

Simple regression exercise using NATIONS dataset

A ploy to enable simple derivation of binary categorical variables.

In SPSS the *Transform/Compute* option allows us to calculate numeric expressions. It also allows us to evaluate expressions which are true/false (for example, that a variable lies above some value). Traditionally (and there are some good reasons for the tradition) 'true' values are represented numerically as one, and 'false' values numerically as zero.

So. Suppose we find, in the nations dataset, by running *Statistics/Summarise/Frequencies* (with 'median' selected through the 'statistics' option box) that the median value for *AgricLbr* (the % of the labour force in agriculture) is 43.2

then, Using *Transform/Compute* with, say *FARMING* as our Target variable and writing

***AgricLbr* >= 43.2**

in the *Numerical Expression* box

gives us a variable *FARMING* which has the value

0 for countries below the world median for percentage labour force in agriculture

1 for countries at or above the median.

Exercise

Use such a *FARMING* variable to test the hypothesis that the interconnection between the Female literacy rate (*FEMALTRC*) in a country and the Percentage of women in the labour force (*WOMENLBR*) is different between countries high and low on 'agriculture'. What do the findings mean?

Comment: so an invitation to run a regression (I ran mine with *WOMENLBR* as the dependent variable, but on reflection one could also defend the causal narrative as running in the opposite direction?)

Additional suggestion: a scattergram with *WOMENLBR* on the Y axis, *FEMALTRC* on the X axis and *FARMING* in the *Set Markers By* box will give you a nice visual image of what is going on.

Further separate exercise

Taking the *NATIONS* dataset, formulate some hypothesis involving four or so variables (since we are 'doing regression' the hypothesis will involve regarding one variable as our dependent variable and the others as some plausible predictors). Perhaps consider the possibility of an interaction effect (between interval level variables). Test your hypothesis. Think about the results.
