

1.3 and 2.1 are both building critical thinking skills

1.3 skepticism about numbers as presented

2.1 Good sampling procedures

News reports of snowstorms say things like "A winter storm spread snow across the area, causing 28 minor traffic accidents." Eric Meyer, a reporter in Milwaukee, Wisconsin, says he often called the sheriff to gather such numbers. One day he decided to ask the sheriff how many minor accidents are typical in good weather: about 48, said the sheriff. Perhaps, says Meyer, the news should say, "Today's winter storm prevented 20 minor traffic accidents."

Crested Butte attracts skiers by advertising that it has the highest average snowfall of ski town in Colorado.

A highway billboard says simply, "Half of all heart disease victims are women."

"Of those aged more than 60 living alone, 54% are women and only 15% are men."

We had a 200% decrease in customers over the last year.

I took groups of 5 groups 10 students and found out how many in each group were democrats. Here are the results for the groups:
% Democrats: 13%, 42%, 62%, 85%, 90%

An advertisement for a home security system says, "When you go on vacation, burglars go to work. According to FBI statistics, over 26% of home burglaries take place between Memorial Day and Labor Day."

n

n

$$\frac{3}{12} \approx 25\%$$

A writer in Science claimed in 1976 that "people over 65, now numbering 10 million, will number 30 million by the year 2000, and will constitute an unprecedented 25 percent of the population." Make sense?

10
out of ?

30
out of ?

120

Colleges know that many prospective students look at popular guidebooks to decide where to apply for admission. The guidebooks print information supplied by the colleges themselves.

Surely no college would simply lie about, say, the average SAT score of its entering students. But we do want our scores to look good. How about leaving out the scores of our international and remedial students? Northeastern University did this, making the average SAT score of its freshman class 50 points higher than with all students included. If we admit economically disadvantaged students under a special program sponsored by the state, surely no one will complain if we leave their SAT scores out of our average? New York University did this.

Use rates not counts
Using rate of change

A rate that doesn't make sense
Numbers that don't compare well
Poor descriptions of the number

Study shows Doctors over worked.

A recent study revealed that the average primary care physician saw 600 patients a week. This was a 3,000% increase over 20 patients that the average doctor saw in 1950s.

→ 120/hy
↳

% Incr and % Decr

$$\frac{E - S}{S} = \frac{600 - 20}{20} = 29.00$$

⇒ 2,900%

Early $\frac{20}{5} = 4$ patients

Elderly overweight

A recent study showed that the elderly are more overweight than ever.
20 years ago the average retiree had a BMI of 23. (Normal range is 18.5 to 24.9).

The study found that in 2012 the average retiree had a BMI of 27, or an increase of 17%.

In terms of actual weight, that is an average increase of 70 lbs.

$$\frac{27-23}{23} = 17\%$$

$$.17 = \frac{S-5}{5}$$

$$.17 = \frac{70}{5}$$

$$\frac{.17}{1} = \frac{70}{5}$$

$$\frac{.17 \cdot 5 = 70}{.17}$$

$$5 = \frac{70}{.17}$$

$$5 = 411. \sim$$

Jurors are less certain than you think!

The law says to jurors that guilty should be "beyond the shadow of a doubt."

A recent study of jurors showed that on average jurors were only 80% certain that lawyers had proved their case.

Things that are true about Fairfield, CT:

Population: 59,404

Total land area: 19,200 acres

Total Value of assessed Real Estate: \$14.78 bil.

$$\frac{59,404 \text{ p}}{19,200 \text{ a}} \approx 3 \text{ p/a} \quad \frac{19,200 \text{ a}}{59,404 \text{ p}} \approx \frac{1}{3} \text{ a/p}$$

Things that are true about Westport, CT:

Population: 26,391

Total land area: 12,800 acres

Total Value of assessed Real Estate: \$14.74 bil.

$$\frac{26,391 \text{ p}}{12,800 \text{ a}} = 2.06$$

Make critical comment on the reasonableness of each of the following claims:

On average the real estate owned by each household in Westport is, roughly, twice as valuable.

Fairfielders put in 21,682,460 working days last year.

$$\frac{4,102}{59,404} \approx 0.069$$

Although Westport property is worth more, the average Fairfield household enjoys an equal amount of living space.

