Article NIMHANS Journal

A Comparison of Statistical Tests for Seasonality

Volume: 05 Issue: 01 January 1987 Page: 53-58

Sanjeev Sarmukaddam ,

Reprints request
&,
- Society for Education, Action and Research in Community Health, P.O. District Gadchiroli, Maharashtra-442 605, India.
Shivaji Rao, *- Department of Biostatistics, National Institute of Mental Health & Neuro Sciences, Bangalore 560 029, India*

Abstract

Seasonal variations in the incidence of a disease is one of the important issues in epidemiological studies. Use of specific statistical tests to detect such variation is widespread. Since powers or efficiencies of most of these tests are not known, criterion efficiencies of twelve original or modified tests have been calculated, keeping the most sophisticated and powerful Walter-Elwood test as the criterion. Suggestions are made regarding choice of a test for small, moderate and large samples.

Key words -
Seasonal variations,
Statistical tests,
Epidemiological studies,
Criterion efficiencies

In epidemiological investigations it is required to look for seasonal or cyclic trends in incidence rates. A sample of individuals is taken and categorized, where the categories have a cyclic order. In most cases categories are periods, e. g., months of a year, in which the individual experienced something specific such as death, birth or entered hospital. Our null hypothesis is that the individuals are equally likely to be allocated to each of the categories. In contrast, the alternative hypothesis is that the frequencies in the categories have some sinusoidal variation over their cyclic order. Usually this alternative arises from a seasonal trend in the frequency of a disease under consideration.

Statistical Methods.

To detect such variation in twelve monthly frequencies from a single year or several aggregated (pooled) years, one can use a general chi-square goodness-of-fit test. Several specific tests [1], [2], [3], [4], [5], [6], [7] also have been proposed by various authors for the detection of the seasonal variation. For illustration, consider the data presented in Table 1, pertaining to female mania admissions in the National Institute of Mental Health & Neuro Sciences, Bangalore during 1979-83.

Table 1 - Monthly female admissions with mania and total female admissions, 1979 - 83

Table 1 - Monthly female admissions with mania and total female admissions, 1979 - 83

X^2 11=19.64, NS, Power=74%

In testing the uniformity of monthly female mania admissions over the year, with expected frequency for each month equals 752/12=62.67, the chi-square value 19.64 with 11 degrees of freedom (df) is not significant at 5% level. While calculating the expected frequencies one can take into consideration the variation in total number of monthly admissions and unequal length of months. But after such adjustment also the chi-square value is not significant at 5% level.

Exact power attained by chi-square test in any analysis can be calculated with the help of Cohen's [8] power tables. For the above analysis the power is 74%, which means that, even if seasonality is present in the above data, chances of its detection by chi-square test are only 74%. Another draw back of chi-square in testing seasonality is that, a chi-square value could be significant even if there is no seasonality, for instance, when frequencies are more in alternate months.

This clearly shows that one must look for more specific tests. One such earliest test is given by Edwards [1]. The method requires arranging twelve monthly frequencies as a set of weights on the rim of a circle representing the year. The data by weights ni (where ni is the i-th month frequency) are placed around a unit circle at points corresponding to the sector mid points at angles θ to an arbitrary diameter, e.g., OA the diameter through 1st January (Fig.1). Angles $\theta i=2\pi(i-½)$ 12 are in radians; to convert them into degrees multiply by $180/\pi=57.3$ degrees.

.Typical components of moments about orthogonal axis OA and OB - $\theta i=2\pi(i-½)/12$***, i=1....12 - (Mean of)*** $x=\Sigma((ni)^{½}\cos\theta i)/12$ ***- (Mean of)*** $y=\Sigma((ni)^{½}\sin\theta i)/12$

In the absence of any cyclic trend the expected centre of gravity of these masses will be at the origin of the circle, and will have a determinate sampling distribution. The actual location of the centre of gravity of these points is ((Mean of) x, (Mean of) y) where (Mean of) $x=\Sigma(\sqrt{ni}\cos\theta i)/12$ and (Mean of) $y=\Sigma(\sqrt{ni}\sin\theta i)/12$. The test compares the observed centre of gravity of the data with its expected position, the origin of the circle.

If we define $S=\Sigma\sqrt{ni}\sin\theta i$, $C=\Sigma\sqrt{ni}\cos\theta i$ and $W=\Sigma\sqrt{ni}$ then the distance of observed centre of gravity from the origin is $d=((S^2+C^2)/W)^{1/2}$ and a measure of the amplitude of the cyclic variation a =4d. On the null hypothesis, test statistic $X^2=½a^2N$ (where $N=\Sigma^{12}$ i=1 ni) is approximately distributed as chi-square with 2 df. Angle corresponding to the peak is given by $\theta^*=\tan^{-1}(S/C)$. For the present data set a=0/147, $\theta^*=169$ which corresponds with late June and $X^2=8.17$ (with 2 df) is significant at 5% level. Walter [9] has given power tables for this test. For the present data set the power of Edwards' test is approximately 88%.

