> **Whenever X and Y are independent, then the rows of the table $p(x, y)$ will be proportional, and so will the columns.** (5-7)

PROBLEMS

5-1 The approximately 100 million adult Americans (age 25 and over in 1985) were roughly classified by education X and age Y as follows (Stat. Abst. of U.S., 1987, p. 122. Actually there were 144 million, but we have conveniently scaled down to 100 million).

Education X (*last completed school*)		Age Y (25–35) 30	(35–55) 45	(55–100) 70
none	0	1,000,000	2,000,000	5,000,000
primary	1	3,000,000	6,000,000	10,000,000
secondary	2	18,000,000	21,000,000	15,000,000
college	3	7,000,000	8,000,000	4,000,000

Suppose the Gallup poll draws an adult American at random.

a. What is the probability of getting a 30-year-old college graduate ($X = 3$ and $Y = 30$, in the lower left cell)? And what is the probability of getting each of the 12 possible combinations of education and age? That is, tabulate the joint probability distribution $p(x, y)$.

b. Calculate $p(x)$ and $p(y)$.

c. Are X and Y independent?

d. Calculate μ_X and σ_X.

e. Graph the joint distribution, as in Figure 5-2.

5-2 **a.** Continuing Problem 5-1, look at just those Americans aged 30. Tabulate their distribution of education X—call it the *conditional* distribution $p(x|y)$ for $Y = 30$. Then calculate their average education—call it $E(X|Y = 30)$.

b. Similarly calculate $E(X|Y = 45)$, and $E(X|Y = 70)$. Mark them all on the graph of $p(x, y)$. Describe in words what this shows.

c. In view of part **b**, would you say education X is independent of age Y? Is this consistent with your answer to Problem 5-1**c**?

5-3 What we calculated in Problem 5-2 can alternatively be done with formulas. The conditional distribution of X is defined as:

$$p(x|y) \equiv \frac{p(x, y)}{p(y)} \qquad \text{like (3-17)}$$

Use this to calculate $p(x|y)$ for $Y = 30$. Does it agree with your answer in Problem 5-2**a**?

5-4 For which of the following joint distributions are X and Y independent?

a.

x	y = 1	y = 2
1	.10	.20
2	.30	.40

b.

x	y = 0	y = 1	y = 2
0	.10	.20	.10
1	.15	.20	.25

c.

x	y = 0	y = 1	y = 2
1	0	.1	.1
2	.1	.4	.1
3	.1	.1	0

d.

x	y = 1	y = 2	y = 3	y = 4
1	.12	.03	.06	.09
2	.20	.05	.10	.15
3	.08	.02	.04	.06

5-5 Using the table in Problem 5-4**c**, calculate and tabulate $p(x|y)$ for $Y = 1$. Is it the same as the unconditional distribution $p(x)$? What does this indicate about the independence of X and Y? Is this consistent with your earlier answer?

5-6 Repeat Problem 5-5, using the table in Problem 5-4**d**.

5-7 A salesman has an 80% chance of making a sale on each call. If three calls are to be made next month, let

$$X = \text{total number of sales}$$

$$Y = \text{total profit from the sales}$$

where the profit Y is calculated as follows: Any sales on the first two calls yield a profit of \$100 each. By the time the third call is made, the original product has been replaced by a new product whose sale yields a profit of \$200. Thus, for example, the sequence (sale, no sale, sale) would give $Y = \$300$.

List the sample space, and then:

a. Tabulate and graph the bivariate distribution.

b. Calculate the marginal distribution of X and of Y. Does the distribution of X agree with Problem 4-4?

c. What is the mean of X and of Y?

d. Are X and Y independent?

a. Find $E(R)$ by first finding its distribution $p(r)$ and then calculating $\Sigma rp(r)$, as in (5-8).

b. Find $E(R)$, that is, $E(X^2 + Y^2)$, by using $\Sigma\Sigma g(x, y)p(x, y)$, as in (5-9).

