DQ 1/3/19
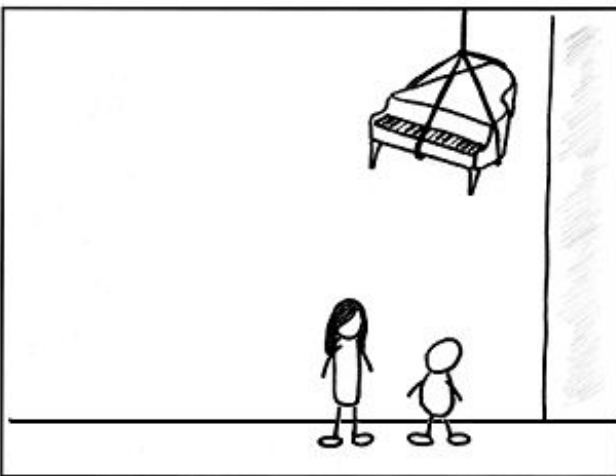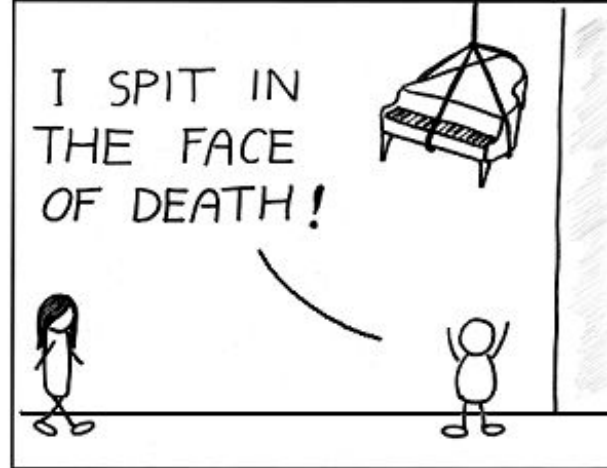
What is statistics? Write a definition IN YOUR OWN words.

DQ 1/4

1. You want to gather information from 100 students at your school about adding a salad bar to the cafeteria. Give three different ways you could collect your data.

NC Math 3
Unit 8: Statistics
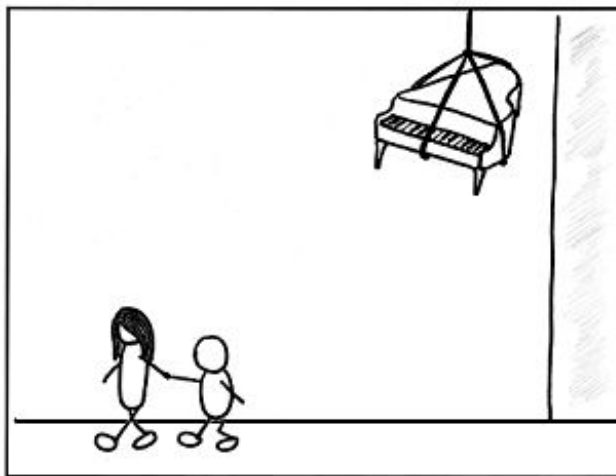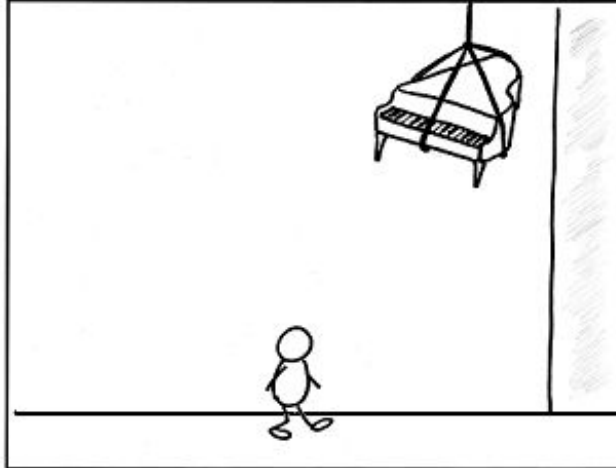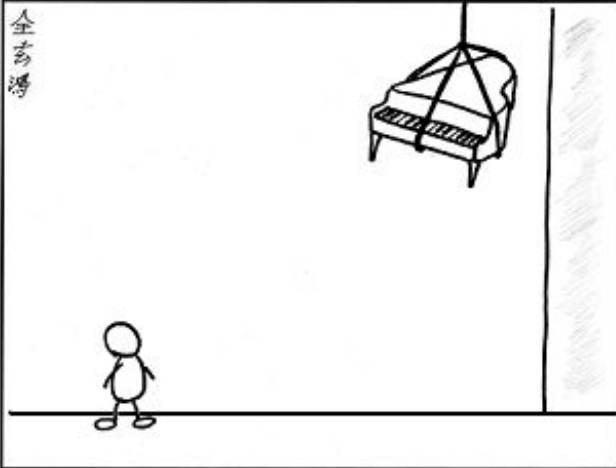
We will discuss the following five topics during this unit:

1. Sampling and Study Design
2. Bias
3. Estimating Population Parameters
4. Margin of Error
5. Simulations

"Statistics say that religious people live longer, so I practice a different religion every day of the week to be sure I'm covered."

Sampling and Study Design

Main Questions

What's the difference between an experiment and an observational study?

What are the different ways that a sample can be collected?

When is a sample considered random?

What is bias and how does it affect the data you collect?

There are three ways to collect data:

I DON'T KNOW HOW TO DO STATISTICS BUT IT DOESN'T MATTER BECAUSE I DIDN'T HAVE DATA.
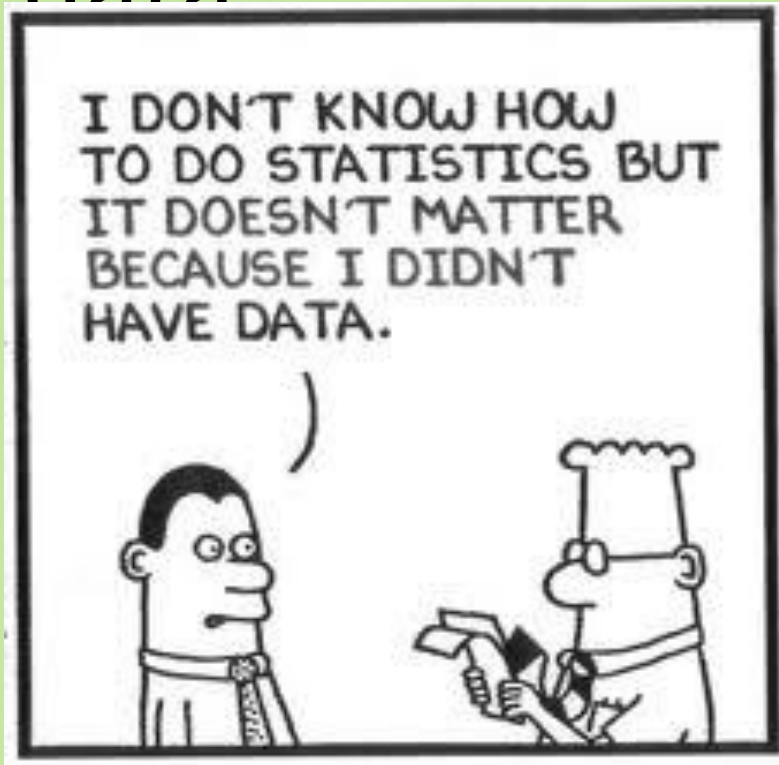
1. Surveys

2. Observational Studies

3. Experiments

Surveys

Surveys most often involve the use of a questionnaire to measure the characteristics and/or attitudes of people.

