

Regression Methods

© 2024, A. C. Davison

Ecole Polytechnique Fédérale de Lausanne
EPFL-FSB-MATH-STAT
Station 8, 1015 Lausanne, Switzerland
<http://stat.epfl.ch>

Everything below is approximate. Please signal any corrections to Anthony.Davison@epfl.ch.

Course web page

Further material, links to data sets, etc., can be found on the course moodle page.

Lecture contents

1. Linear fits/reminders (10/9/24; slides 1–23)
 - Motivation. Linear model. Least squares and decomposition of variance. (1–8)
 - Analysis of variance. ANOVA table. R^2 . Reminders on distributions etc. (9–25)
2. Linear statistics (17/9/24; slides 26–41)
 - Second-order and normal assumptions. Distributional results. (26–29)
 - Distributional results for normal-theory linear model and analysis of variance. Prediction, orthogonality, balance. (30–41)
3. Analysis of variance/diagnostics (24/9/24; slides 40–52)
 - Revision of orthogonality/balance. Diagnostics (start). (40–44)
 - Diagnostic plots, leverage and influence, QQplots. (45–52)
4. Transformation/Model building (1/10/24; slides 53–70)
 - Box–Cox transformation, poisons example. Model building I: generalities (53–62)
 - Variable selection: prediction error and cross-validation (63–70)
5. Variable selection/General models (8/10/24; slides 71–88)
 - AIC and stepwise selection; example. (71–77)
 - Robustness and estimating equations. Quantile and expectile regression. (78–88)
6. General models/IWLS (15/10/24; slides 89–111)
 - General models and motivation. (Likelihood revision.) IWLS. (89–104)
 - Deviance. Diagnostics in general models. (105–111)

7. Generalised linear models II (29/10/24; slides 116–135)
 - GLM densities. IWLS for GLMs. Estimation of ϕ . Jacamar example. (116–127)
 - Chimp example. Binary responses: tolerance distributions and link functions. (128–135)
8. Generalised linear models III (5/11/24; slides 136–160)
 - Nodal involvement example. Count data: Poisson and multinomial distribution. Poisson regression. Premier League example. (136–153).
 - Multinomial data, Poisson response models. Jacamar example. (154–160)
9. Ordinal responses/overdispersion (12/11/24; slides 161–186)
 - Ordinal responses. Overdispersion. Parametric models. (161–174)
 - Quasilikelihood (end). Regularisation; Cox/Battey; collinearity. (175–186)
10. Regularisation (19/11/24; slides 187–201)
 - Bayesian linear model; empirical Bayes estimation; EDF. (187–193)
 - Choice of penalty by CV, GCV, REML. Ridge regression; cement example. (194–201)
11. Scatterplot smoother/Lasso (26/11/24; slides 202–223)
 - Scatterplot smoother; roughness penalty; motorcycle example; knots. (202–212)
 - L_p penalties and geometry. Lasso algorithm, thresholding. Example (213–223)
12. Splines/General approach (3/12/24; slides 225–261)
 - Splines, construction and approaches; natural cubic splines. (225–238)
 - General formulation of smoothers; inference; spring barley example. (239–261)
13. Components of variance/Mixed models (10/12/24; slides 262–285)
 - Examples. Nested and crossed effects. Interpretation of ANOVA, estimation of random effects. (262–271)
 - Mixed model and its likelihood. Estimation, prediction of random effects. Rat data example. (272–285)
14. Generalized additive models (17/12/24; slides 286–302)
 - GAM, PIWLS, iterations, Laplace approximation (286–294)
 - REML for estimation of λ ; AIDS data example (295–302)

Regression Methods: Bibliography

References

- Atkinson, A. C. (1985) *Plots, Transformations, and Regression*. Oxford: Clarendon Press.
- Bates, D. M. and Watts, D. G. (1988) *Nonlinear Regression Analysis and its Applications*. New York: Wiley.
- Chatterjee, S. and Hadi, A. S. (1988) *Sensitivity Analysis in Linear Regression*. New York: Wiley.
- Davison, A. C. (2003) *Statistical Models*. Cambridge: Cambridge University Press.
- Davison, A. C. and Tsai, C.-L. (1992) Regression model diagnostics. *International Statistical Review* **60**, 337–353.
- Dobson, A. J. (1990) *An Introduction to Generalized Linear Models*. London: Chapman & Hall.
- Firth, D. (1991) Generalized linear models. In *Statistical Theory and Modelling: In Honour of Sir David Cox, FRS*, eds D. V. Hinkley, N. Reid and E. J. Snell, pp. 55–82. London: Chapman & Hall.
- Green, P. J. and Silverman, B. W. (1994) *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. London: Chapman & Hall.
- Hastie, T. J., Tibshirani, R. J. and Friedman, J. H. (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second edition. New York: Springer.
- McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*. Second edition. London: Chapman & Hall.
- Rawlings, J. O. (1988) *Applied Regression Analysis: A Research Tool*. Pacific Grove, California: Wadsworth & Brooks/Cole.
- Ruppert, D., Wand, M. P. and Carroll, R. J. (2003) *Semiparametric Regression*. Cambridge: Cambridge University Press.
- Seber, G. A. F. (1977) *Linear Regression Analysis*. New York: Wiley.
- Seber, G. A. F. and Wild, A. J. (1989) *Nonlinear Regression*. New York: Wiley.
- Sen, A. and Srivastava, M. (1990) *Regression Analysis: Theory, Methods and Applications*. New York: Springer.
- Simonoff, J. S. (1996) *Smoothing Methods in Statistics*. New York: Springer.
- Weisberg, S. (1985) *Applied Linear Regression*. Second edition. New York: Wiley.
- Wood, S. N. (2017) *Generalized Additive Models: An Introduction with R*. Second edition. Boca Raton: Chapman and Hall/CRC.