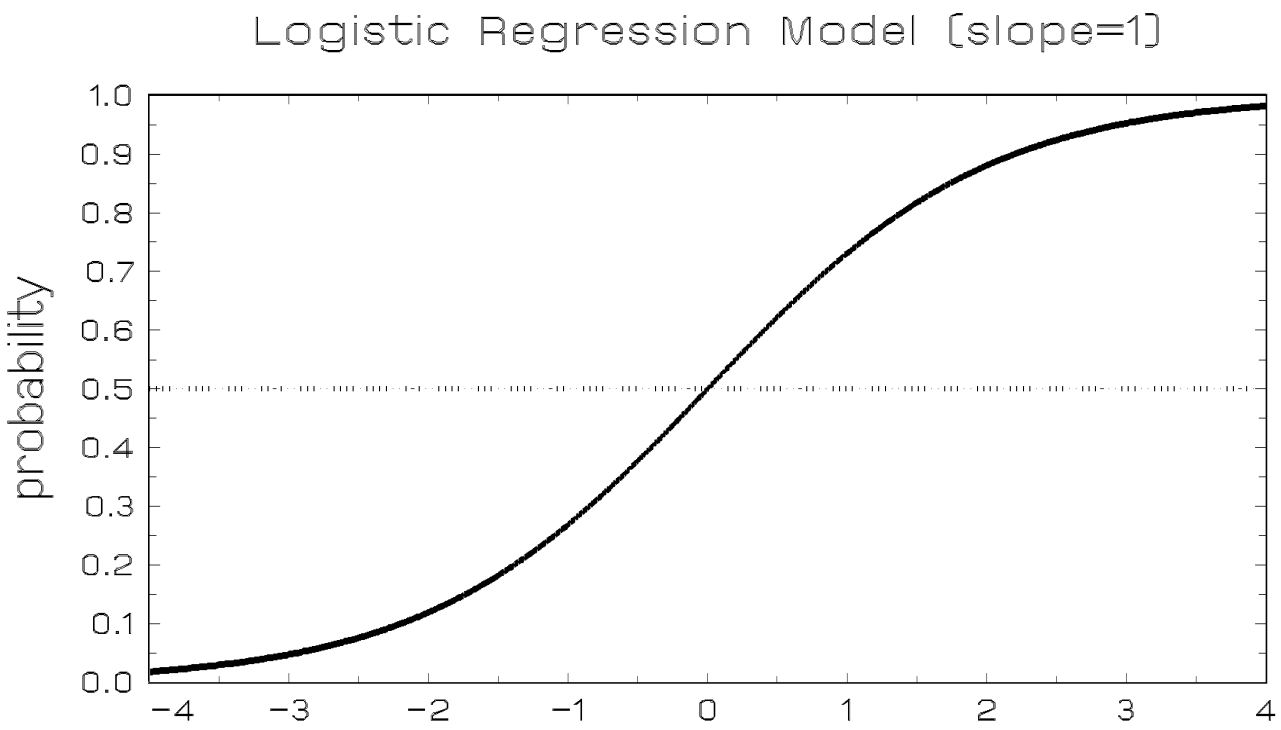
$$\log\left[\frac{P(Y_i = 1)}{1 - P(Y_i = 1)}\right] = \boldsymbol{x}_i'\boldsymbol{\beta}$$

- Dichotomous outcome ($Y = 0$ absence, $Y = 1$ presence).
- Function that links probabilities to regressors is the logit (or log odds) function $\log[P/(1 - P]$. Logit is called the link function.

The model can be written in terms of probabilities:

$$P(Y_i = 1) = \frac{1}{1 + \exp(-\boldsymbol{x}_i'\boldsymbol{\beta})}$$

- Model is a linear model for the logits, not for the probabilities. Logits can take on any values between negative and positive infinity, probabilities can only take on values between 0 and 1.


Logistic Regression Model (slope=1)
probability
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−4
−3
−2
−1
0
1
2
3
4

Logistic Regression Model (slope=1)
logit
4
3
2
1
0
−1
−2
−3
−4
−4
−3
−2
−1
0
1
2
3
4
X

The model can also be written in terms of the odds:

$$\frac{P(Y_i = 1)}{1 - P(Y_i = 1)} = \exp(\boldsymbol{x}_i' \boldsymbol{\beta})$$

