TYPES OF INFERENCES

~~1. IDENTIFYING THE POPULATION MEAN~~

~~2. IDENTIFYING THE 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

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

CASE STUDY: A MEDICAL STUDY

MOST ANALYSIS OR RESEARCH STARTS WITH A HYPOTHESIS

MOST ANALYSIS OR RESEARCH STARTS WITH A HYPOTHESIS

"The average life expectancy of an Indian college graduate is 70 yrs"

"The use of Drug A reduces the risk of heart attack"

YOU NEED TO TEST THIS
HYPOTHESIS
TO KNOW HOW ACCURATE IT IS

LET'S TAKE THIS HYPOTHESIS

"The average life expectancy of an Indian college graduate is 70 yrs"

HOW WILL WE GO ABOUT TESTING IT?

STEP : 1

COLLECT A SAMPLE OF INDIAN COLLEGE GRADUATES

STEP : 2

CALCULATE SAMPLE STATISTICS

MEAN LIFE EXPECTANCY OF THAT SAMPLE = 65

SD OF LIFE EXPECTANCY OF THAT SAMPLE = 10

PEOPLE IN THAT SAMPLE = 100

STEP : 2

CALCULATE SAMPLE STATISTICS

MEAN = 65

SD = 10

PEOPLE = 100

$$\text{STANDARD ERROR} = \frac{\text{SD OF SAMPLE}}{\text{SQRT(SIZE OF SAMPLE)}}$$

$$= \frac{10}{\text{SQRT(100)}}$$

$$= 1$$

SD OF THE DISTRIBUTION YOU GET IF YOU COMPUTED THE MEAN FOR A LARGE NUMBER OF SAMPLES

STEP : 2

CALCULATE SAMPLE STATISTICS

MEAN = 65

SD = 10

PEOPLE = 100

STANDARD ERROR = 1

STEP : 3

PERFORM A TEST OF SIGNIFICANCE

STEP : 3 PERFORM A TEST OF SIGNIFICANCE

"The average life expectancy of an Indian college graduate is 70 yrs"

SAMPLE STATISTICS

MEAN = 65

SD = 10

PEOPLE = 100

STANDARD ERROR = 1

SAMPLE MEAN = 65

WHICH IS NOT THAT FAR FROM 70

SO, IS THE POPULATION MEAN 70 OR <>70 ?

STEP : 3 PERFORM A TEST OF SIGNIFICANCE

"The average life expectancy of an Indian college graduate is 70 yrs"

WE NEED TO UNDERSTAND IF THE DIFFERENCE OF 65 VS 70 IS DUE TO CHANCE, OR IF THE POPULATION MEAN <> 70 TOO

STEP : 3 PERFORM A TEST OF SIGNIFICANCE

"The average life expectancy of an Indian college graduate is 70 yrs"

WE NEED TO UNDERSTAND **IF THE DIFFERENCE OF 65 VS 70 IS DUE TO CHANCE**, OR IF THE POPULATION MEAN <> 70 TOO

NULL HYPOTHESIS

STEP : 3 PERFORM A TEST OF SIGNIFICANCE

"The average life expectancy of an Indian college graduate is 70 yrs"

