$$z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} \tag{4}$$

Example 1 *A population consists of the four numbers 3, 7, 11, 15. Consider all possible samples of size 2 which can be drawn from this population without replacement. Find the mean and the standard deviation in the population, and in the sampling distribution of means. Verify the formullas* (1) *and* (2).

▸ Here, population consists of $4\,(= N_p)$ numbers 3, 7, 11, 15. Therefore, the *mean of the population* is**

$$\mu = \frac{1}{4}(3 + 7 + 11 + 15) = 9 \tag{i}$$

and the *variance of the population* is***

$$\begin{aligned}\sigma^2 &= \frac{1}{4}\left\{(3-9)^2 + (7-9)^2 + (11-9)^2 + (15-9)^2\right\} \\ &= \frac{1}{4}(36 + 4 + 4 + 36) = 20.\end{aligned} \tag{ii}$$

Therefore, the *standard deviation of the population* is

$$\sigma = \sqrt{20}. \tag{iii}$$

We note that the possible samples of size two which can be drawn without replacement from the given population are

$$(3,7),\ (3,11),\ (3,15),\ (7,11),\ (7,15),\ (11,15).$$

The means of these 6 samples are (respectively)

$$5,\ 7,\ 9,\ 9,\ 11,\ 13.$$

These are the items in the sampling distribution of means without replacement. For this (raw) distribution, the mean and variance are given by (respectively)

$$\mu_{\overline{X}} = \frac{1}{6}(5 + 7 + 9 + 9 + 11 + 13) = 9, \tag{iv}$$

---

*Recall the expression (3) of Section 7.7.

**See Formula (1) of Section 8.1.

***See Formula (3) of Section 8.1.

and $$\sigma_{\overline{X}}^2 = \frac{1}{6}\left\{(5-9)^2 + (7-9)^2 + (9-9)^2 + (9-9)^2 + (11-9)^2 + (13-9)^2\right\}$$
$$= \frac{1}{6}(16+4+0+0+4+16) = \frac{20}{3}. \tag{v}$$

Therefore, the standard deviation is

$$\sigma_{\overline{X}} = \sqrt{20/3} = 2.582. \tag{vi}$$

From expressions (i) and (iv) we note that

$$\mu = 9 = \mu_{\overline{X}}.$$

From expressions (iii) and (vi), we find that

$$\frac{\sigma}{\sqrt{N}} \cdot \frac{\sqrt{N_p - N}}{\sqrt{N_p - 1}} = \frac{\sqrt{20}}{\sqrt{2}} \cdot \frac{\sqrt{4-2}}{\sqrt{4-1}} = \frac{\sqrt{20}}{\sqrt{3}} = \sigma_{\overline{X}}.$$

Thus, formulas (1) and (2) are verified. ∎

Example 2 *A population consists of the four numbers* 3, 7, 11, 15. *Consider all possible samples of size* 2 *which can be drawn with replacement from this population. Find* (i) *The population mean and population standard deviation, and* (ii) *The mean and standard deviation of the sampling distribution of means.*

Verify the formulas (1) *and* (3).

▶ Here, the given population is the same as that in Example 1. As found in that Example, the population mean and the population standard deviation are, respectively,

$$\mu = 9, \quad \sigma = \sqrt{20}. \tag{i}$$

The possible samples of size $N = 2$ which can be drawn *with replacement* from the population are

$$\begin{array}{llll} (3,3), & (3,7), & (3,11), & (3,15), \\ (7,3), & (7,7), & (7,11), & (7,15), \\ (11,3), & (11,7), & (11,11), & (11,15), \\ (15,3), & (15,7), & (15,11), & (15,15), \end{array}$$

which are 16 in number. The means of these samples are (respectively)

3, 5, 7, 9,
5, 7, 9, 11,
7, 9, 11, 13,
9, 11, 13, 15.

The frequency distribution of these means is given by the following Table:

Mean (x_i) :	3	5	7	9	11	13	15
Frequency (f_i) :	1	2	3	4	3	2	1

The above frequency distribution is the sampling distribution of means for the given population when the sampling is done with replacement. For this distribution, the mean and variance are given by*

$$\mu_{\overline{X}} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1}{16}(3 + 10 + 21 + 36 + 33 + 26 + 15) = 9, \tag{ii}$$

$$\sigma^2_{\overline{X}} = \frac{\Sigma f_i (x_i - \mu_{\overline{X}})^2}{\Sigma f_i} = \frac{1}{16}\{1 \times (3-9)^2 + 2 \times (5-9)^2 + 3 \times (7-9)^2 + 4 \times (9-9)^2$$
$$+3 \times (11-9)^2 + 2 \times (13-9)^2 + 1 \times (15-9)^2\}$$
$$= \frac{1}{16}(36 + 32 + 12 + 0 + 12 + 32 + 36) = 10 \tag{iii}$$

