

## **Chapter 5: Statistical Thinking and Data Analysis: Enhancing Human Rights Work**

Jorge Luis Romeu

*The role of statistical thinking and data analysis in human rights (HR) work goes beyond detection and standardization of HR violations and the exposing of violators. It includes the development of conceptual models to better understand such violations, as well as to obtain HR analyses of greater validity. This chapter overviews several statistical methodologies that have been successfully applied by the author in previous HR case studies. The author first discusses the data and their characteristics, including measurement scale levels, data collection, information sources, data quality, and their corresponding statistical problems. The author then discusses methods for comparing a case vs. several controls, as well as for longitudinal analyses, using concomitant variables and information. Finally, he defines the concept of “differential increments” (the difference between HR violators’ and control countries’ performance measures), and applies it to comparing the multidimensional aspects of HR problems. Such a differential is used to assess the validity of some HR violators’ claims that HR violations are an unavoidable collateral damage of social advancement.*

### **5.1 Problem Statement**

The detection and assessment of human rights (HR) violations is not a trivial endeavor. On the contrary, it can be considered as an “open-ended” problem because the state of a country's HR situation (as well as its socio-economic development) are multidimensional problems. In addition, the problem components, as well as their relative importance (weights), are defined differently by different analysts. Let's explain that concept in detail.

Usually, there is more than one HR variable or factor in a problem (see the variables defined by Humana (1992)). Those variables can be grouped into three categories: political (e.g. the right to conduct free elections),

economic (e.g. right to work and to receive a fair salary), and social (e.g. the right to an education and health care). Not all analysts agree that all the above are “human rights,” and thus, some are defined as individual rights and others as social rights. However, all categories should be taken into consideration if we want to build a consistent and widely acceptable HR case, i.e. one that a majority will be willing to support and defend because it represents all the aspects (variables) of HR problems.

On the other hand, not all HR categories are simultaneously violated. Often, some HR categories are violated, while others are enhanced. For example, in Mao's China, in Stalin's USSR and in Castro's Cuba, political parties were suppressed by a dictatorship that improved public education and health care for many of its citizens. Such situations may lead different analysts to different conclusions. For example, some analysts may consider only (or mainly) political variables, while others may consider only or mainly socioeconomic ones, and different analysts may weight the same variables differently. Hence, establishing and assessing adequate HR variables ends up being the first, and often the most difficult, part of a HR study.

In addition, seldom does the HR situation of one country (say  $X$ ), completely dominate that of another country (say  $Y$ ), for all  $p$  variables under study. In mathematical parlance this is stated as:  $X_i \geq Y_i$ ,  $i = 1 \dots p$ . Hence, HR variable comparisons are performed through a (conscious or unconscious) process of vector dimension reduction, i.e. through the evaluation of a multivariate function “ $f$ ” defined as:

$$f(X) = \sum_{i=1}^p \alpha_i g_i(X_i) + \alpha_0 g_0(X_1, \dots, X_p) + \varepsilon$$

In non-mathematical terms, such function “ $f$ ” weights each HR and socioeconomic variable into a single value  $f(X)$ . The  $\alpha_i$ 's are the factor weights selected, the  $g_i$ 's are functions of the different HR variables included in the model,  $g_0$  is a function that collects all possible variable interactions, and  $\varepsilon$  is the random error of the statistical model.

Analysts (mathematically or mentally) build such a model and assign weights to each variable, to represent their relative importance in the HR and socioeconomic problem. For example, Stalinists believed that making education widely available to all was much more important (i.e.  $\alpha_i = 10\alpha_j$ ) than holding free and open elections. The resulting mathematical expression  $f(X)$  may not even be conscious for most analysts, but it is at the core of the discrepancies in HR data modeling and analysis.

As a result of all of the above, many HR analyses (e.g. analyses of China, Guatemala, Cuba) reflect the corresponding analysts' conscious or

unconscious philosophies. This situation may help explain why analysts who have conflicting political affiliations often arrive at conflicting HR conclusions that are based on the same HR data.

Examples of such controversial HR assessments abound in the HR literature. Two recent such cases are those of the Cuban revolution of 1959, and the 30-year (1960-90) civil war in Guatemala. In both cases, obvious HR violations were committed. However, very different results were obtained from the available data regarding HR violations, because of the different positions taken by HR analysts.

In the Cuban case, for example, hundreds of university students and faculty have been expelled from or prevented from enrolling in the university. Thousands of others have been jailed, or have been fired from their jobs, for political reasons. And over a period of 48 years of one-party rule, one million citizens, 10% of the Cuban population, have gone into exile. On the positive side, the Cuban government health and education policies have reduced illiteracy to practically zero, and have provided free and universal health care.

In Guatemala, during the long civil war, tens of thousands of civilians have died in obscure circumstances, or in overt massacres perpetrated by the army in their struggle against the guerrillas, and vice versa. On the other hand, after three decades of killings, the country has returned to a more or less pluralistic regime, where former guerrillas and military now jointly participate in public life, instead of fighting an open war.

In both cases, supporters of these regimes justify as “unavoidable collateral damage” actions that opponents of the same regime qualify as HR violations. Both groups claim that “one cannot make an omelet without breaking some eggs.”

To complicate matters even more, HR statistics research stems from essentially two lines of work. One is provided by the highly qualified statistical experts that have become interested in human rights through readings, personal involvement, moral reasons, etc. This book includes examples of many of the most relevant exponents of that sector.

The second line of work stems from researchers such as this author, who have lived for years under regimes that systematically violate human rights. For example, this author was expelled from the University of Havana, then sent for over two years to the Cuban *Unidades Militares para la Ayuda de Producción (UMAP)* forced labor camps; he underwent political trials, was sentenced to jail, and has an extensive dossier with the political police, including detention, interrogation and a suspended twelve-year sentence, all over purely non-violent, political dissent activities, such as writing and publishing abroad a story book about life in the UMAP camps.

After arriving to exile, this author, like other HR workers, has used his statistics training and background to help assess HR situations in his country, as well as in others. The author's HR case studies and work are referenced in the bibliography. This chapter is based on the methods implemented there and reflects the author's past experiences.

Researchers from the above-mentioned second line of work may not always have as deep an understanding of the theory and methods of statistics as researchers from the first line. But they compensate for that with their first-hand experience in what we may call the HR violations body-of-knowledge: experience in HR violations that can never be obtained through readings, interviews, or visits to areas of interest to HR studies. And that experience gives those researchers a sixth sense about where to look, how to assess events, etc. that others lacking that experience may not have.

The author believes that this book on statistics and HR is greatly increased in practical value because it includes the work of researchers from both lines of work mentioned above. For both groups complement each other well and thus will provide unique contributions to the field of HR research. This book also demonstrates how good statistical thinking and modeling can create a more scientific context where HR analyses can take place, where the inevitable trade-offs between different variable categories can be made under more equitable conditions, and where widely acceptable conclusions, leading to positive actions, can be obtained.

