Screening Groundnuts for Seed Resistance to *Aspergillus flavus*: Statistical Approaches to Data Evaluation

Murari Singh , V.K. Mehan , and D. McDonald[1]

Abstract

Environmental factors influence the degree of groundnut seed infection by Aspergillus flavus *and other fungi. This complicates resistance screening over seasons and locations as levels of infection can vary considerably within a genotype. Statistical methods were used to separate genotypes into different resistance/susceptibility categories and to ensure a stable basis for comparisons of control cultivar and test genotypes across environments. An approach was also adopted for comparing the degree and distribution of resistance in spanish and valencia type groundnuts. The establishment of such procedures would facilitate interpretation of screening data from different environments.*

Résumé

Sélection des arachides pour la résistance des graines à *Aspergillus flavus*—analyse statistique des données : *Les facteurs d'environnement influencent l'intensité de l'infection des graines d'arachides par* Aspergillus flavus *et d'autres champignons. Ce phénomène complique la sélection du matériel résistant lorsque les essais sont effectués pendant plusieurs saisons et à divers sites, puisque les niveaux d'infection varient considérablement pour le même génotype. Des méthodes statistiques ont permis de classer les génotypes dans différentes catégories de résistance/sensibilité, et de donner une base sûre de comparaison entre les nouveaux génotypes à tester et les temoins pour différents sites. Cette approche a été également adoptée pour comparer le degré et la distribution de la résistance des types Spanish et Valencia. La mise au point de ces procédés statistiques facilitera l'interprétation des données obtenues dans différents sites expérimentaux.*

Resúmene

La selección de cacahuates basada en la resistencia de sus semillas a *Aspergillus flavus*: Métodos estadísticos para evaluar datos : *Los factores del medio ambiente influyen en la severidad con que ocurre la infección de la semilla de cacahuate por* Aspergillus flavus *y otros hongos. Esto complica la selección para lograr la resistencia a través de varios ciclos de cultivo y en diferentes localidades, debido a que los niveles de infección pueden variar considerablemente dentro de los genotipos. Se utilizaron métodos estadísticos para separar los genotipos en diferentes categorías*

1. Statistician, 2. Groundnut Pathologist, and 3. Principal Groundnut Pathologist, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Andhra Pradesh 502 324, India.

ICRISAT Conference Paper no. CP 436.

Citation: ICRISAT (International Crops Research Institute for the Semi-Arid Tropics). 1989. Aflatoxin contamination of groundnut: proceedings of the International Workshop, 6-9 Oct 1987, ICRISAT Center, India. Patancheru, A.P. 502 324, India: ICRISAT.

de la relación resistencia/susceptibilidad, y en esta forma asegurar una base de comparación estable entre el cultivar testigo y los genotipos bajo prueba, en todos los ambientes. Se adoptó, asimismo, un enfoque específico para la comparación del grado y distribución de la resistencia, en cacahuates de los tipos Spanish y Valencia. La adopción de estos procedimientos facilitarían la interpretación de los resultados de las pruebas de selección efectuados en diferentes condiciones ambientales.

Introduction

By screening groundnut genotypes for resistance to seed colonization by *Aspergillus flavus* Link ex Fries, in vitro, they can easily be classified as resistant, susceptible, or highly susceptible on the basis of arbitrarily set percentages of seeds colonized (LaPrade et al. 1973, Mixon and Rogers 1973, Mehan and McDonald 1980). However, when screening groundnuts for resistance to natural seed infection by the fungus in the field, it is not easy to identify resistant genotypes on the basis of arbitrarily set levels of seed infection, because environmental factors such a moisture, soil temperature in the pod zone, and soil type can influence *A. flavus* infectio.. .. genotype. Levels of seed infection within specific genotypes show very little variation between replicate samples, but levels of infection can vary considerably between trials, locations, or seasons. In such situations, genotypic resistance can best be measured in relation to the reactions of standard resistant and susceptible control genotypes. Thus a genotype can be considered resistant to the fungus if its reaction to seed infection is similar to that of a resistant control genotype in the same environment. In view of this concept, the reaction of a genotype to *A. flavus* can be represented by the probability distribution of *A. flavus* seed infection levels in a given environment, and one of the following approaches used for resistance screening.

