A NOTE ON VICTORIAN WHEAT YIELD DISTRIBUTIONS

TIMOTHY J. RYAN*

Victorian Department of Agriculture

An analysis of wheat yield series from ten localities in Victoria was undertaken to investigate for independence of observations and to test for normality of the distributions. The conclusion was reached that Victorian wheat yield series are composed of random observations and form frequency distributions which display slight negative skewness. The situation may be different for other crops or for wheat in other areas.

Introduction

Simulation modellers commonly obtain values for a stochastic variable by using a pseudo-random number generator to sample from some specified density function. The use of a pseudo-random generator is not valid unless the modellers have demonstrated that the variables are independent between successive values. If the underlying density function of the variable can be approximated by some known function, such as the normal density function, then the computational burden is lessened.

Bostwick [1] investigated dryland crop yield series in several regions in the United States and concluded that the series were not composed of independent values. Bostwick's analysis, however, was criticized by Shaw [7] on the grounds that the Wallis-Moore randomness test [8] was not correctly conducted. A more complete study by Day [2] on Mississippi crop yields concluded that annual yield was independent of the previous year's level. Day also investigated the distributions of the Mississippi crop yields and concluded that they were not normally distributed. In general the crops were positively skewed.

Day stated quite clearly [2, p. 720] that his results applied to the Mississippi Delta Region and should not be generalized. Phillips [6, p. 45], however, supports the inference that crop yields are positively skewed.

The purpose of this paper is to examine wheat yield series from ten sites in Victoria. The yield series are samples obtained on individual farms from the main wheat regions. The first hypothesis was that the yield series from each site was composed of random observations. The second hypothesis was that the yield distributions from each site could be considered as samples obtained from an underlying normal density function.

Data and Methods

Data

The Victorian Department of Agriculture has conducted experiments on the same properties (or adjacent properties) with the same wheat variety in different regions of the State for periods up to 29 years.

* Agricultural Economics Officer, Victorian Department of Agriculture. With the usual *caveat* thanks are extended to Mr R. Jardine for helpful comments and advice on the analysis.

Results from the principal wheat areas, the Mallee, the Wimmera and the East-Northern regions, were chosen for analysis. Table 1 contains

TABLE 1

Trial Sites and Variety Details

Region	Zone[a]	Variety	Site	Years[b]	Median bu/ac	Range
Mallee	2	Insignia	Manangatang	22	32·7	9·6–42·5
			Walpeup	22	28·5	13·7–44·2
Wimmera	4	Pinnacle	Dooen	29	50·2	0–76·2
			Donald	26	39·0	0–60·3
			Kaniva	29	46·4	0–70·5
			Salisbury	29	40·5	0–51·4
			Warracknabeal	26	32·3	0–52·7
East-Northern	6	Sherpa	Dookie	29	30·3	0–54·8
			Rutherglen	22	39·5	0–68·9
			Yarrawonga	18	30·1	0–54·2

[a] Victoria is divided into zones, see [5]. [b] Up to and including 1971.
Source: Wheat Variety Trials, Agronomy Branch, Victorian Department of Agriculture.

details of the experimental locations and of the varieties included. The experiments are situated on soil types which are representative of the district soils, and the plots are positioned each year on the same soil types. The plots are sown on fallow (except Rutherglen) as part of the normal rotation practised by the farmer. No yields from crops sown on stubble are included in the data. The seeding and fertilizer rates used were rates recommended for the district and have been constant over the time period. The yield results are the means of 4 or 6 replicates at the one locality each year.

Non-parametric time series tests for randomness

Before the testing of the second hypothesis it was considered inadmissible to use tests based on assumptions of normality for testing the randomness of the yields. Therefore two non-parametric tests were used to test for trend and for cyclical variation. The first, together with tests of significance, is described by Kendall [4] and involves the calculation of the rank correlation coefficient τ (tau). τ lies between $+1$, indicating perfect agreement, and -1, indicating perfect disagreement between the rankings of time and yield. Table 2 contains τ and the statistic Z. No yield series produced a Z statistic which exceeded the 5 per cent level of significance value, viz. 1·96. Unless a rare event has occurred, the yield series at all sites show little evidence of correlation with time.

