GUIDELINES FOR USE OF THE DLHE LONGITUDINAL SURVEY DATASET

NATIONAL CENTRE FOR SOCIAL RESEARCH
3 AUGUST 2007

1 INTRODUCTION

This paper is designed as a simple guide for users of the DLHE longitudinal survey dataset. It includes guidance on how to generate statistics from the data using the survey weights, how the precision of those statistics can be estimated, how to present tables generated from the data, and some guidance on the interpretation of data from the tables.

The Longitudinal Survey has a complex structure. This means that assistance from a colleague with technical statistical knowledge, especially of statistical software packages, will be needed. The relevant parts of this paper that you should refer him/her to are Sections 2, 4 and 5.

2 AN OVERVIEW OF THE SURVEY DESIGN

The DLHE Longitudinal Survey is a follow-up of the 2003 DLHE census of students who left Higher Education between August 2002 and July 2003. The 2003 study involved sending questionnaires to *all* eligible students (of which there were 412,579) and 307,652 (75%) questionnaires were completed and returned. The Longitudinal Survey was designed as a three-year follow-up of the census, the aim being to collect information on student destinations over this period.

The Longitudinal Survey did not involve sending questionnaires to *all* students but was instead based on a *sub-sample* of 62,039 of the 307,652 students who responded to the census questionnaire. These 62,039 were selected from across all HEIs, but with some groups of students being over-sampled relative to others so that the sample is particularly skewed towards foundation degree and Research Council students, non-English domiciled students and non-white students. The details of the sampling criteria used are:

- All Foundation degree and Research Council students were selected,
- All Black, Mixed and ‘Other ethnic group’ students were selected and 35% of Asians,
- 60% of NI domiciled students and 60% of students who studied at a NI HEI (non-NI domiciled),
- 40% of Welsh domiciled students and 25% of students who studies at Welsh HEIs (non-Welsh domiciled),
- 25% of Scottish domiciled students and 25% of students at Scottish HEIs (non-Scottish domiciled),
- 35% of TTA funded students,
- 15% of healthcare profession students, and
- 5% of all remaining students not already covered.

The rationale for the over-sampling was to ensure that the Longitudinal Survey would have sufficient numbers of students in key sub-groups to allow for separate statistical analyses of these groups. For instance, it means that 7,964 black students were selected, rather than the 1,587 black students that would have arisen from a simple random sample of 60,000.

Inevitably, the over-sampling means that the statistics generated from the Longitudinal Survey are incorrect unless they are adjusted to compensate for the over-sampling. How this is done, and how you should deal with it in your own analyses, is set out in Section 4.

3 THE IMPLICATIONS OF STARTING WITH A SAMPLE RATHER THAN A CENSUS

Because the original DLHE census involved all students and achieved a high response rate, the statistics generated from the census have been treated as exact.

In contrast, because the Longitudinal Survey is based on a sample rather than a census, the statistics from it will be not be exact and the reduction in precision needs to be taken into account in analysis. Put another way, although after accounting for the over-sampling of sub-groups, the sample *should* be representative of all students, there is no guarantee that it is *exactly* representative, and this lack of certainty has to be take into account.

The uncertainty around the statistics from a sample survey is usually presented by:

(a) computing confidence intervals around single statistics; and
(b) testing that differences between two (or more) statistics are 'real' by using formal statistical significance tests.

The interpretation of a confidence interval is that it gives an estimated range of values which is likely to include the true 'all student' value. Confidence intervals are usually calculated as '95% confidence intervals' which means that there is a 95% chance that the interval calculated from the sample covers the true 'all-student' value. It is possible to produce other confidence intervals, such as 90% ones, or 99% ones, but 95% has become the norm for most social surveys. The calculation of confidence intervals is discussed in Section 5.

The width of the confidence interval gives some idea about how uncertain we are about the true all-student value. The wider the interval the greater the uncertainty.

Illustrations:

Example 1: The weighted Longitudinal Survey data can be used to estimate the proportion of graduates working full- or part-time as their main activity at the time of the survey. From the survey it is estimated that 87% of graduates fit the criterion. The 95% confidence interval for the percentage is however 87% +/-1%. This can be interpreted as that we expect the true population value to lie within the range (86%, 88%).

Example 2: Restricting the analysis just to graduates who were studying at the time of the follow-up survey, 20% were doing a higher degree, mainly by research. The 95% confidence interval around this percentage is 20% +/-3%. In other words, we expect the true population value to lie within the range (17%, 23%).

