**Chapter 7**

**Sampling and Sampling Distributions**

**Learning Objectives**

1. Understand the importance of sampling and how results from samples can be used to provide estimates of population characteristics such as the population mean, the population standard deviation and / or the population proportion.

2. Know what simple random sampling is and how simple random samples are selected.

3. Understand the concept of a sampling distribution.

4. Understand the central limit theorem and the important role it plays in sampling.

5. Specifically know the characteristics of the sampling distribution of the sample mean () and the sampling distribution of the sample proportion ().

6. Learn about a variety of sampling methods including stratified random sampling, cluster sampling, systematic sampling, convenience sampling and judgment sampling.

7. Know the definition of the following terms:

parameter target population

sampled population sampling distribution

sample statistic finite population correction factor

simple random sampling standard error

sampling without replacement central limit theorem

sampling with replacement unbiased

point estimator relative efficiency

point estimate consistency

**Solutions:**

1. a. AB, AC, AD, AE, BC, BD, BE, CD, CE, DE

b. With 10 samples, each has a 1/10 probability.

c. E and C because 8 and 0 do not apply; 5 identifies E; 7 does not apply; 5 is skipped since E is already in the sample; 3 identifies C; 2 is not needed since the sample of size 2 is complete.

2. Using the last 3-digits of each 5-digit grouping provides the random numbers:

601, 022, 448, 147, 229, 553, 147, 289, 209

Numbers greater than 350 do not apply and the 147 can only be used once. Thus, the simple random sample of four includes 22, 147, 229, and 289.

3. 459, 147, 385, 113, 340, 401, 215, 2, 33, 348

4. a. 5, 0, 5, 8

Bell South, LSI Logic, General Electric

b. 

5. 283, 610, 39, 254, 568, 353, 602, 421, 638, 164

6. 2782, 493, 825, 1807, 289

7. 108, 290, 201, 292, 322, 9, 244, 249, 226, 125, (continuing at the top of column 9) 147, and 113.

8. Random numbers used: 13, 8, 27, 23, 25, 18

The second occurrence of the random number 13 is ignored.

Companies selected: ExxonMobil, Chevron, Travelers, Microsoft, Pfizer, and Intel

9. 102, 115, 122, 290, 447, 351, 157, 498, 55, 165, 528, 25

10. a. Finite population. A frame could be constructed obtaining a list of licensed drivers from the New York State driver’s license bureau.

b. Infinite population. Sampling from a process. The process is the production line producing boxes of cereal.

c. Infinite population. Sampling from a process. The process is one of generating arrivals to the Golden Gate Bridge.

d. Finite population. A frame could be constructed by obtaining a listing of students enrolled in the course from the professor.

e. Infinite population. Sampling from a process. The process is one of generating orders for the mail-order firm.

11. a. 

b. 

= (-4)2 + (-1)2 + 12 (-2)2 + 12 + 52 = 48

*s* = 

12. a. = 75/150 = .50

b. = 55/150 = .3667

13. a. 

b.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  | 94 | +1 | 1 |
|  | 100 | +7 | 49 |
|  | 85 | -8 | 64 |
|  | 94 | +1 | 1 |
|  | 92 | -1 | 1 |
| Totals | 465 | 0 | 116 |



14. a. Eighteen of the 40 funds in the sample are load funds. Our point estimate is



b. Six of the 40 funds in the sample are high risk funds. Our point estimate is



c. The below average fund ratings are low and very low. Twelve of the funds have a rating of low and 6 have a rating of very low. Our point estimate is



15. a. 

b. 

16. a. The sampled population is U. S. adults that are 50 years of age or older.

b. We would use the sample proportion for the estimate of the population proportion.



c. The sample proportion for this issue is .74 and the sample size is 426.

The number of respondents citing education as “very important” is (.74)426 = 315.

d. We would use the sample proportion for the estimate of the population proportion.



e. The inferences in parts (b) and (d) are being made about the population of U.S. adults who are age 50 or older. So, the population of U.S. adults who are age 50 or older is the target population. The target population is the same as the sampled population. If the sampled population was restricted to members of AARP who were 50 years of age or older, the sampled population would not be the same as the target population. The inferences made in parts (b) and (d) would only be valid if the population of AARP members age 50 or older was representative of the U.S. population of adults age 50 and over.

