Edith Cowan University
Research Online

ECU Research Week | Conferences, Symposia and Campus Events

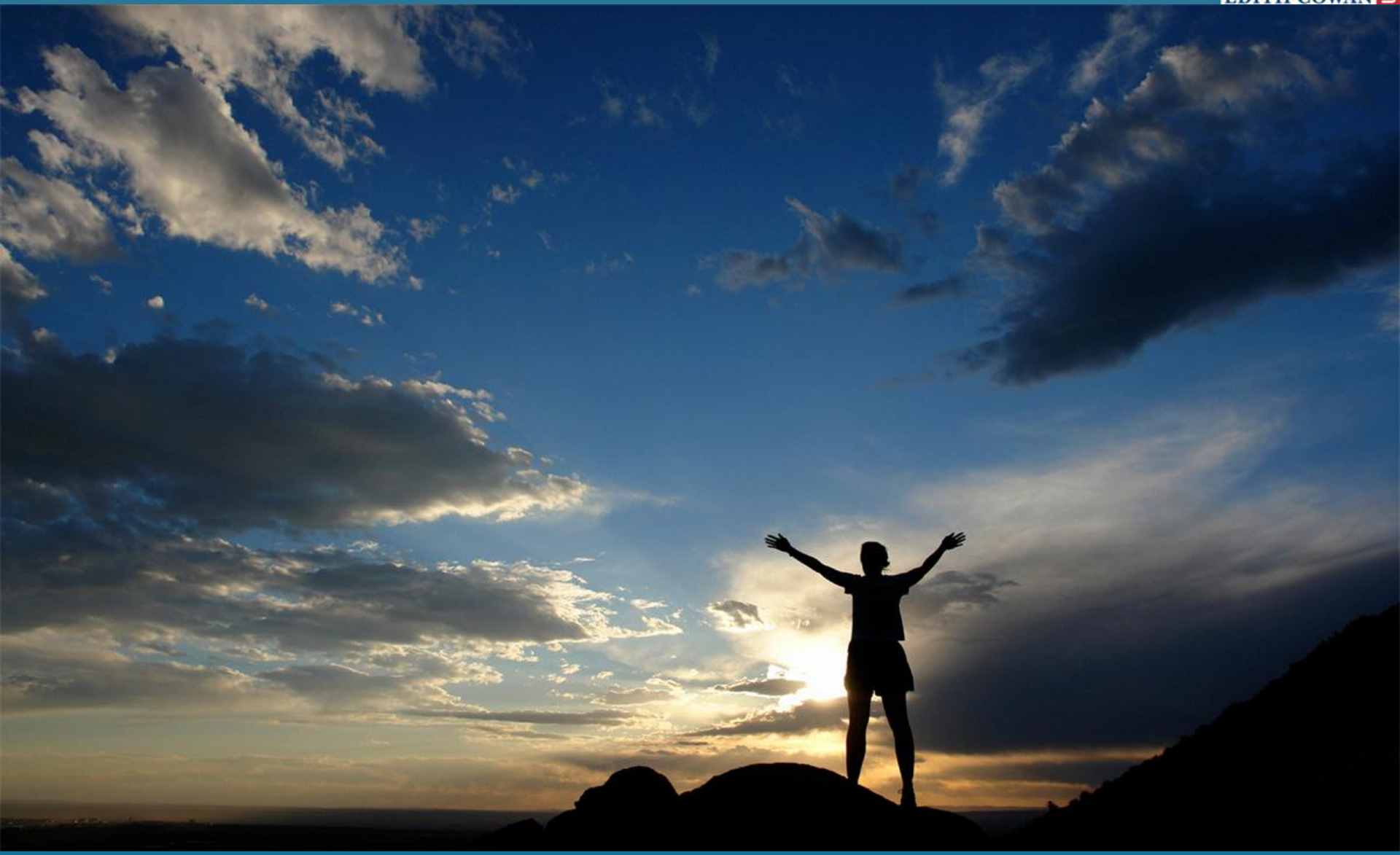
2010

Troubleshooting with Surveys

Sandra Pereira
Edith Cowan University

Presented at the ECU research week 2010, 16th to 20th August 2010
This Presentation is posted at Research Online.
http://ro.ecu.edu.au/creswk/3

