Day 2: More Notation

Restaurants!

1. Pine grove hall
2. RE Farm
3. Faccia Luna
4. Lupitas
5. Juana's
6. Café Alina
7. Meyer Dairy

Unsolicited advice

- Of all the things you do in grad school, nothing will add more value to your degree than research methods.

Review

Downs voting model

$$\mathbf{U = PB - C}$$

If we want our model to account for the utility people gain from doing their civic duty, how would we need to change the model?

Downs voting model

$$\mathbf{U = PB - C}$$

If we want our model to account for the utility people gain from doing their civic duty, how would we need to change the model?

$$\mathbf{U = PB - C + D}$$

Downs voting model

$$\mathbf{U = PB - C}$$

If we believe cost is evaluated differently by people of different socioeconomic status, how would we change our model?

Downs voting model

$$\mathbf{U = PB - SC}$$

If we believe cost is evaluated differently by people of different socioeconomic status, how would we change our model?

- **SC** is a generic expression stating that "socioeconomic status moderates the effect of cost"
- It's mathematical expression may be slightly different depending on how we measure the variables and what the effect is

SC

Simple example:

- Suppose **S** is measured on a 1-10 scale with 10 representing high socioeconomic status.
- Suppose **C** is measured as dollars.
- We believe people of high socioeconomic status will give less consideration to **C**.
- How do we express this?

It’s all addition…

- As a rule of thumb, addition and multiplication are used to express relationships in an equation if you have no priors.
- Multiplication, subtraction, and division can all be reduced to addition, so it doesn’t really matter.
- However, subtraction and division can be used if you have priors about how different terms interact or affect the outcome.

Notation: Models

Common Defined Variables

Variable	Notes
μ	Mu, the population mean
$\overline{X}$	X-bar, the sample mean
β	Beta, the estimated parameters of a linear model. Intercept and coefficients
$\hat{x}, \hat{y}, \hat{u}, \hat{s}$	Hat, the estimated value of a given variable
u, e	The error term in a model. The distance between a predicted value and an observed value.
σ	Sigma, Population standard deviation. σ^2 is the population variance.
s	Sample standard deviation. s^2 is the sample variance
N	Population
n	Sample population

Theorizing Downs

We want to test the Downsian theory of turnout by predicting the percent of elections individuals turn out for.

We hypothesize a model with the following variables:

y: the percent of elections they turned out for in their life

x: The utility gained by if their preferred party wins, weighted by the probability that their vote decides an election ($x = pb$).

c: The average cost of voting across elections

Expressing Downs

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 C_i + u_i$$

This model is expressing a theory. How do I know this?

Expressing Downs

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 C_i + u_i$$

This model is expressing a theory. How do I know this?

- Capital variables indicate random (unobserved) variables. Use lowercase variables if this is vector you have actual values for.
- Absence of hats on parameters

Expressing Downs

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 C_i + u_i$$

This model is expressing a theory. How do I know this?

- Capital variables indicate random (unobserved) variables. Use lowercase variables if this is vector you have actual values for.
- Absence of hats on parameters

Why is cost now being added rather than subtracted?

Expressing Downs

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 C_i + u_i$$

This model is expressing a theory. How do I know this?

- Capital variables indicate random (unobserved) variables. Use lowercase variables if this is vector you have actual values for.
- Absence of hats on parameters

Why is cost now being added rather than subtracted?

- Subtraction is just the addition of negative numbers. In our model we are trying to estimate the effect of cost. We don't assume its negative or positive before estimating the value so we just use addition by convention. If higher cost has a negative effect, β_2 will be a negative number and it will be the same as subtracting the cost.

Expressing Downs

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 C_i + u_i$$

What is u_i ?

Expressing Downs

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 C_i + u_i$$

What is u_i ?

- u_i is random variation that our model does not account for

Expressing Downs

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 C_i + u_i$$

What is u_i ?

- u_i is random variation that our model does not account for

What is the role of β_0 in the equation? In other words, what is its relationship with Y?

Expressing Downs

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 C_i + u_i$$

What is u_i ?

- u_i is random variation that our model does not account for

What is the role of β_0 in the equation? In other words, what is its relationship with Y?

- It is the value Y takes when all other values are zero. Think of it as the starting point.

Expressing Downs

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 C_i + u_i$$

What is u_i ?

- u_i is random variation that our model does not account for

What is the role of β_0 in the equation? In other words, what is its relationship with Y?

- It is the value Y takes when all other values are zero. Think of it as the starting point.

What is the role of β_1 and β_2?

Expressing Downs

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 C_i + u_i$$

What is u_i ?

- u_i is random variation that our model does not account for

What is the role of β_0 in the equation? In other words, what is its relationship with Y?

- It is the value Y takes when all other values are zero. Think of it as the starting point.

What is the role of β_1 and β_2?

- They determine how much of an effect X and C have on Y.

What's the substantive difference between these equations?

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 C_i + u_i$$

$$y_i = \widehat{\beta_0} + \widehat{\beta_1} x_i + \widehat{\beta_2} c_i + \hat{u}_i$$

What does this third equation tell us?

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 C_i + u_i$$

$$y_i = \widehat{\beta_0} + \widehat{\beta_1} x_i + \widehat{\beta_2} c_i + \hat{u}_i$$

$$\hat{y}_i = \widehat{\beta_0} + \widehat{\beta_1} x_i + \widehat{\beta_2} c_i$$

What does this third equation tell us?

$$y_i = \widehat{\beta_0} + \widehat{\beta_1} x_i + \widehat{\beta_2} c_i + u_i$$

$$\hat{y}_i = \widehat{\beta_0} + \widehat{\beta_1} x_i + \widehat{\beta_2} c_i$$

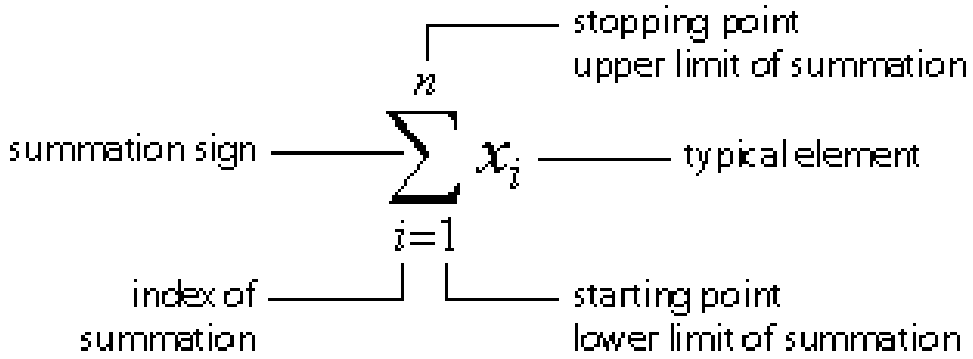
$$y_i = \hat{y}_i + \hat{u}_i$$

Or

$$\hat{u}_i = y_i - \hat{y}_i$$

Some operators you'll need…

Operator	
$\sum_{i=1}^{n} x$	Summation. The sum of numbers in a vector,
$\prod_{i=1}^{n} x$	Product operator. The product of operators in a vector



Population Mean	Sample Mean
$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$	$\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$
N = number of items in the population	n = number of items in the sample