

16

SAMPLING

SYLLABUS

- Concepts of population and simple random sample
- Concept of the sample mean $\bar{X}$ as a random variable with $E(\bar{X}) = \mu$ and $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$
- Distribution of sample mean from a normal population
- Use of the Central Limit Theorem to treat sample mean as having normal distribution when the sample size is sufficiently large (e.g. $n \geq 30$)
- Use of unbiased estimates of the population mean and variance from a sample, including cases where the data are given in summarised form

$$\sum x \text{ and } \sum x^2, \text{ or } \sum (x-a) \text{ and } \sum (x-a)^2$$

CONTENTS

1	Concepts of Population and Sample.....	3
2	Estimation.....	5
2.1	Unbiased Estimate of Population Mean.....	6
2.2	Sample Variance	6
2.3	Unbiased Estimate of Population Variance	7
2.4	Method of Coding	11
3	Sampling Distribution	12
3.1	Expectation and Variance of the Sample Mean $\bar{X}$	14
3.2	Distribution of the Sample Mean (Samples taken from a Population that is Normally Distributed).....	15
3.3	Distribution of the Sample Mean (Samples taken from a Population that is not Normally Distributed).....	17
4	Miscellaneous Examples	19
Annex A:	Proof that s^2 is an Unbiased Estimate of Population Variance σ^2	22
Annex B:	Practice Questions on Sampling	23

LECTURE 1

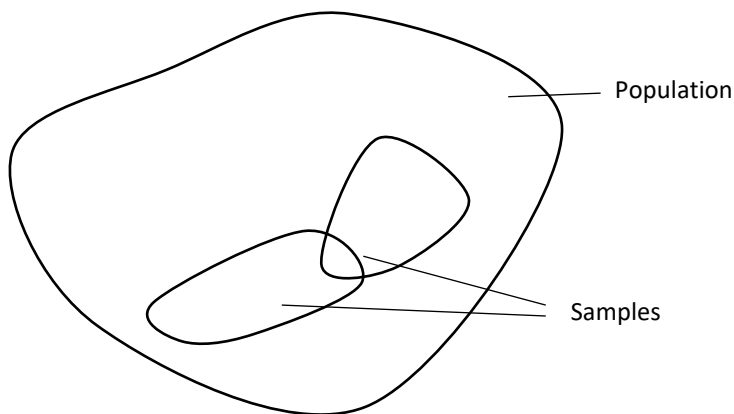
Lesson Outline

- Understand the concepts of population and sample
 - Understand the concept of point estimation of population mean and population variance
 - Find the unbiased estimates of population mean and population variance
-

1 CONCEPTS OF POPULATION AND SAMPLE

We have so far only considered random variables whose probability distribution is known. For example, random variables that follow a binomial distribution, a normal distribution etc. However, the probability distribution of the population is often not completely known. Very rarely in practice can we afford the luxury of examining the complete population. Almost invariably, sample observations are used to make inferences about the population probability distribution. This process is known as **statistical inference**. Sampling theory is the study of this relationship between a population and the samples drawn from that population.

1. A **population** is the complete set of items being investigated.
2. A **sample** is a subset of a population.



3. A sample can be chosen by selecting its members from a **sampling frame** (a suitable list of some sort, e.g. electoral register, class lists) which may or may not contain the entire population.
4. For the sample to be representative of the whole population and free of bias, **random samples** should be taken.

5. A **random sample of size n** is a sample which is taken in such a way that *every sample of size n of the population* has an *equal chance* (equal probability) of being chosen. This also means that for a random sample of size n , *every member of the population* has an equal chance of being chosen. Otherwise, the sample is a **non-random sample**.

Examples of Random Samples

- (a) 50 balls are numbered from 1 to 50 and put into a box. 3 balls are drawn from the box to determine the winning numbers in a game.
- (b) 200 students are numbered from 1 to 200. Select 20 students by the following method: Select the first student by choosing an integer k randomly, where $1 \leq k \leq 10$. Subsequently, select the remaining 19 students by selecting the numbers $k+10$, $k+20$, ..., $k+180$, $k+190$.