c. Find the following expected values any way you like:

i. $E(X - 2)(Y - 2)$

ii. $E(X - 2)^2$

iii. $E(4X + 2Y)$

5-9 In a stand of 380 hardwood trees of marketable size, the diameter at waist height (D) and usable height (H) of each tree was measured (in feet). The 380 pairs of measurements were tallied in the following table, and then converted to relative frequency:

	h	
d	20	25
1.0	.16	.09
1.25	.15	.30
1.50	.03	.17
1.75	.00	.10

Since a tree trunk is roughly a cylinder, the volume V of usable wood that it contains is given approximately by:

$$V = .4D^2H$$

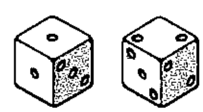
a. Calculate the average volume per tree $E(V)$.

b. Calculate the standard deviation of V.

c. What is the total volume in the whole stand of 380 trees?

d. Calculate the average diameter $E(D)$ and the average height $E(H)$. Is it true that *average* volume obeys the same formula as volume itself, namely $E(V) = .4[E(D)]^2[E(H)]$?

5-10 In a certain gambling game, a pair of dice have their faces renumbered: The sides marked 4, 5, and 6 are renumbered with 1, 2, and 3, so that each die can have only three outcomes—each with the same probability $2/6 = 1/3$.



If the two dice are thrown independently, tabulate the joint distribution of:

X = number on the first die

Y = number on the second die

Then find the mean, variance, and standard deviation for each of the following:

a. X

b. Y

c. The total number of dots, $S = X + Y$

d. Note that S is just the sum of X and Y. Is the mean of S similarly just the sum of the means? Is the variance just the sum of the variances? And is the standard deviation just the sum of the standard deviations?

5-11 Suppose the two dice in Problem 5-10 have their faces renumbered, so that now the probabilities for X and Y are as follows:

x	p(x)
1	1/6
2	2/6
3	3/6

y	p(y)
1	3/6
2	1/6
3	2/6

Answer the same questions as before.

5-3 COVARIANCE

A—COVARIANCE IS LIKE VARIANCE

In this section we will develop the *covariance* to measure how two variables X and Y vary together, using the familiar concept of $E[g(X, Y)]$.

First recall our measure of how X alone varies: We started with the deviations $X - \mu$, squared them, and then took the expectation:

$$\sigma_X^2 = \text{Variance of } X \equiv E(X - \mu)^2 \qquad \text{(4-34) repeated}$$

To measure how two variables X and Y vary together, we again start with the deviations, $X - \mu_X$ and $Y - \mu_Y$. We multiply them, and then take the expectation:

$$\boxed{\sigma_{X,Y} = \text{Covariance of } X \text{ and } Y \equiv E(X - \mu_X)(Y - \mu_Y)} \qquad (5\text{-}10)$$

An example will show how similar the covariance is to the variance. (Incidentally, if you have done Problem 5-8**c**, you have already calculated a covariance as well as a variance, without even realizing it.)

the limiting value of +1.[2] Similarly, if there is a perfect negative linear relation, then ρ would be -1.

To illustrate these bounds, we can calculate ρ for the data in Example 5-3.

$$\rho = \frac{49}{\sqrt{60}\sqrt{70}} = .76$$

This is indeed less than the upper bound of 1.

PROBLEMS

5-12 For each of the following joint distributions, calculate σ_{XY} from the definition (5-10), and from the easier formula (5-11). Then calculate ρ_{XY}.

a.

x \ y	0	1
0	.2	0
1	.4	.2
2	0	.2

b.

x \ y	0	5	10
2	0	.1	.1
4	.1	.4	.1
6	.1	.1	0

c.

x \ y	1	2
0	.06	.04
1	.30	.20
2	.24	.16

d.

x \ y	1	2	3
0	1/8	0	0
1	0	2/8	1/8
2	0	2/8	1/8
3	1/8	0	0

5-13 **a.** In Problem 5-12, parts **c** and **d**, what is ρ? Are X and Y independent?

b. Looking beyond these particular examples, which of the following statements are true for *any* X and Y?

1. If X and Y are independent, then they must be uncorrelated.
2. If X and Y are uncorrelated, then they must be independent.

