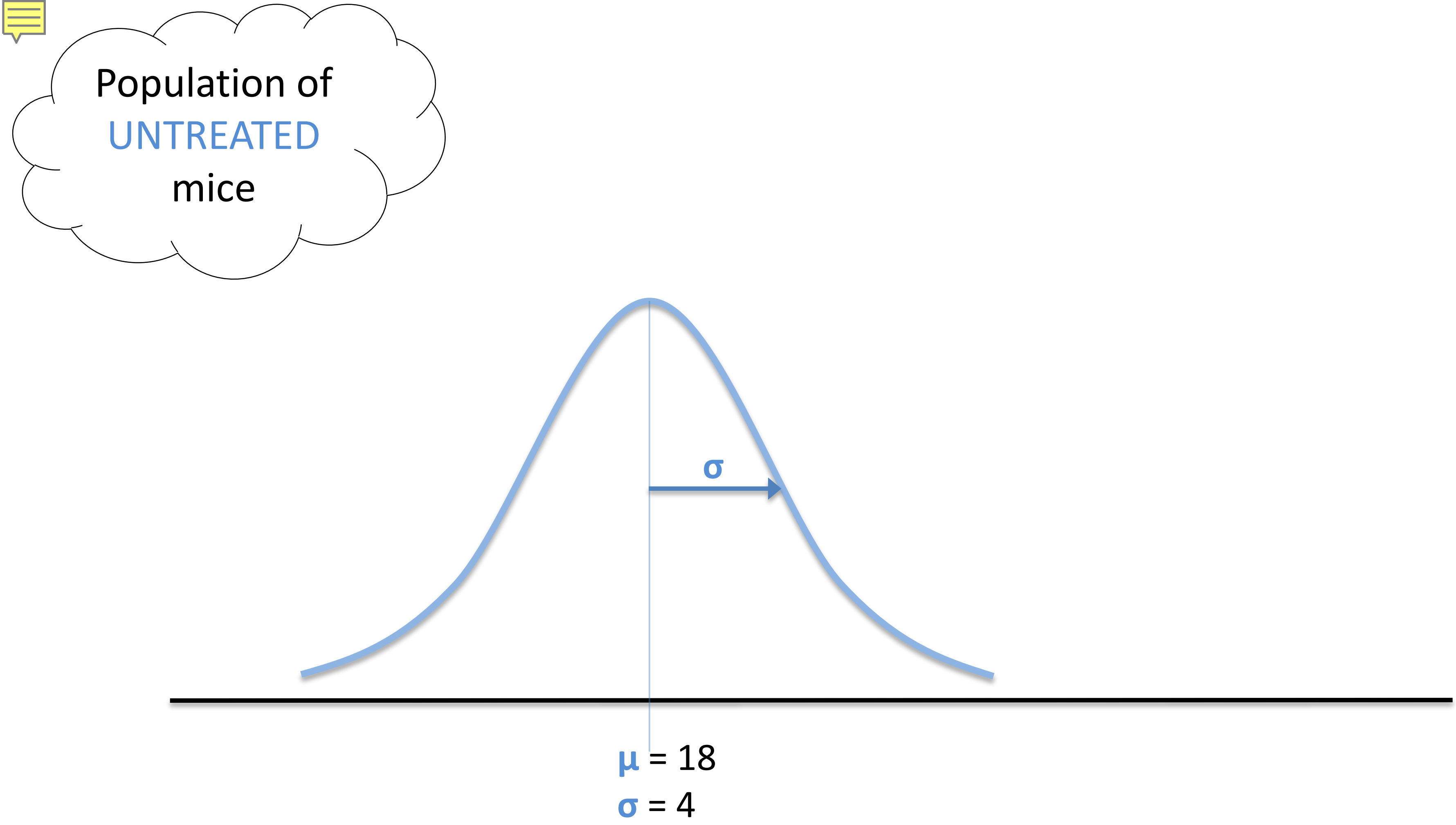
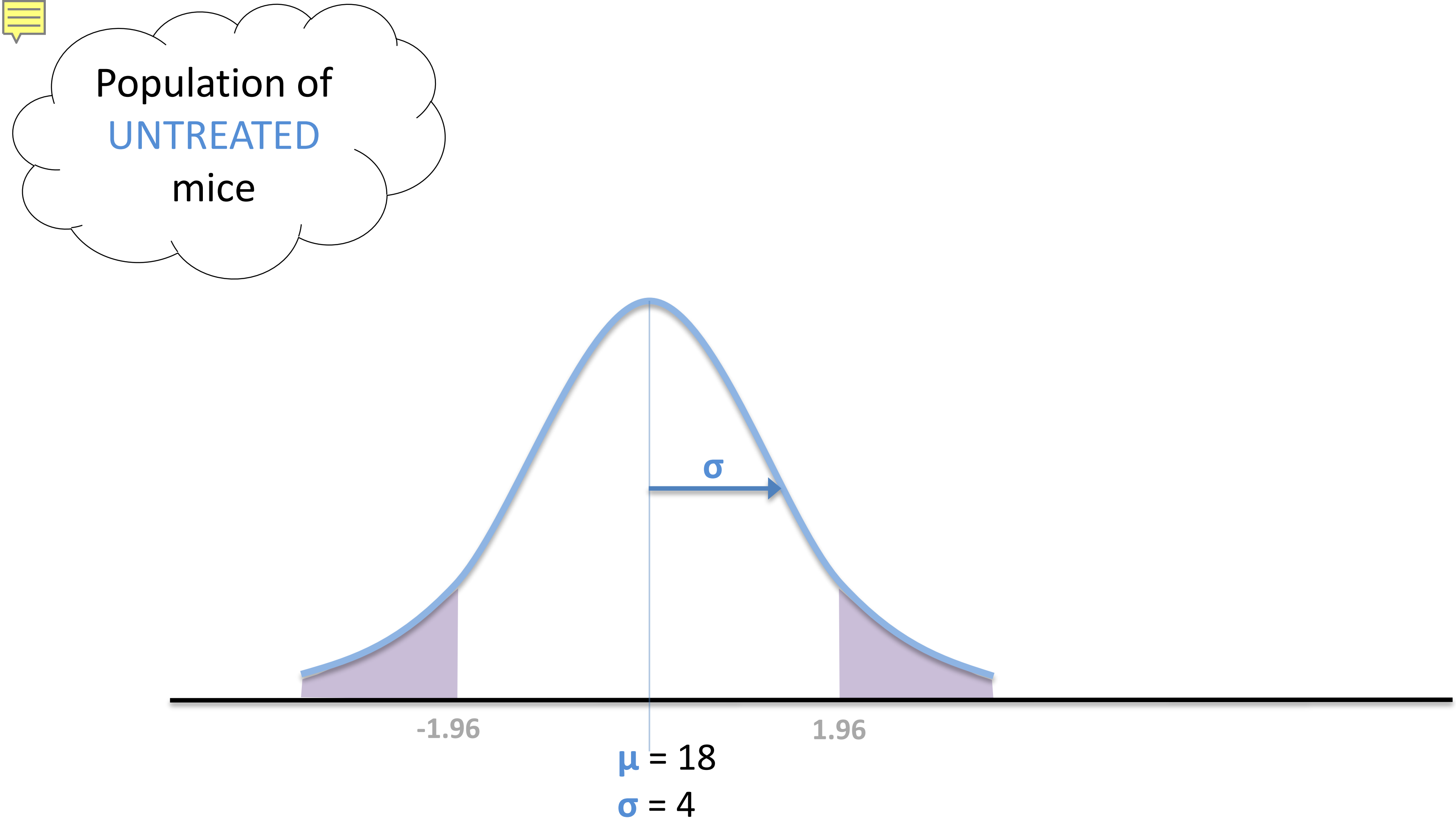
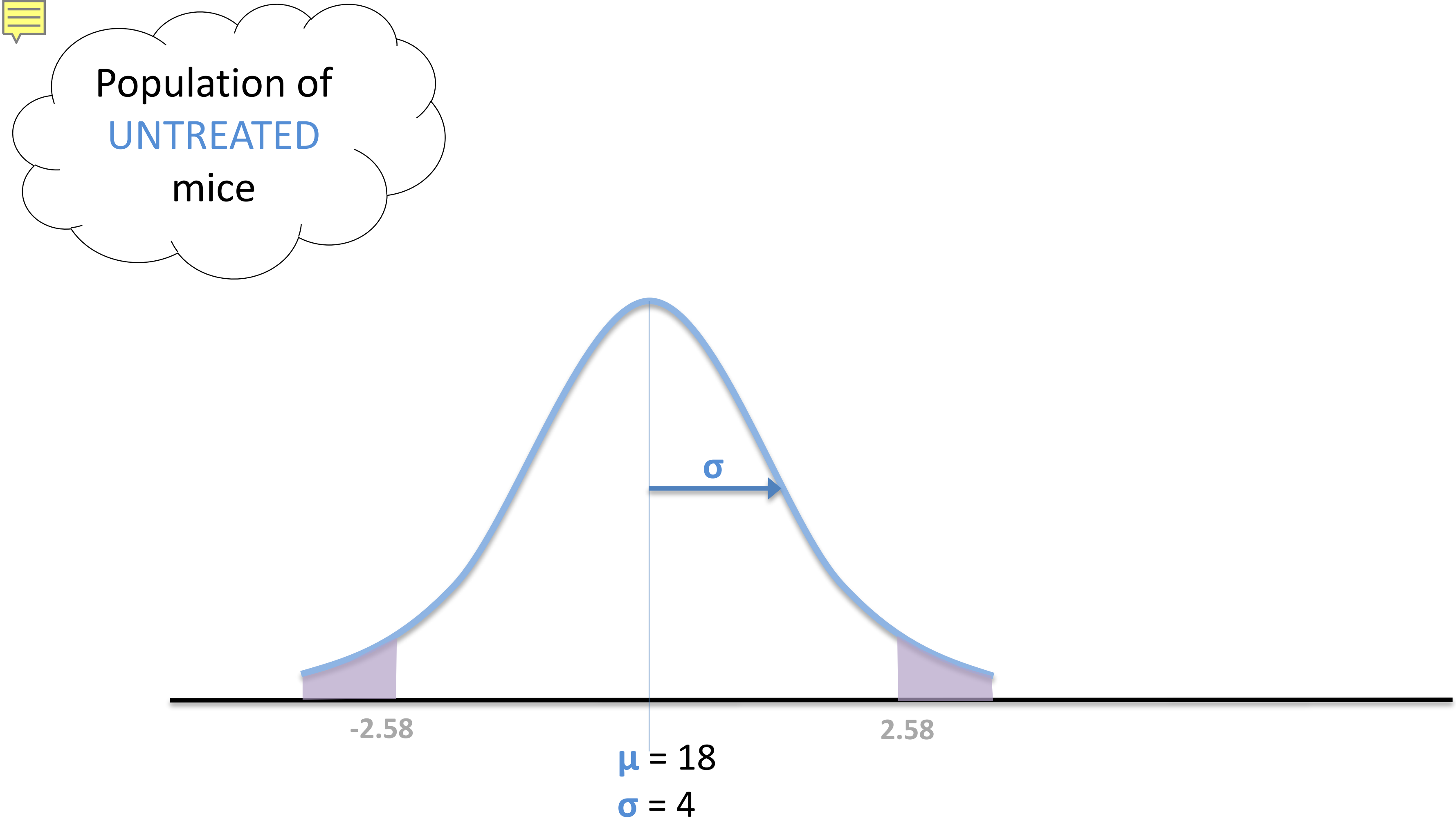
Hypothesis Testing

How many datapoints would you like to buy?

