Quantifying the Qualitative

Information Theory for Comparative Case Analysis

Katya Drozdova • Kurt Taylor Gaubatz

Sensitivity Analysis—Entropy, Inference, and Error

Among the contributions of Claude Shannon's unifying theory of communication was to show that information could be transmitted and decoded accurately, essentially without error. Prior to Shannon, it had been taken for granted that the more information you were sending or the faster you were trying to send it, the greater the probability of error. Shannon's metric allowed for calculating the channel capacity that would be required for the transmission of any given information with "all but an arbitrarily small fraction of the errors" (Shannon, 1948, p. 20). Shannon's insight paved the way for the digital revolution that allows huge amounts of information to zip around without consequential error. The concepts of information and error have, therefore, been intimately linked from the beginning of the Information Age. In this chapter, we turn to the problem of understanding and handling error in the use of information theory for comparative case study analysis.

Here we need to return to a little of the research philosophy we presented in Chapter 1 (Figure 1.1). We can look at all case study work in three phases. The first two phases involve the theoretical conceptualization and design of the case study, the selection of cases, and the actual collection and coding of data (Yin, 2014). These steps are universal to all case studies and are not significantly changed by the use of information metrics, except to the degree that our approach requires setting up of a set of structured-focused comparisons and recording the data as binary {0, 1} measures. The real focus of our approach is the third phase: the systematic analysis of the case study results.

The analytic phase of the case study analysis can be further broken down into three steps. The first of these is to consolidate the case results into a single truth table. The second analytic step is to clearly identify the relationships between the independent variables and the dependent variable in the collected data. This is the primary leverage of the systematic information approach. The information metric provides an exact measure of the amount of information about the variation in the dependent variable that is contained in an independent variable for a given set of measures from a given set of cases.

The final step in the analytic process is to draw inferences from the results of the case study to the real world. This is a core challenge of both large-N and small-n studies but is clearly more acute for small-n work. How can we be confident that our research findings apply in the world?

A gap always exists between the selection of cases we have chosen to analyze and the larger world. In large-*N* statistics, we gain inferential leverage from the possibility that the data set is a reasonably random representative selection from the population. If the data are adequately random, the central limit theorem and other foundational concepts of statistical analysis can help us understand how to infer from a random sample to the population. As we argued in Chapter 1, these standard statistical procedures are often unreliable or inapplicable in small-*n* analysis.

The ability to draw valid inferences from the selected case studies still depends on analytic or contingent generalizations and assessments about the character of case selection. Information theory provides a concrete set of measures with relative magnitudes to systematically guide assessment. The information metric is a rigorous and reproducible basis for understanding the relationship between independent and dependent variables within the selected set of cases. The information metric offers a significant improvement over traditional case study methodologies, but it isn't a magic wand that can overcome all of the limitations of small-*n* analytics.

Traditional small-*n* case analysis often depends on one's qualitative assessment of the connection between the cases and the larger population. Ideally, this assessment should be driven by theory and verified by evidence, but in the absence of systematic tools for evaluating case results, there is a risk of simply making a subjective judgment call. If scholars believe that their cases are representative, they would expect the same information conditions to hold in the world at large. But, there may be many sources of uncertainty in complex broader real-world problems. We may also believe that our cases are especially informative. They may be "most-likely" or "least-likely" cases (Eckstein, 1975; George & Bennett, 2005, p. 121). They may be selected for theoretical purposes in other ways that increase our confidence that they can help us learn about the more general phenomena.

The inferential leverage that comes from the ability to observe the context of the data and the relationships within them is one of the advantages of working with small-*n* analysis. This approach falls within "contingent generalization" (George & Smoke, 1974). George and Bennett (2005) put contingent generalization at the center of case-based techniques. Other scholars have advanced similar arguments for the benefits of case study methods. Ragin (1987, p. 31) uses the term *modest generalizations* to characterize the inferential leverage of case study research, and Yin (2014) uses the term *analytic generalizations.*

Improving the first phase of the analysis step, making the analysis of the cases themselves more systematic and rigorous, does not solve all small-*n* challenges, but it can provide a firmer foundation from which to make inferences. Moreover, these information techniques give us new tools for assessing the impact of the individual cases and coding on our results. In this way, we can better understand the sensitivity of the findings to inclusion or exclusion of

specific cases or to errors in data collection or coding. Understanding the sensitivities can also give us important insights into the data and into the character and quality of our inferences.

