Chapter 3 - Probability

Joshua Registe

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?

A) Probability of getting a sum of 1 is 0%. The minimum value that can be obtained from rolling two dies is 2, obtained from rolling two 1's.

B) probability of getting a sum of 5 is:

(1,4) or (2,3) or (3,2) or (4,1)

4 combinations over 36 possible combinations so the probability is 1/9

C) getting a sum of 12 needs the following combination

(6,6) the probability of rolling 2 6's is 1/36

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?

```
P_belowpoverty <- .146
P_multilingual<- .207
P_belowPov_and_Multiling<- .042
```

A) No, living below the poverty line and speaking a foreign language at home are not disjoint since they can happen at the same time

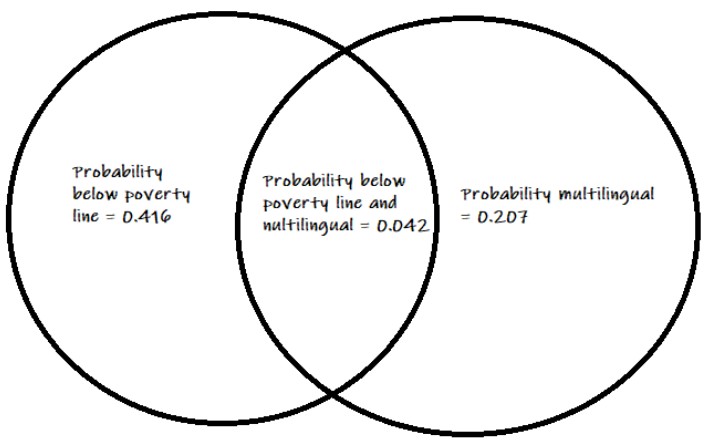

Figure 1: Venn Diagram image

B)

c)

```
P_onlyenglish<- 1-P_multilingual

P_C<- P_belowpoverty - P_belowPov_and_Multiling
```

The probability of speaking english only and being below the poverty line is 0.104

D)

```
P_D<- P_belowpoverty + P_multilingual - P_belowPov_and_Multiling
```

The probability of speaking a foreign language and being below the poverty line is 0.311

E)

```
P_E<- 1- P_D
```

The probability of not speaking a foriegn language and not being below the poverty line is the remainder of the population which is 0.689

F) Yes, these two events are independent

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?

The probability is (114+108-78)/204 which is 70.5%

(b) What is the probability that a randomly chosen male respondent with blue eyes has a partner with blue eyes?

The probability is 78/108 which is 72.2%

(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?

The probability that a male with any eye color has a female partner with blue eyes is 72.2%

(d) Does it appear that the eye colors of male respondents and their partners are independent? Explain your reasoning.

These probabilities are independent. there is no association between what eye color one sex will have vs their partner. the conditional statements eliminate the P(B) from numerator and denominator

---

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.

Probability is (28/95)(59/94) = 18.5%

(b) Determine the probability of drawing a fiction book first and then a hardcover book second, when drawing without replacement.

Probability is (72/95)((59-1*59/72)/94) = 46.9%

(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.

Probability is (72/95)(59/95) = 47%

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

The answers are very similar because the probability that a fictional book that was withdrawn is also a hardcover lowers the probability, but this is offset by the leftover books on the shelf since there is no replacement essentially causing no change to a probability demonstrated in problem C.

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.

p0 = probability no checked bag p1 = probability of 1 checked bag etc.

Average revenue per passenger = (P0xR0) + (P1xR1) + (P2xR2) = (0.54*\$0) + (0.34*\$25) + (0.12*\$35) = \$12.7/passenger Standard deviation

Standard Deviation = Square root [0.54*(0-12.7)$^{2+0.34(25\text{-}12.7)\hat{}2+0.12(35\text{-}12.7)}$2] = \$14.07

(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.

an airline should expect a revenue of \$1524 per passenger with a standard deviation of \$1688.40

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.

The distribution of this data seems to take a bell-shape gaussian form.

(b) What is the probability that a randomly chosen US resident makes less than \$50,000 per year?

The probability that a US resident makes less than 50K is (2.2+4.7+15.8+18.3+21.2)/100 = 62.2%

(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.

assumptions - data are independent and mutually exclusive P_female = 41%, and P_<50K = 62.2%

Probability of both together is .41*.622 = 0.255

(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.

given this new piece of data we know that our assumption is incorrect because their is not a uniform distribution of females in both broke and not broke groups.