SAFOSO

Animal Health Matters.
For Safe Food Solutions.

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs,
Education and Research EAER
State Secretariat for Economic Affairs SECO

Sample size exercise

Marco De Nardi

SAFOSO, Switzerland

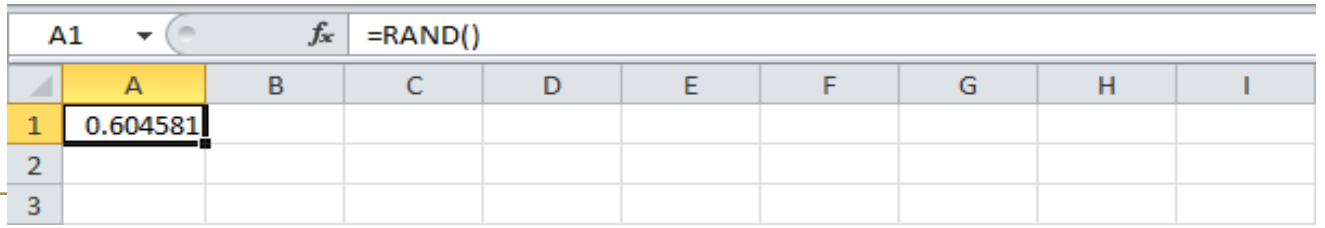
Generation of random numbers in Excel

- Excel has two useful functions when it comes to generating random numbers. The **RAND** and **RANDBETWEEN** function.

- **Rand**
- The RAND function generates a random decimal number between 0 and 1.

- 1. Select cell A1.

- 2. Type RAND() and press Enter. The RAND function takes no arguments.

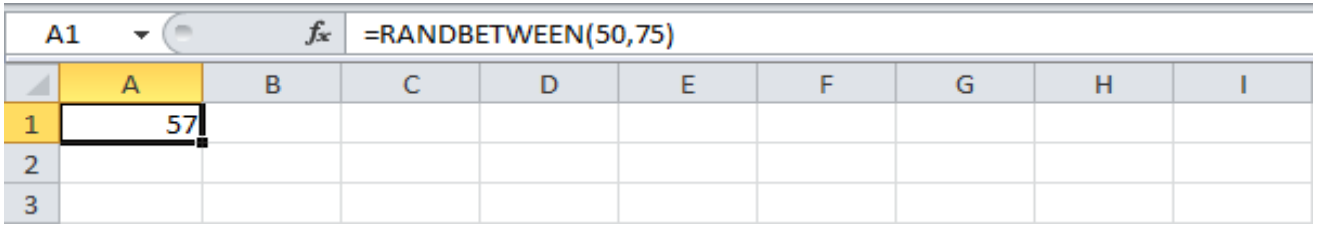


Generation of random numbers in Excel

- **Randbetween**
- The RANDBETWEEN function generates a random whole number between two boundaries.

- 1. Select cell A1.

- 2. Type RANDBETWEEN(50,75) and press Enter



Simple Random Sample (example)

- **Aim:** estimate prevalence of Bovine Brucellosis in a herd of 690 animals.
- Each animal has an eartag with an unique number (from 1 to 690)
- **Sample size**: 100 animals

- **To Do:**

1) Create a database with all 690 eartags

2) Generate a random number for each animal

3) You sample the 100 cattle with the smallest or largest random numbers

ID (eartag)	Random N.
56	1
255	2
145	3
458	3
195	4
690	5
61	6
226	6
253	9
461	10
284	14
291	16
613	16
609	17
156	18
344	18
394	18
15	20
125	20
278	20
87	21
246	21
464	21
505	22
590	22
188	26
432	26