Confidence Intervals and the Information Metric

Anyone coming from a large-*N* background will be familiar with the notion of confidence intervals, or credible intervals if you are of the Bayesian persuasion. In large-*N* statistical work, we take a sufficiently large random sample from a population and then calculate a range around a point prediction in which we have high statistical confidence that the true quantities of interest will be found, based on the observed values in the sample. In other words, we can find a probability that the estimated interval contains the unknown parameter of interest. This approach is a strong tool for gaining leverage on the inference problem in large-*N* research.

There are several ways to calculate confidence intervals for the entropy measure (Esteban & Morales, 1995; Roulston, 1999; Sveshnikov, 1978, p. 288). We are not going to recommend them here. There are four reasons we think these statistical tools are not appropriate for the kinds of case studies we are addressing in this book.

In the first place, our goal is to provide a straightforward approach to case study analysis that is manageable for students and practitioners with widely varying quantitative backgrounds, including scholars who may not be interested in delving into more in-depth statistics. The conceptual workarounds and calculations required for estimating confidence intervals for a small number of cases involve significantly more complex calculations. Thus, approaches to confidence intervals for information measures are beyond the scope appropriate for the introductory level and the intended audience for this book. A single mutual information number for each variable is more elegant and ultimately sufficient for comparative case analysis. This is a simple informative tool to add to a primarily qualitative analysis toolkit for case study research.

Second, these techniques are often designed for the much more challenging process of estimating entropy scores for research problems involving variables that can take on many different or even continuous values, and thus where mutual information scores cannot be directly calculated but must be estimated even in the sample space (Paninski, 2003). In our application for comparative case studies, information metrics do not make point predictions for which confidence intervals are typically designed. Rather, mutual information provides a direct measure, based on the analyzed cases, of the relative information content of the different independent variables about case outcomes of interest. Confidence intervals could provide relative ranges but only at a high computational cost and with dubious additional leverage. In our approach, limiting the analysis to binary coding of

both the independent and dependent variables gives us a complete picture of the conditional probabilities and the ability to calculate entropy and mutual information directly and precisely, with straightforward leverage for interpreting the comparative case results.

Third, the construction of confidence intervals requires an assumption that the cases are a random selection from some larger population. This, as we discuss in Chapter 3, is not often the case for small-*n* studies, where the cases are typically selected for particular theoretical purposes, not for statistical significance. The quality of the confidence intervals will itself depend on our confidence that the cases were selected at random.

Finally, small-*n* case studies remain small *n*—part of a methodology designed explicitly for in-depth qualitative research and particular types of contingent and analytic generalizations that are conceptually distinct from and complementary to traditional statistical techniques. By design, typically there will not be that many cases, which means that even if we are satisfied that the cases are a random selection, the confidence intervals are still likely to be relatively wide. Their primary function will then be to verify that the mutual information number falls inside the confidence interval calculated at a reasonable confidence level (e.g., 99%, 95%, or 90% at least) to establish statistical significance of the results. While helpful as an additional verification by a different method that adds confidence to our overall knowledge, it is not statistical significance that such case studies typically seek.

The benefit of applying information theory to comparative case study is that it helps us precisely understand the information content of the cases we have. It cannot magically overcome all of the challenges of drawing inferences from a very small data set.

How, then, are we to think about inferences with the use of information theory to make small-*n* analysis more systematic?

A real-world example can help demonstrate the procedures for assessing case study results in terms of these sensitivities. This example is about environmental policy.

Analytic Leverage for a Study of Environmental Incentives

The reduction of municipal garbage production has become an important environmental goal for a number of reasons, including the limited availability of suitable landfill space and the impact of toxins in waste disposal strategies. In the early 1990s, a group of researchers at Duke University undertook a study of residential garbage on behalf of the Environmental Protection Agency (EPA) to look at the effects of unit pricing disposal fees on the per capita quantity of municipal solid waste (Miranda & Aldy, 1996). The study authors developed a very careful

structured-focused comparison asking a consistent set of research questions to assess 14 explanatory variables across nine communities. Although the authors provide a summary table for each individual variable and a more detailed descriptive summary for each case, they provide no systematic table to summarize the study as a whole and to provide a direct way to visualize or assess the overarching relationships between the independent and dependent variables.

