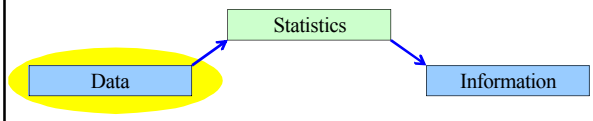
Sampling Methods and Sampling Size

Recall...

Statistics is a tool for converting ***data*** into ***information***:



But where then does ***data*** come from? How is it gathered? How do we ensure its accurate? Is the data reliable? Is it representative of the population from which it was drawn? This chapter explores some of these issues.

Methods of Collecting Data...

There are many methods used to collect or obtain data for statistical analysis. Three of the most popular methods are:

- Direct Observation
- Experiments, and
- Surveys.

 5.3

Topics

Quantitative sampling
- Selecting random samples
- Selecting non-random samples

Qualitative sampling
- Selecting purposive samples

Quantitative Sampling

Purpose – to identify participants from whom to seek some information

Issues
- Nature of the sample
- Size of the sample
- Method of selecting the sample

Quantitative Sampling

Terminology

Population: all members of a specified group
- Target population – the population to which the researcher ideally wants to generalize
- Accessible population – the population to which the researcher has access

Sample: a subset of a population

Subject: a specific individual participating in a study

Sampling technique: the specific method used to select a sample from a population

Obj. 1.1, 1.2, & 1.3

Quantitative Sampling

Important issues

Representation – the extent to which the sample is representative of the population
- Demographic characteristics
- Personal characteristics
- Specific traits

Generalization – the extent to which the results of the study can be reasonably extended from the sample to the population

Obj. 1.4

Quantitative Sampling

Important issues (continued)

Sampling error
- The chance occurrence that a randomly selected sample is not representative of the population due to errors inherent in the sampling technique
- Random nature of errors
- Controlled by selecting large samples

Obj. 6.1

Quantitative Sampling

Important issues (continued)

Sampling bias
- Some aspect of the researcher's sampling design creates bias in the data
- Non-random nature of errors
- Controlled by being aware of sources of sampling bias and avoiding them

Examples
1. Surveying only students who attend additional help sessions in a class
2. Using data returned from only 25% of those sent a questionnaire

Obj. 6.2

Quantitative Sampling

Important issues (continued)

Three fundamental steps
- Identify a population
- Define the sample size
- Select the sample

Obj. 1.5

Quantitative Sampling

Important issues (continued)

General rules for sample size
- As many subjects as possible
- Thirty (30) subjects per group for correlational, causal-comparative, and true experimental designs
- Ten (10) to twenty (20) percent of the population for descriptive designs

Obj. 1.8

Quantitative Sampling

Important issues (continued)

General rules for sample size (continued)

See Table 4.2 for additional guidelines for survey research
- The larger the population size, the smaller the percentage of the population needed to get a representative sample
- For population of less than 100, use the entire population
- If the population is about 500, sample 50%
- If the population is about 1,500, sample 20%
- If the population is larger than 5,000, sample 400

Obj. 1.9

Surveys...

A ***survey*** solicits information from people; e.g. Gallup polls; pre-election polls; marketing surveys.

The ***Response Rate*** (i.e. the proportion of all people selected who complete the survey) is a key survey parameter.

Surveys may be administered in a variety of ways, e.g.
- Personal Interview,
- Telephone Interview,
- Self Administered Questionnaire, and
- Internet

 5.13

Sampling...

Recall that statistical inference permits us to draw conclusions about a population based on a sample.

Sampling (i.e. selecting a sub-set of a whole population) is often done for reasons of ***cost*** (it's less expensive to sample 1,000 television viewers than 100 million TV viewers) and ***practicality*** (e.g. performing a crash test on every automobile produced is impractical).

In any case, the ***sampled population*** and the ***target population*** should be **similar** to one another.

Types of sampling

- Non-probability samples
- Probability samples

 5.15

Selecting Non-Random Samples

Known as non-probability sampling

Use of methods that do not have random sampling at any stage

Useful when the population cannot be described

Three techniques
- Convenience
- Purposive
- Quota

Obj. 5.1

- Convenience samples (ease of access)
  sample is selected from elements of a population that are easily accessible
- Snowball sampling (friend of friend….etc.)
- Purposive sampling (judgemental)
  You chose who you think should be in the study
  Cheaper- but unable to generalise
  potential for bias

 5.17

Selecting Non-Random Samples

Convenience sampling

Selection based on the availability of subjects
- Volunteers
- Pre-existing groups

Concerns related to representation and generalizability

Obj. 5.2 & 5.3

Selecting Non-Random Samples

Purposive sampling

Selection based on the researcher's experience and knowledge of the individuals being sampled

Usually selected for some specific reason

Knowledge and use of a particular instructional strategy

Experience

Being in a specific setting such as a school changing to a teacher-based decision-making process

Need for clear criteria for describing and defending the sample

Concerns related to representation and generalizability

Obj. 5.2 & 5.4

Selecting Non-Random Samples

Quota sampling

Selection based on the exact characteristics and quotas of subjects in the sample when it is impossible to list all members of the population

Concerns with accessibility, representation, and generalizability

Obj. 5.2 & 5.5

Probability samples

- Random sampling
  Each subject has a known probability of being selected
- Allows application of statistical sampling theory to results to:
  - Generalise
  - Test hypotheses
- Probability samples are the best
  Ensure Representativeness and Precision

 5.21

Sampling Plans...

