

Chapter 3 - Probability

Adam Gersowitz

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?

0

(b) getting a sum of 5?

4/36 or 0.111111

```
#install.packages('dice')

library(dice)

getSumProbs(ndicePerRoll = 2,
             nsidesPerDie = 6,
             nkept = 2)
```

```
## $probabilities
##      Sum Probability Ways to Roll
## [1,]  2  0.02777778         1
## [2,]  3  0.05555556         2
## [3,]  4  0.08333333         3
## [4,]  5  0.11111111         4
## [5,]  6  0.13888889         5
## [6,]  7  0.16666667         6
## [7,]  8  0.13888889         5
## [8,]  9  0.11111111         4
## [9,] 10  0.08333333         3
## [10,] 11  0.05555556         2
## [11,] 12  0.02777778         1
##
## $average
## [1] 7
```

(c) getting a sum of 12?

(d) getting a sum of 12?

1/36 or 0.278

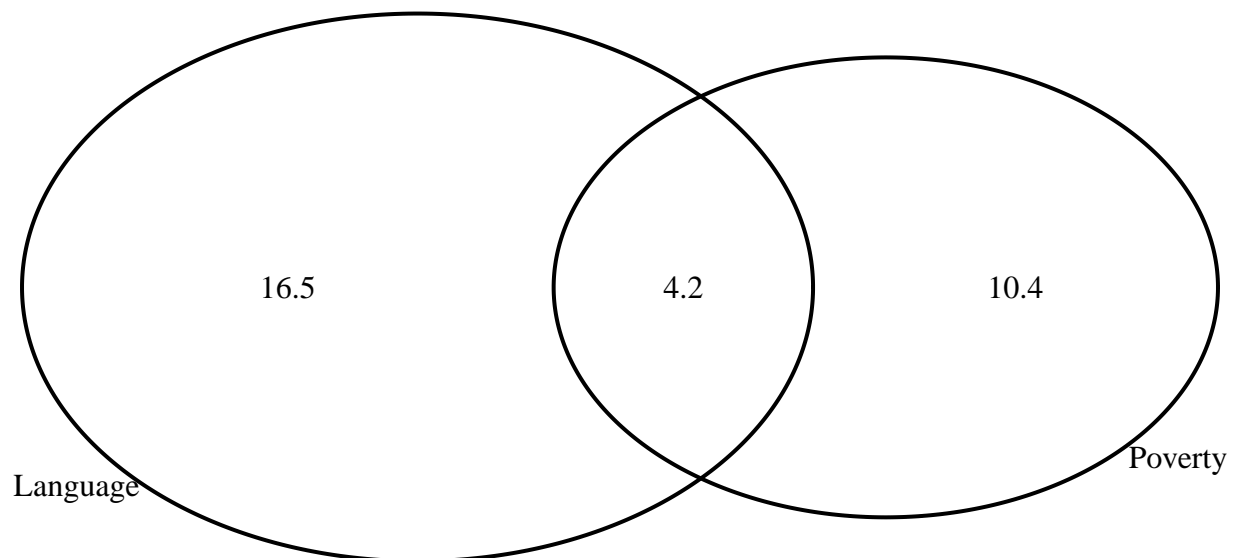
Poverty and language. (3.8, p. 93) The American Community Survey is an ongoing survey that provides data every year to give communities the current information they need to plan investments and services. The 2010 American Community Survey estimates that 14.6% of Americans live below the poverty line,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

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

```
#install.packages('VennDiagram')
library(VennDiagram)
grid.newpage()
draw.pairwise.venn(area1 = 14.6, area2 = 20.7, cross.area = 4.2, category = c("Poverty",
"Language"))
```



```
## (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?

10.4

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

16.5+4.2+10.4

[1] 31.1

31.1%

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

100-31.1

[1] 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?

Yes

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.

		<i>Partner (female)</i>			Total
		Blue	Brown	Green	
<i>Self (male)</i>	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?

$(78+23+13+19+11)/204$

[1] 0.7058824

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

$78/114$

[1] 0.6842105

- (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?

$19/54$

[1] 0.3518519

$11/36$

[1] 0.3055556

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

No, it appears that the partners eye colors are often the same so they are associated with each other and therefore not 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.

	<i>Format</i>		<i>Total</i>
	Hardcover	Paperback	
<i>Type</i>	Fiction	13	59
	Nonfiction	15	8
	Total	28	67

- (a) Find the probability of drawing a hardcover book first then a paperback fiction book second when drawing without replacement.

```
(28/95)*(59/94)
```

```
## [1] 0.1849944
```

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

```
PFHC <-13/95
PFPB<-59/95
PHC_AfterHC<-27/94
PHC_AfterPB<-28/94
(PFHC*PHC_AfterHC)+(PFPB*PHC_AfterPB)
```

```
## [1] 0.2243001
```

- (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.

```
PFHC <-13/95
PFPB<-59/95
PHC_AfterHC<-28/95
PHC_AfterPB<-28/95
(PFHC*PHC_AfterHC)+(PFPB*PHC_AfterPB)
```

```
## [1] 0.2233795
```

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

With a sample of 95 books, replacing the first book on the shelf will not make much of a difference as far as the next selection.

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.

```
EX = (25*0.34)+(35*.12)+(0*.54)
EX
```

```
## [1] 12.7
```

```
xi = c(0,25,35)
PXxi = c(.54,.34,.12)
xipx = xi*PXxi
ximu = xi-EX
xius = ximu^2
xiuspxx = xius*PXxi
variance = sum(xiuspxx)
SD = sqrt(variance)
SD
```

```
## [1] 14.07871
```

- (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.

I assumed that the expected outcome would be the total passengers* the expected baggage fee (120*12.7). I also multiplied the amounts earned per type of baggage by 120 as this would be accurate if one of them were the baggage choice for all 120 passengers. I then ran that through the same process as above (page 118 in the book) to get the Standard Deviation.

```
pop=120
EX = (25*0.34)+(35*.12)+(0*.54)
EX
```

```
## [1] 12.7
```

```
EX*pop
```

```
## [1] 1524
```

```
xi = c(0,25,35)
xip=xi*pop
PXxi = c(.54,.34,.12)
xipx = xip*PXxi
ximu = xip-EX
xius = ximu^2
xiuspxx = xius*PXxi
variance = sum(xiuspxx)
SD = sqrt(variance)
SD
```

```
## [1] 2266.771
```

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.

<i>Income</i>	<i>Total</i>
\$1 to \$9,999 or less	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 total personal income is a binomial distribution that is right-skewed.

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

0.212+0.183+0.158+0.042+0.022

[1] 0.617

- (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.

I'm assuming that Men and Women in this sample make about the same average salary.

(0.212+0.183+0.158+0.042+0.022)*0.41

[1] 0.25297

- (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 information means the assumption I made is not valid. 71.8% of females making under 50k is much higher than the probability that a randomly chosen person would make under 50k (61.7%). This means the probability of a female making under 50k is much higher then the 25.3% determined in part c.