ON MEDIAN TESTS FOR LINEAR HYPOTHESES

G. W. BROWN AND A. M. MOOD
THE RAND CORPORATION

1. Introduction and summary

This paper discusses the rationale of certain median tests developed by us at Iowa State College on a project supported by the Office of Naval Research. The basic idea on which the tests rest was first presented in connection with a problem in [1], and later we extended that idea to consider a variety of analysis of variance situations and linear regression problems in general. Many of the tests were presented in [2] but without much justification. And in fact the tests are not based on any solid foundation at all but merely on what appeared to us to be reasonable compromises with the various conflicting practical interests involved.

Throughout the paper we shall be mainly concerned with analysis of variance hypotheses. The methods could be discussed just as well in terms of the more general linear regression problem, but the general ideas and difficulties are more easily described in the simpler case. In fact a simple two by two table is sufficient to illustrate most of the fundamental problems.

After presenting a test devised by Friedman for row and column effects in a two way classification, some median tests for the same problem will be described. Then we shall examine a few other situations, in particular the question of testing for interactions, and it is here that some real troubles arise. The asymptotic character of the tests is described in the final section of the paper.

2. Friedman's test

In [3] Friedman presented the following test for column effects in a two way classification with one observation per cell. Letting r and s denote the number of rows and columns, the observations in each row are replaced by their ranks in the row; the rows then consist of permutations of the numbers 1, 2, . . . , s. Letting T_j be the column totals of these ranks, it is clear that the distribution of the T_j is independent of the form of the distribution of the observations provided the observations are continuously and identically distributed in rows. Thus under the null hypothesis of no column effects the quantity

$$\chi^2 = \frac{12}{rs(s+1)} \sum_{j=1}^{s} \left(T_j - \frac{r(s+1)}{2}\right)^2$$

is distribution free, and Friedman has shown that it has approximately the chi-square distribution with $s - 1$ degrees of freedom when r is large. Of course the hypothesis of no row effects may be tested by reversing the roles of rows and columns in the test criterion.

It is apparent that this test could be generalized immediately to test for row and column effects when there were several observations per cell; furthermore, differing numbers of observations per cell would raise no conceptual difficulties though the actual computation of significance levels might be troublesome. However a test for interaction between row and column effects seems to require some essential modification of the technique.

3. Median tests

Since the usual hypotheses tested with classified data relate to location parameters it is natural to attempt to set up tests in terms of medians. As a simple illustration we may consider a one way classification with several observations per cell (the number need not be the same for the various cells). Following Friedman's lead one might rank the complete set of observations and use as a test criterion the sum of squares of the deviations of the cell rank totals from their expected values. This would obviously be distribution free. Alternatively one could proceed as follows. Let m_i and n_i be the numbers of observations in the i-th cell which exceed and do not exceed the median of the complete set of observations. If there are k cells then these numbers form a two by k contingency table and one can test the null hypothesis by testing whether $E(m_i) = E(n_i)$ in that contingency table. It is readily shown that the m_i and n_i have the ordinary contingency table distribution (all marginal totals fixed).

The median test has three appealing properties from the practical standpoint. In the first place it is primarily sensitive to differences in location between cells and not to the shapes of the cell distributions. Thus if the observations of some cells were symmetrically distributed while in other cells they were positively skewed, the rank test would be inclined to reject the null hypothesis even though all population medians were the same. The median test would not be much affected by such differences in the shapes of the cell distributions. In the second place the computations associated with the median test are quite simple and the test itself is nothing more than the familiar contingency table test. In the third place when we come to consider more complex experiments it will be found that the median tests are not much affected by differing cell sizes.

However it must be pointed out that the median tests are not always completely satisfactory. For one thing they are not always distribution free, as we shall see in connection with the interaction test. Further, the well known fact that the median of a sum of two random variables is not generally the sum of their medians is a distinct inconvenience in dealing with hypotheses where additivity is a natural *a priori* assumption. Thus if x and y have zero medians and distributions $F(x)$ and $G(y)$, then it is easily shown that

$$.25 \leqq P(x + y < 0) \leqq .75$$

and these limits cannot be improved even if $F = G$, as the following density function shows.

$$\begin{aligned} f(x) &= \frac{1}{2p}, & -p < x < 0 \\ &= \frac{1}{2}, & 0 < x < 1\,. \end{aligned}$$

By making p sufficiently large or small one can make $P(x + y < 0)$ come arbitrarily near either of these limits.

