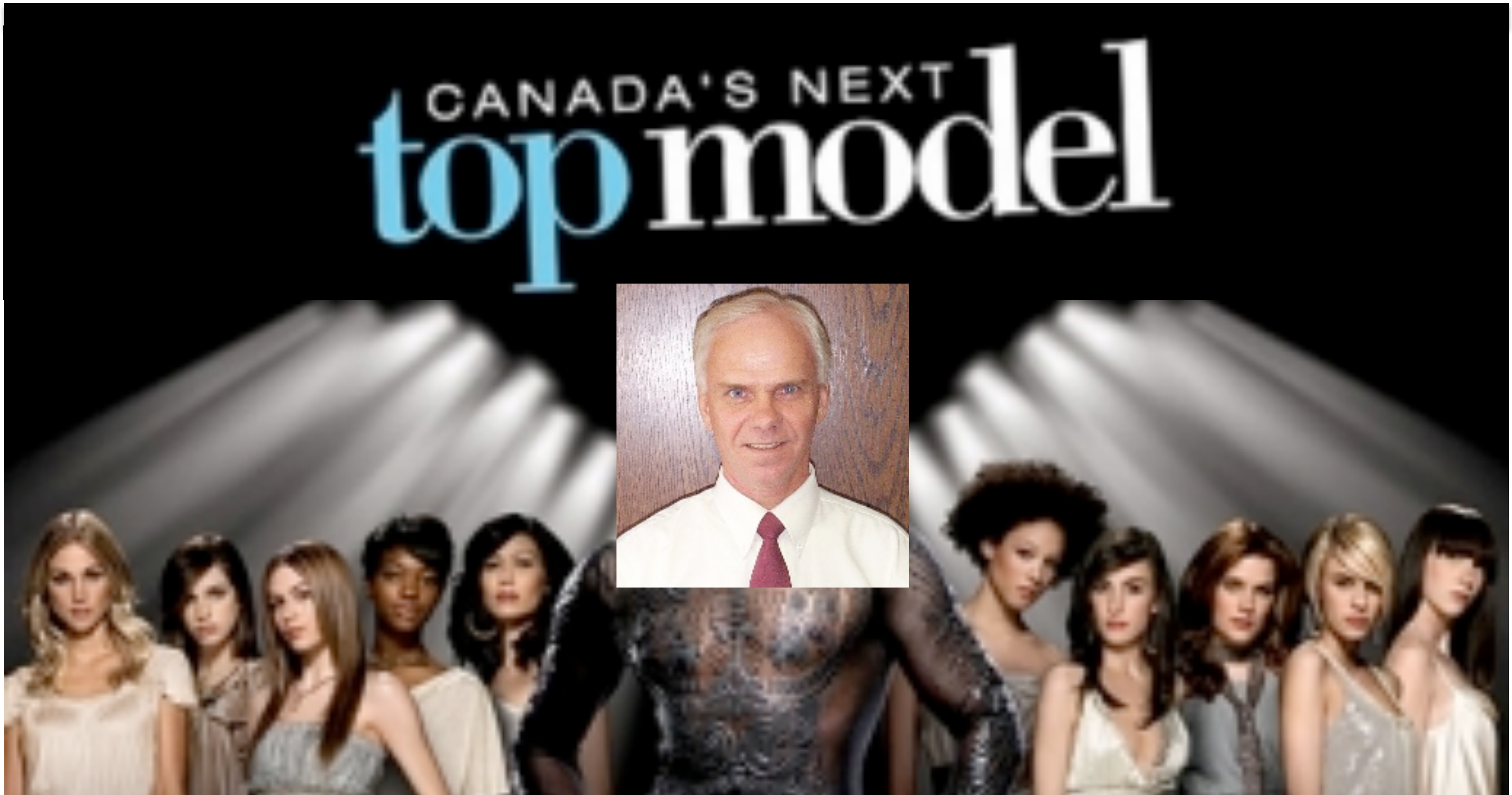


Mixed models

and why sociolinguists should use them

Daniel Ezra Johnson

VARBRUL / GoldVarb	other
dependent variable (DV)	DV, response, y
factor group, independent variable (IV)	IV, factor (categorical), predictor, x
factor	level
factor weight	coefficient, effect, estimate, β
factor weight range	similar to ‘effect size’
input probability	intercept
applications / total	(response) proportion

`lmer`	other
mixed model	mixed-effects, hierarchical, or multilevel model
fixed effect	main effect
(all) fixed-effects model	flat model
conditional modes of random effects	random effect estimates, random effect BLUPs

Terminological ‘translations’