In the remainder of this chapter, the author develops three main analysis approaches. First, he discusses information sources, their measurement scale levels, the characteristics of HR data, and the appropriateness of the statistical methodology employed in typical HR analysis. Then, the author discusses the concept of "differential increments" and, using paired comparisons (treatment vs. control), he assesses HR violators via establishing a set of legitimate "control" countries. Finally, the author uses longitudinal studies, jointly with the graphical juxtaposition of qualitative concomitant variables, to build models that help explain the HR violations process, its origins and some of its causes.

The author illustrates those methodologies and their use via several HR cases he has thoroughly studied previously (details of these Cuban case studies can be found in the original papers, referenced in the bibliography). However, the proposed analysis methodology can be extended or adapted to many other HR studies. The author illustrates that via his references to cases of HR violations in Guatemala during thirty years of civil war.

## 5.2 Examples of Data Problems in HR Studies

In addition to possible analyst bias, there are other non-trivial statistical problems that can affect HR data analyses. They have to do with the origin of the data, the data collectors, unit definitions, etc. All of them can weaken the resulting measurement scale level of the data and, hence, limit the statistical methodologies that can be appropriately used with them. Hence, just as we do with other statistics applications, we need to account for them carefully.

First, when examining HR data, we should be conscious of the type of political system we are assessing, the degree of control the violator country generating the data exerts, etc. In China under Chairman Mao, political and press control were very tight and few independent data were available. But this did not preclude using data from the UN and other international organizations about number of executions, political prisoners, etc. Also, anecdotal and TV coverage information on nation-wide events such as the Tiananmen Square protests or the Cultural Revolution of the 1960s were available, however incomplete or otherwise faulty they might be. On the other hand, in a country where some free press, civil society, etc. still exists, the government is forced to allow greater degree of freedom and some independent data may be compiled, either directly or indirectly. Such occurred under some Latin American dictatorships of the 1980s.

Then, we have the issue of the personnel that gather HR data. As stated above, HR sometimes becomes a football in partisan politics. Both sides of the argument (those interested in justifying, as well as those interested in exposing, HR violations) may exaggerate their counts or magnify the quality of actual counts by adding an “affective” dimension to the data. For example, data on the number of civilian casualties in a war may emphasize number of women and children killed, without including the context (e.g. that they may be used as human shields by one of the sides).

Another problem consists in establishing the threshold beyond which an internationally acceptable action, such as military interrogation of war prisoners, becomes a HR violation (torture). That this is far from trivial is underlined by some recent interrogation procedures, which have been widely discussed, in conjunction with the legal definition of legitimate interrogation versus “torture” (see, for example, Attorney General Alberto Gonzalez and the Abu Ghraib case, in Newsweek International Special Investigation <http://www.msnbc.msn.com/id/4989481/> or the Washington <http://www.washingtonmonthly.com/features/2004/0411.carter.html>).

In addition, there are important data collection problems that arise when analyzing and comparing data from different countries. For example, when using U.N. Yearbooks and third world countries' censuses as sources, this researcher has found the following:

- different definitions for the same variables (does secondary education include normal or vocational schools?);
- different units (gross national product, given in domestic currencies);
- different time periods (results given per year versus per five years);
- overlapping periods (data collected from January to January vs. from June to June);
- vanishing/appearing series (cost of living indices);
- changing bases (index numbers);
- changing definitions within a series (the value of the monetary unit in which (say exports) are reported, fluctuates from year to year); and
- biased, incomplete or revised data.

Then, there are important effects caused by leaving certain concomitant variables out of the analysis of HR data. Some examples of variables excluded are:

- the status of any pre-existing infrastructure (for it is not the same to increase literacy by 20% when the starting level was 10% than when it was 75%);
- consideration of the growth effects of the (*S*-curve) (for there is a steeper rate of growth in the middle of any process, when conditions have been established and needs have been discovered, than at its start or end);
- consideration of the saturation effect (for there are just so much, say miles of road that should be constructed) and
- policy trade-offs to be made (20 kilometers more of roads versus a day care center) in the face of competition for limited resources.

All this will be further discussed in the section of this chapter on concomitant variables, where such information is used to show inconsistencies between a HR violator and a non-violator country.

Finally, a special situation occurs when comparing countries from different social and political regimes. In such comparisons, and to facilitate weight selection in the variable dimension reduction process, the use of "index numbers" has been suggested. However, the use of some such indi-

ces to compare economic achievements of a HR violator with those of a non-violator, such as the Consumer Price Index (CPI), may prove somewhat controversial. Using the CPI would imply favoring the free-enterprise, consumer-oriented system over state-run economies. That would make difficult the fair comparison of nations with non-capitalistic systems, for example.

All the above problems contribute to weaken the measurement scale levels of the resulting HR variables, which may then be assumed to have an ordinal level. The author proposes that non-parametric methods be used when the above-described situations arise in HR data analysis. The author has had similar weak measurement scale experiences when analyzing hardware and software reliability data, and has successfully used non-parametric statistical methods, with excellent results (Romeu et al. 2004; Romeu and Gloss-Soler 1983).

### **5.3 Characterization of “Differential Increments” in Country Comparisons**

In hypothesis testing, we compare either with a standard (Ho: the rate of death during the HR violations period, is the same as an accepted overall rate), or with a peer (Ho: rate of death during the HR violator's control is the same as that of a non-HR violator, peer country). In HR work it is often difficult to establish either of those two null hypotheses in a generally acceptable manner. Here, again, statistics can contribute to deal with the above-mentioned heterogeneity by using the proposed concept of “differential increments.”

A “differential increment” is the difference between the levels attained by a HR violator country, in some socioeconomic areas where it claims success, and similar socioeconomic achievements obtained by other comparable countries that are not HR violators. This approach seeks to show how similarly enabled countries can achieve similar success (be these the prevention of a dictatorship, or the rapid development of socio-economic levels) without having to resort to HR violations.

The author illustrates the derivation of such “differential increments” by comparing Cuban growth rates (used as the “case”), with those of three other “control” countries. Through this example he also shows how to implement that methodology for establishing “acceptable” control countries, a non-trivial but preliminary step in this approach.

The next section compares data from two adjacent, 40-year periods (before and after a claimed HR violation event), taken from Cuba (our case)

and three other control countries, where such an “event” has not occurred. We first compare their socioeconomic data from the period of 1920 to 1960, to establish the four countries' “similarity,” which will allow their use as controls. Then, we compare data from the period of 1960 to 2000, contrasting the growth rate of the case with that of the three controls.

