

SURVEYS

Census

Every member of population is surveyed

Sample

Survey that covers less than 100% population.

POPULATION

Particular group of individual or items.

eg: Students in a school.
Hospitals in Pakistan.

BIAS

Goal is to choose a sample that is not biased.

Each individual in population must get an equal chance.

SAMPLING METHOD

1- Simple random sampling:

Draw Lots (suitable for small populations)

- Write all names
- Put names in a box
- Select a sample without replacement.

Random Number Table

- Assign numbers to all members
- Select sample of required size
- Omit repeats
- Discard out of range values.

Q:

6	8	7	2	5	3	8	1	5	9
2	5	3	4	7	0	5	4	9	5
3	2	6	8	7	4	4	7	0	5

use this random number table to select a

(a) random sample of 8 people from group of 100 people

Person 1	2	3		99	100
01	02	03		99	00

68, 72, 53, 81, 59, 25, 34, 70

(b) random sample of 8 people from group of 60 people

Person 1	2	3	... 60
01	02	03	60

53, 59, 25, 34, 54, 32, 47, 05.

Q: 52 74 54 80 68 72 51 96 08 00
 02 52 09 93 60 43 57 42 13 44

Extract random sample of 12 numbers each to
 2 dp $0 \leq x \leq 10$

5.27, 4.54, 8.06, 8.72, 5.19, 6.08, 0.00, 2.52
 0.99, 3.60, 4.35, 7.42.

STATISTICS OF SAMPLING

X = population



Mean = $E(X) = \mu$

Variance = $V(X) = \sigma^2$

$\bar{X}$ = Sample mean



$E(\bar{X}) = \mu$

$Var(\bar{X}) = \frac{\sigma^2}{n}$

n_s = size of sample

Population

$X \sim N(\mu, \sigma^2)$

$E(X) = \mu$

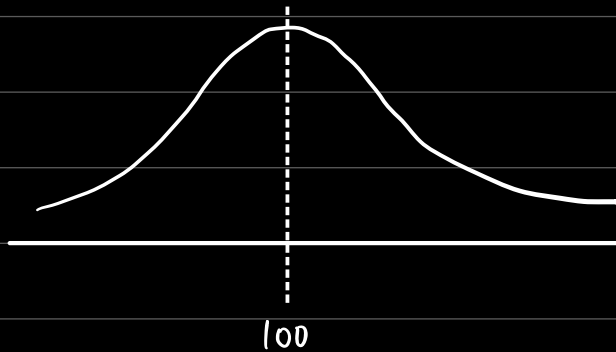
$Var(X) = \sigma^2$

Sampling

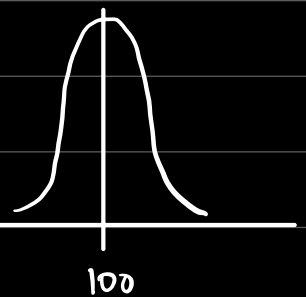
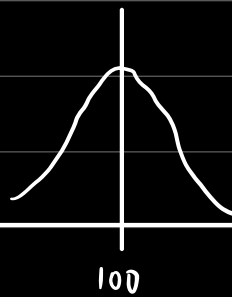
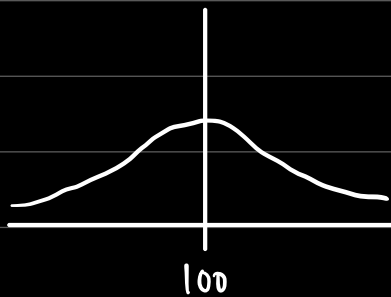
$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n_s}\right)$

