

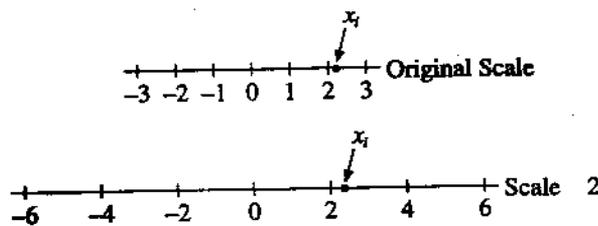
24
27

Changing the scale by a factor c will multiply each data value x_i by c , changing it to cx_i . Again the same individual's value will be at the median and the same individual's value will be at the mode, but these values will be multiplied by c . The geometric mean will be multiplied by c also, as can easily be shown:

$$\begin{aligned}\text{Geometric mean} &= [(cx_1)(cx_2)\cdots(cx_n)]^{1/n} \\ &= (c^n x_1 \cdot x_2 \cdots x_n)^{1/n} \\ &= c(x_1 \cdot x_2 \cdots x_n)^{1/n} \\ &= c \times \text{old geometric mean}\end{aligned}$$

The range will also be multiplied by c .

For example, if $c = 2$ we have:



2.8

We first must compute the midpoint of each group interval in order to compute the grouped mean and variance. These are given as follows:

Degree of astigmatism (diopters)	Midpoint	Frequency
0.0 or less than 0.2	0.075	458
0.2 to 0.3	0.25	268
0.4 to 0.5	0.45	151
0.6 to 1.0	0.80	79
1.1 to 2.0	1.55	44
2.1 to 3.0	2.55	19
3.1 to 4.0	3.55	9
4.1 to 5.0	4.55	3
5.1 to 6.0	5.55	2
		1033

The arithmetic mean is then given by

$$\bar{x} = \frac{458 \times 0.075 + \cdots + 2 \times 5.55}{1033} = 0.39 \text{ diopters}$$

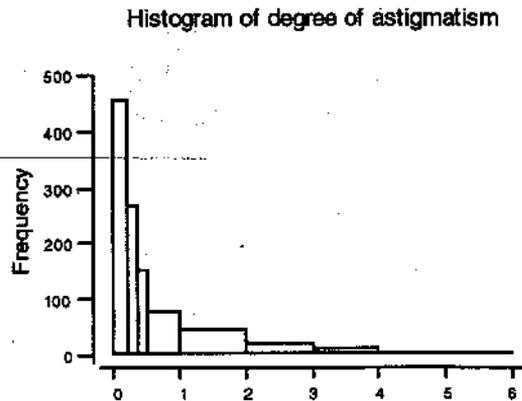
2.9

The variance is given by

$$s^2 = \frac{458(0.075 - 0.39)^2 + \cdots + 2(5.55 - 0.39)^2}{1032} = \frac{407.404}{1032} = 0.395$$

Thus, $s = \sqrt{0.395} = 0.63$ diopters

2.10 The histogram is given as follows.



2.11 We have that

$$\bar{x} = \frac{\sum x_i}{24} = \frac{469}{24} = 19.54 \text{ mg/dL}$$

2.12 We have that

$$s^2 = \frac{\sum_{i=1}^{24} (x_i - \bar{x})^2}{23} = \frac{(49 - 19.54)^2 + \dots + (12 - 19.54)^2}{23} = \frac{6495.96}{23} = 282.43$$

$$s = \sqrt{282.43} = 16.81 \text{ mg/dL}$$

2.13 We provide two rows for each stem corresponding to leaves 5-9 and 0-4 respectively. We have

Stem-and-leaf plot	Cumulative frequency
+4 98	24
+4 1	22
+3 65	21
+3 21	19
+2 78	17
+2 13	15
+1 9699	13
+1 332	9
+0 88	6
+0 2	4
-0	
-0 8	3
-1 03	2

2.14 We wish to compute the average of the $(24/2)$ th and $(24/2 + 1)$ th largest values = average of the 12th and 13th largest points. We note from the stem-and-leaf plot that the 13th largest point counting from the bottom is the largest value in the upper +1 row = 19. The 12th largest point = the next largest value in this row = 19. Thus, the median = $\frac{19+19}{2} = 19$ mg/dL.

- 3.47 Let $C = \{\text{both one male and one female sibling are affected}\}$. The sex of the siblings is only relevant for sex-linked disease. Thus, from Problems 3.31, 3.36, and 3.39,

$$Pr(C|DOM) = \frac{1}{4} \quad Pr(C|AR) = \frac{1}{16} \quad Pr(C|SL) = 0$$

Thus,

$$Pr(DOM|C) = \frac{Pr(C|DOM)}{Pr(C|DOM) + Pr(C|AR) + Pr(C|SL)} = \frac{\frac{1}{4}}{\frac{1}{4} + \frac{1}{16} + \frac{1}{16}} = \frac{\frac{1}{4}}{\frac{5}{8}} = \frac{2}{5}$$

$$Pr(AR|C) = \frac{Pr(C|AR)}{Pr(C|DOM) + Pr(C|AR) + Pr(C|SL)} = \frac{\frac{1}{16}}{\frac{5}{8}} = \frac{1}{5}$$

$$Pr(SL|C) = 0$$

- 3.48 Let $D = \{\text{male sibling affected, female sibling not affected}\}$. $Pr(D|DOM) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

$$Pr(D|AR) = \frac{1}{4} \times \frac{3}{4} = \frac{3}{16} \quad Pr(D|SL) = \frac{1}{2} \times 1 = \frac{1}{2}$$

