Survey Research and Sampling Theory

- Survey research: relies primarily on probability sampling
- Uses findings from a smaller sample to generalize to a larger target population
- Focus on ***sampling theory***

Sampling Theory

- ***Descriptive***: describing attributes of particular sample
- ***Inferential***: generalizing beyond sample to larger population

What is sampling?

- Process of selecting subset of observations from among many possible observations to draw conclusions about a larger population
- ***Representativeness***: each individual in larger population has an equal chance of being chosen for sample

Benefits of sampling

1) Sample is representative

2) Ability to calculate accuracy of sample (standard error)

9 Definitions

1) ***Population***: target population
2) ***Study population***: population used for study
3) ***Sampling unit***: units used for selection
4) ***Sampling frame***: list of sampling units used
5) ***Observation unit***: person from which data are collected
6) ***Binomial variable***: variable with two responses

9 definitions

7) ***Statistic vs. parameter***: summary description of variable in a sample vs. summary description in target population
8) ***Sampling error (s.e.)***: determines accuracy of sample
9) ***Confidence levels***: way of expressing degrees of confidence

Statistic vs. parameter

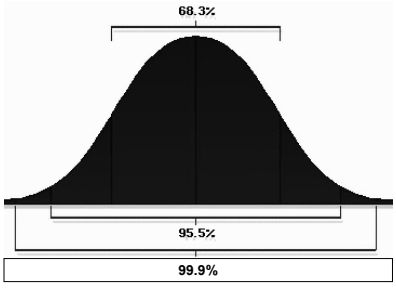
	Sample Statistic	Population Parameter
Mean	$\bar{x}$	μ (mu)
Standard deviation	s	δ (sigma)

Standard error

$$s.e. = \sqrt{\frac{p \times q}{n}}$$

p = proportion saying yes
q = proportion saying no (1-p)
n = sample size

Standard error



μ

Standard error

$$s.e. = \sqrt{\frac{.70 \times .30}{1000}}$$

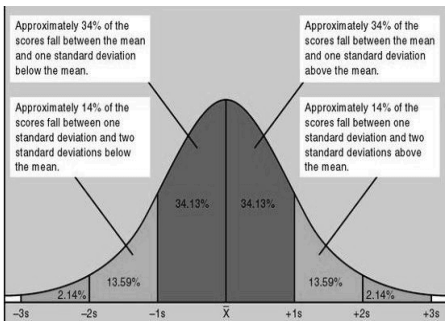
Standard error

$$s.e. = \sqrt{\frac{.21}{1000}}$$

s.e. = .01449

s.e. = 1.449, or 1.4% (.1 decimal place)

Standard error



65.8 67.2 68.6 70.0 71.4 72.8 74.2

Properties of standard errors: different values of p's and q's

p	q	n	s.e.
0.0	1.0	100	0.0
.20	.80	100	.04
.40	.60	100	.049
.50	.50	100	.05
.60	.40	100	.049
.80	.20	100	.04
1.0	0.0	100	0.0

Properties of standard errors: different sample sizes

n	p	q	s.e.
50	.5	.5	.07
100	.5	.5	.05
150	.5	.5	.04
200	.5	.5	.035
250	.5	.5	.032
1000	.5	.5	.016

Examples: Roper Public Opinion Survey

	p	q	n	s.e.
PSRA/ Newsweek (9/26-27)	.65	.35	1011	1.50%
ABC/WP (9/23-26)	.67	.33	1003	1.48%
CBS (9/22-23)	.66	.34	903	1.58%

Probability sampling designs

1) ***Simple random sample (SRS)***:
   --decide on sampling frame
   --give number to everyone on list
   --use table of random numbers to choose people

