Statistical Science
2003, Vol. 18, No. 3, 311–318

John W. Tukey and Data Analysis

David C. Hoaglin

Abstract. From the time that John W. Tukey started to do serious work in statistics, he was interested in problems and techniques of data analysis. Some people know him best for exploratory data analysis, which he pioneered, but he also made key contributions in analysis of variance, in regression and through a wide range of applications. This paper reviews illustrative contributions in these areas.

Key words and phrases: Analysis of variance, exploratory data analysis, regression.

1. INTRODUCTION

To many in statistics and other fields John Tukey may be best known for *Exploratory Data Analysis* (*EDA*), which first appeared in print in 1970, but data analysis played a major role in his work from early on. Indeed, I don't think it would be an exaggeration to say that most of John's contributions to statistics involved or grew out of problems in data analysis. Even if one focuses on data analysis itself, the number is large. Two substantial volumes of *The Collected Works of John W. Tukey* (*CWJWT*) are devoted to Philosophy and Principles of Data Analysis. It would be reasonable to include the volume on Graphics: 1965–1985, and parts of other volumes surely count as well (e.g., Factorial and ANOVA: 1949–1962). And I haven't mentioned the applications, where the object was to analyze the data themselves; nearly all sets of data have some distinctive features, and it's hard to imagine that an analysis with John as a participant would be routine.

This brief account cannot hope to cover more than a small fraction of such a corpus. Thus I offer a selection of topics, chosen to highlight several major areas: exploratory data analysis, of course, and also analysis of variance, regression and applications. In some instances I illustrate the way that John developed techniques and refined them over a period of years. A more comprehensive account would surely include time series and spectrum analysis; fortunately, Brillinger (2002) covers that area in depth.

David C. Hoaglin is Principal Scientist, Abt Associates Inc., 55 Wheeler Street, Cambridge, Massachusetts 02138-1168 (e-mail: Dave_Hoaglin@abtassoc.com).

The way that John got into statistics had a lot to do with his data-analytic orientation. As Fred Mosteller recounts in the biographical sketch that appears in each volume of *CWJWT*, John joined the Department of Mathematics at Princeton University in 1939, after completing his Ph.D. in pure mathematics (topology). His "conversion" to statistics came in part through his work on weapons problems as a member of the Fire Control Research Office at Princeton during World War II. A key influence was the biometrician and data analyst Charles P. Winsor, from whom (as John said in the dedication of *EDA*) he "learned much that could not have been learned elsewhere."

For some of the roots of John's attitude toward data analysis, however, it helps to look a little farther back. When he came to Princeton as a graduate student in 1937, he was in the Chemistry Department (he had completed his undergraduate education and taken a master's degree in chemistry at Brown University). In his foreword to the philosophy volumes of *CWJWT*, John explained,

> A respectable physical-science education—officially in chemistry, but with large doses of physics and substantial doses of geology—probably helped me a lot in understanding the character of the problems to which data brought to me were intended to be relevant. A purely mathematical background would, I believe, have left me at a severe disadvantage. Given a reasonable sensitivity to the underlying issues, seeing many sets of data seems to have made it natural to try to think about techniques in terms of the needs they might fill and the gaps

> that they left. The philosophy that appears in these two volumes is far more based on a "bottom up" approach than on a "top down" one.

This background helps me to understand his approach. So does a discussion in the last part of "The Future of Data Analysis" (Tukey, 1962a):

> If we are to make progress in data analysis, as it is important that we should, we need to pay attention to our tools and our attitudes. If these are adequate, our goals will take care of themselves.
>
> We dare not neglect any of the tools that have proved useful in the past. But equally we dare not find ourselves confined to their use. If algebra and analysis cannot help us, we must press on just the same, making as good use of intuition and originality as we know how.
>
> In particular we must give very much more attention to what specific techniques and procedures do when the hypotheses on which they are customarily developed do not hold. And in doing this we must take a positive attitude, not a negative one. It is not sufficient to start with what it is supposed to be desired to estimate, and to study how well an estimator succeeds in doing this. We must give even more attention to starting with an estimator and discovering what is a reasonable estimand, to discovering what it is reasonable to think of the estimator as estimating. To those who hold the (ossified) view that "statistics is optimization" such a study is hindside before, but to those who believe that "the purpose of data analysis is to analyze data better" it is clearly wise to learn what a procedure really seems to be telling us about. It would be hard to overemphasize the importance of this approach as a tool in clarifying situations.

