7 Sampling Distributions and Point Estimation of Parameters

CHAPTER OUTLINE

LEARNING OBJECTIVES

After careful study of this chapter you should be able to do the following:

1. Explain the general concepts of estimating the parameters of a population or a probability distribution
2. Explain the important role of the normal distribution as a sampling distribution
3. Understand the central limit theorem
4. Explain important properties of point estimators, including bias, variance, and mean square error
5. Know how to construct point estimators using the method of moments and the method of maximum likelihood
6. Know how to compute and explain the precision with which a parameter is estimated
7. Know how to construct a point estimator using the Bayesian approach

7-1 Introduction

- The field of statistical inference consists of those methods used to make decisions or to draw conclusions about a **population.**

- These methods utilize the information contained in a **sample** from the population in drawing conclusions.

- Statistical inference may be divided into two major areas:
  - Parameter estimation
  - Hypothesis testing

7-1 Introduction

Suppose that we want to obtain a point estimate of a population parameter. We know that before the data is collected, the observations are considered to be random variables, say $X_1, X_2, \ldots, X_n$. Therefore, any function of the observation, or any **statistic,** is also a random variable. For example, the sample mean $\overline{X}$ and the sample variance S^2 are statistics and they are also random variables.

Since a statistic is a random variable, it has a probability distribution. We call the probability distribution of a statistic a **sampling distribution.** The notion of a sampling distribution is very important and will be discussed and illustrated later in the chapter.

Definition

A **point estimate** of some population parameter θ is a single numerical value $\hat{\theta}$ of a statistic $\hat{\Theta}$. The statistic $\hat{\Theta}$ is called the **point estimator**.

7-1 Introduction

Estimation problems occur frequently in engineering. We often need to estimate

- The mean μ of a single population
- The variance σ^2 (or standard deviation σ) of a single population
- The proportion p of items in a population that belong to a class of interest
- The difference in means of two populations, $\mu_1 - \mu_2$
- The difference in two population proportions, $p_1 - p_2$

7-1 Introduction

Reasonable point estimates of these parameters are as follows:

- For μ, the estimate is $\hat{\mu} = \bar{x}$, the sample mean.
- For σ^2, the estimate is $\hat{\sigma}^2 = s^2$, the sample variance.
- For p, the estimate is $\hat{p} = x/n$, the sample proportion, where x is the number of items in a random sample of size n that belong to the class of interest.
- For $\mu_1 - \mu_2$, the estimate is $\hat{\mu}_1 - \hat{\mu}_2 = \bar{x}_1 - \bar{x}_2$, the difference between the sample means of two independent random samples.
- For $p_1 - p_2$, the estimate is $\hat{p}_1 - \hat{p}_2$, the difference between two sample proportions computed from two independent random samples.

7-1 Introduction

Example

There are two ponds containing lots of fish, a random sample of 20 fish were selected from each pond and record their weight. The results are as follows:

S1: 1.2, 3.0, 2.3, 1.0; 1.9; 2.1; 1.4; 2.2; 0.7; 1.3; 0.5; 0.8; 2.3; 3.3; 4.1; 3.5; 2.7; 1.3; 3.0; 1.4.

S2: 1.0, 2.3, 1.3; 1.5; 0.3; 1.6; 2.3; 2.6; 3.3; 4.2; 0.8; 2.8; 3.7; 0.5; 4.1; 3.3; 2.1; 3.6; 1.8; 2.1

1) Estimated average weight, variance, standard deviation and standard rate of fish (weight> 2 kg) in each pond.

2) Compare the weight average, variance, standard deviation of the number of fish in two ponds and the standard rate of fish in two ponds.

7.2 Sampling Distributions and the Central Limit Theorem

Statistical inference is concerned with making decisions about a population based on the information contained in a random sample from that population.

Definitions:

The random variables $X_1, X_2, \ldots, X_n$ are a **random sample** of size n if (a) the X_i's are independent random variables, and (b) every X_i has the same probability distribution.

A **statistic** is any function of the observations in a random sample.

The probability distribution of a statistic is called a **sampling distribution.**

DREAM OF INNOVATION

7.2 Sampling Distributions and the Central Limit Theorem

If we are sampling from a population that has an unknown probability distribution, the sampling distribution of the sample mean will still be approximately normal with mean μ and variance σ^2/n, if the sample size n is large. This is one of the most useful theorems in statistics, called the **central limit theorem.** The statement is as follows:

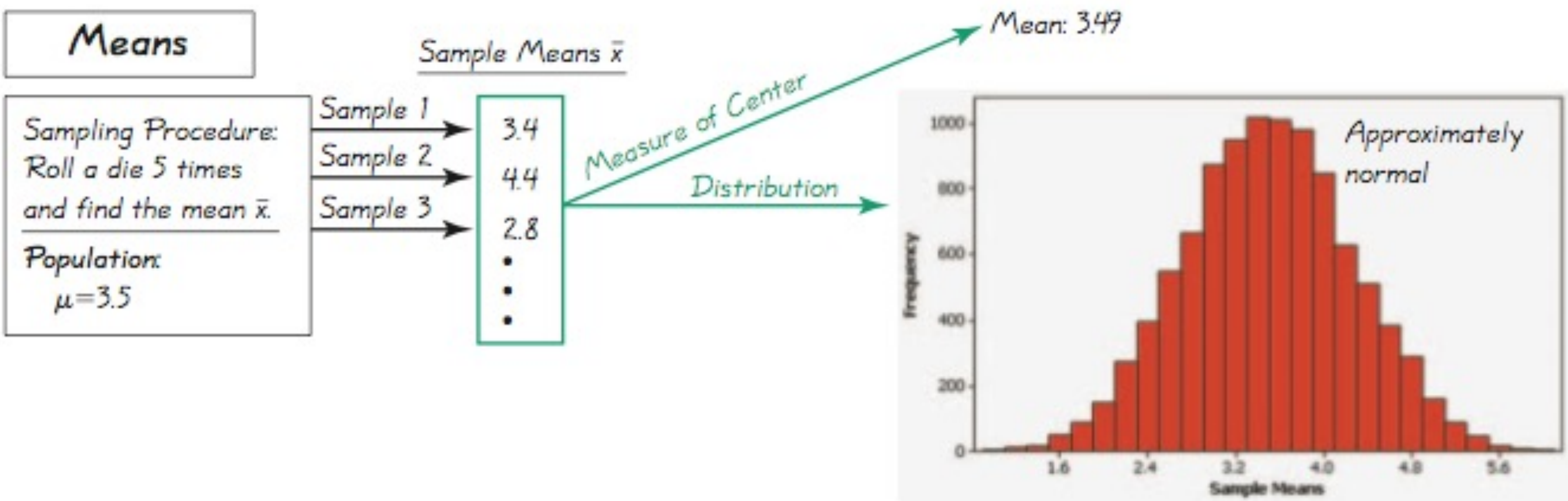
If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population (either finite or infinite) with mean μ and finite variance σ^2, and if $\overline{X}$ is the sample mean, the limiting form of the distribution of

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \qquad (7\text{-}1)$$

as $n \to \infty$, is the standard normal distribution.

Example - Sampling Distributions

Specific results from 10,000 trials



All outcomes are equally likely so the population mean is 3.5; the mean of the 10,000 trials is 3.49. If continued indefinitely, the sample mean will be 3.5. Also, notice the distribution is “normal.”

7.2 Sampling Distributions and the Central Limit Theorem

Example 7-1

An electronics company manufactures resistors that have a mean resistance of 100 ohms and a standard deviation of 10 ohms. The distribution of resistance is normal. Find the probability that a random sample of $n = 25$ resistors will have an average resistance less than 95 ohms.

Note that the sampling distribution of $\overline{X}$ is normal, with mean $\mu_{\overline{X}} = 100$ ohms and a standard deviation of

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

Therefore, the desired probability corresponds to the shaded area in Fig. 7-1. Standardizing the point $\overline{X} = 95$ in Fig. 7-2, we find that

$$z = \frac{95 - 100}{2} = -2.5$$

and therefore,

$$\begin{aligned} P(\overline{X} < 95) &= P(Z < -2.5) \\ &= 0.0062 \end{aligned}$$