Three situations frequently met in practice are discussed, and ways to screen genotypes resistant to seed infection by *A. flavus* considered.

The levels of resistance distributed in spanish and valencia types of groundnut are also compared.

Statistical Approaches for Screening Genotypes for Seed Resistance to Infection by *Aspergillus flavus*

Statistical methods are discussed in relation to three types of situation prevalent experimentation.

Comparison of Distribution Functions

Let T stand for a test genotype and R for the resistant control genotype. The frequency distribution of the level of seed infection by *A. flavus* follows that shown in Figure 1.
A test genotype (T) can be considered resistant if the probability distribution of its seed infection level is identical to that of a resistant control genotype (R) in a given environment. To illustrate this point, let the probability distribution functions of R and T be denoted by $F_R(x)$ and $F_T(x)$,

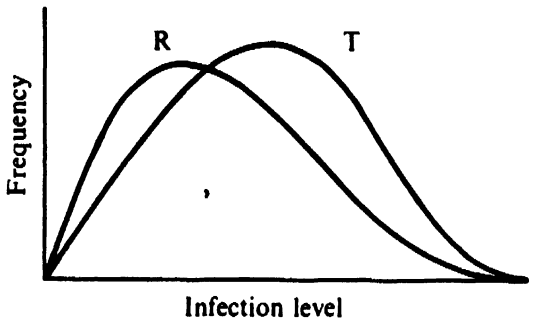

Figure 1. Distribution of infection levels in a resistant (R) and a test (T) groundnut genotype.

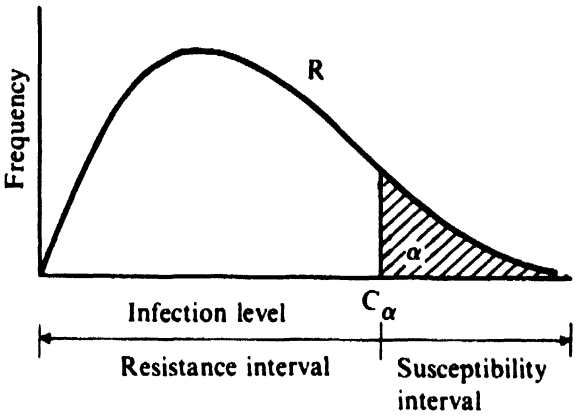

Figure 2. Distribution of infection level in a resistant (R) groundnut genotype and critical point (C_α) used to select resistant genotypes.

...re x is the seed infection level. The test genotype will be resistant relative to the resistant control genotype if the hypothesis.

H_0 : $F_T(x) = F_R(x)$ for all positive x, is not rejected against

H_1 : Shift to right side in the distribution of T.

If the distribution form is known, a parametric test can be applied to the sample observations or a non-parametric test if the distribution is not known, to test the significance of the difference between the two distributions. But this approach appears to be appropriate only for situations with very few genotypes, since the genotypes to be screened will have to be tested in large numbers of plots to examine their distribution functions.

Confidence Interval Method

Test genotypes can be screened by growing them in blocks with a resistant control genotype systematically grown across the field trial. The probability distribution of a resistant control genotype can be calculated based on its seed infection levels across the trial as shown in Figure 2. One can compute the upper 100α percent critical level C_α and define resistance and susceptibility intervals separated by C_α. The quantity C_α can be estimated empirically or by using the timates of parameters of a confirmed distribution to which the sample may belong.

Let the probability distribution of seed infection levels of a resistant control genotype be normal N (μ, σ^2) where the mean μ, and variance can be estimated from sample mean $\bar{x}$, and variance s^2. In this case, a test genotype (T) with a mean $\bar{t}$ (calculated from r replicates) can be considered resistant relative to R if it falls in the resistant region defined by :

$$\bar{t} < C_\alpha$$

where $C_\alpha = \bar{x} + t_{\alpha, e} s/r^{1/2}$

and $t_{\alpha, e}$ is the upper 100α percentage point of t-distribution, with e degrees of freedom.