Calculation of sample size- sample size table

Required Sample Size

Population Size	Confidence = 95% Margin of error 5.0%	3.5%	2.5%	1.0%	Confidence = 99% Margin of Error 5.0%	3.5%	2.5%	1.0%
10	10	10	10	10	10	10	10	10
20	19	20	20	20	19	20	20	20
30	28	29	29	30	29	29	30	30
50	44	47	48	50	47	48	49	50
75	63	69	72	74	67	71	73	75
100	80	89	94	99	87	93	96	99
150	108	126	137	148	122	135	142	149
200	132	160	177	196	154	174	186	198
250	152	190	215	244	182	211	229	246
300	169	217	251	291	207	246	270	295
400	146	265	318	384	250	309	348	391
500	217	306	377	475	285	365	421	485
600	234	340	432	565	315	416	490	579
700	248	370	481	653	341	462	554	672
800	260	396	526	739	363	503	615	763
1,000	278	440	606	906	399	575	727	943
1,200	291	474	674	1,067	427	636	827	1,119
1,500	306	515	759	1,297	460	712	959	1,376
2,000	322	563	869	1,655	498	808	1,141	1,785
2,500	333	597	952	1,984	524	879	1,288	2,173
3,500	346	641	1,068	2,565	558	977	1,510	2,890
5,000	357	678	1,176	3,288	586	1,066	1,734	3,842
7,500	365	710	1,275	4,211	610	1,147	1,960	5,165
10,000	370	727	1,332	4,899	622	1,193	2,098	6,239
25,000	378	760	1,448	6,939	646	1,285	2,399	9,972
50,000	381	772	1,491	8,056	655	1,318	2,520	12,455
75,000	382	776	1,506	8,514	658	1,330	2,563	13,583
100,000	383	778	1,513	8,762	659	1,336	2,585	14,227
250,000	384	782	1,527	9,248	662	1,347	2,626	15,555
500,000	384	783	1,532	9,423	663	1,350	2,640	16,055
1,000,000	384	783	1,534	9,512	663	1,352	2,647	16,317
2,500,000	384	783	1,536	9,567	663	1,353	2,651	16,478
10,000,000	384	784	1,536	9,594	663	1,354	2,653	16,560
100,000,000	384	784	1,537	9,603	663	1,354	2,654	16,584
300,000,000	384	784	1,537	9,603	663	1,354	2,654	16,586

Population size	Sample size required at expected prevalence rate of: 50%	40%	30%	25%	20%	15%	10%	5%	2%	1%	0.5%	0.1%
10	4	5	6	7	8	10	10	10	10	10	10	10
20	4	6	7	9	10	12	16	19	20	20	20	20
30	4	6	8	9	11	14	19	26	30	30	30	30
40	5	6	8	10	12	15	21	31	40	40	40	40
50	5	6	8	10	12	16	22	35	48	50	50	50
60	5	6	8	10	12	16	23	38	55	60	60	60
70	5	6	8	10	13	17	24	40	62	70	70	70
80	5	6	8	10	13	17	24	42	68	79	80	80
90	5	6	8	10	13	17	25	43	73	87	90	90
100	5	6	9	10	13	17	25	45	78	96	100	100
120	5	6	9	10	13	18	26	47	86	111	120	120
140	5	6	9	11	13	18	26	48	92	124	139	140
160	5	6	9	11	13	18	27	49	97	136	157	160
180	5	6	9	11	13	18	27	50	101	146	174	180
200	5	6	9	11	13	18	27	51	105	155	190	200
250	5	6	9	11	14	18	27	53	112	175	228	250
300	5	6	9	11	14	18	28	54	117	189	260	300
350	5	6	9	11	14	18	28	54	121	201	287	350
400	5	6	9	11	14	19	28	55	124	211	311	400
450	5	6	9	11	14	19	28	55	127	218	331	450
500	5	6	9	11	14	19	28	56	129	225	349	500
600	5	6	9	11	14	19	28	56	132	235	379	597
700	5	6	9	11	14	19	28	57	134	243	402	691
800	5	6	9	11	14	19	28	57	136	249	421	782
900	5	6	9	11	14	19	28	57	137	254	437	868
1000	5	6	9	11	14	19	29	57	138	258	450	950
1200	5	6	9	11	14	19	29	57	140	264	471	1102
1400	5	6	9	11	14	19	29	58	141	269	487	1236
1600	5	6	9	11	14	19	29	58	142	272	499	1354
1800	5	6	9	11	14	19	29	58	143	275	509	1459
2000	5	6	9	11	14	19	29	58	143	277	517	1553
3000	5	6	9	11	14	19	29	58	145	284	542	1895
4000	5	6	9	11	14	19	29	58	146	268	556	2108
5000	5	6	9	11	14	19	29	59	147	290	564	2253
6000	5	6	9	11	14	19	29	59	147	291	569	2358
7000	5	6	9	11	14	19	29	59	147	292	573	2437
8000	5	6	9	11	14	19	29	59	147	293	576	2498
9000	5	6	9	11	14	19	29	59	148	294	579	2548
10000	5	6	9	11	14	19	29	59	148	294	581	2588
ϕ	5	6	9	11	14	19	29	59	149	299	598	2995

The approximate sample size required to estimate disease prevalence in large populations.

