User Guide

Coefficient of Variation Extraction Module

Changes in Employment Survey

Cohorts 1 to 10

Statistics Canada Statistique Canada

Canada

Table of Contents

1.0 Introduction 5
1.1 Description 5
1.2 Conditions of Use 5
1.3 Starting the Module 5
1.4 Overview of the Interface 6
1.4.1 Choosing an Estimator 6
1.4.2 Choosing the Cohort 7
1.4.3 Choosing a Variable to Estimate 8
1.4.4 Specifying a Domain 8
1.4.5 Overwriting and Adding Scores 9
1.4.6 The Query Execution Button 9
1.5 Displaying Results 9
1.5.1 Column Descriptions 9
1.5.2 Colour Coding of Coefficients of Variation 10

2.0 Examples of Using the Coefficient of Variation Extraction Module 11
2.1 Example 1: Searching for a Size Estimate 11
2.2 Example 2: Searching for a Ratio Estimate 13

3.0 Category Specifications 15

1.0 Introduction

1.1 Description

The Changes in Employment Survey (CIES), Coefficient of Variation Extraction Module (CVEM) is a scanning tool which provides a direct estimate of quality for totals and proportions for certain variables and subject fields of general interest.

The CVEM serves in producing an approximation of the coefficient of variation (CV) for a given proportion or total in a given subject field.

It should be noted that CV calculations for this application are based on CIES data. For that reason, the module should not be considered to be a generic tool that could be used for other surveys.

1.2 Conditions of Use

The module is programmed using Excel macros; the dialogue box is programmed with Visual Basic for Applications. A message box appears when the module file is opened (see Figure 1-1), warning the user to exercise caution with files containing macros. The user can assume that the macros are virus free. The module will only function correctly if the user enables the macros by clicking the appropriate button (*Enable Macros*).

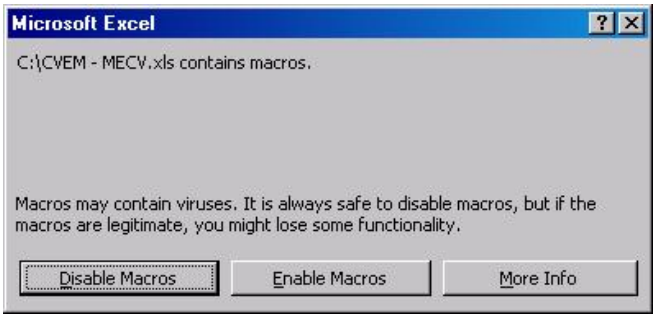


Figure 1-1: Message box at module opening

1.3 Starting the Module

When the module file (*CVEM – MECV.xls*) is opened, only the module's customized toolbar is visible. The usual Excel toolbars are hidden (see Figure 1-2). The toolbar is located in the upper left hand corner of the Excel window and consists of three buttons:

E for the English version of the CVEM,
F for the French version and
? to open the CVEM User Guide.

The user is not required to perform any action on the Excel worksheet itself, but can start the module merely by clicking on the button matching the desired language.

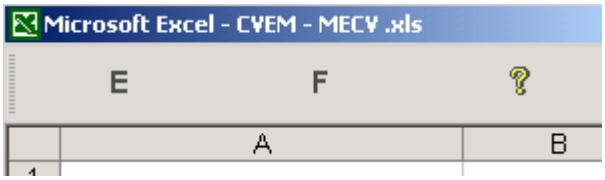


Figure 1-2: CVEM customized toolbar

1.4 Overview of the Interface

1.4.1 Choosing an Estimator

The interface is designed to be user friendly. Users can choose between estimating the CV for the size of the domain or a ratio. Figure 1-3 illustrates the "Size" component, the default when opening the CVEM. Figure 1-4 displays the "Ratio" component. To switch from the "Size" component to the "Ratio" component, the user merely clicks on the "Ratio" button.

