Sampling Distributions

* Statistical methods are used to stay a process by analyzing the data

Population: Population is the set or collection of objects

* Population consists of set of objects, measurements or observations which are of interest

Size: size of the population is denoted by 'N' which represents the number of objects or observations in the population.

* Population is said to be finite or infinite depending on the 'N' finite or infinite.

Sample: Sample is a finite subset of the population.

* The size of the sample is denoted by 'n'
* Sampling is the process of drawing the samples from a given population.

large Sampling: If $n \geq 30$, the sampling (drawing samples) is said to be large sampling and the samples are called large samples.

Small sampling: If $n < 30$, the sampling (drawing samples) is said to be small samplings and the samples are called small size samples

* Statistical inference or inductive statistics deals with the predictions about the population using the information contained in the samples

Parameters: Statistical measures obtained from the population are known as population parameters

Examples: population mean(μ), Population variance(σ), population proportion(p).

statistics: Statistical measures of samples is called as statistics.

Examples: Sample mean($\bar{x}$), Sample variance(s^2), Sample proportions.

Thus parameters refer to population while statistics refer to sample.

* Samples are representatives of the population.

Random Samping: Random sampling is one in which each member of the population has equal chances.

Sampling with replacement: Each member of the population may be chosen more than once, since the member is replaced in the population.

The number of possible ways of choosing the samples of size 'n' from the population of size 'N' is N^n.

let $N = \{a, b, c, d\}$

$n = 2$ choosing two with replacement:

Number of ways $= 4^2 = 16$

$\{(a,a)\ (a,b)\ (a,c)\ (a,d)$
$(b,a)\ (b,b)\ (b,c)\ (b,d)$
$(c,a)\ (c,b)\ (c,c)\ (c,d)$
$(d,a)\ (d,b)\ (d,c)\ (d,d)\}$

Sampling without replacement: An element of the population cannot be chosen more than once, as it is not replaced.

Number of possible ways - $NC_n = \frac{N!}{(N-n)!\,n!}$

Example: let $N = \{a, b, c, d\}$

$n=2$ choosing the samples size two

$$NC_n = 4C_2 = \frac{4!}{(4-2)!\,2!} = 6.$$

Samples = $\{(a,b), (a,c), (a,d), (b,c), (b,d), (c,d)\}$

* Sample mean and sample varience are two important statistics which are measures of a random sample $x_1, x_2, x_3 \cdots x_n$ of size 'n'

$$\text{Sample mean} - \bar{X} = \frac{x_1 + x_2 + x_3 \cdots x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

* Sample mean is called measure of central tendency.

Sample variance:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{n\sum_{i=1}^{n} x_i^2 - \left[\sum_{i=1}^{n} x_i\right]^2}{n}$$

Sample standard deviation

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

Sampling Distribution

A finite population consisting the elements $\{2,3,4,5\}$ samples of size 2 are drawn with replacement then find

(i) Mean of the population

(ii) Varience of the population.

(iii) Sampling distributions of means

(iv) Mean of the sample distribution of means

(v) S.D of sampling Distribution of means

(vi) Verify 4 & 5 with the formula.

Population = $\{2,3,4,5\} \Rightarrow N=4$

Sample of size '2' $\Rightarrow n=2$ (with replacement)

no. of samples $= 4^2 = 16$

Samples are
$$\left\{\begin{array}{cccc} (2,2) & (2,3) & (2,4) & (2,5) \\ (3,2) & (3,3) & (3,4) & (3,5) \\ (4,2) & (4,3) & (4,4) & (4,5) \\ (5,2) & (5,3) & (5,4) & (5,5) \end{array}\right\}$$

(i) Mean of the population $(\mu) = \dfrac{2+3+4+5}{4}$

$= 7/2 = 3.5$

$\boxed{\therefore \mu = 3.5}$

(ii) Variance $(\sigma^2) = \frac{\sum (x_i - \mu)^2}{4}$

$$= \frac{(2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2}{4}$$

$\sigma^2 = 1.25$

Population standard distribution$(\sigma) = \sqrt{1.25}$

$= 1.118$.

(iv) Mean of the sample distribution of means

means of the samples are 2 2.5 3 3.5

2.5 3 3.5 4

3 3.5 4 4.5

3.5 4 4.5 5

x	2	2.5	3	3.5	4	4.5	5
f	1	2	3	4	3	2	1

Sampling distribution of means

$\frac{\sum fx}{f}$

mean of sample distribution of means $(\mu_{\bar{x}})$

$$\mu_{\bar{x}} = \frac{2+5+9+14+12+90+5}{16} = \frac{56}{16} = 3.5$$

iv) S.D sample distribution of means.

