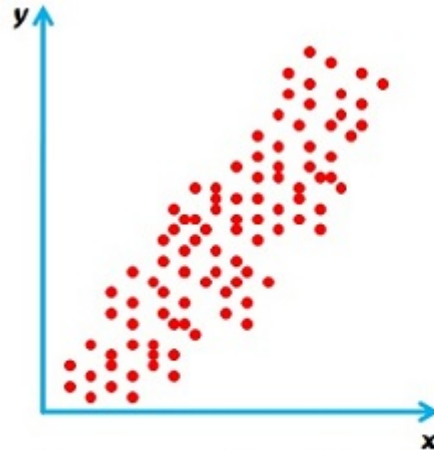


COLLINEARITY IN MIXED MODELS

NOVEMBER 7, 2018

John Hendrickx

Senior Statistical Programmer
Danone Nutricia Research
The Netherlands



COLLINEARITY

Think (almost) identical twins

- Hard to keep apart
- Likely to make mistakes!

Regression models:

- High correlation among independent variables
- High standard error of estimates



COLLINEARITY

Not a big problem in randomised clinical trials

- Classic ANCOVA model with treatment group and baseline Y as independent variables

Newer designs expanded with intercurrent events

Real World Data



COLLINEARITY

Linear regression

- VIF (Variance Inflation)

How much larger is the standard error due to correlations

- Decomposition indices

Better at detecting collinearity among sets of variables



SHORTCOMINGS

Assume unrelated:

- Transformations
- Interactions

Linear models

- Also applicable to generalised linear models?
- Mixed models?
- Generalised mixed models?



PERTURB

Alternative tool

- Add random noise (perturbations) to selected variables
- Recalculate derived terms
- Re-estimate model
- Iterate
- Examine stability of estimates



PERTURB

Ideal for derived terms

Suitable for any regression-like model

Extendable to categorical variables



CATEGORICAL VARIABLES

Categorical variables and collinearity?

- Low frequency → column of zeros with a few ones → mirror image of the intercept
- High percentage → column of ones with a few zeros → almost the same as the intercept



CATEGORICAL VARIABLES

Perturb by reclassifying a certain percentage

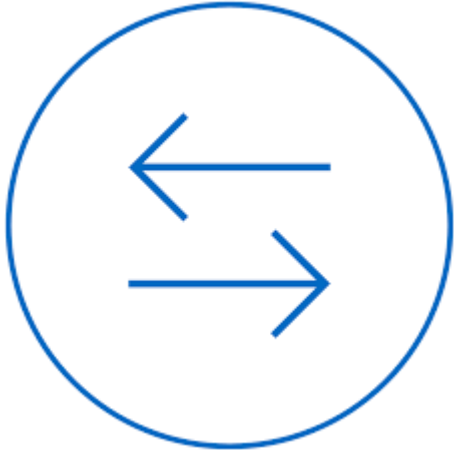
Problem: Large categories shrink, small categories grow



	Reclassified			
Original	a	b	c	total
a	9	0.5	0.5	10
b	3.5	63	3.5	70
c	1	1	18	20
total	13.5	64.5	22	100

CATEGORICAL VARIABLES

Preserve sizes



	Reclassified		
Original	a	b	c
a	0.854	0.096	0.050
b	0.014	0.966	0.020
c	0.025	0.070	0.905

	Reclassified			
Original	a	b	c	total
a	8.54	0.96	0.50	10
b	0.96	67.64	1.40	70
c	0.50	1.40	18.10	20
total	10	70	20	100

MIXED MODELS

Random effects

- Levels variables
- Intraclass correlation

Repeated measures

- Time intervals
- Autocorrelation



PERTURB AND MIXED MODELS

Perturbations in mixed models

- Same basic principle

Adaptations in R

- Different syntax for model formula
- Accessing random effects
- Idiosyncrasies of nlme, lme4



