Homework 3 : Probability

Ramnivas Singh

02/19/2021

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?
b. getting a sum of 5?
c. getting a sum of 12?

Answer: (a) 0, each dice has smallest number as 1 (b) First dice to roll 4 times to get (1,2,3,4) Second dice to roll 4 times to get (1,2,3,4). Sum of these numbers will make 5 1+4=5 2+3=5 3+2=5 4+1=5 Total combinations… (6)(6)=36

4/36=1/9
(c) It should be possibility of getting 6 and 6 which is 1/6*1/6=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?

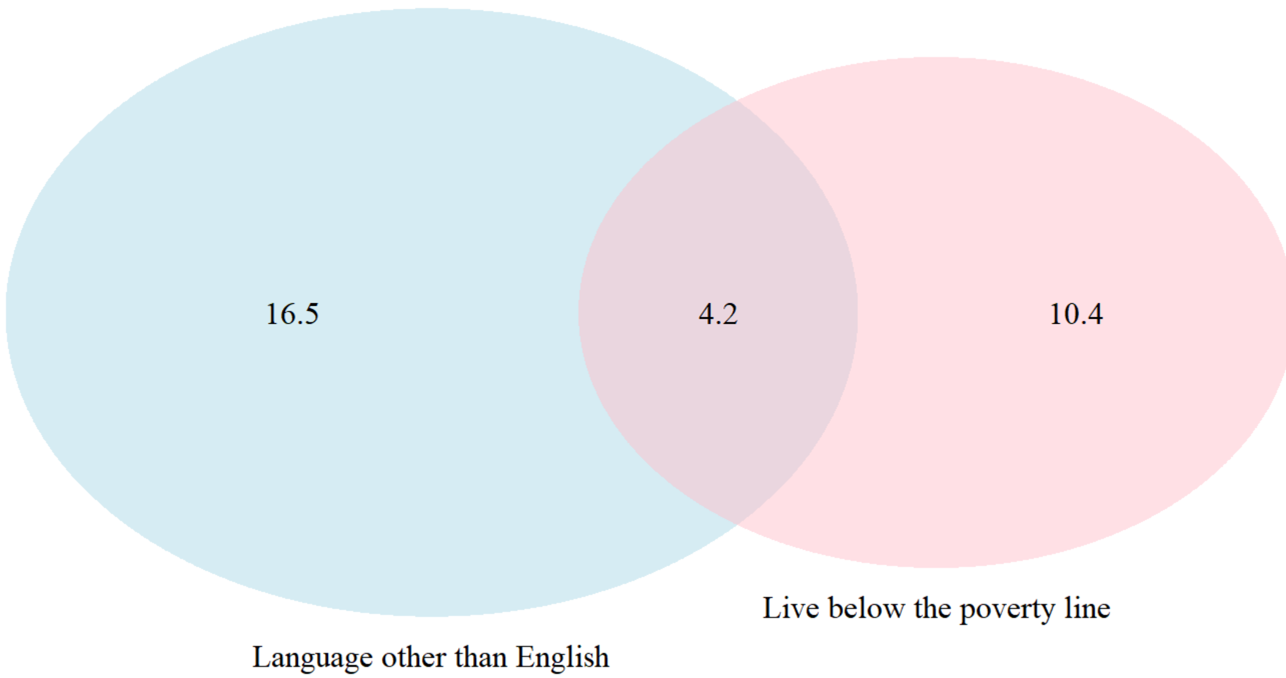
Answers: (a) No, 4.2% fall into both categories. They live below the poverty line and speak a language other than English at home.
(b)

```
library(VennDiagram)
```

```
## Loading required package: grid
```

```
## Loading required package: futile.logger
```

```
grid.newpage()
draw.pairwise.venn(14.6, 20.7, 4.2, category = c("Live below the poverty line", "Language other
 than English"), lty = rep("blank",
    2), fill = c("pink","light blue"),alpha = rep(0.5, 2), cat.pos = c(0,0), cat.dist = rep(0.02
5, 2))
```




```
## (polygon[GRID.polygon.1], polygon[GRID.polygon.2], polygon[GRID.polygon.3], polygon[GRID.poly
gon.4], text[GRID.text.5], text[GRID.text.6], text[GRID.text.7], text[GRID.text.8], text[GRID.te
xt.9])
```

