Lecture 1
Introduction to Multi-level Models

Course Website:
http://www.biostat.jhsph.edu/~ejohnson/multilevel.htm

All lecture materials extracted and further developed from the Multilevel Model course taught by Francesca Dominici:

http://www.biostat.jhsph.edu/~fdominic/teaching/bio656/ml.html

1

Statistical Background on MLMs

- ✓ Main Ideas
- ✓ Accounting for Within-Cluster Associations
- ✓ Marginal & Conditional Models
- ✓ A Simple Example
- ✓ Key MLM components

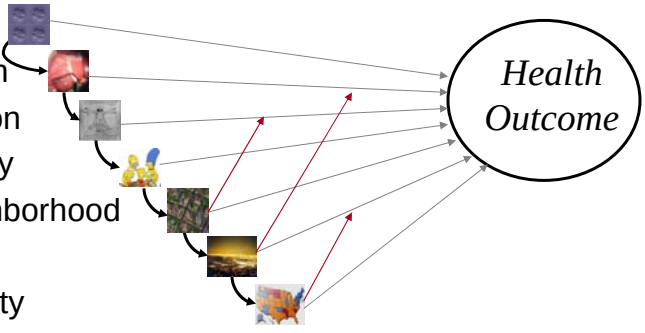
2

The Main Idea…

3

Multi-level Models – Main Idea

- Biological, psychological and social processes that influence health occur at many **levels**:
  - Cell
  - Organ
  - Person
  - Family
  - Neighborhood
  - City
  - Society



- An analysis of risk factors should consider:
  - Each of these levels
  - Their interactions

4

Example: Alcohol Abuse

Level:		
1.	Cell:	Neurochemistry
2.	Organ:	Ability to metabolize ethanol
3.	Person:	Genetic susceptibility to addiction
4.	Family:	Alcohol abuse in the home
5.	Neighborhood:	Availability of bars
6.	Society:	Regulations; organizations; social norms

5

Example: Alcohol Abuse; Interactions between Levels

Level:	
5	Availability of bars *and*
6	State laws about drunk driving
4	Alcohol abuse in the family *and*
2	Person's ability to metabolize ethanol
3	Genetic predisposition to addiction *and*
4	Household environment
6	State regulations about intoxication *and*
3	Job requirements

6

Notation:

