A Study of the Variability Associated With Sampling Peanuts for Aflatoxin

PETER J. TIEMSTRA, Swift Grocery Products Co., Chicago, Illinois 60604

Abstract

One lot of peanuts known to contain aflatoxin was extensively sampled to study the sources of variability. A nested design was used where sections (50 bag units), subsamples and analytical variation were the variables studied. Sample size was the most critical factor in characterizing this lot. Variability from section to section was not significant indicating random distribution of the contaminant. Three 20 lb samples were taken on a number of lots from the 1968 crop, each of which was subdivided into two equivalent subsamples. The aflatoxin was determined on each of these subsamples. The results indicated that all the significant variation came from the subsamples, further enforcing the thesis that sample size is the critical factor in variability, and not lot inhomogeneity. Analysis of 550 lots from the 1967 crop where triple samples and analysis were available indicates that the magnitude of the variability is a little greater than was found on the experimental lot. Using the pooled standard deviation of the 1967 crop data, operating characteristic curves were plotted to demonstrate the improvement that can be expected by increasing the sample size.

Introduction

The wide variation of aflatoxin results obtained on independent samples from a given lot of peanuts have caused concern regarding the adequacy of a 10 lb sample to characterize the true aflatoxin content of one lot of peanuts. Ten pounds has been used as the standard sample size for aflatoxin assay. This study was undertaken to determine the magnitude of the sampling variation and to reconcile the experimental data with theoretical considerations. It is important to know the sampling variation in order to properly characterize peanut lots for their suitability in making manufactured products.

Whitaker (1) and the writer (2) have developed models based on three major premises. The first premise is that only a relatively few kernels in the lot are contaminated. The number of contaminated kernels as compared to the total number of kernels in the lot gives the expected ratio of contaminated kernels to sound kernels to be expected in the sample. If this number is very small, then there will be a great deal of variation in the number of contaminated kernels in the samples. Experimental evidence can give a clue to the relative level of contaminated kernels in the lot.

The second premise is that the level of aflatoxin varies greatly from one kernel to another in the contaminated fraction. From experimental assays on individual kernels (3), the level of aflatoxin in individual kernels that have had mold growing on them ranged from 0 to 300,000 ppb. In the models, this distribution has been assumed to be a hyperbolic function such as the negative binomial or exponential function.

The third premise is that the kernels are randomly distributed throughout the lot. Thus, every kernel has an equal probability of being in the sample. A way of handling nonhomogeneity is to take a good cross section sample such as a continuous sampler would give. An attempt to accomplish this is done by taking the sample from one quarter of the bags in the lot.

Experimental Procedures

We received an 800 bag, 100,000 lb lot of peanuts which had the following aflatoxin results on five independently drawn samples: 0, 200, 0, 13 and 63. This appeared to be a good lot upon which to conduct a sampling study. It was divided into 16 sections consisting of 50 bags per section. Four handfuls of peanuts were removed from each bag in the section and composited to give a sample of about 20 lb. This sample was passed over a riffle making two subsamples, A and B for each section. The subsamples were ground in a Dickens mill (4). The 2 lb discharge was intimately mixed and 100 g portions were drawn for analysis. Each of the subsamples was analyzed by the Best Foods procedure (5) and the CB procedure (6). The variation between methods was used as a measure of variation due to methodology and sample preparation.

The reader will recognize this as a nested design experiment where two major sampling problems were investigated. If there is nonhomogeneity in the lot, i.e., certain bags having more aflatoxin than others, then there would be a significant difference between the sections. However, if the model which is proposed is correct, then the size of the samples or the variations between any two samples, whether from the same section or not, should be the predominant source of variation.

The results of this experiment are shown in Table I. The variation was very great ranging from 0 to 336 ppb. In addition, Section 2 had an unusual spread between sample A and B, 27.5 to 311 ppb. The analysis of variance (ANOVA) is given in Table II. There is a significant variation due to subsamples and the difference between sections is not significant.