A page or two later in that paper John turns to attitudes: "Almost all the most vital attitudes can be described in a type form: *willingness to face up to X*." He discusses a number of X's, including (quotation marks omitted)

- more realistic problems,
- the necessarily approximate nature of useful results in data analysis,
- the need for collecting the results of actual experience with specific data-analytic techniques,
- the need for iterative procedures,
- free use of ad hoc and informal procedures, and
- the fact that data analysis is intrinsically an empirical science.

In the area of attitudes, and with plenty of intuition and originality, John practiced what he preached.

Some of these attitudes contribute to flexibility, an important theme in John's approach to data analysis. The separation between exploratory data analysis and confirmatory data analysis allowed exploratory data analysis to proceed freely, without adherence to a unified framework, and assigned to confirmatory data analysis the systematic task of assessing the strength of the evidence. In what he wrote, John gave much more attention to exploratory. As he explained in the foreword to the philosophy volumes, "there is little doubt that exploratory gets more attention here than would be its fair share, *if these two volumes were to be one's only reference and guide*. Emphasis, however, was rightly placed where the need for more attention was greatest. Exploratory data analysis has, to my joy, been receiving more and more attention, but the pendulum of relative attention has not yet reached the balance point, through which it will, no doubt, overswing." In commenting on Bayesian analysis he mentioned "a natural, *but dangerous* desire for a unified approach" and remarked that "the greatest danger I see from Bayesian analysis stems from the belief that everything that is important can be stuffed into a single quantitative framework." For me these comments illustrate his avoidance of frameworks and unification for data analysis more generally.

2. EXPLORATORY DATA ANALYSIS

John Tukey's qualities and attitudes are nowhere more apparent than in *EDA*. The limited preliminary edition of the book came out, in three xeroxed volumes, in 1970 and 1971 (Tukey, 1970c, d, 1971a), and, after further development, the first edition followed in 1977 (Tukey, 1977a). A few years later the two volumes of "The Statistician's Guide to Exploratory Data Analysis" (Hoaglin, Mosteller and Tukey, 1983b, 1985b) provided conceptual and logical support for selected techniques and explained connections with classical statistical theory.

From the publication of the limited preliminary edition, *EDA* received an enthusiastic welcome, especially in fields that analyze data and apply statistics. In 1975 it was the topic for the first ASA-sponsored short course at the annual meeting in Atlanta. Users embraced such techniques as the stem-and-leaf display and the boxplot (née "schematic plot"), and within a few years the basic techniques, particularly displays, were available in statistical software. By now a number of those techniques have become part of statistical instruction at all levels. So, at the level of tools, the impact of EDA has been broad and lasting. I am not so sure about the attitudes, which require more effort to teach and more reflection, but I am hopeful that they will continue to spread and have a positive impact.

As a basis for a more systematic look, the introduction to *Understanding Robust and Exploratory Data Analysis* (*UREDA*) (Hoaglin, Mosteller and Tukey, 1983b) discusses four main themes that run through exploratory data analysis: resistance, residuals, re-expression and revelation:

- resistance: insensitivity to localized misbehavior in data;
- residual = data MINUS fit;
- re-expression uses a transformation (such as the logarithm or square root) to put the data in a scale that simplifies the analysis;
- revelation relies on displays to show behavior, and thus unexpected features along with familiar regularities.

In presenting the techniques that illustrate and apply these themes, John gave us a torrent of terminology, much of it newly coined. For example, stem-and-leaf display, hinges and other letter values, box-and-whisker plot, the "bulging" rule, running medians, wandering schematic plot, median polish, two-way plot, diagnostic plot, froot and flog, reroughing, double root, product-ratio analysis and pseudospreads. Some of the basic ideas appear much earlier in John's work (though not always in publications). For example, he was interested in resistance and robustness in the 1940s, as well as re-expression—in "One Degree of Freedom for Non-Additivity" (Tukey, 1949h), which I discuss below. From the EDA contributions I have selected a few that illustrate how John developed techniques over a period of time.

