POPULATION AND SAMPLING DISTRIBUTIONS

- Population Distribution
- Sampling Distribution

Population Distribution

Definition

The ***population distribution*** is the probability distribution of the population data.

Population Distribution cont.

- Suppose there are only five students in an advanced statistics class and the midterm scores of these five students are
  - 70 78 80 80 95
- Let x denote the score of a student

Table 1

Population Frequency and Relative Frequency Distributions

x	f	Relative Frequency
70	1	1/5 = .20
78	1	1/5 = .20
80	2	2/5 = .40
95	1	1/5 = .20
	N = 5	Sum = 1.00

Table 2 Population Probability Distribution

x	P (x)
70	.20
78	.20
80	.40
95	.20
	$\Sigma P(x) = 1.00$

Sampling Distribution

Definition

The probability distribution of $\bar{x}$ is called its sampling distribution. It lists the various values that $\bar{x}$ can assume and the probability of each value of $\bar{x}$.

In general, the probability distribution of a sample statistic is called its ***sampling distribution***.

Sampling Distribution cont.

- Reconsider the population of midterm scores of five students given in Table 1
- Consider all possible samples of three scores each that can be selected, without replacement, from that population.
- The total number of possible samples is

$$_5C_3 = \frac{5!}{3!(5-3)!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1 \cdot 2 \cdot 1} = 10$$

Sampling Distribution cont.

- Suppose we assign the letters A, B, C, D, and E to the scores of the five students so that
  - A = 70, B = 78, C = 80, D = 80, E = 95
- Then, the 10 possible samples of three scores each are
  - ABC, ABD, ABE, ACD, ACE, ADE, BCD, BCE, BDE, CDE

Table 3 All Possible Samples and Their Means When the Sample Size is 3

Sample	Scores in the Sample	$\bar{x}$
ABC	70, 78, 80	76.00
ABD	70, 78, 80	76.00
ABE	70, 78, 95	81.00
ACD	70, 80, 80	76.67
ACE	70, 80, 95	81.67
ADE	70, 80, 95	81.67
BCD	78, 80, 80	79.33
BCE	78, 80, 95	84.33
BDE	78, 80, 95	84.33
CDE	80, 80, 95	85.00

Table 4 Frequency and Relative Frequency Distributions of $\bar{x}$ when the Sample Size is 3

$\bar{x}$	f	Relative Frequency
76.00	2	2/10 = .20
76.67	1	1/10 = .10
79.33	1	1/10 = .10
81.00	1	1/10 = .10
81.67	2	2/10 = .20
84.33	2	2/10 = .20
85.00	1	1/10 = .10
	Σf = 10	Sum = 1.00

Table 5 Sampling Distribution of $\bar{x}$ when the Sample Size is 3

$\bar{x}$	$P(\bar{x})$
76.00	.20
76.67	.10
79.33	.10
81.00	.10
81.67	.20
84.33	.20
85.00	.10
	$\Sigma P(\bar{x})$ = 1.00

SAMPLING AND NONSAMPLING ERRORS

Definition

Sampling error is the difference between the value of a sample statistic and the value of the corresponding population parameter. In the case of the mean,

Sampling error = $\bar{x} - \mu$

assuming that the sample is random and no nonsampling error has been made.

SAMPLING AND NONSAMPLING ERRORS cont.

Definition

The errors that occur in the collection, recording, and tabulation of data are called ***nonsampling errors***.

Example 1

Reconsider the population of five scores given in Table 1. Suppose one sample of three scores is selected from this population, and this sample includes the scores 70, 80, and 95. Find the sampling error.

Solution 1

$$\mu = \frac{70+78+80+80+95}{5} = 80.60$$

$$\bar{x} = \frac{70+80+95}{3} = 81.67$$

$$\text{Sampling error} = \bar{x} - \mu = 81.67 - 80.60 = 1.07$$

That is, the mean score estimated from the sample is 1.07 higher than the mean score of the population.

SAMPLING AND NONSAMPLING ERRORS cont.

- Now suppose, when we select the sample of three scores, we mistakenly record the second score as 82 instead of 80
- As a result, we calculate the sample mean as

$$\bar{x} = \frac{70+82+95}{3} = 82.33$$

SAMPLING AND NONSAMPLING ERRORS cont.

