

BASIC DESCRIPTIVE STATISTICS

Monique Golden

DON'T BE FOOLED!



OBJECTIVES

- Understand **distribution**, measures of **central tendency**, and **dispersion**
- Gain preliminary knowledge of assumptions of **normality to make valid results interpretation**
- Apply concepts to examples to **build confidence** in BDS reporting

What is the average?

To what extent?

BUT FIRST...

How many?

How frequent?

What proportion?

what
do *you*
want
to *know*?

What is the most common?

I HAVE MY QUESTIONS AND MY DATA.
NOW WHAT?

ALWAYS PLOT YOUR DATA!

The distribution of a variable is a description of the relative numbers of times each possible outcome will occur in a sampling event.

NORMAL DISTRIBUTION

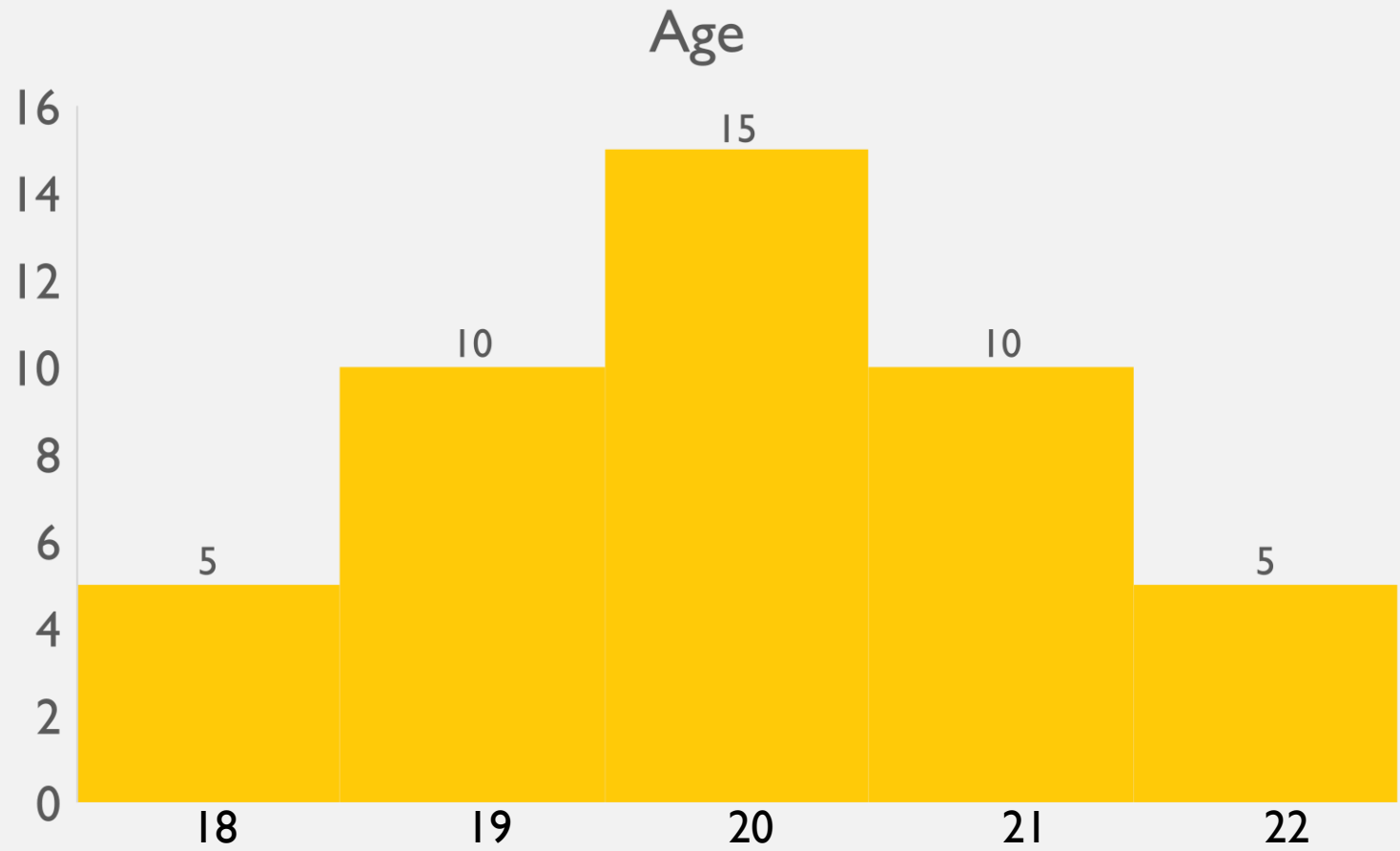
For our purposes...

