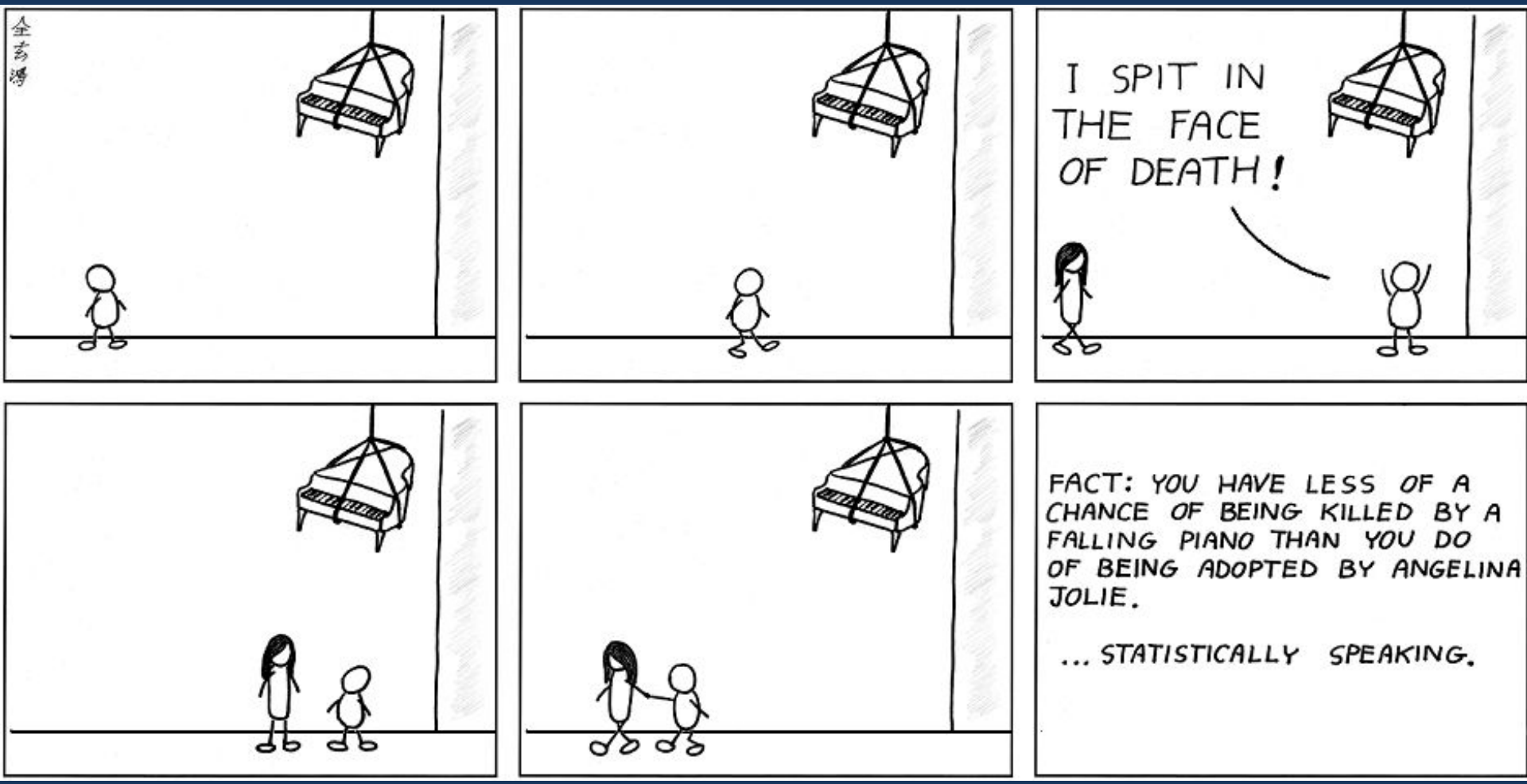
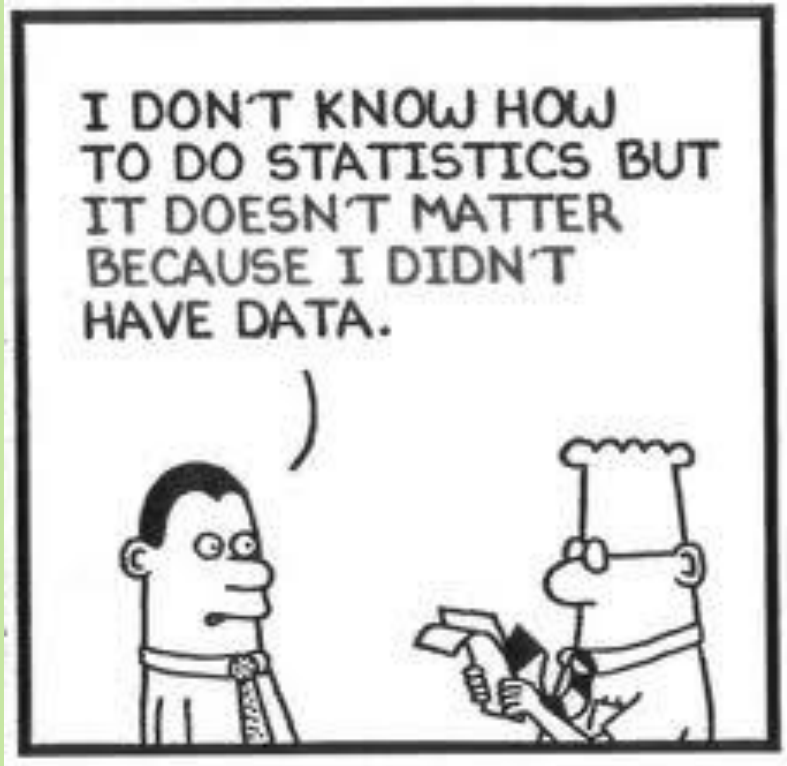
# Statistics!!!!

