Solutions of suggested problems

CH3

Consider a family with a mother, father, and two children. Let A_1 = {mother has influenza}, A_2 = {father has influenza}, A_3 = {first child has influenza}, A_4 = {second child has influenza}, B = {at least one child has influenza}, C = {at least one parent has influenza}, and D = {at least one person in the family has influenza}.

*3.1 What does $A_1 \cup A_2$ mean?

*3.2 What does $A_1 \cap A_2$ mean?

*3.3 Are A_3 and A_4 mutually exclusive?

*3.4 What does $A_3 \cup B$ mean?

*3.5 What does $A_3 \cap B$ mean?

*3.6 Express C in terms of A_1, A_2, A_3, and A_4.

*3.7 Express D in terms of B and C.

*3.8 What does $\bar{A}_1$ mean?

*3.9 What does $\bar{A}_2$ mean?

*3.10 Represent $\bar{C}$ in terms of A_1, A_2, A_3, and A_4.

*3.11 Represent $\bar{D}$ in terms of B and C.

3.1 $A_1 \cup A_2$ means that at least one parent has influenza.

3.2 $A_1 \cap A_2$ means that both parents have influenza.

3.3 No. Both children can have influenza.

3.4 $A_3 \cup B$ means that at least one child has influenza, because if A_3 occurs, then B must occur. Therefore, $A_3 \cup B = B$.

3.5 $A_3 \cap B$ means that the first child has influenza. Therefore, $A_3 \cap B = A_3$.

3.6 $C = A_1 \cup A_2$

3.7 $D = B \cup C$

3.8 $\bar{A}_1$ means that the mother does not have influenza.

3.9 $\bar{A}_2$ means that the father does not have influenza.

3.10 $\bar{C} = \bar{A}_1 \cap \bar{A}_2$

3.11 $\bar{D} = \bar{B} \cap \bar{C}$

Suppose the probability that both members of a married couple, each of whom is 75–79 years of age, will have Alzheimer's disease is .0015.

3.23 What is the conditional probability that the man will be affected given that the woman is affected? How does this value compare with the prevalence in Table 3.5? Why should it be the same (or different)?

3.24 What is the conditional probability that the woman will be affected given that the man is affected? How does this value compare with the prevalence in Table 3.5? Why should it be the same (or different)?

3.25 What is the probability that at least one member of the couple is affected?

Suppose a study of Alzheimer's disease is proposed in a retirement community with people 65+ years of age, where the age–gender distribution is as shown in Table 3.6.

3.26 What is the expected overall prevalence of Alzheimer's disease in the community if the prevalence estimates in Table 3.5 for specific age–gender groups hold?

3.27 Assuming there are 1000 people 65+ years of age in the community, what is the expected number of cases of Alzheimer's disease in the community?

3.23 $Pr(\text{man affected} \mid \text{woman affected}) = \frac{.0015}{.023} = .065$. It is higher than the value in Table 3.5 (.049), indicating that these are dependent events.

3.24 $Pr(\text{woman affected} \mid \text{man affected}) = \frac{.0015}{.049} = .031$. This value is also higher than the unconditional probability in Table 3.5 (.023). If there is some common environmental factor that is associated with Alzheimer's disease, then it would make sense that the conditional probability is higher than the unconditional probability.

3.25 Let $A = \{\text{man affected}\}$, $B = \{\text{woman affected}\}$. We have
$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B) = .049 + .023 - .0015 = .0705$

3.26 Let $Pr(A)$ denote the overall probability of Alzheimer's disease. We have that

$$Pr(A) = Pr(A|65-69M) \times Pr(65-69M) + \ \ldots\ + Pr(A|85+F) \times Pr(85+F)$$
$$= .05 \times .016 + .10 \times .000 + \ \ldots\ + .06 \times .279 = .061$$

Therefore, the expected overall prevalence in the community is 6.1%.

