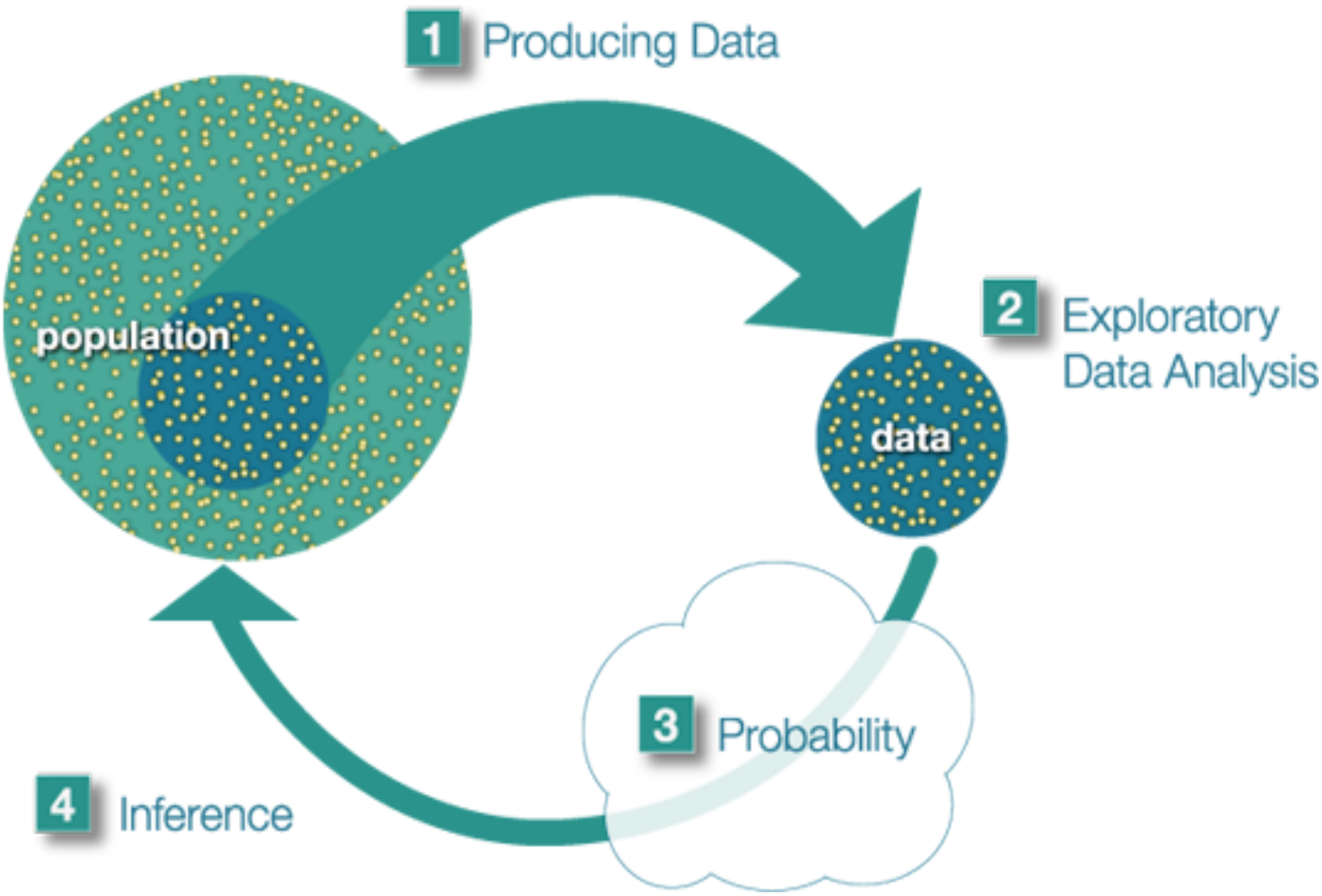
QUANTITATIVE METHODS


1 Producing Data
population
2 Exploratory Data Analysis
data
3 Probability
4 Inference

Statistical Significance

- If a difference in a sample is big enough that we feel confident it didn't happen due to chance, we find the difference ***statistically significant***.

Statistical Significance

- When something is statistically significant, we believe there is a real effect in the population…
- But we accept that this conclusion might be erroneous.

The Performance-Based Scholarship Demonstration

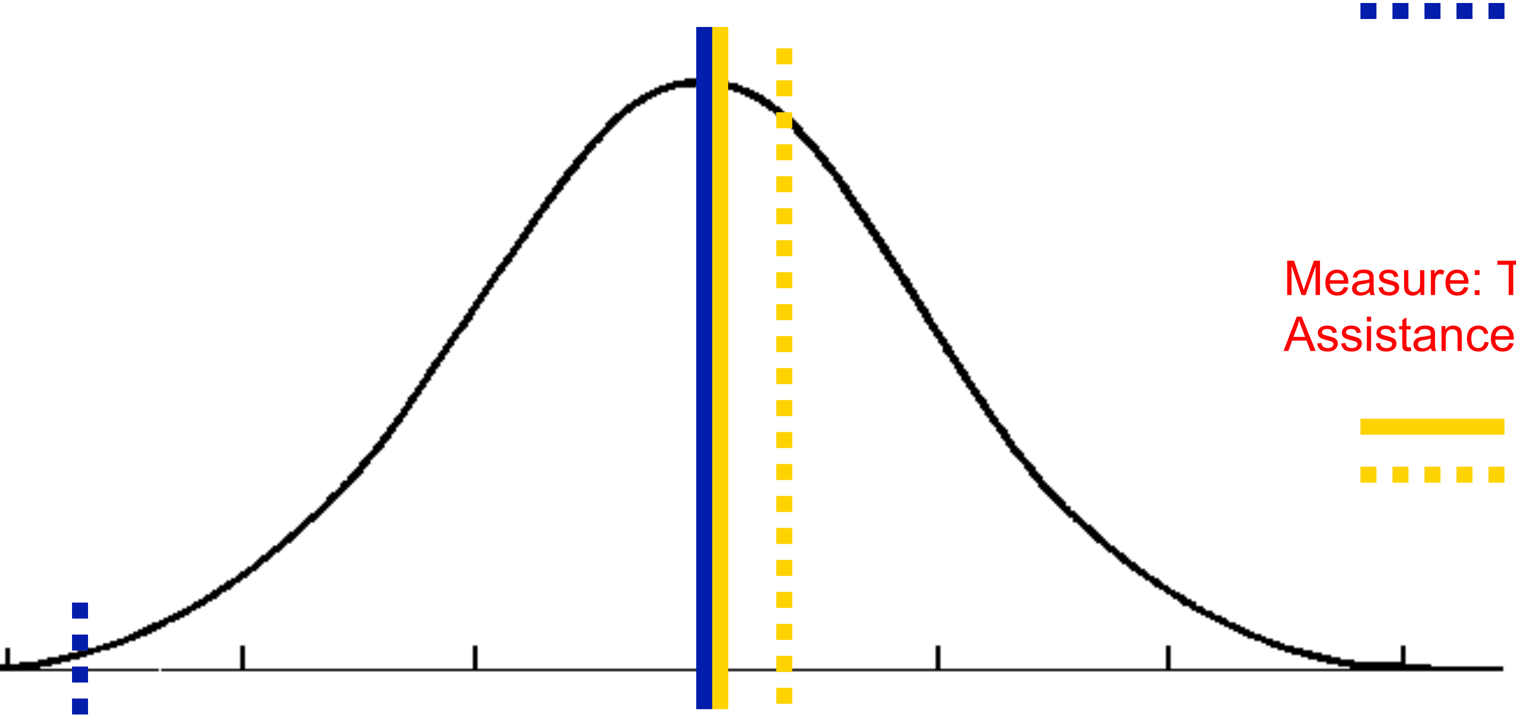
Table 6

Financial Assistance Among Sample Members, by Research Group

University of New Mexico

Outcome	Program Group	Control Group	Difference	Standard Error
Total financial assistance received ($)	5,847	5,250	597 ***	131.1
Pell Grant	1,899	1,965	-66	46.8

Total financial assistance received ($)	5,847	5,250	597 ***	131.1
Pell Grant	1,899	1,965	-66	46.8



Measure: Total Financial Assistance Received

Program Group

Control Group

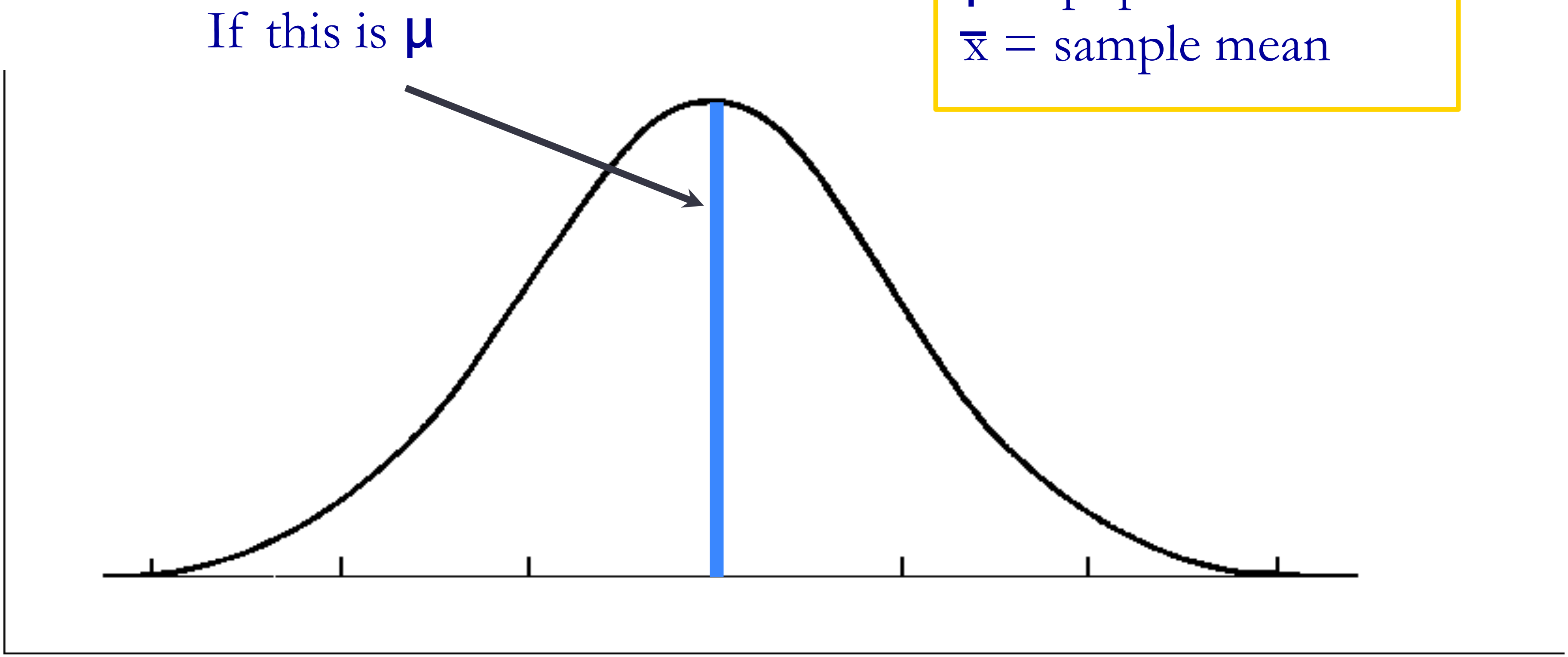
Measure: Total Pell Grant Assistance Received

