Figure 3.14. PLANNED DATA COLLECTING: Sampling and Sampling Distributions

Program 14 in: *Against All Odds: Inside Statistics*

Although simple random sampling (or equiprobable selecting) is the foundation of sampling designs (or sampling protocols), sample surveys in practice usually use more complex designs. This program begins with a look at these designs. In *stratified random sampling* (or EPS from a stratified population), the population is divided into *strata*, groups of units or elements similar to each other in way(s) that may affect their response. A separate SRS is then selected from each stratum. Stratified sampling designs are the analogue of block designs for an experiment. The video illustrates stratified sampling by a survey of recreational fishermen carried out by the National Marine Fisheries Service. The strata in this case are different locations for fishing.

National samples are *multistage samples* that select successively smaller regions within the population in stages. Each stage may employ SRS, stratified sampling, or another type of sampling. The General Social Survey conducted by the National Opinion Research Center is an example of a national sample that provides important information about public sentiment.

Failure to use probability sampling often results in *bias*, or systematic errors in the way the sample represents the population. This program also reminds us of some of the practical difficulties that may cause bias when sampling *people*. A sampling design that systematically misses part of the population, like the *Literary Digest* election poll mentioned in the video, will suffer from bias due to this *undercoverage.*. Misleading results can also be due to *non-response* (some subjects can't be contacted or refuse to answer), to badly worded questions, or even to the race, sex, or appearance of the interviewer.

The deliberate use of chance in producing data is intended to eliminate bias. It also makes the outcomes of an experiment or sample survey subject to the laws of probability. The next four Programs will introduce probability; the final portion of this program takes a first look at the random behaviour of statistical data. First, some basic vocabulary: A number that describes a population is called a *parameter.* A number than can be computed from the data is called a *statistic.* The purpose of sampling or experimentation is usually to use statistics to make statements about unknown parameters.

A statistic from a probability sample or randomized experiment will not take the same value if the sampling or experiment is repeated. The *sampling distribution* describes how that statistic varies in repeated data collection. In the video you can watch a sampling distribution build up under repeated SRS from a population of beads – the notable **quincunx** demonstration, about 20 minutes into the video.. Formal statistical inference is based on the sampling distributions of statistics.

Bias, which we informally described as "favoritism", can be described more exactly in terms of the sampling distribution. Bias means that the centre of the sampling distribution is not equal to the true value of the parameter. We would like to use a statistic that has small bias and also does not vary greatly in repeated sampling or experimenting. The *variability* of the statistic is described by the spread of its sampling distribution. If the sampling distribution is normal, the standard deviation of the distribution describes the variability of the statistic. Properly chosen statistics from randomized data production designs have no bias due to selecting the sample or assigning the experimental units to treatments. The variability of the statistic is determined by the size of the sample or of the experimental groups: *larger* samples give *less* variable results. Notice in particular that as long as the population is much larger than the sample, the variability of sample statistics is influenced *only* by the size of the sample and *not* by the size of the population.

1 Using the terminology of our Course Materials, rewrite the following phrases or sentences from the video summary above.

- *A separate SRS is then selected from each stratum.* (First paragraph)
- *Stratified sampling designs are the analogue of block designs for an experiment.* (First paragraph)
- *The deliberate use of chance in producing data is intended to eleminate bias.* (Fourth paragraph)
- *It also makes the outcomes of an experiment or sample survey subject to the laws of probability.* (Fourth paragraph)
- *..... a first look at the random behaviour of statistical data.* (Fourth paragraph)
- *A number that describes a population is called a* parameter. (Fourth paragraph)
- *A number that can be computed from the data is called a* statistic. (Fourth paragraph)
- *The purpose of sampling or experimentation is usually to use statsitcs to make statements about unknown parameters.* (Fourth paragraph)
- *A statistic from a probability sample or randomized experiment will not take the same value if the sampling or experiment is repeated.* (Fifth paragraph)
- *Properly chosen statistics from randomized data production designs have no bias due to selecting the sample or assigning the experimental units to treatments.* (Last paragraph) [this statement is shortened to ... *random sampling eliminates the bias ...* about 20 minutes into the video commentary.]
- *..... as long as the population is much larger than the sample, the variability of sample statistics is influenced* only *by the size of the sample and* not *by the size of the population.* (Last paragraph)

(*continued overleaf*)

2 Identify the four categories of *error* (in our terminology) involved in the third paragraph overleaf on page 3.71: ..... *Failure to use probability sampling* ..... .

