STUDY OF FREQUENCY DISTRIBUTIONS OF THE CHARACTERS RELATING TO MILK YIELD AND THE CONSEQUENCES OF NON-NORMALITY ON STANDARD TESTS OF SIGNIFICANCE

By Om Parkash and Y. P. Mahajan

Indian Council of Agricultural Research

Received March 1959

Introduction

In statistical analysis it is very often assumed that the character or characters under study follow the normal distribution. In fact normal distribution has played a dominant role both in theoretical and applied statistics since the time of Laplace. However, it has been found that the normal curves could not provide an adequate representation to many of the distributions encountered in statistical practice. Especially, in the field of animal husbandry it has been known for some time past that the various characters, related to milk yield, did not follow the normal law or the normal distribution did not fit them well. In order to carry out the appropriate statistical analysis in such cases, it is necessary to know the nature of such distributions.

Several attempts have been made to construct systems of frequency curves capable of representing a wider variety of distributions occurring in nature, particularly by K. Pearson,[1] Gram-Charlier,[2] Fisher[3] and Edgeworth.[4] Among these, the Pearsonian system of curves is most useful and widely used in practice.

In the past only a few research workers have studied the distributions of characters in the field of animal husbandry and these workers have found that Pearson system of curves fit them well, where the normal distribution fails. Pearl and Miner[5] have made a study of the variation in milk yield and the fat-content of Ayrshire cows from the records of Ayrshire's Cattle Milk Records Committee for the years 1908 and 1909. The investigation was confined to those records of cows which had been 32 weeks or more in milk. The frequency distributions of milk yield and fat-content of cows of the same age were found for each year of age. The authors found that Pearson curves fitting these distribution were of the type I, II, III and IV. Normal curves were also found to fit some of the distributions. J. W. Gowen[6] has published a memoir giving the results of his analysis of the records of a Jersy herd over a number of years. He extracted 1,741 complete eight months' records of healthy cows for analysis. Gowen found that Pearsonian curves of type I, II, III, IV and V prevailed for the distributions of milk yield and butter-fat and the fitted curves were found to be satisfactory. J. F. Tocher[7] made a detailed study of the milk yield data of dairy cows. In the course of statistical analysis of the data of 4,912 Ayrshire cows having complete records and also having a second calf within 60 weeks of previous calving, he found that age of cow followed the Pearson distribution of type I. He found that the regression of milk yield is a function of age and its distribution pooled over all ages will be heterogeneous. He fitted type IV curve to the distribution of milk yield for all ages. The fit was found to be poor owing to the heterogeneity of material with respect to age. He, therefore, fitted Pearson type IV curves to the distributions of milk yield for each year of age and the fits were found to be quite satisfactory.

The paper considers the data relating to yield of four herds, *viz.*, Tharparkar maintained at the Government Cattle Farm, Patna, Kangayam from Livestock Research Station, Hosur, and two Red Sindhi herds maintained at the Livestock Research Station, Hosur, and the Southern Regional Station of the National Dairy Institute, Bangalore, respectively. The data of these herds has already been statistically analysed in the Indian Council of Agricultural Research.[8] The main object of this paper is to examine the data with a view to determine the types of probability distributions of the various characters, *e.g.* (lactation yield, lactation length, age at first calving and calving interval) which fit them best and also to examine to what extent the usual statistical techniques based on the assumption of normality are valid in the case of non-normal populations.

It was found that most of the distributions occurring were non-normal. Of the different types of curves belonging to the Pearsonian system, types I and IV were generally the closest fits but even some of these types were found to be not quite satisfactory. An attempt has been made in the second part of this paper to study the effect of non-normality on the standard tests of significance such as the Standard Normal Deviate test, the Students' '*t*'-test, the Chi-square test and the 'F'-test.

Attempts to study the effect of non-normality have been made by several authors, specially by Pearson.[9] The standard test for testing the equality of variances is the F-test based on the statistic.

$$F = \frac{\sum_{i=1}^{m} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \cdot \frac{(n-1)}{(m-1)}$$

with $(m-1)$, $(n-1)$ degrees of freedom.

Since the F-test is derived on the assumption of normality, it is sensitive to non-normality. This was first pointed out by Pearson on the basis of some experimental investigations. His results were confirmed by Geary[10] who showed that the probabilities derived on the assumption of normality may differ seriously from the true probabilities even in cases where the degree of non-normality is not considerable. Studies undertaken by Gayen[11] and Box[12,13] also indicated the same situation. The effect of non-normality on the test for comparing the means has also been examined by several authors, particularly by Pearson,[14] Gayen[15] and Geary.[16] It was found that the '*t*'-test for comparing means was not so sensitive to non-normality. However, no systematic attempt has been made to study the effect of non-normality on the standard test of significance when samples are drawn from a type IV population which cannot be represented either by Pearsonian system of curves or the Gram-Charlier's series. An attempt has been made in this paper to study empirically the consequences of non-normality on the standard tests of significance mentioned above by computing the power curves particularly in the case of Pearsonian type IV population.

