By: Nabila Hossain
DATA 606

Chapter 2 Homework

Chapter 2: 2.6, 2.8, 2.20, 2.30, 2.38, 2.44 from *OpenIntro Statistics*.

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

Here is a summary of the sum of a pair of fair dice:

Dice sum	1	2	3	3	5	6	7	8	9	10	11	12
Probability	$\frac{0}{36}$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

(a) getting a sum of 1?

Answer: Since the lowest sum you can get while rolling a pair of dice is 2, the probability of getting a sum of 1 is 0.

$$P(\text{sum} = 1) = \frac{0}{36} = 0$$

(b) getting a sum of 5?

Answer: While rolling a pair of dice there is 4 ways you can get the sum dice to be 5. So the probability of getting a sum of 5 is $\frac{1}{9}$.

$$P(\text{sum} = 5) = \frac{4}{36} = \frac{1}{9}$$

(c) getting a sum of 12?

Answer: While rolling a pair of dice there is only 1 way you can get the sum dice to be 12, by rolling two 6. So the probability of getting a sum of 12 is $\frac{1}{36}$.

$$P(\text{sum} = 12) = \frac{1}{36}$$

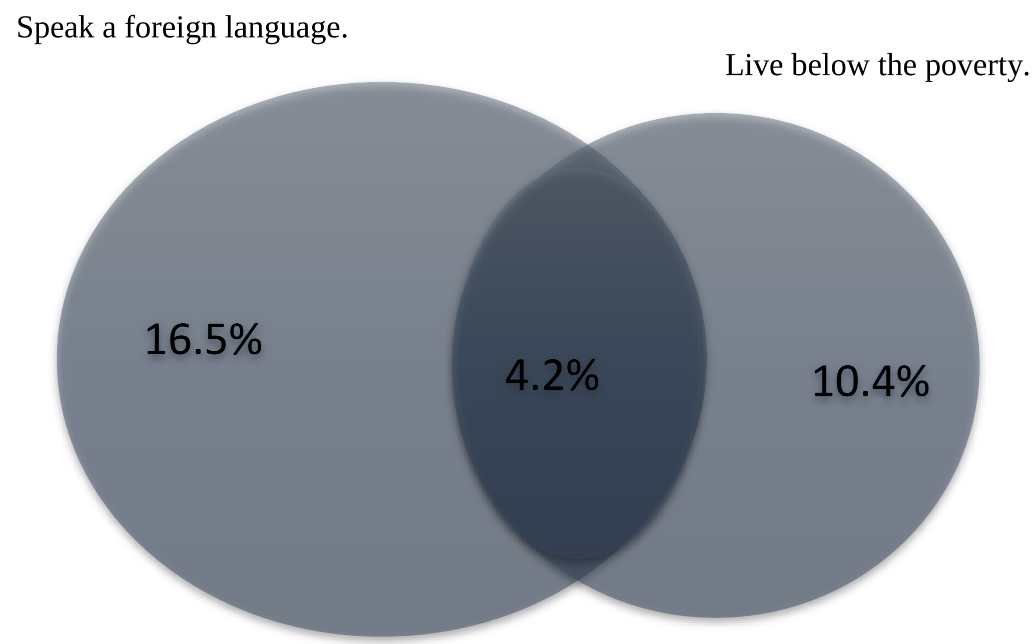
2.8 Poverty and language. The American Community Survey is an ongoing survey that provides data every year to give communities the current information they need to plan investments and services. The 2010 American Community Survey estimates that 14.6% of Americans live below the poverty line, 20.7% speak a language other than English (foreign language) at home, and 4.2% fall into both categories.

(a) Are living below the poverty line and speaking a foreign language at home disjoint?

Answer: No, living below the poverty line and speaking a foreign language at is not disjoint. It is non-disjoint, since 4.2% of Americans live below the poverty line and speak a foreign language.