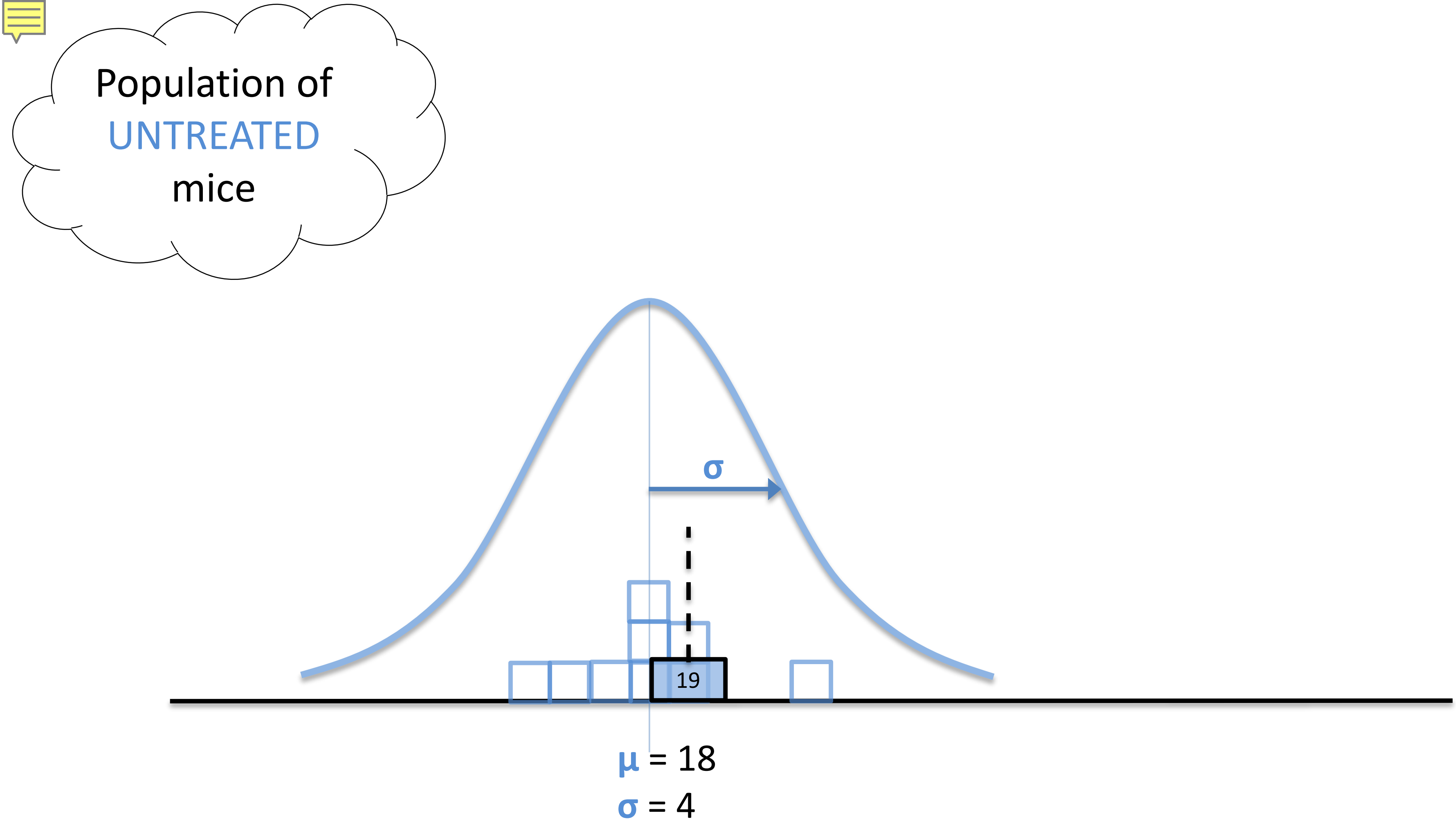
Recall what we know so far

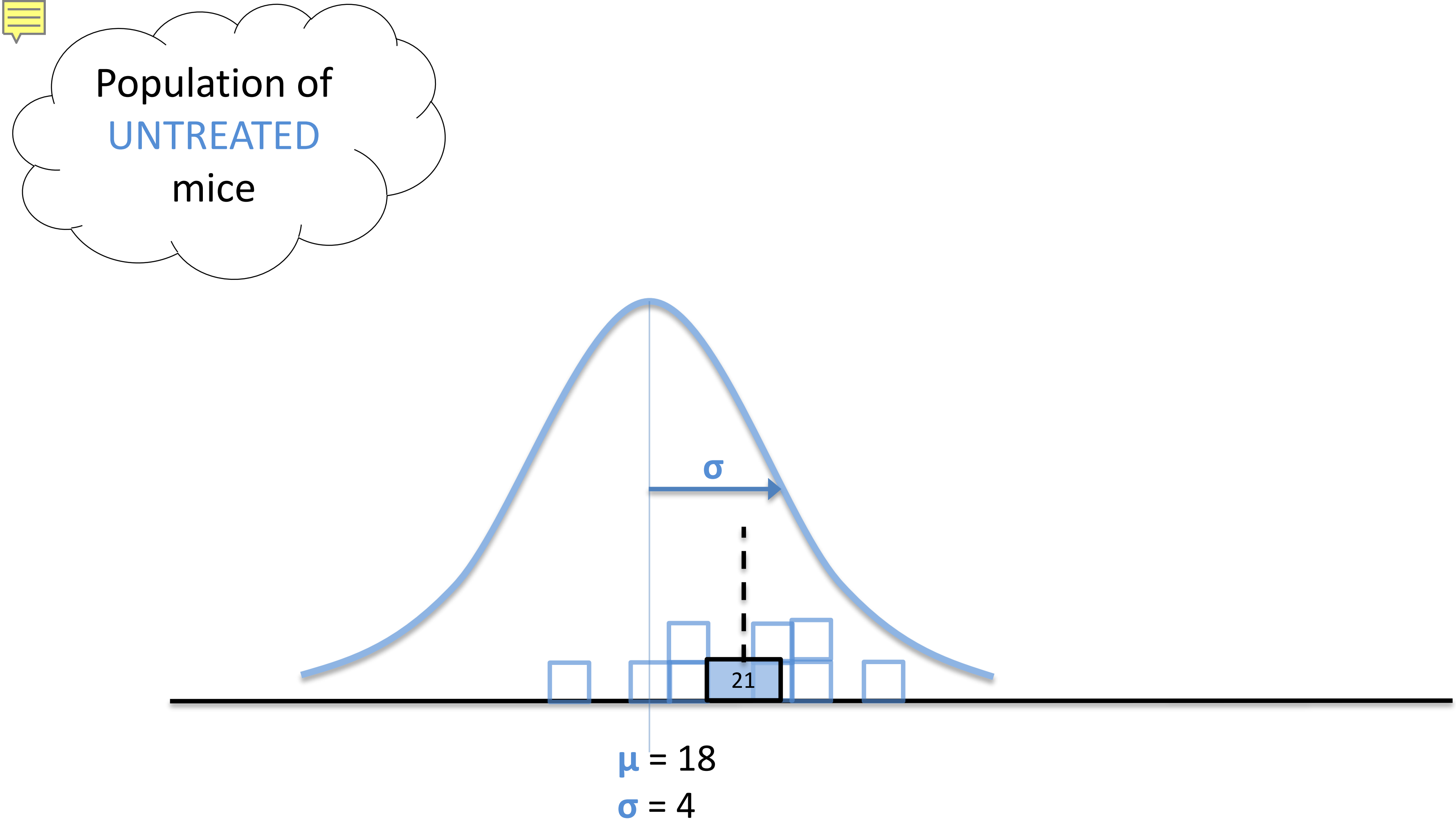

Population of
UNTREATED
mice
σ
μ = 18
σ = 4

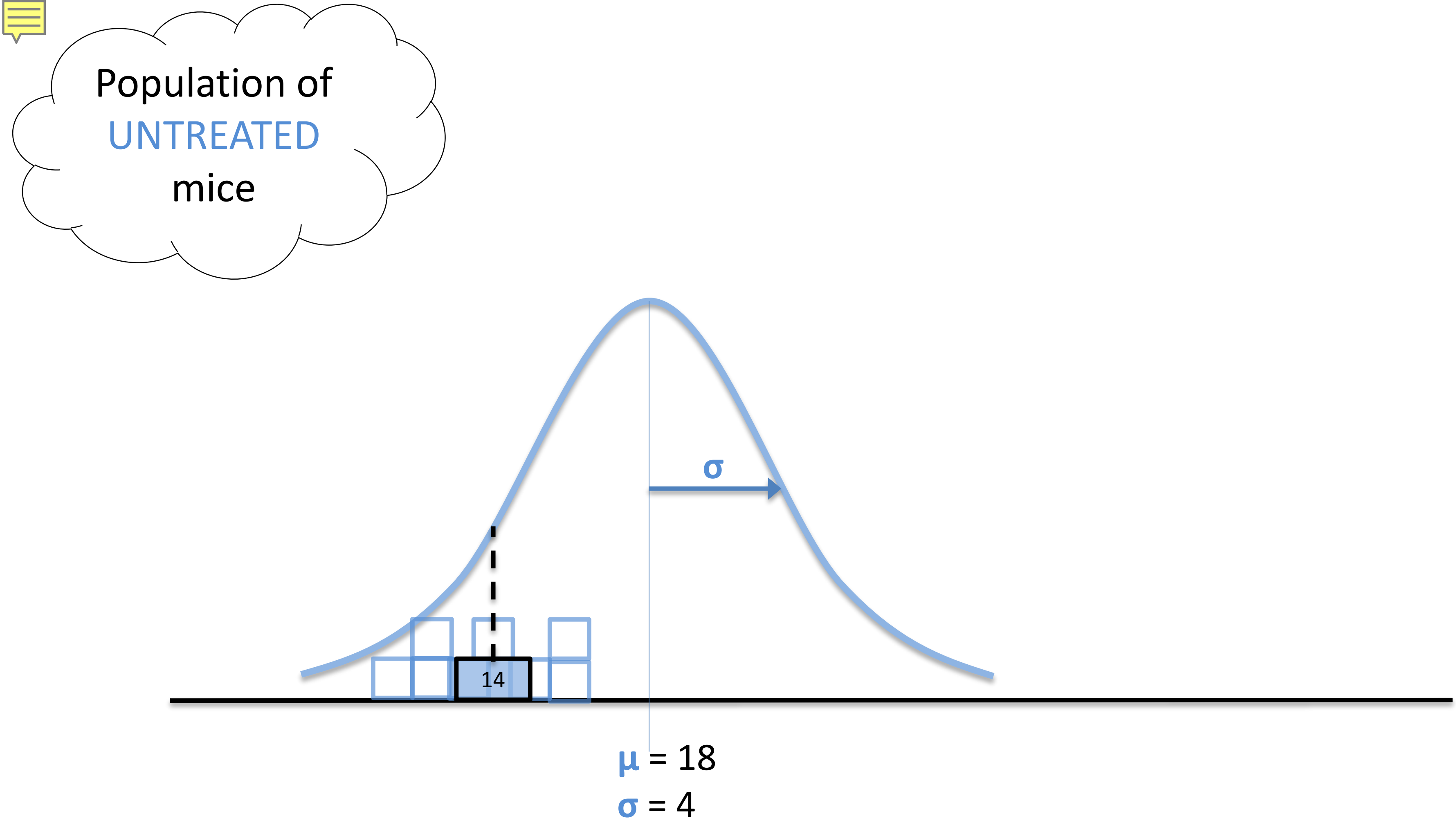

Population of
UNTREATED
mice
σ
-1.96
1.96
μ = 18
σ = 4

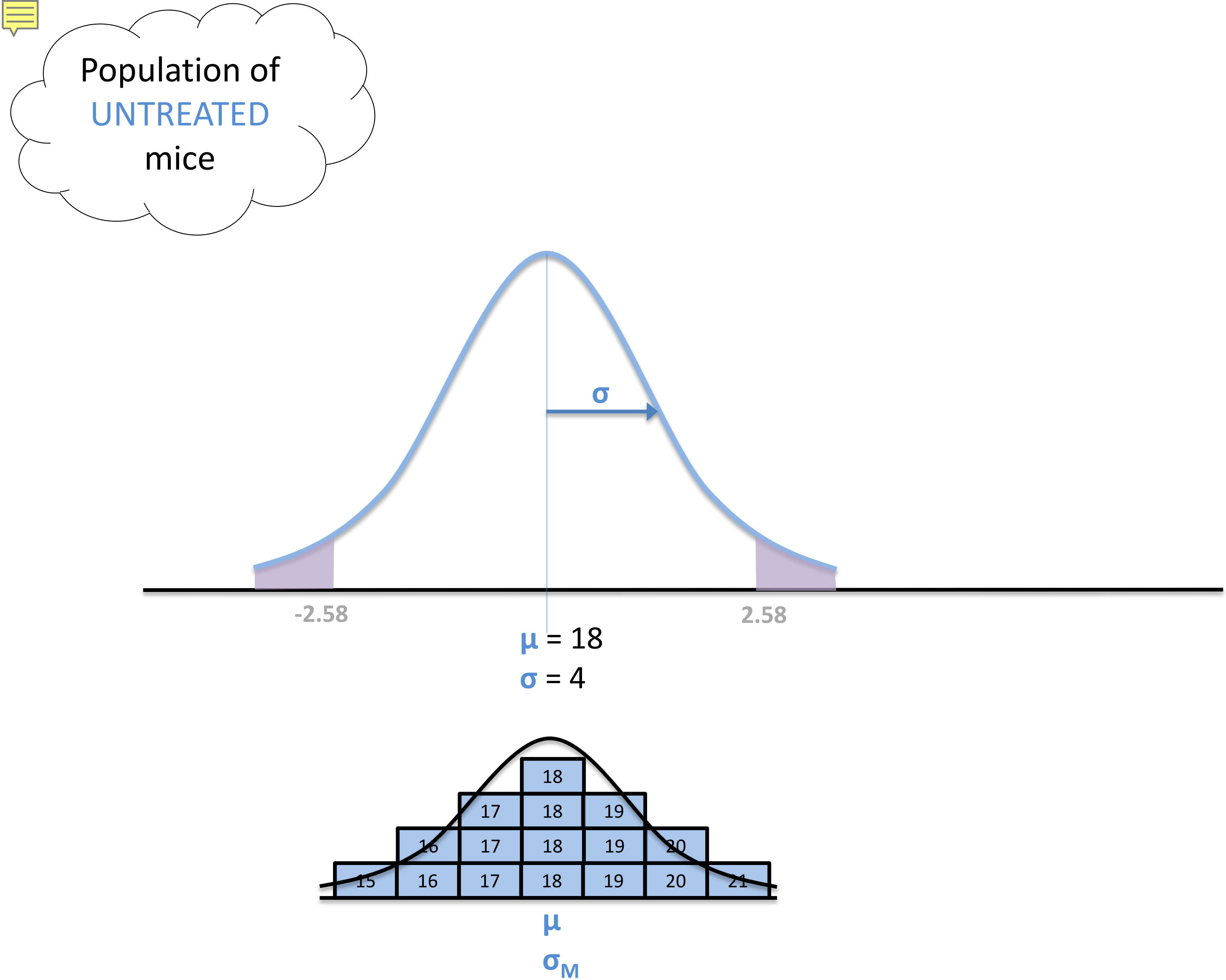


Population of
UNTREATED
mice
σ
-2.58
2.58
μ = 18
