

Sampling People, Records, & Networks

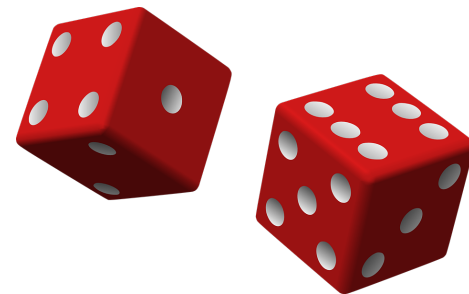
Jim Lepkowski, PhD

Professor & Research Professor *Emeritus*

Institute for Social Research, University of Michigan

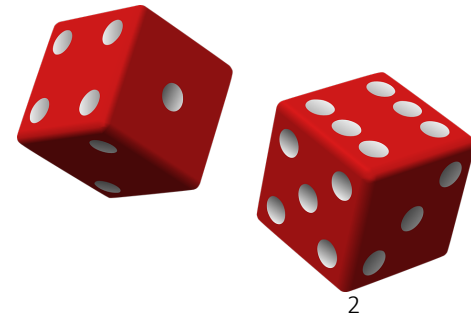
Research Professor,

Joint Program in Survey Methodology, University of Maryland



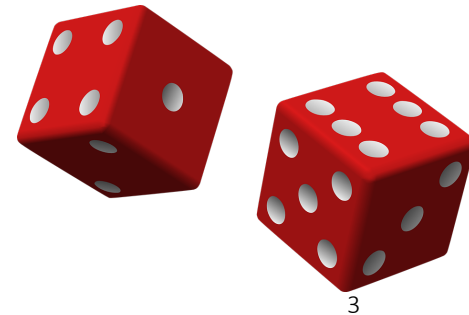
Unit I

- 1 Research designs ...
 - 2 Surveys
 - 3 Why sample?
 - 4 Why randomize?
 - 5 Types of sampling
 - 6 Evaluating samples
 - 7 Units sampled
- **Unit I: Sampling as a research tool**
 - Lecture 1 – Research design & sampling
 - Lecture 2 – Surveys & sampling
 - Lecture 3 -- Why sample at all?
 - Lecture 4 – Why might we randomize, and how do we do it?
 - Lecture 5 – What happens when we randomize?
 - Lecture 6 – How do we evaluate how good the sample is?
 - Lecture 7 – What kinds of things can we sample?
 - **Unit 2: Mere randomization**
 - **Unit 3: Saving money**
 - **Unit 4: Being more efficient**
 - **Unit 5: Simplifying sampling**
 - **Unit 6: Some extensions & applications**

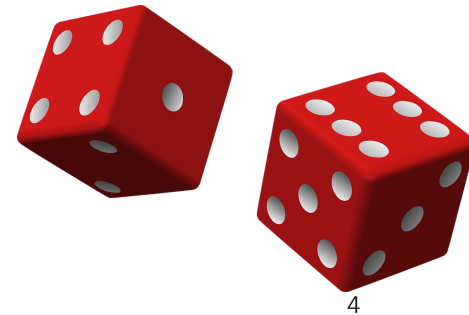


Unit I

- 1 Research designs ...
 - 2 Surveys
 - 3 Why sample?
 - 4 Why randomize?
 - 5 Types of sampling
 - 6 Evaluating samples
 - 7 Units sampled
- **Unit I: Sampling as a research tool**
 - Lecture 1 – Research design & sampling
 - Lecture 2 – Surveys & sampling
 - Lecture 3 -- Why sample at all?
 - Lecture 4 – Why might we randomize, and how do we do it?
 - **Lecture 5 – What happens when we randomize?**
 - Lecture 6 – How do we evaluate how good the sample is?
 - Lecture 7 – What kinds of things can we sample?
 - **Unit 2: Mere randomization**
 - **Unit 3: Saving money**
 - **Unit 4: Being more efficient**
 - **Unit 5: Simplifying sampling**
 - **Unit 6: Some extensions & applications**



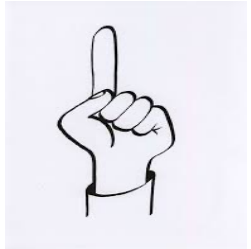
- Population to frame to sample
 - One random sample
 - Many possible samples
- **Unit 1: Sampling as a research tool**
 - Lecture 1 – Research design & sampling
 - Lecture 2 – Surveys & sampling
 - Lecture 3 -- Why sample at all?
 - Lecture 4 – Why might we randomize, and how do we do it?
 - **Lecture 5 – What happens when we randomize?**
 - Lecture 6 – How do we evaluate how good the sample is?
 - Lecture 7 – What kinds of things can we sample?
 - **Unit 2: Mere randomization**
 - **Unit 3: Saving money**
 - **Unit 4: Being more efficient**
 - **Unit 5: Simplifying sampling**
 - **Unit 6: Some extensions & applications**