5-14 Does money make you happy?

Americans were asked in 1971 to rank themselves on a "happiness index" as follows: $H = 0$ (not happy), $H = 1$ (fairly happy), or

[2] Since $X = Y$, (5-10) becomes

$$\sigma_{X,Y} = E(X - \mu_X)(X - \mu_X)$$
$$= E(X - \mu_X)^2 = \sigma_X^2$$

Also note that, since $X = Y$, $\sigma_X = \sigma_Y$. When we substitute all this into (5-13), ρ becomes $\sigma_X^2/\sigma_X\sigma_X = 1$.

$H = 2$ (very happy). Annual household income X (in thousands of dollars) was also recorded for each individual. Then the relative frequencies of various combinations of H and X were roughly as follows (Gallup, 1971):

Joint Distribution of Happiness H and Income X

	h		
x	0	1	2
2.5	.03	.12	.07
7.5	.02	.13	.11
12.5	.01	.13	.14
17.5	.01	.09	.14

a. Calculate $E(H|X)$ for the various levels of X, and mark them on the graph of the bivariate distribution (as in Problem 5-2).

b. Calculate the covariance and correlation.

c. Answer True or False; if False, correct it:

i. As X increases, the average level of H increases. This positive relation is reflected in a positive correlation ρ.

ii. Yet the relation is just a *tendency* (H fluctuates around its average level), so that ρ is less than 1.

iii. This shows that the rich tend to be happier than the poor, that is, money tends to make people happier.

5-15 Does education make you happy?

In Problem 5-14, the amount of education was also measured for each individual: $X = 1$ (elementary school completed), $X = 2$ (high school completed), or $X = 3$ (college completed, or more). Thus X = number of schools completed. Then the relative frequencies of various combinations were roughly as follows (Gallup, 1971):

Joint Distribution of Happiness H and Education X

	h		
x	0	1	2
1	.02	.08	.05
2	.02	.28	.25
3	.01	.13	.16

Repeat the same questions as in Problem 5-14 (with education substituted for income, of course).

TABLE 5-7 The Mean and Variance of Functions of X and Y

Function of X and Y	*Mean*	*Variance*
1. Any function $g(X, Y)$	$E[g(X, Y)] = \sum_x \sum_y g(x, y)p(x, y)$	
2. Linear combination $aX + bY$	$E(aX + bY) = aE(X) + bE(Y)$	$\text{var}(aX + bY) = a^2\text{var } X + b^2\text{var } Y + 2ab \text{ cov }(X, Y)$
3. Simple sum $X + Y$	$E(X + Y) = E(X) + E(Y)$	$\text{var}(X + Y) = \text{var } X + \text{var } Y + 2 \text{ cov}(X, Y)$

PROBLEMS

5-16 Following (5-20), it was explained why the formula for $\text{var}(X + Y)$ includes the covariance term—when the covariance was positive. Give a similar explanation when the covariance is negative.

5-17 A marriage counseling office consisted of 10 couples. The annual incomes (in thousands of dollars) of the men and women were as follows:

Couple	*Man*	*Woman*
MacIntyre	20	15
Sproule	30	35
Carney	30	25
Devita	20	25
Peat	20	25
Matias	30	15
Steinberg	40	25
Aldis	30	25
Yablonsky	40	35
Singh	40	25

A couple is drawn by lot to represent the office at a workshop on personal finances. Let X and Y denote the randomly drawn income of the man and woman. Then find:

a. The bivariate probability distribution, and its graph.

b. The distribution, mean, and variance of X. Also of Y.

c. The covariance $\sigma_{X,Y}$.

d. If S is the combined income of the couple, what is its mean and variance? Calculate two ways: from the distribution of S, as in (5-8), and then using the easy formulas for sums.

e. Suppose (not very realistically) that the couple's income after taxes is $W = .6X + .8Y$. What is its mean and variance?

f. To measure the degree of sex discrimination against wives, a sociologist measured the difference $D = X - Y$. What is its mean and variance?

g. How good a measure of sex discrimination is $E(D)$?