σ = 4

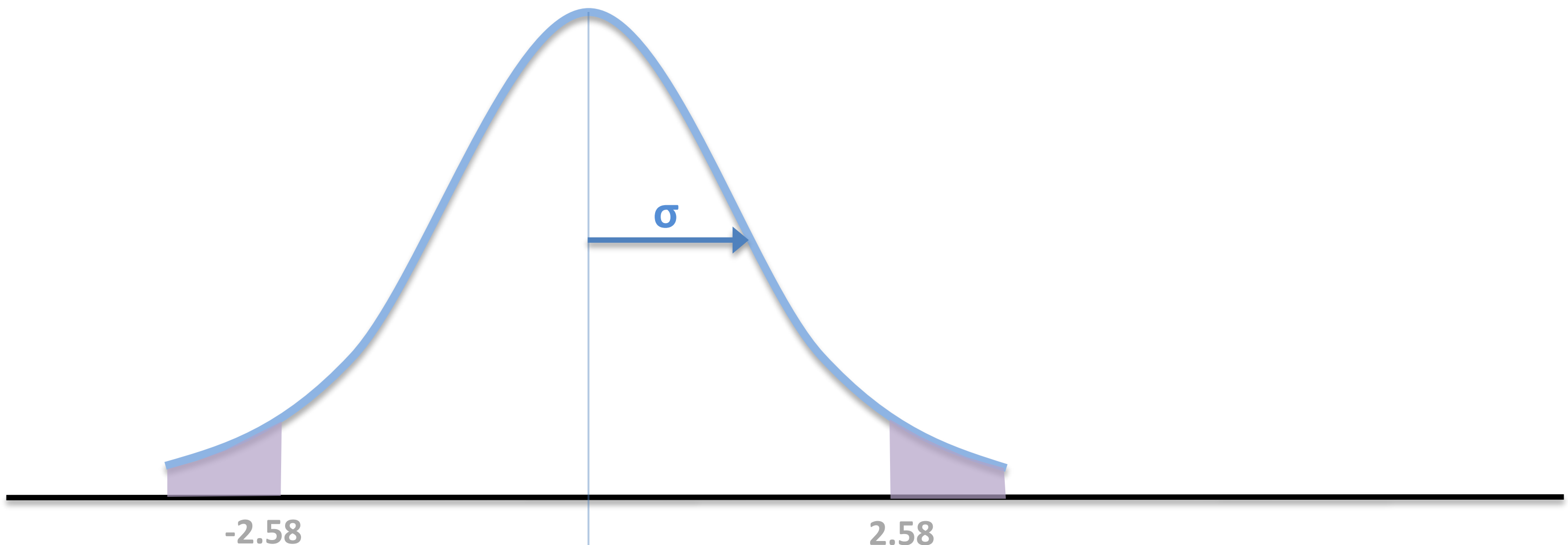

Population of
UNTREATED
mice
σ
19
μ = 18
σ = 4


Population of
UNTREATED
mice
σ
21
μ = 18
σ = 4


Population of
UNTREATED
mice
σ
14
μ = 18
σ = 4

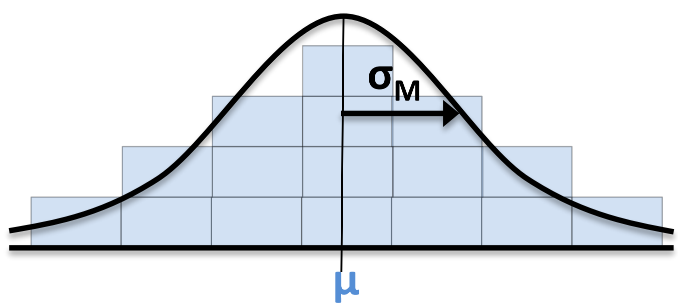

Population of
UNTREATED
mice
σ
-2.58
2.58
μ = 18
σ = 4
18
17
18
19
16
17
18
19
20
15
16
17
18
19
20
21
μ
σ_M

