



Experiment & Data Analysis Part A

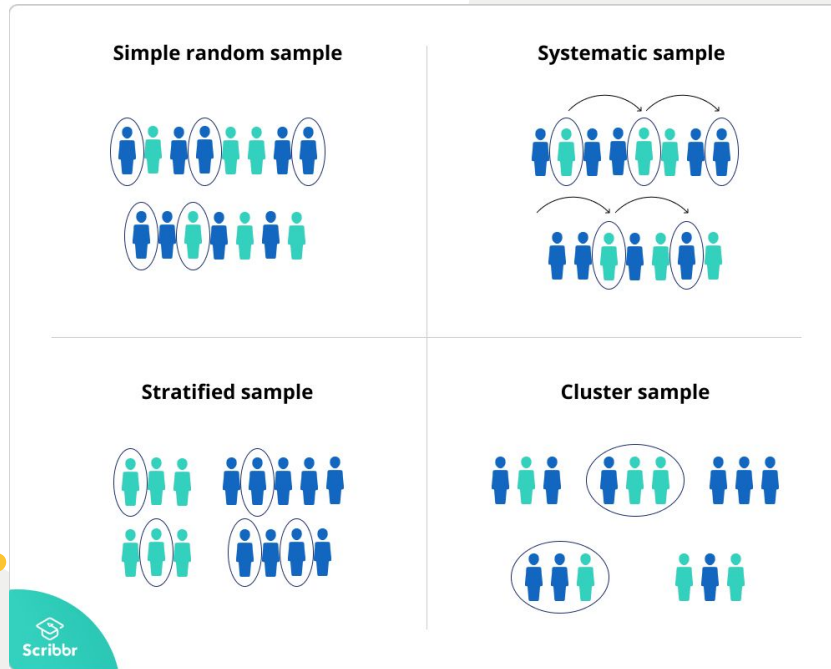
Sampling

Using a subset of a population to make observations/predictions about the larger population



Probability Sampling

Every member of the population has an equal chance of being selected



Example 1

All employees of the company are listed in alphabetical order. From the first 10 numbers, you randomly select a starting point: number 6. From number 6 onwards, every 10th person on the list is selected (6, 16, 26, 36, and so on), and you end up with a sample of 100 people.

Example 2

You want to select a simple random sample of 100 employees of Company X. You assign a number to every employee in the company database from 1 to 1000, and use a random number generator to select 100 numbers.

Example 3

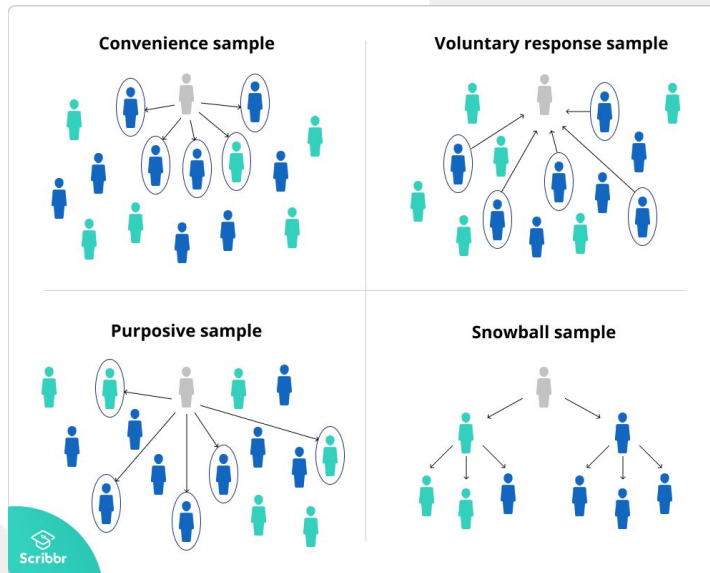
The company has 800 female employees and 200 male employees. You want to ensure that the sample reflects the gender balance of the company, so you sort the population into two strata based on gender. Then you use random sampling on each group, selecting 80 women and 20 men, which gives you a representative sample of 100 people.

Example 4

The company has offices in 10 cities across the country (all with roughly the same number of employees in similar roles). You don't have the capacity to travel to every office to collect your data, so you use random sampling to select 3 offices.

Non-Probability Sampling

Individuals selected using non-random criteria (not an equal chance of selection)



Example 5

You are researching opinions about student support services in your university, so after each of your classes, you ask your fellow students to complete a survey on the topic.

What type of sampling? What issues does it present?

Example 6

You want to know more about the opinions and experiences of disabled students at your university, so you purposefully select a number of students with different support needs in order to gather a varied range of data on their experiences with student services.

What type of sampling? What issues does it present?

Example 7

You send out the survey to all students at your university and a lot of students decide to complete it.

What type of sampling? What issues does it present?

Example 8

You are researching experiences of homelessness in your city. Since there is no list of all homeless people in the city, probability sampling isn't possible. You meet one person who agrees to participate in the research, and she puts you in contact with other homeless people that she knows in the area.

What type of sampling? What issues does it present?