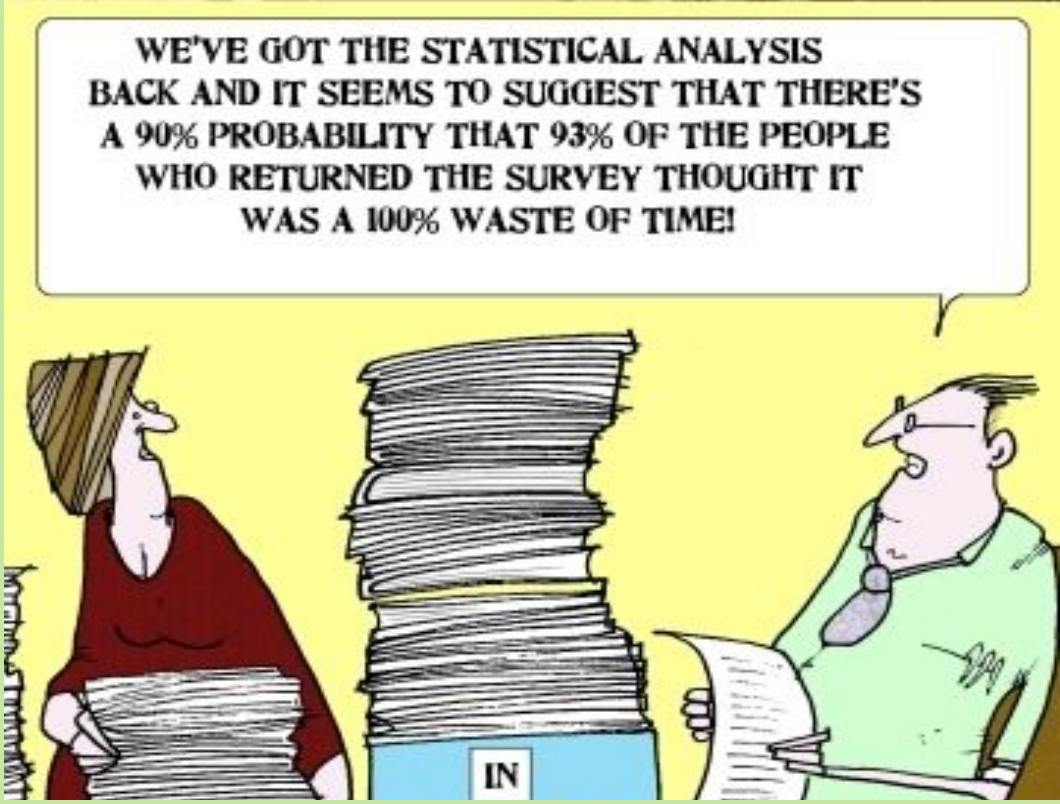

I SPIT IN THE FACE OF DEATH!
FACT: YOU HAVE LESS OF A CHANCE OF BEING KILLED BY A FALLING PIANO THAN YOU DO OF BEING ADOPTED BY ANGELINA JOLIE.
... STATISTICALLY SPEAKING.