2. Material for Study

The material used for study in this paper relates to the four herds, *viz.*, Tharparkar, Kangayam and two Red Sindhi herds referred earlier. The records for study considered in this paper relate only to normal calvings of cows stationed and provided information in respect of the following four characters:—

(i) Lactation yield, (ii) Lactation length, (iii) Calving Interval and (iv) Age at first calving.

The number of observations available for these characters for the four herds are given below;—

Character	Number of observations			
	Tharpar-kar	Kangayam	Sindhi Hosur	Sindhi Bangalore
First lactation ..	421	440	290	261
Second lactation ..	371	265	218	236
Third lactation ..	312	237	151	188
Fourth lactation ..	268	144	..	..
Age at 1st calving	380	440	304	227
Interval between the 1st and 2nd calving	370	296	241	..
Interval between the 2nd and 3rd calving	327	219	175	..

3. STUDY OF TYPES OF FREQUENCY DISTRIBUTION

The process of curve fitting enables us to predict the expected values. When it becomes necessary in practical work to decide on a system of curves for describing frequency distributions it should be borne in mind that the number of constants in the frequency function should be as small as possible. For a large number would involve use of higher order moments which cannot be estimated with high precision.

Each particular curve has its own method of investigation and in order to determine to what system a curve belongs it is necessary to consider certain constants. The constants most commonly used for determining curve types are β_1 and β_2 based on the first four moments about the mean defined by

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}\,; \quad \beta_2 = \frac{\mu_4}{\mu_2^2}\,.$$

On the basis of these coefficients, the distributions are tested for normality. If the distribution is found to be non-normal, the appropriate Pearsonian type is determined by the K-criterion defined as

$$K = \frac{\beta_1 (\beta_2 + 3)^2}{4\,(4\beta_2 - 3\beta_1)\,(2\beta_2 - 3\beta_1 - 6)}\,.$$

The value of the constants and the Pearsonian types for the 38 sets of data in respect of the four characters mentioned in Section 2 are given in Table I and it will be seen that of these, 14 are type I, 15 type IV, 3 type VI, one each of type V and VII. The remaining 4 are found to be normally distributed.

Each of the frequency distributions was then fitted with the appropriate Pearson type curve as determined above following Elderton's approach. It was found that 3 type I, 6 type IV, 1 type V and 3 normal distributions showed good agreement between the observed and the expected frequencies. Two of the type 1 curves were found to be J-shaped. The frequency curves showing good agreement between the observed and the expected frequencies are shown in Diagrams 1–13.

Most of the other distributions encountered had a common peculiarity namely depression at the mean and in these cases no satisfactory fit could be obtained. As seen above, the type IV Pearsonian curve occurs most frequently. Since this type has a complicated mathematical frequency function it is not convenient to use it in the computation of tolerance limits and derivations of appropriate test criteria. An attempt has been made to fit some other systems of curves such as Gram-Charlier's series and Fisher's series to the type IV data relating to fourth lactation yield (Tharparkar). In all the above cases, the overall agreement of the observed frequencies with the expected was found to be satisfactory but the fitted curves could not be used to obtain tolerance limits as there was some discrepancy in the tail area in both the Gram-Charlier's series and Fisher's series.

4. Study of Non-Normality

The purpose of this section is to study the effect of non-normality on standard tests of significance based on the assumption of normality. The tests of significance considered in this section are described below:—

Standard Normal Deviate Test.—This test is based on the statistic

$$d = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

where μ and σ^2 are the population mean and variance respectively.