Program Group

Control Group

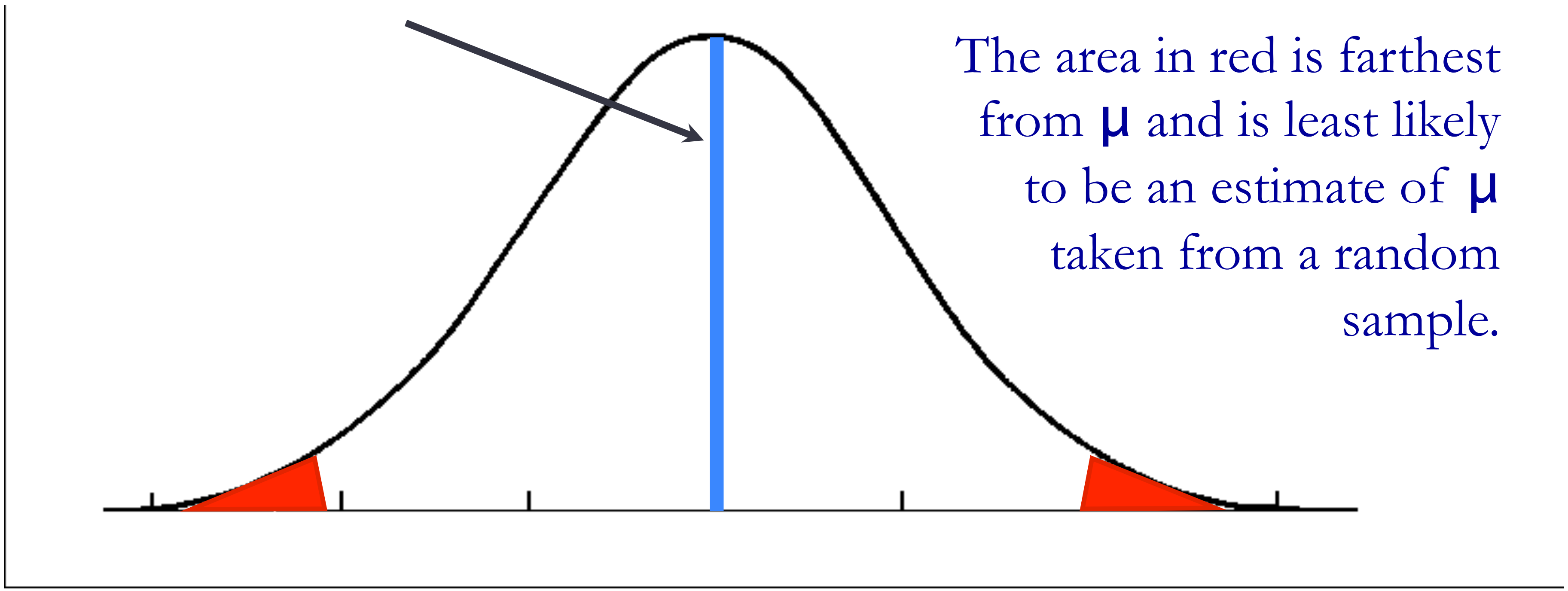
Finding Statistical Significance

μ = population mean
$\bar{x}$ = sample mean



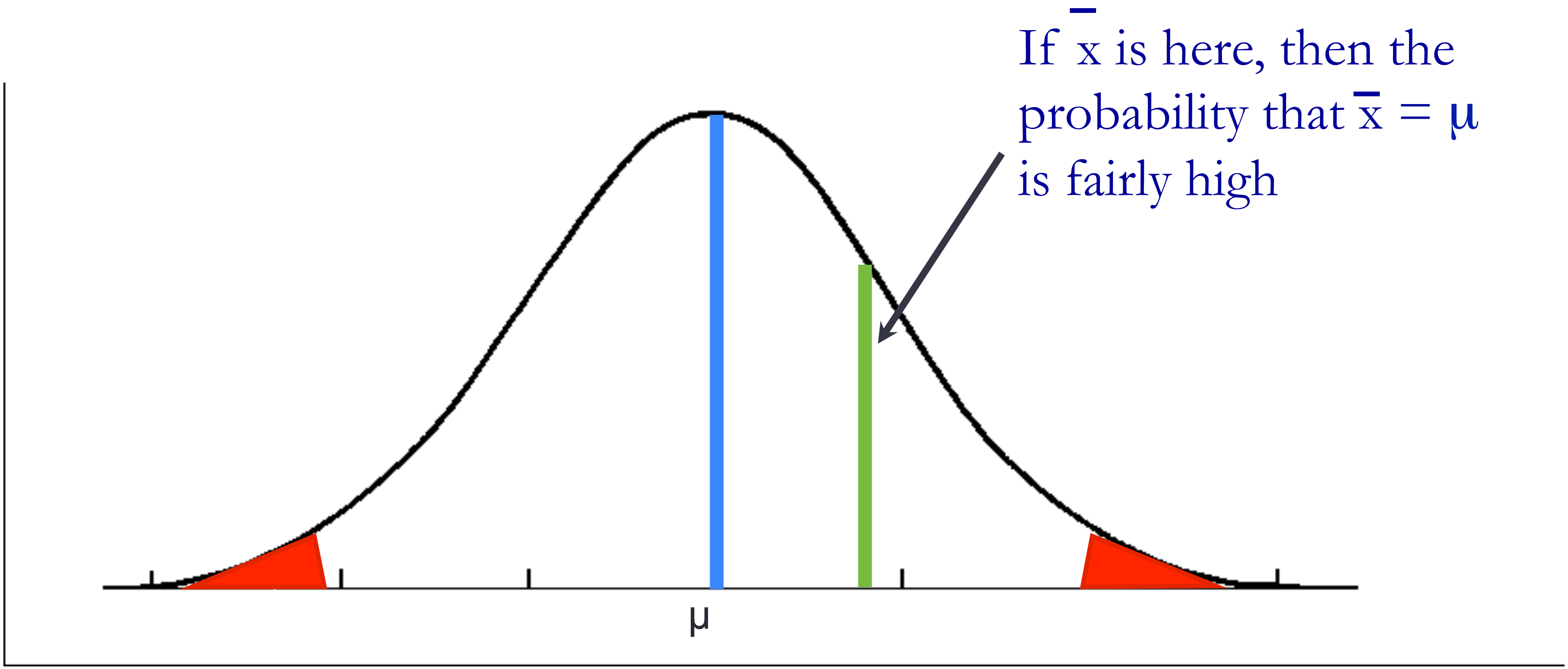
Finding Statistical Significance

If this is μ

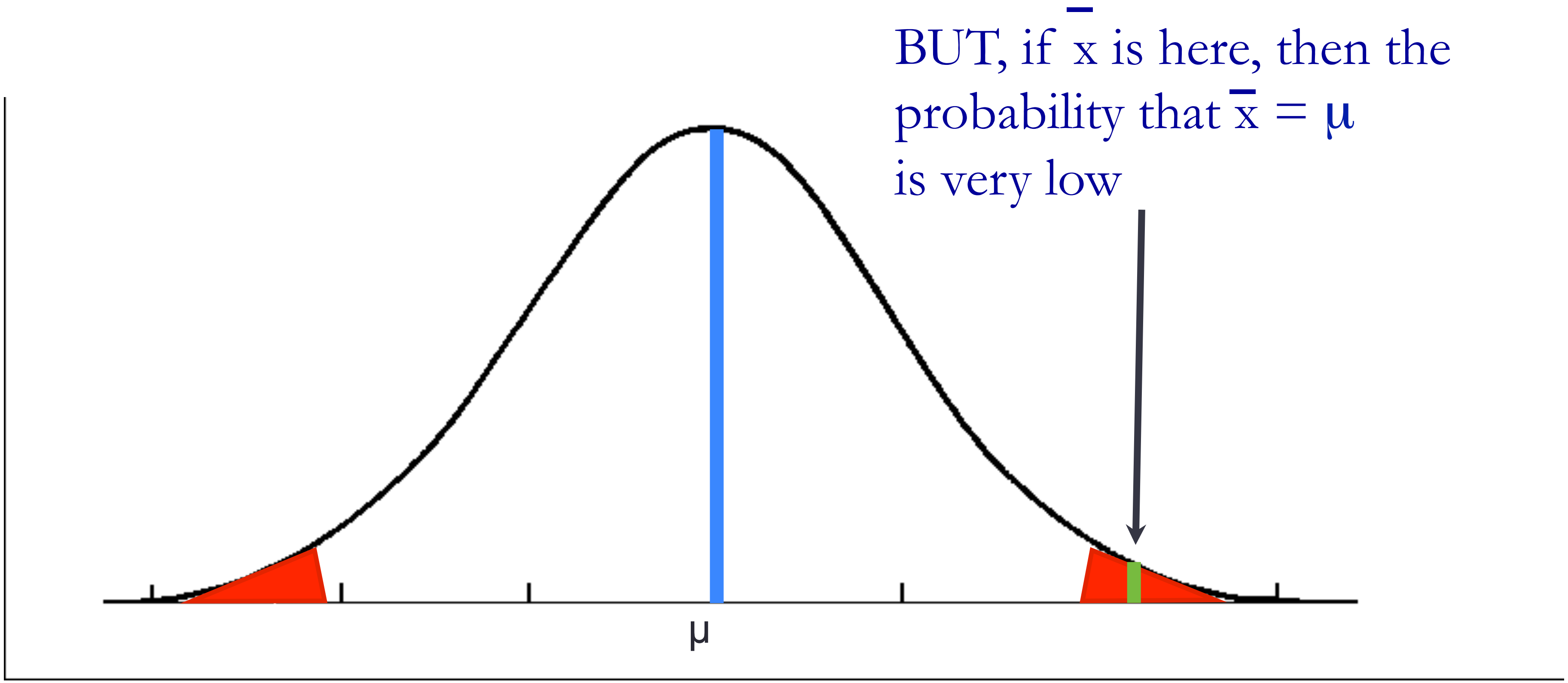


The area in red is farthest from μ and is least likely to be an estimate of μ taken from a random sample.