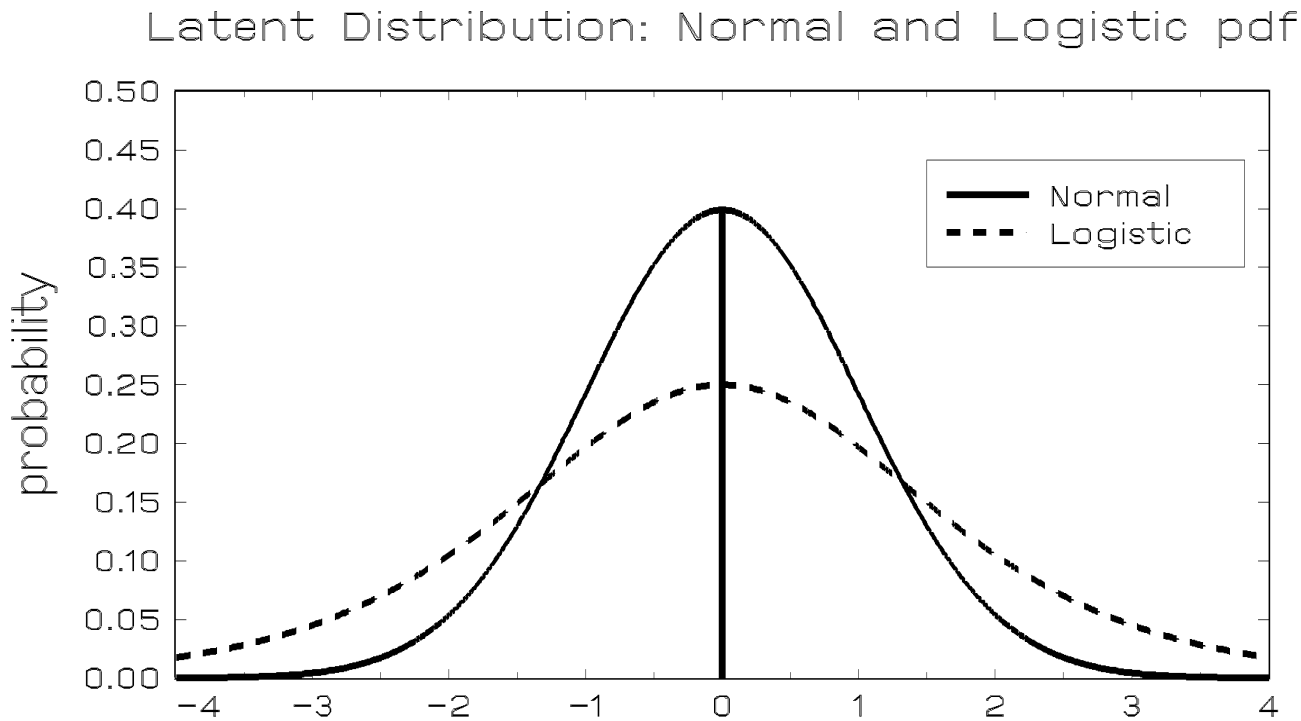
$\exp \beta =$ change in odds for Y per unit change of x

- $\beta = 0$ yields no effect on the odds
- $\beta > 0$ increases odds Y is present with increasing x
- $\beta < 0$ decreases odds Y is present with increasing x

Dichotomous Response and Threshold Concept

Continuous y_i - an unobservable latent variable - related to dichotomous response Y_i via "threshold concept"

Response occurs $(Y_i = 1)$ if $\gamma < y_i$
otherwise, a response does not occur $(Y_i = 0)$



The Threshold Concept in Practice

"How was your day?" (what is your satisfaction level today?)

- Satisfaction may be continuous, but we usually emit a

Great Day!

a day ...

Model for Latent Continuous Responses

Consider the model with p covariates for the latent response strength y_i $(i = 1, 2, \ldots, N)$:

$$y_i = \boldsymbol{x}_i' \boldsymbol{\beta} + \varepsilon_i$$

- probit: $\varepsilon_i \sim$ standard normal (mean=0, variance=1)
- logistic: $\varepsilon_i \sim$ standard logistic (mean=0, variance=$\pi^2/3$)

$\Rightarrow$ $\boldsymbol{\beta}$ estimates from logistic regression are larger (in abs. value) than from probit regression by approximately $\sqrt{\pi^2/3} = 1.8$

Underlying latent variable

- useful way of thinking of the problem
- not an essential assumption of the model

Random-intercept Logistic Regression Model

Consider the model with p covariates for the response Y_{ij} for subject j ($j = 1, 2, \ldots, n_i$) in cluster i ($i = 1, 2, \ldots, N$):

$$\log\left[\frac{P(Y_{ij} = 1)}{1 - P(Y_{ij} = 1)}\right] = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \upsilon_{0i}$$

where

$$\begin{aligned}
Y_{ij} &= \text{dichotomous response for subject } j \text{ in cluster } i \\
\boldsymbol{x}_{ij} &= (p+1) \times 1 \text{ covariate vector (includes 1 for intercept)} \\
\boldsymbol{\beta} &= (p+1) \times 1 \text{ vector of unknown parameters} \\
\upsilon_{0i} &= \text{cluster effects distributed } \mathcal{NID}(0, \sigma_\upsilon^2)
\end{aligned}$$

