Lab Exercise: Data Cleanup & Graphical Summaries

Problem

Quantitative data almost never come to you in the exact format you require. Even if you are working with data that someone else has collected and "cleaned up", you generally need to alter the variables somewhat before you can carry out the statistical analyses that you have planned.

In some cases, this will mean obliterating certain distinctions that the original variable makes, or even declaring that certain values of the original variable are altogether meaningless for your purposes. In others, it will mean computing new variables of your own, perhaps altering the scale on which some variable is measured or combining two or more variables into a single measure.

In every case, it will mean documenting changes you have made, such that you and anyone who comes after you will be able to discern *exactly* what the variables in your dataset measure.

Objectives

For this assignment, you will be proving your ability to perform basic data manipulation functions using SPSS. To be specific, you will be demonstrating the use of the following features: RECODE, COMPUTE, MISSING VALUES, VARIABLE LABELS, and VALUE LABELS. By providing cleaned, validated, and manipulated data, these features serve as the foundation for other data analysis tasks you will perform throughout the semester.

Instructions

- From the course website, **get the data file music.sav**, a subset of the General Social Survey

- Request graphical summaries (i.e., **histograms) for each of 12 variables**:

BIGBAND	BLUES	BLUGRASS	COUNTRY	FOLK	GOSPEL
HVYMETAL	JAZZ	OLDIES	OPERA	RAP	REGGAE

- **Note what the "Values" are** for each of these, and what they mean. ***DO NOT SKIP THIS STEP!***

- **RECODE all of them** so that each has only two valid response categories:

  - **1,** indicating that the respondent **likes** or very much likes that particular style of music; and

  - **0,** indicating that they **dislike** or very much dislike it.

  - (You *could* recode "mixed feelings" & "all other values" to "System missing", but don't need to.)

- **COMPUTE an index called MUSIC by summing** the values of all 12 *new* measures.

  - If and only if you used 1&2, instead of 0&1 as instructed above, subtract 12 when you sum.

  - You may choose to assign an appropriate Variable Label ("Label") to your new MUSIC index.

- **Get output for all 13** variables you created (yes - histogram, central tendency, *and* dispersion ☺)

- **Submit written answers** to the questions on next page. (A few sentences each should suffice.)

- Be sure to include names of anyone *who contributed*, so everyone *who contributed* gets credit.

MUSIC LAB: Submission page

Secretary: _____

Others: _____   _____

                    _____   _____

Questions:

1. What do the 12 initial variables measure (the ones as I gave them, *before* your recoding)?
   (You don't need any output to answer this, or question 3. I've asked here about the *variables*.)

2. What do histograms for the *recoded* measures you created) suggest about interest(s) in music?
   (You *do* need output to answer this, and question 4. I've asked here about the *respondents*.)

3. What does your *index* measure – and how is that different from your answer to question 1?

4. What do the frequency distribution and descriptive statistics for MUSIC suggest?
   (You should be able to talk about the shape, central tendency, *and* dispersion of the index.)

Bonus: Does the valid sample size of MUSIC tell anything interesting about the respondents?