CHAPTER 5

SAMPLING DISTRIBUTIONS

SECTION 5.1

OVERVIEW

One of the most common situations giving rise to a **count *X*** is the **binomial setting**. It consists of four assumptions about how the count was produced. They are

- the number *n* of observations is fixed
- the *n* observations are all independent
- each observation falls into one of two categories called "success" and "failure"
- the probability of success *p* is the same for each observation

When these assumptions are satisfied, the number of successes *X* has a **binomial distribution** with *n* trials and success probability *p*, denoted by *B*(*n*, *p*). For smaller values of n, the probabilities for *X* can be found easily using statistical software. Table C in the text gives the probabilities for certain combinations of *n* and *p*, and there is also an exact formula. For large *n*, the **normal approximation** can be used.

For a large population containing a proportion *p* of successes, the binomial distribution is a good approximation to the number of successes in an SRS of size n, provided the population is at least 10 times larger than the sample. The mean and standard deviation for the binomial count X and the sample proportion $\hat{p} = X/n$ can be found using the formulas,

$$\mu_X = np \qquad\qquad \mu_{\hat{p}} = p$$

$$\sigma_X = \sqrt{np(1-p)} \qquad\qquad \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

When n is large the count X is approximately $N(np, \sqrt{np(1-p)})$ and the proportion $\hat{p}$ is approximately $N(p, \sqrt{\frac{p(1-p)}{n}})$. These approximations should work well when $np \geq 10$ and $N(1 - p) \geq 10$. For the count X, the **continuity correction** improves the approximation, particularly when the values of n and p are closer to these cutoffs.

The exact **binomial probability formula** is given by

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

where $k = 0, 1, 2, ..., n$ and $\binom{n}{k}$ is the **binomial coefficient**

$$\binom{n}{k} = \frac{n!}{k!(n-k)1}$$

and the factorial $n!$ is

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 3 \times 2 \times 1$$

GUIDED SOLUTIONS

Exercise 5.1

KEY CONCEPTS - binomial setting

There are four assumptions that need to be satisfied to ensure that the count X has a binomial distribution. The number of observations or trials is fixed in advance, each trial results in one of two outcomes, the trials are independent and the probability of success is the same from trial to trial. In addition, for a large population with a proportion p of successes, we can use the binomial distribution as an approximation to the distribution of the count X of successes in an SRS of size *n*. For each setting below, see if all four assumptions are satisfied.

a) Think about how this fits in the binomial setting. What is *n*? What are the two outcomes, and why might the trials be considered independent?

b) Think about how many observations or trials there are going to be.

c) Think about whether the trials are independent.

Exercise 5.5

KEY CONCEPTS - binomial probabilities

a) Suppose we let X denote the number of members of the committee that are Hispanic. We are in the binomial setting with $n = 15$ trials, and letting "success" correspond to being Hispanic, we have $p = 0.3$. We want to compute the probability that *exactly* 3 members of the committee are Hispanic, namely we want $P(X = 3)$. Table C can be used to compute this probability. Find the column labelled by 0.30 and look for the row corresponding to $n = 15$ and $k = 3$. Alternatively, you can use statistical software to compute this probability. What do you find?

$P(X = 3) =$

b) Now we want the probability that 3 or fewer members of the committee are Hispanic. Express this probability in terms of X and add the appropriate entries in Table C to compute this probability. Alternatively, you may wish to use statistical software to compute the probability.

Exercise 5.11

KEY CONCEPTS - mean of the binomial, normal approximation for counts

a) Let X denote the number of home runs Mark McGwire will hit in 509 times at bat. The problem tells us to treat X as a count having the $B(509, 0.116)$ distribution. We are asked to compute the mean μ_X of X. What is the formula for μ_X? To answer this, see the Chapter Overview above, or the Summary of Section 5.1 in the text. Now use this formula to compute the mean.

