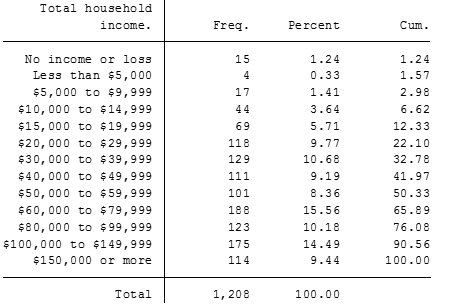
Sociology 328 Social Statistics

Christian Botnick, Andy Xiao and Sharon Michaeli

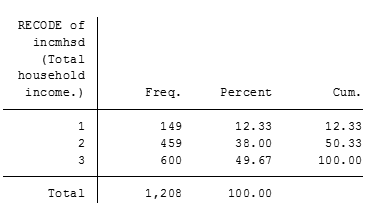
Professor Gerry Veenstra

11/14/2017

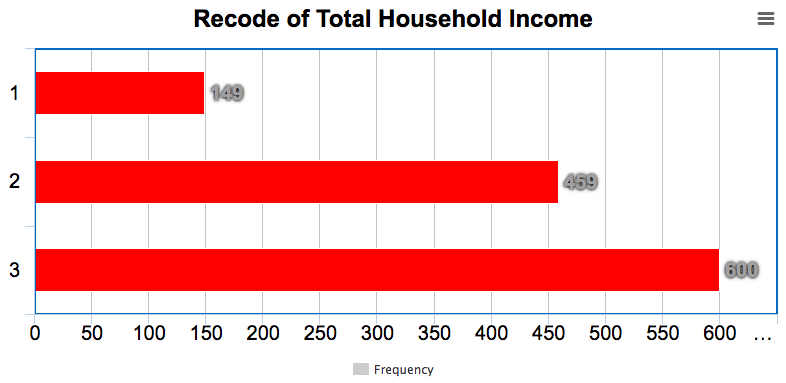
The NOCS 2006\_C10 (National Occupation Classification) constructs 520 different occupational group descriptions composed of 3000 job titles. The NOC is used as a foundational data source for the labour market in Canada. We would theoretically expect the two variables(National Occupational Classification of the respondent and household income) to possess a strong relationship. Household income is often dependent on the type of occupation a person or set of individuals have. Most high paying jobs require a higher level of skill for entry, whereas working as a cashier at a grocery store is a job tailored to almost everyone, requiring minimum job requirement and experience for entry. Firstly, a higher classified job such as: ” Doctor” are obtained through excessive amount of time and money devoted into education. The immense difficulty to achieve such occupational title alienates the average person from the outliers. Therefore, the quantity of positions for the high status occupations are limited. When the occupation is needed by the society and only limited amount of individuals have the ability to perform it, they will undoubtedly be well compensated. On the other end of the spectrum, jobs such as Cashiers, although require significant physical efforts, hold a minimum occupational status and are labeled/tailored towards most individuals. We expect the relationship between the two variables to be: the higher the level of skill required within the occupational group, the higher the household income. As for determining the independent and the dependent variables in this study, we concluded that the NOC strongly influences household income resulting in the NOC being labelled as the independent variable and household income being labelled the dependent variable.

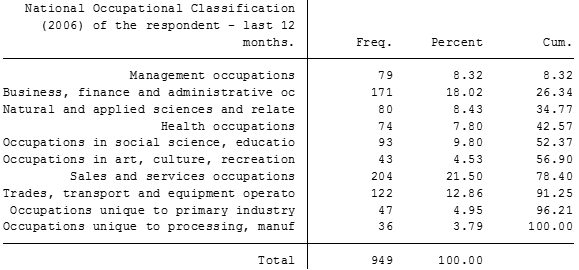
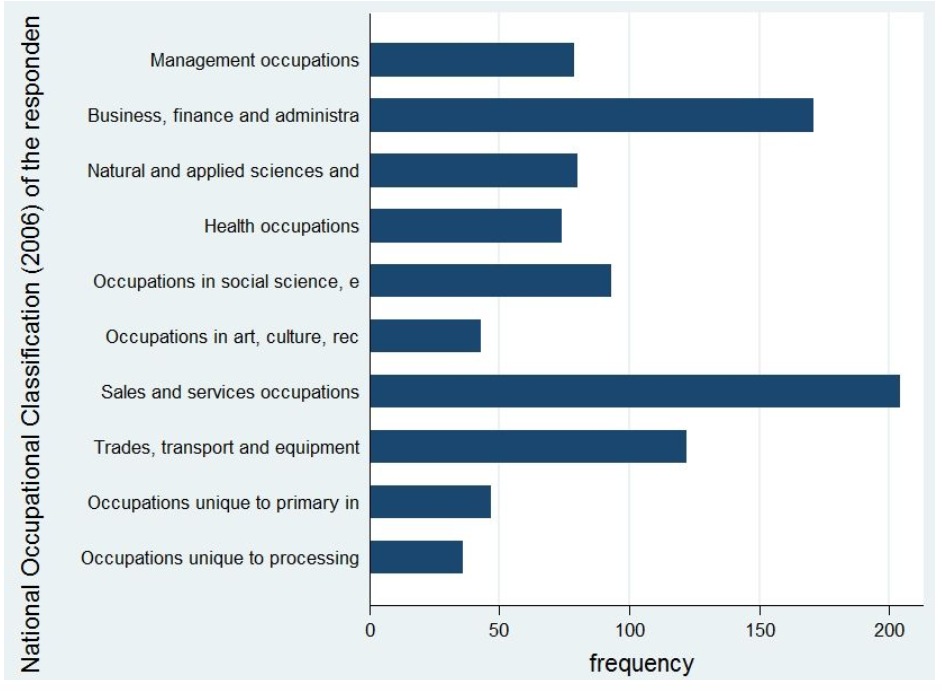
In the frequency table below Household income is an interval-ratio variable categorized into an ordinal variable. This frequency table measures annual income ranging from ‘no income’ to ‘$150,000 or more’. 

