ST3004: Research Methods

Sampling
Why Sampling?...

- When sampling is used, it is possible to generate findings that are **representative** of the whole population rather than collecting the data for the whole population.

- **Cheaper** to collect some data rather than lots.

- **Easier** to make sure that sample is random (if trying a census, non response causes bias- you can at least chase 20 non-responding sample individuals rather than 200 census).

- Often just not possible to do census.
However...

- Have to think carefully about how to choose, and methods of analysis

- Not 100% certain (but ‘certainty’ is just comfort - in practice data errors mean a census is not 100% certain)
Sampling ...

- Many methods exist for picking a sample
- If there is possible bias, it has to be described and accounted for in your methodology section
Population, Sample and Cases

Population

Sample

Case or element
Sampling Techniques

- *Probability* or representative sampling;
- *Non-probability* or judgemental sampling.
Probability Sampling

- Any method of sampling that utilizes some form of random selection.

- The chance, or probability, of each case being selected from the population is known and is usually equal for all cases.

- This means that it is possible to answer research questions and to achieve objectives that require you to estimate statistically the characteristics of the population from the sample.

- Consequently, probability sampling is often associated with survey and experimental research strategies.
Non-Probability Sampling

- Non-random so the probability of each case being selected from the total population is not known.

- It is impossible to answer research questions or to address objectives that require you to make statistical inferences about the characteristics of the population.

- You may still be able to generalise from non-probability samples about the population, but not on statistical grounds.
What kind of Sample?

Sample

Do you need to generalise?

Yes

Probability Sampling

No

Non –Probability Sampling
Deciding Factors

- Do you want the sample to be representative of a larger population?

- In your analysis do you want to use probability-based statistical methods in analysis?

- Does a convenient sampling frame exist?

- What size do you want your sample to be?

- Are you able to afford the time and money to carry out the sample collection?
Steps in Probability Sampling

1. Determine the sampling frame
2. Determine the sampling design
3. Determine the appropriate sample size
4. Execute the sampling process
Sampling Frame

The sampling frame for any probability sample is a complete list of all the cases in the population from which your sample will be drawn.

Examples of existing sampling frames:
- Voting register
- Membership List
- Companies Database
Sampling Frame

- Created/Distilled from lists
- New company start-ups (County Enterprise Boards)

Potential Problems with Sampling Frames:
- Out of date/incomplete (voting register)
- Unit may be inappropriate (household not individuals)
- May contain ineligible elements (teetotallers)
Within probability sampling, by defining the sampling frame you are defining the population about which you want to generalise.

This means that if your sampling frame is a list of all customers of an organisation, strictly speaking you can only generalise, that is apply the findings based upon your sample, to that population.
Sample Size

- Usually assumed that ‘more is better’ (large samples reflect more accurately the characteristics of the population than small samples)

- The standard error of the mean decreases as the samples size $n$ increases but this may depend on the variability of the population sampled

- Even a large sample using a poor sampling design could give you less information than a smaller, more carefully designed sample
## Required Sample Size