$\mu_X =$

b) Now we are asked to compute $P(X \geq 70)$. This can be done using the normal approximation for a count. First check to see that np and $n(1-p)$ are both greater than or equal to 10. Next we need the mean and standard deviation of X. You computed the mean in a). To compute the standard deviation σ_X, use the formula given in the Chaper Overview on the previous page, or the Summary of Section 5.1 in the text. Write the result below.

$\sigma_X =$

Next use the the normal approximation to evaluate $P(X \geq 70)$. This is a normal probability calculation like those discussed in Section 3 of Chapter 1. You may want to review the material there to refresh your memory on how to do such calculations. The first step is to standardize the number 70 (compute its z-score) by subtracting the mean μ_X and dividing the result by σ_X. Next, use Table A to determine the area to the right of this z-score. It may be helpful to draw a normal curve to visualize the area.

$P(X \geq 70) =$

c) To answer this question, we now let X denote the number of home runs Barry Bonds will hit in 476 times at bat. The problem tells us to treat X as a count having the $B(476, 0.0865)$ distribution. We are asked to compute $P(X \geq 73)$. To do so you will again need to use the normal approximation. First compute

$\mu_X =$

$\sigma_X =$

and then use the same method as in (b) to complete the calculation.

$P(X \geq 73) =$

Exercise 5.13

KEY CONCEPTS - proportions, normal approximation for proportions

a) Let X denote the number of students in the SRS of $n = 100$ that support a crackdown on underage drinking. In this case, $X = 62$. To compute the sample proportion use the formula sample proportion = $\hat{p} = X/n =$

b) If the proportion of all students on your campus who support a crackdown is $p = 0.67$, then X is a count having the $B(100, 0.67)$ distribution. You need to compute a probability about $\hat{p}$, namely $P(\hat{p} \leq 0.62)$, because 0.62 is the sample proportion that resulted from the administration's sample. Since $\hat{p} \leq 0.62$ whenever $X \leq 62$, if you have statisticial software which computes binomial probabilities, try to find the exact probability.

If you don't have access to statistical software that computes binomial probabilities, you'll need to use the normal approximation to the sampling distribution of $\hat{p}$ to approximate the probability. Since $np = 100(0.67) = 67$ and $n(1 - p) = 100(0.33) = 33$ are both greater than or equal to 10, we can use the approximation. To use the normal approximation, the mean and standard deviation of $\hat{p}$ must be computed. Do this using the formulas given below.

mean $\mu_{\hat{p}} = p =$

standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} =$

Computing $P(\hat{p} \leq 0.62)$ is a normal probability calculation like those in Section 3 of Chapter 1. You may want to review the material there in order to refresh your memory on how to do such calculations, as they form the basis of solving many of the exercises in this chapter. To compute the probability, we standardize the number 0.62 (compute its z-score) by subtracting the mean $\mu_{\hat{p}}$ and dividing the result by $\sigma_{\hat{p}}$. Next use Table A to determine the area to the left of this z-score. It may be helpful to draw a normal curve and the desired area to help you solve byvisualizing the area.

$P(\hat{p} \leq 0.62) =$

c) You should base your comments on the probability calculated in (b). In particular, is this probability large enough that a value as small as 0.62 could plausibly arise by chance if the actual value is 0.67?

Exercise 5.22

KEY CONCEPTS - binomial probabilities, mean and variance of the binomial

a) Let X = number of truthful persons "failing" the lie detector test. X has a binomial distribution with n = 12 and p = 0.2. The binomial probabilities can be found in Table C in your text or using statistical software. The probabilities below were obtained using statistical software.

k	$P(X = k)$
0	0.068719
1	0.206158
2	0.283468
3	0.236223
4	0.132876
5	0.053150
6	0.015502
7	0.003322
8	0.000519
9	0.000058
10	0.000004
11	0.000000
12	0.000000

Use these (or those in Table C) to evaluate the probability required.

b) Use the formula which expresses the mean and standard deviation of the binomial in terms of n and p.

mean =
standard deviation =

c) Don't be confused by the fact that the mean is not a whole number. Just determine which values of X are less than the mean and then add up their probabilities.

