An **estimator** of a population parameter is a rule that tells us how to use the sample values $x_1, x_2, \ldots, x_n$ to estimate the parameter, and is a statistic.

An **estimate** is the value obtained after the observations $x_1, x_2, \ldots, x_n$ have been substituted into the formula.

Parameter θ	Estimator $\hat{\theta}$	Estimate
Pop. mean μ	Sample mean $\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
Pop. variance σ^2	Sample variance $\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$
Pop. proportion π	Sample proportion $\hat{\pi} = p = \frac{X}{n}$	$\frac{x}{n}$

An estimator $\hat{\theta}$ of θ is called an **unbiased** estimator, if $E(\hat{\theta}) = \theta$.

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators of θ. Then $\hat{\theta}_1$ is said to be **more efficient** than $\hat{\theta}_2$, if $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$.

An estimator $\hat{\theta}$ of θ is a **consistent** estimator, if it becomes almost certain that the value of $\hat{\theta}$ gets closer to the value of θ as the sample size increases.

The estimators in the above table are "good" estimators in the sense that they are unbiased, efficient, and consistent, and they are **point estimates** in that you will get a single number as an estimate.

An **interval estimate** describes a range of values within which a population parameter is likely to lie.

The **upper $\alpha \cdot 100^{\text{th}}$ percentile** of Z, denoted by z_α, is such that $P(Z \geq z_\alpha) = \alpha$. Note that z_α is the $(1-\alpha)\cdot 100^{\text{th}}$ percentile of Z because $P(Z \leq z_\alpha) = 1 - \alpha$.

We know that if X is a normal random variable with mean μ_x and variance σ_x^2, then the sample mean $\bar{X}$ is a normal random variable with mean μ_x and variance $\frac{\sigma_x^2}{n}$. Therefore,

$$P\left(\left|\frac{\bar{X} - \mu_x}{\sigma_x/\sqrt{n}}\right| \leq z_{\alpha/2}\right) = 1 - \alpha$$

From this we have, after going through some algebra,

$$P\left(\bar{X} - z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}} \leq \mu_x \leq \bar{X} + z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}\right) = 1 - \alpha,$$

that is, the probability that μ_x is between $\bar{X} - z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}$ and $\bar{X} + z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}$ is $1 - \alpha$. Note that as $\bar{X} - z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}$ and $\bar{X} + z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}$ are random variables, the interval $\left(\bar{X} - z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}\right)$ is a random interval.

Once we obtain the value of the sample mean $\bar{x}$, we have a fixed interval

$$\left(\bar{x} - z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}, \qquad \bar{x} + z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}\right),$$

and this interval is called a **$(1-\alpha)\cdot$100% confidence interval** for μ_x.
The $(1-\alpha)\cdot$100% is called a **confidence level**. We interpret the confidence interval as follows:
With $(1-\alpha)\cdot$100% confidence the parameter μ_x is between $\bar{x} - z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}$ and $\bar{x} + z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}$.
It is **wrong** to say that the probability that μ_x is between $\bar{x} - z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}$ and $\bar{x} + z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}$ is $1-\alpha$.
This is because $\bar{x} - z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}$ and $\bar{x} + z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}$ are not random variables.
Throughout we will use a confidence interval as an interval estimate.

The $(1-\alpha)$ 100% confidence interval for the population mean μ_x

	σ_x Known	σ_x Unknown
Normal Population	$\bar{x} \pm z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}$	$\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$
Non-normal Population (large *n*)	$\bar{x} \pm z_{\alpha/2}\frac{\sigma_x}{\sqrt{n}}$	$\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$ [1]

Example 1. A random sample of 100 accounts of persons having Royal Credit Cards was taken and the mean balance due was found to be \$72. If the standard deviation of all balances is \$46, find a 90 % C.I. for the mean balance due for all Royal Credit Card holders. Also interpret the C. I.

Example 2. A marketing research organization was hired to estimate the mean prime-lending rate for banks located in the western region of the United States. A random sample of 50 banks was selected and the mean and the standard deviation are 4.1% and 0.24% respectively. Find a 95% C.I. for the population mean prime rate. Interpret the C.I.

MSL Homework: 8.2, 8.3, 8.4
HC Homework: 8.9 (Use Excel/PHStat)

---

[1] You may use $\bar{x} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$ in this case as the textbook does.

We can use MS Excel to obtain "*Z*-intervals" as follows.

Formulas >f_x (Insert Function)>Statistical>CONFIDENCE.NORM

and enter the value of α for **Alpha**, the value of the standard deviation for **Standard_dev**, and the sample size for **Size**.
Then MS Excel will return the value that needs to be added to and subtracted from the sample mean. More specifically, for Example 2, enter 0.05 for Alpha (because we need a 95% C.I.), 0.24 for Standard_dev, and 50 for Size. Then MS Excel will return a value of 0.066523, and the 95% C.I. is 4.1±0.066523. As a reminder when raw data are given instead of summary statistics, we can use MS Excel to compute the sample mean and standard deviation using Formulas >f_x (Insert Function)>Statistical>, then AVERAGE and STDEV.S.

