Week 2 Graded Problems

James Topor

Chapter 2, Problems 2.6, 2.8, 2.20, 2.30, 2.38, 2.44

―――-

2.6

a) It is not possible to get a sum of 1 when rolling a pair of fair dice. Therefore the probability is **zero**.

b) The probability of getting a sum of five can be found by counting the number of instances in which a sum of five can occur and dividing that count by the total number of possible outcomes of rolling two fair dice. The ways in which we can observe a sum of five are as follows, where 'D1' is the value of the first die and 'D2' is the value of the second die:

(D1 = 1, D2 = 4)
(D1 = 4, D2 = 1)
(D1 = 2, D2 = 3)
(D1 = 3, D2 = 2)

So we have 4 possible ways to obtain a sum of 5. We also know that there are 6!= 36 possible sums obtainable by rolling two fair dice, so our calculation is:

$4/36 =$ **1/9**

c) There is only one possible way in which to obtain a sum of 12 when rolling two fair dice: rolling 6 and 6. As such, our probability for getting a sum of 12 is **1/36**.
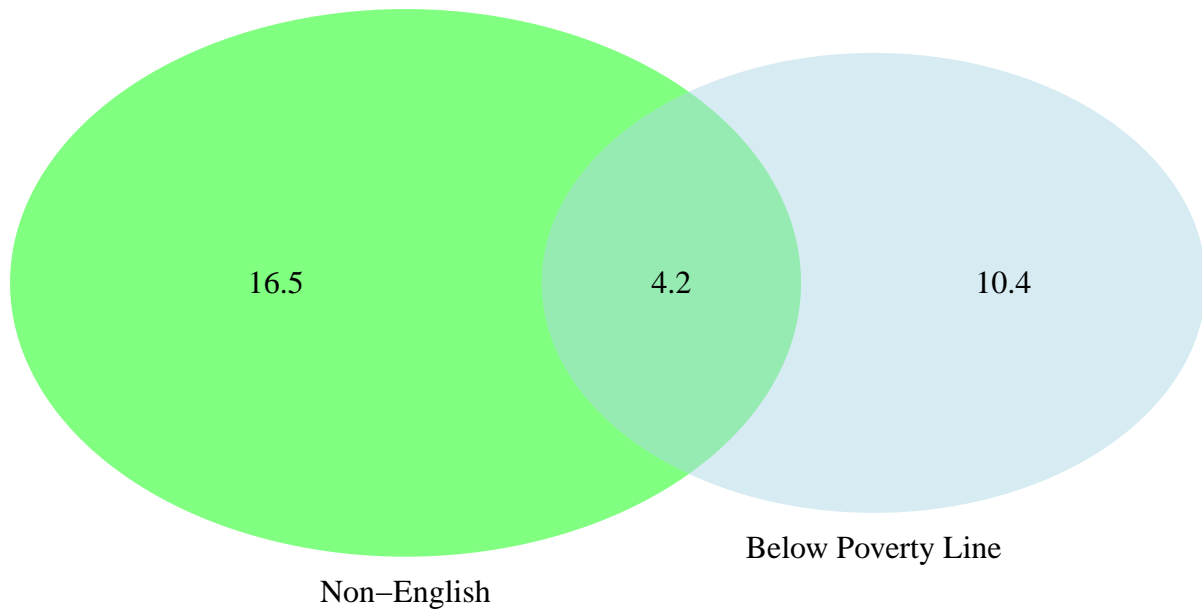
―――-

2.8

a) No they are not disjoint since 4.2% of respondents claim to belong to both groups.

b) Venn diagram

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

16.5   4.2   10.4

Below Poverty Line

Non–English

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

c) The percentage of Americans that live BELOW the poverty line AND speak english at home is:
$.146 - .042 = .104 = 10.4\%$

d) The percentage of Americans that live BELOW the poverty line OR speak a foreign language at home is:

$.146 + .207 - .042 = .311 = 31.1\%$

e) The percentage of Americans that live ABOVE the poverty line AND speak only english at home is found by computing 1 - (% below poverty line OR speak foreign language at home). We calculated (% below poverty line OR speak foreign language at home) in part **d)** above as .311, so our equation is:

$(1 - .311) = .689 = 68.9\%$

- We could also approach this by finding the proportion of Americans speaking English at home, which is $1 - .207 = .793$ and then subtracting the percent of Americans below the poverty line who speak only English at home, which we know to be .104 from part **c)** above: $.793 - .104 = .689 = 68.9\%$

f) We can determine whether the event that someone lives below the poverty line independent of the event that the person speaks a foreign language at home by testing for equivalence of $P(A) * P(B) = P(A, B)$:

- P(below pov) $= .146 =$ P(A)

- P(for lang) $= .207 =$ P(B)

- P(A) * P(B) $= .146 * .207 = .0302$

- P(A,B) = .042 (given to us in text)

As we can see, $(P(A) * P(B) = .0302)! = (P(A, B) = .042)$ so these events are NOT independent.