Population of UNTREATED mice

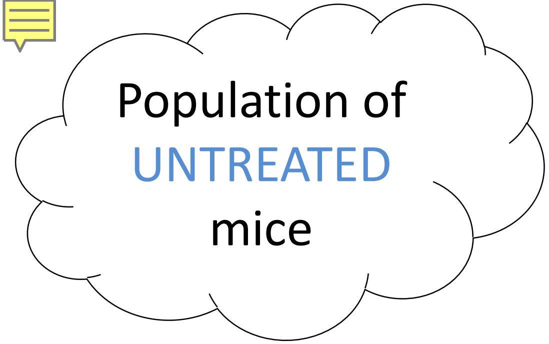


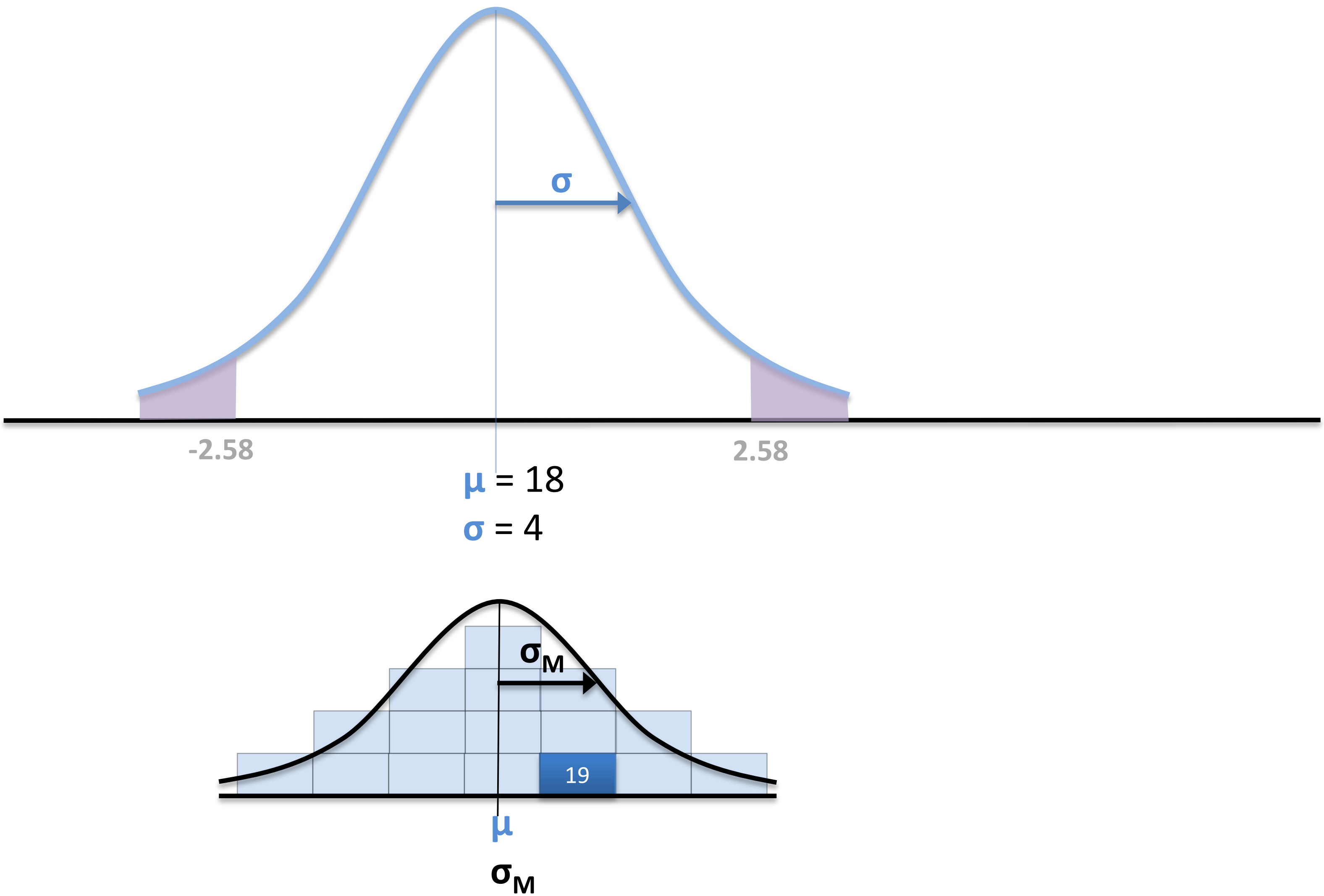
μ = 18

σ = 4

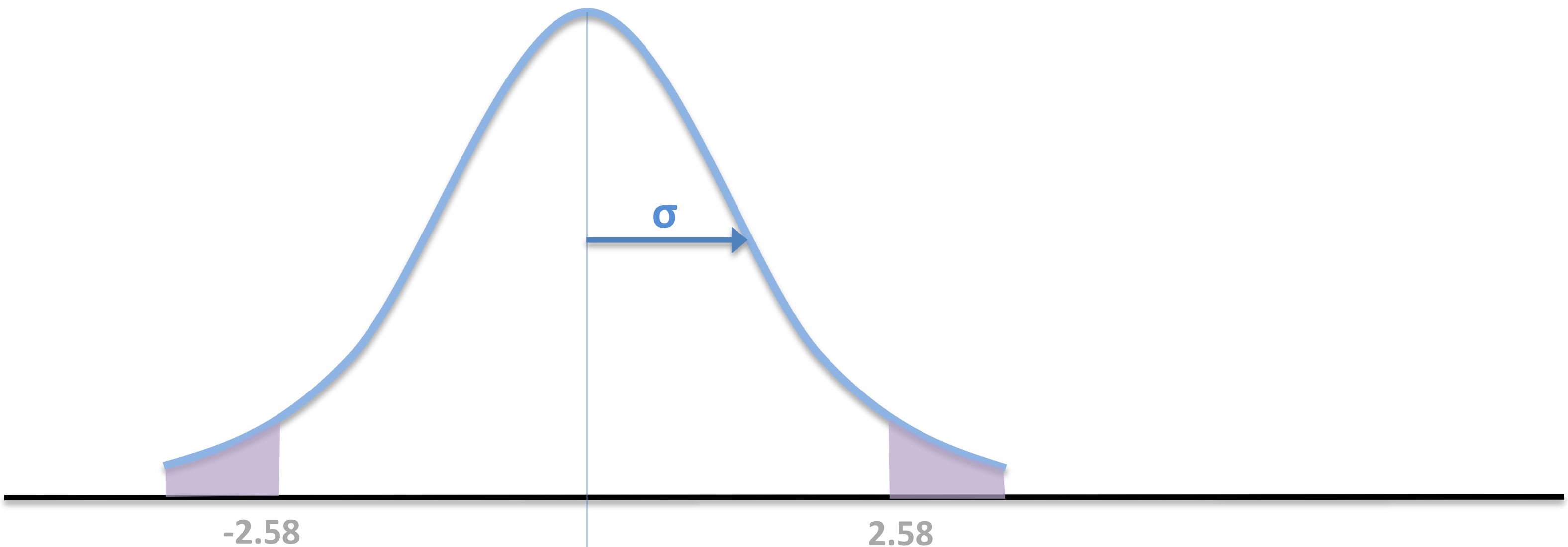


σ_M


Population of
UNTREATED
mice

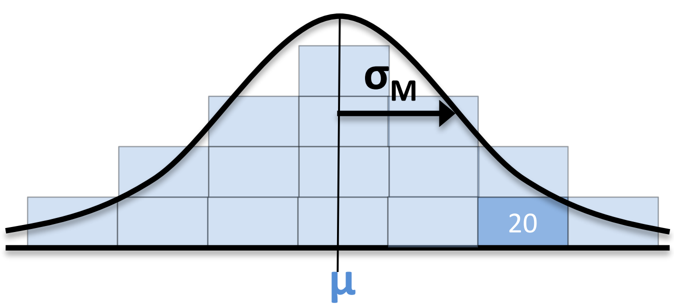

σ
-2.58
2.58
μ = 18
σ = 4
σ_M
19
μ
σ_M

