Theoretical Investigations Into the Accuracy of Sampling Shelled Peanuts for Aflatoxin[1]

T. B. WHITAKER and E. H. WISER, Market Quality Research Division, USDA, and Biological and Agricultural Engineering Dept., N.C. State University, Raleigh, North Carolina 27607

Abstract

Within a population of shelled peanuts, aflatoxin may be concentrated in less than 0.5% of the peanuts. Those peanuts containing aflatoxin might have concentrations up to 1,000,000 μg of aflatoxin per kilogram of peanuts. Because of the distribution pattern, sample means vary widely, and the true average level of aflatoxin in the population is difficult to estimate. The objective of this study was to determine the effect of sample size, N, on sampling accuracy. The negative binomial distribution was used to simulate the actual distribution of aflatoxin since it allowed for a high probability of zero counts along with small probabilities of large counts. Using both the Monte Carlo technique and a direct computation method, the effect of sample size on sampling accuracy was quantitatively described.

Introduction

Aflatoxin, a toxic material produced by the fungus *Aspergillus flavus,* has become a problem in the peanut industry since its discovery in 1960 (1). The average level of aflatoxin in a lot of peanuts is presently estimated by analyzing a sample drawn from the lot. It is important that the average level of aflatoxin in a population of peanuts be accurately determined. Because of the wide range in aflatoxin content among individual kernels in samples of contaminated peanuts, representative sampling is difficult, and variation among replicates tends to be great (2).

The objective of this study was to determine the effects of sample size, N, upon sampling accuracy. The problem was approached by two methods: model simulation, using the Monte Carlo technique, and direct computation of a specified distribution. In both methods, the digital computer was used extensively.

Experimental Procedures

Simulation

To simulate the sampling of a population of peanuts, two requirements must be satisfied: a specified distribution must be selected which matches as closely as possible the probable distribution of aflatoxin in the population, and the method of sampling from the specified distribution must be defined.

The distribution of aflatoxin in shelled peanuts has been described only in general terms. Cucullu et al. (2) indicated that aflatoxin contaminates a small percentage, 5% or less, of the peanuts. On those peanuts which have aflatoxin, however, the level of contamination on a single kernel may reach levels of 1,000,000 μg, or more, of aflatoxin per kilogram. Whitten (3), working with cotton seed, found results similar to those described by Cucullu et al. The negative binomial distribution is a discrete distribution which can match these characteristics. High probabilities of zero counts can be matched with low probabilities of very large counts. This distribution has, in fact, been used successfully in studies of incidence of contagious diseases (4). Since aflatoxin contamination results from the infection of a population of peanuts by *Aspergillus flavus,* the use of this distribution seems reasonable.

The negative binomial probability function may be written as in equation 1.

$$f(x) = (\Gamma(x+k)/(x!\,\Gamma(k)))\,(k/(M+k))^k\,(M/(M+k))^k \quad [1]$$

for x = 0, 1, 2, . . . , where x is the quantity of aflatoxin per peanut, M is the average quantity of aflatoxin, and k is a shape parameter which is related to the degree of contagion d by the relation $k = M/d$. The degree of contagion d represents the degree of dependence of the level of contamination on one peanut to the contamination on an adjacent peanut.

Polya (5) developed a generalized model for contagious events and showed that, for the case of rare events and positive contagion ($d > o$), the negative binomial function [1] was the appropriate distribution. For rare events with no contagion ($d = 0$), the Poisson distribution is obtained.

$$f(x) = ((M)^x/x!)\,(e)^{-M}.$$

For more common events with positive contagion ($d > o$), the gamma distribution is obtained.

$$f(x) = (1/\Gamma(k))\,(k\,x/M)^{k-1}\,(e)^{-k\,x/M}$$

The unit of measurement for aflatoxin was set as 0.001 μg to facilitate conversion to micrograms of aflatoxin per kilogram of peanuts (μg/kg), which is the conventional method of stating aflatoxin con-0.001 μg to facilitate conversion to micrograms of obtaining exactly x units of aflatoxin. The cumulative distribution function is defined as

$$F(x) = \sum_{r=0}^{x} f(r) = \sum_{r=0}^{x} (\Gamma(r+k)/(r!\,\Gamma(k)))\,(k/(M+k))^k\,(M/(M+k))^r, \quad [2]$$

