

$$s_Y = \sqrt{0.445 (-5.55^2) + 0.555 (4.45^2)} = 4.970$$

Thus, the correlation is $r_{XY} = \frac{2.73}{0.917 (4.970)} = 0.60$

7.11. a. The basic setup for this problem is the same as it is for the previous problem. We already have the joint probabilities, so we can start by calculating the expected values of X and Y :

$$E(X) = \frac{(-2) + (-1) + 0 + 1 + 2}{5} = 0$$

$$E(Y) = 0.2 (0) + 0.4 (1) + 0.4 (2) = 1.2$$

Thus, we have the following table:

X	Y	$X - E(X)$	$Y - E(Y)$	$(X - E(X))(Y - E(Y))$
-2	2	-2	0.8	-1.6
-1	1	-1	-0.2	0.2
0	0	0	-1.2	0
1	1	1	-0.2	-0.2
2	2	2	0.8	1.6

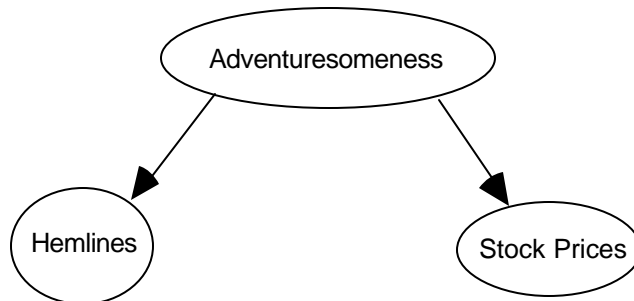
The covariance is the expected value of the numbers in the last column, each of which can occur with probability $1/5$. Calculating this expected value gives a covariance of zero. Likewise, the correlation equals zero.

b. $P(Y = 2) = 0.4$, but $P(Y = 2 | X = -2) = P(Y = 2 | X = 2) = 1.0$ and $P(Y = 2 | X = 0) = 0$

c. $P(X = -2) = 0.2$, but $P(X = -2 | Y = 2) = 0.5$ and $P(X = -2 | Y = 0) = 0$

d. Clearly, X and Y are dependent. In fact, $Y = |X|$. But it is not a linear relationship, and the covariance relationship does not capture this nonlinear relationship.

7.12. The influence diagram would show conditional independence between hemlines and stock prices, given adventuresomeness:



Thus (blatantly ignoring the clarity test), the probability statements would be $P(\text{Adventuresomeness})$, $P(\text{Hemlines} | \text{Adventuresomeness})$, and $P(\text{Stock prices} | \text{Adventuresomeness})$.

7.13. In many cases, it is not feasible to use a discrete model because of the large number of possible outcomes. The continuous model is a “convenient fiction” that allows us to construct a model and analyze it.

7.14.

	B	\bar{B}	
A	0.204	0.476	0.68
\bar{A}	0.006	0.314	0.32
	0.21	0.79	1

$$P(B \text{ and } A) = P(B | A) P(A) = 0.30 (0.68) = 0.204$$

$$P(B \text{ and } \bar{A}) = P(B | \bar{A}) P(\bar{A}) = 0.02 (0.32) = 0.006$$

$$P(\bar{B} | A) = \frac{P(\bar{B} \text{ and } A)}{P(A)} = \frac{0.476}{0.68} = 0.70$$

$$P(\bar{B} | \bar{A}) = \frac{P(\bar{B} \text{ and } \bar{A})}{P(\bar{A})} = \frac{0.314}{0.32} = 0.98$$

$$P(B) = P(B \text{ and } A) + P(B \text{ and } \bar{A}) = 0.204 + 0.006 = 0.21$$

$$P(\bar{B}) = 1 - P(B) = 1 - 0.21 = 0.79$$

$$P(A | B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{0.204}{0.21} = 0.970$$

$$P(\bar{A} | B) = 1 - P(A | B) = 1 - 0.970 = 0.030$$

$$P(A | \bar{B}) = \frac{P(A \text{ and } \bar{B})}{P(\bar{B})} = \frac{0.476}{0.79} = 0.603$$

$$P(\bar{A} | \bar{B}) = 1 - P(A | \bar{B}) = 1 - 0.603 = 0.397$$

7.15. $P(\text{offer}) = 0.50$

$$P(\text{good interview} | \text{offer}) = 0.95$$

$$P(\text{good interview} | \text{no offer}) = 0.75$$

$$P(\text{offer} | \text{good interview}) = P(\text{offer} | \text{good})$$

$$= \frac{P(\text{good} | \text{offer}) P(\text{offer})}{P(\text{good} | \text{offer}) P(\text{offer}) + P(\text{good} | \text{no offer}) P(\text{no offer})}$$

