Leland Randles DATA606

Homework, Chapter 2

2.6 Dice rolls. If you roll a pair of fair dice, what is the probability of

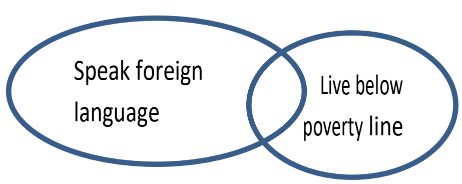
(a) getting a sum of 1?
(b) getting a sum of 5?
(c) getting a sum of 12?

(a) The probability is zero because it is not possible to roll two dice and get a sum of 1
(b) There are four different combinations which could lead to a sum of 5 – 1 & 4, 2 & 3, 4 & 1, or 3 & 2, hence the probability is 4 out of the total combinations, which is 36 (6 times 6). Hence the probability is 4/36 or 1/9.
(c) Only 6 and 6 would yield a sum of 12, and the odds of rolling two sixes is 1/6 * 1/6 or 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.[59]

(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) No, they are not disjoint
(b) Oval labeled "speak foreign language" represents 20.7% of survey participants, oval with "live below poverty line" represents 14.6% of survey participants, and the intersection of the two ovals represents the 4.2% that fall into both categories:



(c) 14.6% minus 4.2% = 10.4%
(d) 14.6% + 20.7% - 4.2% = 31.3%
(e) Seems to be repeat of c above, which was 14.6% minus 4.2% = 10.4%
(f) No, there is overlap between the two

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.

(a) There are three possibilities that fulfill the question – the male has blue eyes but his partner does not, the male does not have blue eyes but his partner does, or both the male and his partner have blue eyes. 36 respondents fit the first scenario, 30 respondents fit the 2nd scenario, and 78 respondents fit the 3rd scenario, which is a total of 144 out of a total of 204 respondents, which means the probability is 144/204 = 0.706.
(b) 78 / 114 = 0.684
(c) 19 / 54 = 0.352; 11 / 36 = 0.306
(d) No. It appears it is more likely that males will pick a partner with the same eye color over a partner with a different eye color, so when you know what color eyes the male respondent has, it effects the likelihood for the eye color of the partner.

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

(a) Probability of drawing a hardcover book is 28 / 95. Probability of drawing a paperback fiction book next without replacement is 59 / 94. Hence the probably of drawing the two in order is 28/95 * 59/94 = 0.185

(b) Probability of drawing a fiction book is 72 / 95. Probability of drawing a hardcover book next without replacement depends on whether the fiction book drawn was paperback or hardcover. There was a 13/72 chance the fiction book was hardcover, and 59/72 chance that it was paperback. Therefore, the chance that the 2nd book drawn is a hardcover book is (13/72)*(27/94) + (59/72)*(28/94), which equals 0.296. Take 72/95 * 0.296 and you get 0.224.

(c) 72/95 * 28/95 = 0.223.

(d) There are fairly similar because the sample size is fairly large (95 books). The effect of sample with or without replacement diminishes based on the size of the sample.

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) Expected value is 15.7. As you can see from table below, the variance is 376.12, so the standard deviation (σ) = 376.12^0.5 = 19.394.

i	**1**	**2**	**3**	**Total**
x_i (in dollars)	0	25	60	
$P(X = x_i)$	0.54	0.34	0.12	
$x_i * P(X = x_i)$	0	8.5	7.2	15.7
$x_i - \mu$	-15.7	9.3	44.3	
$(x_i - \mu)^2$	246.69	86.49	1962.49	
$(x_i - \mu)^2 * P(X = x_i)$	133.21	29.41	235.50	376.12

(b) For 120 passengers, you would expect 120 * \$15.70 = \$1,884. The standard deviation of this amount would also be \$19.394.

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%

(a) It is right-skewed, unimodal

(b) 2.2 + 4.7 + 15.8 + 18.3 + 21.2 = 62.2%

(c) If we assume salaries are equal between the genders (unlikely), then it would be the probability computed in (b) above multiplied times the probability of the resident being female (0.41). 0.622 * 0.41 = 0.255.

(d) If the overall population (males and females combined) has a 62.2% chance of making less than \$50K/year, and 71.8% of females make less than \$50K/year, then females disproportionately make under \$50K/year, and the percentage of males that make \$50K/year or less is going to be significantly below 62.2%, because that is for the whole population. Given the very large sample size and big difference between males and females in the percentages of individuals making less than \$50K/year, my gut instinct is that the difference would easily meet significance thresholds indicating a difference between male and female populations with respect to likelihood they make less than \$50K/year. Therefore the assumption used for (c) is likely untrue.