CHAPTER 8

Elementary Sampling Theory

SAMPLING THEORY

Sampling theory is a study of relationships existing between a population and samples drawn from the population. It is of great value in many connections. For example, it is useful in *estimating* unknown population quantities (such as population mean and variance), often called *population parameters* or briefly *parameters*, from a knowledge of corresponding sample quantities (such as sample mean and variance), often called *sample statistics* or briefly *statistics*. Estimation problems are considered in Chapter 9.

Sampling theory is also useful in determining whether the observed differences between two samples are due to chance variation or whether they are really significant. Such questions arise, for example, in testing a new serum for use in treatment of a disease or in deciding whether one production process is better than another. Their answers involve the use of so-called *tests of significance and hypotheses* that are important in the *theory of decisions*. These are considered in Chapter 10.

In general, a study of the inferences made concerning a population by using samples drawn from it, together with indications of the accuracy of such inferences by using probability theory, is called *statistical inference*.

RANDOM SAMPLES AND RANDOM NUMBERS

In order that the conclusions of sampling theory and statistical inference be valid, samples must be chosen so as to be *representative* of a population. A study of sampling methods and of the related problems that arise is called the *design of the experiment*.

One way in which a representative sample may be obtained is by a process called *random sampling*, according to which each member of a population has an equal chance of being included in the sample. One technique for obtaining a random sample is to assign numbers to each member of the population, write these numbers on small pieces of paper, place them in an urn, and then draw numbers from the urn, being careful to mix thoroughly before each drawing. An alternative method is to use a table of *random numbers* (see Appendix IX) specially constructed for such purposes. See Problem 8.6.

SAMPLING WITH AND WITHOUT REPLACEMENT

If we draw a number from an urn, we have the choice of replacing or not replacing the number into the urn before a second drawing. In the first case the number can come up again and again, whereas in the second it can only come up once. Sampling where each member of the population may be chosen more than once is called *sampling with replacement*, while if each member cannot be chosen more than once it is called *sampling without replacement*.

Populations are either finite or infinite. If, for example, we draw 10 balls successively without replacement from an urn containing 100 balls, we are sampling from a finite population; while if we toss a coin 50 times and count the number of heads, we are sampling from an infinite population.

A finite population in which sampling is with replacement can theoretically be considered infinite, since any number of samples can be drawn without exhausting the population. For many practical purposes, sampling from a finite population that is very large can be considered to be sampling from an infinite population.

SAMPLING DISTRIBUTIONS

Consider all possible samples of size N that can be drawn from a given population (either with or without replacement). For each sample, we can compute a statistic (such as the mean and the standard deviation) that will vary from sample to sample. In this manner we obtain a distribution of the statistic that is called its *sampling distribution*.

If, for example, the particular statistic used is the sample mean, then the distribution is called the *sampling distribution of means*, or the *sampling distribution of the mean*. Similarly, we could have sampling distributions of standard deviations, variances, medians, proportions, etc.

For each sampling distribution, we can compute the mean, standard deviation, etc. Thus we can speak of the mean and standard deviation of the sampling distribution of means, etc.

SAMPLING DISTRIBUTION OF MEANS

Suppose that all possible samples of size N are drawn without replacement from a finite population of size $N_p > N$. If we denote the mean and standard deviation of the sampling distribution of means by $\mu_{\bar{X}}$ and $\sigma_{\bar{X}}$ and the population mean and standard deviation by μ and σ, respectively, then

$$\mu_{\bar{X}} = \mu \qquad \text{and} \qquad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}\sqrt{\frac{N_p - N}{N_p - 1}} \tag{1}$$

If the population is infinite or if sampling is with replacement, the above results reduce to

$$\mu_{\bar{X}} = \mu \qquad \text{and} \qquad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}} \tag{2}$$

For large values of N ($N \geq 30$), the sampling distribution of means is approximately a normal distribution with mean $\mu_{\bar{X}}$ and standard deviation $\sigma_{\bar{X}}$, irrespective of the population (so long as the population mean and variance are finite and the population size is at least twice the sample size). This result for an infinite population is a special case of the *central limit theorem* of advanced probability theory, which shows that the accuracy of the approximation improves as N gets larger. This is sometimes indicated by saying that the sampling distribution is *asymptotically normal*.

In case the population is normally distributed, the sampling distribution of means is also normally distributed even for small values of N (i.e., $N < 30$).

SAMPLING DISTRIBUTION OF PROPORTIONS

Suppose that a population is infinite and that the probability of occurrence of an event (called its success) is p, while the probability of nonoccurrence of the event is $q = 1 - p$. For example, the population may be all possible tosses of a fair coin in which the probability of the event "heads" is $p = \frac{1}{2}$. Consider all possible samples of size N drawn from this population, and for each sample determine the proportion P of successes. In the case of the coin, P would be the proportion of heads turning up in N tosses. We thus obtain a *sampling distribution of proportions* whose mean μ_P and standard deviation σ_P are given by

$$\mu_P = p \qquad \text{and} \qquad \sigma_P = \sqrt{\frac{pq}{N}} = \sqrt{\frac{p(1-p)}{N}} \tag{3}$$

which can be obtained from equations (*2*) by placing $\mu = p$ and $\sigma = \sqrt{pq}$. For large values of N ($N \geq 30$), the sampling distribution is very closely normally distributed. Note that the population is *binomially distributed.*

Equations (*3*) are also valid for a finite population in which sampling is with replacement. For finite populations in which sampling is without replacement, equations (*3*) are replaced by equations (*1*) with $\mu = p$ and $\sigma = \sqrt{pq}$.

Note that equations (*3*) are obtained most easily by dividing the mean and standard deviation (Np and $\sqrt{Npq}$) of the binomial distribution by N (see Chapter 7).

SAMPLING DISTRIBUTIONS OF DIFFERENCES AND SUMS

Suppose that we are given two populations. For each sample of size N_1 drawn from the first population, let us compute a statistic S_1; this yields a sampling distribution for the statistic S_1, whose mean and standard deviation we denote by μ_{S1} and σ_{S1}, respectively. Similarly, for each sample of size N_2 drawn from the second population, let us compute a statistic S_2; this yields a sampling distribution for the statistic S_2, whose mean and standard deviation are denoted by μ_{S2} and σ_{S2}. From all possible combinations of these samples from the two populations we can obtain a distribution of the differences, $S_1 - S_2$, which is called the *sampling distribution of differences of the statistics*. The mean and standard deviation of this sampling distribution, denoted respectively by μ_{S1-S2} and σ_{S1-S2}, are given by

$$\mu_{S1-S2} = \mu_{S1} - \mu_{S2} \qquad \text{and} \qquad \sigma_{S1-S2} = \sqrt{\sigma_{S1}^2 + \sigma_{S2}^2} \tag{4}$$

provided that the samples chosen do not in any way depend on each other (i.e., the samples are *independent*).

If S_1, and S_2 are the sample means from the two populations—which means we denote by $\bar{X}_1$ and $\bar{X}_2$, respectively—then the sampling distribution of the differences of means is given for infinite populations with means and standard deviations (μ_1, σ_1) and (μ_2, σ_2), respectively, by

$$\mu_{\bar{X}1-\bar{X}2} = \mu_{\bar{X}1} - \mu_{\bar{X}2} = \mu_1 - \mu_2 \qquad \text{and} \qquad \sigma_{\bar{X}1-X2} = \sqrt{\sigma_{\bar{X}1}^2 + \sigma_{\bar{X}2}^2} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} \tag{5}$$

using equations (*2*). The result also holds for finite populations if sampling is with replacement. Similar results can be obtained for finite populations in which sampling is without replacement by using equations (*1*).