5-18 Continuing Problem 5-17, we shall consider some alternative schemes for collecting the tax T on the couple's income S. Find the mean and standard deviation of T,

a. If S is taxed at a straight 20%—that is,

$$T = .20S$$

b. If S is taxed at 50%, with the first 15 thousand exempt—that is,

$$T = .5(S - 15)$$

c. If S is taxed according to the following progressive tax table:

S	T	S	T
30	4	55	11
35	5	60	13
40	6	65	16
45	7	70	19
50	9	75	22

5-19 Compare the three tax schemes in Problem 5-18 in terms of the following criteria:

i. Which scheme yields the most revenue to the government?

ii. Which scheme is most egalitarian; that is, which scheme results in the smallest standard deviation in net income left after taxes?

CHAPTER 5 SUMMARY

5-1 A pair of random variables X and Y has a joint probability distribution $p(x, y)$, from which the distributions $p(x)$ and $p(y)$ can be found in the margin. X and Y are then called independent if the simple multiplication rule holds: $p(x, y) = p(x)p(y)$ for all x and y.

5-2 A new random variable $R = g(X, Y)$ has an expected value that can be calculated from the joint distribution $p(x, y)$:

$$E[g(X, Y)] = \Sigma\Sigma\, g(x, y)p(x, y).$$

5-3 Just as variance measures how much one variable varies, so covariance measures how much two variables vary together. Its standardized version is the correlation $\rho = \sigma_{XY}/\sigma_X\sigma_Y$, which measures the degree of linear relation between X and Y.

5-4 The sum $X + Y$ is a particularly convenient function. Its expected value is simply $E(X) + E(Y)$, and its variance is $\text{var}(X) + \text{var}(Y) + 2\,\text{cov}(X, Y)$.

REVIEW PROBLEMS

Problems 5-20 to 5-25 form a sequence that gives a summary of this chapter, with a more applied flavor.

5-20 The approximately 100 million adult Americans (age 25 or over, in 1980) were roughly classified by education X, and sex Z, as follows (U.S. Current Population Reports, Series P-20, No. 390, Mar 1981 and 1980. Actually there were 133 million, but we have conveniently rescaled down to 100 million.):

Education X (last completed school)		*Sex Z*	
		0 (male)	*1 (female)*
none	0	4,000,000	5,000,000
primary	1	10,000,000	12,000,000
secondary	2	23,000,000	29,000,000
college	3	10,000,000	7,000,000

Suppose the Gallup poll draws an adult American at random.

a. What is the chance of getting a college-graduate female ($X = 3$ and $Z = 1$, in the lower-right cell)? Similarly, what is the chance of getting each of the 8 possible combinations of education and sex? That is, tabulate the (joint) probability distribution of X and Z.

b. What is the chance of getting a college graduate ($X = 3$)? Similarly, what is the chance of getting each of the 4 possible levels of education? That is, tabulate the (marginal) probability distribution of X.

c. What is the chance of getting a person with at least secondary school graduation ($X \geq 2$)?

d. Calculate the mean education $E(X)$.

e. Similarly, tabulate the (marginal) probability distribution of sex Z. Then calculate $E(Z)$.

5-21 Continuing Problem 5-20.

a. If the person is male, what are the probabilities of the 4 possible levels of education? That is, find the conditional distribution of X, given $Z = 0$.

b. Is this conditional distribution of X different from the unconditional distribution of X found in Problem 5-20? Is education X therefore dependent or independent of sex Z?

c. Calculate the conditional mean $E(X|Z = 0)$. That is, find the mean education of males.

d. Similarly, calculate $E(X|Z = 1)$. That is, find the mean education of females. Is it higher or lower than for males?

e. How does the unconditional mean $E(X)$ found in Problem 5-20 compare to the two conditional means found above?

5-22 Continuing Problem 5-20, suppose (contrary to fact, fortunately) that each American drew a salary S that depended only upon education and sex as follows (annual salary in \$000):

$$S = 10 + 2X + 4Z$$

(that is, a basic 10 thousand, plus bonuses for education and sex).

a. What would the salary be for each of the 8 combinations (cells) in the given table?