- Continuous
- Smooth
- One “peak”
- Rises then falls just once

CHECK FOR IRREGULARITIES

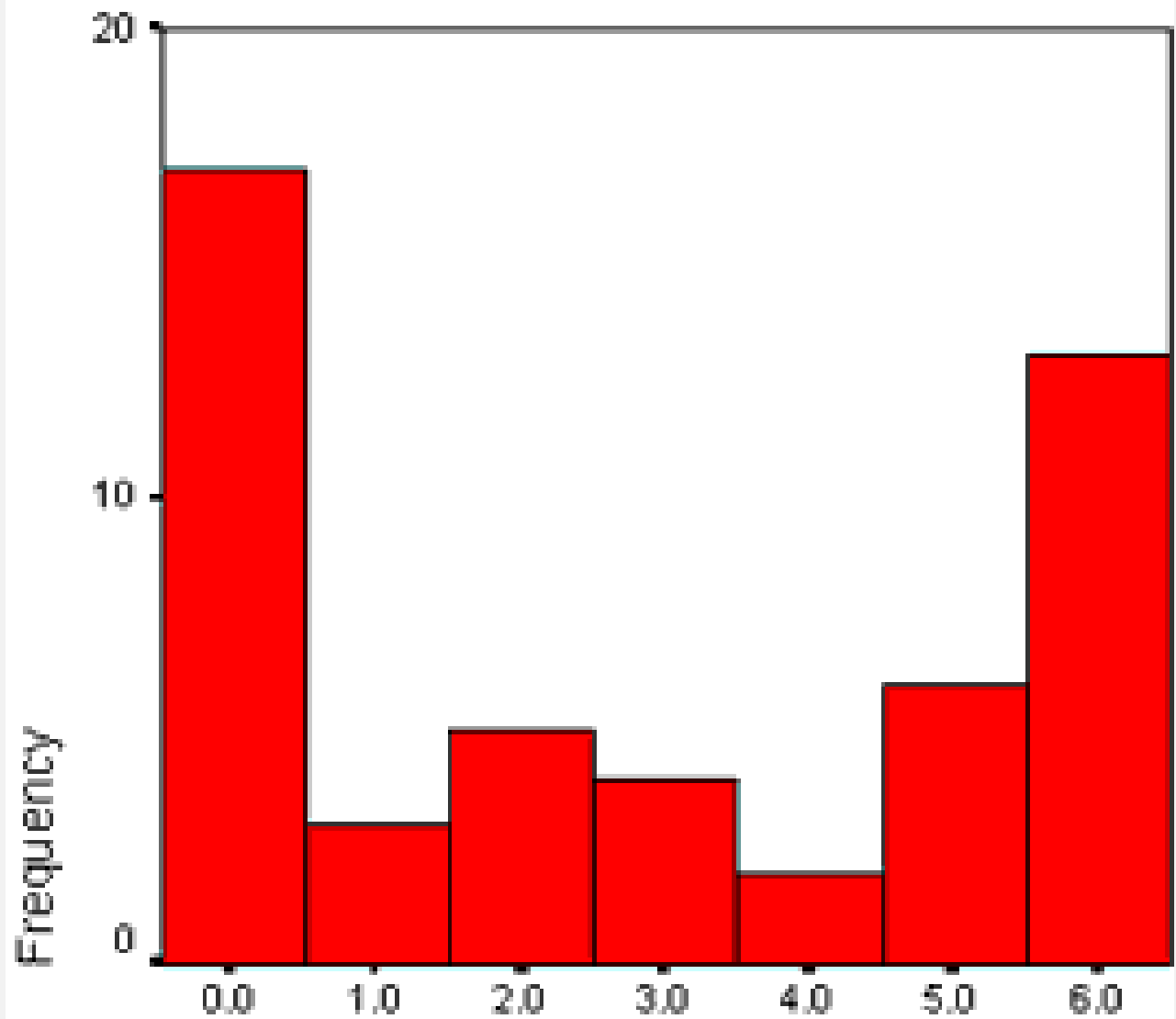
- Use a scatter plot to identify “outliers”; these can “throw off” estimates
- Use a histogram to see “shape” of distribution (smallest to largest values)
- More than one “peak” suggests systematic issues possibly related to sampling or instrumentation

