Chapter 7

Sampling Distributions

7.1 INTRODUCTION: SAMPLING WITH AND WITHOUT REPLACEMENT

Inferential statistics is used to draw conclusions about a population, based on a probability model of random samples of the population. For example, a pollster may want to estimate the proportion of all eligible voters favoring a particular presidential candidate by polling a random sample of eligible voters. Or, a statistician may want to use the mean starting income of a random sample of recent college graduates to estimate the mean starting income of all college graduates. *Since different random samples will most likely give different estimates, some knowledge of the variability of all possible estimates derived from random samples is needed to arrive at reasonable conclusions.* Before investigating this variability, some technical terminology is needed.

In general, a *population* is any finite set of objects being investigated. A sample of objects drawn from a population is a *random sample* if it is selected by a process in which every member of the population has essentially the same chance of being chosen. We consider two types of random sample: those drawn *with replacement* and those drawn *without replacement*. The probability distribution of a random variable defined on a space of random samples is called a *sampling distribution*. Sampling distributions are discussed in this chapter and their application to inferential statistics in the following chapters.

EXAMPLE 7.1 Suppose it is desired to determine the average age of students graduating from colleges in the U.S. in a given year. Here the population is the set of all college graduates in the U.S. for the given year. The age X of each graduate is a random variable defined on the population. The average age $\bar{X}$ of the students in a random sample of n graduates is a random variable defined on the space of all random samples of n graduates. The probability distribution of $\bar{X}$ is a sampling distribution.

Sampling With Replacement

In sampling with replacement, each object chosen is returned to the population before the next object is drawn. We define a random sample of size n, drawn with replacement, as an *ordered n-tuple* of objects from the population, repetitions allowed.

EXAMPLE 7.2 A population consists of the set $S = \{4, 7, 10\}$. The space of all random samples of size 2, drawn with replacement, consists of all ordered pairs (a, b), including repetitions, of numbers in S. There are nine such pairs, which are

$$(4,4),\quad (4,7),\quad (4,10),\quad (7,4),\quad (7,7),\quad (7,10),\quad (10,4),\quad (10,7),\quad (10,10)$$

The Space of Random Samples Drawn With Replacement

In general, if samples of size n are drawn with replacement from a population of size N, then the fundamental principle of counting says there are

$$N \cdot N \cdot \ldots \cdot N = N^n$$

such samples. In any survey involving samples of size n, each of these should have the same probability of being chosen. This is equivalent to making the collection of all N^n samples a probability space in

which each sample has probability $\frac{1}{N^n}$ of being chosen. Hence, in Example 7.2, there are $3^2 = 9$ random samples of size 2, and each of the nine random samples has probability $\frac{1}{9}$ of being chosen.

Sampling Without Replacement

In sampling without replacement, an object chosen is not returned to the population before the next object is drawn. We define a random sample of size n, drawn without replacement, as an *unordered subset* of n objects from the population.

EXAMPLE 7.3 When sampling without replacement from the population $S = \{4, 7, 10\}$, there are only three random samples of size 2, which are the three subsets of S containing two elements, namely

$$\{4, 7\}, \quad \{4, 10\}, \quad \{7, 10\}.$$

The Space of Random Samples Drawn Without Replacement

If samples of size n are drawn without replacement from a population of size N, then there are

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

such samples, which is the number of subsets of the population containing n elements. For instance, in Example 7.3, there are

$$\binom{3}{2} = \frac{3!}{2! \cdot 1!} = 3$$

random samples of size 2. As in the case of sampling with replacement, the collection of all random samples drawn without replacement can be made into a probability space in which any two samples have the same chance of being selected. In Example 7.3, each of the three random samples has probability $\frac{1}{3}$ of being chosen.

EXAMPLE 7.4 Suppose 75 out of the 100 seniors in a high-school senior class prefer candidate A over candidate B for class president. If 20 different seniors, chosen randomly, are polled about their preference, what is the probability that exactly 15 of them favor candidate A? To answer this question, first note that the 20 different seniors can be interpreted as a sample of size 20, drawn without replacement, from a population of size 100. There are $\binom{100}{20}$ such samples. The number of these samples in which 15 seniors favor candidate A is $\binom{75}{15}\binom{25}{5}$, where

$\binom{75}{15}$ = the number of ways 15 seniors can be chosen from the 75 that favor A, and

$\binom{25}{5}$ = the number of ways the remaining 5 seniors of the sample can be chosen from the 25 that do not favor candidate A

Therefore, the probability that exactly 15 seniors in the sample with favor A is

$$P(15) = \frac{\binom{75}{15}\binom{25}{5}}{\binom{100}{20}} \approx 0.226$$

Comparing Sampling With and Without Replacement

We saw in Example 7.4 that if 20 seniors, chosen without replacement, were polled, then the probability that exactly 15 of them would favor candidate A is approximately 0.226. If the 20 seniors were chosen with replacement, then their selection would be a binomial experiment, $b(20, \frac{75}{100})$, and the probability of exactly 15 in the sample favoring candidate A is

$$P(15) = \binom{20}{15}\left(\frac{75}{100}\right)^{15}\left(\frac{25}{100}\right)^{5} \approx 0.202$$

Figure 7-1 shows the complete probability histograms for the number of seniors favoring candidate A when samples of size 20 are drawn with or without replacement, respectively. For $k = 0, 1, 2, \ldots, 9$, and 20, the probability that k seniors favor candidate A is 0 to two decimal places in both types of sampling.

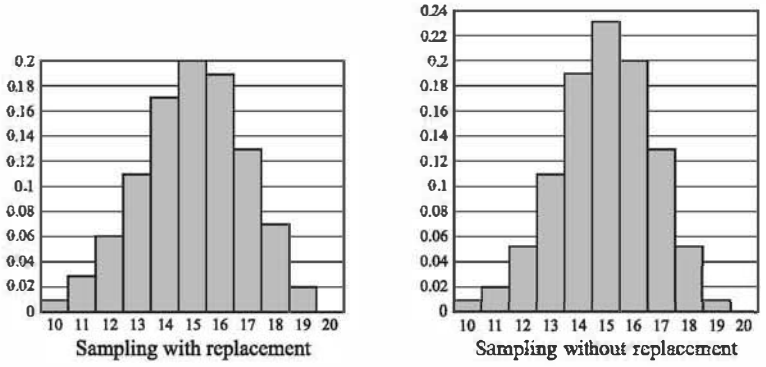

Fig. 7-1

The main difference between the two types of sampling is that when sampling with replacement, the individual outcomes in each sample are *independent*, whereas when sampling without replacement, the outcomes are not independent. For example, if two coins are drawn at random from three dimes and two quarters, then the probability of getting two quarters is $\frac{2}{5} \cdot \frac{2}{5} = 0.16$ if the coins are drawn with replacement, and $\frac{2}{5} \cdot \frac{1}{4} = 0.1$ without replacement. However, when the population is large in comparison with the sample size, results obtained by sampling are very similar whether the sampling is with or without replacement. Therefore, *when the population size is much larger than the sample size, a probability model that assumes the individual outcomes in each sample are independent can be applied to the sampling process regardless of whether the samples are obtained with or without replacement.*

EXAMPLE 7.5 Suppose that 55 percent of all eligible voters in a state favor candidate B for governor. If a random sample of 1000 eligible voters is chosen, find the probability that between 52 percent and 58 percent of the voters in the sample will favor candidate B.

A sample of 1000 voters is small in comparison with the number of eligible voters in any state, so we may use sampling with replacement as a probability model. The sample selection is then a binomial experiment $b(n,p)$, where $n = 1000$ and $p = 0.55$. The probability that between 52 percent and 58 percent of the voters sampled will favor candidate B is the probability that there will be between 520 and 580 successes in the experiment. This probability is equal to

$$\sum_{r=520}^{580} \binom{1000}{r}(0.55)^{r}(0.45)^{1000-r} \approx 0.95$$

The result was determined by using the normal approximation of the binomial (see Problem 7.8). Hence, approximately 95 percent of the time the random sample will be within 3 percentage points of the true percentage of the population favoring candidate B. Also, the result does not depend on the actual size of the total voting population, only that the sample is small by comparison.

7.2 SAMPLE MEAN

Sampling With Replacement

Suppose X is a random variable with mean μ and standard deviation σ, defined on some population. A random sample of size n, drawn with replacement, yields n values, $x_1, x_2, \ldots, x_n$, for X. Since the sample is drawn with replacement, these values are independent of one another. They can therefore be considered to be values of n independent random variables $X_1, X_2, \ldots, X_n$, each with mean μ and standard deviation σ. For example, if X is the age of college graduates in a given year, then X_i would be the age of the ith graduate ($i = 1, 2, \ldots, n$) in a random sample from this population. The random variable

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

is called the *sample mean* of $X_1, X_2, \ldots, X_n$. As a random variable, $\bar{X}$ also has a mean, $\mu_{\bar{X}}$, and a standard deviation, $\sigma_{\bar{X}}$. It can be shown that these sample parameters are related to the corresponding population parameters μ and σ, as stated in Theorem 7.1 below.

Theorem 7.1 (Mean and Standard Deviation of $\bar{X}$: Sampling With Replacement): Suppose a population random variable X has mean μ and standard deviation σ. Then the sample mean $\bar{X}$, for random samples of size n drawn with replacement, has mean $\mu_{\bar{X}}$ and standard deviation $\sigma_{\bar{X}}$ given by

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

Furthermore, if X is approximately normally distributed, then so is $\bar{X}$.

EXAMPLE 7.6 A population consists of the set $S = \{4, 7, 10\}$ as an equiprobable space. Random samples of size 2 are drawn with replacement.

(*a*) Compute the population mean, μ, and standard deviation, σ.
(*b*) Find the sampling distribution (probability distribution) for the sample mean, $\bar{X}$.
(*c*) Compute the mean, $\mu_{\bar{X}}$, and standard deviation, $\sigma_{\bar{X}}$, of $\bar{X}$, and compare with μ and σ.