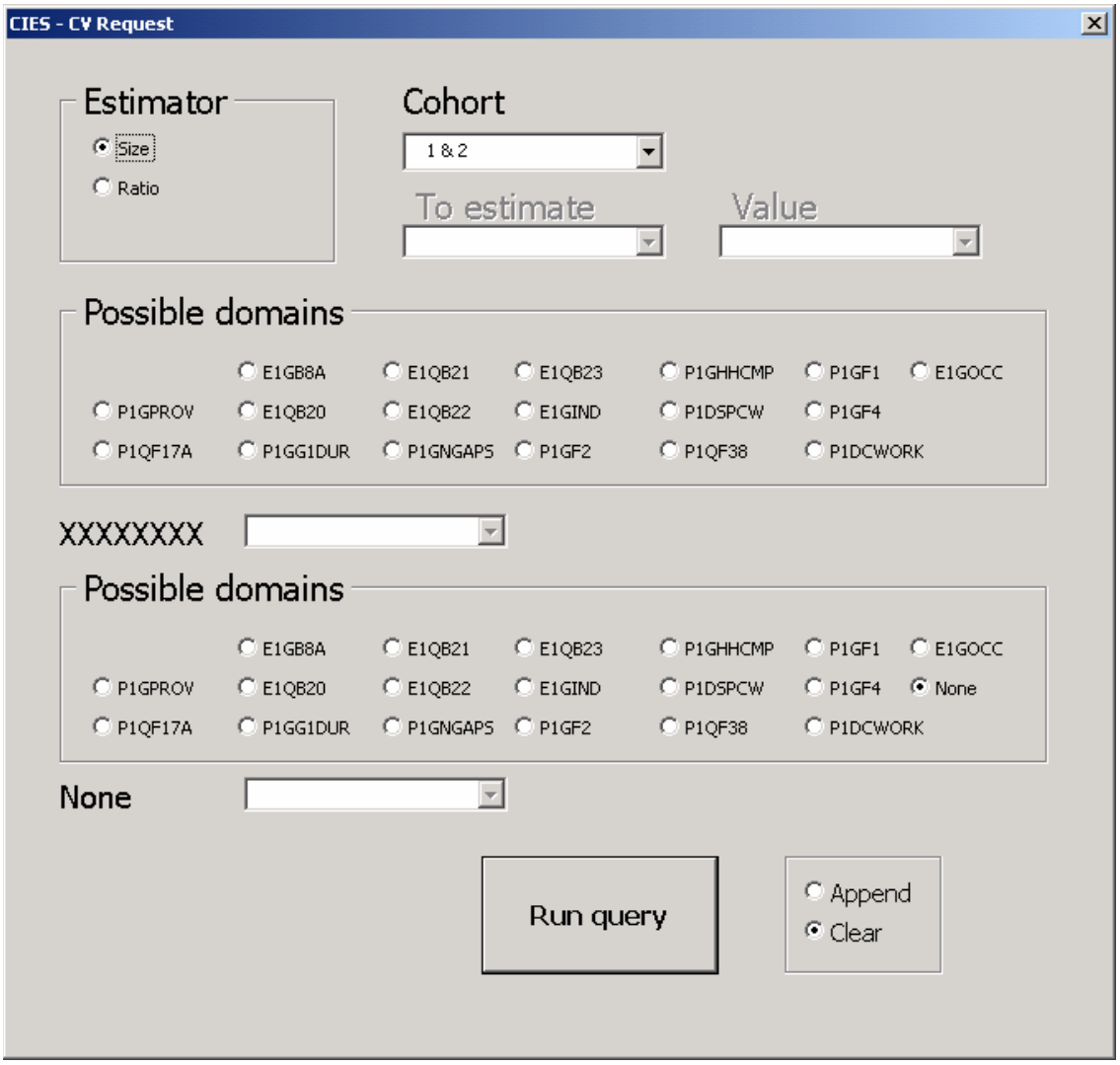


Figure 1-3: The "Size" component of the CV Extraction Module

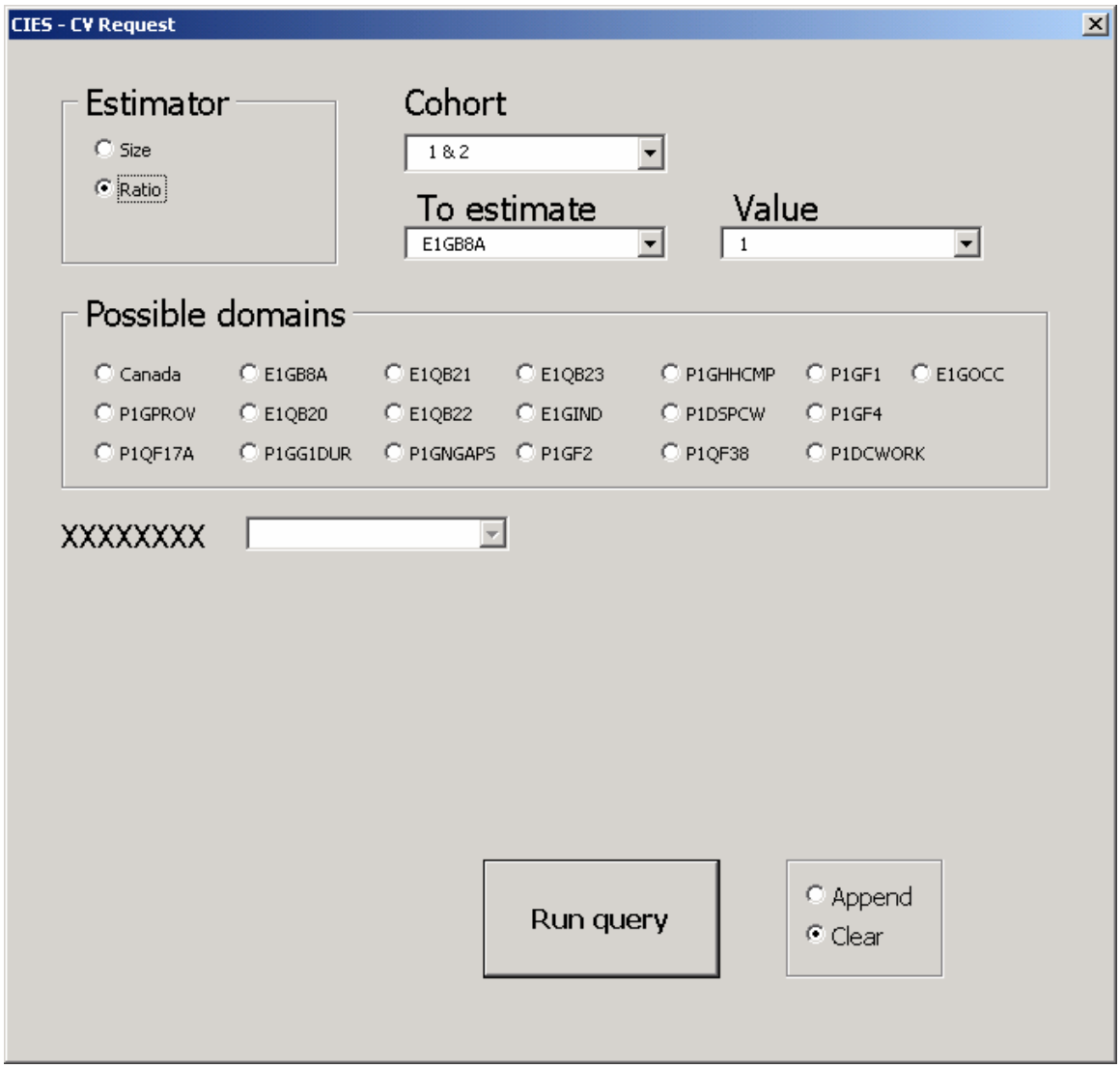


Figure 1-4: The "Ratio" component of the CV Extraction Module

1.4.2 Choosing the Cohort

Whichever estimator has been chosen, the next step is to choose a cohort combination for which estimates are required. By default, Cohort 1 & 2 has been selected. The cohort combinations are:

- Cohort 1 & 2,
- Cohort 3 & 4,
- Cohort 5 & 6,
- Cohort 7 & 8 and
- Cohort 9 & 10

1.4.3 Choosing a Variable to Estimate

If the “Size” component is chosen, the option doesn’t exist to select a “To estimate” variable or a “Value”. These options are only activated when the “Ratio” component has been selected as the estimator (Figure 1-4). When estimating proportions, they indicate the variable and variable values for which coefficients of variation are needed.

Thirty variables have been identified as key variables and are included in the “To estimate” list. The variables are:

E1GB8A	P1DSPCW	P1QF17A	P1QD1
E1QB20	P1GHHCMP	P1GF17B	P1QE1
E1QB21	P1GPROV	P1GF37	P1QE4
E1QB22	P1GF1	P1QF38	P1QF46
E1QB23	P1GF2	P1GF39	P1GG4
E1GIND	P1GF4	P1GG1DUR	P1GG5
P1DJSRCH	P1GF17	P1GNGAPS	P1DCWORK
E1GOCC	P1GVISM*		

* The variable P1GVISM is not available for the Cohort 1 & 2 combination.

By double-clicking on the title “To estimate”, the selected variable’s definition and all its possible values are displayed in a window box.

The user can also select the value for which an estimate of the coefficient of variation is needed by using the “Value” drop-down menu. This list corresponds specifically to the variable selected. By default, the first category for the selected variable is displayed. The “All” option is included to show estimates for all possible values of the selected variable.

1.4.4 Specifying a Domain

