**W2 Mere Randomization**

**2.1 Simple Random Sampling**

Frame & RN’s

• Sample size n from population size N

• Using random numbers from a table, match numbers to unique numbers assigned to each frame element

- for each selection after the 1st, check to see if the frame element has already been selected

- if selected already, reject the selection and use another random number

- continue process until n distinct frame element selected

• Without replacement selection: unique sample elements

• Every element of the population has the same probability of selection (epsem = equal probability selection method) and every combination of size n has the same probability of selection

Selecting a Sample

• Sample size n from population size N

• Using random numbers from a table, or a statistical software system, assign a random number to each of the N frame elements

- sort the frame elements by random number, from smallest to largest

- select the first n frame elements

• This is without replacement & epsem

• Random number generation for n = 500:

RN = TRUNC(URAN(0718)\*500)+1

- URAN: starting point e.g. 0718 🡺 start from page 7 column 18

- TRUN: truncate off the decimal point

Selecting a sample – Another method

• sample size n from population size N

• using random numbers from a table, match numbers to unique numbers assigned to each frame element

- continue process until n frame element selected

- check for duplicates in sample

- If any duplicates occur, reject sample, and draw another sample of size n

- without replacement & epsem

- restricted (simple) RS v.s. unrestricted RS

Definitions of simple random sampling

• Any procedure with fixed sample size n and for which every element of the population has the same probability of selection (epsem) and every combination of size n has the same probability of selection

• All sets of size n distinct elements from N – pick one (N choose n)

Practical Use

• Widely used for simple problems

• But rarely used by practitioners in ‘isolation’

- complicated for ‘lay’ administration

- more efficient methods available

- relies only on randomization

• For practitioner a tool to be used in conjunction with other methods

- random sample of elements within a group

- random sample of groups

**2.2. A short history**

• Sampling practice

- result of attempts to solve practical problems

• Function of theory

- formalize implicit assumptions, and confirm, correct or extend practice

• Origins

- data gathering

- health and social problems

- social physics

- census

- monography

Representative Method

• Kaier: Representative method

- miniature of country

- large number of units

- use prior information in selection

• Von Mayr and others: Census

- no calculation where observation is possible

• Cheysson and others: Monography

- detailed examination of typical cases

Randomization

• Representative

- purposive sampling

- expert choice

- balanced sampling

• Objective

- randomized selection

- Bowley, 1906 (colleague of R.A. Fisher)

• Neyman 1934

- The sampling distribution

- properties of sample under repeated sampling: All possible samples and their associated probabilities of occurrence

- the sampling distribution of an estimator

Comparison

• Conditions under which different procedures will produce valid estimates

- probability sampling

- “unbiased” irrespective of population structure

- purposive/balanced/quota sampling

- tough assumptions about population structure, unlikely to be achieved in practice

Principles

• Probability sampling for objectivity

• Stratification for precision (representativeness)

• Variance estimation from the sample

• Complete and comprehensible description of the sampling procedure

**2.3 SRS sampling distributions**

Basic Framework

• A sample design for which the unit of selection is population element

• Basic framework: Neyman 19334

- must be application to all populations

- must not depend on assumptions about the population structure

- appropriate for large populations of elements

• Repeated sampling

- objective (mechanical) selection of elements

- consider possible outcomes of the sampling process

- evaluation of the whole set of possible outcomes

• The set of all possible values of the estimator that can be obtained with a given sample design

- for a given sample we obtain a particular value, the estimate (such as y\_bar)

• We want to know …

- … how likely is the estimate to be close to the population value?

• In fact, we select just one sample

• The estimate may be correct, or incorrect

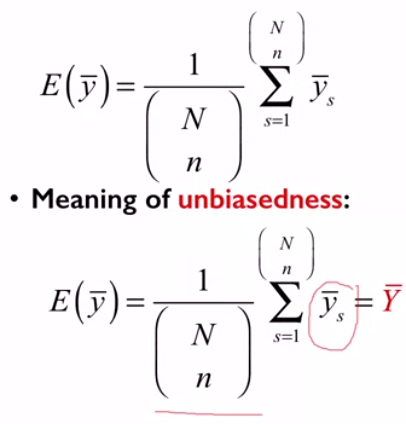
• Want to maximize the probability of a satisfactory estimate