To that end, and for our purposes here, we have translated their results into a set of binary {0, 1} measures of 11 independent variables and 2 dependent variables (3 of the independent variables that had perfectly overlapping values were consolidated—*California, High housing costs,* and *High garbage fees*). The first dependent variable, shown in Table 6.1, is simply whether or not the amount of residential solid waste was reduced after the introduction of unit pricing

Table 6.1   Truth Table for Unit Pricing and Waste Reduction Data

	California (High Housing Costs and Garbage Fees)	Unit Pricing Pre-1990	Large Cities (Pop > 100,000)	Urban Communities	Median Household Income > $35k	Waste Collection Primarily Private	Mandatory Rather Than Opt-In Recycling	Recycling Fee	Use of TV/Radio to Publicize	Tipping Fee > $30/ton	Fee per Gallon/Week < $0.05	Total Waste Volume Reduced
Downers Grove	0	0	0	0	1	1	1	0	1	1	0	1
Glendale	1	0	1	0	0	0	0	0	0	0	1	1
Grand Rapids	0	1	1	1	0	1	0	1	1	1	1	0
Hoffman Estates	0	0	0	0	1	1	1	0	0	1	0	1
Lansing	0	1	1	1	0	0	1	1	1	1	0	0
Pasadena	1	0	1	0	1	0	1	0	0	0	1	0
San Jose	1	0	1	1	1	1	1	0	1	0	0	0
Santa Monica	1	0	0	0	1	0	1	0	1	1	0	1
Woodstock	0	1	0	0	0	1	1	0	0	1	0	1

The binary values shown here were coded by the current authors. California was substituted for the study variables for high garbage fees and housing values, as high garbage fees (over $2.00 per month) and high housing prices (median house value over $150,000) were both perfectly coterminous with the division between the four California cities and the five midwestern cities. More detail on these data and our revisions can be found at http://study.sagepub.com/drozdova.

(i.e., paying fees according to the amount of waste generated). The second dependent variable, to which we will return a little later, is based on the reduction in solid waste sent to a landfill or incinerators. Table 6.2 summarizes the results of the study with the first dependent variable.

The total uncertainty in these data is very high. There is a near-even split in the outcomes with five cases of successful waste reduction and four unsuccessful cases. Total uncertainty (H) is 0.99, which means that without the information from our independent variables, we have nearly perfect uncertainty about which communities were going to successfully reduce their waste volume.

Table 6.2 shows the information structure for these 11 independent variables, sorted from the most to the least informative. Having a large population or being an urban city are the most informative factors for predicting whether unit pricing will reduce solid waste, although both are negatively related to this outcome. Smaller populations and nonurban municipalities are the most likely to have reduced their solid wastes. These two factors have significant overlap, an issue we will turn to in Chapter 7. A recycling fee is one of the few variables over which policy makers have control that is informative about the outcome. As would be expected, the relationship is negative, so having a recycling fee means less diversion of recyclables from the solid waste stream. But, this finding probably merits more investigation since there are only two communities with recycling fees (Lansing and Grand Rapids), and they implement it in quite different ways. These important implications are overlooked in the report for the EPA

Table 6.2 Entropy Metrics for Waste Reduction Data

	Conditional Entropy	Mutual Information	Direction
Large cities	**0.40**	**0.59**	Negative
Urban communities	0.43	0.56	Negative
Recycling fee	0.67	0.32	Negative
Use of TV/radio to publicize	0.90	0.09	Negative
Unit pricing pre-1990	0.92	0.07	Negative
Tipping fee > $30/ton	0.92	0.07	Positive
Fee per gallon/week < $0.05	0.92	0.07	Negative
California	0.98	0.01	Negative
Median household income > $35k	0.98	0.01	Positive
Waste collection primarily private	0.98	0.01	Positive
Mandatory rather than opt-in recycling	0.99	0.00	Positive

but point to potentially important areas for further investigations and policy tools for effecting change. The other variables are essentially noninformative. There is very little reduction in uncertainty about the outcome that comes from knowing their values.