Exercise 5.23

KEY CONCEPTS - normal approximation for counts

a) The population proportion is

$$p = \frac{\text{number of blacks in the 2000 census}}{\text{total number of adults}}$$

Use the numbers given in the statement of the problem to compute p.

$p =$

b) Let X denote the number of blacks in the sample. X is a count and if the sample is a random sample, X will have a binomial distribution. What are n and

p in this case? Now use the formulas which express the mean of X in terms of n and p.

mean =

c) This part can be done using the normal approximation for a count. First check to see that np and $n(1-p)$ are both greater than or equal to 10. Next compute the standard deviation of X using the formula that expresses the standard deviation in terms of n and p.

standard deviation =

Now use the mean from (b), this standard deviation, and the normal approximation to evaluate $P(X \leq 170)$.

$P(X \leq 170)$ =

COMPLETE SOLUTIONS

Exercise 5.1

a) The number of trials 50 is fixed, each child is a boy or girl, whether or not a child is a boy will not alter the probability that any other children are either girls or boys (excluding identical twins to keep it simple), and the probability of any child being a boy should be the same (about 0.5) for each birth. The binomial distribution should be a good probability model for the number of girls.

b) Although each birth is a boy or girl, we are not counting the number of successes in a fixed number of births. The number of observations (births) is random. The assumption of a fixed number of observations is violated.

c) This is not the same as interviewing 50 people at random. The husband and wife may tend to share the same opinion, and the trials will not be independent of each other. If we know the wife agrees, then it is more likely that the husband agrees than if we didn't have this information. The assumption of independent observations is violated.

Exercise 5.5

a) From Table C we find $P(X = 3) = .1700$.

b) We want

$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$

Using Table C to find each of the probabilities on the right hand side of this equation we find

$$P(X \le 3) = .0047 + .0305 + .0916 + .1700 = .2968$$

Exercise 5.11

a) $\mu_X = np = 509 \times 0.116 = 59.044$.

b) First we check that

$np = 509 \times 0.116 = 59.044 \ge 10$, and $n(1\text{-}p) = 509 \times 0.884 = 449.956 \ge 10$

Next we compute

$$\mu_X = np = 509 \times 0.116 = 59.044$$

$$\sigma_X = \sqrt{np(1-p)} = \sqrt{509 \times 0.116 \times 0.884} = \sqrt{52.195} = 7.225$$

When n is large, X is approximately $N(np, \sqrt{np(1-p)}) = N(59.044, 7.225)$. The z-score of 70 is thus

$$z\text{-score of } 70 = \frac{70 - 59.044}{7.225} = 1.52$$

and so using Table A,

$$P(X \ge 70) = P(z \ge 1.52) = 1 - P(z \le 1 .52) = 1 - 0.9357 = 0.0643.$$

c) We compute

$$\mu_X = np = 476 \times 0.0865 = 41.174$$

$$\sigma_X = \sqrt{np(1-p)} = \sqrt{476 \times 0.0865 \times 0.9135} = \sqrt{37.612} = 6.133$$

and so the z-score of 73 is

$$z\text{-score of } 73 = \frac{73 - 41.174}{6.133} = 5.19$$

and so using Table A,

$$P(X \ge 73) = P(z \ge 5.19) = 1 - P(z \le 5.19) \le 1 - 0.9999 = 0.0001$$

Exercise 5.13

a) Sample proportion = $\hat{p} = X/n = \dfrac{62}{100} = 0.62$.

b) Using statistical software we find the exact probability of $X \leq 62$ to be 0.1690. To use the normal approximation we compute

$$\text{mean} = \mu_{\hat{p}} = p = 0.67$$

$$\text{standard deviation} = \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.67 \times 0.33}{100}} = \sqrt{0.0022} = 0.047$$