To reorganize your answer, list the possible salaries in order, and tabulate the corresponding probabilities (the distribution of salary).

b. From the distribution of salary, calculate the expected salary $E(S)$.

c. Since $S = 10 + 2X + 4Z$, try calculating its expected value alternatively as:

$$E(S) = 10 + 2E(X) + 4E(Z)$$

where $E(X)$ and $E(Z)$ were found in Problem 5-20**d** and **e**.

d. Why does your answer to **c** agree with **b**? (Is it a fluke, or is it because S is a special function?)

5-23 Continuing Problem 5-22, now suppose salary S has a different formula, as given below. In each case, calculate $E(S)$ as easily as you can. (Hint: If S is linear, you can use the shortcut in part c above. If S is nonlinear, you can still proceed as in parts **a** and **b** above.)

a. $S = 10 + 10XZ$

b. $S = 10 + 3X - 2Z$

c. $S = (2X + 40)/5$

d. $S = 10 + X^2$

5-24 The approximately 100 million adult Americans (age 25 or over in 1980) were roughly classified by age Y, as well as education X and sex Z, as follows (same source as Problem 5-20):

Education X (last completed school)		Sex Z: $Z = 0$ (male), Age Y: 30	45	70	$Z = 1$ (female), Age Y: 30	45	70
none	0	400,000	1,200,000	2,700,000	400,000	1,100,000	3,300,000
primary	1	1,600,000	3,300,000	4,700,000	1,700,000	3,600,000	6,300,000
secondary	2	8,400,000	9,100,000	5,600,000	9,300,000	11,500,000	8,600,000
college	3	3,700,000	4,200,000	2,100,000	3,000,000	2,800,000	1,400,000

a. If an adult American was drawn at random, what is the probability of drawing a female college graduate about age 30? In general, tabulate the probability for each of the 24 possible combinations of education, age, and sex.

b. Now suppose we did not need the detail of age. Instead, we just wanted the breakdown by education and sex. Find the probability of drawing a female college graduate. And in general, tabulate the probability for each of the 8 combinations of education and sex.

Round your table to two decimals (%), and see if it agrees with Problem 5-20.

c. Suppose we did not need the detail of sex. Instead, we just wanted the breakdown by education and age. Find the probability of drawing a college graduate about age 30. And in general, tabulate the probability for each of the 12 combinations of education and age.

Round your table to two decimals (%), and see if it agrees with Problem 5-1. (Read the table descriptions carefully.)

5-25 The detail (disaggregation) given in Problem 5-24 is often useful to clarify what is going on. For example, we noted in Problem 5-21**d** that the mean education for females was lower than for males. To track down the source of this difference,

a. Compare the mean education of males and females, for age 30.

b. Then repeat for age 45, and for age 70.

c. Summarize in a graph, and in words, what you have discovered.

d. Underline the correct choice:
In part (c) we found females had about .06 units of education less than males, at each age level. Earlier in Problem 5-21(d) the analysis that ignored age showed an overall difference that was [the same of course, surprisingly larger]. This confounding occurs because there are [more, fewer] older women than men, and it is the older people who tend to have [more, less] education.
To restate this point, suppose at each age women achieved the same education as men. Then women's average education overall would [still be somewhat less than men's, now be equal to men's].
Routinely and easily controlling for confounding factors—not just age, but any others as well—is what [multiple regressions, confidence intervals] do, and we will accordingly analyze this data again in Problem 14-8.

