RegVal2025

RegVal2025 is a tool to calculate predicted values based on the results of a moderated regression analysis. I originally programmed it many years ago and now have updated it. RegVal2025 can handle up to five independent variables (IV). These five IVs can be either all continuous variables (e.g. age, height) or all categorical variables (e.g. experimentally manipulated variables) or a mix of categorical and continuous variables. A categorical variable can have up to five levels. The program is written for the case of a full factorial model, that is, the model contains all main effects and all interactions.

When the relevant information to calculate predicted values has been entered via the "Data entry" Mode (see below), an Excel file is saved containing both the entered information and the predicted values. Changes can then also be made in the saved Excel file, and predicted values can be recalculated by reading data from the Excel file (see below "Read data from file" Mode).

Thus, the program may be helpful particularly in those cases in which the same regression equation (i.e., containing the same set of predictors) is tested repeatedly (e.g. with different dependent variables, or when the final sample has been reached).

I believe the tool is relatively easy to use, but it is also very easy to make a mistake. This is true especially when you use the "Read data from file" Mode. So please read through this explanation thoroughly and be careful when you make changes to an existing Excel file to recalculate predicted values.

Rene Ziegler, University of Tübingen, Germany

February 2025

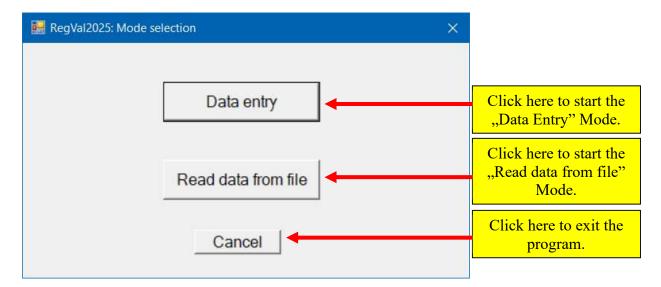
TOC:

Mode selection p. 2
Date entry p. 3
Read data from file p. 17

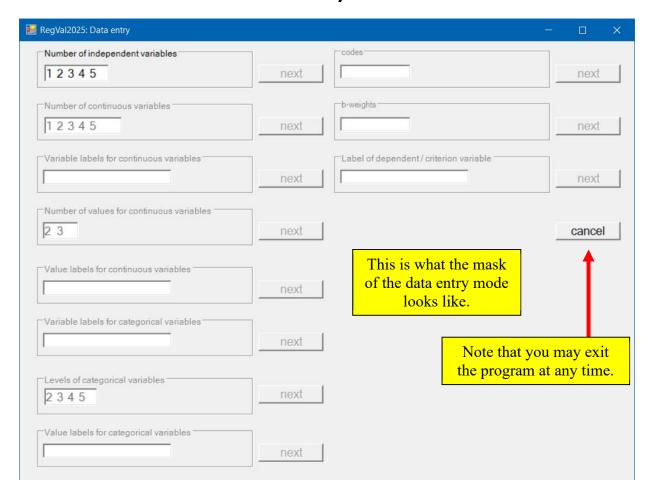
 $^{^1}$ Of course, it is unrealistic to run a five-factorial experiment with each factor having 5 levels (because 5 x 5 x 5 x 5 x 5 x 5 = 3125 cells). Nonetheless, at times, there may be reason to run, for instance, a 5 x 2 factorial design. Or you may be interested in one or the other interaction of 5 continuous variables (e.g. the Big Five personality factors).

Mode selection

Once you have installed the program, when you start it, choose whether you want to enter data via the "Data entry" Mode or whether you want to recalculate predicted values by reading data from file. Note that you need to run the data entry mode at least once before you can read data from file.