(*a*) Since the population is an equiprobable space, the probability of each number in S occurring is $\frac{1}{3}$. Therefore, the mean and standard deviation of the population are

$$\mu = \sum xP(x) = 4 \cdot \frac{1}{3} + 7 \cdot \frac{1}{3} + 10 \cdot \frac{1}{3} = \frac{21}{3} = 7$$

and

$$\sigma = \sqrt{\sum (x - \mu)^2 P(x)} = \sqrt{(4-7)^2 \cdot \frac{1}{3} + (7-7)^2 \cdot \frac{1}{3} + (10-7)^2 \cdot \frac{1}{3}} = \sqrt{\frac{18}{3}} = \sqrt{6}$$

(*b*) Table 7-1 lists the mean value $\dfrac{(a+b)}{2}$ for every possible sample pair, and Table 7-2 gives the sampling distribution for the sample mean, $\bar{X}$.

(*c*) From Table 7-2,

$$\mu_{\bar{X}} = E(\bar{X}) = 4 \cdot \frac{1}{9} + 5.5 \cdot \frac{2}{9} + 7 \cdot \frac{3}{9} + 8.5 \cdot \frac{2}{9} + 10 \cdot \frac{1}{9} = \frac{63}{9} = 7$$

Table 7-1. Samples of size 2, sampling with replacement.

(a, b)	$\bar{x}$
(4, 4)	4
(4, 7)	5.5
(4, 10)	7
(7, 4)	5.5
(7, 7)	7
(7, 10)	8.5
(10, 4)	7
(10, 7)	8.5
(10, 10)	10

Table 7-2. Sampling distribution for $\bar{X}$, sampling with replacement.

$\bar{x}$	$P(\bar{x})$
4	$\frac{1}{9}$
5.5	$\frac{2}{9}$
7	$\frac{3}{9}$
8.5	$\frac{2}{9}$
10	$\frac{1}{9}$

and

$$\sigma_X = \sqrt{\sum(\bar{x} - \mu_X)^2 P(\bar{x})}$$

$$= \sqrt{(4-7)^2 \cdot \frac{1}{9} + (5.5-7)^2 \cdot \frac{2}{9} + (7-7)^2 \cdot \frac{3}{9} + (8.5-7)^2 \cdot \frac{2}{9} + (10-7)^2 \cdot \frac{1}{9}}$$

$$= \sqrt{\frac{27}{9}} = \sqrt{3}$$

Therefore, $\mu_X = 7 = \mu$, and $\sigma_X = \sqrt{3} = \dfrac{\sqrt{6}}{\sqrt{2}} = \dfrac{\sigma}{\sqrt{2}}$, which agree with the formulas of Theorem 7.1, where $n = 2$.

Sampling Without Replacement

If samples are drawn without replacement, then the sample values, $x_1, x_2, \ldots, x_n$, of a random variable X are not independent. Nevertheless, the average of the values, namely

$$\frac{x_1 + x_2 + \cdots + x_n}{n}$$

defines a sample random variable which will also be denoted by $\bar{X}$ and will also be called the *sample mean*. In this case, the mean and standard deviation of $\bar{X}$ are given by Theorem 7.2 below.

Theorem 7.2 (Mean and Standard Deviation of $\bar{X}$: Sampling Without Replacement): Suppose a population random variable X has mean μ and standard deviation σ. Then the sample mean $\bar{X}$, for random sample of size n drawn without replacement, has mean $\mu_{\bar{X}}$ and standard deviation $\sigma_{\bar{X}}$ given by

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

where N is the size of the population and $n < N$. Furthermore, if X is approximately normally distributed, so is $\bar{X}$.

EXAMPLE 7.7 Assume that random samples of size 2 are drawn without replacement from the population $S = \{4, 7, 10\}$ as an equiprobable space.

(*a*) Find the sampling distribution for the sample mean, $\bar{X}$.

(*b*) Compute the mean $\mu_{\bar{X}}$ and standard deviation $\sigma_{\bar{X}}$ of $\bar{X}$, and compare with the population mean μ and standard deviation σ.

(*a*) Table 7-3 lists the mean value $\dfrac{(a+b)}{2}$ for every possible sample pair, and Table 7-4 gives the sampling distribution for the sample mean, $\bar{X}$.

Table 7-3. Samples of size 2, sampling without replacement.

$\{a, b\}$	$\bar{x}$
{4, 7}	5.5
{4, 10}	7
{7, 10}	8.5

Table 7-4. Sampling distribution for $\bar{X}$, sampling without replacement.

$\bar{x}$	$P(\bar{x})$
5.5	$\frac{1}{3}$
7	$\frac{1}{3}$
8.5	$\frac{1}{3}$

(*b*) From Table 7-4,

$$\mu_{\bar{X}} = E(\bar{X}) = 5.5 \cdot \frac{1}{3} + 7 \cdot \frac{1}{3} + 8.5 \cdot \frac{1}{3} = \frac{21}{3} = 7,$$

and

$$\sigma_{\bar{X}} = \sqrt{\sum (\bar{x} - \mu_{\bar{X}})^2 P(\bar{x})} = \sqrt{(5.5-7)^2 \cdot \frac{1}{3} + (7-7)^2 \cdot \frac{1}{3} + (8.5-7)^2 \cdot \frac{1}{3}}$$

$$= \sqrt{\frac{4.5}{3}} = \sqrt{1.5}$$

From Example 7.5, the population mean and standard deviation are $\mu = 7$ and $\sigma = \sqrt{6}$. Hence, $\mu_{\bar{X}} = 7 = \mu$; also, $\sigma_{\bar{X}} = \sqrt{1.5}$ and $\dfrac{\sigma}{\sqrt{2}} \cdot \sqrt{\dfrac{N-n}{N-1}} = \dfrac{\sqrt{6}}{\sqrt{2}} \cdot \sqrt{\dfrac{3-2}{3-1}} = \sqrt{3} \cdot \sqrt{\dfrac{1}{2}} = \sqrt{1.5}$. These equations agree with the formulas of Theorem 7.2.

The Sampling Distribution of $\bar{X}$

The second parts of Theorems 7.1 and 7.2 say that if X is approximately normally distributed, then $\bar{X}$ is also approximately normally distributed. Since we are assuming that the population is finite, X cannot be exactly normal, but many random variables for large populations can, for most practical purposes, be considered to be normally distributed. For example, the national SAT scores in a given year are approximately normally distributed with mean 500 and standard deviation 100. The mean SAT scores for random samples of size n will also be approximatey normal with mean 500 and standard deviation $\dfrac{100}{\sqrt{n}}$. (Since the population size, N, of students taking the SAT is large in comparison to a typical sample size, n, we may assume $\sqrt{\dfrac{N-n}{N-1}} \approx 1$; equivalently, we may assume that the scores in each sample are independent.) The following remarkable theorem says that if the sample size is large, then the sample mean $\bar{X}$ is approximately normally distributed regardless of the distribution of X.

Theorem 7.3 (Central Limit Theorem): Supose X is a random variable with mean μ and standard deviation $\sigma > 0$, defined on some population. If n is large, then the sample mean $\bar{X}$ is approximately normally distributed.

As a rule of thumb, the central limit theorem applies when $n \geq 30$. Note that Theorems 7.1 and 7.2 still apply when n is large. That is, $\bar{X}$ has mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$ if the samples are drawn with replacement, whereas $\bar{X}$ has mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ if the samples are drawn without replacement and $n < N$.

Sampling From a Large Population

As noted earlier, when a random sample is drawn from a large population, it can be assumed that the values $x_1, x_2, \ldots, x_n$ of the sample are independent. The assumption of independence is key to much of the probability theory used as a model for statistical inference. In the following sections, phrases such as "the population is much larger than the sample size" or "the population is large in comparison to the sample size" are meant to convey that the x values obtained in samples are essentially independent. In practice, if the assumption $\sqrt{\frac{N-n}{N-1}} \approx 1$ is reasonable in a given context, then independence may be assumed. Hence the Central Limit Theorem can be rephrased as follows.

Theorem 7.3′ (Central Limit Theorem): Suppose X is a random variable with mean μ and standard deviation $\sigma > 0$, defined on some population. If n is large ($n \geq 30$) and the population size is large in comparison to n, then the sample mean $\bar{X}$ is approximately normally distributed with mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

EXAMPLE 7.8 Suppose that the number of customers entering Dee's Grocery each day over a five-year period is a random variable with mean 100 and standard deviation 10. Then the average number of customers computed over randomly selected 30-day periods can be modeled as a normal random variable with mean 100 and standard deviation $\frac{10}{\sqrt{30}} \approx 1.8$. To see this, first note that the sample size is 30, which is large enough to assume that the sample average is a normal random variable. Also, the population size is the number of days in a 5-year period, which is at least 1826 and sufficiently large compared with the sample size to enable us to assume that the numbers of customers in the days of a sample are independent; equivalently, $\sqrt{\frac{N-n}{N-1}} \geq \sqrt{\frac{1826-30}{1826-1}} \approx 0.9920 \approx 1$.

EXAMPLE 7.9 With reference to Example 7.8, what is the probability that the average number of customers entering Dee's Grocery daily over a 30-day period is between 95 and 105?

As indicated in Example 7.8, the average number of customers, or sample mean $\bar{X}$, can be modeled as a normal random variable with mean 100 and standard deviation 1.8. Then

$$Z = \frac{\bar{X} - 100}{1.8}$$

is a normal random variable with mean 0 and standard deviation 1, that is, a standard normal random variable. Using the standard normal table,

$$P(95 \leq \bar{X} \leq 105) = P\left(\frac{95-100}{1.8} \leq \frac{\bar{X}-100}{1.8} \leq \frac{105-100}{1.8}\right) = P(-2.78 \leq \bar{Z} \leq 2.78) = 0.9946$$

Hence, it is almost certain that the average number of customers entering the store daily over a 30-day period is between 95 and 105.

7.3 SAMPLE PROPORTION

Suppose a proportion p of a population favor candidate A for president. In a random sample of size n drawn from the population, a certain proportion $\hat{p}$ of the sample will favor candidate A, and the

collection of all such proportions defines a random variable $\hat{P}$, called the *sample proportion*. The mean and standard deviation of $\hat{P}$ are given in the next two theorems.

Theorem 7.4 (Mean and Standard Deviation of $\hat{P}$: Sampling With Replacement): Suppose the population proportion is p, and random samples of size n are drawn with replacement. Then the sample proportion $\hat{P}$ has mean p and standard deviation $\sqrt{\dfrac{p(1-p)}{n}}$.

Theorem 7.5 (Mean and Standard Deviation of $\hat{P}$: Sampling Without Replacement): Suppose the population size is N, the population proportion is p, and random samples of size n are drawn without replacement. Then the sample proportion $\hat{P}$ has mean p and standard deviation $\sqrt{\dfrac{p(1-p)}{n}} \cdot \sqrt{\dfrac{N-n}{N-1}}$.

