National Plant Protection Organization (NPPO) Invasive Pest Safeguarding Using JMP® to Profile Sample Design, Expert Opinion and Pest Movement

Ned Jones, MS, Statistician

Abstract:

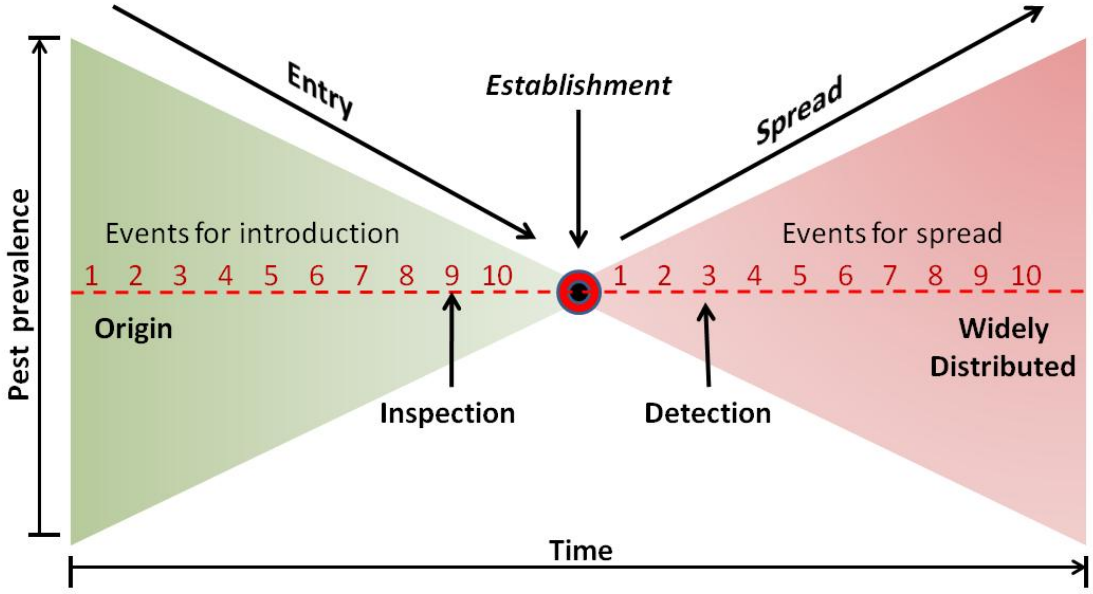
The mission of a National Plant Protection Organization (NPPO) is "to safeguard agriculture and natural resources from the risks associated with the entry, establishment, or spread of animal and plant pests and noxious weeds to ensure an abundant, high-quality, and varied food supply." Inspecting sample commodities for pests before they leave the port accomplishes part of this mission. PPQ uses the JMP Profiler to develop sample designs relying on binomial and hypergeometric distribution, including detection level, detection sensitivity, probability of infestation, confidence and sample size. The Profiler provides a unique view of the interaction of these factors for both single-stage and multiple-stage sampling. Safeguarding also requires the development of exotic pest risk analysis. Often, sufficient data is not available, leaving a reliance on expert opinion to develop these analyses. The Pert distribution helps when quantifying these opinions. Pert is a special case of the beta distribution available in JMP. Using the Profiler, expert opinion inputs are converted to beta parameters while providing visualization of the Pert/beta distribution. Planning programs to eradicate exotic pests that have entered the country requires the ability to predict pest movement. Developing boundaries around new interceptions is critical. JMP 9 provides map-based visualizations of these interceptions. JMP models annual movements by looking at the yearly differences in interception locations. Boundaries can be developed around new interceptions by extending the results to the Profiler and Simulator.

Commodity Sampling for Exotic Pest

The NPPO approves commodities for import on a country commodity basis as part of its mission to safeguard agriculture and natural resources from the risks associated with exotic pests. A commodity approval requires identification of exotic pest associated with the commodity in the country of origin, analysis of the pest and a sound mitigation plan to control the pests as the commodity moves along the path way to the consumer. These plans are referred to as Integrated Pest Management (IPM) plans. Figure 1 provides a conceptual model of the relationship between commodity pathways, events, and pest prevalence(NAPPO 2011). As factors that affect pest entry are known and understood, risk managers may identify either single or multiple risk management options to reduce pest risks associated with the commodity pathway to acceptable levels. An IPM plan includes a range of mitigations such as grower cultural practices, spray programs, harvest culling, packing house culling, washes, brushing, waxing, cold treatment, heat treatment, controlled atmosphere, toxic gas, irradiation and inspection. The preclearance inspection and port of entry inspections include a sample plan. Until recently few of these sample plans had a sound probability basis.

JMP has proven to be a sound tool in the development of these sample plans. Sample plan visualization through the JMP Profiler provides a useful tool to share these plans.

Figure 1: The pathway continuum model which relates change in pest prevalence in a pathway to events (and conditions) along the pathway. This generic model for a pathway begins at the origin where a pest becomes associated with the pathways, proceeds to entry into a new region, establishment, and subsequent spread. A pathway risk analysis can evaluate any set of events along this continuum.



Source: Robert Griffin, USDA-PPQ-CPHST-PERAL

The sample plans used in IPM are based on the presence or absence of pests. The pest presence is usually determined by visual inspection of the commodity in question; however, some pests require PCR, LAMP or Elisa test to determine presence. Presence/absence sampling development usually depends on discrete distributions such as Poisson, binomial or hypergeometric. If the organism presence has been established, estimating the prevalence of rare organism using normal theory can require large sample sizes(Cochran 1977).

Throughout this article we use upper case symbols to refer to variables in the target population and lower case symbols to refer to variables for the sample. In some of the graphs Big N and Big A are used because using N and n or A and a as variable names in the same JMP file causes problems in the JMP profiler. When Big N and Big A are used as variable Big N refers to N and Big A refers to A as used elsewhere in this review. The term sample size refers to the number of sample units. When referring to the size of the sample unit, size of the sample unit will be used.

A consignment of a commodity is presented for import or export. Each consignment is composed of plant units such as fruit, plants, stems, cuttings or other propagules etc. An inspector selects a sample of

plant units following a given protocol based on a well-defined sample design. The sample designs assume clearly defined sample units such that sample units are mutually exclusive.

The sampling design objective as before is to determine that less than P (100%) of the commodity is infested with exotic pests. The primary objective will render a decision whether an unacceptably large portion (fraction, 1-P) of a specified commodity (target population, N) is infested above a specified action level (AL) usually zero infested commodity units or is otherwise defective. It is presumed that suitable actions have been identified to be implemented for either way the decision may go. The action level (AL) in most commodity sampling is 0. Appendix A provides a discussion of the statistical theory used to develop the sample design(Hahn and Meeker 1991).

The binomial, hypergeometric and Poisson distributions are used to develop the sample design in this article. The binomial distribution was used to develop a distribution-free sample design. Then that approach was generalized to the Hypergeometric and Poisson distribution. The binomial distribution represents sampling with replacement while the hypergeometric distribution represents sampling without replacement. The hypergeometric distribution is best used when the sample is >5% of the population. In this range it provides resource saving sample sizes. The binomial's best use is when the sample size is <5% of the population. We can live with the binomial's assumption of sampling with replacement when sample size is <5% of the population. The Poisson distribution is derived from the binomial distribution(Wilks 1966) and provides ease of calculation for small P and very large populations.

Sampling to detect the presence of a pest above a specified level P is based on the binomial. The specified level P can be represented in a binomial distribution. Cochran and others define the binomial distribution as follows(Cochran 1977; Couey and Chew 1986; Hahn and Meeker 1991; Venette, Moon et al. 2002):

$$\Pr(a) = \frac{n!}{a!(n-a)!} P^a Q^{n-a} \qquad (1)$$

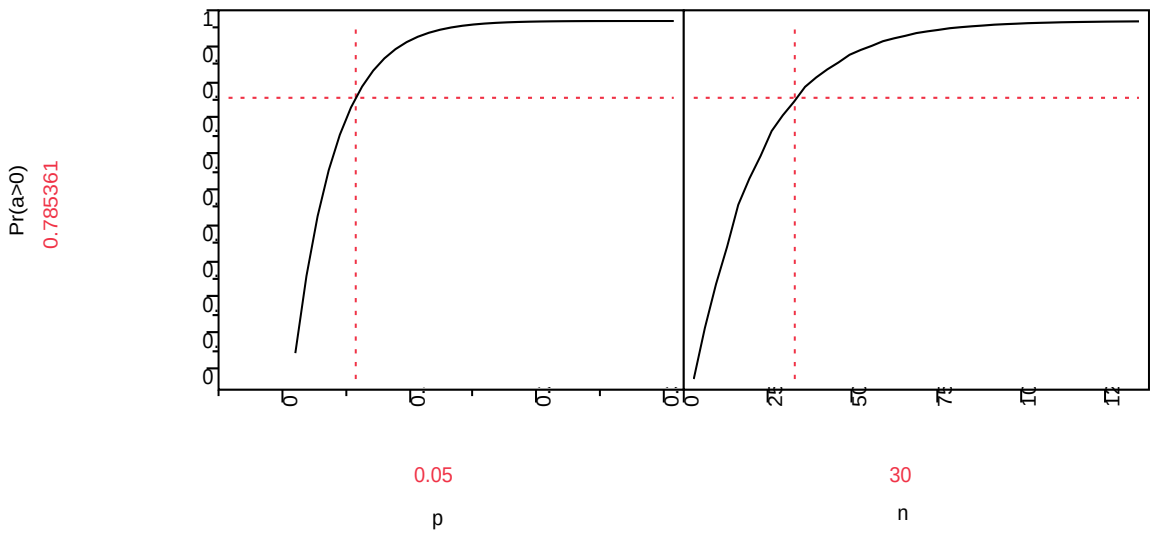
where $\binom{N}{n} = \frac{N!}{n!(N-n)!}$ and $a! = a \cdot (a-1) \cdot (a-2) \cdot \ldots\ldots\ldots \cdot 1$and $0! = 1$.

Where 'n' is the number of commodity units sampled, 'a' is the number of specific occurrences observed in the sample (in our case 'a' is the number of infested commodity units), P is minimum the pest infestation we wish to detect in the commodity and Q=(1-P) is the proportion of the commodity we want to insure is free of pests. We are interested in defining Pr(a) for all a>0, which is Pr(a>0) the probability there is a pests in the commodity sample. The easy way to do this is to define Pr(a=0) and apply basic rules of probability to find Pr(a>0) = 1 – Pr(a=0). From the equation (1) above we know $\Pr(a=0) = Q^n = (1-P)^n$; therefore we can define Pr(a>0) or Pr(of finding >P(100%) pest infestation in the sample size n) is:

$$\Pr(a > 0) = 1 - (1 - \mathrm{P})^{\mathrm{n}} \qquad (2)$$

Figure 2 provides a graph of equation (2); however, because of the functional form the relationship of P to Pr(a>0) changes as n changes and the relationship of n to Pr(a>0) changes as P changes.

Figure 2



The graph shows that the Pr(a>0) increases as either P and/or n increase.

Equation (2) provides an estimate of the confidence the sample will find an infestation of the commodity units >P(100%).

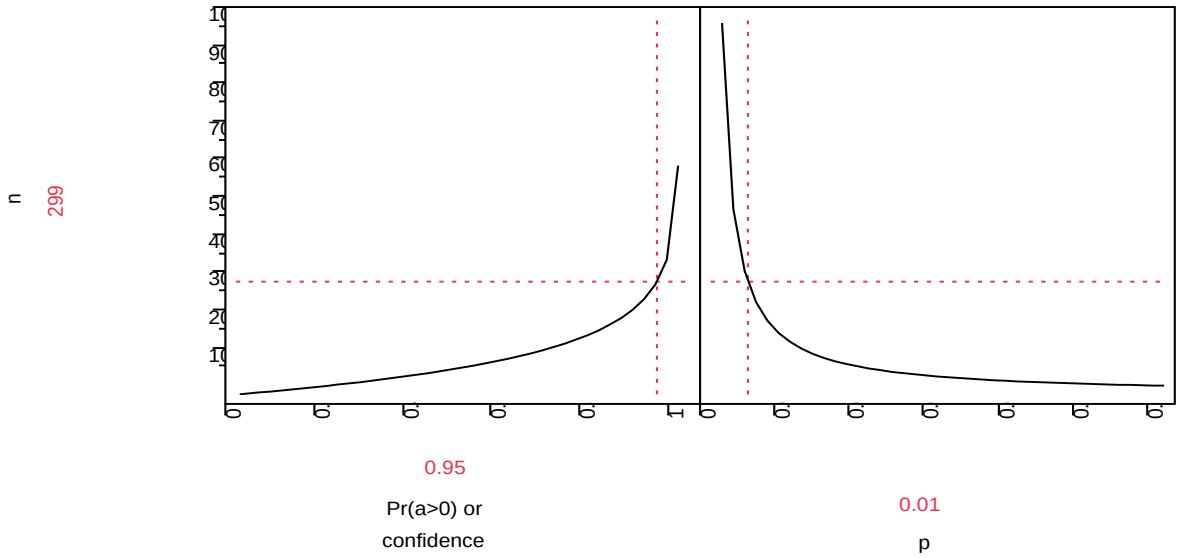
To find the sample size solve equation (2) for n. The result is:

$$n = \frac{\ln(1-\Pr(a>0))}{\ln(1-P)} \qquad \text{for } 0<P<1 \qquad (3)$$

To find for n, we need to know P, the acceptable level of prevalence, and the acceptable level of risk, Pr(a>0) we are willing to take that a>0. The Pr(a>0) is usually set at 0.95 or 95% confidence. Note that the population size, N is not part of the equation when the sample is based on the binomial. The population size is not relevant in the sample design based on the binomial.

Figure 3 provides a graph of equation (3); however, as in equation (2) above because of the functional form the relationship of Pr(a>0) to n changes as P changes and the relationship of P to n changes as Pr(a>0) changes.

Figure 3



The graph shows that n increases as Pa(a>0) increases but n decreases as P increases.

If we need to know P we solve (2) for P. The result is:

$$P = 1 - (1 - \Pr(a > 0))^{1/n} \qquad (4)$$

To solve for P the acceptable level of prevalence, we need to know n the number of units sampled, i.e. the sample size, n and the acceptable level of risk, Pr(a>0) we are willing to take that a>0.

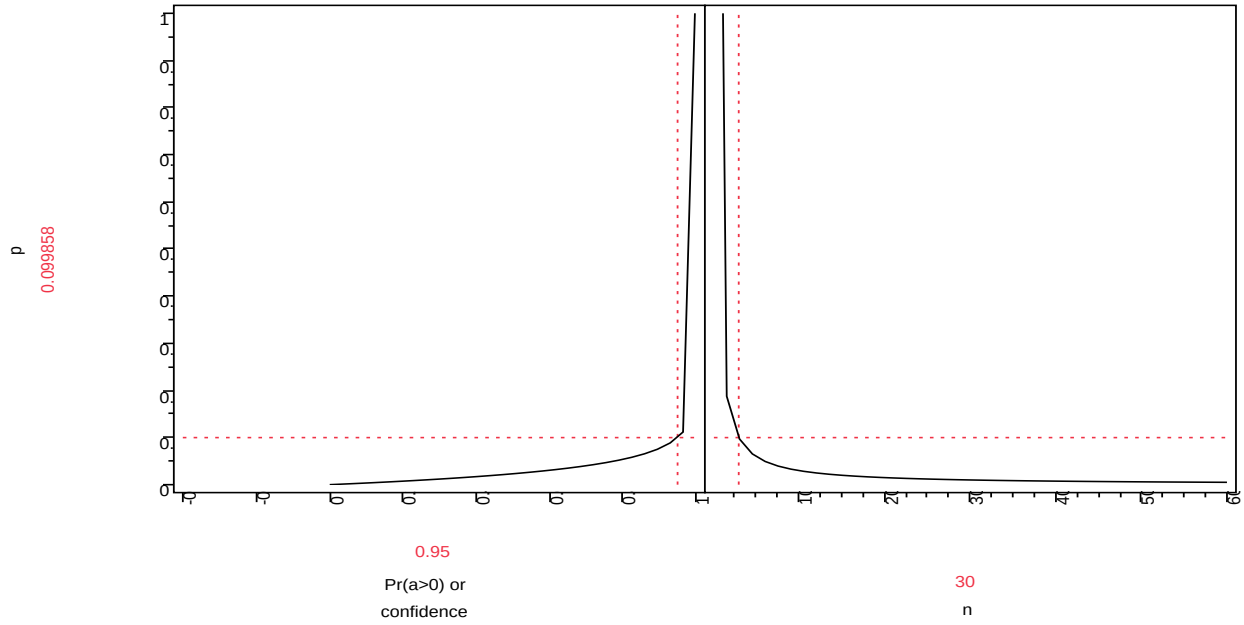
Figure 4 provides a graph of equation (4) ;however, as in equation (2) above because of the functional form the relationship of Pr(a>0) to P changes as n changes and the relationship of n to P changes as Pr(a>0) changes.

The hypergeometric distribution is defined for finite populations. The number of commodity units in the two classes C (infested plant unit) and C'(un-infested plant unit) in the population are A and A', respectively. To calculate the probability corresponding to the numbers a and a' where, $a + a' = n$ and $A + A' = N$ and where N is the number of units in the population and n is the number in sample. Hypergeometric probabilities are conditional probabilities. The probability that a and a' given A and A' is defined as follows(Cochran 1977; Hahn and Meeker 1991):

$$Pr(a, a'|A, A') = \binom{A}{a}\binom{A'}{a'}/\binom{N}{n} \qquad (5)(\text{Cochran 1977})$$

where $\binom{N}{n} = \frac{N!}{n!(N-n)!}$ and $n! = n \cdot (n-1) \cdot (n-2) \cdot \ldots \ldots \ldots \cdot 1$and $0! = 1$.

Figure 4



After setting a to zero and substituting $N - A\ for\ A' and\ n - a\ for\ a'$ the equation can be rewritten as follows:

$$\Pr(a) = \frac{\frac{A!}{a!\,(A-a)!}\frac{(N-A)!}{(n-a)!\,\big(N-A-(n-a)\big)!}}{\frac{N!}{n!\,(N-n)!}}$$

When $a = 0$ the result is as follows:

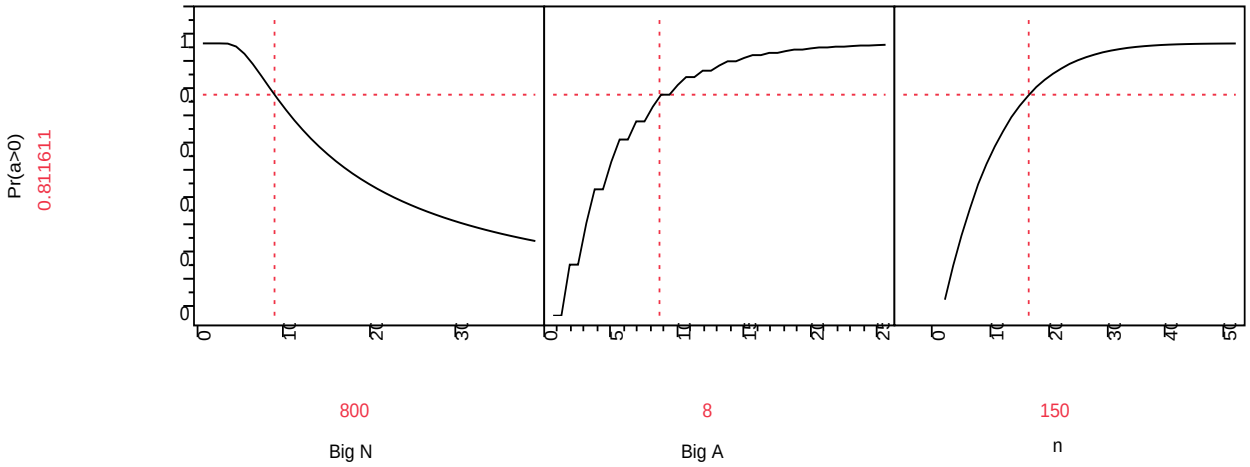
$$\Pr(a=0) = \frac{(N-A)!(N-n)!}{(N-A-n)!N!} = \frac{(N-PN)!(N-n)!}{(N-PN-n)!N!} \quad \text{(5a)} \quad \text{where } PN \text{is adjusted to an integer.}$$

The hypergeometric distribution, equation (5) solves to no convenient form such as the binomial does when a equals zero. Figure 5 provides a graph of the functional relationship presented in equation (5a) when $Pr(a > 0, a'|A.A') = 1 - \Pr(a = 0)\ or\ \Pr(a > 0)$. The JMP hypergeometric function and the JMP Profiler were used to develop the graph in Figure 5 which presents the functional relationship in equation (5a). The graph shows Pr(a>0) decreases as N increases; however, Pr(a<0) increases asymptotically to 1 as either A and/or n increase. As in equation (2) above because of the functional

form the relationship of N to Pr(a>0) changes as n changes this is also true as A changes and the relationship of n to Pr(a>0) changes as N changes this is also true as A changes.

A note of caution, using the hypergeometric distribution with small N can result in lower than desired detection levels, because the relationship of the A (big A) integer input compared to N (big N) results in an implied P much larger than the desired, P, detection level.

Figure 5



JMP provides a hypergeometric function Hypergeometric Distribution(N, K, n, x, <r>). The function returns the probability of the cumulative distribution function at x for the hypergeometric distribution with population size N, K items in the category of interest, sample size n, count of interest x, and optional odds ratio r(SAS_Institute_Inc. 2011). Using the notation from above the functions is coded as follows: $N = big\_N, K = A = P \cdot big\_N, n = n, and\ x = 0$. The coded function appears as

$$Hypergeometric\ Distribution\ (N, ceiling[P \cdot N], n, 0)$$

This function returns the cumulative probability for the function as specified. Note that $P \cdot big\_N$ is the desired detection P multiplied by the population big_N. since this is a discrete distribution JMP automatically rounds down any fraction in the parameters such as $P \cdot big_N$. So if $P \cdot big_N = 7.9$ JMP rounds it down to 7. This is the most conservative approach (JMP_Technical_Support and Archer 2011). This feature can be overridden by using the ceiling function inside the hypergeometric function. Appendix C contains more information on this feature for discrete functions that expect integer inputs.

The ceiling function was used inside the hypergeometric function. This action was chosen because any fraction of a commodity unit infested constitutes an entire unit infested, so any fraction of a unit was rounded up to a commodity unit.

The following estimator for has been used for estimating hypergeometric sample size at the U.S. Nuclear Regulatory Commission(Bennett, Bowen et al. 1988).