$$= \frac{0.95 (0.50)}{0.95 (0.50) + 0.75 (0.50)}$$

$$= 0.5588$$

$$\begin{aligned}
 \mathbf{7.16. a.} \quad E(X) &= 0.05(1) + 0.45(2) + 0.30(3) + 0.20(4) \\
 &= 0.05 + 0.90 + 0.90 + 0.80 \\
 &= 2.65
 \end{aligned}$$

$$\begin{aligned}
 \text{Var}(X) &= 0.05(1-2.65)^2 + 0.45(2-2.65)^2 + 0.30(3-2.65)^2 + 0.20(4-2.65)^2 \\
 &= 0.05(2.72) + 0.45(0.42) + 0.30(0.12) + 0.20(1.82) \\
 &= 0.728
 \end{aligned}$$

$$\sigma_X = \sqrt{0.728} = 0.853$$

$$\begin{aligned}
 \mathbf{b.} \quad E(X) &= 0.13(-20) + 0.58(0) + 0.29(100) \\
 &= -2.60 + 0 + 29 \\
 &= 26.40
 \end{aligned}$$

$$\begin{aligned}
 \text{Var}(X) &= 0.13(-20 - 26.40)^2 + 0.58(0 - 26.40)^2 + 0.29(100 - 26.40)^2 \\
 &= 0.13(2152.96) + 0.58(696.96) + 0.29(5416.96) \\
 &= 2255.04
 \end{aligned}$$

$$\sigma_X = \sqrt{2255.04} = 47.49$$

$$\mathbf{c.} \quad E(X) = 0.368(0) + 0.632(1) = 0.632$$

$$\begin{aligned}
 \text{Var}(X) &= 0.368(0 - 0.632)^2 + 0.632(1 - 0.632)^2 \\
 &= 0.368(0.632)^2 + 0.632(0.368)^2 \\
 &= 0.368(0.632)[0.632 + 0.368] \\
 &= 0.368(0.632) \\
 &= 0.233
 \end{aligned}$$

$$\sigma_X = \sqrt{0.233} = 0.482$$

$$\mathbf{7.17.} \quad E(X) = (1 - p)(0) + p(1) = p$$

$$\begin{aligned}
 \text{Var}(X) &= (1 - p)(0 - p)^2 + p(1 - p)^2 \\
 &= (1 - p)p^2 + p(1 - p)^2 \\
 &= (1 - p)p[p + (1 - p)] \\
 &= (1 - p)p
 \end{aligned}$$

7.18. It is true that $P(A | B) + P(\bar{A} | B) = 1$, because the condition (B) is the same in each probability. Thus, these two probabilities are complements.

However, the question is about $P(A | B) + P(A | \bar{B})$, which can equal anything between 0 and 2. There is no requirement that these two probabilities add up, because the conditions (B and \bar{B}) are different.

$$\begin{aligned}
 \mathbf{7.19. a.} \quad E(\text{Revenue from A}) &= \$3.50 E(\text{Unit sales}) \\
 &= \$3.50(2000) \\
 &= \$7000
 \end{aligned}$$

$$\begin{aligned}
 \text{Var}(\text{Revenue from A}) &= 3.50^2 \text{Var}(\text{Unit sales}) \\
 &= 3.50^2(1000) \\
 &= 12,250 \text{ "dollars squared"}
 \end{aligned}$$

b.

$$\begin{aligned}
 E(\text{Total revenue}) &= \$3.50 (2000) + \$2.00 (10,000) + \$1.87 (8500) \\
 &= \$42,895
 \end{aligned}$$

$$\begin{aligned}
 \text{Var}(\text{Total revenue}) &= 3.50^2 (1000) + 2.00^2 (6400) + 1.87^2 (1150) \\
 &= 41,871 \text{ "dollars squared"}
 \end{aligned}$$

7.20. Let X_1 = random number of breakdowns for Computer 1, and X_2 = random number of breakdowns for Computer 2.

$$\text{Cost} = \$200 (X_1) + \$165 (X_2)$$

$$E(\text{Cost}) = \$200 E(X_1) + \$165 E(X_2) = \$200 (5) + \$165 (3.6) = \$1594$$

If X_1 and X_2 are independent, then

$$\begin{aligned}
 \text{Var}(\text{Cost}) &= 200^2 \text{Var}(X_1) + 165^2 \text{Var}(X_2) = 200^2 (6) + 165^2 (7) \\
 &= 430,575 \text{ "dollars squared"}
 \end{aligned}$$

$$\sigma_{\text{Cost}} = \sqrt{430,575 \text{ "dollars squared"}} = \$656.18$$

The assumption made for the variance computation is that the computers break down independently of one another. Given that they are in separate buildings and operated separately, this seems like a reasonable assumption.

