Applied Generalized Linear Models (FS 20)

ANOVA and ANCOVA (practical)

Viviana Amati

Course structure - schedule

Date	Topic	Assignment
18.02	Introduction to the course	Ass. 1 released on 25.02 due to 19.03
25.02	Introduction to R and review of the linear regression mode	
03.03	The general linear model: ANOVA and ANCOVA	
10.03	**Practical: ANOVA and ANCOVA**	
17.03	Binary outcomes: logistic regression and probit models	Ass. 2 released on 17.03 due to 23.04
24.03	Practical: logistic regression and probit models	
31.03	Nominal outcomes: multinomial logistic regression	
07.04	Practical: multinomial logistic regression	
21.04	Ordinal outcomes: ordered logistic regression and probit models	Ass. 3 released on 21.04 due to 21.05
28.04	Practical: ordered logistic regression and probit models	
05.05	Count outcomes: Poisson and negative binomial models	
12.05	Practical: Poisson and negative binomial models	
19.05	Survival models (lecture+practical)	
26.05	Exam	

Recap: ANOVA

Term used in two different contexts:

- LRMs: the partition of the SST into the SSReg and SSR

$$\underbrace{\sum_{i=1}^{n}(y_i - \overline{y})^2}_{\substack{\text{Total sample variability} \\ \text{(SST)}}} = \underbrace{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}_{\substack{\text{Explained variability} \\ \text{(SSReg)}}} + \underbrace{\sum_{i=1}^{n}\hat{e}_i^2}_{\substack{\text{Residual variability} \\ \text{(SSR)}}},$$

- Design of experiments: statistical methods for testing and fitting linear models in which the explanatory variables are categorical
  - categorical variables are referred to as factors and their categories as levels
  - experiments aim to test whether one (or more factors) have an effect on an outcome variable

Recap: One-way ANOVA

$$Y_{ij} = \mu + \alpha_j + \varepsilon_{ij} \quad ,$$

with

- μ: the population mean of Y
- α_j: effect on the dependent variable in the j-th group
- ε_{ij}: error term
  - independent and normally distributed
  - $E[\varepsilon_{ij}] = 0$ and $\mathrm{Var}[\varepsilon_{ij}] = \sigma^2$

Implication:

$$Y_{ij} \sim N(\mu + \alpha_j, \sigma^2)$$

Recap: One-way ANOVA inference

- The least square estimates for μ and α_j are given by:

$$\hat{\mu} = \frac{1}{M} \sum_{j=1}^{M} \overline{y}_j = \overline{y} \ .$$

$$\hat{\alpha}_j = \overline{y}_j - \overline{y}.$$

  and the fitted values are the group means

$$\hat{y}_{ij} = \hat{\mu} + \hat{\alpha}_j = \overline{y}. + \overline{y}_j - \overline{y}. = \overline{y}_j$$

- We would like to test:

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_{M-1} = 0 \qquad (H_1 : \text{at least one } \alpha_j \text{ differs})$$

Recap: One-way ANOVA table

	Sum of squares (SS)	df	Mean Square (MS)	F-test
Factor	$\sum_{j=1}^{M} n_j(\overline{y}_j - \overline{y})^2$	$M-1$	$\frac{\text{SSReg}}{M-1}$	$\frac{\text{MSReg}}{\text{MSR}}$
Residuals	$\sum_{j=1}^{M}\sum_{i=1}^{n_j}(y_{ij} - \overline{y}_j)^2$	$N-M$	$\frac{\text{SSR}}{N-M}$	
Total	$\sum_{j=1}^{M}\sum_{i=1}^{n_j}(y_{ij} - \overline{y})^2$	$N-1$		

One-way ANOVA: example

Material:

- The folder anova.zip contains the data and the script to illustrate ANOVA and ANCOVA
- The description of the data is in the script as well as in the lecture notes
- The commented output is in the lecture notes (upload tomorrow)

ANCOVA

LRMs with both qualitative and quantitative explanatory variables

$$Y_{ij} = \mu + \alpha_j + \beta(x_{ij} - \overline{x}) + \varepsilon_{ij} \quad ,$$

with

- Y_{ij}: i-th value on Y taken under the j-th level
- μ: overall (population) mean
- α_j: effect of the j-th level
- x_{ij}: quantitative explanatory variable corresponding to y_{ij}
- $\overline{x}$: mean of x_{ij}
- β: coefficient describing the dependence of Y on X
- ε_{ij}: error term
  - independent and normally distributed
  - $E[\varepsilon_{ij}] = 0$ and $\mathrm{Var}[\varepsilon_{ij}] = \sigma^2$