Week 3 - Data Science Math Assignment

Nabila Hossain

July 19, 2015

Assignment: Please complete exercises 2.34, 2.40, 2.42, and 2.46 in OpenIntro Statistics (pp. 114-117).

2.34 Card game. Consider the following card game with a well-shuffled deck of cards. If you draw a red card, you win nothing. If you get a spade, you win $5. For any club, you win $10 plus an extra $20 for the ace of clubs.

- **(a) Create a probability model for the amount you win at this game. Also, find the expected winnings for a single game and the standard deviation of the winnings.**

```
a <- c(0, 5, 10, 30)
b <- c(26/52, 13/52, 12/52, 1/52)
d <- (a[1] * b[1]) + (a[2] * b[2]) + (a[3] * b[3]) + (a[4] * b[4])
e <- ((a[1] - d)^2 * b[1]) + ((a[2] - d)^2 * b[2]) + ((a[3] - d)^2 * b[3]) +
((a[4] - d)^2 * b[4])
```

Equations	Red	Spade	Club(not Ace)	Club(Ace)	Total
x_i	$0	$5	$10	$30	
$P(X = x_i)$	0.5	0.25	0.2307692	0.0192308	
$x_i * P(X = x_i)$	0	1.25	2.3076923	0.5769231	E(X) = 4.1346154
$x_i - E(X)$	-4.1346154	0.8653846	5.8653846	25.8653846	
$(x_i - E(X))^2$	17.0950444	0.7488905	34.4027367	669.0181213	
$(x_i - E(X))^2 * P(X = x_i)$	8.5475222	0.1872226	7.9390931	12.8657331	V(X) = 29.539571

The expected winngings for a single game is E(X) = $4.1346154. The standard deviation is $\sqrt{V(X)} = \$5.4350318$.

- **(b) What is the maximum amount you would be willing to pay to play this game? Explain**

I am willing to pay $4, since the expected value of this game is positive $4.13.

2.40 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 <- c(0, 25, 60)
b <- c(0.54, 0.34, 0.12)
d <- (a[1] * b[1]) + (a[2] * b[2]) + (a[3] * b[3])
e <- ((a[1] - d)^2 * b[1]) + ((a[2] - d)^2 * b[2]) + ((a[3] - d)^2 * b[3])
```

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

Equations	No luggage	1 luggage	2 luggages	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
$x_i - E(X)$	-15.7	9.3	44.3	
$(x_i - E(X))^2$	246.49	86.49	1962.49	
$(x_i - E(X))^2 * P(X = x_i)$	133.1046	29.4066	235.4988	V(X) = 398.01

The average revenue per passanger is E(X) = \$15.7. The standard deviation is $\sqrt{V(X)} =$ \$19.950188.

- **(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.**

Assuming that no one checked in more then 2 pieces of baggage or payed any other extra fee for their baggage (ex: overweight), the airline should be expecting to make about \$1884 in revenue from a flight of 120 passangers. With the standard devition of \$19.950188.

2.42 Selling on Ebay. Marcie has been tracking the following two items on Ebay:
• A textbook that sells for an average of \$110 with a standard deviation of \$4.
• Mario Kart for the Nintendo Wii, which sells for an average of \$38 with a standard deviation of \$5.