Case 1: POPULATION OF X IS NORMAL
(value of variance (σ^2) is known)

$$X \sim N(100, 64)$$



Sample Size: $n=2$ $n=5$ $n=25$



$$\text{Mean}(\bar{X}) = \mu$$

$$Z = \frac{(X - \mu)\sqrt{n}}{\sigma}$$

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

$$SD = \sqrt{\frac{\sigma^2}{n}} = \frac{8}{\sqrt{n}}$$

Q: $X \sim N(25, 340)$

Mean of random sample of size n is drawn from this distribution of $\bar{x}$. Find value of n correct to 2 sf given that $P(\bar{x} > 28)$ is approximately 0.005

$$X \sim N(25, 340)$$

$$\bar{X} \longrightarrow \mu = 25$$

$$\sigma^2 = \frac{340}{n}$$

$$\longrightarrow SD = \sigma = \sqrt{\frac{340}{n}}$$

$$P(\bar{x} > 28) = 0.005$$

$$z = \frac{X - \mu}{\sigma} = \frac{28 - 25}{\sqrt{340/n}}$$

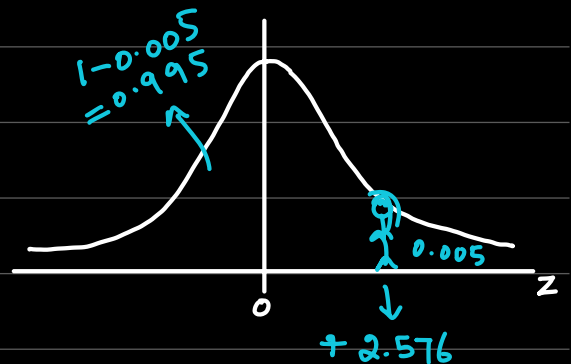
$$P(z > \boxed{+2.576}) = 0.005$$

$$2.576 = \frac{28 - 25}{\sqrt{340/n}}$$

$$\sqrt{\frac{340}{n}} = \frac{28 - 25}{2.576}$$

$$n = 250.68$$

$$n = 250 \text{ (2sf)}$$



Case 2: X is NOT NORMALLY DISTRIBUTED

Central Limit theorem

(CLT) For $n_s \geq 30$ (Large n)

$\bar{X}$ is normally distributed



$$\text{Mean}(\bar{X}) = \lambda$$

$$\text{Var}(\bar{X}) = \frac{\lambda}{n_s}$$

$$\text{Mean}(\bar{X}) = np$$

$$\text{Var}(\bar{X}) = \frac{npq}{n_s}$$

$$\bar{X} \sim N\left(\lambda, \frac{\lambda}{n_s}\right)$$

$$\bar{X} \sim N\left(np, \frac{npq}{n_s}\right)$$

you do not need to do correction of continuity while doing CLT.

DISTRIBUTION OF SAMPLE PROPORTION $P_s = \hat{p}$

Sample n_s



Success = P_s

failure = $q_s = 1 - P_s$

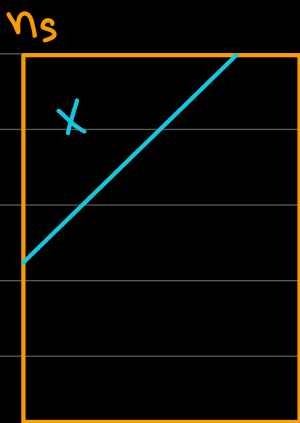
X = no. of success in sample

X will follow Binomial Distribution.

$$X \sim B(n, p)$$

$$E(X) = np$$

$$\text{Var}(X) = npq$$



$$P_s = \frac{x}{n_s}$$

$$E(P_s) = P_s, \quad V(P_s) = \frac{P_s q_s}{n_s}$$

When n is large : Binomial $\longrightarrow$ Normal.

$P_s \sim N\left(p, \frac{pq}{n}\right)$ approximation.
we will need to adjust for continuity correction.

