

Weighting of Results

1. Two kinds of weights are used in processing the answers to qualitative questions. Here they are termed *sample weights* and *size weights*.

Sample weights

2. Sample weights are the inverse of the probability with which each reporting unit has been selected. Suppose, for example, that the target universe has been divided into two groups – large and small reporting units. If all large reporting units are selected for the sample (probability of 1) and if only one in ten are selected from the small reporting units (probability of 0.1), the answers of the reporting units must be multiplied by $1/1 = 1$ and $1/0.1 = 10$ respectively. Higher weights are given to the small reporting units because they have to represent all the other small reporting units that were not selected for the sample.

3. If the reporting units are selected on a simple random basis, e.g. without stratifying the target universe into large and small units as in the example above, the probability of selection is the same for all units and the weights are therefore identical. In this case no sample weights need to be used in processing the answers. The recommendation in this *Manual*, however, is to use a stratified random sample and in this case, sample weights are required reflect the probability of selection for units in the different size-strata.

Size weights

4. Size weights are used in processing qualitative answers because the importance of the answers is assumed to depend on the size of the reporting units. The answers from a large firm carry more weight than answers from a small one.

5. Strictly speaking, the variables to be used as size weights should depend on the survey variable concerned. For example, questions about production in a given reporting unit ought to be weighted by the relative value of production by that unit in the branch as a whole, questions about employment with the number of persons employed, etc. It would be costly, or even impossible, for most countries to obtain such a set of weights for each reporting unit. Furthermore, practical experience has shown that the balances (*B*) are not very sensitive to the choice of weighting variables. In practice it is sufficient to use a single variable reflecting the general economic importance of the enterprise in weighting all the survey answers.

6. If a single variable is to be used, value added should probably be used as the weighting variable because the business tendency survey results will then most closely reflect movements in GDP. At the branch level (e.g. *manufacture of footwear and clothing*) the commonest practice is to use numbers employed. This is the most widely available size measure and, within branches, numbers employed are closely related to value added. However, in aggregating branches to higher levels (e.g. *manufacture of consumer goods* or *total manufacturing*) it is recommended that value added should be used as the weighting variable.

7. Information on value added in the different branches will usually come from national accounts sources industry statistics. Information on numbers employed may be available as one of the characteristics included in the survey frame. Alternatively, the business tendency survey questionnaire used for the first month or quarter of the year can be used to collect information on employment. The information can be requested in size classes rather than exact numbers so that the information can be easily provided without recourse to accounting records.

8. Note that size weights are not generally required in processing answers to quantitative questions because the answers already reflect the size of the reporting unit. Data reported on the value of sales, tons

of output, numbers employed, etc. will be in larger values, volumes or numbers for large firms than for small ones. All the answers to the harmonised questionnaires, however, require size weights. Note this also applies to questions which do not use three point scales, such as questions on capacity utilisation. In assessing the overall level of capacity utilisation for an industry, the percentage utilisation in large firms must be given more weight than the utilisation rates in small ones.

Processing without weights

9. If reporting units have been selected with unequal weights – stratified or PPS sample design – the sample weights and the size weights will work in opposite directions. The sample weights will increase the weight of the small reporting units and the size weights will reduce their weight (and *vice versa* for large units).

10. Because the two types of weights tend to cancel each other out, unweighted results are sometimes calculated. It is simpler and therefore quicker to process without weights and the use of weights may introduce errors if they are out of date or otherwise erroneous. This may be legitimate at a disaggregated, branch level but is not recommended for calculating results at higher levels of aggregation. In combining results for branches to sector totals, weights – preferably value added shares – should always be used.

Weighting methods in practice

11. The Harmonised questionnaires use three point scales for all but a few questions and this *Manual* recommends that the answers to these questions should be expressed as balances (*B*) between the (+) and (-) answers. This section explains how balances are calculated using sample and size weights. It starts by explaining the **general method** which is applicable even in the most demanding case when each respondent is given an individual weight. It is followed by a description of a **simplified method**.¹ This is followed by an illustrative example of the general and simplified methods.

