Vanessa Trujillo
Stats 381
Project 2
February 28, 2018

**Introduction**: The purpose of this assignment is to take sets of data corresponding to different demographics: Gender, Career Sectors, Company sizes, Salaries, and Savings, to determine probabilities amongst variables. All the data presented here has been analyzed and processed through Microsoft Excel.

**Results:**

**5.b**

**A.)** The probability that a person is selected at random out of the 500 employees and is young is 131/500

**B.)** There are 187 women out of the 266 female employees that make over $55,000 so the probability of selecting one of these women at random is from this precondition is 188/266

**C.)** There are 48 women in the Financial sector, the probability of selecting a woman at random who works in the Finance Sector and is paid more than $60,000 is 35/48 since 35 of those women do make more than $60,000.

**D.)**

**1.** P(women) = 266/500 = 0.472

**2.** P (Finance | women) = P(women) * P (women in finance from all women) = (266/500) * (48/266) = 0.096

**3.** P (Finance + 60K | women) = P(women) * P (women in finance from all women) * P (women in finance + 60k) = (266/500) * (48/266) * (35/48) = 0.07

**E.)** The probability that a person is a man if he earns more than 90,000 is 113/500 = 0.226

**F.)**

**1.** P(Man) = 234/500 = 0.468

**2.** P (Earns +90,000 | Man) = [(234/500) *(115*234)] =0.23.

- **i.)** **EVENT A:** P(Man) = Total men/ Total employees **EVENT B:** P (Earns +90,000 | Man) = [(# of male/total employee) *(# of men that earn 90000+/#of male employees)]
- **ii.)** The first one is not conditional, the second one is, we need to set the first one as the condition to solve for a man that earns more than $90,000.
- **iii.)** The prior event is that the selected person is a man.

**G.)** P (Man+85,000 | company size 4 ∩ Sector 5) = 3/500 = 0.006

Vanessa Trujillo
Stats 381
Project 2
February 28, 2018

**H.)** P (Man+85,000 | company size 4 ∩ Sector) = P (men making 85,000) / P (# of men in sector 5) +

P (# of men in company size 4) = (((30/11) / ((102/185) + (33/103))/ 500) = 0.006

5c.

| Company | Variable | Probability |
|---|---|---|
| 1 | 91 | 91/500 |
| 2 | 92 | 92/500 |
| 3 | 133 | 133/500 |
| 4 | 184 | 184/500 |

5d.

A.

| Sectors | employees | probability distribuation |
|---|---|---|
| 1 - Financial | 92 | 92/500 |
| 2- Pharmaceutical | 121 | 121/500 |
| 3 - Manufacturing | 80 | 80/500 |
| 4 - Construction | 40 | 40/500 |
| 5 - Education | 102 | 102/500 |
| 6 - Services | 65 | 65/500 |
| | | |

B.

| Sectors | probability distribuation | mean salary |
|---|---|---|
| 1 - Financial | 92/500 | 89116.43 |
| 2- Pharmaceutical | 121/500 | 85496.35 |
| 3 - Manufacturing | 80/500 | 79893.94 |
| 4 - Construction | 40/500 | 89735.41 |
| 5 - Education | 102/500 | 72987.04 |
| 6 - Services | 65/500 | 83276.36 |

Vanessa Trujillo
Stats 381
Project 2
February 28, 2018

C.

| Expected Value | Variance | Standard Deviation |
|---|---|---|
| 82764.68598 | 33972002.79 | 5828.550659 |

D. The mean is the expected value, and therefore they are equal. In this example the mean = expected value = 82764.69. This means that on average most employees make a salary around $83,000. This is a rough estimate considering how education makes approximately $73,000 annually.

E. The standard deviation was equal to 5828.5069, a relatively low value. This low value signifies that most of the data points studied have a tendency to be closer to the mean of the data set, helping us make the claim that almost all the other average salaries for all the sectors were within a close range of the expected value (mean).