ABSTRACT

JOHANSSON, ANDERS STURE. **Develop Methods To Evaluate the Performance of Aflatoxin Sampling Plans for Shelled Corn**. (Under the direction of Thomas Burton Whitaker.)

Eighteen lots of shelled corn were tested for aflatoxin contamination. The variability and distributional characteristics associated with the aflatoxin testing procedure were investigated. The total variance associated with testing shelled corn was estimated and partitioned into sampling, sample preparation, and analytical variances. All variances were found to increase with an increase in aflatoxin concentration. Using regression analysis, mathematical expressions were developed to model the relationship between aflatoxin concentration and the total, sampling, sample preparation, and analytical variances. The expressions for these relationships were used to estimate the variance for any sample size, subsample size, and number of analyses for a specific aflatoxin concentration. For example, testing a lot with 20 parts per billion (ppb) aflatoxin using a 2.5 lb sample, Romer mill and 50 g subsample, and HPLC analysis, the total, sampling, sample preparation, and analytical variances are 274.9 (CV=82.9%), 214.0 (CV=73.1%), 56.3 (CV=37.5%), and 4.6 (CV=10.7%), respectively. The percentage of the total variance for sampling, sample preparation, and analytical is 77.8, 20.5, and 1.7 %, respectively.

Next, fifteen positively skewed distributions were each fitted to 18 empirical distributions of aflatoxin test results for shelled corn. The compound gamma distribution was selected to model the sample aflatoxin test results for shelled corn. The method of moments technique was chosen to estimate the parameters of the compound gamma distribution. Mathematical expressions were developed to calculate the parameters of the

compound gamma distribution for any lot aflatoxin concentration and test procedure. Observed acceptance probabilities were compared to operating characteristic curves predicted from the compound gamma distribution and all 18 distributions of sample aflatoxin test results were found to lie within a 95% confidence band.

Using the mean and variance relationships to compute the parameters of the compound gamma distribution, 16 sampling plans, based on four sample sizes and four sample acceptance levels were created and analyzed. For a given sample size, decreasing the sample acceptance level, using a sample acceptance level equal to the regulatory guideline: (a) decreases the percentage of lots accepted while increasing the percentage of lots rejected at all aflatoxin concentrations; (b) increases misclassification of lots (both false positives and false negatives) while decreasing the percentage of correct decisions; and (c) decreases the average aflatoxin concentration in the lots accepted and lots rejected. For a given sample size where the sample acceptance level is less than the regulatory guideline, the number of false positives increases and the number of false negatives decreases when compared to the situation where the sample acceptance level equals the regulatory guideline. For a given sample size, where the sample acceptance level is greater than the regulatory guideline, the number of false positives decreases and the number of false negatives increases when compared to the situation where the sample acceptance level equals the regulatory guideline. Increasing the sample size for a given sample acceptance level, where the legal limit equals the sample acceptance level: (a) increases the percentage of lots accepted at lower concentrations while increasing the percentage of lots rejected at higher concentrations; (b) decreases misclassification of lots (both false positives and false negatives) while increasing the percentage of correct

decisions; and (c) decreases the average aflatoxin concentration in the lots accepted while increasing the average aflatoxin concentration in the rejected lots.

DEVELOP METHODS TO EVALUATE THE PERFORMANCE OF AFLATOXIN SAMPLING PLANS FOR SHELLED CORN

by

ANDERS STURE JOHANSSON

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

BIOLOGICAL AND AGRICULTURAL ENGINEERING

Raleigh

1998

APPROVED BY:

_____ _____

Chair of Advisory Committee

_____ _____

DEDICATION

I would like to dedicate this dissertation

to my father

Anders Daniel Johansson

who in our short time together provided me with a lifetime of morals

and values,

to my grandmother

Mammaw

who was always there to support me (many times financially)

through all of life's challenges,

to my parents

Teak and Betsey Edgeworth

who provided motivation, support, and guidance.

BIOGRAPHY

Anders Sture Johansson was born April 4, 1969 to Anders Daniel Johansson and Elizabeth Walters Johansson in Wilmington, Delaware. In 1970, Anders' family moved from Delaware to North Carolina. His father, Anders Daniel, died in 1979. Anders' mother later married Teak Edgeworth and the family lives in Salisbury, North Carolina. His grandfather, Anders Sture, one brother, and his grandfather's three cousins were raised on a farm in Sardol, Sweden where two of the three cousins still farm today. Anders has a grandmother who lives in Salisbury and two step-brothers: Reid, the elder lives in Peach Tree City, Georgia, and Stan, the youngest lives in Boone, North Carolina.

Anders attended Hurley Elementary, Knox Middle, and Salisbury High Schools in Salisbury. Anders continued his education at North Carolina State University in Raleigh, North Carolina. In May 1991, Anders received a Bachelor of Science Degree from North Carolina State University in Biological and Agricultural Engineering. Following graduation, he enrolled in a Masters of Science program at North Carolina State University working with the United States Department of Agriculture's Agriculture Research Service under the supervision of Dr. T. B. Whitaker in peanut quality. He graduated in December 1993. Anders remained at North Carolina State University to pursue a Ph. D. degree working in corn quality under the guidance of Dr. T. B. Whitaker. He became a Ph. D. candidate in June of 1996.

ACKNOWLEGDEMENTS

I would like to express my appreciation to the following people who have contributed, directly or indirectly, to this study.

Dr. T. B. Whitaker for his guidance, encouragement, support, and especially patience throughout this study. He treats me as if I was one of his own children, except when reviewing my papers.

Dr. F. G. Giesbrecht for his guidance to help answer many of the statistical questions I have had. I'm not sure what I will do when he fully retires!

Dr. W. M. Hagler, Jr., Faye Suggs and Hunter Edwards for conducting the analytical procedures and for helping me to understand the details of the analytical methods involved in my study.

Dr. J. H. Young for his guidance and suggestions throughout this study.

Andrew Slate because he has been a daily part of this study from the first day. Andrew has helped with everything from debugging SAS programs to riffle dividing samples. He has also been a very good friend and listener.

United States Department of Agriculture's Agriculture Research Service for the grant money to conduct this study.

Mary Allison Ford for her undying support and praise throughout both, my Master's Thesis and Ph. D. Dissertation.

My grandmother, parents, and the rest of the family for all of their support and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GENERAL INTRODUCTION

Aflatoxin is a naturally occurring mycotoxin that has been proven toxic and carcinogenic (1). This toxin was first discovered in the 1960's when thousands of turkey poults died (1). The deaths of the turkey poults were traced to aflatoxin contaminated feed.

Aflatoxin is mainly produced by two fungi, Aspergillus flavus and Aspergillus parasiticus (2). These fungi can easily invade agricultural commodities when environmental conditions, such as temperature and moisture, are conducive.

The synthesis of aflatoxin in corn by fungi is a potential threat to animal and human health. The U. S. Food and Drug Administration (FDA) has currently set an aflatoxin guideline of 20 parts per billion (ppb) in products for all commodities destined for human consumption (3). Regulatory guidelines that define the maximum concentration of aflatoxin allowable in food and feeds have been established in more than 90 countries (3). To ensure that consumer-ready products meet FDA aflatoxin guidelines, commodity industries and manufacturers use aflatoxin-sampling plans or sample acceptance schemes to either accept or reject a bulk lot based on the lot’s estimated aflatoxin concentration. A sampling plan is defined as an aflatoxin test procedure combined with a sample acceptance limit. The test procedure consists of sampling, sample preparation, and analytical steps. The sample acceptance limit is a threshold concentration that may or may not be equal to the regulatory guideline. If the sample test result is less than or equal to the sample acceptance limit, the bulk lot is accepted; otherwise, it is rejected.

Aflatoxin inspection and sampling programs are usually developed by commodity industries to meet FDA requirements. Processing plants voluntarily test domestic lots of shelled corn, and some states offer voluntary aflatoxin testing programs. Commodity industries also use aflatoxin sampling plans to classify bulk lots into aflatoxin categories to decrease the amount of aflatoxin contaminated corn entering the food chain. In addition, the Federal Grain Inspection Service (FGIS) tests all lots of shelled corn for aflatoxin that are destined for export. The ability to evaluate any aflatoxin sampling plan will provide a means to design the most efficient sampling plan for the resources available.

Estimating the true amount of aflatoxin in a lot of shelled corn is difficult because of the distribution of contaminated kernels in a lot. A small percentage of the kernels are contaminated and some contaminated kernels have extremely high concentrations of aflatoxin (4). Studying the variability associated with testing samples of shelled corn for aflatoxin will provide a base for statistically measuring the effectiveness of sampling plans. Each component of the total variance (sampling, sample preparation, and analysis) must be investigated to show the amount of variance associated with each step of the testing procedure. The variability associated with the testing procedure leads to some lots being misclassified by the sampling plan. The frequency of misclassifications depends upon the regulatory guideline, sample acceptance level and the design of the testing procedure. The degree to which misclassifications will occur can be evaluated with the help of an operating characteristic (OC) curve.

To produce an OC curve, the observed distribution of sample aflatoxin test results must either be measured or adequately predicted by a theoretical distribution. Methods to

predict the parameters of the theoretical distribution from the observed distribution must also be developed.

Once OC curves are produced, they can be used to predict the probability that a lot of shelled corn at a specific aflatoxin concentration will be accepted by a specified sampling plan. The percentages of lots with a given lot concentration that will be accepted and rejected by the specified sampling plan gives an indication of the misclassification errors associated with the sampling plan. The ability to evaluate the performance of an aflatoxin-sampling plan by using an OC curve provides a method to estimate the costs involved with different sampling plans and helps identify the most efficient plan for the resources available.

The purpose of the research was to evaluate the effectiveness of aflatoxin sample plans to correctly classify lots of shelled corn. To accomplish this overall goal, three specific objectives had to be completed:

1. Estimate the variance components associated with testing shelled corn for aflatoxin.
2. Determine a suitable theoretical model to simulate observed distributions of aflatoxin test results in shelled corn.
3. Evaluate aflatoxin sampling plans for shelled corn.

The results of these three objectives are presented as three manuscripts. Each manuscript focuses on a specific objective. The first manuscript estimates the variance associated with each step of the aflatoxin testing procedure and produces relationships between each variance component and aflatoxin concentration. The second manuscript determines a suitable theoretical model to simulate the observed distribution of aflatoxin test results and methods to easily estimate the parameters of the theoretical distribution.

The third manuscript evaluates the effectiveness of aflatoxin sampling plans where different sample sizes and difference acceptance levels are utilized.

References

1. Rodricks, J. V., and H. R. Roberts. 1977. Mycotoxin Regulation in the United States. Pages 753-757 in Mycotoxins: In Human and Animal Health. J. V. Rodricks, C. W. Hesseltine, M. A. Mehlman. (eds). Illinois: Pathotox.

2. Diener, U. L., R. E. Pettit, and R. J. Cole. 1982. Chapter 13: Aflatoxins and Other Mycotoxins in Peanuts in Peanut Science and Technology. H. E. Pattee, and C. T. Young (eds). Yoakum, Texas: American Peanut Research and Education Society, Inc.

3. Food and Agriculture Organization. 1997. Worldwide regulations for mycotoxins 1995. *Food and Nutrition Paper 64*, 4.

4. Curculu, A. F., L. S. Lee, R. Y. Mayne, and L. A. Goldblatt. 1966. Determination of aflatoxins in individual peanuts and peanut sections. *Journal of American Oil Chemists' Society* 43(2):89-92.

CHAPTER 1. ESTIMATION OF VARIANCE COMPONENTS ASSOCIATED WITH TESTING SHELLED CORN FOR AFLATOXIN

ABSTRACT

The variability associated with testing lots of shelled corn for aflatoxin was investigated in this study. Eighteen lots of shelled corn were tested for aflatoxin contamination. The total variance associated with testing shelled corn was estimated and partitioned into sampling, sample preparation, and analytical variances. All variances were found to increase with an increase in aflatoxin concentration. Using regression analysis, mathematical expressions were developed to model the relationship between aflatoxin concentration and the total, sampling, sample preparation, and analytical variances. The expressions for these relationships were used to estimate the variance for any sample size, subsample size, and number of analyses for a specific aflatoxin concentration. Testing a lot with 20 parts per billion (ppb) aflatoxin using a 2.5 lb sample, Romer mill and 50 g subsample, and HPLC analysis, the total, sampling, sample preparation, and analytical variances are 274.9 (CV=82.9%), 214.0 (CV=73.1%), 56.3 (CV=37.5%), and 4.6 (CV=10.7%), respectively. The percentage of the total variance for sampling, sample preparation, and analytical is 77.8, 20.5, and 1.7 %, respectively.

INTRODUCTION

Aflatoxin is a naturally occurring mycotoxin that has been proven toxic and carcinogenic (1). This toxin was first discovered in the 1960's when thousands of turkey

poults died (1). The deaths of the turkey poults were traced to aflatoxin contaminated feed.

Aflatoxin is mainly produced by two fungi, Aspergillus flavus and Aspergillus parasiticus (2). These fungi can easily invade agricultural commodities when environmental conditions, such as temperature and moisture, are conducive.

The synthesis of aflatoxin in corn and its products is a potential threat to animal and human health. The U. S. Food and Drug Administration (FDA) has currently set an aflatoxin guideline of 20 parts per billion (ppb) in products for all commodities destined for human consumption (3).

An aflatoxin testing procedure consists of three steps: a) sampling, b) sample preparation and c) analysis. Aflatoxin inspection and sampling programs are usually developed by commodity industries to meet FDA legal limits. Processing plants voluntarily test domestic lots of shelled corn. Some states offer voluntary aflatoxin testing programs. In addition, the Grain Inspection, Packers and Stockyards Administration's (GIPSA) Federal Grain Inspection Service (FGIS) tests all lots of shelled corn for aflatoxin that are destined for export. The FGIS currently uses 908 g (2 lb) representative test samples for trucks, 1362 g (3 lb) representative test samples for railcars, and 4540 g (10 lb) representative test samples for barges. The test samples are comminuted in a Romer Mill and a 500 g portion is partitioned out. Finally, 50 g subsamples are removed from the 500 g sample for analysis. The aflatoxin in the 50 g subsamples is extracted using solvents such as methanol-water. Aflatoxin in the solvent is quantified by various methods such as high performance liquid chromatography (HPLC) and immunoassay (4).

Combining a threshold or a sample acceptance level (ppb) with a testing procedure produces a sampling plan. Some sampling plans use a sample acceptance level below the FDA legal limit of 20 ppb to insure that a finished product will meet FDA requirements.

Estimating the true amount of aflatoxin in a lot of shelled corn is difficult because of the distribution of contaminated kernels in a lot. Using peanuts, Cucullu et al. (5) showed that a small percentage of peanut pods are contaminated and some contaminated pods have extremely high concentrations of aflatoxin. It is assumed that aflatoxin in shelled corn behaves in the same manor. Studying the variability associated with testing samples of shelled corn for aflatoxin will provide a base for statistically measuring the effectiveness of sampling plans. Each component of the total variance (sampling, sample preparation, and analysis) was investigated to show which step of the testing procedure contributes the most variability (6, 7).

In 1979, Whitaker et al. (7) measured the variability associated with shelled corn for aflatoxin. Whitaker developed artificial mini-lots by combining 1 kg samples from 400 different commercial lots and subdividing into 10 small lots weighing 40 kg each. Whitaker assumed the distribution of aflatoxin among contaminated corn kernels in the artificial mini-lots was typical of the distribution found in commercial lots of corn. In addition, because the negative binomial distribution had been used to describe aflatoxin contaminated kernels in shelled peanuts and cottonseed, Whitaker assumed that it was a suitable distribution to describe aflatoxin contaminated kernels in shelled corn. By assuming the shape parameter, k, of the negative binomial was linearly related to aflatoxin concentration, a functional relationship between aflatoxin concentration and variance components could be determined by the following equations:

Equation 1-1

$$k = aC$$

Equation 1-2

$$\sigma^2 = \frac{C^2}{k} + C$$

where *k* is the shape parameter of the negative binomial distribution, *a* is the coefficient of the regression analysis, *C* is the lot concentration of aflatoxin, and σ^2 is the variance associated with the negative binomial distribution. Substituting Equation 1-1 into Equation 1-2 and simplifying,

Equation 1-3

$$\sigma^2 = (1 + \tfrac{1}{a})C$$

Whitaker showed that the variance is linearly related to the aflatoxin concentration, C.

Using commercial corn lots, the objectives of this study were (a) determine the total variance associated with testing shelled corn for aflatoxin, (b) partition total variance into sampling, sample preparation, and analytical variability components, and (c) determine functional relationships between the variance components and aflatoxin concentration.

EXPERIMENTAL PROCEDURE

Theoretical Considerations

We assumed (a) each lot consists of N individual corn kernels (b) each corn kernel has the same mass and physical characteristics, and (c) that variation of aflatoxin concentration occurs between kernels. With shelled corn, it is common practice to estimate the aflatoxin concentration of a sample of *n* kernels, represented by $\hat{C}$, instead

of analyzing aflatoxin on individual kernels $\hat{C}_i$. In peanuts, Cucullu (5) showed that most individual pods have an aflatoxin concentration of zero, but occasionally, a pod may have an extremely high aflatoxin concentration. It is assumed that aflatoxin in shelled corn behaves in the same manor.

Figure 1-1 shows the relationships between the three major components of the total variation associated with testing shelled corn for aflatoxin: sampling, sample preparation, and analysis.

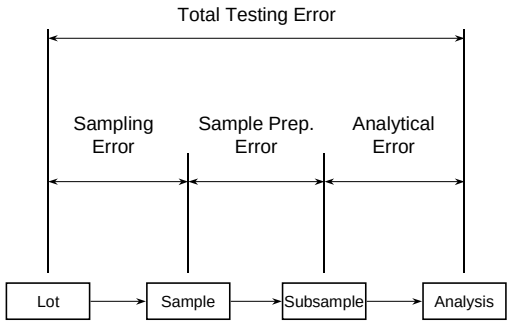


Figure 1-1: Total Variance partitioned into sample, sample preparation, and analytical components.

A statistical model of the aflatoxin test result, $\hat{C}$, can be represented by:

Equation 1-4

$$\hat{C} = \mu + S + SS + A$$

where μ = the true aflatoxin concentration in the lot being tested, S = random deviations of sample concentrations about the true lot concentration with expected value equal to zero and variance $\sigma^2_{\hat{C}(s)}$, SS = random deviations of subsample concentrations about the

comminuted sample concentration with expected value equal to zero and variance $\sigma^2_{\hat{C}(ss)}$, and A = random deviations of analytical assay results about subsample concentration with the expected value zero and variance $\sigma^2_{\hat{C}(a)}$. If independence among the random deviations in Equation 1-4 is assumed, the model for variance can be obtained:

Equation 1-5

$$\sigma^2_{\hat{C}(t)} = \sigma^2_{\hat{C}(s)} + \sigma^2_{\hat{C}(ss)} + \sigma^2_{\hat{C}(a)}$$

where $\sigma^2_{\hat{C}(t)}$ is the total variance associated with the aflatoxin test statistic $\hat{C}$.

Total variance $\sigma^2_{\hat{C}(t)}$ is the estimate of the variance among samples within the same lot of shelled corn, which is specific to mill type, subsample size, number of aliquots, and analytical procedure.

Variance components, $\sigma^2_{\hat{C}(s)}$ and $\sigma^2_{\hat{C}(ss)}$ cannot be measured directly because of the nested design. However, $\sigma^2_{\hat{C}(t)}$, $\sigma^2_{\hat{C}(ssa)}$, and $\sigma^2_{\hat{C}(a)}$ can be measured directly where $\sigma^2_{\hat{C}(ssa)}$ is the combination of sample preparation and analytical variances as shown in Equation 1-6.

Equation 1-6

$$\sigma^2_{\hat{C}(ssa)} = \sigma^2_{\hat{C}(ss)} + \sigma^2_{\hat{C}(a)}$$

Then sampling and sample preparation variances can be calculated by subtraction.

Equation 1-7

$$\sigma^2_{\hat{C}(s)} = \sigma^2_{\hat{C}(t)} - \sigma^2_{\hat{C}(ssa)}$$

Equation 1-8

$$\sigma^2_{\hat{C}(ss)} = \sigma^2_{\hat{C}(ssa)} - \sigma^2_{\hat{C}(a)}$$

The sampling variance, $\sigma^2_{\hat{C}(s)}$, is an estimate of the variability among replicate test samples within each lot of shelled corn. Sample preparation variance $\sigma^2_{\hat{C}(ss)}$ is defined by the variability among replicate subsamples obtained from the same sample comminuted in a suitable mill.

Methods

The first experimental design (see Appendix 1) was an unbalanced nested procedure designed to produce estimates of $\sigma^2_{\hat{C}(t)}$, $\sigma^2_{\hat{C}(ssa)}$, and $\sigma^2_{\hat{C}(s)}$. The notation $s^2_{\hat{C}}$ is denoted as an estimate of $\sigma^2_{\hat{C}}$. A bulk sample weighing approximately 45.4 kg (100 lb) was divided into 32 test samples of 1.13 kg (2.5 lb). Each sample was comminuted in a Romer Mill. Using 50 g subsamples, aflatoxin was extracted with methanol-water (75 + 25, v/v) in a 2:1 ratio. To purify the extract (0.5 ml), it was passed through a Mycosep #224 column (8). The aflatoxins were derivatized using a bromide post-column derivatization process and quantified using HPLC (9).

Two 50 g subsamples were removed from 16 of the 32 samples and one 50 g subsample was removed from the remaining 16 samples. All subsamples were analyzed with a single aliquot per subsample. The unbalanced design was used to keep costs minimal while still providing enough degrees of freedom for adequate variance estimation.

A second experiment was designed to obtain estimates of $\sigma^2_{\hat{C}(a)}$. Ten subsamples were chosen to produce a wide range of aflatoxin concentration. Analytical variance

$\sigma^2_{\hat{C}(a)}$ is the estimate of the variance among 15 replicate aliquots of extract taken from the blender after the extraction process from a single subsample. All aliquot testing was conducted in the same laboratory to produce an analytical variance that reflects within laboratory variance. The results are recorded in total parts per billion (ppb) and contain the sum of aflatoxins B1, B2, G1, and G2.

From the unbalanced nested design, the estimated variance components and the lot aflatoxin concentration associated with each component were determined for each of the 18 lots using the Nested procedure in SAS (10).

RESULTS AND DISCUSSION

Aflatoxin test results for all 18 lots and 32 samples can be seen in Appendix 2 Tables 1-4. Variance estimates were obtained for eighteen lots. Table 1-1 reports aflatoxin concentration, total variance $s^2_{\hat{C}(t)}$, sampling variance $s^2_{\hat{C}(s)}$, and combined sample preparation and analytical variance $s^2_{\hat{C}(ssa)}$ values associated with each contaminated lot of shelled corn. Appendix 3 shows the SAS (10) program used to estimate the variance components. The 18 lots are ranked by aflatoxin concentration, which ranged from about 6 to 677 ppb. In general, as the aflatoxin concentration increases, each variance estimate increases. This reflects the results of similar variance relationship studies conducted on other commodities (6, 11-15).

Table 1-1: Average aflatoxin concentration, sample variance, and combined subsampling and analytical variance components for all 18 lots of shelled corn. **

Lot number	Aflatoxin Concentration ppb	Total Variance	Sample Variance	Subsample + Analytical Variance
1	5.8	77.4	28.2	49.2
2	6.4	121.0	114.7	6.3
3	6.7	150.9	131.8	19.1
4	8.6	149.7	109.4	40.3
5	11.8	203.0	193.0	10.0
6	15.9	353.0	108.4	244.6
7	18.2	194.1	103.9	90.2
8	25.6	413.8	371.9	42.0
9	27.3	590.4	508.2	82.2
10	32.9	557.0	469.5	87.5
11	56.7	370.7	258.9	111.8
12	57.1	887.9	474.8	413.1
13	94.7	1277.4	1106.8	170.5
14	95.6	515.9	444.5	71.5
15	113.8	1452.8	1173.6	279.2
16	276.9	5393.1	2933.3	2459.8
17	298.9	7160.9	4012.7	3148.1
18	676.6	31308.1	9096.1	22212.0

**Testing Plan=1.13kg sample, Romer Mill, 50 g subsample, and 1 aliquot quantified by HPLC.

Sampling Variance

The sampling variance estimates from Table 1-1 show a linear relationship with the mean aflatoxin concentration when plotted in full log scale (Figure 1-2). Therefore, sampling variance was modeled by the following mathematical expression.

Equation 1-9

$$s^2_{\hat{C}(s)} = a\hat{C}^b$$

where $s^2_{\hat{C}(s)}$ is the sample variance component, *a* and *b* are constants determined by regression analysis, and $\hat{C}$ is the estimate of aflatoxin concentration measured in ppb.

Using regression analysis, a relationship between sampling variance and aflatoxin concentration was developed.

Equation 1-10

$$s^2_{\hat{C}(s)} = 11.361\hat{C}^{0.98}$$

with a coefficient of determination of 0.89 in the full log scale.

