**COURSE\_NUM – Multivariable regression - Workshop Activity**

**Task:** The goal of this workshop is to introduce you to running multiple regressions and creating interaction terms in Excel. This will prepare you for running regressions as part of the Module 4 main assignment.

**Preparation**

Begin by opening the Excel file “Multivariable regression – Workshop data”

You may need to add the “Data Analysis” add-in to your version of Excel.

* Click on “Data” in the header ribbon

If you do not see the “Data Analysis” option, you need to install the add-in.

* Click on the Windows symbol in upper left-hand corner OR “File” in the header ribbon.
* Click on “Info”, and then below in the same column, click on “Options.”
* A new box or options will pop up. In the left-hand column, click “Add-ins.”
* Select “Analysis ToolPak” by clicking on it and then click the “Go” button below.
* Another box may pop up with a list of “Add-ins available.” Select “Analysis ToolPak” and click “OK.”
* You will know you are successful when back in your Excel document you click “Data” in the header ribbon and see “Data Analysis” listed in the “Analysis” box.

Other elements that we learned in the Module 3 workshop that we are building on here are:

* Creating dummy variables
* Linear regression

Please go back to the workshop instructions for Module 3 if you need the instructions again about how to conduct those procedures.

**Part 1: Estimate an OLS multivariable regression model**

We are interested in the association between discrimination, material wellbeing, and trust in government/parliament, *net of other factors that might be confounded with trust, material wellbeing, and discrimination.* These include factors such as age, gender, neighbourhood deprivation, and disability status.

For example, we want to know that the association between someone’s material wellbeing and their trust in the government is actually due to their material living standards, and not just explained by the fact they are also more likely to live in high deprivation neighbourhoods where people are also more likely to be distrustful of the government.

Let’s run a multivariable regression model, where:

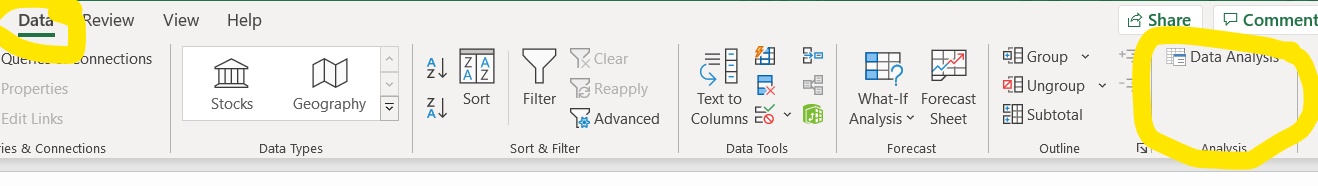
* Y: dependent variable = trust in government
* X: focal independent variables = discrimination and material wellbeing (MWB)
* X: covariates = age, sex, neighbourhood deprivation (NZDEP), disability status

The regression can be expressed like this:

*Preparation: For multivariable analyses, all the “X input” needs to be contiguous—that is, all the variables/columns that you are going to use need to be next to each other. Lucky for you, in this example, the variables are all next to each other so you’re ready to go. But when you do your own analyses with your own data, you may need to reorder your columns by cutting/copying and pasting them beside each other.*

Follow these steps:

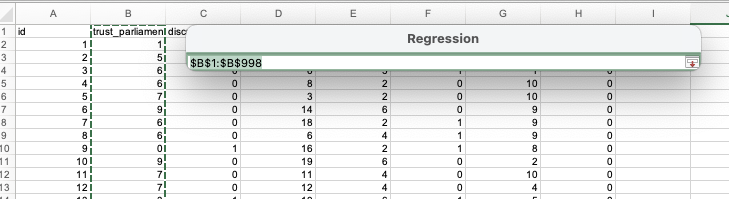
1. Click on “Data” in the header ribbon.
2. Click on “Data Analysis” in the “Analysis” box.



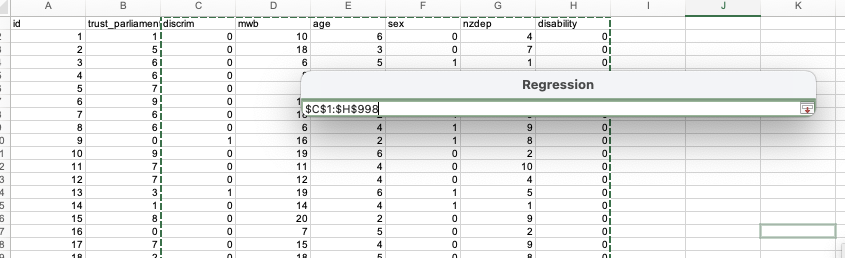
1. A box will pop up. Click on “Regression” in the list of “Analysis Tools”, then click “OK”
2. Input the “Y” data—data in the “trust\_parliament” variable column—by clicking the little arrow button in the “Input Y Range” row.



