Statistical Inference

Abdelhakim Necir

Academic Year 2026–2027

Contents

1 Sampling, Sampling Distributions and Experimental Designs **5**

1.1 Introduction . . . 5
1.2 Population and Sample . . . 5
1.2.1 Population . . . 5
1.2.2 Sample . . . 6
1.3 Parameter and Statistic . . . 6
1.4 Why Do We Use a Sample? . . . 7
1.4.1 Cost . . . 7
1.4.2 Time . . . 7
1.4.3 Destructive Measurements . . . 7
1.4.4 Very Large or Infinite Populations . . . 8
1.4.5 Statistical Efficiency . . . 8
1.5 Methods of Data Collection . . . 8
1.5.1 Observational Studies . . . 8
1.5.2 Surveys . . . 8
1.5.3 Experiments . . . 8
1.6 Types of Sampling . . . 8
1.6.1 Probability Sampling . . . 8
1.6.2 Non-Probability Sampling . . . 9
1.7 Basic Principles of Experimental Design . . . 9
1.7.1 Experimental Unit . . . 9
1.7.2 Factor . . . 10
1.7.3 Treatment . . . 10
1.7.4 Response Variable . . . 10
1.7.5 Randomization . . . 10
1.7.6 Replication . . . 10
1.7.7 Control . . . 10
1.8 Sampling Distributions . . . 10
1.9 Sampling Distribution of the Sample Mean . . . 11
1.9.1 Case of a Normal Population . . . 12
1.10 Sampling Distribution of a Sample Proportion . . . 12
1.11 Sampling Distribution of the Difference of Two Means . . . 13
1.12 Sampling Distribution of the Sample Variance . . . 14
1.13 The Central Limit Theorem . . . 14
1.14 Applications of the Central Limit Theorem . . . 15
1.14.1 Approximation of the Sample Mean . . . 15
1.14.2 Approximation of a Sample Proportion . . . 16
1.15 Standard Errors . . . 16

1.16 Summary 16

2 Point Estimation **19**
2.1 Introduction to Point Estimation 19
2.2 Parametric and Nonparametric Estimation 19
2.2.1 Parametric Estimation 19
2.2.2 Nonparametric Estimation 20
2.3 Point Estimators and Estimates 20
2.4 Random Variables, Estimators, and Unknown Parameters 21
2.5 Unknown Parameters 21
2.6 Methods of Point Estimation 22
2.6.1 An Introductory Example 22
2.6.2 Maximum Likelihood Method 23
2.6.3 The Log-Likelihood Function 23
2.6.4 Likelihood Functions for Common Models 24
2.6.5 Maximum Likelihood Estimator 24
2.6.6 Method of Moments 25
2.7 Properties of Estimators 25
2.7.1 Bias 25
2.7.2 Asymptotic Bias 26
2.7.3 Variance 26
2.7.4 Mean Squared Error 26
2.7.5 Efficiency 27
2.7.6 Admissibility 27
2.7.7 Consistency 27
2.7.8 Convergence in Probability 27
2.7.9 Convergence in Mean of Order k 28
2.7.10 Sufficiency 28
2.7.11 Completeness 28
2.8 Usual Estimators 29
2.8.1 Quantitative Case 29
2.8.2 Qualitative Case: Unknown Proportion 29
2.9 Fisher Information and Score 29
2.9.1 The Score Function 30
2.9.2 Fisher Information 30
2.9.3 Efficiency and Fisher Information 30
2.9.4 Fisher Information and Completeness 31
2.10 Cramér–Rao–Fréchet–Darmois Bound 31
2.10.1 Illustrative Example 31
2.11 Remark on ESBVM 32
2.12 Conclusion 32

3 Confidence Interval Estimation **33**
3.1 Point Estimation versus Confidence Interval Estimation 33
3.2 Confidence Level and Risk Level 34
3.2.1 Interpretation of a Confidence Interval 34
3.3 Construction of Confidence Intervals 34
3.4 Confidence Interval for a Mean 35
3.4.1 Known Variance 35

3.4.2 Unknown Variance . . . 36
3.4.3 Summary for the Mean . . . 37
3.5 Confidence Interval for a Variance . . . 37
3.6 Confidence Interval for a Proportion . . . 38
3.7 Influence of Sample Size on Precision . . . 39
3.8 Determination of the Optimal Sample Size . . . 39
3.8.1 Sample Size for Estimating a Mean . . . 39
3.8.2 Sample Size for Estimating a Proportion . . . 40
3.8.3 Factors Affecting the Required Sample Size . . . 40
3.9 Conclusion . . . 41

4 Introduction to Statistical Tests 43
4.1 Objectives of Statistical Tests . . . 43
4.2 Null and Alternative Hypotheses . . . 43
4.2.1 Two-Sided Alternative . . . 44
4.2.2 Right-Sided Alternative . . . 44
4.2.3 Left-Sided Alternative . . . 44
4.3 Types of Statistical Tests . . . 45
4.4 Neyman–Pearson Principle . . . 45
4.5 Critical and Acceptance Regions . . . 46
4.6 Significance Level . . . 46
4.7 Type I and Type II Errors . . . 47
4.8 Power of a Test . . . 47
4.9 Decision Table . . . 48
4.10 General Steps of a Statistical Test . . . 48
4.11 Standard Statistical Tables . . . 49
4.11.1 Normal Distribution . . . 49
4.11.2 Student's t Distribution . . . 49
4.11.3 Chi-Square Distribution . . . 49
4.11.4 Fisher–Snedecor Distribution . . . 50
4.12 Parametric Tests . . . 50
4.13 Test for a Population Mean . . . 50
4.13.1 Known Variance . . . 50
4.13.2 Unknown Variance . . . 51
4.14 Test for a Population Variance . . . 51
4.15 Test for a Population Proportion . . . 52
4.16 Comparison of Two Population Means . . . 53
4.16.1 Known Variances . . . 53
4.16.2 Unknown Variances . . . 53
4.17 Comparison of Two Population Variances . . . 54
4.18 Comparison of Two Population Proportions . . . 54
4.19 Summary of Parametric Tests . . . 55
4.20 Conclusion . . . 55

Chapter 1

Sampling, Sampling Distributions and Experimental Designs

1.1 Introduction

Statistical inference is based on the principle that information obtained from a sample can be used to draw conclusions about a larger population.

In many practical situations, observing the whole population is impossible, too expensive, too time-consuming, or simply unnecessary. Instead, a subset of units is selected, and the information collected from this subset is used to learn about unknown characteristics of the population.

This chapter introduces the basic concepts of sampling and experimental design, and studies the sampling distributions of some commonly used statistics.

The main objectives are:

- to distinguish between a population and a sample;
- to distinguish between a parameter and a statistic;
- to understand the main methods of data collection;
- to describe the principal sampling designs;
- to introduce the basic principles of experimental design;
- to derive the sampling distributions of important statistics;
- to understand the role of the Central Limit Theorem.

1.2 Population and Sample

1.2.1 Population

Definition 1.1. A *population* is the complete collection of individuals, objects, measurements, or outcomes about which statistical information is required.

A population may be finite or infinite.

Examples include:

- all students registered at a university during a given academic year;
- all manufactured components produced by a factory during one month;
- all households in a given region;
- all possible measurements generated by a probability model.

In a probabilistic formulation, the population is described by a probability distribution, often denoted by P, or by a family

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\},$$

where θ represents an unknown parameter.

1.2.2 Sample

Definition 1.2. A *sample* is a collection of observations selected from a population.

If the sample contains n observations, we write

$$X_1, \ldots, X_n.$$

The integer n is called the *sample size*.

When the observations are independent and identically distributed according to a common probability distribution P, we write

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} P.$$

If the distribution depends on an unknown parameter θ, we may write

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} P_\theta.$$

Remark 1.1. The i.i.d. assumption is fundamental in classical statistical inference. However, in other areas such as time series, spatial statistics, survival analysis, and stochastic processes, observations may be dependent.

1.3 Parameter and Statistic

One of the most important distinctions in statistical inference is the difference between a population parameter and a sample statistic.

Definition 1.3. A *parameter* is a numerical characteristic of a population or of its probability distribution.

Typical examples are:

$$\mu = E(X),$$

the population mean,

$$\sigma^2 = \mathrm{Var}(X),$$

the population variance, and

$$p,$$

a population proportion.
Parameters are generally unknown.

Definition 1.4. A *statistic* is any measurable function of the observed sample that does not depend on unknown parameters.

For example, the sample mean is

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

the sample variance is

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

and, for Bernoulli observations, the sample proportion is

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Remark 1.2. A parameter describes the population, whereas a statistic is computed from the sample.

1.4 Why Do We Use a Sample?

A complete investigation of all units in a population is called a *census*.
Although a census may appear ideal, it is often impractical.
Sampling is commonly preferred for several reasons.

1.4.1 Cost

Collecting information from every member of a large population may require substantial financial resources.

1.4.2 Time

A census may require a long period of data collection, whereas a carefully designed sample can provide useful information much more rapidly.

1.4.3 Destructive Measurements

Some measurements destroy or alter the object being studied.
For example, a component tested until failure cannot be returned to the production process.

