3.5

1. Let events $A$ = “Live to age 60,” $B$ = “Live to age 70,” $C$ = “Live to age 80”; note that event $C$ is a subset of $B$, and that $B$ is a subset of $A$, i.e., they are nested: $C \subset B \subset A$. We are given that $P(A) = 0.90$, $P(B \mid A) = 0.80$, and $P(C \mid B) = 0.75$. Therefore, by the general formula $P(E \cap F) = P(E \mid F) \times P(F)$, we have

See Note → $P(B) = P(B \cap A) = P(B \mid A) \times P(A) = (0.80)(0.90) = 0.72$

$P(C) = P(C \cap B) = P(C \mid B) \times P(B) = (0.75)(0.72) = 0.54$

$P(C \mid A) = \frac{P(C \cap A)}{P(A)} = \frac{0.54}{0.90} = 0.60$

Note: If event $C$ occurs, then event $B$ must have occurred. If event $B$ occurs, then event $A$ must have occurred. Thus, the event $A$ in the intersection of “$B$ and $A$” is redundant, etc.

2. $A$ = “Angel barks” $B$ = “Brutus barks” $P(A) = 0.1$, $P(B) = 0.2$, $P(A \mid B) = 0.3 \Rightarrow P(A \cap B) = 0.06$

(a) Because $P(A) = 0.1$ is not equal to $P(A \mid B) = 0.3$, the events $A$ and $B$ are not independent! Or, equivalently, $P(A \cap B) = 0.06$ is not equal to $P(A) \times P(B) = (0.1)(0.2) = 0.02$.

(b)

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.1 + 0.2 - 0.06 = 0.24$

Via DeMorgan’s Law: $P(A^c \cap B^c) = 1 - P(A \cup B) = 1 - 0.24 = 0.76$

$P(A \cap B^c) = P(A) - P(A \cap B) = 0.1 - 0.06 = 0.04$

$P(A^c \cap B) = P(B) - P(A \cap B) = 0.2 - 0.06 = 0.14$

$P(A \cap B^c) + P(A^c \cap B) = 0.04 + 0.14 = 0.18$, or, $P(A \cup B) - P(A \cap B) = 0.24 - 0.06 = 0.18$

$P(B \mid A) = \frac{P(B \cap A)}{P(A)} = \frac{0.06}{0.1} = 0.6$

$P(B^c \mid A) = \frac{P(B^c \cap A)}{P(A)} = \frac{0.04}{0.1} = 0.4$, or more simply, $1 - P(B \mid A) = 1 - 0.6 = 0.4$

$P(A \mid B^c) = \frac{P(A \cap B^c)}{P(B^c)} = \frac{P(A \cap B^c)}{1 - P(B)} = \frac{0.04}{0.8} = 0.05$

\[
\begin{array}{c|c|c}
A & A^c & \\
\hline
B & 0.06 & 0.14 \\
B^c & 0.04 & 0.76 \\
\hline
0.10 & 0.90 & 1.00
\end{array}
\]

$0.20 = P(B)$

$0.80 = P(B^c)$
3. Urn Model: Events $A$ = “First ball is red” and $B$ = “Second ball is red.” In the “sampling without replacement” case illustrated, it was calculated that, reduced to lowest terms, $P(A) = 4/6 = 2/3$, $P(B) = 2/3$, and $P(A \cap B) = 12/30 = 2/5$. Since $P(A \cap B) = 2/5 \neq 4/9 = 2/3 \times 2/3 = P(A) \times P(B)$, it follows that the two events $A$ and $B$ are not statistically independent. This should be intuitively consistent; as this “population” is small, the probability that event $A$ occurs nontrivially affects that of event $B$, if the unit is not replaced after the first draw. However, in the “sampling with replacement” scenario, this is not the case. For, as illustrated below, $P(A) = 4/6 = 2/3$, $P(B) = 24/36 = 2/3$, and $P(A \cap B) = 16/36 = 4/9$. Hence, $P(A \cap B) = 4/9 = 2/3 \times 2/3 = P(A) \times P(B)$, and so it follows that events $A$ and $B$ are indeed statistically independent.

4. First note that, in this case, $A \subset B$ (event $A$ is a subset of event $B$), that is, if $A$ occurs, then $B$ occurs! (See Venn diagram.) In addition, the given information provides us with the following conditional probabilities: $P(A \mid B) = 0.75$, $P(B^c \mid A^c) = 0.80$. Expanding these out via the usual formulas, we obtain, respectively,

$$0.75 = P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)},$$

i.e., $P(A) = 0.75 \cdot P(B)$

and

$$0.80 = P(B^c \mid A^c) = \frac{P(B^c \cap A^c)}{P(A^c)} = \frac{P(B^c)}{P(A^c)} = \frac{1 - P(B)}{1 - P(A)}$$

i.e., $P(A) = 1.25 \cdot P(B) - 0.25$

upon simplification. Since the left-hand sides of these two equations are identical, it follows that the right-hand sides are equal, i.e., $1.25 \cdot P(B) - 0.25 = 0.75 \cdot P(B)$, and solving yields $P(B) = 0.5$. Hence, there is a 50% probability that any students come to the office hour.

Plugging this value back into either one of these equations yields $P(A) = 0.375$. Hence, there is a 37.5% probability that any students arrive within the first fifteen minutes of the office hour.
5.

<table>
<thead>
<tr>
<th>Cancer stage (A)</th>
<th>Income Level (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low (1)</td>
</tr>
<tr>
<td>1</td>
<td>0.05</td>
</tr>
<tr>
<td>2</td>
<td>0.10</td>
</tr>
<tr>
<td>3</td>
<td>0.15</td>
</tr>
<tr>
<td>4</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
</tr>
</tbody>
</table>

(a) Recall that one definition of statistical independence of A and B is \( P(A \cap B) = P(A) \cdot P(B) \). In particular then, the first cell entry \( P(\text{“A = 1”} \cap \text{“B = 1”}) = P(A = 1) \times P(B = 1) = (0.1)(0.5) = 0.05 \), i.e., the product of the \( i^{th} \) column marginal times \( i^{th} \) first row marginal. In a similar fashion, the cell value in the intersection of the \( i^{th} \) row (\( i = 1, 2, 3 \)) and \( j^{th} \) column (\( j = 1, 2, 3, 4 \)) is equal to the product of the \( i^{th} \) row marginal probability, times the \( j^{th} \) column marginal probability, which allows us to complete the entire table easily, as shown. By definition, this property is only true for independent events (!!!), and is fundamental to the derivation of the “expected value” formulas used in the “Chi-squared Test” (sections 6.2.3 and 6.3.1).