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

Answer:



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

Answer: 10.4% of Americans live below the poverty line and only speak English at home.

$$\begin{aligned} &\mathrm{P(below\ poverty\ line\ and\ only\ English)} \\ &\quad = \mathrm{P(below\ poverty\ line)} - \mathrm{P(foreign\ and\ below\ poverty\ line)} \\ &\quad = 14.6\% - 4.2\% = 10.4\% \end{aligned}$$

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

Answer: 31.1% of Americans live below the poverty line or speak a foreign language at home.

$$\begin{aligned} &\mathrm{P(below\ poverty\ line\ or\ foreign)} \\ &\quad = \mathrm{P(below\ poverty\ line)} + \mathrm{P(foreign\ )} - \mathrm{P(foreign\ and\ below\ poverty\ line)} \\ &\quad = 14.6\% + 20.7\% - 4.2\% = 31.1\% \end{aligned}$$

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

Answer: 68.9% of Americans live above the poverty line and only speak English at home.

$$\mathrm{P(above\ poverty\ line\ and\ english)} = 1 - \mathrm{P(below\ poverty\ line\ or\ foreign)}$$
$$= 1 - 31.1\% = 68.9\%$$

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

Answer: The event that someone lives below the poverty line is not independent of the event that the person speaks a foreign language at home. We can see this if we follow the Bayes' Theorem:

$$P(A|B) = \frac{P(A\ and\ B)}{P(B)} = P(A)$$

$$P(\text{below poverty line} \mid \text{foreign}) = \frac{P(\text{below poverty line } \textit{and} \text{ foreign})}{P(\text{foreign})}$$

$$= \frac{0.042}{0.207} \approx 0.203 \neq P(\text{below poverty line})$$

2.20 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.

	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?

Answer: The probability that a randomly chosen male respondent or his partner has blue eyes is 70.59 %.

$$\mathrm{P(m = blue\ or\ f = blue\ ) = \ P(m = blue) + P(f = blue) - P(m = blue\ and\ f = blue\ )}$$
$$\frac{114}{204} + \frac{108}{204} - \frac{78}{204} = \frac{144}{204} \approx 0.7059 = 70.59\ \%$$

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

Answer: The probability that a randomly chosen male respondent with blue eyes has a partner with blue eyes is 68.42 %.

$$\mathrm{P(f = blue|m = blue)} = \frac{\mathrm{P(f = blue\ and\ m = blue\ )}}{\mathrm{P(m = blue)}} = \frac{78}{114} \approx 0.6842 = 68.42\ \%$$

(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?

Answer: The probability that a randomly chosen male respondent with brown eyes has a partner with blue eyes is 35.19 %. The probability of a randomly chosen male respondent with green eyes having a partner with blue eyes is 30.56%.

$$\mathrm{P(f = blue|m = brown)} = \frac{\mathrm{P(f = blue\ and\ m = brown\ )}}{\mathrm{P(m = brown)}} = \frac{19}{54} \approx 0.3519 = 35.19\ \%$$

$$\mathrm{P(f = blue|m = green)} = \frac{\mathrm{P(f = blue\ and\ m = green\ )}}{\mathrm{P(m = green)}} = \frac{11}{36} \approx 0.3056 = 30.56\ \%$$

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

Answer: The probability of eye colors of male and their partners are dependent. For example the probability of female partner having blue eyes is $P(f = blue) = \frac{108}{204} \approx$ 52.94 %, while the probability female partner having blue eyes given that her male respondents have blue eyes is $\mathrm{P(f = blue|m = blue)} \approx 68.42$ %.

2.30 Books on a bookshelf. The table below shows the distribution of books on a bookcase based on whether they are non-fiction or fiction and hardcover or paperback.

Format

Type		Hardcover	Paperback	Total
	Fiction	13	59	72
	Non-fiction	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.

Answer: The probability of drawing a hardcover book first then a paperback fiction book second when drawing without replacement is 18.50%.

$$P(1st\ HC\ fiction\ then\ PB\ fiction) = \mathrm{P(HC\ fiction) * P(2nd\ PB|1st\ HC)} = \frac{28}{95} * \frac{59}{94}$$
$$\approx 0.1850 = 18.50\ \%$$

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

Answer: The probability of drawing a fiction book first and then a hardcover book second, when drawing without replacement is 22.43%.

