When Big Legal Data isn't Big Enough:

Limitations in Legal Data Analytics

Robert Parnell, CFA, LLB

SettlementAnalytics™
THE SCIENCE OF LITIGATION

September 2016

"Legal data mining can be a data minefield. Tread carefully."

EXECUTIVE SUMMARY

The mass harvesting and storage of court records and other legal data provides an opportunity for corporate litigants and their legal counsel to complement decision making with legal data analytics. But without the use of proper statistical methods, the analysis of data can be invalid or misleading.

Legal professionals do not analyze quantitative legal data merely to observe historical data facts, but rather in an effort to draw a meaningful inference about the present, and to make decisions. Although it is widely understood, it is sometimes forgotten, that we cannot go reliably from past data to some present insight because the past is only a sample of what *could* happen and often it is a very imperfect one.

The central problem is that not all samples of legal data contain sufficient information to be usefully applied to decision making. By the time big data sets are filtered down to the type of matter that is relevant, sample sizes may be too small and measurements may be exposed to potentially large sampling errors. If Big Data becomes 'small data', it may in fact be quite useless.

To be of value in real world decisions, legal data analytics must be able to distinguish between the inherent randomness in historical data samples and statistically meaningful legal track records. This necessarily requires the application of inferential statistics.

In this article we provide legal professionals with an introduction to basic inferential statistical methods so that they will be better able to determine when 'Big Legal Data' is big enough in practice. The reader is introduced to key concepts at an introductory level and a number of online analytical tools are used to show how counsel can evaluate the statistical merit of their data.

Example analyses illustrate how to quantify the uncertainty in the measurement of judicial decision making, and how to determine if a law firm's track record is statistically significant relative to its peer group. The results of statistical analyses are presented graphically.

Using basic inferential statistics such as the methods outlined here, legal professionals will be able to interrogate the statistical validity of their data and evaluate the significance of various quantitative legal metrics.

In practice, although the volume of available legal data will sometimes be sufficient to produce statistically meaningful insights, this will not always be the case. While litigants and law firms would no doubt like to use legal data to extract some kind of informational signal from the random noise that is ever-present in data samples, the hard truth is that there will not always be one. Needless to say, it is important for legal professionals to be able to identify when this is the case.

Overall, the quantitative analysis of legal data is much more challenging and error-prone than is generally acknowledged. Although it is appealing to view data analytics as a simple tool, there is a danger of neglecting the science in what is basically data science. The consequences of this can be harmful to decision making. To draw an analogy, legal data analytics without inferential statistics is like legal argument without case law or rules of precedent — it lacks a meaningful point of reference and authority.

If we are going to examine legal decisions using the quantitative analysis of data, we cannot go halfway. We must make an allowance for the role of inferential statistics – only then will we know if the data have anything to say. With the use of appropriate statistical methods and careful attention to the complexities of data analytics, corporate litigants and Big Law can benefit from this new frontier in Big Data.

Keywords and phrases*:* Legal data analytics, big data, data analytics, statistics, confidence interval, hypothesis test, data science, data mining, legal tech, analytics, predictive analytics, law, Big Law, cognitive bias, data visualization.

When Big Legal Data isn't Big Enough:

Limitations in Legal Data Analytics

Robert Parnell, CFA, LLB
September 27, 2016

Introduction

Much has been made recently of the application of Big Data to the practice of law and litigation in particular. This is the use of very large data sets gleaned from thousands of historical court records to shed light on present legal conflicts. A number of uses have been discussed such as the quantitative analysis of law firm win rates as an aid to counsel selection, or the analysis of individual judges to gauge the probability of a favorable decision. It is an exciting time for the legal profession and the potential to add value through the analysis of data is significant.

Court records can be a valuable source of insight, and if we can analyze them en masse and with great speed, doubtless that is a good thing too. But in the newfound rush to apply 'Big Legal Data' to legal decision making, there is a risk of overlooking a number of important quantitative issues, the neglect of which could result in invalid or misleading analyses. These issues include the age-old concerns from the field of statistics such as the potential for bias in data samples, measurement error and the question of statistical significance among others.

Done well, there is much upside to the quantitative analysis of law, but without proper consideration for its statistical complexities, there is a danger of it being reduced to 'fun with numbers' – amusing, perhaps even interesting, but potentially quite useless or worse.

In this series of articles, we take a close look at some of the quantitative problems that litigants and counsel will encounter as they seek to employ Big

Legal Data. We will explore how these problems can act to undermine the value of data and what can be done about it. Along the way, we will discuss a number of analytic tools and concepts that GCs and outside counsel can employ to evaluate the statistical merit of their data samples. The articles will address some technical issues, but the aim is to present these in an introductory manner which should be accessible to legal professionals without a quantitative background.

This first article investigates the question of when legal data becomes quantitatively meaningful for decision-making purposes. We examine when the data sample pertinent to the legal problem at hand is big enough to be statistically significant and the degree to which we may have confidence in any measurement. In short we ask, when is Big Legal Data big enough?

Data Signal Or Data Noise?

The Internet has been replete with a wide variety of Big Data services for many years, and its extension to the field of law was a natural next step. But with so many data portals now available on the web, there is first a danger of drawing a false equivalence in how they all add value. There is an important distinction to be made if we are to recognize and overcome the problems that can arise in relation to legal data analytics.

Many Big Data services deliver value merely by providing searchable access to specific data targets in a large and comprehensive data set. The most obvious of these is Google, but there are many examples of this needle-in-a-haystack value contribution, such as real estate search engines and hotel reservation websites. In these instances, the search target itself is the subject of interest. While further analysis may be possible, generally the consumer is not using data about one thing to reach a quantitative conclusion about something else.

However, such is not typically the case with the analysis of legal data. Litigants and counsel do not analyze legal data merely to observe some historical data fact, but rather in an effort to extract a quantitative insight about something quite different – i.e., the present or the future. If a company examines its law firm's historical trial win rate, it is not idle curiosity about their past record, it is usually as input to an evaluation of the firm's current capabilities.[1]

It is widely understood, but sometimes forgotten, that we cannot go reliably from past data to some present insight because the past is only a sample of what *could* happen and often it is a very imperfect one. Litigants know all too well that they cannot buy their law firm's past courtroom victories. So when court records are digitized and we aspire to use legal data for decision-making purposes, the value of that data can no longer stand simply on the merits of rapid searchable access to history, as impressive as this may be. The data must be held to a higher standard. When data analytics is used to make a decision, essentially it becomes a data science experiment and the standard should therefore be a scientific one.

The central problem, of course, is that not all samples of legal data contain sufficient information to provide any useful barometer of the present. On the contrary, data – small samples in particular – often contain so much random noise that they can be misleading as a decision aid.

For example, it might be a knowable fact that Law Firm A beat Law Firm B four times out of six in head-to-head court battles. But if it is not a quantitatively meaningful measure then the fact has not yet risen to the level of valuable knowledge. It is still part of the soup of random performance differentials, of which there are many, and which inevitably fluctuate over time.

---

[1] Whether or not performance evaluation can be reduced to one number and whether or not the data is sufficiently unbiased, homogeneous and representative can be debated. This article is agnostic as to these questions. The point here is simply that if historical data is analyzed for decision-making purposes, the justification for its application is an assumed relationship to some present or prospective attribute.

Sam L. Savage writes in The Flaw of Averages, "Information has no value at all unless it has the potential to change a decision."[2] And random variation, we can all agree, should not change decisions. Thus, legal data and the analytics acting on it only become valuable for decision-making purposes when we can use their combination to distinguish meaningful samples of data from potentially meaningless random data facts.[3]

But how much data is enough? At what point can we say that we have enough data such that a useful signal has emerged from all the useless random noise?

We do not need to have a quantitative background to recognize this as a question for statistics. But as Big Data use in general becomes popularized, its proper statistical analysis seems to be increasingly overlooked in favor of simplistic sample summarization – as if the past is always a perfect prologue. The question of sample size is often ignored and as consumers of data we are usually invited, if only by omission, to use our intuition to extrapolate the past into the present or the future. The full consequences of this trend are unknown, but flawed decision making cannot be very far behind.

Where analytics concerns trivialities such as social media metrics, there is probably no great harm to this casual use of data, but where important decisions are on the line, the right statistical methods should always be applied. Unfortunately this is not a popular view today and one purpose of this article is to encourage caution in the face of the current technological fashion.

BIG DATA FASHION

In the present climate there seems to be a greater fascination for the scale and wizardry of Big Data, rather than a focus on its informational content. Rarely

---

[2] Savage, Sam L. *The Flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty*. (Hoboken, NJ, Wiley 2009) 118.

[3] For a thorough examination of the dangers of randomness in decision making see, Taleb, Nassim N. *Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets* (New York, Random House 2005).

are questions raised as to the integrity of the data or its statistical merit. Coverage of the subject is often effusive, and it has become cliché to describe Big Data as 'disruptive', when it might be better to have a conversation about its simple utility and statistical significance before leaping to such grandiose conclusions.

