6.1 TECHNOLOGY CORNER

TI-Nspire instructions in Appendix B; HP Prime instructions on the book's Web site.

Section 6.1 Exercises

1. **Toss 4 times** Suppose you toss a fair coin 4 times. Let X = the number of heads you get.

(a) Find the probability distribution of X.

(b) Make a histogram of the probability distribution. Describe what you see.

(c) Find $P(X \leq 3)$ and interpret the result.

2. **Pair-a-dice** Suppose you roll a pair of fair, six-sided dice. Let T = the sum of the spots showing on the up-faces.

(a) Find the probability distribution of T.

(b) Make a histogram of the probability distribution. Describe what you see.

(c) Find $P(T \geq 5)$ and interpret the result.

3. **Spell-checking** Spell-checking software catches "nonword errors," which result in a string of letters that is not a word, as when "the" is typed as "teh." When undergraduates are asked to write a 250-word essay (without spell-checking), the number X of nonword errors has the following distribution:

Value:	0	1	2	3	4
Probability:	0.1	0.2	0.3	0.3	0.1

(a) Write the event "at least one nonword error" in terms of X. What is the probability of this event?

(b) Describe the event $X \leq 2$ in words. What is its probability? What is the probability that $X < 2$?

4. **Kids and toys** In an experiment on the behavior of young children, each subject is placed in an area with five toys. Past experiments have shown that the probability distribution of the number X of toys played with by a randomly selected subject is as follows:

Number of toys x_i:	0	1	2	3	4	5
Probability p_i:	0.03	0.16	0.30	0.23	0.17	0.11

(a) Write the event "plays with at most two toys" in terms of X. What is the probability of this event?

(b) Describe the event $X > 3$ in words. What is its probability? What is the probability that $X \geq 3$?

5. **Benford's law** Faked numbers in tax returns, invoices, or expense account claims often display patterns that aren't present in legitimate records. Some patterns, like too many round numbers, are obvious and easily avoided by a clever crook. Others are more subtle. It is a striking fact that the first digits of numbers in legitimate records often follow a model known as Benford's law.[7] Call the first digit of a randomly chosen record X for short. Benford's law gives this probability model for X (note that a first digit can't be 0):

pg 349

First digit:	1	2	3	4	5	6	7	8	9
Probability:	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

(a) Show that this is a legitimate probability distribution.

(b) Make a histogram of the probability distribution. Describe what you see.

(c) Describe the event $X \geq 6$ in words. What is $P(X \geq 6)$?

(d) Express the event "first digit is at most 5" in terms of X. What is the probability of this event?

6. **Working out** Choose a person aged 19 to 25 years at random and ask, "In the past seven days, how many times did you go to an exercise or fitness center or work out?" Call the response Y for short. Based on a large sample survey, here is a probability model for the answer you will get:[8]

Days:	0	1	2	3	4	5	6	7
Probability:	0.68	0.05	0.07	0.08	0.05	0.04	0.01	0.02

(a) Show that this is a legitimate probability distribution.

(b) Make a histogram of the probability distribution. Describe what you see.

(c) Describe the event $Y < 7$ in words. What is $P(Y < 7)$?

(d) Express the event "worked out at least once" in terms of Y. What is the probability of this event?

7. **Benford's law** Refer to Exercise 5. The first digit of a randomly chosen expense account claim follows Benford's law. Consider the events A = first digit is 7 or greater and B = first digit is odd.

(a) What outcomes make up the event A? What is $P(A)$?

(b) What outcomes make up the event B? What is $P(B)$?

(c) What outcomes make up the event "A or B"? What is $P(A \text{ or } B)$? Why is this probability not equal to $P(A) + P(B)$?

8. **Working out** Refer to Exercise 6. Consider the events A = works out at least once and B = works out less than 5 times per week.

(a) What outcomes make up the event A? What is $P(A)$?