EDITH COWAN UNIVERSITY
Graduate Research School



www.ecu.edu.au/research/week

ECU RESEARCH WEEK
16-20 August 2010

Troubleshooting with Surveys

Dr Sandra Pereira

Research Consultant

Research and Higher Degrees Office

Faculty of Business and Law

s.pereira@ecu.edu.au

17 August 2010

Notes on "Troubleshooting Surveys" Seminar

Dr S Pereira
Research Consultant
Research and Higher Degrees Office
Faculty of Business and Law
Edith Cowan University

Note that the copyright protected notes are freely available for educational use (only via through the link below). It is customary to cite it by your date of use:

Pereira, S.M.C. Retrieved dd/mm/yyyy from

http://www.ecu.edu.au/research/week/events/research-and-training/2010/seminar-troubleshooting-with-survey

Aims

1. To highlight common problems with survey research.
2. To provide practical advice on how you could avoid these problems.
3. To give you the opportunity to raise any concerns that you have with the planning of your survey research.

Summary

1. Introduction
2. Main steps in survey research
3. Classifying survey errors
4. Minimising errors in survey research
5. Designing your questionnaire

Appendix A: Random Sample Survey (*)

(*) Your homework

1. Introduction

A survey can be anything from a short paper-and -pencil feedback form to an intensive one-on-one, in-depth interview.

A survey encompasses any measurement procedure that involves asking questions of respondents.

Survey research is an important area of measurement in an applied field.

1. Introduction (cont.)

Surveys can be classified broadly into two types: the *interview* and the *questionnaire*.

- *Interviews* are completed by an interviewer based on what the respondent says. (For more information on interviews in qualitative research see slide 24.)
- *Questionnaires* are usually paper-and-pencil instruments that the respondent completes, but on-line surveys are becoming more common.

This seminar will address the questionnaire design in Section 5.

2. Main steps in survey research

1. State research objectives
2. Propose research questions
3. Identify the "target" population
4. "Construct" the list frame

2. Main steps in survey research (cont.)

5. Identify the *sample* (see Appendix A)
6. Select the measurement *method* and *variables*
7. Develop measurement instrument or *questionnaire*
8. Validate of the questionnaire (validity and reliability analyses)

2. Main steps in survey research (cont.)

9. Select and train field-workers ("you")
10. Pre-test the survey (one phase or two phase pilot study)
11. Conduct the test (main study)
12. Organise the fieldwork (data collection and storage)

2. Main steps in survey research (cont.)

13. Organise data management (files, forms, records, software, etc.)
14. Analyse Quantitative and/or Qualitative data
15. Present the findings and results (a chapter or section of a thesis, seminar, article, etc.)

3. Classifying survey errors

(a) Errors of non-observations e.g., when there is a failure to observe the data. That is, the sampled elements make up only part of the target population.

(b) Errors of observations e.g., when the recorded data deviate from reality.

4. Minimising errors in survey research

1. Improve low response rates by using **callbacks**
2. Motivate the participants to fill in the questionnaires by using **rewards and incentives**
3. Provide **good** training skills to the interviewers

4. Minimising errors in survey research (cont.)

4. Carefully **check** the data (twice, three times, four times, and random checks, etc.)
5. Identify **outliers** (discrepant observations) and main reasons for having the outliers

4. Minimising errors in survey research (cont.)

7. Plot your quantitative dataset so that you can a visual understanding of it. (E.g. a scatter plot, bar plot or histogram.)

8. Conduct two separate analyses (one analysis with outliers and another analysis without outliers) and report both results.

5. Design your questionnaire

To develop a new questionnaire, the following topics may be relevant:

- Order the questions (simple to complex)
- Choose the type of responses: open versus closed questions
- How to frame the questions (cultural differences, sensitive issues, etc.)

