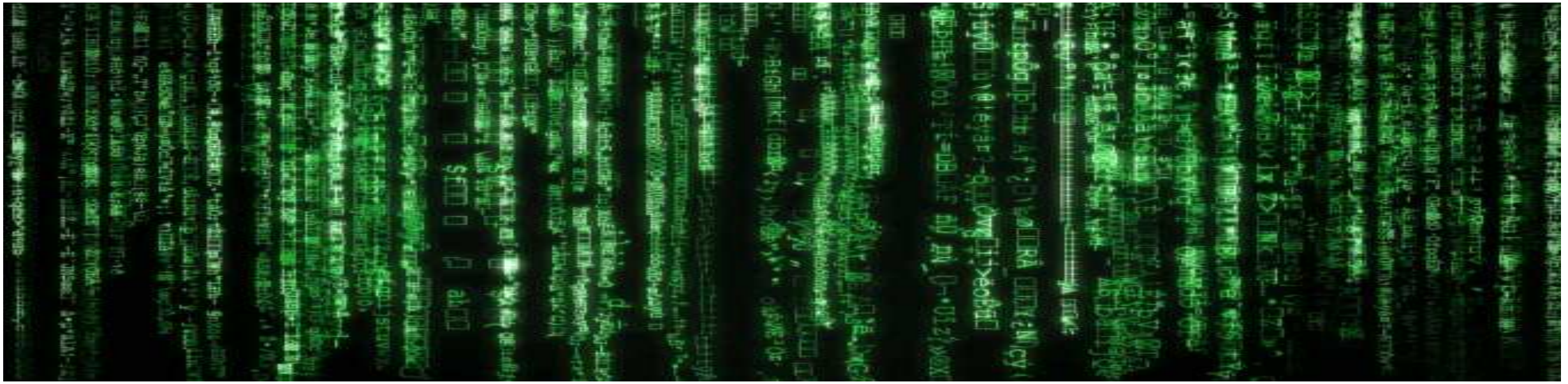


INF 110 Discovering Informatics

Estimation

NAU
NORTHERN
ARIZONA
UNIVERSITY

The Concept of Estimation

- Estimating the characteristics of a population based on available data.

- Often, estimating an unknown parameter.
  - What percent of voters favor candidate A?
  - What is the median annual household income in the US?

- Relies on random samples from a much larger population
  - You can't measure an unknown parameter from ***everyone***.

The Concept of Estimation

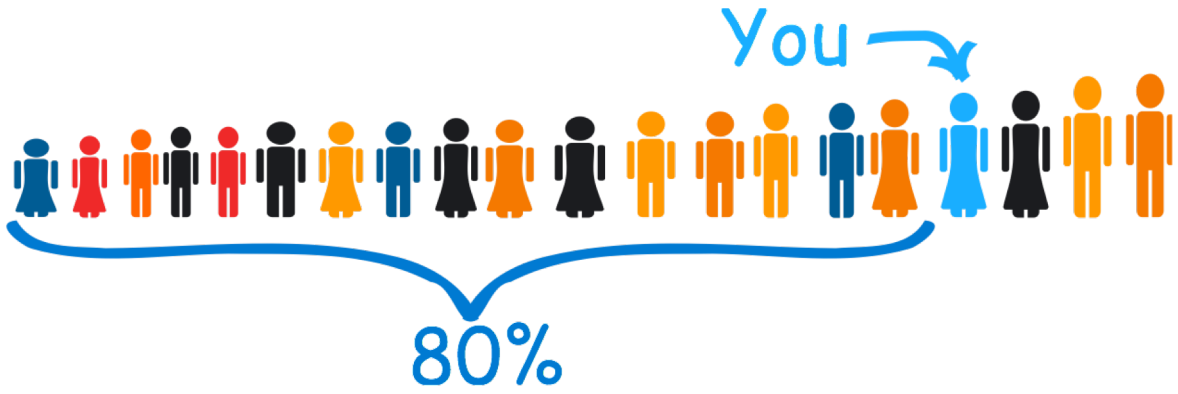
- This leads to a question of inference:
  - ***How can we make justifiable conclusions about an unknown parameter based on a random sample?***
  - We can think of ways to do this based on what we've learned in this course.

