College of Natural Sciences

Department of Statistics

Course Title: Statistical Inference
Course code: **Stat 3052**

Credit: 5 EtCTS

Credit hours: 3 (3Lecture hrs+2 hrs tutorial)

Instructor's Name: Kenenisa T. (MSc.)

Email: kenenisatadesse@gmail.com

April, 2020

Jimma, Ethiopia

Statistical Inference(Stat-3052)

Outline

Chapter 0: Preliminaries

- The aim of **statistical inference** is to make certain determinations with regard to the unknown constants known as parameter(s) in the underlying distribution.
- With the intention of emphasizing the importance of the basic concepts, we begin with a review of the definitions of terms related to random sampling distribution of some estimators in the preliminary chapter.
- The first step in statistical inference is Point Estimation, in which we compute a single value (statistic) from the sample data to estimate a population parameter.
- General concept of **point estimators**, different methods of finding estimators and clarification of their properties are discussed in Chapter 1.
- Then proceed to **Interval Estimation**, a method of obtaining, at a given level of confidence (or probability), two statistics which include within their range an unknown but fixed parameter are discussed in Chapter 2.
- In Chapter 3 we discuss a 2^{nd} major area of statistical inference is **Testing of Hypotheses**. The significance of the differences between estimated parameters from two or more samples are also included in this chapter; such as the significance the difference of two population means.
- **Nonparametric methods** that does not based on sampling distributions are discussed in Chapter 4 (Group Work to be presented by Students).

Definitions of Some Basic Terms

- **Population** refers to all elements of interest characterized by a distribution **F** with some **parameter**, say $\theta \in \Theta$ (where Θ is the set of its possible values called **the parameter space**).
- **Sample** is the set of data $X_1, \ldots, X_n$, selected subset of the population, n is sample size.
- Remember, to use sample data for inference, needs to be representative of population for the question(s) of interest or our study.
- For $X_1, \ldots, X_n$, a random sample (independent and identically distributed, iid) from a distribution with *cumulative distribution function* (*cdf*) $F(x;\theta)$. The cdf admits a probability mass function (pmf) in the discrete case and a probability density function (pdf) in the continuous case, in either case, write this function as $f(x;\theta)$.
- **A parameter** is a number associated with a population characteristic.– value unknown. It is usually assumed to be fixed but unknown. Thus, we estimate the parameter using sample information.
  Examples of population parameters: sample mean (μ) and population variance (σ^2).
- **A statistic or estimate** is a number computed from a sample. A statistic estimates a parameter and it changes with each new sample.
- A statistic is any function of the observations in a random sample, (**no parameter in the function**).

Examples of A Statistic

- **For example**, the sample mean, the sample variance, and the sample proportion p are statistics and they are also random variables.
- Let $X_1, \ldots, X_n$ be random samples taken from a population. **The sample mean**:

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

The sample variance (biased):

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2. \tag{2}$$

The sample variance (unbiased):

$$S^{2*} = \frac{1}{n-1}\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2. \tag{3}$$

The sample proportion p,

$$p = \frac{x}{n}. \tag{4}$$

Sampling Distribution

- Since a statistic is a function of a random variable, it self is also a random variable, and it has a probability distribution. The distribution of a statistic is called **the sampling distribution of the statistic** because is depends on the sample chosen.
- **Example**: Consider $\mathbf{X} = \{X_1, \ldots, X_n\}$ be random samples (iid) taken from a normal population with sample mean μ and variance σ^2, or $\mathbf{X} \sim N\left(\mu,\ \sigma^2\right)$ for each of sample of size n.
  - **The sampling distribution of the sample mean**, $\bar{X}$:

$$
\begin{aligned}
E\left[\bar{X}\right] =& E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[X_i\right] \\
=& \frac{1}{n}\sum_{i=1}^{n}\mu, \quad since\ E\left[X_i\right] = \mu \\
=& \frac{1}{n}\left(n\mu\right) = \mu.
\end{aligned}
$$

$$
\begin{aligned}
Var\left[\bar{X}\right] =& Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \quad \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} Var\left[X_i\right] = \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n}\sigma^2, \quad since X_i^s\ are\ iid \\
=& \frac{1}{n}\left(n\sigma^2\right) = \frac{\sigma^2}{n}.
\end{aligned}
$$

Therefore, $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ **is the sampling distribution of**, $\bar{X}$:

Assessment

- **Problem**: Consider $\mathbf{X} = \{X_1, \ldots, X_n\}$ be random samples (iid) taken from a normal population with mean μ and variance σ^2, or $\mathbf{X} \sim N\left(\mu,\ \sigma^2\right)$ for each of sample of size n.

1. Define $\mathbf{Z} = \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}$.

a. Are there any parameter in the function of **Z**? What are they?

b. Is **Z** a statistic? Why?

c. Derive the $E\left[\mathbf{Z}\right]$.

d. Derive the $Var\left[\mathbf{Z}\right]$.

e. Give the sampling distribution of **Z**.

2. Define $\mathbf{W} = \frac{X-\mu}{\sigma}$.

a. Are there any parameter in the function of **W**? What are they?

b. Is **W** a statistic? Why?

c. Derive the $E\left[\mathbf{W}\right]$.

d. Derive the $Var\left[\mathbf{W}\right]$.

e. Give the sampling distribution of **W**.

What is Statistical Inference?

- **Statistics** is closely related to **probability theory**, but have entirely different goals.
- Recall, from statical theory, that a typical probability problem starts with some assumptions about the distribution of a random variable (e.g., that it's binomial), and the objective is to derive some properties (probabilities, expected values, etc) of said random variable based on the stated assumptions.
- **In statistics**, a sample from a given population is observed, and the goal is to learn something about that population based on the sample.

Statistical Inference/Inferential Statistics

is a conceptually the process of drawing conclusions about population based on the samples that are subject to random variation.

- Every scientific discipline applies statistics to seek relevant information from a given sample of data. The procedure leads to conclusions regarding a **population**, which includes all possible observations of the process or phenomenon, and is called **statistical inference**.
- **The goal of statistical inference** is to develop the mathematical theory of statistics, mostly building on calculus and probability.

Types of Statistical Inference:

- Parameter estimation:

What is Statistical Inference?

- **Statistics** is closely related to **probability theory**, but have entirely different goals.
- Recall, from statical theory, that a typical probability problem starts with some assumptions about the distribution of a random variable (e.g., that it's binomial), and the objective is to derive some properties (probabilities, expected values, etc) of said random variable based on the stated assumptions.
- **In statistics**, a sample from a given population is observed, and the goal is to learn something about that population based on the sample.

Statistical Inference/Inferential Statistics

is a conceptually the process of drawing conclusions about population based on the samples that are subject to random variation.

- Every scientific discipline applies statistics to seek relevant information from a given sample of data. The procedure leads to conclusions regarding a **population**, which includes all possible observations of the process or phenomenon, and is called **statistical inference**.
- **The goal of statistical inference** is to develop the mathematical theory of statistics, mostly building on calculus and probability.

Types of Statistical Inference:

- Parameter estimation:

1. Point estimation;
2. Interval estimation;

What is Statistical Inference?

- **Statistics** is closely related to **probability theory**, but have entirely different goals.
- Recall, from statical theory, that a typical probability problem starts with some assumptions about the distribution of a random variable (e.g., that it's binomial), and the objective is to derive some properties (probabilities, expected values, etc) of said random variable based on the stated assumptions.
- **In statistics**, a sample from a given population is observed, and the goal is to learn something about that population based on the sample.

Statistical Inference/Inferential Statistics

is a conceptually the process of drawing conclusions about population based on the samples that are subject to random variation.

- Every scientific discipline applies statistics to seek relevant information from a given sample of data. The procedure leads to conclusions regarding a **population**, which includes all possible observations of the process or phenomenon, and is called **statistical inference**.
- **The goal of statistical inference** is to develop the mathematical theory of statistics, mostly building on calculus and probability.

Types of Statistical Inference:

- Parameter estimation:
  1. Point estimation;
  2. Interval estimation;
- **Hypothesis testing**;

What is Statistical Inference?

- **Statistics** is closely related to **probability theory**, but have entirely different goals.
- Recall, from statical theory, that a typical probability problem starts with some assumptions about the distribution of a random variable (e.g., that it's binomial), and the objective is to derive some properties (probabilities, expected values, etc) of said random variable based on the stated assumptions.
- **In statistics**, a sample from a given population is observed, and the goal is to learn something about that population based on the sample.

Statistical Inference/Inferential Statistics

is a conceptually the process of drawing conclusions about population based on the samples that are subject to random variation.

- Every scientific discipline applies statistics to seek relevant information from a given sample of data. The procedure leads to conclusions regarding a **population**, which includes all possible observations of the process or phenomenon, and is called **statistical inference**.
- **The goal of statistical inference** is to develop the mathematical theory of statistics, mostly building on calculus and probability.

Types of Statistical Inference:

- Parameter estimation:
  1. Point estimation;
  2. Interval estimation;
- **Hypothesis testing**;
- **Nonparametric Methods**.

Chapter 1: Parametric Point Estimation

- In this chapter, methods of parameter estimation called **point estimation** are introduced.
- One assumes for this purpose that the distribution of the population is known. However, the values of the parameters of the distribution have to be estimated from **a sample of data**, that is, a subset of the population. One also assumes that the sample is **random**.

Definition:

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

- As an example, $\bar{X}$ is a point estimator of μ, that is, $\hat{\mu} = \bar{X}$ and S^2 is a point estimator of σ^2, that is, $\hat{\sigma}^2 = S^2$.
- The main objective of this chapter is to draw a random sample of size **n**, $X_1, \ldots, X_n$, from the underlying distribution, and on the basis of it to construct a point estimate (or estimator) for θ, that is, a statistic $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n) \in \boldsymbol{\Theta}$, which is used for estimating θ.

Methods of Finding Parametric Point Estimators

- There is any number of estimates one may construct, thus, the need to assume certain principles or methods for constructing $\hat{\theta}$.

Methods of Finding Parametric Point Estimators are:

1. **Maximum likelihood estimation (MLE)**.

Methods of Finding Parametric Point Estimators

- There is any number of estimates one may construct, thus, the need to assume certain principles or methods for constructing $\hat{\theta}$.

Methods of Finding Parametric Point Estimators are:

1. **Maximum likelihood estimation (MLE)**.
2. **Estimation by the method of moments**.

Methods of Finding Parametric Point Estimators

- There is any number of estimates one may construct, thus, the need to assume certain principles or methods for constructing $\hat{\theta}$.

Methods of Finding Parametric Point Estimators are:

1. **Maximum likelihood estimation (MLE)**.
2. **Estimation by the method of moments**.
3. **The method of least squares estimation**.

Methods of Finding Parametric Point Estimators

- There is any number of estimates one may construct, thus, the need to assume certain principles or methods for constructing $\hat{\theta}$.

Methods of Finding Parametric Point Estimators are:

1. **Maximum likelihood estimation (MLE)**.
2. **Estimation by the method of moments**.
3. **The method of least squares estimation**.

- The least squares method is commonly used in the so-called, Regression Analysis (Stat-2041), Statistical methods (Stat1013), Time Series Analysis (Stat-2042) and other statistics courses.

