

Detecting Problems in Survey Data using Benford's Law*

George Judge

Laura Schechter

University of California at Berkeley[†]

University of Wisconsin at Madison[‡]

November 1, 2007

Abstract

It is 15:00 in Nairobi. Do you know where your enumerators are??

Good quality data is paramount for applied economic research. If the data are distorted, corresponding conclusions may be incorrect. We demonstrate how Benford's law, the distribution that first digits of numbers in certain data sets should follow, can be used to test for data abnormalities. We conduct an analysis of nine commonly used datasets and find that much data from developing countries is of poor quality while data from the US seems to be of uniformly better quality. Female and male respondents give data of similar quality.

Keywords: Benford's law, first-digit phenomenon, relative frequencies, expected frequencies, data errors, survey quality.

AMS Classification: Primary 62E20.

JEL classification: C10, C24.

*The order of the authors' names has only alphabetical significance. We have benefited from helpful comments from Jennifer Alix-Garcia, Wendy Cho, Joanne Lee, Ted Miguel, Michael Roberts, Elisabeth Sadoulet, two anonymous referees, and seminar participants at BREAD, Purdue, and UW Madison. John Morrow provided excellent research assistance and also created the convenient utility to test your own data against Benford's law which can be found at <http://www.checkyourdata.com>. The second author received funding from USDA Hatch grant 142-1038.

[†]Professor in the Graduate School, Berkeley CA. e-mail: judge@are.berkeley.edu.

[‡]Agricultural and Applied Economics, Madison WI. e-mail: lschechter@wisc.edu.

1 Introduction

In developing countries, much of the social and economic data are collected by surveys. Horror stories are common in which, somebody discovers that one (or more) enumerator is answering the survey himself rather than actually interviewing households. Also prevalent are stories in which, after spending a large sum of money to buy a data set, a researcher realizes that the information of interest to him seems inaccurate. Since information contained in survey data often plays a key role in policy decisions, it is important to have a basis for identifying its quality.¹

In data obtained from economic surveys, questions usually arise pertaining to: i) the quality of the enumerators (what if an enumerator completes the survey questionnaire while enjoying coffee at Starbucks?) and ii) the quality of the responses from those interviewed (what if the questionnaire is poorly designed and elicits answers from respondents that are inconsistent with the objectives of the question?). If either the error of omission or commission occurs, it would be useful to identify it early in the research process. Therefore, a basis upon which one could recognize survey data irregularities, manipulated outcomes, and abnormal digit and number occurrences, would be a valuable tool for researchers designing and using survey data. In this paper, we demonstrate, in the context of large data sets, the use of Benford's first significant digit (FSD) law as one such possibility.²

1.1 Benford's Law

Benford's law characterizes the distribution of FSD observed in large sets of data. In 1881 Simon Newcomb observed that numbers with a first digit of 1 were observed more often than those starting with 2, 3, and so on. Newcomb was able to calculate the probability of a number having a particular nonzero first digit and published this in an article in *The American Journal of Mathematics*. Benford, unaware of Newcomb's article, made the same observation

¹Philipson & Malani (1999) posit that economists tend to pay more attention to the consumption of data rather than the production of data. This is evidenced by the large literature on how to deal with measurement error and the relatively small literature on how to prevent it. The three volumes edited by Grosh & Glewwe (2000), in addition to books by Groves (1989), Sudman et al. (1996), and Biemer et al. (1991) synthesize much of the literature and accepted best practice with regards to survey design and data collection.

²The first significant digit is the first non-zero digit reading a number from left to right.

and published an article in *The Proceedings of the American Philosophical Society* in 1938. This FSD phenomenon was christened as Benford's law.

Newcomb observed the probability of a number having a particular non-zero first digit as roughly

$$P(\text{First digit is } d) = \log_{10}\left(1 + \frac{1}{d}\right) \quad (1)$$

where $d = 1, 2, \dots, 9$. This formula produces a monotonically decreasing FSD distribution and suggests that the quantities expressed in base 10 will be uniformly distributed on a logarithmic scale. Using Newcomb's formula, the probability that the first digit of a number is 1 is about thirty percent ($P(d = 1) = \log_{10}\left(1 + \frac{1}{1}\right) = \log_{10}(2) \approx .30$) while the probability the first digit is 9 is 4.6 percent. Benford's law also has the nice property that it satisfies a scale and base invariance condition (Raimi 1976, Pietronero et al. 2001). This condition means that multiplying the data, such as prices or quantities of farm output, by any positive scalar will lead to the same FSD probability distribution.

Benford's somewhat surprising law, with its monotonic decreasing FSD distribution, has been demonstrated to hold with a large number of data sets that include the populations of towns, budgetary data of corporations, the number of citations received by papers, and the half-lives of radioactive atoms. The range of applications of the Benford phenomenon is impressive. All of these applied FSD distributions represent a dynamic mixture of data outcomes whose resulting combination is unrestricted in terms of the possibility of spanning the nine digit space.

Of course, not all natural data sets can be captured by the Benford FSD distribution. Binary or categorical data are two important economic examples that often occur in surveys. The daily closing price of a particular mutual fund on the Canadian Stock Exchange over a six month period is another example when Benford might not be expected to hold. On the other hand, in line with the dynamic mixture condition noted above, the FSD distribution of the closing prices for 500 stocks on this exchange over a six month period should, and does almost perfectly, track Benford.

Many data sets that fulfill the mixture-combination data condition noted above do not closely follow Benford. In fact, half of Benford's data sets do not closely follow his FSD distribution in terms of the level of the function. However, these data sets are Benford-like in that they are monotonic decreasing functions of the FSD data. Building on this base, we use the Benford

monotonic decreasing FSD distribution as one basis for identifying FSD distributions that may involve possible tampering or falsification of the data in economic surveys. Basically, it is difficult through tampering or human influence to duplicate the FSD's from natural outcomes of data sets, a fact we hope to exploit in this paper.

Like the equally surprising golden ratio (Livio 2002), theories abound as to the basis of the first-digit phenomenon. Consequently, there have been many attempts over the years to explain the logarithmic formula and to provide a theoretical basis for the observed phenomenon. Hill (1995) provided a statistical derivation of the law in the form of a Central Limit Theorem for significant digits: "If distributions are selected at random and random samples are taken from each of these distributions, the significant first digits of the combined sample will converge to the logarithmic (Benford) distribution." For overviews of the history and a sampling of the theoretical and empirical results, the reader is directed to Raimi (1976), Diaconis (1977), Schatte (1988), Hill (1995), Rodriguez (2004), Hill & Schürger (2005) and Berger & Hill (2007).

1.2 Overview of the Paper

Building on Benford's monotonic decreasing FSD distribution for naturally occurring multiplicative data sets, our objective is to exhibit scale-invariant data that may be expected to obey Benford's law and evaluate the behavioral basis of departures.³ Others who have used Benford's law to check the validity of purported scientific data in the social sciences include Varian (1972), Carslaw (1988), Nigrini (1996), Durtschi et al. (2004), Geyer & Williamson (2004), de Marchi & Hamilton (2006), Giles (2007) and Nye & Moul (2007).

To illustrate the use of Benford's FSD law to evaluate enumerator and respondent performance, we carry out a detailed analysis on survey data from rural Paraguayan households. Using this data we find that some enumerators and questions yield higher quality responses than others. We also compare data on crops which are more important for a household's income (with importance defined in multiple ways) with data on less important crops. We find that the former are fairly well in accord with Benford's law, while the latter are much less so. This suggests that Benford's FSD law holds when crop quantities are more salient and so farmers are able to provide their

³For some formal characterizations of when scale invariance occurs, see Morrow (2007).

answers with more accuracy.