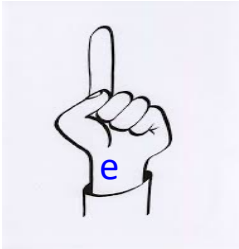
- Sample quality

- Sampling process

1 Population

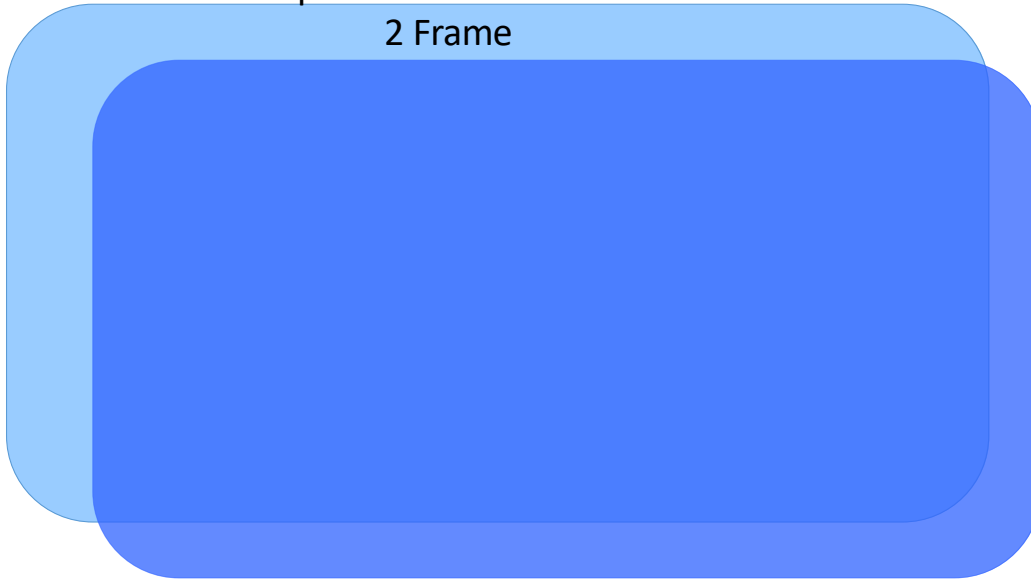


1 Population



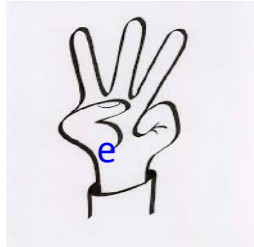
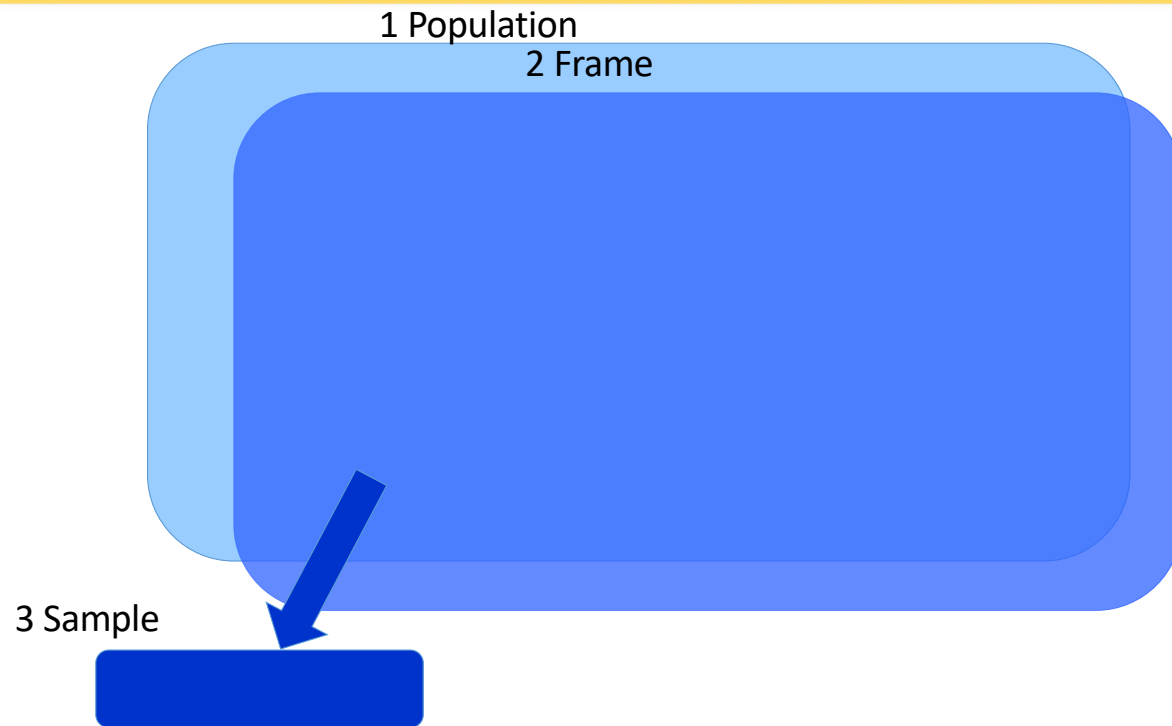
1 Population

2 Frame



Faculty Salaries (in \$1,000)

