4. Sampling Distribution

Suppose we have different samples of size n drawn from the population and for each and every sample of size n compute the means. Obiously these means will not be same. If we group these charecteristics according to its frequency the frequency distribution so generated is called sampling distribution of means.

Note: 1) The sampling distribution of the large samples is assumed to be normal distribution.

2) The standard diviation of sampling distribution is also called standard error.

Case 1 : Sampling with replacement.

Here the items drawn are put back to the population before the next draw. If N is the size of the population and n is size of the sample then we have N^n number of samples. And $\mu_{\bar{x}}$ of the frequency distribution of the sample mean will be equal to the population mean μ. ($\mu_{\bar{x}} = \mu$)

The variance $\sigma_{\bar{x}}^2$ of the frequency distribution of the sample means will be equal to σ^2/n where σ^2 is variace of population

i.e $\sigma_{\bar{x}}^2 = \dfrac{\sigma^2}{n}$.

Case 2: Sampling without replacement.

Here the items are drawn and are not put back to the population before the next draw. In this case $^N C_n$ samples can be collected

Mean : $\mu_{\bar{x}} = \mu$

Variance : $\sigma_{\bar{x}}^2 = \left(\dfrac{N-n}{N-1}\right)\dfrac{\sigma^2}{n}$

Correcting factor $\left(\dfrac{N-n}{N-1}\right) \approx 1$ for large samples (with ~~replacement~~)

① A population consists of 5 numbers 2, 3, 6, 8, 11. all possible samples of size 2 which can be drawn with replacement from this population. F

(i) Find mean and standard deviation of population

(ii) The mean and standard deviation of sampling distribution of means.

(iii) Considering samples without replacements find mean and standard deviation of sampling distribution of means.

Sol^n : $N = 5$

$n = 2$

(i) population mean $(\mu) = 6$ $\left[2+3+6+8+11/5 = 6\right]$

$S.D\ (\sigma) = 3.28$ $\left[\sqrt{\dfrac{\Sigma (x_i - \mu)^2}{n}}\right]$

(ii) Sample with replacement are

(2,2) (2,3) (2,6) (2,8) (2,11)

(3,2) (3,3) (3,6) (3,8) (3,11)

(6,2) (6,3) (6,6) (6,8) (6,11)

(8,2) (8,3) (8,6) (8,8) (8,11)

(11,2) (11,3) (11,6) (11,8) (11,11)

Now individual means of each samples

2 2.5 4 5 6.5

2.5 3 4.5 5.5 7

4 4.5 6 7 8.5

5 5.5 7 8 9.5

6.5 7 8.5 9.5 11

$\bar{x}$:	2	2.5	3	4	4.5	5	5.5	6	6.5	7	8	8.5	9.
f :	2	2	1	2	2	2	2	1	2	4	1	2	2

$$\mu_{\bar{x}} = \frac{\Sigma \bar{x} f}{\Sigma f} = 6$$

standard deviation $= \sigma_{\bar{x}} = 2.32$

$\left\{ \sigma_{\bar{x}}^2 = \frac{\sigma^2}{2} \Rightarrow 5.38 = 5.38 \right\}$

iii) without replacement samples are

$^{N}C_{n} = {}^{5}C_{2} = 10$ samples

(2, 3) (2, 6) (2, 8) (2, 11) (3, 6) (3, 8) (3, 11) (6, 8) (6, 11) (8, 11)

Now individual means are

2.5 4 5 6.5 4.5 5.5 7 7 8.5 9.5

$\mu_{\bar{x}} = \mu = 6$

$$\sigma^2_{\bar{x}} = \left(\frac{N-n}{N-1}\right)\frac{\sigma^2}{n} = \left(\frac{5-2}{5-1}\right)\left(\frac{10.76}{2}\right) = 4.035$$

③ Certain tubes manufactured by a company have mean life time of 800 hrs and standard deviation 60 hrs. Find the probability that a sample of 16 tubes taken from the group will have a mean lifetime