Notice that the event D is not the same as the event that exactly one sibling is affected, since we are specifying which of the two siblings is affected. We have

$$Pr(DOM|D) = \frac{Pr(D|DOM)}{Pr(D|DOM) + Pr(D|AR) + Pr(D|SL)} = \frac{\frac{1}{4}}{\frac{1}{4} + \frac{3}{16} + \frac{1}{2}} = \frac{\frac{1}{4}}{\frac{15}{8}} = \frac{2}{15}$$

$$Pr(AR|D) = \frac{Pr(D|AR)}{Pr(D|DOM) + Pr(D|AR) + Pr(D|SL)} = \frac{\frac{3}{16}}{\frac{1}{4} + \frac{3}{16} + \frac{1}{2}} = \frac{\frac{3}{16}}{\frac{15}{8}} = \frac{1}{5}$$

$$Pr(SL|D) = \frac{Pr(D|SL)}{Pr(D|DOM) + Pr(D|AR) + Pr(D|SL)} = \frac{\frac{1}{2}}{\frac{15}{8}} = \frac{4}{15}$$

Thus, in this situation the sex-linked mode of inheritance is the most likely.

- 3.49 $Pr(LOW) = Pr(LOW \cap < 20 \text{ weeks}) + Pr(LOW \cap 20 - 27 \text{ weeks})$
 $+ Pr(LOW \cap 28 - 36 \text{ weeks})$
 $+ Pr(LOW \cap > 36 \text{ weeks}) = Pr(LOW|< 20 \text{ weeks})Pr(< 20 \text{ weeks})$
 $+ Pr(LOW|20 - 27 \text{ weeks})Pr(20 - 27 \text{ weeks})$
 $+ Pr(LOW|28 - 36 \text{ weeks})Pr(28 - 36 \text{ weeks})$
 $+ Pr(LOW|> 36 \text{ weeks})Pr(> 36 \text{ weeks})$
 $= .540(.0004) + .813(.0059) + .379(.0855) + .035(.9082) = .069$

- 3.50 The dependence of the events $\{\leq 27 \text{ weeks}\}$ and $\{LOW\}$ can be shown by establishing that

$$Pr(\leq 27 \text{ weeks} \cap LOW) \neq Pr(\leq 27 \text{ weeks}) \times Pr(LOW)$$

We have

$$\begin{aligned} Pr\{\leq 27 \text{ weeks} \cap \text{LOW}\} &= Pr\{\text{LOW} \cap 20 \text{ weeks}\} + Pr\{\text{LOW} \cap 20-27 \text{ weeks}\} \\ &= .540(.0004) + .813(.0059) \\ &= .0050 \text{ (from Problem 3.49)} \end{aligned}$$

Similarly, from Problem 3.49

$$\begin{aligned} Pr(\text{LOW}) &= .069 \\ Pr(\leq 27 \text{ weeks}) &= .0004 + .0059 = .0063 \end{aligned}$$

$$\text{Thus, } Pr(\text{LOW}) \times Pr(\leq 27 \text{ weeks}) = .00044 \approx \frac{1}{10} Pr(\leq 27 \text{ weeks} \cap \text{LOW}) = .0050$$

Thus, the two events are dependent.

3.51 $Pr(\leq 36 \text{ weeks} | \text{LOW})$ must be computed. Bayes' theorem is used as follows:

$$\begin{aligned} Pr(\leq 36 \text{ weeks} | \text{LOW}) &= \frac{Pr(\leq 36 \text{ weeks} \cap \text{LOW})}{Pr(\text{LOW})} \\ &= \frac{Pr(\text{LOW} \cap < 20 \text{ weeks}) + Pr(\text{LOW} \cap 20-27 \text{ weeks})}{Pr(\text{LOW})} \\ &\quad + \frac{Pr(\text{LOW} \cap 28-36 \text{ weeks})}{Pr(\text{LOW})} \\ &= \frac{.540(.0004) + .813(.0059) + .379(.0855)}{.069} = \frac{.0374}{.069} = .541 \end{aligned}$$

3.52 Let $A = \{\text{pneumonia and/or bronchitis in the first year of life}\}$

$$\begin{aligned} B_2 &= \{\text{both parents smoke}\} \\ B_1 &= \{\text{one parent smokes}\} \\ B_0 &= \{\text{neither parent smokes}\} \end{aligned}$$

We are given that $Pr(A|B_2) = .176$, $Pr(A|B_1) = .114$, $Pr(A|B_0) = .078$,

$$\begin{aligned} Pr(B_2) &= .40, \quad Pr(B_1) = .25, \quad Pr(B_0) = .35. \text{ Therefore, from the total probability rule,} \\ Pr(A) &= Pr(A|B_2)Pr(B_2) + Pr(A|B_1)Pr(B_1) + Pr(A|B_0)Pr(B_0) \\ &= .176 \times .40 + .114 \times .25 + .078 \times .35 = .126 \end{aligned}$$

3.53 In this group of families, $Pr(B_2) = .10$, $Pr(B_1) = .30$, $Pr(B_0) = .60$. Therefore, from the total probability rule, $Pr(A) = .176 \times .10 + .114 \times .30 + .078 \times .60 = .099$

3.54 The probability of pneumonia and/or bronchitis in the first year of life was .176 without counseling and .099 with the counseling (see Problem 3.53). Therefore, the percentage of cases prevented is $[1 - (.099/.176)] \times 100\% = 44.0\%$.

