Tutorial 4

STAT3015/4030/7030 Generalised Linear Modelling

The Australian National University

Week 4, 2017

Overview

Analysis of Covariance (ANCOVA) model

An ANCOVA model has a continuous predictor x_{ij} and a categorical predictor f_i:

$$Y_{ij} = \mu_i + \beta x_{ij} + \varepsilon_{ij}.$$

With a k level f_i, using the constraint $\tau_1 = 0$ and $k - 1$ indicator variables:

$$\begin{aligned} Y_{ij} =& \mu + \tau_i + \beta x_{ij} + \varepsilon_{ij} \\ =& \beta_0 + \beta_1 z_{1,ij} + \cdots + \beta_{k-1} z_{k-1,ij} + \beta_k x_{ij} + \varepsilon_{ij}, \end{aligned}$$

where $\beta_0 = \mu, \beta_k = \beta$ and $z_{l,ij}$ is the indicator variable for the $(l+1)^{st}$ factor level.

ANCOVA model without interaction terms

A simple ANCOVA model with a continuous predictor x and a factor α:

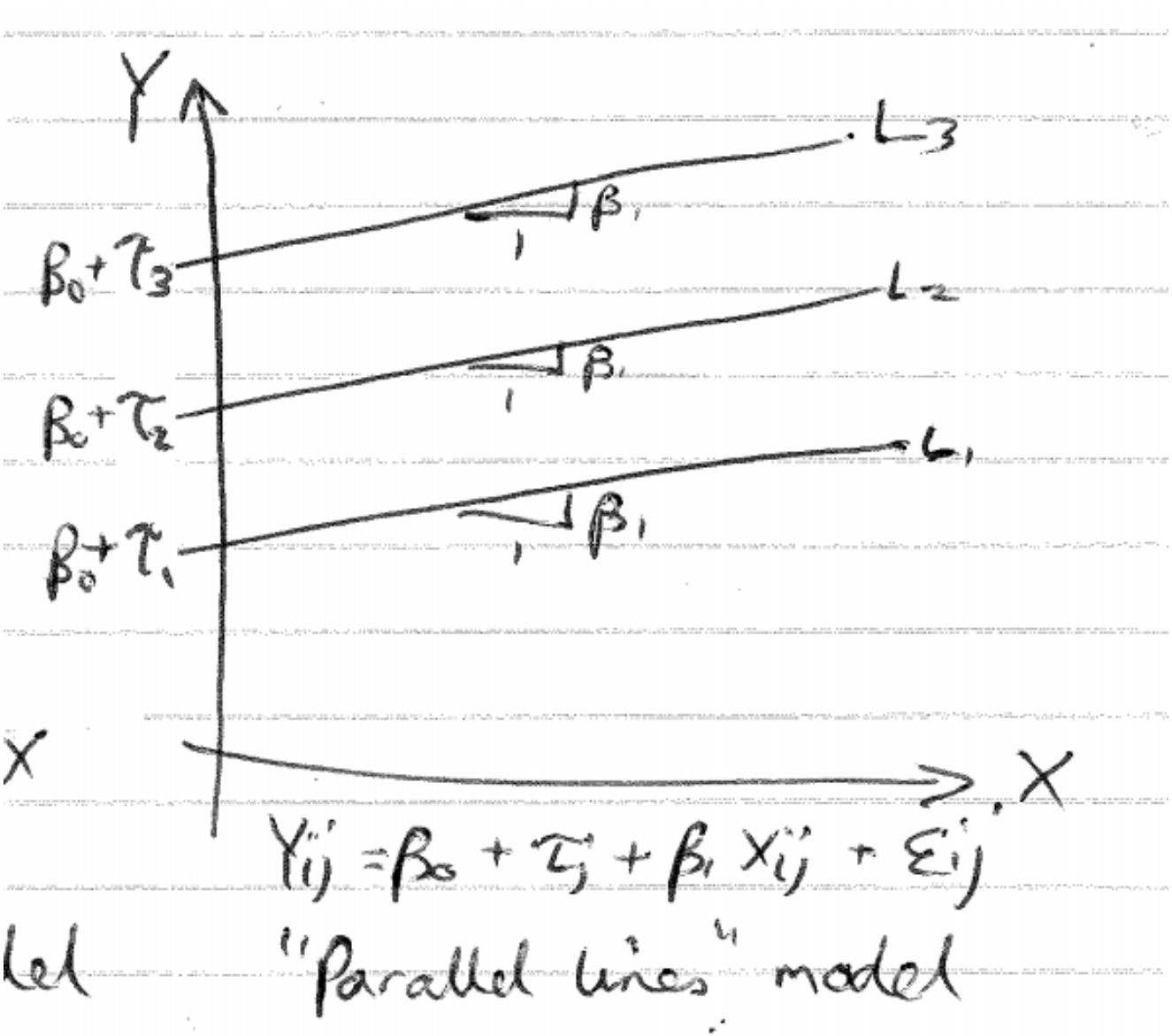
$$Y_{ij} = \beta_0 + \alpha_i + \beta_1 x_{ij} + \varepsilon_{ij}.$$

Using R to fit the above model:

`lm(Y ~ Continuous + factor(categorical))`

(Choose the appropriate treatment in R. Use `options()$contrast`)

Diagram


Y
L_3
β_1
1
$\beta_0 + \tau_3$
L_2
β_1
1
$\beta_0 + \tau_2$
L_1
β_1
1
$\beta_0 + \tau_1$
X
X
$Y_{ij} = \beta_0 + \tau_j + \beta_1 X_{ij} + \varepsilon_{ij}$
"Parallel lines" model

