Statistics 203: Introduction to Regression and Analysis of Variance

Course review

Jonathan Taylor

Today

- Review / overview of what we learned.

General themes in regression models

- Specifying regression models.
  - What is the joint (conditional) distribution of *all* outcomes given *all* covariates?
  - Are outcomes independent (conditional on covariates)? If not, what is an appropriate model?
- Fitting the models.
  - Once a model is specified how are we going to estimate the parameters?
  - Is there an algorithm or some existing software to fit the model?
- Comparing regression models.
  - Inference for coefficients in the model: are some zero (i.e. is a smaller model better?)
  - What if there are two *competing* models for the data? Why would one be preferable to the other?
  - What if there are *many* models for the data? How do we compare models for the data?

Simple linear regression model

- Only one covariate
- $$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \qquad 1 \leq i \leq n$$
- Errors ε are independent $N(0, \sigma^2)$.

Multiple linear regression model

- $$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_{p-1} X_{i,p-1} + \varepsilon_i, \qquad 1 \le i \le n$$
- Errors ε are independent $N(0, \sigma^2)$.
- β_j's: (partial) regression coefficients.
- Special cases: polynomial / spline regression models where extra columns are functions of one covariate.

ANOVA & categorical variables

- Generalization of two-sample tests
- One-way (fixed)

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \qquad 1 \leq i \leq r, 1 \leq j \leq n$$

  α's are constants to be estimated. Errors ε_{ij} are independent $N(0, \sigma^2)$.
- Two-way (fixed):

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \qquad 1 \leq i \leq r, 1 \leq j \leq m, 1 \leq k \leq n$$

  α's, β's, $(\alpha\beta)$'s are constants to be estimated. Errors ε_{ijk} are independent $N(0, \sigma^2)$.
- Experimental design: when balanced layouts are impossible, which is the best design?

Generalized linear models

- Non-Gaussian errors.
- Binary outcomes: logistic regression (or probit).
- Count outcomes: Poisson regression.
- A "link" and "variance" function determine a GLM.
- Link:

$$g(\mathbb{E}(Y_i)) = g(\mu_i) = \eta_i = x_i\beta = X_{i0}\beta_0 + \cdots + X_{i,p-1}\beta_{p-1}.$$

- Variance function:

$$\text{Var}(Y_i) = V(\mu_i).$$

Nonlinear regression models

- Regression function depends on parameters in a nonlinear fashion.
- $$Y_i = f(X_{i1}, \dots, X_{ip}; \theta_1, \dots, \theta_q) + \varepsilon_i, \qquad 1 \leq i \leq n$$
- Errors ε_i are independent $N(0, \sigma^2)$.

Robust regression

- Suppose that we have additive noise, but not Gaussian. Likelihood of the form

$$L(\beta|Y, X_0, \dots, X_{p-1}) \propto \exp\left(-\rho\left(\frac{Y - \sum_{j=0}^{p-1}\beta_j X_j}{s}\right)\right)$$

- Leads to robust regression

$$\sum_{i=1}^{n}\rho\left(\frac{Y_i - \sum_{j=0}^{p-1} X_{ij}\beta_j}{s}\right).$$

- Can downweight residuals with bigger tails than normal random variables.

Random & mixed effects ANOVA

- When the levels of the categorical variables in an ANOVA are a sample from a population, effects should be treated as random.
- One-way (random):

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \qquad 1 \leq i \leq r, 1 \leq j \leq n$$

  where $\alpha_i \sim N(0, \sigma^2_\alpha)$ are random, independent of the errors ε_{ij} which are independent $N(0, \sigma^2)$.
- Introduces correlation in the Y's:

$$\text{Cov}(Y_{ij}, Y_{i'j'}) = \delta_{ii'} \left(\sigma^2_\alpha + \delta_{jj'}\sigma^2\right).$$

Mixed linear models

- Essentially a model of covariance between observations based on "subject" effects.
- General form:

$$Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + Z_{n\times q}\gamma_{q\times 1} + \varepsilon_{n\times 1}$$

  where
  - $\varepsilon \sim N(0, \sigma^2 I)$;
  - $\gamma \sim N(0, D)$ for some covariance D.
- In this model

$$Y \sim N(X\beta, ZDZ' + \sigma^2 I).$$

- Covariance is modelled through "random effect" design matrix Z and covariance D.

