PART TWO: 1 August 2020

Teaching resource: Analysing ethnic differences in health using data from Understanding Society

CITATION FOR THE TEACHING DATASET

University of Essex. Institute for Social and Economic Research. (2020). Understanding Society: Ethnicity and Health Teaching Dataset Wave 1, 2009-2010. [data collection]. UK Data Service. SN: 8465, http://doi.org/10.5255/UKDA-SN-8465-2

CITATION FOR THE TEACHING RESOURCE:

Nandi, Alita and Wiltshire, Deborah. (2020). "Teaching Resource: Analysing ethnic differences in health using data from Understanding Society"

AUTHORS

Alita Nandi

University of Essex

Deborah Wiltshire

University of Essex

Understanding Society

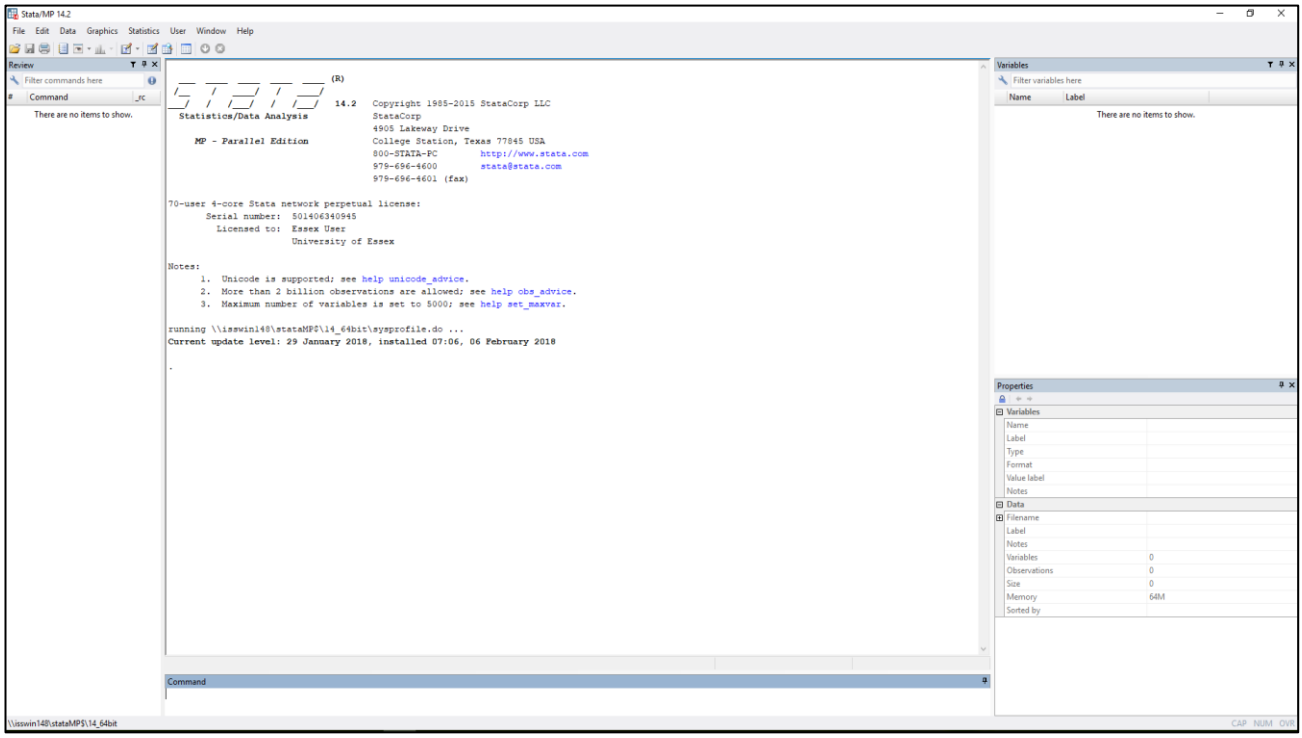
UK Data Service

PART TWO: Data exercise

We want to find out whether and how far health varies across ethnic groups. Are some more healthy than others? Our focus is therefore on two variables: ethnicity and health. Once we know there is such variation we could, at a later stage, add far more variables to try to see what factors are associated with this variation, but below we demonstrate only the first stage.

