PAKISTAN JOURNAL

OF

SCIENTIFIC AND INDUSTRIAL RESEARCH

Vol. 3, No. 4 October 1960

# INVESTIGATIONS ON THE STATISTICAL FREQUENCY DISTRIBUTION OF SOME PHYSICAL CHARACTERISTICS OF COTTONSEED KERNELS

S. HILAL AHMAD* AND MAZHAR M. QURASHI

*Physical Research Division, Central Laboratories, Pakistan Council of Scientific and Industrial Research, Karachi*

(Received June 15, 1960)

Earlier measurements of the statistical distribution of the density of kernels of L.S.S. variety of cottonseed are extended to other varieties and also to cover other characteristics, namely, weight, oil content, and percentage oil content. The resulting distributions show an interesting bimodal character with 25% correlation on the average between weight and density. The percentage oil content distribution shows a ±2% variation about the mean and a possible correlation with the weight distribution.

In an earlier communication[1] dealing with the examination of possible nondestructive methods for the estimation of oil content of a single cottonseed, an exploratory study was made on the statistical distribution of the density of cottonseed kernels of the L.S.S. variety. Although it appeared that the measurement of density was not a suitable basis for the estimation of oil, its statistical distribution (Fig. 1, top) was remarkable in that it is distinctly bimodal, i.e. possesses two maxima. It was therefore considered worthwhile to extend this study to some other varieties and also to cover other characteristics of the kernels, such as weight, volume, oil content, etc.

## 1. Density Measurements on Kernels of Several Varieties

Because it had been found in the earlier measurements that even a sample of 25 kernels gave a fair indication of the phenomena being studied, it has been considered sufficient to use two lots of 25 seeds from each variety examined, and to estimate the standard deviations from the differences between the results for the two lots. Two random lots of 25 kernels were taken from each variety, only the obviously dried and shrivelled up kernels being rejected. The density was measured (as in the earlier work) with a pycknometer or specific gravity bottle, and the measurements were completed within a few days of dehulling; during this period the humidity ranged from 50% to 70%. A reproducibility of ± 0.3 mg. was attained in all the weighings, so that, for kernels with a mean weight of about 50 mg., the calculated values of specific gravity are reliable to ±1%, i.e. to ±0.01.

The results obtained with three varieties of Pakistani cottonseed, namely L.S.S., M4, and 4F, are plotted as frequency distributions in Fig. 1, using intervals of 0.05 in density. The short vertical lines through the plotted points represent the standard deviations as estimated from a comparison between the two lots of 25 kernels for each variety. In order to diminish the drawing errors, and to obtain smooth frequency distribution curves, it was found advisable to calculate the points near the middle with two sets of overlapping interval for the distributions, and this extra set of points is shown by hollow circles in Fig. 1. Using all the points thus plotted, very satisfactory curves can be drawn with an estimated error of less than ±3 units on the percent frequency scale.

The graphs of Fig. 1 all show a more or less bimodal type of distribution. The higher of the two peaks, i.e. the main maximum, occurs at a density centred around 1.04 in the three varieties,

*Now at P.A.F. School, Korangi Creek, Karachi.

while the secondary maximum appears to vary considerably both in position and height, being most prominent in the case of L. S. S. cottonseed kernels. This provides some evidence for the probable existence of two types of seeds within the one variety (L.S.S.), which has been previously believed to show large unexplained variations in its oil content.[2] The distribution for L.S.S. is further seen (Fig. 1, top; broken lines) to be capable of decomposition into two mutually similar distributions centred about the density values 1.08 and 0.97, each with a half-value width of 0.08 units of density.

