AP Statistics
Solutions to Packet 7

<u>Random Variables</u>
Discrete and Continuous Random Variables
Means and Variances of Random Variables

HW #44 2, 3, 6 – 8, 13 – 17

7.2 **THREE CHILDREN** A couple plans to have three children. There are 8 possible arrangements of girls and boys. For example, GGB means the first two children are girls and the third child is a boy. All 8 arrangements are (approximately) equally likely.

(a) Write down all 8 arrangements of the sexes of three children. What is the probability of any one of these arrangements? BBB, BBG, BGB, GBB, GGB, GBG, BGG, GGG. Each has probability 1/8.

(b) Let X be the number of girls the couple has. What is the probability that $X = 2$?
Three of the eight arrangements have two (and only two) girls, so $P(X = 2) = 3/8 = 0.375$.

(c) Starting from your work in (a), find the distribution of X. That is, what values can X take, and what are the probabilities for each value?

Value of X	0	1	2	3
Probability	1/8	3/8	3/8	1/8

7.3 **SOCIAL CLASS IN ENGLAND** A study of social mobility in England looked at the social class reached by the sons of lower-class fathers. Social classes are numbered from 1 (low) to 5 (high). Take the random variable X to be the class of a randomly chosen son of a father in Class I. The study found that the distribution of X is:

Son's class:	1	2	3	4	5
Probability:	0.48	0.38	0.08	0.05	0.01

(a) What percent of the sons of lower-class fathers reach the highest class, Class 5? 1%

(b) Check that this distribution satisfies the requirements for a discrete probability distribution.
All probabilities are between 0 and 1; the probabilities add to 1.

(c) What is $P(X \leq 3)$? $P(X \leq 3) = 0.48 + 0.38 + 0.08 = 1 - 0.01 - 0.05 = 0.94$.

(d) What is $P(X < 3)$? $P(X < 3) = 0.48 + 0.38 = 0.86$.

(e) Write the event "a son of a lower-class father reaches one of the two highest classes" in terms of X. What is the probability of this event?
Write either $X \geq 4$ or $X > 3$. The probability is $0.05 + 0.01 = 0.06$.

(f) Briefly describe how you would use simulation to answer the question in (c).
Read two random digits from Table B. Here is the correspondence: 01 to 48 ⇔Class 1, 49 to 86 ⇔Class 2, 87 to 94 ⇔Class 3, 95 to 99 ⇔ Class 4, and 00 ⇔Class 5.
Repeatedly generate 2 digit random numbers. The proportion of numbers in the range 01 to 94 will be an estimate of the required probability.

7.6 **CONTINUOUS RANDOM VARIABLE, I** Let X be a random number between 0 and 1 produced by the idealized uniform random generator described in Example 7.3 and Figure 7.5 (text p. 398). Find the following probabilities:

(a) $P(0 \leq X \leq 0.4) = 0.4$

(b) $P(0.4 \leq X \leq 1) = 0.6$

(c) $P(0.3 \leq X \leq 0.5) = 0.2$

(d) $P(0.3 < X < 0.5) = 0.2$

(e) $P(0.226 \leq X \leq 0.713) = 0.713 - 0.226 = 0.487$

(f) What important fact about continuous random variables does comparing your answer to (c) and (d) illustrate? A continuous distribution assigns probability 0 to every individual outcome. In this case, the probabilities in (c) and (d) are the same because the events differ by 2 individual values, 0.3 and 0.5, each of which has probability 0.

7.7 **CONTINUOUS RANDOM VARIABLE, II** Let the random variable X be a random number with the uniform density curve as in the previous exercise. Find the following probabilities:

(a) $P(X \leq 0.49) = 0.49$

(b) $P(X \geq 0.27) = 0.73$

(c) $P(0.27 < X < 1.27) = P(0.27 < X < 1) = 0.73$

(d) $P(0.1 \leq X \leq 0.2 \text{ or } 0.8 \leq X \leq 0.9) = 0.1 + 0.1 = 0.2$

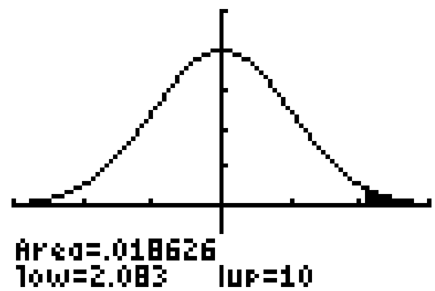
(e) The probability that X is not in the interval 0.3 to 0.8. $P(\text{not}[0.3 \leq X \leq 0.8]) = 1 - 0.5 = 0.5$

