Modeling data with more than two levels

03.30.22

STA 310

Click here for PDF of slides

Announcements

- Mini-project 02
  - Team evaluation & feedback due on Fri, April 1 at 11:59pm.
- HW 04 due **Fri, Apr 1** at 11:59pm
- Final project - optional draft due **Fri, Apr 15**, final report due **Wed, Apr 27**

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

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

Data structure

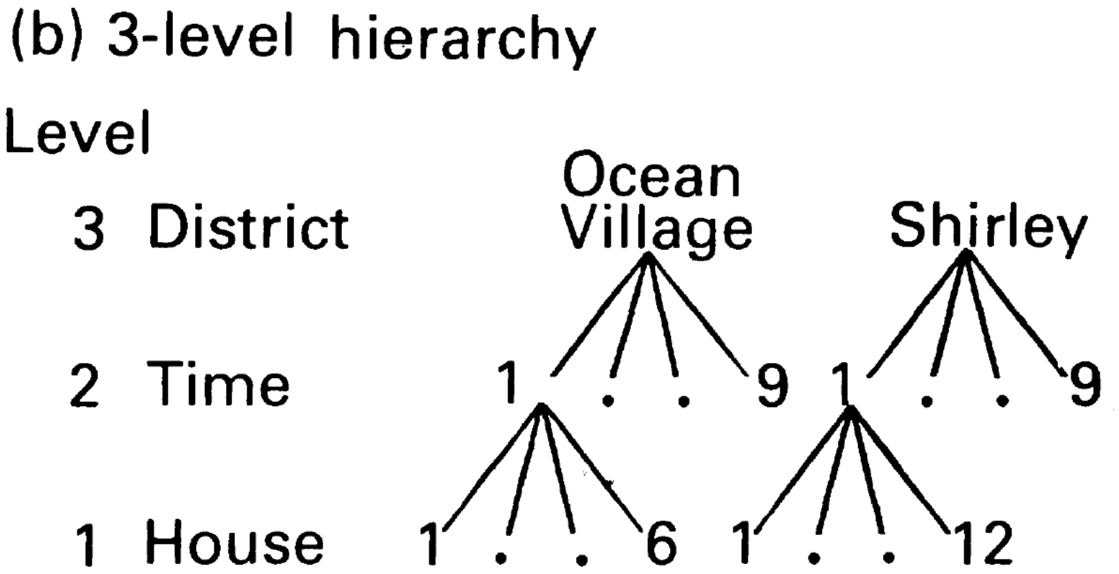

FIGURE 2. Hierarchical data structures

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.

Unconditional means model

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

Level One (house)

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

Level Two (time)

$$a_{ij} = a_i + u_{ij}, \quad u_{ij} \sim N(0, \sigma_u^2)$$

Level Three (district)

$$a_i = \alpha_0 + \tilde{u}_i, \quad \tilde{u}_i \sim N(0, \sigma_{\tilde{u}}^2)$$

STA 310

Label the terms of the composite model

- Y_{ijk}: Price of house k^{th} sold in the j^{th} time period in district i
- α_0:
- ϵ_{ijk}:
- u_{ij}
- $\tilde{u}_i$:

- σ: House-to-house variability within a given time period
- σ_u: Variability over time within a district
- $\sigma_{\tilde{u}}$: District-to-district variability

Model A

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)

	A	B	C
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)

Interpret $\hat{\beta}_0$ (this is α_0 in our model notation)

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	139	127
	(10·3)	(11·6)	(11·7)
Level 3: districts			
Intercept (σ_{φ}^2)	240	201	213
	(6·7)	(6·9)	(7·0)
Bedrooms (σ_{Γ}^2)			58
			(5·7)
Covariance ($\sigma_{\varphi\Gamma}$)			111
			(3·4)

1. Calculate the intraclass correlation for time.
2. Calculate the intraclass correlation coefficient for districts.
3. Is there evidence the multilevel model structure is useful for this data?