(i) b/w 790 hrs to 810 hrs

(ii) Less than 785 hrs

(iii) more than 820 hrs

Given : for population $\mu = 800$ $\sigma = 60$ $n = 16$

converting population data to sample

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$$

$\left\{ \begin{array}{l} \mu_{\bar{x}} = \mu \\ \sigma^2_{\bar{x}} = \sigma^2/n \end{array} \right.$

$$= \frac{x - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$

$$z = \frac{\bar{x} - 800}{60/4} = \frac{\bar{x} - 800}{15}$$

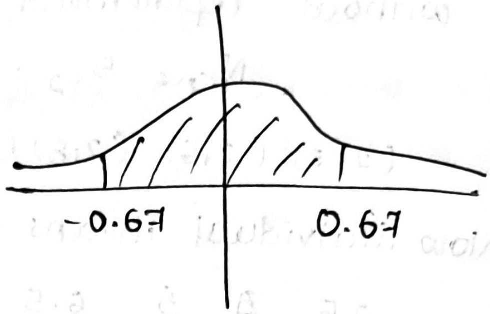
(i) $P(790 < \bar{x} < 810)$

$= P\left(\frac{790-800}{15} < \frac{\bar{x}-800}{15} < \frac{810-800}{15}\right)$

$= P(-0.67 < z < 0.67)$

$= 2A(0.67)$

$= 0.497$



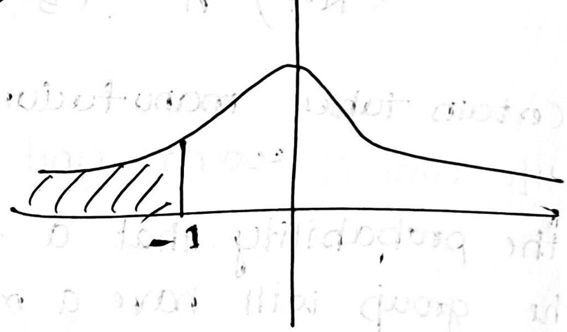
(ii) $P(\bar{x} < 785)$

$= P\left(\frac{\bar{x}-800}{15} < \frac{785-800}{15}\right)$

$= P(z < -1)$

$= 1/2 - A(-1)$

$= 0.158$



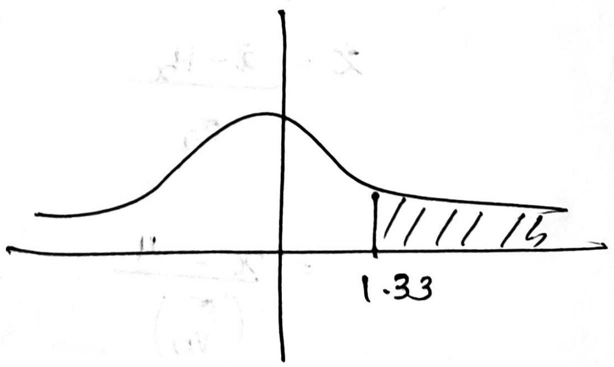
(iii) $P(\bar{x} > 820)$

$= P\left(\frac{\bar{x}-800}{15} > \frac{820-800}{15}\right)$

$= P(z > 1.33)$

$= 1/2 - A(1.33)$

$= 0.091$



(c) If the mean population is 575 with standard deviation 8.3. How large a sample must be in order that there will be one chance in 100 that the mean sample is less than 572.

Population : $\mu = 575$ $\quad \sigma = 8.3$ $\quad n = ?$

$\bar{x} < 572$

For sample

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} \qquad \left\{ \begin{array}{l} \mu_{\bar{x}} = \mu \\ \sigma_{\bar{x}} = \sigma/\sqrt{n} \end{array} \right.$$

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{x} - 575}{8.3/\sqrt{n}}$$

For $\bar{x} = 572$

$$z = \frac{572 - 575}{8.3/\sqrt{n}} = -0.36(\sqrt{n})$$

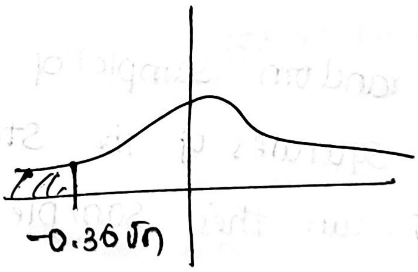
Given, $P(\bar{x} < 572) = 0.01$

$P(z < -0.36\sqrt{n}) = 0.01$

$1/2 - A(0.36\sqrt{n}) = 0.01$

$0.49 = A(0.36\sqrt{n})$

$A(2.35) = A(0.36\sqrt{n})$



By data $A(2.35) = 0.49$

$2.35 = 0.36\sqrt{n}$

$n = 42.6$

$\boxed{n = 43}$

* Test of significance of small samples 26/07/23

Student t-distribution (t-Test)

1) t-test of significance of mean of a random sample

Under the null hypothesis that there is no significance difference between the sample mean $\bar{x}$ and population mean μ.