In these instances his approach was to devise a technique (building on insight and experience), use it on diverse data and modify or fine-tune it (or, perhaps, scrap it). This approach is a natural application of some of the attitudes that I quoted earlier (from "The Future of Data Analysis").

2.1 Fences

EDA uses "fences" to flag possible outliers. These are based on the "hinges," H_L and H_U, which are approximate quartiles of the batch. The basic idea is to calculate the H-spread, $d_H = H_U - H_L$, and lay off a multiple of it below H_L and above H_U:

$$H_L - kd_H \quad \text{and} \quad H_U + kd_H.$$

The limited preliminary edition (Tukey, 1970c) used $k = 1.0$ for the "side values" and $k = 1.5$ for the "three-halves values." By the first edition (Tukey, 1977a) the constants had changed a lot, to $k = 1.5$ for the "inner fences" and $k = 3.0$ for the "outer fences," with the labels "outside" and "far out," respectively, for data values beyond them.

The aim was not to have a formal rule for declaring an observation an outlier, but to call attention to such data for further investigation. The values of k have remained at 1.5 and 3.0, and the "inner fences" naturally see more use in practice. The most frequent application is in boxplots, as illustrated in Figure 1. The inner fences determine which data values should be plotted individually at the ends of a boxplot, and thus how far out the "whiskers" extend.

John did not arrive at 1.5 and 3.0 by setting up some sort of theoretical calculation. That step came later, as a form of evaluation, when we were working on *UREDA*; Boris Iglewicz proposed that we study the performance of the fences for data from the Gaussian and

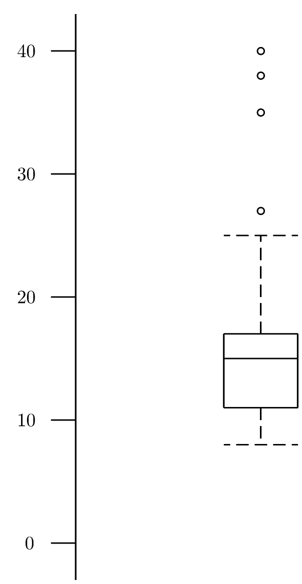

FIG. 1. *Example of a boxplot. The box extends from the lower hinge to the upper hinge and has a line across it at the median. The whiskers show the extent of the data inside the inner fences, and four observations are "outside" at the upper end.*

some heavier-tailed distributions (Hoaglin, Iglewicz and Tukey, 1986m).

2.2 Resistant Smoothing

Resistant smoothing for sequences of data (often indexed by time) illustrates an extended process of development (Tukey, 1977a, Chapters 7, 7+ and 16). The most basic operations work on short segments of the sequence, $\{y_t : t = 1, \ldots, T\}$. For example, for $t = 2, \ldots, T - 1$ the "running median of 3" replaces y_t by the median of $\{y_{t-1}, y_t, y_{t+1}\}$. John put together a favored smoother from a number of building blocks:

- Running medians (3).
- Repeated smoothing or "resmoothing" (R)—use the output of a smoothing operation (most often running medians of 3) as input to that same smoothing operation (and continue until no changes occur).
- An end-value rule—to handle y_1 and y_T.
- Splitting peaks and valleys (2 points wide: $y_{t-1} < y_t = y_{t+1} > y_{t+2}$ or $y_{t-1} > y_t = y_{t+1} < y_{t+2}$) (S)—treat the value on each side of the split as an end value.
- A component to handle steadily increasing sequences: hanning (H) (after von Hann), a weighted average with weights $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$.
- Residuals—the rough.
- Extracting an additional smooth from the rough—"reroughing."
- If both smoothers are the same, "reroughing" is called "twicing."

From all this emerged the favored smoother in *EDA*: "3RSSH, twice."

Theory related to such smoothers and their component operations was subsequently developed by Mallows (1980) and Velleman (1980).

The process of experimentation continued, and by about 1998 John had settled on two simpler choices: "3R" and "3R then 3pR." The operation 3p handles sequences in which two consecutive values are equal:

$$(\ldots, A, B, B, C, \ldots) \to (\ldots, A, M, M, C, \ldots),$$

where M is the median of A, B and C. (Thus 3p replaced the earlier operation of splitting 2-point peaks and valleys, useful because a running median of 3 leaves them unchanged.) In the second edition of *EDA* ("2EDA") he planned to cover only "the simplest useful smoothing."