Properties of the sampling distribution

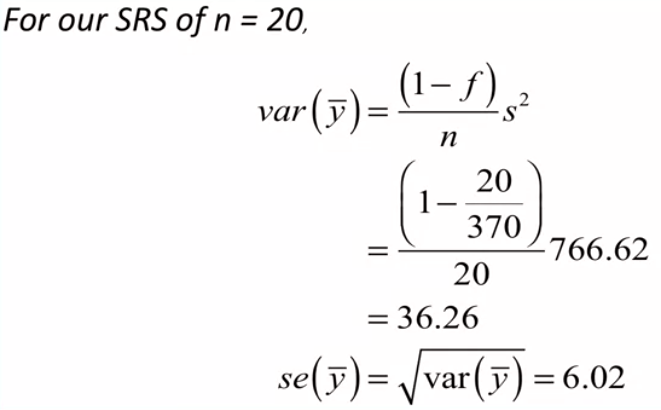
• Unbiasedness

- expected value (average value): E(y\_bar)

- meaning of expected value:



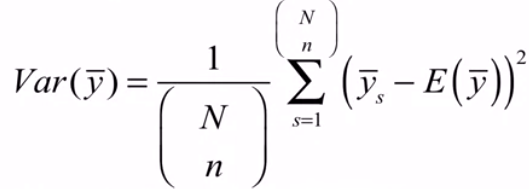
• standard error



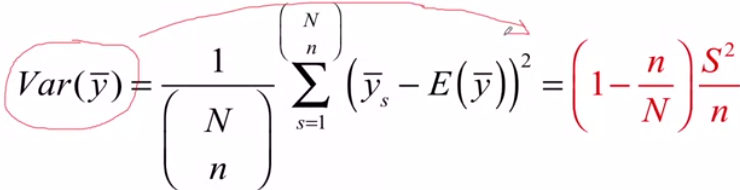
• Variability from one sample to another

- variance of the estimator: Var(y\_bar)

- meaning of the variance:

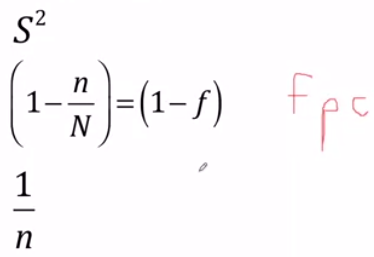


- Algebraically equivalent formula:

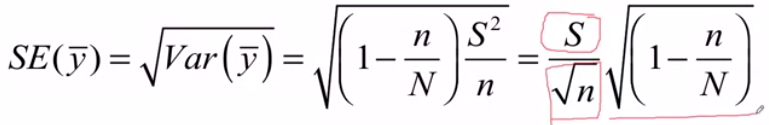


• Variability from one sample to another

- components



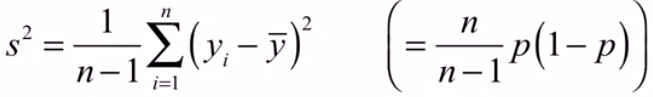
- scale conversion



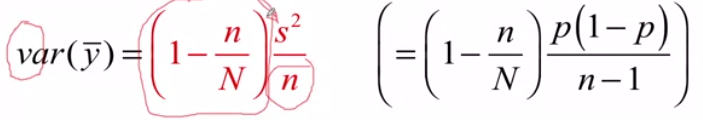
• Estimating variability from one sample to another

- element variance: S^2

- estimated element variance:



- estimated variance & standard error



Confidence Intervals

• For large samples, the sampling distribution of y\_bar is normal

- law of large numbers or central limit theorem

• Form an interval around y\_bar:



• (1-alpha)% or 95% confidence interval

• A statement of uncertainty about our estimated mean

**2.4 Sample Size**

What we need to know

• What sample size do we need to obtain a give standard error of the estimator?

• S^2 population variance known (or guessed)

- census

- other surveys

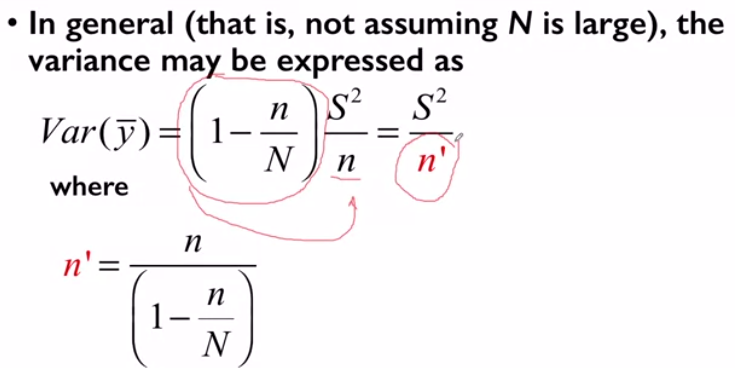
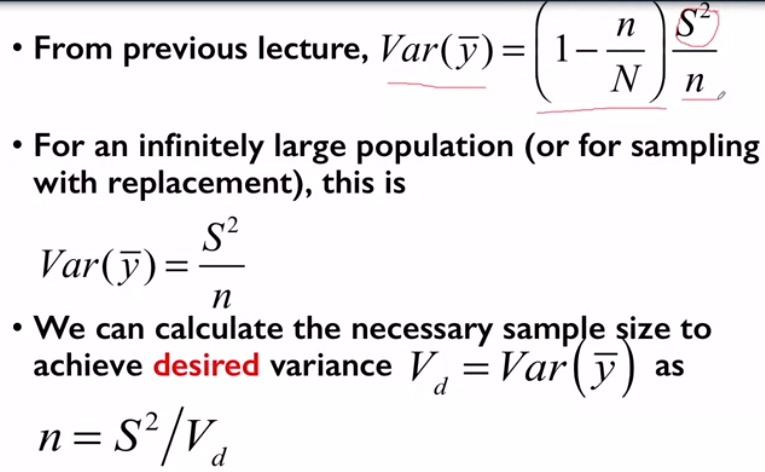
- administrative records