5. Design your questionnaire (cont.)

Practical suggestions

1. Use plain and straightforward language that your target population will understand

2. Search in your research field for commonly used **validity and reliability techniques**

Practical suggestions (cont.)

3. Invite people that are not familiar with your research field to fill in the pre-test questionnaire and consider their feedback about the clarity of the questions

4. Record how many minutes invited people spend filling in the questionnaires (aim for less than 15 minutes)

Practical suggestions (cont.)

5. Each question must address **one** item only (that is, no "AND" or "OR")

6. Do not ask **too** many questions

7. The questions must be relevant to your **main** research questions

Practical suggestions (cont.)

8. If you ask the participant to select from a range of answers then each should be **easily** read, e.g., as a sentence

"… I chose this product because:
(a) it was good value for money (√)
(b) poor availability of other brands (X)
(c) of the poor availability of other brands (√)"

Practical suggestions (cont.)

9. Group related questions together

10. Do not use a crowded layout

11. Acknowledge published source(s) of questions

Practical suggestions (cont.)

12. Do not confuse the participants with sudden reversal of scales, e.g.,

Strongly disagree → *Strongly agree*
(1) *(7)*

then

Strongly agree → *Strongly disagree.*
(1) *(7)*

Examples of questionnaires

- *A problematic questionnaire*, revised by Dr S Pereira and Dr G Maguire, June 2010 (see attached pdf file).

- *Session/Activity Evaluation Sheet, Introduction to Methodology Research and Random Sample Survey Design Workshop*, revised by Dr S Pereira and Dr Greg Maguire, April 2010, (see attached pdf file)

Resources available for you

- Producing a thesis with EndNote (ECU Library)

  http://www.ecu.edu.au/library/endnote/endnote_bibliog.htm

  Getting Started with EndNoteX3 at ECU

  http://www.ecu.edu.au/library/endnote/EndNoteX3%20Demonstration_8Feb10.pdf

  Getting Started with EndNote vX1 at ECU

  http://www.ecu.edu.au/library/endnote/endnote_notes.pdf

- *Interviews in Qualitative Research*

  Cd from Dr Susan Hill (s.hill@ecu.edu.au) or me

Resources available for you (cont.)

- ECU Survey Research Centre (fees)
  http://surveyresearchcentre.ecu.edu.au/

- Dr G Maguire (Research Writing consultant) at g.maguire@ecu.edu.au is also available to review the wording of questionnaires for FB&L and FCHS research students (and staff in some schools)

Faculty of Business and Law: Research and Higher Degree's Office

- Prof Beth Walker (Associate dean)
  b.walker@ecu.edu.au
- Ms Bev Lurie (Acting Research coordinator)
  b.lurie@ecu.edu.au
- Ms Donna Johnston (Administration officer)
  d.johnston@ecu.edu.au
- Dr Greg Maguire (Research Writing consultant)
  g.maguire@ecu.edu.au
- Dr Sandra Pereira (Research consultant)
  s.pereira@ecu.edu.au

References

- APA (2010). Publication manual of the American Psychological Association, 6th edition, Washington, DC.
- Scheaffer, R. L., Mendenhall III, W. and, Ott, R.L. (2006). Elementary survey sampling, Duxbury Advanced Series, 6th edition, Thomson, Brooks/ Cole.

Web references

- National Statistical Service

http://www.nss.gov.au/nss/home.NSF/

- Basic Survey Design (Chapters 1 to 13)

http://www.nss.gov.au/nss/home.nsf/2c4c8bd01df32224ca25713400ea79a/e4e968fc83fb1217ca2571ab00247b3e?OpenDocument

Acknowledgements

Thank you very much to the GRS for their kind invitation!

Thank you so much for your time and participation!

Appendix A: A random sample survey design (*)

(*) Your homework

Appendix A: A random sample survey design

Summary

A.1 Terminology

A.2 How to select a random sample

A.3 Probability sampling techniques

A.4 Non probabilistic sampling

A.1 Terminology

- A sample survey makes inferences about a population from information contained in a selected sample
- The sample estimates population parameters. For example, we estimate the population mean or average (μ), population standard deviation (σ), variance, median, proportion, total, etc.

