Tests of significance for between year comparisons

The purpose of this note is to summarise the two main tests of significance that can be used to compare survey results between years and to identify what needs to be done if we are comparing overlapping multi-year data sets.

There are two main significance tests, comparing individual survey estimates' confidence intervals and testing the confidence interval of the difference in percentages. In the EHCS the first of these methods is normally used, both for within year comparisons between subsets of the data and for between year comparisons. Samples are rolled together to obtain 2-year aggregates with larger samples and therefore we measure the change between two year averages rather than annual change between each individual year. We need to agree which is the correct approach we should be using for between year comparisons. Three examples are considered using the different methods.

Example 1: Percentage of stock non-decent for all tenures.

i) Significance test between sample estimates.

	2001	**02/03**	**03/04**
Percentage stock non-decent	33.3	31.2	29.2
Sample Size	17,532	16,648	16,502
Confidence interval (+/-)	0.70	0.70	0.69
Lower Confidence Interval	32.6	30.5	28.5
Upper Confidence Interval	34.0	31.9	29.9

Using a 2-tailed 95% confidence interval we can say that the percentage of stock non-decent for all tenures is lower in 2002/03 than in 2001. Also it shows that the percentage of stock non-decent for all tenures is lower in 2003/04 than in 2002/03. This does not take into account overlapping samples, surveyor variability etc.
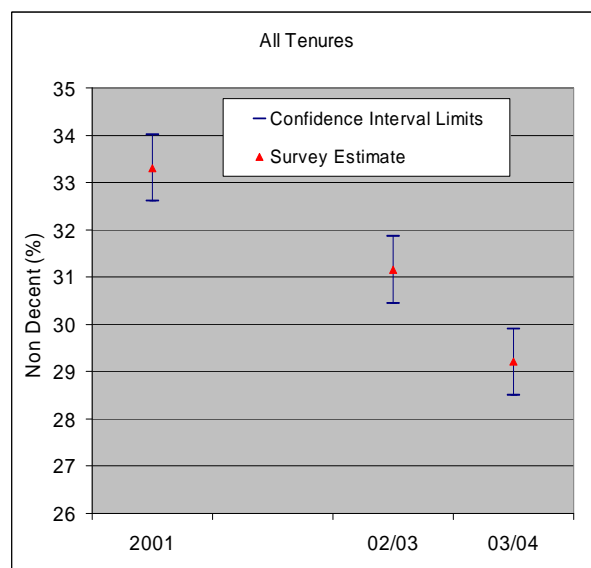
ii) Confidence interval for a difference in percentages.
To calculate the standard error of a difference in percentages between two years, we use the following formula:

$$ SE(p_1 - p_2) = \sqrt{\left(\frac{p_1(100 - p_1)}{n_1} + \frac{p_2(100 - p_2)}{n_2} \right)} $$

Where p_1 is the percentage in year 1 and p_2 is percentage in year 2 and where n_1 and n_2 are the sample sizes in year 1 and 2 respectively.

In the above example, the difference in percentage non-decent between 2001 and 02/03 is 2.15%. Using the above formula the 95% confidence interval for the difference in percentages is between 1.16% and 3.14%. Because this interval does not span zero we can conclude that the two means are significantly different.

The difference in percentage non-decent between 02/03 and 03/04 is 1.95%. Using the above formula the 95% confidence interval for the difference in percentages is between 0.97% and 2.94%. As the interval does not span zero we can conclude that the two means are significantly different. However this method does not take into account overlapping samples.

Standard errors of differences from overlapping multi-year data sets

We can take account of overlapping samples as follows. In the above example 2003 is common to both the 2002/03 and 2003/04 data sets. If the two year datasets are of similar size we can also assume that the variances of the estimates from them are the same, say v. Then:

$$\text{Var} (p_1 - p_2) = 2v (1\text{-}r)$$

where r is the degree of overlap between the two datasets.