General method

12. In situations where the survey units are the reporting units and simple random sampling is used so that the inclusion probability is the same for all reporting units, we get:

$$B = \sum_{i=1}^n (x_i / n) \quad (1)$$

where: *B* is the balance for answers to a specific question,

x_i takes values of +1, 0 or -1 depending upon the answer given by enterprise *i*, and

n is the number of reporting units in the sample

13. By multiplying the *B* s estimated from formula (1) by 100 the result is the balances expressed as percentages.

14. If stratified random sampling or Probably Proportional to Size (PPS) sampling is used, the formula gets slightly more complicated as follows:

1. Mr. Åke Lönquist, former Director at Statistics Sweden, has reported on sampling and weighting methods at several joint OECD-EU workshops and developed the illustrative example used in this Chapter.

$$B = \sum_{i=1}^n (1/f_i) \times (x_i/N) \quad (2)$$

where the additional symbols are defined as follows:

f_i = the sampling probability of reporting unit i ; and

N = the total number of reporting units in the sampling frame (ideally identical with the target universe).

15. Both formulae (1) and (2) result in balance estimates that give equal weight to all respondents. For opinion poll type questions where the “one man, one vote” approach is deemed to be the most appropriate these unweighted estimates are the best choice. In business tendency surveys, however, the economic significance of the replies from different enterprises is linked to the size of the enterprise -- the larger the enterprise the more important is its reply. As noted above, size weights are therefore required.

16. The fact that the survey units will usually be the enterprises but the reporting units will be the *KAUs* belonging to the enterprise, complicates the estimation of the B s. We cannot denote the answer to a question provided by enterprise “ i ” with “ x_i ”, because there is a separate answer for each *KAU* of the enterprise. The notation “ x_{ik} ” has to be used instead to denote the answers of enterprise “ i ” regarding its activity in industry “ k ”.

17. The formula for estimating the balances of industry k , (B_k) assuming simple random sampling and without weighting for the size of individual *KAU* units will be:

$$B_k = \sum_{i=1}^{n_k} (x_{ik} / n_k) \quad (3)$$

where: $x_{ik} = +1, 0$ or -1 depending upon the answer given by enterprise i regarding its activity in industry k ; and

n_k = the number of *KAU*:s in the sample which are classified to industry k .

18. When sampling with varying probabilities is used, the formula will change to:

$$B_k = \frac{\sum_{i=1}^{n_k} (1/f_i \times x_{ik})}{\sum_{i=1}^{n_k} (1/f_i)} \quad (4)$$

where $\sum_{i=1}^{n_k} (1/f_i)$ is an estimate of N_k , the number of *KAUs* classified to industry k in the target universe.

19. If it is intended to show balances separately for size groups (s) within industries (k) the formula for the balance (B_{ks}) will be:

$$B_{ks} = \frac{\sum_{i=1}^{n_{ks}} (1/f_{iks} \times x_{iks})}{\sum_{i=1}^{n_{ks}} (1/f_{iks})} \quad (5)$$

where: f_{iks} = the sampling probability for enterprise i in industry k belonging to size group s ; and

x_{iks} = +1, 0 or -1 depending upon the answer given by enterprise i belonging to size group s for its activity in industry k .

20. When each report unit (KAU) is given a weight according to its size, the estimates of the balances (B_{ks}) from the sample will be:

$$B_{ks} = \frac{\sum_{i=1}^{n_{ks}} (1/f_{iks} \times w_{iks} \times x_{iks})}{\sum_{i=1}^{n_{ks}} (1/f_{iks} \times w_{iks})} \quad (6)$$

$$B_k = \frac{\sum_s (w_{ks} \times B_{ks})}{\sum_s w_{ks}} \quad (7)$$

where $w_{ks} = \frac{\sum_{i=1}^{n_{ks}} (1/f_{iks} \times w_{iks})}{\sum_{i=1}^{n_{ks}} (1/f_{iks} \times w_{iks})} \quad (8)$

$$B = \frac{\sum_k (w_k \times B_k)}{\sum_k w_k} \quad (9)$$

where $w_k = \sum_s w_{ks} \quad (10)$

21. In order to facilitate international comparability, when cut-off limits for sampling varying from country to country, it is recommended that countries apply stratified random sampling with the strata defined by employment, and to use at least the following three strata:

- 1-19 employed
- 20-249 employed
- 250 or more employed

22. This has the additional advantage of allowing the production of international comparisons by size of the enterprises. Countries using weighted estimates may need to stratify further.