TABLE I
Aflatoxin Results on Experimental Lot

Sec.	Sample A Method			Sample B Method			Avg. (ppb)
	B.F.[a] (ppb)	C.B.[b] (ppb)	Avg. (ppb)	B.F.[a] (ppb)	C.B.[b] (ppb)	Avg. (ppb)	
1	22	36	29	9	26	18	23.3
2	20	35	28	336	286	311	166.8
3	65	62	64	30	31	31	47.0
4	8	16	12	11	26	19	15.3
5	10	11	11	20	89	55	32.5
6	24	36	30	20	89	55	42.3
7	27	36	32	42	61	52	41.5
8	32	47	40	10	0	5	22.3
9	49	78	64	18	15	17	40.0
10	16	42	29	10	8	9	19.0
11	19	11	15	31	26	29	21.8
12	42	52	47	39	52	46	46.2
13	57	66	62	10	19	15	38.0
14	20	11	16	30	56	43	29.3
15	10	21	16	45	103	74	44.8
16	15	37	26	40	76	58	42.0
							42.1

[a] Best foods method.
[b] C.B., procedure of Eppley.

TABLE II

Analysis of Variance, Experimental Lot

Source	df	Mean square	F
Total	63		
Sections	15	5043.26	< 1 n.s.[a]
Samples	16	5997.95	17.4[b]
Methods	32	343.98	

[a] n.s. not significant.
[b] Significant at 99% level.

Therefore, the variation as explained by the models appears to be justifiable and there is no evidence of nonhomogeneity.

One very important aspect of these data is the relative magnitude of the standard deviation due to sampling versus that due to the preparation of the sample and analysis. The standard deviation, due to sample preparation and analysis, is 18.5 or a C.V. of 0.44 (44%). This agrees with data obtained on recent collaborative studies where interlaboratory variation was measured (7). The standard deviation of sampling (sample size) is 53.2 or nearly three times that of the analytical deviation. The overall standard deviation is 56.1. Reducing the sample preparation and analysis variation by half to 9.3 would reduce the overall deviation to 54.0 or a difference of only 2.1 units or 3.7%. Reducing the sampling deviation by half to 26.6 would reduce the overall deviation to 32.4 or a difference of 23.7 units or 42.2%. Therefore, the most critical problem in characterizing a lot of peanuts for its aflatoxin content is in improving the sample. Basically, taking a sample four times as big as the 10 lb sample should reduce the sampling deviation by half. A 90 lb sample will reduce the sampling variation to a third. However, caution has to be used in extending this reasoning, since the sample preparation and analytical variation is normal, whereas the sampling variation is skewed and can be best described by a log normal distribution. Therefore, it is impossible to quantitatively express these two types of variation in terms that can be directly compared. The above was done on the assumption that both deviations were normal. The sampling deviation under such techniques is usually inflated over what it should be. Nonetheless, the ANOVA of these data does highlight the importance of the sampling problem.

The damage portion and the rancid, moldy and decayed (RMD) fractions were picked out and measured in each of the samples. The values are given in Table III. Since this was a lot of runner peanuts with splits, we have assumed that there were about

TABLE III

Damage and RMD Results on Experimental Lot

Sec.	Damage		RMD	
	A%	B%	A%	B%
1	.907	.866	.209	.124
2	.704	1.034	.194	.133
3	.727	.410	.146	.164
4	.728	.740	.161	.161
5	.702	.589	.114	.069
6	.989	1.002	.230	.266
7	.914	.923	.277	.201
8	.772	.710	.228	.315
9	.989	.793	.313	.200
10	1.024	.977	.345	.412
11	.830	1.135	.351	.305
12	.831	.917	.242	.258
13	1.170	1.037	.334	.450
14	1.171	1.059	.322	.443
15	1.128	1.044	.438	.387
16	1.096	1.179	.379	.477
Average		.909%		.270%
Standard deviation		.186		.111
Standard deviation theory		.095		.052

