The one-way-table

Laurie Davies
Department of Mathematics and Computer Science
Universität Essen

November 4, 2002

Abstract: The one-way analysis of variance is concerned with comparing the locations of several one-dimensional samples. This paper gives a simple unified semi-graphical and semi-analytical approach to the problem based on approximation intervals for the locations of the samples. The intervals are standardized for normally distributed data to achieve given coverage probabilities. No assumptions placed on the sample sizes or the sample scales. All questions concerning the relationships between the locations of the samples are reduced to questions concerning the approximating intervals. The procedure is stable in that small perturbations either of the data or the standardizing model lead to only small changes in the analysis

1 Introduction

1.1 Procedure orientation

Tukey's work on the analysis of variance spans almost forty-five years from Tukey (1949) to Tukey (1993) with perhaps earlier and later works of which I am not aware. Particularly relevant for the one-way table are Tukey (1952a), Tukey (1952b) and Tukey (1953). Possible reasons for Tukey's long interest in the analysis of variance are its continuing importance for practical data analysis and its use as an exemplar to demonstrate more general problems.

In particular we mention Tukey's emphasis on procedures which, in a narrow sense, may be seen as computer programs designed for some specific task. We quote from Tukey (1993):

> Theoretical results typically have assumptions. However applicable procedures, even those suggested by theoretical results, are typically *never* used where theory's assumptions apply exactly and in detail. Thus applicable procedures do not themselves have assumptions - only some circumstances in which they work (i.e. serve our purposes) better, and some in which they work less well. ... I know some will think these statements heretical, but I find no escape from them and their implications.

Seen in this light one task for the statistician is provide statistical procedures in the form of software which are accompanied by some indication of their range of applicability. Suppose we wish to do this for a procedure for the one-way-analysis of variance. The indication should not be narrow and precise as in

> *Procedure to be used only for i.i.d. Gaussian data with equal variances. Different sample sizes are allowed.*

which would preclude its use but wide and somewhat fuzzy as in

> *Procedure is reliable for (a) different sample sizes but with a minimum size of 2 , (b) at most 50 samples to be compared, (c) different scales, (d) almost symmetric data with up to* 30% *symmetric or extreme outliers or less than* 10% *moderate one-sided outliers.*

Such an indication of the range of applicability of a procedure requires that it has been subject to some form of evaluation. This will include simulations and mathematical probes on various test beds as well as tests using real data.

1.2 Approximating data

In this paper stochastic models are consistently treated as approximations to real data (Davies 1995, Tukey 1993). No assumptions are made concerning a true data generating mechanism. In particular it is not assumed that there exists a true distribution function F or a true mean μ which lie behind the data. Apart from avoiding the ontological excesses of most of the statistical literature this approach has more radical implications in that there is almost no rôle for density based methods such as likelihood. A stochastic model P is regarded as an adequate approximation to a real data set $\boldsymbol{x}_n = (x_1, \ldots, x_n)$ if "typical" data sets $\boldsymbol{X}_n(P) = (X_1(P), \ldots, X_n(P))$ generated under P "look like" $\boldsymbol{x}_n$. One can imagine the following thought experiment. Using the model P we generate 999 samples $\boldsymbol{X}_{in}(P) = (X_{i1}(P), \ldots, X_{in}(P)), i = 1, \ldots, 999$, of size n under P. The real data set are inserted at random giving 1000 samples in all. The word "typical" is quantified by a real number α, $0 < \alpha < 1$, such as 0.95. The statistician is required to specify $0.95 \times 1000 = 950$ of the data sets as being typical. If the real data set is included in the typical data sets then P is regarded as an adequate approximation. In order to specify the typical data sets a decision rule is required which may then be seen as a precise definition of what is meant by approximation. We indicate how to do this when the main interest is location defined by some location functional T_L. The proposed rule requires an auxiliary scale functional T_S. The model we consider is a normal distribution $N(\mu, \sigma^2)$. For each of the 1000 samples we calculate $(T_L - \mu)/T_S$ and define as atypical those samples associated with the 25 smallest and the 25 largest values of $(T_L - \mu)/T_S$. If the real sample is one of the remaining typical samples then μ is regarded as an adequate approximation. We can repeat this not for 999 but for arbitrary many simulated samples. In the limit as the number of samples tends to infinity it is seen that the set of μ-values which are regarded as an adequate approximation is given by the interval