In addition, we conduct a less detailed analysis of nine household surveys across the globe used extensively by research economists, presenting evidence that several widely used household data sets in development economics are not of very high quality. We find this is especially true of the Progres data from Mexico. On the other hand, the data from the United States seems to be of uniformly better quality than a range of surveys from the eight less developed countries. This may be because farmers in the US consult their records when responding to survey questionnaires or because they are more sure of exact quantities. We also find that the quality of data reported by female and male respondents is not very different.

We have access to one data set with data from some clusters which were later suspected to be falsified. These data are less in accord with Benford's law. On the other hand, we have another data set in which enumerators were asked to express their opinion regarding the quality of the data. When they opine that the quality was not good, the data is no less in accord with Benford's law. Perhaps this is because the natural FSD data process underlying Benford's law (like the heads or tails distribution in the flipping of a coin) is hard to duplicate. When an enumerator fakes data in an attempt to make it look real, and when he opines that some data is of better quality, this is exactly the data which is less in accord with Benford, not more.

The paper is organized as follows: Section 2 discusses the Paraguayan data used in this paper and applies Benford's law to both the 2002 and 1999 rounds of data; Section 3 expands the analysis to compare data sets from around the globe; Section 4 discusses the implications of our results for theory and practice; and Section 5 summarizes. Appendix A contains a more detailed description of the online utility www.checkyourdata.com which people can use to carry out the tests shown in this paper. Appendix B contains extra tables.

2 Paraguayan Survey Data

To illustrate the usefulness of Benford's law in assessing data integrity we use survey data from rural Paraguay. We find that some survey questions elicit more errors than others. For example, questions for which farmers know the exact answer generally conform to Benford's law while questions for which farmers may be unsure of the exact answer do not. When unsure, respondents

are more likely to choose numbers starting with 5 than Benford’s law would suggest. Finally, questions about which farmers may have an incentive to hide the truth, such as their donations to church, do not conform with Benford’s law. We also show that data collected by some enumerators are in accord with Benford’s law while data collected by others are not.

In 1991, the Land Tenure Center at the University of Wisconsin in Madison and the Centro Paraguayo de Estudios Sociológicos in Asunción worked together in the design and implementation of a survey of 300 rural Paraguayan households in 16 villages in three departments (comparable to states). The sampling design resulted in a random sample, stratified by land-holdings. The original survey was followed by subsequent rounds of data collection in 1994, 1999, and 2002. Summary statistics about both the households and the respondents can be found in Table B-1.

2.1 Benford’s Law Applied to 2002 Paraguay Data

2.1.1 Enumerated Data

We now examine whether the data collected in Paraguay produces an FSD distribution that is consistent with Benford’s law. We look at data for which the enumerator directly recorded the answer as stated by the farmer. The farmer was asked how much of each crop was harvested by his household in the past year.⁴ The data in Table 1 indicates how the data compares with Benford’s law. From a review of the table, it appears the data generally conform to Benford’s law, although quantities with an FSD of 5 are much more common than suggested by Benford’s law.⁵

In addition to merely visually reviewing the data, we use a χ^2 goodness-of-fit test to check the extent to which the data conforms with Benford’s law. The χ^2 statistic is calculated as $\chi^2 = \sum_{i=1}^9 \frac{(e_i - b_i)^2}{b_i}$ with 8 degrees of freedom where e_i is the observed frequency in each bin in the empirical data and b_i is the frequency expected by Benford. The 10%, 5%, and 1% critical values for χ^2 are 13.36, 15.51, and 20.09.

Following Giles (2007), we use Kuiper’s modified Kolmogorov-Smirnov

⁴If a household produces 150 kilos of corn and 420 kilos of cassava that will count as FSDs of 1 and 4, not 5.

⁵de Marchi & Hamilton (2006) use Benford’s law to test for tampering in self-reported toxics emissions by firms and likewise find that there are a disproportionate share of quantities beginning with 5.

Table 1: Benford’s Law and Production Quantities

Variable	Obs.	1	2	3	4	5	6	7	8	9
Benford		30.10	17.61	12.49	9.69	7.92	6.69	5.80	5.12	4.58
All Products	1632	31.50	19.30	12.81	7.90	12.93	5.82	4.11	3.62	2.02
Enumerator 1	516	30.23	19.19	13.37	8.72	10.66	8.91	2.71	4.46	1.74
Enumerator 2	556	32.73	19.06	12.59	7.73	12.77	6.29	3.42	3.42	1.98
Enumerator 3	560	31.43	19.64	12.50	7.32	15.18	2.50	6.07	3.04	2.32
P.I. sat in	582	33.33	16.49	13.06	7.04	13.92	4.64	4.47	4.64	2.41
P.I. didn’t sit in	1050	30.48	20.86	12.67	8.38	12.38	6.48	3.90	3.05	1.81

goodness-of-fit test (V_N) because it recognizes the ordinality and circularity of the data. This means that first digits of 5 and 6 are close to one another, as are first digits of 9 and 1. Kuiper’s test does not depend on the choice of origin (in contrast to the typical Kolmogorov-Smirnov test). Kuiper’s V_N statistic is calculated as $V_N = \max_x[F_e(x) - F_b(x)] + \max_x[F_b(x) - F_e(x)]$ where $F_e(x)$ is the empirical CDF of the FSD distribution and $F_b(x)$ is Benford’s CDF. Critical values for a modified Kuiper test ($V_N^* = V_N[N^{1/2} + 0.155 + 0.24N^{-1/2}]$) have been given by Stephens (1970). However, both the original and modified Kuiper tests were designed for use with continuous distributions. We use 10%, 5%, and 1% critical values for V_N^* of approximately 1.19, 1.32, and 1.58 shown by Morrow (2007) to be asymptotically valid.

In addition to the two statistical tests, we also include four measures of the distance between the empirical data and Benford’s distribution which are outside of the hypothesis-testing framework and insensitive to sample size. We include the Pearson correlation coefficient between the empirical proportions of first digits in the data and those predicted by Benford. Additionally, as suggested by Leemis et al. (2000), we calculate the distance measure $m = \max_{i=1,2,\dots,9}\{|b_i - e_i|\}$. Third, as suggested by Cho & Gaines (2007) we include a measure which is based on the Euclidean distance between the two distributions. Let $d = \sqrt{\sum_{i=1}^9 (b_i - e_i)^2}$. We then divide d by the maximum possible distance (which would occur when all numbers begin with an FSD of 9) so that the value is bounded between 0 and 1 and we call that measure d^* . Lastly, we note that the mean of the first digits in Benford’s FSD distribution is 3.4402. The maximum possible distance to this would occur when all numbers begin with an FSD of 9. We use a distance measure a^* which is the absolute value of the difference between the average of the

Table 2: Correlations (r), the m Statistic, Distances d^* and a^* , χ^2 Tests, and Kuiper V_N^* Tests between Benford’s Law and Production Quantities

Variable	r	m	d^*	a^*	χ^2	V_N^*
All Products	0.97	0.050	0.065	0.051	101.34**	2.69**
Enumerator 1	0.97	0.031	0.057	0.042	28.20**	1.50*
Enumerator 2	0.98	0.049	0.070	0.061	37.58**	1.68**
Enumerator 3	0.94	0.073	0.092	0.050	67.92**	1.97**
P.I. sat in	0.96	0.060	0.078	0.041	44.94**	1.47*
P.I. didn’t sit in	0.97	0.045	0.067	0.057	67.54**	2.26**

* indicates 95% and ** indicates 99% significantly different from Benford.

empirical FSD distribution and the average of Benford’s FSD distribution divided by the maximum possible difference.⁶

The Pearson correlation coefficient between the empirical data and Benford’s distribution tends to be extremely close to 1 for all of the data sets we look at. The level of variation in r is thus too small to make it a very useful distance measure. The measure a^* , based on the mean of the first digits of the data, is interesting but it is the measure which is most often in disagreement with the other measures and tests. This may be because it is not based on b_i and e_i directly as are the other tests. Thus, for our purposes, of the four distance measures we prefer m and d^* .

