

Chp 8 & 9 (Producing Data: Sampling and Experiments)

Vocab

- population: entire group of individuals about which we want info.
- sample: a subset of population, from which we collect info. (we use this to hopefully draw conclusions for population)
- sampling design: describes how to choose sample from population

- ① a convenience sample: selecting individuals that are easiest to reach (bad design)
- ② voluntary response sample: people who choose themselves by responding to broad appeal; this provides a biased sample (only people w/ strong opinions respond)

- ③ SRS (simple random sample): a sample chosen such that every individual has equal chance of being selected

biased: design of a statistical study is biased if it necessarily favors certain outcomes

- Table of random digits (0, 1, ..., 9)
 - (a) Each entry is equally likely.
 - (b) Entries are independent of all other entries.
- How to use Table B to choose SRS
 - (a) Give each member of population a numerical # (of same length).
 - (b) Read successive groups of #s from table. \Rightarrow gives sample.

Chp 8 (cont)

Ex 1 There are 7014 students taking math classes at a university this semester. The department wants to survey 500 of those students.

(a) How would you label the names of the students to select an SRS?

(b) Use Table B, line 142, to select SRS of only 5 students.

(c) What is the population?

Purpose of a sample \Rightarrow to infer information about the population, based on the sample (statistical inference)

Reasons to rely on random sampling:

① eliminate bias

② Laws of probability allow trustworthy inference

* Larger random samples give more accurate info than small samples.

Chp 8 (cont)

Other sampling designs:

- ④ Stratified Random Sample: (a) classify population into groups of similar individuals (b) then choose separate SRS in each stratum (class) & combine these SRSs to form full sample

Ex 2 A study of academic dishonesty among college students used a 2-stage sampling design. The first stage chose a sample of 30 colleges. Then 200 seniors, 100 juniors and 100 sophomores were chosen from each school. One school has 989 sophomores, 943 juniors, and 895 seniors. You have alphabetical lists of these students. How do you assign labels for stratified sampling? Use Table B (line 122), to select first 5 students.

Chp 8 (cont)

Tidbits:

- undercoverage: when some groups are left out of process of selecting a sample.
- nonresponse: when individual chosen can't be contacted or refuses to participate.
- response bias: lying can occur when getting responses (for various reasons)
- wording of questions is very important, as is order of questions
- don't trust results of sample survey until you know exactly what the question was and the nonresponse rate (and date of question)

Ask a sample of college students these two questions:

"How happy are you with your life in general?" (Answers on a scale of 1 to 5)

"How many dates did you have last month?"

The correlation between answers is $r = -0.012$ when asked in this order. It appears that dating has little to do with happiness. Reverse the order of the questions, however, and $r = 0.66$. Asking a question that brings dating to mind makes dating success a big factor in happiness. ■

TABLE B Random digits

LINE								
101	19223	95034	05756	28713	96409	12531	42544	82853
102	73676	47150	99400	01927	27754	42648	82425	36290
103	45467	71709	77558	00095	32863	29485	82226	90056
104	52711	38889	93074	60227	40011	85848	48767	52573
105	95592	94007	69971	91481	60779	53791	17297	59335
106	68417	35013	15529	72765	85089	57067	50211	47487
107	82739	57890	20807	47511	81676	55300	94383	14893
108	60940	72024	17868	24943	61790	90656	87964	18883
109	36009	19365	15412	39638	85453	46816	83485	41979
110	38448	48789	18338	24697	39364	42006	76688	08708
111	81486	69487	60513	09297	00412	71238	27649	39950
112	59636	88804	04634	71197	19352	73089	84898	45785
113	62568	70206	40325	03699	71080	22553	11486	11776
114	45149	32992	75730	66280	03819	56202	02938	70915
115	61041	77684	94322	24709	73698	14526	31893	32592
116	14459	26056	31424	80371	65103	62253	50490	61181
117	38167	98532	62183	70632	23417	26185	41448	75532
118	73190	32533	04470	29669	84407	90785	65956	86382
119	95857	07118	87664	92099	58806	66979	98624	84826
120	35476	55972	39421	65850	04266	35435	43742	11937
121	71487	09984	29077	14863	61683	47052	62224	51025
122	13873	81598	95052	90908	73592	75186	87136	95761
123	54580	81507	27102	56027	55892	33063	41842	81868
124	71035	09001	43367	49497	72719	96758	27611	91596
125	96746	12149	37823	71868	18442	35119	62103	39244
126	96927	19931	36809	74192	77567	88741	48409	41903
127	43909	99477	25330	64359	40085	16925	85117	36071
128	15689	14227	06565	14374	13352	49367	81982	87209
129	36759	58984	68288	22913	18638	54303	00795	08727
130	69051	64817	87174	09517	84534	06489	87201	97245
131	05007	16632	81194	14873	04197	85576	45195	96565
132	68732	55259	84292	08796	43165	93739	31685	97150
133	45740	41807	65561	33302	07051	93623	18132	09547
134	27816	78416	18329	21337	35213	37741	04312	68508
135	66925	55658	39100	78458	11206	19876	87151	31260
136	08421	44753	77377	28744	75592	08563	79140	92454
137	53645	66812	61421	47836	12609	15373	98481	14592
138	66831	68908	40772	21558	47781	33586	79177	06928
139	55588	99404	70708	41098	43563	56934	48394	51719
140	12975	13258	13048	45144	72321	81940	00360	02428
141	96767	35964	23822	96012	94591	65194	50842	53372
142	72829	50232	97892	63408	77919	44575	24870	04178
143	88565	42628	17797	49376	61762	16953	88604	12724
144	62964	88145	83083	69453	46109	59505	69680	00900
145	19687	12633	57857	95806	09931	02150	43163	58636
146	37609	59057	66967	83401	60705	02384	90597	93600
147	54973	86278	88737	74351	47500	84552	19909	67181
148	00694	05977	19664	65441	20903	62371	22725	53340
149	71546	05233	53946	68743	72460	27601	45403	88692
150	07511	88915	41267	16853	84569	79367	32337	03316

