



ECDE7020

ECDE7012 Session

- Qualtrics & SPSS
- Dr. Charlotte Greenway
- Saturday, 21st March
- Room and time to be confirmed



Research design 5-minute challenge

1. What is your research problem?
2. What is/are your research question/s?
3. What research design may help you respond to your research question(s)?
4. What data collection tools might you use? (Why?)



SAMPLING

What is a sample?

Why use a sample?





What is a sample?

When you conduct research about a group of people, it is rarely possible to collect data from every person in that group.

Instead, you select a **sample**.

The sample is the group of individuals who will actually participate in the research.

Why use a sample?

Limited Resources: It is often impractical to study an entire population due to time, budget and logistical constraints.

Generalisability: The goal (sometimes!) is to draw valid conclusions about the entire population based on the data collected from a sample.

Reducing Bias: A carefully chosen sample can help minimise bias and ensure that research findings are more representative of the population.

Sampling errors and non-sampling errors



The term 'sampling error' doesn't mean mistakes researchers have made when selecting a sample.

Problems like choosing the wrong people, letting bias enter the picture, or failing to anticipate that participants will self-select or fail to respond - these are **non-sampling errors**.



Sampling errors and non-sampling errors

A **sampling error** is a statistical error that occurs when a sample taken from a population does not accurately represent the entire population.

This error arises because the sample is only a subset of the population, and the characteristics of the sample may differ from the characteristics of the entire.

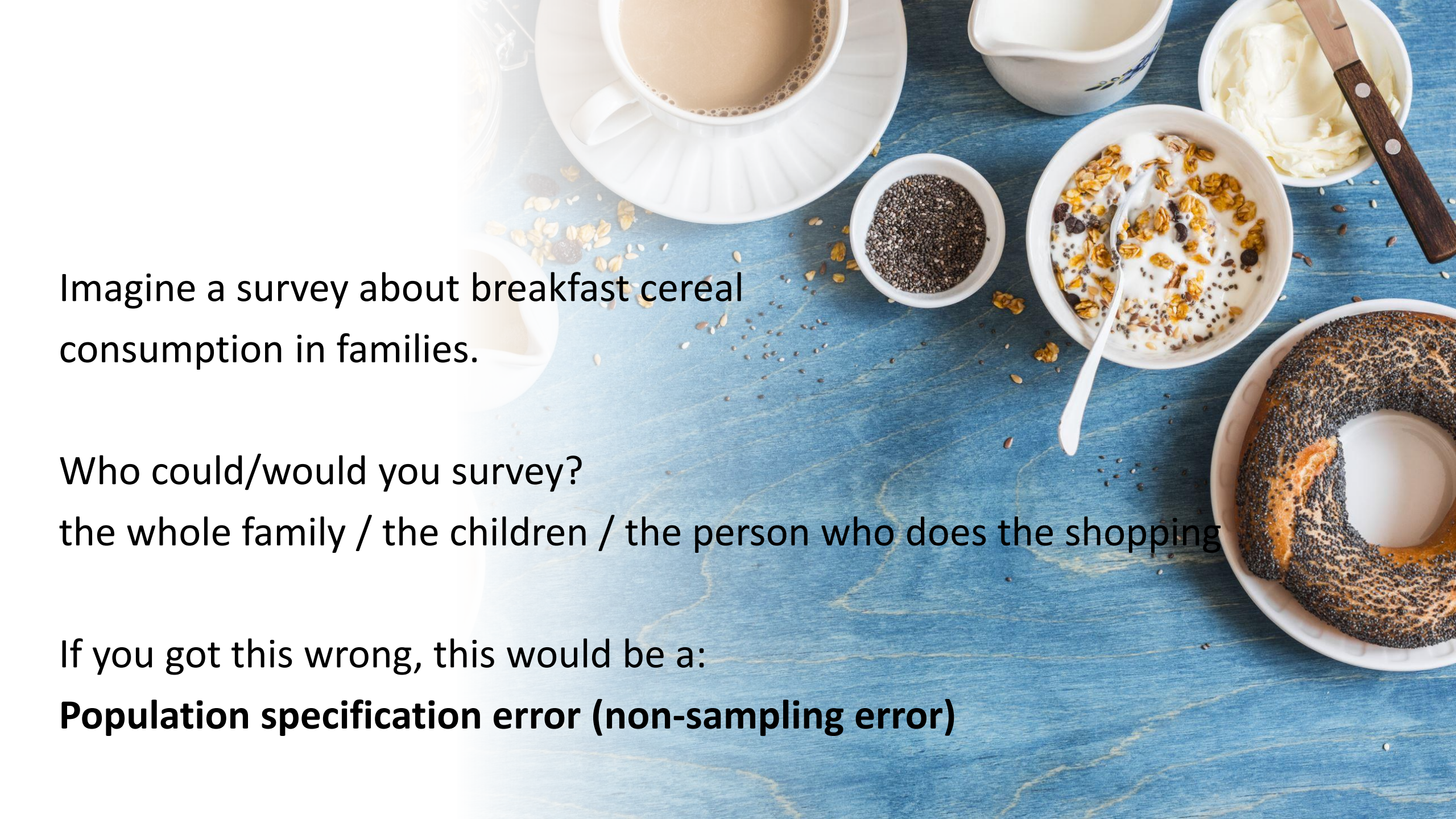
For example, if you measure the height of a thousand individuals from a population of one million, the average height of the thousand is typically not the same as the average height of all one million people.

Sampling errors and non-sampling errors

- Sampling errors occur because of variation in the number or representativeness of the sample that responds.
- Sampling errors can be controlled and reduced by:
 - (1) careful sample designs;
 - (2) large enough samples;
 - (3) multiple contacts to ensure a representative response.

Non-sampling
errors in research





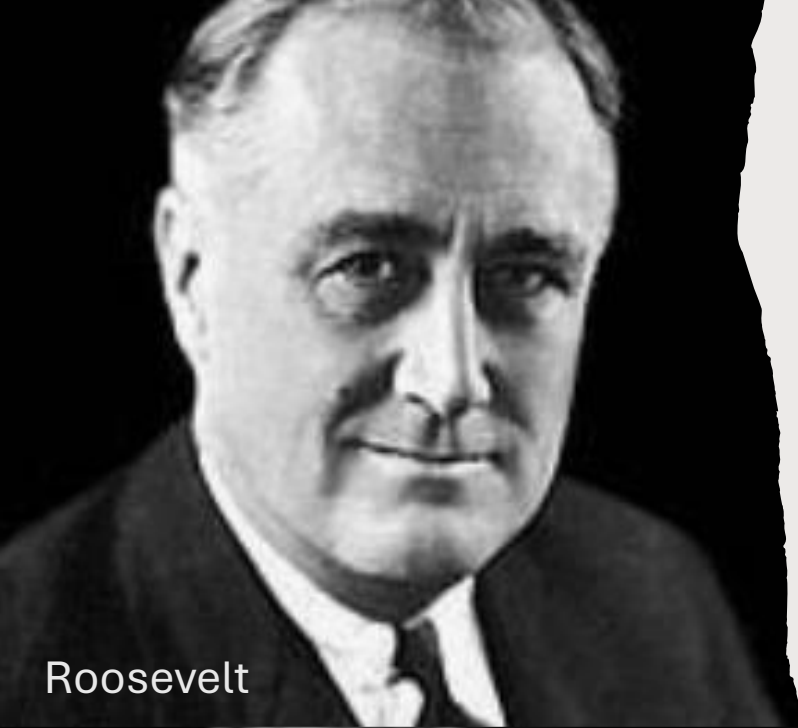
Imagine a survey about breakfast cereal consumption in families.

Who could/would you survey?

the whole family / the children / the person who does the shopping

If you got this wrong, this would be a:

Population specification error (non-sampling error)



Roosevelt



Landon

A famous **xxx** error occurred in the 1936 US presidential election between Roosevelt (the Democratic candidate) and Landon (the Republican candidate).

The sample was taken from car registrations and telephone directories.

In 1936, many Americans did not own cars or telephones. Those who did were largely Republicans.

The results wrongly predicted a Republican victory.

This is an example of a:
Sample frame error (non-sampling error)



In a study, respondents self-selected their participation in the study.