2.3 Hybrid Re-expression

Hybrid re-expression is another more recent development (also intended for inclusion in 2EDA). John sometimes found that he wanted something a bit more complicated than the simplest re-expressions, so he expanded his toolkit by adding hybrids of two re-expressions. These use one re-expression to the left of x_0 and the other to the right of x_0, matching their value and slope at x_0. The one that he used most often (particularly for counts) was the *principal hybrid*, which is made up of the square root for $x \le M$ and the log for $x \ge M$:

$$\begin{aligned} &2\sqrt{Mx} - M && \text{for } x \le M, \\ &M + M \log_e(x/M) && \text{for } x \ge M. \end{aligned}$$

Here the two re-expressions are matched at the median, M, and the hybrid re-expression is matched to the data at M. More experience from use by others would be instructive.

3. ANALYSIS OF VARIANCE

Some of the same attitudes and themes that I discussed earlier and traced through EDA also stand out in John's work in ANOVA. Thus alerted, it is easy for me to see the data-analytic motivation for the pigeonhole model (Cornfield and Tukey, 1956f), but the thrust of that paper and related ones is more methodological. For their direct data-analytic flavor I prefer to focus on "One Degree of Freedom for Non-Additivity" (ODOFFNA; Tukey, 1949h) and on some aspects of "Complex Analyses of Variance: General Problems" (Green and Tukey, 1960b).

3.1 ODOFFNA

A number of John's contributions respond to the practical challenges of dealing with nonadditivity in data that are customarily handled by analysis of variance. The initial publication is "One Degree of Freedom for Non-Additivity" (Tukey, 1949h). Nonconstancy of variability and nonnormality had received considerable attention in the literature. John noted that he had more often needed to be concerned with nonadditivity, and he showed how, in a row-by-column table, to isolate a one-degree-of-freedom piece from the residual sum of squares, with the expectation that this piece would capture:

- discrepant observations;
- systematic behavior associated with analyzing the data in a scale where the effects for rows and columns are not additive.

In terms of the usual breakdown for a two-way table

$$y_{ij} = m + a_i + b_j + e_{ij},$$

the 1-d.f. piece is a multiple of $a_i b_j$.

John indicated how to use this information in choosing a transformation to reduce or remove the nonadditivity, but he did not take this aspect very far, because of limited experience. By the time *EDA* appeared, however, he had accumulated plenty of experience and had refined the approach by defining the *comparison values* $a_i b_j / m$, plotting the e_{ij} against these (the *diagnostic plot*) and interpreting a slope of p as an indication to try something like the $1 - p$ power as the re-expression (Tukey, 1977a, Sections 10F and 10G). The ability to distinguish systematic nonadditivity from discrepant observations was much enhanced by obtaining the fit and residuals from median polish (Tukey, 1977a, Section 11A).

EDA also included a discussion of fitting a multiple of $a_i b_j / m$ (instead of re-expressing the y_{ij}), and John pointed out that the resulting PLUS-one fit (writing the fitted multiple of $a_i b_j / m$ as $a_i b_j / c$),

$$\hat{y}_{ij} = m + a_i + b_j + \frac{a_i b_j}{c},$$

has an equivalent multiplicative form,

$$c A_i B_j + (m - c),$$

with $A_i = 1 + a_i / c$ and $B_j = 1 + b_j / c$. That is,

row-PLUS-column-PLUS-one

is the same family of fits as

row-TIMES-column-PLUS-one.

John also discussed the ODOFFNA fit as an example of the use of the "vacuum cleaner" (Tukey, 1962a).

Looking back on this line of development in the foreword to Volume VII of *CWJWT*, he listed four branches of extensions from ODOFFNA (pages l–li):

- higher-order single degrees of freedom;
- the recognition that a purely multiplicative fit differs from an additive fit by an odoffna single degree of freedom [the PLUS-one fit];
- breakdowns into low-rank (but not single-degree-of-freedom) constituents [as in the "vacuum cleaner," higher-rank fits in McNeil and Tukey (1975c), and work by John Mandel, Ruben Gabriel and others, conveniently summarized by Emerson and Wong (1985)];
- graphical replacement of odoffna by *diagnostic plots*.

