SHERIDAN
INSTITUTE OF HIGHER EDUCATION

UNIVARIATE STATISTICAL ANALYSIS

TUTORIAL SESSION 4

DISCRETE PROBABILITY DISTRIBUTIONS
BINOMIAL AND GEOMETRIC

Chapter 7

Question 7.52 (page.380) – Auto Accidents

Question 7.55 (page.381) – Express Checkout

Question 7.62 (page.382) – Automobiles Inspection
Part a. and part b. only.

Question 7.67 (page.382) – Songs Playlist
Part a. and part b. only.

Question 7.68 (page.383) – Catch a Ball

Because the probability of success is p for each trial, the probability of failure for each trial is $1 - p$. Because the trials are independent,

$$\begin{aligned} p(x) &= P(x \text{ trials to first success}) = P(FF\ldots FS) \\ &= P(F)P(F)\cdots P(F)P(S) \\ &= (1-p)(1-p)\cdots(1-p)p \\ &= (1-p)^{x-1}p \end{aligned}$$

This leads us to the formula for the geometric probability distribution.

Geometric Probability Distribution

If x is a geometric random variable with probability of success $= p$ for each trial, then

$$p(x) = (1-p)^{x-1}p \quad x = 1, 2, 3, \ldots$$

istock.com/Whiteway

Example 7.22 Jumper Cables

Consider the jumper cable scenario described previously. For this problem, $p = 0.4$, because 40% of the students who drive to campus carry jumper cables. The probability distribution of

x = number of students who must be stopped before finding a student with jumper cables

is

$$p(x) = (0.6)^{x-1}(0.4) \quad x = 1, 2, 3, \ldots$$

The probability distribution can now be used to calculate various probabilities. For example, the probability that the first student stopped has jumper cables (that is, $x = 1$) is

$$p(1) = (0.6)^{1-1}(0.4) = (0.6)^0(0.4) = 0.4$$

The probability that three or fewer students must be stopped is

$$\begin{aligned} P(x \le 3) &= p(1) + p(2) + p(3) \\ &= (0.6)^0(0.4) + (0.6)^1(0.4) + (0.6)^2(0.4) \\ &= 0.4 + 0.24 + 0.144 \\ &= 0.784 \end{aligned}$$

EXERCISES 7.51 - 7.69

7.51 CBS News reported that 4% of adult Americans have a food allergy (June 1, 2017, cbsnews.com/news/food-allergies-in-america-new-report-shellfish-peanut-dairy, retrieved March 25, 2018). Consider selecting 10 adult Americans at random. Define the random variable x as

x = number of people in the sample of 10 that have a food allergy.

Find the following probabilities. (Hint: See Examples 7.19 and 7.20.)

a. $p(x < 3)$
b. $p(x \le 3)$
c. $p(x \ge 4)$
d. $p(1 \le x \le 3)$

7.52 The article "Should You Report That Fender-Bender?" (*Consumer Reports*, 2013:15) reported that 7 in 10 auto accidents involve a single vehicle. Suppose 15 accidents are randomly selected. (Hint: See Examples 7.19 and 7.20.)

a. What is the probability that exactly four involve a single vehicle?
b. What is the probability that at most four involve a single vehicle?
c. What is the probability that exactly six involve multiple vehicles?

7.53 FlightView surveyed 2600 North American airline passengers and reported that approximately 80% said that they carry a smart phone when they travel (flightview.com/TravelersSurvey/downloads /survey_infographic_poster.pdf). Suppose that the actual percentage is 80%. Consider randomly selecting six passengers and define the random variable x to be the number of the six selected passengers who travel with a smart phone. The probability distribution of x is the binomial distribution with $n = 6$ and $p = 0.8$.

a. Calculate $p(4)$, and interpret this probability.
b. Calculate $p(6)$, the probability that all six selected passengers travel with a smart phone.
c. Calculate $P(x \geq 4)$.

7.54 Refer to the previous exercise, and suppose that 10 rather than six passengers are selected ($n = 10$, $p = 0.8$). (Hint: Use technology or Appendix Table 9.)

a. Calculate $p(8)$.
b. Calculate $P(x \leq 7)$.
c. Calculate the probability that more than half of the selected passengers travel with a smart phone.

7.55 Twenty-five percent of the customers of a grocery store use an express checkout. Consider five randomly selected customers, and let x denote the number among the five who use the express checkout.

a. Calculate $p(2)$.
b. Calculate $P(x \leq 1)$.
c. Calculate $P(2 \leq x)$. (Hint: Make use of your answer from Part (b).)
d. Calculate $P(x \neq 2)$.

7.56 Example 7.18 described a study in which a person was asked to determine which of three t-shirts had been worn by her roommate by smelling the shirts ("Sociochemosensory and Emotional Functions," *Psychological Science* [2009]: 1118–1123). Suppose that instead of three shirts, each participant was asked to choose among four shirts and that the process was repeated five times. Then, assuming that the participant is choosing at random, $x =$ number of correct identifications is a binomial random variable with $n = 5$ and $p = 1/4$.

a. What are the possible values of x?
b. For each possible value of x, find the associated probability $p(x)$ and display the possible x values and $p(x)$ values in a table.
c. Construct a probability histogram for the probability distribution of x.