Examples of Non-Random Sampling

- (a) A newly opened restaurant wanted to find out their customers' dining experience in the restaurant. The manager of the restaurant suggested to complete the sampling survey in the fastest possible way. On day 1, the staff interviewed the first 20 customers to conduct the survey and collect feedback from the customers' dining experience.

6. A **sample statistic** (or statistic) is any quantity obtained from a sample for the purpose of estimating a population parameter, while a **population parameter** is a quantity that measures some aspect (e.g., mean, variance, proportion) of the whole population. Examples of sample statistics are sample mean and sample variance, while examples of population parameters are population mean and population variance.
7. Often, we are interested in finding two population parameters, namely the **population mean** and **population variance**. Population parameters are fixed and usually unknown. Thus, the study of population often involves finding estimates of these parameters. The population mean is usually denoted by μ and the population variance is denoted by σ^2 .
8. When a sample has been obtained, suitable **sample statistics** are calculated and used to estimate **population parameters**. For example, the sample mean $\bar{x}$ can be used as an estimate for the population mean μ .

2 ESTIMATION

There are many real-life situations where it is difficult, tedious and sometimes impractical to calculate population mean and population variance because of the large population size. However, we can make inferences about these population parameters by examining samples drawn from the population. This process of making inferences or generalisations about a population based on information obtained from samples selected from the population is known as **statistical inference**.

There are two major areas of statistical inferences:

- (a) estimation and
- (b) tests of hypotheses.

Suppose that a population has an unknown parameter (i.e. mean, variance, proportion). Then an **estimate** of this unknown parameter can be made from the information obtained from random sample(s) taken from the population.

Suppose we want to find out the **population mean** of the height of the male students in ACJC. One way of estimating this mean is to use the **sample mean** of a random sample made up of 5 male students from each class.

In this chapter, we will study **point estimation** where a single value is used to infer information about the population.

An **estimate** of a population parameter θ is a **single value** t of a sample statistic T that is used to estimate θ . The sample statistic T is called an **estimator** of θ .

Let X be a random variable and let $X_1, X_2, X_3, \dots, X_n$ be independent observations of X . Then, the sample mean

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i$$

is an **estimator** of the population mean μ .

Consider a population with unknown parameter θ .

Let T be some statistic derived from a random sample taken from a population to estimate the value of θ . Then

$$T \text{ is an } \mathbf{unbiased estimator} \text{ of } \theta \text{ if } E(T) = \theta.$$

2.1 Unbiased Estimate of Population Mean

Consider a random sample of size n taken from a population with unknown mean μ .

For the n independent observations, $X_1, X_2, X_3, \dots, X_n$, let

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

Proof

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{X_1 + X_2 + X_3 + \dots + X_n}{n}\right) \\ &= \frac{1}{n} E(X_1 + X_2 + X_3 + \dots + X_n) \\ &= \frac{1}{n} (E(X_1) + E(X_2) + E(X_3) + \dots + E(X_n)) \\ &= \frac{1}{n} (nE(X)) \\ &= E(X) \\ &= \mu \end{aligned}$$

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

If $x_1, x_2, x_3, \dots, x_n$ are the values that $X_1, X_2, X_3, \dots, X_n$ assume, then an

unbiased estimate of μ is $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum x}{n}$.

Note

- An unbiased estimate for population mean will be a “good” one when the sample size is large.

2.2 Sample Variance

Let X be a random variable and let $X_1, X_2, X_3, \dots, X_n$ be independent observations of X .

Consider a random sample of size n taken from a population and $x_1, x_2, x_3, \dots, x_n$ are the values that $X_1, X_2, X_3, \dots, X_n$ assume.

The **sample variance** is defined by

$$\text{sample variance} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right) - \bar{x}^2$$

The sample variance is *not* an unbiased estimate of the population variance.

2.3 Unbiased Estimate of Population Variance

Let X be a random variable and let $X_1, X_2, X_3, \dots, X_n$ be independent observations of X .

Consider a random sample of size n taken from a population with unknown variance σ^2 .

$x_1, x_2, x_3, \dots, x_n$ are the values that $X_1, X_2, X_3, \dots, X_n$ assume.

It can be shown that

$$s^2 = \frac{n}{n-1} \times \text{sample variance}$$

is an **unbiased estimate** of the population variance σ^2 .

