Date 10.02.22

S-2

Probability and Statistics-2

Sampling and Estimation

SP.20	M-20	S-20	W-20	S-18	W-18
	M-21	S-21	W-21	S-19	W-19

Suresh Goel
(Former Director)
Alliance World School
Noida, Delh. NCR
INDIA

(+91 9810444804)

Scanned with CamScanner

Sampling and Estimation Notes and Revision.

§ **Population and Sample:**

A population is a complete set of items of interest.
A sample is a part of popution.
A sample should be representative of the population.
A sampling technique that results in an unrepresentative sample is said be biased.

§ **Random sample of size n:**

A random sample of size 'n' is a sample selected in such a way that samples of size 'n' that you select from the population have the same chance of being chosen. (Equally likely be chosen)
Otherwise:
The semple technique can be described as biased; the resulting sample is described as unrepresentative.

§ **Genereting a sample:**

Random number can be used to generate a sample in which you have no contol over the selection.
However, using a random sampling method does not guarantee that the resulting sample will be representative of the whole population.

Example 1. Savita wishes to choose a sample of four students from a class of 16 students. The students are numbered from 3 to 18, inclusive. Savita throws three fair dice and adds the scores. Explain why this method of choosing the sample is biased.

Solution: The number (obtained by throwing three dice) 3 to 18 are not equally likely of being selected.

$P(\text{student } 3) = \frac{1}{216}$ (only one case (1,1,1))

$P(\text{Student } 4) = \frac{3}{216} = \frac{1}{72}$ [as (1,1,2), (1,2,1) and (2,1,1) adds to 4]

Scanned with CamScanner

2. Decribe briefly how to use random numbers to choose a sample of 50 employees from a Company with 712 employees.

Solution: Number the employees from 001 to 712.
Generate 50 distinct three-digit numbers (ignoring 000 and numbers above 712, and the numbers that are repeated.)

§ The distribution of Sample means:

The sample mean is the mean of all the items in the chosen sample.
The sample size is the number of items in the Chosen Sample.

§ Expectation and Variance of Sample mean $\overline{X}$.

If we have 'n' samples and calculate their means $\overline{X}$. These means have a distribution, called the distribution sample mean. A sample mean $\overline{X}$ can be regarded as a random Variable.

If a random sample consists of 'n' observations of a random variable X and the mean $\overline{X}$ is found, then,

$$E(\overline{X}(n)) = \mu \quad \text{where } \mu = E(X)$$

and $Var(\overline{X}(n)) = \frac{1}{n} \cdot Var(X)$ or $\frac{\sigma^2}{n}$, $Var(X) = \sigma^2$

§ Central limit theorem (CLT):

The distribution of Sample means of Size 'n' (n is large) is:
(n > 50)

$$\overline{X}(n) \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad \{\text{Normal distribution}\}$$

where the original population has mean μ and Variance σ^2

§ Note: For discrete random variable:
The continuity correction is $\pm \frac{1}{2n}$ { Not $\pm \frac{1}{2}$ }

Scanned with CamScanner

3. It is known that the numbers, N, of words contained in a leading article each day in a certain newspaper can be modelled by a normal distribution with mean 352 and variance 29. A researcher takes a random sample of 10 leading articles and finds the sample mean, $\bar{N}$, of N.

(i) State the distribution of $\bar{N}$, giving the values of any parameters. --[2]

(ii) Find $P(\bar{N} > 354)$ ---[3]

[W-15/73/Q1]

Solution (i) $\bar{N}$ is a normal distribution with mean 352,

number of random samples $= 10 \Rightarrow$ Variance of $\bar{N} = \frac{\sigma^2}{n} = \frac{29}{10}$

$= 2.9$

$\therefore \bar{N} \sim N(352, 2.9)$

(ii) $\therefore$

$$P(\bar{N} > 354) = P\left(Z > \frac{354-352}{\sqrt{2.9}}\right) = P(Z > 1.174)$$

$$= 1 - \phi(1.174) = 1 - 0.880 = 0.120$$

4. The lengths of time people take to complete a certain type of puzzle are normally distributed with mean 48.8 minutes and standard deviation 15.6 minutes. The random variable X represents the time taken, in minutes, by randomly chosen person to solve this type puzzle. The times taken by random samples of 5 people are noted. The mean time $\bar{X}$ is calculated for each sample.

(i) State the distribution of $\bar{X}$, giving the values of any parameters. ---[2]

(ii) Find $P(\bar{X} < 50)$ --[3]

[S-08/07/Q2]

Solution (i) $\bar{X}$ is a normal distribution; with mean $= 48.8$

and Variance $= \frac{\sigma^2}{n} = \frac{15.6^2}{5} = 48.672$

$\therefore \bar{X} \sim N(48.8, 48.672)$

(ii)
$$P(\bar{X} < 50) = P\left(Z < \frac{50-48.8}{\sqrt{48.672}}\right) = P(Z < 0.1720)$$

$$= \phi(0.1720) =$$

$$= 0.568$$

Scanned with CamScanner

5. Dominic wishes to choose a random sample of 5 students from the 150 students in this year. He numbers the students from 1 to 150. Then he uses his calculator to generate five random numbers between 0 to 1. He multiplies each random number by 150 and rounds up to the next whole number to give a student number.

(i) Dominic's first random number is 0.392. Find the student number that is produced by this random number. --[1]

(ii) Dominic's second student number is 104. Find a possible random number that would produced this student number. --[1]

(iii) Explain briefly why five random numbers may not be enough to produce a sample of five student numbers. --[1]

[W-16/73/Q2]

Solution (i) Student number = random number × 150 and rounds upto whole number

$= 0.392 \times 150 = 58.8$ rounds upto 59 ✓

(ii) Let x is the random number for student number 104

$\Rightarrow \quad 103 \leq 150x < 104$

$\Rightarrow \quad 0.6866\ldots \leq x < 0.6933\ldots$

The random number could be any from 0.687 to 0.693, eg 0.692 ✓

(iii) Because of the possibilities of repeats, his calculator may generate two or more random numbers that rounds up to the same student number.

6. Describe briefly how to use random numbers to choose a sample of n students from a year-group of 276 students.

[SP-20/06/Q2] --[3]

Solution: Generate 3 digit numbers. (Number students from 1 to 276).

Ignore number > 276

Ignore repeats.

Scanned with CamScanner

Scanned with CamScanner

7 Andy and Jessica are doing a survey about musical preferences. They plan to choose a representative sample of six students from the 256 students at their college.

(a) Andy suggests that they go to the music building during the lunch hour and choose six students at random from the students who are there. Give a reason why this method is unsatisfactory. ---[1]

(b) Jessica decides to use another method. She numbers all the students in the college from 1 to 256. Then she uses her calculator and generates the following random numbers.