$$[\,T_L(\boldsymbol{x}_n) - qu(0.975, n)T_S(\boldsymbol{x}_n)/\sqrt{n}, T_L(\boldsymbol{x}_n) + qu(0.975, n)T_S(\boldsymbol{x}_n)/\sqrt{n}\,] \quad (1)$$

where $qu(0.975, n)$ denotes the 0.975-quantile of $(T_L - \mu)/T_S$ under the $N(\mu, \sigma^2)$ distribution. In this particular case $qu(0.975, n)$ is independent of the parameters μ and σ.

We illustrate the difference using the concept of estimation. In statistics this typically means estimating the parameter θ of the true underlying distribution P_θ. From our point of view this makes no sense as there is no underlying distribution P_θ. A similar point of view is taken by Donoho (1988) who also expresses his disquiet about statistical inference for objects whose very existence cannot be shown. Donoho was referring to densities but his remarks apply with equal force to theoretical distribution functions. Looked at from the point of view of approximation the problem is to specify those parameter values θ, if any, such that P_θ is an adequate approximation for the data.

There is no need to make untenable assumptions about the real data. We note that if the data have been generated on a test bed, that is with some specified distribution P_{θ_0}, then it does make sense to talk of estimating θ_0. When analysing real data sets the statistician often wishes to estimate some aspect of the real world such as the amount of nitrate in a sample of drinking water. The identification of such real existing quantities with the values of parameters in stochastic models is and will remain speculative: statistics has an irreducible speculative element which is rarely discussed. For a more detailed discussion of this and other matters we again refer to Tukey (1993), Davies (1995) and the introduction of Davies and Kovac (2001).

2 Approximation intervals

2.1 Construction

The procedure we propose is based on calculating an approximation intervals $I_j, j = 1, \ldots, k$, for each sample as given by (1). To operationalize this we must specify the location functional T_L and the scale functional T_S. The performance of the procedure will be determined solely by the choice of T_L and T_S. If stability of analysis is required, that is small to moderate changes in the data or the model should cause only small to moderate changes in the analysis, then it would seem necessary to choose robust functionals T_L and T_S. The default ones we propose will be denoted by T_{tm} and T_{ts} and are as follows. In a first step any possible outliers are eliminated using the following version of Hampel's rule (Hampel (1985)). Values x_i for which

$$|x_i - \mathrm{Med}(\boldsymbol{x}_n)| \geq c(n)\mathrm{Mad}(\boldsymbol{x}_n)$$

are eliminated from the sample. Here Med and Mad denote respectively the median and median absolute deviation of the sample. The factor $c(n)$ is given by

$$c(n) = \begin{cases} 30 & n = 3, \\ 10 & n = 4, 5, 6, 7, \\ 7 & n \geq 8. \end{cases} \tag{2}$$

This choice of $c(n)$ is motivated by Davies and Gather (1993). For the sake of notational simplicity we continue to denote the sample by $\boldsymbol{x}_n$ even after the elimination of the outliers. The median and median absolute deviation of the sample are calculated and a truncated sample $\boldsymbol{x}_n^t = (x_1^t, \ldots, x_n^t)$ is defined as follows;

$$x_i^t = \begin{cases} \text{Med}(\boldsymbol{x}_n) - 3\text{Mad}(\boldsymbol{x}_n) & \text{if } x_i \leq \text{Med}(\boldsymbol{x}_n) - 3\text{Mad}(\boldsymbol{x}_n) \\ \text{Med}(\boldsymbol{x}_n) + 3\text{Mad}(\boldsymbol{x}_n) & \text{if } x_i \geq \text{Med}(\boldsymbol{x}_n) + 3\text{Mad}(\boldsymbol{x}_n) \\ x_i & \text{otherwise.} \end{cases} \tag{3}$$