Correlation vs Causation

Correlation: relationship or pattern between variables

Causation: one event directly **CAUSES** another



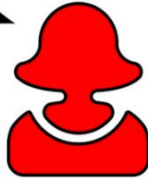
DRY, HOT AND SUNNY
SUMMER WEATHER



ICE CREAM

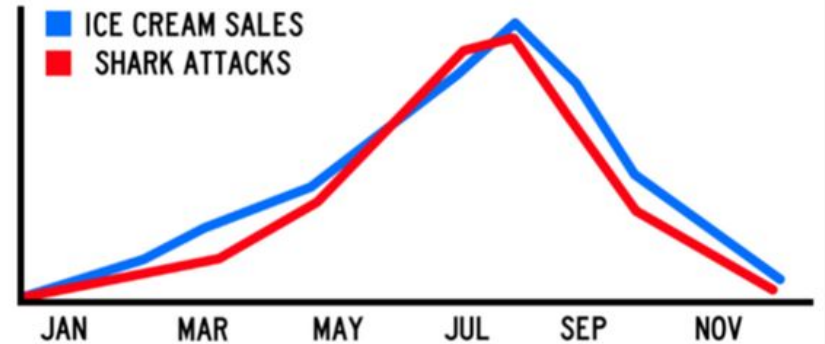


correlation

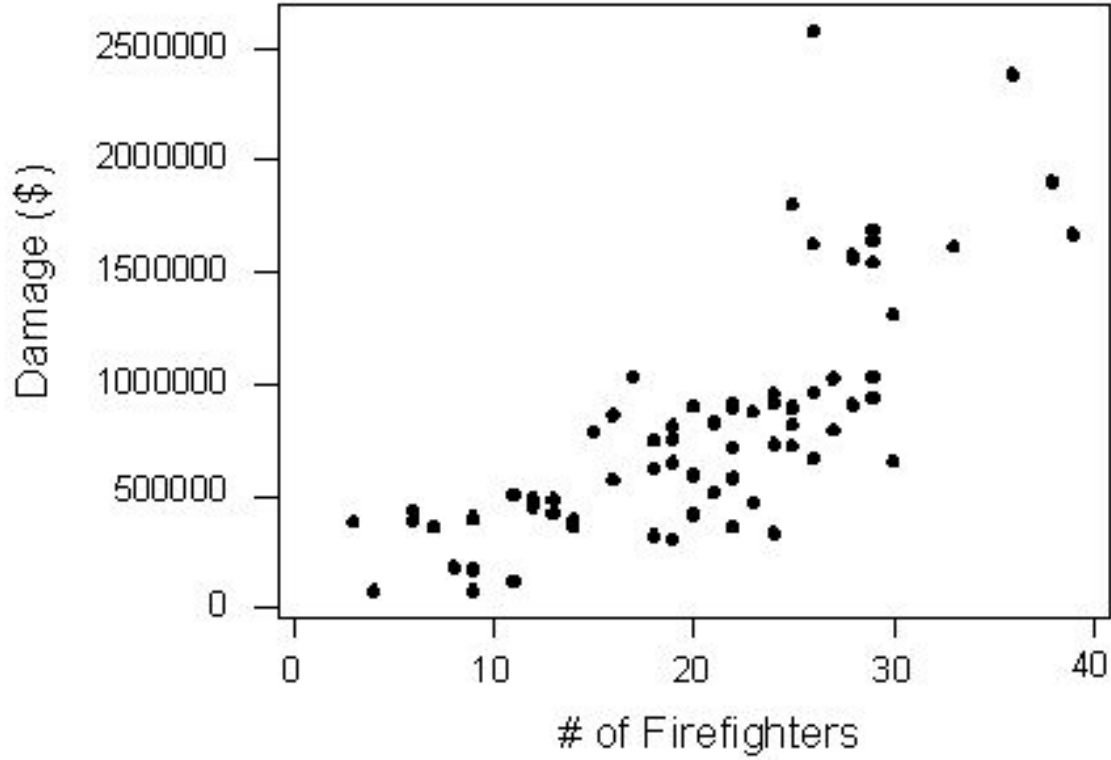


SUNBURN

CORRELATION IS NOT CAUSATION!



Both ice cream sales and shark attacks increase when the weather is hot and sunny, but they are not caused by each other (they are caused by good weather, with lots of people at the beach, both eating ice cream and having a swim in the sea)