3.27 The expected number of cases with Alzheimer's disease $= 1000 \times .061 = 61$.

Suppose that a disease is inherited via a **dominant mode of inheritance** and that only one of the two parents is affected with the disease. The implications of this mode of inheritance are that the probability is 1 in 2 that any particular offspring will get the disease.

3.32 What is the probability that in a family with two children, both siblings are affected?

3.33 What is the probability that exactly one sibling is affected?

3.34 What is the probability that neither sibling is affected?

3.35 Suppose the older child is affected. What is the probability that the younger child is affected?

3.36 If A, B are two events such that A = {older child is affected}, B = {younger child is affected}, then are the events A, B independent?

3.32 The probability that both siblings are affected is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

3.33 The probability that exactly one sibling is affected is $2 \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}$

3.34 The probability that neither sibling will be affected is $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

3.35 The probability that the younger child is affected should not be influenced by whether or not the older child is affected. Thus, the probability of the younger child being affected remains at $\frac{1}{2}$.

3.36 The events A, B are independent because whether or not a child is affected does not influence the outcome for other children in the family.

Research into cigarette-smoking habits, smoking prevention, and cessation programs necessitates accurate measurement of smoking behavior. However, decreasing social acceptability of smoking appears to cause significant underreporting. Chemical markers for cigarette use can provide objective indicators of smoking behavior. One widely used noninvasive marker is the level of saliva thiocyanate (SCN). In a Minneapolis school district, 1332 students in eighth grade (ages 12–14) participated in a study [12] whereby they

(1) Viewed a film illustrating how recent cigarette use could be readily detected from small samples of saliva

(2) Provided a personal sample of SCN

(3) Provided a self-report of the number of cigarettes smoked per week

The results are given in Table 3.10.

TABLE 3.10 Relationship between SCN levels and self-reported cigarettes smoked per week

Self-reported cigarettes smoked in past week	Number of students	Percent with SCN $\geq 100\ \mu g/mL$
None	1163	3.3
1–4	70	4.3
5–14	30	6.7
15–24	27	29.6
25–44	19	36.8
45+	23	65.2

Source: Based on the *American Journal of Public Health*, *71*(12), 1320, 1981.

Suppose the self-reports are completely accurate and are representative of the number of eighth-grade students who smoke in the general community. We are considering using an SCN level $\geq 100\ \mu g/mL$ as a test criterion for identifying cigarette smokers. Regard a student as positive if he or she smokes one or more cigarettes per week.

***3.68** What is the sensitivity of the test for light-smoking students (students who smoke ≤ 14 cigarettes per week)?

***3.69** What is the sensitivity of the test for moderate-smoking students (students who smoke 15–44 cigarettes per week)?

***3.70** What is the sensitivity of the test for heavy-smoking students (students who smoke ≥ 45 cigarettes per week)?

***3.71** What is the specificity of the test?

***3.72** What is the PV^+ of the test?

***3.73** What is the PV^- of the test?

Suppose we regard the self-reports of all students who report some cigarette consumption as valid but estimate that 20% of students who report no cigarette consumption actually smoke 1–4 cigarettes per week and an additional 10% smoke 5–14 cigarettes per week.

***3.74** Assuming the percentage of students with SCN $\geq 100\ \mu g/mL$ in these two subgroups is the same as in those who truly report 1–4 and 5–14 cigarettes per week, compute the specificity under these assumptions.

***3.75** Compute the PV^- under these altered assumptions. How does the true PV^- using a screening criterion of SCN $\geq 100\ \mu g/mL$ for identifying smokers compare with the PV^- based on self-reports obtained in Problem 3.73?

3.68 Let $A = \{\text{test} +\}$, $B_1 = \{\text{no cigarettes}\}$, $B_2 = \{1-4 \text{ cigarettes per week}\}$,

$B_3 = \{5-14 \text{ cigarettes per week}\}$, $B_4 = \{15-24 \text{ cigarettes per week}\}$,

$B_5 = \{25-44 \text{ cigarettes per week}\}$, $B_6 = \{45+ \text{ cigarettes per week}\}$.