From this, we compute the *z*-score of 0.62 to be

$$z\text{-score} = \frac{0.62 - 0.67}{0.047} = -1.06$$

Using Table A we find

$$P(\hat{p} \leq 0.62) = P(z \leq -1.06) = 0.1446$$

(Note: This differs somewhat from the exact value obtained using statistical software. If you use the continuity correction, you get $P(\hat{p} \leq 0.62) = 0.1685$.)

c) It is true that the sample proportion of 0.62 is smaller than the national value of 0.67. However, our probability calculation shows that a sample value of 0.62 (or smaller) has a probability 0.1446 of occurring simply by chance if the actual proportion on campus is 0.67. In most scientific journals this probability would be considered too large to allow one to assert with any degree of confidence that the survey supports the statement "that support for a crackdown is lower on our campus than nationally."

Exercise 5.22

a) X is B(12, 0.2). You are asked to evaluate $P(X \geq 1)$, the probability that the polygraph says that at least one person telling the truth is deceptive. Using the table of probabilities

k	$P(X = k)$
0	0.068719
1	0.206158
2	0.283468
3	0.236223
4	0.132876
5	0.053150
6	0.015502
7	0.003322
8	0.000519
9	0.000058
10	0.000004

```
11   0.000000
12   0.000000
```

you may at first think to add the probabilites for $k = 1, 2, 3, ..., 12$. While this will give the correct answer, in this case it is much simpler to use the rule for complements. $P(X \geq 1) = 1 - P(X = 0) = 1 - 0.068719 = 0.931281$. When determing the complementary event, you must be careful whether an event is of the form greater than or greater than or equal to. The complement of $X \geq 2$ is $X \leq 1$, while the complement of $X > 2$ is $X \leq 2$.

b) The mean is $np = 12(0.2) = 2.4$, and the standard deviation is found using the formula $\sqrt{np(1-p)} = \sqrt{12(0.2)(0.8)} = 1.386$.

c) X is less than the mean 2.4 only if X is 0, 1 or 2. So, we need to find

$$P(X \leq 2) = 0.068719 + 0.206158 + 0.283468 = 0.558345$$

Exercise 5.23

a) $p = \dfrac{23,772,494}{209,128,094} = 0.1137$.

b) X has a binomial distribution with $n = 1500$ and $p = 0.1137$. The mean of X is

$$\text{mean} = np = 1500(0.1137) = 170.55.$$

c) The normal approximation can be used since both $np = 170.55$ and $n(1-p) = 1329.45$ are greater than 10. We compute

$$\text{standard deviation} = \sqrt{np(1-p)} = \sqrt{1500(0.1137)(0.8863)} = 12.29$$

Using the mean in (b) and this standard deviation gives the approximation

$$P(X \leq 170) = P\left(\frac{X - 170.55}{12.29} \leq \frac{170 - 170.55}{12.29}\right) = P(z \leq -0.04) = 0.4840$$

SECTION 5.2

OVERVIEW

This section examines properties of the **sample mean** $\bar{x}$. If we select an SRS of size n from a large population with mean μ and standard deviation σ, the sample mean $\bar{x}$ has a sampling distribution with

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

and

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

This implies that the sample mean is an unbiased estimator of the population mean and is less variable than a single observation.

Linear combinations (such as sums or means) of independent normal random variables have normal distributions. In particular, if the population has a normal distribution, the sampling distribution of $\bar{x}$ is normal. Even if the population does not have a normal distribution, for large sample sizes the sampling distribution of $\bar{x}$ computed from an SRS is approximately normal. In particular, the **central limit theorem** states that for large n, the sampling distribution of $\bar{x}$ computed from an SRS is approximately $N(\mu, \frac{\sigma}{\sqrt{n}})$ for any population with mean μ and finite standard deviation σ.