```
a <- c(110, 38)
b <- c(4, 5)
```

- **(a) Marcie wants to sell the video game and buy the textbook. How much net money (profits losses) would she expect to make or spend? Also compute the standard deviation of how much she would make or spend.**

Marcie should be expecting to have a net profit (loses) of -72, meaning that she should be expected to spend \$72. The standard deviation is \$6.4031242.

- **(b) Lucy is selling the textbook on Ebay for a friend, and her friend is giving her a 10% commission (Lucy keeps 10% of the revenue). How much money should she expect to make? With what standard deviation?**

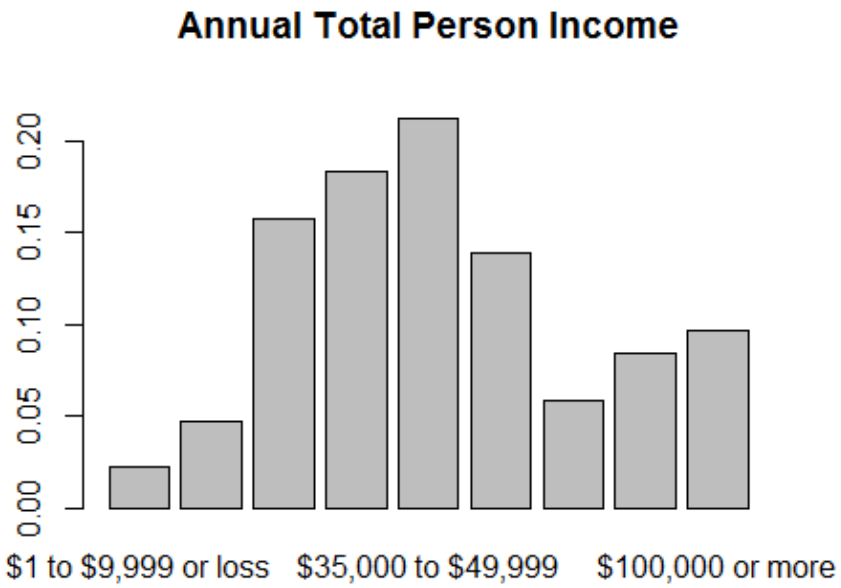
Lucy should be expecting to get a commissin of \$11. With the standard deviation of \$0.4.

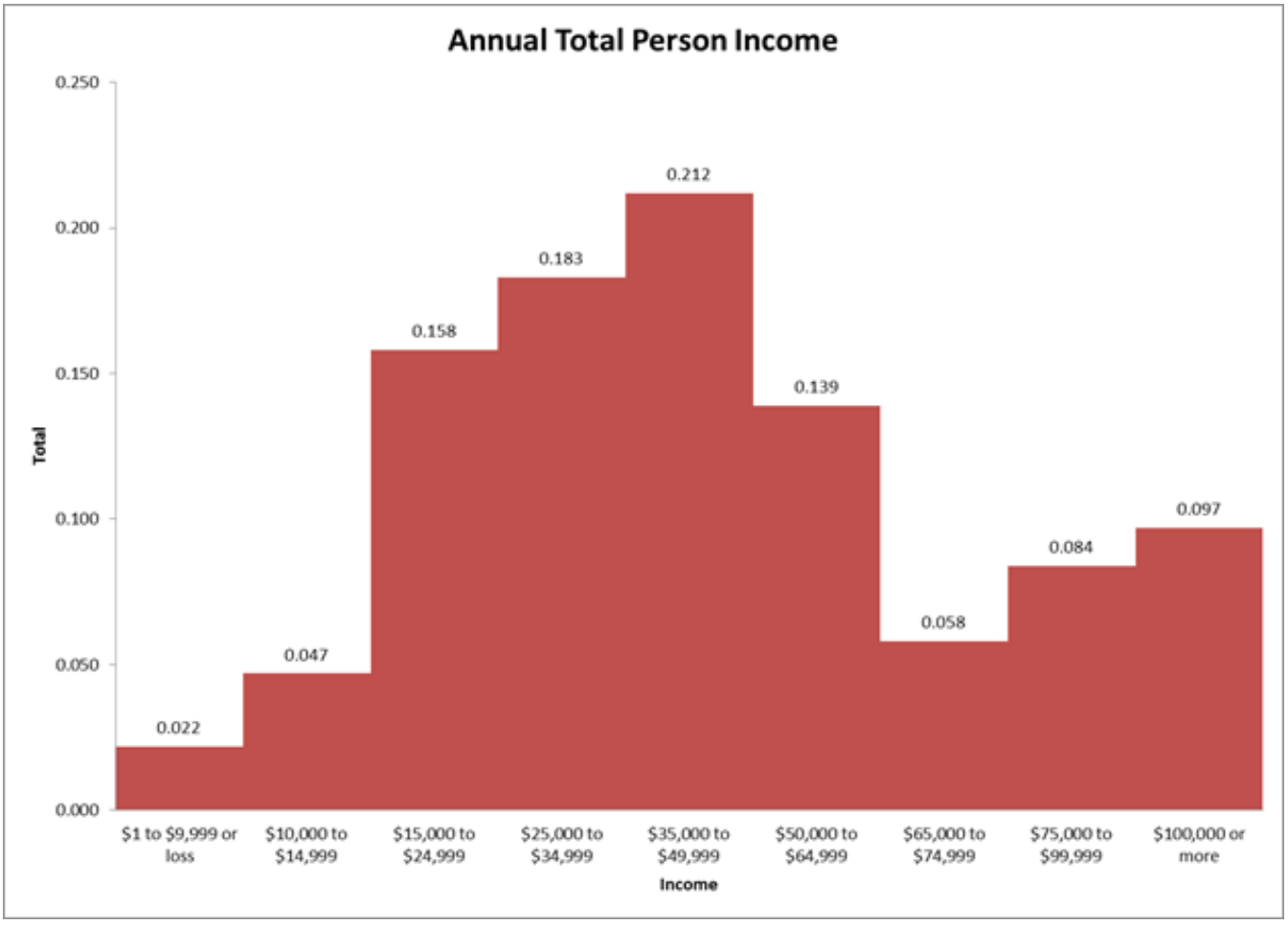
2.46 Income and gender. The relative frequency table below displays the distribution of annual total personal income (in 2009 inflation-adjusted dollars) for a representative sample of 96,420,486 Americans. These data come from the American Community Survey for 2005-2009. This sample is comprised of 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.**

```
##                  Income Total
## 1 $1 to $9,999 or loss 0.022
## 2   $10,000 to $14,999 0.047
## 3   $15,000 to $24,999 0.158
## 4   $25,000 to $34,999 0.183
## 5   $35,000 to $49,999 0.212
## 6   $50,000 to $64,999 0.139
## 7   $65,000 to $74,999 0.058
## 8   $75,000 to $99,999 0.084
## 9     $100,000 or more 0.097
```


Annual Total Person Income
0.00
0.05
0.10
0.15
0.20
$1 to $9,999 or loss
$35,000 to $49,999
$100,000 or more


Annual Total Person Income
Total
0.250
0.200
0.150
0.100
0.050
0.000
0.022
0.047
0.158
0.183
0.212
0.139
0.058
0.084
0.097
$1 to $9,999 or loss
$10,000 to $14,999
$15,000 to $24,999
$25,000 to $34,999
$35,000 to $49,999
$50,000 to $64,999
$65,000 to $74,999
$75,000 to $99,999
$100,000 or more
Income

A we look at the graph on the top we see that the data of total personal income is skewed to the right

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

The probability of a randomly chosen US resident who makes less than $50,000 per year is 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.**

Assuming that the ratio of males to females are the same for each income group. In otherwords 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%.

- **(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.**

The probability of a randomly chosen US resident is a female knowing that 71.8% females make less than $50,000 is 29.438%. The assumption I made in part (c) is valid since we were working with such a large data set, and as we see from the answers they are not very different.