MONASH University

MONASH BUSINESS SCHOOL

ETC3580: Advanced Statistical Modelling

Week 12: Review

Key Concepts

- Sample space of conditional response (Reals, Positive Reals, Non-negative counts, 0/1, proportions, . . .)
- Nature of predictors (numerical, categorical, random)
- Relationship between response and predictors (linear, additive, interacting, nesting)

1. Linear models
2. Generalized linear models
3. Linear mixed-effect models
4. Generalized additive models

Data Analysis Skills

- Interpreting graphs
- Appropriate plots to highlight relationships (e.g., interactions)
- Transformations to allow linear models
- Transformations to limit prediction space
- Dealing with outliers
- Bootstrap
- Permutation tests

Linear Models

- Interactions between 2 numerical predictors
- Interactions between a numerical and categorical predictor
- Outliers
- Leverage
- Cook's distance
- Residual diagnostics
- Conditional variable plots
- F-tests
- QQ plots
- LOO residuals
- Cross-validation

Generalized Linear Models

- Logistic regression, log-odds interpretations
- Binomial regression
- Poisson regression
- Negative binomial regression
- Beta regression
- Over/under dispersion and quasi likelihood
- Deviance residuals
- χ^2 tests for deviance
- Confidence intervals using profile likelihood regions
- AIC
- Zero-inflated models
- Exponential family distributions
- Link functions, canonical link functions

Linear Mixed-Effects Models

- When to use a random effect?
- Nesting and grouping
- Panel/longitudinal data
- REML and MLE
- Interpreting a fitted model
- Bootstrap tests

Generalized Additive Models

- Kernel and local polynomial regression
- Smoothing splines
- Regression splines
- Penalized regression splines
- Mixed model representation of penalized regression splines
- Curse of dimensionality
- Linear smoothers and degrees of freedom
- F test for linearity
- GCV and smoothness selection
- Additive models
- Generalized additive models

Exam

- Closed book, calculators permitted. But only HP 10b11+ calculators as per faculty policy.
- Five questions: 16 + 25 + 16 + 28 + 15 = 100 marks.
- You will need to interpret lots of R output, but only 1 mark is associated with writing R code.
- Focus on data analysis, modelling and interpretation.
- Emphasis on what we actually use for modelling, not on the theory that led up to it.

Theory Questions

- Show something is in the exponential family
- Derive the canonical link function
- Write down the model given the R output, or a description of response and predictors.

Consultation Times

- Tomorrow, 4–5pm
- Monday 23rd, 4–5pm
- Tuesday 24th, 4–5pm
- Monday 30th, 4–5pm
- Tuesday 31st, 4–5pm

EXAM: Tuesday 7 November