3.55 $Pr(\text{mother current smoker} \cap \text{father current smoker})$
 $= Pr(\text{mother current smoker}) \times Pr(\text{father current smoker}) = .4 \times .5 = .20$

3.56 $Pr(\text{father current smoker} | \text{mother not current smoker}) = Pr(\text{father current smoker}) = .5$
 This is a conditional probability compared with the joint probability in Problem 3.55.

3.57 $Pr(\text{father current smoker} \cap \text{mother not current smoker}) = Pr(\text{father current smoker})$
 $\times Pr(\text{mother not current smoker} | \text{father current smoker}) = .5 \times (1 - .6) = .20$

3.58 The smoking habits of the parents are *not* independent random variables because $Pr(\text{mother current smoker} | \text{father current smoker}) = .6 \neq Pr(\text{mother current smoker} | \text{father not current smoker}) = .2$

3.59 Let $A = \{\text{child has asthma}\}$, $M = \{\text{mother current smoker}\}$,

$\bar{M} = \{\text{mother not current smoker}\}$, $F = \{\text{father current smoker}\}$,
 $\bar{F} = \{\text{father not current smoker}\}$. We want $Pr(A)$. We have that

$$Pr(A) = Pr(A | M \cap F)Pr(M \cap F) + Pr(A | M \cap \bar{F})Pr(M \cap \bar{F}) \\ + Pr(A | \bar{M} \cap F)Pr(\bar{M} \cap F) + Pr(A | \bar{M} \cap \bar{F})Pr(\bar{M} \cap \bar{F}).$$

We are given that

$$Pr(A | M \cap F) = .15, Pr(A | M \cap \bar{F}) = .13, \\ Pr(A | \bar{M} \cap F) = .05, Pr(A | \bar{M} \cap \bar{F}) = .04.$$

Also,

$$Pr(M \cap F) = Pr(F)Pr(M | F) = .5 \times .6 = .30 \\ Pr(M \cap \bar{F}) = Pr(\bar{F})Pr(M | \bar{F}) = .5 \times .2 = .10 \\ Pr(\bar{M} \cap F) = Pr(F)Pr(\bar{M} | F) = .5 \times .4 = .20 \\ Pr(\bar{M} \cap \bar{F}) = Pr(\bar{F})Pr(\bar{M} | \bar{F}) = .5 \times .8 = .40$$

Therefore,

$$Pr(A) = .15 \times .30 + .13 \times .10 + .05 \times .20 + .04 \times .40 = .084$$

3.60 We want to compute $Pr(F | A)$. We have from the definition of conditional probability that

$$Pr(F | A) = \frac{Pr(F \cap A)}{Pr(A)} = \frac{Pr(F \cap A)}{.084}$$

Furthermore,

$$Pr(F \cap A) = Pr(M \cap F \cap A) + Pr(\bar{M} \cap F \cap A) \\ = Pr(A | M \cap F)Pr(M \cap F) + Pr(A | \bar{M} \cap F)Pr(\bar{M} \cap F)$$

Referring to problem 3.59, we note that

$$Pr(F \cap A) = .15 \times .30 + .05 \times .20 = .055$$

$$\text{Thus, } Pr(F | A) = \frac{.055}{.084} = .655$$

3.61 We want to compute $Pr(M|A)$. We have that $Pr(M|A) = \frac{Pr(M \cap A)}{Pr(A)}$ where

$$\begin{aligned} Pr(M \cap A) &= Pr(M \cap F \cap A) + Pr(M \cap \bar{F} \cap A) \\ &= Pr(A|M \cap F)Pr(M \cap F) + Pr(A|M \cap \bar{F})Pr(M \cap \bar{F}) \\ &= 15 \times 30 + 13 \times 10 = 058 \end{aligned}$$

$$\text{Thus, } Pr(M|A) = \frac{.058}{.084} = .690$$

3.62 We want to compute $Pr(F|\bar{A})$. We have that

$$Pr(F|\bar{A}) = \frac{Pr(F \cap \bar{A})}{Pr(\bar{A})} = \frac{Pr(F \cap \bar{A})}{.916}$$

where

$$\begin{aligned} Pr(F \cap \bar{A}) &= Pr(M \cap F \cap \bar{A}) + Pr(\bar{M} \cap F \cap \bar{A}) \\ &= Pr(\bar{A}|M \cap F)Pr(M \cap F) + Pr(\bar{A}|\bar{M} \cap F)Pr(\bar{M} \cap F) \\ &= (1-.15) \times 30 + (1-.05) \times 20 = 445 \end{aligned}$$

$$\text{Thus, } Pr(F|\bar{A}) = \frac{.445}{.916} = .486$$

3.63 We want to compute $Pr(M|\bar{A})$. We have that

$$Pr(M|\bar{A}) = \frac{Pr(M \cap \bar{A})}{Pr(\bar{A})} = \frac{Pr(M \cap \bar{A})}{.916}$$

where

$$\begin{aligned} Pr(M \cap \bar{A}) &= Pr(M \cap F \cap \bar{A}) + Pr(M \cap \bar{F} \cap \bar{A}) \\ &= Pr(\bar{A}|M \cap F)Pr(M \cap F) + Pr(\bar{A}|M \cap \bar{F})Pr(M \cap \bar{F}) \\ &= (1-.15) \times 30 + (1-.13) \times 10 = 342 \end{aligned}$$

$$\text{Thus, } Pr(M|\bar{A}) = \frac{.342}{.916} = .373$$

3.64 We found in problem 3.60 that $Pr(F|A) = .655$ and in problem 3.62 that $Pr(F|\bar{A}) = .486$. Since $Pr(F|A) = .655 \neq Pr(F|\bar{A}) = .486$, the father's smoking status and the child's asthma status are *not* independent.