The second test, the Wallis-Moore phase test [8] was developed for application to time series data to detect cyclical variations. A phase is one sequence of like signs of first differences between successive observations. Several authors, including Bostwick [1] have been criticized by Shaw [7] for failing to adhere to this definition of a phase. The test depends on the comparison of the observed number of phases of given

TABLE 2

Kendall's τ and Wallis-Moore's X^2

Site	τ	Z[a]	6/7 X^2[b]
Manangatang	—0·21	—1·38	1·83
Walpeup	0·24	1·56	1·04
Dooen	0·03	0·21	1·11
Donald	0·11	0·79	1·51
Kaniva	—0·15	—1·01	0·32
Salisbury	—0·22	—1·59	3·52
Warracknabeal	—0·11	—0·73	3·22
Dookie	0·20	1·54	0·18
Rutherglen	0·07	0·84	1·55
Yarrawonga	0·13	0·42	2·76

[a] Critical value at 5% level is 1·96.
[b] Critical value at 5 % level is 5·99.

length with the expected number of phases of that length. The resultant statistic, X^2, is such that 6/7 X^2 is distributed as chi square with 2 degrees of freedom. The results are tabulated in Table 2 and show that no series had a statistic which was greater than the critical value at the 5 per cent level of significance. It is concluded that short run regularities of rising and declining yields do not occur more frequently than would be expected in a random sample of observations.

In summary, an examination of wheat yield series from different regions of Victoria showed little evidence of trend or cyclical variation over time. Therefore it appears reasonable to proceed under the assumption that Victorian wheat yield series are composed of independent observations.

Normality tests

Two measures are required to test for normality. The first, g_1, tests for skewness, the second, g_2, for kurtosis [3, p. 52]. Table 3 contains

TABLE 3

Coefficients of Skewness and Kurtosis

Site	g_1	Std. Error	g_2	Std. Error
Manangatang	—0·752	0·491	—0·560	0·953
Walpeup	0·060	0·491	—1·000	0·953
Dooen	—0·662	0·434	—0·028	0·845
Donald	—0·855	0·456	1·016	0·867
Kaniva	—0·640	0·434	—0·141	0·845
Salisbury	—1·434**	0·434	1·758*	0·845
Warracknabeal	—0·834	0·456	1·137	0·867
Dookie	—0·610	0·434	0·728	0·845
Rutherglen	—0·764	0·491	0·670	0·953
Yarrawonga	—0·443	0·536	0·326	1·038

* significant at 5% level.
** significant at 1% level.

the g_1 and g_2 coefficients as well as their standard errors. Only one location displayed strong evidence of non-normality, Salisbury. The

distribution of yield from Salisbury was negatively skewed and displayed positive kurtosis. The negative skewness indicates that the yields are concentrated at the higher levels, with the median being greater than the mean. Hence the use of a symmetrical distribution from which to sample yields would result in constant underprediction. The positive kurtosis (leptokurtic) indicates that the distribution is peaked and elongated at both tails in comparison with the normal distribution.

The g_1 measure was negative for all localities except one, indicating that the underlying distributions may actually be negatively skewed. To test whether the aggregate was significant, Fisher's procedure for combination of probabilities [3, p. 99] was employed. The chi square value obtained was tested as chi square with 20 degrees of freedom. The calculated chi-square was 31·4 and the value of chi-square from the tables at the 5 per cent level was also 31·4. Therefore while the yield series from each site (with one exception) showed little evidence of skewness, the aggregate over all sites tended to negative skewness.

Summary

The hypotheses of randomness and normality for wheat yields were tested with data from ten localities from the three major wheat regions of Victoria.

The conclusion is that the wheat yield series are composed of random observations forming distributions which are slightly negatively skewed. These results could well differ for yield series from other regions and other crops.

References

[1] Bostwick, Don, 'Yield Probabilities as a Markov Process', *Agr. Econ. Res.*, 14 (2): 49-55, 1962.

[2] Day, R. H., 'Probability Distributions of Field Crop Yields', *J. Farm Econ.*, 47: 713-740, 1965.

[3] Fisher, R. A., *Statistical Methods for Research Workers*, 12th ed., Oliver & Boyd, Edinburgh, 1954.

[4] Kendall, M. G., *Rank Correlation Methods*, Griffin, London, 1948.

[5] McCann, J. McC., and Mullaly, J. V., 'Victorian Wheat: New Zones for Variety Recommendations', *J. Agric. (Vic.)*, 69 (11): 293-295, 1971.

[6] Phillips, J. B., 'Statistical Methods in Systems Analysis', *Systems Simulation in Agriculture*, ed., J. B. Dent and J. R. Anderson, Wiley, Sydney, 1971.

[7] Shaw, L. H., 'The Wallis-Moore Non-Parametric Statistical Test', *J. of Farm Econ.*, 46: 249-250, 1964.

[8] Wallis, W. A. and Moore, G. H., 'A Significance Test for Time Series Analysis', *J. Am. Stat. Assoc.*, 36: 401-409, 1941.