1.4.4 Very Large or Infinite Populations

In some situations, the population is extremely large or naturally modeled as infinite.

A complete census is then impossible.

1.4.5 Statistical Efficiency

A properly designed sample can provide sufficiently accurate information with substantially fewer observations than a complete census.

1.5 Methods of Data Collection

Statistical data can be obtained in different ways.

1.5.1 Observational Studies

In an observational study, the investigator records information without deliberately modifying the conditions of the individuals or experimental units.

Observational studies are useful when intervention is impossible, impractical, or ethically inappropriate.

1.5.2 Surveys

A survey collects information directly from individuals, usually through questionnaires, interviews, or electronic forms.

The reliability of the results depends strongly on the sampling design, response rate, and formulation of the questions.

1.5.3 Experiments

In an experiment, the investigator deliberately applies one or more conditions, called *treatments*, and observes their effects on a response variable.

Experiments are especially useful when the objective is to study causal relationships.

Remark 1.3. A statistical analysis cannot compensate for a poor data collection process. The quality of the conclusions depends on both the statistical method and the design by which the data were obtained.

1.6 Types of Sampling

Sampling methods can be broadly divided into probability sampling and non-probability sampling.

1.6.1 Probability Sampling

In probability sampling, each population unit has a known probability of being selected.

This makes it possible to quantify sampling variability.

Simple Random Sampling

Definition 1.5. A *simple random sample* of size n is selected in such a way that every possible sample of size n has the same probability of being chosen.

Simple random sampling is the basic model underlying much of classical statistical inference.

Systematic Sampling

In systematic sampling, a starting unit is selected and then units are chosen according to a fixed interval.

For example, one may select every k-th unit after a random starting point.

Stratified Sampling

The population is divided into subpopulations called *strata*.

A sample is then selected from each stratum.

Stratification is useful when the population contains important subgroups that should all be represented in the sample.

Cluster Sampling

The population is divided into natural groups called *clusters*.

A number of clusters are selected, and observations are collected from units belonging to the selected clusters.

Cluster sampling is often useful when the population is geographically dispersed.

1.6.2 Non-Probability Sampling

In non-probability sampling, the selection probabilities are not controlled by a random mechanism.

Convenience Sampling

A convenience sample consists of observations that are easy to obtain.

Such samples may be strongly biased and are generally unsuitable for formal statistical inference about a population.

Remark 1.4. Probability sampling methods are generally preferred when the objective is to make inferential statements about a population.

1.7 Basic Principles of Experimental Design

An experiment is designed to study the effect of one or more factors on a response variable.

Several concepts are fundamental.

1.7.1 Experimental Unit

An *experimental unit* is the smallest unit to which a treatment is applied.

1.7.2 Factor

A *factor* is an explanatory variable whose effect on the response is being investigated.

The possible values or categories of a factor are called its *levels*.

1.7.3 Treatment

A *treatment* is a specific experimental condition applied to an experimental unit.

1.7.4 Response Variable

The *response variable* is the outcome measured after applying a treatment.

1.7.5 Randomization

Randomization consists of assigning treatments to experimental units by a random mechanism.

Its main purpose is to reduce systematic bias and to help create comparable groups.

1.7.6 Replication

Replication means applying each treatment to several experimental units.

Replication makes it possible to estimate natural variability and generally improves the precision of statistical conclusions.

1.7.7 Control

A control group provides a reference condition against which the effect of a treatment can be compared.

Remark 1.5. Randomization, replication, and control are three basic principles of sound experimental design.

1.8 Sampling Distributions

A statistic is itself a random variable because its value changes from one sample to another.

Definition 1.6. The probability distribution of a statistic is called its *sampling distribution.*

Sampling distributions are fundamental because they describe the variability that results from random sampling.

They provide the theoretical basis for estimation, confidence intervals, and hypothesis testing.

1.9 Sampling Distribution of the Sample Mean

Let

$$X_1, \ldots, X_n$$

be independent and identically distributed random variables such that

$$E(X_i) = \mu$$

and

$$\mathrm{Var}(X_i) = \sigma^2.$$

The sample mean is defined by

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Its expectation is

$$E(\overline{X}) = E\left(\frac{1}{n} \sum_{i=1}^{n} X_i \right).$$

Using linearity of expectation,

$$E(\overline{X}) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \mu.$$

Therefore,

$$E(\overline{X}) = \mu.$$

Its variance is

$$\mathrm{Var}(\overline{X}) = \mathrm{Var}\left(\frac{1}{n} \sum_{i=1}^{n} X_i \right).$$

By independence,

$$\mathrm{Var}(\overline{X}) = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

Hence,

$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}.$$

The standard deviation of the sampling distribution of $\overline{X}$ is

$$\frac{\sigma}{\sqrt{n}}.$$

This quantity is called the *standard error of the sample mean.*

1.9.1 Case of a Normal Population

If

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2),$$

then

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Equivalently,

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

This result is exact for every sample size n when the parent population is normal.

1.10 Sampling Distribution of a Sample Proportion

Suppose that

$$X_1, \ldots, X_n$$

are independent Bernoulli random variables with

$$P(X_i = 1) = p$$

and

$$P(X_i = 0) = 1 - p.$$

Then

$$X_i \sim \text{Bernoulli}(p).$$

The sample proportion is

$$\widehat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Since

$$E(X_i) = p,$$

we obtain

$$E(\widehat{p}) = p.$$

Also,

$$\text{Var}(X_i) = p(1 - p),$$

so

$$\text{Var}(\widehat{p}) = \frac{p(1-p)}{n}.$$

Thus,

$$E(\widehat{p}) = p$$

and

$$\text{Var}(\widehat{p}) = \frac{p(1-p)}{n}.$$

For sufficiently large n, the sampling distribution of $\widehat{p}$ is approximately normal:

$$\widehat{p} \approx \mathcal{N}\left(p, \frac{p(1-p)}{n}\right).$$

Therefore,

$$\frac{\widehat{p} - p}{\sqrt{p(1-p)/n}} \approx \mathcal{N}(0,1).$$

A commonly used practical rule for the normal approximation is that both

$$np$$

and

$$n(1-p)$$

should be sufficiently large.

A frequent rule of thumb is

$$np \geq 5 \qquad \text{and} \qquad n(1-p) \geq 5.$$

Remark 1.6. The conditions above are practical guidelines rather than exact mathematical requirements. Some textbooks use the more conservative threshold 10.

1.11 Sampling Distribution of the Difference of Two Means

Consider two independent populations.

Let

$$X_1, \ldots, X_{n_1}$$

be an i.i.d. sample from a population with mean μ_1 and variance σ_1^2, and let

$$Y_1, \ldots, Y_{n_2}$$

be an independent i.i.d. sample from a second population with mean μ_2 and variance σ_2^2.

Define

$$\overline{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i$$

and

$$\overline{Y} = \frac{1}{n_2} \sum_{j=1}^{n_2} Y_j.$$

Then

$$E(\overline{X} - \overline{Y}) = \mu_1 - \mu_2.$$

Because the two samples are independent,

$$\mathrm{Var}(\overline{X} - \overline{Y}) = \mathrm{Var}(\overline{X}) + \mathrm{Var}(\overline{Y}).$$

Therefore,

$$\mathrm{Var}(\overline{X} - \overline{Y}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Thus,

$$E(\overline{X} - \overline{Y}) = \mu_1 - \mu_2$$

and

$$\text{Var}(\overline{X} - \overline{Y}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

If both populations are normal, then

$$\overline{X} - \overline{Y} \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right).$$

Consequently,

$$Z = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim \mathcal{N}(0, 1).$$

1.12 Sampling Distribution of the Sample Variance

Let

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2).$$

The sample variance is

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

The following result is fundamental.

Theorem 1.1. *If*

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2),$$

then

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

The number

$$n - 1$$

is called the number of *degrees of freedom*.

Remark 1.7. The chi-square distribution above is an exact result under the assumption that the parent population is normal.

This result will later be used to construct confidence intervals and hypothesis tests concerning the population variance.

1.13 The Central Limit Theorem

The Central Limit Theorem is one of the most important results in probability and statistics.

It explains why the normal distribution appears so frequently in statistical inference.

Theorem 1.2 (Central Limit Theorem)**.** *Let*

$$X_1, X_2, \ldots$$

be independent and identically distributed random variables such that

$$E(X_i) = \mu$$

and

$$\mathrm{Var}(X_i) = \sigma^2 < \infty.$$

Then

$$\frac{\sqrt{n}(\overline{X} - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0,1), qquad n \to \infty.$$

Equivalently,

$$\overline{X} \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

for sufficiently large n.

The importance of this theorem is that the original population does not need to be normally distributed.

The approximation improves as the sample size increases, although the rate of improvement depends on the shape of the underlying distribution.

1.14 Applications of the Central Limit Theorem

1.14.1 Approximation of the Sample Mean

For large n,

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \approx \mathcal{N}(0,1).$$

This approximation allows probabilities involving the sample mean to be evaluated using the standard normal distribution.