For normally distributed data this is roughly equivalent to truncating at the mean plus or minus two standard deviations. The location functional T_{tm} is defined to be the mean of the truncated sample and the scale functional T_{ts} to be the standard deviation together with the following finite sample correction

$$T_{ts}(\boldsymbol{x}_n) = SD(\boldsymbol{x}_n^t)/fshscl(n) \tag{4}$$

to make it an unbiased estimate for the standard deviation of a Gaussian distribution. The function $fshscl(n)$ is given by

$$fshscl(n) = \begin{cases} 0.964 - 1.21/n, & \text{if } n \text{ odd,} \\ 0.964 - 0.89/n^{0.85}, & \text{if } n \text{ even.} \end{cases} \tag{5}$$

The quantity $qu(\alpha, n)$ is the α-quantile of the statistic $(T_{tm} - \mu)/T_{ts}$ under the $N(\mu, \sigma^2)$ distribution. This implies

$$P(\mu \in [\,T_{tm} - qu((1+\alpha)/2, n)T_{ts}/\sqrt{n}, T_{tm} + qu((1+\alpha)/2, n)T_{ts}/\sqrt{n}\,]) = \alpha \tag{6}$$

for Gaussian data. An exact analytic expression for $qu(\beta, n)$ is not possible but using simulations a simple analytic approximation of the form

$$\tilde{q}(\beta, n) = qt((1+\beta)/2, n-1)\exp(A(n) + B(N)\log(1-\beta) + C(n)\log^2(1-\beta)) \tag{7}$$

can be derived where $qt(\beta, k)$ denote the β-quantile of the t-distribution with k degrees of freedom. The coefficients $A(n), B(n)$ and $C(n)$ were determined on the basis of 100000 simulations. For β in the range $0.95 \leq \beta \leq 0.9995$ the percentage error compared with the simulated values is less than 2% for $n = 3$ and less than 1% for $n \geq 4$. Our final approximation interval is then given by

$$[\,T_{tm} - \tilde{q}(\alpha, n)T_{ts}/\sqrt{n}, T_{tm} + \tilde{q}(\alpha, n)T_{ts}/\sqrt{n}\,] \tag{8}$$

2.2 Test beds: covering probabilities and efficiency

As mentioned above statistical procedures should in general be designed to be stable. One way of checking this is to examine the behaviour of the procedure on different test beds. The procedure described in the last section is gauged so that the probability that the approximation interval (8) contains the mean μ of a Gaussian distribution is α. If the procedure is stable then this probability should be close to α for symmetric distributions other than the Gaussian. Table 1 gives the results for the slash distribution for various samples sizes and values of α.

n	0.9	0.95	0.99	0.999
3	0.9291	0.9645	0.9930	0.9994
5	0.9352	0.9728	0.9942	0.9993
10	0.9093	0.9650	0.9958	0.9997
25	0.8826	0.9432	0.9905	0.9995
50	0.8747	0.9339	0.9871	0.9989
100	0.8685	0.9272	0.9833	0.9983

Table 1: Covering probabilities for the slash distribution.

The efficacy of a procedure is of importance and can be quantified by comparison with other procedures. In the case of approximation intervals given by (8) the efficacy on a test bed can be measured by the lengths of the approximation intervals for a given probability of the interval covering the location value used for the simulations. This depends to a large extent on the relative efficiency of the location functional T_{tm}. If care is taken the efficiency can also be measured relative to the optimal location estimator on the test bed. A necessary condition for this is that the distribution defining the test bed should be "bland" or "hornless" (Tukey) and not offer any handles allowing a gratuitous increase in efficiency. An example of a hornless distribution is the normal distribution. Because the normal distribution minimizes the Fisher information for a given variance it is difficult to estimate the mean of a normally distributed sample. This is the justification for the wide spread use of the normal distribution as a test bed. An example of a horned distribution is the Cauchy whose peakedness at the origin does permit a gratuitous increase in efficiency (Cohen (1991)). For this reason we use the slash distribution rather than the Cauchy as in Morgenthaler and Tukey (1991). Indeed Morgenthaler and Tukey (1991) argue that the Gaussian and slash test beds are in a sense sufficient for testing robust location functionals and so we confine ourselves to these two challenges. We compare the functional T_{tm} with the mean, the median and with the maximum likelihood estimator T_{sl} for the location parameter of the slash distribution. The functional T_{tm} is calculated *after* the elimination of the outliers. The other three location functionals are calculated with the initial data. The relative efficiencies on the Gaussian and slash test beds for several different sample sizes are shown in Table 2. The results are based on 10000 simulations.