c. P(below the poverty)−P(below the poverty and speak English at home)=14.6−4.2=10.4%
d. P(below the poverty)+P(below the poverty and speak a foreign language at home)-P(below the poverty and speak a foreign language at home)=14.6+20.7−4.2=31.1%
e. 100-(P(below the poverty)+P(below the poverty and speak a foreign language at home)-P(below the poverty and speak a foreign language at home))=100-(14.6+20.7−4.2)=68.9%
f. P(below the poverty) * P(speak foreign language) = 0.146*0.207= 0.030 which does not equal P(below the poverty and speak foreign language) = 0.042, therefore the events are dependent.

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.

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.

Answers: (a) P(self male with blue eyes) + P(partner female with blue eyes) - P(both male female with blue eyes) = 114/204 + 108/204 - 78/204 = 70.59% (b) P(female=blue|male=blue) = 78/114 = 0.68 (c) P(female with blue eyes| male with brown eyes) = 19/54 = 35.2% P(female with blue eyes| male with green eyes) = 11/36 = 30.6% (d) No .From the observation above, males with blue eyes favorite females with blue eyes.

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.

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) P(Hardcover first) * P(paperback fiction) = (28/95) * (59 / 94) = 18.50% (b) P(first:hard cover fiction and second: hard cover) + P(first: papaerback fiction and second: hard cover) = (13/95)(27/94) + (59/95)(28/94) = 22.43% (c) P(first: fiction and second: hardcover) = (72/95) * (28/95) = 22.338% (d) Because the difference of (b) and (c) is replacement of the book which changes the denominator from 1/(9594) to (1/9595). The impact is small.

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.

Answers:
(a)

	X	**P(X)**	**X*P(X)**
No checked bag	0	.54	0 x 0.54 = 0
1 checked bag	$25	.34	25 x 0.34 = 8.5
2 checked bag	$25 + 30 = $60	.12	60 x 0.12 = 7.2

Average revenue per passenger, E(X) = 0+8.5+7.2=15.7

```
# Fees charges for o pieces of Luggage in dollars
bag0 <- 0
# Fees charges for 1st Luggage in dollars
bag1 <- 25
# Fees charges for 2nd Luggage in dollars
bag2 <- bag1 + 35
bag_fees <- c(bag0,bag1,bag2)
baggage_per_bag <- c(0.54, 0.34, 0.12)
# Find Expected value for each x_i
expected_revenue <- bag_fees * baggage_per_bag
Ex <- sum(expected_revenue)
baggage_variance <- bag_fees - Ex
# Calculate the Variance^2 and P(X=x_i)
baggage_EVariance <- baggage_variance^2 * baggage_per_bag
Variance2 <- sum(baggage_EVariance)
# Find the standard deviation by calculating the square root of the variance
sd <- Variance2^(1/2)
sd
```

```
## [1] 19.95019
```

(b) Revenue = 120 * $15.70 = $1,884 ± $20 (or 19.95 rounded up)

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.

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.

Answers:
(a) Right Skewed. (b) P(US resident makes less than $50,000 per year) = (2.2 + 4.7 + 15.8 + 18.3 + 21.2) = 62.2% (c) Assuming income for male and female is similar in all the categories. P(US resident makes less than $50,000 per year and female) = P(US resident makes less than $50,000 per year)* 41% = 62.2* 41/100= 25.50% (d) The assumption in (c) is not correct. If 71.8% of females make lesser than $50k but above model shows 25.5%