Chapter 2 - Probability

CUNY MSDA - IS606 - Homework 2

Completed by: Duubar Villalobos Jimenez mydvtech@gmail.com

February 12, 2017

Homework

OpenIntro Statistics

```
Practice: 2.5, 2.7, 2.19, 2.29, 2.43
Graded: 2.6, 2.8, 2.20, 2.30, 2.38, 2.44
```

2.6 Dice rolls.

(a) getting a sum of 1?

Let's build a probability table

```
# Dice 1 Sample space
dice1 <- c(1,2,3,4,5,6)
Pdice1 <- c(1/6,1/6,1/6,1/6,1/6,1/6)
# Dice 2 sample space
dice2 <- c(1,2,3,4,5,6)
Pdice2 <- c(1/6,1/6,1/6,1/6,1/6,1/6)
# Possible outcomes sums
dsums <- c(2,3,4,5,6,7,8,9,10,11,12)
# Let's calculate the probabilities for each sum
P <- c(1/36, 2/36, 3/36, 4/36, 5/36, 6/36, 5/36, 4/36, 3/36, 2/36, 1/36)
# Let's build the table
dicef <- data.frame(dsums, P)
names(dicef) <- c("Sums", "Probability")
dicef
```

```
##    Sums Probability
## 1     2  0.02777778
## 2     3  0.05555556
## 3     4  0.08333333
## 4     5  0.11111111
## 5     6  0.13888889
## 6     7  0.16666667
## 7     8  0.13888889
## 8     9  0.11111111
## 9    10  0.08333333
## 10   11  0.05555556
## 11   12  0.02777778
```

Based on the possibilities, there's no possible way to obtain a sum of 1, since none of the faces on the dice has a value of zero.

Hence: P(X+Y =0) = 0.

(b) getting a sum of 5?

For this, we have different possibilities:

Let's say X represent the first die and Y the second die.

The possible outcomes will be as follows:

Outcome 1 = P(X=1) * P(Y=4) = 1/6 * 1/6 = 1/36

Outcome 2 = P(X=2) * P(Y=3) = 1/6 * 1/6 = 1/36

Outcome 3 = P(X=3) * P(Y=2) = 1/6 * 1/6 = 1/36

Outcome 4 = P(X=4) * P(Y=1) = 1/6 * 1/6 = 1/36

There are 4 possible ways to obtain a sum of 5.

P(X+Y = 5) = Outcome 1 + Outcome 2 + Outcome 3 + Outcome 4

P(X+Y = 5) = 1/36 + 1/36 + 1/36 + 1/36

P(X+Y = 5) = 4/36

(c) getting a sum of 12?

For this, we have different possibilities:

Let's say X represent the first die and Y the second die.

The possible outcomes will be as follows:

Outcome 1 = P(X=6) * P(Y=6) = 1/6 * 1/6 = 1/36

P(X+Y = 1) = 1/36

2.8 Poverty and language.

Let's define as follows:

A: Americans living below the poverty line.

F: Speak a language other than English (foreign language) at home.

P(A) = 14.6% = 0.146

P(F) = 20.7% = 0.207
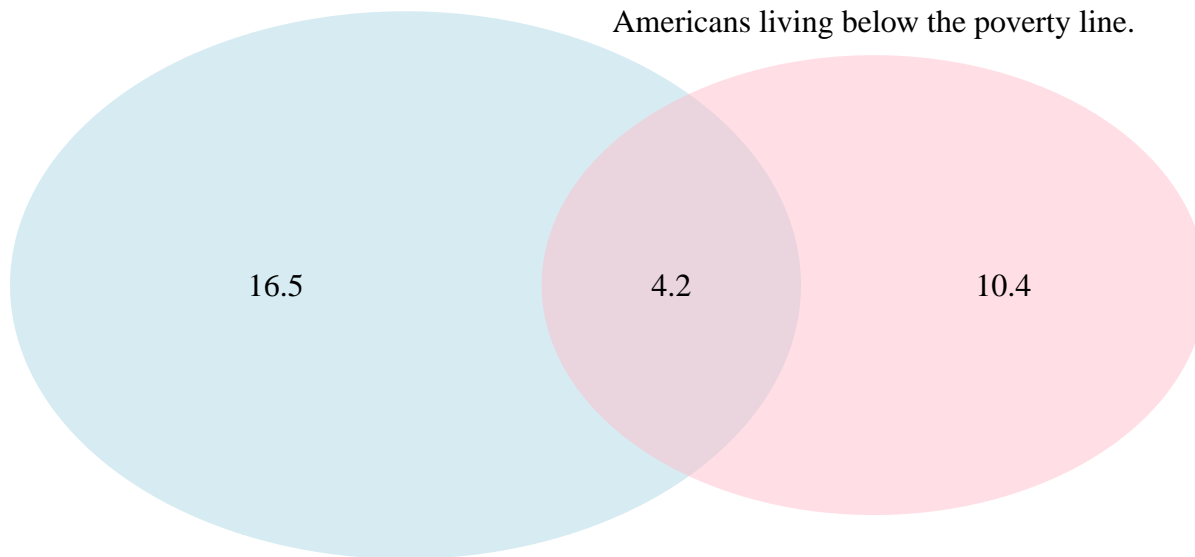
P(A and F) = 4.2% = 0.042