WE NEED TO UNDERSTAND IF THE DIFFERENCE OF 65 VS 70 IS DUE TO CHANCE, OR IF THE POPULATION MEAN <> 70 TOO

ALTERNATIVE HYPOTHESIS

STEP : 3 PERFORM A TEST OF SIGNIFICANCE

NULL HYPOTHESIS

ALL VARIATIONS OBSERVED ARE DUE TO CHANCE I.E. A FLUKE

ALTERNATIVE HYPOTHESIS

THE VARIATIONS OBSERVED CANNOT JUST BE EXPLAINED BY CHANCE

STEP : 3 PERFORM A TEST OF SIGNIFICANCE

NULL HYPOTHESIS VS ALTERNATIVE HYPOTHESIS

A TEST OF SIGNIFICANCE WILL TELL YOU WHICH OF THESE IS BETTER

A TEST OF SIGNIFICANCE

NULL HYPOTHESIS VS ALTERNATIVE HYPOTHESIS

THIS INVOLVES

A) COMPUTING A TEST STATISTIC

SOME VARIABLE WHOSE PROBABILITY DISTRIBUTION IS KNOWN

A TEST OF SIGNIFICANCE

NULL HYPOTHESIS VS ALTERNATIVE HYPOTHESIS

THIS INVOLVES

A) COMPUTING A TEST STATISTIC

B) COMPUTE THE PROBABILITY IF THE NULL HYPOTHESIS IS TRUE

A TEST OF SIGNIFICANCE

NULL HYPOTHESIS VS ALTERNATIVE HYPOTHESIS

THIS INVOLVES

A) COMPUTING A TEST STATISTIC

B) COMPUTE THE PROBABILITY IF THE NULL HYPOTHESIS IS TRUE

C) IF THE PROBABILITY IS TOO LOW, REJECT THE NULL HYPOTHESIS, ELSE ACCEPT IT

STEP : 3 PERFORM A TEST OF SIGNIFICANCE

IN OUR EXAMPLE

"The average life expectancy of an Indian college graduate is 70 yrs"

NULL HYPOTHESIS

Population mean = 70

THE DIFFERENCE BETWEEN 65 AND 70 IS DUE TO CHANCE

ALTERNATIVE HYPOTHESIS

Population mean !=70

THE DIFFERENCE BETWEEN 65 AND 70 IS REAL

STEP : 3 PERFORM A TEST OF SIGNIFICANCE

STEP : 3A

COMPUTE A TEST STATISTIC

WE'LL USE THE Z-STATISTIC

Z REPRESENTS THE NORMALIZED DISTANCE BETWEEN THE SAMPLE MEAN AND THE POPULATION MEAN

WE'LL USE THE Z-STATISTIC

Z REPRESENTS THE NORMALIZED DISTANCE BETWEEN THE SAMPLE MEAN AND THE POPULATION MEAN

WE'LL USE THE Z-STATISTIC

Z REPRESENTS THE NORMALIZED

SAMPLE MEAN - POPULATION MEAN

WE'LL USE THE Z-STATISTIC

Z REPRESENTS THE

NORMALIZED

$$\frac{\text{SAMPLE MEAN} - \text{POPULATION MEAN}}{\text{STANDARD ERROR}}$$

WE'LL USE THE Z-STATISTIC

$$\frac{\text{SAMPLE MEAN - POPULATION MEAN}}{\text{STANDARD ERROR}}$$

$$z = \frac{X - \mu}{\sigma}$$

OK.. BUT HOW WILL THIS HELP?

$$z = \frac{X - \mu}{\sigma}$$

1. THIS VARIABLE REPRESENTS THE DIFFERENCE WHICH WE WANT TO UNDERSTAND

$$z = \frac{X - \mu}{\sigma}$$

WE ALREADY KNOW THE
SAMPLE MEAN IS
NORMALLY DISTRIBUTED

$$z = \frac{X - \mu}{\sigma}$$

2. THE Z-STATISTIC PROBABILITY DISTRIBUTION IS ALSO NORMAL (MEAN =0, SD 1)

$$z = \frac{X - \mu}{\sigma}$$

LET'S COMPUTE THE Z-STATISTIC FOR
THE NULL HYPOTHESIS

Population mean = 70

$$z = \frac{\overset{65}{X} - \overset{70}{\mu}}{\sigma \; 1}$$

$$= -5$$

THE NULL HYPOTHESIS

$\mu = 70$

SAMPLE MEAN = 65

STANDARD ERROR = 1

WE'LL CALCULATE THE PROBABILITY OF THIS VALUE

FINE, WHAT NEXT?