Person: *sijk*

Outcome: Y_{sijk}

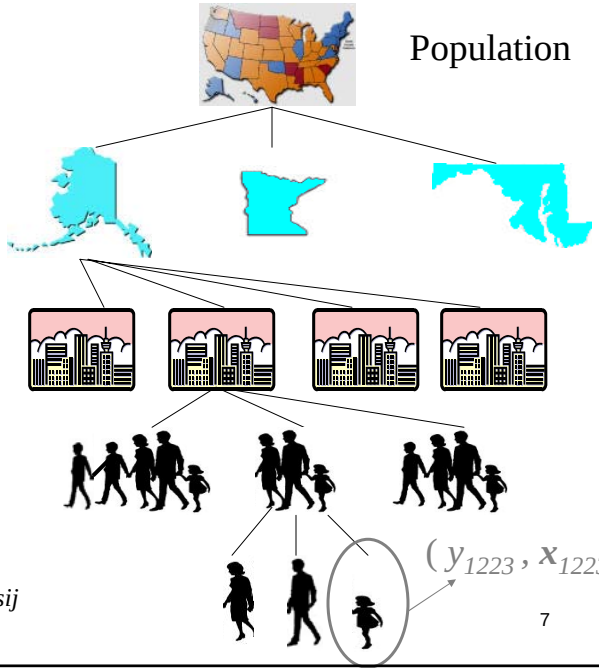
Predictors: $\boldsymbol{X}_{sijk}$

State: $s=1,\ldots,S$

Neighborhood: $i=1,\ldots,I_s$

Family: $j=1,\ldots,J_{si}$

Person: $k=1,\ldots,K_{sij}$



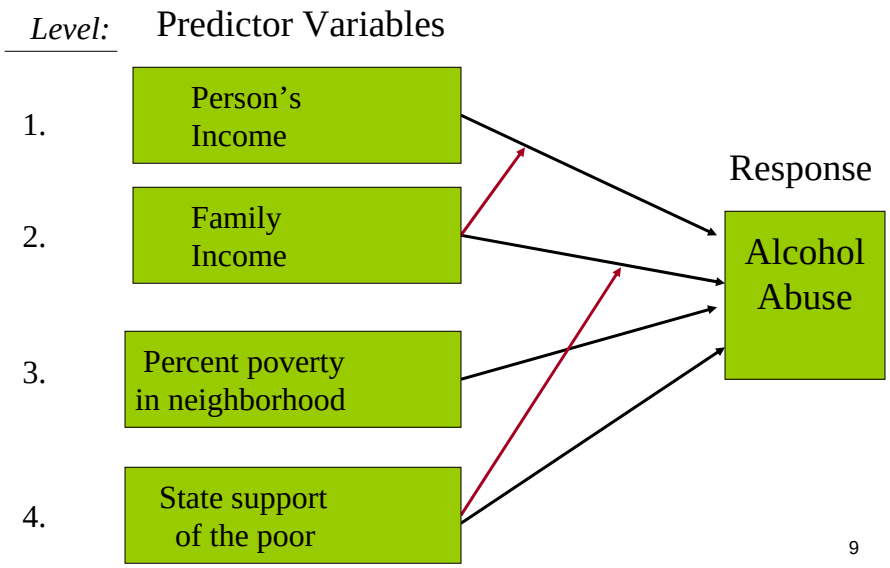
7

Notation (cont.)

- (y_{sijk}, x_{sijk}) are (response, predictors) for
  - person $k = 1, ..., K_{sij}$ in
  - family $j = 1, ..., J_{si}$ in
  - neighborhood $i = 1, ..., I_s$ in
  - state $s = 1, ..., S$
- $\mu_{sijk} = \mathsf{E}(y_{sijk}|x_{sijk})$

8

Multi-level Models: Idea



9

A Rose is a Rose is a…

- Multi-level model
- Random effects model
- Mixed model
- Random coefficient model
- Hierarchical model
- Meta-analysis (in some cases)

Many names for similar models, analyses, and goals.

10

Digression on Statistical Models

- A statistical model is an approximation to reality
- There is not a “correct” model;
  - (forget the holy grail)
- A model is a tool for asking a scientific question;
  - (screw-driver vs. sludge-hammer)
- A useful model combines the data with prior information to address the question of interest.
- Many models are better than one.

11

Generalized Linear Models (GLMs)

$g(\mu) = \beta_0 + \beta_1 * X_1 + \ldots + \beta_p * X_p$

(μ = E(Y|X) = mean)

Model	**Response**	$g(\mu)$	**Distribution**	**Coef Interp**
Linear	Continuous (ounces)	μ	Gaussian	Change in avg(Y) per unit change in X
Logistic	Binary (disease)	$\log\left(\frac{\mu}{(1-\mu)}\right)$	Binomial	Log Odds Ratio
Log-linear	Count/Times to events	$\log(\mu)$	Poisson	Log Relative Risk

12

Generalized Linear Models (GLMs)

$g(\mu) = \beta_0 + \beta_1 * X_1 + \ldots + \beta_p * X_p$

Example: Age & Gender

Gaussian – Linear: $E(y) = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Gender}$

β_1 = Change in Average Response per 1 unit increase in Age, Comparing people of the SAME GENDER.

WHY?

Since: $E(y|\text{Age}+1,\text{Gender}) = \beta_0 + \beta_1(\text{Age}+1) + \beta_2 \text{Gender}$

And: $E(y|\text{Age},\text{Gender}) = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Gender}$

$\Delta E(y) = \beta_1$

13

Generalized Linear Models (GLMs)

$g(\mu) = \beta_0 + \beta_1 * X_1 + \ldots + \beta_p * X_p$

Example: Age & Gender

Binary – Logistic: $\log\{\text{odds}(Y)\} = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Gender}$

β_1 = log-OR of "+ Response" for a 1 unit increase in Age, Comparing people of the SAME GENDER.