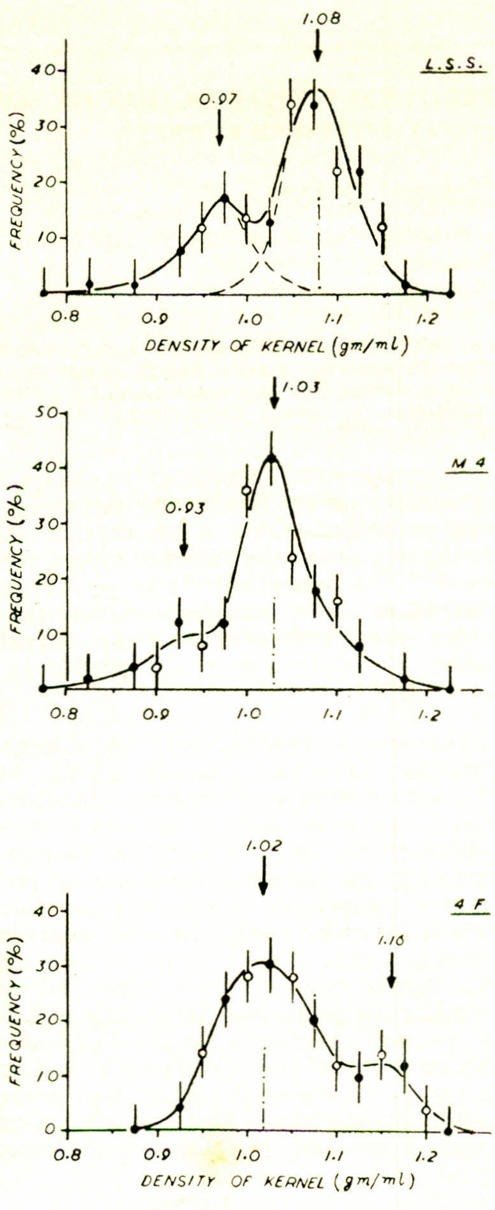


Fig.1.—Statistical frequency distribution graphs for density of kernels of three varieties of cottonseed, namely, L.S.S. (top), M4 (middle) and 4F (bottom). A density interval of 0.05 unit has been used (solid circles) and the hollow circles are intermediate points calculated for increased accuracy in drawing the graphs. The vertical lines through the points represent the estimated standard deviation.

## 2. Frequency Distribution Graphs for Weight of Kernel and their Correlation with Density

In an effort to interpret the above findings, similar distribution graphs of mass or weight have been plotted for the identical kernels used in the density measurements. Since the overall spread

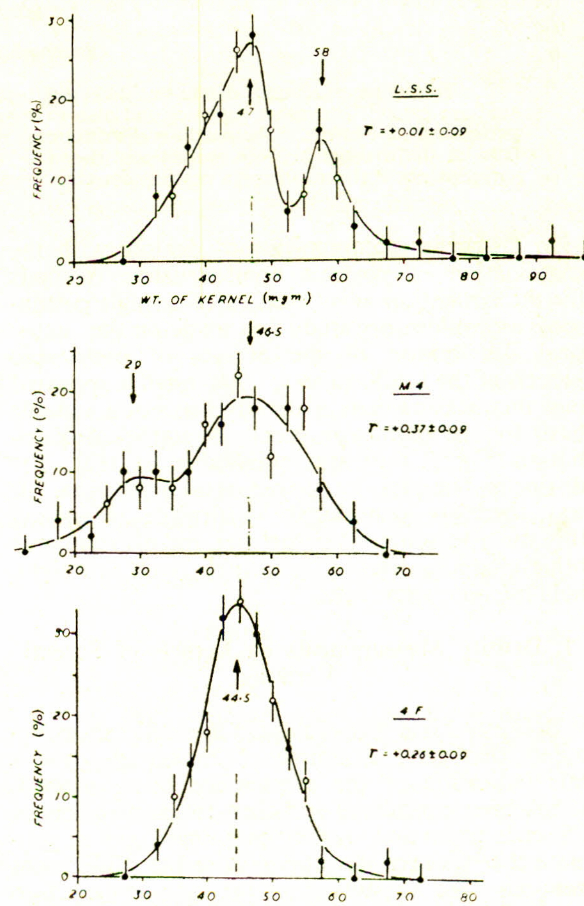


Fig. 2.—Corresponding frequency distribution graphs for weights of the kernels used in the density measurements for the three varieties: L.S.S. (top), M4 (middle) and 4F (bottom). The values of r, the correlation coefficient between weight and density, are given alongside the graphs.

of the weight for a kernel is roughly from 20 mg. to 90 mg., intervals of 5 mg. were chosen for calculating the distribution frequencies which are plotted in Fig. 2. As before, the short vertical lines indicate the estimated standard deviations, while the solid and hollow circles give the values calculated for two sets of overlapping intervals, which facilitate drawing of the smooth curves through the plotted points, giving an accuracy of ±2% frequency.