Finding Statistical Significance



Finding Statistical Significance



Probability & Statistics

- In statistics, we estimate the probability that the data would be at least as extreme as those observed if there were no effect in the population.

Alpha

- Alpha is highest probability we are willing to accept that a difference is due to chance (and not due to a real effect in the population)
- In the social sciences, the de facto standard Alpha is = 0.05
  - 5% probability that a difference occurs by chance

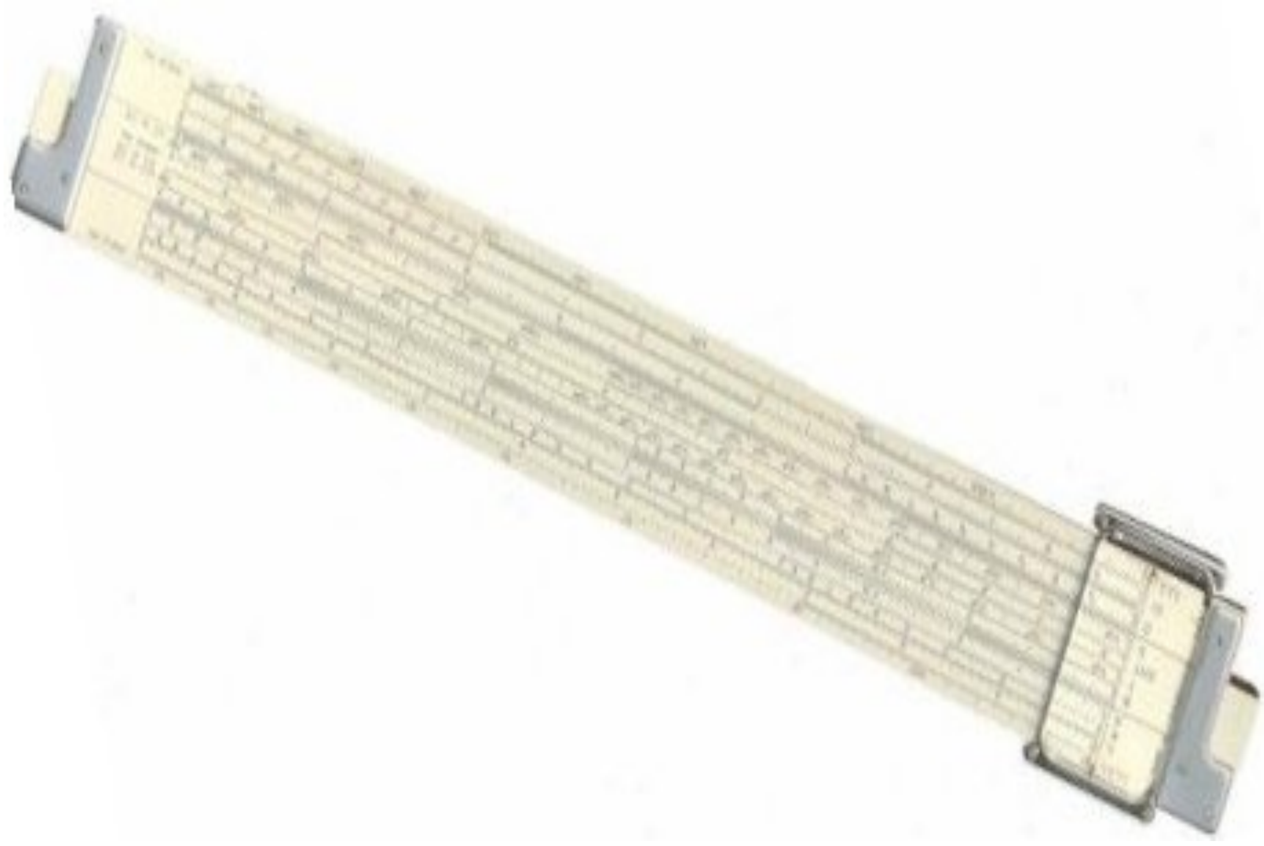
Alpha

"We should be very conservative and only believe a result is genuine when we are 95% confident in our result."

Ronald Fisher

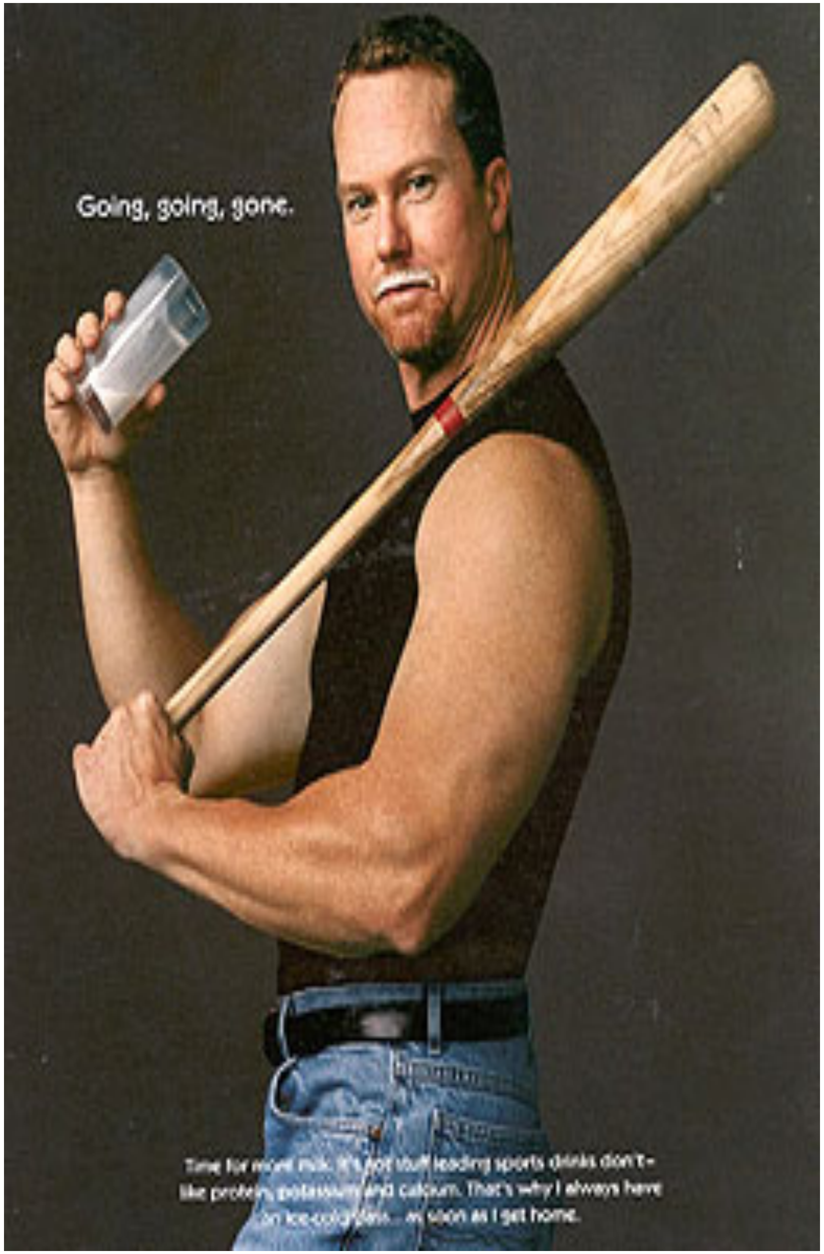
Alpha

- How did Fisher decide on Alpha of 0.05?



Choosing Alpha

- The choice of alpha is extremely important, because it determines the threshold for concluding there is an effect in the population.


Going, going, gone.

Choosing Alpha

- Type I error (α)– finding something statistically significant when there is actually no effect in the population (false positive)
- Type II error (β) – not finding something statistically significant when there really is an effect in the population (false negative)

Choosing Alpha

- Type I and Type II errors are intertwined. You can avoid Type I error by setting a lower alpha, but that increases the chance of Type II error. A higher alpha reduces chance of Type II error, but increases risk of Type I error.

Result of statistical test	Truth	
	Real effect	No real effect
Statistically significant	**True positive**	**Type I Error / False positive**
Not statistically significant	**Type II Error / False negative**	**True negative**

Why are there differences?

Table 5.1 (continued)

Outcome	Transitional Jobs Group	Job Search Group	Difference (Impact)	P-Value[a]
Total earnings[c]				
Quarter 1 (quarter of random assignment)	789	307	482 ***	0.000
Quarter 2	1,467	844	624 ***	0.000
Quarter 3	1,101	902	199 **	0.026
Quarter 4	979	865	113	0.260
Year 1 (Q1 - Q4)	4,336	2,917	1,419 ***	0.000
Sample size (total =1,774)	893	881		

Recap

- In statistics, we estimate the probability that the data would be at least as extreme as those observed if there were no effect in the population.
- If that probability is extremely low, we believe there is an effect in the population (statistical significance).
- Sometimes we're wrong.

Gallup Poll 9/20-9/26/12

- "Suppose the presidential election were held today. If Barack Obama were the Democratic Party's candidate and Mitt Romney were the Republican Party's candidate, who would you vote for Barack Obama, the Democrat or Mitt Romney, the Republican?"
  - 50% Obama
  - 44% Romney

Margin of Error ±2

Gallup Poll 9/20-9/26/12

- "Suppose the presidential election were held today. If Barack Obama were the Democratic Party's candidate and Mitt Romney were the Republican Party's candidate, who would you vote for Barack Obama, the Democrat or Mitt Romney, the Republican?"
  - 50% Obama
  - 44% Romney

Margin of Error ±2

N = 3,050

HYPOTHESIS TESTING

Observations

- Until now, any comparisons we have made were merely observations
- To infer meaning from an observation, a *hypothesis test* is needed

Observations

- An observation:
- In a large General Social Survey sample, men earned an average of $45,273 and women an average of $33,303.

Observations

- It sure *seems* like men make a different amount than women in the U.S. But how do we test that?

