

Telephone No: +61 8 8207 7884

5 December 2006

Catherine Bass
NSW Food Authority
6 Avenue of the Americas
Newington NSW 2127

Dear Catherine

Seed Sampling and Testing

As per our discussion, please find enclosed a copy of the statistical considerations with respect to seed sampling and testing.

Please note that the document refers to the ISS Seed Screening Procedures, which were previously provided by you.

Please don't hesitate to contact me directly should you need clarification or further information on the statistical aspects discussed in the attached document.

Yours sincerely



Andreas Kiermeier
SENIOR STATISTICIAN, SARDI, PIRSA

enc: Seed Sampling and Testing

SARDI



**SOUTH AUSTRALIAN
RESEARCH AND
DEVELOPMENT
INSTITUTE**

SARDI Food Safety
33 Flemington Street
Glenside SA 5065
Correspondence

Telephone
(08) 8207 7886
Facsimile
(08) 8207 7854

International Code +618

**Developing South
Australian Research
and Innovation**



Seed Sampling and Testing

Probability of Capture

The probability of capture, *i.e.* selecting at least one contaminated seed in the sample, assuming a lot size of 800 × 25 kg bags, was calculated for a range of sample sizes and contamination levels. The resulting probabilities, calculated using Binomial and Hypergeometric probability formulae, are displayed in Table 1 and 2, respectively. From these Tables it can be seen that the difference between the two distributions is negligible, and hence either could be used. Probabilities increase as the sample size increases and as the contamination level increases. These probabilities are in agreement with those presented in Table 4 of the *ISS Seed Screening Procedures*.

Tables 1 and 2 were calculated based on a seed weight of 0.000002 kg, as per *ISS Seed Screening Procedures*. However, the same numbers also apply to much larger seeds such as Mung Bean (0.0000625 kg).

Minimum Sampling Requirements

In order to determine the appropriate sampling requirements, a decision has to be made as to what probability of capture, and hence detection, will be required. While no such probability has been explicitly stated in the *ISS Seeds Sampling Procedures*, an explicit level of >99.9999% (based on green shading in Table 4 of the *ISS Seeds Sampling Procedures*) is given. In addition, the recommendation is made that either 3 kg or 1/1000th of the lot size is to be sampled, whichever is greater.

It should be noted that the rule of 1/1000th of the lot size requires larger lots to be sampled more intensively, maintaining the same rate. The Australian Standard AS 1199 also requires larger lots to be sampled more intensively, with respect to total sample size, but allows also for the sampled fraction to be reduced as the lot size increase.

In Table 3 are displayed the capture probabilities for various sample sizes for several lot sizes – 1 and 800 bags, as well as 480 bags of alfalfa and 1152 bags of mung bean seeds (based on industry information) – assuming a minimum level of contamination of 4 contaminated seeds per kg. From this table it can be seen that at this level of contamination there is little advantage, with respect to capture probability, in taking samples larger than 3 kg from a lots exceeding 120 bags (equivalent to 3000 kg), especially when considering the microbiological tests which follow capture. However, advantages could be gained at lower levels of contamination.





Statistical Justifications for Capture Probabilities

There are three main distributions used in sampling problems. These are

- **Hypergeometric distribution:** Used when the sample is large relative to the population (greater than 10%) and particularly when the population is small. There are only two possible outcomes for each unit examined.
- **Binomial distribution:** Used when the sample is small relative to the population and there are only two possible outcomes for each unit examined.
- **Poisson distribution:** Used when sampling units from a population and counting the number of “non-conformities” each unit has.

The Poisson distribution is not applicable for seed sampling as each seed can only have two possible outcomes – contaminated and not contaminated. If a count of the number of organisms was performed.¹

The main use for the Hypergeometric distribution is when subsequent units (seeds) sampled from population (lot) are not independent and hence the probability changes. This is strictly speaking the case – removing a contaminated seed reduces the probability of the next seed being contaminated. However, as indicated in the *ISS Seed Screening Procedures* the effect of this change in probabilities given the very large population of seeds is of negligible effect.

The Binomial distribution can also often be used as a very good approximation to the Hypergeometric distribution. The Binomial (and Hypergeometric) distribution allow calculation of the probability of “capturing” X contaminated seeds in the sample.

The assumptions for Binomial probability calculations to hold are:

1. Fixed number of seeds sampled.
2. Seeds are statistically independent.
3. Only two possible outcomes per seed – contaminated and not contaminated.
4. The probability of a seed being contaminated is the same for each seed.

¹ The Poisson distribution is commonly used in manufacturing environments where the total number of defects or non-conformities are recorded per unit inspected, e.g. 3 scratches and a printing misalignment.



These assumptions are all (approximately) met as discussed below.

1. The number of seeds to be sampled is fixed prior to enrichment, *i.e.* seeds are removed and then enriched.²
2. Statistical independence in this case implies that the probability of selecting a contaminated seeds is not affected by how few or many contaminated seeds have already been selected. While this is not strictly speaking the case, the effect of this violation is minimal and of no practical importance.
3. This assumption is met since seeds are either contaminated or they are not contaminated with the organism/pathogen of interest.
4. This assumption is related to point 2.

The probability of capture is equal to the probability of selecting at least one contaminated seed in the sample.

