Chapter 3 - Probability

Zachary Safir

2/18/2021

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?

As the minimum sum of two dice is two, zero probability of that occurring.

(b) getting a sum of 5?

The possible combinations are one and four, two and three, three and two, and four and one. Four out of thirty six possibilities, therefore a one out of nine possibility.

(c) getting a sum of 12?

Only one possibility, a six and a six therefore, one out of thirty six probability.

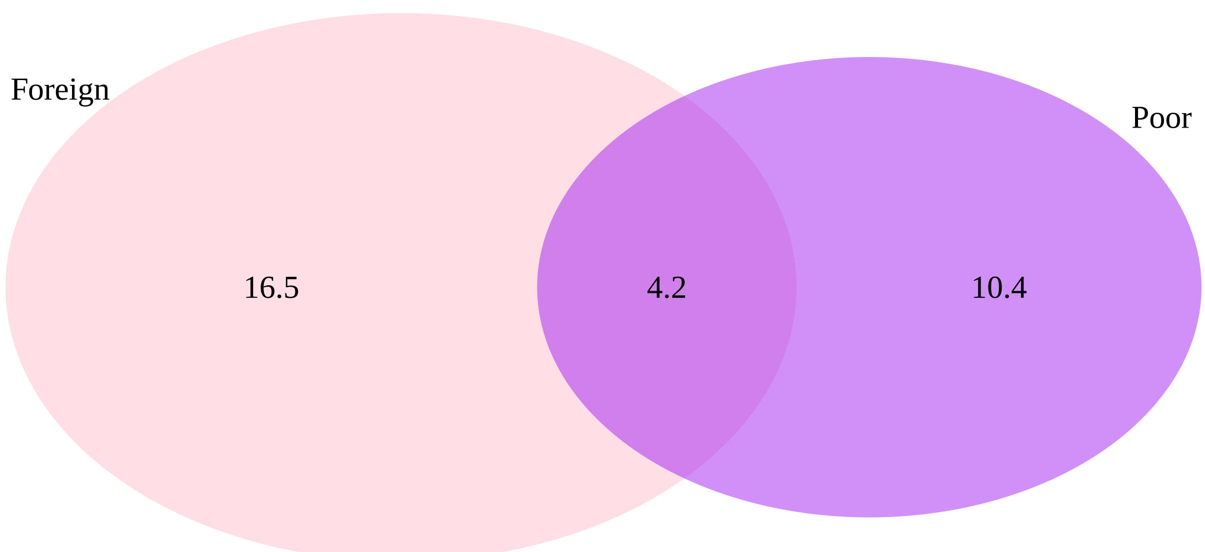
Poverty and language. (3.8, p. 93) The American Community Survey is an ongoing survey that provides data every year to give communities the current information they need to plan investments and services. The 2010 American Community Survey estimates that 14.6% of Americans live below the poverty line, 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, 4.2 percent of people in the survey are both.

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

```
library(VennDiagram)
grid.newpage()
draw.pairwise.venn(area1 =20.7,
                   area2 = 14.6,
                   cross=4.2,
                   fill = c("pink", "purple"),
                  category = c('Foreign','Poor'),
                   lty = "blank")
```




```
## (polygon[GRID.polygon.1], polygon[GRID.polygon.2], polygon[GRID.polygon.3], polygon[GRID.polygon.4],
```

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

As shown in the Venn diagram above, 10.4 percent are poor and speak English at home.

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

14.6 plus 20.7 minus the probably of both which is 4.2, giving us 31.1 percent.

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

100 - 14.6 live above the poverty line, 85.4. 16.5 percent of them speak a foreign language, so 85.4 - 16.5, for a percentage of 68.9

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

Using Bayes' Theorem, if these two events are independent, then the probably of being both below the poverty line and speaking a foreign language over the probably of speaking a foreign language,

$\frac{P(below\ PL\ and\ speak\ FL)}{P(speak\ FL)}\ \frac{.042}{.207} \approx .203$

therefore we can conclude that these are not 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 $\frac{114+19+11}{240}$ which equals .71

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

The probability is $\frac{78}{114}$ which equals .68

(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 for brown eyed male has a blue eyed partner $\frac{19}{54}$ which equals .35

The probability for male with green eyes having a partner with blue eyes is $\frac{11}{36}$ which equals .31

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

We can decide that by looking at the following probability, $P(blue\ male \mid blue\ female) = P(blue\ male) * P(blue\ female)$ $\frac{78}{204} = \frac{114}{204} * \frac{108}{204}$. These are not equal, therefore it does not appear that the eye colors of male respondents and their partners are 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.

The probability is $\frac{28}{95} * \frac{59}{94}$ which equals .18

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

The probability is $\frac{72}{95} * \frac{28}{94}$ which equals .23

(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.