Walter & Elwood [2] have extended this method somewhat more rigorously for variable population at risk and unequal (exact) length of months. Their methods is discussed at length and well illustrated by Walter [10]. For the present data set the methods yields $X^2=9.19$ (with 2 df) which is significant at 5% level. Other estimated parameters are a=0.202 and $\theta^*=186$ which corresponds with early July.

With these estimated parameters one can easily fit a Simple Harmonic Model to the data. For the present data set, the fit is good (X^2=12.99 with 9 df, not significant, Fig.2). Walter-Elwood test results (X^2 (2) for testing seasonality, estimates of a and θ *, X^2 (9) for testing goodness-of-fit of Simple Harmonic Model) for few other interesting data sets are presented in Table 2.

.Graph showing monthly female mania admissions - observed and expected

Table 2 - Results of Walter-Elwood Test

Table 2 - Results of Walter-Elwood Test

Roger [3] has shown that $R=2[(\Sigma\, ni \text{ Sin } \theta i)^2 + (\Sigma\, ni \text{ Cos } \theta i)^2]/N$ has chi-square distribution with 2 df even for small samples. R is nearly equivalent to Edward's test statistic for large samples, but for small samples the approximation is better for the distribution of R as chi-square with 2 df than that of Edward's' test statistic. For the present data set R=9.16 which is significant at 5% level.

Hewitt et al [4] suggested a non-parametric test based on the maximal rank sum of the rates for any six month's segment of the year. For the present data set the test statistic comes form the segment May to October and is equal to 53 which is not significant at 5% level. David & Newell [5] developed a similar method based on the maximum difference in the numbers of cases in two arbitrary halves of the year, but do not allow for the inclusion of a denominator frequency. For the present data set their test statistic come from May to October versus November to April which is again not significant at 5% level.

Freedman [6] suggested a Kolmogorov-Smirnov type statistic where one can adjust for unequal length of months and variation in monthwise total admissions. For the present data set his test statistic takes the values 1.42 and 1.93 for two types of adjustments which are significant at 5% and 1% levels respectively.

Pocock [7] suggested to refer the index of dispersion to the chi-square distribution to test the hypothesis that the variation from month to month is purely random. His test statistic

I=Sample variance /sample mean × (K-l)

is distributed as chi-square with K-l df where K is the number of months in the period under consideration. For the present data set I=82.25 with 59 df is significant at 5% level.

Except Edwards test and approximate calculations of Walter-Elwood test, the powers or efficiencies of any of these tests are not known. To examine their relative performance all these tests have been applied to twenty independent data set of varying sizes (118-3535), and their criterion efficiencies have been calculated, keeping Walter-Elwood test as the criterion. Approximate power of Walter-Elwood test for all the data sets is between 95% to 100%.

Table 3 - Criterion efficiency (%) at 5% significance level - (Criterion: Walter-Elwood Test with exact θ i's)

Table 3 - Criterion efficiency (%) at 5% significance level - (Criterion: Walter-Elwood Test with exact θ i's)

Conclusion

Though the use of quick and easy methods for routine and initial analysis is attractive, rapidity and easiness of application should not be the main or the only criterion for the first choice of a statistical test. On the basis of the above empirical evidence and considering other theoretical aspects of these tests, we recommend the use of Roger's test for small samples and Walter-Elwood's or Edward's test for moderate or large samples. These methods also yields estimates of two important parameters namely, the angle corresponding to the peak and a measure of the amplitude of the cyclic variation. With these parameters one can easily fit a simple harmonic model to the data.

Acknowledgements

The authors are involved in a large study on 'Climatic variables and seasonal variations in mental disorders' along with Dr. K S Mohan and Mr. D M Joseph under the guidance of Dr. P S Gopinath. Data used for the illustration in this paper is taken from the above study. The authors are grateful to all of them.

1.Edwards J H, The recognition and estimation of cyclic trends
Annals of Human Genetics Page: 25: 83-87, 1961
2.Walter S D & Elwood J M, A test for seasonality of events with a variable population at risk
British Journal of Preventive & Social Medicine Page: 29: 18-21, 1975
3.Roger J H, A significance test for cyclic trends in incidence data
Biometrika Page: 64: 152-155, 1977
4.Hewitt D, Milner J, Csima A & Pakula A, On Edward's criterion of seasonality and a non-parametric alternative
British Journal of Preventive & Social Medicine Page: 25: 174-176, 1971
5.David H A & Newell D J, The identification of annual peak periods for a disease
Biometrics Page: 21: 645-650, 1965
6.Freedman L S, The use of a Kolmogorov-Smirnov type statistic in testing hypothesis about seasonal variation
J. Epi. Com. Health Page: 33: 223-228, 1979
7.Pocock S J, Harmonic analysis applied to seasonal variation in sickness absence
Appl. Stat Page: 23: 103-120, 1974
8.Cohen J, *Statistical Power Analysis for the Behavioral Sciences. Academic Press, New York*
Page: 216 & 239-240, 1977
9.Walter S D, The power of test for seasonality
British Journal of Preventive & Social Medicine Page: 31: 137-140, 1977
10.Walter S D, Seasonality of mania, a reappraisal
British Journal of Psychiatry Page: 131: 345-350, 1977