The overall message suggests a data utopia free from anxiety, when in truth there is much to be anxious about. The pitfalls in data analytics are so numerous and commonplace that if you are not worried about your data and the use of it, there is a very good chance you are not using it correctly. As the French biochemist and Nobel Laureate Jacques Monod once said, "It is restlessness, anxiety, dissatisfaction, agony of mind that nourish science."[4] The same could well be said of data analysis.

In general, there is a creeping, almost pop science belief in data as an analytic end in itself, which can sometimes verge on the evangelical. Despite this popular fervor, legal professionals are encouraged to always bring skepticism to their use of data. While there are no doubt many excellent data sources, technology should never be a matter of faith. In reality, data is just the beginning of the analytic process and it should be the role of analytics to evaluate its statistical worth before reaching any conclusive end.

Some have even suggested that data in sufficient volume "makes the scientific method obsolete".[5] This idea of data as a substitute for science had its watershed moment in this excerpt from Chris Anderson's 2008 article for the WIRED magazine website:

> There is now a better way. Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show.[6]

---

4 As quoted in, 'Ariadne', New Scientist (17 Jun 1976) 70, 680.

5 Chris Anderson, 'The end of theory: the data deluge makes the scientific method obsolete', (WIRED.com, June 23, 2008), http://www.wired.com/2008/06/pb-theory/

6 Ibid.

Unfortunately, petabytes do not allow us to say any such thing.[7] The larger the data set, the more likely it is you will find accidental correlation. As Nassim Taleb has pointed out, "big data means anyone can find fake statistical relationships, since the spurious rises to the surface."[8] And as the remainder of this article will hopefully show, data science without the science part is just data – alone, it has very little to say about anything upon which a decision might turn, including legal decisions.

It has to be ironic that the more Big Data begins to permeate every facet of our lives, the more detached it is seeming to become from science and the branch of statistics upon which its credibility rests. In the age of the Internet, statistics sadly appears to be increasingly displaced by colorful infographics, which Georgetown law professor Paul Ohm has rightly described as "the effluent of the information society".[9] In sum, while there is a surfeit of Big Data, it could be argued there is a dearth of good data science.

This general tendency to neglect statistics (and statisticians in the process) is unfortunate and has been recently highlighted by three biostatistics professors from Johns Hopkins University in their blog post at **SimplyStatistics.org**.[10] [11] Attempting to explain the neglect of statistics in Big Data initiatives, the authors suggest the following:

---

[7] Anderson's view of data as a substitute for science was succinctly rebutted by Pigliucci. See, Massimo Pigliucci, The end of theory in science?, EMBO Rep. 2009 Jun; 10(6): 534, **http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2711825/**

[8] Nassim N. Taleb, 'Beware the Big Errors of "Big Data"', (WIRED.com, February 8, 2013), **https://www.wired.com/2013/02/big-data-means-big-errors-people/**

[9] Paul Ohm, Response, The Underwhelming Benefits of Big Data, 161 U. PA. L. REV. ONLINE 339, 346 (2013), **https://www.pennlawreview.com/online/161-U-Pa-L-Rev-Online-339.pdf**

[10] 'Why big data is in trouble: they forgot about applied statistics' (SimplyStatistics.org, May 7, 2014), http://simplystatistics.org/2014/05/07/why-big-data-is-in-trouble-they-forgot-about-applied-statistics/

[11] The Simply Statistics blog post also makes reference to an interesting talk by Terry Speed, a former professor of statistics from UC Berkley and current laboratory head in the field of bioinformatics. Speed debunks the Big Data trend and points to the "puzzling" absence of statisticians in many Big Data initiatives. See, Terry Speed, Speech given at Chalmers University of Technology. (9 Apr 2014). Chalmers Initiative Seminar on Big Data [Seminar Video], **http://www.chalmers.se/en/areas-of-advance/ict/calendar/Pages/Terry-Speed.aspx**

> One reason is that when you actually take the time to do an analysis right, with careful attention to all the sources of variation in the data, it is almost a law that you will have to make smaller claims … [12]

So to analyze data correctly statistics should always be close at hand. And the branch of statistics that governs our ability to make a current decision based on a sample of historical data is called inferential statistics.[13] Whereas in law, an inference is a logical result of facts and reason, in the statistical analysis of data, an inference turns critically on the quantity of data in the sample, among other things.

It can be tempting to imagine that Big Data, by virtue of its very bigness, must always give us a sufficient volume of data to be quantitatively useful. But this is not necessarily the case and should not be assumed. The relevant standard is not whether the entire data source can be regarded as 'Big', but whether the filtered data that is pertinent to the present legal case is big enough.

Recognizing, then, the important role of statistics when applying Big Legal Data, let us take a look now at how it can help us to evaluate the question of data sufficiency.

When Are The Data Big Enough?

A good way to explore this issue is to look at how legal data is actually being used. Often this involves the calculation of an average, a proportion[14] or win rate about some legal track record of interest to GCs and outside counsel as they prepare for trial. For example, counsel may wish to know what proportion

---

[12] Simply Statistics (n 10).

[13] Inferential statistics is distinguished from descriptive statistics. Descriptive statistical methods – such as the mean, standard deviation and the use of box plots – merely summarize the data sample and may not be used to make an inference about the population from which the sample is drawn.

[14] A proportion is essentially an average where the thing being measured has a binary outcome such as grant or deny, win or lose.

of the time a specific judge has ruled one way or the other on a particular type of matter. A hypothetical example will serve to illustrate:

Imagine as defendant's counsel in a securities class action lawsuit, we are considering a motion for dismissal with (the completely fictional) Judge Jones.[15] This judge has earned a reputation for being relatively unreceptive to dismissal motions having granted only 3 of the last 20 motions for a dismissal proportion of 15%. Of all securities class action cases nationwide, approximately 32% are understood to terminate via a motion to dismiss,[16] more than twice the dismissal rate for our judge. As counsel we consider whether to take the data on Judge Jones into consideration. Is the sample big enough to provide an accurate measurement of Judge Jones? Does it differ in a meaningful way from the national average?

On the face of it, the data provide compelling support for the Judge's reputation. The differential with the national average is large. But when we look closer, other explanations suggest themselves.

It could be that the 20 cases were, by chance, an unusual sample or it could be that the Judge's past decisions were not typical of his general propensity to grant motions for dismissal for some other reason unknown. Even without the aid of statistics, we realize that such a small amount of data may not be sufficient to conclude either that Judge Jones is biased against motions to dismiss or that he even was. And yet, we also recognize that as the sample size ('n' in statistical parlance) gets bigger, there is some point at which the analysis would be difficult to ignore; it is common sense and a statistical fact that the more data we have, the more persuasive it is – at least when it comes to making a simple measurement.

---

[15] The idea for this example draws from an article on Forbes.com, see: Daniel Fisher, 'Stanford-Bred Startup Uses Moneyball Stats To Handicap Judges, Lawyers' (Forbes.com, February 2, 2015), http://www.forbes.com/sites/danielfisher/2015/02/02/stanford-bred-startup-uses-moneyball-stats-to-handicap-judges-lawyers/

[16] Couture, Wendy Gerwick, Around the World of Securities Fraud in 80 Motions to Dismiss (January 1, 2014), Loyola University Chicago Law Journal, Vol. 45, 553.

The obvious problem is that when we file a motion to dismiss with Judge Jones, we will not experience his historical average, we will sample from the wider 'population' of his future decisions. For the purpose of this discussion we can think of this unseen population as having a measure of central tendency – the long run 'population proportion' – reflecting the Judge's true propensity to grant such motions.[17]

Since it is this population to which we are really exposed, it would be helpful then if we could evaluate how much uncertainty about the population proportion is implied in our sample. In other words, what do we really know based only on 20 data points? Essentially we need to examine the potential for what in statistics is called 'sampling error' – i.e., the possibility that the small sample that we do have is not a perfect representation of the larger population. This will tell us how useful the sample really is and whether we should place much reliance on it.

Fortunately there is way to do this and you do not need a Ph.D. in applied statistics to run the numbers.

Statistics To The Rescue

Corporate litigants and their counsel can investigate the potential for sampling error in their legal data by computing a 'confidence interval' whenever they look at an average, a proportion or a win rate.[18] All readers will be familiar with confidence intervals even if they have never studied statistics because they are used to compute the familiar 'margin of error' we see in election polls.

---

[17] Again, this article is agnostic as to whether a marginal propensity can be reduced to a quantitative measure. The point is only that if data is going to be used, inferential statistics must be part of the evaluation.

[18] Although we focus here on confidence intervals about the proportion of the time that a judge makes a particular ruling, the same idea can be applied to law firm win rates or the average in some class of compensatory damages. Provided the right formulas are used, confidence intervals have a wide application.

Effectively, confidence intervals provide a range within which we might expect the population proportion to fall with some probability if we were to repeat the sampling process. Confidence intervals therefore provide a measure of the uncertainty we have about the population proportion itself.[19]

Intervals can be computed for different levels of confidence, but it is common practice to examine the 95% confidence interval. When we compute the 95% interval using the 20 data point sample for Judge Jones, we find that the proportion might not be 15% on the nose, but could be as low as 3% or as high as 38% – an uncertainty range of 35 percentage points.