Characteristics of $\upsilon_{0i} \sim \mathcal{NID}(0, \sigma_\upsilon^2)$

- separates model from ususal (fixed-effects) multiple logistic regression model
- takes on $i = 1, 2, \ldots, N$ values
- assess impact of cluster i on individual outcome $logit_{ij}$, represents effect of subject clustering
- common for each cluster member, but changes for each cluster
- if $\upsilon_{0i} = 0$, then cluster has no effect for cluster i
- if $\upsilon_{0i} = 0$ for all clusters, cluster structure has no impact on individual data ($\sigma_\upsilon^2 = 0$)
  - no need for multilevel approach
  - ordinary logistic regression is OK
- if subject clustering has strong effect, estimates of $\upsilon_{0i} \neq 0$ and σ_υ^2 will increase from 0

Model for Latent Continuous Responses

Consider the model with p covariates for the $n_i \times 1$ latent response strength y_{ij}:

$$y_{ij} = \boldsymbol{x}_{ij}'\boldsymbol{\beta} + \upsilon_{0i} + \varepsilon_{ij}$$

where assuming

- $\varepsilon_{ij} \sim$ standard normal (mean 0 and $\sigma^2 = 1$) leads to multilevel probit regression
- $\varepsilon_{ij} \sim$ standard logistic (mean 0 and $\sigma^2 = \pi^2/3$) leads to multilevel logistic regression

Underlying latent variable

- not an essential assumption of the model
- useful for obtaining intra-class correlation (r)

$$r = \frac{\sigma_v^2}{\sigma_v^2 + \sigma^2}$$

and for design effect (d)

$$d = \frac{\sigma_v^2 + \sigma^2}{\sigma^2} = 1/(1 - r)$$

ratio of actual variance to the variance that would be obtained by simple random sampling (holding sample size constant)

Scaling of regression coefficients

Fixed-effects model

$\boldsymbol{\beta}$ estimates from logistic regression are larger (in abs. value) than from probit regression by approximately

$$\sqrt{\frac{\pi^2/3}{1}} = 1.8$$

because

- $V(y) = \sigma^2 = \pi^2/3$ for logistic
- $V(y) = \sigma^2 = 1$ for probit

Mixed-effects model

$\boldsymbol{\beta}$ estimates from mixed-effects model are larger (in abs. value) than from fixed-effects model by approximately

$$\sqrt{d} = \sqrt{\frac{\sigma_v^2 + \sigma^2}{\sigma^2}}$$

because

- $V(y) = \sigma_v^2 + \sigma^2$ in mixed-effects model
- $V(y) = \sigma^2$ in fixed-effects model

difference depends on size of random-effects variance σ_v^2

Within-Clusters / Between-Clusters models

<u>Within-clusters model</u> - level 1 $\quad (j = 1, \ldots, n_i)$

<u>observed response</u>

$$\log\left[\frac{P(Y_{ij}=1)}{1-P(Y_{ij}=1)}\right] = b_{0i} \; + \; b_{1i}\, Sex_{ij}$$