PROPERTIES OF DATA	GoldVarb	Rbrul	R	POSSIBLE ANALYSIS
response / DV: 2 categories	✔	✔	✔	logistic regression
response: 3+ categories			✔	ordinal, multinomial logistic
response: count			✔	Poisson regression, etc.
response: continuous		✔	✔	linear regression
predictor(s) / IV(s) : categorical	✔	✔	✔	(any)
predictor(s): continuous		✔	✔	(any)
predictor(s): have interactions	***hard***		✔	(any)
random intercept(s)	***?***	✔	✔	**mixed model**
random slope(s)	***??***		✔	mixed model
lots of data (need for speed)		✔	✔	
		hard	✔	plots and graphics
			✔	other statistical methods
	✔			"slash" operator
	?	***?***		user friendly

Comparing Software Tools

GoldVarb

Rbrul

R

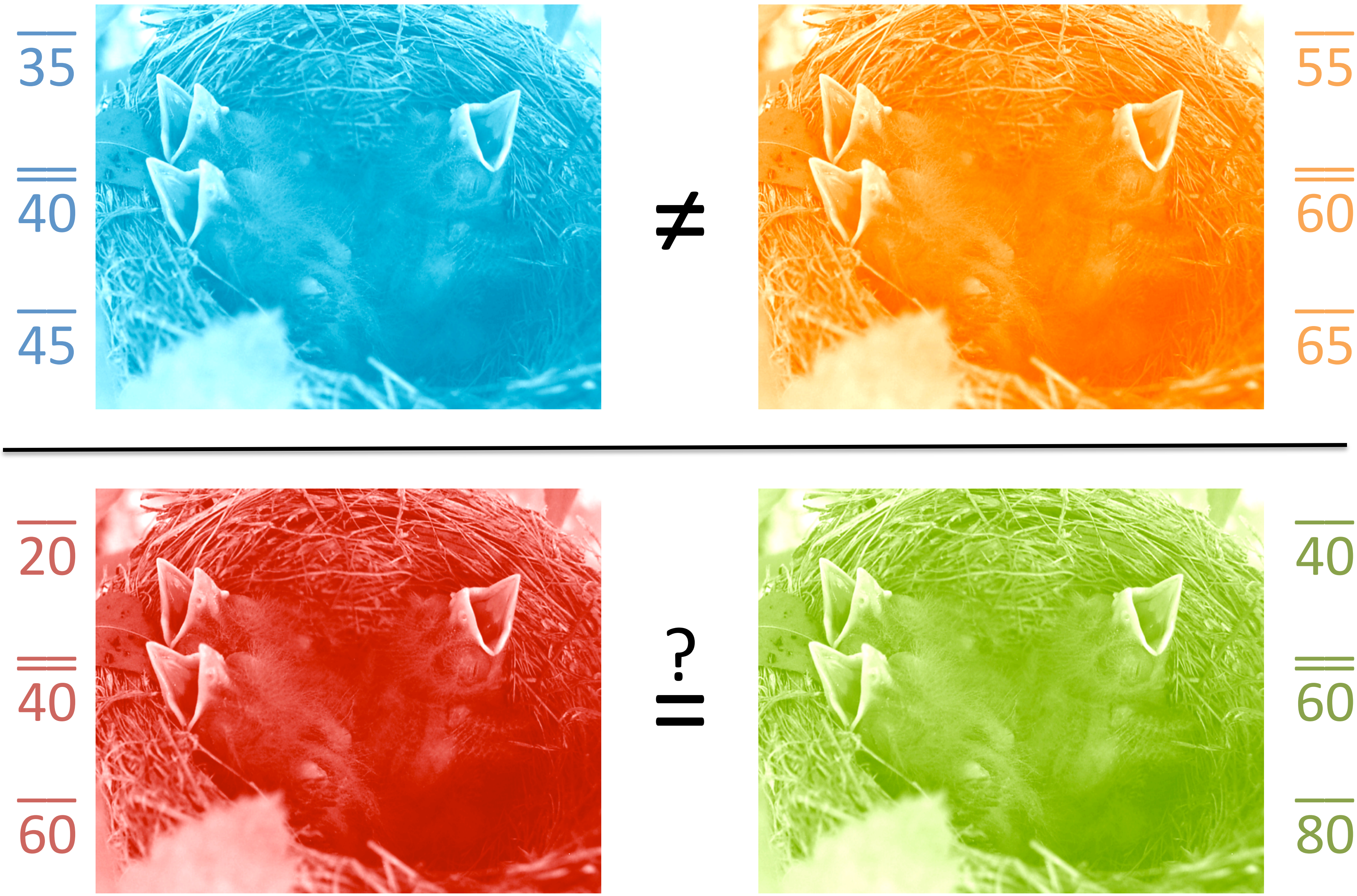
Finding the right tool for the job

- mixed models: both fixed effects and random effects
- fixed effect: ordinary regression predictor (IV)
- random effect: theoretically sampled from a population
    - est. population variance (s.d.) is the real parameter
    - individual estimates (BLUPs) “shrunk” towards mean
    - residual random effects should be normally distributed

- random intercept: individuals “high” or “low” (input prob.)
- random slope: individuals differ w.r.t. predictors (constraints)