For example, the researcher has put out a call for responses on social media.

What are the possible issues with this approach?

- They are going to get responses from people they know;
- Of those people, only the more helpful individuals will reply.

This is an example of a:

Selection error (non-sampling error)

SELECTION BIAS

DEFINITION

Selection bias refers to a distortion in research outcomes caused by the non-random or biased selection of participants for a study, leading to a sample that is not representative of the entire population of interest. This bias can compromise the generalizability and accuracy of research findings, impacting both external and internal validity.

EXAMPLES

- Sampling Bias
- Self-Selection Bias
- Exclusion Bias
- Berkson's bias
- Pre-Screening Of Subjects
- Cherry-Picking Bias
- Undercoverage Bias
- Survivorship Bias
- Attrition Bias

When the research attracts people, e.g. who are ill and seeking treatment rather than healthy individuals, leading to an overrepresentation of one group. (Barker et al., 2016)

When a researcher follows a vetting process to determine whether potential participants are suitable for a study (Walliman, 2021).

Selectively presenting, emphasizing, or excluding data that support a particular conclusion while neglecting significant data that may contradict it (Bryman, 2015).

When some groups of the population are inadequately represented in the sample (Creswell, 2013).

When the focus is solely on the subjects that 'survived' or succeeded, dismissing those that failed or dropped out (Atkinson et al., 2021).

When participants exit a study before its completion, leading to a skewness in the final results (Bryman, 2015).

You're a company doing market research in advance of launching a new product.

Any issues?

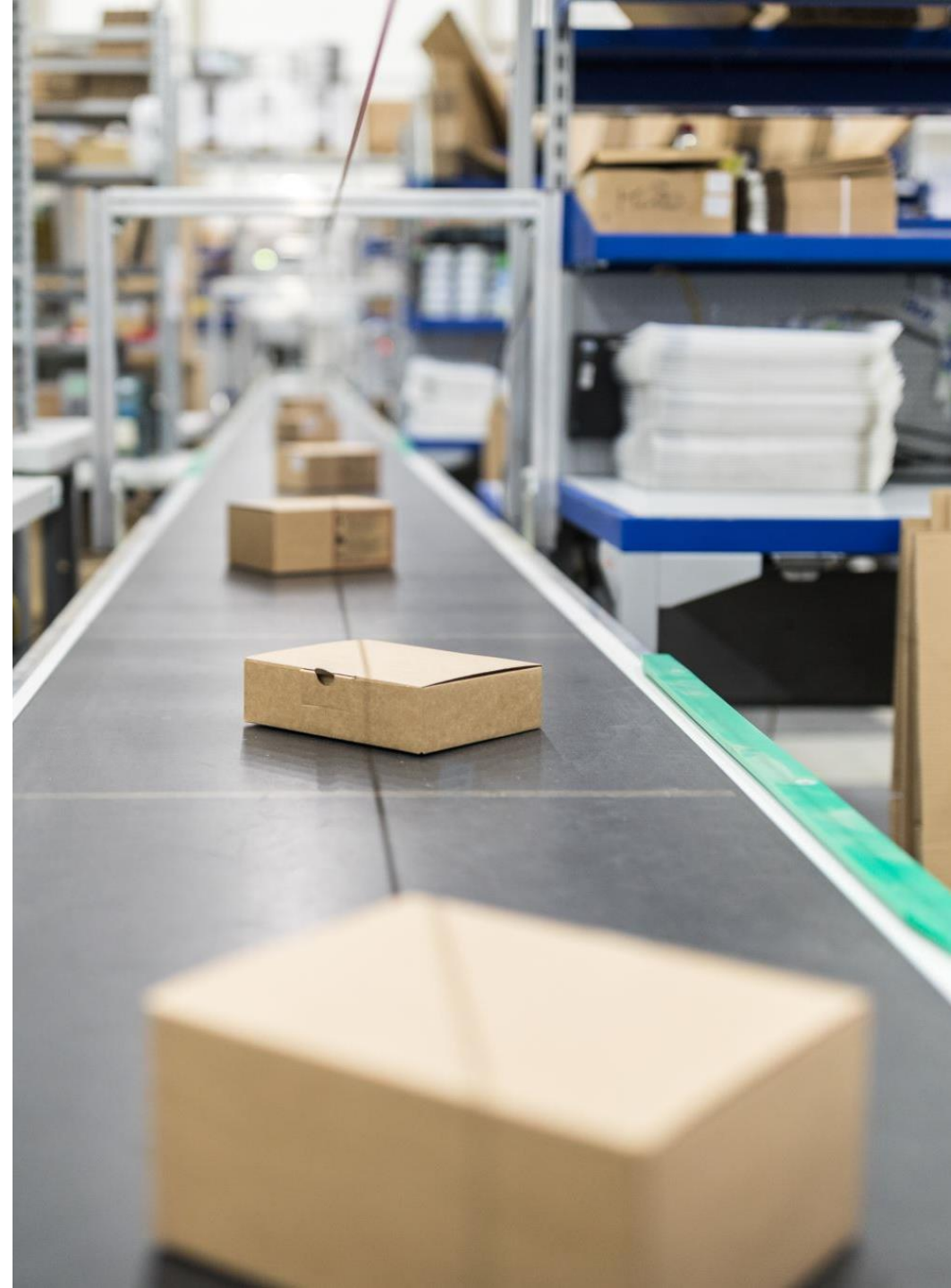
You might get a disproportionate level of participation from your existing customers, since they know who you are.

You might miss out on hearing from a broader pool of people who don't yet buy from you.

This is an example of a:

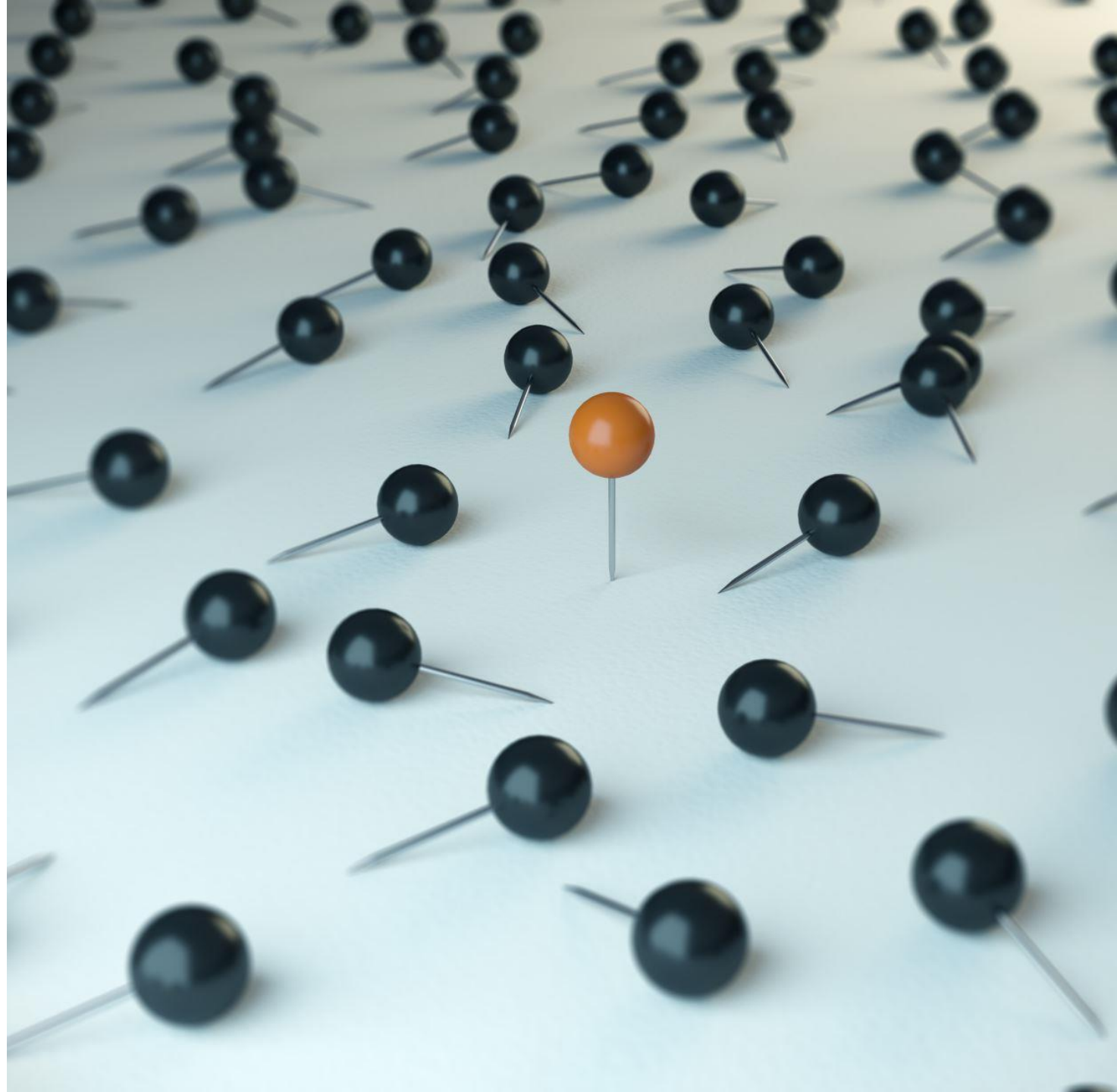
Non-response (non-sampling error)