SAMPLE PROBLEMS

GUIDED SOLUTIONS

EXERCISE 5.33

KEY CONCEPTS - the sampling distribution of the sample mean, normal probability calculations.

a) We take X to be the (numeric) grade of a randomly chosen student. The probability distribution of X is seen to be

Grade (X)	4	3	2	1	0
Probability	0.18	0.32	0.34	0.09	0.07

Refer to Section 4.4 for the formulas for the mean and standard deviation of a discrete random variable. Use these to compute the following.

μ = mean of X=

σ = standard deviation of X =

b) Recall that if we select an SRS of size n from a large population with mean μ and standard deviation σ, the sample mean $\bar{x}$ has a sampling distribution with mean of $\bar{x} = \mu_{\bar{x}} = \mu =$

and

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

In this case, what is n, the sample size?

c) You can compute the probability $P(X \geq 3)$ that a randomly chosen Accounting 201 student gets a grade of B or better directly using the probabilility distribution in part (a). This is just like the calculation in Section 4.3.

$P(X \geq 3) =$

To compute the approximate probability $P(\bar{x} \geq 3)$ that the grade point average for 50 randomly chosen Accounting 201 students is B or better, we use the central limit theorem that states that for large n, the sampling distribution of $\bar{x}$ computed from an SRS is approximately $N(\mu, \frac{\sigma}{\sqrt{n}})$ for any population with mean μ and finite standard deviation σ. Thus, calculating $P(\bar{x} \geq 3)$ is just a normal probability calculation, like those we did in Section 3 of Chapter 1. You may wish to review the material there in order to refresh your memory as to how to do such calculations. We want the probability that the sample mean $\bar{x}$ is 3 or higher. To compute this probability, we standardize the number 3 (compute its z-score) by subtracting the mean $\mu_{\bar{x}}$ and dividing the result by $\sigma_{\bar{x}}$. You computed $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$ in (b). Next use Table A to determine the area to the right of this z-score under a standard normal curve. You may wish to draw a picture of the standard normal to help you visualize the desired area.

$P(\bar{x} \geq 3) =$

EXERCISE 5.36

KEY CONCEPTS - means and standard deviations of random variables, the law of large numbers, the central limit theorem, the sampling distribution of the sample mean, normal probability calculations

a) You may wish to review how to calculate the mean and standard deviation of a random variable. This was discussed in Chapter 4. Recall that if X is a discrete random variable having possible values $x_1, x_2, \ldots, x_k$ with corresponding probabilities p_1, p_2,…, p_k, the mean μ_X is the average of the possible values weighted by the corresponding probabilities, i.e.

$$\mu_X = x_1 p_1 + x_2 p_2 + \ldots + x_k p_k.$$

the variance is

$$\sigma^2_X = (x_1 - \mu_X)^2 p_1 + (x_2 - \mu_X)^2 p_2 + \ldots + (x_k - \mu_X)^2 p_k$$

and the standard deviation σ_X is the positive square root of the variance.

For simplicity, assume that the gambler bets on red. What is the probability he will win, assuming all slots are equally likely to contain the ball? Complete the table below to help you compute the mean and standard deviation of the outcomes.

Outcome (winnings) X	\$1	-\$1
Probability		

$\mu_X =$

$\sigma^2_X =$

$\sigma_X =$

b) Recall that the law of large numbers (see Chapter 4) tells us that the average of the values of X observed in many trials must approach μ_X. Interpret this in the context of this problem, using plain English.

c) The central limit theorem states that for large n, the sampling distribution of the mean winnings computed from an SRS is approximately $N(\mu_X, \frac{\sigma_X}{\sqrt{n}})$. What are n, μ_X, and σ_X here?