	Gauss				Slash			
n	mean	median	T_{tm}	T_{sl}	mean	median	T_{tm}	T_{sl}
3	100.0	73.8	86.2	66.0	0.0	50.2	21.1	100.0
4	100.0	83.5	92.6	75.8	0.1	88.8	64.0	100.0
5	100.0	69.6	91.4	71.9	0.0	98.0	74.5	100.0
6	100.0	76.6	94.2	70.8	0.0	92.9	65.3	100.0
7	100.0	68.3	94.1	73.4	0.0	84.3	64.6	100.0
8	100.0	73.7	95.2	71.1	0.0	90.8	77.3	100.0
9	100.0	66.9	95.1	72.6	0.0	85.7	76.3	100.0
10	100.0	72.0	95.9	71.3	0.0	90.1	78.7	100.0
20	100.0	69.0	97.6	72.9	0.0	85.4	80.2	100.0
50	100.0	63.6	97.9	70.8	0.0	79.8	80.6	100.0

Table 2: Relative efficiencies of the mean, the median, T_{tm} and T_{sl} expressed as a percentage of the most efficient functional.

2.3 Other possibilities

The approximation interval (8) is the default version. The software also allows for approximation intervals based on the mean and standard deviation and also on the M-estimator $T_m = (T_{lm}, T_{sm})$ defined by

$$\int \psi\left(\frac{x - T_{lm}(\boldsymbol{P})}{T_{sm}(\boldsymbol{P})}\right) d\boldsymbol{P} = 0 \tag{9}$$

$$\int \chi\left(\frac{x - T_{lm}(\boldsymbol{P})}{T_{sm}(\boldsymbol{P})}\right) d\boldsymbol{P} = 0 \tag{10}$$

where

$$\begin{aligned} \psi(x) &= (\exp(x/5)-1)/(\exp(x/5)+1) && (11)\\ \chi(x) &= (x^4-1)/(x^4+1). && (12) \end{aligned}$$

The advantage of having several different procedures is that the results can be compared. Tukey recommends two or three.

3 The one-way table

3.1 The problem and a procedure

The one-way table is concerned with differences in location of k samples which we denote by

$$\boldsymbol{x}_{in_i} = (x_{i1}, \ldots, x_{in_i}),\ i = 1, \ldots, k.$$

At the most restrictive level it is assumed that the data are normally distributed with equal variances but possibly different means (see for example Christensen (1987)). Within this framework questions concerning the locations are often formulated in terms of hypotheses which are then tested. For example the question of whether the first three samples have relevantly different locations is answered by testing the hypothesis

$$H_0 : \mu_1 = \mu_2 = \mu_3. \tag{13}$$

Apart from the questionable assumption of normally distributed samples the null hypothesis H_0 of (13) is unlikely to be true or believed in its stated precision. This places the statistician in the somewhat uncomfortable position of formally testing a hypothesis whose falsity is not doubted whatever the result of the formal test.

The ideal of normally distributed samples with equal variances is one not likely to occur in practice: the samples may have different scales and shapes and include possible outliers or exotic observations. We refer to Miller (1986).

