

Session 2-4 Workbook

Exploring Housing Benefits with the English Housing Survey

In the first session, we came up with questions around housing benefits: e.g. who gets it, how much they get, do people in London get more on average than people in other parts of the country etc. Now we'll use the EHS 2010-2011 to examine the facts!

Open SPSS 20 and open the data.

In the cluster we use for these sessions you should be able to SPSS as follows

Click on the  button in the bottom left corner of your screen

Select All Programs> IBM SPSS Statistics> IBM SPSS Statistics 20

Variables and datasets

In the list below are the variables we'll be using today. They are contained in two files. The first you'll use is called 'interviewfs10.sav' and the second is called 'generalfs10.sav'. Both contain variables about households from the English Housing Survey 2010-2011.

Name	Label	File
housbenx	Household (HRP + partner) receives any housing benefit?	Interviewfs10.sav
tenure1	Tenure Group 1	Interviewfs10.sav
hhstype6	Household type - 6 categories	Interviewfs10.sav
amthbenx	Weekly housing benefit	Interviewfs10.sav
hhsizex	Number of persons in the household	Interviewfs10.sav
agehrpx	Age of HRP – continuous	Interviewfs10.sav
agehrp6x	Age of household reference person - 6 band	Interviewfs10.sav
hyeargrx	Household gross annual income	Interviewfs10.sav
nbedsx	Total no of bedrooms household actually has	Interviewfs10.sav
aacode	EHS case number	Interviewfs10.sav
lmd1010	Deprivation - IMD2010 decile ranking of areas (lower SOAs)	generalfs10.sav
GorEHS	Region - EHS order	generalfs10.sav
aagfh10	Household weight (2010-11)	generalfs10.sav

Before we start, a word about measurement levels of variables.

Measurement levels: what they are and why they matter

Each variable has a defined measurement level. The measurement level determines how the variable is treated in analysis as there are recommended methods for handling particular measurement levels and combinations of measurement levels. The three main types of measurement level are summarised below with an example variable and the way in which we will be summarising each kind of variable today.

Type of variable	Example	Summarising the variable
Categorical: nominal	Vehicle type: -Van -Car -Truck -Motorbike -Bicycle	Frequency table Bar chart or pie chart
Categorical: ordinal	Satisfaction with vehicle: -Very high -High -Medium -Low -Very low	Frequency table Bar chart or pie chart
Scalar (or continuous)	Expenditure on vehicles in last year Age of car (in years)	Arithmetic mean (average) Histogram

Session 2: Exploratory analysis

In this session, we're going to examine some variables that should help us to learn a bit about Housing Benefit (HB), what kinds of households HB recipients are in and how much HB they get.

Work through the following analyses and answer the questions. Instructions are given in bold. You can find information about how to do things in SPSS in the separate *Guide to SPSS 20*. The section *About SPSS 20: Data View, Variable View and the Output window* on page 1 explains a bit about SPSS if you have forgotten about how to view data and variables in SPSS.



What proportion of households who rent receive housing benefit (HB)?

Have a guess:

What proportion of households in England rent?

Of these, what proportion receive housing benefit?

See what the person sitting next to you guessed.

The variable that tells us about whether or not households receive HB is *housbenx* "Household (HRP + partner) receives any housing benefit?". **Look at the 'Guide to SPSS 20': use the instructions:**

- **'How to create a one-way frequency table of a variable'** to make a frequency table of *housbenx*. Use ***Paste*** instead of ***OK*** to run your analyses - see Page 3 of the *Guide to SPSS 20* to see why you should use ***Paste*** and how to use it.
- **Note that the sections on Page 2 'Setting SPSS to display variable names or labels'** and Page 4 'Finding and selecting variables quickly from a list in a dialogue box' give you useful tips about using SPSS.

You should have created the frequency table below. It tells you how many survey respondents gave the answers "Yes" and "No", and how many missing responses there were – in this case, all the missing responses are from people for whom the question was irrelevant – i.e. "Not applicable – owner occupier".

Household (HRP + partner) receives any housing benefit?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	2645	15.1	47.2	47.2
	No	2962	16.9	52.8	100.0
	Total	5607	31.9	100.0	
Missing	Not applicable - owner occupier	11949	68.1		
Total		17556	100.0		

You can now answer the questions we asked earlier on: (did your neighbour make better guesses?)

What proportion of households rent (i.e. are not owner occupiers)?

What proportion of renters receive HB?

You should have found that 31.9% of households rent and, of those, 47.2% receive HB.

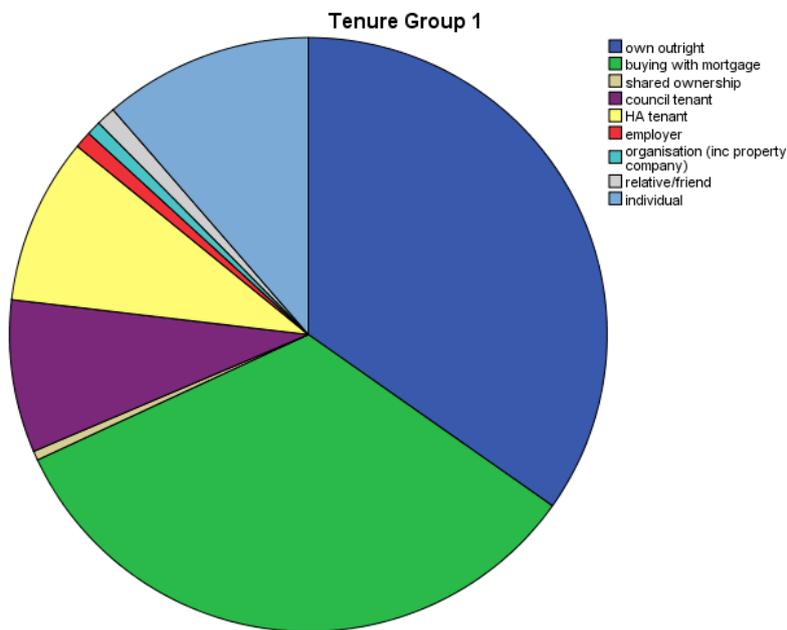
We are interested in comparing those on HB with those not on HB, so to compare like with like, we should restrict ourselves to only those who rent. The *tenure1* variable tells us whether households own outright, are buying with a mortgage or rent in various ways including the shared ownership category.

