

APPLYING STATISTICS IN THE COURTROOM

A New Approach
for Attorneys and
Expert Witnesses

Phillip I. Good

CHAPMAN & HALL/CRC

Table 13.2 Distribution of Incidents by Season

Spring	Summer	Fall	Winter
A	B	C	D
A	C	B	D
AB	C	D	
AC	B	D	
A	CB	D	
			ABCD
...

have occurred in any of the four seasons resulting in a total of $4 \times 4 \times 4 \times 4 = 256$ possibilities.

The probability that each incident occurred during a different season is given by the corresponding number of rows (24) divided by the total number of rows, that is $24/256 < 10\%$. The probability that all four incidents occurred in at most two quarters of the year is $84/256$, a little less than one third of the time.

13.5.2 The Ballot Theorem and the Arc-Sine Law

Fortunately, for the non-mathematically inclined reader, we will not attempt to prove or even enunciate either of the obtuse mathematical results referenced in the title to this section. What we will do is discuss their implications.

Contrary to generally accepted views, the laws governing a prolonged series of individual observations will show patterns and averages far removed from those derived for a whole population. In other words, currently popular psychological tests would lead one to say that in a population of "normal" coins most individual coins are "maladjusted."⁴⁹

Once you get behind in a situation where chance is the predominant factor, it will take you an unexpectedly long time before you catch up. Suppose, for example, that hospital regulations specify that no more than one identifiable "error" is permitted for each ten surgical procedures performed, and by chance the very first set of ten results in two errors. How soon will things even out? If the error rate is indeed one in ten,

⁴⁹ Feller [1968; p. 72].

then the probability is only 35% that the correction will take place in the next set of 10; worse, the probability is 25% that two or more errors will be observed which would only exacerbate the situation. With the next inspection only 6 months away, will the hospital be ready on time? As noted above, the answer is "probably not." If you can't catch up, you may be able to at least stay in place. This time, the probabilities will be on your side. What one can and should argue, particularly if the hospital has instituted a set of remedial procedures, is that the incidence rate during the latter half of the inspection period is satisfactory even if the incidence rate during the initial half was not.⁵⁰

13.6 Coincidence and Ad Hoc-Post Hoc Arguments

Anything can and will happen in the long run. Misleading probabilities, lack of reproducibility, and clusters of events are only a few reasons why ad hoc-post hoc arguments should be rejected out of hand.

13.6.1 Reproducibility

No reputable scientist would ever report his results before he succeeds in reproducing his findings at least twice, once in his own laboratory and once in that of a colleague. An outside test can be particularly telling as all too often some overlooked factor — such as the quality of the laboratory water — is responsible for the results, *not* the factors under investigation. It is better to be found wrong in private than in public. Reproduce a result, then reproduce it again.

13.6.2 Painting the Bull's Eye around the Bullet Holes

Perci Diaconis [1978] spent some years as a statistician investigating paranormal phenomena including those linked to Uri Geller, the man who claimed he could bend spoons with his mind. Diaconis was disappointed but not surprised to find that the hidden "powers" of the occult were more or less those of the average nightclub magician, down to and including forcing a card and taking advantage of ad hoc-post hoc hypotheses. The fact is, as anyone who has played poker will concede, that one out of every two hands contains "something" interesting. The magician's "trick" lies in saying, "Look at this! Isn't this incredible?"

⁵⁰ We are assuming the initial run of bad luck was only that, bad luck.

In Section 13.5.1, we examined the occurrences of “incidents” over the course of four seasons and showed that chance alone could easily be responsible for their uneven distribution throughout the course of the year. Our implicit assumption was that the nature of the incident — misdiagnosis, pilot error, or assembly line failure — had been determined in advance. When rock stars or political figures die, we tend to define the nature of the incident after the fact.

When three buses appear at your stop simultaneously, a stand of cherry trees is found amid a forest of oaks, six cases of leukemia occur in a small town, and the night sky viewed through a telescope is filled with clusters of stars, do you holler “Magic” or remember the Poisson distribution that we studied in Section 10.3.3? The best courtroom defense when your opponent paints a bull’s eye around the bullet holes is to provide a computer simulation producing precisely the same results by chance alone (see Freedman [1983] for an excellent example).