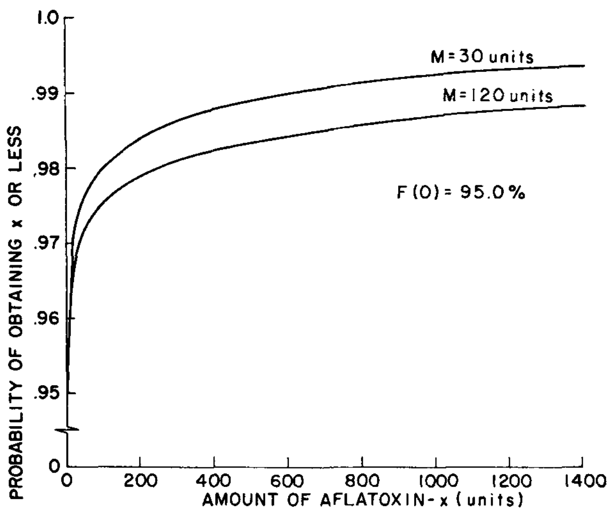

FIG. 1. Cumulative negative binomial distribution.

[1] Journal Paper No. 2775, North Carolina State University Agricultural Experiment Station, Raleigh, N.C.

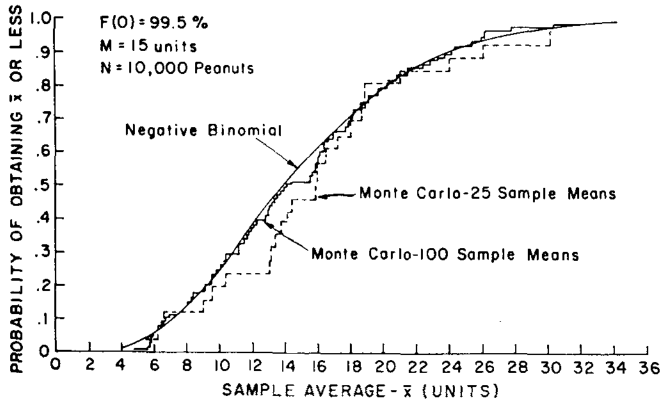

FIG. 2. Comparison of cumulative distribution of sample means as given by Monte Carlo Technique and the negative binomial cumulative distribution function.

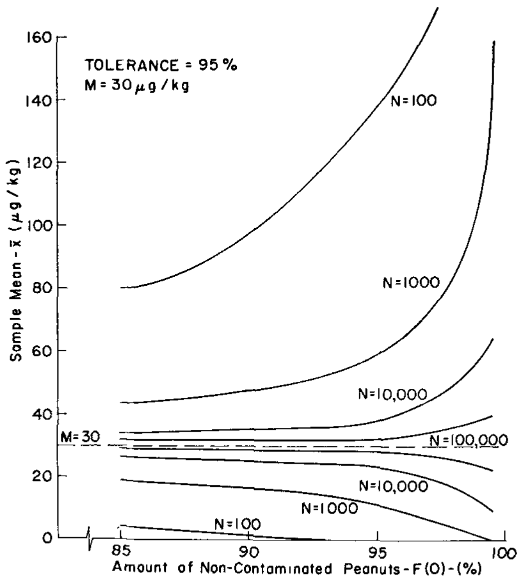

FIG. 4. The effect of the per cent of non-contaminated peanuts on sampling accuracy.

where r is a dummy variable for x and takes the same units as x. Equation 2 gives the probability of obtaining x or less aflatoxin. The per cent of non-contaminated peanuts can be determined by setting x = 0 in 1. Therefore,

$$F(o) = f(o) = (k/(M+k))^{k} \qquad [3]$$

By specifying M and either F(0) or k, the negative binomial distribution is completely described. Two examples of the cumulative distribution, as described by Equation 2, for M = 30 units with F(o) = 95% and M = 120 units with F(o) = 95% are illustrated in Figure 1. The shape of the distribution changes according to the values of M and F(0). For fixed F(0), an increase in M increases the probability of very large amounts. For fixed M, an increase in F(0) has a similar effect.

