Business Statistics: A Decision-Making Approach

6th Edition

Chapter 6

Introduction to Sampling Distributions

Chapter Goals

- To use information from the sample to make inference about the population
  - Define the concept of sampling error
  - Determine the mean and standard deviation for the sampling distribution of the sample mean
  - Describe the Central Limit Theorem and its importance

Sampling Error

- **Sample Statistics are used to estimate Population Parameters**

  ex: $\overline{X}$ is an estimate of the population mean, μ

- **Problems:**
  - **Different samples provide different estimates of the population parameter**
  - **Sample results have potential variability, thus sampling error exits**

Fundamentals of Business Statistics – Murali Shanker Chap 6-3

Statistical Sampling

- *Parameters* are numerical descriptive measures of populations.
- *Statistics* are numerical descriptive measures of samples
- *Estimators* are sample statistics that are used to estimate the unknown population parameter.
- Question: How close is our sample statistic to the true, but unknown, population parameter?

Fundamentals of Business Statistics – Murali Shanker Chap 6-4

Notations

Parameter	Statistic
μ	$\hat{\mu}, \overline{\mathrm{X}}$, Mode, Median
σ	$\hat{\sigma}, s^2$
p	$\hat{p}$

Fundamentals of Business Statistics – Murali Shanker Chap 6-5

Calculating Sampling Error

- Sampling Error:

  The difference between a value (a statistic) computed from a sample and the corresponding value (a parameter) computed from a population

Example: (for the mean)

$$\text{Sampling Error} = \bar{x} - \mu$$

where:

$\bar{x}$ = sample mean
μ = population mean

Fundamentals of Business Statistics – Murali Shanker Chap 6-6

Example

If the population mean is μ = 98.6 degrees and a sample of n = 5 temperatures yields a sample mean of $\bar{x}$ = 99.2 degrees, then the sampling error is

Fundamentals of Business Statistics – Murali Shanker Chap 6-7

Sampling Distribution

A sampling distribution is a distribution of the possible values of a statistic for a given sample size *n* selected from a population

Fundamentals of Business Statistics – Murali Shanker Chap 6-8

Sampling Distributions

Objective: To find out how the sample mean $\overline{X}$ varies from sample to sample. In other words, we want to find out the sampling distribution of the sample mean.

Sampling Distribution Example

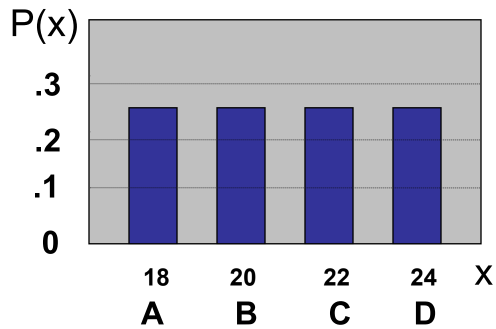
- **Assume there is a population …**
- Population size N=4
- Random variable, X, is age of individuals
- Values of X: 18, 20, 22, 24 (years)

Developing a Sampling Distribution *(continued)*

Summary Measures for the Population Distribution:

$$\mu = \frac{\sum x_i}{N} = \frac{18 + 20 + 22 + 24}{4} = 21$$

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N} = 5$$



Uniform Distribution

Developing a Sampling Distribution *(continued)*

Now consider all possible samples of size n=2

1st Obs	2nd Observation 18	20	22	24
18	18,18	18,20	18,22	18,24
20	20,18	20,20	20,22	20,24
22	22,18	22,20	22,22	22,24
24	24,18	24,20	24,22	24,24

16 possible samples (sampling with replacement)

16 Sample Means

1st Obs	2nd Observation 18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

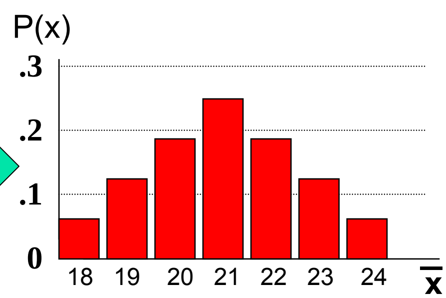
Developing a Sampling Distribution *(continued)*

Sampling Distribution of All Sample Means

16 Sample Means

1st Obs	2nd Observation 18	20	22	24
18	18	19	20	21
20	19	20	21	22
22	20	21	22	23
24	21	22	23	24

Sample Means Distribution



(no longer uniform)

Fundamentals of Business Statistics – Murali Shanker Chap 6-13

Developing a Sampling Distribution *(continued)*

Summary Measures of this Sampling Distribution:

$$\mu_{\bar{x}} = \frac{\sum \bar{x}_i}{N} = \frac{18 + 19 + 21 + \cdots + 24}{16} = 21$$

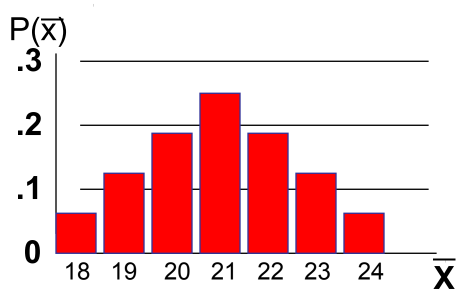
$$\sigma_{\bar{x}}^2 = \frac{\sum (\bar{x}_i - \mu_{\bar{x}})^2}{N}$$