5-26 *A Final Challenge: How to Hedge Your Investment Bets*

You have $10,000 saved up to invest for a year, and are considering stocks and/or short-term Treasury bills. The returns from both sources are judged uncertain, of course, as the following personal probability table indicates:

Bivariate Probability Distribution for Annual Rates of Return on Stocks (*S*) and Treasury Bills (*T*)

	S			
T	−10%	0	10%	20%
6%	0	0	.10	.10
8%	0	.10	.30	.20
10%	.10	.10	0	0

a. Without calculating, state whether the covariance of S and T will be positive, zero, or negative. Now calculate the covariance to confirm.

b. If you invested your $10,000 entirely in stocks, what would be the expected value of your return? And the standard deviation?
Repeat, for investing entirely in Treasury bills, for splitting your investment 50-50, then finally for splitting 20-80. Then record your four answers in the following table:

REVIEW PROBLEMS, CHAPTERS 1 TO 5

5-27 Circle the correct choice in each of the five following brackets. Note the following abbreviations:

RCE = randomized controlled experiment
OS = observational study

To really find out how effective a treatment is, a [*RCE, OS*] is better than a [*RCE, OS*], where feasible. This is because a [*RCE, OS*] actually makes the treatment and control groups equal on average in every respect—except for the treatment itself, of course; whereas a [*RCE, OS*] is usually cluttered up with confounding factors that bias the answer. To the extent these confounding factors can be measured and analyzed with a [*multiple regression, confidence interval*], however, bias can be reduced.

5-28 Suggest some plausible reasons for each of the following relations:

a. Over a period of several years, the relation between the number of colds reported weekly in a city and the amount of beer sold was found to be negative (low numbers of colds associated with high beer sales). Does beer prevent colds?

b. Over a period of 50 years, the relation between clergymen's annual salaries and annual alcohol sales in the U.S. was found to be positive. Would paying clergymen more increase alcohol sales?

c. The Irish humorist George Bernard Shaw once observed that there was a strong positive relation between a young man's income and clothes. His advice to a young man aspiring to be rich was therefore simple: Buy a black umbrella and top hat.

5-29 The distribution of the earned income of American men in 1975 was very roughly as follows:

Annual Income ($000)	*Proportion of Men Earning That Income*
0–5	36%
5–10	23%
10–15	20%
15–20	11%
20–25	5%
25–30	3%
30–35	2%

a. Graph this distribution.

b. Calculate the mean income, and the standard deviation, and show them on the graph.

c. What is the mode? Where will the median be in relation to the mean and mode?

d. What was the total earned income of the whole population of 78 million men?

5-30 A test consists of eight multiple-choice questions, each with a choice of 5 answers. Let X be the number of correct answers for a student who resorts to pure guessing.

a. Graph the distribution of X.

b. Calculate μ_X and σ_X and show them on the graph.

c. If the instructor calculates a rescaled mark $Y = 10X + 20$, what are μ_Y and σ_Y?

d. If the passing mark is $Y = 50$, what is the chance the student who resorts to pure guessing will pass?

5-31 To find out who won the Reagan-Carter TV debate before the 1980 Presidential election, ABC News conducted a phone-in survey. Nearly a million people volunteered to call in: 723,000 believed that Reagan had won the debate, and 244,000 believed Carter had (*New York Times*, Oct. 29, 1980).

CBS on the other hand conducted a very small sample survey. Of all the registered voters who had watched the debate, 1,019 were polled at random by telephone: 44% believed that Reagan had won the debate, 36% believed Carter had, 14% called it a tie, and 6% had no opinion.

Do these two polls agree, allowing for sampling fluctuation? Explain why, or why not. Include a critique of each poll.

5-32 A firm specializing in political polling took a preliminary random sample of a dozen voters from a suburb that is 60% Republican, 20% Democratic, and 20% Independent. What is the chance that a majority in the sample are Republican (6 or more)?

5-33 The following table gives, for a thousand newborn American baby boys (white U.S. males, 1970), the approximate number dying, in successive decades:

a. The third column shows the number surviving at the beginning of each decade. Complete this tabulation of $L(x)$.

b. The mortality rate in the last column is just the number dying during the decade, relative to the number living at the beginning of the decade. Complete the tabulation of the mortality rate $m(x)$. Then answer True or False. If False, correct it.

i. The mortality rate is lowest during the first decade of life.

ii. Roughly speaking, the mortality rate nearly doubles every decade.

c. Ten-year term insurance is an agreement whereby a man pays the insurance company \$x (the "premium," which usually is spread throughout the decade in 120 monthly payments) in return for a payment of $1000 to the man's estate if he dies. In order for this to be a "fair bet" (and ignoring interest), what should the premium x be for a man:

i. In his 20s?

ii. In his 40s?