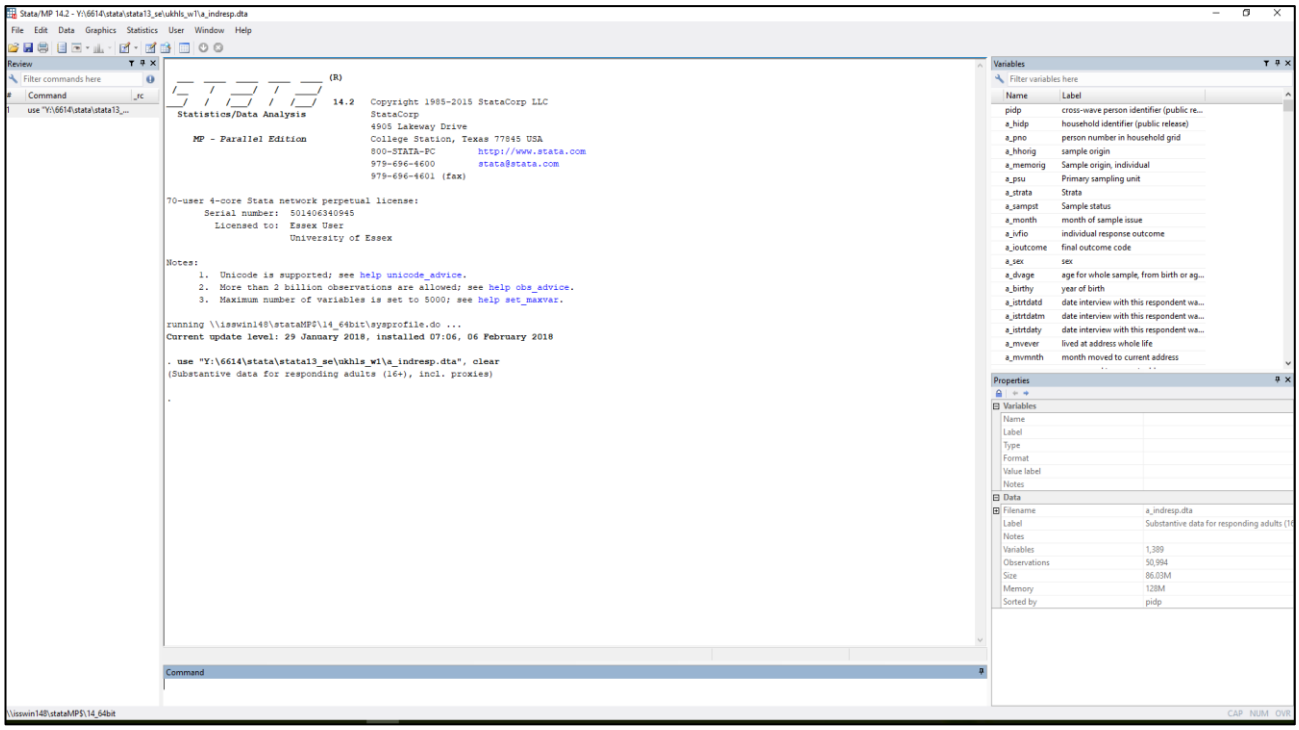
In this part of the exercise you will open the data file using Stata software and then produce frequency tables and graphs.