3.2 Complex Analyses

The paper by Green and Tukey (1960b) has a thoroughly practical orientation and covers a number of important issues. I am grateful to Bert Green for some of the background details. The initial stimulus was a set of data analyzed by Johnson and Tsao (1944) and published by Johnson (1949). The work began in 1950, when John recruited Bert, then a second-year graduate student in psychometrics, to help with reanalyzing the data. As that analysis unfolded and each iteration suggested the next, Bert did the calculations—on an electromechanical calculator! The complexity of the example was genuine: the data layout had 448 observations, and the initial ANOVA table had 39 lines.

The dataset itself is of interest, especially because John used it as a source of examples over the years. The experiment, from psychophysics, measured difference limens for weights "by a method of continuous change. An aluminum pail was attached by a lever system to a ring on the subject's finger. One of seven weights—100, 150, 200, 250, 300, 350, or 400 grams—was placed in the pail. By controlling a flow of water into the pail, the experimenter increased the load in the pail at one of four constant rates—50, 100, 150, or 200 grams per 30 seconds—until the subject reported a change in pull. The difference limen, *DL*, in grams, was measured by the amount of water added. Four men and four women served as subjects. Two of each sex were congenitally blind, the others had normal vision. ... [T]he order of presentation of the weight–rate combinations was randomized. For each weight–rate combination, five determinations of the *DL* were made for each subject. The mean of these five measurements was used in the analysis. The entire experiment was carried out on each of two days, one week apart." Thus the data involve six factors:

- S = sex
- I = sight
- P = person
- R = rate
- W = weight
- D = date

John used a small part of the data throughout the chapter on three-way fits in *EDA* (Tukey, 1977a, Chapter 13) and other parts as examples on other occasions (which I have not attempted to catalog).

Analysis of those data involved a considerable number of questions. For "students" who are prepared to do some homework, the paper is almost a mini-textbook in practical ANOVA. For the present account I would like to single out two "lessons": aggregation and choice of response variable.

Aggregation. Once the initial computations have produced sums of squares and mean squares, according to "the shape of the analysis" (e.g., which classifications are crossed and which are nested), the real work of the analysis can begin. One aim is "to provide a simple summary of the variation in the experimental data." Green and Tukey approach this task by aggregating lines in the analysis, for which they give specific rules. (The "lines in the analysis" are the lines in the ANOVA table, corresponding to the main effects of the factors, two-factor interactions, three-factor interactions, etc., and ending with the residuals. Each line has, at a minimum, a sum of squares, its degrees of freedom and its mean square.) One component generalizes the rule of thumb of Paull (1950): "a line should be aggregated with a basic line only when... the first mean square is less than twice the basic mean square." Two other features of their procedure are important for the present discussion. First, the choice of lines for application of this "Rule of 2" is guided by the expected mean squares. Second, they start with the lowest (i.e., highest-order) line in the ANOVA table (as the "basic line") and lines "above" it. (If the expected mean square for one line contains all the terms in the expected mean square for another line, the first line is "above" the second line.)

Not surprisingly, the question of aggregation receives considerable attention in *Fundamentals of Exploratory Analysis of Variance* (*FEAV*; Hoaglin, Mosteller and Tukey, 1991h, Chapter 11), though in a framework of rather different terminology. The Rule of 2 remains, but other aspects of the procedure have changed substantially. Another rule of thumb now advises: "start by treating all factors as random." The process begins at the top, with the zeroth-order or common term (which usually will not be combined). And for each term of a given order, the mean square is compared with the mean squares for all appropriate terms of the next higher order (e.g., in the example in *FEAV* the mean square for M would be compared with the mean squares for the two-factor interactions $M \times A$, $M \times T$ and $M \times D$). It might be of interest to compare the two procedures, but I am not aware of anything that has been written about this.

Returning to the Green–Tukey paper and the difference limen data, their aggregation procedure reduced the initial 39 lines to 15. In the resulting analysis the main effects of person and rate were the key part of the story. Thus aggregation did much to simplify the summary of the variation in the data.