Therefore, the standard deviation is

$$\sigma_{\overline{X}} = \sqrt{10} \approx 3.162. \tag{iv}$$

Using expressions (ii) and (iv) above, we note that

$$\mu_{\overline{X}} = 9 = \mu \quad \text{and} \quad \sigma_{\overline{X}} = \sqrt{10} = \frac{\sqrt{20}}{\sqrt{2}} = \frac{\sigma}{\sqrt{N}}.$$

Thus, formulas (1) and (3) are verified. ■

Example 3 *The daily wages of* 3000 *workers in a factory are normally distributed with mean equal to Rs.* 68 *and standard deviation equal to Rs.* 3. *If* 80 *samples consisting of* 25 *workers each are obtained, what would be the mean and standard deviation of the sampling distribution of means if sampling were done* (a) *with replacement,* (b) *without replacement?*

In how many samples will the mean is likely to be (i) *between Rs.* 66.8 *and Rs.* 68.3, *and* (ii) *less than Rs.* 66.4?

*using formulas (2) and (4) of Section 8.1.

► Here, the number of items in the population is $N_p = 3000$ and sample size is $N = 25$. Also, the population mean is $\mu = 68$ and the population standard deviation is $\sigma = 3$.

In the case of sampling with replacement, the mean and standard deviation of the sampling distribution of means are given by (see formulas (1) and (3))

$$\mu_{\overline{X}} = \mu = 68 \quad \text{and} \quad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}} = \frac{3}{\sqrt{25}} = 0.6$$

If the sampling is done without replacement, the same quantities are given by (see formulas (1) and (2)).

$$\mu_{\overline{X}} = \mu = 68,$$

and $$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}}\sqrt{\frac{N_p - N}{N_p - 1}} = \frac{3}{\sqrt{25}}\sqrt{\frac{3000-25}{3000-1}} = \frac{3}{5} \times 0.996 = 0.5976 = 0.6.$$

Thus, $\mu_{\overline{X}}$ and $\sigma_{\overline{X}}$ have the same values in both cases.

Since the population (the daily wages of workers) is normally distributed, the sampling distribution of means is also taken as normally distributed. The mean and the standard deviation of this distribution are $\mu_{\overline{X}} = 68$ and $\sigma_{\overline{X}} = 0.6$ in both of the two cases considered above. Therefore, the standard normal variate associated with the sample mean $\overline{X}$ is (by formula (4))

$$z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{X} - 68}{0.6}$$

For $\overline{X} = 66.8$, we get $z = -2$; for $\overline{X} = 68.3$, we get $z = 0.5$, and for $\overline{X} = 66.4$, we get $z = -2.67$.

Therefore, the probability that a sample will have a mean between 66.8 and 68.3 is

$$P(66.8 < \overline{X} < 68.3) = P(-2 < z < 0.5) = P(0 < z < 2) + P(0 < z < 0.5)$$
$$= A(2) + A(0.5) = 0.4772 + 0.1915 = 0.6687,$$

on using the Normal Probability Table.*

Accordingly, in 80 samples, the expected number of samples having means between Rs. 66.8 and Rs. 68.3 is $0.6687 \times 80 = 53$.

Next, the probability that a sample will have a mean less than 66.4 is

$$P(\overline{X} < 66.4) = P(z < -2.67) = P(z > 2.67) = P(0 < z < \infty) - P(0 < z < 2.67)$$

*See Section 7.7.

$$= 0.5 - A(2.67) = 0.5 - 0.4962 = 0.0038.$$

Accordingly, in 80 samples, the expected number of samples having means less than Rs. 66.4 is $0.0038 \times 80 = 0.304 \approx 0$ (nearest non-negative integer). ■

Example 4 *The weights of 1500 ball bearings are normally distributed with a mean of 635 gms and a standard deviation of 1.36 gms. If 300 samples of size 36 are drawn from this populaiton, find the mean and standard deviation of the sampling distribution of means if the sampling is done* (i) *without replacement, and* (ii) *with replacement.*

In the second of the above cases, find how many samples will have their mean greater than 635.5 gms.