Let's look at *tenure1*. Do a frequency table to see how many fit into each category. You should get the following table:

Tenure Group 1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	own outright	6107	34.8	34.8	34.8
	buying with mortgage	5842	33.3	33.3	68.1
	shared ownership	88	.5	.5	68.6
	council tenant	1460	8.3	8.3	76.9
	HA tenant	1589	9.1	9.1	85.9
	employer	162	.9	.9	86.9
	organisation (inc property company)	136	.8	.8	87.6
	relative/friend	182	1.0	1.0	88.7
	individual	1990	11.3	11.3	100.0
	Total	17556	100.0	100.0	

You can see the proportions of different tenure types more easily at a glance by doing a pie chart – select *Pie chart* in the *Charts* button when you run the frequency table above.



See the section in the *Guide to SPSS 20* 'How to filter the data to select certain cases only' (p5-6): apply this to the dataset so that you are selecting only those who rent, i.e. if *tenure1*>2. Make sure you run the frequency table for *tenure1* again to check that the filter has worked!

? What kinds of households get housing benefit?

Household structures vary widely in ways that relate to the lifecycle, housing requirements and propensity to be able to earn. We might ask the question: Are single parents with dependent children more likely to receive HB than one-person households, say?

To look at who lives in households that get HB, we use the household type variable (*hhtype6*). Let's investigate this household type variable: *hhtype6*. It is a categorical variable, so create a frequency table to see how many people are in each category.

Household type - 6 categories

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid couple, no dependent child(ren)	1150	20.5	20.5	20.5
couple with dependent child(ren)	1051	18.7	18.7	39.3
lone parent with dependent child(ren)	908	16.2	16.2	55.4
other multi-person households	550	9.8	9.8	65.3
one person under 60	1010	18.0	18.0	83.3
one person aged 60 or over	938	16.7	16.7	100.0
Total	5607	100.0	100.0	

Can you answer the following questions for renting households:

How many households are there in total?

What is the largest kind of household group?

What proportion are lone parents with dependent child(ren)?

We can now look at whether households who receive HB are of different types than those who don't. **Create a cross-tabulation of *hhype6* by *housbenx* using column percentages (see p7 of the *Guide to SPSS 20*).**

Household type - 6 categories * Household (HRP + partner) receives any housing benefit? Crosstabulation

			Household (HRP + partner) receives any housing benefit?		Total
			Yes	No	
Household type - 6 categories	couple, no dependent child(ren)	Count	311	839	1150
		% within Household (HRP + partner) receives any housing benefit?	11.8%	28.3%	20.5%
	couple with dependent child(ren)	Count	350	701	1051
		% within Household (HRP + partner) receives any housing benefit?	13.2%	23.7%	18.7%
	lone parent with dependent child(ren)	Count	700	208	908
		% within Household (HRP + partner) receives any housing benefit?	26.5%	7.0%	16.2%
other multi-person households	Count	176	374	550	
	% within Household (HRP + partner) receives any housing benefit?	6.7%	12.6%	9.8%	
one person under 60	Count	450	560	1010	
	% within Household (HRP + partner) receives any housing benefit?	17.0%	18.9%	18.0%	
one person aged 60 or over	Count	658	280	938	
	% within Household (HRP + partner) receives any housing benefit?	24.9%	9.5%	16.7%	
Total	Count	2645	2962	5607	
	% within Household (HRP + partner) receives any housing benefit?	100.0%	100.0%	100.0%	

So the highest numbers of HB claimants are lone parents with dependent children (26.5%) and single households over 60 (24.9%). So the suggestion earlier on that households with children are more likely to get HB than single person households was correct, but not by much, and only for one-person households over 60.

Is it possible that the higher percentages we are seeing are simply due to those being the larger proportions of such households in the data? **To see if we come to different conclusions, remove the column percentages and select row percentages instead.**

Household type - 6 categories * Household (HRP + partner) receives any housing benefit? Crosstabulation

			Household (HRP + partner) receives any housing benefit?		Total
			Yes	No	
Household type - 6 categories	couple, no dependent child(ren)	Count	311	839	1150
		% within Household type - 6 categories	27.0%	73.0%	100.0%
	couple with dependent child(ren)	Count	350	701	1051
		% within Household type - 6 categories	33.3%	66.7%	100.0%
	lone parent with dependent child(ren)	Count	700	208	908
		% within Household type - 6 categories	77.1%	22.9%	100.0%
other multi-person households	Count	176	374	550	
	% within Household type - 6 categories	32.0%	68.0%	100.0%	
one person under 60	Count	450	560	1010	
	% within Household type - 6 categories	44.6%	55.4%	100.0%	
one person aged 60 or over	Count	658	280	938	
	% within Household type - 6 categories	70.1%	29.9%	100.0%	
Total	Count	2645	2962	5607	
	% within Household type - 6 categories	47.2%	52.8%	100.0%	



How much HB does each type of household get per week?

The Government are capping HB at £400 pounds a week for all households.

Guess what the average (i.e. mean) amount of weekly HB received is (among HB recipients)?

To find the true value for these data we must restrict the data to just those on HB (i.e. filter the data) using the same method as earlier to select only households that rent: select cases if *housbenx*=1. Do a frequency table of *housbenx* to check the filter has worked as you expect.