A.1 Terminology (cont.)

- The sample estimates population parameters. For example, we estimate the:
  population mean or average (μ) using the sample mean or average ($\overline{x}$);
  population variance (σ^2) using the sample variance (s^2);

A.1 Terminology (cont.)

population standard deviation (σ), using the sample standard deviation (s);

population proportion (Π), using the sample proportion (P);

and population total (τ), using the sample total (T), etc.

A.1 Terminology (cont.)

- An element is an object or item on which a measurement is taken e.g. in research surveys, elements are people or certain types of people

- A population is a collection of elements about which we want to make an inference e.g. all adult population of Australia aged 18 years old or older.

A.1 Terminology (cont.)

- A sampling unit is a non-overlapping collection of elements from the population that covers the entire population.

  That is, an element or set of elements considered for selection in some stage of sampling (same as the elements, in a simple single-stage sample). In a multi-stage sample, the sampling unit could be blocks, households, and individuals within the households.

A.1 Terminology (cont.)

- A frame is a list of sampling units e.g. a list of the study population.
- A sample is a collection of sampling units drawn from a frame or frames e.g., a list of the names of the participants that are (actually) attending the workshop today.

A.2 How to select a random sample

The selected procedure and number of sampling units (observations) depend on how much information we want to know.

For instance, we want to be 95% certain that the true population parameter (e.g. mean) lies in the range of the sample mean ± 2 x standard deviation. (See Figure 2.1).

A.2 How to select a random sample (cont.)