- The difference between this sample mean and the population mean is

$$\bar{x} - \mu = 82.33 - 80.60 = 1.73$$

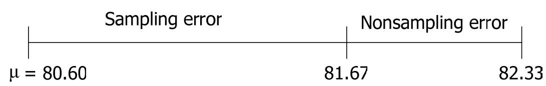
- This difference does not represent the sampling error.
  - Only 1.07 of this difference is due to the sampling error

SAMPLING AND NONSAMPLING ERRORS cont.

- The remaining portion represents the nonsampling error
  - It is equal to 1.70 – 1.07 = .66
  - It occurred due to the error we made in recording the second score in the sample
- Also,

$$\text{Nonsampling error} = \text{Incorrect } \bar{x} - \text{Correct } \bar{x} = 82.33 - 81.67 = .66$$

Figure 1 Sampling and nonsampling errors.



MEAN AND STANDARD DEVIATION OF $\bar{x}$

Definition

The mean and standard deviation of the sampling distribution of $\bar{x}$ are called the ***mean and standard deviation of*** $\bar{x}$ and are denoted by $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$, respectively.

MEAN AND STANDARD DEVIATION OF $\bar{x}$ cont.

Mean of the Sampling Distribution of $\bar{x}$

The ***mean of the sampling distribution of*** $\bar{x}$ is always equal to the mean of the population. Thus,

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

MEAN AND STANDARD DEVIATION OF $\bar{x}$ cont.

Standard Deviation of the Sampling Distribution of $\bar{x}$

The ***standard deviation of the sampling distribution of*** $\bar{x}$ is

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

where σ is the standard deviation of the population and n is the sample size. This formula is used when $n / N \leq .05$, where N is the population size.

Standard Deviation of the Sampling Distribution of $\bar{x}$ cont.

If the condition $n / N \leq .05$ is not satisfied, we use the following formula to calculate $\sigma_{\bar{x}}$:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

where the factor $\sqrt{\frac{N-n}{N-1}}$ is called the finite population correction factor

Two Important Observations

1. The spread of the sampling distribution of $\bar{x}$ is smaller than the spread of the corresponding population distribution, i.e. $\sigma_{\bar{x}} < \sigma_x$
2. The standard deviation of the sampling distribution of $\bar{x}$ decreases as the sample size increases

Example 2

The mean wage for all 5,000 employees who work at a large company is €17.50 and the standard deviation is €2.90. Let $\bar{x}$ be the mean wage per hour for a random sample of certain employees selected from this company. Find the mean and standard deviation of $\bar{x}$ for a sample size of

a) 30 b) 75 c) 200

Solution 2

a)

$$\mu_{\bar{x}} = \mu = €17.50$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.90}{\sqrt{30}} = €.529$$

Solution 2

b)

$$\mu_{\bar{x}} = \mu = €17.50$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.90}{\sqrt{75}} = €.335$$

Solution 2

c)

$$\mu_{\bar{x}} = \mu = €17.50$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.90}{\sqrt{200}} = €.205$$

SHAPE OF THE SAMPLING DISTRIBUTION OF $\bar{x}$

- Sampling from a Normally Distributed Population
- Sampling from a Population That Is Not Normally Distributed

Sampling From a Normally Distributed Population

If the population from which the samples are drawn is normally distributed with mean μ and standard deviation σ , then the sampling distribution of the sample mean, $\bar{x}$,will also be normally distributed with the following mean and standard deviation, irrespective of the sample size:

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

Figure 2 Population distribution and sampling distributions of $\bar{x}$.

(a) Population distribution.

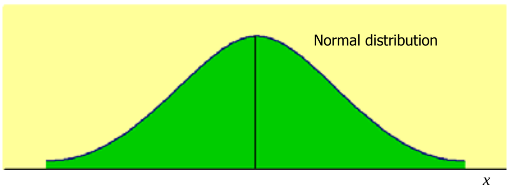


Figure 2 Population distribution and sampling distributions of $\bar{x}$.

(b) Sampling distribution of $\bar{x}$ for n = 5.