$$= \frac{(18-21)^2 + (19-21)^2 + \cdots + (24-21)^2}{16} = 2.5$$

Fundamentals of Business Statistics – Murali Shanker Chap 6-14

Expected Values

$$E(\overline{X}) =$$



$$\sigma^2(\overline{X}) =$$

Fundamentals of Business Statistics – Murali Shanker   Chap 6-15

Comparing the Population with its Sampling Distribution

Population: N = 4

$\mu = 21$

$\sigma^2 = 5$

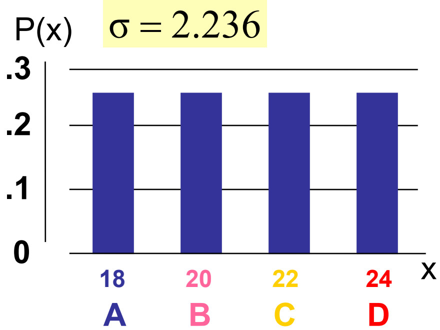
$\sigma = 2.236$

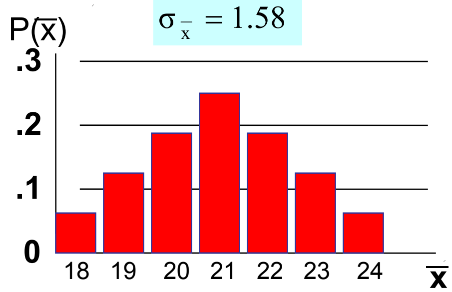
Sample Means Distribution: n = 2

$\mu_{\bar{x}} = 21$

$\sigma_{\bar{x}}^2 = 2.5$

$\sigma_{\bar{x}} = 1.58$





Fundamentals of Business Statistics – Murali Shanker   Chap 6-16

Comparing the Population with its Sampling Distribution

Population: N = 4

$\mu = 21 \quad \sigma^2 = 5 \quad \sigma = 2.236$

Sample Means Distribution: n = 2

$\mu_{\bar{x}} = 21 \quad \sigma_{\bar{x}}^2 = 2.5 \quad \sigma_{\bar{x}} = 1.58$

What is the relationship between the variance in the population and sampling distributions

Fundamentals of Business Statistics – Murali Shanker Chap 6-17

Empirical Derivation of Sampling Distribution

1. Select a random sample of *n* observations from a given population
2. Compute $\bar{X}$
3. Repeat steps (1) and (2) a large number of times
4. Construct a relative frequency histogram of the resulting $\bar{X}$

Fundamentals of Business Statistics – Murali Shanker Chap 6-18

Important Points

1. The mean of the sampling distribution of $\overline{X}$ is the same as the mean of the population being sampled from. That is,

$$E\left(\overline{X}\right) = \mu_{\overline{X}} = \mu_X = \mu$$

2. The variance of the sampling distribution of $\overline{X}$ is equal to the variance of the population being sampled from divided by the sample size. That is,

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$$

Fundamentals of Business Statistics – Murali Shanker Chap 6-19

Imp. Points (Cont.)

3. If the original population is normally distributed, then for **any** sample size *n* the distribution of the sample mean is also normal. That is,

$$X \sim N\left(\mu, \sigma^2\right) \text{then } \overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

4. If the distribution of the original population is not known, but *n* is sufficiently "large", the distribution of the sample mean is **approximately** normal with mean and variance given as $\overline{X} \sim N\left(\mu, \sigma^2/n\right)$. This result is known as the **central limit theorem (CLT)**.

Fundamentals of Business Statistics – Murali Shanker Chap 6-20

Standardized Values

- Z-value for the sampling distribution of $\bar{X}$:

where: $\bar{X}$ = sample mean
μ = population mean
σ = population standard deviation
n = sample size

Example

Let X the length of pregnancy be X~N(266,256).

1. What is the probability that a randomly selected pregnancy lasts more than 274 days. I.e., what is $P(X > 274)$?

2. Suppose we have a random sample of n = 25 pregnant women. Is it less likely or more likely (as compared to the above question), that we might observe a sample mean pregnancy length of more than 274 days. I.e., what is $P(\bar{X} > 274)$

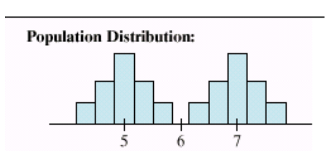
YDI 9.8

- *The model for breaking strength of steel bars is normal with a mean of* 260 *pounds per square inch and a variance of* 400*. What is the probability that a randomly selected steel bar will have a breaking strength greater than* 250 *pounds per square inch?*

- *A shipment of steel bars will be accepted if the mean breaking strength of a random sample of* 10 *steel bars is greater than* 250 *pounds per square inch. What is the probability that a shipment will be accepted?*

YDI 9.9

- The following histogram shows the population distribution of a variable *X*. How would the sampling distribution of $\overline{X}$ look, where the mean is calculated from random samples of size 150 from the above population?



Example

- Suppose a population has mean $\mu = 8$ and standard deviation $\sigma = 3$. Suppose a random sample of size $n = 36$ is selected.
- What is the probability that the sample mean is between 7.8 and 8.2?

Desirable Characteristics of Estimators

- An estimator $\hat{\theta}$ is **unbiased** if the mean of its sampling distribution is equal to the population parameter θ to be estimated. That is, $\hat{\theta}$ is an unbiased estimator of θ if $E(\hat{\theta}) = \theta$. Is $\overline{X}$ an unbiased estimator of μ?

Consistent Estimator

An estimator $\hat{\theta}$ is a **consistent** estimator of a population parameter θ if the larger the sample size, the more likely $\hat{\theta}$ will be closer to θ. Is $\overline{X}$ a consistent estimator of μ?

Efficient Estimator

The **efficiency** of an unbiased estimator is measured by the variance of its sampling distribution. If two estimators based on the same sample size are both unbiased, the one with the smaller variance is said to have greater relative efficiency than the other.