This can go two ways: a)1st draw paperback fiction then draw any hardcover 2nd.
b) 1st draw a hardcover fiction then draw any other hardcover 2nd.

So to get the probability in both scenario we find probability of both cases:

$$P(1st\ fiction\ then\ HC) = \mathrm{P(2nd\ HC\ |\ 1st\ PB\ fiction) + P(2nd\ HC\ |\ 1st\ HC\ fiction)}$$
$$= \left(\frac{59}{95} * \frac{28}{94}\right) + \left(\frac{13}{95} * \frac{27}{94}\right) \approx 0.1850 + 0.0393 = 0.2243 = 22.43\ \%$$

(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.

Answer: 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 is 22.34%

$$P(1st\ fiction\ then\ HC) = \mathrm{P(fiction) * P(HC)} = \frac{72}{95} * \frac{28}{95} \approx 0.2234 = 22.34\ \%$$

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

Answer: The final answers to part (b) 22.43% and (c) 22.34% is very similar. This is due to the fact that the sample size is small (only holds 94 books). If the sample size was larger (over 10,000) than the change in one book (with replacement or without replacement) would be imperceptible.

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.

Answer:

Probability model:

Event	x_i	$P(X = x_i)$	$x_i * P(X = x_i)$
No luggage	\$0	0.54	$0 * .54 = 0$
1 luggage	\$25	0.34	$25 * 0.34 = 8.5$
2 luggage	\$25+\$35 = \$60	0.12	$60 * 0.12 = 7.2$
Total			$E(X) = \sum(x_i * P(X = x_i))$ $= 0 + 8.5 + 7.2 = 15.7$

Equations	No luggage	1 luggage	2 luggage	Total
x_i	\$0	\$25	\$60	
$P(X = x_i)$	0.54	0.34	0.12	
$x_i * P(X = x_i)$	0	8.5	7.2	$E(X) = 15.7$
$x_i - E(X)$	$0 - 15.7$ $= -15.7$	$25 - 15.7$ $= 9.3$	$60 - 15.7$ $= 44.3$	
$(x_i - E(X))^2$	$-15.7^2 = 246.49$	$9.3^2 = 86.49$	44.3^2 $= 1962.49$	
$(x_i - E(X))^2$ $* P(X = x_i)$	$246.49 * 0.54$ $= 133.1046$	$86.49 * 0.34$ $= 29.4066$	1962.49 $* 0.12$ $= 235.4988$	$V(X)$ $= \sum((x_i - E(X))^2 * P(X = x_i))$ $\approx 133.10 + 29.41 + 235.5$ $= 398.01$

The average revenue per passenger is $\boldsymbol{E(X) = \$15.7}$. The standard deviation is $\boldsymbol{\sigma} = \sqrt{V(X)} = \sqrt{398.01} \approx \$\mathbf{19.95}$.

(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.