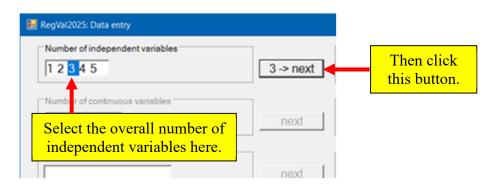


"Data entry" Mode



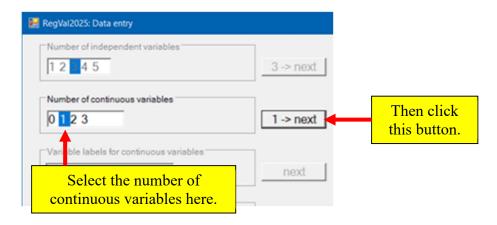
Step 1: Number of independent variables

How many independent variables are there altogether? At this step it doesn't matter whether they are categorical or continuous. For instance, if there is one continuous variable (e.g. "Need for Cognition") and two categorical variables ("Argument Quality" and "Source Expertise"), select "3" and click the "next" button.



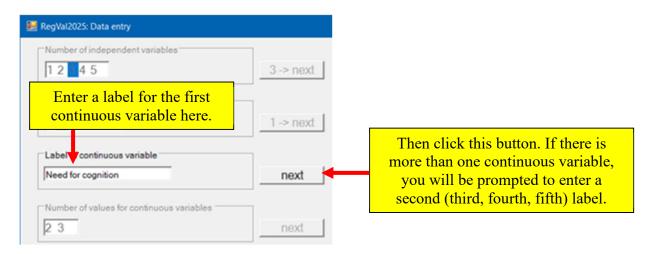
Step 2: Number of continuous variables

How many of the independent variables are continuous variables? For instance, if there is one continuous variable (e.g. "Need for Cognition"), select "1" and click the "next" button. The maximum number you can select is adjusted to the overall number of independent variables specified in Step 1. If you select "0" because there is no continuous independent variable (i.e. only categorical variables), Steps 3, 4, and 5 are skipped.



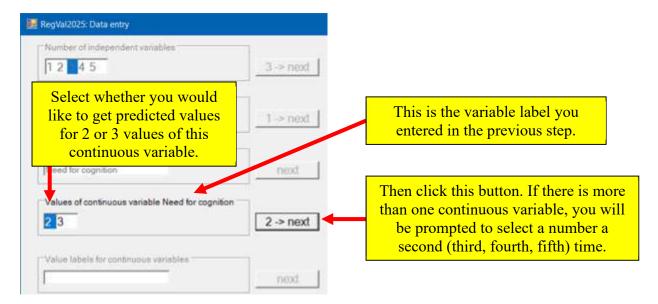
Step 3: Variable labels for continuous variables

Please enter a label for each continuous variable. For instance, type "Need for Cognition" and click the "next" button. If there is more than one continuous variable, you will be prompted to enter a second (third, fourth, fifth) label.



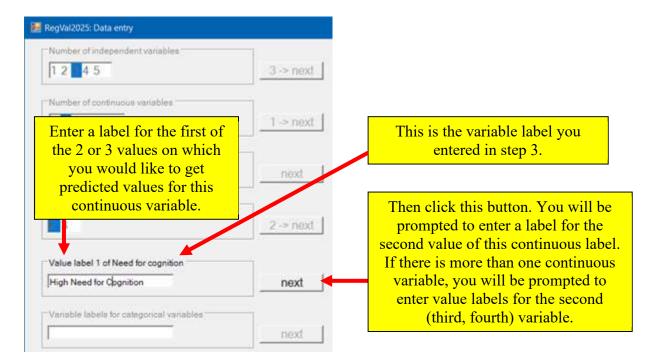
Step 4: Number of values for continuous variables

For how many values (2 or 3) of a continuous variable would you like to get predicted values of your dependent variable? The caption specifies the variable label of the continuous variable entered in Step 3. For instance, you might want to get predicted values for those "high" in Need for Cognition and for those "low" in Need for Cognition. Accordingly, select "2" and click the "next" button. If there is more than one continuous variable, you will be prompted to select a number a second (third, fourth, fifth) time.