(f) $P(X = 0.5) = 0$

7.8 **VIOLENCE IN SCHOOLS, I** An SRS of 400 American adults is asked. "What do you think is the most serious problem facing our schools?" Suppose that in fact 40% of all adults would answer "violence" if asked this question. The proportion $\hat{p}$ of the sample who answer "violence" will vary in repeated sampling. In fact, we can assign probabilities to values of $\hat{p}$ using the normal density curve with mean 0.4 and standard deviation 0.024. Use the normal density curve to find the probabilities of the following events:

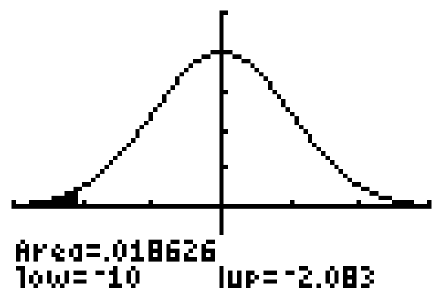
(a) At least 45% of the sample believes that violence is the schools' most serious problem.

$$P(\hat{p} \geq 0.45) = P\left(Z \geq \frac{0.45 - 0.4}{0.024}\right) = P(Z \geq 2.083) = 0.0186$$



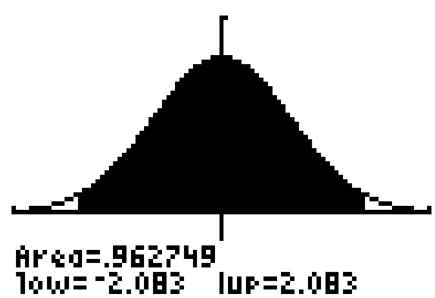
(b) Less than 35% of the sample believes that violence is the most serious problem.

$$P(\hat{p} < 0.35) = P(Z < -2.083) = 0.0186$$



(c) The sample proportion is between 0.35 and 0.45.

$$P(0.35 \leq \hat{p} \leq 0.45) = P(-2.083 \leq Z \leq 2.083) = 0.963$$



7.13 **ROLLING TWO DICE** Some games of chance rely on tossing two dice. Each die has six faces, marked with 1, 2, . . . 6 spots called pips. The dice used in casinos are carefully balanced so that each face is equally likely to come up. When two dice are tossed, each of the 36 possible pairs of faces is equally likely to come up. The outcome of interest to a gambler is the sum of the pips on the two up-faces. Call this random variable *X*.

(a) Write down all 36 possible pairs of faces.

The 36 possible pairs of "up faces" are

(1, 1) (1, 2) (1, 3) (1, 4) (1, 5) (1, 6)
(2, 1) (2, 2) (2, 3) (2, 4) (2, 5) (2, 6)
(3, 1) (3, 2) (3, 3) (3, 4) (3, 5) (3, 6)
(4, 1) (4, 2) (4, 3) (4, 4) (4, 5) (4, 6)
(5, 1) (5, 2) (5, 3) (5, 4) (5, 5) (5, 6)
(6, 1) (6, 2) (6, 3) (6, 4) (6, 5) (6, 6)

(b) If all pairs have the same probability, what must be the probability of each pair?

Each pair must have probability 1/36.

(c) Define the random variable *X*. Then write the value of *X* next to each pair of faces and use this information with the result of (b) to give the probability distribution of *X*. Draw a probability histogram to display the distribution.

Let *X* = sum of up faces. Then

Sum	Outcomes	Probability
$x = 2$	(1, 1)	$p = 1/36$
$x = 3$	(1, 2) (2, 1)	$p = 2/36$
$x = 4$	(1, 3) (2, 2) (3, 1)	$p = 3/36$
$x = 5$	(1, 4) (2, 3) (3, 2) (4, 1)	$p = 4/36$
$x = 6$	(1, 5) (2, 4) (3, 3) (4, 2) (5, 1)	$p = 5/36$
$x = 7$	(1, 6) (2, 5) (3, 4) (4, 3) (5, 2) (6, 1)	$p = 6/36$
$x = 8$	(2, 6) (3, 5) (4, 4) (5, 3) (6, 2)	$p = 5/36$
$x = 9$	(3, 6) (4, 5) (5, 4) (6, 3)	$p = 4/36$
$x = 10$	(4, 6) (5, 5) (6, 4)	$p = 3/36$
$x = 11$	(5, 6) (6, 5)	$p = 2/36$
$x = 12$	(6, 6)	$p = 1/36$

(d) One bet available in craps wins if a 7 or 11 comes up on the next roll of two dice. What is the probability of rolling a 7 or 11 on the next roll? P(7 or 11) = 6/36 + 2/36 = 8/36 or 2/9.