There are 4,102 unemployed fathers in Fairfield.

$$\frac{59,404}{3} \approx 19,800 \quad \frac{4,102}{19,800} \approx 0.207$$

$$\frac{21,682,460}{59,404} = 363$$

	Day 1	Day 2	Day 3		
IBN	\$52	\$55.12	\$58.43	$\frac{6.43}{52}$	12.7%
ipple	\$80	\$84	\$88.2	$\frac{8.2}{80}$	10.25%
S&P 500	1,100	1,155	1,212.75	$\frac{52.75}{1100}$	4.79%

Comment on this claim: ipple is a better investment than IBN because its stock price went up \$8.2 vs IBN's \$6.43

Percentage change calculation:

$$\frac{\text{End} - \text{Start}}{\text{Start}} = \text{Percentage Change}$$

Just like rates are better than counts almost all the time. Percentage changes are more meaningful than absolute changes almost all the time.

Chapter 2 Sampling well and badly

One of the most important steps in any study is deciding how to select your sample.

Vocab

- Bias
- Voluntary response sample
- Convenience sample
- SRS
- Independent

The major topic is **BIAS**.
You will need to know:
Types of biases
Ways to avoid bias

BIAS: The unavoidable sampling enemy

A biased sample:

Selects a group that **systematically favors a certain outcome**,

When a sample is biased you can say:

That you **expect** this group of individuals **to measure differently than the typical group**.

If I ask a set of Republicans what they think of Barack Obama, do I know what to expect?

I expect them to be more unfavorable than average.

If I ask people in this school who will win the Ludlowe vs Warde game, do I expect an accurate prediction?

I expect them to make a stronger prediction of Ludlowe victory than the average person in town.



Key point on bias.

If I ask military people the following question:

Have all the US wars been necessary?

"Potential"

1) Do I expect a biased response?

2) Does that mean I expect the majority of them to say "yes"?

No, the bias means that I just expect them to say yes more than the average person.

If 15% of people would say "Yes" to this question, then I might expect 18% of military people to say "yes".

OR 12%, but I would bet it is not 15%

Remember:

Before => "POTENTIAL bias"

After

"There was bias."

or

"There may have been bias."

The SRS - the ideal sample.

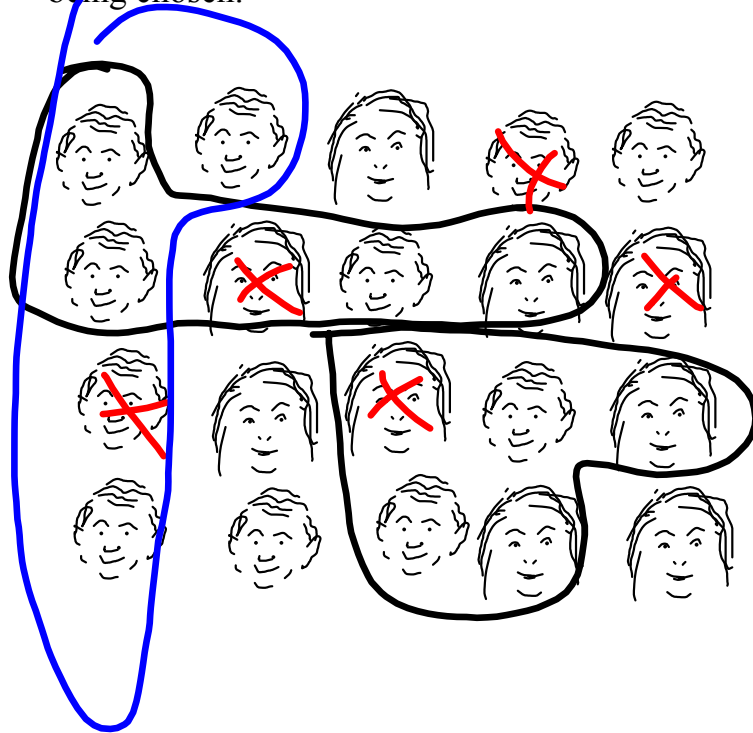
Best unbiased design is to use an SRS



SIMPLE RANDOM SAMPLE definition is tricky.

every group n from the population has the same chance of being chosen.