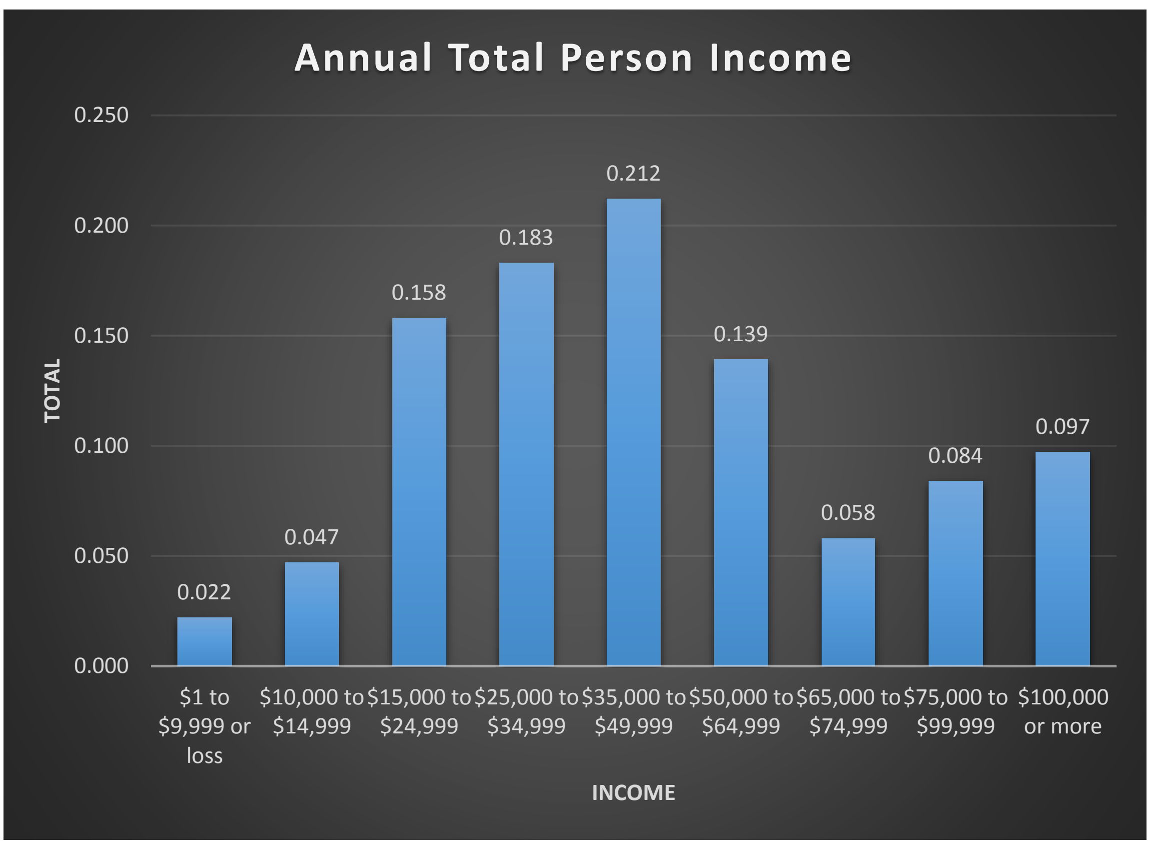
Answer: Assuming that no one checked in more than 2 pieces of baggage or paid any other extra fees for their baggage (ex: overweight), the airline should be expecting to make about $E(120X) = 120 * \text{E(X)} = 120 * \$15.7 = \$\mathbf{1884}$ in revenue from a flight of 120 passengers. With the standard deviation from a flight of 120 passengers is $\sigma = \sqrt{120V(X)} = \sqrt{120 * 398.01} = \sqrt{47{,}761.2} \approx \$\mathbf{218.54}$.

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.

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.

Answer:



As we look at the graph on the top we see that the data of total personal income is skewed to the right. The median is between \$35,000 to 49,999 income. Many people has an income higher than \$14,999. Only 2.2% of people earn less than \$10,000.

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

Answer: The probability of a randomly chosen US resident who makes less than \$50,000 per year is 62.2%.

$$P(Income < \$50{,}000) = 2.2\% + 4.7\% + 15.8\% + 18.3\% + 21.2\% = 62.2\%$$

(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.

Answer: Assuming that there is no relation between gender and income, meaning they are independent. In other words I assumed that 41% of the workforce and each income group is female. Using that assumption the probability of a randomly chosen US resident who makes less than \$50,000 per year and is a female is 25.502%.

$$P(Income < \$50{,}000 \text{ and female}) = P(Income < \$50{,}000\ ) * P(\text{ female})$$
$$= 0.622 * 0.41 = 0.25502 = 25.502\%$$

(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: The assumption I made in part (c) is invalid since we see that there is a relation between income and gender, they are dependent. Since we learn that 71.8% of females make less than \$50,000 per year, not like where gender and income are independent where 62.2% of the both gender makes less than \$50,000 per year.