7.57 Information Security Buzz provides news for the information security community. In an article published on September 24, 2016, it reported that based on a large international survey of Internet users, 60% of Internet users have installed security solutions on all of the devices they use to access the Internet (informationsecuritybuzz.com /articles/21-29-60-kaspersky-lab-presents-first -cybersecurity-index/, retrieved May 2, 2017).

a. Suppose that the true proportion of Internet users who have security solutions on all the devices they use to access the Internet is 0.60. If 20 Internet users are selected at random, what is the probability that more than 10 have security solutions installed on all devices used to access the Internet?
b. Suppose that a random sample of 20 Internet users is selected. Which is more likely—that more than 15 have security solutions on all devices used to access the Internet or that fewer than 5 have security solutions on all devices used to access the Internet? Justify your answer based on probability calculations.

7.58 A breeder of show dogs is interested in the number of female puppies in a litter. If a birth is equally likely to result in a male or a female puppy, give the probability distribution of the variable $x =$ number of female puppies in a litter of size 5.

7.59 *Women's Health Magazine* surveyed 1187 readers to find out how often people wash their sheets (womenshealthmag.com/health/dirty-sheets, March 26, 2015, retrieved May 2, 2017). They found that even though microbiologists recommend that you wash your sheets at least once a week, only 44% said that they wash their sheets that often. Suppose this group is representative of adult Americans and define the random variable x to be the number of adult Americans you would have to ask before you found someone that washes his or her sheets at least once a week.

a. Is the probability distribution of x binomial or geometric? Explain.
b. What is the probability that you would have to ask three people before finding one who washes sheets at least once a week?
c. What is the probability that fewer than four people would have to be asked before finding one who washes sheets at least once a week?
d. What is the probability that more than three people would have to be asked before finding one who washes sheets at least once a week?

7.60 Industrial quality control programs often include inspection of incoming parts from suppliers. If parts are purchased in large lots, a typical plan might be to select 20 parts at random from a lot and inspect them. Suppose that a lot is considered to be acceptable if one or fewer defective parts are found among those inspected. Otherwise, the lot is rejected and returned to the supplier. Use technology or Appendix Table 9 to find the probability of accepting lots that have each of the following (Hint: Identify success with a defective part):

a. 5% defective parts
b. 10% defective parts
c. 20% defective parts

7.61 Suppose that the probability is 0.1 that any given citrus tree will show measurable damage when the temperature falls to 30°F. (Hint: See Example 7.21.)

a. If the temperature does drop to 30°F, what is the expected number of citrus trees showing damage in orchards of 2000 trees?
b. What is the standard deviation of the number of trees that show damage?

7.62 Suppose that 30% of all automobiles undergoing an emissions inspection at an inspection station fail the inspection.

a. Among 15 randomly selected cars, what is the probability that at most 5 fail the inspection?
b. Among 15 randomly selected cars, what is the probability that between 5 and 10 (inclusive) fail to pass inspection?
c. Among 25 randomly selected cars, what is the mean value of the number that pass inspection, and what is the standard deviation of the number that pass inspection?
d. What is the probability that among 25 randomly selected cars, the number that pass is within 1 standard deviation of the mean value? (Hint: See Example 7.21.)

7.63 Suppose that you will take a multiple-choice exam consisting of 100 questions with five possible responses to each question. You have not studied and so must guess (select one of the five answers in a completely random fashion) on each question. Let x represent the number of correct responses on the test.

a. What kind of probability distribution does x have?
b. What is your expected score on the exam? (Hint: Your expected score is the mean value of the x distribution.)
c. Calculate the variance and standard deviation of x.
d. Based on your answers to Parts (b) and (c), is it likely that you would score over 50 on this exam? Explain the reasoning behind your answer.

7.64 Suppose that 20% of the 10,000 signatures on a recall petition are invalid. Would the number of invalid signatures in a sample of 2000 of these signatures have (approximately) a binomial distribution? Explain.

7.65 A city requires that smoke detectors be installed in all houses. There is concern that too many houses are still without detectors, so a costly inspection program is being considered. Let p be the proportion of all houses that have a detector. A random sample of 25 houses is selected. If the sample strongly suggests that $p < 0.80$ (less than 80% have detectors), as opposed to $p \geq 0.80$, the program will be implemented. Let x be the number of residences among the 25 that have a detector, and consider the following decision rule: Reject the claim that $p \geq 0.8$ and implement the program if $x \leq 15$.

a. What is the probability that the program is implemented when $p = 0.80$?
b. What is the probability that the program is not implemented if $p = 0.70$?
c. What is the probability that the program is not implemented if $p = 0.60$?
d. How do the "error probabilities" of Parts (b) and (c) change if the value 15 in the decision rule is changed to 14?

