

Chapter 9: Sampling Distributions

9.1 Sampling Distributions

Sample or Population?

parameter: a number that describes a **p**opulation

- fixed number
- do not know it's exact value

statistic: a number that describes a **s**ample

- not fixed - changes from sample to sample
- used to estimate an unknown parameter

Population (parameters)

μ • the mean of a population

p • a population proportion

Sample (statistics)

\bar{x} • the mean of the sample
• \bar{x} from an experiment or sample is an estimate of μ

\hat{p} • (p-hat) a sample proportion
• it is a count of objects (X) that are "successes" in a sample out of the total sample (n)

Sampling variability - \bar{x} (the value of any statistic) varies from sample to sample within the same population.

That's OK, as long as we:

- Take many samples from the same population
- Calculate \bar{x} or \hat{p} for each sample
- make a histogram
- examine SOCS

Sampling Distribution: the distribution of values for **all possible samples** of the same size from the same population.

(it is an ideal pattern that we will approximate using sampling or simulation.)

Describing Sampling Distributions:

- **Shape**
 - is it symmetric?
 - is it approximately normal?
- **Outliers**
 - are there any potential outliers?
 - are there any important deviations from the overall pattern?
- **Center**
 - Is the center close to the true value of p (if your distribution is of \hat{p}) or μ (if your distribution is of \bar{x})?
- **Spread**
 - How far apart are your the values of your distribution?
 - Can you use σ to describe it?

Bias of a statistic:

Rather than speaking of bias in a sampling method, we can talk about it more precisely using numbers.

Bias occurs when the mean of a statistic (mean \bar{x} or mean \hat{p}) is not the same as the true value of the parameter being estimated.

A statistic is **unbiased** if the mean $\bar{x} = \mu$
or the mean $\hat{p} = p$

Variability of a statistic:

- Described by the spread of its sampling distribution.
- Spread is determined by sampling design and size of sample
- Larger samples give smaller spread
- If population is at least 10 times as large as the sample spread of sampling distribution is about the same for any pop. size.

Bias and Variability

Four combinations:

small bias and small variability (the ideal)

samples are accurate and large enough to be consistently close to the μ or p

small bias and large variability

samples are accurate, but could be larger to decrease variability

large bias and small variability

mean values from samples are not close to the real μ or p

small variability could mean representing only a portion of the population

Poor design in selection method or assignment to treatment

large bias and large variability (the worst possible)

samples are not accurate and and yet, they represent a wide range of the population.

This is a poorly designed experiment or survey.

Samples are probably too small and assignment to treatment is flawed.