The probability is $\frac{72}{95} * \frac{28}{95}$ which equals .22

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

Independent events and a large pool of possible outcomes means that the outcome will not be affected much with or without 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.

The x values we have are 0, \$25, and \$60. Their respective probabilities are .54, .34, and .12 we have

$E(X) = (0 * .54) + (25 * .34) + (60 * .12)$ $E(X) = 15.7$

then we have

$\sigma^2 = \sum_X (x - \mu)^2 * P(X = x)$
$\sigma^2 = (0 - 15.7)^2 * .54 + (25 - 15.7)^2 * .34 + (60 - 15.7)^2 * .12$
$\sigma^2 = 133.1 + 29.4 + 235.5 = 398$

Standard Deviation $= \sigma = \sqrt{\sigma^2} = \sqrt{398} \approx 19.95$ (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.

Rounding 19.95 up to 20 we have $120 * \$15.70 = \$1,884 \pm \$20$ **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.e

---

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.

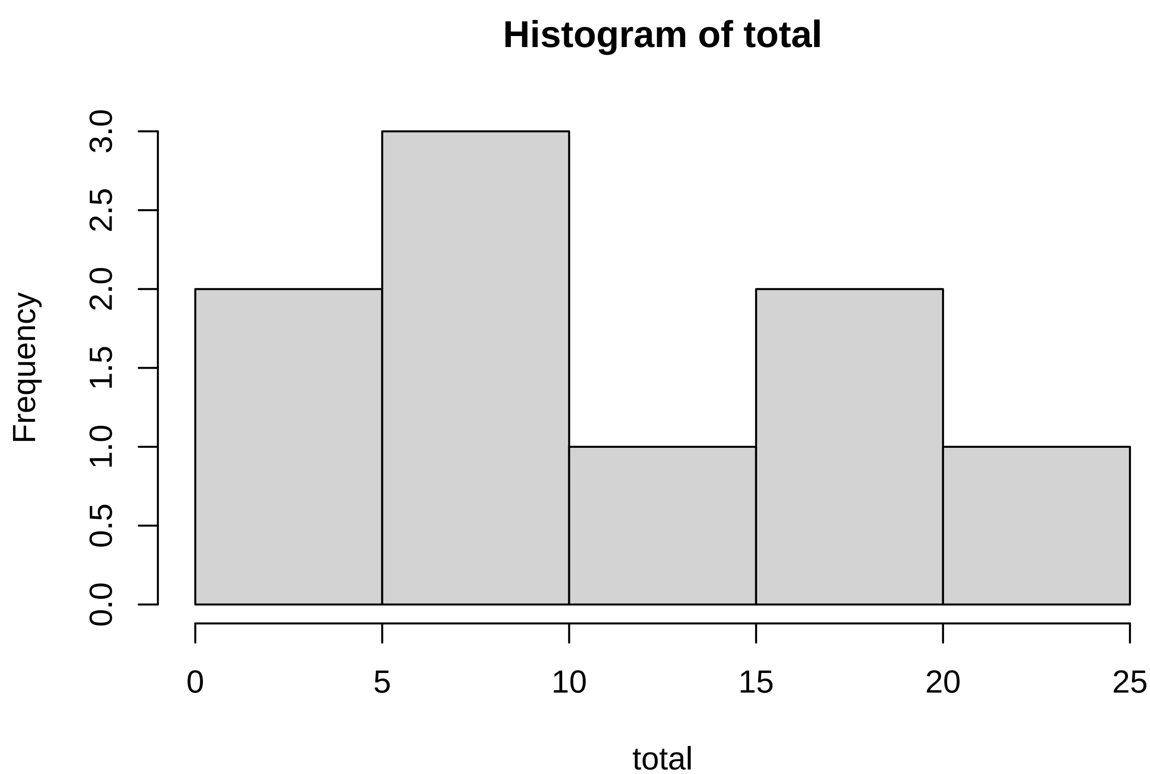
Income	*Total*
\$1 to \$9,999 or loss	2.2%
\$10,000 to \$14,999	4.7%
\$15,000 to \$24,999	15.8%
\$25,000 to \$34,999	18.3%
\$35,000 to \$49,999	21.2%
\$50,000 to \$64,999	13.9%
\$65,000 to \$74,999	5.8%
\$75,000 to \$99,999	8.4%
\$100,000 or more	9.7%

(a) Describe the distribution of total personal income.

The histogram below appears to be bimodal with a right skew.

```
total <- c(2.2,4.7,15.8,18.3,21.2,13.9,5.8,8.4,9.7)

hist(total)
```



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

21.2 + 18.3 + 15.8 + 4.7 + 2.2 which equals 62.2.

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

.6220 * .41 which equals .255 or 25.5%

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.

If 71.8% of females really made less than $50,000 per year then the equation would be 0.718 = 0.622 * 0.41. This is not true, so can conclude that this assumption is incorrect. It appears that the events are not independent.