

User Guide to LAPOP's System for Online Data Analysis (v1)

The following guide is designed to help users navigate through and use LAPOP's System for Online Data Analysis (SODA). The system allows individuals to query and analyze LAPOP survey data free of charge, with no registration requirement.

Introduction

LAPOP's SODA provides free online access to over 100 surveys conducted by LAPOP since the early 1990s (and some earlier surveys). The site supports interactive analyses of data from individual country surveys and, as well, of individual rounds of the *AmericasBarometer* survey. In 2010, LAPOP conducted its largest round of the *AmericasBarometer* survey, which involved national samples of voting age adults in 26 countries: Argentina, Belize, Bolivia, Brazil, Canada, Colombia, Costa Rica, Chile, Dominican Republic, Ecuador, El Salvador, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Trinidad & Tobago, Suriname, the United States, Uruguay, and Venezuela. LAPOP surveys focus on themes relevant to democratic politics and cover topics such as system support, political tolerance, citizen participation, local government, corruption, and authoritarianism.

The data analysis website is maintained by the Centro Centroamericano de Población at the University of Costa Rica. Underlying the user-friendly interface is statistical software that allows an interactive experience. Rather than call up pre-generated charts, the system responds in real time to the user's queries. The system allows anyone around the world with internet access to query and analyze LAPOP data free of charge and with no registration requirement.

LAPOP developed this handbook to assist users of the interactive website. We hope these instructions for how to access and use the website will prove useful to individuals and organizations. We also recommend the system to academic faculty as a way to integrate data analysis into classroom activities and assignments. If you are using this handbook, and find errors or ways we could improve it, please email your question or suggestion to liz.zechmeister@vanderbilt.edu.

The remainder of this handbook is divided into the following sections: Getting Started (p. 2); Beginner Mode (pp. 3-5); Expert Mode (pp. 5-11).

Getting Started

To access LAPOP's free online data analysis system, go to the following website¹:

http://lapop.ccp.ucr.ac.cr/Lapop_English.html

You can also access the online data analysis system by going to

<http://www.vanderbilt.edu/lapop/> and, under "Survey Data", click on "Free Online Data Analysis."

Before accessing the database, you might want to familiarize yourself with the questionnaire from the survey in which you are interested. **All questionnaires (as well as technical information about the study, including the sample) for LAPOP surveys are available on the LAPOP website.**

The homepage for LAPOP's online querying system allows you to access two versions of the online data analysis system: beginner and expert. The "beginner" version of the online data analysis system allows the user to obtain simple summary statistics for variables of interest, while the "expert" version allows the user to employ more complicated statistical analyses, such as regressions. Additionally, the "expert" option allows the user to analyze variables of interest by defining filters that divide the sample into subclasses, recode variables for individual research purposes, and run analyses with multiple variables.

The website also provides links to other important resources. First, you can access the website for Centro Centroamericano de Población from the University of Costa Rica, which is in charge of the management of LAPOP's online data analysis system. Second, you can access the websites for the sponsors of LAPOP's survey data collection: USAID, Vanderbilt University, the Inter-American Development Bank, and the United Nations Development Programme. In addition, the homepage includes direct links to LAPOP and the AmericasBarometer's homepage. Third, you can download a PDF copy of LAPOP's 2010 regional report on the political culture of democracy in the Americas.

To begin the process of querying and analyzing LAPOP survey data, select to enter either the Beginner or Expert Portal.

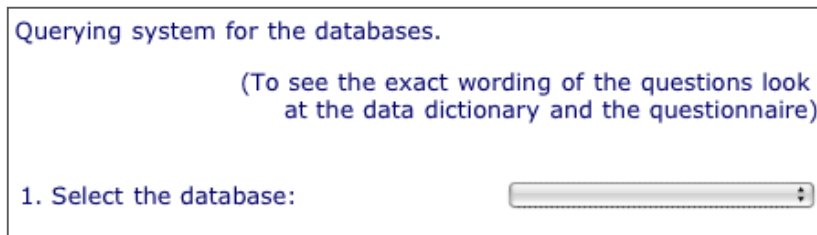
The next pages provide instructions on some of the ways in which users can interact with the system for online data analysis via the Beginner and Expert modes.

