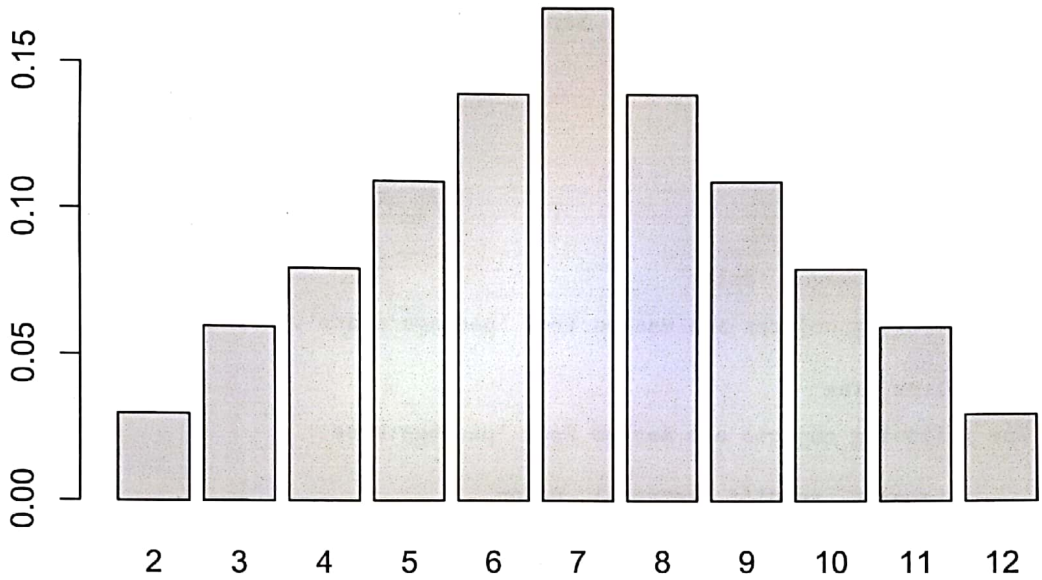
homework_ch2_JimmyNg

Jimmy Ng

September 12, 2018

Question 2.6

```
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

dice <- c(1:6)
combo <- expand.grid(dice, dice) %>%
        mutate(sum = Var1 + Var2)
round(prop.table(table(combo$sum)), 2) %>%
        print %>%
        barplot

##
##    2    3    4    5    6    7    8    9   10   11   12
## 0.03 0.06 0.08 0.11 0.14 0.17 0.14 0.11 0.08 0.06 0.03
```



(a) getting a sum of 1 is 0%
(b) getting a sum of 5 is 11%
(c) getting a sum of 12 is 3%

Question 2.8

(a) No, they are not disjoint. In fact, 4.2% fall into both categories, meaning someone can live below poverty line and speak a foreign language at home.
(b) please see the attached appendix (hand drawn and scanned)
(c) 10.4% of all Americans live below poverty line and only speak English at home.
(d) 4.2% of all Americans live below poverty line and speak foreign language at home; 10.4% live below poverty line and don't speak foreign language at home; and 16.5% live above poverty line and speak foreign language at home. Together (4.2% + 10.4% + 16.5%) that gives us approximately 31.1% of all Americans either live below poverty line or speak a foreign language at home.
(e) 68.9% of all Americans live above the poverty line and only speak English at home.
(f) P(X = below_poverty) = 14.6%; P(X = speak_foreign_language_at_home) = 20.6%; and the P(X = below_povery and speak_foreign_language_at_home) = 4.2%. The conditional probability of someone living below poverty line given he/she speaks another language at home is P(below_povery and speak_foreign_language_at_home) / P(speak_foreign_language_at_home). We can substitute the numbers (4.2% / 20.6%) and got 20.3%. Since the result of 20.3% > P(X = below_poverty) which is 14.6%, living below poverty line is not independent of the event that the person speaks another language at home. In other words, these two events are dependent (related).

Question 2.20

(a) The probability of randomly selecting a blue eye male respondent = (114 / 204) plus the probability of finding a blue eye partner (female) = (108 / 204) minus the ones in common (blue eyes male AND blue eyes partner) = (78 / 204) and together that gives us 70.6%.
(b) Using Bayes' Theorem, the probability that a female partner has blue eyes given we randomly select a blue eye male respondent is 68.4% (please see the appendix for calculation).
(c) Using Bayes' Theorem, the probability that a female partner has blue eyes given we randomly select a brown eye male respondent is 35.2%, whereas the probability that a female partner has blue eyes given we randomly select a green eye male respondent is 30.6% (please see the appendix for calculation).
(d) They are not independent. From above answers, we see that a blue eye male has a higher probability associated with a blue eye partner. Randomly selecting a male respondent with blue eyes give us 68.4% chance that his partner has blue eyes, whereas randomly selecting a brown or green eyes male respondent only gives us 35.2% and 30.6% chance respectively.