First of all look at the average amount of HB paid per week over all (*amthbenx*):

Use *Analyze > Descriptive Statistics > Descriptives*:

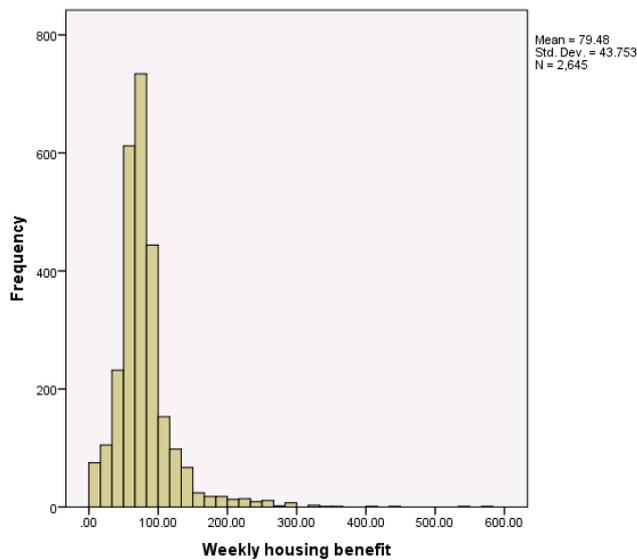
Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Weekly housing benefit	2645	.50	576.92	79.4813	43.75300
Valid N (listwise)	2645				

Among HB recipients, what is the average amount of weekly HB received?.....

What is the largest amount of weekly HB received by anyone in this survey?.....

See the distribution of the weekly HB received by requesting a histogram. You can do this by going into *Graphs> Legacy Dialogs> Histogram...* This should look as follows:



The histogram gives you the shape of the amount of weekly HB, indicating that there is a tendency for HB to peak around 60-80 pounds per week.

Note that the mean and the number of respondents on which the histogram was based are given at the side of the histogram.

We can use the Explore command to look at the relationship between the amount of HB and the type of household. You can find this under *Analyze> Descriptive Statistics> Explore* in the menus. Change the graphs to Histograms.

The following shows the start of the *Descriptive statistics* you get when you run *Explore*. This gives us a lot of information most of which we won't look at today – we will only look at the mean (average) values for each type of household.

Descriptives

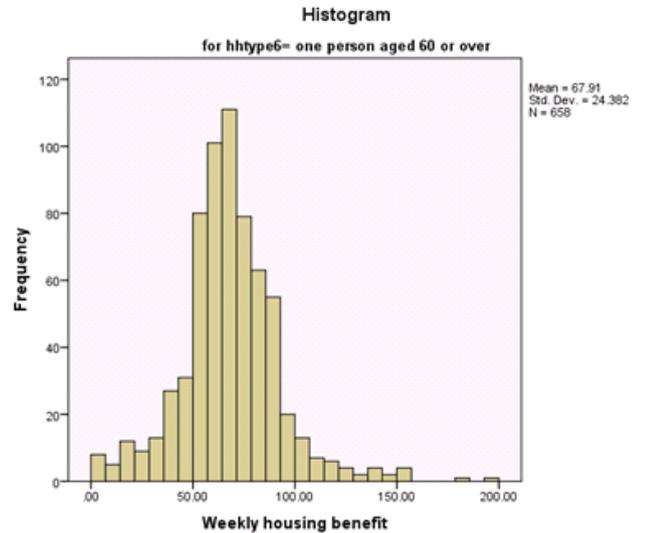
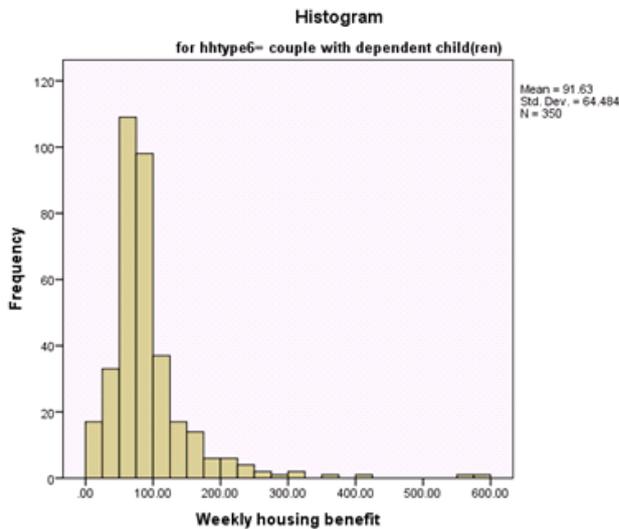
Household type - 6 categories		Statistic	Std. Error	
Weekly housing benefit	couple, no dependent child(ren)	Mean	70.1607	
		95% Confidence Interval for Mean	Lower Bound	66.3255
			Upper Bound	73.9959
		5% Trimmed Mean	68.0829	
		Median	69.0000	
		Variance	1181.538	
		Std. Deviation	34.37350	
		Minimum	1.00	

Look at your results. What is the mean weekly HB benefit for the following types of household:

Couple with dependent children?

One person household aged 60 or over?

Here are the histograms for these two household types:



What can you say about the way that HB is distributed in each case?