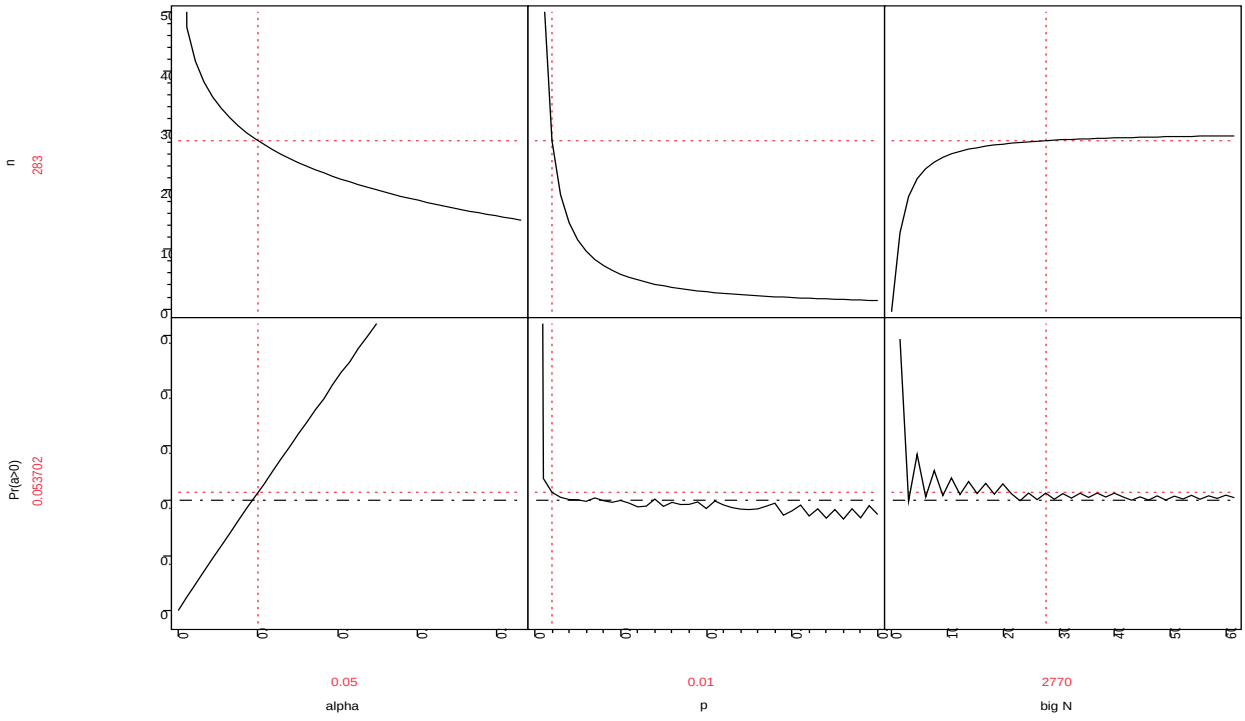
$$n \approx \frac{\left(1-\alpha^{1/V}\right)(2N-V+1)}{2} \qquad (6)$$

Where $V = \max(1, P \cdot N)$

This n and the associated N and A (A=NP) was compared with the hypergeometric distribution function in JMP(SAS_Institute_Inc. 2011). Using the JMP Profiles the graph in figure 6 was produced. The equation (6) n estimate was used in the JMP hypergeometric distribution to estimate Pr(a>0). The graph show that for large population sizes the n estimator provide an n with a close to the desired level for Pr(a>0). As N increases above 4250 the differences from the desired Pr(a>0) decrease, but as N falls below 4250 the actual probability Pr(a>0) varies from the intended Pr(a>0) and the differences in the actual probability Pr(a>0) from the intended probability increase at an increasing rate. Also the actual probabilities for Pr(a>0) tend to be much higher than the intended probability in equation(6). When the intended probability in equation (6) was lowered just a small amount the results improved.

The graph in figure 6 show that as the P the sample detection increases equation (6) provides an n with a more favorable Pr(a>0). The values when P is greater than 0.05 provide a probability Pr(a>0) tend to be lower and be more conservative than the actual Pr(a>0) and those below 0.05 tend to provide less conservative sample size n.

Figure 6



The most reliable method to estimate n for a hypergeometric distribution is by using an iterative approach with a hypergeometric distribution function as in JMP. An iterative function used in a JMP variable function provided results quickly and accurately. A JMP script using an iterative approach in the variable function is provided in Appendix B.

The Poisson distribution can also be used in sample design. The distribution is as follows(Hahn and Meeker 1991; Vose 2008):

$$\Pr(a = a_0) = \frac{e^{-np}(np)^{a_0}}{a_0!} \qquad (7)$$

The Poisson distribution is a very good approximation of the binomial distribution when P is small <0.1 and when n is large. As a result it properties are very similar to the binomial. Since the Poisson is closely related to the binomial the functional relationship matches the binomial. The graphs of the Poisson functional relationship presented in figures 7,8 and 9 are very similar to those of the binomial; however, they are offered here for completeness.

As with the binomial distribution $\Pr(a = 0)$ is easily estimated by the following:

$$\Pr(a = 0) = e^{-np}. \qquad (8)$$

The confidence for a Poisson distribution sample plan when a>0 is as follows:

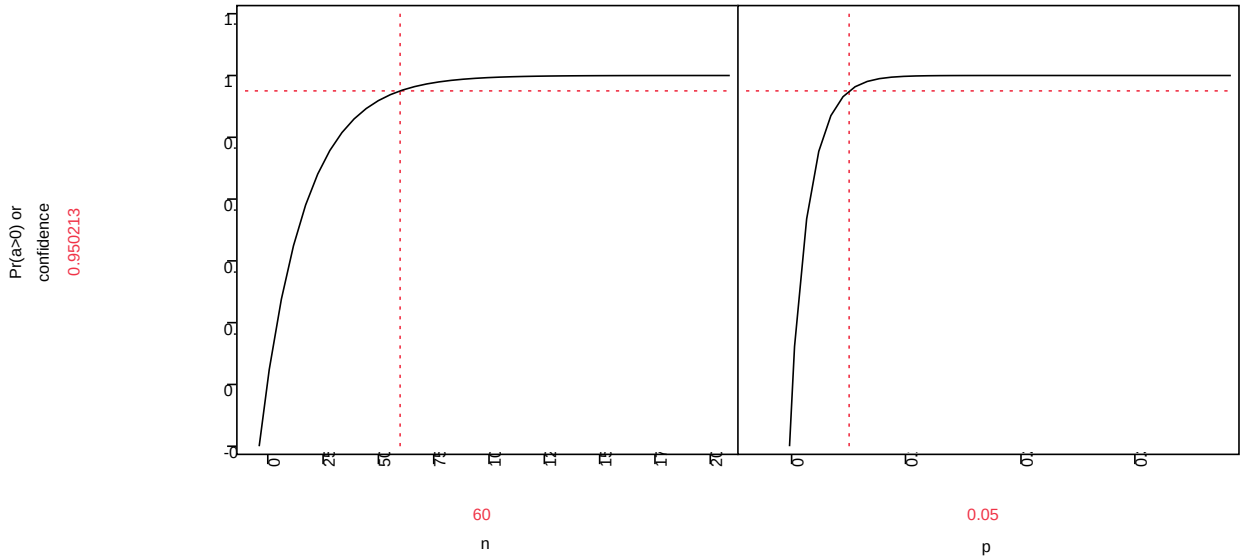
$$\Pr(a > 0) = 1 - e^{-np}. \qquad (9)$$

Figure 7 provides a graph of equation (9); however, as in the binomial equation (2) above because of the functional form the relationship of P to Pr(a>0) changes as n changes and the relationship of n to Pr(a>0) changes as P changes.

The graph shows that the Pr(a>0) increases as either P and/or n increase.

Equation (9) based on the Poisson distribution provides an estimate of the confidence we have that our sample will find an infested unit.

Figure 7



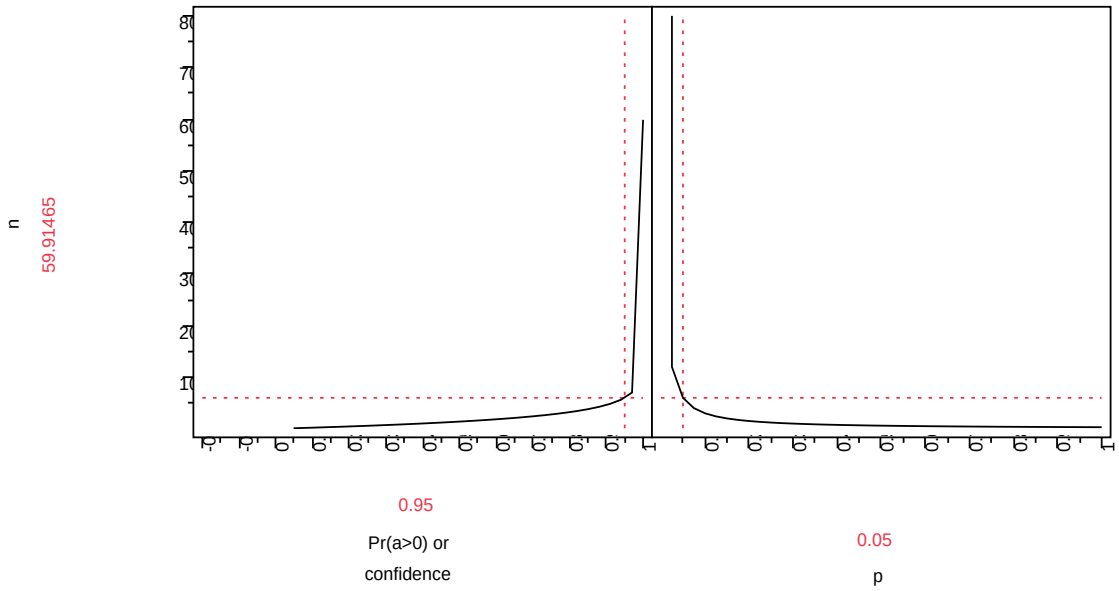
To find the sample size we need for Pr(a>0), given P and Pr(a>0), solve equation (9) for n. The result is a sample n based on the Poisson distribution as follows:

$$n = \frac{\ln(1-\Pr(a>0))}{-P} \qquad \text{for } 0<P<1 \qquad (10)$$

To solve for n, we need to know P, the acceptable level of prevalence, and the acceptable level of risk, Pr(a>0) we are willing to take that a>0. The Pr(a>0) is usually set at 0.95 or 95%. Note that the population size does not come into play when the sample is based on the Poisson.

Figure 8 provides a graph of equation 3); however, as in the binomial equation (2) because of the functional form the relationship of Pr(a>0) to n changes as P changes and the relationship of P to n changes as Pr(a>0) changes.

Figure 8



The graph shows that n increases as Pr(a>0) increases but n decreases as P increases.

If we need to know P we solve equation (9) for P. The result is:

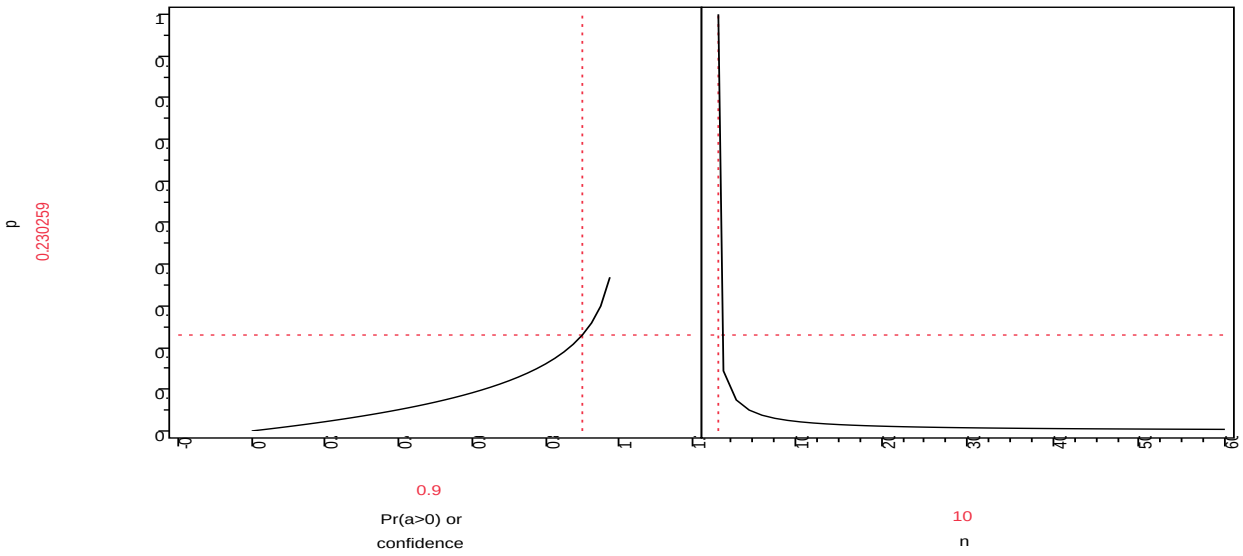
$$P = \frac{\ln(1-\Pr(a>0))}{-n} \qquad (11)$$

To solve for P the acceptable level of prevalence, we need to know n the number of units sampled, i.e. the sample size, and the acceptable level of risk, Pr(a>0) we are willing to take that a>0.

Figure 9 provides a graph of equation (11); however, as in equation (2) above because of the functional form the relationship of Pr(a>0) to P changes as n changes and the relationship of n to P changes as Pr(a>0) changes.

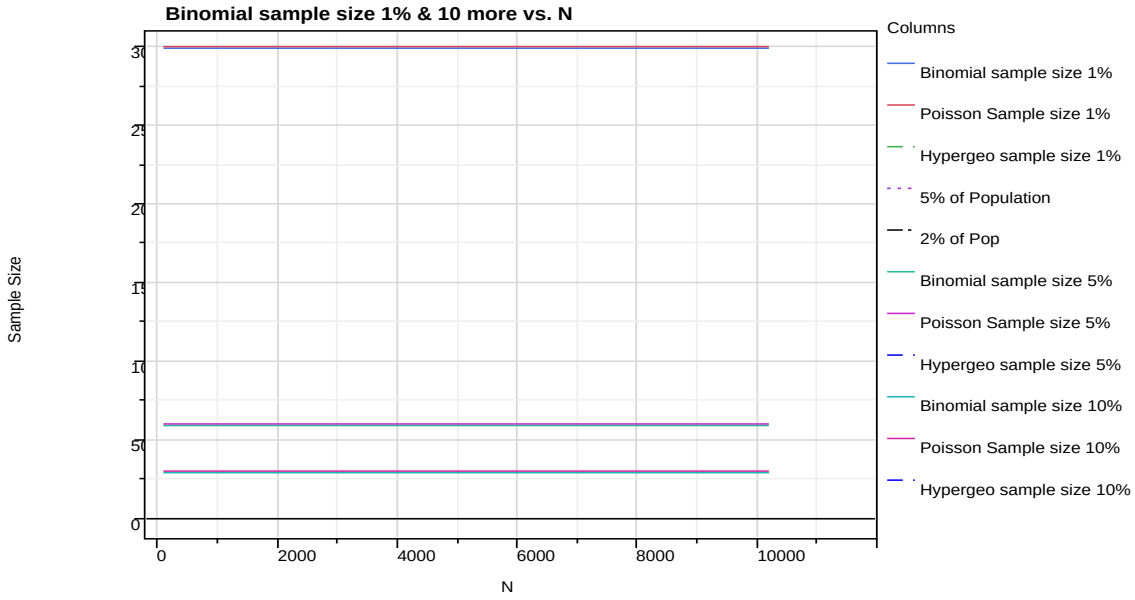
The hypergeometric, binomial or Poisson each play a vital role in developing sample plans in IPM to safeguard against exotic invasive pests. Figure 10 presents the relationship the population size N, the sample size n for the binomial equations (3), the Poisson equations (10) and the results of Hypergeometric sample size result when P the sample detection is set at 0.01, 0.05 and 0.10 each with a 95% confidence. The first relationship we should note is that the at each P detection level the binomial and Poisson sample sizes are unaffected by N while the hypergeometric sample size increases rapidly at first as the population increases but quickly tappers off as the hypergeometric sample size approached the binomial and Poisson sample size level. The 5% of the population line provides a good reference when it intersects the binomial and Poisson, at this point the difference between the hypergeometric and the constant sample sizes binomial and Poisson become negligible.

Figure 9



The 2% line represents the current safeguarding guideline sampling policy for port inspection samples. Its relationship to the probability sample design lines show it is too high when the population is large, too low when the population is small, and rarely just right. The 2% is not consistently applied. Sometimes it is 2% of the commodity, sometimes it's 2% of the boxes in which the commodity is shipped.

Figure 10



The inspection process usually involves the selection and inspection of the commodity. When the hypergeometric or binomial samples are called for in the sampling protocol the inspector selects number of sample units (fruit, plants, etc.) required based on the potential risk of the pests associated with the commodity. These samples are called random but no random selection procedure is followed to direct the selection of the sample. The application of a random sampling process would be difficult logistically and practically. However the commodity population is defined, the fundamental observation is on the commodity unit, which can be infested or clean, and units should be drawn with a probability-based sampling design, because such designs have the desirable property that estimated proportions or means and associated variances are unbiased(Cochran 1977). Unfortunately, the logistically easier and more convenient haphazard method of selecting the sample can yield markedly biased estimates, and if individuals are selected with a systematic procedure, care must be taken that the period of selection does not coincide with an underlying pattern in the population.

The sample selection is frequently a tailgate sample, meaning the commodity or boxes at the tailgate of the truck or shipping container are sampled and inspected. The current sampling is a result of limited resources and limitations on the logistics of performing the sample. Frequently the protocol calls for a random sample but the methodology to perform a random sample is not in place. The application of random sampling results in individual inspectors making an earnest effort using their own judgment to decide what should be inspected. The result is a haphazard sampling approach. However, frequently the inspectors target the inspections toward commodity that show sign of infestation.

A practical alternative to random sampling would be systematic random sampling. Systematic random sampling if applied properly, insures that all of a commodity consignment has a chance to be included in

the sample. The process involves developing a skip interval based on the size of the population N divided by the size of the sample n. Then a random start is selected by multiplying the skip interval by a random number between zero and one. The product of this multiplication should be rounded up to the next whole number if it is not already a whole number. The random number should be provided from a random number table or a computer based random number generator. Then the inspector follows these steps:

1. The inspector counts the commodity units until he reaches the random start. This is the first unit to inspect.

2. Inspect the unit.

3. Then applying the skip interval the inspector counts commodity units until he reaches the skip interval number. This is the next unit to inspect.

4. Inspect the unit.

5. The inspector repeats steps 3 and 4 until n (the sample size) of the commodity units have been inspected.

Following this process the inspector goes through the entire shipment and if he notices a commodity units that should targeted he can inspect that unit also or substitute it for a skip interval unit.

The inspection itself is usually a visual inspection of the commodity. This introduces potential for human error and the possibility of a false negative or false positive identification. Table 1 illustrates the possible outcomes of sampling for rare individuals(Cnnon and Roe 1982; Venette, Moon et al. 2002). Symptoms, in the broad context used here, may also include the results of diagnostic tests. True positives (A) occur when the inspector observes the presence of a pest. False positives (B) occur when pest are present and action is taken, but the pest is not of concern. False negatives result when pests are not apparent, but a pest of concern is present (C). True negatives occur when a pest is neither observed nor is present (D). If perfect correspondence exists between organism and detection, only A and D will result. However, the association between organism symptoms and presence is not perfect. Few if any tests are thought to be perfectly sensitive and specific. Sensitivity (Se) is the probability that a commodity unit with the trait will be judged to be infested. From Table 1, $Se = A/(A + C)$ specificity (Sp) is the probability that an commodity unit without the trait will be judged to be infested $Sp = D/(B + D)$. Where Se<1, true positives will be incorrectly classified as clean, and apparent commodity unit-level infestation prevalence will be an underestimate of the truth. If Sp < 1, true negatives will be incorrectly identified as positive, causing an overestimate of the true prevalence. Errors may be in either direction if both Se and Sp < 1. Where Se and Sp can be characterized with point estimates or empirical distributions, apparent commodity unit-level frequencies can be corrected.

TABLE 1 Relationship between presence of an individual and evidence of its presence (i.e., symptoms) that might be used to guide the search for rare Individuals

	Actual condition	
Results from sampling (symptoms)	Present	Absent
Present	(A) Correct (false positives)	(B) Incorrect
Absent	(C) Incorrect	(D) Correct (false negatives)

Sensitivity of commodity inspection can have a significant impact on sample design to detect pest presence and on estimating pest risk associated with a commodity. Quarantine inspectors search for fruit fly infestations in incoming shipments by visual inspection and by dissecting or cutting a sample of fruit in each shipment. The reliability of the latter procedure for detecting fruit fly larvae is questionable and, therefore, a test was conducted to determine its effectiveness(GOULD 1995). Infested grapefruit, mangoes, guavas, and carambolas were cut open to determine the efficacy of cutting fruit in detecting larval infestations of Caribbean fruit flies, *Anastrepha suspense* (Loew). From 1 to 36% of the larvae were detected by dissection, but 17.9 to 83.5% of the infested fruit were detected. There was considerable variation in the number of larvae found by different inspectors. This is the single study we have on commodity inspection. More work is needed to better the sensitivity of commodity inspection.

Specificity is it not as big an issue in commodity inspection because positives receive so much attention that they are ultimately resolved. When designing a sample to detect pest presence specificity does not play as large role as sensitivity.

The effect of sensitivity on pest detection survey design is as follows:

The probability of an observed inspection positive is $\Pr(A) = \frac{A}{A+B+C+D}$.

The true probability of a positive is $\Pr(A+C) = \frac{A+C}{A+B+C+D}$.

The sensitivity of a positive is $Se = \frac{A}{A+C}$.

The goal of the sample design is for the sample/inspection detection to equal the true probability. When an inspection/sample is implemented we observe A the inspection positives. This represents the observed sample/inspection detection. The sample design must be adjusted by the inspection sensitivity so that our sample/inspection detection measures the true positives. The probability of an observed inspection positives equals the sensitivity, Se multiplied by the desired sample/inspection detection. We present this mathematically as follows:

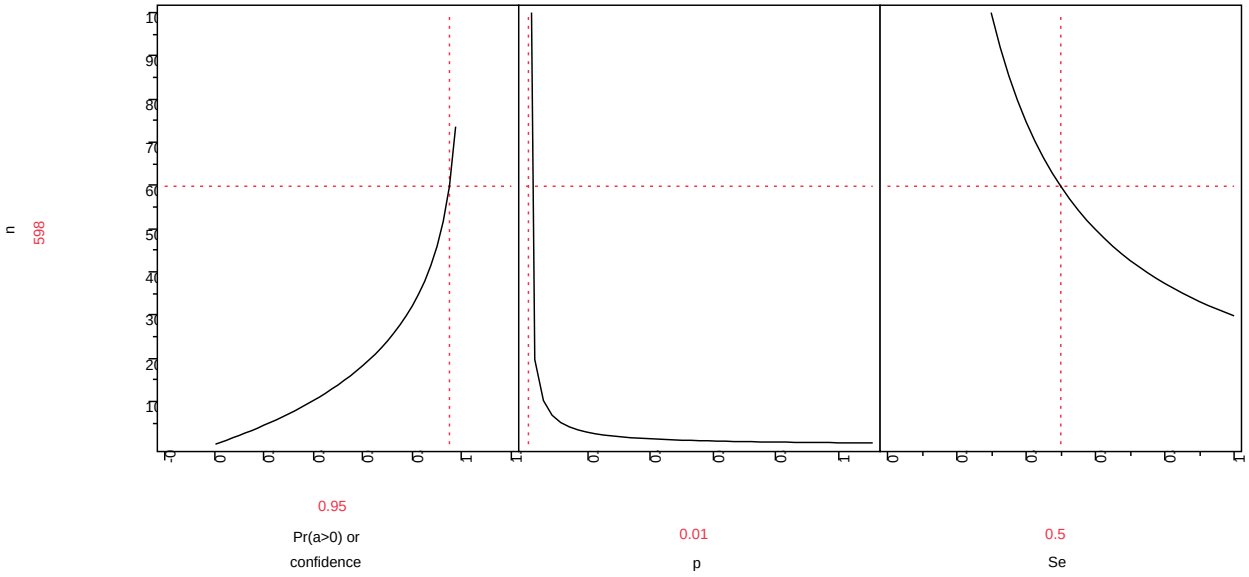
$$Se \cdot \Pr(A+C) = \frac{A}{\cancel{A+C}} \frac{\cancel{A+C}}{A+B+C+D} = \frac{A}{A+B+C+D} = \Pr(A)$$

When the Se adjustment is made our sample size change is inverse proportional to the sensitivity. The binomial sample size formula adjusted for sensitivity is equation (3) with P the sample detection replaced by $Se \cdot P$ as follows:

$$n = \frac{\ln(1-\Pr(a>0))}{\ln(1-(Se \cdot P))} \qquad \text{for } 0<P<1 \qquad (12)$$

Figure 11 provides a graph of equation (12); however, as in equation (2) above because of the functional form the relationship of Pr(a>0) to n changes as P or Se change and the relationship of P or Se to n changes as Pr(a>0) changes.