All of these statistics for testing and measures for summarizing the degree of deviation from Benford’s law are given in Table 2. The goodness-of-fit tests suggest a rejection of the hypothesis that the data on quantities produced was generated by Benford’s distribution. We might worry that one of the three enumerators was not completing surveys properly, thus explaining why the quantity data does not conform with Benford’s law. In addition, the Principal Investigator (P.I.) sat in on interviews with a different enumerator every day, alternating among the three. There was no specific type of household the P.I. tended to visit more often. We analyze whether or not the distribution of FSD changes based on the identity of the enumerator and the presence of the P.I. during the interview. Results suggest that bias remains.

The relative frequency departures are probably due to the fact that the

⁶Note that higher values of V_N^* , χ^2 , m , d^* , and a^* imply the empirical data is less similar to Benford’s distribution, while higher values of the correlation r imply the data is more similar to Benford’s distribution.

farmers were not always sure exactly how much of a crop they had harvested and so tended to choose ‘nice’ round numbers, e.g. a farmer is more likely to claim a harvest of 500 kilos of corn than 422 kilos. From the results in Table 2, we can statistically reject at the 99% level the hypothesis that the production data were generated in accord with Benford’s law. One could argue that the overabundance of observations starting with 5 is not due to guesstimation but rather to the fact that farmers’ plots are similar sizes and thus generate crop production of similar amounts. If this were the case, we would expect that there should also be relatively many observations beginning with 4 and 6 as well. However, when looking at the data one can see that this is not the case - the abundance of 5’s is at the expense of 4’s and 6’s.

This illustrates how Benford’s law may be used to distinguish problems in survey data arising because of an enumerator from those arising because of an ill-phrased question. For researchers who design surveys and collect data, this situation can be identified early in the survey collection process to help avoid major difficulties in subsequent econometric analysis.

2.1.2 Other Variables

In Table 3, we perform the same exercise on other variables which seem likely to follow Benford’s law. We cannot reject that the data on the number of animals owned, and hectares of land owned or used come from Benford’s distribution. The animals-owned variables include separately the quantities owned of each animal species. If households are able to report the assets they own more accurately than the quantities of crops harvested, measures of wealth for rural households in developing countries may be more accurate than measures of income. For research in which either wealth or income can be used, this may suggest a preference for the former. On the other hand, measures of both income and wealth also require accurate reports of prices, which may also contain significant error. We reject the hypothesis that donations to church are in accord with Benford’s law. This may be because respondents are not sure how much they donated to church or because respondents are reluctant to answer honestly.

2.2 Benford’s Law Applied to 1999 Paraguay Data

In 2002 only three enumerators collected the Paraguayan survey data, while in 1999 ten enumerators worked on the survey (with no overlap between the

Table 3: Correlations (r), the m Statistic, Distance d^* and a^* , χ^2 Tests, and Kuiper V_N^* Tests between Benford’s Law and Other Items

Variable	r	m	d^*	a^*	χ^2	V_N^*
Animals Owned	0.99	0.044	0.056	0.042	18.52*	1.14
Has. Land	0.98	0.034	0.050	0.031	8.49	0.72
Church Donations	0.92	0.083	0.123	0.090	38.93**	1.64**

* indicates 95% and ** indicates 99% significantly different from Benford.

Table 4: Correlations (r), the m Statistic, Distance d^* and a^* , χ^2 Tests, and Kuiper V_N^* Tests between Benford’s Law and Animals Owned and Crops Produced

	Surveys	Obs	Obs/Survey	r	m	d^*	a^*	χ^2	V_N^*
Animals (99)	298	827	2.8	0.93	0.061	0.087	0.043	56.21**	3.09**
Animals (02)	223	676	3.0	0.99	0.044	0.056	0.042	18.52*	1.14
Crops (99)	298	1412	4.7	0.99	0.026	0.036	0.025	30.36**	1.21
Crops (02)	223	1632	7.3	0.97	0.050	0.065	0.051	101.34**	2.69**
P.I. sat in (02)	71	582	8.2	0.96	0.060	0.078	0.041	44.94**	1.47*
no P.I. (02)	152	1050	6.9	0.97	0.045	0.067	0.057	67.54**	2.26**

* indicates 95% and ** indicates 99% significantly different from Benford.

two years). Having so many enumerators with only one supervisor might decrease data quality by decreasing supervision. We test if this is the case.

Tables B-2 and B-3 give a detailed analysis of animals-owned data and crops-produced data by enumerator in both rounds of the surveys. For both animals owned and crops produced, the distance measures for enumerators are much more varied in 1999 than 2002. On the other hand, looking at the 1999 data there is no enumerator who does consistently very well, or consistently very poorly for both animals owned and crops produced. In 2002, data from the first enumerator seems to be of consistently very high quality.

Table 4 contains a more succinct analysis. According to Table 4, the production-quantities data in 2002 looks more suspect than the data in 1999. This result is surprising at first, given the previous result that the animals-owned data was better in 2002, but the contradiction can be explained.

Enumerators in 1999 recorded fewer crop-production quantities per household. In 1999 households were recorded as producing an average of 4.7 dif-

ferent crops which rose to 7.3 in 2002. This could be due to an increase in diversification in the three years from 1999 to 2002. The likelier case is that, in 2002, enumerators were encouraged to be quite comprehensive and collect data on all crops produced, not just the most important ones.⁷ Respondents may not be sure about the exact quantity produced of crops which are less important to their livelihood (Groves 1989). This emphasizes the need for caution in using Benford's law. Although quantities of crops produced may be reported less accurately for less important products, ignoring them altogether will not increase the accuracy of measures of total income. This also leads to a bias because the income of more diversified farmers will contain more error than the income of relatively less diversified farmers.

An alternative explanation is that, by asking farmers for a more comprehensive list of crops planted in 2002, they lost patience with us and stopped answering the questions as carefully. This could also lead to the lower quality crop production data in 2002. We test this hypothesis by comparing data on 'important' crops to that on 'non-important' crops. This is shown in Table 5 using three definitions of importance. First, we compare total quantities harvested of crops which were sold by the household, with those grown for home consumption only. Next, we compare crops whose harvests were worth more than 500,000 guaranies in 2002 (342,500 guaranies in 1999 accounting for inflation, a bit less than 100 dollars in 2002) with those which were worth less. Lastly, we compare the four most valuable crops (in terms of the value of total output) for each household with any additional, less valuable, crops.⁸

Looking at Table 5 we see that 'important' crops, no matter how defined, are always more in accord with Benford's law than those defined as 'less important'. These results are quite striking given that there are usually more observations for 'non-important' crops than important ones. There were more non-important crops listed in 2002 due to the P.I.'s insistence on being complete while there were usually more important crops enumerated in 1999 due to the larger sample size in 1999. The inclusion of the less important crops, rather than respondent fatigue, seems to be why the 2002 crop data appears to be of lower quality than the 1999 crop data.⁹

⁷This is supported by the fact that, in 2002, the average number of different crops produced as recorded by the enumerator when the P.I. sat in on the survey was 8.2 versus 6.9 when the P.I. did not sit in on the survey.)

⁸If a household planted fewer than four crops, all were included as 'valuable'.