13.6.3 Data Mining or Searching for Significance

When we perform a statistical test at a 5% significance level, what we are really saying is that, at least once out of each 20 tests (5%), a result that is not significant will appear significant purely by chance. Consider the Baldus study that we examined in Sections 12.4 and 12.7.3. The number of variables, 253, was truly impressive, but a total of 253 variables means that 5% or 13 of them will have statistically significant correlations by chance alone.

The solution, again, is to avoid ad hoc—post hoc hypotheses. Scientists normally work in sequential fashion, taking several small exploratory samples initially, as they develop their hypotheses, and then one or two large samples when they proceed to test them. In the case where we have limited data at hand, as in an audit for scientific integrity, one possibility is to divide the data into two parts, then use the first part to help formulate the hypotheses and the second part to test them. Only in this fashion can we be sure that the variables that test successfully for significance are not simply the results of coincidence.

13.7 Bad Statistics

Primary attacks on statistical results should be and will be directed against the design of the experiment or survey. Consider the factors listed in Chapter 8: bias in selection and response, confounding, observer bias, and inconsistent classifications. Was the population appropriate or germane (see Chapter 1)? What about those who did not respond? Would their responses be the same as those who did?⁵¹

⁵¹ Nonresponders include those who refused to answer or participate as well as those who were not available or were simply overlooked.

Attacks can and will be directed against the methods of collection as they are today in the ongoing review of the decennial census. Does an audit trail allow one to go backward from the numbers stored in the computer to the original observations? At the very least, a random selection (audit) should be made to determine the number and extent of the errors in transcription.

13.7.1 An Example from the National Game

We include the next case because it reiterates important points raised in many of the preceding chapters and, to be honest, because one of the parties is Bud Selig, the current Major League Commissioner; and I am a rabid baseball fan.⁵²

*Selig v. U.S.*⁵³ pitted the commissioner against the Internal Revenue Service (IRS) which disallowed his proportionate share of the amortizations of certain player contracts that the then newly formed Milwaukee Brewers acquired from the ashes of the Seattle Pilots. A two-step regression analysis of player salaries was offered in evidence by Selig and disputed by the IRS.

Though the Court stated, “I decline to resolve this dispute,”⁵⁴ nonetheless, it rejected the regression analysis on the following grounds, all of which we have encountered at least once in previous chapters:

- Wrong population — transaction data from the player market was used as the basis of the regression equation, a market that is highly controlled by the American League, whereas “the relevant market is the club market in which the bundle of assets was purchased.”⁵⁵
- Sample not representative — the sample consisted of transactions in the player market, whereas “testimony was unanimous that higher quality players are rarely transacted for cash in the player market.”⁵⁶
- Wrong predictor variables — the amount by which the player reserve system depressed salary levels was erroneously attributed to the value of the franchise.
- Database not reliable — both parties agreed at trial that the transaction records lacked much information about the true substance of the transactions.

⁵² My daughters and I sat side by side to watch Nolan Ryan, pitching for the Astros, leave our beloved Cubbies hitless in nine.

⁵³ 565 F. Supp. 524 (E.D. Wis. 1983).

⁵⁴ *Id.* at 539.

⁵⁵ *Ibid.*

⁵⁶ *Ibid.* at 540.

- R^2 too small to be of predictive value — the sample size of 36 was inadequate given that 13 coefficients in the regression equation had to be determined.
- Wrong variable predicted — the theoretical salary predicted by the equations consistently underestimated the actual contract values.

Question the Data⁵⁷

- Is the measurement process reliable?
- Is the measurement process valid?
- Were the data recorded correctly?
- Were the units observed the units at issue?

Question the Design

- Were the observations independent?
- What are the confounding variables?
- Can the results be generalized?

Question the Analysis

- Were all test assumptions satisfied?
- Was the most powerful test used?

Question the Presentation

- Do charts and graphs portray data fairly?
- Are rates and percentages properly interpreted?
- Were appropriate measures of location, precision, and association used?

13.7.2 Large Sample Approximations

The majority of statistics in use today represents asymptotic (large-sample) approximations whose use predates the ready availability of personal computers. The chi-square statistic and the F statistic, for example, are valid only for very large samples and, as important, only when each of the subsamples or factor divisions is also sizeable (incorporating ten observations or more).

For example, an analysis of the data in Table 13.3 derived from a study of oral lesions yields a far-from-significant p value of 14% when the chi-

⁵⁷ See Kaye and Freedman [1994] for further detail.