Formula for unbiased estimate of population variance from a single sample in MF27 is as follows:

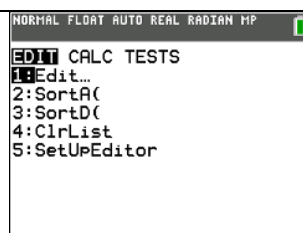
$$s^2 = \frac{n}{n-1} \left(\frac{\sum (x - \bar{x})^2}{n} \right) = \frac{1}{n-1} \left(\sum x^2 - \frac{(\sum x)^2}{n} \right)$$

Example 1

A sample of four matchboxes taken from the output of a filing machine contained 52, 56, 53 and 51 matches. Find unbiased estimates for the mean and variance of the output.

Solution

- 1) Press [STAT] followed by [ENTER] to select 1: Edit.



- 2) Position the cursor at the first row of L1 and enter the data 52, 56, 53 and 51 in the column for L1.

The image shows a calculator screen with a table for data entry. The table has five columns labeled L1, L2, L3, L4, and L5, and a sixth column for the row number. The first four rows of the L1 column contain the values 52, 56, 53, and 51. The cursor is positioned at the first row of L1.

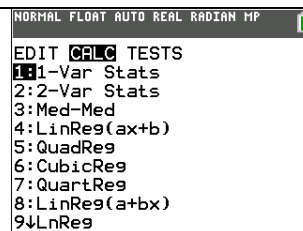
L1	L2	L3	L4	L5	
52					1
56					
53					
51					

Below the table, the text 'L1(5)= ' is displayed.

3) Press [STAT] followed by '→' for the CALC menu.

Press [ENTER] to select 1: 1-Var Stats.

Press [ENTER] to get the results.



NORMAL FLOAT AUTO REAL RADIAN MP

EDIT CALC TESTS

1:1-Var Stats

2:2-Var Stats

3:Med-Med

4:LinReg(ax+b)

5:QuadReg

6:CubicReg

7:QuartReg

8:LinReg(a+bx)

9:LnReg

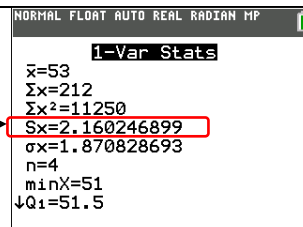
From G.C., Unbiased estimate for population mean = $\bar{x} = 53$, and

Unbiased estimate for population variance →

$$= s^2 = (2.1602\dots)^2$$

$$= 4.67 \text{ (3 sig. fig.)}$$

$$\text{Sample variance} = (1.8708\dots)^2 = 3.5$$



NORMAL FLOAT AUTO REAL RADIAN MP

1-Var Stats

$\bar{x}=53$

$\Sigma x=212$

$\Sigma x^2=11250$

$Sx=2.160246899$

$\sigma x=1.870828693$

$n=4$

$\min X=51$

$\downarrow Q1=51.5$

Note

- $s^2 = \frac{n}{n-1} \times \text{sample variance}$
- Since $\frac{n}{n-1} > 1$ for all positive integers n , $s^2 > \text{sample variance}$.
Hence $s > \text{standard deviation of the data}$.
- From the GC, it can be observed that $s_x > \sigma_x$.
 - $s_x^2 = \text{unbiased estimate of the population variance } (s^2)$
 - $\sigma_x^2 = \text{sample variance}$
 - $\sigma_x = \text{standard deviation of the sample data}$

Example 2

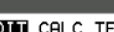
A wholesaler sells oranges in boxes of 50. A sample of 60 boxes is taken during a quality inspection. The table below shows the number of bad oranges in each box.

Number of bad oranges	0	1	2	3	4
Number of boxes	24	15	9	6	6

Find unbiased estimates for the mean and variance of the number of bad oranges in the boxes.

Solution

-
- 1) Press [STAT] followed by [ENTER] to select 1: Edit.



The image shows a TI-84 Plus calculator screen. At the top, the mode settings are displayed: NORMAL, FLOAT, AUTO, REAL, RADIAN, and MP. Below this, the 'EDIT' menu is open, showing the following options: 1:Edit..., 2:SortA(, 3:SortD(, 4:ClrList, and 5:SetUpEditor. The 'EDIT' menu is highlighted with a thick black border.

