multilevel modeling:

concepts, applications and interpretations

lynne c. messer

27 october 2010

overview

- warning – social and reproductive / perinatal epidemiologist
- concepts –
  - why context matters
  - multilevel models – terminology
- applications –
  - issues specific to nested data
  - different types of multilevel models
- interpretations –

concepts – why context matters

- empirically, individual outcomes can't be explained exclusively by individual-level exposures
- persistent contextual effects are observed in all (?) outcomes across populations
- exposures are structured; distributions are differential

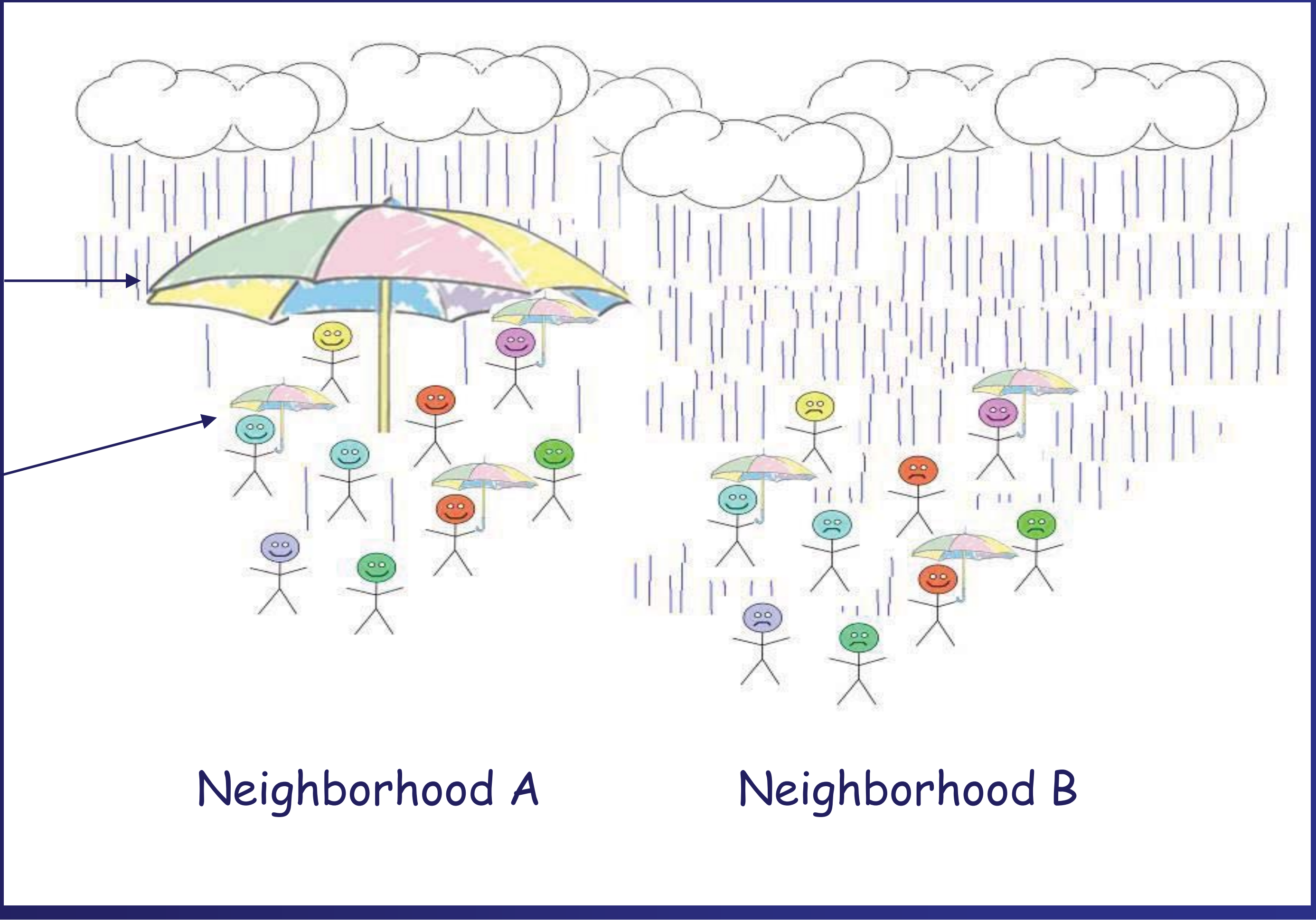
concepts – types of non-individual-level data