#### 5.4 An Example of the Case/Control Approach

If the HR problem were one of *testing a new drug*, our first concern as statisticians would be to find a *suitable* control. Here, the null ( $H_0$ ) is that the treatment (HR violations) does not significantly increase the level of the response (better socioeconomic parameters such as health care and education services) as compared with a non HR violator country.

However, we would not pair, say a final stage, older cancer patient with a young, recently diagnosed one, to implement a clinical trial. We can also compare the case with itself used as its own control, by way of a *before and after* treatment approach. Here, the null ( $H_0$ ) is that before the period of HR violations, the response in question (socioeconomic) was not significantly different than during the HR violations period.

However, finding a suitable control country is not easy. The author uses controls in two successive phases. First, he compares their indices to establish the similarity required to be a valid “control.” Then, he obtains the “differentials” in selected indices in order to assess whether the case (HR violator) has actually achieved higher than the (non violator) control countries, as claimed.

To illustrate the author's selection methodology let's assume we want to use Guatemala as a control, for the study of Cuba (case). That would prove an inadequate choice. For even when size, population, and colonial history (both were Spanish colonies) are similar, they greatly differ in several important variables: ethnic composition (there are no Maya in Cuba, and few Africans in Guatemala), economic development, education levels, production systems, etc. (Cuba was much more advanced, economically and socially, than Guatemala).

The Republic of Chile is a better match. Variables such as population size, socio-economic, ethnic, religious and educational indicators, and even its modern political history are more similar (including alternating periods of pluralism and dictatorship) to those of the case (Cuba).

Three other potential controls are Costa Rica, Hungary, and Mexico. The first is one of the most stable and democratic nations in Latin America, but smaller and poorer than both Cuba and Chile. Hungary is an East



European socialist country that recently evolved toward pluralism (economic system). It is similarly populated, and has similar religious and socioeconomic features. But again, Hungary's history and ethnic composition is different than Cuba's and this may affect Hungary's behavior and social philosophy. We therefore prefer Mexico to Hungary, ending with four Latin American countries that have a similar historic, economic, ethnic and social background.

Chile is by far the closest match. Chile is larger in territory, but has vast, thinly inhabited desert and frigid regions. Its population size is comparable to that of pre/post revolutionary Cuba, as are its racial composition, levels of education and other socioeconomic variables.

Summarizing, Mexico is larger and Costa Rica is smaller than the case country in area and population. But both Mexico and Costa Rica have political and economic systems have had strong similarities with that of Cuba, and the colonial histories of the three countries have many common points. For example, Spain trained many Mexican Viceroy's as Cuban Governors. Even though Costa Rica is smaller than the case in size and population, it has also attained high socioeconomic standards. For many years Mexico, like Cuba, had a state-controlled economy and an hegemonic party political system. Costa Rica, in turn, has had a longstanding pluralistic one. Chile has experienced both pluralistic and authoritarian regimes in the recent past. Cuba has had alternating periods of democratic governments and military regimes, followed by a long dictatorship with a state-controlled economy.

We have discussed the selection of these three control countries in the same spirit as we would discuss the selection of siblings to examine the effects of a treatment versus a control. Through the control countries, we can examine the null hypothesis that the case (HR violator) obtained similar gains as the controls in the responses selected. If we are unable to reject the null, the claim that HR violations are necessary collateral damage in an accelerated process of development is disproven.

In the same spirit that we spend time and effort in a clinical trial to validate the similarities between case and control, we need to spend (possibly much more) time and effort to validate the control countries selected. All details about the countries selected as controls, about the reasons for specific socioeconomic indicators as variables, and about the case/control methodology discussed here, can be found in Romeu's original work, referenced in the bibliography.

### 5.4.1 An Example of Establishing the Validity of the Controls

In Figure 5-1, we present data from The Statesman Yearbook (1929), that quantitatively support these four countries strong similarities and illustrate our validation method. The selected variables provide a snapshot of their economic, political and social conditions, during the mid 1920's.

Variables selected include: total population, population density, primary students, teachers, cattle, kilometers of paved roads, kilometers of railroad tracks, kilometers of telegraph wire, and number of post offices. These measurements are given either in per capita, or per square km.

Some reasons for selecting these specific variables include that they reflect (1) general education level, (2) economic development, (3) communications facilities and (4) are available in the open literature. We don't expect all of them to be at par, in all countries. But their general socioeconomic picture, jointly with the historic-cultural-ethnic one, should point to such similarity. For example, we already showed how Chile and Cuba have similar population sizes, coefficients for primary teachers and students, and percent of post offices.

**Figure 5-1: Socioeconomic Indicators in the Mid-1920's**

Variables	Cuba	Chile	Costa Rica	Mexico
Total Population	3.57	3.75	0.45	14.9
Population Density	31.05	4.9	20.5	19.7
Primary Students	0.139	0.133	0.098	0.084
Teachers	0.00205	0.00288	0.00357	N/A
Cattle	1.337	0.511	0.919	0.375
Kilometers Road	0.0234	0.0466	0.0056	N/A
Kilometers Railroad Tracks	0.0684	0.0187	0.0179	0.0243
Kilometers Telegraph Lines	4.84	0.185	0.060	0.059
Number of Post Offices	0.000221	0.000260	0.000718	0.000044

Source: The Statesman Yearbook (1929)

A snapshot description as that of Figure 5-1 is insufficient to characterize a socioeconomic *process*. As we intend to compare these four countries not only at one instant but during a long period of time (over the twentieth century), we also need to assess the similarities between the *case* (Cuba) and the *controls* (Chile, Costa Rica and Mexico) over the time series for these variables. We will test the null ( $H_0$ ) that the same *process* will continue past the threshold date of 1960, in all performance areas.

In Figure 5-2, we show such data, taken at ten year intervals: circa 1938, 1948 and 1958, respectively. The variables selected include: population density (Dens), infant mortality (Mort), energy (Ener), primary students

(Stud), and number of radio receivers (Rads) given either per capita or per thousand inhabitants. Those variables were selected because they reflect the levels of health, education and nutrition in the population. Data are taken from the corresponding U.N. Yearbooks (1948 to 1993).

Notice, for example, how the 1938 infant mortality rate for the case was comparable to that of 1958 for the three controls. This suggests that the case was historically more advanced in health care than most Latin American countries, including the three controls. In energy levels, the case started behind two of the controls, Mexico and Chile. But the case tripled its energy capacity during these 20 years, surpassing all three controls by the end of this period.

In primary education, Mexico made the most significant gains, while the others remained at constant levels and the case lagged behind. In number of radios (reflecting standard of living as well as level of public information) the case had the highest levels. Finally, illiteracy rates in 1950, which reflect basic education levels, were: Costa Rica (20.6 percent), Chile (19.8 percent), Cuba (22.1 percent), and Mexico (43.2 percent). Only Mexico differed in this index, remaining way behind.