(b) What outcomes make up the event B? What is $P(B)$?

(c) What outcomes make up the event "A and B"? What is $P(A \text{ and } B)$? Why is this probability not equal to $P(A) \cdot P(B)$?

9. pg 352 **Keno** Keno is a favorite game in casinos, and similar games are popular with the states that operate lotteries. Balls numbered 1 to 80 are tumbled in a machine as the bets are placed, then 20 of the balls are chosen at random. Players select numbers by marking a card. The simplest of the many wagers available is "Mark 1 Number." Your payoff is \$3 on a \$1 bet if the number you select is one of those chosen. Because 20 of 80 numbers are chosen, your probability of winning is 20/80, or 0.25. Let X = the net amount you gain on a single play of the game.

(a) Make a table that shows the probability distribution of X.

(b) Compute the expected value of X. Explain what this result means for the player.

10. **Fire insurance** Suppose a homeowner spends \$300 for a home insurance policy that will pay out \$200,000 if the home is destroyed by fire. Let Y = the profit made by the company on a single policy. From previous data, the probability that a home in this area will be destroyed by fire is 0.0002.

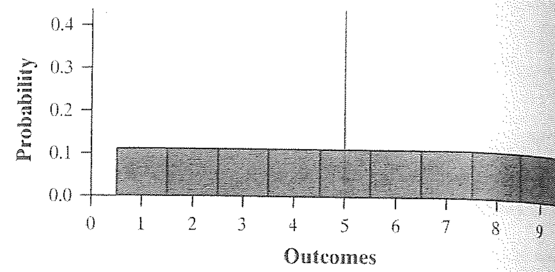
(a) Make a table that shows the probability distribution of Y.

(b) Compute the expected value of Y. Explain what this result means for the insurance company.

11. **Spell-checking** Refer to Exercise 3. Calculate the mean of the random variable X and interpret this result in context.

12. **Kids and toys** Refer to Exercise 4. Calculate the mean of the random variable X and interpret this result in context.

13. **Benford's law and fraud** A not-so-clever employee decided to fake his monthly expense report. He believed that the first digits of his expense amounts should be equally likely to be any of the numbers from 1 to 9. In that case, the first digit Y of a randomly selected expense amount would have the probability distribution shown in the histogram.



(a) Explain why the mean of the random variable Y is located at the solid red line in the figure.

(b) The first digits of randomly selected expense amounts actually follow Benford's law (Exercise 5). According to Benford's law, what's the expected value of the first digit? Explain how this information could be used to detect a fake expense report.

(c) What's $P(Y > 6)$ in the above distribution? According to Benford's law, what proportion of first digits in the employee's expense amounts should be greater than 6? How could this information be used to detect a fake expense report?

14. **Life insurance** A life insurance company sells a term insurance policy to a 21-year-old male that pays \$100,000 if the insured dies within the next 5 years. The probability that a randomly chosen male will die each year can be found in mortality tables. The company collects a premium of \$250 each year as payment for the insurance. The amount Y that the company earns on this policy is \$250 per year, less the \$100,000 that it must pay if the insured dies. Here is a partially completed table that shows information about risk of mortality and the values of Y = profit earned by the company:

Age at death:	21	22	23	24	25	26 or more
Profit:	−\$99,750	−\$99,500	−\$99,250	−\$99,000	−\$98,750	\$1250
Probability:	0.00183	0.00186	0.00189	0.00191	0.00193	

(a) Explain why the company suffers a loss of \$98,750 on such a policy if a client dies at age 25.

(b) Find the missing probability. Show your work.

(c) Calculate the mean μ_Y. Interpret this value in context.

15. **Spell-checking** Refer to Exercise 3. Calculate and interpret the standard deviation of the random variable X. Show your work.
pg 353

16. **Kids and toys** Refer to Exercise 4. Calculate and interpret the standard deviation of the random variable X. Show your work.