Example 1.1. Suppose a population has mean

$$\mu = 50$$

and standard deviation

$$\sigma = 10.$$

For a random sample of size

$$n = 100,$$

the sample mean has approximately the distribution

$$\overline{X} \approx \mathcal{N}\left(50, \frac{10^2}{100}\right).$$

Thus,

$$E(\overline{X}) = 50$$

and

$$\mathrm{SD}(\overline{X}) = 1.$$

For example,

$$P(48 \le \overline{X} \le 52)$$

can be written as

$$P\left(-2 \le \frac{\overline{X} - 50}{1} \le 2\right),$$

and therefore

$$P(48 \leq \overline{X} \leq 52) \approx P(-2 \leq Z \leq 2),$$

where

$$Z \sim \mathcal{N}(0, 1).$$

1.14.2 Approximation of a Sample Proportion

Since the sample proportion is the mean of Bernoulli random variables, the Central Limit Theorem gives

$$\widehat{p} \approx \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

for sufficiently large sample size.

This approximation will later be used in the construction of confidence intervals and hypothesis tests concerning population proportions.

1.15 Standard Errors

The standard deviation of the sampling distribution of a statistic is called its *standard error*.

For the sample mean,

$$\text{SE}(\overline{X}) = \frac{\sigma}{\sqrt{n}}.$$

For the sample proportion,

$$\text{SE}(\widehat{p}) = \sqrt{\frac{p(1-p)}{n}}.$$

For the difference between two independent sample means,

$$\text{SE}(\overline{X} - \overline{Y}) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

When the population quantities involved are unknown, the standard error is usually estimated from the sample.

An important consequence is that increasing the sample size reduces the standard error.

For example,

$$\text{SE}(\overline{X}) = \frac{\sigma}{\sqrt{n}},$$

so multiplying the sample size by 4 divides the standard error by 2.

Thus, larger samples generally provide more precise statistical information.

1.16 Summary

The central framework of this chapter is based on the distinction between population quantities and sample quantities.

Population characteristics include parameters such as

$$\mu, \qquad \sigma^2, \qquad p,$$

whereas samples produce statistics such as

$$\overline{X}, \qquad S^2, \qquad \widehat{p}.$$

Sampling provides a practical way to obtain information about a population, provided that the sampling design is appropriate.

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

For the sample mean,

$$E(\overline{X}) = \mu,$$

and

$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}.$$

For a sample proportion,

$$E(\widehat{p}) = p,$$

and

$$\mathrm{Var}(\widehat{p}) = \frac{p(1-p)}{n}.$$

For two independent sample means,

$$E(\overline{X} - \overline{Y}) = \mu_1 - \mu_2,$$

and

$$\mathrm{Var}(\overline{X} - \overline{Y}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

For a normal population,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Finally, the Central Limit Theorem provides the asymptotic approximation

$$\frac{\sqrt{n}(\overline{X} - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0,1).$$

These results provide the foundation for point estimation, confidence intervals, and hypothesis testing developed in the following chapters.

Chapter 2

Point Estimation

2.1 Introduction to Point Estimation

In statistical inference, one of the main objectives is to learn about unknown characteristics of a population from a sample. When the population distribution is described by a model involving unknown parameters, these parameters must be estimated from the available observations.

Point estimation is concerned with assigning a single numerical value to an unknown parameter on the basis of a random sample.

Let

$$X_1, \ldots, X_n$$

be a random sample from a population whose distribution depends on an unknown parameter θ. A point estimator is a statistic constructed from the sample and used to estimate θ.

Point estimation has two complementary aspects. The first is the construction of estimators, while the second is the study of their properties. Two important methods for constructing estimators are the Maximum Likelihood Method and the Method of Moments. Once an estimator has been obtained, its bias, variance, efficiency, consistency, sufficiency, completeness, and other properties can be investigated.

Point estimation should be distinguished from confidence interval estimation. A point estimator provides a single value, whereas a confidence interval provides a range of values associated with a specified confidence level. Confidence interval estimation will be studied in Chapter 3.

2.2 Parametric and Nonparametric Estimation

Statistical estimation may be considered within either a parametric or a nonparametric framework.

2.2.1 Parametric Estimation

In parametric estimation, the probability distribution of the population is assumed to belong to a specified family of distributions characterized by a finite-dimensional parameter.

A parametric statistical model can be represented as

$$\mathcal{P} = \{P_\theta : \theta \in \Theta\},$$

where Θ is the parameter space and θ denotes the unknown parameter.

For example, suppose that

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2).$$

If σ^2 is known and μ is unknown, the parameter to be estimated is $\theta = \mu$. If both μ and σ^2 are unknown, the parameter is the vector

$$\theta = (\mu, \sigma^2).$$

The advantage of the parametric approach is that the assumed model provides a mathematical structure that can be used to construct and study estimators.

2.2.2 Nonparametric Estimation

In nonparametric estimation, the distribution is not restricted to a finite-dimensional parametric family of the type considered above.

The object of estimation may be, for example, a distribution function, a density function, or another characteristic of the population without specifying a particular finite-dimensional distributional model.

Thus, the main distinction is that parametric estimation is based on a finite-dimensional parameterization of the statistical model, whereas nonparametric estimation does not impose such a restriction.

The present chapter focuses mainly on point estimation in parametric models.

2.3 Point Estimators and Estimates

Let

$$X_1, \ldots, X_n$$

be a random sample and let θ be an unknown parameter.

A **point estimator** of θ is a statistic

$$T = T(X_1, \ldots, X_n)$$

constructed from the random sample with the purpose of estimating θ.

Since T is a function of random variables, it is itself a random variable. Its value is therefore not known before the sample is observed.

After the observations

$$x_1, \ldots, x_n$$

have been obtained, the estimator takes the numerical value

$$t = T(x_1, \ldots, x_n).$$

This numerical value is called the **estimate** of θ.

For example, suppose that μ is the unknown mean of a population. The sample mean

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

is an estimator of μ.

Once the observations

$$x_1, \ldots, x_n$$

are available, the corresponding estimate is

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Therefore, the estimator and the estimate are different objects. The former is random because it depends on the random sample, whereas the latter is a fixed numerical value after the observations have been collected.

2.4 Random Variables, Estimators, and Unknown Parameters

A clear distinction between random variables, estimators, and parameters is essential in statistical inference.

The observations

$$X_1, \ldots, X_n$$

are random variables before the experiment is performed. Their observed values

$$x_1, \ldots, x_n$$

are fixed numbers.

An estimator

$$T(X_1, \ldots, X_n)$$

is also a random variable because it is calculated from the random sample.

In the classical frequentist framework, the parameter θ is regarded as a fixed but unknown quantity. It is not a random variable.

Consequently, before observing the sample,

$$X_1, \ldots, X_n \quad \text{and} \quad T(X_1, \ldots, X_n)$$

are random, while θ is fixed but unknown.

After the observations have been collected, the sample values and the resulting estimate are fixed numerical quantities.

This distinction becomes particularly important when defining the bias, variance, and other properties of an estimator.

2.5 Unknown Parameters

Consider a statistical model with density or probability mass function

$$f(x; \theta), \qquad \theta \in \Theta.$$

The parameter θ describes a characteristic of the population distribution. Depending on the model, it may represent a mean, a variance, a probability, a rate, or several parameters simultaneously.

Suppose that

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} f(x; \theta).$$

The parameter θ is unknown, while the observations are available. Point estimation seeks to construct a statistic

$$T(X_1, \ldots, X_n)$$

that provides useful information about θ.

The quality of an estimator cannot generally be determined from one observed sample alone. Its behavior under repeated sampling must be considered. This leads to the study of the properties of estimators.

2.6 Methods of Point Estimation

Two fundamental methods of point estimation considered in this course are:

1. the Maximum Likelihood Method;

2. the Method of Moments.

The Maximum Likelihood Method constructs an estimator by maximizing the likelihood of the observed data as a function of the unknown parameter.

The Method of Moments constructs estimators by equating sample moments to the corresponding theoretical moments.

2.6.1 An Introductory Example

Suppose that

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p),$$

where p is unknown.

For each observation,

$$P(X_i = x_i) = p^{x_i}(1-p)^{1-x_i}, \qquad x_i \in \{0, 1\}.$$

A natural estimator of p is the proportion of observations equal to one:

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}.$$

This example illustrates the fundamental idea of point estimation: an unknown population parameter is replaced by a statistic calculated from the sample.

The same estimator can also be obtained systematically by both the Maximum Likelihood Method and the Method of Moments.

2.6.2 Maximum Likelihood Method

Let

$$X_1, \ldots, X_n$$

be independent observations from a model with density or probability mass function

$$f(x;\theta), \qquad \theta \in \Theta.$$

After observing

$$x_1, \ldots, x_n,$$

the likelihood function is defined by

$$L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i;\theta),$$

when the observations are independent.

The likelihood is regarded as a function of θ, with the observed data kept fixed.

A maximum likelihood estimator is a value of θ that maximizes the likelihood over the parameter space:

$$\widehat{\theta}_{\mathrm{MLE}} \in \arg\max_{\theta\in\Theta} L(\theta; x_1, \ldots, x_n).$$

The maximum may occur at an interior point of the parameter space or at its boundary. Therefore, solving a likelihood equation is not by itself sufficient; the resulting candidate must be checked against the parameter space and the behavior of the likelihood.

