Sampling Cottonseed Lots for Aflatoxin Contamination

JAMES VELASCO, Agricultural Marketing Research Institute,[1] Beltsville, Maryland 20705,
T.B. WHITAKER, Southern Division, Raleigh, North Carolina 27606, and **M.E. WHITTEN**,
Agricultural Marketing Research Institute, Beltsville, Maryland 20705

ABSTRACT

Large samples called "sublots" were drawn from 41 commercial lots of contaminated cottonseed. Each sublot was subdivided into twenty 5 lb samples which were analyzed for aflatoxin. The mean, median, variance, coefficient of variation, and the estimated range among the sample concentrations were computed. The results indicated that: (A) the variance among sample concentrations was large and was found to be a function of sample concentration and (B) the distribution of sample concentrations was skewed; the density of sample values was greater below the sublot concentration.

INTRODUCTION

The confirmation of aflatoxin in cottonseed cakes by Loosmore, et al., (1) in England in 1964 prompted a survey of cottonseed products in the U.S. The results of a 3 year survey which began in 1964 indicated that aflatoxin could be found in most of the cotton producing areas in the U.S. (2,3). Occurrence appeared to be most common in irrigated areas where temperatures and pink boll worm infestation were high.

Because of this evidence of contamination, the cottonseed processing industry inspects cottonseed by analyzing samples taken from the lots. One of the problems associated with the sampling of agricultural products for aflatoxin is that contamination generally is restricted to a very small percent of the sample. Cucullu, et al., (4) found that in peanut samples of low contamination only ca. 0.24% of the individual kernels contained aflatoxin, while at higher contamination levels ca. 5% of the kernels were contaminated. In cottonseed, Whitten (3) found that, out of 150 cottonseed picked randomly from a highly contaminated sample (8000 ppb), only 18 seeds or 12% contained aflatoxin. Because of this extreme distribution, representative sampling is difficult, variation among replicates tends to be great, and the aflatoxin concentration in a given lot may be exceedingly difficult to estimate with any degree of accuracy. As a result, certain inherent risks are associated with sampling. Samples taken from a good lot may indicate that the lot is bad (processors' risk), and, at other times, samples from a bad lot may indicate that the lot is good (consumers' risk). For peanuts, Whitaker, et al., (5,6) developed a system to evaluate the consumers' risk, processors' risk, and the costs associated with aflatoxin sampling programs used by the peanut industry.

A similar study for cottonseed was needed to help the cottonseed industry evaluate their sampling procedures. It was assumed that the approach Whitaker used for peanuts could be used for other agricultural commodities, such as cottonseed. However, because of the differences between peanuts and cottonseed, such as size and wt of individual kernels, mixing capability, and grinding, system parameters developed for peanuts might not be suitable for cottonseed. This paper describes an empirical study in which the variability of replicated samples taken from cottonseed lots was measured. Results of this study provide the basic information needed to determine system parameters for the evaluation of cottonseed sampling programs.

[1]ARS, USDA.

PROCEDURES

Forty-one "sublots" weighing ca. 200 lb each were drawn in a random fashion from 41 commercial truck lots (ca. 20-25 tons) of cottonseed contaminated with aflatoxins. It was assumed that the distribution of aflatoxin among the cottonseed in the sublots was representative of the distribution found in typical commercial lots. The identities of two sublots were lost during shipment. Cottonseed in these were shipped in eight 50 lb bags of 4 bags/lot. The sublots were recreated by arbitrarily combining the eight bags into two groups of four bags each. Also a 50 lb bag was lost from one sublot during shipment, leaving that particular sublot with 150 lb. Using a riffle divider, each of the forty 200 lb sublots was divided into 20 samples of ca. 10 lb each. The remaining 150 lb sublot was divided into fifteen 10 lb samples.

Each 10 lb sample was passed through a Bauer attrition mill with the blades set to crack the hulls of the seed. The seed then was passed over a small beater to separate the kernels from the hulls.

Because aflatoxin does not contaminate the hulls (7), they were discarded leaving ca. 5 lb kernels/sample. A count/lb indicated an average of 8640 kernels/lb or 43,200 kernels/5 lb. Each 5 lb sample of kernels was passed through a subsampling mill (8), and a subsample of ca. 100 g comminuted material was analyzed for aflatoxin by the method of Velasco (9). As a result, 815 subsamples, each representing 5 lb kernels, were analyzed.

