

The Dominican Republic 2010 Enterprise Surveys Data Set

I. Introduction

1. This document provides additional information on the data collected in the Dominican Republic between March 2011 and September 2011 as part of the Latin America and Caribbean (LAC) Enterprise Survey 2010, an initiative of the World Bank.

The Enterprise Surveys, through interviews with firms in the manufacturing and services sectors, capture business perceptions on the biggest obstacles to enterprise growth, the relative importance of various constraints to increasing employment and productivity, and the effects of a country's business environment on its international competitiveness. They are used to create statistically significant business environment indicators that are comparable across countries. The Enterprise Surveys are also used to build a panel of enterprise data that will make it possible to track changes in the business environment over time and allow, for example, impact assessments of reforms.

The report outlines and describes the sampling methodology, the sample structure as well as additional information that may be useful when using the data, such as information on non-response cases and the appropriate use of the weights.

II. Sampling Structure

2. The sample for the Dominican Republic was selected using stratified random sampling, following the methodology explained in the *Sampling Note*¹. Stratified random sampling² was preferred over simple random sampling for several reasons³:

a. To obtain unbiased estimates for different subdivisions of the population with some known level of precision.

b. To obtain unbiased estimates for the whole population. The whole population, or universe of the study, is the non-agricultural economy. It comprises: all manufacturing sectors according to the group classification of ISIC Revision 3.1: (group D), construction sector (group F), services sector (groups G and H), and transport, storage, and communications sector (group I). Note that this definition excludes the following sectors: financial intermediation (group J), real estate and renting activities (group K, except sub-sector 72, IT, which was added to the population under study), and all public or utilities-sectors.

c. To make sure that the final total sample includes establishments from all different sectors and that it is not concentrated in one or two of industries/sizes/locations.

d. To exploit the benefits of stratified sampling where population estimates, in most cases, will be more precise than using a simple random sampling method (i.e., lower standard errors, other things being equal.)

¹ The complete text can be found at http://www.enterprisesurveys.org/documents/Implementation_note.pdf

² A stratified random sample is one obtained by separating the population elements into non-overlapping groups, called strata, and then selecting a simple random sample from each stratum. (Richard L. Scheaffer; Mendenhall, W.; Lyman, R., "Elementary Survey Sampling", Fifth Edition).

³ Cochran, W., 1977, pp. 89; Lohr, Sharon, 1999, pp. 95

e. Stratification may produce a smaller bound on the error of estimation than would be produced by a simple random sample of the same size. This result is particularly true if measurements within strata are homogeneous.

f. The cost per observation in the survey may be reduced by stratification of the population elements into convenient groupings.

3. Three levels of stratification were used in this country: industry, establishment size, and location. The original sample design with specific information of the industries and locations chosen is described in Appendix E.

4. Industry stratification was designed in the way that follows: the universe was stratified into 1 manufacturing industry, 1 service industry -retail -, and 1 residual sector as defined in the sampling manual. The manufacturing industry, service industry, and residual sectors had a target each of 120 interviews.

5. Size stratification was defined following the standardized definition for the Enterprise Surveys: small (5 to 19 employees), medium (20 to 99 employees), and large (more than 99 employees). For stratification purposes, the number of employees was defined on the basis of reported permanent full-time workers. This seems to be an appropriate definition of the labor force since seasonal/casual/part-time employment is not a common practice, except in the sectors of construction and agriculture.

6. Regional stratification was defined in two locations (city and the surrounding business area): Santo Domingo and the rest of the country (constituted by urban centers around Santiago and Higüey). For the purposes of sampling, the rest of the country was treated as one area.

III. Sampling implementation

7. Given the stratified design, sample frames containing a complete and updated list of establishments as well as information on all stratification variables (number of employees, industry, and location) are required to draw the sample. Great efforts were made to obtain the best source for these listings. However, the quality of the sample frames was not optimal and, therefore, some adjustments were needed to correct for the presence of ineligible units. These adjustments are reflected in the weights computation (*see below*).