Refer to Chapter 1 to refresh your memory concerning the 68-95-99.7 rule. Remember to use μ_X for the mean and $\frac{\sigma_X}{\sqrt{n}}$ for the standard deviation when applying the rule.

d) From (c) we found that the sampling distribution of the gambler's mean winnings $\bar{x}$ if he makes 50 bets is $N(\mu_X = -0.0526, \frac{\sigma_X}{\sqrt{n}} = 0.1412)$. We want to compute the probability that $\bar{x}$ is less than 0. This is just a normal probability calculation of the sort we studied in Section 3 of Chapter 1. Find the *z*-score of 0 and then use Table A to determine the area under the standard normal curve below this *z*-score.

e) Now repeat the same sort of calculations as in (c) and (d), but with $n = 100{,}000$ rather than $n = 50$.

EXERCISE 5.43

KEY CONCEPTS - the sampling distribution of the sample mean, normal probability calculations

We are told that the distribution of individual scores at Southwark Elementary School is approximately normal with mean $\mu = 13.6$ and standard deviation $\sigma = 3.1$. To find L, Mr. Lavin needs to first determine the sampling distribution of the mean score $\bar{x}$ of $n = 22$ children. What is this sampling distribution?

To complete the problem, you need to find L such that the probability of $\overline{x}$ being below L is only 0.05. We did this sort of problem in Section 3 of Chapter 1. First, refer to Table A to find the value z such that the area to the left of z under a standard normal curve is 0.05. What is this value?

This value z is the z-score of L. This means $z = (L - \mu_{\overline{x}})/\sigma_{\overline{x}}$, where $\mu_{\overline{x}}$ and $\sigma_{\overline{x}}$ are the mean and standard deviation, respectively, of the sampling distribution of $\overline{x}$. Solve this equation for L.

EXERCISE 5.45

KEY CONCEPTS - linear combinations of independent random variables

a) We need to recall several facts from Section 4 of Chapter 4. First, recall that for a random variable X

$$\mu_{a+bX} = a + b\mu_X$$

Using a = 0 and b = -1, this implies that $\mu_{-X} = -\mu_X$. Second, recall that for random variables X and Y

$$\mu_{Y+X} = \mu_Y + \mu_X$$

Replacing X by the random variable $-X$ and using $\mu_{-X} = -\mu_X$ we have

$$\mu_{Y-X} = \mu_Y - \mu_X$$

Finally, recall that if X and Y are independent random variables, then

$$\sigma^2_{Y-X} = \sigma^2_Y + \sigma^2_X.$$

We now apply these facts to $\overline{x}$ and $\overline{y}$ to get

$$\mu_{\overline{y}-\overline{x}} = \mu_{\overline{y}} - \mu_{\overline{x}}$$

and

$$\sigma^2_{\bar{y}-\bar{x}} = \sigma^2_{\bar{y}} + \sigma^2_{\bar{x}}$$

We take the square root of $\sigma^2_{\bar{y}-\bar{x}}$ to get the standard deviation of $\bar{y}$ - $\bar{x}$. To complete the problem, what are $\mu_{\bar{y}}$, $\mu_{\bar{x}}$, $\sigma^2_{\bar{y}}$, and $\sigma^2_{\bar{x}}$ here?

b) We know that since weight gains are normally distributed in both populations, $\bar{x}$ and $\bar{y}$ are normally distributed. You determined the mean and standard deviations of the sampling distributions of these means in (a). Since we can assume $\bar{x}$ and $\bar{y}$ are independent, we can assume the sampling distribution of $\bar{y}$ - $\bar{x}$ is also normal. You determined the mean and standard deviation of the sampling distribution of $\bar{y}$ - $\bar{x}$ in (a).

c) Here we want to determine the probability that $\bar{y}$ - $\bar{x}$ is greater than 25. We found that the sampling distribution of $\bar{y}$ - $\bar{x}$ is $N(\mu_{\bar{y}-\bar{x}} = 25, \sigma_{\bar{y}-\bar{x}} = 16.62)$ in (b). The desired probability can be computed using the methods discussed in Section 3 of Chapter 1 for calculating normal probabilities.