Here, continuity correction is not $\pm \frac{1}{2}$

$$\text{Since } p_s = \frac{1}{n_s} \times x \Rightarrow \text{Continuity Correction} = \frac{1}{n_s} \times \frac{1}{2} = \pm \frac{1}{2n_s}$$

$$p = 0.03$$

Q: It is known that 3% of frozen pies delivered to a canteen are broken. Find probability that, on a morning when 500 pies are delivered, 5% or more are broken? $\rightarrow n_s = 500$
proportion

$$p_s \sim N\left(p_s, \frac{p_s q_s}{n_s}\right) \rightarrow p_s \sim N\left(0.03, \frac{0.03 \times 0.97}{500}\right)$$

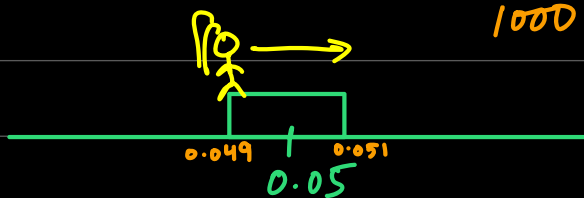
$$P(p_s \geq 5\%)$$

$$\text{continuity correction} = \frac{1}{2n_s} = \frac{1}{2 \times 500}$$

$$P(p_s \geq 0.05)$$

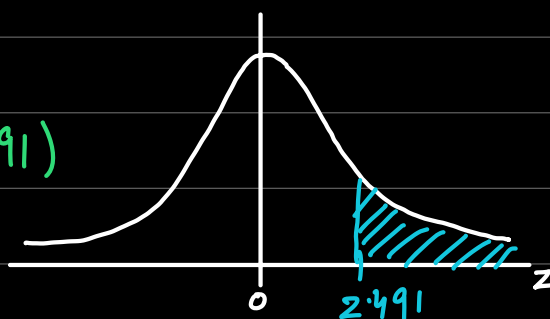
$$= \frac{1}{1000} = 0.001$$

$$P(p_s \geq 0.049)$$



$$Z = \frac{X - \mu}{\sigma} = \frac{0.049 - 0.03}{\sqrt{\frac{0.03 \times 0.97}{500}}} = 2.491$$

$$P(Z \geq 2.491) = 1 - \Phi(2.491) = 0.0064$$



NOTATION

	True Population	Sample	Unbiased estimates of population
Mean	μ	$\bar{x}$	$\hat{\mu}$
Variance	σ^2	s^2	$\hat{\sigma}^2$

UNBIASED ESTIMATES OF POPULATION PARAMETERS

1 POINT ESTIMATES

Sample Size = n_s

UNBIASED ESTIMATE	Symbol	FORMULA
proportion of success (binomial ka p)	$\hat{p}$	$\hat{p} = p_s = \frac{X}{n_s}$
Population Mean	$\hat{\mu}$	$\hat{\mu} = \bar{x} = \frac{\sum x}{n_s}$
Variance	$\hat{\sigma}^2$	$\hat{\sigma}^2 = \frac{n_s}{n_s - 1} \left(\frac{\sum x^2}{n_s} - (\bar{x})^2 \right)$ $\hat{\sigma}^2 = \frac{n_s}{n_s - 1} \left(\frac{\sum (x - \bar{x})^2}{n_s} \right)$

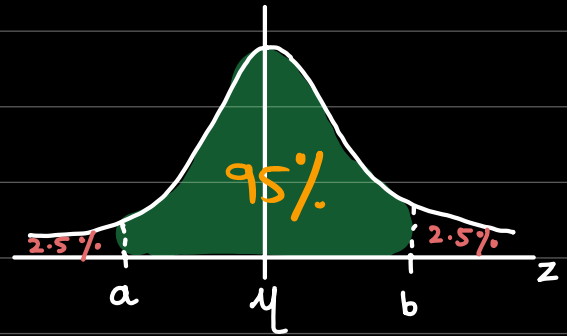
2

CONFIDENCE INTERVAL

Mean $\rightarrow$ 95% $\rightarrow$ (a, b)

CHECKLIST

- 1- Name / Type of Distribution
- 2- Variance (σ^2) $\Rightarrow$ Known / unknown
- 3- $n_s \Rightarrow$ Large / small.

