Check your Understanding (Sampling)

Distributions of Sample Mean and Sample Sum from a Normal Distribution

1. A random sample of size 15 is taken from a normal distribution with mean 60 and standard deviation 4. Find the probability that the sample mean is less than 58.
[0.0264]

Solution:

$X \sim N(60, 4^2)$

$\overline{X} \sim N(60, \frac{4^2}{15})$

$P(\overline{X} < 58) = 0.0264$

2 **PJC H1 Promo 2017/Q5**

The mass of a student in Aishan Secondary School is normally distributed and denoted by *X* kilograms. The masses of the population mean is 56kg and variance is 37.31544.

(iii) In the school, every class is made up of 30 students. Find the probability that a randomly-chosen class has mean weight more than 54kg. **[2]**

(iv) A sample of *n* students is chosen and their weights recorded. Find the least sample size *n* such that the probability that the mean weight of the students weighing more than 57kg is less than 0.3. Explain briefly if *n* needs to be large. **[4]**

Solutions:

(iii)	For a sample of 30, $\overline{X} \sim \text{N}\left(56, \frac{37.31544}{30}\right)$ $\text{P}\left(\overline{X} > 54\right) = 0.964$
(iv)	For a sample of *n*, $\overline{X} \sim \text{N}\left(56, \frac{37.31544}{n}\right)$ $\text{P}\left(\overline{X} > 57\right) < 0.3$ $\text{P}\left(Z > \frac{57-56}{\sqrt{\frac{37.31544}{n}}}\right) < 0.3$

	$\mathrm{P}\left(Z > \frac{\sqrt{n}}{\sqrt{37.31544}}\right) < 0.3$ $1 - \mathrm{P}\left(Z < \frac{\sqrt{n}}{\sqrt{37.31544}}\right) < 0.3$ $\mathrm{P}\left(Z < \frac{\sqrt{n}}{\sqrt{37.31544}}\right) > 0.7$ Using inverse norm, $\mathrm{P}(Z < 0.5244) = 0.7$ Thus, $\frac{\sqrt{n}}{\sqrt{37.31544}} > 0.5244$ $n > 10.3$ Least n is 11. *n* need not be large as Central Limit Theorem is not used since *X* is normally distributed.

3. **2017/IJC/Prelim/Paper I/11(i)**

A supermarket sells Thailand guava. The masses, in kilograms, of the guava have independent normal distributions. The means and standard deviations of these distributions, and the selling prices, in $ per kilogram, are shown in the following table.

	Mean (kg)	Standard deviation (kg)	Selling price ($ per kg)
Thailand	0.19	0.01	4.20

Stating clearly the mean and variance of all distributions that you use, find the probability that the average mass of 8 randomly chosen Thai guavas is less than 0.18 kg, [2]

Ans: 0.00234

3	Let T be random variable denoting the mass (in kg) of a randomly chosen Thai guava respectively. $T \sim \text{N}(0.19, 0.01^2)$ $\overline{T} \sim \text{N}(0.19, \frac{0.01^2}{8})$ $\overline{T} \sim \text{N}(0.19, 0.0000125)$ $\text{P}(\overline{T} < 0.18) = 0.00234$

Distributions of Sample Mean and Sample Sum from a Non-Normal Distribution

4. A random sample of size 50 is taken from a binomial distribution with number of trials as 9 and probability of success as 0.5. Find the probability that the sample mean exceeds 5.

[0.00921]

Solution:

Since $n = 50$ is large, by Central Limit Theorem, $\overline{X} \sim N(9(0.5), \frac{9(0.5)(0.5)}{50})$

i.e. $\overline{X} \sim N(4.5, \frac{2.25}{50})$ approximately.

$P(\overline{X} > 5) = 0.00921$

5. DHS/2018/Q6
The length of a string is a random variable with mean 15 cm and standard deviation √52 cm. A random sample of 30 strings is taken. Find the probability that the sample mean length lies between 180 mm and 270 mm. [0.0113]

Solution:

Let X be the length of a string (in cm).

$\text{E}(X) = 15, \quad \text{Var}(X) = 52$

Since n is large, by Central Limit Theorem, $\overline{X} \sim \text{N}\left(15, \frac{52}{30}\right)$ approximately.

$\text{P}(18 < \overline{X} < 27) = 0.0113$

6. The rate of consumption of petrol of a certain model of a car is known to have a mean of 13km per litre and standard deviation of 2km per litre.
   (i) A random sample of 50 of the same model of cars is taken. Is this sampling distribution of the mean normally distributed? Explain briefly. State the mean and variance of this sampling distribution.
   (ii) Find the probability of the mean consumption rate to be more than 13.8.

