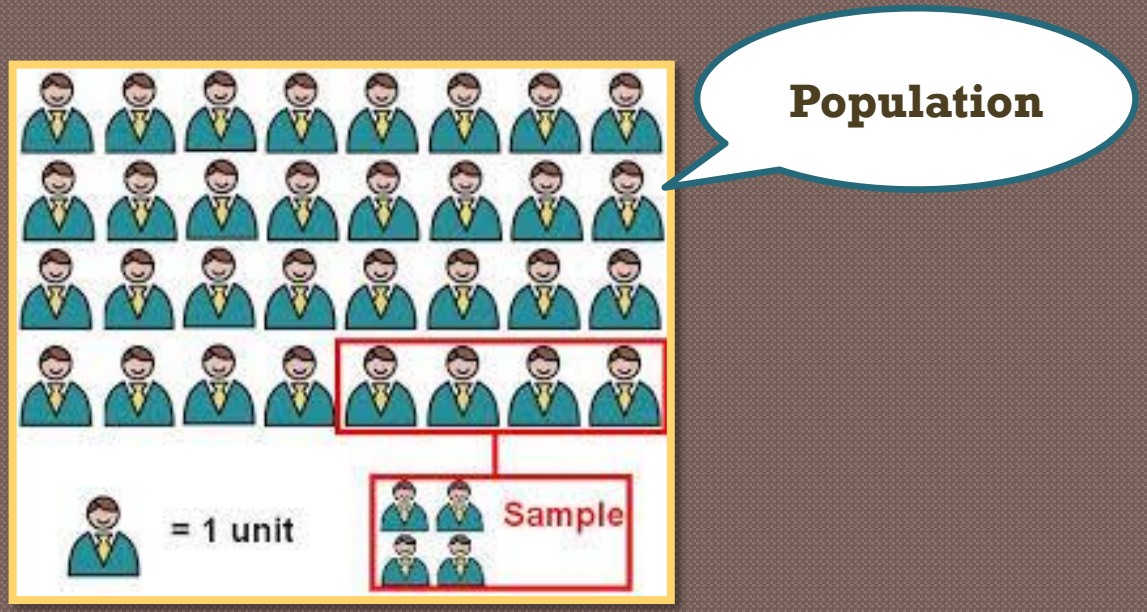
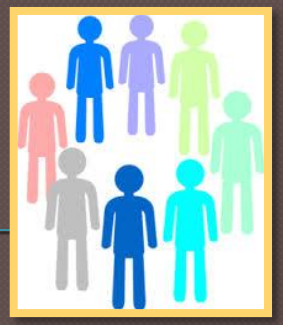
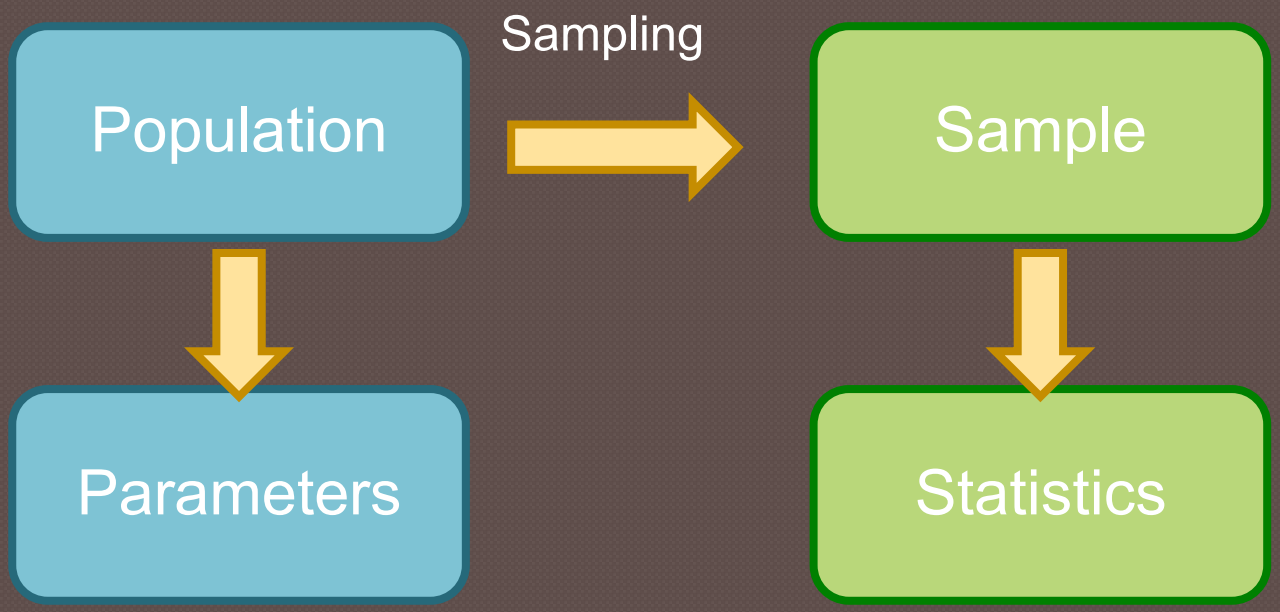
Sampling



