Data606 - Homework 2

Name: Joshua Sturm
Date: 09/04/2017

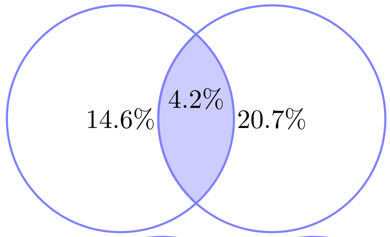
Chapter 2 problems - Page 116

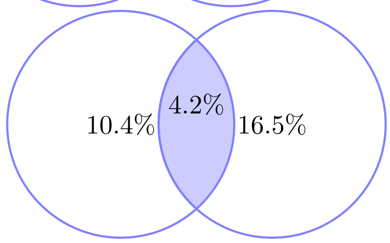
2.6— Dice rolls. If you roll a pair of fair dice, what is the probability of
(a) getting a sum of 1?
(b) getting a sum of 5?
(c) getting a sum of 12?

(a) Since you're rolling two dice, the minimum possible sum is 2, so $P(1) = 0$.
(b) $S = \{(1,4), (2,3), (3,2), (4,1)\} \rightarrow P(\text{sum} = 5) = \frac{4}{36}$ (denominator is 6^2).
(c) $S = \{(6,6)\} \rightarrow \frac{1}{36}$

2.8— Poverty and language. 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?

(a) They are not disjoint, because someone could live below the poverty line **and** speak a foreign language. The survey even includes such a statistic: 4.2%.
(b)





(c)

The left circle is the percentage of Americans living below the poverty line: $14.6\% - 4.2\% = 10.4\%$

The right circle is the percentage of Americans that speak a foreign language: $20.7\% - 4.2\% = 16.5\%$

(d) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$P(A \cup B) = 0.146 + 0.207 - 0.042 = 0.311 \rightarrow 31.1\%$.

(e) Essentially, we want the complement of (d). If they don't live below the poverty line or speak a foreign language, then they must live above, and speak only english.

$P(A^c \cap B^c) = 1 - (A \cup B) = 1 - 0.311 = 0.689 = 68.9\%$.

(f) Using the *Multiplication Rule for Independent Processes*: $P(A \cap B) = P(A) \times P(B)$.

$0.146 \times 0.207 = 0.030222$. Since the actual percent (4.2%) is higher, we can assume that these events are not independent.

2.20 Assortative mating. 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.[65]

		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) What is the probability that a randomly chosen male respondent with brown eyes has a partner 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.

2.20—

(a) We can use the inclusion/exclusion principle: $P(A \cup B) = P(A) +$

$P(B) - P(A \cap B)$.
$\frac{114}{204} + \frac{108}{204} - \frac{78}{204} = \frac{144}{204} = 0.7059 \approx 70.6\%$.

(b) The question is tricky. I interpreted it as a conditional statement, i.e. $P(A|B)$. $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{78}{114} = 0.6842 \approx 68.4\%$.

(c)(i) Similar to (b). $P(A|B) \to \frac{19}{54} = 0.3519 \approx 35.2\%$.

(c)(ii) $P(A|B) = \frac{11}{36} = 0.3056 \approx 30.6\%$.

(d) Just looking at the chart, it is evident that the variables are not independent: men with blue eyes are much more likely to choose a female partner with blue eyes. For a more rigorous proof, we could use the *Multiplication Rule for Independent Processes*.

2.30 Books on a bookshelf. 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.

2.30— (d) The final answers to parts (b) and (c) are very similar. Explain why this is the case.

(a) The probability of picking a hardcover book is $\frac{28}{95}$. Since we're not replacing the book, there are now 94 books. The probability of picking a paperback fiction is $\frac{59}{94} \to \frac{28}{95} \times \frac{59}{94} = 0.6277 \approx 62.8\%$.

(b) We need to account for the possibility of the first fiction book drawn is also a hardcover.

The probability of picking a hardcover fiction first and regular hardcover second is: $\frac{13}{95} \cdot \frac{27}{94} = 0.0393$

The probability of picking a paperback fiction first and a regular hardcover second is: $\frac{59}{95} \cdot \frac{28}{94} = 0.185$.

Since these events are disjoint, we can simply sum their probabilities: $0.393 + 0.185 = 0.2243 \approx 22.4\%$. The $\frac{72}{95} \times \frac{28}{94} = 0.2258 \approx 22.6\%$.

(c) $\frac{72}{95} \times \frac{28}{95} = 0.2234 \approx 22.3\%$.

(d) Because we're only dealing with a difference of one book, so the calculations are similar. If we had withdrawn several books without replacing them, we'd notice a bigger gap between the two probabilities.

2.38— Baggage fees. 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.

(a)

Airline Baggage Fees				
Event	X	P(X)	X$\cdot P(X)$	Variance (σ^2)
No Luggage	0	0.54	0	$(0\text{-}15.7)^2 \cdot 0.54$
1 bag	25	0.34	8.5	$(25\text{-}15.7)^2 \cdot 0.34$
2 bags	60	0.12	7.2	$(60\text{-}15.7)^2 \cdot 0.12$

Since $E(X) = \sum X \cdot P(X)$, the expected earnings is 15.70.
Standard deviation is the square root of the variance: $V(X) = \sigma^2 = \sum (X - \mu)^2 \cdot P(X)$.
$\sigma^2 = (0 - 15.7)^2 \cdot 0.54 + (25 - 15.7)^2 \cdot 0.34 + (60 - 15.7)^2 \cdot 0.12 = 398.01$
sd $= \sqrt{V(X)} = \sqrt{\sigma^2} = \sigma = 19.9502 \approx 19.95$.

(b) We found the average revenue per passenger to be \$15.70, so we can just multiply that by 120: $15.7 \times 120 = \$1,884$.
Similarly, the variance can also be multiplied by 120: $\sigma^2 = 398.01 \times 120 = \$47,761.20$.
$\sigma = \sqrt{\sigma^2} = \sqrt{47,761.20} = \$218.543$.

2.44 Income and gender. 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.[69]

(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.

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%

2.44—

(a) Looking at the numbers at the table, we can see that it's a bimodal distribution, with the peak in the middle - in the 35k-49.9k range.

(b) We can sum up the probabilities of those earning less than 50k. $0.022 + 0.047 + 0.158 + 0.183 + 0.212 = 0.622$.

(c) It's fair to assume that the two events are independent: i.e. the resident can be female and make more than 50,000.
$P(< 50k \cap F) = P(< 50k) \times P(F) = 0.622 \times 0.41 = 0.25502 \approx 25.5\%$.

(d) Using the *Multiplication Rule for Independent Processes*: $P(A \cap B) = P(A) \times P(B)$.
$0.41 \times 0.718 = 0.29348$. Since this is considerably higher than the percentage I found in (c), I'd have to reject my assumption, and conclude that there is some dependency between these variables.