

### 2.2.4 Sampling distribution of $\bar{X}$ (normal population)

If  $X_1, X_2, \dots, X_n$  is a random sample of size  $n$  taken from a normal distribution with mean  $\mu$  and variance  $\sigma^2$  such that  $X \sim N(\mu, \sigma^2)$ , then the distribution of  $\bar{X}$  is also normal and  $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$  where  $\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$ .

Note

1. We use capital letters to represent random variables i.e.  $\bar{X}$  is a random variable,  $\bar{x}$  is an observed value of  $\bar{X}$ .
2.  $\frac{\sigma}{\sqrt{n}}$  is called the standard error of the mean.
3. The mean of  $\bar{X}$  is  $\mu$  i.e.  $E(\bar{X}) = \mu$  or  $\mu_{\bar{X}} = \mu$  and the variance of  $\bar{X}$  is  $\frac{\sigma^2}{n}$  or  $Var(\bar{X}) = \frac{\sigma^2}{n}$  or  $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$  i.e.  $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$  is a standard normal.

Example

A normal distribution has a mean of 40 and a standard deviation of 4. If 25 items are drawn at random, find the probability that their mean is

- (a) 41.4 or more
- (b) between 38.7 and 40.7
- (c) less than 39.5

Soln

(a)

$$X \sim N(40, 4^2), \quad n = 25$$

$$\bar{X} \sim N\left(40, \frac{4^2}{25}\right) \text{ i.e. } \sigma_{\bar{X}} = \frac{4}{5}$$

$$\begin{aligned} P(\bar{X} \geq 41.4) &= P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} > \frac{41.4 - 40}{\frac{4}{5}}\right) \\ &= P(Z > 1.75) \\ &= 1 - P(Z < 1.75) \\ &= 1 - 0.9599 \\ &= 0.0401 \end{aligned}$$

(b)

$$\begin{aligned} P(38.7 < \bar{X} < 40.7) &= P\left(\frac{38.7 - 40}{\frac{4}{5}} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{40.7 - 40}{\frac{4}{5}}\right) \\ &= P(-1.625 < Z < 0.875) \\ &= P(Z < 0.875) - P(Z < -1.625) \\ &= 0.8092 - P(Z > 1.625) \\ &= 0.8092 - (1 - P(Z < 1.625)) \\ &= 0.8092 - (1 - 0.9479) \\ &= 0.7571 \end{aligned}$$

Note:

To find  $P(Z < 0.875)$  we go halfway between 0.87 and 0.88, we get the average of the probabilities for these two i.e.  $\frac{0.8078+0.8106}{2} = 0.8092$ .

Similarly to find  $P(Z < 1.625)$  we find the probability which is halfway between 1.62 and 1.63 i.e.  $\frac{0.9474+0.9484}{2} = 0.9479$

(c)

$$\begin{aligned} P(\bar{X} < 39.5) &= P\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < \frac{39.5 - 40}{\frac{4}{5}}\right) \\ &= P(Z < -0.625) \\ &= P(Z > 0.625) \\ &= 1 - P(Z < 0.625) \\ &= 1 - 0.73405 \\ &= 0.26595 \\ &= 0.2660 \end{aligned}$$

### 2.2.5 Central Limit Theorem

If  $X_1, X_2, \dots, X_n$  is a random sample of size  $n$  taken from a **any** distribution with mean  $\mu$  and variance  $\sigma^2$  then, for large  $n$ , the distribution of the sample mean  $\bar{X}$  is approximately normal and  $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$  where  $\bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$ .

Note

1. Provided  $n$  is large enough, the shape of the sampling distribution of  $\bar{X}$  is approximately normal i.e.

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} Z \sim N(0,1)$$

Read as “converges in distribution to  $N(0,1)$ ”.

The more the underlying population departs from normality, the larger the value of  $n$  necessary to ensure normality.

2. How large should  $n$  be? Generally 30 or more.

### 2.3 One sample Estimation (based on one sample)

Suppose we want to estimate the average salary of all civil servants in Zambia.

- We select a random sample of  $n$  civil servants.
- Calculate  $\bar{x}$  based on our sample
- Use  $\bar{x}$  to estimate  $\mu =$  average salary of all civil servants

The sample mean  $\bar{x}$  is an example of a point estimate for the population mean  $\mu$ . It is obvious that  $\bar{x}$  is (almost) never equal to  $\mu$ . To solve this problem we give a range of possible values of  $\mu$  or construct a confidence interval. This is called interval estimation.

