USING ALL OF THESE YOU CAN STUDY ANY PHENOMENA AND DRAW INFERENCES ABOUT THEM

RANDOM VARIABLES

PROBABILITY DISTRIBUTIONS

SAMPLING

HYPOTHESIS TESTING

SAMPLING IS A LITTLE BIT LIKE FISHING

SAMPLING IS A LITTLE BIT LIKE FISHING

YOU ARE A MARINE BIOLOGIST

YOU LEARN ABOUT FISH

HOW?

SAMPLING IS A LITTLE BIT LIKE FISHING

YOU ARE A MARINE BIOLOGIST
YOU LEARN ABOUT FISH
HOW?

YOU DO THIS BY

1. CATCHING SOME FISH

2. STUDYING THE CAUGHT FISH

3. DRAW CONCLUSIONS ABOUT ALL THE FISH IN THE SEA

SAMPLING IS A LITTLE BIT LIKE FISHING

YOU ARE A MARINE BIOLOGIST
YOU LEARN ABOUT FISH
HOW?

YOU DO THIS BY

POPULATION

1. CATCHING SOME FISH
2. STUDYING THE CAUGHT FISH
3. DRAW CONCLUSIONS ABOUT

ALL THE FISH IN THE SEA

SAMPLING IS A LITTLE BIT LIKE FISHING

YOU ARE A MARINE BIOLOGIST
YOU LEARN ABOUT FISH
HOW?

SAMPLE

YOU DO THIS BY

1. CATCHING SOME FISH
2. STUDYING THE CAUGHT FISH
3. DRAW CONCLUSIONS ABOUT ALL THE FISH IN THE SEA

POPULATION

SAMPLING IS A LITTLE BIT LIKE FISHING

YOU ARE A MARINE BIOLOGIST
YOU LEARN ABOUT FISH
HOW?

SAMPLING

YOU DO THIS BY

1. CATCHING SOME FISH
(SAMPLE)
2. STUDYING THE CAUGHT FISH
3. DRAW CONCLUSIONS ABOUT ALL THE FISH IN THE SEA
POPULATION

DRAWING CONCLUSIONS ABOUT THE POPULATION BY OBSERVING THE SAMPLE IS CALLED

GENERALIZATION

DRAWING CONCLUSIONS ABOUT THE POPULATION BY OBSERVING THE SAMPLE IS CALLED

GENERALIZATION

DRAWING CONCLUSIONS ABOUT THE POPULATION BY OBSERVING THE SAMPLE IS CALLED

GENERALIZATION

INFERENCES

DRAWING CONCLUSIONS ABOUT THE POPULATION BY OBSERVING THE SAMPLE IS CALLED

GENERALIZATION

PSYCHOLOGICAL STUDIES

POLLING

DRUG TRIALS

A/B TESTS

MARKET RESEARCH SURVEYS

PSYCHOLOGICAL STUDIES
POLLING
DRUG TRIALS
A/B TESTS
MARKET RESEARCH SURVEYS

THESE ARE JUST A FEW EXAMPLES OF WHERE SAMPLING MIGHT BE USED

GIVEN ANY SAMPLE

THE SAMPLE CAN BE DESCRIBED USING SAMPLE STATISTICS

THESE ARE NUMBERS THAT CHARACTERIZE THE SAMPLE

SAMPLE STATISTICS

DEPENDING ON THE TYPE OF VARIABLE YOU ARE MEASURING, YOU MIGHT BE INTERESTED IN

SAMPLE MEAN

OR

SAMPLE PERCENTAGE

SAMPLE STATISTICS

SAMPLE MEAN

USED WHEN THE VARIABLE IS CONTINUOUS

HEIGHTS OF A GROUP OF PEOPLE

USER ENGAGEMENT ON A WEBSITE

SAMPLE PERCENTAGE

SAMPLE STATISTICS

SAMPLE MEAN USED WHEN THE VARIABLE IS CONTINUOUS

SAMPLE PERCENTAGE

NORMALLY USED WHEN THE VARIABLE IS BINARY (YES/NO)

DO YOU SUPPORT THIS CANDIDATE?

IS THIS DRUG AN EFFECTIVE TREATMENT?