7.66 Suppose that 90% of all registered California voters favor banning the release of information from exit polls in presidential elections until after the polls in California close. A random sample of 25 registered California voters is to be selected.

a. What is the probability that more than 20 favor the ban?
b. What is the probability that at least 20 favor the ban?
c. What are the mean value and standard deviation of the number of voters in the sample who favor the ban?
d. If fewer than 20 voters in the sample favor the ban, is this inconsistent with the claim that (at least) 90% of California registered voters favors the ban? (Hint: Consider $P(x < 20)$ when $p = 0.9$.)

7.67 Suppose a playlist on a music player consists of 100 songs, of which eight are by a particular artist. Songs are played by selecting a song at random (with replacement) from the playlist. Let the random variable x represent the number of songs played until a song by this artist is played.

a. Explain why the probability distribution of x is not binomial.
b. Find the following probabilities. (Hint: See Example 7.22.)
  i. $p(4)$
  ii. $P(x \leq 4)$
  iii. $P(x > 4)$
  iv. $P(x \geq 4)$
c. Interpret each of the probabilities in Part (b).

7.68 Sophie is a dog that loves to play catch. Unfortunately, she isn't very good, and the probability that she catches a ball is only 0.1. Let x be the number of tosses required until Sophie catches a ball.

a. Does x have a binomial or a geometric distribution?

b. What is the probability that it will take exactly two tosses for Sophie to catch a ball?

c. What is the probability that more than three tosses will be required?

7.69 Suppose that 5% of cereal boxes contain a prize and the other 95% contain the message, "Sorry, try again." Consider the random variable x, where x = number of boxes purchased until a prize is found.

a. What is the probability that at most two boxes must be purchased?

b. What is the probability that exactly four boxes must be purchased?

c. What is the probability that more than four boxes must be purchased?

SECTION 7.6 Normal Distributions

Normal distributions formalize the notion of mound-shaped histograms introduced in Chapter 4. Normal distributions are widely used for two reasons. First, they provide a reasonable approximation to the distribution of many different variables. They also play an important role in many of the inferential procedures that will be introduced in later chapters of this textbook.

Normal distributions are continuous probability distributions that are bell-shaped and symmetric, as shown in Figure 7.16. Normal distributions are sometimes referred to as *normal curves*.

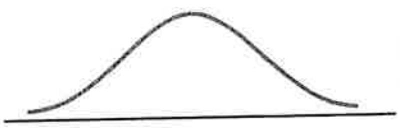
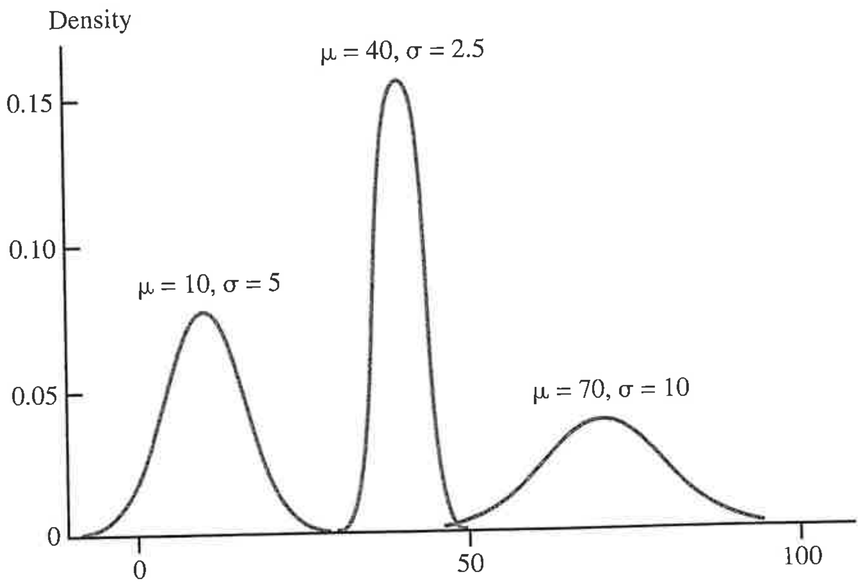


FIGURE 7.16
A normal distribution.

There are many different normal distributions, and they are distinguished from one another by their mean μ and standard deviation σ. The mean μ of a normal distribution describes where the corresponding curve is centered. The standard deviation σ describes how much the curve spreads out around that center. As with all continuous probability distributions, the total area under any normal curve is equal to 1.

Three normal distributions are shown in Figure 7.17. Notice that the smaller the standard deviation, the taller and narrower the corresponding curve. Remember that areas under a continuous probability distribution curve represent probabilities, so when the standard deviation is small, a larger area is concentrated near the center of the curve. This means that the chance of observing a value near the mean is much greater (because μ is at the center).

FIGURE 7.17
Three normal distributions.



The value of μ is the number on the measurement axis lying directly below the top of the curve. The value of σ can be approximated from a picture of the curve. Consider the normal curve in Figure 7.18. Starting at the top (above μ = 100) and moving to