Table 8.1 Standard Error for Sampling Distributions

Sampling Distribution	Standard Error	Special Remarks
Means	$\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{N}}$	This is true for large or small samples. The sampling distribution of means is very nearly normal for $N \geq 30$ even when the population is nonnormal. $\mu_{\bar{X}} = \mu$, the population mean, in all cases.
Proportions	$\sigma_P = \sqrt{\dfrac{p(1-p)}{N}} = \sqrt{\dfrac{pq}{N}}$	The remarks made for means apply here as well. $\mu_P = p$ in all cases.
Standard deviations	(*1*) $\sigma_s = \dfrac{\sigma}{\sqrt{2N}}$ (*2*) $\sigma_s = \sqrt{\dfrac{\mu_4 - \mu_2^2}{4N\mu_2}}$	For $N \geq 100$, the sampling distribution of s is very nearly normal. σ_s is given by (*1*) only if the population is normal (or approximately normal). If the population is nonnormal, (*2*) can be used. Note that (*2*) reduces to (*1*) when $\mu_2 = \sigma^2$ and $\mu_4 = 3\sigma^4$, which is true for normal populations. For $N \geq 100$, $\mu_s = \sigma$ very nearly
Medians	$\sigma_{\text{med}} = \sigma\sqrt{\dfrac{\pi}{2N}} = \dfrac{1.2533\sigma}{\sqrt{N}}$	For $N \geq 30$, the sampling distribution of the median is very nearly normal. The given result holds only if the population is normal (or approximately normal). $\mu_{\text{med}} = \mu$
First and third quartiles	$\sigma_{Q1} = \sigma_{Q3} = \dfrac{1.3626\sigma}{\sqrt{N}}$	The remarks made for medians apply here as well. μ_{Q1} and μ_{Q3} are very nearly equal to the first and third quartiles of the population. Note that $\sigma_{Q2} = \sigma_{\text{med}}$
Deciles	$\sigma_{D1} = \sigma_{D9} = \dfrac{1.7094\sigma}{\sqrt{N}}$ $\sigma_{D2} = \sigma_{D8} = \dfrac{1.4288\sigma}{\sqrt{N}}$ $\sigma_{D3} = \sigma_{D7} = \dfrac{1.3180\sigma}{\sqrt{N}}$ $\sigma_{D4} = \sigma_{D6} = \dfrac{1.2680\sigma}{\sqrt{N}}$	The remarks made for medians apply here as well. $\mu_{D1}, \mu_{D2}, \ldots$ are very nearly equal to the first, second, ... deciles of the population. Note that $\sigma_{D5} = \sigma_{\text{med}}$.
Semi-interquartile ranges	$\sigma_Q = \dfrac{0.7867\sigma}{\sqrt{N}}$	The remarks made for medians apply here as well. μ_Q is very nearly equal to the population semi-interquartile range
Variances	(*1*) $\sigma_{S^2} = \sigma^2\sqrt{\dfrac{2}{N}}$ (*2*) $\sigma_{S^2} = \sqrt{\dfrac{\mu_4 - \dfrac{N-3}{N-1}\mu_2^2}{N}}$	The remarks made for standard deviation apply here as well. Note that (*2*) yields (*1*) in the case that the population is normal $\mu_{S^2} = \sigma^2(N-1)/N$, which is very nearly σ^2 for large N.
Coefficients of variation	$\sigma_V = \dfrac{v}{\sqrt{2N}}\sqrt{1 + 2v^2}$	Here $v = \sigma/\mu$ is the population coefficient of variation. The given result holds for normal (or nearly normal) populations and $N \geq 100$.

Corresponding results can be obtained for the sampling distributions of differences of proportions from two binomially distributed populations with parameters (p_1, q_1) and (p_2, q_2), respectively. In this case S_1 and S_2 correspond to the proportion of successes, P_1 and P_2, and equations (4) yield the results

$$\mu_{P1-P2} = \mu_{P1} - \mu_{P2} = p_1 - p_2 \qquad \text{and} \qquad \sigma_{P1-P2} = \sqrt{\sigma_{P1}^2 + \sigma_{P2}^2} = \sqrt{\frac{p_1q_1}{N_1} + \frac{p_2q_2}{N_2}} \tag{6}$$

If N_1, and N_2 are large $(N_1, N_2 \geq 30)$, the sampling distributions of differences of means or proportions are very closely normally distributed.

It is sometimes useful to speak of the *sampling distribution of the sum of statistics*. The mean and standard deviation of this distribution are given by

$$\mu_{S1+S2} = \mu_{S1} + \mu_{S2} \qquad \text{and} \qquad \sigma_{S1+S2} = \sqrt{\sigma_{S1}^2 + \sigma_{S2}^2} \tag{7}$$

assuming that the samples are *independent*.

STANDARD ERRORS

The standard deviation of a sampling distribution of a statistic is often called its *standard error*. Table 8.1 lists standard errors of sampling distributions for various statistics under the conditions of random sampling from an infinite (or very large) population or of sampling with replacement from a finite population. Also listed are special remarks giving conditions under which results are valid and other pertinent statements.

The quantities μ, σ, p, μ_r and $\bar{X}$, s, P, m_r denote, respectively, the population and sample means, standard deviations, proportions, and rth moments about the mean.

It is noted that if the sample size N is large enough, the sampling distributions are normal or nearly normal. For this reason, the methods are known as *large sampling methods*. When $N < 30$, samples are called *small*. The theory of *small* samples, or *exact sampling theory* as it is sometimes called, is treated in Chapter 11.

When population parameters such as σ, p, or μ_r, are unknown, they may be estimated closely by their corresponding sample statistics namely, s (or $\hat{s} = \sqrt{N/(N-1)}s$), P, and m_r—if the samples are large enough.

SOFTWARE DEMONSTRATION OF ELEMENTARY SAMPLING THEORY

EXAMPLE 1.

A large population has the following random variable defined on it. X represents the number of computers per household and X is uniformly distributed, that is, $p(x) = 0.25$ for $x = 1$, 2, 3, and 4. In other words, 25% of the households have 1 computer, 25% have 2 computers, 25% have 3 computers, and 25% have 4 computers. The mean value of X is $\mu = \Sigma xp(x) = 0.25 + 0.5 + 0.75 + 1 = 2.5$. The variance of X is $\sigma^2 = \Sigma x^2 p(x) - \mu^2 = 0.25 + 1 + 2.25 + 4 - 6.25 = 1.25$. We say that the mean number of computers per household is 2.5 and the variance of the number of computers is 1.25 per household.

EXAMPLE 2.

MINITAB may be used to list all samples of two households taken with replacement. The worksheet would be as in Table 8.2. The 16 samples are shown in C1 and C2 and the mean for each sample in C3. Because the population is uniformly distributed, each sample mean has probability 1/16. Summarizing, the probability distribution is given in C4 and C5.

Note that $\mu_{\bar{x}} = \Sigma \bar{x} p(\bar{x}) = 1(0.0625) + 1.5(0.1250) + \cdots + 4(0.0625) = 2.5$. We see that $\mu_{\bar{x}} = \mu$. Also, $\sigma_{\bar{x}}^2 = \Sigma \bar{x}^2 p(\bar{x}) - \mu_{\bar{x}}^2 = 1(0.0625) + 2.25(0.1250) + \cdots + 16(0.0625) - 6.25 = 0.625$ which gives $\sigma_{\bar{x}}^2 = (\sigma^2/2)$. If MINITAB is used to draw the graph of the probability distribution of xbar, the result shown in Fig. 8-1 is obtained. (Note that $\bar{X}$ and xbar are used interchangeably.)

Table 8.2

C1 household1	C2 household2	C3 mean	C4 xbar	C5 p(xbar)
1	1	1.0	1.0	0.0625
1	2	1.5	1.5	0.1250
1	3	2.0	2.0	0.1875
1	4	2.5	2.5	0.2500
2	1	1.5	3.0	0.1875
2	2	2.0	3.5	0.1250
2	3	2.5	4.0	0.0625
2	4	3.0		
3	1	2.0		
3	2	2.5		
3	3	3.0		
3	4	3.5		
4	1	2.5		
4	2	3.0		
4	3	3.5		
4	4	4.0		

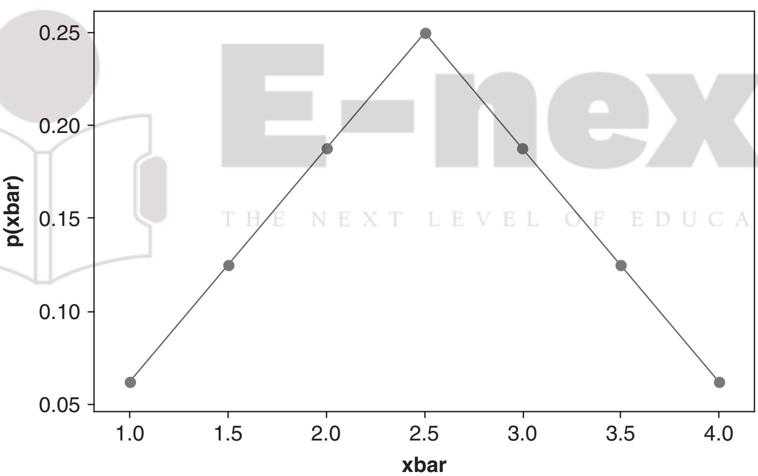

Fig. 8-1 Scatterplot of p(xbar) vs xbar.

Solved Problems

SAMPLING DISTRIBUTION OF MEANS

8.1 A population consists of the five numbers 2, 3, 6, 8, and 11. Consider all possible samples of size 2 that can be drawn with replacement from this population. Find (a) the mean of the population, (b) the standard deviation of the population, (c) the mean of the sampling distribution of means, and (d) the standard deviation of the sampling distribution of means (i.e., the standard error of means).