———-

2.20

a) The probability that a male has blue eyes is 114 / 204.
The probability that a female has blue eyes is 108 / 204.
The probability that both the male and female of a couple have blue eyes is 78/204.

Therefore, the probability that a randomly chosen male respondent or his partner has blue eyes is:

$$(114/204) + (108/204) - (78/204) = (144/204) =$$

[1] 0.7058824

b) The probability that a randomly chosen male respondent with blue eyes has a partner with blue eyes is:

$$(FemaleBlue|MaleBlue) = 78/114 =$$

[1] 0.6842105

c) The probability that a randomly chosen male respondent with brown eyes has a partner with blue eyes is:

$$(FemaleBlue|MaleBrown) = 19/54 =$$

[1] 0.3518519

The probability of a randomly chosen male respondent with green eyes having a partner with blue eyes is:

$$(FemaleBlue|MaleGreen) = 11/36 =$$

[1] 0.3055556

d) The eye colors of male respondents and their partners do not appear to be independent since the conditional probabilities we've calculated are not equal to the corresponding joint probabilities.

For example, if we consider the instance of (Male Blue AND Female Blue) we have $78/204 = .38$.

However when we calculate the conditional probability for (Female Blue | Male Blue) we have $78/114 = .684$ (see part b above). As such, we don't have independence. This can be further assessed by simply reviewing the raw data we were provided: each male of any given eye color clearly appears to prefer females of the same eye color.

2.30

a) The probability of drawing a hardcover book first then a paperback fiction book second when drawing without replacement is:

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

[1] 0.1849944

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

The probability of drawing a fiction book first is (72/95). However, we could have drawn either a hardcover or paperback fiction book first, so we need to calculate the probability of drawing either a HC or PP fiction book first, then calculate the probability of drawing a HC given that a fiction book was drawn first.

The probability of drawing a HC fiction book first is: P(Fiction, HC)= 13/95

If we drew without replacement, the probability of then drawing a hardcover would be: P(HC next) = 27/94, which give us:

(13/95) * (27/94) = 0.0393

[1] 0.03930571

The probability of drawing a paperback fiction book first is: P(Fiction, PP) = 59/95

If we drew without replacement, the probability of then drawing a hardcover would be: P(HC next) = 28/94, which gives us:

(59/95) * (28/94) = .185

[1] 0.1849944

Adding these two together gives us the total probability of drawing any type of fiction book first (either hardcover or paperback) followed by a hardcover if we drew without replacement:

.0393 + .185 = .2243

[1] 0.2243

c) Find the probability of drawing a fiction book first and then a hardcover book second, when drawing WITH replacement:

The probability of drawing a fiction book first is:

P(fiction) = 72/95

If drawing with replacement, the probability of then drawing a HC book is:

P(HC) = 28/95

Therefore, the probability of drawing a fiction book first and then a hardcover book second is:

(72/95) * (28/95) = .223

[1] 0.2233795

d) The results are somewhat similar. This is likely due to the fact that the sample size is large enough to preclude any large impacts on the resulting probabilities when drawing without replacement vs. drawing with replacement.

_____-

2.38 We are given the following:

The fees charged per the number of bags per passenger:
0 bags = \$0
1 bag = \$25
2 bags = \$35 + \$25 (\$25 for the first bag) = \$60

The probabilities of a passenger having zero, one, or two bags:
$P(0) = .54$
$P(1) = .34$
$P(2) = .12$