In the first section of this paper we discussed how “learning curves” and pre-conditions should be taken into consideration (as concomitant variables) to assess the case/control relative growth rates and their socioeconomic gains. Our indicators were obtained from the initial phase, where case and controls are established as relatively homogeneous, as part of the process of preparing a valid case/control country comparison for the second phase (1960-2000).

**Figure 5-2: Pre 1960 Longitudinal Comparison**

Variable	Chile	Costa Rica	Cuba	Mexico
Dens (c. 1938)	6.41	11.29	38.1	9.51
Dens (c. 1948)	7.7	15.23	45.9	12.61
Dens (c. 1958)	9.84	21.1	56.5	16.43
Mort (c. 1938)	235.7	123.1	83.0	128.0
Mort (c. 1948)	160.4	93.3	N/A	99.7
Mort (c. 1958)	126.8	89.0	34.7	80.8
Ener (c. 1938)	0.67	0.17	0.34	0.44
Ener (c. 1948)	0.76	0.22	0.47	0.61
Ener (c. 1958)	0.80	0.26	0.93	0.75
Stud (c. 1948)	0.131	0.144	0.107	0.116
Stud (c. 1958)	0.138	0.156	0.108	0.147
Rads (c. 1938)	0.031	N/A	0.034	0.019
Rads (c. 1948)	0.096	0.029	0.109	0.030
Rads (c. 1958)	0.089	0.070	0.170	0.077

Source: United Nations Yearbooks (1948-1993)

#### **5.4.2 Example of HR Violation Assessments via Pre/Post Test Comparisons**

The initial step in our before/after comparison consists of selecting the data. In the present example, those come for the periods before and after 1960 (sources include Oficina Nacional de Estadísticas 1971, 1981; Oficina Nacional de los Censos Demográfico y Electoral de Cuba 1954-1955; Eberstadt 1986; Gordon 1983; Mesa Lago 1971, 1981, 1987, 2000). With those data, we assess the null hypothesis that improvement levels obtained by the case during the HR violations period are not significantly different than those for the controls, and therefore cannot be used to justify the HR violations as collateral damage associated with such socioeconomic improvements.

Results from Cuba's 1953, 1971 and 1980 censuses yield that the percent of the workforce employed by the government was, respectively, 8.8 percent, 87 percent and 93 percent. There were 740,000 private workers reported in 1962, but only 175,000 in 1978. During that period, however, the population grew from 6 to 8 million. Hence, these data suggest a shift in type of employment patterns, rather than actual employment growth.

In the realm of public education, the 1953 census reports an adult literacy rate of 73 percent. This figure went up to about 96 percent after massive literacy campaigns in the early 1960's. There were only three public universities with 25,000 students and 2,500 professors in 1956. In 1986, there were over a dozen universities, with 256,000 third cycle students. This shows an overwhelming increase in the levels of education, for the country, during the violations period.

On the other hand, there were massive politically and religiously motivated faculty and student purges in the 1960's. And, until the mid-1980s, openly religious students were banned from registering for careers in medicine, economics, engineering, journalism, and others, in a country that, in 1956, was 80 percent self-avowed Catholics and 8 percent Protestants.

In the area of health care, Cuba went from a life expectancy of 64 years (1960) to 74.2 (1984); from an infant mortality rate of 34.7 per 1000 (1959) to 10.2 (1992); from 0.93 physicians per 1000 people (1959) to 4.33 (1992); from 0.74 nurses per thousand people (1959) to 6.83 (1992); from 4.22 hospital beds per thousand people (1959) to 6.1 (1992), Alonso and Lago (1994). Those data show a significant increase in health care levels during the HR violations period.

Internal migration is constrained because the government controls both housing and food ration cards. The 1980 Census reports an average

growth, among the 14 provincial capitals, of 17 percent (the country grew 25 percent), and Havana shows only 7.7 percent growth, half that of any other. Emigration is also severely controlled by the government; in spite of that, 10 percent of the population has left since 1960. Over 120,000 people left during the 1980 Mariel Boatlift and 35,000 more during the 1994 Raft Exodus. Exiles include tens of thousands of university professionals, administrators and technicians, forever lost to contributing toward the country's advance.

HR violations in Cuba are well documented and have occurred as a result of (i) peaceful political or religious activism, or for exercising the right to (ii) economic freedom or (iii) free emigration, when (i) and (ii) were not possible. The claim that such HR violations constitute an unavoidable “collateral damage” in the process of obtaining a larger “good” is often used as a defense by violator supporters. As discussed in our initial section, this is a classic multi-dimensional problem where there are significant gains in some areas, and significant losses in others. It is a matter of which ones the analyst chooses to report, how they are reported, and what kinds of trade-offs and valuations of the variables in question are performed.

#### **5.4.3 An Example of the Use of “Differential Increments”**

Another way of assessing the mentioned claims consists in comparing socioeconomic statistics of the case, with similar ones obtained from the three controls, for the same time periods, using the “differential increments” approach.

For example, in the previous section we showed how the case pre-1960 economic indicators were often better than those of the three controls. It could be reasonably conjectured that such pattern would continue for the following period. It would thus be reasonable to apply a flat 4 percent yearly growth rate to the case, and compare such long term forecasts to the actual values, for the 40-year period growth of 1960-2000. Any difference between both results could be attributed to the effects of the case's policies.

However, this approach is questionable, given that the world significantly changed in the second half of the twentieth century with respect to the first half. There were new technical, geopolitical, economic, etc. factors that did not exist before. Any valid time series analysis is based on the stability of the underlying process, which may not exist here. Hence, we will not pursue this approach.

Instead, we will use the actual levels attained by the three above selected control countries, whose 1920-1960 growth rates were at par or below those of Cuba, for the same period. We will use them to compare the case

actual achievements, with what it could presumably have obtained without the policies that led to the claimed HR violations. The comparison is valid, for both case and controls underwent, in the second period, the same changes and other factors. The difference between their achievements would then provide the *differential increments*.

We present a longitudinal study for the time period after 1960 in Figure 5-3, and a snapshot of the socioeconomic conditions of these four similar countries, at the end of the 1960-2000 period, in

Figure 5-4.

Figure 5-3 includes population density (Dens); infant mortality (Mort); female life expectancy (FLif); energy consumption (Ener), primary students (Std) and radio receivers (Rads), per capita or per thousands.

