Peanut Science (1978) 5, 91-96

Probability Distributions of Peanut Seed Size

James I. Davidson, Jr.,* Paul D. Blankenship[1], and Victor Chew[2]

ABSTRACT

Procedures and mathematical relationships were developed to describe seed size distributions for Florigiant, Florunner, and Starr peanut (*Arachis hypogaea* L.) varieties. Of six standard probability distributions studied, the normal and logistic distributions provided the best fit for the experimental data. These two distributions were therefore fitted to seed size data for several lots of peanuts. For each lot both the normal and logistic distributions provided an excellent fit to the experimental data, but the logistic was slightly superior. Differences between experimental and calculated values were greatest for lots that were the least or most mature. A logistic distribution was also fitted to the average of all data for each variety. These relationships may be used to better relate seed size to quality, marketing, shelling, and processing. They will also be useful in research studies of the effects on seed size of such variables as variety, agronomic practices, climate, soil moisture, and harvest dates.

Key Words: quality, marketing, processing, shelling, seed size, maturity, screening.

Seed size [3] is very important to the peanut (*Arachis hypogaea* L.) industry because it is used to indicate quality and to determine the market value of peanuts (6). Studies conducted during the past 8 years (3, 4, 7) have revealed consistent and definite characteristics in seed size distributions for peanuts grown according to practices recommended by the Cooperative Extension Service. These results indicated that mathematical relationships could be developed to describe seed size distributions.

Mathematical relationships are needed to quantify seed size. These relationships can be used for determining quality and market value of peanuts; for design and operation of shelling and processing equipment; and for studies to determine the effects on seed size of such variables as variety, agronomic practices, climate, soil moisture, and harvest dates.

The purpose of this study was to develop procedures and mathematical relationships for describing seed size distributions of Florigiant, Florunner, and Starr peanut varieties.

The peanuts from Tifton, Georgia, were produced by the Harvesting and Processing Unit, USDA, SEA, FR to determine the physical and shelling properties of several peanut varieties. Seed size distribution data relative to developing the mathematical relationships are presented in this report. Results of the larger study will be published at a future date.

Materials and Methods

Florigiant, Florunner, and Starr, three peanut cultivars, were grown, harvested, and cured by practices recommended by the Cooperative Extension Service. The peanuts were grown in Georgia without irrigation except for one lot of Florunner and one lot of Starr which were grown in Texas with irrigation.

A total of 20 lots (approximately 7 lots per cultivar) were obtained over a period of about 8 years. Lot sizes ranged from about 100 pounds (45.4 kg) from experimental plots to 40 tons (36 tonnes) from field plots. A farmers stock divider was used to remove one 1-10 pound (0.454-4.54 kg) sample from each of the small lots. An automatic spout sampler was used to obtain a 10-50 pound (4.54-22.68 kg) sample from each of the large size lots.

Each sample was handshelled or shelled with the official grade sheller (5). Split kernels (usually less than 3 percent) and hulls were weighed and discarded. The whole kernels were screened over slotted hole screens as described in a recent report (4). The kernel moisture content during screening was 7.0 percent wet basis for all lots.

The screening operations were conducted in the following sequence:

1. Approximately eight slotted screens were stacked on the official grade shaker with the narrowest slot on the bottom and the widest slot on the top. Each screen had slots 2/64 in. (0.08 cm)[4] wider than the slots in the screen immediately below.

2. Part of the seed was weighed and poured on the top screen (being careful not to overload any screen).

3. The shaker was operated for 20 seconds.

4. Each screen was removed from the shaker (top to bottom) and shaken by hand (over a pan) to insure that all seed had been exposed to the slotted openings. Seed that fell through each screen (into pan) were poured onto the next smaller screen.

5. Seed that rode each screen and those that fell through the bottom screen were weighed, and the weights were recorded.

6. Steps 1-5 above were repeated until all seed had been screened.

For verification that the screening method was providing an accurate measurement of seed thickness, the seed thickness of a 1-pound (454 gm) sample for each variety was measured with a micrometer, and the sizing data were compared with those obtained by screening. Seed thickness was obtained by measuring the maximum minor diameter on a plane perpendicular to the longitudinal axis of the seed.