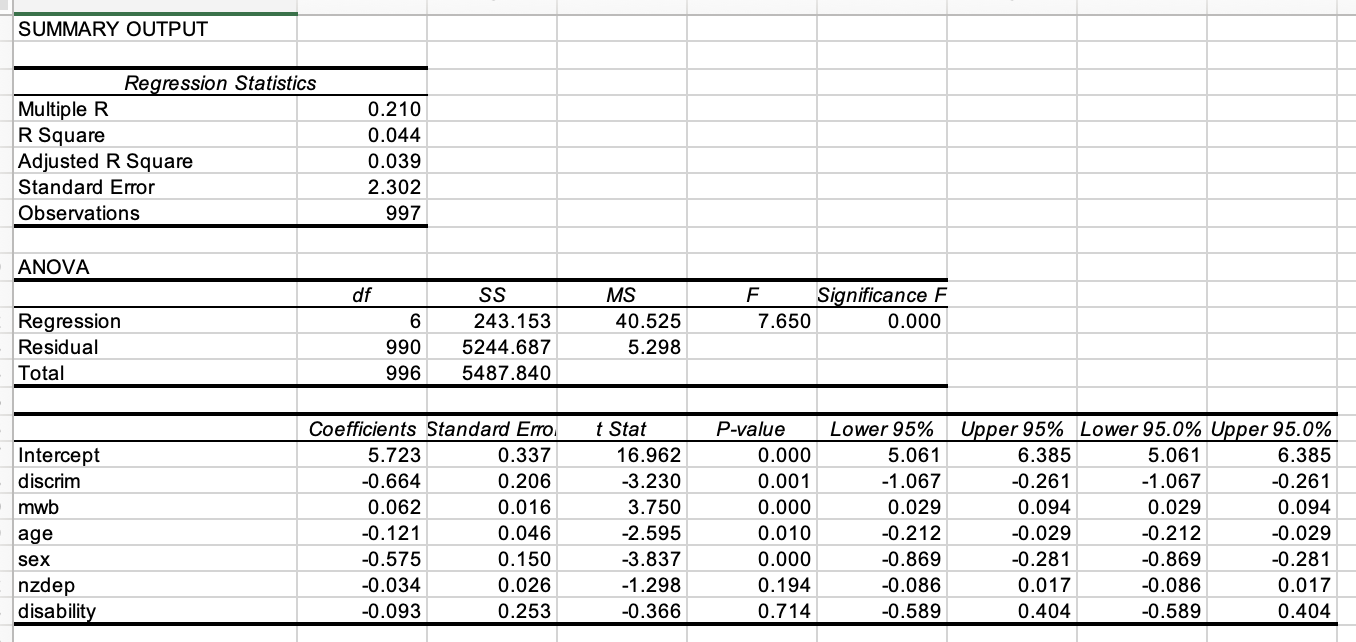
1. A new blank box titled “Regression” will appear. Select the data by clicking the variable name and dragging to the end cell. In this dataset, that will be row 998.



1. Once data are selected, click the little downward arrow button to the right, which will bring you back to the main “Regression” box.
2. Input the “X” data—data in the “discrim”, “mwb”, “age”, “sex”, “nzdep”, and “disability” columns. Do this by selecting the variable name from the first column, dragging down and diagonal across all six columns until you’ve selected all data in all six variable columns.



1. Select “Labels” for including variable names in your output.
2. Select “New Worksheet Ply:” and type “Trust-Multiple” in the text box to the right. This will create a new sheet to store your regression results.



* Remember that you can use the little decimal place adjustment buttons to make it easier to read and interpret your output:

Text

Description automatically generated with medium confidence

***Regression output questions:***

Examine the regression output:

1. What does the R2 tell us about how much variance in the trust in parliament measure was explained by the X variables in our model?
2. Which of our independent variables were *not* statistically associated with trust in parliament?
3. Fill in the gaps: Experiencing discrimination in the past 12 months was \_\_\_\_[negatively/positively]\_\_\_ associated with \_\_\_\_\_\_\_\_\_\_\_, whereby people who have experienced discrimination, on average, report \_[coeff. value]\_\_ \_\_[lower/higher]\_\_ values on the trust scale compared to those who did not experience discrimination, net of other factors included in the model. This finding was statistically significant (*p <* .001).
4. The association between sex (being a woman, vs. being a man) and trust in government, and the association between experiencing discrimination and trust in government, had around the same effect size. Did this surprise you?
5. What would the coefficient contribution to reported trust be for people with a material wellbeing index of 10?
6. The estimated difference in trust in government between someone with a material wellbeing index (0-20 scale) of 0 and someone with a material wellbeing index of 10 is around the same difference as someone who has experienced discrimination in the past year versus someone who has not experienced discrimination in the past year. Does this surprise you?