(b) By construction, we have

\[
\begin{align*}
\pi_{1|1} &= 0.05 / 0.1 = 0.5, \quad \pi_{1|2} = 0.10 / 0.2 = 0.5, \quad \pi_{1|3} = 0.15 / 0.3 = 0.5, \quad \pi_{1|4} = 0.20 / 0.4 = 0.5 \ldots \text{and } P(\text{Low}) = 0.5 \\
\pi_{2|1} &= 0.03 / 0.1 = 0.3, \quad \pi_{2|2} = 0.06 / 0.2 = 0.3, \quad \pi_{2|3} = 0.09 / 0.3 = 0.3, \quad \pi_{2|4} = 0.12 / 0.4 = 0.3 \ldots \text{and } P(\text{Mid}) = 0.3 \\
\pi_{3|1} &= 0.02 / 0.1 = 0.2, \quad \pi_{3|2} = 0.04 / 0.2 = 0.2, \quad \pi_{3|3} = 0.06 / 0.3 = 0.2, \quad \pi_{3|4} = 0.08 / 0.4 = 0.2 \ldots \text{and } P(\text{High}) = 0.2
\end{align*}
\]

(c) Also,

\[
\begin{align*}
\text{… and } P(\text{Stage 1}) &= 0.1 \\
\pi_{2|1} &= 0.10 / 0.5 = 0.2, \quad \pi_{2|2} = 0.06 / 0.3 = 0.2, \quad \pi_{2|3} = 0.09 / 0.3 = 0.3, \quad \pi_{2|4} = 0.12 / 0.4 = 0.3 \ldots \text{and } P(\text{Stage 2}) = 0.2 \\
\pi_{3|1} &= 0.15 / 0.5 = 0.3, \quad \pi_{3|2} = 0.06 / 0.2 = 0.3, \quad \pi_{3|3} = 0.06 / 0.2 = 0.3, \quad \pi_{3|4} = 0.08 / 0.2 = 0.4 \ldots \text{and } P(\text{Stage 3}) = 0.3 \\
\pi_{4|1} &= 0.20 / 0.5 = 0.4, \quad \pi_{4|2} = 0.12 / 0.3 = 0.4, \quad \pi_{4|3} = 0.08 / 0.2 = 0.4, \quad \pi_{4|4} = 0.06 / 0.2 = 0.3 \ldots \text{and } P(\text{Stage 4}) = 0.4
\end{align*}
\]

(d) It was shown in the “Lung cancer” versus “Coffee drinker” example that these two events are independent in the study population; the 2 \times 2 table is reproduced below.

The probability in the first cell (“Yes” for both events), 0.06, is indeed equal to \((0.40)(0.15)\), the product of its row and column marginal sums (i.e., “Yes” for one event, times “Yes” for the other event), and likewise for the probabilities in all the other cells.

Note that this is not true of the 2 \times 2 “Lung Cancer” versus “Smoking” table.
6. The given information can be written as conditional probabilities:

\[ P( A \mid B ) = 0.8, \quad P( B \mid A ) = 0.9, \quad P( B^c \mid A^c ) = 0.85 \]

We are asked to find the value of \( P( A^c \mid B^c ) \). First, let \( P( A ) = a \), \( P( B ) = b \), and \( P( A \cap B ) = c \). Then all of the events in the table can be labeled as shown. Using the definition of conditional probability \( P( E \mid F ) = P( E \cap F ) / P( F ) \), we have

\[ \frac{c}{b} = 0.8, \quad \frac{c}{a} = 0.9, \quad \frac{1-a-b+c}{1-a} = 0.85. \]

Algebraically solving these three equations with three unknowns yields \( a = 0.40 \), \( b = 0.45 \), \( c = 0.36 \), as shown.

Therefore, \( P( A^c \mid B^c ) = \frac{P( A^c \cap B^c )}{P( B^c )} = \frac{0.51}{0.55} = 0.927. \)

7. Let events \( A \), \( B \), and \( C \) represent the occurrence of each symptom, respectively. The given information can be written as:

- \( P(A) = P(B) = P(C) = 0.6 \)
- \( P(A \cap B \mid C) = 0.45 \), and similarly, \( P(A \cap C \mid B) = 0.45 \), \( P(B \cap C \mid A) = 0.45 \) as well.
- \( P(A \mid B \cap C) = 0.75 \), and similarly, \( P(B \mid A \cap C) = 0.75 \), \( P(C \mid A \cap B) = 0.75 \) as well.

(a) We are asked to find \( P(A \cap B \cap C) \). It follows from the definition of conditional probability that \( P(A \cap B \cap C) = P(A \cap B \mid C) \times P(C) \) which, via the first two statements above, results in \( = (0.45)(0.6) = 0.27 \). (The two other equations yield the same value.)

(b) Again, via conditional probability, we have

\[ P(A \cap B \cap C) = P(A \mid B \cap C) \times P(B \cap C) \]

which, via the third statement above and part (a), can be written as \( 0.27 = 0.75 \times P(B \cap C) \), so that \( P(B \cap C) = 0.36 \). So \( P(A^c \cap B \cap C) = 0.36 - 0.27 = 0.09 \), and likewise for the others, \( P(A \cap B^c \cap C) \) and \( P(A \cap B \cap C^c) \).

(See Venn diagram.) Hence, \( P(\text{Two or three}) = (3 \times 0.09) + 0.27 = 0.54 \).

(c) From (b), \( P(\text{Exactly two}) = (3 \times 0.09) = 0.27 \).