Expected prevalence	Confidence level: 90%			95%			99%		
	Tolerable error			Tolerable error			Tolerable error		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
10%	24	97	2435	35	138	3457	60	239	5971
20%	43	173	4329	61	246	6147	106	425	10616
30%	57	227	5682	81	323	8067	139	557	13933
40%	65	260	6494	92	369	9220	159	637	15923
50%	68	271	6764	96	384	9604	166	663	16587
60%	65	260	6494	92	369	9220	159	637	15923
70%	57	227	5682	81	323	8067	139	557	13933
80%	43	173	4329	61	246	6147	106	425	10616
90%	24	97	2435	35	138	3457	60	239	5971

http://epitools.ausvet.com.au/content.php?page=SampleSize

Home

Language English

Epi Tools - Sample size calculations

These utilities can be used to calculate required sample sizes to estimate a population mean or proportion, to detect significant differences between two means or two proportions or to estimate a true herd-level prevalence.

Site Contents

Epidemiological studies

- To estimate a single proportion
- To estimate a single mean
- Two proportions
- Two means with equal sample size and equal variances
- Two means with unequal sample size and unequal variances
- To estimate true prevalence (at animal or herd-level)
- Sample size for a cohort study
- Sample size for a case-control study

Sample size to demonstrate disease freedom

- Sample size assuming perfect test specificity
- Sample size for pooled sampling in a large population
- Sample size to achieve target confidence of freedom
- Design prevalence required to achieve target population sensitivity for given sample size
- FreeCalc sample size calculation for imperfect tests

2-stage sampling, assuming perfect test specificity:

- Least-cost sample sizes where cluster sizes are known (and select clusters for testing).
- Least-cost sample sizes where cluster sizes are *NOT* known.
- Sample sizes for specified cluster sensitivity.

[Home | About this site | Glossary | References | Links]

AusVet Animal Health Services

This site was created by AusVet Animal Health Services with funding from the Australian Biosecurity Cooperative Research Centre. It provides a range of epidemiological tools for the use of researchers and epidemiologists, particularly in animal health. Please send any comments, questions or suggestions to Evan Sergeant

AUSTRALIAN BIOSECURITY CRC FOR EMERGING INFECTIOUS DISEASE

Sample size to estimate a proportion with specified precision

Input Values

Estimated true proportion : 0.5

Confidence level : 0.95

Desired precision (+/-) : 0.05

Population size (for finite populations) : 100000

Submit

This utility calculates the sample size required to estimate a proportion (prevalence) with a specified level of confidence and precision.

Inputs are the assumed true value for the proportion, the desired level of confidence, the desired precision of the estimate and the size of the population for limited population sizes. The desired precision of the estimate (also sometimes called the allowable or acceptable error in the estimate) is half the width of the desired confidence interval. For example if you would like the confidence interval width to be about 0.1 (10%) you would enter a precision of +/- 0.05 (5%).

The program outputs the sample sizes required to estimate the true value with the desired precision and confidence, for both an infinite population and for a population of the specified size. If population size is left blank or zero, only the sample size for an infinite population is calculated.

Sample size is calculated using the formula:

$n = (Z^2 \times P(1-P))/e^2$

where Z = value from standard normal distribution corresponding to desired confidence level (Z=1.96 for 95% CI)
P is expected true proportion
e is desired precision (half desired CI width).
For small populations n can be adjusted so that n(adj) = (Nxn)/(N+n)

Sample size to estimate a single proportion

Analysed: Thu Jun 30, 2016 @ 17:49

Inputs

Estimated Proportion	0.5
Confidence level	0.95
Desired precision of estimate	0.05
Population size	1e+05

Results

	Sample size
Infinite population	385
Population = 1e+05	384

AI Active Surveillance in **Poultry**

- Decision of the **European Commission** of 13 April 2007 on the implementation of **surveillance programmes** for Avian Influenza in poultry and wild birds to be carried out in the Member States.