Statistical significance tests for survey analyses work on a similar principle to confidence intervals, although they might appear very different on the surface. Essentially to test whether two (or more) survey statistics that *appear* to be different are *genuinely* different we start with the hypothesis (called the *null* hypothesis) that the statistics are in fact the same. The test then calculates the probability that the sample survey could, *just by chance*, have generated the size of difference observed in the survey. If this probability is low (and low is usually taken as 0.05 or less) then the null hypothesis is judged unlikely to be true and is rejected. That is, we conclude that the survey difference is a genuine difference.

Note that if the null hypothesis is not rejected then this means that the observed difference *could* have arisen just by chance. That is, the difference could be attributed to sampling error. We cannot however conclude *for certain* that there is no real difference: all we can do is say that the option of 'no real difference' hasn't been ruled out. In other words, statistical tests can be used to reject null hypotheses but generally they cannot prove null hypotheses. It is worth noting that the smaller the sample size the less likely it is that you will be able to positively reject null hypotheses.

Illustrations:

Example 1: As an illustration of testing the difference between two means; we can ask: is there a difference between the gross annual income of men and women graduates after three years?

The mean gross annual income of male respondents is £20,786, the mean for female respondents is £17,861. A t-test can be used to test the difference between these two means, the null hypothesis being that there is in fact no real difference and that the observed difference is just sampling error. The p-value for the t-test turns out to be <0.01 – that is, the probability that the observed gender difference of £3k arose just by chance is less than 0.01.

Since 0.01 is far lower than our cut-off of 0.05 the null hypothesis is rejected and we can conclude that there is very likely to be a gender difference in incomes.

Example 2: As a second illustration, we might ask is there an association between being in work at the time of the Longitudinal Survey and ethnic group?

Running a cross tabulation of ethnicity by employment activity at the Longitudinal Survey shows that 90% of respondents from a white background were in work compared to 86% of respondents from a non-white ethnic group. To test whether this difference is genuine, a chi-squared test is appropriate. The chi-squared test gave a p-value of <0.01. As with Example 1, this suggests there is a significant difference in work rates for white and non-white graduates.

Example 3: Finally, is there an association between being in work at the Longitudinal Survey and domicile?

A cross-tabulation of country of domicile by employment activity shows there to be small differences in the percentage of graduates in work by domicile, ranging from 89% for respondents domiciled in England to 91% for respondents domiciled in Northern Ireland. However a chi-squared test on this returns a p-value of 0.53. This is much larger than our cut-off of 0.05, which suggests these fluctuations in the proportion of respondents in paid work may well just be sampling error.

4 GENERATING UNBIASED STATISTICS FROM THE SURVEY (THE USE OF THE SURVEY WEIGHTS)

The sample design of the Longitudinal Survey, with some groups of students being over-sampled and hence over-represented, means that **the data from the survey cannot be used without adjustment.**

The adjustment is done by weighting the cases in the data files. To illustrate how this is done in principle, black, mixed ethnic group and 'other ethnic' students were selected so that they represented 21% of the selected sample. But from the initial census we know that these students represent just 4% of all students. So, to ensure these students feature in the analysis in their correct proportion, the black, mixed and other ethnicity students in the sample are given a weight of 4/21.

The weight to apply to each case is included in the data files as a separate variable. For analysis at a national level the weight to use is called 'finalwt'. **You must weight the data by finalwt for all *national* analyses**.

The weight to apply for analysis of the data from individual HEIs is called hei_wt. **You must weight the data by hei_wt for all analyses of the data for *individual HEIs***.

Note that the variables finalwt and hei_wt take into account differential non-response rates to the Longitudinal Survey by sub-groups of students as well as the over-sampling. This is done to ensure that the final data set of students who returned a questionnaire is reasonably representative of all students.

The difference between finalwt and hei_wt is that the latter includes an additional adjustment to ensure that the weighted data for each HEI gives a reasonable match to the 2003 DLHE census on a small number of key variables, as shown in the box below. Since adjustment of this kind is only statistically efficient for large sample sizes, the amount of adjustment per HEI is dependent on the Longitudinal Survey sample size.

For HEIs with 400 or more Longitudinal Survey respondents the survey data are weighted so as to give a close percentage match between the survey and the census in terms of broad subject group[1], the part-time/full-time split; and the postgraduate/undergraduate split.

For HEIs with between 200 and 399 Longitudinal Survey respondents the survey data are weighted so as to give a close percentage match between the survey and the census in terms of the part-time/full-time split; and the postgraduate/undergraduate split.

For HEIs with between 100 and 199 Longitudinal Survey respondents the survey data are weighted so as to give a close percentage match between the survey and the census in terms of the postgraduate/undergraduate split.

For HEIs with fewer than 100 Longitudinal Survey respondents no HEI level adjustment has been done.

It is important to use the weights for all analyses including analysis of sub-groups. Only for sub-groups where all the weights are exactly equal is unweighted analysis valid – but in these instances unweighted analysis will simply give the same findings as the weighted analysis so weighted analysis is still appropriate.