Seq. No.	ID	Division	Sex	Ran	Salary	Seq. No.	ID	Division	Sex	Ran	Salary	Seq. No.	ID	Division	Sex	Ran	Salary
1	1	Eng&Prof	m	3	\$88	51	155	Eng&Prof	m	3	\$55	101	217	Lit&SocSci	m	2	\$55
2	2	Medicine	f	3	\$45	52	156	Biol&Sci	m	1	\$49	102	218	Medicine	m	3	\$80
3	9	Medicine	m	3	\$57	53	157	Eng&Prof	m	3	\$57	103	219	Eng&Prof	m	1	\$114
4	11	Medicine	m	1	\$133	54	158	Medicine	m	1	\$118	104	220	Lit&SocSci	m	1	\$63
5	12	Eng&Prof	f	2	\$71	55	159	Medicine	m	3	\$84	105	221	Medicine	m	1	\$112
6	13	Lit&SocSci	m	1	\$113	56	160	Eng&Prof	m	3	\$52	106	222	Medicine	m	1	\$93
7	14	Medicine	f	3	\$65	57	161	Medicine	m	3	\$64	107	223	Lit&SocSci	m	2	\$47
8	15	Biol&Sci	m	3	\$47	58	162	Eng&Prof	m	1	\$75	108	224	Biol&Sci	m	1	\$127
9	16	Lit&SocSci	f	3	\$39	59	163	Medicine	f	1	\$87	109	225	Eng&Prof	m	2	\$121
10	17	Biol&Sci	m	1	\$74	60	164	Eng&Prof	m	3	\$58	110	226	Medicine	m	3	\$58
11	18	Medicine	m	1	\$88	61	165	Medicine	f	3	\$39	111	227	Biol&Sci	f	3	\$97
12	19	Lit&SocSci	m	1	\$62	62	166	Medicine	m	3	\$69	112	228	Lit&SocSci	m	1	\$71
13	37	Lit&SocSci	m	1	\$49	63	167	Medicine	f	2	\$46	113	229	Eng&Prof	m	1	\$72
14	38	Medicine	m	3	\$88	64	179	Eng&Prof	f	1	\$86	114	230	Lit&SocSci	m	3	\$29
15	39	Medicine	m	1	\$181	65	180	Medicine	m	3	\$87	115	231	Medicine	m	2	\$167
16	40	Eng&Prof	m	3	\$63	66	181	Medicine	m	3	\$59	116	232	Lit&SocSci	m	3	\$36
17	41	Medicine	m	2	\$94	67	182	Eng&Prof	f	3	\$44	117	233	Medicine	m	1	\$57
18	42	Eng&Prof	m	1	\$91	68	183	Medicine	m	2	\$123	118	234	Biol&Sci	m	1	\$107
19	43	Medicine	m	1	\$60	69	184	Lit&SocSci	f	3	\$37	119	235	Medicine	m	2	\$88
20	44	Eng&Prof	m	3	\$55	70	185	Lit&SocSci	m	1	\$106	120	236	Medicine	m	2	\$87
21	45	Biol&Sci	m	2	\$55	71	186	Lit&SocSci	m	1	\$91	121	237	Lit&SocSci	f	2	\$43
22	46	Medicine	f	1	\$106	72	187	Lit&SocSci	m	1	\$78	122	238	Lit&SocSci	m	1	\$79
23	47	Medicine	m	1	\$116	73	188	Biol&Sci	m	1	\$77	123	239	Medicine	m	2	\$113
24	48	Medicine	m	3	\$79	74	189	Medicine	m	1	\$90	124	240	Medicine	m	3	\$55
25	49	Lit&SocSci	m	1	\$61	75	190	Eng&Prof	m	2	\$71	125	280	Medicine	m	3	\$57
26	50	Lit&SocSci	f	3	\$37	76	191	Medicine	f	3	\$42	126	281	Eng&Prof	m	3	\$56
27	51	Medicine	m	2	\$72	77	192	Medicine	f	2	\$59	127	282	Eng&Prof	m	2	\$65
28	52	Eng&Prof	m	1	\$105	78	193	Eng&Prof	m	2	\$49	128	283	Medicine	m	2	\$42
29	59	Medicine	m	2	\$79	79	194	Biol&Sci	m	1	\$83	129	284	Medicine	m	1	\$102
30	133	Medicine	m	1	\$61	80	195	Lit&SocSci	m	1	\$34	130	285	Medicine	f	3	\$40
31	134	Medicine	m	1	\$86	81	196	Medicine	f	3	\$42	131	286	Eng&Prof	m	3	\$53
32	135	Biol&Sci	m	1	\$103	82	197	Medicine	m	2	\$97	132	287	Medicine	m	3	\$82
33	136	Lit&SocSci	m	1	\$48	83	198	Medicine	m	1	\$109	133	288	Medicine	m	2	\$64
34	137	Eng&Prof	m	2	\$64	84	199	Lit&SocSci	f	2	\$48	134	289	Eng&Prof	m	1	\$72
35	138	Eng&Prof	m	1	\$78	85	200	Medicine	m	1	\$47	135	290	Biol&Sci	f	3	\$36
36	139	Medicine	f	2	\$53	86	201	Eng&Prof	m	2	\$45	136	291	Lit&SocSci	f	1	\$66
37	140	Biol&Sci	m	1	\$85	87	202	Medicine	m	3	\$83	137	292	Medicine	f	3	\$66
38	141	Eng&Prof	m	1	\$61	88	203	Medicine	m	2	\$51	138	293	Medicine	m	2	\$102
39	142	Medicine	m	1	\$106	89	204	Biol&Sci	m	1	\$78	139	294	Biol&Sci	m	1	\$103
40	143	Lit&SocSci	m	2	\$60	90	205	Lit&SocSci	m	1	\$70	140	295	Medicine	m	1	\$148
41	144	Biol&Sci	f	1	\$73	91	206	Eng&Prof	f	2	\$46	141	296	Lit&SocSci	f	1	\$60
42	145	Medicine	m	1	\$70	92	207	Eng&Prof	m	1	\$85	142	297	Lit&SocSci	f	3	\$46
43	147	Medicine	f	3	\$32	93	208	Lit&SocSci	m	1	\$53	143	298	Lit&SocSci	f	1	\$57
44	148	Lit&SocSci	m	2	\$49	94	209	Medicine	f	3	\$40	144	299	Medicine	f	2	\$50
45	149	Eng&Prof	m	3	\$43	95	210	Eng&Prof	m	1	\$87	145	300	Lit&SocSci	m	1	\$90
46	150	Medicine	m	1	\$75	96	211	Lit&SocSci	m	1	\$71	146	301	Eng&Prof	m	3	\$63
47	151	Lit&SocSci	m	1	\$92	97	212	Medicine	m	1	\$75	147	303	Eng&Prof	m	1	\$80
48	152	Medicine	m	2	\$107	98	214	Biol&Sci	m	1	\$85	148	304	Medicine	m	3	\$56
49	153	Biol&Sci	m	2	\$57	99	215	Lit&SocSci	m	2	\$50	149	305	Medicine	m	1	\$72
50	154	Medicine	m	2	\$114	100	216	Medicine	m	3	\$118	150	306	Eng&Prof	m	1	\$96