Depending on the chosen estimator, different domain possibilities are included for the user. The “Size” component (Figure 1-3) includes two domain possibilities while the “Ratio” component (Figure 1-4) has only one. The domain variables are the same for both components with two exceptions:

1. a “Canada” variable has been added to the “Ratio” component to allow proportion estimates for the entire population, and

2. a “None” variable has been added to the second domain under the “Size” component to allow estimates of coefficient of variation for the size of the domain without crossing any variables.

To select a domain, click on the button to the left of the variable in the “Possible domains” list (E1QB21). The drop-down menu (**XXXXXXXX**) is then activated and the user can choose the desired value for this variable or all possible values by selecting the “All” option. The domains are the following:

P1GPROV	P1GG1DUR	E1QB23	P1DSPCW
P1QF17A	E1QB21	E1GIND	P1QF38
E1GB8A	E1QB22	P1GF2	P1GF1
E1QB20	P1GNGAPS	P1GHHCMP	P1GF4
P1DCWORK	E1GOCC		

All variable definitions and possible values can be seen by double-clicking on the variable name in either of the two domain frames.

1.4.5 Overwriting and Adding Scores

The option of overwriting or adding scores to the output worksheet is available. Once the user clicks on the command button (Run query), the application will search for the requested results and display them in the Excel worksheet. If multiple queries are made, the user can choose between retaining the previous query scores in the worksheet or deleting them. If the user decides to keep them, scores from the next query will appear beneath the preceding scores. The user must choose between these two options by using the “Clear” or “Append” buttons located in the box at the bottom right hand corner (Figure 1-5).

Figure 1-5: Option frame for results display

1.4.6 The Query Execution Button

Once the user has completed the description of the desired subject field and the variable to estimate (if necessary), he/she merely needs to click the “Run query” button (Figure1-6) to execute the query.