χ^2-Test of Significance.—This test is based on the statistic

$$\chi^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{\sigma^2}$$

which is distributed as χ^2 with $(n - 1)$ *d.f.*

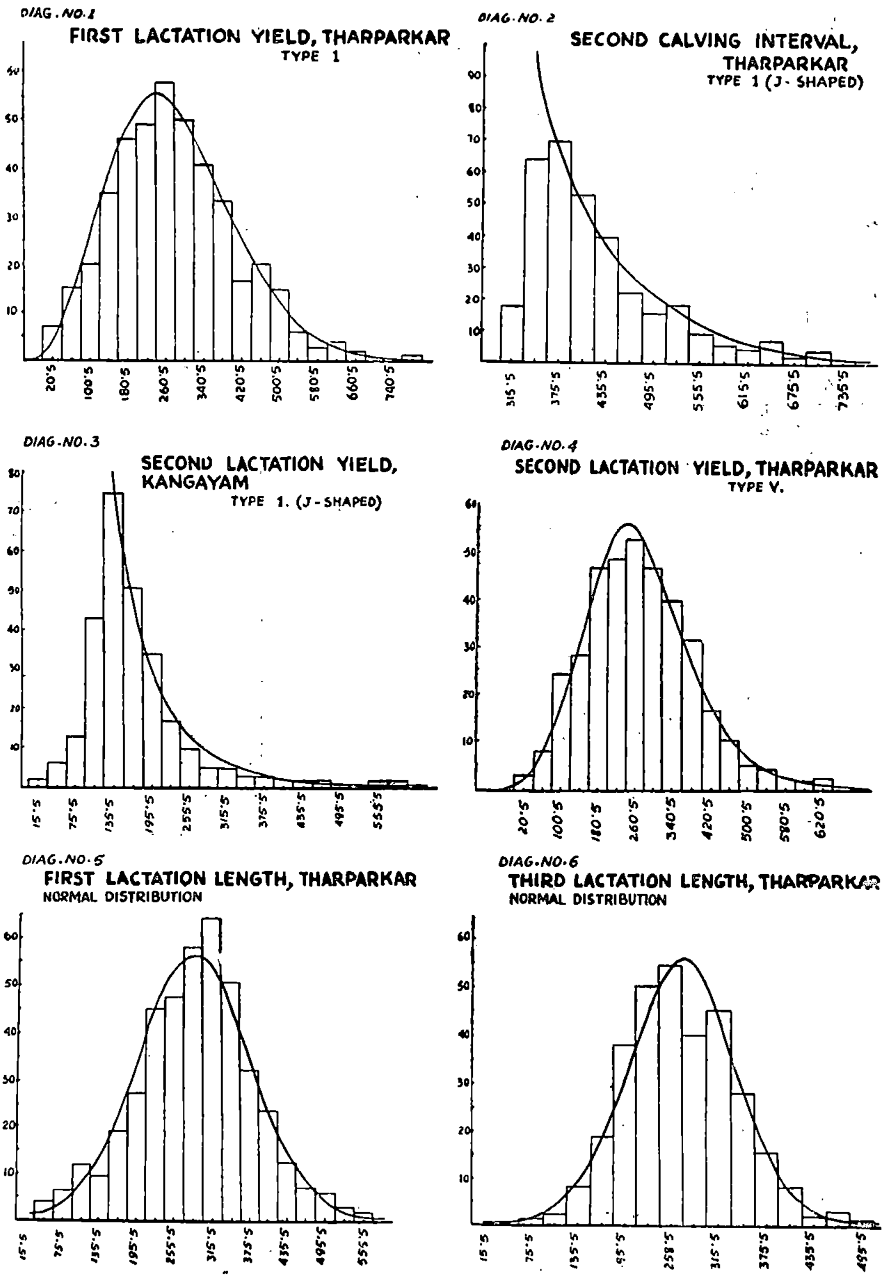
TABLE I

β_1, β_2 and the type of the Pearsonian curve for the characters under study

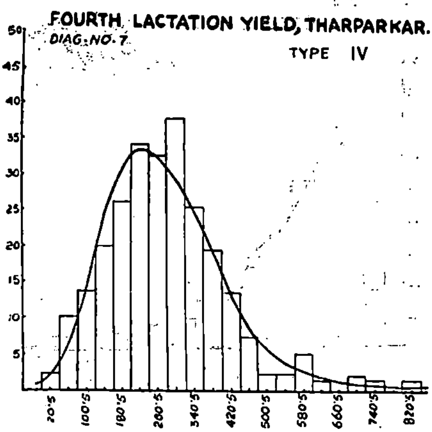
Herd / Character		THARPARKAR			KANGAYAM			SINDHI HOSUR			SINDHI BANGALORE		
		β_1	β_2	Type of Pearsonian curve	β_1	β_2	Type of Pearsonian curve	β_1	β_2	Type of Pearsonian curve	β_1	β_2	Type of Pearsonian curve
First lactation	Yield	·2433	3·1285	I (– ·4099)*	2·2261	8·7156	IV (·5702)	0·0555	3·6035	IV (0·0407)	0·0107	2·4900	Normal Distribution
	Length	·0188	3·2005	Normal distribution	·7162	5·2655	IV (·2715)	2·4613	5·6456	I	0·2507	3·0807	I
Second lactation	Yield	·2589	3·4557	V† (1·5350)	4·7177	9·5627	I (–7·5161)	0·0994	5·4180	IV (0·0181)	0·0533	3·3353	Normal distribution
	Length	·0027	3·8652	VII† (·0007)	·6782	4·8752	IV (·3509)	0·8949	6·9397	IV (0·1696)	0·0422	3·7009	IV (0·200)
Third lactation	Yield	·7273	3·6386	I (– ·7159)	·9944	4·5508	VI (7·8655)	0·6819	4·9596	IV (0·3240)	0·7032	5·3621	IV (0·1884)
	Length	·0700	3·2173	Normal distribution	·40[illegible]4	4·1745	IV (·2897)	0·1968	3·1560	I	0·0314	3·9252	IV (0·0137)
Fourth lactation	Yield	·8700	4·8978	IV (·6738)	1·8106	6·6117	VI (1·1107)	..	..	..	..	..	..
	Length	1·1168	6·9118	IV (·1760)	·1713	3·6702	IV (·1627)	..	..	..	..	..	..
Age at first calving		·4914	3·6525	I (–2·4462)	1·8010	5·3791	I (–3·0414)	1·0318	5·2596	IV (·6889)	1·8433	6·3485	VI (1·7373)
Interval between first and second calving		·3787	2·5232	I (– ·1035)	·3608	2·5839	I (– ·1587)	0·0292	2·0642	I	..	..	..
Interval between second and third calving		1·6543	4·4779	I (– ·8898)	1·3176	4·1782	I (– ·8332)	0·3902	2·4349	I	..	..	..