23. In summary, it is recommended that countries use formulas (6), (7) and (9) for estimating the balances. If the w_{ks} values (total value added or employment in size group s of industry k) and/or the w_k

values (total value added or employment in industry k) are known from other sources, these values should be used for weighting rather than the estimates of them derived from the sample.

Simplified method

24. To simplify the computation of balances for the different industries the following simplified method is sometimes used. It is easier than the general method outlined above because it uses the same weight for all reporting units in each stratum and does not require detailed weighting information for each reporting unit. It is particularly suitable for use by research institutions that may not have access to the more detailed information on enterprises and KAU s that is available to official statistical agencies.

25. First, for every size group of reporting units within an industry the responses of each reporting unit are multiplied by the inverse of their sampling fraction ($1/f_i$). This fraction (f_i) is the sampling probability of the enterprise they belong to. For each question this gives an unbiased estimate of the number of reporting units in the target universe which would have responded (+), (=) and (-) in a comprehensive survey. These numbers - denoted here as $N^*_{ks}^{(+)}$, $N^*_{ks}^{(=)}$ and $N^*_{ks}^{(-)}$ respectively - sum up to N^*_{ks} , which is the estimate from the sample of the number of reporting units in the universe which belong to size category s of industry k^2 . In what follows these groups of reporting units are called *report strata*.

26. Second, for each report stratum (R_{ks}) the relative frequency of each response is computed. Denoting the relative frequency of response (+) by $A^{(+)}$ etc. we get:

$$A_{ks}^{(+)} = \frac{N^*_{ks}^{(+)}}{N^*_{ks}^{(+)} + N^*_{ks}^{(=)} + N^*_{ks}^{(-)}} \tag{11}$$

The analogous expressions apply to $A_{ks}^{(=)}$ and $A_{ks}^{(-)}$.

27. The balance for report stratum R_{ks} is:

$$B_{ks} = A_{ks}^{(+)} - A_{ks}^{(-)} \tag{12}$$

28. Third, from an external source (usually the frame from which the sample was drawn) the number of employees in each report stratum is computed. This number is used as weight for the stratum.

29. Fourth, in each industry (k) the balance (B_{ks}) of each report stratum (R_{ks}) is multiplied by the weight for the stratum (w_{ks}) and the products are added together for the industry. The resulting sum divided by the total weight of the industry ($W_k = \sum w_{ks}$) is the balance for this industry, B_k . The formula for this is:

2. An * after a symbol for a variable indicates an estimate of the size of that variable while the symbol without an * indicates the true size of the variable.

$$B_k = \sum^s (w_{ks} \times B_{ks}) / W_k \quad (13)$$

30. If information is wanted about the weighted relative frequency for each response alternative in industry k , $A_k^{(+)}$ etc. can be estimated in exactly the same way by substituting $A_{ks}^{(+)}$ etc. for B_{ks} in formula (13). Subtracting the estimate for $A_{ks}^{(-)}$ from that for $A_k^{(+)}$ gives the same value for B_{ks} as the direct calculation according to (13) described above.

31. Aggregating industries to groups and totals is done in exactly the same way as when the general method is used, i.e. formula (9) is applied.

32. The simplified method assigns all report units in a given stratum the same weight according to the number of employees in that stratum. Therefore it is recommended that the largest size group (200 or more employees) should be sub-divided into 200-499 and 500 or more employees, when the simplified method is used.

Numerical illustrations

33. This section contains numerical illustrations of the two calculation methods – general and special – described above. In order to show the differences and similarities between the two methods the same set of data has been used in both cases.

The numerical data

34. The assumptions about the size and composition of the universe and the design and size of the sample are recorded in Table 1 below. The sample described there is a stratified random sample in which enterprises are the sampling units. Stratification has been done by industry (Industry A and Industry B) and by two size groups with enterprises employing 250 or more persons in size group 1 and enterprises employing 1 - 249 persons in size group 2. Each enterprise consists of one or more KAUs which are the reporting units. Enterprises have been classified to the industry which accounts for the largest proportion of their employees.