Figure 1-6: Query execution command button

1.5 Displaying Results

1.5.1 Column Descriptions

When opening the file for the first time, an empty Excel spreadsheet will be shown.

Once the first query has been executed, many columns will appear. Column A always contains the estimator (Size or Ratio). The last four columns are always the same, that is:

N - The number of records in this field in the sample;

ESTIMATE - The size of this field population or proportion of the population;

STDERROR - Standard error of the estimate;

COEFFVAR - Coefficient of variation of the estimate.

Their location varies according to the estimator and number of domains chosen. These statistics were obtained using Statistics Canada's Generalized Estimation System (GES). By using characteristics of the sample selection, weighting and calibration, the estimate, its standard error and its coefficient of variation were estimated. These results are contained in a database. The CVEM extracts the estimates from this database for the desired fields and variables. The CVEM has been designed for a selected set of domains and variables. Other estimates can be calculated by Statistics Canada on a cost-recovery basis.

When the "Size" estimator is chosen, the second and possibly the third column contain the selected domain variable values. When the "Ratio" estimator is chosen, column B contains the variable to estimate, column C its value and column D the domain variable.

1.5.2 Colour Coding of Coefficients of Variation

Since the coefficient of variation serves as the main indicator of quality, a simple colour code is used in displaying CVs on the Excel worksheet (Figure 1-7). Based on quality level guidelines (Section 9.5 in the CIES Microdata User Guide):

- estimates with a sample size of less than 30 (the "N" variable column), or very high CVs in excess of 33.3% (*unacceptable*) are displayed on a red background;
- estimates with a sample size of 30 or more (the "N" variable column), and high CVs in the range of 16.6% to 33.3% (*marginal*) appear on a yellow background;
- estimates with a sample size of 30 or more (the "N" variable column), and low CVs in the range of 0.0% to16.5% *(acceptable)* are not colour coded.

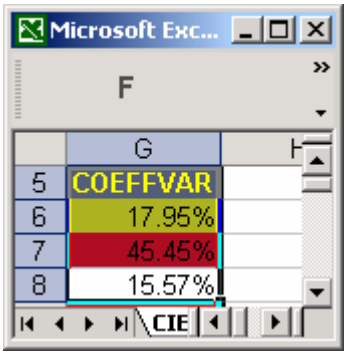


Figure 1-7: Colour coding of coefficients of variation.

Note that for a ratio, the user needs to multiply the estimate by the number of records (the "N" variable column) to obtain an approximation of the number of people in the sample with that particular characteristic.

2.0 Examples of Using the Coefficient of Variation Extraction Module

2.1 Example 1 - Searching for a Size Estimate

Suppose the user is interested in the number of men with a change in employment between July 1995 and December 1995 (Cohort 1 & 2) in every industry. The user first selects the "Size" estimator and Cohort "1 & 2" from the "**Cohort**" drop-down menu. Then he/she defines the subject field in the module by selecting the domain variables E1GIND (industry) and P1GF2 (sex). Since no analysis has yet been performed and the user only wants to observe the quality of the estimates in this subject field, they display the coefficients of variation (CV) for all available proportions by selecting "All" in the drop-down menu to the right of E1GIND. Men are defined in the second subject field by selecting category "1" in the drop-down menu. Figure 2-1 shows how the user arrived at the scores displayed in Figure 2-2.

