Business Statistics: A Decision-Making Approach

6th Edition

Chapter 6
Introduction to Sampling Distributions

Chapter Goals

After completing this chapter, you should be able to:

- Define the concept of sampling error
- Determine the mean and standard deviation for the sampling distribution of the sample mean, $\bar{x}$
- Determine the mean and standard deviation for the sampling distribution of the sample proportion, $\bar{p}$
- Describe the Central Limit Theorem and its importance
- Apply sampling distributions for both $\bar{x}$ and $\bar{p}$

Sampling Error

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

  ex: $\bar{X}$ 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**

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

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 6-4

Review

- Population mean: Sample Mean:

$$\mu = \frac{\sum x_i}{N} \qquad \bar{x} = \frac{\sum x_i}{n}$$

where:

μ = Population mean
$\bar{x}$ = sample mean
x_i = Values in the population or sample
N = Population size
n = sample size

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 6-5

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

$$\bar{x} - \mu = 98.6 - 99.2 = -0.6 \text{ degrees}$$

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 6-6

Sampling Errors

- Different samples will yield different sampling errors
- The sampling error may be positive or negative ($\bar{x}$ may be greater than or less than μ)
- The expected sampling error decreases as the sample size increases

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 6-7

Sampling Distribution

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

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 6-8

Developing a Sampling Distribution

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

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 6-9

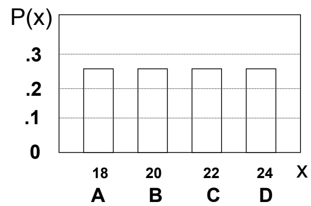
Developing a Sampling Distribution

(continued)

Summary Measures for the Population Distribution:

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

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}} = 2.236$$



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

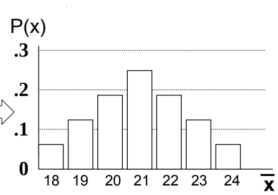
(continued)

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)

Developing a Sampling Distribution

(continued)

Summary Measures of this Sampling Distribution:

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

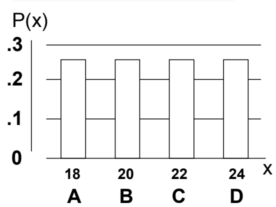
$$\sigma_{\bar{x}} = \sqrt{\frac{\sum (\bar{x}_i - \mu_{\bar{x}})^2}{N}}$$

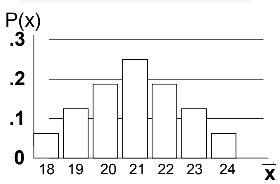
$$= \sqrt{\frac{(18-21)^2 + (19-21)^2 + \cdots + (24-21)^2}{16}} = 1.58$$

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 6-13

Comparing the Population with its Sampling Distribution

Population N = 4	Sample Means Distribution n = 2
$\mu = 21 \quad \sigma = 2.236$	$\mu_{\bar{x}} = 21 \quad \sigma_{\bar{x}} = 1.58$





Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 6-14

If the Population is Normal

(THEOREM 6-1)

If a population is normal with mean μ and standard deviation σ, the sampling distribution of $\bar{x}$ is also normally distributed with

$$\mu_{\bar{x}} = \mu \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc. Chap 6-15

z-value for Sampling Distribution of $\bar{x}$

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

$$z = \frac{(\bar{x} - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

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

Finite Population Correction

- Apply the **Finite Population Correction** if:
  - the sample is large relative to the population (n is greater than 5% of N)

  and…
  - Sampling is without replacement

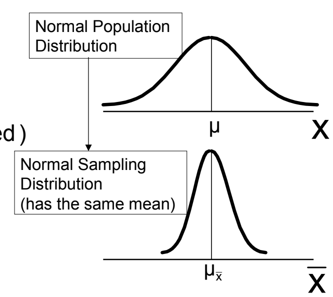
Then

$$z = \frac{(\bar{x} - \mu)}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$$

Sampling Distribution Properties

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

(i.e. $\bar{x}$ is unbiased)

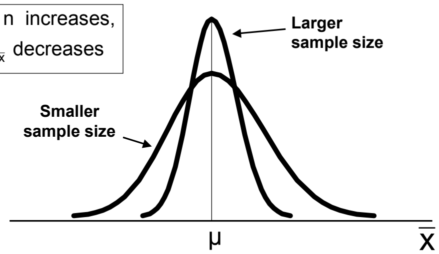


Sampling Distribution Properties

(continued)

- For sampling with replacement:

As n increases, $\sigma_{\bar{x}}$ decreases



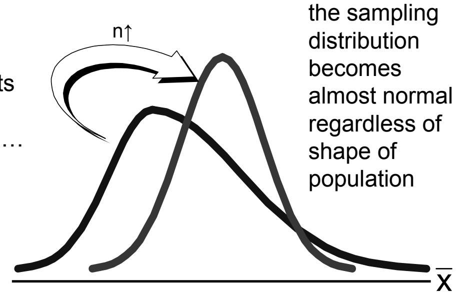
If the Population is **not** Normal

- We can apply the Central Limit Theorem:
  - Even if the population is not normal,
  - …sample means from the population will be approximately normal as long as the sample size is large enough
  - …and the sampling distribution will have