SAMPLE STATISTICS

SAMPLE MEAN / SAMPLE PERCENTAGE

THE NEXT IMPORTANT STATISTIC TO KNOW IS THE STANDARD DEVIATION OF THE SAMPLE

SAMPLE STATISTICS

SAMPLE MEAN / SAMPLE PERCENTAGE

THE STANDARD DEVIATION OF THE SAMPLE

THIS IS DIFFERENT FOR MEAN VS %
MEASUREMENTS

SAMPLE STATISTICS

SAMPLE MEAN / SAMPLE PERCENTAGE

THE STANDARD DEVIATION OF THE SAMPLE

IF WE ARE INTERESTED IN SAMPLE MEAN

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

THIS IS STRAIGHTFORWARD,
THE USUAL FORMULA

SAMPLE STATISTICS

SAMPLE MEAN / SAMPLE PERCENTAGE

THE STANDARD DEVIATION OF THE SAMPLE

IF WE ARE INTERESTED IN SAMPLE %

$$\text{s.d.}(p) = \sqrt{\frac{p\,(1-p)}{n}}$$

P IS THE % OF YESES

LET'S IMAGINE YOU PICKED
100 DIFFERENT SAMPLES

WITH EACH YOU COMPUTED
THE SAMPLE MEAN/%

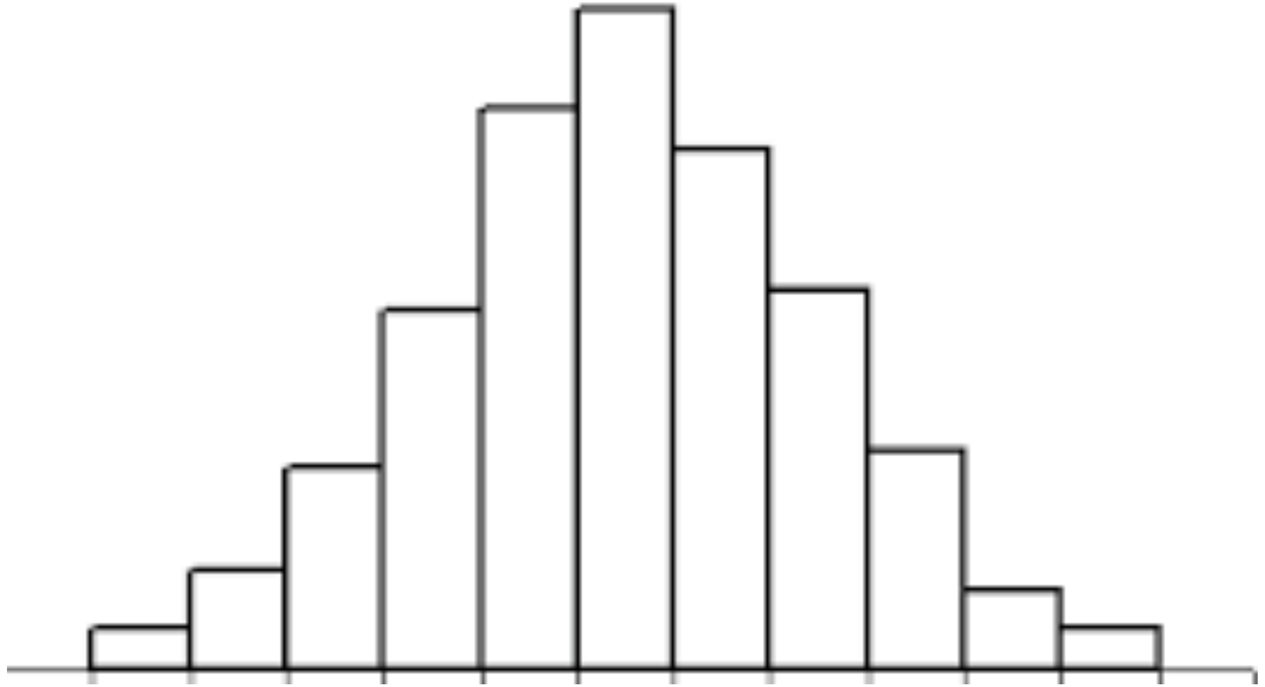
SAMPLE 1	53%
SAMPLE 2	55%
SAMPLE 3	54%
.....	.....
SAMPLE 100	49%

SAMPLE 1	53%
SAMPLE 2	55%
SAMPLE 3	54%
.....	.....
SAMPLE 100	49%

THESE VALUES REPRESENT A VALUES OF RANDOM VARIABLE

SAMPLE 1	53%
SAMPLE 2	55%
SAMPLE 3	54%
.....	.....
SAMPLE 100	49%

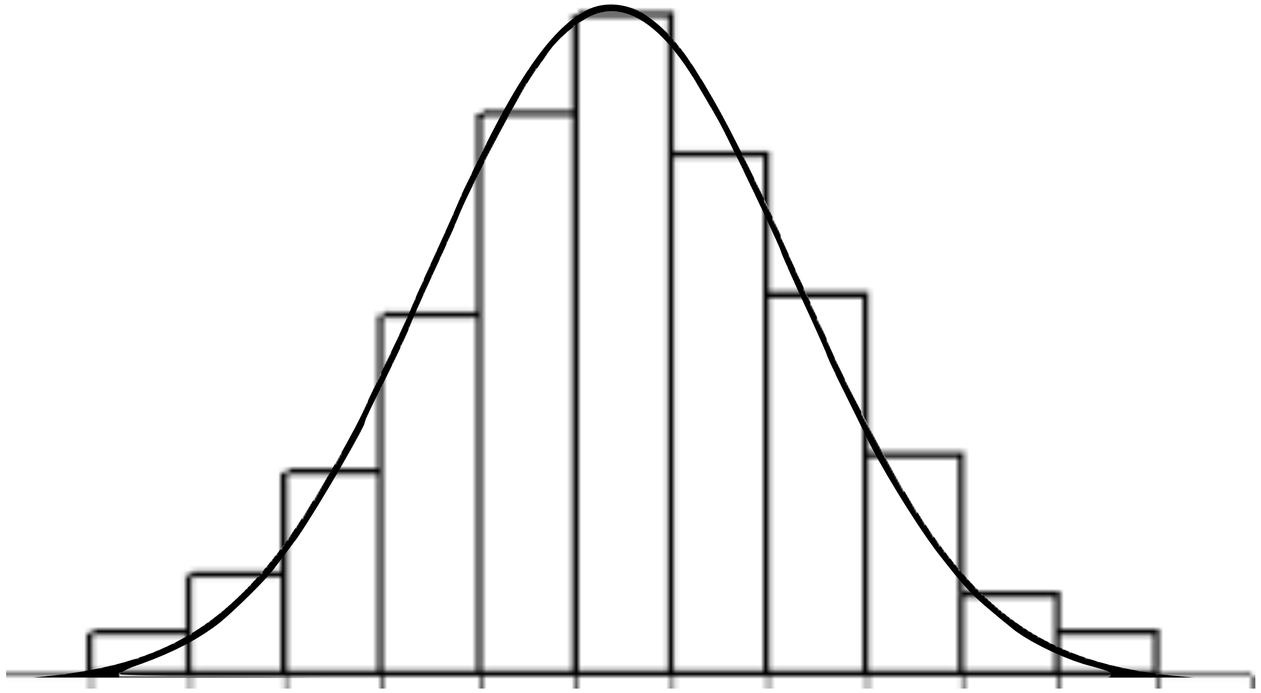
IF YOU PLOT THE HISTOGRAM OF THESE VALUES, THAT WILL REPRESENT IT'S PROBABILITY DISTRIBUTION