- But, then there's another question:
  - ***If the sample was different, would the estimate be different?***

Percentile

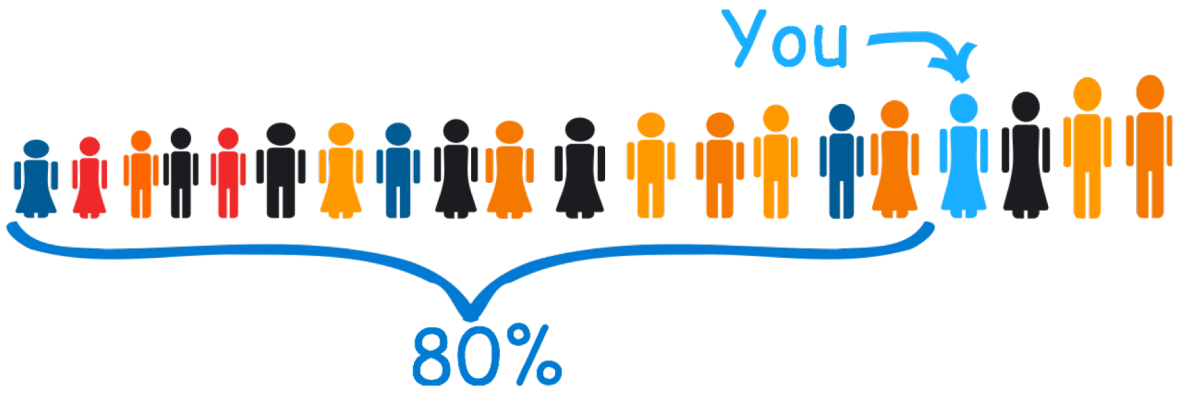
The value below which a percentage of data falls.

Being the 80th percentile of height means 80% of people are shorter than you.



How do you compute percentile?

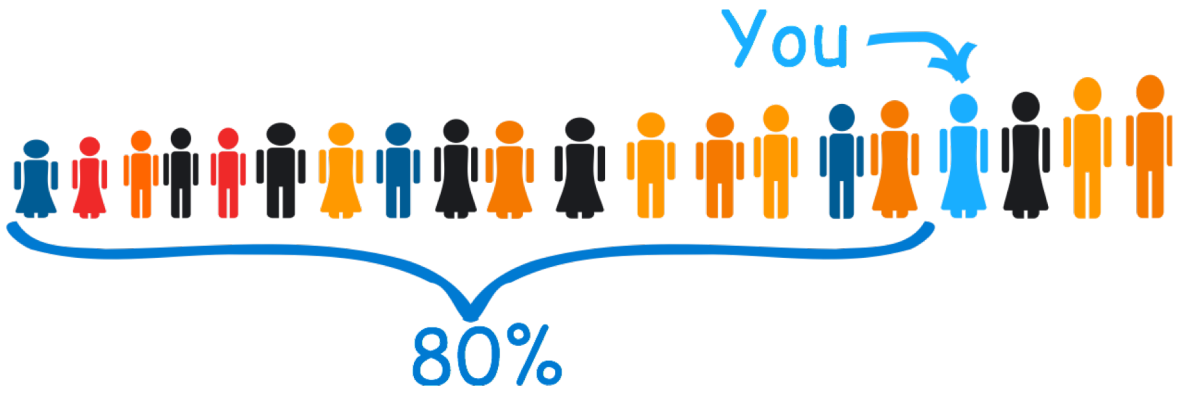
1. Sort the data (short to tall for height)
2. Count the number of elements below the element of interest (you!)
3. What percentage of the total is that number?



You can also do it backward..

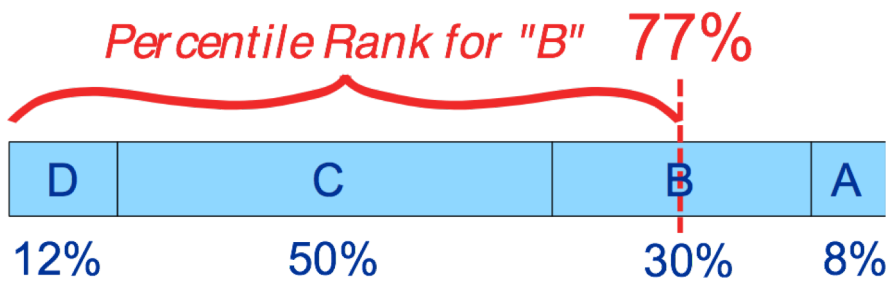
If you have a percentile can you determine which element is at that percentile? Yes.

Find the element after 80% of the data.