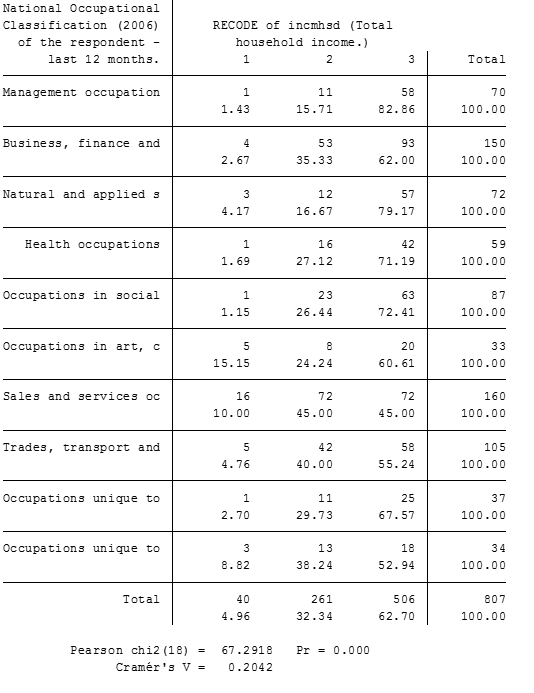
In order to better assess the frequency table above, we recoded the variable from 13 categories of small increments of $5,000 and $10,000 into three larger incremental categories labelled low income, medium income, and high income (1= no income or loss- $19,999, 2= $20,000-59,999, and 3= $60,000 to $150,000 or more). This recoded frequency table is shown below: Frequency Table of Recoded Household Income (incmhsd)

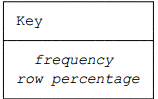
   
 1: Low 2: Medium 3: High

In the table above, 12.33% of respondents fall within the low income category 1 (no income or loss - $19,999), 38.00% fall within the medium income category 2 ($20,000- 59,999), and 49.67% fall within the high income category 3 ($60,000- $150,000) . The majority of respondents fall within the high income increment. The distribution of the data set representing the variable household income, has a higher variance among respondents, the highest frequency of recoded group 3 (600). A better visualisation of this frequency table is shown below in this barchart:



This bar chart better visualises the Household Income variable and shows that the majority of respondents fall within category 3 ($60,000- $150,000). In contrast, the NOC variable assesses the level of occupational classification of the respondent. This variable is grouped into 10 nominal categories shown in the frequency table below: This frequency table demonstrates NOC groups categorized into certain occupational classifications.The majority of respondents (21.50%) fall within the category of ‘Sales and services occupations’, and the lowest amount of respondents (3.78%) fall within the category of ‘Occupational unique to processing, manuf’. The distribution of the data set representing the NOC, has a high variance among respondents, the highest frequency in group Sales and services occupations (204). Below is a barchart that better visualises the distribution from the frequency table:

The bar-chart further shows that the category of ‘Sales and services occupations’ has the highest percentage of respondents, with the category of ‘Business, finance, and administration’ second in the frequency of respondents and the category of ‘Occupational unique to processing, manuf’ with the least amount of respondents. Both variables are categorical variables where Household income is an ordinal level of measurement and National Occupational Classification is a nominal level of measurement. 

To furthermore explain the values in the table: 

In the cross tabulation shown above, it displays the National Occupational classification (2006) of the respondent on the left side, and the Recode of Income on the right. The row totals considered together and the column totals considered together are called the marginal distributions. These are essentially frequency distributions for the two variables. In each row we can calculate the percentage of the row total that pertains to each column. These are the row percentages. The upper portion of each NOC category is the frequency in the 3 groups, and the lower portion of the category is the percentage in each of the 3 groups. For example, in Management occupation, the first income group has a frequency of 1 and a percentage of 1.43, while group 2 has a frequency of 11 and a percentage of 15.71, and group 3 with the majority frequency of 58 with a percentage of 82.86. Which indicates that the wage paid for the management occupations are mostly very high.Through frequency tables 1 and 2, we discovered the relationship between the independent variable classification of occupation and the dependent variable total house income has a cramer’s V value of 0.2042. The value of cramer’s V indicates that the strength of the relationship is modest, as it ranges for 0 -1, no relationship - perfect relationship. Therefore we can conclude that the relationship between household income and occupation class to be modest, as there is a relationship in existence. Furthermore the Pearson’s CHI^2 = 67.2918, with such a high value, we have great evidence against the null hypothesis of statistical independence. The degrees of freedom for a Chi-square grid are (10-1)\*(3-1)= (9)\*(3) = 27, which means 27 of the numbers in the grid is independent.

**4. Provide some interpretive insights regarding the relationship between the variables. Were your theoretical expectations met? NO Present your analyses and insights in sentence and paragraph form (accompanied by attractive graphs and tables) as if you were writing a formal report for public consumption. Please type your double-spaced assignment using Times Roman 12-point font.**

**Conclusion: this is what was expected in hypothesis, this is what actually happened after analyzing the data using cramers v and the chi squared tests**

**4. Provide some interpretive insights regarding the relationship between the variables.email prof to make sure but sounds like reittarating with words what the numbers mean example, it seems household income is modestly related to occupation class. Were your theoretical expectations met? No, expectation was a strong relationship, yet it seems acccording to creamer’s V that it is a modest relationship.**