SOLUTION

(a)
$$\mu = \frac{2+3+6+8+11}{5} = \frac{30}{5} = 6.0$$

(b)
$$\sigma^2 = \frac{(2-6)^2 + (3-6)^2 + (6-6)^2 + (8-6)^2 + (11-6)^2}{5} = \frac{16+9+0+4+25}{5} = 10.8$$

and $\sigma = 3.29$.

(c) There are $5(5) = 25$ samples of size 2 that can be drawn with replacement (since any one of the five numbers on the first draw can be associated with any one of the five numbers on the second draw). These are

(2, 2)	(2, 3)	(2, 6)	(2, 8)	(2, 11)
(3, 2)	(3, 3)	(3, 6)	(3, 8)	(3, 11)
(6, 2)	(6, 3)	(6, 6)	(6, 8)	(6, 11)
(8, 2)	(8, 3)	(8, 6)	(8, 8)	(8, 11)
(11, 2)	(11, 3)	(11, 6)	(11, 8)	(11, 11)

The corresponding sample means are

2.0	2.5	4.0	5.0	6.5	
2.5	3.0	4.5	5.5	7.0	
4.0	4.5	6.0	7.0	8.5	(8)
5.0	5.5	7.0	8.0	9.5	
6.5	7.0	8.5	9.5	11.0	

and the mean of sampling distribution of means is

$$\mu_{\bar{X}} = \frac{\text{sum of all sample means in } (8)}{25} = \frac{150}{25} = 6.0$$

illustrating the fact that $\mu_{\bar{X}} = \mu$.

(d) The variance $\sigma^2_{\bar{X}}$ of the sampling distribution of means is obtained by subtracting the mean 6 from each number in (8), squaring the result, adding all 25 numbers thus obtained, and dividing by 25. The final result is $\sigma^2_{\bar{X}} = 135/25 = 5.40$, and thus $\sigma_{\bar{X}} = \sqrt{5.40} = 2.32$. This illustrates the fact that for finite populations involving sampling with replacement (or infinite populations), $\sigma^2_{\bar{X}} = \sigma^2/N$ since the right-hand side is $10.8/2 = 5.40$, agreeing with the above value.

8.2 Solve Problem 8.1 for the case that the sampling is without replacement.

SOLUTION

As in parts (a) and (b) of Problem 8.1, $\mu = 6$ and $\sigma = 3.29$.

(c) There are $\binom{5}{2} = 10$ samples of size 2 that can be drawn without replacement (this means that we draw one number and then another number different from the first) from the population: (2, 3), (2, 6), (2, 8), (2, 11), (3, 6), (3, 8), (3, 11), (6, 8), (6, 11), and (8, 11). The selection (2, 3), for example, is considered the same as (3, 2).

The corresponding sample means are 2.5, 4.0, 5.0, 6.5, 4.5, 5.5, 7.0, 7.0, 8.5, and 9.5, and the mean of sampling distribution of means is

$$\mu_{\bar{X}} = \frac{2.5 + 4.0 + 5.0 + 6.5 + 4.5 + 5.5 + 7.0 + 7.0 + 8.5 + 9.5}{10} = 6.0$$

illustrating the fact that $\mu_{\bar{X}} = \mu$.

(d) The variance of sampling distribution of means is

$$\sigma_{\bar{X}}^2 = \frac{(2.5 - 6.0)^2 + (4.0 - 6.0)^2 + (5.0 - 6.0)^2 + \cdots + (9.5 - 6.0)^2}{10} = 4.05$$

and $\sigma_{\bar{X}} = 2.01$. This illustrates

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{N}\left(\frac{N_p - N}{N_p - 1}\right)$$

since the right side equals

$$\frac{10.8}{2}\left(\frac{5-2}{5-1}\right) = 4.05$$

as obtained above.

8.3 Assume that the heights of 3000 male students at a university are normally distributed with mean 68.0 inches (in) and standard deviation 3.0 in. If 80 samples consisting of 25 students each are obtained, what would be the expected mean and standard deviation of the resulting sampling distribution of means if the sampling were done (a) with replacement and (b) without replacement?

SOLUTION

The numbers of samples of size 25 that could be obtained theoretically from a group of 3000 students with and without replacement are $(3000)^{25}$ and $\binom{3000}{25}$, which are much larger than 80. Hence we do not get a true sampling distribution of means, but only an *experimental* sampling distribution. Nevertheless, since the number of samples is large, there should be close agreement between the two sampling distributions. Hence the expected mean and standard deviation would be close to those of the theoretical distribution. Thus we have:

(a) $$\mu_{\bar{X}} = \mu = 68.0 \text{ in} \quad \text{and} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}} = \frac{3}{\sqrt{25}} = 0.6 \text{ in}$$

(b) $$\mu_{\bar{X}} = 68.0 \text{ in} \quad \text{and} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}\sqrt{\frac{N_p - N}{N_p - 1}} = \frac{3}{\sqrt{25}}\sqrt{\frac{3000-25}{3000-1}}$$

which is only very slightly less than 0.6 in and can therefore, for all practical purposes, be considered the same as in sampling with replacement.

Thus we would expect the experimental sampling distribution of means to be approximately normally distributed with mean 68.0 in and standard deviation 0.6 in.

8.4 In how many samples of Problem 8.3 would you expect to find the mean (a) between 66.8 and 68.3 in and (b) less than 66.4 in?

SOLUTION

The mean $\bar{X}$ of a sample in standard units is here given by

$$z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - 68.0}{0.6}$$

(a) $$66.8 \text{ in standard units} = \frac{66.8 - 68.0}{0.6} = -2.0$$

$$68.3 \text{ in standard units} = \frac{68.3 - 68.0}{0.6} = 0.5$$

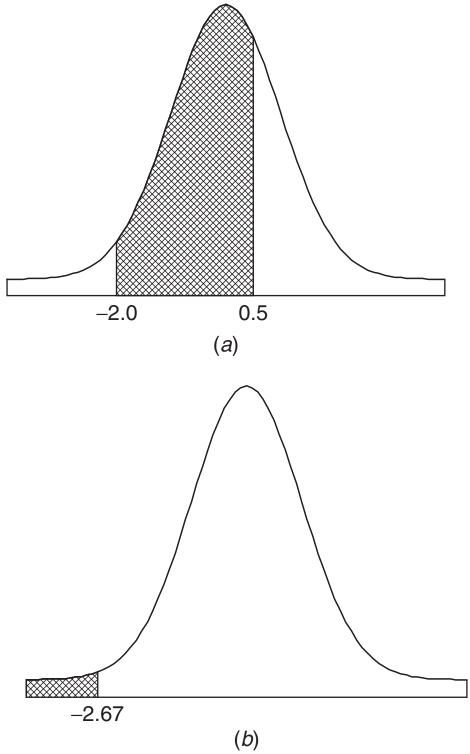

Fig. 8-2 Areas under the standard normal curve. (*a*) Standard normal curve showing the area between $z = -2$ and $z = 0.5$; (*b*) Standard normal curve showing the area to the left of $z = -2.67$.

As shown in Fig. 8-2(*a*),

Proportion of samples with means between 66.8 and 68.3 in

$$= (\text{area under normal curve between } z = -2.0 \text{ and } z = 0.5)$$
$$= (\text{area between } z = -2 \text{ and } z = 0) + (\text{area between } z = 0 \text{ and } z = 0.5)$$
$$= 0.4772 + 0.1915 = 0.6687$$

Thus the expected number of samples is $(80)(0.6687) = 53.496$, or 53.

(*b*)
$$66.4 \text{ in standard units} = \frac{66.4 - 68.0}{0.6} = -2.67$$

As shown in Fig. 8.2(*b*),

Proportion of samples with means less than 66.4 in $= (\text{area under normal curve to left of } z = -2.67)$

$$= (\text{area to left of } z = 0)$$
$$-(\text{area between } z = -2.67 \text{ and } z = 0)$$
$$= 0.5 - 0.4962 = 0.0038$$

Thus the expected number of samples is $(80)(0.0038) = 0.304$, or zero.

8.5 Five hundred ball bearings have a mean weight of 5.02 grams (g) and a standard deviation of 0.30 g. Find the probability that a random sample of 100 ball bearings chosen from this group will have a combined weight of (*a*) between 496 and 500 g and (*b*) more than 510 g.