The distribution curve for the L.S.S. kernels is again found to possess a pronounced bimodal character with a deep minimum between the two peaks, while the graphs for the other varieties deviate but little from a simple nearly Gaussian distribution. Also, it is noteworthy that the major maximum occurs at almost the same weight for all the varieties, with a mean value of 46 ± 1 mg. The subsidiary maximum is, as before, variable in position as well as height.

Further light is thrown on the character of this subsidiary maximum in the graphs of Figs. 1 and 2 by calculating the coefficient of correlation 'r' between the data used for corresponding pairs of graphs. This was done by using the improved Spearman formula[3] for "Rank Correlation," viz.,

$$r = 1 - 6 \sum_{1}^{n} (x - y)^2 \Big/ (r^3 - n)$$

$$= 1 - \frac{6}{n} \sum_{1}^{n} (x - y)^2 \Big/ (n^2 - 1), \quad (1)$$

where 'x' and 'y' are the "ranks" of a given kernel with regard to the two properties, and 'n' is the total number of kernels examined. In order to obtain an estimate of the probable reliability of 'r', the total number of kernels of any one variety was (as before) divided into two groups of 25 each, and the value of 'r' was calculated separately for each group. In this way an overall standard deviation of ±0.09 was derived and the mean values of 'r' calculated from the two groups are given as follows :

$$r_{L.S.S.} = +0.01 \pm 0.09$$
$$r_{M4} = +0.37 \pm 0.09$$
$$r_{4F} = +0.26 \pm 0.09$$

A notable feature is that all these correlation coefficients are positive, and if we make allowance for the indication for abnormal behaviour of the (distribution curves for the) L.S.S. variety by giving it one half the statistical weightage of the others, we get as the weighed mean* value for r,

*It is to be noted that the unweighed mean is not very different, being equal to + 0.21 ± 0.09.

$$r_{mean} = +0.25 \pm 0.06 \quad (2)$$

This shows that there is on the average 25% correlation of the high densities with the larger kernels.

### 3. Investigation of the Correlation between Peaks

Using this mean correlation coefficient as a reference, we can now examine the individual cases of the three varieties. Firstly, we find that the 4F kernels yield a value of r about equal to the mean, and correspondingly we find no subsidiary peak in the weight distribution, and only a small one in the density distribution. On the other hand, the value of 'r' deviates most from the mean value of +0.25 in the case of L.S.S. kernels, which accordingly show the most marked subsidiary maxima in the pair of distribution graphs. Secondly, it appears that (i) the large *negative* deviation (from $r_{mean}$) in the case of L.S.S. seeds corresponds to the subsidiary maxima occurring on *opposite* sides of the main maxima in the density and weight distributions, and (ii) the positive deviation from $r_{mean}$ observed with M4 seeds is associated with these subsidiary peaks being on the *same* side of the main maxima in the two paired distributions. This suggests a strong positive correlation between the two properties of weight and density for the kernels corresponding to these subsidiary peaks. Numerically, the additional area under the subsidiary peaks in the case of L.S.S., is about 0.36±0.02 of the total area under the distribution curve, while $(r_{mean} - r_{L.S.S.}) = 0.24$, thus indicating a (positive) correlation of 0.24/0.36=67%. This result also fits the distributions for M4 seeds, where the subsidiary peaks account for about 0.22±0.09 of the area and for which $(r_{M4} - r_{mean}) = 0.12$. These data together lead to a mean figure of 60±6% for the correlation between the pair of peaks. It is hoped to check on this conclusion by measurements on kernels of other varieties in future.

### 4. Statistical Distribution of Oil Content in L.S.S. Kernels

In the light of the above findings, an attempt was next made to examine the statistical distribution of oil in individual kernels of L.S.S. cottonseeds. For this purpose, we used the filter paper technique,[4] based on expression of the oil between a sandwich of filter papers, which has been developed for this very type of work. Each kernel was weighed, then placed in a sandwich of filter discs, two on either side, and the sandwich pressed at 7 tons between the plates of a 10-ton laboratory press. The area of the oil spot was determined by measuring two mutually perpendicular diameters of the spot on each of two filter papers. The results obtained with 23 L.S.S. kernels are plotted as frequency distribution graphs for area of oil spot and weight of kernel, in Fig. 3 (top and middle, respectively) where the short vertical lines denote the estimated standard errors of the plotted points. While the mass distribution is very similar to the previously found bimodal one in Fig. 2, the distribution of area of oil spot, which is proportional to the total oil in the kernel, is strikingly different.