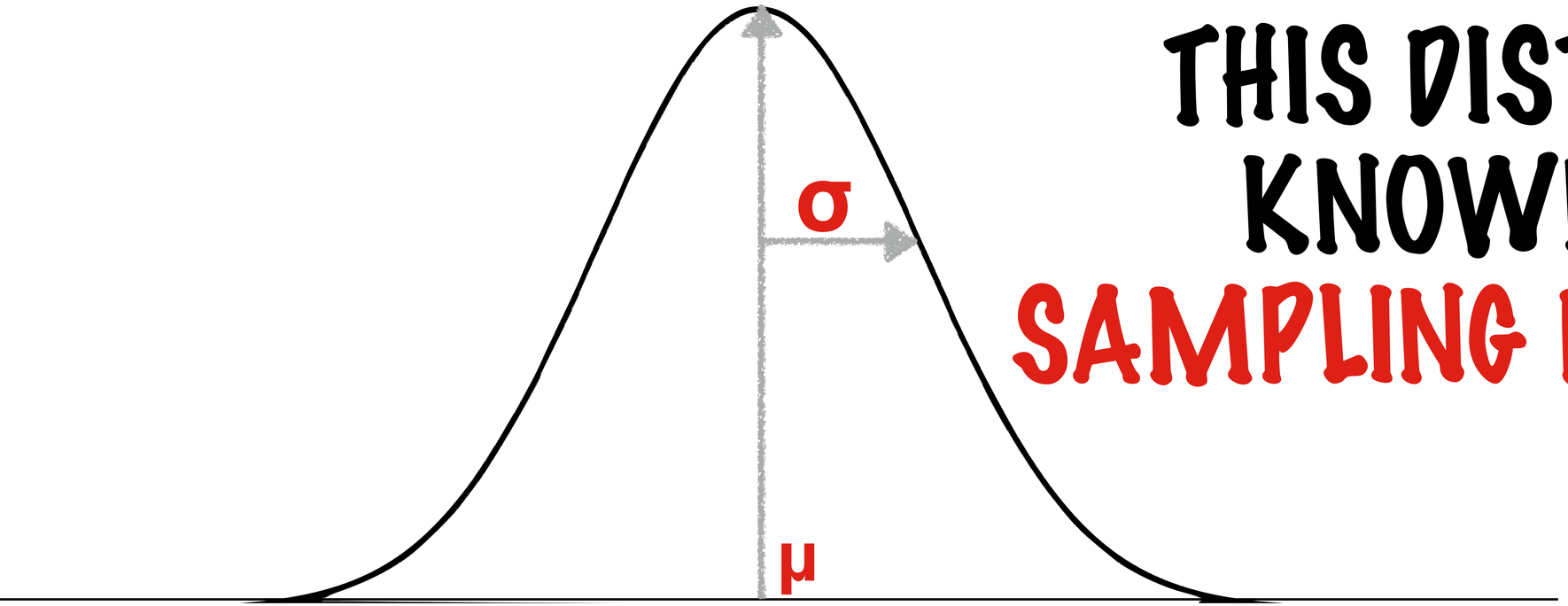
HOW OFF IS THE % OF THE SAMPLE FROM THE % OF THE POPULATION?

SAMPLE 1	53%
SAMPLE 2	55%
SAMPLE 3	54%
.....	.....
SAMPLE 100	49%

IT TURNS OUT, WHENEVER YOU TAKE A LARGE NUMBER OF SAMPLES, THE % (OR MEANS) OF THOSE SAMPLES FOLLOW A NORMAL DISTRIBUTION

