Jour. Ind. Soc. Ag. Statistics
45(2), 1993 : 206-213

Transformation of Mustard Aphid Counts

Narendra Nath Singh and V.N. Rai
Banaras Hindu University, Varanasi - 221 005
(Received : November, 1990)

Summary

The distribution of aphid (Lipaphis erysimi Kalt.) counts on mustard crop was found to confirm to the log-normal distribution. Logarithmic transformation of the counts normalized the data except in those cases where the variability was small and a large number of zero counts were observed. The success of the transformation was tested by the Kolmogorov-Smirnov test. This test is under use in entomological research and its use is recommended for testing the goodness of fit of theoretical distribution of the appropriateness of transformation for small samples.

The experiment was conducted during winter seasons of 1985-1986 and 1986-87 at research farm, Banaras Hindu University, Varanasi, India.

Key Words : Mustard aphid, probability distribution, transformation, Kolmogorov-Smirnov test.

Introduction

Transformation of biological data is often necessary because transformed data satisfies the necessary conditions (such as normality of data and homogeneity of variance) for validity of the parametric statistical methods such as ANOVA. The problem of choosing a suitable transformation is specially important for data on insect counts. The frequently used transformations are the square-root (x replaced by $\sqrt{x}$ or $\sqrt{x + 0.5}$ if x follows a poission distribution, for which the mean = variance), the logarithmic (x replaced by log x or log (x+1) follows a negative binomial or log normal distribution, for which mean < variance) and the arcsine (p replaced by $\sin^{-1} \sqrt{p}$ if p is a Binomial proportion for which mean > variance) (Snedecor and Cochran, [4] and Elliott, [2]).

For the mustard aphid counts the square root transformation is in use by many authors in India. However, Taylor [5] and Hayn & Lowe [3] suggested Logarithmic transformation for other sp of aphid on the basis of mean variance relationship. No pu work is available regarding the suitable transformation o aphid, *Lipaphis erysimi* Kalt. Therefore, under present attempt was made to determine the suitable mustard aphid counts.

Although the suitability of chosen transformation depends upon the actual distribution (poission, Log Normal, Binomial etc.), the usual practice is to compare only the mean and variance and choose the transformation, because the testing of goodness of fit ofa theoretical distribution is either inconvenient or seems to be impossible. The latter possibility often arises if the data are not extensive so that the Chi-square test of goodness of fit (which is a large sample test) can not be applied because there are only a few observations (= replications). For instance, the commercially available statistical package invariably use the Chi-square test. However, there does exist a very good and simple test, the Kolmogorov-Smirnov Test for testing the goodness of fit of a theoretical distribution to data, and the test is applicable to any number of observations (Conover, [1]). But being non-parametric, it does not have as high a power as the Chi-Square test for large samples. This test is unfortunately not as popular in biological work as it should be.

The aim of the present investigation was to examine the choice of a suitable transformation in the case of mustard aphid, *Lipaphis erysimi* Kaltenbach, which is a major pest of mustard crop.

The major constraint was the small number of observations in each case : since the chi-square test could not be used, we used the Kolmogorov-Smirnov test and found it quite effective in such cases.

2. *Material and Methods*

The present experiment was conducted at the Research Farm, Banaras Hindu University, Varanasi - 221 005, India, during 1985-86 and 1986-87 seasons.

During 1985-86 season only one variety of mustard (T-59) was grown, whereas during 1986-87 two sets of experiment were laid to eliminate the effect of different varieties and date of sowing on the distribution and transformation of aphid population. Under first set of experiment T-59 variety was grown with different dates of sowing, *i.e.* 22nd October, 1985 (D_1), 2nd November, 1986 (D_2), 12th November, 1986 (D_3) and 23rd November, 1986 (D4). In second set of experiment 14 different cultivars, *i.e.* B-35, Local raya, RW-33-2, Krishna, Kranti, Rai B-85, RW 15-6, RH-7847, white flower glossy, RK-8401, RW 2-2, RH-7846, B-85 glossy and RC-1324 were grown. Among them the data of only one randomly selected variety at each observation period was considered for statistical analysis.

To record the aphid population 30 plants were randomly selected from experimental field during 1985-86. During 1986-87 two sets

of experiments were laid under randomised block design. In first set of experiment there were 4 replications and each replication may be divided into 4 plots of equal size 5×3 m^2 to accommodate 4 different dates of sowing. In this case 5 plants were randomly selected and tagged from each plot. The observation on aphid population was confined to these plants only on each observational period.

Similarly in second set of experiment 14 different varieties were sown with 3 replications. In each replication 14 lines of different varieties were grown. In second set of experiment 10 plants were randomally selected from each variety in each replication and observations were recorded on these plants only at each observational period.

The observations on mustard aphid population were recorded from top the 10 cm of the central shoot starting with the incidence of aphid population till harvesting of the crop at weekly intervals.