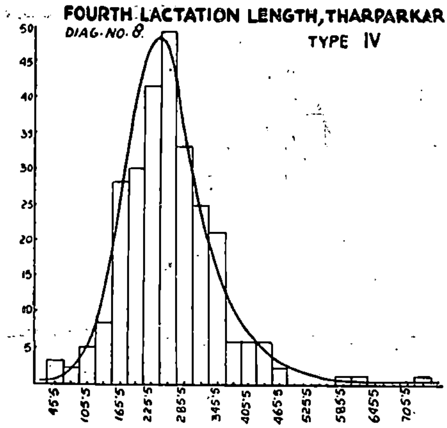
* The figures in the brackets indicate the value of criterion K defined in Section 3.

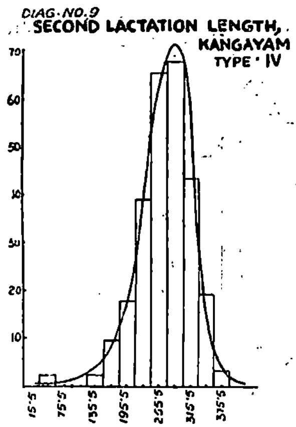
† Types V and VII are the degenerate forms of Type VI and IV respectively.

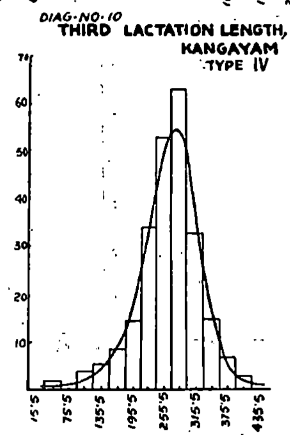

DIAG. NO. 1
FIRST LACTATION YIELD, THARPARKAR
TYPE 1
10 20 30 40 50 60
20·5 100·5 180·5 260·5 340·5 420·5 500·5 580·5 660·5 740·5
DIAG. NO. 2
SECOND CALVING INTERVAL, THARPARKAR
TYPE 1 (J- SHAPED)
10 20 30 40 50 60 70 80 90
315·5 375·5 435·5 495·5 555·5 615·5 675·5 735·5
DIAG. NO. 3
SECOND LACTATION YIELD, KANGAYAM
TYPE 1. (J-SHAPED)
10 20 30 40 50 60 70 80
15·5 75·5 135·5 195·5 255·5 315·5 375·5 435·5 495·5 555·5
DIAG. NO. 4
SECOND LACTATION YIELD, THARPARKAR
TYPE V.
10 20 30 40 50 60
20·5 100·5 180·5 260·5 340·5 420·5 500·5 580·5 620·5
DIAG. NO. 5
FIRST LACTATION LENGTH, THARPARKAR
NORMAL DISTRIBUTION
10 20 30 40 50 60
15·5 75·5 135·5 195·5 255·5 315·5 375·5 435·5 495·5 555·5
DIAG. NO. 6
THIRD LACTATION LENGTH, THARPARKAR
NORMAL DISTRIBUTION
10 20 30 40 50 60
15·5 75·5 155·5 95·5 258·5 315·5 375·5 435·5 495·5