2.6.3 The Log-Likelihood Function

Because products can be difficult to manipulate, it is often more convenient to work with the logarithm of the likelihood.

The log-likelihood function is

$$\ell(\theta) = \log L(\theta; x_1, \ldots, x_n).$$

Since the logarithm is strictly increasing,

$$\arg\max_{\theta\in\Theta} L(\theta) = \arg\max_{\theta\in\Theta} \ell(\theta).$$

Thus, maximizing the log-likelihood gives the same maximum likelihood estimator while usually simplifying differentiation and algebra.

For an interior maximum, a candidate estimator can often be obtained by solving

$$\frac{\partial}{\partial\theta}\ell(\theta) = 0.$$

The second derivative or another appropriate argument is then used to verify that the candidate corresponds to a maximum.

2.6.4 Likelihood Functions for Common Models

The form of the likelihood depends on the statistical model.

For a Bernoulli sample,

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p),$$

the likelihood is

$$L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}.$$

Therefore,

$$L(p) = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}.$$

The log-likelihood is

$$\ell(p) = \left(\sum_{i=1}^{n} x_i\right)\log p + \left(n - \sum_{i=1}^{n} x_i\right)\log(1-p).$$

Differentiating gives

$$\ell'(p) = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1-p}.$$

Solving $\ell'(p) = 0$ gives

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}.$$

Hence the maximum likelihood estimator is

$$\widehat{p}_{\text{MLE}} = \overline{X}.$$

Other common models, such as the normal, exponential, Poisson, and uniform models, lead to likelihood functions of different forms. The general principle remains the same: construct the likelihood from the statistical model and maximize it with respect to the unknown parameter.

2.6.5 Maximum Likelihood Estimator

The maximum likelihood estimator is therefore obtained by

$$\widehat{\theta}_{\text{MLE}} \in \arg\max_{\theta \in \Theta} L(\theta).$$

When the likelihood has a unique maximum, the MLE is uniquely determined.

Maximum likelihood estimation is widely used because it provides a general method that can be applied to a large class of statistical models.

However, maximum likelihood estimators are not automatically unbiased, efficient, or consistent in every problem. These properties must be studied for the particular model under consideration.

2.6.6 Method of Moments

The Method of Moments is another general procedure for constructing estimators.

Let the theoretical moments of a random variable X be

$$\mu_r'(\theta) = \mathbb{E}_\theta[X^r], \qquad r = 1, 2, \ldots$$

The corresponding sample moments are

$$m_r' = \frac{1}{n}\sum_{i=1}^{n} X_i^r.$$

The Method of Moments consists in equating theoretical moments to sample moments.

For a model containing one unknown parameter, the first moment equation is typically

$$\mu_1'(\theta) = m_1'.$$

Solving this equation gives the Method of Moments estimator.

If the parameter is multidimensional, several moment equations can be used. For a parameter vector

$$\theta = (\theta_1, \ldots, \theta_k),$$

one may use the first k moment equations:

$$\mu_r'(\theta) = m_r', \qquad r = 1, \ldots, k.$$

The resulting system is solved for the unknown parameters.

The Method of Moments and the Maximum Likelihood Method may lead to the same estimator in some models, but they are different estimation procedures and need not produce identical results.

2.7 Properties of Estimators

After constructing an estimator, the next question is how well it performs.

Different estimators of the same parameter can differ in their bias, variability, efficiency, and asymptotic behavior. The properties introduced in this section provide criteria for studying and comparing estimators.

2.7.1 Bias

Let T be an estimator of θ. Its bias at θ is defined by

$$\mathrm{Bias}_\theta(T) = \mathbb{E}_\theta[T] - \theta,$$

provided the expectation exists.

An estimator is called **unbiased** if

$$\mathbb{E}_\theta[T] = \theta$$

for every $\theta \in \Theta$.

For example, if

$$X_1, \ldots, X_n$$

have mean μ, then

$$\mathbb{E}[\overline{X}] = \mu.$$

Therefore, $\overline{X}$ is an unbiased estimator of μ.

Bias measures systematic deviation of the average value of an estimator from the parameter being estimated.

2.7.2 Asymptotic Bias

Let T_n be an estimator of θ based on a sample of size n.

Its bias is

$$\text{Bias}_\theta(T_n) = \mathbb{E}_\theta[T_n] - \theta.$$

If

$$\lim_{n\to\infty} \text{Bias}_\theta(T_n) = 0,$$

then the bias vanishes asymptotically.

An estimator may therefore be biased for finite samples while becoming asymptotically unbiased.

Unbiasedness is a finite-sample property, whereas asymptotic unbiasedness concerns the limiting behavior of the estimator.

2.7.3 Variance

The variance of an estimator T is

$$\text{Var}_\theta(T) = \mathbb{E}_\theta\left[(T - \mathbb{E}_\theta[T])^2\right].$$

Variance measures the variability of the estimator under repeated sampling.

For independent observations with variance σ^2,

$$\text{Var}(\overline{X}) = \frac{\sigma^2}{n}.$$

Thus, increasing the sample size generally reduces the variability of the sample mean under the usual independent sampling assumptions.

Bias and variance measure different aspects of estimator performance. An estimator can have a small variance but a nonzero bias, or it can be unbiased while having a relatively large variance.

2.7.4 Mean Squared Error

The mean squared error of an estimator T of θ is

$$\text{MSE}_\theta(T) = \mathbb{E}_\theta\left[(T - \theta)^2\right].$$

It can be decomposed as

$$\text{MSE}_\theta(T) = \text{Var}_\theta(T) + \text{Bias}_\theta(T)^2.$$

This decomposition shows that estimator performance can involve a trade-off between variability and systematic error.

2.7.5 Efficiency

Efficiency provides a way of comparing estimators of the same parameter.

Suppose that T_1 and T_2 are unbiased estimators of θ. If

$$\mathrm{Var}_\theta(T_1) \leq \mathrm{Var}_\theta(T_2)$$

for all $\theta \in \Theta$, then T_1 has no greater variance than T_2.

When the inequality is strict for some parameter values, the two estimators do not have the same efficiency over the whole parameter space.

Efficiency will later be related to Fisher information and the Cramér–Rao–Fréchet–Darmois bound.

2.7.6 Admissibility

Admissibility is a decision-theoretic property.

Suppose that an estimator is evaluated through a loss function and the corresponding risk function. An estimator T_1 is said to dominate an estimator T_2 if its risk is no larger for every parameter value and strictly smaller for at least one parameter value.

An estimator is called **admissible** if no other estimator dominates it under the specified decision criterion.

Admissibility is therefore different from unbiasedness and consistency. It depends on the loss function used to evaluate the statistical procedure.

2.7.7 Consistency

Let T_n be an estimator of θ based on a sample of size n.

The estimator sequence (T_n) is said to be **consistent** for θ if

$$T_n \xrightarrow{\mathbb{P}} \theta \qquad \text{as } n \to \infty.$$

Consistency means that the estimator becomes increasingly concentrated around the true parameter as the sample size grows.

An estimator does not need to be unbiased for every finite sample size in order to be consistent.

2.7.8 Convergence in Probability

The notation

$$T_n \xrightarrow{\mathbb{P}} \theta$$

means that, for every $\varepsilon > 0$,

$$\lim_{n\to\infty} P_\theta\left(|T_n - \theta| > \varepsilon\right) = 0.$$

This is called **convergence in probability**.

Convergence in probability is the principal mode of convergence used in the definition of consistency.

For example, the law of large numbers implies, under appropriate conditions, that

$$\overline{X}_n \xrightarrow{\mathbb{P}} \mu.$$

Thus, the sample mean is consistent for the population mean under those conditions.

2.7.9 Convergence in Mean of Order k

Let $k > 0$. A sequence of random variables (T_n) is said to converge in mean of order k to θ if

$$\lim_{n\to\infty} \mathbb{E}_\theta \left[|T_n - \theta|^k \right] = 0.$$

This is commonly denoted by

$$T_n \xrightarrow{L^k} \theta.$$

For $k = 2$, the convergence is called convergence in quadratic mean.

Under the usual conditions, convergence in mean of order k implies convergence in probability. Thus, convergence in mean provides a stronger type of control over the distance between the estimator and the parameter.

2.7.10 Sufficiency

A statistic can summarize the sample while retaining all the information about a parameter that is relevant for statistical inference.

Let

$$X = (X_1, \ldots, X_n)$$

have a distribution depending on θ, and let

$$T = T(X)$$

be a statistic.

The statistic T is called **sufficient** for θ if the conditional distribution of the sample given T does not depend on θ.

For models having a joint density or probability mass function, sufficiency can often be established using the factorization criterion.

If

$$f(x_1, \ldots, x_n; \theta) = g(T(x_1, \ldots, x_n), \theta) h(x_1, \ldots, x_n),$$

where h does not depend on θ, then T is sufficient for θ.

Sufficiency is therefore related to the possibility of reducing the data without losing information about the parameter.

2.7.11 Completeness

Let T be a statistic whose distribution depends on θ.

