

## D. RELIABILITY OF THE ESTIMATES

The findings presented in this publication are **estimates** based on a sample survey, and may be subject to two types of errors:

### 1. Sampling Errors

Sampling errors occur because only a sample of the population is investigated. The sample used in this survey is one of a large number of possible samples of the same size that could have been selected using the same sampling method. Clearly, estimates based on different samples will differ from one another, and almost all of them will differ from the value that would have been obtained had a complete census been taken (“the census value”).

The **estimate**  $X'$  is the estimated value, based on the specific sample used in this survey, for the corresponding value  $X$  that would have been obtained had a complete census been conducted.

The **sampling error of an estimate**  $\sigma(X')$  is a measure of the average difference between all the estimates that would have been obtained from all possible samples of the same size and derived by the same method, and the value that would have been obtained had a complete census been conducted, under the same conditions of data collection.

Because it is clear that a given estimate (almost always) deviates from the census value, it is not advisable to relate exclusively to the value estimated from a sample. Rather, it is preferable to relate to the interval in which the census value is likely to be found with a given probability (or level of confidence). This can be done on the basis of the **sampling error**.

A **relative sampling error of the estimate** is calculated as the ratio of the sampling error to the estimated value, i.e.,  $\sigma(X')/X'$ .

The **confidence interval** for an estimate is the interval that contains the census value  $X$  at a given level of confidence. Using both the estimate  $X'$  and its estimated sampling error  $\sigma'(X')$ , it is possible to construct a confidence interval at a predetermined confidence level, so that the interval contains the census value  $X$  with the given level of confidence.

Confidence intervals are usually given at a 95% level. The confidence interval at a 95% confidence level is obtained from the interval between  $X' - 2\sigma'(X')$  and  $X' + 2\sigma'(X')$ . Thus, it may be asserted with 95% confidence that the census value lies in this interval.

It is possible to set a higher or lower confidence levels, and to calculate the confidence interval as follows:

$\alpha$	67%	80%	90%	95%	99.5%
$K(\alpha)$	1.0	1.3	1.7	2.0	2.8

$K(\alpha)$  (the number of sampling errors) is determined according to the required level of confidence,  $\alpha$ .

Sampling errors have been calculated for all the estimates presented in this publication. In addition, the CD attached to this publication contains tables parallel to the tables of the estimates, which contain two sampling errors (95% confidence interval).

As a precaution against using estimates subject to high sampling errors, the tables in this publication present estimates with relative sampling errors of 25% to 40% in brackets. Estimates with relative sampling errors over 40% are not presented; instead, the symbol “..” appears in the tables.

Example:

In 2002 the estimate for lone-parent families, with children up to age 17, who live in households with more than one family, is 19,700 (see Table 14). According to the corresponding table in the attached CD, the two sampling errors of the estimate are 3,000 and the confidence interval with a 95% level of confidence is  $19.7 \pm 3.0$ , i.e., it may be asserted with 95% confidence that the number of lone-parent families with children up to age 17, who live in households with more than one family is between 16.7 and 22.7.

**Note:** The confidence intervals are usually symmetrical around the estimate, but for estimates based on a small number of cases in the sample (less than forty cases) the confidence intervals are not symmetrical. In these cases, both the estimate and its estimated sampling error are subject to a high error.

**Comparison of Estimates of Exclusive (i.e., non-Overlapping) Groups**

By using the sampling errors of estimates related to **two exclusive groups** (e.g., men and women), it is possible to test whether the difference between them is statistically significant.

If  $X'_1$  is the estimate for Group 1 and  $X'_2$  is the estimate for Group 2, then the estimated difference between the two groups is:  $D' = X'_1 - X'_2$ .

The confidence interval for the  $D'$  difference at an  $\alpha$  level of confidence is:  $D' \pm k(\alpha)\sigma(D')$ .

If a 95% level of confidence is desired, the confidence interval for the  $D'$  difference is:  $D' \pm 2\sigma(D')$ .

The estimate for two sampling errors of the estimated difference is calculated as follows:

$$2\sigma(D') = \sqrt{(2\sigma(X'_1))^2 + (2\sigma(X'_2))^2}$$

$2\sigma(X'_1)$  and  $2\sigma(X'_2)$  are the two sampling errors which are suitable for each one of the estimates, and which appear on the attached CD.

**Note:** The above formula can also be used to obtain two sampling errors for the estimated difference between two years. However, if the difference is between **parallel** estimates of **two consecutive years**, one should multiply the two sampling errors produced by the formula by a correction factor equal to 0.7. In that case, the estimate for the two sampling errors of the estimated difference will be:

$$2\sigma(D') = 0.7 \cdot \sqrt{(2\sigma(X'_1))^2 + (2\sigma(X'_2))^2}$$

If the confidence interval contains the value zero, we can say that the difference is not significant. I.e., according to the specific sample in the survey, it is **impossible** to say with an  $\alpha$  level of confidence, that  $X_1$  is indeed different from  $X_2$  in the population itself (even if in the sample they are different from each other).

If the interval does not contain the value zero, we can say with an  $\alpha$  level of confidence that there is a significant difference between the two groups.

In any case, it is possible to say with an  $\alpha$  level of confidence, that the difference between the two groups is between  $D' - k(\alpha)\sigma(D')$  and  $D' + k(\alpha)\sigma(D')$ .

Example:

In 2002 approximately 476,000 men aged 25-34 lived in households; 10.5% of them in non-family households. The number of women in this age group who lived in households was 476,500, and 7.6% of them lived in non-family households (see Table 12).

The estimated difference is:  $D' = 10.5\% - 7.6\% = 2.9\%$ .

According to the corresponding table of the two sampling errors in the attached CD, the two sampling errors of the estimates are 1.1% for men and 0.9% for women.

According to the above formula, the two sampling errors of the estimated difference are:  $2\sigma(D') = \sqrt{1.1^2 + 0.9^2} = 1.4\%$ , and therefore the 95% confidence interval for this difference is:  $2.9 \pm 1.4$ . Since the confidence interval does not contain the value zero, we can say with a 95% level of confidence that the difference is significant, and that there is a difference between the percentage of men living in non-family households out of all men aged 25-34, and the corresponding percentage for women. The difference can be found in the interval of  $(2.9 - 1.4 =) 1.5\%$  and  $(2.9 + 1.4 =) 4.3\%$ .

## 2. Non-Sampling Errors

Non-sampling errors in the survey can result from many sources at each stage of data collecting and processing, and they also exist in a complete census.

For additional details, see: Central Bureau of Statistics, *Labour Force Surveys 2002*, Special Publication 1218, Jerusalem, 2004.