(e) After the dice are rolled the first time, several bets lose if a 7 is then rolled. If any outcome other than a 7 occurs, these bets either win or continue to the next roll. What is the probability that anything other than a 7 is rolled?

P (any sum other than 7) = 1 − P (7) = 1 − 6/36 = 30/36 = 5/6 by the complement rule.

7.14 **WEIRD DICE** Nonstandard dice produce interesting distributions of outcomes. You have two balanced, six-sided dice. One is a standard die, with faces having a 1, 2, 3, 4, 5, and 6 spots. The other die has three faces with 0 spots and three faces with 6 spots. Find the probability distribution for the total number of spots Y on the up-faces when you roll these two dice.

Here is a table of the possible observations of Y that can occur when we roll one standard die and one “weird” die. There are 36 possible pairs of faces; however, a number of the pairs are identical to each other.

	1	2	3	4	5	6
0	1	2	3	4	5	6
0	1	2	3	4	5	6
0	1	2	3	4	5	6
6	7	8	9	10	11	12
6	7	8	9	10	11	12
6	7	8	9	10	11	12

The possible values of Y are 1, 2, 3, . . . 12. Each value of Y has probability 3/36 = 1/12.

7.15 **EDUCATION LEVELS** A study of education followed a large group of fifth-grade children to see how many years of school they eventually completed. Let X be the highest year of school that a randomly chosen fifth grader completes. (Students who go on to college are included in the outcome $X = 12$.) The study found this probability distribution for X.

Years:	4	5	6	7	8	9	10	11	12
Probability:	0.010	0.007	0.007	0.013	0.032	0.068	0.070	0.041	0.752

(a) What percent of fifth graders eventually finished twelfth grade? 75.2%

(b) Check that this is a legitimate discrete probability distribution.

All probabilities are between 0 and 1; the probabilities add to 1.

(c) Find $P(X \geq 6)$ = 1 − 0.010 − 0.007 = 0.983.

(d) Find $P(X > 6)$ = 1 − 0.010 − 0.007 − 0.007 = 0.976.

(e) What values of X made up the event “the student completed at least one year of high school”? (High school begins with the ninth grade.) What is the probability of this event?

Either $X \geq 9$ or $X > 8$. The probability is 0.068 + 0.070 + 0.041 + 0.752 = 0.931.

7.16 **HOW STUDENT FEES ARE USED** Weary of the low turnout in student elections, a college administration decides to choose an SRS of three students to form an advisory board that represents student opinion. Suppose that 40% of all students oppose the use of student fees to fund student interest groups and that the opinions of the three students on the board are independent. Then the probability is 0.4 that each opposes the funding of interest groups.

(a) Call the three students A, B, and C. What is the probability that A and B support funding and C opposes it? (0.6)(0.6)(0.4) = 0.144.

(b) List all possible combinations of opinions that can be held by students A, B, and C. (*Hint:* There are eight possibilities) Then give the probability of each of these outcomes. Note that they are not equally likely.

The possible combinations are SSS, SSO, SOS, OSS, SOO, OSO, OOS, OOO
(S = support, O = oppose).
P(SSS) = 0.6^3 = 0.216,
P(SSO) = P(SOS) = P(OSS) = $(0.6^2)(0.4)$ = 0.144,
P(SOO) = P(OSO) = P(OOS) = $(0.6)(0.4^2)$ = 0.096, and
P(OOO) = 0.4^3 = 0.064.

(c) Let the random variable *X* be the number of student representatives who oppose the funding of interest groups. Give the probability distribution of *X*.

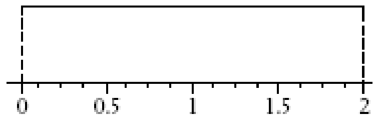
Value of *X*	0	1	2	3
Probability	0.216	0.432	0.288	0.064

(d) Express the event "a majority of the advisory board opposes funding" in terms of *X* and find its probability. Write either $X \geq 2$ or $X > 1$. The probability is 0.288 + 0.064 = 0.352.

7.17 **A UNIFORM DISTRIBUTION** Many random number generators allow users to specify the range of the random numbers to be produced. Suppose that you specify that the range is to be $0 \leq Y \leq 2$. Then the density curve of the outcomes has constant height between 0 and 2, and height 0 elsewhere.

(a) What is the height of the density curve between 0 and 2? Draw a graph of the density curve.

