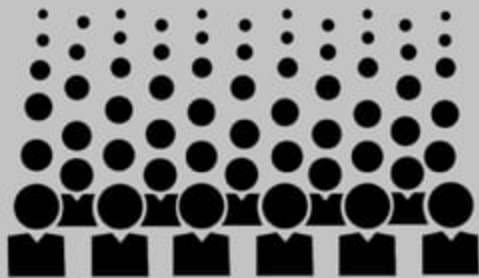


What is sample and Why are samples are used in statistics?

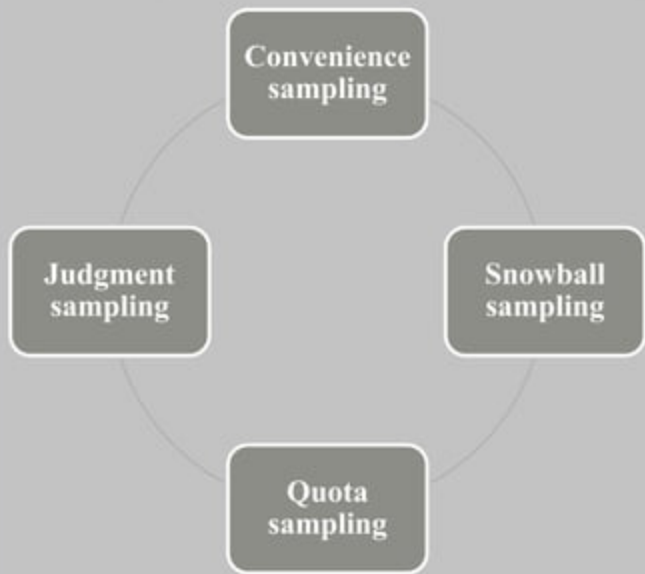
- The population is very large, making a census or a complete enumeration of all the values in the population is either impractical or impossible.
- Elements of a sample:
 - Sample points
 - Sampling units
 - observations
- The data sample may be drawn from:
 - Subset of a population
 - Multi-subset



Probability Sampling Methods



Nonprobability Sampling Methods



What size sample do I need?

The answer to this question is influenced by a number of factors, including:

1. Whether undertaking a **qualitative** or **quantitative** study.
2. Purpose of the study
3. Population size
4. Risk of selecting a "bad" sample
5. Allowable sampling error.



Sample size determination in Qualitative Study

- Probability sampling not appropriate
- Sample should have represent salient characteristics of the population.
- Until point of theoretical saturation
- Sample size is usually small
- A matter of judgement and expertise in evaluating the quality of information
- A flexible, pragmatic approach.



Sample size determination in Qualitative Study

- Researcher actively selects the most productive sample
- Developing a framework of the variables that might influence an individual's contribution
- If the subjects are known to the researcher, they may be stratified according to known public attitudes or beliefs



Sample size determination in Qualitative Study

- It may be advantageous to study a broad range of subjects :
 - *maximum variation sample*
 - *outliers (deviant sample)*
 - *subjects who have specific experiences (critical case sample)*
 - *subjects with special expertise (key informant sample).*



Sample size determination in Qualitative Study

- The iterative process of qualitative study design
- Theoretical sampling necessitates building interpretative theories
- grounded theoretical approach

