APPENDIX IV
CALCULATION FORMULAS FOR STATISTICAL EVALUATION

Appendix IV provides calculation formulas to enable responsible risk assessment personnel to determine the minimum number of samples necessary to meet statistical performance objectives. This appendix also provides statistical guidelines on the probability that a given sampling plan will identify a hot spot, and the probability that no hot spot exists given none was found after sampling.

Calculation Formulas to Determine the Number of Samples Required Given Coefficient of Variation and Statistical Performance Objectives

The minimum number of samples, n, required to achieve a specified precision and confidence level at a defined minimum detectable relative difference may be estimated by the following equation:

For one-sided, one-sample t-test $n \geq [(Z_\alpha + Z_\beta)/D]^2 + 0.5Z^2_\alpha$

For one-sided, two-sample t-test $n \geq 2\,[(Z_\alpha + Z_\beta)/D]^2 + 0.5Z^2_\alpha$

where: Z_α is a percentile of the standard normal distribution such that $P(Z \geq Z_\alpha) = \alpha$, Z_β is similarly defined, and D = MDRD/CV, where MDRD is the minimum detectable relative difference and CV is the coefficient of variation. NOTE: Data must be transformed (Z_α), for example:

Confidence Level			Power		
$1-\alpha$	α	Z_α	$1-\beta$	β	Z_β
0.80	0.20	0.842	0.80	2.00	0.842
0.85	0.15	1.039	0.85	0.15	1.039
0.90	0.10	1.282	0.90	0.10	1.282
0.95	0.05	1.645	0.95	0.05	1.645
0.99	0.01	2.326	0.99	0.01	2.326

As an example of applying the equation above, assume CV = 30%, Confidence Level = 80%, Power = 95%, and Minimum Detectable Relative Difference = 20%. For infinite degrees of freedom (t distribution becomes a normal one), $Z_\alpha = 0.842$ and $Z_\beta = 1.645$. From the data assumed, D = 20% /30%. Therefore,

$$n \geq [(0.842 + 1.645)/(20/30)]^2 + 0.5\,(0.842)^2$$

$$n \geq 13.917 + 0.354 = 14.269$$

$n \geq 15$ samples required (round up)

Source: Adapted from EPA 1989c.

APPENDIX IV
(continued)

Calculation Formulas For The Statistical Evaluation Of The Detection Of Hot Spots

Hot Spot Will Be Identified: Example # 1

These formulas are useful in evaluating the probability that a particular sampling plan will identify a hot spot. Let R represent the radius of a hot spot and D be the distance between adjacent grid points where samples will be collected. The probability that a grid point will fall on a hot spot is easily obtained from a geometrical argument since at least one grid point must fall in any square of area D^2 centered at the center of the hot spot. From this concept, it follows that the probability of sampling a hot spot P(H/E) is given by:

$$P(H/E) = (\pi R^2)/D^2 \qquad \text{if } R \leq D/2$$

$$= \{R^2[\pi - 2 \arccos (D/(2R))] + (D/4)\sqrt{(4R^2 - D^2)}\}/D^2 \qquad \text{if } D/2 < R < D/\sqrt{2}$$

$$= 1 \qquad \text{if } R \geq D/\sqrt{2}$$

where the angle D/(2R) is expressed in radian measure, H is the case that a hot spot is found, and E is the case that a hot spot exists.

An example is if the grid spacing is $D = 2R$, then the probability of a hit is $\pi/4 = 0.785$, which implies that the probability that this grid spacing would not hit a hot spot if it exists is 0.215.