17. a. 409/999 = .41

b. 299/999 = .30

c. 291/999 = .29

d. The sampled population is all subscribers to the *American Association of Individual Investors Journal.* This is also the target population for the inferences made in parts (a), (b), and (c). There is no statistical basis for making inferences to a target population of all investors. That is not the group from which the sample is drawn.

18. a. 

b. 

c. Normal with *E* () = 200 and  = 5

d. It shows the probability distribution of all possible sample means that can be observed with random samples of size 100. This distribution can be used to compute the probability that  is within a specified  from **

19. a. The sampling distribution is normal with

*E* = ** = 200



For 5, 

Using Standard Normal Probability Table:

At = 205,  = .8413

At = 195,  = .1587

= .8413 - .1587 = .6826

b. For  10, 

Using Standard Normal Probability Table:

At = 210,  = .9772

At = 190,  = .0228

= .9772 - .0228 = .9544

20. 









The standard error of the mean decreases as the sample size increases.

21. a. 

b. *n* / *N* = 50 / 50,000 = .001

Use 

c. *n* / *N* = 50 / 5000 = .01

Use 

d. *n* / *N* = 50 / 500 = .10



Note: Only case (d) where *n* /*N* = .10 requires the use of the finite population correction factor.

22. a. *E*() = 51,800 and 







The normal distribution foris based on the Central Limit Theorem.

b. For *n* = 120, *E* () remains $51,800 and the sampling distribution of  can still be approximated by a normal distribution. However,  is reduced to = 365.15.

c. As the sample size is increased, the standard error of the mean, , is reduced. This appears logical from the point of view that larger samples should tend to provide sample means that are closer to the population mean. Thus, the variability in the sample mean, measured in terms of, should decrease as the sample size is increased.

23. a. With a sample of size 60 

At = 52,300, 

*P*(≤ 52,300) = *P*(*z* ≤ .97) = .8340

At = 51,300, 

*P*(< 51,300) = *P*(*z* < -.97) = .1660

*P*(51,300 ≤≤ 52,300) = .8340 - .1660 = .6680

b. 

At = 52,300, 

*P*(≤ 52,300) = *P*(*z* ≤ 1.37) = .9147

At = 51,300, 

*P*(< 51,300) = *P*(*z* < -1.37) = .0853

*P*(51,300 ≤ ≤ 52,300) = .9147 - .0853 = .8294

24. a. Normal distribution, 



b. Within 1 week means 16.5 ≤≤ 18.5

At = 18.5,  *P*(*z* ≤ 1.75) = .9599

At = 16.5, *z* = -1.75. *P*(*z* < -1.75) = .0401

So *P*(16.5 ≤ ≤ 18.5) = .9599 - .0401 = .9198

c. Within 1/2 week means 17.0 ≤≤ 18.0

At = 18.0,  *P*(*z* ≤ .88) = .8106

At = 17.0, *z* = -.88 *P*(*z* < -.88) = .1894

*P*(17.0 ≤≤ 18.0) = .8106 - .1894 = .6212

25.  This value for the standard error can be used for parts (a) and (b) below.

a.  *P*(*z* ≤ .95) = .8289

 *P*(*z* < -.95) = .1711

probability = .8289 - .1711 =.6578

b.  *P*(*z* ≤ .95) = .8289

 *P*(*z* < -.95) = .1711

probability = .8289 - .1711 =.6578

The probability of being within 10 of the mean on the Mathematics portion of the test is exactly the same as the probability of being within 10 on the Critical Reading portion of the SAT. This is because the standard error is the same in both cases. The fact that the means differ does not affect the probability calculation.

c.  The standard error is smaller here because the sample size is larger.

 *P*(*z* ≤ 1.00) = .8413

 *P*(*z* < -1.00) = .1587

probability = .8413 - .1587 =.6826

The probability is larger here than it is in parts (a) and (b) because the larger sample size has made the standard error smaller.

