Quantitative Data Basic Introduction

Dr Kevin Morrell
www.kevinmorrell.org.uk

Session Overview

- Quantitative data
- Hypotheses
- Surveys
- Sampling
- Some basic statistics
- Examples

Language and learning objectives

1. By the end of the session you will be able to define the following: nominal, ordinal, interval and ratio scales; hypothesis, significance, test; validity, reliability, generalizability.
2. You will be able to identify some of the advantages and disadvantages associated with quantitative data, and the use of surveys.
3. You will have a basic understanding of two related issues: sampling and basic statistics.
4. You will have a nodding acquaintance with some common statistical tests.

Quantitative Data

"…it is *not* possible to become a social scientist without an understanding of statistics."
(Diamond and Jefferies, 2001: 1)

Working with quantitative data

- Primary data (questionnaires, measurement)
- Secondary data (existing research, library searches of statistical data)
- In almost any research project you will need to study, and write about, quantitative research

Advantages of Quantitative Data - 1

- Specific
- Replicable
- Generalizable
- Breadth… aggregate across multiple subjects, summarizes findings

But

- 'data' are constructed what you look for and see are subjective as are interpretations
- Danger of being seduced by numbers

Advantages of Quantitative Data - 2

"…in non-laboratory situations where experiments are often neither feasible nor ethically defensible, surveys give that reassuring scientific ring of confidence. Associated with surveys is a satisfyingly complex set of technological concerns about sampling, question-wording, answer-coding etc."

(Robson RWR: 230).

Surveys

Surveys

There is a sense in which surveys are more like a research strategy… than a specific tactic or method. However, many of the concerns involved in doing a survey are not so much with questions of overall strategic design as with highly practical and tactical matters to do with the detailed design of the instrument to be used… determining the sample to be surveyed and ensuring high response rates.

(Robson, 2002: 228).

Typical features of surveys

- Use of a fixed, stable quantitative design
- Collection of specific data in quantifiable form
- Comparatively larger sample than qualitative studies
- Attempt to seek representative sample as the basis for wider generalisation

Scales

Scales And Scale Types

Responses are measured on the following

Ratio
Interval
Ordinal
Nominal

Why are they important?

They determine the type of analysis that can be conducted

Scales And Scale Types: Non-metric

NOMINAL

- Values assigned to categories
- The categories can not be ranked
- The numbers assigned have no intrinsic meaning

ORDINAL

- Assign numbers to categories that correspond to ranks
- Responses can be ranked from high to low (vice versa)
- The distance between the first and second category is not necessarily the same as the distance between the second and third categories

Scales And Scale Types: Metric

INTERVAL

- Have the properties of ordinal scales
  PLUS
- The distance between the first and second category IS the same as the second to third etc

RATIO

- Have the properties of interval scales
- Also have items classified into a 'zero' category

Dichotomous Scales

A yes/no response scale is a **dichotomous** scale

They are typically treated as nominal but they can be ordinal too

It's harder to predict these because there is no room for error (e.g. business failure)

Scales - Examples

Nationality, race, religion, gender, region, type of employer, are nominal measures

Write down 3 more nominal measures

Questions where respondents are asked to evaluate a service, product or experience are usually ordinal measures (though composite scales mean we can sometimes treat these as interval)

Write down 3 more nominal measures

Physical attributes like time, height, weight, distance (and things like cost and profit) are ratio measures, but there are fewwer of these in 'real life'

Which scale is it again?

- Example - In a census, properties could be classified according to how many occupants they have in them
- This appears to have the properties of a ratio scale – there is a true zero (when houses are empty) and there is an equal interval between data points.
- But there might be a *categorical* difference between an empty house and a house with one person in (for noise levels for instance). This little example shows how sometimes by 'measuring' we actually 'construct' what we measure.