No Hot Spot Exists: Example # 2

This set of formulas addresses the probability that no hot spot exists (given that none were found). This argument requires the use of a subjective probability, P(E) (where P(E) is the probability that a hot spot exists), based on historical and perhaps geophysical evidence. Then, if E is the case that there are no hot spots at the study site and if H is the case that no hot spot is found in the sample, Bayes formula gives:

$$P(E \mid \bar{H}) = P(\bar{H} \mid E)\, P(E) \,/\, [P(\bar{H} \mid E)\, P(E) + P(\bar{H} \mid \bar{E})\, P(\bar{E})]$$

$$= P(\bar{H} \mid E)\, P(E) \,/\, [P(\bar{H} \mid E)\, P(E) + P(\bar{E})].$$

For the case where D = 2R, it was found from Example 1 that P(H|E) =0.215. Therefore, if one is given that the chance P(E) of a hot spot is thought to be 0.25 prior to the investigation, the probability of a hot spot existing if the study does not find one is:

$$P(E \mid \text{no hit}) = 0.215\,(0.25) \,/\, [0.215\,(0.25) + 0.75] = 0.067.$$

Hence, the probability that no hot spot exists is (1-0.067) = 0.933.

Source: Adapted from EPA 1989c.

Appendix IV (continued)
Number of Samples Required in a One-Sided One-Sample t-Test to Achieve a Minimum Detectable Relative Difference at Confidence Level (1-α) and Power of (1-β)

Coefficient of Variation (%)	Power (%)	Confidence Level (%)	Minimum Detectable Relative Difference (%)				
			5	10	20	30	40
10	95	99	66	19	7	5	4
		95	45	13	5	3	3
		90	36	10	3	2	2
		80	26	7	2	2	1
	90	99	55	16	6	5	4
		95	36	10	4	3	2
		90	28	8	3	2	2
		80	19	5	2	1	1
	80	99	43	13	6	4	4
		95	27	8	3	3	2
		90	19	6	2	2	2
		80	12	4	2	1	1
15	95	99	145	39	12	7	5
		95	99	26	8	5	3
		90	78	21	6	3	3
		80	57	15	4	2	2
	90	99	120	32	11	6	5
		95	79	21	7	4	3
		90	60	16	5	3	2
		80	41	11	3	2	1
	80	99	94	26	9	6	5
		95	58	16	5	3	3
		90	42	11	4	2	2
		80	26	7	2	2	1
20	95	99	256	66	19	10	7
		95	175	45	13	9	5
		90	138	36	10	5	3
		80	100	26	7	4	2
	90	99	211	55	16	9	6
		95	139	36	10	6	4
		90	107	28	8	4	3
		80	73	19	5	3	2
	80	99	164	43	13	8	6
		95	101	27	8	5	3
		90	73	19	6	3	2
		80	46	12	4	2	2

Source: EPA 1989c

B21-002-80,1

Appendix IV (continued)
Number of Samples Required in a One-Sided One-Sample t-Test to Achieve a Minimum Detectable Relative Difference at Confidence Level (1-α) and Power of (1-β)
(continued)

Coefficient of Variation (%)	Power (%)	Confidence Level (%)	Minimum Detectable Relative Difference (%)				
			5	10	20	30	40
25	95	99	397	102	28	14	9
		95	272	69	19	9	6
		90	216	55	15	7	5
		80	155	40	11	5	3
	90	99	329	85	24	12	8
		95	272	70	19	9	6
		90	166	42	12	6	4
		80	114	29	8	4	3
	80	99	254	66	19	10	7
		95	156	41	12	6	4
		90	114	30	8	4	3
		80	72	19	5	3	2
30	95	99	571	145	39	19	12
		95	391	99	26	13	8
		90	310	78	21	10	6
		80	223	57	15	7	4
	90	99	472	120	32	16	11
		95	310	79	21	10	7
		90	238	61	16	8	5
		80	163	41	11	5	3
	80	99	364	84	26	13	9
		95	224	58	16	8	5
		90	164	42	11	6	4
		80	103	26	7	4	2
35	95	99	775	196	42	25	15
		95	532	134	35	17	10
		90	421	106	28	13	8
		80	304	77	20	9	6
	90	99	641	163	43	21	13
		95	421	107	28	14	8
		90	323	82	21	10	6
		80	222	56	15	7	4
	80	99	495	126	34	17	11
		95	305	78	21	10	7
		90	222	57	15	7	5
		80	140	36	10	5	3

Source: EPA 1989C

B21-002-80,2