8. TNS Opinion was hired to implement the LAC 2010 enterprise surveys roll out. In the Dominican Republic the contractor was EEC Canada.

9. The sample frame was provided by the Oficina Nacional de Estadística (ONE), dated 2009. The database contained the following information:

- Coverage;
- Up to datedness;
- Availability of detailed stratification variables ;
- Location identifiers- address, phone number, email;
- Electronic format availability;
- Contact name(s).

Sample Frames

10. The sample frame was then used for the selection of a sample with the aim of obtaining interviews with 360 establishments with five or more employees

11. The quality of the frame was assessed at the outset of the project through visits to a random subset of firms and local contractor knowledge. The sample frame was not immune from the typical problems found in establishment surveys: positive rates of non-eligibility, repetition, non-existent units, etc. In addition, the sample frame contains no telephone/fax numbers so the local contractor had to screen the contacts by visiting them. Due to response rate and ineligibility issues, additional sample had to be extracted by the World Bank in order to obtain enough eligible contacts and meet the sample targets.

12. Given the impact that non-eligible units included in the sample universe may have on the results, adjustments may be needed when computing the appropriate weights for individual observations. The percentage of confirmed non-eligible units as a proportion of the total number of sampled establishments contacted for the survey was 3.95% (40 out of 1,013 establishments).⁴

IV. Data Base Structure:

13. The structure of the data base reflects the fact that 3 different versions of the questionnaire were used. The basic questionnaire, the Core Module, includes all common questions asked to all establishments from all sectors (manufacturing, retail, and other services). The second expanded variation, the Manufacturing Questionnaire, is built upon the Core Module and adds some specific questions relevant to the sector. The third expanded variation, the Services Questionnaire, is also built upon the Core Module and adds to the core specific questions relevant to either retail or IT. Each variation of the questionnaire is identified by the index variable, *a0*.

14. All variables are named using, first, the letter of each section and, second, the number of the variable within the section, i.e. *a1* denotes section A, question 1. Variable names preceded by a prefix “LAC” indicate questions specific to LAC, therefore, they may not be found in the implementation of the rollout in other countries. All other suffixed variables are global and are present in all country surveys over the world. All variables are numeric with the exception of those variables with an “x” at the end of their names. The suffix “x” denotes that the variable is alpha-numeric.

15. There are 2 establishment identifiers, *idstd* and *id*. The first is a global unique identifier. The second is a country unique identifier. The variables *a2* (sampling location), *a6a* (sampling establishment’s size), and *a4a* (sampling sector) contain the establishment’s classification into the strata chosen for each country using information

⁴ Based on out of target contacts and impossible to contact establishments

from the sample frame. The strata were defined according to the guidelines described above.

16. There are three levels of stratification: industry, size and location. Different combinations of these variables generate the strata cells for each industry/location/size combination. A distinction should be made between the variable *a4a* and *d1a2* (industry expressed as ISIC rev. 3.1 code). The former gives the establishment's classification into one of the chosen industry-strata, whereas the latter gives the actual establishment's industry classification (four digit code) in the sample frame.

17. All of the following variables contain information from the sampling frame. They may not coincide with the reality of individual establishments as sample frames may contain inaccurate information. The variables containing the sample frame information are included in the data set for researchers who may want to further investigate statistical features of the survey and the effect of the survey design on their results.

-*a2* is the variable describing sampling locations

-*a6a*: coded using the same standard for small, medium, and large establishments as defined above. The code -9 was used to indicate units for which size was undetermined in the sample frame.

-*a4a*: coded using ISIC codes for the chosen industries for stratification. These codes include most manufacturing industries (15 to 37), other manufacturing (2), retail (52), and (45, 50, 51, 55, 60, 63, 72) for other Services.

18. The surveys were implemented following a 2 stage procedure. Typically first a screener questionnaire is applied over the phone to determine eligibility and to make appointments. Then a face-to-face interview takes place with the Manager/Owner/Director of each establishment. However, the phone numbers were unavailable in the sample frame, and thus the enumerators applied the screeners in person. The variables *a4b* and *a6b* contain the industry and size of the establishment from the screener questionnaire. Variables *a8* to *a11* contain additional information and were also collected in the screening phase.

