Data Science Math Week 3 Homework

Complete exercises 2.34, 2.40, 2.42 and 2.46 in OpenIntro Statistics pp. 114 – 117

Exercise 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.

1. 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.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | X | P(X) | X P(X) | (X – E(X))2 | P(X) (X – E(X))2 |
| Red card | 0 | 26/52 | 0 | (0 – 4.13)2 = 18.5761 | (26/52) (18.5761) = 9.28805 |
| Spade | 5 | 13/52 | 65/52 = 1.25 | (5 – 4.13)2 = 0.7569 | (13/52) (0.7569) = 0.189225 |
| Club | 10 | 12/52 | 120/52 = 2.31 | (10 – 4.13)2 = 34.4569 | (12/52) (34.4569) = 7.951592 |
| Ace of clubs | 30 | 1/52 | 30/52 = 0.58 | (30 – 4.13)2 = 669.2569 | (1/52) (669.2569) = 12.870325 |
|  |  |  | E(X) = 215/52  = 4.13 |  | V(X) = 30.299192  SD(X) = 5.50 |

1. What is the maximum amount you would be willing to pay to play this game? Explain.

You would want the expected earning to equal the amount you pay to play the game, so that you will at least break even.

Profit = winnings – cost or E(X – cost) = E(X) – cost = 0 = 4.13 – cost. The most one should pay to play the game is $4.13 so that on average one would not lose money playing the game. Ideally you pay less than this amount, to increase your profit.

Exercise 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.

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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *i* | 1 (no checked bags) | 2 (one checked bag) | 3 (two checked bags) | Total |
| *xi* | 0 | 25 | 60 |  |
| P(X = *xi*) | 0.54 | 0.34 | 0.12 |  |
| *xi* \* P(X = *xi*) | 0 | 8.5 | 7.2 | E(X) = 15.7 |
| *xi* - *µ* | -15.7 | 9.3 | 44.4 |  |
| (*xi* - *µ*)2 | 246.49 | 86.49 | 1962.49 |  |
| (*xi* - *µ*)2 \* P(X = *xi*) | 133.1046 | 29.4066 | 235.4988 | V(X) = 398.01  SD(X) = 19.95 |

The average revenue per passenger is $15.70 with a standard deviation of $19.95.

1. 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.

For a flight of 120 passengers the airline should expect to make 120 \* $15.70 = $ 1884.00

The variance of the airline luggage profit is 120x 19.952 = 47760.3, and the standard deviation is $218.54

This estimate relies on the assumption that the amount of baggage each passenger checks is independent of the amount checked by other passengers. For most passengers this assumption may be justified, but if anyone is traveling with babies, the amount of luggage checked will not be independent, because there is just way too much stuff and their baggage gets spread out among the other travelers in the party, making the relationship dependent.

Exercise 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.

1. 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 can expect ($38 – $110 = -$72) a net loss of $72

The variance of what she would spend is (-1)2 x 52 + (1)2 x 42 = 41, the standard deviation of what she would spend is $6.40.

1. 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 expect to make $11, with a standard deviation of $0.40

Exercise 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.

1. Describe the distribution of total personal income.
2. What is the probability that a randomly chosen US resident makes less than $50,000 per year?
3. 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.
4. 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.

|  |  |
| --- | --- |
| Income | Total |
| $1 to $9,999 or less | 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% |

The distribution above is centered with a median at $35,000 to $49,999. The spread of the data encompasses the entire range, with no gaps, and no obvious skew. It appears to be unimodal, but not symmetric.

1. 0.022 + 0.047 + 0.158 + 0.183 + 0.212 = 0.622
2. Assuming that the individual income ranges are distributed between males and females in the same ratio as the entire sample ( not an assumption likely to be true), the probablilty that a randomly chosen US resident makes less than $50,000 and is female would be 0.622 x 0.41 = 0.25502
3. If the assumption in c were true, this would mean that approximately 24, 589,152 females in the survey make less than $50,000. If the survey is comprised of 41% females, this is 39,532,399 women, and if 71.8 % of them make less than $50,000, that comes to 28,384,263 women. Given that 62.2% of the respondents (59,973,542) make less than $50K and 28,384,263 women make less than $50K, this indicates that 47.3% of the less than $50K income bracket is women, a significantly higher percentage than their overall representation in the survey population. The reason for this is not clear from the data given, perhaps more women are working part time than men, or it could be that they are working in lower paid industries, or it could be attributable to other factors.

At any rate, it shows the assumption made in part c is not valid.