Using this information we can create a dataframe in R to assist with our calculations:

```r
# create a dataframe with the info given in the problem
n_bags <- c(0, 1, 2)
bag_fees <- c(0, 25, 60)
bag_prob <- c(0.54, 0.34, 0.12)
bags_df <- data.frame(n_bags, bag_fees, bag_prob)

# add a column tracking the weighted revenue for each number of bags
bags_df$wght_rev <- bags_df$bag_prob * bags_df$bag_fees

# display dataframe
bags_df
```

```
##   n_bags bag_fees bag_prob wght_rev
## 1      0        0     0.54      0.0
## 2      1       25     0.34      8.5
## 3      2       60     0.12      7.2
```

a) To calculate the average revenue per passenger we can simply sum the weighted revenue amounts we computed while creating the dataframe. Specifcally, the average revenue per passenger is: $(.54 * 0) + (.34 * 25) + (.12 * 60)$:

```r
meanRev = sum(bags_df$wght_rev)
meanRev
```

```
## [1] 15.7
```

To calculate the standard deviation we first calculate the variance and then take its square root:

```r
varRev <- sum(bags_df$bag_prob * ((bags_df$wght_rev - meanRev)^2))
varRev
```

```
## [1] 159.4002
```

```r
sdRev <- sqrt(varRev)
sdRev
```

```
## [1] 12.62538
```

b) If we have 120 passengers, we should expect approximately $1884 in baggage revenue for the flight.

```
meanRev * 120
```

```
## [1] 1884
```

The standard deviation will be $138.30, We calculate this result by multiplying the number of passengers (120) by the square of the standard deviation per passenger:

```
# NOTE: See page 59 of Chap 2 slide deck for formula to use here
varRev120 <- 120 * (sdRev^2)
sdRev120 <- sqrt(varRev120)
sdRev120
```
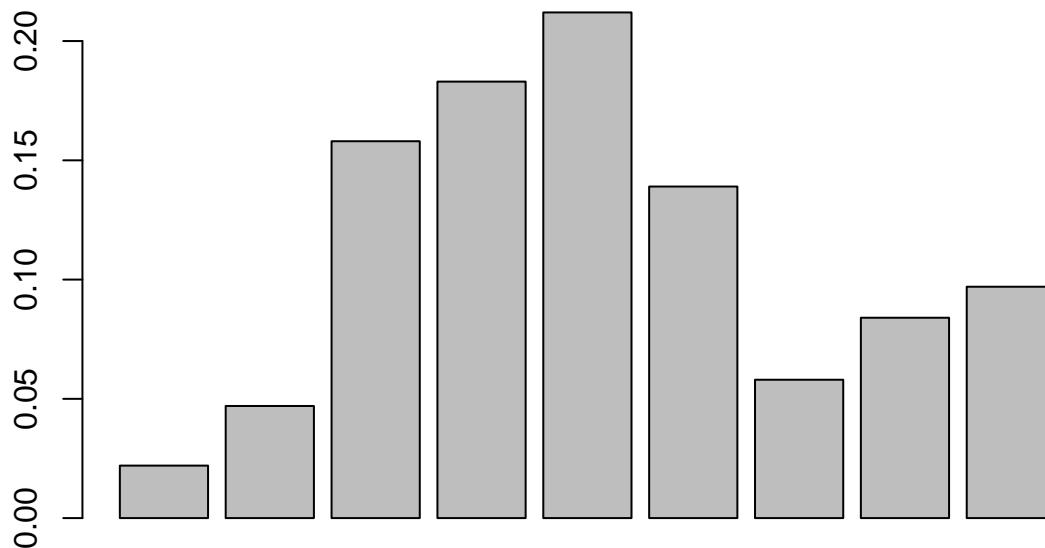
```
## [1] 138.3041
```

This calculation for the standard deviation is valid assuming each individual passenger's bag count is independent of every other passenger's bag count.

2.44

a) We can examine the distribution of total personal income via a bar plot:

```
vec <- c(.022, .047, .158, .183, .212, .139, .058, .084, .097)
barplot(vec)
```

The distibution is bimodal as shown in the above barplot.

b) The probability that a randomly chosen person earns < \$50K per year is:

.022 + .047 + .158 + .183 + .212 = .622

```
.022 + .047 + .158 + .183 + .212
```

```
## [1] 0.622
```

c) We would assume that the proportion of female respondents is .41 as stated in the problem and that income is independent of gender. We already know the probability of a randomly chosen person earning < \$50K from part **b)** above as .622. This gives us the probability that a randomly chosen US resident makes less than \$50,000 per year and is female as:

.622 * .41 = .255.

```
.622 * .41
```

```
## [1] 0.25502
```

d) If the proportion of females earning less than \$50,000 per year is actually 71.8% then the assumption of income being independent of gender made in part **c)** above is incorrect since independence would indicate that only .41 of women should earn < \$50K.