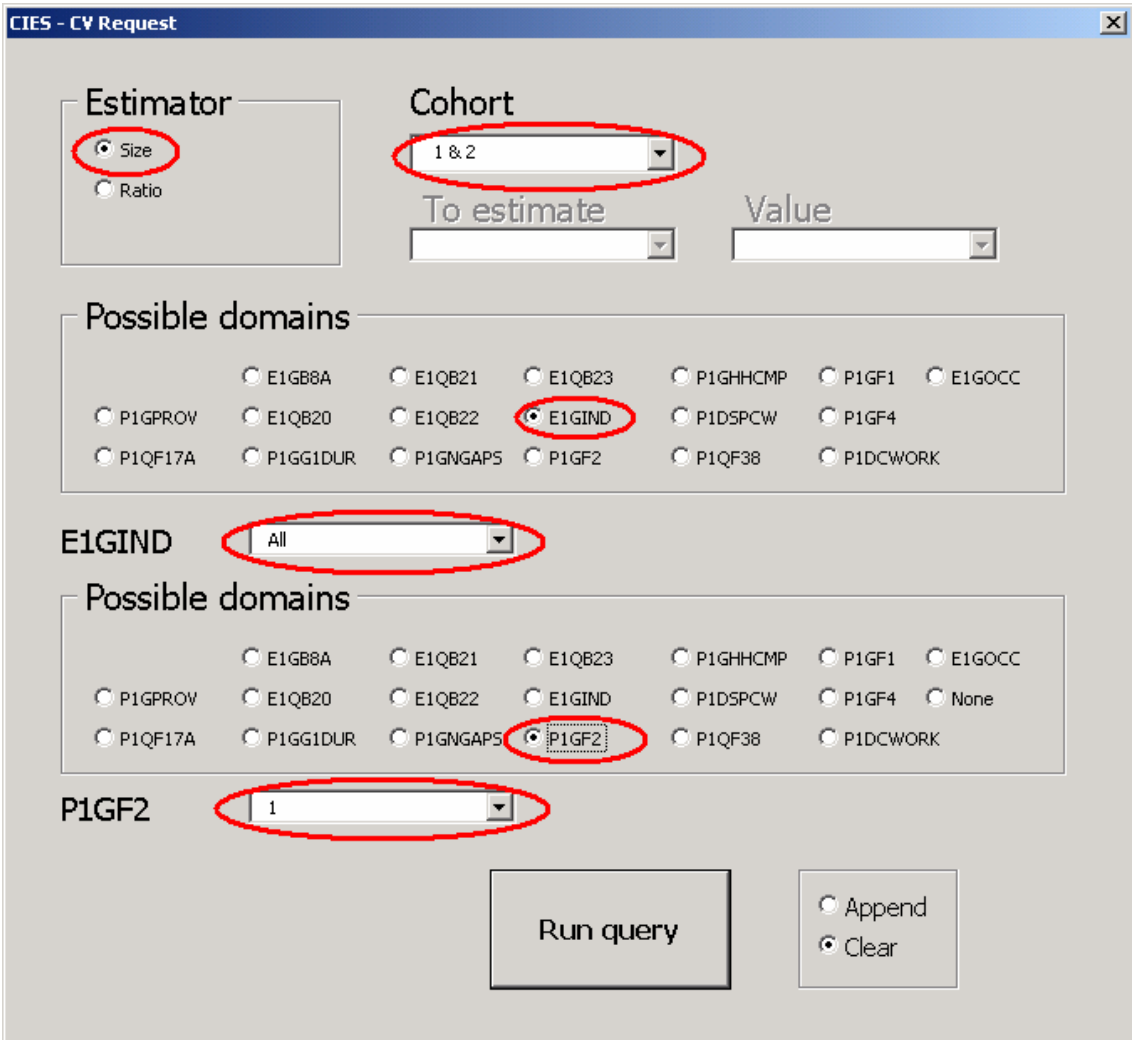
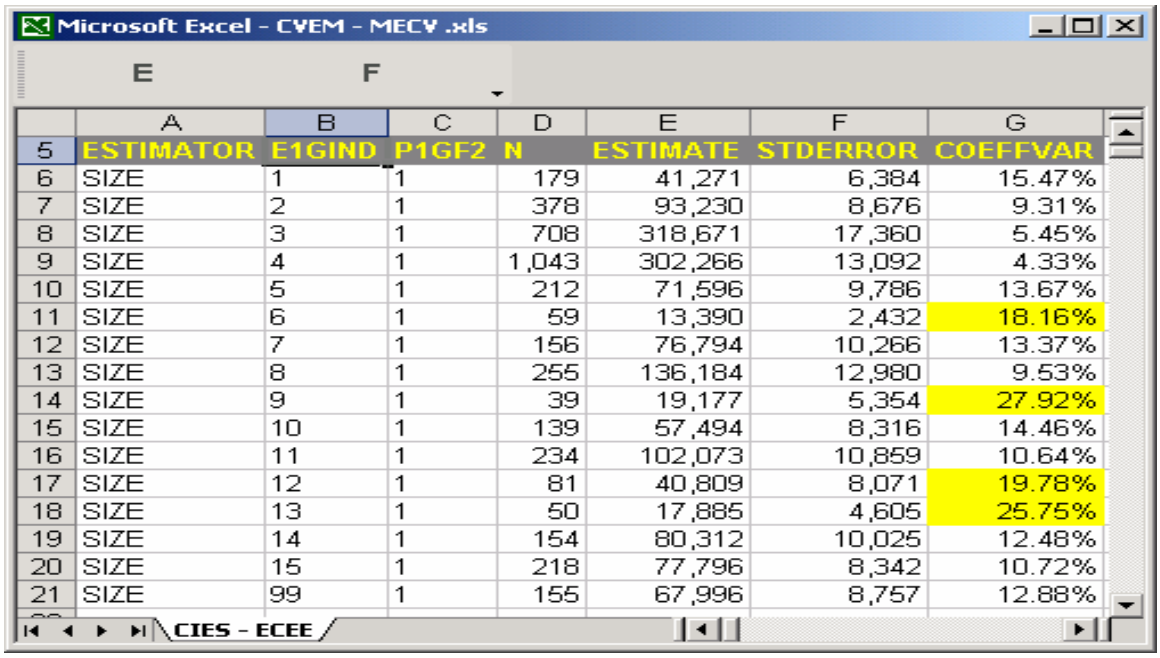