### 2.3.1 Confidence intervals for $\mu$ (one sample problem)

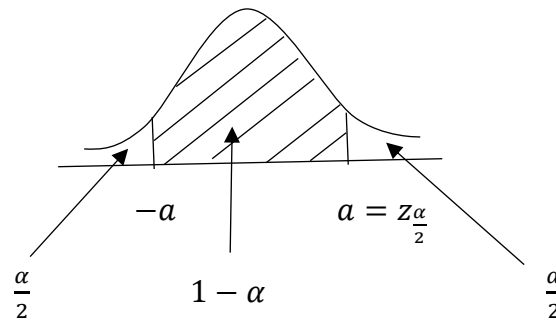
To construct a confidence interval for the population mean  $\mu$ , we consider three cases:

1.  $X_1, X_2, \dots, X_n$  is a random sample from a population random variable  $X$  and  $X \sim N(\mu, \sigma^2)$  where  $\mu$  is unknown and  $\sigma^2$  is known.
2.  $X_1, X_2, \dots, X_n$  is a random sample from a population random variable  $X$  and  $X \sim N(\mu, \sigma^2)$  where  $\mu$  and  $\sigma^2$  are unknown.
3.  $X_1, X_2, \dots, X_n$  is a random sample from a population random variable  $X$  and  $X$  is not normally distributed or its distribution is unknown.

#### Case 1: $X \sim N(\mu, \sigma^2)$ $\sigma^2$ is known

Let

$$P\left(-a < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < a\right) = 1 - \alpha \quad (0 < \alpha < 1)$$



Then

$$\begin{aligned} P\left(-z_{\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < z_{\frac{\alpha}{2}}\right) &= 1 - \alpha \\ P\left(-z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \bar{X} - \mu < z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) &= 1 - \alpha \quad (\text{multiply by } \frac{\sigma}{\sqrt{n}}) \\ P\left(z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} > \mu - \bar{X} > -z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) &= 1 - \alpha \quad (\text{multiply by } -1) \\ P\left(-z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu - \bar{X} < z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) &= 1 - \alpha \\ P\left(\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) &= 1 - \alpha \end{aligned}$$

i.e. the probability that  $\mu$  is in the interval  $\left(\bar{X} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$  is  $1 - \alpha$ . When a particular sample is picked from the population we have the interval  $\left(\bar{x} - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$ .

**Defn**

If  $\bar{x}$  is the mean of a random sample of size  $n$  taken from a normal population with known variance  $\sigma^2$ , then a  $100(1 - \alpha)\%$  confidence interval for  $\mu$  is given by

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

Note

1. Suppose  $\alpha = 0.05$  then we have a 95% confidence interval. What does a 95% confidence interval mean? It means if we were to take 100 samples of size  $n$  from the population and construct 100 confidence intervals then approximately 95 of them will contain  $\mu$ .
2. If  $\bar{x}$  is used to estimate  $\mu$ , we can be  $100(1 - \alpha)\%$  confident that the error will not exceed a specified value  $e$  when the sample size is

$$n = \left( \frac{\sigma z_{\frac{\alpha}{2}}}{e} \right)^2$$

i.e.

$$e = z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$e^2 = \frac{z_{\frac{\alpha}{2}}^2 \sigma^2}{n}$$

$$n = \frac{z_{\frac{\alpha}{2}}^2 \sigma^2}{e^2} = \left( \frac{\sigma z_{\frac{\alpha}{2}}}{e} \right)^2$$

**Example**

After a rainy night, 12 worms surfaced in the garden. The average length of the 12 worms is calculated to be 10.39cm. Assuming that the sample came from a normal population with variance 4,

- (a) find a 95% confidence interval for the mean length of all worms in the garden.
- (b) find a 90% confidence interval for the mean length of all worms in the garden.
- (c) how large a sample is required if we want to be 95% confident that our estimate of  $\mu$  is off by less than 0.5cm.

Soln

(a)

$$X \sim N(\mu, 4)$$

$$\bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \alpha = 0.05$$

$$10.39 \pm z_{0.025} \frac{2}{\sqrt{12}}$$

$$10.39 \pm 1.96 \left( \frac{2}{\sqrt{12}} \right)$$

$$10.39 \pm 1.1316$$

$$(9.26, 11.52)$$

(b)

$$\begin{aligned} \bar{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \alpha = 0.1 \\ 10.39 \pm z_{0.05} \frac{2}{\sqrt{12}} \\ 10.39 \pm 1.645 \left( \frac{2}{\sqrt{12}} \right) \\ 10.39 \pm 0.94974 \\ (9.44, 11.34) \end{aligned}$$

(c)

$$\begin{aligned} n &= \left( \frac{\sigma z_{\frac{\alpha}{2}}}{e} \right)^2, e = 0.5 \\ n &= \left( \frac{2z_{0.025}}{0.5} \right)^2 \\ n &= \left( \frac{2(1.96)}{0.5} \right)^2 \\ n &= 61.46 \text{ (we usually "round up")} \\ n &= 62 \end{aligned}$$