3.65 We found in problem 3.61 that $Pr(M|A) = .690$ and in problem 3.63 that $Pr(M|\bar{A}) = .373$. Since $Pr(M|A) = .690 \neq Pr(M|\bar{A}) = .373$, the mother's smoking status and the child's asthma status are *not* independent.

3.66 Let \bar{A} = {did not quit smoking over the 14-year period, 1962–1975}. We have

$$\begin{aligned} Pr(\bar{A}) &= Pr(\text{did not quit from } 1/1/62 \text{ to } 12/31/62) \\ &\quad \times Pr(\text{did not quit from } 1/1/63 \text{ to } 12/31/63 | \text{smoker on } 1/1/63) \\ &\quad \times \dots \times Pr(\text{did not quit from } 1/1/75 \text{ to } 12/31/75 | \text{smoker on } 1/1/75) \\ &= (1-.031)^5 \times (1-.071)^4 \times (1-.047)^5 = .8543 \times .7448 \times .7861 = .500 \end{aligned}$$

It follows that $Pr(A) = Pr(\text{quit smoking over the 14-year period}) = 1 - Pr(\bar{A}) = .500$.

3.67 We use the same methodology as in problem 3.66 except that we use the quitting rates for heavy smokers rather than light smokers. We have

$$Pr(\bar{A}) = (1-.020)^5 \times (1-.050)^4 \times (1-.041)^5 = .9039 \times .8145 \times .8111 = .597$$

Thus, $Pr(\text{quit smoking over the 14-year period}) = 1 - Pr(\bar{A}) = .403$.

3.68 Let A = {test +}, B_1 = {no cigarettes}, B_2 = {1–4 cigarettes per week},

B_3 = {5–14 cigarettes per week}, B_4 = {15–24 cigarettes per week},

B_5 = {25–44 cigarettes per week}, B_6 = {45+ cigarettes per week}.

We wish to compute the sensitivity = $Pr(A|B)$ where $B = B_2 \cup B_3$. We have

$$\begin{aligned} Pr(A|B) &= \frac{Pr(A \cap B)}{Pr(B)} = \frac{Pr(A \cap B_2) + Pr(A \cap B_3)}{Pr(B_2) + Pr(B_3)} \\ &= \frac{Pr(A|B_2)Pr(B_2) + Pr(A|B_3)Pr(B_3)}{\frac{(70+30)}{1332}} = \frac{.043\left(\frac{70}{1332}\right) + .067\left(\frac{30}{1332}\right)}{\frac{100}{1332}} \\ &= \frac{.043 \times 70 + .067 \times 30}{100} = \frac{5}{100} = .050 \end{aligned}$$

3.69 We wish to compute the sensitivity = $Pr(A|C)$ where $C = B_4 \cup B_5$. We have

$$\begin{aligned} Pr(A|C) &= \frac{Pr(A \cap C)}{Pr(C)} = \frac{Pr(A|B_4)Pr(B_4) + Pr(A|B_5)Pr(B_5)}{Pr(B_4) + Pr(B_5)} \\ &= \frac{.296\left(\frac{27}{1332}\right) + .368\left(\frac{19}{1332}\right)}{\frac{(27+19)}{1332}} = \frac{.296 \times 27 + .368 \times 19}{46} = \frac{15}{46} = .326 \end{aligned}$$

3.70 Sensitivity = $Pr(A|B_6) = .652$

3.71 Specificity = $Pr(\text{test -} | \text{nonsmoker}) = Pr(\bar{A}|B_1) = 1 - Pr(A|B_1) = 1 - .033 = .967$

3.72 We have that $PV+ = Pr(\text{true } + | \text{test } +) = Pr(\text{smoker } 1+ \text{ cigarettes per day } | A) = 1 - Pr(B_1 | A)$

$$\begin{aligned} &= 1 - \frac{Pr(B_1 \cap A)}{Pr(A)} = 1 - \frac{Pr(A|B_1)Pr(B_1)}{\sum_{i=1}^6 Pr(A|B_i)Pr(B_i)} \\ &= 1 - \frac{.033\left(\frac{1163}{1332}\right)}{.033\left(\frac{1163}{1332}\right) + \dots + .652\left(\frac{23}{1332}\right)} \\ &= 1 - \frac{.033 \times 1163}{.033 \times 1163 + \dots + .652 \times 23} \\ &= 1 - \frac{38}{73} = \frac{35}{73} = .479 \end{aligned}$$