If the population is much larger than the sample size, then $\sqrt{\dfrac{N-n}{N-1}} \approx 1$, and if the sample size itself is also large ($n \geq 30$), then the central limit theorem (Theorem 7.3′) can be used to obtain the following result.

Theorem 7.6 (Central Limit Theorem for Sample Proportions): Suppose the sample size n is large ($n \geq 30$), and the population size is large in comparison to n. Then the sample proportion $\hat{P}$ is approximately normally distributed with mean p and standard deviation $\sqrt{\dfrac{p(1-p)}{n}}$.

Theorems 7.4 and 7.5 follow from Theorems 7.1 and 7.2, respectively, and Theorem 7.6 follows from theorem 7.3′ (see Problems 7.63–7.65).

EXAMPLE 7.10 Suppose 25 percent of all U.S. workers belong to a labor union. What is the probability that in a random sample of 100 U.S. workers, at least 20 percent will belong to a labor union?

The sample size, $n = 100$, is greater than 30, and the total number of U.S. workers is much larger than 100. Therefore, the sample proportion $\hat{P}$ of workers that belong to a labor union can be modeled as a normal random variable with mean $p = 0.25$ and standard deviation $\sqrt{\dfrac{p(1-p)}{n}} = \sqrt{\dfrac{0.25 \times 0.75}{100}} \approx 0.0433$. Then

$$Z = \frac{\hat{P} - 0.25}{0.0433}$$

is a standard normal random variable. Using the standard normal table,

$$\begin{aligned} P(\hat{P} \geq 0.2) &= P\left(\frac{\hat{P} - 0.25}{0.0433} \geq \frac{0.2 - 0.25}{0.0433}\right) \\ &\approx P(Z \geq -1.15) \\ &= P(Z \leq 1.15) \\ &= 0.8749 \end{aligned}$$

Hence, it is very likely that there will be at least 20 percent union workers in a random sample of 100 workers.

7.4 SAMPLE VARIANCE

Let X be a population random variable with mean μ and standard deviation σ. We assume that random samples of size n are taken with replacement, or if they are taken without replacement, we assume that the population size is much larger than n. Then the values $x_1, x_2, \ldots, x_n$ of X in a random

sample are, in effect, values of n independent random variables $X_1, X_2, \ldots, X_n$, each with mean μ and standard deviation σ. The random variable

$$S^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n - 1}$$

where $\bar{X}$ is the sample mean, is called the *sample variance.*

Since S^2 is intended to be an average of the square deviations from $\bar{X}$, it may seem more natural to divide by n rather than $n - 1$. In fact, some statisticians do define S^2 with n as the denominator, and there are pros and cons for each choice. A reason in favor of dividing by $n - 1$, as above, is that the expected value of S^2 is then equal to σ^2, the variance of X (see Problem 7.31). In technical terms, the above S^2 is an *unbiased estimator* of σ^2. Before discussing a sampling distribution related to S^2, we must introduce the chi-square random variable.

The Chi-Square Distribution

Because of the central limit theorem, the normal distribution plays a major role in inferential statistics. Another continuous probability distribution that plays an important role in inferential statistics is the chi-square distribution, which can be defined as follows.

Definition: Let $Z_1, Z_2, \ldots, Z_k$ be k independent normal random variables, each with mean 0 and standard deviation 1. Then, the random variable

$$\chi^2 = Z_1^2 + Z_2^2 + \cdots + Z_k^2$$

is called a *chi-square random variable with k degrees of freedom.*

Properties of the Chi-Square Distribution

A chi-square random variable χ^2 with k degrees of freedom is often denoted by $\chi^2(k)$ to emphasize its dependence on the parameter k, which can be any positive integer, including 1. There is a density curve for each value of k, several of which are illustrated in Fig. 7-2. Note that $\chi^2(k)$ assumes only non-negative values (since it is a sum of squares). Also, as k increases, the corresponding density curve becomes less skewed to the right and more symmetric about the mode, which is $k - 2$; $\chi^2(k)$ has mean k and standard deviation $\sqrt{2k}$.

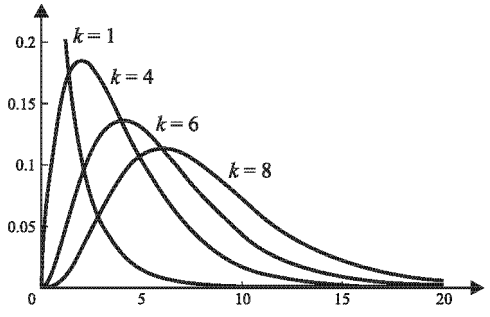

Fig. 7-2 Chi-square distribution for k degrees of freedom

EXAMPLE 7.11 Suppose X_1, X_2, and X_3 are independent normal random variables, each with mean 100 and standard deviation 15, and let $Z_i = (X_i - 100)/15$ for $i = 1, 2, 3$. Then Z_1, Z_2, and Z_3 are independent normal random variables, each with mean 0 and standard deviation 1. Therefore, Z_1^2, Z_2^2, and Z_3^2 are each $\chi^2(1)$, with mean 1 and standard deviation $\sqrt{2}$; and $Z_1^2 + Z_2^2 + Z_3^2$ is $\chi^2(3)$, with mean 3 and standard deviation $\sqrt{6}$.

The Sampling Distribution of $(n-1)S^2/\sigma^2$

We are now in a position to determine a sampling distribution related to the sample variance S^2. Note that if $X_1, X_2, \ldots, X_n$ are n random variables, each with mean μ and standard deviation $\sigma > 0$, then (see Problem 7.30)

$$\sum (X_i - \bar{X})^2 = \sum (X_i - \mu)^2 - n(\bar{X} - \mu)^2$$

Dividing both sides by σ^2 gives

$$\sum \frac{(X_i - \bar{X})^2}{\sigma^2} = \sum \frac{(X_i - \mu)^2}{\sigma^2} - \frac{(\bar{X} - \mu)^2}{\sigma^2/n}$$

If the X_is are independent random variables, the left side of the above equation is $(n-1)S^2/\sigma^2$; and if the X_is are also normally distributed, the right side is the difference of a $\chi^2(n)$ random variable (by definition) and a $\chi^2(1)$ random variable (by the Central Limit Theorem 7.3′). The following result can then be established.

Theorem 7.7: Suppose random samples of size n corresponding to some random variable X are drawn from a population whose size is much larger than n. Suppose also that X is (approximately) a normal random variable with mean μ and standard deviation $\sigma > 0$. Then $(n-1)S^2/\sigma^2$ is (approximately) a chi-square random variable with $n-1$ degrees of freedom.

Mean and Standard Deviation of S^2

As stated above, the expected value of S^2 is σ^2, the variance of X. That is, the mean of S^2 is σ^2. Also, since $(n-1)S^2/\sigma^2$ is a chi-square random variable with $n-1$ degrees of freedom, the standard deviation of $(n-1)S^2/\sigma^2$ is $\sqrt{2(n-1)}$. Therefore, the standard deviation of S^2 is $[\sqrt{2(n-1)}]\sigma^2/(n-1)$, which is equal to $[\sqrt{2/(n-1)}]\sigma^2$.

EXAMPLE 7.12 The annual college SAT scores are (approximately) normally distributed with mean $\mu = 500$ and standard deviation $\sigma = 100$. If S^2 is the sample variance on the space of all random samples of 50 SAT scores, then $49S^2/\sigma^2$ is (approximately) a $\chi^2(49)$ random variable, which has mean 49 and standard deviation $\sqrt{2 \cdot 49} = 7\sqrt{2} \approx 9.9$. S^2 itself has mean $\mu_{S^2} = \sigma^2 = 10{,}000$ and standard deviation $\sigma_{S^2} = [\sqrt{2/49}] \cdot 100^2 \approx 2020$.

BASIC ASSUMPTION REGARDING FUTURE SAMPLING

In Chapters 8, 9, 10, and 11, unless otherwise stated, we will assume, for simplicity, that either sampling is done with replacement or that the population size N is much larger than the sample size n. This will ensure that the individual outcomes of a random sample are essentially independent, and make the correction factor $\sqrt{\dfrac{N-n}{N-1}}$ for the sample variance unnecessary.

Solved Problems

SAMPLING WITH AND WITHOUT REPLACEMENT

7.1. Let $S = \{1, 5, 6, 8\}$.

(*a*) List all samples of size 3, with replacement.

(*b*) How many samples, with replacement, are there of size 4, size 5, size n?

(*a*) A sample with replacement is a 3-tuple of numbers from S, repetitions allowed. By the fundamental counting principle, there are $4 \times 4 \times 4 = 64$ such samples:

(1, 1, 1), (1, 1, 5), (1, 1, 6), (1, 1, 8) (1, 5, 1), (1, 5, 5,), (1, 5, 6), (1, 5, 8)
(1, 6, 1), (1, 6, 5), (1, 6, 6), (1, 6, 8), (1, 8, 1), (1, 8, 5), (1, 8, 6), (1, 8, 8)
(5, 1, 1), (5, 1, 5), (5, 1, 6), (5, 1, 8), (5, 5, 1), (5, 5, 5), (5, 5, 6), (5, 5, 8)
(5, 6, 1), (5, 6, 5), (5, 6, 6), (5, 6, 8), (5, 8, 1), (5, 8, 5), (5, 8, 6), (5, 8, 8)
(6, 1, 1), (6, 1, 5), (6, 1, 6), (6, 1, 8), (6, 5, 1), (6, 5, 5), (6, 5, 6), (6, 5, 8)
(6, 6, 1), (6, 6, 5), (6, 6, 6), (6, 6, 8), (6, 8, 1), (6, 8, 5), (6, 8, 6), (6, 8, 8)
(8, 1, 1), (8, 1, 5), (8, 1, 6), (8, 1, 8), (8, 5, 1), (8, 5, 5), (8, 5, 6), (8, 5, 8)
(8, 6, 1), (8, 6, 5), (8, 6, 6), (8, 6, 8), (8, 8, 1), (8, 8, 5), (8, 8, 6), (8, 8, 8)

(*b*) There are $4^4 = 256$ samples of size 4, $4^5 = 1024$ samples of size 5, and, in general, 4^n samples of size n for any positive integer n.

7.2. Let $S = \{1, 5, 6, 8\}$.

(*a*) List all samples of size 3, without replacement.