Any genotype with a mean seed infection level exceeding the critical boundary point C_α will be susceptible.

Data from the 1986 rainy-season trial at ICRISAT Center were subjected to the above

analysis to select genotypes resistant to seed infection by *A. flavus* in relation to the resistant control genotype J 11.

Confidence Interval and Clustering Methods

Various genotypes including a resistant control are usually grown in an experimental design. Each genotype will be observed in r plots (r = number of replications). As r is generally small, one can not accurately obtain the distribution of R (a resistant control genotype). In this situation one can apply the analysis of variance to the observations to estimate error variance and means. Such a data set can be used to select genotypes similar to R in one of the following ways:

Using confidence intervals

A genotype (T) can be regarded as resistant if its mean seed infection level ($\bar{t}$) does not differ significantly from the mean ($\bar{r}$) of R using upper tail t-test as in situation (2), i.e., when

$$\bar{t} - \bar{r} < t_{\alpha,e}\, \hat{SE}\,(\bar{t} - \bar{r})$$

or

$\bar{t}$ lies in the 100 $(1-\alpha)$% one-sided confidence interval $(0, \bar{r} + t_{\alpha,e}\, \hat{SE}(\bar{t}-\bar{r}))$.

Where $t_{\alpha,e}$ is upper 100α% point of the t-distribution with degrees of freedom e used to estimate standard error SE $(\bar{t} - \bar{r})$ of difference $\bar{t} - \bar{r}$, by $\hat{SE}(\bar{t} - \bar{r})$.

In cases where the standard error varies with extreme levels of seed infection, some modification may be required to obtain more precise estimates of error variance associated with the genotypes close to R. This can be done by splitting the genotypes into two groups—one group with genotypes close to R and the standard error can be computed for this group alone, while the other group may contain the rest of the genotypes with a different standard error.

Clustering

The replicate-wise data on genotypes can be used in the form of r-variate information to cluster genotypes based on their similarity in reaction to *A. flavus* infection as assessed in individual replicate plots. The numerical and graphical results in clustering methods can be obtained by using such statistical packages as GENSTAT and SAS. Without these packages, it is also easy to determine the genotypes that fall in a similar cluster with a resistant control at a specif similarity level as follows:

Let Z_{Rj} and Z_{Tj} be the infection levels of the resistant (R) and test (T) genotypes in j-th block. The distance between R and T is

$$D(R,T) = \left(\sum_{j=1}^{r} d\,(R,T,j)\right)^{1/2}$$

where

$$d(R,T,j) = (Z_{Tj} - Z_{Rj})^2.$$

Further, the following modification in computing distance will be required so that a genotype which outperforms the resistant control in the block(s), is not rejected.
Thus, the difference $d(R,T,j) = 0$ if $Z_{Tj} < Z_{Rj}$.

The similarity would then be proportional to the negative of D(R,T). The range of similarity computed for all the genotypes in this manner can be set on a 0 to 100 scale. The cluster of genotypes at a specified level consists of those genotypes for which the similarity percentage is less than, or equal to, that level.

The comparison of the two methods may be rather difficult, as there is no obvious link between the confidence coefficient ($1-\alpha$) of the confidence interval method and the percentage similarity level in the clustering technique. The confidence interval method is very sensitive to the estimator of experimental errors. This method has been grouping more genotypes (with higher susceptibility level, in some of our examples) at $\alpha = 0.05$ compared to the cluster method at 95% similarity level (see Table 1). Furthermore, the clustering method is able to pick up differences between test and control infection levels within each block, and hence may reject genotypes for susceptibility more often than does the confidence interval method where these differences (between R and T) across a block may average very close to zero. This feature would appear to be more useful when resistance screening is done across diverse environments, because the genotype × environment (g × e) interaction is taken into account by the clustering technique.

The confidence interval method and clustering technique can be illustrated using data on the percentage of seed infected by *A. flavus* from the following at ICRISAT Center.