[(i)13, 2/25 (ii) 0.00234]

Solution:

(i) Let X be the random variable "rate of consumption of petrol of the car."
Since X is not normally distributed and $n = 50$ is large, by Central Limit Theorem, $\overline{X}$ is normally distributed.

$\Rightarrow \overline{X} \sim N(13, \frac{2^2}{50})$ approximately

(ii) $P(\overline{X} > 13.8) = 0.00234$

7. David spends time relaxing by watching television everyday. The time X, in hours, which he spends watching television has mean 1.6 hours and standard deviation 1.1 hours.
   (i) If $\overline{X}$ denotes the average time over a randomly chosen period of 100 days state the approximate distribution of $\overline{X}$ and find P($\overline{X} > 1.85$).
   (ii) Find the probability that the total time David spend watching television in a year is more than 590 hours.

[(i) 0.0115 (ii) 0.00215]

Solution:

(i) By Central Limit Theorem, $\overline{X} \sim N(1.6, \frac{1.1^2}{100})$ approximately

i.e. $\overline{X} \sim N(1.6, 0.0121)$ approximately

$P(\overline{X} > 1.85) = 0.011521$
≈ 0.0115

(ii) Let Y denote the random variable denoting the total time David spend watching television in a year.

$Y \sim N(1.6, \times 365, 0.0121 \times 365)$

$Y \sim N(584, 4.4165)$

$\text{P}(Y > 590) = 0.00215$

8. Rachel is a softball player for her school team, and everyday she attends a training session to train on batting softballs. The probability that she successfully bats a softball is p, and each hit is independent of one another. In each training session, Rachel will do 60 batting attempts.
   (i) Let X denotes the number of successful hits out of 60 batting attempts in each training session. State, with a reason, the distribution of the mean number of successful hits, $\overline{X}$, in 85 training sessions.
   (ii) It is given that the probability Rachel will have at most 17 successful hits on average in 85 training sessions is 0.5. Find the value of p, and the value of the standard deviation of $\overline{X}$.

`[(ii) 0.283, 0.379]

Solutions:

(i) Let Y be the random variable "the number of successfully bats a softball out of 60 attempts."

Then $Y \sim B(60, p)$

$E(Y) = 60p$

$Var(Y) = 60p(1 - p)$

Since the sample size 85 is large, therefore by Central Limit Theorem,

$\overline{X} \sim N\left(60p,\ \frac{60p(1-p)}{85}\right)$ approximately.

(ii) Since $P(\overline{X} \leq 17) = 0.5$, therefore $E(\overline{X}) = 17$.

Therefore $60p = 17$

$\Rightarrow p = 0.283$.

Since $p = 0.283$, therefore $Var(\overline{X}) = \frac{60(0.283)(1-0.283)}{85} = 0.143$

Therefore standard deviation of $\overline{X} = \sqrt{0.143} = 0.379$

9 [For this question, define all random variables clearly and show all working clearly.]

An experiment is conducted where a fair 6-sided die is thrown 8 times and the number of sixes obtained is recorded. Find the expected number of sixes obtained and the variance of the number of sixes obtained. **[2]**

60 such experiments are conducted. Find the probability that the mean number of sixes obtained is more than 1.5. **[3]**

9	Let X be the r.v no of sixes obtained for a fair die in 8 throws. $X \sim B\left(8, \frac{1}{6}\right)$ $E(X) = 8\left(\frac{1}{6}\right) = \frac{4}{3}$ $Var(X) = 8\left(\frac{1}{6}\right)\left(\frac{5}{6}\right) = \frac{10}{9}$
	$\overline{X} \sim N\left(\frac{4}{3}, \frac{\frac{10}{9}}{60}\right) \sim N\left(\frac{4}{3}, \frac{1}{54}\right)$ approx by Central Limit Theorem since n = 60 large

	$\text{P}\left(\overline{X} > 1.5\right) = 0.110$ (to 3 sf)

10. **2016/FE/IJC/Qn1**

At a nature reserve, the weight of sloths is found to have mean 4 kg and standard deviation 0.9 kg.