- compositional data
  - properties of individuals
  - aggregation of individual-level variables—such as census data
- contextual data
  - properties of places
  - integral variables; no individual-level analogs—services, resources
- directly observed data (can be combination)
  - survey direct observation of the built environment such as “walkability” or broken windows

concepts - partitioning variance

neighborhood - level protection

individual - level protection



concepts – definition and synonyms

- ml modeling: a method that allows researchers to investigate the effect of group or place characteristics on individual outcomes while accounting for non-independence of observations
- synonyms:
  - multilevel models
  - contextual models
  - hierarchical analysis

  different models:
  - fixed effects
  - random effects
  - marginal models (e.g., GEE)
- longitudinal (panel) data, repeated measures designs use ml methods as well

concepts – when are observations dependent?

- dependence arises when data are collected by cluster / aggregating unit
  - children within schools
  - patients within hospitals
  - pregnant women within neighborhoods
  - cholesterol levels within a patient
- why care about clustered data?
  - two children / observations within one school are probably more alike than two children / observations drawn from different schools
  - knowing one outcome informs your understanding about another outcome (i.e., statistical dependence)

concepts – why use multilevel models?

- standard regression models are mis-specified for clustered data
  - $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$; $\varepsilon \sim N(0,\sigma^2)$ i.i.d.
  - more on this
- hierarchical models out-perform unbiased models (result in lower mean squared errors)
  - more on this

concepts summary – why use multilevel models?

- outcomes may be clustered by some unit of aggregation (contextual unit)
- individuals within contexts may be similar in ways that are unmeasured
- to take into account clustering / non-independence of observations
- to partition the observed variability into within-context and between- context variables
- to allow for different types of policy or interventions to change population values / distributions

concepts – how to tell if you need ml models

- reality 1: anytime you have data collected from some aggregate unit / clusters, you will have to use ml models
- reality 2: calculating an intraclass correlation coefficient will quantify your clustering (in absence of running a ml model)
- reality 3: even if your 'clustered data' aren't empirically clustered, article and grant reviewers will demand it

application – linear and logistic regression

- linear model review:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} ... + \varepsilon_i$$

β_0 = intercept β_1 = slope for exposure X_1
β_2 = slope for covariate X_2...

ε_i = error term (assumed normal and i.i.d.)

- logistic model review:

$$\ln [P(y_i) / (1-P((y_i))] = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} ...$$

α = constant β_1 = slope for exposure X_1
β_2 = slope for covariate X_2

application – model assumptions

- baseline outcome means (mean values when exposure and covariates = 0) differs only due to variability between subjects
- individuals, and their errors, independent and identically distributed (i.i.d. assumption)
- all non-specified variables (e.g., area-level variables; those confounders you did not measure) assumed = 0

(inappropriate) application – add group-level variables

$$Y_{ij} = \beta_0 + \beta_{1ij}X_1 + \beta_{2ij}X_2 ... + \beta_j G_j + \varepsilon_{ij}$$

Y_{ij} = outcome for individual i in context j
β_{1ij}= slope for exposure X_1 for individual $_i$ in context $_j$ ...
β_j = slope for community variable G_j ε_{ij} = error term

- problem: making cross-level inferences [drawing inferences regarding factors associated with variability in outcome at one level based on data collected at another level]
- e.g., making individual inferences based on group-level associations

(inappropriate) application – interact group-level variables

$\ln [P(Y_{ij}) / (1-P(Y_{ij}))] = \alpha_i + \beta_{1ij}X_{1ij} + \beta_{2ij}X_{2ij} + \beta_{3j}G_j + \beta_{2ij}X_2{*}\beta_{3j}G_j$