Maximum Likelihood (ML) Method

- Perhaps, the most widely accepted principle is the so-called **principle of Maximum Likelihood (ML)**.
- Let **X** be a r.v. with p.d.f. $f(.;\theta)$, where θ is unknown a parameter lying in a parameter space $\mathbf{\Theta}$.
- Then the objective is to estimate θ on the basis of a random sample of size n from $f(.;\theta)$, $X_1, X_2, \ldots, X_n$.
- Then, replacing θ in $f(.;\theta)$ by a "good" estimate of it, one would expect to be able to use the resulting p.d.f. for the purposes.
- This principle dictates that we form the joint p.d.f. of the observed values of the $X_i^{\prime}s$, is a function of θ (and call it the likelihood function), and maximize the likelihood function with respect to θ.
- The maximizing point (assuming it exists and is unique) is a function of $X_1, X_2, \ldots, X_n$, and is what we call the Maximum Likelihood Estimate (MLE) of θ.
- The notation used for the likelihood function is $L(\theta|X_1, X_2, \ldots, X_n)$. Then, we have that:

$$L\left(\theta|X_1, X_2, \ldots, X_n\right) = f\left(X_1;\theta\right) \times f\left(X_2;\theta\right) \times \ldots \times f\left(X_n;\theta\right) = \prod_{i=1}^{n} f\left(X_i;\theta\right), \quad \theta \in \mathbf{\Theta}. \tag{5}$$

- A value of θ which maximizes $L\left(\theta|\mathbf{X}\right)$ is called **a Maximum Likelihood Estimate(MLE)** of θ.
- Clearly, the MLE depends on X, and we usually write $\hat{\theta} = \hat{\theta}\left(\mathbf{X}\right)$.
- Thus, $L\left(\hat{\theta}|\mathbf{X}\right) = max\left\{L\left(\theta|\mathbf{X}\right);\ \theta \in \mathbf{\Theta}\right\}$.

MLE Cont...

- Once we decide to adopt the Maximum Likelihood Principle it is done through **differentiation**.
- It must be stressed that, whenever a maximum is sought by differentiation, the second-order derivative(s) must also be examined in search of a maximum.
- Also, it should be mentioned that maximization of the likelihood function, which is the product of n factors, is equivalent to maximization of its logarithm (always with base e), which is the sum of n summands, thus much easier to work with.
- **REMARK 1:** Let us recall that a function $y = g(X)$ **attains a maximum at a point** $X = x_0$, if

$$\frac{d}{dx} g(X) \mid_{x=x_0} = 0 \text{ and } \frac{d^2}{dx^2} g(X) \mid_{x=x_0} < 0.$$

- **Example:** Consider the following Bernoulli pmf of discrete random variables $X = \{0, 1\}$ and parameter p, with parametric space $\boldsymbol{\Theta} = (p)$

$$f\left(X_j|p\right) = p^{x_j}\left(1-p\right)^{1-x_j}, \text{ where } x_j = 0, 1.$$

where X is a discrete variable and p is a parameter.

$$\begin{aligned} L\left(p|X\right) &= \prod_{j=1}^{n} f\left(X_j; p\right) \\ &= \prod_{j=1}^{n} p^{x_j}\left(1-p\right)^{1-x_j} \quad \text{where } x_j = 0, 1. \end{aligned}$$

MLE Example 1

- Taking the logarithm of both sides to get:

$$\begin{aligned}
ln\left(L\left(p|X\right)\right) =& ln\left(\prod_{j=1}^{n} p^{x_j}\left(1-p\right)^{1-x_j}\right)\\
=& ln\left(p\right)\sum_{j=1}^{n} x_j + ln\left(1-p\right)\sum_{j=1}^{n}\left(1-x_j\right)\\
=& ln\left(p\right)\sum_{j=1}^{n} x_j + ln\left(1-p\right)\left(n-\sum_{j=1}^{n} x_j\right), \ \ and\\
\frac{d\left(ln\left(L\left(p|X\right)\right)\right)}{dp} =& \frac{d\left(ln\left(p\right)\sum\limits_{j=1}^{n} x_j + ln\left(1-p\right)\left(n-\sum\limits_{j=1}^{n} x_j\right)\right)}{dp}\\
=& \frac{\sum\limits_{j=1}^{n} x_j}{p} + \frac{\left(n-\sum\limits_{j=1}^{n} x_j\right)}{1-p}.
\end{aligned}$$

- Hence by equating the foregoing equation to zero, the estimate of the parameter is obtained as

$$\hat{p} = \frac{\sum\limits_{j=1}^{n} x_j}{n}.$$

MLE Examples 2

- Let $X_1, \ldots, X_n$ be a continuous random sample from the $N\left(\mu, \sigma^2\right)$ distribution, with parametric space $\boldsymbol{\Theta} = \left\{\mu, \sigma^2\right\}$, where only one of the parameters is known. Determine the MLE of the other (unknown) parameter.

$$f\left(X_j; \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi}\sigma} exp\left[\frac{-1}{2\sigma^2}\left(x_i - \mu\right)^2\right], \; -\infty < x_j < \infty, \; \sigma^2 > 0$$

- **Case 1: Let μ be unknown**

$$\begin{aligned} L\left(\mu|X_j\right) =& \prod_{j=1}^{n}\left(\frac{1}{\sqrt{2\pi}\sigma} exp\left[\frac{-1}{2\sigma^2}\left(x_i - \mu\right)^2\right]\right) \\ =& \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n exp\left[\frac{-1}{2\sigma^2}\sum_{j=1}^{n}\left(x_i - \mu\right)^2\right] \\ ln\left(L\left(\mu|X_j\right)\right) =& ln\left(\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n exp\left[\frac{-1}{2\sigma^2}\sum_{j=1}^{n}\left(x_i - \mu\right)^2\right]\right) \\ =& -n\, ln\left(\sqrt{2\pi}\right) - n\, ln\left(\sigma\right) - \frac{1}{2\sigma^2}\sum_{j=1}^{n}\left(x_i - \mu\right)^2 \qquad (6) \end{aligned}$$

MLE Examples 2 cont...

- Taking the partial derivative of Equation (6) in terms of μ both sides and then equating the result in to zero:

$$\frac{\partial\left[ln\left(L\left(\mu|X_j\right)\right)\right]}{\partial\mu} = \frac{\partial\left[-n\,ln\left(\sqrt{2\pi}\right) - n\,ln\left(\sigma\right) - \frac{1}{2\sigma^2}\sum_{j=1}^{n}\left(x_i-\mu\right)^2\right]}{\partial\mu}$$

$$= 0 - \frac{2}{2\sigma^2}\sum_{j=1}^{n}\left(x_i-\mu\right),\ \textit{equating to zero}$$

$$\sum_{j=1}^{n}\left(x_i-\mu\right) = 0$$

$$\sum_{j=1}^{n}x_i - n\,\mu = 0$$

$$\hat{\mu} = \frac{\sum_{j=1}^{n}x_i}{n} = \bar{X}$$

MLE Examples 2 cont...

- **Case 2: Let σ be unknown**
- Taking the partial derivative of Equation (6) in terms of σ both sides and then equating the result in to zero:

$$\frac{\partial\left[\ln\left(L\left(\sigma^2|X_j\right)\right)\right]}{\partial\sigma} = \frac{\partial\left[-n\ln\left(\sqrt{2\pi}\right) - n\ln(\sigma) - \frac{1}{2\sigma^2}\sum_{j=1}^{n}(x_i-\mu)^2\right]}{\partial\sigma}$$

$$= 0 - \frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{j=1}^{n}(x_i-\mu)^2\,,\ \textit{equating to zero}$$

$$\sum_{j=1}^{n}(x_i-\mu)^2 = n\,\sigma^2$$

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{n}(x_i-\mu)^2}{n} = S^2$$

Properties of MLE

Properties of Maximum Likelihood Estimation

- Under very general and not restrictive conditions, when the sample size n is large and $\hat{\theta}$ if is the maximum likelihood estimator of the parameter θ, then:

Properties of MLE

Properties of Maximum Likelihood Estimation

- Under very general and not restrictive conditions, when the sample size n is large and $\hat{\theta}$ if is the maximum likelihood estimator of the parameter θ, then:

1. $\hat{\theta}$ is an approximately unbiased estimator for θ.

Properties of MLE

Properties of Maximum Likelihood Estimation

- Under very general and not restrictive conditions, when the sample size n is large and $\hat{\theta}$ if is the maximum likelihood estimator of the parameter θ, then:

1. $\hat{\theta}$ is an approximately unbiased estimator for θ.
2. The variance of $\hat{\theta}$ is nearly as small as the variance that could be obtained with any other estimator, and

Properties of MLE

Properties of Maximum Likelihood Estimation

- Under very general and not restrictive conditions, when the sample size n is large and $\hat{\theta}$ if is the maximum likelihood estimator of the parameter θ, then:

1. $\hat{\theta}$ is an approximately unbiased estimator for θ.
2. The variance of $\hat{\theta}$ is nearly as small as the variance that could be obtained with any other estimator, and
3. $\hat{\theta}$ has an approximate normal distribution.

Method of Moments

- **Definition: Moments** Let $X_1, \ldots, X_n$ be a random sample from either a probability mass function or probability density function with r unknown parameters $\theta_1, \ldots, \theta_r$. The moment estimators $\hat{\theta_1}, \ldots, \hat{\theta_r}$ are found by equating the first r population moments to the first r sample moments and solving the resulting equations for the unknown parameters.
- This methodology applies in principle also in the case that there are r parameters involved, $\mathbf{\Theta} = \{\theta_1, \ldots, \theta_r\}$, or, as we say, when $\mathbf{\Theta}$ has r coordinates, $r \geq 1$.
- In such a case, we have to assume that the r first moments of the $X_i' s$ are finite; that is, the first k^{th} population moment,

$$m_k\left(\theta_1, \ldots, \theta_r\right) = E\left[X^K\right], \ \left(\theta_1, \ldots, \theta_r\right) \in \Re, \quad k = 1, 2, \ldots, r. \tag{7}$$

- Then form the first k^{th} sample moments

$$\mu_k = \frac{1}{n}\sum_{j=1}^{n} X^k \tag{8}$$

$k = 1, \ldots, r$, and equate Equations (7) and (9) to the corresponding (population) moments; that is,

$$m_k = \mu_k, \ for\ k = 1, \ldots, r \tag{9}$$

- that is, we solve for each parameter by equating $m_1 = \mu_1, \ m_2 = \mu_2, \ldots, \ m_k = \mu_k$
- Assuming that we can solve for $\theta_1, \ldots, \theta_r$ in Equation (9), and that the solutions are unique, we arrive at what we call the moment estimates of the parameters $\theta_1, \ldots, \theta_r$.

Examples of Moment Methods

- **Example 1** Let $X_1, \ldots, X_n$ be a continuous random sample from the $N\left(\mu, \sigma^2\right)$ distribution, with parametric space $\mathbf{\Theta} = \left(\mu, \sigma^2\right)$, where only one of the parameters is known.

$$\mu_1 = \frac{1}{n}\sum_{j=1}^{n} X_i = \bar{X} \text{ and}$$

$$m_1 = E\left[X\right] = \mu$$

Equating $\mu_1 = m_1$

$$\mu_1 = m_1 \Rightarrow \frac{1}{n}\sum_{j=1}^{n} X_i = \mu$$

Therefore, $\hat{\mu} = \frac{1}{n}\sum_{j=1}^{n} X_i = \bar{X}$

Examples of Moments Method cont...

- For moment estimate of σ^2

$$\mu_2 = \frac{1}{n}\sum_{j=1}^{n} X_i^2 \text{ and}$$

$$m_2 = E\left[X^2\right] = \sigma^2 + \mu^2 \quad (\textbf{Verfy!})$$

Equating $\mu_2 = m_2$

$$\begin{aligned}
\mu_2 &= m_2 \\
\frac{1}{n}\sum_{j=1}^{n} X_i^2 &= \sigma^2 + \mu^2 \\
\frac{1}{n}\sum_{j=1}^{n} X_i^2 &= \sigma^2 + \bar{X}^2, \; since\ \bar{X} = \mu \\
\frac{1}{n}\sum_{j=1}^{n} X_i^2 - \bar{X}^2 &= \sigma^2 \\
\frac{1}{n}\sum_{j=1}^{n} \left(X_i - \bar{X}\right)^2 &= \sigma^2
\end{aligned}$$

Thus, $\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n} \left(X_i - \bar{X}\right)^2$

Assessment

1 Let $X_1, \cdots, X_n$ be a discrete random sample from the *Poisson* (λ), $\lambda > 0$ distribution, with parametric space $\mathbf{\Theta} = (\lambda)$.

$$f\left(X_j;\lambda\right) = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!},\ \lambda > 0,\ x_i = 0,1,\cdots\ \textit{and}\ i = 1,2,\cdots,n$$

Determine the MLE of the unknown parameter λ.