Figure 2-1: Using the module for a size estimate



Microsoft Excel - CVEM - MECV .xls

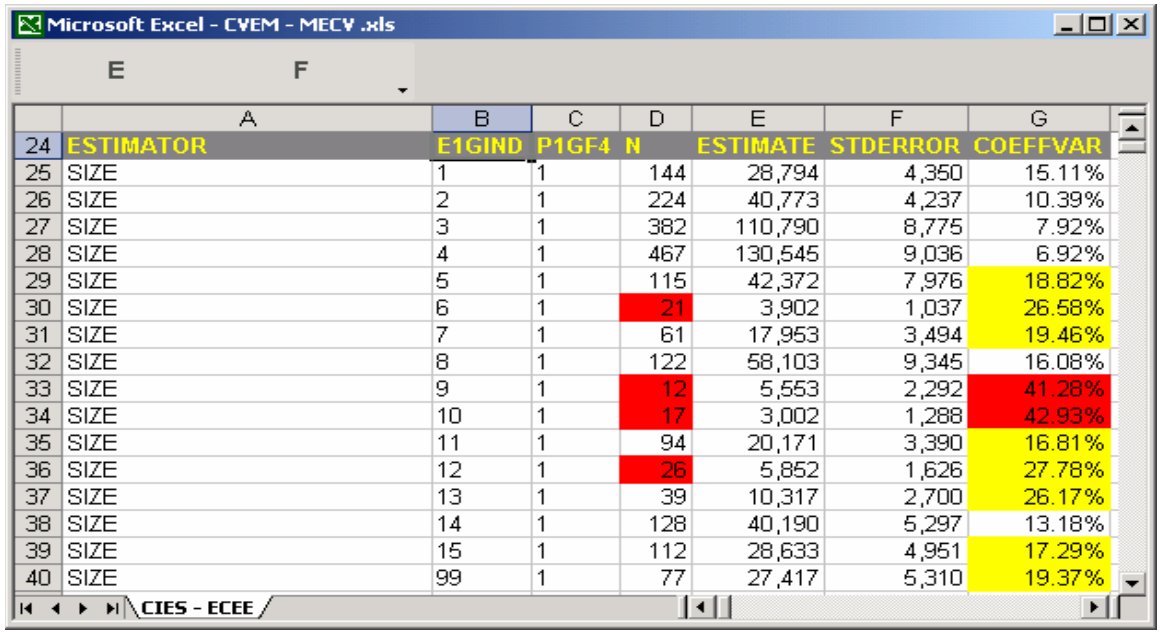
	A	B	C	D	E	F	G
5	ESTIMATOR	E1GIND	P1GF2	N	ESTIMATE	STDERROR	COEFFVAR
6	SIZE	1	1	179	41,271	6,384	15.47%
7	SIZE	2	1	378	93,230	8,676	9.31%
8	SIZE	3	1	708	318,671	17,360	5.45%
9	SIZE	4	1	1,043	302,266	13,092	4.33%
10	SIZE	5	1	212	71,596	9,786	13.67%
11	SIZE	6	1	59	13,390	2,432	18.16%
12	SIZE	7	1	156	76,794	10,266	13.37%
13	SIZE	8	1	255	136,184	12,980	9.53%
14	SIZE	9	1	39	19,177	5,354	27.92%
15	SIZE	10	1	139	57,494	8,316	14.46%
16	SIZE	11	1	234	102,073	10,859	10.64%
17	SIZE	12	1	81	40,809	8,071	19.78%
18	SIZE	13	1	50	17,885	4,605	25.75%
19	SIZE	14	1	154	80,312	10,025	12.48%
20	SIZE	15	1	218	77,796	8,342	10.72%
21	SIZE	99	1	155	67,996	8,757	12.88%

CIES - ECEE

Figure 2-2: Example 1 query results

This subject field (E1GIND) can clearly provide good estimates but the user must be cautious with industries 6, 9, 12 and 13.

If the user wishes to perform the same kind of analysis for the 15 to 24 age group (by selecting P1GF4 and the value 1), it becomes clear that the estimates are unreliable for industry 6 (Figure 2-3). If the user is interested in a subset of this subject field, he/she cannot expect to obtain quality estimates. Furthermore, industries 9, 10 and 12 also have sample sizes too small to lead to quality estimates.



Microsoft Excel - CVEM - MECV .xls

	A	B	C	D	E	F	G
24	ESTIMATOR	E1GIND	P1GF4	N	ESTIMATE	STDERROR	COEFFVAR
25	SIZE	1	1	144	28,794	4,350	15.11%
26	SIZE	2	1	224	40,773	4,237	10.39%
27	SIZE	3	1	382	110,790	8,775	7.92%
28	SIZE	4	1	467	130,545	9,036	6.92%
29	SIZE	5	1	115	42,372	7,976	18.82%
30	SIZE	6	1	21	3,902	1,037	26.58%
31	SIZE	7	1	61	17,953	3,494	19.46%
32	SIZE	8	1	122	58,103	9,345	16.08%
33	SIZE	9	1	12	5,553	2,292	41.28%
34	SIZE	10	1	17	3,002	1,288	42.93%
35	SIZE	11	1	94	20,171	3,390	16.81%
36	SIZE	12	1	26	5,852	1,626	27.78%
37	SIZE	13	1	39	10,317	2,700	26.17%
38	SIZE	14	1	128	40,190	5,297	13.18%
39	SIZE	15	1	112	28,633	4,951	17.29%
40	SIZE	99	1	77	27,417	5,310	19.37%