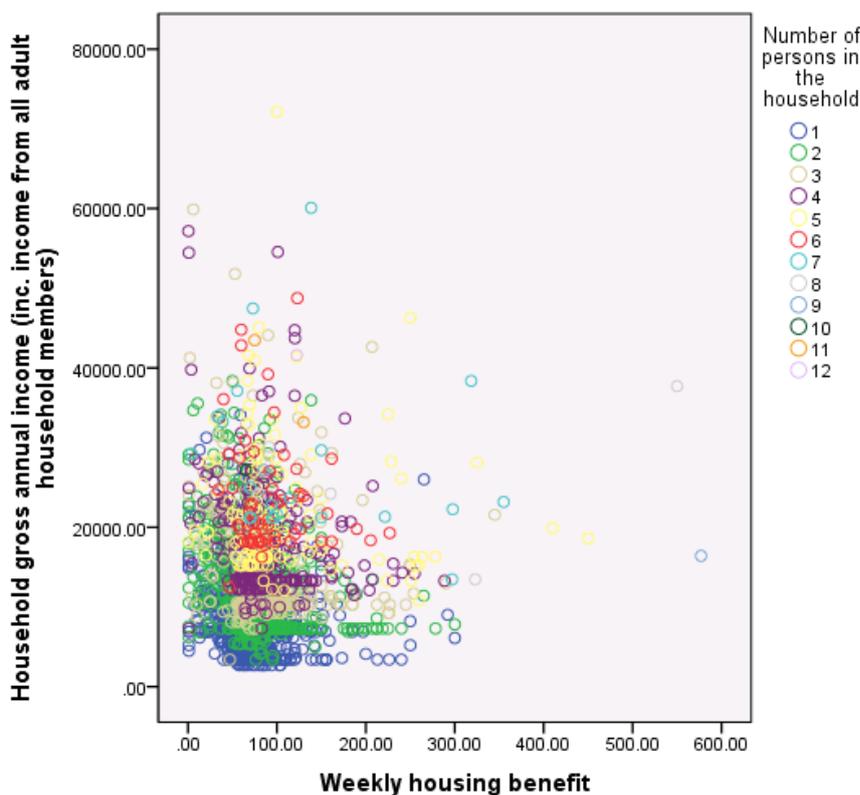
You have probably concluded that although they look a bit different, these two histograms actually tell a similar story about how HB is distributed for most respondents in these types of households. In both sets of households, most get somewhere between 50 and 100 pounds a week. However, among couple with children households, there are quite a few who get over 150 pounds per week. In some cases, they get very large amounts per week – approximately 550-600 pounds per week! These households are likely to be among those affected by the HB cap therefore.

Thinking about the distribution of amounts received, do you think that the mean value of benefit received per week is a good measure of what is typical? Is it a better measure for some groups than others? What other measures might you explore as way of expressing some sort of average?



What is the relationship between amount of HB and household income?

Now if we want to look at the relationship between the amount of HB received and another continuous variable such as the household income (*hyeargrx*) among HB recipients, we can create a scatterplot (use *Graph > Legacy dialogs > Scatter/Dot...*). Note that there is an option to make the dots on the graph different colours for different groups within the data. For example, it might be interesting to show different markers for the different sizes of household (*hhsizex*).



Now save your syntax file to the desktop as housing1.sps (use *File> Save as...* in the Syntax Editor). You will continue to add to this file today and can keep saving it again under the same name as you go along. You will need it again for the last session so make sure you don't delete it.

Group exercise:

Are households with younger heads of household more likely to get HB, or to get larger amounts of HB than households with older heads?

In your groups, examine the relationship between age of the head of household and receipt of HB and the amount of HB received. The head of household is called the HRP, or Household Representative Person. Age is presented in the dataset as both a continuous variable (*agehrpx*) or a grouped variable (*agehrp6x*). Because one is continuous/scalar, and the other is a categorical variable (now that it has been grouped), the techniques you use to examine this relationship will vary.

In groups:

- discuss the possible ways to look at the relationship between HB (*housbenx* and *amtbenx*) and age of head of household (*agehrpx* and *agehrp6x*) (Think about the level of measurement of each variable: What kind of variable is *agehrp6x*?)
- divide up the tasks among your group and try them out
- look at what your group has produced. What do you think works best?

There is no right or wrong answer to whether you should use age as a continuous or as a grouped variable – what works best depends entirely on the particular variable and what you want to show with your analyses or graphs.

Session 3: Working with data files

Simple variable derivation

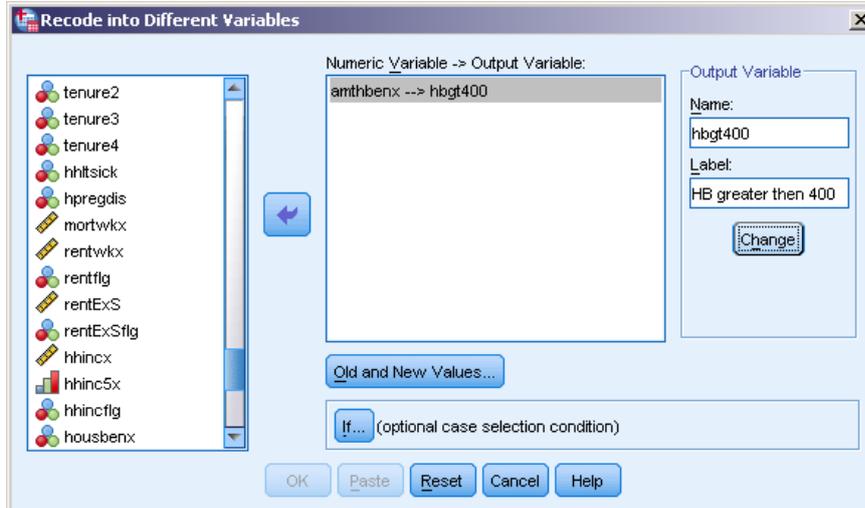
Secondary analysts typically have to work with the data to get it into the right format for their research. Often this is simply a case of grouping values or combining information from variables in a single file.

Recoding a variable

In order to identify the proportion who obtain over £400 pw in HB, one simple approach is to create a variable which indicates whether the responding household receives more than £400 or not.