17. **Benford's law and fraud** Refer to Exercise 13. It might also be possible to detect an employee's fake expense records by looking at the variability in the first digits of those expense amounts.

(a) Calculate the standard deviation σ_Y. This gives us an idea of how much variation we'd expect in the employee's expense records if he assumed that first digits from 1 to 9 were equally likely.

(b) Now calculate the standard deviation of first digits that follow Benford's law (Exercise 5). Would using standard deviations be a good way to detect fraud? Explain.

18. **Life insurance**

(a) It would be quite risky for you to insure the life of a 21-year-old friend under the terms of Exercise 14. There is a high probability that your friend would live and you would gain \$1250 in premiums. But if he were to die, you would lose almost \$100,000. Explain carefully why selling insurance is not risky for an insurance company that insures many thousands of 21-year-old men.

(b) The risk of an investment is often measured by the standard deviation of the return on the investment. The more variable the return is, the riskier the investment. We can measure the great risk of insuring a single person's life in Exercise 14 by computing the standard deviation of the income Y that the insurer will receive. Find σ_Y using the distribution and mean found in Exercise 14.

19. **Housing in San Jose** How do rented housing units differ from units occupied by their owners? Here are the distributions of the number of rooms for owner-occupied units and renter-occupied units in San Jose, California:[9]

	Number of Rooms									
	1	2	3	4	5	6	7	8	9	10
Owned	0.003	0.002	0.023	0.104	0.210	0.224	0.197	0.149	0.053	0.035
Rented	0.008	0.027	0.287	0.363	0.164	0.093	0.039	0.013	0.003	0.003

Let X = the number of rooms in a randomly selected owner-occupied unit and Y = the number of rooms in a randomly chosen renter-occupied unit.

(a) Make histograms suitable for comparing the probability distributions of X and Y. Describe any differences that you observe.

(b) Find the mean number of rooms for both types of housing unit. Explain why this difference makes sense.

(c) Find and interpret the standard deviations of both X and Y.

20. **Size of American households** In government data, a household consists of all occupants of a dwelling unit, while a family consists of two or more persons who live together and are related by blood or marriage. So all families form households, but some households are not families. Here are the distributions of household size and family size in the United States:

	Number of Persons						
	1	2	3	4	5	6	7
Household probability	0.25	0.32	0.17	0.15	0.07	0.03	0.01
Family probability	0	0.42	0.23	0.21	0.09	0.03	0.02

Let X = the number of people in a randomly selected U.S. household and Y = the number of people in a randomly chosen U.S. family.

(a) Make histograms suitable for comparing the probability distributions of X and Y. Describe any differences that you observe.

(b) Find the mean for each random variable. Explain why this difference makes sense.

(c) Find and interpret the standard deviations of both X and Y.

21. **Random numbers** Let X be a number between 0 and 1 produced by a random number generator. Assuming that the random variable X has a uniform distribution, find the following probabilities:

(a) $P(X > 0.49)$

(b) $P(X \geq 0.49)$

(c) $P(0.19 \leq X < 0.37 \text{ or } 0.84 < X \leq 1.27)$

22. **Random numbers** Let Y be a number between 0 and 1 produced by a random number generator. Assuming that the random variable Y has a uniform distribution, find the following probabilities:

(a) $P(Y \leq 0.4)$

(b) $P(Y < 0.4)$

(c) $P(0.1 < Y \leq 0.15 \text{ or } 0.77 \leq Y < 0.88)$

23. **Running a mile** A study of 12,000 able-bodied male students at the University of Illinois found that their times for the mile run were approximately Normal with mean 7.11 minutes and standard deviation 0.74 minute.[10] Choose a student at random from this group and call his time for the mile Y. Find $P(Y < 6)$ and interpret the result.
pg 357

24. **ITBS scores** The Normal distribution with mean $\mu = 6.8$ and standard deviation $\sigma = 1.6$ is a good description of the Iowa Test of Basic Skills (ITBS) vocabulary scores of seventh-grade students in Gary, Indiana. Call the score of a randomly chosen student X for short. Find $P(X \geq 9)$ and interpret the result.