The percent of kernels by weight, that rode each screen was calculated and plotted to provide the seed size distribution. Definite characteristics of seed size distribution plots (such as symmetry and peakness) were used to select the probability density functions that had the greatest potential for fitting the experimental data. The best fit was the probability density function that provided the lowest absolute deviation (difference) between the experimental and calculated values.

The cumulative probability function was also developed for the best fit to allow a wider application of the research results.

[1]Mechanical Engineer and Agricultural Engineer, respectively, U. S. Department of Agriculture, Science and Education Administration, Federal Research, Southern Region, Georgia-South Carolina Area, National Peanut Research Laboratory, Dawson, Georgia 31742.

[2]Mathematical Statistician, U. S. Department of Agriculture, Science and Education Administration Federal Research, Southern Region, University of Florida, Gainesville, Florida 32611.

[3]In this report seed size is considered to be synonymous with seed thickness.

[4]Screen sizes will be presented in the English units for the benefit of industry users.

Results

Seed size distributions determined by the screening procedure were in good agreement with the distributions obtained from micrometer measurements; see Table 1. (Regression analysis of the screening and micrometer data showed that these were consistent with a true slope of unity and an intercept of zero). The screening procedure permitted rapid evaluations of relatively large samples and was preferred over the micrometer measurement method.

Table 1. Performance of slotted-hole vibrating screens in sizing peanut seed.

Seed size range 1/ (64th in.)	Percentage of seed (by weight) 2/ Florigiant		Florunner	
	Screening	Micrometer	Screening	Micrometer
< 14	1.3	1.4	0.6	0.6
14-16	1.4	1.5	1.9	2.1
16-18	7.6	7.4	4.3	3.8
18-20	15.2	15.6	14.9	15.4
20-22	33.2	30.2	38.6	35.2
22-24	24.9	29.1	26.0	25.8
24-26	15.3	12.3	13.4	14.8
26-28	1.2	2.4	0.3	2.3

1/ 1/64 in. = 0.04 cm.

2/ One 1-pound sample was used for each comparison.

Seed size data for the three peanut varieties are presented in Appendix Tables 1, 2, and 3. These data indicated that the distributions were approximately symmetrical. The probability density function of six symmetrical distributions (1), uniform, triangular, Laplace, logistic, cosine, and normal were fitted to the average seed size data. The shape of the uniform and triangular distributions were not characteristic of the data plots, and peaks of the Laplace and cosine distributions did not fit the data plots. The logistic and normal distributions had the best potential for fitting the data. These distributions were selected for more detailed study.

The probability density function for the normal distribution is:

$$y = [100i/\,(\sigma\sqrt{2\pi})]e^{-(x-\mu)^2/(2\sigma^2)} \quad [1]$$

where y = percentage of kernels, by weight (theoretical)

i = width of sizing interval in 64ths of an inch (0.04 cm) = 2

f = percentage of kernels, by weight (experimental)

σ = population standard deviation, estimated by the sample standard deviation(s)

$s = \sqrt{[\Sigma fx^2 - (\Sigma fx)^2/\Sigma f]/[(\Sigma f)-1]}$

e = base of natural log = 2.7183

x = seed size in 64ths of an inch

μ = population mean estimated by sample mean $\bar{x} = \Sigma fx/\Sigma f = \Sigma fx/100$ Equation [1] will be estimated by

$$y = [(100i)/(s\sqrt{2\pi})]e^{-(x - x^2/(2s^2)}$$

$$= [(100)\,(i)/(s\ \sqrt{2(3.1416)})]e^{-Z^2/2}$$

$$= (39.8942i/s)e^{-Z^2/2} \quad [2]$$

where $Z = (x - \bar{x})/s$ and $\bar{x} = \Sigma fx/100$

Similarly the probability density function for the logistic distribution is given by

$$y = 100i/[2.2054\sigma \cosh^2[(x-\mu)/1.1027\sigma]] \quad [3]$$

where $\cosh\ [(x-\mu)/1.1027\sigma] =$

$$[e^{(x-\mu)/1.1027\sigma} + e^{-(x-\mu)/1.1027\sigma}]/2$$

Equation [3] for the logistic density function will be estimated by

$$y = 100i/[(2.2054s)\cosh^2[(x-\bar{x})\,/\,(1.1027s)]]$$

$$= (100)\,(i)/[2(1.1027s)\cosh^2(Z/1.1027)]$$

$$= 45.3432i/[s\cosh^2(Z/1.1027)] \quad [4]$$

Appendix Table 4 presents the procedure for fitting the normal distribution to the average data of Florunner peanuts. Appendix Table 5 presents the procedure for fitting the logistic distribution. Pertinent data for all lots are presented in Table 2. Plots of the density function for the average data of each variety are presented in Figs. 1, 2, 3 and 4.

