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

ANSWER

(a). It's impossible to roll a sum of 1 with two fair dice, as the minimum sum you can get is 2. Therefore, the probability of rolling a sum of 1 is zero.

(b). The probability of rolling a sum of 5 with two fair dice is 4/36, which is approximately 11.11%. There are 4 combinations that result in a sum of 5: (1+4), (2+3), (3+2), and (4+1). Since there are 36 possible outcomes when rolling two dice, the probability is calculated as 4/36 or 1/9.

(c). The probability of rolling a sum of 12 with two fair dice is 1/36, which is about 2.78%. This sum can be achieved in only one way: (6+6). Since there are 36 possible outcomes when rolling two dice, the probability is 1/36.

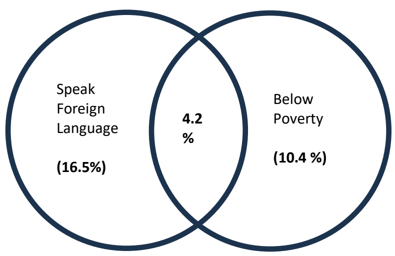
Poverty and language. (3.8, p. 93) The American Community Survey is an ongoing survey that provides data every year to give communities the current information they need to plan investments and services. The 2010 American Community Survey estimates that 14.6% of Americans live below the poverty line,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

ANSWER

(a). No, there are people who speak a foreign language at home and live below the poverty line. However, the number of Americans who speak a foreign language at home exceeds the number of those living in poverty. Since, 4.2% of the population lives below the poverty line and speaks a language other than English.

(b).



(c). The percentage of Americans who live below the poverty line and only speak English at home is 10.4%, calculated by subtracting the 4.2% who live below the poverty line and speak a foreign language from the total 14.6% who live below the poverty line.

P(P) - P(P & F) = 14.6 - 4.2 = 10.4%

(d). The percentage of Americans who either live below the poverty line or speak a foreign language at home is 31.1%. This is found by adding 14.6% (those below the poverty line) to 20.7% (those who speak a foreign language) and then subtracting 4.2% (to account for those counted twice because they fall into both categories).

P(P) + P(F) - P(P & F) = 14.6 + 20.7 - 4.2 = 31.1%

(e). The percentage of Americans who live above the poverty line and only speak English at home is 68.9%, calculated by subtracting the 31.1% who either live below the poverty line or speak a foreign language from the total 100%.

P(E and NoPov)= 100 - P(P) + P(Other) - P(P and Other) = 100 - (20.7 + 14.6 - 4.2) = 68.9 %

(f). The events are not independent, as there are various ways one variable can intersect with the other, indicating that one variable can influence or overlap with the other.

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

ANSWER

(a).

Let's consider the following:

- **P(MB)**: Probability that a male has blue eyes.
- **P(FB)**: Probability that a female partner has blue eyes.
- **P(MB AND FB)**: Probability that both the male and his partner have blue eyes.
- **P(MB or FB)**: Probability that either the male or the female partner has blue eyes.
-

Using the addition rule for probabilities:
P(MB or FB) = P(MB)+P(FB)−P(MB AND FB)

Substituting the values:
P(MB or FB) = 114/204 + 108/204 −78/204 = 144/204 ≈ 70.59 %

Thus, the probability that either the male or the female partner has blue eyes is approximately 70.59%.

(b). The probability that a randomly selected male respondent with blue eyes has a partner with blue eyes is 78/114, which equals approximately 68.42%.

(c). The likelihood of a male with brown eyes having a partner with blue eyes is 19 out of 204. The probability that a male has brown eyes is 54 out of 204. When dividing these two probabilities, we get approximately 35.19%.

For a male with green eyes, the probability of having a blue-eyed partner is 11 out of 204. The probability that a male has green eyes is 36 out of 204. Dividing these figures yields about 30.56%.

(d). The probability that a female has blue eyes, given that her male partner has blue eyes, is 78 out of 114. Meanwhile, the overall probability of a female having blue eyes is 108 out of 204. Because these probabilities differ, it suggests that the eye colors of blue-eyed individuals are not independent.
This indicates that, in general, the eye colors of male respondents and their partners are 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.

		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.

ANSWERS

(a). The probability of selecting a hardcover book first and then a paperback fiction book is given by:
P(Hardcover first) × P(Paperback fiction) = (28/95) × (59/94) ≈ 18.50%

(b). The probability of drawing a fiction book first and then a hardcover book second, without replacement, is given by:
72/95 × 28/94 ≈ 22.58 %

(c). The probability of drawing a fiction book first and a hardcover book second, with both events occurring in sequence, is calculated as: 72/95 × 28/95 ≈ 22.34%

(d). The difference between (b) and (c) arises from whether the book is replaced or not, which changes the denominator from 1/(95 x 94) to 1/(95 x 95). This alteration has a minor impact on the overall percentage.

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.
(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.

<u>ANSWER</u>

(a).