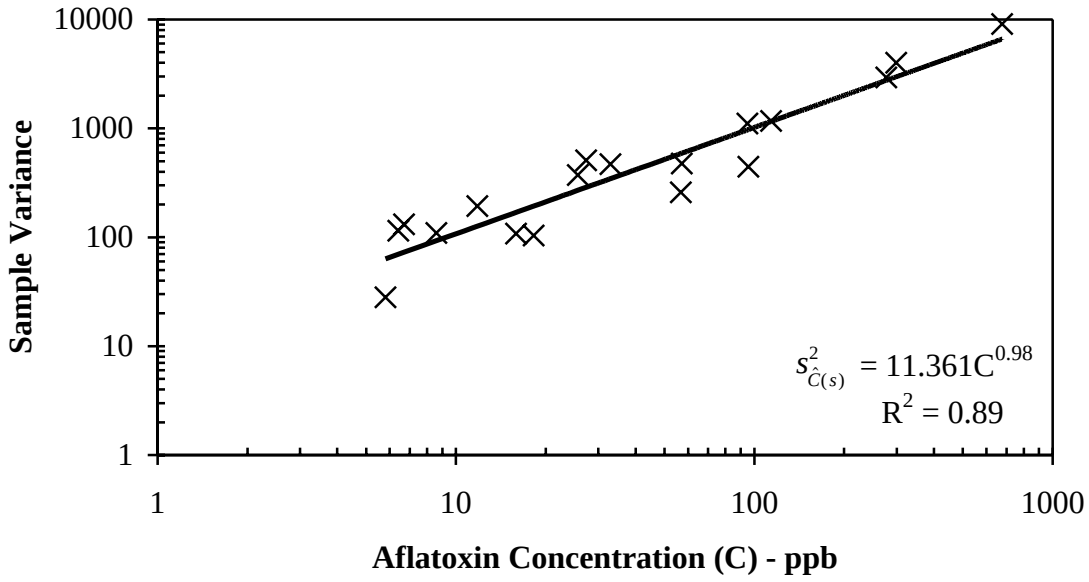


Figure 1-2: Sampling variance versus aflatoxin concentration for 1.13 kg test samples of shelled corn.

Combined Sample Preparation and Analytical Variance

Table 1-1 reports the combined sample preparation and analytical variance estimates $s^2_{\hat{C}(ssa)}$. Figure 1-3 shows the combined sample preparation and analytical variance estimates versus aflatoxin concentration plotted in full log scale. Generally, combined sample preparation and analytical variance estimates increased with an increase in aflatoxin concentration. Using regression analysis, the following relationship

was developed between the combined sample preparation and analytical variance and aflatoxin concentration.

Equation 1-11

$$s^2_{\hat{C}(ssa)} = 1.383\hat{C}^{1.27}$$

with a coefficient of determination of 0.78 in the full log scale.

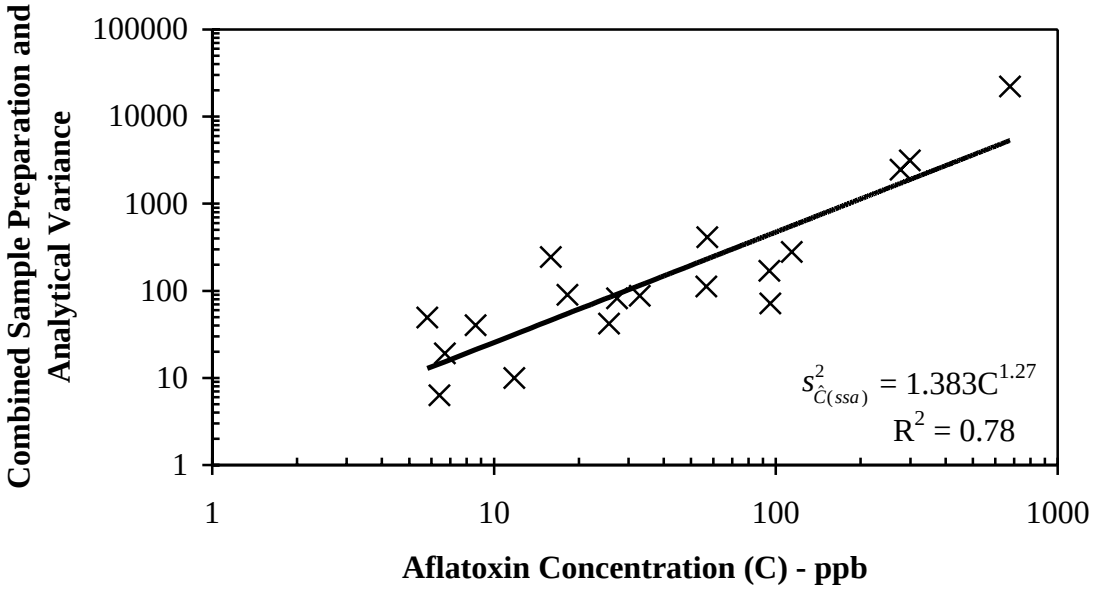


Figure 1-3: Combined sample preparation and analytical variance versus aflatoxin concentration for test subsamples of shelled corn using 50 g subsamples comminuted in the Romer mill, 1 aliquot per subsample, and HPLC.

Analytical Variance

Table 1-2 shows analytical variance estimates $s^2_{\hat{C}(a)}$ among the 15 replicated test results for each of the 10 subsamples analyzed. Generally, as the aflatoxin concentration increased, analytical variance also increased. Figure 1-4 shows the linear relationship

between the analytical variance and the aflatoxin concentration when plotted in the full log scale. Using regression analysis, the following mathematical expression provided a suitable relationship between the analytical variance and aflatoxin concentration.

Equation 1-12

$$s^2_{\hat{C}(a)} = 0.143\hat{C}^{1.16}$$

with a coefficient of determination of 0.92 in the full log scale.

Table 1-2: Average aflatoxin concentration, analytical variance, and coefficient of variation among replicate aflatoxin test results on 15 aliquots quantified by HPLC.**

Subsample Number	Aflatoxin Concentration	Analytical Variance	Coefficient of Variation
1	28.3	9.1	10.6%
2	58.1	20.7	7.8%
3	58.7	24.5	8.4%
4	68.4	14.5	5.6%
5	103.7	22.7	4.6%
6	117.3	15.8	3.4%
7	189.0	63.0	4.2%
8	433.2	230.9	3.5%
9	876.7	266.6	1.9%
10	937.8	608.4	2.6%

** Analytical Procedure = 15 aliquots taken from blender after extraction and quantifed by HPLC.

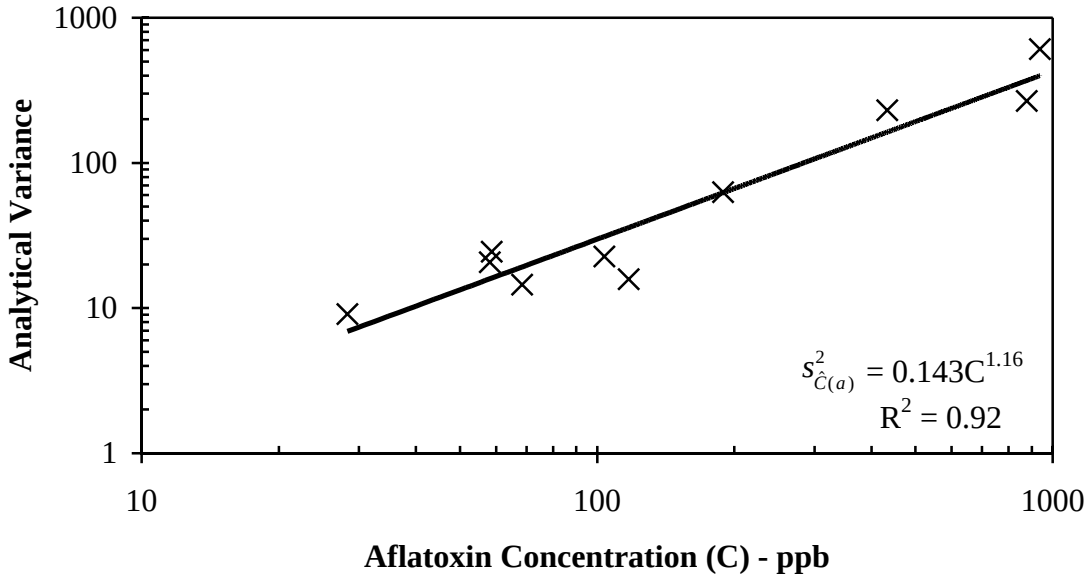


Figure 1-4: Analytical variance versus aflatoxin concentration for test subsamples of shelled corn using 15 aliquots per subsample and HPLC.

Sample Preparation Variance

Once relationships are developed for $s^2_{\hat{C}(ssa)}$ and $s^2_{\hat{C}(a)}$, Equation 1-8 can be used to determine sample preparation variance $s^2_{\hat{C}(ss)}$. An equation to estimate sample preparation variance can be calculated by subtraction of Equation 1-12 from Equation 1-11.

Equation 1-13

$$s^2_{\hat{C}(ss)} = 1.383\hat{C}^{1.27} - 0.143\hat{C}^{1.16}$$

Equation 1-13 can be simplified by regressing the difference, $s^2_{\hat{C}(ss)}$, on aflatoxin concentration $\hat{C}$. A suitable expression is shown in Equation 1-14.

Equation 1-14

$$s^2_{\hat{C}(ss)} = 1.254\hat{C}^{1.27}$$

Application of Results

Equations 1-10 through 1-14 estimate variances associated with testing a lot of shelled corn for aflatoxin using a 1.13 kg sample, Romer Mill, 50 g subsample, and HPLC. Reducing one or more of the three variance components, sampling, sample preparation, or analytical, will reduce the total variance associated with a testing procedure. Statistical theory indicates that an increase in quantity of material tested will decrease variance associated with that step of the testing procedure. For example, increasing sample size or number of sample units can reduce sampling variance; increasing subsample size or number of subsample units can reduce sample preparation variance; and increasing the size of the aliquot or number of aliquots taken from the blender after the extraction process to be quantified by HPLC can reduce analytical variance.

Equation 1-10 can be modified to predict the sampling variance for a given sample size.

Equation 1-15

$$s^2_{\hat{C}(s)} = \left(\frac{1.13}{ns}\right) \cdot 11.361\hat{C}^{0.98}$$

where *ns* is the sample size in kg.

The sample preparation variance in Equation 1-14 can be modified to predict the effect of any size subsample comminuted in the Romer mill.

Equation 1-16

$$s^2_{\hat{C}(ss)} = \left(\frac{50}{nss}\right) \cdot 1.254 \hat{C}^{1.27}$$

where *nss* is the subsample size in g.

A similar expression can be derived for the analytical variance described in Equation 1-12. Modification of Equation 1-12 shows the effect of any number of aliquots quantified by HPLC.

Equation 1-17

$$s^2_{\hat{C}(a)} = \left(\frac{1}{na}\right) \cdot 0.143 \hat{C}^{1.16}$$

where *na* is the number of aliquots.

Total variance can be estimated for any sample size, subsample size comminuted in a Romer Mill and number of aliquots quantified by HPLC by summing Equations 1-15 through 1-17.

Equation 1-18

$$s^2_{\hat{C}(t)} = \left[\left(\frac{12.95}{ns}\right)\hat{C}^{0.98}\right] + \left[\left(\frac{62.70}{nss}\right)\hat{C}^{1.27}\right] + \left[\left(\frac{0.143}{na}\right)\hat{C}^{1.16}\right]$$

With Equation 1-18, the total variance associated with testing a contaminated lot of shelled corn for aflatoxin at 20 ppb using a 1.13 kg sample, Romer Mill, 50 g subsamples, and quantifying 1 aliquot per subsample using HPLC is 274.9 (CV=82.9%). Sampling, sample preparation, and analytical variances are 214.0 (CV=73.1%), 56.3 (CV=37.5%), and 4.6 (CV=10.7%), respectively and account for 77.8, 20.5, and 1.7 % of the total variation. Sampling variance accounts for the majority of the total variance,

with sample preparation producing the next largest amount, and analytical with the least. This follows the same pattern observed with other commodities (6, 11-15).

The effect of increasing sample size on reducing testing variability can be demonstrated with Equation 1-18. Testing a contaminated lot of shelled corn for aflatoxin at 20 ppb using a 5 kg sample, Romer Mill, 100 g subsamples, and quantifying 1 aliquot per subsample using HPLC produces a total variance value of 81.6 (CV=45.2%). Sampling, sample preparation, and analytical variances are 48.8 (CV=34.9%), 28.2 (CV=26.5%), and 4.6 (CV=10.7%), respectively and account for 59.8, 34.5, and 5.7 % of the total variation.

Assuming that aflatoxin test results from shelled corn follow the theory of normally distributed variables, a lot with an aflatoxin concentration of 20 ppb and a total variance of 81.6 implies that aflatoxin test results will fall in the range of 20 ± 18 ppb, or 2-38 ppb, 95% of the time.

Whitaker's (7) study used similar sampling techniques, but different sample preparation and analytical methods than this study. Only sampling variances can be compared between the two studies. Figure 1-5 shows that Whitaker's sampling variance is considerably less than in this study. This suggests Whitaker's assumption that his mini-lots were representative of commercial lots may be incorrect. In Figure 1-5, the slopes of the two lines are almost parallel, which implies that Whitaker's assumption about a linear relationship between aflatoxin concentration and sampling variance may be correct.

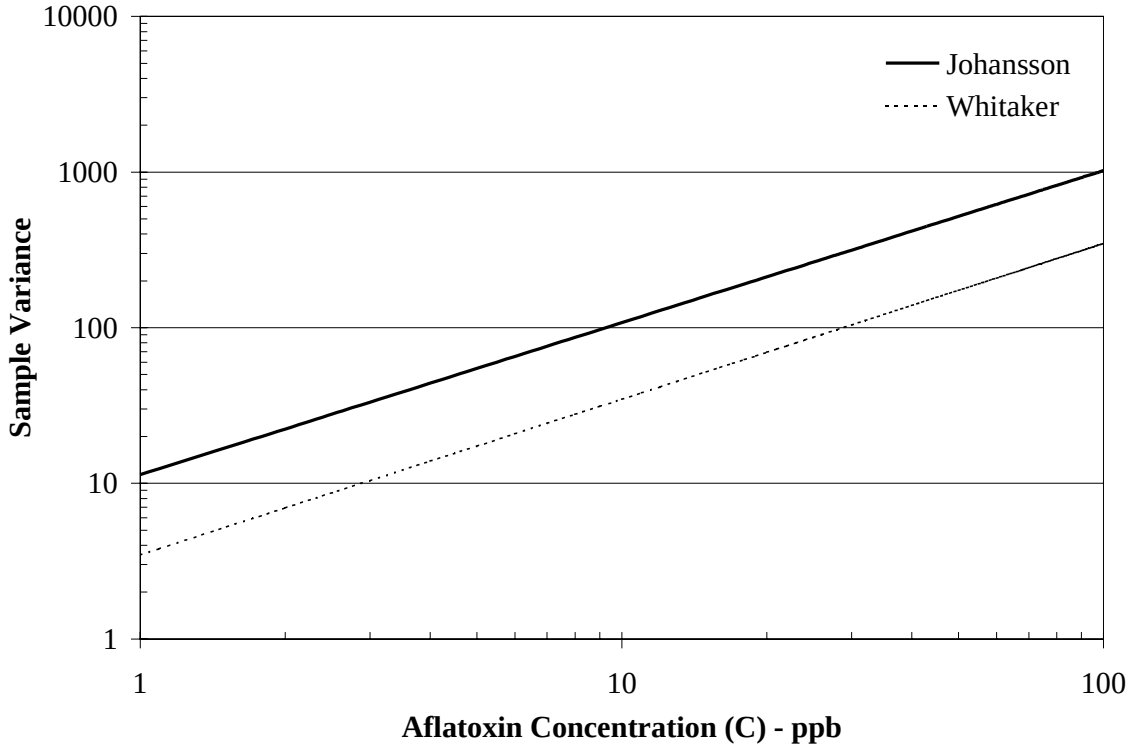


Figure 1-5: Comparison of sampling variances for Whitaker's 1979 study to Johansson's 1998 study of variability associated with testing shelled corn for aflatoxin.

The next step would be to calculate which theoretical distribution describes the sample distribution of shelled corn data in this study.

SUMMARY

Eighteen lots of shelled corn were tested for aflatoxin using an unbalanced nested design. Estimates of the total variability associated with testing shelled corn for aflatoxin were shown to increase as aflatoxin concentration increased. This also held true for each step of the test procedure: sampling, sample preparation, and analytical variability. Using regression analysis, mathematical expressions were developed to model all three variance components. The expressions were used to estimate the variance for any sample size, subsample size, and number of analyses for a specific aflatoxin concentration. For

example, testing a lot with 20 parts per billion (ppb) aflatoxin using a 2.5 lb sample, Romer mill and 50 g subsample, and HPLC analysis, the total, sampling, sample preparation, and analytical variances are 274.9 (CV=82.9%), 214.0 (CV=73.1%), 56.3 (CV=37.5%), and 4.6 (CV=10.7%), respectively. The percentage of the total variance for sampling, sample preparation, and analytical is 77.8, 20.5, and 1.7 %, respectively. As with testing of aflatoxins in other commodities, sampling variance contributes the most variability followed by sample preparation and then analytical variability.

REFERENCES

1. Rodricks, J. V., and Roberts, H. R. 1977. Mycotoxin Regulation in the United States. Pages 753-757 in Mycotoxins: In Human and Animal Health. Rodricks, J. V., Hesseltine, C. W., Mehlman, M. A. (eds). Illinois: Pathotox.

2. Diener, U. L., R. E. Pettit, and R. J. Cole. 1982. Chapter 13: Aflatoxins and Other Mycotoxins in Peanuts in Peanut Science and Technology. H. E. Pattee, and C. T. Young (eds). Yoakum, Texas: American Peanut Research and Education Society, Inc.

3. Anonymous. 1997. Worldwide regulations for mycotoxins 1995. *FAO Food and Nutrition Paper 64.* FAO Viale della Terme di Caracalla. 00100 Rome, Italy.

4. Marshall, J. W. 1992. U. S. Department of Agriculture Federal Grain Inspection Service Aflatoxin Handbook. Sec. 4.2-4.8.

5. Cucullu, A. F., L. S. Lee, R. Y. Mayne, and L. A. Goldblatt. 1966. Determination of aflatoxins in individual peanuts and peanut sections. *Journal of American Oil Chemists' Society* 43(2):89-92.

6. Whitaker, T. B., F. E. Dowell, W. M. Hagler, Jr., F. G. Giesbrecht, and J. Wu. 1994. Variability Associated with Sampling, Sample Preparation, and Chemical Testing for Aflatoxin in Farmer's Stock Peanuts. *Journal of AOAC International.* 77:107-116.

7. Whitaker, T. B., J. W. Dickens, and R. J. Monroe. 1979. Variability Associated with Testing Corn for Aflatoxin. *Journal of the American Oil Chemists' Society.* 56:789-794.

8. Trucksess, M. W., M. E. Stack, S. Nesheim, R. H. Albert, and T. R. Romer. 1994. Multi functional column coupled with liquid chromatography for determination of

aflatoxins B1, B2, G1, and G2 in corn, almonds, Brazil nuts, peanuts, and pistachio nuts. *Journal of AOAC International.* 77:1512-1521.

9. Traag, W. A., J. M. P. Van Trijp, L. G. M. Th. Tuinstra, and W. Th. Kok. 1987. Sample clean-up and post-column derivatization for the determination of aflatoxin B1 in feedstuffs by liquid chromatography. *J. Chromatogr.* 396:389-394.

10. Statistical Analysis System Institute, Inc. 1996. SAS Program 6.12. Cary, NC 27513.

11. Whitaker, T. B., W. Horwitz, R. Albert, and S. Nesheim. 1996. Variability Associated with Analytical Methods Used to Measure Aflatoxin in Agricultural Commodities. *Journal of AOAC International.* 79:476-485.

12. Cambell, A. D., T. B. Whitaker, A. E. Pohland, J. W. Dickens, and D. L. Park. 1986. Sampling, sample preparation, and sampling plans for foodstuffs for mycotoxin analysis. *Pure and Appl. Chem.* Vol. 58, pp. 305-314.

13. Whitaker, T. B., M. E. Whiten, and R. J. Monroe. 1976. Variability Associated with Testing Cottonseed for Aflatoxin. *Journal of the American Oil Chemists' Society.* Vol. 53, No. 7, pp. 502-505.

14. Whitaker, T. B., J. W. Dorner, F. E. Dowell, and F. G. Giesbrecht. 1992. Variability Associated with Chemical Testing Screened Farmer's Stock Peanuts for Aflatoxin. *Peanut Science.* 19:88-91.

15. Whitaker, T. B., J. W. Dickens, and R. J. Monroe. 1974. Variability of Aflatoxin Test Results. *Journal of the American Oil Chemists' Society.* 51:214-218.

CHAPTER 2. DETERMINATION OF A SUITABLE STATISTICAL MODEL TO SIMULATE OBSERVED DISTRIBUTIONS OF AFLATOXIN TEST RESULTS IN SHELLED CORN

ABSTRACT

Suitability of several theoretical distributions to predict the observed distribution of aflatoxin test results in shelled corn is investigated in this study. Fifteen positively skewed distributions were each fitted to 18 empirical distributions of aflatoxin test results for shelled corn. The compound gamma distribution was selected to model the sample aflatoxin test results for shelled corn. The method of moments technique was chosen to estimate the parameters of the compound gamma distribution. Mathematical expressions were developed to calculate the parameters of the compound gamma distribution for any lot aflatoxin concentration and test procedure. Observed acceptance probabilities were compared to operating characteristic curves predicted from the compound gamma distribution and all 18 observed acceptance probabilities were found to lie within a 95% confidence band.

The parameters of compound gamma were used to calculate the fraction of aflatoxin-contaminated kernels in a lot at 20 ppb, which was estimated to be about six in 10,000.

INTRODUCTION

Regulatory guidelines that define the maximum concentration of aflatoxin allowable in food and feeds have been established in more than 90 countries (1). In the United States, the Food and Drug Administration (FDA) has a regulatory guideline of 20

parts per billion (ppb) for foods destined for human consumption. To ensure that consumer-ready products meet FDA aflatoxin guidelines, commodity industries and manufacturers use aflatoxin-sampling plans or sample acceptance schemes to either accept or reject a bulk lot based on the lot's estimated aflatoxin concentration. A sampling plan is defined as an aflatoxin test procedure combined with a sample acceptance limit. The test procedure consists of sampling, sample preparation, and the analytical steps. The sample acceptance limit is a threshold concentration that may or may not be equal to the regulatory guideline. If the sample test result is less than or equal to the sample acceptance limit, the bulk lot is accepted; otherwise, it is rejected.

Risks

Ideally, a lot with aflatoxin concentration, C, greater than a defined guideline, C_g, should be rejected from entering the food chain, and a lot with C less than or equal to C_g should be accepted. In reality, the lot aflatoxin concentration C is estimated by quantifying the aflatoxin $\hat{C}$ in a sample taken from the lot. The values of $\hat{C}$ for a specified lot will differ from C because of random variation associated with sampling (only a sample of kernels is tested), sample preparation and the analytical procedure (2-7). The discrepancy between C and $\hat{C}$ lead to some lots being misclassified.

Normally lots of shelled corn are inspected with an aflatoxin-sampling plan and are classified either as unacceptable when the sample test result, $\hat{C}$ is greater than the sample acceptance limit C_a, or acceptable when $\hat{C}$ is less than or equal to the sample acceptance limit. Note that in practice C_a and C_g do not have to be equal.

In a sampling plan, two types of misclassification can occur. The first type of misclassification, a false positive, occurs when $\hat{C} > C_a$ and the lot is rejected when in reality, $C \leq C_g$ or the lot is acceptable. This is also called the seller's risk because a good lot would be diverted from the food chain at an expense to the seller. The second type of misclassification, a false negative occurs when $\hat{C} \leq C_a$ and the lot is accepted when in reality $C > C_g$ that is, the lot is not acceptable. This misclassification is known as a buyer's risk because a bad lot may require additional cleaning and processing to lower the lot aflatoxin concentration level. Also, if the unacceptable lot goes undetected, the lot enters the food chain and becomes a potential health risk to the consumer. The frequency of these two misclassifications occurring depends upon C_g, C_a, and the design of the sampling plan. The degree to which these two misclassifications will occur can be evaluated with the help of an operating characteristic curve.

Operating Characteristic Curve

Let $P\{C\}$ be the probability that a lot of shelled corn with an aflatoxin concentration C will be accepted by a specified sampling plan. Clearly, any reasonable sampling plan will give a large probability of accepting a lot with a small C and a small (hopefully zero) probability of accepting a lot with a large C. An operating characteristic (OC) curve is simply a plot of $P\{C\}$ versus C. Figure 2-1 shows a typical OC curve. An OC curve is uniquely defined by a specific testing plan and sample acceptance limit (6, 8-14). In most sampling plans, $C_a \leq C_g$.