19. Note that there are additional variables for location size by population (*a3*) and firm size by number of workers (*11*, *16* and *18*) that reflect more accurately the reality of each establishment. Advanced users are advised to use these variables for analytical purposes.

20. Variables *11*, *16* and *18* were designed to obtain a more accurate measure of employment accounting for permanent and temporary employment. Special efforts were made to make sure that this information was not missing for most establishments.

21. Variables *a17x* gives interviewer comments, including problems that occurred during an interview and extraordinary circumstances which could influence results. Please note that sometimes this variable is removed due to privacy issues.

V. Universe Estimates

22. Universe estimates for the number of establishments in each cell in the Dominican Republic were produced for the strict, weak and median eligibility definitions. The estimates were the multiple of the relative eligible proportions.

23. Appendix B shows the overall estimates of the numbers of establishments in the Dominican Republic based on the sample frame.

24. For some establishments where contact was not successfully completed during the screening process (because the firm has moved and it is not possible to locate the new location, for example), it is not possible to directly determine eligibility. Thus, different assumptions about the eligibility of establishments result in different adjustments to the universe cells and thus different sampling weights.

25. Three sets of assumptions on establishment eligibility are used to construct sample adjustments using the status code information.

26. Strict assumption: eligible establishments are only those for which it was possible to directly determine eligibility. The resulting weights, which include adjustments applied to panel firms (see below), are included in the variable *w_strict_panadj*.

Strict eligibility = (Sum of the firms with codes 1,2,3,4,&16) / Total

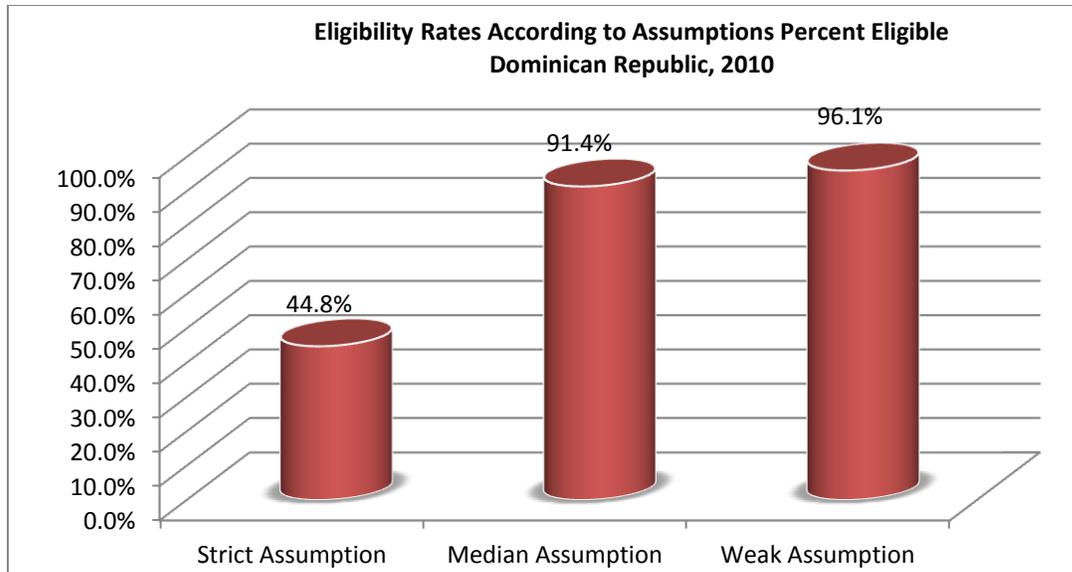
27. Median assumption: eligible establishments are those for which it was possible to directly determine eligibility and those that rejected the screener questionnaire or an answering machine or fax was the only response. The resulting weights are included in the variable *w_median_panadj*.