A ***sampling plan*** is just a method or procedure for specifying how a sample will be taken from a population.

We will focus our attention on these three methods:

•Simple Random Sampling,
•Stratified Random Sampling, and
•Cluster Sampling.

•Random sampling,by far, is the most common one used.

 5.22

Simple Random Sampling...

A ***simple random sample*** is a sample selected in such a way that every possible sample of the same size is equally likely to be chosen.

Drawing three names from a hat containing all the names of the students in the class is an example of a simple random sample: any group of three names is as equally likely as picking any other group of three names.

VERY EASY TO DEFINE!

VERY, VERY DIFFICULT TO DO!

•Random sample of 100 cokes bottles today at the coke plant.
•Random sample of 50 pine trees in a 1000 acre forest.
•Random sample of 5 deer in a national forest.

 5.23

Simple Random Sampling...

A government income tax auditor must choose a sample of 5 of 11 returns to audit…[Can do many different ways]

Person	Generate Random #
baker	0.87487
george	0.89068
ralph	0.11597
mary	0.58635
sally	0.34346
joe	0.24662
andrea	0.47609
mark	0.08350
greg	0.53542
aaron	0.37239
kim	0.73809

	Person	Sorted Random #
1	mark	0.08350
2	ralph	0.11597
3	joe	0.24662
4	sally	0.34346
5	aaron	0.37239
	andrea	0.47609
	greg	0.53542
	mary	0.58635
	kim	0.73809
	baker	0.87487
	george	0.89068

 5.24

Stratified Random Sampling...

A ***stratified random sample*** is obtained by separating the population into mutually exclusive sets, or strata, and then drawing simple random samples from each stratum.

Strata 1 : Gender	Strata 2 : Age	Strata 3 : Occupation
Male	< 20	professional
Female	20-30	clerical
	31-40	blue collar
	41-50	other
	51-60	
	> 60	

We can acquire about the total population, make inferences within a stratum or make comparisons across strata

 5.25

Stratified Random Sampling...

After the population has been stratified, we can use ***simple random sampling*** to generate the complete sample:

Income Category	Population Proportion	Sample Size n = 400	Sample Size n = 1000
under $25,000	25%	100	250
$25,000 - $39,999	40%	160	400
$40,000 – $60,000	30%	120	300
over $60,000	5%	20	50

If we only have sufficient resources to sample 400 people total, we would draw 100 of them from the low income group...

...if we are sampling 1000 people, we'd draw 50 of them from the high income group.

 5.26

Cluster Sampling...

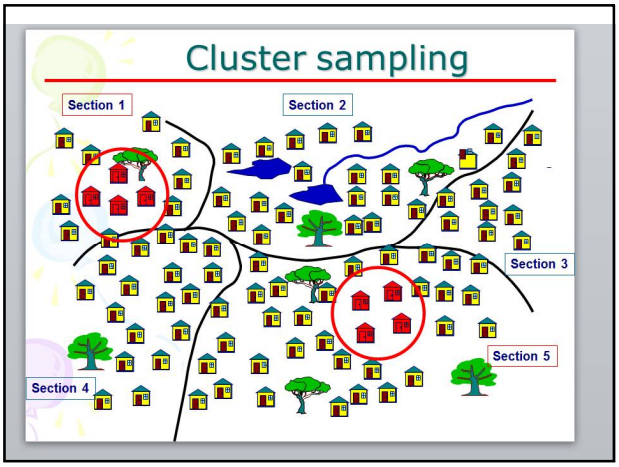
Cluster: a group of sampling units close to each other i.e. crowding together in the same area or neighborhood

A ***cluster sample*** is a simple random sample of groups or clusters of elements (vs. a simple random sample of individual objects).

This method is useful when it is difficult or costly to develop a complete list of the population members or when the population elements are widely dispersed geographically. Used more in the "old days".

Cluster sampling may increase sampling error due to similarities among cluster members.

 5.27



Sample Size...

Numerical techniques for determining sample sizes will be described later, but suffice it to say that the larger the sample size is, the more accurate we can expect the sample estimates to be.

 5.29

Quantitative Sampling

Important issues (continued)

General rules for sample size

As many subjects as possible

Thirty (30) subjects per group for correlational, causal-comparative, and true experimental designs

Ten (10) to twenty (20) percent of the population for descriptive designs

Obj. 1.8

Quantitative Sampling

Important issues (continued)

General rules for sample size (continued)

The larger the population size, the smaller the percentage of the population needed to get a representative sample

For population of less than 100, use the entire population

If the population is about 500, sample 50%

If the population is about 1,500, sample 20%

If the population is larger than 5,000, sample 400

Obj. 1.9

Sampling Size

For descriptive statistics, to have 95% confidence level in estimating population parameters using a sample, can use:

1. Krejcie and Morgan (1970) Table. (Pg 295, Sekaran and Bougie).

2. Bartlett's Table

Bartlett, J.E., Kotrlik, J.W., Higgins, C.C. (2001). Determining appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal*, 19 (1), pp. 43-50.