2 Let $X_1, \cdots, X_n$ be a continuous random sample from the Negative Exponential distribution *exp* (λ), $\lambda > 0$, with parametric space $\mathbf{\Theta} = (\lambda)$.

$$f\left(X_j;\lambda\right) = \lambda e^{-\lambda x_i},\ \lambda > 0,\ x_i > 0,\ \textit{and}\ i = 1,2,\cdots,n.$$

Derive the the MLE of the unknown parameter λ.

3 Given the pdf

$$f\left(X_j;\theta\right) = \theta^2 X_j e^{-\theta x_i},\ \lambda > 0,\ x_i > 0,\ \textit{and}\ i = 1,2,\cdots,n,$$

Derive the the MLE of the unknown parameter θ.

4 Given the pdf

$$f\left(X_j;\alpha,\beta\right) = \frac{1}{\beta}e^{-\left(X_j-\alpha\right)/\beta},\ \alpha \in \Re,\ \beta > 0,\ x_i > \alpha,\ \textit{and}\ i = 1,2,\cdots,n,$$

Derive the the MLE of the unknown parameter α *and* β when

a α unknown when β is known.

b β unknown when α is known.

- Suppose that $X_1, \cdots, X_n$ is a random sample from an exponential distribution with parameter λ. Now there is only one parameter to estimate. Show that the moment estimator of λ is $\hat{\lambda} = 1/\bar{X}$.

Properties of Point Estimators

- **Note that** we may have several different choices for the point estimator of a parameter. Thus, in order to decide which point estimator of a particular parameter is the best one to use, we need to examine their statistical properties and develop some criteria for comparing estimators.

Properties of Point Estimators

- A point estimator can be evaluated based on:

1. **Unbiasedness** (mean): whether the mean of this estimator is close to the actual parameter?

Properties of Point Estimators

- **Note that** we may have several different choices for the point estimator of a parameter. Thus, in order to decide which point estimator of a particular parameter is the best one to use, we need to examine their statistical properties and develop some criteria for comparing estimators.

Properties of Point Estimators

- A point estimator can be evaluated based on:

1. **Unbiasedness** (mean): whether the mean of this estimator is close to the actual parameter?
2. **Efficiency** (variance): whether the standard deviation of this estimator is close to the actual parameter.

Properties of Point Estimators

- **Note that** we may have several different choices for the point estimator of a parameter. Thus, in order to decide which point estimator of a particular parameter is the best one to use, we need to examine their statistical properties and develop some criteria for comparing estimators.

Properties of Point Estimators

- A point estimator can be evaluated based on:

1. **Unbiasedness** (mean): whether the mean of this estimator is close to the actual parameter?
2. **Efficiency** (variance): whether the standard deviation of this estimator is close to the actual parameter.
3. **Consistency** (size): whether the probability distribution of the estimator becomes concentrated on the parameter as the sample sizes increases

Unbiased Estimators

Definition:

1. **Unbiasedness:** An estimator $\hat{\theta}$ of an unknown parameter θ is unbiased if

$$E\left[\hat{\theta}\right] = \theta, \text{ for } \theta \in \boldsymbol{\Theta}$$

Otherwise, it is a **Biased Estimator** of θ.

Unbiased Estimators

Definition:

1. **Unbiasedness:** An estimator $\hat{\theta}$ of an unknown parameter θ is unbiased if
$$E\left[\hat{\theta}\right] = \theta, \ \text{for } \theta \in \mathbf{\Theta}$$
Otherwise, it is a **Biased Estimator** of θ.
2. **Bias (B)** if an estimator $\hat{\theta}$ of a parameter θ is **biased**, then
$$\textbf{Bias (B)} = E\left[\hat{\theta}\right] - \theta$$
is called **Bias (B)** of $\hat{\theta}$.

- The main point here is that, how an estimator should be close to the true value of the unknown parameter.
- When an estimator is **unbiased**, the **bias is zero**.
- **Example 1:** Suppose that **X** is a random variable with mean μ and variance σ^2. Let $X_1, \cdots, X_n$ be a random sample of size n from the population represented by **X**. Show that $\bar{X}$ and S^{2*} defined in Equations (1) and (3) are unbiased estimator of μ and σ^2, respectively.

Discussion 1:

$$E\left[\bar{X}\right] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} E\left[X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\mu, = \quad \frac{1}{n}(n\mu) = \mu, \ \text{since } E\left[X_i\right] = \mu.$$

- Therefore, $\bar{X}$ is unbiased estimator of the population mean μ.

Discussion 2:

$$
\begin{aligned}
E\left[S^{2*}\right] =& E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right] \\
=& \frac{1}{n-1}E\left[\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right] \\
=& \frac{1}{n-1}E\left[\sum_{i=1}^{n}\left(X_i^2-2\bar{X}X_i+\bar{X}^2\right)\right] \\
=& \frac{1}{n-1}E\left[\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right] \\
=& \frac{1}{n-1}\sum_{i=1}^{n}E\left[X_i^2\right]-nE\left[\bar{X}^2\right], \textit{since } E\left[X_i^2\right]=\mu^2+\sigma^2 \textit{ and } E\left[\bar{X}^2\right]=\mu^2+\frac{\sigma^2}{n}(^{**}) \\
=& \frac{1}{n-1}\sum_{i=1}^{n}\left(\mu^2+\sigma^2\right)-n\left(\mu^2+\frac{\sigma^2}{n}\right) \\
=& \frac{1}{n-1}n\left(\mu^2+\sigma^2\right)-n\left(\mu^2+\frac{\sigma^2}{n}\right) \\
=& \frac{1}{n-1}\left(n-1\right)\sigma^2=\sigma^2
\end{aligned}
$$

- Therefore, S^{2*} is unbiased estimator of the population variance σ^2.

- To show that $E\left[X_i^2\right] = \mu^2 + \sigma^2 (**)$,

$$
\begin{aligned}
Var\left[X_i\right] =& E\left[\left(X_i - \mu\right)^2\right] \\
\sigma^2 =& E\left[X_i^2 - 2\mu X_i + \mu^2\right] \\
=& E\left[X_i^2\right] - 2\mu E\left[X_i\right] + \mu^2 \\
=& E\left[X_i^2\right] - 2\mu^2 + \mu^2 \\
=& E\left[X_i^2\right] - \mu^2 \\
\Rightarrow E\left[X_i^2\right] =& \sigma^2 + \mu^2
\end{aligned}
$$

- To show that $E\left[\bar{X}^2\right] = \mu^2 + \frac{\sigma^2}{n} (**)$,

$$
\begin{aligned}
Var\left[\bar{X}\right] =& E\left[\bar{X}^2\right] - \left(E\left[\bar{X}\right]\right)^2 \\
\frac{\sigma^2}{n} =& E\left[\bar{X}^2\right] - \left(E\left[\bar{X}\right]\right)^2 \\
=& E\left[\bar{X}^2\right] - \mu^2 \\
\Rightarrow E\left[\bar{X}^2\right] =& \mu^2 + \frac{\sigma^2}{n}
\end{aligned}
$$

- **Example 2:** Suppose that **X** is a random variable with mean μ and variance σ^2. Let $X_1, \cdots, X_n$ be a random sample of size n from the population represented by **X**. Show that S^2 defined in Equation (2) is biased estimator of σ^2.

Discussion 3:

$$
\begin{aligned}
E\left[S^2\right] &= E\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right] \\
&= \frac{1}{n}E\left[\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right] \\
&= \frac{1}{n}E\left[\sum_{i=1}^{n}\left(X_i^2-2\bar{X}X_i+\bar{X}^2\right)\right] \\
&= \frac{1}{n}E\left[\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n}E\left[X_i^2\right]-nE\left[\bar{X}^2\right]\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n}\left(\mu^2+\sigma^2\right)-n\left(\mu^2+\frac{\sigma^2}{n}\right)\right], \; see(**)\; above \\
&= \frac{1}{n}\left[n\left(\mu^2+\sigma^2\right)-n\left(\mu^2+\frac{\sigma^2}{n}\right)\right] = \frac{1}{n}\left(n-1\right)\sigma^2 = \sigma^2-\frac{\sigma^2}{n}
\end{aligned}
$$

- Therefore, S^2 is a biased estimator of the population variance σ^2, with $bias = B = -\frac{\sigma^2}{n}$. Bias is negative. MLE for tends to underestimate σ^2

Mean Square Error (MSE) of an Estimator

- Sometimes it is necessary to use a biased estimator. In such cases, the mean square error of the estimator can be important.
- The MSE of an estimator is the expected squared difference between $\hat{\theta}$ and θ.

Definition: Mean Square Error (MSE)

The mean square error of an estimator of the parameter $\hat{\theta}$ is defined as

$$MSE(\hat{\theta}) = E\left[\left(\hat{\theta} - \theta\right)^2\right]$$

Assertion:

The mean square error of $\hat{\theta}$ is equal to the variance of the estimator plus the squared bias. That is,

$$MSE(\hat{\theta}) = Var\left[\hat{\theta}\right] + (Bias)^2$$

Proof:

$$\begin{aligned}
MSE(\hat{\theta}) =& E\left[\left(\hat{\theta} - \theta\right)^2\right] = E\left[\left(\hat{\theta} - E\left[\hat{\theta}\right] + E\left[\hat{\theta}\right] - \theta\right)^2\right] \\
=& E\left[\left(\hat{\theta} - E\left[\hat{\theta}\right]\right)^2 + \left(E\left[\hat{\theta}\right] - \theta\right)^2 - 2\left(\hat{\theta} - E\left[\hat{\theta}\right]\right)\left(E\left[\hat{\theta}\right] - \theta\right)\right] \\
=& E\left[\left(\hat{\theta} - E\left[\hat{\theta}\right]\right)^2\right] + E\left[\left(E\left[\hat{\theta}\right] - \theta\right)^2\right] - \underbrace{2E\left[\left(\hat{\theta} - E\left[\hat{\theta}\right]\right)\left(E\left[\hat{\theta}\right] - \theta\right)\right]}_{=0--------(***)} \qquad (10) \\
=& E\left[\left(\hat{\theta} - E\left[\hat{\theta}\right]\right)^2\right] + E\left[\left(E\left[\hat{\theta}\right] - \theta\right)^2\right] = Var\left[\hat{\theta}\right] + (Bias)^2.
\end{aligned}$$

Efficiency of an Estimator

- The mean square error is an important criterion for comparing two estimators.
- The term efficiency is used as a relative measure of the variance of the sampling distribution, with the efficiency increasing as the variance decreases.
- One may search unbiased estimators to find the one with the smallest variance and call it the most efficient.

Definition: Efficiency

An estimator that has minimum mean square error among all possible unbiased estimators is called **an efficient estimator**.

- The mean square error of an estimator, which is equivalent to the sum of its variance and the square of its bias, can be used as **a relative measure of efficiency (RE)** when comparing two or more estimators.

Definition: Relative Efficiency

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two estimators of the parameter θ, and let $MSE(\hat{\theta}_1)$ and $MSE(\hat{\theta}_2)$ be the mean square errors of $\hat{\theta}_1$ and $\hat{\theta}_2$. Then the **RE** of $\hat{\theta}_2$ to $\hat{\theta}_1$ is defined as

$$RE = \frac{MSE\left(\hat{\theta}_1\right)}{MSE\left(\hat{\theta}_2\right)} \tag{11}$$

Remark:

If this relative efficiency is **less than 1**, we would conclude that $\hat{\theta}_1$ is a more efficient estimator of θ than $\hat{\theta}_2$, in the sense that it **has a smaller mean square error**.

