

## CHAPTER 7

### THE SAMPLING DISTRIBUTION OF THE MEAN

#### 7.1 Sampling Error; The need for Sampling Distributions

**Sampling Error** – the error resulting from using a sample characteristic (statistic) to estimate a population characteristic (parameter).

**Example 7.1** (Household Income): Try to estimate the population mean income,  $\mu$ , of all U.S. households by the sample mean income,  $\bar{x}$ , of the 60,000 households surveyed. In 1998, it was reported to be \$51,855 ( $\bar{x}$ ) in CPR (Current Population Reports).

**Q:** How accurate is our sample mean (estimate) likely to be? Is it within \$1000 of the population mean households income?

- **Sampling Distribution:** The distribution of a statistic, or the distribution of all possible observations of the statistic for samples of a given size.

**Example 7.2** (Unrealistic small population case): Heights of the five starting players on a men's basketball team data are given as follows.

Player	A	B	C	D	E
Height	76	78	79	81	86

How confident of our sample mean of size two is within 1 inch of the population mean height?

**Solution:** The population mean height is

$$\mu = \frac{\sum x}{N} = \frac{76 + 78 + 79 + 81 + 86}{5} = 80 \text{ inches.}$$

Possible samples of size two and its means are summarized in the following table (TABLE 7.2).

Samples	A, B	A, C	A, D	A, E	B, C	B, D	B, E	C, D	C, E	D, E
Heights	76,78	76,79	76,81	76,86	78,79	78,81	78,86	79,81	79,86	81,86
$\bar{x}$	77.0	77.5	78.5	81.0	78.5	79.5	82.0	80.0	82.5	83.5

Our sample will be one of the above 10 possible samples. Since there are only 3 samples (C,D and B,D and A,E) which lie within 1 inch of the population mean 80. The confidence will be 30%.

**Sampling Distribution of the sample mean** - the distribution of the variable  $\bar{x}$  (i.e., of all possible sample means) for a given variable x.

## Sample size and Sampling Error

As **sample size** increases, the more **sample means** cluster around the **population mean**, and the **sampling error** of estimating  $\mu$ , by  $\bar{x}$  is smaller.

### What we do in practice?

Large and unknown population -> Obtaining the sampling distribution is not feasible. But we can find approximate sampling distribution of the sample mean for any underlying population distributions, which we discuss in 7.2 and 7.3.

## 7.2 The Mean and Standard deviation of $\bar{x}$

We use the sampling distribution of the sample mean to make inferences about a population mean based on the mean of a sample from the population. But generally we do not know the exact distribution of the sample mean (sampling distribution)

Under certain conditions, we can approximate the sampling distribution of the sample mean ( $\bar{x}$ ) by the normal distribution. Normal distribution is determined by its mean and standard deviation. So let's denote its mean is  $\mu_{\bar{x}}$  and its standard deviation is  $\sigma_{\bar{x}}$ .

### Mean of the variable $\bar{x}$

For samples of size  $n$ , the mean of the variable  $\bar{x}$  equals the mean of the variable  $x$  under consideration: (the mean of all possible sample mean equals the population mean)

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

### Standard Deviation of the variable $\bar{x}$

For samples of size  $n$ , the standard deviation of the variable  $\bar{x}$  equals the standard deviation of the variable under consideration divided by the square root of the sample size: (the standard deviation of all possible sample means equals the population standard deviation divided by the square root of the sample size)

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

## Sample Size and Sampling Error

1. The larger the sample size, the smaller the standard deviation of  $\bar{x}$ .
2. The smaller the standard deviation of  $\bar{x}$ , the more closely its possible values cluster around the mean of  $\bar{x}$ .

3. The mean of  $\bar{x}$  is the same as the population mean:  $\mu_{\bar{x}} = \mu$

**NOTE:** The *standard deviation of  $\bar{x}$*  determines the amount of sampling error to be expected when a population mean is estimated by a sample mean. So often it is referred to as the **standard error** of the sample mean.

**Standard error (SE)** of a statistic – standard deviation of a statistic

### 7.3 The Sampling Distribution of the Mean

#### Sampling Distribution of the Mean for a Normally Distributed Variable

If the variable  $x$  of a population is normally distributed with mean  $\mu$  and standard deviation  $\sigma$ , then, for any sample of size  $n \geq 1$ , the variable  $\bar{x}$  is also *normally distributed* with mean  $\mu$  and standard deviation  $\frac{\sigma}{\sqrt{n}}$ .

- **The Central Limit Theorem (CLT)** – one of the most important theorems in statistics

For a relatively large sample size, the variable  $\bar{x}$  is *approximately normally distributed*, regardless of the distribution of the variable  $x$  under consideration. The approximation becomes better and better with increasing sample size.

#### The Sampling Distribution of the Mean

If a variable  $x$  of a population has mean  $\mu$  and standard deviation  $\sigma$ , then for samples of size  $n$ ,

1. The mean of  $\bar{x}$  equals the population mean  $\mu$ :  $\mu_{\bar{x}} = \mu$
2. The standard deviation of  $\bar{x}$  (standard error of  $\bar{x}$ ) equals the population standard deviation divided by the square root of the sample size:  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$
3. If  $x$  is normally distributed, then so is  $\bar{x}$ , regardless of sample size.
4. If the sample size is large (approximately bigger than 30), then  $\bar{x}$  is approximately normally distributed, regardless of the distribution of  $x$ .