(d) From (a) and (c), it follows that \( P(A \cap B^c \cap C^c) = 0.6 - (0.27 + 0.09 + 0.09) = 0.15 \), and likewise for the others, \( P(A^c \cap B^c \cap C) \) and \( P(A^c \cap B \cap C^c) \). Hence \( 3 \times 0.15 = 0.45 \).

(e) From (b), (c), and (d), we see that \( P(A \cup B \cap C) = 0.27 + 3(0.9) + 3(0.15) = 0.99 \), so that \( P(A^c \cap B^c \cap C^c) = 1 - 0.99 = 0.01 \). (See Venn diagram.)

(f) Working with \( A \) and \( B \) for example, we have \( P(A) = P(B) = 0.6 \) from the given, and \( P(A \cap B) = 0.36 \) from part (b). Since it is true that \( 0.36 = 0.6 \times 0.6 \), it does indeed follow that \( P(A \cap B) = P(A) \times P(B) \), i.e., events \( A \) and \( B \) are statistically independent.
8. With events $A =$ Accident, $B =$ Berkeley visited, and $C =$ Chelsea visited, the given statements can be translated into mathematical notation as follows:

i. $P(B \cap C) = P(B)P(C)$

ii. $P(B) = .80$

iii. $P(C) = .75$

Therefore, substituting ii and iii into i yields $P(B \cap C) = (.8)(.75)$, i.e., $P(B \cap C) = .60$. (purple + gray)

Furthermore, it also follows from statistical independence that

- $P(B$ only$) = P(B \cap C^c) = (.8)(1-.75)$, i.e., $P(B \cap C^c) = .20$ (blue + green)
- $P(C$ only$) = P(B^c \cap C) = (1-.8)(.75)$, i.e., $P(B^c \cap C) = .15$ (red + orange)
- $P($Neither $B$ nor $C$) = P($B^c \cap C^c$) = $(1-.8)(1-.75)$, i.e., $P(B^c \cap C^c) = .05$ (yellow + white)

iv. $P(A \mid B \cap C) = .90$, which implies $P(A \cap B \cap C) = P(A \mid B \cap C)P(B \cap C) = (.9)(.6)$, i.e., $P(A \cap B \cap C) = .54$, hence $P(A^c \cap B \cap C) = .06$.

v. $P(A \mid B^c \cap C) = .35$, which implies $P(A \cap B^c \cap C) = P(A \mid B^c \cap C)P(B^c \cap C) = (.35)(.2)$, i.e., $P(A \cap B^c \cap C) = .07$, hence $P(A^c \cap B^c \cap C) = .13$.

vi. $P(A \mid B \cap C^c) = .20$, which implies $P(A \cap B \cap C^c) = P(A \mid B \cap C^c)P(B \cap C^c) = (.2)(.15)$, i.e., $P(A \cap B \cap C^c) = .03$, hence $P(A^c \cap B \cap C^c) = .12$.

vii. $P(A \mid B^c \cap C^c) = .02$, which implies $P(A \cap B^c \cap C^c) = P(A \mid B^c \cap C^c)P(B^c \cap C^c) = (.02)(.05)$, i.e., $P(A \cap B^c \cap C^c) = .01$, hence $P(A^c \cap B^c \cap C^c) = .049$. 

![Venn Diagram](attachment://venn_diagram.png)
9. The given information tells us the following.

(i) \( P(A \cup B) = .99 \)
(ii) \( P(B \mid A) = .60 \), which implies that \( P(B \cap A) = .6 \, P(A) \)
(iii) \( P(A \mid B) = .75 \), which implies that \( P(A \cap B) = .75 \, P(B) \)

Because the left-hand sides of (ii) and (iii) are the same, it follows that \( .6 \, P(A) = .75 \, P(B) \), or
(iv) \( P(B) = .8 \, P(A) \).

Now, substituting (ii) and (iv) into the general relation \( P(A \cup B) = P(A) + P(B) - P(A \cap B) \) gives

\[
.99 = P(A) + .8 \, P(A) - .6 \, P(A),
\]

or \( .99 = 1.2 \, P(A) \), i.e., \( P(A) = .825 \). Thus, \( P(B) = .66 \) via (iv), and \( P(B \cap A) = .495 \) via (ii). The two events \( A \) and \( B \) are certainly **not independent**, which can be seen any one of three ways:

\( P(A \mid B) = .75 \) from (iii), is not equal to \( P(A) = .825 \) just found;

\( P(B \mid A) = .60 \) from (ii), is not equal to \( P(B) = .66 \) just found;

\( P(A \cap B) = .495 \) is not equal to \( P(A) \times P(B) = .825 \times .66 = .5445 \).
10. **Switch!** It is tempting to believe that it makes no difference, since once a zonk door has been opened and supposedly ruled out, the probability of winning the car should then be *equally likely* (i.e., 1/2) between each of the two doors remaining. However, it is important to remember that the host does not eliminate one of the original three doors *at random*, but always – i.e., “with probability 1” – a door other than the one chosen, and *known (to him) to contain a zonk*. Rather than discarding it, this nonrandom choice conveys useful information, namely, if indeed that had been the door originally chosen, then not switching would certainly have resulted in losing. As exactly one of the other doors also contains a zonk, the same argument can be applied to that door as well, whichever it is. Thus, as it would only succeed if the winning door was chosen, the strategy of **not switching would result in losing two out of three times**, on average.

This very surprising and counterintuitive result can be represented via the following table. Suppose that, for the sake of argument, Door 1 contains the car, and Doors 2 and 3 contain goats, as shown.