 5.32

http://api.ning.com/files/dDaMclZ3KnGUTT6nb2fPTHLjun-LLZEfrgdcoswcvTsB60CmiaZ93cmYBLFd1wUyFBUK4H9eT767qY8mUR7PWj88cc1Xw6h/SampleSizeDetermination.pdf

 5.33

Sample Size for Inferential Statistical Analysis

Statistical power is the probability of not missing an effect, due to sampling error, when there really is an effect to be found.

Power is the probability (prob = 1 - β) of correctly rejecting H_o when it really is false.

 5.34

Conventions And Decisions About Statistical Power

Acceptable risk of a Type II error is often set at 1 in 5, i.e., a probability of 0.2.

The conventionally uncontroversial value for "adequate" statistical power is therefore set at 1 - 0.2 = 0.8.

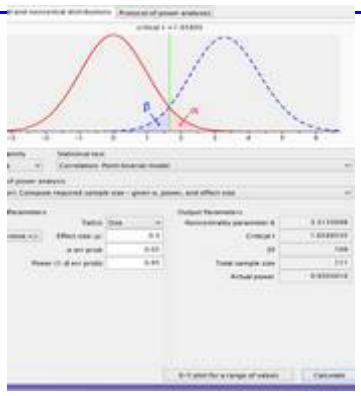
People often regard the minimum acceptable statistical power for a proposed study as being an 80% chance of an effect that really exists showing up as a significant finding.

 5.35

Sample Size for Inferential Statistical Analysis can be determined using a software, GPower.

http://www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/download-and-register

 5.36



 5.37

Sampling and Non-Sampling Errors...

Two major types of error can arise when a sample of observations is taken from a population:

sampling error and ***nonsampling error***.

Sampling error refers to differences between the sample and the population that exist only because of the observations that happened to be selected for the sample. Random and we have no control over.

Nonsampling errors are more serious and are due to mistakes made in the acquisition of data or due to the sample observations being selected improperly. Most likely caused be poor planning, sloppy work

 5.38

Sampling Error...

Sampling error refers to differences between the sample and the population that exist only because of the observations that happened to be selected for the sample.

Increasing the sample size **will** reduce this type of error.

 5.39

Nonsampling Error...

Nonsampling errors are more serious and are due to mistakes made in the acquisition of data or due to the sample observations being selected improperly. Three types of nonsampling errors:

Errors in data acquisition,
Nonresponse errors, and
Selection bias.

Note: increasing the sample size **will not** reduce this type of error.

 5.40

Errors in data acquisition...

…arises from the recording of incorrect responses, due to:

— incorrect measurements being taken because of faulty equipment,
— mistakes made during transcription from primary sources,
— inaccurate recording of data due to misinterpretation of terms, or
— inaccurate responses to questions concerning sensitive issues.

 5.41

Nonresponse Error...

…refers to error (or ***bias***) introduced when responses are not obtained from some members of the sample, i.e. the sample observations that are collected may not be representative of the target population.

As mentioned earlier, the ***Response Rate*** (i.e. the proportion of all people selected who complete the survey) is a key survey parameter and helps in the understanding in the validity of the survey and sources of nonresponse error.

 5.42

Selection Bias…

…occurs when the sampling plan is such that some members of the target population cannot possibly be selected for inclusion in the sample.

 5.43

Qualitative Sampling

Unique characteristics of qualitative research
- In-depth inquiry
- Immersion in the setting
- Importance of context
- Appreciation of participant's perspectives
- Description of a single setting

The need for alternative sampling strategies

Obj. 7.2

Qualitative Sampling

Purposive techniques – relying on the experience and insight of the researcher to select participants
- Intensity – compare differences of two or more levels of the topics
  - Students with extremely positive and extremely negative attitudes
  - Effective and ineffective teachers

Obj. 7.3

Qualitative Sampling

Purposive techniques (continued)
- Homogeneous – small groups of participants who fit a narrow homogeneous topic
- Criterion – all participants who meet a defined criteria
- Snowball – initial participants lead to other participants

Obj. 7.4, 7.5, & 7.6

Qualitative Sampling

Purposive techniques (continued)
- Random purposive – given a pool of participants, random selection of a small sample

Combinations of techniques

Inherent concerns related to generalizability and representation

Obj. 7.7 & 7.8

Qualitative Sampling

Sample size

Generally very small samples given the nature of the data, collection methods and the data itself

Two general guidelines
- Redundancy of the information collected from participants
- (Maturation point)

Representation of the range of potential participants in the setting

Obj. 7.9

Generalizability

Probability sampling

- Begins with a population and selects a sample from it.
- Generalizability to the population is relatively easy.

Non-probability and purposive sampling

- Begins with a sample that is NOT selected from some larger population
- Must consider the population hypothetical as it is based on the characteristics of the sample
- Generalizability is often very limited

Obj. 7.10