Figure 11



The Poisson formula for n adjusted for sensitivity is as follows:

$$n = \frac{\ln(1-\Pr(a>0))}{-Se \cdot P} \qquad \text{for } 0<P<1 \qquad (13)$$

The hypergeometric formula for the n approximation adjusted for sensitivity is as follows:

$$n \approx \frac{\left(1-\alpha^{1/V}\right)(2N-V+1)}{2} \qquad (14)$$

Where $V = \max(1, Se \cdot P \cdot N)$

Using the notation from above, the functions is coded as follows: $N = big_N, K = A = Se \cdot P \cdot big_N, n = n, and\ x = 0$. The coded function appears as

$$Hypergeometric\ Distribution\ (N, ceiling[Se \cdot P \cdot N], n, 0)$$

which returns the cumulative probability for the function as specified. Note that $P \cdot big\_N$ is the desired probability of detection P multiplied by the population big_N. Since this is a discrete distribution function JMP automatically rounds down any fraction in the parameters such as $Se \cdot P \cdot big_N$. So if $Se \cdot P \cdot big_N = 7.9$ JMP rounds it down to 7. This is be the most conservative approach(JMP_Technical_Support and Archer 2011). This feature can be overridden by using the ceiling function inside the hypergeometric function. Appendix C contains more information on this feature for discrete functions that expect integer inputs.

Table 2 shows the sample sizes for detection level of 1% 5% and 10% with 95% confidence for these distributions binomial, Poisson, and hypergeometric for populations of 100, 300 600 and 1000. When the detection is higher the hypergeometric converges to the binomial and Poisson more quickly than it does when detection is lower. This effect is slightly reduced as sensitivity decreases.

Table 2 Shows Samples based on detection of 1%, 5%, and 10% with 95% of confidence
For the Binomial, Poisson, and Hypergeometric for Population sizes 100, 300, 600 and 1,000

		Sample Size					
				Pop 100	Pop 300	Pop 600	Pop 1,000
Detection	Sensitivity	Binomial	Poisson	Hypergeo	Hypergeo	Hypergeo	Hypergeo
0.01	1.00	299	300	95	189	236	258
0.05	1.00	59	60	45	54	56	57
0.10	1.00	29	30	25	28	29	29
0.01	0.67	448	450	95	189	270	348
0.05	0.67	89	90	52	71	79	83
0.10	0.67	44	45	34	39	41	43
0.01	0.50	598	600	95	233	379	450
0.05	0.50	119	120	63	93	108	112
0.10	0.50	59	60	45	54	56	57

The sample sizes in Table 2 are for direct sampling of the commodity units such as plants, cuttings, stems, fruit, etc. Most commodities are shipped in some sort of packaging with a specified average number of commodity units in each package.

NPPO, Plant Inspection Stations see plant parts presented for inspection on a continuous basis. The plant parts units are usually packed in boxes, bags or bags in boxes, bundles, wrapped in newspaper, etc. The objective of the inspection is to insure the consignment is free of exotic pest per plant unit below a specific P(100%). The country plant genera combinations are assigned pest risk of levels (high, medium and low). The risks levels (high, medium and low) are assigned detection level P values (1%, 5%, and 10% respectively). For simplicity of discussion we will use a box as the packaging unit and propagules as the plant unit within the box. When a box is selected for sampling the inspector first

empties the entire box contents on to a white paper to see if pests appear on the paper. Next they inspect the empty box to see if any pests were left behind. Then the propagules are inspected for pests. This presents a classic cluster sampling situation. The population is composed of N boxes and M propagules in each box. The total number of propagules in the consignment is N boxes multiplied by M propagules per box. Inspecting the propagules one by one or selecting a sample of propagules is just not operationally practical or feasible. How many boxes of the N with M propagules per box should be selected to insure with 95% confidence $\leq P(100\%)$ of the propagules are pest free. Based on historic inspection results we assume that the pests would be evenly distributed over the propagules. If the pests are clustered, the sampling plan would need to be modified to take this clustering into consideration. Based on expert opinion we expect inspectors to detect an infested box at least 50% of the time(Schuler 2010).

A consignment invoice at the Plant Inspection Station usually provides the number of boxes and the total number of propagules for each plant genus species and country of origin. From this information we can assign risk (low, medium or high) and P, detection values (10%, 5% or 1%, respectively). The number of propagules per box, M, can be calculated. Given the assigned P we can calculate the probability of one or more pests in the box, $\Pr(a_M > 0)$, using equation (2) substituting M for n. The result is as follows:

$$Pr(a_M > 0) = 1 - (1 - P)^M \quad (15)$$

This gives the probability of an infested box $Pr(a_M > 0)$ base on P and M the number of propagules in the box. The binomial distribution was used because it provides and exact result. The hypergeometric was not used because as noted above when N (big N) is small the relationship of A (big A) integer input compared to N (big N) implies a P much larger than the desired, P, detection level.

To calculate the number of boxes to sample the sample sizes assuming a binomial, Poisson or hypergeometric distributions of pests over the boxes will be evaluated.

The binomial sample size can be found by using equation (12) above with the results in equation (15) $1 - (1 - P)^M$ substituted for P as follows:

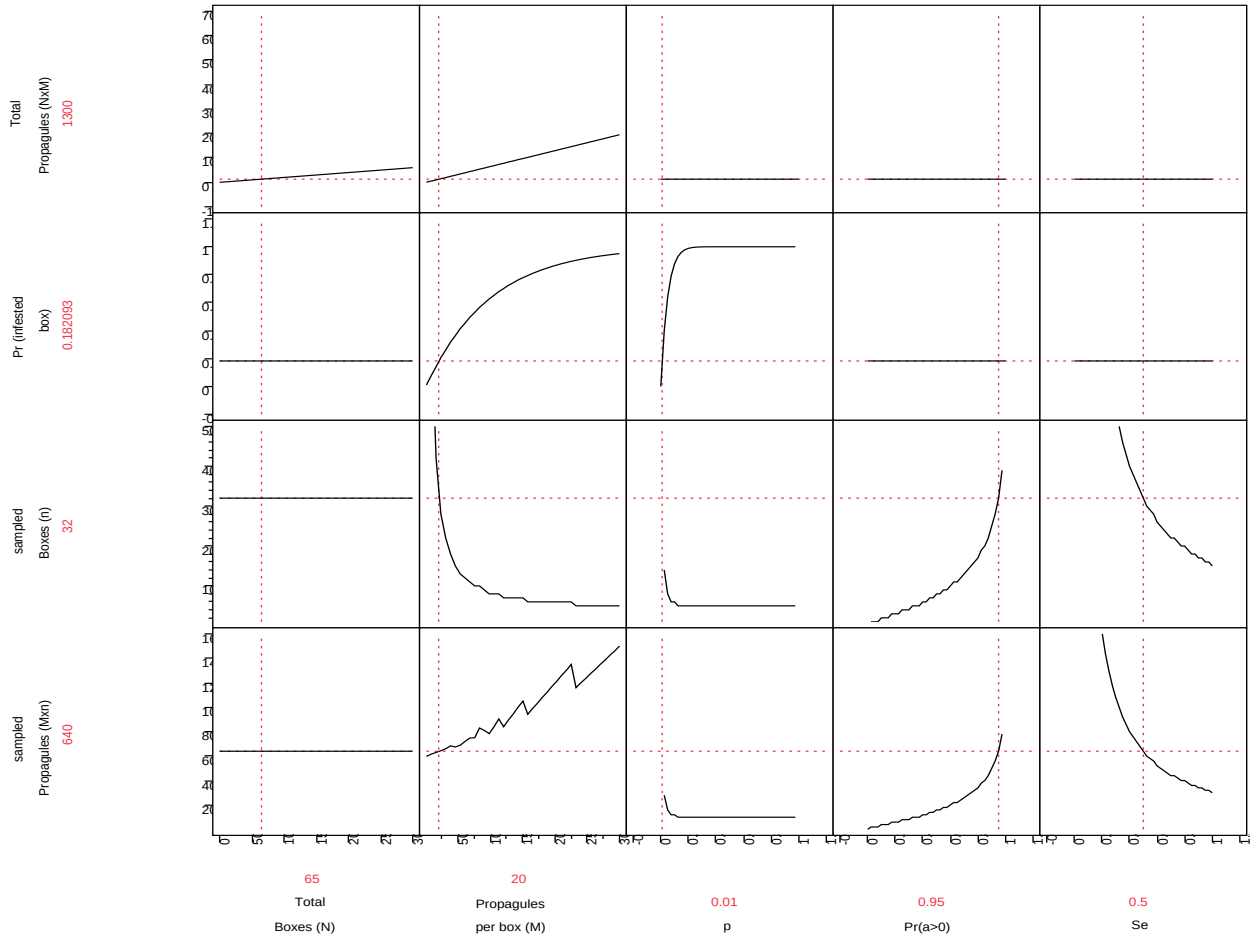
$$n = \frac{\ln(1-\Pr(a>0))}{\ln\left(1-(Se\cdot(1-(1-p)^M))\right)} \quad \text{for } 0<P<1 \quad (16)$$

Figure 12 shows a graph of the functional relationship for sample size presented in equation (16). The Total Boxes (N) only has a relationship with Total Propagules (NxM). The Propagules per Box (M) has a positive relationship with Pr(infested box), Total Propagules and Propagules sampled; however, the Propagules sampled not smooth because of the discrete binomial relationship. The Propagules per Box (M) has a positive asymptotic to 1 relationship with Pr(infested box). Also P the desired detection has a positive asymptotic to 1 relationship with Pr(infested box). As P increases Sampled Boxes (n) and Samples Propagules (nxM) decrease to some minimum number of boxes sampled which is dependent on

confidence Pr(a>0) and sensitivity. As confidence increases the sample Boxes (n) and sample Propagules (nxM) increase. As sensitivity increases the sample Boxes (n) and sample Propagules (nxM) decrease.

The number of Propagules per box (M) has an interactive relationship with all the other variables. The desired detection P interacts with every variable except Total Propagules (NxM). Both confidence Pr(a>0) and sensitivity, Se interact with Propagules per box and the desired detection, P.

Figure 12



If the Poisson distribution to the pest spread over the boxes the sample size can be found by using equation (13) above with the results in equation (15) $1-(1-P)^M$ substituted for P as follows:

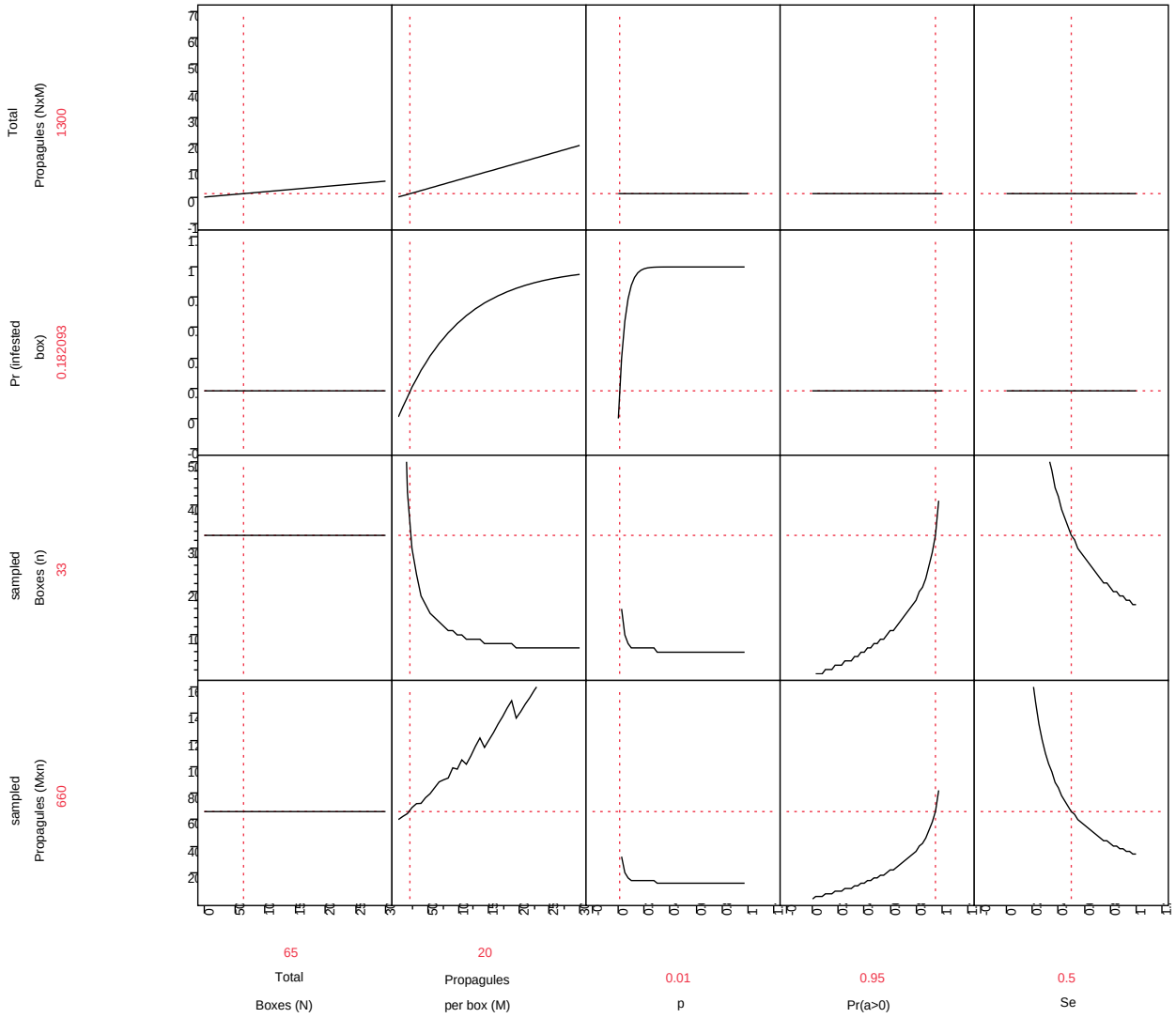
$$n = \frac{\ln(1-\Pr(a>0))}{-Se\cdot(1-(1-P)^M)} \qquad \text{for } 0<P<1 \qquad (17)$$

Since the Poisson is directly related to the binomial the relationships observed are similar to the binomial. They are presented here for completeness. Figure 14 shows a graph of the functional relationship for sample size presented in equation (17). The Total Boxes (N) only has a relationship with Total Propagules (NxM). The Propagules per Box (M) has a positive relationship with Pr(infested box),

Total Propagules and Propagules sampled; however, the Propagules sampled are not smooth because of the discrete binomial relationship. The Propagules per Box (M) has a positive asymptotic to 1 relationship with Pr(infested box). Also P the desired detection has a positive asymptotic to 1 relationship with Pr(infested box). As P increases Sampled Boxes (n) and Samples Propagules (nxM) decrease to some minimum number of boxes sampled which is dependent on confidence Pr(a>0) and sensitivity. As confidence increases the sample Boxes (n) and sample Propagules (nxM) increase. As sensitivity increases the sample Boxes (n) and sample Propagules (nxM) decrease.

The number of Propagules per box (M) has an interactive relationship with all the other variables. The desired detection P interacts with every variable except Total Propagules (NxM). Both confidence Pr(a>0) and sensitivity, Se interact with Propagules per box and the desired detection, P.

Figure 14



Using this combination the Poisson for the distribution of pest over the boxes may not be the best choice unless there are many boxes and the detection level P is very low.

The hypergeometric approximation of sample size can be found by using equation (14) above with the results in equation (15) $1-(1-P)^M$ substituted for P as follows:

$$n \approx \frac{\left(1-\alpha^{1/V}\right)(2N-V+1)}{2} \qquad (18)$$

where $V = \max(1, Se \cdot (1-(1-P)^M \cdot N)$.

Caution must be used when applying equation (18) as demonstrated above. Frequently the assumed $\propto = \Pr(a=0)$ is smaller than the actual sample $\Pr(a=0)$.

Or when hypergeometric iterative approach is used the JMP hypergeometric function is coded as follows:

$$Hypergeometric\ Distribution\ (N, ceiling[Se \cdot 1-(1-P)^M \cdot N], n, 0)$$

As in the examples of this function above the ceiling function was used inside the hypergeometric function. This action was chosen because any fraction of a box infested constitutes an entire box infested, so any fraction of a box was rounded up to a whole box.

The $\Pr(a>0)$ can be presented in the JMP Profiler using hypergeometric distribution function to represent the spread of pests over the boxes. Figure 15 displays the graph of this relationship. The graph presents a similar relationship among the variables sampled Propagules, Pr(infested box) and Total Propagules as seen for the binomial and Poisson distributions. The last variable $\Pr(\mathrm{a}>0)$ is presented as a function of the other variables because of the difficulty of solving equation (5a) for sample size n. Figure 16 provides a comparison of the $\Pr(\mathrm{a}>0)$ for both the hypergeometric and the binomial distribution. The general shape of the relationships is the same; however the smaller sample size produced by the hypergeometric is a result of the population of boxes, N being included in the functional relationship. The effect of N in the hypergeometric $\Pr(\mathrm{a}>0)$ relationship in the first two graphs where produces a small decrease in the hypergeometric as N increases while the binomial $\Pr(\mathrm{a}>0)$ relationship is flat. This difference is reflected in each of the other inputs p, the detection level, Propagules per Box (M), sensitivity, Se and Samples Boxes (n). Also the steps in the hypergeometric $\Pr(\mathrm{a}>0)$ functions with Propagules per Box (M), sensitivity, Se are much more pronounced than with the binomial $\Pr(\mathrm{a}>0)$. And as expected the Hypergeometric produces a smaller sample size than the binomial.

Figure 15

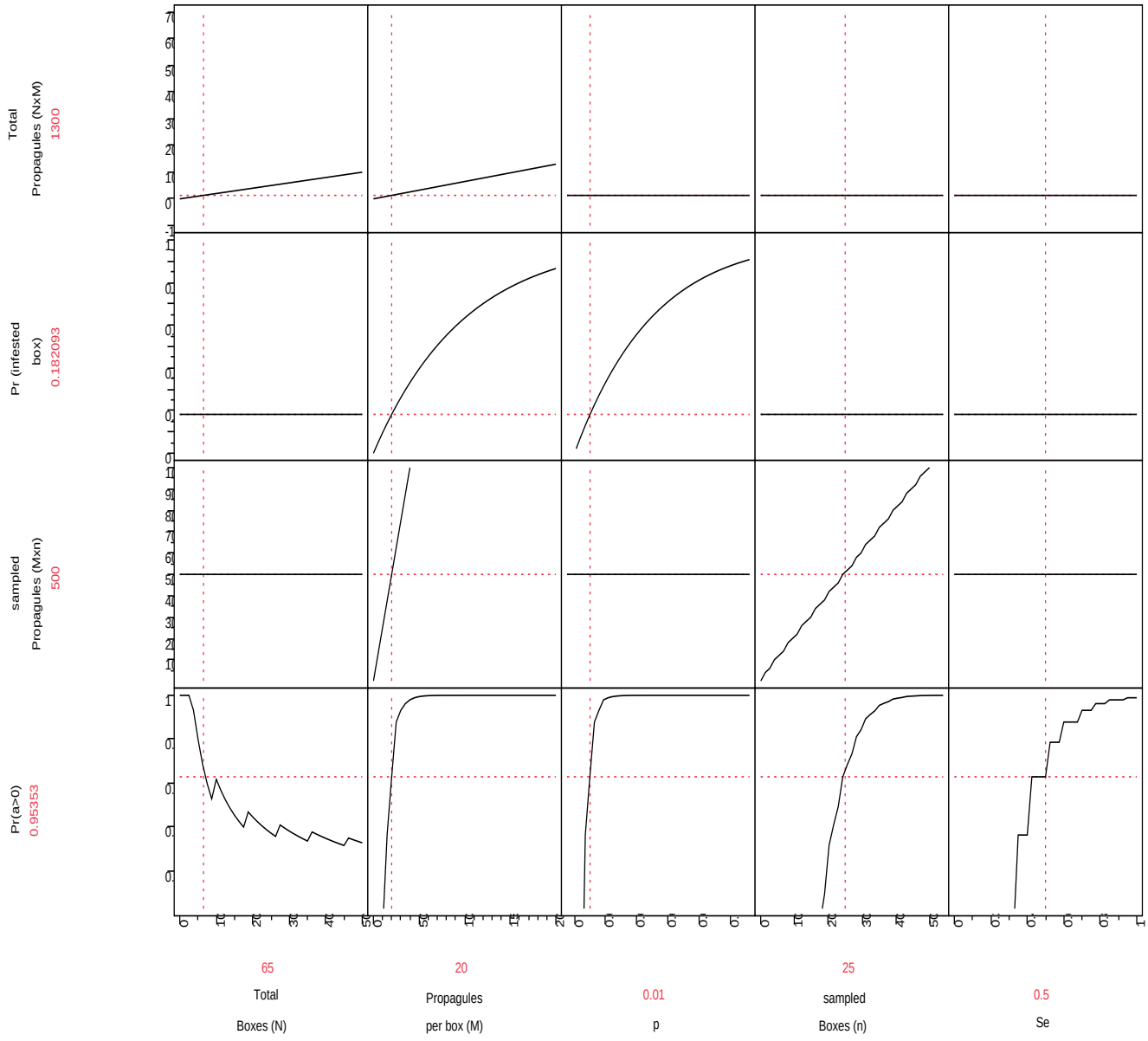
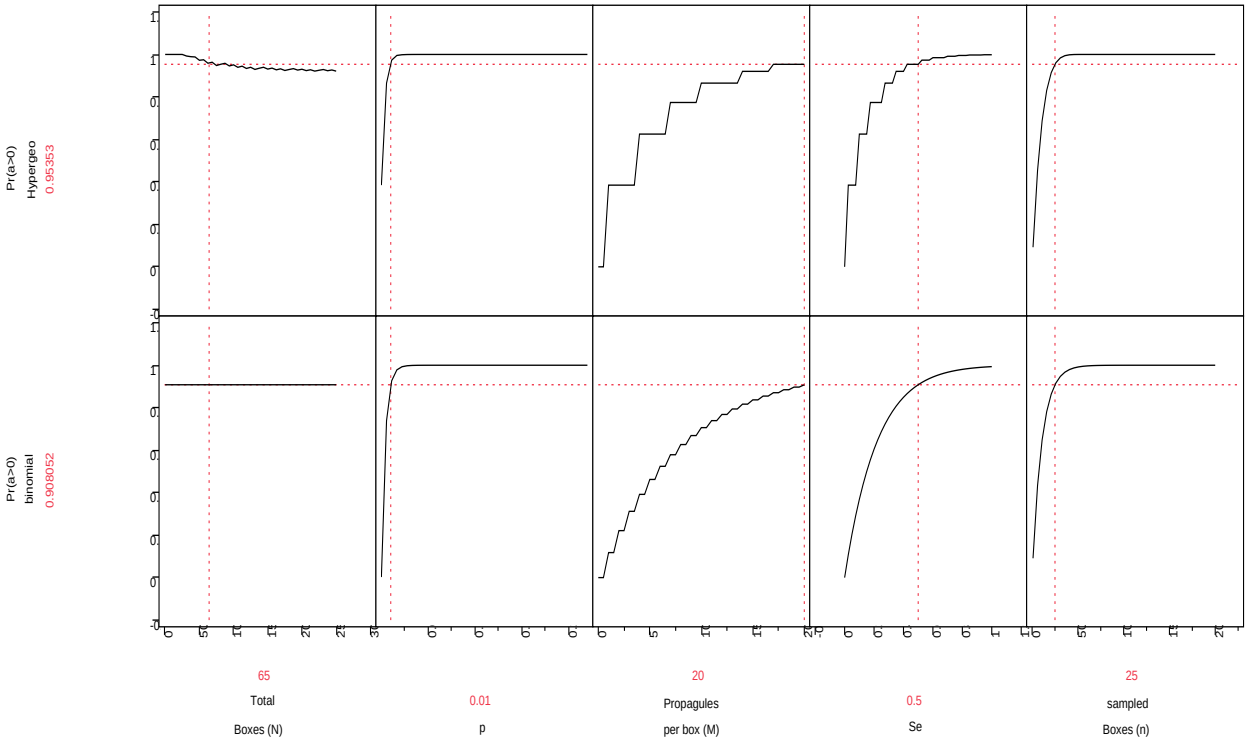

Total Propagules (NxM)
1300
Pr (infested box)
0.182093
sampled Propagules (Mxn)
500
Pr(a>0)
0.95353
65
Total Boxes (N)
20
Propagules per box (M)
0.01
p
25
sampled Boxes (n)
0.5
Se

Figure 16



After considering the alternatives the NPPO Plant Inspection Station management wanted to use the hypergeometric distribution when selecting boxes to inspect, because it produces the smallest sample size which provides the most efficient use of their scarce personnel resources. The Plant Inspection Stations sample design uses the binomial distribution to estimate the probability of an infested box and then uses this probability to inform the hypergeometric distribution of boxes to estimate A the number of infested boxes in the population of boxes. This sample uses a cluster sample approach with boxes selected as clusters and the entire cluster (box) inspected. No subsampling of the propagules in the box, so no two stage sampling was used. The sampling assumes that the pests are evenly distributed over the boxes and there is little or no clustering of pests.