(*b*) How many samples, without replacement, are there of size 4, size n?

(*a*) A sample of size 3, without replacement, is a subset of S containing 3 elements. There are $\binom{4}{3} = 4$ subsets: {1, 5, 6}, {1, 5, 8}, {1, 6, 8}, {5, 6, 8}.

(*b*) For $n = 1, 2, 3, 4$, there are $\binom{4}{n}$ samples of size n; for $n > 4$, there are no samples of size n.

7.3. Five different banks draw a name at random from the same list of 100 names to send a credit-card application. How many random samples of five applications, one application for each bank, are possible? How many of the samples contain the same name more than once?

Let the banks be denoted by A, B, C, D, E. Each sample of five applications can be considered as a 5-tuple of names, where the first name is chosen by bank A, the second by bank B, and so on. Since repetitions are allowed, the sampling is with replacement, and there are $100^5 = 10{,}000{,}000{,}000$, or 10 billion, possible samples. By the fundamental counting principle, the number of samples with five different names is $100 \times 99 \times 98 \times 97 \times 96 = 9{,}034{,}502{,}400$. Subtracting this number from 10 billion, we find that there are 965,497,600 applications with the same name appearing more than once.

7.4. How many committees of 5 people can be randomly selected from a group of 10 women and 15 men. How many of the committees will have all men? How many will have all women? How many have three women and two men?

The number of 5-person committees is the number of ways that 5 people can be chosen from a group of 25 people, or the number of samples of size 5 that can be chosen, without replacement, from a population of size 25, which is $\binom{25}{5} = 53{,}130$. The number that have all men is $\binom{15}{5} = 3003$, and the number that have all women is $\binom{10}{5} = 252$. The number that have three women and two men is $\binom{10}{3}\binom{15}{2} = 12{,}600$.

7.5. What is the most likely breakdown of men and women in a committee of five randomly chosen from 15 men and 10 women?

Since the ratio of 15 men to 10 women is 3 to 2, it seems reasonable that a committee of 3 men and 2 women would be the most likely to occur at random. This expectation can be checked by simply counting the number of each type of committee. From Problem 7.4 we have the following counts.

5 men	3003
5 women:	252
3 women, 2 men:	12,600

Similarly, we get the following counts.

$$\text{1 man, 4 women:} \quad \binom{15}{1}\binom{10}{4} = 3150$$

$$\text{3 men, 2 women:} \quad \binom{15}{3}\binom{10}{2} = 20{,}475$$

$$\text{4 men, 1 woman:} \quad \binom{15}{4}\binom{10}{1} = 13{,}650$$

As expected, a committee with 3 men and 2 women is the most likely to occur.

7.6. A professor asks her class to determine the number of random samples of size 3 that can be selected, without replacement, from a population of three Democrats and two Republicans. James answers that there are three random samples, one consisting of 3 Democrats, one consisting of 2 Democrats and 1 Republican, and 1 consisting of 1 Democrat and 2 Republicans. Is James right? If not, how many are there?

All random samples of size 3 should have the same chance of occurring. However, since there are more Democrats than Republicans, a Democrat is more likely to be selected than a Republican. Therefore, a sample consisting of 2 Democrats and 1 Republican is more likely than a sample consisting of 1 Democrat and 2 Republicans. So James's answer is incorrect. To arrive at the correct answer, label the Democrats as D_1, D_2, D_3 and the Republicans as R_1, R_2. Then, there are $\binom{5}{3} = 10$ random samples of size 3, without replacement, namely:

$$\{D_1, D_2, D_3\}, \quad \{D_1, D_2, R_1\}, \quad \{D_1, D_2, R_2\}, \quad \{D_1, D_3, R_1\}, \quad \{D_1, D_3, R_2\},$$
$$\{D_2, D_3, R_1\}, \quad \{D_2, D_3, R_2\}, \quad \{D_1, R_1, R_2\}, \quad \{D_2, R_1, R_2\}, \quad \{D_3, R_1, R_2\}$$

Note that the probability that a random sample of size 3 will have 2 Democrats and 1 Republican is $\frac{6}{10}$, whereas the probability that the sample will have 1 Democrat and 2 Republicans is only $\frac{3}{10}$.

7.7. How many random samples of size 3, with replacement, are there for the population in Problem 7.6? How many are there in each of the categories: 3 Democrats; 2 Democrats, 1 Republican; 1 Democrat, 2 Republicans; 3 Republicans?

There are $5^3 = 125$ such random samples, broken down as follows.

3 Democrats: 27 random samples. They are: (D_1, D_1, D_1), (D_2, D_2, D_2), (D_3, D_3, D_3); 3 permutations each of (D_1, D_1, D_2), (D_1, D_1, D_3), (D_2, D_2, D_1), (D_2, D_2, D_3), (D_3, D_3, D_1), (D_3, D_3, D_2); and 6 permutations of (D_1, D_2, D_3).

2 Democrats, 1 Republican: 54 random samples. They are: 3 permutations each of (D_1, D_1, R_1), (D_1, D_1, R_2), (D_2, D_2, R_1), (D_2, D_2, R_2), (D_3, D_3, R_1), (D_3, D_3, R_2); and 6 permutations each of (D_1, D_2, R_1), (D_1, D_2, R_2), (D_1, D_3, R_1), (D_1, D_3, R_2), (D_2, D_3, R_1), (D_2, D_3, R_2).

1 Democrat, 2 Republicans: 36 random samples. They are: 6 permutations each of (D_1, R_1, R_2), (D_2, R_1, R_2), (D_3, R_1, R_2); and 3 permutations each of (D_1, R_1, R_1), (D_1, R_2, R_2), (D_2, R_1, R_1), (D_2, R_2, R_2), (D_3, R_1, R_1), (D_3, R_2, R_2).

3 Republicans: 8 random samples. They are: (R_1, R_1, R_1), (R_2, R_2, R_2); and 3 permutations each of (R_1, R_1, R_2), (R_1, R_2, R_2).

7.8. In Example 7.5 it was stated that $\sum_{r=520}^{580} \binom{1000}{r} (0.55)^r (0.45)^{1000-r} \approx 0.95$. Show that this is true.

The result says that $P(520 \le X \le 580) \approx 0.95$, where X is a binomial random variable with mean $np = 1000(0.55) = 550$, and standard deviation $\sqrt{np(1-p)} = \sqrt{1000(0.55)(0.45)} \approx 15.73$. By approximating X by a normal random variable with the same mean and standard deviation, and using the continuity correction, we get

$$P(520 \le X \le 1000) = P\left(\frac{519.5 - 550}{15.73} \le \frac{X - 550}{15.73} \le \frac{580.5 - 550}{15.73}\right)$$
$$\approx P(-1.94 \le Z \le 1.94)$$

where Z is the standard normal random variable. Then, from the standard normal table,

$$P(-1.94 \le Z \le 1.94) = 2P(0 \le Z \le 1.94)$$
$$\approx 2(0.4738)$$
$$\approx 0.95$$

SAMPLE MEAN

7.9. A population random variable X has mean 100 and standard deviation 16. What are the mean and standard deviation of the sample mean $\bar{X}$ for random samples of size 4 drawn with replacement?

For the population, $\mu = 100$ and $\sigma = 16$. By Theorem 7.1 the mean $\mu_{\bar{X}}$ and standard deviation $\sigma_{\bar{X}}$ of $\bar{X}$ are:

$$\mu_{\bar{X}} = \mu = 100 \qquad \text{and} \qquad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{16}{\sqrt{4}} = 8$$

7.10. With reference to Problem 7.9, what are the mean and standard deviation of $\bar{X}$ if the population size is 250, and the samples of size 4 are drawn without replacement?

By Theorem 7.2, where $N = 250$ and $n = 4$,

$$\mu_{\bar{X}} = \mu = 100 \qquad \text{and} \qquad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \frac{16}{\sqrt{4}} \sqrt{\frac{246}{249}} \approx 7.95$$

7.11. Suppose the random variable X in Problem 7.9 is approximately normally distributed. What is $P(95 \le \bar{X} \le 105)$ for samples of size 4 drawn with replacement?

By Problem 7.9, the mean and standard deviation of $\bar{X}$ are $\mu_{\bar{X}} = 100$ and $\sigma_{\bar{X}} = 8$. By Theorem 7.1, $\bar{X}$ is approximately normally distributed. Therefore,

$$P(95 \le \bar{X} \le 105) = P\left(\frac{95 - 100}{8} \le \frac{\bar{X} - 100}{8} \le \frac{105 - 100}{8}\right)$$
$$= P(-0.625 \le Z \le 0.625),$$

where Z is the standard normal random variable. Using a standard normal table,

$$P(-0.625 \le Z \le 0.625) = 2P(0 \le Z \le 0.625)$$
$$\approx 2(0.2324)$$
$$\approx 0.46$$

7.12. Suppose the random variable X in Problem 7.9 is approximately normally distributed. What is $P(95 \leq \bar{X} \leq 105)$ for samples of size 4 drawn without replacement?

By Problem 7.10, the mean and standard deviation of $\bar{X}$ are $\mu_{\bar{X}} = 100$ and $\sigma_{\bar{X}} \approx 7.95$. By Theorem 7.2, $\bar{X}$ is approximately normally distributed. Therefore,

$$P(95 \leq \bar{X} \leq 105) = P\left(\frac{95 - 100}{7.95} \leq \frac{\bar{X} - 100}{7.95} \leq \frac{105 - 100}{7.95}\right)$$
$$\approx P(-0.63 \leq Z \leq 0.63)$$

where Z is the standard normal random variable. Using a standard normal table,

$$P(-0.63 \leq Z \leq 0.63) = 2P(0 \leq Z \leq 0.63)$$
$$\approx 2(0.2357)$$
$$\approx 0.47$$

7.13. Let $S = \{1, 5, 6, 8\}$. Find the probability distribution of the sample mean $\bar{X}$ for random samples of size 2 drawn with replacement.

Since S has 4 elements, there are $4^2 = 16$ random samples of size 2 drawn with replacement. These pairs and their average values are given in the following table.