- interacting group- and individual-level variables will get you close to the right answer
- problem: error structure is multilevel, but errors only specified at the individual-level
- individuals within contexts are correlated with each other
- errors not independent and identically distributed

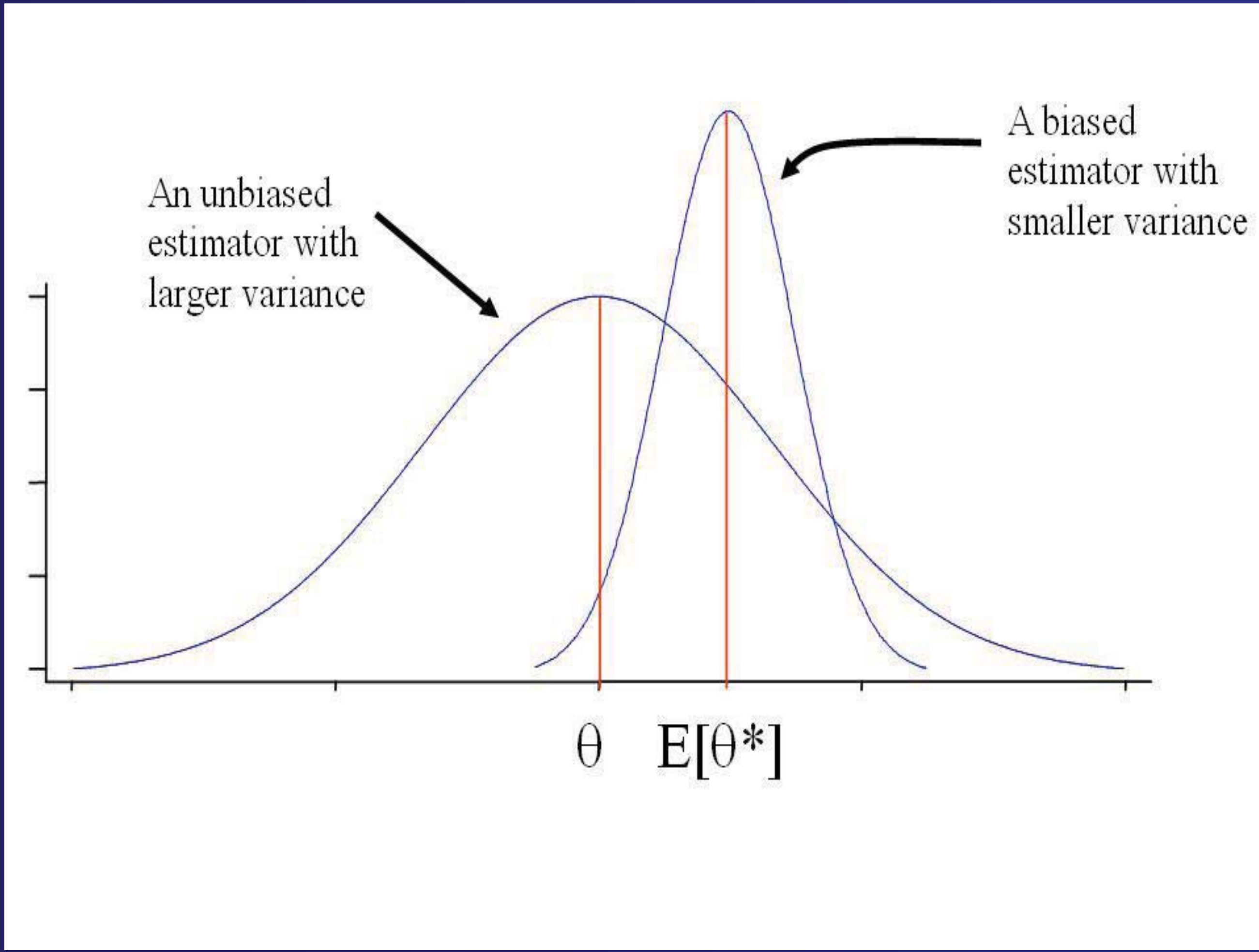
application – multilevel error structure?