ex. asking students their opinion about extending the school day
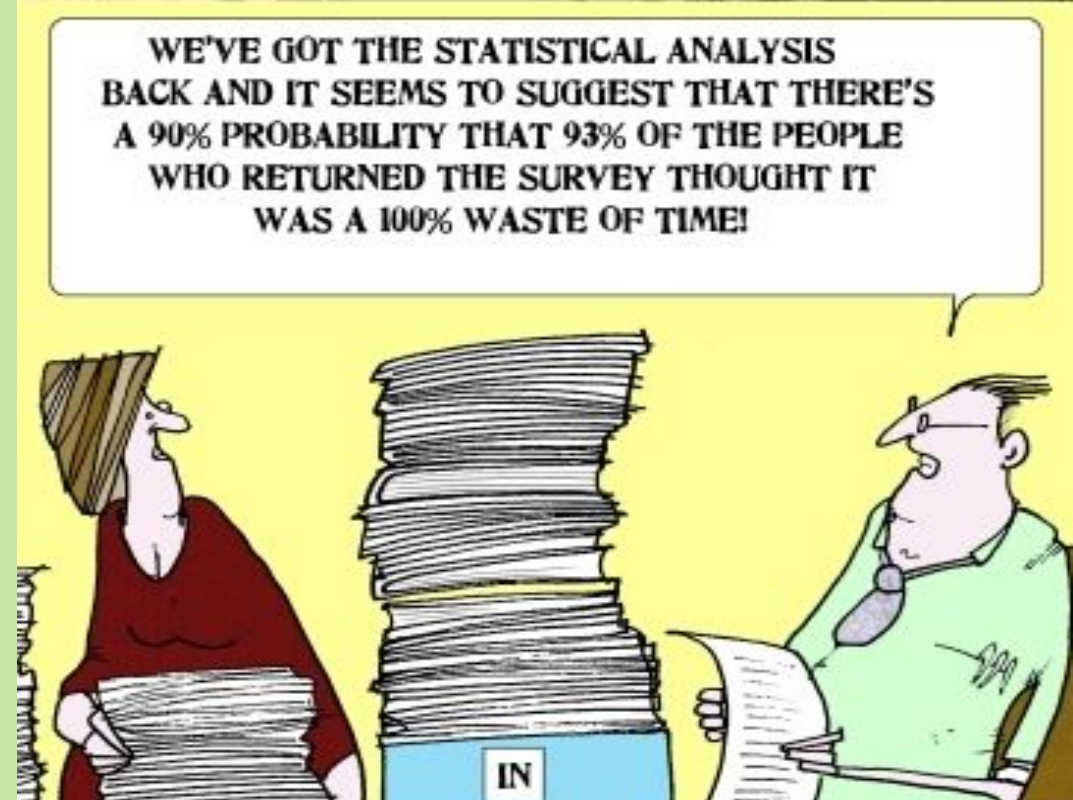
Observational Studies
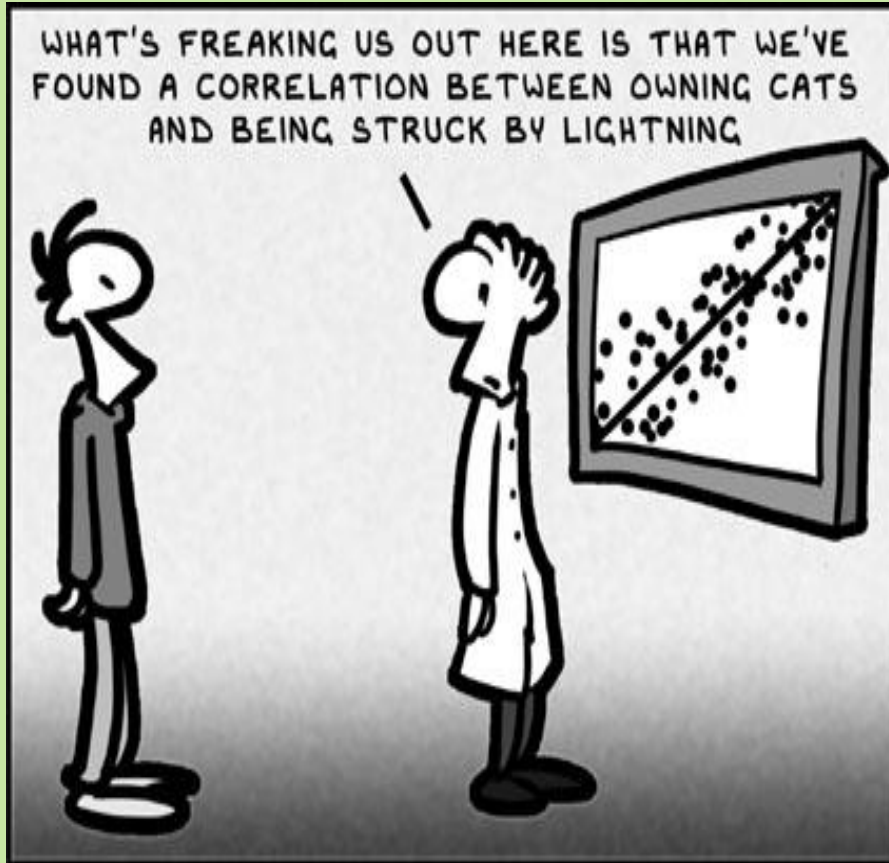
Individuals are observed and certain outcomes are measured, but <u>no attempt is made to affect the outcome.</u>