Theory Comes First

- Based on our observation that ‘men make a different amount of money than women’ you develop a theory:
- *in the U.S., men and women do not have the same average level of income*

Hypothesis

- A hypothesis is an explicit statement of your theory:
- In this case, the research hypothesis is:

$$H_1: \mu_{men} \neq \mu_{women}$$

Hypothesis

- A hypothesis involves a dependent variable and one or more independent variables
- Dependent Variable -- the variable that measures the outcome you are studying.
- Independent Variable -- a variable that you believe will *predict* or *affect* the outcome variable.

Hypothesis

- Suppose I suspect that those who drink 5 cups of coffee or more each day have higher stress levels than those that consume less coffee.
- Independent Variable *predicts* Dependent Variable (outcome)
- Coffee drinking (IV) *predicts* stress level (DV)

Test Yourself

- Decide which are the dependent and the independent:
  - Exam grade, study hours
  - Number of puppies, happiness level
  - Daily food intake, weight
  - Life expectancy, development category

Answers

- Exam grade (DV), study hours (IV)
- Number of puppies (IV), happiness level (DV)
- Daily food intake (IV), weight (DV)
- Life expectancy (DV), development category (IV)

Types of Hypothesis Tests

- Two types of Hypothesis tests (for now):
  - One-sample Hypothesis Test-- testing the difference of a mean or proportion of a sample against the mean or proportion of the population
  - Two-sample Hypothesis Test-- testing the difference of means or proportions for two samples

The Research Process

- ‘Pre’ Steps:
    - Formulate and articulate your *theory* and *hypothesis*, usually based on *observations*.
    - Get to know the data you have available to test a hypothesis. What variables do you have? Which is the dependent and which the independent? How are they measured (nom/ord/IR)? How are they distributed?

The Research Process

- Step 1: Decide which hypothesis test to use. State your assumptions, and verify that your data meet the test requirements:
  - The sample is EPSEM
  - The variables are appropriate level of measurement
  - N is appropriate for the test

Which hypothesis test to use?

- How to know which to use? Well, we haven't yet covered *any*, but, the rest of the semester will be about hypothesis testing. Prior to the midterm, we will only be covering one- and two-sample hypothesis tests (Chapters 8 & 9).

Test Yourself

- For our observation about differences in earnings levels between men and women:
  - How many variables are involved?
  - Is it ‘one-sample’ or ‘two-sample’?
  - Which is the dependent and which is/are the independent?
  - How are they measured?

Answer

- There are two variables: sex and earnings
- It would be a ‘two-sample’ test because men make one sample and women the other sample
- Sex *predicts* earnings
- Sex is independent, earnings is dependent
- Sex is nominal, earnings is interval-ratio

Test Yourself

- Do all variables used in a hypothesis test need to be normally distributed? Why, or why not?

Answer

- *Do all variables used in a hypothesis test need to be normally distributed? Why, or why not?*
- The variables do *not* need to be normally distributed, thanks to the Central Limit Theorem. As long as the sample is EPSEM (Equal Probability of Selection Method) and N is 100 or greater, the CLT applies.

The Research Process

- Step 2: State the null and research hypotheses
- In research, we assume no difference when we approach a problem. With our men/women earnings theory, this ‘null hypothesis (H_0)’ this is communicated as:
- H_0: $\mu_{men} = \mu_{women}$

The Research Process

- Step 2:
- H_0: $\mu_{men} = \mu_{women}$
- H_1: $\mu_{men} \neq \mu_{women}$
- The null hypothesis is that there is no difference in the earnings of men compared to the earnings of women. The research hypothesis is that men and women have different levels of income.

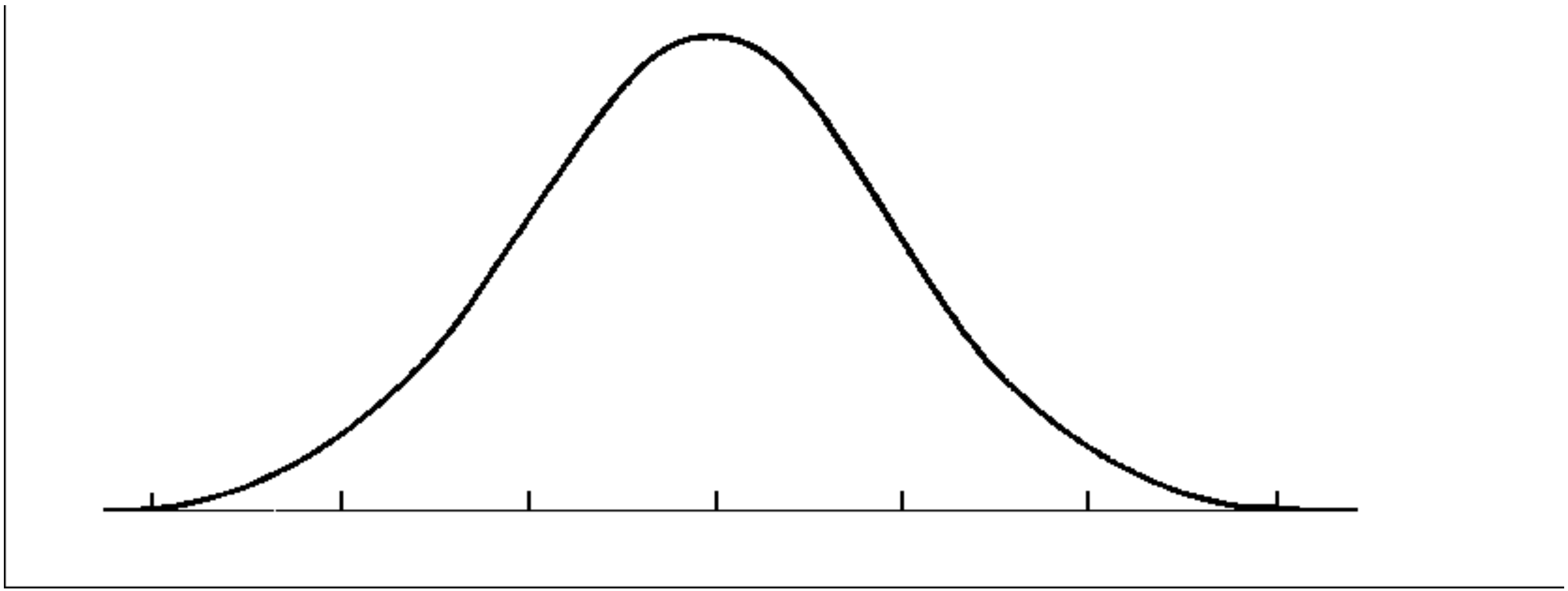
The Research Process

- Later in the process we will test the null hypothesis.
- If we ‘reject the null hypothesis,’ we are concluding that there *is* an effect. This is commonly communicated as *statistically significant*.

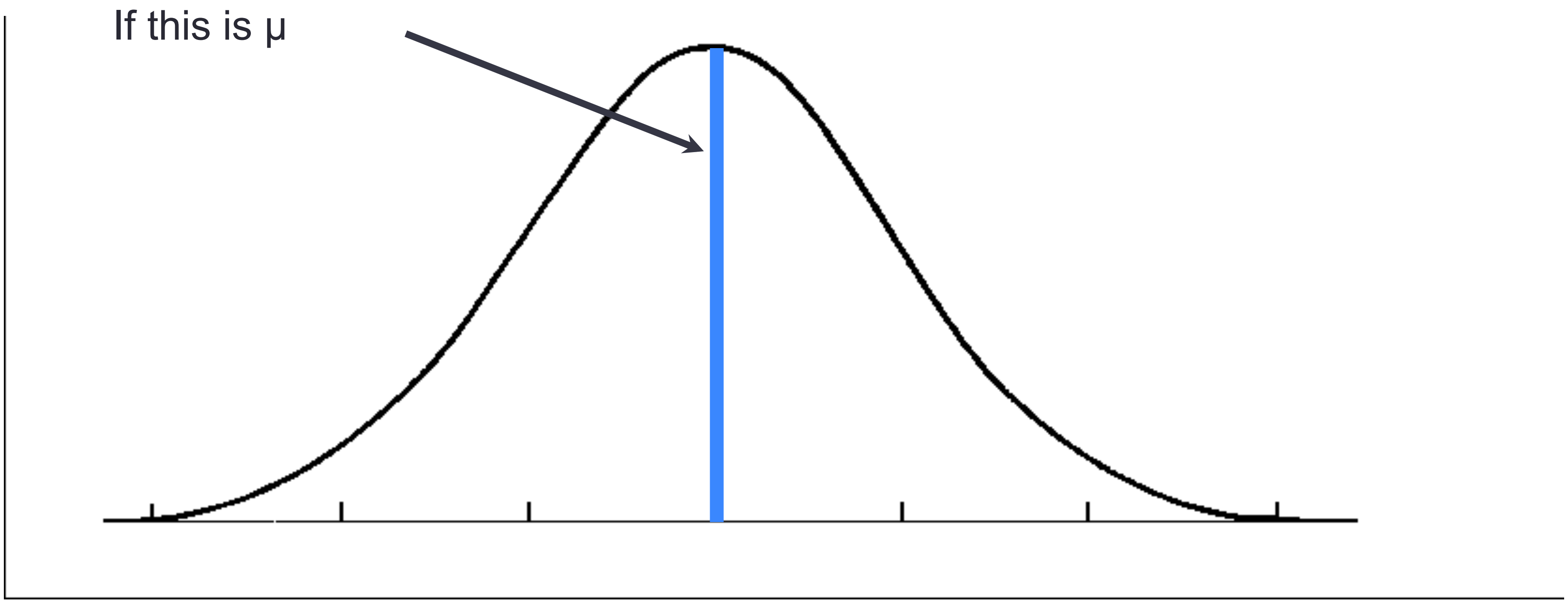
The Research Process

- Step 3: Select alpha and establish the *critical region.*
- The critical region is the part of the distribution (for example Z / normal curve) that is farthest from the center. The area of the critical region is equal to alpha.

