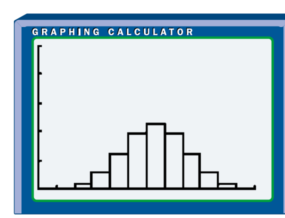
8.4 Normal Approximation to the Binomial Distribution

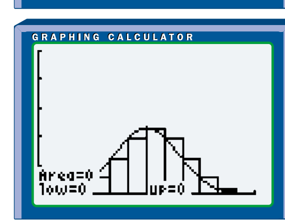
The normal distribution is a continuous distribution. Many real-life situations involve discrete data, such as surveys of people or testing of units produced on an assembly line. As you saw in section 8.3, these situations can often be modelled by a normal distribution. If the discrete data have a binomial probability distribution and certain simple conditions are met, the normal distribution makes a very good approximation. This approximation allows the probabilities of value ranges to be calculated more easily than with the binomial formulas.

INVESTIGATE & INQUIRE: Approximating a Binomial Distribution

1. On your graphing calculator, enter the integers from 0 to 12 in L1. This list represents the number of successes in 12 trials.
2. With the cursor on L2, enter the **binompdf(function**. Use 12 for the number of trials and 0.5 for the probability of success. The calculator will place into L2 the binomial probabilities for each number of successes from L1.
3. Use **STAT PLOT** to construct a histogram using L1 as Xlist and L2 as Freq. Your result should look like the screen on the right.



4. Now, construct a normal distribution approximation for this binomial distribution. From the DISTR menu, select the **ShadeNorm(function** and use 0, 0, 6, and $\sqrt{3}$ as the parameters. The normal approximation should now appear, superimposed on the binomial histogram.



5. Investigate the effect of changing the binomial probability of success, p. Use p values of 0.3, 0.1, 0.95, and 0.7. Keep the number of trials at 12 and repeat steps 2 through 4 for each value of p. In step 4, use the **ShadeNorm(function** and enter 0, 0, $12p$, $\sqrt{12pq}$ for each value of p. You will have to adjust the **window settings** for some of these situations.

6. For each different p value in step 5, make a subjective estimate of how good an approximation the normal distribution is for the underlying binomial distribution. Summarize your findings.
7. Statistical theory states that for a binomial distribution with n trials and probability of success p, a normal distribution with $\mu = np$ and $\sigma = \sqrt{npq}$ is a reasonable approximation as long as both np and nq are greater than 5. Recall that $q = 1 - p$ is the probability of failure. Do your results meet these criteria?

Example 1 Bank Loans

A bank found that 24% of its loans to new small businesses become delinquent. If 200 small businesses are selected randomly from the bank's files, what is the probability that at least 60 of them are delinquent? Compare the results from the normal approximation with the results from the calculations using a binomial distribution.

Solution 1: Using a Normal Distribution Table or Graphing Calculator

Here,
$$np = (200)(0.24) = 48 > 5 \quad \text{and} \quad nq = (200)(0.76) = 152 > 5$$

Therefore, the normal approximation should be reasonable.

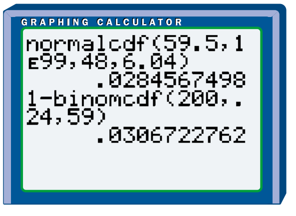
$$\begin{aligned} \mu &= np \\ &= (200)(0.24) \\ &= 48 \end{aligned} \qquad \begin{aligned} \sigma &= \sqrt{npq} \\ &= \sqrt{(200)(0.24)(0.76)} \\ &= 6.04 \end{aligned}$$

Using the normal approximation with continuity correction, and referring to the table of Areas Under the Normal Distribution Curve on pages 606 and 607, the required probability is

$$\begin{aligned} P(X > 59.5) &= P\left(Z > \frac{59.5 - 48}{6.04}\right) \\ &= P(Z > 1.90) \\ &= 1 - P(Z < 1.90) \\ &= 0.029 \end{aligned}$$

Using a graphing calculator, this normal approximation probability can be calculated using the function normalcdf(59.5,1E99,48,6.04). Alternatively, the binomial probability can be calculated as 1–binomcdf(200,.24,59). The calculator screen shows the results of the two methods.

The normal approximation gives a result of 0.0285. The binomial cumulative density function gives 0.0307. The results of the two methods are close enough to be equal to the nearest percent. So, the probability that at least 60 of the 200 loans are delinquent is approximately 3%.



