GAU Res. J. 1997. 23 (1) : 67-72

Frequency Distribution of Biological Measurements and Harvest Index of Chickpea

V. P. Singh*, M. L. Lakhera and N. M. Patel

Department of Agricultural Statistics, Gujarat Agricultural University, Anand Campus, Anand-388 110

ABSTRACT

The frequency distributions of grain yield (X), biomass yield (Y) and harvest index (Z) of chickpea have been investigated for the data set of a uniformity trial conducted during the *rabi* season of the year 1993 at the Agricultural Research Station, Gujarat Agricultural University, Derol, Gujarat, India. Three probability distributions, viz., erlang, logistic and normal have been graduated and tested for their goodness of fit. Normal distribution gave good fit for X and Y but for Z, no distribution gave good fit.

(Received on 9th Janu. 1996)

INTRODUCTION

The individual measurements of the experimental units of grain yield (X) are denoted by x_i, and the individual values of the experimental units of total biomass weight (Y) are y_i (i = 1, 2, ..., N). The harvest indices are calculated by $z_i = (x_i / y_i)$. It is of particular interest for application in agronomy and in plant breeding (Donald and Hamblin, 1976).

In many agricultural applications, e.g., selection procedures in plant breeding, analysis of yield and yield components etc., the experimenter is particularly interested in three aspects in handling the harvest indices, viz., calculation of mean harvest indices, variability of harvest indices and correlation among grain yield, biomass weight and harvest index. These aspects have been investigated theoretically by Huan (1990) and Huan *et al.*, (1991) and applied to the data sets of winter oilseed rape. They reported that normal distribution gave a good fit.

The determination of frequency distribution of ratios, viz., coefficient of variation (CV), land equivalent ratio (LER), harvest indices etc. is of particular interest for the field experimenters [Oyejola and Mead (1989), Huan (1992) and Singh (1995)] due to its major role and importance for the development of statistical test vis-a-vis-test of significance.

In the field of theoretical statistics, the distribution of Z = X/Y must not be necessarily known and normally distributed though X and Y are bivariate normally distributed. A general solution of this problem exists (Fieller, 1932) but it can not be simply applied to the

Present address : RBI Endowment Unit, Institute of Rural Management Anand, (IRMA)), Post Box No. 60, Anand, Gujarat-388 001.

biological measurements (harvest indices) because of its complex solution and some restricting conditions imposed on two variables X and Y : $X > 0$, $Y > 0$ and $X < Y$. A specification of general solution for application to harvest indices is calling attention of Agricultural Statisticians and yet to be established (Huan, 1992). Therefore, it is assumed distribution free assumptions for X and Y with a subsequent derivation for the distribution of Z, but to postulate distributional assumptions for Z directly and to prove their validity empirically by using experimental data set based on a uniformity trial on chickpea.

MATERIALS AND METHODS

Data collection

A uniformity trial on chickpea crop was conducted during the *rabi* season of the year 1993 at the Agricultural Research Station, Gujarat Agricultural University, Derol with recommended package of practices. The crop was sown in rows along North-South direction at a spacing of 30 cm between rows. A block of 25 m wide and 30 m long, in a contiguous unit of 1 m^2 giving in all 750 units, was harvested to collect the data on grain yield (X) and biomass weight (Y) for each unit, separately. Thereafter, the harvest index (Z) was derived in per cent for each unit.

Fitting of probability distribution

The study variates (X,Y, Z) were used in preparing frequency distributions. The measures of skewness (β_1) and kurtosis (β_2) based on moments about mean were obtained by the following methods (Hastings and Peacock, 1975).

$$\beta_1 = (\mu_3)^2 / (\mu_2)^3$$

and,

$$\beta_2 = (\mu_4) / (\mu_2)^2$$

where,

$$\mu r = \frac{1}{N} \sum_{j=1}^{k} fj\, q(j - \bar{q})^r, \quad N = \sum_{j=1}^{k} fj,$$

$$(r = 2, 3, 4)$$

where, qj/fj, (j = 1, 2, ..., k), be the frequency distribution of the study variates, then its arithmetic mean $\bar{q}$ will be given by

$$\bar{q} = \left(\sum_{j=1}^{k} fj\, qj \right) / \sum_{j=1}^{k} fj = \text{Arithmetic mean}$$

where, k = number of classes in the frequency distribution.