The statistic T is called **complete** if, for every measurable function g for which the expectation exists,

$$\mathbb{E}_\theta[g(T)] = 0 \qquad \text{for every } \theta \in \Theta$$

implies

$$g(T) = 0 \qquad \text{almost surely.}$$

Completeness is a property of the family of distributions generated by the statistic.

It is particularly useful when combined with sufficiency, because a complete sufficient statistic provides strong uniqueness results for unbiased estimators of functions of the parameter.

2.8 Usual Estimators

Several estimators occur repeatedly in statistical applications. The most common examples concern quantitative variables and qualitative variables.

2.8.1 Quantitative Case

Let

$$X_1, \ldots, X_n$$

be observations of a quantitative variable.

The sample mean is

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

It is the usual estimator of the population mean μ.

The usual unbiased estimator of the population variance σ^2 is

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

The corresponding sample standard deviation is

$$S = \sqrt{S^2}.$$

The denominator $n-1$ is important in the unbiased estimation of σ^2 under the standard independent sampling assumptions.

2.8.2 Qualitative Case: Unknown Proportion

Suppose that each observation belongs to one of two categories.

Define

$$X_i = \begin{cases} 1, & \text{if the observation belongs to the category of interest,} \\ 0, & \text{otherwise.} \end{cases}$$

If

$$X_i \sim \text{Bernoulli}(p),$$

then p represents the population proportion of observations belonging to the category of interest.

The natural estimator of p is the sample proportion

$$\widehat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}.$$

In the Bernoulli model, this estimator is also the maximum likelihood estimator of p.

2.9 Fisher Information and Score

The likelihood contains information about the unknown parameter. The score function and Fisher information provide two important tools for quantifying this information.

2.9.1 The Score Function

Let X have density or probability mass function

$$f(x;\theta).$$

The **score function** is defined by

$$U(\theta;X) = \frac{\partial}{\partial\theta}\log f(X;\theta),$$

whenever the derivative exists.

For a sample of independent observations,

$$X_1,\ldots,X_n,$$

the sample score is

$$U_n(\theta) = \frac{\partial}{\partial\theta}\log L(\theta) = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(X_i;\theta).$$

The score describes the local sensitivity of the log-likelihood to changes in the parameter.

2.9.2 Fisher Information

Under appropriate regularity conditions, the Fisher information in one observation is

$$I(\theta) = \mathbb{E}_\theta\left[U(\theta;X)^2\right].$$

Under the same type of regularity conditions, it can also be written as

$$I(\theta) = -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f(X;\theta)\right].$$

For independent observations,

$$I_n(\theta) = nI(\theta).$$

Thus, the total information increases linearly with the sample size in the independent case.

2.9.3 Efficiency and Fisher Information

Fisher information provides a theoretical benchmark for the precision of estimators.

For a regular one-parameter model, the Cramér–Rao–Fréchet–Darmois bound uses Fisher information to provide a lower bound for the variance of an unbiased estimator.

This establishes an important connection:

$$\text{Fisher information} \longrightarrow \text{variance bound} \longrightarrow \text{efficiency}.$$

An estimator that attains the relevant lower bound is efficient in the Cramér–Rao sense.

2.9.4 Fisher Information and Completeness

Fisher information and completeness are distinct concepts.

Fisher information is a measure of the local information about a parameter contained in a statistical model.

Completeness is a property of the family of distributions of a statistic. It is mainly connected with uniqueness properties of unbiased estimators.

Both concepts are important in the study of estimators, but they should not be interpreted as equivalent notions.

2.10 Cramér–Rao–Fréchet–Darmois Bound

The Cramér–Rao–Fréchet–Darmois bound provides a lower bound for the variance of unbiased estimators under suitable regularity conditions.

Consider a regular one-parameter model with Fisher information $I_n(\theta)$ based on a sample of size n.

If T is an unbiased estimator of θ, then, under the required regularity conditions,

$$\mathrm{Var}_\theta(T) \geq \frac{1}{I_n(\theta)}.$$

For independent observations,

$$I_n(\theta) = nI(\theta),$$

so that

$$\mathrm{Var}_\theta(T) \geq \frac{1}{nI(\theta)}.$$

The bound provides a benchmark against which the variance of an unbiased estimator can be compared.

If an unbiased estimator attains the bound, then it has the smallest variance allowed by the bound under the stated conditions and is efficient in the Cramér–Rao sense.

The regularity assumptions are essential. The bound cannot be applied mechanically to every statistical model.

2.10.1 Illustrative Example

Suppose that

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2),$$

where σ^2 is known and μ is unknown.

For one observation,

$$\log f(X;\mu) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(X-\mu)^2}{2\sigma^2}.$$

The score function is

$$U(\mu; X) = \frac{X-\mu}{\sigma^2}.$$

Consequently,

$$I(\mu) = \mathbb{E}_\mu\left[\frac{(X-\mu)^2}{\sigma^4}\right] = \frac{1}{\sigma^2}.$$

For n independent observations,

$$I_n(\mu) = \frac{n}{\sigma^2}.$$

The Cramér–Rao bound therefore gives

$$\mathrm{Var}_\mu(T) \geq \frac{\sigma^2}{n}$$

for every unbiased estimator T of μ satisfying the required conditions.

On the other hand,

$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}.$$

Thus, the sample mean attains the bound in this model and is efficient in the Cramér–Rao sense.

2.11 Remark on ESBVM

The official program supplied for the course mentions "ESBVM" immediately after the Cramér–Rao–Fréchet–Darmois bound. However, the program does not provide the expansion, definition, or mathematical description of this abbreviation.

Consequently, no specific mathematical interpretation is introduced here. The notation should be clarified from the original departmental source before a corresponding section is added to the course.

2.12 Conclusion

Point estimation provides a systematic framework for learning about unknown population parameters from a sample.

The chapter has introduced the distinction between parametric and nonparametric estimation, point estimators and estimates, and random variables and unknown parameters.

Two important methods for constructing estimators were then presented: Maximum Likelihood and the Method of Moments.

The statistical properties of estimators were subsequently introduced, including bias, asymptotic bias, variance, efficiency, admissibility, consistency, sufficiency, completeness, convergence in probability, and convergence in mean of order k.

The usual estimators of a population mean, variance, standard deviation, and proportion were also considered.

Finally, the score function and Fisher information were introduced and linked to the Cramér–Rao–Fréchet–Darmois bound and estimator efficiency.

These concepts form the theoretical foundation for confidence interval estimation and statistical testing, which will be developed in the following chapters.

Chapter 3

Confidence Interval Estimation

3.1 Point Estimation versus Confidence Interval Estimation

In point estimation, an unknown population parameter is estimated by a single numerical value obtained from a sample.

Let

$$X_1, \ldots, X_n$$

be a random sample from a population depending on an unknown parameter θ. A point estimator of θ is a statistic

$$T = T(X_1, \ldots, X_n).$$

After observing the sample, the estimator takes a numerical value

$$t = T(x_1, \ldots, x_n),$$

which is used as an estimate of θ.

A point estimate does not, however, indicate the uncertainty associated with the estimation. Two samples drawn from the same population may produce different estimates.

Confidence interval estimation provides a way of quantifying this sampling uncertainty.

A confidence interval for θ is an interval whose endpoints are calculated from the sample and which is constructed so that the procedure has a specified probability of covering the unknown parameter.

A confidence interval can be written in the general form

$$[L(X_1, \ldots, X_n),\, U(X_1, \ldots, X_n)],$$

where L and U are statistics and

$$P_\theta\left(L \leq \theta \leq U\right) = 1 - \alpha.$$

The quantity $1 - \alpha$ is called the **confidence level**, while α is called the **risk level** or **significance level of the confidence procedure**.

Thus, point estimation and interval estimation have different objectives:

- point estimation provides a single value;
- confidence interval estimation provides a range of plausible values together with a specified confidence level.

3.2 Confidence Level and Risk Level

Let

$$0 < \alpha < 1.$$

A confidence interval with confidence level $1 - \alpha$ is constructed so that

$$P_\theta \left(L(X_1, \ldots, X_n) \leq \theta \leq U(X_1, \ldots, X_n) \right) = 1 - \alpha.$$

The quantity

$$\alpha$$

is called the **risk level**.

For example, the commonly used confidence levels are

$$90\%, \qquad 95\%, \qquad 99\%.$$

They correspond respectively to

$$\alpha = 0.10, \qquad \alpha = 0.05, \qquad \alpha = 0.01.$$

For a fixed sample size, increasing the confidence level generally produces a wider interval. Conversely, decreasing the confidence level generally produces a narrower interval.

This illustrates the fundamental compromise between confidence and precision: a procedure designed to cover the parameter more frequently generally requires a wider interval.

3.2.1 Interpretation of a Confidence Interval

Suppose that a confidence interval

$$[L, U]$$

has been obtained using a procedure with confidence level $1 - \alpha$.

The confidence level is a property of the procedure used to construct the interval. In the frequentist framework, the parameter θ is fixed, whereas the interval endpoints are random before the sample is observed.

Thus, the statement

$$P_\theta(L \leq \theta \leq U) = 1 - \alpha$$

refers to the long-run behavior of the confidence interval procedure over repeated samples.