Survey Data Collection & Analytic Specialization

Sampling People, Records, & Networks

Table of Random Numbers

	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45	46-50
1	57982	03250	09488	12085	38325	01091	03826	77648	57881	04677
2	34104	88576	32549	09603	73860	43758	05470	38344	04492	10730
3	01943	29119	72315	70334	98029	80936	37792	45017	18819	84787
4	25329	02138	52670	11692	03654	91109	17357	27013	32767	38642
5	23869	85541	29993	57693	84597	43545	23014	06089	52569	93160
6	69419	03843	33691	19541	80307	96350	62760	07467	12970	83870
7	14078	76539	08538	17888	65481	03995	63362	29308	23196	51186
8	14339	07381	13958	46562	75860	66994	97734	83135	77941	39411
9	97670	49015	97498	01260	63189	29044	11089	81803	87889	81923
10	15279	11006	82780	74184	34105	09246	31268	91081	99324	68381
11	97947	01993	12603	63485	16937	06190	77536	18598	92322	21257
12	00631	55073	73183	05961	16402	39801	96560	88053	70254	73836
13	30246	12514	24792	46736	61024	87174	63045	35815	57531	41129
14	74016	75832	78113	18374	78851	46324	78629	54165	73712	08415
15	53669	78559	66231	29085	78506	80934	88520	86912	17774	29711
16	98280	13316	23734	63720	50355	74201	17325	69300	28342	91793
17	32396	30527	10030	69065	87877	81658	23343	87272	74683	86065
18	05708	99540	11289	09030	94001	05587	41643	40782	25630	48366
19	04004	49911	47941	81165	14505	07432	79939	01309	79165	57871
20	84424	61863	81655	87420	76149	74712	81555	82645	05850	54779
21	84546	16063	09051	31787	57235	27302	83322	48361	11630	80174
22	76882	00300	45868	17098	53385	83121	70986	70598	91543	44137
23	98169	25789	99043	73890	35327	91723	77932	77971	41355	19562
24	33621	78638	27009	29610	20171	74951	34482	89391	81210	86209
25	61507	79038	38331	70320	73347	57565	19920	41574	68935	79233
26	20920	81284	01289	66936	60782	71442	63257	45294	11141	15333
27	78861	84019	31418	64943	42169	88995	48423	95960	27581	11679
28	08307	99947	51173	87234	60693	06143	13459	59514	79507	38474
29	99441	02441	66284	05523	19477	51695	58978	92845	69511	81076
30	79407	56191	54855	83823	33322	96478	39901	82641	38556	56564
31	34643	05408	67404	45254	37951	12189	76651	51762	59986	83559
32	79071	00901	91791	64053	47239	64456	10573	91720	32152	70961
33	39776	75808	19351	39680	75676	29758	11181	48497	06845	08877
34	39404	45516	97619	49788	96815	22561	01201	50766	19787	53701
35	48094	09146	14671	56419	47127	53938	80707	35195	31953	27177
36	18410	40494	13931	61713	27618	34620	72264	38473	21435	08084
37	67160	05126	99743	18590	94419	32835	22290	76000	81239	43728
38	08555	37978	24194	23612	83749	24952	05356	66661	57819	99209
39	68035	97917	09711	82240	05453	43354	44223	19470	20455	93951
40	12989	19518	39150	20121	60930	48430	29572	66745	10380	89898
41	54634	90880	59787	00767	25687	88036	32752	27228	52429	14533
42	22603	31768	32309	68558	90881	87299	30044	06782	43324	65189
43	59743	43696	26720	27716	39721	74953	90138	90812	86546	07429
44	97687	86485	58473	00533	12837	74736	31091	42786	44844	10143
45	45106	28013	06476	46786	61360	15114	00315	45583	85248	59929
46	05655	11433	54983	86786	26287	75827	57443	70597	38351	91876
47	51610	48745	02202	10037	94816	87585	15188	64636	70648	50783
48	86277	69022	37506	91519	22285	04022	49127	77337	93702	80773
49	51243	81450	22865	61165	40356	58407	82449	40460	28950	00041
50	72699	19280	12900	57824	58612	00859	35035	82783	15422	45453

One possible sample

No.	Seq. No.	Income	
1	341	42	
2	19	60	
3	253	134	
4	238	80	
5	140	148	
6	143	57	
7	152	37	
8	6	113	
9	302	89	
10	323	81	
11	57	64	
12	40	60	
13	336	41	
14	209	44	
15	83	109	
16	346	106	
17	184	87	
18	85	47	
19	129	102	
20	226	71	
Sum		1572	
Mean		78.6	(74.07)

1 Population

2 Frame

3 Sample

4 Estimate

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$



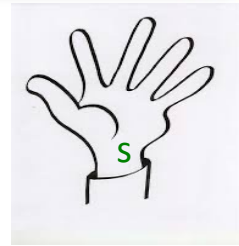
Sample mean

$$\bar{y} = 78.6$$

The means differ because they come from different samples, and the salaries differ across the sample elements in the two samples.