The normal and logistic distributions provided excellent fits to the experimental data for each lot, as well as for the average data for all lots of each variety. Fits were better for the Florunner and Florigiant varieties than for the Starr varieties, probably because of poorer screening efficiencies associated with the narrower size range (steeper density plots) of the Starr peanuts and the assumption that all Starr seed that fell through a 14/64 in. (0.556 cm) slotted screen had an average size of 13/64 in. (0.516 cm). Use of 12/64 in. (0.476 cm) and 10/64 in. (0.397 cm) screens would undoubtedly improve the fit for the Starr peanuts. Within each variety the logistic distribution generally provided a better fit than the normal. For the Florigiant in four of six instances, the logistic distribution provided a better fit than the normal, even though the average data indicated the normal fitted equally as well.

The Florunner and Florigiant varieties had about the same seed size distribution. The longer seed of the Florigiant accounted for the differences in seed count per unit weight normally obtained for the two varieties. Because of fruiting characteristics, the Starr variety had a narrow range of seed size and displayed a steeper density plot than the Florunner or Florigiant.

Table 2. Summary of calculated values for fitting normal and logistic density functions to seed size data.

Lot no.	Mean seed size ($\bar{x}$) (64th in.)	Standard deviation(s)	Average absolute deviation Normal	Logistic
		Florigiant		
1966-4 1/	20.215	2.916	2.86	2.46
1967-4	19.748	2.325	0.71	1.01
1969-4	20.638	2.331	1.72	1.36
1973-1 2/	22.355	2.404	1.99	2.20
1974-1	21.100	2.651	1.86	1.82
1976-1	20.385	2.728	1.69	1.42
Composite 3/	20.750	2.696	1.22	1.22
		Florunner		
1972-1	21.060	2.101	0.91	1.08
1972-2	21.188	2.518	1.98	1.12
1972-3	20.686	2.276	1.19	0.62
1973-1 2/	21.442	2.346	2.35	1.66
1974-1	20.646	2.775	1.74	1.21
1975-1	20.760	2.427	1.96	1.15
1976-1 1/	19.716	2.523	2.49	1.58
Composite 4/	20.801	2.400	1.45	1.00
		Starr		
1969-4	17.992	1.959	2.23	2.33
1970-4 1/	17.740	1.965	3.67	3.53
1972-1	18.698	1.813	1.53	1.50
1972-3	18.654	1.754	2.70	1.88
1973-1	18.939	1.650	2.17	1.51
1974-1	17.919	1.765	4.02	3.52
1976-1	17.586	1.841	3.67	3.05
Composite 5/	18.221	1.895	2.85	2.27

1/ Least mature.

2/ Most mature.

3/ Calculated for data in last column of Appendix Table 1.

4/ Calculated for data in last column of Appendix Table 2.

5/ Calculated for data in last column of Appendix Table 3.

The maturity of the seed affected size distribution. The mean size of lots of mature seed was generally larger than that of lots of less mature seed, and the average standard deviation was smaller (Table 2). The average absolute deviation between the normal and logistic values was generally greatest (poorest fit) for the least and most mature lots. For the individual seed lots most of the average absolute deviation appeared to result from a skewness of the data to the right for the most mature lots and a skewness to the left for immature lots.

No consistent correlation of sample size to standard deviation or to absolute deviation was found. This

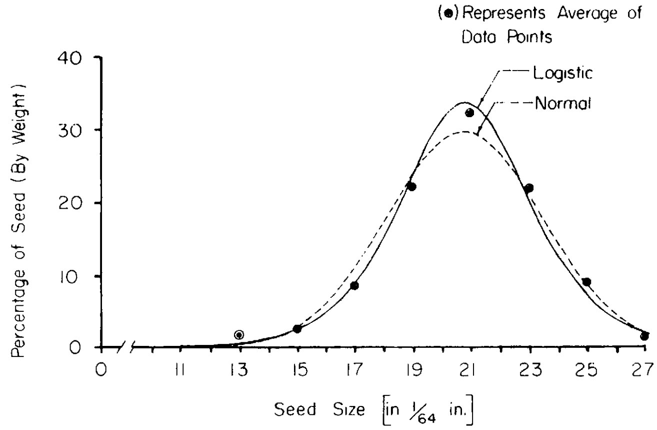

Fig. 1. Probability density curves for average seed size distribution of Florigiant peanuts.