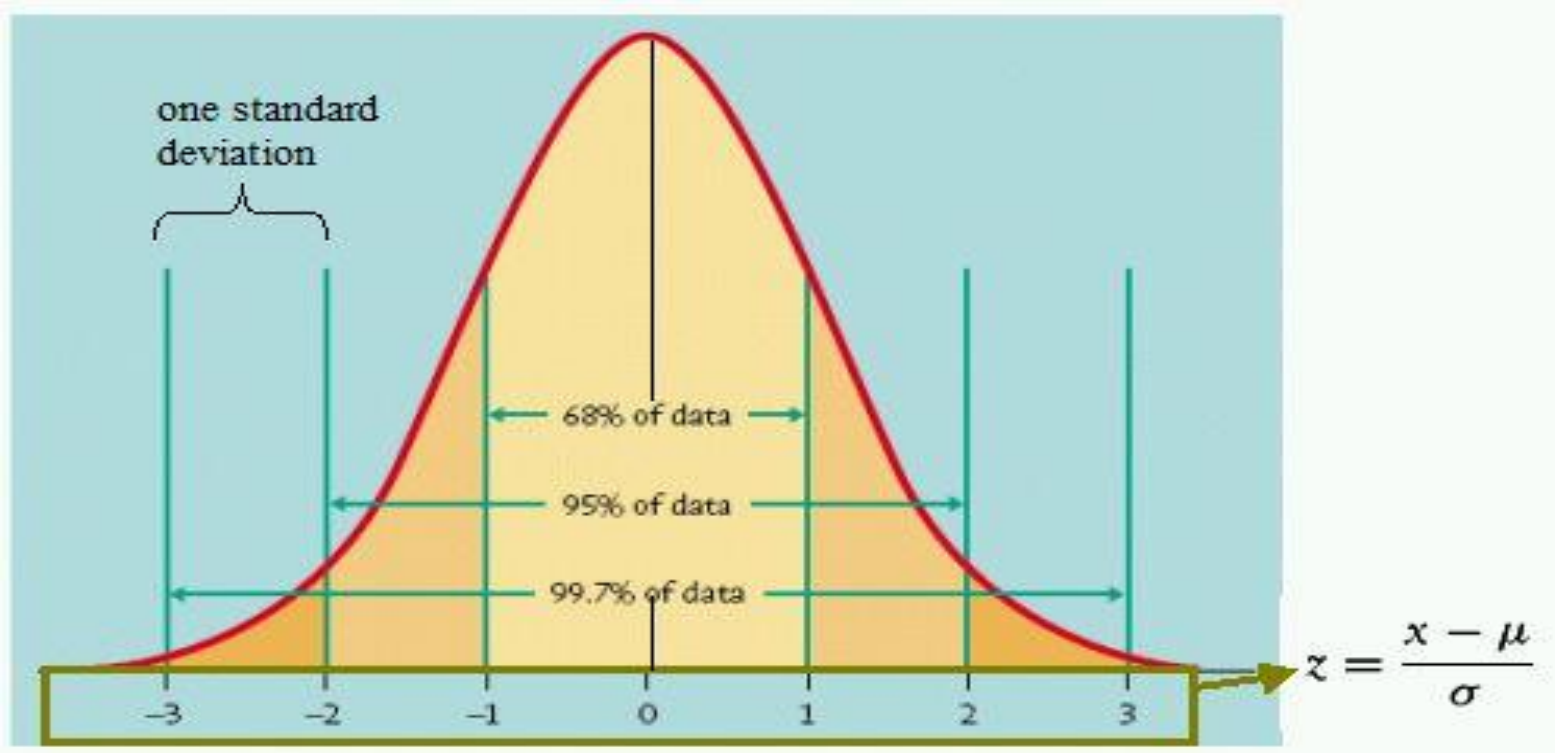


Figure 2.1 : The standard normal distribution.

Source: http://academic.reed.edu/math/hancock/images/stdnormaldist.jpg

A.3 Probability sampling techniques

Randomness is built into the sampling design so that the properties of the estimators (sample mean, sample variance, sample standard deviation, sample proportion, sample total, etc.) can be assessed a on probability basis.

Sample designs based on planned randomness are named probability samples.

Examples of techniques for collecting probability samples:

- Simple Random Sample (SRS)
- Stratified Random Sample (StrRS)
- Systematic Random Sample (SyRS)
- Cluster Sample (CS)

A.3.1 Simple random sample (SRS)

A SRS consists of selecting a group of n sampling units such that each sample of size n has an equal chance of being selected.

Practical Example A.1

There are N=1000 client records from which a SRS of n=20 is to be drawn. Using a **Table of Random Numbers** (See Table A.1), the randomly selected file numbers are:

049	711	196	856
223	995	635	023
241	963	094	010
421	895	103	521
375	854	005	024

Practical Example A.1 (cont.)

Part of a Table of Random Numbers

61424	20419	86546	00517
90222	27993	04952	66762
50349	71146	97668	86523
85676	10005	08216	25906
02429	19761	15370	43882
90519	61988	40164	15815
20631	88967	19660	89624
89990	78733	16447	27932

Table A.1 "An Example of a Table of Random Numbers"

Source: http://www.gifted.uconn.edu/siegle/research/Samples/RANTBLE.JPG

Practical Example A.1 (cont.)

So that the client records with these numbers are the SRS of size n=20.

That is, the client record numbers: 049, 711,… , 005, and 024 (see previous page) are selected and analysed.

Using a statistical software (SPSS, SAS, SPLUS, R, etc.), we can compute *sample mean, sample standard deviation, sample variance, sample median,* etc. for the random sample.

A.3.1 Simple random sample (SRS) (cont.)

How can we calculate the appropriate sample size (*n*) for the **population mean (average)** based on a survey and using the SRS technique?

The computation of the sample size *n*, for estimating **population means** (averages) with a bound on the error of estimation B, is given by the Equation A.1. This takes into account the size and variability of the population.

A.3.1 Simple random sample (SRS) (cont.)

Equation A.1:

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2}$$

where N is the population size, $D = \frac{B^2}{4}$ and σ^2 is the population variance.

Practical Example A.2

For instance, the **population mean** for the amount of money for a business's accounts receivable must be estimated.

Although no prior data are available to estimate the **population variance**, it is known that most accounts lie within a $100 **range**. There are N = 1000 open accounts.

Compute the **sample size** needed to estimate the population mean with a bound on the error of estimation B = $3.

