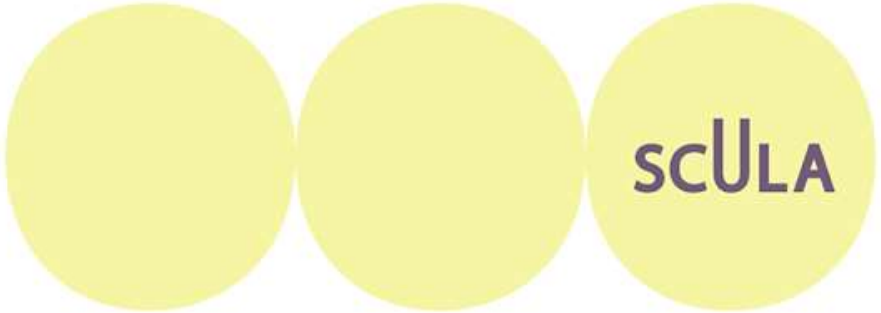
SAT MATH SECTION

Statistics II

What is the goal of statistic?

We are making predictions and estimations but based on limited time and information.

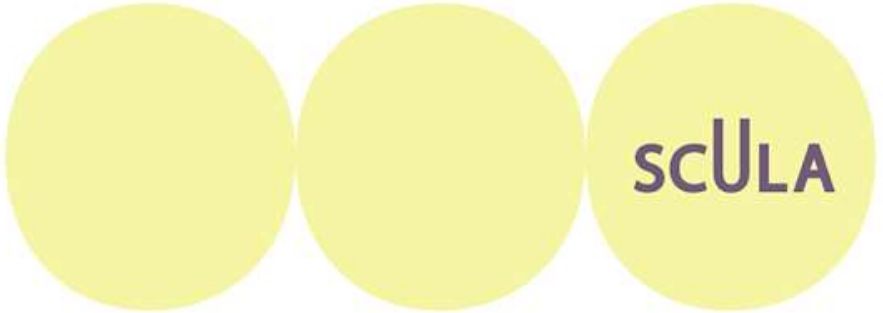
- Statistical Sampling
- Interpreting the line of the best fit



Data Sampling

Because we cannot test all the data available (usually data sets are huge), we can only consider a representative sample .

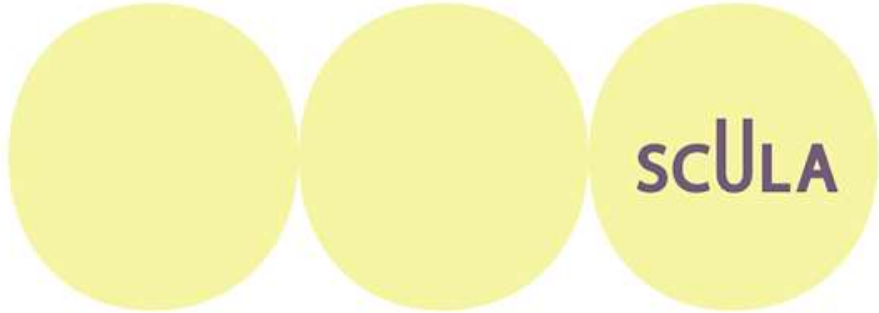
For example, if we want to survey the citizens of a country about a certain phenomenon, we should only select a sample from all the citizens of the country .



EXAMPLE 1: A pet food store chose 1,000 customers at random and asked each customer how many pets he or she has. The results are shown in the table below.