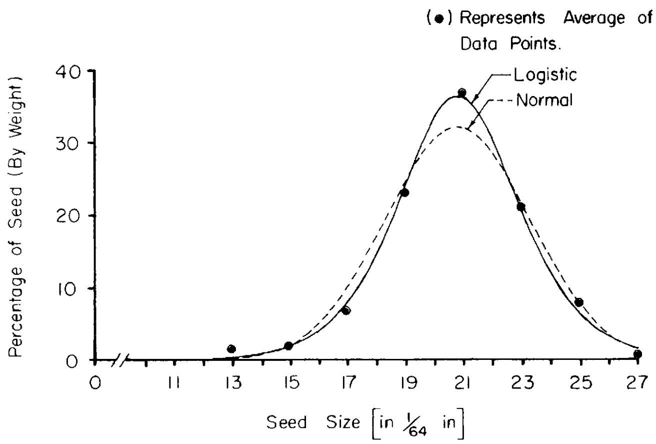

Fig. 2. Probability density curves for average seed size distribution of Florunner peanuts.

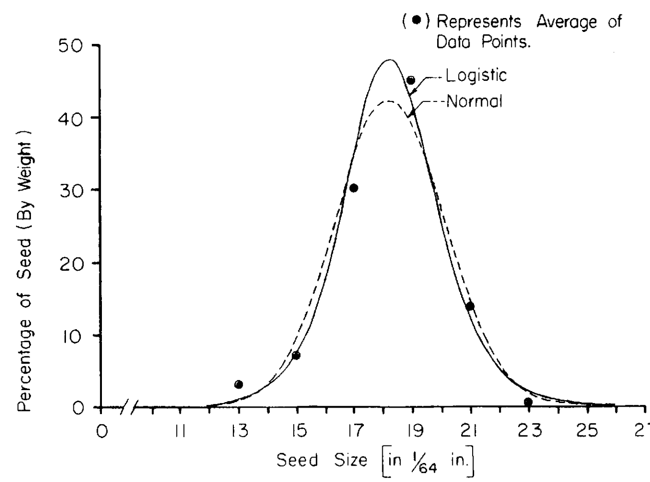

Fig. 3. Probability density curves for average seed size distribution of Starr peanuts.

indicated that sampling methods provided a representative sample for each lot.

Although the probability density function is useful in describing the seed size distribution, the cumulative probability function is more convenient for many practical applications, such as determining the percentage of seed that will ride or fall through a particular size screen. The cumulative function for the logistic distribution is

$$y = 100/[1 + e^{-1.8137 (x - \bar{x})/s}] = 100/[1 + e^{-1.8137 Z}] \quad [5]$$

In this form y is the percentage of seed, by weight, that fell through a particular size screen, and 100 -y = y′ is the percentage of seed that rode a particular size screen. The cumulative curves for the average percentage of seed that rode a certain size screen (y′) for each variety are presented in Fig. 5. The equations for these curves are listed as follows:

$$\text{Florigiant} \; — y' = 100[1 - 1/(1 + e^{-0.673(x - 20.75)})]$$

$$\text{Florunner} \; — y' = 100[1 - 1/(1 + e^{-0.756(x - 20.801)})]$$

$$\text{Starr} \; — y' = 100[1 - 1/(1 + e^{-0.957(x - 18.221)})]$$

For varieties or lots with seed size distributions similar to the ones presented herein, screening information can be extracted from the tables or cumulative curves or the information may be obtained by calculations. For lots that have different seed size distributions than those presented here, the seed size distributions may be developed as described herein.

