Exact significance levels for Hewitt's test for seasonality

S. D. WALTER

From the Department of Epidemiology and Public Health, Yale University, New Haven, USA

SUMMARY   Hewitt proposed an easily computed test for seasonality in ranked monthly data. This paper provides a tabulation of the exact distribution of the test statistic, which previously could be estimated only from simulated data. The use of this test when ties may occur in the data is also discussed.

Several recent papers have compared the power of various methods to detect seasonal variation in epidemiological data. St. Leger[1] contrasted a likelihood ratio test with Edwards's[2] well-known geometric test, for sinusoidal cyclic fluctuations, using a computer simulation. Walter[3] obtained an asymptotic expression for the power of Edwards's test, giving tables of typical values. Freedman,[4] using simulated data, compared the power of Edwards's test, a new Kolmogorov-Smirnov type test, Hewitt's[5] test, and the usual χ^2 test for any variation, not necessarily seasonal, among the 12 months. Of these, Hewitt's test is the easiest to compute.

For all these methods, except Hewitt's, the distribution of the appropriate test statistic is known, at least approximately, under the null hypothesis of constant incidence. In order to determine the significance of the seasonal effect using Hewitt's test, however, one must consult an empirical distribution established from simulated data[5]; the significance levels thus obtained are therefore only approximate and subject to sampling variation. This implies some uncertainty in the precise conclusions from Hewitt's test; a fair comparison of its power with that of alternative tests also relies on knowing the true significance levels. For these reasons it was decided to establish the exact probabilities of Hewitt's statistic in the tail of its distribution associated with the usually adopted significance levels.

Method

Data to be analysed by Hewitt's method are expressed in terms of 12 monthly ranks for the variable of interest (for example, the frequency or rate of a disease), with rank 12 assigned to the month with highest incidence, 11 to the next highest, and so on down to rank 1 for the month with the lowest incidence. The test statistic T is the maximum rank-sum for any 6 consecutive months; this corresponds to testing for a seasonal fluctuation at any time during the year. (Hewitt *et al* also discuss the use of this statistic when a 6-month period for the rank sum is specified *a priori;* this approach would be used much less often in practice). To determine the significance of a particular value of T for a given data set, we thus require the probabilities $P(T = t)$ for various values of t.

Let d_t be a *decomposition* of t into 6 distinct integral parts corresponding to any 6 monthly ranks; for example, T could assume the value 54 by some arrangement of the 6 consecutive monthly ranks 12, 11, 10, 9, 7, and 5, which is therefore one possible decomposition of 54. Now by basic probability we have

$$P(T = t) = \sum_i P(T;d_{ti}) - \sum\sum_{i \neq j} P(T;d_{ti},d_{tj})$$
$$= \sum_i P(T;d_{ti} \mid d_{ti})\, P(d_{ti})$$
$$- \sum\sum_{i \neq j} P(T;d_{ti},\ d_{tj} \mid d_{ti},d_{tj})\, P(d_{ti},d_{tj}) \quad (1)$$

where $P(T;d_{ti})$ denotes the probability that T has the (unique) decomposition d_{ti}, and $P(d_{ti})$ is the probability that d_{ti} occurs in the data. The second terms in (1) correspond to cases where T has two possible decompositions, d_{ti} and d_{tj}, that is, there are two 6-month periods of the year with equal rank sums. (It is not possible to have three such decompositions for the values of T to be considered). The summations of (1) are over all possible decompositions of t. The conditional probabilities reflect that although a particular decomposition of t may occur in a certain 6-month period of the data, this

is not necessarily the period for which the rank sum is maximised. Noting that, under the null hypothesis of no seasonality, all sets of 6 ranks are equally likely to occur in a particular 6-month period, we have $P(d_{ti}) = 12/{}_{12}C_6 = 1/77$ for all values of t and i; ${}_{12}C_6 = 924$ is the number of ways of selecting any 6 ranks from 12.

Although somewhat tedious, an enumeration of the distribution of T using (1) is considerably easier than considering all the 12! possible rank arrangements. A few typical calculations are described below, and the complete set is available on request.

The largest possible value of T is 57, for which there is only one decomposition (12,11,10,9,8,7). If this decomposition occurs in the data, it must constitute the maximum, and thus $P(T = 57) = 1/77$. For T = 56 there is also only one decomposition, namely (12,11,10,9,8,6), but this is not the maximum in all cases. Firstly, if rank 6 occurs at either end of the 6-month period containing the decomposition (for example, in an arrangement such as 8,11,10,9,12,6), then this period will not contain the maximum T if rank 7 occurs in the month adjacent to the other end of the period. (The maximum in this case occurs in the period starting one month earlier or later). Hence the probability is 5/6 that such a permutation of this decomposition, if it occurs, is the maximum. Secondly, if the rank 6 is not at either end of the period, this decomposition must be the maximum. There are 2 × 5! permutations of the first type, and 4 × 5! of the second. Thus from (1)

$$P(T = 56) = [(5/6) \times 2 \times 5! + 1 \times 4 \times 5!]/(6! \times 77) = 17/1386$$

