Chapter 3 - Probability

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?

Zero. The lowest possible sum is 2.

(b) getting a sum of 5?

The first die can be 1,2,3 or 4. The second can only one number based on the first $\frac{4}{6} * \frac{1}{6} = \frac{1}{9}$ or 11.11%

(c) getting a sum of 12?

There is only one way to get a sum of 12, double 6s. $\frac{1}{36}$ or 2.77%

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?

No they are not disjoint. You can be both a non-native speaker and below the poverty line.

(b) Draw a Venn diagram summarizing the variables and their associated probabilities.

```
 ---------------- - - - - - - -
|poverty  |both  | non-english |
|_14%_____|_4%___|_ _20%_ _ _ _|
```

(c) What percent of Americans live below the poverty line and only speak English at home?

P(pvoerty and non-english) = 14.6% x 79.3% = 11.57%

(d) What percent of Americans live below the poverty line or speak a foreign language at home?

the stats definition of OR is inclusive so this would be poverty or 'flh' or both 14.2% + 20.7% + 4.2% = 39.1%

(e) What percent of Americans live above the poverty line and only speak English at home?

85.4% * 79.3% = 67.8%

(f) Is the event that someone lives below the poverty line independent of the event that the person speaks a foreign language at home?

No. If you know something about either percentage (below poverty of speaking non-english at home) you can determine information about the other event.

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?

There are 114 males with blue eyes and 108 females with blue eyes. This captures all couples with blues eye somewhere, but we need to subrtract the double counted blue/blue couples. $114 + 108 - 78 = 144$ and $144/204 = 70.5\%$ have blue eyes in either of the partners.

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

This is conditional probability, what is the probability of a male=blue given female=blue

$P(f = blue|m = blue) = \frac{\frac{78}{204}}{\frac{114}{204}} = 69.4\%$

(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?

$P(f = blue|m = brown) = \frac{\frac{19}{204}}{\frac{54}{204}} = 35.2\%$

$P(f = blue|m = green) = \frac{\frac{11}{204}}{\frac{36}{204}} = 30.5\%$

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

They seem dependent. You can determine information about possible eye color of one partner from the other.

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.

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.

72/95 * 27/94 = 21.7%

(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.

72/95 * 28/95 = 22.3%

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

Because the same book is both ficition and hardcover so both the numerator and denominator decreased by the same amount in part (b). Also one book out of 95 isn't going be make much change.

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.

$E(X) = 0 * (.54) + 25 * (.34) + (35 + 25) * (.12) = 15.7$ dollars per customer is expected.

$\sigma^2 = (x_i - \mu)^2 * P(x_i) = (-15.7)^2 + (25 - 15.7)^2 * .34 + (60 - 15.7)^2 * (.12) = 398.01,$

$\sigma = 19.95$

(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.

$\sigma_{120} = \sqrt{}(120 * Var) = \sqrt{}(120 * 398.01) = 218.54$

We except to make $\$15.7 * 120 = \$1884 \pm 218.54$

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.

This appears to be relatively normal (Gaussian). It is mostly symetric and does not appear to be noticably skewed.

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

This would be 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.

Assuming this survey discriminates against gender non-conforming people such that there are only two genders represented. (.41)*(.622) = 25.5%

(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.

This is a huge change. Something has to be wrong, either there is another group of people who were lumped into the female category and who make more money or there is something wrong with how income is being reported or collected.