The result of the aflatoxin test for each 5 lb sample is denoted as $\bar{x}$. For a given sublot, the true aflatoxin concentration, μ, is estimated by averaging all sample results, $\bar{x}$, from that sublot. This estimated aflatoxin concentration of each sublot is denoted by $\bar{\bar{x}}$. Therefore:

$$\bar{\bar{x}} = \sum_{i=1}^{ns} \bar{x}_i/ns, \qquad (I)$$

where ns is the number of samples analyzed for the given sublot. For each sublot, the variance among the sample results $\bar{x}$ is denoted as $s^2_{\bar{x}}$ and is an estimate of the lot variance $\sigma^2_{\bar{x}}$.

RESULTS AND DISCUSSION

Observed values of $\bar{x}$ from all sublots are tabulated in Table I along with the estimated aflatoxin concentration $\bar{\bar{x}}$. The sublots in Table I are ranked according to their $\bar{\bar{x}}$ values. Aflatoxin was not found in two sublots, leaving results for 39 sublots in Table I. From the values in Table I, the median (MD), variance ($s^2_{\bar{x}}$), the coefficient of variation (CV), and deviation estimates (D) or $\bar{x}$ for 95% confidence limits for each sublot were computed (Table II). By inspecting the $\bar{x}$ values in Table I and the statistics in Table II, the nature of aflatoxin test results for a wide range of sublot concentration $\bar{\bar{x}}$ can be described for conditions specified in the procedure.

Inspection of Table I indicates that sublot concentrations $\bar{\bar{x}}$ varied from 0.45-218.7 ppb, while individual sample concentrations ranged from 0-400 ppb. Within a given sublot, the widest range of sample concentrations occur in sublot no. 29 where $\bar{x}$ varied from 1-300 ppb. The above observations indicate that the variability among $\bar{x}$ values for each sublot was large.

Inspection of Table II indicates that the variance $s^2_{\bar{x}}$ is greater than the sublot concentration $\bar{\bar{x}}$ and that $s^2_{\bar{x}}$ tends to

TABLE I

Aflatoxin Test Results on 5 lb Samples of Cottonseed Kernels $\bar{x}$ and Estimated Sublot Concentration $\bar{\bar{x}}$