Calculation of the Probability of Detection

Detection of contamination using the *ISS Seeds Screening Procedures* is a two step process. Firstly, contaminated seed has to be captured, and secondly, the microbiological test has to detect the contamination (given that contaminated seed has been selected). Consequently, the probability of detecting contamination can be obtained by multiplying the probability of capture and the sensitivity of the microbiological test.

For example, consider a microbiological test for a pathogen with sensitivity of 0.9. Then, assuming the probability of capturing contaminated seeds is 0.99, the probability of detection of contamination is $0.99 \times 0.9 = 0.891$. The interpretation is that 89.1% of all contaminated lots will be identified as contaminated.

² In contrast, different distributional assumption would need to be made if sampling was performed *until* a certain number of contaminated seeds are captured. This scenario is more common on manufacturing situations, where each part can be fully inspected for compliance directly after sampling.

Table 1: Probability of capturing, expressed as a percentage, contaminated seed in a sample of given weight/number of seeds when the lot (800 x 25 kg bags) is contaminated at a given level of contamination, assuming Binomial distribution assumptions hold.

		Level of Contamination (Contaminated Seeds per kg)									
Sample Weight	Seeds Sampled	0.1	1	2	3	4	5	10	100		
0.00002	1	0.000020	0.000200	0.000400	0.000600	0.000800	0.001000	0.002000	0.020000		
0.00002	10	0.000200	0.002000	0.004000	0.006000	0.008000	0.010000	0.019998	0.199820		
0.0002	100	0.002000	0.019998	0.039992	0.059982	0.079968	0.099951	0.199802	1.980329		
0.002	1,000	0.019998	0.199800	0.399202	0.598205	0.796812	0.995022	1.980152	18.128562		
0.02	10,000	0.199800	1.980135	3.921064	5.823564	7.688395	9.516303	18.127088	86.469178		
0.2	100,000	1.980133	18.126941	32.968049	45.118935	55.067247	63.212240	86.466742	100.000000		
1	500,000	9.516259	63.212093	86.466526	95.021338	98.168465	99.326222	99.995460	100.000000		
2	1,000,000	18.126926	86.466499	98.168451	99.752129	99.966455	99.995460	100.000000	100.000000		
3	1,500,000	25.918180	95.021308	99.752128	99.987659	99.999386	99.999969	100.000000	100.000000		
5	2,500,000	39.346937	99.326209	99.995460	99.999969	100.000000	100.000000	100.000000	100.000000		
10	5,000,000	63.212060	99.995460	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000		
12	6,000,000	69.880582	99.999386	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000		
25	12,500,000	91.791502	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000		
50	25,000,000	99.326206	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000		
100	50,000,000	99.995460	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000		





Table 2: Probability of capturing, expressed as a percentage, contaminated seed in a sample of given weight/number of seeds when the lot (800 x 25 kg bags) is contaminated at a given level of contamination, assuming Hypergeometric distribution assumptions hold.

		Level of Contamination (Contaminated Seeds per kg)									
Sample Weight	Seeds Sampled	0.1	1	2	3	4	5	10	100		
0.000002	1	0.000020	0.000200	0.000400	0.000600	0.000800	0.001000	0.002000	0.020000		
0.00002	10	0.000200	0.002000	0.004000	0.006000	0.008000	0.010000	0.019998	0.199820		
0.0002	100	0.002000	0.019998	0.039992	0.059982	0.079968	0.099951	0.199802	1.980329		
0.002	1,000	0.019998	0.199800	0.399202	0.598205	0.796812	0.995022	1.980152	18.128563		
0.02	10,000	0.199800	1.980136	3.921066	5.823566	7.688399	9.516308	18.127097	86.469192		
0.2	100,000	1.980143	18.127023	32.968183	45.119100	55.067427	63.212424	86.466878	100.000000		
1	500,000	9.516485	63.213012	86.467202	95.021711	98.168649	99.326306	99.995462	100.000000		
2	1,000,000	18.127745	86.467852	98.168817	99.752204	99.966468	99.995463	100.000000	100.000000		
3	1,500,000	25.919847	95.022428	99.752239	99.987668	99.999386	99.999969	100.000000	100.000000		
5	2,500,000	39.350728	99.326630	99.995466	99.999969	100.000000	100.000000	100.000000	100.000000		
10	5,000,000	63.221258	99.995471	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000		
12	6,000,000	69.891428	99.999388	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000		
25	12,500,000	91.804329	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000		
50	25,000,000	99.330411	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000		
100	50,000,000	99.995572	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000	100.000000		



Table 3: Effect of different sample sizes and lot sizes on the probability of capture, assuming a constant minimum level of contamination. For each lot size greater than or equal to 120 bags, the left-most column relates to the minimum of 3 kg requirement; the right-most column relates to a sample size of 1/1000th of the lot size.

25 kg Bags per lot	1	1	120	800	800	480	480	1152	1152
Lot size (kg)	25	25	20000	20000	20000	12000	12000	28800	28800
Seed Type	Alfalfa	Alfalfa	Alfalfa	Alfalfa	Alfalfa	Alfalfa	Alfalfa	Mung Bean	Mung Bean
Seed Weight (g)	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.0625	0.0625
Contaminated Seeds per kg	4	4	4	4	4	4	4	4	4
Contaminated Seeds per lot	100	100	12000	80000	80000	48000	48000	115200	115200
Seeds sampled (kg)	0.025	1.5	3	3	20	3	12	3	28.8
Probability of capture (%)	9.5163	99.7521	99.9994	99.9994	100.0000	99.9994	100.0000	99.9994	100.0000

