Introduction to Statistical Modelling

STAT2507D

Chapter 8 - I

Large Sample Estimation

REVIEW

- Use of descriptive statistics, both graphical and numerical, to describe and interpret sets of measurements (chapter 1-3)
- Probability and probability distributions – the basic tools used to describe populations of measurements (chapter 4-6)
- Link between probability and statistical inference (chapter 7)

REVIEW

- Populations are described by their probability distributions and parameters
  - For quantitative populations, the location and shape are described by μ and σ
  - For binomial populations, the location and shape are determined by p
- If the values of the parameters are unknown, we make inferences about them using sample information

INTRODUCTION

- Many Statistics are either sums or averages calculated from sample measurements.
- The CLT states that, even if the sampled population are not normal, the sampling distribution of these statistics will be approximately normal when sample size n is large

INTRODUCTION

- These statistics are the tools you use for inferential statistics – making inferences about a population using information contained in a sample
- Statistical inference is concerned with making decisions or predictions about parameters – numerical descriptive measures that characterize the population

WHERE WE'RE GOING: STATISTICAL INFERENCE

- Methods of making inferences about population fall into one of two categories:
  - Estimation: estimating or predicting the value of the parameter
    - "What are the most likely value of μ or p?"
  - Hypothesis testing: Making a decision about the value of a parameter based on some preconceived idea
    - "Did the sample come from a population with $\mu = 5$ or $p = 0.2$?"

EXAMPLES

- A consumer wants to estimate the average price of similar homes in her city before putting her home on the market
  - Estimation: Estimate μ, the average home price
- A manufacturer wants to know if a new type of steel is more resistant to high temperatures than an old type was
  - Hypothesis test: Is the new average resistance μ_N , equal to the old average resistance, μ_o? i.e. $\mu_N = \mu_o$ or $\mu_N > \mu_0$?

TYPES OF INFERENCE AND ESTIMATORS

- Whether you are estimating parameters or testing hypotheses, statistical methods are important because they provide:
  - Method for making the inference
  - A numerical measure of the goodness or reliability of the inference

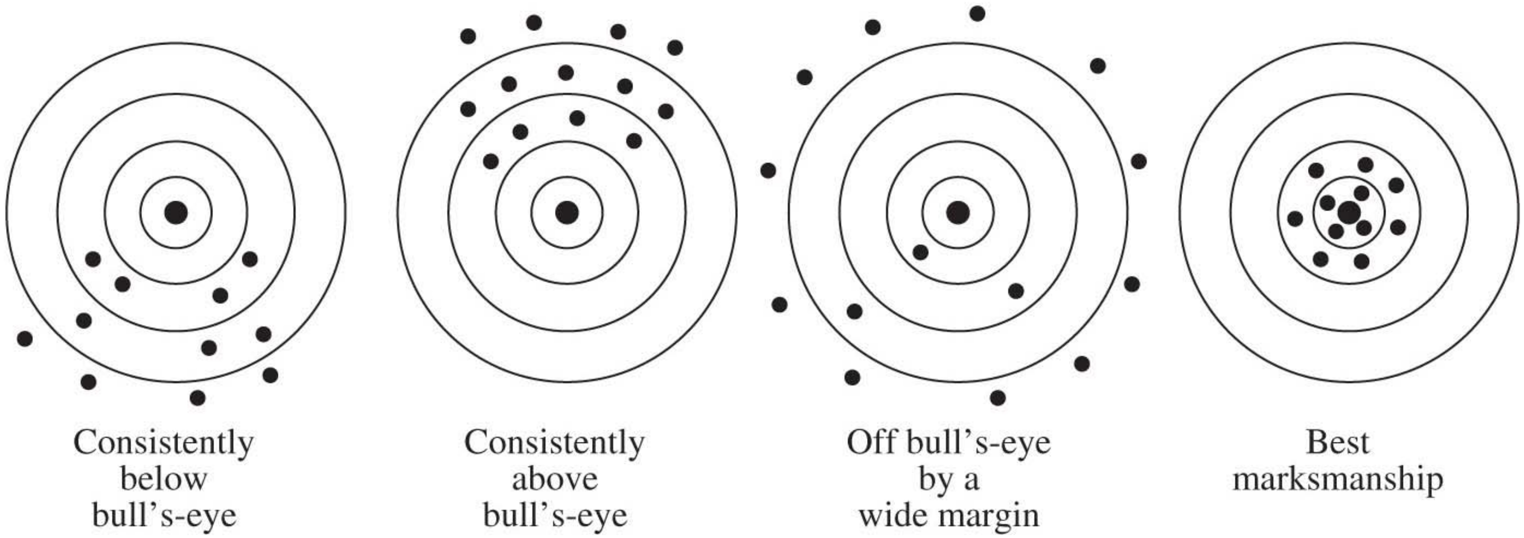
TYPES OF INFERENCE AND ESTIMATORS

- An estimator is a rule, usually a formula, that tells you how to calculate the estimate based on the sample
  - Point estimation: A single number is calculated to estimate the parameter
  - Interval estimation: two numbers are calculated to create an interval within which the parameter is expected to lie