Concept of Sampling

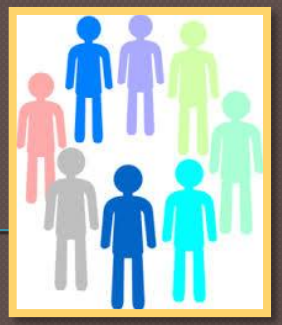




Note: Numbers computed from a population called "Parameter"

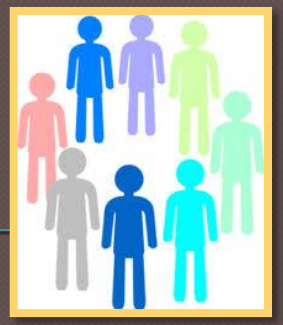
Numbers computed from a sample called "Statistics"

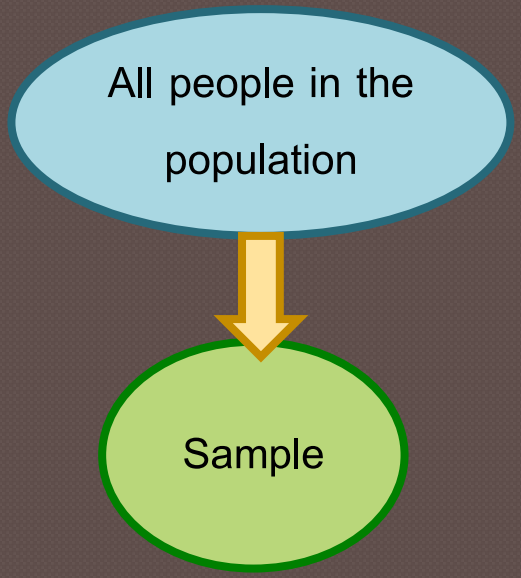
Forms of Sampling



- Random Sampling
  - Simple Random Sampling
  - Stratified Random Sampling
  - Cluster Sampling
- Non-random Sampling
  - Convenience Sampling
  - Quota Sampling
  - Known Group Sampling
  - Snowball Sampling