4. Tests for row and column effects

In an r by s table with one observation per cell the column effects may be tested by first finding the row medians, then counting the number of observations, say m_j, in the j-th column which exceed their respective row medians. The m_j will obviously be distributed independently of the population distributions if the observations are continuously and identically distributed within rows. In fact, as is shown in [2, p. 398], the numbers m_i and $r - m_i$ form a two by s table which has very nearly the ordinary contingency table distribution.

When one considers the case of several observations per cell in a two way classification the above test is still valid if it is assumed that the interactions are zero. And it is plain that differing cell sizes would introduce no essential difficulty in this case. However when interactions must be taken into account the test must be changed. In this case there are three tests that might be of interest:

(a) Test of column effects against interaction,
(b) Test of column effects against error,
(c) Joint test of column and interaction effects.

The second of these has not been investigated but the first and third offer no difficulty if the cell sizes are equal.

To test column effects against interaction effects one needs merely to find the cell medians and apply the test described in the first paragraph of this section to those medians. It is assumed in this test that the observations are of the form

$$x_{ijk} = \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

where $k = 1, 2, \ldots, t$, the cell size. Since we are here considering a test of main effects against interactions it is natural to suppose that all the quantities α, β, γ, ϵ are random variables, say with distributions $F_1(\alpha)$, $F_2(\beta)$, $F_3(\gamma)$, $F_4(\epsilon)$ respectively. The null hypothesis supposes that $F_2(\beta)$ is concentrated at zero and we may assume that γ and ϵ have zero medians. Letting $G_t(\tilde{\epsilon})$ represent the distribution of medians of samples of size t from F_4, it is apparent that, under the null hypothesis, the cell medians in the i-th row constitute a random sample of sample of size s from a population distributed by the convolution of F_3 and G_t translated by an amount α_1. Thus the situation is precisely equivalent to the case of one observation per cell.

It is also apparent that the distribution free character of the test will not be spoiled if the cell sizes differ only between rows for it is required merely that the cell medians in a given row be identically distributed. If the cell sizes differ within rows the test is no longer distribution free, but even in this case one would not often go far wrong in behaving as though it were. Certainly if one has to contend only with a missing observation here and there he need have no qualms about using the test. Furthermore, if F_3 has a large "variance" so to speak relative to that of F_4, the convolutions for the different cells will all be essentially F_3 even for widely different cell sizes. Hence if the minimum cell size is large the test will be nearly distribution free in any case.

Turning now to the joint test of column and interaction effects, the natural extension of the foregoing tests consists of first finding the row medians and counting the number m_{ij} of observations in the i, j-th cell which exceed the i-th row median; then under the null hypothesis (F_2 and F_3 both concentrated at zero) a test criterion which is a function of the m_{ij} will be distribution free. Of course one expects the m_{ij} to be half the cell size on the average and in fact the ordinary chi-square criterion for testing independence in a 2 by r by s table may ordinarily be used [2] since it has approximately the chi-square distribution with $r(s - 1)$ degrees of freedom. Here differing cell sizes do not destroy the distribution free character of the test since all observations are identically distributed in any case (except for row wise translations) under the null hypothesis.

5. Tests for interactions

To simplify the discussion let us consider a two by two table with observations identically distributed except for cellwise location. The cell population medians will be represented by:

$\alpha + \beta + \gamma$	$\alpha - \beta - \gamma$
$-\alpha + \beta - \gamma$	$-\alpha - \beta + \gamma$

where α is the row effect, β the column effect, and γ the interaction effect; the fact that their sum is zero introduces no essential simplification. The quantities α, β, and γ may be regarded as fixed numbers or as random variables. Also, assuming for the moment that the cell sizes are equal, the observations will be denoted by:

$u_1, u_2, \ldots, u_t$	$v_1, v_2, \ldots, v_t$
$x_1, x_2, \ldots, x_t$	$y_1, y_2, \ldots, y_t$

and they are assumed to be indexed in order of magnitude within cells.

In order to test whether the interaction is zero it is first necessary to remove the row and column effects somehow. One obvious method for doing this is to add corresponding elements in the diagonal cells. Thus under the null hypothesis, $u_1 + y_1$ is distributed the same as $v_1 + x_1$, $u_2 + y_2$ the same as $v_2 + x_2$, and so on, whatever may be the row and column effects. The interaction may therefore be tested by applying the sign test to these t pairs of sums. This test is entirely distribution free but suffers certain disadvantages. In the first place it seems incapable of a satisfactory generalization to the case of unequal cell sizes. In the second place one must suspect that a more powerful test might exist because the addition of observations has automatically increased the error "variance."