(i) Find the probability that the total weight of 60 randomly selected sloths is at most 250 kg. [3]

(ii) If there is a probability of more than 0.83 that the mean weight of a large sample of n randomly selected sloths is less than 4.1 kg, find the least value of n. [3]

Ans: (i) 0.924 (ii) 74

10(i)	Let X be the random variable denoting the weight of a randomly chosen sloth. Consider $S = X_1 + X_2 + ... + X_{60}$ $\text{E}(S) = 60 \times 4 = 240$ $\text{Var}(S) = 60 \times 0.9^2 = 48.6$ Since the sample size is large (i.e. $n = 60$), by Central Limit Theorem, $S \sim \text{N}\left(240, 48.6\right)$ approximately. $\therefore \text{P}(S \le 250) = 0.92428 = 0.924$ (3 sig fig)
(ii)	Consider $\overline{X} = \dfrac{X_1 + X_2 + ... + X_n}{n}$ Since the sample size is assumed to be large, by Central Limit Theorem, $\overline{X} \sim \text{N}\left(4, \dfrac{0.9^2}{n}\right)$ *<u>Method 1: Setting up a table</u>* Using GC, When $n = 73$, $\text{P}(\overline{X} < 4.1) = 0.82877$ (< 0.83) When $n = 74$, $\text{P}(\overline{X} < 4.1) = 0.83042$ (> 0.83) When $n = 75$, $\text{P}(\overline{X} < 4.1) = 0.83204$ (> 0.83) $\therefore n \ge 74$. Thus the least value of n is 74. *<u>Method 2: Use of standardization</u>* Given: $\text{P}(\overline{X} < 4.1) > 0.83$ $\text{P}\left(Z < \dfrac{4.1 - 4}{0.9/\sqrt{n}}\right) > 0.83$

	$\text{P}\left(Z < \frac{0.1\sqrt{n}}{0.9}\right) > 0.83$ $\frac{0.1\sqrt{n}}{0.9} > 0.954165$ $\sqrt{n} > 8.58748$ $n > 73.7449$ $\therefore$ the least value of n is 74.

11. **2017/AJC/Prelim/Paper1/Q7**

The heights of a certain species of plant at maturity is known to have a mean of 100 cm and standard deviation of 5 cm.

(i) Gardener Jeremy planted 60 seedlings and measured them upon maturity. Find the probability that the average height of his plants is less than 99.1 cm. [2]

(ii) A large number of gardeners take part in a competition. Each gardener planted 60 seedlings and measured them upon maturity. Gardeners whose plants have an average height of more than *k* cm will be awarded a prize. If 5% of the gardeners are awarded a prize, find the minimum value of *k*. [2]

11 (i)	Let X be the height of a randomly selected plant. Since n=60>50 is large, By Central Limit Theorem, $\overline{X} \sim N\left(100, \frac{5^2}{60}\right)$ approximately $P(\overline{X} < 99.1) = 0.0816$
(ii)	$P(\overline{X} > k) = 0.05$ $P(\overline{X} < k) = 0.95$ Min k =102.16 = 102 (3sf)

Ans: (i) 0.0816 (ii) 102

Unbiased Estimates of Population Mean and Variance

12. A company is selling a type of Multi-Purpose Vehicle (MPV). The number of MPV sold in a day, X, over a period of 60 randomly chosen days are recorded and the data obtained is summarised by

$$\sum(x-10)=-399 \qquad \sum(x-10)^2=2845$$

(i) Find the unbiased estimates of the population mean and variance of the number of MPV sold.

(ii) Find the probability that the mean number of MPV sold per day is at least 3 over a period of 100 days.

[(i)3.35, 3.25, (ii) 0.974]

Solution:
(i) Let X be the number of MPVs sold.

Unbiased estimate of the population mean of X,

$$\bar{x}=\left(\frac{-399}{60}\right)+10=3.35$$

Unbiased estimates of the population variance of X,

$$s^2=\frac{1}{59}\left[2845-\frac{(-399)^2}{60}\right]=3.2483\approx 3.25 \text{ (to 3 s.f.)}$$

(ii) Since n is large, by Central Limit Theorem, $\bar{X}\sim N(3.35,\frac{3.2483}{100})$ approximately

$$P(\bar{X}\geq 3)=0.97393\approx 0.974$$

13. The following table illustrates the number of cars sold in a day during a period of 50 days.

No. of cars sold	1	2	3	4	5	6
No. of days	18	8	8	5	7	4

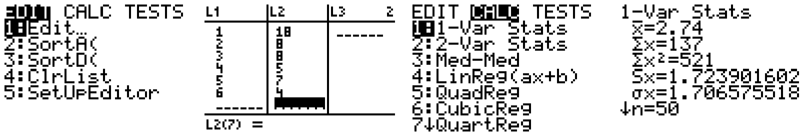
(i) Find the unbiased estimates of the mean and variance of the number of cars sold.

(ii) Find the probability that the mean number of cars sold per day is greater than 3 over a period of 70 days.