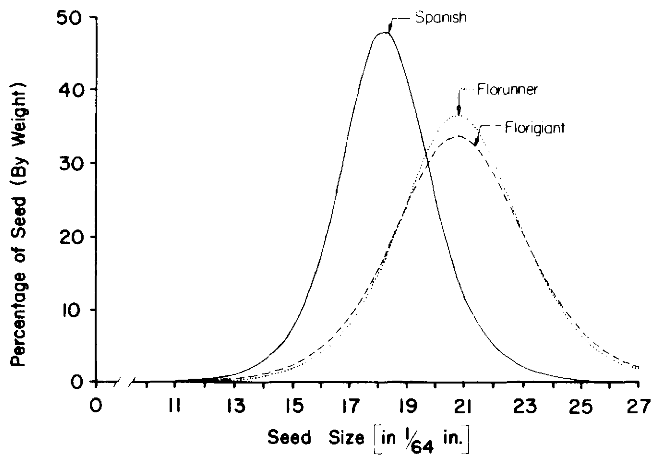

Fig. 4. Logistic density curves for average seed size distributions of Florigiant, Florunner, and Starr peanuts.

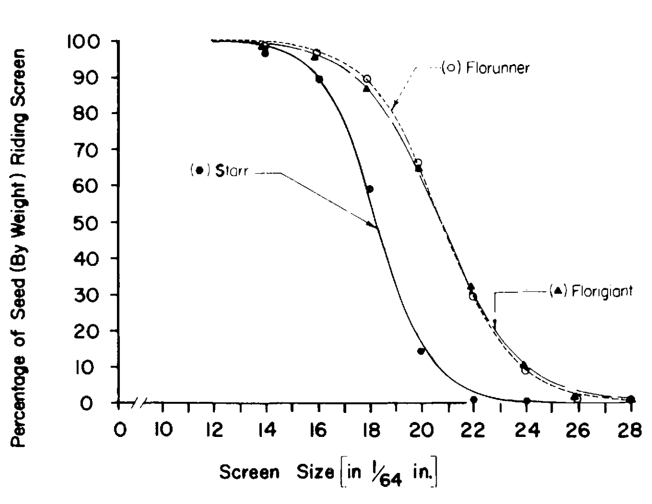

Fig. 5. Logistic probability cumulative curves for average seed size distributions of Florigiant, Florunner and Starr peanuts.

Discussion

In recent years, the industry has often requested the development of better methods for correlating seed size to quality factors such as maturity and taste. Considerable interest has also been expressed in the development of scientific procedures for specifying screen sizes by which the different market grades can be obtained.

Mathematical relationships that describe seed size distributions are beneficial in the above applications, especially if the indication of a relationship between maturity and seed size distribution characteristics (deviation and skewness) are verified. Perhaps the seed size distribution characteristics and difference in seed count per unit weight (4) will provide accurate estimates of market quality.

In addition, the mathematical relationships will be useful in determining screening requirements and equipment performance for sizing shelled peanuts in shelling and processing plants. Researchers will also find the relationships useful in their studies to better relate seed size to variables such as variety, agronomic practices, climate, soil types, soil moistures, and harvest dates.

The most important requirement in fitting seed size distribution data to the probability function is to obtain a representative sample of farmers stock peanuts and shell them to obtain a maximum whole seed outturn. The large seed split frequently during the shelling process, and splitting could result in a biased sample. The handshelling of samples larger than 1 pound (454 gm) is often impractical. In such instances, splitting can best be minimized by using gentle drying treatments and then shelling the peanuts at relatively high moistures (2) with the sheller used by the Federal State Inspection Service. Before the seed are screened, moisture should be gradually reduced to 7 percent wet basis to minimize errors resulting from the effects of moisture on seed size (8).

Conclusions

1. Seed size (thickness) distributions can be accurately obtained by exposing representative samples of seed to slotted-hole vibrating screens.
2. Seed size distributions for Florigiant, Florunner, and Starr peanuts can be adequately described by logistic and normal probability functions.
3. Characteristics of the seed size distributions are related to variety, and probably maturity and other important variables.

Literature Cited

1. Chew, V. 1968. Some useful alternatives to the normal distribution. The American Statistician, June p. 22-24.
2. Davidson, J. I. Jr., P. D. Blankenship, and R. S. Hutchison. 1970. Shelling and storage of partially dried (cured) peanuts. J. Amer. Peanut Res. and Educ. Assoc., Inc. 2 (1): 57-64.

3. Davidson, J. I., Jr., R. O. Hammons, W. Dillard, and R. B. Moss. 1973 Selected physical, shelling, and germination properties of the new Spanish peanut varieties "Spancross" and "Tifspan." ARS-S-24, 11 p.