As you might imagine, it is never a good thing when the range within which we can be said to have confidence in our measurement is more than twice as big as the number we are trying to measure.

With a potential high of 38%, Judge Jones's propensity to grant a motion for dismissal could in fact be materially higher than our national average – not lower. Simply put, the Honourable Judge Jones might be unreceptive to dismissal motions, but he might not be. The range of uncertainty is so wide that this judge may, in the fullness of time, show no particular tendency at all relative to the national average for all judges. The high level of uncertainty in our measurement is caused by the small sample size.

---

[19] For an example of confidence intervals used in the analysis of US Supreme Court Justices, see: Lee Epstein, Andrew D. Martin, Kevin M. Quinn, and Jeffrey A. Segal, Circuit Effects: How the Norm of Federal Judicial Experience Biases the Supreme Court (2008), 157 U. Pa. L. Rev. 833, 879, **http://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=1185&context=penn_law_review**

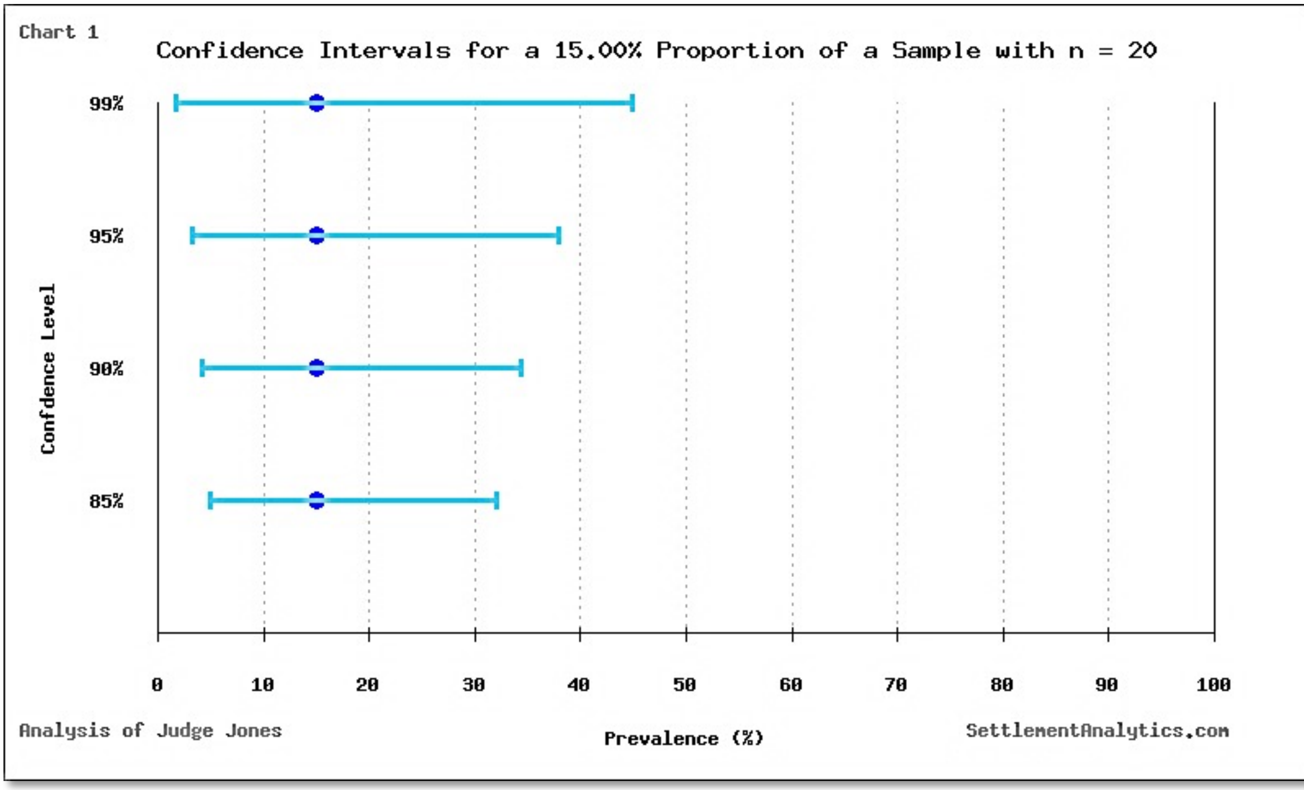


Figure 1.

We could of course compute intervals for less demanding degrees of confidence, but this does not really solve the problem. On the one hand the interval would be narrower, but on the other hand we would have less confidence in what it has to say.

Still, it can be informative to look at how confidence intervals change for different confidence levels as it provides a more complete picture of the data. As a convenience to the reader, we provided a simple legal data analytics tool for calculating and visualizing a range of confidence intervals on the SettlementAnalytics website.[20]

[20] SettlementAnalytics, "Confidence Intervals," (accessed August 31, 2016), https://settlementanalytics.com/law-stats/confidence-interval/

Technical Note: There are several different methods for computing confidence intervals and each have their advantages and disadvantages. However, because of the mostly binary nature of judicial responses to motions, here we have used a confidence interval for a binomial proportion. There are then several different methods of computing binomial proportion confidence intervals. In this analysis we have used the Clopper-Pearson method (sometimes called the 'exact method'), which also has advantages and disadvantages. A discussion of these is beyond the scope of this article, but readers should be aware that results will vary somewhat depending on the choice of method.

Using this analytic technique we can visually compare intervals for the standard confidence levels of 90%, 95% and 99%. The chart in Figure 1 (above) illustrates how these look for Judge Jones based on the 20 data point sample size.[21] The user can also add a custom confidence level, which for this particular analysis we have set at 85%.

As the reader can see, the intervals are generally quite wide. Even if we accept a low 85% level of confidence, the range of uncertainty is more than 27 percentage points. And if we want to be 99% confident, the uncertainty increases to more than 43 percentage points. These results are summarized in Table 1.

In short, the analysis is telling us that there is a great deal of uncertainty about our measurement of Judge Jones based on this small amount of data. We may know the historical facts about this judge's record, but they do not contain much information about the measurement we are really interested in: his marginal propensity to grant motions for dismissal.

[21] The input assumptions for this chart are provided in Appendix 1.

This example serves to illustrate an important point about legal data analytics: simple summary measures of sample data such as a proportion or a win rate do not reveal the whole story. This is true whether we are looking at judicial track records or whether we are comparing the performance of one law firm with another. Data, which at first blush may appear to indicate a difficult judge or a successful law firm, may not contain very much valuable information at all when it comes to making a decision. Without careful statistical analysis, summary legal metrics like proportions or averages can be misleading and our quantitative intuition can fail us.

Your results

Confidence Level	85% (custom)	90%	95%	99%
Proportion	15.00%	15.00%	15.00%	15.00%
Upper Confidence Bound	32.12%	34.37%	37.89%	44.95%
Lower Confidence Bound	4.99%	4.22%	3.21%	1.76%
Confidence Interval (% points)	27.13%	30.15%	34.69%	43.18%

Table 1

The point here is not to decry every calculation of an average or every casual evaluation of quantitative legal data, but rather to emphasize that once we begin to analyse the law quantitatively, we must consider statistical issues like measurement error and confidence. When the law is transported from the world of *qualitative* legal argument into the world of *quantitative* analysis, it is imperative that we take the discipline of inferential statistics along for the journey.

More Data Please

If counsel should discover that they do not have a big enough sample of data to be particularly meaningful, it can be instructive to ask just how much data would be necessary. It can help to know at what point we should consider taking the data seriously. Here again confidence intervals can help.

The analytic trick is to simply pretend that we do in fact have additional data and to compute several confidence intervals for a range of increasing sample sizes, keeping the observed proportion as a constant. By visually inspecting how the confidence intervals narrow as the sample size increases, counsel can get a good sense as to when the volume of data starts to mean something.

We include an application to automatically run this analysis on **our website** as part of our confidence interval tools.[22] This application automatically computes a range of 95% confidence intervals for sample sizes ranging from n = 5 to 5,000 for the given proportion. Figure 2 illustrates what this looks like using Judge Jones's proportion of 15%.

As you might expect, uncertainty in the measurement diminishes as the amount of available data increases. Generally, the range of uncertainty for sample sizes greater than 300 is fairly narrow, whereas the range of uncertainty for sample sizes less than 300 is quite large. And not surprisingly, uncertainty expands rapidly as n approaches single digits. All of this reconciles with what we do know intuitively: when we don't have a lot of data, it doesn't really mean very much.