PLOT HISTOGRAM

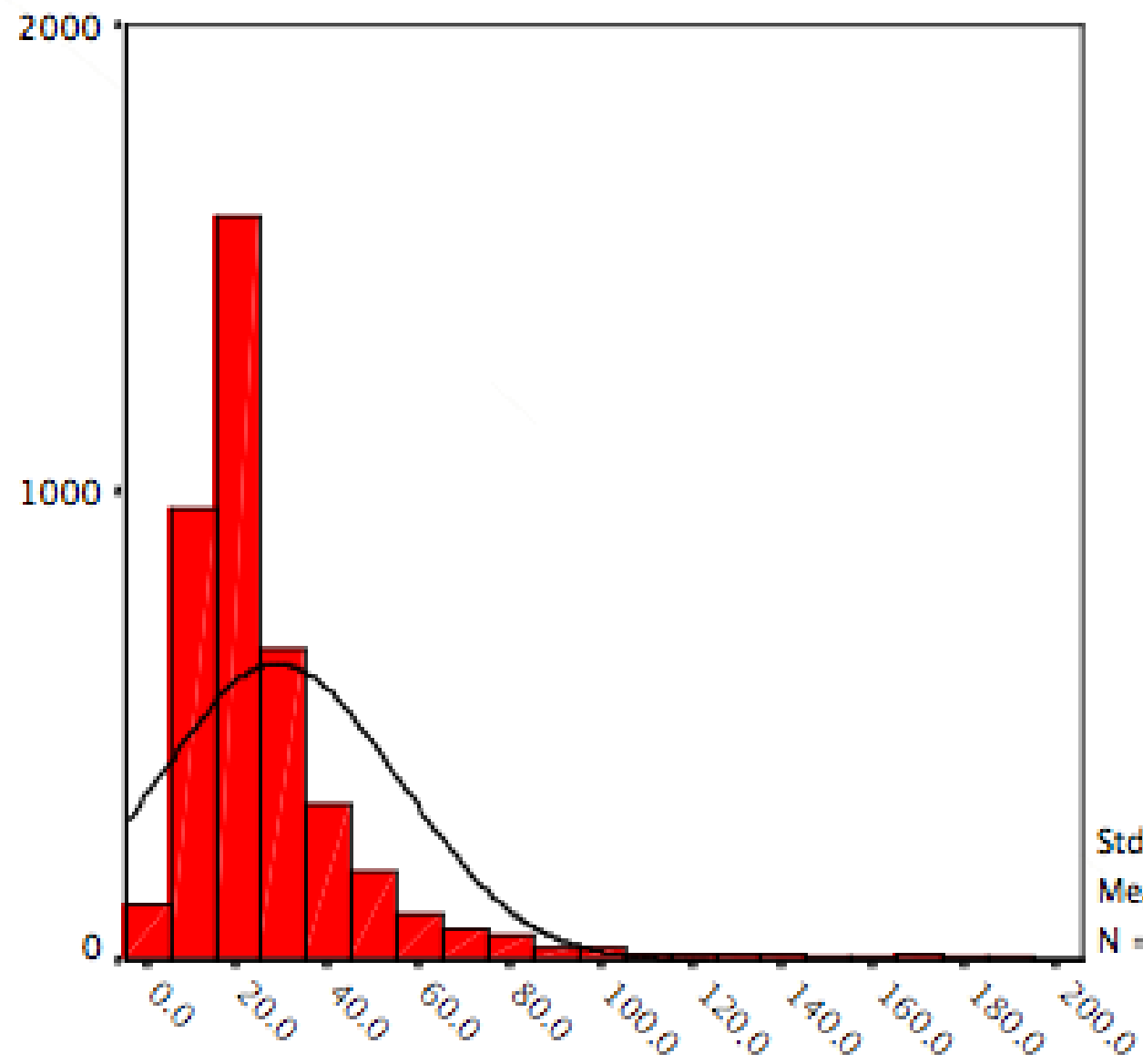


EXAMPLES OF NON-NORMAL DISTRIBUTIONS

NIC

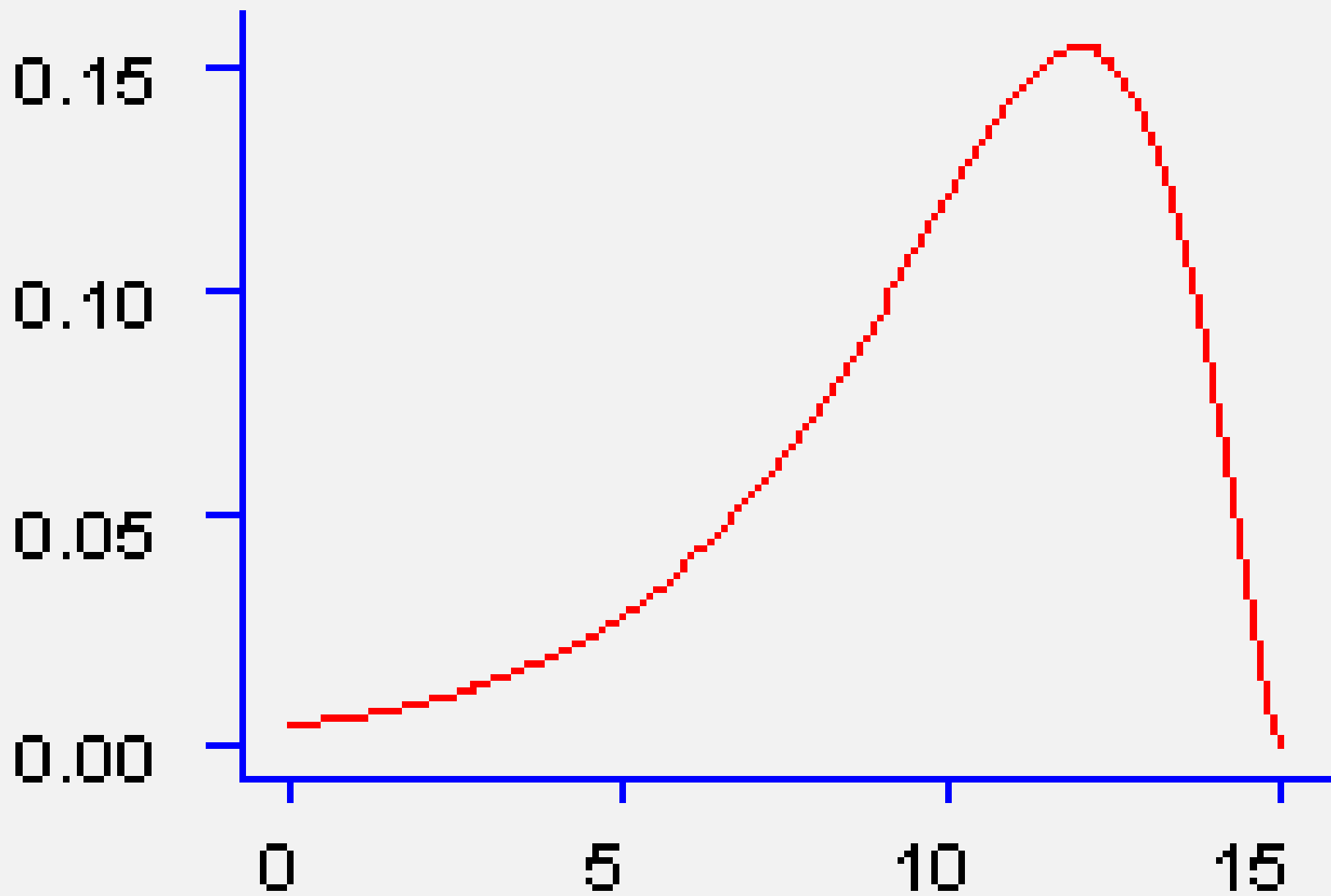


NIC



Std. Dev = 26.92
Mean = 28.9
N = 4269.00

Current PE