**Figure 5-3: Post 1960 Longitudinal Comparison**

Variables	Chile	Costa Rica	Cuba	Mexico
Dens (1970)	13.2	34.1	73.3	24.9
Dens (1980)	14.9	44.0	86.3	36.5
Dens (1990)	17.7	58.7	92.6	43.7
Mort (1980)	47	30	22	58
Mort (1990)	18	16	15	41
FLif (1980)	70.6	73.1	74.8	68.4
FLif (1990)	75.1	77.7	77.0	72.1
Ener (1970)	86	67	74	66
Ener (1980)	135	145	150	155
Ener (1986)	170	193	200	221
Std (1975)	0.224	0.183	0.192	0.190
Std (1980)	0.197	0.155	0.148	0.204
Std (1989)	0.151	0.141	0.083	0.168
Rads (1975)	164	77	194	111
Rads (1985)	330	246	326	189
Rads (1990)	340	259	343	242

Source: United Nations Statistical yearbooks (1949 to 1993)

From Figure 5-3, we see how Cuba and Chile have contained population growth, an indicator of social advance, as opposed to Mexico (who has only slowed it). Next, we see how Chile and Costa Rica have gone from a level of infant mortality much higher than Cuba, to one comparable to it. Mexico also made advances, but still lags behind in this indicator. In life expectancy, however, all three control countries have attained much higher levels, and Costa Rica has matched those of the case. These health indicators signal out how there have also been large health improvements

throughout Latin America (and the world) brought by new technical and medical advances independently of political regimes.

As the null hypothesis of no difference is not rejected, it is then possible to conjecture that the case's larger advances in health care would have been obtained anyway, especially when Cuba already had, since the 1950's, higher health standards and a longer health care tradition.

In energy consumption (indicator of industrial development and general standards of living) Costa Rica maintained a growth level similar to that of Cuba, but Mexico had a larger growth rate and Chile a smaller growth rate than Cuba. Primary students have decreased as a percentage of the general population. But this has been a general phenomenon in developing countries. Higher and technical education level percentages continue to be larger in Cuba. But the three controls are rapidly closing the gap. Finally, the number of radio receivers is now similar in Cuba and Chile while Costa Rica and Mexico are lagging behind.

Figure 5-3 allows us to compare, not only the *level* attained by each country, but also its *growth rate*. Mexico is still the one with the highest population growth and lowest socioeconomic indices among these four countries. Notice how Chile, Costa Rica and Cuba departed from different levels (case is generally ahead). However, all four have practically closed the gap by the end of the twentieth century.

Figure 5-4 shows some 1990s United Nations statistics. Notice the close agreement in indicators from Cuba, Chile and Costa Rica, showing how significant socioeconomic advances have been achieved by all countries in the region during the last forty years, independent of the type of socioeconomic system that prevails in each.

**Figure 5-4: Socioeconomic Indicators in the 1990s**

Variables	Chile	Costa Rica	Cuba	Mexico
Illiteracy (per 100 people)	8.9	7.4	3.8	17.0
Infant Mortality (per 1,000 infants)	17.1	13.9	11.1	43.0
Expected Male Life Span (years)	68.1	72.4	72.6	62.1
Expected Female Life Span (years)	75.1	77.0	76.1	66.0
Calorie Intake (daily average)	2,480	2,711	3,153	2,986
Protein Intake (daily average)	69.6	64	71.6	81.5
Cement (000s Tons)	2,115	N/A	3,696	24,683
Energy (000s KW)	1,270	602	1,461	1,788
Phones (per 1,000 people)	8.3	14.9	5.8	11.8
TV sets (per 1,000 people)	201	136	203	127
Radios (per 1,000 people)	340	259	343	242
Newspapers (units)	47	6	15	216

Students-1 (000s)	1,991	422	885	14,508
Students-2 (000s)	742	123	1,073	6,704
Population (per square kilometer)	13.1	3.0	10.6	86.2

Source: United Nations Statistical Yearbook 1993

Figure 5-4 also shows how Chile and Costa Rica reduced their illiteracy rate to single digits, and Mexico from 44 percent to 17 percent. With regard to intake of calories and proteins (nutrition indicators), Cuba is still about 10 percent above the three control countries in the former, and about the same in the latter. However, given the pre-1960 high health standards of the case, this is no surprise.

In energy consumption, basis of an industrial policy, the case is positioned between controls Mexico and Chile. Costa Rica, a mainly agricultural and rural nation, lags behind. Finally, in standard of living indicators (phones, TV, radios), the case levels of pre-1960 years have been surpassed or attained by the control countries. Given its higher levels in the 1950's, it seems that a higher growth rate was achieved by the three controls, with regards to standards of living. But we would still need to consider the saturation effect of the "S-curve."

We now show how we estimate the *differential increments* with respect to Costa Rica, for the variable *infant mortality*. Such differential is 2.8 (per 1,000). With respect to literacy, the differential is 3.6 percent. Those two values are obtained by assuming that the case would have attained, at least, the illiteracy rate of the control Costa Rica (7.4 percent) using the pre-1960 model, instead of its own value of 3.8 percent.

This premise is based on the fact that, under the previous model, the case was always at par or above Costa Rica in educational indicators. Therefore, any difference between the two may be attributed to the effect of the current Cuban socioeconomic development model. This would constitute a legitimate estimate of the *differential* for the variable *eradication of illiteracy*. All other differentials can be obtained in a similar manner.

Another approach would consist in using non-parametric regression to estimate the *differentials*. For illustration, we show the percent indices of total production of electricity for 1970-1986 in Figure 5-5, taken from Wilkie (1990). The year 1975 corresponds to 100 percent. The four country average values (per year) are also given. In Figure 5-5 we have also included the slope and Index of Fit (IoF) of the parametric linear regression, and the slope (NPSlp) and confidence interval (C.I.) for the non-parametric slope obtained from these data. Regressions were first obtained separately for each country, then for country averages, and finally, for the combined three controls. Hence, the combined regression column corresponds to the



(3x4=12) control data points. We obtain the parametric and non-parametric slopes (8.02 and 8.5) from the combined regression, and use them to estimate the general growth for variable electric power. We then compare them with the growth rates (slopes) for the case alone (7.7 and 7.5). Using the slope of the combined data, we obtain a higher index for the 1985 electricity production than the case actually achieved.

We also obtain a 90 percent non-parametric confidence interval for the slope of the combined data, and note the case slope is included within that confidence interval. We cannot reject the null hypothesis that the case growth in electricity production during the 1960-2000 period is similar to that of the combined three controls, none of which followed the case development model (which involved violating HR).

This analysis supports that regional growth, in general, has been similar in countries with different socioeconomic systems, for all have caught up with the case's growth rates. Such results lead us to question the case's claim that to obtain higher achievements in health and education, HR violations were unavoidable.

This argument is not unique to Cuba. The same has been used in Guatemala, with respect to the guerrilla insurrection. Costa Rica and Honduras, for example, two regional and similarly small and developing countries, did not resort to Guatemala's levels of violence and were not overrun by guerrilla insurrection either.