We propose here a very simple partly analytic and partly graphical procedure which allows a decision about any question concerning the location values. As a first step we construct the approximation intervals $I_i, i = 1, \ldots k$, as described in Section 2.1 for each of the k samples $\boldsymbol{x}_{in_i}, i = 1, \ldots k$. The default value of α we use is

$$\alpha = \alpha_k = 0.95^{1/k}. \tag{14}$$

The approximation (7) is accurate for $k \leq 50$ for α_k as in (14). Consider now the case of Gaussian test beds where the samples are independent and the i-th sample is distributed as $N(\mu_i, \sigma_i^2)$. On such test beds and with the choice of α in (14) it follows that

$$\boldsymbol{P}(\mu_i \in I_i, i = 1, \ldots, k) = 0.95. \tag{15}$$

Having constructed the intervals all questions concerning relationships between the location values are reduced to questions concerning the intervals. Using the terminology of approximation we can replace the testing of the hypothesis (13) by the question as to whether the first three samples can be adequately approximated by some common location values. As I_i is the set of approximating values for the i-th sample the set of joint approximating values is the intersection $I_1 \cap I_2 \cap I_3$. We now describe the procedure in more detail.

3.2 Grouping the samples

In forming groups considerations of simplicity are taken as a guiding principle. The number of groups is the number of different values of the location parameters required to give an adequate description of the data. The simplest situation is where there is at least one value which is simultaneously adequate for all the k samples. This is the case if and only if the intersection of the $I_i, i = 1, \ldots, k$, is non-empty. For data generated on Gaussian test beds with

$$\mu_1 = \mu_2 = \ldots = \mu_k = \mu$$

it follows from (15) that

$$\boldsymbol{P}(\mu \in \bigcap_{i=1}^{k} I_i) = 0.95. \tag{16}$$

Thus with probability 0.95 there exists at least one value adequate for all k intervals and furthermore the intersection of the intervals gives a 0.95-confidence interval for μ.

If the intersection of all the intervals is empty then at least two different values of the location parameter will be required to adequately approximate the data. We describe how the minimum number of values can be calculated. Suppose that $I_i = [a_i,\, b_i]$ and without loss of generality we assume that the

a_i are monotone increasing, that is $a_1 \leq a_2 \leq \ldots \leq a_k$. If not then we simply relabel them so that they are. Consider the smallest i_1 such that

$$\min_{1 \leq i < i_1} b_i < a_{i_1}.$$

Such an i_1 exists as the intersection of all the intervals is empty. The first group G_1 is given by

$$G_1 = \{i : 1 \leq i < i_1 \text{ and } b_i < a_{i_1}\}.$$

It is clear that $\cap_{1 \leq i < i_1} I_i \neq \emptyset$ so that all the samples $\boldsymbol{x}_{in_i}$ with $1 \leq i < i_1$ can be approximated by a common location value. We now define i_2 as the smallest integer such that

$$\min_{i_1 \leq i < i_2} b_i < a_{i_2}$$

and define the group G_2 by

$$G_2 = \{i : i_1 \leq i < i_2 \text{ and } b_i < a_{i_2}\}.$$

Again it is clear that $\cap_{i_1 \leq i < i_2} I_i \neq \emptyset$ so that all the samples $\boldsymbol{x}_{in_i}$ with $i_1 \leq i < i_2$ can be approximated by a common location value. We continue in this manner until all the samples have been treated. If the procedure results in g groups $G_i, i = 1, \ldots, g$, then it is clear from the construction that the samples can be approximated by g different values of the location parameter and that this is the minimum number required. The final division into groups gives then the minimum number of groups required, the defining samples for each group, the approximation intervals for each group and all possible allocations to the groups for the remaining samples. For each sample the set of other samples which cannot be approximated by the same parameter value is listed.

3.3 Linear combinations of location

Let m_i and s_i denote respectively the value of the location and the scale functional for the i-th sample. Given a set of k numbers $c_i, i = 1, \ldots, k$, the set of points

$$\sum_1^k c_i l_i$$

where each l_i ranges over I_i forms an interval of length

$$2\sum_1^k |c_i| s_i \tilde{q}(\alpha_k, n_i)/\sqrt{n_i} \tag{17}$$

with $\tilde{q}(\alpha, n)$ given by (7) and centred at

$$\sum_1^k c_i m_i.$$

If this interval is denoted by $I(c_1, \ldots, c_k)$ then the following holds on Gaussian test beds

$$\boldsymbol{P}\left(\sum_1^k c_i\mu_i \in I(c_1,\ldots,c_k) \quad \text{for all} \quad c_1,\ldots,c_k\right) \geq 0.95 \tag{18}$$