Sampling strategy

a) Sampling shall be stratified throughout the territory of the whole country, taking into account the number of holdings to be sampled;

that number shall be defined so as to ensure the identification of at least one infected holding if the prevalence of infected holdings is ≥ 5 % (95 % confidence interval- 99% for ducks, geese and turkeys);

What do we know?

- Aim: detect **at least one** infected holding
- Prevalence of infected holdings is ≥ 5 %
- 95 % confidence interval
- 99% for ducks, geese and turkeys

- **Which formula would you use?**

- To estimate prevalence? $n = 1.96^2\ P(1-P)/d^2$
- To detect disease? $n = (1-(1-P)^{1/d})((N-d/2)+1)$

Home

Language English

Sample size to achieve specified population level (or herd, flock, cluster, etc) sensitivity

Input Values

This utility calculates the sample size required to achieve a target population or cluster level sensitivity for a survey. It replaces the previous options for calculating sample size for both a large (unknown) population and a finite population.

For these calculations unit specificity is assumed to be 100%. For cluster (herd, flock, etc) level calculations, enter test sensitivity, unit-level design prevalence and required cluster-level sensitivity. Alternatively, for population level calculations enter cluster-level (herd) sensitivity, cluster-level design prevalence and required population-level sensitivity.

Calculations use the hypergeometric approximation if population size is provided, or binomial method if population size is not specified.

Design prevalence (proportion or units): 0.05

(•) Proportion ( ) Unit(s)

Unit (test or cluster) sensitivity: 0.99

Required population sensitivity: 0.95

Population size (if known): 251

Submit Reset

Inputs are:

- Design prevalence as a proportion or number of units (animals for cluster level sensitivity and clusters for population-level sensitivity);
- Unit sensitivity (test sensitivity to calculate cluster (herd) sensitivity or cluster (herd) sensitivity tocalculate population sensitivit;
- Required population-level sensitivity; and
- Population size (optional if design prevalence is specified as a proportion, required if design prevalence is a number of units). Leave population size blank if not known.

Outputs are:

- Required sample size for the given target population sensitivity, design prevalence and unit sensitivity; and
- A table and graph of sample sizess for varying population and design prevalence values and the given target population sensitivity and unit sensitivity.

Sample size for freedom surveys

Analysed: Thu Jun 30, 2016 @ 18:14

Results

Design prevalence (Pstar)	5%
Unit (test or cluster) sensitivity	0.99
Required population sensitivity	0.95
Population size (N)	251
Sample size (n)	53

Home

Language English

Sample size to achieve specified population level (or herd, flock, cluster, etc) sensitivity

Input Values

This utility calculates the sample size required to achieve a target population or cluster level sensitivity for a survey. It replaces the previous options for calculating sample size for both a large (unknown) population and a finite population.

For these calculations unit specificity is assumed to be 100%. For cluster (herd, flock, etc) level calculations, enter test sensitivity, unit-level design prevalence and required cluster-level sensitivity. Alternatively, for population level calculations enter cluster-level (herd) sensitivity, cluster-level design prevalence and required population-level sensitivity.

Calculations use the hypergeometric approximation if population size is provided, or binomial method if population size is not specified.

Design prevalence (proportion or units): 0.05

(•) Proportion ( ) Unit(s)

Unit (test or cluster) sensitivity: 0.99

Required population sensitivity: 0.95

Population size (if known): 100000

Submit Reset

Inputs are:

- Design prevalence as a proportion or number of units (animals for cluster level sensitivity and clusters for population-level sensitivity);
- Unit sensitivity (test sensitivity to calculate cluster (herd) sensitivity or cluster (herd) sensitivity tocalculate population sensitivit;
- Required population-level sensitivity; and
- Population size (optional if design prevalence is specified as a proportion, required if design prevalence is a number of units). Leave population size blank if not known.

Outputs are:

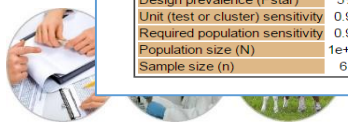
- Required sample size for the given target population sensitivity, design prevalence and unit sensitivity; and
- A table and graph of sample sizess for varying population and design prevalence values and the given target population sensitivity and unit sensitivity.