- in model fitting, there is a penalty on the random effects
    - as much variance as possible assigned to fixed effects
    - only the left-over variance is assigned to random effects
- this random effect penalty allows nested models to fit
    - sometimes fixed vs. random (or separate runs) is a valid choice
    - but nested predictors must be random effects in a mixed model

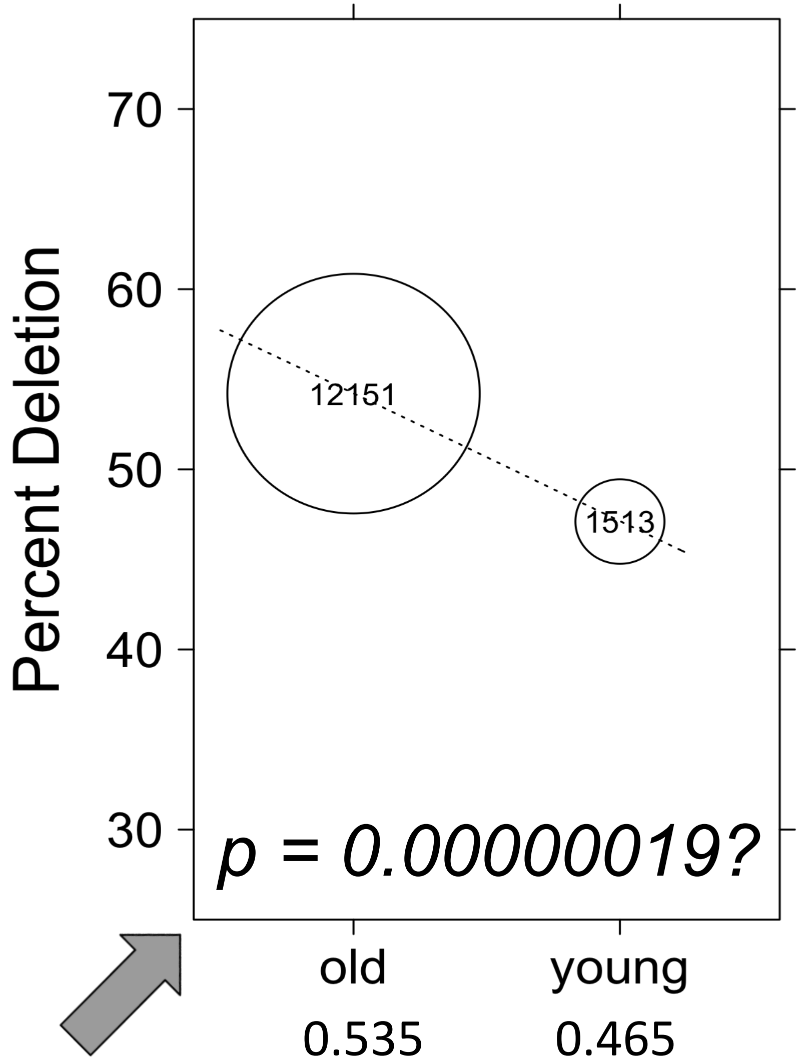
What are mixed models?