COLLINEARITY IN MIXED MODELS

Fixed and random effects interdependent

Random slope

- Same variable for fixed and random effects

Sparse levels

- Small research sites

Convergence issues due to collinearity

FIXED

RANDOM



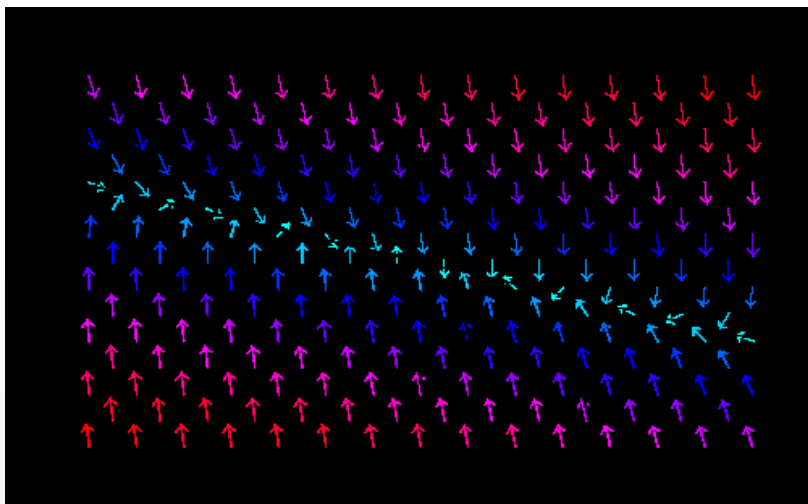
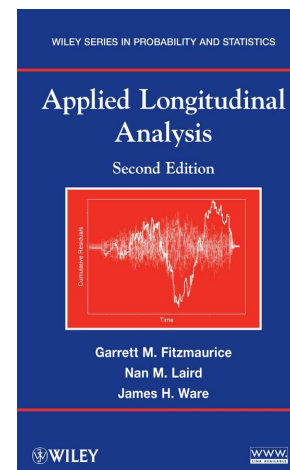
images.jpg

EXAMPLE

From “Applied Longitudinal Analysis”

Random slope and fixed effects of a derived variable

Warning in R “Model is nearly unidentifiable: very large eigenvalue”



EXAMPLE

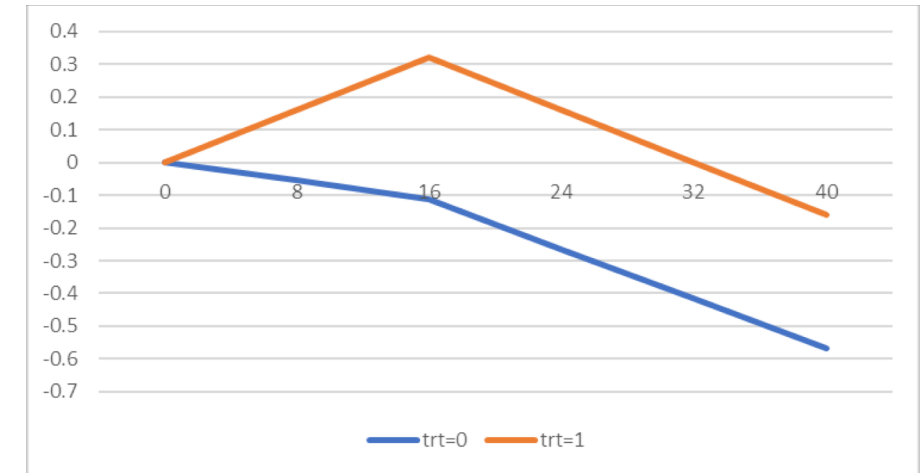
ACTG Study 193A

- Randomised, double-blind study of 1309 AIDS patients
- Dual versus triple combinations of HIV-1 inhibitors
- Dependent variable: log of CD4 counts
- Assessments at baseline, 1 to 9 times during follow-up (40 weeks)

Data shows peak in CD4 counts at 16 weeks

Piecewise linear spline model

- Knot at 16 weeks



MODEL

`lmer(logcd4 ~ week+week_16+trt:week+trt:week_16 + (1+week+week_16 | id),data=cd4)`

logcd4	log of CD4 counts +1
trt	1=triple therapy, 0=dual therapy
week	Assessment week (baseline=0)
week_16	max(0,week-16)
id	Subject identifier