¹ Please note that the website runs best on Internet Explorer but will also run using Mozilla Firefox and might accommodate other web browsers.

Beginner Mode

To access the beginner mode click on “Beginner” on the homepage referenced above. The beginner version of the querying system consists of a series of options that the user can follow to examine frequencies/distributions, means, and cross tabulations of variables of interest. In order to better explain how to use this version of the online data analysis system, we will walk through an example query that focuses on presidential approval in the Americas in 2010.

Step 1. *Select the database of interest.* Select the database from the drop-down menu. In our example, we select “AmericasBarometer 2010.”

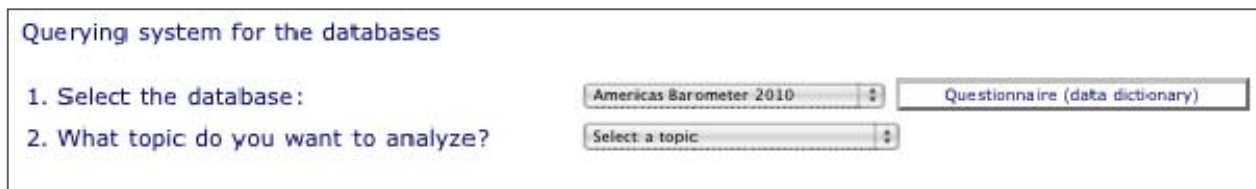


Querying system for the databases.

(To see the exact wording of the questions look at the data dictionary and the questionnaire)

1. Select the database:

Step 2. *Select the topic you want to analyze.* Variables are arranged according to eighteen different topics. If you have trouble finding the variable in which you are interested, skim through the topics you think are most relevant and this will likely lead you to the topic that contains the particular question in which you are interested. In our example, our topic of interest is “10. Political System Support.”



Querying system for the databases

1. Select the database:

2. What topic do you want to analyze?

Notice that a questionnaire/data dictionary button will now appear next to our database selection. Clicking on this button brings up a new window in which variables (survey items/questions) are described.²

Step 3. *Select the variable of interest.* From the menu, select the specific survey question for which you are interested in obtaining information. The menu contains both the question code (number) and the first part of the question wording (the full question can be seen in the questionnaire; see the note under Getting Start on accessing questionnaires). In our example, our variable of interest is “(m1) Speaking in general of the current administration, how would you rate the job performance of President NAME CURRENT PRESIDENT?” Once you select the variable of interest, the system automatically opens a new tab in your browser (titled “Query results...”) that displays the frequency statistics of the variable.

² Country questionnaire are also available and can be accessed by going to <http://www.vanderbilt.edu/lapop/core-surveys.php>

Querying system for the databases

1. Select the database: Americas Barometer 2010 Questionnaire (data dictionary)
2. What topic do you want to analyze? 10. Political System Support
3. Select the variable: (m1) speaking in general of the current administration/government, how would you rate
4. What process do you want to run? Frequency Mean Cross Tabulation

Step 4. *Select a mode of analysis.* As mentioned above, the system automatically provides the frequency data for the variable. To obtain the mean statistic for the variable of interest, select “Mean.” Now, the “Query results...” table will display the variable’s number of observations (respondents), mean, standard deviation, minimum value, and maximum value. To obtain a cross-tabulation with another variable of interest, click “Cross Tabulation.” This brings up another set of steps (5 and 6) for which you must select the topic and variable of interest for the cross-tabulation. For our example, say we are interested in a cross-tabulation using presidential approval and evaluation of the country’s economic situation.