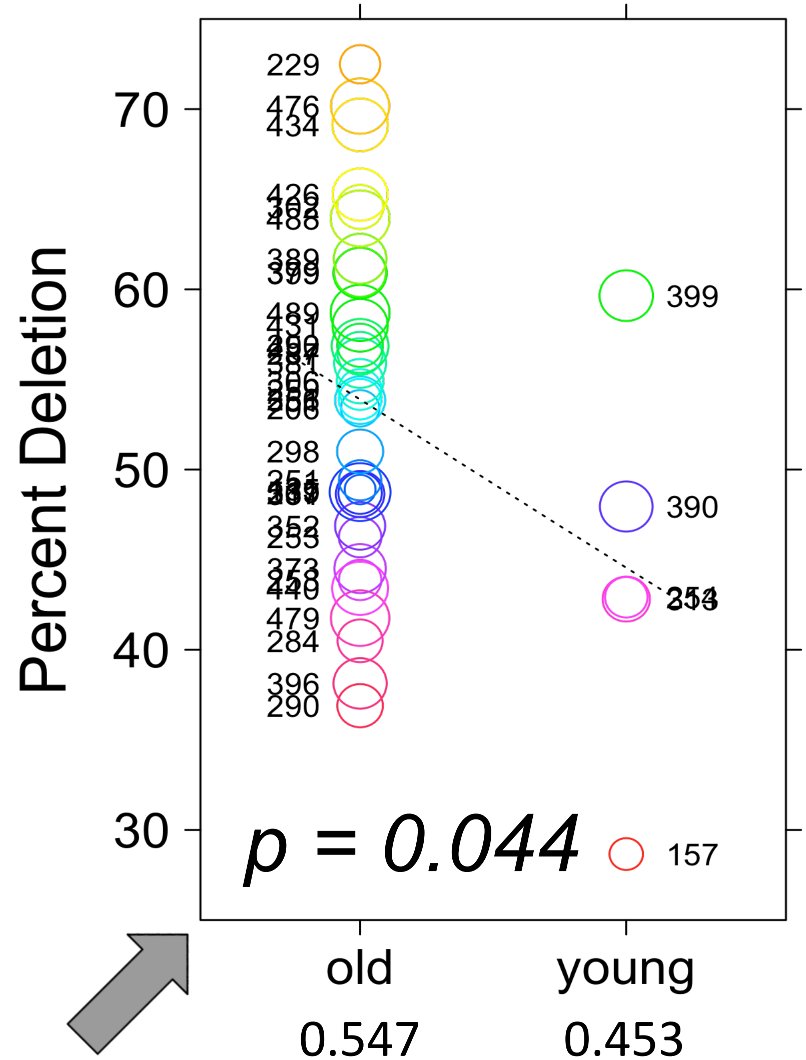
Mixed models for nested data

When we don’t need mixed models

And when we might need them



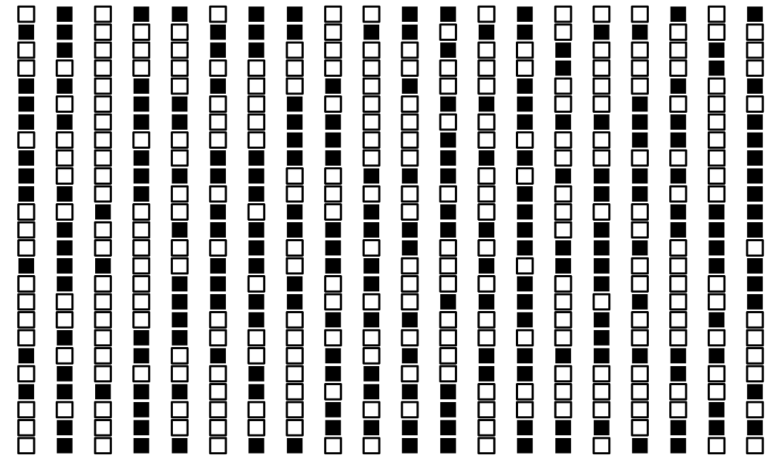
age w/ no random effect

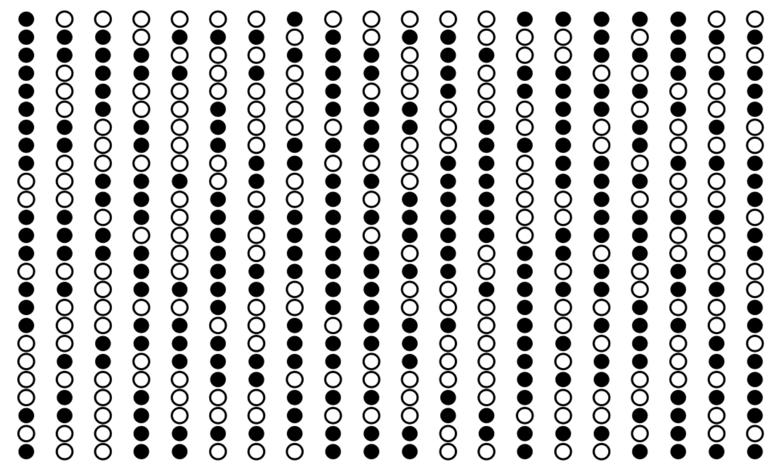


age + random intercept: speaker

Random effects and significance

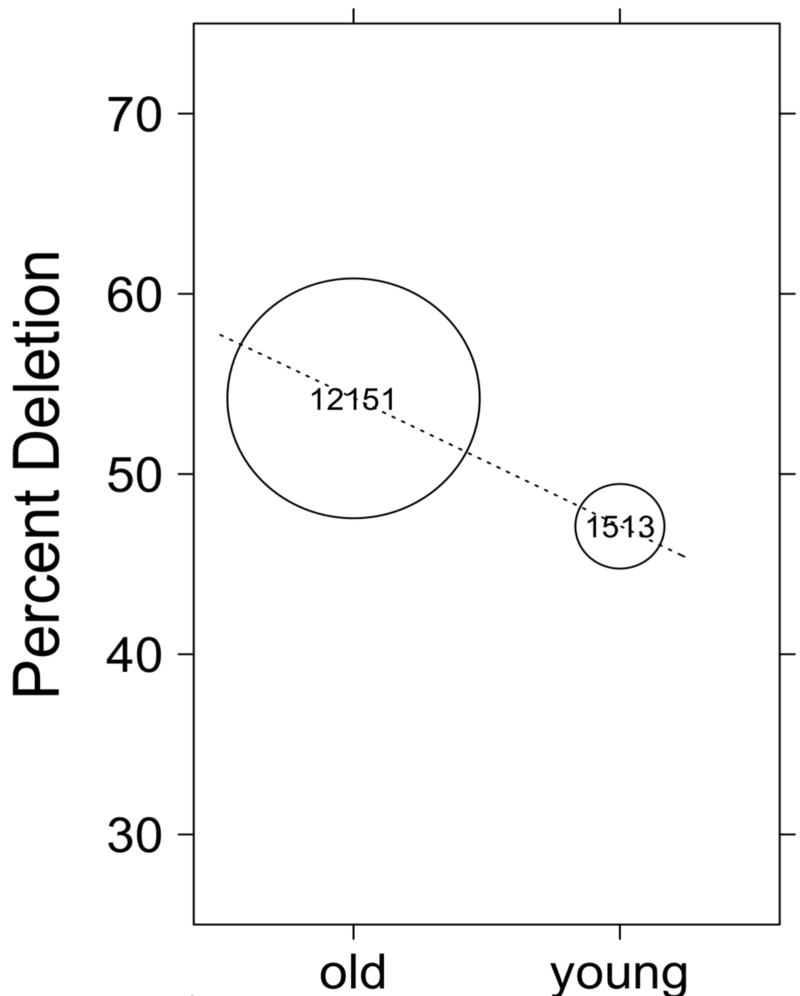
large effect size: 0.167 vs. 0.833
small significance: p = 0.08





