

# A practical guide to



**ASSIGN  
BUSTER**

A Practical Guide to Sampling Statistical & Technical Team This guide is brought to you by the Statistical and Technical Team, who form part of the VFM Development Team. They are responsible for advice and guidance on quantitative, analytical and technical issues. For further information about the matters raised in this guide, please contact: Alison Langham on ext. 7171

This guide is the latest in a series on sampling. It has been produced in response to a large number of requests received by the Statistical and Technical Team relating to sampling matters. The guide aims to consolidate the information required for you to complete the survey process from design to reporting. It provides this advice in an informal and practical way which should also help you understand the work of your consultants, and ask informed questions of the audited body. This guide replaces the previous guidance “ Use of Sampling - VFM Studies” published in 1992. Other guides related to this matter: Taking a Survey (1999) Presenting Data in Reports (1998) Collecting, Analysis and Presenting Data (1996) Contents Why sample? Sample design 4 5 Defining the population 6 Data Protection Act issues 6 Contracting out 6 Sample size 7 Weighting a sample 9 Sampling methods 11 Methods, their use and limitations Selecting an appropriate method Extracting the sample Interpreting and reporting the results Interpreting the results Reporting the results Glossary of terms Appendix 1 Relevant formulae for simple random sampling 11 13 14 15 15 17 18 19 Why sample? Recent examples VFM reports require reliable forms of evidence from which to draw robust conclusions. It is usually not cost effective or practicable to collect and examine all the data that might be available. Instead it is often necessary to draw a sample of information from the whole population to enable the detailed “ Sampling provides a means of gaining

information about the population without the need to examine the population in its entirety. " examination required to take place. Samples can be drawn for several reasons: for example to draw inferences across the entire population; or to draw illustrative examples of certain types of behavior. Caveats Sampling can provide a valid, defensible methodology but it is important to match the type of sample needed to the type of analysis required. The auditor should also take care to check the quality of the information from which the sample is to be

Excerpt from Highways Agency: Getting best value from the disposal of property HC58 Session 1999-00 drawn. If the quality is poor, sampling may not be justified. Do we really use them? Of the 31 reports published by the end of July of the 1999-2000 session, there are 7 examples of using judgmental sampling for illustrative case studies and 24 examples of sampling to draw inferences across the population, of which 19 were the basis for surveys. Can they provide strong evidence? In the Health area, four studies made extensive use of sampling and survey techniques to form the majority of the evidence which identified the potential for a one off saving of up to £400 million and possible annual savings of £150 million. 4

Excerpt from Charitable funds associated with NHS Bodies HC516 Session 1999-00

Sample design Sample design covers the method of selection, the sample structure and plans for analysing and interpreting the results. Sample designs can vary from simple to complex and depend on the type of information required and the way the sample is selected. The design will impact upon the size of the sample and the way in which analysis is carried out. In simple terms the tighter the required precision and the more complex the design, the larger the sample size. The design may make use of the characteristics of the population, but it does not

have to be proportionally representative. It may be necessary to draw a larger sample than would be expected from some parts of the population; for example, to select more from a minority grouping to ensure that we get sufficient data for analysis on such groups. Many designs are built around random selection. This permits justifiable inference from the sample to the population, at quantified levels of precision. Given due regard to other aspects of design, random selection guards against bias in a way that selecting by judgement or convenience cannot. However, a random selection may not always be either possible or what is required, in these cases care must be taken to match clear audit objectives to the sample design to The aim of the design is to achieve a balance between the required precision and the available resources. prevent introducing unintended bias. If you are sampling for the purposes of a survey then you should also be aware of the Taking a Survey guidance issued in 1999. 5 Defining the population Data Protection Act issues The first step in good sample design is to ensure that the specification of the target population is as clear and complete as possible to ensure that all elements within the population are represented. The target population is sampled using a sampling frame. Often the units in the population can be identified by existing information; for example, pay-rolls, company lists, government registers etc. A sampling frame could also be geographical; for example postcodes have become a well-used means of selecting a sample. Try to obtain the sample frame in the most automated way possible for ease of sampling; for example a database spreadsheet file. Often a government database or computer file can be All sampling frames will have some defects, despite assurances you may receive from the holder of the data. Usually there are ways to deal with this, for example amending