Number of pets	Number of customers
1	600
2	200
3	100
4 or more	100

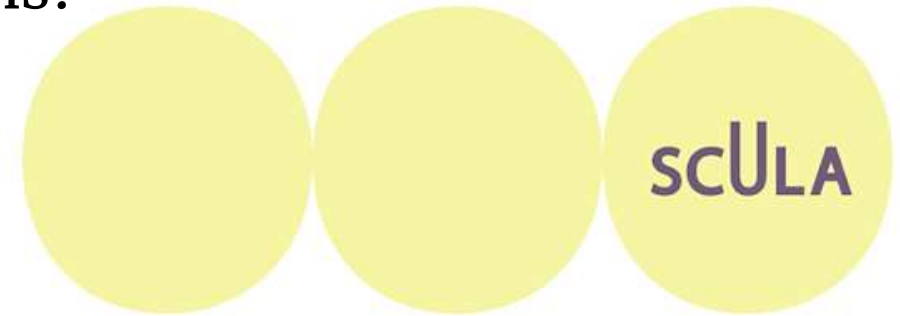
There are a total of 18,000 customers in the store's database. Based on the survey data, what is the expected total number of customers who own 2 pets?



Each 200 customers have 2 pets. We can can estimate the number who own 2 pets to be :

$$1800 \times \frac{200}{1000} = 3600$$

We can also deduce from this example that we can only generalize the results of a sampling on the samples that adhere to the same criteria. So, how do we evaluate statistical claims?



Evaluating Statistical Claims :

We routinely conduct research to answer questions such as "how many residents are in favor of a new law" or "is a new medical treatment effective?" While research results can give us powerful insights, we must carefully consider how the research is conducted, which in turn affects what conclusions can be drawn.

For example:

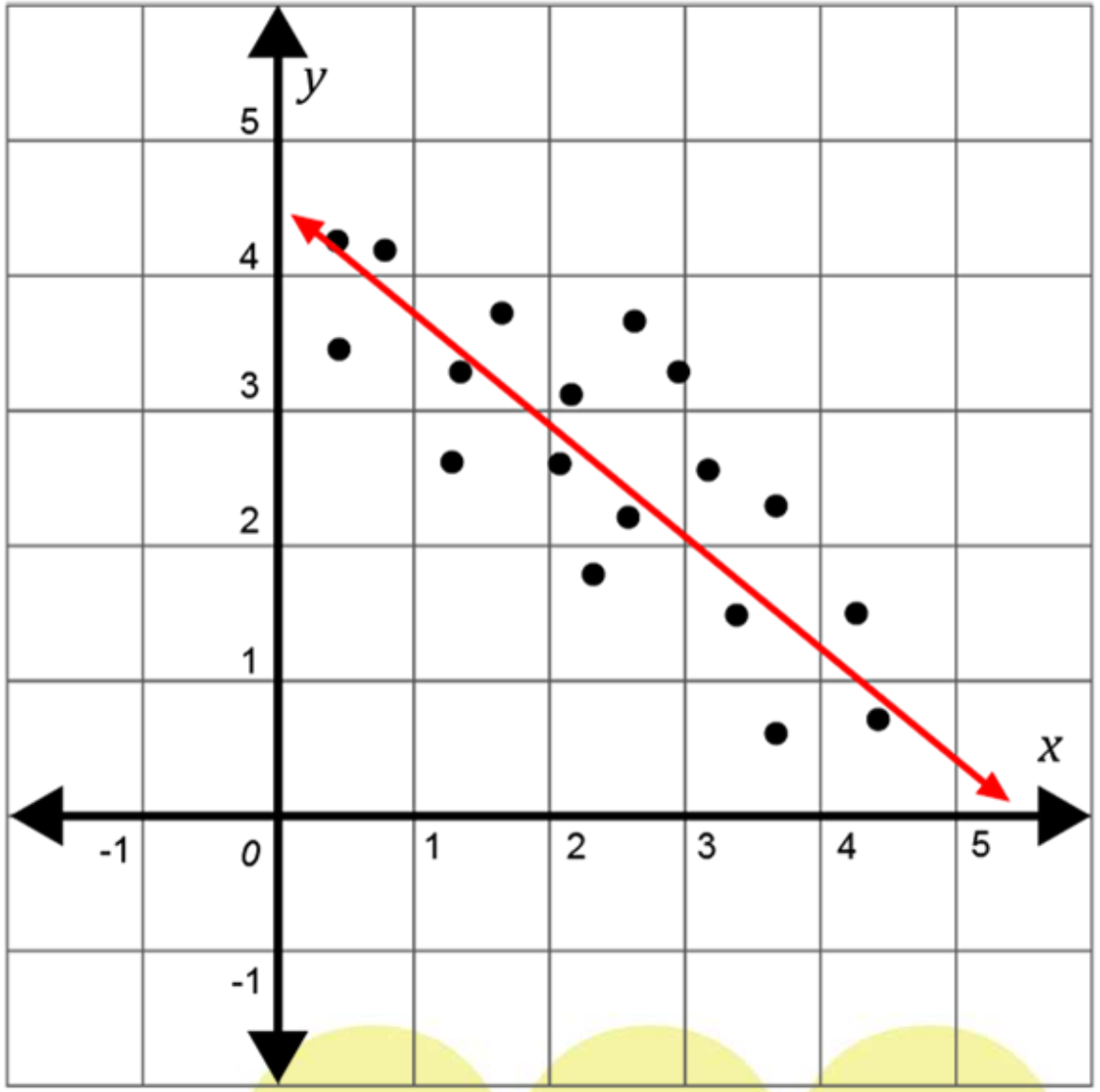
If a survey was given to individuals of one ethnicity, then the results of the survey are not representative of individuals of other ethnicities.