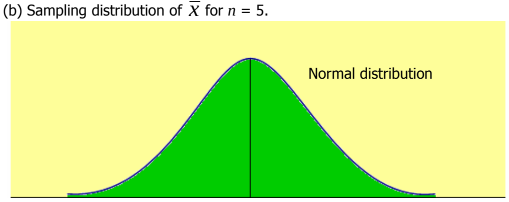


Figure 2 Population distribution and sampling distributions of $\bar{x}$.

(c) Sampling distribution of $\bar{x}$ for n = 16.

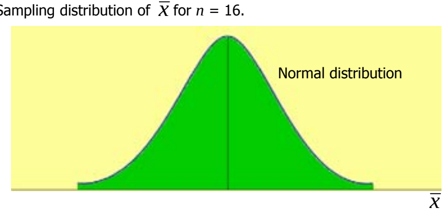


Figure 2 Population distribution and sampling distributions of $\bar{x}$.

(d) Sampling distribution of $\bar{x}$ for n = 30.

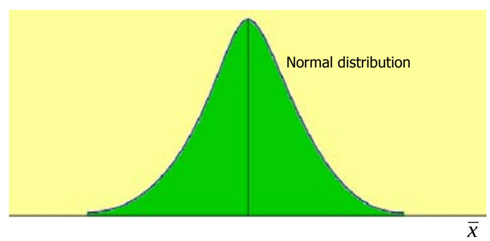
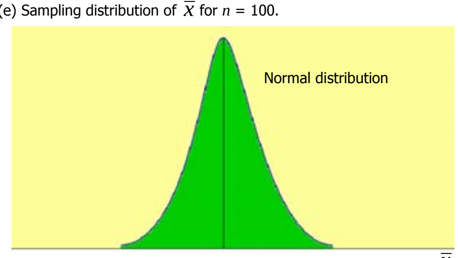


Figure 2 Population distribution and sampling distributions of $\bar{x}$.

(e) Sampling distribution of $\bar{x}$ for n = 100.



Example 3

In a recent exam, the mean score for all examinees was 1,020. Assume that the distribution of scores of all examinees is normal with the mean of 1,020 and a standard deviation of 153. Let $\bar{x}$ be the mean score of a random sample of certain examinees. Calculate the mean and standard deviation of $\bar{x}$ and describe the shape of its sampling distribution when the sample size is

a) 16 b) 50 c) 1000

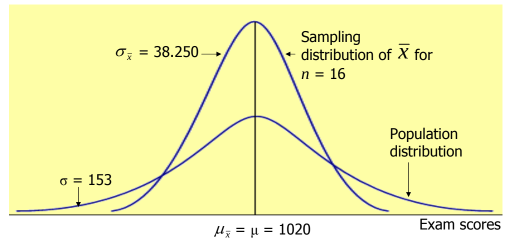
Solution 3

a)

$$\mu_{\bar{x}} = \mu = 1020$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{153}{\sqrt{16}} = 38.250$$

Figure 3



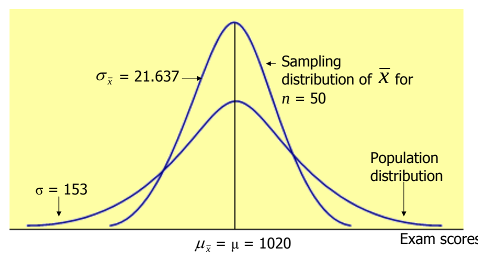
Solution 3

b)

$$\mu_{\bar{x}} = \mu = 1020$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{153}{\sqrt{50}} = 21.637$$

Figure 4



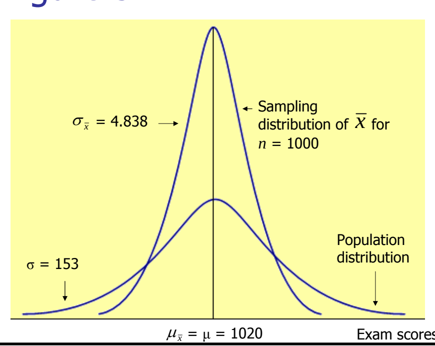
Solution 3

c)

$$\mu_{\bar{x}} = \mu = 1020$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{153}{\sqrt{1000}} = 4.838$$

Figure 5



Sampling From a Population That Is Not Normally Distributed

Central Limit Theorem

According to the ***central limit theorem***, for a large sample size, the sampling distribution of $\bar{x}$ is approximately normal, irrespective of the shape of the population distribution. The mean and standard deviation of the sampling distribution of $\bar{x}$ are

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

The sample size is usually considered to be large if $n \geq 30$.