Population of UNTREATED mice



$\mu = 18$

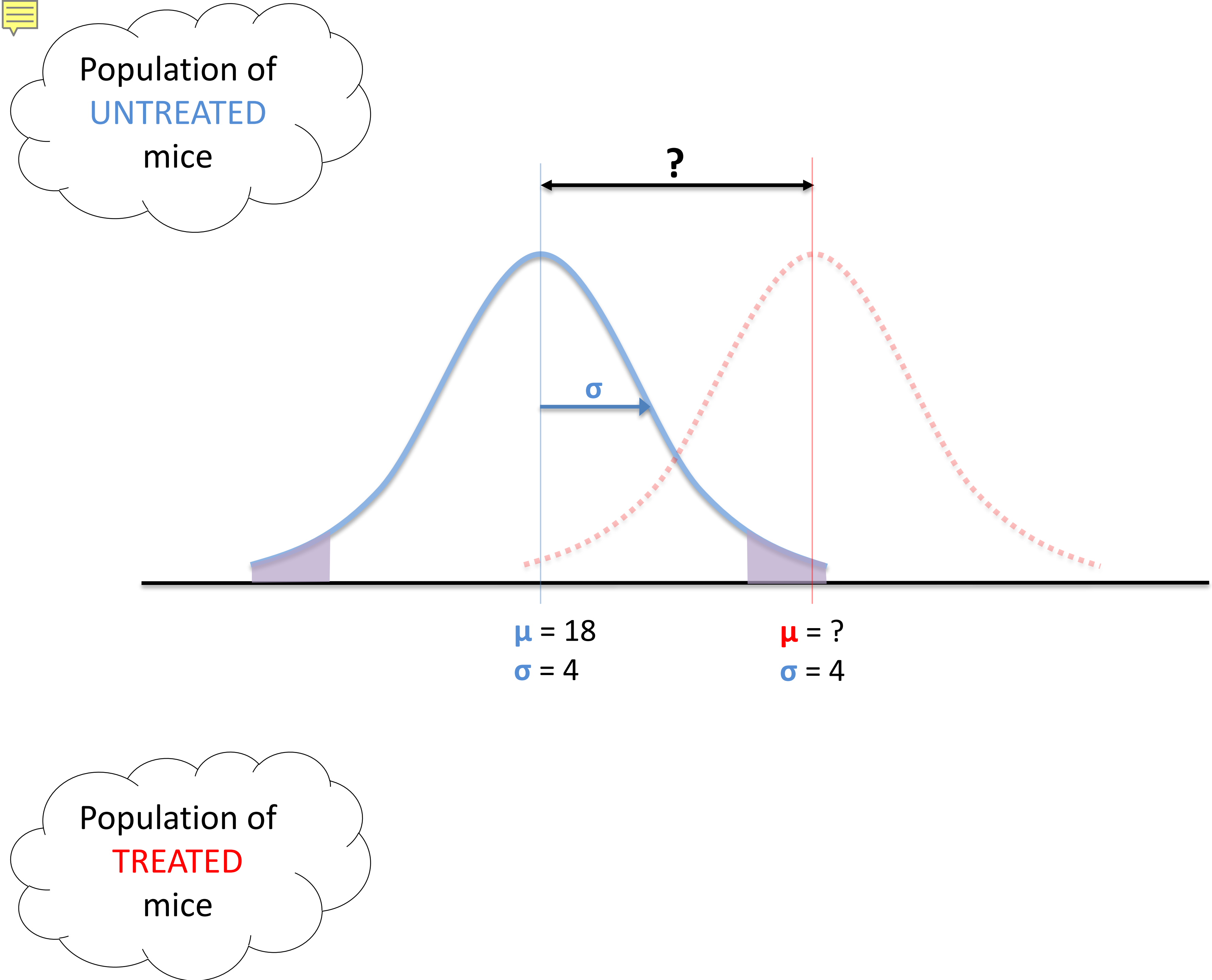
$\sigma = 4$

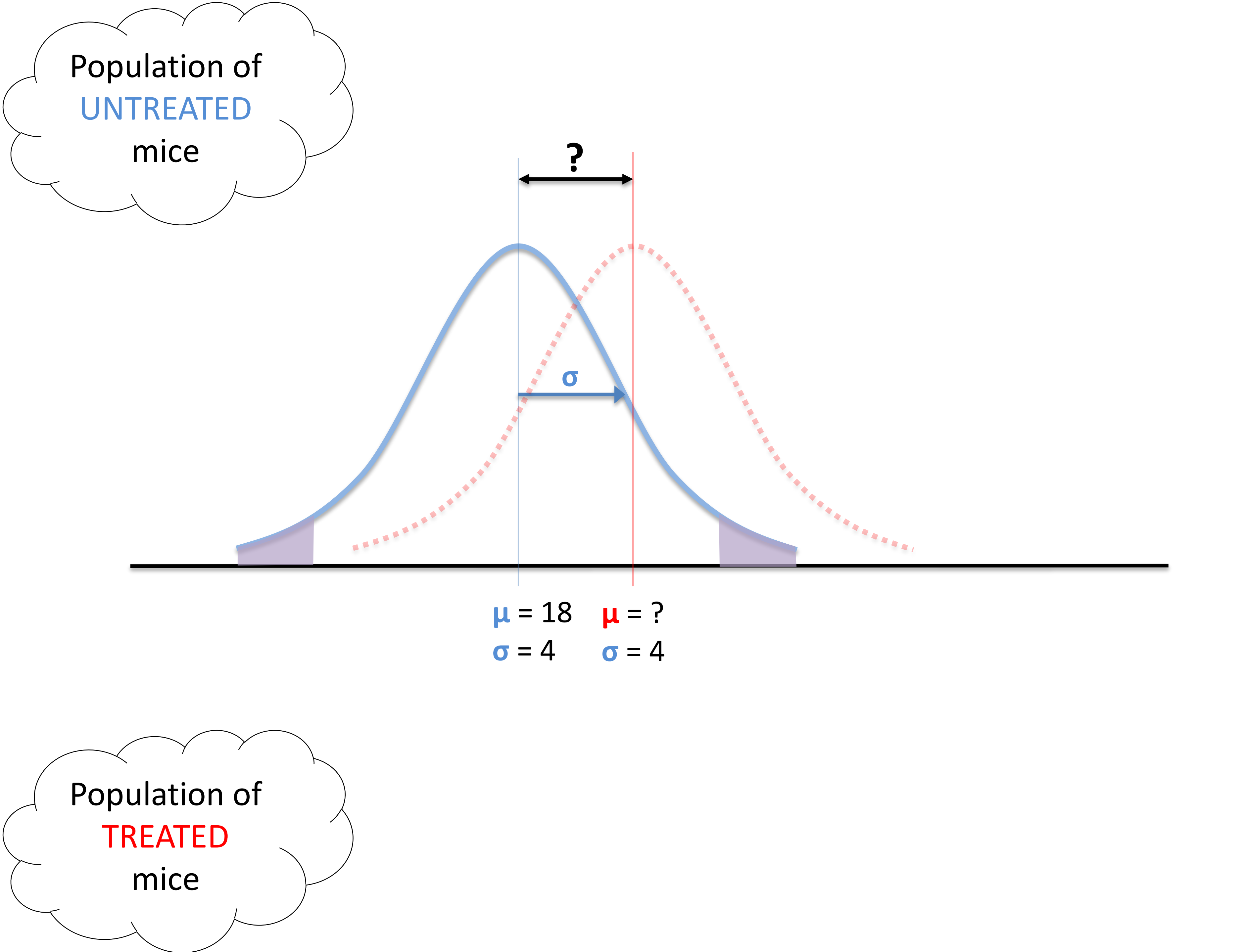


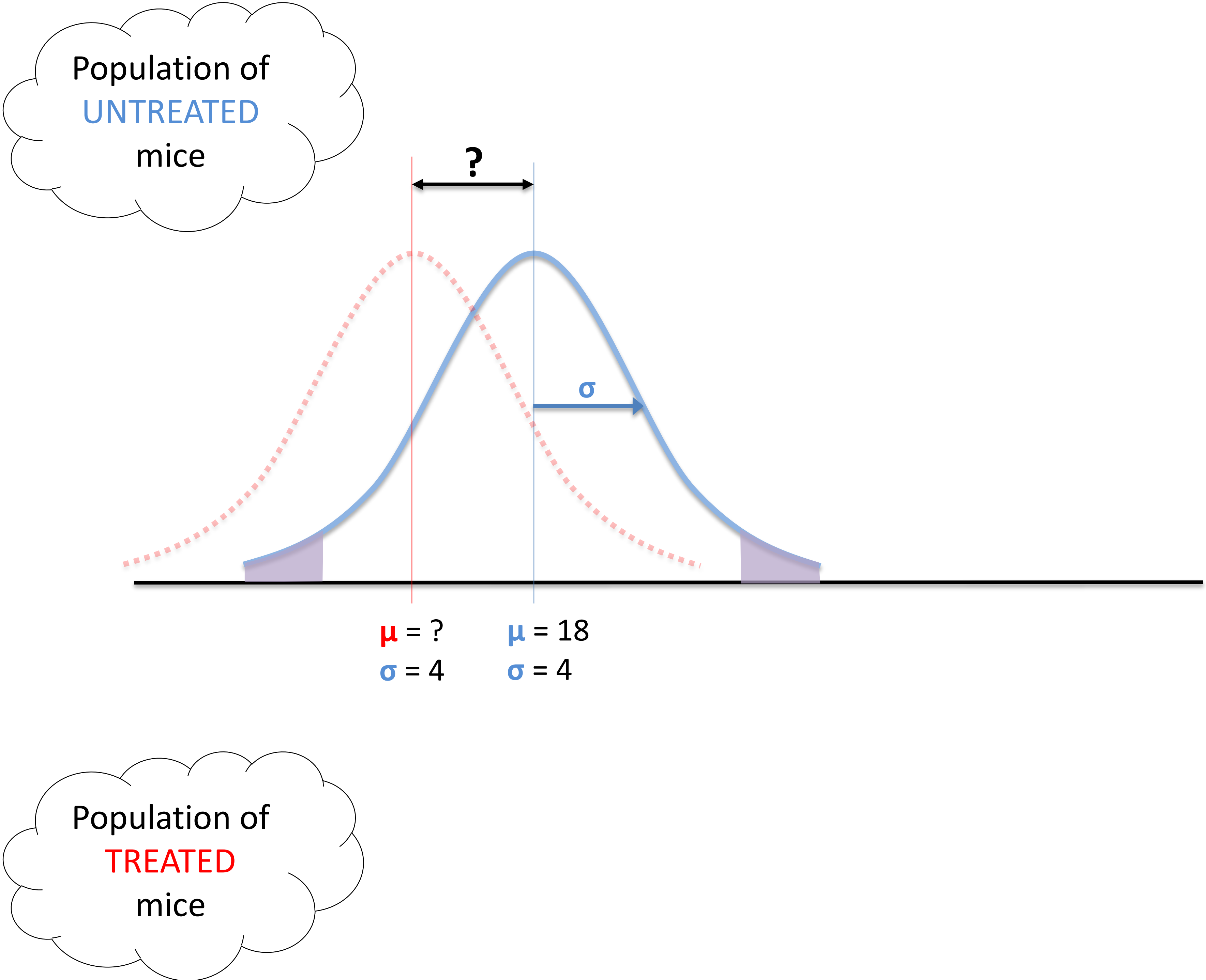
σ_M

Think design

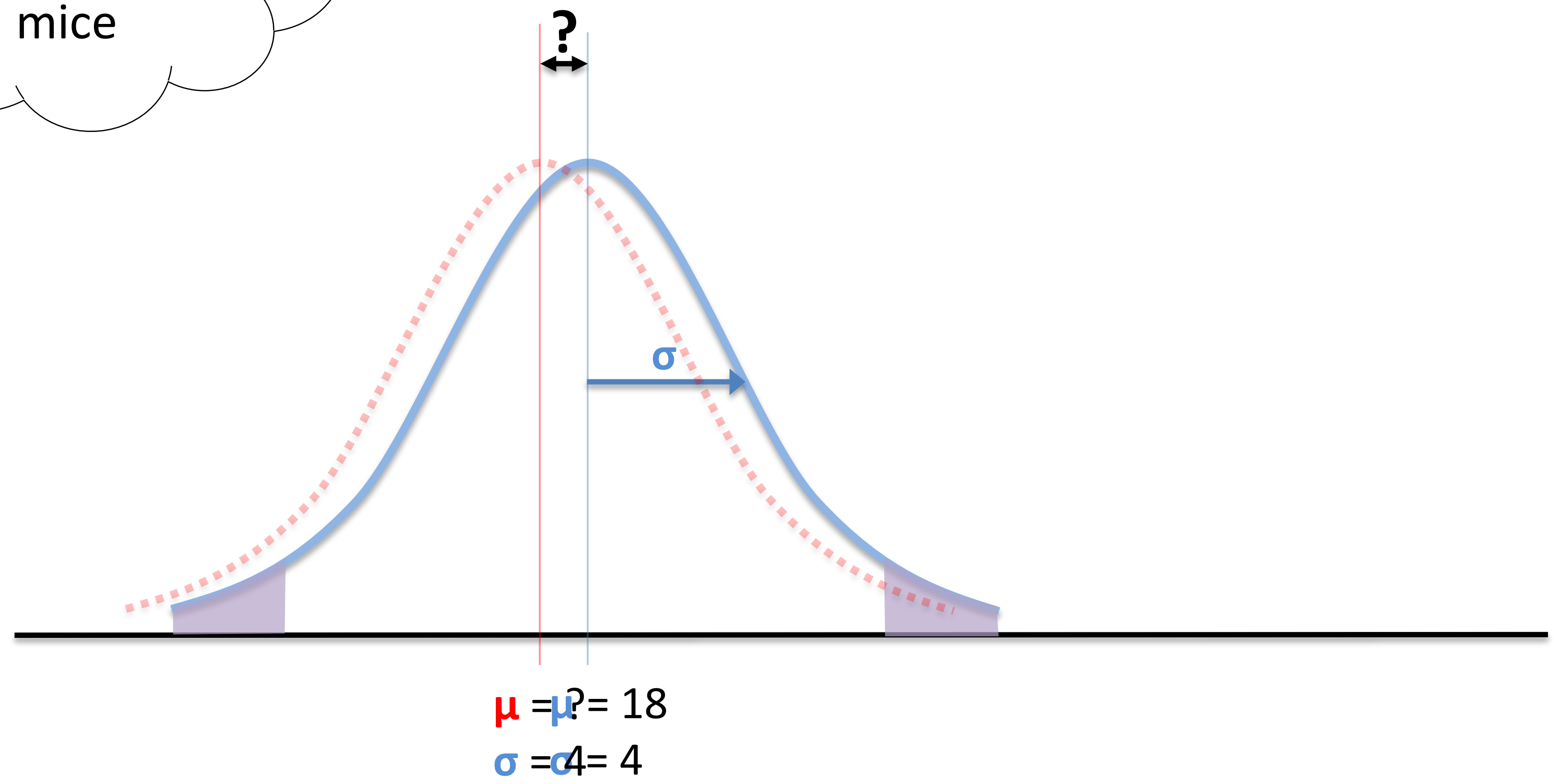
how to make statements about reality