Non-response errors occur when those who **do** respond are different than those who **don't**.



Probability and non-probability sampling techniques

Flick, U. (2014) An Introduction to Qualitative Research. 5th Edition, London: Sage Publications.



$$F = G \frac{m_1 m_2}{d^2}$$

Probability sampling methods

$$F - E + V = 2$$

$$i\hbar \frac{\partial}{\partial t} \psi = \hat{H} \psi$$

$$\phi(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}}$$

$$E = mc^2$$

$$ds \geq 0$$

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}$$

$$\frac{df}{dt} = \lim_{h \rightarrow 0} \frac{f(t+h) - f(t)}{h}$$

Probability sampling methods

Probability sampling means that every member of the population has a chance of being selected.

It is often used in quantitative research.

If you want to produce results that are representative of the whole population, probability sampling techniques are the most valid choice.

Simple random sample

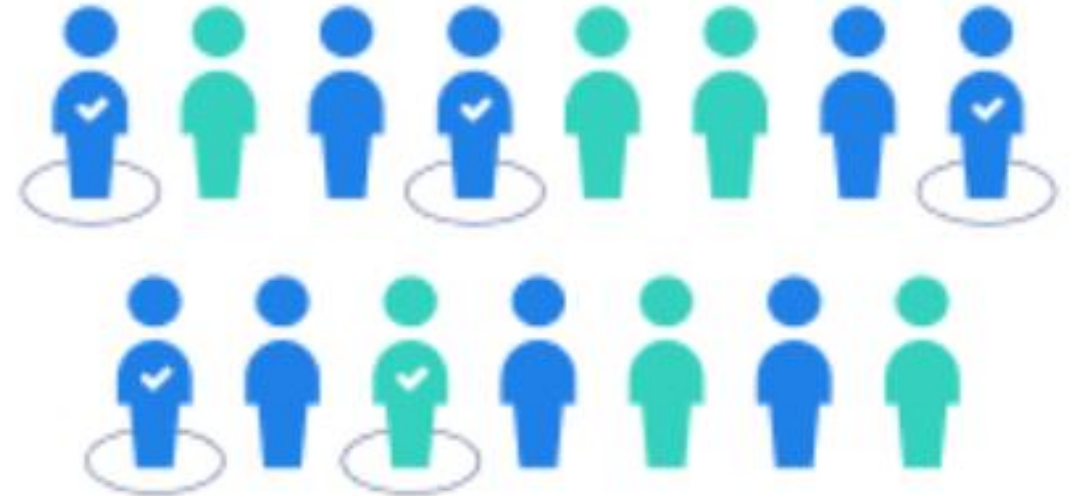


Every member of the population has an equal chance of being selected. Your sampling frame should include the whole population. Use tools like random number generators that are based entirely on chance.

Benefit:

Simple random samples are usually representative of the population we are interested in since every member has an **equal chance** of being included in the sample.

Simple random sample





You want to select a **simple random sample** of 10% of 1000 employees of a marketing company.

Can you do it? How?

Assign a number to every employee in the company database from 1 to 1000.

Use a random number generator to select 100 numbers.

Simple random sample

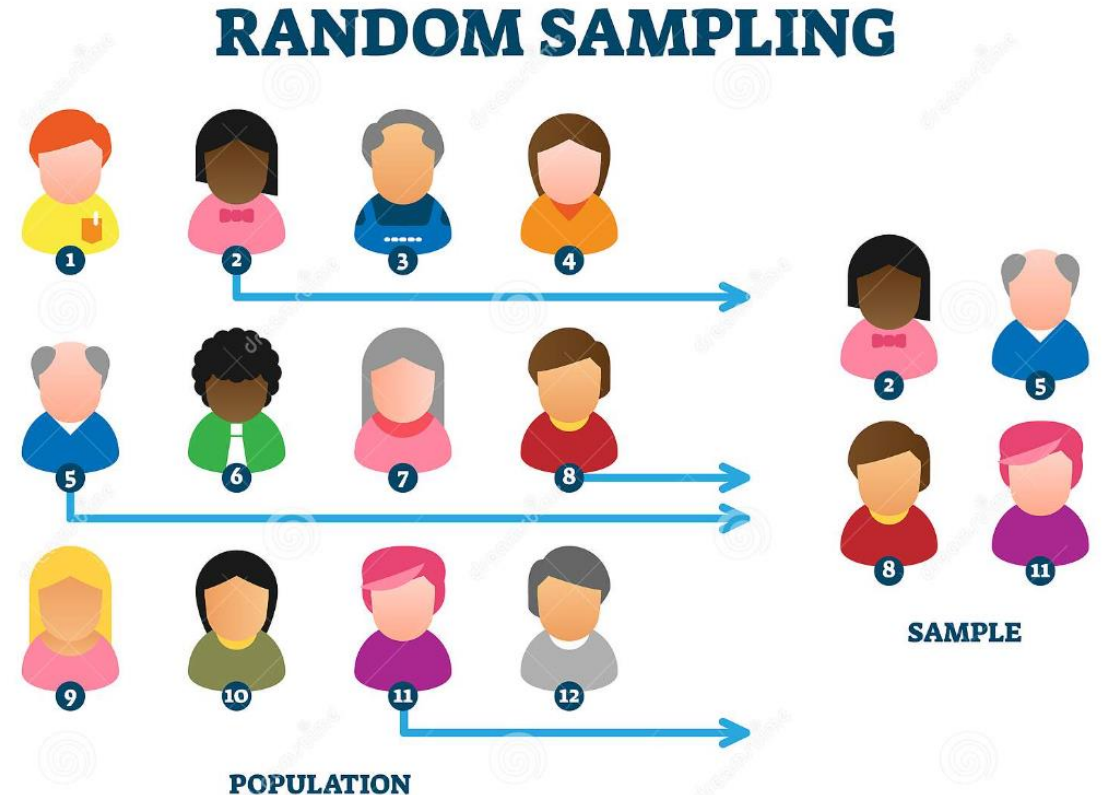


Systematic random sample

Similar to simple random sampling, but slightly easier to conduct. Every member of the population is listed with a number, but instead of randomly generating numbers, individuals are chosen at regular intervals.

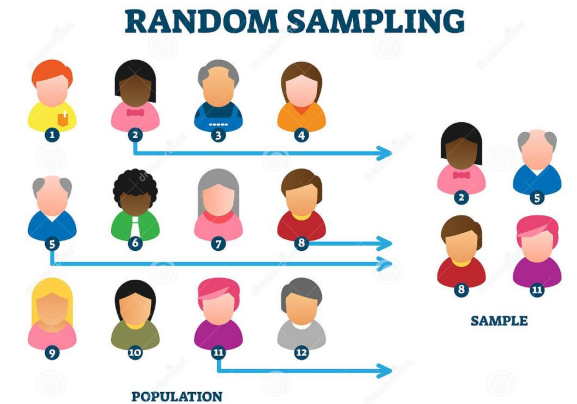
Benefit:

Systematic random samples are usually representative of the population we are interested in since every member has an equal chance of being included in the sample.