Table 1. Confidence interval and clustering methods for groundnut genotypes similar to J 11, ICRISAT Center, rainy season 1985.

Clustering method:					
Similarity (%)	No. of genotypes	Cluster unit J 11[2]	Mean	SD	Range
> 94	6	GNP104, ICG 3700, ICG 4106 ICG 3660, ICG 2359, ICG 1326	0.5	0.18	0.33-0.67
89-93	10	ICGS(E) 119, ICG7 101[1], ICG 8666, ICG 7633	0.73	0.35	0.33-1.33
< 89	61	Many[1]	2.10	1.12	0.33-4.33
Confidence interval:					
Confidence efficient	No. of genotypes	Resistant group of J 11	Mean	SD	Range
95 (at $P = 0.05$ one sided)	27	ICG 1323, ICG 1436, ICG 1720, ICG 1811, ICG 2359, ICG 3241, ICG 3251, ICG 3478, ICG 3499, ICG 3660, ICG 4106, ICG 6321, ICG 1684, ICG 3700, ICG 4749, ICG 7633, ICG 4502, ICG 4681, ICG 7101, ICG 3263, ICG 7886, ICG 8631, ICG 8666, ICG 8991, GNP 104, GNP 1020, ICG S(E)-119	1.0	0.395	0.33-1.67

1. In addition to genotypes of above group.
2. Mean (J 11) = 0.67±0.39

1986 rainy-season trial

At ICRISAT Center 196 genotypes were grown in a triple lattice design with J 11 as a standard resistant control and JL 24 as a susceptible control genotype systematically sown after every 7th test entry, and appearing thrice in each block. In order to explain the confidence interval method, we estimated the following parameters on the distribution of infection level in J 11.
mean ($\hat{\mu}$) = 1.71, standard deviation ($\hat{\sigma}$) = 0.99
Coefficient of skewness ($\hat{\beta}_1$) = 0.45 ± 0.22
Coefficient of kurtosis ($\hat{\beta}_2$–3) = 0.19 ± 0.43

In view of the low values of $\hat{\beta}_1$ and $\hat{\beta}_2$, it is reasonable to represent the distribution of J 11 as a normal distribution. At α = 0.05 (or 95% confidence coefficient), $t_{\alpha,e}$ = 1.645 (e is large) and confidence interval is {0, 2.65}.
The following genotypes fall in this interval; ICG 1910, ICG 9820, ICG 10021, ICG 10927, and ICG 10147
with mean = 1.73, standard deviation = 0.35, and range = 1.33 - 2.33.

1985 rainy-season trial

At ICRISAT Center 100 genotypes were grown in a triple lattice design with J 11, the stanc resistant control as one of the entries. The lattice blocks did not show any better control of variation, so data were handled as if observed from a randomized complete block design. A set of 61 genotypes were found to cluster together around a 90% similarity level (with infection levels varying up to 4.33), while others had quite high levels of susceptiblity. For the analysis of variance to estimate experimental error variances, this group of 61 entries were separately analyzed, and the remaining entries were not included. The previously confidence interval and clustering methods were applied, and the results are presented in Table 1 for a 95% confidence interval, and >90 percent similarity level. We used the GENSTAT package (Lawes Agricultural Trust 1986) for cluster analysis modified as required for the similarity (or distance).

Distribution of *A. flavus* Seed Infection in Spanish and Valencia Genotypes

To study the nature of distributions of *A. flavus* infection levels in both spanish and valencia groundnut genotypes, the Kolmogorov-Smirnov single sample test (Pearson and Hartley 1976) was applied on original and log-transformed values. Mean *A. flavus* infection levels for trials in the 1986 rainy and 1985/86 postrainy seasons were separately used for this analysis. The distributions of genotypes in the two groups were also compared using the Kolmogorov-Smirnov two-sample test. The Kolmogorov-Smirnov single-sample test was also performed on seed infection levels in the two systematic control cultivars J 11, and JL 24 tested in the 1986 rainy season. The plot of the infection levels prompted us to look into the distribution of normal and log normal types.