204393, 162007, 204028, 587119, 207395.

From these numbers, she obtains six student numbers. The first three of her student numbers are 204, 162 and 7.

Continue Jessica's method to obtain the next three student numbers. --[2]

[W-21/62/Q2]

Solution (a) Bias towards students who play instruments or only music students or the six will possibly be friends/have similar music preferences.

(b) 28, 119, 207. (Choose first three digits/or last three digits < 257)

8. Luis has to chose one person at random from four people, A, B, C and D. He throws a fair six-sided die. If the score is 1, he will choose A. If the score is 2 he will choose B. If the score is three, he will choos C. If the score is 4 or more he will choose D.

(i) Explain why the choice made by this method is not random. ---[1]

(ii) Describe how Luis could use a single throw of the die to make a random choice. --[1]

(iii) Another day Luis has to choose two people at random from the same four people A, B, C and D. List the possible choices of two people and hence describe how Luis could use a single throw of the die to make this random choice. [S-19/73/Q3]---[2]

Solution (i) D is more likely to be chosen.

(ii) Reject scores of 5 or 6. (or choose D when score is 4)

(iii) AB, AC, AD, BC, BD, CD.; Allocate 1: AB; 2: AC; 3: AD; 4: BC; 5: BD; 6: CD.

Scanned with CamScanner

9. A residents' association has 654 members, numbered 1 to 654. The secretary wishes to send a questionnaise to a random sample of members. In order to choose the members for the sample she uses a table of random numbers. The first line in the table is a follows:
1096, 4357, 3765, 0431, 0928, 9264.
The numbers of the first two members in the sample are 109 and 643. Find the numbers of the next three members in the sample.
[S-17|73|Q1] --[3]

Solution: 573; 43 (043), 289. (How is it done – First three digit, next three digit, -- so not but 'reject' number > 654 and repeated no)

10. Renu wishes to choose a representative sample of six employees from 78 employees at her place of work.
(a) Renu consider taking as her sample the first six people arriving at work one morning. Give two reasons why this method is unsatisfactory.
(b) Renu decides to use the following method to choose her sample. She numbers each employee at her place of work and generates the following random numbers on her calculator:
642, 784, 034, 796, 313, 215, 950, 850, 565, 013, 311, 170, 929
from these random numbers, she chooses employees 40, 47, 63, 32, 59 and 8. Explain how she choose these employees.

Solution (a) Only those who arrive at work early can be selected. Anyone who is absent or working away from the workplace cannot be selected.
(b) Starting with second group of three, she takes the last digit of that group with the first digit of the next group of three.

Scanned with CamScanner

§ For a sample size 'n' drawn from a normal distribution with known variance σ^2, and sample mean $\bar{x}$, the test statistic is: $z = \left(\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\right)$

11. The time, in minutes, taken by students to complete a test has the distribution $N(125, 36)$

(a) Find the probability that the mean time taken to complete the test by a random sample of 40 students is less than 123 minutes. ---[3]

(b) Explain whether it was necessary to the Central Limit theorem in the solution to part (a). ---[1]

[S-21/61/Q2]

Solution: $N(125, 36) \rightarrow \mu = 125, \sigma^2 = 36, \; E(\bar{X}) = 125$, random variable mean time $= \bar{x}$

(a) for $n = 40$, Variance $(\bar{X}(n)) = \frac{\sigma^2}{n} = \frac{36}{40}$ ✓

$$P(\bar{x} < 123) = P\left(z < \frac{123-125}{\sqrt{\frac{36}{40}}}\right) = P(z < -2.108)$$

$$= 1 - \phi(2.108) = 1 - 0.9825 = 0.0175 ✓$$

(b) No, population is normal.

12. The score on one spin of a 5-sided spinner is denoted by the random variable X with prob. distribution as shown in the table:

x	0	1	2	3	4
$P(X=x)$	0.1	0.2	0.4	0.2	0.1

(a) Show that $Var(X) = 1.2$ ---[2]

The spinner is spun 200 times. The score on each spin is noted and mean, $\bar{x}$, of the 200 scores is found.

(b) Given that $P(\bar{X} > a) = 0.1$, find the value of a. ---[4]

(c) Explain whether it was necessary to use central limit theorem in your answer to part (b). [S-20/62/Q4] ---[1]

Solution: $E(X) = \Sigma p_i x_i = 0\times.1 + 1\times.2 + 2\times.4 + 3\times.2 + 4\times0.1 = 2$ ✓

(a) $E(X^2) = \Sigma p x^2 = 0^2\times.1 + 1^2\times.2 + 2^2\times.4 + 3^2\times.2 + 4^2\times.1 = 5.2$

$Var(X) = E(X^2) - (E(X))^2 = 5.2 - 2^2 = 1.2$ ✓

(b) Now $n = 200$ (large), $E(\bar{X}) = 2$, $Var(\bar{X}) = \frac{\sigma^2}{n} = \frac{1.2}{200}$; using Normal distribution

$$P(\bar{X} > a) = P\left(z > \frac{a-2}{\sqrt{1.2 \div 200}}\right) = .1 \Rightarrow 1 - \phi\left(\frac{a-2}{\sqrt{1.2 \div 200}}\right) = 0.1$$

$$\Rightarrow \frac{a-2}{\sqrt{1.2 \div 200}} = \phi^{-1}(0.9) = 1.282 \Rightarrow a = 2 + \sqrt{1.2 \div 200} \times 1.282$$

$$\Rightarrow a = 2.10 \text{ (3 sf)}$$

(c) Yes, use CLT, as X is not normally distributed.

Scanned with CamScanner

13. The random variable X has mean 372 and standard deviation 54.
(i) describe fully the distribution of the mean of a random sample of 36 values of X. --[3]
(ii) The distribution in part (i) might be either exact or approximate. State the condition under which the distribution is exact. --[1]
[S-19/72/Q2]

Solution: Given a random variable $X \rightarrow$ mean $\mu = 372$, and $SD = 54$
(i) Now of random sample of mean $\bar{x}$, $E(\bar{X}) = 372$ ✓ ; $n = 36$ (Given)
The distribution is Normal ; Variance $= \frac{\sigma^2}{n} = \frac{54^2}{36} \Rightarrow S.D = \frac{54}{\sqrt{36}} = 9$ ✓
(ii) X is normal $\rightarrow \bar{x}$ is normal.

