Longitudinal Data Analysis, including Categorical Outcomes

ICPSR Summer Program2017

Instructor: Donald Hedeker, University of Chicago

Course Description:
This workshop will focus on analysis of longitudinal data using mixed models. Models for continuous outcomes will first be presented, including description of the multilevel or hierarchical representation of the model. Use of polynomials for expressing change across time, treatment of time-invariant and time-varying covariates, and modeling of the variance-covariance structure of the longitudinal outcomes will be described.

For dichotomous, ordinal and nominal outcomes, this workshop will focus next on the mixed logistic regression model, and generalizations of it. Specifically, the following models will be described: mixed logistic regression for dichotomous outcomes, mixed logistic regression for nominal outcomes, and mixed proportional odds and non-proportional odds models for ordinal outcomes. The latter models are useful because the proportional odds assumption of equal covariate effects across the cumulative logits of the model is often unreasonable.

Finally, missing data issues will be covered. Mixed models allow incomplete data across time and assume that these missing observations are "missing at random" (MAR) under maximum likelihood estimation. Approaches that can go further, and don't necessarily assume MAR, are through the use of pattern mixture and selection models. Applications will be described of mixed pattern mixture and selection models.

In all cases, methods will be illustrated using software. Stata will be used for most examples, with some use of SAS, and SuperMix for categorical outcomes.

Prerequisites:
Participants should be thoroughly familiar with multiple linear regression, and have some knowledge of logistic regression.

Text:
Hedeker and Gibbons (2006). Longitudinal Data Analysis, New York: Wiley.

Daily Schedule:
9-12:30 lecture
12:30-1:30 lunch
1:30-5:00 lab/lecture

Class materials are available at

https://hedeker-sites.uchicago.edu/page/longitudinal-data-analysis-including-categorical-outcomes

Course Sequence

Monday August 7 - Continuous Outcomes

Mixed models
Multilevel or HLM representation
Time-varying covariates
Between-subjects and within-subject effects
Reading: Hedeker & Gibbons chapter 4

Slides: Continuous_Mixed_Stata_SAS.pdf

Datasets:
RIESBY.DAT.txt
RIESBYT4.DAT.txt

Stata Files:
riesbydesc.do
riesbyplot.do
riesbymu.do
riesby_mataest.do
riesby_mataest2.do
riesby_obsest_plot.do
riesby_BSWS.do

SAS Files:
riesbydesc.sas
riesbyplot.sas
riesbym.sas
riesbym2.sas
riesbsws.sas

Tuesday August 8 - Continuous outcomes

Covariance pattern models
Selection of variance-covariance matrix
Reading: Hedeker & Gibbons chapter 6

Slides: Continuous_CPM_Stata_SAS.pdf

Dataset: bockrrm.dat.txt

Stata file: bockcpm.do

SAS file: bockcpm.sas

Mixed models with autocorrelated errors
Reading: Hedeker & Gibbons chapter 7

Slides: Continuous_ACerrors_Stata_SAS.pdf

Dataset: bockrrm.dat.txt

Stata file: bockac.do

SAS file: bockac.sas

Wednesday August 9 - Dichotomous Outcomes

Mixed logistic regression model
Threshold concept
Scaling and marginalization of regression parameters
Multilevel or HLM representation
Reading: Hedeker & Gibbons chapter 9

Slides: Binary_Supermix_Stata_SAS.pdf

Dataset: SCHIZX1.DAT.txt

Stata files:
schizb_plots.do
schizb.do

SAS file:
schizb.sas

Thursday August 10 - Ordinal & Nominal Outcomes

Mixed ordinal logistic regression model
Proportional odds assumption
Partial and non-proportional odds models
Mixed nominal logistic regression model
Reference cell and Helmert contrasts
Reading: Hedeker & Gibbons chapters 10 & 11

Slides: Ordinal_Nominal_Supermix_Stata_SAS.pdf

Dataset: SCHIZX1.DAT.txt
sdhouse.DAT.txt

Stata files:
schiz_ordinal_logits.do
schiz_ordinal_mataest.do
schizo.do
sdhouse_ordinal_logits.do
sdo.do
sdhouse_nominal_logits.do
sdn.do

SAS files:
schzofit.sas
sdo.sas
sdn.sas

Friday August 11 - Missing data in longitudinal studies

Missing data mechanisms MCAR, MAR, NMAR
Testing for MCAR vs MAR using grouped-time survival analysis
NMAR models
Pattern-mixture models
Selection models
Reading: Hedeker & Gibbons chapter 14

Slides: MissingData.pdf

Dataset: schizxdrop.dat
schizrep.dat.txt
schizrep_surv.dat

Stata files:
schiz_mcartest.do
schiz_MixGee.do
schiz_PatternMixture.do
schiz_surv.do

SAS files:
SCHZ_MCARtest.sas
schiz_MixGee.sas
schiz_PatternMixture.sas
schiz_SharedParm.sas

Additional topics if time permits:

Sample Size Determination for Longitudinal studies
Reading: Hedeker, D., Gibbons, R.D., & Waternaux, C. (1999). Sample size estimation for longitudinal designs with attrition: comparing time-related contrasts between two groups. Journal of Educational and Behavioral Statistics, 24:70-93. pdf file RMASS2 instructions RMASS2 program (rename file extension to exe to use)

Slides: Pow2g.pdf

Stata files
longt.do
longt_contrast.do
longp.do
ttestpow.do
lregpow.do
longreg.do

SAS files
Ttest_power_N=63.sas
LReg_power_N=137.sas
RandInt_power_N=33.sas
LR_RandInt_power_N=70.sas

Multiple Imputation for binary outcomes under NMAR
Reading: Hedeker, D., Mermelstein, R.J., & Demirtas, H. (2007). Analysis of Binary Outcomes with Missing Data: Missing=Smoking, Last Observation Carried Forward, and a Little Multiple Imputation. *Addiction*, 102:1564-1573. pdf file SAS code SAS code description dataset

Modeling of between- and within-subject variance using mixed location-scale models
Reading: Hedeker, D., Mermelstein, R.J., & Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of Ecological Momentary Assessment (EMA) data. *Biometrics*, 64:627-634. pdf file supplemental materials