MY PLOT LOOKS NORMAL.
WHAT'S NEXT?

MEASURES OF CENTRAL TENDENCY

MODE



MEDIAN



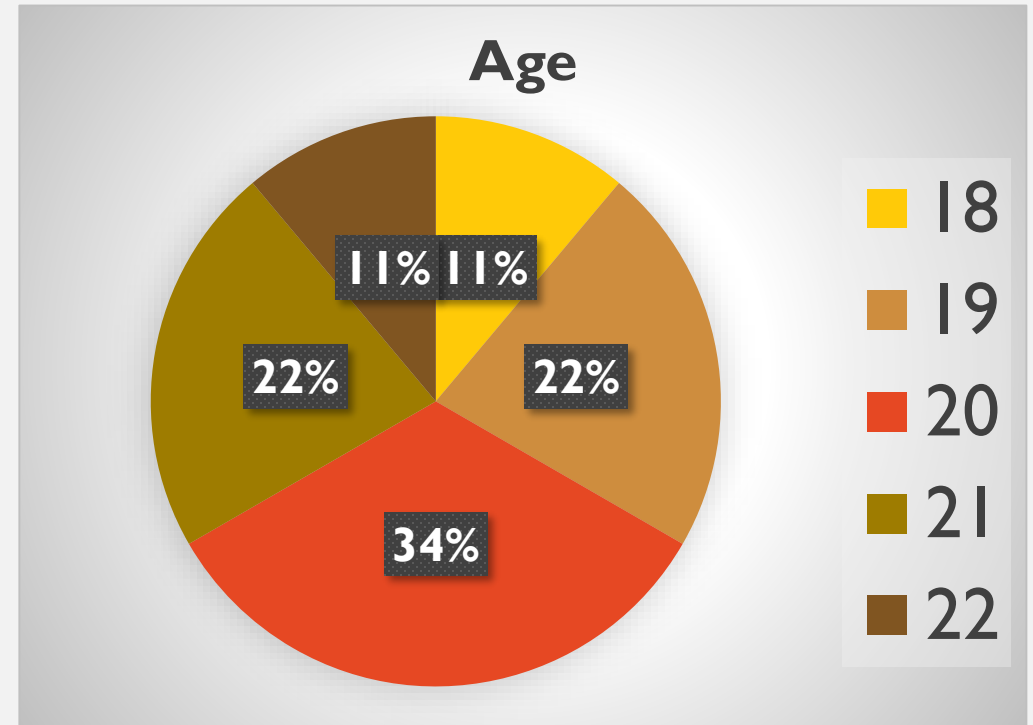
MEAN



REPRESENTING FREQUENCIES

The most commonly reported age (mode) is 20.