Equation 2, along with the Monte Carlo technique, may be used to simulate actual sampling of a population of peanuts. The sampling procedure is analogous to drawing a peanut at random from the population and determining the amount of aflatoxin on the peanut. It is assumed that if sufficient peanuts were examined, the aflatoxin amounts would be distributed according to the negative binomial distribution with fixed values of M and F(0). This procedure can therefore be simulated by selecting random numbers which are distributed according to the negative binomial distribution with specified M and F(0), each value thus determined is assumed to be a corresponding amount of aflatoxin on an imaginary peanut.

The actual procedure for obtaining random numbers requires generation of random numbers uniformly distributed between 0 and 1. Any such number substituted in Equation 2 for F(x), allows solution for x which is the simulated amount.

If a simple random sample of N peanuts is to be simulated, then N random numbers are generated with N values of aflatoxin $x_1, x_2, \ldots, x_N$ being calculated. The digital computer was used to generate the random numbers and solve Equation 2 for x.

Once a sample of size N was drawn, the sample mean of aflatoxin was calculated. Since the sample means are themselves distributed about the mean of the distribution M, numerous samples of size N must be taken. The sample means $\bar{x}$ from the Monte Carlo method should be normally distributed for sufficiently large sample sizes, and this seemed to be true for the largest values of N. However, for smaller values of N, the normal approximation was unsatisfactory. A

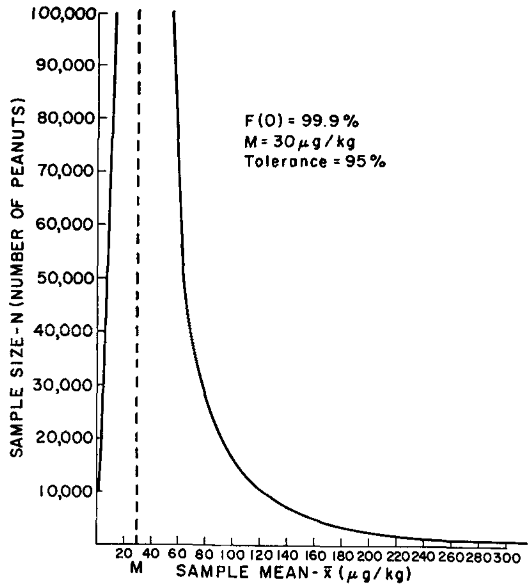

FIG. 3. The effect of sample size on sampling accuracy.

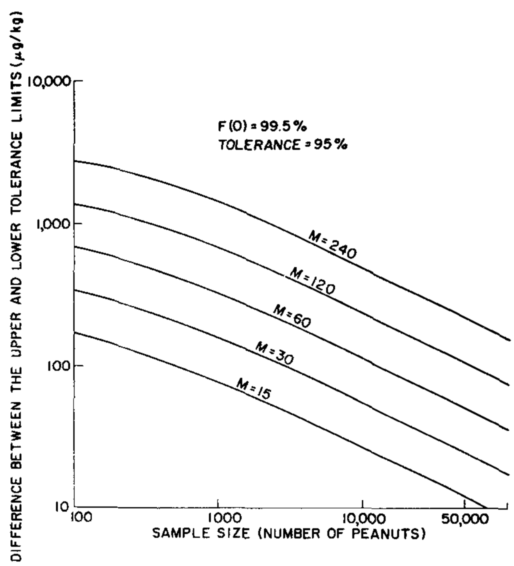

FIG. 5. The effect of the average level of aflatoxin in a population on sampling accuracy.

more complete description of Monte Carlo Simulation using the digital computer as a tool is described by Naylor *et al.* (6).

Direct Computation

The Monte Carlo approach is very valuable when the distribution of the sample means cannot be calculated directly. Results will be more accurate, however, if the exact distribution of the sample means is available. For the negative binomial distribution, the exact cumulative distribution of the sum of N identically distributed values can be calculated and is given in Equation 4. This distribution is also negative binomial. The exact cumulative distribution of the sample means can then be determined by a simple scale transformation.

$$F(N\bar{x}) = \sum_{r=0}^{N\bar{x}} (\Gamma(r+Nk)/(r!\,\Gamma(Nk)))\,(Nk/(Nk+NM))^{Nk}\,(NM/(Nk+NM))^{r} \quad [4]$$

The cumulative distributions of the sample means from both the Monte Carlo method and Equation 4 are illustrated in Figure 2. The two distributions, one containing 25 sample means and the other containing 100 sample means, from the Monte Carlo method vary about the exact cumulative distribution. The figure illustrates that at least 100 sample means are required to adequately approximate the exact distribution.