Choice of response variable. The other "lesson" in simplifying an analysis is choice of the response variable. After considerable work on the difference limen data, Green and Tukey discovered, to their chagrin, that analysis of response time (in seconds) was much better than analysis of difference limen (in grams). The interpretation is that "each person in the experiment responds after a constant Time, regardless of the Rate." The transformation of the data divides each value of the dependent variable by the corresponding value of Rate. This operation is different from re-expression (discussed above as part of EDA), and it has its own name: "reformulation." In the actual re-analysis (Can you hear the whir of Bert's calculator?!), Green and Tukey also used a re-expression, the logarithm of response time, to stabilize variability.

In summary, for ANOVA John contributed powerful techniques that can be brought to bear when one analyzes data.

4. REGRESSION

In the general area of regression John made a number of key contributions. Chapters 12–16 of *Data Analysis and Regression* (Mosteller and Tukey, 1977b) should be required reading for both students and practitioners.

Some of his ideas paved the way for a number of effective regression diagnostics. The leave-out-one mechanism associated with the jackknife (Tukey, 1958g) underlies various measures of the influence of single observations, including Cook's distance (Cook, 1977) and DBETAS and DFITS of Belsley, Kuh and Welsch (1980). And the information on leverage is contained in the "hat matrix," $H = X(X^T X)^{-1} X^T$, so named because it takes y into $\hat{y} = Hy$. Of course, $X(X^T X)^{-1} X^T$ has been around as long as we have had regression in matrix notation, but a compact and evocative name makes it more accessible to some audiences. Again we see John's penchant for terminology. As I recall, I first saw the name "hat matrix" in mimeographed notes for a course, around 1968. For the case of a two-way table Anscombe and Tukey (1963b) gave the formulas for all elements of H.

Besides *Data Analysis and Regression* John's writings indicate that he gave considerable attention to regression. Beaton and Tukey (1974c) discussed, at length, issues that arise in fitting polynomials to equally spaced data. They were motivated (in connection with the Dartmouth Conference on Critical Evaluation of Chemical and Physical Structural Information) by data on the band spectrum of hydrogen fluoride—a prototype situation because the equal spacing is theoretically precise and the data are accurate to many decimal places. Their topics include orthogonalization, various aspects of least-squares fitting, robust-resistant fitting, nonlinear smoothing, the degree of the polynomial, indicators for stopping and balancing bias and variance.

Earlier Anscombe and Tukey (1963b) examined an extensive kit of techniques for working with residuals. Though their focus was mainly on one-way and two-way classifications, several of the techniques are applicable to regressions with less structure and have been widely used.

Also, I am aware of two substantial unpublished pieces: "Introduction to the Dilemmas and Difficulties of Regression" (1979) and "Practical Regression for 1983" (1982, with Paul Velleman). And in 1991 John circulated two series of short notes related to regression.

Some of the scientific applications suggest that it is appropriate to end this discussion of regression by quoting briefly from the unpublished latter part of the manuscript whose published part appeared as "Causation, Regression, and Path Analysis" (Tukey, 1954b): "Most of us, I fear, are far from being at home with determinate functions of several variables, so that when these determinate functions become covered up with fluctuations, and at best exist as regression surfaces, it is not surprising that we have discomfort and difficulty. Such comfort and ease as I have in multiple-variable statistical situations is due, I believe, to learning about the determinate case as a student of physical chemistry." John's scientific training had a sizable impact on his statistical work.

5. APPLICATIONS

In the work that I have been discussing, specific applications are often in view. Throughout his career John also put great effort into a wide variety of projects in which the focus was primarily the applications themselves. Some of these left no published trace (nor any public trace—for example, for reasons of national security). For others we do have a published account. I mention only a few of these activities as illustrations:

- The Kinsey Report (Cochran, Mosteller and Tukey, 1953b, 1954c);
- Panel on Seismic Improvement (Tukey, 1959e);
- Environmental pollution (Tukey et al., 1965n; Tukey, 1966a);
- The National Halothane Study (Subcommittee on the National Halothane Study, 1966d; Gentleman, Gilbert and Tukey, 1969a);
- National Assessment of Educational Progress (Tukey, 1970a);
- Impacts of Stratospheric Change (Tukey, 1976h);
- Adjustment of the U.S. Census (Ericksen, Kadane and Tukey, 1989x; Tukey, 1990t).

Many of these activities involved John's participation for a number of years, during which he contributed in many ways besides advising on or guiding data analysis.