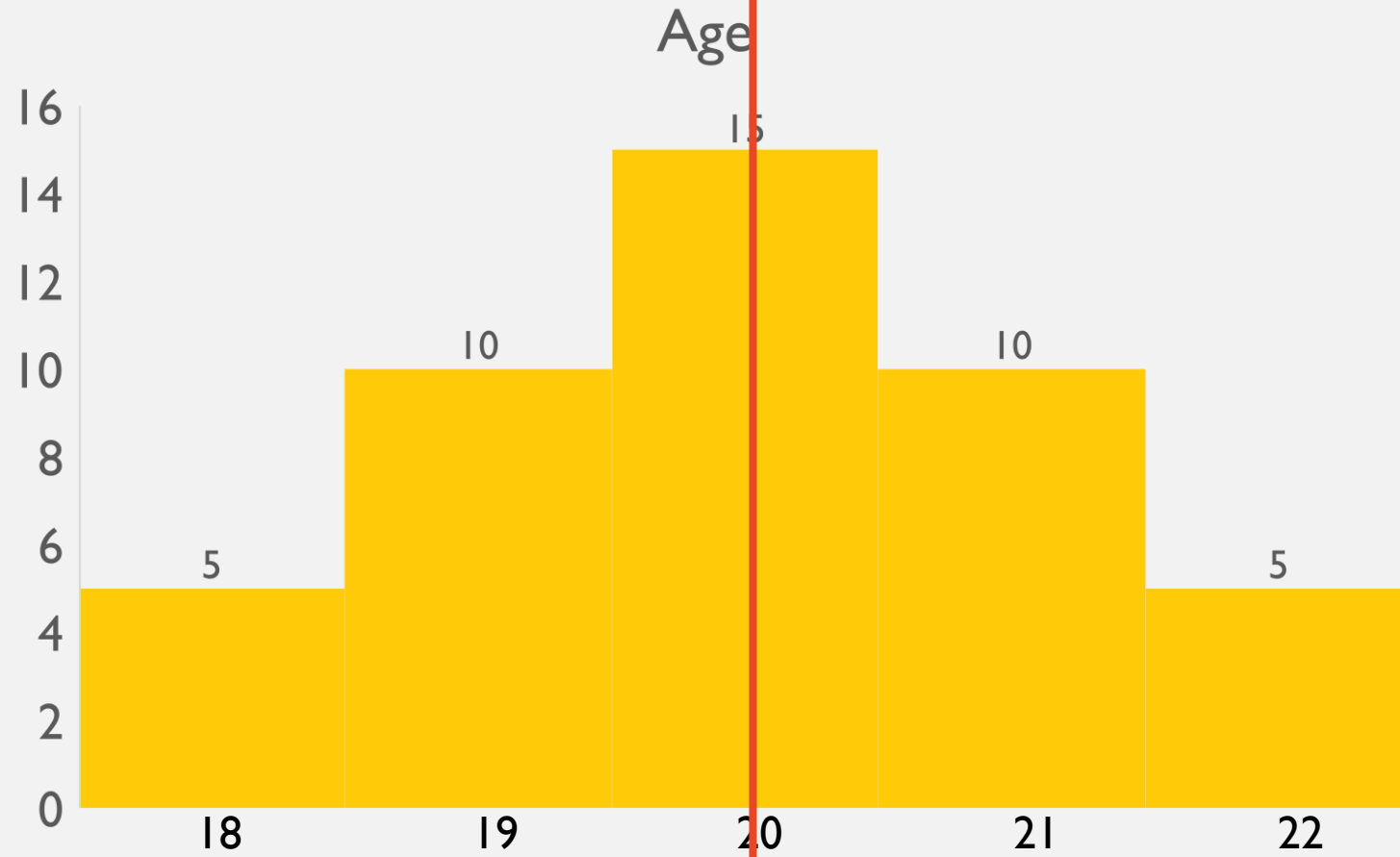
AGE	n
18	5
19	10
20	15
21	10
22	5
Total	45



Thirty-four percent of respondents indicated their age at the time of the study was 20.