We always recode into a different variable so that we can check the new variable against the old variable to ensure that it has worked the way we would have hoped.

Use the command `Transform > Recode into different variable` to open the recode dialogue box shown below:



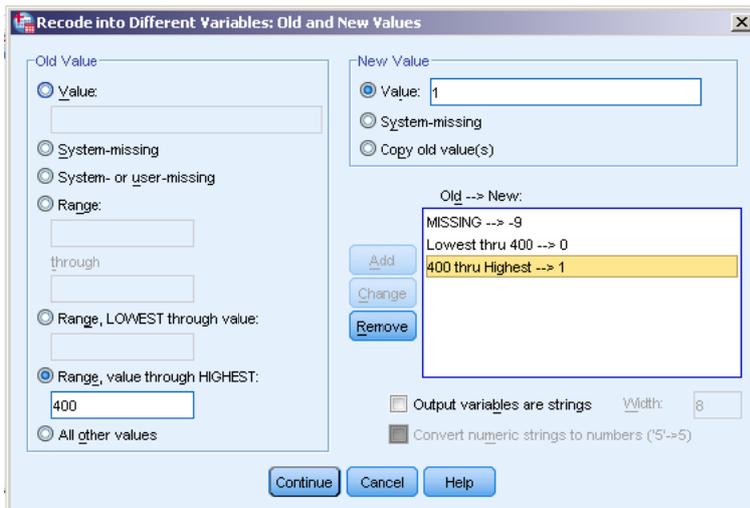
Click **C**hange to apply the new name (hbgt400) and label (HB greater than 400) to the output variable. Then use **O**ld and New Values to define the values.

In Old Value click on Range lowest through value to group those who receive those £400 pw or less. In New Value add a new value of 0 (0 is a good choice of value where we want a value to mean no). Click Add to store this.

Repeat this process, this time using the Range value through highest option in the Old Value panel. It is OK to use £400 as the boundary again because any cases which receive exactly £400pw will be reassigned by the previous recode statement. In the new value panel assign the value to 1 (which again, is a good choice to indicate 'yes').

Finally, set all missing values to -9. Select 'System missing or User missing' in the Old Value panel and -9 to the new value field.

Once all three values are in the Old→New: field click Continue.

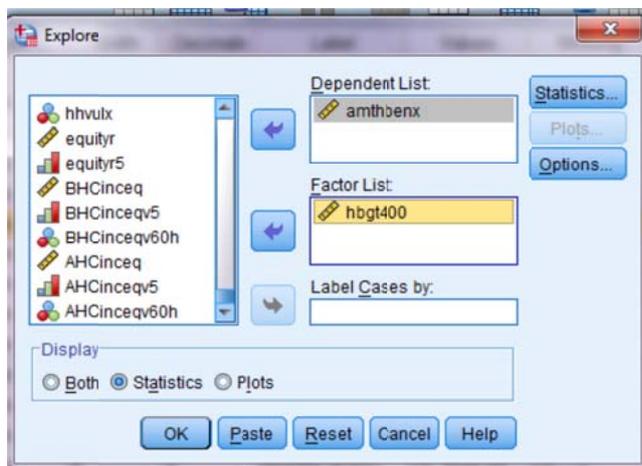


Click Paste then run as usual to execute the transformation.

The easiest way to check that this has worked as expected is to use the Explore procedure

You can find this procedure by selecting:

Analyze > Descriptive Statistics > Explore.



The dependent variable is the ungrouped variable *amthbenx* as this is a scalar variable.

The factor list is the value of the new variable as this is in categories. **We will not need to see plots, so we can limit the output to statistics in the Display area. Paste then run in the usual way.**

When considering the resulting descriptive statistics view the minimum and maximum value of *amthbenx* for each value of *hbgt400*.

- Is the minimum value of *amthbenx* between 0 and 400 where *hbgt400* = 0?
.....
- Is the minimum value of *amthbenx* over 400 where *hbgt400* = 1?
.....
- Does it make sense that there are no valid values of *amthbenx* for cases with the value of -9 for *hbgt400*?
.....

If the answer to all of these is yes, the evidence suggests that your recode worked as expected.



What percentage of those receiving housing benefit would be affected by a cap of housing benefit at £400?

How to link files

Do people claiming Housing Benefit live in poorer areas than renters not claiming HB?

Some datasets are provided in multiple files. The English Housing Survey is an example of this sort of data and you will normally need to link files in order to get all of the variables you need for an analysis.

At the start of this workbook the variables to be used were listed. They come from two different files, but up until now we have only looked at those variables that come from the *interviewfs10* file. From this point on we will also use variables that are drawn from the *generalfs10* file.

We will be adding a number of variables from the general file to the interview file. This is because the general file contains information about the area which we need for our analysis and because the weight variable that we will be using in Session 4 are in this file but not in the interview file.

A good way of approaching this question would be to use the IMD deciles and to produce a two-way table of IMD decile by whether or not the household claims Housing Benefit.

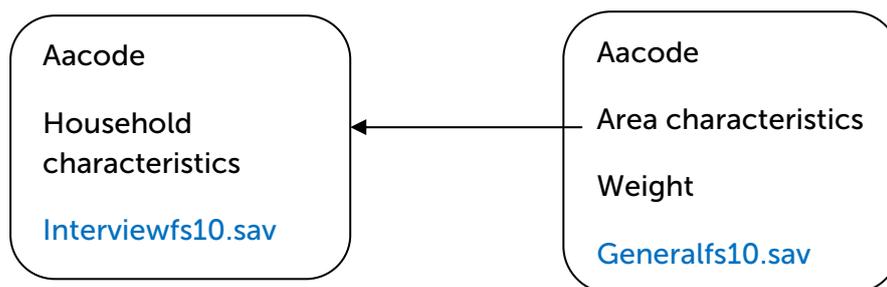
The Index of Multiple Deprivation (IMD) is a country-specific official measure of deprivation across a number of dimensions at the local (super output area) level. So for example it combines information on under-education, poor health and low incomes. Separate measures are available for each member country of the UK, so they are not comparable across countries. More information about IMD is available at <https://www.gov.uk/government/publications/english-indices-of-deprivation-2010>.