Querying system for the databases

1. Select the database: Americas Barometer 2010 Questionnaire (data dictionary)
2. What topic do you want to analyze? 10. Political System Support
3. Select the variable: (m1) speaking in general of the current administration/government, how would you rate
4. What process do you want to run? Frequency Mean Cross Tabulation
5. Topic for cross-tabulation: Select a topic

Step 5. *Select topic of interest for cross-tabulation.* As in step 2, select the topic that contains your variable of interest. In our example, our topic of interest is “4. Economic.”

Querying system for the databases

1. Select the database: Americas Barometer 2010 Questionnaire (data dictionary)
2. What topic do you want to analyze? 10. Political System Support
3. Select the variable: (m1) speaking in general of the current administration/government, how would you rate
4. What process do you want to run? Frequency Mean Cross Tabulation
5. Topic for cross-tabulation: 4. Economic
6. Select the variable for cross-tabulation: Selecciona una variable

Step 6. *Select variable of interest for cross-tabulation.* As in step 3, select the specific survey question in which you are interested. In our example, our variable of interest is “(soct1) How would you describe the country’s economic situation? Would you say that is it very good, good, neither good nor bad, bad or very bad?” Once you select the variable, the system calculates the cross-tabulation and displays the results in the “Query results...” tab or in a split screen.

Querying system for the databases

1. Select the database: Americas Barometer 2010
2. What topic do you want to analyze? 10. Political System Support
3. Select the variable: (m1) speaking in general of the current administration/government, how would you rate
4. What process do you want to run? Frecuency Mean Cross Tabulation
5. Topic for cross-tabulation 4. Economic
6. Select the variable for cross-tabulation (isoc1) how would you describe the country's economic situation? would you say that it i

Once you have analyzed the data using one or more of the above methods, you might want to display the data in a more visually-interesting formatting. This can be done in a number of ways, for example by cutting and pasting the data into Excel. Paste the data into Excel, use the mouse to select Column A (or whatever column the data have been pasted into) and then select the “Text to Columns” option under the Data tab. If you choose “Fixed Width”, you will be able to adjust the lines for the columns to separate the data into columns. Once you have done this, you can proceed to create bar charts, pie graphs, etc.

Expert Mode

To access the expert mode click on “Expert” on the homepage referenced above. The website for the expert version of the querying system allows users to conduct a variety of different analyses including simple tabulations and linear and logistic regressions. Additionally, the interface allows a number of options for recoding variables and modifying the presentation of results in a table format.

Tabulation:

Step 1. *Select the database of interest.* As an example, click on the menu provided and select “AmericasBarometer 2010.”

Step 2. *Select “Tabulation” as the statistical analysis (process) of interest.* Users can tabulate a single variable, cross-tabulate two variables, or cross-tabulate two variables and a control variable. Users may also summarize continuous variables to obtain informative summary statistics, and “filter” the calculation of results (tabulation and summary statistics) for a specific sub-sample. All results will be weighted according to the sample design.

The screenshot shows a web interface for selecting a database and a process. At the top, there is a dropdown menu for "Select the database:" with "Americas Barometer 2010" selected. Below it is another dropdown menu for "Select the process:" with "Tabulation" selected. In the center, there is a table with five rows and two columns. The first column contains buttons labeled "ROW", "COLUMN", "CONTROL", "SUMMARIZE", and "FILTER". The second column contains empty text input fields. Below the table, there is explanatory text: "Row" "Column" and "Control" buttons are for selecting the variables you want to tabulate. "Summarize" is for selecting continuous variables you want summary statistics (averages). "Filter" button is for selecting a sub-class. The resulting table will contains information for this sub-sample. For further information, click on "Tutorial" below. All results will be weighted according the sample design. In most surveys weighting has no consequences since the fact the sample is self-weighted. At the bottom of the interface, there are four buttons: "Process", "Reset", "Questionnaire (data dictionary)", and "Help".

To tabulate a single variable, click on “Row.” This will redirect you to the “Assistant for expression building” where you can select and modify the variable of interest. For example, in the assistant select “q1” (sex) on the variables menu. Then click accept for the variable to be selected.