MEDIAN

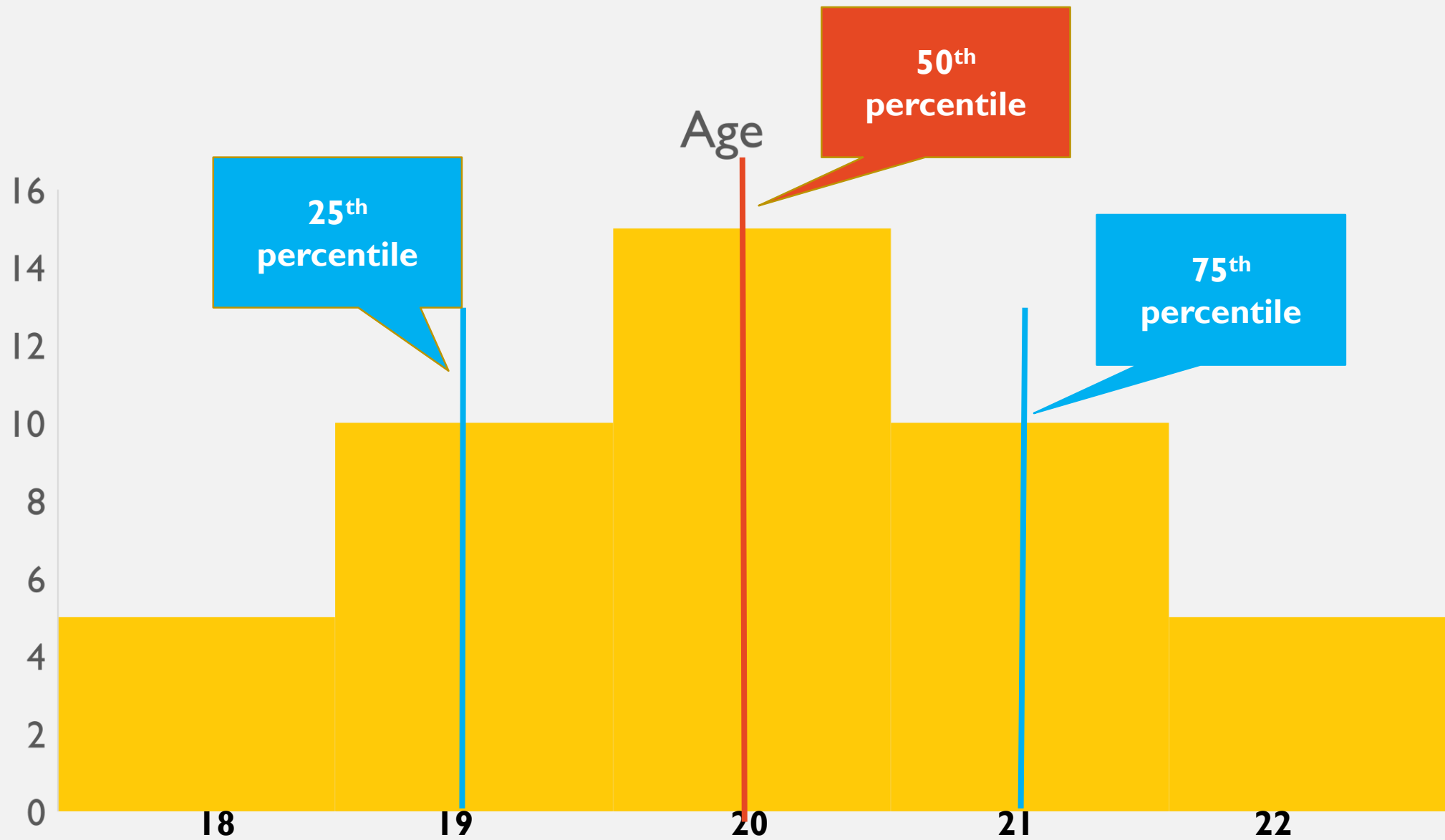
The median age of our sample is 20.



LET'S DISCUSS QUARTILES

three values of the random variable that divide a population into four equal groups

Using the quartile function in Excel...