14. The length, in centimetres, of a certain type of snake is modelled by the random variable X with mean 52 and standard deviation 6.1. A random sample of 75 snakes is selected, and the sample mean, $\bar{X}$, is found.
(i) Find: $P(51 < \bar{X} < 53)$ --[4]
(ii) Explain why it was necessary to use Central Limit theorem in the solution to part (i). --[1]
[M-17/72/Q3]

Solution: For random variable X: mean $\mu = 52$ and SD $\sigma = 6.1$
(i) For sample mean $\bar{x}$, $n = 75$, $E(\bar{X}) = 52$; $SD = \frac{\sigma}{\sqrt{n}} = \frac{6.1}{\sqrt{75}}$ $\left(\text{Var or } \frac{\sigma^2}{n}\right)$
using Normal distribution for $\bar{x}$,

$$\therefore P(51 < \bar{x} < 53) = P\left(\left(z < \frac{53-52}{6.1/\sqrt{75}}\right) - \left(z < \frac{-52}{6.1/\sqrt{75}}\right)\right)$$

$$= \phi(1.420) - \phi(-1.420)$$

$$= \phi(1.420) - \{1 - \phi(1.420)\}$$

$$= 0.9222 - 1 + 0.9222$$

$$= \underline{0.844} \checkmark \ (3sf)$$

(ii) As X is not stated to be normal,
hence we need to use CLT for $\bar{X}$ to be normal distribution.

Scanned with CamScanner

15 The mass, in tonnes, of iron ore produced per day at a mine is normally distributed with mean 7.0 and standard deviation 0.46. Find the probability that the total amount of iron ore produced in 10 randomly chosen days is more than 71 tonnes. --[5]

[S-17/73/Q3]

Solution: For random variable $X \sim N(7, 0.46^2)$

Now sample for $n = 10$, Total $= 71 \Rightarrow$ Sample mean $E(\overline{X}) = \frac{71}{10} = 7.0$

and $Var(\overline{X}) = \frac{0.46^2}{10}$

for $\overline{X} \sim N\left(7, \frac{0.46^2}{10}\right)$

$\therefore P(\overline{X} > 7.1) = P\left(Z > \frac{7.1 - 7}{0.46/\sqrt{10}}\right) = P(Z > 0.687)$

$= 1 - \phi(0.687) = 1 - 0.754 = 0.246$ (3sf.

⊗ 16. A population has mean 12 and standard deviation 2.5. A large sample of size n is chosen from this population and the sample mean is denoted by $\overline{X}$. Given that $P(\overline{X} < 12.2) = 0.975$, correct to 3 significant figures, find the value of n. --[4]

[W-18/71/Q3]

Solution: X is such that: $\mu = 12$ and $SD = 2.5$

Now for large value of n for sample mean $\overline{X}$, $E(\overline{X}) = 12$

using Normal distribution (CLT); $Var(\overline{X}) = \frac{2.5^2}{n}$ $\left(\text{or } SD = \frac{2.5}{\sqrt{n}}\right)$

$$P(\overline{X} < 12.2) = P\left(Z < \frac{12.2 - 12}{\frac{2.5}{\sqrt{n}}}\right) = 0.975 \text{ (Given)}$$

$$\Rightarrow \frac{12.2 - 12}{\frac{2.5}{\sqrt{n}}} = \phi^{-1}(0.975) = 1.96$$

$$\Rightarrow 0.2 = \frac{1.96 \times 2.5}{\sqrt{n}}$$

$$\Rightarrow \sqrt{n} = \frac{1.96 \times 2.5}{0.2} = 24.5$$

$$\Rightarrow n = (24.5)^2 = 600.25$$

$n = 600$ (or 601) whole number.

Scanned with CamScanner

Estimation

§ Sample statistics and Population statistics:

Population parameters, mean and variance are denoted by μ & σ^2 resp. estimates of population's mean and variance are called population statistics.
Estimates of population parameters from a sample are sample mean $\bar{x}$ and sample variance s^2.
If $\hat{U}$ is some statistic derived from a random sample taken from a population, then $\hat{U}$ is an unbiased estimate for U if $E(\hat{U}) = U$.

The most efficient estimate is one that is unbiased and has the smallest variance.

§ For sample size 'n' taken from a population, σ^2 is the population variance, the expection of population variance $E(V)$ is:
$E(V) = \frac{n-1}{n} \times \sigma^2$; The population mean μ is the sample mean $\bar{x}$.

An unbiased estimate of population variance σ^2 is:

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

$$\sigma = \frac{1}{(n-1)}\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right)$$

Example 17: The daily takings, $\$x$, for a shop were noted on 30 randomly chosen days. The takings are summarised by, $\Sigma x = 31500$ and $\Sigma x^2 = 33\,141\,816$; Calculate unbaised estimates of the population mean and variance of the shops daily takings. --[3]
[S-07/7/Q6]

Solution: Mean $\bar{x} = \frac{\Sigma x}{n} = \frac{31500}{30} = 1050$ ✓

$s^2 = \frac{1}{n-1}\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right]$
$= \frac{1}{(30-1)}\left(33141816 - \frac{(31500)^2}{30}\right)$
$= 2304$ ✓

Scanned with CamScanner

18. The mass in, in kilograms, of a block of cheese sold in a supermarket is denoted by the random variable M. The masses of a random sample of 40 blocks are summarised as follows.

$n = 40, \quad \Sigma m = 20.50, \quad \Sigma m^2 = 10.7280$ --- [3]

(a) Calculate unbiased estimate of the population mean and variance of M.

(b) The price, \$P, of a block of cheese of mass M kg is found using the formula P = 11

Find estimates of the population mean and variance of P ---[3]

[W-21/62/Q1]

Solution (a) Unbiased estimate of mean of M, $\mu = \frac{\Sigma m}{n} = \frac{20.50}{40} = 0.5125$ ✓ $(=\mu)$

$Var(est\ M) = \frac{n}{(n-1)}\left[\frac{\Sigma m^2}{n} - \mu^2\right]$

$s^2 = \frac{40}{39}\left[\frac{10.7280}{40} - (0.5125)^2\right] = 0.0056859$

$= 0.00569$ ✓ (3 sf)

(b) for $P = 11M + 0.50$.

est Mean of $P = (11\mu + 0.5) = 11 \times 0.5125 + 0.5 = 6.1375$

$= 6.14$ ✓ (3 sf)

est Var $P = 11^2 \times est\sigma^2 = 11^2 \times 0.0056859 = 0.688$ ✓ (3 sf)

---

19. A random sample of 100 values of X is taken. These value are summarised as: $n = 100, \ \Sigma x = 1556, \ \Sigma x^2 = 29004$

Calculate unbiased estimates of the population mean and variance of X

[S-20/63/Q1] --[3]

Solution: Est $\mu = \frac{\Sigma x}{n} = \frac{1556}{100} = 15.56$ ✓

$Est\ \sigma^2 = \frac{n}{n-1}\left[\frac{\Sigma x^2}{n} - \mu^2\right]$

$= \frac{100}{99}\left[\frac{29004}{100} - (15.56)^2\right] = 48.4105$

$Est\ \sigma^2 = s^2 = 48.4$ ✓ (3 sf)

Scanned with CamScanner

19. A magazine conducted a survey about the sleeping time of adults. A random sample of 12 adults was chosen from the adults travelling to work on a train.