Sample	$\bar{x}$	Sample	$\bar{x}$	Sample	$\bar{x}$	Sample	$\bar{x}$
(1, 1)	1	(1, 5)	3	(1, 6)	3.5	(1, 8)	4.5
(5, 1)	3	(5, 5)	5	(5, 6)	5.5	(5, 8)	6.5
(6, 1)	3.5	(6, 5)	5.5	(6, 6)	6	(6, 8)	7
(8, 1)	4.5	(8, 5)	6.5	(8, 6)	7	(8, 8)	8

The probability distribution of $\bar{X}$ is given in the following table:

$\bar{x}$	1	3	3.5	4.5	5	5.5	6	6.5	7	8
$p(\bar{x})$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

7.14. Let $S = \{1, 5, 6, 8\}$. Find the probability distribution of the sample mean $\bar{X}$ for random samples of size 2 drawn without replacement.

Since S has 4 elements, there are $\binom{4}{2} = 6$ random samples of size 2 drawn without replacement.

These, their average value, and the probability distribution of $\bar{X}$ are given in the following two tables:

Sample	$\bar{x}$
{1, 5}	3
{1, 6}	3.5
{1, 8}	4.5
{5, 6}	5.5
{5, 8}	6.5
{6, 8}	7

$\bar{x}$	$p(\bar{x})$
3	$\frac{1}{6}$
3.5	$\frac{1}{6}$
4.5	$\frac{1}{6}$
5.5	$\frac{1}{6}$
6.5	$\frac{1}{6}$
7	$\frac{1}{6}$

7.15. Let $S = \{1, 5, 6, 8\}$. Compute the population mean μ and standard deviation σ. Also, compute the mean $\mu_{\bar{X}}$ and standard deviation $\sigma_{\bar{X}}$ of the sample mean $\bar{X}$ for random samples of size 2 drawn with replacement. Verify that $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}} = \sigma/\sqrt{2}$, as stated in Theorem 7.1.

The population, taken as an equiprobable space, has mean $\mu = \dfrac{1+5+5+8}{4} = 5$, and standard deviation $\sigma = \sqrt{\dfrac{(1-5)^2+(5-5)^2+(6-5)^2+(8-5)^2}{4}} = \sqrt{\dfrac{26}{4}} = \dfrac{\sqrt{26}}{2}$. Using the probability distribution table in Problem 7.13, the mean of $\bar{X}$ is

$$\begin{aligned}\mu_{\bar{X}} &= 1\times\frac{1}{16} + 3\times\frac{2}{16} + 3.5\times\frac{2}{16} + 4.5\times\frac{2}{16} + 5\times\frac{1}{16} + 5.5\times\frac{2}{16} + 6\times\frac{1}{16} \\ &\quad + 6.5\times\frac{2}{16} + 7\times\frac{2}{16} + 8\times\frac{1}{16} \\ &= \frac{80}{16} = 5\end{aligned}$$

which is the same as the population mean. The variance of $\bar{X}$ is

$$\begin{aligned}\sigma_{\bar{X}}^2 &= (1-5)^2\frac{1}{16} + (3-5)^2\frac{2}{16} + (3.5-5)^2\frac{2}{16} + (4.5-5)^2\frac{2}{16} + (5-5)^2\frac{1}{16} \\ &\quad + (5.5-5)^2\frac{2}{16} + (6-5)^2\frac{1}{16} + (6.5-5)^2\frac{2}{16} + (7-5)^2\frac{2}{16} + (8-5)^2\frac{1}{16} \\ &= \frac{1}{16}(16+8+4.5+0.5+0+0.5+1+4.5+8+9) \\ &= \frac{52}{16} = \frac{13}{4}\end{aligned}$$

Therefore, the standard deviation of $\bar{X}$ is $\sigma_{\bar{X}} = \dfrac{\sqrt{13}}{2}$. Since $\sigma = \dfrac{\sqrt{26}}{2}$ and $n = 2$, it follows that $\dfrac{\sigma}{\sqrt{n}} = \dfrac{\sqrt{26}}{2\sqrt{2}} = \dfrac{\sqrt{13}}{2} = \sigma_{\bar{X}}$, as stated in Theorem 7.1.

7.16. Let $S = \{1, 5, 6, 8\}$. Compute the mean $\mu_{\bar{X}}$ and standard deviation $\sigma_{\bar{X}}$ of $\bar{X}$ for random samples of size 2 drawn without replacement. Verify that $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}\sqrt{\dfrac{N-n}{N-1}}$, as stated in Theorem 7.2.

As computed in Problem 7.15, the population mean and standard deviation are $\mu = 5$ and $\sigma = \dfrac{\sqrt{26}}{2}$. Using the probability distribution table in Problem 7.14, the mean of $\bar{X}$ is

$$\begin{aligned}\mu_{\bar{X}} &= 3\times\frac{1}{6} + 3.5\times\frac{1}{6} + 4.5\times\frac{1}{6} + 5.5\times\frac{1}{6} + 6.5\times\frac{1}{6} + 7\times\frac{1}{6} \\ &= \frac{30}{6} = 5\end{aligned}$$

as stated in Theorem 7.2. The standard deviation of $\bar{X}$ is

$$\begin{aligned}\sigma_{\bar{X}} &= \sqrt{(3-5)^2\frac{1}{6} + (3.5-5)^2\frac{1}{6} + (4.5-5)^2\frac{1}{6} + (5.5-5)^2\frac{1}{6} + (6.5-5)^2\frac{1}{6} + (7-5)^2\frac{1}{6}} \\ &= \sqrt{\frac{13}{6}}\end{aligned}$$

Since $N = 4$ and $n = 2$, we have $\dfrac{\sigma}{\sqrt{n}}\sqrt{\dfrac{N-n}{N-1}} = \dfrac{\sqrt{26}}{2\sqrt{2}}\sqrt{\dfrac{2}{3}} = \dfrac{\sqrt{13}}{2}\dfrac{\sqrt{2}}{\sqrt{3}} = \dfrac{\sqrt{13}}{\sqrt{2}\sqrt{3}} = \sqrt{\dfrac{13}{6}}$, which is equal to $\sigma_{\bar{X}}$, as stated in Theorem 7.2.

7.17. Let $S = \{1, 5, 6, 8\}$. Find the probability distribution of the sample mean $\bar{X}$ for random samples of size 3 drawn (*a*) with replacement, (*b*) without replacement.

(*a*) There are $4^3 = 64$ random samples of size 3 drawn with replacement. These are shown in Problem 7.1. By finding the average of the three entries in each triple, we arrive at the following probability distribution.

$\bar{x}$	$p(\bar{x})$	$\bar{x}$	$p(\bar{x})$	$\bar{x}$	$p(\bar{x})$
1	1/64	13/3	3/64	19/3	6/64
7/3	3/64	14/3	6/64	20/3	3/64
8/3	3/64	5	7/64	7	3/64
10/3	3/64	16/3	3/64	22/3	3/64
11/3	3/64	17/3	6/64	24/3	1/64
4	6/64	6	4/64		

(*b*) There are $\binom{4}{3} = 4$ random samples of size 4 drawn without replacement. They are $\{1, 5, 6\}$, $\{1, 5, 8\}$, $\{1, 6, 8\}$, $\{5, 6, 8\}$. Computing the average of the entries in each of these, we arrive at the following probability distribution table.

$\bar{x}$	4	14/3	5	19/3
$p(\bar{x})$	1/4	1/4	1/4	1/4

7.18. Find $P(4 \le \bar{X} \le 6)$, where $\bar{X}$ is the sample mean for random samples of size 3 drawn with replacement from the population $\{1, 5, 6, 8\}$.

Using the probability distribution table in Problem 7.17(*a*), we find that

$$\begin{aligned} P(4 \le \bar{X} \le 6) &= p(4) + p\left(\frac{13}{3}\right) + p\left(\frac{14}{3}\right) + p(5) + p\left(\frac{16}{3}\right) + p\left(\frac{17}{3}\right) + p(6) \\ &= \frac{6}{64} + \frac{3}{64} + \frac{6}{64} + \frac{7}{64} + \frac{3}{64} + \frac{6}{64} + \frac{4}{64} \\ &= \frac{35}{64} \approx 0.55 \end{aligned}$$

7.19. Find $P(4 \le \bar{X} \le 6)$, where $\bar{X}$ is the sample mean for random samples of size 3 drawn without replacement from the population $\{1, 5, 6, 8\}$.

Using the probability distribution table in Problem 7.17(*b*), we find that

$$\begin{aligned} P(4 \le \bar{X} \le 6) &= p(4) + p\left(\frac{14}{3}\right) + p(5) \\ &= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4} \end{aligned}$$

7.20. Does the Central Limit Theorem (Theorem 7.3) apply to the sample mean $\bar{X}$ for random samples of size 36 drawn with replacement from the population $\{1, 5, 6, 8\}$? If so, use the theorem to compute $P(4 \le \bar{X} \le 6)$.

Since the sample size 36 is larger than 30, Theorem 7.3 does apply. Hence, we may assume that $\bar{X}$ is approximately normally distributed. Also, by Theorem 7.1, $\bar{X}$ has mean $\mu_X = 5$ and standard deviation

$\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{36}}$, where $\sigma = \dfrac{\sqrt{26}}{2}$ is the population standard deviation (as computed in Problem 7.15). Therefore, $\sigma_{\bar{X}} = \dfrac{\sqrt{26}}{2\sqrt{36}} = \dfrac{\sqrt{26}}{12} \approx 0.4249$. Then

$$P(4 \le \bar{X} \le 6) = P\left(\frac{4-5}{0.4249} \le \frac{\bar{X}-5}{0.4249} \le \frac{6-5}{0.4249}\right)$$
$$\approx P(-2.35 \le Z \le 2.35)$$

where Z is the standard normal random variable. Using a standard normal table, we find that $P(-2.35 \le Z \le 2.35) = 2P(0 \le Z \le 2.35) \approx 2(0.4906) \approx 0.98$.

7.21. Does the Central Limit Theorem (Theorem 7.3′) apply to the sample mean $\bar{X}$ for random samples drawn without replacement from the population $\{1, 5, 6, 8\}$?

No, only samples of size 4 or less can be drawn without replacement. Furthermore, the population size, 4, can never be much larger than the sample size.

SAMPLE PROPORTION

7.22. The proportion of unmarried men between ages 21 and 30 years in a town is $\frac{2}{3}$. Suppose random samples of size 16 are drawn with replacement from all men in the town between ages 21 and 30. What are the mean and standard deviation of the proportion $\hat{P}$ for all such samples?