TABLE IV

Analysis on Subsections From Section 2

Sub sec.	Sound kernel		RMD		Net damage		Total ppb
	%	ppb	%	ppb	%	ppb	
1	98.97	4	.30	1137	.73	5	7.4
2	99.17	5	.28	45000	.55	120	130.3
3	99.23	0	.25	687	.52	0	1.7
4	98.91	8	.30	3080	.79	18	17.4
5	99.17	0	.25	70700	.58	348	178.7
							67.1

1,000 kernels per pound or a total of about 10,000 kernels in the sample. At an average damaged content of 0.91%, the standard deviation of the sampling which can be calculated theoretically is 0.095%. For the RMD average of 0.27%, the theoretical standard deviation is 0.052%. The standard deviation determined from the analytical results was 0.186% for the total damage and 0.111% for the RMD portion. The difference between these values and the theoretical values can be attributed to the variation caused by the human error of determining the damage and not to the sampling. When these results were compared to the aflatoxin results, no positive correlation of significance was found indicating that the level of aflatoxin in the contaminated kernels is quite diverse.

Section 2 was further divided into five subsections, each subsection consisting of 10 bags. A composited 10 lb sample was taken from each subsection in which each bag was represented. The damaged kernels were picked out of each and the RMD fraction was, in turn, picked out of the damaged portion. However, the fractions were analyzed individually and the aflatoxin content for the whole sample was calculated from the weighted amount in each fraction. The aflatoxin is given in Table IV as it was determined on each of the three fractions. It can be seen that the aflatoxin does reside primarily in the RMD fraction with some being present in the rest of the damage and practically none in the sound kernel portion. These values demonstrate the variation of aflatoxin in the damage and RMD fractions. The two sections with the greatest damage and RMD (1 and 4) are much lower in aflatoxin than Sections 2 and 5.

Results and Discussion

Before examining the data, it will be helpful to look at the models that have been proposed. They indicate that the distribution of the aflatoxin results due to sampling is skewed, thus providing an explanation for the inclination of the results seen in the data.

The expected distribution of contaminated kernels in the sample can be calculated from the Poisson distribution. Furthermore, the level of the aflatoxin in the contaminated kernels can be described by the exponential distribution function. When the aflatoxin content of a number of kernels is averaged out according to the exponential theory, the result will be within a given range of the true value as described by the Poisson distribution (8). Thus, an expected frequency curve can be plotted for the distribution of aflatoxin results due to sampling variation. Figure 1 shows two such distribution curves for contamination levels of 0.10% and 0.05% and an average aflatoxin content of 60 ppb. It can be seen that the curve becomes broader and more skewed with smaller contamination levels. The median and the 95% confidence intervals are given in the figure for a 10 lb

	MODEL		MEDIAN	%<60	95% CONF. INT.
A	TIEMSTRA	F(0) 99.90	55	58.1	12 - 115
B	"	" 99.95	49	61.9	0 - 130
C	WHITAKER	" 99.90	36	65.3	0 - 253
D	LOG. .35		40	66.6	9 - 183