The work that we have done so far could be just as well expressed in probability terms. For example, find:

i. The probability of a man dying in his 40s.

ii. The probability of a man surviving to age 40.

iii. The conditional probability of a person dying in his 40s, given that he has survived to 40.

Age, x	$n(x)$ = *Number Dying Within the Decade*	$L(x)$ = *Number Living to the Beginning of the Decade*	*Mortality Rate, Per Decade* $m(x) = \frac{n(x)}{L(x)}$
0 to 10	26	1,000	.026
10 to 20	9	974	.009
20 to 30	18	965	.019
30 to 40	21	947	.022
40 to 50	49	·	·
50 to 60	117	·	·
60 to 70	219	·	·
70 to 80	282		
80 to 90	208		
90 to 100	51		
	Total = 1000		

5-34 Referring to the mortality table in Problem 5-33:

a. Find the mean age at death (i.e., mean length of life, also called life expectancy).

*b. If the mortality rate were 50% higher after age 40, how much would this reduce life expectancy? (Incidentally, heavy cigarette smoking—two or more packs per day—seems to be associated with roughly this much increase in mortality according to the U.S. Department of Health, Education, and Welfare, 1964.)

5-35 An apartment manager in Cincinnati orders three new refrigerators, which the seller guarantees. Each refrigerator has a 20% probability of being defective.

a. Tabulate the probability distribution for the total number of defective refrigerators, X.

b. What are the mean and variance of X?

c. Suppose that the cost of repair, in order to honor the guarantee, consists of a fixed fee ($10) plus a variable component ($15 per defective refrigerator). That is:

$$c(x) = \begin{cases} 0 & \text{if } x = 0 \\ 10 + 15x & \text{if } x > 0 \end{cases}$$

Find the average cost of repair.

5-36 A company produces TV tubes with a length of life that is normally distributed with a standard deviation of 5 months. How large should the mean μ be in order that 90% of the tubes last for the guarantee period (18 months)?

5-37 A clinic plans to set up a mass testing program for diabetes using an inexpensive test of high reliability: Of all people with diabetes, 95% are correctly diagnosed positive (as having diabetes); of all people free of diabetes, 98% are correctly diagnosed negative (the remaining 2% being erroneously diagnosed positive).

The community served by the clinic has about 10,000 patients, and the undiagnosed diabetes rate runs at about 1%. The clinic director wants to know three things:

a. About how many patients will have diabetes and be missed (i.e., get a negative diagnosis)?

b. About how many patients will be diagnosed positive (and therefore require follow up)?

c. What proportion of the patients in part **b** will actually have diabetes?

d. The director cannot believe your answer to part **c.** Explain why it is so low, as simply as you can.

5-38 When is it true, or approximately true, that:

a. $E(X^2) = [E(X)]^2$?

b. $E(XY) = E(X)E(Y)$?

5-39 Because of unforeseen delays, the last two stages for manufacturing a complex plastic were of very uncertain duration. In fact, the times required for completion (X and Y, in hours) had the following probability distributions:

x	$p(x)$
1	.4
2	.2
3	0
4	.4

y	$p(y)$
0	.4
1	.2
2	0
3	0
4	.4

Two questions were of interest to management: (1) the total time T required; and (2) the cost C of running the first process, which was \$200 per hour, plus a fixed cost of \$300. In formulas,

$$T = X + Y$$

$$C = 200X + 300$$

Assuming X and Y are independent, tabulate the distribution of T, and of C, and then answer the following questions:

a. What are the medians $M(X)$, $M(Y)$, $M(T)$, and $M(C)$.

b. Is it true that $M(T) = M(X) + M(Y)$? That is,

$$M(X + Y) = M(X) + M(Y)?$$

c. Is it true that $M(C) = 200M(X) + 300$? That is,

$$M(200X + 300) = 200M(X) + 300?$$

5-40 Repeat Problem 5-39 for the means, instead of the medians.