You want to select a **systematic random sample** of 10% of 1000 employees of a marketing company.

Can you do it? How?



All employees of the company are listed in alphabetical order.

From the first 10 numbers, you randomly select a starting point: number 6.

From number 6 onwards, every 10th person on the list is selected (6, 16, 26, 36, and so on).

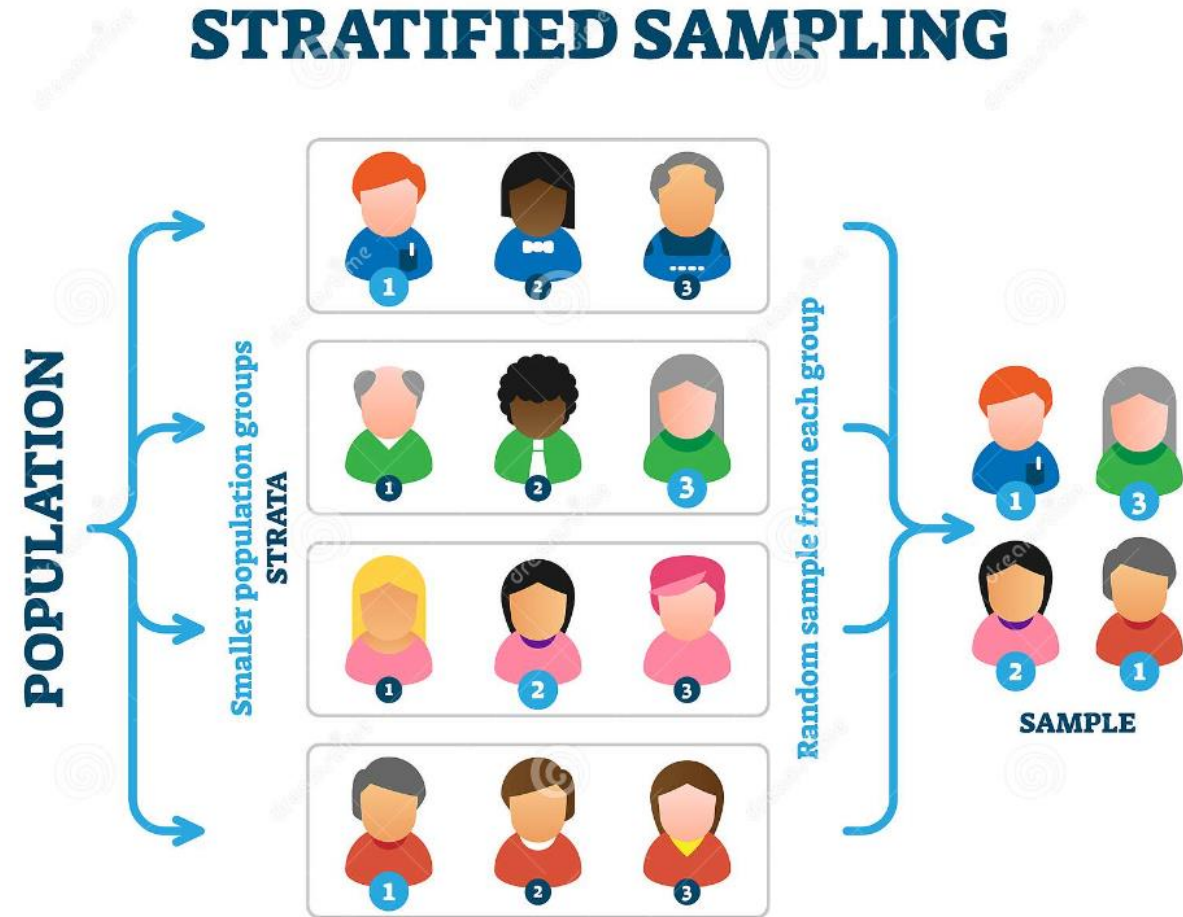
You end up with a sample of 100 people.

Stratified random sample

This involves dividing the population into sub-populations that may differ in important ways.

Benefit:

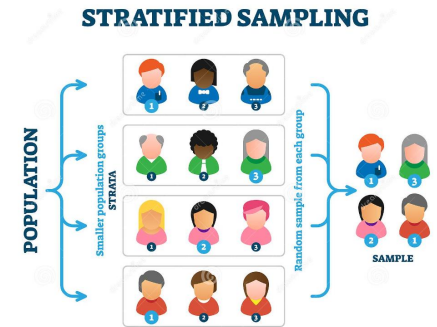
Stratified random samples ensure that an equal number of people from each subpopulation are included in the sample so it allows you draw more precise conclusions.



You want to select a **stratified random sample** of 10% of 1000 employees of a marketing company.

The company has 800 female employees and 200 male employees.

Can you do it? How?



You want to ensure that the sample reflects the gender balance of the company, so you sort the population into two strata based on gender.

Then you use random sampling on each group, selecting 80 women and 20 men, which gives you a representative sample of 100 people.

Cluster sampling

Involves dividing the population into sub-groups, but each sub-group should have similar characteristics to the whole sample. Instead of sampling individuals from each subgroup, you randomly select entire subgroups.

Benefit:

This method is good for dealing with large and dispersed populations.

Cluster sample



Limitation

There is more risk of error in the sample, as there could be substantial differences between clusters. It's difficult to guarantee that the sampled clusters are really representative of the whole population.

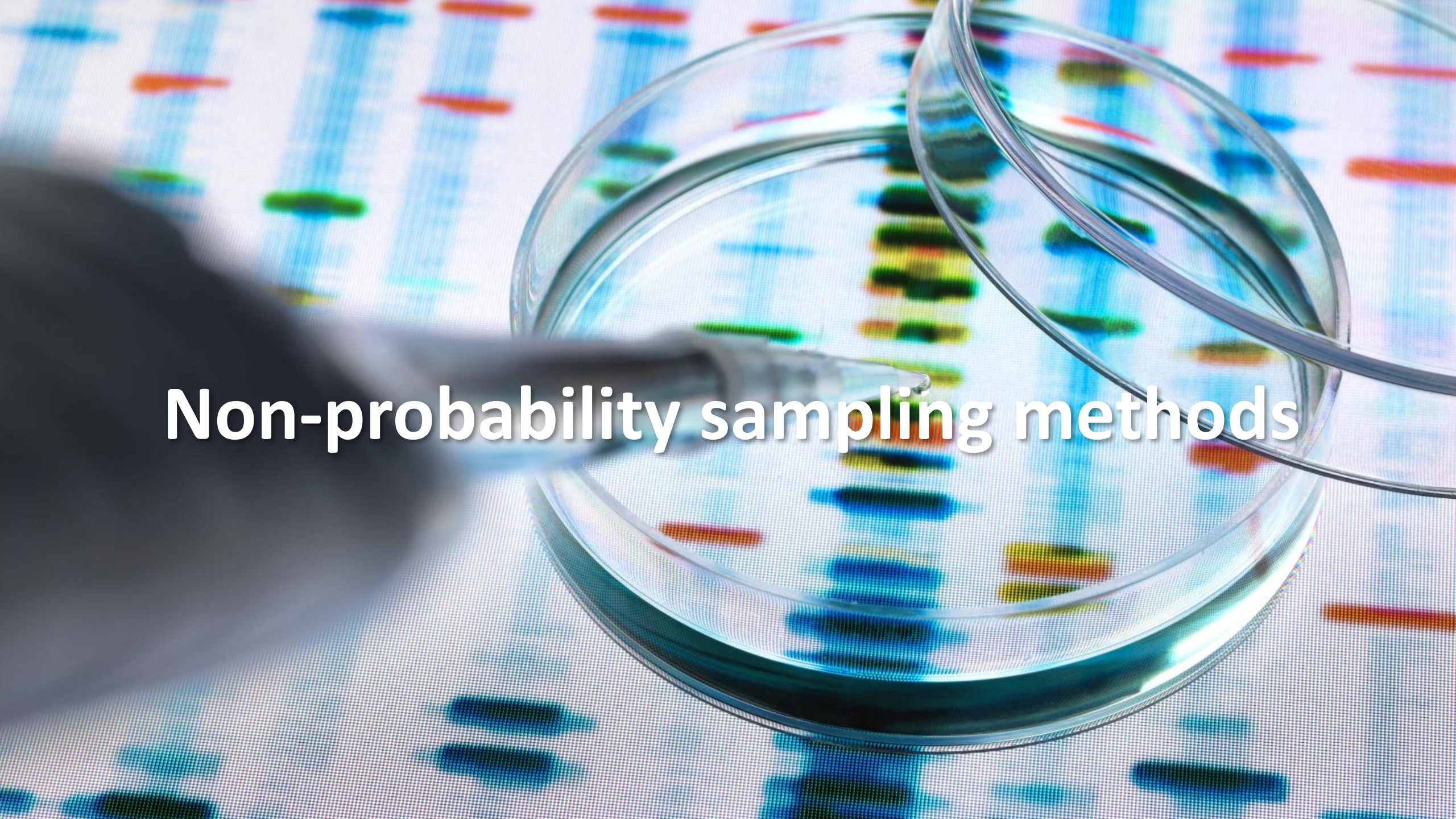


You want to select a **cluster sample** of 10% of 1000 employees of a marketing company.