Already we have gained some analytic leverage for these cases. Neither of these insights about the challenges of encouraging waste reduction in large and urban cities was addressed in the original report. We can go further with this example, however, by also looking at the effects of the different cases on the mutual information scores.

The Information Metric and the Problem of Inference

The reason for analyzing a set of cases is almost always to draw inferences about the world at large and often about informing policy or practical decisions. We select a set of cases that we believe can tell us something about the larger world, especially when—following the principles set out in Chapter 3—they are systematically drawn and scientifically analyzed within a theoretic framework. In the solid waste unit pricing study, it is clear that the EPA was looking for an actionable understanding of the interventions or other best practices that might improve waste management outcomes in other cities across the United States. These nine communities were studied in the hopes of generating inferences about things that could be done in other communities.

Our argument is that even just applying the information metric to a set of cases is already a significant enhancement for the subjective analysis that has heretofore been the primary approach to assessing case study results. As we saw earlier, information analysis gave us new insights about the data collected for understanding the factors that might contribute to municipal solid waste reduction. Beyond that, however, the information approach gives us additional tools, of only slightly increased complexity, that can help us assess some of the sources and the extent of potential inferential error.

Sensitivity Analysis

The very fact that one is using comparative case studies is a strong indicator that the domain of interest has the potential for analytical error. If we were working with a mechanically deterministic system, there would be no need for comparative cases. One would just trace through the mechanism and then perfectly characterize the chain of causes and effects. Instead, we use small-n or large-N studies to help us understand the broader world through a closer study of some smaller part of it. This introduces the possibility of inferential error.

All small-*n* studies have to cross this inferential gap. For this, they will require some degree of subjective assessment. Information metrics cannot eliminate this gap. But to the degree that they provide a more structured understanding of the data we have, they can discipline the subjective move across the inference gap and strengthen it with some objective indicators of the information content of the results drawn from that data based on the analyzed cases and factors. Beyond the basic information metrics, we can get a better sense of our data and results by looking at how sensitive the information metrics are to possible variation.

Sensitivity analysis can improve our understanding of the information contours in the cases we do have. It can provide additional insights and fundamentally strengthen a well-designed case study. The most straightforward and intuitive way to do this is to look at the effects of each of the individual cases on the entropy and mutual information measures.

Dropped-Case Analysis

Our first approach is to look at the effect of removing each individual case from the information analysis. How sensitive is the mutual information score to the inclusion or exclusion of any particular case? Cases that have a particularly large effect are worth exploring more closely. They may be anomalies for which it could be fruitful to look more carefully to try to identify what makes them different. It could also be, however, that outliers are mistakes. If a case does not fit with the other data, it is more likely that it was mismeasured or miscoded. Of course, it is important not to make post hoc judgments to reject inconvenient data. Indeed, these may be the most critical cases that warrant the most analytic care. At a minimum, creating a truth table can make the data patterns explicit to help identify such mismatched cases.

To assess the sensitivity of the results to the particular cases, we simply rerun the analysis repeatedly, leaving one case out each time. While this analysis is conceptually straightforward, to do it efficiently probably requires the use of more sophisticated computational tools. You could do it with Excel by simply duplicating your worksheet several times and dropping a different case each time. This could be a little tedious but with very few cases is probably manageable.

Alternatively, this is a relatively easy task for a scripted computer language such as R, which is rapidly becoming the dominant computer language for statistical analysis. R has a steep learning curve, but it can be effectively used by beginners (Gaubatz, 2014). In Appendix B and on the website that accompanies this book (http://study.sagepub.com/drozdova), we have provided the R code for doing all of the analyses in this book, including automating dropped-case analysis.