6. CONCLUSION

Over the years John received much recognition for his many accomplishments—in the form of medals (including the National Medal of Science), awards and honorary degrees. As his student (1966–1970) and, later, collaborator, I was aware of many of his contributions, though not always of the details. The process of reviewing part of his work in data analysis has given me a renewed appreciation of their number, breadth, depth and impact. For nearly 60 years, statistics, science and the nation benefited enormously from the efforts of John W. Tukey.

ACKNOWLEDGMENTS

I am grateful to Bert F. Green for background information on the paper by Green and Tukey (1960b) and to F. R. Anscombe, Karen Kafadar, Frederick Mosteller, Howard Wainer and a referee for helpful comments and suggestions.

REFERENCES

(In items with John Tukey as author or co-author the letter following the year matches the bibliography in Volume VII of *CWJWT*.)

ANSCOMBE, F. J. and TUKEY, J. W. (1963b). The examination and analysis of residuals. *Technometrics* **5** 141–160, 536–537, **6** 127.

BEATON, A. E. and TUKEY, J. W. (1974c). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics* **16** 147–185.

BELSLEY, D. A., KUH, E. and WELSCH, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Wiley, New York.

BRILLINGER, D. R. (2002). John W. Tukey's work on time series and spectrum analysis. *Ann. Statist.* **30** 1595–1618.

CLEVELAND, W. S., ed. (1988a). *The Collected Works of John W. Tukey V. Graphics: 1965–1985*. Wadsworth, Belmont, CA.

COCHRAN, W. G., MOSTELLER, F. and TUKEY, J. W. (1953b). Statistical problems of the Kinsey report. *J. Amer. Statist. Assoc.* **48** 673–716.

COCHRAN, W. G., MOSTELLER, F. and TUKEY, J. W. (1954c). *Statistical Problems of the Kinsey Report on Sexual Behavior in the Human Male*. Amer. Statist. Assoc., Washington, DC.

COOK, R. D. (1977). Detection of influential observations in linear regression. *Technometrics* **19** 15–18.

CORNFIELD, J. and TUKEY, J. W. (1956f). Average values of mean squares in factorials. *Ann. Math. Statist.* **27** 907–949.

COX, D. R., ed. (1992a). *The Collected Works of John W. Tukey VII. Factorial & ANOVA*: *1949–1962*. Wadsworth, Pacific Grove, CA.

EMERSON, J. D. and WONG, G. Y. (1985). Resistant nonadditive fits for two-way tables. In *Exploring Data Tables*, *Trends*, *and Shapes* (D. C. Hoaglin, F. Mosteller and J. W. Tukey, eds.) 67–124. Wiley, New York.

ERICKSEN, E., KADANE, J. B. and TUKEY, J. W. (1989x). Adjusting the 1980 Census of Housing and Population. *J. Amer. Statist. Assoc.* **84** 927–944.

GENTLEMAN, W. M., GILBERT, J. P. and TUKEY, J. W. (1969a). The smear-and-sweep analysis. In *The National Halothane Study* (J. P. Bunker et al., eds.) 287–315. National Institutes of Health/National Institute of General Medical Sciences, Bethesda, MD.

GREEN, B. F., JR. and TUKEY, J. W. (1960b). Complex analyses of variance: General problems. *Psychometrika* **25** 127–152.

HOAGLIN, D. C., IGLEWICZ, B. and TUKEY, J. W. (1986m). Performance of some resistant rules for outlier labeling. *J. Amer. Statist. Assoc.* **81** 991–999.

HOAGLIN, D. C., MOSTELLER, F. and TUKEY, J. W., eds. (1983b). *Understanding Robust and Exploratory Data Analysis*. Wiley, New York.

HOAGLIN, D. C., MOSTELLER, F. and TUKEY, J. W., eds. (1985b). *Exploring Data Tables*, *Trends*, *and Shapes*. Wiley, New York.

HOAGLIN, D. C., MOSTELLER, F. and TUKEY, J. W., eds. (1991h). *Fundamentals of Exploratory Analysis of Variance*. Wiley, New York.

JOHNSON, P. O. (1949). *Statistical Methods in Research*. Prentice-Hall, New York.