SOLUTION

For the sample distribution of means, $\mu_{\bar{X}} = \mu = 5.02\,\text{g}$, and

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{N}}\sqrt{\frac{N_p - N}{N_p - 1}} = \frac{0.30}{\sqrt{100}}\sqrt{\frac{500-100}{500-1}} = 0.027\,\text{g}$$

(*a*) The combined weight will lie between 496 and 500 g if the mean weight of the 100 ball bearings lies between 4.96 and 5.00 g.

$$4.96 \text{ in standard units} = \frac{4.96 - 5.02}{0.0027} = -2.22$$

$$5.00 \text{ in standard units} = \frac{5.00 - 5.02}{0.027} = -0.74$$

As shown in Fig. 8-3(*a*),

$$\begin{aligned}\text{Required probability} &= (\text{area between } z = -2.22 \text{ and } z = -0.74)\\ &= (\text{area between } z = -2.22 \text{ and } z = 0) - (\text{area between } z = -0.74 \text{ and } z = 0)\\ &= 0.4868 - 0.2704 = 0.2164\end{aligned}$$

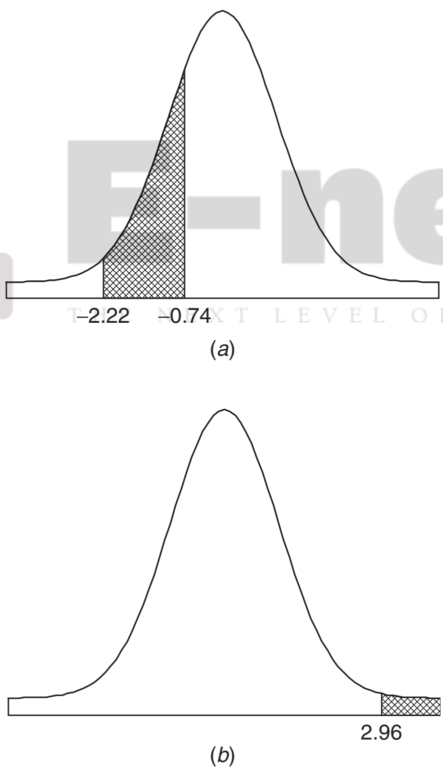

Fig. 8-3 Sample probabilities are found as areas under the standard normal curve. (*a*) Standard normal curve showing the area between $z = -2.22$ and $z = -0.74$; (*b*) Standard normal curve showing the area to the right of $z = 2.96$.

(*b*) The combined weight will exceed 510 g if the mean weight of the 100 bearings exceeds 5.10 g.

$$5.10 \text{ in standard units} = \frac{5.10 - 5.02}{0.027} = 2.96$$

As shown in Fig. 8-3(*b*),

$$\begin{aligned}\text{Required probability} &= (\text{area to right of } z = 2.96)\\ &= (\text{area to right of } z = 0) - (\text{area between } z = 0 \text{ and } z = 2.96)\\ &= 0.5 - 0.4985 = 0.0015\end{aligned}$$

Thus there are only 3 chances in 2000 of picking a sample of 100 ball bearings with a combined weight exceeding 510 g.

8.6 (*a*) Show how to select 30 random samples of 4 students each (with replacement) from Table 2.1 by using random numbers.

(*b*) Find the mean and standard deviation of the sampling distribution of means in part (*a*).

(*c*) Compare the results of part (*b*) with theoretical values, explaining any discrepancies.

SOLUTION

(*a*) Use two digits to number each of the 100 students: 00, 01, 02, ..., 99 (see Table 8.3). Thus the 5 students with heights 60–62 in are numbered 00–04, the 18 students with heights 63–65 in are numbered 05–22, etc. Each student number is called a *sampling number*.

Table 8.3

Height (in)	Frequency	Sampling Number
60–62	5	00–04
63–65	18	05–22
66–68	42	23–64
69–71	27	65–91
72–74	8	92–99

We now draw sampling numbers from the random-number table (Appendix IX). From the first line we find the sequence 51, 77, 27, 46, 40, etc., which we take as random sampling numbers, each of which yields the height of a particular student. Thus 51 corresponds to a student having height 66–68 in, which we take as 67 in (the class mark). Similarly, 77, 27, and 46 yield heights of 70, 67, and 67 in, respectively.

By this process we obtain Table 8.4, which shows the sample numbers drawn, the corresponding heights, and the mean height for each of 30 samples. It should be mentioned that although we have entered the random-number table on the first line, we could have started *anywhere* and chosen any specified pattern.

(*b*) Table 8.5 gives the frequency distribution of the sample mean heights obtained in part (*a*). This is a *sampling distribution of means*. The mean and the standard deviation are obtained as usual by the coding methods of Chapters 3 and 4:

$$\text{Mean} = A + c\bar{u} = A + \frac{c\sum fu}{N} = 67.00 + \frac{(0.75)(23)}{30} = 67.58 \text{ in}$$

$$\text{Standard deviation} = c\sqrt{\overline{u^2} - \bar{u}^2} = c\sqrt{\frac{\sum fu^2}{N} - \left(\frac{\sum fu}{N}\right)^2} = 0.75\sqrt{\frac{123}{30} - \left(\frac{23}{30}\right)^2} = 1.41 \text{ in}$$

(*c*) The theoretical mean of the sampling distribution of means, given by $\mu_{\bar{X}}$, should equal the population mean μ, which is 67.45 in (see Problem 3.22), in agreement with the value 67.58 in of part (*b*).

The theoretical standard deviation (standard error) of the sampling distribution of means, given by $\sigma_{\bar{X}}$, should equal $\sigma/\sqrt{N}$, where the population standard deviation $\sigma = 2.92$ in (see Problem 4.17) and the sample size $N = 4$. Since $\sigma/\sqrt{N} = 2.92/\sqrt{4} = 1.46$ in, we have agreement with the value 1.41 in of part (*b*). The discrepancies result from the fact that only 30 samples were selected and the sample size was small.

Table 8.4

Sample Number Drawn	Corresponding Height	Mean Height	Sample Number Drawn	Corresponding Height	Mean Height
1. 51, 77, 27, 46	67, 70, 67, 67	67.75	**16.** 11, 64, 55, 58	64, 67, 67, 67	66.25
2. 40, 42, 33, 12	67, 67, 67, 64	66.25	**17.** 70, 56, 97, 43	70, 67, 73, 67	69.25
3. 90, 44, 46, 62	70, 67, 67, 67	67.75	**18.** 74, 28, 93, 50	70, 67, 73, 67	69.25
4. 16, 28, 98, 93	64, 67, 73, 73	69.25	**19.** 79, 42, 71, 30	70, 67, 70, 67	68.50
5. 58, 20, 41, 86	67, 64, 67, 70	67.00	**20.** 58, 60, 21, 33	67, 67, 64, 67	66.25
6. 19, 64, 08, 70	64, 67, 64, 70	66.25	**21.** 75, 79, 74, 54	70, 70, 70, 67	69.25
7. 56, 24, 03, 32	67, 67, 61, 67	65.50	**22.** 06, 31, 04, 18	64, 67, 61, 64	64.00
8. 34, 91, 83, 58	67, 70, 70, 67	68.50	**23.** 67, 07, 12, 97	70, 64, 64, 73	67.75
9. 70, 65, 68, 21	70, 70, 70, 64	68.50	**24.** 31, 71, 69, 88	67, 70, 70, 70	69.25
10. 96, 02, 13, 87	73, 61, 64, 70	67.00	**25.** 11, 64, 21, 87	64, 67, 64, 70	66.25
11. 76, 10, 51, 08	70, 64, 67, 64	66.25	**26.** 03, 58, 57, 93	61, 67, 67, 73	67.00
12. 63, 97, 45, 39	67, 73, 67, 67	68.50	**27.** 53, 81, 93, 88	67, 70, 73, 70	70.00
13. 05, 81, 45, 93	64, 70, 67, 73	68.50	**28.** 23, 22, 96, 79	67, 64, 73, 70	68.50
14. 96, 01, 73, 52	73, 61, 70, 67	67.75	**29.** 98, 56, 59, 36	73, 67, 67, 67	68.50
15. 07, 82, 54, 24	64, 70, 67, 67	67.00	**30.** 08, 15, 08, 84	64, 64, 64, 70	65.50

Table 8.5

Sample Mean	Tally	f	u	fu	fu^2
64.00	/	1	-4	-4	16
64.75		0	-3	0	0
65.50	//	2	-2	-4	8
66.25	~~////~~ /	6	-1	-6	6
$A \rightarrow$ 67.00	////	4	0	0	0
67.75	////	4	1	4	4
68.50	~~////~~ //	7	2	14	28
69.25	~~////~~	5	3	15	45
70.00	/	1	4	4	16
		$\sum f = N = 30$		$\sum fu = 23$	$\sum fu^2 = 123$

SAMPLING DISTRIBUTION OF PROPORTIONS

8.7 Find the probability that in 120 tosses of a fair coin (*a*) less than 40% or more than 60% will be heads and (*b*) $\frac{5}{8}$ or more will be heads.