CIES - ECEE

Figure 2-3: Query results of the variant of Example 1

2.2 Example 2 - Searching for a Ratio Estimate

Suppose the user is interested in knowing the proportion of men with a change in employment between July 1995 and December 1995 (Cohort 1 & 2) in every industry. First, select the "Ratio" estimator and Cohort "1 & 2" from the "**Cohort**" drop-down menu. The "**To estimate**" and "**Value**" boxes are now activated. Now choose the variable P1GF2 (sex) from the "**To estimate**" drop-down menu and category 1 from the "**Value**" drop-down menu. Then define the subject field in the module by selecting the domain variable E1GIND (industry). To display the CVs for all available proportions select "All" in the drop-down menu to the right of E1GIND and click on the "**Run query**" button. The resulting scores are shown in Figure 2-5.

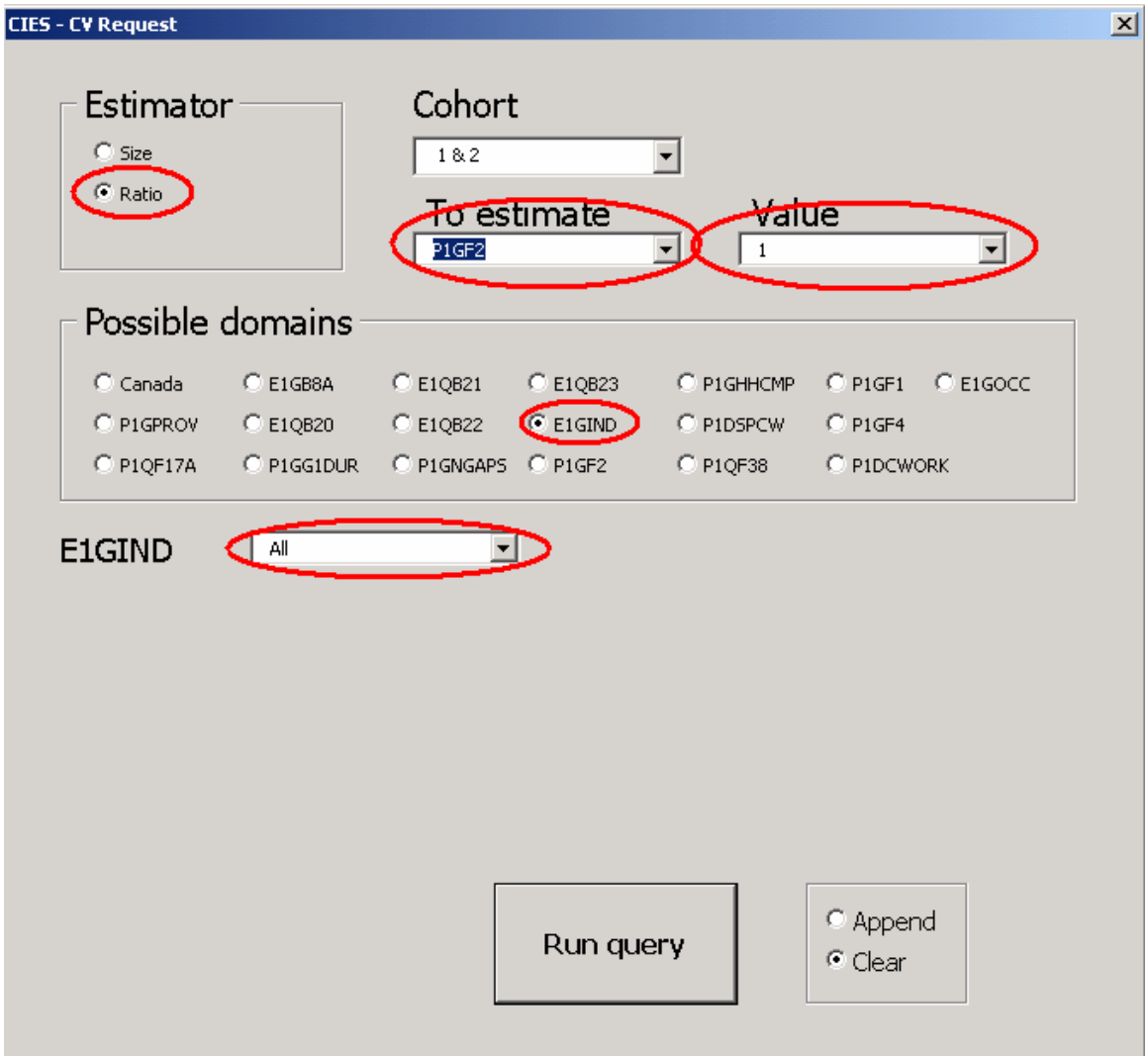
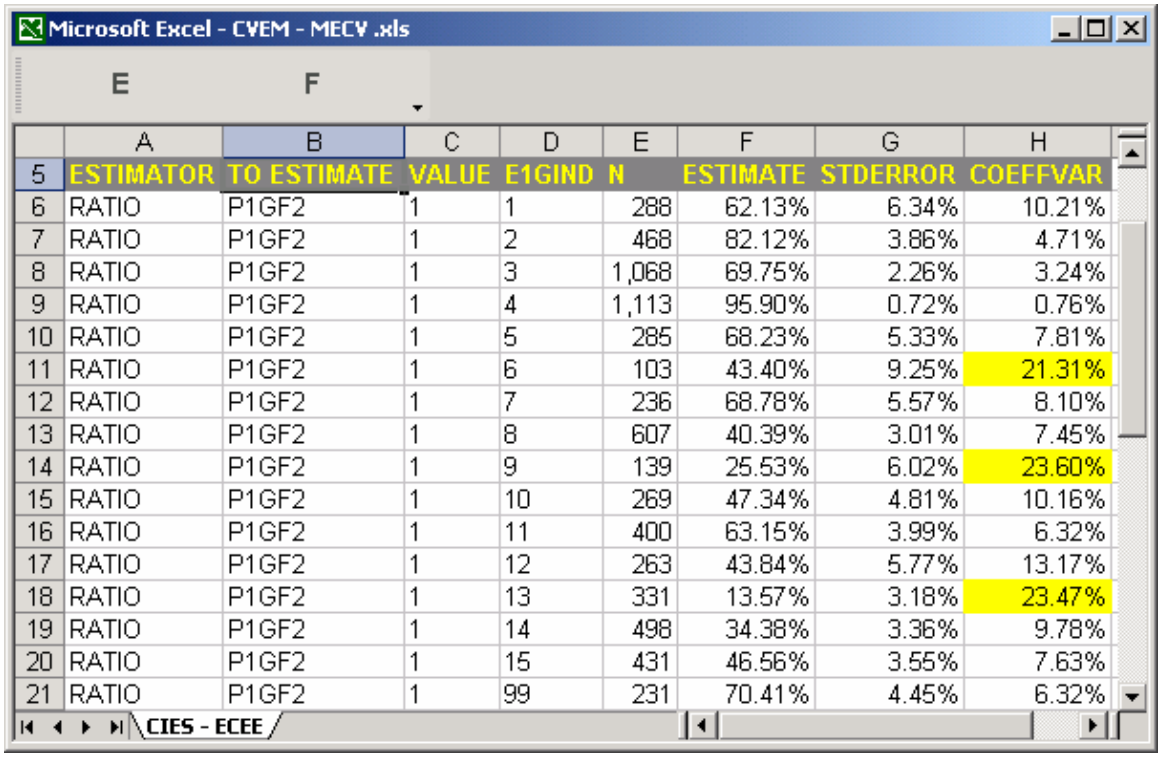