$\mu_{\bar{x}} = \mu$ i.e mean of the sampling distribution of means is always equal to mean of the population distribution.

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum(\bar{x}_i - \mu_{\bar{x}})^2 f}{16}}$$

$$\sigma_{\bar{x}} = \sqrt{\frac{(2-3.5)^2\times 1 + (2.5-3.5)^2\times 2 + (3-3.5)^2(3) + (4-3.5)^2\times 3 + (4.5-3.5)^2\times 2 + (5-3.5)^2\times 1}{16}}$$

$$= \frac{1}{4}\sqrt{10} = 0.79$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.118}{\sqrt{2}} = 0.79$$

* Solve the above question by considering without replacement case

$\{2,3,4,5\}$

${}^N C_n = {}^4C_2 = 6$

Samples $\{(2,3)\ (2,4)\ (2,5)\ (3,4)\ (3,5)\ (4,5)\}$

Sample means ($\bar{x}$) : 2.5, 3, 3.5, 3.5, 4, 4.5

Sampling distribution of means

$\bar{x}$	2.5	3	3.5	4	4.5
f	1	1	2	1	1

$$\mu_{\bar{x}} = \frac{\sum \bar{x} f}{\sum f} = \frac{(2.5)1 + (3\times1) + (3.5)2 + 4(1) + (4.5)(1)}{6} = 3.5$$

$$\sigma_{\bar{x}} = \sqrt{\frac{\sum (x_i - \mu_{\bar{x}})^2 f}{n}}$$

$$= \sqrt{\frac{(2.5-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (4.5-3.5)^2}{6}}$$

$$\sigma_{\bar{x}} = \sqrt{\frac{1 + (0.5)^2 + (0.5)^2 + (1)}{6}} = \sqrt{5/12} = 0.645$$

$\boxed{\mu_{\bar{x}} = \mu}$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \left[\sqrt{\frac{N-n}{N-1}}\right]$$

$$= \frac{1.118}{\sqrt{2}} \left(\sqrt{\frac{2}{3}}\right) = 0.645$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \left(\sqrt{\frac{N-n}{N-1}}\right)$$

1. Determine the mean and standard deviation of sampling distribution of means of 300 random samples each of size 36 are drawn from population of 1500 which is normally distributed with mean 22.4 and standard deviation 0.048 if the sampling is done

(i) with replacement

(ii) without replacement

(iii) Between 22.39 and 22.41

(iv) > 22.42

(v) < 22.37

(vi) < 22.38 and > 22.41

$n = 36$, $N = 1500$

$\mu = 22.4$, $\sigma = 0.048$

$\mu_{\bar{x}} = ?$ $\sigma_{\bar{x}} = ?$

① with replacement

Since $\mu_{\bar{x}} = \mu = 22.4$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.048}{\sqrt{36}} = 0.008$$

② without replacement

$\mu_{\bar{x}} = \mu = 22.4$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$= \frac{0.048}{\sqrt{36}} \sqrt{\frac{1500-36}{1500-1}} = 0.0079$$

(iii)

$P(22.39 < x < 22.41)$ with $x_1 = 22.39$, $x_2 = 22.41$

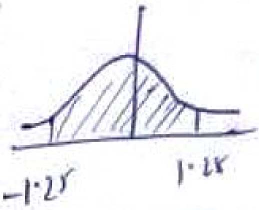
$$z_1 = \frac{x_1 - \mu}{\sigma/\sqrt{n}} = \frac{22.39 - 22.4}{0.048/\sqrt{36}} = -1.25$$

$$z_2 = \frac{x_2 - \mu}{\sigma/\sqrt{n}} = \frac{22.41 - 22.4}{0.048/\sqrt{36}} = 1.25$$

$$P(22.39 < x < 22.4) = P(-1.25 < z < 1.25)$$

$$= 2(0.3944)$$

$$= 0.7888$$

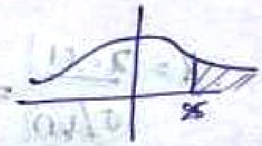


∴ no. of samples having mean b/w 22.39 & 22.41

$$= 300 \times 0.7888$$

$$= 236.64$$

(iv) > 22.42



$P(22.42 < x < \infty)$

$$z_1 = \frac{x_1 - \mu}{\sigma/\sqrt{n}} = \frac{22.42 - 22.4}{0.048/\sqrt{36}} = \frac{20}{8} = 2.5$$

$P(z > 2.5) = 0.5 - P(z = 0 \text{ to } z = 2.5)$

$= 0.5 - 0.4938 = 0.0062$

no. of samples $= 300 \times 0.0062$

$= 1.86$

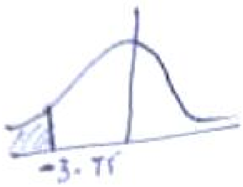
$= 2$ samples

5. Less than 22.37

$P(\bar{x} < 22.37)$

$z = \frac{x - \mu}{\sigma/\sqrt{n}} = \frac{22.37 - 22.4}{0.048/\sqrt{36}} = -3.75$