Establishing ‘Critical Region’

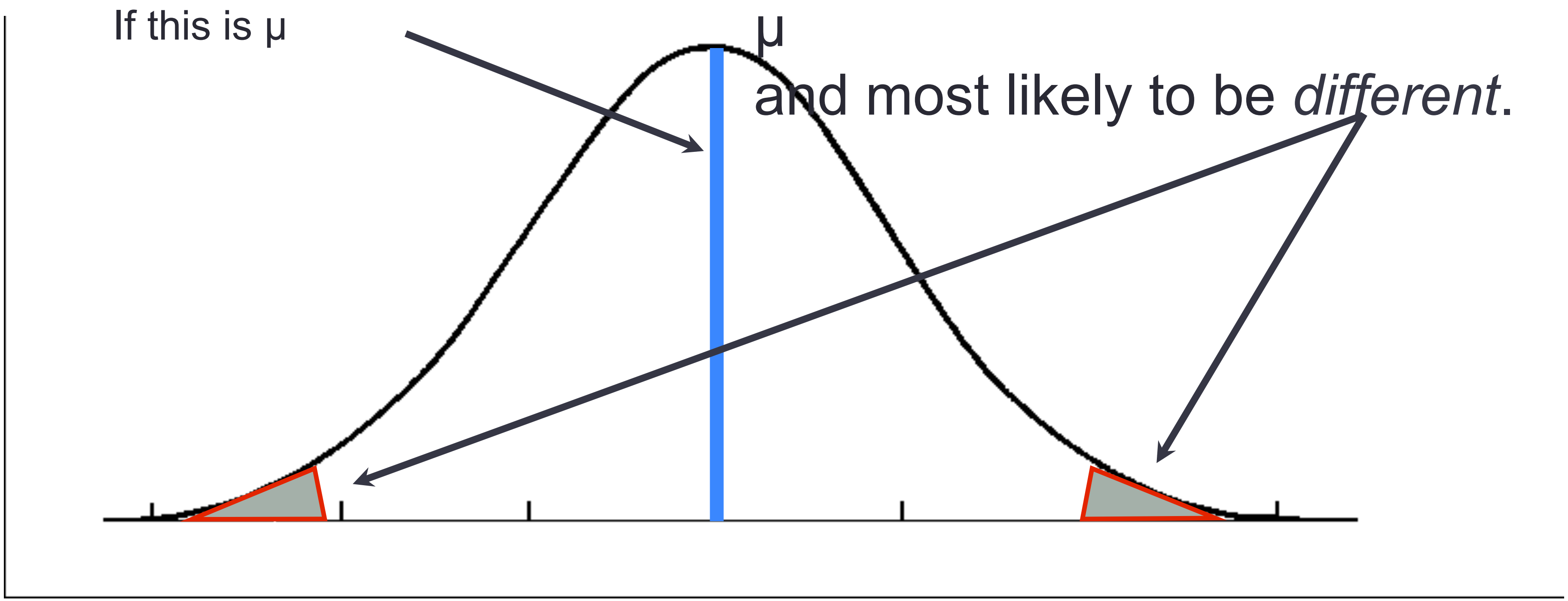


Establishing ‘Critical Region’

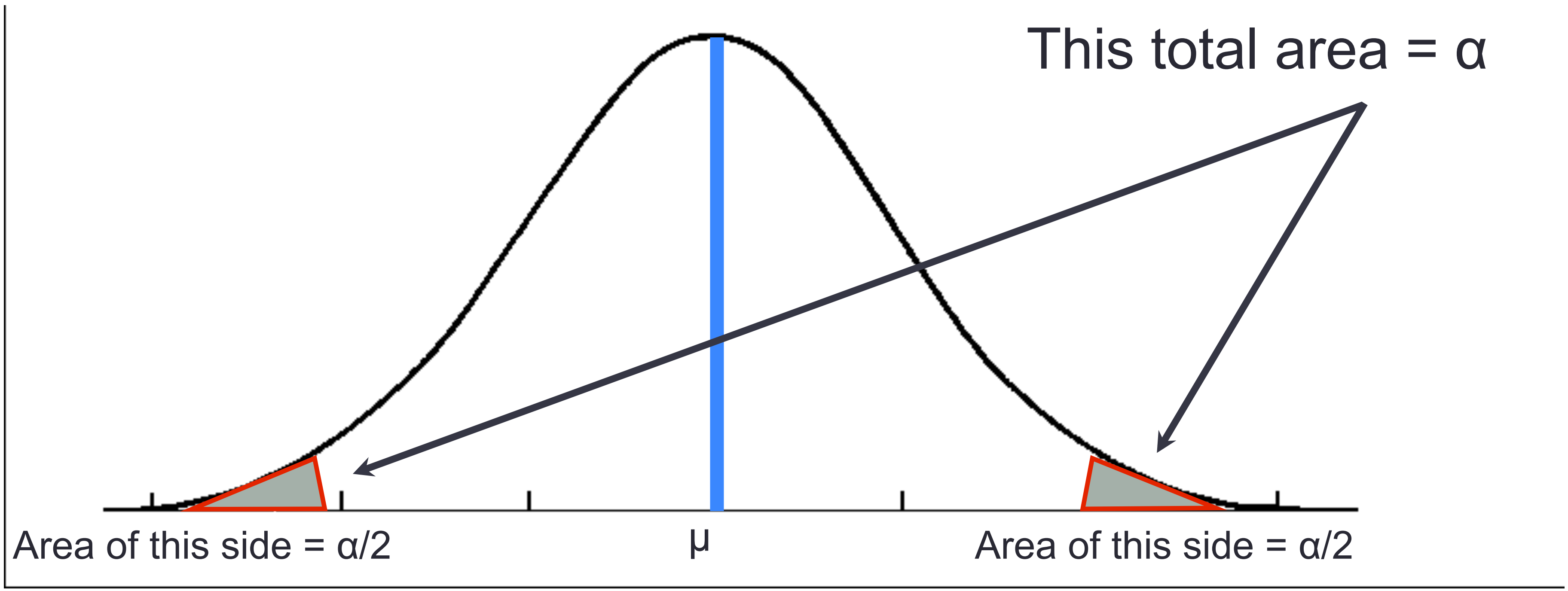


Establishing ‘Critical Region’

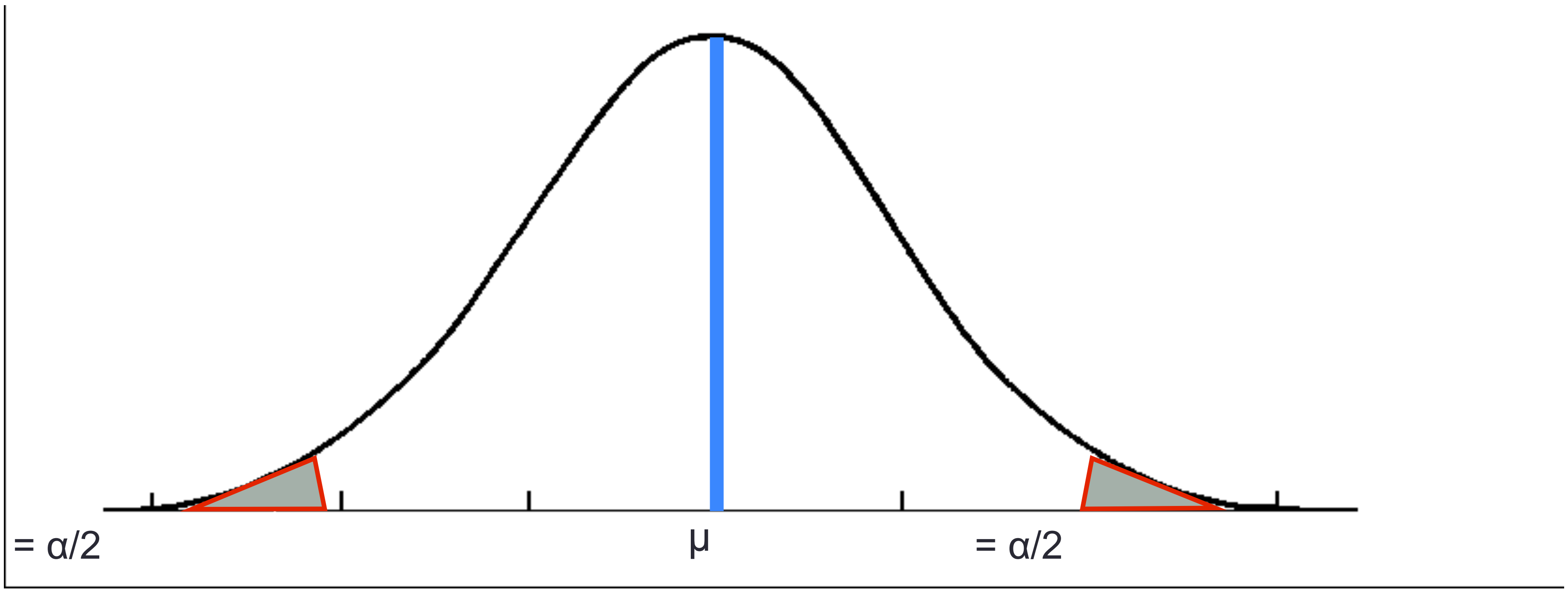
Then area is red is farthest from μ
and most likely to be *different*.



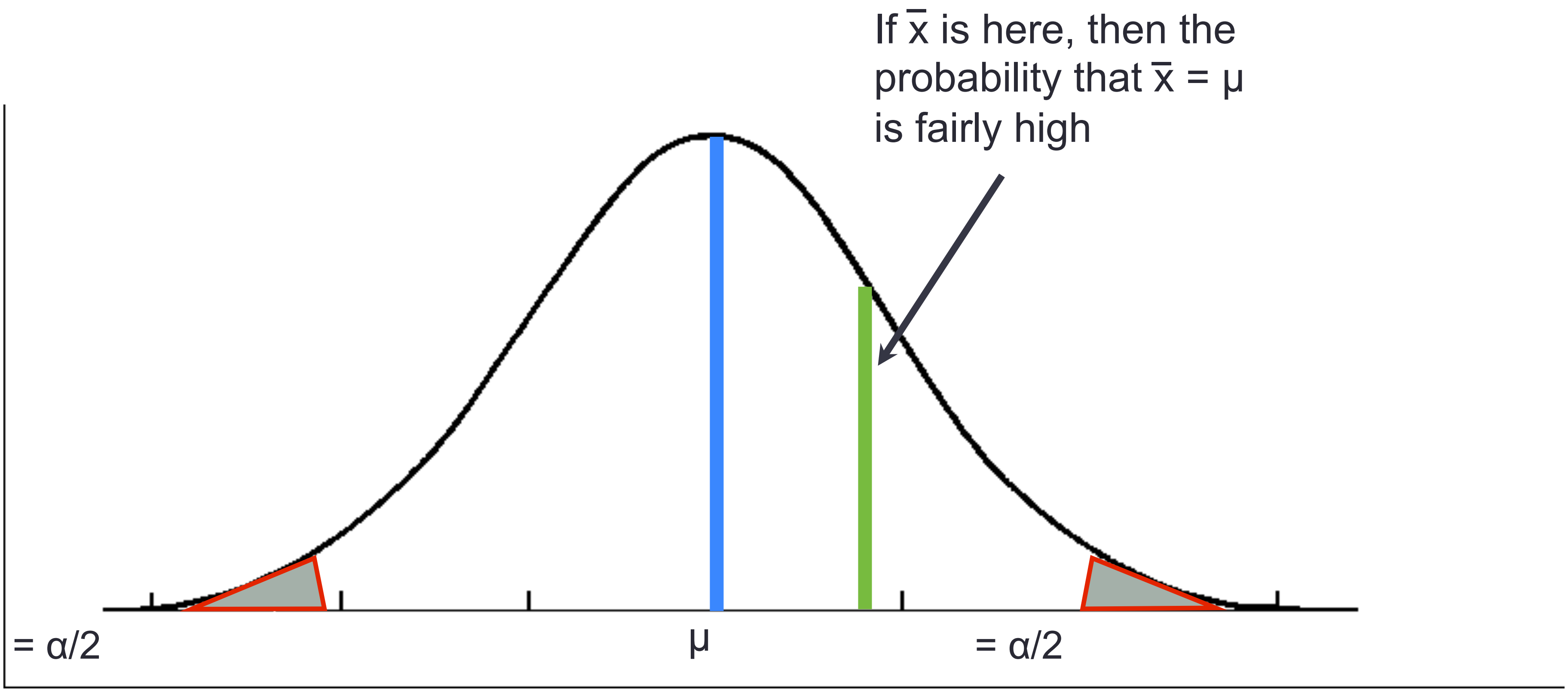
Establishing 'Critical Region'