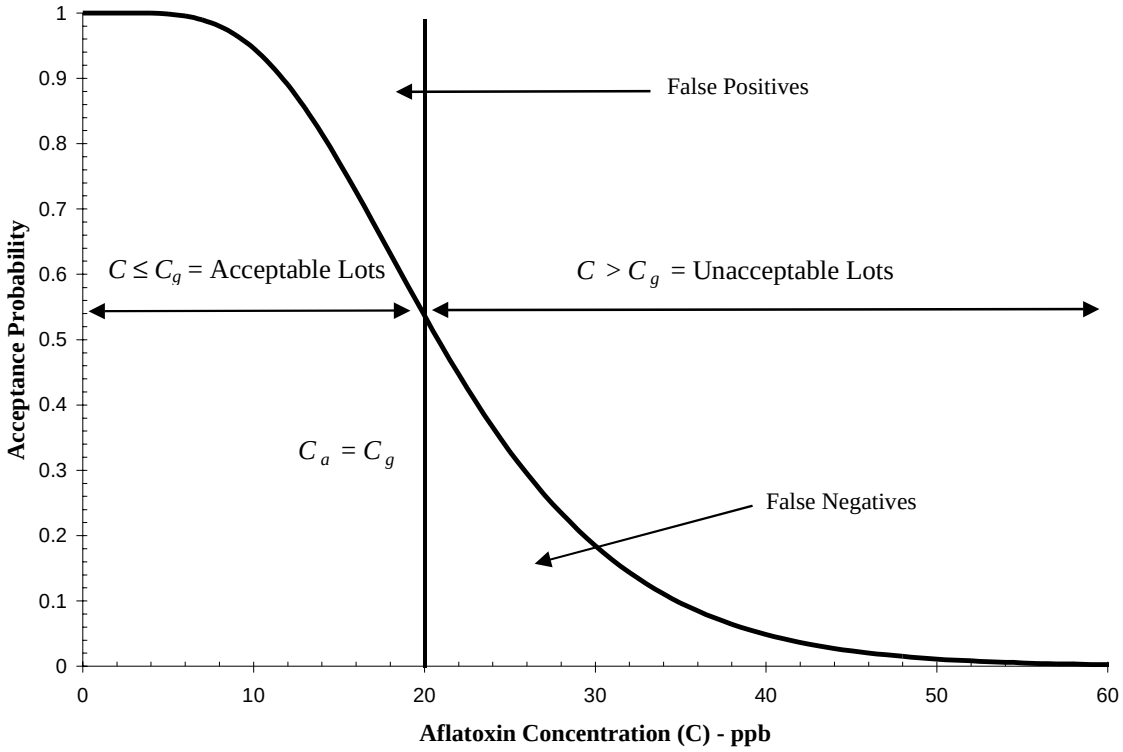


Figure 2-1: Typical operating characteristic curve showing the performance of an aflatoxin sampling plan when testing lots with aflatoxin concentration C.

Knowing the percentage of lots with a given lot concentration *C* that will be accepted and rejected by the specified sampling plan gives an indication of the misclassification errors associated with the sampling plan. The ability to evaluate the performance of an aflatoxin-sampling plan will provide a method to estimate the costs involved with different sampling plans and help identify the most efficient plan for the resources available.

Theoretical Distribution

In order to determine an OC curve, the observed distribution of sample aflatoxin test results must either be measured or adequately described by a theoretical distribution. When modeling the observed distribution of aflatoxin test results, a theoretical

distribution should be able to (1) accurately describe the observed distribution of aflatoxin test results for a specific test procedure over a wide range of aflatoxin concentrations, (2) contain parameters that are easily computed from experimental data and (3) be able to be applied to practical situations.

The distribution of sample aflatoxin test results for different commodities has been studied by several different scientists (3, 8, 15-29). Shelled corn, cottonseed, and pistachio nuts have been investigated, but the majority of studies have been on peanuts.

These studies have shown similar characteristics about the observed distribution of sample aflatoxin test results for the above commodities: (a) aflatoxin sample test results taken from contaminated lots were positively skewed (>50% of sample test results were below the lot concentration); (b) for a given lot concentration, the range between minimum and maximum sample test results is large which indicates a large variance associated with the test procedure; (c) the range between minimum and maximum increases as aflatoxin concentration increases which suggests that the variance associated with sample test results increases with aflatoxin concentration; and (d) as aflatoxin concentration increases, the distribution of sample test results becomes less skewed and more like the normal distribution (6).

The following distributions have been investigated as possible models of observed distributions of sample aflatoxin test results: negative binomial (6, 15, 21, 22), compound gamma (21-23), log normal (23, 24), truncated normal (23), Waibel (25), and three-parameter Weibull (26-28).

Because several different distributions have been suggested to model the observed distribution of aflatoxin test results, the objective of this study was to (1) compare the

suitability of several theoretical distributions to accurately predict observed distributions of sample aflatoxin test results for shelled corn, (2) develop methods to estimate the parameters of the theoretical distribution and (3) compare observed and predicted OC curves to evaluate how accurately OC curves can be determined from theoretical models.

METHODS

Theoretical Distributions

Four distributions were evaluated: the compound gamma, negative binomial, three-parameter log normal and truncated normal. See Appendix 4 for the probability density functions associated with each distribution. Several cases of the compound gamma distribution were evaluated where the shape parameter α was unrestricted (CG), set equal to 0.5 (CG0.5), 1 (CG1), 1.5 (CG1.5), 2 (CG2), and 2.5 (CG2.5). The compound gamma, negative binomial, and truncated normal have parameters that relate to a sample size term in the density function to allow the distribution to be adjusted for different sample sizes (i.e. different test procedures). The three-parameter log normal density function does not contain a sample size term, but was considered because of its ability to fit positively skewed data.

Parameter Estimation Methods

Method of moments (MM) and maximum likelihood (ML) are the two parameter estimation methods used in this study. The MM technique provides an easy and intuitive method of estimation, but does not always lead to the best estimators of distributional parameters. The ML method tends to produce good estimators, but these often can be very difficult to compute (30).

The ML method was used to estimate the parameters for all distributions. For comparison, the method of moments was also used to estimate parameters of the compound gamma and negative binomial distributions. SAS (31) procedures were used to calculate estimates under both methods (See Appendix 5).

Goodness of Fit

The power divergence (PD) (32) test statistic was selected as the criterion to evaluate the goodness of fit (GOF) of the theoretical distributions to the observed distributions. The PD statistic was selected because this test is generally found to have reasonable power against a broad range of alternatives (32). The test statistic is actually very similar to the familiar chi-square goodness of fit test statistic. The fit was considered acceptable if the test statistic failed to exceed the 95% critical value. The test statistic was converted to a GOF probability where the lower the GOF probability, the better the fit.

Observed Distribution

Bulk samples weighing approximately 45.4 kg (100 lb) were taken from each of 18 lots of shelled corn suspected of aflatoxin contamination. Each bulk sample was divided in 32 test samples of 1.13 kg (2.5 lb) and comminuted in the Romer Mill. A 50 g subsample was removed from each comminuted sample. Each subsample was analyzed using high performance liquid chromatography (HPLC) (7). This provided 18 observed distributions with 32 sample test results per distribution (See Appendix 6). An observed distribution refers to the empirical cumulative density function of sample test results.

RESULTS

Goodness of Fit

Table 2-1 shows the frequency that each theoretical distribution acceptably fit the 18 observed distributions. The three-parameter log normal and six cases of the compound gamma distribution gave acceptable fits 100% of the time. The compound gamma fit all 18 observed distributions using both the MM and ML parameter estimation methods. Using the MM technique, the compound gamma distribution provided 100% acceptable fits for CG1MM (compound gamma distribution with $\alpha = 1$ using method of moments), CG1.5MM, CG2MM, and CG2.5MM. When using the ML technique, the compound gamma distribution produced 100% acceptable fits for the CG2ML (compound gamma distribution with $\alpha = 2$ using maximum likelihood) and CG2.5ML. The negative binomial using MM gave the fewest acceptable fits.

Table 2-1: The number of times each theoretical distribution provided an acceptable fit to the 18 observed distributions of sample test results using the power divergence test.

Distribution[a]	Parameter Estimation Method[b]	Acceptable Fit[c]	
		Number	%
Three-Parameter Log Normal	ML	18	100%
CG1	MM	18	100%
CG1.5	MM	18	100%
CG2	MM	18	100%
CG2.5	MM	18	100%
CG2	ML	18	100%
CG2.5	ML	18	100%
CG1	ML	17	94%
CG1.5	ML	17	94%
CG	ML	17	94%
Truncated Normal	ML	15	83%
CG0.5	MM	14	78%
CG0.5	ML	14	78%
Negative Binomial	ML	10	56%
Negative Binomial	MM	6	33%

[a] CG = compound-gamma;
CG0.5 = special case of the compound-gamma where shape parameter = 0.5;
CG1 = special case of the compound-gamma where shape parameter = 1;
CG1.5 = special case of the compound-gamma where shape parameter = 1.5;
CG2 = special case of the compound-gamma where shape parameter = 2;
CG2.5 = special case of the compound-gamma where shape parameter = 2.5;

[b] ML = maximum likelihood
MM = method of momemnts

[c] Acceptable Fit = null hypothesis not rejected at the 5% confidence level.

The negative binomial distribution has a problem predicting the percent zero aflatoxin sample values when a lot contains a high percentage of zero sample values. The negative binomial distribution does however model the positive aflatoxin test results adequately. The compound gamma and the three-parameter log normal distributions were able to model the positively skewed distributions with high percent zero aflatoxin sample values.

The three-parameter log normal cannot be statistically modified to predict the distribution for different sample sizes. If the log normal distribution were to be used as the model, extensive lab testing would be required to determine how the distributional

parameters change as the sample size is varied. Statistically incorporating a sample size term into the log normal distribution was investigated, but no acceptable results were found. The remaining theoretical distributions investigated in this study can be modified to predict the distribution of sample test results for sample sizes other than that used in this study (15, 23).

In Table 2-1, four out of the six compound gamma distributions that acceptably fit all 18 observed distributions were fitted using the MM as the parameter estimation technique. Only these four theoretical distributions were considered since the MM provides GOF results equivalent to the ML technique and has simpler computational formulas to adjust for changes in testing procedure.

The four compound gamma distributions (CG1MM, CG1.5MM, CG2MM and CG2.5MM) where parameters were estimated using the MM technique were evaluated to determine which of these four distributions had the most best fits. A best fit is defined as the theoretical distribution that fit an observed distribution with the lowest GOF probability. Table 2-2 shows the percentage of best fits each of the theoretical distributions produced. The CG2.5MM tied with CG2MM with the highest number of best fits followed by CG1.5MM and CG1MM.

Table 2-2: Number of times each theoretical distribution provided a best fit to the 18 observed distributions of sample test results when using the power divergence test.

Distribution[a]	Parameter Estimation Method[b]	Best Fit[c] Number[d]	%
CG2.5	MM	10	56%
CG2	MM	10	56%
CG1.5	MM	8	44%
CG1	MM	7	39%

[a] CG1 = special case of the compound-gamma where shape parameter = 1;
CG1.5 = special case of the compound-gamma where shape parameter = 1.5;
CG2 = special case of the compound-gamma where shape parameter = 2;
CG2.5 = special case of the compound-gamma where shape parameter = 2.5;

[b] MM = method of momemnts

[c] Best Fit = theoretical distribution with lowest GOF probability.

[d] Sum of Number of Best Fits > 18 due to ties

Table 2-3 shows the average GOF probability for each distribution. Averaging the GOF probability compares how well a theoretical distribution fit all observed distributions as a group. A better fit would entail a lower average GOF probability. The CG2.5MM produced the lowest average GOF probability followed by CG2MM, CG1.5MM and CG1MM respectively.

Table 2-3: Average of all 18 power divergence test results for each theoretical distribution.

Distribution[a]	Parameter Estimation Method[b]	Average GOF Tests[c]
CG2.5	MM	0.51
CG2	MM	0.52
CG1.5	MM	0.57
CG1	MM	0.63

[a] CG1 = special case of the compound-gamma where shape parameter = 1;
CG1.5 = special case of the compound-gamma where shape parameter = 1.5;
CG2 = special case of the compound-gamma where shape parameter = 2;
CG2.5 = special case of the compound-gamma where shape parameter = 2.5;

[b] MM = method of momemnts

[c] Average of GOF test = average of the GOF probabilities for all 18 distributional fits.

Table 2-4 shows the maximum GOF probability that each of the four theoretical distributions produced. The lower the maximum GOF probability value, the more robust

a theoretical distribution is. The CG2.5MM and CG2MM tied with the lowest maximum GOF probability value of 0.87.

Table 2-4: Maximum value of the power divergence test for each theoretical distribution.

Distribution[a]	Parameter Estimation Method[b]	Maximum of GOF Tests[c]
CG2.5	MM	0.87
CG2	MM	0.87
CG1	MM	0.89
CG1.5	MM	0.94

[a] CG = compound-gamma;
CG0.5 = special case of the compound-gamma where shape parameter = 0.5;
CG1 = special case of the compound-gamma where shape parameter = 1;
CG1.5 = special case of the compound-gamma where shape parameter = 1.5;
CG2 = special case of the compound-gamma where shape parameter = 2;
CG2.5 = special case of the compound-gamma where shape parameter = 2.5;

[b] MM = method of momemnts

[c] Maximum of GOF Tests = maximum GOF probability that a theoretical distribution produced.

Table 2-5 summarizes results of Tables 2-1 through 2-4 to help determine which of the four theoretical distributions would best model sample test results of shelled corn for aflatoxin. Each theoretical distribution acceptably fit all 18 observed distributions. The percentage of best fits is tied between CG2.5MM and CG2MM with 56% each. The CG2.5MM distribution has the lowest average GOF probability (0.51) while CG2.5MM and CG2MM are tied with the lowest maximum GOF probability (.87).

Table 2-5: Summary comparison for each of the theoretical distributions. Percentage of acceptable fits, percentage of best fits, average of GOF tests, and maximum values of GOF tests are included.

	Distribution			
	CG2.5MM[a]	**CG2MM[b]**	**CG1.5MM[c]**	**CG1MM[d]**
Acceptable Fits[e]	100%	100%	100%	100%
Best Fits[f]	56%	56%	44%	39%
Average of GOF tests[g]	0.51	0.52	0.57	0.63
Maximum of GOF test[h]	0.87	0.87	0.89	0.94

[a] CG2.5MM = special case of the compound-gamma where shape parameter = 2.5 and method of moments is the parameter estimation technique.

[b] CG2MM = special case of the compound-gamma where shape parameter = 2 and method of moments is the parameter estimation technique.

[c] CG1.5MM = special case of the compound-gamma where shape parameter = 1.5 and method of moments is the parameter estimation technique.

[d] CG1MM = special case of the compound-gamma where shape parameter = 1 and method of moments is the parameter estimation technique.

[e] Acceptable Fit = null hypothesis not rejected at the 5% confidence level.

[f] Best Fit = theoretical distribution with GOF acceptance probability.

[g] Average of GOF tests = sum of the GOF probabilities for all 18 distributional fits.

[h] Maximum of GOF Tests = maximum GOF probability that a theoretical distribution produced.

The CG2.5MM was chosen to model the observed distribution of aflatoxin sample test results for shelled corn. Table 2-6 shows for each lot, the three compound gamma distribution parameters determined by the method of moments and GOF probability results as determined by the power-divergence test. Figure 2-2 shows an example of the CG2.5MM theoretical distribution compared to the observed distribution for lot 10. Very little difference between the two distributions can be noted (See Table 2-6, lot 10, PD test = .07).

Table 2-6: Lot concentration, compound gamma distribution parameters, and results of the power-divergence test for each lot.

Lot	Aflatoxin Concentration	Compound Gamma Parameters[a]			Goodness of Fit Probability[b]
		α	β	λ	
1	4.8	2.5	4.13	0.46	0.20
2	6.8	2.5	4.54	0.60	0.68
3	7.6	2.5	6.10	0.50	0.37
4	9.6	2.5	4.47	0.86	0.69
5	11.8	2.5	4.95	0.95	0.70
6	15.2	2.5	2.15	2.83	0.75
7	18.2	2.5	6.49	1.12	0.07
8	24.7	2.5	4.31	2.29	0.10
9	26.1	2.5	6.44	1.62	0.63
10	35.1	2.5	4.44	3.16	0.87
11	58.8	2.5	1.88	12.51	0.35
12	60.9	2.5	4.24	5.75	0.76
13	95.5	2.5	3.12	12.27	0.78
14	96.4	2.5	1.47	26.15	0.43
15	114.8	2.5	3.67	12.52	0.66
16	276.8	2.5	5.06	21.86	0.07
17	293.1	2.5	4.84	24.20	0.37
18	677.4	2.5	10.99	24.67	0.73

[a] α = shape parameter
β = scale parameter
λ = expected number of contaminated kernels

[b] Goodness of Fit Probability is based on the power-divergence test

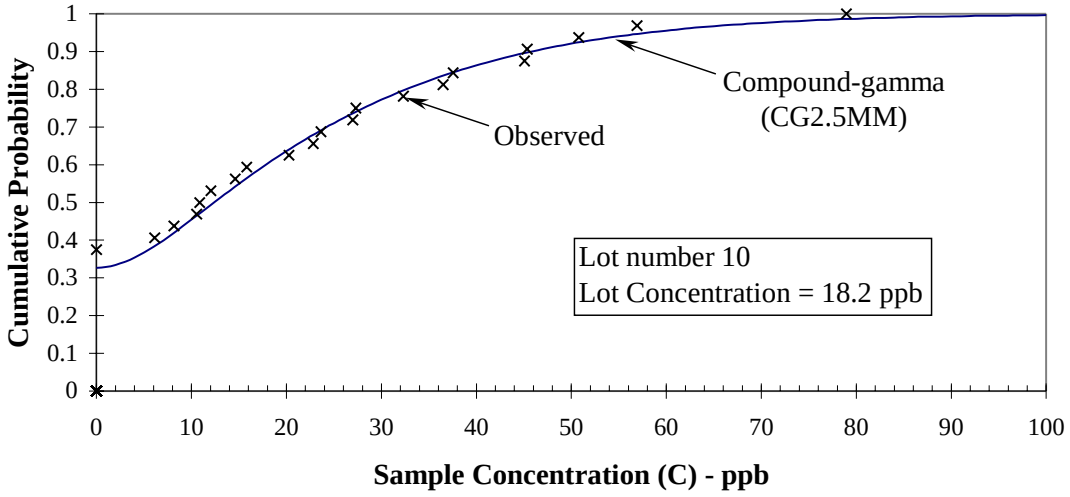


Figure 2-2: Comparison of compound gamma (CG2.5MM) theoretical distribution to the observed distribution of 32 aflatoxin test results.

Giesbrecht and Whitaker (23) suggest that a shape parameter $\alpha \leq 2$ seems most appropriate when using the compound gamma distribution to model the observed distribution of shelled peanuts for aflatoxin. Waibel (25) suggests a compound gamma distribution with a shape parameter $\alpha = 2$ to describe aflatoxin test results of shelled peanuts. The empirical data in this study suggests using a compound gamma distribution with a shape parameter $\alpha = 2.5$ to describe aflatoxin test results of shelled corn. An $\alpha > 2$ implies that the distribution of aflatoxin content on individual kernels in shelled corn behaves differently than the distribution of aflatoxin content on peanut kernels.

Parameter Estimation

It is important to be able to calculate the parameters of the compound gamma distribution for any lot aflatoxin concentration. Only the expected number of contaminated kernels λ and the scale parameter β need to be calculated because the shape

parameter α is set equal to 2.5. As shown in the following equation, λ is a function of lot aflatoxin concentration *C* and variance s_C^2 of the aflatoxin testing procedure.

Equation 2-1

$$\lambda = \left[\frac{\alpha + 1}{\alpha}\right]\frac{C^2}{s_C^2}$$

Johansson et al. (7) estimated s_C^2 as

Equation 2-2

$$s_C^2 = \left[\left(\frac{12.95}{ns}\right)C^{0.98}\right] + \left[\left(\frac{62.70}{nss}\right)C^{1.27}\right] + \left[\left(\frac{0.143}{na}\right)C^{1.16}\right]$$

where *ns* is any sample size in kg, *nss* is any subsample size in g, and *na* is any number of aliquots. Figure 2-3 shows a plot of the observed parameter λ and the predicted λ (Equations 2-1 and 2-2) versus aflatoxin concentration *C*. As the aflatoxin concentration *C* increases, both the observed and predicted λ's increase. Figure 2-3 also shows that the predicted λ is a good estimate of the observed λ.

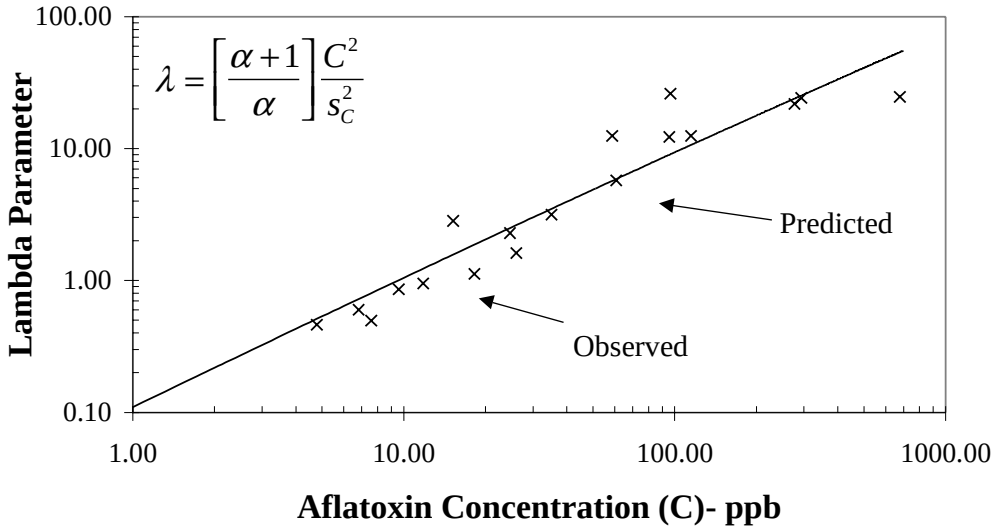


Figure 2-3: Lambda parameter (compound gamma distribution using method of moments and alpha = 2.5) versus aflatoxin concentration for 2.5 lb. samples.

Once the λ parameter has been determined, the scale parameter β can be calculated using the following equation.

Equation 2-3

$$\beta = \frac{C}{\alpha\lambda}$$

An interesting fact about the λ parameter is that it can be used to predict the number of aflatoxin-contaminated kernels found in the sample.

Equation 2-4

$$p = \frac{\lambda}{ns}$$

where *p* is the fraction of aflatoxin-contaminated kernels and *ns* in this study equals 3,456 kernels (1.13 kg sample). Equation 2-4 can be used to estimate *p* for any lot concentration *C* by substituting Equation 2-4 into Equation 2-1 and solving for *p*.

Figure 2-4 shows that as aflatoxin concentration *C* increases, *p* also increases. Using Equations 2-1, 2-2, and 2-4, a lot with aflatoxin concentration equal to 20 ppb is estimated to have approximately six contaminated kernels per 10,000 kernels. The fraction of aflatoxin-contaminated kernels for shelled corn in this study was lower than that in shelled peanuts discussed by Giesbrecht and Whitaker (23). They indicate that one contaminated kernel in 1000 is a reasonable estimate of *p*. Waibel (25) estimates one contaminated peanut kernel in 10,000. For a test procedure modeled by the compound gamma distribution at a given lot aflatoxin concentration, the fraction of contaminated particles *p* depends on the value chosen for the shape parameter α if all other variables remain constant.

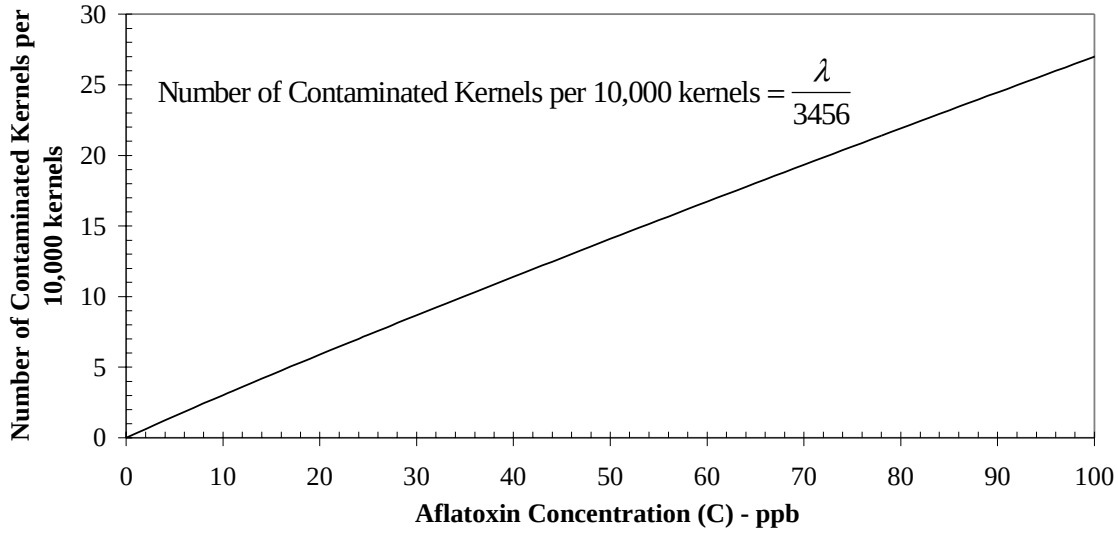


Figure 2-4: Number of contaminated kernels per 10,000 kernels versus aflatoxin concentration for 2.5 lb. samples.

Operating Characteristic Curve

Figure 2-5 shows a predicted OC curve for a sampling plan that uses the test procedure described in this study, which consists of 1.13 kg samples comminuted in a Romer mill, aflatoxin extraction from 50 g subsamples, quantification by HPLC and a sample acceptance level of 20 ppb (FDA guideline). The predicted OC curve was calculated using the compound gamma distribution. Parameters λ and β were calculated from Equations 2-1, 2-2, and 2-3 for various lot concentrations *C* and an $\alpha = 2.5$. Figure 2-5 also shows the observed acceptance probabilities *P{C}* for the sampling plan described above.