- Rewrite the paragraph using our terminology.

---

1 Phrases from the video summary overleaf on page 3.71, rewritten in the terminology of these Course Materials, are as follows.

+ A sample is selected from each stratum by EPS; these stratum (sub)samples together comprise the (overall) sample.
+ Stratifying and blocking are *similar* in that both involve forming groups of elements (or units) [the stata or the blocks] with the same (or similar) values for one or more (non-focal) explanatory variates, but stratifying and blocking *differ* in their (primary) intents.
  Stratifying (in the Plan for an investigation with a *descriptive* aspect) aims to increase the *precision* of estimators and (often) also to provide Answers for population *sub*groups (the strata) as well as for the whole population.
  Blocking (when feasible) [in the Plan for an investigation with a *causative* aspect], in conjuction with equiprobable assigning, aims to manage *confounding* by *non*-focal explanatory variates and so decrease the limitation on an Answer involving *causation* by the *focal* variate.
  – Management of non-focal explanatory variates by blocking often results in *in*creased precision of estimators, but this is an *incidental*, although useful, benefit.
+ The next three statements are all encompassed by:
  *Equi*probable selecting (and *probability* selecting more generally) provides a basis for quantifying sampling imprecision and so, *in conjunction with adequate replicating* (or an *adequate sample size*), allows an Answer to be obtained with acceptable limitation imposed by sample error in the context of a particular investigation.
+ The next two statements are both covered by:
  Attribute: a quantity defined as a function of the response (and, perhaps, explanatory) variate(s) over a *group* of elements/units, typically: – the target population/process, – the respondent population, – the sample. – the study population/process, – the non-respondent population,
  Examples of attributes (commonly of interest) are averages, proportions, totals and ratios of two averages.
+ In the usual situation of an investigation involving a *sample* (as distinct from a *census*), an Answer is usually an estimate (based on a model for the behaviour under repetition of a sample attribute) for (the value of) a population attribute.
+ Under repetition of equiprobable selecting, the value of a sample attribute will (usually) vary from sample to sample.
+ Under EPS, the random variable in the model representing the sample average or the sample proportion is an *un*biased estimator of the corresponding population attribute.
  For estimating a ratio of two averages under EPS, there *is* estimating bias in the model; however, *un*like (real-world) inaccuracy, estimating bias *de*creases in magnitude with increasing sample size.
+ Under EPS, provided the sample size is a *small* proportoin of the population size (*i.e*, $n \ll N$), the population size has little or no practically important effect on the standard deviation of the random variable representing the sample attribute.
  The *un*intuitive consequence is that the (model) random variable representing the sample attribute under EPS has essentially the *same* precision, as an estimator of the population attribute for a sample size of (say) 1,500 respondents, *regardless* of whether the population size is 30 million (as for Canada, say) or 300 million (as for the U,S.).

2 The four categories of *error* (in our terminology) involved in the third paragraph overleaf on page 3.71: ..... *Failure to use probability sampling* ..... are (in order):

– sample error; – study error; – non-response error; – measurement error.

[For an investigation with a *descriptive* aspect, we recall the pictorial representation of how these four categories of error comprise *overall* error on the lower half of page 3.13 on the first side of Figure 3.3.]

+ Non-probability sample selecting methods impose *severe* limitation on Answers because they provide no basis for quantifying the likely magnitude of *sample* error. Further, even under EPS, a Plan for an investigation that fails to adequately manage
  – *study* error (arising from inadequate overlap of the study and target populations),
  – *non-response* error [arising from a non-respondent population that is appreciable in size in relation to that of the respondent population, and/or whose attribute(s) of interest differ appreciably between the two populations],
  – *measurement* error [*e.g.*, arising from the questionnaire and its administration including, for instance, badly worded questions or the race, sex, or appearance of the interviewer (if there is one)],

  can impose limitations on Answer(s) that compromise their usefulness – *i.e*, Answer(s) whose departure from the true state of affairs may be large enough to be practically important in the investigation context.
  We are reminded that severe limitation on Answer(s) can arise from *in*adequate management of *any one* category of error, even if all the other categories are managed adequately.