In the table on page 2, $t_{\alpha,\nu}$ denotes the upper α-100th percentile of a (Student's) *t*- random variable with ν degrees of freedom, that is $P(T_\nu > t_{\alpha,\nu}) = \alpha$ for a *t* random variable *T* with ν d.f. Table E.3 on pages 800 and 801 of the textbook contains $t_{\alpha,\nu}$ values for selected $1-\alpha$ and ν .
To get $t_{\alpha,\nu}$ in Excel **Formulas >f_x f_x (Insert Function)>Statistical>T.INV**, then enter $1-\alpha$ for **Probability** and ν for **Deg_freedom**. For example, to find $t_{0.025,9}$, enter $0.975(=1-0.025)$ for Probability and 9 for Deg_freedom.

We note that $\frac{\bar{X}-\mu_x}{s/\sqrt{n}}$ is a *t* random variable with $n-1$ degrees of freedom.

Characteristics of a (**Student's**) ***t* random variable**

1. The probability distribution of a *t* random variable is symmetric about zero, and t random variable takes values from $-\infty$ to ∞.
2. The *t* distribution is bell shaped and has a similar appearance to the standard normal distribution but flatter.
3. The parameter of a *t* distribution is called the degrees of freedom, usually denoted by ν.
4. The mean is zero and variance is $\frac{\nu}{\nu-2}$.
5. As ν increases, the distribution gets closer to that of *Z*.

Example 3. Tests on a popular brand of paint gave the following results on the square feet of coverage per gallon:

150, 159, 162, 144, 150, 162, 155, 164, 157, 140.

Assuming that the coverages are normally distributed find a 95% C.I. for the mean square feet coverage of the brand of paint. Interpret the C.I.

You may obtain "t-intervals" using MS Excel as follows:
Tools>Data Analysis…>Descriptive Statistics Then enter Input range, and check off Summary statistics and Confidence Level for Mean.
By default MS Excel computes 95% C.I. If you need a C.I. with a confidence level other than 95%, then change 95 to the level you need.
More specifically, for Example 3, first enter the data in column A, cells a1 through a10. Next click Tools>Data Analysis…>Descriptive Statistics. Then enter a1:a10 for Range, and check off Summary statistics and Confidence Level of Mean, and you will see the output on the next page.

Column1	
Mean	154.3
Standard Error	2.560599
Median	156
Mode	150
Standard Deviation	8.097325
Sample Variance	65.56667
Kurtosis	-0.77634
Skewness	-0.58242
Range	24
Minimum	140
Maximum	164
Sum	1543
Count	10
Confidence Level(95.0%)	5.792482

Here Mean is the sample mean, $\bar{x}$, Standard Error is the (estimated) standard error, $\frac{s}{\sqrt{n}}$, Standard deviation is the sample standard deviation *s*, and the number in the line of Confidence Level (95%) is the value added to and subtracted from the sample mean for the C.I. Thus 95% C.I. is 154.3 ± 5.792482 .

If the sample standard deviation is given, you may use
Formulas >f_x (Insert Function)>Statistical>CONFIDENCE.T
and enter the value of α for **Alpha**, the value of the standard deviation for **Standard_dev**, and the sample size for **Size**.

MSL Homework: 8.11, 8,12, 8.15, 8.16
HC Homework:, 8.18 (Use Excel/PHStat)

The $(1-\alpha)$ 100% confidence interval for the population proportion π

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

provided $n \geq 25$, $np \geq 5$, and $n(1-p) \geq 5$.

Example 4. A tobacco company conducts a market study by sampling and interviewing 1000 smokers to determine their brand preference. If 313 of the 1000 smokers preferred the company's brand, find the 90% confidence interval for the proportion of all smokers who prefer the company's brand.

MSL Homework: 8.27, 8.29, 8.30
HC Homework: 8.33 (Use Excel/PHStat)

Sample sizes:

For estimation of the population mean

$$n \geq \left(\frac{z_{\alpha/2} \cdot \sigma_x}{e}\right)^2$$

Here, *e* is a predetermined bound on error of estimation, also called an acceptable sampling error or a tolerance. If σ_x is unknown, use the sample standard deviation *s* from a pilot sample or use *range*/6 as an approximation of. σ_x.

Example 5. Past experience has indicated that the salaries of factory workers in a certain industry are approximately normally distributed with a standard deviation of $500. How large a sample of factory workers would be required if we wish to estimate the population mean salary to within $60 with a confidence of 99%?