7.21. The possible values for revenue are 100 (\$3) = \$300 and 300 (\$2) = \$600, each with probability 0.5. thus, the expected revenue is 0.5 (\$300) + 0.5 (\$600) = \$450. The manager's mistake is in thinking that the expected value of the product is equal to the product of the expected values, which is true only if the two variables are independent, which is not true in this case.

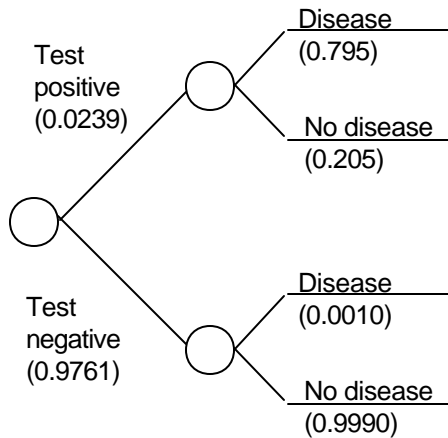
7.22. Notation: "Pos" = positive
 "Neg" = negative
 "D" = disease
 " \bar{D} " = no disease

$$\begin{aligned}
 P(\text{Pos}) &= P(\text{Pos} | D) P(D) + P(\text{Pos} | \bar{D}) P(\bar{D}) \\
 &= 0.95 (0.02) + 0.005 (0.98) \\
 &= 0.0239
 \end{aligned}$$

$$P(\text{Neg}) = 1 - P(\text{Pos}) = 1 - 0.0239 = 0.9761$$

$$P(D | \text{Pos}) = \frac{P(\text{Pos} | D) P(D)}{P(\text{Pos})} = \frac{0.95 (0.02)}{0.0239} = 0.795$$

$$P(D | \text{Neg}) = \frac{P(\text{Neg} | D) P(D)}{P(\text{Neg})} = \frac{0.05 (0.02)}{0.9761} = 0.0010$$



Probability Table

	Pos	Neg	
D	0.0190	0.0010	0.02
\bar{D}	0.0049	0.9751	0.98
	0.0239	0.9761	1

7.23. Test results and field results are conditionally independent given the level of carcinogenic risk. Alternatively, given the level of carcinogenic risk, knowing the test results will not help specify the field results.

$$\begin{aligned}
 7.24. P(\text{TR+ and FR+} \mid \text{CP high}) &= P(\text{TR+} \mid \text{FR+ and CP high}) P(\text{FR+} \mid \text{CP high}) \\
 &= P(\text{TR+} \mid \text{CP high}) P(\text{FR+} \mid \text{CP high})
 \end{aligned}$$

The second equality follows because FR and TR are conditionally independent given CP. In other words, we just multiply the probabilities together. This is true for all four of the probabilities required:

$$P(\text{TR+ and FR+} \mid \text{CP high}) = 0.82 (0.95) = 0.779$$

$$P(\text{TR+ and FR-} \mid \text{CP high}) = 0.82 (0.05) = 0.041$$

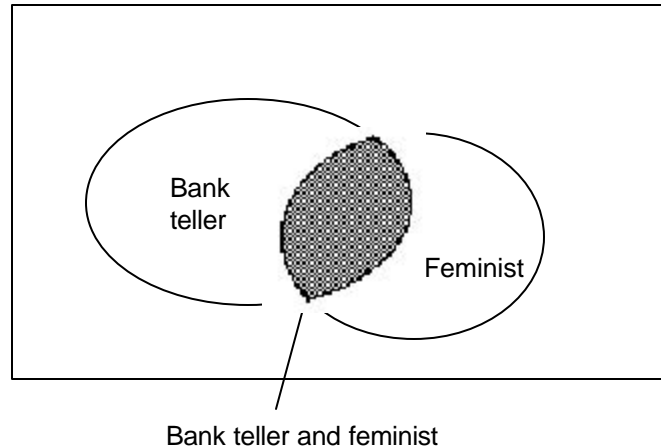
$$P(\text{TR- and FR-} \mid \text{CP low}) = 0.79 (0.83) = 0.6557$$

$$P(\text{TR- and FR+} \mid \text{CP low}) = 0.79 (0.17) = 0.1343$$

7.25. Students' answers will vary considerably here, depending on their opinions. However, most will rate h as more likely than f. Tversky and Kahneman (1982) (see reference in text) found that as many as 85% of experimental subjects ranked the statements in this way, which is inconsistent with the idea of joint probability (see the next question). Moreover, this phenomenon was found to occur consistently regardless of the degree of statistical sophistication of the subject.