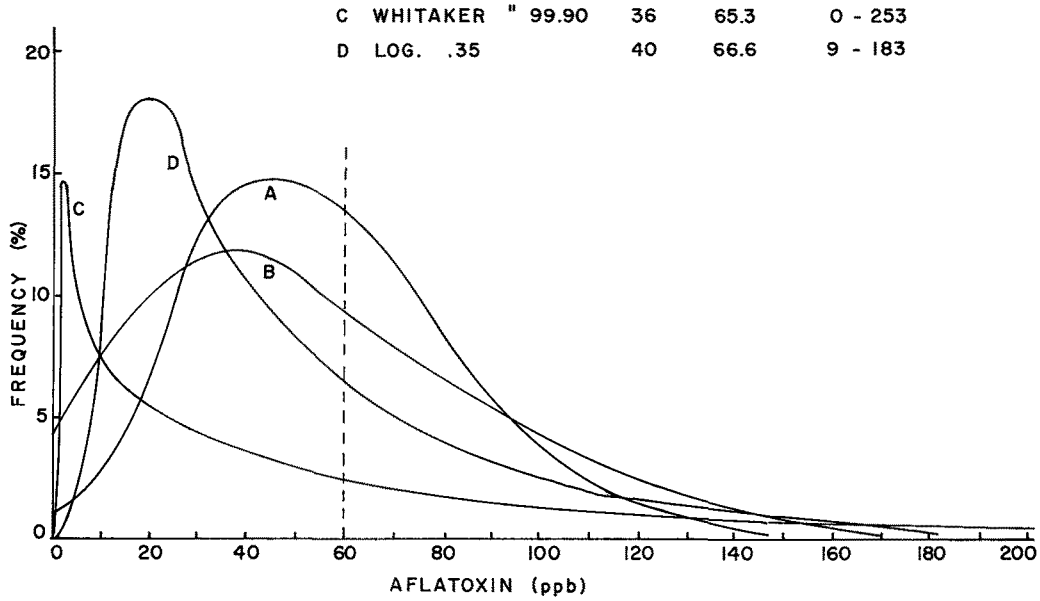

FIG. 1. Distribution curves of models.

sample along with the percentage of tests that would be obtained which would analyze under the average. The median drops from 55 ppb to 49 ppb as the contamination level decreases from 0.10 to 0.05%. A distribution curve for the Whitaker model is also given in Figure 1 with a contamination level of 0.10% which shows a very high probability of obtaining a result under 5 ppb, i.e., 20% of the test values on a lot of 30 ppb would be 5 or under and of 9% of a lot of 120 ppb, 65% of the results would be below the average. The 95% confidence interval is given for a 10 lb sample for a lot with an average of 60 ppb. Therefore, the models provide insight into how the distribution pattern for aflatoxin results should be. A log normal distribution curve (D) is included for comparison purposes. Its significance will be evident in the ensuing discussion.

The important aspect of these distribution patterns is that the probability of getting a result on a 10 lb sample below the true average is about twice as great as that of getting a result above the average. On the other hand, there is a good probability of obtaining a high result further from the average than there is of getting a low one.

Figure 2 shows the distribution pattern of the results from the experimental lot. The skewed distribution is evident, which agrees with the models. A log normal distribution curve is superimposed on the data which are obtained from the values by transforming them to logarithms and then determining their normal distribution. The log normal distribution is also skewed and it can be seen that the curve fits the data very well. It should be noted that it is impossible to handle a value of 0 when