26. a. 

Within 25 means - 939 must be between -25 and +25.

The *z* value for - 939 = -25 is just the negative of the *z* value for - 939 = 25. So we just show the computation of *z* for - 939 = 25.

*n* = 30  *P*(-.56 ≤ *z* ≤ .56) = .7123 - .2877 = .4246

*n* = 50  *P*(-.72 ≤ *z* ≤ .72) = .7642 - .2358 = .5284

*n* = 100  *P*(-1.02 ≤ *z* ≤ 1.02) = .8461 - .1539 = .6922

*n* = 400  *P*(-2.04 ≤ *z* ≤ 2.04) = .9793 - .0207 = .9586

b. A larger sample increases the probability that the sample mean will be within a specified distance of the population mean. In the automobile insurance example, the probability of being within 25 of ** ranges from .4246 for a sample of size 30 to .9586 for a sample of size 400.

27. a. 

At = 22.18,  *P*(*z* ≤ 1.54) = .9382

At = 21.18, *z* = -1.54

*P*(*z* < -1.54) = .0618, thus

*P*(21.18 ≤≤ 22.18) = .9382 - .0618 = .8764

b. 

At = 19.30,  *P*(*z* ≤ 1.72) = .9573

At = 18.30, *z* = -1.72, *P*(*z* < -1.72) = .0427, thus

*P*(18.30 ≤ ≤ 19.30) = .9573 - .0427 = .9146

1. In part (b) we have a higher probability of obtaining a sample mean within $.50 of the population mean because the standard error for female graduates (.2899) is smaller than the standard error for male graduates (.3253).

d. With *n* = 120, 

At = 18.50, 

*P*(< 18.50) = *P*(*z* < -1.60) = .0548

28. a. This is a graph of a normal distribution with = 95 and



b. Within 3 strokes means 92 ≤ ≤ 98

 

*P*(92 ≤ ≤ 98) = *P*(-1.17 ≤ *z* ≤ 1.17) = .8790 - .1210 = .7580

The probability the sample means will be within 3 strokes of the population mean of 95 is .7580.

c. 

Within 3 strokes means 103 ≤ ≤ 109

 

*P*(103 ≤ ≤ 109) = *P*(-1.44 ≤ *z* ≤ 1.44) = .9251 - .0749 = .8502

The probability the sample means will be within 3 strokes of the population mean of 106 is .8502.

1. The probability of being within 3 strokes for female golfers is higher because the sample size is larger.

29. *μ* = 183 *σ* = 50

a. *n* = 30 Within 8 means 



*P*(175 ≤ ≤ 191) = *P*(-.88 ≤ *z* ≤ .88) = .8106 - .1894 = .6212

b. *n* = 50 Within 8 means 



*P*(175 ≤ ≤ 191) = *P*(-1.13 ≤ *z* ≤ 1.13) = .8708 - .1292 = .7416

c. *n* = 100 Within 8 means 



*P*(175 ≤ ≤ 191) = *P*(-1.60 ≤ *z* ≤ 1.60) = .9452 - .0548 = .8904

1. None of the sample sizes in parts (a), (b), and (c) are large enough. The sample size will need to

be greater than *n* = 100, which was used in part (c).

30. a. *n* / *N* = 40 / 4000 = .01 < .05; therefore, the finite population correction factor is not necessary.

b. With the finite population correction factor



Without the finite population correction factor



Including the finite population correction factor provides only a slightly different value for than when the correction factor is not used.

c.  *P*(*z* ≤ 1.54) = .9382

*P*(*z* < -1.54) = .0618

Probability = .9382 - .0618 = .8764

31. a. *E*() = *p* = .40

b. 

c. Normal distribution with *E*() = .40 and  = .0490

d. It shows the probability distribution for the sample proportion .

32. a. *E*() = .40



Within ± .03 means .37 ≤≤ .43

 *P*(*z* ≤ .87) = .8078

*P*(*z* < -.87) = .1922

*P*(.37 ≤≤ .43) = .8078 - .1922 = .6156

b.  *P*(*z* ≤ 1.44) = .9251

*P*(*z* < -1.44) = .0749

*P*(.35 ≤≤ .45) = .9251 - .0749 = .8502

33. 