The height should be ½ since the area under the curve must be 1.



(b) Use your graph from (a) and the fact that probability is area under the curve to find $P(Y \leq 1)$

$P(Y \leq 1) = \frac{1}{2}$

(c) Find $P(0.5 < Y < 1.3)$. $P(0.5 < Y < 1.3) = 0.4$

(d) Find $P(Y \geq 0.8)$. $P(Y \geq 0.8) = 0.6$

HW #45 ***2.28, 29, 31, 32 **7.22, 25, 26, 29

2.28 **NORMAL REVIEW** Use your calculator to find the proportion of observations from a standard normal distribution that falls in each of the following regions. *In each case, sketch a standard normal curve and shade the area representing the region.*

(a) $z \leq -2.25$ P($z \leq -2.25$) = [normalcdf(-10, -2.25)] = 0.0122

(b) $z \geq -2.25$ P($z \geq -2.25$) = [normalcdf(-2.25, 10)] = 0.9878

(c) $z > 1.77$ P($z > 1.77$) = [normalcdf(1.77, 10)] = 0.0385

(d) $-2.25 < z < 1.77$ P($-2.25 < z < 1.77$) = [normalcdf(-2.25, 1.77)] = 0.9494

2.29 **MORE NORMAL REVIEW** Use your calculator to find the value z of a standard normal variable that satisfies each of the following conditions. *In each case, sketch a standard normal curve with your value of z marked on the axis.*

(a) The point z with 70% of the observations falling below it. About 0.52 [invNorm(0.7)]

(b) The point z with 85% of the observations falling above it. About -1.04 [invNorm(0.15)]

(c) Find the number z such that the proportion of observations that are less than z is 0.8 About 0.84 [invNorm(0.8)]

(d) Find the number z such that 90% of all observations are greater than z. About -1.28 [invNorm(0.1)]

2.31 **GESTATION PERIOD** The length of human pregnancies from conception to birth varies according to a distribution that is approximately normal with mean 266 days and standard deviation 16 days.

(a) What percent of pregnancies last less than 240 days (that's about 8 months)? About 5.21%

$$P(X < 240) = P\left(Z < \frac{240-266}{16}\right) = P(z < -1.625) = 0.0521$$ [normalcdf(-10, -1.625)]

(b) What percent of pregnancies last between 240 and 270 days (roughly between 8 months and 9 months)? About 55%

$$P(240 < X < 270) = P\left(\frac{240-266}{16} < Z < \frac{270-266}{16}\right) = P(-1.625 < Z < 0.25) = 0.5466$$

[normalcdf (-1.625, -0.25)]

(c) How long do the longest 20% of pregnancies last? Approximately 279 days or longer.

invNorm(0.8) = 0.8416 Solving $\frac{x-266}{16} = 0.8416$, *gives* $x = 279$ *days.*

2.32 **ARE WE GETTING SMARTER?** When the Stanford-Binet "IQ test" came into use in 1932, it was adjusted so that scores for each age group of children followed roughly the normal distribution with mean μ = 100 and standard deviation σ = 15. The test is readjusted from time to time to keep the mean at 100. If present-day American children took the 1932 Stanford-Binet test, their mean scores would be about 120. The reasons for the increase in IQ over time are not known but probably include better childhood nutrition and more experience in taking tests.

(a) IQ scores above 130 are often called "very superior." What percent of children had very superior scores in 1932? $P(X > 130) = P\left(Z > \frac{130-100}{15}\right) = P(Z > 2) = 0.0228$

About 2.3% of children had very superior scores in 1932.

(b) If present-day children took the 1932 test, what percent would have very superior scores? (Assume that the standard deviation σ = 15 does not change.)

$$P(X > 130) = P\left(Z > \frac{130-120}{15}\right) = P(Z > 0.67) = 0.2514$$

About 25.1% of present-day children taking the 1932 test would have very superior scores.

7.22 **GRADE A DISTRIBUTION** The table below gives the distribution of grades {A = 4, B = 3, and so on} in a large class:

Grade:	0	1	2	3	4
Probability:	0.10	0.15	0.30	0.30	0.15

Find the average (that is, the mean) grade in this course.

μ_X = (0)(0.10) + (1)(0.15) + (2)(0.30) + (3)(0.30)+ (4)(0.15) = 2.25.

7.25 **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 the 80 numbers are chosen, your probability of winning is 20/80, or 0.25.

(a) What is the probability distribution (the outcomes and their probabilities) of the payoff X on a single play?

Value of X	0	3
Probability:	0.75	0.25

The payoff is either \$0 or \$3.