⁹More important crops, about which there is probably less guessing, conform more closely to Benford's law. This may be used as evidence to help convince those skeptical of

Table 5: Correlations (r), the m Statistic, Distance d^* and a^* , χ^2 Tests, and Kuiper V_N^* Tests for More and Less ‘Important’ Crops

	# Obs	r	m	d^*	a^*	χ^2	V_N^*
Crops Sold in 1999	378	0.99	0.027	0.038	0.000	5.58	0.72
Crops Not Sold in 1999	1034	0.98	0.036	0.050	0.034	40.12**	1.51*
Crops Sold in 2002	384	0.97	0.038	0.054	0.007	12.75	0.82
Crops Not Sold in 2002	1248	0.97	0.054	0.075	0.065	100.82**	2.87**
Crops Value > 342,500 in 1999	889	1.00	0.018	0.025	0.016	9.81	0.58
Crops Value \leq 342,500 in 1999	523	0.95	0.066	0.086	0.040	53.81**	1.57*
Crops Value > 500,000 in 2002	667	0.99	0.027	0.037	0.005	10.54	1.17
Crops Value \leq 500,000 in 2002	965	0.94	0.085	0.113	0.083	164.08**	3.34**
Hh’s top 4 crops in 1999	1026	0.99	0.019	0.028	0.009	13.73	0.77
Hh’s other crops in 1999	386	0.98	0.045	0.071	0.068	28.53**	1.47*
Hh’s top 4 crops in 2002	828	0.98	0.032	0.046	0.030	27.25**	1.04
Hh’s other crops in 2002	804	0.96	0.069	0.092	0.073	95.08**	2.78**

* indicates 95% and ** indicates 99% significantly different from Benford.

3 Comparing High-Profile Data Sets

In this section we analyze the quality of nine data sets which have been used in a multitude of academic papers and have been used to make policy prescriptions. We look at two data sets collected under the supervision of academic economists and seven data sets collected under the supervision of government or international agencies with input from academic economists. In recent years, it has become more popular for researchers to supervise their own data collection, but nothing is known about the relative quality of these homegrown data sets. *A priori* one could argue why either should be of higher quality. In this study, the data sets collected by academic researchers without aid from government or international agencies seem to be more free of distortions than those collected by the government. On the other hand, we only consider a small number of variables across a small number of surveys, so it is difficult to make any definite statements about relative quality.

The seven data sets we examine are:

1. The Matlab Health and Socioeconomic Survey (MHSS) was collected in 1996 as a collaborative effort by RAND, multiple universities in the

Benford’s ability to describe the distribution of FSD in naturally occurring data.

United States, and research centers in Bangladesh. There is data on over 4,500 rural Bangladeshi households.

2. Data was collected in Ghana from 1996-1998 under the supervision of Chris Udry (a professor at Yale) and Markus Goldstein (then a graduate student at UC Berkeley). The data set includes information on 294 households.
3. The Progres data from Mexico consists of panel data for 24,000 rural Mexican households collected every 6 months beginning in November of 1997. This data was collected by Progres, which is part of the Mexican government, with consultations from the International Food Policy Research Institute (IFPRI).
4. The IFPRI Pakistan data includes 14 rounds of panel data covering rural households and villages spanning 1986-91. The survey was jointly produced by IFPRI, the Government of Pakistan, and the U.S. Agency for International Development (USAID). Large fluctuations in agricultural production observations across rounds are in part due to the particular season in which that round was conducted.
5. The 2002 round of Paraguayan data discussed thus far was collected under the supervision of Laura Schechter when she was a graduate student at UC Berkeley and includes 223 households. The 1999 round of data was collected under the supervision of Diana Fletschner, then a graduate student at UW Madison, and includes 298 households.
6. The Peru Living Standards Measurement Survey (LSMS) (called PLSS or ENNIV) contains information on both rural and urban households. We have excluded urban households to maintain comparability with the other data sets. The survey has data on 2,349 rural households in 1985, 594 rural households in 1991, and 1,336 rural households in 1994. The 1985 data was collected by the Statistical Institute of Peru (“Instituto Nacional de Estadística e Informática del Perú (INEI)”), with technical and financial support from the World Bank and the Central Reserve Bank of Peru. The 1991 and 1994 data were collected by the Peruvian research enterprise Cuánto S.A. with technical and financial assistance from the World Bank (with additional assistance from the Interamerican Development Bank in 1994).

7. The KwaZulu-Natal Income Dynamics Study (KIDS) surveys individuals in the South African province of KwaZulu-Natal in 1993, 1998, and 2004. KIDS was a collaborative project between researchers at IFPRI, the University of KwaZulu-Natal, the University of Wisconsin, London School of Hygiene and Tropical Medicine, the Norwegian Institute of Urban and Regional Studies, and the South African Department of Social Development.
8. The United States Department of Agriculture produces the Agricultural Resource Management Survey (ARMS). The sample size was, more or less, 10,000 farm households in the 1990's but has gone up to approximately 20,000 households in the twenty-first century. In the ARMS data, missing observations are created and outliers replaced using a complicated interpolation process. Unfortunately, the interpolated observations are not flagged and so the analysis here includes all observations.
9. The Vietnam Living Standards Survey (VLSS) contains information on 4,800 households in 1992 and 6,000 households in 1998. This data was collected by the General Statistical Office of Vietnam with help from the World Bank.

We have chosen to look at two variables for which Benford's law may be applicable and are comparable across surveys, the quantities of crops harvested and the number of animals owned. We expect these data to be in accord with Benford's law for two reasons, the first being that they concern, arguably, naturally random occurring phenomena. Although farmers plant seeds, nature conditions the output, and although farmers sell and buy animals, they reproduce at their own pace. Secondly, both variables are mixtures of distributions. Observations of crops produced are the combination of distributions for all the crops that households could produce while the animals-owned variable combines the distributions for all the types of animals households could own.

The results for crop production quantities in the eight surveys which ask about quantity produced and the results for the nine surveys which ask about animal ownership are presented in Tables 6 and 7. (The U.S. data asks crop yield per acre, which is not analyzed here since it is not comparable and would not necessarily be expected to follow Benford's law.) We see that, as

expected, the two tests (the χ^2 statistic and V_N^* test) tend to rise with sample size, while the distance measures of r , m , d^* , and a^* do not.

The data from Bangladesh, Ghana, Paraguay, and the United States are consistently relatively close to Benford's law; the data from Mexico and Pakistan seem to be the least in accord with Benford's law; and the data from Peru, South Africa, and Vietnam are somewhere in the middle.¹⁰ The data from Mexico and Pakistan are the only ones with correlations below .90 and m and d^* statistics higher than 0.10 for both animals owned and crops produced. Comparing the χ^2 and V_N^* from Mexico with those of other surveys of comparable sizes such as Vietnam, Peru in 1985, and Bangladesh we find that the test statistics for the data from Mexico are much higher than those in the other three data sets. The test statistics for the Pakistan data are more in line with those of the other smaller data sets.

Comparing the quality of data on crop production (a component of income) with that on animals owned (a component of wealth), neither seems to be clearly better in all of the data sets. Crops are more in accord with Benford than animals in the data sets from Pakistan and Vietnam while the opposite is true in the data from Mexico and South Africa. In the other data sets the two types of variables perform equally well.

The data from the United States performs extremely well in comparison with the data sets from developing countries. This could be because farmers in the U.S. tend to consult their records while answering surveys and so their answers involve less guesstimation and are more 'correct' than the corresponding answers given by farmers in developing countries. It could also be due to the fact that missing values and outliers in the U.S. data set are replaced with interpolated values. Since the National Agricultural Statistics Service (NASS) does not flag interpolated values, it is impossible to distinguish between these two hypotheses.

¹⁰We asked Markus Goldstein why he thought the data collected in Ghana was of such high quality. Markus replied that, as do many other organizations, they paid enumerators high wages so they would be afraid of losing their jobs. More unusually, they made a point of hiring almost all high school graduates, with only one college graduate. Lastly, Chris and Markus sat in on interviews for an average of one and a half days a week.