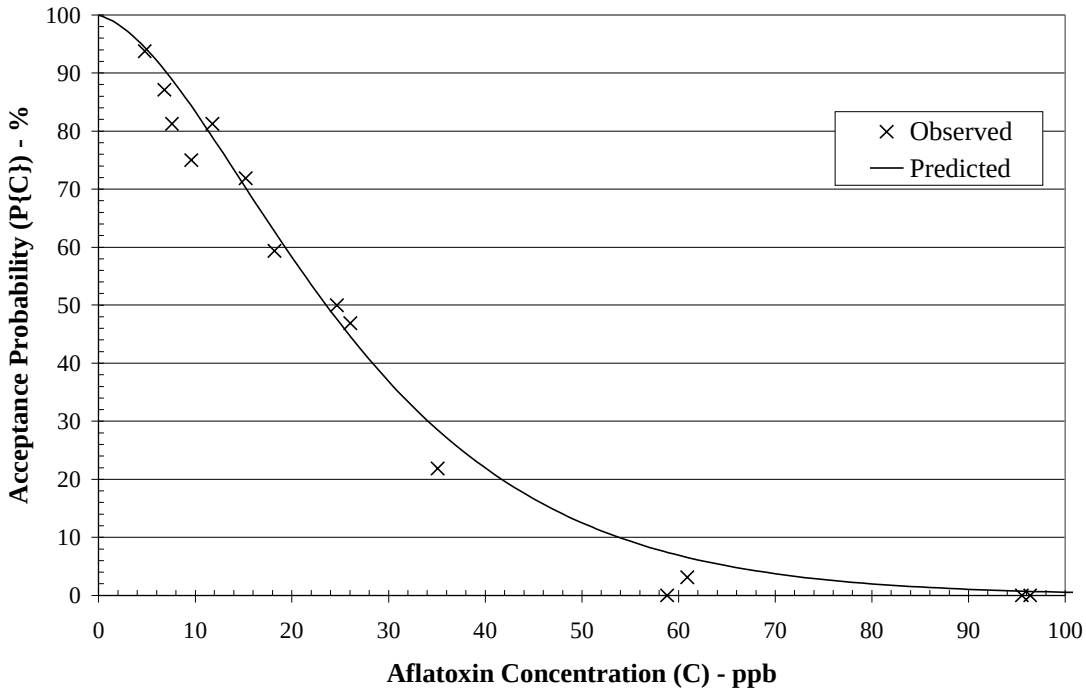


Figure 2-5: Observed and predicted acceptance probabilities for a 2.5 lb. sample and a 20 ppb acceptance level.

With eighteen observed distributions of sample aflatoxin test results, 18 point estimates of acceptance probability *P{C}* can be determined and compared to the predicted OC curve in Figure 2-5. The observed acceptance probability for a given lot is defined as a ratio of the number of sample aflatoxin test results that are 20 ppb or less divided by 32 or the total number of sample aflatoxin test results. For $C > 100$ ppb, acceptance probabilities are not shown because both the predicted and observed OC curves are equal to zero. All observed acceptance probabilities lie within the 95% confidence band computed from Equation 2-5.

Equation 2-5

$$95\%CI = P\{C\} \pm t_{.05,18df}\sqrt{\frac{P\{C\}\left[1-P\{C\}\right]}{18}}$$

where $t_{.05,\ 18df} = 2.101$ (33).

SUMMARY

To develop an aflatoxin-sampling plan, a theoretical model must be used to predict the distribution of sample test results associated with a specific aflatoxin test procedure. In this study, fifteen distributional cases based upon four theoretical distributions were each compared to 18 observed distributions. Parameters were estimated by ML for all distributions. For comparison, MM was also used to estimate parameters for the compound gamma and negative binomial distributions. The power-divergence goodness of fit test statistic was used to evaluate the suitability of each theoretical distribution. Six special cases of the compound gamma distribution and the log normal distribution had acceptable fits for all 18 observed distributions. The compound gamma distribution was selected to model the sample aflatoxin test results for shelled corn because the log normal distribution had no means of extrapolating between sample sizes. The MM technique was chosen to estimate the parameters of the compound gamma distribution because it estimated similar results to the ML technique and has simpler computational formulas. The compound gamma distribution, CG2.5MM (shape parameter $\alpha = 2.5$, and method of moments parameter estimation method) was chosen to model the observed distribution of aflatoxin test results for shelled corn. Mathematical expressions were developed to calculate the parameters of the compound gamma distribution for any lot aflatoxin concentration and aflatoxin test procedure. The compound gamma distribution was

utilized to create operating characteristic (OC) curves, which were compared to observed acceptance probabilities *P{C}*. All 18 observed acceptance probabilities *P{C}* were found to lie within a 95% confidence band of the predicted OC curve.

Using the parameters from the compound gamma distribution, the number of aflatoxin-contaminated kernels was estimated and found to increase as aflatoxin concentration increased. At 20 ppb, results show about six aflatoxin-contaminated kernels in 10,000.

REFERENCES

16. Food and Agriculture Organization. 1997. Worldwide regulations for mycotoxins 1995. *Food and Nutrition Paper 64*, 4.

17. Whitaker, T. B., J. W. Dickens, and R. J. Monroe. 1974. Variability of Aflatoxin Test Results. *Journal of the American Oil Chemists' Society*. 51:214-218.

18. Whitaker, T. B., J. W. Dickens, and R. J. Monroe. 1979. Variability Associated with Testing Corn for Aflatoxin. *Journal of the American Oil Chemists' Society*. 56:789-794.

19. Whitaker, T. B., M. E. Whiten, and R. J. Monroe. 1976. Variability Associated with Testing Cottonseed for Aflatoxin. *Journal of the American Oil Chemists' Society*. Vol. 53, No. 7, pp. 502-505.

20. Whitaker, T. B., F. E. Dowell, W. M. Hagler, Jr., F. G. Giesbrecht, and J. Wu. 1994. Variability Associated with Sampling, Sample Preparation, and Chemical Testing for Aflatoxin in Farmer's Stock Peanuts. *J. AOAC Int.* 77:107-116.

21. Whitaker, T. B., F. G. Giesbrecht, and J. Wu. 1996. Suitability of several statistical models to simulate observed distribution of sample test results in inpections of aflatoxin-contaminated peanuts lots. *J. AOAC Int.* 79, 981-988.

22. Johansson, A. S., T. B. Whitaker, F. G. Giesbrecht, W. M. Hagler, and J. H. Young. 1998. Estimation of variance components associated with testing shelled corn for aflatoxin. Submitted.

23. deKoe, W. J., and P. R. Defize. 1990. Reaction to a paper by Whitaker and Dickens on aflatoxin testing plans for shelled peanuts in the U. S. and the export market. *J. AOAC Int.* 73, 809-811.

24. Whitaker, T. B. 1990. Reply to "Reaction to a paper by Whitaker and dickens on aflatoxin testing plans for shelled peanuts in the U. S. and the export market." *J. AOAC Int.* 73, 812-813.

25. Cambell, A. D., T. B. Whitaker, A. E. Pohland, J. W. Dickens, and D. L. Park. 1986. Sampling, sample preparation, and sampling plans for foodstuffs for mycotoxin analysis. *Pure and Appl. Chem.* Vol. 58, pp. 305-314.

26. Cucullu, A. F., L. S. Lee, R. Y. Mayne, and L. A. Goldblatt. 1966. Determination of aflatoxins in individual peanuts and peanut sections. *Journal of American Oil Chemists' Society* 43(2):89-92.

27. Whitaker, T. B., and D. L. Park. 1994. The Toxicology of Aflatoxins, D. L. Eaton and J. D. Groopman (Eds.), Academic Press, San Diego, CA, pp. 433-450.

28. Dorner, J. W., and R. J. Cole. 1993 *J. AOAC Int.* 76, 983-987.

29. Horwitz, W., R. Albert, and S. Nesheim. 1993. Reliability of mycotoxin assays-An update. *J. AOAC Int.* 76, 461-489.

30. Whitaker, T. B., and E. H. Wiser. 1969. Theoretical investigation into the accuracy of sampling shelled peanuts for aflatoxin. *J. Assoc. Off. Anal. Chem.* 46, 377-379.

31. Whitaker, T. B., J. W. Dickens, R. J. Monroe, and E. H. Wiser. 1972. Comparison of the observed distribution of the aflatoxin in shelled peanuts to the negative binomial distribution. *J. Assoc. Off. Anal. Chem.* 49, 590-593.

32. Wilks, S. S. 1962. Mathematical Statistics, John Wiley and Sons, New York, NY, pp. 121-122.

33. Whitaker, T. B., F. G. Giesbrecht, J. Wu, W. M. Hagler, and F. E. Dowell. 1994. Predicting the distribution of aflatoxin rest results from farmer's stock peanuts. *J. AOAC Int.* 77, 659-666.

34. Whitaker, T. B., and M. E. Whitten. 1977. Evaluation of cottonseed aflatoxin testing programs. *J. Am. Oil Chem. Soc.* 54, 436-441.

35. Whitaker, T. B., J. W. Dorner, F. E. Dowell, and F. G. Giesbrecht. 1992. Variability Associated with Chemical Testing Screened Farmer's Stock Peanuts for Aflatoxin. *Peanut Science.* 19:88-91.

36. Knutti, R., and C. Schlatter. 1978. Problems of assessing aflatoxin in peanuts-Proposal for a sampling and analysis plan for the control of imports. *Mitt. Geb. Lebensmittelunters. Hyg.* 69, 264-274.

37. Knutti, R., and C. Schlatter. 1982. Distribution of aflatoxin in whole peanut kernels, sampling plans for small samples. *Z. Lebensm. Unters. Forsch.* 174, 122-128.

38. Giesbrecht, F. G., and T. B. Whitaker. 1998. Investigations of the problems of assessing aflatoxin levels in peanuts. *Biometrics.* 54, 739-753.

39. Brown, G. H. 1984. The distribution of total aflatoxin levels in composited samples of peanuts. *Food Technol. Aust.* 36, 128-130.

40. Waibel, Von J. 1977. Sample size for the determination of aflatoxin in peanuts. *Deutsche Lebensmitten-Rundschaw.* 73, No. 11: 353-357.

41. Coker, R. D. 1989. Aflatoxin Contamination of Groundnut, Proceedings of the International Workshop, October 6-9, 1987, ICRISAT Center, India, pp. 123-132.

42. Jewers, K., R. D. Coker, G. Blunden, M. J. Jazwinski, and A. J. Sharkey. 1986. in Spoilage and Mycotoxins of Cereals and Other Stored Products, B. Flannigan (Ed.) C. A. B. International, Farnham Royal, Slough, UK, pp. 83-88.

43. Sharkey, A. J., O. G. Roch, and R. D. Coker. 1994. A case-study on the development of a sampling and testing protocol for aflatoxin levels in edible nuts and oil-seeds. *The Statistician*. 43, 267-275.

44. Schatzki, T. F. 1995. Distribution of aflatoxin in pistachios. *J. Agric. Food Chem.* 1566-1569.

45. Mendenhall, W., D. D. Wackerly, R. L. Scheaffer. 1990. Mathematical Statistics with Applications, Duxbury Press, Belmont CA, pp. 413-425.

46. Statistical Analysis System Institute, Inc. 1996. SAS Program 6.12, Cary, NC 27513.

47. Read, T. R. C., and N. A. C. Cressie. 1988. Goodness-of-Fit Statistics for Discrete Multivariate Data, Springer-Verlag, New York, NY, pp. 1-18.

48. Steel, R. G. D., J. H. Torrie, and D. A. Dickey. 1997. Principles and Procedures of Statistics, A Biometrical Approach. McGraw Hill, New York, NY, pp. 553-556.

49. Steel, R. G. D., J. H. Torrie, and D. A. Dickey. 1997. Principles and Procedures of Statistics, A Biometrical Approach. McGraw Hill, New York, NY, pp. 553-556.

CHAPTER 3. EVALUATION OF AFLATOXIN SAMPLING PLANS FOR SHELLED CORN

ABSTRACT

Evaluating the effects of changes in sample size and/or acceptance level on the performance of aflatoxin sampling plans for shelled corn are discussed in this study. Sixteen sampling plans were evaluated. For a given sample size, decreasing the sample acceptance level, where the sample acceptance level equals the regulatory guideline: (a) decreases the percentage of lots accepted while increasing the percentage of lots rejected at all aflatoxin concentrations; (b) increases misclassification of lots (both false positives and false negatives) while decreasing the percentage of correct decisions; and (c) decreases the average aflatoxin concentration in the lots accepted and lots rejected. For a given sample size where the sample acceptance level is less than the regulatory guideline, the number of false positives increases and the number of false negatives decreases when compared to the situation where the sample acceptance level equals the regulatory guideline. For a given sample size, where the sample acceptance level is greater than the regulatory guideline, the number of false positives decreases and the number of false negatives increases when compared to the situation where the sample acceptance level equals the regulatory guideline. Increasing the sample size for a given sample acceptance level, where the sample acceptance level equals the regulatory guideline: (a) increases the percentage of lots accepted at lower concentrations while increasing the percentage of lots rejected at higher concentrations; (b) decreases misclassification of lots (both false positives and false negatives) while increasing the percentage of correct decisions; and

(c) decreases the average aflatoxin concentration in the lots accepted while increasing the average aflatoxin concentration in the rejected lots.

INTRODUCTION

Over 90 countries across the world have regulations that define the maximum concentration of aflatoxin allowed in food and feeds (1). In the United States, the Food and Drug Administration (FDA) has set a regulatory guideline at 20 parts per billion (ppb) for all foods destined for human consumption (2). Nationwide aflatoxin testing services are available upon request but are not mandatory on domestic lots of shelled corn. Currently the Grain Inspection, Packers and Stockyards Administration's (GIPSA) Federal Grain Inspection Service (FGIS) administers an aflatoxin sampling plan for all lots of exported corn (3). Commodity industries also use aflatoxin sampling plans to classify bulk lots into aflatoxin categories to decrease the amount of aflatoxin contaminated corn entering the food chain. The ability to evaluate any aflatoxin sampling plan will provide a means to design the most efficient sampling plan for the resources available.

An effective sampling plan minimizes the misclassification of lots at a reasonable cost. Aflatoxin sampling plans are uniquely defined by a test procedure and sample acceptance level. The test procedure consists of specifying sample size, sample preparation (including mill type and subsample size) and analytical method. Because of the variability associated with each step of the testing procedure, some lots will be misclassified by a given sampling plan. The fraction of lots accepted and rejected, either correctly or incorrectly, by a sampling plan can be determined from an operating characteristic (OC) curve. The shape of an OC curve is unique for an aflatoxin testing

procedure and sample acceptance level. The sample acceptance level is a threshold value that determines how a lot of shelled corn is classified and may or may not be equal to the regulatory guideline.

Johansson et al. (4) measured the variability associated with sampling, sample preparation, and analytical steps of an aflatoxin test procedure and found that the compound gamma distribution (5) adequately described the distribution of sample test results of shelled corn. Additionally, Johansson et al. (5) developed a method to compute OC curves that evaluate aflatoxin sampling plans for shelled corn using both the variability and the compound gamma distribution. The objective of this study was to evaluate the effects of various sample sizes and various sample acceptance levels on the performance of aflatoxin sampling plans for shelled corn.

METHODS

Risks

Ideally, all lots with aflatoxin concentration, C, greater than a defined regulatory guideline or guideline, C_g, should be rejected from entering the food or feed chain, and all lots with C less than or equal to C_g should be accepted. In reality, the lot aflatoxin concentration C is estimated by quantifying the aflatoxin in a sample taken from the lot. We denote the estimate of the lot by $\hat{C}$. The values of $\hat{C}$ for a specified lot may differ from C because of random variation associated with sampling, sample preparation and analytical steps of the aflatoxin test procedure (4, 6-10). Discrepancies between C and $\hat{C}$ lead to some lots being misclassified. Clearly, the objective is to minimize such

misclassifications. Aflatoxin concentrations are measured in *ng* of total aflatoxin per *g* of product or parts per billion (ppb).

Normally lots of shelled corn are inspected with an aflatoxin-sampling plan and are classified either as unacceptable when the sample test result, $\hat{C}$ is greater than the sample acceptance limit C_a, or acceptable when $\hat{C}$ is less than or equal to the sample acceptance limit. Note that in practice, C_a and C_g do not have to be equal. Nations that have no formal regulatory guideline C_g may be interested in investigating various C_a values in order to help establish a regulatory guideline. In the case where C_g is established, such as with grain processors and manufacturers, investigating various C_a values relative to C_g to reduce misclassifications may be of importance.

In a sampling plan, two types of misclassification can occur. First, a false positive occurs when an acceptable lot $C \leq C_g$ is rejected by the sampling plan because $\hat{C} > C_a$. This is also called the seller's risk because a good lot would be diverted from the food or feed chain at an expense to the seller. Second, a false negative occurs when an unacceptable lot $C > C_g$ is accepted by the sampling plan because $\hat{C} \leq C_a$. This misclassification is known as a buyer's risk. If the unacceptable lot goes undetected, the lot enters the food chain and becomes a potential health risk to the consumer. If the lot is detected at a later stage, additional cleaning and processing are required to lower the aflatoxin level. The frequency of these two misclassifications occurring depends upon C_g, C_a, the design of the aflatoxin test procedure, and the true distribution of lot concentrations. The frequency to which these two misclassifications occur at various lot concentrations can be evaluated from an OC curve.

Operating Characteristic Curve

Let *P{C}* be the probability that a lot of shelled corn with a specific aflatoxin concentration *C* is accepted by a specified sampling plan. Ideally, a sampling plan will give a large probability of accepting a lot with a small *C* and a small (zero) probability of accepting a lot with a large *C*. An OC curve is a plot of *P{C}* versus *C*. Figure 3-1 shows a typical OC curve. An OC curve is uniquely defined by a specific testing procedure, sample acceptance limit and the distribution of $\hat{C}$ values for a given *C*. In this study, *P{C}* was computed using the compound gamma distribution to model the distribution of $\hat{C}$ values (5). The distributional parameters of the compound gamma distribution are estimated from the variability observed among $\hat{C}$ values determined for repeated samples taken from lots of shelled corn (4).

Figure 3-1 predicts the percentage of lots with a given lot concentration *C* that will be accepted and rejected by the specified sampling plan and gives an indication of the misclassification errors associated with the sampling plan. An ideal OC curve would be equal to one when $C \leq C_g$ and equal to zero when $C > C_g$, implying no false positives and no false negatives. Reducing the variability associated with the test procedure helps to minimize discrepancies between $\hat{C}$ and *C*, which will reduce the misclassification of lots.

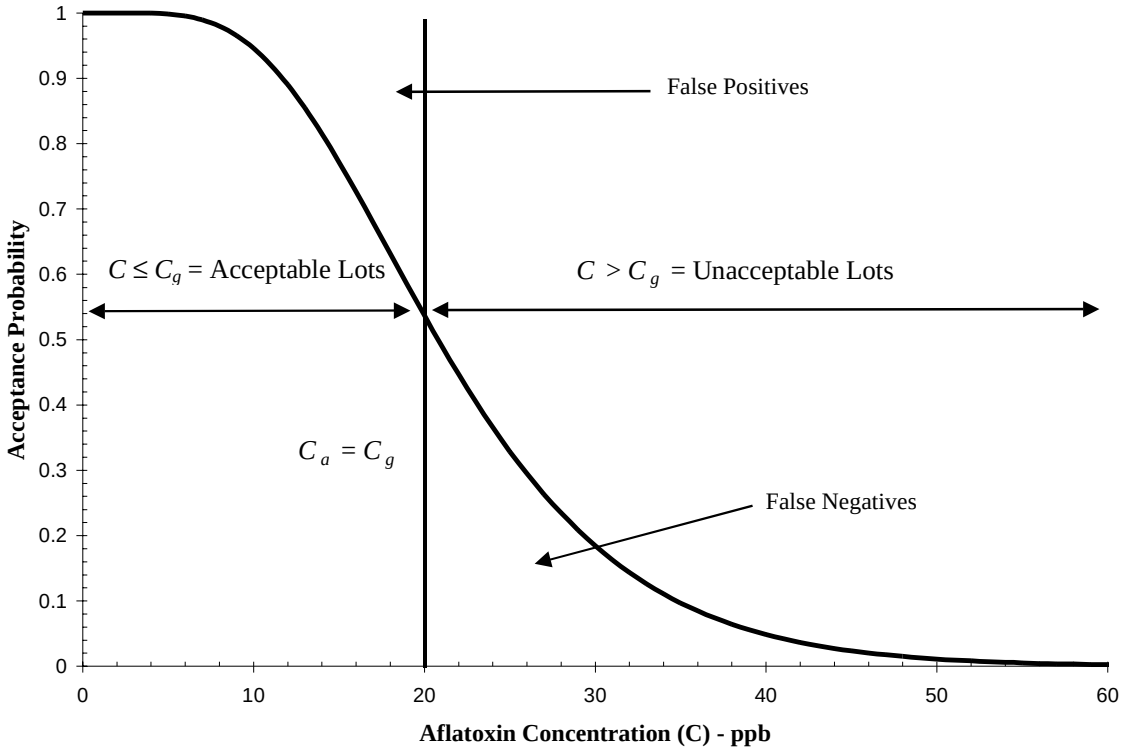


Figure 3-1: Typical operating characteristic curve showing the performance of an aflatoxin sampling plan when testing lots with aflatoxin concentration C.

Crop Distribution

The *P{C}* curve illustrated in Figure 3-1 only shows the percentage of lots accepted (or rejected) for a lot with a given aflatoxin concentration *C*. To convert the probability to number of lots accepted, number of lots rejected, false positives, or false negatives in a given crop year, the distribution of lots for a specific crop year according to aflatoxin concentrations *C* must be known. Crop distributions may vary from one extreme to another. If no lots contain aflatoxin, i.e., *C = 0* for all lots, then all lots will be accepted with no false positives or false negatives. Conversely, if all lots had very high *C* values, there would be no lots accepted and again no false positives or false negatives.

Problems arise when there is a mixture of lot concentrations, some lots with small amounts or no aflatoxin, some lots with extremely high amounts of aflatoxin, and some in between. In this study, a highly contaminated crop distribution i.e., a crop distribution where an appreciable number of lots contained aflatoxin was used to better illustrate the effects that various sample sizes and various sample acceptance levels have on a sampling plan. The crop distribution used in this study was obtained from the FDA (11).

The discussion below describes the process used to convert the percentage of lots accepted and rejected (OC curve) to the number of lots accepted, rejected and misclassified. The calculation for the average aflatoxin concentration in both accepted lots and rejected lots is also discussed.

Let L be the total number of lots tested in a crop year. The total number of lots having a specific aflatoxin concentration C is $L \times f(C)$ where $f(C)$ is the fraction of lots with concentration C obtained from the crop distribution. Note that $f(C)$ represents the lot distribution for the crop year. For a given sampling plan, the number of accepted lots can be determined using the following equation:

Equation 3-1

$$L_a = L\sum_{C=0}^{C_{\max}} f(C) \times P\{C\}$$

where $P\{C\}$ is obtained from the OC curve and C_{max} is the maximum lot concentration in the crop distribution. The number of rejected lots is defined as:

Equation 3-2

$$L_r = L\sum_{C=0}^{C_{\max}} f(C) \times \left[1 - P\{C\}\right]$$

or

Equation 3-3

$$L_r = L - L_a$$

The number of lots that are misclassified can also be calculated. The number of false positives L_{fp} can be determined from Equation 3-4.

Equation 3-4

$$L_{fp} = L\sum_{C=0}^{C_g} f(C)\times\left[1-P\{C\}\right]$$

The number of false negatives L_{fn} can be determined from Equation 3-5.

Equation 3-5

$$L_{fn} = L\sum_{C>C_g}^{C_{max}} f(C)\times P\{C\}$$

The percentage of correct decisions (CD) associated with a sampling plan is

Equation 3-6

$$CD = \frac{\left[L-\left(L_{fp}+L_{fn}\right)\right]}{L}\times 100$$

The average aflatoxin concentration among accepted lots (A_{acc}) for a range of C values from C_1 to C_2 where $C_1 < C_2$ can be determined with the following equation.

Equation 3-7

$$A_{acc} = \frac{\sum_{C=C_1}^{C_2} C\times f(C)\times P\{C\}}{\sum_{C=C_1}^{C_2} f(C)\times P\{C\}}$$

For example, if C_1 = 0 ppb and $C_2 = C_{max}$, then equation 3-6 estimates the average aflatoxin concentration among all lots of shelled corn that were accepted by the aflatoxin sampling plan.

The average aflatoxin concentration among rejected lots (A_{rej}) for a range of C values from C_1 to C_2 where $C_1 < C_2$ can be determined with the following equation.

Equation 3-8

$$A_{rej} = \frac{\sum_{C=C_1}^{C_2} C \times f(C) \times \left[1 - P\{C\}\right]}{\sum_{C=C_1}^{C_2} f(C) \times \left[1 - P\{C\}\right]}$$

RESULTS

Sixteen OC curves shown in Figures 3-2 through 3-9 were computed for 16 sampling plans that use four sample sizes (2.5, 5, 10, and 20 kg) and four sample acceptance levels (5, 10, 15, and 20 ppb). OC curves are based upon the compound gamma distribution used to describe the distribution of replicate sample test results from contaminated lots (5). OC curves were created using the SAS program in Appendix 7. Figures 3-2 through 3-5 show the effect of different sample sizes for a given sample acceptance level on the percentage of lots accepted and rejected at various C values. Figures 3-6 through 3-9 show the effect of sample acceptance levels for a given sample size on the percentage of lots accepted and rejected at various C values. The cumulative crop distribution is shown in Table 3-1. The lot concentration values for the crop distribution range from 0 to 1150 ppb and the average aflatoxin concentration for all lots in the crop distribution was 13.5 ppb.