Table 13.3 Oral Lesions in Three Regions of India

Site of Lesion	Kerala	Gujarat	Andhra Pradesh
Labial Mucosa	0	1	0
Buccal Mucosa	8	1	8
Commissure	0	1	0
Gingiva	0	1	0
Hard Palate	0	1	0
Soft Palate	0	1	0
Tongue	0	1	0
Floor of Mouth	1	0	1
Alveolar Ridge	1	0	1

square statistic is employed, whereas the true significance level as determined by permutation methods is 2%.

The burden of proof is on the proposer of a given statistic to demonstrate that all underlying assumptions are satisfied. For example, to apply an F statistic to the analysis of variance, the individual data items must be (1) independent of one another, (2) identically distributed, and (3) taken from a normal distribution. Counter-examples abound. Here are just a few:

- Repeated measurements on the same individual are not independent.
- Successive prices of a stock are not independent, while day-to-day changes may be.
- Volunteers differ from the average, more passive, participant; thus, the first mouse selected from a cage tends to be more active, more aggressive, and have higher corticosteroid levels than the last mouse selected.

13.7.3 Multiple Statistics, Multiple Conclusions

In some situations, a choice of statistics is available. For example, in a two-sample comparison one might use either a t test or a permutation test. The two tests yield roughly equal results in the long run, accepting and rejecting in the same proportion of cases, but in any specific case, they may yield quite different results. It is important to determine during discovery whether the opposing statistician has gone “shopping” for the most favorable result.⁵⁸

⁵⁸ Chapter 15 discusses questions to ask during discovery.

13.8 Counterattack

Four factors considered in earlier chapters apply to any statistical technique:

- Data must be complete and accurate.
- Appropriate populations must be used for comparison purposes.
- Samples sizes must be adequate.
- Assumptions underlying the statistical model must be satisfied.

More precise measurements almost always yield more powerful tests. For example, a test that is based on the actual values of a variable will require far fewer samples for a given power and significance level than one for which our only knowledge is whether or not its value exceeded a certain critical threshold.

Do not accept single-valued (point) estimates. Insist on being provided with confidence intervals.

Use distribution-free statistics in preference to those that are distribution-dependent; fewer assumptions provide fewer grounds for challenge.

A result based on many factors is superior to one based on only a few; the courts recognize that a multivariate regression provides results superior to a contingency table or a correlation coefficient.

Use one statistical expert to offset the reports of another. In many cases, the courts will simply step away from the controversy.

Question the accuracy and precision of the data, the appropriateness of the survey or experimental design, and the validity and appropriateness of the analysis and of the presentation. Timely objections must be made at trial.

Chapter 14

What Every Statistician Should Know about Courtroom Procedure

The graduations and ranks of the courts are infinite, extending beyond the ken even of initiates. The proceedings in the courts of law are generally a mystery to the lower officials as well; therefore, they can almost never follow the progress of the cases they are working on throughout their course; the case enters their field of vision, often they know not whence, and continues on, they know not where.¹

Kafka thoroughly understood the trial process. Consider, Joseph K.'s experiences in Kafka's *The Trial*. First, came a lengthy process of behind-the-scenes activity in which the authorities decided whether a trial would be appropriate and the defendant cast about for possible lines of defense. Then came an equally lengthy process of discovery in which various essential documents were elicited from the opposing parties and subjected to intense scrutiny. Witnesses were made to undergo lengthy interrogations while the pretrial activity ground on. The trial was over in a few short days and statisticians who may have labored for years on trial preparation were not invited inside. In a small fraction of real-life cases, regrettably not in the case of Joseph K., a trial could be followed by an appeal.

¹ Somewhat modified from Kafka's *The Trial*.

In an informal poll of the statisticians I know who have done legal work I found that 30% worked with attorneys prior to filing, 70% participated at some point during the discovery process though only 35% actually gave depositions, 10% testified during trials, and 14% worked on appeals. We review the statistician's role during each of these phases in what follows.

14.1 Selecting the Case

You may want to decline some or all of the cases you are offered because you find the client morally objectionable or the case is without merit, or merely to avoid the embarrassment and humiliation to which an expert witness may be subjected during a hotly contested trial.

The U.S. system of justice is adversarial. You are not asked to serve in a dispassionate search for justice, but are expected to represent your client's interests to the fullest. Once you accept a case, professional ethics require that you present only one point of view and prepare substantive arguments rebutting other views. While the fees you can demand from attorneys are near the top of the scale, you should turn down any clients you feel you would be unable to represent adequately.