The company has offices in 10 cities across the country - all with roughly the same number of employees in similar roles.

Can you do it? How?

You don't have the time or money to travel to every office to collect your data, so you use random sampling to select 3 offices. These are your clusters.

A magnifying glass is held over a DNA gel electrophoresis image. The gel shows multiple lanes with horizontal bands of various colors (blue, green, yellow, orange, red) representing different DNA fragments. The magnifying glass is positioned over one of the lanes, and its handle is visible on the left side of the frame. The text "Non-probability sampling methods" is overlaid in white, bold font across the center of the image.

Non-probability sampling methods

Non-probability sampling methods

Individuals are selected based on **non-random criteria**.

Strengths

- *Easy - convenient
- *Accessible - useful when the population of interest is difficult to access or to define
- *Cost effective

Limitations

- *High risk of **sampling bias** = the inferences you can make about the population are weaker than with probability samples and your conclusions may be limited.
- *Lack of generalisability

Non-probability sampling techniques are often used in exploratory and **qualitative research**, where the aim is to develop an initial understanding of a small or under-researched population.

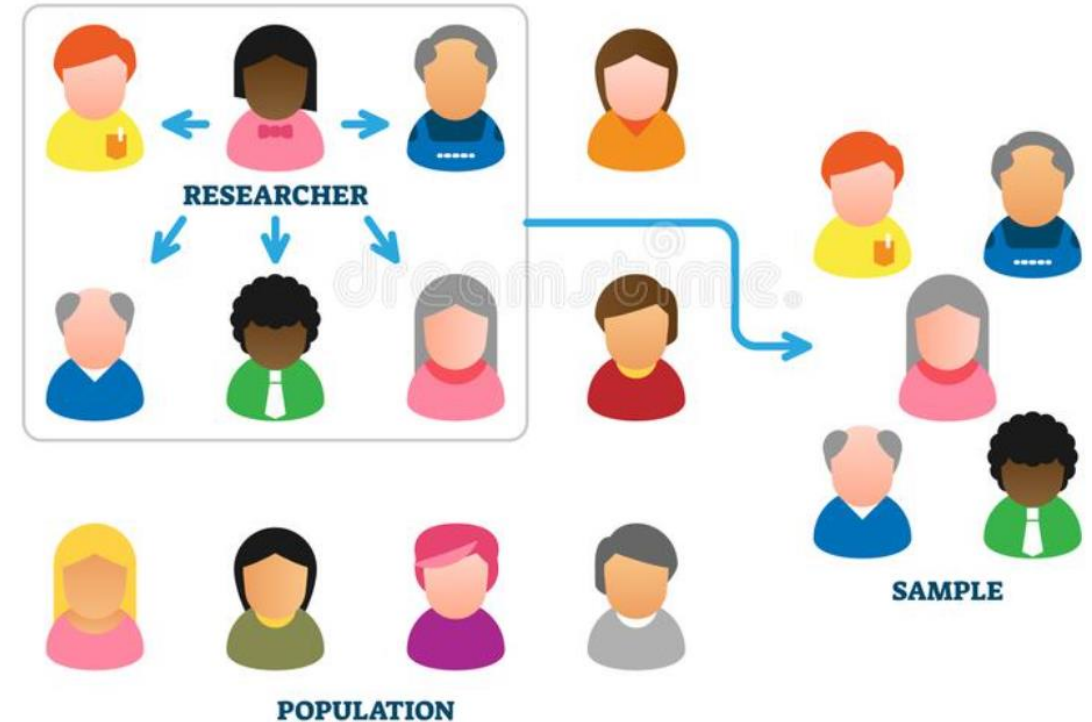
CONVENIENCE SAMPLING

Convenience sample

Deliberately choose a group of people who are 'most convenient' = people who happen to be most accessible to you.

Benefit:

You have participants that are likely to give you the information you want to get



Limitations:

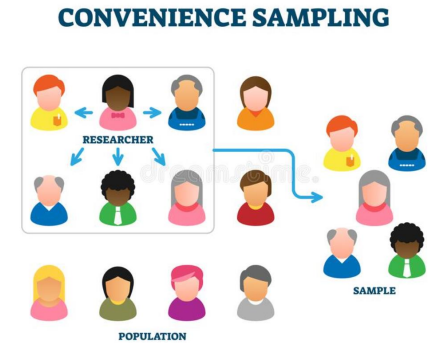
Does not produce a representative sample of the population and this can distort the results.

Findings cannot be generalised to the wider population.

Creswell, J. (1994) Research Design: Qualitative, Quantitative, and Mixed Methods Approaches.

You want to research opinions about student support services in your university through a **convenience sample**.

Can you do it? How?



After our lecture today, ask students to complete a questionnaire on the topic. This is a convenient way to gather data.

However, as you only surveyed students taking the same class as you at the same level, the sample is not representative of all the students at the university.

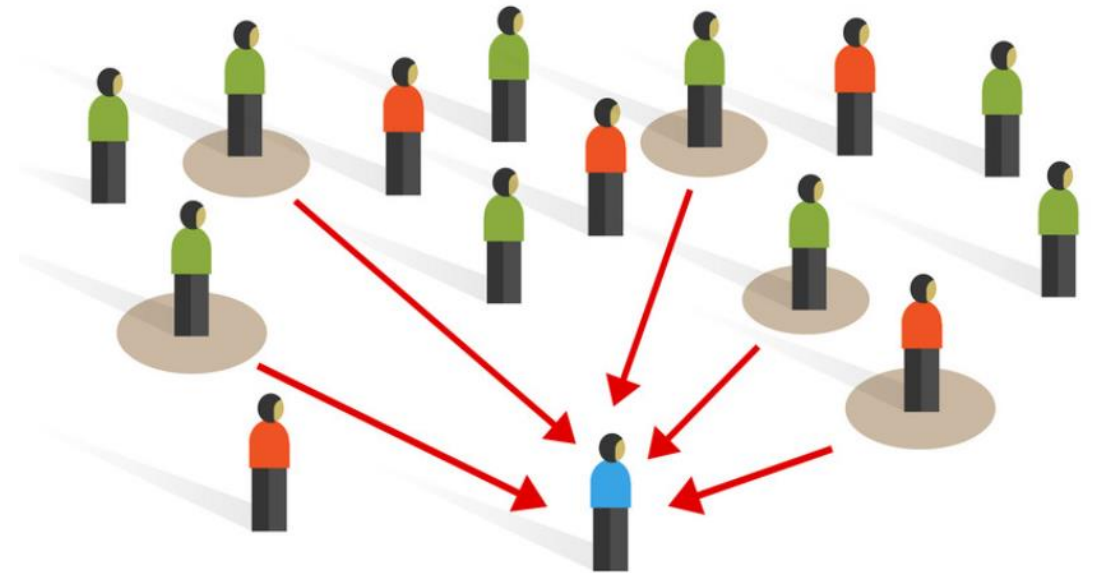
Voluntary response sampling

Mainly based on ease of access.

Instead of the researcher choosing participants and directly contacting them, people volunteer themselves.