WHY?

Since: $\log\{\text{odds}(y|\text{Age}+1,\text{Gender})\} = \beta_0 + \beta_1(\text{Age}+1) + \beta_2 \text{Gender}$

And: $\log\{\text{odds}(y|\text{Age},\text{Gender})\} = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Gender}$

Δ log-Odds $= \beta_1$

➡ log-OR $= \beta_1$

14

Generalized Linear Models (GLMs)

$g(\mu) = \beta_0 + \beta_1 * X_1 + \ldots + \beta_p * X_p$

Example: Age & Gender

Counts – Log-linear: $\log\{E(Y)\} = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Gender}$

β_1 = log-RR for a 1 unit increase in Age,
Comparing people of the SAME GENDER.

WHY?

Self-Check: Verify Tonight

15

"Quiz": Most Important Assumptions of Regression Analysis?

A. Data follow normal distribution

B. All the key covariates are included in the model

C. Xs are fixed and known

D. Responses are independent

16

Non-independent responses
(Within-Cluster Correlation)

- Fact: two responses from the same family tend to be more like one another than two observations from different families
- Fact: two observations from the same neighborhood tend to be more like one another than two observations from different neighborhoods
- Why?

17

Why? (Family Wealth Example)

Great-Grandparents → Grandparents → Parents → You

GOD → Great-Grandparents, Grandparents, Parents, You

18

Key Components of Multi-level Models

- Specification of predictor variables from multiple levels **(Fixed Effects)**
  - Variables to include
  - Key interactions
- Specification of correlation among responses from same clusters **(Random Effects)**
- Choices must be driven by scientific understanding, the research question and empirical evidence.

19

Correlated Data…
(within-cluster associations)

20

Multi-level analyses

- Multi-level analyses of social/behavioral phenomena: an important idea
- Multi-level models involve predictors from multi-levels and their interactions
- They must account for associations among observations within clusters (levels) to make efficient and valid inferences.

21

Regression with Correlated Data

Must take account of correlation to:

- Obtain valid inferences
  - standard errors
  - confidence intervals
- Make efficient inferences

22

Logistic Regression Example: Cross-over trial

- Response: 1-normal; 0- alcohol dependence
- Predictors: period (x_1); treatment group (x_2)
- Two observations per person (cluster)
- Parameter of interest: log odds ratio of alcohol dependence: placebo vs. treatment

Mean Model: $\log\{\text{odds(AD)}\} = \beta_0 + \beta_1 \text{Period} + \beta_2 \text{Placebo}$

23

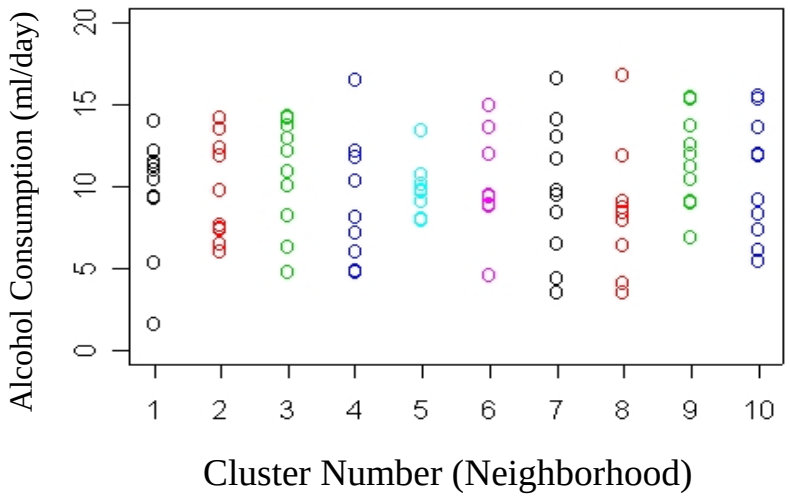
Results: estimate (standard error)