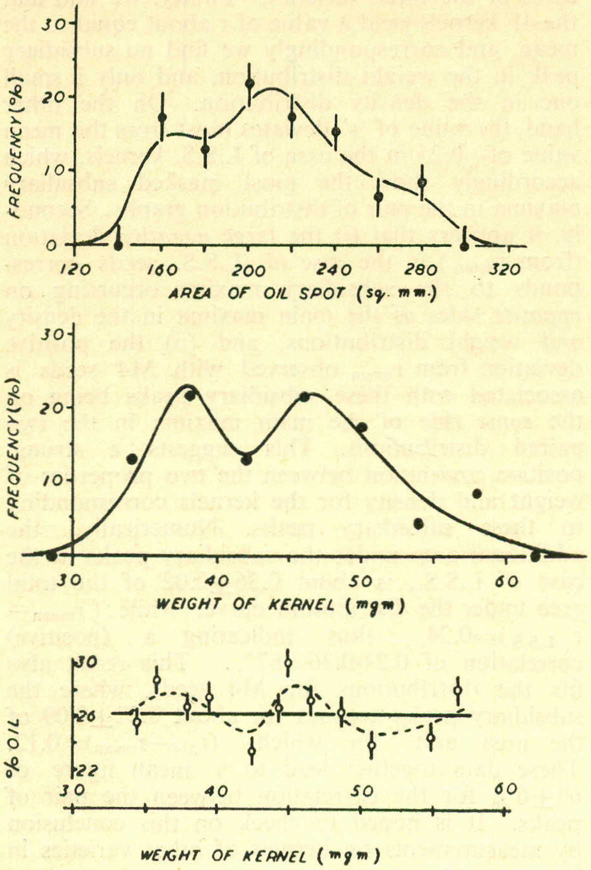


Fig. 3.—Graphs showing the frequency distribution of of (i) oil content measured as area of oil spot (top), and (ii) weight of kernel (middle), together with the percentage oil content against weight of kernel (bottom) for a sample of L.S.S. cottonseed kernels. The broken line shows the best curve through the plotted points in the bottom graph.

In an effort to elucidate this difference, the deduced percentage oil content (estimated[4] as 5.6 $\times$ area of spot/mass of kernel) has been plotted against the weight of the kernels in Fig. 3 (bottom), in which the points are means for two kernels, and the vertical lines indicate the standard deviation as estimated from the departure about these means. Although the broken line curve, which gives the best graph through the points, is seen to follow the general trend of the mass distribution curve in the middle of Fig. 3, it is noteworthy that the horizontal line drawn through the experimental points and shown by the full line is also a valid interpretation of the present data, because it is cut by 40% of the vertical lines and missed by 60%. It can therefore be concluded from Fig. 3 that (i) the variation in percentage oil content from kernel to kernel is at most of the order of $\pm 2\%$ on the weight of the kernel, and (ii) more accurate and extensive measurements are necessary before any correlations of percentage oil content with other constants of the seeds can be ascertained. Further studies in this direction are planned.

**Acknowledgement.**—The authors are grateful to Dr. M. Kamal, Mr. M.A. Ansari and members of the staff of the Cotton Research Institute for supplying samples of the different varieties of cottonseed and for useful discussions.

## References

1. S. H. Ahmad and M. M. Qurashi, Pakistan J. Sci. Ind. Research, **2,** 223 (1959).
2. Private communication from the Pakistan Cotton Committee.
3. H. Jeffreys, *Theory of Probability* (Clarenden Press, Oxford, 1948), second edition, p. 204.
4. (a) S. N. Ahmad and M. M. Qurashi, Pakistan J. Sci. Ind. Research, **2,** 13 (1959).
   (b) M. M. Qurashi, S. N. Ahmad, S. A. K. Lodhi, ibid., **3,** 87 (1960).