

GENERAL HOUSEHOLD SURVEY 2006: GUIDELINES FOR USING THE DATA

1) GHS questionnaires

The 2006 questionnaires have been provided in Word format. The household section is asked first and is called "Hshldsec06.doc". The individual section is called "Indsec06.doc".



2) GHS data files

The data files include households where a response from one or more individuals was missing. On person level files such individuals can be identified by the variable Schedtyp (where Schedtyp= 3). Two data files have been supplied; one at person level and the other at household level.

Person-level file

Ghs06client.sav contains person-level data. There is one record for every individual, including children, living in each responding household. This file contains schedule variables (that is, variables which were on the original questionnaire) and the derived variables which have been programmed. All household-level variables have also been attached to all individuals in each household on the person level file.

The file includes four serial numbers which denote households, families and people. Hserial is a household identifier; each household has a unique value at Hserial, and this value is attached to everyone in the household. Fserial is a family identifier. Each family unit has a unique value at Fserial, and this value is attached to everyone in the family unit.

Persno is a person identifier; each person within a household has a separate person number. Each person also has a unique PID number.

This file can be used to run tables at person level, household level or family level (see below), but for ease of access we have also supplied a household level file.

Household file

Ghs06clientHhld.sav contains household data. There is one record per responding household. The file contains schedule variables and derived variables. Some schedule variables have been deleted or grouped to maintain confidentiality.

Creating household and family files from the person level file

The variable HRP is the person number of the HRP so to identify households you need to select on HRP=PERSNO.

For families you use FUH=PERSNO (FUH is the person number of the family unit head).

Weight variable: Weight06

The data set is unweighted. Weight06 is the variable you should use to weight the data (see http://www.statistics.gov.uk/downloads/theme_compendia/GHS06/AppendixD2006.pdf). This weight applies to both household and individual level data.

3) Missing values

For each variable, there are three possible missing values. For the majority of variables, these are -6, -8, and -9. No values have been declared missing on the files supplied.

-6 means that the person was not eligible to answer a particular section and is therefore not eligible for inclusion in the relevant derived variable. For example, smoking questions are only asked of adults, so children are coded as -6 for smoking variables. Adults aged 16-59 are asked the family information questions, so children and adults aged 60 and over are coded as -6 for these variables.

-8 (NA) means 'No answer'. This is used where a respondent either does not know the answer or refuses to answer an individual question.

-9 (DNA) means 'Does not apply'. This is used when a household or respondent is eligible to be asked a question, but is routed past it by the flow of the questionnaire. For example, if a respondent is asked whether he or she is eligible to belong to their employer's pension scheme and the answer is 'no', he or she will be coded as -9 on the follow-up question which asked whether he or she belongs to the scheme.

Missing values are defined differently for the income section to make it possible to identify those who have refused the whole section. As SPSS only allows the use of three missing values, we have combined -6 and -9. For income variables, the three missing values are -7, -8 and -9.

-7 means that a respondent has refused the whole income section.

-8 (NA) has the same meaning as for other variables.

-9 combines -6 and -9, and therefore includes people who are not eligible for the income section, such as children, and respondents who are routed past questions by the flow of the questionnaire.

Some words of warning

If you are using SPSS-NT, then it is very easy to save amended versions of SPSS files when you don't want to. When you try to quit SPSS, the dialogue box asks if you want to save the file. The intuitive reaction is to click on 'yes', but you should only do this if you want to save any amendments you have made. If you have used a 'keep' or 'select' statement, clicking 'yes' will save only the cases or variables you have selected. We recommend that you keep the CDROM or original file safe, just in case you inadvertently change the data files.

Both data files are large, and it takes time to run frequencies and tabulations on the full files. We recommend that you use a 'keep' statement to select relevant variables.

The Social and Vital Statistics division at ONS is currently using version 12.0 of SPSS. You may find you have problems with missing values when trying to create derived variables. The solution we have used is to declare no missing values, then create the new variables, then declare missing values.

4) Excel spreadsheet (Variables – changes)

The list of variables on the Excel spreadsheet denotes changes made between the 2005 and the 2006 datasets including:

- Changes to variable names
- Variables (either derived or questionnaire) that have been dropped
- New variables (either derived or questionnaire) that have been introduced

All changes are listed on the Excel spreadsheet.

5) Specifications for derived variables

Also in the zipped file is a specification for derived variables on the data file.



GHS Team
Social and Vital Statistics Division
February 2008

Changes to GHS questionnaire for 2006

Household Questionnaire

Household Information

The variables MarStat, LiveWith and dvmardf have been altered to include civil partnerships. To highlight this change their names have been changed to MarSta, LivWth and dvmrdf respectively.

Individual Questionnaire

Pensions

Variables: pppamnt and ppper

Renamed: PenAmnt and PenPer

Health

Variable: illness

Renamed: LSIII

Childcare

New childcare questions: ChAtt and ChPeo

Drinking

Although not a change to the questionnaire, it is worth noting here that there has been a change in the methodology for deriving the number of units consumed by respondents from their descriptions of what they drink. The data set contains two sets of variables for number of units of alcohol consumed, one using the previous method of calculation and one using the revised method.

Old method:

alcwkg1 alcwkg2 alcwkg3 drating l7alcg1 l7alcg2 l7alcgrp l7alctot nbl7unit nbunits qnbeer qsbeer qwine sbl7unit sbunits wl7tot

Revised method:

ralcwkg1, ralcwkg2, ralcwkg3, rdrating, rl7alcg1, rl7alcg2, rl7algrp, rl7altot, rnbl7uni, rnbunits, rqnbeer, rqsbeer, rqwine, rsbl7uni, rsbunits, rwl7tot

The revisions affect the calculations for units of alcohol in wine and beer. The calculations for spirits, sherry and alcopops are unaffected (spl7tot shyl7tot pol7tot qspirit qsherry qpops).

Income

Variables: btowamt btowdk btowpd dropped. Back to work bonus no longer paid

Smoking

tarlev06, tar06g1, tar06g2 - new tar levels

Trailers

Trailer on "Social Mobility" removed.

soc2kf
soc2km
famsitu
nbrossis
dadyofb
mumyofb
fthecsta
mthecsta
svisef
svisem
nempleef
emp1to9f
dvempnof
nempleem
emp1to9m
dvempnom
ftheduq
mtheduq
finprobl

Trailer on "Social and Cultural Participation" added.

Care
FreqArt
FreqCin
FreqCnF
FreqCnR
FreqLSpt
FreqMtF
FreqMtR
FreqPerf
Grplnf01
Grplnf02
Grplnf03
Grplnf04
Grplnf05
Grplnf06
Grplnf07
Grplnf08
Grplnf09
Grplnf10
GrplIntro
IIIBed
Money
VolFreq