Chapter 3 - Probability

SOLUTIONS ARE FOR PERSONAL USE ONLY. NOT NOT DISTRIBUTE OR SHARE.

Dice rolls. (3.6, p. 92) If you roll a pair of fair dice, what is the probability of

(a) getting a sum of 1?
(b) getting a sum of 5?
(c) getting a sum of 12?

Solution

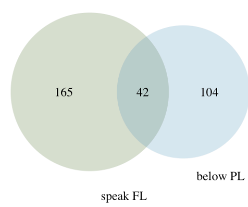
(a) P(sum of 1) = 0. Since there are two dice being rolled, the minimum possible sum is 2.
(b) P(sum of 5) = P(1,4) + P(2,3) + P(3,2) + P(4,1) = $\left(\frac{1}{6} \times \frac{1}{6}\right) \times 4 = \frac{4}{36} \approx 0.11$.
(c) P(sum of 12) = P(6,6) = $\left(\frac{1}{6} \times \frac{1}{6}\right) = \frac{1}{36} \approx 0.0278$.

Poverty and language. (3.8, p. 93) The American Community Survey is an ongoing survey that provides data every year to give communities the current information they need to plan investments and services. The 2010 American Community Survey estimates that 14.6% of Americans live below the poverty line, 20.7% speak a language other than English (foreign language) at home, and 4.2% fall into both categories.

(a) Are living below the poverty line and speaking a foreign language at home disjoint?
(b) Draw a Venn diagram summarizing the variables and their associated probabilities.
(c) What percent of Americans live below the poverty line and only speak English at home?
(d) What percent of Americans live below the poverty line or speak a foreign language at home?
(e) What percent of Americans live above the poverty line and only speak English at home?
(f) Is the event that someone lives below the poverty line independent of the event that the person speaks a foreign language at home?

Solution

(a) No, there are people who are both living below the poverty line and speak a language other than English at home.
(b) The Venn diagram is shown below.



(c) Each person living below the poverty line either speaks only English at home or doesn't. Since 14.6% of Americans live below the poverty line and 4.2% speak a language other than English at home, the other 10.4% only speak English at home.
(d) Using the General Addition Rule:

$$P(below\ PL\ or\ speak\ FL) = P(below\ PL) + P(speak\ FL) - P(both) \quad = 0.146 + 0.207 - 0.042 = 0.311$$

(e) P(neither below PL nor speak FL) = 1 - P(below PL or speak FL) = 1 - 0.311 = 0.689.
(f) Two approaches:

(1) Using the multiplication rule: $P(below\ PL) \times P(speak\ FL) = 0.146 \times 0.207 = 0.030$, which does not equal $P(below\ PL\ and\ speak\ FL) = 0.042$, therefore the events are dependent.

(2) Using Bayes' theorem: If the two events are independent, then $P(below\ PL \mid speak\ FL) = P(below\ PL)$. Using Bayes' theorem,

$$\begin{aligned} P(below\ PL \mid speak\ FL) &= \frac{P(below\ PL\ and\ speak\ FL)}{P(speak\ FL)} \\ &= \frac{0.042}{0.207} \approx 0.203 \end{aligned}$$

Since this probability is different than $P(below\ PL) = 0.146$, we determine that the two events are dependent.

Assortative mating. (3.18, p. 111) Assortative mating is a nonrandom mating pattern where individuals with similar genotypes and/or phenotypes mate with one another more frequently than what would be expected under a random mating pattern. Researchers studying this topic collected data on eye colors of 204 Scandinavian men and their female partners. The table below summarizes the results. For simplicity, we only include heterosexual relationships in this exercise.

		Partner (female)			
		Blue	Brown	Green	Total
Self (male)	Blue	78	23	13	114
	Brown	19	23	12	54
	Green	11	9	16	36
	Total	108	55	41	204

(a) What is the probability that a randomly chosen male respondent or his partner has blue eyes?
(b) What is the probability that a randomly chosen male respondent with blue eyes has a partner with blue eyes?
(c) Whatistheprobabilitythatarandomlychosenmalerespondentwithbrowneyeshasapartner with blue eyes? What about the probability of a randomly chosen male respondent with green eyes having a partner with blue eyes?
(d) Does it appear that the eye colors of male respondents and their partners are independent? Explain your reasoning.

Solution

(a) P(man or partner has blue eyes) = (108 + 114 - 78) / 204 = 0.7059
(b) P(partner with blue eyes | man with blue eyes) = 78 / 114 = 0.6842
(c) P(partner with blue eyes | man with brown eyes) = 19 / 54 = 0.3519 P(partner with blue eyes | man with green eyes) = 11 / 36 = 0.3056
(d) It is much more likely for a man with blue eyes to have a partner with blue eyes than a man with another eye color to have a partner with blue eyes. Therefore it appears that eye colors of males and their partners are not independent.

Books on a bookshelf. (3.26, p. 114) The table below shows the distribution of books on a bookcase based on whether they are nonfiction or fiction and hardcover or paperback.