!!!DOES NOT SAY EVERY INDIVIDUAL HAS THE SAME CHANCE!!!!



How could I pick a simple random sample of 5 kids from this class?

- Randomly pick one from each row?
- Send them out in the hall and take the first five who come in.
- Take five kids whose birthday date number are smallest.

Have them each pick a number from a bag and then use a random number generator to pick 5 numbers from the set.

:Have them each pick a number from a bag and then use a random number generator to pick 5 numbers from the set.

SRS = NO RULE = random choice.

- Smell a rule, a decision factor, an influence, not an SRS

- Groups not individuals =>

Could any two people be prevented from being in the group?

Pick by a random process - use a random number generator

The Vietnam lottery example shows how randomness does not just happen but takes a lot of thought.

Probably good enough does not cut it.

Rules, patterns, biases are hard to fight off.

RECAP:

How to pick your sample.

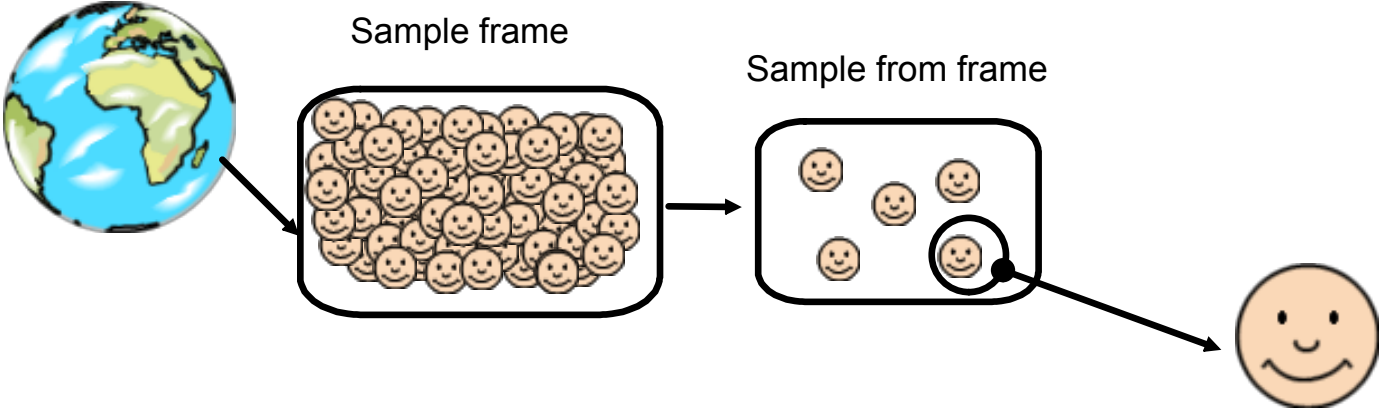
Avoid bias from bad sample picking.

The right way SRS

Assign numbers to sample frame

Pick randomly

Wrong way



Labels for biased samples

Voluntary response sample: The sample choose itself by responding to a general appeal.
See call-in shows and internet surveys.

Convenience sample: You selecting those individuals easily accessed by you.

Why would a voluntary response sample be biased?

- 1) Followers of whose asking. "Fox News watchers say..."
- 2) People with an interest more likely to invest time.
"What do you think of the new highway being put through Fairfield?"

2.2, 2.4

Isn't every survey voluntary response? You can't make people answer.

Not a voluntary response sample if you choose the individuals

Voluntary response if they individuals chose themselves.

The is a name for the people you approach refusing to cooperate: Nonresponse. That is also a potential source of bias.

Why would a convenience sample be biased?

All individuals have some similar circumstance:

What made them convenient gives them things in common.

Neighborhoods:

- Similar wealth
- Same geographic features (close to town, highway etc.)
- Size/value/age of real estate similar
- They all fall under the same zoning law.

Convenience not just people. I could study plants in my lawn.

Describe the bias of standing in downtown Fairfield and asking:

"Should the current deficit be addressed with increased revenue, taxes, or spending cuts?"



Voluntary response and convenience sample are not SRS.

Not equally likely to get any group size n from the population.