Median eligibility = (Sum of the firms with codes 1,2,3,4,16,10,11, & 13) / Total

28. Weak assumption: in addition to the establishments included in points a and b, all establishments for which it was not possible to contact or that refused the screening questionnaire are assumed eligible. This definition includes as eligible establishments with dead or out of service phone lines, establishments that never answered the phone, and establishments with incorrect addresses for which it was impossible to find a new address. Under the weak assumption only observed non-eligible units are excluded from universe projections. The resulting weights are included in the variable *w_weak_panadj*.

Weak eligibility = (Sum of the firms with codes 1,2,3,4,16,91,92,93,10,11,12,&13) / Total

29. The indicators computed for the Enterprise Survey website use the median weights. The following graph shows the different eligibility rates calculated for firms in the sample frame under each set of assumptions.



30. Universe estimates for the number of establishments in each industry-location-size cell in the Dominican Republic were produced for the strict, weak and median eligibility definitions. Appendix D shows the universe estimates of the numbers of registered establishments that fit the criteria of the Enterprise Surveys.

31. Once an accurate estimate of the universe cell projection was made, weights for the probability of selection were computed using the number of completed interviews for each cell.

VI. Weights

32. Since the sampling design was stratified and employed differential sampling, individual observations should be properly weighted when making inferences about the population. Under stratified random sampling, unweighted estimates are biased unless sample sizes are proportional to the size of each stratum. With stratification the probability of selection of each unit is, in general, not the same. Consequently, individual observations must be weighted by the inverse of their probability of selection (probability weights or *pw* in Stata.)⁵

33. Special care was given to the correct computation of the weights. It was imperative to accurately adjust the totals within each location/industry/size stratum to account for the presence of ineligible units (the firm discontinued business or was unattainable, education or government establishments, establishments with less than 5 employees, no reply after having called in different days of the week and in different business hours, no tone in the phone line, answering machine, fax line⁶, wrong address or moved away and could not get the new references) The information required for the

⁵ This is equivalent to the weighted average of the estimates for each stratum, with weights equal to the population shares of each stratum.

⁶ For the surveys that implemented a screener over the phone.

adjustment was collected in the first stage of the implementation: the screening process. Using this information, each stratum cell of the universe was scaled down by the observed proportion of ineligible units within the cell. Once an accurate estimate of the universe cell (projections) was available, weights were computed using the number of completed interviews.

The selection of panel firms required additional adjustments to account for varying probabilities of selection between fresh and panel sample universes. For additional information on this methodology, please refer to Enterprise Survey documentation of weighting methodology.

34. Appendix C shows the cell weights for registered establishments in the Dominican Republic.

VII. Appropriate use of the weights

35. Under stratified random sampling weights should be used when making inferences about the population. Any estimate or indicator that aims at describing some feature of the population should take into account that individual observations may not represent equal shares of the population.

36. However, there is some discussion as to the use of weights in regressions (see Deaton, 1997, pp.67; Lohr, 1999, chapter 11, Cochran, 1953, pp.150). There is not strong large sample econometric argument in favor of using weighted estimation for a common population coefficient if the underlying model varies per stratum (stratum-specific coefficient): both simple OLS and weighted OLS are inconsistent under regular conditions. However, weighted OLS has the advantage of providing an estimate that is independent of the sample design. This latter point may be quite relevant for the Enterprise Surveys as in most cases the objective is not only to obtain model-unbiased estimates but also design-unbiased estimates (see also Cochran, 1977, pp 200 who favors the used of weighted OLS for a common population coefficient.)⁷

37. From a more general approach, if the regressions are descriptive of the population then weights should be used. The estimated model can be thought of as the relationship that would be expected if the whole population were observed.⁸ If the models are developed as structural relationships or behavioral models that may vary for different parts of the population, then, there is no reason to use weights.