The mean-variance relation was examined for all cases (Tables 1, 2 and 3) and it was found that mean < variance and also variances are proportional to mean in (almost) all cases. This led to the possibility of examining the log transformation. The mean and variance of transformed values in parentheses (Tables 1, 2, and 3) suggested that the transformation log x or log (x+1) follows the normal distribution. Since the number of observations in each cell was small, the Kolmogorov-Smirnov test was applied to all cases. From the frequency distribution for the number of aphids x, the transformation log (x+1) was applied and the relative cumulative probability distribution, f(x) was determined. The absolute differences (D) between F(x) and NF(x) (normal probability distribution) were calculated. Then the maximum D (Kolmogorov-Smirnov statistic) was compared with tabulated D for a fixed N (where, N is the number of observations) at both 0.05 and 0.01 level.

3. *Results*

The mean and the variance of the data recorded on aphid population from different sets of experiments during 1985-86 and 1986-87 are presented in Tables 1, 2 and 3.

The figures contained in Tables 1, 2 and 3 show that the variances are greater than the corresponding means and also proportional to mean at each date of observations during both the years. This indicates that the logarithmic transformation log (x+1) may be applied in the present case. Now two different questions arise:

(a) Are the variances homogeneous ? We find that the (Table 2) original variable shows more heterogeneity of variances than the transformed variable (Tables 1, 2 and 3). However, the variances are not stabilized completely; for instance, the first and last row of Table 2.

(b) Does x follow the log normal (or negative binomial) distribution ? In other words, does log (x+1) follow the normal distribution ? This was tested by Kolmogorov–Smirnov test because of the small sample sizes. Further, the relationship between transformed-mean (mean of transformed values) and variance (of transformed values) indicates that the transformation applied on the basis of untransformed mean–variance relationship may be suitable, because after the transformation the variances (Tables 1, 2 and 3) are more stable than before, except in the cases where the original vari ability was small and there were a large number of zero counts (*i.e.* mean was small) first and last rows of Table 2.

The maximum D (Kolmogorov–Smirnov statistic) was compared with the tabulated D for a fixed N at both the levels at every interval for each set of experiment during both the years and also presented in Tables 1, 2 and 3.

It was observed that the calculated D at both 0.05 and 0.01 levels at each date of observations (except at first and last dates of observations in the experiments laid with different dates of sowing and different varieties during 1986–87) is less than the tabulated D values.

Thus in both cases (homogeneity of variances and falling of normal distribution), the exceptional cases fall in the first and last rows of table 2. In our opinion, there is a valid reason for such behaviour. When the insect population is scarce (*i.e.*, there are plenty of zero counts together with a few small values) we can expect that the variance will be very small, and that the log normal distribution can not fit the data.

Hence, we can draw the inference that the aphid counts follow the normal distribution better after the logarithmic transformation, irrespective of dates of sowing and varieties. In other words, the aphid population data were well described by log normal distribution.

Table 1. The mean, variance and Kolmogorov-Smirnov Statistic (D) on T-59 variety of mustard during 1985-86 (N=30)

Date of observation during 1986	Mean (X)	Variance (s^2)	Calculated Max. D
5th January	13.47 (1.12)	40.98 (0.04)	0.058
11th January	21.90 (1.30)	191.27 (0.05)	0.133
18th January	35.73 (1.55)	111.09 (0.01)	0.099
26th January	215.53 (2.31)	5650.53 (0.03)	0.091
2nd February	286.63 (2.43)	9529.66 (0.03)	0.137
9th February	426.43 (2.62)	14699.14 (0.01)	0.168
16th February	22.33 (0.65)	339.66 (0.10)	0.073
24th February	4.97 (0.65)	17.47 (0.13)	0.132

1. Figures in parentheses are mean and variance of transformed values : Log (x+1).

2. Tabulated D values at 0.05 and 0.01 levels of significance are 0.242 and 0.290 respectively.

Table 2. The mean, variance and calculated Kolmogorov Smirnov Statistic (D) of aphid counts recorded on T59 variety sown at different dates during 1986-87. (here N = 20)