\bar{y}

Since each mean is based on a sample, and not a census, it will not be equal to the overall population mean, nor will means from different samples be equal to one another.



1 Population

2 Frame

3 Sample

3 Sample

3 Sample

4 Estimate

4 Estimate

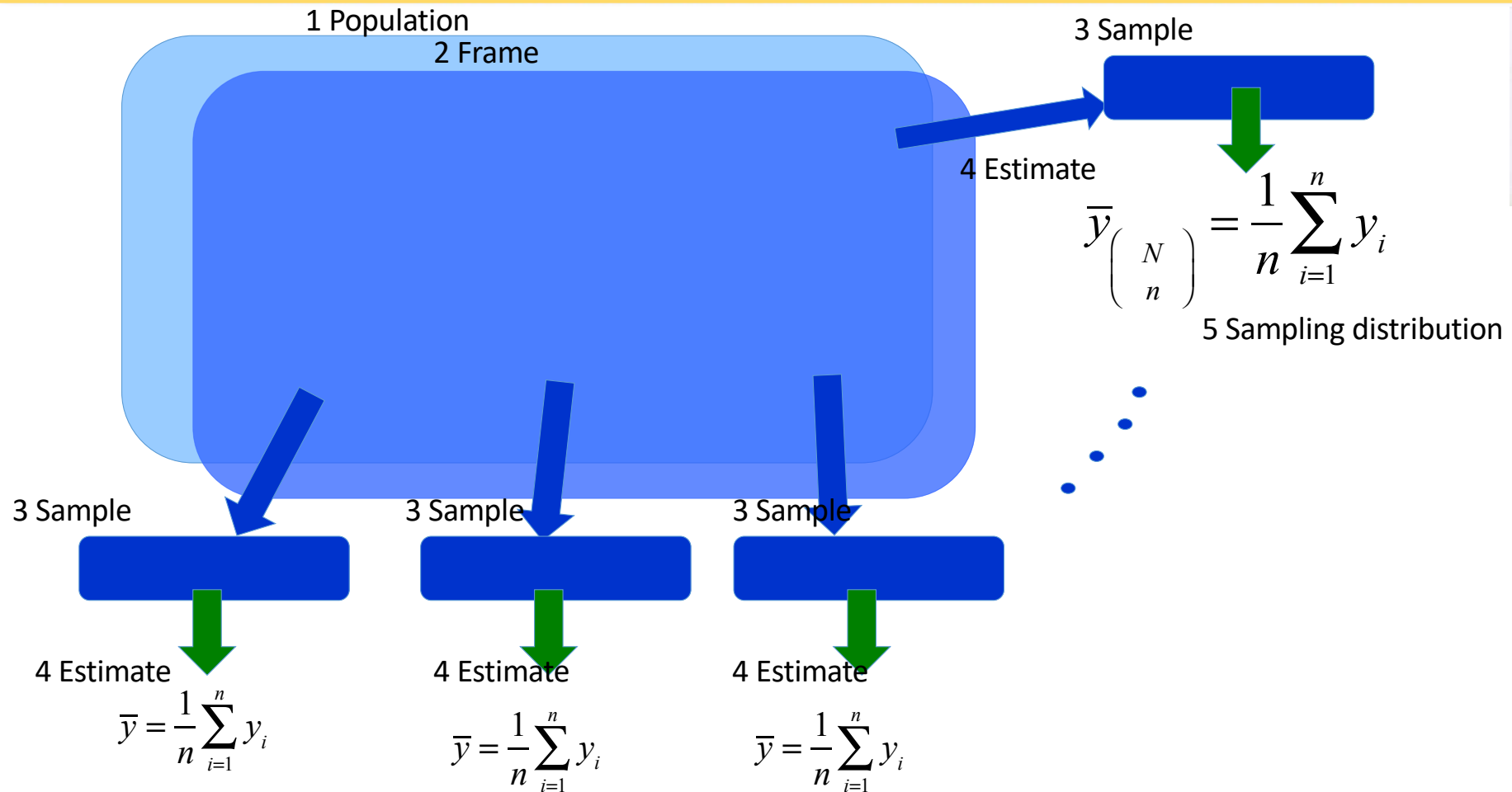