Sample size for freedom surveys

Analysed: Thu Jun 30, 2016 @ 18:18

Results

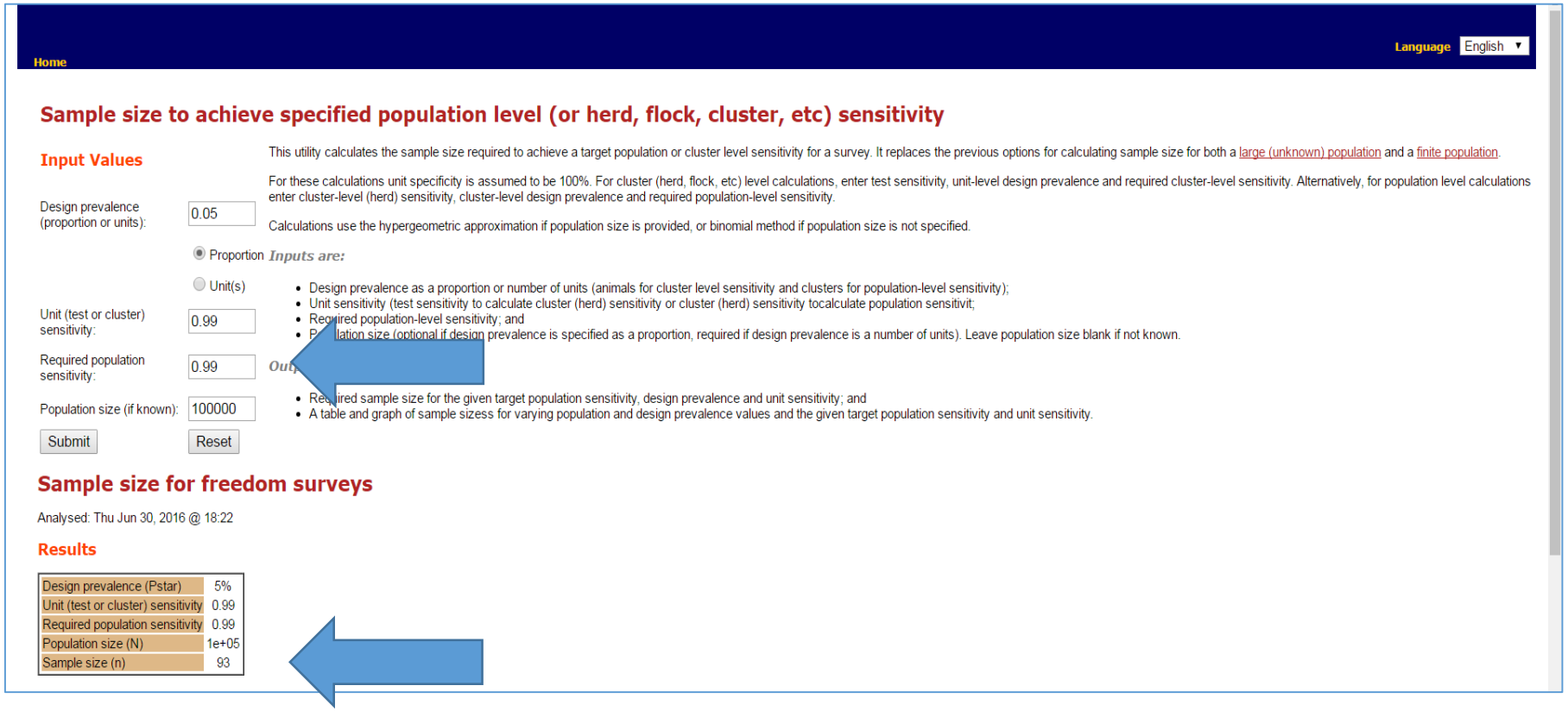
Design prevalence (Pstar)	5%
Unit (test or cluster) sensitivity	0.99
Required population sensitivity	0.95
Population size (N)	1e+05
Sample size (n)	61



Number of holdings to be sampled of each poultry category (except turkey, duck and goose holdings)

Number of holdings per poultry category per Member State	Number of holdings to be sampled
Up to 34	All
35-50	35
51-80	42
81-250	53
>250	60

99% for ducks, geese and turkeys



Home

Language English

Sample size to achieve specified population level (or herd, flock, cluster, etc) sensitivity

Input Values

This utility calculates the sample size required to achieve a target population or cluster level sensitivity for a survey. It replaces the previous options for calculating sample size for both a large (unknown) population and a finite population.