PERTURBATION ANALYSIS

Re-estimate model 100 times

Add perturbations to week at each iteration

- Mean 0, SD 0.5
- 95.4% of perturbations ± 1 week

Derive week₁₆ at each iteration

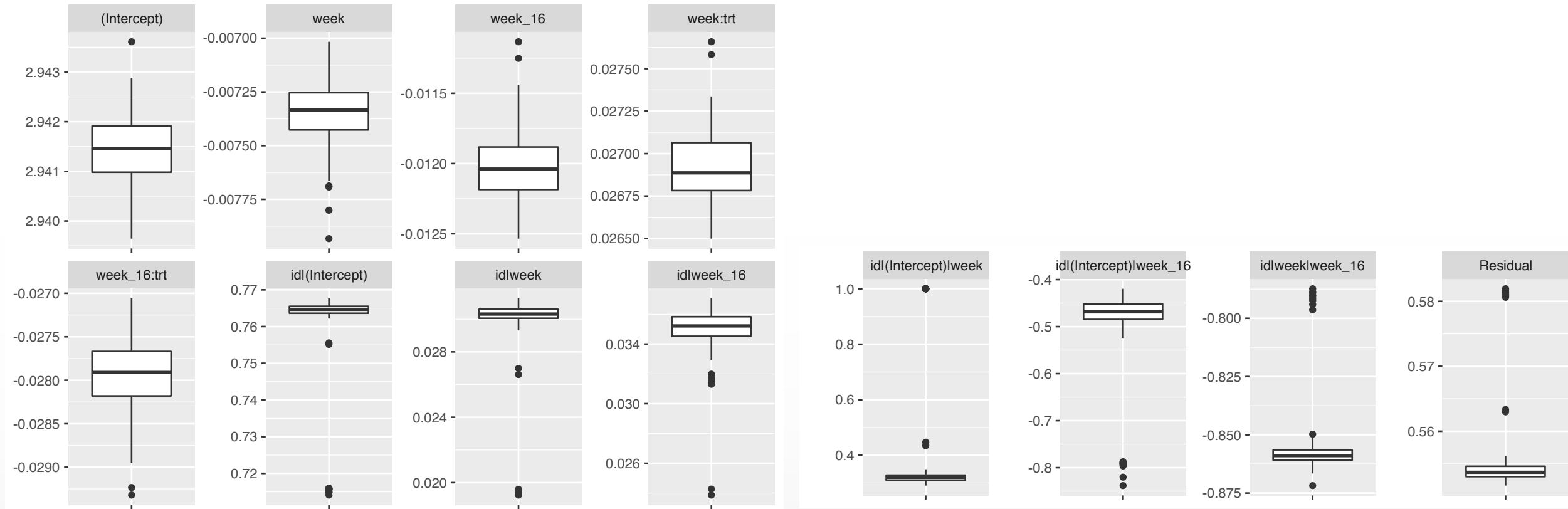
Re-run the model



RESULTS

	mean	s.d.	min	max
(Intercept)	2.941	0.001	2.940	2.944
week	-0.007	0.000	-0.008	-0.007
week_16	-0.012	0.000	-0.013	-0.011
week:trt	0.027	0.000	0.026	0.028
week_16:trt	-0.028	0.000	-0.029	-0.027
id (Intercept)	0.760	0.014	0.714	0.768
id week	0.029	0.003	0.019	0.031
id week_16	0.035	0.002	0.024	0.037
id (Intercept) week	0.380	0.196	0.291	1.000
id (Intercept) week_16	-0.500	0.106	-0.838	-0.419
id week week_16	-0.853	0.020	-0.872	-0.787
Residual	0.556	0.008	0.552	0.582

BOX PLOTS OF PERTURBATION RESULTS



RESULTS

11 warning messages “degenerate Hessian with negative eigenvalues”

9 correlations of 1

- Random intercept/random slope for week

BUT:

- Fixed effects are very stable
- No correlations between 0.45 and 1



CONCLUSION

Model is unstable

- Small changes to the data could lead to different results
- Drop correlation random intercept/random slope
- Fixed effects are stable



WRAPPING UP

Perturbation analysis

- Useful for evaluating mixed models
- Stability of results with changes to data
- Also possible with categorical variables

Downside

- No clear-cut criteria
- What magnitude?
- Experiment!



THANK YOU!



EXAMPLE 2

Small sites (levels variable)

LipiDiDiet trial

- Effects of medical food on prodromal Alzheimer
- Nutricia product (Souvenaid)
- Study set up by a consortium
- Randomised, double-blind, parallel-group trial
- 11 Sites, small sites were pooled



EXAMPLE 2

Dependent variable: CDR-SB (Clinical Dementia Rating Sum of Boxes)

Analysis of change from baseline

Visits at 12 and 24 months

Repeated measures model

- Compound symmetry covariance structure

Random intercept for pooled sites

Baseline Y as a covariate

Baseline MMSE as a covariate



SMALL SITES

Pooled Site	Count	Percent
01	99	35.0%
02	45	15.9%
03	38	13.4%
04	35	12.4%
08	32	11.3%
10	18	6.4%
11	16	5.7%



MODEL

```
mod1<-lme(chg ~ base+trt01p+avismod+avismod:trt01p+blmmse,  
          random=~1|poolsite/usubjid,correlation=corCompSymm(),  
          na.action=na.exclude, data=m_eff)
```

chg	Change from baseline for CDR-SB
trt01p	Planned treatment group (Active vs Control)
avismod	Visit (values 12, 24)
base	Baseline Y
blmmse	Baseline MMSE
poolsite	Site (small sites pooled)
usubjid	Subject id

RECLASSIFY

Pooled site reclassified

- 5% overall probability of reclassification
- Large sites have higher probability, small sites lower
- Prevent growth of small sites, shrinkage of large