7.26. a. The students' explanations will vary, but many of them argue on the basis of the degree to which Linda's description is consistent with the possible classifications. Her description makes her sound not much like a bank teller and a lot like an active feminist. Thus, statement h (bank teller and feminist) is more consistent with the description than f (bank teller). Tversky and Kahneman claim that the conjunction effect observed in the responses to problem 7.25 stem from the representativeness heuristic. This heuristic is discussed in Chapter 8 of *Making Hard Decisions with DecisionTools*.

b.



$P(\text{Bank teller and feminist}) = P(\text{Feminist} \mid \text{Bank teller}) P(\text{Bank teller})$. Since $P(\text{Feminist} \mid \text{Bank teller})$ must be less than or equal to one, $P(\text{Bank teller and feminist})$ must be less than or equal to $P(\text{Bank teller})$. The area for the intersection of the two outcomes cannot be larger than the area for Bank teller.

c. The friend is interpreting h as a conditional outcome instead of a joint outcome. Statement h clearly is a joint outcome, because both outcomes (bank teller and feminist) occur.

7.27. To start, we need some labels. Let us say that we have chosen Door A. The host has opened Door B, revealing the goat, and Door C remains closed. The question is whether we should switch to C. The decision rule is simple: switch if the probability of the car being behind Door C is greater than the probability that it is behind A. Let “Car C” denote the outcome that the car is behind C, and likewise with the goats and the other doors. We want to calculate $P(\text{Car C} \mid \text{Goat B})$. Use Bayes theorem:

$$P(\text{Car C} \mid \text{Goat B}) = \frac{P(\text{Goat B} \mid \text{Car C}) P(\text{Car C})}{P(\text{Goat B} \mid \text{Car A}) P(\text{Car A}) + P(\text{Goat B} \mid \text{Car B}) P(\text{Car B}) + P(\text{Goat B} \mid \text{Car C}) P(\text{Car C})}$$

The prior probabilities $P(\text{Car A})$, $P(\text{Car B})$, and $P(\text{Car C})$ are all equal to $1/3$. For the conditional probabilities, the key is to think about the host’s behavior. The host would never open a door to reveal the car. Thus, $P(\text{Goat B} \mid \text{Car C}) = 1$ and $P(\text{Goat B} \mid \text{Car B}) = 0$. Finally, what if the car is behind A? What is $P(\text{Goat B} \mid \text{Car A})$? In this case, we assume that the host would randomly choose B or C, so $P(\text{Goat B} \mid \text{Car A}) = 0.5$. Plug these numbers into the formula to get:

$$P(\text{Car C} \mid \text{Goat B}) = \frac{1 (1/3)}{0.5 (1/3) + 0 (1/3) + 1 (1/3)} = \frac{2}{3}$$

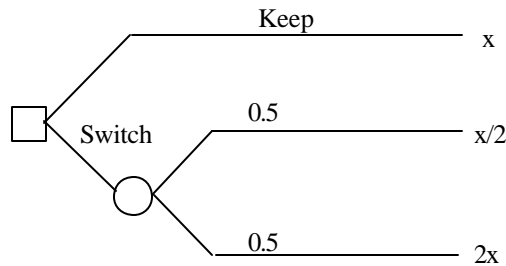
Thus, you should always switch when the host reveals the goat!

Here’s another way to think about it: You had a one-third chance of getting the correct door in the first place. Thus, there is a two-thirds chance that the goat is behind B or C. By showing the goat behind B, the host has effectively shifted the entire two-thirds probability over to Door C.

Still another way to think about the problem: If you played this game over and over, one-third of the time the car would be behind A, and two-thirds of the time it would be behind one of the other doors. Thus, two-thirds of the time, the host shows you which door the car is *not* behind. If you always switch to the door that the host did not open, you will find the car $2/3$ of the time. The other $1/3$ the car is behind the door you chose in the first place.