4. Davidson, J. I. Jr., V. Chew, and R. O. Hammons. 1976. An objective method for evaluating size and shape characteristics of peanut seed. J. Amer. Peanut Res. and Educ. Assoc., Inc. 8(1):27-43.

5. Dickens, J. W. and D. D. Mason. 1962. A peanut sheller for grading samples: An application of statistics in design. Trans. of the Amer. Society of Agric. Engineers. 2(11):42-45.

6. Hinds, M. K. and G. W. Kromer. 1973. Peanut Marketing, pp. 657-677. In: Peanuts-Culture and Uses. Published by Amer. Peanut Res. and Educ. Assoc., Inc. Stillwater, OK.

7. McIntosh, F. P. and J. I. Davidson, Jr. 1971. Selected physical and shelling properties of Florunner peanuts. ARS 52-68. 24 p.

8. Slay, W. O. 1976. Dimensional changes in peanut pods, kernels, and hulls as moisture is removed during drying. ARS-S-81, 8 p.

Appendix Table 1. Seed size distribution of several lots of Florigiant peanuts.

Seed size 2/ (64th in.)	Percentage (by weight) of seed in each size range						
	Lot 1966-4	Lot 1967-4	Lot 1969-4	Lot 1973-1	Lot 1974-1	Lot 1976-1	Average
13	4.6	0.3	0.9	0.0	1.7	3.1	1.9
15	4.6	4.1	2.3	1.1	2.3	2.7	2.8
17	9.5	15.0	8.9	3.3	7.2	9.2	8.8
19	22.6	32.6	22.7	9.4	18.3	27.7	22.2
21	31.6	32.0	37.7	30.2	31.4	30.6	32.2
23	20.6	12.6	22.4	29.0	28.0	18.6	21.9
25	5.9	2.3	4.8	23.4	9.6	7.4	8.9
27	0.7	0.1	0.3	3.8	1.4	0.8	1.2

1/ The first four numbers of the lot number represent the crop year. The last number represents the growing location (1 - Tifton, Georgia, 4 - Dawson, Georgia).

2/ The seed size was considered to be the midpoint of the range (e.g. 14/64 in. - 16/64 in. = 15/64 in.). Also the fall through the 14/64 in. slotted-hole screen was considered to be in the size range of 12/64 in. - 14/64 in. (seed size of 13/64 in.). 1/64 in. = 0.04 cm.

Appendix Table 2. Seed size distributions of several lots of Florunner peanuts.

Seed size 2/ (64th in.)	Percentage (by weight) of seed in each size range							
	Lot 1972-1	Lot 1972-2	Lot 1972-3	Lot 1973-1	Lot 1974-1	Lot 1975-1	Lot 1976-1	Average
13	0.4	1.3	0.8	0.6	2.7	1.6	3.7	1.6
15	0.9	2.1	1.5	1.9	3.2	0.9	4.0	2.1
17	5.3	5.6	7.8	4.3	7.6	6.8	11.4	7.0
19	21.2	17.6	25.4	14.9	23.6	26.3	33.2	23.2
21	40.0	36.1	38.5	38.6	33.1	37.6	32.2	36.6
23	25.8	26.3	19.4	26.0	19.5	17.9	12.4	21.0
25	6.4	9.3	6.2	13.4	8.9	7.6	3.1	7.9
27	0.1	1.6	0.4	0.3	1.4	1.3	0.0	0.7

1/ The first four numbers of the lot number represent the crop year. The last number represents location. (1 - Tifton, Georgia, 2 - McRae, Georgia, 3 - Commanche County, Texas)

2/ The seed size was considered to be the midpoint of the range (e.g. 14/64 in. - 16/64 in. = 15/64 in.). Also the fall through the 14/64 in. slotted-hole screen was considered to be in the size range of 12/64 in. - 14/64 in. (seed size of 13/64 in.). 1/64 in. = 0.04 cm.

Appendix Table 3. Seed size distributions of several lots of Starr peanuts.