4.1 Software implications

Because the survey data has to be weighted this means that generating statistics from the data has to be done in a package that can handle weighted data. This means you probably need a statistical software package such as SPSS (versions 12 onwards), Stata, or SAS for your analysis. Preferably you should also use a package that can handle complex survey data (that is, data generated using non-standard sample designs).

Note that in SPSS the complex samples module has to be purchased as a separate module. Although weighted analysis in earlier versions or outside of the module will give correct means and percentages, confidence intervals and significance tests will be incorrect.

In Stata the 'SVY' commands should be used; in SAS the 'PROC SURVEY' commands are the appropriate ones to use, although as with SPSS, any weighted analysis will give correct means and percentages. Confidence intervals and significance tests may however be slightly incorrect[2].

[1] The broad subject groups used are: health and welfare; science and agriculture; engineering, manufacture and construction; social science, business, law and combined; humanities and arts; education.

[2] SPSS calculates weighted confidence intervals, standard errors and statistical tests rather differently to Stata and SAS. Whereas the latter two can generate slightly inaccurate CIs and tests if the complex survey routines are not used, SPSS will, under some circumstances, be very inaccurate.

5 ESTIMATING THE PRECISION OF STATISTICS FROM THE SURVEY

The greatest influence on the confidence intervals and standard errors for statistics from the survey will be the sample size being used. The smaller the sample size the lower the precision and the wider the confidence intervals. This means that statistics for individual HEIs will have much wider confidence intervals than national statistics, and statistics for sub-groups will have wider confidence intervals than statistics based on all students.

Beyond sample size though the precision will also be affected by the design of the Longitudinal Survey sample, with the sample being selected as a series of separate samples from a large number of census sub-groups (which means that the survey is stratified). Related to the stratification, the precision is also affected by the subsequent weighting of the data. This means that the survey needs to be handled as a *stratified weighted sample*.

Generating confidence intervals and significance tests that take the stratification and weighting into account is not something that can be done easily by hand: it needs to be done within an appropriate statistical software package (see Section 4.1).

The dataset provided has a stratification variable (**strata**) as well as the weights.

6 PRESENTING TABLES FROM THE SURVEY DATA

In this section we include some good practice guidelines on how to present tables based on survey data. The example table below is used to illustrate the points.

Table 1.1 Activity on the 27th November 2006, by age and sex

Aged 18 and over *2006*

Activity on the 27th November 2006	**Age group**				**Total**
	18-20	21-24	25-29	30+	
	%	%	%	%	%
Men					
Work full time or part time	85	88	91	90	89
Not in paid work	15	12	9	10	11
Women					
Work full time or part time	85	89	91	92	90
Not in paid work	15	11	9	8	10
Bases (unweighted)					
Men	*292*	*5144*	*1163*	*2262*	*8861*
Women	*433*	*8724*	*1823*	*4971*	*15951*
Bases (weighted)					
Men	*566*	*9672*	*2487*	*5541*	*18266*
Women	*941*	*18975*	*4029*	*12727*	*36673*

1. ***Headings should be clear and specific***

2. ***The population that the figures in the table refer to should be specified***

   In the example above it is given in the left hand side of the second row of the table (as the population of graduates aged 18 and over).

3. ***Avoid spurious precision.***

   For a sample survey, where all statistics are to a degree, imprecise, it is misleading to generate percentage to several decimal places. Standard practice is to round percentages to the nearest integer, but for large samples (of several thousand) one decimal place would be acceptable.

   There is no default for other statistics, such as means. Judgement about what is reasonable needs to be applied here, although calculating the 95% confidence interval can give some indication of the 'margin of error' and hence of a reasonable level of precision to present.

4. ***Present both unweighted and weighted bases***

   In the example table above both unweighted and weighted bases are shown in italics at the bottom of the table. The two bases serve different purposes:

   - the unweighted bases give the raw sample counts that the statistics in the table are based on. This gives some indication of likely precision, and allows readers to make their own judgement about how to interpret statistics based on small sample sizes;

   - the weighted bases give the numbers of the students in the column in their correct proportions (relative to other columns). For example, in the table above, the weighted base for men aged 21-24 is 9,672 from a total of 18,266. This implies that around 53% of men in the survey population are aged 21-24.

5. ***Add footnotes to the bottom of the table if they are needed***

   Use footnotes to guide readers through the tables. For instance, a footnote might say that, because of rounding, percentages do not add to 100 exactly.