▸ Here, $N_p = 1500$, $\mu = 635$, $\sigma = 1.36$ and $N = 36$. In the case of sampling without replacement the mean and the standard deviation of the sampling distribution of means are given by (see formulas (1) and (2))

$$\mu_{\overline{X}} = \mu = 635$$

and
$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}}\frac{\sqrt{N_p - N}}{\sqrt{N_p - 1}} = \frac{1.36}{\sqrt{36}} \cdot \frac{\sqrt{1500 - 36}}{\sqrt{1500 - 1}} = (0.2267) \times \frac{38.2623}{38.717} = 0.224.$$

If the sampling is done with replacement, the above quantities are given by (see formulas (1) and (3))

$$\mu_{\overline{X}} = \mu = 635$$

and
$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}} = \frac{1.36}{6} = 0.2267.$$

Since the population (- the weights of ball bearings being considered) is normally distributed, the sampling distribution of means is also taken to be normally distributed. In the case of sampling with replacement, the mean and the standard deviation of the distribution of means have been found as $\mu_{\overline{X}} = 635$ and $\sigma_{\overline{X}} = 0.2267$. The standard normal variate associated with the sample mean $\overline{X}$ is

$$z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{X} - 635}{0.2267}.$$

For $\overline{X} = 635.5$, we get $z = 2.2056$. Therefore, the probability that a sample will have a mean greater than 635.5 gms is

$$P(\overline{X} > 635.5) = P(z > 2.2056) = P(0 < z < \infty) - P(0 < z < 2.2056)$$

8.3.2. Sampling Distribution of Proportions

It can be proved that for samples of large size, the sampling distribution of proportions is a normal distribution. Then the standard normal variate associated with the sample proportion is given by

$$z = \frac{\mathcal{P} - \mu_{\mathcal{P}}}{\sigma_{\mathcal{P}}} = \frac{\mathcal{P} - p}{\sqrt{pq/N}} \tag{3}$$

Example 1 *Find the probability that in 100 tosses of a fair coin between 45% and 55% of the outcomes are heads.*

▶ We treat 100 tosses of a fair coin as a sample of size $N = 100$ from the infinite population of all possible tosses of the coin.

Since the probability of getting a head in a toss is $p = 1/2$, the mean and the standard deviation for the distribution of the proportion $\mathcal{P}$ of sucess (getting head) in the given sample are (see formulas (1) and (2))

$$\mu_{\mathcal{P}} = p = \frac{1}{2} = 0.5,$$

$$\sigma_{\mathcal{P}} = \left[\frac{p(1-p)}{N}\right]^{1/2} = \left[\frac{(1/2)(1/2)}{100}\right]^{1/2} = 0.05$$

The corresponding standard normal variate is

$$z = \frac{\mathcal{P} - \mu_{\mathcal{P}}}{\sigma_{\mathcal{P}}} = \frac{\mathcal{P} - 0.5}{0.05}.$$

For $\mathcal{P} = 45\%$, we have $z = \dfrac{0.45 - 0.5}{0.05} = -1.$

For $\mathcal{P} = 55\%$, we have $z = \dfrac{0.55 - 0.5}{0.05} = 1.$

Therefore, in the chosen sample of tosses, the probability that between 45% and 55% of the outcomes are heads is

$$\begin{aligned} P(0.45 < \mathcal{P} < 0.55) &= P(-1 < z < 1) = 2P(0 < z < 1) \\ &= 2A(1.00) = 2 \times (0.3413) = 0.6826, \end{aligned}$$ ■

Example 2 *Out of 1000 samples of 200 children each, in how many would you expect to find that* (a) *less than 40% are boys,* (b) *between 40% and 60% are boys,* (c) *55% or more are girls.*

▶ Let $\mathcal{P}$ be the proportion of boys in a sample. The probability that a child chosen at random is a boy is $p = 1/2$. Therefore, for the distribution of proportion of boys in a sample of $N = 200$ children,

$$\text{Mean} = \mu_{\mathcal{P}} = p = \frac{1}{2} = 0.5,$$

Scanned with CamScanner

and $$\text{S.D.} = \sigma_{\mathscr{P}} = \left(\frac{pq}{N}\right)^{1/2} = \left[\frac{(1/2)\times(1/2)}{200}\right]^{1/2} = \frac{1}{\sqrt{800}} = 0.0354.$$

The corresponding standard normal variate is

$$z = \frac{\mathscr{P} - \mu_{\mathscr{P}}}{\sigma_{\mathscr{P}}} = \frac{\mathscr{P} - 0.5}{0.0354}.$$

For $\mathscr{P} = 40\%$, we have $z = \dfrac{0.4 - 0.5}{0.0354} = -2.82$,

For $\mathscr{P} = 60\%$, we have $z = \dfrac{0.6 - 0.5}{0.0354} = 2.82$,

For $\mathscr{P} = 45\%$, we have $z = \dfrac{0.45 - 0.5}{0.0354} = -1.41$.

