

Lecture 18

①

We have been talking about sampling and how we use a sample to infer something about the population from which it is drawn. Today we start putting this ~~into~~ ^{into} a more formal statistical framework.

Definitions

parameter a number that describes the population. In practice unknown ^{but fixed value}

statistic a number which describes the sample. Its value is known based on any given sample. May change value from sample to sample.

Note we use a statistic to estimate the unknown population parameter

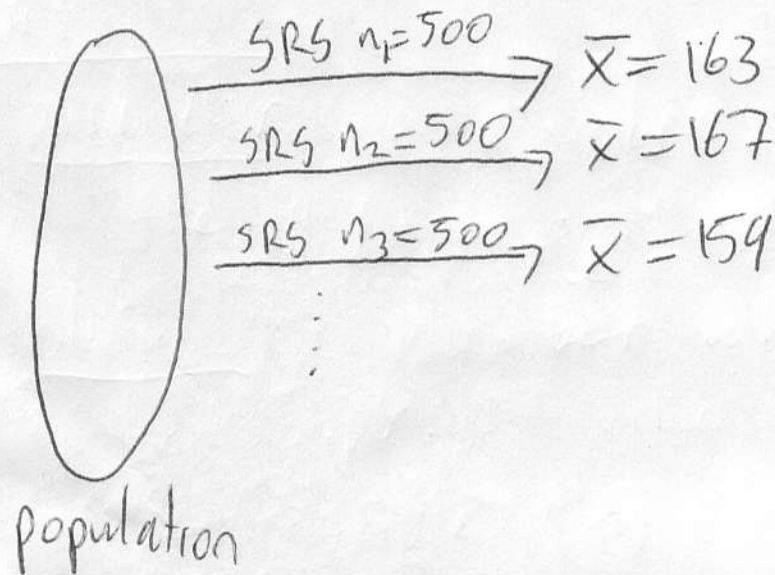
<u>Examples</u>	population	statistic
mean	μ	\bar{X}
std deviation	σ	S
proportion	p	\hat{p}

Sampling variability

(2)

Surprisingly statistics have variability. Why is this? It is because different samples (even of the same size) will give different values of the statistic. i.e. if I sample ^{out of} different individuals I will get different value of my statistic.

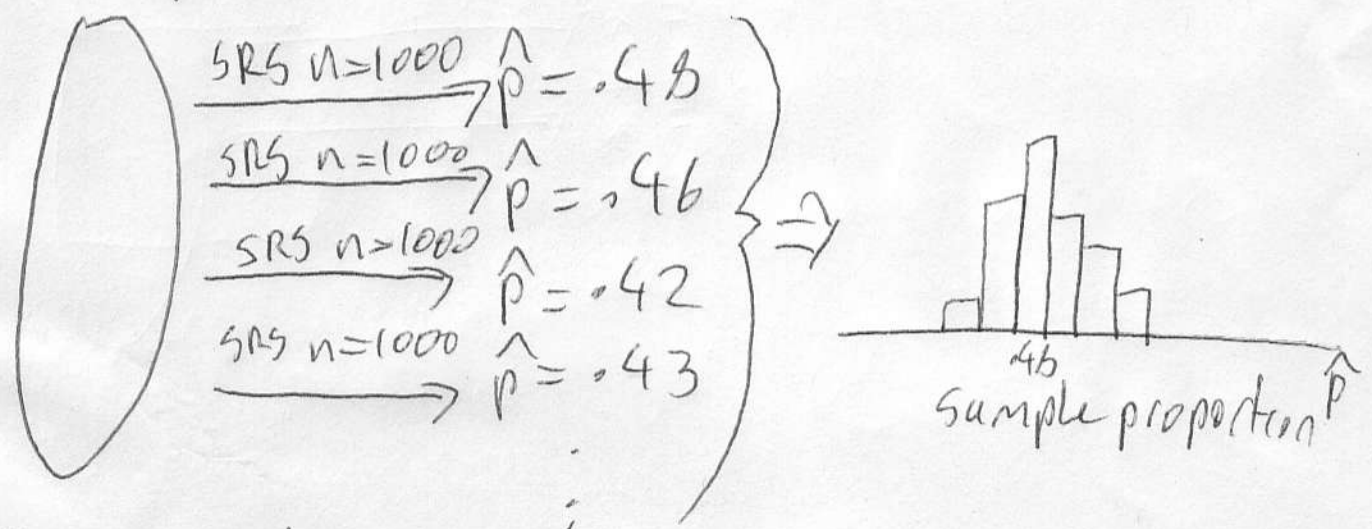
eg



The sampling distribution of a statistic is the distribution of values taken by the statistics in all possible samples of some fixed size from same population.

eg Take samples of size 1000 From the population.
 Ask "Do you agree with the Presidents Iraq policies?" Population parameter p is
 The proportion of people in population who say "yes I agree". Sample statistic
 $\hat{p} = \frac{\# \text{ people in sample who say "I agree"}}{1000}$