Question 2.30

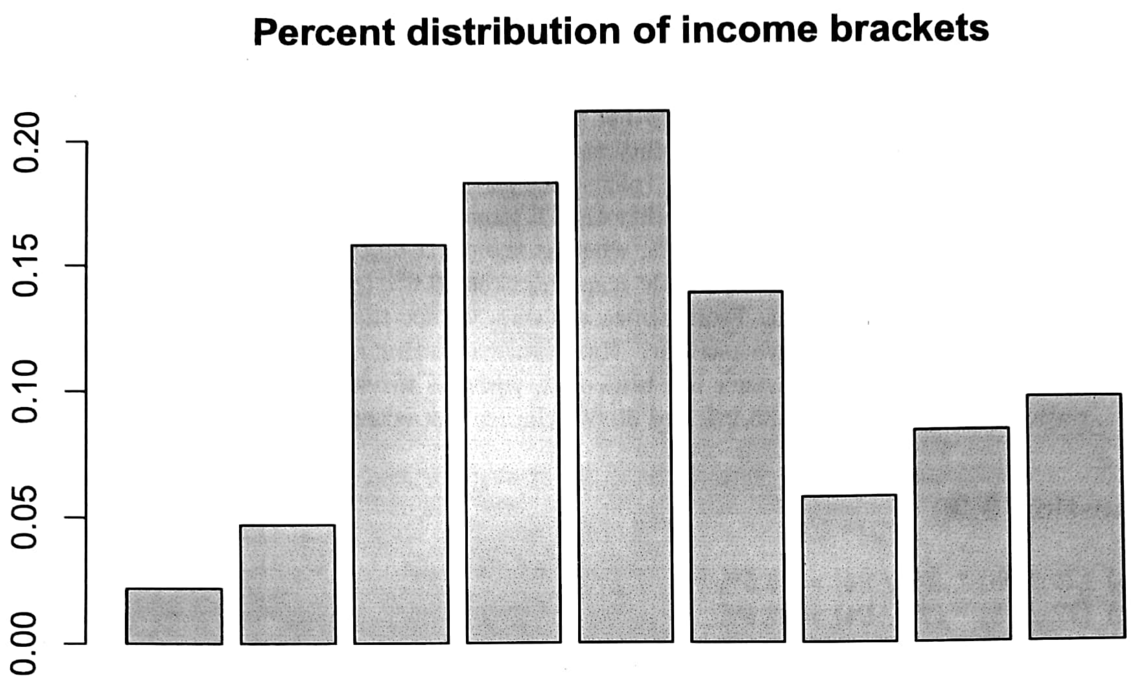
(a) (28 / 95) * (59 / 94) = 18.5%
(b) (72 / 95) * (28 / 94) = 22.6%
(c) (72 / 95) * (28 / 95) = 22.3%
(d) It is very similar because the drawing with replacement has made very small increase to the denominator of the second probability (from 94 to 95). Such small change will not result in dramatic shift in overall probability, especially when we have large enough sample size (with or without replacement won't matter - imagine if we have 100000 samples).

Question 2.38

(a) Please see the appendix for detail calculation. The average revenue per passenger (mu) is $15.7; the variance is 389.01; and the standard deviation is 19.95019.
(b) 120 * $15.7 = $1884. We will still retain the SD from (a).

Question 2.44

```
percent <- c(0.022, 0.047, 0.158, 0.183, 0.212, 0.139, 0.058, 0.084, 0.097)
barplot(percent, main = "Percent distribution of income brackets")
```



(a) Please see above barplot of the percent distribution of income brackets. This does not appear to be a normal distribution. It has a right skew with very few of high networth or ultra high networth individuals. Most likely it is a bimodal distribution.

(b) 62.2%

(c) 62.2% * 50% = 31.1%. We have to multiply 62.2% by 50% because we should expect that the chance of randomly finding a female resident is equal to the chance of finding a male resident.