Therefore, the probability that a sample contains less than 40% of boys is

$$\begin{aligned} P(\mathscr{P} < 0.4) &= P(z < -2.82) = P(z > 2.82) = P(z > 0) - P(0 < z < 2.82) \\ &= 0.5 - A(2.82) = 0.5 - 0.4974 = 0.0026. \end{aligned}$$

Accordingly, out of 1000 samples, the expected number of samples containing less than 40% of boys is $0.0026 \times 1000 = 2.6 \approx 3$.

Next, the probability that a sample contains between 40% and 60% of boys is

$$\begin{aligned} P(0.4 < \mathscr{P} < 0.6) &= P(-2.82 < z < 2.82) \\ &= 2P(0 < z < 2.82) = 2A(2.82) = 2 \times 0.4974 = 0.9948. \end{aligned}$$

Accordingly, out of 1000 samples, the expected number of samples containing 40% to 60% of boys is $0.9948 \times 1000 \approx 995$.

Lastly, the probability that a sample contains 55% or more of girls is the same as the probability of having less than 45% of boys. This probability is

$$\begin{aligned} P(\mathscr{P} < 0.45) &= P(z < -1.41) = P(z > 1.41) \\ &= 0.5 - A(1.41) = 0.5 - 0.4192 = 0.0808. \end{aligned}$$

Accordingly, out of 1000 samples, the expected number of samples containing 55% or more of girls is $0.0808 \times 1000 = 81$. ∎

Here, p_1 and p_2 are the probabilites of successes in the samples from the two populations, and $q_1 = 1 - p_1$ and $q_2 = 1 - p_2$.

Similarly, one can define the sampling distributions of sums of means and sums of proportions.

Example 1 *The electric light bulbs of a manufacturer A have a mean life of 1300 hours with a standard deviation of 120 hours, while those of a manufacturer B have a mean life of 1200 hours with a standard deviation of 90 hours. If random samples of 150 bulbs of each brand are tested, what is the probability that brand A bulbs have a mean life which is atleast 80 hours more than brand B bulbs.*

▶ Let $\overline{X}_A$ and $\overline{X}_B$ denote the mean life of samples of brands A and B respectively. Then, by formulas (7) and (8), we have

$\mu_A = 1300$
$\mu_B = 1200$
$\sigma_A = 120$
$\sigma_B = 90$
$N_A = 150$
$N_B = 150$

$$\mu_{(\overline{X}_A - \overline{X}_B)} = 1300 - 1200 = 100$$

and
$$\sigma_{(\overline{X}_A - \overline{X}_B)} = \left(\frac{\sigma_A^2}{N_A} + \frac{\sigma_B^2}{N_B}\right)^{1/2} = \left(\frac{120^2}{150} + \frac{90^2}{150}\right)^{1/2} = \sqrt{(96+54)} = \sqrt{150}.$$

Therefore, the standard normal variate for the difference in means is

$$z = \frac{(\overline{X}_A - \overline{X}_B) - \mu_{(\overline{X}_A - \overline{X}_B)}}{\sigma_{(\overline{X}_A - \overline{X}_B)}} = \frac{(\overline{X}_A - \overline{X}_B) - 100}{\sqrt{150}}.$$

For $\overline{X}_A - \overline{X}_B = 80$, we get $z = \frac{80-100}{\sqrt{150}} = -1.63$.

Therefore, the probability that brand A bulbs will have a mean life which is atleast 80 hours more than brand B bulbs is

$$\begin{aligned} P(\overline{X}_A - \overline{X}_B \geq 80) &= P(z > -1.63) = P(-1.63 < z < 0) + P(z > 0) \\ &= P(0 < z < 1.63) + P(z > 0) \\ &= A(1.63) + 0.5 = 0.445 + 0.5 = 0.945. \end{aligned}$$ ■

Example 2 *Two friends A and B play a game of "heads and tails", each tossing a coin 50 times. A will win the game if he tosses 3 or more heads than B; otherwise B wins. Determine the probability that A wins.*

▶ Let $\mathscr{P}_A$ and $\mathscr{P}_B$ denote the proportions of heads obtained by A and B respectively. The probability of getting a head in a toss is $p = 1/2$ (for both A and B). The number of tosses made by A and B (which are the sample sizes here) are $N_A = N_B = 50$. Therefore, by formulas (11) and (12),

$$\mu_{(\mathscr{P}_A - \mathscr{P}_B)} = p - p = 0,$$