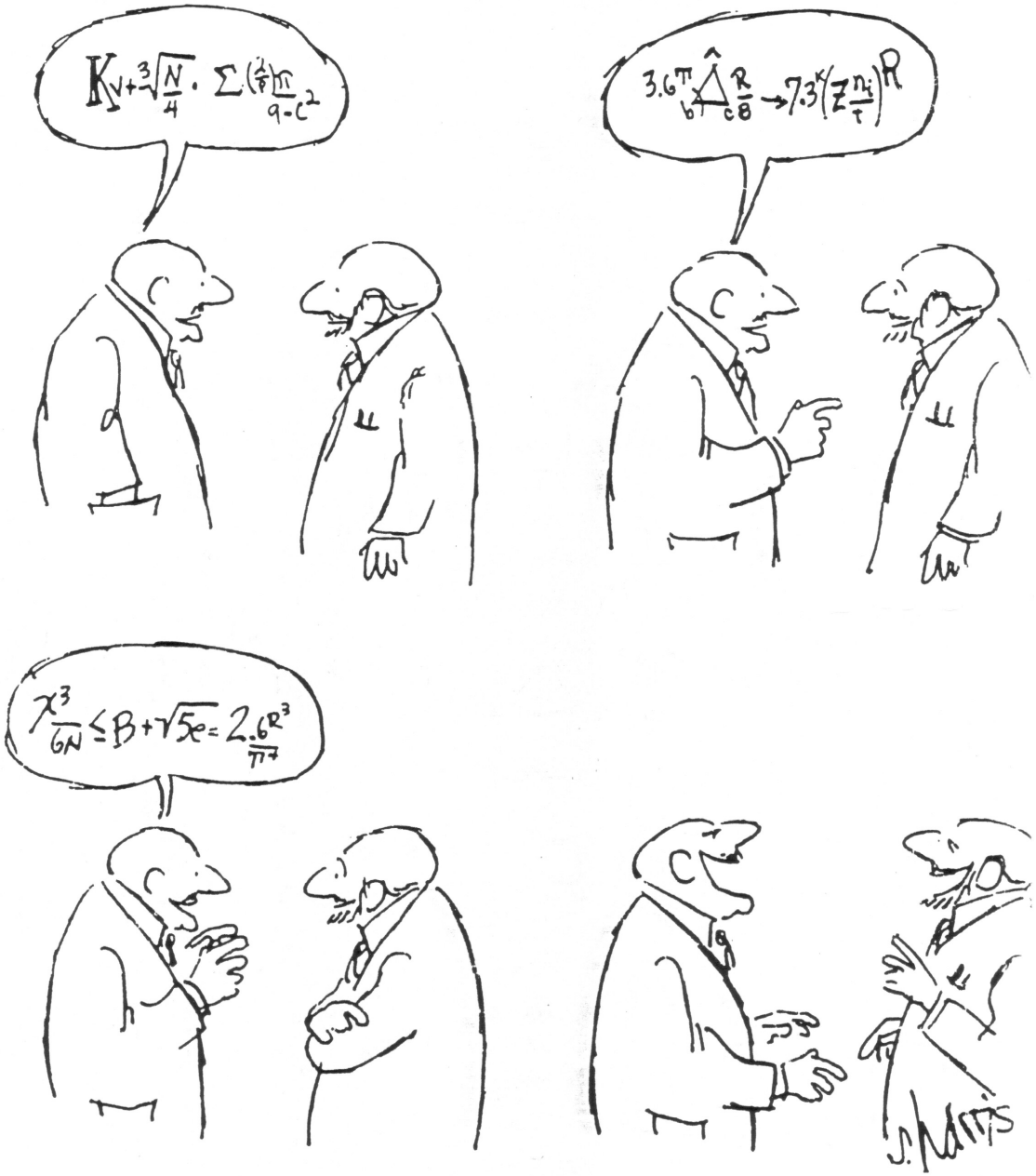
<u>latent response</u>

$$y_{ij} = b_{0i} \; + \; b_{1i}\, Sex_{ij} \; + \; \varepsilon_{ij}$$

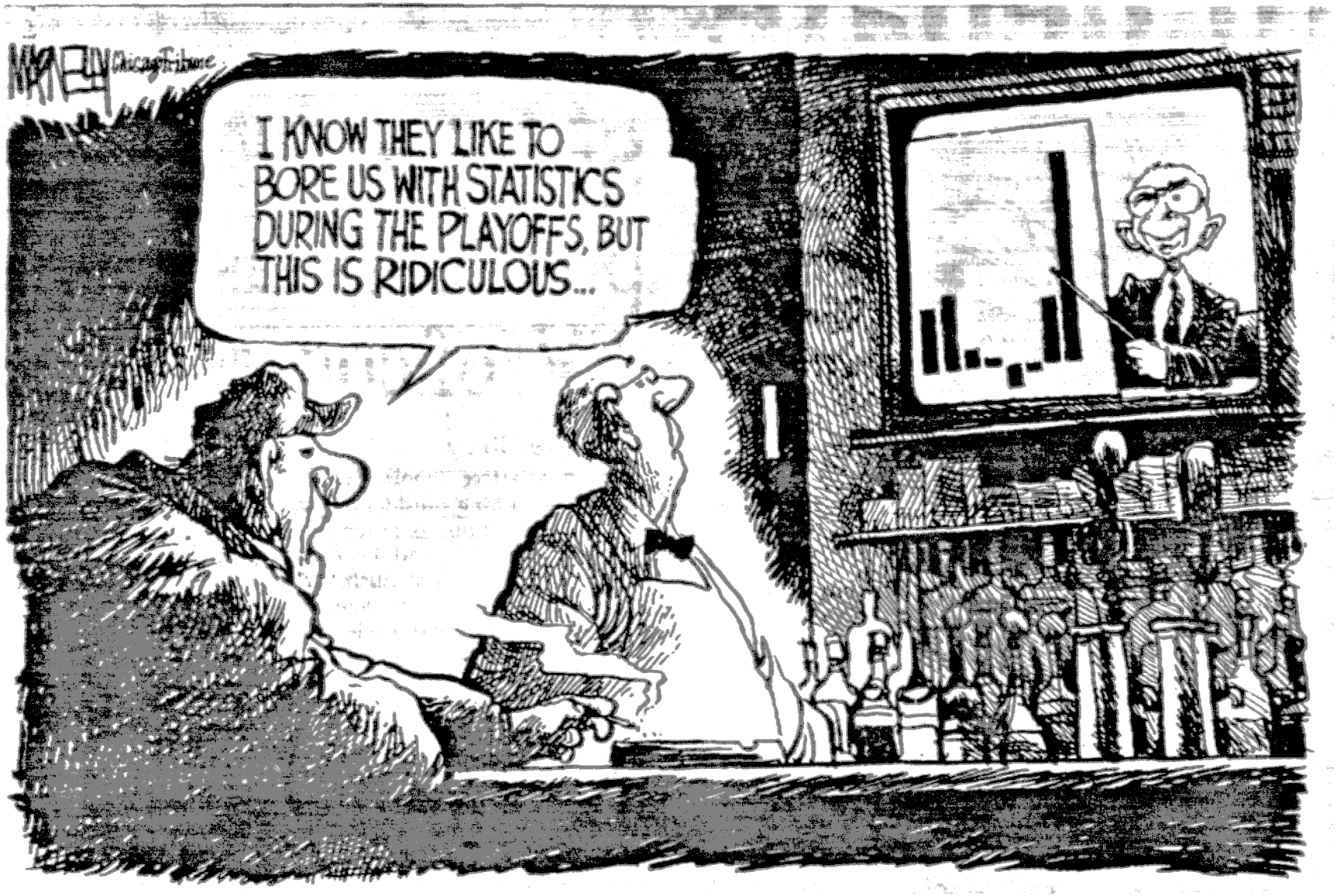
<u>Between-clusters model</u> - level 2 $\quad (i = 1, \ldots, N)$