<table>
<thead>
<tr>
<th>If contestant chooses:</th>
<th>Door 1</th>
<th>Door 2</th>
<th>Door 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>then host reveals:</td>
<td>Door 2 or Door 3 <em>(at random)</em></td>
<td>Door 3 <em>(not at random)</em></td>
<td>Door 2 <em>(not at random)</em></td>
</tr>
<tr>
<td>Switch?</td>
<td>Yes</td>
<td>LOSE</td>
<td>WIN</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>WIN</td>
<td>LOSE</td>
</tr>
</tbody>
</table>

\[
P(\text{Win | Switch}) = \frac{2}{3} \quad P(\text{Lose | Switch}) = \frac{1}{3}
\]

\[
P(\text{Win | Stay}) = \frac{1}{3} \quad P(\text{Lose | Stay}) = \frac{2}{3}
\]

Much mathematical literature has been devoted to the Monty Hall Problem – which has a colorful history – and its numerous variations. In addition, many computer programs exist on the Internet (e.g., using Java applets), that numerically simulate the Monty Hall Problem, and in so doing, verify that the values above are indeed correct. Despite this however, many people (including more than a few professional mathematicians and statisticians) heatedly debate the solution in favor of the powerfully intuitive, but incorrect, “switching doesn’t matter” answer. Strange but true...
11. (a) We know that for any two events $E$ and $F$, \[ P(E \cup F) = P(E) + P(F) - P(E \cap F) \]. Hence, 
\[ P(A \cup B) = P(A) + P(B) - P(A \cap B), \] i.e., \[ 0.69 = P(A) + P(B) - 0.19, \] or \[ 0.88 = P(A) + P(B) \].
Likewise,
\[ P(A \cup C) = P(A) + P(C) - P(A \cap C), \] i.e., \[ 0.70 = P(A) + P(C) - 0.20, \] or \[ 0.90 = P(A) + P(C) \]
and
\[ P(B \cup C) = P(B) + P(C) - P(B \cap C), \] i.e., \[ 0.71 = P(B) + P(C) - 0.21, \] or \[ 0.92 = P(B) + P(C) \].
Solving these three simultaneous equations yields \[ P(A) = 0.43, \ P(B) = 0.45, \ P(C) = 0.47 \].

(b) Events $E$ and $F$ are **statistically independent** if \[ P(E \cap F) = P(E)P(F) \]. Hence,
\[ P(A \cap C \cap B) = P(A \cap C \cap B) = (0.20)(0.45), \] i.e., \[ P(A \cap B \cap C) = 0.09 \], from which the entire Venn diagram can be reconstructed from the triple intersection out, using the information above.
12. Odds Ratio and Relative Risk

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Diseased ((D^+))</th>
<th>Nondiseased ((D^-))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed ((E^+))</td>
<td>(p_{11})</td>
<td>(p_{12})</td>
</tr>
<tr>
<td>Unexposed ((E^-))</td>
<td>(p_{21})</td>
<td>(p_{22})</td>
</tr>
<tr>
<td>(p_{11} + p_{21})</td>
<td>(p_{12} + p_{22})</td>
<td>1</td>
</tr>
</tbody>
</table>

In a **cohort study** design...

\[
OR = \frac{\text{odds of disease, given exposure}}{\text{odds of disease, given no exposure}} = \frac{P(D^+ \mid E^+)}{P(D^- \mid E^-)} = \frac{p_{11} / p_{12}}{p_{21} / p_{22}} = \frac{p_{11} p_{22}}{p_{12} p_{21}}.
\]

In a **case-control study** design...

\[
OR = \frac{\text{odds of exposure, given disease}}{\text{odds of exposure, given no disease}} = \frac{P(E^+ \mid D^+)}{P(E^- \mid D^-)} = \frac{p_{11} / p_{21}}{p_{12} / p_{22}} = \frac{p_{11} p_{22}}{p_{12} p_{21}}.
\]

Both of these quantities agree, so the **odds ratio** can be used in either type of **longitudinal study**, although the interpretation must be adjusted accordingly. This is not true of the **relative risk**, which is only defined for cohort studies. (However, it is possible to estimate it using Bayes’ Law, provided one has an accurate estimate of the disease prevalence.)
13. Following the hint, we have

\[
\begin{array}{c|c|c}
  & x & y \\
  \hline
  .10 & .25 & \hline
  .20 & \\
\end{array}
\]

from which we can derive two equations in the two unknowns \( x \) and \( y \). The fact that the column marginal probabilities must sum to 1 yields the first equation \( x + .25 + y = 1 \), i.e.,

\[
x + y = .75.
\]

Moreover, we have statistical independence between the rows and columns. Therefore,

- the product of the first column marginal \( x \), times the first row marginal, equals .10, and
- the product of the third column marginal \( y \), times the second row marginal, equals .20.

Thus the row marginals must be as shown below.

\[
\begin{array}{c|c|c|c}
  & x & .25 & y \\
  \hline
  .10 & \hline
  .20 & .20/y & 1 \\
\end{array}
\]

This yields our second equation, \( \frac{.10}{x} + \frac{.20}{y} = 1 \), i.e.,

\[
\frac{1}{x} + \frac{2}{y} = 10.
\]

Substituting \( y = .75 - x \) from the first equation into the second yields

\[
x^2 - .65x + .075 = 0.
\]

when simplified. This factors: \( (x - 0.5)(x - 0.15) = 0 \), so that either \( x = 0.50 \) or \( x = 0.15 \), from which it follows that either \( y = 0.25 \) or \( y = 0.60 \), respectively. Thus, the two solutions are:

\[
\begin{array}{ccccccccc}
  .10 & .05 & .05 & .20 & .10 & .16677 & .40 & .66667 \\
  .40 & .20 & .20 & .80 & .05 & .08333 & .20 & .33333 \\
  .50 & .25 & .25 & 1 & .15 & .25 & .60 & 1 \\
\end{array}
\]

14. Under construction…
15. Events: $A = \text{Aspirin use}$, $B_1 = \text{GI bleeding}$, $B_2 = \text{Primary stroke}$, $B_3 = \text{CVD}$

**Prior probabilities:**

\[
P(B_1) = 0.2, \quad P(B_2) = 0.3 \quad P(B_3) = 0.5
\]

Conditional probabilities:

\[
P(A \mid B_1) = 0.09, \quad P(A \mid B_2) = 0.04 \quad P(A \mid B_3) = 0.02
\]