- 2) Enter the data 0, 1, 2, 3 and 4 into L1. Then position the cursor at the second column (L2) and enter the frequencies 24, 15, 9, 6 and 6.

[illegible]

- 3) Press [STAT] followed by '→' for the CALC menu.
Press [ENTER] to select 1: 1-Var Stats.

NORMAL FLOAT AUTO REAL RADIAN MP

EDIT **CLC** TESTS

1:1-Var Stats

2:2-Var Stats

3:Med-Med

4:LinReg(ax+b)

5:QuadReg

6:CubicReg

7:QuartReg

8:LinReg(a+bx)

9:LnReg

- 4) Press $[2^{\text{nd}}][1][,][2^{\text{nd}}][2]$ for L1, L2 to get the results.

NORMAL FLOAT AUTO REAL RADIAN MP

1-Var Stats

List:L1
FreqList:L2
Calculate

From the output, we get:

Unbiased estimate for population mean

$$= \bar{x} = 1.25.$$

Unbiased estimate for population variance

$$= s^2 = (1.3482\dots)^2 = 1.82 \text{ (3 s.f.)}$$

(Sample variance = $(1.3369)^2 = 1.79$ (3 s.f.))

1-Var Stats

$\bar{x}=1.25$
 $\Sigma x=75$
 $\Sigma x^2=201$
 $Sx=1.348256878$
 $\sigma x=1.336974196$
 $n=60$
 $\min X=0$
 $\downarrow Q_1=0$

Example 3

A random sample of 12 boys is taken and the heights, x cm are measured. It is found that $\sum x = 1638$ and $\sum (x - \bar{x})^2 = 1453$. Find the unbiased estimates of the population mean and variance.

Solution

From MF27, we have

$$s^2 = \frac{n}{n-1} \left(\frac{\sum (x - \bar{x})^2}{n} \right) = \frac{1}{n-1} \left(\sum x^2 - \frac{(\sum x)^2}{n} \right)$$

Unbiased estimate of population mean,

Unbiased estimate of population variance,

■

Example 4

A random sample of 250 adult men undergoing a routine medical inspection had their height (x cm) measured to the nearest cm, and the following data was obtained,

$$\sum x = 43205, \sum x^2 = 7\,469\,107.$$

Calculate unbiased estimates of the population mean and the population variance.

Solution

Unbiased estimate of population mean,

Unbiased estimate of population variance,

■

Extra Practice

Annex B: Practice Question 1

LECTURE 2

Lesson Outline

- Method of coding
- Understand that the sample mean $\bar{X}$ is a random variable with $E(\bar{X}) = \mu$ and $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$
- Distribution of $\bar{X}$ given that X has a normal distribution, i.e. if $X \sim N(\mu, \sigma^2)$, then $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

2.4 Method of Coding

Consider a random sample of size 5:

5010, 5020, 5030, 5040, 5050

Instead of calculating the mean and variance directly, we can subtract 5000 from each value and consider these values instead:

10, 20, 30, 40, 50

This is to keep the recorded values small so that it is easier to manage when calculating the unbiased estimates of the population mean and variance. This method of considering $x - 5000$ instead of x is known as **method of coding**.

For a sample of random variable X where the values are summarised in the form $\sum(x - c)$ and $\sum(x - c)^2$, where c is a constant, we let $Y = X - c$. Then we have the following results:

	$Y = X - c$	$y = x - c$
Mean	$E(Y) = E(X) - c$	$\bar{y} = \bar{x} - c \Rightarrow \bar{x} = \bar{y} + c$ $\bar{x} = \frac{\sum y}{n} + c = \frac{\sum (x - c)}{n} + c$
Variance	$\text{Var}(Y) = \text{Var}(X - c)$ $= \text{Var}(X)$	$s_y^2 = s_x^2$ $= \frac{1}{n-1} \left(\sum y^2 - \frac{(\sum y)^2}{n} \right)$ (MF27) $= \frac{1}{n-1} \left(\sum (x - c)^2 - \frac{(\sum (x - c))^2}{n} \right)$