Practical Example A.2 (solution)

We know that N = 1000, B = 3, range = 100. Note that a range ~= 4 x standard deviation = 4 x sd therefore sd = range/4. That is, sd = 100/4 = 25.

We can now compute the population variance, denoted by σ^2. The estimated value of the population variance is

σ^2 = (sd x sd) = (25 x 25) = 625.

We now need to calculate D, given by $D = \frac{B^2}{4}$

Practical Example A.2 (solution)

D= (3x3)/4= 2.25 . Substituting the values into Equation A.1, we are finally able to compute the sample size *n*

$$n = \frac{N\sigma^2}{(N-1)D+\sigma^2} = \frac{1000\text{x}625}{999\text{x}2.25+625} = 217.56 = 218$$

Thus, we need approximately *n* = 218 random elements to estimate the **population mean** (the mean account receivable) with a *bound on the error of estimation* of $3.

Practical Example A.2 (software solution)

1. Click at the sample size calculator available at Employee Survey Solutions at Custominsight.com website http://www.custominsight.com/articles/random-sample-calculator.asp
2. Type 5.87 in at the *tolerated error*
3. Type 1000 in at the *population size*,
4. Click at Calculate and check your results at *95% confidence level.*

Practical Example A.2 (software solution)

Note that at *95% confidence level*, the sample size is approximately $n = 218$ random elements with a *bound on the error of estimation* of \$ (5.87)/2 = 2.935.

- Could you please explain why have we got different errors?

  That is, website calculator: 2.935

  my calculation: 3

A.3.2 Stratified random sample (StRS)

A StRS is obtained by separating the population elements into non-overlapping groups, called strata, and then selecting a simple random sample from each stratum.

A.3.2 Stratified random sample (StRS) (cont.)

How can we compute the sample size for the **population mean** based on a survey and using the StRS technique?

The calculation of the sample size *n* for estimating population means (averages) with a bound on the error of estimation B is given by the Equation A.2.

Equation A.2:

$$n = \frac{\sum_{i=1}^{L} N_i^2 \frac{\sigma_i^2}{a_i}}{N^2 D + \sum_{i=1}^{L} N_i \sigma_i^2}$$

where a_i is the fraction of observations allocated to stratum i,

σ_i^2 is the population variance for stratum i,

L is the total number of strata, and

$D = \frac{B^2}{4}$.

Practical Example A.3

An advertising firm, interested in determining how much television advertising is needed in a small village, decides to use a stratified sample survey to estimate the **mean number of hours** each week that households within the village watch television.

The village contains two main regions, A and B, and a rural area. (Note that the number of strata is L=3). There are 155 households in region A, 62 households in region B , and 93 households in the rural area.

Practical Example A.3 (cont.)

A prior survey suggests that the stratum variances are approximately $\sigma_1^2 \approx 25$, $\sigma_2^2 \approx 225$, and $\sigma_3^2 \approx 100$.

We choose the sample size to obtain a bound of error of estimation equal to 2 hours if the allocation fractions are given by $a_1 = \frac{1}{3}$, $a_2 = \frac{1}{3}$, and $a_3 = \frac{1}{3}$. In other words, take an equal number of observations from each stratum.

Practical Example A.3 (cont.)

Note that **1/3** is just an easy approximation if the regions differed greatly in population we would calculate the actual fraction of observations allocated to each stratum.

Scheaffer et al. (2006), pages 130 and 132, suggest few rules for sampling allocation.

Our task is to estimate the sample size *n* for the **population mean** using the StRS technique.

Practical Example A.3 (solution)

We know that a bound on the error of 2 hours means

$$2\sqrt{\mathrm{Var}(\bar{y}_{st})} = 2 \quad \text{or} \quad \mathrm{Var}(\bar{y}_{st}) = 1$$

It is known that $2\sqrt{\mathrm{Var}(\bar{y}_{st})} = B$ or $\mathrm{Var}(\bar{y}_{st}) = \dfrac{B^2}{4}$ or $B^2 = 4$

Thus $D = \dfrac{B^2}{4} = \dfrac{4}{4} = 1$ or $D = 1$

We also know that $N_1 = 155$, $N_2 = 62$, and $N_3 = 93$.