JOHNSON, P. O. and TSAO, F. (1944). Factorial design in the determination of differential limen values. *Psychometrika* **9** 107–144.

JONES, L. V., ed. (1986c). *The Collected Works of John W. Tukey III. Philosophy and Principles of Data Analysis*: *1949–1964*. Wadsworth, Belmont, CA.

JONES, L. V., ed. (1986e). *The Collected Works of John W. Tukey IV. Philosophy and Principles of Data Analysis*: *1965–1986*. Wadsworth, Belmont, CA.

MALLOWS, C. L. (1980). Some theory of nonlinear smoothers. *Ann. Statist.* **8** 695–715.

MCNEIL, D. R. and TUKEY, J. W. (1975c). Higher-order diagnosis of two-way tables, illustrated on two sets of demographic empirical distributions. *Biometrics* **31** 487–510.

MOSTELLER, F. and TUKEY, J. W. (1977b). *Data Analysis and Regression*: *A Second Course in Statistics*. Addison-Wesley, Reading, MA.

PAULL, A. E. (1950). On a preliminary test for pooling mean squares in the analysis of variance. *Ann. Math. Statist.* **21** 539–556.

Subcommittee on the National Halothane Study of the Committee on Anesthesia, National Academy of Sciences–National Research Council (1966d). Summary of the National Halothane Study: Possible association between halothane anesthesia and postoperative hepatic necrosis. *J. Amer. Med. Assoc.* **197** 775–788.

TUKEY, J. W. (1949h). One degree of freedom for non-additivity. *Biometrics* **5** 232–242.

TUKEY, J. W. (1954b). Causation, regression, and path analysis. In *Statistics and Mathematics in Biology* (J. W. Gowen, O. Kempthorne and J. Lush, eds.) 35–66. Iowa State College Press, Ames.

TUKEY, J. W. (1958g). Bias and confidence in not-quite large samples (abstract). *Ann. Math. Statist.* **29** 614.

TUKEY, J. W. (1959e). Appendix 9: Equalization and pulse shaping techniques applied to the determination of initial sense of Rayleigh waves. In *The Need for Fundamental Research in Seismology* 60–129. Report of the Panel on Seismic Improvement, U.S. State Department, Washington.

TUKEY, J. W. (1962a). The future of data analysis. *Ann. Math. Statist.* **33** 1–67, 812.

TUKEY, J. W. (1966a). Statement on pollution. In *Federal Research and Development Programs*, *Hearings before a Sub-Committee of the Committee on Government Operations*, *House of Representatives*, *89th Congress*, *Second Session*, *January 1966* 121–126. U.S. Government Printing Office, Washington.

TUKEY, J. W. et al. (1970a). *1969–1970 Science*: *National Results and Illustrations of Group Comparisons*, *July 1970.* Report I. National Assessment of Educational Progress, Ann Arbor, MI.

TUKEY, J. W. (1970c). *Exploratory Data Analysis* **1** (limited preliminary edition). Addison-Wesley, Reading, MA.

TUKEY, J. W. (1970d). *Exploratory Data Analysis* **2** (limited preliminary edition). Addison-Wesley, Reading, MA.

TUKEY, J. W. (1971a). *Exploratory Data Analysis* **3** (limited preliminary edition). Addison-Wesley, Reading, MA.

TUKEY, J. W. et al. (1976h). *Halocarbons*: *Environmental Effects of Fluoromethane Release*. Committee on Impacts of Stratospheric Change, National Academy of Sciences, Washington.

TUKEY, J. W. (1977a). *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.

TUKEY, J. W. et al. (1990t). Proposed guidelines for statistical adjustment of the 1990 Census. In *Hearing before the Sub-Committee on Census and Population of the Committee on Post Office and Civil Service*, *House of Representatives*, *101st Congress*, *Second Session*, *January 30*, *1990* 177–190. U.S. Government Printing Office, Washington.

TUKEY, J. W., ALEXANDER, M., BENNETT, H. S. et al. (1965n). *Restoring the Quality of Our Environment*. Report of the Environmental Pollution Panel, President's Science Advisory Committee. U.S. Government Printing Office, Washington.

VELLEMAN, P. F. (1980). Definition and comparison of robust nonlinear data smoothing algorithms. *J. Amer. Statist. Assoc.* **75** 609–615.