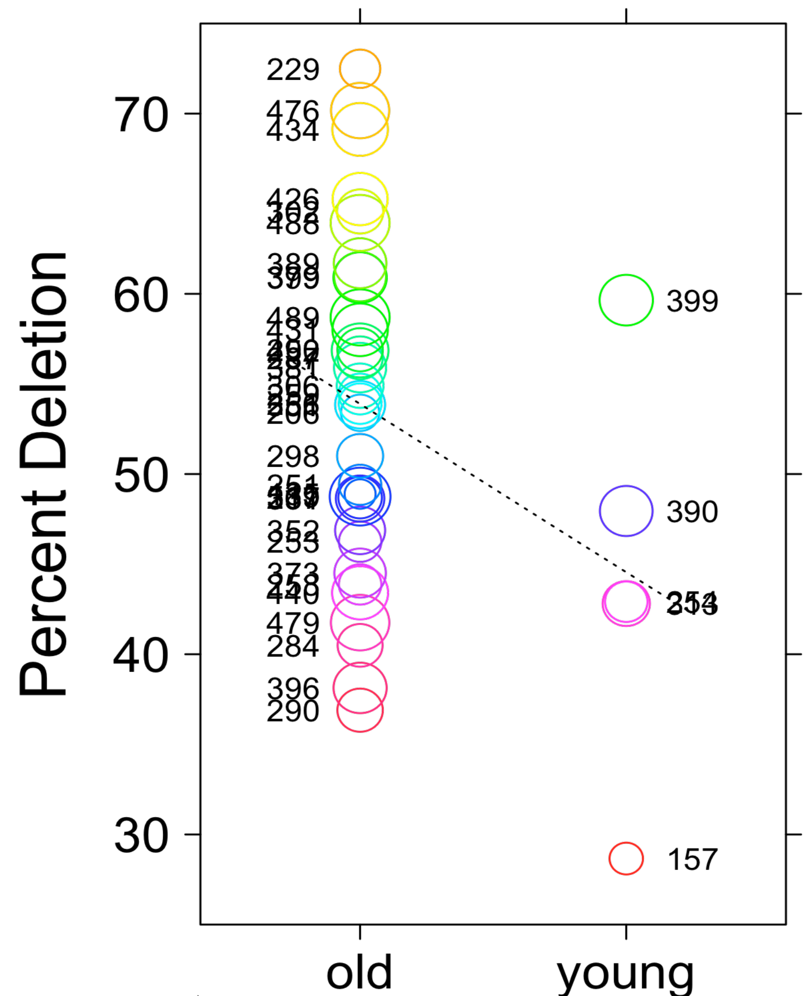
small effect size: 0.45 vs. 0.55
larger significance: p = 0.002

Significance vs. ‘effect size’