After the sample has been observed, the resulting interval is fixed.

3.3 Construction of Confidence Intervals

The construction of a confidence interval generally relies on a statistic whose sampling distribution is known or can be determined.

The basic principle is to find a quantity involving the unknown parameter whose distribution does not depend on that parameter, or whose distribution is known after standardization.

Suppose that a statistic T satisfies

$$P_\theta \left(a \leq T(X, \theta) \leq b \right) = 1 - \alpha.$$

If the inequalities can be rearranged in terms of θ, then one obtains a confidence interval

$$[L(X), U(X)]$$

such that

$$P_\theta\left(L(X) \leq \theta \leq U(X)\right) = 1 - \alpha.$$

The form of the confidence interval depends on the parameter being estimated and on the assumptions concerning the population distribution.

3.4 Confidence Interval for a Mean

Let

$$X_1, \ldots, X_n$$

be a random sample from a population with mean μ and variance σ^2.

The parameter of interest is the population mean

$$\mu.$$

The construction of a confidence interval for μ depends on whether the population variance σ^2 is known or unknown.

3.4.1 Known Variance

Assume first that σ^2 is known.

Under the normal model,

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2),$$

the sample mean satisfies

$$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

Therefore,

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1).$$

Let $z_{\alpha/2}$ denote the positive number satisfying

$$P(Z > z_{\alpha/2}) = \frac{\alpha}{2},$$

where $Z \sim \mathcal{N}(0, 1)$.

Then

$$P\left(-z_{\alpha/2} \leq \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1 - \alpha.$$

Rearranging gives

$$P\left(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

Hence the confidence interval for μ is

$$\boxed{\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]}.$$

The margin of error is

$$E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

The interval can therefore be written as

$$\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

3.4.2 Unknown Variance

Suppose now that σ^2 is unknown.

For a normal population,

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2),$$

define

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

The statistic

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

has a Student distribution with $n-1$ degrees of freedom:

$$T \sim t_{n-1}.$$

Let $t_{n-1,\alpha/2}$ denote the positive critical value satisfying

$$P(T > t_{n-1,\alpha/2}) = \frac{\alpha}{2}.$$

Then

$$P\left(-t_{n-1,\alpha/2} \le \frac{\overline{X} - \mu}{S/\sqrt{n}} \le t_{n-1,\alpha/2}\right) = 1 - \alpha.$$

Therefore,

$$P\left(\overline{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}} \le \mu \le \overline{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right) = 1 - \alpha.$$

Hence the confidence interval is

$$\boxed{\left[\overline{X} - t_{n-1,\alpha/2}\frac{S}{\sqrt{n}},\ \overline{X} + t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}\right]}.$$

The corresponding margin of error is

$$E = t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}.$$

3.4.3 Summary for the Mean

For a normal population, the confidence interval for μ is obtained as follows:

Variance	Confidence interval for μ
σ^2 known	$\overline{X} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$
σ^2 unknown	$\overline{X} \pm t_{n-1,\alpha/2}\dfrac{S}{\sqrt{n}}$

The normal distribution is used when σ is known, whereas the Student distribution is used when σ is unknown under the normal population assumption.

3.5 Confidence Interval for a Variance

Consider a random sample

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2),$$

where both μ and σ^2 may be unknown.

The parameter of interest is the population variance σ^2.

The sample variance is

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

For a normal population, the statistic

$$\frac{(n-1)S^2}{\sigma^2}$$

has a chi-square distribution with $n-1$ degrees of freedom:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}.$$

Let

$$\chi^2_{n-1,\alpha/2}$$

and

$$\chi^2_{n-1,1-\alpha/2}$$

denote the appropriate chi-square quantiles, so that

$$P\left(\chi^2_{n-1,\alpha/2} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{n-1,1-\alpha/2}\right) = 1-\alpha.$$

Because σ^2 appears in the denominator, reversing the inequalities when solving for σ^2 gives

$$P\left(\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}\right) = 1-\alpha.$$

Thus, the confidence interval for the population variance is

$$\boxed{\left[\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}\right]}.$$

This construction relies on the normality assumption. In particular, the chi-square distribution of the standardized sample variance is a specific property of the normal model.

3.6 Confidence Interval for a Proportion

Suppose that a population is divided into two categories and let

$$p$$

denote the proportion of individuals belonging to the category of interest.

Let

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p).$$

The sample proportion is

$$\widehat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Since

$$\mathbb{E}[\widehat{p}] = p$$

and

$$\text{Var}(\widehat{p}) = \frac{p(1-p)}{n},$$

the standardized statistic can be approximated by a standard normal variable for sufficiently large samples:

$$\frac{\widehat{p} - p}{\sqrt{p(1-p)/n}} \approx \mathcal{N}(0,1).$$

Using the normal approximation, a confidence interval for p can be constructed.

Replacing the unknown p in the standard error by $\widehat{p}$ gives the usual approximate interval

$$\boxed{\left[\widehat{p} - z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}},\ \widehat{p} + z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right]}.$$

Equivalently,

$$\widehat{p} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}.$$

The approximation is justified when the sample is sufficiently large for the normal approximation to the binomial distribution to be appropriate.

The resulting endpoints may sometimes fall outside the interval $[0,1]$. Since a proportion must satisfy

$$0 \le p \le 1,$$

the assumptions and approximation should be checked before interpreting the interval.

3.7 Influence of Sample Size on Precision

The sample size has a direct influence on the precision of an estimator and of the corresponding confidence interval.

For the confidence interval for a mean with known variance, the margin of error is

$$E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Thus,

$$E \propto \frac{1}{\sqrt{n}}.$$

For a fixed confidence level and fixed population variance, increasing the sample size decreases the margin of error.

For example, if the sample size is multiplied by four,

$$n \longrightarrow 4n,$$

then

$$\frac{1}{\sqrt{4n}} = \frac{1}{2\sqrt{n}}.$$

Therefore, the margin of error is divided by two.

This illustrates an important property of sampling precision: increasing the sample size improves precision, but the improvement occurs at a rate proportional to the inverse square root of the sample size.

For a proportion, the approximate margin of error is

$$E = z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}.$$

Again, the margin of error decreases as the sample size increases.

The same general principle applies to many confidence interval procedures, although the exact form of the margin of error depends on the parameter and the sampling distribution.

3.8 Determination of the Optimal Sample Size

In many statistical studies, the sample size must be determined before the data are collected.

A common objective is to choose n so that the margin of error does not exceed a prescribed value E at a specified confidence level.

3.8.1 Sample Size for Estimating a Mean

Suppose that the population variance σ^2 is known and that the desired margin of error is E.

For a confidence level $1-\alpha$,

$$E = z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Solving for n gives

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2.$$

Since the sample size must be an integer, the required sample size is

$$\boxed{n = \left\lceil \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2 \right\rceil}.$$

The ceiling function ensures that the prescribed margin of error is not exceeded.

If σ is unknown, a preliminary estimate of the standard deviation may be used when planning the study.

3.8.2 Sample Size for Estimating a Proportion

For the approximate confidence interval for a proportion, the margin of error is

$$E = z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}.$$

Solving for n gives

$$n = \frac{z_{\alpha/2}^2 p(1-p)}{E^2}.$$

Since p is generally unknown before the study, a preliminary estimate $\widetilde{p}$ may be used:

$$n = \frac{z_{\alpha/2}^2 \widetilde{p}(1-\widetilde{p})}{E^2}.$$

When no prior information about p is available, the value

$$p = \frac{1}{2}$$

is commonly used because

$$p(1-p) \le \frac{1}{4}$$

for all $p \in [0,1]$.

This gives the conservative sample size

$$\boxed{n = \left\lceil \frac{z_{\alpha/2}^2}{4E^2} \right\rceil}.$$

This choice guarantees the required sample size for the normal-approximation formula over the possible values of p.

3.8.3 Factors Affecting the Required Sample Size

The required sample size depends mainly on:

- the confidence level;
- the desired margin of error;

- the variability of the population;
- the parameter being estimated.

For a fixed confidence level, requiring greater precision means choosing a smaller margin of error. Since the sample size is proportional to $1/E^2$, achieving twice the precision in terms of halving the margin of error generally requires approximately four times the sample size.

Similarly, increasing the confidence level increases the critical value and therefore increases the required sample size for a fixed margin of error.

3.9 Conclusion

Confidence interval estimation extends point estimation by incorporating sampling uncertainty into the estimation procedure.

A confidence interval is constructed so that the associated procedure contains the unknown parameter with a specified confidence level.

The chapter considered confidence intervals for three fundamental parameters:

- the population mean;
- the population variance;
- the population proportion.

For the population mean, the normal distribution is used when the population variance is known, while the Student distribution is used when the variance is unknown under the normal population model.

For the population variance, the chi-square distribution provides the corresponding confidence interval under normality.

For a population proportion, the normal approximation to the binomial model leads to an approximate confidence interval under appropriate sample-size conditions.

Finally, the influence of sample size on precision was examined, together with the determination of a sample size required to achieve a prescribed margin of error at a specified confidence level.