Experiments

<u>Treatments are imposed</u> prior to observation. Experiments are the only way to show a cause-and-effect relationship.

© Mike Baldwin / Cornered

"Do a double-blind test. Give the new drug to rich patients and a placebo to the poor. No sense getting their hopes up. They couldn't afford it even if it works."

Remember:

Correlation is not causation!

Observational Study or Experiment?

Fifty people with clinical depression were divided into two groups. Over a 6 month period, one group was given a traditional treatment for depression while the other group was given a new drug. The people were evaluated at the end of the period to determine whether their depression had improved.

Experiment

Observational Study or Experiment?

To determine whether or not apples really do keep the doctor away, forty patients at a doctor's office were asked to report how often they came to the doctor and the number of apples they had eaten recently.

Observational Study

Observational Study or Experiment?

To determine whether music really helped students' scores on a test, a teacher who taught two U. S. History classes played classical music during testing for one class and played no music during testing for the other class.

Experiment

Types of Sampling

In order to collect data, we must choose a **sample**, or a group that represents the population.

The goal of a study will determine the type of sampling that will take place.

Simple Random Sample (SRS)

All individuals in the population have the same probability of being selected, and all groups in the sample size have the same probability of being selected.

Putting 100 kids' names in a hat and picking out 10 - SRS

Putting 50 girls' names in one hat and 50 boys' names in another hat and picking out 5 of each – not a SRS

Stratified Random Sample

If a researcher wants to highlight specific subgroups within the population, they divide the entire population into different subgroups, or strata, and then randomly selects the final subjects proportionally from the different strata.

Systematic Random Sample

The researcher selects a number at random, n, and then selects every nth individual for the study.

Convenience Sample

Subjects are taken from a group that is conveniently accessible to a researcher, for example, picking the first 100 people to enter the movies on Friday night.

Cluster Sample

The entire population is divided into groups, or clusters, and a random sample of these clusters are selected. All individuals in the selected clusters are included in the sample.
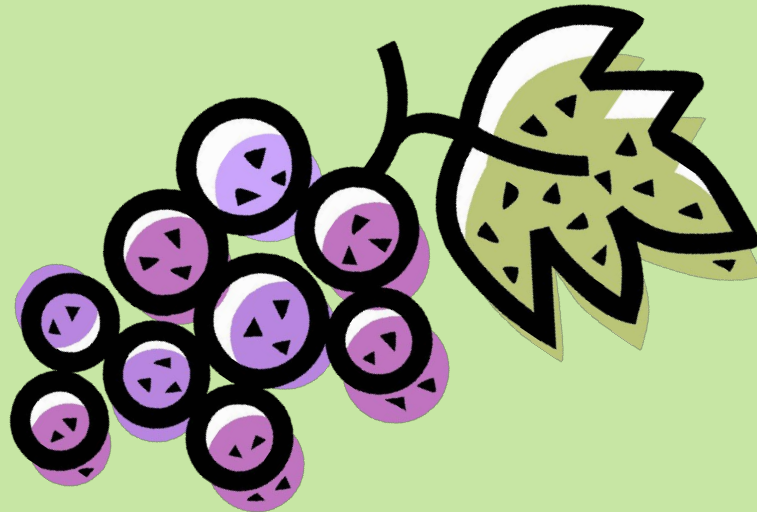
Name that sample!