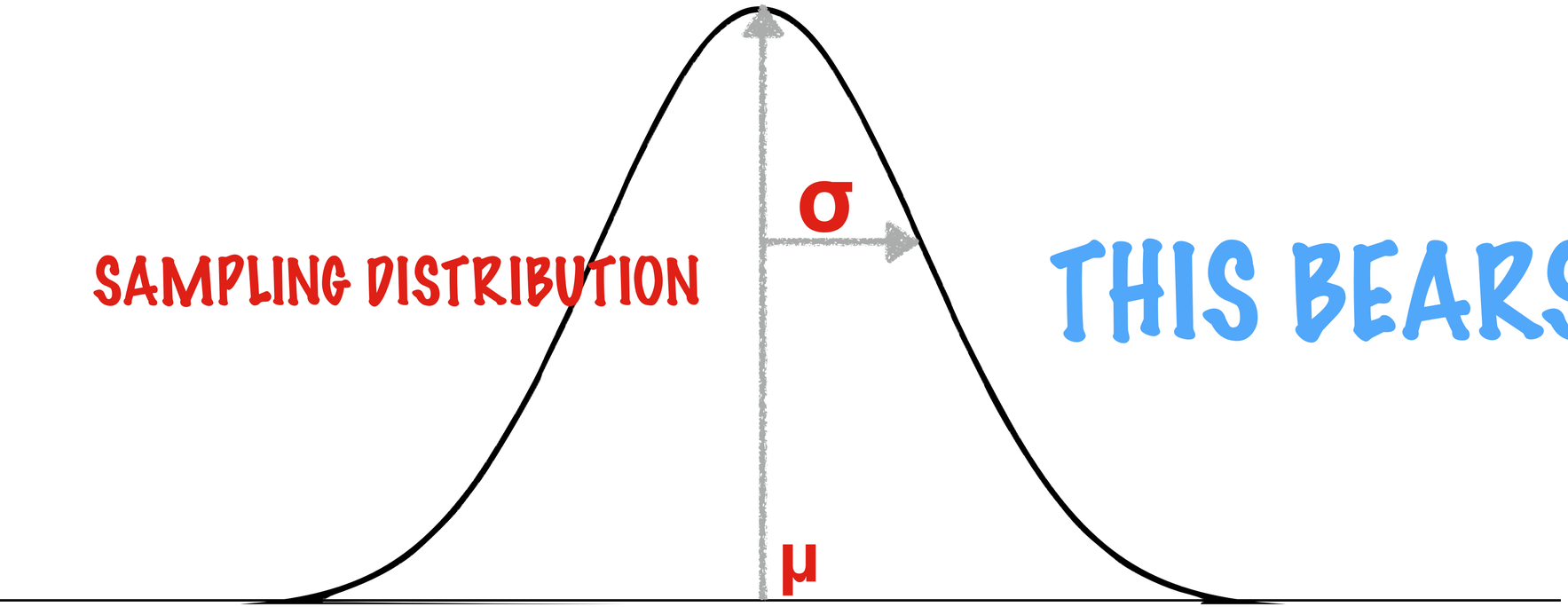


HOW OFF IS THE % OF THE SAMPLE FROM THE % OF THE POPULATION?



THIS DISTRIBUTION KNOWN AS THE SAMPLING DISTRIBUTION

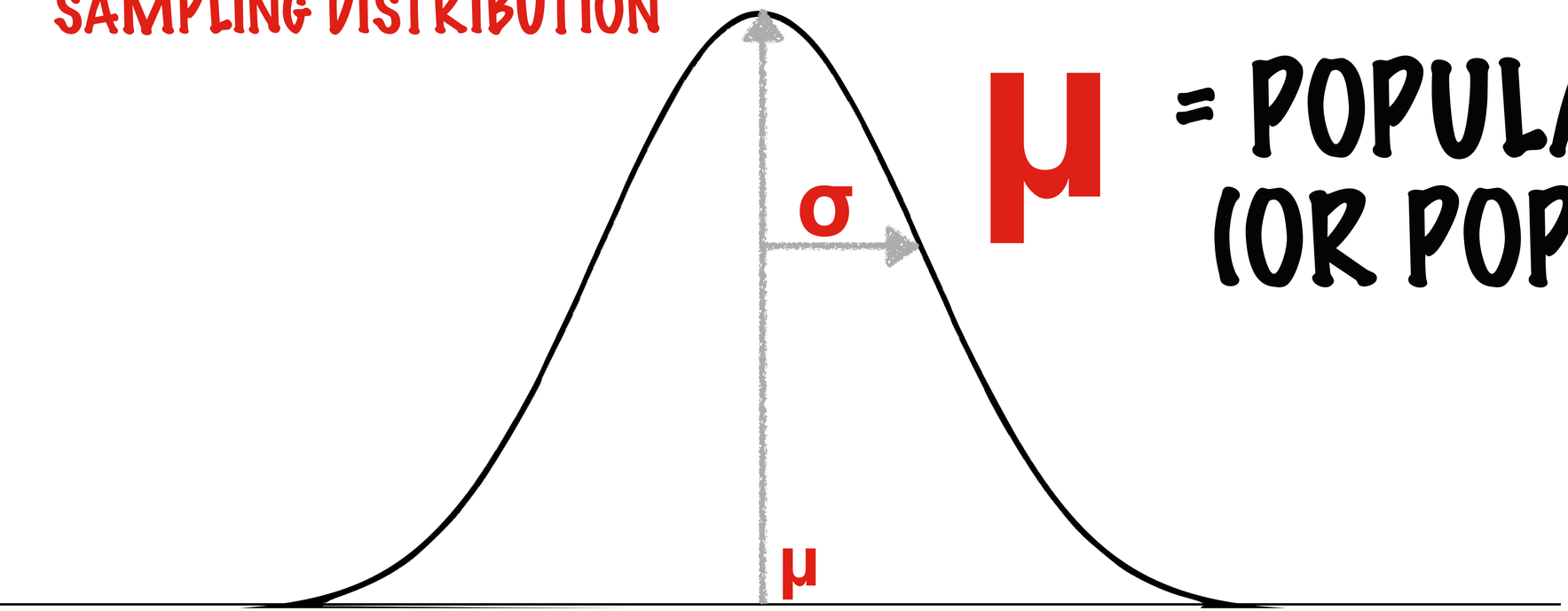
HOW OFF IS THE % OF THE SAMPLE FROM THE % OF THE POPULATION?



