Modeling data with more than two levels

cont’d

04.04.22

STA 310

Click here for PDF of slides

Announcements

- Final project - optional draft due **Fri, Apr 15**, final report due **Wed, Apr 27**

Click here for Google slides for exercises.

Learning goals

- Write form of model for models with more than two levels
- Interpret fixed and random effects at each level
- See how three-level models are used in data analysis example
- Use the model to understand the covariance structure among observations

Data: Housing prices in Southampton

The data includes the price and characteristics for 918 houses sold between 1986 and 1991 in Southampton, England. The data were originally collected from a local real estate agency and were analyzed in the 1991 article "Specifying and Estimating Multi-Level Models for Geographical Research" by Kelvyn Jones. The primary variables of interest are

- **price**: Sales price in thousands of £
- **Age**: Age of the house
- **Bedrooms**: Number of bedrooms
- **House Type**: (semi-detached, detached, bungalow, terrace, flat)
- **Central heating**: Whether house has central heating (0: yes, 1: no)
- **Garage**: Number of garages (none, single, double)
- **Districts**: one of 34 districts (baseline:)
- **Half-years**: Half-year periods beginning the second half of 1986

STA 310

Data structure

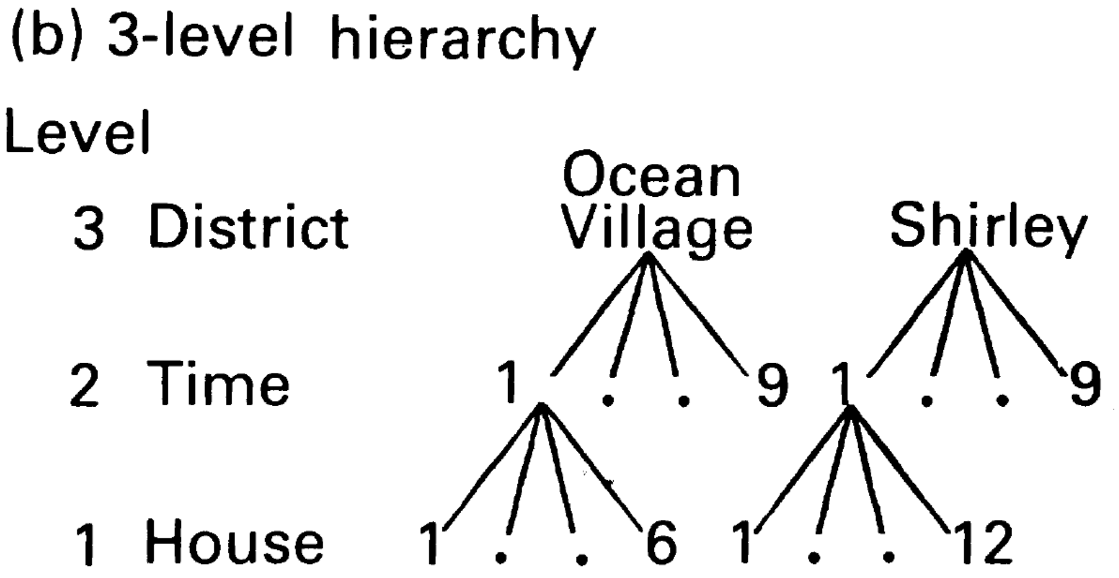

FIGURE 2. Hierarchical data structures

Adapted from Figure 2b from Jones (1991)

Note: The paper uses different symbols to represent parameters than what is in the textbook. The slides will follow the textbook.