(d) 71.8% * 50% = 35.9%. The chance of randomly finding a female resident is still 50% (equally likely to find a male resident); the probability is now updated and known to us that 71.8% of the females make less than 50k. Thus, we multiply the two probabilities together and that gives us 35.9%. This is not too much different from (c) and therefore (c) is still valid. From the beginning, we should have applied a weight to the odds since only 41% of the data that we collected came from females and that would have affected the calculatoin of probability.

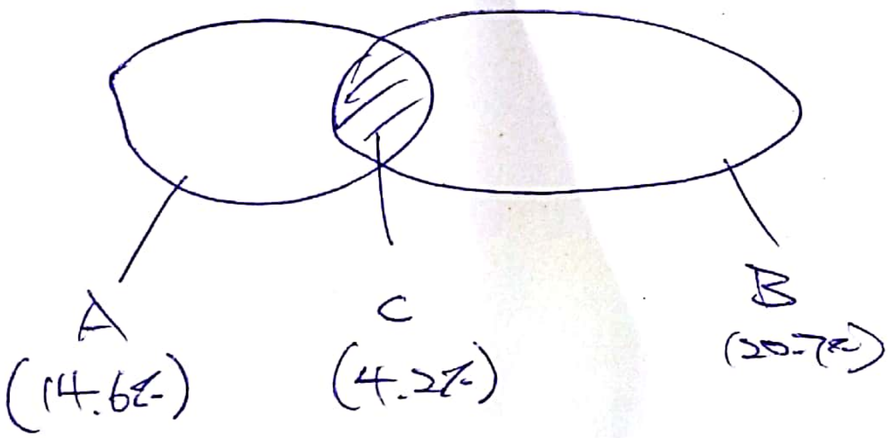
Question 2.8

2.8 (b) Venn Diagram

A - 14.6% below poverty line

B - 20.7% speak another language at home

C - 4.2% fall into both categories



TREE DIAGRAM

Poverty

speak another language at home

Below 14.6%
Y — 4.2%
N — 10.4%

Above 85.4%
Y — 16.5%
N — 68.9%

Question 2.20

2.20(b) — Bayesian Theorem

Female Blue Eye		Male Blue Eye	
Y	108/204	Y	78/108
		N	30/108
N	96/204	Y	36/96
		N	60/96

$$P(F^+ \mid M^+)$$

$$= \frac{P(M^+ \mid F^+)\ P(F^+)}{P(M^+ \mid F^+)\ P(F^+) + P(M^+ \mid F^-)\ P(F^-)}$$

$$= \frac{(78/108) \times (108/204)}{(78/108) \times (108/204) + (36/96)(96/204)}$$

$$= \frac{0.382353}{(0.382353) + (0.176471)}$$

$$\approx 68.4\%$$

2.20(c) Brown Eye Male

Female Blue		Male Brown	
Y	108/204	Y	19/108
		N	89/108
N	96/204	Y	35/96
		N	61/96

$$P(F^+ \mid M^+)$$

$$= \frac{P(M^+ \mid F^+)\ P(F^+)}{P(M^+ \mid F^+) P(F^+) + P(M^+ \mid F^-) P(F^-)}$$

$$= \frac{(19/108) \times (108/204)}{(19/108) \times (108/204) + (35/96)(96/204)}$$

$$= \frac{0.093137}{(0.093137) + (0.171569)}$$

$$\approx 35.2\%$$

2.20 (c) Green Eye Male

$$P(F^+ \mid M^+)$$

Female Blue		Male Green	
Y	108/204	Y	11/108
		N	97/108
N	96/204	Y	25/96
		N	71/96

$$= \frac{P(M^+ \mid F^+)\ P(F^+)}{P(M^+ \mid F^+)P(F^+) + P(M^+ \mid F^-)P(F^-)}$$

$$= \frac{(11/108) \times (108/204)}{(11/108)(108/204) + (25/96)(96/204)}$$

$\approx 30.6\%$

Question 2.38

i	Baggages # 0	1	2	Total
x_i	\$ 0	\$ 25	\$ 60	
$P(X=x_i)$	54%	34%	12%	
$x_i \cdot P(X=x_i)$	0	8.5	7.2	μ 15.7
$x_i - \mu$	-15.7	9.3	44.3	
$(x_i - \mu)^2$	246.49	86.49	1962.49	
$(x_i-\mu)^2 \cdot P(X=x_i)$	133.1046	29.4066	235.4988	σ^2 398.01

$\mu = 15.7$

$\sigma^2 = 398.01$

$\sigma = 19.95019$