Why we do this

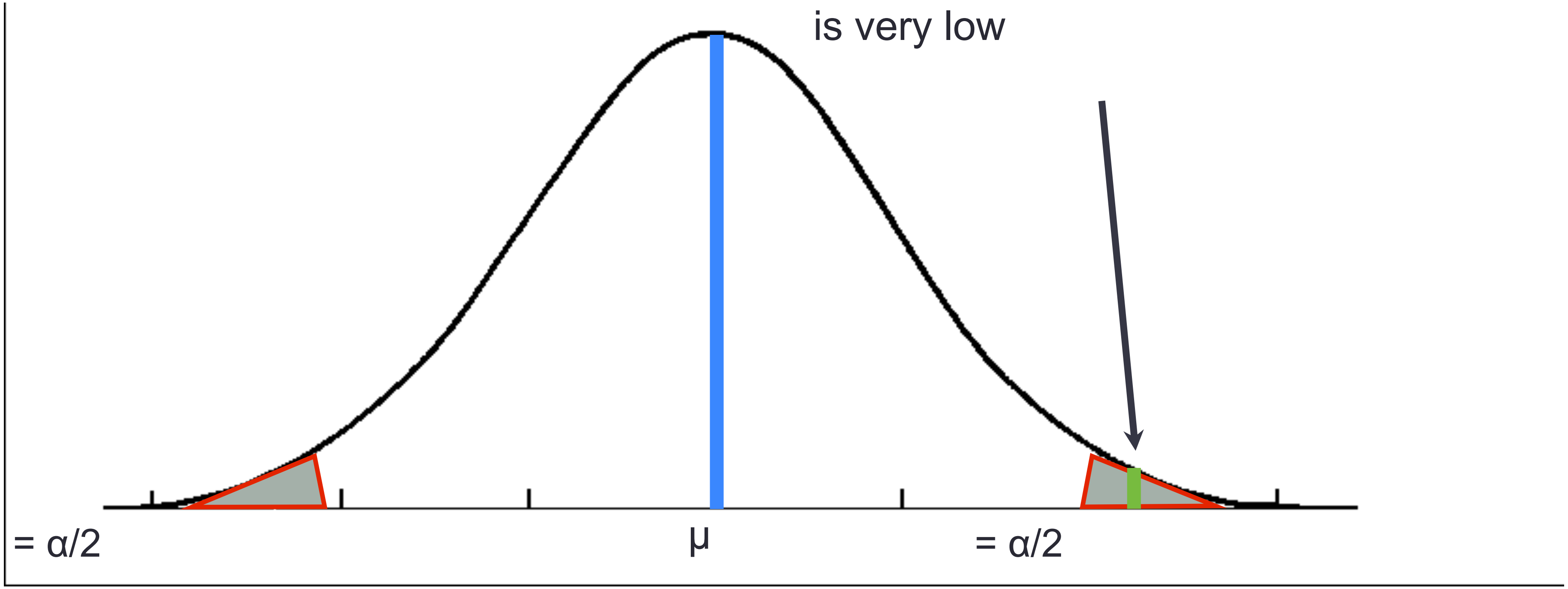


Why we do this



Why we do this

BUT, if $\bar{x}$ is here, then the probability that $\bar{x} = \mu$ is very low

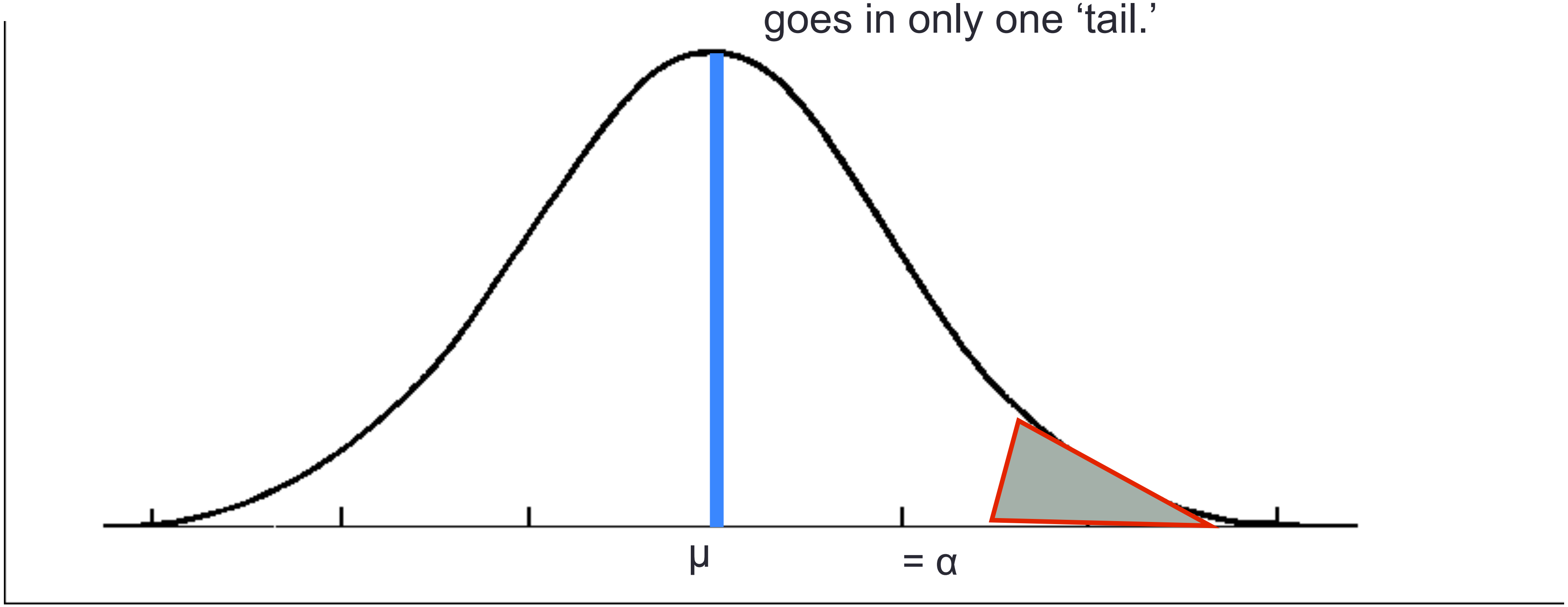


Why we do this

- Hypothesis tests determine the probability that $\bar{x} = \mu$. If the probability is less than α, then we consider them to be different (aka, statistically significant)

Establishing ‘Critical Region’

Sometimes, the critical region goes in only one ‘tail.’



One tail, two tails

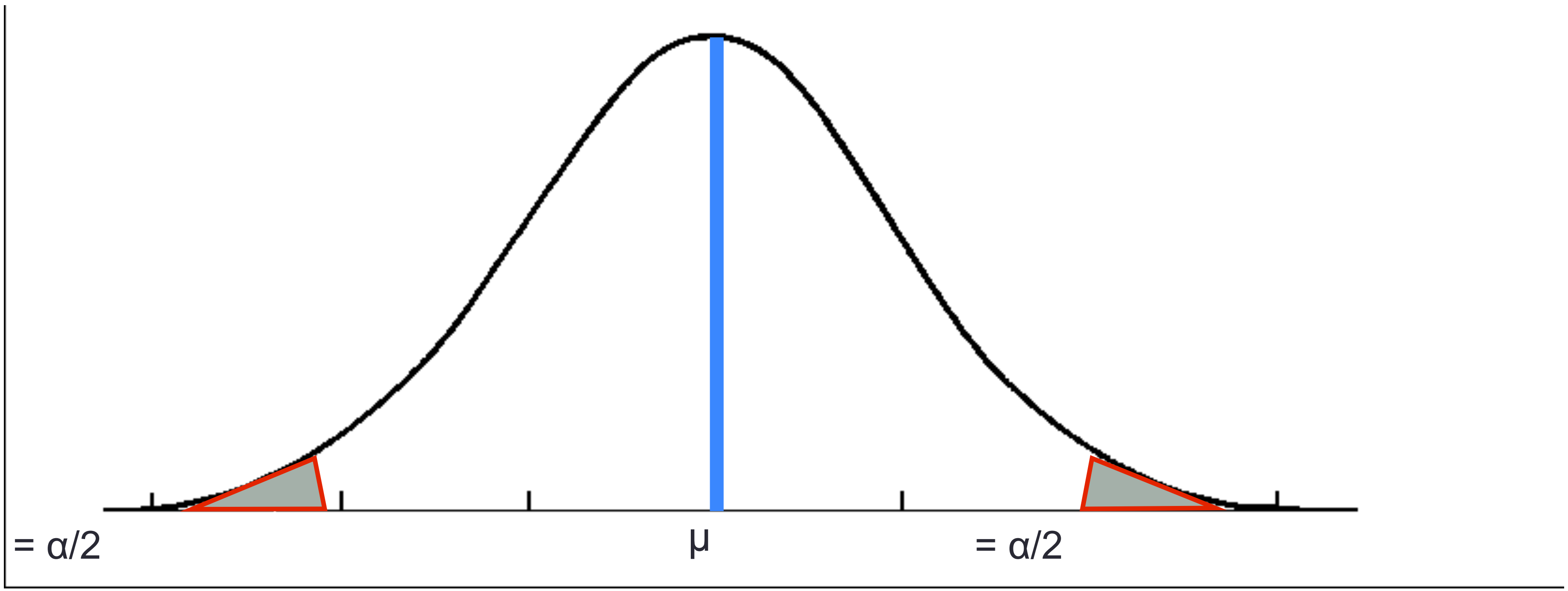
- Two-tailed tests are for testing ‘any means difference’
- One-tailed tests are for testing ‘a mean that is *greater than* or *less than* μ’

One tail, two tails

- Although one-tailed tests seem appealing (men earn *more than* women), they are often viewed with skepticism. As such, two-tailed tests are the most common (men earn *a different amount than* women).