Table 1. The target universe, the sampling and the reporting units

Industry and size group	Number of enterprises in the universe	Number of enterprises (survey units) in the sample	Number of KAUs (reporting units) in the sample
Industry A: size group 1	2	2	6
size group 2	20	2	3
Industry B: size group 1	6	2	3
size group 2	30	5	8
TOTAL	58	11	20

35. It is assumed that the information recorded in Table 2 is available from external sources.

Table 2. Report unit data for the target universe. Value added by industry and employment by industry and size group

Industry and size group of establishments	Value added	Number of employees
Industry A: size group 1	4.000 mln \$	2.000
size group 2		2.500
Industry B: size group 1	16.000 mln \$	5.000
size group 2		3.000
TOTAL	20.000 mln \$	12.500

36. The 11 enterprises and 20 reporting units, which the sample consists of according to Table 1, are listed in Table 3 below. The enterprises are numbered 1 - 11 and the reporting units belonging to enterprise 1 are numbered 1.1, 1.2 and 1.3, those belonging to enterprise 2 are numbered 2.1 and 2.2. etc. The reporting units belonging to the same enterprise all have the same sampling probability as the enterprise to which they belong. Therefore the contents of column 8 is the same for all reporting units belonging to the same enterprise and identical with column 4 for that enterprise. On the other hand, reporting units belonging to the same enterprise might belong to different industries and size groups. Examples of this are included in the list.

Table 3. List of enterprises and report units with data for numerical illustration of two weighting and estimation methods in BT-surveys

Enterprises				Reporting units					
No	Industry	Size group	Sampling Probability	No	Industry	Size group	Sampling Probability	Weight (Number of employees)	Response to question Q
Col 1	Col 2	Col 3	Col 4	Col 5	Col 6	Col 7	Col 8	Col 9	Col 10
1	A	1	1/1	1.1	A	1	1/1	600	+ 1
				1.2	A	2	1/1	100	0
				1.3	B	1	1/1	300	- 1
2	A	1	1/1	2.1	A	2	1/1	160	0
				2.2	B	1	1/1	250	+ 1
				2.3	A	2	1/1	140	- 1
	A	2	1/10	3.1	A	2	1/10	70	- 1
4	A	2	1/10	4.1	A	2	1/10	60	- 1
				4.2	B	2	1/10	50	+ 1
5	B	1	1/3	5.1	B	1	1/3	1000	- 1
6	B	1	1/3	6.1	B	1	1/3	400	0
				6.2	A	1	1/3	400	- 1
7	B	2	1/6	7.1	B	2	1/6	5	0
8	B	2	1/6	8.1	B	2	1/6	30	+ 1
9	B	2	1/6	9.1	B	2	1/6	140	+ 1
				9.2	A	2	1/6	20	0
				9.3	A	2	1/6	10	- 1
10	B	2	1/6	10.1	A	2	1/6	40	- 1
				10.2	B	2	1/6	80	- 1
11	B	2	1/6	11.1	B	2	1/6	170	+ 1

General method

37. Balances (B) are calculated as follows.

Table 4. Calculating the balance for industry A (B_a)

Report unit no.	Response [x_{ik}]	Sampling probability [f_i]	Adjustment factor to universe level [$1/f_i$]	Response adjusted to universe level [$x_{ik} \times 1/f_i$]	Weight [w_{ik}]	Weight adjusted to universe level [$1/f_i \times w_{ik}$]	Weighted response at universe level [$1/f_i \times w_{ik} \times x_{ik}$]
Col 1	Col 2	Col 3	Col 4 (1/Col 3)	Col 5 (Col 2* Col 4)	Col 6	Col 7 (Col 4*Col 6)	Col 8 (Col 5* Col 6)
1.1	+1	1/1	1	+ 1	600	600	600
1.2	0	1/1	1	0	100	100	0
2.1	0	1/1	1	0	160	160	0
2.3	- 1	1/1	1	- 1	140	140	- 140
3.1	- 1	1/10	10	- 10	70	700	- 700
4.1	- 1	1/10	10	- 10	60	600	- 600
6.2	- 1	1/3	3	- 3	400	1200	-1200
9.2	0	1/6	6	0	20	120	0
9.3	- 1	1/6	6	- 6	10	60	- 60
10.1	- 1	1/6	6	- 6	40	240	-1200
SUM	••	••	44	- 35	1600	3920	- 3300

•• = Irrelevant information

38. From the SUM row the balance for industry A is calculated as the quotient of column 8 / column 7 which amounts to $-3300 / 3920 = -0.84$. Expressed in percentage points this gives the balance - 84. The number of reporting units adjusted to universe level being 44, the unweighted balance is the quotient of column 5 / column 4 which amounts to $-35 / 44 = -0.80$ or, in percentage points, - 80.