Two stage sampling provides the next logical step in safeguarding Agriculture and Natural Resources from invasive exotic pests. Most commodities shipped in boxes, bags etc. qualify for two stage sampling. Two stage sampling requires identifying and defining the population primary sampling units, boxes, bags, pallets etc. and population of secondary sampling units, fruit, flowers stems, tile etc. Remember the primary units must be defined so that they are mutually exclusive from each other and the secondary units must be defined so that they are mutually exclusive. Most commodity consignments presented for inspection can be defined to meets these requirements. In addition to being mutually exclusive the primary units should be the same size, or nearly so(Cochran 1977). And the secondary sample units should be uniform of size or at least as uniform as possible. Sampling proceeds as follows; first select a sample from the primary units, boxes; next select a subsample from the secondary units contained within the selected primary sample units. In detection sampling deciding how to characterize

the sampling distribution of primary and secondary sample units provides a matter of some concern. The primary sample can be base on either a binomial or a hypergeometric distribution and the same consideration need to be made for the secondary sample. The sample design can consider the primary sample distribution and the secondary sample distribution to be binomial and binomial respectively, or binomial and hypergeometric, hypergeometric and hypergeometric or hypergeometric and binomial.

The sampling design objective as before is to determine that less than $P\ (100\%)$ of the commodity is infested with exotic pests. We will assume that the processing of the commodity resulted in sufficient shuffling of the commodity so that the pests are evenly distributed through the commodity, i.e. pests clusterings are not an issue in the sample design. The sample design considers the following:

1. The desired detection, P, specifies the initial condition per secondary sampling unit.

2. The secondary stage probability of detection is developed using the assumed sample distribution. The result is $\Pr(\ infestation\ \ sec.\ \ sample)$ the probability of infestation in the secondary sample.

3. The $\Pr(\ infestation\ \ sec.\ \ sample)$ provide the initial condition per primary sampling unit.

4. The two-stage probability of detection is developed using the assumed primary sample distribution. The result is $\Pr(\ a > 0)$ the probability of infestation for the commodity.

5. Then a specific sample size, n, can be developed given the following information:

    1. $P\ =\ the\ tolerance\ limit\ for\ commodity\ infestation, P\ (100\%)$

    2. $N\ =\ the\ number\ of\ primary\ sample\ units, boxes,$

    3. $M\ =\ the\ number\ of\ secondary\ units\ per\ primary\ unit, fruit,$

    4. $Pr(a > 0) = the\ desired\ confidence\ Pr(a > 0) \cdot 100\%$, and

    5. $Se\ =\ an\ estimate\ of\ the\ sensitivity$, ranging from 17 to 83%(GOULD 1995) .

For the remainder of the two-stage sampling discussion boxes will be used as the primary sample unit and fruit will be the secondary sample unit.

If we assume a binomial distribution for the boxes and a binomial distribution for the fruit, equation (2) and equation (12) solved for $Pr(a > 0)$ are combined to provide the following result:

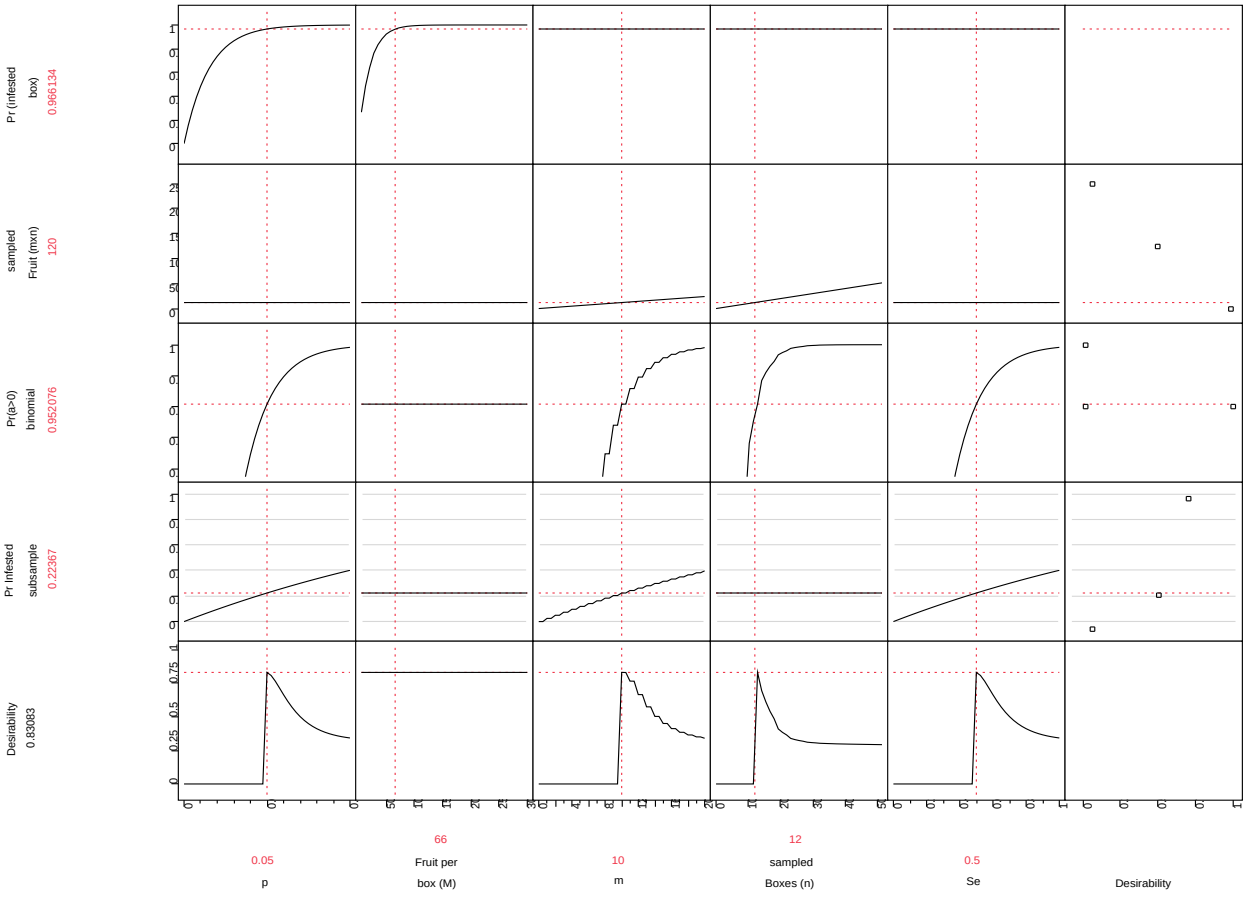
$$Pr(a > 0) = \ 1 - \left(1 - (\ 1 - (1 - (\mathrm{Se} \cdot \mathrm{P})^{\mathrm{m}})\right)^{\mathrm{n}} \qquad (19)$$

At the first observation we notice the relationship's functional form of the magnifying effect of the probability of and infested sample of fruit nested within the sample of boxes. Also we note there is neither a population of boxes, N, nor population of fruit, M in the relationship. Unless the populations

are large this could lead to an overestimate of sample size. On the other it provides a general form for a set P desired detection and Se sensitivity.

Figure 17 provides a graph showing a snapshot of the functional relationship $Pr(a > 0)$binomial in equation (18). It also shows the relationship for Pr(infested box) the probability of an infested box, sampled fruit (mxn) the number of sampled fruit, Pr(infested subsample) the probability of an infested subsample of m fruit, and the JMP Desirability function. The functional relationship of $Pr(a > 0)$ binomial increases asymptotically to 1 as P the desired detection, m fruit sampled in the box, n boxes sampled, or Se sensitivity as anyone or combination factors increase. The JMP desirability function are smooth piecewise functions crafted to fit the control points(SAS_Institute_Inc. 2011). The minimize and maximize functions are three-part piecewise smooth functions that have exponential tails and a cubic middle. The target function is a piecewise function that is a scale multiple of a normal density on either side of the target (with different curves on each side), which is also piecewise smooth and fit to the control points. When the desirability graph to the right has a positive slope the desirability function objective is maximize when the slope is negative the objective is minimize and when the

Figure 17 Two-stage with binomial distribution for boxes and binomial for the fruit.



function forms an arrow point to the right the objective target is the point of the arrow. The relationships across the bottom of the graph show where the desirability is maximized for each input. This is very helpful for meeting objectives in two-stage sampling. With the detection P and sensitivity levels Se set we can explore combinations of n boxes to sample and m fruit to sample within the selected boxes. The maximum number of fruit to sample within the selected boxes can be found. Also find the maximum number of boxes to sample if only one fruit per box were sampled. Plus any m, n combinations of interest between these two extremes can be evaluated. For the $P = 0.05\ and\ Se = 0.5$ the minimum number of fruit to inspect stays around 120.

The next two-stage distribution combination to consider will be hypergeometric for the boxes and binomial for the fruit within the box. The equation (2) with sensitivity adjustment and hypergeometric function are combined to provide the following result:

$$\Pr(a > 0) = 1 - (Hypergeometric\ Distribution\ (N, ceiling[big\ A\ boxes], n, 0)) \quad (20)$$

Where $big\ A\ boxes = (1 - (1 - (Se \cdot \mathrm{P})^{\mathrm{m}}) \cdot N$

Figure 18 provides a graph showing a snapshot of the functional relationship $Pr(a > 0)$hypergeometric in equation (20). It also shows the relationship for Pr(infested box) the probability of an infested box, sampled fruit (mxn) the number of sampled fruit, Pr(infested subsample) the probability of an infested subsample of m fruit, big A boxes and the JMP Desirability function. The functional relationship of $Pr(a > 0)$ binomial increases asymptotically to 1 as P the desired detection, m fruit sampled in the box, n boxes sampled, or Se sensitivity as anyone or combination factors increase. The relationship of $Pr(a > 0)$ with Total Boxes (N) approached 0.95 from above as the number of boxes increases. With the detection P and sensitivity levels Se set we can use the desirability function to explore combinations of n boxes to sample and m fruit to sample within the selected boxes. The maximum number of fruit to sample within the selected boxes can be found. Also find the maximum number of boxes to sample if only one fruit per box were sampled. Plus any m, n combinations of interest between these two extremes can be evaluated. For the $P = 0.05\ and\ Se = 0.5$ the minimum number of fruit to inspect stays around 120.

The third two-stage distribution combination to consider will be binomial for the boxes and hypergeometric for the fruit within the box. The equation (2) and hypergeometric function with sensitivity adjustment are combined to provide the following result:

$$\Pr(a > 0) = 1 - (1 - (Hypergeometric\ Distribution\ (M, round[big\ A\ fruit, 0], m, 0)))^{\mathrm{n}} \quad (20)$$

Where $big\ A\ fruit = Se \cdot \mathrm{P} \cdot M$

Figure 18 Two-stage with hypergeometric distribution for boxes and binomial for the fruit.

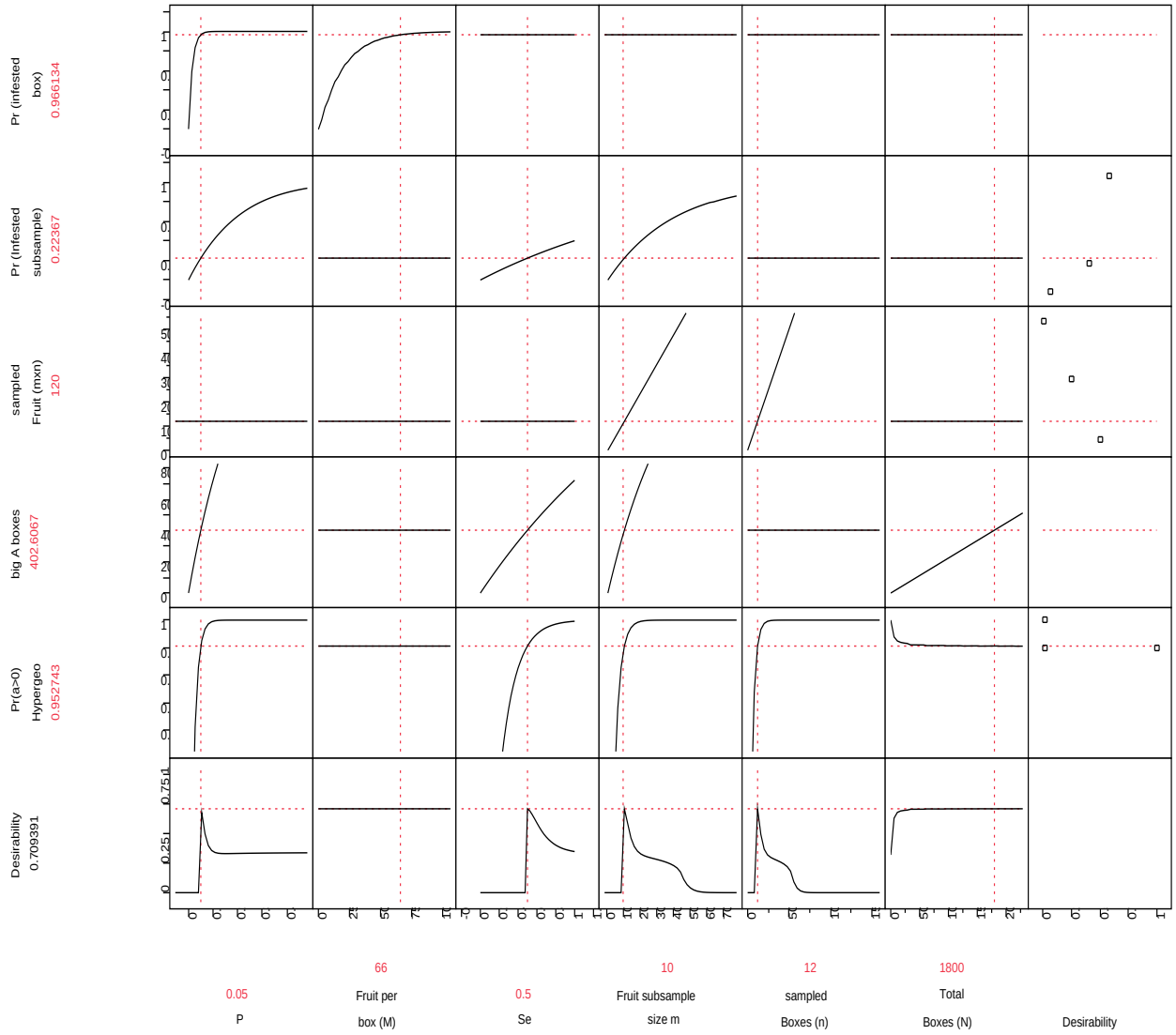
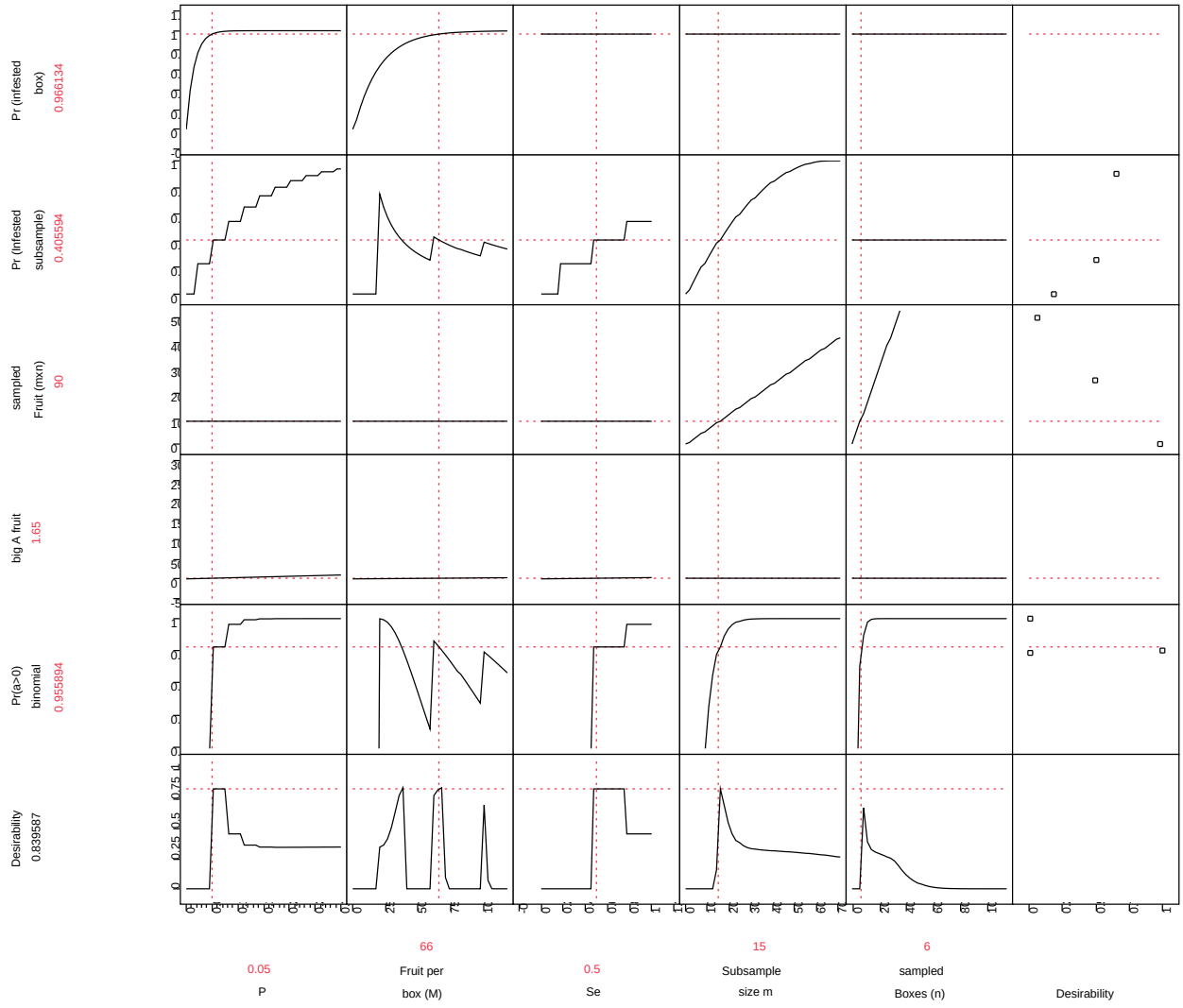


Figure 19 provides a graph showing a snapshot of the functional relationship $Pr(a > 0)$binomial in equation (20). It also shows the relationship for Pr(infested box) the probability of an infested box, sampled fruit (mxn) the number of sampled fruit, Pr(infested subsample) the probability of an infested subsample of m fruit, big A fruit and the JMP Desirability function. The functional relationship of $Pr(a > 0)$ binomial increases asymptotically to 1 as P the desired detection, m fruit sampled in the box, n boxes sampled, or Se sensitivity as anyone or combination factors increase; however, the P and the Se are step functions with the Se being the more pronounced step. With the detection P and sensitivity levels Se set we can use the desirability function to explore combinations of n boxes to sample and m fruit to sample within the selected boxes. The maximum number of fruit to sample within the selected boxes can be found. Also find the maximum number of boxes to sample if only one fruit per box were

sampled. Plus any m, n combinations of interest between these two extremes can be evaluated. For the $P = 0.05\ and\ Se = 0.5$ the minimum number of fruit to inspect stays around 90.

Figure 19 Two-stage with binomial distribution for boxes and hypergeometric for the fruit.