25. **Ace!** Professional tennis player Rafael Nadal hits the ball extremely hard. His first-serve speeds follow a Normal distribution with mean 115 miles per hour (mph) and standard deviation 6 mph. Choose one of Nadal's first serves at random. Let Y = its speed, measured in miles per hour.

(a) Find $P(Y > 120)$ and interpret the result.

(b) What is $P(Y \geq 120)$? Explain.

(c) Find the value of c such that $P(Y \leq c) = 0.15$. Show your work.

26. **Pregnancy length** The length of human pregnancies from conception to birth follows a Normal distribution with mean 266 days and standard deviation 16 days. Choose a pregnant woman at random. Let X = the length of her pregnancy.

(a) Find $P(X \geq 240)$ and interpret the result.

(b) What is $P(X > 240)$? Explain.

(c) Find the value of c such that $P(X \geq c) = 0.20$. Show your work.

Multiple choice: Select the best answer for Exercises 27 to 30.

Exercises 27 to 29 refer to the following setting. Choose an American household at random and let the random variable X be the number of cars (including SUVs and light trucks) they own. Here is the probability model if we ignore the few households that own more than 5 cars:

Number of cars X:	0	1	2	3	4	5
Probability:	0.09	0.36	0.35	0.13	0.05	0.02

27. What's the expected number of cars in a randomly selected American household?

(a) 1.00 (b) 1.75 (c) 1.84 (d) 2.00 (e) 2.50

28. The standard deviation of X is $\sigma_X = 1.08$. If many households were selected at random, which of the following would be the best interpretation of the value 1.08?

(a) The mean number of cars would be about 1.08.

(b) The number of cars would typically be about 1.08 from the mean.

(c) The number of cars would be at most 1.08 from the mean.

(d) The number of cars would be within 1.08 from the mean about 68% of the time.

(e) The mean number of cars would be about 1.08 from the expected value.

29. About what percentage of households have a number of cars within 2 standard deviations of the mean?

(a) 68% (b) 71% (c) 93% (d) 95% (e) 98%

30. A deck of cards contains 52 cards, of which 4 are aces. You are offered the following wager: Draw one card at random from the deck. You win \$10 if the card drawn is an ace. Otherwise, you lose \$1. If you make this wager very many times, what will be the mean amount you win?

(a) About −\$1, because you will lose most of the time.

(b) About \$9, because you win \$10 but lose only \$1.

(c) About −\$0.15; that is, on average you lose about 15 cents.

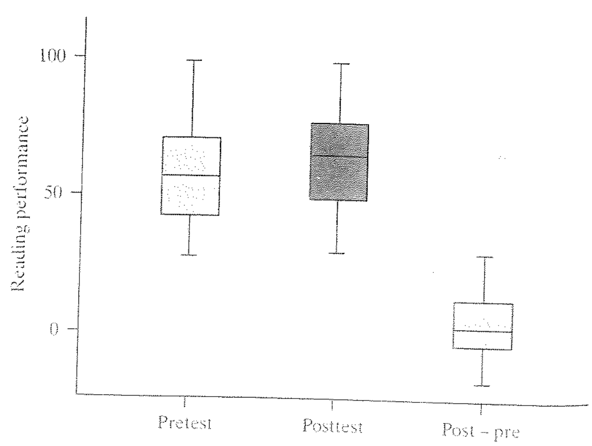
(d) About \$0.77; that is, on average you win about 77 cents.

(e) About \$0, because the random draw gives you a fair bet.