Sublot no.	Aflatoxin test results for samples - $\bar{x}$ (ppb)																				Sublot concentration $\bar{\bar{x}}$
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	4	.45
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	10	.55
3	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	3	7	.80
4	0	0	1	1	1	1	1	1	1	1	1	1	1	1	2	4	6	6	8	12	2.50
5	0	0	0	0	0	0	0	0	0	1	1	1	1	1	3	5	7	9	10	14	2.65
6	0	0	0	0	0	0	0	0	0	0	0	1	1	1	8	17	21	21	25	50	7.25
7	0	0	1	1	1	1	1	1	1	1	1	1	1	1	2	7	7	13	23	100	8.20
8	0	0	0	1	1	1	1	1	1	1	4	6	10	10	12	13	16	27	40	44	9.45
9	0	0	0	0	1	1	1	1	1	1	4	5	5	7	9	13	20	25	41	70	10.25
10	0	1	1	1	1	1	1	1	1	1	1	1	2	7	11	16	21	27	27	90	10.60
11	0	0	0	0	0	1	1	1	1	2	7	9	10	16	25	26	30	33	39	48	12.45
12	0	0	0	0	1	1	1	1	1	1	8	9	12	23	24	24	25	30	40	50	12.55
13	1	1	1	1	1	3	8	10	10	10	13	13	13	16	20	22	27	33	34	36	13.65
14	0	1	1	1	1	3	3	9	10	19	12	21	34	41	58						13.67
15	0	0	0	0	0	0	2	2	4	5	6	7	9	11	13	14	20	37	53	92	13.90
16	0	1	1	1	1	1	1	4	4	11	12	14	19	20	21	22	24	28	38	56	13.95
17	1	1	1	1	2	4	5	7	8	10	10	11	12	14	17	31	34	36	38	83	16.30
18	0	0	1	1	1	2	3	8	9	10	13	20	21	25	33	40	45	50	71	82	21.75
19	1	1	1	3	5	6	7	10	10	14	15	20	20	24	29	30	35	40	50	133	22.70
20	1	2	2	3	5	6	6	9	15	22	30	32	33	37	48	54	55	70	86	90	29.65
21	0	4	9	14	16	16	25	27	30	30	31	32	32	32	34	37	40	42	42	100	30.30
22	0	3	4	6	26	28	30	31	34	36	37	38	40	42	43	53	61	63	81	98	37.70
23	2	4	5	8	12	24	31	36	36	38	40	41	51	51	52	66	70	74	84	84	40.75
24	6	6	7	14	20	22	24	24	31	33	38	40	42	45	54	60	67	68	68	165	41.70
25	2	9	10	13	15	15	17	21	22	26	29	38	40	41	41	52	59	100	130	166	42.30
26	6	8	11	14	20	20	22	23	26	28	50	50	56	62	64	74	75	80	100	138	46.35
27	12	13	14	28	29	31	32	32	34	41	44	49	65	70	74	77	77	77	93	119	50.55
28	10	10	14	20	25	31	32	34	37	37	55	61	61	65	70	74	83	86	101	117	51.15
29	1	1	13	17	20	24	28	29	31	31	36	41	42	55	55	67	91	100	166	300	57.40
30	3	9	13	17	25	33	33	40	46	46	51	59	60	67	67	90	100	100	145	204	60.40
31	24	30	32	34	35	35	36	40	42	46	46	46	51	69	75	76	90	98	133	195	61.65
32	15	16	20	21	27	30	48	52	57	67	70	80	80	90	111	118	133	136	144	160	73.75
33	45	53	61	65	70	76	87	88	89	90	92	92	93	94	96	102	113	114	115	134	88.45
34	1	16	29	40	53	73	85	89	100	104	113	118	120	121	128	143	157	175	260	266	109.85
35	25	42	61	65	70	99	102	120	132	146	157	167	171	200	218	267	270	270	300	380	162.40
36	70	80	91	114	116	120	127	130	133	150	178	178	192	196	200	201	206	237	252	269	169.75
37	62	74	76	87	99	119	128	150	154	154	167	178	179	179	206	207	233	280	303	365	170.00
38	84	95	119	127	128	132	150	160	167	173	175	201	201	205	214	217	218	234	238	263	175.05
39	80	85	99	99	105	106	109	169	220	236	240	247	260	269	289	295	307	375	384	400	218.70

TABLE II

Estimated Sublot Concentration $\bar{\bar{x}}$, Median MD, Variance $s^2_{\bar{x}}$, Coefficient of Variation CV, and Deviation $D_{.05}$ for 95% Confidence Limits

Minilot	$\bar{\bar{x}}$ (ppb)	MD (ppb)	$s^2_{\bar{x}}$	CV (%)	$D_{.05}$ (ppb)
1	0.45	0.0	0.9	209.879	1.977
2	0.55	0.0	5.0	406.436	4.679
3	0.80	0.0	2.7	205.206	3.436
4	2.50	1.0	9.8	125.488	6.566
5	2.65	1.0	17.1	155.964	8.650
6	7.25	0.0	174.3	182.102	27.633
7	8.20	1.0	498.3	272.220	46.720
8	9.45	2.5	174.0	139.606	27.613
9	10.25	2.5	309.3	171.566	36.806
10	10.60	1.0	431.2	195.900	43.462
11	12.45	4.5	237.3	123.735	32.243
12	12.55	4.5	234.3	121.957	32.035
13	13.65	11.5	135.2	85.179	24.335
14	13.67	9.0	306.2	128.015	37.534
15	13.90	5.5	520.1	104.069	47.732
16	13.95	13.0	222.8	106.997	31.240
17	16.30	10.0	397.1	122.248	41.706
18	21.75	11.5	600.7	112.688	51.299
19	22.70	14.5	872.0	130.087	61.806
20	29.65	26.0	424.3	69.476	43.115
21	30.30	30.5	819.3	94.465	59.908
22	37.70	36.5	636.7	66.933	52.814
23	40.75	39.0	699.7	64.911	55.363
24	41.70	35.5	1259.8	85.117	74.288
25	42.30	27.5	1831.4	101.169	89.569
26	46.35	39.0	1228.7	75.625	73.364
27	50.55	52.5	863.6	58.136	61.508
28	51.15	46.0	958.8	60.536	64.808
29	57.40	33.5	4747.6	120.040	144.214
30	60.40	48.5	2401.1	81.127	102.559
31	61.65	46.0	1751.6	67.887	87.597
32	73.75	68.5	2183.3	63.356	97.796
33	88.45	91.0	590.5	27.474	50.861
34	109.85	108.5	4990.0	64.306	147.850
35	162.40	151.5	9185.9	59.017	200.600
36	169.75	164.0	4740.7	40.561	144.109
37	170.00	160.5	6400.3	47.060	167.444
38	175.05	174.0	2509.3	28.616	104.845
39	218.70	238.0	11309.4	48.626	222.581

