STATISTICAL NOTE ON THE SIGNIFICANT CHARACTER OF LOCAL VARIATION IN PROPORTION OF DEXTRAL AND SINISTRAL SHELLS IN SAMPLES OF THE SNAIL *BULIMINUS DEXTROSINISTER* FROM THE SALT RANGE, PUNJAB.

By P. C. MAHALANOBIS, *B.Sc., M.A. (Cantab.), Professor of Physics, Presidency College, Calcutta.*

Dr. Annandale, Director, Zoological Survey of India, sent me the following data for a statistical report on a collection of shells of the snail *Buliminus dextrosinister* from the Salt Range.

Locality.	DEXTRAL.		SINISTRAL.		TOTAL
	No.	Percentage.	No.	Percentage.	
1. Chalisa	2	100	...	...	2
2. Katas	10	83·3	2	16·7	12
3. Kallar Kahar ...	21	33·3	49	66·7	70
4. Sardhi	2	4·2	46	95·8	48
TOTAL ...	35	37·7	97	62·3	132

Two distinct questions arise.

(*a*) Do dextral and sinistral shells occur in nature in equal proportions or do they occur in different proportions ?

(*b*) If in different proportion, is this proportion the same for all localities, or is there local variation in the proportion ?

Let us assume that the shells occur in equal proportions in nature. Then the " theoretical proportion " of each is $\frac{1}{2}$. For example in the total sample, the " theoretical proportion " would be 66 of each. The " observed proportion " is 35 dextral and 97 sinistral.

Of course we cannot expect to get 66 dextral shells in *each* sample of 132. The " observed proportions "will vary on account of fluctuations of sampling. The question is whether difference between the " observed " and the " theoretical " proportion is merely due to such fluctuations of sampling or whether it is indicative of a real difference in the proportions in which dextral and sinistral shells occur in nature.

To put it in a slightly different way : A sample of 35 dextral and 97 sinistral shells may sometimes occur on account of fluctuations of sampling. But how often ? In other words, the precise mathematical question is : What is the exact probability of such occurrence ?

The solution is well-known and may be stated quite generally.

If the " theoretical proportion " of dextral and sinistral shells is p and q respectively ($p + q = 1$) then the probability of occurrence of a sample of s dextral and r sinistral ($s + r = m$) in a sample of m is given by

$$C_s = \frac{|\underline{m}}{|\underline{s}\;|\underline{r}} \cdot (p)^s\,(q)^r$$

By direct arithmetical calculation I obtain the following table.

TABLE II.—*Probability of occurrence of "observed" samples for* $p=q=\frac{1}{2}$.

Locality.	Number of dextral shells.	Probability.	odds against
1. Chalisa	2	0·25	4 to 1
2. Katas	10 or more.	0·01 92 86	51 to 1
3. Kallar Kahar	21 or less	$5{\cdot}46\times 10^{-4}$	18 28 to 1
4. Sardhi	2 or less	$4{\cdot}16\ 21\times 10^{-12}$	$24{\cdot}10^{10}$ to 1
TOTAL ...	35 or less	$3{\cdot}15\ 27\times 10^{-8}$	$32{\cdot}10^{6}$ to 1

In the case of *Chalisa* the evidence is quite inconclusive. *Katas* also is doubtful ; odds of 51 to 1 are not sufficiently high to justify us in asserting that the proportion of dextral shells is really greater than $\frac{1}{2}$. On the other hand the odds are very considerable in the case of *Kallar Kahar* and practically overwhelming in the case of the *Sardhi* and the *Total.* Thus on the whole we may reasonably argue that dextral shells occur in substantially lower proportion in nature.

I may now pass on to the second question, namely, is there any local variation in the proportion ? The problem may be stated quite generally as follows.

In a first sample of n shells, p are found to be dextral and q sinistral ($p + q = n$), what is the chance of occurrence of s dextral and r sinistral shells in another sample of m ($m = r + s$)?

The solution is given by Bayes' theorem and the result can be easily calculated with the help of formulae given by M. Greenwood[1] and Karl Pearson.[2]

$$C_s = \frac{|\underline{n+1}\;|\underline{m}}{|\underline{p}\;|\underline{q}\;|\underline{m+n+1}} \cdot \frac{|\underline{p+s}\;|\underline{q+r}}{|\underline{s}\;|\underline{r}}$$

Greenwood[3] says " Let the chance of a_2 or more successes in n_2 (s or more successes in m_1 in our notation) after a_1 successes in n_1 (*i.e.* p successes in n) be p_2 and the chance of a_1 or less successes in n_1 trials (p or less in n) after a_2 successes in n_2 (s success in m) be p_2. Then, since either n_1 or n_2 might have been drawn first, a measure of the probability of the observed result will be $\frac{1}{2}(p_1 + p_2)$."

[1] *Biometrika*, IX, pp. 69-90.

[2] *Phil. Mag.* 1907, p. 365 also *Biometrika*, XIII, 1920, pp. 1-16, and *Tables for Statisticians and Biometricians*, pp. lxx-lxxiii.

[3] *Ibid.* p. 81.

I am not quite sure about the validity of Greenwood's argument, at least, in its application to the present problem. It seems desirable to give full weight to the most favourable probability for agreement before differentiation is definitely asserted. On this principle the more numerous sample should invariably be taken as the first sample. Doing this we shall find the most favourable probability that the samples are in agreement. If this most favourable probability itself is found to be too small then we shall be justified in asserting differentiation. In other words we should chose the case most unfavourable for our conclusion, thus erring on the safe side.