Exercises 31 to 34 refer to the following setting. Many chess masters and chess advocates believe that chess play develops general intelligence, analytical skill, and the ability to concentrate. According to such beliefs, improved reading skills should result from study to improve chess-playing skills. To investigate this belief, researchers conducted a study. All of the subjects in the study participated in a comprehensive chess program, and their reading performances were measured before and after the program. The graphs and numerical summaries below provide information on the subjects' pretest scores, posttest scores, and the difference (post – pre) between these two scores.

Descriptive Statistics: Pretest, Posttest, Post – pre

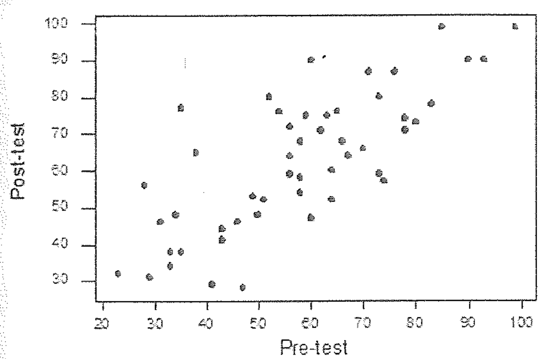
Variable	N	Mean	Median	StDev	Min	Max	Q_1	Q_3
Pretest	53	57.70	58.00	17.84	23.00	99.00	44.50	70.50
Posttest	53	63.08	64.00	18.70	28.00	99.00	48.00	76.00
Post-pre	53	5.38	3.00	13.02	-19.00	42.00	-3.50	14.00

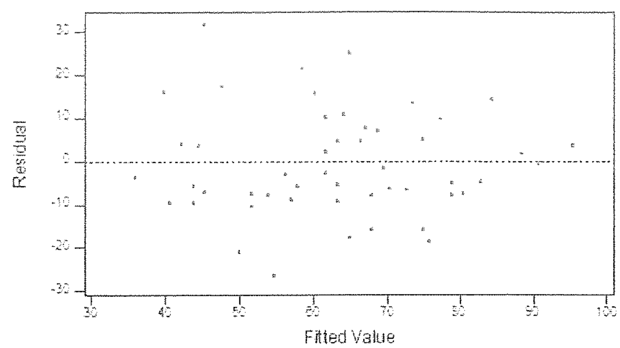


31. **Better readers?** (1.3) Did students have higher reading scores after participating in the chess program? Give appropriate statistical evidence to support your answer.

32. **Chess and reading** (4.3) If the study found a statistically significant improvement in reading scores, could you conclude that playing chess causes an increase in reading skills? Justify your answer.

Some graphical and numerical information about the relationship between pretest and posttest scores is provided below.





Regression Analysis: Posttest versus Pretest

Predictor	Coef	SE Coef	T	P
Constant	17.897	5.889	3.04	0.004
Pretest	0.78301	0.09758	8.02	0.000
S = 12.55	R-Sq = 55.8%		R-Sq(adj) = 54.9%	

33. **Predicting posttest scores** (3.2) What is the equation of the linear regression model relating posttest and pretest scores? Define any variables used.

34. **How well does it fit?** (3.2) Discuss what s, r^2, and the residual plot tell you about this linear regression model.

6.2 Transforming and Combining Random Variables

WHAT YOU WILL LEARN By the end of the section, you should be able to:

- Describe the effects of transforming a random variable by adding or subtracting a constant and multiplying or dividing by a constant.
- Find the mean and standard deviation of the sum or difference of independent random variables.
- Find probabilities involving the sum or difference of independent Normal random variables.

In Section 6.1, we looked at several examples of random variables and their probability distributions. We also saw that the mean μ_X and standard deviation σ_X give us important information about a random variable. For instance, for X = the amount gained on a single \$1 bet on red in a game of roulette, we already showed that $\mu_X = -\$0.05$. You can verify that the standard deviation is $\sigma_X = \$1.00$. That is, a player can expect to lose an average of 5 cents per \$1 bet if he plays many games. But if he plays only a few games, his actual gain could be much better or worse than this expected value.