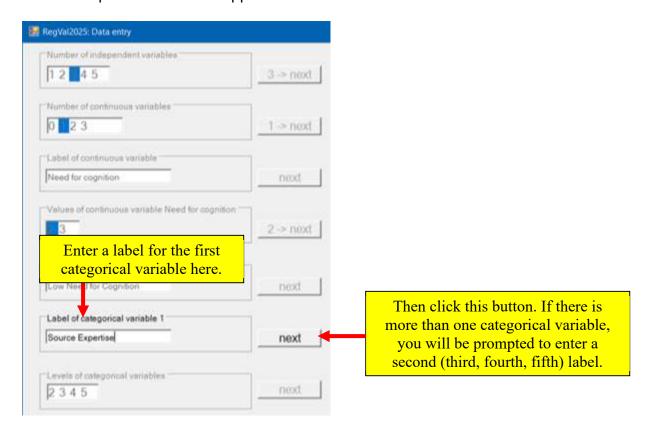
Step 5: Value labels for continuous variables

Please label the values of the continuous variable for which you will get predicted values. The caption specifies the variable label of the continuous variable entered in Step 3 (e.g. "Need for Cognition"). For instance, if you selected "2" in Step 4, now you might label them as "High NC" and "Low NC". Accordingly, first enter "High NC" and click the "next" button. Then, enter "Low NC" and click the "next" button. If there is more than one continuous variable, you will be prompted to enter value labels for the second (third, fourth) variable.



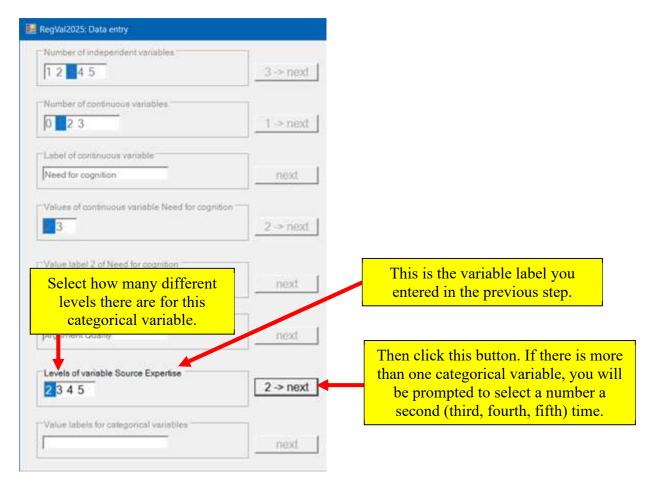
Step 6: Variable labels for categorical variables

Please enter a label for each categorical variable. For instance, type "Source Expertise" and click the "next" button. If there is more than one categorical variable, you will be prompted to enter a second (third, fourth, fifth) label. If there is no categorical independent variable (i.e. only continuous variables), this Step as well as Steps 7 and 8 are skipped.



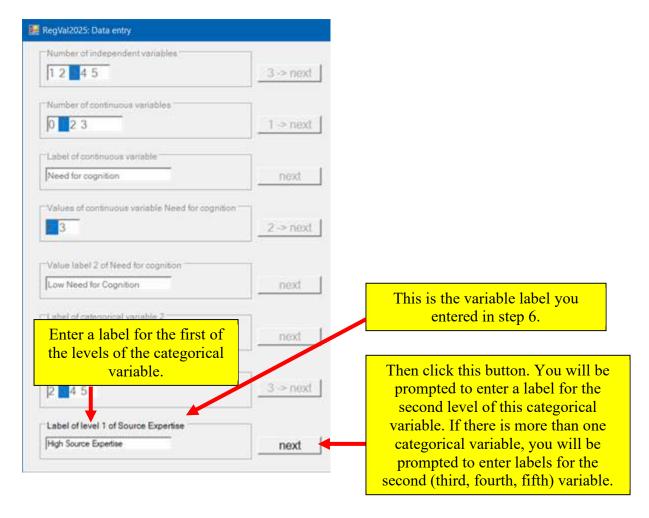
Step 7: Levels of categorical variables

How many different levels of the categorical variable are there? The caption specifies the variable label of the categorical variable entered in Step 6. For instance, "Source Expertise" might be either high or low. Accordingly, select "2" and click the "next" button. If there is more than one categorical variable, you will be prompted to select a number a second (third, fourth, fifth) time.