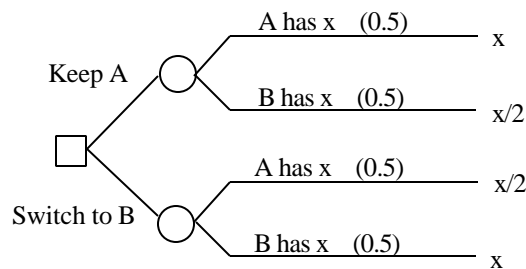
This question was asked of Marilyn Vos Savant, the person with the highest recorded I.Q. Her published answer was correct, but it created quite a stir because many people (including PhDs) did not understand how to solve the problem.

7.28. The host is proposing a decision tree that looks like this:



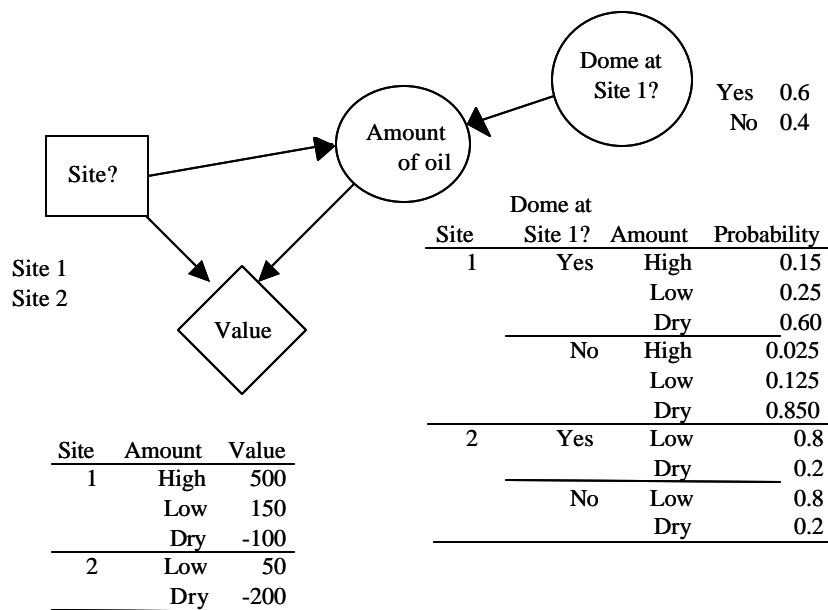
But this is not correct. Suppose that x is equal to \$100. Then the host is saying that if you swap, you have equally likely chances at an envelope with \$200 and an envelope with \$50. But that's not the case! (If it were true, you would definitely want to switch.)

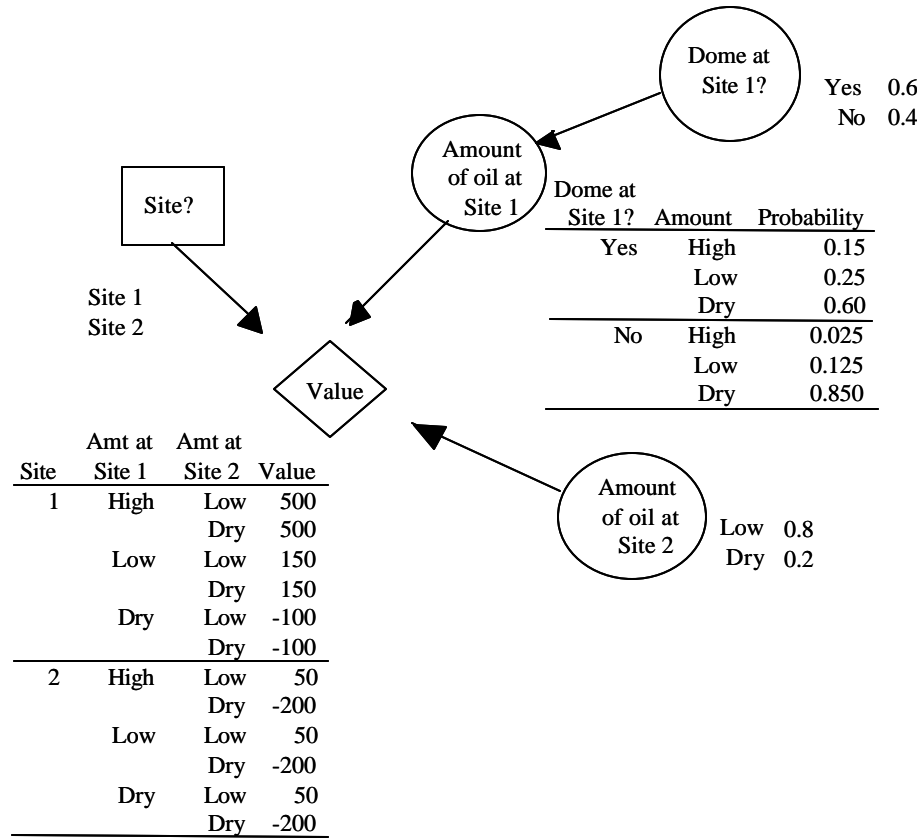
Labeling the two envelopes A and B, the contestant correctly understands that the decision tree is as follows:



The two decision branches are equivalent from the point of view of the decision maker.

