

GRAPHING THE INTERACTION BETWEEN A DICHOTOMOUS AND A QUANTITATIVE VARIABLE

Throughout this document, the following symbols will be used

X_1 : predictor variable 1 that codes the dichotomous variable

X_2 : the quantitative predictor variable

X_1X_2 = the product of X_1 and X_2

C_i : a variable or set of variables used as statistical controls

Y : outcome variable

\hat{Y} : estimate of Y from regression model

The interaction between X_1 and X_2 is tested by estimating the coefficients of the following regression model:

$$\hat{Y} = a + b_1X_1 + b_2X_2 + b_3(X_1X_2) + \sum b_iC_i$$

where the number of variables being controlled (C) may be as small as 0 (in which case the last term disappears). If b_3 is statistically different from zero, it is said that X_1 and X_2 interact.

To construct a graphical depiction of the interaction,

(1) Select at least two but preferably several values of the quantitative variable X_2 that will be used in the figure on the X axis. Call the number of values used k . For example, if X_1 is measured on a 1 to 5 scale, you might use the values 1, 2, 3, 4, and 5, and so $k = 5$. Alternatively, you could use the mean of X_2 , one standard deviation below the mean of X_2 , and one standard deviation above the mean of X_2 , in which case $k = 3$. Now create a data file containing $2k$ rows, such that the first k rows contain, for a variable named $X2$, the values of X_2 you decided to use on the X axis, in sequential order. Repeat this for the next k rows. For example, if $k = 5$ using numbers 1 through 5, the first row would have $X2 = 1$, the second would be $X2 = 2$, etc., the fifth row would have $X2 = 5$, the sixth row would have $X2$ set to 1, and so forth., and the tenth row would have $X2$ set to 5.

(2) For the first k rows of this data file you started creating in (1) above, set a new variable $X1$ equal to the numerical code used to define group 1. And for the next k rows, set that variable $X1$ to the value used to define group 2. For instance, if your dichotomous variable (X_1) was coded 0 for group 1 and 1 for group 2, then set $X1 = 0$ for the first k rows and $X1 = 1$ for the next 2 rows.

(3) Create a set of additional variables that contain the mean of the covariates in the model for the data you have analyzed. For example, if you have two covariates Z and W , create two columns that contain, for every case, the mean of Z in the first column and the mean of W in the second column. (If you have no covariates in the model, then this step is unnecessary).

Suppose, for example, that you were interested in plotting the interaction between X_2 (measured on a 1 to 5 scale) and X_1 (1 for males and 0 for females) controlling for Z and W (with means of 6 and 8.5, respectively), your data file might look like this:

X2	X1	Z	W
1	0	6	8.5
2	0	6	8.5
3	0	6	8.5
4	0	6	8.5
5	0	6	8.5
1	1	6	8.5
2	1	6	8.5
3	1	6	8.5
4	1	6	8.5
5	1	6	8.5

(4) Using the regression model estimated on the original data (not the data in the table you just created), generate \hat{Y} but *applying that regression model to the data in the table you did just create*. For example, suppose the regression model from the analysis of the original data was

$$\hat{Y} = 3.4 + 8.3X_1 - 0.2X_2 - 7.3X_1X_2 - 0.1W + 0.3Z$$

Then in SPSS, the command would be

```
compute yhat = 3.4+8.3*X1-0.2*X2+7.3*X1*X2-0.1*W+0.3*Z.
```

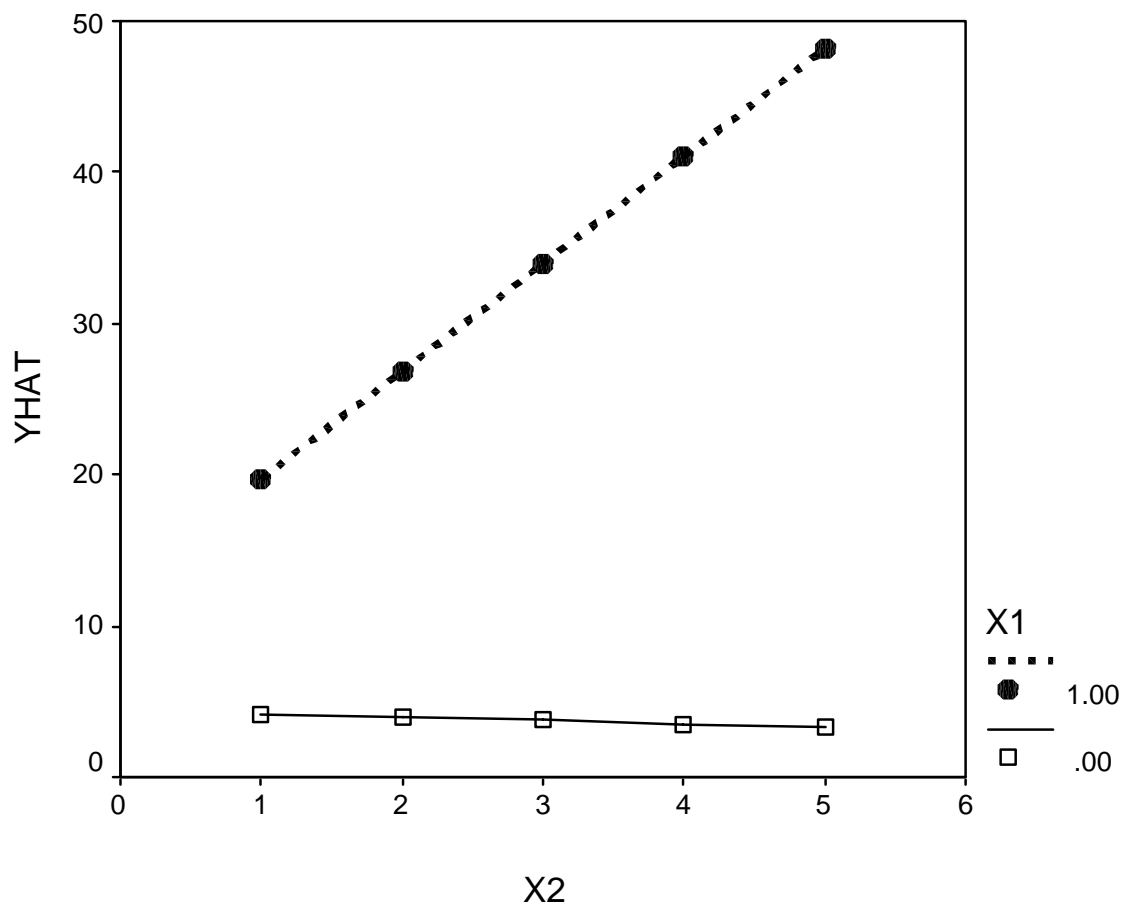
where the regression coefficients and intercept in the compute command are from the regression model estimated in your data (not the data in the table you just created). This will create a new variable in the data set you just created called “yhat.” These are the estimated values of Y using the regression model from the data analyzed.

(5) In SPSS, the following command will generate the estimated values of \hat{Y} from the model, using different markers for the two groups defined by X_1 .

```
GRAPH
  /SCATTERPLOT(BIVAR)=x2 WITH yhat BY x1.
```

Using the editing features of SPSS, it is possible to edit the graph the command above produces to create a figure that looks like on the next page

This same procedure can be used to graph an interaction between a quantitative and a multicategorical variable. In that case, you’d have several dummy variables coding group membership and several corresponding interaction terms in the regression model and your compute command.



From the CD accompanying Hayes, A. F. (2005). *Statistical Methods for Communication Science*. Mahwah, NJ: Lawrence Erlbaum Associates.