By Theorem 7.4, the mean of $\hat{P}$ is $\frac{2}{3}$, and the standard deviation of $\hat{P}$ is

$$\sqrt{\frac{\frac{2}{3}(1-\frac{2}{3})}{16}} = \sqrt{\frac{\frac{2}{3}\times\frac{1}{3}}{16}} = \frac{\sqrt{2}}{12} \approx 0.1179$$

7.23. Suppose the town in Problem 7.22 has 225 men between ages 21 and 30 years, and the sampling is without replacement. Then what are the mean and standard deviation of $\hat{P}$?

By Theorem 7.5, the mean of $\hat{P}$ is still $\frac{2}{3}$, but the standard deviation is the standard deviation without replacement, 0.1179, multiplied by

$$\sqrt{\frac{225-16}{225-1}} = \sqrt{\frac{209}{224}} \approx 0.9659$$

Hence, the new standard deviation is approximately $0.1179 \times 0.9659 \approx 0.1139$.

7.24. The proportion of Democrats in a population consisting of three Democrats, D_1, D_2, D_3, and two Republicans, R_1, R_2 is $p = \frac{3}{5}$. There are 125 random samples of size $n = 3$ that can be drawn with replacement from the population. Find the probability distribution for the sample proportion $\hat{P}$ of Democrats defined by the collection of all 125 such samples (see also Problem 7.60).

Let $\hat{p}$ denote the proportion of Democrats in a given sample; $\hat{p}$ can assume the values: 1 (three Democrats), $\frac{2}{3}$ (two Democrats, one Republican), $\frac{1}{3}$ (one Democrat, two Republicans), 0 (three Republicans). Problem 7.7 gives the breakdown of the samples into these categories, which results in the following probability distribution table.

Category	$\hat{p}$	Frequency	$P(\hat{p})$
3 Democrats	1	27	$\frac{27}{125}$
2 Democrats, 1 Republican	$\frac{2}{3}$	54	$\frac{54}{125}$
1 Democrat, 2 Republicans	$\frac{1}{3}$	36	$\frac{36}{125}$
3 Republicans	0	8	$\frac{8}{125}$

7.25. Verify that the sample proportion $\hat{P}$ in Problem 7.24 has mean $p = \frac{3}{5}$ and standard deviation $\sqrt{\dfrac{p(1-p)}{n}}$, as stated in Theorem 7.4.

From the probability distribution table in Problem 7.24, the mean of $\hat{P}$ is

$$\sum \hat{p}P(\hat{p}) = 1 \times \frac{27}{125} + \frac{2}{3} \times \frac{54}{125} + \frac{1}{3} \times \frac{36}{125} + 0 \times \frac{8}{125} = \frac{75}{125} = \frac{3}{5}$$

The variance of $\hat{P}$ is

$$\begin{aligned}\sum (\hat{p} - p)^2 P(\hat{p}) &= \left(1 - \frac{3}{5}\right)^2 \frac{27}{125} + \left(\frac{2}{3} - \frac{3}{5}\right)^2 \frac{54}{125} + \left(\frac{1}{3} - \frac{3}{5}\right)^2 \frac{36}{125} + \left(0 - \frac{3}{5}\right)^2 \frac{8}{125} \\ &= \frac{4}{25} \cdot \frac{27}{125} + \frac{1}{225} \cdot \frac{54}{125} + \frac{16}{225} \cdot \frac{36}{125} + \frac{9}{25} \cdot \frac{8}{125} \\ &= \frac{4}{25} \cdot \frac{27}{125} + \frac{1}{25} \cdot \frac{6}{125} + \frac{16}{25} \cdot \frac{4}{125} + \frac{9}{25} \cdot \frac{8}{125} \\ &= \frac{250}{25 \times 125} \\ &= \frac{2}{25}\end{aligned}$$

Therefore, the standard deviation of $\hat{P}$ is $\dfrac{\sqrt{2}}{5}$. Also,

$$\sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{\frac{3}{5}(1 - \frac{3}{5})}{3}} = \frac{\sqrt{2}}{5}$$

7.26. There are only $\binom{5}{3} = 10$ random samples of size $n = 3$ that can be drawn without replacement from the population D_1, D_2, D_3, R_1, R_2. Find the probability distribution for the sample proportion $\hat{P}$ of Democrats defined by the collection of all 10 random samples.

The ten random samples of size $n = 3$, drawn without replacement are:

$$\{D_1, D_2, D_3\}, \quad \{D_1, D_2, R_1\}, \quad \{D_1, D_2, R_2\}, \quad \{D_1, D_3, R_1\}, \quad \{D_1, D_3, R_2\},$$
$$\{D_2, D_3, R_1\}, \quad \{D_2, D_3, R_2\}, \quad \{D_1, R_1, R_3\}, \quad \{D_2, R_1, R_2\}, \quad \{D_3, R_1, R_3\}$$

Let $\hat{p}$ denote the proportion of Democrats in a given sample; $\hat{p}$ can assume the values: 1 (three Democrats), $\frac{2}{3}$ (two Democrats, one Republican), or $\frac{1}{3}$ (one Democrat, two Republicans). We obtain the following probability distribution table.

Category	$\hat{p}$	Frequency	$P(\hat{p})$
3 Democrats	1	1	$\frac{1}{10}$
2 Democrats, 1 Republican	$\frac{2}{3}$	6	$\frac{6}{10}$
1 Democrat, 2 Republicans	$\frac{1}{3}$	3	$\frac{3}{10}$

7.27. Verify that the sample proportion $\hat{P}$ in Problem 7.26 has mean $p = \frac{3}{5}$ and standard deviation $\sqrt{\frac{p(1-p)}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$, as stated in Theorem 7.5.

From the probability distribution table in Problem 7.26, the mean of $\hat{P}$ is

$$\sum \hat{p}P(\hat{p}) = 1 \times \frac{1}{10} + \frac{2}{3} \times \frac{6}{20} + \frac{1}{3} \times \frac{3}{10} = \frac{6}{10} = \frac{3}{5}$$

The standard deviation is

$$\sqrt{\left(1-\frac{3}{5}\right)^2 \frac{1}{10} + \left(\frac{2}{3}-\frac{3}{5}\right)^2 \frac{6}{10} + \left(\frac{1}{3}-\frac{3}{5}\right)^2 \frac{3}{10}} = \sqrt{\left(\frac{2}{5}\right)^2 \frac{1}{10} + \left(\frac{1}{15}\right)^2 \frac{6}{10} + \left(\frac{4}{15}\right)^2 \frac{3}{10}}$$

$$= \sqrt{\frac{90}{2250}} = \sqrt{\frac{1}{15}} = \frac{1}{5}$$

Furthermore,

$$\sqrt{\frac{p(1-p)}{n}} \cdot \sqrt{\frac{N-n}{n-1}} = \sqrt{\frac{\frac{3}{5}(1-\frac{3}{5})}{3}} \cdot \sqrt{\frac{5-3}{5-1}}$$

$$= \frac{\sqrt{2}}{5} \cdot \sqrt{\frac{1}{2}} = \frac{1}{5}$$

SAMPLE VARIANCE

7.28. Suppose Z_1, Z_2, Z_3 are three independent standard normal random variables. Use these to generate three chi-square random variables, each with 2 degrees of freedom. What are the mean and variance of each of the three chi-square random variables?

$Z_1^2 + Z_2^2$, $Z_1^2 + Z_3^2$, and $Z_2^2 + Z_3^2$ are each chi-square random variables with $k = 2$ degrees of freedom. Each one has mean 2 and variance $2k = 4$.

7.29. Suppose Z is a standard normal random variable. Is $Z^2 + Z^2$ a chi-square random variable with 2 degrees of freedom?

No. If $Z^2 + Z^2$ were $\chi^2(2)$, it would have variance 4, as in Problem 7.28, but $Z^2 + Z^2 = 2Z^2$, and $\text{Var}(2Z^2) = 2^2 \text{Var}(Z^2) = 4 \times 2 = 8$.

7.30. Let X_1, X_2 be two random variables, each with mean μ. Show that $\sum_{i=1}^{2}(X_i - \bar{X})^2 = \sum_{i=1}^{2}(X_i - \mu)^2 - 2(\bar{X} - \mu)^2$, where $\bar{X}$ is the sample mean.

$$\sum_{i=1}^{2}(X_i - \bar{X})^2 = (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2$$

$$= [(X_1 - \mu) + (\mu - \bar{X})]^2 + [(X_2 - \mu) + (\mu - \bar{X})]^2$$

$$= (X_1 - \mu)^2 + 2(X_1 - \mu)(\mu - \bar{X}) + (\mu - \bar{X})^2$$

$$+ (X_2 - \mu)^2 + 2(X_2 - \mu)(\mu - \bar{X}) + (\mu - \bar{X})^2$$

$$= \sum_{i=1}^{2}(X_i - \mu)^2 + (\mu - \bar{X})(2X_1 - 2\mu + \mu - \bar{X} + 2X_2 - 2\mu + \mu - \bar{X})$$

$$= \sum_{i=1}^{2}(X_i - \mu)^2 + (\mu - \bar{X})(2\bar{X} - 2\mu)$$

$$= \sum_{i=1}^{2}(X_i - \mu)^2 - 2(\bar{X} - \mu)^2$$

The same procedure can be used to show that, for any positive integer n, $\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2$.

7.31. Let $X_1, X_2, \ldots, X_n$ be n independent random variables, each with mean μ and standard deviation σ. Show that the expected value of the sample variance,

$$S^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n - 1}$$

is equal to σ^2.

First note that $\sigma_{X_i}^2 = E(X_i - \mu)^2 = \sigma^2$ and $\sigma_{\bar{X}}^2 = E(\bar{X} - \mu)^2 = \sigma^2/n$ by Theorem 7.1. Then

$$\begin{aligned} E(S^2) = E\left(\frac{\sum(X_i - \bar{X})^2}{n-1}\right) &= \frac{1}{n-1} E(\sum(X_i - \mu)^2 - n(\bar{X} - \mu)^2) \\ &= \frac{1}{n-1}\sum E(X_i - \mu)^2 - \frac{n}{n-1} E(\bar{X} - \mu)^2 \\ &= \frac{1}{n-1}\sum \sigma^2 - \frac{n}{n-1}\cdot\frac{\sigma^2}{n} \\ &= \frac{n\sigma^2}{n-1} - \frac{\sigma^2}{n-1} \\ &= \sigma^2 \end{aligned}$$

7.32. Let $S = \{1, 5, 6, 8\}$. Find the probability distribution of the sample variance S^2 for random samples of size 3 drawn without replacement.