This raises an ethical issue. Should you as a professional be prepared to see all sides of an argument? The answer is yes until you form your initial opinion and agree to accept ongoing payments from an attorney. Your initial decision should not be made lightly or without a thorough, careful review of all the evidence at hand. Should you later alter your opinion as a result of additional evidence or additional insight, you should immediately notify the attorney who engaged you, but should this additional insight come to you suddenly on the second day of a brutal cross-examination by an opposing attorney who bears an uncommon resemblance to the schoolyard bully, your only ethical outlet is tears or feigned insanity. Your best defense against bullies is thorough, careful preparation.

14.2 Prefiling

Prior to filing, the statistician serving the plaintiff can assist in answering three questions:

- Does the data support the charge?
- What is the magnitude of any as-yet-unrealized damages?
- What additional data is needed to answer the preceding questions?

The statistician for the defense, assuming he or she has been well briefed by his or her attorney, will address these same issues and, anticipating the answers of his or her opposite number, will be asked to outline lines of defense.

You may be asked to collect and analyze the data. If so, make sure to anticipate any and all objections to the collection process. Favor the simple analysis (and the simple explanation) over the complex. The sensitivity analysis described in Section 9.3.1 is an excellent example.

14.3 Discovery

The objectives of discovery are threefold:

- To get all the relevant facts before the court
- To encourage settlement prior to trial
- To streamline the trial process by committing each side to particular facts and procedures

Statisticians will agree with the American Bar Association that the "need for full and fair disclosure is especially apparent with respect to scientific proof and the testimony of experts." This sort of evidence is practically impossible to test or rebut at trial without the opportunity to examine it closely.²

"There are no satisfactory grounds for withholding information in the discovery process."³ If you are to be called as a witness, virtually every facet of your career is open to discovery (including any unpleasant or best-overlooked items). If you have merely aided the attorney in the organization and presentation of the case, examining and commenting on documents, then everything you have done is privileged, considered part of the attorney's own work product, and not subject to examination.

My best advice in the face of these conflicting rulings is to keep your notes "barebones," but not so skimpy that you will be unable to flesh them out at trial should you be called to testify.

14.4 Depositions

As part of the discovery process, the opposing attorney may take the opportunity to question you prior to trial, usually in his office, but possibly on neutral ground. (Yes, you will get paid, but by the other attorney.)

² *ABA Standards Relating to Discovery and Procedures before Trial*, 66.

³ *State v. Tankersley*, 191 Ariz. 359, 956 P.2d 486, 495 (1998).

Both the time and the place of the deposition should be at least partially at your convenience.

Whether testifying before trial during the taking of a deposition or testifying at trial, you have one primary guideline: keep it simple. This does not mean that the statistics must be simple. It means that your well-rehearsed explanation should be straightforward and coherent.

The object of the deposition on the attorney's part is to probe his opponent's case for weaknesses. Your own objective while being deposed is to reveal as little as possible while conforming to the rules of discovery.

You may be tempted to stray from your rehearsed opinion, particularly if you feel that your honor is at stake or your judgment is questioned. Do not stray. The opposing attorney who will examine you is not your friend, will never be your friend, and cannot be won over to your side, however reasonable your arguments are.

The opposing attorney's questions will focus on the weaknesses rather than the strengths of your position.⁴ Your best defense is to keep your answers brief and, if pressed, to repeat the arguments you have already made. Never volunteer information. Whether or not the opposition has employed a statistician, the worst thing you as an expert witness can do is to volunteer a defense for an objection that has not been raised.

Listen to the questions carefully. Answer as briefly as possible. Take your time. You are not on a quiz show; you get no bonus points for the quickest or even the best answer. Do not qualify your answers or provide analogies or counter-examples. If you think of something that might clarify your answer, suppress the thought.

You do not need an answer for every question. It is better to be thought a fool than to open your mouth and prove you are one. Very likely, you will be asked, "What other methodologies were considered and why were they rejected?" The best answer you can give, at least from your own attorney's point of view, is "none." Do not lie; answer "none" only if this is truly the case. If a better answer occurs to you later in the interrogation, suppress it; take it home and tell it to your dog. To paraphrase warnings stated on cop shows, your answers are being taken down in evidence and will be used against you.