If a medical treatment is effective when tested on mice, **we cannot conclude that the treatment is just as effective on humans without additional testing.**

Line of the best fit

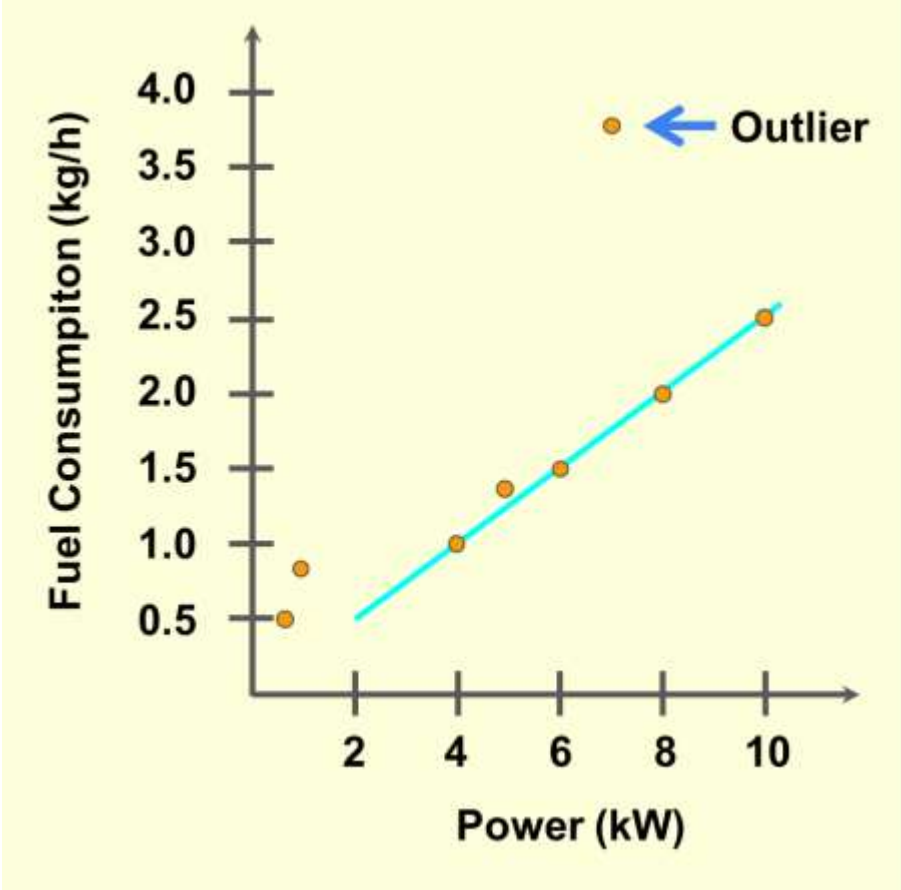
When we collect data that we represent graphically, most of the time, the allure of the data is not regular .