The screenshot shows the "Assistant for expression building" interface. At the top, there is a text input field containing "q1". Below it, there is a dropdown menu for "Variables" and a dropdown menu for "Operator". Below these, there is a dropdown menu for "Val". At the bottom, there is a row of buttons: "(", ")", "AND", "OR", "1", "2", "3", "4", "5", "6", "7", "8", "9", "0". Below this row, there are four buttons: "Recode", "Reset", "Accept", and "Cancel".

Click “Process.” Results are displayed in a new window. New results will replace old results.

Select the database: Americas Barometer 2010

Select the process: Tabulation

ROW	a1
COLUMN	
CONTROL	
SUMMARIZE	
FILTER	

"Row" "Column" and "Control" buttons are for selecting the variables you want to tabulate.
"Summarize" is for selecting continuous variables you want summary statistics (averages).
"Filter" button is for selecting a sub-class. The resulting table will contains information for this sub-sample.
For further information, click on "Tutorial" below.
All results will be weighted according the sample design. In most surveys weighting has no consequences since the fact the sample is self-weighted.

Process Reset Questionnaire (data dictionary) Help

Notice that a questionnaire/data dictionary button is available when you are selecting/running an analysis (or “process”). Clicking on this button brings up a new window in which variables (survey items/questions) are described.

Also notice that a “Reset” button will always be available to clear variable options.

To cross-tabulate two variables, click on “ROW.” This will redirect you to the “Assistant for expression building” where you can select a variable of interest. Following the examples used in the explanation for using the beginner mode, select your variable of interest, for example “soct1” (national economic situation), and click “Accept”. Once redirected back to the process menu, click on “COLUMN.” This will again redirect you to the “Assistant for expression building” where you can select another variable of interest. For example, select “m1” (presidential approval) and click “Accept”. Click “Process.” Statistical results for the cross-tabulation perceptions of the two variables (in this case, national economy and presidential approval) are displayed in a new window.

Select the database: Americas Barometer 2010

Select the process: Tabulation

ROW	soct1
COLUMN	m1
CONTROL	
SUMMARIZE	
FILTER	

"Row" "Column" and "Control" buttons are for selecting the variables you want to tabulate.
"Summarize" is for selecting continuous variables you want summary statistics (averages).
"Filter" button is for selecting a sub-class. The resulting table will contains information for this sub-sample.
For further information, click on "Tutorial" below.
All results will be weighted according the sample design. In most surveys weighting has no consequences since the fact the sample is self-weighted.

Process Reset Questionnaire (data dictionary) Help

To cross-tabulate two variables and a control variable, click on “ROW.” This will redirect you to the “Assistant for expression building” where you can select a variable of interest. Following the examples used in the explanation for using the beginner mode, select “soct1” (national economic situation) and click “Accept”. Once redirected back to the process menu, click on “COLUMN.” This will again redirect you to the “Assistant for expression building” where you can select another variable of interest. Select “m1” (presidential approval) and click “Accept”. Once redirected back to the process menu, click on “CONTROL.” As an example, select q1 (sex) as the control variable of interest. Click “Process.” Statistical results for the cross-tabulation perceptions of the national economy and presidential approval by sex are displayed in a new window.

Select the database: Americas Barometer 2010

Select the process: Tabulation

ROW	soct1
COLUMN	m1
CONTROL	q1
SUMMARIZE	
FILTER	

"Row" "Column" and "Control" buttons are for selecting the variables you want to tabulate.
 "Summarize" is for selecting continuous variables you want summary statistics (averages).
 "Filter" button is for selecting a sub-class. The resulting table will contain information for this sub-sample.
 For further information, click on "Tutorial" below.
 All results will be weighted according to the sample design. In most surveys weighting has no consequences since the fact the sample is self-weighted.