Do not guess. You may be asked for your opinions of various statistical methodologies, for example, Good's Test for Type I censored data. Your best answer is, "I am not familiar with a test under that name."⁵

⁴ You should know; you prepared the list of nasty questions for your attorney to ask the opposing party's statistician.

⁵ Of course if you're one of the few familiar with that rather obscure contribution to the statistical literature, then fire away.

In summary, your goal throughout the deposition should be the same as Dorothy's in Oz — to go home as quickly as possible where there are people who love you and care for you.

14.5 Post-Deposition, Pretrial Activities

Your attorney will expect you to prepare questions for use in his depositions of the opposing party's statistician, along with a list of documents to be elicited during discovery.⁶ Afterward, you will be asked to review the responses and prepare a further list of questions to use at trial. Your responsibilities extend from the design of forms, through the collection and storage of data, to the actual analysis. You may need to recommend further analyses or the gathering of additional data to advance your client's case.

14.6 In the Courtroom

Your appearance at trial will have three phases:

1. You will sit for hours waiting to be called as a witness only to be sent home — again — without testifying.
2. You will undergo direct examination by the attorney who engaged you or one of his or her colleagues.
3. You will be cross-examined by the same attorneys who did their best to humiliate you during your deposition.

The rules for cross-examination are essentially the same as those that guided you during the taking of your deposition. Listen to the questions. Do not be quick to respond. When you respond, be truthful, brief, and to the point. Do not embellish.

Apart from the continuing admonishment to keep it simple, the rules for direct testimony during trial are quite different from the rules for a deposition. During direct testimony, you will be doing your best to instruct a judge and jury. Teach — use all the tricks you learned in the classroom to engage the listeners. Use analogies. Will visual aids be more effective? Your attorney can arrange to have an easel or a projector and screen brought into the courtroom so that you can present tables and graphs you prepared earlier. After displaying a graph or a pie chart, start listing main points on the easel (until and unless you are asked to sit down).

⁶ See Section 15.3.1 for some suggestions.

Emphasize the positive. Talk about the procedures you used and your contribution to the analysis. Even if the data was a mess and you went through hell to clean it up and make it usable, a judge does not need to know that. If data had to be discarded, be prepared to account for it briefly and simply. Do not make bad data the focal point of your testimony.⁷

14.7 Appeals

Your work during appeals is much the same as it was during prefling and pretrial: to recognize and comment on bad data and faulty analyses.

14.8 Summary

The statistician can make effective contributions at all stages of a trial by focusing on three issues:

- Does the data support the charge?
- What is the magnitude of any as-yet-unrealized damages?
- What additional data is needed to answer the preceding questions?

Whether testifying before trial during the taking of a deposition or testifying at trial, the statistician has one primary guideline: keep it simple. This does not mean that the statistics themselves need be simple. It means that your well-rehearsed explanation must be straightforward and coherent.

The opposing attorney who will be examining you is not your friend, will never be your friend, and cannot be won over to your side, however reasonable your arguments. His questions will focus on the weaknesses rather than the strengths of your position. Be cool. Remember, when it's all over, you'll have your doctorate — oops! That was an entirely different ordeal.

Chapter 15

Making Effective Use of Statistics and Statisticians

For the lawyers — and even the least important of them has at least a partial overview of the circumstances — are far from wishing to introduce or carry out any sort of improvement in the court system, while — and this is quite characteristic — almost every statistician, even the most simple-minded among them, starts thinking up suggestions for improvement from the moment the trial starts, and in doing so often wastes time and energy that would be better spent in other ways.¹

In what follows, we consider the contributions a statistician can make at various points in the trial process, the likely areas of ignorance, and some guidelines to ensure a statistician will prove a help, not a hindrance.

15.1 Selecting a Statistician

You need a statistician if:

- The word *sample* is used.
- You need to conduct a survey or an experiment or appraise the results of a survey or an experiment.
- The opposing side has hired a statistician.

¹ Somewhat modified from Kafka's *The Trial*.

⁷ Unless you are testifying concerning the other statistician's data.

You should be looking for a statistician who:

- Is a great communicator.
- Possesses a breadth of knowledge and experience.

A good communicator can both explain and listen. He or she can explain statistical concepts so that you and your paralegal can understand them. He or she is willing to accept criticism and suggestions from you for improvement, can keep cool during depositions, and knows when to stop talking. A good communicator will listen to and follow the other person's arguments before attempting to rebut them.