The two difficulties mentioned above can be overcome by a median test to be described now, but the test is unhappily not distribution free. However examination of a few special cases indicates that the test is rather insensitive to the form of the population distribution so that it seems to merit some investigation. Certainly for large samples it is a satisfactory test.

Referring to the two by two table of observations above, there exist three numbers a, a', and b such that if the observations are translated by minus a in the first row, by minus a' in the second row, minus b in the first column, and plus b in the second column, then the transformed observations will have zero medians in both rows and columns. Thus $(a - a')/2$ is a point estimate of α, and b is an estimate of β. Letting m_{ij} $(i, j = 1, 2)$ be the numbers of positive residuals in the four cells, the test of zero interaction may be made by using a criterion such as the chi-square criterion for a two by two contingency table with marginal totals fixed.

For a long time we supposed that the distribution of the m_{ij} was independent of the population distribution, but after several unsuccessful attempts to demonstrate this proposition we finally examined two simple cases to find that it was not true. If one analyzes the various possibilities for a two by two table with two observations per cell he finds that the m_{ij} will be

2	0
0	2

or

0	2
2	0

when

$$u_1 + y_1 > x_2 + v_2 \quad \text{or} \quad x_1 + v_1 > u_2 + y_2 \quad \text{respectively}.$$

Otherwise all m_{ij} will be equal to one. The probability that one or the other of these inequalities will hold is found to be

$$\frac{11}{63} \cong .175 \text{ for the uniform density},$$

$$\frac{5}{27} \cong .185 \text{ for the density } e^{-x},\ x > 0 .$$

This probability has also been investigated empirically for the normal distribution by Bernice Brown. For 1000 tables constructed from random normal deviates she found one or the other of the above inequalities to hold in 172 cases.

We may observe that this test is much more sensitive than the one described earlier using the sign test. Here the significance level for two observations per cell is somewhat less than 0.2 while with the sign test the level would be 0.5.

This interaction test generalizes immediately to the case of several rows and columns. One determines the m_{ij} by finding joint row and column medians which, when subtracted from the observations, make them have zero medians in both rows and columns. Unfortunately we have not found any simple method for determining these medians and can only offer an iterative procedure. This consists merely of finding the row medians and subtracting them from the observations, then finding the column medians of the residuals and subtracting them out. This procedure is repeated until the signs of the residuals are balanced in both rows and columns. Sometimes this method does not converge in a finite number of steps but one can readily determine the proper values after a few steps have been carried through because the crucial observations are identified by the fact that their signs alternate from step to step.

6. General method

The interaction test points the way to the general method of making median tests in more complex situations. The procedure is to fit those parameters which are free to vary under the null hypothesis by their median estimates and remove their effects by subtracting out these medians. Then one tests whether the signs of the residuals are split fifty-fifty when the residuals are classified by the parameters being tested. This scheme has been tried on a great many sets of data from more complex experiments and with good results. That is to say, the probability levels for these median tests usually differed but little from the levels given by the ordinary analysis of variance tests.

The linear regression problem, for example, may be dealt with as follows. Let y be distributed with median

$$a_0 + \sum_1^k a_r z_r$$

and suppose we have a sample of n observations

$$y_i, z_{1i}, z_{2i}, \ldots, z_{ki} \quad \text{with} \quad i = 1, 2, \ldots, n .$$

The point estimates of the coefficients a_r may be defined as those numbers $\tilde{a}_r$ such that

$$\underset{z_{ri} \leqq \tilde{z}_r}{\text{median}} \left(y_i - \tilde{a}_0 - \sum \tilde{a}_r z_{ri} \right) = \underset{z_{ri} > \tilde{z}_r}{\text{median}} \left(y_i - \tilde{a}_0 - \sum \tilde{a}_r z_{ri} \right) = 0 ,$$

$$r = 1, 2, \ldots, k ,$$

where $\tilde{z}_r$ is the median of the n observations z_{ri}. Thus the whole set of observations is divided into two groups on the median of z_1 and the $\tilde{a}$'s chosen so that the deviations have zero medians in both groups. Simultaneously the two groups formed by dividing the observations on $\tilde{z}_2$ must have zero medians, etc. Actually there are only $k + 1$ conditions here, k which require the medians to be the same in a given pair of groups, and one which requires the over all median to be zero.

To test whether the last $(k - m)$ a's have values a_{s0} $(s = m + 1, \ldots, k)$ one would fit the other coefficients by m relations like those above except that a_s would be replaced by a_{s0}. Then the resulting deviations divided into 2^{m-k} groups on the $\tilde{z}_s$ should have their signs about evenly divided in each group under the null hypothesis. The 2^{m-k+1} numbers obtained by counting the positive and negative deviations in each group form a contingency table with all marginal totals fixed and may be tested by the ordinary chi-square criterion with $m - k$ degrees of freedom when n is large.

