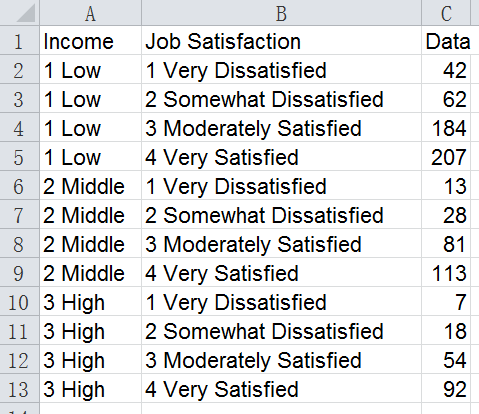
|  |
| --- |
|  |
| STAT 445 Assignment 9 |
| Kun Yang 301178299 |
|  |
|  |
|  |

|  |
| --- |
|  |

**1. Consider the data in Table 12.15 on page 753 of the text.**

1. ***Construct a reformatted table of these data in Excel in a format suitable for exporting into a JMP spreadsheet for doing a correspondence analysis.***

**

1. ***The following mosaic plots were constructed by JMP. They contain evidence of a data entry error. Describe in as much detail as you can what this error appears to have been.***



The error data is the group of high income with moderate level of job satisfaction. From the left plot, we can see that the color of this group is brown but it should be blue. And from the right plot, it make this group as a separate row itself but it should be in the row of high income.

1. ***The following mosaic plots were constructed after the error was fixed. Can you detect any pattern of interest in these? If so, describe it, and comment on what insight this might provide on the relationship between income and job satisfaction.***

****

In the left plot, the low income is relatively most abundant in very dissatisfied level then it decreases from somewhat dissatisfied to very satisfied levels. And the high income is relatively most abundant in very satisfied level then it decreases from moderately satisfied to very dissatisfied levels.

In the right plot, the very satisfied is relatively most abundant in the high income level then it decreases from middle to low income levels. In addition, very dissatisfied is the most abundant in the low level income then it decrease from the middle to high income levels.

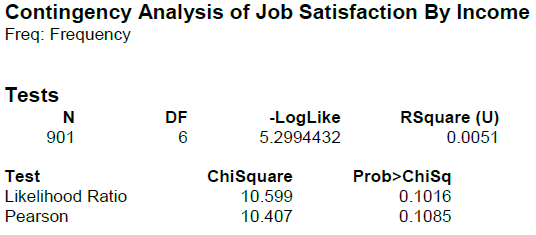
Therefore, it might provide a relationship that the job satisfaction will increase as the income level increase.

1. ***Provide two distinct reasons why the evidence in these plots does not prove that higher income levels create more job satisfaction.***

We can notice that the highest income group has the smallest sample size and the lowest income group has the largest sample size. The distribution of the highest income group may be skewed because of the small sample size.

Moreover, we cannot identify whether the job satisfaction of each level income are affected by other possible factors. For example, maybe the health condition affects the job satisfaction level. But in the mosaic plot we cannot figure it out.

1. ***Following is a summary of the test results that JMP provides for these data. What null hypothesis is being tested here? Is it rejected? What relationship, if any, does this test result have to your answer to part (d) above?***



In this case, the null hypothesis is that the distribution of the job satisfaction is the same across each income level. The statistic of the Chi-square test shows that it **CANNOT be reject**. Therefore, we have enough evidence that the job satisfaction is the same across each income level.

And these dataset do **NOT** provide enough evidence that the higher income levels create more job satisfaction.

***2. The following output is from the same JMP analysis, but this time, for the related, correspondence analysis.***

1. ***How are the Inertia Values related to the Singular Values?***

The inertia value reflects the relative variation which is accounted for by each dimension.

The singular Value displays the canonical correlation between the two variables for each dimension.

***The inertia is the square of the singular values***

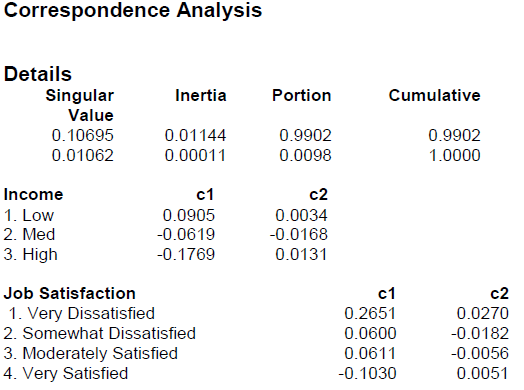
1. ***How many dimensions are needed to capture almost all of the variation in this contingency table?***

***Justify your answer by reference to the appropriate portion of the output.***

We need one dimension to capture almost all of the variation in this contingency table. From the column of cumulative, we know that the first dimension has already capture 99.02% of the variation. (We also can calculate the result from the column of inertia. The 1st-dimenion inertia value divided by the sum of the inertia value in each of the rows= 0.01144/(0.01144+0.00011)=0.99, same result. )

1. ***Use the output below to construct a standard, two-dimensional correspondence analysis graph. Make it as neat and professional as you reasonably can. Marks will be deducted for sloppy graphs.***

*(The graph generated by* JMP *is almost impossible to read. It is important to be able to draw better ones yourself.)*

****

1. ***Does your graph confirm your answer to part (b)? Explain in at most one sentence.***

Yes, it confirms my answer.

All the points on the C1-axis go from the top to the bottom and scatter vertically alone the c2-axis as a nearly straight line, it means C1-axis (dimension one) capture almost all variation.