

Sampling

The Basics

Sampling is used when you do not want to conduct census surveys

If the sample is randomly generated and of sufficient size the results should be representative of the total population

Sampling error (also known as confidence interval or margin of error) measures the accuracy of survey results in relation to the total population

Sampling error relevant for quantitative, not small-scale qualitative, surveys

The results from a sample survey should be checked for representativeness and weighted if necessary. Weighting is a statistical technique in which a number is assigned to each data result to reflect its relative importance, for example to boost the results from one particulate subgroup who returned fewer surveys than other subgroups in the survey, to reflect the true proportion in the total population.

Standard benchmark sampling errors for STAR surveys and HouseMark for benchmarking are :

- ± 5% for populations of up to 1,000 residents (up to 278 responses needed)
- ± 4% for populations of between 1,000 and 10,000 residents (375 – 566 responses)
- ± 3% for populations of over 10,000 residents (over 965 responses needed)

(Margins of error at the 95% confidence level)

It is best to achieve at least 100 responses at the sub-group level (for example management area, neighbourhood)



Sampling error - look up table

Population	Numbers of achieved responses required to meet sampling errors at 95% confidence level							
	±3.0%	±4.0%	±5.0%	±6.0%	±7.0%	±8.0%	±9.0%	±10.0%
250	203	177	152	129	110	94	81	70
500	341	273	217	174	141	116	96	81
750	441	334	254	197	156	125	103	85
1000	516	375	278	211	164	131	106	88
1250	576	406	294	220	170	134	108	89
1500	624	428	306	227	174	137	110	91
2000	695	462	322	236	179	140	112	92
2500	748	485	333	241	182	142	114	93
3000	786	500	341	245	184	143	114	93
3500	819	513	346	248	186	144	115	94
4000	843	523	350	250	187	145	116	94
4500	862	530	354	252	188	146	116	94
5000	880	535	357	253	189	146	116	95
7500	935	555	365	258	191	147	117	95
10000	965	567	370	260	193	147	118	95
15000	999	578	375	262	194	149	118	96
20000	1015	584	377	263	194	149	118	96
25000	1025	587	379	264	195	150	118	96
30000 +	1032	588	380	265	195	150	118	96

Sample design for a STAR survey

Tenure	Population	Proposed sample	Sample as % of total population	Estimated response rate	Estimated replies	Sampling error
General needs tenants	7000	1383	20%	40%	553	±4.0%
Housing for Older People	1000	750	75%	50%	375	±4.0%
Supported housing tenants	300	300	100%	40%	120	±7.0%
Sub total tenants	8300	2433	29%		1048	±2.9%
Leaseholders	575	575	100%	30%	173	±6.3%
Shared owners	225	225	100%	30%	68	±9.9%
Sub total homeowners	800	800	100%		240	±5.3%
ALL RESIDENTS	9100	3233	36%		1288	±2.6%



Sample design – subgroup sampling

Example 1 : Proportionate sampling based on overall accuracy of $\pm 3.0\%$

- Data does not have to be weighted
- Data accuracy higher in largest areas
- Can be more cost effective as sample size is smaller

- Different levels of accuracy between subgroups being analysed
- Not suitable if greater reliability or census required in particular subgroups

	Population	Sample size (number of achieved interviews/replies)	Sampling error
Subgroup			
Neighbourhood 1	1,500	166	$\pm 7.2\%$
Neighbourhood 2	1,750	194	$\pm 6.7\%$
Neighbourhood 3	2,500	277	$\pm 5.6\%$
Neighbourhood 4	1,000	111	$\pm 8.8\%$
Neighbourhood 5	1,800	200	$\pm 6.6\%$
All areas	8,550	949	$\pm 3.0\%$

Example 2 : Disproportionate sampling based on accuracy of $\pm 6.5\%$ at subgroup level

- Weighting has to be applied at group level
- Accuracy in largest areas the same as smallest area
- Likely to increase sampling size
- Reliability same across subgroups

	Population	Sample size (number of achieved interviews/replies)	Sampling error
Subgroup			
Neighbourhood 1	1,500	198	$\pm 6.5\%$
Neighbourhood 2	1,750	201	$\pm 6.5\%$
Neighbourhood 3	2,500	208	$\pm 6.5\%$
Neighbourhood 4	1,000	185	$\pm 6.5\%$
Neighbourhood 5	1,800	202	$\pm 6.5\%$
All areas	8,550	994	$\pm 3.0\%$