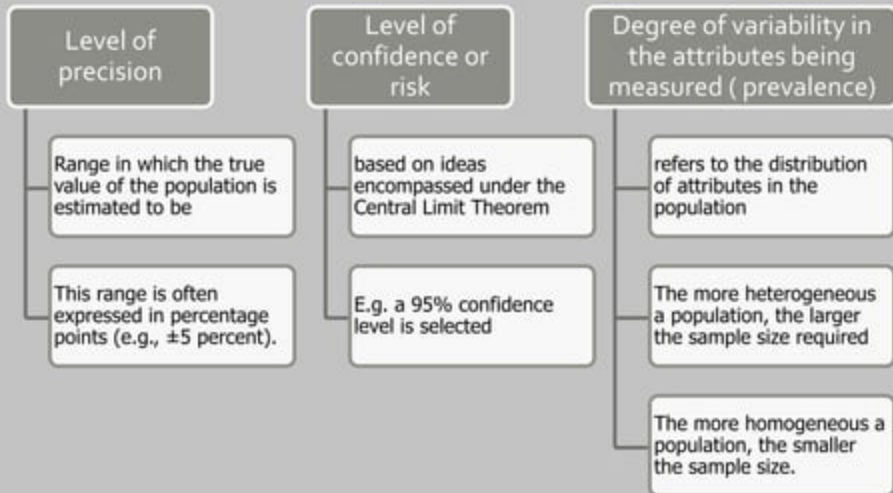


Some suggestions of sample size in qualitative studies

- The smallest number of participants could be 15
- Number of participants Should lie under 50
- 6-8 participants for FGDs AND at least 2 FGDs per population group
- Attainment of saturation
- Justification of choice of number



Sample size determination in quantitative study



Sample size determination in quantitative study

- A proportion of 50 % indicates a greater level of variability than either 20% or 80%
- proportion of 0.5
- Sample size affects accuracy of representation
- Minimum suggested sample is 30 and upper limit is 1,000



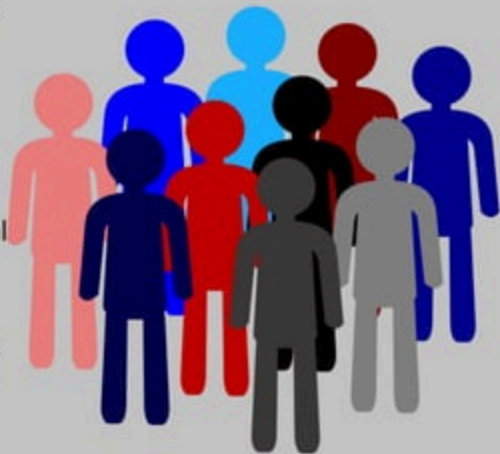
Strategies for Determining Sample Size

1. Using a census for small populations
2. Imitating a sample size of similar studies
3. Using published tables
4. Applying formulas to calculate a sample size



1. Using a Census for Small Populations

- One approach is to use the entire population as the sample
- Cost considerations make this impossible for large populations
- Attractive for small populations (e.g., 200 or less).
- Eliminates sampling error and provides data on all the individuals
- Some costs are "fixed"
- the entire population would have to be sampled in small populations to achieve a desirable level of precision



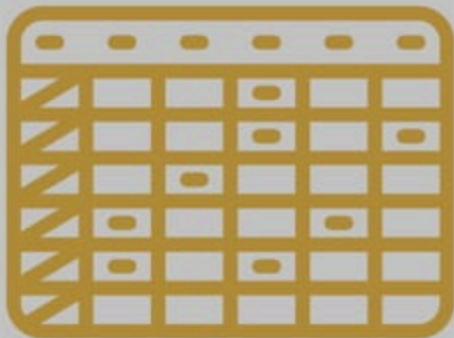
2. Imitating a sample size of similar studies

- Use the same sample size as those of studies similar to the one you plan(Cite reference)
- Procedures employed in these studies should be reviewed
- A review of the literature in your discipline can provide guidance about "typical" sample sizes



3. Using published tables

- Published tables provide the sample size for a given set of criteria.
- Necessary for given combinations of precision, confidence levels and variability.
- The sample sizes presume that the attributes being measured are distributed normally or nearly so.
- you may need to calculate the necessary sample size for a different combination of levels of precision, confidence, and variability.



3. Using published tables

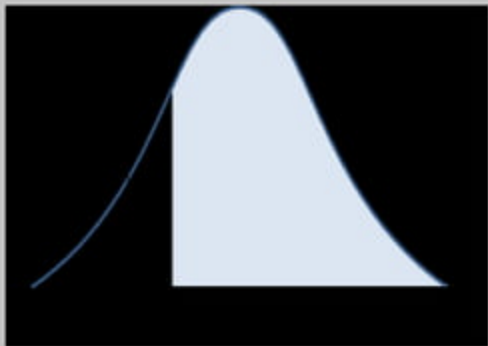
- Sample Size for $\pm 5\%$, $\pm 7\%$ and $\pm 10\%$ Precision Levels where Confidence Level Is 95% and $P=.5$.