For these calculations unit specificity is assumed to be 100%. For cluster (herd, flock, etc) level calculations, enter test sensitivity, unit-level design prevalence and required cluster-level sensitivity. Alternatively, for population level calculations enter cluster-level (herd) sensitivity, cluster-level design prevalence and required population-level sensitivity.

Calculations use the hypergeometric approximation if population size is provided, or binomial method if population size is not specified.

Design prevalence (proportion or units): 0.05

Proportion

Unit(s)

Inputs are:

- Design prevalence as a proportion or number of units (animals for cluster level sensitivity and clusters for population-level sensitivity);
- Unit sensitivity (test sensitivity to calculate cluster (herd) sensitivity or cluster (herd) sensitivity tocalculate population sensitivit;
- Required population-level sensitivity; and
- Population size (optional if design prevalence is specified as a proportion, required if design prevalence is a number of units). Leave population size blank if not known.

Unit (test or cluster) sensitivity: 0.99

Required population sensitivity: 0.99

Outputs are:

- Required sample size for the given target population sensitivity, design prevalence and unit sensitivity; and
- A table and graph of sample sizess for varying population and design prevalence values and the given target population sensitivity and unit sensitivity.

Population size (if known): 100000

Submit Reset

Sample size for freedom surveys

Analysed: Thu Jun 30, 2016 @ 18:22

Results

Design prevalence (Pstar)	5%
Unit (test or cluster) sensitivity	0.99
Required population sensitivity	0.99
Population size (N)	1e+05
Sample size (n)	93

Number of turkey, duck and goose holdings to be sampled

Number of holdings per poultry category per Member State	Number of holdings to be sampled
Up to 46	All
47-60	47
61-100	59
101-350	80
>350	90

Sampling strategy

(b) the number of birds sampled from each holding shall be defined so as to ensure 95 % probability of identifying at least one positive bird if the prevalence of sero-positive birds is ≥ 30 %.

Blood samples for serological examination shall be collected from at least 5 to 10 birds per holding for each poultry category (except ducks geese and quail),

If more than one shed is present on a holding it is recommended to take at least 5 birds per shed.

What do we know?

- **Aim:** 95 % probability of identifying at least one positive bird
- Prevalence of infected holdings is ≥ 30 %.
- 95 % confidence interval

Home

Language English

Sample size to achieve specified population level (or herd, flock, cluster, etc) sensitivity

Input Values

Design prevalence (proportion or units): 0.3

Proportion

Unit(s)

Unit (test or cluster) sensitivity: 0.95

Required population sensitivity: 0.95

Population size (if known):

Submit Reset

This utility calculates the sample size required to achieve a target population or cluster level sensitivity for a survey. It replaces the previous options for calculating sample size for both a large (unknown) population and a finite population.

For these calculations unit specificity is assumed to be 100%. For cluster (herd, flock, etc) level calculations, enter test sensitivity, unit-level design prevalence and required cluster-level sensitivity. Alternatively, for population level calculations enter cluster-level (herd) sensitivity, cluster-level design prevalence and required population-level sensitivity.

Calculations use the hypergeometric approximation if population size is provided, or binomial method if population size is not specified.

Inputs are:

- Design prevalence as a proportion or number of units (animals for cluster level sensitivity and clusters for population-level sensitivity);
- Unit sensitivity (test sensitivity to calculate cluster (herd) sensitivity or cluster (herd) sensitivity tocalculate population sensitivit;
- Required population-level sensitivity; and
- Population size (optional if design prevalence is specified as a proportion, required if design prevalence is a number of units). Leave population size blank if not known.

Outputs are:

- Required sample size for the given target population sensitivity, design prevalence and unit sensitivity; and
- A table and graph of sample sizess for varying population and design prevalence values and the given target population sensitivity and unit sensitivity.

Sample size for freedom surveys

Analysed: Thu Jun 30, 2016 @ 18:34

Results

Design prevalence (Pstar)	30%
Unit (test or cluster) sensitivity	0.95
Required population sensitivity	0.95
Population size (N)	Unknown
Sample size (n)	9

SAFOSO

Animal Health Matters.
For Safe Food Solutions.

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs,
Education and Research EAER
State Secretariat for Economic Affairs SECO

Thanks