(b) What is the mean payoff m_X? For each \$1 bet, m_X = (\$0)(.75) + (\$3)(0.25) = 0.75.

(c) In the long run, how much does the casino keep from each dollar bet?

The casino makes 25 cents for every dollar bet (in the long run).

7.26 **GRADE DISTRIBUTION, II** Find the standard deviation s_X of the distribution of grades in Exercise 7.22. In 7.22, we had μ = 2.25, so

$$s_X^2 = (0-2.25)^2(0.10) + (1-2.25)^2(0.15) + (2-2.25)^2(0.30) + (3-2.25)^2(0.30) + (4-2.25)^2(0.15) = 1.3875$$

$$s_X = \sqrt{1.3875} = 1.178$$

7.29 **KIDS AND TOYS** In an experiment on the behavior of young children, each subject is placed in an area with five toys. The response of interest is the number of toys that the child plays with. Past experiments with many subjects have shown that the probability distribution of the number X of toys played with 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) Calculate the mean m_X and the standard deviation

m_X = (0)(0.03) + (1)(0.16) + (2)(0.30) + (3)(0.23) + (4)(0.17) + (5)(0.11) = 2.68

$s_X^2 = (0-2.68)^2(0.03) + (1-2.68)^2(0.16) + (2-2.68)^2(0.30) + (3-2.68)^2(0.23) + (4-2.68)^2(0.17) + (5-2.68)^2(0.11) = 1.7176$, and $s_X = \sqrt{1.7176} = 1.3106$

(b) Describe the details of a simulation you could carry out to approximate the mean number of toys m_X and the standard deviation s_X. Then carry out your simulation. Are the mean and standard deviation produced from your simulation close to the values you calculated in (a)?

To simulate (say) 500 observations of *x*, using the T1-84, we will first simulate 500 random integers between 1 and 100 by using the command: randInt(1,100,500) →L1
The command sortA(L1) sorts these random observations in increasing order.
We now identify 500 observations of *x* as follows:

Integers		correspond to
	1 to 3	$x = 0$
	4 to 19	$x = 1$
	20 to 49	$x = 2$
	50 to 72	$x = 3$
	73 to 89	$x = 4$
	90 to 100	$x = 5$

For a sample run of the simulation, we obtained

	observations of
12	$x = 0$
86	$x = 1$
155	$x = 2$
118	$x = 3$
75	$x = 4$
54 .	$x = 5$

These data yield a sample mean and standard deviation of $\bar{x} = 2.64$ *and* $s = 1.292$ very close to μ and σ.

HW #46 32, 33, 35, 37, 38, 41

7.32 (a) A gambler knows that red and black are equally likely to occur on each spin of a roulette wheel. He observes five consecutive reds and bets heavily on red at the next spin. Asked why, he says that "red is hot" and that the run of reds is likely to continue. Explain to the gambler what is wrong with this reasoning. The wheel is not affected by its past outcomes—it has no memory; outcomes are independent. So on any one spin, black and red remain equally likely.

(b) After hearing you explain why red and black remain equally probable after five reds on the roulette wheel, the gambler moves to a poker game. He is dealt five straight red cards. He remembers what you said and assumes that the next card dealt in the same hand is equally likely to be red or black. Is the gambler right or wrong? Why? He is wrong. Removing a card changes the composition of the remaining deck, so successive draws are not independent. If you hold 5 red cards, the deck now contains 5 fewer red cards, so your chance of another red decreases.

7.33 **OVERDUE FOR A HIT** Retired baseball player Tony Gwynn got a hit about 35% of the time over an entire season. After he failed to hit safely in six straight at-bats, a TV commentator said "Tony is due for a hit by the law of averages." Is that right? Why?

No: Assuming all "at-bat"s are independent of each other, the 35% figure only applies to the "long run" of the season, not to "short runs."

7.35 **CHECKING INDEPENDENCE, II** In which of the following games of chance would you be willing to assume independence of *X* and *Y* in making a probability model? Explain your answer in each case.

(a) In blackjack, you are dealt two cards and examine the total points *X* on the cards (face cards count 10 points). You can choose to be dealt another card and compete based on the total points *Y* on all three cards. Dependent: since the cards are being drawn from the deck without replacement, the nature of the third card (and thus the value of *Y*) will depend upon the nature of the first two cards that were drawn (which determine the value of *X*).

(b) In craps, the betting is based on successive rolls of two dice. *X* is the sum of the faces on the first roll, and *Y* is the sum of the faces on the next roll. Independent: *X* relates to the outcome of the first roll, *Y* to the outcome of the second roll, and individual dice rolls are independent (the dice have no memory).