- First open Stata



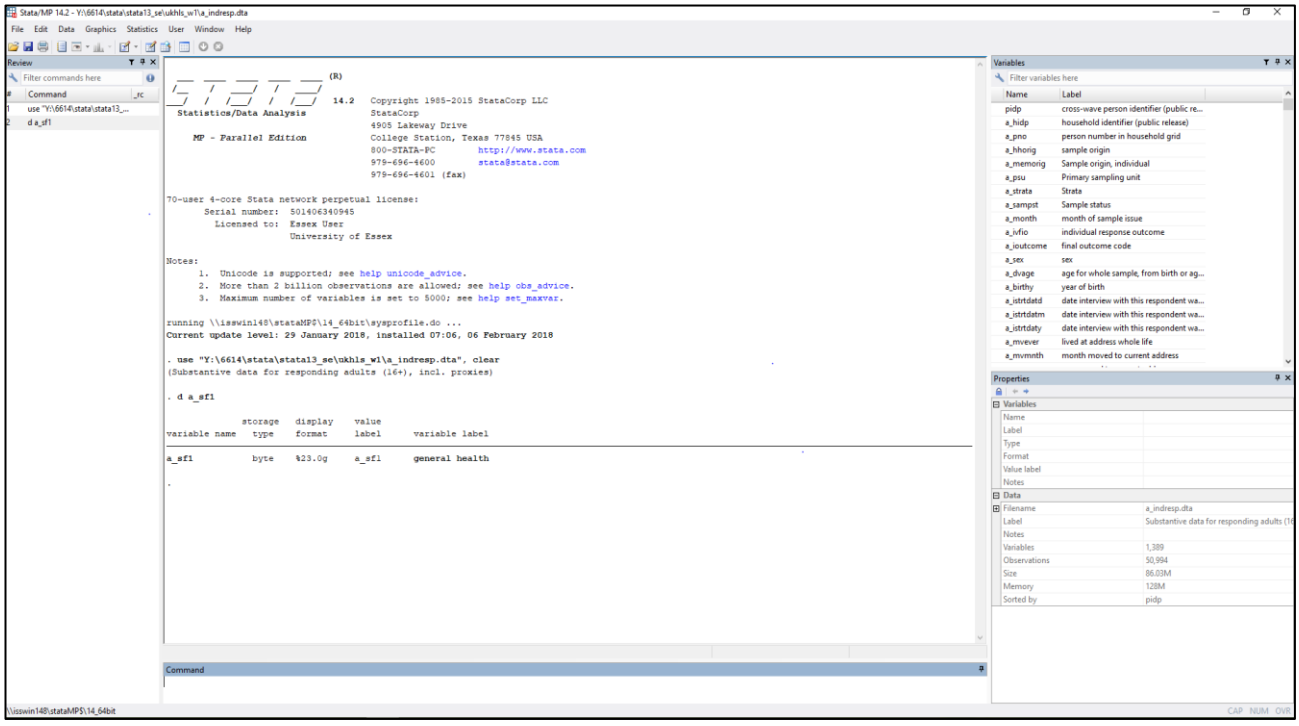
- Then open the data file called **a_indresp_td.dta** from the location where you have downloaded this file. This is a truncated version of the original file, **a_indresp.dta**. You can either open it using the OPEN icon on the top menu bar or you can type the following in the command box:

```
use a_indresp_td, clear
```



- Now look for the variables discussed in Part One. If you just want to see whether the variable exists in the file, or see the variable label, type the following in the “Command” box

```
describe a_sf1
```



Stata is case sensitive. If you type,

```
describe A_sf1
```

Stata will say no such variable exists!

```
. des A_sf1
variable A_sf1 not found
r(111);
```

- You can produce the frequency table you saw online by typing,

```
tabulate a_sf1
```

```
. tabulate a_sf1

     general health |      Freq.     Percent        Cum.
--------------------+-----------------------------------
            missing |         63        0.12        0.12
            refusal |         18        0.04        0.16
         don't know |         13        0.03        0.18
          excellent |      9,583       18.79       18.98
          very good |     16,285       31.94       50.91
               good |     14,035       27.52       78.43
               fair |      7,309       14.33       92.77
               poor |      3,688        7.23      100.00
--------------------+-----------------------------------
              Total |     50,994      100.00
```

Does that mean that 18.79% of the sample has reported that they have excellent general health? No, because Stata considers the missing values (-1 -2 -9) as valid values. So, the first thing you do is tell Stata that this variable will be ignored for those people with one of these three negative values. This is done by recoding to 'system missing' (denoted by a full stop):

```
recode a_sf1 -9/-1=.
```

Now type,

```
tabulate a_sf1
```

```
. tabulate a_sf1
```

general health	Freq.	Percent	Cum.
excellent	9,583	18.83	18.83
very good	16,285	31.99	50.82
good	14,035	27.57	78.39
fair	7,309	14.36	92.75
poor	3,688	7.25	100.00
Total	50,900	100.00	

And now you can see 18.83% of the sample reported that they have excellent general health.