4 Estimate

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

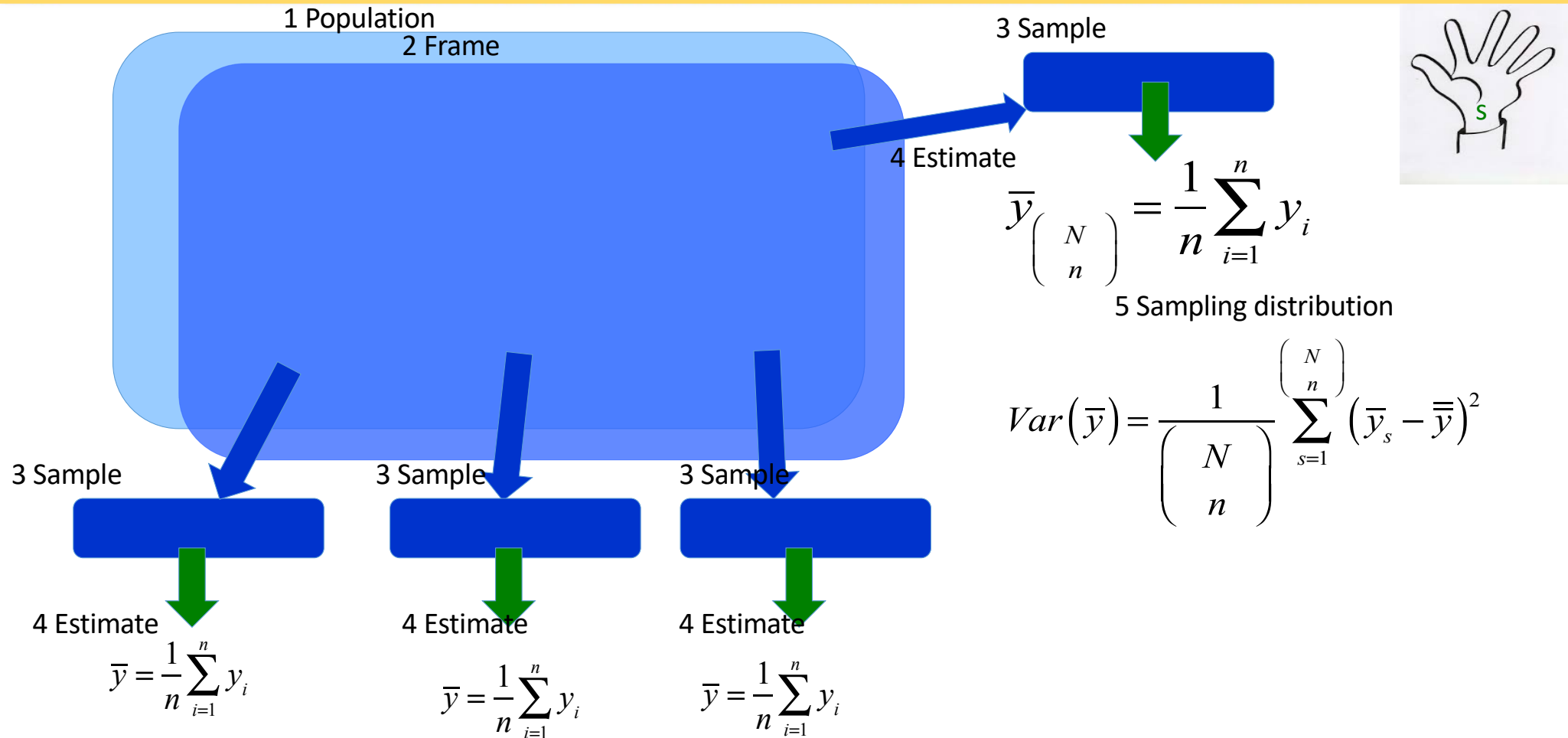
$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

Probability sampling principles



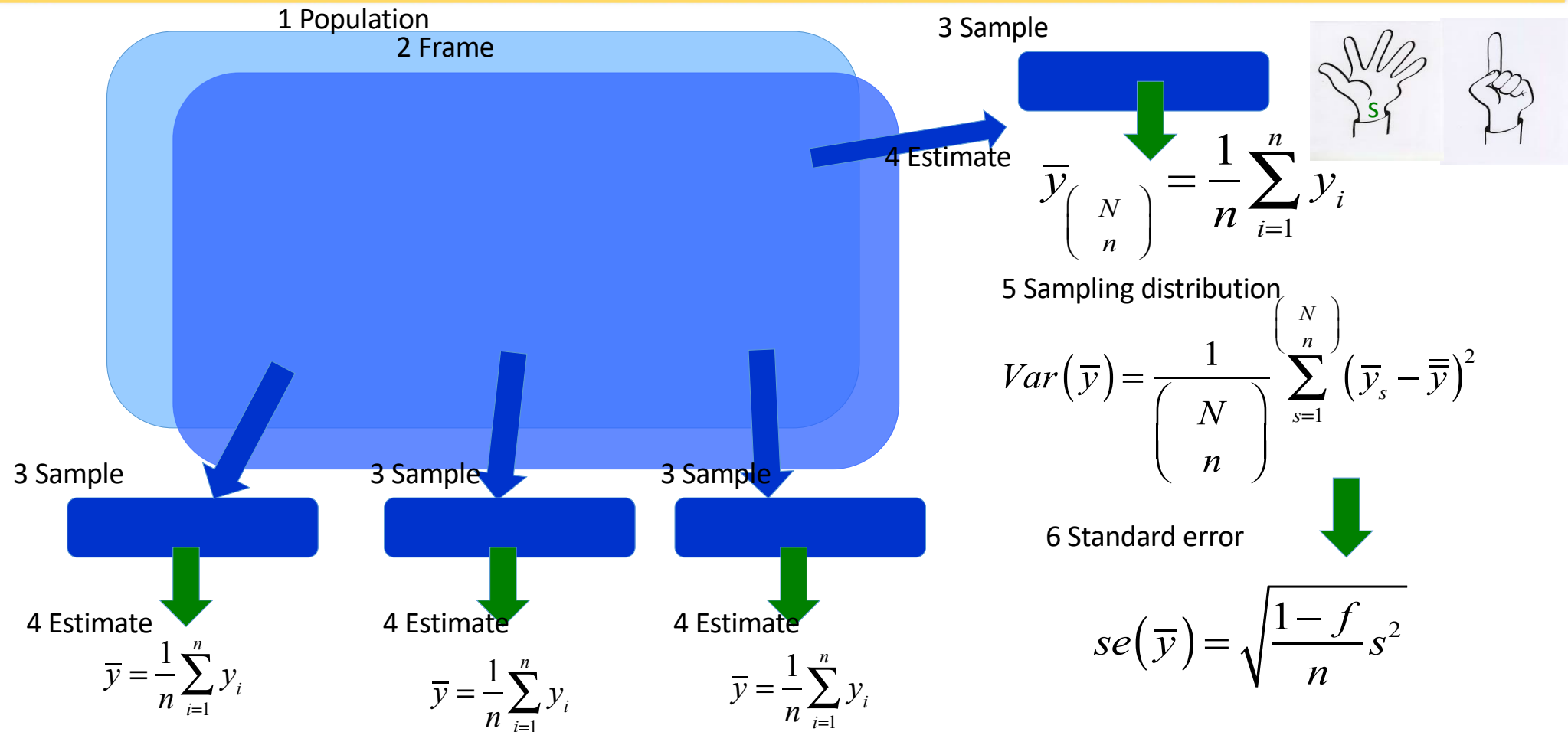
Survey Data Collection & Analytic Specialization