Table 6: Correlations (r), the m Statistic, Distance d^* and a^* , χ^2 Tests, and Kuiper V_N^* Tests between Benford's Law and Crop Quantities Produced

	# Surveys	# Obs	r	m	d^*	a^*	χ^2	V_N^*
Bangladesh (1996)	4,522	7,357	0.93	0.065	0.085	0.039	440.92**	8.27**
Ghana (96-98)	294	799	0.99	0.016	0.026	0.010	8.45	0.85
Mexico (Oct 98)	24,067	10,028	0.84	0.154	0.169	0.059	3959.21**	15.45**
Mexico (May 99)	22,328	9,659	0.88	0.134	0.155	0.075	3145.46**	13.14**
Pakistan (July 86)	927	874	0.97	0.040	0.058	0.036	31.62**	1.84**
Pakistan (Oct 86)	909	181	0.86	0.100	0.128	0.016	23.37**	1.40*
Pakistan (Jan 87)	882	563	0.99	0.047	0.062	0.046	20.71**	1.13
Pakistan (Mar 87)	854	46	0.94	0.051	0.090	0.028	4.79	0.73
Pakistan (Apr 87)	845	191	0.96	0.041	0.078	0.059	15.90*	1.13
Pakistan (July 87)	831	658	0.97	0.033	0.061	0.022	43.75**	0.95
Pakistan (Jan 88)	813	728	0.96	0.046	0.069	0.001	41.45**	1.41*
Pakistan (Mar 88)	809	69	0.92	0.048	0.096	0.050	8.11	1.16
Pakistan (Aug 88)	804	559	0.99	0.022	0.038	0.027	15.44	0.77
Pakistan (Jan 89)	802	708	0.99	0.034	0.052	0.054	26.23**	1.27
Pakistan (Mar 89)	766	200	0.78	0.100	0.141	0.037	29.67**	2.11**
Pakistan (Aug 89)	759	691	0.98	0.025	0.045	0.020	22.48**	1.06
Pakistan (Oct 91)	726	1,095	0.99	0.022	0.033	0.017	18.72*	1.16
Paraguay (1999)	298	1,412	0.99	0.026	0.036	0.025	30.36**	1.21
Paraguay (2002)	223	1,632	0.97	0.050	0.065	0.051	101.34**	2.69**
Peru (1985)	2,349	8,050	0.99	0.026	0.042	0.030	287.76**	4.23**
Peru (1991)	594	2,495	0.98	0.047	0.086	0.102	228.02**	5.47**
Peru (1994)	1,336	4,101	0.98	0.038	0.064	0.059	246.39**	4.63**
South Africa (1993)	1,412	543	0.99	0.097	0.134	0.159	79.34**	3.89**
South Africa (1998)	1,075	788	0.96	0.082	0.108	0.118	97.48**	3.68**
Vietnam (1992)	4,800	20,936	0.99	0.034	0.049	0.045	762.36**	7.80**
Vietnam (1998)	6,002	24,420	0.98	0.035	0.055	0.053	1092.93**	9.64**

* indicates 95% and ** indicates 99% significantly different from Benford.

Table 7: Correlations (r), the m Statistic, Distance d^* and a^* , χ^2 Tests, and Kuiper V_N^* Tests between Benford's Law and Animals Owned

	# Surveys	# Obs	r	m	d^*	a^*	χ^2	V_N^*
Bangladesh (1996)	4,522	6,807	0.99	0.057	0.073	0.078	333.56**	7.25**
Ghana (Nov 96)	294	332	0.98	0.023	0.042	0.013	7.79	0.79
Ghana (Dec 97)	294	335	0.99	0.016	0.028	0.006	2.76	0.60
Ghana (Aug 98)	294	306	0.99	0.017	0.033	0.007	3.83	0.75
Mexico (Nov 97)	24,077	48,042	0.99	0.021	0.044	0.007	1142.88**	6.64**
Mexico (Oct 98)	24,067	37,130	1.00	0.098	0.119	0.138	4052.83**	27.23**
Mexico (May 99)	22,328	34,539	1.00	0.108	0.127	0.142	4065.22**	28.15**
Mexico (Nov 99)	23,266	38,817	0.99	0.114	0.133	0.141	4750.71**	30.85**
Mexico (May 00)	22,627	37,580	1.00	0.115	0.133	0.146	4703.09**	29.88**
Pakistan (July 86)	927	1,562	0.96	0.119	0.155	0.183	358.08**	8.07**
Pakistan (Oct 86)	909	1,981	0.94	0.120	0.141	0.153	386.43**	7.36**
Pakistan (Jan 87)	882	2,010	0.77	0.128	0.171	0.061	395.04**	8.82**
Pakistan (July 87)	831	1,961	0.84	0.116	0.147	0.078	326.96**	7.95**
Pakistan (Mar 88)	809	1,786	0.77	0.171	0.193	0.095	456.86**	8.58**
Pakistan (Mar 89)	766	1,838	0.81	0.154	0.174	0.092	402.10**	8.28**
Pakistan (Oct 91)	726	559	0.71	0.135	0.158	0.067	89.93**	3.83**
Paraguay (1999)	298	827	0.93	0.061	0.087	0.043	56.21**	3.09**
Paraguay (2002)	223	676	0.99	0.044	0.056	0.042	18.52*	1.14
Peru (1985)	2,349	8,007	0.98	0.072	0.089	0.092	696.76**	8.73**
Peru (1991)	594	2,369	0.97	0.069	0.090	0.104	220.62**	4.82**
Peru (1994)	1,336	3,392	0.99	0.048	0.059	0.064	126.40**	3.72**
South Africa (1993)	1,412	696	1.00	0.019	0.027	0.010	4.35	0.52
South Africa (1998)	1,075	694	1.00	0.039	0.045	0.042	11.26	1.18
United States (1996)	9,573	8,549	1.00	0.009	0.016	0.002	33.29**	1.81**
United States (1997)	11,724	20,345	1.00	0.016	0.025	0.031	188.41**	4.40**
United States (1998)	11,812	11,127	1.00	0.013	0.021	0.020	56.90**	2.12**
United States (1999)	10,251	8,749	1.00	0.011	0.017	0.021	36.58**	1.77**
United States (2000)	10,309	8,944	1.00	0.008	0.014	0.013	26.00**	1.69**
United States (2001)	7,699	6,651	0.99	0.018	0.025	0.012	41.85**	2.46**
United States (2002)	12,391	16,257	1.00	0.012	0.017	0.019	85.63**	2.84**
United States (2003)	18,459	14,066	1.00	0.014	0.021	0.027	99.90**	3.25**
United States (2004)	20,579	15,798	1.00	0.009	0.011	0.008	22.54**	1.11
United States (2005)	22,843	17,228	1.00	0.019	0.017	0.005	56.06**	2.66**
Vietnam (1992)	4,800	8,005	0.99	0.102	0.148	0.171	1372.73**	16.79**
Vietnam (1998)	6,002	8,351	0.97	0.092	0.120	0.142	1154.57**	13.95**

* indicates 95% and ** indicates 99% significantly different from Benford.

Note that the November 1997 round of data from Mexico appears to be of much higher quality than the later rounds. This is interesting because the first round of the Mexican Progres data, the Encaseh, was a census used for targeting the households. This collected only the most easily collectable information and lasted only fifteen to twenty minutes. The later surveys, the Encels, were classic household surveys collecting detailed information on consumption, labor, health, income, and more. These surveys lasted one to two hours. These results suggest that longer surveys may tire out enumerators and/or respondents and can have a serious effect on data quality.

3.1 Male vs Female Respondents

Five of the surveys identify which household member responded to the questionnaire. If women in developing countries are in charge of livestock while men are in charge of crop production, one might think that women can answer questions about livestock more accurately, while men can answer questions about crop production more accurately. On the other hand, this ignores the fact that households endogenously chose which household member answers the survey questions. In Tables 8 and 9 we test this idea.