Step 8: Value labels for categorical variables

Please label the levels of the categorical variables. The caption specifies the variable label of the categorical variable entered in Step 6 (e.g. "Source Expertise"). For instance, if you selected "2" in Step 7, now you might label them as "High Expertise" and "Low Expertise". Accordingly, first enter "High Expertise" and click the "next" button. Then enter "Low Expertise" and click the "next" button. If there is more than one categorical variable, you will be prompted to enter level labels for the second (third, fourth, fifth) variable.



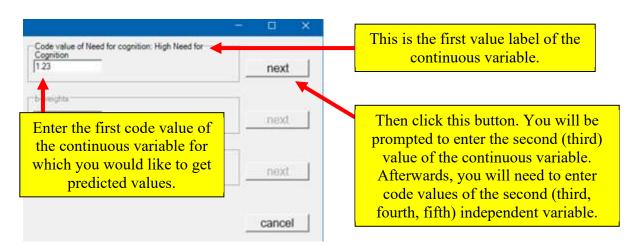
Step 9: Code values

Enter the code values for the independent variables. You will be prompted to enter the code values in the order the variables were declared. For instance, one of the three independent variables may have been specified as continuous (Step 2), and labeled as "Need for Cognition" (Step 3). Further, you may have selected that you would like to get predicted values for two values of this independent variable (Step 4), with value labels "High Need for Cognition" and "Low Need for Cognition" (Step 5). Then you will first be prompted to enter a code for "High Need for Cognition". Given centered NC scores with higher scores indicating higher Need for Cognition, you might enter the standard deviation (e.g. "1.23") and then click the "next" button. Next, you will be prompted to enter a code value for "Low Need for Cognition", so you might enter minus the standard deviation (e.g. "-1.23"). If NC scores were standardized (with higher scores indicating higher Need for Cognition) you might enter the standard deviations "1" and "-1".

The second independent variable may be a categorical variable, labeled as "Source Expertise" (Step 6), with two levels (Step 7) labeled "High Expertise" and "Low Expertise" (Step 8). Then you will be prompted to enter a code value for "High Expertise". For instance, in the case of effect coding (High Expertise = 1; Low Expertise = -1) enter "1" and click the "next" button. Next, you will be prompted to enter a code value for "Low Expertise", so you might enter "-1" and click the "next" button.

Let's assume the third independent variable is a categorical variable with three levels. For instance, "Argument Quality" is good, medium, or bad. In this case, you will be prompted to enter code values in the order: first level on the first vector, second level on the first vector, third level on the first vector, followed by the same order for the second vector.

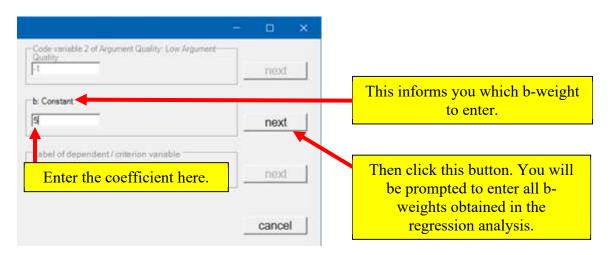
Please note that the program accepts only digits, the minus sign, and the decimal indicator (dot or comma, depending on your computer system) as entries.



Step 10: regression coefficients

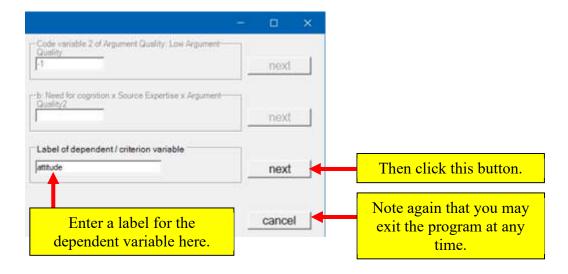
Please enter the regression coefficients (b-weights) pertaining to the predictor specified in the caption. For example, there may be three independent variables, one continuous variable CON1 (e.g. "Need for Cognition"), and two categorical variables CAT1 ("Source Expertise") and CAT2 ("Argument Strength"). The first categorical variable has two levels ("High Expertise" and "Low Expertise"), the second categorical variable has three levels ("Good Arguments", "Medium Arguments", and "Bad Arguments"). Accordingly, there are two vectors for the second categorical variable CAT21 and CAT22. The order in which regression-coefficients have to be entered is as follows:

- Constant (b₀)
- CON1, CAT1, CAT21, CAT22 (i.e., all "main effects" first)
- CON1xCAT1, CON1xCAT21, CON1xCAT22
 CAT1xCAT21, CAT1xCAT22 ("two-way interactions" next)
- CON1xCAT1xCAT21,
 CON1xCAT1xCAT22 ("three-way interactions" last)



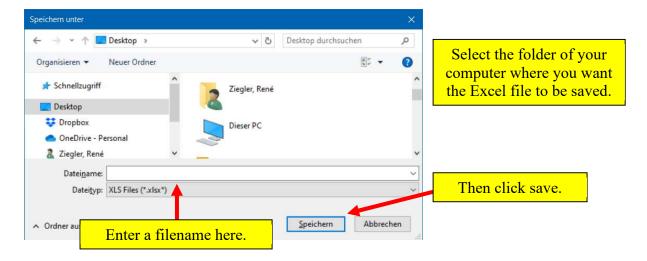
Step 11: Label of dependent / criterion variable

Please enter a label for the dependent / outcome / predicted / criterion variable and click the "next" button.



Step 12: Save Excel file

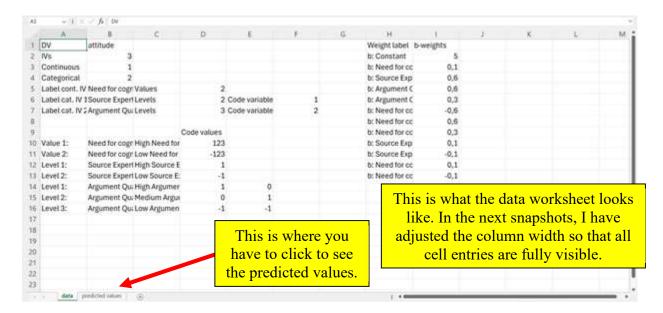
You will be prompted to enter a filename. All entries and the calculated predicted values will be saved in an Excel file. RegVal2025 will then automatically be terminated. The Excel file consists of two worksheets. The first one (labeled "data") contains all the entries made in Steps 1 through 11. The second one contains the predicted values and is labeled respectively (see below).



This has been the final step of the data entry mode. Next, I describe what you find in the Excel file.

The Excel file should open automatically once you have saved it in Step 12. In any case, you should find the following information in the Excel file:

Worksheet "data":



In column B you find:

Row 1: Label for dependent variable (Step 11)

Row 2: Number of independent variables (Step 1)

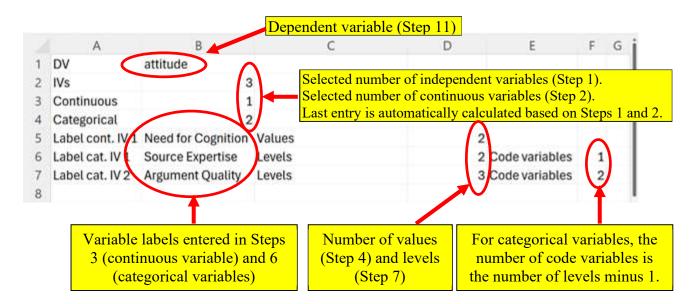
Row 3: Number of continuous variables (Step 2)

Row 4: Number of categorical variables (derived from Steps 1 and 2)

Subsequent rows:

Column B / column D / column F:

Variable labels for continuous IVs (Step 3) / Number of values (Step 4) / - Variable labels for categorical IVs (Step 6) / Number of levels (Step 7) / Number of code variables