Assuming that the data from field experiments follow the continuous distribution, the continuous probability distributions which had no parametric relationship among themselves (i.e., no relationship between mean and variance) were considered to judge the distribution of harvest indices. In this regard three distributions, viz., normal, logistic and erlang distributions were examined (Hastings and Peacock, 1975). The corresponding probability density functions (p.d.f.) and distribution functions (D.F.) are given below :

(a) Normal distribution

$$\text{p.d.f.} = f(q) = \frac{1}{\sigma\,(2\pi)^{1/2}} \exp\left[-\frac{(q-\mu)^2}{2\sigma^2}\right]$$

$$\text{D.F.} = \Phi(m) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{m} \exp(-m^2/2)\, dm$$

where $|q| \leq \infty$, $|\mu| \leq \infty$, $\sigma > 0$ and m follows N (o, 1).

(b) Logistic distribution

$$\text{p.d.f.} = f(q) = \frac{\exp\{(q-a)/k\}}{k[1+\exp\{(q-a)/k\}]^2}$$

$$\text{D.F.} = \Phi(q) = 1 - [1+\exp\{(q-a)/k\}]^{-1},$$

where $|q| \leq \infty$, $|a| \leq \infty$, and $k > 0$.

(c) Erlang distribution

$$\text{p.d.f.} = f(q) = \frac{(q/b)^{(c-1)} \exp(-q/b)}{b\{(c-1)!\}}$$

$$\text{D.F.} = \Phi(q) = 1 - [\exp(-q/b)]\left[\sum_{i=0}^{c-1} (q/b)^i / i!\right]$$

where, $q \geq 0$, $b > 0$ and $c > 0$ (c is a positive integer).

In order to fit a continous distribution function to a given frequency distribution (q_i/f_i), mean and variance (second order moment about mean) from the data set were calculated. The estimates of the constant terms (parameters) of probability distribution functions were obtained through the empirical values of mean and variance of the frequency distribution of study variates. Now, at $q = q_i$ (say), the distribution function $\Phi(q_i)$ are computed. Finally, the successive intervals are obtained by subtraction, viz., $\Phi(qi+1) - \Phi(qi)$, (i = 1, 2, ..., k) and on multiplying these k subtracted values by N, i.e., $(\sum_{i=1}^{k} f_i)$, we get the expected frequencies of fitted probability distribution functions (Gomez and Gomez, 1984). The observed and fitted frequency distributions of study variates were tested for their goodness of fit by Karl Pearson's chi-square test (Finney, 1989).

RESULTS AND DISCUSSION

The estimates of mean, range, variance, CV (in per cent), β_1 and β_2 of chickpea data are presented in table 1.

The mean value of grain weight (X) and biomass weight (Y) were obtained 231.80 g and 519.67 g whereas it was 44.98 per cent for harvest index (Z). The correlation coefficient (r) between X and Y was obtained 0.79 which showed a significant positive association between the variables. Scrutiny of raw data exhibited a considerable range of variation, causing a high value of CV which indicated a poor precision (Gomez and Gomez, 1984).

TABLE 1

Estimates of statistical parameters.

Variate	*Mean*	*Range*	*Variance*	*CV*	β_1	β_2
X	231.80	70-510	4868.63	30.102	0.172	3.176
Y	519.67	150-1200	19810.03	27.084	0.302	3.745
Z	44.98	16.17-92.00	79.28	19.795	0.828	7.074

For the normality of the data, β_1 and β_2 should be 0 and 3 but in the present study the values obtained crossed these limits, i.e., $\beta_1 > 0$ and $\beta_2 > 3$. Generally, with field experiment, several workers reported that it is not always possible to get such values of β_1 and β_2 (Bose *et al.;* 1936, Huan; 1992 and Singh; 1995). In a theoretical investigation, Srivastva (1959) reported that test of significance for F and t test were hardly affected if β_1 was between -0.5 and 0.5 and β_2 was between 2 and 4. Therefore, findings summarized in Table 1 for β_1 and β_2 would not affect the test of significance for the variables X and Y except for Z causing a serious problem of non-normality.

All the variables were graduated for goodness of fit for the erlang, logistic and normal probability distributions to get an exact idea about the distributional properties of the frequency distributions of X, Y and Z. The results are being presented in Table 2 for grain weight (X), biomass weight (Y) and harvest index (Z).