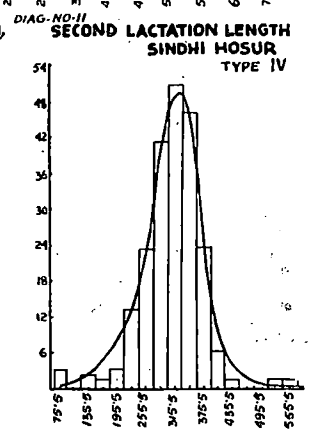
DIAGRAMS 1–6

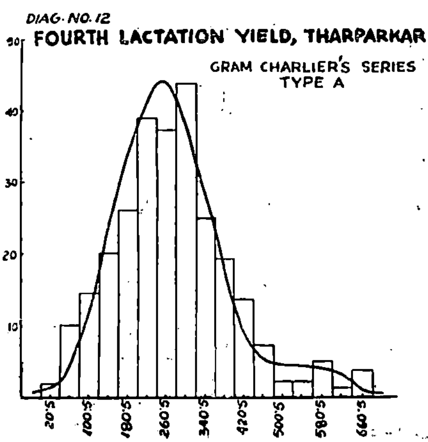


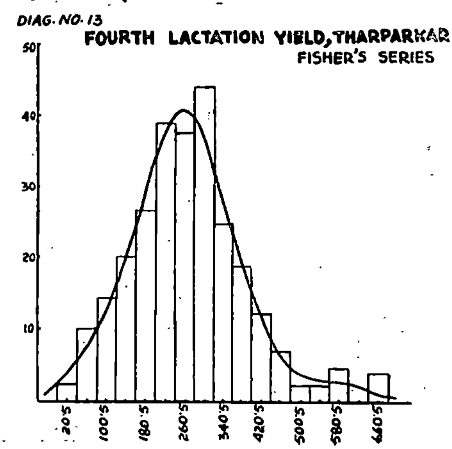












DIAGRAMS 7–13

Students' (t)-Test.—This test is based on the statistic

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

which is distributed as students t with $(n - 1)$ *d.f.* $\bar{x}$ and s^2 being the sample mean and variance respectively.

The Variance Ratio F-Test.—This test is based on the statistic

$$F = \frac{\sum_{i=1}^{m} (x_i - \bar{x})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \frac{(n - 1)}{(m - 1)}$$

which is distributed as F with $(m - 1)$, $(n - 1)$ *d.f.*

4.1. *Effect of Non-Normality on the Distributions of Standard Test Statistics*

All the tests of significance described above are based on the assumption of normality. In this sub-section the effect of non-normality on the distribution of standard test statistics will be studied empirically. For this purpose samples of varying sizes were drawn from *five* different populations. The test statistics described above were computed for all these samples. Frequency distributions of each of these statistics were then obtained for each of the *five* different populations and for different sample sizes. The agreement of these with the expected on the assumption of normality was then tested by the χ^2-test for goodness of fit. It was generally found that the distribution of 't' statistic remained insensitive to non-normality while the distribution of χ^2 statistic was affected significantly in some populations. The summary of these results is given in Table II.

It was also found that the distribution of sample means tends to normal distribution as the sample size tends to infinity, the rate of convergence to normality depending on the skewness of the population. This can be seen from Table III giving the sample size for which the distribution of means tends to normal for different non-normal populations.

TABLE II

Study of non-normality for distributions of different statistics for various populations.

Population considered	β_1 of the population	β_2 of the population	Type of Pearsonian curve	Sample size	No. of samples	Distribution of sample means	Value of goodness of fit for sample variance	Value of goodness of fit for sample '*t*'
First lactation yield Sindhi Hosur	0·0555	3·6035	IV	15	100	Normal	8·75 for 7 d.f. (14·067)	3·52 for 8 d.f. (15·507)
				10	160	Normal	7·81 for 7 d.f. (14·067)	4·409 for 9 d.f. (16·919)
				5	300	Normal	12·151 for 7 d.f. (14·067)	7·585 for 7 d. (14·067)
Second lactation yield Sindhi Hosur	0·0994	5·4180	IV	10	100	Normal	10·335 for 6 d.f. (12·592)	4·16 for 8 d.f. (15·507)
				5	250	Normal	14·129 for 7 d.f.* (14·067)	10·607 for 7 d.f. (14·067)
Second lactation length Sindhi Hosur	0·8947	6·9397	IV	5	300	IV	Highly significant† value	4·451 for 7 d.f. (14·067)
First lactation length Sindhi Bangalore	0·2507	3·0807	I	10	200	Normal	Highly significant† value	Significant value*

Values in the brackets are the tabulated values. * Indicates significance at 5% level. † Indicates significance at 1% level.