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

No, they are not disjoint since both are happening mutually.

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

Speak a language other than English (foreign language) at home.

Americans living below the poverty line.

16.5   4.2   10.4

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

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

P(A and Speak English at Home) = P(A) - P(A and F)

P(A and Speak English at Home) = 0.146 - 0.042

P(A and Speak English at Home) = 0.104

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

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

P(A or F) = P(A) + P(F) - P(A and F)

P(A or F) = 0.146 + 0.207 - 0.042

P(A or F) = 0.311

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

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

AC: Complement of A

FC: Complement of C

P(AC and FC) = 1 - (P(A) + P(F) - P(A and F))

3

P(AC and FC) = 1 - 0.311

P(AC and FC) = 0.689

Answer: The percent of Americans live above the poverty line and only speak English at home is 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?

Let's build our independence condition:

P(A and F) = P(A) * P(F)

0.042 = 0.146 * 0.207

Since 0.042 ≠ 0.030

We conclude that these events a not independent, since the independency multiplication rule is not satisfied.

2.20 Assortative mating.

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

By applying the general addition rule:

P(M_Blue or F_Blue) = P(M_Blue) + P(F_Blue) - P(M_Blue and F_Blue)

P(M_Blue or F_Blue) = 108/204 + 114/204 - 78/204

P(M_Blue or F_Blue) = 0.7059

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

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

P(F_Blue | M_Blue) = P(F_Blue and M_Blue) / P(M_Blue)

P(F_Blue | M_Blue) = (78/204) / (114 / 204)

P(F_Blue | M_Blue) = 0.6842

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

P(F_Blue | M_Brown) = P(F_Blue and M_Brown) / P(M_Brown)

P(F_Blue | M_Brown) = (23/204) / (54/204)

P(F_Blue | M_Brown) = 0.4259

Answer: The probability that a randomly chosen male respondent with brown eyes has a partner with blue eye is 42.59%

P(F_Blue | M_Green) = P(F_Blue and M_Green) / P(M_Green)

P(F_Blue | M_Green) = (13/204) / (36/204)

P(F_Blue | M_Green) = 0.3611

Answer: The probability of a randomly chosen male respondent with green eyes having a partner with blue eyes is 36.11%

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

Let's build our independence condition:

P(M_Blue and F_Blue) = P(M_Blue) * P(F_Blue)

(78/204) = (114/204) * (108/204)

Since $0.3824 \neq 0.2958$

We conclude that these events a not independent, since the independency multiplication rule is not satisfied.

2.30 Books on a bookshelf.

(a) Find the probability of drawing a hardcover book first then a paperback fiction book second when drawing without replacement.

P(First H and Second being P_fiction) = P(First H) * P(Second being P_fiction)

P(First H and Second being P_fiction) = (28/95) * (59/94)

P(First H and Second being P_fiction) = = 0.185

Answer: The probability of drawing a hardcover book first then a paperback fiction book second when drawing without replacement is 18.5%

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

P(Fiction and second being H) = P(Fiction) * P(second being H)

P(Fiction and second being H) = (72/95) * (28/94)

P(Fiction and second being H) = 0.2258

Answer: The the probability of drawing a fiction book first and then a hardcover book second, when drawing without replacement is 22.58%

(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.

P(Fiction and second being H) = P(Fiction) * P(second being H)

P(Fiction and second being H) = (72/95) * (28/95)

P(Fiction and second being H) = 0.2234

Answer: The probability of drawing a fiction book first and then a hardcover book second, when drawing with replacement is 22.34%

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

Answer: In this case the answers are very similar, this is because when the possible events are considerable large, the outcome will not be affected by much when there's no replacement in random drawings.