increase with the sublot concentration $\bar{\bar{x}}$. Because the sublot concentration $\bar{\bar{x}}$ varied so widely, the relative variability, or coefficient of variation, was calculated. The coefficient of variation is defined as:

$$CV = s_{\bar{x}} * 100/\bar{\bar{x}}, \quad \text{(II)}$$

where $s_{\bar{x}}$ is the standard deviation or the square root of $s^2_{\bar{x}}$. The CV is large, especially at low sublot concentrations (Table II). For ca. half of the sublots, the CV is greater than 100%, indicating that the standard deviation $s_{\bar{x}} > \bar{\bar{x}}$ for these sublots. The CV tends to decrease as $\bar{\bar{x}}$ increases. Therefore, even though the variance $s^2_{\bar{x}}$ tends to increase with $\bar{\bar{x}}$, the variability of $\bar{x}$ values relative to the sublot concentration $\bar{\bar{x}}$ tends to decrease as $\bar{\bar{x}}$ increases.

The practical implications that can be made from the above discussion concerning variance and CV values is that it would be very difficult to estimate, with any reliability, the true lot concentration by drawing one 5 lb sample of cottonseed kernels. From Table I, a rough estimate of the maximum deviation expected in the $\bar{x}$ values, ca. $\bar{\bar{x}}$, can be made. However, using the variance estimates in Table II, the deviation, for 95% confidence limits, can be estimated more precisely by using the t distribution (9). The interval for 95% confidence would be:

$$D_{0.5} = t_{.05} * s_{\bar{x}}, \quad \text{(III)}$$

where $t_{.05}$ is the tabulated value of Student's t-test for 5% probability level and 19 degrees of freedom. For sublot 14 where there were 15 $\bar{x}$ values, the degrees of freedom would be 14. The interval in which sample values would lie for 95% confidence would be:

$$I_{.05} = \bar{\bar{x}} \pm D_{.05}. \quad \text{(IV)}$$

In Table II, the value of $D_{.05}$, or the magnitude of the deviation expected in the $\bar{x}$ values, is given for each sublot. The deviation, $D_{.05}$, is greater than the sublot concentration $\bar{\bar{x}}$ in all but four sublots. These four exceptions fell among the seven sublots having the highest $\bar{\bar{x}}$ values. The fact that, for all sublots below no. 33, the deviation is greater than $\bar{\bar{x}}$ at the 95% confidence level indicates that 5 lb samples might assay "negative" (zero ppb) when drawn from lots with concentrations up to ca. 100 ppb.

The underlying assumption concerning the deviation estimates is that the distribution of $\bar{x}$ values is normal or symetric about the lot concentration μ. However, the fact that $D_{.05}$ is greater than $\bar{\bar{x}}$, especially at low $\bar{\bar{x}}$ values, indicates that the distribution of $\bar{x}$ values may be skewed to the low side of the sublot concentration $\bar{\bar{x}}$. This implies that there is a greater probability of obtaining a sample result $\bar{x}$ less than $\bar{\bar{x}}$ than a result $\bar{x}$ greater than $\bar{\bar{x}}$.

Comparison of the median MD with the arithmetic mean $\bar{\bar{x}}$ for each sublot in Table II also indicates that the distribution is skewed. The median is the middle item in any array, or that value for which 50% of the $\bar{x}$ observations, when arranged in order or magnitude, lie on each side (10,11). For all but three sublots, MD < $\bar{\bar{x}}$ which indicates again that there is a greater density of $\bar{x}$ values on the low side of $\bar{\bar{x}}$. The $\bar{x}$ values in Table I and the difference between $\bar{\bar{x}}$ and MD in Table II, indicate that there is more skewness at low $\bar{\bar{x}}$ values and the distribution becomes more symetrical as $\bar{\bar{x}}$ increases. This observation agrees with the behavior of the coefficient of variation with $\bar{\bar{x}}$, in that, the higher the CV value, the more skewed the distribution of $\bar{x}$ appear to be about $\bar{\bar{x}}$.