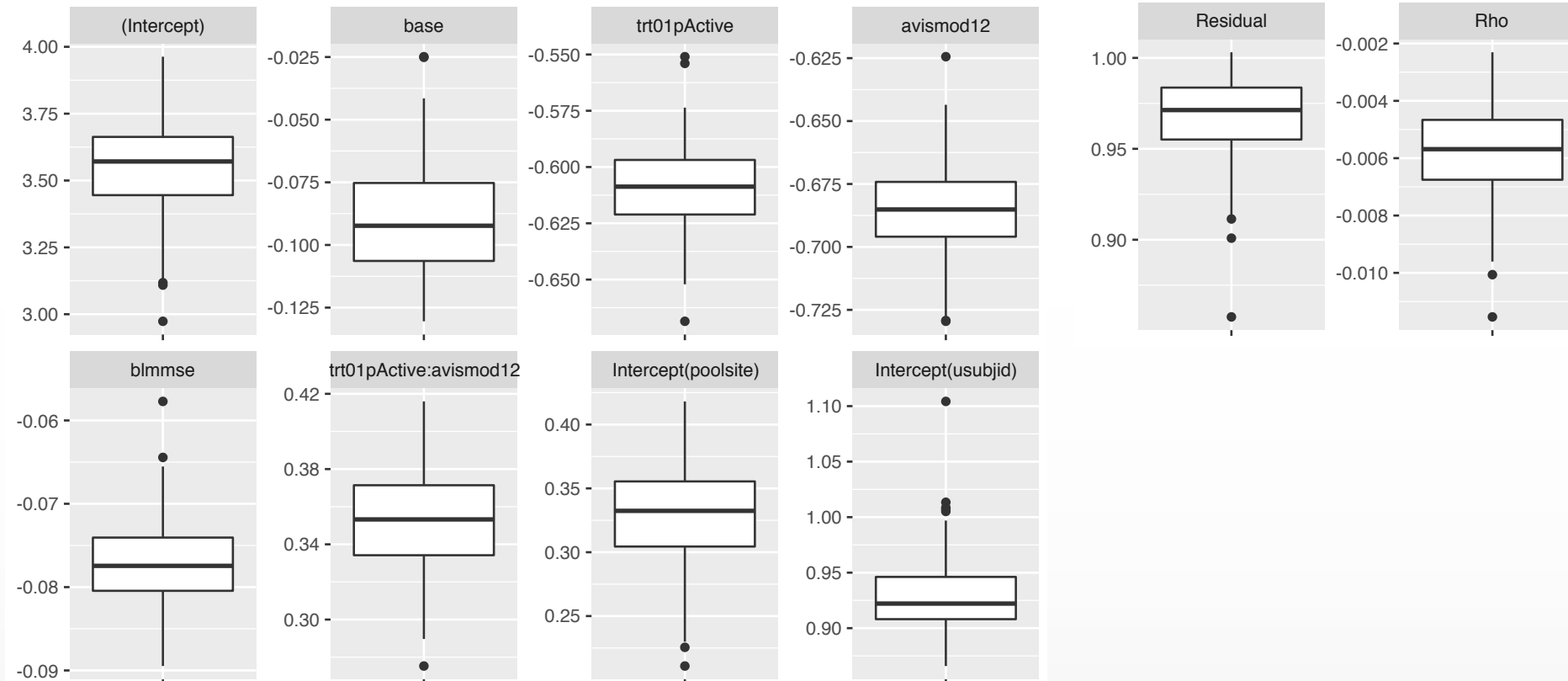
Re-estimate the model 100 times

- Reclassify 5% of cases at each iteration
- Evaluate impact on coefficients

RESULTS

	mean	s.d.	min	max
(Intercept)	3.544	0.186	2.974	3.963
Baseline	-0.089	0.024	-0.130	-0.025
Treatment(Active)	-0.608	0.020	-0.669	-0.551
Visit	-0.686	0.019	-0.730	-0.624
Baseline MMSE	-0.077	0.006	-0.089	-0.058
Treatment(Active):Visit	0.352	0.028	0.275	0.416
Intercept(poolsite)	0.326	0.043	0.211	0.418
CS	0.932	0.036	0.866	1.104
Residual	0.967	0.025	0.858	1.003
Rho	-0.006	0.002	-0.012	-0.002

RESULTS



RESULTS

Intercept is mainly affected

Random intercept for site

- Smaller magnitude

Compound symmetry coefficient

Fixed effects are quite stable

Conclusion:

Reclassification has negligible impact

- More interesting to examine impact on sites without pooling



CATEGORICAL VARIABLES

Loglinear models for square tables

- Preserve sizes
- Suitable for ordered categories