3.73 We have that $PV- = Pr(\text{true } - | \text{test } -) = Pr(\text{nonsmoker } | \bar{A}) = Pr(B_1 | \bar{A}) = \frac{Pr(B_1 \cap \bar{A})}{Pr(\bar{A})}$

$$\begin{aligned} &= \frac{Pr(\bar{A} | B_1)Pr(B_1)}{1 - Pr(A)} = \frac{[1 - Pr(A | B_1)]Pr(B_1)}{1 - Pr(A)} \\ &= \frac{(1 - .033)\left(\frac{1163}{1332}\right)}{1 - \left(\frac{73}{1332}\right)} = \frac{.8443}{.9452} = .893 \end{aligned}$$

3.74 Let $B_{1, \text{none}}$ = no cigarettes actually consumed and reported non-smoker

$B_{1, 1-4}$ = 1-4 cig/wk actually consumed and reported non-smoker

$B_{1, 5-14}$ = 5-14 cig/wk actually consumed and reported non-smoker

It follows that

$$\begin{aligned} \text{observed specificity} &= Pr(\bar{A} | B_1) = Pr(\bar{A} | B_{1, \text{none}}) \times Pr(B_{1, \text{none}} | B_1) \\ &\quad + Pr(\bar{A} | B_{1, 1-4}) \times Pr(B_{1, 1-4} | B_1) \\ &\quad + Pr(\bar{A} | B_{1, 5-14}) \times Pr(B_{1, 5-14} | B_1) \end{aligned}$$

We have from Table 3.8 that

$$Pr(\bar{A} | B_{1, 1-4}) = 1 - .043 = .957, \quad Pr(\bar{A} | B_{1, 5-14}) = 1 - .067 = .933.$$

We wish to compute the true specificity = $Pr(\bar{A} | B_{1, \text{none}})$ if the observed specificity = $1 - .033 = .967$. If x = true specificity, then, we have

$$.967 = x(.70) + .957(.20) + .933(.10)$$

Thus,

$$x = \frac{.967 - .957(.20) - .933(.10)}{.70} = \frac{.6823}{.70} = .975.$$

3.75 We wish to compute $Pr(T | \bar{A})$ where T = true non-smoker, \bar{A} = $SCN < 100 \mu\text{g/mL}$. We have from Bayes'

Theorem that

$$Pr(T|\bar{A}) = \frac{Pr(\bar{A}|T)Pr(T)}{Pr(\bar{A}|T)Pr(T) + Pr(\bar{A}|\bar{T})Pr(\bar{T})}$$

From Problem 3.74, we have $Pr(\bar{A}|T) = x = .975$. Also,

$$Pr(T) = \frac{1163}{1332} \times .70 = .611$$

It remains to compute $Pr(\bar{A}|\bar{T})$. We have from Table 3.8 that there are 70 students who report 1-4 cig/wk, 30 who report 5-14 cig/wk, 27 who report 15-24 cig/wk, 19 who report 25-44 cig/wk, and 23 who report 45+ cig/wk. However, in addition there are $(.20)(1163) = 232.6$ students who report not smoking but who actually smoke 1-4 cig/wk and $.10(1163) = 116.3$ students who report not smoking but who actually smoke 5-14 cig/wk. Thus, there are a total of 302.6 students who actually smoke 1-4 cig/wk and 146.3 students who actually smoke 5-14 cig/wk. Thus, there are a total of 517.9 smoking students.

Furthermore,

$$\begin{aligned} Pr(\bar{A}|\bar{T}) &= \sum_{i=2}^6 Pr(\bar{A}|B_i)Pr(B_i|\bar{T}) = (1-.043)\left(\frac{302.6}{517.9}\right) + (1-.067)\left(\frac{146.3}{517.9}\right) \\ &\quad + (1-.296)\left(\frac{27}{517.9}\right) + (1-.368)\left(\frac{19}{517.9}\right) + (1-.652)\left(\frac{23}{517.9}\right) \\ &= .898 \end{aligned}$$

It follows that

$$Pr(T|\bar{A}) = \frac{.975(.611)}{.975(.611) + .898(.389)} = \frac{.596}{.945} = .630 = \text{true predictive value negative.}$$

This is lower than the PV^- based on self-report of .893 given in Problem 3.73.

3.76 The sensitivity is given by

$$Pr(\Delta DBP \text{ automated} \geq 10 \text{ mm Hg} | \Delta DBP \text{ manual} \geq 10 \text{ mm Hg}) = \frac{6}{13} = .462.$$

3.77 The specificity is given by

$$Pr(\Delta DBP \text{ automated} < 10 \text{ mm Hg} | \Delta DBP \text{ manual} < 10 \text{ mm Hg}) = \frac{51}{66} = .773.$$

3.78 We know that $PV^+ = Pr(\text{true}^+ | \text{test}^+)$

$$\begin{aligned} &= Pr(\Delta DBP \text{ manual} \geq 10 \text{ mm Hg} | \Delta DBP \text{ automated} \geq 10 \text{ mm Hg}) \\ &= \frac{6}{21} = .286. \text{ Furthermore, } PV^- = Pr(\text{true}^- | \text{test}^-) \\ &= Pr(\Delta DBP \text{ manual} < 10 \text{ mm Hg} | \Delta DBP \text{ automated} < 10 \text{ mm Hg}) \\ &= \frac{51}{58} = .879. \end{aligned}$$

- 4.25 The distribution of the number of gonorrhea cases that occurred over a 3-month period is approximated by a Poisson distribution with parameter $\mu = np = 10,000 \times \left(\frac{50}{100,000}\right) = 5.0$. From the Poisson tables, (Table 2) compute

$$Pr(X \geq 10 | \mu = 5) = .0181 + .0082 + .0034 + .0013 + .0005 + .0002 = .032$$

Thus, the number of gonorrhea cases in this county over this time period is unusual, since the probability of observing at least 10 cases is quite small (i.e., $< .05$).