Simple Random Sampling

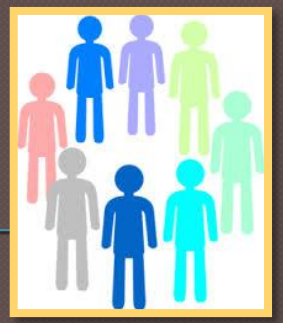


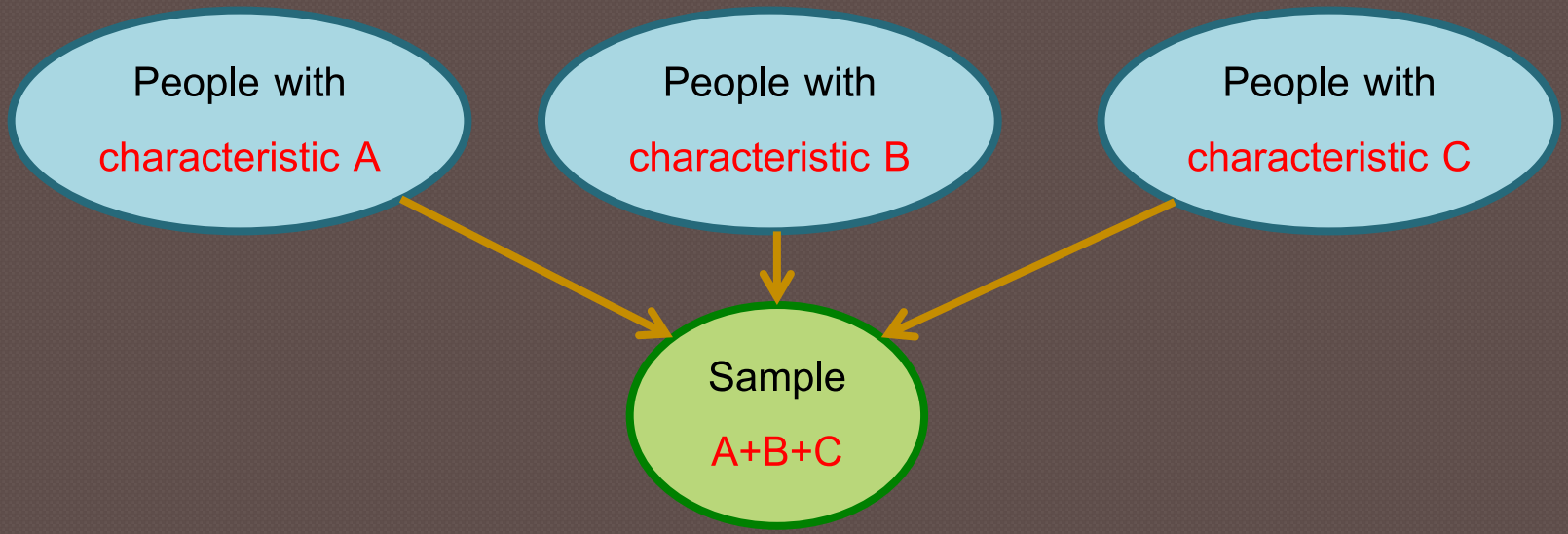


Remark:

- Best assurance of representativeness
- Time-consuming

Stratified Random Sampling

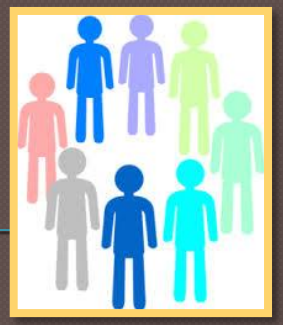


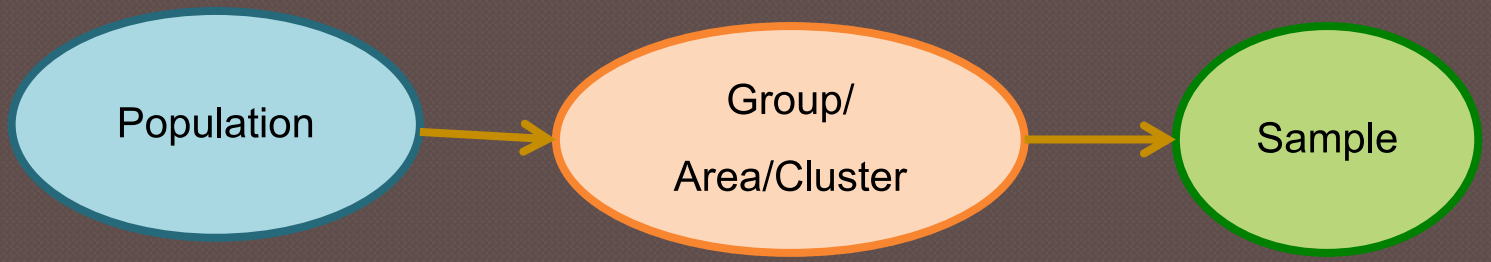


Remark:

- Easier than Simple Random Sampling
- Come very close to the representativeness
- Weights for stratification are often unknown

Cluster Sampling

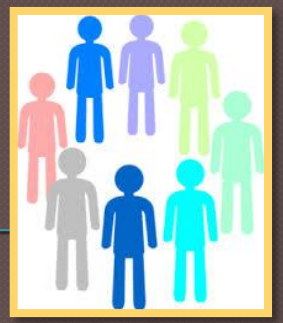




Remark:

- Easier than Simple Random Sampling
- Come very close to the representativeness
- Weights for cluster are often unknown

Non-random Sampling



- Convenience Sampling

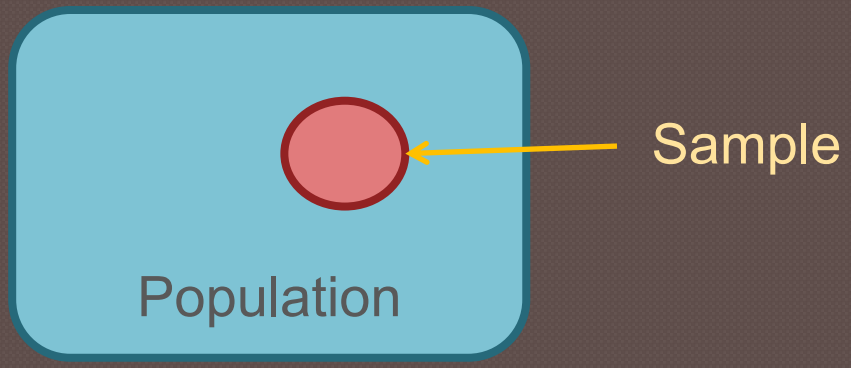
  (Select events that are most readily available)
- Quota Sampling

  (Like stratified sampling but without the random selection)
- Known Group Sampling

  (Selection of events from groups that are known to possess a particular characteristics under investigation)
- Snowball Sampling

  (Selection of events based on referrals from initial information)

Sample Size



- A sample is a percentage of the total population in statistics. You can use the data from a sample to make inferences about a population as a whole.

How to find a sample Size

Step 1: Conduct a census if you have a small population

Step 2: Use a sample size from a similar study

Step 3: Use a table to find your sample size

Step 4: Use a sample size calculator

Step 5: Use a formula

- Sample size in Descriptive studies
- Sample size in Inferential studies

Statistical Sample Size Table

Confidence Level

Confidence Interval

Required Sample Size								
	Confidence = 95%				Confidence = 99%			
Population Size	Margin of error				Margin of Error			
	5.0%	3.5%	2.5%	1.0%	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

Sample Size Calculator

Determine Sample Size

Confidence Level: (•) 95% ( ) 99%

Confidence Interval: 5

Population: 1050292

Calculate Clear

Sample size needed: 384

Formulas for Sample Size Calculation

Requirements:

- ✓ Confidence Level
  - → Ex. "95% confident" = 5% significance level (α)
    "99% confident" = 1% significance level (α)
- ✓ Confidence Interval
  - → Degree of precision (Margin for Random Errors <MRE>)
    Ex. +/- 1%, +/- 2%, +/- 3%, +/- 5%, +/- 10%