0.535 0.465

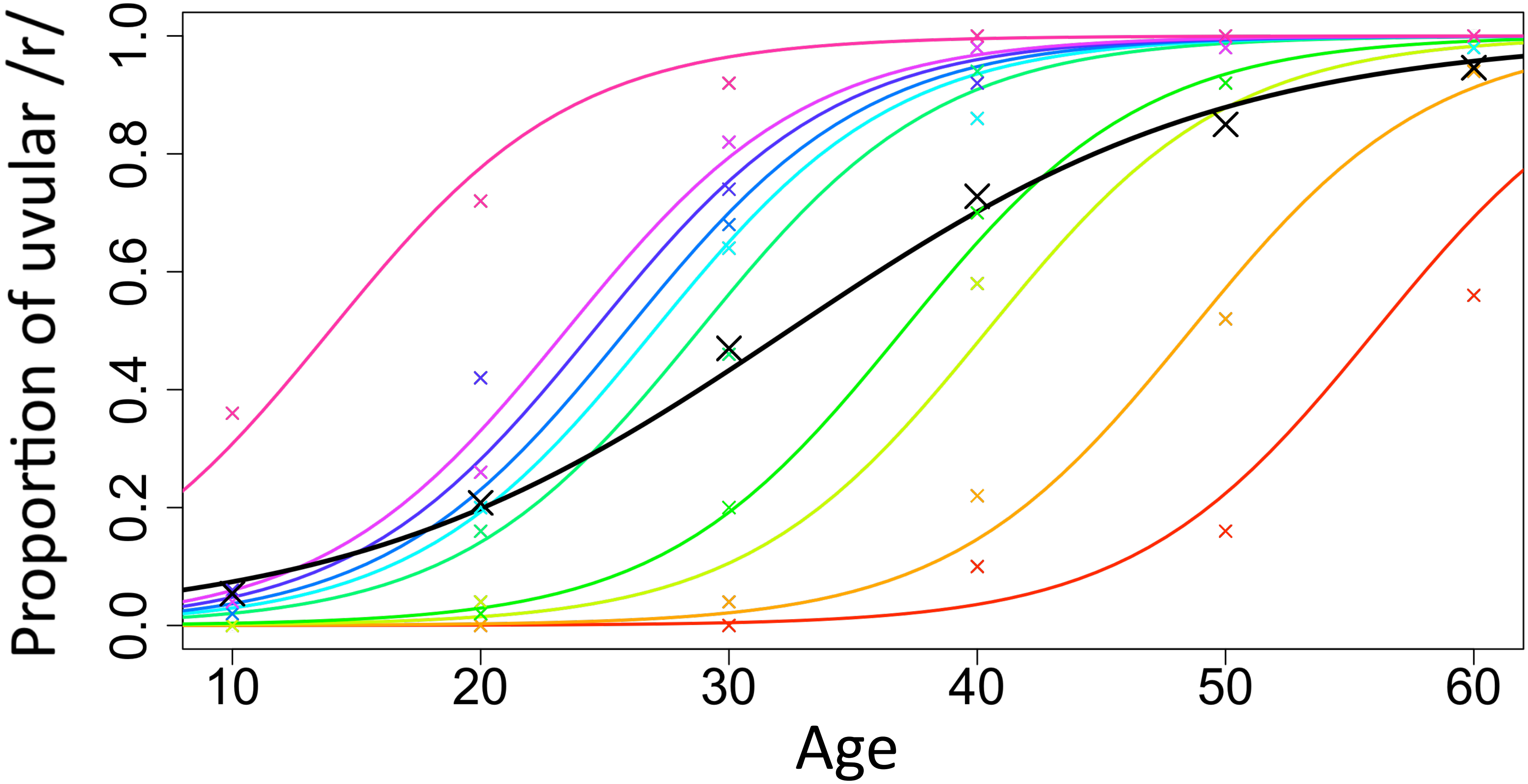
age w/ no random effect



0.547 0.453

age + random intercept: speaker

Unbalanced data and effect size



age coefficient w/ no random effect: 0.113 log-odds/year

age coeff. w/ speaker random effect: 0.205 log-odds/year

Crossed factors and effect size

Summary: speaker effect's effects

speaker-nesting predictors

constant within (data from) each speaker

age? gender race class c.o.p. ...

- significance more accurate:
  p = larger, "no longer significant"?
- effect sizes more accurate with unbalanced data: larger/smaller

speaker-crossed predictors

vary within (data from) each speaker

age? style phon./gram. context...

- effect sizes more accurate:
  larger (logistic regression only)

speaker-nesting predictors	word-nesting predictors
constant within (data from) each speaker	*constant within (data from) each word*
age? gender race class c.o.p. ...	frequency gram. cat. int. phon. ..

- significance more accurate:
  p = larger, "no longer significant"?
- effect sizes more accurate with
  unbalanced data, larger/smaller

speaker-crossed predictors	word-crossed predictors
vary within (data from) each speaker	*vary within (data from) each word*
age? style phon./gram. context...	stress style ext. phon. ...

- effect sizes more accurate:
  larger (logistic regression only)

Word effect just like speaker effect

speaker-nesting predictors

constant within (data from) each speaker

age? gender race class c.o.p. ...

word-nesting predictors

constant within (data from) each word

frequency gram. cat. int. phon. ..