**Figure 5-5: Analysis of Total Electricity Production: 1975=100%**

Countries	1970	1975	1980	1985	Slope	IoF	NPSlp	C.I.
Chile	86	100	135	161	5.2	97	5.1	4.9-6.1
CostaRica	67	100	202	185	9.1	81	8.2	6.6-13.5
Cuba	74	100	150	185	7.7	98	7.5	7.0-8.5
Mexico	66	100	155	216	10.1	98	10.5	8.9-11.6
Average	73	100	164	187	8.04	97	8.1	5.4-8.7
Combined					8.02	86	8.5	6.9-11.0

Source: Statistical Abstract of Latin America, Vol. 28 (1990)

### 5.5 An Example of the Longitudinal Study Approach

This methodology consists of modeling, both graphically and analytically, the time series of HR violations within its context, in the manner physicians do with an electrocardiogram during a stress test, or in the manner the advance and retreat of Napoleon's forces were contrasted with number of casualties and dates, during his invasion of Russia in 1808, on a milestone statistical chart (see <http://www.edwardtufte.com/tufte/posters>).

First, we look at the changes and anomalies in the time series pattern. Then, we associate those with the changes in their contextual conditions, and try to infer causes, effects, etc. Let's illustrate such approach with an example.

Let the Cuban Gross Domestic Product (GDP) 1980-2000 time series  $\{X_t\}$ , be modeled as a function of social domestic policies, Soviet and Western European subsidies and businesses, and the effect of the U.S. economic embargo. In mathematical parlance we write:

$$X_t = f(g_1(W_t), g_2(Y_t), g_3(Z_t), g_4(W, Y, Z)) + E_t$$

where  $E_t \sim F$  ( $1980 \leq t \leq 2000$ ), and with the factors:

- (i)  $W_t$ , the Case economic/social domestic policies,
- (ii)  $Y_t$ , Soviet and Western European subsidies and businesses, and
- (iii)  $Z_t$ , effect of the U.S. economic embargo.

As usual,  $E_t$  is the statistical model error term, distributed F, and  $g_4$  collects all possible interactions between the model variables.

This conceptual model helps understand and assess the mentioned claims, made by the case, about its HR violations and its causes, by relating them to the overall 1980-2000 socioeconomic period (Figure 5-7), as well as to their historical context. Let's give the background of such context.

In 1980, 10,000 persons, in 48 hours, sought asylum in the Peruvian Embassy in Havana. Shortly after, 125,000 people sailed to the United States via the Mariel Boatlift. After these events, the Cuban government allowed some forms of private enterprise and free markets, up to then totally forbidden in Cuba. As a result, the economy grew considerably in the 1981-1985 period. In 1986, during the period of "Rectification of Errors," the Cuban government again prohibited most forms of private enterprise, and the economy shrank again.

At the end of the 1980's, internal struggles in the communist party led to the trial and execution of General A. Ochoa and other military, as well as to an increase in the internal dissidence movement. Then came, in 1989, the Glastnost and Perestroika movements, followed by the break-up of the Soviet Union, the end of the Cold War, and the loss of all Soviet subsidies: the Cuban economy collapsed.

The Cuban government then reversed its traditional economic policy. It established joint ventures with large European tourism companies, starting

a new “state capitalism” that saved the economic situation but exacerbated the internal dissident movement.

As a result of the large influx of foreign tourists, an explosion in male and female prostitution, drug use, etc. reappeared, all of which had disappeared from Cuba since the mid 1960s. All of this triggered another mass movement to leave the country at all costs, which caused extensive riots and the ensuing 1994 Raft Exodus.

Jointly with this social, political and economic situation, an explosion of peaceful dissent rapidly spread (e.g. publication of the document “La Patria es de Todos,” opening of “free libraries,” organization of independent labor unions and peasant cooperatives, etc.). Those activities culminated with the visit to Cuba of Pope John Paul II, and the rally of one million people in Havana to attend the Pope's mass. Such political events had not been witnessed in Cuba since the beginning of the revolution.

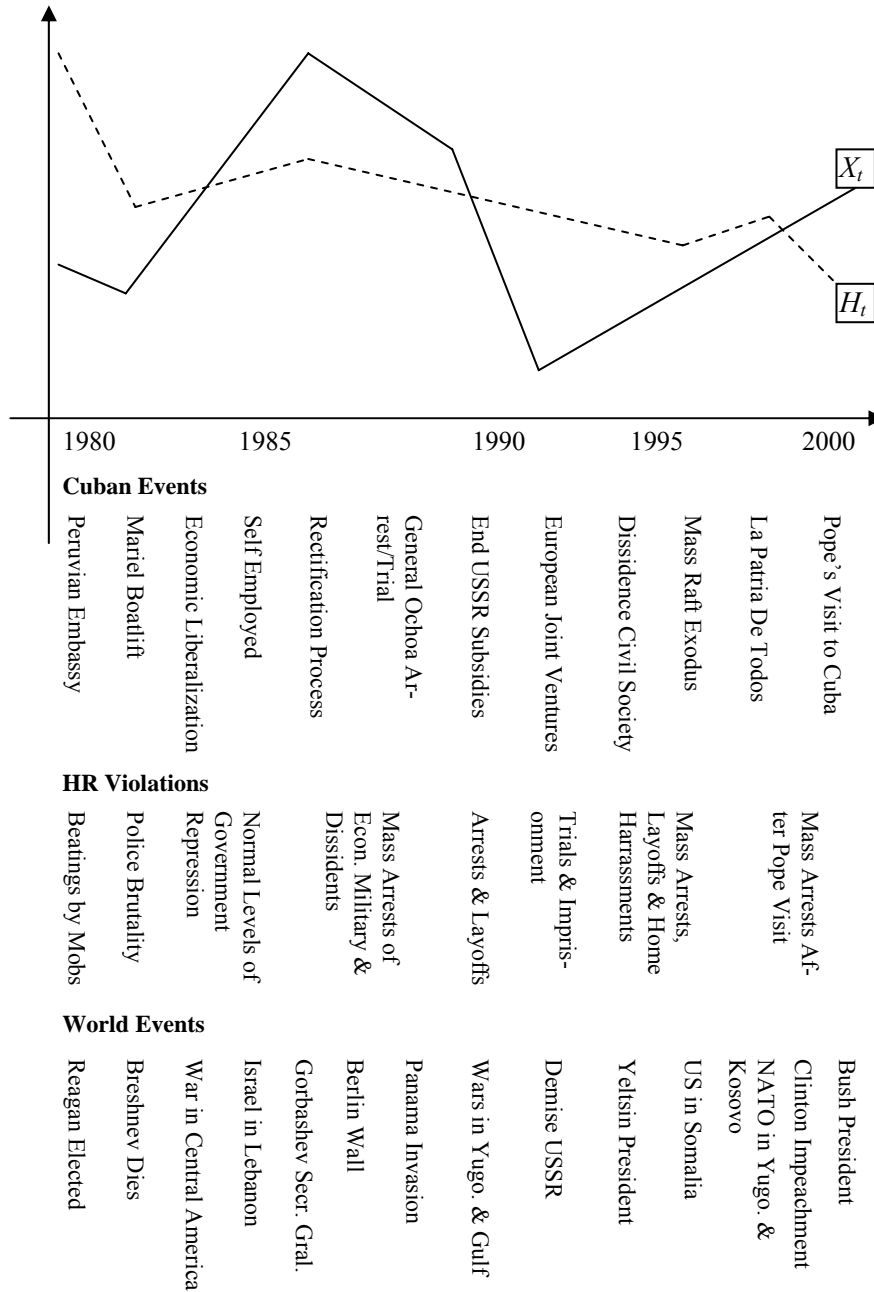
Throughout that same period (1980-2000), the American Embargo remained in place. Its effect on the overall Cuban economy can be assessed through the model. It appears to be non-significant, as the response is barely affected by it.