PROPERTIES OF POINT ESTIMATORS

- Since an estimator is calculated from sample values, it varies from sample to sample according to its sampling distribution
- An estimator is unbiased if the mean of its sampling distribution equals the parameter of interest. Otherwise, the estimator is said to be biased.
  - It does not systematically overestimate or underestimate the target parameter. Such an estimator is said to be unbiased

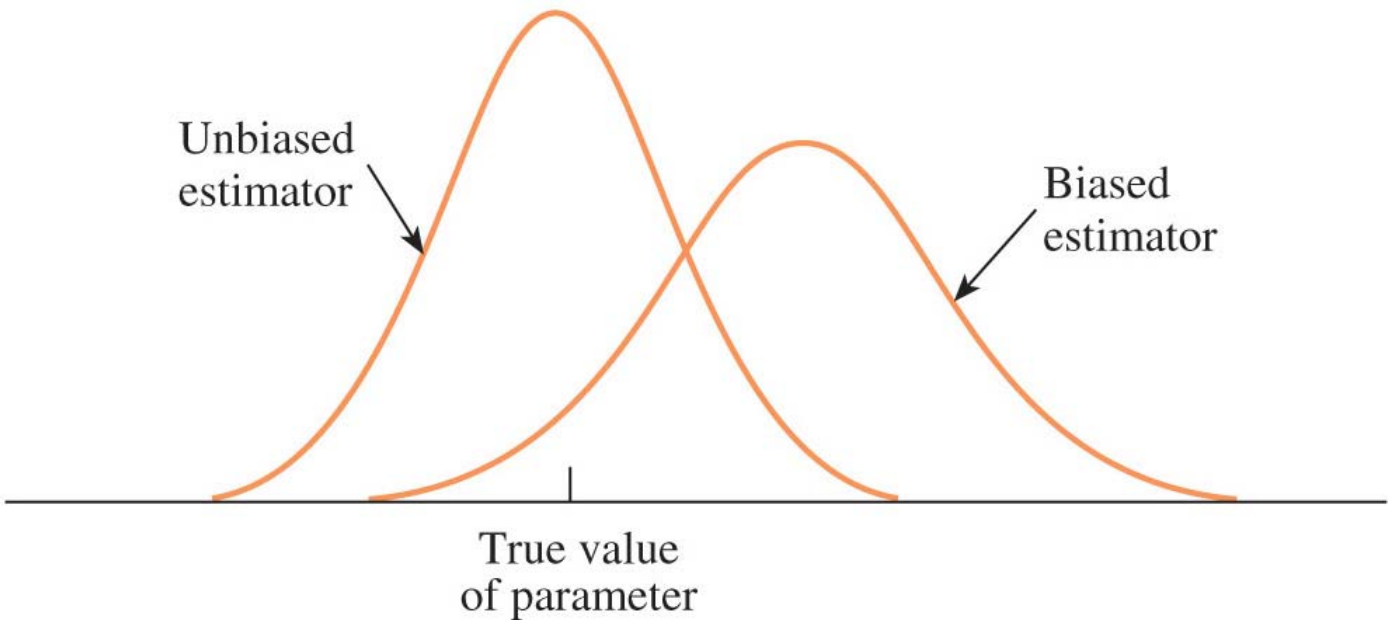
PROPERTIES OF POINT ESTIMATORS



PROPERTIES OF POINT ESTIMATORS

- Sampling distribution for an unbiased estimator and biased estimator are shown below. The sampling distribution for biased estimator is shifted to the right of the true value of the parameter. This biased estimator is more likely than an unbiased one to overestimate the value of the parameter.