When you download the EHS 2010-2011 household files from UK Data Service, you will see that the data are in two folders called *Interview* and *Derived*. *Interview* contains a number of files based on the raw data from the interviews. The data you are more likely to want to use is the derived data in the *Derived* folder. If you were to download the data from the UK Data Service and open the *Derived folder*, you will find two files with names:

- *Generalfs10.sav* contains information about household tenure, the Government Office Region, deprivation scales and the household weight.
- *Interviewfs10.sav* contains results of the household interviews with information such as household composition, age, gender and employment status of the HRP and partner, the number of bedrooms etc.

To use the interview data with the weight, you must combine the two files. The files are at the same level (both household) so this is relatively easily done. You need a matching variable to link the two files together. The matching variable is common to both files and it acts as a key to make sure that the right information for the household is combined. The matching variable is the household identifier, which is called *aacode* in both files.

For cases where *aacode* in *interview.sav* = *aacode* in *general.sav*



We will add to the individual file some variables from the general file for the case with the same value of *aacode*.

To combine the two files

Data > Merge Files > Add Variables

The following uses the drop-down menus but if you are more familiar with SPSS, you can use syntax. The syntax file to link the files is shown at the end of this section.

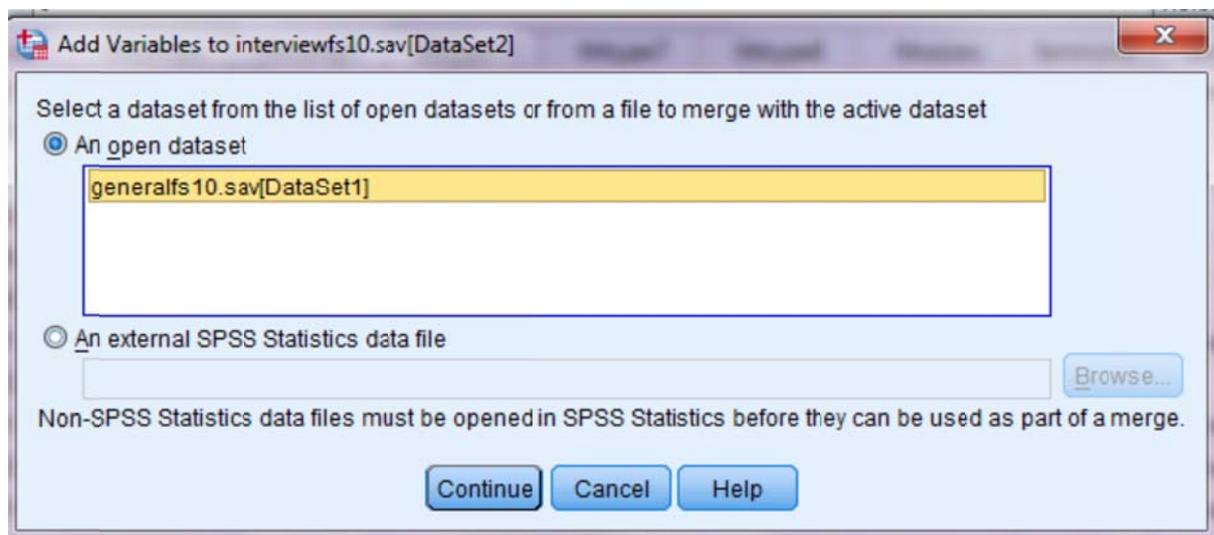
The matching variable is the household identifier *aacode*. Before merging data sets, the matching variable must be ordered in the same way in both data sets. The data in these EHS files are sorted by *aacode* already so there is no need to sort the files.

Start by opening *generalfs10.sav*

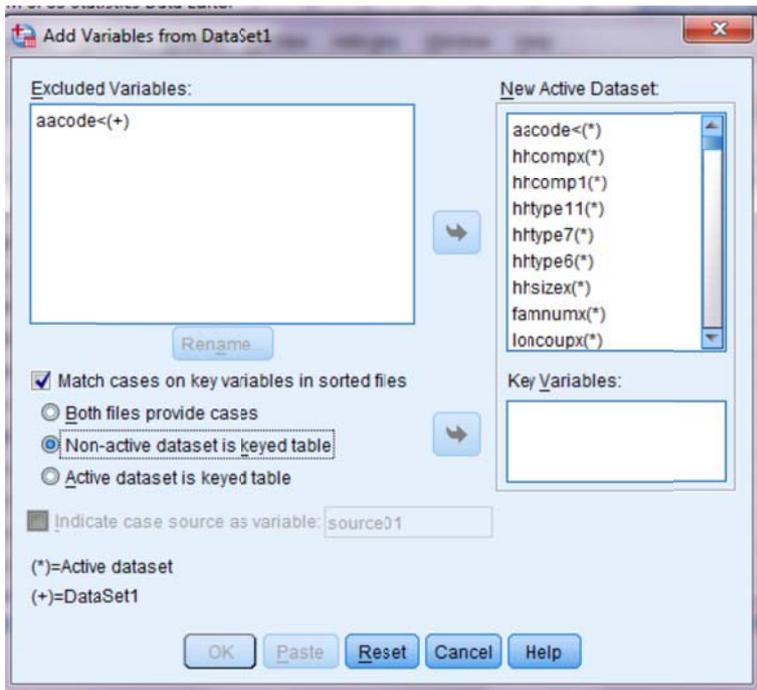
Order the file by clicking *Data > Sort Cases*. Select *aacode* as your ordering variable, click *Paste*. Save the file as its current name.