- significance more accurate:
  p = larger, "no longer significant"?
- effect sizes more accurate with
  unbalanced data, larger/smaller

speaker-crossed predictors

vary within (data from) each speaker

age? style phon./gram. context...

word-crossed predictors

vary within (data from) each word

stress style ext. phon. ...

word

speaker

- effect sizes more accurate:
  larger (logistic regression only)

Crossed random effects for speaker & word

- use random effect estimates to identify ‘new’ fixed effects
  - modeled subject/word variation may include true individual variation, as well as unmodeled fixed effects
  -
- use random effect estimates to (empirically) build groups

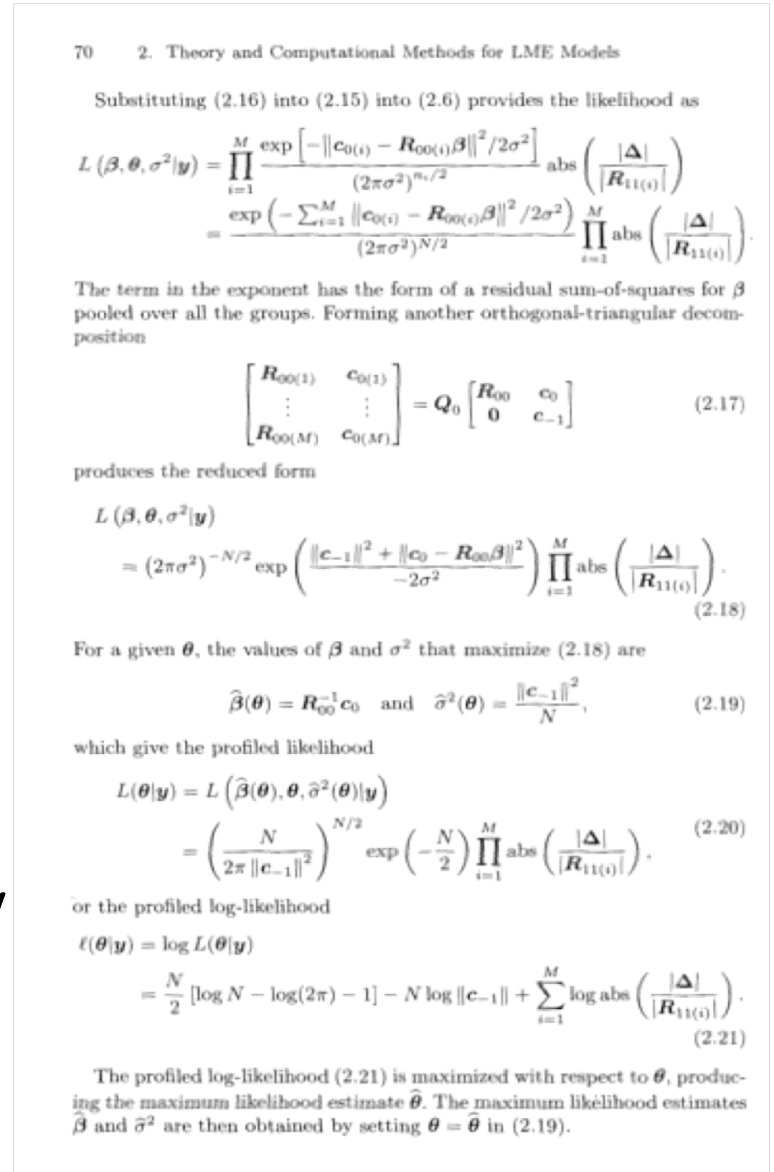
- use random effect estimates as predictors in new models

- use random effect population variances to predict behavior of new subjects and words not in the original sample

- can perform an easy transformation into the ‘language’ of GoldVarb (with some caveats) – this is not a real problem

Other benefits of mixed models

- cutting-edge statistics,
  like VARBRUL was in the 1970's
  - follow evolution on R-sig-ME
- double debate over p-values:
  - best way to calculate them
  - should they be used at all?
- convergence problems
  - requires more data (1000's > 100's)
- mixed model tool can be used
  well or badly, just like any model
  - still need to address multicollinearity
- should not be the only tool
  - mixed models are a better hammer,
    but everything is still not a nail



70 2. Theory and Computational Methods for LME Models

Substituting (2.16) into (2.15) into (2.6) provides the likelihood as

$$L\left(\boldsymbol{\beta},\boldsymbol{\theta},\sigma^2|\boldsymbol{y}\right) = \prod_{i=1}^{M} \frac{\exp\left[-\left\|\boldsymbol{c}_{0(i)} - \boldsymbol{R}_{00(i)}\boldsymbol{\beta}\right\|^2/2\sigma^2\right]}{(2\pi\sigma^2)^{n_i/2}} \operatorname{abs}\left(\frac{|\boldsymbol{\Delta}|}{\left|\boldsymbol{R}_{11(i)}\right|}\right)$$
$$= \frac{\exp\left(-\sum_{i=1}^{M}\left\|\boldsymbol{c}_{0(i)} - \boldsymbol{R}_{00(i)}\boldsymbol{\beta}\right\|^2/2\sigma^2\right)}{(2\pi\sigma^2)^{N/2}} \prod_{i=1}^{M} \operatorname{abs}\left(\frac{|\boldsymbol{\Delta}|}{\left|\boldsymbol{R}_{11(i)}\right|}\right).$$