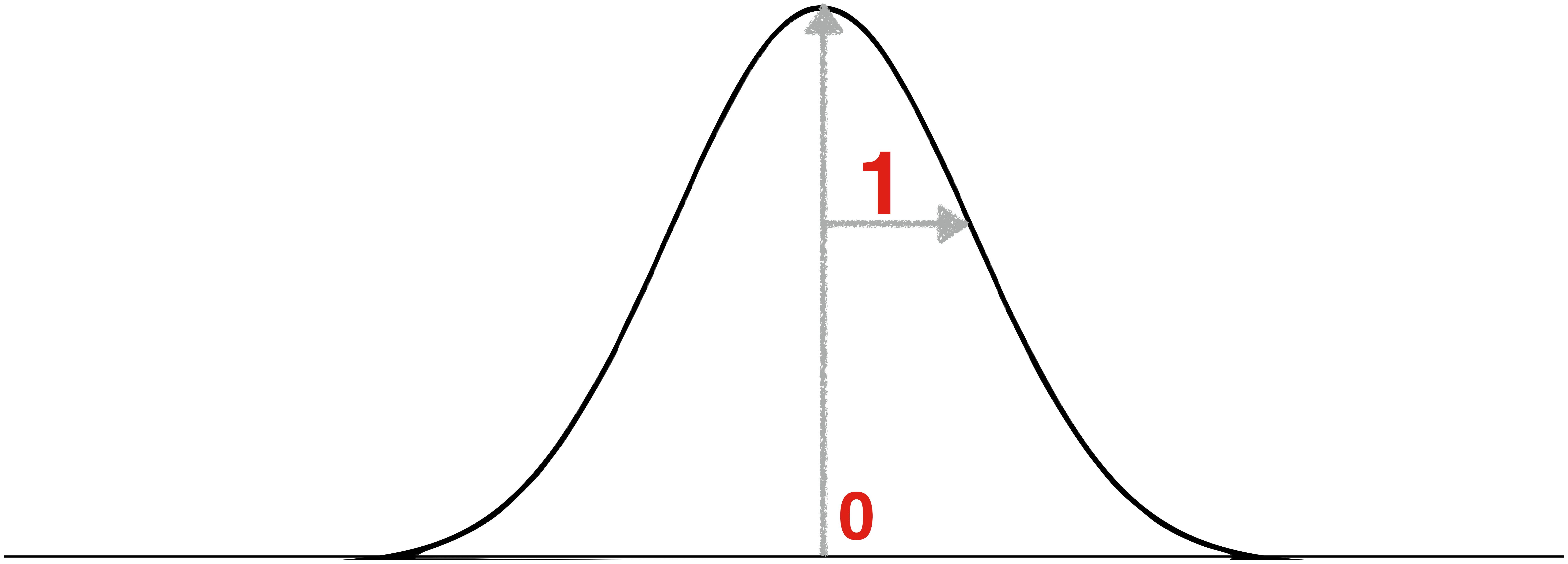
STEP : 3 PERFORM A TEST OF SIGNIFICANCE

STEP : 3B

COMPUTE THE PROBABILITY IF
THE NULL HYPOTHESIS IS TRUE

Z-STATISTIC FOR THE NULL HYPOTHESIS = -5

THE PROBABILITY
DISTRIBUTION OF Z IS
NORMAL (MEAN =0, SD 1)



