

STAT 100

Chapter 4

Richard Lockhart

Simon Fraser University

Fall 2014 — Surrey



Jargon to learn

- *Undercoverage, overcoverage, frame.*
- *response bias, non-response bias, selection bias.*
- *Weighting, editing, imputation.*
- *stratum/strata, multi-stage, cluster sampling,*
- *sampling errors, non-sampling errors.*
- *CATI, CAPI*



Summary of key ideas from last time

- Parameters are numbers describing populations.
- Statistics describe samples.
- We usually never find out the value of *parameters*.
- A statistic used to guess the value of a parameter is an *estimate*.
- The margin of error in a survey describes the likely size of random variation.
- Usually cited in a *confidence interval*.



Frames

- Sampling requires a *frame*.
- *Undercoverage* occurs when some members of the target population are NOT on the frame.
- *Overcoverage* occurs when some population members are on the frame more than once.
- Frame for National Household Survey is provided by Census.
- Many frames are *area frames* — lists of dwellings or properties rather than lists of people.



Non-sampling error

- *Questionnaire effects*. Exact wording matters, question order matters, subject of previous questions on the survey matters.
- *Response errors*. Some answers are probably wrong: father on census form younger than child? “Son” listed as female.
- *Interviewer bias*. Answers given to female interviewers may be different than those given to male interviewers.
- *Mode effects*. Same question asked in person, by telephone, over the net gets different answers from same person.



Close look at Labour Force Survey

- **Target Population.** Notice about 2% exclusion and special treatment of North.
- *Stratified, multi-stage, cluster* rotating panel sample.
- *Strata* are formed by breaking up provinces into Economic Areas or Census Metropolitan Areas, then into smaller groups.
- Within strata break up into small areas of a few hundred dwellings; these are *clusters*.
- Pick some clusters at random.
- Make list of dwellings within selected cluster.
- Sample dwellings within cluster.
- Collect data on all adults in each dwelling.



Panel Design

- Once sampled dwelling stays in for 6 months.
- 1/6 of dwellings replaced each month.
- Makes estimates of month over month change more precise.
- Also rotate clusters out of sample after longer period.



Interviewing techniques

- Computer Assisted Personal Interview (CAPI) in first month.
- Then 5 months of Computer Assisted Telephone Interview (CATI).
- Questionnaire.
- Things to notice.
- Follow-up for non-response. Response legally required. STC reports 10% non-response.



Data Processing

- *Edit* stage: answers checked for accuracy and reasonableness.
- Many errors caught by computer flags during CATI/CAPI.
- *Imputation* stage: missing or rejected answers replaced by “imputed” value.
- *Weighting*. Each person has *raw weight* from random sampling. (If 1 household in 50 chosen each person gets weight 50.)
- Then weights adjusted for non-response.
- Then adjusted again to match census.
- No longer wise to match long form census data.



Weighting

- Weighting does 2 things: shrinks bias we hope, and lowers variability of estimate.
- Scottish Independence.

	15-24		25-39		40-59		60+	
Group	M	F	M	F	M	F	M	F
Unwtd	52	59	115	114	212	223	156	153
WTD	80	79	125	129	185	195	132	158

- Notice that young people are under-represented – so we *adjust* for this known source of bias.



Further issues

- Publication is suppressed for potentially identifying information.
- Unemployment results are seasonally adjusted.
- Weather in Canada is bad in the winter – actual numbers unemployed always rise in winter!
- Seasonal adjustment increases number employed in winter, lowers it in summer.
- Reference period: week containing 15th of month.
- Published about 3 weeks later.
- Revised later as errors corrected, updated information received.