Breadth of knowledge is more important than depth; you want someone who can find the holes in his or her own arguments as well as the opposition's, and who is not so wedded to techniques (the penalty sometimes of too much depth) as to be unable to utilize a statistical test or estimator more appropriate to the task at hand.

The statistician's credentials are not as critical as they are with other expert witnesses — because the statistician's explanations, not the statistician, will fall under scrutiny. A doctorate is probably preferable to a master's degree, and a doctorate in mathematical statistics coupled with some knowledge of the subject area is preferable to a doctorate in the subject coupled with some knowledge of statistics.

You can benefit by employing a statistician to supplement and interpret the testimony of other expert witnesses. In a trial pivoting on the analysis of DNA data, for example, consider employing both an expert on DNA analysis² and a statistician to provide significance levels for the expert's findings.

15.2 Prefiling Preparation

Most attorneys fail to make adequate use of statisticians during the prefiling process. A statistician can assist in estimating the extent of recoverable damages and assessing the validity of the evidence. He or she can dictate what further evidence must be collected and suggest possible lines of counterattack.

Recently, I worked for an attorney whose client was undergoing a government audit. A large fine for the client was the likely outcome. I sketched a possible line of defense based on possible differences among the various lines of business the client was engaged in. A single sample

² When judges weigh one expert's opinion against another's, credentials are essential; see, for example, *People v. Axell*, 235 Cal. App. 3d 836.

would not do because the error rates uncovered at audit would surely vary from product line to product line.³ At the attorney's request, the client's staff set about gathering the supporting evidence needed to demonstrate the differences. The government would either have to broaden its inquiry — an expensive process — or limit its claim for damages to the business lines actually sampled.

Statisticians can fulfill similar roles during the discovery process, but only assuming it is not already too late for them to be effective.

15.3 Discovery

Witnesses who understand the discovery and deposition process, who appreciate the differences in goals between depositions and courtroom testimony, and who are well versed in the techniques for effectively answering questions, survive this process without creating pitfalls to overcome at trial.⁴

A statistician is needed during the discovery process to specify what reports and documents are required, to comment on the completeness and adequacy of the documents retrieved, to help formulate a set of questions, and to aid in interpreting the responses.

Normally, any interchange, oral or written, with your statistician is exempt from discovery. The exception to the work product rule comes when the expert testifies as a witness.⁵ If you anticipate that your statistician may testify, he or she should be admonished to limit the extent of what he or she commits to paper.

I aroused the ire of my attorney client (and rightly so) when I failed to heed his admonishment to provide oral reports only. (In my defense, I only committed my points to writing when I began to feel my oral reports were being tuned out. As with any other expert witness, a little attention to the statistician's findings on the attorney's part goes a long way to ensure a positive relationship.)

15.3.1 Questions

Here is a list of questions to use during discovery. Your statistician can help with interpretation and with rephrasing questions for recalcitrant witnesses.

³ See Section 3.2.

⁴ Starr [1998, p. 121].

⁵ *U.S. v. Nobles*, 422 U.S. 225, 239 (1975).

Documents

Be sure to obtain copies of all reports and have your statistician examine them for completeness. The ideal report will permit the reader to replicate the study in its entirety. Those aspects of reports that fall short of this ideal warrant further inquiry.

The Witness (Credentials and Experience)

Jurists have made it clear that the credentials of an expert witness are important to them and help determine both admissibility and the weight to be given the evidence.⁶ As always, avoid asking questions if you are not going to like the answers.

What is your educational background?

Is statistics your primary occupation?

Have you participated in studies like this in the past?

Get the details of the expert's experience. Have you previously determined sample sizes?

Have you previously supervised the collection and storage of data?

How did you acquire your knowledge of the specific statistical techniques employed in this case?

The Sample

How were individuals or items chosen for inclusion in the sample?

Was a random mechanism employed?

What was the nature of the mechanism?

Was the sample simple random? Clustered? Stratified?

Were there controls?

What was their nature?

Were the controls adequate?

In your expert opinion was the final sample representative of the population from which it was drawn?

How were the sample and subsample sizes determined?

In your expert opinion and in retrospect, are these sample sizes adequate?

Data Collection Forms

Obtain copies of all forms used. Ask whether these forms were pretested. If they were, ask for copies of all pretest results and forms. (The opposing

⁶ See, for example, *People v. Axell supra*.

party may have kept changing the survey form or the experimental design until they got the results they wanted.)