Time series regression models

- Another model of covariance between observations, based on dependence in time.
- $$Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \varepsilon_{n\times 1}$$
- In these models, $\varepsilon \sim N(0, \Sigma)$ where the covariance Σ depends on what kind of time series model is used (i.e. which $ARMA(p,q)$ model?).
- Example, if ε is $AR(1)$ with parameter ρ then
$$\Sigma_{ij} = \sigma^2 \rho^{|i-j|}.$$

Functional linear model

- We talked about a functional two-sample t-test.
- General form

$$Y_{i,t} = \beta_{0,t} + \sum_{j=1}^{p-1} X_{ij}\beta_{j,t} + \varepsilon_{i,t}$$

  where the noise $\varepsilon_{i,t}$ is a random function, independent across "observation" (curves) $Y_{i,\cdot}$.
- Parameter estimates are curves: leads to nice inference problems for smooth random curves.

Least squares

- Multiple linear regression – OLS

$$\widehat{\beta} = (X^t X)^{-1} X^t Y.$$

- Non-constant variance but independent – WLS

$$\widehat{\beta} = (X^t W X)^{-1} X^t W Y, \qquad W_i = 1/\text{Var}(Y_i)$$

- General correlation – GLS

$$\widehat{\beta} = (X^t \Sigma^{-1} X)^{-1} X^t \Sigma^{-1} Y, \qquad \text{Cov}(Y) = \sigma^2 \Sigma.$$

Maximum likelihood

- In the Gaussian setting, with Σ known least squares is MLE.
- In other cases, we needed iterative techniques to solve MLE:
  - nonlinear regression: iterative projections onto the tangent space;
  - robust regression: IRLS with weights determined by $\psi = \rho'$
  - generalized linear models: IRLS, Fisher scoring
  - time series regression models: two-stage procedure approximates MLE (can iterate further, though)
  - mixed models: similar techniques (though we skipped the details)

Diagnostics: influence and outliers

- Diagnostic plots:
  - Added variable plots.
  - QQplot.
  - Residuals vs. fitted.
  - Standardized residuals vs. fitted.
- Measures of influence
- Cook's distance.
- $DFFITS$.
- $DFBETAS$.
- Outlier test with Bonferroni correction.
- Techniques are most developed for multiple linear regression model, but some can be generalized (using "whitened" residuals).

Penalized regression

- We looked at ridge regression, too.
- A generic example of the “bias-variance” tradeoff in statistics.
- Minimize

$$SSE_\lambda(\beta) = \sum_{i=1}^n \left(Y_i - \sum_{j=1}^{p-1} X_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p-1} \beta_j^2.$$

- Other penalties possible: basic idea is that the penalty is a measure of “complexity” of the model.
- Smoothing spline: ridge regression for scatterplot smoothers.

Hypothesis tests: multiple linear regression

- Multiple linear regression: model R has j less coefficients than model F – equivalently there are j linear constraints on β's.
- $$F = \frac{\frac{SSE(R)-SSE(F)}{j}}{\frac{SSE(F)}{n-p}} \sim F_{j,n-p} (\text{if } H_0 \text{ is true})$$
- Reject $H_0 : R$ is true at level α if $F > F_{1-\alpha,j,n-p}$.

Hypothesis tests: general case

- Other models: $DEV(\mathcal{M}) = -2\log L(\mathcal{M})$ replaces $SSE(\mathcal{M})$.
- Difference $D = DEV(R) - DEV(F) \sim \chi^2_j$ (asymptotically).
- Denominator in the F statistic is usually either known or based on something like Pearson's X^2:

$$\widehat{\phi} = \frac{1}{n-p}\sum_{i=1}^n r_i(F)^2.$$

  In general, residuals are "whitened"

$$r(\mathcal{M}) = \Sigma(\mathcal{M})^{-1/2}(Y - X\widehat{\beta}(\mathcal{M})).$$

- Reject $H_0 : R$ is true at level α if $D > \chi^2_{1-\alpha,n-p}$.

Model selection: AIC, BIC, stepwise

- Best subsets regression (`leaps`): adjusted R^2, C_p.
- Akaike Information Criterion (AIC)

$$AIC(\mathcal{M}) = -2\log L(\mathcal{M}) + 2 \cdot p(\mathcal{M}).$$

  in model $\mathcal{M}$ evaluated at the MLE (Maximum Likelihood Estimators).

- Schwarz's Bayesian Information Criterion (BIC)

$$BIC(\mathcal{M}) = -2\log L(\mathcal{M}) + p(\mathcal{M}) \cdot \log n$$

- Penalized regression can be thought of as model selection as well: choosing the "best" penalty parameter on the basis of

$$GCV(\mathcal{M}) = \frac{1}{\text{Tr}(S(\mathcal{M}))} \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i(\mathcal{M})\right)^2$$

  where

$$\widehat{Y}(\mathcal{M}) = S(\mathcal{M})Y.$$

That's it!

Thanks!