---

[22] SettlementAnalytics, "Confidence Intervals," (accessed August 31, 2016), https://settlementanalytics.com/law-stats/confidence-interval/

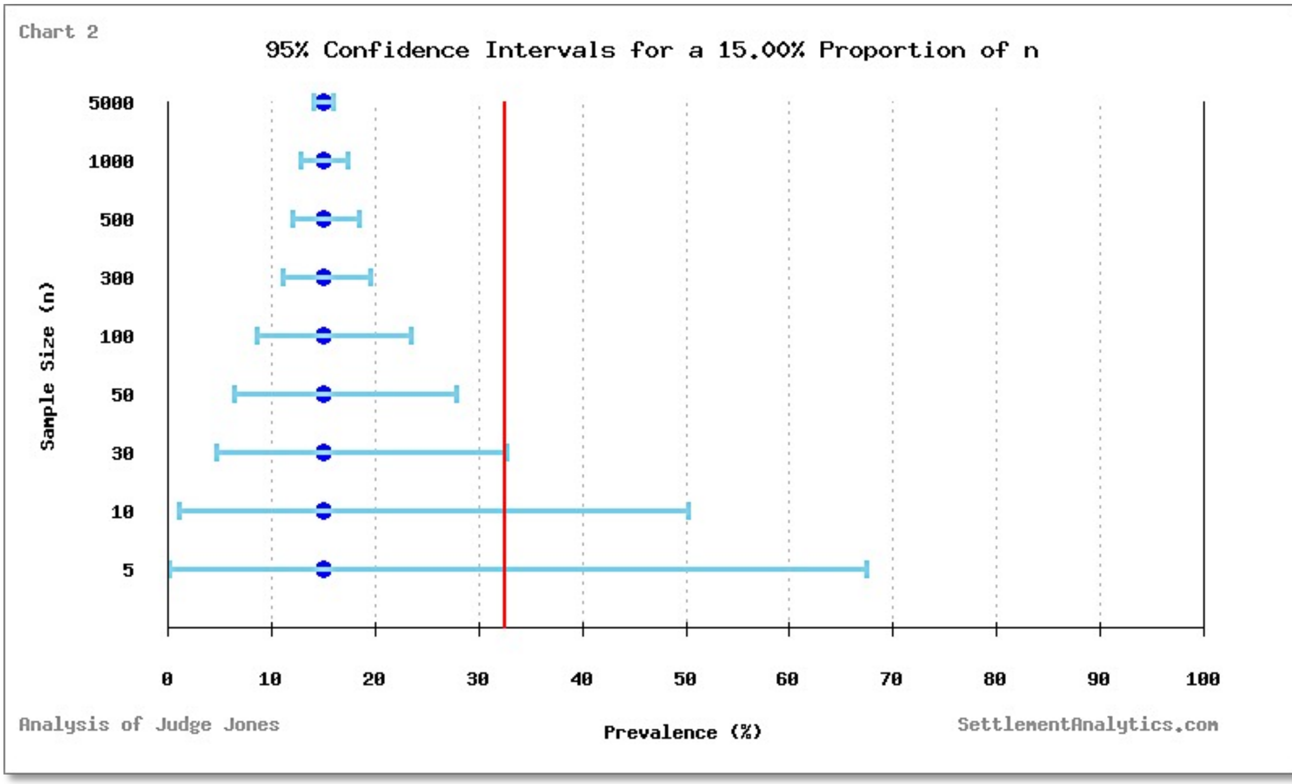


Figure 2.

Overlaid on the chart in Figure 2 is a vertical red line for the national average proportion we have used for all judges. This illustrates the extent to which the various sample measurements diverge from the national average after taking measurement uncertainty into consideration. It stands to reason that Judge Jones could be more convincingly described as unreceptive to dismissal motions if we had enough data so that the range of uncertainty about the measurement at least excluded the national average. But notice this does not occur until n reaches about 30. Until this point there is simply not enough data to support the claim that the national average for all judges is not an acceptable possibility for Judge Jones's population proportion.

Of course 30 will not always be the magic number – this will depend on the underlying proportion at issue, the average to which it is being compared and the choice of method for computing the confidence interval. The point is to show how this technique can help litigants get a sense as to when they have enough legal data points for their analyses to contain meaningful insights, and when they do not.

Differentiated, But Still Uncertain

However, it is important to keep in mind that even if our sample size is big enough so that the confidence interval actually excludes some benchmark (such as the national average in this case), the interval itself can still be disconcertingly wide. For example, notice that even if we had a sample size of 50 decisions for Judge Jones, although the confidence interval would now comfortably exclude the national average, uncertainty about the measurement would still be more than 20 percentage points.

So even with 50 data points it could not be said that we have an accurate reading for this particular judge. By visually inspecting the chart in Figure 2, the reader will see that we do not achieve any reasonable degree of certainty in this measurement until the sample size reaches about 300, and it is only at some point over 1,000 that we have something we might describe as precision.

The problem that litigants and their law firms will face in using Big Legal Data as an aid to legal decision-making is that by the time they have filtered the data down to the court, the judge, the law firm and/or the type of matter that is relevant to the case on their desks, the sample size will not always be big enough to produce something statistically meaningful. And when it is, it still may not be enough to produce a high degree of measurement accuracy.

The bottom line is that while litigants and law firms would no doubt like to use Big Legal Data to extract some kind of informational signal from amid the noise that is present in data samples, the hard truth is that there will not always be one. Therefore, before using legal data to inform any decision, counsel should first look at whether they have enough of it. Confidence intervals can be a good way of doing that.

However, as we will see later, they are not the only way.

Cognitive Biases

If the only problem was that legal professionals will sometimes be dealing with small, statistically insignificant samples of legal data, this would be concern enough. But it has been found that, as human beings, we are particularly susceptible to being irrationally influenced by small amounts of data. Examples of this are the cognitive biases known as 'Representativeness', and 'Insensitivity to Sample Size'.[23]

Psychologists Daniel Kahneman and Amos Tversky studied these phenomena and other related biases in the 1970s. In their 1971 paper, 'Belief in the Law of Small Numbers', the authors found that people expect samples drawn from a population will be "more similar to one another and to the population than sampling theory predicts…"[24] Essentially they found evidence of a cognitive bias in which people believe irrationally that small samples are excessively representative of the population. A believer in this law of small numbers, according to Kahneman and Tversky, places "undue confidence in early trends", and "underestimates the breadth of confidence intervals".[25] Thus, when Big Legal Data is filtered and yields a relevant but small sample of data, counsel may be subjecting themselves to this cognitive bias.

Belief in the law of small numbers also shares some similarity with the 'Hot Hand Fallacy' – the irrational belief that recent success in random tasks will be followed by further success. Gilovich et al. studied this phenomenon in 1985 and found evidence of a belief in the 'hot hand', which they attributed to, "a general misconception of chance according to which even short random

---

23 Tversky, Amos; Kahneman, Daniel, 'Judgment under Uncertainty: Heuristics and Biases' (27 September 1974), Science, Vol. 185, No. 4157. pp. 1124-1131, 1124, 1125.

24 Tversky, Amos; Kahneman, Daniel, 'Belief in the law of small numbers' (August 1971) Psychological Bulletin, Vol 76(2),105-110, 105.

25 Ibid 109.

sequences are thought to be highly representative of their generating process".[26]

In the context of evaluating legal data, the Hot Hand Fallacy creates a danger of attaching undue significance to a short streak of positive or negative results. This could be problematic for legal professionals when attempting to interpret a string of judicial decisions, law firm track records or the performance of an individual lawyer.

Cognitive biases of this sort influence all of us and have probably been at the root of many questionable decisions in general. While we need to guard against these biases at the best of times, ready access to small subsets of quantitative data heightens the legal profession's exposure to its own otherwise perfectly human tendency to err in this regard.

Data Visualization

Data visualization is another promising area of the legal data landscape. Visualization usually involves graphical and schematic renderings of data in two and three dimensions. It often combines impressive data imagery with an almost infinite ability to filter, connect and display data objects. Such visualization techniques can give counsel the ability to see large volumes of data at one time and examine its order and structure.

Where the use of data visualization is limited to better comprehension of the data *sample*, there can be few analytic quibbles. Visualization, for example, might be validly used to efficiently survey discovery documentation, to understand its sources and completeness. However, if data visualization goes from understanding the sample to promoting an inference about a *population*, then, again, it must be held to a higher analytic standard before insights can be said to be valid. Questions of sampling error, uncertainty and statistical

[26] Gilovitch, Thomas, Robert Villone, and Amos Tversky, 'The Hot Hand in Basketball: On the Misperception of Random Sequences' (1985) Cognitive Psychology, 17, 295–314, 295.

significance become vital considerations, and these will require statistical analysis.

Advocates of data visualization in general often cite as a benefit the ability of users to discover patterns in data images that would otherwise lie hidden – a sort of ‘visual data mining’. There may be significant potential to develop visual pattern recognition applications in this branch of legal data analytics. But will the discovered patterns always be valid?