Figure 6.1 is a straightforward dot chart (Cleveland, 1993) showing the individual effects of dropping each case on the mutual information score for several of the independent variables. This exercise provides a number of potentially important insights. We see here,
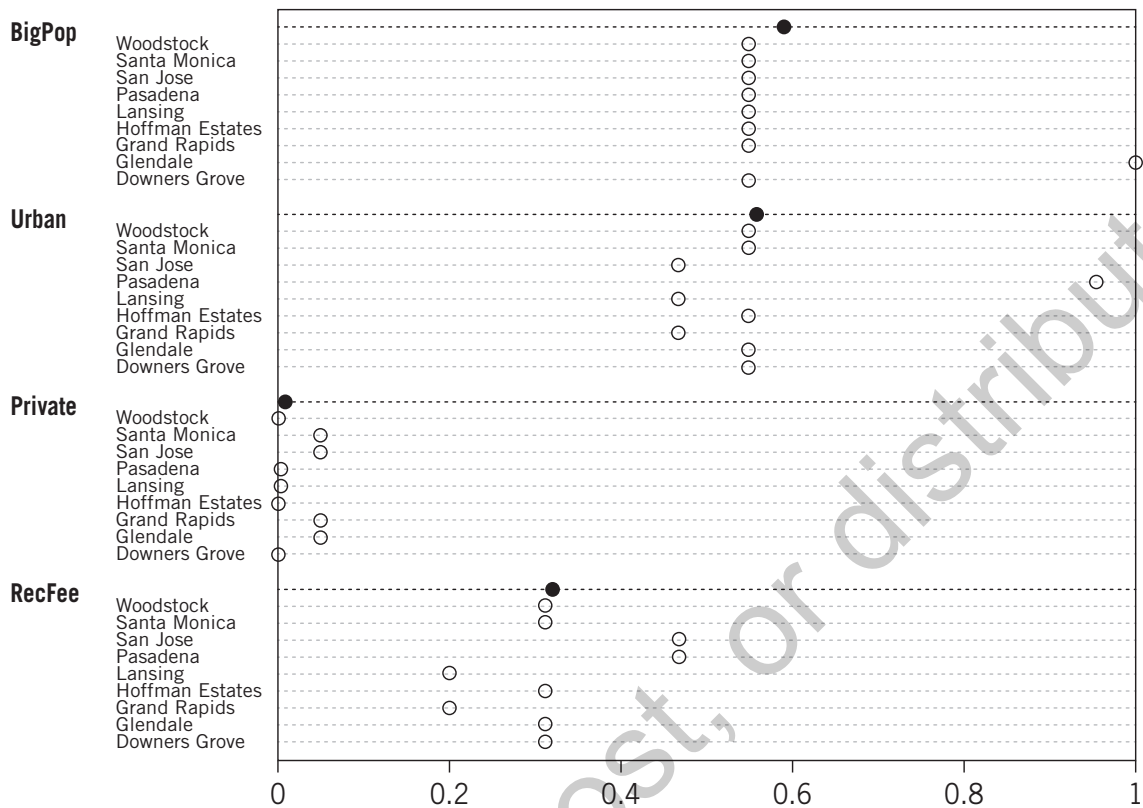
This figure shows the effect on the entropy measure of dropping each case for a selected set of variables. The solid dots represent the mutual information score when all cases are included. The open dots show the mutual information score when each corresponding case on the y-axis is dropped from the analysis. Variable names have been abbreviated and include Large cities (BigPop), Urban communities (Urban), Waste collection primarily private (Private), and Recycling fee (RecFee).

Figure 6.1 Dropped-Case Analysis

for example, that the mutual information score for the variable indicating that garbage service is private is very stable. Its very low contribution to uncertainty reduction is relatively unaffected by the exclusion of any particular case. On the other hand, we can see from the top panel of Figure 6.1 that Glendale is a critical outlier for the *Big population* variable. Glendale is the only large city with a successful outcome. Without Glendale, population size would be perfectly informative of successful solid waste reduction programs. All the small cities in this study succeeded in reducing their waste output. None of the large cities, except for Glendale, pulled this off. This policy-relevant implication is overlooked in the report for the EPA but points to a potentially fruitful area for further investigation. How Glendale succeeded where all other large municipalities failed is an important puzzle in these data.

Similarly, we can see that Pasadena has an outsized effect on the *Urban* variable. No urban area succeeded in reducing its solid waste production. All the nonurban municipalities had reductions with the singular exception of Pasadena. A more careful examination of what might have gone wrong in Pasadena could be very important.

The other variables exhibit less variation in the dropped-case analysis. Even so, we can see in Figure 6.2 that Glendale again stands out for its low tipping costs (the disposal fees imposed by landfills) but relatively high fees for waste producers. The implications of that combination may be another area ripe for further analysis.