Similar, although more complex, working applies for the other values of T. There are respectively 2, 3 and 5 distinct decompositions of T = 55, 54, and 53, which is as far as it is necessary to enumerate the distribution if the conventional significance levels are used. For each permutation of each decomposition, one must consider the possible adjacent ranks (if any), and their associated probabilities, which would cause the decomposition not to be the maximum. For example, the permutation (9,4,12,11,10,7) of a decomposition of 53 is not the maximum if either rank 8 is adjacent to rank 9 or if ranks 6 and 8 (in either order) are adjacent to 7 (because $6 + 8 > 9 + 4$). There is thus a 4/5 probability that such an arrangement is the maximum. An example of T having two decompositions is provided by the sequence (7,6,12,11,10,9,5,8), which contains two maximal rank sums equal to 55.

The Table shows the exact distribution of T thus obtained, and the simulated distribution obtained by Hewitt. The discrepancies between the two are modest, and within the expected range of sampling variation. One convenient finding is that T = 55 is

Table   *Exact and simulated distributions for Hewitt's test statistic T*

T	Exact probability	Simulated probability*	Exact cumulative probability	Simulated cumulative probability*
57	1/77 = 0·0130	0·0134	0·0130	0·0134
56	17/1 386 = 0·0123	0·0114	0·0253	0·0248
55	19/825 = 0·0230	0·0216	0·0483	0·0464
54	223/6 930 = 0·0322	0·0302	0·0805	0·0766
53	571/11 550 = 0·0494	0·0494	0·1299	0·1260

* Estimated from 5000 Monte Carlo samples.

significant with a probability closer to the usual 5% level than is indicated by Hewitt's data.

In Freedman's[4] comparison of power for Hewitt's and various other tests, data were simulated to contain a sinusoidal seasonal effect, and nominal 10%, 5%, and 1% significance levels were used; for Hewitt's test the critical values of T were taken as 53, 55, and 57 respectively, these having type I error rates closest to the nominal levels. The exact probabilities indicate that T = 53 and T = 57 have type I error rates somewhat higher than their nominal values, implying that Edwards's test compares less favourably with the other tests. On the other hand, T = 55 is only slightly conservative compared to its nominal 5% significance level. In total, these results suggest that the power of Kuiper's[6] Kolmogorov-Smirnov type test is approximately the same as that of Edwards's test, at least for the sinusoidal variation used by Freedman.

EFFECT OF TIES

Hewitt's test as proposed, and as discussed here, makes no provision for the possibility of two or more months having equal (or tied) incidence. Although ties are relatively unlikely if the variable of interest is a rate, this is not so if raw monthly frequencies are being analysed. It has been pointed out to me by a student, Dr. O. Marrero, that in fact quite large samples are needed in this case to avoid ties completely. In particular, there *must* be a tie if the total sample size is less than $0 + 1 + 2 + .. + 11 = 66$, this being the smallest set of 12 distinct monthly frequencies. In those practical problems requiring statistical analysis, the seasonal fluctuation is usually quite small (typically less than 20%), and then the chance of a tie is not negligible, even with samples of several hundred. This problem does not appear to have been recognised previously in the literature, which is surprising when we recall that Hewitt's test was proposed originally as an alternative to Edwards's test for small samples.

One solution to this problem is to adjust the raw frequencies according to the number of days in each month (that is, rank 'cases per day' by month). This is

good general practice anyway, because it precludes the possibility of a spurious seasonal 'calendar effect'. If this is done, then to prevent ties one must have distinct frequencies in the 4 months with 30 days and also in the 7 months with 31 days; this implies a minimum total sample of 27, for all practical purposes, to avoid ties, but once again ties will occur frequently in practice with all but quite large samples. If ties do occur, one might modify the test statistic by assigning average ranks to tied months (for example, rank 11·5 to two months tied with the highest incidence), and interpolate the probability level of T if necessary. Although this manoeuvre is used with other non-parametric methods, the precise distributional implications of doing so in Hewitt's test are not known.

Reprints from Dr. S. D. Walter, Department of Epidemiology and Public Health, Yale University School of Medicine, 60 College Street, New Haven, Connecticut 06510, USA.

References

[1] St. Leger AS. Comparison of two tests for seasonality in epidemiological data. *Appl Stat* 1976; **25:** 280–286.

[2] Edwards JH. The recognition and estimation of cyclic trends. *Ann Hum Genet* 1961; **25:** 83–6.

[3] Walter SD. The power of a test for seasonality. *Br J Prev Soc Med* 1977; **31:** 137–40.

[4] Freedman LS. The use of a Kolmogorov-Smirnov type statistic in testing hypotheses about seasonal variation. *J Epidemiol Community Health* 1979; **33:** 223–8.

[5] Hewitt D, Miller J, Csima A, Pakula A. On Edwards's criterion of seasonality and a non-parametric alternative. *Br J Prev Soc Med* 1971; **25:** 174–6.

[6] Kuiper NH. Tests concerning random points on a circle. *Proceedings, Koninklijke Nederlandse Akademie van Wetenschappen (Series A)* 1962; **63:** 38–47.