Sampling People, Records, & Networks



Survey Data Collection & Analytic Specialization

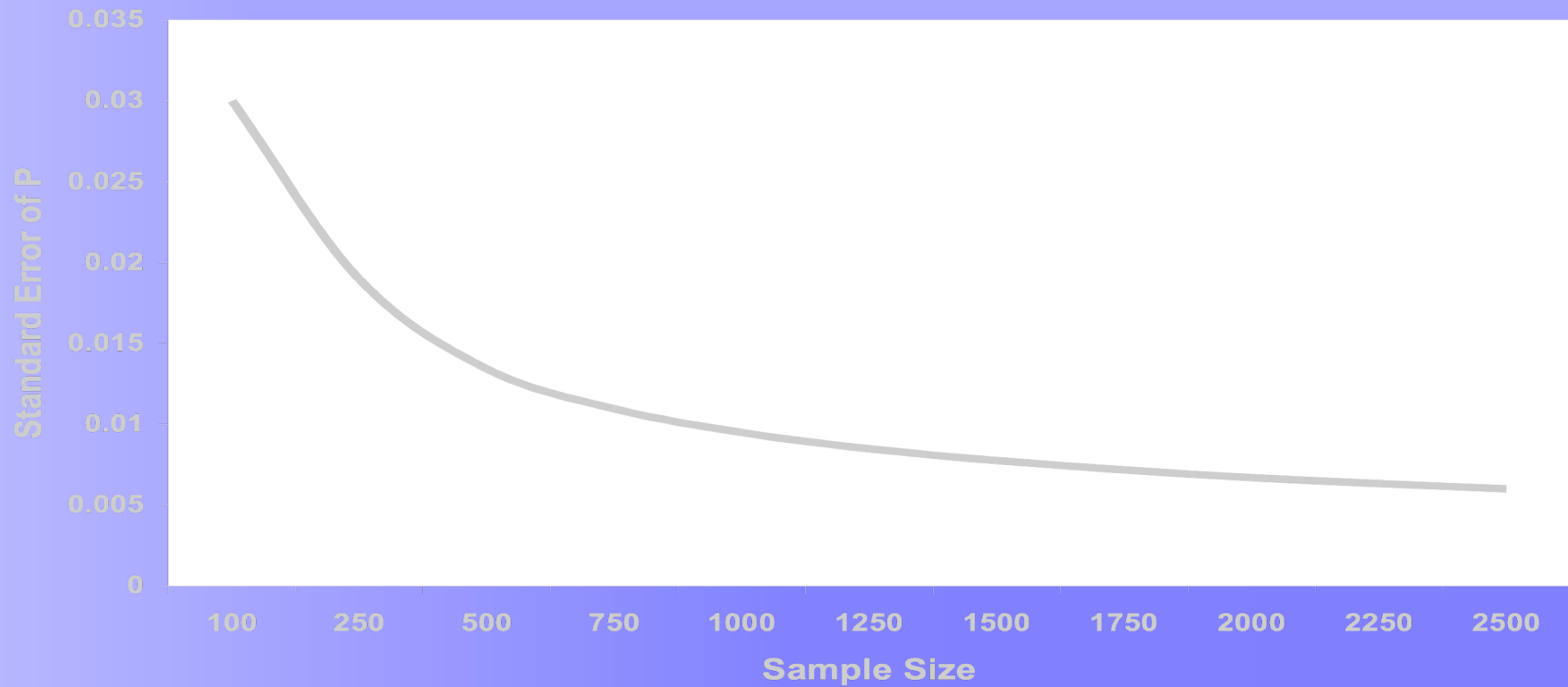
Sampling People, Records, & Networks



Standard errors

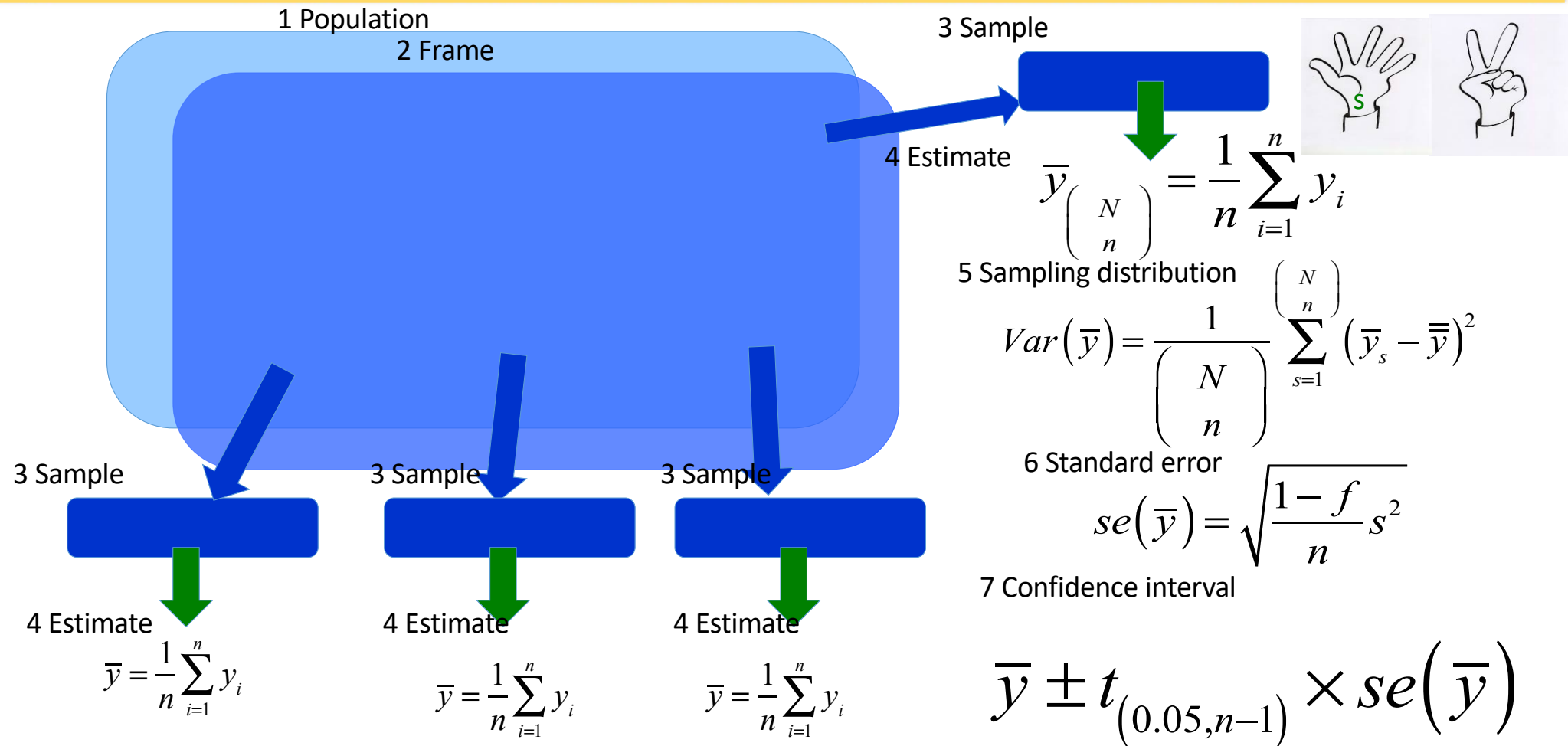
For our SRS of $n = 20$,

$$\begin{aligned} \text{var}(\bar{y}) &= \frac{(1-f)}{n} s^2 \\ &= \frac{\left(1 - \frac{20}{370}\right)}{20} 766.62 \\ &= 36.26 \\ \text{se}(\bar{y}) &= \sqrt{\text{var}(\bar{y})} = 6.02 \end{aligned}$$



Survey Data Collection & Analytic Specialization

Sampling People, Records, & Networks



Probability sampling principles

95% Confidence interval

For our SRS of $n = 20$,

$$\bar{y} \pm t_{(1-\alpha/2, n-1)} se(\bar{y})$$

$$78.6 \pm t_{(0.975, 19)} \times 6.02$$

$$78.6 \pm 2.09 \times 6.02$$

$$(66.0, 98.2)$$

- Population to frame to sample
- One random sample
- Many possible samples
- **A random process ...**
 - Using random digits applied to a frame to generate, in theory, a large number of possible samples
 - And we can measure the variance across all possible samples from just a single randomly drawn sample
 - But only random samples allow us to do this without making any assumptions about either ...
 - The sampling mechanism
 - The population distribution

Unit I

- 1 Research designs ...
 - 2 Surveys
 - 3 Why sample?
 - 4 Why randomize?
 - 5 Types of sampling
 - 6 Evaluating samples
 - 7 Units sampled
- **Unit I: Sampling as a research tool**
 - Lecture 1 – Research design & sampling
 - Lecture 2 – Surveys & sampling
 - Lecture 3 -- Why sample at all?
 - Lecture 4 – Why might we randomize, and how do we do it?
 - Lecture 5 – What happens when we randomize?
 - **Lecture 6 – How do we evaluate how good the sample is?**
 - Lecture 7 – What kinds of things can we sample?
 - **Unit 2: Mere randomization**
 - **Unit 3: Saving money**
 - **Unit 4: Being more efficient**
 - **Unit 5: Simplifying sampling**
 - **Unit 6: Some extensions & applications**