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

$$Pr(A \mid B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{Pr(A \cap B_2) + Pr(A \cap B_3)}{Pr(B_2) + Pr(B_3)}$$

$$= \frac{Pr(A \mid B_2)Pr(B_2) + Pr(A \mid B_3)Pr(B_3)}{\frac{(70+30)}{1332}} = \frac{.043\left(\frac{70}{1332}\right) + .067\left(\frac{30}{1332}\right)}{\frac{100}{1332}}$$

$$= \frac{.043 \times 70 + .067 \times 30}{100} = \frac{5}{100} = .050$$

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

$$Pr(A \mid C) = \frac{Pr(A \cap C)}{Pr(C)} = \frac{Pr(A \mid B_4)Pr(B_4) + Pr(A \mid B_5)Pr(B_5)}{Pr(B_4) + Pr(B_5)}$$

$$= \frac{.296\left(\frac{27}{1332}\right) + .368\left(\frac{19}{1332}\right)}{\frac{(27+19)}{1332}} = \frac{.296 \times 27 + .368 \times 19}{46} = \frac{15}{46} = .326$$

3.70 Sensitivity $= Pr(A \mid B_6) = .652$

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

3.72 We have that

$$PV+ = \Pr(\text{true} + \mid \text{test}+) = \Pr(\text{smoker } 1+ \text{ cigarettes per day} \mid A) = 1 - \Pr(B_1 \mid A)$$

$$= 1 - \frac{\Pr(B_1 \cap A)}{\Pr(A)} = 1 - \frac{\Pr(A \mid B_1)\Pr(B_1)}{\Pr(A)} = 1 - \frac{0.033 \times \frac{1163}{1332}}{0.033 \times \frac{1163}{1332} + \cdots + 0.652 \times \frac{23}{1332}}$$

$$= 1 - \frac{0.033 \times 1163}{0.033 \times 1163 + \cdots + 0.652 \times 23} = 1 - \frac{38}{73} = \frac{35}{73} = 0.479$$

3.73 We have that

$$PV- = \Pr(\text{true -} \mid \text{test -}) = \Pr(\textit{non}\text{smoker} \mid \bar{A}) = \Pr(B_1 \mid \bar{A})$$

$$= \frac{\Pr(B_1 \cap \bar{A})}{\Pr(\bar{A})} = \frac{\Pr(\bar{A} \mid B_1)\Pr(B_1)}{1 - \Pr(A)} = \frac{1 - \Pr(A \mid B_1)\Pr(B_1)}{1 - \Pr(A)} = \frac{(1 - 0.033) \times \frac{1163}{1332}}{1 - \frac{73}{1332}}$$

$$= \frac{0.8443}{0.9452} = 0.893$$

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

$$\begin{aligned}\text{Observed specificity} = \Pr\left(\bar{A} \mid B_1\right) &= \Pr\left(\bar{A} \mid B_{1,none}\right) \times \Pr\left(B_{1,none} \mid B_1\right)\\ &+ \Pr\left(\bar{A} \mid B_{1,1-4}\right) \times \Pr\left(B_{1,1-4} \mid B_1\right)\\ &+ \Pr\left(\bar{A} \mid B_{1,5-14}\right) \times \Pr\left(B_{1,5-14} \mid B_1\right)\end{aligned}$$

We have from Table 3.10 that

$$\Pr\left(\bar{A} \mid B_{1,1-4}\right) = 1-0.043 = 0.957, \quad \Pr\left(\bar{A} \mid B_{1,5-14}\right) = 1-0.067 = 0.933$$

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

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

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

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

Solutions of suggested problems

CH4

4.9 Suppose 6 of 15 students in a grade-school class develop influenza, whereas 20% of grade-school students nationwide develop influenza. Is there evidence of an excessive number of cases in the class? That is, what is the probability of obtaining at least 6 cases in this class if the nationwide rate holds true?