[Ans: 2.74; 2.97; 0.103]

Solutions:

(i) From the GC,



unbiased estimate for population mean $= \bar{x} = 2.74$

unbiased estimate for population variance is $s^2 = \frac{n}{n-1} S_x{}^2 = 1.723901602^2 = 2.97$ (3 s.f.)

(ii) By Central Limit Theorem, as sample size n = 70 is large,

$\bar{X} \sim N(2.74, \frac{2.9718}{70})$ approximately

$P(\bar{X} > 3) = 0.103$

14. The exhaustion time of an alkaline battery is defined as the time it takes to discharge completely under controlled experimental conditions. Research has shown that the exhaustion time of AA-size alkaline batteries manufactured by the market leaders, Durazer and Energi-cell, have the following population parameters:

	Mean (in hours)	**Standard deviation (in hours)**
Durazer	19.87	0.24
Energi-cell	20.16	0.19

(i) 50 Durazer and 80 Energi-cell batteries are tested. Determine the distribution of the sample mean exhaustion time for each brand.

(ii) Hence find the probability that the difference in the mean exhaustion times of the two brands of batteries is less than 0.2 hours.

[(ii) 0.0123]

Solutions:

(i) Let X and Y be the exhaustion time (in hours) of a Durazer and an Energi-cell battery respectively.

Since $n_1 = 50$ is large, by the Central Limit Theorem, $\bar{X} \sim \mathrm{N}\left(19.87, \frac{0.24^2}{50}\right)$ approximately.

Since $n_2 = 80$ is large, by the Central Limit Theorem, $\bar{Y} \sim \mathrm{N}\left(20.16, \frac{0.19^2}{80}\right)$ approximately.

(ii) $\mathrm{E}\left(\bar{X} - \bar{Y}\right) = \mathrm{E}\left(\bar{X}\right) - \mathrm{E}\left(\bar{Y}\right) = 19.87 - 20.16 = -0.29$

$\mathrm{Var}\left(\bar{X} - \bar{Y}\right) = \mathrm{Var}\left(\bar{X}\right) + \mathrm{Var}\left(\bar{Y}\right) = \frac{0.24^2}{50} + \frac{0.19^2}{80} = 0.00160325$

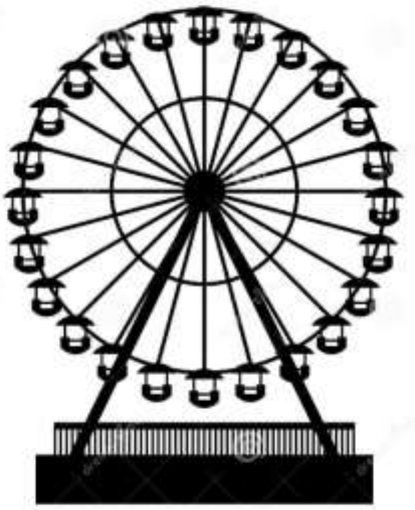
$$\Rightarrow \qquad \overline{X} - \overline{Y} \sim \mathrm{N}\left(-0.29, 0.00160325\right)$$

$$\mathrm{P}\left(\left|\overline{X} - \overline{Y}\right| < 0.2\right) = \mathrm{P}\left(-0.2 < \overline{X} - \overline{Y} < 0.2\right)$$

$$= 0.012297$$

$$= 0.0123 \qquad \text{(to 3 s.f.)}$$

15 TJC/2018/Q7



The diagram shows an observation wheel with 24 capsules. Each capsule can carry passengers up to a maximum load of 3000 kg. The weights of male passengers have mean 70 kg and standard deviation 8.9 kg.

(i) The operator of the observation wheel allows n randomly chosen male passengers to enter a capsule. Find the greatest value of n such that the probability that the total weight of the n male passengers exceed the maximum load is less than 0.01. [Greatest n is 40]

(ii) Explain whether it is necessary to assume that the weights of male passengers are normally distributed.

Solutions:

Let M be the weight of a male passenger

$\text{E}(M) = 70$, $\text{Var}(M) = 8.9^2$

Let $T = M_1 + M_2 + M_3 + \ldots + M_n$

Assuming that n is large, by Central Limit Theorem,

$T \sim \text{N}(70n, 8.9^2 n)$ approximately

(i) $\text{P}(T > 3000) < 0.01$

From GC

n	$\text{P}(T > 3000)$
40	0.00019 < 0.01
41	0.0113 > 0.01

Greatest value of n is 40

(ii) It is not necessary to assume that the weights of male passengers are normally distributed as the total weight of n male passengers is approximately normally distributed by the Central Limit Theorem since n is large.