7.37 **TIME AND MOTION, I** A time and motion study measures the time required for an assembly-line worker to perform a repetitive task. The data show that the time required to bring a part from a bin to its position on an automobile chassis varies from car to car with mean 11 seconds and standard deviation 2 seconds. The time required to attach the part to the chassis varies with mean 20 seconds and standard deviation 4 seconds.

(a) What is the mean time required for the entire operation of positioning and attaching the part?
The total mean is 11 + 20 = 31 seconds.

(b) If the variation in the worker's performance is reduced by better training, the standard deviations will decrease. Will this decrease change the mean you found in (a) if the mean times for the two steps remain as before? No, changing the standard deviations does not affect the means.

(c) The study finds that the times required for the two steps are independent. A part that takes a long time to position, for example, does not take more or less time to attach than other parts. How would your answers to (a) and (b) change if the two variables were dependent with correlation 0.8? With correlation 0.3? The answers would not change. The total mean does not depend on dependence or independence of the two variables.

7.38 **TIME AND MOTION, II** Find the standard deviation of the time required for the two-step assembly operation studied in the preceding exercise, assuming that the study shows the two times to be independent. Redo the calculation assuming that the two times are dependent, with correlation 0.3. Can you explain in nontechnical language why positive correlation increases the variability of the total time. Assuming that the two times are independent, the total variance is

$$s_{total}^2 = s_{pos}^2 + s_{att}^2 = 2^2 + 4^2 = 20, \text{ so } s_{total} = \sqrt{20} = 4.472 \text{ seconds.}$$

Assuming that the two times are dependent with correlation 0.3, the total variance is

$$s_{total}^2 = s_{pos}^2 + s_{att}^2 + 2r\, s_{pos} s_{att} = 2^2 + 4^2 + 2(0.3)(2)(4) = 24.8,$$

$$\text{so } s_{total} = \sqrt{24.8} = 4.98 \text{ seconds.}$$

The positive correlation of 0.3 indicates that the two times have some tendency to either increase together or decrease together, which increases the variability of their sum.

7.41 Leona and Fred are friendly competitors in high school. Both are about to take the ACT college entrance examination. They agree that if one of them scores 5 or more points better than the other, the loser will buy the winner a pizza. Suppose that in fact Fred and Leona have equal ability, so that each score varies normally with mean 24 and standard deviation 2. (The variation is due to luck in guessing and the accident of the specific questions being familiar to the student.) The two scores are independent. What is the probability that the scores differ by 5 or more points in either direction?

If F and L are their respective scores, then F − L has a $N\left(0, \sqrt{2^2 + 2^2}\right)$ distribution, so

P(Fred scores 5 points more than Leona) + P(Leona scores 5 points more than Fred) =

$$P(F - L > 5) + P(L - F > 5) = P\left(Z > \frac{5-0}{\sqrt{8}}\right) + P\left(Z < \frac{0-5}{\sqrt{8}}\right) =$$

$$P(Z > 1.7678) + P(Z < -1.7678) = 2P(Z > 1.7678) = 2(0.03854) = 0.0771$$

HW #47   44, 49, 55 – 61, 62(skip a)

7.44 **WEIRD DICE** You have two balanced, six-sided dice. The first has 1, 3, 4, 5, 6, and 8 spots on its six faces. The second die 1, 2, 2, 3, 3, and 4 spots on its faces.

(a) What is the mean number of spots on the up-face when you roll each of these dice?

First die: $m = (1)(1/6) + (3)(1/6) + (4)(1/6) + (5)(1/6) + (6)(1/6) + (8)(1/6) = 4.5$

Second die: $m = (1)(1/6) + (2)(1/6) + (2)(1/6) + (3)(1/6) + (3)(1/6) + (4)(1/6) = 2.5$

(b) Write the probability model for the outcomes when you roll both dice independently. From this, find the probability distribution of the sum of the spots on the up-faces of the two dice.

	1	3	4	5	6	8
1	2	4	5	6	7	9
2	3	5	6	7	8	10
2	3	5	6	7	8	10
3	4	6	7	8	9	11
3	4	6	7	8	9	11
4	5	7	7	9	10	12

The probability distribution of *X* is:

x	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

(c) Find the mean number of spots on the two up-faces in two ways: from the distribution you found in (b) and by applying the addition rule to your results in (a). You should of course get the same answer.

$$m = (2)(1/36) + (3)(1/18) + (4)(1/12) + (5)(1/9) + (6)(5/36) + (7)(1/6)$$
$$(8)(5/36) + (9)(1/9) + (10)(1/12) + (11)(1/18) + (12)(1/36) = 7$$

Using addition rule for means: μ = mean from 1^{st} die + mean from 2^{nd} die = 7.

