

ACM/ESE 118 (Winter 2007)

Methods in Applied Statistics and Data Analysis

Lectures Tuesdays and Thursdays 10:30–11:55 am, 155 Arms
Recitation Wednesdays 3–4 pm, 162 S. Mudd

www.gps.caltech.edu/~tapio/acm118/

- Week 1:** Simple linear regression (least squares estimation, analysis of residuals)
- 2:** Inferences about model parameters, multiple linear regression
 - 3:** Analysis of variance, comparison of models, model selection
 - 4:** Assessing goodness-of-fit, outliers, influential observations
 - 5:** Collinearity and rank-deficiency, singular value decomposition, regularization
 - 6:** Ridge regression, regularization by truncated singular value decomposition
 - 7:** Choosing regularization parameters (generalized cross-validation, L-curve)
 - 8:** Principal component analysis, linear discriminant analysis
 - 9:** Hierarchical cluster analysis
 - 10:** Resampling methods and the bootstrap

Texts (on reserve in Sherman-Fairchild Library):

- Weisberg, S, 2005: *Applied Linear Regression*, 3rd ed. Wiley. (**Required.**)
- Efron, B., and R. J. Tibshirani, 1993: *An Introduction to the Bootstrap*. Chapman & Hall.
- Hansen, P. C., 1998: *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. Society for Industrial and Applied Mathematics.
- Johnson, R. A., and D. W. Wichern, 2002: *Applied Multivariate Statistical Analysis*, 5th ed. Prentice Hall.
- Venables, W. N., and B. D. Ripley, 2002: *Modern Applied Statistics with S*, 4th ed. Springer.

Homework: Homework will be distributed on Thursdays and is due in class the following Thursday

Office hours: Agostino Capponi (acapponi@caltech): Wednesdays 6:30–7:30, 160 Jorgensen
Roger Donaldson (rdonald@acm): Mondays 1:30–2:30, Red Door
Hannes Helgason (hannes@acm): Wednesdays 2–3, 226 Guggenheim
Tapio Schneider (tapio@gps): Tuesdays 3–4, 112 N. Mudd