Assignment 1: Use Stata to investigate the bivariate relationship between the categorical variables NOCS2006\_C10 (occupational classification) and MSS\_Q110 (self-reported stress) in the GSS dataset.

1. Provide a theoretical rationale to guide your analysis. Why would you expect the two variables to be related? How would you expect them to be related? Are they likely to be strongly or weakly related, do you think? Is one likely to cause or influence the other, i.e., does it make sense to designate one as the independent variable and the other as the dependent variable?

2. As far as you can tell, what, exactly, do the variables measure or assess? What are their levels of measurement? What do their distributions look like in this dataset?

Occupational classification is measured at nominal level while self-reported stress is measured at an ordinal level. Occupational classification is distributed in a wide spread that doesn't offer much insight, over a third of the respondents’ information is missing and those who have responded are spread between so many different categories that none offer a particularly high frequency.

|  |  |  |
| --- | --- | --- |
| **National Occupational Classification of the respondent** | **Frequency** | **Percent** |
| Management occupations | 79 | 5.27 |
| Business, finance and administrative | 171 | 11.40 |
| Natural and applied sciences and related | 80 | 5.33 |
| Health occupations | 74 | 4.93 |
| Occupations in social science, education | 93 | 6.20 |
| Occupations in art, culture, recreation | 43 | 2.87 |
| Sales and services occupations | 204 | 13.60 |
| Trades, transport and equipment operation | 122 | 8.13 |
| Occupations unique to primary industry | 47 | 3.13 |
| Occupations unique to processing, manufacturing | 36 | 2.40 |
| Missing | 551 | 36.73 |
| **Total** | **1,500** | **100.00** |

|  |  |  |
| --- | --- | --- |
| **Thinking about the amount**  **of stress in your life,**  **would you say most days are?:** | **Frequency** | **Percent** |
| Not at all stressful | 275 | 18.33 |
| Not very stressful | 351 | 23.40 |
| A bit stressful | 586 | 39.07 |
| Quite a bit/extremely stressful | 248 (34) | 16.53 |
| Missing | 40 | 2.67 |
| **Total** | **1,500** | **100.00** |

\*collapsed categories to create combined quite a bit/extremely\*



1: Not at all stressful 2: Not very stressful 3: A bit stressful 4: Quite a bit/extremely stressful

3. Using statistical techniques covered in the lectures and textbook, describe and summarize the relationship between the two variables. You may have to recode or transform a variable first to make the relationship more intelligible. How are the variables related to one another (if at all)? If they are related, is the relationship strong or weak (or something in between)?

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **National Occupational Classification of the respondent** | **Not at all stressful** | **Not very stressful** | **A bit stressful** | **Quite a bit/extremely stressful** | **Total** |
| Management occupation | 7  8.97 | 13  16.67 | 37  47.44 | 21  26.92 | 78  100.00 |
| Business, finance | 15  8.77 | 39  22.81 | 80  46.78 | 37  21.64 | 171  100.00 |
| Natural and applied sciences | 6  7.50 | 20  25.00 | 41  51.25 | 13  16.25 | 80  100.00 |
| Health occupations | 2  2.78 | 19  26.39 | 31  43.06 | 20  27.78 | 72  100.00 |
| Occupations in social sciences | 6  6.52 | 11  11.96 | 51  55.43 | 24  26.09 | 92  100.00 |
| Occupations in art, culture | 7  16.28 | 8  18.60 | 17  39.53 | 11  25.58 | 43  100.00 |
| Sales and services | 22  10.84 | 45  22.17 | 93  45.81 | 43  21.18 | 203  100.00 |
| Trades, transport | 16  13.11 | 28  22.95 | 57  46.72 | 21  17.21 | 122  100.00 |
| Occupations unique to primary | 14  29.79 | 10  21.28 | 17  36.17 | 6  12.77 | 47  100.00 |
| Occupations unique to processing | 7  20.00 | 3  8.57 | 17  48.57 | 8  22.86 | 35  100.00 |
| **Total** | **102**  **10.82** | **196**  **20.78** | **441**  **46.77** | **204**  **21.63** | **943**  **100.00** |

Cramer's V = 0.1304

|  |  |
| --- | --- |
| Not at all stressful | 275 (102 = 37.09) |
| Not very stressful | 351 (196 = 55.84) |
| A bit stressful | 586 (441 = 75.26) |
| Quite a bit/extremely stressful | 248 (204 = 82.26) |







4. Provide some interpretive insights regarding the relationship between the variables. Were your theoretical expectations met?

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.