Recap

STA 310

Unconditional means model

$$Y_{ijk} = \alpha_0 + \tilde{u}_i + u_{ij} + \epsilon_{ijk}$$

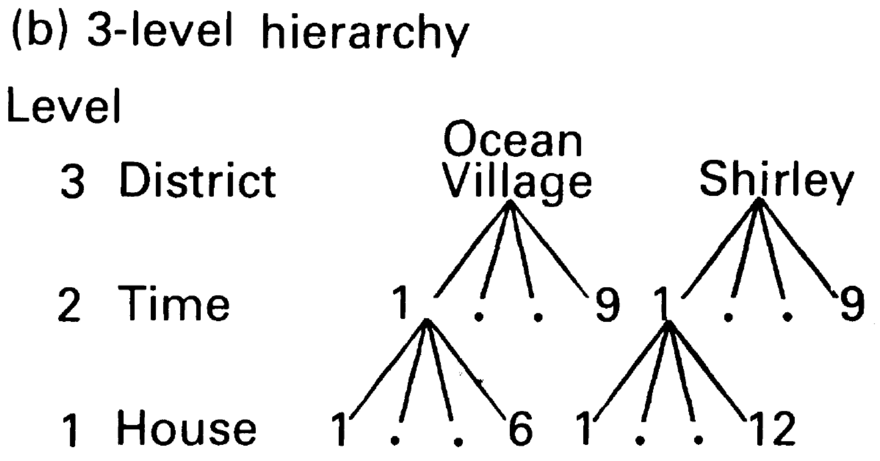

FIGURE 2. Hierarchical data structures

Level Three

$$a_i = \alpha_0 + \tilde{u}_i, \qquad \tilde{u}_i \sim N(0, \sigma^2_{\tilde{u}})$$

Level Two

$$a_{ij} = a_i + u_{ij}, \qquad u_{ij} \sim N(0, \sigma^2_u)$$

Level One

$$Y_{ijk} = a_{ij} + \epsilon_{ijk}, \qquad \epsilon_{ijk} \sim N(0, \sigma^2)$$

STA 310

Model A: Unconditional means model

TABLE I. *ML estimates for house price variation*

	Model A	B	C
Fixed effects			
Level 1: house			
Intercept (β_0)	58·1	57·0	56·7
Age (β_1)		0·0	0·0
		(0·1)	(0·1)
House type			
detached (β_2)		22·3	21·1
		(11·2)	(11·0)
bungalow(β_3)		17·9	15·7
		(5·6)	(5·2)
terrace (β_4)		2·7	3·0
		(1·8)	(2·0)
flat (β_5)		1·1	−0·9
		(0·5)	(0·4)
Bedrooms (β_6)		9·5	8·0
		(9·5)	(4·7)
Central heating (β_7)		−3·1	−3·1
		(2·6)	(2·6)
Garage			
single (β_8)		6·7	6·6
		(4·6)	(4·8)
double (β_9)		26·1	24·1
		(6·1)	(5·8)
Random effects			
Level 1: house			
Intercept (σ^2_ε)	397	236	215
Level 2: half-years			
Intercept (σ^2_μ)	169	139	127
	(10·3)	(11·6)	(11·7)
Level 3: districts			
Intercept (σ^2_φ)	240	201	213
	(6·7)	(6·9)	(7·0)
Bedrooms (σ^2_Γ)			58
			(5·7)
Covariance ($\sigma_{\varphi\Gamma}$)			111
			(3·4)

The overall mean price of houses in this data set is £ 58,100

Table 1 from Jones (1991)

STA 310

Model A: Random effects

Random effects			
Level 1: house			
Intercept (σ_ε^2)	397	236	215
Level 2: half-years			
Intercept (σ_μ^2)	169 (10·3)	139 (11·6)	127 (11·7)
Level 3: districts			
Intercept (σ_φ^2)	240 (6·7)	201 (6·9)	213 (7·0)
Bedrooms (σ_Γ^2)			58 (5·7)
Covariance ($\sigma_{\varphi\Gamma}$)			111 (3·4)

$$\hat{\rho}_{time} = \frac{169}{397 + 169 + 240} = 0.210$$

$$\hat{\rho}_{district} = \frac{240}{397 + 169 + 240} = 0.298$$

STA 310

Model with covariates

Model B: Add covariates + random intercepts

TABLE I. *ML estimates for house price variation*

	Model A	Model B	Model C
Fixed effects			
Level 1: house			
Intercept (β_0)	58·1	57·0	56·7
Age (β_1)		0·0	0·0
		(0·1)	(0·1)
House type			
detached (β_2)		22·3	21·1
		(11·2)	(11·0)
bungalow(β_3)		17·9	15·7
		(5·6)	(5·2)
terrace (β_4)		2·7	3·0
		(1·8)	(2·0)
flat (β_5)		1·1	−0·9
		(0·5)	(0·4)
Bedrooms (β_6)		9·5	8·0
		(9·5)	(4·7)
Central heating (β_7)		−3·1	−3·1
		(2·6)	(2·6)
Garage			
single (β_8)		6·7	6·6
		(4·6)	(4·8)
double (β_9)		26·1	24·1
		(6·1)	(5·8)
Random effects			
Level 1: house			
Intercept (σ^2_ε)	397	236	215
Level 2: half-years			
Intercept (σ^2_μ)	169	139	127
	(10·3)	(11·6)	(11·7)
Level 3: districts			
Intercept (σ^2_φ)	240	201	213
	(6·7)	(6·9)	(7·0)
Bedrooms (σ^2_Γ)			58
			(5·7)
Covariance ($\sigma_{\varphi\Gamma}$)			111
			(3·4)