- 4.26 We have that $Pr(k$ episodes of otitis media over time $t)$ is given by

$$\frac{e^{-1.6t} (1.6t)^k}{k!}, \quad k = 0, 1, 2, \dots$$

Thus,

$$Pr(0 \text{ episodes in 2 years}) = \frac{e^{-3.2} 3.2^0}{0!} = e^{-3.2} = .0408$$

$$Pr(1 \text{ episode in 2 years}) = \frac{e^{-3.2} (3.2)}{1!} = .1304$$

$$Pr(2 \text{ episodes in 2 years}) = \frac{e^{-3.2} (3.2)^2}{2!} = .2087$$

Thus, $Pr(X \leq 2) = .0408 + .1304 + .2087 = .380$, $Pr(X \geq 3) = 1 - .38 = .62$

- 4.27 $Pr(k$ episodes in 1 year) = $\frac{e^{-1.6} (1.6)^k}{k!}$. Thus, $Pr(0$ episodes in 1 year) = $e^{-1.6} = .202$.

- 4.28 From Problem 4.26, $Pr(1$ sibling has 3+ episodes of otitis in 2 years) = .62. Thus, if the number of episodes for two different siblings are independent random variables, then the probability that two siblings will each have 3+ episodes in the first two years of life = $.62^2 = .385$.

- 4.29 $Pr(1$ sibling will have 3+ episodes in the first 2 years of life) = ${}_2C_1 .62^1 .38^1 = .471$.

- 4.30 $Pr(\text{neither sibling will have 3+ episodes in the first 2 years of life}) = .38^2 = .144$.

- 4.31 The number of siblings with 3+ episodes of otitis media in the first 2 years of life is a binomial random variable with parameters $n = 2$, $p = .620$. Thus, the expected value = $np = 2 \times .620 = 1.24 =$ expected number of siblings with 3+ episodes of otitis media in the first 2 years of life.

- 4.32 The number of hypertensives in the sibships will be binomially distributed with parameters $n = 3$, $p = .18$. Therefore,

$$Pr(0 \text{ hypertensives}) = .82^3 = .5514$$

$$Pr(1 \text{ hypertensive}) = \binom{3}{1} (.18)^1 (.82)^2 = .3631$$

$$Pr(2 \text{ hypertensives}) = \binom{3}{2} (.18)^2 (.82)^1 = .0797$$

$$Pr(3 \text{ hypertensives}) = \binom{3}{3} (.18)^3 = .0058$$

- 4.33** The probability of 2+ affected siblings under the independence model $= .0797 + .0058 = .086$. Let X = the number of sibships with 2+ affected siblings. If independence holds, then X will be binomially distributed with parameters $n = 25$, $p = .086$. We must calculate $Pr(X \geq 5)$. Using the BINOMDIST function of Excel, we have $Pr(X \leq 4 | n = 25, p = .086) = .943$, $Pr(X \geq 5 | n = 25, p = .086) = .057$.

Thus, the occurrence of 5+ hypertensive households out of 25 households is somewhat unusual under the assumption of independence, and we may suspect that this assumption does not hold. One possible alternative assumption is that if one sibling is hypertensive, then it is more likely that other siblings in the household will be hypertensive, due to common environmental factors.

- 4.34** We wish to compute $Pr(X \geq 4) = 1 - Pr(X \leq 3)$, where X is binomially distributed with parameters $n = 100$, $p = .025$. We use the BINOMDIST function of Excel. We have:

$$Pr(X \leq 3 | n = 100, p = .025) = .759, Pr(X \geq 4 | n = 100, p = .025) = .241.$$

Thus, since this probability is not small, there is no excess risk of malformations in this group.

- 4.35** We wish to compute $Pr(X \geq 8) = 1 - Pr(X \leq 7)$, where X is binomially distributed with parameters $n = 75$, $p = .025$. We use the BINOMDIST function of Excel as follows:

$$Pr(X \leq 7 | n = 75, p = .025) = .9994, Pr(X \geq 8 | n = 75, p = .025) = .0006.$$

Since this probability is so low, we can conclude that there is an excess risk of major malformations among offspring of women who use marijuana.

- 4.36** The probability that a hypertensive is being treated appropriately and is complying with the treatment is

$$\begin{aligned} & Pr(\text{hypertensive is told he or she has high blood pressure}) \\ & \quad \times Pr(\text{adequately treated} | \text{told}) \\ & \quad \times Pr(\text{complying} | \text{adequately treated}) \\ & = \left(\frac{1}{2}\right)^3 = .125 \end{aligned}$$

Thus, we want

$$\begin{aligned} Pr(X \geq 5) &= 1 - Pr(X \leq 4) \\ &= 1 - \sum_{k=0}^4 \binom{10}{k} .125^k .875^{10-k} \end{aligned}$$

We use Excel as follows $Pr(X \leq 4 | n = 10, p = .125) = .996$, $Pr(X \geq 5 | n = 10, p = .125) = .004$.