Population of UNTREATED mice
?
σ
μ = 18
σ = 4
μ = ?
σ = 4
Population of TREATED mice


Population of
UNTREATED
mice
?
σ
μ = 18
σ = 4
μ = ?
σ = 4
Population of
TREATED
mice

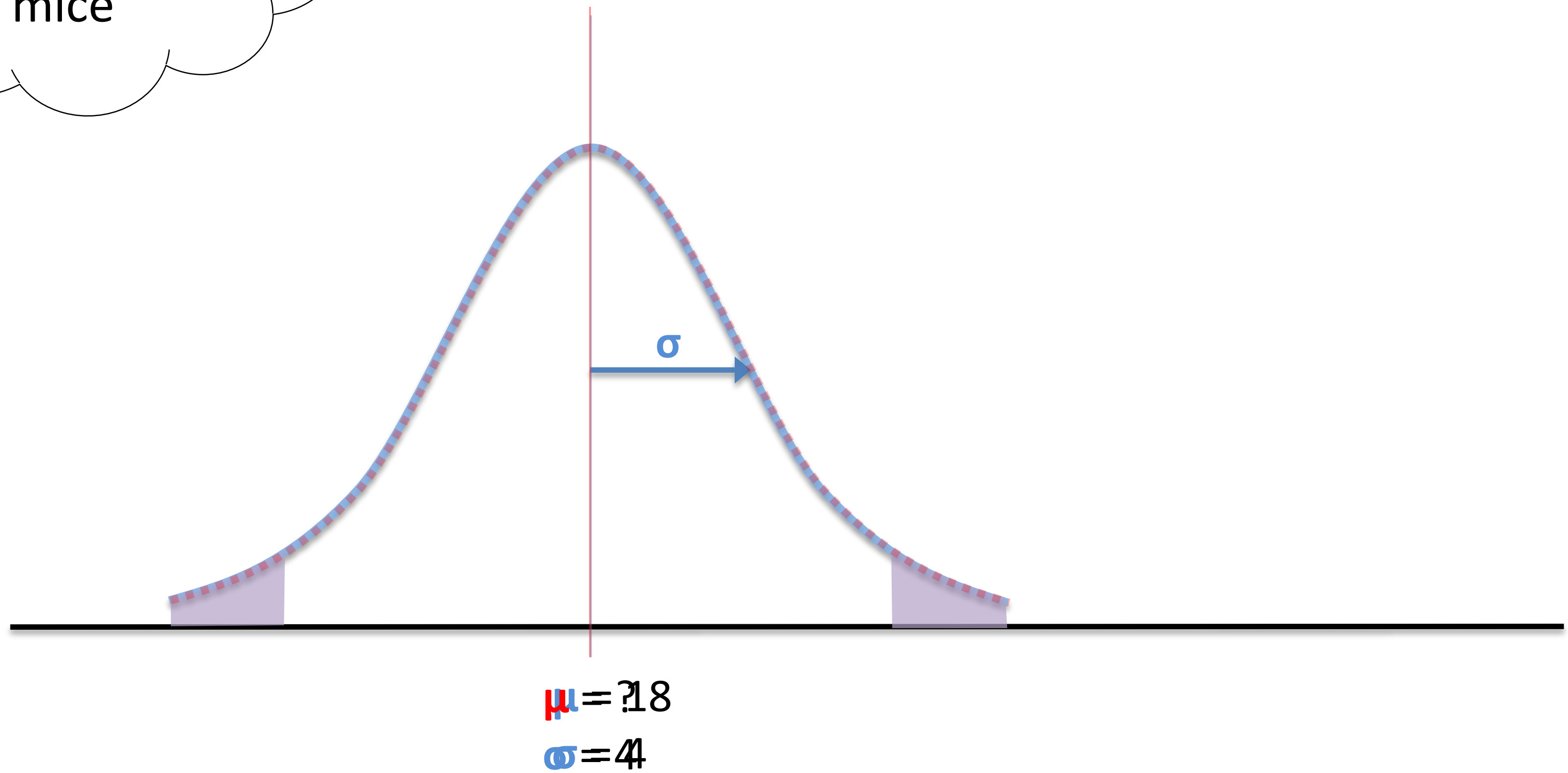

Population of
UNTREATED
mice
?
σ
μ = ?
σ = 4
μ = 18
σ = 4
Population of
TREATED
mice