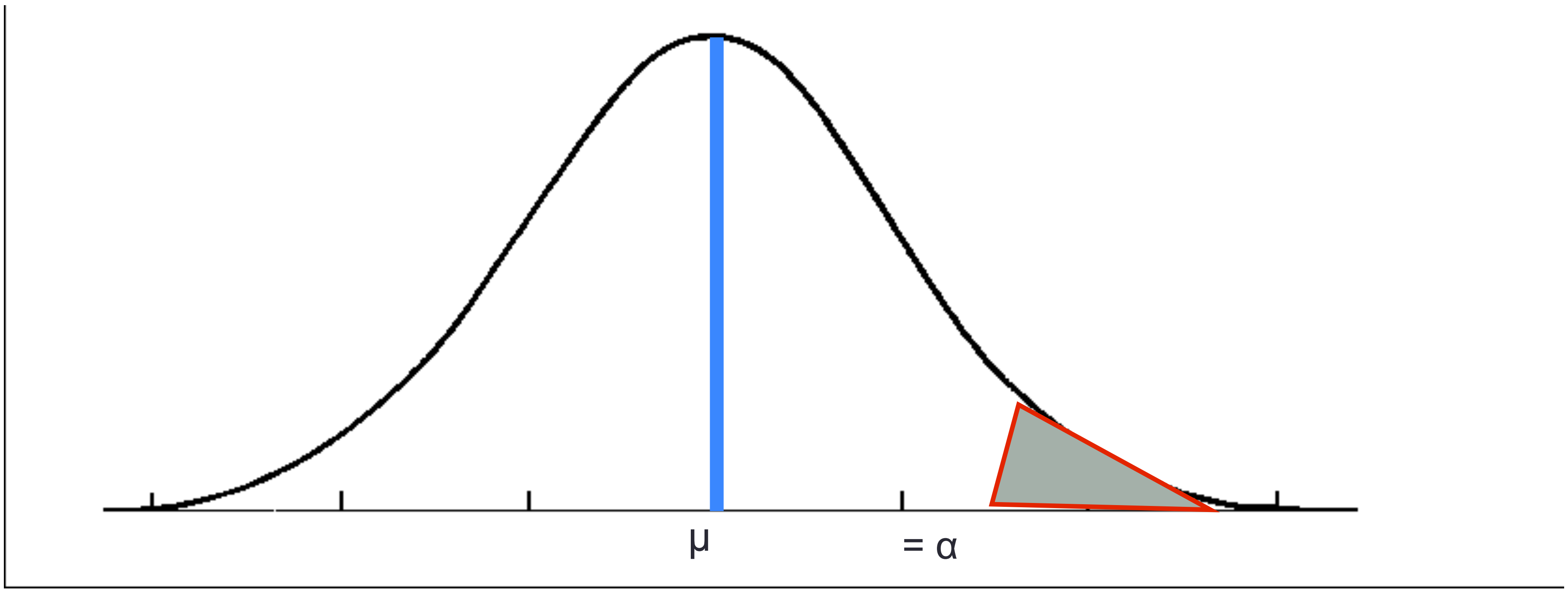
Two tails:

$\mu_{men} \neq \mu_{women}$



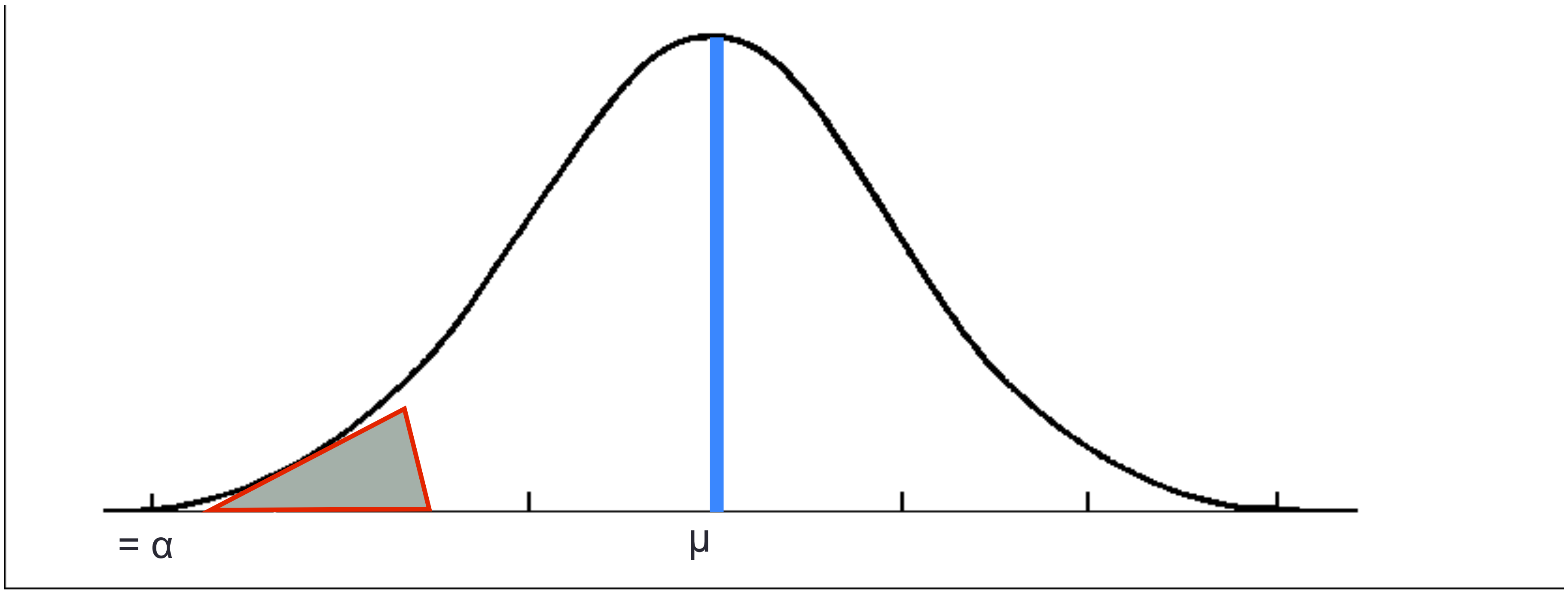
One tail:

$\mu_{men} > \mu_{women}$



One tail:

$\mu_{men} < \mu_{women}$



Test Yourself

- Our observation was:
- In a large General Social Survey sample, men earned an average of $45,273 and women an average of $33,303.
- Should we make this a one-tailed or two-tailed test?

Answer

- Even though it is very tempting to test the hypothesis that men earn *more than* women, the best approach is a two-tailed test (men earn a *different amount* than women). This is a more conservative approach, and is open to *any* research finding.

Critique of one-tailed tests

- One-tailed tests make it easier to detect statistical significance
- Dismisses the possibility that the finding is opposite of what you expected (example: people on low-fat diets weigh *less than* those not on low-fat diets)

Test Yourself

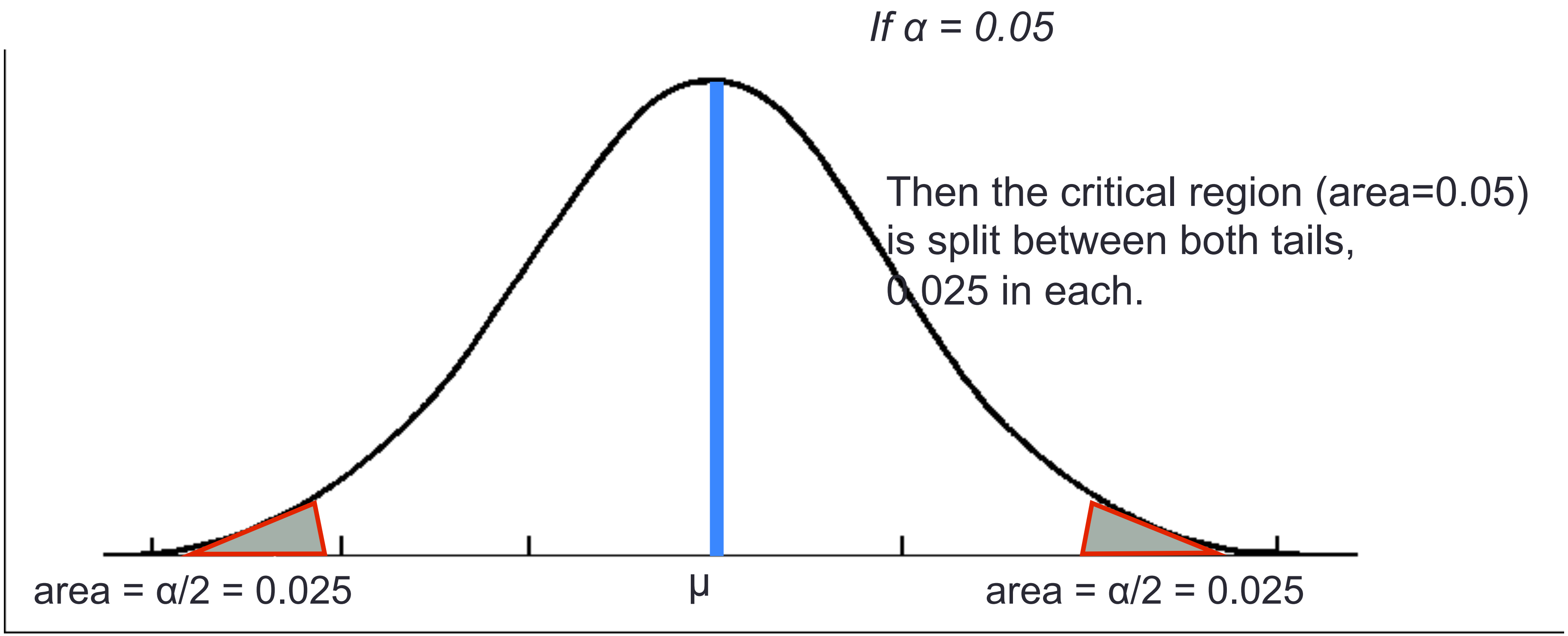
- If α is = 0.10
- For a two-tailed test, how much of alpha will go in the left tail? And the right tail?
- For a one-tailed test, how much of alpha will go in the left tail? And the right tail?