(i) Give a reason why this is an unsatisfactory sample for the purposes of survey. --[1]

(ii) State a population for which this would be satisfactory. [1]

A satisfactory sample of 12 adults gave numbers of hours of sleeping as shown below;

4.6, 6.8, 5.2, 6.2, 5.7, 7.1, 6.3, 5.6, 7.0, 5.8, 6.5, 7.2

(iii) Calculate unbiased estimates of the mean and variance of the sleeping time of adults. --[3]

[S-08/7/Q1]

Solution: (i) Commuters are not representative of the whole population.

(ii) people who travel to work on (this) train.

(iii) Est Mean $\bar{x} = \frac{4.6 + 6.8 + \text{-------} + 6.5 + 7.2}{12} = \frac{74}{12} = \underline{6.17}$

Unbiased variance (estimate of) $= \frac{1}{(n-1)}\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right)$ $\left[s^2 = \frac{n}{(n-1)}\sigma^2\right]$

$$s^2 = \frac{1}{(12-1)}\left(463.56 - \frac{74^2}{12}\right) \qquad \left\{\Sigma x^2 = 463.56\right.$$

$$s^2 = 0.657$$

§§ If population mean and population variance are unknown:

Sample data can be used to conduct a hypothesis test that the population mean has a particular value, as follows:

For a large sample size n drawn with unknown variance and sample mean $\bar{x}$.

The test statistic is: $z = \left(\frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}\right)$

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

Scanned with CamScanner

20 Amy has to choose a random sample from 265 students in her year at College. She numbers the students from 1 to 265 and then uses random numbers generated by her calculator. The first two random numbers produced by her calculator are 0.213 165 448 and 0.073 165 196. --- [2]

(i) Use these figures to find the numbers of the first four students in her sample.

These are 25 students in Amy's sample. She asked each of them how much money, \$x, they earned in a week, on average. Her results are summarised below.

$$n = 25 \quad \Sigma x = 510 \quad \Sigma x^2 = 13225$$

(ii) Find unbiased estimates of the population mean and variance. --- [3]

(iii) Explain briefly what is meant by 'population' in this question: --- [1]

S-18/73/Q2

Solution (i) 213, 165, 73 (073), 196

[Take a group of 3 digits in continuation (number > 265 and repeated numbers should be rejected).

(ii) est mean $= \frac{\Sigma 510}{25} = 20.4 \checkmark = \mu$ (let)

est variance $s^2 = \frac{n}{n-1}\left[\frac{\Sigma x^2}{n} - \mu^2\right]$

$= \frac{25}{4}\left[\frac{13225}{25} - (20.4)^2\right]$

$= \underline{118} \checkmark$ (3 sf)

(iii) (Average) weekly earnings of all students in Amy's year.

Scanned with CamScanner

21. The diameter, in cm, of pistons made in a certain factory is denoted by X, where X is normally distributed with mean μ and variance σ^2. The diameters of a random sample of 100 pistons were measured, with the follwing results.

$$n = 100, \quad \Sigma x = 208.7, \quad \Sigma x^2 = 435.57$$

(i) Calculate unbiased estimates of μ and σ^2. ---[3]

The pistons are designed to fit into cylinders. The internal diameter, in cm, of the cylinders is denoted by Y, where Y has an independent normal distribution with mean 2.12 and variance 0.000144. A piston will not fit into a cylinder if $Y - X < 0.01$

(ii) Using your answers to part (i) find the probability that a randomly chosen piston will not fit into a randomly chosen cylinder.

[W-15/73/Q7] --[6]

Solution (i) Est mean $\mu = \frac{\Sigma x}{n} = \frac{208.7}{100} = \underline{2.087}$ ✓

$$s^2 = \text{Est } \sigma^2 = \frac{n}{(n-1)}\left[\frac{\Sigma x^2}{n} - \mu^2\right]$$

$$= \frac{100}{99}\left[\frac{435.57}{100} - (2.087)^2\right] = \underline{0.000132} ✓$$

(ii) $E(Y-X) = 2.12 - 2.087 = 0.033$ ✓

$Var(Y-X) = Var\,Y + Var\,X = 0.000144 + 0.00013232 = 0.000276$ ✓

$$P(Y-X < 0.01) = P\left(z < \frac{0.01 - 0.033}{\sqrt{0.00027632}}\right)$$

$$= P(z < -1.384)$$

$$= 1 - \phi(1.384) = 1 - 0.9168$$

$$= \underline{0.0832} ✓$$

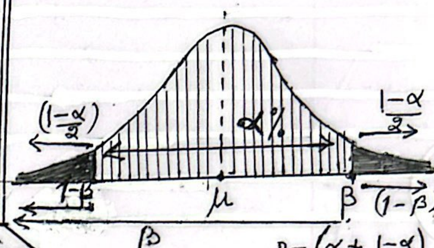
Scanned with CamScanner

Confidence interval (C.I $= \alpha\%$) for population mean.

§ **$\alpha\%$ of confidence interval:**

If we work out sample means for a large number of samples.

$\alpha\%$ of the time we would expect the sample mean $\bar{x}$ to lie within shaded area.



$\beta = \left(\alpha + \frac{1-\alpha}{2}\right)$

$z = \phi^{-1}(\beta)$

$\alpha = \beta - (1-\beta)$

§ For $\alpha\%$ confidence interval (C.I) for a population mean from a normal population with given variance

$$\bar{x} - z \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z \cdot \frac{\sigma}{\sqrt{n}}$$ ✓

Example: For 95% confidence interval.

$$\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}$$

$\alpha = 95\% = 0.95$

$1 - \alpha = 1 - 0.95 = 0.05$

$\beta = \alpha + \frac{(1-\alpha)}{2} = 0.95 + \frac{0.05}{2}$

$= 0.95 + 0.025$

$\beta = 0.975$

$z = \phi^{-1}(\beta) = \phi^{-1}(0.975) = 1.96$

§ For large sample:

C.I is $\bar{x} \pm z \cdot \frac{s}{\sqrt{n}}$

using unbaised estimation Variance s^2,

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

C.I $= \alpha\%$	90%	95%	98%	99%
$\beta = p$	0.95	0.975	0.99	0.995
z	1.645	1.960	2.326	2.576

$P(Z \le z) = p \Rightarrow z = \phi^{-1}(p)$

$z = \phi^{-1}(\beta)$

$\beta = \left(\alpha + \frac{1-\alpha}{2}\right)$

$\alpha = \beta - (1-\beta)$

(*) "Given in the normal distribution table."