TABLE III

Sample size for which the distribution of the means tends to normal for different non-normal populations

Sl. No.	Character	Pearsonian Type	β_1	β_2	Size of sample for normality
1	First lactation yield, Tharparkar	I	·2433	3·1285	5
2	First calving interval, Kangayam	,,	·3608	2·5839	10
3	First calving interval, Tharparkar	,,	·3787	2·5232	10
4	Age at first calving, Tharparkar	,,	·4914	3·6525	10
5	Third lactation yield, Tharparkar	,,	·7273	3·6386	15
6	Second calving interval, Kangayam	,,	1·3176	4·1782	25
7	Second calving interval, Tharparkar	,,	1·6543	4·4779	35
8	Age at first calving, Kangayam	,,	1·8010	5·3791	40
9	Second lactation yield, Kangayam	,,	4·7177	9·5627	100
10	Fourth lactation length, Kangayam	IV	·1713	3·6702	10
11	Third lactation length, Kangayam	,,	·4004	4·1745	20
12	Second lactation length, Kangayam	,,	·6782	4·8752	30
13	First lactation length, Kangayam	,,	·7162	5·2655	35
14	Fourth lactation yield, Tharparkar	,,	·8700	4·8978	30
15	Fourth lactation length, Tharparkar	,,	1·1168	6·9118	60
16	First lactation yield, Kangayam	,,	2·2216	8·7156	90
17	Third lactation yield, Kangayam	VI	·9944	4·5508	25
18	Fourth lactation yield Kangayam	,,	1·8106	6·6117	70
19	Second lactation yield, Tharparkar	V	·2589	3·4557	10
20	Second lactation length, Tharparkar	VII	·0027	3·8652	5

4.2. *Effect of Non-Normality on the Level of Significance*

All the standard tests of significance are based on the assumption that the parent distribution is normal. For example, consider students' 't' test to test the null hypothesis $H_0 : \mu = \mu_0$ against the alternative $H_1 : \mu \neq \mu_0$. Let t_0 be the tabulated value of t at level of significance α for $(n - 1)$ *d.f.* If the observed t value exceeds t_0 we reject H_0 and accept H_1 at level of significance α. To study the effect of nor-normality on the level of significance, samples of varying sizes were drawn from different non-normal populations and the percentage frequency of rejection was found for different levels of significance. Similar procedure was carried out in regard to other tests. The results pertaining to this analysis are given in Table IV. It is seen from the table that the level of significance was affected significantly mostly in the case of χ^2 and F-test.

4.3. *Effect of Non-Normality on the Power of the Tests of Significance*

Let $x_1, x_2, \ldots, x_n$ be a random sample of n independent observations from a population with a frequency function $f(x, \mu)$ where μ is the population mean. Consider the problem of testing hypothesis H_0 that the population μ has a specified value μ_0 against the alternative H_1 that $\mu > \mu_0$. A test of hypothesis H_0 is a region W in the sample space called the critical region such that if the observed sample point falls in the critical region, the hypothesis H_0 is rejected. The region W is so chosen that the probability of rejecting the hypothesis when the hypothesis is true is minimum and the probability of rejecting the hypothesis when the hypothesis is not true is maximum. The probability of rejecting the hypothesis when the hypothesis is true is called type I error of the test or in symbols.

$$P\{(x_1, x_2, \ldots, x_n) \in W/H_0\} = \alpha \text{ (say)} = \int_W f(x_1, x_2, \ldots, x_n/H_0)\, dx_1 . dx_2 \ldots dx_n \quad (1)$$

where α is the level of significance of the test. Usually in practice, α is fixed and then W is so chosen that the probability of rejecting the hypothesis when the hypothesis is not true is maximum. This probability of rejection or in symbols

$$P\{(x_1, x_2, \ldots, x_n) \in W/H_1\} = \int_W f(x_1, x_2, \ldots, x_n/H_1)\, dx_1 . dx_2 \ldots dx_n \quad (2)$$