Answer

If α is = 0.10

- For a two-tailed test, how much of alpha will go in the left tail? And the right tail?
- *Half of alpha will go in each tail (0.05 in the left, and 0.05 on the right)*
- For a one-tailed test, how much of alpha will go in the left tail? And the right tail?
- *Depends on which tail; whichever tail you hypothesize, all of alpha will go in that tail.*

Threshold for critical region, two tails (α=0.05)



Threshold for critical region, two tails (α=0.05)

- If area in each tail is = 0.025
- Go to Appendix A...
- Scan column c (area beyond Z) until you find an area of 0.025. The corresponding Z Score is the threshold for the critical area.

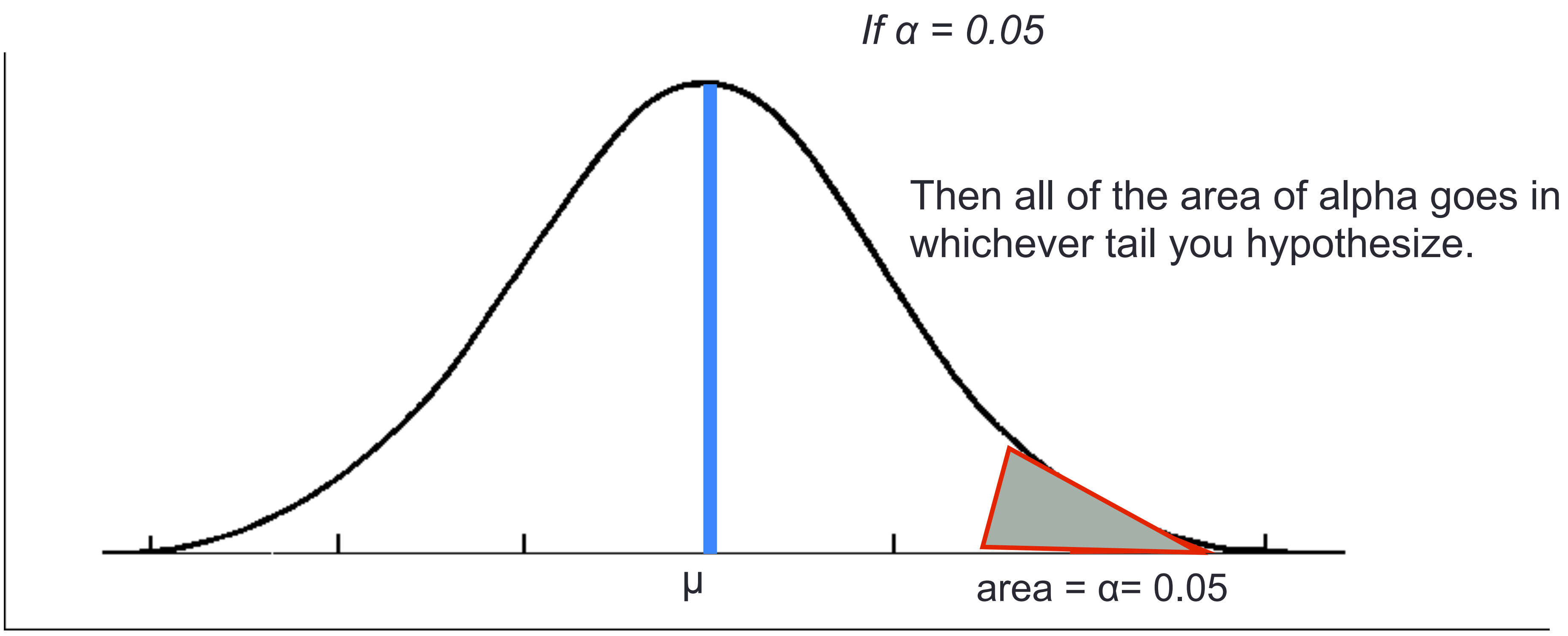
Threshold for critical region, two tails (α=0.05)

(a)	*(b)*	*(c)*
Z	*Area Between Mean and Z*	*Area Beyond Z*
1.95	0.4744	0.0256
1.96	0.4750	0.0250
1.97	0.4756	0.0244

Threshold for critical region, two tails (α=0.05)

(a)	*(b)*	*(c)*
Z	*Area Between Mean and Z*	*Area Beyond Z*
1.95	0.4744	0.0256
$\boxed{1.96}$	$\boxed{0.4750}$	$\boxed{0.0250}$
1.97	0.4756	0.0244

Threshold for critical region, one tail (α=0.05)



Threshold for critical region, two tails (α=0.05)

- If area in the tail is = 0.05
- Go to Appendix A...
- Scan column c (area beyond Z) until you find an area of 0.05. The corresponding Z Score is the threshold for the critical area.

Threshold for critical region, two tails (α=0.05)

(a)	*(b)*	*(c)*
Z	*Area Between Mean and Z*	*Area Beyond Z*
1.64	0.4495	0.0505
1.65	0.4505	0.0495
1.66	0.4515	0.0485

Threshold for critical region, two tails (α=0.05)

(a)	*(b)*	*(c)*
Z	*Area Between Mean and Z*	*Area Beyond Z*
1.64	0.4495	0.0505
1.65	0.4505	0.0495
1.66	0.4515	0.0485

Test Yourself

- If $\alpha = 0.01$, what is the Z threshold for a two tailed test? And for a one-tailed test?

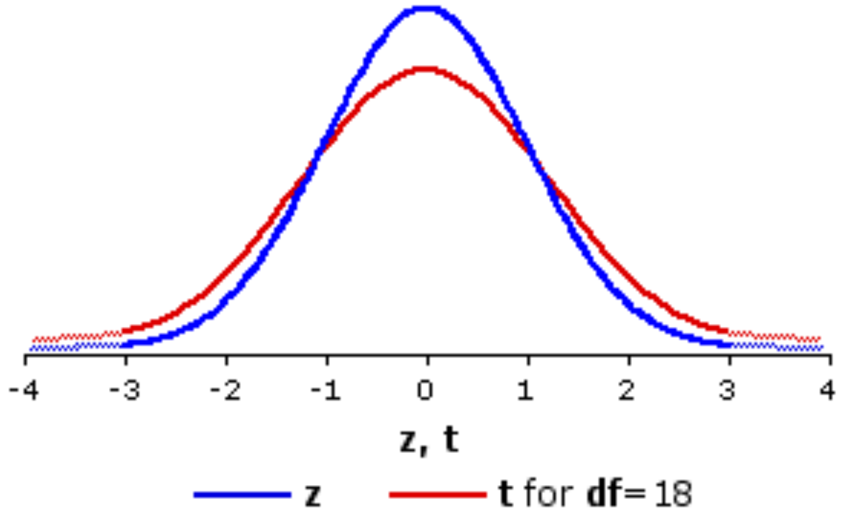
Answer

- Z = 2.58 for a two-tailed test
- Z = -2.33 or +2.33 for a one-tailed test (negative if it goes in left tail, positive if right tail)
- Hint: A table of the relevant Z Scores for different levels of alpha is on the sheet of formulas that you can reference during exams.

Small Samples

- For small samples ($N < 100$), Z is not appropriate. When working with small samples, we will instead use the *t* distribution.
- Same process as Z, but different distribution (Appendix B)

Z versus *t*



The Research Process

- Step 1: Decide which hypothesis test to use. State your assumptions, and verify that your data meet the test requirements:
- Step 2: State the null and research hypotheses
- Step 3: Select alpha and establish the *critical region.*

The Research Process

- Step 4: Compute the Test Statistic
- Step 5: Interpret the results

Back to Our Observation

- In a large General Social Survey sample, men earned an average of $45,273 and women an average of $33,303.
- More info:

Men	Women
x = 45,273	x = 33,303
s = 29,061	s = 22,485
N = 671	N = 589

Do Men and Women Have Different Income Levels?

- Two samples (men and women)
- Large sample (N > 100)
- Step 1: Conducting a two-sample hypothesis test of the mean, for a large sample. The dependent variable, income, is interval-ratio. The independent variable, sex, is nominal with two categories (m/w).

Do Men and Women Have Different Income Levels?

- Step 2:
- H_0: $\mu_{men} = \mu_{women}$
- H_1: $\mu_{men} \neq \mu_{women}$

Do Men and Women Have Different Income Levels?

- Step 3:
  - Alpha = 0.05, two-tailed
  - Z (critical) = ±1.96

Do Men and Women Have Different Income Levels?

- Step 4: Compute the test statistic.

Do Men and Women Have Different Income Levels?

- Step 5: Interpret the Results
- If Z or t obtained is in the critical area, then the finding is *statistically significant*.

Practice

- Do college students that live on campus have different levels of involvement in university activities than commuters?
- Results of a random sample (x= # activities):

Residents	Commuters
$\bar{x} = 12.4$	$\bar{x} = 10.2$
$s = 2.0$	$s = 1.9$
$N = 158$	$N = 173$

Practice

- A marriage counseling center needs to evaluate one of their newest programs, so they conducted a random sample survey. Among those who participated in their regular program, 59% of the marriages ended in divorce (among 182 couples). But among those who participated in their new program, 53% of the marriages ended in divorce (among 178 couples). Are the divorce rates different in the new program?

Practice

- In a random sample of senior citizens, is there a different rate of daily face-to-face interactions, by community? (x = number of daily face-to-face interactions)

Community 1	Community 2
$\bar{x} = 1.42$	$\bar{x} = 1.58$
s = 0.10	s = 0.78
N = 43	N = 37

Practice

- A random sample of 105 workers in a factory in Corvallis, OR found that workers make on average $24,375, with a standard deviation of $799. IRS data show that among all Corvallis residents, the average earnings are $24,230, with a standard deviation of $845. Do the factory workers have *higher* earnings than the rest of Corvallis?
- Also test whether the earnings are *different*.

Common probability distributions