7. Asymptotic character of the tests

When the cell sizes are large all the contingency tables arising in these median tests may be tested by the ordinary chi-square criterion for such tables. The proof of this statement is so similar for all the various cases that we shall discuss only one example, namely, the two by two table exhibited in the fifth section. Here, however, we may suppose that the cell sizes are different, say t_{ij} $(i, j = 1, 2)$, without encountering any difficulty.

When all the rows and columns have even numbers of observations or when any pair of the four have even numbers of observations, the numbers a, a', and b are uniquely determined in terms of four of the observations (except in degenerate cases) provided the usual definition of the median as the average of the middle pair is employed. When all the marginal sums are odd a similar definition is needed in order to get a unique set (a, a', b). Suppose, for example, we have found a set such that

$$u_{r+1} - a - b = 0\,, \quad y_{R+1} - a' + b = 0\,, \quad R = \tfrac{1}{2}(t_{22} - t_{11}) + r$$

and the signs are balanced in rows and columns. It is clear then that

$$v_{T-r} < a - b < v_{T-r+1}\,, \qquad x_{T'-r} < a' + b < x_{T'_1-r+1}$$

where

$$T = \tfrac{1}{2}\,(t_{11} + t_{12} - 1)\,, \qquad T' = \tfrac{1}{2}\,(t_{11} + t_{21} - 1)\,.$$

It follows that (a, a', b) can be varied subject to $u_{r+1} + y_{R+1} = a + a'$ and subject to these inequalities. One could make the set unique, for example, by putting $a - b$ equal to $\frac{1}{2}(v_{T-r} + v_{T-r+1})$ if the $a - b$ interval were shorter than the $a' + b$ interval. Or one could simply use one of the end points of the $a - b$ interval. In any case the m_{ij} of the resulting contingency table are uniquely determined.

Now let us write down the joint density function for (r, a, a', b) for the case just discussed, defining

$$a - b = v_{T-r}\,.$$

Letting $F(u)$ represent the distribution of u and using the notation given at the beginning of section 5 for the population medians, the desired density is the product of the four quantities

$$(r+1)\binom{t_{11}}{r+1} F^r(u_{r+1})\,[1 - F(u_{r+1})]^{t_{11}-r-1}dF(u_{r+1})\,,$$

$$(T-r)\binom{t_{12}}{T-r} F^{T-r-1}(v_{T-r} + 2\beta + 2\gamma)\,[1 - F(v_{T-r} + 2\beta + 2\gamma)]^{t_{12}-T+r} \times dF(v_{T-r} + 2\beta + 2\gamma)\,,$$

$$(R+1)\binom{t_{22}}{R+1} F^R(y_{R+1} + 2\alpha + 2\beta)\,[1 - F(y_{R+1} + 2\alpha + 2\beta)]^{t_{22}-R-1} \times dF(y_{R+1} + 2\alpha + 2\beta)\,,$$

$$\binom{t_{21}}{T'-r} F^{T'-r}(x_{T'-r} + 2\alpha + 2\gamma)\,[1 - F(x_{T'-r} + 2\alpha + 2\gamma)]^{t_{21}-T'+r}.$$

Using the technique and the assumptions ordinarily employed (see, for example, Wilks [4, p. 91]) to obtain the asymptotic distribution of a sample median, one finds that (r, a, a', b) are jointly normally distributed as the t_{ij} become infinite with fixed ratios. On integrating out (a, a', b) it is found that r is normally distributed with just the same mean and variance as z in the contingency table

z	$\frac{1}{2}(t_{11} + t_{12}) - z$
$\frac{1}{2}(t_{11} + t_{21}) - z$	$\frac{1}{2}(t_{22} - t_{11}) + z$

with fixed marginal totals.

REFERENCES

[1] G. W. BROWN and A. M. MOOD, "Homogeneity of several samples," *American Statistician*, Vol. 2 (1948), p. 22.

[2] A. M. MOOD, *Introduction to the Theory of Statistics*, McGraw-Hill, New York, 1950.

[3] M. FRIEDMAN, "Use of ranks to avoid the assumption of normality in analysis of variance," *Jour. Amer. Stat. Assoc.*, Vol. 32 (1937), pp. 675–701.

[4] S. S. WILKS, *Mathematical Statistics*, Princeton University Press, Princeton, 1943.