The last two-stage distribution combination to consider will be hypergeometric for the boxes and hypergeometric for the fruit within the box. The hypergeometric function for boxes with big A for boxes estimated Pr(infested subsample and hypergeometric function with sensitivity adjustment are combined to provide the following result:

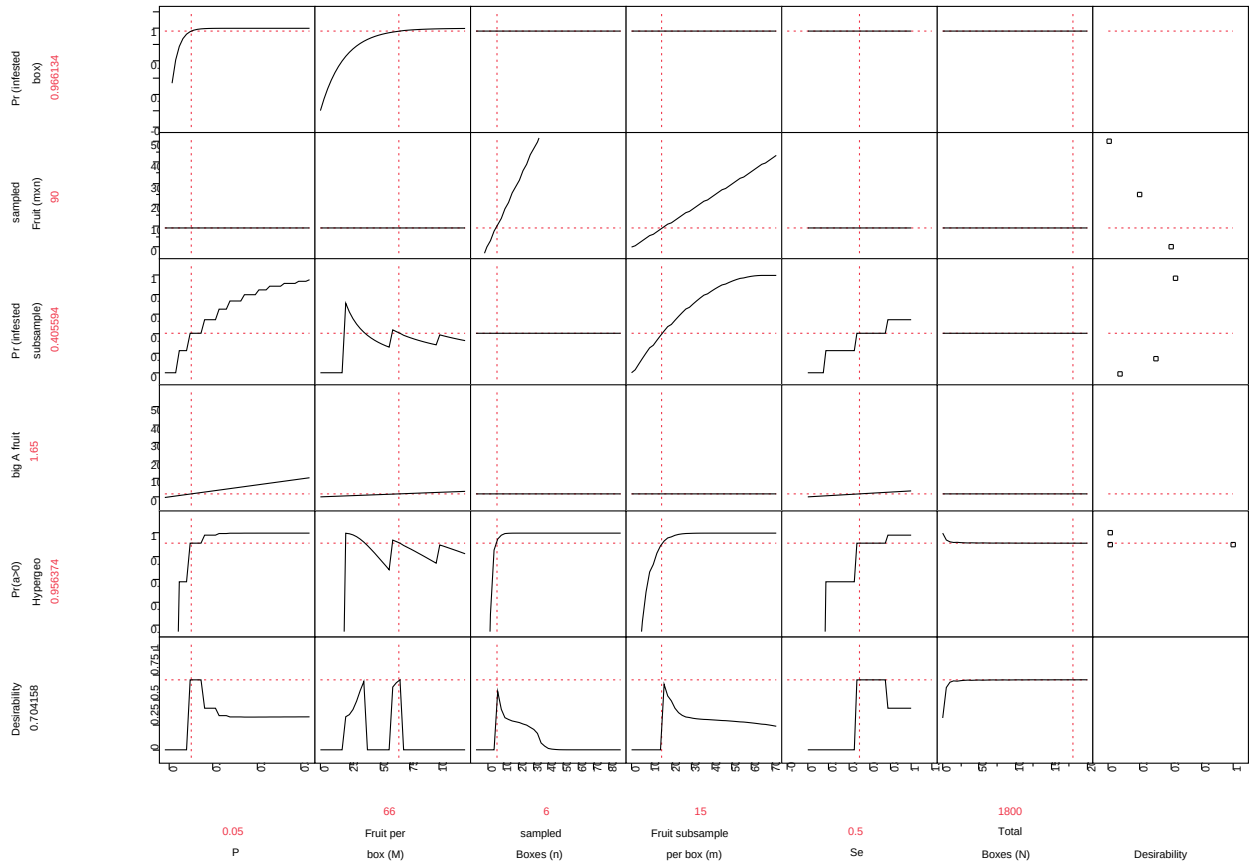
$$\Pr(a > 0) = 1 - (Hypergeometric\ Distribution\ (N, ceiling[big\ A\ boxes]\ , n, 0) \qquad (20)$$

where $big\ A\ boxes = \left(1 - Hypergeometric\ Distribution\ (M, round[big\ A\ fruit, 0], m, 0)\right) \cdot N,$

and, $big\ A\ fruit = Se \cdot \mathrm{P} \cdot M.$

Figure 20 provides a graph showing a snapshot of the functional relationship $Pr(a > 0)$binomial in equation (21). It also shows the relationship for Pr(infested box) the probability of an infested box, sampled fruit (mxn) the number of sampled fruit, Pr(infested subsample) the probability of an infested subsample of m fruit, big A fruit and the JMP Desirability function. The functional relationship of $Pr(a > 0)$ binomial increases asymptotically to 1 as P the desired detection, m fruit sampled in the box, n boxes sampled, or Se sensitivity as anyone or combination factors increase; however, the P and the Se are step functions with the Se being the more pronounced step. The relationship of $Pr(a > 0)$ with Total Boxes (N) and Fruit per box (M) approached 0.95 from above as the number of boxes or fruit per box increases. With the detection P and sensitivity levels Se set we can use the desirability function to explore combinations of n boxes to sample and m fruit to sample within the selected boxes. The maximum number of fruit to sample within the selected boxes can be found. Also find the maximum number of boxes to sample if only one fruit per box were sampled. Plus any m, n combinations of interest between these two extremes can be evaluated. For the $P = 0.05\ and\ Se = 0.5$ the minimum number of fruit to inspect stays around 90.

Figure 20 Two-stage with hypergeometric distribution for boxes and hypergeometric for the fruit.



The difference in these four plans is quite dramatic. The two-stage which uses a hypergeometric distribution for the boxes and hypergeometric for the fruit and binomial/hypergeometric combination produced sample sizes which were 25% smaller than the plans based on using a hypergeometric distribution for the boxes and binomial for the fruit and hypergeometric/binomial combination. The plans using a hypergeometric distribution on the fruit within the boxes produced smaller samples than the plans using the binomial. For the populations there were 1800 boxes and 66 fruit per box, with a sample designed to detection a 5% ($P = 0.05$) infestation with 95% confidence. We should expect there to be very little difference between the binomial and hypergeometric on the 1800 box population because with a population that large there is very little difference in sample needed; however, hypergeometric based sample plan diverges from the binomial as the population decreases and becomes even more sensitive to population changes as the population becomes smaller and smaller.

Figure 21

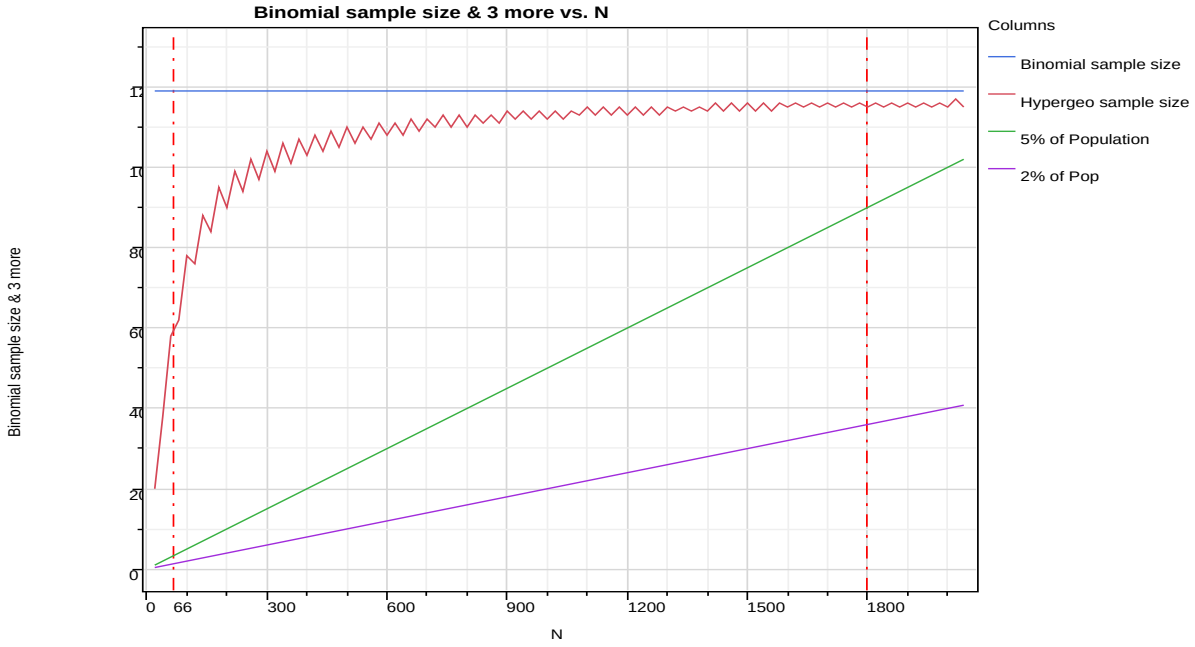
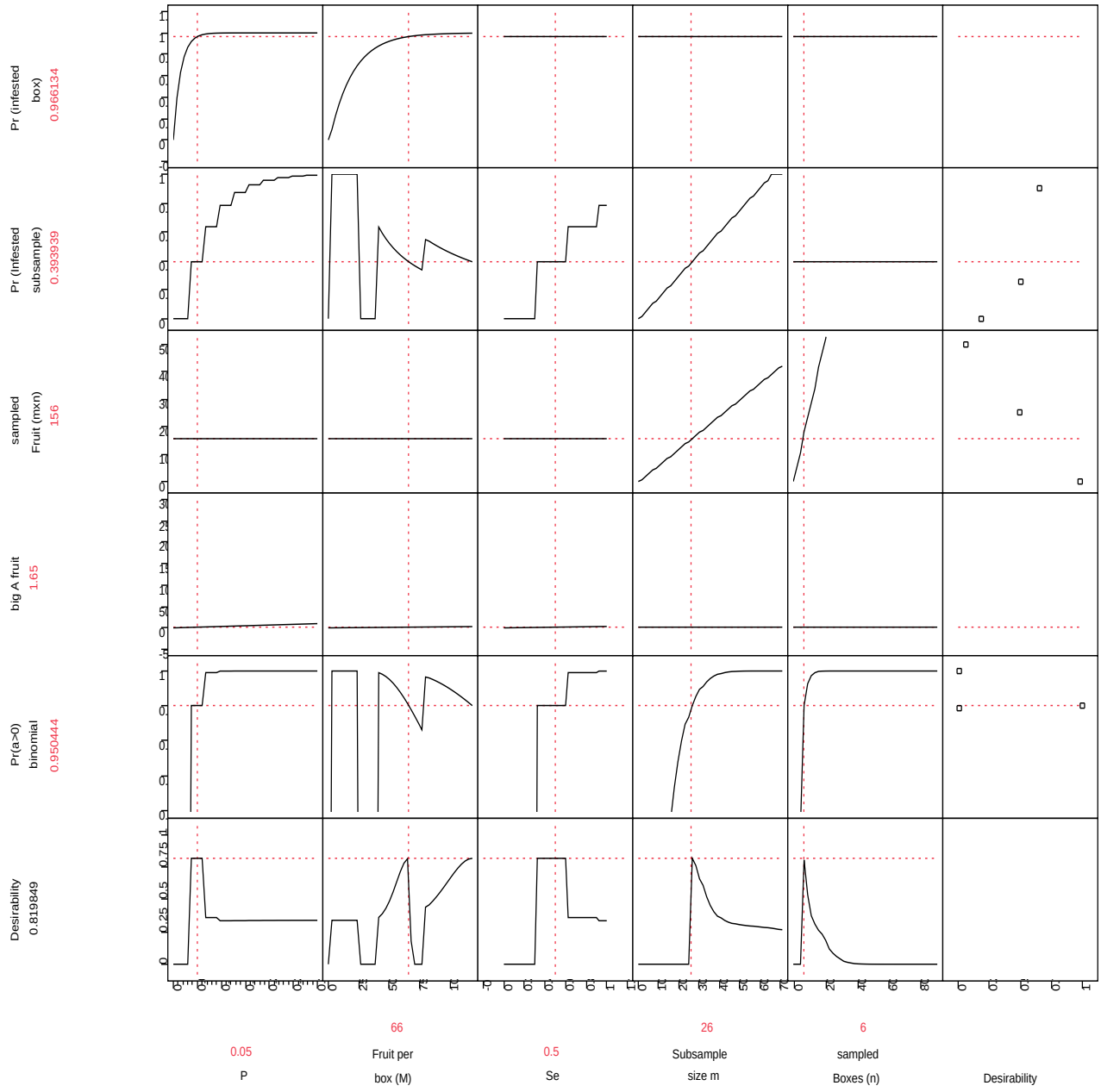


Figure 21 details 5% detection lines for the binomial and hypergeometric distributions presented in figure 10, where the relationship between the population size N, the sample size n for distributions were presented. The figure 21 sample size relationships have been adjusted for sensitivity. The graph shows why there was so little a difference between the binomial and hypergeometric based sample plans for the populations of 1800 boxes. There is very little difference between sample size needed for the hypergeometric based sample plan and the binomial based plan. This can be seen by following the red dashed line for a population of 1800 up to where it intersects the Hypergeometric and binomial sample size lines. This is why when there were a large number of boxes, either the binomial or the hypergeometric sample plan work well; however, the binomial would be preferred because of ease of working with functional relationship. When the population size decreases the hypergeometric based

sample size diverges from binomial sample size. This is because the hypergeometric detection efficiency relative to sample size increases as the population size decreases. Figures 20 and 18 show the Pr(infested subsample), the probability of an infested subsample for the hypergeometric is 0.41 when sampling 15 of the 66 fruit per box and the binomial is 0.22 with a sample size of 10 fruit per box. The difference is binomial calculates probability of infestation based on the sample size regardless of population size. The population size could be 10, 66, 100 or 1,000 or whatever the binomial estimate would be .22 for a sample of 6. While the hypergeometric probability of infestation is based on the sample size taken from a population of 66 in which 2 fruit were identified as infested. Since the hypergeometric probability includes a consideration of population size and the sample size is large relative to the population we feel intuitively that the hypergeometric based sample design should provide a better estimate of the probability of infestation than the binomial. In this case selecting subsamples from boxes of 66 fruit the binomial distribution seems to be an underestimate. When selected from the small population the hypergeometric provides a better estimate of the probability of infestation; however, the hypergeometric provides at best erratic estimates of infestation for small populations which can be over estimates, under estimates and exact estimates when used with the P detection parameter. There will be more discussion of the sample design problems caused when P is used with the use hypergeometric distribution.

The decreased sample sizes observed when the sample design applies the hypergeometric distribution to the fruit is very sensitive to the estimate of A (big A Fruit) for fruit. In equation (20) the round function is applied to A (big A Fruit) for fruit in the hypergeometric function. This overrides the JMP default to round down all non-integers inputs in the hypergeometric function. Using the round function is less conservative than the JMP approach but this decision was made to try to minimize the errors when P is used with the hypergeometric distribution in the sample design. The logic was that if the estimate of A infested fruit in the box is a non-integer then follow the rounding rules would result in the A used being closest to the estimate $big\ A\ fruit = Se \cdot P \cdot M$. We should also note that $Se = 0.5$ is a very conservative approach to inspection sensitivity. This assumes that the inspectors detect infested fruit 50% of the time they are faces with infested fruit. The small change of using round in the hypergeometric makes a large difference in sample size. The two-stage sample design for without the rounding the function on $big\ A\ fruit = P \cdot Se \cdot N$ is presented in Figure (22). Without using the round function the sample size increases to 156 fruit inspected with 6 boxes selected and 26 fruit selected within each box. The same result is observed when the sample design is hypergeometric for the boxes and hypergeometric for the fruit within the boxes. This change puts the sample plans with hypergeometric applied to the fruit within the boxes more costly as far as number of samples than the plans that use the binomial applied to the fruit within the boxes. This brings us to the questions. What are the differences in detection between the alternative plans? What approach is best?

Figure 22 Two-stage with binomial distribution for boxes and hypergeometric for the fruit without rounding A for fruit in the hypergeometric function.



P is the desired detection of the sample plans. Sample plan designs which apply the binomial to a small population (number of fruit in the boxes) provide a very straight forward application of P in the plan; however, when P is applied to the hypergeometric sample plan design the A (number of infested fruit per box) is the product of the sensitivity, the desired detection and the population $A = Se \cdot P \cdot M$. This product can produce an integer or a non-integer. If the product is a non-integer it must be adjusted to an integer, A', for use in the hypergeometric distribution sample plan. The true detection of the sample

plan changes from the original desired detection, P, when this adjustment to an integer occurs. The new sample-p' after the adjustment is as follows:

$$p' = A'/(M \cdot Se). \qquad (21)$$

In our examples above Figures 19 and 22 P=0.05,Se=0.5 and M=66. The product is 1.65. In the Figure 19 example 1.65 rounds to 2 and true detection of the sample plan is 0.061 which means the sample/inspection would detect an infestation larger than 6.1% with 95% confidence rather than the planned 5%. The absolute change is 1.1% fewer pests would be detected. The relative increase is 21% more pests getting through the sample/inspection undetected. The two-stage sample is 6 boxes with a 15 fruit subsampling per sample box. Total fruit sampled is 90. The Pr (infested subsample), the probability of an infested subsample equals 0.406.

In the Figure 22 example 1.65 rounds down to 1 and true detection of the sample plan is 0.03 which means the sample/inspection would detect an infestation larger than 3% rather than the planned 5%. The absolute change is 2% more pests would be detected. The relative decrease is 39% fewer pests getting through the sample/inspection undetected. The two-stage sample is 6 boxes with a 26 fruit subsampling per sample box. The probability of an infested subsample, Pr (infested subsample), equals 0.394.

The average of these two plans is 20.5 which would round to a 21 fruit subsample per box. With 6 boxes sampled the total fruit sampled would be 126.

If a straight line interpolation method is applied using the 0.65 from the 1.65 above the result would be 6 boxes sampled with a subsample of 19 fruit per sampled box. Total fruit sampled is 114.

These compare to the sample design applying binomial to the boxes and binomial to the fruit with in the box with a detection, P of 0.05 (5%) which when 6 boxes are sampled require a subsample of 20 fruit per sampled box. Total fruit sampled is 120. The probability of an infested subsample, Pr (infested subsample) equals 0.397.

Equation (5a) is hypergeometric $\Pr(a = 0)\ and \Pr(a > 0) = 1 - \Pr(a = 0)$ to maintain the hypergeometric relationship the inputs must be integers. If A is set to $(Se \cdot P \cdot M)$ substituting M for N and m for n and not adjusting the A an integer creates the following result:

$$\Pr(infested\ subsample) \approx 1 - \frac{(M-SePM)!(M-m)!}{(M-SePM-m)!M!} \qquad (22)$$

where $SePM$ not adjusted to an integer.

The true hypergeometric probability relationship no longer exists when $A = (Se \cdot P \cdot M)$ is not an integer, which occurs frequently. However, equation (22) works well when A is an integer as in the true hypergeometric relationship in equation (20). If the equation (22) result is substituted into the equation for the probability of an infested subsample the result is as follows:

$$\Pr(a > 0) = 1 - (1 - \frac{(M - SePM)!(M-m)!}{(M - SePM - m)!M!})^{\mathrm{n}}, \qquad (23)$$

where $big\ A\ fruit = Se \cdot \mathrm{P} \cdot M$.

Note equation (23) has limited application, because most computers cannot compute >170!.

When $\Pr(a > 0) = 0.95\ with\ Se = 0.5, P = 0.05, n = 6$ the relationship has an $m = 18$ and the total samples are 108. It is interesting that fruit per box sample is closest to the interpolated results above.

Foe the calculations in this document the hypergeometric distribution function has based on the JMP function $Hypergeometric\ Distribution\ (M, round[Se \cdot P \cdot M, 0], m, 0)$ which is the $\Pr(a = 0)$; however, the $Hypergeometric\ Probability\ (M, round[Se \cdot P \cdot M, 0], m, 0)$ can serve the same purpose because both provide equal results for $\Pr(a = 0)$ for identical inputs. Additionally the $Hypergeometric\ Probability\ (M, [Se \cdot P \cdot N], m, 0)$ function accepts non-integers and yield results which match equation (23), and the hypergeometric probability function is not limited by the population size. The JMP $Hypergeometric\ Probability$ function appears to be using the Gamma function to calculate probability. Sample designs developed using the JMP hypergeometric probability function provide solutions when A is not an integer. It appears to find solutions between the hypergeometric integer solutions. Using the JMP hypergeometric probability function issues involving sample-p' and A' because the actual $A\ (big\ A\ fruit) = Se \cdot P \cdot M$ can be applied in the JMP hypergeometric probability function. JMP Support was contacted concerning the use of the JMP hypergeometric probability function the contacted the complete response is in Appendix D. A summary of the response is as follows(JMP_Technical_Support and Archer 2011);

- The Gamma function used to compute factorials (and combinations) in the Hypergeometric probability functions,
- k! = Gamma(k+1) where Gamma is as defined here http://en.wikipedia.org/wiki/Gamma_function,
- JMP expects the hypergeometric probability function to only take integer values,
- use of non-integer values little concerning and should be careful how the results are used,
- When non-integer values are used can't really be interpreted as probabilities,

Using the equation 23 non-integer application may make sense because the non-integer represents a set of boxes where the average infestation per box is not exactly an integer. An example of this would be 165 infested fruit in 100 boxes of 66 fruit per box. Assuming the sample plan is binomial for boxes and the rounded hypergeometric is applied to the fruit in the box, the rounded representation is 200 infested fruit in 100 boxes of 66 fruit per box. If the JMP hypergeometric probability function is used on the fruit in the boxes the representation is 165 infested fruit. Taking advantage of the hypergeometric probability function's use of non-integers seems reasonable. But caution must be used. Further study using simulation seems to be warranted and a better understanding of how it performs on small populations is necessary.

Figure 23 displays graphs of the functional relationships for $\Pr(a > 0)$ the probability of a two-stage detection sample with 95% confidence. The inputs are P, the desired detection, Se, the inspection sensitivity, m, the number of boxes sampled and n, the number of fruit subsampled. In the sample plans the binomial distribution was applied to boxes for all $\Pr(a > 0)$ with the following different distribution strategies applied to the fruit.

- hypergeometric with A rounded down,
- hypergeometric with A rounded,
- binomial applied to the fruit, and
- equation 23 approximation applied to the fruit (JMP $Hypergeometric\ probability$ function).