22. A doctor wishes to investigate the mean fat content in low-fat burgers. He takes a random sample of 15 burgers and sends them to a laboratory where the mass, in grams, of fat in each burger is determined. The results are as follows:

9, 7, 8, 9, 6, 11, 7, 9, 8, 9, 8, 10, 7, 9, 9.

Assume that the mass, in grams, of fat in low-fat burgers is normally distributed with mean μ and that the population standard deviation is 1.3.

(i) Calculate a 99% confidence interval of μ. --[4]

(ii) Explain whether it was necessary to use the central limit theorem in the calculation in part (i). --[2]

(iii) The manufactures claims that the mean mass of fat in burgers of this type is 8g.
Use your answer to part (i) to comment on this claim. --[2]

[S-11/72/Q4]

Solution:

(i) mean $\bar{x} = \frac{\Sigma x_i}{n} = \frac{126}{15} = 8.4$, $n = 15$, $\sigma = 1.3$

$\alpha = 99\% = .99 \Rightarrow \beta = \alpha + \frac{1-\alpha}{2} = .99 + \frac{1-.99}{2}$

at 99% C.I $\rightarrow$ $\beta = 0.995$, $z = \phi^{-1}(0.995) = 2.576$, $\phi(\beta) = z$

At 99% C.I; $8.4 - 2.576 \times \frac{1.3}{\sqrt{15}} < \mu < 8.4 + 2.576 \times \frac{1.3}{\sqrt{15}}$

$\Rightarrow$ CI; $7.54 < \mu < 9.26$ ✓

At 99% C.I $\left\{ \bar{x} - 2.576 \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2.576 \cdot \frac{\sigma}{\sqrt{n}} \right\}$

(ii) It is not necessary to use the Central Limit Theorem as given the population is normally distributed, so $\bar{X}$ is normally distributed.

(iii) The confidence interval (7.54, 9.26) contains a mass of 8g, so the manufacturer's claim is justified at this level.

Note: $\beta = \alpha + \frac{(1-\alpha)}{2}$

$\phi(z) = \beta \Rightarrow z = \phi^{-1}(\beta)$

Date __ / __ / __
Page No. P-17

23. Leaves from a certain type of tree have lengths that are distributed with standard deviation 3.2 cm. A random sample of 250 of these leaves is taken and the mean length of this sample is found to be 12.5 cm.

(a) Calculate a 99% confidence interval for the population mean length. [3]

(b) Write down the probability that whole of a 99% confidence interval will lie below the population mean. [1]

[SP-20/06/Q1]

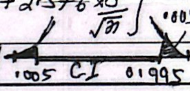
Solution: Standard deviation $\sigma = 3.2$; $\bar{x} = 12.5$, $n = 250$,

(a) for 99% confidence interval; ($z = 2.576$ for $\beta = 0.99 + 0.005 = 0.995$)

CI: $12.5 - 2.576 \times \frac{3.2}{\sqrt{250}} < \mu < 12.5 + 2.576 \times \frac{3.2}{\sqrt{250}}$ $\left\{ \begin{aligned} \beta &= \alpha + \frac{(1-\alpha)}{2} \\ &= 0.99 + \frac{(1-0.99)}{2} \end{aligned} \right.$

$\Rightarrow \underline{12.0 < \mu < 13.0 \; (3 \text{ sf})}$ ✓ $\left\{ \bar{x} - 2.576 \times \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 2.576 \times \frac{\sigma}{\sqrt{n}} \right\}$

(b) $\underline{0.005}$ (or $\underline{0.5\%}$) $\left[\frac{(1-0.99)}{2} = \frac{0.01}{2} = 0.005 \right]$



Scanned with CamScanner

24 100 randomly chosen adults each throw a ball once. The length, l, metres, of each throw is recorded. The results are summarised as: $n = 100$, $\Sigma l = 3820$, $\Sigma l^2 = 182200$
Calculate a 94% confidence interval for the population mean length of throws by adults. ---[6]

S-21 / 61 / Q4

Solution: $n = 100$, $\Sigma l = 3820$, $\Sigma l^2 = 182200$

mean. $\bar{l} = \frac{\Sigma l}{n} = \frac{3820}{100} = \underline{38.2}$ ✓

Unbiased estimated variance $s^2 = \frac{1}{(n-1)}\left(\Sigma x^2 - n\bar{x}^2\right)$

$= \frac{1}{(100-1)}\left(182200 - 100 \times 38.2^2\right) = \frac{12092}{33}$

or $s^2 = \underline{366.424}$ ✓

for 94% CI;

$\left(CI: \bar{x} - z \cdot \frac{s}{\sqrt{n}} < \mu < \bar{x} + z \cdot \frac{s}{\sqrt{n}}\right)$

for 94% C.I $\Rightarrow \alpha = 0.94$
$\beta = \alpha + (1-\alpha)/2$
$= .94 + \frac{0.6}{2} = .97$
0.3
$\beta = 0.97$ 0.3

CI: $38.2 - 1.881 \times \sqrt{\frac{366.424}{100}} < 38.2 + 1.881 \times \sqrt{\frac{366.424}{100}}$

Check;
$\phi^{-1}(0.97)$
$z = 1.881$ ✓
$[z = \phi^{-1}(\beta)]$

$\Rightarrow$
CI: $\underline{34.6 < \mu < 41.8}$ (3 sf)

25. The masses, m kilograms, of flour in a random sample of 90 sacks of flour are summaried as follows:

$$n = 90, \quad \Sigma m = 4509, \quad \Sigma m^2 = 225950$$

(a) Find unbaised estimates of the population mean and variance. --[3]

(b) Calculate a 98% confidence interval for the population mean. --[3]

(c) Explain why it was necessary to use Central Limit theorem in answering part (b). --[1]

(d) Find the probability that the confidence interval found in part (b) is wholly above the true value of the population mean. --[2]

S-21 | 63 | Q4

Solution: (a) est mean $\bar{x} = \frac{\Sigma m}{n} = \frac{4509}{90} = 50.1$ ✓

$$s^2 = \text{est Var}(m) = \frac{n}{n-1}\left(\frac{\Sigma m^2}{n} - \mu^2\right) = \frac{90}{89}\left(\frac{225950}{90} - (50.1)^2\right) = \frac{481}{890}$$
$$= \underline{0.552} \text{ (3sf)} ✓$$

(b) C.I $\quad \bar{x} \pm z\sqrt{\frac{\sigma^2}{n}}$

$$= 50.1 \pm 2.326 \times \sqrt{\frac{481/890}{90}}$$

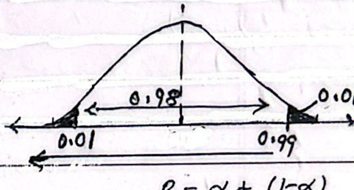
$$\left\{\begin{array}{l} 98\% \text{ C.I} \\ \Rightarrow z = \phi^{-1}(98 + 1\%) \\ z = \phi^{-1}(0.99) = 2.326 \end{array}\right.$$

$$= 49.9 \text{ to } 50.3 \text{ (3 sf)} \Rightarrow \text{CI: } \underline{49.9 < \mu < 50.3} ✓$$

(c) Population of masses is not given to be Normal distribution.