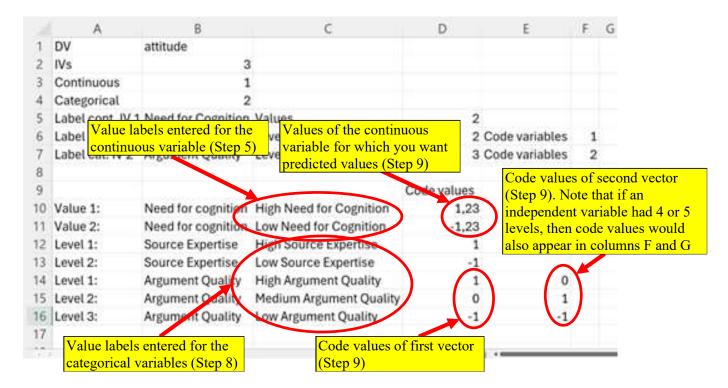


Further rows:

Column C / column D (E, F, G):

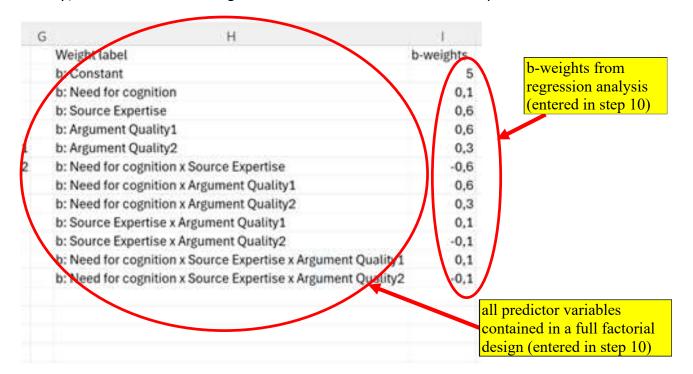
Value labels of continuous IVs (Step 5) / Code values (Step 9) / -

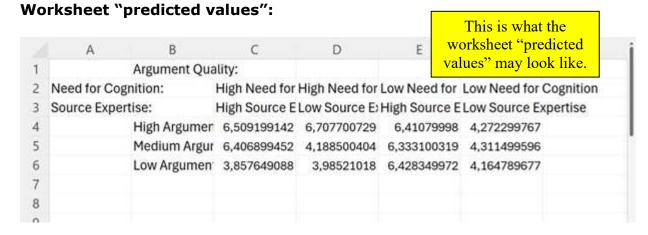
Value labels of categorical IVs (Step 8) / Code values of first vector (second, third, fourth vector) (Step 9)



In column H, labels of all predictor variables contained in a full factorial design (i.e., all main effects and interaction terms that are possible given the number and kind of independent variables) are listed.

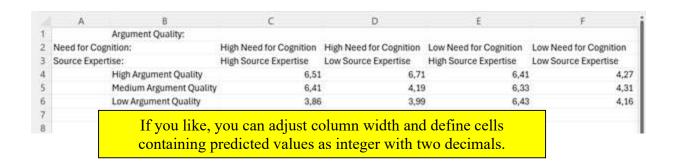
Finally, column I lists the regression coefficients entered in Step 10.

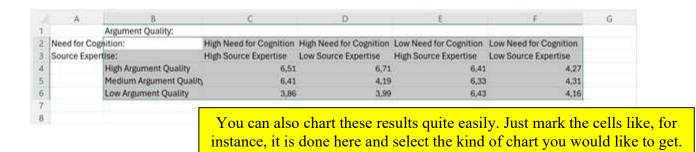


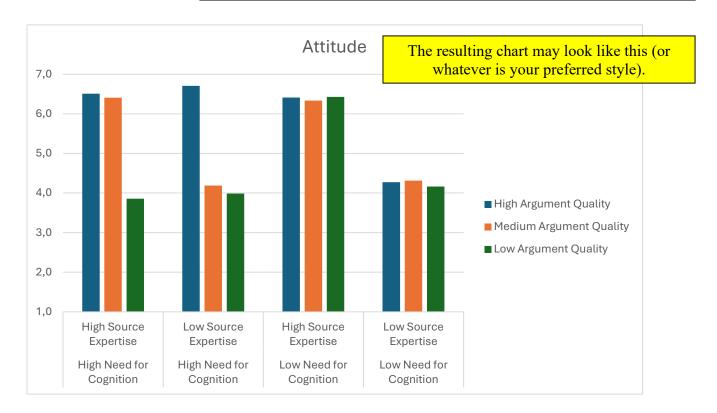


The entries depend on the number of IVs (organized similar to what is common for tables in scientific papers in case of two or three IVs).