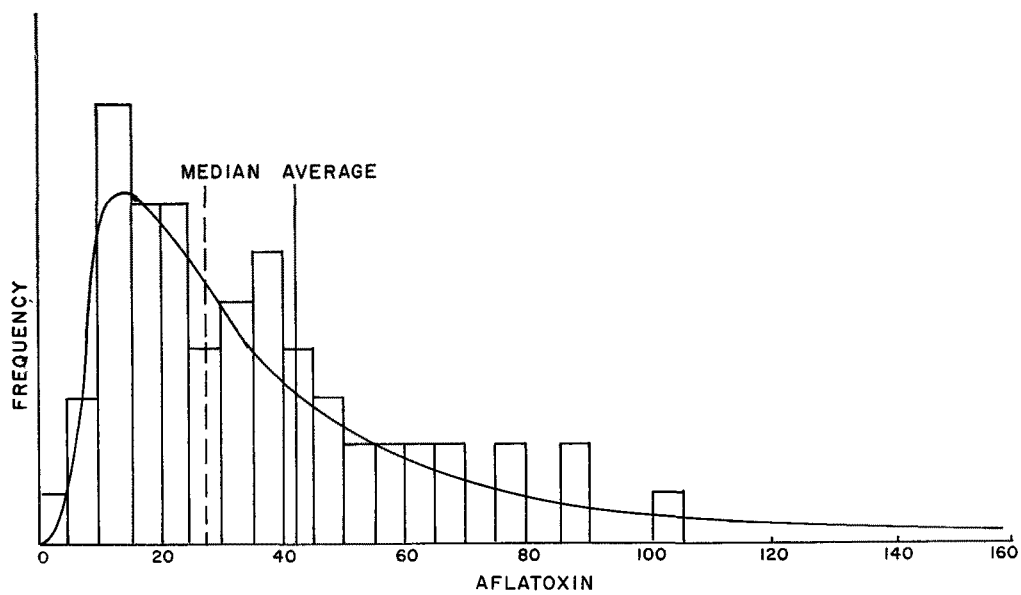

FIG. 2. Distribution curve of experimental lot.

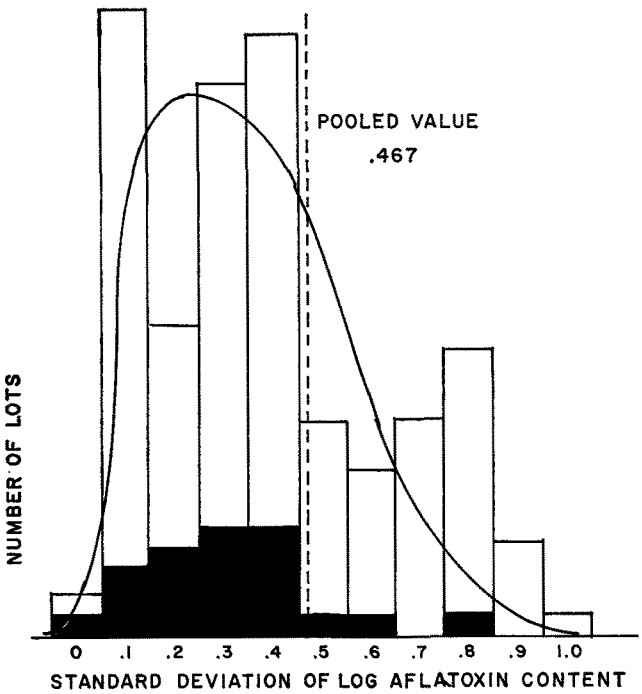

FIG. 3. Chi-square distribution for experimental lot and 1967 crop data.

log transformations are taken since the log of 0 is minus infinity. Therefore, the value 1 was assigned to all the 0 values.

The average found by the log transformation process happens to be the median, which in this case, was 27.9 ppb. In the log normal model, 66.5% of the results would be expected to be above the arithmetic average. Actually, 72% of the data was below the arithmetic average of 42.1 ppb. The 95% confidence interval is 5 to 170 ppb, according to the log analysis. Comparing this curve with the one of the model which most closely resembles it will give an indication of the ratio of contaminated to sound kernels in the lot. It may be argued that this lot represents only one example and does not reflect the true condition in all lots. Certainly it is to be expected that there will be some variation in the level of contaminated nuts from lot to lot. But

TABLE V

Aflatoxin in 1968 Crop

Lot	Original		A		B		Avg. (ppb)
	1 (ppb)	2 (ppb)	1 (ppb)	2 (ppb)	1 (ppb)	2 (ppb)	
1	53	0	55	0	80	90	46.3
2	90	13	24	0	0	0	21.2
3	90	23	0	90	0	90	48.8
4	60	65	0	60	62	90	56.2
5	45	116	29	30	0	30	41.7
6	120	105	0	30	0	30	47.5
7	135	0	29	23	0	120	51.2
8	60	43	31	8	141	8	48.5
9	45	13	0	15	128	8	34.8
10	60	13	13	45	20	53	34.0
11	90	90	0	45	0	30	43.0
12	90	0	79	230	143	30	95.3
13	38	26	40	8	58	8	29.7
14	38	54	44	30	0	0	27.7
15	75	63	23	0	13	8	30.3
16	46	19	192	23	53	15	59.7
17	45	40	0	60	0	16	26.8
18	60	126	68	75	42	45	69.3
19	53	64	63	213	46	13	75.3
20	60	0	90	88	15	12	44.2
21	60	43	90	72	45	41	58.5
22	68	26	30	16	26	0	27.7
23	75	0	23	51	0	30	29.8
24	140	20	75	22	60	60	62.8
25	60	0	80	38	63	75	52.7
26	230	10	0	8	0	0	41.3
27	45	0	67	98	112	56	63.0
28	37	45	15	130	75	0	50.3

