Sampling Techniques and Questionnaire Design

Department of Political Science and Government
Aarhus University

September 29, 2014

Review: Stratified Sampling

- What is it?
- Why do we do it?
- Most useful when subpopulations are:
  1. identifiable in advance
  2. differ from one another
  3. have low within-stratum variance

Review: Outline of Process

1. Identify our population
2. Construct a sampling frame
3. Identify variables we already have that are related to our survey variables of interest
4. Stratify or subset or sampling frame based on these characteristics
5. Collect an SRS (of some size) within each stratum
6. Aggregate our results

Review: Estimates from a stratified sample

- Within-strata estimates are calculated just like an SRS
- Within-strata variances are calculated just like an SRS
- Sample-level estimates are weighted averages of stratum-specific estimates
- Sample-level variances are weighted averages of strataum-specific variances

Review: Design effect

- Ratio of variances in a design against a same-sized SRS
- $d^2 = \frac{Var_{stratified}(y)}{Var_{SRS}(y)}$
- Possible to convert design effect into an *effective sample size*:
- $n_{effective} = \frac{n}{d}$

Example Setup

- Interested in individual-level rate of crime victimization in Denmark
- We think rates differ among native-born and immigrant populations
- Assume immigrants make up 12% of population
- Compare uncertainty from different designs ($n = 1000$)

SRS

- Assume equal rates across groups ($p = 0.10$)
- Overall estimate is just $\frac{Victims}{n}$
- $SE(p) = \sqrt{\frac{p(1-p)}{n-1}}$
- $SE(p) = \sqrt{\frac{0.09}{999}} = 0.0095$

SRS

- Assume equal rates across groups ($p = 0.10$)
- Overall estimate is just $\frac{Victims}{n}$
- $SE(p) = \sqrt{\frac{p(1-p)}{n-1}}$
- $SE(p) = \sqrt{\frac{0.09}{999}} = 0.0095$
- SEs for subgroups (native-born and immigrants)?

SRS

- Assume equal rates across groups ($p = 0.10$)
- Overall estimate is just $\frac{Victims}{n}$
- $SE(p) = \sqrt{\frac{p(1-p)}{n-1}}$
- $SE(p) = \sqrt{\frac{0.09}{999}} = 0.0095$
- SEs for subgroups (native-born and immigrants)?
- What happens if we don't get any immigrants in our sample?

Proportionate Allocation I

- Assume equal rates across groups
- Sample 880 native-born and 120 immigrant individuals
- $SE(p) = \sqrt{Var(p)}$, where
  - $Var(p) = \sum_{h=1}^{H} (\frac{N_h}{N})^2 \frac{p_h(1-p_h)}{n_h-1}$
  - $Var(p) = (\frac{0.09}{879})(.88^2) + (\frac{0.09}{119})(.12^2)$
  - $SE(p) = 0.0095$
- Design effect: $d^2 = \frac{0.0095^2}{0.0095^2} = 1$

Proportionate Allocation I

- Note that in this design we get different levels of uncertainty for subgroups
- $SE(p_{native}) = \sqrt{\frac{p(1-p)}{879}} = \sqrt{\frac{0.09}{879}} = 0.010$
- $SE(p_{imm}) = \sqrt{\frac{p(1-p)}{119}} = \sqrt{\frac{0.09}{119}} = 0.028$

Proportionate Allocation IIa

- Assume different rates across groups (immigrants higher risk)
- $p_{native} = 0.1$ and $p_{imm} = 0.3$ (thus $p_{pop} = 0.124$)
- $Var(p) = \sum_{h=1}^{H} (\frac{N_h}{N})^2 \frac{p_h(1-p_h)}{n_h-1}$
- $Var(p) = (\frac{0.09}{879})(.88^2) + \frac{0.21}{119})(.12^2))$
- $SE(p) = 0.01022$

Proportionate Allocation IIa

- $SE(p) = 0.01022$
- Compare to SRS:
  - $SE(p) = \sqrt{\frac{0.124(1-0.124)}{n-1}} = 0.0104$
- Design effect: $d^2 = \frac{0.01022^2}{0.0104^2} = 0.9657$
- $n_{effective} = \frac{n}{sqrt(d^2)} = 1017$