USA population

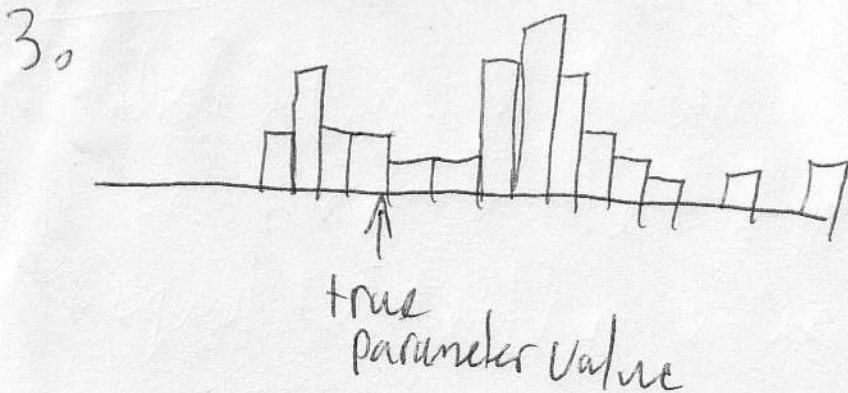
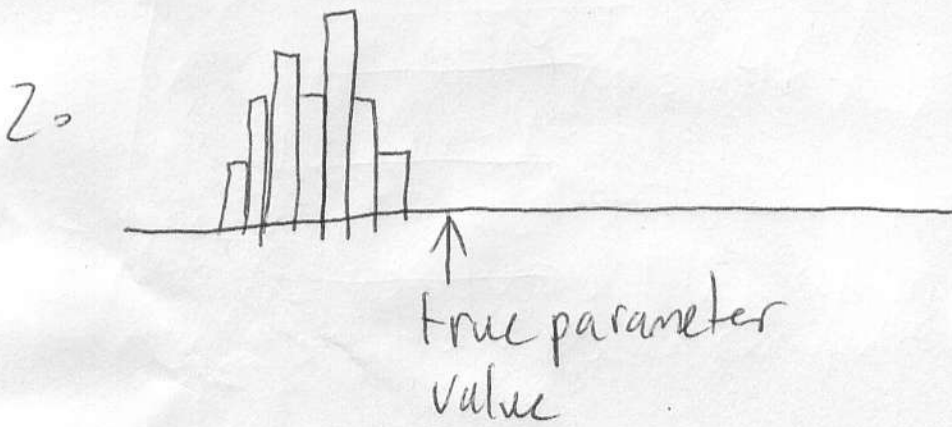
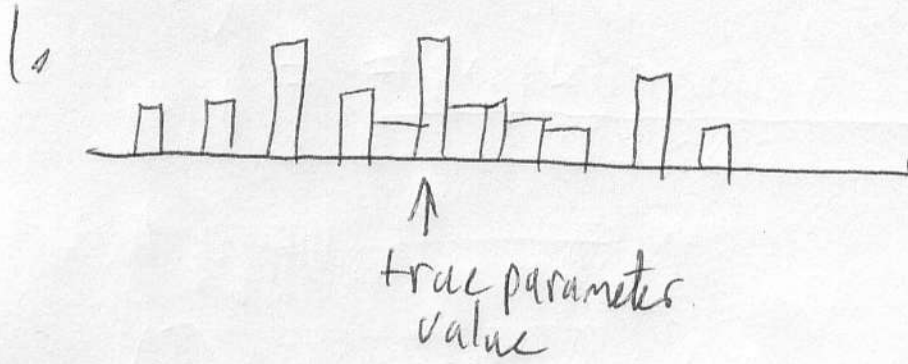


Ex 3.67 gets you to simulate a sampling distribution.

Bias Online there is an example created using a computer.
 - concerns the center of the sampling distribution. A statistic is unbiased if the mean of the sampling distribution is equal to the true value of the parameter.

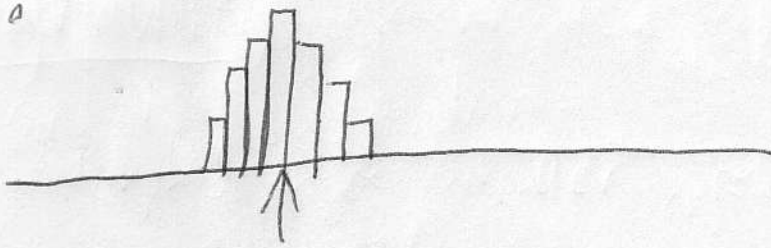
variability concerns the spread of the sampling distribution. Statistics based on larger samples have smaller spreads

examples Variable Yes/no? ^P Based Yes/no? ^P
 high/low high/low



G₀

5



true parameter
value

which of one of these cases is the ideal?

G₀ has low bias and low variability so it is best. In real life we only take 1 sample so we would like our statistic to have low bias and low variability and so likely to above or near the true value.

Dealing with Bias and Variability

Bias - random sampling reduces risk of bias

variability - larger sized samples reduce variability

Note Population size has no effect on variability of statistic (small sample size relative to population size)

Example

(6)

An agency of the federal govt takes a SRS of residents in each state to estimate the proportion of owners of real estate in ^{each of} the states population. Populations range from about 494K people in Wyoming to 35 million in California.

- (a) Does the variability of the sample proportion vary from state to state if a SRS of size 2000 is taken in each state? why?
- (b) Will the variability of the sample proportion change if a SRS of 0.1% of the states population is taken in each state. Explain. which state will have the highest variability? the lowest?

Answers

- (a) No. Variability depends only on sample size. Since the sample sizes are the same size in each state there is no difference in variability.
- (b) Yes. The estimates in the smaller states will be more variable than those in the larger states because they have smaller samples. California with the largest sample size will be least variable. Wyoming largest variability.