The test results are always much higher for males than females, but that is due to the higher sample size of male respondents. The four distance measures, on the other hand, are quite similar for males and females in the developing country data. These results suggest that there is not much of a difference in the overall quality of information given by male versus female respondents in developing countries. On the other hand, in the US data, according to the distance measures female respondents consistently give data less in accord with Benford's law. Perhaps women in the US are less involved with agriculture than are women in developing countries and so give less-well informed answers. Women in the US also give a much higher share of answers starting with the first digit of five than do men or women in other countries. Additionally, males and females do not seem to perform differently in answering questions related to crops versus livestock. We might interpret these results as showing that enumerators can get reasonable quality data from both male and female household members.

Table 8: Gender: Correlations (r), the m Statistic, Distances d^* and a^* , χ^2 Tests, and Kuiper V_N^* Tests between Benford's Law and Crop Quantities Produced

	Sex	# Surveys	# Obs	r	m	d^*	a^*	χ^2	V_N^*
Bangladesh (1996)	Male	3,546	6,481	0.93	0.066	0.086	0.039	389.42**	7.98**
Bangladesh (1996)	Female	976	876	0.94	0.058	0.084	0.035	63.01**	2.49**
Paraguay (2002)	Male	138	997	0.98	0.050	0.069	0.057	65.76**	2.24**
Paraguay (2002)	Female	43	294	0.96	0.057	0.074	0.037	22.64**	1.12
Paraguay (2002)	Male and Female	42	341	0.97	0.044	0.060	0.045	20.44**	1.24
Peru (1985)	Male	1,632	6,750	0.99	0.025	0.042	0.029	211.44**	3.57**
Peru (1985)	Female	500	1,300	0.99	0.030	0.045	0.044	42.50**	1.66**
Peru (1994)	Male	1,026	3,546	0.98	0.048	0.062	0.056	197.86**	3.97**
Peru (1994)	Female	180	555	0.96	0.047	0.081	0.079	57.27**	2.57**
Vietnam (1992)	Male	2,648	14,629	0.99	0.035	0.050	0.047	565.90**	7.00**
Vietnam (1992)	Female	1,390	6,307	0.99	0.033	0.048	0.038	210.37**	3.63**
Vietnam (1998)	Male	2,763	15,485	0.99	0.034	0.055	0.055	696.09**	8.06**
Vietnam (1998)	Female	1,671	8,922	0.98	0.036	0.055	0.049	402.43**	5.34**

* indicates 95% and ** indicates 99% significantly different from Benford.

Table 9: Gender: Correlations (r), the m Statistic, Distance d^* and a^* , χ^2 Tests, and Kuiper V_N^* Tests between Benford's Law and Animals Owned

	Sex	# Surveys	# Obs	r	m	d^*	a^*	χ^2	V_N^*
Bangladesh (1996)	Male	3,546	5,657	0.98	0.059	0.073	0.077	281.84**	6.34**
Bangladesh (1996)	Female	976	1,150	0.99	0.056	0.083	0.082	59.64**	3.60**
Paraguay (2002)	Male	138	414	0.99	0.069	0.078	0.066	17.77*	1.41*
Paraguay (2002)	Female	43	126	0.96	0.056	0.062	0.000	6.60	0.64
Paraguay (2002)	Male and Female	42	136	0.95	0.065	0.078	0.012	12.63	0.87
Peru (1985)	Male	1,632	6,422	0.97	0.078	0.094	0.102	563.81**	8.28**
Peru (1985)	Female	500	1,585	0.99	0.048	0.076	0.097	99.77**	3.55**
Peru (1994)	Male	1,026	2,915	0.99	0.048	0.059	0.061	105.57**	3.48**
Peru (1994)	Female	180	477	0.98	0.048	0.068	0.081	28.85**	1.88**
United States (1996)	Male	8,721	7,870	1.00	0.010	0.016	0.001	31.29**	1.88**
United States (1996)	Female	422	396	0.98	0.029	0.047	0.015	9.61	0.85
United States (1997)	Male	11,290	19,642	1.00	0.016	0.025	0.030	179.81**	4.29**
United States (1997)	Female	418	696	0.99	0.034	0.043	0.041	13.80	1.18
United States (1998)	Male	11,312	10,637	1.00	0.012	0.021	0.020	56.08**	1.99**
United States (1998)	Female	500	430	0.99	0.025	0.036	0.011	5.69	0.79
United States (1999)	Male	9,786	8,329	1.00	0.011	0.017	0.022	35.57**	1.80**
United States (1999)	Female	465	420	0.99	0.018	0.033	0.008	4.88	0.67
United States (2000)	Male	7,685	6,663	1.00	0.011	0.021	0.010	34.82**	2.50**
United States (2000)	Female	418	421	0.98	0.038	0.048	0.033	9.35	0.84
United States (2001)	Male	5,456	4,634	0.99	0.018	0.026	0.011	31.62**	2.12**
United States (2001)	Female	307	328	1.00	0.025	0.034	0.032	3.75	0.54
United States (2002)	Male	11,769	15,446	1.00	0.013	0.017	0.019	79.88**	2.72**
United States (2002)	Female	622	811	0.99	0.017	0.028	0.015	9.40	0.95
United States (2003)	Male	17,586	13,379	1.00	0.014	0.021	0.029	93.35**	3.26**
United States (2003)	Female	873	687	0.98	0.030	0.049	0.003	18.68*	1.62**
United States (2004)	Male	19,458	14,980	1.00	0.009	0.011	0.009	22.05**	1.08
United States (2004)	Female	1,121	818	1.00	0.017	0.024	0.010	5.01	0.69
United States (2005)	Male	21,694	16,395	1.00	0.009	0.017	0.006	54.59**	2.59**
United States (2005)	Female	1,149	833	0.99	0.031	0.037	0.017	12.15	0.96
Vietnam (1992)	Male	2,648	5,536	0.99	0.100	0.144	0.169	913.25**	13.57**
Vietnam (1992)	Female	1,390	2,469	0.99	0.107	0.156	0.176	464.11**	9.93**
Vietnam (1998)	Male	2,763	5,338	0.97	0.090	0.115	0.136	669.87**	10.94**
Vietnam (1998)	Female	1,671	3,009	0.98	0.095	0.131	0.152	443.34**	9.00**

* indicates 95% and ** indicates 99% significantly different from Benford.

3.2 Enumerators' Opinions and Data Fabrication

The survey from Bangladesh asked enumerators to judge both the accuracy of the respondents' answers as well as the seriousness and attentiveness of the respondent. Possible answers were: excellent, good, fair, not so bad, and very bad. In Table 10 we compare those surveys which were judged to be fair, not so bad, or very bad in terms of either accuracy or attentiveness (or both) with those that were good or excellent in both categories. As there are many more 'good' observations than 'bad', the test statistics for the good data are higher. More suggestively, the correlation with Benford is also higher in the 'bad' data while the m , d^* , and a^* distance measures are lower. If anything, the bad data seems to be better than the good data! Although this analysis is only applied to one data set, it suggests that enumerator evaluations of the respondents' data should be taken with a grain of salt.

When carrying out field work in May 2001 researchers working on the KIDS South Africa data found evidence that some of the 1998 households were fabricated by the enumerators. In 2004 it was concluded that only six clusters of data might have been fabricated and these have been removed from the version of the data available to the public. We have gained access to these deleted clusters and compare the data from those households to those in the rest of the survey. While only some of the households in the cluster may have been found to have been fabricated, all of them were dropped from the data set to be conservative.

Again, it is quite difficult to make comparisons due to the small number of observations in the clusters categorized as potentially being fabricated. Still, for the animal data, the 'fabricated' data performs much worse than the non-fabricated data according to all measures and tests. The results for the crop production data are less stark but are still suggestive of the low quality of data collected in the clusters in which the households were potentially fabricated.