2.38 Baggage fees.

(a) Build a probability model, compute the average revenue per passenger, and compute the corresponding standard deviation.

```
# Number of bags
bags <- c(0, 1, 2)
# Fees charges for o pieces of luggage in dollars
Luggage_0 <- 0
# Fees charges for 1st luggage in dollars
Luggage_1 <- 25
# Fees charges for 2nd luggage in dollars
Luggage_2 <- Luggage_1 + 35

# Baggage fees table
baggage_fees <- c(Luggage_0,Luggage_1,Luggage_2)

# Percentage of passangers that check baggage in decimal form
baggage_percent_per_pax <- c(0.54, 0.34, 0.12)

# Find Expected value for each x_i
E_revenue <- baggage_fees * baggage_percent_per_pax

# Find the overall Expected value AKA mu
Ex <- sum(E_revenue)

# Expected Revenue per passenger
Ex
```

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

```
# Create mu collumn
mu <- c(Ex, Ex, Ex)

# Create data frame
baggage <- data.frame(bags, baggage_fees, baggage_percent_per_pax)

# Find  The variance_i of x_i and mu
baggage_variance <- baggage_fees - Ex

# Calculate the Variance^2 and P(X=x_i)
baggage_EVariance <- baggage_variance^2 * baggage_percent_per_pax

# Create visual representation of the table
baggage <- cbind(baggage, E_revenue, mu, baggage_variance, baggage_variance^2,  baggage_EVariance)

# Name columns for the baggage data frame
names(baggage) <- c("bags", "x_i", "P(X=x_i)", "E(X_i)", "mu", "Variance", "Variance^2", "Variance^2*P(X

# View Table
```

```
baggage
```

```
##   bags x_i P(X=x_i) E(X_i)   mu Variance Variance^2 Variance^2*P(X=x_i)
## 1    0   0     0.54    0.0 15.7    -15.7     246.49             133.1046
## 2    1  25     0.34    8.5 15.7      9.3      86.49              29.4066
## 3    2  60     0.12    7.2 15.7     44.3    1962.49             235.4988
```

```r
# Find the overall value for the Variance^2
Variance2 <- sum(baggage_EVariance)

# Print the overall Variance
# Variance2

# Find the standard deviation by calculating the square root of the variance
sd <- Variance2^(1/2)

# Prinr the standard deviation
sd
```

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

(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.

```r
# Number of bags
bags <- c(0, 1, 2)
# Fees charges for o pieces of luggage in dollars
Luggage_0 <- 0
# Fees charges for 1st luggage in dollars
Luggage_1 <- 25
# Fees charges for 2nd luggage in dollars
Luggage_2 <- Luggage_1 + 35

# Baggage fees table
baggage_fees <- c(Luggage_0,Luggage_1,Luggage_2)

# Percentage of passangers that check baggage in decimal form
baggage_percent_per_pax <- c(0.54, 0.34, 0.12)

# Number of passengers
pax <- 120

# Find Expected value for each x_i
E_revenue <- baggage_fees * baggage_percent_per_pax * pax

# Find the overall Expected value AKA mu
Ex <- sum(E_revenue)

#Expected revenue
Ex
```

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

```r
# Create mu collumn
mu <- c(Ex, Ex, Ex)

# Create data frame
baggage <- data.frame(bags, baggage_fees, baggage_percent_per_pax)

# Find  The variance_i of x_i and mu
baggage_variance <- baggage_fees - Ex

# Calculate the Variance^2 and P(X=x_i)
baggage_EVariance <- baggage_variance^2 * baggage_percent_per_pax

# Create visual representation of the table
baggage <- cbind(baggage, E_revenue, mu, baggage_variance, baggage_variance^2,  baggage_EVariance)

# Name columns for the baggage data frame
names(baggage) <- c("bags", "x_i", "P(X=x_i)", "E(X_i)", "mu", "Variance", "Variance^2", "Variance^2*P(X

# View Table
baggage
```

```
##   bags x_i P(X=x_i) E(X_i)   mu Variance Variance^2 Variance^2*P(X=x_i)
## 1    0   0     0.54      0 1884    -1884    3549456           1916706.2
## 2    1  25     0.34   1020 1884    -1859    3455881           1174999.5
## 3    2  60     0.12    864 1884    -1824    3326976            399237.1
```

```r
# Find the overall value for the Variance^2
Variance2 <- sum(baggage_EVariance)

# Print the overall Variance
#Variance2

# Find the standard deviation by calculating the square root of the variance
sd <- Variance2^(1/2)

# Prinr the standard deviation
sd
```

```
## [1] 1868.407
```

Answer: The expected revenue for the 120 passengers is $1884.00. The standard deviation will be $1868.41.