Hypotheses

What's a hypothesis

- *Hypothesis*: a predicted, or expected answer to a research question
- e.g. Tenure [of 'x', at time 't', in place 'y'] is positively correlated with salary

Ideal, or idealised practice of hypothesis testing

- Study literature for 'gaps' / areas for contribution
- Choose a context & frame research questions
- Stipulate hypotheses
- Design (instrument, methods, sample frame etc)
- Collect data
- Analysis / test hypotheses
- Write up

Although…

- Null results don't get published as readily and aren't interesting (we don't see all the tests)
- It's much messier than that
- Often people gather first, and test afterwards (Ready, Fire, Aim)
- Funded research comes with constraints

What does 'significant' mean

- General everyday sense, worth drawing attention to, indicative
- Technical sense – threshold used to demonstrate the probability that a given result (usually a 'difference'-type result) cannot be explained simply by chance
- Difference between 'significant' and significant, i.e. is it theoretically meaningful – use sparingly and carefully misuse / thoughtless use is a quick way to show up ignorance

Sampling

Why is sampling important?

- Ideally, social scientists working with quantitative data use statistics to test hypotheses.
- Hypotheses are usually 'law-like' and refer to a wider population than the one sampled.
- So, when social scientists examine quantitative data, they do so to draw *inferences* about the wider population.
- As well as considering how well their study is designed (internal validity), they need to consider how well their data / sample allows them to make these inferences (external validity).
- Remember… it's critically important in case study and qualitative research too.

Key terms

- *Sample*: "the units chosen to be included in a study… to represent a population statistically… random selection methods should be used" > sampling is how you choose those units
- *Population*: "The universe of elements from which the sample elements are drawn" can be literal e.g. population of GB, or metaphorical e.g. NHS Trusts in the UK
- *Sample frame*: "the source of the eligible population from which the survey sample is drawn" e.g. NHS Trusts in the South-East (sources of bias), those with land-line telephones (may not be representative)

The main types of random sampling

- Simple random - Random sample from sample frame (lottery method, computer etc.)
- Stratified random - Random sample from defined groups or strata
- Cluster - Random sample of whole groups or units (geographical clusters – e.g. all the teachers in 10 secondary schools: social capital / networks)
- Multistage - Random sample of clusters in strata or levels, then random sample of individuals (geographical clusters, then random samples within these – e.g. 20% of teachers in each of 50 secondary schools: more self-contained, attitudinal data)

The main types of non-random sampling

- Purposive – choose on the basis of desired characteristics
- Quota - predefine categories and get enough in each
- Snowball - obtain names of further subjects from those already contacted
- Convenience - take what you can get

Sampling errors: sources

- Data collected on sample, not population (need to constantly qualify interpretation)
- Random sampling reduces but does not eliminate error
- Some assumed natural variation among samples
- Types of bias – common method variance, or bias, problem of apathy
- Problematic cases, or issues in terms of sampling:
  - Literacy / familiarity with paper-work (problematic if comparing community and office based staff)
  - Voter participation

Non-sampling errors, such as...

- Poor choice, definition or operationalisation of sampling frame
- Poor measuring instruments or data collection procedures
- Errors in data processing or analysis
- Non-response by subjects

Basic Statistics

Some common statistical tests

- Chi-Square Test – is the observed result different to what one would expect?
- Bivariate Correlation – to what extent are two variables related?
- Discriminant Analysis – can one discriminate between groups on the basis of a given variable?
- Linear Regression – how well does a set of variables 'predict' the value of another variable?
- Logistic Regression - how well does a set of variables 'predict' the value of another (dichotomous) variable?
- Cluster Analysis – are there identifiable groups, or clusters in the dataset?

Language and learning objectives - recap

1. By the end of the session you will be able to define the following: nominal, ordinal, interval and ratio scales; hypothesis, significance, test; validity, reliability, generalizability.
2. You will be able to identify some of the advantages and disadvantages associated with quantitative data, and the use of surveys.
3. You will have a basic understanding of two related issues: sampling and basic statistics.
4. You will have a nodding acquaintance with some common statistical tests.

Quantitative Data Basic Introduction

Dr Kevin Morrell
www.kevinmorrell.org.uk