Procedure

Using Equation 4, the cumulative distribution of the sample means was calculated for different values of F(0), M and N. Several cases were studied using F(0) values of 99.9%, 99.5%, 95.0%, 90.0% and 85.0%, M values of 15, 30, 60, 120, 240, 480 and 960 units and N values of 100, 1,000, 10,000, 50,000 and 100,000 peanuts. Using the exact cumulative distribution of the sample means, sampling accuracy was quantitatively described in terms of upper and lower tolerance limits between which 95% of the sample means would lie. The upper and lower tolerance limits, which are around the true mean of the population M, were determined by finding the values of the sample mean corresponding to the cummulative probabilities of 0.975 and 0.025, respectively.

The sample means are an average of the units of aflatoxin per peanut within the sample. The units of aflatoxin per peanut were converted into μg/kg by dividing the units of aflatoxin by the weight of the peanut.

$$\mu g/kg = (units/peanut)/(g/peanut) \quad [5]$$

This conversion makes it possible to simulate the sampling of different varieties of peanuts. Woodruff (7) gives the number of peanuts per pound for several varieties which can be used to determine the average weight per peanut in a particular variety. For convenience, the results of this study are expressed in μg/kg by assuming an average weight of 1 g.

Results and Discussion

For given F(0) and M values, sampling accuracy increases as the sample size N increases. Figure 3 illustrates the effect of sample size on sampling accuracy for F(0) = 99.9% and M = 30 μg/kg. Samples of 4,500 peanuts (approximately 10 lb) drawn from the described population would have means falling between 0 and 144 μg/kg 95% of the time. However, samples of 90,000 peanuts (approximately 200 lb) drawn from the same population would have means with a 95% chance of lying between 12 and 57 μg/kg. Increasing the sample size N to 20,000 peanuts reduces appreciably the upper tolerance limit. The lower tolerance limit changes much less since it has a lower bound at $\bar{x} = 0$.

The effect of the per cent of non-contaminated peanuts F(0) on sampling accuracy is illustrated in Figure 4. For given M and N values, sampling accuracy increases as F(0) decreases. As the fraction of contaminated peanuts decreases with M and N constant, the inclusion of contaminated peanuts in a sample becomes more and more rare thereby causing greater variation among sample means $\bar{x}$.

The effect of the average level of aflatoxin M in the population on sampling accuracy is illustrated in Figure 5. For given F(0) and N values, sampling accuracy increases as M decreases. The difference between the upper and lower tolerance limits is given for several M values.

The ability to quantitatively describe the effect of sample size on sampling accuracy gives added insight into the problems of sampling shelled peanuts for aflatoxin. This study provides a foundation from which an efficient sampling procedure (sample size and number of samples) can be formulated to determine if the true average level of aflatoxin M in a population of shelled peanuts is above or below a certain critical level.

REFERENCES

1. Bampton, S. S., Trop. Sci. *5*, 74–81, (1963).
2. Cucullu, A. F., L. S. Lee, R. Y. Mayne and L. A. Goldblatt, JAOCS *43*, 89–92, (1966).
3. Whitten, M. E., The Cotton Gin and Oil Mill Press, *67*, 7 (1966).
4. Anscombe, F. J., Biometrika *37*, 358–382, (1950).
5. Polya, G., Ann. Inst. Henri Poincare *1*, 117–161, (1931).
6. Naylor, T. H., J. L. Balintfy, S. B. Donald, Kong Chu, "Computer Simulation Techniques," John Wiley & Sons, Inc., New York (1966).
7. Woodroof, J. G., "Peanuts: Production, Processing, Products," The Avi Publishing Co., Inc., Westport, Connecticut, 176, (1966).

[Received December 13, 1968]