Chp 9

Vocab

- vs
- Observational study: observes individuals + measures variables w/o influencing responses; purpose is to describe.
 - Experiment: deliberately imposes some "treatment" on individuals + measures response; purpose is to study treatment response (to see if it causes a change); useful when trying to measure cause & effect.

- randomized comparative experiment: uses comparison of 2 or more treatments and random assignment of subjects

- completely randomized experiment: all subjects are allocated at random among all treatments

- block design: random assignment of individuals to treatments carried out separately w/in each block. (a block is a group of individuals that are known to be similar in some way that's expected to affect response to treatment).

- double-blind experiment: neither the subjects nor people doing experiment know which treatment subject is receiving

Two variables are confounded when their effects on a response variable cannot be distinguished from each other.

* uncontrolled experiments often yield worthless results because of confounding w/ lurking variables.

subject: individual in an experiment

treatment: experimental condition applied to subjects

factor: explanatory variable in an experiment

Chp 9 (cont)

Tidbits

- In ^{logic} randomized comparative experiments
 - (1) random assignment yields groups that are similar in all respects
 - (2) comparative design makes sure that influences other than treatment are equal among groups
 - (3) Thus, differences in average response must be due to treatment or chance (of random assignment).
- Basic Principles of Statistical Design of Experiments
 - (1) control effects of lurking variables (by comparing 2 or more groups)
 - (2) randomize (use chance for assignments)
 - (3) use enough subjects to reduce chance of variation
- Statistically significant: an observed effect too large to be explained by chance; if a well-designed experiment has statistically significant result, it does imply causation.
- Statistical analysis of an experiment cannot necessarily be generalized!
- For block design, form blocks based on most important unavoidable sources of variability; randomization will then average out effects of remaining variation & allow unbiased comparison of treatments.

Chp 9 (cont)

Ex 1

Cell phones and brain cancer. A study of cell phones and the risk of brain cancer looked at a group of 469 people who have brain cancer. The investigators matched each cancer patient with a person of the same sex, age, and race who did not have brain cancer, then asked about use of cell phones.² Result: "Our data suggest that use of handheld cellular telephones is not associated with risk of brain cancer." Is this an observational study or an experiment? Why? What are the explanatory and response variables?

Ex 2

How long did I work? A psychologist wants to know if the difficulty of a task influences our estimate of how long we spend working at it. She designs two sets of mazes that subjects can work through on a computer. One set has easy mazes and the other has hard mazes. Subjects work until told to stop (after 6 minutes, but subjects do not know this). They are then asked to estimate how long they worked. The psychologist has 30 students available to serve as subjects.

- Describe the design of a completely randomized experiment to learn the effect of difficulty on estimated time.
- Describe the design of a matched pairs experiment using the same 30 subjects.

Chp 9 (cont)

Ex 3

Attitudes toward homeless people. Negative attitudes toward poor people are common. Are attitudes more negative when a person is homeless? To find out, read to subjects a description of a poor person. There are two versions. One begins

Jim is a 30-year-old single man. He is currently living in a small single-room apartment.

The other description begins

Jim is a 30-year-old single man. He is currently homeless and lives in a shelter for homeless people.

After reading the description, ask subjects what they believe about Jim and what they think should be done to help him. The subjects are 544 adults interviewed by telephone.¹² Outline the design of this experiment.

Ex 4

Experimental design. The clinical trial was a completely randomized experiment that assigned 240 patients at random among 4 treatments as follows:

	Antibiotic pill	Placebo pill
Steroid spray	53	64
Placebo spray	60	63

- (a) Outline the design of the experiment.
- (b) How will you label the 240 subjects?
- (c) Explain briefly how you would do the random assignment of patients to treatments. Assign the first 5 patients who will receive the first treatment.

Chp 9 (cont)

Good example!

EXAMPLE 9.9 Cell phones and driving

Does talking on a hands-free cell phone distract drivers? Undergraduate students "drove" in a high-fidelity driving simulator equipped with a hands-free cell phone. The car ahead brakes: how quickly does the subject react? Let's compare two designs for this experiment. There are 40 student subjects available.

In a *completely randomized design*, all 40 subjects are assigned at random, 20 to simply drive and the other 20 to talk on the cell phone while driving. In the *matched pairs design* that was actually used, all subjects drive both with and without using the cell phone. The two drives are on separate days to reduce carryover effects. The *order* of the two treatments is assigned at random: 20 subjects are chosen to drive first with the phone, and the remaining 20 drive first without the phone.⁸

Some subjects naturally react faster than others. The completely randomized design relies on chance to distribute the faster subjects roughly evenly between the two groups. The matched pairs design compares each subject's reaction time with and without the cell phone. This makes it easier to see the effects of using the phone. ■