- standard methods generally produce unbiased point estimates
  - your betas or ORs will be ~correct
- standard errors too small
  - confidence intervals will be wrong (too precise)
- unless you can demonstrate there are no correlations between the following:
  - individual-level predictors
  - group-level predictors
  - unobserved characteristics

application – introduction to "shrinkage"

- trade-off between bias and precision in the estimation of parameter θ using estimator θ^*
- $MSE(\theta^*) = E[\theta^* - \theta]^2$
- $VAR(\theta^*) = E[\theta^* - E[\theta^*]]^2$
- $BIAS(\theta^*) = (E[\theta^*] - \theta)$
- $MSE(\theta^*) = VAR(\theta^*) + BIAS(\theta^*)^2$

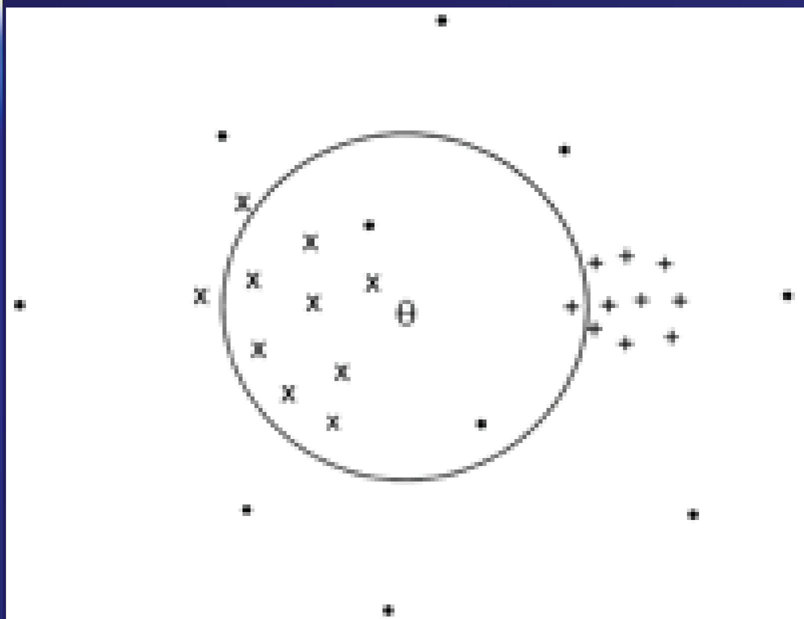
application – bias versus variance

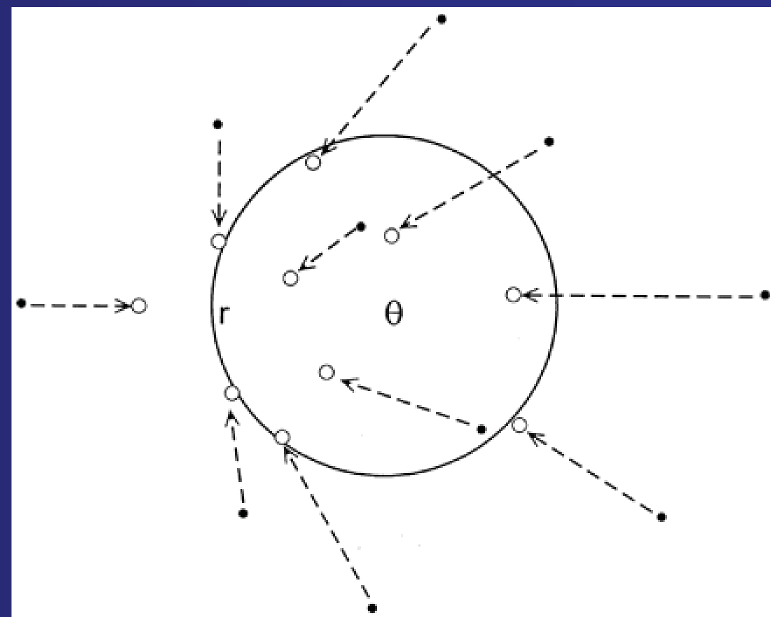


application – bias versus variance

- it is possible for the variance of a biased estimator to be sufficiently smaller than the variance of an unbiased estimator to more than compensate for the bias introduced.
- in this case, the biased estimator is closer, on average, to the parameter being estimated than is the unbiased estimator.

application – greenland 2000





Greenland 2000; Figure 1

• = Rifle 1 shots
X = Rifle 2 shots
+ = Rifle 3 shots

Greenland 2000; Figure 2

How cluster from Rifle 1 could be made better by pulling toward a point r.

application – to ml models?