Figure 23

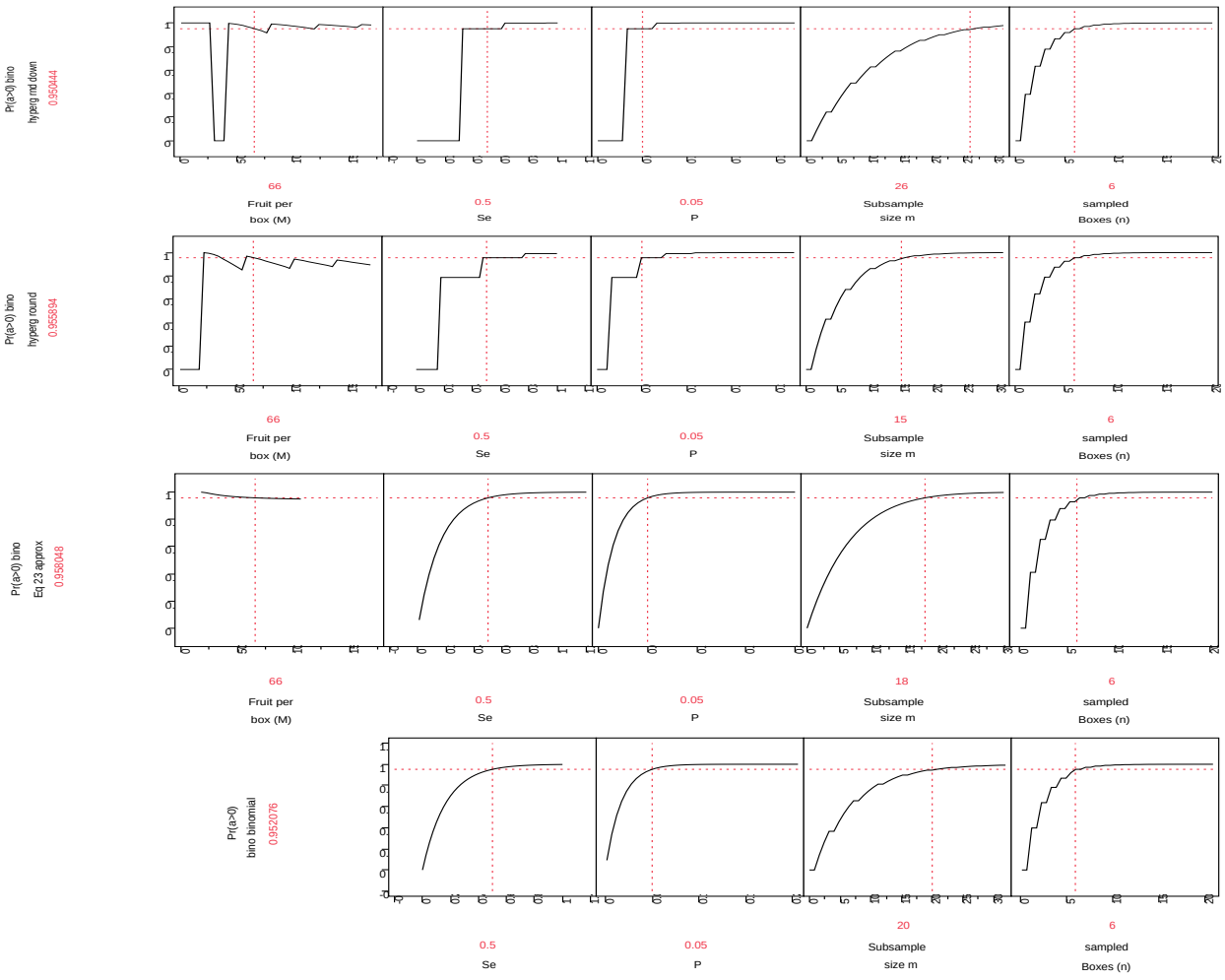
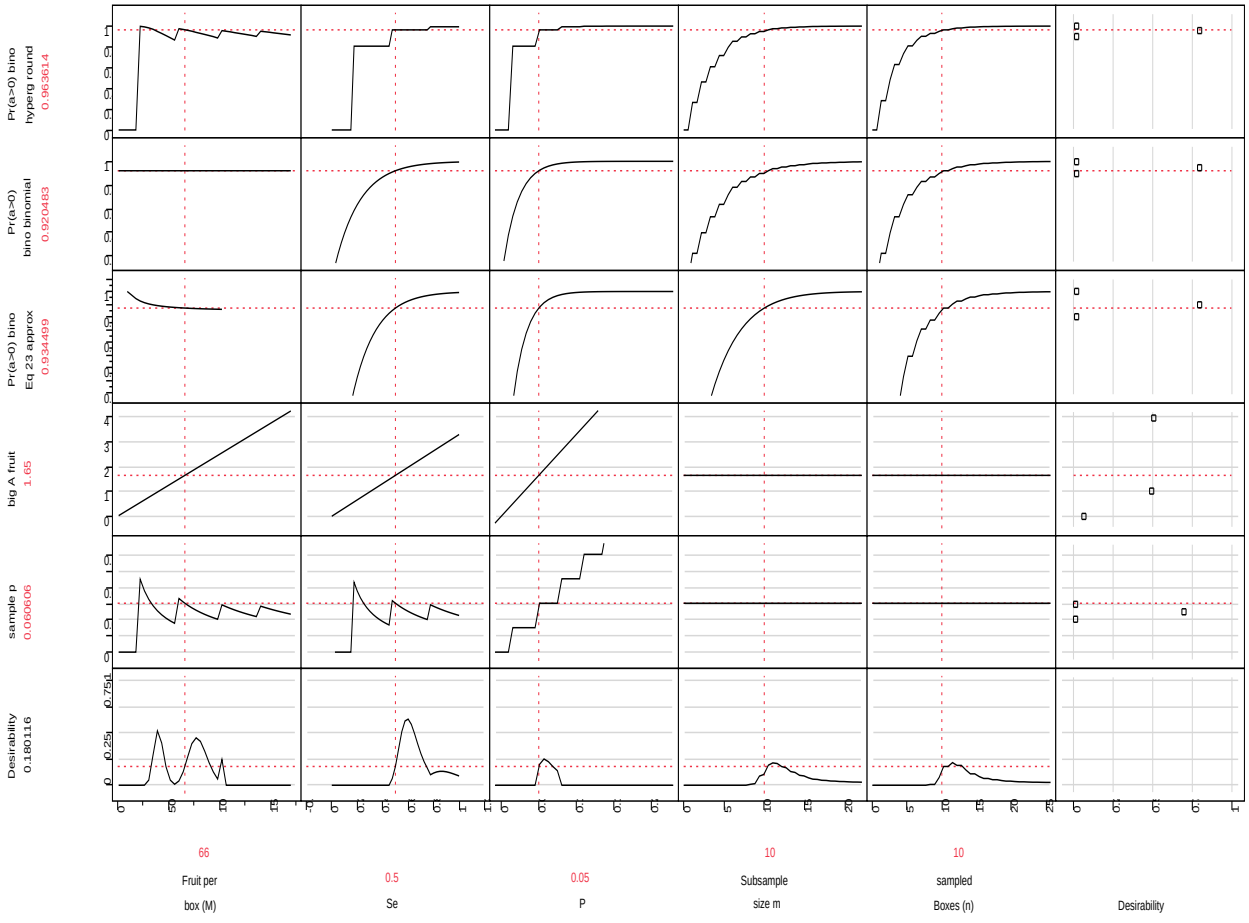


Figure 23 compares four strategies to develop a two-stage detection sample design. The functional relationships show points ($A = 0$) where the $Pr(a > 0)$ for the hypergeometric sample design on fruit within the box would not respond to increases, n boxes or m fruit per box. This is because A for fruit is rounding to zero. If there is no $A\ in\ M$ then the sample probability of $a\ in\ m\ equals\ zero$. If any part of

an integer in the hypergeometric function is rounded down then $big\ A\ fruit = Se \cdot P \cdot M$ could be less than one and result in $Pr(a > 0) = 0$. This could happen frequently when M is small. It depends on $big\ A\ fruit = Se \cdot P \cdot M$.

Sample plan development evaluation accommodate the example with 1800 boxes and 66 fruit per box as an example; however, it must also be applicable to shipments with 35 boxes with 20 fruit per box or any other combination. Sample design needs to be flexible to apply to all inspection/sample situations. The issues the plan must address are the actual sample-p' for fruit within the box and small values of A for fruit with in the box. When consignments contain few boxes, a situation frequent encountered at the plant inspection stations; applying the hypergeometric distribution to the boxes provides an advantage, as long as the number of boxes is not too small. Also, when the sample within the box is small the actual sample-p' for the boxes and small values of A for boxes could easily become issues the sample plan must address.

Figure 24



A development structure is needed to account for the issues mentioned above when analyzing various sample plans. The JMP Profiler provides an excellent platform for this analysis (see Figure 24). Using this

approach the performance of sample plans can be compared across a wide range of alternatives. The number of boxes in the consignment, number of fruit per box, and the desired detection level can be observed to see now they interact. The effect of adjusting, Se the sensitivity could be evaluated in cases when inspectors vary in their performance (Gould 1995)or the detection of specific pests varies from commodity to commodity. The samples per box and the boxes sampled can be observed at various levels to determine and understand their effect on other variables. Sample-p's functional relationship is set by A' the adjusted big A for fruit. A' is a function of P, the desired detection, Se, the sensitivity and M, the number of fruit within the box. Observing the interactions of variables plays a critical role in understanding the pros and cons of the sample plans.

A combination of plans should provide a good solution. The binomial distribution and the hypergeometric rounded seem to provide several workable sample design applications. The binomial applied to boxes works well when the population of boxes is large for example 1800 boxes with 66 fruit per box. But this result does not work well as the number of boxes decreases. When the number of boxes decreases applying the hypergeometric distribution to the boxes becomes a resource saver; however, defining a point where this transition occurs is dependent on the variables that define the probability of an infested subsample. By studying various combinations of distributions area's can be defined where the combinations work the best and areas where small population sizes cause relationships to break down.

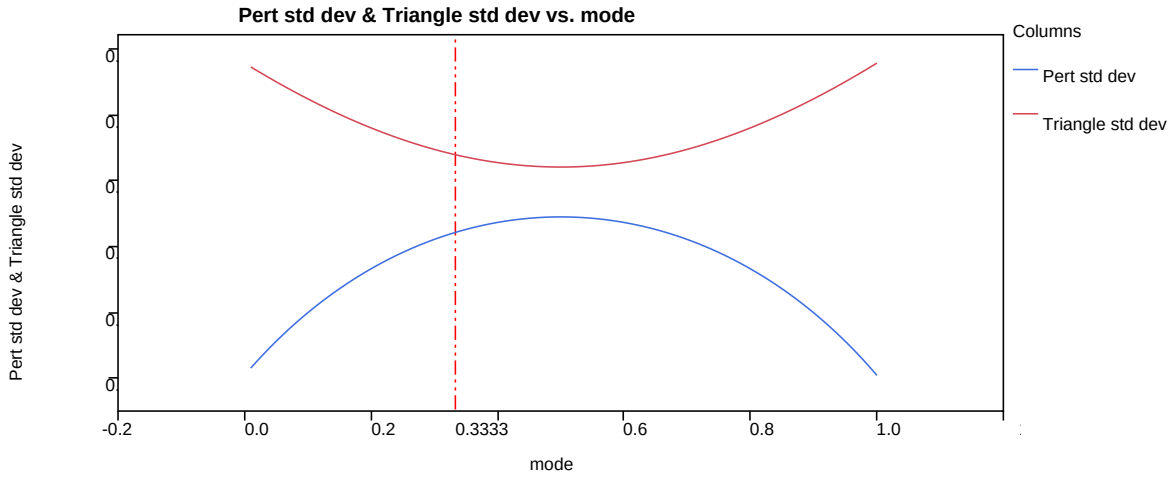
Modeling Expert Opinion

Sometimes it is necessary to rely on expert opinion when building a model. When experts have defined the minimum, maximum and most likely observation, the triangular distribution is logical to use as the basis for the model. The Pert distribution may offer a better alternative. The pert distribution is a special case of the beta distribution; however, the input parameters are much more easily understood than the beta parameters. It was developed in conjunction with project plan management. The Pert Project plan was where the Pert distribution was developed. For each task in the project plan the minimum, maximum, and the most likely time to complete the task were used to model the task completion.

The Pert Distribution should be compared with the Triangle distribution where the mean is equally sensitive to each parameter. The PERT distribution therefore does not suffer to the same extent the potential systematic bias problems of the Triangle distribution, that is in producing too great a value for the mean of the risk analysis results where the maximum for the distribution is very large(Vose 2008).

The standard deviation of a PERT distribution is also less sensitive to the estimate of the extremes. Although the equation for the PERT standard deviation is rather complex, the point can be illustrated very well graphically. Figure 26 compares the standard deviations of the Triangle and PERT distributions with minimum a=0, maximum b= 1, and varying most likely value c(Vose 2008).

Figure 26



The observed pattern extends to any set of values. The PERT distribution produces a systematically lower standard deviation than the Triangle distribution, particularly where the distribution is highly skewed (i.e. b is close to the minimum or maximum). As a general rule of thumb, cost and duration distributions for project tasks often have a ratio of about 2:1 between the (maximum - most likely) and (most likely - minimum), equivalent to c = 0.3333 on the figure above. The standard deviation of the PERT distribution at this point is about 88% of that for the Triangle distribution. This implies that using

PERT distributions throughout a cost or schedule model, or any other additive model with similar ratios, will display about 10% less uncertainty than the equivalent model using Triangle distributions.

Although the Pert distribution is not available in JMP as such, it can be modeled in JMP since it is a special case of the beta distribution. JMP defines the Beta distribution as follows(SAS_Institute_Inc. 2011):

Beta Distribution

> The beta distribution has two shape parameters: $\alpha > 0$ and $\beta > 0$. A threshold parameter (θ) and a scale parameter (σ) are additional arguments, where $\theta \leq x \leq \theta + \sigma$. The default value for θ is 0. The default value for σ is 1.

Developing the functional relationships needed to translate the Pert parameters a, minimum, b, maximum, and c, most likely into the bets distribution parameters is as follows(Vose 2008):

$$PERT\ (a, b, c)\ =\ BetaDistribution(x, \alpha, \beta, \theta(threshold), \sigma(scale))$$

where:

$\theta = a$,

$\sigma = a - b$,

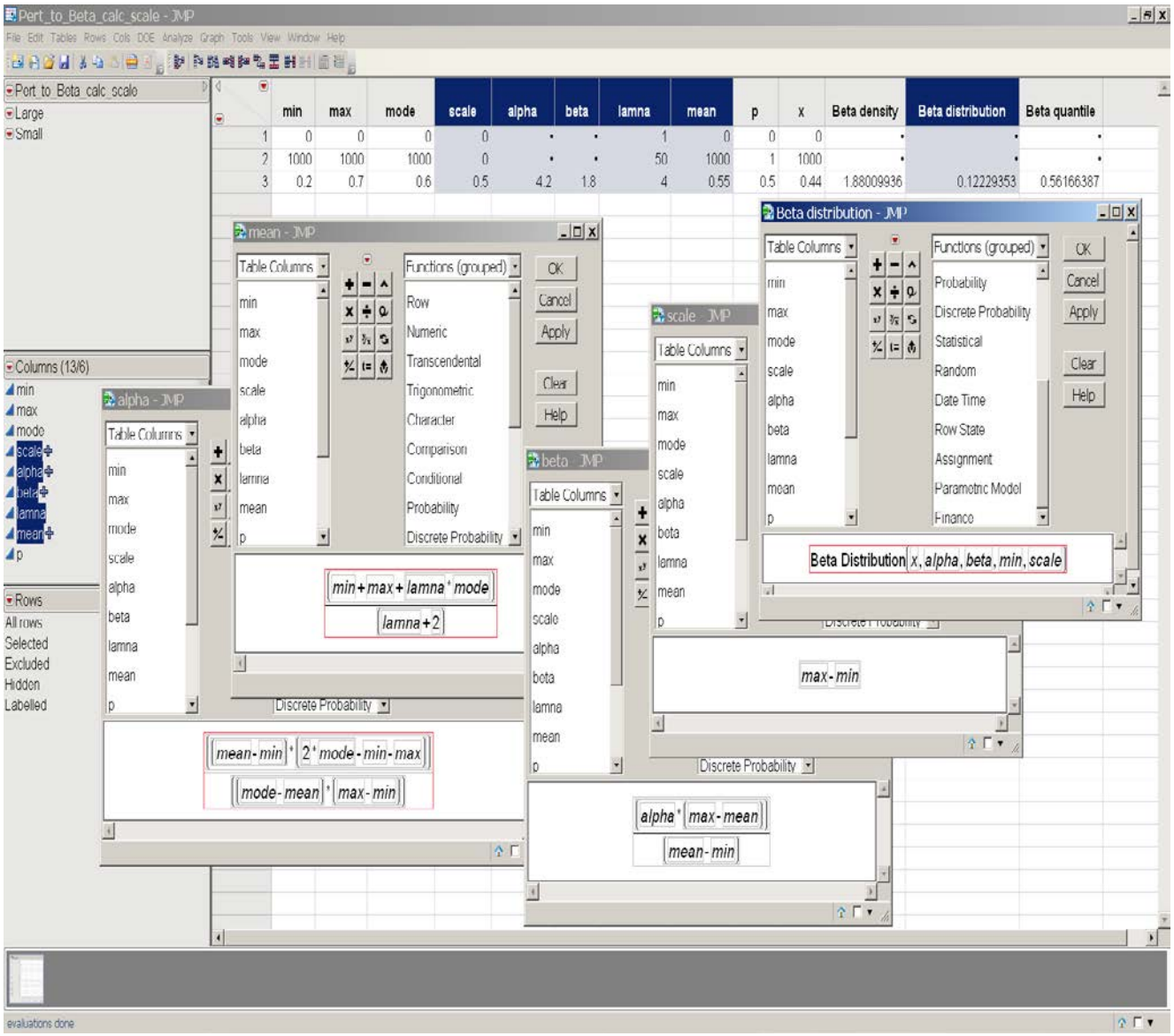
$\alpha = \frac{(\mu - a)(2c - a - b)}{(c - \mu)(b - a)}$,

$\beta = \frac{\alpha(b - \mu)}{(\mu - a)}$, and

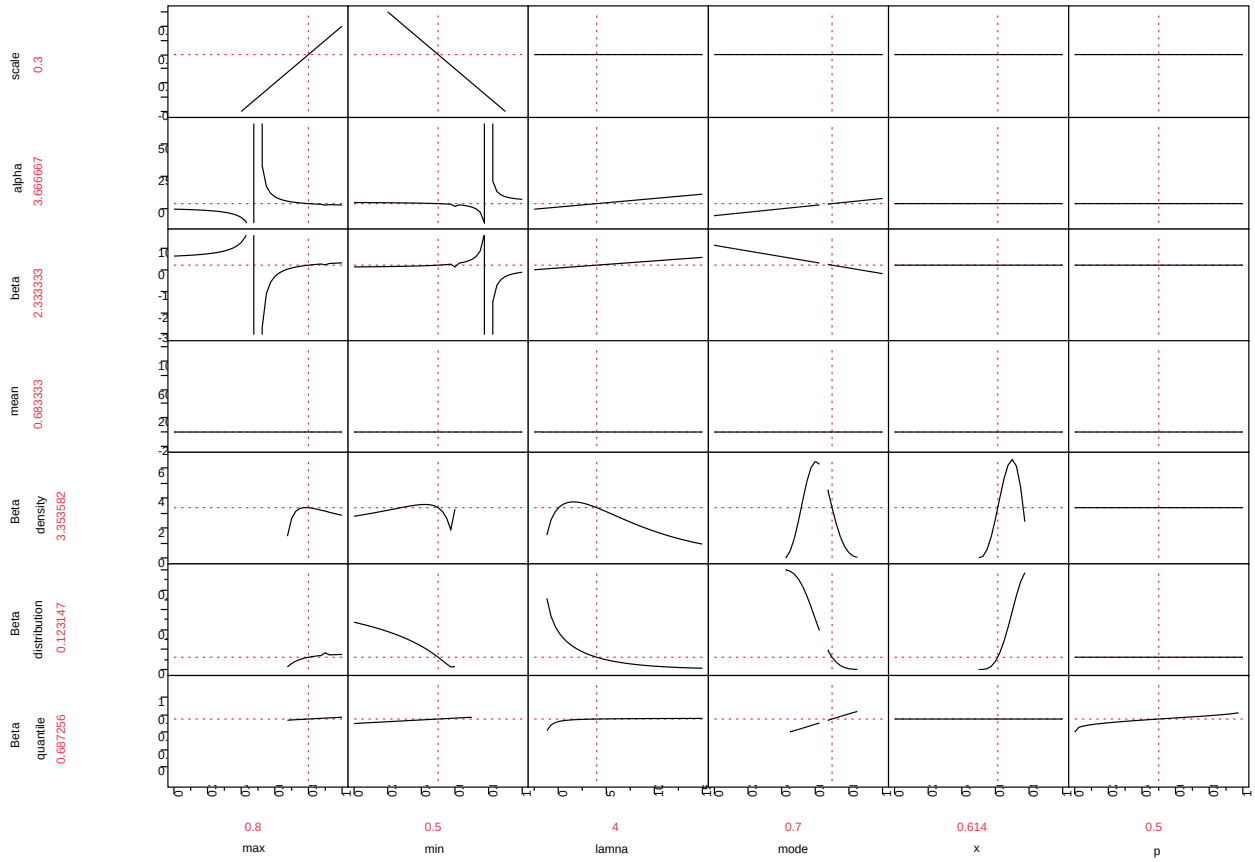
$\mu = \frac{a + 4\ c + b}{6}$.

The last equation for the mean is a restriction that is assumed in order to be able to determine values for α and β. It also shows how the mean for the PERT distribution is four times more sensitive to the most likely value than to the minimum and maximum values.

With these relationships a JMP table can be setup as follows:


Pert_to_Beta_calc_scale - JMP
File Edit Tables Rows Cols DOE Analyze Graph Tools View Window Help
Pert_to_Beta_calc_scale
Large
Small
Columns (13/6)
min
max
mode
scale
alpha
beta
lamna
mean
p
Rows
All rows
Selected
Excluded
Hidden
Labelled
min max mode scale alpha beta lamna mean p x Beta density Beta distribution Beta quantile
1 0 0 0 0 . . 1 0 0 0 . . .
2 1000 1000 1000 0 . . 50 1000 1 1000 . . .
3 0.2 0.7 0.6 0.5 4.2 1.8 4 0.55 0.5 0.44 1.88009936 0.12229353 0.56166387
mean - JMP
Table Columns
Functions (grouped)
OK
Cancel
Apply
Clear
Help
Row
Numeric
Transcendental
Trigonometric
Character
Comparison
Conditional
Probability
Discrete Probability
min+max+lamna*mode / lamna+2
alpha - JMP
Discrete Probability
(mean-min)*(2*mode-min-max) / (mode-mean)*(max-min)
scale - JMP
max-min
beta - JMP
alpha*(max-mean) / mean-min
Beta distribution - JMP
Probability
Discrete Probability
Statistical
Random
Date Time
Row State
Assignment
Paramotric Model
Finance
Beta Distribution x, alpha, beta, min, scale
evaluations done

The formulas can be entered into the profiler to get the following result:



The Profiler displays functional relationships and interactions. The notch in the mode-beta density panel represents the point where the functions are undefined. The mode equals the mean at that point the denominator of the alpha function is zero. The Profiler also shows that some nonsense relationships can be defined for alpha and beta where the mode can be larger than the maximum or less than the minimum. Caution must be used when using this tool.