TABLE VI

Analysis of Variance, 1968 Crop Data

Source	df	Mean square	F
Total	167		
Lots	27	1699.757274	<1 n.s.
Samples (sections)	56	2199.142857	1.059 n.s.
Subsamples	84	2075.946429	

hopefully there are none with a high average aflatoxin level having a lower level of contaminated kernels, which would give a greater variation than this particular lot.

Data were collected on a number of lots from the 1968 crop, which gives an indication of the source of variation as well as its magnitude. Table V gives the analysis on 28 lots from which three independent samples were taken, each containing portions from one fourth of the bags. Approximately 25 lb of sample were drawn from the original sample, of which 5 lb were used for the grade sample and the remainder split into two equivalent portions. Each portion was prepared and analyzed separately, usually in different laboratories. The A and B samples contained only 20 lb of sample, which were split and each half analyzed separately in different laboratories. The samples are designated original, A and B and the subsamples, 1 and 2. The analysis of variance is given in Table VI. The variation of the subsamples (sample size) is significant, whereas the variation between samples (nonhomogeneity) is not. It is interesting that there appeared to be no significant differences between lots. However, using the standard deviation which these data gave, a difference of 45 ppb between two lots would be significant at the 95% confidence level. Several sets of consecutively numbered lots from one sheller are shown in brackets in Table V. The averages of these lots show greater uniformity within the set than do the individual analyses on the subsamples, especially if the lots were all processed from one barn of farmers stock and the contamination had ample opportunity to be mixed throughout the load. This strengthens the theory that sample size is critical in reducing variation when nonhomogeneity is not present.

These data were also analyzed by logarithmic transformation. The pooled standard deviation of the data was 0.69, which is almost twice the value of the experimental lot. Part of this problem may be in the number of zeros that are in these data. Wherever a zero would appear in a lot, the calculated standard deviation would be much larger than where there were no zeros.

Over 550 lots of the 1967 crop were sampled and analyzed in triplicate. Since only one 10 lb sample was taken and analyzed, no estimate of the variation, which was due to sample size rather than nonhomogeneity could be made. However, the general magnitude of the standard deviation 0.44 on the log transformed data, could be determined. This is close to the 0.35 value obtained on the experimental lot. One interesting observation noted in these data was that 68.5% of the lots had two results under the average. This is close to the expected number of 66.6%, according to the log normal distribution.

To further substantiate the true variation in the lots, as compared to the experimental test and the theoretical models, the chi-squared test was applied. Chi-squared is defined as:

$$\chi^2 = n\,(\sigma/\sigma')^2$$

where n is the number of tests used to calculate σ,σ is the standard deviation calculated from the three tests on each lot, σ' is the true standard deviation of the entire lot. The distribution of this function can be obtained from tables in most statistical texts, including Juran (9). To understand this statistic, it might be easier to take an example of a single lot, such as the experimental lot tested above. If the samples were taken randomly in groups of three, an apparent standard deviation could be calculated from each group. Sometimes this standard deviation would be low and sometimes it would be high, but the pooled values would approach the true standard deviation. This was done and the distribution is shown as the solid bars in Figure 3.