- **Efficiency**

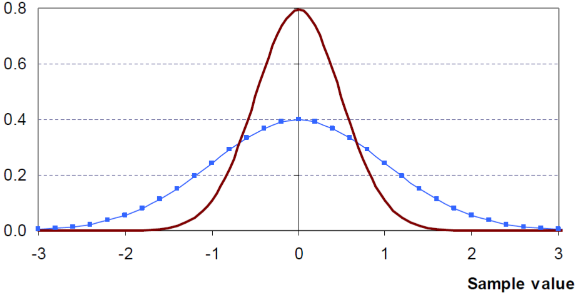


Figure: The density function of the efficient estimator is exemplified by a normal density with ($\sigma = 0.5$). The dotted line indicates a less efficient estimator ($\sigma = 1$).

Example of RE

- **Example** The unbiased estimated mean of of the densities of 40 concrete test cubes is $2445kg/m^3$. However, if we had only the first five test cubes, the second unbiased estimated mean would be $2431kg/m^3$. Hence the relative efficiency, as given by the ratio of the MSE values, bears inversely with the ratio of variances:
$$RE = \frac{MSE(\hat{\theta}_1)}{MSE(\hat{\theta}_2)} = \frac{\sigma^2/n_1}{\sigma^2/n_2} = \frac{\sigma^2/40}{\sigma^2/4} = \frac{1}{8} < 1.$$
- This result confirms what we already know, that is, the large-sample estimator for the mean is more efficient than that based on a smaller sample.
- The efficiency is seen to be proportional to the sample size n.

Consistency of an Estimator

- A consistent estimator of a parameter θ produces statistics that converge to θ, in terms of probability.

Definition: Consistency

- An estimator θ_n , based on a sample size n, is a consistent estimator of a parameter θ, if for any positive number ϵ,

$$\lim_{n\to\infty} Pr\left[\left|\hat{\theta}_n - \theta\right| \leq \epsilon\right] = 1 \tag{12}$$

- As n grows, the estimator will collapse on the true value of the parameter: thus, we do have **asymptotic unbiasedness.**
- One finds, however, that sometimes an unbiased estimator may not be consistent.
- **Example** In Equations (2) and (3) we considered two methods (S^2 *and* S^{2*}) of estimating the variance σ^2.

Assessment

1. Suppose we have independently distributed random samples of size $2n$ from a population denoted by $\mathbf{X}$, and $E[\mathbf{X}] = \mu$ and $Var[\mathbf{X}] = \sigma^2$. Let

$$\bar{X}_1 = \frac{1}{2n} \sum_{i=1}^{2n} X_i \text{ and } \bar{X}_2 = \frac{1}{n} \sum_{i=1}^{n} X_i,$$

be two estimators of μ. Which is the better estimator of μ? Explain your choice.

2. Let $X_1, \ldots, X_n$ denote a random sample from a population having mean μ and variance σ^2. Consider the following estimators of μ:

$$\hat{\theta}_1 = \frac{X_1 + X_2 + \ldots + X_7}{7} \text{ and}$$

$$\hat{\theta}_2 = \frac{2X_1 - X_6 + X_4}{2}.$$

a. Is either estimator unbiased?

b. Which estimator is best? In what sense is it best?

c. Calculate the relative efficiency of the two estimators.

3. Suppose that $\hat{\theta}_1$ and $\hat{\theta}_2$ are estimators of the parameter θ We know that $E\left[\hat{\theta}_1\right] = \theta$, $E\left[\hat{\theta}_2\right] = \theta/2, Var\left[\hat{\theta}_1\right] = 10$ and $Var\left[\hat{\theta}_2\right] = 4$. Which estimator is best? In what sense is it best?

Chapter 2: Parametric Interval Estimation

- So far we have discussed the point estimation of a parameter, or more precisely, point estimation of several real valued parametric functions in the previous chapter.
- In case of continuous distributions the probability that the point estimator actually equaled the value of the parameter being estimated is zero.
- Hence, it seems desirable that a point estimate should be accompanied by some measure of the possible error of estimate.
- For instance, a point estimate must be accompanied by some interval about the point estimate together with some measure of assurance that the true value of the parameter lies within the interval.
- Instead of making the inference of estimating the true value of the parameter to be a point, we might make the inference that the true value of the parameter is contained in some interval. This is called **the problem of interval estimation.**
- In this chapter, methods of parameter estimation called **interval estimation** are introduced.
- An interval estimate for a population parameter θ is called a **confidence interval (CI)**.
- We cannot be certain that the interval contains the true, unknown population parameter—we only use a sample from the full population to compute the point estimate and the interval estimation too.
- However, the confidence interval is constructed so that we have high confidence that it **does contain the unknown population parameter** θ.
- Surprisingly, it is easy to determine such intervals in many cases, and the same data that provided the point estimate are typically used.

Basics of Parametric Interval Estimation

- A confidence interval estimate for θ is an interval of the form $L \leq \theta \leq U$ where the endpoints L and U are statistic computed from the sample data.
- Because different samples will produce different values of L and U, these end-points are values of random variables L and U, respectively.

Parametric Confidence Interval

Suppose that we can determine values of L and U such that the following probability statement is true:

$$Pr\left[L \leq \theta \leq U\right] = 1 - \alpha; \quad where \quad 0 \leq \alpha \leq 1. \tag{13}$$

- There is a probability of $(1 - \alpha)$ of selecting a sample for which the CI will contain the true-value θ.
- The end-points or bounds L and U are called the **lower- and upper-confidence limits**, respectively, and $1 - \alpha$ is called **the confidence coefficient**.

- **Remark: The length of confidence interval** is the difference of **lower- and upper-confidence limits**, given by **U- L**.
- **Example:** Suppose that $X_1, \ldots, X_n$ is a random sample from a normal distribution with unknown mean μ and known variance σ^2. From the results of Chapter 1 we know that the sample mean $\bar{X}$ is normally distributed with mean μ and variance $\frac{\sigma^2}{n}$. We may standardize by subtracting the mean and dividing by the standard deviation, which results:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}. \tag{14}$$

Now Z has **a standard normal distribution, that is** $Z \sim N(0, 1)$.

- Creating a new random variable Z by this transformation is referred to as **standardizing** quantity Z with pdf:

$$f(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \tag{15}$$

which is independent of the true value of the unknown parameters μ and σ^2.

Confidence interval for the mean μ (when σ^2 is known)

- The random variable Z represents the distance of $\bar{X}$ from its mean μ in terms of standard error $\sigma/\sqrt{n}$.
- It is the key step to calculate a probability for an arbitrary normal random variable.
- From Equations (19) and (14) where only one of μ or σ^2 is unknown.
- To construct a confidence interval for it with confidence coefficient $1-\alpha$.

1. **Let μ be unknown**. Consider any two points $L < U$ from the Normal tables for which $Pr[L \le Z \le U] = 1-\alpha$ where $Z \sim N(0,1)$. In particular, for $U = Z_{\alpha/2}$ *and* $L = -Z_{\alpha/2}$

- It follows that:

$$
\begin{aligned}
&Pr[L \le Z \le U] = 1-\alpha \\
&Pr\left[-Z_{\alpha/2} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le Z_{\alpha/2}\right] = 1-\alpha, \ \textit{for all } \mu \\
&Pr\left[-Z_{\alpha/2}\sigma/\sqrt{n} \le \bar{X}-\mu \le Z_{\alpha/2}\sigma/\sqrt{n}\right] = 1-\alpha \\
&Pr\left[-\bar{X}-Z_{\alpha/2}\sigma/\sqrt{n} \le -\mu \le -\bar{X}+Z_{\alpha/2}\sigma/\sqrt{n}\right] = 1-\alpha \\
&Pr\left[\bar{X}-Z_{\alpha/2}\sigma/\sqrt{n} \le \mu \le \bar{X}+Z_{\alpha/2}\sigma/\sqrt{n}\right] = 1-\alpha.
\end{aligned}
$$

Definition: Confidence Interval for the unknown mean parameter μ (when σ^2 is known)

Let $\bar{X}$ be the mean of a random sample of size n drawn from a normal population with known standard deviation σ. The $100(1-\alpha)\%$ central two-sided confidence interval for the population mean μ is given by:

$$Pr\left[\bar{X}-Z_{\alpha/2}\sigma/\sqrt{n} \le \mu \le \bar{X}+Z_{\alpha/2}\sigma/\sqrt{n}\right] = 1-\alpha. \quad (16)$$

$$\textit{That is, } \mu \textit{ lies in the interval } \left(\bar{X}-Z_{\alpha/2}\sigma/\sqrt{n},\ \bar{X}+Z_{\alpha/2}\sigma/\sqrt{n}\right)$$

Plot of two sided CI for Standard Normal

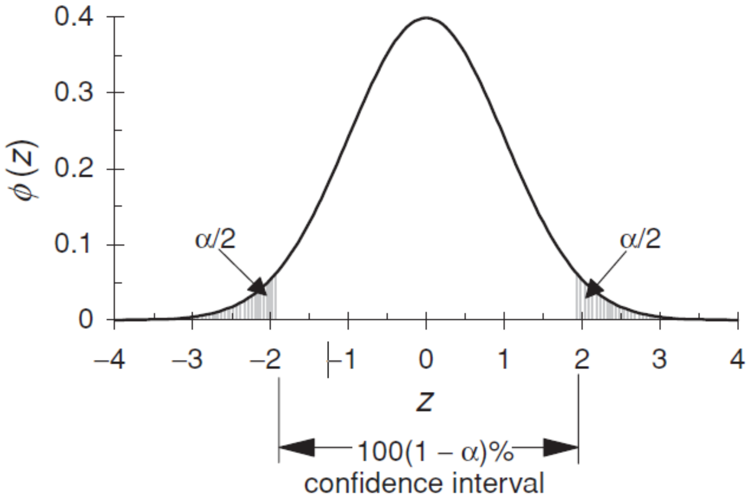
- 

Figure: Standard Normal pdf showing two-sided Confidence Interval.

Confidence Interval for the Variance σ^2 (when μ is known)

2. Let σ^2 be unknown. Set $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\mu)^2$.

- Recall that $\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n-1}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2_{(n-1)}$.
- From the Chi-Square tables, determine any pair $0 < L < U$ for which $Pr\,[L \le X \le U] = 1-\alpha$, where $X \sim \chi^2_{n-1}$. Then we have

$$Pr\left[L \le \frac{(n-1)S^2}{\sigma^2} \le U\right] = 1-\alpha, \ \textit{for all } \sigma^2$$

$$Pr\left[\frac{1}{U} \le \frac{\sigma^2}{(n-1)S^2} \le \frac{1}{L}\right] = 1-\alpha$$

$$Pr\left[\frac{(n-1)S^2}{U} \le \sigma^2 \le \frac{(n-1)S^2}{L}\right] = 1-\alpha$$

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

Definition: Confidence Interval for Variance, σ^2 (when μ is known)

- Let $\hat{S}^2$ be the variance of a random sample of size n drawn from a normal distribution with unknown variance. The $100(1-\alpha)\%$ percent equi-tailed two-sided confidence interval for the population variance σ^2 is as follows:

$$Pr\left[\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}\right] = 1-\alpha;\ \textit{or} \tag{17}$$

That is, σ^2 *lies in the interval* $\left(\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}\right)$,

where $\chi^2_{n-1,\alpha/2}$ and $\chi^2_{n-1,1-\alpha/2}$ are the values that a χ^2_{n-1} variate exceeds with probabilities $\alpha/2$ and $(1-\alpha/2)$, respectively.

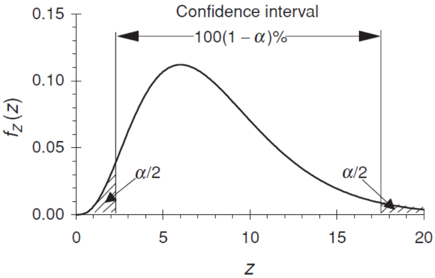


Figure: Equal-tails Confidence Interval for Variance (chi-squared distribution).