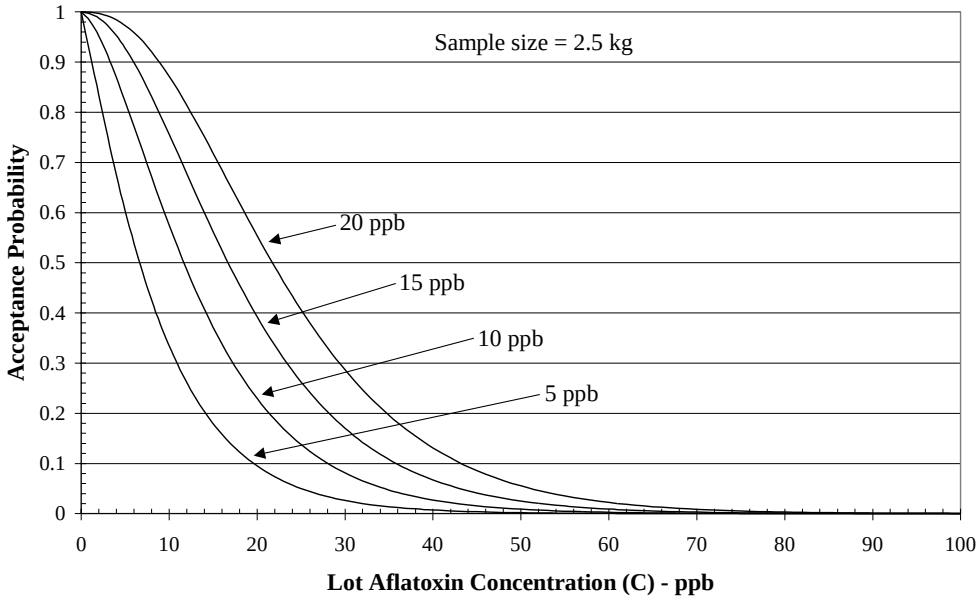


Figure 3-2: OC curves for 5, 10, 15, and 20 ppb sample acceptance level when using a 2.5 kg sample size.

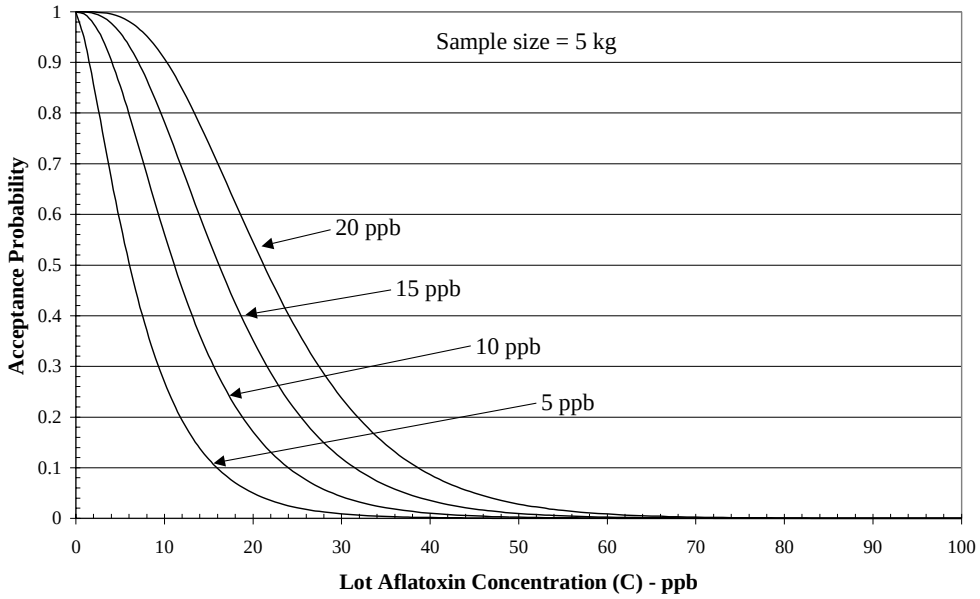


Figure 3-3: OC curves for 5, 10, 15, and 20 ppb sample acceptance level when using a 5 kg sample size.

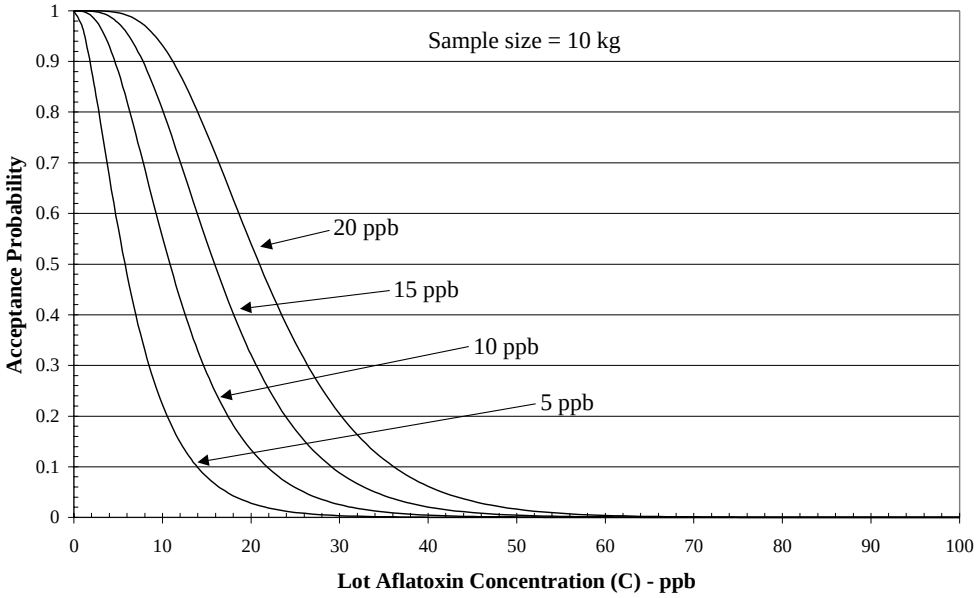


Figure 3-4: OC curves for 5, 10, 15, and 20 ppb sample acceptance level when using a 10 kg sample size.

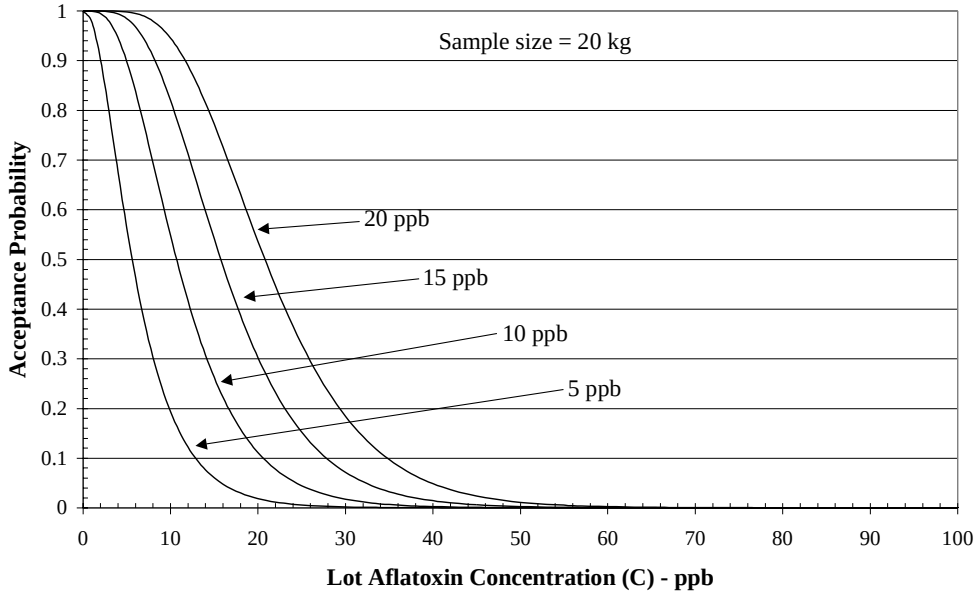


Figure 3-5: OC curves for 5, 10, 15, and 20 ppb sample acceptance level when using a 20 kg sample size.

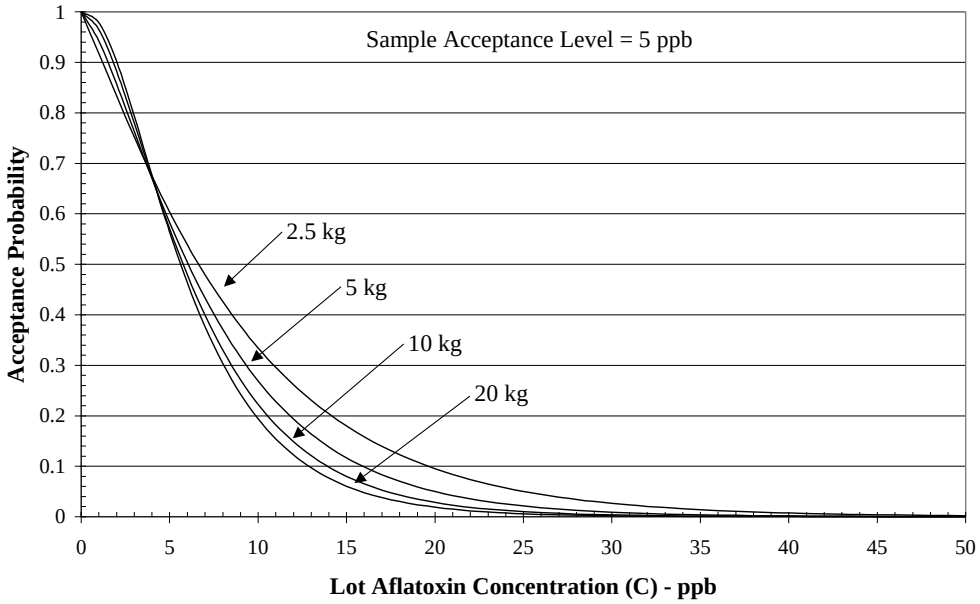


Figure 3-6: OC curves for 2.5, 5, 10, and 20 kg sample sizes when using a 5 ppb sample acceptance level.

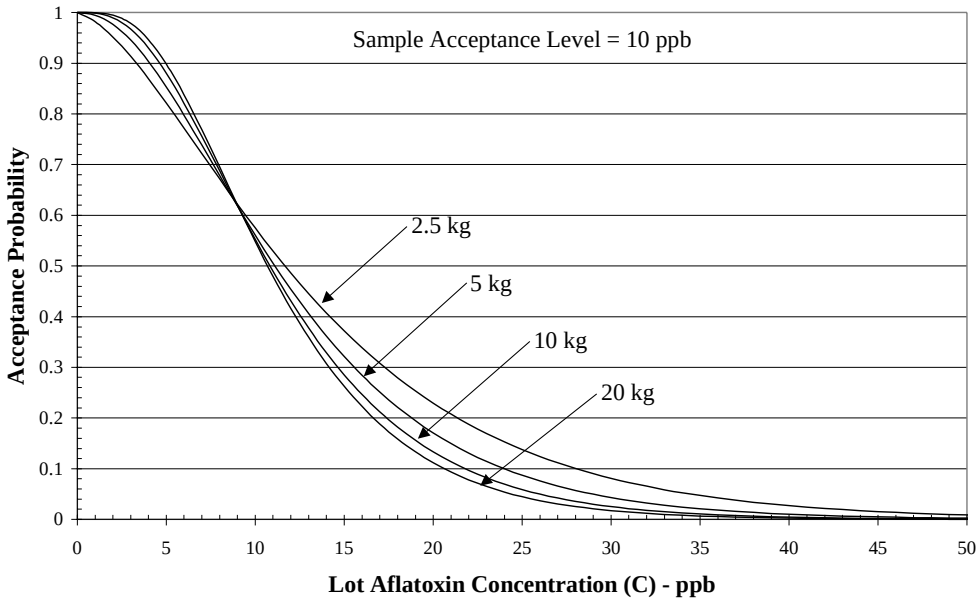


Figure 3-7: OC curves for 2.5, 5, 10, and 20 kg sample sizes when using a 10 ppb sample acceptance level.

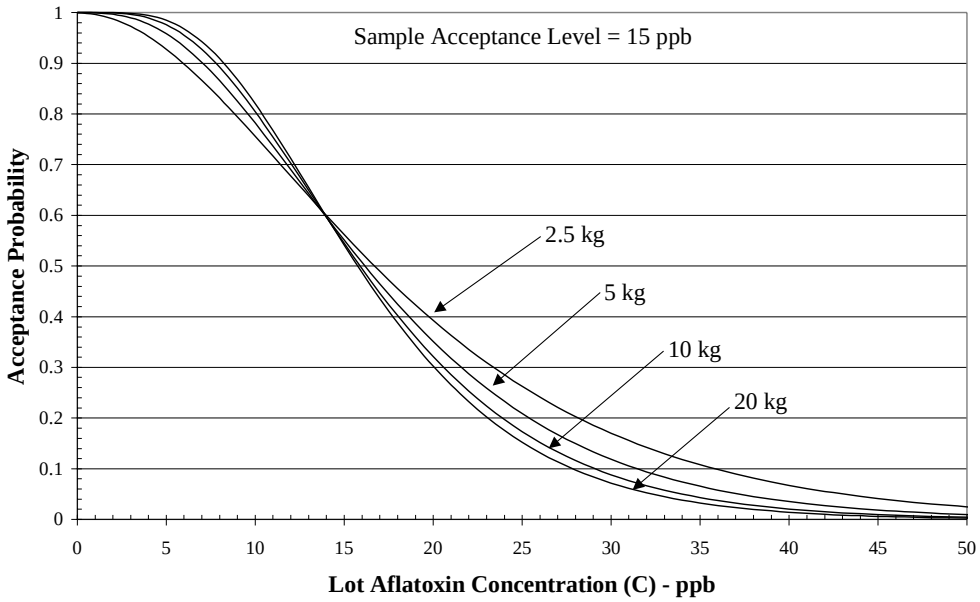


Figure 3-8: OC curves for 2.5, 5, 10, and 20 kg sample sizes when using a 15 ppb sample acceptance level.

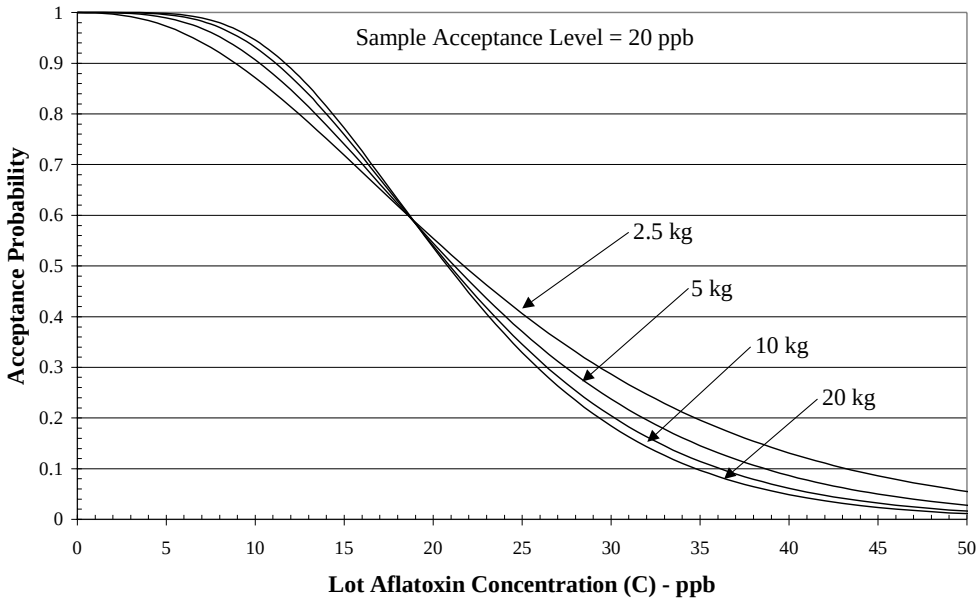


Figure 3-9: OC curves for 2.5, 5, 10, and 20 kg sample sizes when using a 20 ppb sample acceptance level.

Table 3-1: Cumulative distribution among lot aflatoxin concentrations estimated from aflatoxin sample results from FDA crop data.

Lot Concentration (ppb)	Cumulative Frequency (%)
0	40.00
5	57.48
10	85.87
20	91.00
30	93.28
50	95.48
75	96.78
100	97.50
200	98.70
300	99.14
400	99.37
500	99.50
750	99.68
1000	99.78
1150	100.00

Sample Acceptance Level

Figures 3-2 through 3-5 show the effect of sample acceptance levels 5, 10, 15, and 20 ppb on the percentage of lots accepted and rejected at various aflatoxin concentrations *C*. Each figure represents a different sample size (2.5, 5, 10, and 20 kg). For a given sample size, as the sample accept level decreases from 20 ppb to 5 ppb, the OC curves shift to the left on the aflatoxin concentration scale. Shifting the OC curve to the left indicates that the percentage of lots accepted decreases and the percentage of lots rejected increases at all lot aflatoxin concentrations. Figure 3-2 (2.5 kg sample) for example, shows as the sample acceptance level C_a decreases from 20 ppb to 5 ppb, the percentage of lots accepted at 10 ppb, decreases from 87.2% to 33.4% and the percentage of lots rejected at 10 ppb increases from 12.8% to 66.6%.

Tables 3-2 through 3-5 show the effect of changing sample acceptance level from 20 ppb to 5 ppb ($C_g = C_a$) for each sample size on the number of lots accepted and the number of lots rejected per 100 lots tested when using the crop distribution in Table 3-1. For all sample sizes, the number of lots accepted decreases and the number of lots rejected increases as sample acceptance level decreases from 20 to 5 ppb. Using Tables 3-2 through 3-5 with a sample size = 2.5 kg, decreasing the sample acceptance level from 20 ppb to 5 ppb ($C_g = C_a$) will decrease the number of lots accepted from 90.2 to 76.7 per 100 lots tested and the number of lots rejected will increase from 9.8 to 23.3 per 100 lots tested.

Table 3-2: Effect of increasing sample size using a 5 ppb sample acceptance level on the percentage of lots accepted and rejected, false positives, false negatives, correct decisions, and average ppb among lots accepted and lots rejected.

	Sample Size, kg			
Parameter	2.5	5.0	10.0	20.0
Regulatory Guideline, ppb	5	5	5	5
Sample Acceptance Level, ppb	5	5	5	5
Lots Accepted (per 100 lots tested)	76.67	76.58	76.84	77.06
Lots Rejected (per 100 lots tested)	23.33	23.42	23.16	22.94
False Positives (per 100 lots tested)	6.61	5.96	5.23	4.73
False Negatives (per 100 lots tested)	4.24	3.50	3.03	2.76
Correct Decisions %	89.15	90.54	91.74	92.51
Avg ppb Among Accepted Lots	1.34	1.23	1.17	1.14
Avg ppb Among Rejected Lots	53.64	53.80	54.59	55.22

Table 3-3: Effect of increasing sample size using a 10 ppb sample acceptance level on the percentage of lots accepted and rejected, false positives, false negatives, correct decisions, and average ppb among lots accepted and lots rejected.

	Sample Size, kg			
Parameter	2.5	5.0	10.0	20.0
Regulatory Guideline, ppb	10	10	10	10
Sample Acceptance Level, ppb	10	10	10	10
Lots Accepted (per 100 lots tested)	84.00	84.53	84.85	85.01
Lots Rejected (per 100 lots tested)	16.00	15.47	15.15	14.99
False Positives (per 100 lots tested)	4.29	3.37	2.82	2.52
False Negatives (per 100 lots tested)	2.43	2.04	1.80	1.66
Correct Decisions %	93.28	94.59	95.38	95.82
Avg ppb Among Accepted Lots	1.77	1.69	1.65	1.63
Avg ppb Among Rejected Lots	75.36	78.32	80.14	81.08

Table 3-4: Effect of increasing sample size using a 15 ppb sample acceptance level on the percentage of lots accepted and rejected, false positives, false negatives, correct decisions, and average ppb among lots accepted and lots rejected.

	Sample Size, kg			
Parameter	2.5	5.0	10.0	20.0
Regulatory Guideline, ppb	15	15	15	15
Sample Acceptance Level, ppb	15	15	15	15
Lots Accepted (per 100 lots tested)	87.89	88.28	88.47	88.56
Lots Rejected (per 100 lots tested)	12.11	11.72	11.53	11.44
False Positives (per 100 lots tested)	2.89	2.24	1.89	1.71
False Negatives (per 100 lots tested)	1.70	1.44	1.29	1.20
Correct Decisions %	95.41	96.32	96.82	97.09
Avg ppb Among Accepted Lots	2.12	2.06	2.03	2.02
Avg ppb Among Rejected Lots	96.41	100.00	101.82	102.71

Table 3-5: Effect of increasing sample size using a 20 ppb sample acceptance level on the percentage of lots accepted and rejected, false positives, false negatives, correct decisions, and average ppb among lots accepted and lots rejected.

	Sample Size, kg			
Parameter	2.5	5.0	10.0	20.0
Regulatory Guideline, ppb	20	20	20	20
Sample Acceptance Level, ppb	20	20	20	20
Lots Accepted (per 100 lots tested)	90.20	90.46	90.59	90.65
Lots Rejected (per 100 lots tested)	9.80	9.54	9.41	9.35
False Positives (per 100 lots tested)	2.11	1.66	1.42	1.29
False Negatives (per 100 lots tested)	1.31	1.12	1.00	0.94
Correct Decisions %	96.58	97.23	97.58	97.77
Avg ppb Among Accepted Lots	2.42	2.38	2.36	2.35
Avg ppb Among Rejected Lots	115.82	119.41	121.16	122.02

Tables 3-2 through 3-5 show the number of misclassifications, false positives L_{fp} and false negatives L_{fn} per 100 lots tested, and how they are affected by a change in sample acceptance level ($C_g = C_a$). For all sample sizes, both L_{fp} and L_{fn} increase as sample acceptance level decreases. For example in Tables 3-2 through 3-5, as the sample acceptance level decreases from 20 to 5 ppb ($C_g = C_a$) for a sample size of 2.5 kg, the L_{fp} increases from 2.1 to 6.6 per 100 lots tested, and L_{fn} increases from 1.3 to 4.2 per 100 lots tested.

The percentage of CD was estimated using Equation 3-6. Tables 3-2 through 3-5 show for a given sample size, as the sample acceptance level decreases from 20 to 5 ppb, CD decreases ($C_g = C_a$). For example, Tables 3-2 through 3-5 show that for a sample size = 2.5 kg, as the sample acceptance level decreases from 20 ppb to 5 ppb ($C_g = C_a$), the CD value decreases from 96.6% to 89.2%.

Decreasing the sample acceptance level (sample size remaining constant and $C_g = C_a$) decreases the average aflatoxin concentration among lots accepted and lots rejected.

Tables 3-2 through 3-5 (2.5 kg sample) show that as the sample acceptance level is decreased from 20 ppb to 5 ppb, the average aflatoxin concentration for lots accepted changes from 2.42 ppb to 1.34 ppb, and the average aflatoxin concentration for lots rejected changes from 115.82 ppb to 53.64 ppb. Reducing the sample acceptance level shifts the higher lot aflatoxin concentration values in the lots accepted to the lower aflatoxin concentrations in the lots rejected.

Table 3-6 shows the number of misclassifications, false positives (L_{fp}) and false negatives (L_{fn}) per 100 lots tested, and how they are affected by a change in sample acceptance level C_a when the regulatory guideline C_g remains constant at 20 ppb i.e., $C_g \neq C_a$. For a given sample size of 20 kg and a regulatory guideline of 20 ppb, L_{fp} increases from 0.2 to 6.1 per 100 lots tested as sample acceptance level decreases from 30 ppb to 10 ppb. In contrast, given a sample size of 20 kg and a regulatory guideline of 20 ppb, L_{fn} decreases from 2.3 to 0.1 per 100 lots tested as sample acceptance level decreases from 30 ppb to 10 ppb.

Table 3-6: Effect of decreasing sample acceptance level where the regulatory guideline equals 20 ppb and using a 20 kg sample size on the number of false positives, false negatives and correct decisions.

	Sample Acceptance Level, ppb		
Parameter	30	20	10
Regulatory Guideline, ppb	20	20	20
False Positives (per 100 lots tested)	0.21	1.29	6.12
False Negatives (per 100 lots tested)	2.27	0.94	0.13
Correct Decisions %	97.53	97.77	93.75

For a given sample size, where the sample acceptance level is less than the regulatory guideline ($C_a < C_g$), the number of false positives increases and the number of

false negatives decreases when compared to the situation where the sample acceptance level equals the regulatory guideline. For a given sample size, where the sample acceptance level is greater than the regulatory guideline ($C_a > C_g$), the number of false positives decreases and the number of false negatives increases when compared to the situation where the sample acceptance level equals the regulatory guideline.

For a given sample size of 20 kg and regulatory guideline of 20 ppb, Table 3-6 shows that a sample acceptance level above or below the regulatory guideline ($C_g \neq C_a$) decreases the percentage of CD. For example, Table 3-6 shows that for a given sample size of 20 kg and regulatory guideline of 20 ppb, a sample acceptance level of 10 ppb gives 93.7% CD, and a sample acceptance level of 30 ppb gives 97.5% CD, whereas a sample acceptance level of 20 ppb ($C_g = C_a$) gives 97.8% CD.