Voluntary response and convenience are popular ways to sample badly
Many other sources of POTENTIAL BIAS

Quick vocab book doesn't yet:

The Sample Frame: The group you are selecting your sample from.

If I want to poll Fairfield, I might use the Fairfield phone book even though that is not everyone in Fairfield. That is the sample frame.

How to get an SRS:

- 1) Assign numbers to the individuals in the sample frame. ("Label")
- 2) Use a source of random numbers to pick. ("Table")

Two practical sources of random numbers:

Random number table

Computer



Using a random number table

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

Key decision: How many digits do you need? Answer: How many did you label in your sample frame?

10 things in sample frame, 1 digit (0-9)

11 - 100, 2 digits (10-99)

101-1000, 3 digits (100-999)

Start left to right, taking sets of that many digits.

You don't get a winner every time and sometimes have to move to the next digit(s).

- If you already picked the item skip.
- If the digits give you a number bigger than everyone in your frame.

Let's randomly pick people in this class.



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	75552
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	11957
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	91546
125	96746	12149	37823	71868	18442	35119	62103	37244
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	87200
129	36759	58984	68288	22913	18638	54303	00735	08710
130	69051	64817	57174	09517	84534	06480	87200	97245
131	05007	16632	81194	14873	04097	85576	45295	48801
132	68732	55259	84292	08796	47065	47120	71001	71001

Statistics for East Cupcake

	Percentage of population	Total earnings
Adult Males	42%	\$36 mil.
Adult Females	24%	\$21.6 mil.
Children	36%	\$2 mil.

A) The average number of children per household is 4.3.

B) The much larger total earnings for men show that there is clear discrimination in the wages for men and women.

C) The total earnings of the males in West Cupcake is \$36.5 mil., so males in West Cupcake are slightly better earners than males in East Cupcake.



Examination on Wednesday

Reliability vs. Bias vs. validity.
Finding issues with data
Calculate percent increase and decrease

Sampling Bias

Voluntary response sample

Convenience sample

SRS

Using a random number table

Notes:

Great questions on random number table.

You pick the number of digits based on the group you are picking from, not how many you pick.

The 000 thing is good to understand. I will not try to trip you up on that.

1) mini- two to discuss

2) Mini in progress

3) exercise

Explaining your steps:


1) Say what numbers you assigned and how you assigned them to the sample frame (specifying the sample frame).

I assigned the number 01 to 24 to the students on the class attendance list in alphabetical order.

2) Name your random number generator

I used the random number table to select 5 people.

3) Describe fully how you used it, so that some one could repeat your steps.

You could use this ●  model for any random number table exercise.

I started on line 109.
I took the two digits at a time left to right.
If they matched a number I hadn't picked yet I selected the student with that number for my sample.
If the number was too large or already picked, I read the next to numbers.
I did this until I had picked five students.

I had my calculator generate integers from 1 to 24 until I had 5 different numbers.
I selected the students with those five numbers for my sample.

If it is not an SRS and you smell a rat, but we haven't named that kind of bias, how do you explain the potential bias?

Say:

- 1) There is POTENTIAL bias
- 2) Give a specific example of how the bias could arise
- 3) Say what the affect would be on your measurement
- 4) Say what the affect would be on your conclusion.

A medical firm wants to study sexual behavior. They recruit young people and train them to do phone interviews on certain sexual behavior. The training is all about making sure the interviewer eases into the subject, giving proper background and using a sequence that gives the interviewee plenty of opportunity to back out before being offended. The interviewers are given a street corner and told they will be paid by how many full interviews they complete.

Foreshadowing two important issues on SRSs.

1) Just because it is an SRS does not mean it is unbiased. Pick a bad sample frame and you have a problem.

What are the problems with using a phone book these days?
What specific bias could arise?

The fix: Just take care the sample frame truly represents the population.

2) Sometimes you don't want an SRS.
(You never want bias, but....)

What is the problem with randomly selecting 20 students from the high school to ask an opinion on almost anything?
(Hint: What unlucky thing might happen?)

The fix: SRS the same amount out of the subgroups.
We will discuss later.

Independent: Two things are independent when knowing anything about one does not make you expect anything about the other.