SOLUTION

First method

We consider the 120 tosses of the coin to be a sample from the infinite population of all possible tosses of the coin. In this population the probability of heads is $p = \frac{1}{2}$ and the probability of tails is $q = 1 - p = \frac{1}{2}$.

(*a*) We require the probability that the number of heads in 120 tosses will be less than 48 or more than 72. We proceed as in Chapter 7, using the normal approximation to the binomial. Since the number of heads is a discrete variable, we ask for the probability that the number of heads is less than 47.5 or greater than 72.5.

$$\mu = \text{expected number of heads} = Np = 120(\tfrac{1}{2}) = 60 \qquad \text{and} \qquad \sigma = \sqrt{Npq} = \sqrt{(120)(\tfrac{1}{2})(\tfrac{1}{2})} = 5.48$$

$$47.5 \text{ in standard units} = \frac{47.5 - 60}{5.48} = -2.28$$

$$72.5 \text{ in standard units} = \frac{72.5 - 60}{5.48} = 2.28$$

As shown in Fig. 8-4,

$$\begin{aligned}\text{Required probability} &= (\text{area to the left of } -2.28 \text{ plus area to the right of } 2.28)\\ &= (2(0.0113) = 0.0226)\end{aligned}$$

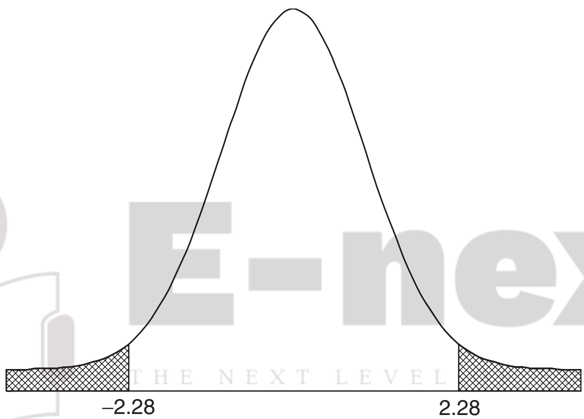

Fig. 8-4 Normal approximation to the binomial uses the standard normal curve.

Second method

$$\mu_P = p = \tfrac{1}{2} = 0.50 \qquad \sigma_P = \sqrt{\frac{pq}{N}} = \sqrt{\frac{(\frac{1}{2})(\frac{1}{2})}{120}} = 0.0456$$

$$40\% \text{ in standard units} = \frac{0.40 - 0.50}{0.0456} = -2.19$$

$$60\% \text{ in standard units} = \frac{0.60 - 0.50}{0.0456} = 2.19$$

$$\begin{aligned}\text{Required probability} &= (\text{area to the left of } -2.19 \text{ plus area to the right of } 2.19)\\ &= (2(0.0143) = 0.0286)\end{aligned}$$

Although this result is accurate to two significant figures, it does not agree exactly since we have not used the fact that the proportion is actually a discrete variable. To account for this, we subtract

$1/2N = 1/2(120)$ from 0.40 and add $1/2N = 1/2(120)$ to 0.60; thus, since $1/240 = 0.00417$, the required proportions in standard units are

$$\frac{0.40 - 0.00417 - 0.50}{0.0456} = -2.28 \qquad \text{and} \qquad \frac{0.60 + 0.00417 - 0.50}{0.0456} = 2.28$$

so that agreement with the first method is obtained.

Note that $(0.40 - 0.00417)$ and $(0.60 + 0.00417)$ correspond to the proportions 47.5/120 and 72.5/120 in the first method.

(*b*) Using the second method of part (*a*), we find that since $\frac{5}{8} = 0.6250$,

$$(0.6250 - 0.00417) \text{ in standard units} = \frac{0.6250 - 0.00417 - 0.50}{0.0456} = 2.65$$

$$\begin{aligned}\text{Required probability} &= (\text{area under normal curve to right of } z = 2.65)\\ &= (\text{area to right of } z = 0) - (\text{area between } z = 0 \text{ and } z = 2.65)\\ &= 0.5 - 0.4960 = 0.0040\end{aligned}$$

8.8 Each person of a group of 500 people tosses a fair coin 120 times. How many people should be expected to report that (*a*) between 40% and 60% of their tosses resulted in heads and (*b*) $\frac{5}{8}$ or more of their tosses resulted in heads?

SOLUTION

This problem is closely related to Problem 8.7. Here we consider 500 samples, of size 120 each, from the infinite population of all possible tosses of a coin.

(*a*) Part (*a*) of Problem 8.7 states that of all possible samples, each consisting of 120 tosses of a coin, we can expect to find 97.74% with a percentage of heads between 40% and 60%. In 500 samples we can thus expect to find about (97.74% of 500) = 489 samples with this property. It follows that about 489 people would be expected to report that their experiment resulted in between 40% and 60% heads.

It is interesting to note that $500 - 489 = 11$ people who would be expected to report that the percentage of heads was not between 40% and 60%. Such people might reasonably conclude that their coins were loaded even though they were fair. This type of error is an everpresent *risk* whenever we deal with probability.

(*b*) By reasoning as in part (*a*), we conclude that about $(500)(0.0040) = 2$ persons would report that $\frac{5}{8}$ or more of their tosses resulted in heads.

8.9 It has been found that 2% of the tools produced by a certain machine are defective. What is the probability that in a shipment of 400 such tools (*a*) 3% or more and (*b*) 2% or less will prove defective?

SOLUTION

$$\mu_P = p = 0.02 \qquad \text{and} \qquad \sigma_P = \sqrt{\frac{pq}{N}} = \sqrt{\frac{(0.02)(0.98)}{400}} = \frac{0.14}{20} = 0.007$$

(*a*) **First method**

Using the correction for discrete variables, $1/2N = 1/800 = 0.00125$, we have

$$(0.03 - 0.00125) \text{ in standard units} = \frac{0.03 - 0.00125 - 0.02}{0.007} = 1.25$$

$$\text{Required probability} = (\text{area under normal curve to right of } z = 1.25) = 0.1056$$

If we had not used the correction, we would have obtained 0.0764.

Another method

(3% of 400) = 12 defective tools. On a continuous basis 12 or more tools means 11.5 or more.

$$\bar{X} = (2\% \text{ of } 400) = 8 \qquad \text{and} \qquad \sigma = \sqrt{Npq} = \sqrt{(400)(0.02)(0.98)} = 2.8$$

Then, 11.5 in standard units $= (11.5 - 8)/2.8 = 1.25$, and as before the required probability is 0.1056.

(*b*)
$$(0.02 + 0.00125) \text{ in standard units} = \frac{0.02 + 0.00125 - 0.02}{0.007} = 0.18$$
$$\text{Required probability} = (\text{area under normal curve to left of } z = 0.18)$$
$$= 0.5000 + 0.0714 = 0.5714$$

If we had not used the correction, we would have obtained 0.5000. The second method of part (*a*) can also be used.

8.10 The election returns showed that a certain candidate received 46% of the votes. Determine the probability that a poll of (*a*) 200 and (*b*) 1000 people selected at random from the voting population would have shown a majority of votes in favor of the candidate.

SOLUTION

(*a*)
$$\mu_P = p = 0.46 \qquad \text{and} \qquad \sigma_P = \sqrt{\frac{pq}{N}} = \sqrt{\frac{(0.46)(0.54)}{200}} = 0.0352$$

Since $1/2N = 1/400 = 0.0025$, a majority is indicated in the sample if the proportion in favor of the candidate is $0.50 + 0.0025 = 0.5025$ or more. (This proportion can also be obtained by realizing that 101 or more indicates a majority, but as a continuous variable this is 100.5, and so the proportion is $100.5/200 = 0.5025$.)

$$0.5025 \text{ in standard units} = \frac{0.5025 - 0.46}{0.0352} = 1.21$$

$$\text{Required probability} = (\text{area under normal curve to right of } z = 1.21)$$
$$= 0.5000 - 0.3869 = 0.1131$$

(*b*)
$$\mu_p = p = 0.46 \qquad \text{and} \qquad \sigma_P = \sqrt{\frac{pq}{N}} = \sqrt{\frac{(0.46)(0.54)}{1000}} = 0.0158$$

$$0.5025 \text{ in standard units} = \frac{0.5025 - 0.46}{0.0158} = 2.69$$

$$\text{Required probability} = (\text{area under normal curve to right of } z = 2.69)$$
$$= 0.5000 - 0.4964 = 0.0036$$

SAMPLING DISTRIBUTIONS OF DIFFERENCES AND SUMS

8.11 Let U_1 be a variable that stands for any of the elements of the population 3, 7, 8 and U_2 be a variable that stands for any of the elements of the population 2, 4. Compute (*a*) μ_{U1}, (*b*) μ_{U2}, (*c*) μ_{U1-U2}, (*d*) σ_{U1}, (*e*) σ_{U2}, and (*f*) σ_{U1-U2}.