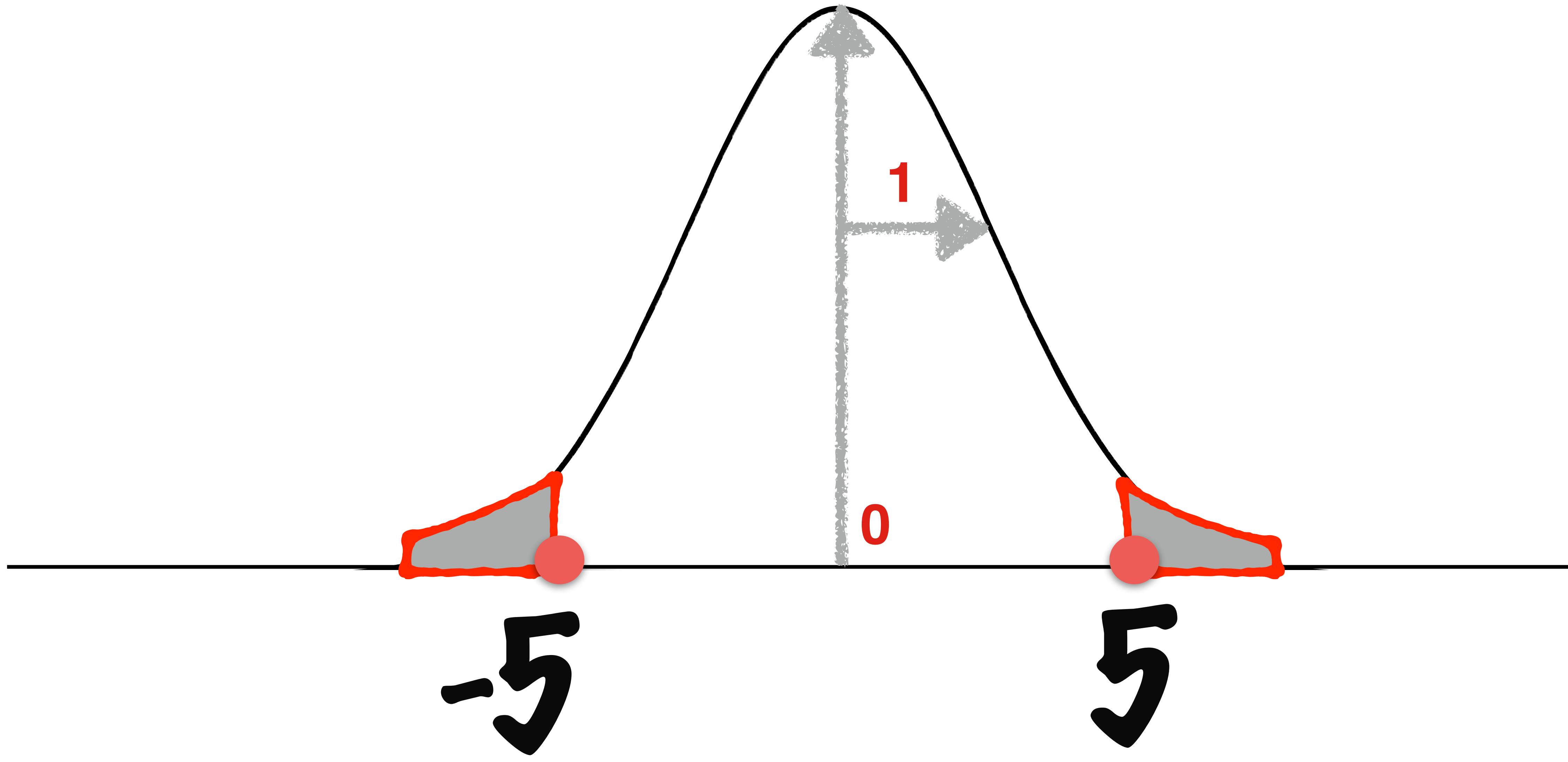
Z-STATISTIC FOR THE NULL HYPOTHESIS = -5

P-VALUE

P(|Z| > 5)

= P (THE DIFFERENCE >5)