There are four random samples of size 3 drawn without replacement: $\{1, 5, 6\}$, $\{1, 5, 8\}$, $\{1, 6, 8\}$, $\{5, 6, 8\}$. There are four corresponding values of

$$S^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + (X_3 - \bar{X})^2}{2}$$

and each has probability $\frac{1}{4}$, as indicated in the following table.

Sample	X_1	X_2	X_3	$\bar{X}$	S^2	p
$\{1, 5, 6\}$	1	5	6	4	7	$\frac{1}{4}$
$\{1, 5, 8\}$	1	5	8	$\frac{14}{3}$	$\frac{37}{3}$	$\frac{1}{4}$
$\{1, 6, 8\}$	1	6	8	5	13	$\frac{1}{4}$
$\{5, 6, 8\}$	5	6	8	$\frac{19}{3}$	$\frac{7}{3}$	$\frac{1}{4}$

7.33. Use the probability distribution determined in Problem 7.32 to compute the mean μ_{S^2} and the standard deviation σ_{S^2} of the sample variance S^2 for random samples of size 3 drawn without replacement from the population $\{1, 5, 6, 8\}$.

The mean of S^2 is $\mu_{S^2} = 7 \times \frac{1}{4} + \frac{37}{3} \times \frac{1}{4} + 13 \times \frac{1}{4} + \frac{7}{3} \times \frac{1}{4} = \frac{104}{12} = \frac{26}{3}$; and the variance of S^2 is

$$\begin{aligned} \sigma_{S^2}^2 &= \left(7 - \frac{26}{3}\right)^2 \times \frac{1}{4} + \left(\frac{37}{3} - \frac{26}{3}\right)^2 \times \frac{1}{4} + \left(13 - \frac{26}{3}\right)^2 \times \frac{1}{4} + \left(\frac{7}{3} - \frac{26}{3}\right)^2 \times \frac{1}{4} \\ &= \frac{676}{36} \\ &= \frac{169}{9} \end{aligned}$$

Therefore, the standard deviation of S^2 is $\sigma_{S^2} = \sqrt{\frac{169}{9}} = \frac{13}{3}$.

7.34. It can be shown that when sampling without replacement, the mean of the corresponding sample variance is $\mu_{S^2} = \frac{N}{N-1}\sigma^2$, where σ^2 is the population variance, taken as an equiprobable space, and N is the population size. Use Problem 7.33 to verify this result for the population $\{1, 5, 6, 8\}$, when samples of size 3 are drawn without replacement.

In Problem 7.15 it was determined that the population variance for $\{1, 5, 6, 8\}$, taken as an equiprobable space, is $\sigma^2 = \frac{26}{4}$. From Problem 7.33, we have $\mu_{S^2} = \frac{26}{3}$. Since $N = 4$, we get $\frac{N}{N-1}\sigma^2 = \frac{4}{3} \cdot \frac{26}{4} = \frac{26}{3}$, as was to be shown.

7.35. Suppose samples of size 10 corresponding to a population random variable X are drawn without replacement. Suppose also that X is normal, with mean 75 and standard deviation 5. What are the mean μ_{S^2} and standard deviation σ_{S^2} of the sample variance S^2?

The mean μ_{S^2} of S^2 is equal to the variance of X regardless of the sample size. Therefore, $\mu_{S^2} = 5^2 = 25$. Also, by Theorem 7.7, $9s^2/25$ is chi-square with 9 degrees of freedom. Therefore, the standard deviation of $9s^2/25$ is $\sqrt{2 \times 9} = 3\sqrt{2}$; and the standard deviation of S^2 is $\sigma_{S^2} = \frac{25}{9} \times 3\sqrt{2} = \frac{25\sqrt{2}}{3}$.

Supplementary Problems

INTRODUCTION: SAMPLING WITH AND WITHOUT REPLACEMENT

7.36. How many samples of size 3 can be drawn from $S = \{2, 4, 8, 10, 12\}$, (*a*) with replacement, (*b*) without replacement?

7.37. In Problem 7.36, how many of the samples drawn with replacement have three different numbers?

7.38. If a population has size 10, what is the sample size n for which there are the most samples drawn (*a*) with replacement, (*b*) without replacement?

7.39. Repeat Problem 7.38 for a population of size N.

7.40. If a student guesses each answer in a 5-question True–False test, what is the most likely number of correct answers the student will get? What is the least likely number of correct answers the student will get?

7.41. Suppose there are 20 business majors in a statistics class of 32 students. If a random sample of 4 students is chosen without replacement, what is the probability that at least 2 of them will be business majors?

7.42. Repeat Problem 7.41 if the samples are chosen with replacement.

SAMPLE MEAN

7.43. A population random variable X has mean 75 and standard deviation 8. Find the mean and standard deviation of $\bar{X}$, based on random samples of size 25 taken with replacement.

7.44. Repeat Problem 7.43 if the random samples are taken without replacement.

7.45. Suppose the random variable X in Problem 7.43 is approximately normally distributed. Find $P(72 \leq \bar{X} \leq 78)$.

7.46. Repeat Problem 7.45 if the samples are taken without replacement, and the population has size 400.

7.47. SAT scores are approximately normally distributed with mean 500 and standard deviation 100. If a random sample of size 50 is taken, what is $P(\bar{X} \geq 525)$?

7.48. With reference to Problem 7.47, how large must the sample size be so that $P(475 \leq \bar{X} \leq 525) = 0.95$?

7.49. A population random variable X has mean 250 and standard deviation 75. Suppose $\bar{X}$ has standard deviation 13.5, based on random samples of size 25 taken without replacement. How large is the population?

7.50. Let $S = \{2, 4, 8, 16, 32\}$. Find the probability distribution of the sample mean $\bar{X}$ for samples of size 2 drawn without replacement.

7.51. Find the mean and standard deviation of $\bar{X}$ in Problem 7.50.

7.52. Repeat Problem 7.51 if the samples are drawn with replacement.

7.53. A population random variable X has mean 25 and standard deviation 5. Samples of size 40 are drawn with replacement. Find $P(24 \leq \bar{X} \leq 26)$.

7.54. Suppose the waiting time for a bus is a random variable with mean 8 minutes and standard deviation 4 minutes. In a given month, what is the probability that the average waiting time is less than 6 minutes?

7.55. Let X be a 4-place decimal number drawn at random from the interval [0, 10]. X has mean 5 and standard deviation 2.89. Suppose 100 numbers are drawn at random from the interval. What is the probability that the average of the numbers is between 4.8 and 5.2?

SAMPLE PROPORTION

7.56. Thirty-three percent of the first-year students at an urban university live in university housing. What are the mean and standard deviation of the proportion $\hat{P}$ of first-year students in university housing for all samples of size 50, drawn with replacement, from the population of first-year students?

7.57. With reference to Problem 7.56, suppose there are a total of 3970 first-year students. What are the mean and standard deviation of the proportion $\hat{P}$ if the samples are drawn without replacement?

7.58. With reference to Problem 7.56, what is the probability that between 15 and 18 of first-year students in a random sample of 50 live in university housing?

7.59. Show that if the random variable $\hat{P}$ is the sample proportion, with mean p and variance $\dfrac{p(1-p)}{n}$, corresponding to random samples of size n, then $n\hat{P}$ is a binomial random variable with mean np and variance $np(1-p)$.

7.60. In Problem 7.24, the probability distribution of the sample proportion $\hat{P}$ of Democrats in random samples of size 3, drawn with replacement from a population of three Democrats and two Republicans, was obtained by listing the frequency of all possible proportions in samples of size 3. Use the fact that $n\hat{P}$ is a binomial random variable (see Problem 7.59) to obtain the probability distribution $\hat{P}$ without listing all possible frequencies.

7.61. A population is broken down into two categories, A and B. Suppose the proportion of the population in category A is 0.7, and let $\hat{P}$ be the proportion in category A in random samples of size 5 drawn with replacement from the population. Use the fact that $n\hat{P}$ is a binomial random variable (Problem 7.59) to find the probability distribution of $\hat{P}$.

7.62. A population is broken down into two categories A and B, and p is the proportion in category A. Selecting a single individual from the population can be modeled as a Bernoulli random variable X, where $X = 1$ if the individual is in category A, and $X = 0$ if the individual is in category B. Show that the sample mean $\bar{X}$, corresponding to random samples of size n, is the proportion $\hat{P}$ of individuals in the sample that are in category A.

7.63. Since the random variable X in Problem 7.62 is a Bernoulli random variable, X has mean $\mu = p$ and standard deviation $\sigma = \sqrt{p(1-p)}$. Use these equations and the fact that $\bar{X} = \hat{P}$ to show that Theorem 7.4 follows from Theorem 7.1.

7.64. Use the equations in Problem 7.63 and the fact that $\bar{X} = \hat{P}$ (Problem 7.62) to show that Theorem 7.5 follows from Theorem 7.2.

7.65. Use the results of the previous two problems to show that Theorem 7.6 follows from Theorem 7.3′.

SAMPLE VARIANCE

7.66. Let $X_1, X_2, \ldots, X_n$ be n independent normal random variables, each with mean 20 and variance 4. Explain why $\dfrac{(X_1-20)^2}{4} + \dfrac{(X_2-20)^2}{4} + \cdots + \dfrac{(X_n-20)^2}{4}$ is a chi-square random variable with n degrees of freedom.

7.67. Let X_1, X_2, X_3 be three random variables, each with mean 25 and variance 7, and let $\bar{X}$ be the sample mean. Show that

$$\sum \frac{(X_i-\bar{X})^2}{7} = \sum \frac{(X_i-25)^2}{7} - \frac{(\bar{X}-25)^2}{7/3}$$

7.68. As stated in Example 7.12, the annual SAT scores are approximately normally distributed with mean $\mu = 500$ and standard deviation $\sigma = 100$. Let S^2 be the sample variance defined for random samples of 25 SAT scores. Find the mean and standard deviation of S^2.

7.69. With reference to Problem 7.68, for what value $\hat{S}^2$ of S^2 is $P(S^2 \leq \hat{S}^2) = 0.95$?

7.70. With reference to Problem 7.68, for what value $\hat{S}^2$ of S^2 is $P(S^2 \geq \hat{S}^2) = 0.95$?

Answers to Supplementary Problems

7.36. (*a*) $5^3 = 125$; (*b*) $\binom{5}{3} = 10$

7.37. $5 \cdot 4 \cdot 3 = 60$

7.38. (*a*) There is no such n since the number of samples drawn with replacement, which is 10^n, increases as n increases. (*b*) $\binom{10}{5} = 252$.