- What is the ethnic composition of the sample?

```
recode a_racel_dv -9/-1=.

tabulate a_racel_dv
```

Ethnic group incorp. all waves, codings, modes and bhps	Freq.	Percent	Cum.
british/english/scottish/welsh/northern	36,920	75.18	75.18
irish (white)	732	1.49	76.67
gypsy or irish traveller (white)	18	0.04	76.70
any other white background (white)	1,503	3.06	79.76
white and black caribbean (mixed)	356	0.72	80.49
white and black african (mixed)	158	0.32	80.81
white and asian (mixed)	203	0.41	81.22
any other mixed background (mixed)	188	0.38	81.61
indian (asian or asian british)	1,984	4.04	85.65
pakistani (asian or asian british)	1,538	3.13	88.78
bangladeshi (asian or asian british)	1,185	2.41	91.19
chinese (asian or asian british)	331	0.67	91.87
any other asian background (asian or as	653	1.33	93.20
caribbean (black or black british)	1,183	2.41	95.60
african (black or black britih)	1,501	3.06	98.66
any other black background (black or bl	91	0.19	98.85
arab (other ethnic group)	291	0.59	99.44
any other ethnic group (other ethnic gr	276	0.56	100.00
Total	49,111	100.00	

- What proportion of Irish people report excellent health and how does that compare with Caribbean people? To see this we need a cross-tabulation of two variables, ethnicity and health. Type

```
tabulate a_racel_dv a_sf1
```

Ethnic group incorp. all waves, codings, modes and bhps	General health excellent	very good	good	fair	or Poor?	Total
british/english/scott	6,467	11,787	10,331	5,579	2,721	36,885
irish (white)	157	214	180	111	69	731
gypsy or irish travel	2	1	5	4	6	18
any other white backg	335	572	418	128	49	1,502
white and black carib	56	116	105	55	24	356
white and black afric	42	41	51	13	11	158
white and asian (mixe	44	69	52	31	6	202
any other mixed backg	43	65	50	19	11	188
indian (asian or asia	410	638	566	247	118	1,979
pakistani (asian or a	276	450	432	211	168	1,537
bangladeshi (asian or	223	356	359	138	108	1,184
chinese (asian or asi	66	127	94	35	8	330
any other asian backg	139	232	191	63	27	652
caribbean (black or b	193	339	300	254	95	1,181
african (black or bla	464	497	332	146	59	1,498
any other black backg	16	28	29	11	7	91
arab (other ethnic gr	54	89	104	29	13	289
any other ethnic grou	57	83	73	30	30	273
Total	9,044	15,704	13,672	7,104	3,530	49,054

157 out of 731 Irish people say that they have excellent health, while 193 out of 1,181 Caribbean people say the same.

- What does it mean in percentage terms? Instead of computing by hand, let Stata do that. Type

```
tabulate a_racel_dv a_sf1, row
```