4.10 What is the expected number of students in the class who will develop influenza?

4.9 We wish to compute

$$Pr(X \geq 6) = 1 - Pr(X \leq 5) = 1 - \sum_{k=0}^{5} {}_{15}C_k (.2)^k (.8)^{15-k}$$

We refer to the exact binomial tables (Table 1) under $n = 15$, $p = .2$, to obtain

$$\begin{aligned} Pr(X \geq 6) &= 1 - (.0352 + .1319 + .2309 + .2501 + .1876 + .1032) \\ &= 1 - .9389 = .061 \end{aligned}$$

4.10 $E(X) = 15 \times .2 = 3.0$

Newborns were screened for human immunodeficiency virus (HIV) or acquired immunodeficiency syndrome (AIDS) in five Massachusetts hospitals. The data [8] are shown in Table 4.15.

4.14 If 500 newborns are screened at the inner-city hospital, then what is the exact binomial probability of exactly 5 HIV-positive test results?

4.17 Answer Problem 4.14 for a mixed urban/suburban hospital (hospital C).

4.14 Let $X_1 =$ the number of inner-city newborns with HIV-positive test results. We have that

$$\Pr(X_1 = 5) = {}_{500}C_5\left(\frac{30}{3741}\right)^5\left(\frac{3711}{3741}\right)^{495}$$
$$= \frac{500\times 499\times 498\times 496}{5\times 4\times 3\times 2\times 1}\left(\frac{30}{3741}\right)^5\left(\frac{3711}{3741}\right)^{495}$$
$$= \left(2.5524\times 10^{11}\right)\left(3.3164\times 10^{-11}\right)(.0186)$$
$$= .1573$$

4.17 Let $X_2 =$ number of mixed urban/suburban hospital infants that are positive for the HIV virus. We have that

$$Pr(X_2 = 5) = {}_{500}C_5\left(\frac{11}{5006}\right)^5\left(\frac{4995}{5006}\right)^{495}$$
$$= \left(2.5524\times 10^{11}\right)\left(5.1228\times 10^{-14}\right)(.3366)$$
$$= .0044 \cong .004$$

The presence of bacteria in a urine sample (bacteriuria) is sometimes associated with symptoms of kidney disease in women. Suppose a determination of bacteriuria has been made over a large population of women at one point in time and 5% of those sampled are positive for bacteriuria.

***4.33** If a sample size of 5 is selected from this population, what is the probability that 1 or more women are positive for bacteriuria?

***4.34** Suppose 100 women from this population are sampled. What is the probability that 3 or more of them are positive for bacteriuria?

One interesting phenomenon of bacteriuria is that there is a "turnover"; that is, if bacteriuria is measured on the same woman at two different points in time, the results are not necessarily the same. Assume that 20% of all women who are bacteriuric at time 0 are again bacteriuric at time 1 (1 year later), whereas only 4.2% of women who were not bacteriuric at time 0 *are* bacteriuric at time 1. Let X be the random variable representing the number of bacteriuric events over the two time periods for 1 woman and still assume that the probability that a woman will be positive for bacteriuria at any one exam is 5%.

***4.35** What is the probability distribution of X?

***4.36** What is the mean of X?

***4.37** What is the variance of X?

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

4.34 $Pr(3 \text{ or more } +) = 1 - Pr(2 \text{ or less } +)$

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

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

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

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

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

4.35 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

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

4.37 Variance of $X = E\left(X^2\right) - \mu^2$

$$E\left(X^2\right) = 0(.910) + 1(.080) + 4(.010) = 0.120$$

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

Solutions of suggested problems

CH5

Because serum cholesterol is related to age and sex, some investigators prefer to express it in terms of *z*-scores. If X = raw serum cholesterol, then $Z = \dfrac{X - \mu}{\sigma}$, where μ is the mean and σ is the standard deviation of serum cholesterol for a given age–gender group. Suppose Z is regarded as a standard normal random variable.

*__5.1__ What is $Pr(Z < 0.5)$?