1986 rainy-season trial

Out of 196 genotypes cited, 98 were valencia and 98 spanish type. The values of some basic statistics on the distribution for both types are presented in Table 2 for original as well as log

Table 2. Parameters of distribution of *Aspergillus flavus* infection levels in 98 valencia and 98 spanish groundnut genotypes, ICRISAT Center, rainy season 1986.

	Valencia		Spanish	
	Original	Log transformed	Original	Log transformed
Mean	10.19	2.17	9.23	2.08
SD	5.45	0.57	5.70	0.54
Skewnes	0.71	-0.45	2.28	0.07
SE	±0.24	±0.24	±0.24	±0.24
Kurtosis	-0.78	0.81	0.72	0.90
SE	±0.48	±0.48	±0.48	±0.48
Dmax	1.99	1.01	1.90	1.02
Prob	0.001	0.26	0.001	0.25

transformed observations. The low coefficient of skewness and kurtosis and high value of bability level (Prob) for Kolmogorov-Simirnov statistics (Dmax) for log transformed data cate that the level of infection is log normally distributed. The mean infection level was ound to be similar in the two groups (see also Fig. 3).

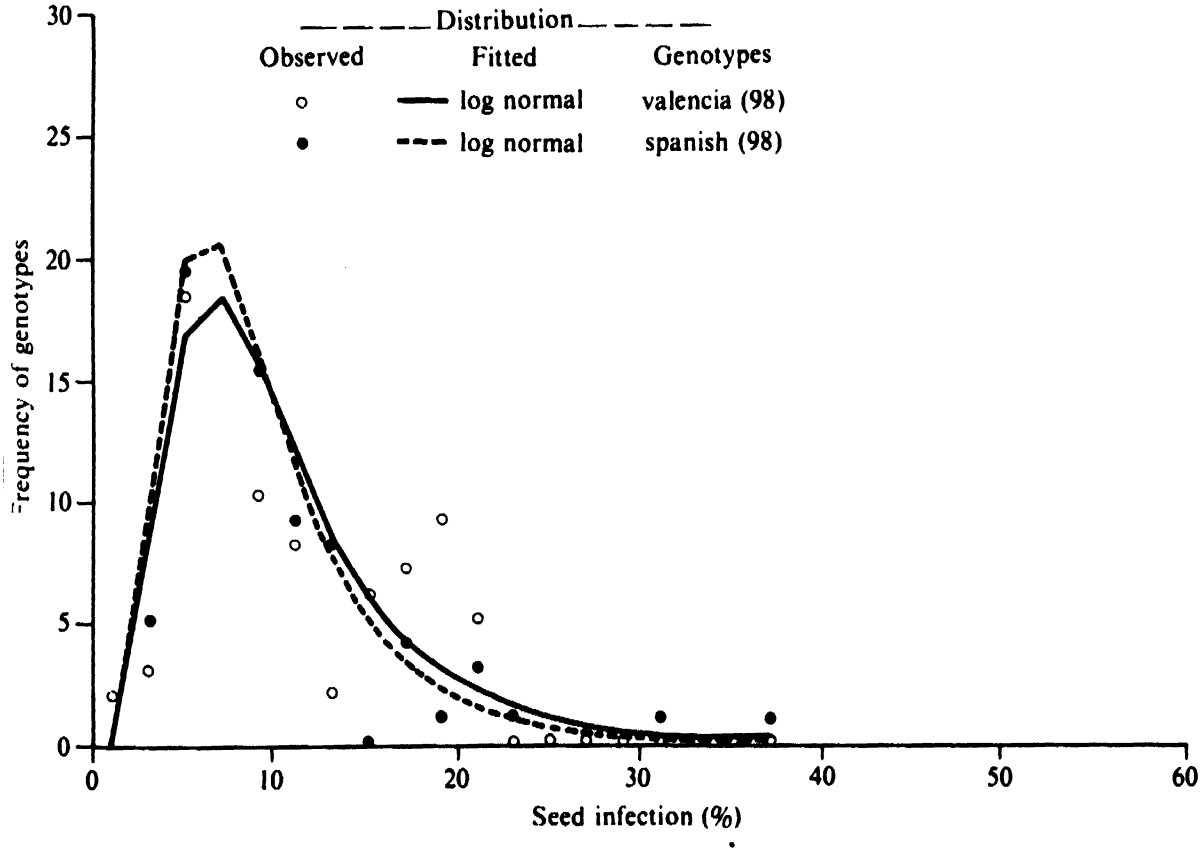

Figure 3. Observed and fitted log normal distributions of seed infection in 98 valencia and 98 spanish groundnut genotypes in Trial 2, ICRISAT Center, rainy season 1986.