<table>
<thead>
<tr>
<th>Population Size</th>
<th>Confidence = 95%</th>
<th></th>
<th></th>
<th></th>
<th>Confidence = 99%</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.0%</td>
<td>3.5%</td>
<td>2.5%</td>
<td>1.0%</td>
<td>5.0%</td>
<td>3.5%</td>
<td>2.5%</td>
<td>1.0%</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>20</td>
<td>19</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>19</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>30</td>
<td>28</td>
<td>29</td>
<td>29</td>
<td>30</td>
<td>29</td>
<td>29</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>50</td>
<td>44</td>
<td>47</td>
<td>48</td>
<td>50</td>
<td>47</td>
<td>48</td>
<td>49</td>
<td>50</td>
</tr>
<tr>
<td>75</td>
<td>63</td>
<td>69</td>
<td>72</td>
<td>74</td>
<td>67</td>
<td>71</td>
<td>73</td>
<td>75</td>
</tr>
<tr>
<td>100</td>
<td>80</td>
<td>89</td>
<td>94</td>
<td>99</td>
<td>87</td>
<td>93</td>
<td>96</td>
<td>99</td>
</tr>
<tr>
<td>150</td>
<td>108</td>
<td>126</td>
<td>137</td>
<td>148</td>
<td>122</td>
<td>135</td>
<td>142</td>
<td>149</td>
</tr>
<tr>
<td>200</td>
<td>132</td>
<td>160</td>
<td>177</td>
<td>196</td>
<td>154</td>
<td>174</td>
<td>186</td>
<td>198</td>
</tr>
<tr>
<td>250</td>
<td>152</td>
<td>190</td>
<td>215</td>
<td>244</td>
<td>182</td>
<td>211</td>
<td>229</td>
<td>246</td>
</tr>
<tr>
<td>300</td>
<td>169</td>
<td>217</td>
<td>251</td>
<td>291</td>
<td>207</td>
<td>246</td>
<td>270</td>
<td>295</td>
</tr>
<tr>
<td>400</td>
<td>196</td>
<td>265</td>
<td>318</td>
<td>384</td>
<td>250</td>
<td>309</td>
<td>348</td>
<td>391</td>
</tr>
<tr>
<td>500</td>
<td>217</td>
<td>306</td>
<td>377</td>
<td>475</td>
<td>285</td>
<td>365</td>
<td>421</td>
<td>485</td>
</tr>
<tr>
<td>600</td>
<td>234</td>
<td>340</td>
<td>432</td>
<td>565</td>
<td>315</td>
<td>416</td>
<td>490</td>
<td>579</td>
</tr>
<tr>
<td>700</td>
<td>248</td>
<td>370</td>
<td>481</td>
<td>653</td>
<td>341</td>
<td>462</td>
<td>554</td>
<td>672</td>
</tr>
<tr>
<td>800</td>
<td>260</td>
<td>396</td>
<td>526</td>
<td>739</td>
<td>363</td>
<td>503</td>
<td>615</td>
<td>763</td>
</tr>
<tr>
<td>1,000</td>
<td>278</td>
<td>440</td>
<td>606</td>
<td>906</td>
<td>399</td>
<td>575</td>
<td>727</td>
<td>943</td>
</tr>
<tr>
<td>1,200</td>
<td>291</td>
<td>474</td>
<td>674</td>
<td>1067</td>
<td>427</td>
<td>636</td>
<td>827</td>
<td>1119</td>
</tr>
<tr>
<td>1,500</td>
<td>306</td>
<td>515</td>
<td>759</td>
<td>1297</td>
<td>460</td>
<td>712</td>
<td>959</td>
<td>1376</td>
</tr>
<tr>
<td>2,000</td>
<td>322</td>
<td>563</td>
<td>869</td>
<td>1655</td>
<td>498</td>
<td>808</td>
<td>1141</td>
<td>1785</td>
</tr>
<tr>
<td>2,500</td>
<td>333</td>
<td>597</td>
<td>952</td>
<td>1984</td>
<td>524</td>
<td>879</td>
<td>1288</td>
<td>2173</td>
</tr>
<tr>
<td>3,500</td>
<td>346</td>
<td>641</td>
<td>1068</td>
<td>2565</td>
<td>558</td>
<td>977</td>
<td>1510</td>
<td>2890</td>
</tr>
<tr>
<td>5,000</td>
<td>357</td>
<td>678</td>
<td>1176</td>
<td>3288</td>
<td>586</td>
<td>1066</td>
<td>1734</td>
<td>3842</td>
</tr>
<tr>
<td>7,500</td>
<td>365</td>
<td>710</td>
<td>1275</td>
<td>4211</td>
<td>610</td>
<td>1147</td>
<td>1960</td>
<td>5165</td>
</tr>
<tr>
<td>10,000</td>
<td>370</td>
<td>727</td>
<td>1332</td>
<td>4899</td>
<td>622</td>
<td>1193</td>
<td>2098</td>
<td>6239</td>
</tr>
<tr>
<td>25,000</td>
<td>378</td>
<td>760</td>
<td>1448</td>
<td>6939</td>
<td>646</td>
<td>1285</td>
<td>2399</td>
<td>9972</td>
</tr>
<tr>
<td>50,000</td>
<td>381</td>
<td>772</td>
<td>1491</td>
<td>6056</td>
<td>655</td>
<td>1318</td>
<td>2520</td>
<td>12455</td>
</tr>
<tr>
<td>75,000</td>
<td>382</td>
<td>776</td>
<td>1506</td>
<td>6514</td>
<td>658</td>
<td>1330</td>
<td>2563</td>
<td>13583</td>
</tr>
<tr>
<td>100,000</td>
<td>383</td>
<td>778</td>
<td>1513</td>
<td>6762</td>
<td>659</td>
<td>1336</td>
<td>2585</td>
<td>14227</td>
</tr>
<tr>
<td>250,000</td>
<td>384</td>
<td>782</td>
<td>1527</td>
<td>9248</td>
<td>662</td>
<td>1347</td>
<td>2626</td>
<td>15555</td>
</tr>
<tr>
<td>500,000</td>
<td>384</td>
<td>783</td>
<td>1532</td>
<td>9423</td>
<td>663</td>
<td>1350</td>
<td>2640</td>
<td>16055</td>
</tr>
<tr>
<td>1,000,000</td>
<td>384</td>
<td>783</td>
<td>1534</td>
<td>9512</td>
<td>663</td>
<td>1352</td>
<td>2647</td>
<td>16317</td>
</tr>
<tr>
<td>2,500,000</td>
<td>384</td>
<td>784</td>
<td>1536</td>
<td>9567</td>
<td>663</td>
<td>1353</td>
<td>2651</td>
<td>16478</td>
</tr>
<tr>
<td>10,000,000</td>
<td>384</td>
<td>784</td>
<td>1538</td>
<td>9594</td>
<td>663</td>
<td>1354</td>
<td>2653</td>
<td>16560</td>
</tr>
<tr>
<td>100,000,000</td>
<td>384</td>
<td>784</td>
<td>1537</td>
<td>9603</td>
<td>663</td>
<td>1354</td>
<td>2654</td>
<td>16584</td>
</tr>
<tr>
<td>300,000,000</td>
<td>384</td>
<td>784</td>
<td>1537</td>
<td>9603</td>
<td>663</td>
<td>1354</td>
<td>2654</td>
<td>16586</td>
</tr>
</tbody>
</table>
Sample Size - minimum

