Q-STEP STATA 'HOW TO' GUIDES:

HOW TO RECODE VARIABLES IN STATA

Creator: Joshua Townsley

Before analysing quantitative data, sometimes we might want to recode data. This might be because we want to reverse the ordering of the coding, or simplify a survey variable by grouping responses together.

This guide offers a step-by-step approach to recoding a variable in STATA. The example used in this guide is based on the Welsh Election Study (WES) – a survey of registered voters in Wales.

Step 1: Deciding How to Recode

The first step is to look at the variable in question and decide what recoding we would like to carry out. For example, we have a variable called "profile_gross_household" that asks respondents what their gross household income is.

codebook profile_gross_household, tab(100)

The responses are as follows:

Q-Step A step-change in quantitative social science skills Funded by the Nuffield Foundation, ESRC and HEFCE
Nuffield Foundation
Sheffield Methods Institute.
The University Of Sheffield.

```
          type:  numeric (double)
         label:  profile_gross_household

         range:  [1,17]                       units:  1
unique values:  17                         missing .:  2/4,165

   tabulation:  Freq.   Numeric  Label
                  129         1  under £5,000 per year
                  253         2  £5,000 to £9,999 per year
                  395         3  £10,000 to £14,999 per year
                  340         4  £15,000 to £19,999 per year
                  399         5  £20,000 to £24,999 per year
                  317         6  £25,000 to £29,999 per year
                  260         7  £30,000 to £34,999 per year
                  199         8  £35,000 to £39,999 per year
                  189         9  £40,000 to £44,999 per year
                  162        10  £45,000 to £49,999 per year
                  173        11  £50,000 to £59,999 per year
                  115        12  £60,000 to £69,999 per year
                  121        13  £70,000 to £99,999 per year
                   34        14  £100,000 to £149,999 per year
                    9        15  £150,000 and over
                  287        16  Don't know
                  781        17  Prefer not to answer
                    2         .
```

As we can see, the variable coding ranges from 1-17, with higher values indicating a higher level of income.

Firstly, we might want to set some values as 'missing'. This means we tell STATA to ignore certain values we deem irrelevant. In this case, we will set 16 ('Don't Know') and 17 ('Prefer not to answer') as missing.

mvdecode profile_gross_household, mv(16 17)

Next, we might want to recode this variable so that some income categories are grouped thus:

1 "less than £20k" 2 "£20K - £40K" 3 "£40K - £60K" 4 "£60K +"

Step 2: Using the Recode Command

Once we have decided how we would like our new, recoded variable to look, we can use the RECODE command in STATA to do it. As a rule, we put the original values on the left-hand side of the '=' sign, and the new values on the right-hand side.

We also want to stipulate that we want to generate ("gen") a new variable from our recoding, called "IncomeCat".

recode profile_gross_household (1 2 3 4=1 "LessThan£20k") (5 6 7 8=2 "£20kto£40k") (9 10 11=3 "£40kto£60k") (12 13 14 15=4 "£60kandOver"), gen(incomecat)

```
(2966 differences between profile_gross_household and incomecat)
```

STATA then confirms that the new variable had been generated. STATA tells us that this has changed 2,966 data points.

Note: if we wanted STATA to group a range of responses together, we can use the "/" term, thus:

recode profile_gross_household (1/4=1 "LessThan£20k") (5/8=2 "£20kto£40k") (9/11=3 "£40kto£60k") (12/15=4 "£60kandOver"), gen(incomecat)

Step 3: Checking the Recoded Variable

Now we have recoded the old variable containing every income category into a new variable of 4 larger income groups, we can check the new variable using the CODEBOOK command.

codebook incomecat, tab(100)

Tasks:

1. Recode the 'age' variable into an 'agecat' variable that groups the ages of respondent into an intuitive system. Remember to consider carefully how you would like to group ages.

2. Recode the 'profile_education_level' variable into the following categories:

    a. No Quals
    b. GCSE/Equiv
    c. A Level/Equiv
    d. University
    e. Other