The test statistic is given by

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

where $S^2 = \frac{\Sigma (x-\bar{x})^2}{n-1}$

With degree of freedom n-1 (D.f)

Here the formula is used when the population standard deviation is not known

The confidence limit of population mean is given by

$$\bar{x} - t_\alpha (S/\sqrt{n}) < \mu < \bar{x} + t_\alpha (S/\sqrt{n})$$

1) A random sample of size 16 has 53 as mean. The sum of the squrares of the standard deviation from mean is 135. Can this sample be regarded as taken from the population having 56 as the mean. Also obtain 95% and 99% confidence limit of the mean of the population.

$n = 16$

$\bar{x} = 53$

$\Sigma (x-\bar{x})^2 = 135$

$\mu = 56$

Null hypothesis $H_0: \mu = 56$

alternate hypothesis $H_1: \mu \neq 56$ / two tailed test)

test statistics

t or $z = \dfrac{\bar{x} - \mu}{(s/\sqrt{n})}$

$$s^2 = \frac{\Sigma(x - \bar{x})^2}{n-1} = \frac{135}{15} = 9$$

$$\boxed{s = 3}$$

$$t \text{ or } z = \left(\frac{53 - 56}{3/\sqrt{16}}\right)$$

$z = -4$

$|z| = |-4| = 4 > 2.13$ at 5% level of significance

and n-1 i.e 15 as degree to freedom

(check under 0.05 at 15 DoF freedom in given table)

$|z| = 4 > 2.13$ so H_0 is rejected. at 5% level of significance.

~~$|z| = 4 > 2.95$ at 1% level of significance and 15 as degree of freedom.~~

~~so H0 is reject at 1% significant level~~

The confidence limit at 95%

$$\bar{x} - t_\alpha(s/\sqrt{n}) < \mu < \bar{x} + t_\alpha(s/\sqrt{n})$$

at 95% : ie $\bar{x} \pm t_{0.05}(s/\sqrt{n})$

$53 \pm 2.13\left(3/4\right)$

53 ± 1.5975

$\Rightarrow [54.59 \quad (51.25) \quad 51.4025)$

at 99% : $\bar{x} \pm t_{0.01}(s/\sqrt{n})$

$\Rightarrow \quad 53 \pm 2.95\,(3/4)$

$\Rightarrow \quad 53 \pm 2.2125$

$[55.2125 \quad \cancel{2.2125} \quad 50.7875]$

2) 9 items of the sample have following values 45, 47, 50, 52, 48, 47, 49, 53, 51. Does the mean of these value differ significantly from assumed mean of 47.5.

$n = 9 \quad \mu = 47.5 \quad d.f = 9-1 = 8 \quad \bar{x} = 49.11$ (calculator)

$H_0 : \mu = 47.5$

$H_1 : \mu \neq 47.5$ (two tailed test)

test - statistics

$$t \text{ or } z = \frac{\bar{x}-\mu}{(s/\sqrt{n})}$$

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

x	45	47	50	52	48	47	49	53	51
$(x-\bar{x})$	-4.1	-2.1	0.9	2.9	-1.1	-2.1	-0.1	3.9	1.9
$(x-\bar{x})^2$	16.81	4.41	0.81	8.41	1.21	4.41	0.01	15.21	3.61

$\Sigma(x-\bar{x})^2 = 54.89$

$$s^2 = \frac{54.89}{8} = 6.861$$

$\boxed{s = 2.619}$

assumed mean is nothing but population mean

$$t \text{ or } z = \frac{\bar{x}-\mu}{(s/\sqrt{n})} = \frac{49.1-47.5}{(2.619/\sqrt{8})}$$

$z = 1.832 < 2.31$ at 5% level of significance at d.f 8 as degree of freedom.

$\therefore$ Ho is accepted.

3) A sample of 20 items has mean 42 units and standard deviation 5 units. Test the hypothesis that it is random sample from the normal population with mean 45 units.