THIS BEARS REPEATING

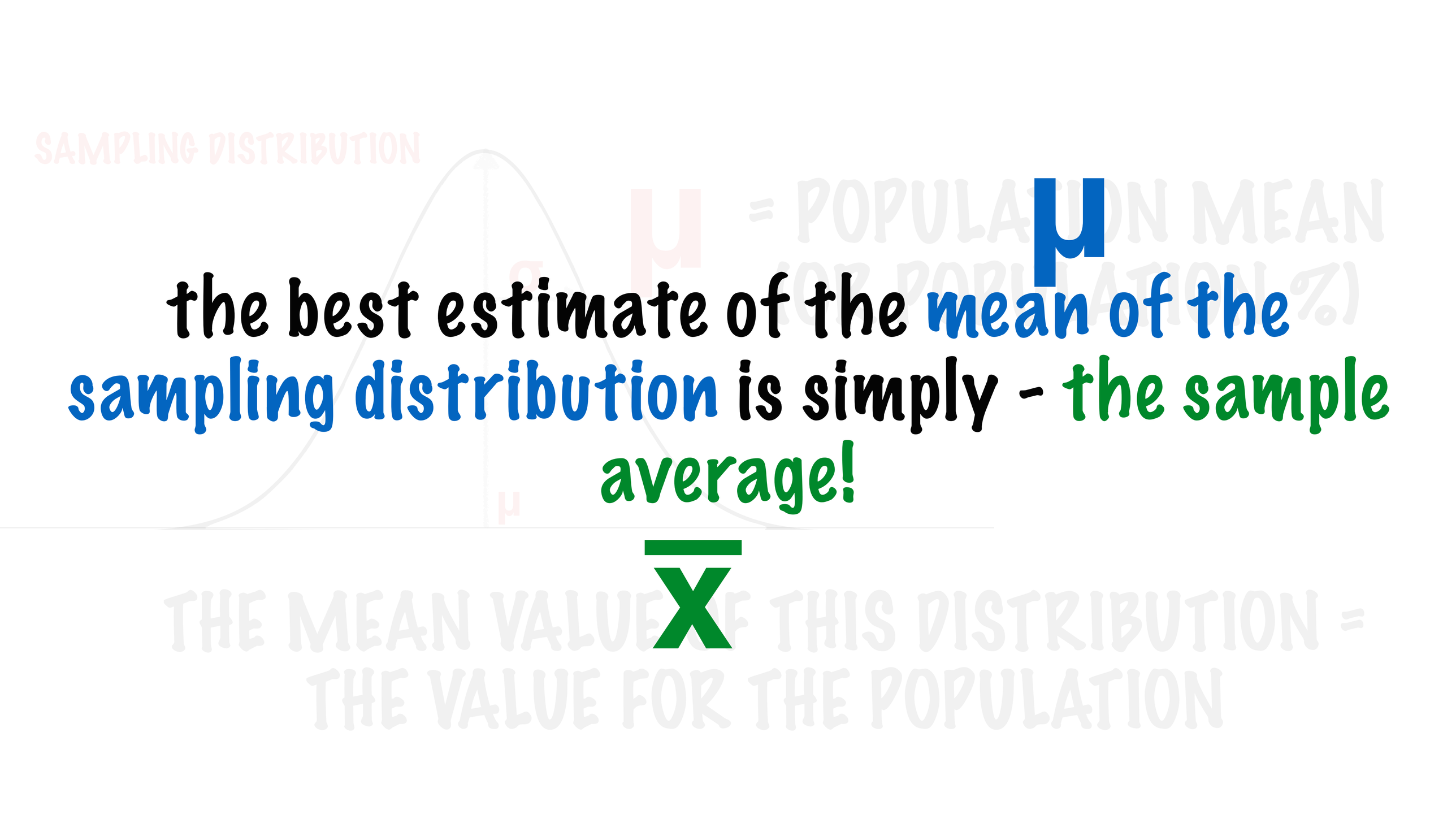
THE SAMPLING DISTRIBUTION OF A % (OR MEAN) IS THE DISTRIBUTION YOU GET WHEN YOU COMPUTE THAT VALUE FOR A LARGE NUMBER OF SAMPLES

SAMPLING DISTRIBUTION



μ = POPULATION MEAN
(OR POPULATION %)