*__5.2__ What is $Pr(Z > 0.5)$?

*__5.3__ What is $Pr(-1.0 < Z < 1.5)$?

5.1 .6915

5.2 .3085

5.3 .7745

Suppose that total carbohydrate intake in 12- to 14-year-old boys is normally distributed, with mean = 124 g/1000 cal and standard deviation = 20 g/1000 cal.

5.6 What percentage of boys in this age range have carbohydrate intake above 140 g/1000 cal?

5.7 What percentage of boys in this age range have carbohydrate intake below 90 g/1000 cal?

Suppose boys in this age range who live below the poverty level have a mean carbohydrate intake of 121 g/1000 cal with a standard deviation of 19 g/1000 cal.

5.8 Answer Problem 5.6 for boys in this age range and economic environment.

5.9 Answer Problem 5.7 for boys in this age range and economic environment.

5.6 If X represents total carbohydrate intake in 12–14-year-old males, then we compute

$$\Pr(Y > 140) = 1 - \Phi\left(\frac{140-124}{20}\right)$$
$$= 1 - \Phi(0.80) = 1 - 0.7881 = 0.212$$

5.7 We compute

$$\Pr(Y < 90) = \Phi\left(\frac{90-124}{20}\right)$$
$$= \Phi(-1.70) = 1 - \Phi(1.70) = 0.0446$$

5.8 Let Y represent carbohydrate level in 12–14-year-old boys below the poverty level. We wish to compute

$$\Pr(Y > 140) = 1 - \Phi\left(\frac{140-121}{19}\right)$$
$$= 1 - \Phi(1.00) = 0.1587$$

5.9 We compute

$$\Pr(Y < 90) = \Phi\left(\frac{90-121}{19}\right)$$
$$= \Phi(-1.632) = 1 - \Phi(1.632) = 0.514$$

Serum cholesterol is an important risk factor for coronary disease. We can show that serum cholesterol is approximately normally distributed, with mean = 219 mg/dL and standard deviation = 50 mg/dL.

***5.14** If the clinically desirable range for cholesterol is < 200 mg/dL, what proportion of people have clinically desirable levels of cholesterol?

***5.15** Some investigators believe that only cholesterol levels over 250 mg/dL indicate a high-enough risk for heart disease to warrant treatment. What proportion of the population does this group represent?

***5.16** What proportion of the general population has borderline high-cholesterol levels—that is, > 200 but < 250 mg/dL?

5.14 Let $X =$ serum cholesterol. Compute

$$Pr(X \le 200) = \Phi\left(\frac{200-219}{50}\right) = \Phi(-0.38) = 1-\Phi(0.38) = 1-.6480 = .352$$

5.15 We want

$$Pr(X \ge 250) = 1-\Phi\left(\frac{250-219}{50}\right) = 1-\Phi(0.62) = 1-.7324 = .268.$$

5.16 We want

$$\begin{aligned} Pr(200 < X < 250) &= \Phi\left(\frac{250-219}{50}\right) - \Phi\left(\frac{200-219}{50}\right) \\ &= \Phi(0.62) - \Phi(-0.38) \\ &= \Phi(0.62) - [1-\Phi(0.38)] \\ &= \Phi(0.62) + \Phi(0.38) - 1 \\ &= 0.7324 + .6480 - 1 = .380 \end{aligned}$$

In pharmacologic research a variety of clinical chemistry measurements are routinely monitored closely for evidence of side effects of the medication under study. Suppose typical blood-glucose levels are normally distributed, with mean = 90 mg/dL and standard deviation = 38 mg/dL.

5.31 If the normal range is 65–120 mg/dL, then what percentage of values will fall in the normal range?

5.32 In some studies only values at least 1.5 times as high as the upper limit of normal are identified as abnormal. What percentage of values would fall in this range?

5.33 Answer Problem 5.32 for values 2.0 times the upper limit of normal.

5.34 Frequently, tests that yield abnormal results are repeated for confirmation. What is the probability that for a normal person a test will be at least 1.5 times as high as the upper limit of normal on two separate occasions?