Model B: Add covariates + random intercepts

TABLE I. *ML estimates for house price variation*

	A	*Model 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. Write the composite model.
2. Which variables appear to have a statistically significant effect on price?
3. Use this model to interpret the effect of bedrooms on the price of houses in Southampton.

STA 310

Model C: Additional random effect

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. 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

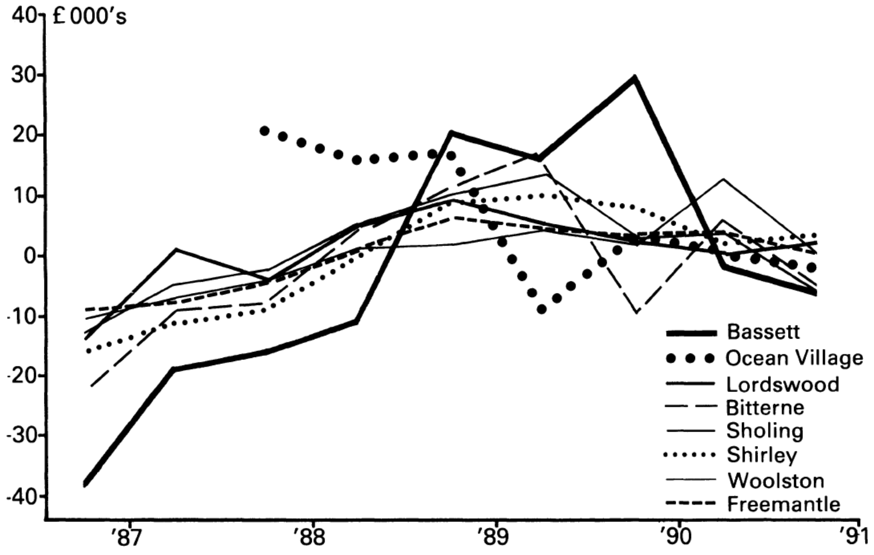

Figure 3 from Jones (1991)

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?

Price by district and bedrooms

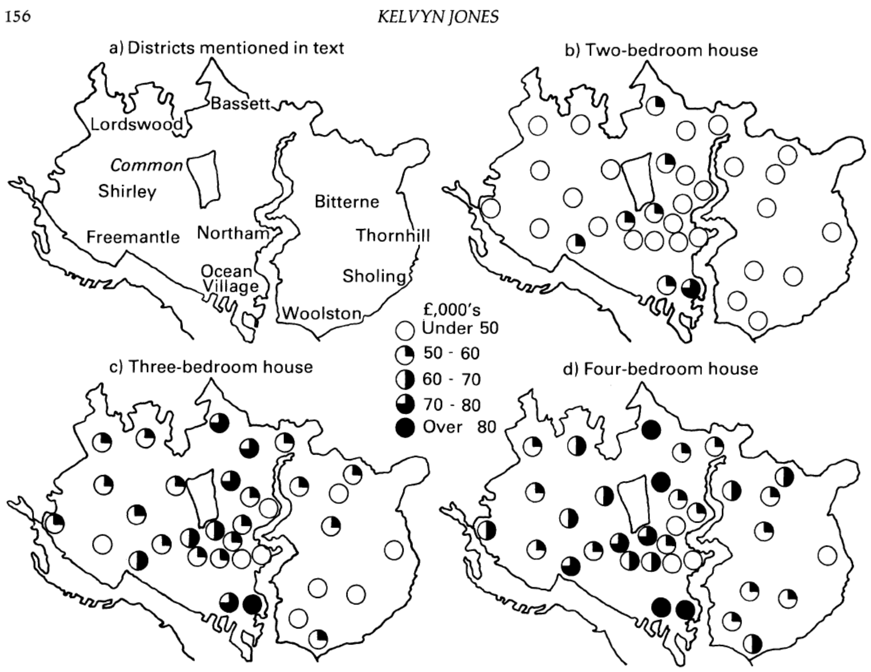

FIGURE 4. District house price variations, 1986–1990, based on level-3 random effects for model C

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?

How does our understanding of the effect of bedrooms differ in this model compared to Model B?

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
  - Chapter 10: Multilevel data with more than two levels

STA 310