Benefit:

A sense of engagement and involvement among respondents as they actively choose to participate.



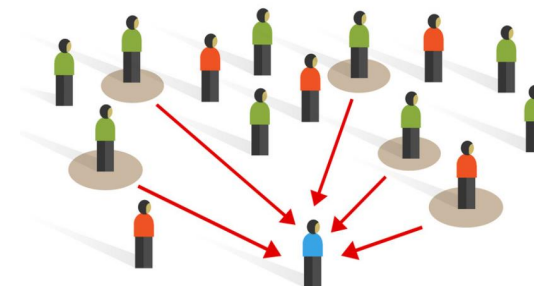
Limitation

Voluntary response samples are always at least somewhat **biased**, as some people will inherently be more likely to volunteer than others, leading to **self-selection bias**.*

*Rosenbaum, P. R. (2010). Design of Observational Studies. Springer.

You want to research opinions about student support services in your university through a **voluntary response sample**.

Can you do it? How?



Send out the survey to all students at your university and hope that a lot of students decide to complete it.

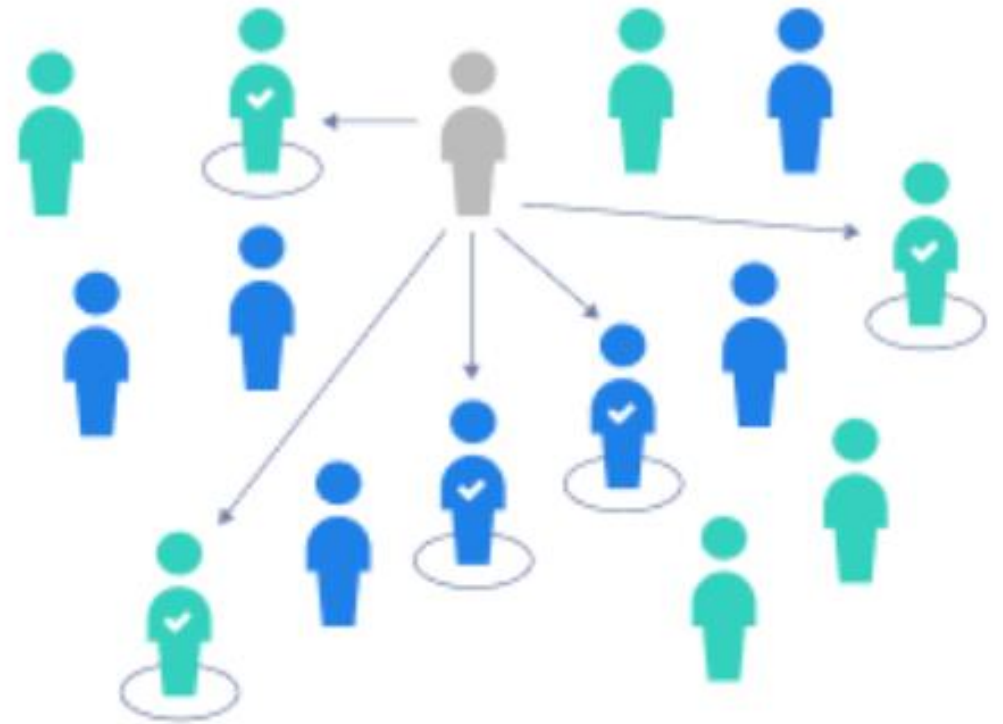
However, the people who responded are more likely to be those who have strong opinions about student support services, so you can't be sure that their opinions are representative of all students.

Purposive sampling

Involves the researcher using their expertise to select a sample that is most useful to the purposes of the research.

Must have clear criteria and rationale for inclusion - always make sure to describe your **inclusion and exclusion criteria** and beware of **observer bias** affecting your arguments.

Purposive sample



You want to research opinions about student support services in your university through **purposive sampling**.

You want to know more about the opinions and experiences of disabled students at your university.

Can you do it? How?

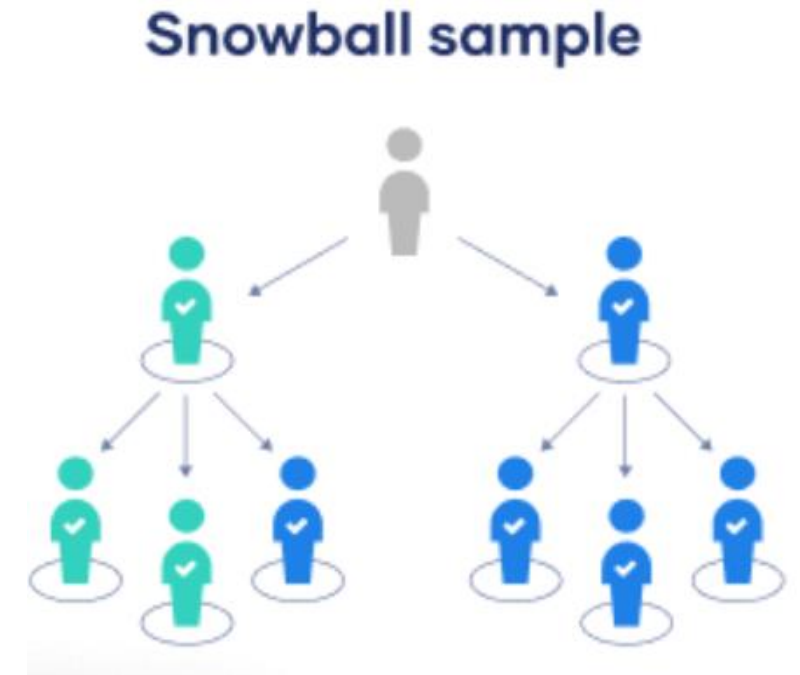


Purposefully select a number of students with different support needs in order to gather a varied range of data on their experiences with student services.

Snowball sampling

If the population is challenging to access, snowball sampling can be used to recruit participants via other participants.

The number of people you have access to “snowballs” as you get in contact with more people.



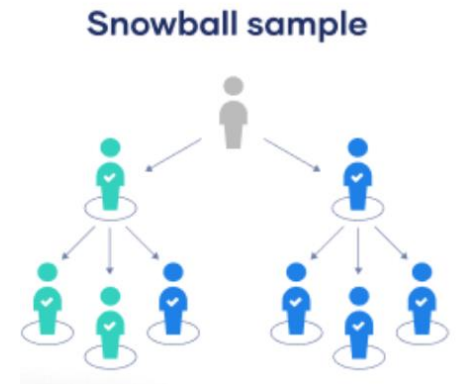
Limitation

You have no way of knowing how representative your sample is due to the reliance on participants recruiting others. This can lead to **sampling bias**.

You want to research opinions about student support services in your university through **snowballing sampling**.

You want to know more about the opinions and experiences of students who have blue eyes at your university. There is no list of people with blue eyes available.

Can you do it? How?



You meet one person who agrees to participate in the research and (s)he puts you in contact with other people who have blue eyes.

Quota Sample

Relies on the non-random selection of a predetermined number or proportion of units. This is called a quota.

Divide the population into meaningful subgroups (called strata) that are based on certain characteristics.

Then, select a specific number of participants from each subgroup until you reach your quota.

The aim of quota sampling is to control who makes up your sample.



You want to research opinions about student support services in your university through **quota sampling**.

You want to gauge interest in new English language sessions. You want the sessions to be suitable for students who have reached different levels of fluency.

Can you do it? How?



Divide the population into groups – beginners, intermediate, advanced.

Set a quota of 200 people for each group.

Continue recruiting until you reach the quota of 200 participants for each sub-group.

Follow-up thinking

Follow up thinking task:

Consider Assessment Component 2: Research Methods / Tools

- What sampling method(s) do you propose to use? Search for relevant literature in order to **justify your choice**.
- How will you use the sampling method(s) **within your specific research context?**

Assignment Component 2

Sampling	What is/are your sampling method(s)? Argue the case for its/their use, including the strengths and limitations of your approach. Refer to literature where possible.	500
----------	--	-----



Presenting Sampling Information (in a thesis)



In which section of thesis should a discussion about sampling go?

Where is it?