In other words the intervals are simultaneous 0.95-confidence intervals for the linear combination $\sum_1^k c_i\mu_i$ of the means μ_i. For large n_i the length of the interval is approximately

$$2z\left(0.975^{1/(2k)}\right)\sum_1^k |c_i|\sigma_i/\sqrt{n_i}$$

where $z(\alpha)$ is the α-quantile of the standard Gaussian distribution.

3.4 Further options

The software allows for other procedures. Apart from the default construction of approximation intervals given by (8) they can also be constructed as described in Section 2.3. In all cases a global scale functional can be used, that is the same scale is used for each sample. This is a case of "borrowing strength". The factor $c(n)$ of (2) can be defined by the user and the value 0.95 used in (14) can be altered. Finally if the samples are in some given order the minimum modality of a function required to describe the approximation values is calculated. By this we mean a function

$$f : [1, 2, \ldots, k] \to \mathbb{R} \text{ with } f(i) \in I_i \text{ for all } i. \tag{19}$$

A simple example is the question as to whether there exists a nondecreasing set of approximation values. This is the case if and only if it is possible to choose the function f in (19) to be monotone increasing.

3.5 Real data

We give an example of real data. It is taken from Rice (1988), page 397 and give the results of an inter-laboratory test. The left panel of Figure 1 shows the box-plots, the center panel shows the default approximation intervals whereas the right panel shows the approximation intervals base on the mean and standard deviation and under the standard assumption that the samples have the same variance. The default location and scale values are given in Table 3 as well as the end points of the corresponding approximation intervals.

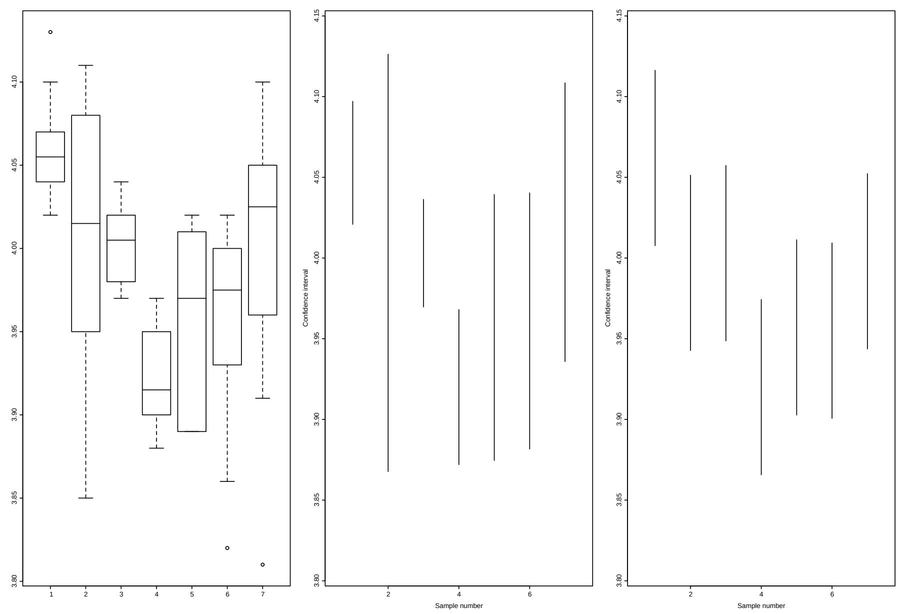

Figure 1: Box-plots and approximation intervals for the data from Rice (1988). Second panel shows intervals based on the default functionals with local scales, the third panel shows intervals based on the mean and a global standard deviation.

