**Using SPSS with crosstabs and chi square**

**The type of variable you have to work with influences strongly the statistical tools you can use to analyze potential relationships. This exercise is designed to show that you can identify the type of variables we have chosen and then shows you can control SPSS to do the analysis.**

1. Look at the following variables in the dataset GSS2012A\_student. What does each one measure? Which one is nominal level? Which is ordinal level? Which is interval level? (10 points total)

1a. race2 (named race\_2 in dataset instead)

The variable “race\_2” measures which race does the respondent belong to. It has 2 value labels, each coded as “1” and “2” for White” and “Black”, respectively. Because the coding only records differences in race, this variable is nominal-level.

1b. age

The variable “age” measures how old the respondent is. It has 3 value labels, each coded as “89”, “98”, and “99” for “89+”, “DK”, and “NA”, respectively. Because the coding represents numeric quantities that denote one’s age (also shown by other “age” variables), this variable is interval-level.

1c. income06

The variable “income06” measures the respondent’s total family income. It has many values – ranging from 0 to 99, each indicating how much money a given respondent’s family earns – also ranging from “under 1000” to “150000 or over”. Since the coding represents numeric quantities of ascending income levels, this variable is interval-level.

1d. polviews

The variable “polviews” measures how the respondent’s places themselves ideologically. It has many values – ranging from 0 to 9, each indicating a level of how liberal or conservative the respondent is. Since the coding distinguishes the relative amount of how liberal or conservative one is (ranging from “extremely liberal”, to “moderate”, to “extremely conservative”), the variable is ordinal-level.

1e. educ4 (named educ\_4 in dataset instead)

The variable “educ\_4” measures how much education does the respondent have. It has 4 values – ranging from 1 to 4, each indicating how much education did the respondent complete – also ranging from “below high school” (denoted as <HS) to “above college” (denoated as Coll+). Because the coding distinguishes the level of educational attainment one has, the variable is ordinal-level.

1. Now do a cross-tab of two of these variables, setting out the relationship expressed in the following hypothesis: “In a comparison of individuals, those who identify as “white” are more likely to describe themselves as liberal than those who identify as “black.” Put the appropriate row or column percentages in so you can evaluate the results. (5 points). Do the data support the hypothesis? (5 points)

Based on the cross-tabular data of the race and ideological self-placement variables (from top to bottom), it seems that 1657 white and black respondents indicated their ideological self-placement. Of these 1657 persons, 62 self-identify as “extremely liberal”, 45 and 17 of whom are white and black, respectively. Furthermore, of these same 1657 persons, 189 self-identify as just “liberal”, 161 and 28 of whom are white and black, respectively. Lastly, of these same 1657 persons, 195 self-identify as “slightly liberal”, 158 and 37 of whom are white and black, respectively. These numbers, based on the “extremely liberal”, “liberal”, and “slightly liberal” rows, seem to suggest that whites are more likely to describe themselves as “liberal” than are blacks – even at varying levels of being “liberal”. This observation, then, seems to support the hypothesis.

However, it should be noted that, for every row, the amount of whites is disproportionately higher than that of blacks. The total number of whites who indicated their ideological self-placement is 1379, while the total number of blacks who indicated their ideological self-placement is 278. This numerical disparity in respondents can lead to inflated results for whites’ responses. Thus, while the hypothesis is supported by the numbers stated earlier, it could be as such due to whites outnumbering blacks in every row.

**Crosstabs**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Case Processing Summary** | | | | | | |
|  | Cases | | | | | |
| Valid | | Missing | | Total | |
| N | Percent | N | Percent | N | Percent |
| Ideological Self-Placement \* Race: Black / White | 1657 | 83.9% | 318.001 | 16.1% | 1975.001 | 100.0% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ideological Self-Placement \* Race: Black / White Crosstabulation** | | | | | |
|  | | | Race: Black / White | | Total |
| White | Black |
| Ideological Self-Placement | ExtrmLib | Count | 45 | 17 | 62 |
| % within Ideological Self-Placement | 72.6% | 27.4% | 100.0% |
| Liberal | Count | 161 | 28 | 189 |
| % within Ideological Self-Placement | 85.2% | 14.8% | 100.0% |
| SlghtLib | Count | 158 | 37 | 195 |
| % within Ideological Self-Placement | 81.0% | 19.0% | 100.0% |
| Moderate | Count | 507 | 130 | 637 |
| % within Ideological Self-Placement | 79.6% | 20.4% | 100.0% |
| SlghtCons | Count | 219 | 29 | 248 |
| % within Ideological Self-Placement | 88.3% | 11.7% | 100.0% |
| Conserv | Count | 242 | 29 | 271 |
| % within Ideological Self-Placement | 89.3% | 10.7% | 100.0% |
| ExtrmCons | Count | 47 | 8 | 55 |
| % within Ideological Self-Placement | 85.5% | 14.5% | 100.0% |
| Total | | Count | 1379 | 278 | 1657 |
| % within Ideological Self-Placement | 83.2% | 16.8% | 100.0% |

1. Now do the same analysis using the chi square statistical test. (5 points) What do the results tell you about the relationship? Is it supported or not? (5 points)

Below is the tabular data of the *χ*2 test below produced by SPSS. The alternative hypothesis is: whites are more likely to describe themselves as liberal than blacks. The null hypothesis, however, is: whites are not more likely to describe themselves as liberal than blacks. If the null hypothesis is correct that, in the population from which the sample was drawn, there is no relationship between race and political views, then random sampling error will produce the observed data 0.000 of the time. In other words, the tabular data above and its observed differences between race and ideological self-placement will yield a test statistic of 24.191 about 0% of the time. Since the p-value is practically zero, and is less than the significance level of 0.05 (derived from the 95% confidence interval standard), it can asserted that the null hypothesis should be rejected. The alternative hypothesis (the hypothesis stated earlier), then, should be kept and is supported by the chi-square test. There is, indeed, a correlated relationship between race and ideological self-placement.

|  |  |  |  |
| --- | --- | --- | --- |
| **Chi-Square Tests** | | | |
|  | Value | df | Asymptotic Significance (2-sided) |
| Pearson Chi-Square | 24.191a | 6 | .000 |
| Likelihood Ratio | 24.544 | 6 | .000 |
| Linear-by-Linear Association | 9.378 | 1 | .002 |
| N of Valid Cases | 1657 |  |  |
| a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 9.23. | | | |