Course	:	Diploma in Electronic Systems Diploma in Telematics & Media Technology Diploma in Aerospace Systems & Management Diploma in Electrical Engineering with Eco-Design Diploma in Mechatronics Engineering Diploma in Digital & Precision Engineering Diploma in Aeronautical & Aerospace Technology Diploma in Biomedical Engineering Diploma in Nanotechnology & Materials Science Diploma in Engineering with Business Diploma in Information Technology Diploma in Financial Informatics Diploma in Cybersecurity & Forensics Diploma in Infocomm & Security Diploma in Chemical & Pharmaceutical Technology Diploma in Biologics & Process Technology Diploma in Chemical & Green Technology		
Module	:	Engineering Mathematics 2B / Mathematics 2B/ Computing Mathematics 2	–	EG1761/2008/2681/2916/2961 EGB/D/F/H/J/M207 IT1201/1531/1631/1761 CLB/C/G201

Topic 8 : Distribution of sample means

Objectives :

At the end of this lesson, the student should be able to:

1 identify distribution of sample means
2 apply the Central Limit Theorem to find the probability of a sample mean for sufficiently large samples

Topic 8: Distribution of Sample Means

8.1 Introduction of $\overline{X}$

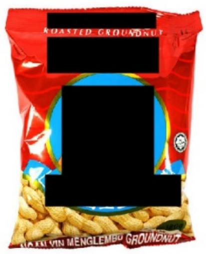
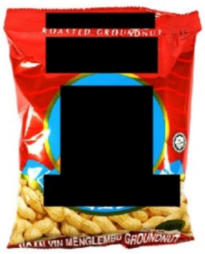
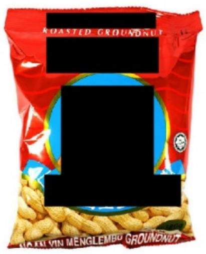
- In Chapter 1 we introduced the concept of sample mean, $\bar{x}$, for **a single sample data set**. In this case $\bar{x}$ is a single **value**. In this chapter we will look at the sample means of **multiple samples** and the **distribution of the sample means**.

- Introduction example:

  Let X denote the weight of a single peanut from a packet of peanuts. Suppose we weigh each peanut in that packet, we can calculate the sample mean of the weight of the peanuts, $\bar{x}$, in that single packet.

  1 packet:  $\bar{x}$

  Suppose we have n packets of peanuts and we will calculate the sample mean of each packet of peanuts. The sample mean for a packet of peanut will most likely be different from the sample mean of other packets.

   $\bar{x}_1$,  $\bar{x}_2$, …,  $\bar{x}_n$

  Hence in general, the sample mean, $\overline{X}$, is a **random variable** (as we cannot determine the actual value of $\bar{x}$ for a randomly chosen sample.)

8.2 Distribution of $\overline{X}$

- Since $\overline{X}$ is a random variable, we can calculate probabilities involving $\overline{X}$ once we know its distribution which is shown below:

- Let μ and σ^2 be the mean and variance of a random variable X (single quantity). If we have a sample of n objects,

(a) If $X \sim N(\mu, \sigma^2)$, then $\overline{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$.

(b) If the distribution of X is **not normal** (or unknown), and $\underline{n \geq 30}$

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

This is known as the **Central Limit Theorem**.

- From above, we see that the mean of $\overline{X} = \mu$, variance of $\overline{X} = \dfrac{\sigma^2}{n}$.

- Standard deviation of $\overline{X} = \dfrac{\sigma}{\sqrt{n}}$ is also known as the **standard error of the mean**.

Example 8.2-1

The mass of garlic bulbs produced by a particular farm is approximately normally distributed, with a mean of 60 g and a standard deviation of 5 g. State the distribution of the sample mean of a random sample of 16 garlic bulbs.

Solution:

Let X be the mass of a garlic bulb. $X \sim N\left(60, 5^2\right)$

The sample mean of 16 garlic bulbs, $\overline{X} \sim N(\qquad , \qquad)$

Example 8.2-2

The waistline of forty-year-old male Singaporeans is known to have a mean of 33 inches and a variance of 9 square inches. A random sample of 36 forty-year-old male Singaporeans was selected. Find the probability that the sample mean

(a) is greater than 31.5 inches,

(b) lies between 32 and 34 inches,

(c) differs from the population mean by more than one inch.

Solution:

Let X be the waistline of a forty year old male.

$n = 36$

Sample mean of 36 male, $\overline{X} \sim N\left(33, \frac{9}{36} \right) approx \Rightarrow \overline{X} \sim N\left(33, \frac{1}{4} \right) approx$ (by CLT)

a) $P\left(\overline{X} > 31.5\right) = 1 - P\left(\overline{X} \le 31.5\right) = 1 - P\left(Z \le \frac{31.5 - 33}{\sqrt{1/4}} \right)$

=

b) $P\left(32 \le \overline{X} \le 34\right) = P\left(\overline{X} \le 34\right) - P\left(\overline{X} < 32\right) = P\left(Z \le \frac{34 - 33}{\sqrt{1/4}} \right) - P\left(Z < \frac{32 - 33}{\sqrt{1/4}} \right)$

=

c) $P\left(\overline{X} < 32 \text{ or } \overline{X} > 34\right) = 1 - P\left(32 \le \overline{X} \le 34\right) =$

Example 8.2-3

The body length (excluding the tail) of a particular species of mice is approximately normally distributed, with a mean of 12 cm and a standard deviation of 2.4 cm.