Chapter 4: Methodology.....	57
-----------------------------	----

4.1 Overview	57
--------------------	----

iii

4.2 Research paradigm	57
-----------------------------	----

4.3 Research design	58
---------------------------	----

4.3.1 Introduction	58
--------------------------	----

4.3.2 Sampling approach.....	60
------------------------------	----

4.3.3 Research tools.....	63
---------------------------	----

4.3.4 Interview preparation	69
-----------------------------------	----

4.3.5 Pilot study	70
-------------------------	----

4.4 Approach to data analysis.....	71
------------------------------------	----

4.5 Ethical considerations	75
----------------------------------	----

4.6 Reflexivity.....	77
----------------------	----

Chapter 4: Methodology.....	51
------------------------------------	-----------

4.1 Introduction.....	51
-----------------------	----

4.2 Context of the Study.....	51
-------------------------------	----

4.3 Sequence of Data Collection and Analysis.....	54
---	----

4.4 Ethical Considerations.....	56
---------------------------------	----

4.4.1 Consent.....	56
--------------------	----

4.4.2 Power Relations.....	58
----------------------------	----

4.4.3 Gatekeeper Consent.....	61
-------------------------------	----

4.4.4 Confidentiality and Anonymity.....	62
--	----

4.5 Research Methods: Whole-cohort Questionnaires.....	63
--	----

4.6 Research Methods: Case Study.....	71
---------------------------------------	----

4.6.1 Case Study Participants.....	73
------------------------------------	----

4.6.2 Case Study Data Sources.....	77
------------------------------------	----

4.6.3 Case Study Observation.....	77
-----------------------------------	----

4.6.4 Case Study Trainee Questionnaires.....	78
--	----

4.6.5 Case Study Classteacher Data Collection.....	80
--	----

4.6.6 Case Study Document Scrutiny.....	83
---	----

4.6.7 Case Study Interviews.....	84
----------------------------------	----

4.8 Data Analysis.....	89
------------------------	----

4.8.1 Presentation of Case Study data.....	89
--	----

4.8.2 Thematic Analysis.....	90
------------------------------	----

4.8.3 Ensuring trustworthiness.....	94
-------------------------------------	----

4.8.4 Questionnaire data presentation and analysis.....	97
---	----

4.9 Summary.....	98
------------------	----

Sampling - Thesis 1

The four school leaders were selected using purposive sampling. A number of contributing factors relating to both the school leaders and the researcher were considered. With regards to the former, it is important to note that all participants were known to me, and I already had a working, professional relationship with them. This brought with it strengths and weaknesses. On the one hand, it allowed me to recruit participants myself I was more prone to researcher bias (Shaw, 2009). However, by my careful structuring of interview questions, detailed coding and a well-defined approach to data analysis. I also recognise that having

Given the nature of my role in education, and my many years' experience of working with school leaders on a variety of projects, it was not difficult to source from my existing network a diverse group of possible participants from a range of schools across Wales. Participants were purposively sampled and chosen on the basis that they and their school met the following criteria:

- The school was last inspected within the last complete inspection cycle (2016-2022);
- The participating school leader was in post at the time of the last inspection;
- The participating school leader holds a senior position within the school's staffing structure, and thus forms part of the school's senior management team. They must hold one of the following posts:
 - Headteacher;
 - Deputy headteacher;
 - Assistant headteacher;
 - A role similar to those identified above;
- There has been no significant structural change to the school (e.g. merger/federation) since the last inspection.

Sampling - Thesis 1

- Two primary and two secondary schools, all English-medium in type (Welsh-medium schools were available and met the criteria, but I am not a Welsh speaker);
- Schools of different sizes (e.g. fewer/more than 500 pupils; fewer/more than 1,000 pupils);
- Schools in different geographical locations (e.g. north, south, east, west and in areas of high deprivation/low deprivation);
- Schools with contrasting performance in their last inspection (e.g. range of 'Excellent' and/or 'Good'; 'Adequate and needs improvement' and/or 'Unsatisfactory and needs urgent improvement').

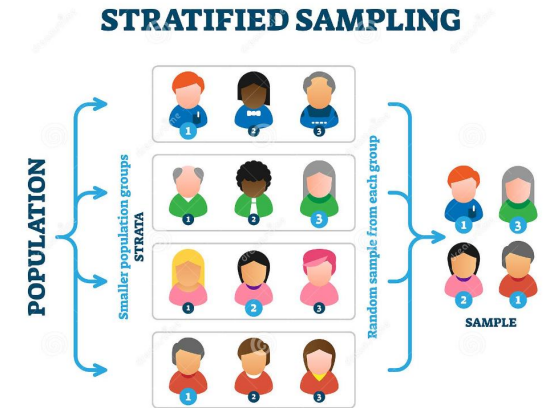
Sampling - Thesis 2

In order to ensure a representative sample, a process of random stratified sampling was carried out to, where trainees were divided “into homogenous groups, each group containing subjects with similar characteristics” (Cohen, Manion & Morrison, 2011:154). It was decided that there should be three elements that would determine the sample:

1. Prior achievement in Welsh language
2. Male: Female ratio
3. Geographical area

There were 61 Trainee Teachers at this point.

1. Trainees with no prior knowledge of Welsh;
2. Trainees who had studied Welsh up to Year 9;
3. Trainees who had studied GCSE W2L Short Course;
4. Trainees who had studied GCSE W2L Full Course;
5. Trainees who had studied AS level W2L;
6. Trainees who had studied A level W2L;
7. Trainees who had studied GCSE Welsh First Language.



Sampling - Thesis 2

The proportions of trainees within each category can be seen in Figure 2.