Table 2 showed that the grain weight (X) gave best fit for the normal distibution. It gave lowest chi-square value in comparison to other two distributions. This was also supported by the lowest values of β_1 and β_2 (Table 1). In addition to it, if a variable was tested for its goodness of fit by chi-square test for two theoretical probability distributions and both revealed a non-significant chi-square simultaneously, vis-a-vis, a good fit to both probability distributions. Then, the best fit would be one with which chi-square value would be least among these distributions. In view to it, the erlang distribution gave best fit out of all three probability distributions for the biomass weight (Y). But at the same place, biomass weight had also a good fit (non-significant chi-square) for the normal distribution. Hence, the frequency distribution of biomass weight can be considered to follow the normal distribution. The harvest index (Z) was found highly significant at 5 per cent probability level for goodness of fit revealing that no distribution was giving a good fit to it

TABLE 2

Frequency distribution and chi-square values for goodness of fit.

Grain weight (X)		*Biomass Weight (Y)*		*Harvest index (Z)*	
Class	*Observed Frequency*	*Class*	*Observed Frequency*	*Class*	*Observed Frequency*
< 20	0	< 300	32	> 15	0
20 - 60	0	300-340	35	15-20	2
60-100	2	340-380	45	20-25	10
100-140	62	380-420	70	25-30	14
140-180	96	420-460	79	30-35	44
180-220	152	460-500	85	35-40	103
220-260	169	500-540	87	40-45	208
260-300	127	540-580	75	45-50	193
300-340	87	580-620	61	50-55	107
340-380	38	620-660	57	55-60	32
380-420	9	660-700	31	60-65	15
420-460	4	700-740	39	65-70	9
460-500	3	740-780	23	70-75	5
≥ 500	1	≥ 780	31	≥ 75	8
Total	750		750		750
Chi-Square values					
Erlang	23.29*		12.85		66.95*
Logistic	15.50		29.48*		30.92*
Normal	4.58		17.25		74.10*

*Significant at 5% probability level.

(Table 2). The results also supported that the ratio of two normal variates would not be necessarily normal in every case as envisaged by Steel and Torrie (1960).

Conclusion

From the foregoing data analysis and discussion, the present study concluded that the frequency distributions of grain yield and biomass yield gave a good fit to the normal distribution whereas harvest index did not gave a good fit to it. Hence, to have a statistically valid inference from the harvest index based on the ratio of biological measurements in field experimentation, the test of significance should be attempted very carefully.

ACKNOWLEDGMENT

The authors are grateful to the anonymous referee (s) for valuable suggestions leading to the improvement in the paper and wish to put on record their due thanks to the Council of Scientific & Industrial Research (CSIR), New Delhi for providing the Senior Research Fellowship (SRF) to the senior author of paper and the Department of Agricultural Statistics, Gujarat Agricultural University, Anand Campus, Anand for providing all necessary experimental assistance.

REFERENCES

Bose, S. S., Ganguli, P. M. and Mahalanobis, P . C. (1936). Statistical notes for agricultural workers-No. 19. The frequency distribution of plot yields and the optimum size of plots in a uniformity trial with rice in Assam. *Ind. Jour. Agric. Sci.,* **6** (5) : 1107-1121.

Donald, C. M. and Hamblin, J. (1976). The biological yield and harvest index of cereals as agronomic and plant breeding criteria. *Advances in Agronomy,* **28** : 361-405.

Feller, E. C. (1932). The distribution of the index in a normal bivariate population. *Biometrika,* **24** : 428- 440.

Finney, D. J. (1989). Was this in your textbook ? : IV frequency data. *Experimental Agriculture,* **25** (1) : 11- 26.

Gomez, A. K. and Gomez, A. A. (1984). Statistical procedures for agricultural research. John Wiley and Sons, New York.

Hastings, N. A. J. and Peacock, J. B. (1975). Statistical distributions. Butterworth and Comp. Ltd., London.

Huhn, M. (1990). On the variability of harvest indices. *Jour. Agron. and Crop Sci.,* **164** : 271-281.

Huhn, M. (1992). Experimental results on the frequency distribution of harvest indices of winter oil seed rape. *Jour. Agron. and Crop. Sci,* **169** (1-2) : 61-69.

Huhn, M., Grosse, F. and Leon, J. (1991). On harvest indices of winter oilseed rape. *Agron. and Crop Sci.,*. **167** : 299-309.

Oyejola, B. A. and Mead, R. (1989). On the standard errors and other moments for ratios of biological measurements. *Experimental Agriculture,* **25** 473-484.

Singh, V. P. (1995). Variability analysis in field experimentation - A case study on maize. A Ph.D thesis submitted to the Gujarat Agricultural University, Anand Campus, Anand.

Srivastva, A. B. L. (1959). Effect of non-normality on the power of analysis of variance test. *Biometrika*, **46** : 114-122.

Steel, R. G. D. and Torrie, J. H. (1960). Principles and procedures of statistics, McGraw Hill Book Co., Inc. New York.