$$b_{0i} = \beta_0 \; + \; \beta_2\, Grp_i \; + \; \upsilon_{0i}$$
$$b_{1i} = \beta_1 + \; \beta_3\, Grp_i$$

$$\upsilon_{0i} \sim \mathcal{NID}(0, \sigma_\upsilon^2) \quad \text{and} \quad \varepsilon_{ij} \sim \mathcal{LID}(0, \pi^2/3)$$



© 1973 *The New Yorker* Magazine.


MANLEY Chicago Tribune
I KNOW THEY LIKE TO BORE US WITH STATISTICS DURING THE PLAYOFFS, BUT THIS IS RIDICULOUS...

Effects of a School-based Intervention

The Television School and Family Smoking Prevention and Cessation Project (Flay, *et al.*, 1988); a subsample:

- *sample* - 1600 7th-graders - 135 classes - 28 schools
  - 1 to 13 classes per school, 2 to 28 students per class
- *outcome* - knowledge of the effects of tobacco use
- *timing* - students tested at pre and post-intervention
- *design* - schools exposed to
  - a social-resistance classroom curriculum (CC)
  - a media (television) intervention (TV)
  - CC combined with TV
  - a no-treatment control group

Main question of interest:

- Influence of the intervention on the tobacco health knowledge scores (THKS) ?

Challenges in the analysis:

- outcome variable (THKS) is number correct of 7 items
- controlling for intra-school and intra-class variability
- potential explanatory variables are at different levels

Tobacco and Health Knowledge Scale
Post-Intervention Scores ≥ 3 (out of 7)
Subgroup Descriptive Statistics

	CC = no		CC = yes	
	TV=no	TV=yes	TV=no	TV=yes
n	421	416	380	383
proportions	.416	.483	.632	.603
odds	.711	.935	1.714	1.520
logits	-.341	-.067	.539	.419

Within-Clusters / Between-Clusters components

<u>*Within-clusters model*</u> *- level 1* $(j = 1, \ldots, n_i \text{ subjects})$

$$logit_{ij} = b_{0i}$$

<u>*Between-clusters model*</u> *- level 2* $(i = 1, \ldots, N \text{ clusters})$

$$b_{0i} \;=\; \beta_0 \;+\; \beta_1 CC_i \;+\; \beta_2 TV_i \;+\; \beta_3 (CC_i \times TV_i) \;+\; \upsilon_{0i}$$

$$\upsilon_{0i} \sim \mathcal{NID}(0, \sigma_\upsilon^2)$$

β_0 = THKS logit for CC=no TV=no subgroup

β_1 = logit diff. between CC=yes vs CC=no (for TV=no)

$$b_{0i} = \beta_0 + (\beta_1 + \beta_3 TV_i)CC_i + \beta_2 TV_i + \upsilon_{0i}$$

β_2 = logit diff. between TV=yes vs TV=no (for CC=no)

$$b_{0i} = \beta_0 + (\beta_2 + \beta_3 CC_i)TV_i + \beta_1 CC_i + \upsilon_{0i}$$

β_3 = difference in logit attributable to interaction

υ_{0i} = random cluster deviation

note: interpretation depends on coding of variables, and βs are adjusted for the cluster effects (cluster-specific effects)

3-level model

Within-classrooms (and schools) model - *level 1*
($k = 1, \ldots, n_{ij}$ students)

$$logit_{ijk} = b_{0ij}$$

Between-classrooms (within-schools) model - *level 2*
($j = 1, \ldots, n_i$ classrooms)

$$b_{0ij} \;=\; b_{0i} \;+\; v_{0ij}$$

Between-schools model - *level 3* ($i = 1, \ldots, N$ schools)

$$b_{0i} \;=\; \beta_0 \;+\; \beta_1 CC_i \;+\; \beta_2 TV_i \;+\; \beta_3 (CC_i \times TV_i) \;+\; v_{0i}$$

$$v_{0ij} \sim \mathcal{NID}(0, \sigma^2_{v(2)}) \quad \text{and} \quad v_{0i} \sim \mathcal{NID}(0, \sigma^2_{v(3)})$$

$$
\begin{aligned}
\beta_0 &= \text{THKS logit for CC=no TV=no subgroup} \\
\beta_1 &= \text{logit diff. between CC=yes vs CC=no (for TV=no)} \\
\beta_2 &= \text{logit diff. between TV=yes vs TV=no (for CC=no)} \\
\beta_3 &= \text{difference in logit attributable to interaction} \\
\upsilon_{0ij} &= \text{random classroom deviation} \\
\upsilon_{0i} &= \text{random school deviation}
\end{aligned}
$$