In our example r is approximately 0.5 so:

$$\text{Var} (p_1 - p_2) \quad = 2v (0.5) \qquad = v$$
$$= \text{Var} (p_1)$$

Therefore $\text{SE} (p_1 - p_2) = \sqrt{\text{Var} (p_1)} \quad = \text{SE} (p_1)$

If there is no overlap between the years than r is 0, so:

$$\text{Var} (p_1 - p_2) \quad = 2\text{Var} (p_1)$$
$$\text{and} \quad \text{SE} (p_1 - p_2) \quad = \sqrt{2} \text{ x } \text{SE}(p_1)$$

Therefore the reduction in $\text{SE} (p_1 - p_2)$ due to the overlapping year $= (\sqrt{2} -1) /\sqrt{2}$
$$= 0.293$$

I.e. when the overlapping year is taken into account the standard error is reduced by around 29%.

In this example:

Difference between percentages in 02/03 and 03/04 =		-1.95
Standard error of the difference	=	0.50
Revised standard error of the difference	=	0.36

Applying the reduction in standard errors to the above example, then the difference between the two percentages is between 1.25% and 2.65% using a 95% confidence interval. As the interval does not span zero, we would still conclude that the two means are significantly different.

This method can be tested using a slightly different method (comparing a normal test statistic against normal tables), which is covered in the annex.

Example 2: Percentage of stock non-decent for private sector in South East.

i) Significance test between sample estimates.

	2001	02/03	03/04
Percentage stock non-decent	29.5	28.6	26.3
Sample Size	1110	1870	1988
Confidence interval (+/-)	2.68	2.05	1.94
Lower Confidence Interval	26.8	26.6	24.4
Upper Confidence Interval	32.2	30.7	28.2

Using a 2-tailed 95% confidence interval we cannot say that the percentage of stock non-decent in private sector in South East is lower in 2002/03 than in 2001. There is also no evidence to suggest the percentage of stock non-decent in private sector in South East is lower in 2003/04 than in 2002/03 or in 2001 at the 95% confidence level. This does not take into account overlapping samples, surveyor variability etc.

Private Sector South East
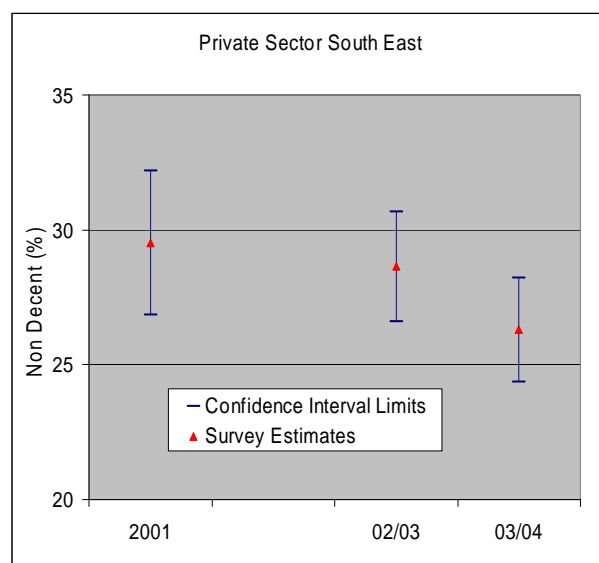
ii) Confidence interval for a difference in percentages.

In the above example, the difference in percentage non-decent between 2001 and 02/03 is 0.88%. Using the above formula the 95% confidence interval for the difference in percentages is between -2.50% and 4.26%. As this interval spans zero we can conclude that the two means are not significantly different.

The difference in percentage non-decent between 02/03 and 03/04 is 2.35%. Using the above formula the 95% confidence interval for the difference in percentages is between -0.47% and 5.17%. As this interval spans zero we can conclude that the two means are not significantly different. However this method does not take into account overlapping samples.

Taking into account the common year which is 2003, the standard error reduces to (100-29)% = 71% of what it would be if these two-year datasets were not overlapping. In this example:

|		
Difference between percentages in 02/03 and 03/04 =		-2.35
Standard error of the difference	=	1.44
Revised standard error of the difference	=	1.02

Applying the reduction in standard errors, the difference between the two percentages is between 0.36% and 4.34% using a 95% confidence interval. As the interval does not span zero we would conclude from this test that the two means *are* significantly different.