Proportionate Allocation IIa

- Subgroup variances are still different
- $SE(p_{native}) = \sqrt{\frac{p(1-p)}{879}} = \sqrt{\frac{.09}{879}} = 0.010$
- $SE(p_{imm}) = \sqrt{\frac{p(1-p)}{119}} = sqrt\frac{.21}{119} = 0.040$

Proportionate Allocation IIb

- Assume different rates across groups (immigrants lower risk)
- $p_{native} = 0.3$ and $p_{imm} = 0.1$ (thus $p_{pop} = 0.276$)
- $Var(p) = \sum_{h=1}^{H} (\frac{N_h}{N})^2 \frac{p_h(1-p_h)}{n_h-1}$
- $Var(p) = (\frac{0.21}{879})(.88^2) + \frac{0.09}{119})(.12^2))$
- $SE(p) = 0.014$

Proportionate Allocation IIb

- $SE(p) = 0.014$
- Compare to SRS:
  - $SE(p) = \sqrt{\frac{0.276(1-0.276)}{n-1}} = 0.0141$
- Design effect: $d^2 = \frac{0.014^2}{0.0141^2} = 0.9859$
- $n_{effective} = \frac{n}{sqrt(d^2)} = 1007$

Proportionate Allocation IIb

- Subgroup variances are still different
- $SE(p_{native}) = \sqrt{\frac{p(1-p)}{879}} = \sqrt{\frac{.21}{879}} = 0.0155$
- $SE(p_{imm}) = \sqrt{\frac{p(1-p)}{119}} = sqrt\frac{.09}{119} = 0.0275$

Proportionate Allocation llc

- Look at same design, but a different survey variable (household size)
- Assume: $\bar{y}_{native} = 4$ and $\bar{Y}_{imm} = 6$ (thus $\bar{Y}_{pop} = 4.24$)
- Assume: $Var(Y_{native}) = 1$ and $Var(Y_{imm}) = 3$ and $Var(Y_{pop}) = 4$
- $Var(\bar{y}) = \sum_{h=1}^{H} (\frac{N_h}{N})^2 \frac{s_h^2}{n_h}$
- $SE(\bar{y}) = \sqrt{\frac{1^2}{880}(.88^2) + \frac{3^2}{120}(.12^2)} = 0.0443$

Proportionate Allocation llc

- $SE(\bar{y}) = 0.0443$
- Compare to SRS:
  - $SE(\bar{y}) = \sqrt{\frac{s^2}{n}} = \sqrt{4/1000} = 0.0632$
- Design effect: $d^2 = \frac{0.0443^2}{0.0632^2} = 0.491$
- $n_{effective} = \frac{n}{sqrt(d^2)} = 1427$

Proportionate Allocation llc

- $SE(\bar{y}) = 0.0443$
- Compare to SRS:
  - $SE(\bar{y}) = \sqrt{\frac{s^2}{n}} = \sqrt{4/1000} = 0.0632$
- Design effect: $d^2 = \frac{0.0443^2}{0.0632^2} = 0.491$
- $n_{effective} = \frac{n}{sqrt(d^2)} = 1427$
- Why is d^2 so much larger here?

Disproportionate Allocation I

- Previous designs obtained different precision for subgroups
- Design to obtain stratum-specific precision (e.g., $SE(p_h) = 0.02$)
- $n_h = \frac{p(1-p)}{v(p)} = \frac{p(1-p)}{SE^2}$
- $n_{native} = \frac{0.09}{0.02^2} = 225$
- $n_{imm} = \frac{0.21}{0.02^2} = 525$
- $n_{total} = 225 + 525 = 750$

Disproportionate Allocation II

- Neyman optimal allocation
- How does this work?
  - Allocate cases to strata based on within-strata variance
  - Only works for one variable at a time
  - Need to know within-strata variance