Date of Observation during 1987	Different Dates of Sowing											
	D_1			D_2			D_3			D_4		
	Mean	Variance	Max. D	Mean	Variance	Max. D	Mean	Variance	Max. D	Mean	Variance	Max. D
5th Jan.	4.4 (0.44)	48.59 (0.24)	.316	2.4 (0.25)	23.62 (0.19)	.466	0.05 (0.02)	0.05 (0.01)	.564	2.0 (0.28)	16.08 (0.13)	.333
12th Jan	26.00 (1.15)	1020.80 (0.29)	.083	46.95 (1.34)	2564.41 (0.36)	.156	45.45 (1.36)	5368.49 (0.21)	.252	7.10 (0.66)	79.39 (0.25)	.207
19th Jan	72.95 (1.71)	4106.25 (0.14)	.116	91.00 (1.63)	6096.49 (0.42)	.142	137.05 (1.95)	11778.70 (0.24)	.206	20.60 (0.82)	1276.63 (0.50)	.225
26th Jan	109.35 (1.95	5181.12 (0.11)	.068	139.00 (1.99)	12334.32 (0.16)	.130	140.90 (2.09)	5234.52 (0.48)	.098	71.20 (1.76)	3034.91 (0.20)	.159
2nd Feb	189.60 (2.25)	5251.90 (0.12)	.109	244.85 (2.36)	7269.27 (0.13)	.121	236.10 (2.34)	7598.61 (0.40)	.159	242.75 (2.32)	15202.89 (0.13)	.093
9th Feb	175.05 (2.21)	4569.76 (0.15)	.140	223.73 (2.43)	6564.24 (0.18)	.153	281.60 (2.42)	9070.66 (0.63)	.010	221.10 (2.32)	5191.20 (0.22)	.152
16th Feb	4.15 (0.35)	77.09 (0.26)	.392	13.45 (0.84)	210.25 (0.37)	.216	73.00 (1.35)	8486.09 (0.70)	.148	13.53 (0.73)	473.50 (0.38)	.181
23rd Feb	0.00 (0.00)	0.00 (0.00)	.500	0.65 (0.13)	1.82 (0.06)	.455	1.00 (0.19)	3.10 (0.08)	.352	0.40 (0.10)	0.64 (0.03)	.462

1. Figures in parentheses are mean and variance of transformed values; Log (X+1).
2. Tabulated D values at 0.05 and 0.01 levels of significance are 0.294 and 0.352 respectively.

Table 3. The mean, variance and calculated Kolmogorov-Smirnov Statistic of aphid counts recorded on different varieties at different observational periods during 1986-87. (N = 30)

Date of Observation during 1986 & 1987	Name of the Varieties	Mean ($\bar{X}$)	Variance (s^2)	Max. D
28th Dec. 86	B-35	1.77 (0.29)	7.67 (0.12)	0.336
5th Jan. 87	Local raya	6.40 (0.61)	59.14 (0.25)	0.189
12th Jan. 87	RW 33-2	15.67 (0.93)	595.36 (0.27)	0.174
19th Jan. 87	Krishna	11.00 (0.86)	88.92 (0.27)	0.184
26th Jan. 87	Kranti	246.57 (1.90)	162941.40 (0.36)	0.185
2nd Feb. 87	Rai B-85	265.93 (2.22)	141948.10 (0.15)	0.141
9th Feb. 87	RW 15-6	156.30 (2.05)	18594.05 (0.14)	0.072
16th Feb. 87	RH 7847	133.37 (1.85	16355.85 (0.40)	0.148
23rd Feb. 87	White Flower Glossy	4.20 (0.55)	14.59 (0.18)	0.238
2nd Mar. 87	RK-8401	2.13 (0.23)	16.81 (0.18)	0.476

1. Figures in parentheses are mean and variance of transformed values : Log (x+1).

2. Tabulated D values at 0.05 and 0.01 levels of significance are 0.242 and 0.290 respectively.

4. *Discussion*

Under present investigation, the relationship of mean and variance reveals that the logarithmic transformation [log (x+1)] may be appropriate for mustard aphid counts.

The Kolmogorov–Smirnov test was used to judge the suitability of the transformation chosen because of the small sample size. This test also confirms that the transformed data followed the normal distribution, which means that the original counts of aphid population follow the log normal distribution. The present finding is in agreement with those of Taylor [5] and Hayman and Lowe [3]. They reported that log (x+1) is the most appropriate transformation for the counts of bean aphid, *aphis fabae* and cabbage aphid, *Brevicoryne brassicae* L. respectively. However, they have not tested the goodness of fit of transformation, whereas, under present investigation it is tested by Kolmogorov–Smirnov test.

The log (x+1) transformation was not found suitable at first and last observational dates (in the case of different dates of sowing and different varieties). This may be due to the occurrence of large number of zero-counts. This simply indicates that there were very few number of aphids which are not able to reflect the original distribution.

From the present study it was concluded that the mustard aphid, *Lipaphis erysimi* (Kalt.) follows log normal distribution and therefore logarithmic transformation [log (x+1)] is the most appropriate transformation for mustard aphid counts.

REFERENCES

[1] Conover, W.J., 1971. Practical Nonparametric Statistics, John Willey & Sons, Inc., New York, London, Sydney, Toronto.

[2] Elliott, J.M., 1979. Some methods for the Statistical Analysis of Samples of Benthic Invertabrates. Freshwater Biological Association, Scientific Publication No. 25.

[3] Hayman, B.L. and A.D. Lowe, 1961. The Transformation of Counts of the Cabbage Aphid. *Newzealand Journal of Science*, **4**, 271–278.

[4] Snedecor, W.G.& W.G. Cochran, 1967. Statistical Methods. Sixth Edition. Ames (Iowa). Iowa State College Press.

[5] Taylor, L.R., 1970. Aggregation and the Transformation of Counts of *Aphis fabae* Scopoli on beans. *Ann. appl. Biol.*, **66**, 181–189.