Return to *interviewfs10.sav* and repeat this procedure so that we can be certain that each file is sorted by *aacode*, save it.

To merge the *interviewfs10.sav* data with this file, ensure that you are in the *interviewfs10.sav* window and use the menu: *Data > Merge Files > Add Variables* to get to the following screen:

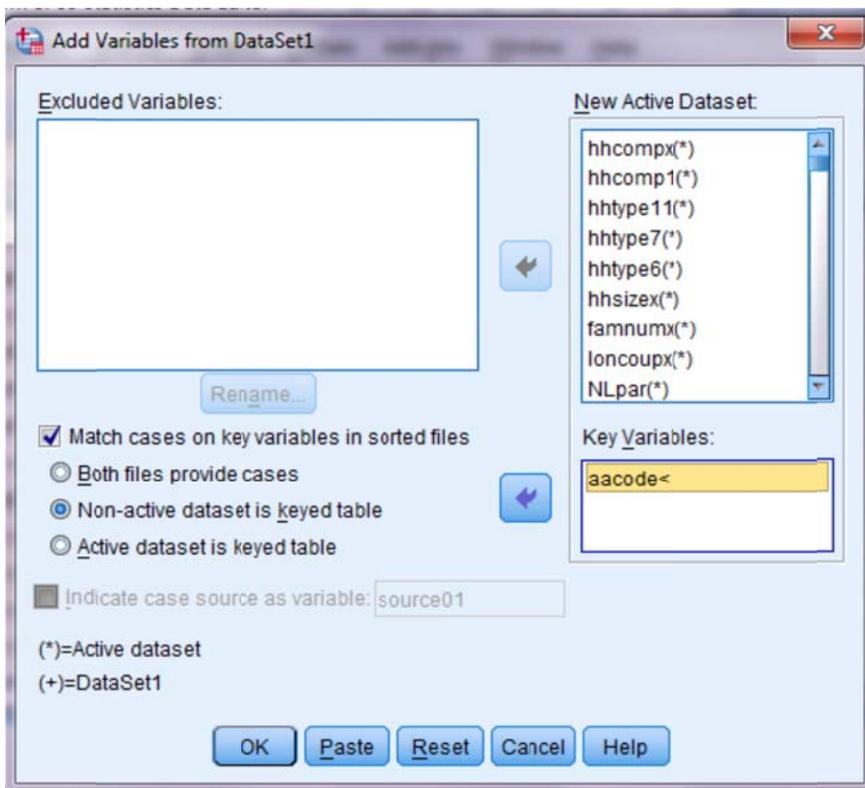


Select *generalfs10.sav* as the dataset to draw variables from and click Continue.



In the new window, tick *Match cases on key variables in sorted files*, and *Non active dataset is a keyed table* (in other words, we are going to use the current Interview file as our file, then we're going to treat the general file as the look up table we're pulling in variables from).

Then click on *aacode* in the left-hand window and click on the arrow to move it to the *Key Variable* box as shown below:



Then press Paste to save. Then select the syntax and run this to merge the files.

The new merged data set is now displayed: you should see that there are variables in this file which were not present before, most notably for our purposes imd1010 and afgh10 are present. **Save the new dataset as EHS10all.sav.**

We can now proceed to answer our research question.



Are people on housing benefit more likely to live in deprived areas?

A two-way table will give us an indication of whether this is so.

Request a crosstab which you think would be in the best format to answer the question.

Did yours look anything like the one below?

**Renters & Shared Ownership tenant households
Household Benefit receipt by HRP and partner by IMD decile**

Deprivation - IMD2010 decile ranking of areas (lower SOAs) * Household (HRP + partner) receives any housing benefit? Crosstabulation

% within Deprivation - IMD2010 decile ranking of areas (lower SOAs)

		Household (HRP + partner) receives any housing benefit?		Total
		Yes	No	
Deprivation - IMD2010 decile ranking of areas (lower SOAs)	most deprived 10% of areas	60.2%	39.8%	100.0%
	2nd	55.5%	44.5%	100.0%
	3rd	51.9%	48.1%	100.0%
	4th	46.5%	53.5%	100.0%
	5th	44.0%	56.0%	100.0%
	6th	40.4%	59.6%	100.0%
	7th	37.7%	62.3%	100.0%
	8th	38.1%	61.9%	100.0%
	9th	36.4%	63.6%	100.0%
	least deprived 10% of areas	33.8%	66.2%	100.0%
Total	47.2%	52.8%	100.0%	

At first glance it would appear that there is a clear relationship between housing benefit receipt levels and type of area. The more deprived an area the respondent lives in, the more likely they are to be claiming HB.

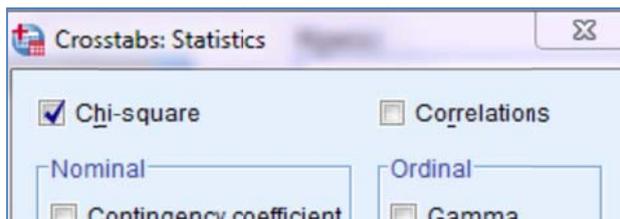
If there were no relationship at all between the two variables, we would see the same percentages (47% and 53%) in each row.

Apparent relationships can be due to small numbers. The smaller the sample size the less accurate our estimates are.

If we drill down into a dataset either explicitly by looking at sub-populations using a filter variable, or because (as is the case here) we are using a variable which only applies to some of the respondents, we may end up with much smaller numbers of respondents in our data than in the survey as a whole.

One way of handling this possible problem is to obtain a chi-squared statistic. This indicates whether variations away from the pattern we would expect if there were no relationship between our row and column variables is likely to be due to sampling variability alone.