Average Revenue

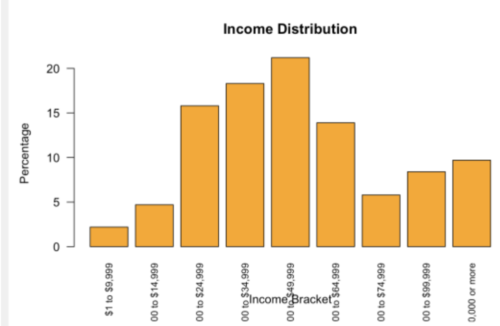
```
```{r}
# Define the fees and passengers
fees <- c(0, 25, 60)
passengers <- c(54, 34, 12)

# Create a data frame
baggage <- data.frame(fees, passengers)

# Calculate the total revenue
total_revenue <- sum(baggage$fees * baggage$passengers)

# Calculate the total number of passengers
total_passengers <- sum(baggage$passengers)

# Calculate the average revenue per passenger
average_revenue <- total_revenue / total_passengers

# Print the average revenue
print(average_revenue)

```

[1] 15.7
```

Standard Deviation

```
```{r}
# Define the fees, passengers, and average revenue
fees <- c(0, 25, 60)
passengers <- c(54, 34, 12)
average_revenue <- (sum(fees * passengers)) / sum(passengers)

# Calculate the squared deviations from the average revenue
squared_deviations <- (fees - average_revenue)^2

# Compute the weighted sum of squared deviations
weighted_squared_deviations <- squared_deviations * passengers

# Calculate the variance
variance <- sum(weighted_squared_deviations) / sum(passengers)

# Compute the standard deviation
standard_deviation <- sqrt(variance)

# Print the standard deviation
print(standard_deviation)
```

[1] 19.95019
```

Event	**X**	**P(X)**	**X * P(X)**
0 Checked bag	0	.54	0
1st checked bag	$25	.34	8.5
2nd checked bags	$60	.12	7.2
			E(X) = 15.7

= X*P(X)
0 Checked bags = 0 * .54=0
1st Checked bags = 0.34 * 25=$8.50
2nd Checked bags = 0.12 * 60=$7.20
Total sum or $15.70 per passenger

The variance is: 15.70^2 * 0.54 = 133.10
+ (25 - 15.70)^2 * 0.34 = 29.41
+ (60 - 15.70)^2 * 0.12 = 235.50
Total = (133.10 + 29.41 + 235.50) $^2 = 398.01 $^2
SQRT of above is the standard deviation = $19.95

(b).

For 120 passengers, the expected value is 120 * $15.70 or $1884

Total Expected Revenue

```
```{r}
# Define the number of passengers and corresponding fees
total_passengers <- 120
percent_fees_25 <- 0.34
percent_fees_35 <- 0.12
percent_fees_25_other <- 0.12

fee_25 <- 25
fee_35 <- 35
fee_25_other <- 25

# Calculate the number of passengers for each fee type
passengers_fee_25 <- total_passengers * percent_fees_25
passengers_fee_35 <- total_passengers * percent_fees_35
passengers_fee_25_other <- total_passengers * percent_fees_25_other

# Calculate the revenue for each fee type
revenue_fee_25 <- passengers_fee_25 * fee_25
revenue_fee_35 <- passengers_fee_35 * fee_35
revenue_fee_25_other <- passengers_fee_25_other * fee_25_other

# Calculate the total expected revenue
total_revenue <- revenue_fee_25 + revenue_fee_35 + revenue_fee_25_other

# Print the total expected revenue
print(total_revenue)
```

[1] 1884
```

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.
(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.

ANSWER

(a).

```
```{r}
# Define income brackets and corresponding percentages
income_brackets <- c("$1 to $9,999", "$10,000 to $14,999", "$15,000 to $24,999",
                     "$25,000 to $34,999", "$35,000 to $49,999", "$50,000 to $64,999",
                     "$65,000 to $74,999", "$75,000 to $99,999", "$100,000 or more")

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

# Create a data frame with income brackets and percentages
income_distribution <- data.frame(
  Income_Bracket = income_brackets,
  Percentage = percentages
)

# Display the frequency table
print(income_distribution)

```
```

Income_Bracket <chr>	Percentage <dbl>
$1 to $9,999	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

9 rows

```
```{r}
   # Creating a bar plot of the percentages
barplot(income_distribution$Percentage,
        names.arg = income_distribution$Income_Bracket,
        las = 2,
        col = "orange",                 # Color of the bars
        main = "Income Distribution", # Title of the plot
        xlab = "Income Bracket",        # Label for the x-axis
        ylab = "Percentage",            # Label for the y-axis
        cex.names = 0.8)                # Adjust text size for x-axis labels

```
```



The distribution is roughly normal but shows significantly more kurtosis in the right tail compared to the left tail.

(b).

To determine the probability that a randomly selected US resident earns less than $50,000 annually, we sum the percentages of residents in income brackets below $50,000. These brackets are:

- **$1 to $9,999**: 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%

By adding these percentages together:

2.2% + 4.7% + 15.8% + 18.3% + 21.2% = 62.2%

Thus, there is a 62.2% chance that a randomly selected resident earns less than $50,000 per year.

(c).

To find the probability that a randomly selected US resident earns less than $50,000 per year and is female, I followed below steps:

1. **Calculating the Probability of Earning Less Than $50,000**: From the data, 62.2% of residents fall into income brackets below $50,000 annually.

2. **Determining the Proportion of Females**: The female population constitutes 41% of the total population.

3. **Computing the Joint Probability at last**: Assuming that the distribution of income is the same across genders, we multiply the probability of earning less than $50,000 by the proportion of females:

Probability = 62.2% × 41% = 25.5%

Thus, there is a 25.5% chance that a randomly selected resident is both a female and earns less than $50,000 per year.

(d). If the data shows that 71.8% of females earn less than $50,000 a year, it means our assumption that income is evenly distributed between genders is wrong. This difference suggests that income levels for females are not the same as for the overall population, indicating that income is not evenly spread between males and females.