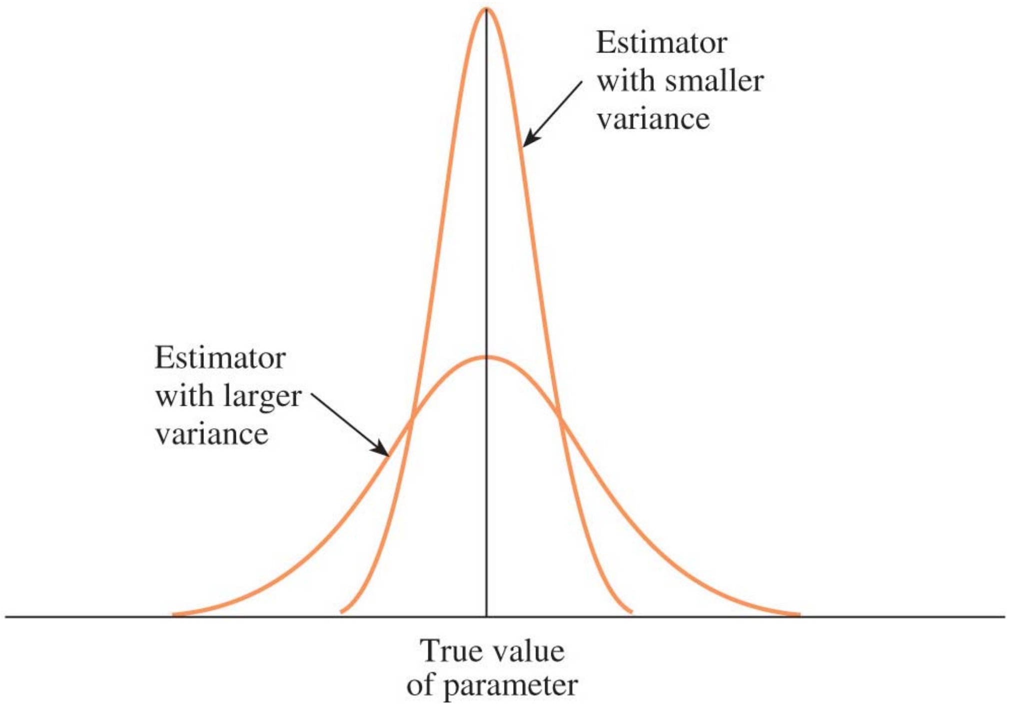
Properties of Point Estimators



Properties of Point Estimators

- The second desirable characteristics of an estimator is that the spread (as measured by the variance) of the sampling distribution should be as small as possible.