Process Reset Questionnaire (datadictionary) Help

When conducting tabulations, users can also obtain summary statistics for a particular variable and/or restrict the analyses to a specific sub-sample. For example, after selecting the “ROW”, click “SUMMARIZE” to select a continuous variable for which you want summary statistics. Then, click “Process.” Results will be displayed in a new window. Note that the “SUMMARIZE” option can only be performed when selecting only when selecting “ROW” variable. Note that the variable in “ROW” has to be the same in the “SUMMARIZE” field; otherwise, the result page mixes the labels for both questions.

To restrict the tabulation analyses to a specific sample, after selecting the “ROW”, “COLUMN” and/or “CONTROL”, click “FILTER” to select a variable for a sub-class. For example, select “ur” (urban/rural) and use the “Operator” and “Value” buttons to select an urban sub-sample. Under “Operator” select “==”, and under “Value” select “urbano 1”.³ Click “Accept”, and once directed back to the process menu, click “Process.” Statistical results for the tabulations analyses will be displayed only for those living in urban areas. See tables that follow.

³Further information on how to modify the value of variables will be explained below.

Assistant for expression building

Select the database:

Select the process:

ROW	sort1
COLUMN	m1
CONTROL	q1
SUMMARIZE	
FILTER	ur==1

"Row" "Column" and "Control" buttons are for selecting the variables you want to tabulate.
 "Summarize" is for selecting continuous variables you want summary statistics (averages).
 "Filter" button is for selecting a sub-class. The resulting table will contain information for this sub-sample.
 For further information, click on "Tutorial" below.
 All results will be weighted according to the sample design. In most surveys weighting has no consequences since the fact the sample is self-weighted.

To modify how tabulation results are presented, scroll down to "Options." Here you can label the "Process title," indicate whether you want results to display missing values, indicate whether and what kind of percentages you want the results to show, select whether you want results to be labeled according to question wording or coding, and select the type of summary statistic you want the tabulation analyses to show (frequencies, means, and/or standard deviations).

Options

Process title	<input type="text"/>
Missing values	<input type="text" value="without missing values"/>
Percentages	<input type="text" value="Without percentages"/>
Labels	<input type="text" value="Labels"/>
Cell's content	<input checked="" type="checkbox"/> Frequencies <input checked="" type="checkbox"/> Mean <input checked="" type="checkbox"/> Standard deviation

OLS (linear) Regression:

Step 1. *Select the database of interest.* As an example, click on the menu provided and select “AmericasBarometer 2010.”

Step 2. *Select “OLS Regression” as the statistical analysis (process) of interest.* Click on “Y” to select a continuous dependent variable. This will redirect you to the “Assistant for expression building.” To illustrate with the earlier example regarding presidential approval, select “m1” (or your variable of interest) as the dependent variable and click “Accept.”⁴ Now, click on “X’s Categorical” to select a categorical independent variable (this assumes you have a categorical variable to add; if not, proceed to selecting a continuous X). For our example, select “soct1” (perceptions of national economic situation) and click “Accept.” You can now click “Process” to calculate this simple two-variable regression or select another independent variable. If you want to add another independent variable to the analysis, you can either click again on “X’s Categorical” or, to add a continuous variable click on “X’s Continuous.” Notice that you can only select a continuous variable here. For our example, select “ed” (years of education) and click “Accept”. Once you finish adding variables, click “Process” to run the regression analyses. Statistical results for how perceptions of the national economy and years of education affect levels of presidential approval will be displayed in a new window.

Select the database:

Select the process:

<input type="button" value="Y"/>	<input type="text" value="m1"/>
<input type="button" value="X's Categorical"/>	<input type="text" value="i,soct1"/>
<input type="button" value="X's Continuous"/>	<input type="text"/>
<input type="button" value="FILTER"/>	<input type="text"/>

Click the “Y” bottom to select the dependent variable. Click the “X’s” bottoms to select the independent variables (either categorical or metric variables).
Click the “Filter” bottom to select a subsample.
For further information, click on “Help” or “Example”.
All results will be weighted according to the sampling design.
In most surveys, this has no consequences since they're self-weighted.