For estimation of the population proportion π

$$n \geq \frac{\tilde{p}(1-\tilde{p})z_{\alpha/2}^2}{e^2}$$

$\tilde{p}$ is an "estimate" of π based on a pilot sample or *a priori* knowledge.
If $\tilde{p}$ is unknown, use $\tilde{p} = \frac{1}{2}$

Example 6. The S corporation is interested in the proportion of defective units, π, produced of its Super Betamax VCR model. The quality control manager wishes to estimate π within 0.05 unit with a confidence of 99 %.

a. Find the sample size required to meet the criteria.
b. Suppose it is decided to take a small pilot sample of VCR's to provide a preliminary estimate of π. In the pilot sample of 25 VCR's one defective unit is found. Using this information, find the sample size now required.

MSL Homework: 8.34, 8.36, 8.37, 8.38
HC Homework: 8.44(Use Excel/PHStat) , 8.45 (Use Excel/PHStat)

Using PHStat we can also obtain sample sizes.

For **Example 8.5**, click

PHStatSample Size > Determination for the Mean…

Then: enter 500 for Population Standard Deviation and 60 for Sampling Error, and change the Confidence Level to 99. Then you will get the following:

Sample Size Determination

Data	
Population Standard Deviation	**500**
Sampling Error	**60**
Confidence Level	**99%**

Intemediate Calculations	
Z Value	-2.5758293
Calculated Sample Size	460.7567084

Result	
Sample Size Needed	**461**

For **Example 8.6-b**, click

PHStat>Sample Size > Determination for the Proportion…

Then: enter 0.04 (or 0.5 for a) for Estimate of True Proportion and 0.05 for Sampling Error, and change the Confidence Level to 99. Then you will get the following:

Sample Size Determination

Data	
Estimate of True Proportion	0.04
Sampling Error	0.05
Confidence Level	99%

Intermediate Calculations	
Z Value	-2.5758293
Calculated Sample Size	101.9120118

Result	
Sample Size Needed	102

Using PHStat we can also obtain confidence intervals learned in Chapter 8.

For **Example 8.1**, click

PHStat>Confidence Intervals > Estimate for the Mean, sigma known…

Then: enter 46 for the Population Standard Deviation; change the Confidence Level to 90; enter 100 for the Sample Size and 72 for the Sample Mean. Then you will get the following

Confidence Interval Estimate for the Mean

Data	
Population Standard Deviation	46
Sample Mean	72
Sample Size	100
Confidence Level	90%

Intermediate Calculations	
Standard Error of the Mean	4.6
Z Value	-1.644853
Interval Half Width	7.566323802

Confidence Interval	
Interval Lower Limit	64.4336762
Interval Upper Limit	79.5663238

For **Example 8.2**, click

PHStat>Confidence Intervals > Estimate for the Mean, sigma known…

(Although in this case we use the sample standard deviation, *s*, in place of the population standard deviation, σ, we "pretend" that we know the standard deviation. The is because we need to compute the *z*-interval.)

Then: enter 0.24 for the Population Standard Deviation; change the Confidence Level to 95; enter 50 for the Sample Size and 4.1 for the Sample Mean. Then you will get the following

Confidence Interval Estimate for the Mean

Data	
Population Standard Deviation	**0.24**
Sample Mean	**4.1**
Sample Size	**50**
Confidence Level	**95%**

Intermediate Calculations	
Standard Error of the Mean	0.033941125
Z Value	-1.95996398
Interval Half Width	0.066523384

Confidence Interval	
Interval Lower Limit	**4.033476616**
Interval Upper Limit	**4.166523384**

For **Example 8.3**,

First enter the data,

PHStat>Confidence Intervals > Estimate for the Mean, sigma unknown…

Then: click Sample Statistics Unknown and enter the Sample Cell Range. Then you will get the following

Confidence Interval Estimate for the Mean

Data	
Sample Standard Deviation	**8.097324661**
Sample Mean	**154.3**
Sample Size	**10**
Confidence Level	**95%**

Intermediate Calculations	
Standard Error of the Mean	2.560598888
Degrees of Freedom	9
t Value	2.262158887
Interval Half Width	5.792481531

Confidence Interval	
Interval Lower Limit	**148.51**
Interval Upper Limit	**160.09**

For **Example 8.4**

PHStat>Confidence Intervals > Estimate for the Proportion…

Then: enter 1000 for the Sample Size and 313 for the Number of Successes and change the Confidence level to 90. Then you will get the following

Confidence Interval Estimate for the Mean

Data	
Sample Size	**1000**
Number of Successes	**313**
Confidence Level	**90%**

Intermediate Calculations	
Sample Proportion	0.313
Z Value	-1.644853
Standard Error of the Proportion	0.014663935
Interval Half Width	0.024120018

Confidence Interval	
Interval Lower Limit	**0.288879982**
Interval Upper Limit	**0.337120018**