SAS for multilevel analysis of dichotomous outcomes

PROC GLIMMIX (version 9.1.3 and thereafter)

- Multiple levels of nesting, crossed random effects
- Pseudo-likelihood estimation (by default)
  - Linearization to avoid integration over the random effects
  - Produces biased estimates if number of level-1 or level-2 units is small and/or ICC is large
- Full likelihood estimation using numerical quadrature for integration over the random effects (METHOD=QUAD)

PROC NLMIXED

- Full likelihood estimation using numerical quadrature for integration over the random effects
- Only for 2-level models; allows programming features

SAS Example: tvsfpb_GLIMMIX_NLMIXED.sas

```
DATA one; INFILE 'C:\data\tvsfpors.dat';
INPUT sid cid thkso thksb int thkspre cc tv cctv;

/* logistic regression ignoring clustering */
PROC LOGISTIC DESCENDING;
MODEL thksb = cc tv cctv;
RUN;

/* GLIMMIX: students in classrooms - Quasi-Like */
PROC GLIMMIX NOCLPRINT;
CLASS cid;
MODEL thksb (DESCENDING) = cc tv cctv /
DIST=BINARY SOLUTION;
RANDOM INTERCEPT / SUBJECT = cid TYPE=CHOL;
RUN;
```

```
/* GLIMMIX: students in classrooms - Full-Like */
PROC GLIMMIX NOCLPRINT METHOD=QUAD(QPOINTS=21);
CLASS cid;
MODEL thksb (DESCENDING) = cc tv cctv /
DIST=BINARY SOLUTION;
RANDOM INTERCEPT / SUBJECT = cid TYPE=CHOL;
RUN;

/* NLMIXED: students in classrooms - Full-Like */
PROC NLMIXED;
PARMS b0=-.34 b_cc=.88 b_tv=.27 b_cctv=-.39 sd=1;
z = b0 + b_cc*cc + b_tv*tv + b_cctv*cctv + sd*u;
IF (thksb=1) THEN p = 1/(1 + EXP(-z));
ELSE IF (thksb=0) THEN p = 1 - (1/(1 + EXP(-z)));
logl = LOG(p);
MODEL thksb ~ GENERAL(logl);
RANDOM u ~ NORMAL(0,1) SUBJECT=cid;
RUN;
```

```
/* GLIMMIX: 3-level - quasi-likelihood */
PROC GLIMMIX NOCLPRINT;
CLASS cid sid;
MODEL thksb (DESCENDING) = cc tv cctv /
DIST=BINARY SOLUTION;
RANDOM INTERCEPT / SUBJECT = cid(sid) TYPE=CHOL;
RANDOM INTERCEPT / SUBJECT = sid TYPE=CHOL;
RUN;

/* GLIMMIX: 3-level - full-likelihood */
PROC GLIMMIX NOCLPRINT METHOD=QUAD(QPOINTS=1);
CLASS cid sid;
MODEL thksb (DESCENDING) = cc tv cctv /
DIST=BINARY SOLUTION;
RANDOM INTERCEPT / SUBJECT = cid(sid) TYPE=CHOL;
RANDOM INTERCEPT / SUBJECT = sid TYPE=CHOL;
RUN;
```

THKS Post-Int (dichotomized) Scores - LR Estimates (std errs)

	Fixed	*GLIMMIX full*		*GLIMMIX quasi*	
intercept	-.341 ***	-.384 ***	-.391 ***	-.370 ***	-.383 ***
	(.099)	(.140)	(.190)	(.137)	(.204)
CC	.880 ***	.887 ***	.981 **	.856 **	.970 **
	(.145)	(.203)	(.277)	(.199)	(.293)
TV	.273 **	.232	.324	.225	.331
	(.139)	(.199)	(.268)	(.195)	(.286)
CC× TV	-.394 *	-.324	-.505	-.314	-.521
	(.204)	(.287)	(.387)	(.281)	(.409)
class sd		.524	.396	.513	.383
		(.083)	(.097)	(.093)	(.078)
school sd			.346		.402
			(.110)		(.115)
-2 log L	2162.53	2138.15	2134.43		

***$p < .01$ **$p < .05$ *$p < .10$ (Wald-tests not done for sds)

Calculation of ICC - 2 level model

$$r = \frac{\sigma_v^2}{\sigma_v^2 + \sigma^2}$$

Random classrooms model

$$r = \frac{.524^2}{.524^2 + \pi^2/3} = .077$$

$\Rightarrow$ 7.7% of the unexplained variation is at the classroom level

Calculation of ICC - 3 level model

Level-3 (likeness of students in the same school)

$$r = \frac{\sigma^2_{\upsilon(3)}}{\sigma^2_{\upsilon(3)} + \sigma^2_{\upsilon(2)} + \sigma^2} = \frac{.346^2}{.346^2 + .396^2 + \pi^2/3} = .034$$