7.49 A study of working couples measures the income *X* of the husband and the income *Y* of the wife in a large number of couples in which both the partners are employed. Suppose that you knew the means m_X and m_Y and the variances s_X^2 and s_Y^2 of both variables in the population.

(a) Is it reasonable to take the mean of the total income *X* + *Y* to be $m_X + m_Y$? Explain your answer.

Yes: This is always true; it does not depend on independence.

(b) Is it reasonable to take the variance of the total income to be $s_X^2 + s_Y^2$? Explain your answer.

No: It is not reasonable to believe that *X* and *Y* are independent.

7.55 **LIFE INSURANCE, I** 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 *X* 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 the distribution of *X*. Fill in the missing probability in the table and calculate the mean profit m_X.

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

The missing probability is 0.99058 (so that the sum is 1). This gives mean earnings m_X = $303.3525.

7.56 **LIFE INSURANCE, II** It would be quite risky for you to insure the life of a 21-year-old friend under the terms of the previous exercise. There is a high probability that your friend would live and you would gain in $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.

The mean μ of the company's "winnings" (premiums) and their "losses" (insurance claims) is positive. Even though the company will lose a *large* amount of money on a *small* number of policyholders who die, it will gain a *small* amount on the majority. The law of large numbers says that the average "winnings" minus "losses" should be close to μ, and overall the company will almost certainly show a profit.

7.57 **LIFE INSURANCE, III** 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 larger σ is), the riskier the investment. We can measure the great risk of insuring a single person's life in Exercise 7.55 by computing the standard deviation of the income *X* that the insurer will receive. Find s_X, using the distribution and mean found in Exercise 7.55.

$s_X^2 = 94,236,826.64, \textit{ so that } s_X = \$9707.57$

7.58 **LIFE INSURANCE, IV** The risk of insuring one person's life is reduced if we insure many people. Use the result of the previous exercise and the rules for means and variances to answer the following questions.

(a) Suppose that we insure two 21-year-old males, and that their ages at death are independent. If X and Y are the insurer's income from the two insurance policies, the insurer's average income on the two policies is:

$$Z = \frac{X+Y}{2} = 0.5X + 0.5Y$$

Find the mean and standard deviation of Z. You see that the mean income is the same as for a single policy but the standard deviation is less.

$$m_Z = \frac{1}{2} m_T = m_X = \$303.3525 \quad s_Z = \sqrt{\frac{1}{4} s_X^2 + \frac{1}{4} s_Y^2} = \sqrt{\frac{1}{2} s_X^2} = \$6864.29$$

(b) If four 21-year-old mean are insured, the insurer's average income is

$$Z = \frac{1}{4}(X_1 + X_2 + X_3 + X_4)$$

Where X_i is the income from insuring one man. The X_i are independent and each has the same distribution as before. Find the mean and standard deviation of Z. Compare your results with the results of (a). We see that averaging over many insured individuals reduces risk.

Using the new definition of Z, we have

$$m_Z = \$303.3525\,(\textit{unchanged})\ \textit{and}\ s_Z = \sqrt{\frac{1}{4} s_z^2} = \frac{1}{2} s_X = \$4853.78\ \ (\textit{smaller by a factor of}\ 1/\sqrt{2}\,)$$

7.59 **AUTO EMISSIONS** The amount of nitrogen oxides (NOX) present in the exhaust of a particular type of car varies from car to car according to the normal distribution with mean 1.4 grams per mile (g/mi) and standard deviation 0.3 g/mi. Two cars of this type are tested. One has 1.1 g/mi of NOX, the other 1.9. The test station attendant finds this much variation between two similar cars surprising. If X and Y are independent NOX levels for cars of this type, find the probability

$$P(X - Y \geq 0.8 \quad or \quad X - Y \leq -0.8)$$

that the difference is at least as large as the value the attendant observed.

$X - Y$ *is* $N(0, \sqrt{0.3^2 + 0.3^2}\,) = N(0, 0.4243)$,

$$so\ \ P(|X - Y| \geq 0.8) = P\left(|Z| \geq \frac{0.8-0}{0.4243}\right) = 2P(Z \geq 1.8856) = 0.0594$$

2[normalcdf(1.8856, 10)]

7.60 **MAKING A PROFIT** Rotter Partners is planning a major investment. The amount of profit *X* is uncertain but a probabilistic estimate gives the following distribution (in millions of dollars):

Profit:	1	1.5	2	4	10
Probability:	0.1	0.2	0.4	0.2	0.1

(a) Find the mean profit m_X and the standard deviation of the profit.