$P(\bar{x} < 22.37) = P(z < -3.75)$



Area $= 0.5 - \begin{pmatrix} z = 0 \text{ to} \\ z = 3.75 \end{pmatrix}$

$= 0.5 - 4999$

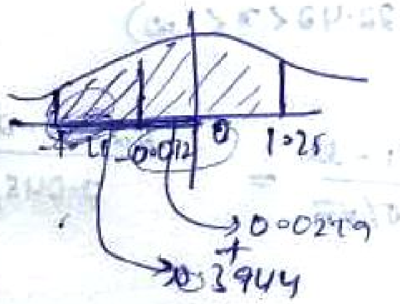
$= 0.0001$

$\therefore$ $300 \times 0.0001 =$ zero samples

(vi) < 22.38 > 22.4

$z = \frac{x - \mu}{\sigma/\sqrt{n}} = \frac{22.38 - 22.4}{0.048/\sqrt{36}} = -0.0721$

$z = \frac{x - \mu}{\sigma/\sqrt{n}} = \frac{22.4 - 22.41}{0.048/\sqrt{36}} = 1.25$



0.0279

0.0279

$P(\bar{x} < 2.5)$ 0.0279×300 $\underline{0.3665}$

$+ P(z > 1.25)$

$1 - (0.4938 + 0.3944)$

$= 0.1118$

$0.1118 \times 300 = 33.54$

Determine the probability that the sample mean area covered by a sample of 40 litre paint boxes will be between 510 to 520 sq.ft given that a 1 litre of such paint box covers on the avarage 513.3 sq.ft with a standard deviation of 31.5 sqft.

$\mu = 513.3$ sqf $\sigma = 31.5$

$P(510 < x < 520)$ $n = 40$

x_1 x_2

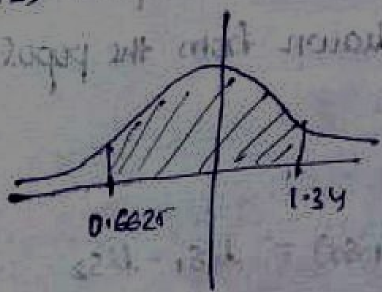
$$z = \frac{x-\mu}{\sigma/\sqrt{n}} = \frac{510-513.3}{31.5/\sqrt{40}} = -0.6625$$

$$z = \frac{x_2-\mu}{\sigma/\sqrt{n}} = \frac{520-513.3}{31.5/\sqrt{40}} = 1.3452.$$

$P(-0.6625 < x < 1.3452)$

$= 0.2454 + 0.4099$

$= 0.6553$



no of samples brought

$0.6553 \times$ sample population

Sampling distribution of differences and sums.

Let μ_{S_1} and σ_{S_1} be the mean and standard deviation of a statistic S_1 obtain by computing S_1 for all possible samples of size n_1 drawn from the population A.

Similarly μ_{S_2} and σ_{S_2} be the mean and standard deviation of sampling distribution of statistic S_2 for all samples of size n_2 drawn from the population B.

$(S_1 - S_2)$ represents the difference of the statistic for all possible samples drawn from the population A and B.

$$\mu_{(S_1 - S_2)} = \mu_{S_1} - \mu_{S_2}$$

$$\sigma_{(S_1 - S_2)} = \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2}$$

$$\mu_{S_1 + S_2} = \mu_{S_1} + \mu_{S_2}$$

$$\sigma_{(S_1 + S_2)} = \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2}$$

If $U_1 = \{2, 7, 9\}$ and $U_2 = \{3, 8\}$

find μ_{U_1}, μ_{U_2}, σ_{U_1}, σ_{U_2}, $\mu_{U_1 - U_2}$, $\mu_{U_1 + U_2}$, $\sigma_{U_1 - U_2}$, $\sigma_{U_1 + U_2}$

$$\mu_{U_1} = \frac{2+7+9}{3} = 18/3 = 6$$

$$\mu_{U_2} = \frac{3+8}{2} = 11/2 = 5.5$$

$$\sigma_{U_1} = \sqrt{\frac{\sum (x_i - \mu)}{3}} = \sqrt{\frac{(-4)^2 + (1)^2 + (3)^2}{3}}$$