2.44 Income and gender.

(a) Describe the distribution of total personal income.

This is a smooth continuous distribution with what seems to be a multi modal shape.

```r
income <- c("$1 to $9,999","$10,000 to $14,999","$15,000 to $24,999","$25,000 to $34,999","$35,000 to $4

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

dist <- data.frame(income, total)
dist
```

```
##               income total
```

```
## 1        $1 to $9,999    2.2
## 2 $10,000 to $14,999    4.7
## 3 $15,000 to $24,999   15.8
## 4 $25,000 to $34,999   18.3
## 5 $35,000 to $49,999   21.2
## 6 $50,000 to $64,999   13.9
## 7 $65,000 to $74,999    5.8
## 8 $75,000 to $99,999    8.4
## 9   $100,000 or more    9.7
```

```
barplot(dist$total, names.arg=income)
```
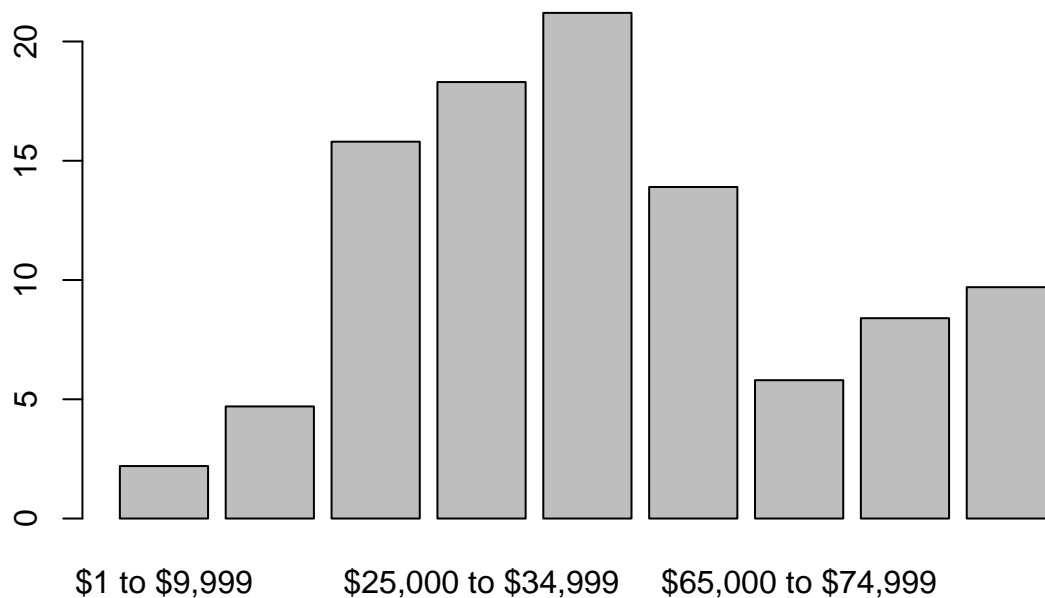

(b) What is the probability that a randomly chosen US resident makes less than $50,000 per year?

P(Resident < $50000) = P($1 to $9,999) + P($10,000 to $14,999) + P($15,000 to $24,999) +P($25,000 to $34,999) + P($35,000 to $49,999)

P(Resident < $50000) = 0.022 + 0.047 + 0.158 + 0.183 + 0.212

P(Resident < $50000) = 0.622

```
sum(dist[1:5,2])
```

```
## [1] 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.

This sample is comprised of 59% males and 41% females.

Assumption:

Since we don't know the relationship between the probability of an income of less than $50,000 and being female.

I will be assuming that they are independent events then P(A and B) = P(A) x P(B).

P(Resident < $50000 and F) = P(Resident < $50000) * P(F)

P(Resident < $50000 and F) = 0.622 * 0.41

P(Resident < $50000 and F) = 0.2550

Answer: The probability that a randomly chosen US resident makes less than $50,000 per year and is female is 25.50%.

(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.

P(Resident < $50000 and F) = P(Resident < $50000) * P(F)

$0.718 = 0.622 * 0.41$

Since $0.718 \neq 0.2550$

We conclude that these events are not independent, since the independency multiplication rule is not satisfied.