The standard error of the proportion,decreases as *n* increases

34. a. 

Within ± .04 means .26 ≤≤ .34

 *P*(*z* ≤ .87) = .8078

*P*(*z* < -.87) = .1922

*P*(.26 ≤≤ .34) = .8078 - .1922 = .6156

b. 

 *P*(*z* ≤ 1.23) = .8907

*P*(*z* < -1.23) = .1093

*P*(.26 ≤≤ .34) = .8907 - .1093 = .7814

c. 

 *P*(*z* ≤ 1.95) = .9744

*P*(*z* < -1.95) = .0256

*P*(.26 ≤≤ .34) = .9744 - .0256 = .9488

d. 

 *P*(*z* ≤ 2.76) = .9971

*P*(*z* < -2.76) = .0029

*P*(.26 ≤≤ .34) = .9971 - .0029 = .9942

e. With a larger sample, there is a higher probability  will be within  .04 of the population proportion *p*.

35. a.

The normal distribution is appropriate because *np* = 100(.30) = 30 and *n*(1 - *p*) = 100(.70) = 70 are both greater than 5.







b. *P* (.20    .40) = ?

 *P*(*z* ≤ 2.18) = .9854

*P*(*z* < -2.18) = .0146

*P*(.20 ≤≤ .40) = .9854 - .0146 = .9708

c. *P* (.25    .35) = ?

 *P*(*z* ≤ 1.09) = .8621

*P*(*z* < -1.09) = .1379

*P*(.25 ≤≤ .35) = .8621 - .1379 = .7242

36. a. This is a graph of a normal distribution with a mean of = .55 and



b. Within ± .05 means .50 ≤ ≤ .60

 

*P*(.50 ≤ ≤ .60) = *P*(-1.42 ≤ *z* ≤ 1.42) = .9222 - .0778 = .8444

c. This is a graph of a normal distribution with a mean of = .45 and



d. 

Within ± .05 means .40 ≤ ≤ .50

 

*P*(.40 ≤ ≤ .50) = *P*(-1.42 ≤ *z* ≤ 1.42) = .9222 - .0778 = .8444

e. No, the probabilities are exactly the same. This is because , the standard error, and the width of the interval are the same in both cases. Notice the formula for computing the standard error. It involves . So whenever *p* = 1 - *p* the standard error will be the same. In part (b), *p* = .45 and 1 – *p* = .55. In part (d), *p* = .55 and 1 – *p* = .45.

f. For *n* = 400, 

Within ± .05 means .50 ≤ ≤ .60

 

*P*(.50 ≤ ≤ .60) = *P*(-2.01 ≤ *z* ≤ 2.01) = .9778 - .0222 = .9556

The probability is larger than in part (b). This is because the larger sample size has reduced the standard error from .0352 to .0249.

37. a. Normal distribution





b.  *P*(*z* ≤ 1.94) = .9838

*P*(*z* < -2.14) = .0162

*P*(.09 ≤≤ .15) = .9838 - .0162 = .9676

c.  *P*(*z* ≤ 1.07) = .8577

*P*(*z* < -1.07) = .1423

*P*(.105 ≤≤ .135) = .8577 - .1423 = .7154

38. a. It is a normal distribution with

*E*() = .42



b.  *P*(*z* ≤ 1.05) = .8531

*P*(*z* < -1.05) = .1469

*P*(.39 ≤≤ .44) = .8531 - .1469 = .7062

c.  *P*(*z* ≤ 1.75) = .9599

*P*(*z* < -1.75) = .0401

*P*(.39 ≤≤ .44) = .9599 - .0401 = .9198

d. The probabilities would increase. This is because the increase in the sample size makes the standard

error, , smaller.