Remark:

The corresponding one-sided upper confidence limit for σ^2 is defined as:

$$\left[\sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha}}\right] = 1-\alpha. \tag{18}$$

- **Example:** The compressive strengths of 40 test cubes of concrete samples with the sample mean and sample standard deviation of 60.14 and 5.02 N/mm^2, respectively. We also assume that the compressive strengths are normally distributed. To facilitate the application, let us assume that the estimated standard deviation of $5.02N/mm^2$ is the true on known value.

a. Construct a 95% confidence interval for the population mean μ.
b. Construct an upper one-sided 99% confidence limit for the population variance.
c. Construct a 95% two-sided confidence limit for the population variance.

- **Discussion:** Given: n=40, $\bar{X} = 60.4$ and $\hat{S} = 5.02$

a. From standardized normal Table $Z_{\alpha/2} = Z_{0.025} = 1.96$. Using Equation (16), we have

$$Pr\left[\bar{X} - Z_{\alpha/2}\hat{S}/\sqrt{n} \leq \mu \leq \bar{X} + Z_{\alpha/2}\hat{S}/\sqrt{n}\right] = 1-\alpha$$

$$Pr\left[60.4 - 1.96 * 5.02/\sqrt{40} \leq \mu \leq 60.4 + 1.96 * 5.02/\sqrt{40}\right] = 0.95$$

$$Pr\left[58.58 \leq \mu \leq 61.70\right] = 0.95$$

- **Therefore we are** 95% **confident that the interval** (58.58, 61.70) **includes the true population mean** μ**.**
- **The length of confidence interval** is $61.70 - 58.58 = 3.12$.

Example Cont...

b. From χ^2 Table $\chi^2_{n-1,\ \alpha} = \chi^2_{39,\ 0.01} = 21.426$. Using Equation (18), we have

$$Pr\left[\sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha}}\right] = 1-\alpha$$

$$Pr\left[\sigma^2 \leq \frac{39(5.02)^2}{\chi^2_{39,0.01}}\right] = 0.99$$

$$Pr\left[\sigma^2 \leq \frac{39(25.2004)}{21.426}\right] = 0.99 \Rightarrow Pr\left[\sigma^2 \leq 45.87\right] = 0.99.$$

Hence the 99% upper confidence limit for σ is $6.76 N/mm^2$.

c. From χ^2 Table $\chi^2_{n-1,\ \alpha/2} = \chi^2_{39,\ 0.025} = 58.120$ and $\chi^2_{n-1,\ 1-\alpha/2} = \chi^2_{39,\ 0.975} = 23.654$. Using Equation (17), we have

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

$$Pr\left[\frac{(39)(25.2004)}{\chi^2_{39,\ 0.025}} \leq \sigma^2 \leq \frac{(39)(25.2004)}{\chi^2_{39,0.975}}\right] = 0.95$$

$$Pr\left[\frac{982.816}{58.120} \leq \sigma^2 \leq \frac{982.816}{23.654}\right] = 0.95 \Rightarrow Pr\left[16.9 \leq \sigma^2 \leq 41.55\right] = 0.95.$$

Hence the 95% two-sided confidence limit for σ is $(4.11, 6.45)$ *in* N/mm^2.

- This interval is fairly wide because there is a lot of variability in the compressive strengths of cubes of concrete samples measured.

Simultaneous Confidence Interval for the Mean and Variance (Small Sample)

- In **Example 1 and 2**, the position was adopted that only one of the parameters in the $N\left(\mu, \sigma^2\right)$ distribution was unknown. In practice, both μ and σ^2 are most often unknown.
- In this sub section we try pave the way to solving the problem.

Simultaneous Confidence Interval for the Mean

- Let $X_1, \ldots, X_n$ be a random sample from the $N\left(\mu, \sigma^2\right)$ distribution, where both μ and σ^2 are unknown.
- To construct confidence intervals for μ and σ^2, each with confidence coefficient $1-\alpha$, we have

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ and}$$

$$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n-1}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2_{n-1},$$

where $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\mu)^2$ and these two r.v.'s are independent.

- It follows that their ratio

$$\frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}} = \frac{(\bar{X}-\mu)}{S/\sqrt{n}} \sim t_{n-1}\ (t-\textit{distribution with } (n-1)\ D.F.).$$

Simultaneous Confidence Interval cont...

- As usual from the t-tables, determine any pair L and U with $L < U$ such that $P[L \leq X \leq U] = 1 - \alpha$, where $X \sim t_{n-1}$.
- Let $L = -t_{n-1;\alpha/2}$ *and* $U = t_{n-1;\alpha/2}$ it follows that:

$$P\left[L \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq U\right] = 1-\alpha$$

$$P\left[-t_{n-1;\alpha/2} \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq t_{n-1;\alpha/2}\right] = 1-\alpha$$

$$P\left[-t_{n-1;\alpha/2}\frac{S}{\sqrt{n}} \leq \bar{X}-\mu \leq t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}\right] = 1-\alpha$$

$$P\left[\bar{X}-t_{n-1;\alpha/2}\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X}+t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}\right] = 1-\alpha \qquad (19)$$

Definition:

If $\bar{X}$ and S are the sample mean and sample standard deviation of a random samples $X_1, X_2, \cdots, X_n$ from a normal distribution with unknown variance σ^2, a $100(1-\alpha)\%$ confidence interval for population mean μ is

$$\left(\bar{X}-t_{n-1;\alpha/2}\frac{S}{\sqrt{n}},\ \ \bar{X}+t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}\right)$$

where $t_{n-1;\alpha/2}$ is the upper $100\alpha/2$ percentage point of the t distribution with $n-1$ degrees of freedom.

- One-sided confidence bounds for the mean of a $t-$ distribution are also of interest and are simply to use only the appropriate lower or upper confidence limit from Equation (19) and replace $t_{n-1;\alpha/2}$ by $t_{n-1;\alpha}$.

Simultaneous Confidence Interval cont...

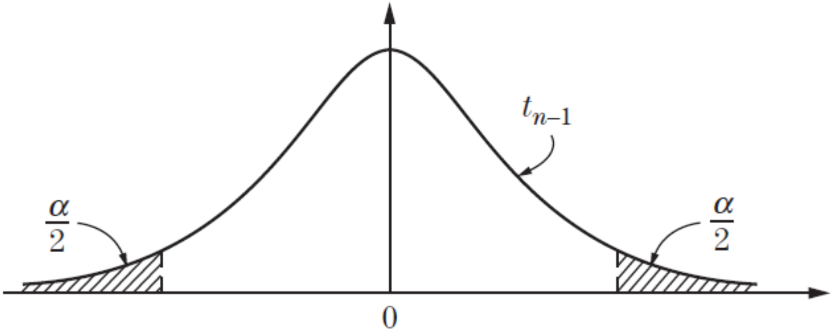
- 

Figure: Two-Sided Confidence Interval for the Population Mean Using t-Distribution

Confidence Interval for σ^2

- The construction of a confidence interval for σ^2 in the presence of (an unknown) μ is easier.
- We have already mentioned that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$ and repeat the process to obtain the confidence interval as

$$\left(\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}} \right), \tag{20}$$

where, $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$

- **Note that:** The confidence interval in Equation (16) is different from Equation (19) in that σ in Equation (16) is replaced by an estimate S, and then the constant $Z_{\alpha/2}$ in Equation (16) is adjusted to $t_{n-1;\alpha/2}$.
- Likewise, the confidence intervals in Equation (17) and (20) are of the same form, with the only difference that (the unknown) μ in Equation (17) is replaced by its estimate $\bar{X}$ in Equation (20).

Example

- An article in the Journal of Heat Transfer (Trans. ASME, Sec. C, 96, 1974, p. 59) described a new method of measuring the thermal conductivity of Armco iron. Using a temperature of $100^0 F$ and a power input of $550 watts$, the following 10 measurements of thermal conductivity (in $Btu/hr - ft -^0 F$) were obtained:

$$41.60, 41.48, 42.34, 41.95, 41.86, 42.18, 41.72, 42.26, 41.81, 42.04$$

- A point estimate of the sample mean thermal conductivity at $100^0 F$ and 550 watts is
$$\bar{X} = \frac{41.60+41.48+\cdots+42.04}{10} = 41.924 Btu/hr - ft -^0 F$$

- And a point estimate of the sample standard deviation is:

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$
$$= \sqrt{\frac{(41.60-41.924)^2 + (41.48-41.924)^2 + \cdots + (42.04-41.924)^2}{9}} = 0.284 Btu/hr -$$

- The estimated standard error of $\bar{X}$ is
$$\sigma_{\bar{X}} = \frac{S}{\sqrt{n}} = \frac{0.284}{\sqrt{10}} = 0.0898$$

- If we can assume that thermal conductivity is normally distributed, σ^2 is unknown and n=10 is small it is advisable to use $t - distribution$.

- From the student t-distribution table $t_{n-1,\alpha/2} = t_{9,0.025} = 2.262$

- Thus, at 95% confidence level the true mean μ of thermal conductivity is with the interval
$$\bar{X} \pm t_{n-1,\alpha/2}\frac{S}{\sqrt{n}} = 41.924 \pm 2.262(0.0898) = (41.721, \;\; 42.127).$$

Example cont...

- To construct a 95% confidence interval for σ^2, $\chi^2_{n-1,\alpha/2} = \chi^2_{9,0.025} = 19.02$ and $\chi^2_{n-1,1-\alpha/2} = \chi^2_{9,0.975} = 2.70$

$$\left(\frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}} , \frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}} \right)$$
$$= \left(\frac{9(0.284^2)}{\chi^2_{9,0.025}} , \frac{9(0.284^2)}{\chi^2_{9,0.975}} \right)$$
$$= \left(\frac{0.726}{19.02} , \frac{0.726}{2.70} \right) = (0.038 \ , \ 0.269)$$

- This last expression may be converted into a confidence interval on the standard deviation σ by taking the square root of both sides, resulting in $(0.195 \ , \ 0.518)$.
- Therefore, at the 95% level of confidence, the thermal conductivity data indicate that the process standard deviation could be as small as 0.195 $Btu/hr - ft -^0 F$ and large as 0.518 $Btu/hr - ft -^0 F$

- It is often necessary to construct confidence intervals on a population proportion.

Population Proportion

- Suppose that a random sample of size n has been taken from a large (possibly infinite) population and that $X \leq n$ observations in this sample belong to a class of interest. Then

$$\hat{P} = \frac{X}{n} \tag{21}$$

is a point estimator of the proportion of the population p that belongs to this class, where n and p are the parameters of **a binomial distribution**.

- The sampling distribution of $\hat{P}$ is approximately normal with mean p and variance if $p\,(p-1)$ is not too close to either 0 or 1 and if n is relatively large.
- Assuming that np and $n(p-1)$ are greater than 5,

$$\begin{aligned} Z &= \frac{X - np}{\sqrt{np(p-1)}} \\ &= \frac{\frac{X}{n} - p}{\sqrt{\frac{p(p-1)}{n}}}, \ \textit{devide by } n \\ &= \frac{\hat{P} - p}{\sqrt{\frac{p(p-1)}{n}}} \approx N\,(0,\ 1) \end{aligned}$$

A Large-Sample CI for Population Proportion

- To construct the confidence interval on p, note that

$$Pr\left[-Z_{\alpha/2} \leq Z \leq Z_{\alpha/2}\right] = 1 - \alpha$$

$$Pr\left[-Z_{\alpha/2} \leq \frac{\hat{P} - p}{\sqrt{\frac{p(p-1)}{n}}} \leq Z_{\alpha/2}\right] = 1 - \alpha$$

$$Pr\left[\hat{P} - Z_{\alpha/2}\sqrt{\frac{p(p-1)}{n}} \leq p \leq \hat{P} + Z_{\alpha/2}\sqrt{\frac{p(p-1)}{n}}\right] = 1 - \alpha$$

Definition: Confidence Interval for Population Proportion, $\hat{P}$

A $100(1-\alpha)\%$ confidence interval for p is

$$\left(\hat{P} - Z_{\alpha/2}\sqrt{\frac{p(p-1)}{n}},\ \hat{P} + Z_{\alpha/2}\sqrt{\frac{p(p-1)}{n}}\right) \tag{22}$$

where the quantity $\sqrt{\frac{p(p-1)}{n}}$ in Equation (22) is called **the standard error of the point estimator** $\hat{P}$.

