Multilevel Logistic Regression with Dyadic Data

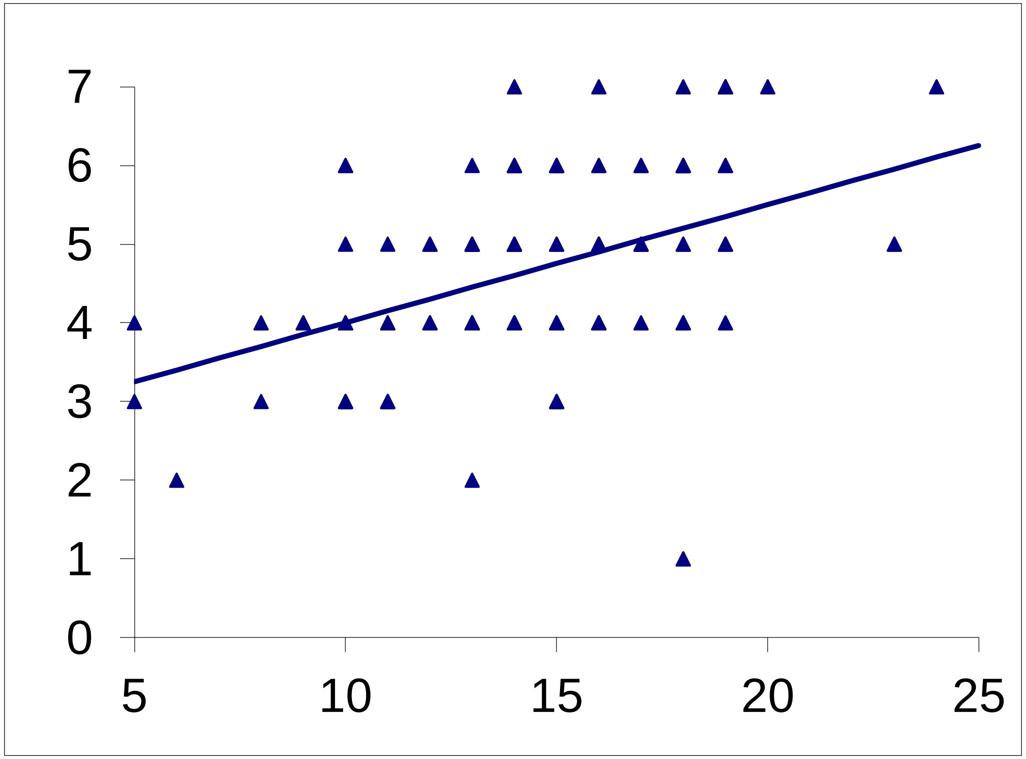
Scott Liening

(11/8/2011)