6. ***Alert readers to small unweighted sample bases by putting square brackets around the figures in the relevant column***

   Standard practice for UK official social survey statistics is that statistics are put into square brackets if they are based on a unweighted sample size of less than 50. This gives readers a very simple means of identifying statistics with very low precision. However, the threshold of 50 is just a convention: to avoid all HEI level tables being cluttered with square brackets we recommend that brackets are used for samples sizes of less than 25.

7. ***Avoid presenting statistics for very small sub-groups of students where there is a risk that the individual students in the sub-group will be identifiable***

   For example, for an individual HEI, students divided into sub-groups based on ethnic group and subject area of degree might give sample sizes that are so small that there is a high risk that individuals are identifiable. This needs to be avoided.

8. ***If percentages are rounded, adopt a convention that allow for genuine zeros to be distinguished from 'less than 0.5%'.***

   The standard is to use '–' for percentages that are genuinely zero and '0' for percentages of between 0 and 0.5%.

   Note that if this convention is used there should be a note of it included with the tables (either at the front of the report or as a footnote to all relevant tables)

7 INTERPRETING THE DATA FROM TABLES

In this section, we include a few pointers on how to interpret and report on statistics from tables in an interpretative report.

The aim of analysis will usually be to describe *the student population*, insofar as we can infer its characteristics from the sample, and not to describe the particular sample we happen to have selected. So if something is true of the sample, but likely to be a random sampling effect and thus not true of the population, it need not be commented on (except sometimes simply to explain that an apparently interesting difference is not significant and should be ignored).

There are no hard and fast rules that can be applied to table interpretation. But a few good-practice 'rules' have evolved over time:

Avoid reporting on small sample sizes where possible.

Statistics based on small sample sizes will be imprecise, and comparing two (or more) statistics based on small sample sizes can lead to some very misleading interpretations of data if this imprecision is ignored.

The best way to avoid mis-interpretation is to simply accept the limitations of the data: if some sub-groups have sample sizes that are too small for meaningful analysis then don't use these sub-groups in your analysis.

Use significance tests to test whether an observed difference between two groups is likely to be genuine given the size and design of the sample

If you identify a difference between groups that looks of interest then test that it is likely to be a genuine difference by using a formal statistical test.

To test for differences between percentages a chi-squared test is appropriate. But note that the weighting and the sample design need to be taken into account, so a statistical software package is needed.

Do not conclude that non-significance equates to equal

In contrast, if a difference proves not be statistically significant take care to report this as ‘no evidence of a real difference’ rather than as ‘no difference’.

Avoid reporting on outliers

Statistical testing is not a panacea. Even with large sample sizes it is possible that ‘blips’ in the data can occur (with, for instance, a percentage for one sub-group being unusually high relative to other groups) and some of these ‘blips’ may register as ‘significant’. Analysis that concentrates on the outliers, is less likely to give an adequate summary of the survey results than analysis that concentrates on identifying patterns in the data. To help with this we suggest:

1. Look for patterns rather than isolated differences.

If looking at, say, the relationship between degree award and destinations then look for a trend rather than concentrating on peaks and troughs in the distribution.

2. Look for corroboration in the national sample.

If, as an individual HEI, you see a potentially interesting relationship for your institution’s sample, check whether the relationship also arose in the national data. If it did then the relationship is more likely to be genuine.

3. Consider the inherent plausibility of the observed effect.

A difference that is plausible should be given more credence than one that is unexpected and implausible.

Take account of the magnitude of the difference.

Finally, for differences that are ‘statistically significant’ caution is also needed. For significant differences you should take account of the magnitude of that difference.

A difference can be “significant” – that is, indicative of a real student difference – without necessarily having much interest or substantive importance. In a sense, significance test results say more about the size of the sample than anything else: the larger the sample size the smaller the differences that are significant. Over-emphasis on significance can diminish the attention paid to magnitude. In interpreting differences it is useful to pose two questions: (a) ‘Is it significant’ and (b) is the difference large enough to be important.

8 DESCRIBING THE SURVEY IN PUBLISHED REPORTS

If using the data in reports we recommend that you include an appropriately edited version of the following paragraph.

‘Statistics in this report are based on data from the DLHE Longitudinal Survey. The DLHE Longitudinal Survey is a three-year follow-up of the 2003 DLHE census of students who left Higher Education between August 2002 and July 2003, the aim being to collect information on student destinations over the three-year period. The survey was based on a stratified sample of 62,039 respondents to the census questionnaire. In total 24,823 of the 62,039 selected graduates completed and returned a questionnaire.

The Longitudinal Survey sample deliberately over-sampled some groups of students relative to others and the survey data has been weighted to adjust for this. The survey weights also include an adjustment for differential response rates across sub-groups of students.

Tables show both unweighted and weighted bases, although all statistics are based on weighted data. The unweighted bases show the raw sample sizes; the weighted bases show the sample sizes after the weights have been applied.’