$m_X =$ (1)(0.1) + (1.5)(0.2) + (2)(0.4) + (4)(0.2) + (10)(0.1) = 3 million dollars.

$s_X^2 =$ (4)(0.1) + (2.25)(0.2) + (1)(0.4) + (1)(0.2) + (49)(0.1) = 6.35 million dollars

So $s_X =$ 2.52 million dollars.

(b) Rotter Partners owes its source of capital a fee of $200,000 plus 10% of the profits *X*. So the firm actually retains

$$Y = 0.9X - 0.2$$

from the investment. Find the mean and standard deviation of *Y*.

$m_Y = 0.9m_X - 0.2 = 2.5$ *million dollars*

$s_Y = 0.9s_X = 2.268$ *million dollars*

7.61 **A BALANCED SCALE** You have two scales for measuring weights in a chemistry lab. Both scales give answers that vary a bit in repeated weighings of the same item. If the true weight of a compound is 2.00 grams (g), the first scale produces readings *X* that have a mean of 2.000 g and standard deviation 0.002 g. The second scale's readings *Y* have a mean of 2.001 g and standard deviation 0.001 g.

(a) What are the mean and standard deviation of the difference $Y - X$ between the readings? (The readings *X* and *Y* are independent.)

$m_{Y-X} = m_Y - m_X = 2.001 - 2.000 = 0.001\ g$

$s_{Y-X}^2 = s_Y^2 + s_X^2 = 0.002^2 + 0.001^2 = 0.000005,\ so\ s_{Y-X} = 0.002236\ g.$

(b) You measure once with each scale and average the readings. Your result is *Z* = *(X* + *Y)*/2. What are m_Z and s_Z? Is the average *Z* more or less variable than the reading *Y* of the less variable scale?

$m_Z = \frac{1}{2}m_X + \frac{1}{2}m_Y = 2.0005\ g.$

$s_Z^2 = \frac{1}{4}s_X^2 + \frac{1}{4}s_Y^2 = 0.00000125,\ so\ s_Z = 0.001118\ g.$

Z is slightly more variable than Y, so $s_Y < s_Z$

7.62 **IT'S A GIRL!** A couple plans to have children until they have a girl or until they have four children, whichever comes first. Example 5.24 (text p. 313) estimated the probability that they will have a girl among their children. Now we ask a different question: How many children, on the average, will couples who follow this plan have?

(a) To answer this question, construct a simulation similar to that in Example 5.24 but this time keep track of the number of children in each repetition. Carry out 25 repetitions and then average the results to estimate the expected value.

To do one repetition, start at any point in Table B and begin reading digits. Let the digits 0, 1, 2, 3, 4 = girl and 5, 6, 7, 8, 9 = boy, and read a string of digits until a "0 to 4" (girl) appears or until four consecutive "5 to 9"s (boys) have appeared, whichever comes first. Then let the observation of X = number of children for this repetition = the number of digits in the string you have read. Repeat this procedure 25 times to obtain your 25 observations.

(b) Construct the probability distribution table for the random variable X = number of children.

The possible outcomes and their corresponding values of X = number of children are as follows:

	Outcome	
$x = 1$	G	(first child is a girl)
$x = 2$	BG	(second child is a girl)
$x = 3$	BBG	(third child is a girl)
$x = 4$	BBBG,BBBB	(four children)

Using the facts that births are independent, the fact that B and G are equally likely to occur on any one birth, and the multiplication rule for independent events, we find that

$P(x = 1) = 1/2$

$P(x = 2) = (1/2)(1/2) = 1/4$

$P(x = 3) = (1/2)(1/2)(1/2) = 1/8$

$P(x = 4) = (1/2)(1/2)(1/2)(1/2) + (1/2)(1/2)(1/2)(1/2) = 1/16 + 1/16 = 1/8$

The probability distribution of X is therefore:

x_i	1	2	3	4
p_i	1/2	1/4	1/8	1/8

(c) Use the table from (b) to calculate the expected value of X. Compare this number with the result from your simulation in (a).

$$m_X = \sum x_i p_i$$

$= (1)(1/2) + (2)(1/4) + (3)(1/8) + (4)(1/8)$

$= 1/2 + 1/2 + 3/8 + 1/2$

$= 1.875$