CONFIDENCE INTERVAL FOR μ

CASE A	CASE B	CASE C	CASE D
normal	non-normal	normal	non-normal
Var (σ^2) Known	Var (σ^2) Known	σ^2 not known	σ^2 not known
$n_s =$ Any size	$n_s \geq 30$ (large)	$n_s \geq 30$ (large)	$n_s \geq 30$ (large)
CLT not applied	CLT applied	CLT not applied	CLT applied
$X \sim N(\mu, \sigma^2)$		$X \sim N(\mu, \sigma^2)$	
$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n_s}\right)$	$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n_s}\right)$	$\bar{X} \sim N\left(\mu, \frac{\hat{\sigma}^2}{n_s}\right)$	$\bar{X} \sim N\left(\mu, \frac{\hat{\sigma}^2}{n_s}\right)$
Perfectly Normal	Approx. Normal	Perfectly Normal	Approx. Normal
$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n_s}}}$	$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n_s}}}$	$z = \frac{\bar{x} - \mu}{\frac{\hat{\sigma}}{\sqrt{n_s}}}$	$z = \frac{\bar{x} - \mu}{\frac{\hat{\sigma}}{\sqrt{n_s}}}$
C.I = $\bar{x} \pm z \frac{\sigma}{\sqrt{n_s}}$	C.I = $\bar{x} \pm z \frac{\sigma}{\sqrt{n_s}}$	C.I = $\bar{x} \pm z \frac{\hat{\sigma}}{\sqrt{n_s}}$	C.I = $\bar{x} \pm z \frac{\hat{\sigma}}{\sqrt{n_s}}$
width = $2z \frac{\sigma}{\sqrt{n_s}}$	width = $2z \frac{\sigma}{\sqrt{n_s}}$	width = $2z \frac{\hat{\sigma}}{\sqrt{n_s}}$	width = $2z \frac{\hat{\sigma}}{\sqrt{n_s}}$

CONFIDENCE INTERVALS FOR SAMPLE PROPORTION

$$\hat{p} = p_s = \frac{X}{n_s}$$

$$\text{For: } n \geq 30, \quad p_s \sim N\left(p_s, \frac{p_s q_s}{n_s}\right)$$

$$\text{CI for } p_s \longrightarrow p_s \pm z \sqrt{\frac{p_s q_s}{n_s}}$$

$$\text{Width} = 2z \sqrt{\frac{p_s q_s}{n_s}}$$

Do continuity correction of $\pm \frac{1}{2n_s}$ if finding probability;

Q. Mass of a tablet is normally distributed with standard deviation of 0.042 mg. A random sample of 5 tablets analysed and the mean of mass of tablets was found to be 5.12 mg. Calculate a symmetric 95% confidence interval for population mean of mass of tablets. Give values of end points of intervals to 3 sf.

Check list: 1) Normal

2) $SD = 0.042 \Rightarrow \text{Var} = \sigma^2 = (0.042)^2$

3) $n_s = 5$

Case A

$$C.I = \bar{x} \pm z \frac{\sigma}{\sqrt{n_s}}$$

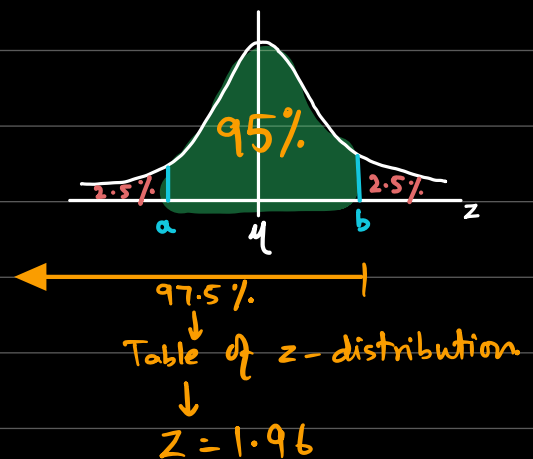
$$= 5.12 \pm \frac{(1.96)(0.042)}{\sqrt{5}}$$

$$= 5.12 \pm 0.0368$$

$$(5.0832, 5.1568)$$

↓

$$(5.08, 5.16)$$



(iii) Find an estimate for mean:

$$\text{Mean} = \text{Midpoint} = \frac{5.08 + 5.16}{2} = 5.12$$

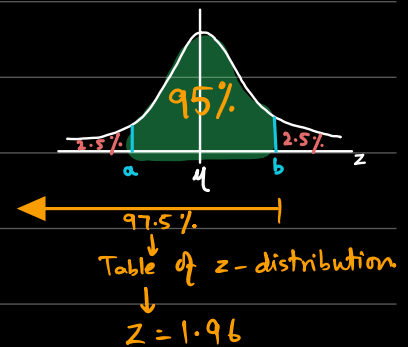
Q: The heights of men in a particular district are distributed with mean μ , and standard deviation σ .