(d) $1 - 0.98 = 0.02$

Req Prob $= 0.02 \div 2 = 0.01$ ✓



$$\beta = \alpha + \frac{(1-\alpha)}{2}$$
$$= 0.98 + \frac{(1-0.98)}{2}$$
$$= 0.98 + .01$$
$$\beta = 0.99$$

26. A construction company notes the time, t days, that it takes to build each house of certain design. The results of a random sample of 60 such houses are summarised as follows: $\Sigma t = 4820$, $\Sigma t^2 = 392050$

(a) Calculate a 98% confidence interval for the population mean time. ---[6]

(b) Explain why it was necessary to use Central Limit Theo. in part (a) ---[1]

M-21/62/Q1

Solution: (a) $\bar{x} = \text{Est}(\mu) = \frac{\Sigma t}{n} = \frac{4820}{60} = \frac{241}{3}$ (or 80.3); $\text{Est}\,\sigma^2 = \frac{60}{59}\left(\frac{392050}{60} - \left(\frac{241}{3}\right)^2\right)$

$s^2 = \text{Est}\,\sigma^2 = 82.0904$

for 98% C.I. $z = \phi^{-1}(0.98 + 0.01)$

$z = \phi^{-1}(0.99) = 2.326$

$\text{C.I} = \bar{x} \pm z\sqrt{\frac{s^2}{n}} = \frac{241}{3} \pm 2.326 \times \sqrt{\frac{82.0904}{60}} = 77.6 \text{ to } 83.1$

$\text{C.I} = 77.6 < \mu < 83.1$

(b) Population distribution of times is unknown. $(n = 60 > 50)$

27. The lengths, X centimeters, of a random sample of 7 leaves from a certain variety of tree are as follows: 5.2, 4.8, 5.5, 6.1, 4.8, 3.9, 4.4.

(a) Calculate unbiased estimates of the population mean and variance of X. ---[3]

It is now given that the true value of the population variance of X is 0.55, and that X has normal distribution.

(b) Find a 95% confidence interval for the population mean of X ---[3]

S-20/61/Q1

Solution (a) Est mean $\bar{x} = \frac{\Sigma x}{n} = \frac{34.7}{7} = 4.9571$ or 4.96 (3sf) ; $\Sigma x^2 = 175.15$

$s^2 = \text{Est Var.}\,\sigma^2 = \frac{n}{n-1}\left(\frac{\Sigma x^2}{n} - (\bar{x})^2\right) = \frac{7}{6}\left(\frac{175.15}{7} - (4.9571)^2\right) = 0.523$ (3sf)

(b) 95% C.I $\rightarrow z = 1.96$

$z = \phi^{-1}\left(0.95 + \frac{0.05}{2}\right)$

$= \phi^{-1}(0.975)$

$z = 1.96$

$\text{C.I} = \bar{x} \pm z\sqrt{\frac{s^2}{n}} = 4.96 \pm 1.96\sqrt{\frac{0.523}{7}}$

C.I: 4.42 to 5.49 (3sf)

$\beta = \alpha + \frac{(1-\alpha)}{2}$

$z = \phi^{-1}(\beta)$

28. The height, h centimetres, of a random sample of 100 fully grown animals of a certain species were measured. The results are summarised as: $n=100$, $\Sigma h = 7570$, $\Sigma h^2 = 588050$

(a) Find unbiased estimates of population mean and variance. --[3]

(b) Calculate a 99% confidence interval for the mean height of animals of this species. --[3]

Four random samples were taken and a 99% confidence interval for the population mean, μ, was found from each sample.

(c) Find the probability that all four of these confidence intervals contain the true value of μ. --[2]

S-21/62/Q6

Solution: $n = 100$, $\Sigma h = 7570$, $\Sigma h^2 = 588050$

(a) $est(\mu) = \frac{\Sigma h}{n} = \frac{7570}{100} = \underline{75.70}$ ✓

$$est(\sigma^2) = s^2 = \frac{1}{(n-1)}\left(\Sigma h^2 - \frac{(\Sigma h)^2}{n}\right) = \frac{1}{(100-1)}\left(588050 - \frac{(7570)^2}{100}\right)$$

$$\Rightarrow s^2 = 152 \text{ (3sf)} \quad (=151.525)$$

(b) CI: $\bar{h} - z\cdot\frac{s}{\sqrt{n}} < \mu < \bar{h} + z\cdot\frac{s}{\sqrt{n}}$ $\left[z = 2.576 \text{ (for 99\% C.I)}^{*} \; \beta = 0.995\right]$

$$\Rightarrow 75.7 - 2.576\times\sqrt{\frac{151.525}{100}} < \mu < 75.7 + 2.576\times\sqrt{\frac{151.525}{100}}$$

C.I: $\underline{72.5 < \mu < 78.9}$ ✓

(c) $P(\text{four C.I}) = (0.99)^4$

$= \underline{0.961 \text{ (3sf)}}$

$*$
$$\left\{\begin{aligned} \beta &= \alpha + \left(\frac{1-\alpha}{2}\right) \\ &= .99 + \left(\frac{1-0.99}{2}\right) \\ &= 0.99 + 0.005 \\ &= 0.995 \\ z &= \phi^{-1}(\beta) = \phi^{-1}(0.995) \\ &= 2.576 \end{aligned}\right.$$

§ Confidence interval (CI) for population proportion:

For a large random sample, size n, an approximate confidence interval for the population proportion, $\hat{p}$, is:

$$\left(\hat{p} - z\cdot\sqrt{\frac{p(1-p)}{n}},\ p + z\cdot\sqrt{\frac{p(1-p)}{n}}\right) \qquad \left\{z = \phi^{-1}\left(\alpha + \frac{(1-\alpha)}{2}\right)\right.$$