the list, selecting a larger sample and eliminating ineligible items, combining information from varying sources, or using estimated or proxy data. If you are having difficulties identifying a suitable sampling frame come and discuss this with the Statistical and Technical Team. used to identify the population and select a sample. You will need to ensure that this data is accurate, reliable, can be accessed, and that you have permission to draw a sample. The Data Protection Act requires us to obtain agreement to use data which also hold individuals' details. Many databases cannot be accessed because of this or other security reasons. However, it may be possible to extract selected information which is sufficient for the purposes of the study; for example using summarised data so that the individual cannot be identified. If you are in any doubt as to your position in this matter please refer to the Policy Unit. Contracting out If you use an outside contractor to carry out the sample they will normally put forward their proposed sample design. The design will often depend on whether you can obtain a suitable sampling frame from which the sample can be selected. If you cannot provide a database the contractor may be able to suggest a sampling frame to use. The contractor may well use a more complex sampling design than simple random sampling and it is important to check that what they have done is reasonable. The Statistical and Technical Team hold a database of contractors previously used by the Office, or you may wish to search for specific contractors who specialise in certain fields. A useful starting point for this is the British Market Research Association's selectline web page at: [www.bmra.org.uk/selectline](http://www.bmra.org.uk/selectline) The Team offer their service as a reference partner when drafting the tender for the work, evaluating the bids, A sampling frame is a list of all units in your population. 6 or assessing the <https://assignbuster.com/a-practical-guide-to/>

quality of the work. seeking an indication of likely population value a Sample size lower level such as 90 per cent is acceptable. For any sample design deciding upon the appropriate sample size will depend on five key factors and these are shown below. It is important to consider these factors together to achieve the right balance and ensure that the sample objectives are met. No estimate taken from a sample is expected to be exact, inference to the population will have an attached margin of error. The better the design, the less the margin of error and the tighter the precision but in most cases the larger the sample size. The amount of variability in the population i. e. the range of values or opinions, will also affect accuracy and therefore the size of sample required when estimating a value. The more variability the less accurate the estimate and the larger the sample size required. Population size does not normally affect sample size. In fact the larger the population size the lower the proportion of that population that needs to be sampled to be representative. It is only when the proposed sample size is more than 5 per cent of the population that the population size becomes part of the formulae to calculate the sample size. The effect is to slightly reduce the required sample size. If you are in this position please refer to the Team. If seeking to sample for attributes as opposed to the calculation of an average value, the proportion of the population displaying the attribute you are seeking to identify is the final factor for consideration. This can be estimated from the information that is known about the population, for example the proportion of hospitals who consider long waiting lists to be a problem. The confidence level is the likelihood that the results obtained from the sample lie within the associated precision. The higher the confidence level, that is the more certain you wish to be that the results are not atypical, the larger the

sample size. We normally use 95 per cent confidence to provide forceful conclusions, however, if you are only Margin of error or precision - a measure of the possible difference between the sample estimate and the actual population value. Variability in the population - the standard deviation is the most usual measure and often needs to be estimated. Confidence level how certain you want to be that the population figure is within the sample estimate and its associated precision. The population proportion - the proportion of items in the population displaying the attributes that you are seeking. Population size - total number of items in the population - only important if the sample size is greater than 5% of the population in which case the sample size reduces. Sample size 7 Our samples tend to be one-off exercises carried out with limited resources. Sometimes that means that the results can only be representative of the population in broad terms and breakdowns into smaller sub-groups may not always be meaningful. Practical limitations will often be the chief determinant of the sample size. A sample size of between 50 and 100 should ensure that the results are sufficiently reliable for the majority of purposes, although there will be occasions when a sample as small as 30 may be sufficient. Samples smaller than this fall into the category of case studies where statistical inferences to the population cannot be made, however, they can still form part of a valid and defensible methodology. The decisions surrounding the sample design and methodology should be discussed with all the parties involved to ensure their agreement to the process and avoid problems during clearance. Figure 1 (opposite) contains a sample size lookup table for samples selected using simple random sampling, the most frequently used method in the Office. If sampling for attributes then read off the sample size for the population