Because the variance $s^2_{\bar{x}}$ appears to increase with the sub-

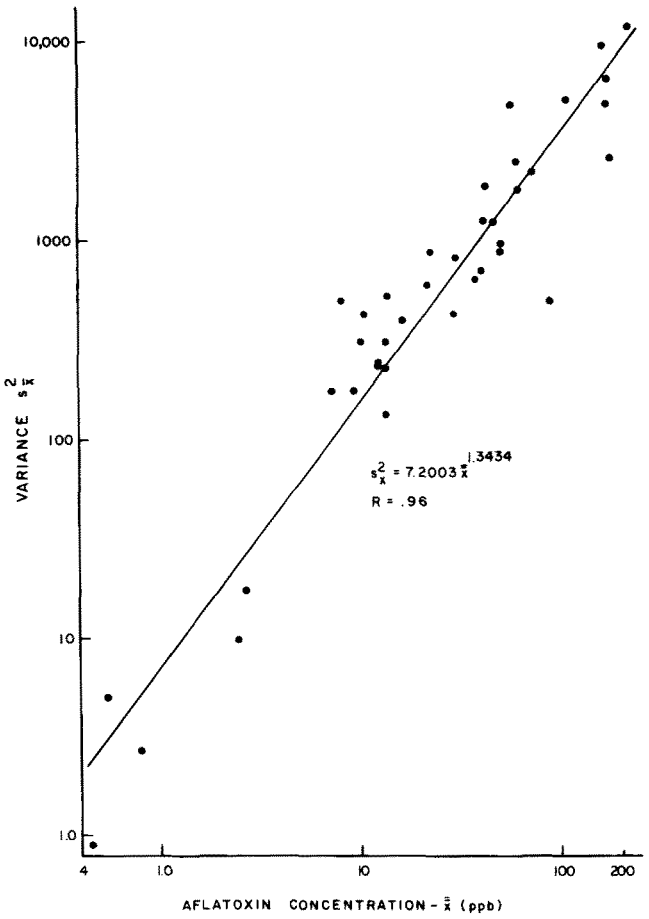

FIG. 1. Relationship between the variance of 5 lb samples $s^2_{\bar{x}}$ and the estimated sublot concentration $\bar{\bar{x}}$. Correlation coefficient r = .96.

lot concentration $s^2_{\bar{x}}$ may be a function of $\bar{\bar{x}}$. Cochran (12) reported that studies suggest the relationship to be:

$$s^2_{\bar{x}} = A\,\bar{\bar{x}}^B, \qquad (V)$$

where A and B are constants independent of $\bar{\bar{x}}$. If equation V is an appropriate function, a plot of $s^2_{\bar{x}}$ vs $\bar{\bar{x}}$ on a log-log graph should result in a linear relationship (Fig. 1). The regression equation for the plot is:

$$\ln(s^2_{\bar{x}}) = 1.9741 + 1.3434 * \ln(\bar{\bar{x}}), \qquad (VI)$$

with a correlation coefficient of 0.961 in the lob scale. Equation VI can be transformed to give:

$$s^2_{\bar{x}} = 7.2003\,\bar{\bar{x}}^{1.3434}. \qquad (VII)$$

Using Figure 1 or equation VII, the variance of 5 lb samples can be estimated for any sublot concentration $\bar{\bar{x}}$. For example, the estimated variance of 5 lb samples from a lot with a concentration of 20 ppb is ca. 400.

The coefficient of variation (equation II) also can be expressed as a function of $\bar{\bar{x}}$ by combining equation II with equation VII:

$$CV = 268.33\,\bar{\bar{x}}^{-0.3283}. \qquad (VIII)$$

A plot of CV values in Table II and equation VIII is shown in Figure 2. Equation VIII is not the result of a regression analysis on CV values (Table II) but of the regression analysis on the variance values which are substituted into the CV equation. As Figure 2 indicates, the CV decreases as $\bar{\bar{x}}$ increases, and the decrease in CV values is more rapid at low $\bar{\bar{x}}$ than at high $\bar{\bar{x}}$ values.