SOLUTION

(*a*) μ_{U1} = mean of population $U_1 = \frac{1}{3}(3 + 7 + 8) = 6$

(*b*) μ_{U2} = mean of population $U_2 = \frac{1}{2}(2 + 4) = 3$

(*c*) The population consisting of the differences of any member of U_1 and any member of U_2 is

$$\begin{matrix} 3-2 & 7-2 & 8-2 \\ 3-4 & 7-4 & 8-4 \end{matrix} \quad \text{or} \quad \begin{matrix} 1 & 5 & 6 \\ -1 & 3 & 4 \end{matrix}$$

Thus
$$\mu_{U1-U2} = \text{mean of } (U_1 - U_2) = \frac{1+5+6+(-1)+3+4}{6} = 3$$

This illustrates the general result $\mu_{U1-U2} = \mu_{U1} - \mu_{U2}$, as seen from parts (*a*) and (*b*).

(*d*)
$$\sigma^2_{U1} = \text{variance of population } U_1 = \frac{(3-6)^2 + (7-6)^2 + (8-6)^2}{3} = \frac{14}{3}$$

or
$$\sigma_{U1} = \sqrt{\frac{14}{3}}$$

(*e*)
$$\sigma^2_{U2} = \text{variance of population } U_2 = \frac{(2-3)^2 + (4-3)^2}{2} = 1 \qquad \text{or} \qquad \sigma_{U2} = 1$$

(*f*)
$$\sigma^2_{U1-U2} = \text{variance of population } (U_1 - U_2)$$
$$= \frac{(1-3)^2 + (5-3)^2 + (6-3)^2 + (-1-3)^2 + (3-3)^2 + (4-3)^2}{6} = \frac{17}{3}$$

or
$$\sigma_{U1-U2} = \sqrt{\frac{17}{3}}$$

This illustrates the general result for independent samples, $\sigma_{U1-U2} = \sqrt{\sigma^2_{U1} + \sigma^2_{U2}}$, as seen from parts (*d*) and (*e*).

8.12 The electric light bulbs of manufacturer *A* have a mean lifetime of 1400 hours (h) with a standard deviation of 200 h, while those of manufacturer *B* have a mean lifetime of 1200 h with a standard deviation of 100 h. If random samples of 125 bulbs of each brand are tested, what is the probability that the brand *A* bulbs will have a mean lifetime that is at least (*a*) 160 h and (*b*) 250 h more than the brand *B* bulbs?

SOLUTION

Let $\bar{X}_A$ and $\bar{X}_B$ denote the mean lifetimes of samples *A* and *B*, respectively. Then

$$\mu_{\bar{X}_A - \bar{X}_B} = \mu_{\bar{X}_A} - \mu_{\bar{X}_B} = 1400 - 1200 = 200\text{ h}$$

and
$$\sigma_{\bar{X}_A - \bar{X}_B} = \sqrt{\frac{\sigma^2_A}{N_A} + \frac{\sigma^2_B}{N_B}} = \sqrt{\frac{(100)^2}{125} + \frac{(200)^2}{125}} = 20\text{ h}$$

The standardized variable for the difference in means is

$$z = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_{\bar{X}_A - \bar{X}_B})}{\sigma_{\bar{X}_A - \bar{X}_B}} = \frac{(\bar{X}_A - \bar{X}_B) - 200}{20}$$

and is very closely normally distributed.

(*a*) The difference 160 h in standard units is $(160 - 200)/20 = -2$. Thus

$$\text{Required probability} = (\text{area under normal curve to right of } z = -2)$$
$$= 0.5000 + 0.4772 = 0.9772$$

(*b*) The difference 250 h in standard units is $(250 - 200)/20 = 2.50$. Thus

$$\text{Required probability} = (\text{area under normal curve to right of } z = 2.50)$$
$$= 0.5000 - 0.4938 = 0.0062$$

8.13 Ball bearings of a given brand weigh 0.50 g with a standard deviation of 0.02 g. What is the probability that two lots of 1000 ball bearings each will differ in weight by more than 2 g?

SOLUTION

Let $\bar{X}_1$ and $\bar{X}_2$ denote the mean weights of ball bearings in the two lots. Then

$$\mu_{\bar{X}_1-\bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = 0.50 - 0.50 = 0$$

and
$$\sigma_{\bar{X}_1-X_2} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} = \sqrt{\frac{(0.02)^2}{1000} + \frac{(0.02)^2}{1000}} = 0.000895$$

The standardized variable for the difference in means is

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - 0}{0.000895}$$

and is very closely normally distributed.

A difference of 2 g in the lots is equivalent to a difference of $2/1000 = 0.002$ g in the means. This can occur either if $\bar{X}_1 - \bar{X}_2 \geq 0.002$ or $\bar{X}_1 - X_2 \leq -0.002$; that is,

$$z \geq \frac{0.002 - 0}{0.000895} = 2.23 \qquad \text{or} \qquad z \leq \frac{-0.002 - 0}{0.000895} = -2.23$$

Then $\Pr\{z \geq 2.23 \text{ or } z \leq -2.23\} = \Pr\{z \geq 2.23\} + \Pr\{z \leq -2.23\} = 2(0.5000 - 0.4871) = 0.0258$.

8.14 A and B play a game of "heads and tails," each tossing 50 coins. A will win the game if she tosses 5 or more heads than B; otherwise, B wins. Determine the odds against A winning any particular game.

SOLUTION

Let P_A and P_B denote the proportion of heads obtained by A and B. If we assume that the coins are all fair, the probability p of heads is $\frac{1}{2}$. Then

$$\mu_{P_A-P_B} = \mu_{P_A} - \mu_{P_B} = 0$$

and

$$\sigma_{P_A-P_B} = \sqrt{\sigma_{P_A}^2 + \sigma_{P_B}^2} = \sqrt{\frac{pq}{N_A} + \frac{pq}{N_B}} = \sqrt{\frac{2(\frac{1}{2})(\frac{1}{2})}{50}} = 0.10$$

The standardized variable for the difference in proportions is $z = (P_A - P_B - 0)/0.10$.

On a continuous-variable basis, 5 or more heads means 4.5 or more heads, so that the difference in proportions should be $4.5/50 = 0.09$ or more; that is, z is greater than or equal to $(0.09 - 0)/0.10 = 0.9$ (or $z \geq 0.9$). The probability of this is the area under the normal curve to the right of $z = 0.9$, which is $(0.5000 - 0.3159) = 0.1841$.

Thus the odds against A winning are $(1 - 0.1841):0.1841 = 0.8159:0.1841$, or 4.43 to 1.

8.15 Two distances are measured as 27.3 centimeters (cm) and 15.6 cm with standard deviations (standard errors) of 0.16 cm and 0.08 cm, respectively. Determine the mean and standard deviation of (*a*) the sum and (*b*) the difference of the distances.

SOLUTION

If the distances are denoted by D_1 and D_2, then:

(*a*)
$$\mu_{D1+D2} = \mu_{D1} + \mu_{D2} = 27.3 + 15.6 = 42.9 \text{ cm}$$
$$\sigma_{D1+D2} = \sqrt{\sigma_{D1}^2 + \sigma_{D2}^2} = \sqrt{(0.16)^2 + (0.08)^2} = 0.18 \text{ cm}$$

(*b*)
$$\mu_{D1-D2} = \mu_{D1} - \mu_{D2} = 27.3 - 15.6 = 11.7 \text{ cm}$$
$$\sigma_{D1-D2} = \sqrt{\sigma_{D1}^2 + \sigma_{D2}^2} = \sqrt{(0.16)^2 + (0.08)^2} = 0.18 \text{ cm}$$

8.16 A certain type of electric light bulb has a mean lifetime of 1500 h and a standard deviation of 150 h. Three bulbs are connected so that when one burns out, another will go on. Assuming that the lifetimes are normally distributed, what is the probability that lighting will take place for (*a*) at least 5000 h and (*b*) at most 4200 h?