The names of 70 contestants are written on 70 cards, the cards are placed in a bag, the bag is shaken, and three names are picked from the bag.

Simple random sample   Stratified sample

Convenience sample   Cluster sample

Systematic sample

Name that sample!

To avoid working late, the quality control manager inspects the last 10 items produced that day.

Simple random sample   Stratified sample

Convenience sample   Cluster sample

Systematic sample

Name that sample!

A researcher for an airline interviews all of the passengers on five randomly selected flights.

Simple random sample   Stratified sample

Convenience sample   Cluster sample

Systematic sample

Name that sample!

A researcher randomly selects and interviews fifty male and fifty female teachers.

Simple random sample   Stratified sample

Convenience sample   Cluster sample

Systematic sample

Name that sample!

Every fifth person boarding a plane is searched thoroughly.

Simple random sample   Stratified sample

Convenience sample   Cluster sample

Systematic sample

Types of Bias in Survey Questions

Bias occurs when a sample systematically favors one outcome.

1. **Question Wording Bias**

In a survey about Americans' interest in soccer, the first 25 people admitted to a high school soccer game were asked, "How interested are you in the world's most popular sport, soccer?"

2. **Undercoverage bias** occurs when the sample is not representative of the population.

3. **Response bias** occurs when survey respondents lie or misrepresent themselves.

4. **Nonresponse bias** occurs when an individual is chosen to participate, but refuses.

5. **Voluntary response bias** occurs when people are asked to call or mail in their opinion.

Name that bias!

On the twelfth anniversary of the death of Elvis Presley, a Dallas record company sponsored a national call-in survey. Listeners of over 1000 radio stations were asked to call a 1-900 number (at a charge of $2.50) to voice an opinion concerning whether or not Elvis was really dead. It turned out that 56% of the callers felt Elvis was alive.

Voluntary response bias

Name that bias!

In 1936, *Literary Digest* magazine conducted the most extensive public opinion poll in history to date. They mailed out questionnaires to over 10 million people whose and addresses they had obtained from telephone books and vehicle registration lists. More than 2.4 million people responded, with 57% indicating that they would vote for Republican Alf Landon in the upcoming Presidential election. However, Democrat Franklin Roosevelt won the election, carrying 63% of the popular vote.

Undercoverage bias

Why is this question biased?

Do you think the city should risk an increase in pollution by allowing expansion of the Northern Industrial Park?

Can you rephrase it to remove the bias?

Why is this question biased?

If you found a wallet with $100 in it on the street, would you do the honest thing and return it to the person or would you keep it?

