AP Statistics – Lesson Notes - Chapter 11: Understanding Randomness

How do we obtain data to analyze?

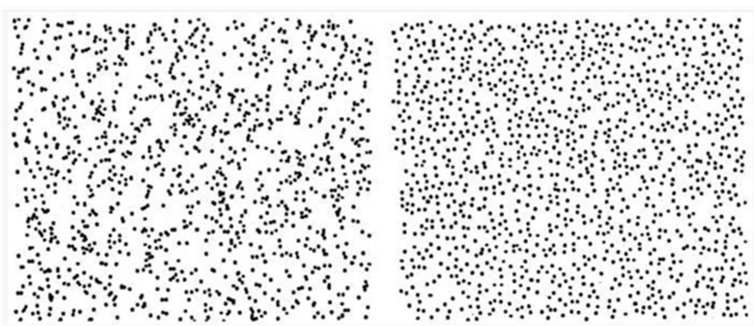
- **Taking samples of the population you are studying** **Ch12**
  - **Observational Studies:** Observing what is already happening. May suggest associations, but can't be used to establish cause-and-effect due to uncontrolled factors. **Ch13**
  - **Experiments:** Controlled situation designed to eliminate uncontrolled factors. Can be used to establish cause-and-effect. **Ch13**
- **Simulations:** Using randomness to generate a fake but representative data set which matches reality. Used when it is unethical or too costly to collect actual data. **Ch11**

Experiments and Simulations always involve the use of randomization, so we talk about randomness and simulations first.

The Federalist Papers Activity

Which dot grouping is random?

A B



You can't rely on humans to produce random results

People tend to assume random means 'evenly distributed' but that is not always the case. Our brains are wired to quickly see patterns and to view them as unusual.

<u>You can't rely on selections by people to be truly random</u>.

If we can't trust our intuition, how can we get truly random information?

- Write selections on identical pieces of paper, put in a box, mix thoroughly, and randomly select some of the papers.
- Use fair dice rolls.
- Use a calculator's random data functions.
- Use a random digit table.
- Use spinners.

Random Digit Tables

We use charts of truly random numerical digits. These are available in the back of our textbook and on websites:

40603	16152	83235	37361	98783	24838	39793	80954	76865	32713
40941	53585	69958	60916	71018	90561	84505	53980	64735	85140
73505	83472	55953	17957	11446	22618	34771	25777	27064	13526
39412	16013	11442	89320	11307	49396	39805	12249	57656	88686
57994	76748	54627	48511	78646	33287	35524	54522	08795	56273
61834	59199	15469	82285	84164	91333	90954	87186	31598	25942
91402	77227	79516	21007	58602	81418	87838	18443	76162	51146
58299	83880	20125	10794	37780	61705	18276	99041	78135	99661
40684	99948	33880	76413	63839	71371	32392	51812	48248	96419
75978	64298	08074	62055	73864	01926	78374	15741	74452	49954

These tables are sometimes generated by monitoring truly random physical processes (like time between radioactive particle ejections) or by running tests on generated digit sequences to verify the absence of patterns.

We can assign digits or combinations of digits to particular outcomes and to run 'trials' by using a random digit table to dictate what randomly happens for each trial.

Calculator Random Functions

You can also use your calculator to generate random numbers individually or to populate lists with random numbers.

[Math, PRB]
randInt(0,9) produces random integers from 0 to 9.
randInt(0,56,3) produces a group of 3 random integers between 0 and 56.
randInt(0,99,100)->L1 loads list 1 with 100 random integers between 0 and 99.

randNorm(80,7,100)->L1 loads list 1 with 100 random real numbers sampled from a normal population with mean = 80, SD = 7.

Your calculator has a very long random digit table stored and is just reading numbers from it. To make it random from calculator to calculator, every person needs to start at a different place in the table by 'seeding' the calculator:

(any positive integer), STO ->, Math/PRB/rand

Simulations

Consider this scenario: a cereal manufacturer puts pictures of famous athletes on cards in boxes of cereal. 20% of the boxes contain Tiger Woods cards, 30% contain Lance Armstrong, and the rest contain Serena Williams.

If you want all three cards, how many boxes of cereal would you expect to have to buy?

One way to answer this question would be to actually buy hundreds of boxes of cereal and keep track of how many boxes it took to get a set. But this is expensive and inefficient. It would be better to 'simulate' buying a large number of boxes.

We can use a model of the situation to imitate cards found in random boxes of cereal.

Some terms...

Simulation - a sequence of random outcomes that model a situation (buying a number of random cereal boxes).

Component - the most basic event in a simulation (the selection of a particular box of cereal).

Outcome - possible results a component can have (Tiger, Lance, or Serena).

Model - shows a linkage between results from the random process and outcomes.

Trial - a sequence of events we want to investigate (simulating buying boxes until we get one of each card type).

Response variable - what we record to be able to answer the question.

Steps for setting up a simulation:

1) Identify the component to be repeated.
Selection of a box of cereal.

2) Explain how you will model the outcome.
In a random digit table, the digits 0-9 are equally likely to occur. The cards are distributed as follows: 20% Tiger, 30% Lance, 50% Serena so we will assign digits to outcomes as follows:

0,1=Tiger   2,3,4 = Lance   5,6,7,8,9=Serena

3) Explain how you will simulate a trial.
We are simulating a trial of opening boxes until we have one of each card. We'll do this by looking at digits in sequence from the random digit table, and determining by the digit what card we obtained. We'll stop when we've obtained one of each card.

4) State clearly what the response variable is.
The response variable will measure how many boxes were opened to obtain all 3 cards.

5) Run many trials.

An example simulation

40603 16152 83235 37361 98783 24838 39793 80954 76865 32713

Trial #	Outcomes	Response Variable value
1		
2		
3		
4		
5		

The values of the response variable form a distribution:

6) Analyze the response variable.

We wanted to know how many boxes we might expect to buy to get all three cards. To answer this question, we might report a measure of center for this data, either a median or a mean. (Of course, the simulation's results will be more accurate if we include many more trials.)

7) State your conclusion in the context of the problem.

With this small number of trials, we might say we expect to have to buy 4 boxes (median value) or we might say 8 boxes (7.4 = mean value). This number would be more convincing if we ran hundreds of trials.