SOLUTION

Assume the lifetimes to be L_1, L_2, and L_3. Then

$$\mu_{L1+L2+L3} = \mu_{L1} + \mu_{L2} + \mu_{L3} = 1500 + 1500 + 1500 = 4500\text{ h}$$

$$\sigma_{L1+L2+L3} = \sqrt{\sigma_{L1}^2 + \sigma_{L2}^2 + \sigma_{L3}^2} = \sqrt{3(150)^2} = 260\text{ h}$$

(*a*) $$50000\text{ h in standard units} = \frac{5000 - 4500}{260} = 1.92$$

$$\text{Required probability} = (\text{area under normal curve to right of } z = 1.92)$$
$$= 0.5000 - 0.4726 = 0.0274$$

(*b*) $$4200\text{ h in standard units} = \frac{4200 - 4500}{260} = -1.15$$

$$\text{Required probability} = (\text{area under normal curve to left of } z = -1.15)$$
$$= 0.5000 - 0.3749 = 0.1251$$

SOFTWARE DEMONSTRATION OF ELEMENTARY SAMPLING THEORY

8.17 Midwestern University has 1/3 of its students taking 9 credit hours, 1/3 taking 12 credit hours, and 1/3 taking 15 credit hours. If X represents the credit hours a student is taking, the distribution of X is $p(x) = 1/3$ for $x = 9$, 12, and 15. Find the mean and variance of X. What type of distribution does X have?

SOLUTION

The mean of X is $\mu = \sum xp(x) = 9(1/3) + 12(1/3) + 15(1/3) = 12$. The variance of X is $\sigma^2 = \sum x^2 p(x) - \mu^2 = 81(1/3) + 144(1/3) + 225(1/3) - 144 = 150 - 144 = 6$. The distribution of X is uniform.

8.18 List all samples of size $n = 2$ that are possible (with replacement) from the population in Problem 8.17. Use the chart wizard of EXCEL to plot the sampling distribution of the mean to show that $\mu_{\bar{x}} = \mu$, and show that $\sigma_{\bar{x}}^2 = \sigma^2/2$.

SOLUTION

A	B	C	D	E	F	G
		mean	xbar	p(xbar)	xbar × p(xbar)	xbar^2 × p(xbar)
9	9	9	9	0.111111	1	9
9	12	10.5	10.5	0.222222	2.333333333	24.5
9	15	12	12	0.333333	4	48
12	9	10.5	13.5	0.222222	3	40.5
12	12	12	15	0.111111	1.666666667	25
12	15	13.5			12	147
15	9	12				
15	12	13.5				
15	15	15				

The EXCEL worksheet shows the possible sample values in A and B and the mean in C. The sampling distribution of xbar is built and displayed in D and E. In C2, the function =AVERAGE(A2:B2) is entered and a click-and-drag is performed from C2 to C10. Because the population is uniform each sample has probability 1/9 of being selected. The sample mean is represented by xbar. The mean of the sample mean is $\mu_{\bar{x}} = \Sigma \bar{x} p(\bar{x})$and is computed in F2 through F6. The function =SUM(F2:F6) is in F7 and is equal to 12 showing that $\mu_{\bar{x}} = \mu$. The variance of the sample mean is $\sigma^2_{\bar{x}} = \Sigma \bar{x}^2 p(\bar{x}) - \mu^2_{\bar{x}}$ and is computed as follows. $\Sigma \bar{x}^2 p(\bar{x})$ is computed in G2 through G6. The function =SUM(G2:G6) is in G7 and is equal to 147. When 12^2 or 144 is subtracted from 147, we get 3 or $\sigma^2_{\bar{x}} = \sigma^2/2$. Figure 8-5 shows that even with sample size 2, the sampling distribution of $\bar{x}$ is somewhat like a normal distribution. The larger probabilities are near 12 and they tail off to the right and left of 12.

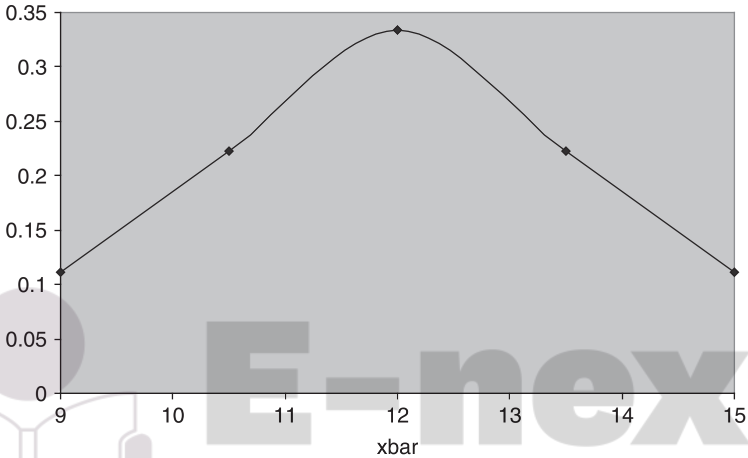

Fig. 8-5 Distribution of xbar for $n = 2$.

8.19 List all samples of size $n = 3$ that are possible (with replacement) from the population in Problem 8.17. Use EXCEL to construct the sampling distribution of the mean. Use the chart wizard of EXCEL to plot the sampling distribution of the mean. Show $\mu_{\bar{x}} = \mu$, and $\sigma^2_{\bar{x}} = \frac{\sigma^2}{3}$.

SOLUTION

A	B	C	D	E	F	G	H
			mean	xbar	p(xbar)	xbar*p(xbar)	xbar^2p(xbar)
9	9	9	9	9	0.037037037	0.333333333	3
9	9	12	10	10	0.111111111	1.111111111	11.11111111
9	9	15	11	11	0.222222222	2.444444444	26.88888889
9	12	9	10	12	0.259259259	3.111111111	37.33333333
9	12	12	11	13	0.222222222	2.888888889	37.55555556
9	12	15	12	14	0.111111111	1.555555556	21.77777778
9	15	9	11	15	0.037037037	0.555555556	8.333333333
9	15	12	12			12	146
9	15	15	13				
12	9	9	10				

Continued

A	B	C	D	E	F	G	H
			mean	xbar	p(xbar)	xbar*p(xbar)	xbar^2p(xbar)
12	9	12	11				
12	9	15	12				
12	12	9	11				
12	12	12	12				
12	12	15	13				
12	15	9	12				
12	15	12	13				
12	15	15	14				
15	9	9	11				
15	9	12	12				
15	9	15	13				
15	12	9	12				
15	12	12	13				
15	12	15	14				
15	15	9	13				
15	15	12	14				
15	15	15	15				

The EXCEL worksheet shows the possible sample values in A, B, and C, the mean in D, the sampling distribution of xbar is computed and given in E and F. In D2, the function =AVERAGE(A2:C2) is entered and a click-and-drag is performed from D2 to D28. Because the population is uniform each sample has probability 1/27 of being selected. The sample mean is represented by xbar. The mean of the sample mean is $\mu_{\bar{x}} = \Sigma \bar{x} p(\bar{x})$ and is computed in G2 through G8. The function =SUM(G2:G8) is in G9 and is equal to 12 showing that $\mu_{\bar{x}} = \mu$ for samples of size $n = 3$. The variance of the sample mean is $\sigma_{\bar{x}}^2 = \Sigma \bar{x}^2 p(\bar{x}) - \mu_{\bar{x}}^2$ and is computed as follows. $\Sigma \bar{x}^2 p(\bar{x})$ is computed in H2 through H8. The function =SUM(H2:H8) is in H9 and is equal to 146. When 12^2 or 144 is subtracted from 146 we get 2. Note that $\sigma_{\bar{x}}^2 = \sigma^2/3$. Figure 8-6 shows the normal distribution tendency of the distribution of xbar.

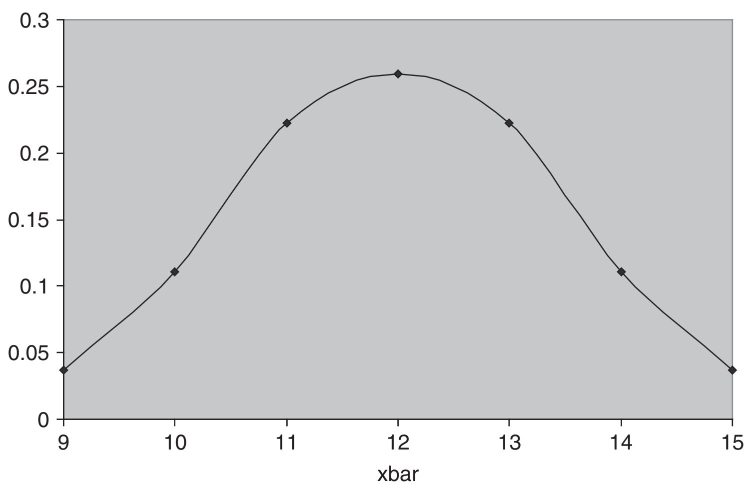

Fig. 8-6 Distribution of xbar for $n = 3$.

8.20 List all 81 samples of size $n = 4$ that are possible (with replacement) from the population in Problem 8.17. Use EXCEL to construct the sampling distribution of the mean. Use the chart wizard of EXCEL to plot the sampling distribution of the mean, show that $\mu_{\bar{x}} = \mu$, and show that $\sigma_{\bar{x}}^2 = \sigma^2/4$.