$\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$

Central Limit Theorem

As the sample size gets large enough…



the sampling distribution becomes almost normal regardless of shape of population

If the Population is **not** Normal

(continued)

Sampling distribution properties:

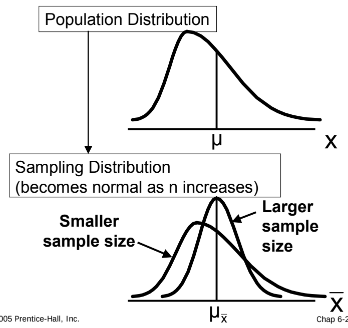
Central Tendency

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

Variation

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

(Sampling with replacement)



 Chap 6-22

How Large is Large Enough?

- For most distributions, n > 30 will give a sampling distribution that is nearly normal
- For fairly symmetric distributions, n > 15
- For normal population distributions, the sampling distribution of the mean is always normally distributed

 Chap 6-23

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?

 Chap 6-24

Example

(continued)

Solution:

- Even if the population is not normally distributed, the central limit theorem can be used (n > 30)
- … so the sampling distribution of $\bar{x}$ is approximately normal
- … with mean $\mu_{\bar{x}} = 8$
- …and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{36}} = 0.5$

 Chap 6-25

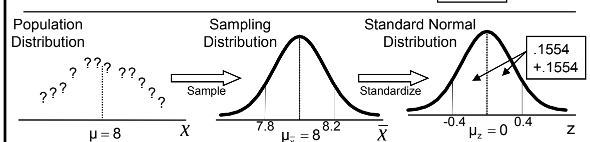
Example

(continued)

Solution (continued):

$$P(7.8 < \mu_{\bar{x}} < 8.2) = P\left(\frac{7.8-8}{3/\sqrt{36}} < \frac{\mu_{\bar{x}} - \mu}{\sigma/\sqrt{n}} < \frac{8.2-8}{3/\sqrt{36}} \right)$$

$$= P(-0.4 < z < 0.4) = \boxed{0.3108}$$



 Chap 6-26

Population Proportions, p

p = the proportion of population having some characteristic

- Sample proportion ($\bar{p}$) provides an estimate of p:

$$\bar{p} = \frac{x}{n} = \frac{\text{number of successes in the sample}}{\text{sample size}}$$

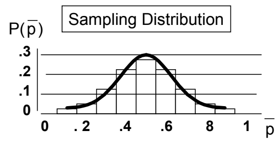
- If two outcomes, $\bar{p}$ has a binomial distribution

 Chap 6-27

Sampling Distribution of p

- Approximated by a normal distribution if:
  - $np \geq 5$

    $n(1-p) \geq 5$



where

$$\mu_{\bar{p}} = p$$ and $$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$$

(where p = population proportion)

z-Value for Proportions

Standardize $\bar{p}$ to a z value with the formula:

$$z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{\bar{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

- If sampling is **without replacement** and n is greater than 5% of the population size, then $\sigma_{\bar{p}}$ must use the **finite population correction factor:**

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}\sqrt{\frac{N-n}{N-1}}$$

Example

- If the true proportion of voters who support Proposition A is p = .4, what is the probability that a sample of size 200 yields a sample proportion between .40 and .45?
- i.e.: **if p = .4 and n = 200, what is $P(.40 \leq \bar{p} \leq .45)$?**

Example

(continued)

- **if p = .4 and n = 200, what is $P(.40 \le \bar{p} \le .45)$?**

Find $\sigma_{\bar{p}}$:

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{.4(1-.4)}{200}} = .03464$$

Convert to standard normal:

$$P(.40 \le \bar{p} \le .45) = P\left(\frac{.40-.40}{.03464} \le z \le \frac{.45-.40}{.03464}\right)$$

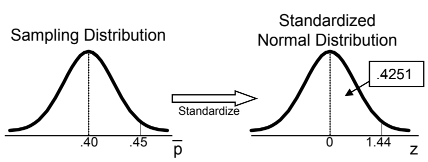
$$= P(0 \le z \le 1.44)$$

Example

(continued)

- if p = .4 and n = 200, what is $P(.40 \le \bar{p} \le .45)$?

Use standard normal table: $P(0 \le z \le 1.44) =$ .4251



Chapter Summary

- Discussed sampling error
- Introduced sampling distributions
- Described the sampling distribution of the mean
  - For normal populations
  - Using the Central Limit Theorem
- Described the sampling distribution of a proportion
- Calculated probabilities using sampling distributions
- Discussed sampling from finite populations