COMPLETE SOLUTIONS

EXERCISE 5.33

a) μ = mean of X = $4\times 0.18 + 3\times 0.32 + 2\times 0.34 + 1\times 0.09 + 0\times 0.07 = 2.45$

To compute the standard deviation we first compute the variance

$$\sigma^2 = (4\text{-}2.45)^2 \times 0.18 + (3\text{-}2.45)^2 \times 0.32 + (2\text{-}2.45)^2 \times 0.34 + (1\text{-}2.45)^2 \times 0.09$$

$$+ (0\text{-}2.45)^2 \times 0.07 = 1.2075$$

and so

$$\sigma = \text{standard deviation of } X = \sqrt{1.2075} = 1.099$$

b) Using the results in part (a) and the fact that the sample size is n = 50 we obtain

$$\text{mean of } \bar{x} = \mu_{\bar{x}} = \mu = 2.45$$

$$\text{standard deviation of } \bar{x} = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{1.099}{\sqrt{50}} = 0.155$$

c) $P(X \geq 3) = P(X = 3) + P(X = 4) = 0.32 + 0.18 = 0.50$.

Using the results of (b), the sampling distribution of $\bar{x}$ is approximately $N(2.45, 0.155)$. Thus

$$P(\bar{x} \geq 3) = P\left(\frac{\bar{x} - 2.45}{0.155} \geq \frac{3 - 2.45}{0.155}\right) = P(z \geq 3.55) = 1 - P(z \leq 3.55)$$

The value 3.55 is outside the range of Table A, and all we can say is that $P(z \leq 3.55)$ is larger than 0.9998, hence $P(\bar{x} \geq 3)$ is smaller than 1 - 0.9998 = 0.0002.

EXERCISE 5.36

a) Since 18 of the 38 slots are red, the probability of winning (if you bet on red) is 18/38. Thus we have

Outcome (winnings) X	\$1	-\$1
Probability	18/38	20/38

and so

mean winnings = μ_X = (\$1)(18/38) + (-\$1)(20/38) = -\$2/38 = -\$0.0526

variance of winnings (in dollars squared) = σ^2_X
= $(1 - [-2/38])^2(18/38) + (-1 - [-2/38])^2(20/38)$
= 0.5249 + 0.4724
= 0.9973

standard deviation of winnings = σ_X = $\$\sqrt{0.9973}$ = \$ 0.9986

b) The law of large numbers tells us that if a gambler makes a large number of bets on red, his mean winnings per bet will be approximately $\mu_X = -\$0.0526$. In other words, he will lose $0.0526 on the average per bet.

c) Here $n = 50$, $\mu_X = -\$0.0526$, and $\sigma_X = \$0.9986$. The approximate distribution of the gambler's mean winnings in 50 bets is thus normal with

$$\text{mean} = \mu_X = -\$0.0526$$

and

$$\text{standard deviation} = \frac{\sigma_X}{\sqrt{n}} = \frac{\$0.9986}{\sqrt{50}} = \$0.1412$$

The 68-95-99.7 rule says that the middle 95% of the gambler's mean winnings on nights when he places 50 bets is within two standard deviations of the mean. In other words, it lies between $\mu_X \pm 2\frac{\sigma_X}{\sqrt{n}} = -\$0.0526 \pm 2(\$0.1412)$ or between -$0.3350 and $0.2298. Multiplying by 50 to convert to total winnings, we get that the middle 95% of the gambler's total winnings on nights when he places 50 bets is between -$16.75 and $11.49.

d) The sampling distribution of the gambler's mean winnings $\bar{x}$ if he makes 50 bets is $N(\mu_X = -0.0526, \frac{\sigma_X}{\sqrt{n}} = 0.1412)$. To find the probability that $\bar{x}$ is less than 0, we compute the z-score of 0, which is

$$z\text{-score} = (0 - [-0.0526])/(0.1412) = 0.0526/0.1412 = 0.37$$

According to Table A, the area under the standard normal curve to the left of 0.37 is 0.6443. Thus the probability that the gambler will lose money if he makes 50 bets is 0.6443.

e) If $n = 100{,}000$ bets are made on red, the sampling distribution of the mean winnings is again normal but now with

$$\text{mean} = \mu_X = -\$0.0526$$

and

$$\text{standard deviation} = \frac{\sigma_X}{\sqrt{n}} = \frac{\$0.9986}{\sqrt{100{,}000}} = \$0.00316$$

The 68-95-99.7 rule says that the middle 95% of the mean winnings of gamblers on these 100,000 bets is within two standard deviations of the mean.