To illustrate this modeling approach, we show the Cuban Gross Domestic Product (GDP) by year in Figure 5-6. Concomitant events included in the graphical model of Figure 5-7 are: the Peruvian Embassy event, the Mariel Exodus, Free Markets, Rectification Process, the Ochoa trials, internal dissidence, the end of the Soviet Union and its subsidies, European tourism industry, and the Pope's visit.

**Figure 5-6: Example of Time Series (Cuba) with Concomitant Variables**

Year	1979	1980	1981	1982	1983	1984
$X_t$	4.2	2.2	16.0	3.9	4.9	7.2
Concomitant Events		Mariel Exodus	Free Markets			Free Markets
Year	1985	1986	1987	1989	1992	1995
$X_t$	4.6	1.2	-3.6	N/A	N/A	N/A
Concomitant Events	Free Markets	Rectification Process	Glastnost	Ochoa Trials	USSR Ends	EU Tourism

Source: Statistical Abstract of Latin America, Vol. 28, Table 3407.



**Figure 5-7: Graphical/Analytical representation of HR data, with Concomitant (historical events) variables. Note:  $H_t$  describes HR violation levels;  $X_t$  describes economic development levels.**

Figure 5-7 shows the economic variable  $X_t$ , and a second generic HR variable  $H_t$  that encompasses incarcerations, detentions without trial, mob attacks on dissidents, etc. Both variables capture the effects of the factors and events of interest. We use them to study their association with, as well as how they are impacted by, the concomitant socioeconomic and historical variables.

A careful look at this time series and its concomitant variables, as well as to the history of the entire 1960-2000 period, shows how, as the economy improves and the people become less dependent on the government, there is an increase in repression (HR violations) followed by a prohibition of the independent economic activities that bring about such independence. This reaction allows the government to recuperate its control of the political and economic life of the country.

Such cyclic periods of economic freedoms, followed by periods of independence and internal dissent, and then of government crack-downs on the independent economy, have occurred throughout 1980-2000: from the Mariel Boatlift to the Pope's visit. Such an analysis and modeling approach can help us understand the complexities in the HR situation of a country, and through that understanding devise policies that may eventually lead to finding a solution to the HR problems.

## 5.6 Other Statistical Studies and Researchers

HR is a very elusive concept, but a very concrete reality. Like beauty, HR issues are difficult to define but easy to recognize.

Many American Statistical Association (ASA) statisticians have been active, both as interested professionals as well as committed researchers, in HR work. Some of them have even been mentors to the author, encouraging him in his work, providing him readings and direction, and co-authoring some of his papers.

The ASA as an institution has also been very active in HR through its Scientific Freedom and Human Rights (SFHR) Committee, to which the author belonged for several years, as well as by promoting good statistical research. Much of this research has been published in the ASA Proceedings of the Social Statistics Section. A Compendium, including papers from 1984 to 2001, is available on the Web: [http://www.amstat.org/sections/ssoc/SSS\\_Human\\_Rights\\_Papers.pdf](http://www.amstat.org/sections/ssoc/SSS_Human_Rights_Papers.pdf). The Compendium is a great starting point to read the many excellent papers appearing in all those sources; those papers include interesting case studies and implement many statistical procedures in very unusual ways.

To avoid needless repetition, we do not give in this paper, which is part of a monograph on the HR subject, references to said work. This in no way diminishes the valuable contributions of all of these researchers to the area of the statistical study of HR violations.

The AAAS has also been in the forefront of HR work. In its Web site there is a HR Directory that includes a search engine and many links to other HR sources (<http://shr.aaas.org/dhr/>). Jana Asher, a co-editor of this book, has also compiled a partial list of research papers on statistics and HR assessments ([www.geocities.com/janalynnasher/hrbooklist.html](http://www.geocities.com/janalynnasher/hrbooklist.html)). Asher has also conducted several AAAS in-depth studies about HR violations, where she incorporates some time series charts with analytical and concomitant variable analyses like the ones we have discussed in this paper ([http://shr.aaas.org/pubs/author.php?\\_id=61](http://shr.aaas.org/pubs/author.php?_id=61)).

Without diminishing the work of any others, the author wants to acknowledge those who have mentored him, and have contributed in one form or another to his research. They include D. Banks, T. Jabine, F. Leone, H and L. Spirer, D. Samuelson and J. Asher, among others.

It would not be fair to conclude without mentioning the series of reports by Humana, as well as those of Human Rights Watch, America's Watch, Amnesty International, Freedom House, the US State Department, the United Nations, and other international organizations that the author often uses to obtain data to pursue his work. The HR monitoring program of Columbia University, among others, has produced some monographs and reports that are worth examination. The Human Rights Quarterly special issue, and the book to which it led (Claude and Jabine 1992), are major milestones in the vast work of statistics in support of HR. Finally, ASA Past-President F. Scheuren has written about the use of survey sampling in HR research. That attests to the degree and comprehensive commitment that the ASA has for HR causes and HR assessment.

## 5.7 Conclusions

Assessing HR violations is difficult, due to the multi-dimensionality of the problem, the lack of HR data, and measurement scale weaknesses, among other issues. Statistical analysis, with its methodological tools, can be very helpful in uncovering and testing HR violation patterns. In addition, statistical thinking can be useful in interpreting HR data analysis results and in providing an unbiased context in which to conduct a constructive discussion; one where HR violations will not be “justified” as part of the cost of obtaining a “greater” public good.