ANCOVA model with an interaction term

An ANCOVA model with a continuous predictor x, a factor α and also an interaction term:

$$Y_{ij} = \beta_0 + \alpha_i + \beta_1 x_{ij} + \gamma_i x_{ij} + \varepsilon_{ij}.$$

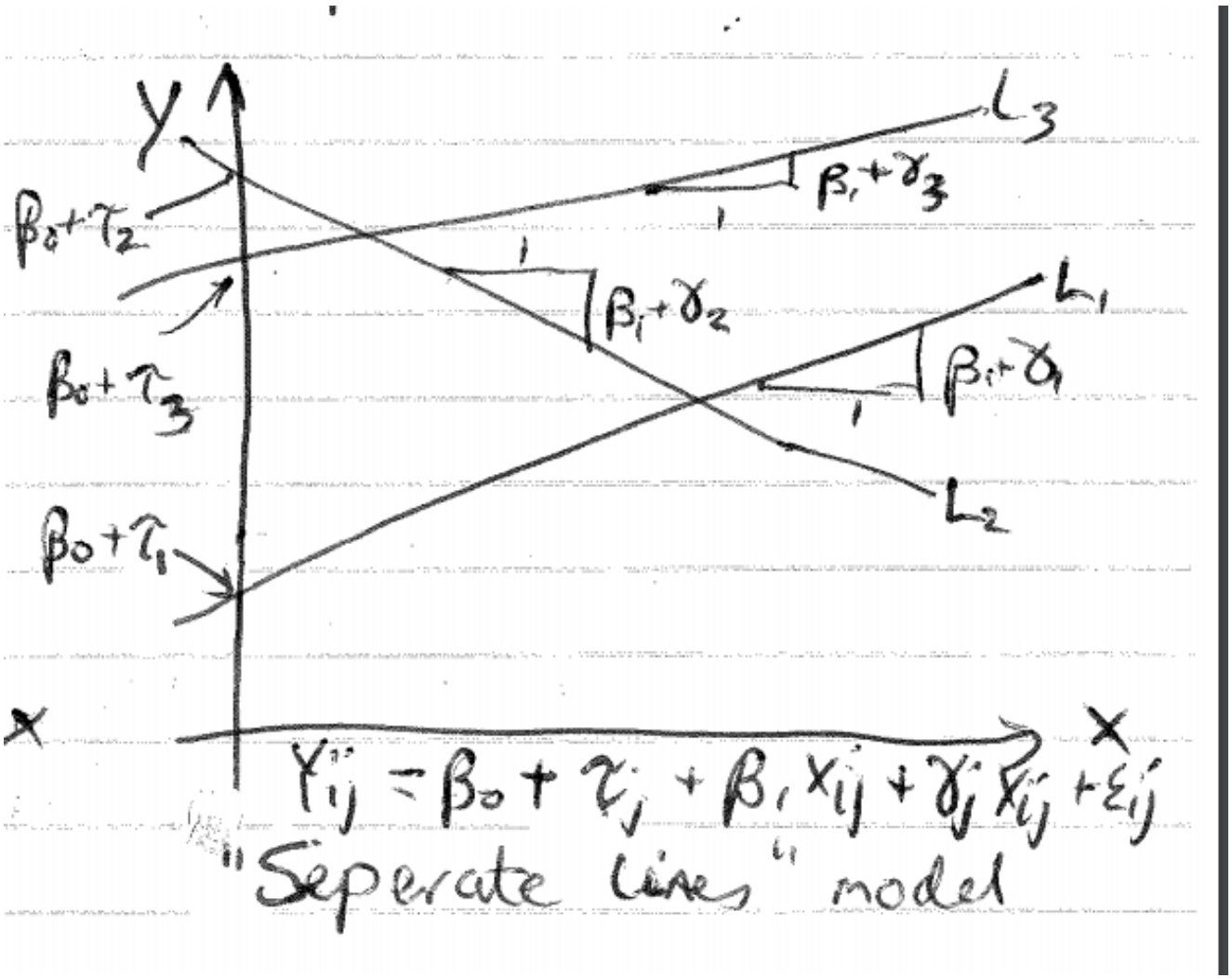
If we use indicator variables for the categorical predictor (e.g. encoding 1 for "Yes" and 0 for "No"), we can have the following parameterisation:

$$\begin{cases} \hat{Y}_{ij} = (\hat{\beta}_0 + \alpha_Y) + (\hat{\beta}_1 + \alpha_Y) X_{ij}, & X_{ij} \in \{\text{Yes}\} \\ \hat{Y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 X_{ij}, & X_{ij} \in \{\text{No}\} \end{cases}$$

Using R to fit the above model:

```
lm(Y ~ Continuous + factor(categorical) + Interaction)
```

Diagram


Y
X
L_3
L_2
L_1
$\beta_0+\tau_2$
$\beta_0+\tau_3$
$\beta_0+\tau_1$
$\beta_1+\gamma_3$
$\beta_1+\gamma_2$
$\beta_1+\gamma_1$
1
1
1
$Y_{ij} = \beta_0 + \tau_j + \beta_1 x_{ij} + \gamma_j x_{ij} + \varepsilon_{ij}$
"Seperate Lines" model

Question 1

- Check the diagnostic plots for unusual observations.
- Expected difference in price and the corresponding CI between a three- and a four-bedroom house?
  $\left(h = c(0, 0, 0, 0, 0, -1, 1, 0), \quad SE = \sigma\sqrt{h^T(X^TX)^{-1}h}\right)$
- some reference materials on cross-validation:
  http://robjhyndman.com/hyndsight/crossvalidation/