- when you have information for j different clusters, you can use the grand mean as the "prior" to shrink toward
- translation: R = sum of context-specific estimates or grand mean
- just need to know weights for each estimate
- translation: How much do you trust the cluster-specific proportions, versus how much you trust the grand proportion?
- answer: depends on N and ICC

application – intraclass correlation coefficient

- estimates the degree of clustering by unit of aggregation
- icc = between cluster variance / total variance*
  - icc = 0 : no clustering - people within a cluster are just the same as people in another cluster
  - icc > 0 : people in same cluster are more similar to each other than to people in other clusters

*total variance = within cluster + between cluster variance

application – modeling clustered data

- two main approaches:
  - population average models with robust variance estimators
    - marginal models that account for cross level correlation across all units of aggregation
    - not conditional on being in a certain cluster; does not model clustering directly
    - provides robust tests, corrected standard errors, corrects for heteroskedasticity*
  - ml models
    - random effects models (unit-specific models that condition on specific units of aggregation for inference)
    - fixed effects models (area-level coefficients held constant across units of aggregation)
    - mixed models (models that combine some fixed and random effects; not going into any more detail about mixed models in this lecture)

* heteroskedasticity results from errors not having constant variance

application – population average models

- developed by Eicker (1963, 1967), Huber (1967) and White (1980); often referred to as "huber-white" or "sandwich" variance
- does not specify the population distribution; only specifies the marginal distribution
- examples: generalized estimating equations with robust errors
- pros
  - model response change as function of covariates 'averaged' over group to group heterogeneity
- cons
  - do not explicitly account for heterogeneity across higher-level units / contexts; therefore no examination of group to group variation

application – population average models

$\Pr(Y_{ij}=1 \mid X_{ij}) = f(X_{ij}\,\beta^*)$ ← note: no conditioning on cluster

- Y_{ij} = preterm birth (1) versus term birth (0) for woman *i* in tract *j*
- X_{ij} = low (1) or high (0) ses for woman *i* in tract *j*
- no locations specified, just averaged over all tracts
- allows you to compare 'average low' versus 'average high' ses women

application - multilevel models random and fixed

random intercept models: context specific mean realized from a random distribution

- random effects models

random intercept

random slope

random slope and random intercept

random slope models: exposure effect realized from a random distribution

application – random effects model – simplest (1)

$\ln [P(y_i) / (1-P((y_i))] = \beta_{0j}$

$\beta_{0j} = \gamma_{00} + \mu_j$ ← γ_{00} distribution mean of random coefficients, estimated as weighted average of tract intercepts; μj = cluster-specific parameter

- simplest hierarchical logistic model expresses context-level intercepts β_{0j} as function of overall intercept γ_{00} and context-specific random deviations μ_j

application – random effects model – next (2)

add individual-level or neighborhood-level covariates to explain some of the between tracts variance.

for probability of preterm delivery $p_{ij} = Pr(y_{ij} = 1)$ for individuals i in tracts j:

$$\ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \beta_{0j} + \beta_1 X_{ij}$$

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + \mu_{0j}, \qquad \mu_{0j} \sim N(0, \tau_{00})$$

application - random effects model - next (2)

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + \mu_{0j}, \qquad \mu_{0j} \sim N(0, \tau_{00})$$

- γ_{00} is the mean of the distribution of random coefficients, estimated as the weighted average of tract intercepts.
- both the log-odds of outcome in each tract and γ_{00} (the weighted average of tract-specific log-odds) are estimates for the true tract-specific log-odds.
- an optimal (minimum MSE) estimator for β_{0j} is formed by taking the weighted average of these two quantities, with intra-class correlations for weights

application – logistic random intercept models [3]

replacing the second-level equation into the first level equation yields the combined equation:

note: conditioning on cluster ↓

$$\ln\left(\frac{p_{ij}}{1-p_{ij}}\right) = \gamma_{00} + \gamma_{01} Z_j + \beta_1 X_{ij} + \mu_{0j}$$

these models have random effects only for the intercept, but one could also specify models with random effects for one or more of the slope terms.