One reason why they might not be is the cognitive bias known as ‘apophenia’. This is the human tendency to see meaning and patterns in images that might actually be quite random.[27] Psychologist, Peter Brugger described apophenia as, “the pervasive tendency of human beings to see order in random configurations.”[28] Apophenia explains why we see a face in the moon or a buy signal in the random ups and downs of a stock price chart. The problem is that with endless options to slice, dice and creatively display legal data it would be a miracle if you did not find patterns routinely. Wickham et al. echo this concern: “When visualizing data, how do we avoid falling into the trap of apophenia where we see patterns in random noise?”[29]

This is not to say that meaningful patterns will not sometimes exist or that counsel will not see them, but rather that they may sometimes be difficult to distinguish from the patterns that can occur by chance and those that may be imagined or magnified by apophenia.

---

[27] There are several variants of apophenia including, pareidolia, overfitting, the gambler’s fallacy, see Wikipedia, The Free Encyclopedia, "Apophenia," (accessed August 31, 2016), https://en.wikipedia.org/wiki/Apophenia

[28] Brugger, P. *From haunted brain to haunted science: A cognitive neuroscience view of paranormal and pseudoscientific thought. In: Hauntings and Poltergeists: Multidisciplinary Perspectives*, Edited by J. Houran and R. Lange (North Carolina: McFarland & Company, Inc. Publishers, 2001), pp. 195–213, 196

[29] Hadley Wickham, Dianne Cook, Heike Hofmann, Andreas Buja, "Graphical inference for infovis", IEEE Transactions on Visualization & Computer Graphics, vol.16, no. 6, pp. 973-979, November/December 2010, 973, 973.

Independent of the role of apophenia, the idea of visual pattern recognition may warrant caution because it inserts the researcher into the data science experiment. For example, in the same data image Bob may see one pattern and Alice may see a completely different pattern. But if Bob cannot see Alice's pattern or Alice cannot see Bob's pattern, it must raise a question about the validity of the patterns themselves.

In this way, data visualization taps into on the very modern notion that it is your unique user experience of the data that counts – a pattern becomes valid merely because you discern it. This reverence for the subjective may be appropriate in many things, but data analytics probably should not be one of them. Analytics, after all, is not a Rorschach test.[30]

Another concern with the idea of analytics-as-visualization is that it also tends to flatter our investigative ego. You will be more likely to find a pattern if indeed you wanted to find one, which has to be quite likely – why else would you be looking at the data images in the first place? Unfortunately, this is an example of something called confirmation bias,[31] which is "the tendency to search for, interpret, favor, and recall information in a way that confirms one's pre-existing beliefs…"[32] Clearly, confirmation biases could act to undermine the validity of perceived patterns if we are predisposed to see them and are motivated to believe in their existence.

Whatever their cause, cognitive biases such as those discussed above provide additional reason to check and balance our gut-feel interpretation of legal data with rigorous statistical analysis.

---

[30] Or is it? Wickam et al. have proposed and developed a two-stage process of graphical statistical inference to balance the benefits of data visualization with the rigour of statistical scepticism. They propose a "Rorschach Protocol", to "calibrate user vision to the natural variability in [data] plots". The vital point of distinction here is the essential role of inferential statistics. See Hadley Wickham, Dianne Cook, Heike Hofmann, Andreas Buja, 975 (n 29).

[31] Wason, Peter, "On The Failure to Eliminate Hypotheses in a Conceptual Task" (1960), Quarterly Journal of Experimental Psychology. 12 (3): 129–140.

[32] Wikipedia, The Free Encyclopedia, "Confirmation Bias," (accessed August 31, 2016), https://en.wikipedia.org/wiki/Confirmation_bias

STATISTICALLY INSIGNIFICANT 'BIG' DATA

Another way to examine the statistical merit of a sample of legal data is to use a hypothesis test. In some ways a hypothesis test can be thought of as the complement of a confidence interval, but this time it starts with the assumption that there is no difference between the population being sampled and some standard or benchmark such as a national average.[33] This starting assumption is known as the 'null hypothesis', and it holds that the observed difference between the sample measurement and the null value derives only from the random variation that occurs in sampling.

The reason for framing the problem this way is so that the observed divergence (relative to the null) can then be examined in terms of its probability of occurring by chance. Using this approach allows that at some extreme point the divergence can be deemed so improbable that the null hypothesis can then be rejected. Essentially, a hypothesis test gives us a standardized way of being able to say when a sample of observations is meaningfully different from the norm. This is what it means when scientists say that a test result is 'statistically significant'.

Provided the data satisfy certain technical requirements, we can use hypothesis tests to say whether a sample of quantitative legal data is significantly different from, say, a wider peer group or a national average. An analysis of a law firm's performance track record will help to illustrate:

Imagine you are General Counsel for a technology company. As owner of a patent, your company has identified infringement and intends to litigate. The patent is critical to your company's future and you are considering a tactical

[33] In medical research this is usually the control group or the placebo.

allocation of this case to a new law firm because you suspect that your current law firm may have underperformed in litigation on behalf of patentees.[34]

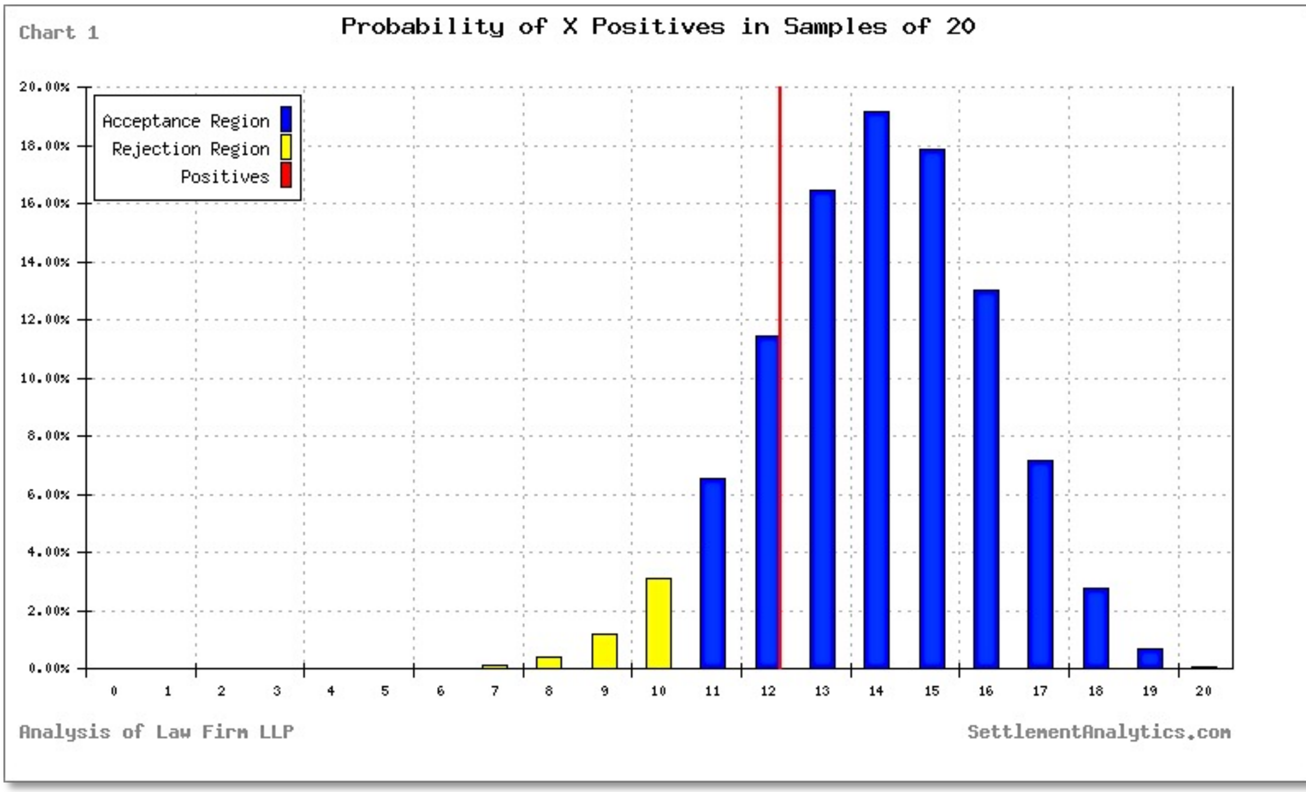


Figure 3.

Data indicates that your current attorneys, (the completely fictional) 'Law Firm LLP', have won only 12 out of 20 patent infringement cases when representing patent owners at trial. Law Firm's win rate of 60% appears to compare

[34] The idea for this example draws from an article on Law.com, see: Scott Graham, 'Firms Claim Bragging Rights in New Field of Patent Litigation' (Law.com, March 6, 2015) http://www.law.com/sites/lawcomteam/2015/03/06/firms-claim-bragging-rights-in-new-field-of-patent-litigation/

unfavourably with the overall 70% rate at which patentees are understood to prevail at trial in patent disputes nationwide.[35]

Although Law Firm seems to have underperformed, you recognize that chance may be playing a factor in this small sample of data. You decide to run a hypothesis test to see if the underperformance is statistically significant when compared to the national average win rate for patentees at first-instance.