Hence the unbiased estimates of the population mean and variance of X are given by

$$\bar{x} = \frac{\sum(x-c)}{n} + c$$

and

$$s^2 = \frac{1}{n-1} \left(\sum(x-c)^2 - \frac{(\sum(x-c))^2}{n} \right)$$

Example 5

An electronic device is advertised as being able to retain information stored in it for 70 to 90 hours after the power has been switched off. In experiments carried out to test this claim, the retention time in hours, X , was measured on 250 occasions, and the data obtained is summarised by $\sum(x-76) = 683$ and $\sum(x-76)^2 = 26132$. Find the unbiased estimates for the population mean and the population variance.

Solution

Unbiased estimate of population mean,

Unbiased estimate of population variance,

■

3 SAMPLING DISTRIBUTION

If we repeatedly take samples from a population and calculate the same statistic each time, there is a range of values that the statistic can take. The statistic will have its own distribution, which we call **sampling distribution**.

Worked Example 6

A manufacturer of light bulbs sells 60-watt bulbs and 100-watt bulbs in the ratio of 3:1. Let X be the random variable “Power of a bulb, in watts”. The distribution of population is

x	60	100
$P(X = x)$	$\frac{3}{4}$	$\frac{1}{4}$

A random sample of 3 bulbs is taken from a store containing bulbs in this ratio. The possible samples are as follows:

Samples	X_1	X_2	X_3
1	60	60	60
2	60	60	100
3	60	100	60
4	100	60	60
5	60	100	100
6	100	60	100
7	100	100	60
8	100	100	100

Computing the mean in each sample and its probability,

Samples	$\bar{X}$	Probability
1	$\frac{60+60+60}{3} = 60$	$\frac{3}{4} \times \frac{3}{4} \times \frac{3}{4} = \frac{27}{64}$
2	$\frac{60+60+100}{3} = 73\frac{1}{3}$	$\frac{3}{4} \times \frac{3}{4} \times \frac{1}{4} = \frac{9}{64}$
3	$\frac{60+100+60}{3} = 73\frac{1}{3}$	$\frac{3}{4} \times \frac{1}{4} \times \frac{3}{4} = \frac{9}{64}$
4	$\frac{100+60+60}{3} = 73\frac{1}{3}$	$\frac{1}{4} \times \frac{3}{4} \times \frac{3}{4} = \frac{9}{64}$
5	$\frac{60+100+100}{3} = 86\frac{2}{3}$	$\frac{3}{4} \times \frac{1}{4} \times \frac{1}{4} = \frac{3}{64}$
6	$\frac{100+60+100}{3} = 86\frac{2}{3}$	$\frac{1}{4} \times \frac{3}{4} \times \frac{1}{4} = \frac{3}{64}$
7	$\frac{100+100+60}{3} = 86\frac{2}{3}$	$\frac{1}{4} \times \frac{1}{4} \times \frac{3}{4} = \frac{3}{64}$
8	$\frac{100+100+100}{3} = 100$	$\frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} = \frac{1}{64}$

The distribution of the sample mean $\bar{X}$ is

$\bar{x}$	60	$73\frac{1}{3}$	$86\frac{2}{3}$	100
$P(\bar{X} = \bar{x})$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

■

3.1 Expectation and Variance of the Sample Mean $\bar{X}$

If $X_1, X_2, \dots, X_n$ are independent observations of X that form a sample (of size n) taken from a population, then the **sample mean** is the random variable

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

If the population has mean μ and variance σ^2 , then

- (a) the **expectation of the sample mean**, $\bar{X}$, is given by $E(\bar{X}) = \mu$;
- (b) the **variance of the sample mean**, $\bar{X}$, is given by $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$.