Key

frequency
row percentage

Ethnic group incorp. all waves, codings, modes and bhps	General health excellent	very good	good	fair	or Poor?	Total
british/english/scott	6,467 17.53	11,787 31.96	10,331 28.01	5,579 15.13	2,721 7.38	36,885 100.00
irish (white)	157 21.48	214 29.27	180 24.62	111 15.18	69 9.44	731 100.00
gypsy or irish travel	2 11.11	1 5.56	5 27.78	4 22.22	6 33.33	18 100.00
any other white backg	335 22.30	572 38.08	418 27.83	128 8.52	49 3.26	1,502 100.00
white and black carib	56 15.73	116 32.58	105 29.49	55 15.45	24 6.74	356 100.00
white and black afric	42 26.58	41 25.95	51 32.28	13 8.23	11 6.96	158 100.00
white and asian (mixe	44 21.78	69 34.16	52 25.74	31 15.35	6 2.97	202 100.00
any other mixed backg	43 22.87	65 34.57	50 26.60	19 10.11	11 5.85	188 100.00
indian (asian or asia	410 20.72	638 32.24	566 28.60	247 12.48	118 5.96	1,979 100.00
pakistani (asian or a	276 17.96	450 29.28	432 28.11	211 13.73	168 10.93	1,537 100.00
bangladeshi (asian or	223 18.83	356 30.07	359 30.32	138 11.66	108 9.12	1,184 100.00
chinese (asian or asi	66 20.00	127 38.48	94 28.48	35 10.61	8 2.42	330 100.00
any other asian backg	139 21.32	232 35.58	191 29.29	63 9.66	27 4.14	652 100.00
caribbean (black or b	193 16.34	339 28.70	300 25.40	254 21.51	95 8.04	1,181 100.00
african (black or bla	464 30.97	497 33.18	332 22.16	146 9.75	59 3.94	1,498 100.00
any other black backg	16 17.58	28 30.77	29 31.87	11 12.09	7 7.69	91 100.00
arab (other ethnic gr	54 18.69	89 30.80	104 35.99	29 10.03	13 4.50	289 100.00
any other ethnic grou	57 20.88	83 30.40	73 26.74	30 10.99	30 10.99	273 100.00
Total	9,044 18.44	15,704 32.01	13,672 27.87	7,104 14.48	3,530 7.20	49,054 100.00

Now you can see that 16.34% of Caribbean people say they have excellent health as compared to 21.48% of Irish people.

- But ethnic groups are not homogeneous. Will you find the same difference if you look at men and women?

  For women, type

```
tabulate a_racel_dv a_sf1 if a_sex_dv==2, row
```

Key
frequency *row percentage*

Ethnic group incorp. all waves, codings, modes and bhps	General health					
	excellent	very good	good	fair	or Poor?	Total
british/english/scott	3,536 17.21	6,601 32.13	5,773 28.10	3,089 15.04	1,546 7.52	20,545 100.00
irish (white)	79 19.70	122 30.42	103 25.69	64 15.96	33 8.23	401 100.00
gypsy or irish travel	1 8.33	0 0.00	3 25.00	4 33.33	4 33.33	12 100.00
any other white backg	181 20.59	349 39.70	233 26.51	85 9.67	31 3.53	879 100.00
white and black carib	27 12.33	65 29.68	73 33.33	36 16.44	18 8.22	219 100.00
white and black afric	22 25.58	21 24.42	29 33.72	6 6.98	8 9.30	86 100.00
white and asian (mixe	27 23.48	39 33.91	27 23.48	18 15.65	4 3.48	115 100.00
any other mixed backg	25 20.83	43 35.83	33 27.50	13 10.83	6 5.00	120 100.00
indian (asian or asia	167 17.36	293 30.46	289 30.04	142 14.76	71 7.38	962 100.00
pakistani (asian or a	113 14.07	213 26.53	241 30.01	141 17.56	95 11.83	803 100.00
bangladeshi (asian or	84 14.48	178 30.69	194 33.45	71 12.24	53 9.14	580 100.00
chinese (asian or asi	32 19.63	64 39.26	42 25.77	20 12.27	5 3.07	163 100.00
any other asian backg	77 21.10	119 32.60	119 32.60	35 9.59	15 4.11	365 100.00
caribbean (black or b	109 15.53	193 27.49	172 24.50	167 23.79	61 8.69	702 100.00
african (black or bla	219 25.58	298 34.81	202 23.60	99 11.57	38 4.44	856 100.00
any other black backg	11 19.64	18 32.14	16 28.57	6 10.71	5 8.93	56 100.00
arab (other ethnic gr	24 18.60	46 35.66	42 32.56	13 10.08	4 3.10	129 100.00
any other ethnic grou	25 17.24	49 33.79	37 25.52	16 11.03	18 12.41	145 100.00
Total	4,759 17.54	8,711 32.10	7,628 28.11	4,025 14.83	2,015 7.43	27,138 100.00