Example

- In a random sample of 85 automobile engine crankshaft bearings, 10 have a surface finish that is rougher than the specifications allow. Construct A 95% two-sided confidence interval for p.
- **Discussion:**
- A point estimate of the proportion of bearings in the population that exceeds the roughness specification is

$$\hat{P} = \frac{10}{85} = 0.12$$

and

$$\hat{P} - Z_{0.025}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \leq p \leq \hat{P} + Z_{0.025}\sqrt{\frac{\hat{P}(\hat{P}-1)}{n}}$$

$$0.12 - 1.96\sqrt{\frac{0.12(0.88)}{85}} \leq p \leq 0.12 + 1.96\sqrt{\frac{0.12(0.88)}{85}}$$

$$0.05 \leq p \leq 0.19$$

Confidence Interval for the Difference Between two Samples Means cont...

- **Assumption**: $\sigma_1^2 = \sigma_2^2 = \sigma^2$; that is, we have to assume that the variances, although unknown, are equal.
- Recall that: $\bar{X} - \mu_1 \sim N\left(0, \ \frac{\sigma_1^2}{m}\right)$ and $\bar{Y} - \mu_2 \sim N\left(0, \ \frac{\sigma_2^2}{n}\right)$.
- By independence of X and Y it follows that:

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim N(0, \ 1) \tag{23}$$

- Further recall that if $S_X^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X})^2$ and $S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$, then $\frac{(m-1)S_X^2}{\sigma^2} \sim \chi^2_{(m-1)}$ and $\frac{(n-1)S_Y^2}{\sigma^2} \sim \chi^2_{(n-1)}$.
- By independence of X and Y,

$$\frac{(m-1)S_X^2 + (n-1)S_Y^2}{\sigma^2} \sim \chi^2_{(m+n-2)} \tag{24}$$

- From Equations (23) and (24)

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}\left(\frac{1}{m} + \frac{1}{n}\right)}} \sim t_{m+n-2}. \tag{25}$$

Definition: CI for Difference of Two Means from Two Independent Population

- Let $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ be **two independent random samples** from the $N\left(\mu_1,\ \sigma_1^2\right)$ and $N\left(\mu_2,\ \sigma_2^2\right)$ distributions, respectively, with all $\mu_1,\ \mu_2,\ \sigma_1^2$ *and* σ_2^2 unknown.
Thus, from a t-distribution in Equation (25), the confidence interval for the difference of the true parameter means $(\mu_1 - \mu_2)$ is

$$(\bar{X} - \bar{Y}) \pm t_{m+n-2,\alpha/2}\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}\left(\frac{1}{m} + \frac{1}{n}\right)} \tag{26}$$

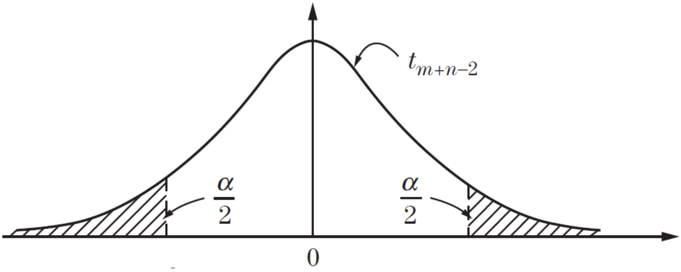


Figure: Confidence Interval for the Difference of Two Means

CI for Ratio of Variances, $\frac{\sigma_1^2}{\sigma_2^2}$ for Two Samples from Independent Normal Populations

- Recall once more that $\frac{(m-1)S_X^2}{\sigma_1^2} \sim \chi^2_{(m-1)}$ and $\frac{(n-1)S_Y^2}{\sigma_2^2} \sim \chi^2_{(n-1)}$.
- By independence of X and Y,
$$\frac{\frac{S_X^2}{\sigma_1^2}}{\frac{S_Y^2}{\sigma_2^2}} = \frac{\sigma_1^2}{\sigma_2^2} \times \frac{S_Y^2}{S_X^2} \sim F_{n-1,m-1}. \tag{27}$$
- From the F-tables, determine any pair (L, U) with $0 < L < U$ such that $P(L \leq X \leq U) = 1 - \alpha$, where $X \sim F_{n-1,m-1}$.
- Then,
$$Pr\left[L \leq \frac{\sigma_1^2}{\sigma_2^2} \times \frac{S_Y^2}{S_X^2} \leq U\right] = 1 - \alpha$$
$$Pr\left[L\frac{S_X^2}{S_Y^2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq U\frac{S_X^2}{S_Y^2}\right] = 1 - \alpha$$
$$Pr\left[F_{n-1,m-1;1-\alpha/2}\frac{S_X^2}{S_Y^2} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq F_{n-1,m-1;\alpha/2}\frac{S_X^2}{S_Y^2}\right] = 1 - \alpha$$

CI for Ratio of Variances $\frac{\sigma_1^2}{\sigma_2^2}$

Let $X_1, \cdots, X_m$ and $Y_1, \cdots, Y_n$ be **two independent random samples** from the $N(\mu_1, \ \sigma_1^2)$ and $N(\mu_2, \ \sigma_2^2)$ distributions, respectively, with all $\mu_1, \ \mu_2, \ \sigma_1^2$ *and* σ_2^2 unknown.

A $100(1-\alpha)\%$ confidence interval required for $\frac{\sigma_1^2}{\sigma_2^2}$ is then,
$$\left(F_{n-1,m-1;1-\alpha/2}\frac{S_X^2}{S_Y^2}, \ F_{n-1,m-1;\alpha/2}\frac{S_X^2}{S_Y^2}\right). \tag{28}$$

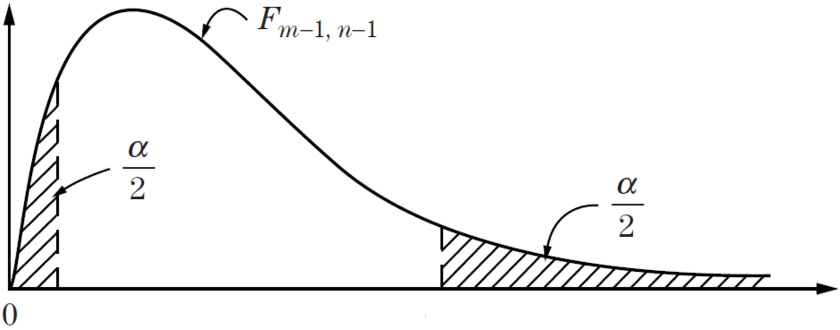


Figure: Confidence Interval for Ratio of Variances from two Independent Population (F-Distribution)

Example

- The summary statistics given below from two catalysts types in which 8 samples in the pilot plant are take from each are being analyzed to determine how they affect the mean yield of a chemical process. Specifically, the 1^{st} catalyst is currently in use, but the 2^{nd} catalyst is acceptable.

Observation Number	Catalyst 1	Catalyst 2
1	91.50	89.19
2	94.18	90.95
3	92.18	90.46
4	95.39	93.21
5	91.79	97.19
6	89.07	97.04
7	94.72	91.07
8	89.21	92.75
	$\bar{x}_1 = 92.255$	$\bar{x}_2 = 92.733$
	$s_1 = 2.39$	$s_2 = 2.98$

Table: Catalyst Yield Data

Discussion

- **Construct a confidence interval for difference between the mean yields. Use $\alpha = 0.05$, and assume equal variances.**
- Using Equation (26) and $t_{m+n-2,\alpha/2} = t_{14,0.025} = 2.145$, m=n=8, m+n-2=14.

$$
\begin{aligned}
&(\bar{X} - \bar{Y}) \pm t_{m+n-2,\alpha/2}\sqrt{\frac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}\left(\frac{1}{m} + \frac{1}{n}\right)} \\
&= (92.225 - 92.733) \pm t_{14;0.025}\sqrt{\frac{7(2.39)^2 + 7(2.98)^2}{14}\left(\frac{1}{8} + \frac{1}{8}\right)} \\
&= -0.478 \pm 2.145\sqrt{\frac{7(2.39^2) + 7(2.98)^2}{14}\left(\frac{1}{8} + \frac{1}{8}\right)} \\
&= -0.478 \pm 2.145(1.350) \\
&= -0.478 \pm 2.897 = (-3.375,\ 2.419)\,.
\end{aligned}
$$

- **Construct a confidence interval for the ratio variance $\frac{\sigma_1^2}{\sigma_2^2}$ of yields. Use $\alpha = 0.05$.**
- Using Equation (28) and $F_{n-1,m-1;1-\alpha/2} = F_{7,7;0.975} = \frac{1}{F_{7,7,0.025}} = \frac{1}{4.99} = 0.200$ and $F_{n-1,m-1;\alpha/2} = F_{7,7,0.025} = 4.99$.

$$
\left(F_{n-1,m-1;1-\alpha/2}\frac{S_X^2}{S_Y^2},\ F_{n-1,m-1;\alpha/2}\frac{S_X^2}{S_Y^2}\right) = \left(F_{7,7;0.975}\frac{2.39^2}{2.98^2},\ F_{7,7,0.025}\frac{2.39^2}{2.98^2}\right)
$$

$$
= (0.200(0.643), 4.99(0.643)) = (0.129, 3.209)
$$

Assessment

1. Let $\bar{X} = 102$ and that $n = 50$ *and* $\sigma^2 = 10$. What is a 95% confidence interval for μ?
2. A survey was made in the core course, asking (among other things) the annual salary of the jobs that the students had before enrolling as a full–time PhD students. Here is a subset ($n = 10$) of those responses (in thousands of dollars):

   $$20,\ 34,\ 52,\ 21,\ 26,\ 29,\ 71,\ 41,\ 23,\ 67$$

a. Construct a 95% confidence interval for the true average income for incoming full–time PhD students.

b. Construct a 95% confidence interval for the true standard deviation income for incoming full–time PhD students.

3. A forester wishes to estimate the average number of "count trees" per acre (trees larger than a specified size) on a 2,000-acre plantation. She can then use this information to determine the total timber volume for trees in the plantation. A random sample of $n = 50$ one-acre plots is selected and examined. The average (mean) number of count trees per acre is found to be **27.3**, with a standard deviation of **12.1**. Use this information to construct a 99% confidence interval for μ, the mean number of count trees per acre for the entire plantation.

Chapter 3: Basics of Hypothesis Testing

- In the previous chapter we illustrated how to construct a confidence interval estimate of a parameter from sample data.
- However, many problems in decision making require that we decide whether **to accept or reject a statement about some parameter**. The statement is called a **hypothesis**, and the action of decision-making procedure about the hypothesis is called **hypothesis testing**.
- **Definition: A statistical hypothesis** is a statement about the parameters of one or more populations.
- A random sample is taken from the population and statistical hypotheses, called **null** and its **alternative**, are declared. Then a statistical test is made.
- If the observed random sample do not support the model or theory postulated, **the null hypothesis is rejected in favor of the alternative one**, which may be considered to be true.
- However, if the observations are in agreement, then the **null hypothesis is not rejected**. **This does not necessarily mean that it is accepted**. **It suggests that there is insufficient evidence from the data against the null hypothesis in favor of the alternative one**.

Two-Sided Hypotheses Test

- A test of any hypothesis such as
$$H_0 : \hat{\theta} = \theta_0 \text{ versus } H_a : \hat{\theta} \neq \theta_0$$
is called **a two-sided test**, because it is important to detect differences from the hypothesized value θ_0 of the parameter that lie on either side of θ_0.
- In such a test, the critical region is split into two parts, with (usually) **equal probability placed in each tail of the distribution of the test statistic**.
- For example, if say Z_0 is standardized normally distributed random variable, then the critical regions can be visualized as in the figure follows.