MEASURES OF CENTRAL TENDENCY

MODE



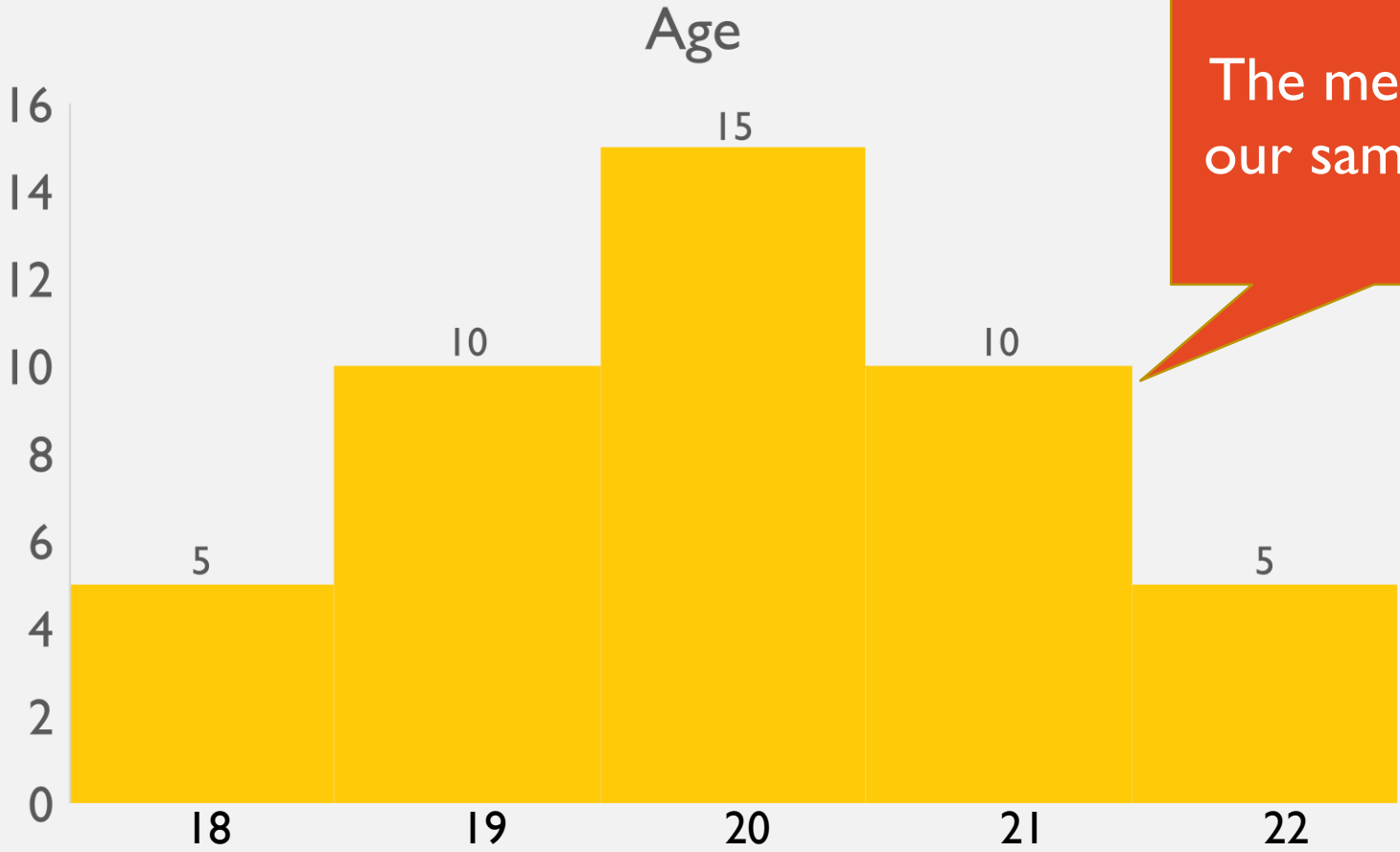
MEDIAN



MEAN



MEAN



The average age of our sample is 20.

The expected age of this community is 20 years.

The mean age of our sample is 20.

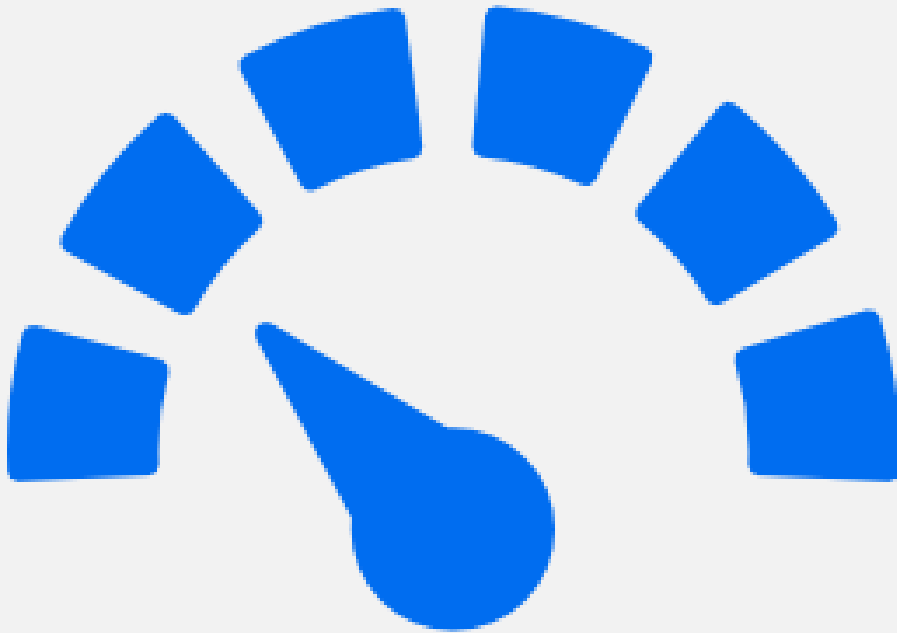
On average, respondents are 20 years old.

I GET THE **POINT**, BUT *DO YOU?*

I MEAN IT'S ALRIGHT

- The mean is a valuable statistic, but merely a point estimate...
- It gives us no information about how “good” of an estimate it is based on our data and RQs/AQs/Interest
- Variance gives us this information!

VARIANCE AND STANDARD DEVIATION



- Estimates suggest imprecise or inexact measure of variable
- Variance is a measure of inexactness, impreciseness wrt data
- Sometimes more variation is better; sometimes it is not!