Table 1 from Jones (1991)

1. Write the composite model.
2. Click here to interpret the covariate assigned to your group.

Model C: Additional random effect

TABLE I. *ML estimates for house price variation*

	Model A	B	C
Fixed effects			
Level 1: house			
Intercept (β_0)	58·1	57·0	56·7
Age (β_1)		0·0	0·0
		(0·1)	(0·1)
House type			
detached (β_2)		22·3	21·1
		(11·2)	(11·0)
bungalow(β_3)		17·9	15·7
		(5·6)	(5·2)
terrace (β_4)		2·7	3·0
		(1·8)	(2·0)
flat (β_5)		1·1	−0·9
		(0·5)	(0·4)
Bedrooms (β_6)		9·5	8·0
		(9·5)	(4·7)
Central heating (β_7)		−3·1	−3·1
		(2·6)	(2·6)
Garage			
single (β_8)		6·7	6·6
		(4·6)	(4·8)
double (β_9)		26·1	24·1
		(6·1)	(5·8)
Random effects			
Level 1: house			
Intercept (σ^2_ε)	397	236	215
Level 2: half-years			
Intercept (σ^2_μ)	169	139	127
	(10·3)	(11·6)	(11·7)
Level 3: districts			
Intercept (σ^2_φ)	240	201	213
	(6·7)	(6·9)	(7·0)
Bedrooms (σ^2_Γ)			58
			(5·7)
Covariance ($\sigma_{\varphi\Gamma}$)			111
			(3·4)

Table 1 from Jones (1991)

1. How does this model differ from Model B?
2. Write the composite model.
3. Write the Level One, Level Two, and Level Three models.

STA 310

Visualizing price by district over time

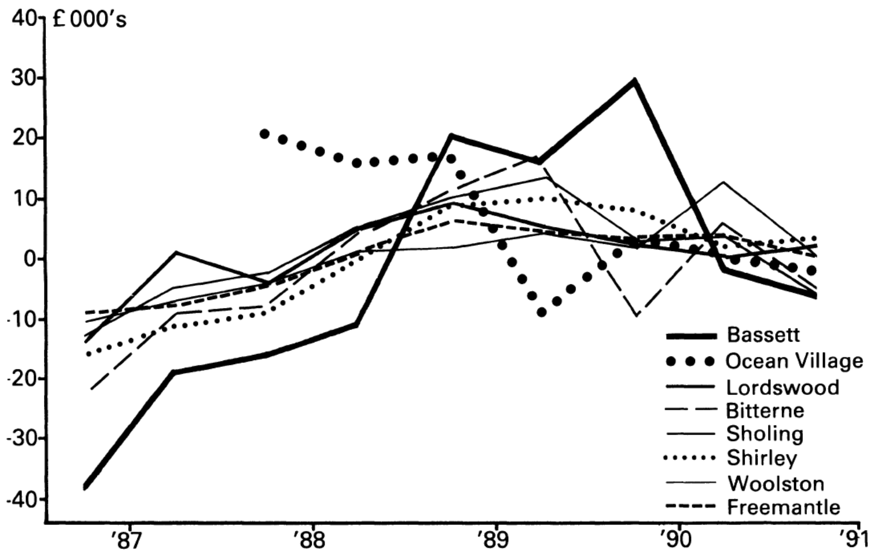

Figure 3 from Jones (1991)

1. What do you observe from the plot?
2. What terms in the model can be understood from the plot?
3. How might you use this type of plot to support decisions you make in the analysis?

Click here to submit your group's response.