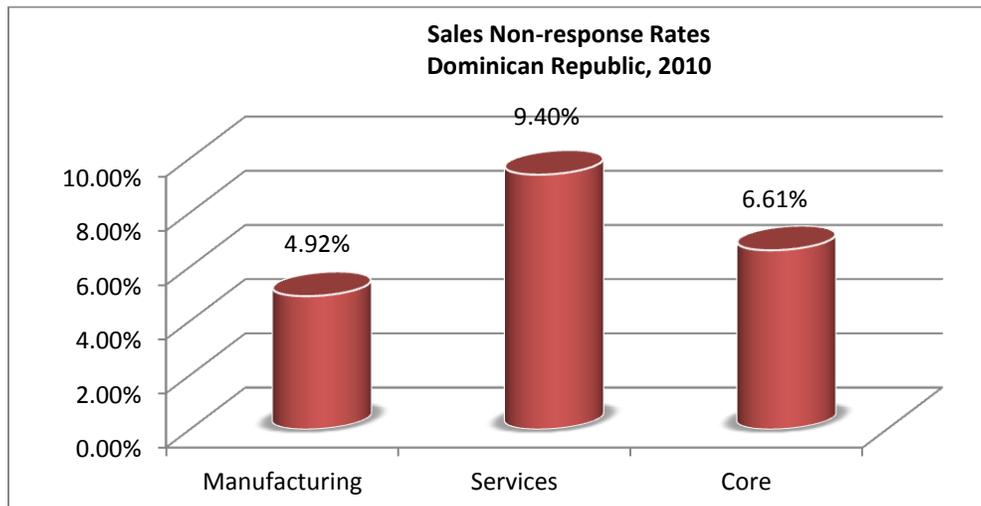
⁷ Note that weighted OLS in Stata using the command regress with the option of weights will estimate wrong standard errors. Using the Stata survey specific commands svy will provide appropriate standard errors.

⁸ The use weights in most model-assisted estimations using survey data is strongly recommended by the statisticians specialized on survey methodology of the JPSM of the University of Michigan and the University of Maryland.

VIII. Non-response

38. Survey non-response must be differentiated from item non-response. The former refers to refusals to participate in the survey altogether whereas the latter refers to the refusals to answer some specific questions. Enterprise Surveys suffer from both problems and different strategies were used to address these issues.

39. Item non-response was addressed by two strategies:
- a- For sensitive questions that may generate negative reactions from the respondent, such as corruption or tax evasion, enumerators were instructed to collect the refusal to respond as a different option from don't know (-7).
 - b- Establishments with incomplete information were re-contacted in order to complete this information, whenever necessary. However, there were clear cases of low response. The following graph shows non-response rates for the sales variable, *d2*, by sector. Please, note that the coding utilized in this dataset does not allow us to differentiate between "Don't know" and "refuse to answer", thus the non-response in the chart below reflects both categories (DKs and NAs).