(a) Therefore, the *posterior probabilities* are, respectively,

\[
P(B_1 \mid A) = \frac{(0.09)(0.2)}{(0.09)(0.2) + (0.04)(0.3) + (0.02)(0.5)} = \frac{0.018}{0.040} = 0.45
\]

\[
P(B_2 \mid A) = \frac{(0.04)(0.3)}{(0.09)(0.2) + (0.04)(0.3) + (0.02)(0.5)} = \frac{0.012}{0.040} = 0.30
\]

\[
P(B_3 \mid A) = \frac{(0.02)(0.5)}{(0.09)(0.2) + (0.04)(0.3) + (0.02)(0.5)} = \frac{0.010}{0.040} = 0.25
\]

(b) The probability of gastrointestinal bleeding ($B_1$) increases from 20% to 45%, in the event of aspirin use ($A$); similarly, the probability of primary stroke ($B_2$) remains constant at 30%, and the probability of cardiovascular disease ($B_3$) decreases from 50% to 25%, in the event of aspirin use. Therefore, although it occurs the least often among the three given vascular conditions, gastrointestinal bleeding occurs in the highest overall proportion among the patients who used aspirin in this study. Furthermore, although it occurs the most often among the three conditions, cardiovascular disease occurs in the lowest overall proportion among the patients who used aspirin in this study, suggesting a protective effect. Lastly, as the prior probability $P(B_2)$ and posterior probability $P(B_2 \mid A)$ are equal (0.30), the two corresponding events “Aspirin use” and “Primary stroke” are statistically independent. Hence, the event that a patient has a primary stroke conveys no information about aspirin use, and vice versa (although aspirin does have a protective effect against secondary stroke). The following Venn diagram shows the relations among these events, drawn approximately to scale.

![Venn Diagram](image-url)
16. Events:  \( S \) = “Five year survival”  \( \Rightarrow \)  \( S^c \) = “Death within five years”  
\( T \) = “Treatment”  \( \Rightarrow \)  \( T^c \) = “No Treatment”

Prior probability \( P(S) = 0.4 \) \( \Rightarrow \)  \( P(S^c) = 1 - P(S) = 0.6 \)

Given

Conditional probability \( P(T \mid S) = 0.8 \) \( \Rightarrow \)  \( P(T^c \mid S) = 1 - P(T \mid S) = 0.2 \)

Conditional probability \( P(T \mid S^c) = 0.3 \) \( \Rightarrow \)  \( P(T^c \mid S^c) = 1 - P(T \mid S^c) = 0.7 \)

Posterior probabilities (via Bayes’ Formula):

(a) \( P(S \mid T) = \frac{P(T \mid S) P(S)}{P(T \mid S) P(S) + P(T \mid S^c) P(S^c)} \)

\[ \begin{align*}
(0.8)(0.4) & \quad = 0.32 \\
(0.8)(0.4) + (0.3)(0.6) & \quad = 0.50 = \boxed{0.64}
\end{align*} \]

\( P(S \mid T^c) = \frac{P(T^c \mid S) P(S)}{P(T^c \mid S) P(S) + P(T^c \mid S^c) P(S^c)} \)

\[ \begin{align*}
(0.2)(0.4) & \quad = 0.08 \\
(0.2)(0.4) + (0.7)(0.6) & \quad = 0.50 = \boxed{0.16}
\end{align*} \]

Given treatment \( (T) \), the probability of five-year survival \( (S) \) increases from a prior of 0.40 to a posterior of 0.64. Moreover, given no treatment \( (T^c) \), the probability of five-year survival \( (S) \) decreases from a prior of 0.40 to a posterior of 0.16. Hence, in this population, treatment is associated with a four-fold increase in the probability of five-year survival. (This is the relative risk.) Note, however, that this alone may not be enough to recommend treatment. Other factors, such as adverse side effects and quality of life issues, are legitimate patient concerns to be decided individually.

(b) Odds of survival, given treatment \( \frac{P(S \mid T)}{P(S^c \mid T)} = \frac{0.64}{1 - 0.64} = \boxed{1.778} \)

Odds of survival, given no treatment \( \frac{P(S \mid T^c)}{P(S^c \mid T^c)} = \frac{0.16}{1 - 0.16} = \boxed{0.190} \)

\[ \therefore \text{Odds Ratio} = \frac{1.778}{0.190} = \boxed{9.33} \]

“The odds of survival given treatment are 9.33 times greater than the odds of survival given no treatment.”
17. Let $P(A) = a$, $P(B) = b$, $P(A \cap B) = c$, as shown. Then it follows that

(1) \[ 0 \leq c \leq a \leq 1 \quad \text{and} \quad 0 \leq c \leq b \leq 1 \]
as well as

(2) \[ 0 \leq a + b - c \leq 1. \]

Therefore, $\Delta = |c - ab|$ in this notation. It thus suffices to show that

\[ \frac{-1}{4} \leq ab - c \leq \frac{1}{4}. \]

From (2), we see that

\[ ab - c \leq a(1 - a + c) - c = a - a^2 - (1 - a)c \leq a - a^2 \leq \frac{1}{4}. \]

Clearly, this inequality is sharp when $c = 0$ and $a = 1/2$, i.e., when $P(A \cap B) = 0$ (e.g., $A$ and $B$ are disjoint) and $P(A) = 1/2$. Moreover, because the definition of $\Delta$ is symmetric in $A$ and $B$, it must also follow that $P(B) = 1/2$. (See first figure below.) Furthermore, from (1),

\[ ab - c \geq (c)(c) - c = c^2 - c \geq -\frac{1}{4}. \]

This inequality is sharp when $a = b = c = 1/2$, i.e., when $P(A) = P(B) = P(A \cap B) = 1/2$, which implies that $A = B$, both having probability 1/2. (See second figure.)
18.  