Example 3: Percentage of stock non-decent for Social Sector in West Midlands.

i) Significance test between sample estimates.

	2001	02/03	03/04
Percentage stock non-decent	43.5	37.1	31.6
Sample Size	726	633	586
Confidence interval (+/-)	3.61	3.76	3.77
Lower Confidence Interval	39.9	33.3	27.9
Upper Confidence Interval	47.2	40.9	35.4

Using a 2-tailed 95% confidence interval, we cannot say that the percentage of stock non-decent in social sector stock non-decent in West Midlands is lower in 2002/03 than in 2001, nor that it is lower in 2003/04 than in 2002/03. However it is lower in 03/04 than in 2001 at the 95% confidence level. This does not take into account overlapping samples, surveyor variability etc.

Social Sector West Midlands
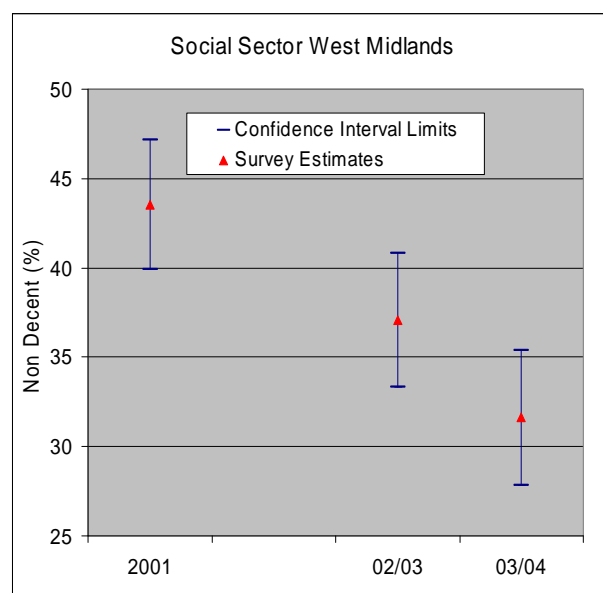
ii) Confidence interval for a difference in percentages.

In this example, the difference in percentage non-decent between 2001 and 02/03 is 6.46%. Using the above formula the 95% confidence interval for the difference in percentages is between 1.23% and 11.67%. As this interval does not span zero we can conclude that the two means are significantly different.

The difference in percentage non-decent between 02/03 and 03/04 is 5.46. Using the above formula the 95% confidence interval for the difference in percentages is between 0.13% and 10.78%. As the interval does not span zero we can conclude that the two means are significantly different. However this method does not take into account overlapping samples.

When comparing 02/03 and 03/04 then 2003 is a common year, so the standard error of the difference reduces to (100-29)% = 71% of what it would be if these two-year datasets were not overlapping. The table below shows the key statistics for our example:

Difference between percentages in 02/03 and 03/04 =	-5.46
Standard error of the difference =	2.72
Revised standard error of the difference =	1.92

Applying the reduction in standard errors, the difference between the two percentages is between 1.69% and 9.22% using a 95% confidence interval. As the interval does not span zero we would conclude that the two means are significantly different using this test, which is the opposite to our conclusion when testing the two sample estimates independently (see above).

Conclusion

When choosing between the two main methods we have described, the key point is that we want to test for the change in the same variable over time e.g. percentage non-decent in London in 2002/03 against percentage non-decent in London in 2003/04. As the two time-points are related (not independent) then it is appropriate to use the second method as the two datasets are interrelated. The first method is better suited to comparing independent variables/indicators, e.g. percentage non-decent in London against percentage non-decent in North West.

If a common year is involved between the two years being tested, then this should be accounted for, with the survey error of the difference being adjusted accordingly. In our examples, our conclusion is only different between the two tests in example 3, and as we have outlined that the second method is the one we should be using then we would conclude that the percentage of stock non-decent in social sector in West Midlands is significantly different between 02/03 and 03/04.