5.35 Suppose that in a pharmacologic study involving 6000 patients, 75 patients have blood-glucose levels at least 1.5 times the upper limit of normal on one occasion. What is the probability that this result could be due to chance?

5.31 If the random variable X represents blood glucose, then we wish to compute

$$\begin{aligned}\Pr(65 \le X \le 120) &= \Phi\left(\frac{120-90}{38}\right) - \Phi\left(\frac{65-90}{38}\right)\\ &= \Phi(0.789) - \Phi(-0.658)\\ &= \Phi(0.789) - (1-\Phi(0.658))\\ &= 0.785 - 0.255 = 0.53\end{aligned}$$

Thus, 53% of people will fall in the normal range.

5.32 1.5 times the upper limit of normal $= 1.5 \times 120 = 180$. We wish to compute

$$\begin{aligned}\Pr(X \ge 180) &= 1-\Phi\left(\frac{180-90}{38}\right)\\ &= 1-\Phi(2.37) = 0.0089\end{aligned}$$

Thus, 0.9% of normal people should have values that are at least 1.5 times the upper limit of normal.

5.33 Since $2 \times 120 = 240$, we evaluate

$$\begin{aligned}\Pr(X \ge 240) &= 1-\Phi\left(\frac{240-90}{38}\right)\\ &= 1-\Phi(3.95) < 0.0001\end{aligned}$$

This would be an extremely unusual level in a normal person.

5.34 The probability is given by $.0089^2 = .00008$. This is the reason why abnormal lab tests are repeated, since the probability of abnormal results on two separate occasions for normal people is very low.

5.35 Let the random variable X represent the number of patients in the study with abnormal blood-glucose levels on one occasion. X is binomially distributed with parameters $n = 6000$ and $p = .0089$. We approximate X by a normal random variable Y with mean $\mu = np = 6000 \times .0089 = 53.6$ and variance $npq = 6000 \times .0089 \times .9911 = 53.11$. We wish to compute

$$\begin{aligned}\Pr(X \ge 75) \cong \Pr(Y \ge 74.5) &= 1-\Phi\left(\frac{74.5-53.6}{\sqrt{53.11}}\right)\\ &= 1-\Phi(2.87) = 0.0021\end{aligned}$$

A treatment trial is proposed to test the efficacy of vitamin E as a preventive agent for cancer. One problem with such a study is how to assess compliance among participants. A small pilot study is undertaken to establish criteria for compliance with the proposed study agents. In this study, 10 patients are given 400 IU/day of vitamin E and 10 patients are given similar-sized placebo capsules over a 3-month period. Their serum vitamin E levels are measured before and after the 3-month period, and the change (3-month – baseline) is shown in Table 5.2.

TABLE 5.2 Change in serum vitamin E (mg/dL) in pilot study

Group	Mean	*sd*	*n*
Vitamin E	0.80	0.48	10
Placebo	0.05	0.16	10

***5.36** Suppose a change of 0.30 mg/dL in serum levels is proposed as a test criterion for compliance; that is, a patient who shows a change of ≥ 0.30 mg/dL is considered a compliant vitamin E taker. If normality is assumed, what percentage of the vitamin E group would be expected to show a change of at least 0.30 mg/dL?

***5.37** Is the measure in Problem 5.36 a measure of sensitivity, specificity, or predictive value?

***5.38** What percentage of the placebo group would be expected to show a change of not more than 0.30 mg/dL?

***5.39** Is the measure in Problem 5.38 a measure of sensitivity, specificity, or predictive value?

***5.40** Suppose a new threshold of change, Δ mg/dL, is proposed for establishing compliance. We wish to use a level of Δ such that the compliance measures in Problems 5.36 and 5.38 for the patients in the vitamin E and placebo groups are the same. What should Δ be? What would be the compliance in the vitamin E and placebo groups using this threshold level?