7.29. Here are two possible influence diagrams:

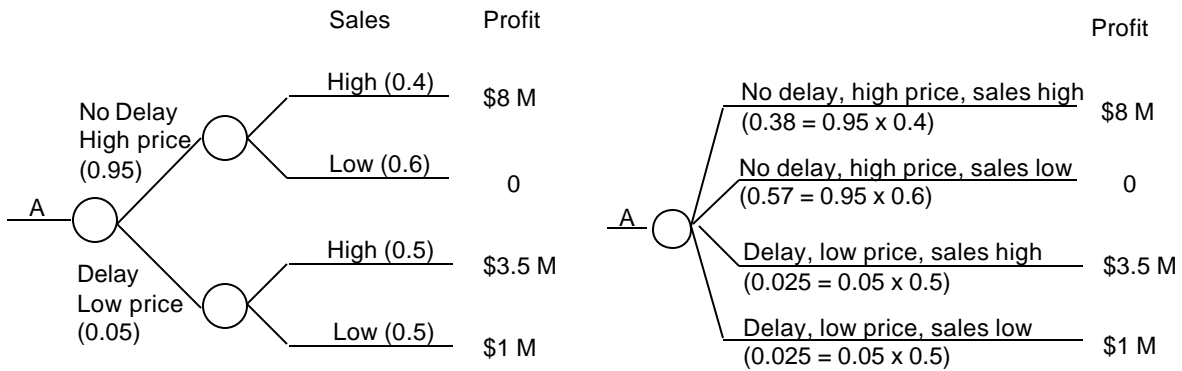




The two diagrams are equivalent. However, while the first is somewhat more compact, the second may represent the uncertainty in the problem better. Also, with the second it would be more straightforward to calculate the expected value of information for the uncertain events.

The second version is modeled in the Excel file "Problem 7.29.xls". The optimal solution is to choose the first site, and the expected value is \$10,000.

7.30.



From the collapsed tree, we can calculate

$$E(\text{Profit}) = 0.38 (8) + 0.57 (0) + 0.025 (3.5) + 0.025 (1) = \$3.1525 \text{ M}$$

$$\begin{aligned} \text{Var}(\text{Profit}) &= 0.38 (8 - 3.1525)^2 + 0.57 (0 - 3.1525)^2 \\ &\quad + 0.025 (3.5 - 3.1525)^2 + 0.025 (1 - 3.1525)^2 \\ &= 14.7130 \text{ million-dollars-squared} \end{aligned}$$

$$\sigma_{\text{Profit}} = \sqrt{14.7130} = \$3.8358 \text{ M}$$

7.31.

$$E(\text{Payoff}) = \$4.56 \text{ billion (calculated previously)}$$

$$\begin{aligned} \text{Var}(\text{Payoff}) &= 0.2 (10.3 - 4.56)^2 + 0.5 (5 - 4.56)^2 + 0.3 (0 - 4.56)^2 \\ &= 12.9244 \text{ billion-dollars-squared} \end{aligned}$$

$$\sigma_{\text{Payoff}} = \sqrt{12.9244} = \$3.5951 \text{ billion}$$

7.32. Let “+” indicate positive results, and “-” indicate negative results.

$$\begin{aligned} P(+) &= P(+ | \text{Dome}) P(\text{Dome}) + P(+ | \text{No dome}) P(\text{No Dome}) \\ &= 0.99 (0.6) + 0.15 (0.4) \\ &= 0.654 \end{aligned}$$

$$\begin{aligned} P(\text{Dome} | +) &= \frac{P(+ | \text{Dome}) P(\text{Dome})}{P(+ | \text{Dome}) P(\text{Dome}) + P(+ | \text{No dome}) P(\text{No Dome})} \\ &= \frac{0.99 (0.6)}{0.99 (0.6) + 0.15 (0.4)} \\ &= 0.908 \end{aligned}$$

$$P(\text{No Dome} | +) = 1 - 0.908 = 0.092$$

We can now calculate the EMV for Site 1, given test results are positive:

$$\begin{aligned} \text{EMV}(\text{Site 1} | +) &= (\text{EMV} | \text{Dome}) P(\text{Dome} | +) + (\text{EMV} | \text{No dome}) P(\text{No dome} | +) \\ &= (\$52.50 \text{ K}) 0.908 + (-\$53.75 \text{ K}) 0.092 \\ &= \$42.725 \text{ K} \end{aligned}$$

[EMV|Dome and EMV|No dome have been calculated and appear in Figure 7.15.]