Can you rephrase it to remove the bias?
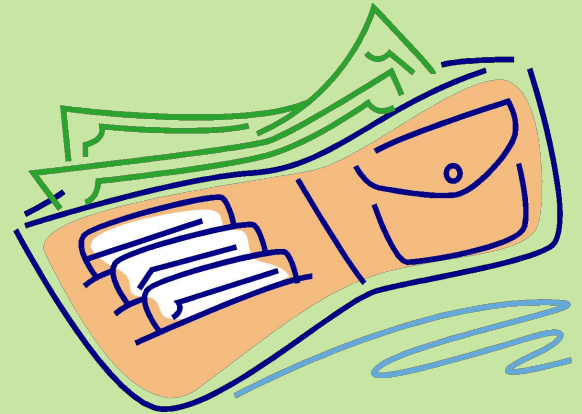
Questions about sampling?
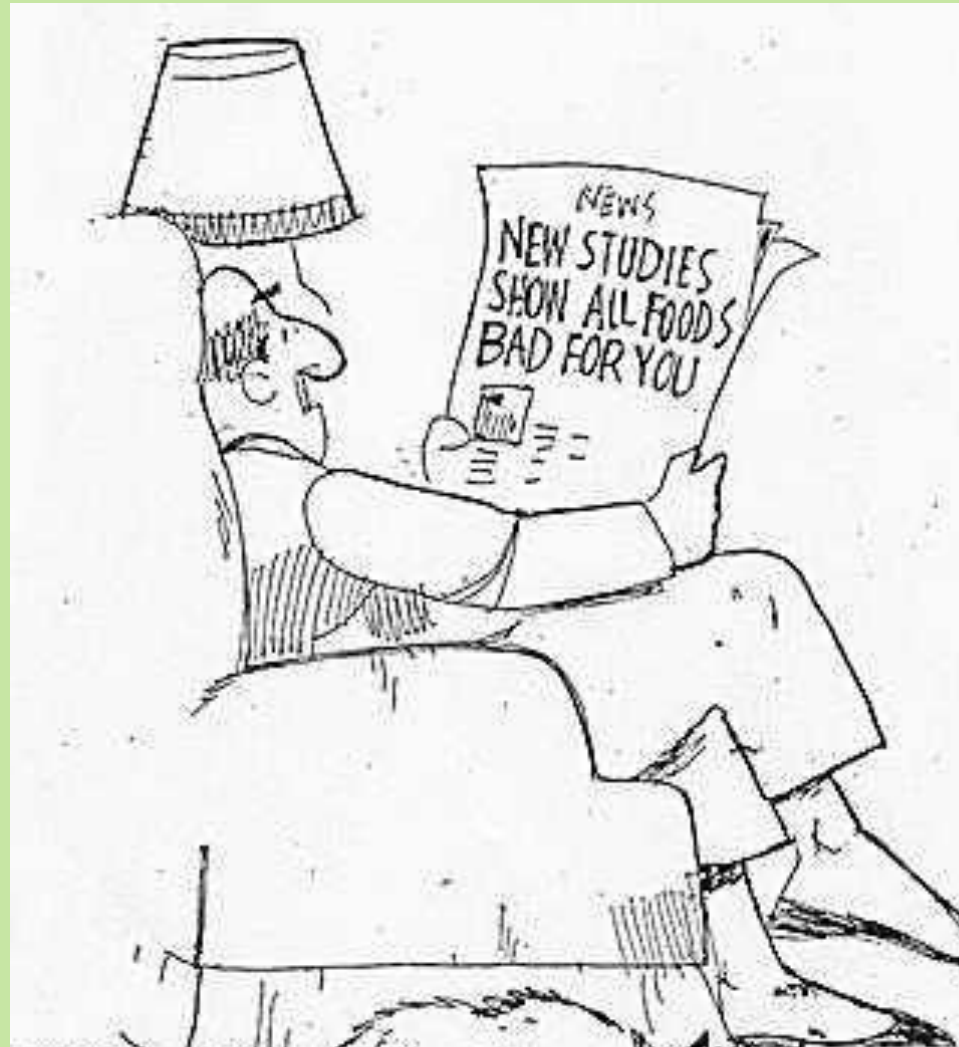
Estimating Population Parameters

Vocabulary for this lesson is important!

Parameter   a value that represents a population

Statistic   a value based on a sample and used to estimate a parameter

	population ⟶ parameter	sample ⟶ statistic
mean	μ	$\overline{x}$
proportion	p	$\hat{p}$
standard deviation	σ	s

Finding a Margin of Error

"I'm going to need a Margin of Error or I can't publish your prediction of six more weeks of winter."

Margin of error is a "cushion" around a statistic

$$ME = \frac{1}{\sqrt{n}}$$

n = sample size

Suppose that 900 American teens were surveyed about their favorite event of the Winter Olympics. Ski jumping was the favorite of 20% of those surveyed. This result can be used to predict the proportion of **ALL** American teens who favor ski jumping.

$$\hat{p} = 0.2 \qquad ME = \dfrac{1}{\sqrt{900}}, \ or \ 0.03$$

$$0.2 \pm 0.03 = 0.17 - 0.23$$

We can confidently state that the true proportion of American teens who favor ski jumping falls between 17% and 23%.