	Model	
Variable	Ordinary Logistic Regression	Account for correlation
Intercept (β_0)	0.66 (0.32)	0.67 (0.29)
Period (β_1)	-0.27 (0.38)	-0.30 (0.23)
Placebo (β_2)	0.56 (0.38)	0.57 (0.23)

Similar Estimates,

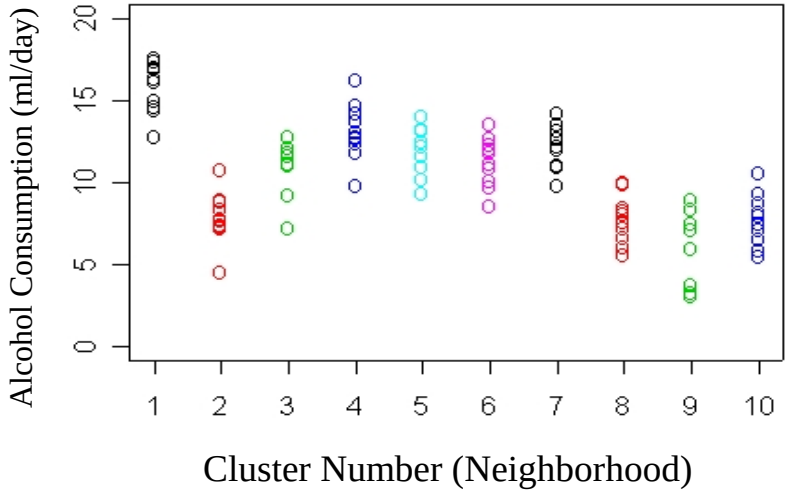
WRONG Standard Errors (& Inferences) for OLR 24

Simulated Data: Non-Clustered



25

Simulated Data: Clustered



26

Within-Cluster Correlation

- Correlation of two observations from same cluster =

$$\frac{\text{Tot Var - Var Within}}{\text{Tot Var}}$$

- Non-Clustered = (9.8-9.8) / 9.8 = 0
- Clustered = (9.8-3.2) / 9.8 = 0.67

27

Models for Clustered Data

- Models are tools for inference
- Choice of model determined by scientific question
- Scientific Target for inference?
  - *Marginal* mean:
    - Average response across the population
  - *Conditional* mean:
    - Given other responses in the cluster(s)
    - Given unobserved random effects
- We will deal mainly with conditional models
  (but we'll mention some important differences)

28

Marginal vs Conditional Models…

29

Marginal Models

- Focus is on the “mean model”: E(Y|X)
- Group comparisons are of main interest, i.e. neighborhoods with high alcohol use vs. neighborhoods with low alcohol use
- Within-cluster associations are accounted for to correct standard errors, but are not of main interest.

$$\log\{ \text{odds(AD)} \} = \beta_0 + \beta_1 \text{Period} + \beta_2 \text{Placebo}$$

30

Marginal Model Interpretations

- $\log\{ \text{odds(AD)} \} = \beta_0 + \beta_1 \text{Period} + \beta_2 \text{Placebo}$

  $= 0.67 + (-0.30)\text{Period} + (0.57)\text{Placebo}$

TRT Effect: (placebo vs. trt)

OR = exp(0.57) = 1.77, 95% CI (1.12, 2.80)

Risk of Alcohol Dependence is almost twice as high on placebo, regardless of, (adjusting for), time period

WHY?

Since: $\log\{\text{odds(AD|Period, placebo)}\} = \beta_0 + \beta_1 \text{Period} + \beta_2$

And: $\log\{\text{odds(AD|Period, trt)}\} = \beta_0 + \beta_1 \text{Period}$

Δ log-Odds $= \beta_2$

OR $=$ exp(β_2)

31

Random Effects Models

- Conditional on unobserved latent variables or "random effects"
  - Alcohol use within a family is related because family members share an unobserved "family effect": common genes, diets, family culture and other unmeasured factors
  - Repeated observations within a neighborhood are correlated because neighbors share: common traditions, access to services, stress levels,…
  - $\log\{ \text{odds(AD)} \} = b_i + \beta_0 + \beta_1 \text{Period} + \beta_2 \text{Placebo}$

32

Random Effects Model Interpretations

WHY?