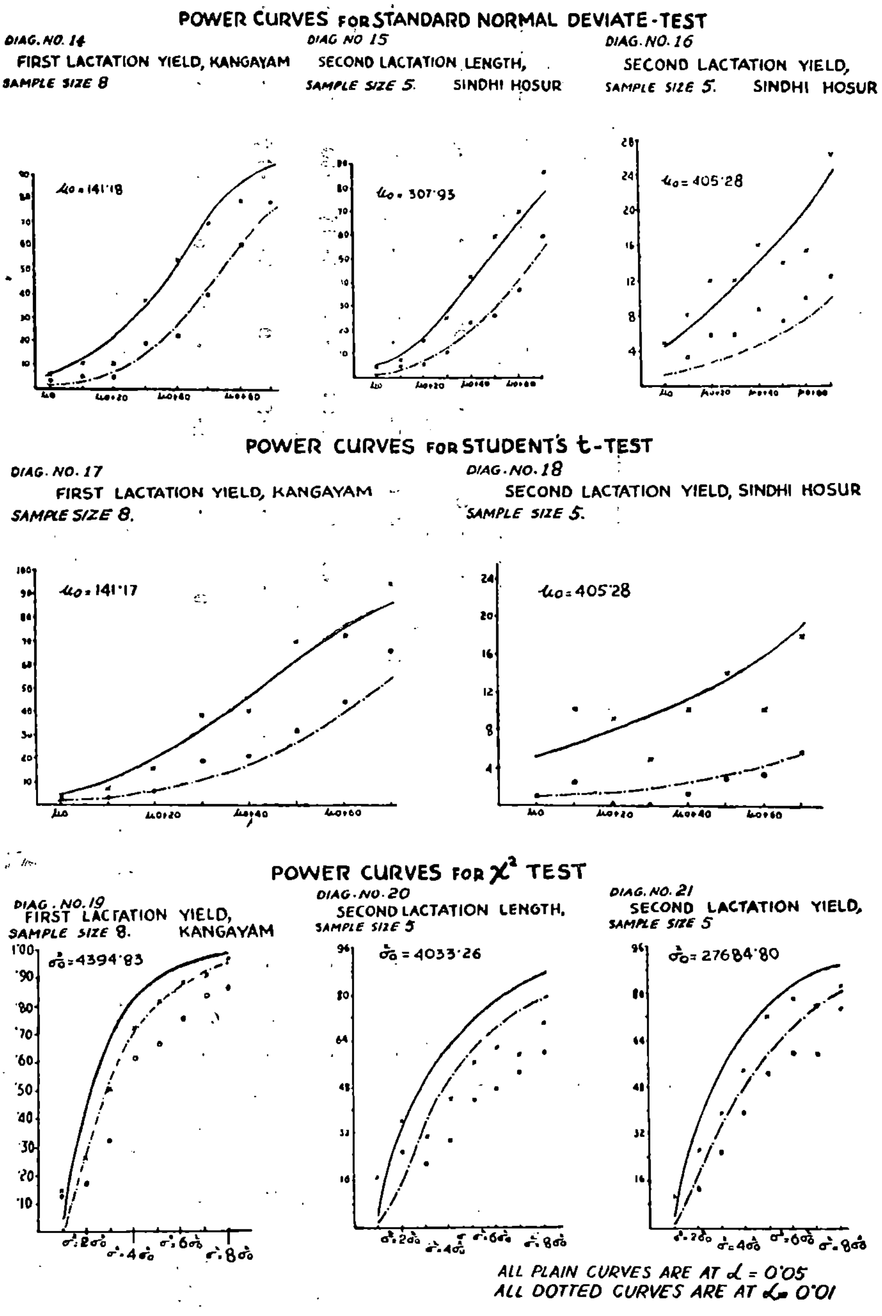
is called the power of the test. From (1) and (2) it is evident that the level of significance and the power of the test depend upon the frequency function of the parent populations.

TABLE IV

Effect of non-normality on the level of significance

Population considered	β_1 of the population	β_2 of the population	Type of Pearsonian curve	Size of the sample	Total no. of samples = N	Percentage no. of rejected samples at α level of significance for S.N.D. test		Percentage no. of rejected samples at α level of significance for 't'-test		Percentage no. of rejected samples at α level of significance for χ^2-test		Percentage no. of rejected samples at α level of significance for 'F'-test	
						$\alpha=0\cdot01$	$\alpha=0\cdot05$	$\alpha=0\cdot01$	$\alpha=0\cdot05$	$\alpha=0\cdot01$	$\alpha=0\cdot05$	$\alpha=0\cdot01$	$\alpha=0\cdot05$
First lactation yield Sindhi Hosur	0·0555	3·6035	IV	5	300	1·00	4·66	1·00	3·66	2·66†	6·33	1·33	6·66
Second lactation yield Sindhi Hosur	0·0994	5·4181	IV	5	300	0·66	6·00	1·66	5·66	5·00†	11·00†	3·33†	8·33†
Second lactation length Sindhi Hosur	0·8947	6·9397	IV	5	300	3·33†	8·66†	1·66	5·00	5·60†	12·60†	1·66	9·66†
First lactation length Sindhi Bangalore	0·2507	3·0807	I	10	200	0·50	6·00	2·00	10·50†	0·00	3·50	2·50	8·50*

*Indicates significance at 5% level. † Indicates significance at 1% level.