Data Collection

How was the data collected?

How did you ensure the individual results were independent of one another?

What were the backgrounds of the individuals who collected the data? What training did the data collectors receive?

Was the training the same for all the data collectors?

How did you ensure the prescribed methods were actually followed?

Were all answers entered as given or was any interpretation of answers made prior to entry? (Get details of any interpretations.)

What percentage of forms was validated by resampling?

Who conducted the resampling?⁷

These questions are essential as the majority of errors is introduced during the collection process.⁸ If the statistician being deposed denies any knowledge of the collection process, confirm and document the lack of knowledge as it can be quite revealing at trial.

What was the percentage of non-responders? (You may want to distinguish those who were simply unavailable from those who refused to answer.)

What attempts if any were made to reduce the number of non-responders?

Were the non-responders subsampled by other means?

Data Entry

How was the data entered into the computer?

Was its correctness validated on entry?

If so, how?

What percentage of responses was validated?

⁷ In *Rust Environment & Infrastructure, Inc. v. Teunissen*, 131 F.3d 1210, 1218 (7th Cir. 1997), the court criticized a survey in part because it "did not comport with accepted practice for independent validation of the results."

⁸ The census is a prime example; see Section 3.2.2.

Data Storage

- How was the data stored?
- What procedures did you employ to ensure the integrity of the data?
- When did you last verify the integrity of the database?

Data Analysis

- What statistical techniques did you employ? (Obtain coherent written descriptions of all techniques and formulas employed along with their justification.)
- What was the power of your tests?⁹
- What assumptions underlie these procedures?
- Does the data have to be distributed in a certain way?
- Do the observations need to be independent?
- Did you verify your assumptions were satisfied?
- What alternative statistical techniques did you consider?
- What are the alternative techniques?

The Witness (Trial Experience)

- Have you testified in other trials?
- How many?
- What was their nature?
- Did you employ similar data collection and statistical procedures in those trials?
- What occasioned the differences in the present case?

15.3.2 Depositions

Many statisticians (and other expert witnesses) consider themselves more intelligent than others. They labor under the illusion that once they have explained their interpretation of the data, the other side, awestruck by their brilliance, will pack it in and slink away. The best approach is to give them some indication of the mixture of indifference and antagonism that actually lies ahead, though, inevitably, they won't quite believe it until it happens.

⁹ This question has a dual purpose. The concept of power, central to a formal study of statistical theory, is often a foreign concept to those who have only a passing knowledge of statistics. A blank look on the face of the so-called expert suggests this question and the expert's lack of knowledge need be followed up at trial.

On the other hand, your statistician can assist you in reviewing the depositions of those who are collecting and analyzing the data for your opponent. Your statistician can prepare a list of questions for trial, and, if warranted, perform further analyses and/or gather additional data.

15.4 Presentation of Evidence

The questions in Section 15.3.1 and those in the sidebar included with Section 13.6. should form the basis of both your examination and cross-examination. Hopefully, your statistician will have provided you with a list of questions specific to the case at hand.

If you do not understand your statistician's presentation, neither will the judge and jury. In that situation, have him or her read Chapter 14 and try again.

Not all statisticians understand the need for visual aids and know how to use them effectively, but all statisticians now have access to statistics computer programs that can generate outstanding charts and graphs. Show your statistician examples of the types of graphic aids you require and he or she will be able to generate them for you.

As with other expert witnesses, statisticians can testify only to their methods and findings and a confidence interval or significance level. They cannot testify to any conclusions about other aspects of the trial. In *Mahan v. Farmers Union Central Exchange Inc.*,¹⁰ for example, the court ruled:

Statisticians may testify that their statistical tests show or do not show patterns of discrimination based on age, but may not testify to the ultimate conclusion that discrimination was or was not exercised [in the case at hand]. The jury should be the final arbiter of that issue.

In *State v. Jackson*,¹¹ the court ruled the trial judge erred in allowing the expert to state a conclusion that the defendant was "probably" the father of the victim's child.

15.5 Appeals

The material in this and preceding chapters has indicated the bases on which appeals may be filed on statistical grounds. Lynda Axell appealed

¹⁰ 235 Mont. 410, 786 P.2d 850, 857 (1989).

¹¹ 320 N.C. 452 (1987).