		Format		
		Hardcover	Paperback	Total
Type	Fiction	13	59	72
	Nonfiction	15	8	23
	Total	28	67	95

(a) Find the probability of drawing a hardcover book first then a paperback fiction book second when drawing without replacement.
(b) Determine the probability of drawing a fiction book first and then a hardcover book second, when drawing without replacement.
(c) Calculate the probability of the scenario in part (b), except this time complete the calculations under the scenario where the first book is placed back on the bookcase before randomly drawing the second book.
(d) The final answers to parts (b) and (c) are very similar. Explain why this is the case.

Solution

(a) P(first hardcover, second paperback fiction) = $\frac{28}{95} \times \frac{59}{94} = 0.1850$

(b) Break this into two disjoint statements (let $F = fiction$, $N = nonfiction$, $H = hardcover$, $P = paperback$):

$$\begin{aligned} P(\text{1st F, 2nd H}) &= P((\text{1st FH and 2nd H}) \text{ OR } (\text{1st FP and 2nd H})) \\ &= P(\text{1st FH and 2nd H}) + P(\text{1st FP and 2nd H}) \\ &= \frac{13}{95}\frac{27}{94} + \frac{59}{95}\frac{28}{94} = 0.2243 \end{aligned}$$

(c) Same probability statements as part (c), except now the calculations are $\frac{13}{95}\frac{28}{95} + \frac{59}{95}\frac{28}{95} = 0.2234$.

(d) There are so many books on the bookcase and we are only drawing two books, so the probability associated with the second book will be almost entirely unaffected by whatever book is drawn first. This makes the second draw under the “without replacement” setting almost independent of the first draw, meaning it is about equivalent to drawing with replacement.

Baggage fees. (3.34, p. 124) An airline charges the following baggage fees: \$25 for the first bag and \$35 for the second. Suppose 54% of passengers have no checked luggage, 34% have one piece of checked luggage and 12% have two pieces. We suppose a negligible portion of people check more than two bags.

(a) Build a probability model, compute the average revenue per passenger, and compute the corresponding standard deviation.

(b) About how much revenue should the airline expect for a flight of 120 passengers? With what standard deviation? Note any assumptions you make and if you think they are justified.

Solution

(a) The probability model and the calculation of average revenue per passenger (expected value) are as follows:

Event	X	P(X)	X · P(X)	$(X - E(X))^2$	$(X - E(X))^2 \cdot P(X)$
No baggage	0	0.54	0	$(0 - 15.70)^2 = 246.49$	$246.49 \times 0.54 = 133.10$
1 checked bag	25	0.34	8.5	$(25 - 15.70)^2 = 86.49$	$86.49 \times 0.34 = 29.41$
2 checked bags	60	0.12	7.2	$(60 - 15.70)^2 = 1962.49$	$1962.49 \times 0.12 = 235.50$
			E(X) = \$15.70		V(X) = \$398.01
					$SD(X) = \sqrt{V(X)} = \$19.95$

(b) We assume independence between individual fliers. This probably is not exactly correct, but it would provide a helpful first approximation for the true revenue.

$$
\begin{aligned}
E(X_1 + \cdots + X_{120}) &= E(X_1) + \cdots + E(X_{120}) = 120 \times 15.70 = \$1,884 \\
V(X_1 + \cdots + X_{120}) &= V(X_1) + \cdots + V(X_{120}) = 120 \times 398.01 = \$47,761.20 \\
SD(X_1 + \cdots + X_{120}) &= \sqrt{47,761.20} = \$218.54
\end{aligned}
$$

Income and gender. (3.38, p. 128) The relative frequency table below displays the distribution of annual total personal income (in 2009 inflation-adjusted dollars) for a representative sample of 96,420,486 Americans. These data come from the American Community Survey for 2005-2009. This sample is comprised of 59% males and 41% females.

Income	*Total*
\$1 to \$9,999 or loss	2.2%
\$10,000 to \$14,999	4.7%
\$15,000 to \$24,999	15.8%
\$25,000 to \$34,999	18.3%
\$35,000 to \$49,999	21.2%
\$50,000 to \$64,999	13.9%
\$65,000 to \$74,999	5.8%
\$75,000 to \$99,999	8.4%
\$100,000 or more	9.7%

(a) Describe the distribution of total personal income.
(b) What is the probability that a randomly chosen US resident makes less than \$50,000 per year?
(c) What is the probability that a randomly chosen US resident makes less than \$50,000 per year and is female? Note any assumptions you make.
(d) The same data source indicates that 71.8% of females make less than \$50,000 per year. Use this value to determine whether or not the assumption you made in part (c) is valid.

Solution

(a) The distribution is right skewed, with a median between \$35,000 and \$49,999. The IQR of the distribution is very roughly about \$30,000. The distribution is skewed to the high end, and there are probably outliers on the high end due to the nature of the data.
(b) P(less than \$50,000) = 2.2 + 4.7 + 15.8 + 18.3 + 21.2 = 62.2%
(c) Assuming that gender and income are independent: P(less than \$50,000 and female) = P(less than \$50,000) ?? P(female) = 0.622 ?? 0.41 = 0.255 = 25.5%
(d) If these variables were independent, then the percentage of females who earn less than \$50,000 (71.8%) would equal the percentage of all people who make less than \$50,000 (62.2%). Since this is not the case, gender and income are dependent.