To run this hypothesis test, we begin by establishing the null hypothesis that, in the long run, Law Firm's population performance would be the same as the national average rate of 70%. If this were true, then out of its 20 cases, Law Firm would have typically prevailed 14 times (70% of 20). This assumption thus centres the analysis around a null level of 14, and we can now look at how unlikely it would be to achieve Law Firm's actual win rate if only chance was at work.

To do this, a hypothesis test models the random variation we would expect to find in the number of wins if samples of 20 could be drawn randomly from a population characterized by the null win rate. The chart in Figure 3 (above) illustrates this model as the theoretical probability of different numbers of wins centred about the null.[36] As a convenience to the reader, we provide an application to run this type of analysis on **our website**.[37]

35 Barry, C., Arad, R., Ansell, L., Cartier, M., Lee H., "2015 Patent Litigation Study - A change in patentee fortunes," PWC, May 2015, 20, https://www.pwc.com/us/en/forensic-services/publications/assets/2015-pwc-patent-litigation-study.pdf

36 The data for this chart and the input assumptions are provided in Appendix 2.

37 SettlementAnalytics, "Hypothesis Tests," (accessed August 31, 2016), https://settlementanalytics.com/law-stats/hypothesis-test/

Technical Note: There is a technical question here as to the theoretical model we choose to describe the randomness in the sample. In scientific investigations, often the normal distribution or the *t*-distribution are suitable in hypothesis testing. However, because court cases decided on the merits have mostly binary outcomes (win or lose), it is more appropriate to use a binomial distribution in this particular case.

The next step is to establish some extreme degree of improbability in the divergence from the null, below or equal to which we would be comfortable rejecting the null hypothesis. There is no hard and fast rule as to what this level should be, but by tradition it is often set at 5% or 1%.[38] This is essentially the number at which point we could say that chance is unlikely to be the cause of the divergence between the null and the actual observation. The technical term for this threshold is the 'significance level' of the test and in our analysis of Law Firm we will set this at a level of 5%.

This significance level is then used to demarcate a 'Rejection Region' on the chart so that any number of wins observed inside this region can be said to occur with a probability less than or equal to our 5% significance level. The Rejection Region relevant to our analysis of Law Firm is illustrated by the yellow-shaded 'tail' of the distribution in Figure 3. The area in this tail represents approximately 5% of all possible outcomes.[39]

The whole point of a hypothesis test is that if the actual observed number of wins falls inside this Rejection Region, its occurrence can be deemed so

---

[38] The choice of significance level should be guided in part by managing the risk of 'Type I' and 'Type II' errors. A Type I error occurs when we incorrectly reject a null hypothesis that is true. And a Type II error occurs when we fail to reject a null hypothesis that is not true. The more demanding is the significance level of the test, the less likely it is that a Type I error has occurred.

[39] Only "approximately", because the binomial distribution is discrete and not continuous. If we wanted to refine this analysis we could consider something called a continuity correction, but this is beyond the scope of this introduction.

improbable that we can reject the idea that it has occurred by chance.[40] If this were true, clearly it would provide some support for the idea that Law Firm really has underperformed.

Having marked the Rejection Region in yellow in Figure 3, it is now a trivial matter to look at where the actual performance for Law Firm is located – marked by the vertical red line. We can see that the firm's actual performance of 12 wins is clearly not inside the Rejection Region.[41] Therefore, we cannot reject the null hypothesis and so it is not possible to say that Law Firm's win rate is significantly lower than the norm.[42]

Technical Note: In this example we have assumed that there is one sample (Law Firm's performance), and that it is being compared to a known or fixed standard (the national average win rate for all patentees). But in reality, the national average is also based on a historical sample that will change over time. If we wanted to be more exact in our analysis, we could look at the problem as the difference between two sample proportions and ask whether the *difference* is statistically significant. We could also compute the confidence interval for the difference between two proportions. This difference of proportions approach could also be used to quantitatively compare the performance of two law firms. For a hypothetical example of how this might look, see Appendix 3.

When we factor the relatively small sample size and compare it to below par performance that might have arisen merely by chance relative to the national

---

[40] There are two ways of interpreting a hypothesis test in statistics: the Critical Values method and the p-value method. The reader will notice that we are using the Critical Values method because it lends itself to a visual explanation. The advantage of this method is that once readers are sufficiently comfortable with the technical issues, they can find the result of a hypothesis test by simply looking at whether the red line is inside the yellow region.

[41] Note that the Rejection Region extends all the way to 0 on the x-axis – the yellow tail is not visible all the way because the probabilities become vanishingly small.

[42] For the purposes of introducing the concepts of confidence intervals and hypothesis testing we have skipped over some of the finer technical details. Needless to say, readers should familiarize themselves with these details before applying any statistical test. Informed use of statistical methods should always be the goal. For the moment, two comments are warranted regarding this example of a hypothesis test: a) Not being able to reject the null hypothesis does not mean we accept the null hypothesis as true; b) Whenever we fail to reject a null hypothesis, it is still possible that a Type II error has occurred as discussed above (n 38).

average, our fictional law firm's track record appears to be fairly ordinary. While the win rate is lower than average, with only 20 data points it is within the bounds of commonplace sample variation. The underperformance is not significant at the 5% level.[43]

Using this example, we are able to see how a hypothesis test can help to put performance data into perspective.

Critical Mass Data

Clearly, the number of data points and not just the win rate is an important factor in achieving statistical significance. The less data we have, the less likely it is that we would be able to distinguish a sample average from any given null. Conversely the more data we have, the more likely it is that the measurement would become statistically significant. So in addition to being a useful barometer of performance relative to some benchmark, a hypothesis test is useful because it also says something about whether we have sufficient data in the first place.

Using the above example, it is an interesting exercise to ask how many data points would be necessary at Law Firm's ratio of 12:20 for their win rate to become significant at the 5% level. At what point would we be able to reject the null hypothesis and legitimately question this firm's performance level? Would 18 wins in 30 cases be sufficient? What about 24 in 40? Or 30 in 50? The reader can use our **hypothesis test calculator**[44] to re-run this data

[43] Note it has become fashionable in the past few years to simply cite the probability of an observation (its 'p-value') as opposed to declaring significance relative to some arbitrary level. There are advantages and disadvantages to both approaches and for a discussion of this issue in relation to law see, David H. Kaye, Is Proof of Statistical Significance Relevant?, 61 Wash. L. Rev. 1333 (1986). In the present article we focus on the use of hypotheses and significance levels for the purposes of introduction only and not as an attempt to settle the dispute between the two methods.

[44] SettlementAnalytics, "Hypothesis Tests," (accessed August 31, 2016), https://settlementanalytics.com/law-stats/hypothesis-test/

experiment and discover what data volume they would need in order to find something statistically significant at this win rate.

Although it is obvious by now, it is worth repeating: not all data samples are created equally. Sometimes the available legal data volumes will be sufficient to produce statistically meaningful analyses, but sometimes they will not. Statistical methods like confidence intervals and hypothesis tests can help litigants and their counsel to distinguish these situations and shed light on when it is that Big Legal Data is big enough in practice.

CONSULTANTS AND COURTS AGREE

In their May 2011 report on Big Data, McKinsey Global Institute identified the "scientific process of controlled experimentation that includes the formulation of specific hypotheses" as a way that companies will use Big Data to create value.[45] However, as we have seen, the scientific use of data is not just a method of *creating* value, it is imperative to *the validity* of that value. This is true in the use of legal data generally and it is particularly true in the burgeoning field of legal tech, where data is often the bedrock of products and services.

McKinsey are not alone in their opinions. Even the courts themselves have emphasized the importance of applying rigorous statistical methods when weighing quantitative evidence. In *Moultrie, Appellant, v. Joseph R. Martin,* the Court reasoned as follows:

> When a litigant seeks to prove his point exclusively through the use of statistics, he is borrowing the principles of another discipline, mathematics, and applying these principles to the law. In borrowing from another discipline, a litigant cannot be selective in which principles

[45] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H., "Big data: The next frontier for innovation, competition, and productivity," McKinsey Global Institute, May 2011, 98, http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

> are applied. He must employ a standard mathematical analysis. Any other requirement defies logic to the point of being unjust. Statisticians do not simply look at two statistics, … and make a subjective conclusion that the statistics are significantly different. Rather, statisticians compare figures through an objective process known as hypothesis testing.[46]

It is ironic that in the courtroom there is an insistence that data be held to the high standard of inferential statistics, but outside the courtroom data used in pretrial litigation often seem to get a 'statistical hall pass'. Yet, the issues are the same and with the advent of Big Legal Data, the frequent pretrial use of quantitative methods must vastly dominate the occasional courtroom application. The truth is that both are important, because no matter the venue, where key legal decisions are at stake, it is imperative to be able to distinguish informational signals from noise.[47]

This article is not intended to be the definitive introduction to inferential statistical methods for law. Rather, the intention is to highlight the dangers of using quantitative legal data armed only with sample descriptions, and to encourage the critical evaluation of data with more appropriate methods. While there certainly are more advanced statistical methods available, a basic understanding of confidence intervals and hypothesis testing lays a good foundation upon which to build.