39. a. Normal distribution with  and



b.  *P*(*z* ≤ 1.96) = .9750

*P*(*z* < -1.96) = .0250

*P*(.71 ≤ ≤ .79) = *P*(-1.96 ≤ *z* ≤ 1.96) = .9750 - .0275 = .9500

c. Normal distribution with  and



d.  *P*(*z* ≤ 1.31) = .9049

*P*(*z* < -1.31) = .0951

*P*(.71 ≤ ≤ .79) = *P*(-1.31 ≤ *z* ≤ 1.31) = .9049 - .0951 = .8098

e. The probability of the sample proportion being within .04 of the population mean was reduced from .9500 to .8098. So there is a gain in precision by increasing the sample size from 200 to 450. If the extra cost of using the larger sample size is not too great, we should probably do so.

40. a. *E* () = .76



Normal distribution because *np* = 400(.76) = 304 and *n*(1 - *p*) = 400(.24) = 96

b.  *P*(*z* ≤1.40) = .9192

*P*(*z* < -1.40) = .0808

*P*(.73 ≤ ≤ .79) = *P*(-1.40 ≤ *z* ≤ 1.40) = .9192 - .0808 = .8384

c. 

 *P*(*z* ≤ 1.92) = .9726

*P*(*z* < -1.92) = .0274

*P*(.73 ≤ ≤ .79) = *P*(-1.92 ≤ *z* ≤ 1.92) = .9726 - .0274 = .9452

41. a. *E*() = .17



Distribution is approximately normal because *np* = 800(.17) = 136 > 5

and *n*(1 – *p*) = 800(.83) = 664 > 5

b.  *P*(*z* ≤ 1.51) = .9345

*P*(*z* < -1.51) = .0655

*P*(.15 ≤ ≤ .19) = *P*(-1.51 ≤ *z* ≤ 1.51) = .9345 - .0655 = .8690

c. 

 *P*(*z* ≤ 2.13) = .9834

*P*(*z* < -2.13) = .0166

*P*(.15 ≤ ≤ .19) = *P*(-2.13 ≤ *z* ≤ 2.13) = .9834 - .0166 = .9668

42. The random numbers corresponding to the first seven universities selected are

122, 99, 25, 55, 115, 102, 61

The third, fourth and fifth columns of Table 7.1 were needed to find 7 random numbers of 133 or less without duplicate numbers.

Author’s note: The universities identified are: Clarkson U. (122), U. of Arizona (99), UCLA (25), U. of Maryland (55), U. of New Hampshire (115), Florida State U. (102), Clemson U. (61).

43. a. With *n* = 100, we can approximate the sampling distribution with a normal distribution having

*E*() = 8086



b.  *P*(*z* ≤ .80) = .7881

*P*(*z* < -.80) = .2119

*P*(7886 ≤ ≤ 8286) = *P*(-.80 ≤ *z* ≤ .80) = .7881 - .2119 = .5762

The probability that the sample mean will be within $200 of the population mean is .5762.

c. At 9000, 

*P*( ≥ 9000) = *P*(*z* ≥ 3.66) 0

Yes, the research firm should be questioned. A sample mean this large is extremely unlikely (almost 0 probability) if a simple random sample is taken from a population with a mean of $8086.

44. a. Normal distribution with

*E*= 406



b.  *P*(*z* ≤ 1.50) = .9332

*P*(*z* < -1.50) = .0668

*P*(391 ≤≤ 421) = *P*(-1.50 ≤ *z* ≤ 1.50) = .9332 - .0668 = .8664

c. At = 380, 

*P*( ≤ 380) = *P*(*z* ≤ -2.60) = .0047

Yes, this is an unusually low performing group of 64 stores. The probability of a sample mean annual sales per square foot of $380 or less is only .0047.

45. With *n* = 60 the central limit theorem allows us to conclude the sampling distribution is approximately normal.

a. This means 14 ≤≤ 16

At = 16,  *P*(*z* ≤ 1.94) = .9738

*P*(*z* < -1.94) = .0262

*P*(14 ≤≤ 16) = *P*(-1.94 ≤ *z* ≤ 1.94) = .9738 - .0262 = .9476

b. This means 14.25 ≤≤ 15.75

At = 15.75,  *P*(*z* ≤ 1.45) = .9265

*P*(*z* < -1.45) = .0735

*P*(14.25 ≤≤ 15.75) = *P*(-1.45 ≤ *z* ≤ 1.45) = .9265 - .0735 = .8530

46. *μ* = 27,175 *σ* = 7400

a. 

b. 