SOLUTION

The method used in Problems 8.18 and 8.19 is extended to samples of size 4. From the EXCEL worksheet, the following distribution for xbar is obtained. In addition, it can be shown that $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}}^2 = \sigma^2/4$.

xbar	p(xbar)	xbar*p(xbar)	xbar^2p(xbar)
9	0.012345679	0.111111111	1
9.75	0.049382716	0.481481481	4.694444444
10.5	0.12345679	1.296296296	13.61111111
11.25	0.197530864	2.222222222	25
12	0.234567901	2.814814815	33.77777778
12.75	0.197530864	2.518518519	32.11111111
13.5	0.12345679	1.666666667	22.5
14.25	0.049382716	0.703703704	10.02777778
15	0.012345679	0.185185185	2.777777778
	1	12	145.5

The EXCEL plot of the distribution of xbar for samples of size 4 is shown in Fig. 8-7.

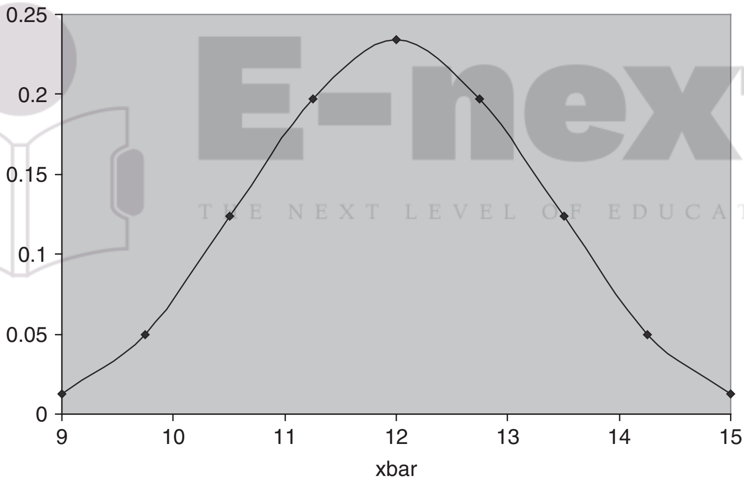

Fig. 8-7 Distribution of xbar for $n = 4$.

Supplementary Problems

SAMPLING DISTRIBUTION OF MEANS

8.21 A population consists of the four numbers 3, 7, 11, and 15. Consider all possible samples of size 2 that can be drawn with replacement from this population. Find (*a*) the population mean, (*b*) the population standard deviation, (*c*) the mean of the sampling distribution of means, and (*d*) the standard deviation of the sampling distribution of means. Verify parts (*c*) and (*d*) directly from (*a*) and (*b*) by using suitable formulas.

8.22 Solve Problem 8.21 if the sampling is without replacement.

8.23 The masses of 1500 ball bearings are normally distributed, with a mean of 22.40 g and a standard deviation of 0.048 g. If 300 random samples of size 36 are drawn from this population, determine the expected mean and standard deviation of the sampling distribution of means if the sampling is done (*a*) with replacement and (*b*) without replacement.

8.24 Solve Problem 8.23 if the population consists of 72 ball bearings.

8.25 How many of the random samples in Problem 8.23 would have their means (*a*) between 22.39 and 22.41 g, (*b*) greater than 22.42 g, (*c*) less than 22.37 g, and (*d*) less than 22.38 g or more than 22.41 g?

8.26 Certain tubes manufactured by a company have a mean lifetime of 800 h and a standard deviation of 60 h. Find the probability that a random sample of 16 tubes taken from the group will have a mean lifetime of (*a*) between 790 and 810 h, (*b*) less than 785 h, (*c*) more than 820 h, and (*d*) between 770 and 830 h.

8.27 Work Problem 8.26 if a random sample of 64 tubes is taken. Explain the difference.

8.28 The weights of packages received by a department store have a mean of 300 pounds (lb) and a standard deviation of 50 lb. What is the probability that 25 packages received at random and loaded on an elevator will exceed the specified safety limit of the elevator, listed as 8200 lb?

RANDOM NUMBERS

8.29 Work Problem 8.6 by using a different set of random numbers and selecting (*a*) 15, (*b*) 30, (*c*) 45, and (*d*) 60 samples of size 4 with replacement. Compare with the theoretical results in each case.

8.30 Work Problem 8.29 by selecting samples of size (*a*) 2 and (*b*) 8 with replacement, instead of size 4 with replacement.

8.31 Work Problem 8.6 if the sampling is without replacement. Compare with the theoretical results.

8.32 (*a*) Show how to select 30 samples of size 2 from the distribution in Problem 3.61.

(*b*) Compute the mean and standard deviation of the resulting sampling distribution of means, and compare with theoretical results.

8.33 Work Problem 8.32 by using samples of size 4.

SAMPLING DISTRIBUTION OF PROPORTIONS

8.34 Find the probability that of the next 200 children born, (*a*) less than 40% will be boys, (*b*) between 43% and 57% will be girls, and (*c*) more than 54% will be boys. Assume equal probabilities for the births of boys and girls.

8.35 Out of 1000 samples of 200 children each, in how many would you expect to find that (*a*) less than 40% are boys, (*b*) between 40% and 60% are girls, and (*c*) 53% or more are girls?

8.36 Work Problem 8.34 if 100 instead of 200 children are considered, and explain the differences in results.

8.37 An urn contains 80 marbles, of which 60% are red and 40% are white. Out of 50 samples of 20 marbles, each selected with replacement from the urn, how many samples can be expected to consist of (*a*) equal numbers of red and white marbles, (*b*) 12 red and 8 white marbles, (*c*) 8 red and 12 white marbles, and (*d*) 10 or more white marbles?

8.38 Design an experiment intended to illustrate the results of Problem 8.37. Instead of red and white marbles, you may use slips of paper on which R and W are written in the correct proportions. What errors might you introduce by using two different sets of coins?

8.39 A manufacturer sends out 1000 lots, each consisting of 100 electric bulbs. If 5% of the bulbs are normally defective, in how many of the lots should we expect (*a*) fewer than 90 good bulbs and (*b*) 98 or more good bulbs?

SAMPLING DISTRIBUTIONS OF DIFFERENCES AND SUMS

8.40 *A* and *B* manufacture two types of cables that have mean breaking strengths of 4000 lb and 4500 lb and standard deviations of 300 lb and 200 lb, respectively. If 100 cables of brand *A* and 50 cables of brand *B* are tested, what is the probability that the mean breaking strength of *B* will be (*a*) at least 600 lb more than *A* and (*b*) at least 450 lb more than *A*?

8.41 What are the probabilities in Problem 8.40 if 100 cables of both brands are tested? Account for the differences.

8.42 The mean score of students on an aptitude test is 72 points with a standard deviation of 8 points. What is the probability that two groups of students, consisting of 28 and 36 students, respectively, will differ in their mean scores by (*a*) 3 or more points, (*b*) 6 or more points, and (*c*) between 2 and 5 points?

8.43 An urn contains 60 red marbles and 40 white marbles. Two sets of 30 marbles each are drawn with replacement from the urn and their colors are noted. What is the probability that the two sets differ by 8 or more red marbles?

8.44 Solve Problem 8.43 if the sampling is without replacement in obtaining each set.

8.45 Election returns showed that a certain candidate received 65% of the votes. Find the probability that two random samples, each consisting of 200 voters, indicated more than a 10% difference in the proportions who voted for the candidate.

8.46 If U_1 and U_2 are the sets of numbers in Problem 8.11, verify that (*a*) $\mu_{U1+U2} = \mu_{U1} + \mu_{U2}$ and (*b*) $\sigma_{U1+U2} = \sqrt{\sigma^2_{U1} + \sigma^2_{U2}}$.

8.47 Three masses are measured as 20.48, 35.97, and 62.34 g, with standard deviations of 0.21, 0.46, and 0.54 g, respectively. Find the (*a*) mean and (*b*) standard deviation of the sum of the masses.

8.48 The mean voltage of a battery is 15.0 volts (V) and the standard deviation is 0.2 V. What is the probability that four such batteries connected in series will have a combined voltage of 60.8 V or more?

8.49 The credit hour distribution at Metropolitan Technological College is as follows:

x	6	9	12	15	18
$p(x)$	0.1	0.2	0.4	0.2	0.1

Find μ and σ^2. Give the 25 (with replacement) possible samples of size 2, their means, and their probabilities.

8.50 Refer to problem 8.49. Give and plot the probability distribution of xbar for $n = 2$.

8.51 Refer to problem 8.50. Show that $\mu_{\bar{x}} = \mu$ and $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{2}$.

8.52 Refer to problem 8.49. Give and plot the probability distribution of xbar for $n = 3$.