Data Analysis for the Social Sciences

Quantitative Data Analysis I

2022-09-19

Welcome everyone to the first of two lectures on quantitative data analysis.
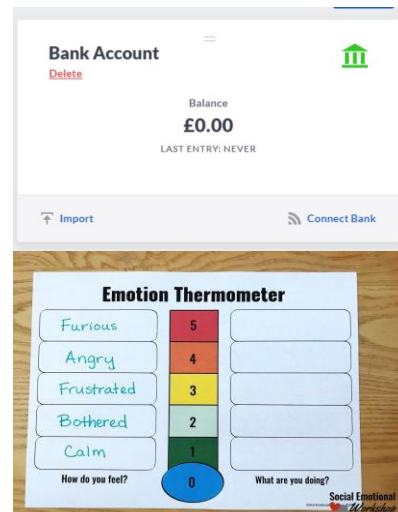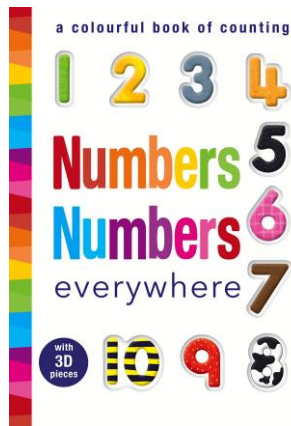
Our focus today is on two important topics:
1. How we design a piece of analysis
2. How quantitative data are structured

Data

Let's start with a simple definition (Agresti, 2018).

This definition applies equally to qualitative or quantitative data.

Numbers, numbers everywhere

Numbers are everywhere (whether you like them or not). On the positive side, you are much more adept at understanding and using numbers than you may give yourself credit for (think of your ability to manage finances, bills, academic and personal lives) – this translates into research also. On the negative side, people and institutions have been known to manipulate numbers to their advantage (and thus your disadvantage).

Some numbers are real i.e., the thing you are measuring is naturally numeric (bank balance, height, distance from Paisley campus to your accommodation).

Some numbers are labels i.e., we use numbers to represent characteristics, qualities or emotions.

For quantitative research, a collection of numbers is called data. But these data need to be structured properly in order to analyse them.

What are data?

A collection of observations (Agresti, 2018):

- **An observation** is a set of measurements for a case.

- **A measurement** is a description of some characteristic of a case.

- **A case** is the entity we are observing e.g., individuals, countries, animals, companies, networks etc.

Let's start with a simple definition (Agresti, 2018).

This definition applies equally to qualitative or quantitative data.

What are quantitative data?

A collection of observations in a particular format **[Variable-by-case matrix]**

• **Case** = the entity we are observing e.g., individuals, countries, animals, companies, networks etc.

• **Variable** = what we are measuring about a case e.g., a characteristic

• **Value** = the result of measuring a variable for a given case (MacInnes, 2017)

Quantitative data are data arranged in a particular format.

What are quantitative data?

unique_id	name	age	sex	employed
1001	John	21	1	1
1002	Jane	20	2	0
1003	Jerry	45	1	0
1004	Jessica	37	2	0
1005	Joanne	22	2	1
1006	Jordie	19	1	1
1007	Jacob	20	1	0
1008	Jason	25	1	1
1009	Jenny	28	2	1
1010	Jerome	59	1	0

Here is what a variable-by-case matrix looks like.

We'll see real examples in the lab later today.

Designing

Let's start with a simple definition (Agresti, 2018).

This definition applies equally to qualitative or quantitative data.

What is your estimand?

"*The estimand is the object of inquiry—it is the precise quantity about which we marshal data to draw an inference.*" (Lundberg et al., 2021: 532)

1. Who are you studying? **[Unit of analysis]**

2. What are you trying to quantify? **[Conceptualisation]**

3. How will you quantify it? **[Operationalisation]**

4. What does your attempt to quantify it look like? **[Level of measurement]**

Conceptualisation

*"We must convert our ideas about social phenomena into data **by deciding what to measure** and how to measure it."* (Agresti, 2018: 23)

Unit of analysis = Individual

Object / Topic	Conceptualisation
Fuel poverty	Spending more than 10% of net household income on energy costs
Social mobility	Being in a higher occupational class than your father / mother

Conceptualisation

Unit of analysis = Individual

Object / Topic	Conceptualisation
Food insecurity	
General wellbeing	

[Launch padlet:]

Operationalisation

"*We must convert our ideas about social phenomena into data by deciding what to measure and **how to measure it**.*" (Agresti, 2018: 23)

Unit of analysis = Individual

Object / Topic	Conceptualisation	Operationalisation
Fuel poverty	Spending more than 10% of net household income on energy costs	?

The trick here is think about what values go in the column for fuel poverty.

Operationalisation

unique_id	name	age	sex	employed	%_spent_fuel	in_fuel_poverty
1001	John	21	1	1	5	
1002	Jane	20	2	0	25	
1003	Jerry	45	1	0	31	
1004	Jessica	37	2	0	11	
1005	Joanne	22	2	1	15	
1006	Jordie	19	1	1	58	
1007	Jacob	20	1	0	60	
1008	Jason	25	1	1	9	
1009	Jenny	28	2	1	10	
1010	Jerome	59	1	0	36	

Operationalisation

"*We must convert our ideas about social phenomena into data by deciding what to measure and **how to measure it**.*" (Agresti, 2018: 23)

Unit of analysis = Individual

Object / Topic	Conceptualisation	Operationalisation
Fuel poverty	Spending more than 10% of net household income on energy costs	Categorical variable with two categories: • In fuel poverty (1) • Not in fuel poverty (0)

Remember how you conceptualised fuel poverty in order to decide what values the 'in_fuel_poverty' column can take.

Operationalisation

unique_id	name	age	sex	employed	%_spent_fuel	in_fuel_poverty
1001	John	21	1	1	5	0
1002	Jane	20	2	0	25	1
1003	Jerry	45	1	0	31	1
1004	Jessica	37	2	0	11	1
1005	Joanne	22	2	1	15	1
1006	Jordie	19	1	1	58	1
1007	Jacob	20	1	0	60	1
1008	Jason	25	1	1	9	0
1009	Jenny	28	2	1	10	0
1010	Jerome	59	1	0	36	1

Measuring

Let's start with a simple definition (Agresti, 2018).

This definition applies equally to qualitative or quantitative data.

Measurement

Measurement Scale	Level of Measurement	Description	Example
Categorical	Nominal	Presence of some attribute	Sex at birth, Ethnicity
Categorical	Ordinal	More or less of some attribute	Social class, Degree classification
Numeric	Interval	More or less of some attribute (and by how much)	Income, Number of deaths

Identifying and understanding which level(s) of measurement are present in your data is crucial, as it determines which analytical techniques you can use (e.g., does it make sense to calculate average marital status?).

Measurement

unique_id	name	age	sex	employed	%_spent_fuel	in_fuel_poverty
1001	John	21	1	1	5	0
1002	Jane	20	2	0	25	1
1003	Jerry	45	1	0	31	1
1004	Jessica	37	2	0	11	1
1005	Joanne	22	2	1	15	1
1006	Jordie	19	1	1	58	1
1007	Jacob	20	1	0	60	1
1008	Jason	25	1	1	9	0
1009	Jenny	28	2	1	10	0
1010	Jerome	59	1	0	36	1

Can you identify the level of measurement for each column in the table?

Measurement

Properties of good measures:

1. Reliable

2. Valid

Regardless of the measurement scale / level of measurement, we want our measures to possess two properties:

Reliable: produces the same results under equivalent conditions.

Valid: actually measures the concept we are interested in.