Sample Size in Descriptive Studies

A sample to calculate the mean value of the population

b) The formula to calculate the sample size for a mean (or point) estimate is:

$$N = \left(\frac{SD}{SE}\right)^2$$

where N = the required sample size,

SD = the standard deviation, and

SE = the standard error of the mean

SE = MRE / Z* value

for a confidence level

Confidence level	z* value
90%	1.645
95%	1.96
98%	2.33
99%	2.58

Sample Size in Descriptive Studies

A sample to estimate a proportion or percentage

$$N = \frac{P\,(100\% - P)}{(SE)^2}$$

SE = MRE / Z* value

for a confidence level

Confidence level	z* value
90%	1.645
95%	1.96
98%	2.33
99%	2.58

Sample Size in Inferential Studies

A sample size for Chi-squared test

$$N = K \times \frac{p_1(1-p_1) + p_2(1-p_2)}{(p_1 - p_2)^2}$$

K = constant which is a function of α and β

Table 2: Values of K, as used for sample size calculations

		Power: 50% $\beta = 0.5$	80% $\beta = 0.2$	90% $\beta = 0.1$	95% $\beta = 0.05$
α:	0.10	2.7	6.2	8.6	10.8
	0.05	3.8	7.9	10.5	13.0
	0.02	5.4	10.0	13.0	15.8
	0.01	6.6	11.7	14.9	17.8

Sample Size in Inferential Studies

A sample size for t-test

$$N = 2 \times K \times \left(\frac{\sigma}{\mu_1 - \mu_2} \right)^2$$

K = constant which is a function of α and β

Table 2: Values of K, as used for sample size calculations

		Power: 50% β = 0.5	80% β = 0.2	90% β = 0.1	95% β = 0.05
α:	0.10	2.7	6.2	8.6	10.8
	0.05	3.8	7.9	10.5	13.0
	0.02	5.4	10.0	13.0	15.8
	0.01	6.6	11.7	14.9	17.8

Validity and Reliability

Validity

- Conclusion Validity
- Internal Validity
- External Validity
  - Population Validity
  - Ecological Validity

- Test Validity
  - Criterion Validity
    - Concurrent Validity
    - Predictive Validity
  - Content Validity
  - Construct Validity
- Face Validity

Validity

Types	Meaning
Population Validity	Generalized based on the sample
Ecological Validity	Generalized based on the testing environment
Concurrent Validity	Measures the test against a benchmark test
Predictive Validity	Measures how well a test predicts abilities
Content Validity	Measures how well a test compares to the real world using scores by experts
Face Validity	Measures how representative a research project is at face value using personal judgments

Reliability

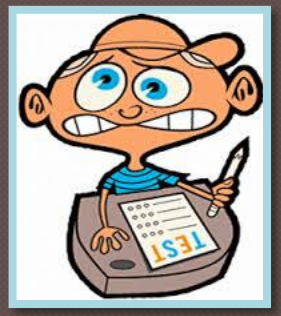
- The idea behind reliability is that any significant results must be more than a one-off finding and be inherently repeatable.
- Types of Reliability of Instruments:
  - Test-Retest Reliability
  - Internal Consistency Reliability

Validity and Reliability

Test-Retest Reliability

- Meaning: Involves administering the same test with the same group of people, after a period of time, and comparing the results to ensure that there is a correlation between the results.

Now ⟶ Future



Note: Assuming that nothing has changed in that time period

Internal Consistency Reliability

- Meaning: Involves measuring two different versions of the same item within the same test to ensure that there is a correlation and that they measure the same thing.
- 3 main techniques for measuring the internal consistency reliability: 1) Split-Halves Test
  2) Kuder-Richardson Test
  3) Cronbach's Alpha Test

Internal Consistency Reliability

- Split-Halves Test

→ Ex. A questionnaire could be divided into odd and even questions. The results from both halves are statistically analyzed, and if there is weak correlation between the two, then there is a reliability problem with the test.

Internal Consistency Reliability

- Kuder-Richardson Test
- → Allow only right or wrong questions (1, 0)

$$r_{KR-20} = \left(\frac{K}{K-1}\right)\left(1 - \frac{\sum pq}{s^2}\right)$$

Internal Consistency Reliability

- Cronbach's Alpha Test
- → Allow multi-level responses (Rating scale: 1, 2, 3, 4, 5)
- → The Higher the score, The better

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\Sigma Vi}{Vtest}\right)$$


Q&A