We can summarize the effects of the dropped-case analysis in terms of either the cases or the variables. Table 6.3 shows the mean and the maximum change in the mutual information score for each variable. We can see here that the mean variation in the information scores is relatively small for all of the variables. The maximum variation is more variable. The California, high income, private, and mandatory recycling information scores change very
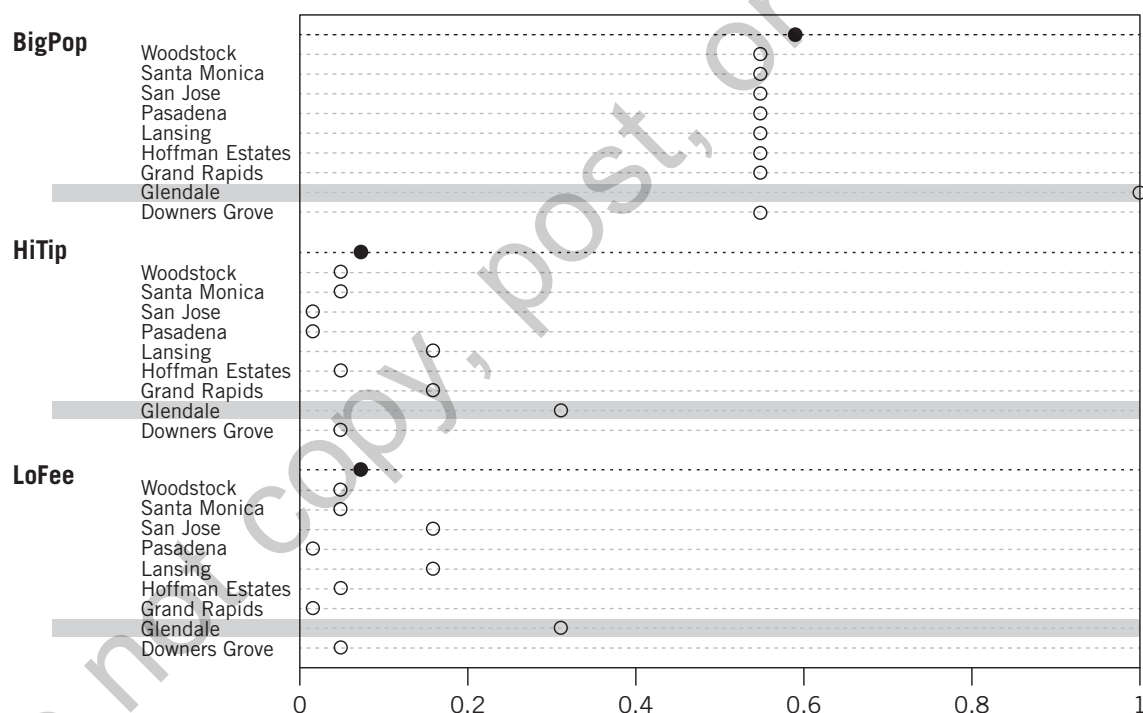
This figure highlights the effect of dropping each city on the mutual information scores for the three variables where the city of Glendale has a particularly large effect. If Glendale had not been included in the study, the information effects of high tipping costs (HiTip) *and low garbage fees* (LoFee) *would be much higher. Without Glendale in the study, the variable for being a large city* (BigPop) *would be perfectly informative about the success of solid waste reduction programs.*

Figure 6.2   Garbage in Glendale

Table 6.3 Mean and Maximum Variation in Mutual Information Scores From Dropping Individual Cases

Variable	Mean Variation	Maximum Variation
California	0.02	0.04
Unit pricing pre-1990	0.07	0.24
Large cities	0.08	0.41
Urban communities	0.08	0.40
Median household income > $35k	0.02	0.04
Waste collection primarily private	0.02	0.04
Mandatory rather than opt-in recycling	0.03	0.14
Recycling fee	0.06	0.15
Use of TV/radio to publicize	0.08	0.26
Tipping fee > $30/ton	0.07	0.24
Tipping fee per gallon/week < $0.05	0.07	0.24

This table shows the mean and maximum change by variable in the mutual information score when each case is dropped.

little with the inclusion or exclusion of any one case. Each of the rest of the variables has about the same amount of mean change, between 0.06 and 0.08. They are all relatively stable, with the exceptions of the large effects we see in Figure 6.1 from Glendale and Pasadena for the *Large cities* and *Urban communities* variables, respectively.