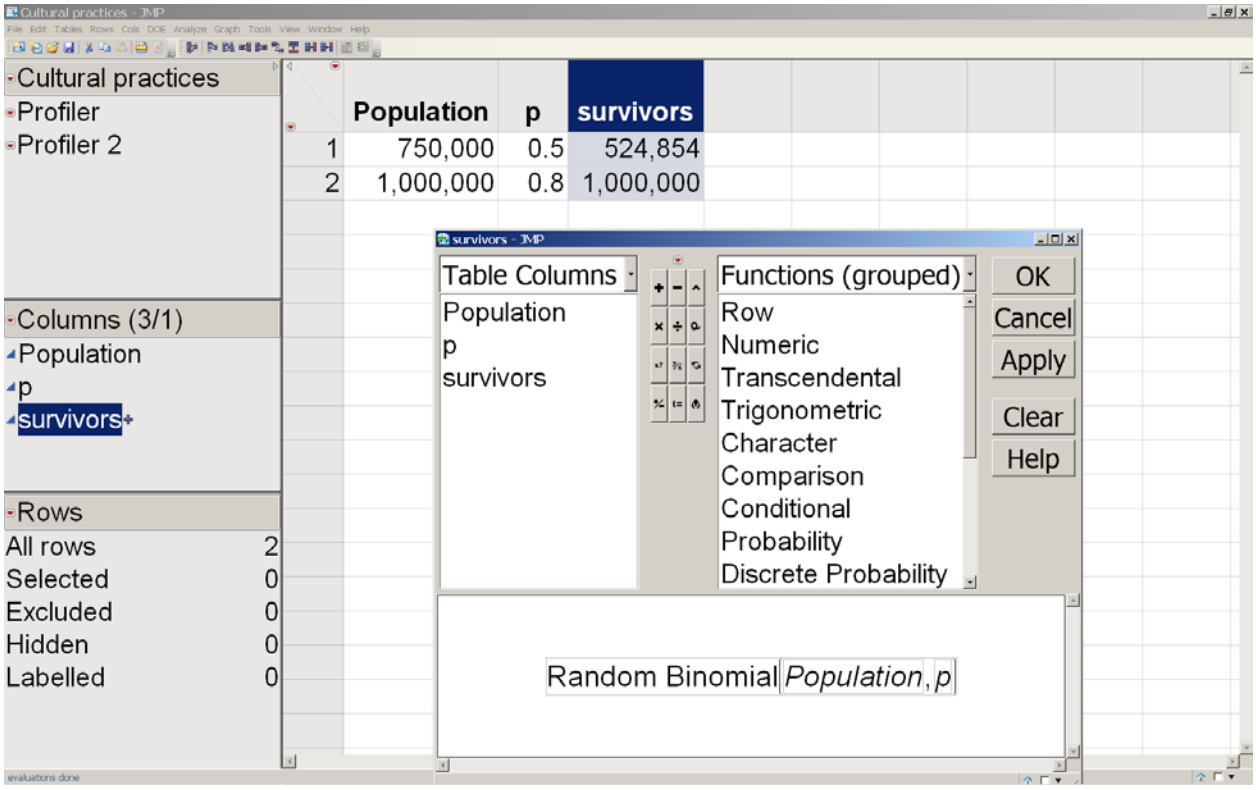
The use the Pert distribution can be demonstrated with a simulation. A team of experts indicates that implementing a certain set of cultural practices in the growing of fruit will curtail; pest survival. The team indicates that the probability of survival with the cultural practices ranges from 0.5 to 0.8 and is most likely to be 0.7. The beta parameters are as follows:

$$\alpha = \frac{(\mu-a)(2c-a-b)}{(c-\mu)(b-a)} = 3.66667, \text{ and } \beta = \frac{\alpha(b-\mu)}{(\mu-a)} = 2.33333.$$

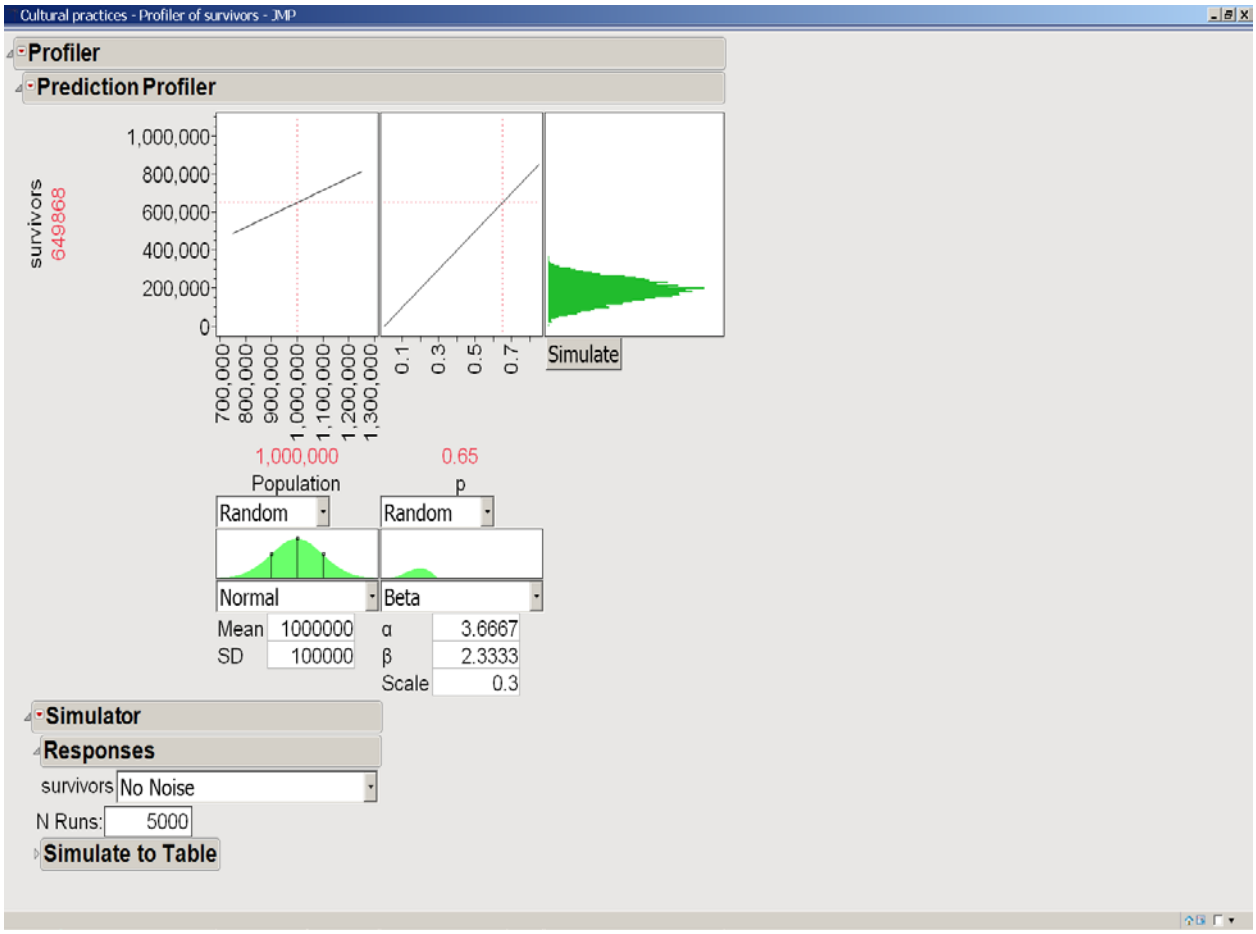
When the Pert distribution is defined with the beta distribution the sum of alpha and beta will equal 6. This can be used as a check that the parameters have been correctly estimated. If the population of fruit ranges from 750,000 to 1,250,000 the pest survival in the fruit can be simulated by applying the Pert

distribution to the probability of survival and the normal distribution to the population of fruit. Then the random binomial function can be used to simulate the number of infested fruit.

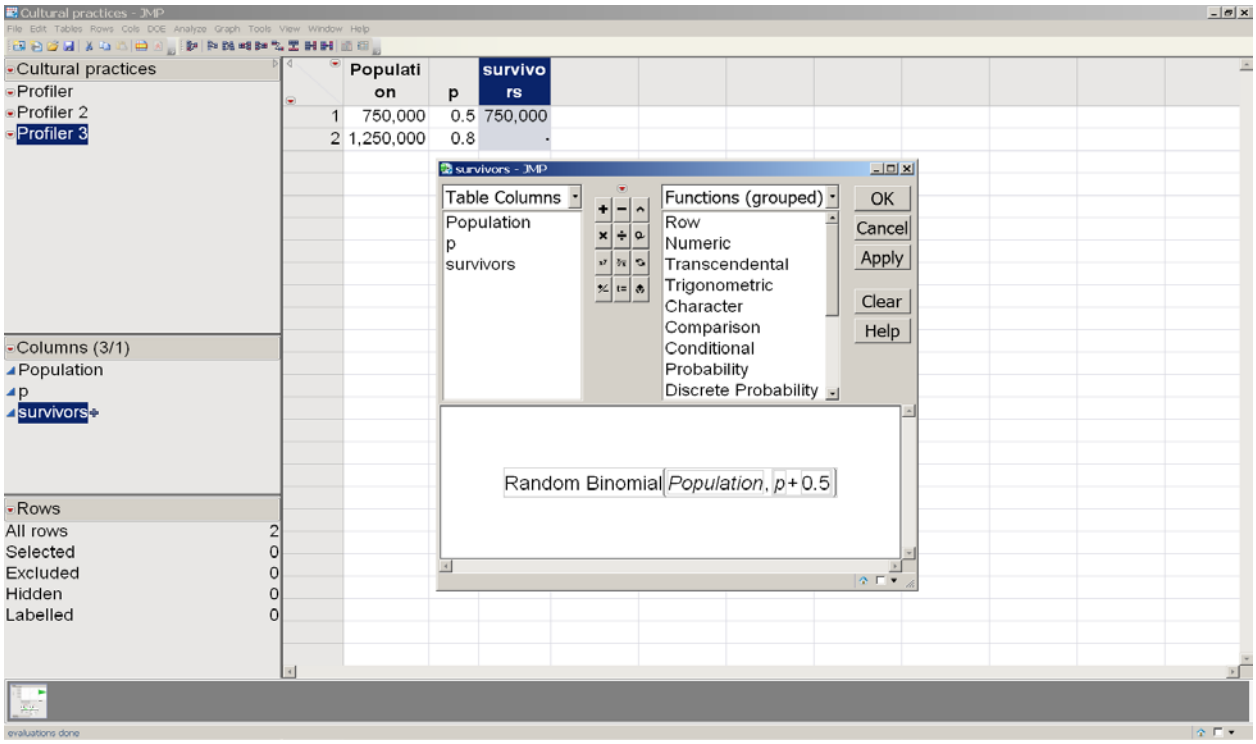
The JMP table to accomplish this would appear as follows:



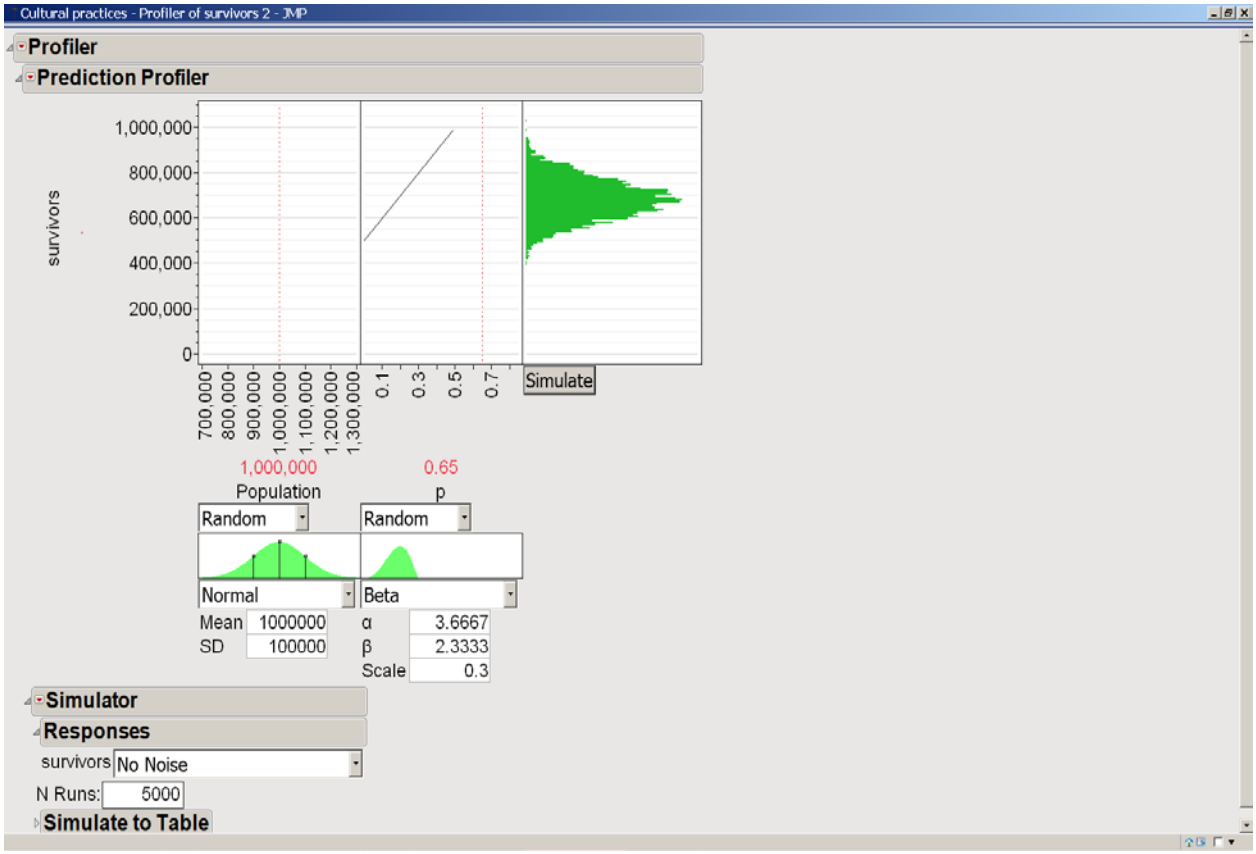
To start bring the survivors into the Profiler and choose the simulator option. Next change the population from fixed to random and select normal distribution. For the p, probability of survival change fixed to random and select beta for the distribution then enter alpha=3.6667 and beta=2.333 and set the scale to 0.3. Then after some axis adjustment click the simulate button to get the result below:



The simulator result does not match what we expect when there are suppose to be 50% to80% survivors in 1,000,000 with a most likely value of 70%. When the Pert to beta conversion was done there was a position parameter theta; however, that option was not available in the simulator. The simulator assumes the starting position of zero. There is a way to adjust the simulation to the correct starting position. The formula for survivors in the data table needs to be adjusted by adding 0.5 the p in the binomial function. The updated table and formula are as follows:



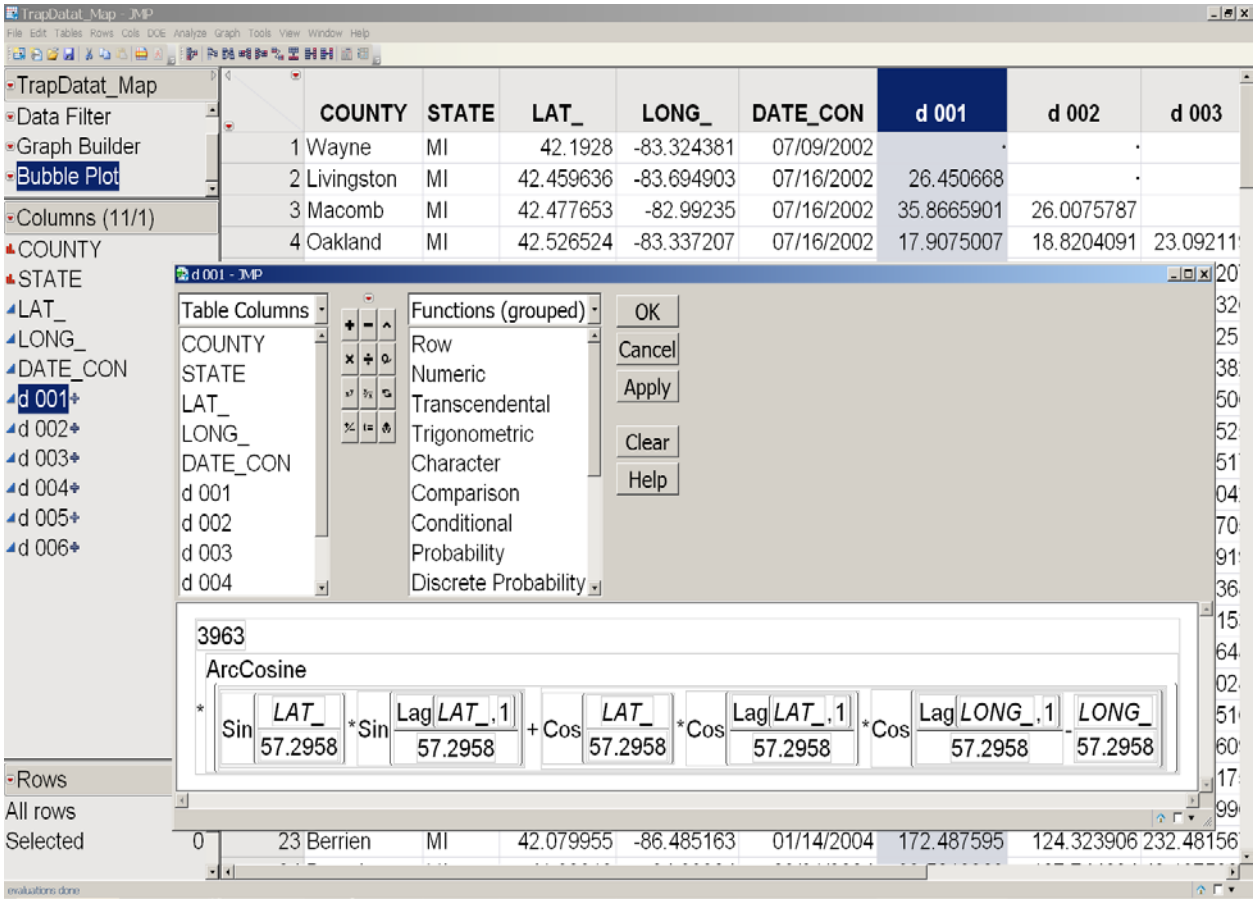
The new simulation has results in the expected range as follows:



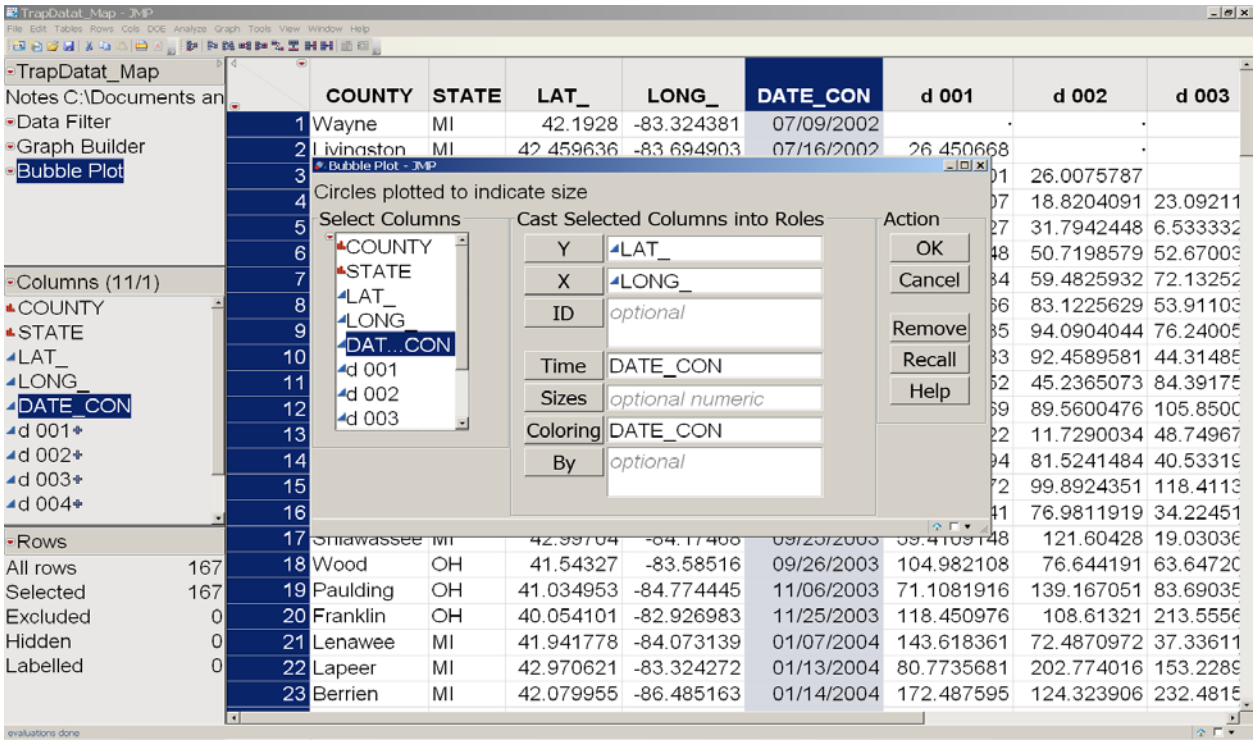
Pest Movement Mapping and Modeling

The NPPO tracks exotic pest movement within the United States in conjunction with eradication efforts. Putting a point on a map provides a visual image of pest locations. This has proven to be a very effective method of putting the pest in context relative to population centers, commercial agriculture and natural resources. In some pest programs the date and location of each new interception is recorded and the data is mapped using map based software. The introduction of JMP 9 has allowed pest mapping in JMP plus the maps the animated of pest movement.

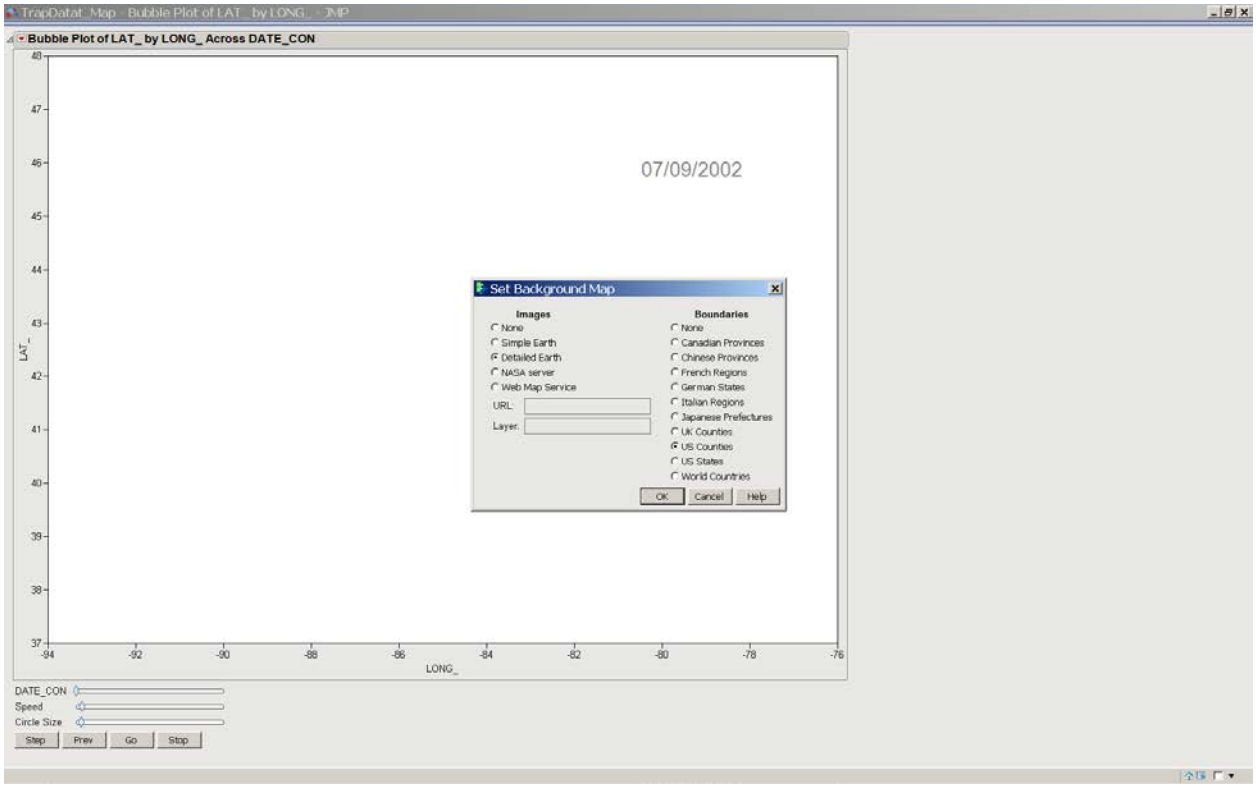
The data set for Emerald Ash Borer (EAB) is provided below.