EMV(Site 1 | +) is greater than EMV(Site 2 | +). If the test gives a positive result, choose Site 1.

If the results are negative:

$$\begin{aligned} P(-) &= 1 - P(+) \\ &= 1 - 0.654 \\ &= 0.346 \end{aligned}$$

$$\begin{aligned} P(\text{Dome} | -) &= \frac{P(- | \text{Dome}) P(\text{Dome})}{P(- | \text{Dome}) P(\text{Dome}) + P(- | \text{No dome}) P(\text{No Dome})} \\ &= \frac{0.01 (0.6)}{0.01 (0.6) + 0.85 (0.4)} \end{aligned}$$

$$= 0.017$$

$$P(\text{No Dome} | -) = 1 - 0.017 = 0.983$$

We can now calculate the EMV for Site 1, given test results are negative:

$$\begin{aligned} \text{EMV}(\text{Site 1} | -) &= (\text{EMV} | \text{Dome}) P(\text{Dome} | -) + (\text{EMV} | \text{No dome}) P(\text{No dome} | -) \\ &= (\$52.50 \text{ K}) 0.017 + (-\$53.75 \text{ K}) 0.983 \\ &= -\$51.944 \text{ K} \end{aligned}$$

EMV(Site 1 | -) is less than the EMV(Site 2 | -). If the test gives a negative result, choose Site 2.

$$7.33. \quad P(+ \text{ and Dome}) = P(+ | \text{Dome}) P(\text{Dome}) = 0.99 (0.60) = 0.594.$$

$$P(+ \text{ and Dome and Dry}) = P(\text{Dry} | + \text{ and Dome}) P(+ \text{ and Dome})$$

But $P(\text{Dry} | + \text{ and Dome}) = P(\text{Dry} | \text{Dome}) = 0.60$. That is, the presence or absence of the dome is what matters, not the test results themselves. Therefore:

$$P(+ \text{ and Dome and Dry}) = 0.60 (0.594) = 0.356$$

Finally,

$$P(\text{Dome} | + \text{ and Dry}) = \frac{P(\text{Dome and } + \text{ and Dry})}{P(+ \text{ and Dry})}$$

But

$$P(+ \text{ and Dry}) = P(+ \text{ and Dry} | \text{Dome}) P(\text{Dome}) + P(+ \text{ and Dry} | \text{No dome}) P(\text{No dome})$$

and

$$\begin{aligned} P(+ \text{ and Dry} | \text{Dome}) &= P(\text{Dry} | + \text{ and Dome}) P(+ | \text{Dome}) \\ &= P(\text{Dry} | \text{Dome}) P(+ | \text{Dome}) \\ &= 0.6 (0.99) \end{aligned}$$

$$\begin{aligned} P(+ \text{ and Dry} | \text{No dome}) &= P(\text{Dry} | + \text{ and No dome}) P(+ | \text{No dome}) \\ &= P(\text{Dry} | \text{No dome}) P(+ | \text{No dome}) \\ &= 0.85 (0.15) \end{aligned}$$

Now we can substitute back in:

$$P(+ \text{ and Dry}) = 0.6 (0.99) (0.6) + 0.85 (0.15) (0.4) = 0.407$$

and

$$\begin{aligned} P(\text{Dome} | + \text{ and Dry}) &= \frac{0.356}{0.6 (0.99) (0.6) + 0.85 (0.15) (0.4)} \\ &= 0.874. \end{aligned}$$

$$7.34. \quad \text{EMV}(\text{Site 1}) = p(52.50) + (1 - p)(-53.75)$$

Set this equal to EMV(Site 2) = 0, and solve for p:

$$p(52.50) + (1 - p)(-53.75) = 0$$

$$52.50p + 53.75(1-p) = 53.75$$

$$p = \frac{53.75 - 53.75}{52.50 - 53.75}$$

$$p = 0.5059.$$

If $0.55 < P(\text{Dome}) < 0.65$, then the optimal choice for the entire region is to drill at Site #1.