To facilitate dealing with collected data, we approximate it to a linear relationship that we represent through a line. We call this line the line of the best fit .



Outliers :

The line of best fit is determined by the correlation between the two variables on a scatter plot. In the case that there are a few outliers **era taht stniop atad) tser eht morf yawa raf detacol** tsujda lliw enil eht **(atad eht fo** esoht stneserper ti taht os .llew sa stniop



Example :

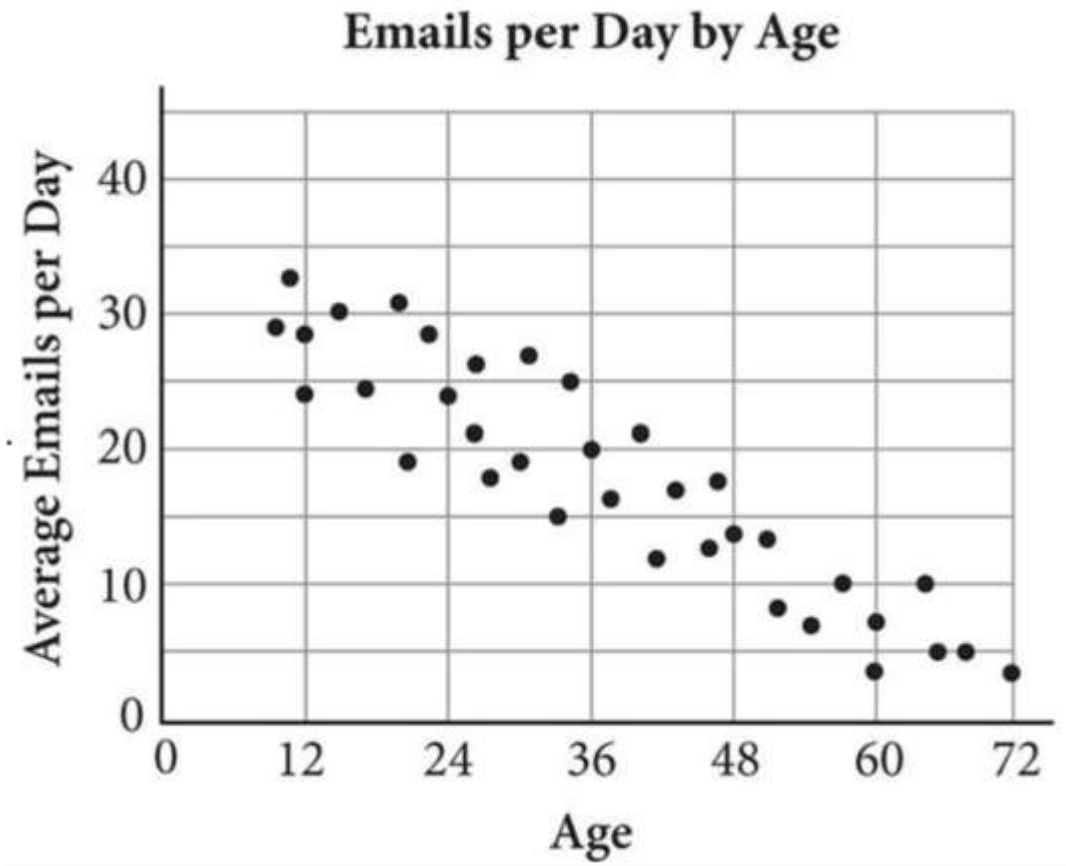
Which of the following equations best represents the trend of the data shown in the figure above?

A. $y = -2.4x + 30$

B. $y = -1.2x + 40$

C. $y = -0.8x + 40$

D. $y = -0.4x + 36$



SCULA

PRACTICE

https://drive.google.com/file/d/1zo2HQ-Ro_tYGm2BqQT-32FSPtuNE2lAr/view?usp=drive_link

THANK YOU!

DO YOU HAVE ANY QUESTIONS?