Since: $\log\{\text{odds}(AD_i|\text{Period}, \text{Placebo}, b_i)\} = \beta_0 + \beta_1\text{Period} + \beta_2 + b_i$

And: $\log\{\text{odds}(AD_i|\text{Period}, \text{TRT}, b_i)\} = \beta_0 + \beta_1\text{Period} + b_i$

Δ log-Odds = β_2

➡ OR = $\exp(\beta_2)$

- In order to make comparisons we must keep the subject-specific latent effect (b_i) the same.
- In a Cross-Over trial we have outcome data for each subject on both placebo & treatment
- In other study designs we may not.

33

Marginal vs. Random Effects Models

- For linear models, regression coefficients in random effects models and marginal models are identical:

  average of linear function = linear function of average

- For non-linear models, (logistic, log-linear,…) coefficients have different meanings/values, and address different questions
  - Marginal models -> *population-average* parameters
  - Random effects models -> cluster-*specific* parameters

34

Marginal -vs- Random Intercept Models; Cross-over Example

	Model		
Variable	Ordinary Logistic Regression	Marginal (GEE) Logistic Regression	Random-Effect Logistic Regression
Intercept	0.66 (0.32)	0.67 (0.29)	2.2 (1.0)
Period	-0.27 (0.38)	-0.30 (0.23)	-1.0 (0.84)
Placebo	0.56 (0.38)	0.57 (0.23)	1.8 (0.93)
Log OR (assoc.)	0.0	3.56 (0.81)	5.0 (2.3)

35

Comparison of Marginal and Random Effect Logistic Regressions

- Regression coefficients in the random effects model are roughly 3.3 times as large
  - Marginal: population odds (prevalence with/prevalence without) of AD is exp(.57) = 1.8 greater for placebo than on active drug;
    population-average parameter
  - Random Effects: a person's odds of AD is exp(1.8)= 6.0 times greater on placebo than on active drug;
    cluster-specific, here person-specific, parameter

Which model is better? They ask different questions.

36

Refresher: Forests & Trees

Multi-Level Models:

- Explanatory variables from multiple levels
  - i.e. person, family, n'bhd, state, …
  - Interactions
- Take account of correlation among responses from same clusters:
  - i.e. observations on the same person, family,…
  - Marginal: GEE, MMM
  - Conditional: RE, GLMM ← Remainder of the course will focus on these.

37

Key Points

- "Multi-level" Models:
  - Have covariates from many levels and their interactions
  - Acknowledge correlation among observations from within a level (cluster)
- Random effect MLMs condition on unobserved "latent variables" to account for the correlation
- Assumptions about the latent variables determine the nature of the within cluster correlations
- Information can be borrowed across clusters (levels) to improve individual estimates

38

Examples of two-level data

- Studies of health services: assessment of quality of care are often obtained from patients that are clustered within hospitals. Patients are level 1 data and hospitals are level 2 data.
- In developmental toxicity studies: pregnant mice (dams) are assigned to increased doses of a chemical and examined for evidence of malformations (a binary response). Data collected in developmental toxicity studies are clustered. Observations on the fetuses (level 1 units) nested within dams/litters (level 2 data)
- The “level” signifies the position of a unit of observation within the hierarchy

39

Examples of three-level data

- Observations might be obtained in patients nested within clinics, that in turn, are nested within different regions of the country.
- Observations are obtained on children (level 1) nested within classrooms (level 2), nested within schools (level 3).

40

Why use marginal model when I can use a multi-level model?

- Public health problems: what is the impact of intervention/exposure on the population?
  - Most translation into policy makes sense at the population level
- Clinicians may be more interested in subject specific or hospital unit level analyses
  - What impact does a policy shift within the hospital have on patient outcomes or unit level outcomes?