[46] Moultrie v Martin, 690 F.2d at 1082 [as cited in David H. Kaye, Is Proof of Statistical Significance Relevant?, 61 Wash. L. Rev. 1333, 1335 (1986)].

[47] For a thorough discussion of the difference see, Silver, N. *The Signal and the Noise: Why Most Predictions Fail — but Some Don't* (Penguin Group 2012).

THE DATA MINEFIELD

The aim of this article has not been to endorse the data-driven analysis of law[48] or to suggest that statistical testing alone will make legal data fit for some particular purpose. Rather, the aim has been to show that if legal data is going to be used quantitatively, it cannot be used validly so without recourse to inferential statistics. To draw an analogy, legal data analytics without inferential statistics is like legal argument without case law or rules of precedent — it lacks a meaningful point of reference and authority.

But while statistical tests are necessary, they are still not sufficient.

Even when there is a large volume of data to hand, when confidence intervals are narrow and measurements appear to be statistically significant, this still does not mean we should accept the results of quantitative analyses without further enquiry. There are a number of important issues yet to consider.

In the next article in this series we will look at several other factors that can potentially undermine the valid use of legal data including things like sample bias and the lack of stability in the data over time (something called 'stationarity'), among other things.[49]

Occasionally, even inferential statistics will not be able to save data from what might be its inappropriate application. It may happen that the conditions necessary for certain statistical tests are not present. Perhaps the sample is not statistically random or perhaps the sample size is just too small for a particular type of test. However, the inapplicability of statistical methods should not be seen as creating a bye for the data. It just makes it even more likely that the data simply have nothing to say.

---

[48] In fact there are several reasons why quantitative metrics can fail even when they appear to be statistically significant. For example, in the analysis of law firms, the informational value of win rates can be undermined by cherry-picking behavior in case selection, differences in average case difficulty, and conservatism in settlement advice etc.

[49] Strictly speaking some of these considerations should precede statistical testing. We have taken a few liberties with sequence in order to quickly get to the issue of data sample size.

As the reader can see, the obstacles to drawing even a simple inference from legal data are many. And when it comes to more ambitious applications of legal data such as predictive modelling and machine-based learning, these difficulties are only compounded. Predictive models are frequently undermined by problems such as spurious correlation,[50] over-fitting, multiple testing,[51] heterogeneity,[52] and the presence of something called 'confounders' – unseen factors, which may be the real cause of apparent relationships in data.[53]

We will take a closer look at these issues in due course. For the moment, it is sufficient to observe that legal data analytics is much more complex and error-prone than is generally acknowledged. While access to quantitative legal data can be a force for good, with the modern convenience of rapid data search and retrieval and the ease of data summarization, users may be tempted to oversimplify the problem. There is, simply put, a risk of forgetting the science in what is basically data science.

However, with the use of inferential statistics and careful attention to the complexities of data analytics, GCs and outside counsel can benefit from this new frontier in Big Data. But as we have already seen, this analytic path can still be hazardous. Although tools of statistical analysis can help, legal data mining can still be a data minefield. Tread carefully.

---

[50] For a useful discussion of the problem of spurious correlation see, Nassim N. Taleb, 'Beware the Big Errors of "Big Data"', (WIRED.com, February 8, 2013), **https://www.wired.com/2013/02/big-data-means-big-errors-people/**

[51] For a brief and accessible discussion of the multiple testing problem see, '10 things statistics taught us about big data analysis' (SimplyStatistics.org, May 22, 2014), **http://simplystatistics.org/2014/05/22/10-things-statistics-taught-us-about-big-data-analysis/**

[52] For an interesting overview of the problem of heterogeneity in Big Data see, Kristina Lerman, 'The Curses Of Heterogeneity In Big Data', (ACM SIGMOD Blog, Oct 30, 2013), **http://wp.sigmod.org/?p=960**

[53] For a helpful non-technical survey of the problems with data mining in general see, Tim Harford, 'Big data: are we making a big mistake?' (FT Magazine, March 28, 2014), **http://www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html**

SUGGESTED ACTION PLAN

- Corporate litigants and law firms should institute internal practices so that quantitative data is never used to make a vital legal decision unless there has been a thorough examination of its statistical merit.
- Before running statistical tests or building predictive models using legal data, consider important preliminary issues such as sample bias, heterogeneity, the potential for confounders and the problem of spurious correlation.
- Understand the advantages and disadvantages of alternative methods and distributions used in statistical testing and make appropriate tradeoffs.
- Provide corporate training so that all those who have access to quantitative legal data, also have a working knowledge of statistical considerations.
- Always consult with qualified statisticians before using legal data sets and statistical tools.

The Author:

Robert Parnell is President and CEO of **SettlementAnalytics**

Acknowledgments:

The author would like to thank Martin Petts and Mohammad Ellahi for their valuable comments and suggestions.

BIBLIOGRAPHY

Anderson, C., ‘The end of theory: the data deluge makes the scientific method obsolete’, *WIRED.com* (WIRED.com, June 23, 2008), accessed Aug 31, 2016

'Ariadne', New Scientist (17 Jun 1976) 70, 680

Barry, C., Arad, R., Ansell, L., Cartier, M., Lee H., “2015 Patent Litigation Study - A change in patentee fortunes,” PWC, May 2015, 20

Brugger, P. *From haunted brain to haunted science: A cognitive neuroscience view of paranormal and pseudoscientific thought. In: Hauntings and Poltergeists: Multidisciplinary Perspectives*, Edited by Houran. J., and Lange R., (North Carolina: McFarland & Company, Inc. Publishers, 2001), pp. 195–213, 196

Couture, W.G., Around the World of Securities Fraud in 80 Motions to Dismiss (January 1, 2014), Loyola University Chicago Law Journal, Vol. 45, 553

Epstein, L., Martin, A.D., Quinn, K.M., and Segal, J.A., Circuit Effects: How the Norm of Federal Judicial Experience Biases the Supreme Court (2008), 157 U. Pa. L. Rev. 833

Fisher, D., ‘Stanford-Bred Startup Uses Moneyball Stats To Handicap Judges, Lawyers’ *Forbes.com*, (Forbes.com, February 2, 2015), accessed August 31, 2016

Gilovitch, T., Villone R., and Tversky, A., ‘The Hot Hand in Basketball: On the Misperception of Random Sequences’ (1985) Cognitive Psychology, 17, 295–314

Graham, S., ‘Firms Claim Bragging Rights in New Field of Patent Litigation’, *Law.com,* (Law.com, March 6, 2015), accessed August 31, 2016

Harford, T., 'Big data: are we making a big mistake?' *FT.com*, (FT Magazine, March 28, 2014), accessed August 31, 2016

Kaye, D.H., Is Proof of Statistical Significance Relevant?, 61 Wash. L. Rev. 1333 (1986)

Lerman, K., ‘The Curses Of Heterogeneity In Big Data’, *wp.sigmod.org* (ACM SIGMOD Blog, Oct 30, 2013), accessed August 31, 2016

Moultrie v Martin, 690 F.2d at 1082 [as cited in Kaye. D.H., Is Proof of Statistical Significance Relevant?, 61 Wash. L. Rev. 1333, 1335 (1986).]

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H., “Big data: The next frontier for innovation, competition, and productivity,” McKinsey Global Institute, May 2011, 98

Ohm, P., Response, The Underwhelming Benefits of Big Data, 161 U. PA. L. REV. ONLINE 339, 346 (2013)

Pigliucci, M., The end of theory in science?, EMBO Rep. 2009 Jun; 10(6): 534, http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2711825/ (accessed August 31, 2016)

Savage, S.L., *The Flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty*. (Hoboken, NJ, Wiley 2009) 118

Silver, N., *The Signal and the Noise: Why Most Predictions Fail — but Some Don’t* (Penguin Group 2012)

Speed, T., Speech given at Chalmers University of Technology. (9 Apr 2014). Chalmers Initiative Seminar on Big Data [Seminar Video]

Taleb, N.N., ‘Beware the Big Errors of “Big Data”’, (WIRED.com, February 8, 2013), https://www.wired.com/2013/02/big-data-means-big-errors-people/ (accessed August 31, 2016)

Taleb, N.N., *Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets* (New York, Random House 2005)

Tversky, A., Kahneman, D., ‘Judgment under Uncertainty: Heuristics and Biases’ (27 September 1974), Science, Vol. 185, No. 4157. pp. 1124-1131

Tversky, A., Kahneman, D., ‘Belief in the law of small numbers’ (August 1971) Psychological Bulletin, Vol 76(2),105-110

Wason, P.C., "On The Failure to Eliminate Hypotheses in a Conceptual Task" (1960), Quarterly Journal of Experimental Psychology. 12 (3): 129–140

Wickham, H., Cook, D., Hofmann, H., Buja, A., "Graphical inference for infovis", IEEE Transactions on Visualization & Computer Graphics, vol.16, no. 6, pp. 973-979, November/December 2010, 973, 973

— — "Confidence Intervals," *SettlementAnalytics.com.* SettlementAnalytics, https://settlementanalytics.com/law-stats/confidence-interval/ (accessed August 31, 2016)