In other words, lies between $\mu_X \pm 2\frac{\sigma_X}{\sqrt{n}}$ = -\$0.0526 ± 2(\$0.00316) or between -\$0.05892 and -\$0.04628. Multiplying by 100,000 to convert to total winnings, we get that the middle 95% of the mean winnings on 100,000 bets is between -\$5892 and -\$4628.

EXERCISE 5.43

The sampling distribution of the mean score $\bar{x}$ of 22 children is approximately normal with mean

$$\mu_{\bar{x}} = \mu = 13.6$$

and standard deviation

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3.1}{\sqrt{22}} = 0.66$$

Next, we note that from Table A the value of z such that the area to the left of it under a standard normal curve is 0.05 is z = -1.65. Thus

$$-1.65 = (L - 13.6)/0.66$$

Solving for L gives

$$L = (-1.65)(0.66) + 13.6 = 12.51$$

EXERCISE 5.45

a) We know that the mean and standard deviation of X are 360g and 55g and the mean and standard deviation of Y are 385g and 50g. For an average based on a sample size of n = 20, we have

$$\mu_{\bar{x}} = 360$$

$$\sigma_{\bar{x}}^2 = (55)^2/(20) = 151.25$$

$$\mu_{\bar{y}} = 385$$

$$\sigma_{\bar{y}}^2 = (50)^2/(20) = 125$$

Therefore

$$\mu_{\bar{y}-\bar{x}} = \mu_{\bar{y}} - \mu_{\bar{x}} = 385 - 360 = 25$$

$$\sigma^2_{\bar{y}-\bar{x}} = \sigma^2_{\bar{y}} + \sigma^2_{\bar{x}} = 151.25 + 125 = 276.25$$

$$\sigma_{\bar{y}-\bar{x}} = \sqrt{276.25} = 16.62.$$

b) From (a) we have

$$\mu_{\bar{x}} = 360$$

$$\sigma^2_{\bar{x}} = (55)^2/(20) = 151.25 \text{ (hence } \sigma_{\bar{x}} = \sqrt{151.25} = 12.30)$$

$$\mu_{\bar{y}} = 385$$

$$\sigma^2_{\bar{y}} = (50)^2/(20) = 125 \text{ (hence } \sigma_{\bar{y}} = \sqrt{125} = 11.18)$$

$$\mu_{\bar{y}-\bar{x}} = 25$$

$$\sigma_{\bar{y}-\bar{x}} = 16.62.$$

Thus,

the distribution of $\bar{x}$ is $N(\mu_{\bar{x}} = 360, \sigma_{\bar{x}} = 12.30)$

the distribution of $\bar{y}$ is $N(\mu_{\bar{y}} = 385, \sigma_{\bar{y}} = 11.18)$

the distribution of $\bar{y} - \bar{x}$ is $N(\mu_{\bar{y}-\bar{x}} = 25, \sigma_{\bar{y}-\bar{x}} = 16.62)$.

c) To determine the probability that $\bar{y} - \bar{x}$ is greater than 25, we compute the z-score of 25. We get

$$z\text{-score} = (25 - \mu_{\bar{y}-\bar{x}})/\sigma_{\bar{y}-\bar{x}} = (25 - 25)/16.62 = 0.$$

The area to the right of 0 under a standard normal curve is 0.5, so this is the probability that $\bar{y} - \bar{x}$ is greater than 25.