$n = 20$ ~~S=5~~   S.D = 5 $H_0 : \mu = 45$

$\bar{x} = 42$ D.of.f = 19 $H_1 : \mu \neq 45$ (two tailed test)

$\mu = 45$

t- test

$$t = \frac{\bar{x}-\mu}{s/\sqrt{n-1}}$$ → (Here in formula we have to take instead of 'S' we take 's' because S.D is mentioned)

$$= \frac{42-45}{5/\sqrt{20-1}}$$

$t = -2.61$

$|t| = 2.61 \not< 2.09$ at 5% level of significance as 19 as degree of freedom.

Ho is rejected.

* t-test for diff of means of two samples (Small)

This test is used to check whether the two samples $x_1, x_2 \ldots x_{n_1}$ and $y_1, y_2 \ldots y_{n_2}$ which are of the sizes n_1 and n_2 respectively. and are drawn from two normal population with mean μ_1 and μ_2 under the assumption that population variance are equal (i-e $\sigma_1^2 = \sigma_2^2 = \sigma$).

Under the null hypothesis that the samples are drawn from the normal population with mean μ_2 and μ_2

$$t = \frac{\bar{x} - \bar{y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where $\bar{x}$ and $\bar{y}$ are means of the samples and degree of freedom is $n_1 + n_2 - 2$.

Note: 1) If the two samples standard deviation s_1 and s_2 are given $S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$

2) If S_1 and S_2 are not given

$$S^2 = \frac{\Sigma (x - \bar{x})^2 + \Sigma (y - \bar{y})^2}{n_1 + n_2 - 2}$$

1) Two samples of sodium vapour bulbs were tested for length of life and the following results were obtained.

	Size	Sample mean	Sample S.D
Type 1	8	1234 hrs	36 hrs
Type 2	7	1036 hrs	40 hrs

is the difference in the mean significant to generelise that type 1 is supirior to type 2. regarding length of life.

$n_1 = 8$ $n_2 = 7$

$H_0: \mu_1 = \mu_2$

$\bar{x} = 1234$ $\bar{y} = 1036$

$H_1: \mu_1 > \mu_2$ (one tailed test)

$S_1 = 36$ $S_2 = 40$

$$t = \frac{\bar{x} - \bar{y}}{S\sqrt{1/n_1 + 1/n_2}}$$

$$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$$

$S = 40.73$

$t = 9.392$

$t = 2 \times 9.39$

$t = 18.79$ (Here we are considering single tailed test and formula is on two tailed test so we have to double it)

$t = 18.79 > 1.77$ at 2α i.e $2 \times 0.05 = 0.1$ significance level at df 13.

$\therefore$ Ho is rejected.

2) Two horses A and B were tested according to time to run a particular race with the following results

Horse A	28	30	32	33	33	29	34
Horse B	29	30	30	24	27	29	

Test whether you can discriminate b/w two horses.

$n_1 = 7$ $\quad \bar{x} = 31.28$

$n_2 = 6$ $\quad \bar{y} = 28.16$

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$ (two tailed test)

$$t = \frac{\bar{x} - \bar{y}}{S\sqrt{1/n_1 + 1/n_2}}$$

$$S^2 = \frac{\Sigma (x - \bar{x})^2 + \Sigma (y - \bar{y})^2}{n_1 + n_2 - 2}$$

x	28	30	32	33	33	29	34
$(x-\bar{x})$	-3.28	-1.28	0.72	1.72	1.72	-2.28	2.72
$(x-\bar{x})^2$	10.75	1.63	0.51	2.95	2.95	5.19	7.39

$$\Sigma (x-\bar{x})^2 = 31.37$$

y	29	30	30	24	27	29
$(y-\bar{y})$	0.84	1.84	1.84	-4.16	-1.16	0.84
$(y-\bar{y})^2$	0.70	3.38	3.38	17.3	1.34	0.70

$$\Sigma (y-\bar{y})^2 = 26.8$$

$$s^2 = \frac{\Sigma(x-\bar{x})^2 + \Sigma(y-\bar{y})^2}{n_1+n_2-2}$$

$$= \frac{31.37 + 26.8}{11}$$

$$s = 2.29$$

$$t = \frac{\bar{x}-\bar{y}}{s\sqrt{1/n_1 + 1/n_2}} = \frac{31.28 - 28.16}{2.29\sqrt{1/7+1/6}}$$

$t = 2.44 > 2.20$ at 5% significance level of degree of freedom 11

$\therefore$ Ho is rejected.