Seed size 2/ (64th in.)	Percentage (by weight) of seed in each size range							
	Lot 1969-4	Lot 1970-4	Lot 1972-1	Lot 1972-3	Lot 1973-1	Lot 1974-1	Lot 1976-1	Average
13	3.5	4.8	1.5	1.7	1.1	3.9	6.0	3.2
15	10.5	12.6	4.2	4.8	2.8	7.0	7.6	7.1
17	32.4	32.1	26.1	23.2	18.5	35.5	43.5	30.2
19	40.6	41.4	46.1	50.1	53.2	46.5	36.9	45.0
21	12.5	8.7	20.5	19.8	21.8	6.8	6.0	13.7
23	0.5	0.1	1.7	0.4	1.3	0.2	0.0	0.7

1/ The first four numbers of the lot number represent the crop year. The last number represents the growing location (1 = Tifton, Georgia, 3 = Erath County, Texas, 4 = Dawson, Georgia).

2/ The seed size was considered to be the midpoint of the range (e.g. 14/64 in. - 16/64 in. = 15/64 in.). Also the fall through the 14/64 in. slotted-hole screen was considered to be in the size range of 12/64 in. - 14/64 in. (seed size of 13/64 in.). 1/64 in. = 0.04 cm.

Appendix Table 4. Calculations for fitting a normal curve to seed size distribution.

Average data for Florunner peanuts (see last column of Appendix Table 2)				
x	f	z	y	Absolute deviation |y -f|
13	1.6	-3.2501	0.17	1.43
15	2.1	-2.4169	1.79	0.31
17	7.0	-1.5836	9.49	2.49
19	23.2	-0.7504	25.09	1.89
21	36.6	0.0829	33.13	3.47
23	21.0	0.9162	21.85	0.85
25	7.9	1.7494	7.20	0.70
27	0.7	2.5827	1.18	0.48

Total absolute deviation = $\Sigma|y-f|$ = 11.60
Average absolute deviation = $\Sigma|y-f|$/no. of screens = 11.60/8 = 1.45

$$\bar{x} = \frac{\Sigma fx}{100} = \frac{2080.1}{100} = 20.801$$

$$s = \sqrt{\frac{\Sigma fx^2 - \frac{(\Sigma fx)^2}{100}}{99}} = \sqrt{\frac{43838.5 - \frac{4326816.01}{100}}{99}} = \sqrt{\frac{570.3399}{99}} = 2.4002$$

$$x_1 = 13, z_1 = \frac{x - \bar{x}}{s} = \frac{13. - 20.801}{2.4002} = -3.2501, \frac{z_1^2}{2} = \frac{(-3.2501)^2}{2} = 5.2817 \text{ 1/}$$

$$y_1 = \frac{79.7884e^{-5.2817}}{2.4002} = 0.17 \text{ 1/}, \; |y-f| = |0.17-1.6| = 1.43 \text{ 1/}$$

1/ Similar calculations were made for other values of x and the results entered in the above table.

Appendix Table 5. Calculations for fitting a logistic curve to seed size distribution.

Average data for Florunner peanuts (see last column of Appendix Table 2)				
x	f	z/1.1027	y	Absolute deviation |y -f|
13	1.6	-2.9474	0.41	1.19
15	2.1	-2.1918	1.84	0.26
17	7.0	-1.4361	7.66	0.66
19	23.2	-0.6805	24.55	1.35
21	36.6	0.0752	37.57	0.97
23	21.0	0.8308	20.27	0.73
25	7.9	1.5866	5.83	2.07
27	0.7	2.3422	1.37	0.67

Total absolute deviation = $\Sigma|y-f|$ = 7.90
Average absolute deviation = $\Sigma|y-f|$/no. of screens = 7.90/8 = 0.99

$\bar{x}$ = 20.801, s = 2.4002 (from Appendix Table 4)

x_1 = 13, z_1 = -3.250 (from Appendix Table 4)

$$y_1 = \frac{90.6865}{2.4002 \cosh^2\left(\frac{z_1}{1.1027}\right)}$$

$$\frac{z_1}{1.1027} = \frac{-3.250}{1.1027} = -2.9474 \text{ 1/}$$

$$\cosh^2(-2.9474) = [(e^{-2.9474} + e^{2.9474})/2]^2 = (9.5548)^2$$

$$y_1 = \frac{90.6865}{(2.4002)(9.5548)^2} = 0.41 \text{ 1/}$$

$$|y-f| = |0.41-1.6| = 1.19 \text{ 1/}$$

1/ Similar calculations were made for other values of x and the results were entered in above table.

Accepted June 1, 1978