- For men, type

```
tabulate a_racel_dv a_sf1 if a_sex_dv==1, row
```

Key
frequency *row percentage*

Ethnic group incorp. all waves, codings, modes and bhps	General health excellent	very good	good	fair	or Poor?	Total
british/english/scott	2,931 17.94	5,186 31.74	4,557 27.89	2,490 15.24	1,175 7.19	16,339 100.00
irish (white)	78 23.64	92 27.88	77 23.33	47 14.24	36 10.91	330 100.00
gypsy or irish travel	1 16.67	1 16.67	2 33.33	0 0.00	2 33.33	6 100.00
any other white backg	154 24.72	223 35.79	185 29.70	43 6.90	18 2.89	623 100.00
white and black carib	29 21.17	51 37.23	32 23.36	19 13.87	6 4.38	137 100.00
white and black afric	20 27.78	20 27.78	22 30.56	7 9.72	3 4.17	72 100.00
white and asian (mixe	17 19.54	30 34.48	25 28.74	13 14.94	2 2.30	87 100.00
any other mixed backg	18 26.47	22 32.35	17 25.00	6 8.82	5 7.35	68 100.00
indian (asian or asia	243 23.89	345 33.92	277 27.24	105 10.32	47 4.62	1,017 100.00
pakistani (asian or a	163 22.21	237 32.29	191 26.02	70 9.54	73 9.95	734 100.00
bangladeshi (asian or	139 23.01	178 29.47	165 27.32	67 11.09	55 9.11	604 100.00
chinese (asian or asi	34 20.36	63 37.72	52 31.14	15 8.98	3 1.80	167 100.00
any other asian backg	62 21.60	113 39.37	72 25.09	28 9.76	12 4.18	287 100.00
caribbean (black or b	84 17.57	145 30.33	128 26.78	87 18.20	34 7.11	478 100.00
african (black or bla	245 38.16	199 31.00	130 20.25	47 7.32	21 3.27	642 100.00
any other black backg	5 14.29	10 28.57	13 37.14	5 14.29	2 5.71	35 100.00
arab (other ethnic gr	30 18.75	43 26.88	62 38.75	16 10.00	9 5.63	160 100.00
any other ethnic grou	32 25.00	34 26.56	36 28.13	14 10.94	12 9.38	128 100.00
Total	4,285 19.55	6,992 31.91	6,043 27.58	3,079 14.05	1,515 6.91	21,914 100.00

The same pattern shows up for men and women. You can see that 17.57% of Caribbean men report excellent health compared to 23.64% of Irish men. And 15.53% of Caribbean women report excellent health compared to 19.7% of Irish women.

Another interesting pattern becomes evident. In each of these groups, a lower percentage of women say they have excellent health as compared to men: 15.53% Vs 17.57% (Caribbean) and 19.7% Vs 23.64% (Irish).

Exercise 3: Repeat cross-tabulations for general health by marital status and country of birth separately.

Exercise 4: If the variable **a_ivfio** records the type of interview – full interview or proxy interview – then how can you check that 863 people who answered yes to **a_health** were interviewed by proxy?