Logistic Regression

- Linear regression: $y = \beta_0 + \beta_1 x$

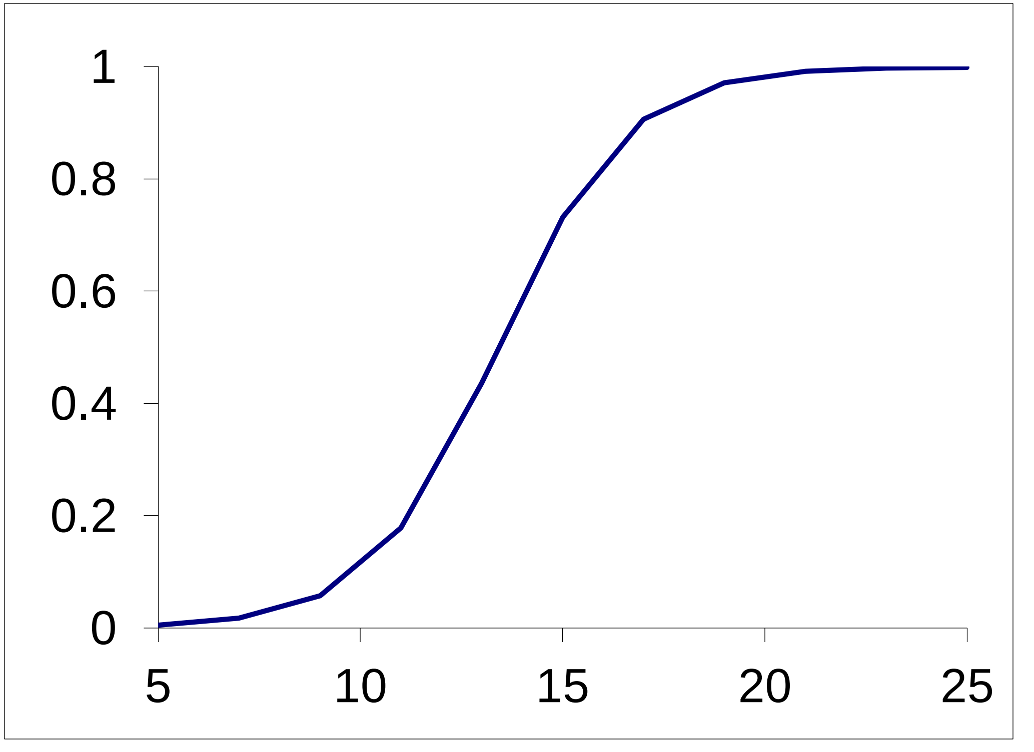
$$y = 2.50 + 0.15x$$



Logistic Regression

- Logistic regression: $y = {}^{1}\!/_{1+e^{-(\beta_0+\beta_1 x)}}$

$$y = {}^{1}\!/_{1+e^{-(-8.50+0.63x)}}$$



Multilevel Models

- Individuals within clusters: $y = \alpha_{ji} + \beta_{ji}x + \epsilon_{ji}$
  - Students within classrooms
  - Dyads (e.g., spouses within couples)
  - Assumes continuous outcomes

Multilevel Logistic Regression

- Individuals within clusters: $y = logit^{-1}(\alpha_{ji} + \beta_{ji}x + \epsilon_{ji})$
  - Students within classrooms
  - Dyads (e.g., spouses within couples)
  - Dichotomous outcomes
    - Probability of an event happening

Dyadic Data

- Violations of independence assumption
  - One member's value often depends on the other member's values
  - Underestimation of standard errors, increased Type I Errors

Example

- DV
  - Task choice
- IVs
  - Status condition, Testosterone, Cortisol
- N = 115 (61 dyads)

Example

- Actor-partner interdependence model (APIM)
  - Structure data such that each case is both an actor (e.g., that individual's values) and a partner (e.g., the other half of the dyad's values)

N	Dyad	ZT1	ZTChange	CChange	LCort1	shDOM	Choice	Choice2	Cond	pZT1	pZTChang	pCChange
101	1	-1.48712	-0.23054	-0.04755	-0.92	16	1	0	0	-1.77243	0.76113	0.08666
102	1	-1.77243	0.76113	0.08666	-0.76	9	1	0	1	-1.48712	-0.23054	-0.04755
103	2	0.07437	-1.00207	-0.01384	-0.76	19	0	1	1			
105	3	-1.05619	0.15298	0.02725	-1.18	11	1	0	1	-1.5535	-0.16534	-0.04174
106	3	-1.5535	-0.16534	-0.04174	-0.7	17	1	0	0	-1.05619	0.15298	0.02725
109	4	0.42356	0.43281	0.07321	-0.49	15	1	0	0	-0.95282	0.08938	-0.03082
110	4	-0.95282	0.08938	-0.03082	-0.24	10	0	1	1	0.42356	0.43281	0.07321
117	5	-0.41985	-0.79003	-0.00384	-0.51	16	0	1	1	0.13386	-0.75519	-0.00886
118	5	0.13386	-0.75519	-0.00886	-0.77	18	0	1	0	-0.41985	-0.79003	-0.00384
119	6	-0.54025	0.21357	0.16534	-0.45	15	0	1	0	-0.45146	-0.17406	-0.00908
120	6	-0.45146	-0.17406	-0.00908	-0.95	8	0	1	1	-0.54025	0.21357	0.16534