4.37 $Pr(\text{hypertensive knows he or she has high blood pressure}) = \frac{1}{2}$. We want

$$Pr(X \geq 7) = \binom{10}{7} \left(\frac{1}{2}\right)^{10} + \binom{10}{8} \left(\frac{1}{2}\right)^{10} + \binom{10}{9} \left(\frac{1}{2}\right)^{10} + \binom{10}{10} \left(\frac{1}{2}\right)^{10}$$

Refer to the binomial tables (Table 1 of the Appendix) with $n = 10$, $p = .50$ and find that

$$Pr(X = 10) = .0010$$

$$Pr(X = 9) = .0098$$

$$Pr(X = 8) = .0439$$

$$Pr(X = 7) = .1172$$

Thus, $Pr(X \geq 7) = .0010 + .0098 + .0439 + .1172 = .172$.

4.38 If the rates are each decreased to 40%, then $(.6)^3 = .216 = 21.6\%$ of hypertensives will be appropriately treated as opposed to $(.5)^3 = .125 = 12.5\%$. Thus, if the current annual mortality rate for *untreated* hypertensives is x and for *treated* hypertensives is $.8x$, then the *current overall mortality rate* for hypertensives is

$$.125(.8x) + .875x = .975x$$

The *new overall mortality rate* would be

$$.216(.8x) + .784x = .957x$$

Thus, the ratio of the new mortality to the current mortality rate is

$$\frac{.957x}{.975x} = 0.981 = 98.1\%$$

Thus, the overall mortality rate among hypertensives would be reduced by 1.9%.

4.39 We have

$$Pr(k \text{ positives}) = \binom{5}{k} .05^k .95^{5-k}$$

Thus, $Pr(1 \text{ or more } +) = 1 - Pr(0+)$

$$Pr(0+) = .95^5 = .77$$

Thus, $Pr(1 \text{ or more } +) = .23$

4.40 $Pr(3 \text{ or more } +) = 1 - Pr(2 \text{ or less } +)$

$$Pr(2 \text{ or less } +) = Pr(0) + Pr(1) + Pr(2)$$

$$Pr(0) = .95^{100} = .0059$$

$$Pr(1) = \binom{100}{1} (.05)^1 (.95)^{99} = .0312$$

$$Pr(2) = \binom{100}{2} (.05)^2 (.95)^{98} = .0812$$

Hence, $Pr(2 \text{ or less } +) = .118$ and $Pr(3 \text{ or more } +) = .882$.

4.41 We know that X can only take on the values 0, 1, or 2.

$$\begin{aligned} Pr(0) &= Pr(2 \text{ negatives}) = Pr(\text{negative at time 0}) \\ &\quad \times Pr(\text{negative at time 1} | \text{negative at time 0}) \\ &= .95(1-.042) = .95(.958) = .910 \end{aligned}$$

$$\begin{aligned} Pr(1) &= Pr(1 \text{ positive}) \\ &= Pr(\text{negative at time 0} \cap \text{positive at time 1}) \\ &\quad + Pr(\text{positive at time 0} \cap \text{negative at time 1}) \\ &= Pr(\text{negative at time 0}) \times Pr(\text{positive at time 1} | \text{negative at time 0}) \\ &\quad + Pr(\text{positive at time 0}) \times Pr(\text{negative at time 1} | \text{positive at time 0}) \\ &= .95(.042) + .05(.80) = .080 \end{aligned}$$

$$\begin{aligned} Pr(2) &= Pr(2 \text{ positives}) = Pr(\text{positive at time 0}) \times Pr(\text{positive at time 1} | \text{positive at time 0}) \\ &= .05(.20) = .010 \end{aligned}$$

Thus, the probability of distribution X is

X	$Pr(X)$
0	.910
1	.080
2	.010

4.42 Mean of $X = E[X] = 0(.910) + 1(.080) + 2(.010) = 0.100 = \mu$.

4.43 Variance of $X = E[X^2] - \mu^2$

$$E[X^2] = 0(.910) + 1(.080) + 4(.010) = 0.120$$

$$Var(X) = 0.120 - 0.100^2 = 0.110$$

4.44 The 11 males are ages 82, 60, 50, 4, 32, 36, 69, 47, 22, 11, and 19. Based on the life table data, the first male has probability $p_1 = \frac{(l_{82} - l_{83})}{l_{82}} = \frac{(28,705 - 25,712)}{28,705} = \frac{2,993}{28,705} = .104$ of dying in the next year.

Similarly, the following probabilities of death (p_i) are obtained for the i th man, $i = 1, \dots, 11$ in the next year

i	Age(x)	l_x	l_{x+1}	P_i	i	Age(x)	l_x	l_{x+1}	P_i
1	82	28,705	25,712	.104	7	69	65,161	62,881	.035
2	60	80,908	79,539	.017	8	47	91,413	90,960	.005
3	50	89,919	89,320	.007	9	22	97,372	97,196	.002
4	4	98,658	98,620	.000	10	11	98,464	98,443	.000
5	32	95,606	95,405	.002	11	19	97,852	97,702	.002
6	36	94,771	94,546	.002					

The expected number of deaths over 1 year = $\sum p_i = .104 + \dots + .002 = .176$.