39. Repeating the calculations for industry B we get the information recorded in Table 5.

40. From the SUM row in Table 5 the balance for industry B is calculated as the quotient of column 8 / column 7 which amounts to $-990 / 7800 = -0,13$. Expressed in percentage points this gives the balance - 13. The unweighted balance is derived from the SUM row as the quotient of column 5 / column 4 which amounts to $+19 / 48 = +0,40$ or in percentage points + 40.

Table 5. Calculation of balances for industry B, (B_B)

Report unit no.	Response [x _{ik}]	Sampling probability [f _i]	Adjustment factor to universe level [1/f _i]	Response adjusted to universe level [x _{ik} × 1/f _i]	Weight [w _{ik}]	Weight adjusted to universe level [1/f _i × w _{ik}]	Weighted response at universe level [1/f _i × w _{ik} × x _{ik}]
Col 1	Col 2	Col 3	Col 4 (1/Col 3)	Col 5 (Col 2* Col 4)	Col 6	Col 7 (Col 4*Col 6)	Col 8 (Col 5* Col 6)
1.3	- 1	1/1	1	- 1	300	300	- 300
2.2	+ 1	1/1	1	+ 1	250	250	250
4.2	+ 1	1/10	10	+ 10	50	500	500
5.1	- 1	1/3	3	- 3	1000	3000	- 3000
6.1	0	1/3	3	0	400	1200	0
7.1	0	1/6	6	0	5	30	0
8.1	+ 1	1/6	6	+ 6	30	180	180
9.1	+ 1	1/6	6	+ 6	140	840	840
10.2	- 1	1/6	6	- 6	80	480	- 480
11.1	+ 1	1/6	6	+ 6	170	1020	1020
SUM	+ 4	• •	48	+ 19	2425	7800	- 990

• • = Irrelevant information

41. For calculating the balance for industry A and B combined, equation (9) is used. In our numerical example the value added data in Table 2 (4.000 mln \$ for industry A and 16.000 miln \$ for industry B) are used as weights. In the formula numerical values are used as follows:

$$w_k \text{ for industry A} = 4.000 \quad B_k \text{ for industry A} = - 84$$

$$w_k \text{ for industry B} = 16.000 \quad B_k \text{ for industry B} = - 13$$

Consequently:

$$\sum^k w_k = 4.000 + 16.000 = 20.000 \text{ and}$$

42. Applying the same calculation method and industry weights to the unweighted balances we get

$$B = [4000 \times (- 80) + 16.000 \times (+ 40)] / 20.000 = [- 320.000 + 640.000] : 20.000 = + 16$$

Simplified method

43. Estimates from the sample of the total number of reporting units in each industry/size group of the universe responding (+), (=) and (-) respectively - labelled $N_{ks}^{*(+)}$, $N_{ks}^{*(=)}$ and $N_{ks}^{*(-)}$ in formula (11) - are computed as follows using the data recorded in Table 3.