- The larger the absolute size of a sample, the more closely its distribution will be to the normal distribution and thus the more robust it will be.

- A sample size of 30 or more will usually result in a sampling distribution for the mean that is very close to a normal distribution.

- For this reason, a minimum number of 30 for statistical analyses provides a useful rule of thumb.

- Where the population in the category is less than 30 you should normally collect data from all cases in that category.
Probability Sampling Techniques

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

- Based on the use of random numbers
- Each unit of the population is given the same probability of independent selection
- Ideally requires a sample size of over a few hundred
- Only suitable for geographically dispersed cases if you do not require face-to-face contact when collecting your data.
Simple Random Sampling

- In each case:
  - The chances of obtaining an unrepresentative sample are small
  - This chance decreases as the sample size increases
  - This chance can be calculated (sampling error)
Systematic Sampling

Sampling fraction  =  Actual sample size/total population
Systematic Sampling

- Systematic selection of cases from sampling frame (every nth name on database/list)

- Regular intervals with random starting point

- It is probability-based because you can determine the probability of each unit being in the sample

- BUT can be subject to selection bias - need to ensure that the lists do not contain periodic patterns.

- Works equally well for a small or large number of cases.
Stratified Random Sampling

• Divide the population into two or more relevant and significant strata based on one or a number of attributes.

• A random sample (simple or systematic) is then drawn from each of the strata.

• E.g., you may wish to stratify a sample of an organisation’s employees by both department and salary grade.
Cluster Sampling

• Divide the population into discrete groups (clusters) prior to sampling (e.g., you could group your data by type of manufacturing firm or geographical area.

• Your sampling frame is now the complete list of clusters rather than of individual cases within the population.

• You then select a few clusters, normally using simple random sampling. Data are then collected from every case within the selected clusters.
Multi-stage Sampling

Cluster Population

Sample (2 clusters)
Non-Probability Sampling Techniques

- Quota Sampling
- Judgemental/Purposeful Sampling
- Snowball Sampling
Sample Size

- For all non-probability sampling techniques, other than for quota samples (later) there are no rules.

- Your sample size is dependent on your research question(s) and objectives.
Quota Sampling

- Common in market data research, especially interview surveys
- Quota calculated often based on census data
- Interviewer selects people to reach quota
- Commonly opinion polls of c. 1,000 of the voting population are targeted
Purposive / Judgement

The most subjective method based on judgement derived from previous experience

1. Extreme case/deviant sampling
   - Studying the outliers that diverge from the norm as regards a particular phenomenon, issue, or trend.
   - By studying the deviant cases, researchers can often gain a better understanding of the more regular patterns of behaviour.
   - E.g., research into the relationship between study habits and high academic achievement, should purposively sample students considered high achievers.
2. Heterogeneous/maximum variation sampling

- Selected to provide a diverse range of cases relevant to a particular phenomenon or event.
- The purpose of this kind of sample design is to provide as much insight as possible into the event or phenomenon under examination.

- E.g., when conducting a street poll about an issue, a researcher would want to ensure that he or she speaks with as many different kinds of people as possible in order to construct a robust view of the issue from the public's perspective.
3. **Homogeneous sampling**
   - Selected for having a shared characteristic or set of characteristics.
   