You can request the chi-squared statistic to be produced as part of the crosstabs procedure. Simply hit the Statistics button when you are defining the two-way table and request chi-squared.



	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	171.587 ^a	9	.000
Likelihood Ratio	172.923	9	.000
Linear-by-Linear Association	162.898	1	.000
N of Valid Cases	5607		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 153.31.

If there were no relationship between the variables, the probability of a Pearson Chi-Squared statistic as large as this, with a table this size and a sample size of 5,607 is <0.001 (which is what SPSS means when it says .000 – this probability, no matter how small is never =0).

Task if you have time: Can you repeat this procedure for different regions?

GorEHS will allow you to distinguish between regions.

You can produce tables separately for each region by adding GorEHS as a layer variable to the crosstabs procedure.

Session 4: Weighting

So far, we have learned who gets HB, whether it is linked to deprivation or the geographical area in which you live. However, everything we have done so far applies only to the individuals who responded to the survey unless we apply a weight to the data and then run the analyses. First, we go back to the EHS documentation to see what it says about weights in the EHS.

Look at the documentation for the English Housing Survey, household data, 2010-2011 on the UK Data Service website. You can search pdf documents using Control-F. Search for "weight".

Do you see anything about weights when you search the questionnaire?

Is this what you'd expect?

What is the name of the weight in the 2010-11 dataset, according to the user guide?

.....

For more detail about the weights, you can look for the technical documents related to weighting. These documents are normally supplied to the UK Data Service with the data, but if not, you can normally find this information by searching using an internet search engine (e.g. search for "English housing survey 2010-2011 technical documents" in Google).

Go to the dataset. Have a look at the values of the weighting variable. You can look either in the Data View window or you can use Descriptives to summarise this variable. For the weight, what is its:

- **minimum?**
- **maximum?**
- **mean?**

This is the first table you created at the start of the practical session.

Household (HRP + partner) receives any housing benefit?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	2645	15.1	47.2	47.2
	No	2962	16.9	52.8	100.0
	Total	5607	31.9	100.0	
Missing	Not applicable - owner occupier	11949	68.1		
Total		17556	100.0		

Answer the questions (some you answered earlier):

How many households were included in this survey?.....

What proportion of households rent (i.e. are not owner occupiers)? **31.9%**

What proportion of renters receive HB? **47.2%**

Now apply the weight to the data using *Data > Weight cases...* . In SPSS, once you give the command to weight the data, all subsequent analyses will be weighted unless you remove the weight.

Then re-run the analysis to get the following table:

Household (HRP + partner) receives any housing benefit?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	3295792	15.1	43.5	43.5
	No	4280963	19.6	56.5	100.0
	Total	7576754	34.6	100.0	
Missing	Not applicable - owner occupier	14315925	65.4		
Total		21892679	100.0		

Now answer the following:

How many households are there in England? (round to the nearest thousand – the weighted data aren't accurate to the individual household!).....

What proportion of households rent (i.e. are not owner occupiers)?

What proportion of renters receive HB?

Note the difference between the results. Sometimes weighting make a little difference to the estimates but sometimes it makes a very big difference. Always quote weighted results (though see the section below). Note that some surveys have weights that don't gross up to the population size. You still use them in the same way, but you can't then see estimates of the numbers in the population so easily.

To weight or not to weight?

Generally, always use the weighted results when you present results of analyses. You will probably want to note the raw numbers in the sample so that readers can gauge how reliable your results are. Large numbers of respondents are better than small numbers i.e. a survey of 200 gives much less reliable results than one with 20,000 respondents.

However, when you first start looking at your data, you may want to do so with the data unweighted. This is because this allows you to see how many actual responses there were to a particular question or a combination of questions. This is important because when you weight the data using a weight that grosses up to the population size, the numbers will look large. But would you base your results on the answer or just 1 or 2 people?

For example, would you be confident that if there were only 2 people of a particular ethnic minority, say, in your survey and 1 lives in a bungalow and the other in a semi-detached house, that in the wider population 50% of people of that ethnic group live in bungalows and the other 50% live in semi-detached houses?

Rerunning your syntax

You should have been pressing *Paste* (not *OK*) to run your commands today. This should have created a syntax file with all the commands you have used. This provides a record of what you have done. But more importantly, by running this syntax file, you can re-run the commands to redo all the analyses in a matter of seconds.

This is important because when we started today we didn't weight the data. If we want the results to be nationally representative (which we do), we must weight the data.

You are going to re-run all your analyses again using syntax and with the weight on – this will take about 10 seconds to run!

You should now have the weight on. If you haven't, then apply it now (*Data> Weight cases...*).

Now re-run the whole syntax file you have created from your analyses today by selecting all the commands (Control-A) in the syntax editor window and then pressing the arrow button .

Congratulations! – you have just created results that you could publish because they are weighted. You can now make statements about receipt of HB for the population of England. If you want to, you can save your results by using *File> Save as...*

If there is time: In groups:

Other research questions

You came up with some questions about HB and other benefits in the UK at the start of the day. **Have a look at some of them and discuss how you would go about answering them using these or other datasets.**

- Which variables do you need?
- How would you find a relevant dataset?
- What kinds of analyses would you use?

Choose a research question from the morning and start investigating it. You can't necessarily answer all research questions and some may be answerable but very complicated. Restrict yourselves to a straightforward question if you want to make progress in limited time!

Note that you don't need to decide everything about your research question in advance. You'll need a topic and interesting question but you can decide whether it applies to England only, say, or to the UK or England and Wales etc., or to all adults or an ethnic group only or other subset of the population based on what data are available. If the English Housing Survey is the only or the best survey to use, then you'll probably decide to restrict yourself to England only!