The same approach was used for the individual lots of the 1967 crop. If all the lots had had the same level of contamination and the same variation in the sampling, then the Chi-square distribution should have approached the theoretical curve. Only lots that averaged over 100 ppb were graphed since there were fewer zeros to handle. As noted previously, zeros appear to inflate the values of the standard deviation. Therefore, it was felt that the lots with higher aflatoxin content would give a more uniform and reasonable estimate of the deviation. One hundred and ten lots that averaged over 100 ppb were present in this series. The frequency distribution of these standard deviations is shown in Figure 3, along with the theoretical distribution of the Chi-square distribution. The pooled value for the standard deviation of these data was .367, which is nearly the same as what we found on the experimental lot. Since the Chi-square distribution is approached and the pooled value of the standard deviation is approximately the same as that of the experimental lot (0.367 vs. 0.352), it is reasonable to assume that on the average, the level of contaminated kernels in most of the contaminated lots from the 1967 crop were approximately the same as what we had in the experimental lot.

The important aspect of this study is the ability to obtain some measure of the improvement in characterizing a lot by increasing the sample size. Knowing that the sampling deviation for a 10 lb sample is approximately 0.35 (log basis), the probability of obtaining an analysis of less than 20 can be calculated for any lot of a given aflatoxin level. Thus, a lot that has a true level of 20 ppb would be accepted 66.7% of the time, since the frequency distribution indicates that that is the expected number of assays at 20 or less. Likewise, a lot that had a true level of 50 ppb would have a frequency distribution that indicates that 31% of the assays would be under 20 ppb. Plotting these figures would give an operating characteristic (OC) curve as is shown in Figure 4. Increasing the sample size to 40 lb will reduce the sampling error by half ($\sigma_s = \sigma_{10}/\sqrt{n_{10}}$, where σ_s is the sampling deviation, σ_{10} is the deviation of a 10 lb sample and n_{10} is the number of 10 lb units in the sample). Since the sampling deviation contributes the major share of the deviation, it will also improve the total deviation by nearly a like amount, as was indicated earlier. On this premise, the OC curves for 30, 50, 70 and 100 lb samples were constructed and are shown in Figure 4.

One way of determining the accuracy of a sample to characterize a lot is to compare the original analysis with the analysis of the various fractions after it has been processed and cleaned up. The analysis of the pickout fraction will be more accurate since there is a much higher level of contaminated kernels, in fact, virtually all will be contaminated. Thus, the variation due to low levels of contaminated kernels is practically nil. Likewise, a 10 lb sample from the pickouts will resemble a much greater proportion of the lot, equalizing much of the bad kernel to bad