Population of
UNTREATED
mice



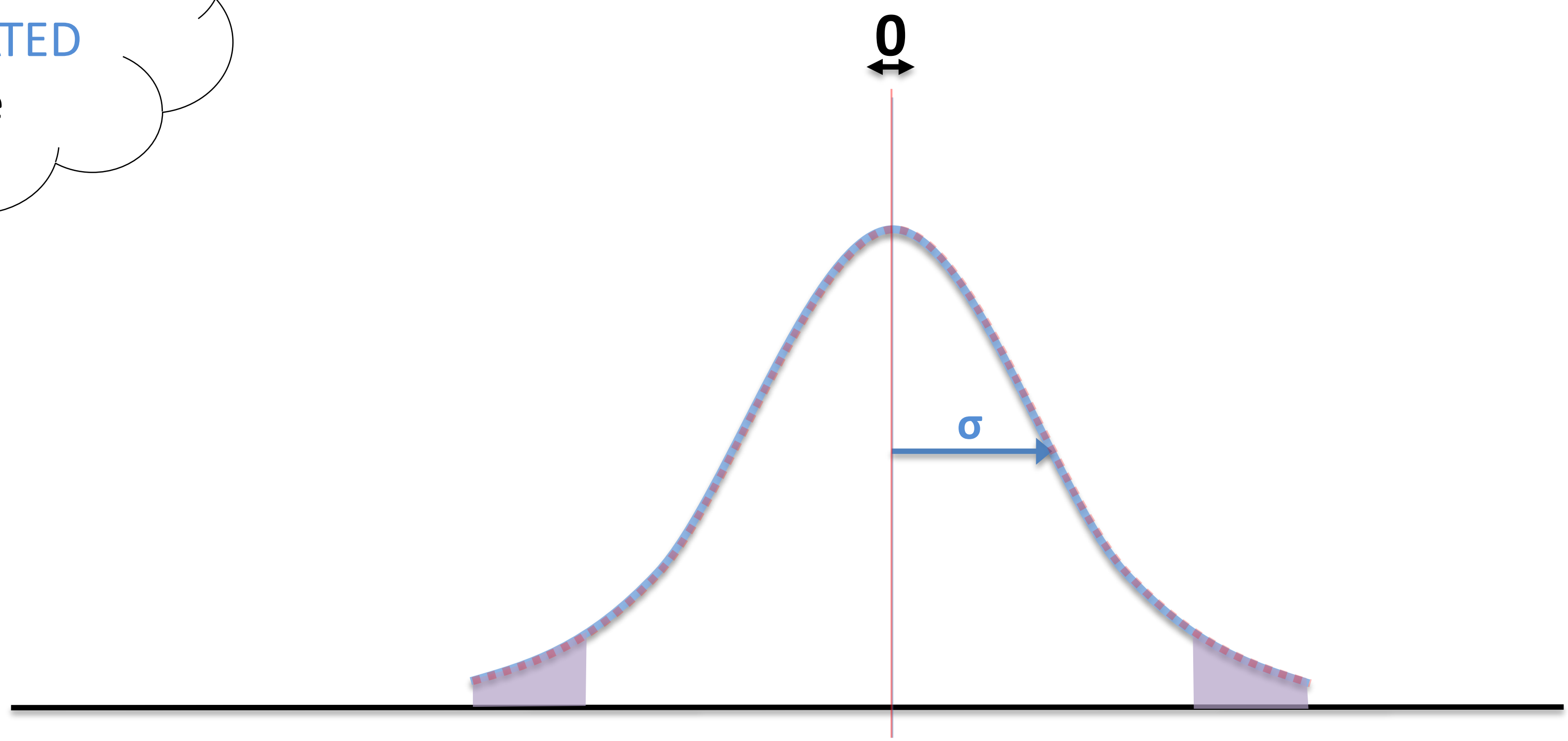
Population of
TREATED
mice

Population of
UNTREATED
mice



Population of
TREATED
mice

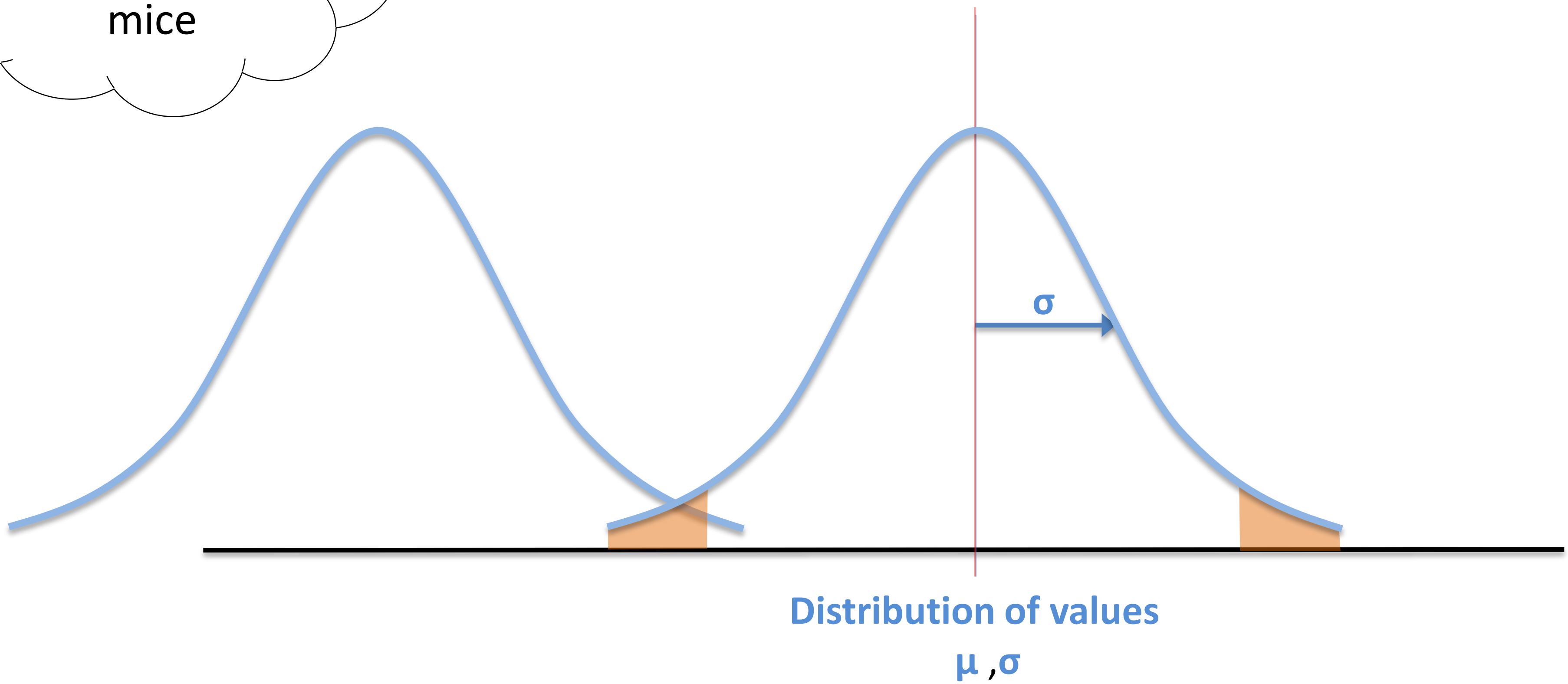
Population of
UNTREATED
mice



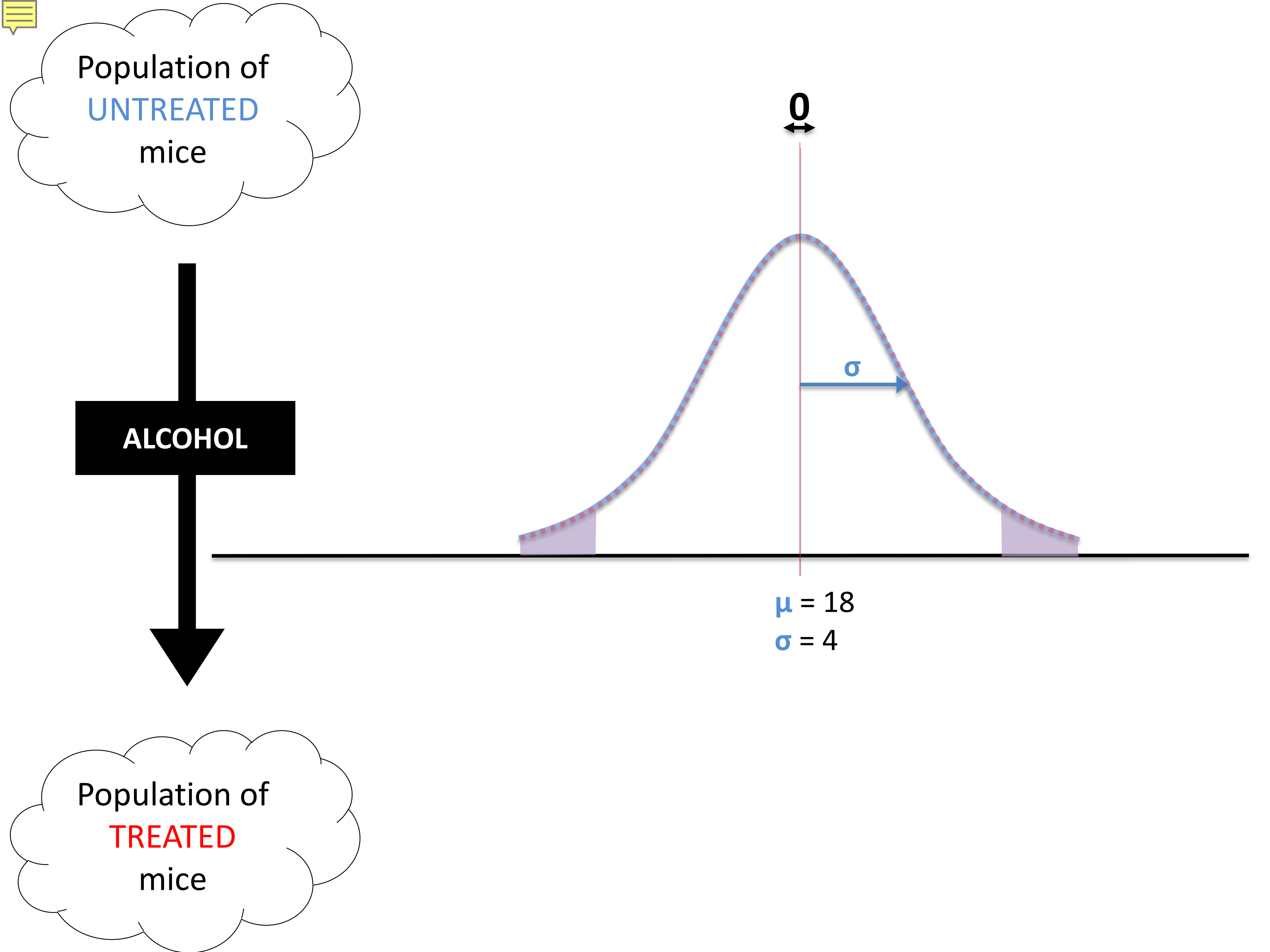
$\mu = 18$
$\sigma = 4$

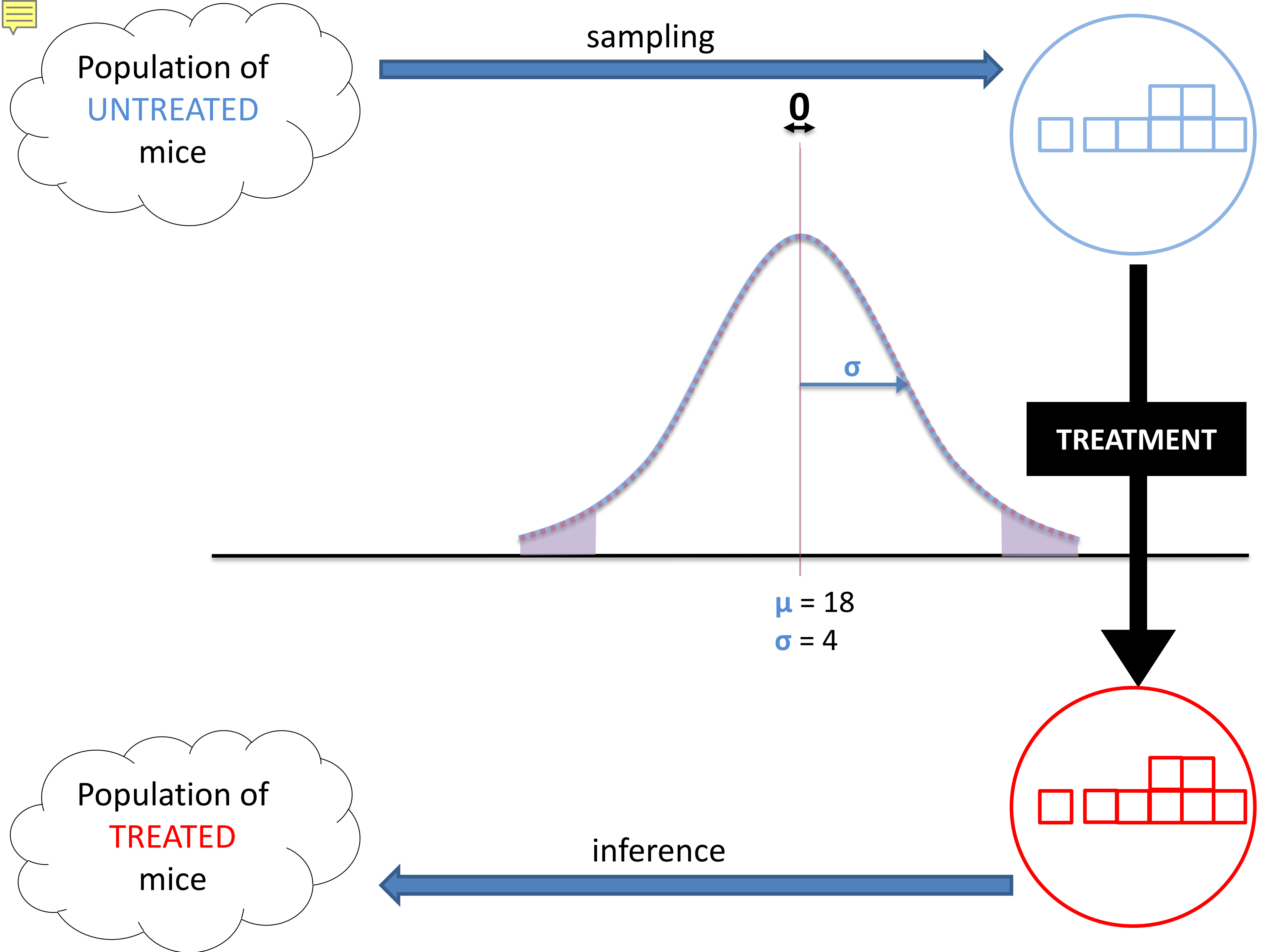
Population of
TREATED
mice

Population of
UNTREATED
mice



Population of
TREATED
mice

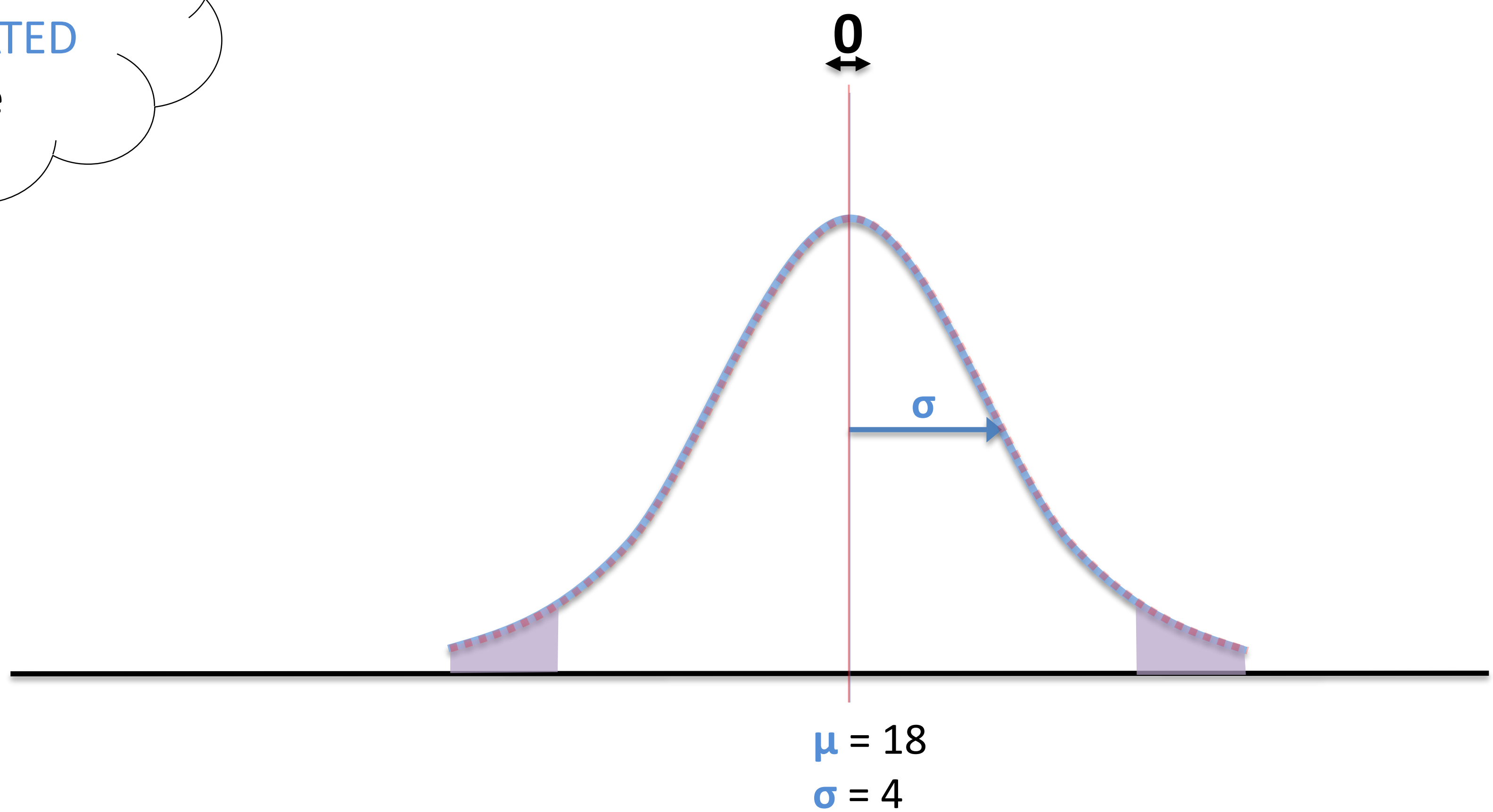

Population of
UNTREATED
mice
ALCOHOL
Population of
TREATED
mice
0
σ
μ = 18
σ = 4


Population of
UNTREATED
mice
sampling
0
σ
TREATMENT
μ = 18
