Quantitative Research: Defining features

Dr. Syed Owais
Assistant Professor (Sociology)
University of Peshawar

1

Main preoccupation of Quant research

1. **Concern with Measurement**
2. **Concern with causality**
   - "Quantitative researchers are rarely concerned merely to ***describe how things are*** [emphasis added], but are keen to say why things are the way they are" (Bryman, 2012: 175).
     - E.g., not just describe how much racial discrimination exists or who (in a population holds it) but ***why*** does it exist or what ***causes*** it.
   - It's because the arguments/model is derived from *natural science* tradition.
3. **Concern with Generalisation**
   - findings can be generalized beyond the confines of the particular context in which the research was conducted.
   - Given that it is rarely feasible to send questionnaires to or interview whole populations, it is necessary to **sample,** which is **as representative as possible** to state that the obtained results from the survey/experiment are not unique to the people studied but to the larger pool (**population**) that make up the sample.

2

Main preoccupation of Quant research

4. **Replicability**

- natural sciences aim to reduce to a minimum the contaminating influence of the scientist's biases and values.
  - To check the influence of potentially damaging problems, scientists may seek to replicate—that is, to reproduce—each other's experiments.
- Consequently, it is often regarded as important that the researcher spells out clearly his or her procedures so that they can be replicated by others, even if the research does not end up being replicated.

Sampling: some central concepts

- ***Population*:** the universe of units from which the sample is to be selected.
  - Units = people, nations, cities, regions, firms, newspapers, etc.
- ***Sample*:** the segment of the population that is selected for investigation. It is a subset of the population. The method of selection may be based on a probability or a non-probability approach.
- ***Sampling frame*:** the listing of all units in the population from which the sample will be selected. E.g., school records, telephone director, directory from cell-phone operators etc.
- ***Representative sample*:** a sample that reflects the population accurately so that it is a microcosm of the population.

Developing a Sampling Plan

1. Define the Population of Interest
2. Identify a Sampling Frame (if possible)
3. Select a Sampling Method
4. Determine Sample Size
5. Execute the Sampling Plan

Probability Sampling

- *A* sample that has been selected using random selection so that each unit in the population has a known chance of being selected. The aim of probability sampling is to keep *sampling error* to a minimum.

Types of probability sampling

- Simple Random Sampling
- Systematic Sampling
- Stratified Random Sampling
- Multi-stage Cluster Sampling

1) Simple random sampling

- Each unit of the population has an equal probability of inclusion in the sample.
- E.g., *Assessing Quality of Education in HEIs: A Survey/~~Case Study~~ of UoP*
  - *Only enough Resources (money)* available to interview/survey students!
  - Information/statistical data from **Director Admission's office** = 13200 students.
  - Decision → survey of full-time BS students in UoP = 8000.
  - So if sample = 400; total population of (the above-defined students) = 8000, then
  - n/N (where 'n' means the *sample* of our interest and 'N' means the population of our interest)
  - → 400/8000 = 0.05
  - Multiple 0.05 with 100 = 5, i.e., 5 students in every 100 students or 1 student in every 20 students (in BS) has the chance of being selected in our sample
  - This is called **Sampling Fraction.**

1) Simple random sampling: steps

- Define **population** of interest (in our case: 8000)
- Select/devise **sampling frame** (students' record from director admissions, controller of examinations etc. or your own collection of students records from each department in the Uni)
- Decide **sample size** (in our case: 400)
- **List** all the units (students) in **population B** (in our case: 1 to 8000).
- Use **random numbers table (RNT)** (see below) or use a computer programme to generate random numbers for you.

Random no. table

73735	45963	78134	63873
02965	58303	90708	20025
98859	23851	27965	62394
33666	62570	64775	78428
81666	26440	20422	05720
15838	47174	76866	14330
89793	34378	08730	56522
78155	22466	81978	57323
16381	66207	11698	99314
75002	80827	53867	37797
99982	27601	62686	44711
84543	87442	50033	14021
77757	54043	46176	42391
80871	32792	87989	72248
30500	28220	12444	71840

Are made of columns of 5-digit numbers

Available in most statistics textbooks

Simple random sampling

- Using RNT, select sample of your interest (in our case: 400 students).
  - First, almost no opportunity for human bias to manifest itself (i.e., Students not selected subjective criteria, e.g., whether they looked friendly or not)
  - Second, the process is not dependent on the students' availability. Selection is done without respondents' knowledge
- In using RNT for our population of 8000 (which has four digits), we can decide whether to select the 4 digit numbers from the left-hand side or the right-hand side in the columns making 5-digit numbers in the table!
  - Let's say, we decide to chose the four digits on the right-hand side in each column (begging randomly!) . . .

Simple random sampling

1. 3735 (i.e., a student whom we have serial numbered as 3735 in our list of population will be surveyed/interviewed using (1) face-to-face, (2) postal, (3) telephonic or (4) web-based survey tool/questionnaire!
2. 2965
3. 3666
4. 1666
5. And … so on until we have selected 400 students

2) Systematic sampling

- If we feel the procedure is tortuous or we don't have adequate time etc. for SRS, then we use systematic sampling, which is:
- We select units from sampling frame without using random no. table/computer programmes.
- Since from our calculation of **sampling frame (n/N),** we know that we should select 1 in every 20 students from the total pool of 8000 BS students.
- Using RNT, we decide to choose the first two digits to the left-hand side of a column in it, looking for a figure between 1and 20
- In the above RNT, we find **02965** as the suitable and accurate serial no. for selection.
  - From this point onwards, we select every 20th student in our sampling frame thus: 02+20 = 22) → 22+20 = 42 → 42+20 = 62 and so on.