CALCULATING SAMPLE VARIANCE

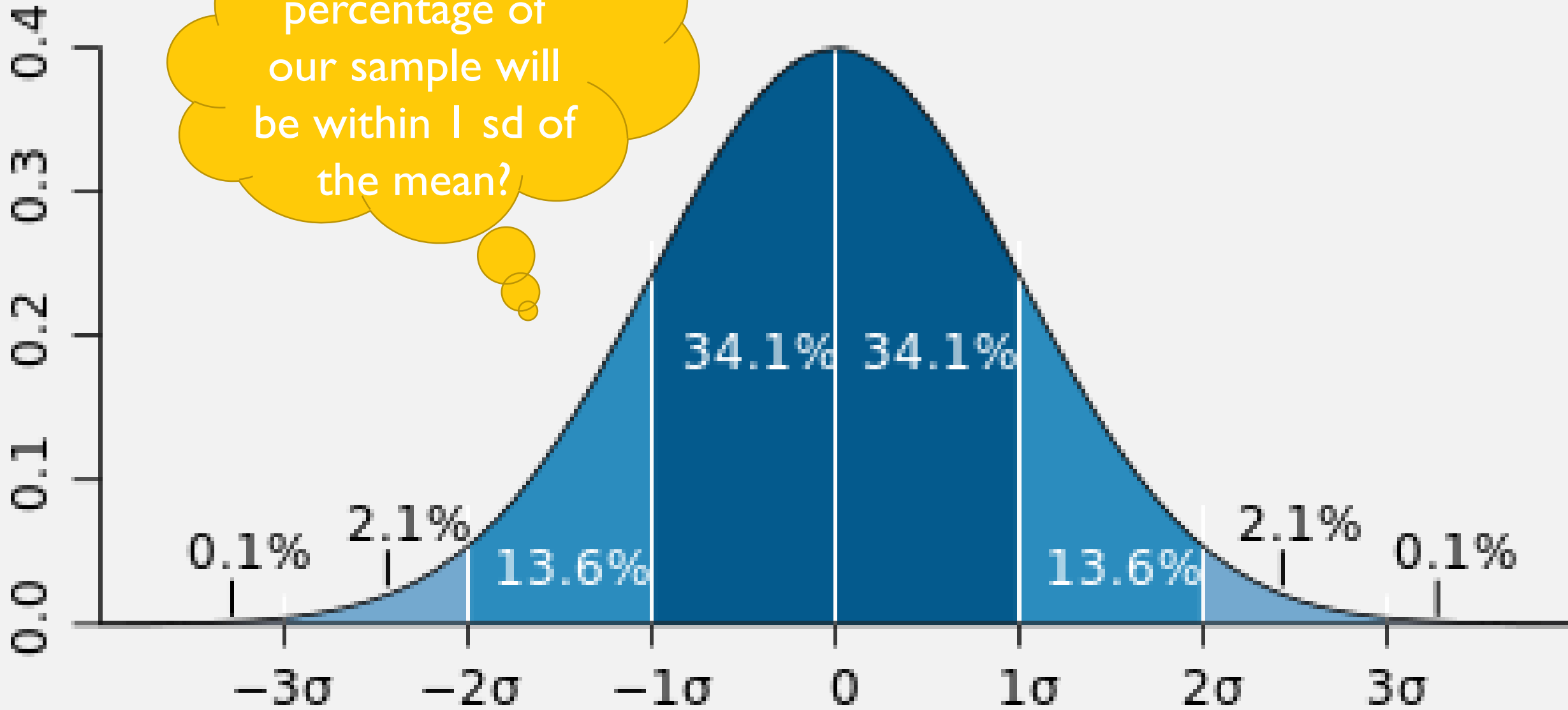
Age	Mean-Age	(Mean-Age) ²
19	20-19=1	1
20	20-20=0	0
18	20-18=2	4
18	20-18=2	4
22	20-22=-2	4
21	20-21=-1	1
For total sample, sample var= 1.36		

The variance is calculated by the total squared deviation divided by one less the number of observations

STANDARD DEVIATION

- Standard deviation (sd) is the average (expected) impreciseness of an estimate (how wide or narrow the distribution is)
- $\sqrt{\text{sample variance}}$
- Represents area under curve
- Formulate confidence intervals

What percentage of our sample will be within 1 sd of the mean?



SUMMING IT UP

THE GOLDEN KEYS TO SUCCESS



Don't be scared to analyze data! Start with your RQs/AQs/Interests



Plot your data, check for irregularities



Based on RQs, use most appropriate statistic



Before you report, ask *does this make sense?* Be practical

THANK YOU AND GOOD LUCK WITH
YOUR DATA ANALYSIS ENDEAVORS.

QUESTIONS?

I can be reached via email

Monique.golden@uconn.edu