= AREA UNDER THE CURVE ABOVE 5 AND BELOW -5



Z-STATISTIC FOR THE NULL HYPOTHESIS = -5

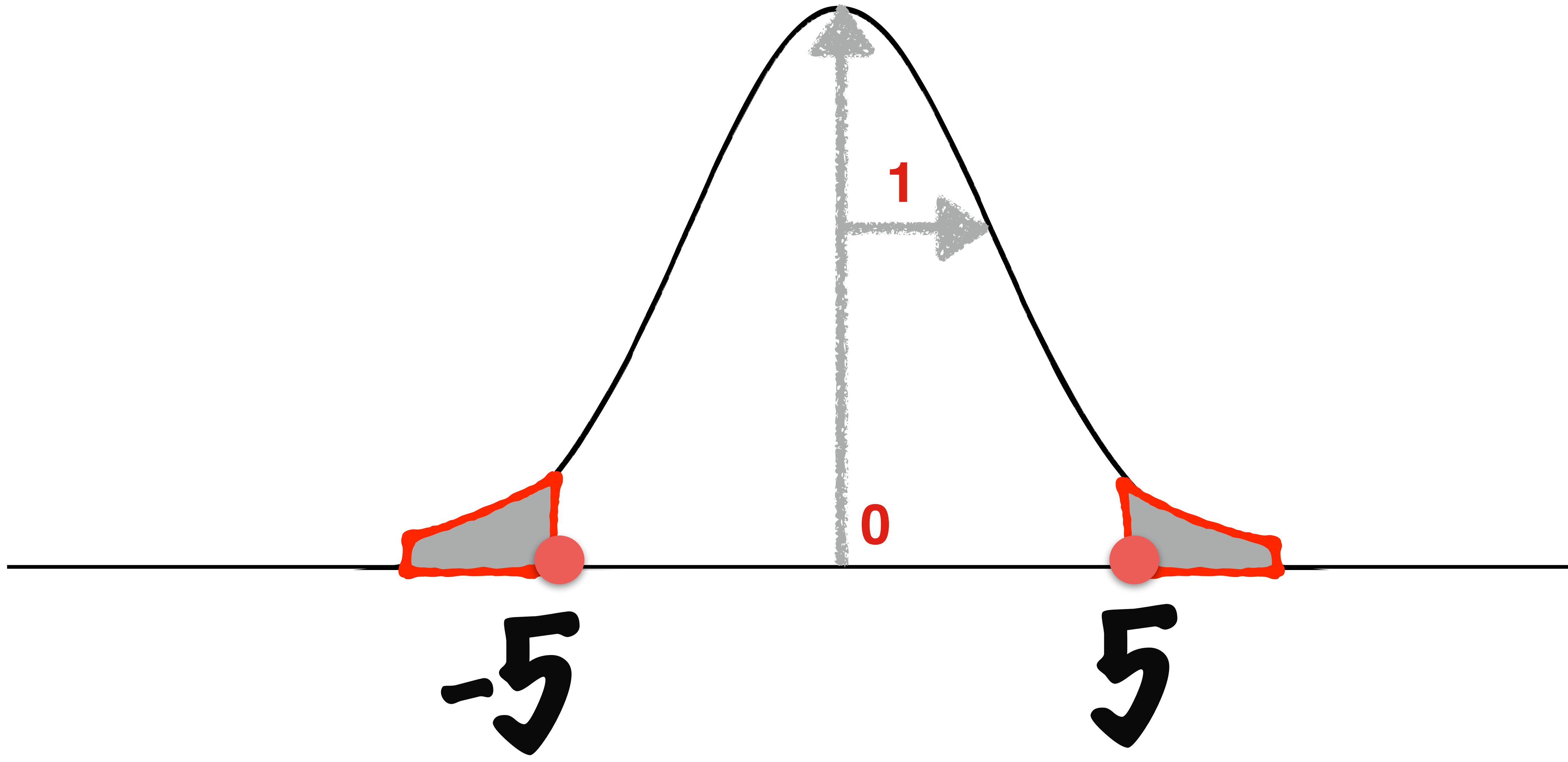
P-VALUE