Price by district and bedrooms

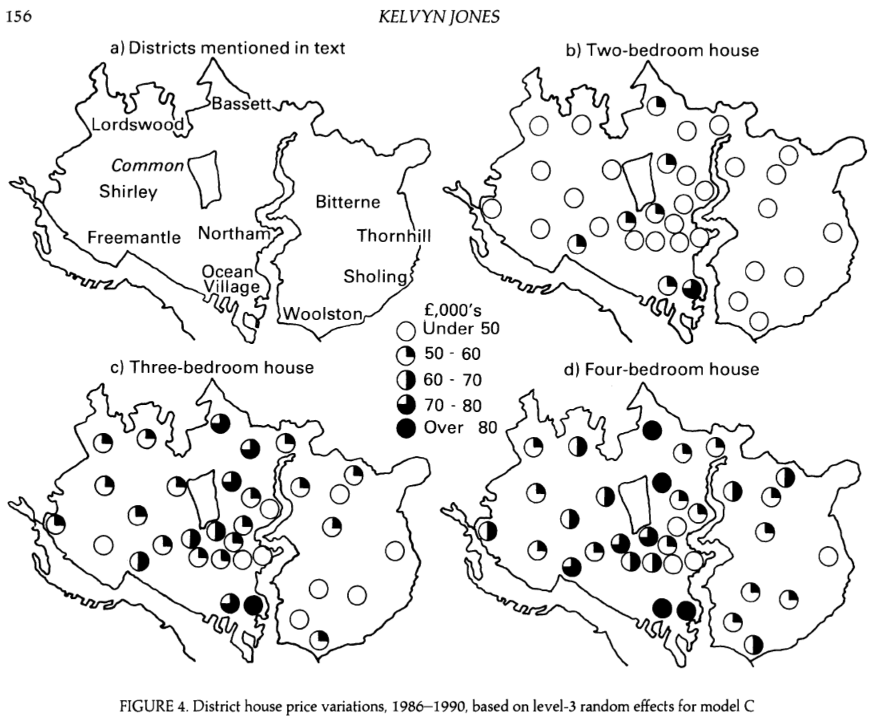

Figure 4 from Jones (1991)

1. What do you observe from the plot?
2. What terms in the model can be understood from the plot?
3. How might you use this type of plot to support decisions you make in the analysis?

Click here to submit your group's response.

Based on the output and visualizations, how does our understanding of the effect of bedrooms differ in Model C compared to Model B?

Covariance structure of observations

What we've done

So far we have discussed...

- the covariance structure between error terms at a given level, e.g. the relationship between u_i and v_i from a Level Two model:

$$\begin{bmatrix} u_i \\ v_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{bmatrix}\right)$$

- how to use the intraclass correlation coefficient to get an idea of the average correlation among observations nested in the same Level Two and/or Level Three observations

Questions we want to answer

Now we want to be able to answer more specific questions about the covariance structure between *observations* at different levels.

- What is the covariance structure of houses in the same district sold in different time periods? $(Y_{ij}, Y_{ij'})$
- What is the covariance structure of houses in the same district sold in the same time period? $(Y_{ijk}, Y_{ijk'})$
  - How does this structure differ between Model B and Model C in Jones (1991)?

Calculating variance and covariance

Suppose $Y_1 = a_1X_1 + a_2X_2 + a_3$ and $Y_2 = b_1X_1 + b_2X_2 + b_3$ where X_1 and X_2 are random variables and a_i and b_i are constants for $i = 1, 2, 3$, then we know from probability theory that:

$$Var(Y_1) = a_1^2 Var(X_1) + a_2^2 Var(X_2) + 2a_1a_2Cov(X_1, X_2)$$

$$Cov(Y_1, Y_2) = a_1b_1Var(X_1) + a_2b_2Var(X_2) + (a_1b_2 + a_2b_1)Cov(X_1, X_2)$$

Note: This extends beyond two random variables

We will use these properties to define the covariance structure of the observations in the model.

from BMLR: Section 9.7.5

Covariance structure under Model B

Let Y_{ijk} be the sales price for the house k in district i sold in time period j, and $x_1, \ldots, x_9$ be the house-level covariates.

$$Y_{ijk} = \alpha_0 + \sum_{i=1}^{9} \alpha_i x_i + [\tilde{u}_i + u_{ij} + \epsilon_{ijk}]$$

$$\tilde{u}_i \sim N(0, \sigma^2_{\tilde{u}}), \quad u_{ij} \sim N(0, \sigma^2_u), \quad \epsilon_{ijk} \sim N(0, \sigma^2)$$

Only use the random effects portion of the model in the variance and covariance derivations.

Variance and covariance derivations

Use Model B to write the derivation of $Var(Y_{ijk})$, the variance of an individual observation.

STA 310

Variance and covariance derivations

Use Model B to write the derivation of $Cov(Y_{ijk}, Y_{ijk'})$, the covariance between houses sold in the same time period that are in the same district.

Variance and covariance derivations

Write the derivation of $Cov(\mathbf{Y}_{ij})$, the covariance matrix for houses sold in the same time period that are in the same district.

STA 310

Looking ahead

- Calculate the correlation matrix
- Data analysis with three-level models
- Multilevel GLMs

Acknowledgements

- Jones, K. (1991). Specifying and estimating multi-level models for geographical research. Transactions of the institute of British geographers, 148-159.

- Beyond Multiple linear regression
  - Section 9.7: Covariance structure among observations
  - Chapter 10: Multilevel data with more than two levels

STA 310