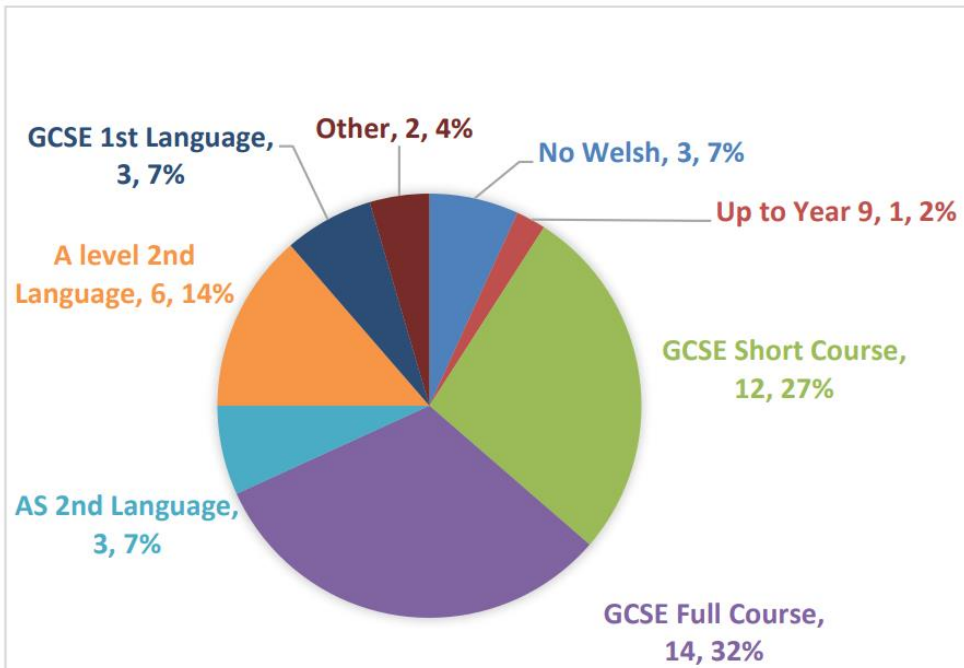


Figure 2: Welsh Previously Studied

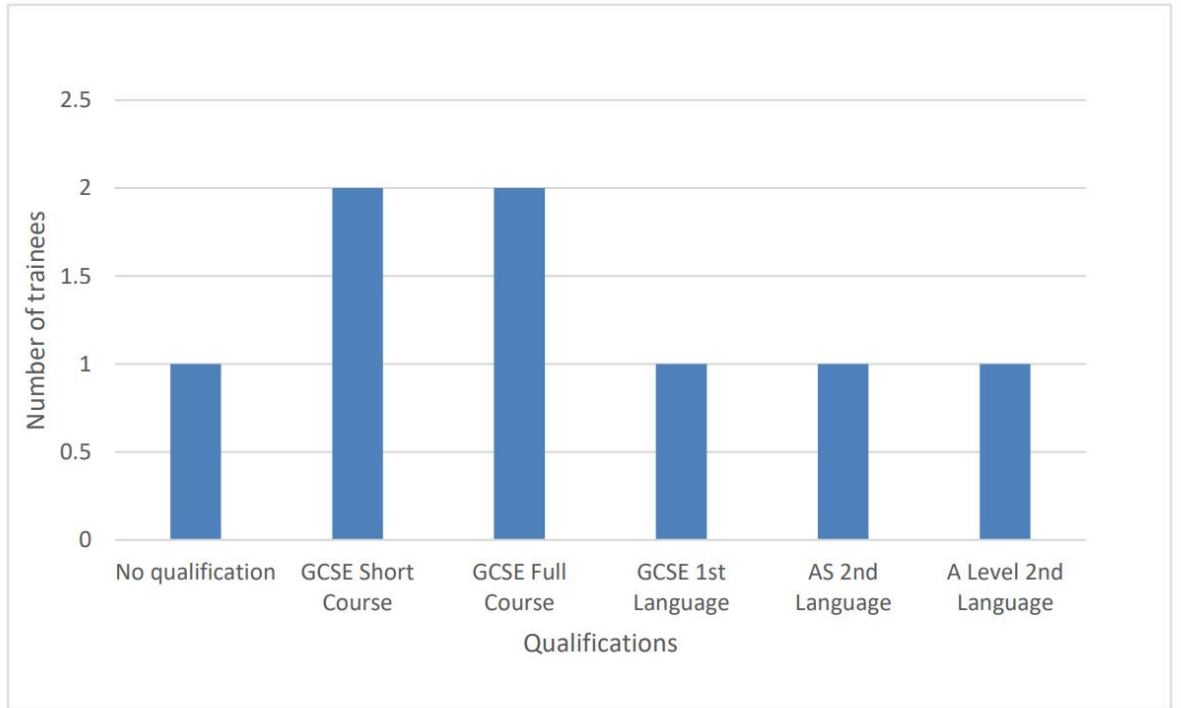


Figure 3: Welsh Language Qualifications of Case Study Trainees

n=61

Sampling - Thesis 2

In order to ensure a representative sample, a process of random stratified sampling was carried out to, where trainees were divided “into homogenous groups, each group containing subjects with similar characteristics” (Cohen, Manion & Morrison, 2011:154). It was decided that there should be three elements that would determine the sample:

1. Prior achievement in Welsh language
2. Male: Female ratio
3. Geographical area

Secondly, it was considered that the numbers of male and female trainees within the sample to be observed should be representative of the whole cohort. The total number in the cohort being placed in a school for PTE2 was 61, with 48 female trainees and 13 male trainees. Therefore, it was decided that six female trainees and two male trainees would be chosen as a representative sample of the whole cohort.

Sampling - Thesis 2

In order to ensure a representative sample, a process of random stratified sampling was carried out to, where trainees were divided “into homogenous groups, each group containing subjects with similar characteristics” (Cohen, Manion & Morrison, 2011:154). It was decided that there should be three elements that would determine the sample:

1. Prior achievement in Welsh language
2. Male: Female ratio
3. Geographical area

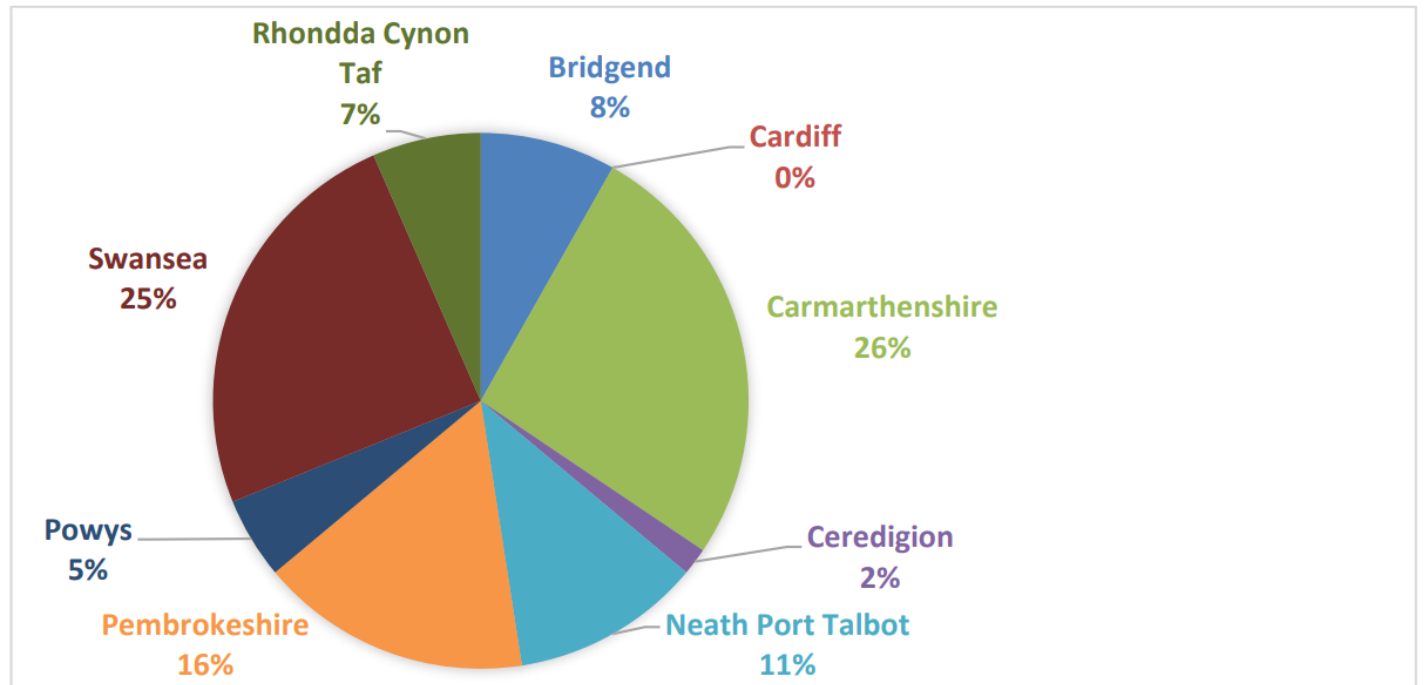


Figure 4: PTE2 Locations

Sampling - Thesis 2

	Trainee 1 Group 6, Female		Trainee 2 Group 4, Female		Trainee 3 Group 1, Female		Trainee 4 Group 7, Female		Trainee 5 Group 3, Female		Trainee 6 Group 4, Male		Trainee 7 Group 5, Male		Trainee 8 Group 3, Female	
	Year2	Year3	Year2	Year3	Year2	Year3	Year2	Year3	Year2	Year3	Year2	Year3	Year2	Year3	Year2	Year3
Bridgend												X	X	X		
Cardiff		X														
Carms.	X			X												
Ceredigion			X													
NPT									X	X						
Pembs.					X	X										
Powys																
RCT											X					
Swansea							X	X							X	X
V. of Glam.																

Carms. = Carmarthenshire
 Pembs. = Pembrokeshire
 V. of Glam. = Vale of Glamorgan

NPT = Neath Port Talbot
 RCT = Rhondda Cynon Taf

Figure 5: Characteristics of case study trainees

Next week: 05/03/26

Individual Tutorials, online

- 11-11.30
- 11.30-12

12/03/05

- Ethics (please download the ethics form onto your laptop)
- IQ002c, 9am

A photograph of a dining table with several plates of food. In the center, a white bowl contains a salad with white rice, green leafy vegetables, and red tomatoes. To the left, a plate has a tomato-based salad with green herbs. To the right, another plate shows a salad with white rice and red tomatoes. In the foreground, there are plates with yellow and red vegetables, possibly zucchini and tomatoes, and a plate with green vegetables and orange carrots. Hands are visible holding silver forks, indicating people are eating. The text "See you next week!" is overlaid in white, centered on the image, with a white horizontal line underneath it.

See you next week!