It is interesting to note that data which enumerators qualify as 'good' is actually less in accord with Benford's law, as is data that enumerators fabricate. Perhaps this is due to enumerators' mistaken perception as to what realistic data ought to look like.

Table 10: Enumerator Opinion and Data Fabrication: Correlations (r), the m Statistic, Distance d^* and a^* , χ^2 Tests, and Kuiper V_N^* Tests between Benford's Law and Animals Owned and Crop Quantities Produced

	Quality	# Surveys	# Obs	r	m	d^*	a^*	χ^2	V_N^*
Crops Produced									
Bangladesh (1996)	'Bad'	1,384	2,096	0.95	0.052	0.070	0.029	94.33**	3.46**
Bangladesh (1996)	'Good'	3,107	5,204	0.92	0.070	0.093	0.042	360.55**	7.53**
Animals Owned									
Bangladesh (1996)	'Bad'	1,384	2,053	1.00	0.046	0.068	0.074	80.56**	3.84**
Bangladesh (1996)	'Good'	3,107	4,716	0.98	0.066	0.078	0.079	262.15**	6.16**
Crops Produced									
South Africa	'Fabricated'	283	62	0.99	0.118	0.163	0.204	14.60*	1.78**
South Africa	'Not Fabricated'	2,487	1,331	0.98	0.076	0.113	0.134	167.20**	5.22**
Animals Owned									
South Africa	'Fabricated'	283	50	0.70	0.156	0.253	0.061	29.87**	1.90**
South Africa	'Not Fabricated'	2,487	1,390	1.00	0.029	0.033	0.026	12.54	1.09

* indicates 95% and ** indicates 99% significantly different from Benford.

4 Implications for Theory and Practice

When respondents are asked for answers of which they are unsure they tend to estimate and round to ‘nice’ numbers. In addition, at least in this study, larger data sets collected by government statistical offices seem to be of lower quality than data collected by academic researchers. The Progres data in Mexico and the IFPRI data from Pakistan are particularly inconsistent with Benford’s law. These are issues which should not be ignored. Although sophisticated econometric techniques are available to deal with measurement error once it is identified, we should be much more careful and serious about both enumerator quality and designing questionnaires that elicit data with minimal respondent errors.

We have shown evidence that suggests data errors increase for more-diversified farmers. If certain questions are more prone to errors than others, then we will find that surveys for households which are more active in those areas will contain more errors, which can cause serious problems if the data is used in an estimation and inference context. We have also shown that there are certain questions which are more or less susceptible to response errors. Questions for which people tend to be unsure of the answer or for which people may have an incentive to answer dishonestly, such as donations to church or production of secondary crops, are more susceptible to errors.

Although the exact questions which lead to departures from Benford’s FSD distribution may be different in each country and situation, Benford’s law provides a simple means of testing for such irregularities in data. Researchers can easily and quickly test whether the variable that is of most interest to their research follows Benford’s law or exhibits errors. They can also test whether certain enumerators are collecting more irregular data, and whether households in certain clusters appear more irregular. The data which is presumed to be fabricated in the South African KIDS data was only noticed when researchers tried to go back to resurvey those households. Benford’s law could be used for a similar purpose, especially in cross-sectional data.

Glewwe & Dang (2005) show how having computers available for data-input at the district level, so that mistakes can be found more quickly and households reinterviewed sooner, can improve data quality. These computers could easily be programmed to include a Benford’s law component to test for the quality of responses to different questions and from different enumerators.

These results demonstrate why one should not consider only Benford’s law when evaluating enumerators or data sets. For example, while the data

collected in 1999 on crop quantities produced are much more in accord with Benford's law than that collected in 2002, the evidence suggests that this is not because the enumerators were better in 1999. The enumerators in 1999 seem to have only collected data on the most important crops, while the enumerators in 2002 collected data on many more crops, but for which there was more respondent error. Hence, while Benford's law suggests that the 1999 data contains less measurement error, other evidence suggests that this is because the 1999 data includes fewer crops. This is a warning against using Benford's law in isolation when judging the quality of a data set.

5 Conclusions

We have demonstrated how Benford's law can be used to detect data abnormalities arising both from questions that are difficult to answer and from enumerator errors. While econometricians and applied economists spend much energy correcting for measurement error in pre-existing data sets, they should also try to avoid it by detecting these problems early in the data-collection process.

There remains much room for future research on topics related to survey design and enumerator contracts. Can researchers articulate which types of questions and situations will lead to more accurate answers in general? For example, the following situations may affect error: use of interpreters, presence of non-family members during the interview, participation of more than one family member in the interview, and participation of female rather than male household members.

Furthermore, while Philipson & Malani (1999) show how random enumerator audits with prizes for accurate reporting can be used to decrease errors when direct data verification is possible, a contract has not yet been designed for data tests such as Benford's law which may be more prone to both Type I and Type II errors. These are important steps that should be taken to increase the quality of data production in addition to that of data consumption.

Finally, Scott & Fasli (2001) note that even in Benford's original paper only half of the data sets provide a reasonably close fit with Benford's law. Consequently, it seems possible that a family of data-based FSD distributions may be more compatible with observed data sets than Benford's distribution itself. To this end, Grendar et al. (2007) use information-theoretic methods to

develop a family of alternative Benford-like distributions. As these methods are refined, new tests of data quality may arise that provide insights on Benford's law and other scale invariant natural phenomenon. As a side note, our ongoing research using insights from this research concerning survey data has turned to the use of Benford's FSD to identify falsification in clinical trials (a life and death matter) and manipulation and collusion in market data.

References

- Benford, F. (1938), 'The law of anomalous numbers', *Proceedings of the American Philosophical Society* **78**(4), 551–572.
- Berger, A. & Hill, T. P. (2007), 'Newton's method obeys Benford's law', *American Mathematical Monthly* **114**(7), 588–601.
- Biemer, P. P., Groves, R. M., Lyberg, L. E., Mathiowetz, N. A. & Sudman, S., eds (1991), *Measurement Errors in Surveys*, New York: John Wiley & Sons.
- Carslaw, C. A. P. N. (1988), 'Anomalies in income numbers: Evidence of goal oriented behavior', *Accounting Review* **63**(2), 321–327.
- Cho, W. K. T. & Gaines, B. J. (2007), 'Breaking the (Benford) law: Statistical fraud detection in campaign finance', *The American Statistician* **61**(3), 1–6.
- de Marchi, S. & Hamilton, J. T. (2006), 'Assessing the accuracy of self-reported data: An evaluation of the toxics release inventory', *Journal of Risk and Uncertainty* **32**(1), 57–76.
- Diaconis, P. (1977), 'The distribution of leading digits and uniform distribution mod 1', *The Annals of Probability* **5**(1), 72–81.
- Durtschi, C., Hillison, W. & Pacini, C. (2004), 'The effective use of Benford's law to assist in detecting fraud in accounting data', *Journal of Forensic Accounting* **5**(1), 17–34.
- Geyer, C. L. & Williamson, P. P. (2004), 'Detecting fraud in data sets using Benford's law', *Computation in Statistics: Simulation and Computation* **33**(1), 229–246.

