

Hunter College

STAT 707 General Linear Models II

Course Content (Spring 2017)

Reference books

1. Advanced Linear Modeling: Multivariate, Time Series, and Spatial Data; Non-parametric Regression and Response Surface Maximization, 2nd edition, by Ronald Christensen.
2. Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd edition, by Stephen W. Raudenbush and Anthony S. Bryk.

Topic List

1. Regression and linear models (2 lectures)
 - Definitions of Chi-square and F -distribution.
 - Orthogonality theorem, Helmert's transformation, Geary's theorem, Fisher's lemma.
 - One-way analysis of variance, two-way analysis of variance without replications. two-way analysis of variance with replications and interaction.
 - Normal equation and QR decomposition, Gauss-Markov theorem.
 - sum-of-square decomposition, trace formula, chi-square decomposition theorem.
 - generalized lest squares and weighted regression, Kolodziejczyk's thoerem,
 - multicollinearity.
 - R functions: `aov()` in stats, `aovp()` in lmPerm,
 - Reference:
 - Box, George E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, **71**, 791-799.
 - Neyman, J., Iwaszkiewicz, K., & Kolodziejczyk, St. (1935). Statistical problems in agricultural experimentation. *Journal of the Royal Statistical Society*, **2**, 107-180.
 - Sabbaghi, A., & Rubin, Donald B. (2014). Comments on the Neyman-Fisher controversy and its consequences. *Statistical Science* **29**, 267-284.

2. Akaike information criterion (AIC) (1 lecture)

- Mean discrimination, AIC, AICc.
- Reference:
 - Akaike, Hirotogu (1973). Information theory and the maximum likelihood principle. in Petrov, B.N., Csaki, F., 2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR, September 2-8, 1971, Budapest: Akademiai Kiad, pp. 267-281.

3. Analysis of covariance (ANCOVA) (0.5 lecture)

- Review: inverse of a partitioned matrix.
- Introducing further explanatory variables, augmented regression model, sum of squares decomposition for additional covariates
- Covariates, concomitant variable, ANCOVA, nested model.

4. Multivariate analysis of variance (MANOVA) (0.5 lecture)

- Homoscedasticity.
- Hotelling's T-square, paired Hotelling's T-square.
- Bartlett's test.
- MANOVA, Wilk's lambda, Hotelling-Lawley's trace, Pillai's trace, Roy's maximum root.
- Reference:
 - Pillai, K. C. S. (1955). Some new criteria in multivariate analysis. *Annals of Mathematical Statistics*, **26**, 117-121.

5. Multilevel analysis: application (5 lectures)

- Multi-stage sampling, primary unit, secondary unit.
- Clustered data, aggregation, disaggregation, intraclass correlation coefficient (ICC), random effect ANOVA model, design effect.
- Random intercept model, nesting structure, fixed and random parameters, empty model, model with one explanatory variable, residual intraclass correlation coefficient, standardized coefficients, model with multiple explanatory variables, within-group regression, between-group regression, within-group deviation score, parameter estimation (maximum likelihood, restricted/residual maximum likelihood, empirical Bayes), three-level model.
- Hierarchical linear model, random slopes, heteroscedasticity, grand-mean centered variable, cross-level interaction, latent regression.

- Tests for fixed parameters, multi-parameter tests for fixed effect, Wald test, deviance test, halved p -value for variance parameters, model specification.
- Assumptions of multilevel modeling, cross-validation, model checks, specification of the fixed and the random part, transformation, tests for heteroscedasticity based on standardized residual dispersions of weighted average of the logarithms of estimated residual variance, inspection of level-1 residuals.
- Crossed random effects and nested random effects, crossed random effect modeling, handling crossed random effects by using dummy variables, examples of crossed factor modeling (multiple roles, social network).
- Social network, reciprocity, popularity, activity/outgoingness, directed and undirected relations.
- Longitudinal data analysis, fixed occasion designs and variable occasion designs, compound symmetry model, fully multivariate model, incomplete paired data.
- Multivariate multilevel model.
- R functions: `lme()` in `nlme`, `lmer()`, `glmer()` in `lme4`.
- Reference:
 - Raudenbush, Stephen W., & Bryk, Anthony S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*, 2nd edition, Sage Publications.
 - Woltman, H., Feldstain, A., MacKay, J. C., and Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology*, **8**, 52-69.
 - Hausman, J. A., & Taylor W. E. (1981). Panel data and unobservable individual effects. *Econometrica*, **49**, 1377-1398.
 - Kreft, I. G. G., de Leeuw, J., & Aiken, L. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, **30**, 1-22.
 - Self, G. S., & Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, **82**, 605-610.
 - Snijders, T. A. B., & Bosker, R. J. (1994). Modeled variance in two-level models. *Sociological Methods and Research*, **22**, 342-363.
 - Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, **73**, 43-56.
 - Goldstein, H. (1987). Multilevel covariance component models. *Biometrika*, **74**, 430-431.

6. Generalized estimating equation (GEE) (1 lecture)

- Comparison between GLMM and GEE.
- Review: GLM and quasi-likelihood, link function, information, Fisher's scoring method.
- Marginal models and random effect models.
- GEE estimate of model coefficients, information sandwich estimate, covariance matrix structure (independent, exchangeable, AR1, unstructured).
- R functions: `gee()` in `gee`, `geeglm()` in `geepack`.
- Reference:
 - Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22.
 - Zeger, S. L., Liang, K.-Y., & Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, **44**, 1049-1060.

7. Kriging (1 lecture)

- Non-parametric regression in spatial problems in multiple dimensions.
- Tobler's first law of Geography.
- Inverse distance weighted (IDW) interpolation, semi-variogram, coefficient solving by using Lagrange multiplier, sill, partial sill, range, nugget.
- Simple Kriging and ordinary Kriging.
- R functions: `krige()`, `vgm()` in `gstat` and `sp`.
- Reference:
 - Christensen, Ronald (2001). *Advanced Linear Modeling: Multivariate, Time Series, and Spatial Data; Nonparametric Regression and Response Surface Maximization, 2nd edition*. Springer.
 - Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *semiparametric Regression*. Cambridge University Press.

8. Causal inference (1.5 lecture)

- Simpson's paradox, Cornfield's inequality, relative risk, risk ratio.
- Propensity score, balancing score, estimation of propensity score, propensity score matching.
- Reference:
 - Cornfield, J. et al. (1959). Smoking and lung cancer: recent evidence and a discussion of some questions. *Journal of the National Cancer Institute*, **22**, 173-203.

- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role fo the propensity score in observational studies for causal effects. *Biometrika*, **70**, 41-55.

9. Principal components analysis (0.5 lecture)

- Rayleigh quotient, PCA.
- Reference:
 - Christensen, Ronald (2001). *Advanced Linear Modeling: Multivariate, Time Series, and Spatial Data; Nonparametric Regression and Response Surface Maximization*, 2nd edition. Springer.

10. Group sequential methods (0.5 lecture)

- Pocock't test, O'Brien & Fleming's test.
- R functions: `refinedBoundary()` in `gsrsb`.
- Reference:
 - O'Brien, P. C., & Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics* **35**, 549-556.
 - Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika*, **64**, 191-199.

11. Final review (0.5 lecture)