Proof

$$\begin{aligned} E(\bar{X}) &= E\left(\frac{X_1 + X_2 + X_3 + \dots + X_n}{n}\right) \\ &= \frac{1}{n} E(X_1 + X_2 + X_3 + \dots + X_n) \\ &= \frac{1}{n} (E(X_1) + E(X_2) + \dots + E(X_n)) \\ &= \frac{1}{n} (nE(X)) = \frac{1}{n} (n\mu) = \mu \end{aligned}$$

$$\begin{aligned} \text{Var}(\bar{X}) &= \text{Var}\left(\frac{X_1 + X_2 + X_3 + \dots + X_n}{n}\right) \\ &= \frac{1}{n^2} \text{Var}(X_1 + X_2 + X_3 + \dots + X_n) \\ &= \frac{1}{n^2} (\text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n)) \\ &= \frac{1}{n^2} (n\text{Var}(X)) = \frac{\sigma^2}{n} \end{aligned}$$

■

A related random variable is the sum of n independent observations of X .

If the population has mean μ and variance σ^2 , then

- (a) the **expectation of the sum of the n independent observations of X** , $X_1 + X_2 + \dots + X_n$ is given by $E(X_1 + X_2 + \dots + X_n) = n\mu$;
- (b) the **variance of the sum of the n independent observations of X** , $X_1 + X_2 + \dots + X_n$ is given by $\text{Var}(X_1 + X_2 + \dots + X_n) = n\sigma^2$.

Proof

$$\begin{aligned} E(X_1 + X_2 + X_3 + \dots + X_n) &= E(X_1) + E(X_2) + \dots + E(X_n) \\ &= nE(X) = n\mu \end{aligned}$$

$$\begin{aligned} \text{Var}(X_1 + X_2 + X_3 + \dots + X_n) &= \text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n) \\ &= n\text{Var}(X) = n\sigma^2 \end{aligned} \quad \blacksquare$$

Note

- The above properties hold regardless of the distribution of X .

3.2 Distribution of the Sample Mean (Samples Taken from a Population that is Normally Distributed)

If a random sample $X_1, X_2, \dots, X_n$ of size n is drawn from a normal distribution such that $X \sim N(\mu, \sigma^2)$, then,

- (1) the **sampling distribution of the sample mean, $\bar{X}$** , is also normal with mean μ and variance $\frac{\sigma^2}{n}$, i.e.

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

- (2) the **distribution of the sum of the n observations, $X_1 + X_2 + \dots + X_n$** , is also normal with mean $n\mu$ and variance $n\sigma^2$, i.e.,

$$X_1 + X_2 + \dots + X_n \sim N(n\mu, n\sigma^2).$$

Example 7

The height of a particular species of plant follows a normal distribution with mean 21 cm and standard deviation $\sqrt{90}$ cm. A random sample of 10 plants is taken and the mean height calculated. Find the probability that this sample mean lies between 18 cm and 27 cm.

Solution

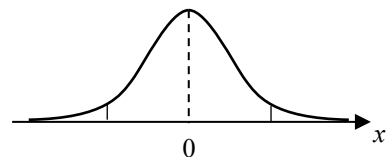
■

Example 8

If $X_1, X_2, \dots, X_n$ is a random sample from $N(\mu, 1)$, state the distribution of the sample mean, $\bar{X}$. Find the least sample size required to ensure that the probability that $\bar{X}$ is within ± 0.1 of μ is greater than 0.95.

Solution

Note that $\frac{1}{\sqrt{\frac{1}{n}}} = \sqrt{n}$



■

LECTURE 3

Lesson Outline

- Apply Central Limit Theorem in problems where X is not normally distributed and the sample size is sufficiently large, i.e.

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ **approximately** by Central Limit Theorem for large } n.$$

3.3 Distribution of the Sample Mean (Samples Taken from a Population that is Not Normally Distributed)

CENTRAL LIMIT THEOREM

If $X_1, X_2, \dots, X_n$ is a random sample of size n drawn from **any** distribution (not normally distributed) with mean $E(X) = \mu$ and variance $\text{Var}(X) = \sigma^2$, then, **for large n ($n \geq 30$)**,

The **distribution of the sample mean**, $\bar{X}$, is **approximately** normal, i.e.,

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ **approximately**}$$

where $\mu = E(X)$, $\sigma^2 = \text{Var}(X)$

and

The **distribution of the sum of the n observations**, $X_1 + X_2 + \dots + X_n$, is **approximately** normal, i.e.,

$$X_1 + X_2 + \dots + X_n \sim N(n\mu, n\sigma^2) \text{ **approximately** .}$$

where $\mu = E(X)$, $\sigma^2 = \text{Var}(X)$

Any distribution has mean and variance, therefore it is wrong to conclude $X \sim N(\mu, \sigma^2)$ when both parameters are given.