application – logistic random effects output

```
. xi: xtlogit single_preterm  AS_stddepriv_8_new i.cat_momage i.cat_momedu, i(c_tract) or nolog
i.cat_momage      _Icat_momag_0-5     (naturally coded; _Icat_momag_0 omitted)
i.cat_momedu      _Icat_momed_1-3     (naturally coded; _Icat_momed_1 omitted)

Random-effects logistic regression              Number of obs      =     28158
Group variable (i): c_tract                     Number of groups   =       105

Random effects u_i ~ Gaussian                   Obs per group: min =        34
                                                               avg =     268.2
                                                               max =      1148

                                                Wald chi2(8)       =    104.08
Log likelihood  = -7825.5801                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
single_pre~m |         OR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
AS_stddepr~w |   1.209596   .0436054     5.28   0.000     1.127081    1.298153
_Icat_moma~1 |   1.017455   .1558981     0.11   0.910     .7535132     1.37385
_Icat_moma~2 |   .9998798   .1140788    -0.00   0.999      .799525    1.250442
_Icat_moma~3 |   1.110174   .0762816     1.52   0.128     .9702947    1.270218
_Icat_moma~4 |   1.028658   .0739753     0.39   0.694     .8934238    1.184363
_Icat_moma~5 |   1.266427   .0988397     3.03   0.002     1.086793    1.475751
_Icat_mome~2 |    1.42299   .0845803     5.93   0.000     1.266508    1.598807
_Icat_mome~3 |   1.317904    .097865     3.72   0.000     1.139398    1.524377
-------------+----------------------------------------------------------------
    /lnsig2u |  -4.466719   .7646109                     -5.965329   -2.968109
-------------+----------------------------------------------------------------
     sigma_u |   .1071678   .0409708                      .0506577    .2267166
         rho |   .0034789   .0026507                      .0007794    .0153835
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =     2.43 Prob >= chibar2 = 0.060
```

application – random effects models

- pros
  - helps explain variance in area or context effects
  - allows observation of group to group variation and interaction with individual-level variables
- cons
  - don't account for selection bias
  - don't account for omitted variable bias

application- why partition variance?

- random-effects models allow you to decompose the total variance in individual-level outcomes into within-group and between-group components

- In the ANOVA context, has an explanatory interpretation as identifying the mechanism as being contextual or compositional

application – linear fixed effect multilevel model

$$Y_{ij} = \beta_0 + \beta_1 X_1 + \beta_2 X_1 \ldots + \beta_{286} X_1 + \varepsilon_{ij}$$

- β_1 = context – level estimate for exposure X_1 in context 1…
- context-specific variables not allowed to vary; held fixed
- controls for observed and unobserved / unmeasured context variables
- usually accomplished by creating an indicator for each contextual variable

application – fixed effects models

- pros
  - controls for neighborhood effects
  - controls for selection bias
  - controls for omitted variable bias
- cons
  - doesn't explain variance between contexts
  - can't identify independent contextual effects
  - statistically inefficient

application – linear fixed effects output

```
. xi: xtreg term_bwt i.cat_momedu, i(tract) fe
i.cat_momedu      _Icat_momed_0-2     (naturally coded; _Icat_momed_0 omitted)

Fixed-effects (within) regression               Number of obs      =       968
Group variable (i): tract                       Number of groups   =       100

R-sq:  within  = 0.0023                         Obs per group: min =         1
       between = 0.1332                                        avg =       9.7
       overall = 0.0066                                        max =        39

                                                F(2,866)           =      1.00
corr(u_i, Xb)  = 0.0953                         Prob > F           =    0.3684

------------------------------------------------------------------------------
    term_bwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
_Icat_mome~1 |  -6.692403   41.73948    -0.16   0.873    -88.61478    75.22997
_Icat_mome~2 |  -61.08572   46.44495    -1.32   0.189    -152.2435     30.0721
       _cons |   3425.676   27.98791   122.40   0.000     3370.744    3480.608
-------------+----------------------------------------------------------------
     sigma_u |  219.86417
     sigma_e |  505.85722
         rho |  .15889284   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(99, 866) =     1.04             Prob > F = 0.3692
```