Level-2 (likeness of students in same classroom & school)

$$r = \frac{\sigma^2_{\upsilon(3)} + \sigma^2_{\upsilon(2)}}{\sigma^2_{\upsilon(3)} + \sigma^2_{\upsilon(2)} + \sigma^2} = \frac{.346^2 + .396^2}{.346^2 + .396^2 + \pi^2/3} = .078$$

Level-2 (likeness of classes in the same school)

$$r = \frac{\sigma^2_{\upsilon(3)}}{\sigma^2_{\upsilon(3)} + \sigma^2_{\upsilon(2)}} = \frac{.346^2}{.346^2 + .396^2} = .433$$

- $r < .5$: the school level contributes slightly less to variability than the class level
- average classroom post THKS scores are moderately similar within schools

CC	TV	logistic $\Psi(z) = [1 + \exp(-z)]^{-1}$	estimate
Fixed-effects model			
0	0	$\Psi(-.341)$	.416
0	1	$\Psi(-.341 + .273)$	.483
1	0	$\Psi(-.341 + .880)$	.632
1	1	$\Psi(-.341 + .273 + .880 - .394)$	.603
Random-classrooms model		$\hat{d} = (.524^2 + \pi^2/3)/(\pi^2/3)$	
0	0	$\Psi((-.384)/\sqrt{\hat{d}})$	.409
0	1	$\Psi((-.384 + .232)/\sqrt{\hat{d}})$	.464
1	0	$\Psi((-.384 + .887)/\sqrt{\hat{d}})$	.619
1	1	$\Psi((-.384 + .232 + .887 - .324)/\sqrt{\hat{d}})$	.597
3-level model		$\hat{d} = (.346^2 + .396^2 + \pi^2/3)/(\pi^2/3)$	
0	0	$\Psi((-.391)/\sqrt{\hat{d}})$	.407
0	1	$\Psi((-.391 + .324)/\sqrt{\hat{d}})$	.484
1	0	$\Psi((-.391 + .981)/\sqrt{\hat{d}})$	.638
1	1	$\Psi((-.391 + .324 + .981 - .505)/\sqrt{\hat{d}})$	.597

d = design effect $= (\sigma_\upsilon^2 + \sigma^2)/\sigma^2$ or $= (\sigma_{\upsilon(3)}^2 + \sigma_{\upsilon(2)}^2 + \sigma^2)/\sigma^2$

Within-Clusters / Between-Clusters components

<u>*Within-clusters model*</u> *- level 1* $(j = 1, \ldots, n_i \text{ subjects})$

$$logit_{ij} = b_{0i} + b_{1i} PRETHKS_{ij}$$

<u>*Between-clusters model*</u> *- level 2* $(i = 1, \ldots, N \text{ clusters})$

$$\begin{aligned} b_{0i} &= \beta_0 + \beta_2 CC_i + \beta_3 TV_i + \beta_4 (CC_i \times TV_i) + \upsilon_{0i} \\ b_{1i} &= \beta_1 \end{aligned}$$

$$\upsilon_{0i} \sim \mathcal{NID}(0, \sigma_\upsilon^2)$$

β_0 = (PRETHKS adjusted) logit for CC=no TV=no subgroup

β_1 = effect of PRETHKS on POSTTHKS

β_2 = (PRETHKS adjusted) logit diff. between CC=yes vs CC=no (for TV=no)

β_3 = (PRETHKS adjusted) logit diff. between TV=yes vs TV=no (for CC=no)

β_4 = (PRETHKS adjusted) difference in logit attributable to interaction

υ_{0i} = random cluster deviation

3-level model

Within-classrooms (and schools) model - *level 1*
($k = 1, \ldots, n_{ij}$ students)

$$logit_{ijk} = b_{0ij} + b_{1ij} PRETHKS_{ijk}$$

Between-classrooms (within-schools) model - *level 2*
($j = 1, \ldots, n_i$ classrooms)

$$\begin{aligned} b_{0ij} &= b_{0i} + \upsilon_{0ij} \\ b_{1ij} &= b_{1i} \end{aligned}$$

Between-schools model - *level 3* ($i = 1, \ldots, N$ schools)

$$\begin{aligned} b_{0i} &= \beta_0 + \beta_2 CC_i + \beta_3 TV_i + \beta_4 (CC_i \times TV_i) + \upsilon_{0i} \\ b_{1i} &= \beta_1 \end{aligned}$$

$$\upsilon_{0ij} \sim \mathcal{NID}(0, \sigma^2_{\upsilon(2)}) \quad \text{and} \quad \upsilon_{0i} \sim \mathcal{NID}(0, \sigma^2_{\upsilon(3)})$$