Percentile for Grouped Data

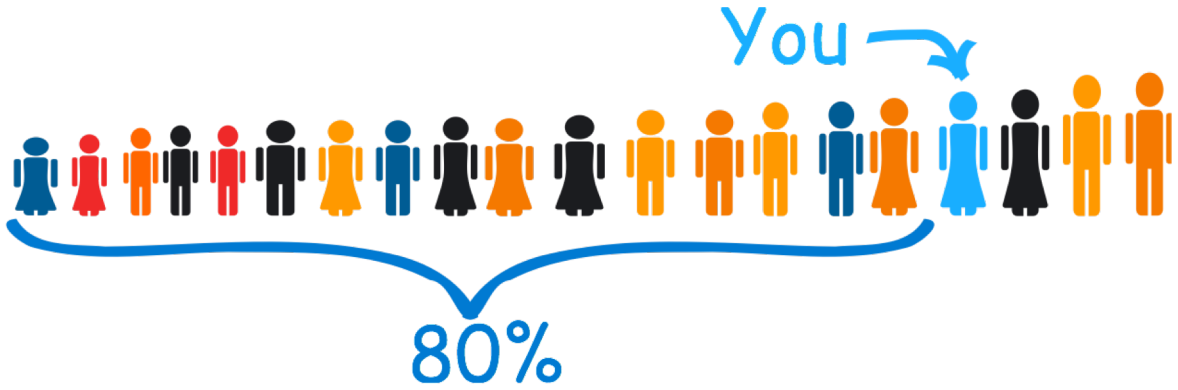
- Add up the groups below the group of interest
- Then add half of the group of interest



12 + 50 + (30/2) = 77%

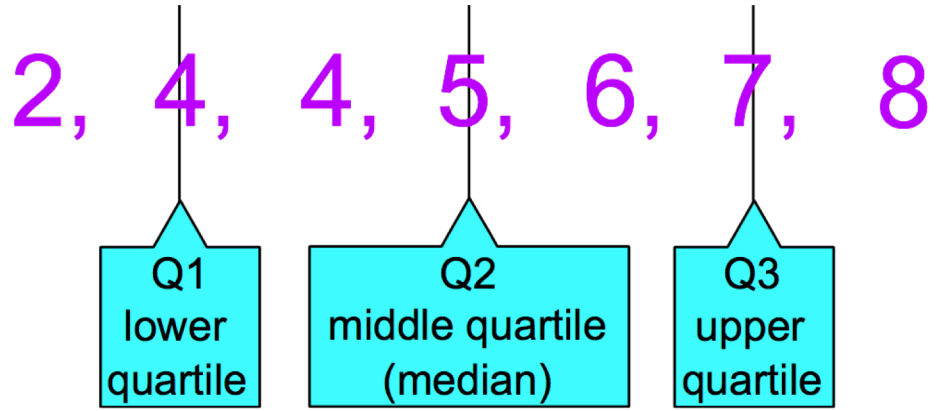
Deciles

- Deciles are just percentiles divided by 10
- In this example, you are the 8th decile:



Quartiles

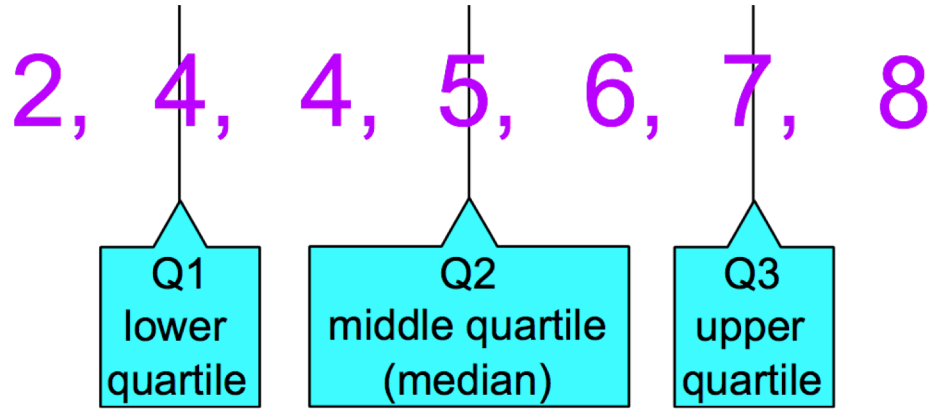
- Quartiles are percentiles divided by 25



- There is a lot of disagreement about what quartiles actually are:
  - ***Quartiles are the ranges that fall within the boundaries***
  - ***Quartiles are the boundaries themselves***

