

Socio-economic Classifications

Supplemental: Deriving the NS-SEC

David Pevalin

Hello, and welcome to a supplemental video to this resource on socio economic classifications. This supplemental deals with the practicalities of deriving NS-SEC on datasets using the derivation matrices that are supplied on the ONS website. So, if we start off with the minimum requirements that you have to have in your data, at an individual level, you have to have the SOC, the Standard Occupational Classification 2020 value to occupational unit level, which is OUG level and that's the four digits at level of SOC to do that, that's what the NS-SEC is designed to do. And you need to have that for the individuals in your dataset. Obviously, if you've got a historical data that uses SOC 2010, then you use the 2010 derivation matrices for NS-SEC. And similarly, for Soc 2000. So, there are three derivation methods to get NS-SEC depending on the level of information that you have, in your data on employment status.

And the full method can be used when you've got enough data to create a seven-category employment status variable that has these seven categories. So, first self-employed with 25 or more employees, so there won't be many of these within your data set, self-employed with the less than 25 employees. So these are the employers who we talked about in the in the NS-SEC video. Category three is self-employed, but with no employees also only count workers. And then the employees are split into four categories, managers in a larger establishment 25 or more employees, managers in a small establishment less than 25 employees, a supervisor and an employee there are, you know, issues of response bias when it comes to asking people whether they're a supervisor or not, and they're there. So, you have to be careful with what you've got in your in your data about what actually a supervisor is and who's who's categorised as such, and similarly for manager.

So, if you don't have enough data to create a seven-category version, the one you're less likely to have is the number of employees within the workplace and this is particularly so for registration data, official registration data such as birth registration, marriage death registration data. So, as part of the original process, we had to create a reduced method of derivation that only used five category employment status variables. So, it was missing the the number of employees or number of employees within the establishment. So, the five categories in this version here, so number one is just self-employed with employees and self, then secondly, self-employed with no employees and then the employees are split into three, manager, supervisor and employee. So, the process of of deriving NSC practically within your data set and whatever software packages and your users are using is similar whether you're using the seven category or the five category, there's there's just a minor adjustments to the code. But there is a way that if you have no data available on employment status, but you must still have the SOC Code, there is a way called the simplified method, which is simply a straight match on to the occupational unit group level for, for NS-SEC.

So, this is an extract of the table that's available on the derivation matrix that's available on the ONS website for SOC 2020. So, this is the full derivation matrices. As you can see here, these are the seven employment status variables. And then with the full method and the reduced method table there is a column for SS-CC which is the simplified method, which is what you use if you don't have any information. To create this, this variable for an individual. So, as you can see here in this extract, some of the cells are shaded. What this is, is the shaded cells represent an ineligible occupation by employment status combination. Now, you may well have people in your data set that come up with this, but that's in regard, this is based on who's meant to go into these occupational unit groups. And then in terms of the standard occupational classification, but then people are self-reporting. Normally, they're the date the variables that go into into the employment status variable, so some you might get some ineligible mixes. So, for example, if we look at occupational unit group 3115, here quality assurance technicians, they are ineligible to be managers, because it's not within the definition of that occupational unit group in SOC. So, they, so these cells are shaded, and then they are filled with a value that gets you to an operational category known SSEC. But they're based on a set of priority rules, derived in conjunction with ONS that are on the website about how these ineligible cells are filled. So conversely, from this where quality assurance technicians are not allowed to be managers, we see here for occupational unit group 2494, advertising accounts managers and creative directors, these are managers. And actually, in accordance with the definition of the occupational unit group, they shouldn't be supervisors or employees. So that's why these are shaded. So, apologies.

So then, if we look at an extract of the reduced derivation matrices, we still have the SSEC, the simplified method put in there in pale orange colour. And then there are five columns for the five-category employment status variable, and the matrix is laid out in a similar way, with shaded combinations, but filled with the priority rules again. So, to use the full derivation matrix, you match it onto your data by the SOC, OUG at the individual level. So, what that does, is it puts a row of the derivation matrix alongside each individual in your main data at the individual level. And it adds eight new variables to your data set, which are the eight columns in the full derivation matrix. So SSEC. su25p, this this row here of these eight data's these eight variables. And then to create an NS-SEC value for each of them. This code will vary depending on what software you're using, but basically, if your employment status variable with seven categories is one, then the NS-SEC value for that person that, or those people with a value of one comes from the se25p column and similarly down through this and so, you you write this code and whatever is appropriate syntax for your your software, and if an individual does not have any data, so, if your empstat seven variable is missing, but you still have a SOC, OUG, then you can say that the NS-SEC is equal to the SSEC the simplified method in your data. You may want to do checks on that and on to see how many is coming up for each one just to make sure you're using as much data as possible with your employment status to get the best result.

And so, similarly, for the reduced method. You match on your reduced derivation matrix in a similar way. This time the merge creates six new variables, the these listed here, and then the code is similar if your employment status variable with five categories is equal to 12345, then your NS-SEC are these corresponding new variables. And similarly, if you don't have the data to create the employment status five variable, then you can use the SSEC simplified column in your data.

There's more information available on the ONS website about this and about the question if you're designing a survey, like which questions to ask in the survey, so you can create different values for your employment status. And so it's done in much more details down to the wording of the question to ask as a little bit of an aside, there's there's a, there's a gap between the simplified method which is no, no, no data for employment status, and then the reduced method which you can you have enough information to create five categories. Not done in an official way for ONS. But I've, I've, I have experimented with a two-column matrix, which if you only have information, whether it's somebody self-employed or employed, but with no further information, and that gives an improvement over the simplified methods.

So my contact details are below if you if you find yourself in a position where you've got data with with self-employed or employed but nothing else, then contact me and I can I can talk you through the process of creating a two, a two column matrix to derivative but because that wasn't officially done for ONS that's why it's not available on their on their website. Anyway, best of luck with all of your projects, and I hope this has been useful to you. Thank you. Bye bye