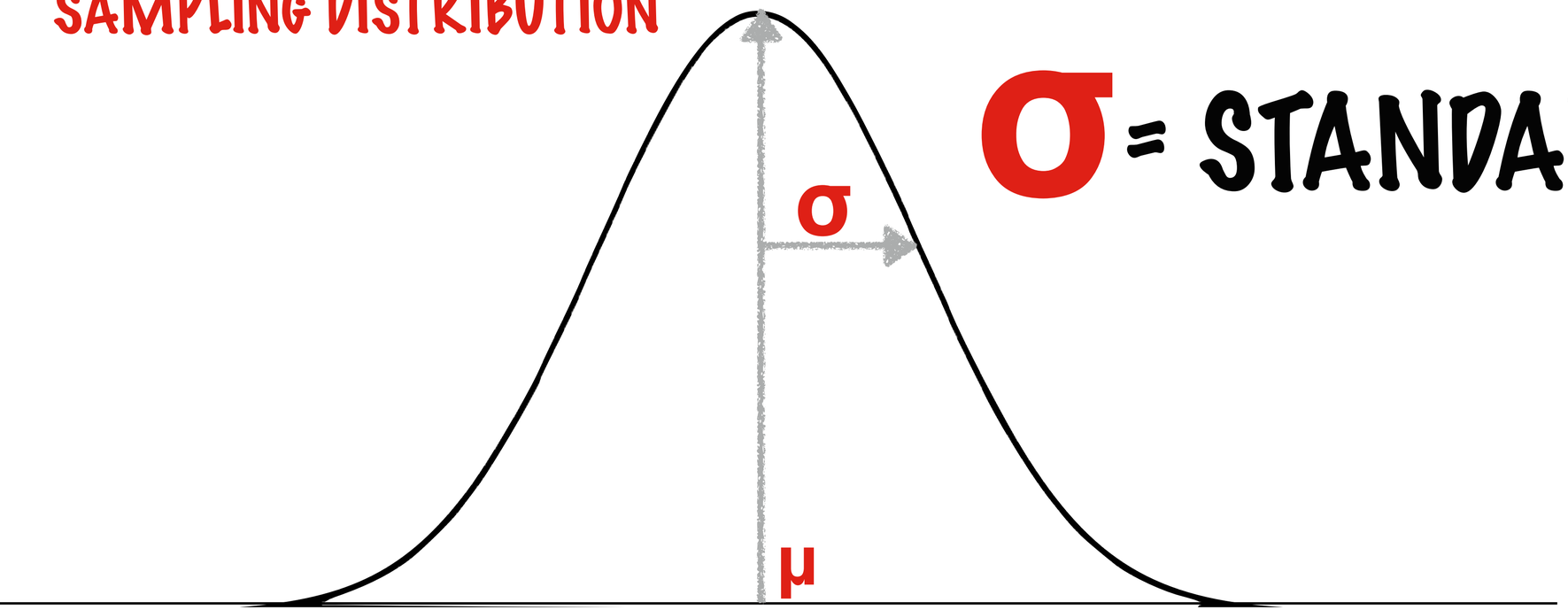
THE MEAN VALUE OF THIS DISTRIBUTION =
THE VALUE FOR THE POPULATION



the best estimate of the mean of the sampling distribution is simply - the sample average!