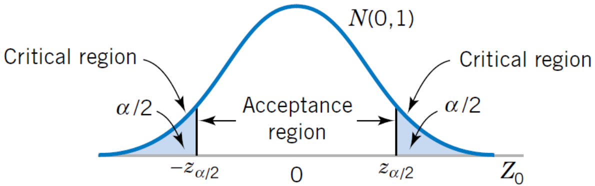


Figure: The distribution of Z when $H_0 : \hat{\mu} = \mu_0$ is true, with critical region for the two-sided alternative $H_a : \hat{\mu} \neq \mu_0$

One-Sided Hypotheses Test

- We may also develop procedures for testing hypotheses on the mean μ where the alternative hypothesis is one-sided.

$$H_0 : \hat{\theta} = \theta_0 \text{ versus } H_a : \hat{\theta} < \theta_0 \text{ or}$$
$$H_0 : \hat{\theta} = \theta_0 \text{ versus } H_a : \hat{\theta} > \theta_0$$

- If the alternative hypothesis is $H_a : \hat{\theta} > \theta_0$, the critical region should lie in the upper tail of the distribution of the test statistic, whereas if the alternative hypothesis is $H_a : \hat{\theta} < \theta_0$, the critical region should lie in the lower tail of the distribution. Consequently, these tests are called **one-tailed tests**.

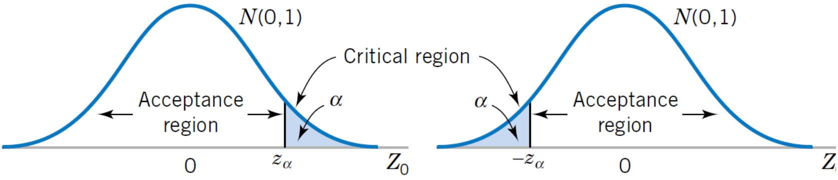


Figure: Critical Regions for One-sided Alternative $H_a : \hat{\theta} > \theta_0$ (left), and the One-sided Alternative $H_a : \hat{\theta} < \theta_0$ (right), for Standardized Normal Distributed Z.

Rejection Regions

- **Critical values:** The values of the test statistic that separate the rejection and non-rejection regions. They are the boundary values obtained corresponding to the preset level.
- **Rejection region:** The set of values for the test statistic that leads to rejection of H_0.
- **Non-rejection region:** the set of values not in the rejection region that leads to non-rejection of H_0.

The Procedure for Hypothesis Tests

- As just outlined, hypothesis testing concerns one or more parameters and also the related probability distribution. **The basic steps, in applying hypothesis-testing methodology** is recommended as follows.

1. From the problem context, identify the parameter of interest.
2. State the null hypothesis, H_0 in terms of a population parameter, such as μ or σ^2.
3. Specify an appropriate alternative hypothesis, H_a in terms of the same population parameter.

4 Choose a significance level, α.

5. Determine an appropriate test statistic, substituting quantities given by the null hypothesis but not the observed values. State what statistical distribution is being used, so that we may need to make an assumption regarding the underlying distribution.
6. Compute any necessary sample quantities, assuming that the null hypothesis is true and substitute these into the equation for the test statistic.

7 State **the rejection region** or also called **a critical region**, for the test statistic.

8. Decide whether or not H_0 should be rejected and report that in the problem context based on the observed level of significance p-value.
9. State a conclusion, that might be either to accept the null hypothesis, or else to reject the null hypothesis in favour of the alternative hypothesis.

Types of Possible Error

- We may decide to take some action on the basis of the test of significance, such as adjusting the process if a result is statistically significant. **But we can never be completely certain we are taking the right action**.
- There are two types of possible error which we must consider.

Table: *Types of Possible Error*

	H_0 True	H_0 False
Fail to reject H_0	Correct Decision	Type II error
Reject H_0	Type I error	Correct Decision

- **The type I error** specification is the probability of making errors **when the null hypothesis is true**. This specification is commonly represented with the symbol α.
- For example if we say that a test has $\alpha \leq 0.05$ we guarantee that if the null hypothesis is true the test will not make more than $1/20$ mistakes.
- **P(of type I error) = P(of rejecting H_0 whereas H_0 is true)**$=\alpha$ (**the significance level**).
- **The type II error** specification is the probability of making errors **when the null hypothesis is false**. This specification is commonly represented with the symbol β.
- P(of type II error) = P(of accepting H_0 whereas H_0 is false)$=\beta$.
- For example if we say that for a test β is unknown we say that we cannot guarantee how it will behave when the null hypothesis is actually false.
- **The power of test:** specification is the probability of correctly rejecting the null hypothesis when it is false, or **the power of a test** is the probability of making the correct decision when the alternative hypothesis is true.
- **Thus the power specification is** $1-\beta$.

P-Value

- **The p-value is the probability that the null hypothesis is true** (based on the data) or p-value is **the smallest significance level at which the null hypothesis can be rejected**.
- We noted previously that reporting the results of a hypothesis test in terms of a P-value is very useful because it conveys more information than just the simple statement "reject H_0" or "fail to reject H_0".
- The p-value is a number between **0** and **1** that represents a probability.
- The observed level of significance or p-value is the probability of obtaining a result as far away from the expected value as the observation is, or farther, purely by chance, when the null hypothesis is true.
- Notice that a smaller observed level of significance indicates that the null hypothesis is less likely.
- If this observed level of significance is small enough, we conclude that the null hypothesis is not plausible.
- In many instances we choose a critical level of significance before observations are made.
- The most common choices for the critical level of significance are 10%, 5%, and 1%.
- If the observed level of significance is smaller than a particular critical level of significance, we say that the result is statistically significant at that level of significance.
- If the observed level of significance is not smaller than the critical level of significance, we say that the result is not statistically significant at that level of significance.

Test of the Mean of Normal Population when the Variance of the Population is known.

- Assume that a random sample $X_1, X_2, \ldots, X_n$ has been taken from the population. Based on our previous discussion, the sample mean $\bar{X}$ is an **unbiased point estimator** of μ with variance σ^2/n.

1. **The test of hypotheses:**
$$H_0 : \hat{\mu} = \mu_0 \text{ versus } H_a : \hat{\mu} \neq \mu_0$$
where μ_0 is a specified constant.
2. **The test statistic:**
$$Z_{cal} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \tag{29}$$
3. If the null hypothesis $H_0 : \hat{\mu} = \mu_0$ is true,$E\left[\bar{X}\right] = \mu_0$, and it follows that the distribution of $Z \sim N(0, 1)$ is the standard normal the probability is $1 - \alpha$ that the test statistic Z_{cal} falls between $-Z_{\alpha/2}$ and $Z_{\alpha/2}$, where $Z_{\alpha/2}$ is the $100\alpha/2$ percentage point of the standard normal distribution. That is,
$$Pr\left[-Z_{\alpha/2} \leq Z_{cal} \leq Z_{\alpha/2}\right] = 1 - \alpha.$$
4. **Rejection Region**: Reject H_0 if the observed value of the test statistic Z_{cal} is either $Z_{cal} > Z_{\alpha/2}$ or $Z_{cal} < -Z_{\alpha/2}$.

Example

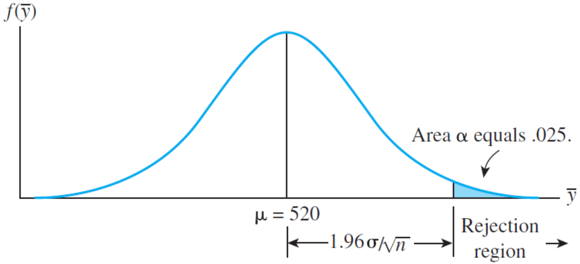
- The Texas A & M agricultural extension service wants to determine whether the mean yield per acre (in bushels) for a particular variety of soybeans has increased during the current year over the mean yield in the previous 2 years when $\mu = 520$ bushels per acre. The research statement is that yield in the current year has increased above 520 with $\alpha = 0.025$ confidence level. Suppose we have decided to take a sample of $n = 36$ one-acre plots, and from these data we compute $\bar{y} = 573$ and $S = 124$. Can we conclude that the mean yield for all farms is above 520?
- **Discussion:** Assume that σ can be estimated by S.

1. **The test of hypothesis**:

$$H_0 : \hat{\mu} \leq 520 \text{ versus } H_a : \hat{\mu} > 520$$

2. **The test statistic:**

$$Z_{cal} = \frac{\bar{y} - \mu_0}{S/\sqrt{n}} = \frac{573 - 520}{124/\sqrt{36}} = \frac{573 - 520}{124/\sqrt{36}} = 2.56$$



3. **Rejection region:** $Z_{cal} = 2.56 > Z_{tabulated} = Z_{\alpha} = Z_{0.025} = 1.96$
4. **Conclusion:** We reject the null hypothesis in favor of the research hypothesis and conclude that **the average soybean yield per acre is greater than 520**.

Summary of Statistical Test of Hypothesis of the Mean for Large $n \geq 30$, (σ is known)

Hypotheses:

Case 1. H_0: $\mu \leq \mu_0$ vs. H_a: $\mu > \mu_0$ (right-tailed test)
Case 2. H_0: $\mu \geq \mu_0$ vs. H_a: $\mu < \mu_0$ (left-tailed test)
Case 3. H_0: $\mu = \mu_0$ vs. H_a: $\mu \neq \mu_0$ (two-tailed test)

T.S.: $z = \dfrac{\bar{y} - \mu_0}{\sigma/\sqrt{n}}$

R.R.: For a probability α of a Type I error,

Case 1. Reject H_0 if $z \geq z_\alpha$.
Case 2. Reject H_0 if $z \leq -z_\alpha$.
Case 3. Reject H_0 if $|z| \geq z_{\alpha/2}$.

Note: These procedures are appropriate if the population distribution is normally distributed with σ known. In most situations, if $n \geq 30$, then the Central Limit Theorem allows us to use these procedures when the population distribution is nonnormal. Also, if $n \geq 30$, then we can replace σ with the sample standard deviation s. The situation in which $n < 30$ is presented later in this chapter.

.

Hypothesis Testing on the Mean of a Population with Unknown Variance σ^2

- The important point upon which the test procedure relies is that if $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with mean μ and unknown variance σ^2, then the random variable

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \tag{30}$$

has a t distribution with $n - 1$ degrees of freedom.

- Based on our previous discussion, the sample mean $\bar{X}$ is an **unbiased point estimator** of μ and estimated sample standard deviation $S/\sqrt{n}$ we have:

1. **The test of hypotheses:**

$$H_0 : \hat{\mu} = \mu_0 \text{ versus } H_a : \hat{\mu} \neq \mu_0$$

where μ_0 is a specified constant.

2. **The test statistic:**

$$T_c = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \tag{31}$$

- The **critical region** to control the type I error probability at the desired level, the t percentage points $t_{\alpha/2,\ n-1}$ and as the boundaries of the critical region $-t_{\alpha/2,\ n-1}$ and $t_{\alpha/2,\ n-1}$ so that we would reject $H_0 : \hat{\mu} = \mu_0$ if

$$T_c > t_{\alpha/2,\ n-1} \text{ or } T_c < -t_{\alpha/2,\ n-1}$$

Example

- An airline wants to evaluate the depth perception of its pilots over the age of 50. A random sample of $n = 14$ airline pilots over the age of 50 are asked to judge the distance between two markers placed 20 feet apart at the opposite end of the laboratory. The sample data listed here are the pilots' error (recorded in feet) in judging the distance.

$$2.7,\ 2.4,\ 1.9,\ 2.6,\ 2.4\ 1.9,\ 2.3$$
$$2.2,\ 2.5,\ 2.3,\ 1.8,\ 2.5,\ 2.0,\ 2.2$$

  Use the sample data to test the hypothesis that the average error in depth perception for the company's pilots over the age of 50 is 2.00 at $\alpha = 0.05$ confidence level on μ.
- **Discussion:** The sample $n = 14$ is small and assuming that the data sets are normally distributed. **Verify that** $\bar{X} = 2.26$ and $S = 0.28$.