These concepts provide the basis for the construction and interpretation of confidence intervals in practical statistical inference.

Chapter 4

Introduction to Statistical Tests

4.1 Objectives of Statistical Tests

Statistical inference is not limited to estimating unknown population parameters. In many applications, one also needs to make decisions about population characteristics or about assumptions concerning a statistical model.

A statistical test provides a formal procedure for using sample information to evaluate a statistical hypothesis.

The general objective of a statistical test is to determine whether the observed data provide sufficient evidence to reject a specified hypothesis about the population.

A statistical test does not prove a hypothesis to be true or false in an absolute sense. It provides a decision based on the information contained in the sample and on a prescribed decision rule.

The main elements of a statistical test are:

- the formulation of hypotheses;
- the choice of a test statistic;
- the specification of a significance level;
- the determination of a critical region;
- the calculation of the observed value of the test statistic;
- the statistical decision;
- the interpretation of the decision in the context of the problem.

4.2 Null and Alternative Hypotheses

A statistical test begins with the formulation of two hypotheses.

The **null hypothesis**, denoted by

$$H_0,$$

represents the hypothesis that is tested directly.

The **alternative hypothesis**, denoted by

$$H_1$$

or sometimes H_a, represents the competing hypothesis.

For example, suppose that μ denotes a population mean and that a reference value μ_0 is given. One may consider

$$H_0 : \mu = \mu_0$$

against

$$H_1 : \mu \neq \mu_0.$$

The alternative hypothesis determines the form of the statistical test.

4.2.1 Two-Sided Alternative

A two-sided test is used when deviations from the null value in either direction are of interest.

For example,

$$H_0 : \theta = \theta_0$$

against

$$H_1 : \theta \neq \theta_0.$$

Evidence against H_0 may then arise from values of the test statistic that are either sufficiently small or sufficiently large.

4.2.2 Right-Sided Alternative

A right-sided test is used when values larger than the null value are of interest.

The hypotheses take the form

$$H_0 : \theta \leq \theta_0$$

against

$$H_1 : \theta > \theta_0.$$

In some parametric settings, the null hypothesis may be written simply as

$$H_0 : \theta = \theta_0$$

when the test is constructed against the specified one-sided alternative.

4.2.3 Left-Sided Alternative

A left-sided test is used when values smaller than the null value are of interest:

$$H_0 : \theta \geq \theta_0$$

against

$$H_1 : \theta < \theta_0.$$

The direction of the alternative hypothesis determines the location of the critical region.

4.3 Types of Statistical Tests

Statistical tests can be classified according to the hypotheses and the parameter being studied.

For a scalar parameter θ, the three basic forms are:

- two-sided test:
$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0;$$
- right-sided test:
$$H_0 : \theta \leq \theta_0 \quad \text{versus} \quad H_1 : \theta > \theta_0;$$
- left-sided test:
$$H_0 : \theta \geq \theta_0 \quad \text{versus} \quad H_1 : \theta < \theta_0.$$

Statistical tests may also be classified according to the statistical model and the parameter under investigation.

In this course, the principal parametric tests concern:

- a population mean;
- a population variance;
- a population proportion;
- the difference between two population means;
- the comparison of two population variances;
- the difference between two population proportions.

The appropriate test depends on the assumptions concerning the population and on whether the population variance is known or unknown.

4.4 Neyman–Pearson Principle

The Neyman–Pearson principle provides a fundamental framework for the construction of statistical tests.

Consider testing a simple null hypothesis

$$H_0 : \theta = \theta_0$$

against a simple alternative

$$H_1 : \theta = \theta_1.$$

Suppose that the observations have density or probability mass function

$$f(x; \theta).$$

The likelihood ratio compares the likelihood under the null hypothesis with the likelihood under the alternative hypothesis.

The likelihood ratio can be written as

$$\Lambda(x) = \frac{f(x;\theta_0)}{f(x;\theta_1)}.$$

A likelihood-ratio test rejects H_0 when $\Lambda(x)$ is sufficiently small, or equivalently when

$$\frac{f(x;\theta_1)}{f(x;\theta_0)}$$

is sufficiently large.

The threshold is chosen so that the probability of rejecting H_0 when H_0 is true is controlled at the prescribed significance level.

The Neyman–Pearson lemma establishes, under the simple-versus-simple setting, that the likelihood-ratio test provides the most powerful test of a given size among the corresponding class of tests.

This principle is fundamental for understanding the construction of statistical tests.

4.5 Critical and Acceptance Regions

Let $T = T(X_1, \ldots, X_n)$ be a test statistic.

A statistical test divides the possible values of the test statistic into two regions.

The **critical region**, also called the rejection region, is the set of values for which the null hypothesis is rejected.

The **acceptance region** is the complementary region in which the null hypothesis is not rejected.

If $\mathcal{C}$ denotes the critical region, then the decision rule is

$$\begin{cases} \text{Reject } H_0, & T \in \mathcal{C}, \\ \text{Do not reject } H_0, & T \notin \mathcal{C}. \end{cases}$$

The critical region depends on the form of the alternative hypothesis.

For a right-sided test, the critical region generally lies in the upper tail of the distribution of the test statistic.

For a left-sided test, it lies in the lower tail.

For a two-sided test, the critical region is generally divided between the two tails.

4.6 Significance Level

The **significance level** of a test is denoted by

$$\alpha, \qquad 0 < \alpha < 1.$$

It is the probability of rejecting the null hypothesis when the null hypothesis is true.

For a simple null hypothesis, this can be written as

$$\alpha = P_{\theta_0}(T \in \mathcal{C}).$$

For a composite null hypothesis, the size of the test is generally defined through the largest rejection probability over the parameter values allowed under H_0.

Common significance levels are

$$0.10, \qquad 0.05, \qquad 0.01.$$

The significance level is fixed before applying the test.

A smaller value of α requires stronger evidence against H_0 before the null hypothesis is rejected.

4.7 Type I and Type II Errors

A statistical decision can lead to two types of errors.

A **Type I error** occurs when the null hypothesis is rejected even though it is true.

The probability of a Type I error is controlled by the significance level α.

A **Type II error** occurs when the null hypothesis is not rejected even though the alternative hypothesis is true.

For a particular alternative parameter value θ, the probability of a Type II error is denoted by

$$\beta(\theta).$$

Thus,

$$\beta(\theta) = P_\theta(T \notin \mathcal{C}),$$

for parameter values belonging to the alternative hypothesis.

The two errors are conceptually different:

- Type I error: rejecting a true null hypothesis;
- Type II error: not rejecting a false null hypothesis.

In general, reducing the probability of a Type I error while keeping the sample size fixed can increase the probability of a Type II error.

4.8 Power of a Test

The **power** of a statistical test at a parameter value θ is the probability of rejecting the null hypothesis when that parameter value is true.

It is defined by

$$\pi(\theta) = P_\theta(T \in \mathcal{C}).$$

For parameter values in the alternative hypothesis,

$$\pi(\theta) = 1 - \beta(\theta).$$

Thus, the power measures the ability of a test to detect departures from the null hypothesis.

A test with greater power at a particular alternative parameter value has a greater probability of rejecting the null hypothesis when that alternative is true.

The power depends on the parameter value, the significance level, the sample size, the test statistic, and the underlying statistical model.

4.9 Decision Table

The possible decisions of a statistical test can be summarized as follows.

	H_0 true	H_1 true
Reject H_0	Type I error	Correct decision
Do not reject H_0	Correct decision	Type II error

The probability of a Type I error is controlled by α, whereas the probability of a Type II error is represented by $\beta(\theta)$ for a specific alternative parameter value.

The power is

$$1 - \beta(\theta).$$

This table emphasizes that a statistical decision is made under uncertainty. The test provides a rule for making the decision based on the sample.

4.10 General Steps of a Statistical Test

A statistical test can be carried out through the following general steps.

Step 1: Formulate the hypotheses

Specify the null and alternative hypotheses:

$$H_0 \qquad \text{and} \qquad H_1.$$

Step 2: Choose the significance level

Select a value

$$\alpha$$

before examining the result of the test.

Step 3: Choose the test statistic

Construct a statistic whose sampling distribution is known, or can be determined or approximated, under the null hypothesis.

Step 4: Determine the critical region

Use the distribution of the test statistic under H_0 to determine the critical value or critical values corresponding to the chosen significance level.

Step 5: Calculate the observed test statistic

Use the sample observations to calculate the numerical value of the test statistic.

Step 6: Make the statistical decision

If the observed value belongs to the critical region, reject H_0.

Otherwise, do not reject H_0.

Step 7: State the conclusion

The statistical conclusion should be expressed in the context of the original problem.

It is important to distinguish between "reject H_0" and "do not reject H_0". Failure to reject H_0 does not establish that H_0 is true.

4.11 Standard Statistical Tables

The construction of many classical statistical tests requires critical values from standard probability distributions.

The main distributions used in this course are:

- the standard normal distribution;
- Student's t distribution;
- the chi-square distribution;
- the Fisher–Snedecor distribution.

4.11.1 Normal Distribution

If

$$Z \sim \mathcal{N}(0,1),$$

then critical values can be obtained from the standard normal distribution table.