I have also used the " four-fold X^2 method " of Pearson [1] as a check. In the present notation.

$$X^2 = \frac{(pr-qs).^2\ (m+n)}{n.\ m.\ (p+s)\ (q+r)}$$

and may be easily calculated. The probability of occurrence, P, is then found from Tables XII and XVII (pp. 26, 31) of Tables for Biometricians and Statisticians. Q gives the probability on Bayes' theorem.

Chalisa : Katas

Q = ·6286	...	...	...	6 to 10 in favour of agreement.
P = ·312	...	...	...	1 to 3 in favour of agreement.

There is no evidence of differentiation. The probability is that both represent the same content in nature.

Chalisa : Kallar Kahar

Q = ·0927	...	...	...	94 to 10 against agreement.
P = ·2271	...	...	...	44 to 10 against.

Differentiation is not improbable but the evidence is not conclusive.

Chalisa : Sardhi

Q = ·0047	...	...	...	...	212 to 1 against.
P = $2{\cdot}55 \times 10^{-5}$		...	...	...	40,000 to 1 against.

Differentiation is probable ; but the smallness of the number (2 only) in the case of Chalisa renders the conclusion somewhat doubtful.

Katas : Kallar Kahar

Q = ·00065	..	...	...	...	1528 to 1 against.
P = ·00630	...	...	..	...	157 to 1 against.

It seems fairly certain that Katas has got a higher proportion of dextral shells than Kallar Kahar.

Kallar Kahar and Sardhi

Q = ·000177	...	...	...	...	5662 to 1 against.
P = ·007083	...	...	...	...	140 to 1 against

These also seem to be quite significantly differentiated.

---

[1] *Draper's Company Research Memoirs, Biometric Series VIII.* " On a Novel Method of regarding the Association of two varities classed solely in Alternate Categories."

Katas : Sardhi

$Q=5{\cdot}5{\cdot}10^{-6}$ ... ... 180,800 to 1 against.

$P=1{\cdot}6{\cdot}10^{-9}$ ... ... 10^8 to 1 against.

The probability is now overwhelming that the samples are different. Summing up we may say: Chalisa is uncertain; Katas shows distinctly higher proportion of dextral shells than either Kallar Kahar or Sardhi, while Kallar Kahar itself shows significant excess over Sardhi.

The following table will give some idea about the reliability of the observed proportion in each case :—

TABLE III.—*Percentage frequency of dextral shells in samples of the same size m.*

1. *Chalisa m= 2*		3. *Kallar Kahar m= 70*		4. *Sardhi m= 48*		*Total m= 132*	
0	10·00	10	0·679	0	13·350	22	0·9137
1	30·00	11	1·087	1	19·143	23	1·2244
2	60·00	12	1·634			24	1·5927
		13	2·316	2	19·349	25	2·0137
		14	3·113			26	2·4781
		15	3·985	3	16·124	27	2·9713
2. *Katas m= 12*		16	4·874	4	11·960	28	3·4571
		17	5·711	5	8·186	29	3·9682
		18	6·430	6	5·273	30	4·4278
2	0·084	19	6·970	7	3·236	31	4·8319
3	0·302	20	7·287	8	1·962	32	5·1606
4	0·866			9	1·084	33	5·3979
5	2·079	21	7·360	10	0·597	34	5·5332
6	4·312						
7	7·854	22	7·193			35	5·5616
8	12·622	23	6·809				
9	17·764	24	6·251			36	5·4843
		25	5·569			37	5·3085
10	21·317	26	4·819			38	5·0457
		27	4·054			39	4·7119
11	20·348	28	3·316			40	4·3244
12	12·435	29	2·638			41	3.9019
		30	2·043			42	3·4626
		31	1·540			43	3·0230
		32	1·131			44	2·5971
		33	0·808			45	2·1962
						46	1·8285
						47	1·4992
						48	1·2107
						49	0·9632

Taking 1% (one per cent of the samples) as a limiting value, we find the range to be 0 to 2 (*i.e.* 0% to 100%) dextral shells in the case of Chalisa, 5 to 12 (*i.e.* 42% to 100%) in the case of Katas, 11 to 32 (16% to 46%) in the case of Kallar Kahar, 0 to 9 (0% to 19%) in the case of Sardhi and 23 to 48 (17% to 29%) in the case of the total sample.

Finally we may compare each sample with the three others taken together.

		Dextral.	*Sinistral.*	*Total.*
1. Others	...	33	97	130
Chalisa	...	2	0	2
		35	97	132

$P=\cdot0678$. The evidence is inconclusive.

2. *Katas and others*

Others	...	25	95	120
Katas	...	10	2	12

$P = 2 \cdot 38 \times 10^{-5}$ Katas seems to be really differentiated from the rest.

3. *Kallar Kahar and others*

Kallar Kahar	...	21	49	70
Others	...	14	48	62

$P = \cdot 1863$ Kallar Kahar agrees well with the rest of the sample and may be considered fairly typical.

4. *Sardhi and others*

Sardhi	...	2	46	48
Others	...	33	51	84

$P = 2 \cdot 38 \times 10^{-5}$ Sardhi also is different from the rest.

Conclusion.—We may on the whole conclude that dextral shells occur in less frequency in nature, the proportion probably being roughly from a fifth to a third; that Kallar Kahar is a fairly typical sample, while Katas and Sardhi seem to be somewhat differentiated from the rest.