Note

- This approximation of using the Central Limit Theorem gets better as the sample size n gets larger. The larger the sample size, the closer the sampling distribution gets to the normal distribution.

Example 9

The random variable X has mean 15 and variance 25. The random variable $\bar{X}$ is the mean of a random sample of 70 observations on X . State the approximate distribution of $\bar{X}$ and hence find, approximately, $P(15 < \bar{X} < 16)$.

Solution

■

Example 10

Many samples of size n , where n is large, are taken from $B(20, 0.2)$. It is found that approximately 90% of the sample means are less than 4.314. Estimate n .

Solution:

Let X be the binomial random variable. $X \sim B(20, 0.2)$

■

Extra Practice

Annex B: Practice Question 2

4 MISCELLANEOUS EXAMPLES

Example 11

In a small town, there are two clinics, Lee's Clinic and Hope Clinic. At Lee's Clinic, the waiting time of a patient follows a normal distribution with mean 25 minutes and standard deviation 8 minutes. The waiting time for patient at Hope Clinic follows a normal distribution with mean 37 minutes and standard deviation 4 minutes. [AJC/2012/2/10]

- (i) A random sample of five patients at Lee's Clinic is selected. Find the probability that the mean waiting time of these five patients differs from the waiting time of a randomly selected patient at Hope Clinic by at least five minutes. [4]
State an assumption needed for your calculations to be valid. [1]
- (ii) On a particular day, n (where $n > 40$) patients were treated at Lee's Clinic. Find the least value of n such that the probability that at most 40 of the patients have waiting times of more than 25 minutes is less than 0.95. [3]
- (iii) At Lee's Clinic, on average, one in five patients is treated for influenza. Sixty samples of 20 patients are chosen. Find the probability that the average number of patients treated for influenza per sample is more than 3.5. [3]

Solution

Let L be random variable for waiting time of a patient at Lee's Clinic and H be random variable for waiting time of a patient at Hope Clinic.

$$L \sim N(25, 8^2), H \sim N(37, 4^2)$$

$$(i) \quad \text{Let } \bar{L} = \frac{L_1 + L_2 + L_3 + L_4 + L_5}{5} \sim$$

$$\text{So we have } \bar{L} - H \sim$$

$$\text{Required probability} =$$

Assumption: The waiting times of all patients are independent.

- (ii) Let Y be random variable of number of patients with waiting time more than 25 minutes.

$$P(L > 25) = 0.5$$

$$Y \sim B(n, 0.5)$$

By G.C.,

$$n = 67, P(Y \leq 40) = 0.9568$$

$$n = 68, P(Y \leq 40) = 0.9429$$

$$n = 69, P(Y \leq 40) = 0.9260$$

X	Y1
63	0.9887
64	0.9836
65	0.9768
66	0.968
67	0.9568
68	0.9429
69	0.926
70	0.9059
71	0.8825
72	0.8556
73	0.8254

- (iii) Let W be random variable for number of patients treated for influenza in a sample of 20. $W \sim$

Since $n = 60$, by Central Limit Theorem, $\overline{W} \sim$

■

Example 12

In a glass factory, it is found that 20% of glass panels produced by machine A are more than 3 mm thick. Given that the thickness of glass panels produced by machine A follows a normal distribution with mean 2.56 mm and standard deviation σ mm, show that $\sigma = 0.523$. [2]

- (i) For the manufacture of a certain type of windscreen, two of the glass panels produced by machine A are used to form a double panel. Find the probability that the thickness of a double panel is between 4 mm and 6 mm. [2]
- (ii) One hundred samples, each consisting of 5 glass panels produced by machine A , are tested for their thickness. In how many of these samples would you expect to find the mean thickness to be greater than 3 mm? [3]

[ACJC/2011/2/8/modified]

Solution

Let X be the thickness of the glass panel produced by machine A .