Summarizing, statistics can contribute to the HR work in at least three important ways. Firstly, by raising *awareness* among analysts and the public, to HR different factors and their complexities, and to the appropriate use of specific data analysis methodologies. Secondly, by providing a *scientific framework* (statistical thinking and philosophy) where data analyses may be performed in a more unbiased and acceptable fashion. Finally, by incorporating useful *statistical tools* such as case/control methodology, longitudinal studies and the use of historical and socioeconomic information, jointly with graphical analysis, to include those concomitant variables. Such inclusion can shed additional light, enlarging the problem dimension and facilitating the search for solutions. The methodologies used here were illustrated via the case of Cuba. However, they may be modified or extended and used in the study of other countries.

Finally, we also discussed the problem of interpreting HR results. The weights  $\alpha_i$ 's, assigned to reduce the problem of dimensionality, are of crucial importance, for different analysts may use different weights in their conscious or unconscious efforts to justify or condemn HR violations. We believe that it is in this arena, that statistical thinking can contribute most to HR work.

In his 22 years of direct exposure to HR violations, as well as his 25+ years working in favor of the recognition of HR for all, the author has found one issue of continuous re-occurrence. It is the one related to the "justification" of HR violations by way of pursuing "a greater good." In the parlance of such HR violators, "violations are an unfortunate necessary evil" in the quest of a "higher, nobler objective." Such justification pursues two well-defined objectives: one internal and another external. It tries to convince fellow countrymen that certain despicable actions have some moral validity and they should partake. It also pursues to convince the international community that such HR violations are necessary evils that are forced upon the violator by the actions of foreign governments or by the violators desire to raise socioeconomic or political performance measures, at all costs, in a short period of time. The author finds these concepts morally unacceptable.

The author can summarize his HR research work as follows: finding methods for fairly evaluating such HR ambiguities. HR violations are sometimes uncovered and standardized. But that does not necessarily lead to public condemnation for the reasons above-discussed and, thus, little action follows. As a result, people suffering HR violations do not improve their situation, which is the main objective of the author's HR research work. By contributing to take "wind away from the sails" of HR violator supporters, whoever those may be, the author will be actively contributing

towards improving HR causes, for the HR violators will then become isolated.

It is not enough to point out that HR violations occur, or to quantify them. It is also necessary to move others to act. And this is only possible when we can also show that there are no HR violations that can be justified under the “greater good” umbrella.

## 5.8 Acknowledgements

Many friends and colleagues have contributed to the author’s research in various ways throughout time. He gratefully acknowledges all of them, both in the Association for the Study of the Cuban Economy (ASCE) and ASA. We also thank the two reviewers, for their constructive criticism and comments, and particularly the two monograph editors, for their careful English editing of an earlier version of this paper, and for their invitation to contribute an article to this important book.

## References

- Alonso, J. F and A. Lago (1994) A First Approximation Design of the Social Safety Net of a Democratic Cuba. Cuba in Transition, Vol. 14. Assoc. for the Study of the Cuban Economy (ASCE). Miami, FL.
- Claude RP, Jabine TB, eds (1992) Human Rights and Statistics: Getting the Record Straight. University of Pennsylvania Press, Philadelphia.
- Eberstadt N (1986) Did Fidel Fudge the Figures? Literacy and Health: The Cuban Model. *Caribbean Review* 15:4-7, 37-38
- Gordon AM (1983) The Nutriture of Cubans: Historical Perspective and Nutritional Analysis. *Cuban Studies* 13:1-34
- Humana, Ch. (1992) World Human Rights Guide (3<sup>rd</sup>. Ed). Oxford University Press. New York. For more information, visit the Humana Web Site: <http://www.bsos.umd.edu/gvpt/lpbr/subpages/reviews/humana.htm>
- Mesa Lago C (1987) Cuba in the 1970’s. University of New Mexico Press, Albuquerque
- Mesa Lago C (2000) Market, Socialist and Mixed Economies: Comparative Policy and Performance of Chile, Cuba and Costa Rica. Johns Hopkins Press, Baltimore
- Mesa Lago C, ed (1971) Revolutionary Changes in Cuba. University of Pittsburgh Press, Pittsburgh
- Mesa Lago C (1981) The Economy of Socialist Cuba: a Two Decade Appraisal. University of New Mexico Press, Albuquerque



- Oficina Nacional de Estadísticas (1971) Censo de Población y Viviendas. Comité Estatal de Estadísticas. La Habana, Cuba.
- Oficina Nacional de Estadísticas (1981) Censo de Población y Viviendas. Comité Estatal de Estadísticas. La Habana, Cuba.
- Cuba en Cifras (1989). Comité Estatal de Estadísticas. La Habana, Cuba.
- Anuario Demografico (1979). Comité Estatal de Estadísticas. La Habana, Cuba
- Oficina Nacional de los Censos Demográfico y Electoral de Cuba (1954-1955) Censos de población. Electoral y de viviendas de 1953. La Habana, Cuba.
- Romeu JL (1993) Statistical Thinking Can Enhance Cuban Human Rights Assessments. Proceedings of the Social Statistics Section of the American Statistical Association, Alexandria, pp. 118-122
- Romeu JL (1994) A Statistical Comparison of Cuban Socioeconomic Development. Proceedings of the Social Statistics Section of the American Statistical Association, Alexandria, pp. 159-163
- Romeu JL (1996) More on the statistical comparison of Cuban socioeconomic development. Proceedings of 1995 Association for the Study of the Cuban Economy Conference, Miami, pp. 293-301
- Romeu JL (2005) On the Juarez Lincoln Marti International Education Project. Amer. Jour. Math & Mgmt. Sci. Vol 25, No. 1 & 2 pp 1-20. For more info. visit <http://web.cortland.edu/matresearch>.
- Romeu JL, Ciccimaro J, Trinkle J (2004) Measuring Cost Avoidance in the Face of Messy Data. Proceedings of the Reliability and Maintainability Symposium (RAMS), Los Angeles, CA, pp. 157-162
- Romeu JL, Gloss-Soler S (1983) Some Measurement Problems Detected in Software Data Analysis. Proceedings of Computer Software and Applications Conference (COMPSAC), Chicago IL
- Romeu JL, Samuelson D, Spierer H (1992) An Assessment of the Human Rights Situation in Cuba. Proceedings of the Social Statistics Section of the American Statistical Association, Alexandria, pp. 292-297
- The Statesman Yearbook (1929-55-92) Palgrave Macmillan, Basingstoke and New York. For more: <http://www.statesmansyearbook.com/public/>
- United Nations Yearbook (1948-52-59-69-71-81-93) United Nations, New York. More info: <http://www.un.org/Pubs/whatsnew/ycd02.htm>
- Wilkie J. W., Ed (1990) Statistical Abstract of Latin America, Volume 28. UCLA Latin American Center Publications, Los Angeles