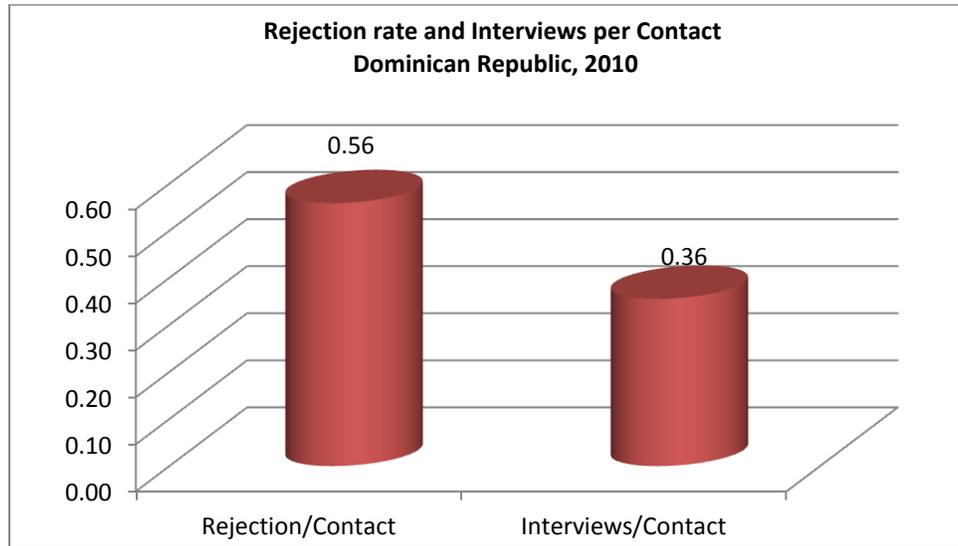


40. Survey non-response was addressed by maximizing efforts to contact establishments that were initially selected for interview. Attempts were made to contact the establishment for interview at different times/days of the week before a replacement establishment (with similar strata characteristics) was suggested for interview. Survey non-response did occur but substitutions were made in order to potentially achieve strata-specific goals. Further research is needed on survey non-response in the Enterprise Surveys regarding potential introduction of bias.

41. As the following graph shows, the number of realized interviews per contacted establishment was 0.36⁹. This number is the result of two factors: explicit refusals to participate in the survey, as reflected by the rate of rejection (which includes rejections of

⁹ The estimate is based on the total no. of firms contacted including ineligible establishments.

the screener and the main survey) and the quality of the sample frame, as represented by the presence of ineligible units. The number of rejections per contact was 0.56.



42. Details on the rejection rate, eligibility rate, and item non-response are available at the level strata. This report summarizes these numbers to alert researchers of these issues when using the data and when making inferences. Item non-response, selection bias, and faulty sampling frames are not unique to the Dominican Republic. All enterprise surveys suffer from these shortcomings, but in very few cases they have been made explicit.

References:

Cochran, William G., Sampling Techniques, 1977.

Deaton, Angus, The Analysis of Household Surveys, 1998.

Levy, Paul S. and Stanley Lemeshow, Sampling of Populations: Methods and Applications, 1999.

Lohr, Sharon L. Sampling: Design and Techniques, 1999.

Scheaffer, Richard L.; Mendenhall, W.; Lyman, R., Elementary Survey Sampling, Fifth Edition, 1996.

Appendix A

Status Codes:

	Status	#
Eligibles	1. Eligible establishment (Correct name and address)	454
	2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	0
	3. Eligible establishment (Different name but same address - the firm/establishment changed its name)	0
	4. Eligible establishment (Wrong address - the firm/establishment has changed address and the address could be found)	0
Ineligibles	5. The establishment has less than 5 permanent full time employees	6
	6. The firm discontinued businesses	23
	7. Not a business: private household	1
	8. Ineligible activity: education, agriculture, finances, governments...	2
Unobtainable	91. No reply (<i>after having called in different days of the week and in different business hours</i>)	0
	92. Line out of order	0
	93. No tone	0
	94. Phone number does not exist	0
	10. Answering machine	0
	11. Fax line - data line	0
	12. Wrong address/ moved away and could not get the new references	47
	13. Refuses to answer the screener	472
	14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)	0
	151. Out of target - outside the covered regions, firm moved abroad	8
	152. Out of target - firm moved abroad	0
	152. Out of target - Not registered with SAT	0
	Total	1013

Sample Target	Dominican Republic
	360
Complete interviews (Total)	360
Incomplete interviews	0
Eligible in process	0
Refusals	94
Out of target	32
Impossible to contact	47
Ineligible - coop.	8
Refusal to the Screener	472
Total	1013

Appendix B

Universe Estimate, Dominican Republic:

Source: Oficina Nacional de Estadística, 2009

Universe Figures

Dominican Republic		Manufacturing	Retail	Services	Grand Total
Santo Domingo	1-Small	924	992	2392	4308
	2-Medium	429	232	796	1457
	3-Large	197	46	226	469
Total		1550	1270	3414	6234
Rest of the country	1-Small	197	297	615	1109
	2-Medium	138	63	163	364
	3-Large	93	11	39	143
Total		428	371	817	1616
Grand Total		1978	1641	4231	7850

Appendix C

Strict Cell Weights Dominican Republic:

Strict Assumption Weights				
Dominican Republic		Manufacturing	Retail	Services
Santo Domingo	1-Small	27.38	14.81	51.83
	2-Medium	9.30	5.41	17.25
	3-Large	3.99	1.40	6.40
Rest of the country	1-Small	4.38	5.16	10.25
	2-Medium	2.65	1.08	2.85
	3-Large	1.69	1.00	1.40