Stratified Random Sampling

- It's not necessary that the BS students' population of 8000 will be equally divided into 5 faculties (physical sciences, natural sciences, social sciences, humanities, management sciences) in UoP. E.g., it might be
  - … AND
- Let's say we would like adequate proportion of each faculty in our BS students' sample of 400.

Stratified random sample

faculty	BS students in faculty	Stratified sample required
Arts & Humanities	800	(if 1 in 20, then) 40 (in 800)
Social Sciences	1800	… 90
Physical sciences	2200	… 110
Natural science	3000	… 150
Management sciences	200	… 10
Total	8000	… 400 sample

Stratified random sampling

- We can use multiple criteria for **stratifying the population of our interest,** *e.g.*, gender, age, region, CGPA records, etc.
  - Provided data/records on such aspects of the population is available.
  - If no, then headache!

Multi-stage cluster sampling

- LET'S SAY…
- We want to conduct a national survey of BS students in all *public sector universities.*
  - We know simple random sample might prove too costly and time-consuming even with a sample of 400
  - Imagine, RNT gives you serial no. of students who are in *Baluchistan* and *Sindh* only!
- To do multi-stage cluster sampling, say for the selection of 400 students from all public sector unis…
- We first decide to sample 10 Unis from the **population of Unis** using RNT, giving us 10 clusters!
- Then, we sample 400 students in each sampled Uni using RNT (again!).

17

Multi-stage cluster sampling

- It is possible that we end up with the following 10 Unis
  1. Peshawar Uni
  2. Sindh Uni
  3. Turbat Uni
  4. Shiringal Uni
  5. Bannu Uni
  6. Baluchistan Uni
  7. Karachi Uni
  8. Kashmir Uni
  9. Swat Uni
  10. Chitral Uni

18

Multi-stage cluster sampling

- To save costs and time, we may, e.g., further divide Pakistan into administrative units/provinces
  - → then **randomly** sample 2 provinces
  - → sample 5 Unis from each province using **RNT**
  - → Sample 800 students in each sampled Uni in both the provinces, using **RNT**

Factors to Consider in Sample Design

- Research objectives
- Resources
- Knowledge of target population
- Degree of accuracy
- Time frame
- Research scope
- Statistical analysis needs

Determining Sample Size

- How many completed questionnaires needed for a **representative sample**?
  - No absolute answer but generally the larger the better (takes more time and money!).
    - A 1000 sample of purely random from population of 70 million is better than a sample of 7000 that is inaccurately sampled using simple random sampling or sampled using non-probability sample.
- Usually, decision depends on:
  - How different, heterogenous or dispersed the population is.
  - Desired level of (statistical) confidence (a BIG & IMPORTANT topic covered in the discussion on inferential statistics).
  - Desired degree of accuracy.
  - AND response rate: $$\frac{\text{number of usable questionnaires}}{\text{total sample} - \text{unsuitable or uncontactable members of the sample}} \times 100$$

Sampling: additional concepts

- ***Non-response*:** a source of non-sampling error that is particularly likely to happen when individuals are being sampled. It occurs whenever some members of the sample refuse to cooperate, cannot be contacted, or for some reason cannot supply the required data (for example, because of mental incapacity).
- ***Census*:** the enumeration of an entire population. Thus, if data are collected in relation to all units in a population, rather than in relation to a sample of units of that population, the data are treated as census data. The phrase '*the* census' typically refers to the complete enumeration of all members of the population of a nation state—that is, a national census. This form of enumeration currently occurs once every ten years in the UK, although there is some uncertainty at the time of writing about whether another census will take place. However, in a statistical context, like the term *population*, the idea of a census has a broader meaning than this.

Sampling: additional concepts

- ***Sampling bias*:** a distortion in the representativeness of the sample that arises when some members of the population (or more precisely the sampling frame) stand little or no chance of being selected for inclusion in the sample.
- ***Sampling error*:** error in the findings deriving from research due to the difference between a sample and the population from which it is selected. This may occur even though probability sampling has been employed.
  - arises because it is extremely unlikely that one will end up with a truly representative sample, even when probability sampling is employed.
- ***Non-sampling error*:** error in the findings deriving from research due to the differences between the population and the sample that arise either from deficiencies in the sampling approach, such as an inadequate sampling frame or **non-response**, or from such problems as poor question wording, poor interviewing, or flawed processing of data.

Sampling: additional concepts

- ***sampling-related error*.**
  - This is error that is subsumed under the category *non-sampling error* but that arises from activities or events that are related to the sampling process and that are connected with the issue of generalizability or external validity of findings. Examples are an inaccurate sampling frame and non-response.

Errors in Survey Research

- ***data-collection error .***
  - error that is connected with the implementation of the research process. includes such factors as: poor question wording in self- completion questionnaires or structured interviews; poor interviewing techniques; and flaws in the administration of research instruments.
- ***data-processing error.***
  - This arises from faulty management of data, in particular, errors in the *coding* of answers.

The third and fourth sources of error relate to factors that are not associated with sampling and instead relate much more closely to concerns about the validity of measurement.

References

- Neuman, W. Lawrence (2013). Social Research Methods: Qualitative and Quantitative Approaches. Available from: https://cleavermonkey.files.wordpress.com/2015/10/w-lawrence-neuman-social-research-methods_-qualitative-and-quantitative-approaches-pearson-education-limited-2013.pdf
- Bryman, Alan (4th Ed.) (2012). *Social Research Methods* (3rd ed.). Oxford: Oxford University Press. Available from: http://en.bookfi.net/book/2266020