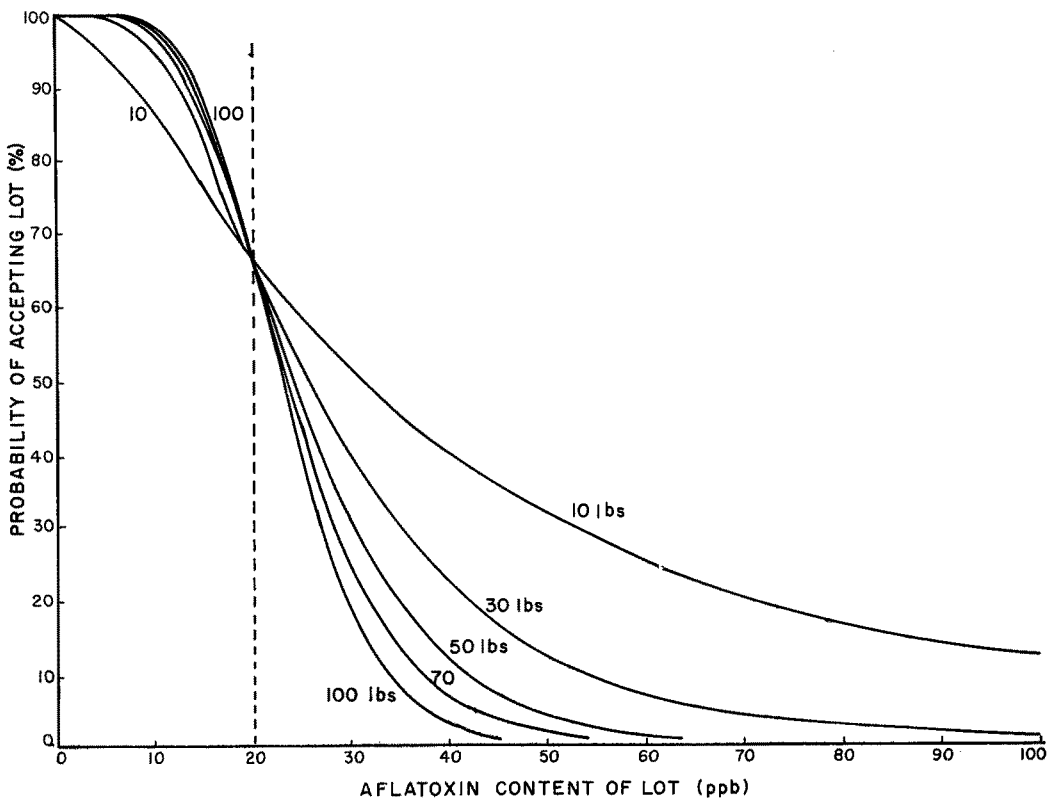

FIG. 4. Operating characteristic curves.

TABLE VII

Lot	Results before processing		Results after processing				
			Sound kernels		Pickouts		Overall Aflatoxin Avg. (ppb)
	No. of tests	Aflatoxin Avg. (ppb)	No. of tests	Aflatoxin Avg. (ppb)	%	Aflatoxin Avg. (ppb)	
A	6	75	5	0	10.8	712	77
B	6	69	5	3	6.5	690	54
C	6	58	5	10	2.3	2100	59
D	6	59	6	8	4.5	1080	57
E	6	63	5	0	2.9	1080	32
F	6	95	8	0	4.1	1960	84
G	4	143	8	2	3.0	750	23

kernel variation. The results of several lots on which such data could be conveniently obtained are given in Table VII. These lots were blanched followed by electronic and hand sorting. A number of 10 lb samples were taken from each sound kernel fraction (indicated in the table) and 1 or 2–10 lb samples from the pickouts. The weighed aflatoxin for the entire lot, as determined from these two fractions, is given in the last column. With the exception of lots E and G, the aflatoxin accounted for in the finished materials was close to the average of the determinations on the original material. One of the reasons lot G did not check well may have been the fact that only four samples and analyses were made on the original material instead of six, and of these four, two were very high.

REFERENCES

1. Whitaker, T. B., and E. H. Wiser, 5th National Peanut Research Conference, Peanut Improvement Working Group, Norfolk, Va. (1968).
2. Tiemstra, P. J., 5th National Peanut Research Conf. Peanut Improvement Working Group, Norfolk, Va. (1968).
3. Cucullu, A. F., L. S. Lee, R. Y. Mayne and L. A. Goldblatt, JAOCS *43*, 89–92 (1966).
4. Dickens, J. W., and J. B. Satterwhite, Food Technol. *23*, 950–952 (1969).
5. Waltking, A. E., G. Bleffert and M. Kiernan, JAOCS *45*, 880–884 (1968).
6. Eppley, R. M., J. Assoc. Offic. Anal. Chemists *49*, 1218–1222 (1966); amended in Ibid. *51*, 485–488 (1968).
7. Waltking, A. E., 83rd Annual Meeting, Association of Official Analytical Chemists, Washington, D.C., 1969.
8. Davis, D. J., J. Am. Statistical Assoc. *47*, 113–132 (1952).
9. Juran, J. M., "Quality Control Handbook," 2nd Edition, McGraw-Hill Book Co., Inc., New York, 1962.

[Received May 16, 1969]