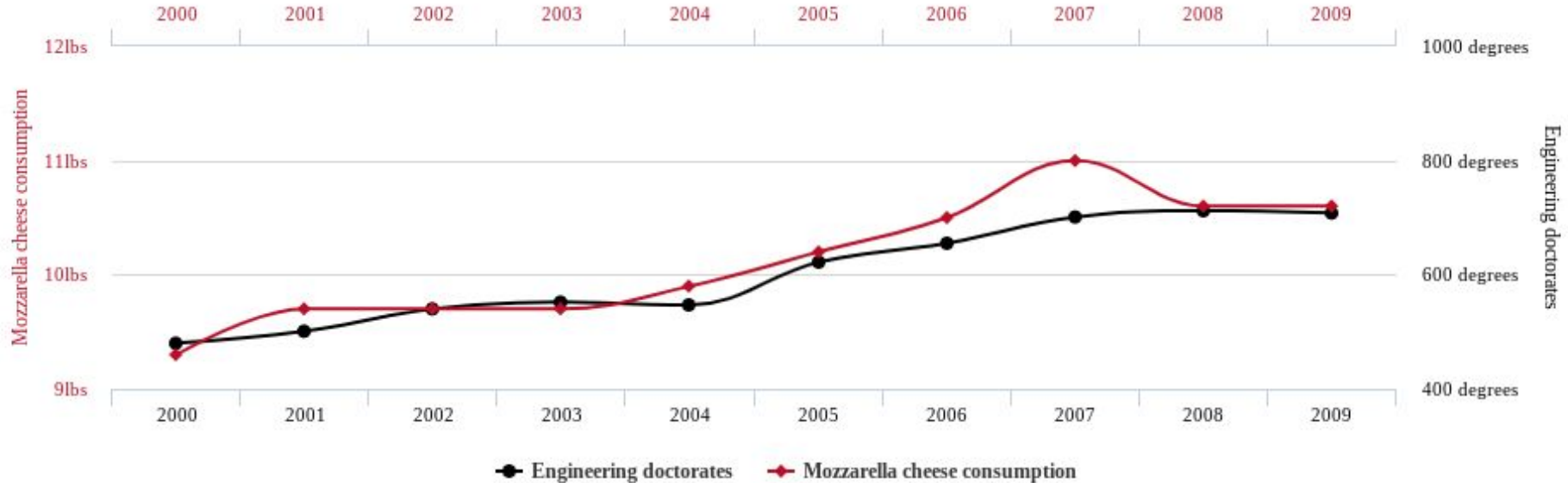


Per capita consumption of mozzarella cheese

correlates with

Civil engineering doctorates awarded

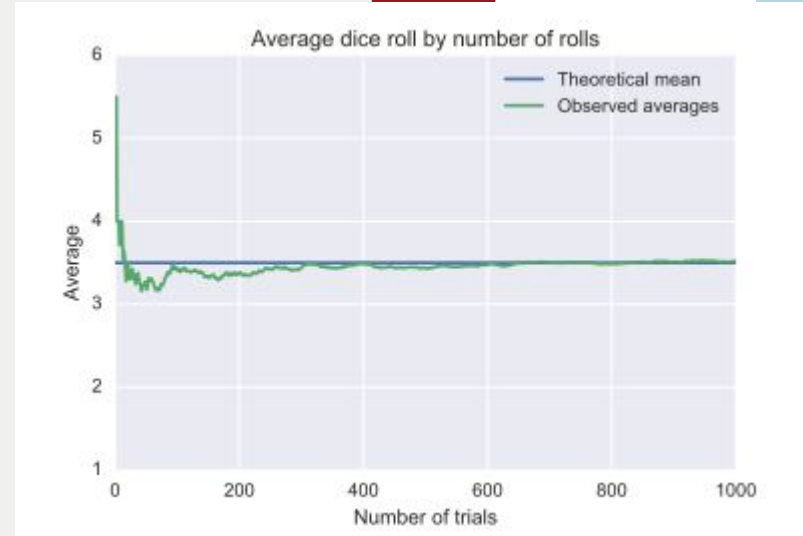
Correlation: 95.86% (r=0.958648)



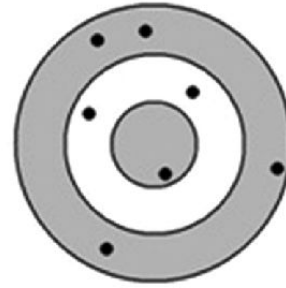
Law of Large Numbers

Sample mean converges to the theoretical mean as sample size increases

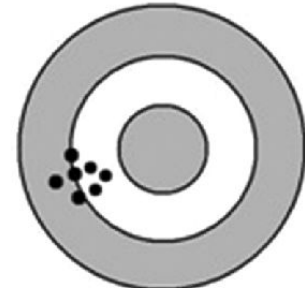
As number of trials or observations increases, observed probability approaches expected probability



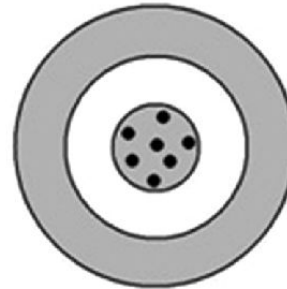
Precision vs Accuracy



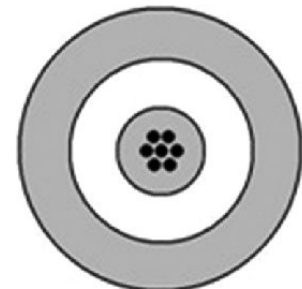
**Neither accurate
nor precise**



**Precise but
not accurate**



**Accurate but
not precise**



**Both accurate
and precise**