Disproportionate Allocation II

- Assume big difference in victimization
- $p_{native} = 0.01$ and $p_{imm} = 0.50$ (thus $p_{pop} = 0.0688$)
- Allocate according to: $n_h = n\frac{W_h S_h}{\sum_{h=1}^{H} W_h S_h}$
- $\sum_{h=1}^{H} W_h S_h = (0.88 * 0.0099) + (0.12 * 0.25) = 0.0387$
- $n_{native} = 1000\frac{0.0087}{0.0387} = 225$
- $n_{imm} = 1000\frac{0.03}{0.0387} = 775$

Disproportionate Allocation II

- $SE(p_{native}) = \sqrt{\frac{p(1-p)}{225}} = \sqrt{\frac{0.0099}{225}} = 0.00663$
- $SE(p_{imm}) = \sqrt{\frac{p(1-p)}{775}} = \sqrt{\frac{.25}{775}} = 0.01796$
- $Var(p) = \sum_{h=1}^{H} (\frac{N_h}{N})^2 \frac{p_h(1-p_h)}{n_h-1}$
- $Var(p) = (\frac{0.0099}{225})(.88^2) + (\frac{0.25}{775})(.12^2)$
- $SE(p) = 0.00622$

Disproportionate Allocation II

- $SE(p) = 0.00622$
- Compare to SRS:
  - $SE(p) = \sqrt{\frac{0.0688(1-0.0688)}{n-1}} = 0.008$
- Design effect: $d^2 = \frac{0.00622^2}{0.008^2} = 0.6045$
- $n_{effective} = \frac{n}{sqrt(d^2)} = 1286$

Final Considerations

- Reductions in uncertainty come from creating homogeneous groups
- Estimates of design effects are variable-specific
- Sampling variance calculations do not factor in time, costs, or feasibility

Questions about stratified sampling?

Cluster Sampling

- What is it?
- Why do we do?

Cluster Sampling

- What is it?
- Why do we do?
- Most useful when:
  1. Population has a clustered structure
  2. Unit-level sampling is expensive or not feasible
  3. Clusters are similar

Cluster Sampling

- Advantages

Cluster Sampling

- Advantages
  - Cost savings!
  - Capitalize on clustered structure

Cluster Sampling

- Advantages
  - Cost savings!
  - Capitalize on clustered structure
- Disadvantages

Cluster Sampling

- Advantages
  - Cost savings!
  - Capitalize on clustered structure
- Disadvantages
  - Units tend to cluster for complex reasons (self-selection)
  - Major increase in uncertainty if clusters differ from each other
  - Complex to design (and possibly to administer)
  - Analysis is much more complex than SRS or stratified sample

Cluster Sampling

- Number of stages
  - One-stage sampling
  - Two- or more-stage sampling
- Number of clusters
- Sample size w/in clusters
- Everything depends on variability of clusters

Sampling Variance for Cluster Sampling

- Sampling variance depends on *between*-cluster variation:
  $Var(\bar{y}) = (\frac{1-f}{a})(\frac{1}{a-1})(\sum_{\alpha=1}^{a}(\bar{y}_\alpha - \bar{y})^2)$
- When *between*-cluster variance is high, *within*-cluster variance is likely to be low
  - "Cluster homogeneity"

Design Effect for Cluster Sampling

- Cluster samples almost always less *statistically* efficient than SRS
- Design Effect depends on cluster homogeneity:
  - $d^2 = \frac{Var_{clustered}(y)}{Var_{SRS}(y)}$
  - $d^2 = 1 + (n_{cluster} - 1)roh$
- *roh* (*intraclass correlation coefficient*):
  - Proportion of unit-level variance that is between-clusters
  - Generally positive and small (about 0.00 to 0.10)

Questions about cluster sampling?

Example: Burnham et al.

- What is the research question?

Example: Burnham et al.

- What is the research question?
- What are the population and unit of analysis?

Example: Burnham et al.

- What is the research question?
- What are the population and unit of analysis?
- What is the sampling strategy? Why?

Example: Burnham et al.

- What is the research question?
- What are the population and unit of analysis?
- What is the sampling strategy? Why?
- What do they find?