1. **Hypothesis:**

$$H_0 : \hat{\mu} = 2.00 \text{ versus } H_a : \hat{\mu} \neq 2.00$$

1. **Test Statistic:**

$$T_c = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{2.26 - 2.00}{0.28/\sqrt{14}} = \frac{0.26}{0.28/3.742} = 3.474$$

3. **Critical region**From the t-distribution table, $t_{\alpha/2,n-1} = t_{0.025,13} = 2.16$.
4. **Conclusion:** Since $T_c = 3.474 > t_{0.025,13} = 2.16$, H_0 is rejected. the average error in depth perception for the company's pilots over the age of 50 is different from 2.00

- **Exercise:**

a. Compute the upper and lower one-sided tests at the same significance level.
b. Compute a 95% confidence interval on μ, the average error in depth perception for the company's pilots over the age of 50.

Hypothesis Testing on the Mean of a Population with Unknown Variance σ^2

Null hypothesis: $H_0\colon \mu = \mu_0$

Test statistic: $$T_0 = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

Alternative hypothesis	Rejection criteria
$H_1\colon \mu \neq \mu_0$	$t_0 > t_{\alpha/2,n-1}$ or $t_0 < -t_{\alpha/2,n-1}$
$H_1\colon \mu > \mu_0$	$t_0 > t_{\alpha,n-1}$
$H_1\colon \mu < \mu_0$	$t_0 < -t_{\alpha,n-1}$

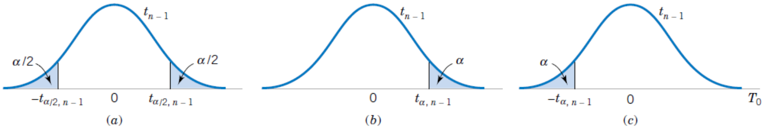


Figure: Critical Regions for two-sided (a), One-sided Alternative $H_a : \hat{\theta} > \theta_0$ (left), and the One-sided Alternative $H_a : \hat{\theta} < \theta_0$ (right), for Student t Distributed T.

Tests for a Population Variance, σ^2

- Let $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with mean μ and unknown variance σ^2. We have already mentioned that $\frac{(n-1)S^2}{\sigma^2} = \sim \chi^2_{(n-1)}$ and repeat the process to obtain the confidence interval as

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

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

1. **The test of hypotheses:**

$$H_0 : \hat{\sigma}^2 = \sigma_0^2 \text{ versus } H_a : \hat{\sigma}^2 \neq \sigma_0^2$$

where σ_0^2 is a specified constant population variance.

2. **The test statistic:**

$$\chi^2_{cal} = \frac{(n-1)S^2}{\sigma_0^2} \tag{32}$$

- The **critical region** to control the type I error probability at the desired level, the χ^2 percentage points $\chi^2_{n-1,\alpha/2}$ and as the boundaries of the critical region $\chi^2_{n-1,\alpha/2}$ and $\chi^2_{n-1,1-\alpha/2}$ so that we would reject $H_0 : \hat{\sigma}^2 = \sigma_0^2$ if

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

Tests for a Population Variance cont...

H_0: **1.** $\sigma^2 \le \sigma_0^2$ H_a: **1.** $\sigma^2 > \sigma_0^2$

2. $\sigma^2 \ge \sigma_0^2$ **2.** $\sigma^2 < \sigma_0^2$

3. $\sigma^2 = \sigma_0^2$ **3.** $\sigma^2 \neq \sigma_0^2$

T.S.: $\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$

R.R.: For a specified value of α,

1. Reject H_0 if χ^2 is greater than χ_U^2, the upper-tail value for α and df $= n - 1$.
2. Reject H_0 if χ^2 is less than χ_L^2, the lower-tail value for $1 - \alpha$ and df $= n - 1$.
3. Reject H_0 if χ^2 is greater than χ_U^2, based on $\alpha/2$ and df $= n - 1$, or less than χ_L^2, based on $1 - \alpha/2$ and df $= n - 1$.

Check assumptions and draw conclusions.

.

Tests on a Population Proportion

- **Recall** that a random sample of size n has been taken from a large (possibly infinite) population and that $X(\leq n)$ observations in this sample belong to a class of interest.
- Then is a point estimator of the proportion of the population p that belongs to this class. Note that **n** and **p** are the **parameters of a binomial distribution** with mean **p** and variance $np(1-p)$, if p is not too close to either 0 or 1 and if n is relatively large.

1. **The test of hypotheses:**
$$H_0 : \hat{p} = p_0 \text{ versus } H_a : \hat{p} \neq p_0$$
where p the binomial parameter and assuming that $X \sim N\left(np_0,\ np_0\left(1-p_0\right)\right)$.
2. **The test statistic:**
$$Z_{cal} = \frac{X - np_0}{\sqrt{np_0(1-p_0)}} \tag{33}$$
3. **Rejection Region**: Reject H_0 if the observed value of the test statistic Z_{cal} is either $Z_{cal} > Z_{\alpha/2}$ or $Z_{cal} < -Z_{\alpha/2}$.

Example

- **Example:** A semiconductor manufacturer produces controllers used in automobile engine applications. The customer requires that the process fallout or fraction defective at a critical manufacturing step not exceed 0.05 and that the manufacturer demonstrate process capability at this level of quality using $\alpha = 0.05$. The semiconductor manufacturer takes a random sample of 200 devices and finds that four of them are defective. Can the manufacturer demonstrate process capability for the customer?
- **Discussion:** $X = 4$, $\alpha = 0.05$, $n = 200$ and $p_0 = 0.05$.

1. **The test of hypotheses:**
$$H_0 : \hat{p} = 0.05 \text{ versus } H_a : \hat{p} < 0.05$$
2. **Rejection Region**: Reject H_0 if the observed value of the test statistic Z_{cal} is or $Z_{cal} < -Z_\alpha = -Z_{0.05} = -1.645$.
3. **The test statistic:**
$$\begin{aligned} Z_{cal} &= \frac{X - np_0}{\sqrt{np_0(1-p_0)}} \\ &= \frac{4 - 200(0.05)}{\sqrt{200(0.05)(1-0.05)}} = -1.95 \end{aligned}$$
4. **Conclusion**: Reject H_0 since $Z_{cal} = -1.95 < -Z_\alpha = -Z_{0.05} = -1.645$. We conclude that the process is capable.

Hypotheses Tests for a Difference in Means Distributions, Variances Unknown

- Let $X_{11}, X_{12}, \ldots, X_{1n_1}$ be a random sample of n_1 observations from the first population and $X_{21}, X_{22}, \ldots, X_{2n_2}$ be a random sample of n_2 observations from the second population.
- Let $\bar{X}_1, \bar{X}_2, S_1^2$ and S_2^2 be the sample means and sample variances, respectively.
- The expected value of the difference in sample means is $E\left[\bar{X}_1 - \bar{X}_2\right] = \mu_1 - \mu_2$, so is an unbiased estimator of the difference in means (**Verify!**).
- Tests of hypotheses on the difference in means $\mu_1 - \mu_2$ of two normal distributions where the variances and are unknown. **A t-statistic** will be used to test these hypotheses.
- Two different situations must be treated. In the first case, we assume that the **variances of the two normal distributions are unknown but equal**; that is, $\sigma_1^2 = \sigma_2^2 = \sigma^2$. In the second, **we assume that σ_1^2 and are unknown and σ_2^2 not necessarily equal**.
- **Case 1**: $\sigma_1^2 = \sigma_2^2 = \sigma^2$

1. The variance of $\bar{X}_1 - \bar{X}_2$ is

$$\begin{aligned} Var\left[\bar{X}_1 - \bar{X}_2\right] =& \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \\ =& \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \end{aligned}$$

2. **The pooled estimator** of σ^2, denoted by S_p^2, is defined by

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

$$T_c = \frac{\left(\bar{X}_1 - \bar{X}_2\right) - (\mu_1 - \mu_2)}{S_p\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

has a t- distribution with $n_1 + n_2 - 2$ degrees of freedom.

Hypotheses Tests for a Difference in Means for Equal Variance

1. **Test hypothesis**: $H_0 : \mu_1 - \mu_2 = D_0$
2. **Test Statistics**

$$T_c = \frac{\bar{X}_1 - \bar{X}_2 - D_0}{S_p \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \tag{34}$$

Alternative Hypothesis	Rejection Criterion
$H_1\colon \mu_1 - \mu_2 \neq \Delta_0$	$t_0 > t_{\alpha/2, n_1+n_2-2}$ or
	$t_0 < -t_{\alpha/2, n_1+n_2-2}$
$H_1\colon \mu_1 - \mu_2 > \Delta_0$	$t_0 > t_{\alpha, n_1+n_2-2}$
$H_1\colon \mu_1 - \mu_2 < \Delta_0$	$t_0 < -t_{\alpha, n_1+n_2-2}$

Example

- The summary statistics given below from two catalysts types in which 8 samples in the pilot plant are take from each are being analyzed to determine how they affect the mean yield of a chemical process. Specifically, the 1^{st} catalyst is currently in use, but the 2^{nd} catalyst is acceptable.

Observation Number	Catalyst 1	Catalyst 2
1	91.50	89.19
2	94.18	90.95
3	92.18	90.46
4	95.39	93.21
5	91.79	97.19
6	89.07	97.04
7	94.72	91.07
8	89.21	92.75
	$\bar{x}_1 = 92.255$	$\bar{x}_2 = 92.733$
	$s_1 = 2.39$	$s_2 = 2.98$

Table: Catalyst Yield Data

Example

- A test is run in the pilot plant and results in the data shown in Table above. Is there any difference between the mean yields? Use $\alpha = 0.05$, and assume equal variances.

1. **Test hypothesis**: $H_0 : \mu_1 - \mu_2 = 0$ versus $H_a : \mu_1 - \mu_2 \neq 0$

$$S_p^2 = \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}$$
$$= \frac{7\left(2.39^2\right) + 7\left(2.98^2\right)}{8 + 8 - 2}$$
$$= \frac{7\left(2.39^2\right) + 7\left(2.98^2\right)}{8 + 8 - 2} = 7.30 \Rightarrow S_p = \sqrt{7.30} = 2.70$$

3. **Rejection region:** Reject H_0 if $T_c > t_{\alpha/2,14} = t_{0.025}, 14 = 2.145$ or $T_c < -t_{\alpha/2,14} = -t_{0.025,14} = -2.145$
2. **Test Statistics**

$$Z_c = \frac{\mu_1 - \mu_2 - 0}{S_p\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$
$$= \frac{92.255 - 92.733 - 0}{2.70\sqrt{\left(\frac{1}{8} + \frac{1}{8}\right)}} = -0.35$$

4. **Conclusion:** Since $-t_{\alpha/2,14} = 2.145 < T_c = -0.35 < t_{0.025,14} = 2.145$ H_0 does not be rejected. That is, at the $\alpha = 0.05$ level of significance, we do not have strong evidence to conclude that catalyst 2 results in a mean yield that differs from the mean yield when catalyst 1 is used.

Tests on Two Population Proportions

- **Recall** that a random sample of size n_1 and n_2 has been taken from a large (possibly infinite) populations and that $X_1(\leq n)$ and $X_2(\leq n)$ observations in this sample belong to a class of interest.
- There is a point estimator of the proportion of the population p that belongs to this class. Note that **n** and **p** are the **parameters of a binomial distribution** with mean **p** and variance $np(1-p)$, if p is not too close to either 0 or 1 and if n is relatively large.

1. **The test of hypotheses:**

$$H_0 : \hat{p}_1 = \hat{p}_2$$

Test statistic:
$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Alternative Hypotheses	Rejection Criterion
$H_1: p_1 \neq p_2$	$z_0 > z_{\alpha/2}$ or $z_0 < -z_{\alpha/2}$
$H_1: p_1 > p_2$	$z_0 > z_\alpha$
$H_1: p_1 < p_2$	$z_0 < -z_\alpha$