Sample Size

Figures 3-6 through 3-9 show for a given sample acceptance level the effects of increasing sample size from 2.5 to 20 kg on the percent of lots accepted and rejected at various lot concentrations. Each OC curve in Figures 3-6 through 3-9 is the same OC curve as in Figures 3-2 through 3-5 but are grouped according to sample acceptance level. For each sample acceptance level, the four OC curves associated with each sample size intersect at a lot concentration similar to the sample acceptance level. For a given sample acceptance level, the slope of the OC curve increases in the lot concentration region near the sample acceptance level as the sample size increases. As the slope increases, the percentage of lots accepted at a concentration below the sample acceptance level should increase and the percentage of lots rejected at a concentration above the sample acceptance level should also increase. For example, using a sample acceptance

level of 20 ppb and increasing the sample size from 2.5 to 20 kg (Figure 3-9), the percentage of lots accepted at 10 ppb increases from 87.2% to 94.6%, and the percentage of lots rejected at 30 ppb increases from 71.4% to 81.6%.

Tables 3-2 through 3-5 show the effect of changing sample size from 2.5 to 20 kg for each sample acceptance level on the number of lots accepted and lots rejected per 100 lots tested when using the crop distribution in Table 3-1 and assuming ($C_a = C_g$). For a given sample acceptance level, the number of lots accepted increases and the number of lots rejected decreases as sample size increases. Using a sample acceptance level of 20 ppb (Table 3-5), increasing sample size from 2.5 to 20 kg will increase the number of lots accepted from 90.2 to 90.7 per 100 lots tested and the number of lots rejected will decrease from 9.8 to 9.3 per 100 lots tested. Using the sample acceptance level equal to 5 ppb, the percentage of lots accepted decreases slightly and then increases as the sample size increases and vice versa with the percentage of lots rejected. The inconsistency is due to the skewed nature of the distribution of contaminated kernels in the lot and has been noted by Whitaker et al. (12) in shelled peanuts.

Ideally, increasing the sample size increases the slope of an OC curve, which increases the number of lots accepted and decreases the number of lots rejected. This implies that the number of false positives and the number of false negatives should be reduced, thus increasing the percentage of CD. Table 3-5 shows that using a sample acceptance level of 20 ppb, as the sample size is increased from 2.5 to 20 kg, L_{fp} decreases from 2.1 to 1.3 per 100 lots tested and L_{fn} decreases from 1.3 to 0.9 per 100 lots tested. Table 3-5 also shows that using a sample acceptance level of 20 ppb, the percentage of CD increases from 96.6% to 97.8% as the sample size increases.

For a given sample acceptance level, increasing sample size decreases the average aflatoxin concentration of the accepted lots and increases the average aflatoxin concentration of the rejected lots of shelled corn. For example, Table 3-2 (sample acceptance level equal to 5 ppb) shows that as sample size increases from 2.5 to 20 kg, the average aflatoxin concentration among the accepted lots decreases from 1.34 to 1.14 ppb and the average aflatoxin concentration among the rejected lots increases from 53.64 to 55.22 ppb.

SUMMARY

Evaluating the effect of sample size and sample acceptance level on aflatoxin sampling plans for shelled corn was investigated in this study. Four sample sizes (2.5, 5, 10, and 20 kg) combined with four sample acceptance levels (5, 10, 15, and 20 ppb) were used to create 16 sampling plans. Using the compound gamma distribution, operating characteristic curves were developed to evaluate the 16 sampling plans. For a given sample size, if the sample acceptance level equals the regulatory guideline and sample acceptance level decreases: (a) the percentage of lots accepted decreases and percentage of lots rejected increases; (b) the percentage of correct decisions decreases because misclassification of lots (both false positives and false negatives) increases; and (c) the average aflatoxin concentration in both lots accepted and lots rejected decreases. For a given sample size where the sample acceptance level is less than the regulatory guideline, the number of false positives increases and the number of false negatives decreases when compared to the situation where the sample acceptance level equals the regulatory guideline. For a given sample size, where the sample acceptance level is greater than the regulatory guideline, the number of false positives decreases and the number of false

negatives increases when compared to the situation where the sample acceptance level equals the regulatory guideline.

For a given sample acceptance level, if the sample acceptance level equals the regulatory guideline and sample size increases: (a) increases the percentage of lots accepted at lower concentrations while increasing the percentage of lots rejected at higher concentrations; (b) the percentage of correct decisions increases because the misclassification of lots (both false positives and false negatives) decreases; and (c) the average aflatoxin concentration for the lots accepted decreases while the average aflatoxin concentration for the lots rejected increases.

REFERENCES

1. Food and Agriculture Organization. 1997. Worldwide regulations for mycotoxins 1995. *Food and Nutrition Paper 64*, 4.

2. Food and Agriculture Organization. 1993. Sampling plans for aflatoxin analysis in peanuts and corn. *Food and Nutrition Paper 55*.

3. USDA Grain Inspection, Packers and Stockyards Administration. 1998. Aflatoxin Fact Sheet. http://www.usda.gov/gipsa/progser/inspwgh/aflafact.htm (10/05/98).

4. Johansson, A. S., T. B. Whitaker, F. G. Giesbrecht, W. M. Hagler, and J. H. Young. 1998. Estimation of variance components associated with testing shelled corn for aflatoxin. Submitted

5. Johansson, A. S., T. B. Whitaker, F. G. Giesbrecht, W. M. Hagler, and J. H. Young. 1998. Determination of a suitable statistical model to simulate observed distributions of aflatoxin test results in shelled corn. Submitted.

6. Whitaker, T. B., J. W. Dickens, and R. J. Monroe. 1974. Variability of Aflatoxin Test Results. *Journal of the American Oil Chemists' Society*. 51:214-218.

7. Whitaker, T. B., J. W. Dickens, and R. J. Monroe. 1979. Variability Associated with Testing Corn for Aflatoxin. *Journal of the American Oil Chemists' Society*. 56:789-794.

8. Whitaker, T. B., M. E. Whiten, and R. J. Monroe. 1976. Variability Associated with Testing Cottonseed for Aflatoxin. *Journal of the American Oil Chemists' Society.* Vol. 53, No. 7, pp. 502-505.

9. Whitaker, T. B., F. E. Dowell, W. M. Hagler, Jr., F. G. Giesbrecht, and J. Wu. 1994. Variability Associated with Sampling, Sample Preparation, and Chemical Testing for Aflatoxin in Farmer's Stock Peanuts. *J. AOAC Int.* 77:107-116.

10. Whitaker, T. B., F. G. Giesbrecht, and J. Wu. 1996. Suitability of several statistical models to simulate observed distribution of sample test results in inpections of aflatoxin-contaminated peanuts lots. *J. AOAC Int.* 79, 981-988.

11. Troxell, Terry. Food and Drug Administration. 1998. Personal communication.

12. Whitaker, T. B., J. Wu, F. E. Dowell, W. M. Hagler, and F. G. Giesbrecht. 1994. Effects of sample size and sample acceptance level on the number of aflatoxin-contaminated farmer's stock lots accepted and rejected at the buying point. *J. AOAC Int.* 77, 1672-1680.

CONCLUSIONS

The variability and distributional characteristics associated with testing shelled corn for aflatoxin were investigated to evaluate the effectiveness of aflatoxin sampling plans for shelled corn. First, the variability associated with testing samples of shelled corn was studied to provide a base for statistically measuring the effectiveness of aflatoxin sampling plans for shelled corn. The total variance associated with testing lots of shelled corn for aflatoxin was shown to increase as aflatoxin concentration increased. This also held true for each step of the test procedure: sampling, sample preparation, and analytical variability. Linear regressions were shown to model all three components well in the full-log scale. Testing a lot with 20 parts per billion (ppb) aflatoxin using a 2.5 lb sample, Romer mill and 50 g subsample, and HPLC analysis, the total, sampling, sample preparation, and analytical variances are 274.9 (CV=82.9%), 214.0 (CV=73.1%), 56.3 (CV=37.5%), and 4.6 (CV=10.7%), respectively. The percentage of the total variance for sampling, sample preparation, and analytical is 77.8, 20.5, and 1.7 %, respectively. As with testing of aflatoxins in other commodities, sampling variance contributes the most variability followed by sample preparation and then analytical variability.

Next, statistically modeling the distribution of sample test results associated with a specific aflatoxin test procedure was investigated. Fifteen distributional cases based upon four theoretical distributions were each compared to 18 observed distributions. Parameters were estimated by maximum likelihood method, ML, for all distributions. For comparison, the method of moments, MM was also used to estimate parameters for the compound gamma and negative binomial distributions. The power-divergence goodness of fit test statistic was used to evaluate the suitability of each theoretical

distribution. Six special cases of the compound gamma distribution and the log normal distribution had acceptable fits for all 18 observed distributions. The compound gamma distribution was selected to model the sample aflatoxin test results for shelled corn because the log normal distribution had no means of extrapolating between sample sizes. The MM technique was chosen to estimate the parameters of the compound gamma distribution because it calculated similar results to the ML technique and has simpler computational formulas. The compound gamma distribution, CG2.5MM (shape parameter $\alpha = 2.5$, and method of moments parameter estimation method) was chosen to model the observed distribution of aflatoxin test results for shelled corn. Mathematical expressions, utilizing the mean and variance relationships, were developed to calculate the parameters of the compound gamma distribution for any lot aflatoxin concentration and aflatoxin test procedure. The compound gamma distribution was utilized to create operating characteristic curves, which were compared to observed acceptance probabilities *P{C}*. All 18 observed acceptance probabilities *P{C}* were found to lie within a 95% confidence band of the predicted OC curve.

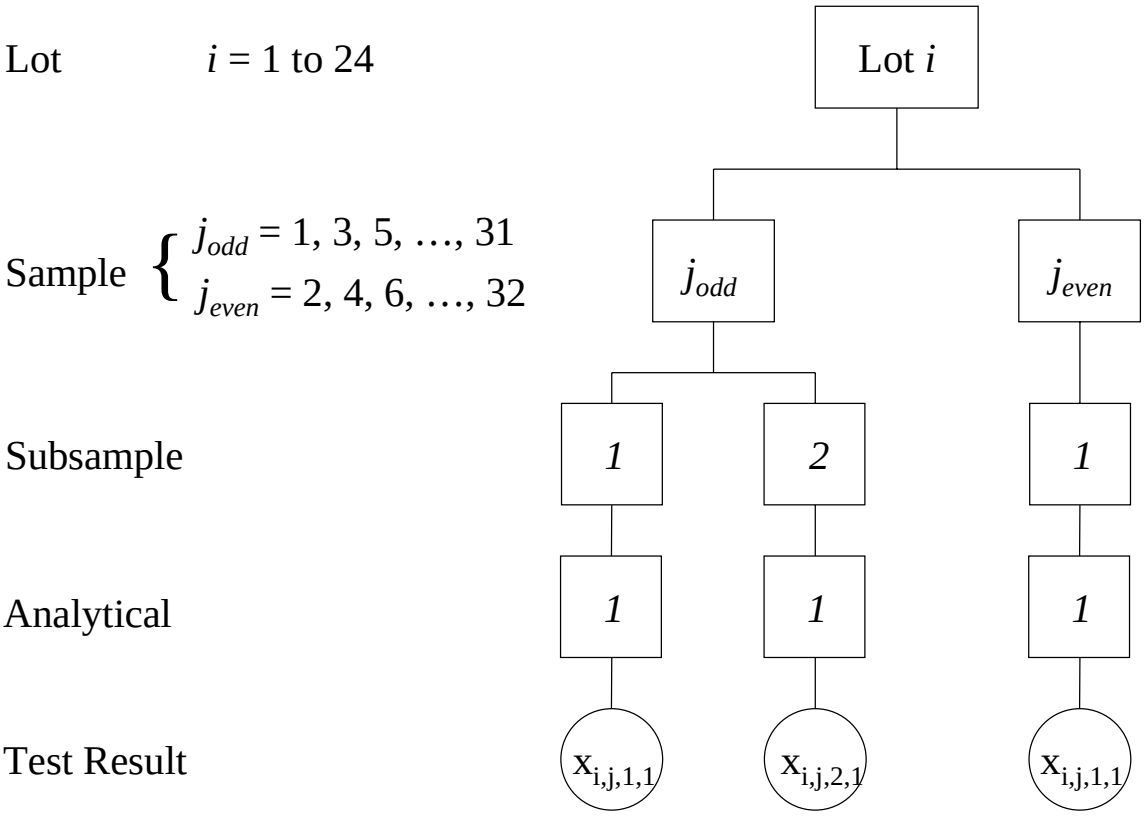
To evaluate the effect of sample size and sample acceptance level on the performance of aflatoxin sampling plans for shelled corn, four sample sizes (2.5, 5, 10, and 20 kg) combined with four sample acceptance levels (5, 10, 15, and 20 ppb) were used to create 16 sampling plans. Using the compound gamma distribution, operating characteristic curves were developed to evaluate the 16 sampling plans. For a given sample size, if the sample acceptance level equals the regulatory guideline and sample acceptance level decreases: (a) the percentage of lots accepted decreases and percentage of lots rejected increases; (b) the percentage of correct decisions decreases because

misclassification of lots (both false positives and false negatives) increases; and (c) the average aflatoxin concentration in both lots accepted and lots rejected decreases. For a given sample size where the sample acceptance level is less than the regulatory guideline, the number of false positives increases and the number of false negatives decreases when compared to the situation where the sample acceptance level equals the regulatory guideline. For a given sample size, where the sample acceptance level is greater than the regulatory guideline, the number of false positives decreases and the number of false negatives increases when compared to the situation where the sample acceptance level equals the regulatory guideline.

For a given sample acceptance level, if the sample acceptance level equals the regulatory guideline and sample size increases: (a) increases the percentage of lots accepted at lower concentrations while increasing the percentage of lots rejected at higher concentrations; (b) the percentage of correct decisions increases because the misclassification of lots (both false positives and false negatives) decreases; and (c) the average aflatoxin concentration for the lots accepted decreases while the average aflatoxin concentration for the lots rejected increases.

APPENDICES

APPENDIX 1. EXPERIMENTAL DESIGN.



Appendix 1 Figure 1: Schematic of the experimental design using 24 lots of shelled corn.

Twenty-four lots of shelled corn were riffle divided into 32 samples. An unbalanced nested design was used where 16 of the 32 samples had two subsamples taken while the remaining 16 samples only had one sample taken. This provided 1,152 test results.

APPENDIX 2. SAMPLE AFLATOXIN TEST RESULTS FOR 24 LOTS OF SHELLED CORN.

Appendix 2 Table 1: Sample aflatoxin test results for lots 1 - 6. Table includes 32 samples per lot with subsamples A and B for odd numbered lots and subsample A for even numbered lots[a].

	Lot 1		Lot 2		Lot 3		Lot 4		Lot 5		Lot 6	
Sample #	**Sub A**	**Sub B**	**Sub A**	**Sub B**	**Sub A**	**Sub B**	**Sub A**	**Sub B**	**Sub A**	**Sub B**	**Sub A**	**Sub B**
1	231.0	317.9	289.3	350.2	20.0	28.2	170.5	189.1	30.8	52.1	831.8	1111.2
2	378.9		376.6		75.0		80.6		30.6		1205.5	
3	350.5	331.4	293.9	314.4	57.9	51.6	115.3	106.5	0.0	0.0	886.4	603.7
4	165.2		300.3		77.3		138.4		23.4		694.5	
5	212.4	166.9	337.1	193.2	55.4	38.7	91.7	70.7	64.6	42.0	604.1	378.3
6	267.9		255.4		115.3		100.2		0.0		695.4	
7	314.7	339.9	479.7	502.7	58.6	51.2	87.4	86.5	22.3	27.7	516.5	645.1
8	222.9		208.8		36.3		127.7		41.0		740.8	
9	319.4	307.9	192.8	221.6	54.8	59.1	118.0	157.5	23.9	12.4	596.2	704.4
10	359.3		154.4		79.3		93.7		8.1		578.4	
11	444.4	292.5	289.1	269.5	58.9	43.6	120.1	96.3	16.5	0.0	868.5	648.7
12	257.1		255.2		47.1		62.2		46.9		563.4	
13	307.3	357.9	214.9	236.1	67.6	94.7	101.1	125.6	62.8	66.8	630.3	751.4
14	228.1		320.6		76.5		75.9		59.2		588.1	
15	236.3	389.4	271.9	527.0	33.8	37.8	69.7	105.1	23.5	31.6	595.6	1178.5
16	259.0		344.0		26.5		92.4		91.6		647.9	
17	216.5	118.6	227.0	197.6	50.8	32.3	107.0	116.2	9.0	7.3	490.6	537.9
18	295.1		380.9		60.3		134.3		45.1		566.1	
19	221.3	171.1	240.1	236.1	22.7	30.6	111.0	120.0	59.5	48.5	727.9	695.4
20	346.7		303.3		68.6		125.1		31.4		761.9	
21	157.0	152.8	370.7	377.1	44.2	51.8	83.0	78.4	65.2	55.5	579.5	534.9
22	271.7		338.7		44.2		137.9		26.4		627.9	
23	326.1	289.6	387.4	382.0	65.6	82.7	204.7	151.0	0.0	0.0	684.2	641.1
24	407.5		394.3		186.0		248.0		82.0		1150.5	
25	378.3	353.1	203.4	190.1	76.2	85.2	107.4	85.1	21.7	0.0	668.5	581.7
26	190.4		350.2		54.0		117.8		31.8		642.2	
27	256.1	283.0	338.4	458.8	92.3	0.0	109.5	117.8	29.2	29.2	584.0	670.6
28	290.9		266.6		57.2		120.9		33.2		548.2	
29	194.1	277.2	233.1	228.7	38.2	58.0	80.6	75.6	27.4	36.3	598.1	599.1
30	210.7		236.6		63.4		82.7		13.4		604.8	
31	279.6	282.1	239.2	280.2	39.2	47.7	94.1	106.8	69.3	45.9	489.6	518.6
32	261.8		286.4		45.2		164.7		32.9		709.6	

[a] ---- denotes missing data.

Appendix 2 Table 2: Sample aflatoxin test results for lots 7 - 12. Table includes 32 samples per lot with subsamples A and B for odd numbered lots and subsample A for even numbered lots[a].

	Lot 7		Lot 8		Lot 9		Lot 10		Lot 11		Lot 12	
Sample #	**Sub A**	**Sub B**	**Sub A**	**Sub B**	**Sub A**	**Sub B**	**Sub A**	**Sub B**	**Sub A**	**Sub B**	**Sub A**	**Sub B**
1	0.0	0.0	0.0	0.0	0.0	0.0	110.1	93.6	14.3	20.7	11.4	8.1
2	20.3		8.0		41.1		98.2		15.2		13.2	
3	0.0	0.0	0.0	0.0	0.0	0.0	73.2	71.3	16.5	33.3	0.0	0.0
4	27.3		0.0		0.0		132.3		0.0		0.0	
5	22.8	0.0	0.0	0.0	18.5	0.0	128.7	153.2	21.2	13.9	0.0	0.0
6	0.0		19.1		0.0		112.4		0.0		41.6	
7	0.0	0.0	0.0	0.0	0.0	0.0	42.4	57.1	18.3	48.3	26.8	28.4
8	0.0		32.5		0.0		88.6		9.7		0.0	
9	56.9	2.2	31.6	40.6	0.0	17.6	104.2	103.3	34.6	40.2	0.0	0.0
10	0.0		0.0		0.0		108.2		6.3		9.3	
11	0.0	0.0	0.0	0.0	0.0	0.0	71.7	80.8	6.6	0.0	5.4	0.0
12	0.0		0.0		16.4		80.4		17.7		6.1	
13	0.0	19.3	0.0	0.0	0.0	0.0	57.0	56.0	35.7	57.8	9.8	12.3
14	32.3		0.0		21.7		113.5		32.5		6.3	
15	0.0	0.0	8.1	0.0	0.0	0.0	95.9	74.6	27.7	42.3	9.0	13.0
16	0.0		0.0		23.1		85.4		7.5		0.0	
17	45.1	27.2	18.7	14.4	7.7	11.7	71.0	84.8	2.6	10.4	0.0	12.1
18	0.0		0.0		0.0		105.5		27.2		18.3	
19	23.6	20.4	0.0	0.0	19.2	27.5	112.6	94.2	16.8	21.7	0.0	7.0
20	15.8		27.4		0.0		66.2		6.1		8.5	
21	6.1	6.8	3.3	2.4	26.4	26.3	119.4	116.7	5.9	0.0	15.1	19.1
22	36.5		0.0		21.9		99.8		39.2		0.0	
23	45.4	0.0	0.0	0.0	20.5	0.0	62.7	65.1	0.0	22.7	0.0	0.0
24	79.0		0.0		0.0		99.8		17.0		40.4	
25	37.6	32.2	0.0	0.0	0.0	0.0	102.5	110.7	0.0	0.0	0.0	0.0
26	12.0		17.6		0.0		95.2		20.6		37.1	
27	10.9	15.5	10.3	7.4	2.6	0.0	103.7	103.3	15.2	13.0	38.5	32.3
28	10.5		0.0		39.5		84.1		18.5		6.8	
29	14.6	36.1	----	20.2	6.8	13.1	112.2	113.2	13.8	23.0	48.5	47.2
30	8.2		6.2		10.4		87.5		8.6		14.8	
31	27.0	23.1	2.6	6.3	5.2	8.0	139.6	128.6	21.6	40.1	10.1	12.0
32	50.8		26.1		26.0		120.2		10.5		0.0	

[a] ---- denotes missing data.

Appendix 2 Table 3: Sample aflatoxin test results for lots 13 - 18. Table includes 32 samples per lot with subsamples A and B for odd numbered lots and subsample A for even numbered lots[a].

	Lot 13		Lot 14		Lot 15		Lot 16		Lot 17		Lot 18	
Sample #	**Sub A**	**Sub B**	**Sub A**	**Sub B**	**Sub A**	**Sub B**	**Sub A**	**Sub B**	**Sub A**	**Sub B**	**Sub A**	**Sub B**
1	48.0	47.4	0.0	0.0	15.0	17.0	0.0	0.0	0.0	12.2	46.3	55.5
2	112.3		0.0		0.0		28.5		37.9		54.6	
3	73.5	74.7	0.0	0.0	19.9	37.4	62.2	53.3	----	----	59.4	44.6
4	116.7		0.0		0.0		20.0		13.8		33.5	
5	97.6	97.1	0.0	4.1	27.9	6.7	52.6	12.9	0.0	0.0	51.1	19.6
6	105.2		0.0		0.0		0.0		25.3		85.0	
7	170.3	191.5	0.0	0.0	0.0	0.0	50.6	48.4	49.6	58.6	48.0	37.9
8	49.0		0.0		0.0		23.9		15.7		56.9	
9	110.8	86.4	0.0	0.0	31.5	0.0	30.1	32.5	40.2	41.8	110.3	74.1
10	102.8		24.4		0.0		11.9		30.7		40.6	
11	57.4	59.7	0.0	0.0	0.0	7.2	10.6	14.2	8.3	7.0	53.2	47.1
12	54.9		34.7		0.0		0.0		0.0		50.5	
13	94.6	91.3	0.0	0.0	0.0	4.9	11.4	8.3	64.6	87.1	62.5	52.4
14	151.3		10.6		0.0		16.6		13.6		36.7	
15	106.8	101.9	0.0	0.0	0.0	0.0	17.8	19.7	16.1	13.0	49.8	50.0
16	89.0		0.0		0.0		22.7		0.0		89.5	
17	69.8	84.2	0.0	0.0	8.1	14.5	52.0	64.5	19.5	22.1	78.4	78.6
18	93.4		0.0		0.0		21.3		61.1		60.8	
19	88.9	44.4	19.5	14.5	0.0	6.3	0.0	7.6	23.5	32.3	59.0	66.8
20	73.0		4.9		0.0		13.9		5.6		74.5	
21	132.7	163.8	0.0	0.0	0.0	0.0	27.6	25.1	42.9	35.8	65.3	56.1
22	57.1		24.0		0.0		90.0		34.9		47.4	
23	130.2	145.8	0.0	12.5	0.0	0.0	0.0	0.0	0.0	12.2	60.9	54.0
24	147.2		0.0		0.0		0.0		----		69.6	
25	45.4	34.0	0.0	15.9	0.0	0.0	31.9	22.6	24.7	23.3	111.7	91.6
26	105.7		24.6		4.1		38.8		53.6		68.1	
27	75.9	90.2	5.7	7.0	10.1	9.2	54.9	72.6	42.7	31.4	45.8	53.8
28	129.7		0.0		7.8		0.0		7.7		47.9	
29	87.9	65.2	38.8	26.8	11.9	9.9	89.9	74.8	8.1	6.8	38.5	27.8
30	129.5		42.7		4.9		33.1		27.0		48.7	
31	99.4	112.5	0.0	0.0	11.8	12.1	11.0	21.5	18.5	28.6	20.4	31.3
32	50.9		12.8		0.0		10.5		54.1		56.9	

[a] ---- denotes missing data.

APPENDIX 3. SAS PROGRAM TO DETERMINE TOTAL AND SAMPLE PREPARATION COMBINED WITH ANALYTICAL VARIANCE COMPONENTS FOR THE SAMPLE AFLATOXIN TEST RESULTS.

```
data a; infile 'd:\aj\bigjohn\bjsasa.csv' dlm=',';
          input type $ lot sample subsam $ conc;
data sdaj.bjsasa; set a;
 run;

proc sort; by lot sample;
proc print;
proc means; by lot; var conc;
proc mixed /*noitprint*/; class sample;
  model conc=;
  by lot;
  make 'covparms' out=sdaj.mixed;
    random sample;
run;
```

APPENDIX 4. DISTRIBUTIONAL CHARACTERISTICS.