Table 3. Parameters of distribution of *Aspergillus flavus* infection levels of 269 spanish and 163 valencia groundnut genotypes, ICRISAT Center, postrainy season trials 1985/86.

	Valencia		Spanish	
	Original	Log transformed	Original	Log transformed
Mean	24.55	3.07	18.42	2.76
SD	11.97	0.56	9.31	0.60
Skewness	0.69	-0.98	0.75	-1.04
SE	±0.19	±0.19	±0.15	±0.15
Kurtosis	-0.17	2.35	0.91	2.13
SE	±0.37	±0.37	±0.30	±0.30
Dmax	1.41	0.85	1.04	1.46
Prob	0.037	0.47	0.23	0.028

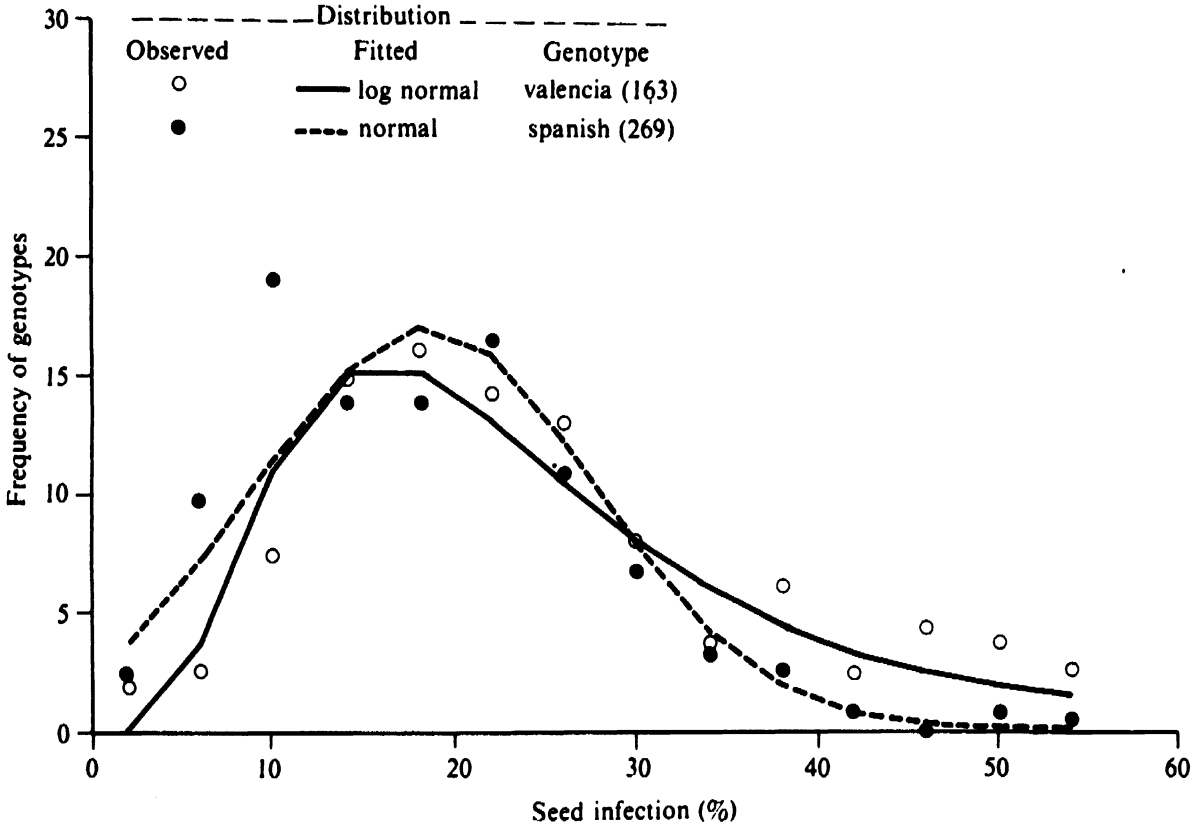

Figure 4. Observed distributions and fitted distributions of seed infection in 163 valencia (log normal distribution) and 269 spanish (normal distribution) groundnut genotypes in Trial 3, ICRISAT Center, postrainy season 1985/86.

1985/86 postrainy-season trial

Three sets each of 144 genotypes were sown in randomized blocks in the same field at ICRISAT Center. Of the 432 genotypes, 163 were of valencia, and 269 of spanish type. The basic statistics for on infection levels for the two groups are presented in Table 3, while the graphical presentation of the distributions (observed and fitted) are shown in Figure 4. Infection levels in the spanish types followed a normal distribution unlike those for the 1986 rainy-season trial and the two botanical groups appear to possess varying potential to provide genotypes with seed resistance to *A. flavus*.

Discussion

The selection of genotypes on the basis of their performance relative to a standard resistant control genotype allows for the flexibility/variability in infection levels that might result from variation in the environmental conditions under field experimentation. The comparison of distribution functions requires large numbers of plots and hence cannot be used to select several entries. While using the confidence method, the experimental error variance requires precise estimation or it may group susceptible entries along with resistant ones. In preparing similarities for cluster methods, one-sided distances should only be considered, since genotypes with infection levels below that of the resistant control are always desirable. Cluster analysis separates susceptible genotypes using differences within blocks while the confidence interval method may not. The application of the clustering method to data from International Cooperative trials would be more sound because genotype × environmental interactions would be successfully reflected in the form of distances (than differences in means if the confidence interval method is applied).