Size of Population	Sample Size (n) for Precision (e) of:		
	$\pm 5\%$	$\pm 7\%$	$\pm 10\%$
100	81	67	51
125	96	78	56
150	110	86	61
175	122	94	64
200	134	101	67
225	144	107	70
250	154	112	72
275	163	117	74
300	172	121	76
325	180	125	77
350	187	129	78
375	194	132	80
400	201	135	81
425	207	138	82
450	212	140	82

Normal Distribution

The **normal distribution** is the most important and most widely used **distribution in statistics**. It is sometimes called the "bell curve," although the tonal qualities of such a bell would be less than pleasing

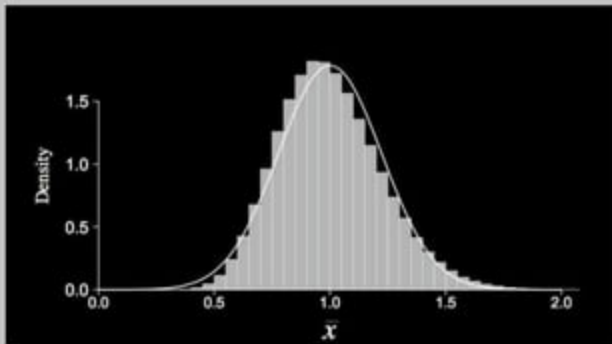
$$\frac{1}{\sqrt{2\sigma^2\Pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

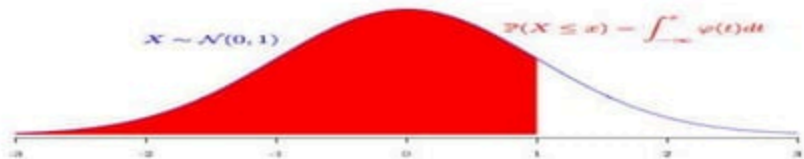


Central Limit theorem

The theorem says that if we take samples of size n from any arbitrary population (with any arbitrary distribution) and calculate \bar{X} then sampling distribution of \bar{X} will approach the normal distribution as the sample size n increases with mean μ and standard error

$$\frac{\sigma}{\sqrt{n}}$$





	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Cochran equation

Where n_0 is the sample size,

$Z_{\alpha/2}$ is the abscissa of the normal curve that cuts off an area α at the tails;

$(1 - \alpha)$ equals the desired confidence level, e.g., 95%);

e is the desired level of precision,

p is the estimated proportion of an attribute that is present in the population, and q is $1-p$.

The value for Z is found in statistical tables which contain the area under the normal curve. e.g $Z = 1.96$ for 95 % level of confidence

$$n = \frac{z_{\alpha/2}^2 \cdot p \cdot q}{e^2}$$

Finite population

N is the population size

$$n = \frac{n_0}{1 + \frac{(n_0 - 1)}{N}}$$

A Simplified Formula For Proportions

Yamane (1967:886) provides a simplified formula to calculate sample sizes.

ASSUMPTION:

95% confidence level

$P = .5$;

Where n is the sample size,

N is the population size,

e is the level of precision.

$$n = \frac{z_{\alpha}^2 \cdot p \cdot q \cdot N}{z_{\alpha}^2 \cdot p \cdot q + N(e^2)}$$

$$n = \frac{N}{1 + N(e^2)}$$

Allocation methods in Stratified sampling

Sample size through proportional allocation method :

In this method, the sampling fraction, $\frac{n}{N}$ is same in all strata. This allocation was used to obtain a sample that can estimate size of the sample with greater speed and a higher degree of precision.



$$n_i = n \cdot \frac{N_i}{N}$$

Allocation methods in Stratified sampling

The sample size through optimum allocation method :

The allocation of the sample units to the different stratum is determined with a view to minimize the variance for a specified cost of conducting the survey or to minimize the cost for a specified value of the variance. The cost function is given by

Where, a is the observed cost which is constant, c_i is the average cost of surveying one unit in the i th stratum.

Therefore, the required sample size in different stratum is given by

$$c = a + \sum_i^k n_i c_i \qquad n_i = n \frac{\frac{N_i S_i}{\sqrt{c_i}}}{\sum_i^k \frac{N_i S_i}{\sqrt{c_i}}}$$

THANK YOU
!