$\bar{x}$

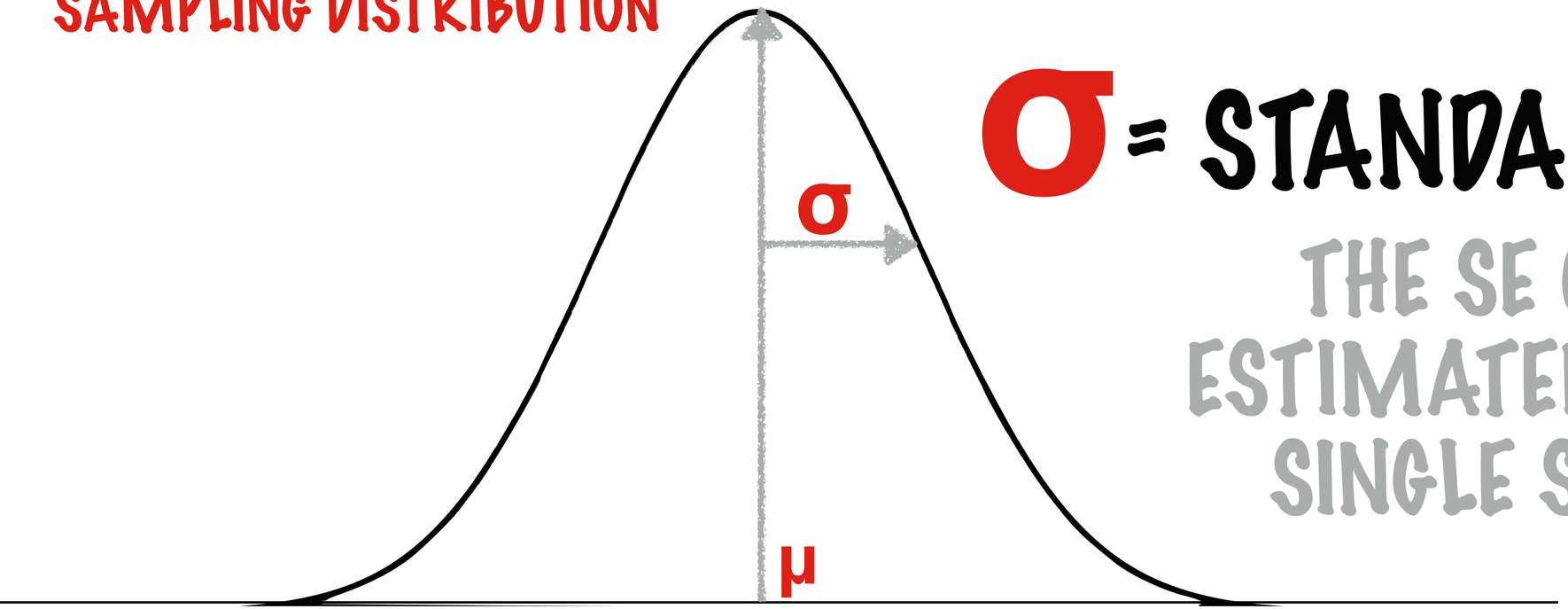
SAMPLING DISTRIBUTION



σ = STANDARD ERROR

THE SE CAN BE ESTIMATED FROM A SINGLE SAMPLE

SAMPLING DISTRIBUTION



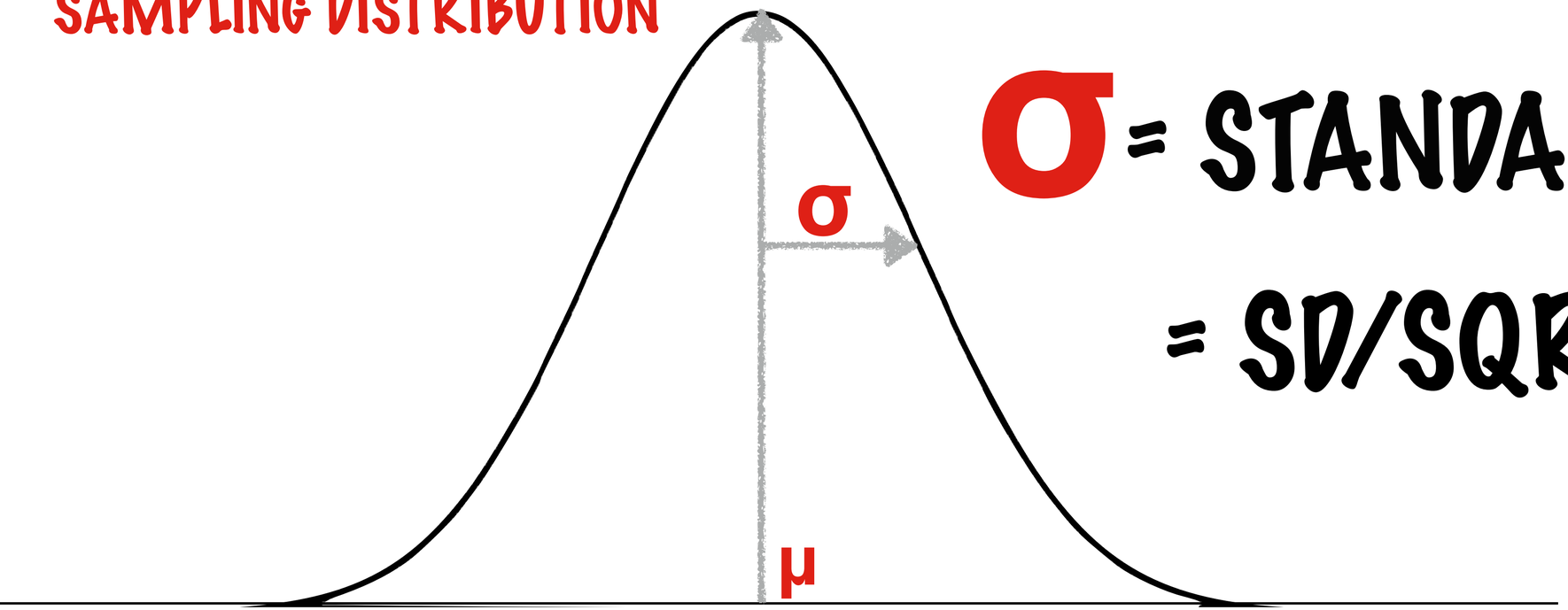
σ = STANDARD ERROR

THE SE CAN BE ESTIMATED FROM A SINGLE SAMPLE

IT DEPENDS ON THE SD AND THE SIZE OF THE SAMPLE

IF WE ARE INTERESTED IN SAMPLE MEAN

SAMPLING DISTRIBUTION



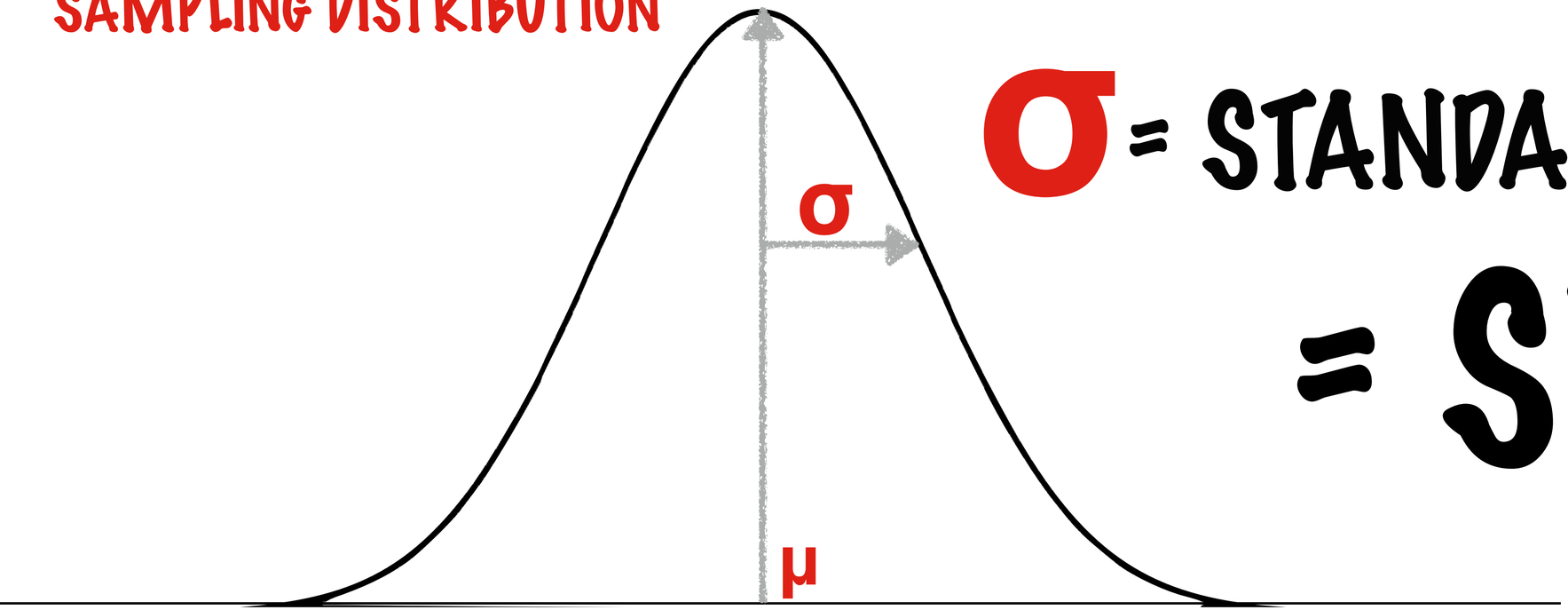
σ = STANDARD ERROR

= SD/SQRT(N)

IT DEPENDS ON THE SD AND THE SIZE OF THE SAMPLE

IF WE ARE INTERESTED IN SAMPLE %

SAMPLING DISTRIBUTION



σ = STANDARD ERROR

= SD

IT DEPENDS ON THE SD AND THE SIZE OF THE SAMPLE

SAMPLE STATISTICS

SAMPLE MEAN / SAMPLE PERCENTAGE

THE STANDARD DEVIATION OF THE SAMPLE

STANDARD ERROR

ONCE YOU HAVE THESE 3 NUMBERS,
THEY BECOME THE MEANS FOR
DRAWING INFERENCES ABOUT THE
POPULATIONS

DRAWING INFERENCES ABOUT THE POPULATIONS

MOST INFERENCES FALL UNDER FEW SPECIFIC TYPES

DRAWING INFERENCES ABOUT THE POPULATIONS

THERE IS A STANDARD PROCEDURE INVOLVED FOR EACH TYPE

DRAWING INFERENCES ABOUT THE POPULATIONS

TO DRAW ANY INFERENCE

1. IDENTIFY THE TYPE

2. FOLLOW THE STANDARD PROCEDURE

TYPES OF INFERENCES

1. IDENTIFYING THE POPULATION MEAN

“Airline pilots are on average 5’5” tall +/- 1” ”

TYPES OF INFERENCES

2. IDENTIFYING THE POPULATION %

“ 40%
+/- 1%
of youngsters in cities like Sci-fi ”