PROPERTIES OF POINT ESTIMATORS



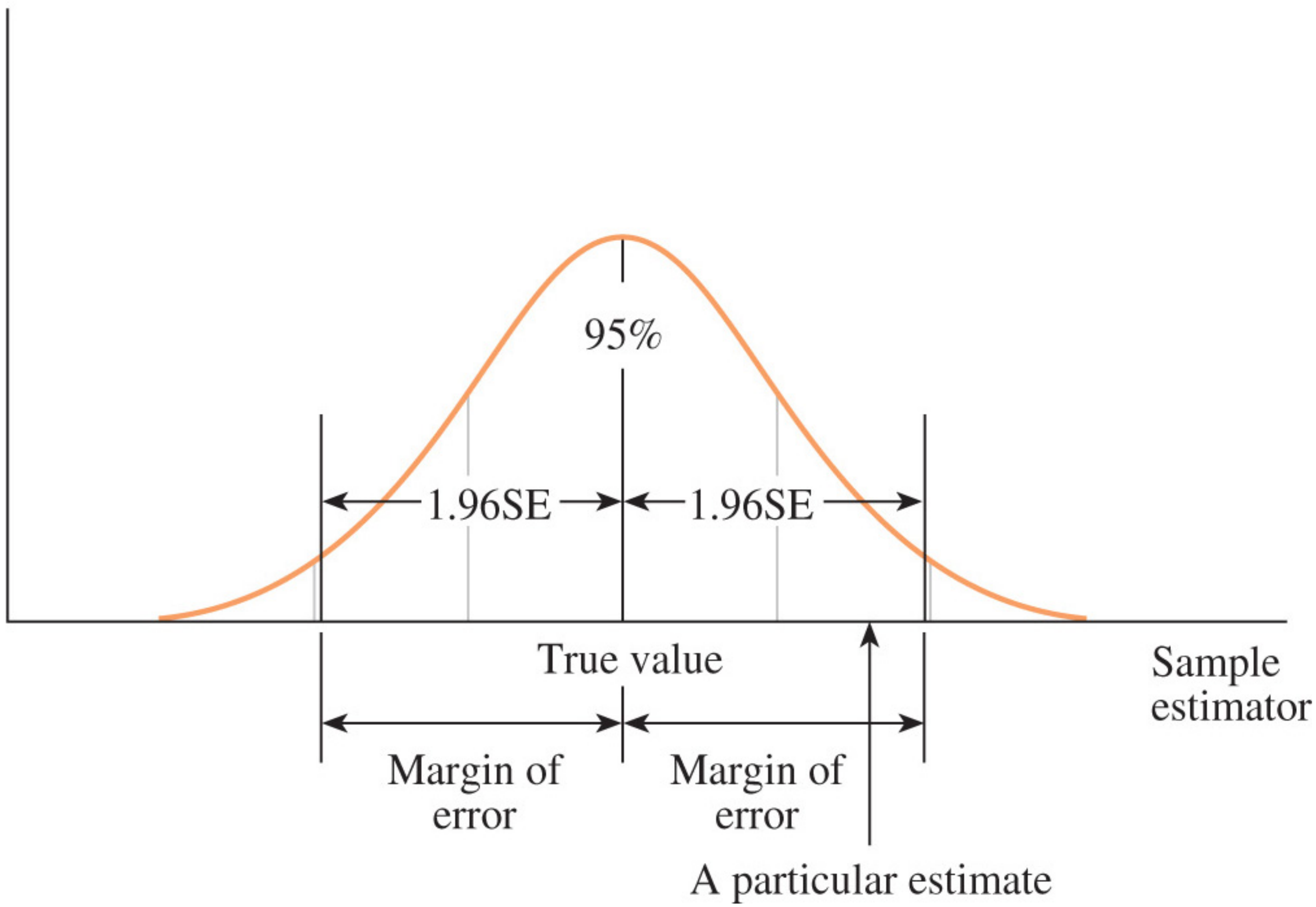
MEASURING THE GOODNESS OF AN ESTIMATOR

- The distance between an estimate and the true value of the parameter is the error of estimation
  - In the previous example, it is the distance between bullet and the bull's eye.
- In this chapter, the sample sizes are large. So, the estimators are unbiased and will have normal distributions (due to Central Limit Theorem)

The Margin of Error

- For unbiased estimators with normal sampling distributions, 95% of all point estimates will lie within 1.96 standard deviations of the parameter of interest
- Margin of Error (95%): the maximum error of estimation, calculated as 1.96 standard errors of the estimator

Margin of Error



ESTIMATING MEANS AND PROPORTIONS

- For a quantitative population:
  - Point estimator of population mean μ: $\bar{x}$
  - Margin of error (n ≥ 30): $\pm 1.96\frac{s}{\sqrt{n}}$
- For a binomial population:
  - Point estimator of population proportion p: $\hat{p} = \frac{x}{n}$
  - Margin of error ($n\hat{p} > 5\ and\ n\hat{q} > 5$): $\pm\ 1.96\ \sqrt{\frac{\hat{p}\hat{q}}{n}}$

EXERCISE

8.1) Calculate the margin of error in estimating a population mean μ for these values:

a) $n = 30, \sigma^2 = 0.2$

b) $n = 30, \sigma^2 = 0.9$

c) $n = 30, \sigma^2 = 1.5$

d) What effect does a larger population variance have on the margin of error?

EXERCISE

8.2) Calculate the margin of error in estimating a population mean μ for these values.

a) $n = 50, s^2 = 4$

b) $n = 500, s^2 = 4$

c) $n = 5000, s^2 = 4$

d) What effect does an increased sample size have on the margin of error?

EXERCISE

8.3) Suppose a sample of 75 one-square-metre plots, randomly chosen in North America's boreal (northern) forests, produced a mean biomass of 4.2 kilograms per square metre (kg/m^2), with a standard deviation of $1.5 kg/m^2$. Estimate the average biomass for the boreal forests of North America and find the margin of error for your estimate.

EXERCISE

8.4) Do our children spend as much time enjoying the outdoors and playing with family and friends as previous generations did? Or are our children spending more and more time glued to the television, computer, and other multimedia equipment? A random sample of 250 children between the ages of 8 and 18 showed that 170 children had a TV in their bedroom and that 120 of them had a video game console in their bedroom.

EXERCISE CONT'D

a) Estimate the proportion of all 8- to 18- year –olds who have a TV in their bedroom, and calculate the margin of error for your estimate

b) Estimate the proportion of all 8- to 18-year-olds who have a video game console in their bedroom, and calculate the margin of error for your estimate.