On basis of results from a random sample of 100 men from district, the 95% confidence interval for μ was (177.22, 179.18)

Calculate:

(a) Value of sample mean ($\bar{x}$) and σ

checklist: $\rightarrow$ Not Normal
 $\rightarrow$ Var Known
 $\rightarrow n_s \geq 30$ } Case B.



$$C.I = \bar{x} \pm z \frac{\sigma}{\sqrt{n_s}}$$

$$\bar{x} - z \frac{\sigma}{\sqrt{n_s}} = 177.22$$

$$\bar{x} + z \frac{\sigma}{\sqrt{n_s}} = 179.18$$

$$\bar{x} - \frac{1.96\sigma}{\sqrt{100}} = 177.22$$

$$\bar{x} + \frac{1.96\sigma}{\sqrt{100}} = 179.18$$

$$2\bar{x} = 356.4$$

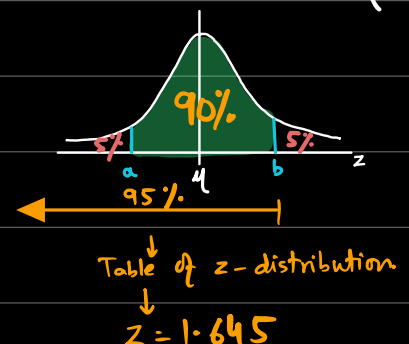
$$\bar{x} = 178.2$$

$$\sigma = 5$$

(b) Calculate a symmetric 90% confidence interval of μ .

$$C.I = \bar{x} \pm z \frac{\sigma}{\sqrt{n_s}}$$

$$= 178.2 \pm \frac{(1.645)(5)}{\sqrt{100}}$$



$$= (177.33, 179.02)$$

Q3) The fuel consumption of a new model of car is being tested. In one trial 50 cars are chosen at random. The tests under identical conditions for distance covered on 1 litre gave following results:

$$\sum x = 525$$

$$\sum x^2 = 5625$$

Calculate 95% confidence interval for mean.

Checklis: 1) Non-Normal
2) σ^2 not known
3) $n_s \geq 30$ } Case D

$$\bar{x} = \frac{\sum x}{n_s} = \frac{525}{50} = 10.5$$

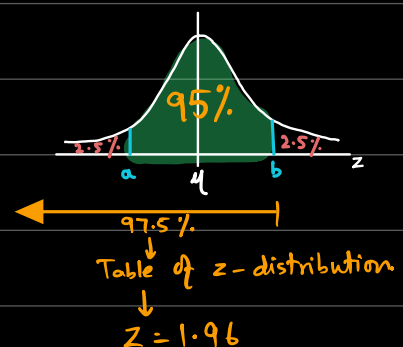
σ^2 not known use estimate:

$$\begin{aligned}\hat{\sigma}^2 &= \frac{n_s}{n_s - 1} \left(\frac{\sum x^2}{n_s} - (\bar{x})^2 \right) \\ &= \frac{50}{50 - 1} \left(\frac{5625}{50} - (10.5)^2 \right)\end{aligned}$$

$$\hat{\sigma}^2 = 2.2959 \dots$$

$$\hat{\sigma} = 1.515 \dots$$

$$\begin{aligned}\text{C.I.} &= \bar{x} \pm z \frac{\hat{\sigma}}{\sqrt{n_s}} \\ &= 10.5 \pm \frac{(1.96)(1.515 \dots)}{\sqrt{50}}\end{aligned}$$



(10.08 , 10.92)

12 — 23