Compound Gamma Distribution Density Function (1)

Appendix 4 Equation 1

$$f_n(x) = \begin{cases} e^{-\lambda} & \text{if x = 0} \\ \sum_{k=1}^{n} \frac{\lambda^k e^{-\lambda}}{\Gamma(k+1)} \frac{x^{k\alpha-1} e^{-x/\beta}}{\beta^{k\alpha}\Gamma(k\alpha)} & \text{if x > 0} \end{cases}$$

α = shape parameter
β = scale parameter
λ = expected number of contaminated kernels in the lot
n = total number of kernels in the lot

The first two moments can be calculated:

Appendix 4 Equation 2

$$\hat{\mu}_1' = \frac{\left(\sum x_i\right)}{m} = \mu$$

μ = mean of the samples
m = number of samples

Appendix 4 Equation 3

$$\hat{\mu}_2' = \frac{\left(\sum x_i^2\right)}{m}$$

All three parameters of the compound gamma distribution can be calculated if α is fixed.

Appendix 4 Equation 4

$$\hat{\beta} = \frac{\hat{\mu}_2' - \left(\hat{\mu}_1'\right)^2}{(\alpha+1)\hat{\mu}_1'} = \frac{\hat{\sigma}^2}{(\alpha+1)\hat{\mu}'}$$

σ^2 = observation-to-observation variance

Appendix 4 Equation 5

$$\lambda = \frac{\hat{\mu}_1'}{\alpha\beta}$$

Negative Binomial Distribution Density Function (1)

Appendix 4 Equation 6

$$\Pr\{x;\mu,k\} = \frac{\Gamma(k+x)}{\Gamma(x+1)\Gamma(k)}\, p^k (1-p)^x \quad \text{for x = 0, 1, ...}$$

Appendix 4 Equation 7

$$p = \frac{k}{k+\mu}$$

Appendix 4 Equation 8

$$q = \frac{\mu}{\mu+k} = 1-p$$

Appendix 4 Equation 9

$$k = \frac{\mu^2}{(\sigma^2-\mu)}$$

μ = mean
σ^2 = sample-to-sample variance

Appendix 4-Equations 3-9 describe individual kernels according to their aflatoxin concentration within a lot of shelled corn, where *x* is the quantity of aflatoxin on a single kernel, *p* is the average quantity of aflatoxin in the total population or lot, *k* is the shape parameter, and Γ is the gamma function.

To create a negative binomial distribution that describes the distribution of sample aflatoxin test results, *N* independent negative binomial variables must be summed, where *N* is the number of kernels in the sample. If the random variable *X* has a negative binomial distribution with parameters μ and *K*, then the distribution of the sum of *N* independent observations has a negative binomial distribution with mean $N\mu$ and shape parameter *NK* (2).

Three-parameter Log Normal Distribution Density Function (3)

Appendix 4 Equation 10

$$f(x) = \begin{cases} \left(\frac{1}{(x+\alpha)\sigma\sqrt{2\Pi}}\right) e^{\frac{-(\ln(x+\alpha)-\mu+\alpha)^2}{2\sigma^2}}, & x > 0 \\ 0, & \text{elsewhere} \end{cases}$$

α = threshold parameter
μ = mean
σ = standard deviation

Truncated Normal Distribution Density Function (4)

Appendix 4 Equation 11

$$f(x) = \left(\frac{1}{\sigma\sqrt{2\Pi}}\right) e^{\frac{-(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < +\infty$$

μ = mean
σ = standard deviation

References

1. Giesbrecht, F. G., and T. B. Whitaker. 1998. Investigations of the problems of assessing aflatoxin levels in peanuts. *Biometrics.* 54, 739-753.

2. Whitaker, T. B., J. W. Dickens, R. J. Monroe, and E. H. Wiser. 1972. Comparison of the observed distribution of the aflatoxin in shelled peanuts to the negative binomial distribution. *J. Assoc. Off. Anal. Chem.* 49, 590-593.

3. Aitchison, J. and J. A. C. Brown. 1973. The Lognormal Distribution with special reference to its uses in economics. Cambridge University Press, London, pp. 14-16.

4. Mendenhall, W., D. D. Wackerly, R. L. Scheaffer. 1990. Mathematical Statistics with Applications. Duxbury Press, Belmont CA, pp. 413-425.

APPENDIX 5. SAS PROGRAM TO DETERMINE THE DISTRIBUTIONAL PARAMETERS AND GOODNESS OF FIT FOR EACH OF THE THEORETICAL DISTRIBUTIONS USING THE VARIANCE ESTIMATES FROM CHAPTER 1.

Compound Gamma Distribution

```
%MACRO SUMMARY ;
U0 + 1 ;  IF PRED < .0625 THEN GOTO SKIP ;
U0 = U0 - 1 ;   U1 + 1 ;  IF PRED < .125 THEN GOTO SKIP ;
U1 = U1 - 1 ;   U2 + 1 ;  IF PRED < .1875 THEN GOTO SKIP ;
U2 = U2 - 1 ;   U3 + 1 ;  IF PRED < .25 THEN GOTO SKIP ;
U3 = U3 - 1 ;   U4 + 1 ;  IF PRED < .3125 THEN GOTO SKIP ;
U4 = U4 - 1 ;   U5 + 1 ;  IF PRED < .375 THEN GOTO SKIP ;
U5 = U5 - 1 ;   U6 + 1 ;  IF PRED < .4375 THEN GOTO SKIP ;
U6 = U6 - 1 ;   U7 + 1 ;  IF PRED < .5 THEN GOTO SKIP ;
U7 = U7 - 1 ;   U8 + 1 ;  IF PRED < .5625 THEN GOTO SKIP ;
U8 = U8 - 1 ;   U9 + 1 ;  IF PRED < .625 THEN GOTO SKIP ;
U9 = U9 - 1 ;   U10 + 1 ; IF PRED < .6875 THEN GOTO SKIP ;
U10 = U10 - 1 ; U11 + 1 ; IF PRED < .75 THEN GOTO SKIP ;
U11 = U11 - 1 ; U12 + 1 ; IF PRED < .8125 THEN GOTO SKIP ;
U12 = U12 - 1 ; U13 + 1 ; IF PRED < .875 THEN GOTO SKIP ;
U13 = U13 - 1 ; U14 + 1 ; IF PRED < .9375 THEN GOTO SKIP ;
U14 = U14 - 1 ; U15 + 1 ;
SKIP : ;
IF LAST THEN DO ;
  OUTPUT;
  END ;
KEEP LOT DISTRIB parm1 parm2 parm3 U0-U15 ;

PROC APPEND BASE = sdaja.sumcomgm FORCE ;
RUN ;

%MEND SUMMARY ;

/*********************************************/
/*   This code picks up the initial data.
/*********************************************/
data start;
infile 'd:\aj\bigjohn\aflatox\bjaldat2.csv' dlm=',';
    input lot sam tppb;
 if lot = 1 and tppb ^= .
 then output ;
```

```
data start; set start;
  proc sort; by tppb;
/*********************************************/
/*   This code computes the first three
/*   moments of the data set and obtains
/*   the count.
/*********************************************/
data moments ; set start end = lastone ;
keep lot u1 u2 u3 noobs maxtppb ;
u1 + tppb ;
u2 + tppb*tppb ;
u3 + tppb*tppb*tppb ;
if lastone then do ;
   u1 = u1/_n_ ;
   u2 = u2/_n_ ;
   u3 = u3/_n_ ;
   noobs = _n_ ;
   maxtppb = tppb ;
   output ;
   end ;
/*********************************************/
/*   Merge the moments data with the base
/*   data.
/*********************************************/
data init ; merge start moments ;
by lot ;
DATA INIT ; SET INIT ;
cum = _n_/noobs ;
expuord = (_n_-.375)/(noobs+.25) ;
run;
/*********************************************************/
/*   USES THE METHOD OF MOMENTS TO COMPUTE
/*   ESTIMATES OF LAMBDA AND BETA WITH ALPHA = .5,
/*    1, 1.5, 2, 2.5   FOR THE COMPOUND GAMMA DISTRIBUTION.
/*********************************************************/
DATA PRED ; SET INIT ;
DROP INTLAM SUM RATIO I ;
ALPHA = 0.5 ;  /****alpha can equal .5, 1, 1.5, 2, 2.5****/
LAMBDA = ((ALPHA+1)/ALPHA)*(U1*U1)/(U2-U1*U1) ;
BETA = (U2-U1*U1)/((ALPHA+1)*U1) ;
INTLAM = INT(LAMBDA) ; MAXLAM = MAX( 2*INTLAM,300 ) ;
RATIO = EXP( - LAMBDA ) ;
SUM = RATIO ;
DO I = 1 TO INTLAM ;
     RATIO = RATIO*LAMBDA/I ;
```

```
    SUM = SUM+RATIO*PROBGAM((TPPB+.5)/BETA,ALPHA*I) ;
    END ;
DO I = INTLAM+1 TO MAXLAM ;
    RATIO = RATIO*LAMBDA/I ;
    SUM = SUM+RATIO*PROBGAM((TPPB+.5)/BETA,ALPHA*I) ;
    IF RATIO < .00000001 THEN GOTO DONELOOP ;
    END ;
DONELOOP: ;
PRED = SUM ;
title 'Compound Gamma alpha=0.5 (mm)' ;
/*********************************************************/;
/*   THIS CODE COMPUTES SUMMARY STATISTICS
/*   DOCUMENTING THE QUALITY OF FIT.
/*********************************************************/;
DATA summary1 ;
SET  PRED (KEEP = LOT SAM NOOBS ALPHA LAMBDA
       BETA PRED CUM EXPUORD
        RENAME = ( ALPHA  = PARM1
                LAMBDA = PARM2
                 BETA = PARM3 )) END = LAST;
LENGTH DISTRIB $ 25 ;
DISTRIB = 'COMP GAMMA (cg05mm)' ;
 %SUMMARY
/*******************************************************/
/*   This code uses maximum likelihood to fit       */ ;
/*   lambda and beta (alpha = .5) of the Compound   */ ;
/*   Gamma distribution. This distribution is a     */ ;
/*   mixture of Gamma distributions with Poisson    */ ;
/*   weights.                                       */ ;
/*******************************************************/ ;
PROC NLIN DATA=INIT noprint METHOD = DUD SMETHOD = GOLDEN ;
PARMS LAMBDA = 3. 4.  BETA = 100 110 ;
BOUNDS .1 < LAMBDA ;
BOUNDS 5 < BETA ;
ALPHA = .5 ;
IF _ITER_ = 0 THEN IF _OBS_ = 1 THEN DO ;
  LAMBDA = ((ALPHA+1)/ALPHA)*(U1*U1)/(U2-U1*U1) ;
BETA = (U2-U1*U1)/((ALPHA+1)*U1) ;
  END ;
Y = 0 ;
IF TPPB = 0 THEN DO ;
   FX = EXP(-LAMBDA) ;
   END ;
        ELSE DO ;
        LMULT = LOG( LAMBDA ) + ALPHA*LOG( TPPB/BETA ) ;
```

```
                LRATIO = LOG( LAMBDA ) + (ALPHA-1)*LOG (TPPB/BETA) -
                LGAMMA(ALPHA);
   SUM = EXP( LRATIO ) ;
   LGM = LGAMMA(ALPHA) ;
   DO I = 2 TO MAX( 2*LAMBDA , 300 ) ;
      LGMOLD = LGM ; LGM = LGAMMA(I*ALPHA) ;
      LRATIO = LRATIO + LMULT - LOG( I ) + LGMOLD - LGM ;
      IF LRATIO > 100 THEN DO ;
            T = .25*LRATIO ;
            TE = EXP( T ) ;
            SUM = SUM + TE*TE*TE*TE ;
            END ;
            ELSE SUM = SUM + EXP( LRATIO ) ;
      IF LRATIO < -16. THEN GOTO DONELOOP ;
      END ;
   DONELOOP : FX = SUM*EXP(-TPPB/BETA)*EXP(-LAMBDA)/BETA ;
   END ;
MODEL Y = SQRT( -LOG( MAX( FX/10 , .0000000001 ) ) ) ;
OUTPUT OUT=FIT PARMS = LAMBDA BETA ;
title 'Compound Gamma alpha=.5 (ml)' ;
 /*********************************************************/
/*   COMPUTES THE CDF OF THE COMPOUND GAMMA          */
/*   USING THE ESTIMATES OF LAMBDA AND BETA          */
/*   (WITH ALPHA = .5, 1, 1.5, 2, 2.5) OBTAINED FROM THE
/*   PRECEEDING STEP. */
/*********************************************************/
DATA PRED ; SET FIT ;
DROP INTLAM SUM RATIO I ;
INTLAM = INT(LAMBDA) ; MAXLAM = MAX( 2*INTLAM , 300 ) ;
ALPHA = .5 ;  /*alpha can equal .5, 1, 1.5, 2, 2.5*/
RATIO = EXP( - LAMBDA ) ;
SUM = RATIO ;
DO I = 1 TO INTLAM ;
    RATIO = RATIO*LAMBDA/I ;
    SUM = SUM+RATIO*PROBGAM((TPPB+.5)/BETA,ALPHA*I) ;
    END ;
DO I = INTLAM+1 TO MAXLAM ;
    RATIO = RATIO*LAMBDA/I ;
    SUM = SUM+RATIO*PROBGAM((TPPB+.5)/BETA,ALPHA*I) ;
    IF RATIO < .0000001 THEN GOTO DONELOOP ;
    END ;
DONELOOP: ;
PRED = SUM ;
/*********************************************************/;
/*    THIS CODE COMPUTES SUMMARY STATISTICS          */;
```

```
/*    DOCUMENTING THE QUALITY OF FIT.                */;
/*****************************************************/;
DATA summary2 ;
SET PRED (KEEP = LOT SAM NOOBS ALPHA LAMBDA
        BETA PRED CUM EXPUORD
         RENAME = ( ALPHA  = PARM1
                  LAMBDA = PARM2
                    BETA = PARM3 )) END = LAST;
LENGTH DISTRIB $ 25 ;
DISTRIB = 'COMP GAMMA (cg05ml)' ;
 %SUMMARY
/*****************************************************/
/*  PRINTS OUT THE PARAMETERS AND RANGE OF BINS
/* ***************************************************/
DATA SUMMARY; SET SDAJA.SUMCOMGM;
PROC PRINT DATA=SUMMARY;
  VAR LOT DISTRIB PARM1 PARM2 PARM3 U0-U15 ;
  RUN;
PROC PRINT DATA=SDAJA.PREDCGMM; RUN;
/*****************************************************/
/*  GOODNESS OF FIT CALCULATIONS
/*****************************************************/
DATA ONE ; SET SDAJA.SUMCOMGM ;  ARRAY U (I) U0 - U15 ;
PDL = 0 ; E = 0 ;
ITEST = 0 ; C = NOOBS/16 ;
DO I = 1 TO 16 ;
   IF ITEST = 0 THEN DO ;
E = E + C ;
     IF U > 0 THEN DO ;
       ITEST = I ;
       PDL = U*((U/E)**.667 - 1 ) ;
       END ;
     END ;
           ELSE DO ;
     IF U > 0 THEN PDL = PDL + U*((U/C)**.667 - 1 ) ;
     END ;
   END ;
PDL = 9*PDL/5 ;
if distrib = 'COMP GAMMA (cg05mm)' then prob = probchi(pdl,16-3-itest+1) ;
if distrib = 'COMP GAMMA (cg1mm)' then prob = probchi(pdl,16-3-itest+1) ;
if distrib = 'COMP GAMMA (cg1.5mm)' then prob = probchi(pdl,16-3-itest+1) ;
if distrib = 'COMP GAMMA (cg2mm)' then prob = probchi(pdl,16-3-itest+1) ;
if distrib = 'COMP GAMMA (cg2.5mm)' then prob = probchi(pdl,16-3-itest+1) ;
if distrib = 'COMP GAMMA (cg05ml)' then prob = probchi(pdl,16-3-itest+1) ;
if distrib = 'COMP GAMMA (cg1ml)' then prob = probchi(pdl,16-3-itest+1) ;
```

```
if distrib = 'COMP GAMMA (cg1.5ml)' then prob = probchi(pdl,16-3-itest+1) ;
if distrib = 'COMP GAMMA (cg2ml)' then prob = probchi(pdl,16-3-itest+1) ;
if distrib = 'COMP GAMMA (cg2.5ml)' then prob = probchi(pdl,16-3-itest+1) ;
if distrib = 'COMP GAMMA (cgml)' then prob = probchi(pdl,16-4-itest+1) ;
KEEP  LOT DISTRIB ITEST PDL PROB ;
DATA MERGE1; MERGE SDAJA.SUMCOMGM ONE; RUN;
DATA SDAJA.SUMCOMGM; SET MERGE1;
PROC PRINT DATA=SDAJA.SUMCOMGM; RUN;
```

Negative Binomial Distribution

```
/*********************************************/
/*   THIS CODE PICKS UP THE INITIAL DATA.   */
/*********************************************/
DATA START;
INFILE 'C:\AJ\BIGJOHN\AFLATOX\BJALDAT2.CSV' DLM=',';
    INPUT LOT SAM TPPB;
 IF LOT = 1 AND TPPB ^= .
 THEN OUTPUT ;
DATA START; SET START;
  PROC SORT; BY TPPB;
/*********************************************/
/*   THIS CODE COMPUTES THE FIRST THREE     */
/*   MOMENTS OF THE DATA SET AND OBTAINS    */
/*   THE COUNT.                             */
/*********************************************/
DATA MOMENTS ; SET START END = LASTONE ;
KEEP LOT U1 U2 U3 NOOBS MAXTPPB ;
U1 + TPPB ;
U2 + TPPB*TPPB ;
U3 + TPPB*TPPB*TPPB ;
IF LASTONE THEN DO ;
  U1 = U1/_N_ ;
  U2 = U2/_N_ ;
  U3 = U3/_N_ ;
  NOOBS = _N_ ;
  MAXTPPB = TPPB ;
  OUTPUT ;
  END ;
/*********************************************/
/*   MERGE THE MOMENTS DATA WITH THE BASE   */
/*   DATA.                                  */
/*********************************************/
DATA INIT ; MERGE START MOMENTS ;
BY LOT ;
```

```
PROC SORT DATA=INIT; BY TPPB;
RUN;
/*******************************************************/
/*    THIS CODE USES THE SAMPLE MEAN AND VARIANCE    */ ;
/*    TO ESTIMATE THE PARAMETERS OF THE NEGATIVE     */ ;
/*    BINOMIAL AND COMPUTES THE CORRESPONDING CDF.   */ ;
/*    THE CDF IS STORED AS PRED.                     */ ;
/*******************************************************/ ;
DATA PRED; SET INIT;
DROP MEAN VAR Q TPPBOLD INCR KINCR FAC;
RETAIN MEAN VAR PK INCR Q FAC KINCR TPPBOLD PRED P K;
IF _N_ = 1 THEN DO;
  MEAN =U1;
  VAR =(32/31)*(U2 - U1*U1) ;
  P = MEAN/VAR ; Q = 1 - P ;
  K = MEAN*MEAN/(VAR-MEAN) ;
  TPPBOLD = 0 ; PRED = 0 ;
PK = P**K ; INCR = 1 ;
  KINCR = K-1 ; FAC = 0 ;
  END;
IF TPPB= 0 THEN DO ;
PRED = PK ;
  END ;
        ELSE DO ;
  DO X = TPPBOLD+1 TO TPPB ;
    FAC = FAC + 1 ;
    KINCR = KINCR + 1 ;
    INCR = INCR*KINCR*Q/FAC ;
    PRED = PRED + PK*INCR ;
    END ;
  TPPBOLD = TPPB ;
  END ;
TITLE 'NEGATIVE BINOMIAL (MM)' ;
/*******************************************************/ ;
/*    THIS CODE USES THE NLIN PROCEDURE TO OBTAIN     */ ;
/*    THE MAXIMUM LIKELIHOOD ESTIMATES OF THE        */ ;
/*    PARAMETERS OF THE NEGATIVE BINOMIAL.           */ ;
/*    THE METHOD OF MOMENTS ESTIMATORS ARE USED AS   */ ;
/*    STARTING VALUES FOR THE ITERATIVE FITTING.     */ ;
/*    THE PARAMETER ESTIMATES ARE STORED AS P & K    */ ;
/*    IN THE DATA SET FITTED1.                       */ ;
/*******************************************************/ ;
PROC NLIN DATA = INIT METHOD = DUD NOPRINT;
PARMS P = .01 K = 3 ;
BOUNDS .00001 < P < .99999 ;
```

```
BOUNDS .150 < K ;
IF _ITER_ = 0 THEN IF _OBS_ = 1 THEN DO ;
        P = U1/( U2-U1*U1 ) ;
        K = U1*U1/( (U2-U1*U1) - U1) ;
        END ;
RETAIN PK INCR Q FAC KINCR TPPBOLD PRED;
Y = 0 ;
Q = 1 - P ;
IF _OBS_ = 1 THEN DO ;
  TPPBOLD = 0 ;
  PRED = 0 ;
  PK = P**K ; INCR = 1 ; KINCR = K-1 ; FAC = 0 ;
END ;
IF TPPB = 0 THEN DO ;
  PRED = PK ;
  END ;
         ELSE DO ;
  DO X = TPPBOLD+1 TO TPPB ;
    FAC = FAC + 1 ;
    KINCR = KINCR + 1 ;
    INCR = INCR*KINCR*Q/FAC ;
    PRED = PK*INCR ;
    END ;
  TPPBOLD = TPPB ;
  END ;
MODEL Y = SQRT( -LOG( PRED ) ) ;
OUTPUT OUT = FIT PARMS = P K ;
TITLE 'NEGATIVE BINOMIAL (ML)' ;
/*****************************************************/
/*    THIS SECTION OF CODE USES THE FITTED         */ ;
/*    PARAMETERS P & K TO COMPUTE THE FITTED       */ ;
/*    DISTRIBUTION AND STORE IT AS THE VARIABLE    */ ;
/*    PRED.                                        */ ;
/*****************************************************/ ;
DATA PRED ; SET FIT ;
RETAIN PRED PK INCR FAC KINCR TPPBOLD ;
DROP Q PK INCR KINCR TPPBOLD FAC X ;
Q = 1 - P ;
IF _N_ = 1 THEN DO ;
   TPPBOLD = 0 ; PRED = 0 ;
   PK = P**K ; INCR = 1 ;
   KINCR = K-1 ; FAC = 0 ;
   END ;
IF TPPB <= 0 THEN DO ;
   PRED = PK ;
```

```
   END ;
         ELSE DO ;
   DO X = TPPBOLD+1 TO TPPB ;
      FAC = FAC + 1 ;
      KINCR = KINCR + 1 ;
      INCR = INCR*KINCR*Q/FAC ;
      PRED = PRED + PK*INCR ;
END ;
   TPPBOLD = TPPB ;
   END ;
/*********************************************************/;
/*   THIS CODE COMPUTES SUMMARY STATISTICS          */;
/*   DOCUMENTING THE QUALITY OF FIT.                */;
/*********************************************************/;
DATA SUMMARY ;
SET PRED (KEEP = LOT SAM NOOBS P K
       PRED CUM EXPUORD) END = LAST;
PARM3 = . ;
LENGTH DISTRIB $ 25 ;
DISTRIB = 'NEG BINOMIAL (ML)' ;
         U0 + 1 ;  IF PRED < .0625 THEN GOTO SKIP ;
U0 = U0 - 1 ;   U1 + 1 ;  IF PRED < .125 THEN GOTO SKIP ;
U1 = U1 - 1 ;   U2 + 1 ;  IF PRED < .1875 THEN GOTO SKIP ;
U2 = U2 - 1 ;   U3 + 1 ;  IF PRED < .25 THEN GOTO SKIP ;
U3 = U3 - 1 ;   U4 + 1 ;  IF PRED < .3125 THEN GOTO SKIP ;
U4 = U4 - 1 ;   U5 + 1 ;  IF PRED < .375 THEN GOTO SKIP ;
U5 = U5 - 1 ;   U6 + 1 ;  IF PRED < .4375 THEN GOTO SKIP ;
U6 = U6 - 1 ;   U7 + 1 ;  IF PRED < .5 THEN GOTO SKIP ;
U7 = U7 - 1 ;   U8 + 1 ;  IF PRED < .5625 THEN GOTO SKIP ;
U8 = U8 - 1 ;   U9 + 1 ;  IF PRED < .625 THEN GOTO SKIP ;
U9 = U9 - 1 ;   U10 + 1 ; IF PRED < .6875 THEN GOTO SKIP ;
U10 = U10 - 1 ; U11 + 1 ; IF PRED < .75 THEN GOTO SKIP ;
U11 = U11 - 1 ; U12 + 1 ; IF PRED < .8125 THEN GOTO SKIP ;
U12 = U12 - 1 ; U13 + 1 ; IF PRED < .875 THEN GOTO SKIP ;
U13 = U13 - 1 ; U14 + 1 ; IF PRED < .9375 THEN GOTO SKIP ;
U14 = U14 - 1 ; U15 + 1 ;
SKIP : ;
IF LAST THEN DO ;
  OUTPUT;
  END ;
KEEP LOT DISTRIB P K U0-U15 ;
PROC PRINT DATA=SUMMARY;
 VAR LOT DISTRIB P K U0-U15 ;
 RUN;
```

```
DATA ONE ; SET SUMMARY;  ARRAY U (I) U0 - U15 ;
*    IF DISTRIB = 'NEG BINOMIAL (MM)' THEN DELETE ;
*    IF DISTRIB = 'NEG BINOMIAL (ML)' THEN DELETE ;
PDL = 0 ; E = 0 ;
ITEST = 0 ; C = NOOBS/16 ;
DO I = 1 TO 16 ;
  IF ITEST = 0 THEN DO ;
    E = E + C ;
    IF U > 0 THEN DO ;
      ITEST = I ;
      PDL = U*((U/E)**.667 - 1 ) ;
      END ;
    END ;
          ELSE DO ;
    IF U > 0 THEN PDL = PDL + U*((U/C)**.667 - 1 ) ;
    END ;
  END ;
PDL = 9*PDL/5 ;
                IF DISTRIB = 'NEG BINOMIAL (MM)' THEN PROB = 1-
                PROBCHI(PDL,16-3-ITEST+1) ;
                IF DISTRIB = 'NEG BINOMIAL (ML)' THEN PROB = 1-
                PROBCHI(PDL,16-3-ITEST+1) ;
KEEP  LOT DISTRIB ITEST PDL PROB ;
DATA MERGE1; MERGE SUMMARY ONE; BY LOT; RUN;
PROC PRINT DATA=MERGE1; RUN;
```