- 1 or 2 IVs: The second row shows the variable label of the first IV in the first column ("A") and its value labels starting from the third column ("C"). The second column ("B") shows the variable label of the second IV in the first row and its value labels starting from the third row.
- 3 IVs: The second and third row show the variable labels of the first and second IVs in the first column ("A") and their value labels starting from the third column ("C"). The second column ("B") shows the variable label of the third IV in the first row and its value labels starting from the fourth row.
- 4 or 5 IVs: The second, third and fourth column ("B", "C", and "D") show the variable labels of the third, fourth and fifth IV, respectively (first row) and their value labels starting from the fourth row. The second and third row show the variable labels of the first and second IVs in the first column ("A") and their value labels starting from either the fourth column ("D", given 4 IVs) or the fifth column ("E", given 5 IVs).





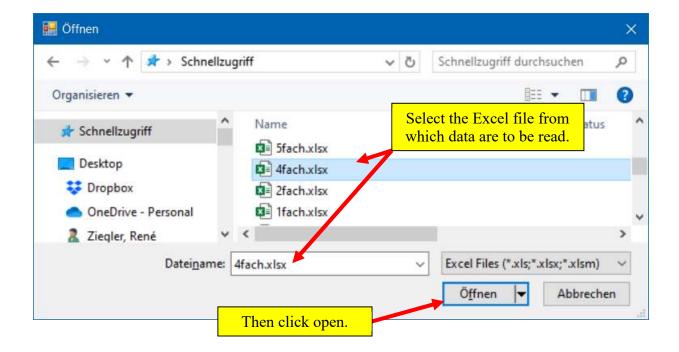


This is all there is to say on the data entry mode and the saved Excel file. The following pages describe how to make use of the "Read data from file" Mode.

"Read data from file" Mode

When all the information necessary to calculate predicted values has been saved in an Excel file, it is possible to read information from such a file and recalculate predicted values. For instance, you may have conducted a regression analysis before the last few participants have been run in the lab. Or you run the same regression analysis on another dependent variable. Rather than re-entering all the information via the data entry mode, you may change only the information in the Excel file that is different from the original data entry.

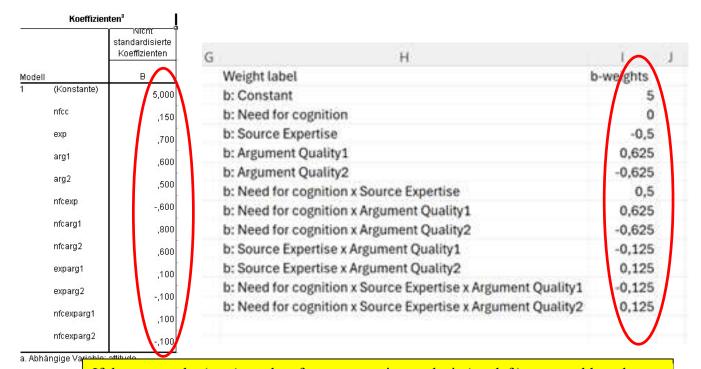
When these changes have been made, it is neither necessary to save the file nor to close the file. Just start the program, select the read-from-file mode and the appropriate Excel file, and the recalculation takes effect immediately, opening a copy of the Excel file (you need to save the changes when closing the file).



On the next three pages I describe which kinds of changes can be made in an existing Excel file created via the data entry mode.

1. Change of regression coefficients. If the order in which they are listed in the output of your statistics software is identical to the order in which the weights are listed in the worksheet, you may simply copy and paste the obtained regression coefficients.