41

Why use marginal model when I can use a multi-level model?

- Your study design may induce a correlation structure that you are not interested in
  - Sampling individuals within neighborhoods or households
  - Outcome: population mortality
  - Marginal model allows you to adjust inferences for the correlation while focusing attention on the model for mortality
- Dose-response or growth-curve
  - Here we are specifically interested in an individual trajectory
  - And also having an estimate of how the individual trajectories vary across individuals is informative.

42

Additional Points: Marginal Model

- We focus attention on the population level associations in the data and we try to model these best we can (mean model)
- We acknowledge that there is correlation and adjust for this in our statistical inferences.
- These methods (GEE) are robust to misspecification of the correlation

- We are obtaining estimates of the target of interest and valid inferences even when we get the form of the correlation structure wrong.

43

Multi-level Models

- Suppose you have hospital level summaries of patient outcomes
  - The fixed effect portion of your model suggests that these outcomes may differ by whether the hospital is teaching/non-teaching or urban/rural
  - The hospital level random effect represents variability across hospitals in the summary measures of patient outcomes; this measure of variability may be of interest
  - Additional interest lies in how large the hospital level variability is relative to a measure of total variability; what fraction of variability is attributable to hospital differences?

44

Additional considerations:

- Interpretations in the multi-level models can be tricky!
- Think about interpretation of gender in a random effects model:
  - E(Y|gender,bi) = b0 + b1gender + bi
  - Interpretation of b1:

    Among persons with similar unobserved latent effect bi, the difference in average Y if those same people had been males instead of females

  - Imagine the counter-factual world….does it make sense?

45

Comparison of Estimates: Linear Model and Non-linear model

- A hypothetical cross-over trial
  - N = 15 participants
  - 2 periods
  - treatment vs placebo
- Two outcomes of interest
  - Continuous response: say alcohol consumption (Y)
  - Binary response: say alcohol dependence (AD)

46

Linear model

E(Y|Period,Treatment) = b0 + b1Period + b2Treatment

	Ordinary Least Squares	GEE (Indep)	GEE (Exchange)	Random subject effect
Intercept (b0)	15.2 (1.22)	15.2 (1.16)	15.2 (1.07)	15.2 (1.13)
Period (b1)	2.57 (1.38)	2.57 (1.31)	2.57 (1.01)	2.57 (1.08)
Treatment (b2)	-0.43 (1.38)	-0.43 (1.31)	-0.43 (1.01)	-0.43 (1.08)

SAME estimates . . . DIFFERENT standard errors . . .

47

Non-Linear model

Log(Odds(AD|Period,Treatment)) = b0 + b1Period + b2Treatment

	Ordinary Logistic Regression	GEE (Indep)	GEE (Exchange)	Random subject effect
Intercept (b0)	-1.14 (0.75)	-1.14 (0.75)	-1.11 (0.83)	-1.14 (0.75)
Period (b1)	0.79 (0.83)	0.79 (0.83)	0.76 (1.02)	0.79 (0.83)
Treatment (b2)	1.82 (0.83)	1.82 (0.83)	1.80 (1.03)	1.82 (0.83)

SAME estimates and standard errors

Estimates and standard errors change (a little)

48

What happened in the GEE models?

- In non-linear models (binary, count, etc), the mean of the outcome is linked to the variance of outcome:
  - X ~ Binomial, mean p, variance p(1-p)
  - X ~ Poisson, mean λ, variance λ
- When we change the structure of the correlation/variance, we change the estimation of the mean too!
- The target of estimation is the same and our estimates are unbiased.

49

Why similarity between GEE and random effects here?

- No association in AD within person
- Little variability across persons
- Odds ratio of exposure across persons ~ 1

```
tab AD0 AD1

           |         1 AD
      0 AD |         0          1 |     Total
-----------+----------------------+----------
         0 |         1          7 |         8
         1 |         5          2 |         7
-----------+----------------------+----------
     Total |         6          9 |        15
```

50