How does sample size affect margin of error?

If your sample size is 400 and you wish to cut the margin of error in half, what will your new sample size be?

1600

What sample size produces a given margin of error?

If you want your margin of error to be 5%, what size sample will you need?

400

Simulation

Simulation is a way to model random events, so that simulated outcomes closely match real-world outcomes.


"We've done a computer simulation of your projected performance in five years. You're fired."

Why run a simulation?

Some situations may be difficult, time-consuming, or expensive to analyze. In these situations, simulation may approximate real-world results while requiring less time, effort, or money.

Carole and John are playing a dice game. Carole believes that she can roll six dice and get each number, one through six, on a single roll. John knows the probability of this occurrence is low. He bets Carole that he will wash her car if she can get the outcome she wants in twenty tries.
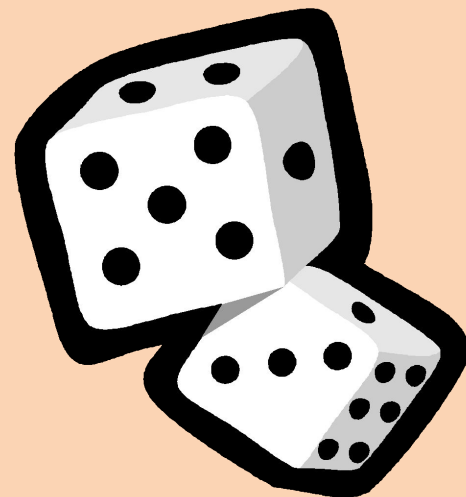
What is the problem that we are simulating?

Can Carole get one of each number in a roll of six dice?

What random device will you use to simulate the problem and how will you use it?

We will use the calculator to generate random numbers.

BEFORE YOU START!

You are running 20 trials, so make 20 blanks on your paper. This will keep you from losing count of how many trials you've run. It also makes recording easy!

✓ ✓ ✗ ✓ ✗ __ __ __ __ __ __ __

__ __ __ __ __ __ __ __

Seeding

Since a calculator is a type of computer, it can never be truly random.

For this reason, we can configure our calculators to give everyone the same set of "random" data (so we can all work together!).

The process of calibrating our calculators in this way is called seeding.

How to seed the calculator:

```
5→rand
            5
■
```

```
5→rand
            5
rand
    .7278038625
■
```

If you wish, you can control the starting "seed" value.
To seed the random generator, choose a seed value and store it into the rand command.

Now, start generating your random values.
If you feed two calculators the same seed value, they will each produce the same result when rand is entered.

How will you conduct each trial? How many trials will you conduct?

I will use the **randInt(** command in my calculator to generate random integers.

randInt(min value, max value, number of data in set)

randInt(1, 6, 6)

What are the results of these trials?

We received all 6 numbers only 1 out of 20 times.

What predictions can be made based on these results?

There's approximately a 5% chance of this occurring.

The more trials you run, the closer you will get to the theoretical probability (Law of Large Numbers).

On a certain day the blood bank needs 4 donors with type O blood. If the hospital brings in groups of five, what is the probability that a group would arrive that satisfies the hospitals requirements, assuming that 45% of the population has type O blood?

Let 1-45 represent people with type O blood.

Let 46-100 represent people with other blood types.

Remember to seed the calculator to 5!
Then, run RandInt(1, 100, 5) twenty times.
Record how many trials satisfy the hospital's requirements.

To seed your calculator:

MATH, PRB, 5, STO →, RAND, ENTER

To run the simulation:

MATH, PRB, RandInt(1, 100, 5), ENTER

Five of the twenty groups have four or more members with type O blood. Therefore, there is a 25% chance that they hospital will get the Type O blood they need.

"Statistics say that religious people live longer, so I practice a different religion every day of the week to be sure I'm covered."

Questions about simulations?

Any questions about statistics?

"Mr. Osborne, may I be excused? My brain is full."