TYPES OF INFERENCES

3. VERIFYING WHETHER THE POPULATION MEAN IS EQUAL TO A CERTAIN VALUE

“ The average life expectancy of college graduates is 70 years ”

TYPES OF INFERENCES

4. VERIFYING WHETHER THE POPULATION % IS EQUAL TO A CERTAIN VALUE

“ 30% of people who took the drug had a side effect”

TYPES OF INFERENCES

5. VERIFYING WHETHER 2 POPULATION MEANS ARE DIFFERENT

“ Indians are on average shorter than Americans”

TYPES OF INFERENCES

6. VERIFYING WHETHER 2 POPULATION % ARE DIFFERENT

“ Only 10% of people who don’t take the drug get better, but 80% of people who take the drug get better”

TYPES OF INFERENCES

1. IDENTIFYING THE POPULATION MEAN
2. IDENTIFYINGTHE POPULATION %
3. VERIFYING WHETHER THE POPULATION MEAN IS EQUAL TO A CERTAIN VALUE
4. VERIFYING WHETHER THE POPULATION % IS EQUAL TO A CERTAIN VALUE
5. VERIFYING WHETHER 2 POPULATION MEANS ARE DIFFERENT
6. VERIFYING WHETHER 2 POPULATION % ARE DIFFERENT

TYPES OF INFERENCES

1. IDENTIFYING THE POPULATION MEAN
2. IDENTIFYINGTHE POPULATION %
3. VERIFYING WHETHER THE POPULATION MEAN IS EQUAL TO A CERTAIN VALUE
4. VERIFYING WHETHER THE POPULATION % IS EQUAL TO A CERTAIN VALUE
5. VERIFYING WHETHER 2 POPULATION MEANS ARE DIFFERENT
6. VERIFYING WHETHER 2 POPULATION % ARE DIFFERENT

LET'S TAKE SOME CASE STUDIES AND GO THROUGH THE STANDARD PROCEDURE FOR EACH OF THESE