<table>
<thead>
<tr>
<th>Treatment A</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment B</td>
<td>Yes</td>
<td>0.14</td>
</tr>
<tr>
<td>No</td>
<td>0.21</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>0.35</td>
<td>0.65</td>
</tr>
</tbody>
</table>

(a) Given: \( P(A) = .35, \ P(B) = .40, \ P(A \cap B) = .14 \)

Then \( P(\text{A only}) = P(A \cap B^c) = .35 - .14 = .21 \), \( P(\text{B only}) = P(A^c \cap B) = .40 - .14 = .26 \), and \( P(\text{Neither}) = P(A^c \cap B^c) = 1 - (.21 + .14 + .26) = 1 - .61 = .39 \), as shown in the first Venn diagram above. Since \( P(A \cap B) = .14 \) and \( P(A) P(B) = (.35)(.40) = .14 \) as well, it follows that the two treatments are indeed statistically independent in this population.

\[ P(\text{A or B}) = .61 \text{ (calculated above)} \]

(b) Given: \( P(\text{A only}) = .35, \ P(\text{B only}) = .40, \ P(A \cap B) = .14 \)

Then \( P(\text{Neither}) = P(A^c \cap B^c) = 1 - (.35 + .14 + .40) = 1 - .89 = .11 \), as shown in the second Venn diagram above. Since \( P(A \cap B) = .14 \) and \( P(A) P(B) = (.49)(.54) = .2646 \), it follows that the two treatments are not statistically independent in this population.

\[ P(\text{A or B}) = .89 \text{ (calculated above)} \]

\[ P(\text{A xor B}) = .35 + .40, \text{ or } .89 - .14, = .75 \]
19. Let events $A = \text{Adult}$, $B = \text{Male}$, $C = \text{White}$. We are told that

\[(1) \quad P(A \cap B \mid C) = 0.3, \quad \text{i.e.,} \quad \frac{P(A \cap B \cap C)}{P(C)} = 0.3, \quad \text{so that} \quad P(A \cap B \cap C) = 0.3 P(C), \]

\[(2) \quad P(A \cap C \mid B) = 0.4, \quad \text{i.e.,} \quad \frac{P(A \cap C \cap B)}{P(B)} = 0.4, \quad \text{so that} \quad P(A \cap B \cap C) = 0.4 P(B), \]

and finally,

\[(3) \quad P(A \mid B \cap C) = 0.5, \quad \text{i.e.,} \quad \frac{P(A \cap B \cap C)}{P(B \cap C)} = 0.5, \quad \text{so that} \quad P(A \cap B \cap C) = 0.5 P(B \cap C). \]

Since the left-hand sides of all three equations are the same, it follows that all the right-hand sides are equal as well.

(a) Therefore, equating (1) and (3) yields

\[0.5 \frac{P(B \cap C)}{P(C)} = 0.3, \quad \text{i.e.,} \quad \frac{P(B \cap C)}{P(C)} = \frac{0.3}{0.5}, \quad \text{or by definition,} \quad P(B \mid C) = 0.6, \quad \text{i.e.,} \quad 60\%. \]

and

(b) equating (2) and (3) yields

\[0.5 \frac{P(B \cap C)}{P(B)} = 0.4, \quad \text{i.e.,} \quad \frac{P(B \cap C)}{P(B)} = \frac{0.4}{0.5}, \quad \text{or by definition,} \quad P(C \mid B) = 0.8, \quad \text{i.e.,} \quad 80\%. \]

20. Again, let events $A = \text{Adult}$, $B = \text{Male}$, $C = \text{White}$. We are here told that

- $P(B \mid A) = .1$, $P(C \mid B) = .2$, $P(A \mid C) = .3$
- $P(A \mid B) = .4$, $P(B \mid C) = .5$, $P(C \mid A) = ?$

However, it is true that

\[P(A \mid B) \times P(B \mid C) \times P(C \mid A) = P(B \mid A) \times P(C \mid B) \times P(A \mid C) \]

because

\[\frac{P(A \cap B)}{P(B)} \times \frac{P(B \cap C)}{P(C)} \times \frac{P(C \cap A)}{P(A)} = \frac{P(B \cap A)}{P(A)} \times \frac{P(C \cap B)}{P(B)} \times \frac{P(A \cap C)}{P(C)}, \]

since the numerators of each side are simply rearrangements of one another, as likewise are the denominators. Therefore,

\[0.4 \times 0.5 \times P(C \mid A) = 0.1 \times 0.2 \times 0.3,\]

i.e., $P(C \mid A) = 0.03$, or $3\%$. 

21. **The Shell Game**

(a) With 20 shells, the probability of winning exactly one game is 1/20, or .05; therefore, the probability of losing exactly one game is .95. Thus (reasonably assuming independence between game outcomes), the probability of losing all \( n \) games is equal to \(.95^n\), from which it follows that the probability of not losing all \( n \) games – i.e., \( P(\text{winning at least one game}) \) – is equal to \( 1 - (.95^n) \).

In order for this probability to be greater than \(.5 \) – i.e., \( 1 - (.95^n) > .5 \) – it must be true that \(.95^n < .5 \), or \( n > \frac{\log (.5)}{\log (.95)} = 13.51 \), so \( n \geq 14 \) games. As \( n \to \infty \), it follows that \(.95^n \to 0 \), so that \( P(\text{win at least one game}) = 1 - (.95)^n \to 1 \) (“certainty”).

(b) Using the same logic as above with \( n \) shells, the probability of winning exactly one game is \( \frac{1}{n} \); therefore, the probability of losing exactly one game is \( 1 - \frac{1}{n} \). Thus (again, tacitly assuming independence between game outcomes), the probability of losing all \( n \) games is equal to \( \left(1 - \frac{1}{n}\right)^n \), from which it follows that the probability of not losing all \( n \) games – i.e., \( P(\text{win at least one game}) \) – is equal to \( 1 - \left(1 - \frac{1}{n}\right)^n \), which approaches \( 1 - e^{-1} = .632 \) as \( n \to \infty \).

22. In progress...

23. Recall that \( RR = \frac{p}{q} \) and \( OR = \frac{p/(1-p)}{q/(1-q)} = \frac{p(1-q)}{q(1-p)} \), with \( p = P(D+|E+) \) and \( q = P(D+|E-) \).

The case \( RR = 1 \) is trivial, for then \( p = q \), hence \( OR = 1 \) as well; this corresponds to the case of no association.