Weak Cell Weights Dominican Republic:

Weak Assumption Weights				
Dominican Republic		Manufacturing	Retail	Services
Santo Domingo	1-Small	60.23	39.15	113.62
	2-Medium	20.74	11.21	39.14
	3-Large	7.88	2.25	11.30
Rest of the country	1-Small	11.07	13.02	26.65
	2-Medium	6.33	2.58	7.88
	3-Large	3.32	1.38	1.85

Median Cell Weights Dominican Republic:

Median Assumption Weights				
Dominican Republic		Manufacturing	Retail	Services
Santo Domingo	1-Small	57.49	37.03	109.63
	2-Medium	20.02	11.02	37.15
	3-Large	7.77	2.25	11.30
Rest of the country	1-Small	9.53	11.22	22.55
	2-Medium	6.21	2.54	7.20
	3-Large	3.27	1.38	1.75

Appendix D

Strict Universe Estimates

Strict Assumption Universe Estimate

Dominican Republic		Manufacturing	Retail	Services	Grand Total
Santo Domingo	1-Small	410.67	370.35	1036.53	1817.55
	2-Medium	185.90	108.27	344.93	639.10
	3-Large	99.81	28.00	128.07	255.88
Total		696.38	506.61	1509.53	2712.53
Rest of the country	1-Small	74.42	108.43	205.00	387.85
	2-Medium	52.90	26.00	57.05	135.95
	3-Large	42.16	8.00	28.00	78.16
Total		169.48	142.43	290.05	601.96
Grand Total		865.86	649.04	1799.58	3314.49

Weak Universe Estimates

Weak Assumption Universe Estimate

Dominican Republic		Manufacturing	Retail	Services	Grand Total
Santo Domingo	1-Small	903.47	978.77	2272.40	4154.64
	2-Medium	414.70	224.27	782.73	1421.70
	3-Large	197.00	45.00	226.00	468.00
Total		1515.17	1248.04	3281.13	6044.34
Rest of the country	1-Small	188.24	273.43	533.00	994.67
	2-Medium	126.50	62.00	157.57	346.07
	3-Large	83.08	11.00	37.00	131.08
Total		397.82	346.43	727.57	1471.82
Grand Total		1912.99	1594.47	4008.70	7516.16

Median Universe Estimates

Median Assumption Universe Estimate

Dominican Republic		Manufacturing	Retail	Services	Grand Total
Santo Domingo	1-Small	862.40	925.87	2192.67	3980.93
	2-Medium	400.40	220.40	742.93	1363.73
	3-Large	194.37	45.00	226.00	465.37
		1457.17	1191.27	3161.60	5810.04
Rest of the country	1-Small	161.98	235.71	451.00	848.69
	2-Medium	124.20	61.00	143.98	329.18
	3-Large	81.84	11.00	35.00	127.84
		368.02	307.71	629.98	1305.72
Grand Total		1825.19	1498.98	3791.58	7115.75

Appendix E

Original Sample Design, Dominican Republic:

Dominican Republic		Manufacturing	Retail	Services	Grand Total
Santo Domingo	1-Small	15	25	20	60
	2-Medium	20	20	20	60
	3-Large	25	20	20	65
Location total		60	65	60	185
Rest of the country	1-Small	15	21	20	56
	2-Medium	20	23	20	63
	3-Large	25	11	20	56
Location total		60	55	60	175
Grand Total		120	120	120	360

Completed Interviews, Dominican Republic:

Realized Interviews

Dominican Republic		Manufacturing	Retail	Services	Grand Total
Santo Domingo	1-Small	15	25	20	60
	2-Medium	20	20	20	60
	3-Large	25	20	20	65
Location total		60	65	60	185
Rest of the country	1-Small	17	21	20	58
	2-Medium	20	24	20	64
	3-Large	25	8	20	53
Location total		62	53	60	175
Grand Total		122	118	120	360