As with the tabulation analyses, the user can also filter the OLS regression results. To restrict the regression analysis to a sub-sample, click on “FILTER” to select a sub-sample variable. For example, select “ur” (urban/rural) and use the “Operator” and “Value” buttons to select an urban sub-sample. Under “Operator” select “==”, and under “Value” select

⁴ As it is coded in the questionnaire, “m1” is technically an ordered categorical variable ranging from 1 to 5. However, for simplicity purposes and ease of explanations, this example assumes that the variable can be interpreted as continuous.

“urbano 1”. Click “Accept” and then click “Process.” Results for how perceptions of the national economy and years of education affect levels of urban residents’ presidential approval will be displayed in a new window.

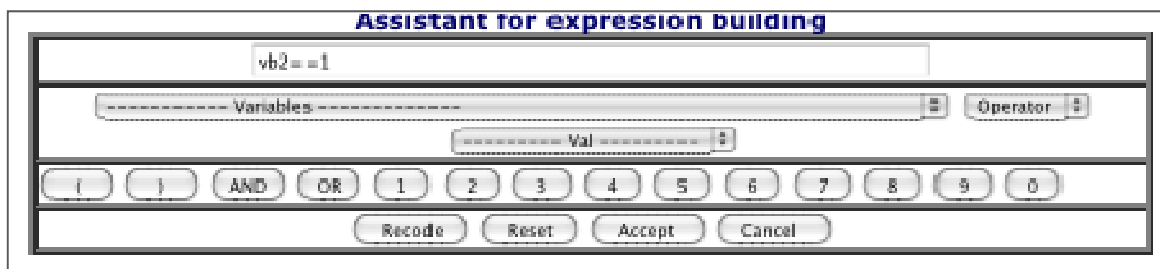
Logistic Regression:

Step 1. *Select the database of interest.* As an example, click on the menu provided and select “AmericasBarometer 2010.”

Step 2. *Select “Logistic Regression” as the statistical analysis (process) of interest.*⁵ Click on “Y” to select a dichotomous dependent variable. This will redirect you to the “Assistant for expression building.” From here, the steps to be followed are the same as those for conducting an OLS regression.

More on Selecting Variables:

When using the “Assistant for expression building”, you can use the operator and values options to specify or restrict the values of your variables of interest. For example, if you wanted to include a control for or filter the analyses only for people who voted in the previous election, you would select variable “vb2” and restrict this variable to include only those who did vote. To do this, click select “==” under “Operator” and “voted 1” under “Value.”



Notice there are various operator options you can use to restrict the value of your variable. Also, you can use the number buttons to assign values to variables.

Recoding Variables:

LAPOP’s online data analysis website also allows users to define recoding rules for the variables selected on “ROW”, “COLUMN”, and “CONTROL.” After you select a variable, click on “RECODE”. This will automatically bring you to the recoding assistant. The recoding program only allows users to recode variables into new categorical ones; therefore, you

⁵ Please note that some variables with answer categories “yes/no” are appropriate for logistic regression but must first be recoded to a 1, 0 dichotomous variables. This is because yes/no variables tend to be coded 1=yes, 2=no.

will need to select ranges for each new category in the recoding process. Ranges include: Low value ("FROM"), high value ("TO"), new value ("VALUE") and a label ("LABEL"). For example, if you want to recode the variable age (q2) in three new categories ("Young", "Adults", and "Elderly"), you will need to type the following ranges:

FROM 18 TO 35 VALUE 1 LABEL Young (click "Include")
FROM 36 TO 64 VALUE 2 LABEL Adults (click "Include")
FROM 65 TO 99 VALUE 3 LABEL Elderly (click "Include")

Click "Accept" when you are done recoding. This will redirect you to the assistant for expression building, where you would again click "Accept."

FROM	TO	VALUE	LABEL
0	0	0	

To change the coding of a category, select the category, click "Change" and recode. To clear a category, select the category and click "Exclude." When you are done recoding click "ACCEPT" and the system will take you back to the tabulation/regression assistant.