Practical Example A.3 (solution) (cont.)

The computed sums are

$$\sum_{i=1}^{3}\frac{N_i\sigma_i^2}{a_i}=\frac{N_1^2\sigma_1^2}{a_1}+\frac{N_2^2\sigma_2^2}{a_2}+\frac{N_3^2\sigma_3^2}{a_3}=\frac{(155)^2(25)}{(1/3)}+\frac{(62)^2(225)}{(1/3)}+\frac{(93)^2(100)}{(1/3)}$$

$$=(24025)(75)+(3844)(675)+(8649)(300)=6991275$$

$$\sum_{i=1}^{3}N_i\sigma_i^{\,2}=N_1\sigma_1^2+N_2\sigma_2^2+N_3\sigma_3^2=(155)(25)+(62)(225)+(93)(100)=271125$$

$$N^2D=(N_1+N_2+N_3)D=(155+62+93)1=(310)^2 1=96100$$

Practical Example A.3 (solution) (cont.)

From Equation A.2, we then have the following sample sizes:

$$n = \frac{\sum_{i=1}^{L} N_i^2 \frac{\sigma_i^2}{a_i}}{N^2 D + \sum_{i=1}^{L} N_i \sigma_i^2} = \frac{6991275}{96100 + 27125} = \frac{6991275}{123225} = 56.7 = 57$$

$n_1 = n(a_1) = 57(1/3) = 19$ 19 elements for region A,

$n_2 = n(a_2) = 57(1/3) = 19$ 19 elements for region B, and

$n_3 = n(a_3) = 57(1/3) = 19$ 19 elements for rural area.

A.3.3 Systematic random sample (SyRS)

A SyRS is obtained by randomly selecting one element from the first k elements in the frame and every kth element thereafter.

This technique is known as a 1-in-k systematic sample with a *random* start.

E.g. in a 1-in-5 systematic sample, choose the first element randomly, let's say, the third element then Choose the eighth element (3+5 th), the thirteenth (8+5 th), the eigtheenth element (13+5 th), etc.

A.3.4 Cluster random sample (CRS)

A CRS is a probability sample in which each sampling unit is a collection of elements, or a a cluster of elements.

This design is more complex and it is not Shown here. However, you are **most welcome** to **consult me** for sampling size calculations.

A.4 Non-probabilistic sampling (NPS)

A non-probabilistic sampling is when choice of the sampling units is not based on mathematical criteria. For instance:

- Volunteer sampling
- Judgment sampling
- Judgmental sampling or purposive sampling

A.4 Non-probabilistic sampling (NPS)

- Quota sampling
- Purposive sampling
- Snowball or chain sampling, etc.

Details on these types of sampling techniques are available at

http://www.statcan.gc.ca/edu/power-pouvoir/ch13/nonprob/5214898-eng.htm

http://www.experiment-resources.com/non-probability-sampling.html

http://www.socialresearchmethods.net/kb/sampnon.php

Examples of sample size calculators that are available on the internet (*)

- NSS sample size calculator for proportion at http://www.nss.gov.au/nss/home.nsf/NSS/0A4A642C712719DCCA2571AB00243DC6?opendocument
- TMS random sample calculator at http://www.custominsight.com/articles/random-sample-calculator.asp
- UCSF DB power and sample size programs at http://www.epibiostat.ucsf.edu/biostat/sampsize.html
- DBVU power and sample size calculation at http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize

(*) Bear in mind that sample size calculations are **most reliable** when you rely on advice provided by research consultants, statisticians and biostatisticians.

Please be **careful** when using other sample size calculators available on the internet. (So far the ones that are stated in this presentation have given similar results to mine.)

Other calculators could have been programmed by non-experts so that the resulting sample size could underestimate stated confidence levels or population parameters (means, variances, standard deviations, proportions, totals, etc).

Acknowledgement

Thank you very much for your time and participation!