Annex

An alternative test statistic for a difference in percentages is:

$$z_1 = \frac{p_1 - p_2}{\sqrt{\left(\frac{p_1(100 - p_1)}{n_1} + \frac{p_2(100 - p_2)}{n_2}\right)}}$$

which has a Normal distribution with mean =0 and standard error =1

Using example 1, between 02/03 and 03/04 the difference is -1.95, and the revised standard error of that difference of 0.36, gives a test statistic of 5.42. Using a 2 tail normal test at the 5% level gives a critical value of +/-1.96. Our test statistic is above the critical value, i.e. there is a significant difference between the two percentages.

The advantage of this method is it allows you to calculate the probability of the difference being significant and it is easy to test at different levels of confidence, by looking up different values from Normal Tests. E.g. In example 2 (private sector in South East) between 02/03 and 03/04 the difference is 2.35, with a standard error of that difference of 1.02, which gives a test statistic of 2.30. Using a 2 tail normal test at the *1%* level gives a critical value of +/-2.576. Our test statistic is below the critical value, I.e. the evidence does not suggest a significant difference between the two percentages at the 1% level.

Calculating confidence intervals for crosstabulations

Note that these confidence intervals are based on a simple random survey (ie excluding design effects from stratification and surveyor variability) and therefore ***underestimate*** the size of the interval.

Example

1. SPSS syntax and output to produce required tables

```
GET
  FILE='D:\DATA\EHCS2001\Database\Provisional\coretgo3.sav'.
MATCH FILES /FILE=*
 /TABLE='D:\DATA\EHCS2001\Database\Derived\physical summary.sav'
 /BY aacod01.
EXECUTE.
MATCH FILES /FILE=*
 /TABLE='D:\DATA\EHCS2001\Database\Derived\tenure.sav'
 /BY aacod01.
EXECUTE.
MATCH FILES /FILE=*
 /TABLE='D:\DATA\EHCS2001\Database\Derived\unfit01x.sav'
 /BY aacod01.
EXECUTE.

*** The first table calculates 'p' (the probability of unfitness
    occurring in specified populations – in this example tenure groups
    and the whole population of dwellings) expressed as a percentage .

WEIGHT
  BY grdwtotg .
CROSSTABS
 /TABLES=ten401x  BY unfit01x
 /FORMAT= AVALUE TABLES
 /CELLS= ROW .
```

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Tenure 2001 - 4 categories * Fitness - final 2001	21355925	100.0%	0	.0%	21355925	100.0%

Tenure 2001 - 4 categories * Fitness - final 2001 Crosstabulation

% within Tenure 2001 - 4 categories

		Fitness - final 2001		Total
		unfit	fit	
Tenure 2001 - 4 categories	owner occupied	3.2%	96.8%	100.0%
	private rented	11.0%	89.0%	100.0%
	LA	4.7%	95.3%	100.0%
	RSL	3.4%	96.6%	100.0%
Total		4.2%	95.8%	100.0%

```
WEIGHT
  OFF.
CROSSTABS
  /TABLES=ten401x  BY unfit01x
  /FORMAT= AVALUE TABLES
  /CELLS= COUNT .
```

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Tenure 2001 - 4 categories * Fitness - final 2001	17532	100.0%	0	.0%	17532	100.0%

Tenure 2001 - 4 categories * Fitness - final 2001 Crosstabulation

Count

		Fitness - final 2001		
		unfit	fit	Total
Tenure 2001 - 4 categories	owner occupied	295	8413	8708
	private rented	187	1358	1545
	LA	223	4324	4547
	RSL	85	2647	2732
Total		790	16742	17532

2) Use of standard tables to calculate approximate confidence intervals for each tenure and for all dwellings, and a comparison with the actual confidence intervals produced by applying standard formula.

The standard formula to derive the 95% confidence interval (expressed as a percentage) is given by:

$$CI = (1.96 \text{ x } SQRT((p \text{ X } (1 - p/n)) \text{ x } 100.$$

Unfitness rate by tenure, 2001

tenure 2001:	unfitness p(%)	marginal count n(raw cases)	approx CI from tables +/-	calculated CI (simple random sample) +/-
owner occupied	3.2	8,708	0.36	0.37
private rented	11.0	1,545	1.58	1.56
LA	4.7	4,547	0.64	0.62
RSL	3.4	2,732	0.67	0.68
all tenures	4.2	17,532	0.29	0.30