$P(|Z| > 5)$

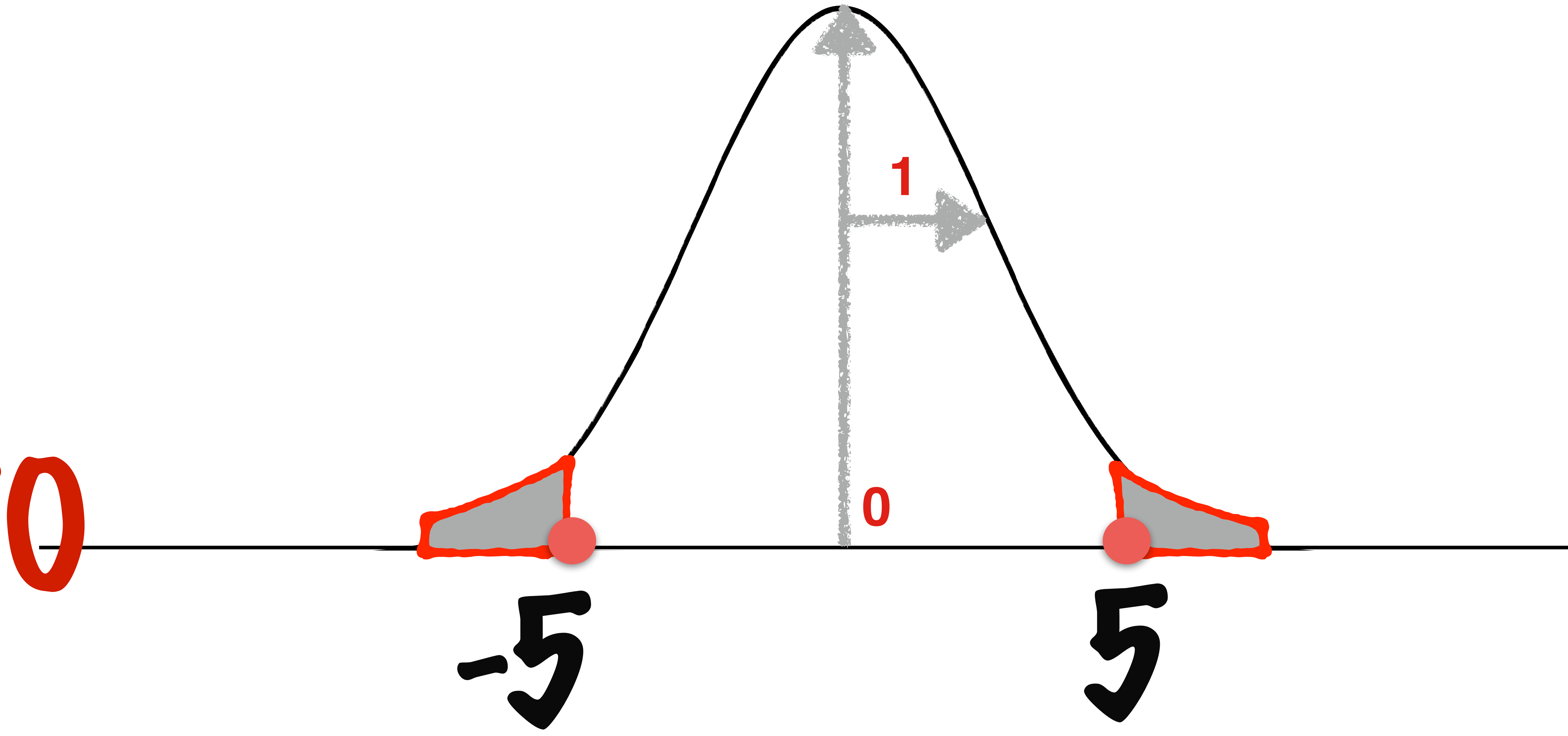
= P (THE DIFFERENCE >5)

= AREA UNDER THE CURVE ABOVE 5 AND BELOW -5

= ?



$P(|Z| > 5)$