*P*(> 27,175) = *P*(*z* > 0) = .50

Note: This could have been answered easily without any calculations ; 27,175 is the expected value of the sampling distribution of .

c.  *P*(*z* ≤ 1.05) = .8531

*P*(*z* < -1.05) = .1469

*P*(26,175 ≤≤ 28,175) = *P*(-1.05 ≤ *z* ≤ 1.05) = .8531 - .1469 = .7062

d. 

 *P*(*z* ≤ 1.35) = .9115

*P*(*z* < -1.35) = .0885

*P*(26,175 ≤≤ 28,175) = *P*(-1.35 ≤ *z* ≤ 1.35) = .9115 - .0885 = .8230

47. a. 

*N* = 2000



*N* = 5000



*N* = 10,000



Note: With *n* / *N*  .05 for all three cases, common statistical practice would be to ignore the finite population correction factor and use for each case.

b. *N* = 2000

 *P*(*z* ≤ 1.24) = .8925

*P*(*z* < -1.24) = .1075

Probability = *P*(-1.24 ≤ *z* ≤ 1.24) = .8925 - .1075 = .7850

*N* = 5000

 *P*(*z* ≤ 1.23) = .8907

*P*(*z* < -1.23) = .1093

Probability = *P*(-1.23 ≤ *z* ≤ 1.23) = .8907 - .1093 = .7814

*N* = 10,000

 *P*(*z* ≤ 1.23) = .8907

*P*(*z* < -1.23) = .1093

Probability = *P*(-1.23 ≤ *z* ≤ 1.23) = .8907 - .1093 = .7814

All probabilities are approximately .78 indicating that a sample of size 50 will work well for all 3 firms.

48. a. 

= 500/20 = 25 and *n* = (25)2 = 625

b. For  25,

 *P*(*z* ≤ 1.25) = .8944

*P*(*z* < -1.25) = .1056

Probability = *P*(-1.25 ≤ *z* ≤ 1.25) = .8944 - .1056 = .7888

49. Sampling distribution of 





**





The area below= 2.1 must be 1 - .05 = .95. An area of .95 in the standard normal table shows

*z* = 1.645.

Thus,



Solve for 



50. *p* = .28

a. This is the graph of a normal distribution with *E*() = *p* = .28 and



b. Within ± .04 means .24 ≤ ≤ .32

 

*P*(.24 ≤ ≤ .32) = *P*(-1.38 ≤ *z* ≤ 1.38) = .9162 - .0838 = .8324

c. Within ± .02 means .26 ≤ ≤ .30

 

*P*(.26 ≤ ≤ .30) = *P*(-.69 ≤ *z* ≤ .69) = .7549 - .2451 = .5098

51. 

*P* (  .375) = ?

 *P*(*z* < -1.02) = .1539

*P* (  .375) = 1 - .1539 = .8461

52. a. 

Within ± .04 means .36 ≤ ≤ .44

 

*P*(.36 ≤ ≤ .44) = *P*(-1.59 ≤ *z* ≤ 1.59) = .9441 - .0559 = .8882

b. We want *P*(.45)



*P*(.45) = *P*(*z*  1.99) = 1 - .9767 = .0233

53. a. Normal distribution with *E* () = .15 and



b. *P* (.12   .18) = ?

 *P*(*z* ≤ 1.03) = .8485

*P*(*z* < -1.03) = .1515

*P*(.12 ≤≤ .18) = *P*(-1.03 ≤ *z* ≤ 1.03) = .8485 - .1515 =.6970

54. a. 

Solve for *n*



b. Normal distribution with *E*() = .25 and = .0625

(Note: (48)(.25) = 12 > 5, and (48)(.75) = 36 > 5)

c. *P* (  .30) = ?

 *P*(*z* ≤ .80) = .7881

*P* (  .30) = 1 - .7881 = .2119