Producing graphs

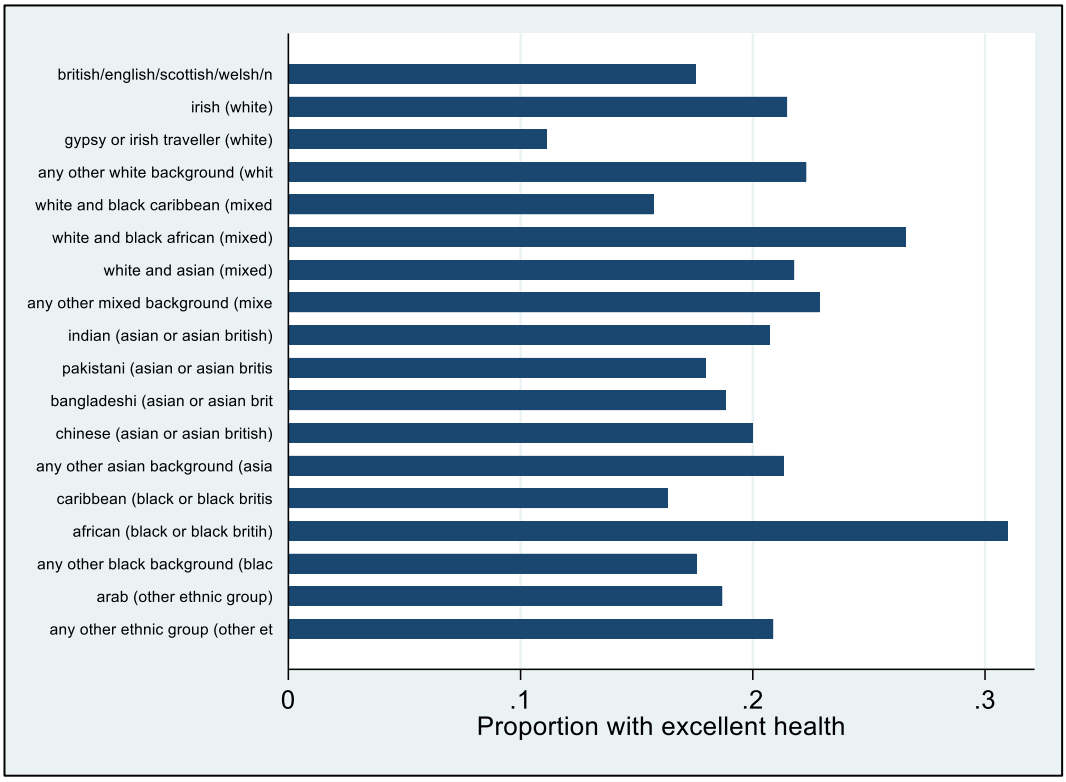
You can also produce nice graphs in Stata to show distributions. For example, to show the distribution of adults with excellent health across ethnic groups you can produce bar graphs.

First produce a 0-1 variable that takes on the value 1 if a person reports excellent health

```
generate excellent_health=0 if a_sf1>=2 & a_sf1<=5
replace  excellent_health=1 if a_sf1==1
```

To produce a horizontal bar graph showing the proportion with excellent health across ethnic groups in Stata type

```
graph hbar excellent_health, over(a_racel_dv) ytitle("Proportion with
excellent health")
```



Population estimates and weights

Can you use these estimates to say something about the population? For example, can you say that 75.18% of adults are “white: British/English/Scottish/Welsh/Northern Irish” (see Page 4)? The answer is no. That is because the 50,994 adults who responded in Wave 1 include more people of *some types* than in the population.

One reason for this is non-response: not everyone who was invited to participate agreed to participate. This could be for a variety of reason such as they were too busy to participate or were not at home etc. Research has found that older people are more likely to participate (Lynn et al 2012). So, if there are more older people in the sample than in the population and as older people are more likely to have poorer health it means that the estimate of % of people with poor health will be higher in the sample than in the population.

Another reason for some sub-populations being over/under represented in the sample is because of how the sample was designed. For example, in *Understanding Society* sample, some minority ethnic groups and Northern Ireland residents are over-represented. As a result the sample will have a higher percentage (%) of ethnic minorities than in the population (and a lower percentage of “white: British/English/Scottish/Welsh/ Northern Irish”).

Weights are designed to reverse the effect of having a sample that includes more/less people of certain types than the population. That is, weights are higher for individuals under-represented in the sample and lower for individuals over-represented in the sample. So if you produce weighted estimates you will have a more accurate population estimate.

Which weights to use? How to use weights?