Figure 2-4: Using the module for a ratio estimate



Microsoft Excel - CVEM - MECV .xls

	A	B	C	D	E	F	G	H
5	ESTIMATOR	TO ESTIMATE	VALUE	E1GIND	N	ESTIMATE	STDERROR	COEFFVAR
6	RATIO	P1GF2	1	1	288	62.13%	6.34%	10.21%
7	RATIO	P1GF2	1	2	468	82.12%	3.86%	4.71%
8	RATIO	P1GF2	1	3	1,068	69.75%	2.26%	3.24%
9	RATIO	P1GF2	1	4	1,113	95.90%	0.72%	0.76%
10	RATIO	P1GF2	1	5	285	68.23%	5.33%	7.81%
11	RATIO	P1GF2	1	6	103	43.40%	9.25%	21.31%
12	RATIO	P1GF2	1	7	236	68.78%	5.57%	8.10%
13	RATIO	P1GF2	1	8	607	40.39%	3.01%	7.45%
14	RATIO	P1GF2	1	9	139	25.53%	6.02%	23.60%
15	RATIO	P1GF2	1	10	269	47.34%	4.81%	10.16%
16	RATIO	P1GF2	1	11	400	63.15%	3.99%	6.32%
17	RATIO	P1GF2	1	12	263	43.84%	5.77%	13.17%
18	RATIO	P1GF2	1	13	331	13.57%	3.18%	23.47%
19	RATIO	P1GF2	1	14	498	34.38%	3.36%	9.78%
20	RATIO	P1GF2	1	15	431	46.56%	3.55%	7.63%
21	RATIO	P1GF2	1	99	231	70.41%	4.45%	6.32%

CIES - ECEE

Figure 2-5: Example 2 query results

Note the ”TO ESTIMATE” and “VALUE” columns in the query results table. The user should also note that for the variable P1GF2 = 1 (Male), the sample sizes (N) for the subject field (E1GIND) are all greater than 30. To obtain the approximate number of men in that domain, multiply the variable value in the “N” column by the value in the “ESTIMATE” column (for E1GIND = 1, the number of men would be 288 * 62.13% = 179). When possible, use the size estimates to obtain the precise sample size of the domain defined by crossing the variables.

3.0 Category Specifications

To ensure that scores are properly compared, users should check that the categories they use correspond to those used in developing the Coefficient of Variation Extraction Module (CVEM). Double-clicking on the title “To estimate” when a variable has been selected shows the variable definition and possible values. Double-clicking on variables in either domain frame will give the same results.