Quartiles

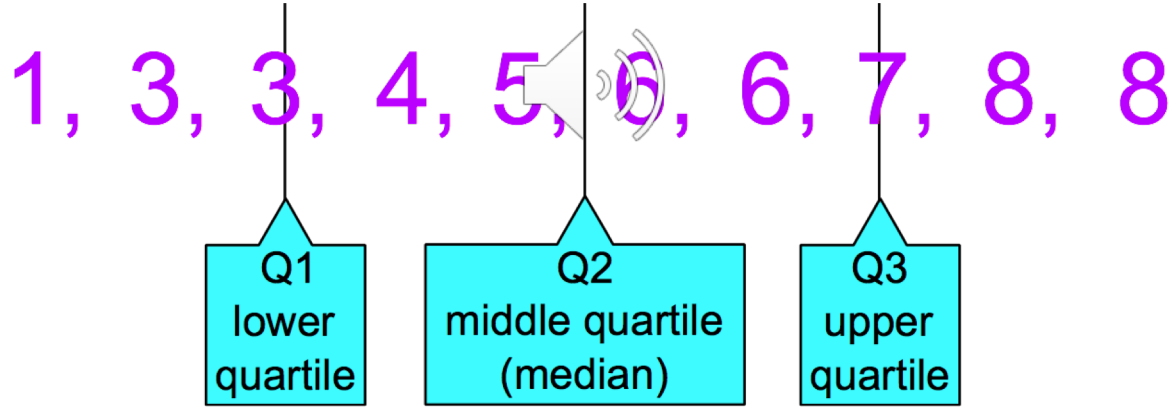
- Quartiles are percentiles divided by 25
- In this example, 5 is second quartile



For INF 110, Quartiles are values at boundaries not ranges!

Quartiles

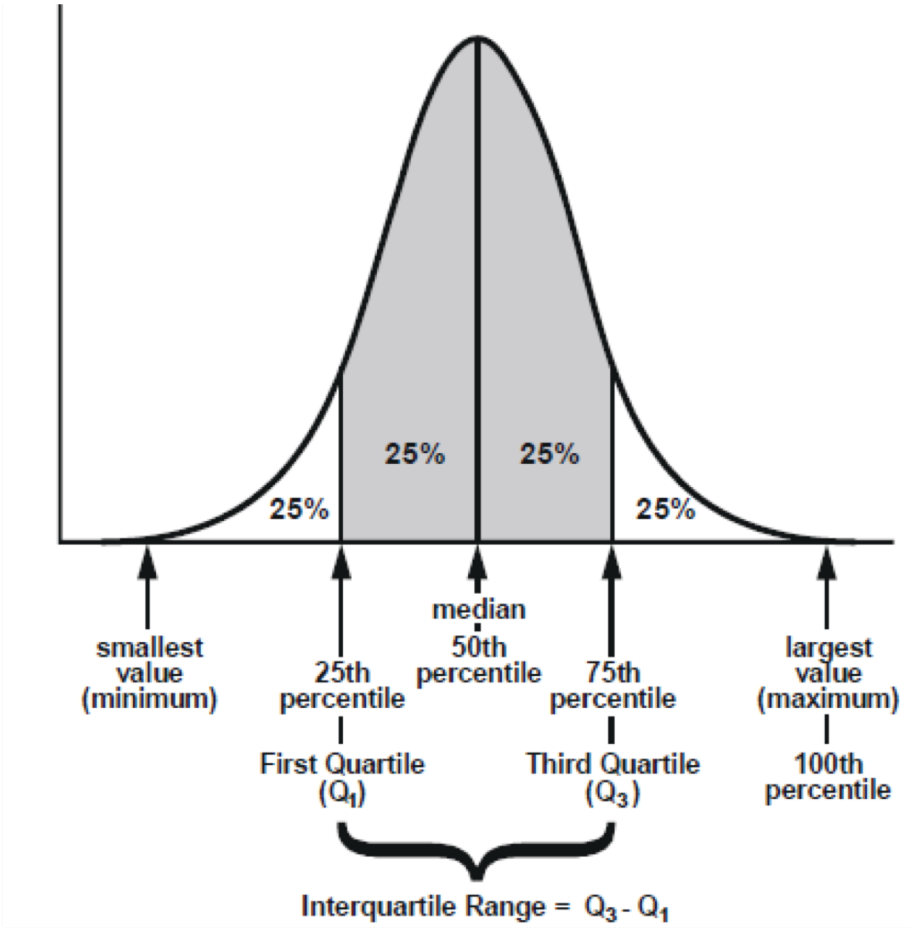
If a quartile falls between two values the quartile is the average of the two



Q2 = 5.5

Quartiles

- The middle half of values in a normal distribution centered at Q2 fall between Q1 and Q3
- This is called the Interquartile Range (IQR)



Percentile Gotchas

- If all the values are the same (e.g., 100, 100, 100, 100), what are the quartiles?
- What if there are too few samples (e.g., 100, 100)?