Equations VII and VIII reflect the total variability of aflatoxin test results on 5 lb samples of cottonseed kernels, 100 g subsamples, and the analytical method developed by

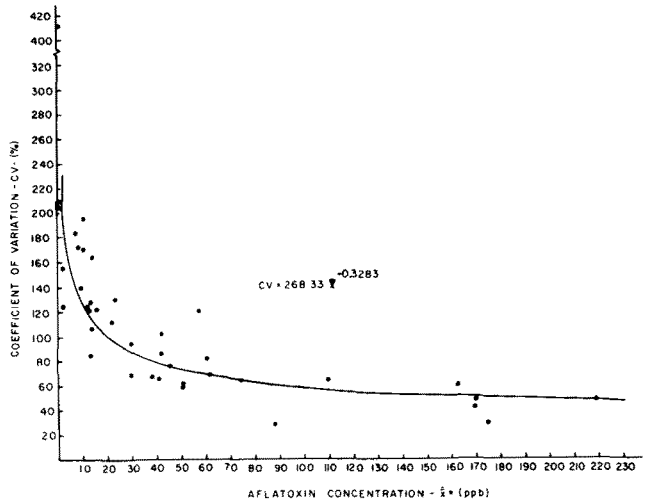

FIG. 2. Relationship between the coefficient of variation CV of 5 lb samples and the estimated sublot concentration $\bar{\bar{x}}$.

Velasco. The effect of sample size ω (where ω is the wt) upon the variability of aflatoxin test results can be estimated if it is assumed that all observed variability is associated with sampling and that no variability is associated with the subsampling and analytical procedures. Based upon the above assumption the variability of aflatoxin test results for samples of size ω is:

$$s^2_{\bar{x}}\Big|_{\omega} = (5/\omega) * s^2_{\bar{x}}\Big|_{5}, \qquad (IX)$$

where $s^2_{\bar{x}}|_5$ is the variance among samples of size 5 lb which is given by equation VII. Substituting equation VII into equation IX gives:

$$s^2_{\bar{x}}\Big|_{\omega} = (36.0015/\omega) * \bar{\bar{x}}^{1.3434}. \qquad (X)$$

Equation X will tend to underestimate $\sigma^2_{\bar{x}}$ when $\omega > 5$ lb and overestimate $\sigma^2_{\bar{x}}$ when $\omega < 5$ lb, due to the assumption that all variability is confined to sampling. However, until the variability of the subsampling and analytical procedures can be estimated, equation X can serve as a guide to the effect of sample size upon the variability of aflatoxin test results.

The sampling results presented in this paper indicate the nature of the sampling problems that may be encountered in the inspection of cottonseed lots for aflatoxin. Also the data will provide the basis for the estimation of system parameters needed to evaluate the costs and risks associated with various sampling procedures. Future investigations will be concerned with determining a model that will accurately simulate the observed distribution of $\bar{x}$ values given in Table I. Once a simulation model is decided upon, operating characteristic curves representing specific sampling procedures will be evaluated.

ACKNOWLEDGMENTS

E.L. Cox, ARS, gave statistical advice, and L. Lanzafame and S. Morris provided technical assistance.

REFERENCES

1. Loosmore, R.M., R. Allcroft, E.A. Tutton, and R.B.A. Carnaghan, Vet. Res. 76:64 (1964).
2. USDA, Report to the National Cottonseed Products Association, Memphis, Tenn., December 1964.
3. Whitten, M.E., Cotton Gin Oil Mill Press 67:7 (1966).
4. Cucullu, A.F., L.S. Lee, R.Y. Mayne, and L.A. Goldblatt, JAOCS 43:89 (1966).
5. Whitaker, T.B., and E.H. Wiser, Ibid. 46:377 (1969).
6. Whitaker, T.B., J.W. Dickens, and R.J. Monroe, Ibid. 49:590 (1972).

7. Whitten, M.E., Ibid. 47:5 (1969).
8. Dickens, J.W., and J.B. Satterwhite, Food Technology 23:90 (1969).
9. Velasco, J., J. Ass. Offic. Anal. Chem. 55:1360 (1972).
10. Steel, R.G., and J.H. Torrie, "Principles and Procedures of Statistics," McGraw-Hill, New York, N.Y., 1960.
11. Snedecor, G.W., "Statistical Methods," Iowa State University Press, Ames, Iowa, 1961.
12. Cochran, W.G., "Sampling Techniques," John Wiley & Sons, New York, N.Y., 1963.

[Received September 3, 1974]