Table 6.4 shows the mean change in the mutual information score for each case. That is, it shows the mean effect of dropping one case across all of the variables. Here we see that Glendale and Pasadena are the most unusual. Their inclusion in this study has significant effects on the information metrics for several of the independent variables. Again, a closer look into these two cases might be particularly appropriate for clarifying the results and enhancing comparative rigor. Such clarity will be especially important if policy conclusions are to be drawn from the studies.

Dropped-case analysis allows us to see the impact of each case on our analytic conclusions. Importantly, these are insights that do not depend on randomized case selection. Ultimately, whether a given case warrants additional attention has to be a substantive question rather than a simple response to the dropped-case analysis, but understanding the impact of each case can be an important starting point for thinking through that issue.

At the end of the day, it remains critical to remember the principles of case selection outlined in Chapter 3. It is interesting and valuable in terms of potential policy implications to see the effects of each case, and this analysis can, again, help guide a more nuanced and subjective assessment of the cases. It is essential not to simply engage in post hoc case selection. The move from specifying the

Table 6.4 Mean and Maximum Change in All Variables for Each Dropped Case

City	Mean Change in Mutual Information Without City	Maximum Change in Mutual Information Without City
Downers Grove	0.02	0.10
Glendale	0.11	0.41
Grand Rapids	0.06	0.12
Hoffman Estates	0.02	0.04
Lansing	0.05	0.12
Pasadena	0.10	0.40
San Jose	0.06	0.15
Santa Monica	0.03	0.10
Woodstock	0.04	0.24

This table shows the mean and maximum effect of dropping each case on the information scores across all of the variables.

relationships between variables in the set of analyzed cases to inferring that those same relationships hold in the general population requires confidence that the cases are appropriately selected.

Outcome Coding Sensitivity

The dropped-case analysis focuses on the effects of each case on the mutual information scores for the relationship between the outcome and each independent variable. Another sensitivity measure would be to look at how the results change if the outcome scoring changes. This allows us to consider the critical question of whether the outcomes have been properly coded.

As discussed earlier, the environmental study offers a second dependent variable, looking at the reduction in landfill and incinerated solid waste. For the first dependent variable, the overall amount of solid waste, turning the outcome into a binary {0, 1} measure was straightforward: Cities that reduced their solid waste output were coded 1, and others (i.e., those that did not change or those that increased their solid waste output) were coded 0. The reduction in landfill/incinerator use is a little more subjective to code as binary since all of the studied cities had some level of reduction. Subject matter experts may have strong views about the appropriate cutoff threshold for identifying a successful or unsuccessful solid waste reduction outcome. We could start with two different cutoff levels: a greater than 10% reduction or a greater than 25% reduction. These two coding conventions, along with the original continuous outcome, are set out in Table 6.5.

Table 6.5 Landfill and Incinerated Waste Reduction Outcomes

City	LIR	LIR10 (10% or Greater Reduction)	LIR25 (25% or Greater Reduction)
Downers Grove	–0.24	1	0
Glendale	–0.34	1	1
Grand Rapids	–0.22	1	0
Hoffman Estates	–0.39	1	1
Lansing	–0.50	1	1
Pasadena	–0.60	0	0
San Jose	–0.22	1	0
Santa Monica	–0.70	0	0
Woodstock	–0.30	1	1

This table shows the actual reduction in landfill/incinerator solid waste with two different cutoff points for the binary coding. The first, LIR10, identifies those communities that decreased their solid waste by more than 10%. The second, LIR25, identifies those communities that achieved a greater than 25% reduction. LIR means landfill/incinerator reduction.

The 10% landfill or incinerator reduction cutoff yields an outcome variable with seven successes and just two failures. This yields a lower degree of outcome uncertainty, $H(Y) = 0.76$, because success would usually be a pretty good guess about the outcome. The 25% landfill or incinerator reduction cutoff yields an outcome variable with four successes and five failures. This is a nearly perfect degree of outcome uncertainty, $H(Y) = 0.99$. The mutual information (uncertainty reduction) metrics for these alternative outcome variables are displayed in Table 6.6 with abbreviated independent variable names.

For the 10% reduction breakpoint, being in California (which, remember, means high fees and high housing costs), having mostly private collection programs, and being a high-income city are the most informative independent variables. High income and being in California make success less likely, while private programs increase the likelihood of success.