Two groups are required for the data. Group 1 consists of laboratory 4 and the set of allowable values of the location parameter is [3.872, 3.968]. Group 2 consists of the laboratories 1 and 3 and the set of acceptable location parameters is [4.021, 4.036]. The result of the procedure described in Section

3.2 is as follows. The locations of the laboratories 1 and 3 can be simultaneously approximated by any value in the interval [4.021, 4.036]. The remaining laboratories can be simultaneously approximated by any value in the interval [3.936, 3.968]. There are other possibilities. Thus laboratories 1,2,3, 5, 6 and 7 can be simultaneously approximated by any value in the interval [3.872, 3.968] and laboratory 4 by any value in the interval [3.872, 3.968].

Lab	location	scale	lower	upper
1	4.059	0.0300	4.021	4.097
2	3.997	0.1015	3.868	4.126
3	4.003	0.0262	3.970	4.036
4	3.920	0.0377	3.872	3.968
5	3.957	0.0647	3.875	4.039
6	3.961	0.0623	3.882	4.040
7	4.022	0.0565	3.936	4.109

Table 3: Default locations, scales and intervals for the data from Rice (1988).

We consider the difference in locations between Laboratory 1 with the largest location and Laboratory 4 with the smallest location values. Using the results of Section 3.3 with $c_1 = 1$, $c_4 = -1$ and the remaining c_i zero we obtain [0.053, 0.225] as the approximation interval for the difference in locations.

3.6 Software

The software is available on my web site

`http://www.stat-math.uni-essen.de/~davies/f1way.html`

3.7 Acknowledgments

I thank two anonymous referees for their comments on the first version of this article. One referee in particular gave a long and detailed criticism of the first version which lead to an almost complete revision of the article. The present version has profited considerably. The financial support of the Deutsche Forschungsgesellschaft (SFB 475) is gratefully acknowledged.

References

[Christensen, 1987] Christensen, R. (1987). *Plane Answers to Complex Questions*. Springer-Verlag, New York.

[Cohen, 1991] Cohen, M. (1991). The background of configural polysampling: a historical perspective. In Morgenthaler, S. and Tukey, J. W., editors, *Configural Polysampling: A Route to Practical Robustness*, chapter 2. Wiley, New York.

[Davies and Gather, 1993] Davies, L. and Gather, U. (1993). The identification of multiple outliers (with discussion). *Journal of the American Statistical Association*, 88:782–801.

[Davies and Kovac, 2001] Davies, P. L. and Kovac, A. (2001). Local extremes, runs, strings and multiresolution. *Annals of Statistics*, 29(1):1–65.

[Donoho, 1988] Donoho, D. L. (1988). One-sided inference about functionals of a density. *Annals of Statistics*, 16:1390–1420.

[Hampel, 1985] Hampel, F. R. (1985). The breakdown points of the mean combined with some rejection rules. *Technometrics*, 27:95–107.

[Miller, 1986] Miller, R. G. (1986). *Beyond ANOVA, Basics of Applied Statistics*. Wiley, New York.

[Morgenthaler and Tukey, 1991] Morgenthaler, S. and Tukey, J. W., editors (1991). *Configural Polysampling: A Route to Practical Robustness*. Wiley, New York.

[Rice, 1988] Rice, J. A. (1988). *Mathematical Statistics and Data Analysis*. Wadsworth and Brookes/Cole, Belmont, California.

[Tukey, 1949] Tukey, J. W. (1949). One degree of freedom for non-additivity. *Biometrics*, 5:232–242.

[Tukey, 1952a] Tukey, J. W. (1952a). Allowance for various types of error rates. Unpublished IMS address, Virginia Polytechnic Institute, Blacksburg, Virginia.

[Tukey, 1952b] Tukey, J. W. (1952b). The problem of multiple comparisons. Unpublished manuscript.

[Tukey, 1952c] Tukey, J. W. (1952c). Various methods from a unified point of view. Unpublished IMS address, Chicago, Illinois.

[Tukey, 1993] Tukey, J. W. (1993). Exploratory analysis of variance as providing examples of strategic choices. In S.Morgenthaler, E.Ronchetti, and W.A.Stahel, editors, *New Directions in Statistical Data Analysis and Robustness*, Basel. Birkhäuser.