application – how to decide which model to use

- depends on what you want to say…
  - if you want to look at risk / odds for the average individual with some exposure compared with average individual with some other exposure, use a population averaged model (e.g., GEE with robust estimator)
  - if you want to talk about how changes in context specific exposures will change the risk / odds in that context, use the unit-specific / random effects model
  - if you want to want to consider the effect of some variable holding all observed and unobserved contextual factors constant, use a context fixed effect model

interpretation – unemployment and PTB; 3 models

	Logistic		Logistic (PA)		Logistic (RE)	
	OR	95% CI	OR	95% CI	OR	95% CI
>5% unemployment	1.29	(1.08, 1.55)	1.29	(1.04, 1.61)	1.31	(1.04, 1.64)
Age 25-29	1.31	(1.05, 1.64)	1.31	(1.04, 1.61)	1.31	(1.05, 1.64)
Age 30-34	1.69	(1.33, 2.15)	1.70	(1.35, 2.10)	1.68	(1.32, 2.14)
Age 35+	2.10	(1.60, 2.76)	2.10	(1.60, 2.77)	2.10	(1.60, 2.75)
High school	1.37	(1.13, 1.66)	1.37	(1.10, 1.70)	1.38	(1.14, 1.67)
< High school	1.74	(1.36, 2.26)	1.74	(1.33, 2.27)	1.76	(1.34, 2.29)
Not married	1.49	(1.23, 1.80)	1.49	(1.25, 1.77)	1.49	(1.23, 1.80)

interpretation – logistic model

- standard logistic model (>5% unemployment versus ≤5% unemployment)

  OR = 1.29 (95% CI: 1.08, 1.55)

assuming that women are evenly distributed across neighborhoods with regard to preterm birth, the odds of preterm delivery will increase by 29% for a randomly selected woman living in a high unemployment tract compared with a randomly selected woman living in a low unemployment tract

interpretation – population average model

- population average logistic model (>5% unemployment versus ≤5% unemployment)

  OR = 1.29 (95% CI: 1.04, 1.61)

the odds of preterm delivery will increase by 29% for a randomly selected woman (read: average woman) in a low unemployment tract if she were to be relocated to a high unemployment tract

interpretation – random effects logistic model

- random effects logistic model (>5% unemployment versus ≤5% unemployment)

  OR = 1.31 (95% CI: 1.04, 1.64)

the odds of preterm delivery will increase by 31% for a randomly selected woman in a specific census tract with low unemployment if that tract is somehow manipulated to have high unemployment

interpretation - fixed effects logistic model

- fixed effects logistic model (low personal ses versus high personal ses)

OR= 1.29 (95% CI: 1.04, 1.61)

holding all observed and unobserved contextual effects fixed, the odds of preterm delivery will increase by 29% for a randomly selected woman with low individual-level ses compared with a randomly selected woman with high individual-level ses

interpretation - conclusion

- standard regression models assume that data is not clustered by a higher level grouping
- one can model clustered data by either using methods robust to this violation of assumptions, or else by modeling this clustering directly
- random effects models estimate *conditional* parameters (i.e., the effect of exposure *given* a particular cluster)

thank you

and

questions?

fallacies associated with different levels of data / inference

- atomistic fallacy - drawing inferences regarding variability across contexts based on aggregation of individual responses
  - a.k.a., individualistic fallacy
- ecological fallacy – drawing inferences regarding variability across individuals based on group-level / aggregated data
- psychologistic fallacy – failure to consider group-level characteristics in drawing inferences regarding causes of individual variability
- sociologistic fallacy – failure to consider individual-level characteristics in drawing inferences regarding causes of group variability