7.39. (*a*) There is no sample size n which gives the most samples drawn with replacement; the number of such samples, N^n, increases as n increases.
(*b*) If N is even, then the maximum number of samples of size n, drawn without replacement, occurs when $n = N/2$. If N is odd, then the maximum number occurs when $n = (N-1)/2$ and when $n = (N+1)/2$.

7.40. The probability of exactly n correct answers is $P(n) = \binom{5}{n} \times (0.5)^5$, which is a maximum when n is either 2 or 3, and is a minimum when n is either 0 or 5. Hence the most likely number of correct answers is 2 or 3, and the least likely is 0 or 5.

7.41. $1 - [P(0) + P(1)] = 1 - \left[\binom{20}{0}\binom{12}{4}\Big/\binom{32}{4} + \binom{20}{1}\binom{12}{3}\Big/\binom{32}{4}\right] \approx 0.86$

7.42. $1 - [P(0) + P(1)] = 1 - \left[\left(\frac{12}{32}\right)^4 + 4 \times \frac{20}{32} \times \left(\frac{12}{32}\right)^3\right] \approx 0.85$

7.43. $\mu_{\bar{X}} = 75$; $\sigma_{\bar{X}} = 8/\sqrt{25} = 1.6$

7.44. $\mu_{\bar{X}} = 75$; $\sigma_{\bar{X}} = \frac{8}{\sqrt{25}}\sqrt{\frac{N-25}{N-1}} = 1.6\sqrt{\frac{N-25}{N-1}}$, where N is the size of the population.

7.45. $P(72 \le \bar{X} \le 78) = P\left(\frac{72-75}{1.6} \le \frac{\bar{X}-75}{1.6} \le \frac{78-75}{1.6}\right) \approx P(-1.875 \le Z \le 1.875) \approx 0.94$

7.46. $P(72 \le \bar{X} \le 78) = P\left(\frac{72-75}{1.55} \le \frac{\bar{X}-75}{1.55} \le \frac{78-75}{1.55}\right) \approx P(-1.935 \le Z \le 1.935) \approx 0.95$

7.47. $P(\bar{X} \ge 525) = P\left(\frac{\bar{X}-500}{100/\sqrt{50}} \ge \frac{525-500}{100/\sqrt{50}}\right) \approx P(Z \ge 1.77) \approx 0.04$

7.48. $P(475 \le \bar{X} \le 525) = P\left(\frac{475-500}{100/\sqrt{n}} \le \frac{\bar{X}-500}{100/\sqrt{n}} \le \frac{525-500}{100/\sqrt{n}}\right) \approx P\left(-\frac{\sqrt{n}}{4} \le Z \le \frac{\sqrt{n}}{4}\right) = 0.95$ for $\frac{\sqrt{n}}{4} = 1.96$, or $n = 61.5$; round up to $n = 62$.

7.49. $\frac{75}{\sqrt{25}}\sqrt{\frac{N-25}{N-1}} = 13.5$; $\sqrt{\frac{N-25}{N-1}} = \frac{13.5}{15} = 0.9$; $N = 128$

7.50.

$\bar{x}$	3	5	6	9	10	12	17	18	20	24
$P(\bar{x})$	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

7.51. $\mu_{\bar{X}} = 12.4$; $\sigma_{\bar{X}} = 6.68$

7.52. $\mu_{\bar{X}} = 12.4$; $\sigma_{\bar{X}} = 7.715$

7.53. $P(24 \le \bar{X} \le 26) = P\left(\dfrac{24-25}{5/\sqrt{40}} \le \dfrac{\bar{X}-25}{5/\sqrt{40}} \le \dfrac{26-25}{5/\sqrt{40}}\right) \approx P(-1.26 \le Z \le 1.26) = 0.79$

7.54. For a 30 day month, $P(\bar{X} < 6) = P\left(\dfrac{\bar{X}-8}{4/\sqrt{30}} < \dfrac{6-8}{4/\sqrt{30}}\right) \approx P(Z < -2.74) = 0.003$

7.55. $P(4.8 \le \bar{X} \le 5.2) = P\left(\dfrac{4.8-5}{2.89/\sqrt{100}} \le \dfrac{\bar{X}-5}{2.89/\sqrt{100}} \le \dfrac{5.2-5}{2.89/\sqrt{100}}\right) \approx P(-0.69) \le Z \le 0.69) = 0.51$

7.56. $\mu_{\hat{P}} = 0.33$; $\sigma_{\hat{P}} = \sqrt{\dfrac{0.33(1-0.33)}{50}} = 0.0665$

7.57. $\mu_{\hat{P}} = 0.33$; $\sigma_{\hat{P}} = \sqrt{\dfrac{0.33(1-0.33)}{50}} \cdot \sqrt{\dfrac{3970-50}{3970-1}} = 0.0661$

7.58. $P\left(\dfrac{15}{50} \le \hat{P} \le \dfrac{18}{50}\right) = P\left(\dfrac{0.3-0.33}{0.0665} \le \dfrac{\hat{P}-0.33}{0.0665} \le \dfrac{0.36-0.33}{0.0665}\right) \approx P(-0.45 \le Z \le 0.45) = 0.35$

7.59. $n\hat{P}$ is the number of "successes" in n trials, where p is the probability of success, due to sampling with replacement, in each trial.

7.60. $p = \dfrac{3}{5} = 0.6$; $P(\hat{P} = 0) = P(3\hat{P} = 0) = (0.4)^3 = 0.064$

$P\left(\hat{P} = \dfrac{1}{3}\right) = P(3\hat{P} = 1) = 3 \times 0.6 \times (0.4)^2 = 0.288$

$P\left(\hat{P} = \dfrac{2}{3}\right) = P(3\hat{P} = 2) = 3 \times (0.6)^2 \times 0.4 = 0.432$

$P(\hat{P} = 1) = P(3\hat{P} = 3) = (0.6)^3 = 0.216$

7.61. $P(\hat{P} = 0) = P(5\hat{P} = 0) = (0.3)^5 = 0.002\,43$

$P\left(\hat{P} = \dfrac{1}{5}\right) = P(5\hat{P} = 1) = 5 \times 0.7 \times (0.3)^4 = 0.028\,35$

$P\left(\hat{P} = \dfrac{2}{5}\right) = P(5\hat{P} = 2) = 10 \times (0.7)^2 \times (0.3)^3 = 0.1323$

$P\left(\hat{P} = \dfrac{3}{5}\right) = P(5\hat{P} = 3) = 10 \times (0.7)^3 \times (0.3)^2 = 0.3087$

$P\left(\hat{P} = \dfrac{4}{5}\right) = P(5\hat{P} = 4) = 5 \times (0.7)^4 \times 0.3 = 0.360\,15$

$P(\hat{P} = 1) = P(5\hat{P} = 5) = (0.7)^5 = 0.168\,07$

7.62. $\bar{x} = \dfrac{x_1 + x_2 + \cdots + x_n}{n} = \dfrac{\text{number of } x_i\text{s equal to 1}}{n} = \hat{p}$

7.63. $\mu_{\hat{P}} = \mu_X = p$ by Theorem 7.1, and $\sigma_{\hat{P}} = \sigma_X = \dfrac{\sigma}{\sqrt{n}} = \sqrt{\dfrac{p(1-p)}{n}}$, also by Theorem 7.1.

7.64. $\mu_{\hat{P}} - \mu_X = p$ by Theorem 7.2, and $\sigma_{\hat{P}} = \sigma_X = \dfrac{\sigma}{\sqrt{n}} \cdot \sqrt{\dfrac{N-n}{N-1}} = \sqrt{\dfrac{p(1-p)}{n}} \cdot \sqrt{\dfrac{N-n}{N-1}}$, also by Theorem 7.2.

7.65. $\hat{P} = \bar{X}$, and $\bar{X}$ is approximately normally distributed with mean p and standard deviation $\sigma_X = \dfrac{\sigma}{\sqrt{n}} = \sqrt{\dfrac{p(1-p)}{n}} = \sigma_{\hat{P}}$ by Theorem 7.6.

7.66. $\dfrac{(X_1 - 20)^2}{4}, \dfrac{(X_2 - 20)^2}{4}, \ldots, \dfrac{(X_n - 20)^2}{4}$ are n independent normal random variables, each with mean 0 and standard deviation 1. By definition, their sum is a chi-square random variable with n degrees of freedom.

7.67.
$$\sum \frac{(X_i - 25)^2}{7} = \sum \frac{(X_i - \bar{X} + \bar{X} - 25)^2}{7}$$
$$= \sum \left[\frac{(X_i - \bar{X})^2}{7} + \frac{2(X_i - \bar{X})(\bar{X} - 25)}{7} + \frac{(\bar{X} - 25)^2}{7} \right]$$
$$= \sum \frac{(X_i - \bar{X})^2}{7} + \frac{2(\bar{X} - 25)}{7} \sum (X_i - \bar{X})^2 + \sum \frac{(\bar{X} - 25)^2}{7}$$
$$= \sum \frac{(X_i - \bar{X})^2}{7} + 0 + 3 \sum \frac{(\bar{X} - 25)^2}{7}$$

7.68. $\mu_{S^2} = \sigma^2 = 10{,}000$; $\sigma_{S^2} = [\sqrt{2/(n-1)}]\sigma^2 = \sqrt{\dfrac{2}{24}} \times 10{,}000 = \dfrac{5000}{\sqrt{3}}$

7.69. $P(S^2 \le \hat{S}^2) = P\left(\dfrac{24S^2}{10{,}000} \le \dfrac{24\hat{S}^2}{10{,}000}\right) \approx P(\chi^2(24) \le 0.0024\hat{S}^2) = 0.95$ for $0.0024\hat{S}^2 = 36.4$, or $\hat{S}^2 = 15{,}166.67$

7.70. $P(S^2 \ge \hat{S}^2) = P\left(\dfrac{24S^2}{10{,}000} \ge \dfrac{24\hat{S}^2}{10{,}000}\right) \approx P(\chi^2(24) \ge 0.0024\hat{S}^2) = 0.95$

$P(\chi^2(24) \le 0.0024\hat{S}^2) = 0.05$ for $0.0024\hat{S}^2 = 13.8$, or $\hat{S}^2 = 5750$