# There are three ways to collect data:




1. Surveys
2. Observational Studies
3. Experiments

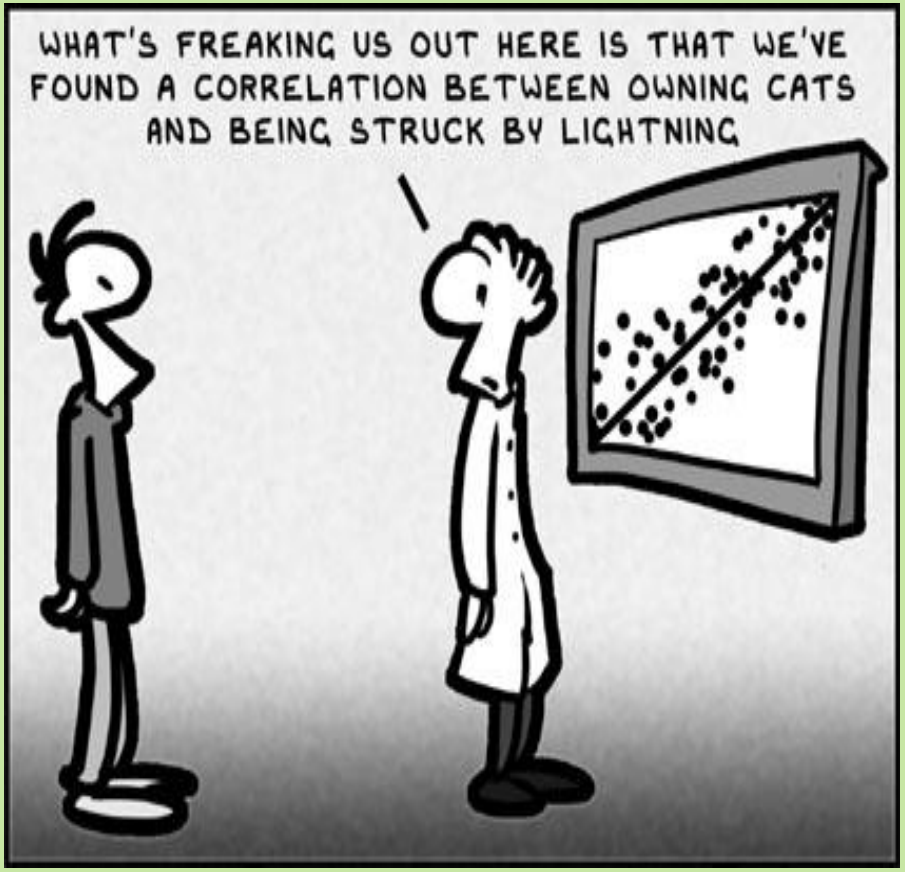
# Surveys




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.

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

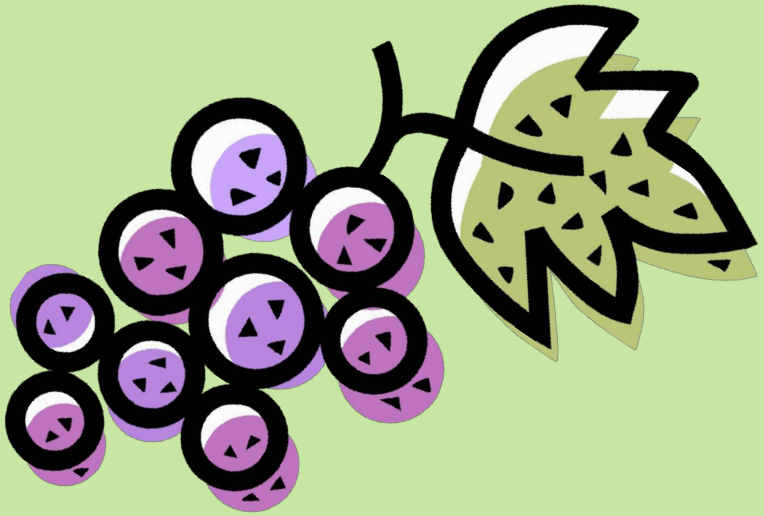
The researcher selects a number at random, n, and then selects every nth individual for the study.

# 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.



# Voluntary Response Sample

When people are self-selected volunteers into 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

Voluntary response 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

Voluntary response 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

Voluntary response 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

Voluntary response sample

# Name that sample!

Every fifth person boarding a plane is searched thoroughly.

Simple random sample

Stratified sample

Convenience sample

Cluster sample

Systematic sample

Voluntary response sample

# Population and Parameter VS Sample and Statistic

***Population*** all the members of the group you want information about

***Parameter*** a value that represents a population

# Population and Parameter VS Sample and Statistic

***Sample*** the selected members of the group that you get information from (it is not WHOLE group)

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

population

sample

| | parameter | statistic |
|---|---|---|
| mean | μ | $\bar{x}$ |
| proportion | $p$ | $\hat{p}$ |
| standard deviation | σ | s |

# Name that: Population? Sample? Parameter? Statistic?

The average salary of 500 employees at MyersCorp is $63,000.

# Name that: Population? Sample? Parameter? Statistic?

A survey of 400 students at Rolesville High School found that 63% of them prefer to eat pizza for lunch.

# 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

$$ME = 2\frac{s}{\sqrt{n}}$$

$s$ = sample standard deviation

$n$ = 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 percentage of **ALL** American teens who favor ski jumping, using a standard deviation of 0.13

# How does sample size affect margin of error?

If your sample size is 400 and your standard deviation is 2.4, what is your margin of error?

If you want to cut your margin of error in half, what will your new sample size be?

# 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 for a standard deviation of .77?

# 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.

# 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.

# 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?

# Why is this question biased?

Last year teachers went on a one day strike in order to protest, forcing numerous school systems to shut down and forcing many kids to go without their lone source of nutrition for the day. They plan to do this once again on May 1. Do you support this strike?

Can you rephrase it to remove the bias?