proportion and precision required to give your sample size. If there is more than the one outcome, for example A, B, C or D and the proportions were say 20 per cent, 10 per cent, 30 per cent and 40 per cent then the necessary sample size would be the one for the highest i. e. 40 per cent at the required confidence level and precision. If you are unsure of the population proportion then a 50 per cent proportion provides the most conservative sample size estimate and can also be used to provide an approximate sample size when determining a numeric estimate. The table shows the sample size needed to achieve the required precision depending on the population proportion using simple random sampling. For example, for 5 per cent precision with a population proportion of 70 per cent a sample size of 323 is required at the 95 per cent confidence level. Should you wish to calculate an exact simple random sample size for your own circumstances the formulae to do this are at appendix 1. However, should you elect to carry out a sampling methodology other than that based on a simple random sample please contact the Statistical and Technical Team who will be able to help you calculate an appropriate sample size. As a general rule, a statistical sample should contain 50 to 100 cases for each sample or sub-group to be analysed.

8 Figure 1: Sample size lookup table Population Proportion Precision (at the 95 per cent confidence level)  $\pm 12\%$   $\pm 10\%$   $\pm 8\%$   $\pm 5\%$   $\pm 4\%$   $\pm 3\%$   $\pm 2\%$   $\pm 1\%$

50%	66	96	150	384	600	1,067	2,401	9,604
45% or 55%	66	95	148	380	594	1,056	2,376	9,507
40% or 60%	64	92	144	369	576	1,024	2,305	9,220
35% or 65%	60	87	136	349	546	971	2,184	8,739
30% or 70%	56	81	126	323	504	896	2,017	8,067
25% or 75%	50	72	112	288	450	800	1,800	7,203
20% or 80%	42	61	96	246	384	683	1,536	6,147
15% or 85%	34	48	76	195	306	544	1,224	4,898
10% or 90%	24	35	54	138	216	384	864	3,457



95% 12 18 28 72 114 202 456 1, 824 If you are expecting non-response or a difficulty in locating your sample selections then it is prudent to over sample to ensure that the sample size achieved provides the required level of precision. The figures in bold and italics denote sample sizes of less than the recommended minimum. Weighting a sample A simple random sample of 384 cases might give the If a normal sample would be insufficient to reflect reflect the population characteristics it may still be the population characteristics then it may be perfectly valid if you are interested in the locations as necessary to look at ways in which this can be well as the workload. In the last column the sample improved. One way of doing this is to weight the has been weighted to reflect the population sample. If, for example, you are looking to sample characteristics. This approach would be more three regional offices and they have varying suitable if you are interested more in the actual cases workloads, you may want the sample to reflect the than the locations. The method of calculating the workloads at each location. Figure 2 shows an results for a weighted sample are different than for example where a total sample size of 384 (50 per cent the simple random sample. breakdown shown in figure 2. Whilst this does not proportion for a 5 per cent precision at 95 per cent confidence) is required.

Figure 2: Example of a weighted sample

Location	Population workload % of population workload	A simple random sample of the total workload	The effect on sample size at each location when the sample is weighted
North	50, 000 13%	128	51
South East	250, 000 67%	153	256
South West	75, 000 20%	103	77
TOTAL	375, 000 100%	384	384

A weighted sample more accurately reflects the workloads at the regional locations. 9 Post-weighting the sample done by applying the population proportions to the Should this

<https://assignbuster.com/a-practical-guide-to/>

weighting be required, but had not taken adjusted result. In this case an unweighted result of place at the sample selection stage then it is possible 37 per cent becomes a weighted proportion of 49 per to weight the sample in the results phase. This is cent. results of the unweighted sample to produce an Figure 3: Example of a post-weighted sample Location Simple Random Sample (A) Displaying Required Attribute (B) Unweighted Proportion Location Percentage Weighted Proportion (B/A) (C) (B/A) $\times$ C North 128 21 16% 13% 2% South East 153 98 64% 67% 43% South West 103 23 22% 20% 4% TOTAL 384 142 37% -- 49% A weighted sample more accurately reflects the workloads at the regional locations. This example shows that the most important thing is to gather sufficient information to enable you to make judgements about the population that you are sampling, whether it be that the information comes to light prior to selecting the sample or as a result of selecting the sample. Always be aware of what the results are saying and how true a reflection of the population they are. Ensure that the sample reflects the population characteristics whether before or after the sample selection.