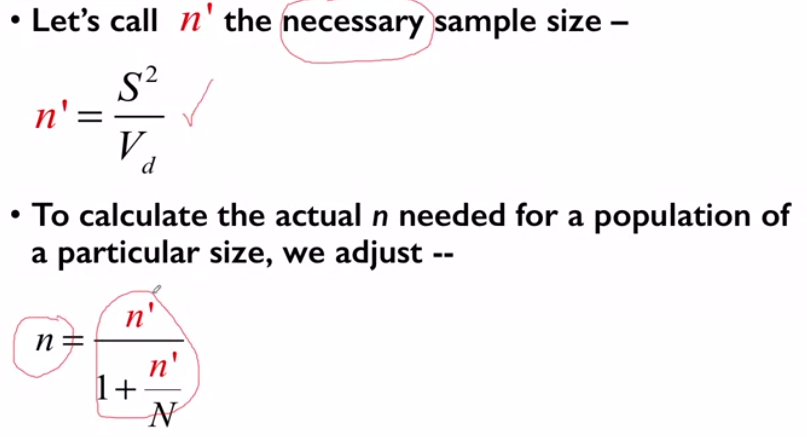
• Desired standard error

- policy requirements in terms of root(Var(y\_bar))

- decision making requirements

Sample Size Formula





Example

• Interested in U.S. population attitudes about how well its current president is doing his or her job

- “Do you approve or disapprove of the job President Obama is doing as President?” (if approve/disapprove, ask: ) “Do you approve/ disapprove strongly or somewhat?”

• Estimate the proportion P approving strongly or somewhat in a new survey

• Suppose p = 0.6 in the last survey

- then s^2 = p(1-p) = 0.6(1-0.6) = 0.24

• for our new survey about to be conducted, “project” that S^2 = 0.24

• Also need to specify precision of the new survey estimate … in advance … the V\_d = Var(y\_bar)

• Suppose we would like to end up with an uncertainty statement that says that between 58% and 62% of the U.S. population think President Obama is doing a good job … at a 95% level of confidence

• Recall that the upper confidence limit, the 62% value, is the proportion, 60% in this case, plus a multiplier times the standard error

• That is, 62% = 60% + z x se(60%)

• For a 95% confidence interval, z = 1.96, say z = 2

• Then, if 62% = 60% + 2 x se(60%), se(60%) = 1%

• If that’s the kind of confidence interval we want, then we want a standard error of 1%

• of course, se(p) = 0.01 is another way to say this, in terms of what we want to have happened

• proportions are better to work with than percentages

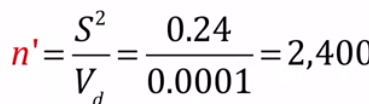
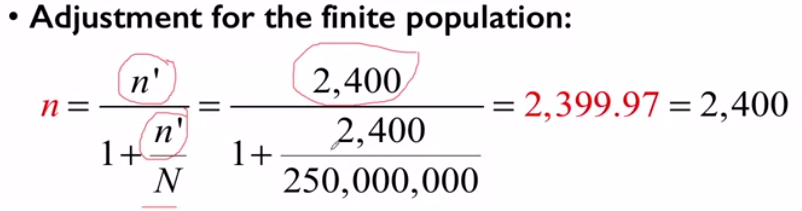
• We need the square of the standard error, or the variance

V\_d = Var(p) = (SE(p))^2

That is V\_d = (0.01)^2 = 0.0001

• Hence, we have S^2 = 0.24 and V\_d = 0.0001

• This yields a necessary sample size of:

**2.5 Margin on Error**

Two questions to consider following up from the previous section…

• Is there a more direct way to figure this out from a projected confidence interval?

• Why doesn’t the population size have a big effect on the sample size?