Solution 2: Using Fathom™

Open Fathom, and open a new document if necessary. Drag a new **collection** box to the workspace. Rename it Loans.

Double click on the collection to open the **inspector**. Choose the Measures tab.
Create two measures: Binomial and Normal.
Edit the Binomial formula to 1 – binomialCumulative(59,200,.24).
Edit the Normal formula to 1 – normalCumulative(59.5,48,6.04).
These functions give the same probability values as the equivalent graphing calculator functions.

Example 2 Market Share

QuenCola, a soft-drink company, knows that it has a 42% market share in one region of the province. QuenCola's marketing department conducts a blind taste test of 70 people at the local mall.

a) What is the probability that fewer than 25 people will choose QuenCola?

b) What is the probability that exactly 25 people will choose QuenCola?

Solution 1 Using a Normal Distribution Table

a) For this binomial distribution,

$$\begin{aligned} np &= 70(0.42) \\ &= 29.4 \\ &> 5 \end{aligned} \quad \text{and} \quad \begin{aligned} nq &= 70(0.58) \\ &= 40.6 \\ &> 5 \end{aligned}$$

Therefore, you can use the normal approximation.

$$\begin{aligned} \mu &= np \\ &= (70)(0.42) \\ &= 29.4 \end{aligned} \qquad \begin{aligned} \sigma &= \sqrt{npq} \\ &= \sqrt{(70)(0.42)(0.58)} \\ &= 4.13 \end{aligned}$$

Using the normal approximation with continuity correction,

$$\begin{aligned} P(X < 24.5) &= P(Z < -1.19) \\ &= 0.117 \end{aligned}$$

So, there is a 12% probability that fewer than 25 of the people surveyed will choose QuenCola.

b) The probability of exactly 25 people choosing QuenCola can also be calculated using the normal approximation.

$$\begin{aligned} P(24.5 < X < 25.5) &= P(-1.19 < Z < -0.94) \\ &= P(Z < -0.94) - P(Z < -1.19) \\ &= 0.1736 - 0.1170 \\ &= 0.057 \end{aligned}$$

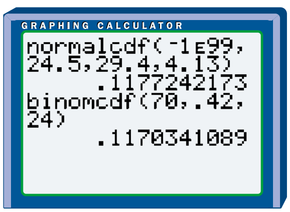
The probability that exactly 25 people will choose QuenCola is approximately 5.7%.

Solution 2: Using a Graphing Calculator

a) To find the probability that 24 or fewer people will choose QuenCola, you need to use a parameter of 24.5 for the normal approximation, but 24 for the cumulative binomial function. The **normalcdf(** and **binomcdf(functions** differ by less than 0.001.

$P(X_{\text{normal}} < 24.5) = 0.118$
$P(X_{\text{binomial}} \leq 24) = 0.117$

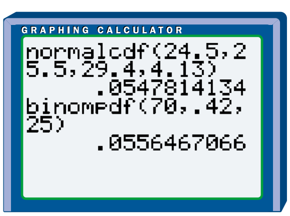
There is a 12% probability that fewer than 25 of the people surveyed will choose this company's product.



b) The **normalcdf(** and **binompdf(functions** return the following values.

$P(24.5 < X_{\text{normal}} < 25.5) = 0.0548$

$P(X_{\text{binomial}} = 25) = 0.0556$



Here, the normal approximation is a bit smaller than the value calculated using the actual binomial distribution. The manual calculation gave a slightly higher value because of rounding. The probability that exactly 25 of the people surveyed will choose this company's product is 5.6%.

Key Concepts

- A discrete binomial probability distribution can be approximated with a continuous normal distribution as long as np and nq are both greater than 5.
- To approximate the mean and standard deviation, the values $\mu = np$ and $\sigma = \sqrt{npq}$ are used.
- As with other discrete data, continuity correction should be applied when approximating a binomial distribution by a normal distribution.

Communicate Your Understanding

1. The probability of an airline flight arriving on time is 88%. Explain how to use the normal approximation to find the probability that at least 350 of a random sample of 400 flights arrived on time.

2. Construct an example of a binomial distribution for which np is less than 5, and for which the normal distribution is *not* a good approximation. Then, show that the condition $np > 5$ and $nq > 5$ is needed.