Figure 6 Population distribution and sampling distributions of $\bar{x}$.

(a) Population distribution.

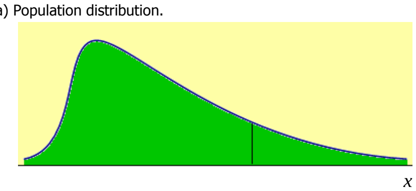


Figure 6 Population distribution and sampling distributions of $\bar{x}$.

(b) Sampling distribution of $\bar{x}$ for $n = 4$.

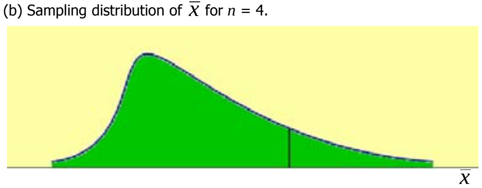


Figure 6 Population distribution and sampling distributions of $\bar{x}$.

(c) Sampling distribution of $\bar{x}$ for $n = 15$.

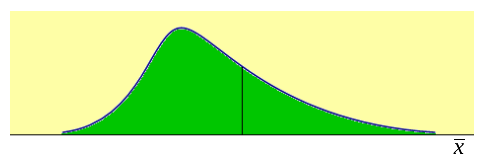


Figure 6 Population distribution and sampling distributions of $\bar{x}$.

(d) Sampling distribution of $\bar{x}$ for $n = 30$.

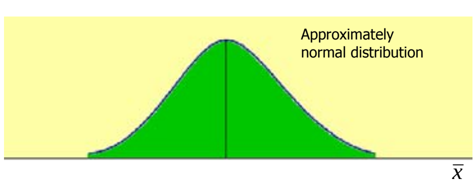
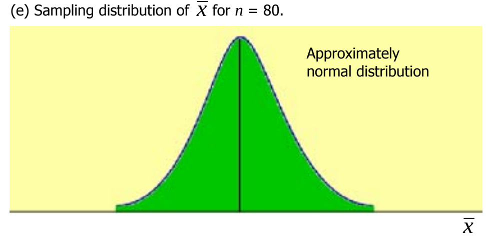


Figure 6 Population distribution and sampling distributions of $\bar{x}$.

(e) Sampling distribution of $\bar{x}$ for $n = 80$.



Example 4

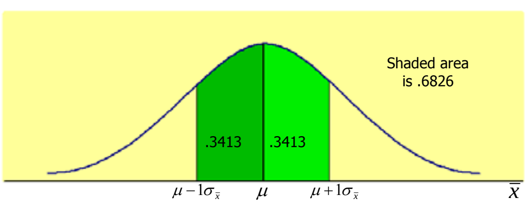
The mean rent paid by all tenants in a large city is €1,550 with a standard deviation of €225. However, the population distribution of rents for all tenants in this city is skewed to the right. Calculate the mean and standard deviation of $\bar{x}$ and describe the shape of its sampling distribution when the sample size is

a) 30 b) 100

APPLICATIONS OF THE SAMPLING DISTRIBUTION OF $\bar{x}$

1. If we make all possible samples of the same (large) size from a population and calculate the mean for each of these samples, then about 68.26% of the sample means will be within one standard deviation of the population mean.

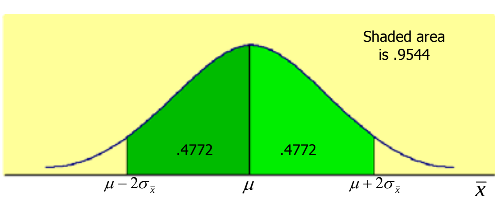
Figure 9 $P(\mu - 1\sigma_{\bar{x}} \le \bar{x} \le \mu + 1\sigma_{\bar{x}})$



APPLICATIONS OF THE SAMPLING DISTRIBUTION OF $\bar{x}$ cont.