Where z is determined by the percentage level of the (α%) confidence interval. (Width of $CI = 2z\sqrt{\frac{p(1-p)}{n}}$)

29. In a random sample of 200 share holders of a company, 103 said that they wanted a change in the management.

(i) Find an approximate 92% confidence interval for the proportion, p, of all share holders who want a change in the management. --[3]

(ii) State the probability that 92% confidence interval does not contain p. --[1]

[S-17/73/Q2]

Solution: $p = 103/200$

$z = \phi^{-1}\left[.92 + \frac{1-.92}{2}\right]$

$\left[92 \pm 4 \right.$

$4\% < \mu < 96\%$

(i) At 92% $CI \Rightarrow z = \phi^{-1}(96\%) = \phi^{-1}(0.96) = 1.751$

or $z = 1.751$ ✓

$$CI: \quad \frac{103}{200} - 1.751 \times \sqrt{\frac{\frac{103}{200}\left(1 - \frac{103}{200}\right)}{200}},\quad \frac{103}{200} + 1.751 \times \sqrt{\frac{\frac{103}{200}\left(1 - \frac{103}{200}\right)}{200}}$$

$\Rightarrow CI$: <u>0.453 to 0.577</u> (3 sf) as final answer.

(ii) $P(92\%\ CI \text{ does not contain } p) = 1 - 0.92 = 0.08$ $\left\{ \begin{array}{l} (100-92)\% \\ = 8\% \end{array}\right.$

$\alpha = 92\% = 0.92$

$\beta = \alpha + \frac{(1-\alpha)}{2}$

$= 0.92 + \frac{(1-0.92)}{2}$

$= 0.92 + 0.04$

$= 0.96$

$z = \phi^{-1}(\beta) = \phi^{-1}(0.96)$

$= 1.751$

3. Q. Ramesh plans to carry out a survey in order to find out what adults in his town thinks about local sports facilities. He chooses a random sample from the adult members of tennis club and gives each one of them a questionnaire.

(i) Give a reason why this will not result in Ramesh having a random sample of adults who live in the town. ---[1]

(ii) Describe briefly a valid method that Ramesh could use to choose a random sample of adults in the town. ---[2]

Ramesh now uses a valid method to choose a random sample of 350 adults from the town. He finds that 47 adults think that the local sports facilities are good.

(iii) Calculate an approximate 90% confidence interval for the proportion of all adults in the town who thinks that the local sports facilities are good. ---[4]

(iv) Ramesh calculates a confidence interval whose width is 1.25 times the width of his 90% confidence interval. Ramesh's new interval is an x% confidence interval. Find the value of x. [S-19/71/Q6] ---[3]

Solution: (i) Biased towards people who like tennis, Excludes people who don't like tennis.

(ii) Obtain a list of all people in the town, use random numbers.

(iii) $\hat{p} = 47/350$

$\text{Var}(\hat{p}) = \dfrac{p(1-p)}{n} = \dfrac{\frac{47}{350}\left(1-\frac{47}{350}\right)}{n}$

for 90% confidence interval, $z = 1.645$

$CI = 90\% = 0.9$

$z = \phi^{-1}\left(0.9 + \dfrac{(1-.9)}{2}\right)$

$= \phi^{-1}(0.9 + 0.5)$

$= \phi^{-1}(0.95)$

$= 1.645$

CI: $\left(p - z\sqrt{\dfrac{p(1-p)}{n}},\ p + z\sqrt{\dfrac{p(1-p)}{n}}\right)$

$\dfrac{47}{350} - 1.645\sqrt{\dfrac{\frac{47}{350}\left(1-\frac{47}{350}\right)}{350}};\ \dfrac{47}{350} + 1.645\sqrt{\dfrac{\frac{47}{350}\left(1-\frac{47}{350}\right)}{350}}$

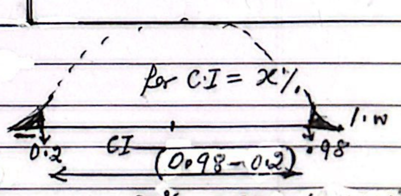
0.104 to 0.164 ✓

(iv) New C.I $= 1.25 \times 1.645 = 2.056$ ✓

$\phi(2.056) = 0.98$

$\therefore x = 96$ $[x\% = 0.96]$

for C.I = x%.



$x\% = 0.98 - (1 - 0.98)$
$= 0.96$

31. A random sample of 75 students at a large college was selected for a survey. 15 of those students said they owned a car. From this result an approximate $\alpha\%$ confidence interval from the proportion of all students at the college who own a car was calculated. The width of this interval was found to to be 0.162.
Calculate the value of α correct to 2 significant figures. --[5]
[W-21/61/Q3]

Solution: $\text{est}(p) = \frac{15}{75} = 0.2$ and $(1-p) = 1 - 0.2 = 0.8$, $n = 75$

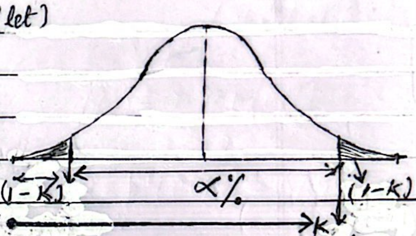
Let z corresponds to $\alpha\%$ CI

Given width of C.I $= 2z\sqrt{\frac{p(1-p)}{n}} = 0.162$ (Given)

$$\Rightarrow 2z\sqrt{\frac{0.2\times0.8}{75}} = 0.162 \Rightarrow z = 1.754$$

Now $\phi(z) = \phi(1.754) = 0.96 = k$ (let)

$\therefore$



$\alpha = 0.96 - (1 - 0.96)$

$\alpha = 2\times 0.96 - 1 = 0.92$

CI: $\alpha = 92\%$

$\alpha = k - (1-k)$

$= .96 - (1 - .96) = .92$

(or $k\% = \frac{1+\alpha\%}{2}$)

32. Sunita has a six-sided die with faces marked 1, 2, 3, 4, 5, 6. The probability that the die shows a six on any throw is 'p'. Sunita throws the die 500 times and finds that it shows a six 70 times.