For a two-sided test at significance level α, the critical values are

$$-z_{\alpha/2} \qquad \text{and} \qquad z_{\alpha/2}.$$

For a right-sided test, the critical value is

$$z_\alpha,$$

and for a left-sided test the corresponding lower critical value is

$$-z_\alpha.$$

4.11.2 Student's t Distribution

The Student's t distribution depends on the number of degrees of freedom.

It is used, in particular, for tests concerning a population mean when the population variance is unknown and the normal population assumption is appropriate.

The critical values depend on both the significance level and the degrees of freedom.

4.11.3 Chi-Square Distribution

The chi-square distribution depends on its degrees of freedom.

It is used in particular for tests concerning a population variance under the normal population assumption.

If

$$X \sim \chi^2_\nu,$$

then its distribution is determined by the degrees of freedom ν.

4.11.4 Fisher–Snedecor Distribution

The Fisher–Snedecor distribution, usually denoted by F, depends on two degrees of freedom.

It is used in particular for comparing two population variances under the appropriate normality and independence assumptions.

4.12 Parametric Tests

Parametric tests are based on a specified statistical model involving one or more unknown parameters.

The tests considered in this section concern means, variances, proportions, and comparisons between two populations.

4.13 Test for a Population Mean

Let

$$X_1, \ldots, X_n$$

be a random sample from a normal population with mean μ and variance σ^2.

Suppose that the objective is to test

$$H_0 : \mu = \mu_0.$$

4.13.1 Known Variance

Suppose that σ^2 is known.

The test statistic is

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}.$$

Under H_0,

$$Z \sim \mathcal{N}(0, 1).$$

For the two-sided alternative

$$H_1 : \mu \neq \mu_0,$$

the critical region is

$$|Z| > z_{\alpha/2}.$$

For the right-sided alternative

$$H_1 : \mu > \mu_0,$$

the critical region is

$$Z > z_{\alpha}.$$

For the left-sided alternative

$$H_1 : \mu < \mu_0,$$

the critical region is

$$Z < -z_{\alpha}.$$

4.13.2 Unknown Variance

Suppose that σ^2 is unknown.

For a normal population, use

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}.$$

Under H_0,

$$T \sim t_{n-1}.$$

For a two-sided alternative,

$$H_1 : \mu \neq \mu_0,$$

the critical region is

$$|T| > t_{n-1,\alpha/2}.$$

For a right-sided alternative,

$$H_1 : \mu > \mu_0,$$

the critical region is

$$T > t_{n-1,\alpha}.$$

For a left-sided alternative,

$$H_1 : \mu < \mu_0,$$

the critical region is

$$T < -t_{n-1,\alpha}.$$

4.14 Test for a Population Variance

Consider

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2).$$

Suppose that the objective is to test

$$H_0 : \sigma^2 = \sigma_0^2.$$

The test statistic is

$$Q = \frac{(n-1)S^2}{\sigma_0^2}.$$

Under H_0,

$$Q \sim \chi^2_{n-1}.$$

For the two-sided alternative

$$H_1 : \sigma^2 \neq \sigma_0^2,$$

the critical region is

$$Q < \chi^2_{n-1,\alpha/2} \quad \text{or} \quad Q > \chi^2_{n-1,1-\alpha/2}.$$

For a right-sided alternative

$$H_1 : \sigma^2 > \sigma_0^2,$$

the critical region is

$$Q > \chi^2_{n-1,1-\alpha}.$$

For a left-sided alternative

$$H_1 : \sigma^2 < \sigma_0^2,$$

the critical region is

$$Q < \chi^2_{n-1,\alpha}.$$

The use of the chi-square distribution here relies on the normality assumption.

4.15 Test for a Population Proportion

Let

$$X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Bernoulli}(p),$$

and let $\widehat{p}$ denote the sample proportion:

$$\widehat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Suppose that we want to test

$$H_0 : p = p_0.$$

For sufficiently large samples, the standardized statistic

$$Z = \frac{\widehat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

is approximately standard normal under H_0.

Thus,

$$Z \approx \mathcal{N}(0, 1).$$

For the two-sided alternative

$$H_1 : p \neq p_0,$$

the approximate critical region is

$$|Z| > z_{\alpha/2}.$$

For

$$H_1 : p > p_0,$$

the critical region is

$$Z > z_\alpha.$$

For

$$H_1 : p < p_0,$$

the critical region is

$$Z < -z_\alpha.$$

The validity of this approximation depends on the sample size and on the adequacy of the normal approximation to the binomial distribution.

4.16 Comparison of Two Population Means

Suppose that two independent populations have means μ_1 and μ_2.

Let

$$X_1, \ldots, X_{n_1}$$

be a sample from the first population and

$$Y_1, \ldots, Y_{n_2}$$

a sample from the second population.

The parameter of interest is

$$\mu_1 - \mu_2.$$

A common hypothesis is

$$H_0 : \mu_1 - \mu_2 = 0,$$

which is equivalent to

$$H_0 : \mu_1 = \mu_2.$$

4.16.1 Known Variances

Suppose that the two population variances σ_1^2 and σ_2^2 are known.

The test statistic is

$$Z = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

Under the null hypothesis,

$$Z \sim \mathcal{N}(0, 1).$$

For

$$H_0 : \mu_1 = \mu_2$$

against

$$H_1 : \mu_1 \neq \mu_2,$$

the two-sided critical region is

$$|Z| > z_{\alpha/2}.$$

4.16.2 Unknown Variances

When the variances are unknown, the test procedure depends on the assumptions made about the two populations.

If the two normal populations have equal variances, a pooled-variance Student test can be used.

The pooled variance estimator is

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

The test statistic is

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)_0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

Under the null hypothesis,

$$T \sim t_{n_1+n_2-2}.$$

The critical region is then determined from the Student distribution according to the chosen alternative hypothesis.

4.17 Comparison of Two Population Variances

Suppose that two independent normal populations have variances σ_1^2 and σ_2^2.

The hypothesis of equal variances is

$$H_0 : \sigma_1^2 = \sigma_2^2.$$

The statistic

$$F = \frac{S_1^2}{S_2^2}$$

has an F distribution under the null hypothesis:

$$F \sim F_{n_1-1,n_2-1}.$$

For a two-sided alternative

$$H_1 : \sigma_1^2 \neq \sigma_2^2,$$

the critical region is determined from the two tails of the Fisher–Snedecor distribution.

For one-sided alternatives, the critical region is placed in the corresponding tail.

The procedure relies on the independence and normality assumptions for the two samples.

4.18 Comparison of Two Population Proportions

Let

$$\widehat{p}_1 = \frac{X_1}{n_1}, \qquad \widehat{p}_2 = \frac{X_2}{n_2}$$

be the sample proportions from two independent populations.

The parameter of interest is

$$p_1 - p_2.$$

To test

$$H_0 : p_1 = p_2,$$

a pooled estimate of the common proportion is used:

$$\widehat{p} = \frac{n_1\widehat{p}_1 + n_2\widehat{p}_2}{n_1 + n_2}.$$

The approximate test statistic is

$$Z = \frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}.$$

Under the null hypothesis and suitable large-sample conditions,

$$Z \approx \mathcal{N}(0,1).$$

For the two-sided alternative

$$H_1 : p_1 \neq p_2,$$

the critical region is

$$|Z| > z_{\alpha/2}.$$

For

$$H_1 : p_1 > p_2,$$

the critical region is

$$Z > z_{\alpha}.$$

For

$$H_1 : p_1 < p_2,$$

the critical region is

$$Z < -z_{\alpha}.$$

The normal approximation should only be used when the sample sizes and observed counts are sufficiently large for the approximation to be appropriate.

4.19 Summary of Parametric Tests

The principal parametric tests introduced in this chapter can be summarized as follows.

Parameter	Test statistic	Reference distribution
μ, σ^2 known	$\dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$	$\mathcal{N}(0,1)$
μ, σ^2 unknown	$\dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$	t_{n-1}
σ^2	$\dfrac{(n-1)S^2}{\sigma_0^2}$	χ^2_{n-1}
p	$\dfrac{\widehat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$	$\mathcal{N}(0,1)$ approximately

For two populations, the corresponding procedures involve the normal, Student, Fisher–Snedecor, or normal approximation distributions depending on the parameter and assumptions.

4.20 Conclusion

Statistical testing provides a formal framework for making decisions about population characteristics using sample information.

The construction of a statistical test begins with the formulation of a null hypothesis and an alternative hypothesis. The significance level determines the allowed probability of a Type I error, while the power of the test describes its ability to detect departures from the null hypothesis.

The Neyman–Pearson principle provides a fundamental basis for constructing powerful tests in the simple-versus-simple setting.

The concepts of critical region, acceptance region, Type I error, Type II error, and power are central to the interpretation of statistical tests.

The chapter also introduced the main probability distributions used in classical parametric testing: the normal, Student, chi-square, and Fisher–Snedecor distributions.

Finally, tests concerning a population mean, variance, and proportion were presented, together with procedures for comparing two population means, variances, and proportions.

These methods provide the basic tools required for applying statistical hypothesis testing to common inference problems.