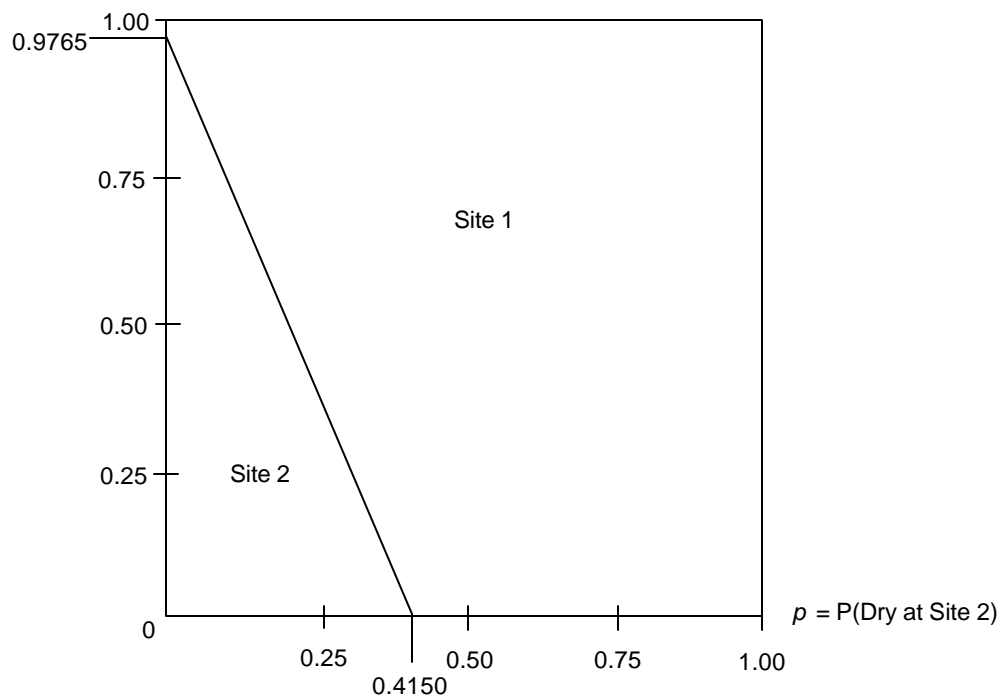
7.35. Choose Site 1 if

$$EMV(\text{Site 1}) > EMV(\text{Site 2})$$

$$q(52.50) + (1 - q)(-53.75) > p(-200) + (1-p)(50)$$

$$q > -2.3529p + 0.9765$$

$$q = P(\text{Dome at Site 1})$$



7.36. $P(\text{FR pos}) = P(\text{FR pos} | \text{CP High}) P(\text{CP High}) + P(\text{FR pos} | \text{CP Low}) P(\text{CP Low})$
 $= 0.95(0.27) + 0.17(0.73)$
 $= 0.3806$

$$P(\text{FR} + | \text{TR} +) = P(\text{FR} + | \text{CP High and TR} +) P(\text{CP High} | \text{TR} +)$$

$$+ P(\text{FR} + | \text{CP Low and TR} +) P(\text{CP Low} | \text{TR} +)$$

But FR and TR are conditionally independent given CP, so

$$P(\text{FR} + | \text{CP High and TR} +) = P(\text{FR} + | \text{CP High}) = 0.95$$

and

$$P(\text{FR} + | \text{CP Low and TR} +) = P(\text{FR} + | \text{CP Low}) = 0.17$$

We can calculate $P(\text{CP High} | \text{TR} +)$ using Bayes' theorem:

$$\begin{aligned} P(\text{CP High} | \text{TR} +) &= \frac{P(\text{TR} + | \text{CP High}) P(\text{CP High})}{P(\text{TR} + | \text{CP High}) P(\text{CP High}) + P(\text{TR} + | \text{CP Low}) P(\text{CP Low})} \\ &= \frac{0.82 (0.27)}{0.82 (0.27) + 0.21 (0.73)} \\ &= 0.5909. \end{aligned}$$

Therefore

$$P(\text{CP Low} | \text{TR} +) = 1 - P(\text{CP High} | \text{TR} +) = 1 - 0.5909 = 0.4091.$$

Substitute back to obtain

$$P(\text{FR} + | \text{TR} +) = 0.95 (0.5909) + 0.17 (0.4091) = 0.6309.$$

It is important to note that $P(\text{FR} + | \text{TR} +) \neq P(\text{FR} +) = 0.3806$. Thus, the two are not fully independent, even though they are conditionally independent given CP. Another way to say it is that conditional independence does not necessarily imply regular (marginal) independence.