We can see, however, that these results are quite sensitive to the choice of cutoff point for coding a successful outcome. If we increase the cutoff point to a 25% reduction, Downers Grove, Grand Rapids, and San Jose fall out of the success category. At the 25% level, there are not many strong predictors. High-income communities are again less likely to reduce their garbage output. This is intuitive, as they are less sensitive to the usage fees that are the subject of this study. More surprising is that the use of media to publicize the programs is also associated with reduced program efficacy. This variable was not very informative for the 10% reduction outcome variable or for the original outcome variable, which was coded 1 for any

Table 6.6 Information Metrics for LIR10 and LIR25

	LIR10	LIR10	LIR25	LIR25
Variable	Mutual Information	Direction	Mutual Information	Direction
CA	0.32	Negative	0.09	Negative
Private	0.32	Positive	0.01	Negative
HiInc	0.23	Negative	0.23	Negative
Pre-1990	0.15	Positive	0.07	Positive
Urban	0.15	Positive	0.02	Negative
MandRec	0.09	Positive	0.00	Negative
RecFee	0.09	Negative	0.00	Positive
HiTip	0.03	Positive	0.02	Positive
LoFee	0.03	Negative	0.02	Negative
BigPop	0.00	Positive	0.00	Negative
Media	0.00	Positive	0.23	Negative

reduction in total solid waste output (Table 6.2). This set of findings clearly warrants more detailed study. If it holds up, it would be an obvious and more straightforward area for policy change than many of the other variables.

Several of the explanatory variables in Table 6.6 change direction from the 10% case to the 25% case. This is not surprising since in every instance at one or the other end, the mutual information is very close to zero, so it can easily flip above or below zero to change the sign.

Ultimately, interpreting Table 6.6 requires some degree of substantive expertise. What counts as a large or very large reduction in landfill or incinerator waste should be determined contextually with theoretically, empirically, and experientially informed expert assessment. Indeed, it could be that the explanatory factors for a very large reduction in landfill or incinerator waste are different from the explanatory factors for just a normal reduction in waste. In any case, we see from Table 6.6 that several of these results are sensitive to this coding choice. The analyst will likely want to draw on the original and more in-depth data from the case studies, which remain available to provide the advantages of the more nuanced understanding available in the small-*n* environment.

The same techniques used for the sensitivity of the information metrics to outcome coding can also be used to assess the sensitivity to the coding of the independent variables. This can be labor intensive, depending on the number and character of the independent variables.

It will be most appropriate for those factors where the criterion for making the binary coding {0, 1} is more ambiguous. Here, again, substantive expertise is required to make judgments about the appropriate cutoff points and their meaning. And, as discussed in Chapters 3 and 4, whichever meaningful cutoff points are selected, the binary coding criteria should clearly and measurably define two mutually exclusive categories so that the underlying data are precisely classified.

Conclusion

Information metrics can give us significant insights into the relationships within a set of small-*n* data. Considering the sensitivity of the entropy metrics to the inclusion or exclusion of particular cases, as well as the sensitivity of the metrics to particular coding decisions, can help cast further light on the data and the conclusions in a structured-focused case study.

In the EPA garbage study, the application of information metrics both made the study results more systematic and rigorous and provided additional insights that were not included in the original report. We saw that waste reduction appears to be considerably more difficult in high-income, large, urban cities. But, analyzing the sensitivity of these results to the specific cases helped us quickly identify the cases that stand out as having a large impact on the results. We saw, in particular, that the Glendale case may merit further attention as the only large-population city that was successful in reducing its solid waste output. But, Glendale is also coded as nonurban. All of the nonurban municipalities were successful in reducing waste, with the exception of Pasadena. That case, too, may require a closer look.

This example study also highlights that further insights might depend on understanding the patterns of how the different independent factors combine. Thus far, these measures have been used in a strictly bivariate sense: Entropy is a measure of uncertainty, and mutual information is a measure of the information value of each individual factor for understanding an outcome. It is likely that we would also benefit from knowing about the relationships between the independent variables. For this purpose, we can push the information metric further by combining it with the technique called qualitative case analysis (QCA). That is the subject of the next chapter.

Additional resources are provided at http://study.sagepub.com/drozdova.