5.17 We have that $X \sim N(12.8, 5.1^2)$. We wish to compute $Pr(X > 20)$. We have

$$\begin{aligned} Pr(X > 20) &= 1 - Pr(X \leq 20) \\ &= 1 - \Phi\left(\frac{20 - 12.8}{5.1}\right) \\ &= 1 - \Phi(1.412) \\ &= 1 - .9210 = .079 \end{aligned}$$

5.18 Here, $X \sim N(9.3, 3.2^2)$. We wish to compute $Pr(X > 20)$. We have that

$$\begin{aligned} Pr(X > 20) &= 1 - Pr(X \leq 20) \\ &= 1 - \Phi\left(\frac{20 - 9.3}{3.2}\right) \\ &= 1 - \Phi(3.34) \\ &= 1 - .9996 = .0004 \end{aligned}$$

5.19 Suppose the rate given in previous census data is correct. The number of cases of cervical cancer, X , out of 100,000 women would be binomially distributed with parameters $n = 100,000$ and $p = .002$. This random variable can be approximated by a normal random variable Y with mean $= 100,000 \times .002 = 200$, and variance $= 100,000 \times .002 \times .998 = 199.6$. We wish to compute

$$\begin{aligned} Pr(X \leq 100) &\cong Pr(Y \leq 100.5) \\ &= \Phi\left(\frac{100.5 - 200}{\sqrt{199.6}}\right) \\ &= 1 - \Phi(7.04) < .001 \end{aligned}$$

Thus, it is very unlikely that the old rate of 2 per 1,000 is still valid.

5.20 Let $X =$ serum cholesterol. Compute

$$\begin{aligned} Pr(X \leq 200) &= \Phi\left(\frac{200 - 219}{50}\right) \\ &= \Phi(-0.38) = 1 - \Phi(0.38) \\ &= 1 - .6480 = .3520 \end{aligned}$$

5.21 We want

$$\begin{aligned} Pr(X \geq 250) &= 1 - \Phi\left(\frac{250 - 219}{50}\right) \\ &= 1 - \Phi(0.62) = 1 - .7324 = .2676 \end{aligned}$$

5.22 We want

$$\begin{aligned} Pr(200 < X < 250) &= \Phi\left(\frac{250 - 219}{50}\right) - \Phi\left(\frac{200 - 219}{50}\right) \\ &= \Phi(0.62) - \Phi(-0.38) \\ &= \Phi(0.62) - [1 - \Phi(0.38)] \\ &= \Phi(0.62) + \Phi(0.38) - 1 \\ &= 0.7324 + .6480 - 1 = .3804 \end{aligned}$$

5.46 We wish to compute $Pr(X \geq 0.30)$ where $X \sim N(0.80, 0.48^2)$. We have

$$\begin{aligned} Pr(X \geq 0.30) &= 1 - Pr(X < 0.30) \\ &= 1 - \Phi\left(\frac{0.30 - 0.80}{0.48}\right) \\ &= 1 - \Phi\left(\frac{-0.50}{0.48}\right) \\ &= 1 - \Phi(-1.04) = \Phi(1.04) = .851 \end{aligned}$$

Thus, 85% of the vitamin E group would be expected to show a change of ≥ 0.30 mg/dL.

5.47 This probability = $Pr(\text{test} + | \text{vitamin E taker}) = \text{sensitivity}$.

5.48 We wish to compute $Pr(X < 0.30)$ where $X \sim N(0.05, 0.16^2)$. We have

$$\begin{aligned} Pr(X < 0.30) &= \Phi\left(\frac{0.30 - 0.05}{0.16}\right) \\ &= \Phi\left(\frac{0.25}{0.16}\right) = \Phi(1.56) = .941 \end{aligned}$$

Thus, 94% of the placebo group would be expected to show a change of not more than 0.30 mg/dL.

5.49 This probability = $Pr(\text{test} - | \text{placebo taker}) = \text{specificity}$.

5.50 If our threshold = Δ , then

$$\begin{aligned} \text{sensitivity} &= 1 - \Phi\left(\frac{\Delta - 0.80}{0.48}\right) = \Phi\left(\frac{0.80 - \Delta}{0.48}\right) \\ \text{specificity} &= \Phi\left(\frac{\Delta - 0.05}{0.16}\right) \end{aligned}$$

If we want sensitivity = specificity, then we require

$$\frac{0.80 - \Delta}{0.48} = \frac{\Delta - 0.05}{0.16}$$

Solving for Δ , we get $\Delta = 0.2375$ mg/dL. The estimated compliance in each group is given by

$$\begin{aligned} \text{Vitamin E: sensitivity} &= \Phi\left(\frac{0.80 - 0.2375}{0.48}\right) = \Phi(1.17) = .88 \\ \text{Placebo: specificity} &= \Phi\left(\frac{0.2375 - 0.05}{0.16}\right) = \Phi(1.17) = .88 \end{aligned}$$

Thus, the compliance would be 88% in each group using this measure of compliance.

5.51 From Problem 3.27, the expected overall prevalence of Alzheimer's disease is .061. If we study n people, then the number of people (X) with Alzheimer's disease will be binomially distributed with parameters n and $p = .061$. We approximate this distribution by a normal distribution Y with mean = $.061n$ and variance = $.061 \times .939 \times n = .0573n$. We wish to find n such that $Pr(X \geq 100)$