Complex Survey Designs

- Often stratification and clustering are used together
- The choice of design must do at least one of:
  - Improve statistical efficiency
  - Improve ease/cost of implementation
- Design effects
- Weights

Concept definition and Operationalization

- Questionnaires start with concept definition
- Multiple ways to operationalize any concept
- Important concepts may require multiple measures

Topics of questions

- Evaluations (opinions, attitudes, etc.)
- Recall (behavior, events, knowledge, etc.)
  - Demographics (age, sex, ethnicity, etc.)

Structure of a question

- Survey mode
- Survey context
- Vignette or introductory text
- Question itself
- Response format and options
- Follow-ups, branches, checks, validation, clarification

Evaluative questions

- Name an object of evaluation
- Possibly describe that object
- Ask for a transformation of the evaluation onto a set of responses
- Individuals differ in how they form opinions
  - Memory-based processing
  - Online processing

Response options for evaluative questions

- Ratings
  - Bipolar
  - Branching
  - Unipolar
- Scales/Thermometers
- Agree-disagree
- Forced choices
- Open-ended
- Rankings (note: need alternatives to rank against)

Extended Example

- Public opinion survey in Denmark
- Construct: Opinion toward Danish involvement in air strikes on Islamic State militants in Iraq and Syria

Example: Rating (bipolar)

Do you support or oppose Denmark's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria?

- Strongly support
- Somewhat support
- Neither support nor oppose
- Somewhat oppose
- Strongly oppose

Example: Rating (branching)

Do you support or oppose Denmark's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria?

- Support
- Neither support nor oppose
- Oppose

Would you say that you strongly [support|oppose] or somewhat [support|oppose] Denmark's participation?

- Strongly
- Somewhat

Example: Rating (bipolar)

Are you favorable or unfavorable toward Denmark's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria?

- Very favorable
- Somewhat favorable
- Neither favorable nor unfavorable
- Somewhat unfavorable
- Strongly unfavorable

Example: Rating (unipolar)

To what extent do you support Denmark's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria?

- Strongly
- Moderately
- Somewhat
- Not at all

Example: Rating (unipolar)

How favorable are you toward Denmark's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria?

- Extremely favorable
- Very favorable
- Moderately favorable
- Somewhat favorable
- Not at all favorable

Example: Numbered Scale

On a scale from 1 to 5, with 1 being “strongly oppose” and 5 being “strongly support,” to what extent do you support Denmark’s participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria?

1. Strongly oppose
2.
3.
4.
5. Strongly support

Example: Thermometer

We would like to get your feelings toward some of political policies. Please rate your support for the policy using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward the policy. Ratings between 0 degrees and 50 degrees mean that you don't feel favorable toward the policy. You would rate the policy at the 50 degree mark if you don't feel particularly favorable or unfavorable toward.

Denmark's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria.

- 0–100 slider

Example: Agree/Disagree (bipolar)

To what extent do you agree with the following statement:
I support Denmark's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria.

- Strongly agree
- Somewhat agree
- Neither agree nor disagree
- Somewhat disagree
- Strongly disagree

Example: Agree/Disagree (unipolar)

To what extent do you agree with the following statement:
I support Denmark's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria.

- Agree completely
- Agree to a large extent
- Agree to a moderate extent
- Agree a little bit
- Agree not at all

Example: Forced choice

When thinking about Denmark's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria, which of the following comes closer to your opinion:

- Denmark should participate in air strikes
- Denmark should not participate in air strikes

Example: Open-ended

In your own words, how would you describe your opinion on Denmark's participation in U.S.-led air strikes on Islamic State (IS) in Iraq and Syria?

Additional Considerations

- How many response categories?
- Middle category (presence and label)
- “no opinion” and/or “don’t know” options
- Probe if “no opinion” or “don’t know”?
  - Encourage guessing?
  - Clarify/describe object of evaluation?
- Branching format?
- Order of response categories
- Changes based on survey mode

Questions about writing evaluative questions?

Activity!

- Generate questions in pairs
- Discuss with the class

Assignment for next week

- What constructs/concepts do you intend to measure in your survey?
- How do you plan to measure these?
- How have these constructs been operationalized in other research?

Next week's agenda

- Continue questionnaire design
- Measuring sensitive information
- Measuring knowledge
- Reference periods