3. Has technology reduced the usefulness of the normal approximation to a binomial distribution? Justify your answer.

Practise

A

1. For which of the binomial distributions listed below is the normal distribution a reasonable approximation?

   a) $n = 60$, $p = 0.4$ **b)** $n = 45$, $p = 0.1$

   c) $n = 80$, $p = 0.1$ **d)** $n = 30$, $p = 0.8$

B

2. Copy the table. Use the normal approximation to complete the table.

Sample Size, n	Probability of Success, p	μ	σ	Probability
60	0.4			$P(X < 22) =$
200	0.7			$P(X < 160) =$
75	0.6			$P(X > 50) =$
250	0.2			$P(X > 48) =$
1000	0.8			$P(780 < X < 840) =$
90	0.65			$P(52 < X < 62) =$
100	0.36			$P(X = 40) =$
3000	0.52			$P(X = 1650) =$

Apply, Solve, Communicate

Use the normal approximation to the binomial distribution, unless otherwise indicated.

3. It is estimated that 62% of television viewers "channel surf" during commercials. A market-research firm surveyed 1500 television viewers. What is the probability that at least 950 of them were channel surfing?

4. **Application** Salespeople sometimes advertise their products by telephoning strangers. Only about 1.5% of these "cold calls" result in a sale. Toni makes cold calls 8 h per day for 5 days. The average time for a cold call is 90 s. What is the probability that Toni gets at least 30 new customers for the week?

5. A theatre found that 7% of people who purchase tickets for a play do not show up. If the theatre's capacity is 250 people, what is the probability that there are fewer than 20 "no shows" for a sold-out performance?

6. A magazine reported that 18% of car drivers use a cellular phone while driving. In a survey of 200 drivers, what is the probability that exactly 40 of them will use a cellular phone while driving? Compare the results of using the binomial distribution and the normal approximation.

7. The human-resources manager at a company knows that 34% of the workforce belong to a union. If she randomly surveys 50 employees, what is the probability that exactly 30 of them do not belong to a union? Compare the results of using the binomial distribution with the results of using the normal approximation.

8. **Application** A recent survey of a gas-station's customers showed that 68% paid with credit cards, 29% used debit cards, and only 3% paid with cash. During her eight-hour shift as cashier at this gas station, Serena had a total of 223 customers.

a) What is the probability that

i) at least 142 customers used a credit card?

ii) fewer than 220 customers paid with credit or debit cards?

b) What is the expected number of customers who paid Serena with cash?

9. A computer-chip manufacturer knows that 72% of the chips produced are defective. Suppose 3000 chips are produced every hour.

a) What is the probability that

i) at least 800 chips are acceptable?

ii) exactly 800 chips are acceptable?

b) Compare the results of using the binomial distribution with those found using the normal approximation.

10. Calculate the probability that 200 rolls of two dice rolls will include

a) more than 30 sums of 5

b) between 30 and 40, inclusive, sums of 5

11. On some busy streets, diamond lanes are reserved for taxis, buses, and cars with three or more passengers. It is estimated that 20% of cars travelling in a certain diamond lane have fewer than three passengers. Sixty cars are selected at random.

a) Use the normal approximation to find the probability that

i) fewer than 10 cars have fewer than three passengers

ii) at least 15 cars have fewer than three passengers

b) Compare these results with those found using the binomial distribution.

c) How would the results compare if 600 cars were selected?

ACHIEVEMENT CHECK

Knowledge / Understanding	Thinking / Inquiry / Problem Solving	Communication	Application

12. The probability of winning a large plush animal in the ring-toss game at the Statsville School Fair is 8%.

a) Find the probability of winning in at least 10% of 300 games, using

i) a binomial distribution

ii) a normal distribution

b) Predict how the probabilities of winning at least 50 times in 500 games will differ from the answers in part a). Explain your prediction.

c) Verify your prediction in part b) by calculating the probabilities using both distributions. Do your calculations support your predictions?

d) When designing the game, one student claims that having $np > 3$ and $nq > 3$ is a sufficient test for the normal approximation. Another student claims that np and nq both need to be over 10. Whom would you agree with and why?

C

13. **Inquiry / Problem Solving**

a) A newspaper knows that 64% of the households in a town are subscribers. If 50 households are surveyed randomly, how many of these households should the newspaper expect to be subscribers?

b) The marketing manager for the newspaper has asked you for an upper and lower limit for the number of subscribers likely to be in this sample. Find upper and lower bounds of a range that has a 90% probability of including this number.