Table of Random Numbers

353	531	893	109	782	283	383	699	927	271	572	665	272	033	256	822	646	599
326	551	815	937	908	698	209	303	911	025	788	311	792	837	739	552	234	572
487	127	026	313	341	479	722	602	236	382	151	011	778	951	709	060	258	536
754	985	948	804	025	997	562	674	876	870	799	032	043	526	100	957	539	370
227	147	613	857	859	941	274	415	620	697	002	340	959	915	626	297	533	572
806	751	870	677	373	854	094	958	012	395	381	862	250	614	683	747	628	551
513	402	866	696	142	063	252	818	477	081	526	028	338	702	819	679	829	909
806	644	284	010	686	076	915	751	214	470	879	935	141	398	387	734	589	212
207	958	295	175	396	117	918	037	319	432	641	020	612	068	547	948	278	020
620	690	036	654	078	918	721	454	671	995	596	286	377	567	237	520	244	694
466	427	395	393	520	074	634	578	023	246	718	851	870	216	107	387	621	509
058	787	706	094	603	303	300	185	705	825	727	849	501	551	001	123	873	926
503	584	221	176	116	309	975	910	535	258	450	962	183	666	156	454	239	551
705	579	790	244	547	495	973	805	251	235	783	701	378	460	398	223	688	574
026	919	327	267	616	572	620	245	257	525	008	997	885	063	340	0666	328	412
779	134	709	373	332	507	525	640	840	825	739	457	999	739	068	829	336	148
461	149	798	070	930	862	672	718	849	440	769	864	022	494	826	339	910	303
885	219	668	005	418	832	416	220	692	820	641	375	242	093	364	145	848	792
874	034	514	628	693	628	200	006	795	114	842	554	881	377	427	216	193	042
790	386	783	689	565	565	349	410	216	558	301	096	577	520	923	717	188	545
028	549	529	434	083	800	569	290	298	345	937	569	279	951	183	287	808	149
212	355	367	397	638	282	720	178	695	430	074	427	422	082	629	971	496	649
960	399	700	253	375	594	024	223	383	030	345	389	739	911	022	189	565	982
648	561	528	870	907	713	608	682	576	272	718	849	715	156	823	174	733	600
416	957	547	553	534	707	206	963	459	894	847	611	763	755	388	114	274	681
416	603	967	591	013	878	424	452	659	676	984	806	692	012	934	436	869	557
719	637	969	450	489	528	364	459	708	305	074	378	670	284	431	361	912	251
595	969	582	627	920	772	560	892	500	138	461	213	905	775	881	782	272	032
769	536	611	069	694	254	195	799	928	452	200	674	202	812	986	143	243	264
355	587	878	446	137	690	647	407	362	882	033	746	390	609	144	531	944	869
543	594	002	496	648	999	262	702	811	087	336	020	166	472	293	904	949	465
727	070	996	660	024	135	799	414	136	666	841	134	588	915	116	803	917	993
160	707	361	339	054	251	397	480	805	790	228	702	690	170	511	937	723	505
147	360	150	990	380	789	436	781	337	250	860	294	684	572	415	750	726	647
151	064	609	878	095	737	897	510	891	450	992	950	890	434	306	781	006	681

Source: http://www.irs.ustreas.gov/prod/bus_info/tax_pro/irm-part/part03/34740020.GIF

Probability sampling designs

2) ***Systematic sample with a random start:***

--simpler than SRS

--choose every k^{th} element, where k = # in population / # in sample (sampling interval)

Example: systematic sampling

- ✓Target population = 1000
- ✓Sample population = 100
- ✓k = 1000/100
- ✓k = 10

Example: systematic sampling

- ✓k = 10
- ✓Start randomly with number between 1 and k
- ✓Randomly select 6, then: 16, 26, 36 . . . through 996
- ✓Watch out for periodicity!

Periodicity

Platoon 1	**Platoon 2**	**Platoon 3**	**Platoon 4**
Sgt.	Sgt.	Sgt.	Sgt.
Corporal	Corporal	Corporal	Corporal
Corporal	Corporal	Corporal	Corporal
Private	Private	Private	Private
Private	Private	Private	Private
Private	Private	Private	Private
Private	Private	Private	Private
Private	Private	Private	Private
Private	Private	Private	Private
Private	Private	Private	Private

Probability sampling designs

3) ***Stratified sample***

--ensures different groups are adequately represented in sample

--increases accuracy in estimating population parameter

--reduces s.e.

Example: stratified sample

	population %	sample n
Anglo	70%	70
Black	20%	20
Hispanics	10%	10
Total	100%	100

Probability sampling designs

4) ***Multistage cluster sample***:

--used when there is no list of names

--create a sampling frame via set of "stages"

Multistage cluster sampling: 3 stages

1) ***Stage 1***: define area using map (divided into blocks)
   --Choose blocks via SRS or systematic sampling
   --Blocks = PSU (primary sampling unit)

Multistage cluster sampling: 3 stages

2) ***Stage 2***: list and number all dwelling on selected blocks only
   --SRS or systematic sampling of dwellings
   --Dwellings = secondary sampling unit

Multistage cluster sampling: 3 stages

3) ***Stage 3***:
   --interview HH or randomly selected member of dwelling

Multistage cluster sampling: Implications

- ✓ Sampling error increased:
  --Each sampling unit increases sampling error
- ✓ General guideline: maximize number of clusters and minimize number of elements within cluster
- ✓ Why?