DIAGRAMS 14–21

The standard test of significance described earlier are all based upon the assumption of normality. These tests are known to be most powerful when the parent population is normal. As most of distributions investigated in this paper are of type IV, we are particularly interested in developing the most powerful tests for comparing means and variances when the parent population is of type IV. Since the type IV frequency function is rather complicated, this problem involves formidable difficulties and does not lead to any simple solution. An attempt has therefore been made to estimate empirically the loss in power resulting from the use of the standard tests of significance based on the assumption of normality when the parent population is non-normal. For this purpose, hundred samples of a given size were drawn from a similar non-normal population with mean $\mu_0 + \delta$, generated from the original population with mean μ_0. The S.N.D. test was then performed at 5% and 1% level of significance and the frequency of rejection was taken to be an estimate of the power of the test at the point $\mu_0 + \delta$ where $\delta = 10$. Similar analysis was carried out to estimate the power at the points $\mu_0 + 2\delta$, $\mu_0 + 3\delta$, ..., $\mu_0 + 7\delta$. A similar procedure was also carried out to estimate the power curves of the other two tests when the parent population was non-normal. All these points of empirical power curves were plotted along with their corresponding theoretical power curves calculated on the basis of normal parent populations. It was observed that the power of the S.N.D. test and 't'-test remained insensitive to non-normality while the power of χ^2-test was affected considerably. This can be seen from the power curves given in Diagrams 14–21. For the sake of convenience the investigation was confined to the data on

(I) First lactation yield, Kangayam.

(II) Second lactation yield, Sindhi Hosur.

(III) Second lactation length, Sindhi Hosur.

The empirical points of power curves for No. I, II, III populations are shown by the symbols 'X' and '0' at $\alpha = \cdot 05$ and $\alpha = \cdot 01$ respectively along with the corresponding expected power curves.

5. Summary

The object of this paper was to study the distributions of milk yield data and its related characters like lactation yield, lactation length, age at first calving and calving interval. The data related to four herds, *viz.*, Tharparkar, Kangayam, Sindhi Hosur, and Sindhi Bangalore. The distributions of these characters were mostly found to be

nonnormal. Among the 38 sets of data studied, 14 were found to be Pearsonian type I, 15 to be type IV, 3 type VI, one each of type V and VII. The remaining 4 were found to be normally distributed.

Effect of these non-normal distributions on standard tests of significance such as Standard Normal Deviate test, Student's t-test, χ^2-test and F-test was studied empirically. This study was made in respect of the sampling distributions of these statistics, the level of significance and the power of the standard tests. It was found that S.N.D. test and 't'-test were insensitive to non-normality while χ^2- and F-tests were affected significantly.

6. Acknowledgement

The authors avail this opportunity of acknowledging their sincere thanks to Dr. B. V. Sukhatme, Professor of Statistics, Indian Council of Agricultural Research, for his valuable guidance, assistance and constant encouragement which have been very helpful in preparation of this paper.

The authors are also highly indebted to Dr. V. G. Panse, Statistical Adviser to Indian Council of Agricultural Research, for his invaluable criticism and suggestions.

References

1. Pearson, K. .. "Contribution to mathematical theory of evolution," *Phil. Trans. Roy. Soc.*, 1895, **186**, Part I.
2. Gram-Charlier .. *A New Form of Frequency Functions.* Meddelande from Lands Astronomiska Observatorium, Series II, No. 51.
3. Fisher, Arne .. *Frequency Curves*, MacMillan, 1928.
4. Edgeworth .. *The Law of Error*, Cambridge Philosophical Transactions, 20.
5. Pearl and Miner .. "Study of variation of lactation yield and fat-contents Ayrshire cows," *Jour. of Agric. Res.*, September 1919, **16**.
6. Gowen, J. W. .. "Study of variation of lactation yield of Jersy herd," *Genetics*, March 1920, **5** (2); *Ibid.*, May 1920, **5** (3).
7. Tocher, J. F. .. "Investigation of milk yield of dairy cows," *Biometrika*, 1928, **20 B**.
8. I.C.A.R. .. *Report on the Scheme for Statistical Analysis of the Breeding Data Collected at the Government Cattle Farm, Patna, Livestock Research Station, Hosur and the I.D.R.I., Bangalore.*

9. Pearson, E. S. .. *Biometrika*, 1928, **20 A.**

10. Geary, R. C. .. *Ibid.*, 1937, **34.**

11. Gayen, A. K. .. *Ibid.*, 1950, **37.**

12. Box, E. P. .. *Ibid.*, 1953, **40.**

13. ——— .. *Jour. Roy. Stat. Soc.*, 1955.

14. Pearson, E. S. .. *Biometrika*, 1928, **20A.**

15. Gayen, A. K. .. *Ibid.*, 1950, **36.**

16. Geary, R. C. .. *Jour. Roy. Stat. Soc.*, September 3, 1936, **69.**