- If this happens, just document it and mention it when it's time to present results.

Live Code percentile

Tasks: The datascience module comes with a function that finds a value at a given percentile. Try this out on some random data. Use the sort function to check your work.

Learning Outcomes

- Using the percentile method
- Understanding random sets

Bootstrapping

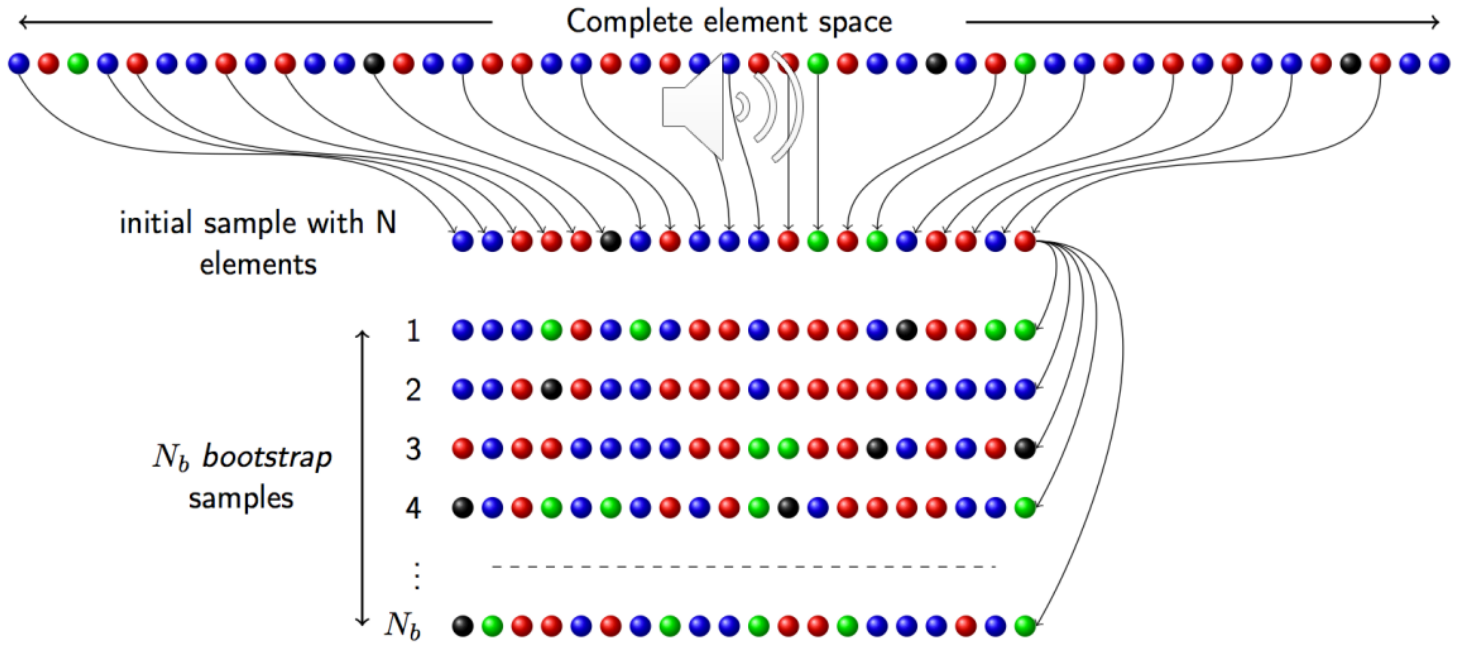
Sometimes we have a sample that is big and that we believe reflects the population but it's not big enough.

That is, we need more data.

We can use the data we have to create more pseudodata through a bootstrap.

Bootstrapping

Idea: Sample with replacement (don't take it out) for as much data as you need:



Live Code Travel Time

Tasks: Create a table of travel times from home to school in minutes (round by 10). Use bootstrapping to create a population of 100 student's travel times using the sample method.

Learning Outcomes

- Using the sample method
- Understanding random sets

Live Code Travel Time Part 2

Tasks: Using the bootstrapped data, what are the quartiles in the set?

Learning Outcomes

- Using the percentile method
- Understanding random sets


end