

## 2. Estimation and Sampling Distributions

### 2.1 Introduction

We have defined statistical inference as the process of making judgements about a population based on properties of a sample from the population. There are two branches of statistical inference – **estimation** and **hypothesis testing**.

**Estimation** involves approximating the value of an unknown parameter. Some commonly used statistics are:

Measure	Parameter	Statistic
Mean	$\mu$	$\bar{x}$
Variance	$\sigma^2$	$S^2$
Standard deviation	$\sigma$	$S$
Proportion	$P$	$\hat{P}$

Estimation can either be done by giving a point estimate or interval estimate (confidence interval).

**Hypothesis testing** involves choosing between two opposing statements concerning a population.

### 2.2 Sampling distributions

To evaluate the reliability of our inferences, we will need to know the probability distribution for the statistic we are using.

#### 2.2.1 Defn

The distribution of a statistic is called its sampling distribution.

#### Example

Consider a small population of board members of an organization. Let  $x$  = number of children that each board member has.

Board member	1	2	3	4	5	6	7	8
Number of children	2	2	0	0	2	2	0	3

i.e. board member 1 has 2 children, board member 2 has 2 children, etc...

The average number of children for this population is

$$\mu = \frac{\sum_{i=1}^n x_i}{n} = \frac{2 + 2 + 0 + 0 + 2 + 2 + 0 + 3}{8} = \frac{11}{8} = 1.38$$

To estimate  $\mu$  we pick simple random samples of size 2. How many samples exist?

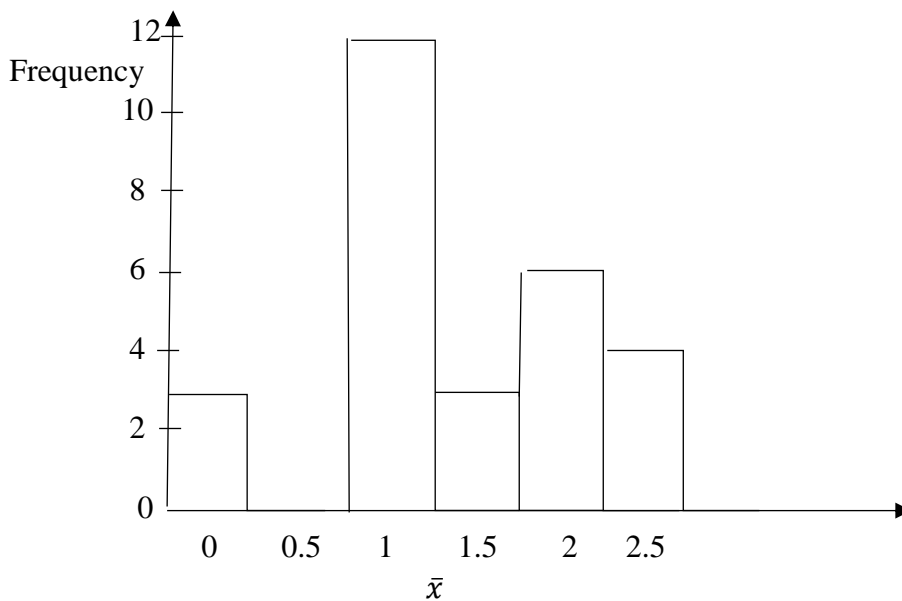
$$\binom{8}{2} = \frac{8!}{2!6!} = 28$$

We can list them and calculate the mean of each of them as follows:

	Sample	$\bar{x}$	
1	(1, 2)	2	i.e. mean for board members 1 and 2 = $\frac{2+2}{2} = 2$
2	(1, 3)	1	i.e. mean for board members 1 and 3 = $\frac{2+0}{2} = 1$
3	(1, 4)	1	i.e. mean for board members 1 and 4 = $\frac{2+0}{2} = 1$
4	(1, 5)	2	i.e. mean for board members 1 and 5 = $\frac{2+2}{2} = 2$
	⋮	⋮	
28	(7, 8)	1.5	i.e. mean for board members 7 and 8 = $\frac{0+3}{2} = 1.5$

(You can complete the whole table as an exercise)

A histogram of the  $\bar{x}$  values is given as



Observations:

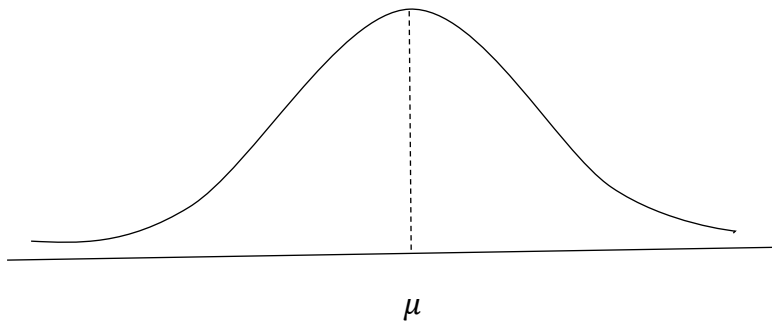
- The histogram tends to be centered around the mean  $\mu$ .
- The histogram resembles a bell-shaped distribution.
- Generally the sampling distribution of  $\bar{x}$  resembles the normal distribution (actually the distribution will look more normal as the number of samples increase).

### 2.2.2 Normal distribution

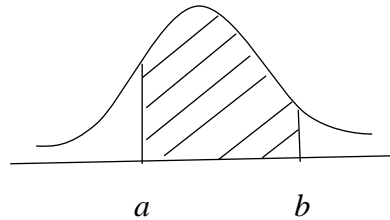
A normal distribution is very important in statistics. A normal distribution is characterized by its mean  $\mu$  and variance  $\sigma^2$ . If a random variable  $X$  is normally distributed with mean  $\mu$  and variance  $\sigma^2$ , then we write  $X \sim N(\mu, \sigma^2)$ . The graph of a normal distribution is called a normal curve.

### Properties of a normal curve

1. The normal curve is bell shaped and has a peak over the mean  $\mu$ .

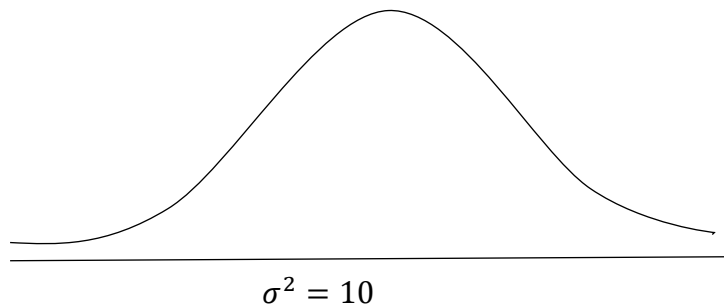
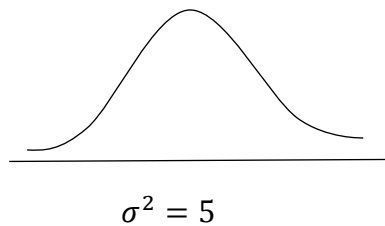


2. It is symmetrical about a vertical line through  $\mu$ .
3. The probability that  $X$  is between  $a$  and  $b$  is given by the area under the curve i.e.



$$\begin{aligned} P(a < X < b) &= \text{Area below the curve between } a \text{ and } b \\ &= P(a < X \leq b) \\ &= P(a \leq X \leq b) \\ &= P(a \leq X < b) \end{aligned}$$

4. The parameters  $\sigma^2$  controls the spread of the curve e.g.



i.e. the second curve is more variable so it will have a bigger variance.