2. If we take all possible samples of the same (large) size from a population and calculate the mean for each of these samples, then about 95.44% of the sample means will be within two standard deviations of the population mean.

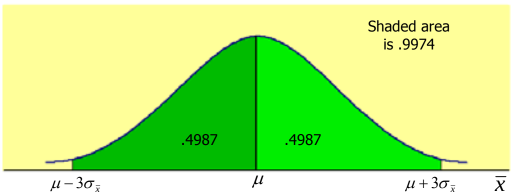
Figure 10 $P(\mu - 2\sigma_{\bar{x}} \le \bar{x} \le \mu + 2\sigma_{\bar{x}})$



APPLICATIONS OF THE SAMPLING DISTRIBUTION OF $\bar{x}$ cont.

3. If we take all possible samples of the same (large) size from a population and calculate the mean for each of these samples, then about 99.74% of the sample means will be within three standard deviations of the population mean.

Figure 11 $P(\mu - 3\sigma_{\bar{x}} \le \bar{x} \le \mu + 3\sigma_{\bar{x}})$



Example 5

Assume that the weights of all packages of a certain brand of cookies are normally distributed with a mean of 32 ounces and a standard deviation of .3 ounce. Find the probability that the mean weight, $\bar{x}$, of a random sample of 20 packages of this brand of cookies will be between 31.8 and 31.9 ounces.

Solution 5

$$\mu_{\bar{x}} = \mu = 32 \text{ ounces}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{.3}{\sqrt{20}} = .06708204 \text{ ounce}$$

z Value for a Value of $\bar{x}$

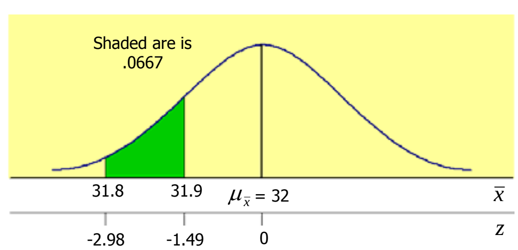
The ***z value for a value of*** $\bar{x}$ is calculated as

$$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

Solution 5

- For $\bar{x}$ = 31.8: $z = \frac{31.8 - 32}{.06708204} = -2.98$
- For $\bar{x}$ = 31.9: $z = \frac{31.9 - 32}{.06708204} = -1.49$
- $P(31.8 < \bar{x} < 31.9) = P(-2.98 < z < -1.49)$
  = .99856 - .93189
  = **.0667**

Figure 12



POPULATION AND SAMPLE PROPORTIONS

The ***population and sample proportions***, denoted by P and p, respectively, are calculated as

$$P = \frac{X}{N} \quad \text{and} \quad p = \frac{x}{n}$$

POPULATION AND SAMPLE PROPORTIONS cont.

where

- N = total number of elements in the population
- n = total number of elements in the sample
- X = number of elements in the population that possess a specific characteristic
- x = number of elements in the sample that possess a specific characteristic

Example 7

Suppose a total of 789,654 families live in a city and 563,282 of them own homes. A sample of 240 families is selected from this city, and 158 of them own homes. Find the proportion of families who own homes in the population and in the sample. Find the sampling error.

Solution 7

$$P = \frac{X}{N} = \frac{563{,}282}{789{,}654} = .71$$

$$p = \frac{x}{n} = \frac{158}{240} = .66$$

$$\text{Sampling error} = p - P = .66 - .71 = -.05$$

MEAN, STANDARD DEVIATION, AND SHAPE OF THE SAMPLING DISTRIBUTION OF p

- Sampling Distribution of p
- Mean and Standard Deviation of p
- Shape of the Sampling Distribution of p

Sampling Distribution of p

Definition

The probability distribution of the sample proportion, p, is called its ***sampling distribution***. It gives various values that p can assume and their probabilities.

Example 8

Boe Consultant Associates has five employees. Table 6 gives the names of these five employees and information concerning their knowledge of statistics.