Table 6. Calculation of balances by the simplified weighting method

Report unit no	Industry	Size group	Response [x_{ik}]	Sampling probability [f_i]	Adjustment factor to universe level [$1/f_i$]	Response adjusted to universe level [$x_{ik} \times 1/f_i$]	Industry A Size group 1 N^*_{ks}	Industry A Size group 2 N^*_{ks}	Industry B Size group 1 N^*_{ks}	Industry B Size group 2 N^*_{ks}
Col 1	Col 2	Col 3	Col 4	Col 5	Col 6 (1/Col 4)	Col 7	Col 8	Col 9	Col 10	Col 11
1.1	A	1	+ 1	1/1	1	+ 1	1			
1.2	A	2	0	1/1	1	0		1		
1.3	B	1	- 1	1/1	1	- 1			1	
2.1	A	2	0	1/1	1	0		1		
2.2	B	1	+ 1	1/1	1	+ 1			1	
2.3	A	2	- 1	1/1	1	- 1		1		
3.1	A	2	- 1	1/10	10	- 10		10		
4.1	A	2	- 1	1/10	10	- 10		10		
4.2	B	2	+ 1	1/10	10	+ 10				10
5.1	B	1	- 1	1/3	3	- 3			3	
6.1	B	1	0	1/3	3	0			3	
6.2	A	1	- 1	1/3	3	- 3	3			
7.1	B	2	0	1/6	6	0				6
8.1	B	2	+ 1	1/6	6	+ 6				6
9.1	B	2	+ 1	1/6	6	+ 6				6
9.2	A	2	0	1/6	6	0		6		
9.3	A	2	- 1	1/6	6	- 6		6		
10.1	A	2	- 1	1/6	6	- 6		6		
10.2	B	2	- 1	1/6	6	- 6				6
11.1	B	2	+ 1	1/6	6	+ 6				6
SUM	••	••	••	••	93	••	4	41	8	40

•• = Irrelevant information

44. Table 7 is derived from Table 6. The frequencies in absolute numbers presented in Table 7 add up to the SUM for columns 8 to 11 in Table 6.

Table 7. Distribution within report strata of responses to question Q adjusted to universe level . Absolute frequencies.

SUM OF RESPONSES				
REPORT UNIT STRATUM	(+)	(=)	(-)	SUM
Industry A, Size group 1	1	0	3	4
Size group 2	0	8	33	41
Industry B, Size group 1	1	3	4	8
Size group 2	28	6	6	40

45. The absolute numbers in this table are converted to percentages as shown in Table 8.

Table 8. Distribution within report strata of responses to question Q adjusted to universe level . Percentage frequencies

REPORT UNIT STRATUM	SUM OF RESPONSES			SUM	BALANCE [(+) - (-)]
	(+)	(=)	(-)		
Industry A, Size group 1	25	0	75	100	- 50
Size group 2	0	20	80	100	- 60
Industry B, Size group 1	13	37	50	100	- 37
Size group 2	70	15	15	100	+ 55

46. The figures in the cells of the columns (+), (=) and (-) are the A_{ks} values according to formula (11) above expressed as percentages and the figures in the cells of column BALANCE are the B_{ks} values in formula (12). Applying formula (11) for estimating the balances for our two industries and using the numerical data presented in Table 2 we get the information recorded in Table 9.

Table 9. Aggregation of size groups to industries and industries to industry aggregates

INDUSTRY	SIZE GROUP	w_{ks}	VALUE ADDED	B_{ks}	B_k	BV
A	1	2000		- 50		
A	2	2500		- 60		
A	Total	4500	4000 mln \$	- 56		(1)
B	1	5000		- 37		
B	2	3000		+ 55		
B	Total	8000	16000 mln \$	- 3		(1)
A + B	TOTAL	12500	20000 mln \$	- 14		(2)

(1) The balances by industry (B_k) have been computed by applying formula (12) above to the numerical data in the table. Thus:

For industry A we get $B_k = [2000 \times (-50) + 2500 \times (-60)] / 4500 = - 56$

For industry B we get $B_k = [5000 \times (-37) + 3000 \times (+55)] / 8000 = - 3$

(2) The balance for aggregates of industries - i.e. industry A plus industry B - is computed applying formula (9) above to the numerical data in the table as follows:

$B = [4000 \times (-56) + 16000 \times (-3)] / 20000 = - 14$

47. In this simplified numerical example, the three different procedures - size-weighted general method, general method without size-weights and the simplified method - give very different results. Using real-life data, the usual finding is that the choice of weighting procedure has a limited effect on the results.

Recommended weighting and estimation procedure

48. The recommended approach is to use reliable external information to come as close as possible to the correct weighting of size groups and industries for reporting units. When this condition is fulfilled a mix between the simplified and general method should be used. The general method is used to estimate the response distribution at the report stratum level and from there on external information is used as indicated in Table 2. In this manner the error due to estimating the importance of different size groups and of different industries is reduced.