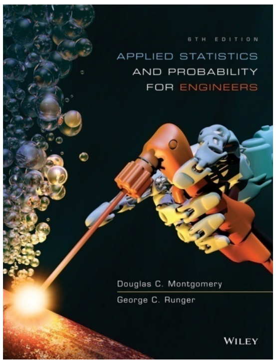
WILEY



Applied Statistics and Probability for Engineers

Sixth Edition

Douglas C. Montgomery George C. Runger

Chapter 7

Point Estimation of Parameters and Sampling Distributions

7 Point Estimation of Parameters and Sampling Distributions

CHAPTER OUTLINE

7-1 Point Estimation
7-2 Sampling Distributions and the Central Limit Theorem
7-3 General Concepts of Point Estimation
7-3.1 Unbiased Estimators
7-3.2 Variance of a Point Estimator
7-3.3 Standard Error: Reporting a Point Estimate
7-3.4 Mean Squared Error of an Estimator
7-4 Methods of Point Estimation
7-4.1 Method of Moments
7-4.2 Method of Maximum Likelihood
7-4.3 Bayesian Estimation of Parameters

Learning Objectives for Chapter 7

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

1. General concepts of estimating the parameters of a population or a probability distribution.
2. Important role of the normal distribution as a sampling distribution.
3. The central limit theorem.
4. Important properties of point estimators, including bias, variances, and mean square error.
5. Constructing point estimators using the method of moments, and the method of maximum likelihood.
6. Compute and explain the precision with which a parameter is estimated.
7. Constructing a point estimator using the Bayesian approach.

Point Estimation

- A point estimate is a reasonable value of a population parameter.
- $X_1, X_2, \ldots, X_n$ are random variables.
- Functions of these random variables, x-bar and s^2, are also random variables called statistics.
- Statistics have their unique distributions which are called sampling distributions.

Point Estimator

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

As an example, suppose the random variable X is normally distributed with an unknown mean μ. The sample mean is a point estimator of the unknown population mean μ. That is, $\hat{\mu} = \bar{X}$. After the sample has been selected, the numerical value $\bar{x}$ is the point estimate of μ.
Thus if $x_1 = 25, x_2 = 30, x_3 = 29,$ and $x_4 = 31,$ the point estimate of μ is

$$\bar{x} = \frac{25+30+29+31}{4} = 28.75$$

Some Parameters & Their Statistics

Parameter	Measure	Statistic
μ	Mean of a single population	x-bar
σ^2	Variance of a single population	s^2
σ	Standard deviation of a single population	s
p	Proportion of a single population	p-hat
$\mu_1 - \mu_2$	Difference in means of two populations	x bar$_1$ - x bar$_2$
$p_1 - p_2$	Difference in proportions of two populations	p hat$_1$ - p hat$_2$

- There could be choices for the point estimator of a parameter.
- To estimate the mean of a population, we could choose the:
  - Sample mean.
  - Sample median.
  - Average of the largest & smallest observations in the sample.

Some Definitions

- The random variables X_1, X_2,…,X_n are a random sample of size n if:
  a) The X_i 's are independent random variables.
  b) Every X_i has the same probability distribution.
- A statistic is any function of the observations in a random sample.
- The probability distribution of a statistic is called a sampling distribution.

Central Limit Theorem

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

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

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

Example 7-2: Central Limit Theorem

Suppose that a random variable X has a continuous uniform distribution:

$$f(x) = \begin{cases} 1/2, & 4 \le x \le 6 \\ 0, & \text{otherwise} \end{cases}$$

Find the distribution of the sample mean of a random sample of size $n = 40$.

By the CLT the distribution $\bar{X}$ is normal .

$$\mu = \frac{b+a}{2} = \frac{6+4}{2} = 5$$

$$\sigma^2 = \frac{(b-a)^2}{12} = \frac{(6-4)^2}{12} 1/3$$

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} = \frac{1/3}{40} = \frac{1}{120}$$

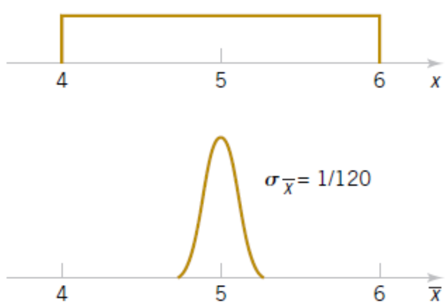


Figure 7-5 The distribution of X and $\bar{X}$ for Example 7-2.

Sampling Distribution of a Difference in Sample Means

- If we have two independent populations with means μ_1 and μ_2, and variances σ_1^2 and σ_2^2, and
- If X-bar$_1$ and X-bar$_2$ are the sample means of two independent random samples of sizes n_1 and n_2 from these populations:
- Then the sampling distribution of:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

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

Example 7-3: Aircraft Engine Life

The effective life of a component used in jet-turbine aircraft engine is a random variable with mean 5000 and SD 40 hours and is close to a normal distribution. The engine manufacturer introduces an improvement into the Manufacturing process for this component that changes the parameters to 5050 and 30. Random samples of size 16 and 25 are selected.

What is the probability that the difference in the two sample means is at least 25 hours?