Table 6 Information on the Five Employees of Boe Consultant Associates

Name	Knows Statistics
Ally	yes
John	no
Susan	no
Lee	yes
Tom	yes

Example 8

- If we define the population proportion, P, as the proportion of employees who know statistics, then
- $P = 3 / 5 = .60$

Example 8

- Now, suppose we draw all possible samples of three employees each and compute the proportion of employees, for each sample, who know statistics.

$$\text{Total number of samples} = {}_5C_3 = \frac{5!}{3!(5-3)!} = \frac{5\cdot4\cdot3\cdot2\cdot1}{3\cdot2\cdot1\cdot2\cdot1} = 10$$

Table 7 All Possible Samples of Size 3 and the Value of p for Each Sample

Sample	Proportion Who Know Statistics p
Ally, John, Susan	1/3 = .33
Ally, John, Lee	2/3 = .67
Ally, John, Tom	2/3 = .67
Ally, Susan, Lee	2/3 = .67
Ally, Susan, Tom	2/3 = .67
Ally, Lee, Tom	3/3 = 1.00
John, Susan, Lee	1/3 = .33
John, Susan, Tom	1/3 = .33
John, Lee, Tom	2/3 = .67
Susan, Lee, Tom	2/3 = .67

Table 8 Frequency and Relative Frequency Distribution of p when the Sample Size is 3

p	f	Relative Frequency
.33	3	3/10 = .30
.67	6	6/10 = .60
1.00	1	1/10 = .10
	$\Sigma f = 10$	Sum = 1.00

Table 9 Sampling Distribution of $\bar{p}$ When the Sample Size is 3

$\bar{p}$	$P(\bar{p})$
.33	.30
.67	.60
1.00	.10
	$\Sigma P(\bar{p}) = 1.00$

Mean and Standard Deviation of $\bar{p}$

Mean of the Sample Proportion

The ***mean of the sample proportion***, $\bar{p}$, is denoted by $\mu_{\bar{p}}$ and is equal to the population proportion, P. Thus,

$$\mu_{\bar{p}} = P$$

Mean and Standard Deviation of $\bar{p}$ cont.

Standard Deviation of the Sample Proportion

The ***standard deviation of the sample proportion***, $\bar{p}$, is denoted by $\sigma_{\bar{p}}$ and is given by the formula

$$\sigma_{\bar{p}} = \sqrt{\frac{PQ}{n}}$$

Where P is the population proportion, Q = 1 – P, and n is the sample size. This formula is used when $n/N \leq .05$, where N is the population size.

Mean and Standard Deviation of $\bar{p}$ cont.

If the $n/N \leq .05$ condition is not satisfied, we use the following formula to calculate $\sigma_{\bar{p}}$:

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

where the factor $\sqrt{\frac{N-n}{N-1}}$ is called the finite population correction factor

Shape of the Sampling Distribution of $\bar{p}$

Central Limit Theorem for Sample Proportion

According to the central limit theorem, the sampling distribution of $\bar{p}$ is approximately normal for sufficiently large sample size. In the case of proportion, the sample size is considered to be sufficiently large if nP and nQ are both greater than 5 – that is if

$$nP > 5 \quad \text{and} \quad nQ > 5$$

Example 9

A National Survey of Students shows about 87% of first year students and final year students rate their college experience as "good" or "excellent" in 2002. Assume this result is true for the current population of all first year students and final year students. Let $\bar{p}$ be the proportion of first year students and final year students in a random sample of 900 who hold this view. Find the mean and standard deviation of $\bar{p}$ and describe the shape of its sampling distribution.

Solution 9

$$P = .87 \quad \text{and} \quad Q = 1 - P = 1 - .87 = .13$$
$$\mu_p = P = .87$$
$$\sigma_p = \sqrt{\frac{PQ}{n}} = \sqrt{\frac{(.87)(.13)}{900}} = .011$$
$$nP = 900(.87) = 783 \quad \text{and} \quad nQ = 900(.13) = 117$$

Solution 9

- nP and nQ are both greater then 5
- Therefore, the sampling distribution of p is approximately normal with a mean of .87 and a standard deviation of .011, as shown in Figure 15

Figure 15