The term in the exponent has the form of a residual sum-of-squares for $\boldsymbol{\beta}$ pooled over all the groups. Forming another orthogonal-triangular decomposition

$$\begin{bmatrix} \boldsymbol{R}_{00(1)} & \boldsymbol{c}_{0(1)} \\ \vdots & \vdots \\ \boldsymbol{R}_{00(M)} & \boldsymbol{c}_{0(M)} \end{bmatrix} = \boldsymbol{Q}_0 \begin{bmatrix} \boldsymbol{R}_{00} & \boldsymbol{c}_0 \\ \boldsymbol{0} & \boldsymbol{c}_{-1} \end{bmatrix} \tag{2.17}$$

produces the reduced form

$$L\left(\boldsymbol{\beta},\boldsymbol{\theta},\sigma^2|\boldsymbol{y}\right) = \left(2\pi\sigma^2\right)^{-N/2} \exp\left(\frac{\|\boldsymbol{c}_{-1}\|^2 + \|\boldsymbol{c}_0 - \boldsymbol{R}_{00}\boldsymbol{\beta}\|^2}{-2\sigma^2}\right) \prod_{i=1}^{M} \operatorname{abs}\left(\frac{|\boldsymbol{\Delta}|}{\left|\boldsymbol{R}_{11(i)}\right|}\right). \tag{2.18}$$

For a given $\boldsymbol{\theta}$, the values of $\boldsymbol{\beta}$ and σ^2 that maximize (2.18) are

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \boldsymbol{R}_{00}^{-1}\boldsymbol{c}_0 \quad \text{and} \quad \widehat{\sigma}^2(\boldsymbol{\theta}) = \frac{\|\boldsymbol{c}_{-1}\|^2}{N}, \tag{2.19}$$

which give the profiled likelihood

$$L(\boldsymbol{\theta}|\boldsymbol{y}) = L\left(\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}),\boldsymbol{\theta},\widehat{\sigma}^2(\boldsymbol{\theta})|\boldsymbol{y}\right)$$
$$= \left(\frac{N}{2\pi\|\boldsymbol{c}_{-1}\|^2}\right)^{N/2} \exp\left(-\frac{N}{2}\right) \prod_{i=1}^{M} \operatorname{abs}\left(\frac{|\boldsymbol{\Delta}|}{\left|\boldsymbol{R}_{11(i)}\right|}\right), \tag{2.20}$$

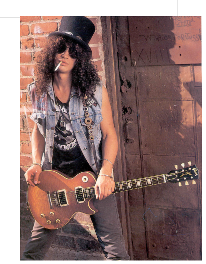
or the profiled log-likelihood

$$\ell(\boldsymbol{\theta}|\boldsymbol{y}) = \log L(\boldsymbol{\theta}|\boldsymbol{y})$$
$$= \frac{N}{2}\left[\log N - \log(2\pi) - 1\right] - N\log\|\boldsymbol{c}_{-1}\| + \sum_{i=1}^{M} \log \operatorname{abs}\left(\frac{|\boldsymbol{\Delta}|}{\left|\boldsymbol{R}_{11(i)}\right|}\right). \tag{2.21}$$

The profiled log-likelihood (2.21) is maximized with respect to $\boldsymbol{\theta}$, producing the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$. The maximum likelihood estimates $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma}^2$ are then obtained by setting $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$ in (2.19).

- "All models are wrong ... but some are useful." – Box

Drawbacks to mixed models



- it is fixed-effect models that make an assumption:
  - that residual subject and word variances are zero
  - i.e. that word-specific phonology is wrong
- mixed models are agnostic
  - random effects can be zero
  - they do not *assume* a word-specific (or speaker-specific) phonology, they *allow* for it *if it is supported by the data*
- must model speaker/word
  - with random effects, if nested
  - often crossed r. effects for both
- or other results will be wrong
  - maybe not very far wrong?
- as quantitative linguists, we strive for right numbers

Sali Tagliamonte
fellow panelists
Josef Fruehwald
Maryam Bakht
Meghan Armstrong
Kyle Gorman
Kirk Hazen
David Sankoff
Florian Jaeger
Rbrul testers
R developers

Doug Bates `lmer`
Qdoba on Bleecker

Pinheiro, José C. and Douglas M. Bates. 2000. *Mixed-Effects Models in S & S-PLUS.* New York: Springer.

Baayen, R. Harald, Douglas J. Davidson and Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59, 390-412. [I recommend this whole special issue on Emerging Data Analysis.]

Johnson, Daniel Ezra. 2009. Getting off the GoldVarb Standard: introducing Rbrul for mixed-effect variable rule analysis. *Language and Linguistics Compass* 3/1: 359-383.

Rbrul (a work in progress) is at: www.danielezrajohnson.com/Rbrul.R

Conclusions, thanks, references