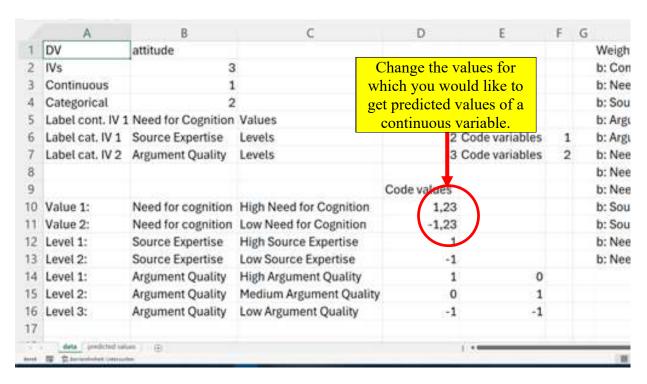
A friendly but urgent warning is in place here: You _cannot_ simply reorder the weights in the worksheet. This would lead to completely wrong predicted values because the program assumes that certain b-weights (i.e., those pertaining to specific effects) are found in certain cells of column I.



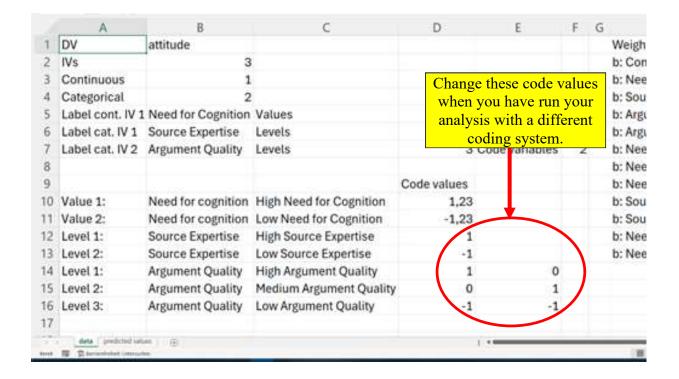
If these were the (new) results of your regression analysis (see left), you could mark these new coefficients and copy and paste them into column I of the worksheet "data" (see old b-weights right).

Please note that this is only possible, when the order in the output table is identical to the order in the worksheet. If this is not the case, make sure that each coefficient is entered into the cell in column I that pertains to the predictor that is mentioned in column H.

2. Change of standard deviation of a continuous variable. In this case, entries in those cells of the worksheet that contain the values for the continuous variable may have to be changed (e.g., if predicted values at one standard deviation above and below the mean are of interest).

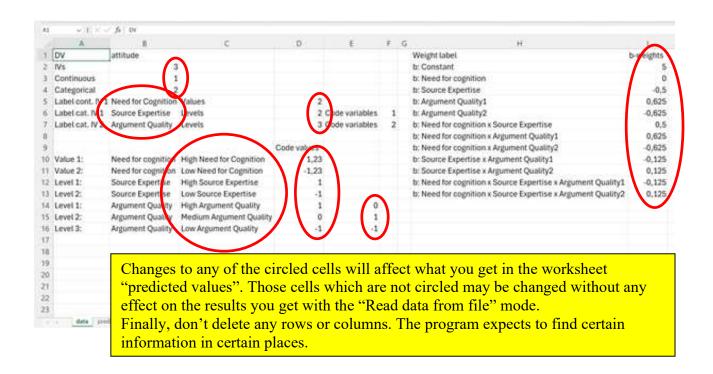


Change of the kind of coding of a categorical variable. In this case, make sure to adjust the code values for the categorical variable contained in the worksheet (e.g. dummy coding instead of effect coding).



In general, when data are read from an existing file, the program makes use of the following information:

- Number of independent variables (Column "B", second row), number of continuous variables (Column "B", third row), number of categorical variables (Column "B", fourth row)
- Variable labels for continuous IVs (if there are any; Column "B", starting fifth row), number of values (Column "D", same row)
- Variable labels for categorical IVs (if there are any; Column "B"), number of levels (Column "D", same row)
- Value labels of continuous IVs (Column "C"), code values (Column "C", same row)
- Value labels of categorical IVs (Column "C"), code values of first vector (second, third, fourth vector (Column "D", "E, "F"; same row)
- Regression coefficients / b-weights (Column "I", starting second row)



That's it. I hope the tool is of help for the kind of analysis you want to run.