- Giles, D. E. (2007), ‘Benford’s law and naturally occurring prices in certain ebaY auctions’, *Applied Economics Letters* **14**(3), 157–161.
- Glewwe, P. & Dang, H.-A. H. (2005), The impact of decentralized data entry on the quality of household survey data in developing countries: Evidence from a randomized experiment in Vietnam. Unpublished Manuscript.
- Grendar, M., Judge, G. & Schechter, L. (2007), ‘An empirical non-parametric likelihood family of data-based Benford-like distributions’, *Physica A: Statistical Mechanics and its Applications* **380**, 429–438.
- Grosh, M. & Glewwe, P., eds (2000), *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 years of the Living Standards Measurement Study*, Washington DC: The World Bank.
- Groves, R. M. (1989), *Survey Errors and Survey Costs*, New York: John Wiley & Sons.
- Hill, T. P. (1995), ‘A statistical derivation of the significant-digit law’, *Statistical Science* **10**(4), 354–363.
- Hill, T. P. & Schürger, K. (2005), ‘Regularity of digits and significant digits of random variables’, *Stochastic Processes and Their Applications* **115**(10), 1723–1743.
- Leemis, L. M., Schmeiser, B. W. & Evans, D. L. (2000), ‘Survival distributions satisfying Benford’s law’, *The American Statistician* **54**(4), 236–241.
- Livio, M. (2002), *The Golden Ratio: The Story of Phi, the World’s Most Astonishing Number*, New York: Broadway.
- Morrow, J. (2007), Benford’s law, families of distributions, and a test basis. Unpublished Manuscript.
- Newcomb, S. (1881), ‘Note on the frequency of use of the different digits in natural numbers’, *American Journal of Mathematics* **4**(1), 39–40.
- Nigrini, M. J. (1996), ‘A taxpayer compliance application of Benford’s law’, *Journal of the American Taxation Association* **18**(1), 72–91.

- Nye, J. & Moul, C. (2007), ‘The political economy of numbers: On the application of Benford’s law to international macroeconomic statistics’, *The B.E. Journal of Macroeconomics* **7**(1 (Topics)).
- Philipson, T. & Malani, A. (1999), ‘Measurement errors: A principal investigator-agent approach’, *Journal of Econometrics* **91**(2), 273–298.
- Pietronero, L., Tosatti, E., Tosatti, V. & Vespignani, A. (2001), ‘Explaining the uneven distribution of numbers in nature: The laws of Benford and Zipf’, *Physica A: Statistical Methods and its Applications* **293**(1-2), 297–304.
- Raimi, R. (1976), ‘The first digit problem’, *American Mathematical Monthly* **83**(7), 521–538.
- Rodriguez, R. J. (2004), ‘First significant digit patterns from mixtures of uniform distributions’, *The American Statistician* **58**(1), 64–71.
- Schatte, P. (1988), ‘On mantissa distributions in computing and Benford’s law’, *Journal of Information Processing and Cybernetics* **24**(10), 443–455.
- Scott, P. D. & Fasli, M. (2001), Benford’s law: An empirical investigation and a novel explanation. Unpublished Manuscript.
- Stephens, M. A. (1970), ‘Use of the Kolmogorov-Smirnov, Cramer-Von Mises and related statistics without extensive tables’, *Journal of the Royal Statistical Society, Series B* **32**(1), 115–122.
- Sudman, S., Bradburn, N. M. & Schwarz, N. (1996), *Thinking About Answers*, San Francisco: Jossey-Bass Publishers.
- Varian, H. (1972), ‘Benford’s law’, *The American Statistician* **26**(3), 65.

A Appendix: <http://www.checkyourdata.com>

A website is available with an online utility into which you can easily input variables from any data set (as a comma separated value (csv) file) and see if your data is in accord with Benford’s law. At <http://www.checkyourdata.com>,

there is a tutorial explaining how to get data from Stata into the online utility and a video tutorial on the basics of using the utility. In addition, in the online utility itself, there is a help tab which gives more detailed information on the available commands.

When using the utility, researchers can either use the pull-down tab commands or can type commands directly into the command window. After inputting data, all of the tests and measures discussed in this paper are calculated. It is also possible to combine multiple variables, for example to look at corn, soy, wheat, and cotton production variables together. One can split variables as well, for example to look at data collected from male and female respondents separately, or by different enumerators separately. The program also can make graphs of the data and gives other summary statistics.

B Appendix: Extra Tables and Figures

Table B-1: Summary Statistics for 2002 Paraguay Data

Household Variables		
Variable	Mean	(Std. Dev.)
Theft Experienced	111,000	(336,000)
Gifts Given	306,000	(524,000)
Annual Income	28,300,000	(72,100,000)
Median Annual Income	9,046,000	
Family Size	5.6	(2.4)
Land Owned (hectares)	36.6	(95.8)
Respondent Variables		
Variable	Mean	(Std. Dev.)
Male	79%	
Age	52.2	(14.8)
Years of Education	4.9	(2.7)
Obs	223	

The relevant exchange rate is approximately 4,800 Guaranies to the dollar.

Table B-2: Correlations (r), the m Statistic, Distances d^* and a^* , χ^2 Tests, and Kuiper V_N^* Tests between Benford's Law and Animals Owned in 1999 and 2002

	Surveys	Obs	r	m	d^*	a^*	χ^2	V_N^*
Total in 1999	298	827	0.93	0.061	0.087	0.043	56.21**	3.09**
Enumerator 1	33	105	0.94	0.081	0.098	0.076	10.73	0.88
Enumerator 2	36	103	0.97	0.050	0.078	0.086	8.18	1.09
Enumerator 3	33	91	0.83	0.079	0.125	0.010	14.87	1.39*
Enumerator 4	37	94	0.78	0.099	0.152	0.027	16.24*	1.80**
Enumerator 5	27	78	0.90	0.068	0.109	0.065	9.35	1.13
Enumerator 6	8	29	0.86	0.071	0.121	0.026	6.43	0.61
Enumerator 7	32	93	0.96	0.039	0.062	0.016	4.85	0.60
Enumerator 8	10	31	0.93	0.147	0.184	0.178	9.15	1.37*
Enumerator 9	32	84	0.82	0.098	0.143	0.036	14.86	1.33*
Enumerator 10	22	50	0.61	0.184	0.246	0.047	20.44**	1.76**
Total in 2002	223	676	0.99	0.044	0.056	0.042	18.52*	1.14
Enumerator 1	71	211	0.97	0.030	0.058	0.006	7.58	0.83
Enumerator 2	75	223	0.96	0.051	0.083	0.034	15.81*	0.77
Enumerator 3	77	242	0.99	0.063	0.078	0.082	11.77	1.34*

* indicates 95% and ** indicates 99% significantly different from Benford.

Table B-3: Correlations (r), the m Statistic, Distances d^* and a^* , χ^2 Tests, and Kuiper V_N^* Tests between Benford's Law and Quantities Produced

	Surveys	Obs	Obs/Survey	r	m	d^*	a^*	χ^2	V_N^*
Total in 1999	298	1412	4.7	0.99	0.026	0.036	0.025	30.36**	1.21
Enumerator 1	33	159	4.8	0.90	0.065	0.098	0.023	20.80**	1.57*
Enumerator 2	36	139	3.9	0.98	0.033	0.064	0.057	6.40	0.85
Enumerator 3	33	153	4.6	0.97	0.087	0.052	0.074	13.61	1.23
Enumerator 4	37	162	4.4	0.97	0.048	0.064	0.051	6.84	0.71
Enumerator 5	27	136	5.0	0.98	0.052	0.083	0.087	11.12	1.06
Enumerator 6	8	43	5.4	0.78	0.084	0.144	0.084	9.21	0.71
Enumerator 7	32	149	4.7	0.96	0.095	0.112	0.042	12.15	1.42*
Enumerator 8	10	60	6.0	0.96	0.049	0.090	0.052	7.03	0.44
Enumerator 9	32	202	6.3	0.95	0.043	0.076	0.045	14.33	1.47*
Enumerator 10	22	89	4.0	0.96	0.055	0.069	0.045	5.55	0.78
Total in 2002	223	1632	7.3	0.97	0.050	0.065	0.051	101.34**	2.69**
Enumerator 1	71	516	7.3	0.97	0.031	0.057	0.042	28.20**	1.50*
Enumerator 2	75	556	7.4	0.98	0.049	0.070	0.061	37.58**	1.68**
Enumerator 3	77	560	7.3	0.94	0.073	0.092	0.050	67.92**	1.97**

*-95% and **-99% significantly different from Benford.