$$X \sim N(2.56, \sigma^2).$$

$$\text{Given } P(X > 3) = 0.2$$

(i) $W = X_1 + X_2 \sim N(2(2.56), 2(0.5228^2))$

(ii) Since $X \sim N(2.56, 0.5228^2)$, $\bar{X} \sim N\left(2.56, \frac{0.5228^2}{5}\right)$

Therefore, expected number =

■

ANNEX A

Proof that s^2 is an unbiased estimate of population variance σ^2

$$E(S^2) = \sigma^2 \text{ where } s^2 = \frac{n}{n-1} \left(\frac{\sum (x - \bar{x})^2}{n} \right) = \frac{1}{n-1} \left(\sum x^2 - \frac{(\sum x)^2}{n} \right) \text{ in MF27}$$

Proof

$$\begin{aligned} E(S^2) &= E \left(\frac{\sum (X - \bar{X})^2}{n-1} \right) \\ &= \frac{1}{n-1} E \left(\sum (X^2 - 2X\bar{X} + \bar{X}^2) \right) \\ &= \frac{1}{n-1} E \left(\sum X^2 - 2\bar{X} \sum X + \sum \bar{X}^2 \right) \\ &= \frac{1}{n-1} E \left(\sum X^2 - 2n\bar{X}^2 + n\bar{X}^2 \right) \text{ since } \bar{X} = \frac{\sum X}{n} \\ &= \frac{1}{n-1} E \left(\sum X^2 - n\bar{X}^2 \right) \\ &= \frac{1}{n-1} \left(E \left(\sum X^2 \right) - nE \left(\bar{X}^2 \right) \right) \\ &= \frac{1}{n-1} \left(\sum E \left(X^2 \right) - nE \left(\bar{X}^2 \right) \right) \text{ since } E \left(\sum (X) \right) = \sum E \left(X \right) \text{ as } E \left(X + Y + \dots \right) = E \left(X \right) + E \left(Y \right) + \dots \end{aligned}$$

Since

$$\begin{aligned} \text{Var}(X) &= E(X^2) - (E(X))^2 & \text{Var}(\bar{X}) &= E(\bar{X}^2) - (E(\bar{X}))^2 \\ \sigma^2 &= E(X^2) - \mu^2 & \frac{\sigma^2}{n} &= E(\bar{X}^2) - \mu^2 \\ E(X^2) &= \sigma^2 + \mu^2 & E(\bar{X}^2) &= \frac{\sigma^2}{n} + \mu^2 \end{aligned}$$

Thus,

$$\begin{aligned} E(S^2) &= \frac{1}{n-1} \left(\sum E(X^2) - nE(\bar{X}^2) \right) \\ &= \frac{1}{n-1} \left(\sum (\sigma^2 + \mu^2) - n \left(\frac{\sigma^2}{n} + \mu^2 \right) \right) \\ &= \frac{1}{n-1} \left(n(\sigma^2 + \mu^2) - (\sigma^2 + n\mu^2) \right) \\ &= \frac{1}{n-1} (n-1)\sigma^2 \\ &= \sigma^2 \end{aligned}$$

ANNEX B

Practice Questions on Sampling

- 1 Find unbiased estimates of the population mean and variance for each of the samples below:
 - (a) 53.7, 39.9, 46.0, 62.8, 44.9, 55.3, 49.6, 48.2, 53.0 and 66.9
 - (b) $n = 150$, $\sum x = 7170$, $\sum x^2 = 364182$
 - (c) $n = 250$, $\sum (x - 4) = -106$, $\sum (x - 4)^2 = 1104$

- 2 Write down the distributions of $\bar{X}$ and $X_1 + X_2 + \cdots + X_n$ for each of the following scenarios:
 - (a) $n = 16$, $X \sim N(11, 3^2)$
 - (b) $n = 1000$, $X \sim B(15, 0.6)$

Answers

- 1 (a) 52.03, 67.4 (b) 47.8, 144 (c) 3.576, 4.25

- 2 (a) $\bar{X} \sim N\left(11, \frac{9}{16}\right)$, $X_1 + X_2 + \cdots + X_n \sim N(176, 144)$

- (b) $\bar{X} \sim N\left(9, \frac{3.6}{1000}\right)$, $X_1 + X_2 + \cdots + X_n \sim N(9000, 3600)$