(a) If a random sample of 16 mice is selected, what is the probability that it will have an **average** body length of between 11 and 13 cm?

(b) If a random sample of 25 mice is selected, what is the probability that it will have an **average** body length of between 11 and 13 cm?

(c) Comment on the answers obtained in part (a) and (b).

Solution:

Let X be the body length of a mouse.

$$X \sim N(12, 2.4^2)$$

a) $n = 16$

Sample mean of 16 mice, $\overline{X} \sim N\left(12, \dfrac{2.4^2}{16}\right) \Rightarrow \overline{X} \sim N(12, 0.36)$

$$P\left(11 \le \overline{X} \le 13\right) = P\left(\overline{X} \le 13\right) - P\left(\overline{X} < 11\right) = P\left(Z \le \frac{13-12}{\sqrt{0.36}}\right) - P\left(Z < \frac{11-12}{\sqrt{0.36}}\right)$$

$=$

b) $n = 25$

Sample mean of 25 mice, $\overline{X} \sim N\left(12, \dfrac{2.4^2}{25}\right) \Rightarrow \overline{X} \sim N(12, 0.2304)$

$$P\left(11 \le \overline{X} \le 13\right) = P\left(\overline{X} \le 13\right) - P\left(\overline{X} < 11\right) = P\left(Z \le \frac{13-12}{\sqrt{0.2304}}\right) - P\left(Z < \frac{11-12}{\sqrt{0.2304}}\right)$$

$=$

c) The required probability becomes ____________ when sample size ____________.

Tutorial 8: Distribution of Sampling Means

A Self Practice Questions

1 Let $X_1, X_2, ..., X_n$ be independent random variables. Write down the mean and variance of $\overline{X}$ for each of the following:

(i) $n = 15$, mean of $X = 4$, variance of $X = 7$.

(ii) $n = 30$, mean of $X = 5$, standard deviation of $X = 3$.

2 Let $X_1, X_2, ..., X_n$ be independent random variables with mean 3 and variance 5. Write down the distribution of $\overline{X}$ (with explanation if necessary) when:

(i) X_i's are normal, $n = 10$, (ii) X_i's are normal, $n = 60$,

(iii) Distribution of X_i's are unknown, $n = 35$.

3 Calculate $P\left(\overline{X} < 4\right)$ for Question 2(iii).

B Discussion Questions

1 In a certain population of swordtail fish, the lengths of the individual fish follow approximately a normal distribution with mean 52.0 mm and standard deviation of 6.0 mm. Find the probability that a random sample of 25 swordtail fishes will have an average length of

(i) less than 48.6 mm

(ii) between 52.4 and 54.4 mm.

2 According to an article, root-canal therapy costs from \$200 to \$700. Suppose the mean cost for root-canal therapy is \$450 and the standard deviation is \$125. If a sample of 100 dentists was selected across the country, find the probability that the mean cost per root canal for the sample would fall between \$425 and \$475.

3 The average number of days spent in a particular hospital for a coronary bypass in 2013 was 9 days and the standard deviation was 4 days. What is the probability that a random sample of 30 patients will have an average stay longer than 9.5 days? State any assumptions required on the distribution on the days spent.

4 The intelligence quotient (IQ) score of a certain population of children is approximately normally distributed with a mean of 102 and a standard deviation of 10. Let Y be the random variable 'the IQ score of children'.

(i) If a random sample of n children is selected, find the value of n given that $P(\bar{Y} > 103) = 0.3446$.

(ii) Using the value of n found in part(i), find the value of k if $P(k < \bar{Y} < 105) = 0.6730$.

5 The heartbeat rate of a certain population of babies follows a normal distribution with mean 70 beats/min and standard deviation of 10 beats/min.

(i) Find the probability that a baby randomly selected from this population has a heartbeat rate of less than 66 beats/min.

(ii) If a sample of 8 babies is randomly selected, find the probability 3 of them will have a heartbeat rate of less than 66 beats/min

(iii) If a random sample of 36 babies is selected, what is the probability that it will have a mean heartbeat rate of more than 68 beats/min.

6 The masses of Giant apple follow a normal distribution with mean 700 g and standard deviation of 100 g.

(i) Find the probability that the total mass of 10 Giant apples will be more than 7.2 kg.

(ii) A random sample of n apples is chosen. Find the least value of n such that there is a probability of not more than 0.25 that the sample mean differs from its mean mass by more than 20 g.

Answers

A1 i mean = 4, variance = $\frac{7}{15}$, ii mean = 5, variance = $\frac{3}{10}$.

A2 i $\overline{X} \sim N\left(3, \frac{1}{2}\right)$ ii $\overline{X} \sim N\left(3, \frac{1}{12}\right)$

iii $\overline{X} \sim N\left(3, \frac{1}{7}\right)$ approx

A3 0.9960

B1 i 0.0023 ii 0.3479

B2 0.9544

B3 0.2483

B4 i $n = 16$ ii $k = 100$

B5 i 0.3446 ii 0.2771 iii 0.8849

B6 i 0.2643 ii least $n = 34$