Log Normal Distribution

```
%MACRO SUMMARY ;
        U0 + 1 ;  IF PRED < .0625 THEN GOTO SKIP ;
U0 = U0 - 1 ;   U1 + 1 ;  IF PRED < .125 THEN GOTO SKIP ;
U1 = U1 - 1 ;   U2 + 1 ;  IF PRED < .1875 THEN GOTO SKIP ;
U2 = U2 - 1 ;   U3 + 1 ;  IF PRED < .25 THEN GOTO SKIP ;
U3 = U3 - 1 ;   U4 + 1 ;  IF PRED < .3125 THEN GOTO SKIP ;
U4 = U4 - 1 ;   U5 + 1 ;  IF PRED < .375 THEN GOTO SKIP ;
U5 = U5 - 1 ;   U6 + 1 ;  IF PRED < .4375 THEN GOTO SKIP ;
U6 = U6 - 1 ;   U7 + 1 ;  IF PRED < .5 THEN GOTO SKIP ;
U7 = U7 - 1 ;   U8 + 1 ;  IF PRED < .5625 THEN GOTO SKIP ;
U8 = U8 - 1 ;   U9 + 1 ;  IF PRED < .625 THEN GOTO SKIP ;
U9 = U9 - 1 ;   U10 + 1 ; IF PRED < .6875 THEN GOTO SKIP ;
U10 = U10 - 1 ; U11 + 1 ; IF PRED < .75 THEN GOTO SKIP ;
U11 = U11 - 1 ; U12 + 1 ; IF PRED < .8125 THEN GOTO SKIP ;
U12 = U12 - 1 ; U13 + 1 ; IF PRED < .875 THEN GOTO SKIP ;
U13 = U13 - 1 ; U14 + 1 ; IF PRED < .9375 THEN GOTO SKIP ;
U14 = U14 - 1 ; U15 + 1 ;
```

```
SKIP : ;
IF LAST THEN DO ;
  OUTPUT;
  END ;
KEEP LOT DISTRIB PARM1 PARM2 PARM3 U0-U15 ;
PROC APPEND BASE = SDAJA.SUMLOGN FORCE ;
RUN ;
%MEND SUMMARY ;
/*********************************************/
/*   THIS CODE PICKS UP THE INITIAL DATA.   */
/*********************************************/
DATA START;
INFILE 'D:\AJ\BIGJOHN\AFLATOX\BJALDAT2.CSV' DLM=',';
    INPUT LOT SAM TPPB;
 IF LOT = 1 AND TPPB ^= .
 THEN OUTPUT ;
PROC SORT; BY TPPB;
/*********************************************/
/*   THIS CODE COMPUTES THE FIRST THREE     */
/*   MOMENTS OF THE DATA SET AND OBTAINS    */
/*   THE COUNT.                             */
/*********************************************/
DATA MOMENTS ; SET START END = LASTONE ;
KEEP LOT U1 U2 U3 NOOBS MAXTPPB ;
U1 + TPPB ;
U2 + TPPB*TPPB ;
U3 + TPPB*TPPB*TPPB ;
IF LASTONE THEN DO ;
  U1 = U1/_N_ ;
  U2 = U2/_N_ ;
  U3 = U3/_N_ ;
  NOOBS = _N_ ;
  MAXTPPB = TPPB ;
  OUTPUT ;
  END ;
/*********************************************/
/*   MERGE THE MOMENTS DATA WITH THE BASE   */
/*   DATA.                                  */
/*********************************************/
DATA INIT ; MERGE START MOMENTS ;
BY LOT ;
RUN;
/*****************************************************/
/*    THIS CODE USES MAXIMUM LIKELIHOOD             */ ;
/*    TO ESTIMATE THE PARAMETERS OF THE THREE-      */ ;
```

```
/*    PARAMETER LOGNORMAL DISTRIBUTION.            */ ;
/*****************************************************/
PROC NLIN DATA = INIT NOPRINT METHOD = DUD ;
PARMS ALPHA = 0 MU = 1 SIGMA = 1 ;
BOUNDS .1 < SIGMA   ;
IF _ITER_=0 THEN IF _OBS_=1 THEN DO;
  MU=2*LOG(U1)-.5*LOG(U2);
  SIGMA=LOG(U2)-2*LOG(U1);
    END;
Y = 0 ;
TSIGMA = MAX( SIGMA , .001 ) ;
IF TPPB = 0 THEN DO ;
   FX = PROBNORM( ( LOG( MAX(ALPHA+.5 , .001) )-MU )
                  /TSIGMA  ) ;
   END ;
   ELSE DO ;
   FX = PROBNORM( ( LOG( MAX(TPPB+ALPHA+.5 , .001) )-MU )
                  /TSIGMA  )
     - PROBNORM( ( LOG( MAX(TPPB+ALPHA-.5 , .001) )-MU )
                  /TSIGMA  ) ;
   END ;
MODEL Y = SQRT( ABS( LOG( MAX(FX,.00000000001) ) ) ) ;
OUTPUT OUT = FIT PARMS = ALPHA MU SIGMA ;
TITLE 'LOGNORMAL (ML)' ;
/*****************************************************/
/*    THIS CODE COMPUTES THE CDF OF THE THREE-       */ ;
/*    PARAMETER LOGNORMAL DISTRIBUTION, USING THE    */ ;
/*    ESTIMATED THE PARAMETERS VALUES.               */ ;
/*    THE CDF IS STORED AS PRED.                     */ ;
/*****************************************************/
DATA PRED ; SET FIT;
PRED = PROBNORM( ( LOG( MAX(TPPB+ALPHA+.5 , .001) )-MU )
                  /SIGMA  ) ;
DATA SUMMARY1; SET PRED
        (KEEP = LOT SAM NOOBS ALPHA MU SIGMA
                  PRED CUM EXPUORD
         RENAME = (ALPHA=PARM1
                             MU=PARM2
                           SIGMA=PARM3))END=LAST;
LENGTH DISTRIB $ 25 ;
DISTRIB = '3-PAR LOGNORMAL' ;
%SUMMARY

PROC PRINT DATA=SDAJA.SUMLOGN;
  VAR LOT DISTRIB PARM1 PARM2 PARM3 U0-U15;
```

```
 RUN;

DATA ONE ; SET SDAJA.SUMLOGN ;  ARRAY U (I) U0 - U15 ;
PDL = 0 ; E = 0 ;
ITEST = 0 ; C = NOOBS/16 ;
DO I = 1 TO 16 ;
  IF ITEST = 0 THEN DO ;
    E = E + C ;
    IF U > 0 THEN DO ;
      ITEST = I ;
      PDL = U*((U/E)**.667 - 1 ) ;
      END ;
    END ;
         ELSE DO ;
    IF U > 0 THEN PDL = PDL + U*((U/C)**.667 - 1 ) ;
    END ;
  END ;
PDL = 9*PDL/5 ;
        IF DISTRIB = '3-PAR LOGNORMAL' THEN PROB =
        PROBCHI(PDL,MAX(16-4-ITEST+1,1)) ;
KEEP  LOT DISTRIB ITEST PDL PROB ;

DATA MERGE1; MERGE SDAJA.SUMLOGN ONE; RUN;
DATA SDAJA.SUMLOGN; SET MERGE1;
PROC PRINT DATA=SDAJA.SUMLOGN; RUN;
```

Truncated Normal Distribution

```
%MACRO SUMMARY ;
        U0 + 1 ;  IF PRED < .0625 THEN GOTO SKIP ;
U0 = U0 - 1 ;   U1 + 1 ;  IF PRED < .125 THEN GOTO SKIP ;
U1 = U1 - 1 ;   U2 + 1 ;  IF PRED < .1875 THEN GOTO SKIP ;
U2 = U2 - 1 ;   U3 + 1 ;  IF PRED < .25 THEN GOTO SKIP ;
U3 = U3 - 1 ;   U4 + 1 ;  IF PRED < .3125 THEN GOTO SKIP ;
U4 = U4 - 1 ;   U5 + 1 ;  IF PRED < .375 THEN GOTO SKIP ;
U5 = U5 - 1 ;   U6 + 1 ;  IF PRED < .4375 THEN GOTO SKIP ;
U6 = U6 - 1 ;   U7 + 1 ;  IF PRED < .5 THEN GOTO SKIP ;
U7 = U7 - 1 ;   U8 + 1 ;  IF PRED < .5625 THEN GOTO SKIP ;
U8 = U8 - 1 ;   U9 + 1 ;  IF PRED < .625 THEN GOTO SKIP ;
U9 = U9 - 1 ;   U10 + 1 ; IF PRED < .6875 THEN GOTO SKIP ;
U10 = U10 - 1 ; U11 + 1 ; IF PRED < .75 THEN GOTO SKIP ;
U11 = U11 - 1 ; U12 + 1 ; IF PRED < .8125 THEN GOTO SKIP ;
U12 = U12 - 1 ; U13 + 1 ; IF PRED < .875 THEN GOTO SKIP ;
U13 = U13 - 1 ; U14 + 1 ; IF PRED < .9375 THEN GOTO SKIP ;
U14 = U14 - 1 ; U15 + 1 ;
```

```
SKIP : ;
IF LAST THEN DO ;
  OUTPUT;
  END ;
KEEP LOT DISTRIB PARM1 PARM2 PARM3 U0-U15 ;
PROC APPEND BASE = SDAJA.SUMTNORM FORCE ;
RUN ;
%MEND SUMMARY ;
/*********************************************/
/*   THIS CODE PICKS UP THE INITIAL DATA.   */
/*********************************************/
DATA START;
INFILE 'D:\AJ\BIGJOHN\AFLATOX\BJALDAT2.CSV' DLM=',';
    INPUT LOT SAM TPPB;
 IF LOT = 1 AND TPPB ^= .
 THEN OUTPUT ;
PROC SORT; BY TPPB;
/*********************************************/
/*   THIS CODE COMPUTES THE FIRST THREE     */
/*   MOMENTS OF THE DATA SET AND OBTAINS    */
/*   THE COUNT.                             */
/*********************************************/
DATA MOMENTS ; SET START END = LASTONE ;
KEEP LOT SAM U1 U2 U3 NOOBS MAXTPPB ;
U1 + TPPB ;
U2 + TPPB*TPPB ;
U3 + TPPB*TPPB*TPPB ;
IF LASTONE THEN DO ;
  U1 = U1/_N_ ;
  U2 = U2/_N_ ;
  U3 = U3/_N_ ;
  NOOBS = _N_ ;
  MAXTPPB = TPPB ;
  OUTPUT ;
  END ;
/*********************************************/
/*   MERGE THE MOMENTS DATA WITH THE BASE   */
/*   DATA.                                  */
/*********************************************/
DATA INIT ; MERGE START MOMENTS ;
BY LOT ;
RUN;
/*****************************************************/
/*    THIS CODE USES MAXIMUM LIKELIHOOD             */ ;
/*    TO ESTIMATE THE PARAMETERS OF THE             */ ;
```

```
/*    TRUNCATED NORMAL DISTRIBUTION.                */ ;
/*****************************************************/
PROC NLIN DATA = INIT METHOD = DUD NOPRINT;
PARMS MU = 500 SIGMA = 200 ;
BOUNDS 50  < MU   ;
BOUNDS 10  < SIGMA  ;
IF _ITER_ = 0 THEN IF _OBS_ = 1 THEN DO
  MU = U1 ;
  SIGMA = SQRT( U2 - U1*U1) ;
  END ;
Y = 0 ;
TSIGMA = MAX( SIGMA , .01 ) ;
IF TPPB = 0 THEN DO ;
  FX = PROBNORM( ( +5 - MU )/TSIGMA  ) ;
  END ;
        ELSE DO ;
  FX = .398942280401*EXP(-.5*((TPPB-MU)/TSIGMA)**2 )
                  /TSIGMA  ;
  END ;
MODEL Y = SQRT( ABS( LOG( MAX(FX,.00000000001) ) ) ) ;
OUTPUT OUT = FIT PARMS = MU SIGMA ;
TITLE 'TRUNCATED NORMAL (ML)' ;
/*****************************************************/
/*    THIS CODE COMPUTES THE CDF OF THE TRUNCATED    */ ;
/*    NORMAL DISTRIBUTION, USING THE                 */ ;
/*    ESTIMATED THE PARAMETERS VALUES.               */ ;
/*    THE CDF IS STORED AS PRED.                     */ ;
/*****************************************************/
DATA PRED ; SET FIT ;
PRED = PROBNORM( (TPPB+.5 -MU )/SIGMA  ) ;
/*****************************************************/;
/*    THIS CODE COMPUTES SUMMARY STATISTICS          */;
/*    DOCUMENTING THE QUALITY OF FIT.                */;
/*****************************************************/;
DATA SUMARY1 ;
SET PRED (KEEP = LOT SAM NOOBS MU SIGMA
          PRED CUM EXPUORD
        RENAME = (   MU   = PARM1
                 SIGMA = PARM2 )) END = LAST;
PARM3 = . ;
LENGTH DISTRIB $ 25 ;
DISTRIB = 'TRUNCATED NORMAL' ;
 %SUMMARY

PROC PRINT DATA=SDAJA.SUMTNORM;
```

```
 VAR LOT DISTRIB PARM1 PARM2 PARM3 U0-U15;
 RUN;

DATA ONE ; SET SDAJA.SUMTNORM ;  ARRAY U (I) U0 - U15 ;
PDL = 0 ; E = 0 ;
ITEST = 0 ; C = NOOBS/16 ;
DO I = 1 TO 16 ;
  IF ITEST = 0 THEN DO ;
    E = E + C ;
    IF U > 0 THEN DO ;
      ITEST = I ;
      PDL = U*((U/E)**.667 - 1 ) ;
      END ;
    END ;
          ELSE DO ;
    IF U > 0 THEN PDL = PDL + U*((U/C)**.667 - 1 ) ;
    END ;
  END ;
PDL = 9*PDL/5 ;
         IF DISTRIB = 'TRUNCATED NORMAL' THEN PROB =
         PROBCHI(PDL,MAX(16-3-ITEST+1,1)) ;
KEEP  LOT DISTRIB ITEST PDL PROB ;
DATA MERGE1; MERGE SDAJA.SUMTNORM ONE; RUN;
DATA SDAJA.SUMTNORM; SET MERGE1;
PROC PRINT DATA=SDAJA.SUMTNORM; RUN;
```

APPENDIX 6. SAMPLE AFLATOXIN TEST RESULTS USED TO DETERMINE THE DISTRIBUTIONAL PARAMETERS FOR EACH OF THE THEORETICAL DISTRIBUTIONS.

Appendix 6 Table 1: Sample aflatoxin test results used to determine the distributional parameters for each of the theoretical distributions[a].

	Lot																							
Sample	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	20.4	20.0	45.4	42.4	62.2	157.0	154.4	489.6
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	33.5	22.7	48.0	57.0	69.7	165.2	192.8	490.6
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	36.7	26.5	49.0	62.7	75.9	190.4	203.4	516.5
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.1	38.5	33.8	50.9	66.2	80.6	194.1	208.8	548.2
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.6	0.0	0.0	0.0	9.0	40.6	36.3	54.9	71.0	80.6	210.7	214.9	563.4
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.9	0.0	5.6	0.0	13.4	45.8	38.2	57.1	71.7	82.7	212.4	227.0	566.1
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.1	0.0	7.7	0.0	16.5	46.3	39.2	57.4	73.2	83.0	216.5	233.1	578.4
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.3	0.0	8.1	10.5	21.7	47.4	44.2	69.8	80.4	87.4	221.3	236.6	579.5
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.6	0.0	8.3	10.6	22.3	47.9	44.2	73.0	84.1	91.7	222.9	239.2	584.0
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.5	0.0	13.6	11.0	23.4	48.0	45.2	73.5	85.4	92.4	228.1	240.1	588.1
11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.6	0.0	13.8	11.4	23.5	48.7	47.1	75.9	87.5	93.7	231.0	255.2	595.6
12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.7	0.0	15.7	11.9	23.9	49.8	50.8	87.9	88.6	94.1	236.3	255.4	596.2
13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.4	10.5	6.1	16.1	13.9	26.4	50.5	54.0	88.9	95.2	100.2	256.1	266.6	598.1
14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.1	13.8	8.2	18.5	16.6	27.4	51.1	54.8	89.0	95.9	101.1	257.1	271.9	604.1
15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.3	14.3	10.5	19.5	17.8	29.2	53.2	55.4	93.4	98.2	107.0	259.0	286.4	604.8
16	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.8	15.2	10.9	23.5	20.0	30.6	54.6	57.2	94.6	99.8	107.4	261.8	289.1	627.9
17	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.5	15.2	12.0	24.7	21.3	30.8	56.9	57.9	97.6	99.8	109.5	267.9	289.3	630.3
18	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.2	9.0	16.5	14.6	25.3	22.7	31.4	56.9	58.6	99.4	102.5	111.0	271.7	293.9	642.2
19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.6	0.0	6.8	9.3	16.8	15.8	27.0	23.9	31.8	59.0	58.9	102.8	103.7	115.3	279.6	300.3	647.9
20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.3	0.0	7.7	9.8	17.0	20.3	30.7	27.6	32.9	59.4	60.3	105.2	104.2	117.8	290.9	303.3	668.5
21	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.2	0.0	10.4	10.1	17.7	22.8	34.9	28.5	33.2	60.8	63.4	105.7	105.5	118.0	295.1	320.6	684.2
22	0.0	0.0	0.0	0.0	0.0	0.0	4.1	8.0	4.9	16.4	11.4	18.3	23.6	37.9	30.1	41.0	60.9	65.6	106.8	108.2	120.1	307.3	337.1	694.5
23	0.0	0.0	0.0	0.0	0.0	0.0	4.9	8.1	5.7	18.5	13.2	18.5	27.0	40.2	31.9	45.1	62.5	67.6	110.8	110.1	120.9	314.7	338.4	695.4
24	0.0	0.0	0.0	0.0	0.0	0.0	7.8	10.3	10.6	19.2	14.8	20.6	27.3	42.7	33.1	46.9	65.3	68.6	112.3	112.2	125.1	319.4	338.7	709.6
25	0.0	0.0	0.0	0.0	0.0	0.0	8.1	17.6	12.8	20.5	15.1	21.2	32.3	42.9	38.8	59.2	68.1	75.0	116.7	112.4	127.7	326.1	344.0	727.9
26	0.0	0.0	0.0	0.0	0.0	0.0	10.1	18.7	19.5	21.7	18.3	21.6	36.5	49.6	50.6	59.5	69.6	76.2	129.5	112.6	134.3	346.7	350.2	740.8
27	0.0	0.0	0.0	0.0	0.0	0.0	11.8	19.1	24.0	21.9	26.8	27.2	37.6	53.6	52.0	62.8	74.5	76.5	129.7	113.5	137.9	350.5	370.7	761.9
28	0.0	0.0	0.0	0.0	0.0	0.0	11.9	26.1	24.4	23.1	37.1	27.7	45.1	54.1	52.6	64.6	78.4	77.3	130.2	119.4	138.4	359.3	376.6	831.8
29	0.0	0.0	0.0	0.0	0.0	0.0	15.0	27.4	24.6	26.0	38.5	32.5	45.4	61.1	54.9	65.2	85.0	79.3	132.7	120.2	164.7	378.3	380.9	868.5
30	0.0	0.0	0.0	0.0	0.0	3.5	19.9	31.6	34.7	26.4	40.4	34.6	50.8	64.6	62.2	69.3	89.5	92.3	147.2	128.7	170.5	378.9	387.4	886.4
31	0.0	0.0	0.0	0.0	25.0	11.2	27.9	32.5	38.8	39.5	41.6	35.7	56.9	.	89.9	82.0	110.3	115.3	151.3	132.3	204.7	407.5	394.3	1150.5
32	0.0	0.0	0.0	9.4	.	21.7	31.5	.	42.7	41.1	48.5	39.2	79.0	.	90.0	91.6	111.7	186.0	170.3	139.6	248.0	444.4	479.7	1205.5

[a] . denotes missing data.

APPENDIX 7. SAS PROGRAM USED TO PRODUCE OPERATING CHARACTERISTIC CURVES FOR THE COMPOUND GAMMA DISTRIBUTION USING THE METHOD OF MOMENTS AND AN ALPHA EQUAL TO 2.5.

```
*********************************************************
*  OPERATING CHARACTERISTIC CURVES FOR SHELLED CORN    *
*  ANDERS JOHANSSON        3/3/97                *
*********************************************************;
*********************************************************
*  INPUT PARAMETER VALUES:                        *
*                                                 *
*    MSA     = STARTING MEAN AFLATOXIN LEVEL (PPB)     *
*                                                 *
*    MEA     = ENDING MEAN AFLATOXIN LEVEL (PPB)       *
*                                                 *
*    MIA     = INCREMENTAL AFLATOXIN LEVEL (PPB)       *
*                                                 *
*    NS      = INITIAL SAMPLE SIZE (KG)             *
*                                                 *
*    NSS     = SUBSAMPLE SIZE (G)                  *
*                                                 *
*    NA      = NUMBER OF ANALYSES                   *
*                                                 *
*     XH     = SAMPLE GUIDELINE LEVEL               *
*                                                 *
*********************************************************;
DATA INTIAL;
PRODUCT='SHELLED CORN';
MSA=1; MEA=200; MIA=1;
NS=2.5; NSS=50; NA=1;
XH=100;

DATA A; SET INTIAL;
ARRAY CPROB{100};
DO I=1 TO 100;   CPROB{I}=0;  END;
*********************************************************
*                                                 *
*  BEGIN LOOP OF MEAN AFLATOXIN LEVELS FOR COMPUTING    *
*    PROBABILITY VALUES.                          *
*                                                 *
*  THE SHAPE PARAMETER MAY BE SET AS AN INPUT VALUE     *
*    OR COMPUTED AS A FUNCTION OF THE MEAN AFLATOXIN    *
*    LEVEL.                                       *
```

```
*                                                    *
*****************************************************;
M=0; ALPHA=0; LAMBDA=0; BETA=0; RATIO=0;
DO J=1 TO 100;   CPROB{J}=1;     END;
DO M=MSA TO MEA BY MIA;
*****************************************************
*                                                    *
*  COMPUTE VARIANCE COMPONENTS - WILL NEED TO DOCUMENT *
*    REFERENCES FOR PARAMETER VALUES.                *
*                                                    *
*  VS   = VARIANCE OF SAMPLING                       *
*                                                    *
*  VSS   = VARIANCE OF SUBSAMPLING                   *
*                                                    *
*  VA   = VARIANCE OF ANALYTICAL METHOD              *
*                                                    *
*  VAR   = TOTAL VARIANCE                            *
*        = VSAM + VSUB + VANA                        *
*                                                    *
*****************************************************;

IF PRODUCT='SHELLED CORN' THEN DO;
VS=((2.5*0.45359)/NS)*(EXP(2.430175286)*M**0.976870993);
          VSS=(50/NSS)*((EXP(0.32418533)*M**1.266793664)-(EXP(-
          1.944936)*M**1.159129));
VA=(1/NA)*(EXP(-1.944936)*M**1.159129);

DO X=1 TO XH;
CUM=0;
ALPHA = 2.5 ;
LAMBDA = ((ALPHA+1)/ALPHA)*(M*M)/(VAR) ;
BETA = (VAR)/((ALPHA+1)*M) ;
INTLAM = INT(LAMBDA) ; MAXLAM = MAX( 2*INTLAM,300 ) ;
RATIO = EXP( - LAMBDA ) ;
SUM = RATIO ;
DO I = 1 TO INTLAM ;
    RATIO = RATIO*LAMBDA/I ;
    SUM = SUM+RATIO*PROBGAM((X)/BETA,ALPHA*I) ;
    END ;
DO I = INTLAM+1 TO MAXLAM ;
    RATIO = RATIO*LAMBDA/I ;
    SUM = SUM+RATIO*PROBGAM((X)/BETA,ALPHA*I) ;
    IF RATIO < .00000001 THEN GOTO DONELOOP ;
    END ;
DONELOOP: ;
```

```
PRED = SUM ;
CUM=CUM+PRED;
CPROB{X}=CUM;  END;
OUTPUT;
END;
RUN;

DATA ZERO;   L=0;
   ARRAY CPROB{100};
   DO L=1 TO 100;  CPROB{L}=1;  END;
   M=0;

DATA OCCG; SET A ZERO ; RUN;
PROC SORT DATA=OCCG; BY M;
PROC PRINT DATA=OCCG;
  TITLE 'OC CURVES FOR VARIOUS ACCEPT LEVELS';
  VAR NS M CPROB5 CPROB10 CPROB15 CPROB20 ;
RUN;
```