7.2 Sampling Distributions and the Central Limit Theorem

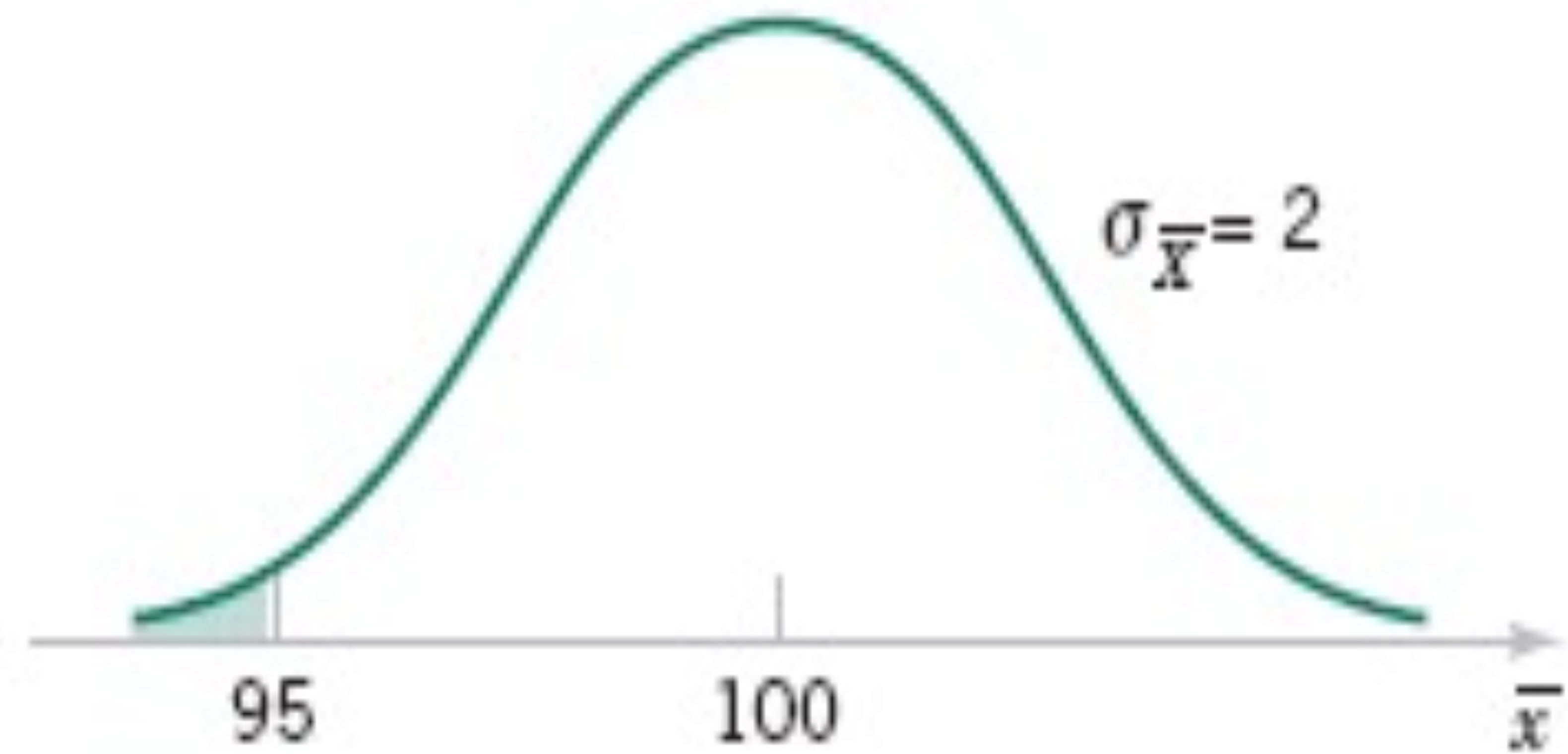


Figure 7-2 Probability for Example 7-1

7.2 Sampling Distributions and the Central Limit Theorem

Approximate Sampling Distribution of a Difference in Sample Means

If we have two independent populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 and if $\overline{X}_1$ and $\overline{X}_2$ are the sample means of two independent random samples of sizes n_1 and n_2 from these populations, then the sampling distribution of

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \qquad (7\text{-}4)$$

is approximately standard normal, if the conditions of the central limit theorem apply. If the two populations are normal, the sampling distribution of Z is exactly standard normal.

The effective life of a component used in a jet-turbine aircraft engine is a random variable with mean 5000 hours and standard deviation 40 hours. The distribution of effective life is fairly close to a nomal distribution. The engine manufacturer introduces an improvement into the manufacturing process for this component that increases the mean life to 5050 hours and decreases the standard deviation to 30 hours. Soppose that a random sample of n_1 =16 components is selected from the "old" process and a random sample of n_2=25 components is selected from the "improve" process. What is the probability that the difference in the two sample means $\bar{X}_1 - \bar{X}_2$ is at least 25 hours? Assume that the old and improve processes can be regarded as independent population.

Example – Water Taxi Safety

Assume the population of weights of men is normally distributed with a mean of 172 lb and a standard deviation of 29 lb.

a) Find the probability that if an *individual* man is randomly selected, his weight is greater than 175 lb.

b) Find the probability that *20 randomly selected men* will have a mean weight that is greater than 175 lb (so that their total weight exceeds the safe capacity of 3500 pounds).

IMPORTANT TERMS AND CONCEPTS

- Bayes estimator
- Bias in parameter estimation
- Central limit theorem
- Estimator versus estimate
- Likelihood function
- Maximum likelihood estimator
- Mean square error of an estimator
- Minimum variance unbiased estimator
- Moment estimator
- Normal distribution as the sampling distribution of a sample mean
- Normal distribution as the sampling distribution of the difference in two sample means
- Parameter estimation
- Point estimator
- Population or distribution moments
- Posterior distribution
- Prior distribution
- Sample moments
- Sampling distribution
- Standard error and estimated standard error of an estimator
- Statistic
- Statistical inference
- Unbiased estimator