**Part 2: Create an interaction term and re-estimate the OLS multivariable regression model**

As well as hypothesising that there is an association between discrimination, material wellbeing, and trust in government, we also hypothesise that experiences of discrimination may matter more so for people who have fewer resources to fall back on or buffer them from those discrimination experiences. Put simply, we think the association between discrimination and trust in government will be different for those with more material resources compared to those with fewer material resources (i.e., the slope between material wellbeing and trust will differ by discrimination).

We can test this hypothesis through adding an *interaction term* into our model. Our new equation can be expressed like this:

*Create the interaction term*

First, let’s create the interaction term between discrimination and material wellbeing:

1. Label column I (i.e., a new column): discrim\_mwb
2. In row 2, column I, enter the formula that multiplies the discrim variable with the mwb variable: “=C2\*D2”. The \* symbol tells excel to multiply the values.

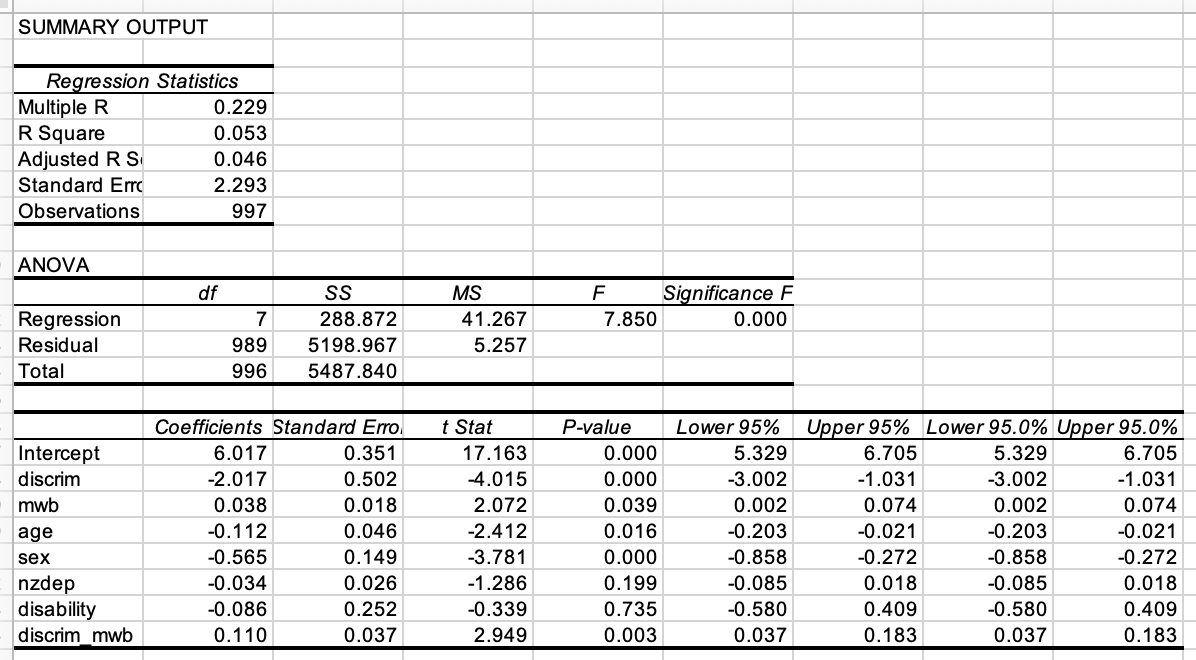
Table

Description automatically generated

1. Fill the remainder of the column with that equation, altering the formula for each row (e.g., = C2\*D2, then the next row is =C3\*D3, the next row is =C4\*D4). There are two ways to do this. Move the cursor to the bottom right-hand corner until it turns into a cross, then either:
   * Double-click the corner. It will populate to the end of the column; or,
   * Hold the mouse button down while dragging the cross down to the end of the column.

*Run the multivariable regression*

1. Follow the same steps as above for running a regression model, but this time, also include the interaction term in your models.
2. Create a new sheet for output. Label it “Trust-Interact”



***Regression output questions:***

Examine the regression output:

1. What difference do you see between the R2 in the first regression model (without the interaction term) and the R2 in this one with the interaction term? Did the interaction term improve the model fit?
2. What does the p-value of the discrim\_mwb interaction term tell us about whether the association between discrimination and trust in government is different if you have more or less material resources?
3. Intuitively, what does the 0.110 interaction term coefficient tell us about the association between material wellbeing and trust in government?
4. Plug in the coefficient values (just use the intercept/alpha, discrim, mwb, and discrim\_mwb) to estimate the predicted trust values for someone who:

* Has experienced discrimination and has a material wellbeing index of 1:
* Has experienced discrimination and has a material wellbeing index of 10:
* Has not experienced discrimination and has a material wellbeing index of 1:
* Has not experienced discrimination and has a material wellbeing index of 10:

1. What do the estimated values produced in Q4 above say about the association between discrimination and trust in government?