THERE ARE MANY WAYS TO COMPUTE THIS VALUE

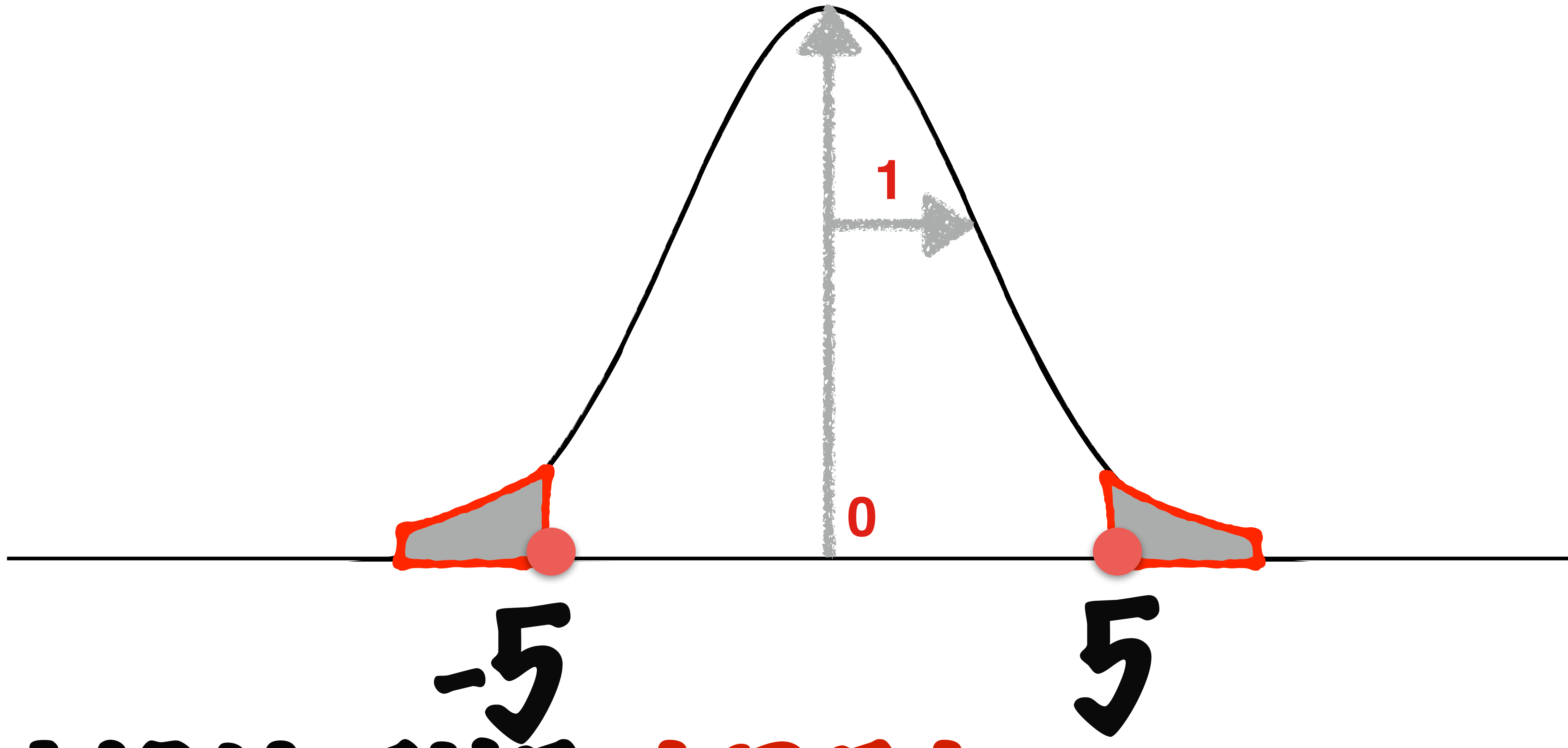
1. LOOK UP A TABLE OF NORMAL DISTRIBUTION VALUES