   E.g., a team of researchers wanted to understand of what significance having red hair is to red-heads, so they asked red-heads about this. This is a homogenous sample created on the basis of hair colour.
4. **Critical case sampling**

   - Just one case is chosen for study because the researcher expects that studying it will reveal insights that can be applied to other like cases.

   - E.g. If you wanted to study sexuality and gender identity develop among secondary school students, you could select what is considered to be an average secondary school in terms of population and family income, so that your findings from this case could be more generally applicable.
5. Typical case sampling

- Useful when a researcher wants to study a phenomenon or trend as it relates to what are considered "typical" or "average" members of the effected population.

- E.g., if you want to study how a type of educational curriculum affects the average student, then you could choose to focus on average members of a student population.
Snowball Sampling

- Used when the members of a population are difficult to locate.

- The researcher collects data on the few members of the target population he or she can locate, then asks those individuals to provide information needed to locate other members of that population whom they know.
Self-Selection Sampling

Wanted!!!

Research participants

5 Euro Voucher reward
Convenience Sampling
Managing Group Work

Stable working environment:
- Group organisation: defined roles/flexible
- Agree on clear guidelines on how group will work

Run good group meetings:
- What was done
- What needs to be done
- Who/when will do it
<table>
<thead>
<tr>
<th>Activity</th>
<th>Week number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Holiday</td>
<td>1-27,38</td>
</tr>
<tr>
<td>2 Read literature</td>
<td>3-28</td>
</tr>
<tr>
<td>3 Finalise objectives</td>
<td>4-29</td>
</tr>
<tr>
<td>4 Draft literature review</td>
<td>5-30</td>
</tr>
<tr>
<td>5 Read methodology literature</td>
<td>6-31</td>
</tr>
<tr>
<td>6 Devise research approach</td>
<td>7-32</td>
</tr>
<tr>
<td>7 Draft research strategy and method</td>
<td>8-33</td>
</tr>
<tr>
<td>8 Develop questionnaire</td>
<td>9-34</td>
</tr>
<tr>
<td>9 Pilot test and revise questionnaire</td>
<td>10-35</td>
</tr>
<tr>
<td>10 Administer questionnaire</td>
<td>11-36</td>
</tr>
<tr>
<td>11 Enter data into computer</td>
<td>12-37</td>
</tr>
<tr>
<td>12 Analyse data</td>
<td>13-38</td>
</tr>
<tr>
<td>13 Draft findings chapter</td>
<td></td>
</tr>
<tr>
<td>14 Update literature read</td>
<td></td>
</tr>
<tr>
<td>15 Complete remaining chapters</td>
<td></td>
</tr>
<tr>
<td>16 Submit to tutor and await feedback</td>
<td></td>
</tr>
<tr>
<td>17 Revise draft, format for submission</td>
<td></td>
</tr>
<tr>
<td>18 Print, bind</td>
<td></td>
</tr>
<tr>
<td>19 Submit</td>
<td></td>
</tr>
</tbody>
</table>
Managing Group Work

Good group dynamics

Idea Generation
- Ensure everyone has contributed
- Listen to everyone’s contribution
- Identify individual’s key skills

Participation
- Ensure everyone participates
- Use methods such as ‘Round Robin’ to ensure all contributions are captured

Listening
- Negotiate, Compromise, Consensus
Project Management Skills

- Maintain conceptual understanding of project and ensure all work addresses goals of project

- Ensure project plan is manageable and tasks are assigned fairly

- Prioritise tasks effectively

- Keep projects on track – anticipate risks and re-prioritise accordingly
  - Eg. If tasks are running late, re-organise/prioritise

- Manage the project plan

- Tools: google drive, dropbox, mind maps, wiki
Project management - communications

Listen and Clarify

- Avoid interrupting
- Draw out opinions and ideas from whole group
- Use techniques to ensure equal chance to speak
- Pay attention to group dynamics and feelings
- Summarize ideas and feelings frequently

Communicate

- Summarise and synthesise information from meetings
- Document and track decisions and suggestions
Being a Leader

Maintain strategic view of project

Creating constructive Conflict
- Encourage and protect minority opinion
- Seek multiple solutions to problem
- Deal openly with disagreement

Bring out best in people
- Make most of group member's skills, ambition and talent
- Ensure open communication between group members
- Ensure everyone is clear on next steps

Know your subject
Running a Good Meeting

**Preparation**
Always create and distribute an **agenda**

Take notes or ask someone to take **notes**

**Meeting Content**
1. **Review** progress on project to-date
2. Ensure meeting covers all items on agenda within time
3. Each agenda item should have next-steps/solution
4. **Assign** next-steps to group

**Completion**
Write and distribute **overview** of meeting (including task assignments with deadlines)