THKS Post-Int (dichotomized) Scores - LR Estimates (std err)

	Fixed		*GLIMMIX full*				*GLIMMIX quasi*	
intercept	-1.217	***	-1.253	***	-1.246	***	-1.217 ***	-1.212 ***
	(.140)		(.170)		(.195)		(.167)	(.207)
PRETHKS	.400	***	.401	***	.395	***	.390 ***	.383 ***
	(.043)		(.046)		(.046)		(.045)	(.046)
CC	.973	***	.988	***	1.039	***	.959 ***	1.032 ***
	(.151)		(.197)		(.243)		(.195)	(.261)
TV	.316	**	.287		.334		.279	.343
	(.144)		(.192)		(.234)		(.184)	(.253)
CC× TV	-.413	**	-.369		-.467		-.358	-.490
	(.219)		(.277)		(.340)		(.274)	(.363)
class sd			.468		.394		.464	.379
			(.086)		(.099)		(.081)	(.095)
school sd					.250			.317
					(.121)			(.117)
-2 log L	2073.3		2057.18		2056.25			

$^{***}p < .01$ $^{**}p < .05$ $^{*}p < .10$ (Wald-tests not done for sds)

Calculation of ICC - 2 level models

$$r = \frac{\sigma_v^2}{\sigma_v^2 + \sigma^2}$$

Random classrooms model

$$r = \frac{.468^2}{.468^2 + \pi^2/3} = .062$$

$\Rightarrow$ 6.2% of the unexplained variation is at the classroom level

Calculation of ICC - 3 level model

Level-3 (likeness of students in the same school)

$$r = \frac{\sigma^2_{\upsilon(3)}}{\sigma^2_{\upsilon(3)} + \sigma^2_{\upsilon(2)} + \sigma^2} = \frac{.250^2}{.250^2 + .394^2 + \pi^2/3} = .018$$

Level-2 (likeness of students in same classroom & school)

$$r = \frac{\sigma^2_{\upsilon(3)} + \sigma^2_{\upsilon(2)}}{\sigma^2_{\upsilon(3)} + \sigma^2_{\upsilon(2)} + \sigma^2} = \frac{.250^2 + .394^2}{.250^2 + .394^2 + \pi^2/3} = .062$$

Level-2 (likeness of classes in the same school)

$$r = \frac{\sigma^2_{\upsilon(3)}}{\sigma^2_{\upsilon(3)} + \sigma^2_{\upsilon(2)}} = \frac{.250^2}{.250^2 + .394^2} = .287$$

- $r < .5$: the school level contributes less to variability than the class level
- average classroom post THKS scores are moderately similar within schools

CC	TV	logistic $\Psi(z) = [1+\exp(-z)]^{-1}$	estimate
Fixed-effects model			
0	0	$\Psi(-1.217 + 2.152 \times .400)$	.412
0	1	$\Psi(-1.217 + .316 + 2.087 \times .400)$	.483
1	0	$\Psi(-1.217 + .973 + 2.050 \times .400)$	.640
1	1	$\Psi(-1.217 + .316 + .973 - .413 + 1.979 \times .400)$	.610
Random-classrooms model		$\hat{d} = (.468^2 + \pi^2/3)/(\pi^2/3)$	
0	0	$\Psi((-1.253 + 2.152 \times .401)/\sqrt{\hat{d}})$	.407
0	1	$\Psi((-1.253 + .287 + 2.087 \times .401)/\sqrt{\hat{d}})$	.469
1	0	$\Psi((-1.253 + .988 + 2.050 \times .401)/\sqrt{\hat{d}})$	.632
1	1	$\Psi((-1.253 + .287 + .988 - .369 + 1.979 \times .401)/\sqrt{\hat{d}})$	.606
3-level model		$\hat{d} = (.250^2 + .394^2 + \pi^2/3)/(\pi^2/3)$	
0	0	$\Psi((-1.246 + 2.152 \times .395)/\sqrt{\hat{d}})$	.405
0	1	$\Psi((-1.246 + .334 + 2.087 \times .395)/\sqrt{\hat{d}})$	.479
1	0	$\Psi((-1.246 + 1.039 + 2.050 \times .395)/\sqrt{\hat{d}})$	.642
1	1	$\Psi((-1.246 + .334 + 1.039 - .467 + 1.979 \times .395)/\sqrt{\hat{d}})$	.605

d = design effect $= (\sigma_\upsilon^2 + \sigma^2)/\sigma^2$ or $= (\sigma_{\upsilon(3)}^2 + \sigma_{\upsilon(2)}^2 + \sigma^2)/\sigma^2$

Summary

- Mixed logistic regression model is direct extension of ordinary logistic regression
- Useful approach for multilevel data
- Software is available in SAS (and other programs)
- (Extended) methods are available for ordinal, nominal, count outcomes
- Similar models can be used for longitudinal, albeit more issues
  - more random effects are typically necessary
  - missing data and attrition