See the User Guide to find the correct weights to use.

If you want to produce the weighted estimate of adults with poor health living in the UK in 2009-10, given that the general health question, **a_sf1**, was asked of all adults including proxy respondents, the correct weight to use is **a_indpxus_xw**. To produce this weighted estimate in Stata type,

```
tabulate a_sf1 [aweight= a_indpxus_xw]
```

General health	Freq.	Percent	Cum.
excellent	9,451.0825	18.82	18.82
very good	16,261.1791	32.38	51.21
good	13,813.904	27.51	78.72
fair	7,145.4922	14.23	92.95
or Poor?	3,541.342	7.05	100.00
Total	50,213	100.00	

As you can see the weighted estimate of the percentage of adults who report poor health is 7.05% which is less than the unweighted estimate of 7.25%.

If you want to produce the weighted estimate of adults who report their ethnic group as "white: British/English/Scottish/Welsh/Northern Irish" in 2009-10, given that this question was not asked of the proxy respondents the correct weight is **a_indinus_xw**. To produce this weighted estimate in Stata type,

```
tabulate a_racel_dv [aweight=a_indinus_xw]
```

Ethnic group incorp. all waves, codings, modes and bhps	Freq.	Percent	Cum.
british/english/scottish/welsh/northern	40,302.483	85.72	85.72
irish (white)	636.748502	1.35	87.07
gypsy or irish traveller (white)	17.5459751	0.04	87.11
any other white background (white)	1,737.4301	3.70	90.80
white and black caribbean (mixed)	156.56419	0.33	91.14
white and black african (mixed)	76.10520055	0.16	91.30
white and asian (mixed)	134.084508	0.29	91.58
any other mixed background (mixed)	105.173772	0.22	91.81
indian (asian or asian british)	1,127.5886	2.40	94.21
pakistani (asian or asian british)	602.302852	1.28	95.49
bangladeshi (asian or asian british)	233.050492	0.50	95.98
chinese (asian or asian british)	179.885683	0.38	96.37
any other asian background (asian or as	346.728009	0.74	97.10
caribbean (black or black british)	403.202547	0.86	97.96
african (black or black britih)	603.751538	1.28	99.24
any other black background (black or bl	37.43267389	0.08	99.32
arab (other ethnic group)	175.69271	0.37	99.70
any other ethnic group (other ethnic gr	142.229807	0.30	100.00
Total	47,018	100.00	

As you can see the weighted estimate of the percentage of adults who report their ethnic group as "white: British/English/Scottish/Welsh/Northern Irish" in 2009-10 is 85.72% which is more than the unweighted estimate of 75.18%.

Exercise 5: Produce weighted estimate of proportion of Caribbean and Irish women reporting excellent health.

ADDITIONAL INFORMATION

If you want to create the teaching datafile a_indresp_td.dta from the datafile a_indresp.dta, then use the Stata syntax file **data_prep.do**

The Stata code for examples and the exercises discussed here are in the Stata syntax file **data_analysis.do** and the output is recorded in the text file **data_analysis.log**

HELP AND SUPPORT

Understanding Society has a wealth of information on our website:

It is a highly comprehensive online source of information regarding its variables, methodology, survey design and implementation. It is also an up to date source of training courses, webinars, videos, data releases and other relevant news regarding longitudinal research.

Email User Support if you have any comments or suggestions about this teaching resource.

ACKNOWLEDGEMENTS

This is a collaborative project by Understanding Society and the UK Data Service.

Acknowledgements go to Malcom Brynin (University of Essex) and Annette Pasotti (University of Essex).

Understanding Society
Institute for Social and Economic Research (ISER)
University of Essex
Wivenhoe Park
Colchester
CO4 3SQ
Tel: +44 (0) 1206 872957

www.understandingsociety.ac.uk

@usociety
Understanding Society –
UK Household Longitudinal Study

Understanding Society has been commissioned by the Economic and Social Research Council (ESRC). The Scientific Research Team is led by the Institute for Social and Economic Research (ISER) at the University of Essex.