Suppose \( RR > 1 \). Then \( p > q \), which implies \( 1 < \frac{1-q}{1-p} \), or \( \frac{p}{q} < \frac{p(1-q)}{q(1-p)} \), i.e., \( RR < OR \).

Thus we have \( 1 < RR < OR \). ✓ For the case \( RR < 1 \), simply reverse all the inequalities. □

24. (a) Let event \( A \) = “perfect square” = \( \{1^2, 2^2, 3^2, \ldots, (10^6)^2\} \); then \( P(A) = 10^6/10^{12} = 10^{-6} \).

Likewise, let \( B \) = “perfect cube” = \( \{1^3, 2^3, 3^3, \ldots, (10^4)^3\} \); then \( P(B) = 10^4/10^{12} = 10^{-8} \).

Thus, \( A \cap B \) = “perfect sixth” = \( \{1^6, 2^6, 3^6, \ldots, (10^2)^6\} \); hence \( P(A \cap B) = 10^2/10^{12} = 10^{-10} \).

Therefore, \( P(A \cup B) = P(A) + P(B) - P(A \cap B) = 10^{-6} + 10^{-8} - 10^{-10} \).

(b) Let event \( C \) = “perfect fourth” = \( \{1^4, 2^4, 3^4, \ldots, (10^3)^4\} \); then \( P(C) = 10^3/10^{12} = 10^{-9} \).

Likewise, let \( D \) = “perfect sixth” = \( \{1^6, 2^6, 3^6, \ldots, (10^2)^6\} \); then \( P(D) = 10^2/10^{12} = 10^{-10} \).

Thus, \( C \cap D \) = “perfect twelfth” = \( \{1^{12}, 2^{12}, 3^{12}, \ldots, 10^{12}\} \); hence \( P(C \cap D) = 10/10^{12} = 10^{-11} \).

Therefore, \( P(C \cup D) = P(C) + P(D) - P(C \cap D) = 10^{-9} + 10^{-10} - 10^{-11} \).
25. We construct a counterexample as follows. Let $0 < \varepsilon < 1$ be a fixed but arbitrary number. Also, with no loss of generality, assume the first toss results in 1 (i.e., Heads), as shown. Now suppose that the next $n_0 - 1$ tosses all result in 0 (i.e., Tails), where $n_0 > \frac{1}{\varepsilon}$. It then follows that the proportion of Heads in the first $n_0$ tosses is $\left(\frac{1}{n_0} < \varepsilon\right)$, i.e., *arbitrarily close to 0*. Now suppose that the next $n_1 - n_0$ tosses all result in 1 (i.e., Heads), where $n_1 > \frac{-1+n_0}{\varepsilon}$. It then follows that the proportion of Heads in the first $n_1$ tosses is $\left(\frac{1-n_0+n_1}{n_1} > 1-\varepsilon\right)$, i.e., *arbitrarily close to 1*. By continuing to attach sufficiently large blocks of zeros and ones in this manner – i.e., $n_2 > \frac{1-n_0+n_1}{\varepsilon}$, $n_3 > \frac{-1+n_0-n_1+n_2}{\varepsilon}$, … – an infinite sequence is generated that does not converge, but forever oscillates between values which come arbitrarily close to 0 and 1, respectively.

<table>
<thead>
<tr>
<th>1 toss</th>
<th>$n_0$ tosses</th>
<th>$n_1$ tosses</th>
<th>$n_2$ tosses</th>
<th>$n_3$ tosses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$n_0 - 1$</td>
<td>$n_1 - n_0$</td>
<td>$n_2 - n_1$</td>
<td>$n_3 - n_2$</td>
</tr>
<tr>
<td>1</td>
<td>000...0</td>
<td>111........1</td>
<td>0000000............0</td>
<td>111111111...........1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># Heads:</th>
<th>$X_0 = 1$</th>
<th>$X_1 = 1-n_0+n_1$</th>
<th>$X_2 = 1-n_0+n_1$</th>
<th>$X_3 = 1-n_0+n_1-n_2+n_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of Heads</td>
<td>$\frac{1}{n_0} &lt; \varepsilon$</td>
<td>$\frac{1-n_0+n_1}{n_1} &gt; 1-\varepsilon$</td>
<td>$\frac{1-n_0+n_1}{n_2} &lt; \varepsilon$</td>
<td>$\frac{1-n_0+n_1-n_2+n_3}{n_3} &gt; 1-\varepsilon$</td>
</tr>
</tbody>
</table>

**Exercise:** Prove that $n_k > \max\left\{n_{k-1}, \frac{(1-\varepsilon)^k}{\varepsilon}\right\}$ for $k = 0, 1, 2, \ldots$

**Hint:** By construction, $n_k > \frac{n_{k-1}-n_{k-2}+n_{k-3}-\ldots+(-1)^{k-1}}{\varepsilon}$. From this, show that $n_{k+1} > \left(\frac{1-\varepsilon}{\varepsilon}\right)n_k$. 
26. 

(a) From the given, the first column marginal probability \( P(A) = 0.60 \), and the upper right cell (NE corner) is \( P(A^c \cap B) = 0.30 \), as shown below. From \( P(A) = 0.60 \), it follows that the second column marginal has complement probability \( P(A^c) = 1 - 0.60 = 0.40 \). Therefore, via independence, the first row marginal probability must be \( P(B) = 0.30/0.40 \), or \( 0.75 \), which is the answer. However, if desired, the rest of the table can also be determined, since the second row marginal probability must be \( P(B^c) = 1 - 0.75 = 0.25 \). The three remaining values can be obtained by subtraction.

\[
\begin{array}{ccc}
A & A^c \\
B & 0.45 & 0.30 & 0.75 \\
B^c & 0.15 & 0.10 & 0.25 \\
0.60 & 0.40 & 1 \\
\end{array}
\]