2. USE A FUNCTION IN R OR EXCEL

WE'LL USE THE SECOND METHOD HERE

P(|Z| > 5)

FUNCTION IN R



PNORM() IN R WILL TELL YOU THE AREA UNDER THE CURVE FROM -INF TO Z (CUMULATIVE DISTRIBUTION FUNCTION)

```
> 1-(pnorm(5)-pnorm(-5))
[1] 5.733031e-07
```

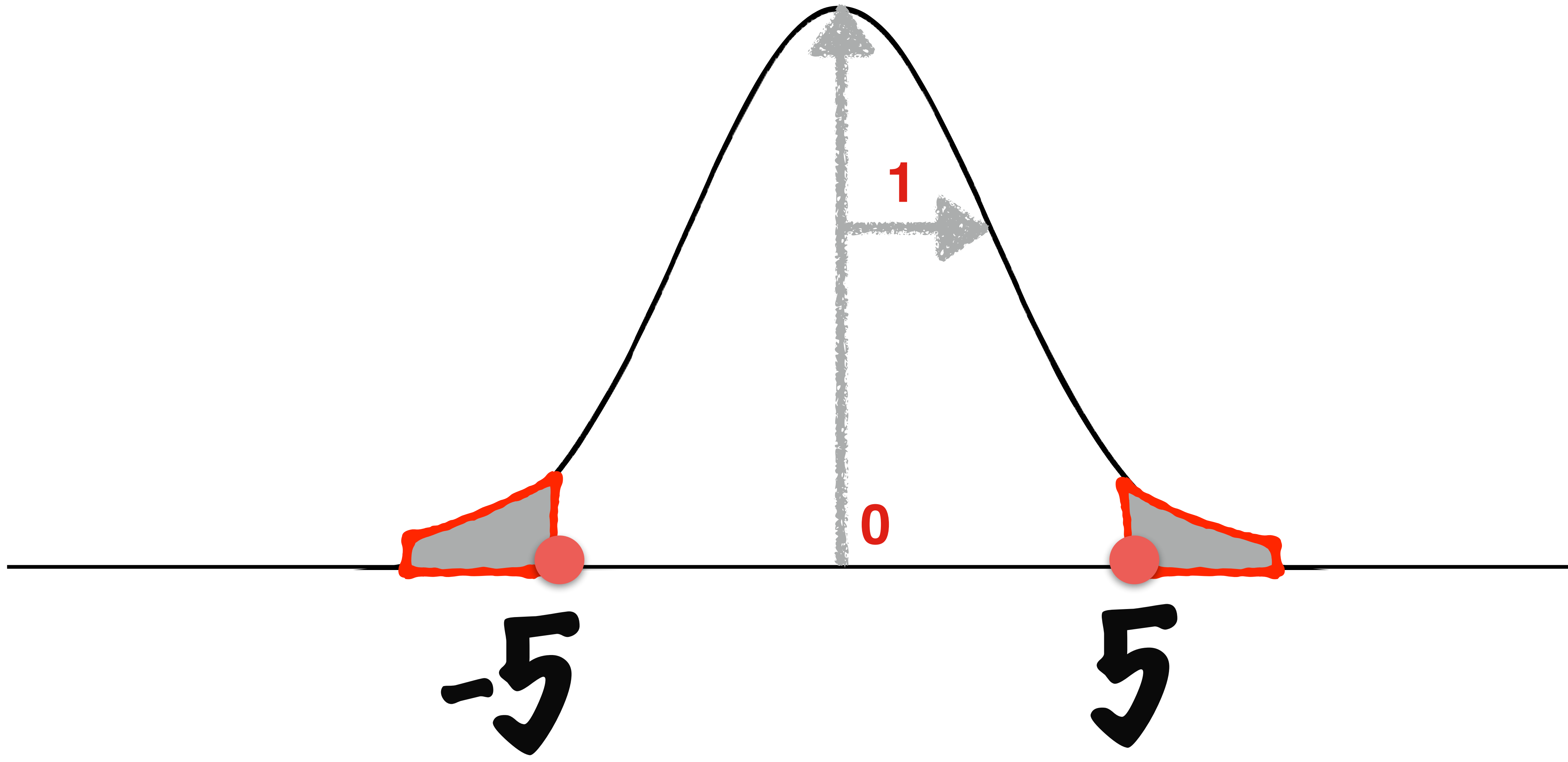
Z-STATISTIC FOR THE NULL HYPOTHESIS = -5

P-VALUE

P(|Z| > 5)

= P (THE DIFFERENCE >5)

= AREA UNDER THE CURVE ABOVE 5 AND BELOW -5



= 0.0000005

HOW LOW IS LOW? IT DEPENDS

0.1 90% CONFIDENCE THAT NULL HYPOTHESIS IS FALSE OR

90% CONFIDENCE THAT THE DIFFERENCE OBSERVED IS STATISTICALLY SIGNIFICANT

0.05 95% CONFIDENCE THAT NULL HYPOTHESIS IS FALSE

95% CONFIDENCE THAT THE DIFFERENCE OBSERVED IS STATISTICALLY SIGNIFICANT

0.01 99% CONFIDENCE THAT NULL HYPOTHESIS IS FALSE

99% CONFIDENCE THAT THE DIFFERENCE OBSERVED IS STATISTICALLY SIGNIFICANT

STEP : 3 PERFORM A TEST OF SIGNIFICANCE

IN OUR EXAMPLE

"The average life expectancy of an Indian college graduate is 70 yrs"

NULL HYPOTHESIS

Population mean = 70

THE DIFFERENCE BETWEEN 65 AND 70 IS DUE TO CHANCE

ALTERNATIVE HYPOTHESIS

Population mean !=70

THE DIFFERENCE BETWEEN 65 AND 70 IS REAL

HOW LOW IS LOW? IT DEPENDS

0.1 90% CONFIDENCE THAT NULL HYPOTHESIS IS FALSE OR

90% CONFIDENCE THAT THE DIFFERENCE OBSERVED IS STATISTICALLY SIGNIFICANT

0.05 95% CONFIDENCE THAT NULL HYPOTHESIS IS FALSE

95% CONFIDENCE THAT THE DIFFERENCE OBSERVED IS STATISTICALLY SIGNIFICANT

0.01 99% CONFIDENCE THAT NULL HYPOTHESIS IS FALSE

99% CONFIDENCE THAT THE DIFFERENCE OBSERVED IS STATISTICALLY SIGNIFICANT

TYPES OF INFERENCES

~~1. IDENTIFYING THE POPULATION MEAN~~

~~2. IDENTIFYING THE 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