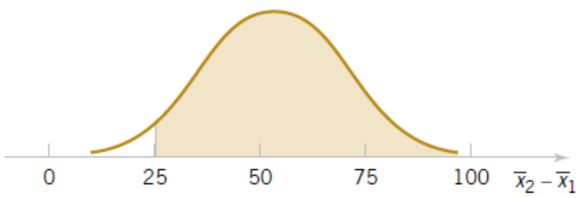


Figure 7-6 The sampling distribution of $\overline{X}_2 - \overline{X}_1$ in Example 7-3.

	Process		
	Old (1)	New (2)	Diff (2-1)
x-bar =	5,000	5,050	50
s =	40	30	
n =	16	25	
Calculations			
$s / \sqrt{n}$ =	10	6	11.7
		z =	-2.14
P(xbar$_2$-xbar$_1$ > 25) = P(Z > z) =			0.9838
			= 1 - NORMSDIST(z)

Unbiased Estimators Defined

The point estimator $\hat{\Theta}$ is an unbiased estimator for the parameter θ if:

$$E\left(\hat{\Theta}\right) = \theta$$

If the estimator is not unbiased, then the difference:

$$E\left(\hat{\Theta}\right) - \theta$$

is called the bias of the estimator $\hat{\Theta}$.

The mean of the sampling distribution of $\hat{\Theta}$ is equal to θ.

Example 7-4: Sample Mean & Variance Are Unbiased-1

- X is a random variable with mean μ and variance σ^2. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n.
- Show that the sample mean (X-bar) is an unbiased estimator of μ.

$$E(\bar{X}) = E\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right)$$

$$= \frac{1}{n}\left[E(X_1) + E(X_2) + \ldots + E(X_n)\right]$$

$$= \frac{1}{n}\left[\mu + \mu + \ldots + \mu\right] = \frac{n\mu}{n} = \mu$$

Example 7-4: Sample Mean & Variance Are Unbiased-2

Show that the sample variance (S^2) is a unbiased estimator of σ^2.

$$E(S^2) = E\left(\frac{\sum_{i=1}^{n}(X - \bar{X})^2}{n-1}\right) = \frac{1}{n-1}E\left[\sum_{i=1}^{n}\left(X_i^2 + \bar{X}^2 - 2\bar{X}X_i\right)\right]$$

$$= \frac{1}{n-1}\left[E\left(\sum_{i=1}^{n}X_i^2 - n\bar{X}^2\right)\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n}E\left(X_i^2\right) - nE\left(\bar{X}^2\right)\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}\left(\mu^2 + \sigma^2\right) - n\left(\mu^2 + \sigma^2/n\right)\right]$$

$$= \frac{1}{n-1}\left[n\mu^2 + n\sigma^2 - n\mu^2 - \sigma^2\right] = \frac{1}{n-1}\left[(n-1)\sigma^2\right] = \sigma^2$$

Minimum Variance Unbiased Estimators

- If we consider all unbiased estimators of θ, the one with the smallest variance is called the minimum variance unbiased estimator (MVUE).
- If X_1, X_2,…, X_n is a random sample of size n from a normal distribution with mean μ and variance σ^2, then the sample X-bar is the MVUE for μ.

Standard Error of an Estimator

The standard error of an estimator $\hat{\Theta}$ is its standard deviation, given by

$$\sigma_{\hat{\Theta}} = \sqrt{V\left(\hat{\Theta}\right)}.$$

If the standard error involves unknown parameters that can be estimated,

substitution of these values into $\sigma_{\hat{\Theta}}$

produces an estimated standard error, denoted by $\hat{\sigma}_{\hat{\Theta}}$.

Equivalent notation: $\hat{\sigma}_{\hat{\Theta}} = s_{\hat{\Theta}} = se\left(\hat{\Theta}\right)$

If the X_i are $\sim N\left(\mu, \sigma^2\right)$, then standard error of $\bar{X}$ is

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}.$$

If σ is not known, then $\hat{\sigma}_{\bar{X}} = \dfrac{s}{\sqrt{n}}$.

Example 7-5: Thermal Conductivity

- These observations are 10 measurements of thermal conductivity of Armco iron.
- Since σ is not known, we use s to calculate the standard error.
- Since the standard error is 0.2% of the mean, the mean estimate is fairly precise. We can be very confident that the true population mean is 41.924 ± 2(0.0898) or between 41.744 and 42.104.

x_i	
41.60	
41.48	
42.34	
41.95	
41.86	
42.18	
41.72	
42.26	
41.81	
42.04	
41.924	= Mean
0.284	= Std dev (*s*)
0.0898	= Std error

Mean Squared Error

The mean squared error of an estimator $\hat{\Theta}$ of the parameter θ is defined as:

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

$$\text{Can be rewritten as} = E\left[\hat{\Theta} - E\left(\hat{\Theta}\right)\right]^2 + \left[\theta - E\left(\hat{\Theta}\right)\right]^2$$

$$= V\left(\hat{\Theta}\right) + \left(\text{bias}\right)^2$$

Conclusion: The mean squared error (MSE) of the estimator is equal to the variance of the estimator plus the bias squared.