$$= \sqrt{\frac{16+1+9}{3}} = \sqrt{\tfrac{26}{3}}$$

$= 2.94$

$$\sigma_{U_2} = \sqrt{\frac{(2.5)^2 + (2.5)^2}{2}}$$

$$\sigma_{U_2} = 2.5$$

$$\mu_{S_1 - S_2} = \mu_{S_1} - \mu_{S_2} = 0.5$$

$$\mu_{S_1 + S_2} = \mu_{S_1} + \mu_{S_2} = 11.5$$

$$\sigma_{S_1 - S_2} = \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2}$$

$$= 3.861$$

$\mu_{U_1 + U_2} = \{2+3, 2+8, 7+3, 7+8, 9+3, 9+8\}$

$\{ \ldots 17\}$

$\mu_{U_1-U_2} = \{-1, -6, 4, -1, 6, 1\}$

$\mu_{U_1-U_2} = \frac{-1-6+4-1+6+1}{6} = 3/6$

$= 0.5$

$\sigma_{U_1+U_2} = \sqrt{\frac{(5-11.5)^2 + 2(10-11.5)^2 + (11-11.5)^2 + (12-11.5)^2 + (17-11.5)^2}{6}} = 3.861$

$\sigma_{U_1} = \sqrt{\frac{\Sigma(x_i - \mu_{U_1})}{3}} = \sqrt{\frac{(2-6)^2 + (7-6)^2 + (9-6)^2}{3}}$

$= 2.94$

$\sigma_{U_2} = \sqrt{\frac{\Sigma x_i - \mu_{U_2}}{2}} = \sqrt{\frac{(3-5.5)^2 + (8-5.5)^2}{2}}$

$= 2.5$

$\sigma_{U_1-U_2} = \sqrt{\frac{2(-1-0.5)^2 + (-6-0.5)^2 + (4-0.5)^2 + (6-0.5)^2 + (1-0.5)^2}{6}}$

$\sigma_{U_1-U_2} = 3.861$

* The mean voltage of a battery is 15 volts with a s.d of 0.2 V find the probability that 4 of such batteries connected in series will have a combained voltage of 60.8 volts or more

Let A, B, C, D are batteries

$\mu_{x_A} = \mu_{x_B} = \mu_{x_C} = \mu_{x_D} = 15V$

$\sigma_A = \sigma_B = \sigma_C = \sigma_D = 0.2V$

$\mu_{x_A+x_B+x_C+x_D} = 60V$

$\sigma_{A+B+C+D} = \sqrt{\sigma_A^2 + \sigma_B^2 + \sigma_C^2 + \sigma_D^2}$

$= \sqrt{4(0.2)^2}$

$= 0.2 \times 2 = 0.4V$

$P(X > 60.8)$

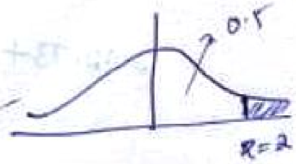
$z = \frac{\bar{x} - \mu_{x_A+x_B+x_C+x_D}}{\sigma_{A+B+C}/\sqrt{n}}$

$= \frac{60.8-60}{0.4/\sqrt{1}} = \frac{0.8}{0.4}$

$= 2$

$P(X > 60.8) = P(z > 2)$

$= 0.5 - 0.4772 = 0.0228$



Statistical estimation:

Statistical estimation is a part of statistical inference where the population parameter is estimated from the corresponding sample statistics

Estimations are two types, point estimation and interval estimation

and

$n = 40$ $\bar{x} = 12.73$ $\sigma = 2.06$

(i) $\bar{x}$ 12.73 99% of confidence

$LOS = 0.01 = \alpha$

0.005

2.575

$E = z_{\alpha/2} \cdot \sigma/\sqrt{n}$

$= 2.575 \cdot 2.06/\sqrt{40}$

(ii) 2% LOS. $= 0.02 =$

$\bar{x} - z_{\alpha/2} \cdot \sigma/\sqrt{n} < \mu < \bar{x} + z_{\alpha/2} \sigma/\sqrt{n}$

$\alpha/2 = 0.01$

$z = 0.49$

$z_{\alpha/2} = 2.33$

$12.73 - 2.33 \cdot \frac{2.06}{\sqrt{40}} < \mu$

$< 12.73 + 2.33 \cdot \frac{2.06}{\sqrt{40}}$

$12.73 - 0.759 < \mu <$

$12.73 + 0.759$

$11.971 < \mu < 13.489$

'mc limit for μ

at 98% confidence

$(11.971 < \mu < 13.489)$

$E = z_{\alpha/2} \cdot \sigma/\sqrt{n}$

$\frac{1}{2} = z_{\alpha/2} \cdot 2.06/\sqrt{40}$

$\frac{1}{2} \times \frac{\sqrt{40}}{2.06} = z_{\alpha/2}$

$z_{\alpha/2} = 1.5350$

accepted region

$= 0.485$

$2(0.437)$