For most situations there appear to be no problems of discontinuous distribution of inoculum and associated levels of seed infection by *A. flavus*. Within trials levels of infection for specific genotypes showed little variation between replicate samples. If a situation occurred where inoculum pressure showed greater variation, it would be worth following the design and analysis approach recommended by Gilliver et al (1985) for sorghum resistance screening against *Striga* since this involves control with cultivars in close juxtaposition to test lines.

It is useful to compare distributions when examining the level of resistance in various botanical groups.

Acknowledgments

Authors thank N. Ramakrishna and V.R. Prabhakar for computational, and C.P. Jaiswal for secretarial assistance.

References

Gilliver, B., Vasudeva Rao, M.J., and Venkateswarlu, P. 1985. A design and methods of analysis to monitor crop growth conditions illustrated with sorghum screening trials for resistance to *Striga*. Experimental Agriculture 21:233-240.

LaPrade, J.C., Bartz, J.A., Norden, A.J., and Demuynk, T.J. 1973. Correlation of peanut seed coat surface wax accumulations with tolerance to colonization by *Aspergillus flavus*. Journal, American Peanut Research and Education Association 5:89-94.

LAT (Lawes Agricultural Trust). GENSTAT (A General Statistical Program). 1986. Release 4.04, Rothamsted Experimental Station, Harpenden, UK: Numerical Algorithms Group Limited.

Mehan, V.K., and McDonald, D. 1980. Screening for resistance to *Aspergillus flavus* invasion and aflatoxin production in groundnuts. ICRISAT Groundnut Improvement Program Occasional Paper no.2. Patancheru, A.P. 502 324, India: International Crops Research Institute for the Semi-Arid Tropics. 15 pp. (Limited distribution.)

Mixon, A.C., and Rogers, K.M. 1973. Peanut accessions resistant to seed infection by *Aspergillus flavus*. Agronomy Journal 65:560-562.

Pearson, E.S., and Hartley, H.O. 1976. Biometrika tables for statisticians. vol.2. University College, London, UK: University College, Biometrika Trust.