10 Select a method that fulfils your objectives and matches the information and resources available. Sampling methods Methods, their use and limitations There are many different ways in which a sample can be selected. Nine of the most common methods are illustrated below. Method Definition Uses Cluster sampling Units in the population can often be found in geographical groups or clusters eg. schools, households etc. | Quicker, easier and cheaper than other forms of random sampling. | Larger sampling error than other forms of random sampling. | Does not require complete population information. | If clusters are not small it can become expensive. | Useful for face-to-face interviews. | A larger sample size may be needed to

<https://assignbuster.com/a-practical-guide-to/>

compensate for greater sampling error. A random sample of clusters is taken, then all units within those clusters are examined. | Works best when each cluster can be regarded as a microcosm of the population. Convenience sampling Judgement sampling Using those who are willing to volunteer, or cases which are presented to you as a sample. Based on deliberate choice and excludes any random process. | Readily available. | The larger the group, the more information is gathered. | Normal application is for small samples from a population that is well understood and there is a clear method for picking the sample. | Is used to provide illustrative examples or case studies. Limitations | Sample results cannot be extrapolated to give population results. | May be prone to volunteer bias. | It is prone to bias. | The sample is small and can lead to credibility problems. | Sample results cannot be extrapolated to give population results. 11 Method Definition Uses

Limitations Multi-stage sampling The sample is drawn in two or more stages (eg. a selection of offices at the first stage and a selection of claimants at the second stage). | Usually the most efficient and practical way to carry out large surveys of the public. | Complex calculations of the estimates and associated precision. Probability proportional to size Samples are drawn in proportion to their size giving a higher chance of selection to the larger items (eg. the more claimants at an office the higher the office's chance of selection). | Where you want each element (eg. claimants at an office) to have a equal chance of selection rather than each sampling unit (eg. offices). | Can be expensive to get the information to draw the sample. The aim is to obtain a sample that is representative of the population. | It is a quick way of obtaining a sample. | Not random so stronger possibility of bias. | It can be fairly cheap. | Good knowledge of population characteristics is essential.

Quota sampling The population is stratified by important variables and the required quota is obtained from each stratum. If there is no sampling frame it may be the only way forward. Additional information may improve the credibility of the results.

Simple random sampling Ensures every member of the population has an equal chance of selection. Produces defensible estimates of the population and sampling error.

Simple sample design and interpretation.

Stratified sampling Ensures units from each main group are included and may therefore be more reliably representative. The strata can have equal sizes or you may wish a higher proportion in certain strata.

Systematic sampling The population is sub-divided into homogenous groups, for example regions, size or type of establishment. Should reduce the error due to sampling. After randomly selecting a starting point in the population between 1 and  $n$ , every  $n$ th unit is selected, where  $n$  equals the population size divided by the sample size.

Ensures cases are spread across the population. Only appropriate if you are interested in the elements. Estimates of the sampling error and confidence limits probably can't be calculated. Need complete and accurate population listing. May not be practicable if a country-wide sample would involve lots of audit visits. Selecting the sample is more complex and requires good population information. The estimates involve complex calculations. Can be costly and timeconsuming if the sample is not conveniently located. Can't be used where there is periodicity in the population.

Selecting an appropriate method As you can see there are many methods available for use with varying degrees of complexity. Certain methods suit circumstances better than others and the following diagram is designed to help you select an appropriate method.

sample For simple random sampling it is possible to use either Excel or SPSS to select the sample for you. An illustration of how to extract a sample using both of these methods is shown below: In Excel In SPSS In SPSS use “ Data > Select Cases” then use the option “ Random” and complete the In Excel use “ Tools > Data Analysis > Sampling” to bring up the dialogue box shown. Enter the population value range as the input range and the number of samples. You can simply put a single cell at the start of an adjacent blank column for the output range. This will create a filter column which when selected will only allow any analysis or printing functions etc. to be carried out on the sample data rather than the population. The sample items will then be extracted and placed in the output range. It is also possible to use IDEA to extract the sample, If you are not intending to use a simple random contact the Statistical and Technical Team if you sample then the Statistical and Analytical Team can want help to do this. If the population is not held advise on how to extract the sample. electronically then an interval sample from a random starting point could be used as an alternative.