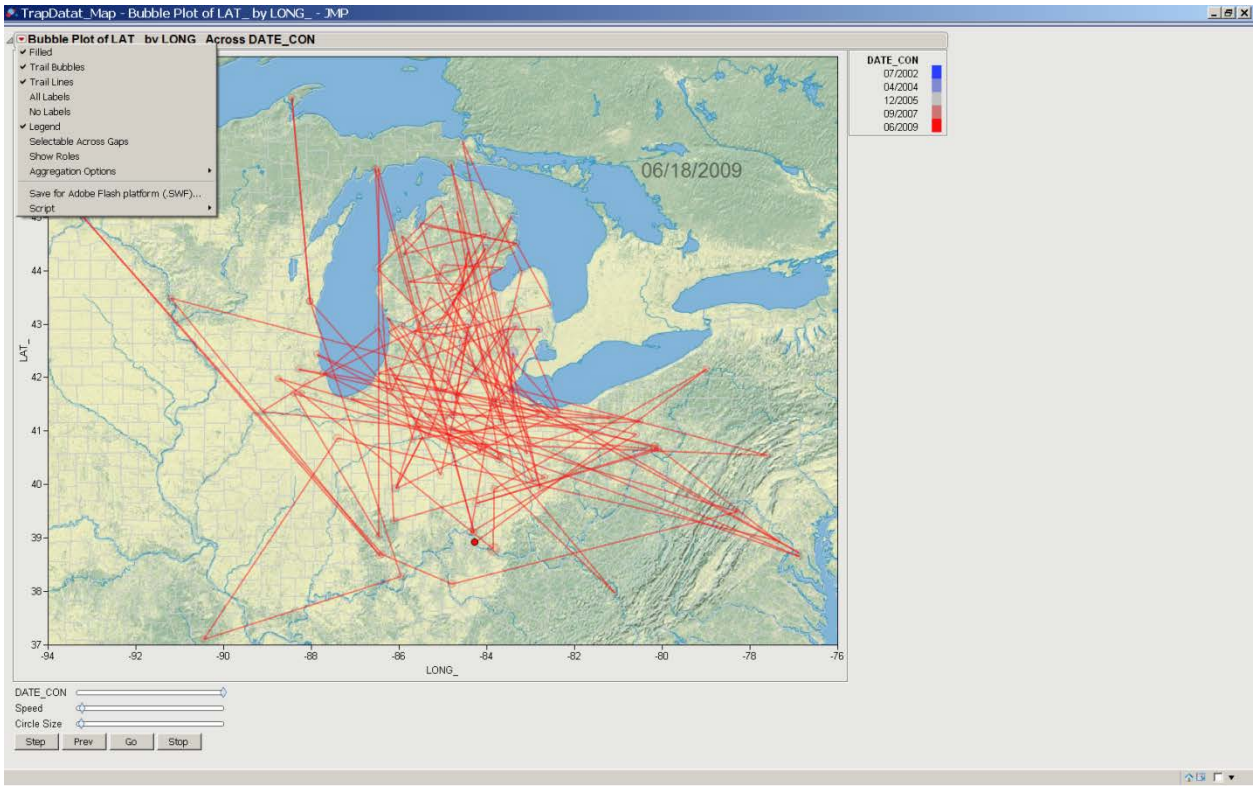
It includes county, state, latitude, longitude and date. The added variables are the distances between interceptions in miles. These distances were calculated using great circle navigation. The formula is displayed. The most obvious use of this data is to make a map of the interceptions. Maps can be made in several of the graph options. Using the bubble plot allow the inclusion of motion or animation in the maps. To create the bubble plot click on the graph button and select bubble plot. In the bubble plot build window move ‘Lat_’ to ‘Y’, ‘Long’ to ‘X’ and ‘Date_CON’ to ‘Time’ and optionally to ‘Color’ then clock ‘OK’.



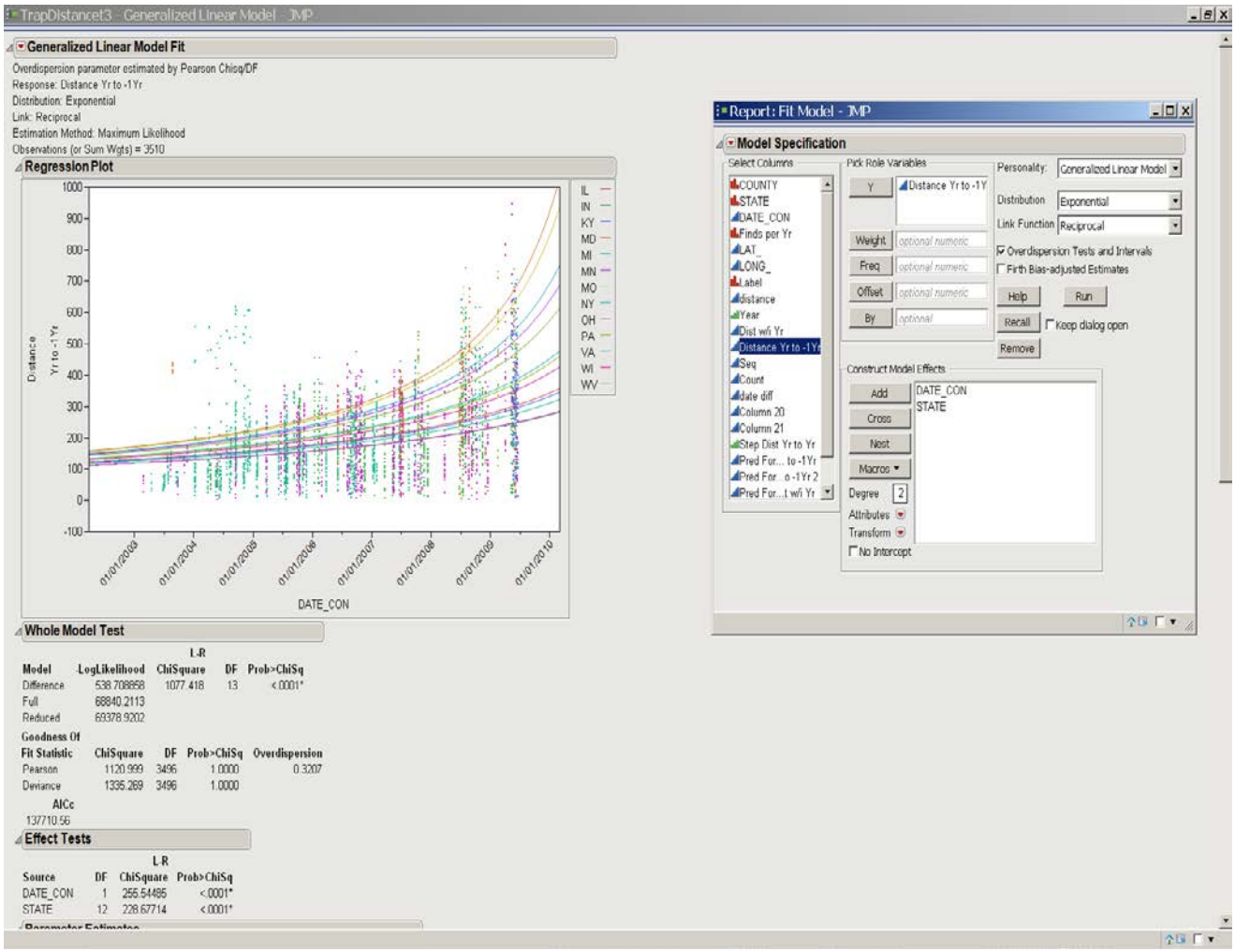
The result is as follows:



Right click on the graph and choose ‘Background Map’. The set background map window opens. Select either or both an image option (Detailed Earth) and Boundaries (US Counties), and click ‘OK’. The result is as follows:

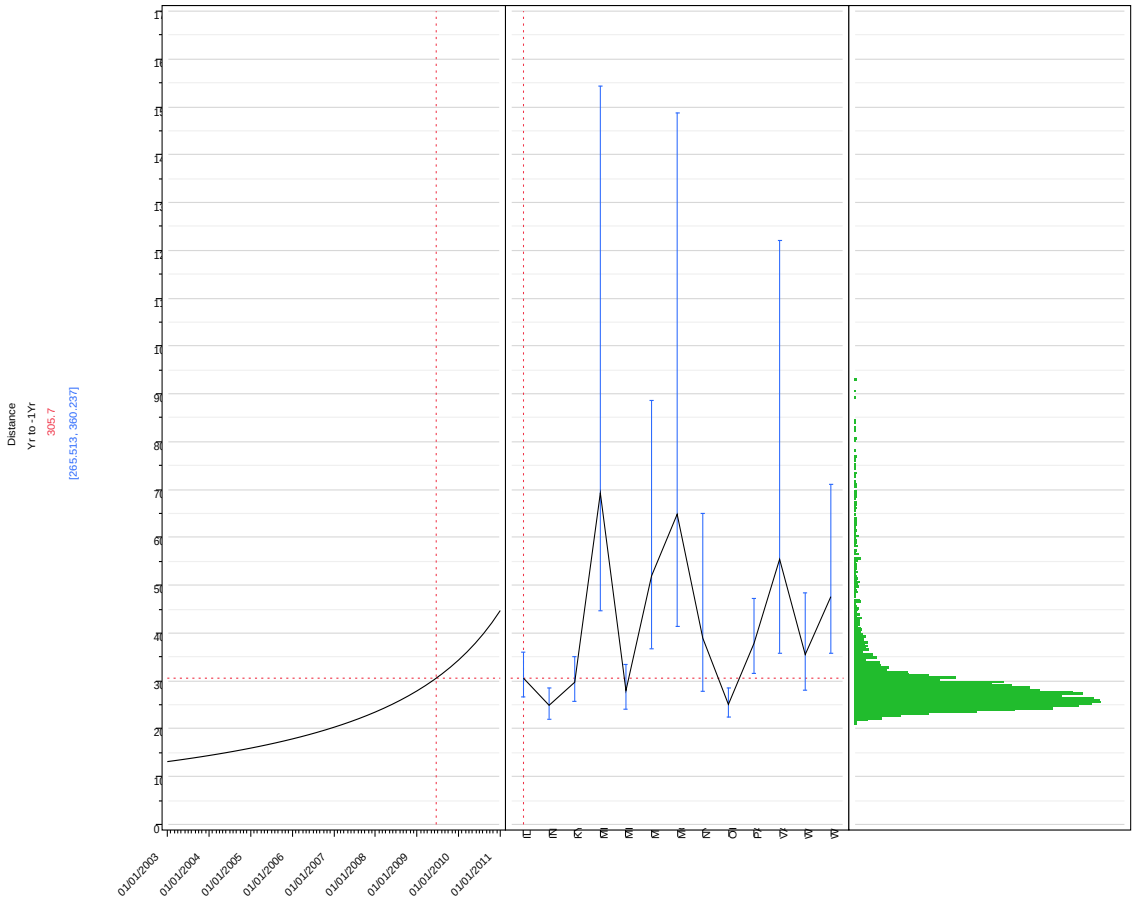


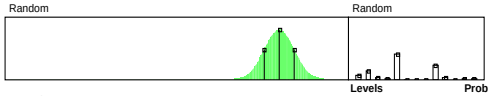
One question to be answered was how far from the previous year’s interceptions should surveyors search for EAB. A model was built to predict the movement of EAB from year to year. The original data set was manipulated so that each interception observation could be compared with every observation in successive years. A variable was set-up that calculated the distances from the current year’s interceptions to the previous year’s interceptions. Then a generalized linear model was run on the distances based on year and state. Since the distances were all positive the exponential distribution and a reciprocal link function were used based on recommendations in JMP® 9 Modeling and Multivariate Methods (SAS_Institute_Inc. 2010). The model was set up in Fit Model and model results were as follows:



The model was significant with good fit. The effects tests were significant for both year and state. The parameter estimates were significant with the exception of NY, PA and WV. The profiler with simulator was created and the simulator was set up with the states random probability proportional to the number of observations for each state and the year was moved out to the end of the regression lines (June-July 2008) and the distribution was visually adjusted to cover a little more than a year. The results are shown below.

Next the data were simulated to a JMP table. Remember the question to answer was how far out from last year's interceptions should we be looking for EAB. The year to year distances distribution was analyzed and the results are provided below. The upper 95% tolerance limit was estimated to include 95% of the distances and 99%, the distances were 431 miles and 487 miles respectively. The fix distribution all was run on the data. The best fit was a Normal 3 mix with means and sigmas of 268 (23), 342 (51) and 576(167). The probabilities of the mix were 0.82, 0.13 and 0.05. After a presentation of the analysis and the results it was decided to increase the distance around last year's interceptions to 450 miles.

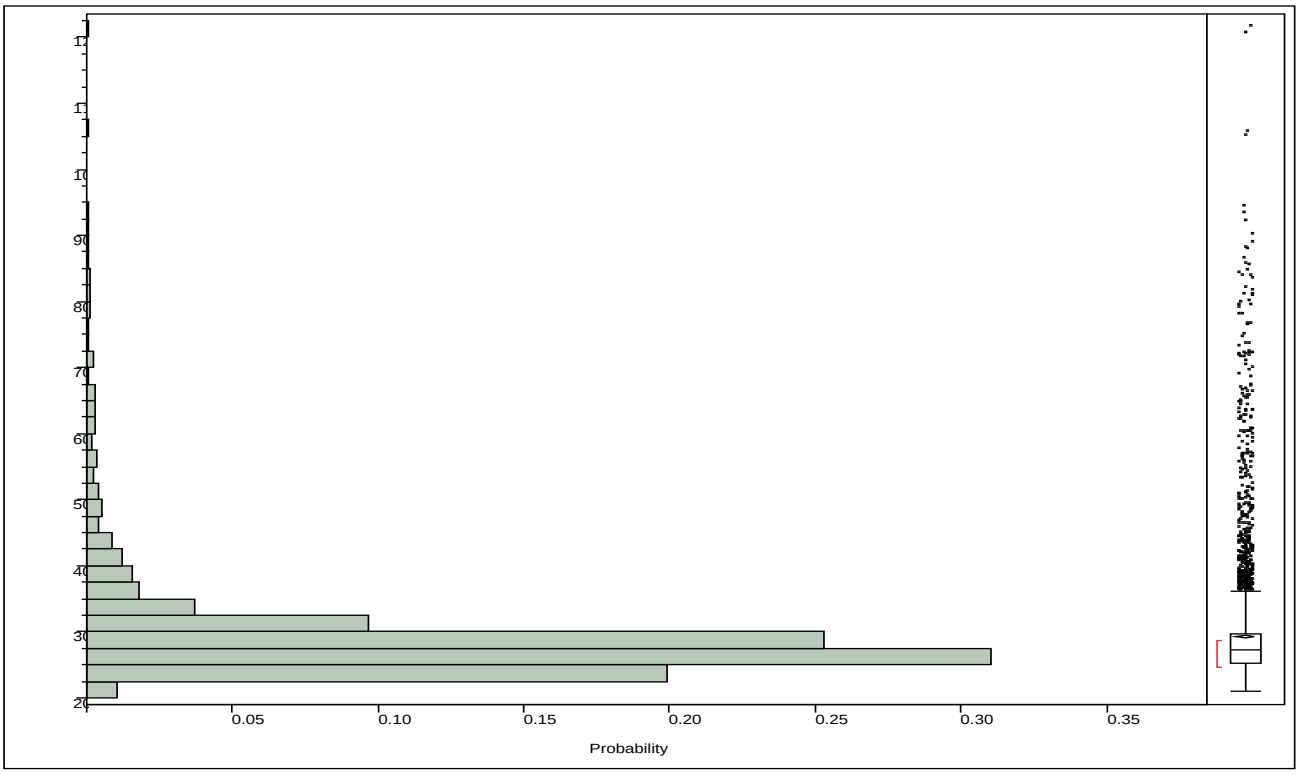




Normal

Mean	3.33e+9
SD	1.13e+7

Levels	Prob
IL	0.06587
IN	0.13772
KY	0.04192
MD	0.01198
MI	0.41916
MN	0.00599
MO	0.00599
NY	0.00599
OH	0.23353
PA	0.03593
VA	0.00599
WI	0.01796
WV	0.01198



Normal 3 Mixture

One-sided Tolerance Interval

Proportion	Lower TI	Upper TI	1-Alpha
0.950	.	430.542	0.950

One-sided Tolerance Interval

Proportion	Lower TI	Upper TI	1-Alpha
0.990	.	487.3245	0.950

Show

Bennett, C. A., W. M. Bowen, et al. (1988). Statistical Methods for Nuclear Material Management.

Cnnon, R. M. and R. T. Roe (1982). Livestock Disease Surveys: A Field Manual for Veterinarians. C. A. G. P. Serv.**:** 35.

Cochran, W. G. (1977). Sampling Techniques, WILEY.

Couey, H. M. and V. Chew (1986). "Confidence Limits and Sample Size in Quarantine Research." Journal of Economic Entomology **79**(4): 887-890.

Gould, W. P. (1995). "PROBABILITY OF DETECTING CARIBBEAN FRUIT FLY (DIPTERA: TEPHRITIDAE) INFESTATIONS BY FRUIT DISSECTION." Florida Entomologist **78**(3): 502-507.

Hahn, G. J. and W. Q. Meeker (1991). Statistical Intervals: A Guide for Practitioners John Wiley & Sons.

JMP_Technical_Support and L. Archer (2011). [SAS 7610638494] Follow up questions to 7610612348. E. M. Jones. Raleigh, NC.

JMP_Technical_Support and L. Archer (2011). [SAS 7610612348] how to with hyper geometric function in JMP. E. M. Jones. Raleigh, NC.

NAPPO, N. A. P. P. O. (2011). General Guidelines for Pathway Risk Analysis (Draft).

SAS_Institute_Inc. (2010). JMP® 9 Modeling and Multivariate Methods. Cary, NC, SAS Institute Inc.

SAS_Institute_Inc. (2011). JMP.

Schuler, C. (2010). Supvy Sr RPM AQI, PPQ, APHIS. E. M. Jones. Raleigh, NC.

Venette, R. C., R. D. Moon, et al. (2002). "STRATEGIES AND STATISTICS OF SAMPLING. FOR RARE INDIVIDUAL." Annu. Rev Entomology **47**: 143-174.

Vose, D. (2008). Risk Analysis: A Quantitative Guide West Essex, England, John Wiley and Sons.

Wilks, S. S. (1966). Elementary Statistical Analysis. Princeton, NJ, Princeton University Press.

Wilson, J. E. (2011). Visual Salple Plan, Battelle Memorial Instutute.

Appendix A

A one-sided distribution-free tolerance bound Is equivalent to a one-sided distribution-free confidence bound for a percentile of that population. That is, a one-sided distribution-free lower (upper) 100(1-α)% tolerance bound that will be exceeded by (that will exceed) at lease 100p% of the population is the same as a distribution-free lower (upper) 100(1-α)% confidence bound for the 100pth percentile of the population(Hahn and Meeker 1991; Wilson 2011).

The commodity population parameter of interest is the true P^{th} percentile of the commodity population of infested units, where $0 < P < 100$. The true P^{th} percentile is the value above which (100 - P)% of the population lies and below which P% of the population lies. The objective is to reject the null hypothesis if the true P^{th} percentile exceeds the specified action level (AL). But, the true P^{th} percentile will never be known with 100% confidence because all possible measurements from the population cannot be obtained. Hence the decision whether to reject the null hypothesis is made using the computed upper tolerance limit (UTL) for the P^{th} percentile, that is, by computing the upper 100(1-α)% confidence limit on the P^{th} percentile (see Decision Rule below). A design with an α of 0.05 and a P^{th} percentile of 0.01 would mean that the decision will be made using the computed UTL for the 95% confidence limit on the 99th percentile.

Hypothesis Being Tested is as follows:

The null hypothesis (baseline assumption) is as follows:

H_o: The true P^{th} percentile $\leq$ AL

or equivalently,

H_o: Less than P% of the population < AL

The H_o is rejected if UTL < AL, in which case the alternative hypothesis (H_a) is accepted as being true, where:

H_a: More than P% of the population < AL

The action level (AL) in most commodity sampling is 0.

The sample designs require that samples are selected either using simple random sampling (SRS), or systematic sampling with a random start location to determine the commodity plant units for inspection or samples are collected and subsequently measured.

Decision Rule and Number of Samples, n are determined as follows:

The null hypothesis is rejected and the alternative hypothesis is accepted if the nonparametric (distribution-free) UTL for the P^{th} percentile is less than the specified action level (AL). The nonparametric UTL is simply the maximum of the n measurements obtained from the population of interest, where n is computed using the assumed probability distribution.

Appendix B

Hypergeometric sample size without sensitivity adjustment

The following JMP script requires a population variable N and a sample detection variable P.

```
t1 = :N; While(Hypergeometric Distribution(:N, :N * :P, t1, 0) < 0.05, t1 -= 1); If(Ceiling(t1 += 1) >= :N, :N,
    Ceiling(t1))
```

Hypergeometric sample size with sensitivity adjustment

The following JMP script requires a population variable N, a sample detection variable P and inspection sensitivity.

```
t1 = :N; While(Hypergeometric Distribution(:N, :N * :P, t1, 0) < 0.05, t1 -= 1); If(Ceiling((t1 += 1) / :SE) >=
    :N, :N, Ceiling(t1 / :SE))
```

Appendix C
Email from JMP Technical Support

Hi Ned,

It seems that JMP consistently takes the floor of non-integer vales EXCEPT in the negative binomial. Please see below for the full breakdown:

Gamma Poisson - for non integer x JMP takes the floor of x, so Gamma Poisson Probability(6, 2, 2) = Gamma Poisson Probability(6.01, 2, 2) = Gamma Poisson Probability(6.9, 2, 2)

Binomial - for non integer k JMP takes the floor of k. This is also true for sample size N - JMP will take the floor of a non-integer N

neg binomial - for non-integer k JMP takes the floor of k. However for sample size n seems to change with all values so

neg binomial distribution(.4, 20, 15);

neg binomial distribution(.4, 20, 15.2); neg binomial distribution(.4, 20, 15.8); all produce the same value, but:

neg binomial distribution(.4, 20, 15);

neg binomial distribution(.4, 20.7, 15); neg binomial distribution(.4, 20.2, 15); all produce different values. This actually seems like a bug - I am reporting it to development

Beta Binomial - for non-integer x JMP takes the floor of x. This is also true for sample size n - JMP takes the floor for non-integer n.

Hypergeometric - JMP takes the floor for any non-integer value (N, K, n, x)

Poisson - for non-integer x JMP takes the floor of x

I hope this information is helpful. Let me know if you have follow-up questions, or need clarification.

I will be happy to continue working with you should you have any follow-up questions regarding this matter simply reply to this email within 5 business days. If this response has fully answered your question, there is no need to reply although I encourage you to respond to let me know if I have resolved this issue to your satisfaction.

Best regards, Laura Archer

JMP Technical Support

SAS® … THE POWER TO KNOW

Appendix D

Hi Ned,

I heard back from the developer and he said the following

"This is happening because we are using the Gamma function to compute factorials (and combinations) in the Hypergeometric probability functions. When k is an integer

k! = Gamma(k+1) where Gamma is as defined here http://en.wikipedia.org/wiki/Gamma_function

I would expect the hypergeometric probability function to only take integer values. If the user is using these values, that's a little concerning and they should be careful how they use the results. The values that they are getting can't really be interpreted as probabilities."

So basically it does look like they are aware that the values can be noninteger and the profile plot appears smooth, but they do not intend for non-integer values to be used. Does this answer your question? Give you the information you need? Please let me know.

Best regards,

Laura Archer
JMP Technical Support
SAS® … THE POWER TO KNOW®