— — "Hypothesis Tests," *SettlementAnalytics.com.* SettlementAnalytics, https://settlementanalytics.com/law-stats/hypothesis-test/ (accessed August 31, 2016)

— — “Why big data is in trouble: they forgot about applied statistics” *SimplyStatistics.org,* Simply Statistics, May 7, 2014, (accessed Aug 31, 2016)

— — ‘10 things statistics taught us about big data analysis' *SimplyStatistics.org,* Simply Statistics, May 22, 2014, (accessed Aug 31, 2016)

— — "Apophenia," *Wikipedia.org,* Wikipedia, The Free Encyclopedia, (accessed August 31, 2016)

— — "Confirmation Bias," *Wikipedia.org,* Wikipedia, The Free Encyclopedia, (accessed August 31, 2016)

About the Author

Robert Parnell is the president and CEO of SettlementAnalytics, which he established in 2011 to bring the application of quantitative and scientific methods to the economic analysis of litigation and settlement bargaining strategy. Prior to founding SettlementAnalytics, Mr. Parnell had over 18-years of experience in finance and institutional investment management in analytic, risk management, portfolio management and executive roles. Mr. Parnell holds a BSc in physics from King's College London, an MBA from the Ivey Business School and a law degree from the University of London International Programmes. Mr. Parnell is also a holder of the right to use the Chartered Financial Analyst® designation.

About SettlementAnalytics

SettlementAnalytics™ is a legal-economic research and software development firm focused on the application of game theory and model-based analytics to the strategic analysis of litigation and settlement decision making. Software applications developed by SettlementAnalytics combine canonical and proprietary game theory models of legal conflict, which also integrate ideas from information economics, financial economics, bargaining theory and Monte Carlo simulation. The firm's flagship software service – **OptiSettle**™ – is the first commercially available software application to provide a game theory analytics platform for legal dispute. To enquire about OptiSettle, or onsite seminars on the game theory of legal conflict email **info@settlementanalytics.com**.

For more information please contact:

Robert Parnell
President and CEO
Email: info@settlementanalytics.com
SettlementAnalytics.com

Appendix 1

Input Assumptions for Figure 1, Figure 2 and Table 1

Sample size	20
Number of the sample that satisfy some criteria	3
Custom confidence required (e.g. for 80% input 80)	85
Null Level - Optional (e.g. for 50% input 50)	32
Chart Label - Optional	Analysis of Judge Jones

When you have entered these inputs, click "Calculate". Your results will display in the table below along with the three standard confidence intervals.

APPENDIX 2

Data for Figure 3

Your results

X of 20 Samples	Probability of x Positives in 20 Samples	Cumulative Probability <= x	Cumulative Probability >= x
0	0.00%	0.00%	100%
1	0.00%	0.00%	100.00%
2	0.00%	0.00%	100.00%
3	0.00%	0.00%	100.00%
4	0.00%	0.00%	100.00%
5	0.00%	0.00%	100.00%
6	0.02%	0.03%	100.00%
7	0.10%	0.13%	99.97%
8	0.39%	0.51%	99.87%
9	1.20%	1.71%	99.49%
10	3.08%	4.80%	98.29%
11	6.54%	11.33%	95.20%
12	11.44%	22.77%	88.67%
13	16.43%	39.20%	77.23%
14	19.16%	58.36%	60.80%
15	17.89%	76.25%	41.64%
16	13.04%	89.29%	23.75%
17	7.16%	96.45%	10.71%
18	2.78%	99.24%	3.55%
19	0.68%	99.92%	0.76%
20	0.08%	100.00%	0.08%

Note: The blue range is the null hypothesis Acceptance Region; the yellow range is the null hypothesis Rejection Region; the red row marks the observed number of wins for Law Firm LLP.

APPENDIX 2 (CONTINUED)

Input Assumptions for Figure 3

Sample size (n)	20
(max n = 169)	
Observed Positives in Sample	12
Number of the sample that satisfy some criteria (must be < n)	
Null Hypothesis % (e.g. for 50% input 50)	70
One Tail Significance % (e.g. for 5% input 5)	5
(Note: Two tail significance = 2 x one tail significance.)	
Significance Test	One Tail Left
Chart Label - Optional	Analysis of Law Firm LLP

When you have entered these inputs, click "Calculate". Your results will display in the table below.

APPENDIX 3

Sample Illustration of Inputs for Confidence Interval for Difference in Proportions

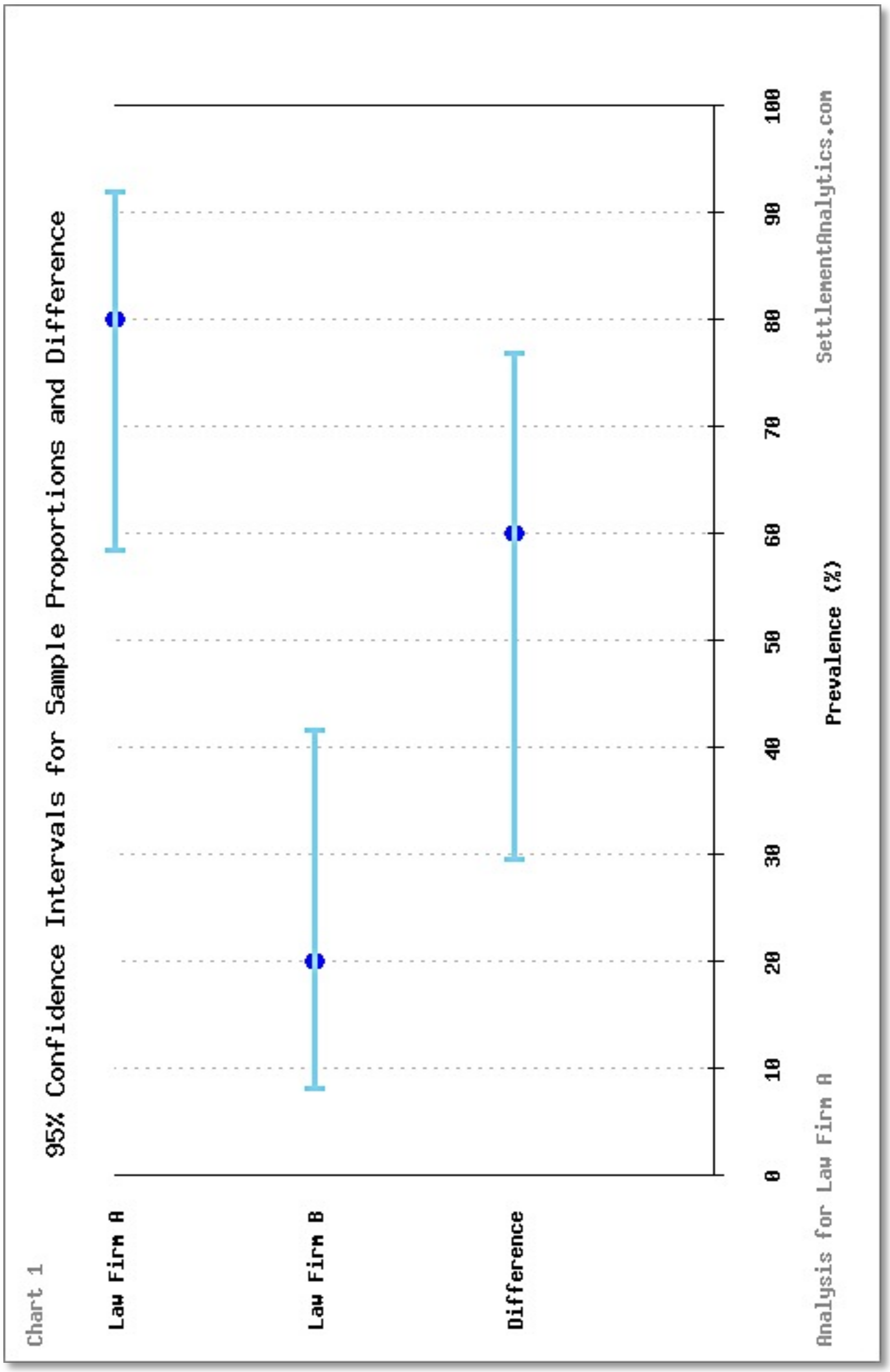
Sample Labels (Edit Optional)	Law Firm A	Law Firm B
Sample Size (n)	20	20
Observed Positives (k)	16	4
Proportion	80.00	20.00
Confidence Level	95	
Chart Label (optional)	Analysis for Law Firm A	

Proportion for Sample A must be greater than proportion for Sample B.

When you have entered these inputs, click "Calculate". Your results will display in the table below.

APPENDIX 3 (CONTINUED)

Sample Illustration of Outputs for Confidence Interval for Difference in Proportions



SettlementAnalytics™
Macro Research Associates Limited
1301 Bank of America Tower
12 Harcourt Road, Suite 1479
Central, Hong Kong

SettlementAnalytics™
THE SCIENCE OF LITIGATION