(b) From the given, the upper left cell (NW corner) has probability \( P(C \cap D) = 0.72 \), and the lower right cell (SE corner) has probability \( P(C^c \cap D^c) = 0.02 \), as shown below in the first table. Now let \( c = P(C) \) and \( d = P(D) \), and thus \( 1-c \) and \( 1-d \) are their respective complement probabilities. Then via independence, we have the two equations \( cd = 0.72 \) and \( (1-c)(1-d) = 0.02 \). Solving these simultaneously yields two possible solutions: either \( c = 0.9, d = 0.8 \), or \( c = 0.8, d = 0.9 \). These then result in the two remaining tables below, respectively. Note they are transposes of one another, i.e., the rows of one are just the columns of the other. (This is because of symmetry; events \( C \) and \( D \) are interchangeable in the original question.)

\[
\begin{array}{ccc}
C & C^c \\
D & 0.72 & 0.08 & 0.80 \\
D^c & 0.02 & 1-d \\
C & 0.90 & 0.10 & 1 \\
C^c & 0.80 & 0.20 & 1 \\
\end{array}
\]
27. Label the empty cells as shown.

<table>
<thead>
<tr>
<th></th>
<th>.01</th>
<th>x</th>
<th>.02</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>?</td>
<td>z</td>
<td>.50</td>
</tr>
<tr>
<td>.03</td>
<td>w</td>
<td>.04</td>
<td></td>
</tr>
<tr>
<td>.60</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It then follows that:

1. \(0.01 + x + 0.02 + y + ? + z + 0.03 + w + 0.04 = 1\), i.e., \(x + y + ? + z + w = 0.90\)
2. \(x + ? + w = 0.60\)
3. \(y + ? + z = 0.50\)

Adding equations (2) and (3) together yields \(x + y + 2? + z + w = 1.10\). Subtracting equation (1) from this yields \(? = 0.20\).

28. In progress…

29. Careful calculation shows that \(P(A) = a\), \(P(B) = b\), \(P(C) = c\), and \(P(A \cap B) = ab\), \(P(A \cap C) = ac\), \(P(B \cap C) = bc\), so that the events are indeed pairwise independent. However, the triple intersection \(P(A \cap B \cap C) = d\), an arbitrary value. Thus \(P(A \cap B \cap C) \neq P(A)P(B)P(C)\), unless \(d = abc\). In that case, the Venn diagram simplifies to the following unsurprising form.
Bar Bet

(a) **Absolutely not!** To see why, let us start with the simpler scenario of drawing four cards from a fair deck, *with replacement*. In this case, all cards have an *equal likelihood* of being selected (namely, $1/52$). *This being the case,* and the fact that there are 12 face cards in a standard deck, it follows that the probability of selecting a face card is $12/52$, *and* the outcome of any selection is *statistically independent* of any other selection. To calculate the probability of at least one face card, we can subtract the probability of the complement – no face cards – from 1. That is, $1 – \text{the probability of picking 4 non-face cards: } 1 – (40/52)^4 = 0.65$.

Now suppose we modify the scenario to selecting $n = 4$ cards, *without replacement*. Unlike the above, the probability of selecting a face card now *changes* with every draw, making the outcomes *statistically dependent*. Since the number of cards decreases by one with each draw, the probability of picking all 4 non-face cards is no longer simply $(40/52)^4 = (40/52)(40/52)(40/52)(40/52)$, but $(40/52)(39/51)(38/50)(37/49)$.*

Therefore, the probability of picking $\geq 1$ face card $= 1 – (40/52)(39/51)(38/50)(37/49) = 0.6624$. This means that I will win the bet approximately 2 out of 3 times! Counterintuitive perhaps, but true nonetheless.

(b) **No**, you should still not take the bet. Using the same logic with $n = 3$ draws, the probability of picking at least one face card $= 1 – (40/52)(39/51)(38/50) = 0.5529$. Thus, I still enjoy about a 5+\% advantage over “even money” (i.e., 50\%). On average, I will win 11 out of every 20 games played, and make one dollar.

(c) The \texttt{R} simulation should be consistent with the result found in part (a), namely, that the proportion of wins $\approx 0.6624$, and therefore the proportion of losses $\approx 0.3376$.

---

*Note:* For those who remember “combinatorics,” another way to arrive at this value is the following: There are $\binom{52}{4}$ ways of randomly selecting 4 cards from the deck of 52. Of this number, there are $\binom{40}{4}$ ways of randomly selecting 4 non-face cards. The ratio of the two, $\frac{\binom{40}{4}}{\binom{52}{4}}$, yields the same value as the underlined four-factor product above. (This is an illustration of the so-called *Hypergeometric distribution*.)
31.  
(a) False; let $B = C$, for example. A nontrivial counterexample: see bottom left Venn diagram.
(b) False. See counterexample on bottom right.

32.  
(a) False. The same Venn diagram on the bottom left serves as a counterexample here too.
(b) True!  \textbf{Proof:}

Given that event $A$ is statistically independent of events $B$, $C$, and $B \cap C$. Therefore,

\begin{itemize}
  \item $P(A \cap B) = P(A) P(B)$
  \item $P(A \cap C) = P(A) P(C)$
  \item $P(A \cap (B \cap C)) = P(A) P(B \cap C)$
\end{itemize}

Now consider $P(A \cap (B \cup C))$. From the “distributive law” for sets, we have

$$P(A \cap (B \cup C)) = P(A \cap B) \cup (A \cap C)$$

$$= P(A \cap B) + P(A \cap C) - P(A \cap B \cap C)$$

$$= P(A) P(B) + P(A) P(C) - P(A) P(B \cap C)$$

$$= P(A) P(B) + P(C) - P(B \cap C)$$

$$= P(A) P(B \cup C),$$

i.e., $A$ is statistically independent of $B \cup C$.  QED

(c) True! The proof above just has to be modified a bit:

$$P(A \cap (B \cup C)) = P(A \cap B) \cup (A \cap C)$$

$$= P(A \cap B) + P(A \cap C) - \overline{P(A \cap B \cap C)}$$

$$= P(A) P(B) + P(A) P(C)$$

$$= P(A) \left[ P(B) + P(C) - P(B \cap C) \right]$$

$$= P(A) P(B \cup C).$$