Example

- Test for within cluster dependence
  - Pairwise intraclass correlation coefficient (PICC)

Statistics for Table of Choice by pChoice

Statistic	Value	ASE	95% Confidence Limits	
Gamma	0.0704	0.3229	-0.5625	0.7034
Kendall's Tau-b	0.0324	0.1498	-0.2612	0.3260
Stuart's Tau-c	0.0296	0.1370	-0.2389	0.2981
Somers' D C¦R	0.0315	0.1457	-0.2540	0.3170
Somers' D R¦C	0.0333	0.1541	-0.2687	0.3353
Pearson Correlation	0.0324	0.1498	-0.2612	0.3260
Spearman Correlation	0.0324	0.1498	-0.2612	0.3260
Lambda Asymmetric C¦R	0.0000	0.0000	0.0000	0.0000
Lambda Asymmetric R¦C	0.0000	0.0000	0.0000	0.0000
Lambda Symmetric	0.0000	0.0000	0.0000	0.0000
Uncertainty Coefficient C¦R	0.0008	0.0076	0.0000	0.0157
Uncertainty Coefficient R¦C	0.0008	0.0073	0.0000	0.0151
Uncertainty Coefficient Symmetric	0.0008	0.0074	0.0000	0.0154

Example

- Establishing starting parameter values

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)	Dyad	0.02153
Residual		0.2029

Between dyad variance found via mixed linear model procedure

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	-1.2585	0.8712	-2.9662	0.4491	-1.44	0.1486
Cond	-0.3763	0.4159	-1.1915	0.4389	-0.90	0.3656
ZT1	-0.2416	0.2944	-0.8185	0.3353	-0.82	0.4118
LCort1	-1.0164	1.0994	-3.1711	1.1383	-0.92	0.3552

Beta estimates found via generalized linear model procedure's generalized estimating equations formulation

Example

- Running the multilevel logistic model (MLLM)

Starting parameter values

```
proc mixed data = dele method = reml;
class dyad;
model choice = cond zt1 lcort1 /solution;
random intercept /subject = dyad type = un;
run;
```

```
proc genmod data = dele descending;
class dyad;
model choice = cond zt1 lcort1 /dist = bin link = logit;
repeated subject = dyad /type = un;
run;
```

MLLM

```
proc nlmixed data = dele qpoints = 20 tech = newrap;
parms beta0 = -1.2585 beta1 = -0.3763 beta2 = -0.2416 beta3 = -1.0164 s2u = 0.02153;
eta = beta0 + beta1*cond + beta2*zt1 + beta3*lcort1 + u;
mu = exp(eta)/(1+exp(eta));
model choice ~binary(mu);
random u ~normal(0, s2u) subject = dyad;
run;
```

Example

- Results of the MLLM

Overall model fit

Fit Statistics

-2 Log Likelihood	118.7
AIC (smaller is better)	128.7
AICC (smaller is better)	129.4

Parameter Estimates

	Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
	beta0	-1.3724	1.0630	57	-1.29	0.2019	0.05	-3.5011	0.7563	-1.08E-7
Condition	beta1	-0.4100	0.4718	57	-0.87	0.3884	0.05	-1.3547	0.5347	-6.23E-8
Testosterone	beta2	-0.2576	0.2658	57	-0.97	0.3367	0.05	-0.7899	0.2748	1.876E-8
Cortisol	beta3	-1.1066	1.3622	57	-0.81	0.4200	0.05	-3.8344	1.6212	7.179E-8
	s2u	0.4341	1.0095	57	0.43	0.6688	0.05	-1.5874	2.4556	-1.89E-7

Individual predictors

References

- Der, G. & Everitt, B.S. (2009). *A handbook of statistical analysis using SAS* (3rd Ed.). Boca Raton, FL: CRC Press.

- Gelman, A. & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* New York, NY: Cambridge University Press.

- McMahon, J.M., Pouget, E.R., & Tortu, S. (2006). A guide for multilevel modeling of dyadic data with binary outcomes using SAS PROC NLMIXED. *Computational Statistics and Data Analysis, 50 (12),* 3663-3680.