(a) Calculate an approximate 99% confidence interval for p. ---[4]

(b) Sunita believes that the die is fair. Use your answer to part (a) to comment on her belief. ---[1]

(c) Sunita uses the result of her 500 throws to calculate an α% confidence interval for p. This interval has width 0.04. Find the value of α. ---[5]

S-20/63/Q5

Solution: $n = 500$, $p = \frac{70}{500} = 0.14$, Confidence interval $= 99\%$, $(1 - .99) = .01$

(a) $\phi(z) = k = .99 + \frac{0.01}{2} = 0.995$

$z = \phi^{-1}(0.995) = 2.576$

$$CI = p \pm z \times \sqrt{\frac{p(1-p)}{n}}$$

$$CI = 0.14 + 2.576 \times \sqrt{\frac{0.14(1-0.14)}{500}} = 0.14 \pm 0.04 \Rightarrow 0.100 \leq p \leq 0.180$$

(b) $P(X=6) = 1/6 = 0.1666$--- which lies within the confidence interval, hence the belief is justified.

(c) $$\text{width of } CI = 2z \cdot \sqrt{\frac{p(1-p)}{n}} = 0.04$$

$$\Rightarrow 2z\sqrt{\frac{0.14 \times (1-0.14)}{500}} = 0.04 \Rightarrow z = 1.289$$

$$\phi(1.289) = 0.9013$$

$$C.I \ \alpha\% = 0.9013 - (1 - 0.9013) = 0.803$$

$$\therefore CI = 80.3\%$$

or $\alpha = 80.3$ ✓

33. In a survey, a random sample of 250 adults in Fromleigh were asked to fill a questionnaire about their travel.

(a) It was found that 102 adults in the sample travel by bus. Find an approximate 90% confidence interval for the proportion of all the adults in Fromleigh who travel by bus. ---[3]

(b) The survey included a question about the amount, x dollars, spent on travel per year. The results are summarised as follows: $n = 250$, $\Sigma x = 50460$, $\Sigma x^2 = 19854200$

Find unbiased estimates of the population mean and variance of the amount spent per year on travel. ---[3]

A councillor wanted to select a random sample of houses in Fromleigh. He planned to select the first house on each of the 143 streets in Fromleigh.

(c) Explain why this would not provide a random sample. ---[1]

[W-20/61/Q2]

Solution: (a) $n = 250$, Travel by bus $= 102$, $\Big\{$ 90% Confidence interval

$p = \frac{102}{250}$

$\phi(z) = 0.90 + \frac{(0.1)}{2}$

$= 0.95$

$z = \phi^{-1}(0.95) = 1.645$ ✓

$CI = p \pm z\sqrt{\frac{p(1-p)}{n}}$

$= \frac{102}{250} \pm 1.645\sqrt{\frac{\frac{102}{250}\left(1-\frac{102}{250}\right)}{250}}$

CI is 0.357 to 0.459 (3s.f) ✓

(b) $n = 250$, $\Sigma x = 50460$, $\Sigma x^2 = 19854200$

Estimate of mean $= \frac{\Sigma x}{n} = \frac{50460}{250} = \$201.84$ ✓

Estimate of Var. $s^2 = \frac{1}{(n-1)}\left(\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right) = \frac{1}{249}\left(19854200 - \frac{50460^2}{250}\right)$

$= 38832.75$ dollars^2. ✓

(c) Every house does not have an equal chance of being selected, (or most houses have no chance of being selected). ✓

34. The probability that a certain spinner lands on red on any spin is p. The spinner is spun 140 times and it lands on red 35 times.

(a) Find an approximate 96% confidence interval for p. --- [3]

From further three experiments, Jack finds a 90% confidence interval, a 95% confidence interval and a 99% confidence interval for p.

(b) Find the probability that exactly two of these confidence intervals contain the true value of p. --- [3]

[W-21/62/Q3]

Solution (a) For 96% C.I ; $p = \frac{35}{140} = 0.25$

for 96% C.I
$z = \phi^{-1}(0.96 + 0.02)$
$= \phi^{-1}(0.98)$
$z = 2.054$ ✓

CI is: $p \pm z \sqrt{\frac{p(1-p)}{n}}$

$= 0.25 \pm 2.054 \sqrt{\frac{0.25 \times 0.75}{140}}$

$=$ 0.175 to 0.325 (3 sf)

(b) P(Exactly 2 CI out of 3 contain true p) $= 0.90 \times 0.95 \times 0.01 + 0.1 \times .95 \times .99 + 0.90 \times 0.05 \times 0.99 = 0.147$ ✓

35. A six-sided die has faces marked 1, 2, 3, 4, 5, 6. When the die is thrown 300 times it shows a six on 56 throws.

(a) Calculate an approximate confidence interval for the probability that the die shows a six on one throw.

(b) Marsilla claims the die is biased. Use your answer to part (a) to comment on this claim. -- [1]

[W-20/62/Q2]

Solution (a) $CI = \frac{56}{300} \pm 2.054 \sqrt{\frac{\frac{56}{300} \times \frac{244}{300}}{300}}$; $p = \frac{56}{300}$ and $z = \phi^{-1}[0.96 + 0.02]$
$= \phi^{-1}(0.98) = 2.054$

$=$ 0.14 to 0.233 ✓

(b) $\frac{1}{6} = 0.167$ [Prob. of getting six on a fair die]

This within confidence interval, so no reason to believe die is biased.

Date 10.02.22

Unit-4

S-2

Probability and Statistics-2

Sampling and Estimation-

SP.20	M-20	S-20	W-20	S-18	W-18
	M-21	S-21	W-21	S-19	W-19

Contents:

1. Expectation and variance of sample mean $\bar{X}$. $E(\bar{X}) = \mu$; $\mu = E(X)$; $Var(\bar{X}) = \frac{\sigma^2}{n}$; $Var(X) = \sigma^2$ — Page 2
2. Central Limit theorem C.L.T. $\bar{X}(n) = N(\mu, \frac{\sigma^2}{n})$; $n > 50$.
3. Test statistic: $z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ for normal distribution. --- Page-7

4. Estimation:

Unbiased estimate of population variance σ^2:

$$s^2 = \frac{n}{n-1}\sigma^2 = \frac{n}{n-1}\left(\frac{\Sigma x^2}{n} - \bar{x}^2\right) \text{ --- Page 10}$$

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

5. Confidence Interval, $\bar{x} - z\cdot\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z\cdot\frac{\sigma}{\sqrt{n}}$ $\left(\beta = \alpha + \frac{(1-\alpha)}{2}\right)$ & $z = \phi^{-1}(\beta)$ → Page-15

$\alpha =$ for 96%, $z = \phi^{-1}(.96 + 0.02) = \phi^{-1}(.98) = 2.054$

6. Confidence interval (C.I) for population proportion 'p'

$$p - z\sqrt{\frac{p(1-p)}{n}},\quad p + z\sqrt{\frac{p(1-p)}{n}} \quad ; z = \phi^{-1}\left(\alpha + \left(\frac{1-\alpha}{2}\right)\right) \text{ ---- Page-22}$$

→ $\alpha = C.I$