5.41 Suppose we consider the serum vitamin E assay as a screening test for compliance with vitamin E supplementation. Participants whose change in serum vitamin E is $\geq \Delta$ mg/dL will be considered vitamin E takers, and participants whose change is $< \Delta$ mg/dL will be considered placebo takers. Choose several possible values for Δ, and construct the receiver operating characteristic (ROC) curve for this test. What is the area under the ROC curve? (*Hint:* The area under the ROC curve can be computed analytically from the properties of linear combinations of normal distributions.)

5.36 We wish to compute $Pr(X \geq 0.30)$ where $X \sim N\left(0.8, 0.48^2\right)$. We have

$$\begin{aligned}\Pr\left(X \geq 0.30\right) &= 1 - \Pr\left(X < 0.30\right) = 1 - \Phi\left(\frac{0.30 - 0.80}{0.48}\right)\\ &= 1 - \Phi\left(\frac{-0.50}{0.48}\right)\\ &= 1 - \Phi\left(-1.04\right) = \Phi\left(1.04\right)\\ &= 0.851\end{aligned}$$

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

5.37 This probability = Pr(test + |vitamin E taker) = sensitivity.

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

$$\begin{aligned}\Pr\left(X < 0.30\right) &= \Phi\left(\frac{0.30 - 0.05}{0.16}\right)\\ &= \Phi\left(\frac{0.25}{0.16}\right)\\ &= \Phi\left(1.56\right) = 0.941\end{aligned}$$

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

5.39 This probability = Pr(test−|placebo taker) = specificity.

5.40 If our threshold $= \Delta$, then

$$\text{sensitivity} = 1 - \Phi\left(\frac{\Delta - 0.80}{0.48}\right) = \Phi\left(\frac{0.80 - \Delta}{0.48}\right)$$

$$\text{specificity} = \Phi\left(\frac{\Delta - 0.05}{0.16}\right)$$

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

$$\text{Vitamin E: sensitivity} = \Phi\left(\frac{0.80 - 0.2375}{0.48}\right) = \Phi\left(1.17\right) = 0.88$$

$$\text{Placebo: specificity} = \Phi\left(\frac{0.2375 - 0.05}{0.48}\right) = \Phi\left(1.17\right) = 0.88$$

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

5.41 We will choose $\Delta = 0, 0.10, \ldots, 2.0$ and compute the sensitivity and (1 – specificity) for each cutpoint based on the formula in problem 5.50 based on Excel. The results are given in the accompanying table and the ROC curve is plotted in the figure. The area under the ROC curve = *Pr*(serum vitamin E for a random vitamin E taker is greater than the serum vitamin E for a random placebo taker) $= Pr(X - Y > 0)$, where $X \sim N(0.80, 0.48^2)$ and $Y \sim N(0.05, 0.16^2)$. From Equation 5.10, $X - Y$ is normally distributed with mean $= 0.80 - 0.05 = 0.75$ and variance $= 0.48^2 + 0.16^2 = 0.256$. Thus, $\Pr(X - Y > 0) = 1 - \Phi\left[(0 - 0.75)/\sqrt{0.256}\right] = 1 - \Phi(-1.48) = \Phi(1.48) = .931$. Hence, the area under the ROC curve is 93%.

Sensitivity vs. 1-Specificity for Vitamin E Example

Cutpoint	Sensitivity	1-Specificity
0	0.952	0.623
0.1	0.928	0.377
0.2	0.894	0.174
0.3	0.851	0.059
0.4	0.798	0.014
0.5	0.734	0.002
0.6	0.662	0.000
0.7	0.583	0.000
0.8	0.500	0.000
0.9	0.417	0.000
1	0.338	0.000
1.1	0.266	0.000
1.2	0.202	0.000
1.3	0.149	0.000
1.4	0.106	0.000
1.5	0.072	0.000
1.6	0.048	0.000
1.7	0.030	0.000
1.8	0.019	0.000
1.9	0.011	0.000
2	0.006	0.000