#### 14 Interpreting & reporting the results

##### Interpreting the results

The choice of sample design and how well it mimics The simple random sample gave an average price of the population will impact on the results. The closer £35, 630 ranging between £30, 480 and £40, 780 i. e. at the sample design to the population characteristics the 95 per cent confidence level the average property the more precise the estimate from the sample. It is value is £35, 630 plus or minus 14 per cent. The therefore important to match the calculation of the stratified sample gave an average of £36, 260 ranging results from the sample to the design of the sample. between £34, 370 and £38, 150 i. e. at the 95 per cent

The following shows an example of

<https://assignbuster.com/a-practical-guide-to/>

a population of 146 properties priced between £176, 000 and £17, 750 with an average value of £35, 760. The population distribution is illustrated and shows that the confidence level the average property value is £36, 260 plus or minus 5 per cent. In this case, the stratified sample provides a more precise estimate of the population average. majority of the population is priced below £50, 000. To obtain the sample estimate for a simple random Two samples of 50 properties were selected, one sample you can use a package such as Excel or SPSS using simple random sampling and the other which will return not only the average but also the stratifying the population into above and below standard deviation, and the precision at the 95 per £50, 000. cent confidence level. 15 In Excel In SPSS In Excel use “ Tools > Data Analysis > In SPSS use “ Analyze > Descriptive Descriptive Statistics” to bring up this Statistics > Explore” to bring up this dialogue box. Put the sample values as the dialogue box. Select the required variable. input range. The output produced is shown above and provides all the required information. The output produced is shown above and provides all the required information. For attribute sampling the results are often quoted as The formulae used to calculate the results in this “ 70 per cent agreed that cleanliness would reduce section are given in appendix 1. infection”. If the sample size was 250 then what would be the precision of the answer. You could use the table at Figure 1 to provide an estimate, looking along the 70 per cent proportion row you will find that a sample size of 250 lies between 5 and 8 per cent precision. The accurate result is 6 per cent precision. 16 If you are using a sample other than simple random then seek advice from the Statistical and Technical Team when it comes to calculating the results. Reporting the results When reporting the results of a sample it is important

to cover several key facts: | the sample size; | the sample selection methodology; | the estimates resulting from the sample, and | the precision and confidence intervals for the estimates. Excerpt from Managing Finances in English Further Education Colleges HC454 Session 1999-00 Excerpt from Compensating the Victims of Violent Crime HC454 Session 1999-00 For advice on graphical presentation of the data see Presenting Data in Reports (1998) or contact the Statistical and Technical Team. 17 Glossary of terms

Confidence level Margin of error A measure of the difference between the estimate from the sample and the population value. Population The number of items from which to draw your sample. Population proportion The proportion of items within the population which exhibit the characteristics you are seeking to examine, this is only required when sampling for attributes. Precision A measurement of the accuracy of the sample estimate compared to the population value. Sample A selection of items from which you may estimate a feature of the population. Sample size The number of items in the sample. Standard deviation 18 The certainty with which the estimate lies within the margin of error. A measure of the variability in the population values, this is only required when sampling for values. Appendix 1 Relevant formulae for simple random sampling Sampling for proportions Z score values where Z is the z score associated with the confidence level required, E is the required precision, and p is the occurrence rate within the population. Confidence level Z score value 80% 1. 28 85% Sample size 1. 44 90% 1. 65 95% 1. 96 99% 2. 58 Estimate of proportions where  $y_i = 0$  or  $1$ , so that the estimate becomes a count of all the relevant cases divided by the number of cases in the sample. Precision Sampling for values Sample size If the sample size,  $n$ , is at least 5% of the population size,  $N$ , then the

calculation becomes: Adjusted sample size where  $s$  is an estimate of the standard deviation. Estimate of the average where  $y_i$  are the individual values from the sample. Precision 19 Designed and produced by the National Audit Office Design Group