σ = 4
Population of
TREATED
mice
inference

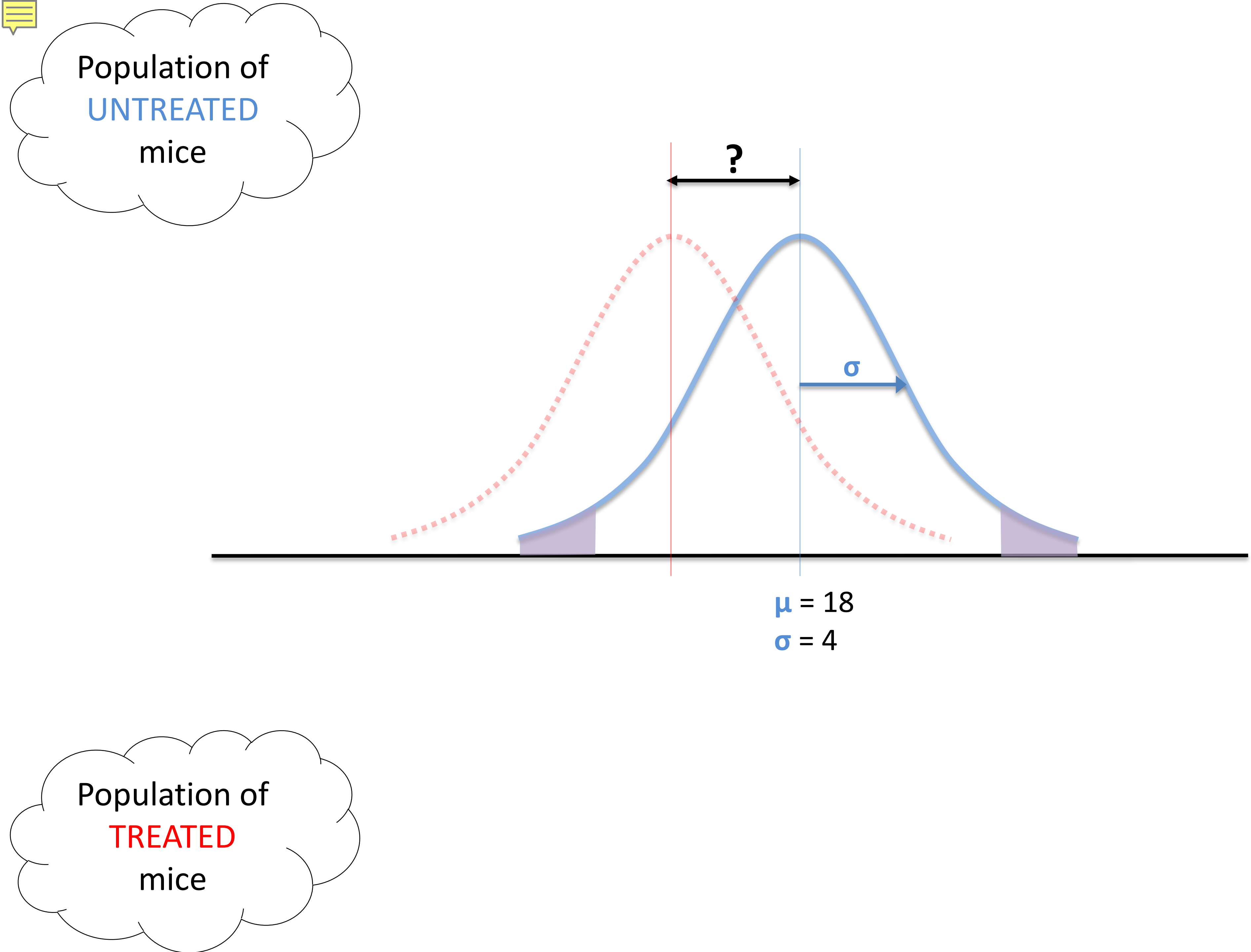
Formulating Hypotheses

making statements

Population of UNTREATED mice



Population of TREATED mice


Population of
UNTREATED
mice
?
σ
μ = 18
σ = 4
Population of
TREATED
mice

Hypotheses Pair

H_0 - Null Hypothesis (no effect)

H_A – Alternative Hypothesis (some effect)

Hypotheses Pair

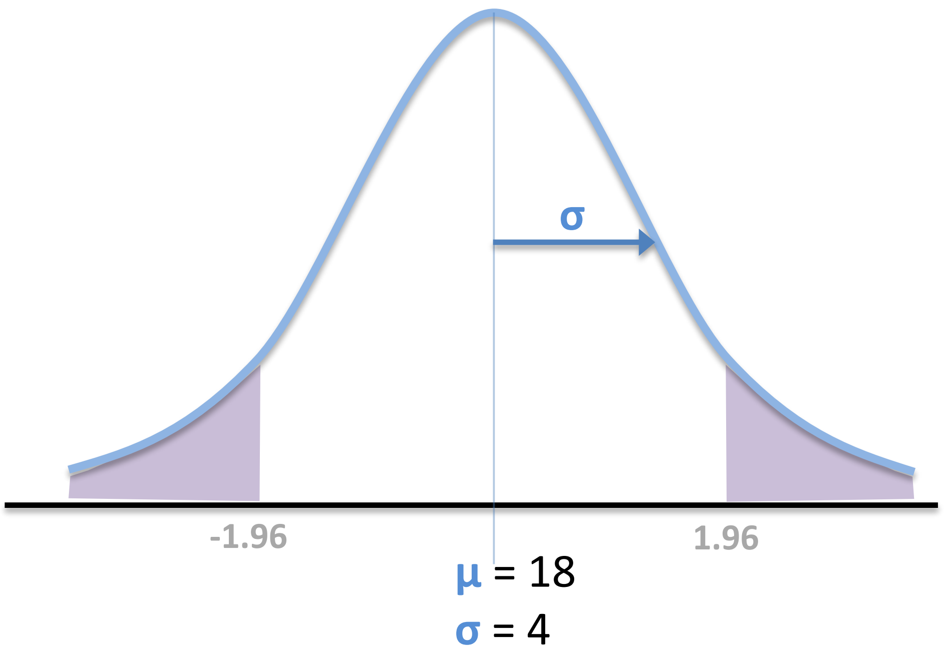
H_0 - Null Hypothesis (no effect)

H_A – Alternative Hypothesis (some effect)

$H_0 : \mu_1 = \mu_2$

$H_A : \mu_1 \neq \mu_2$

Non-Directional



Hypotheses Pair

H_0 - Null Hypothesis (no effect)

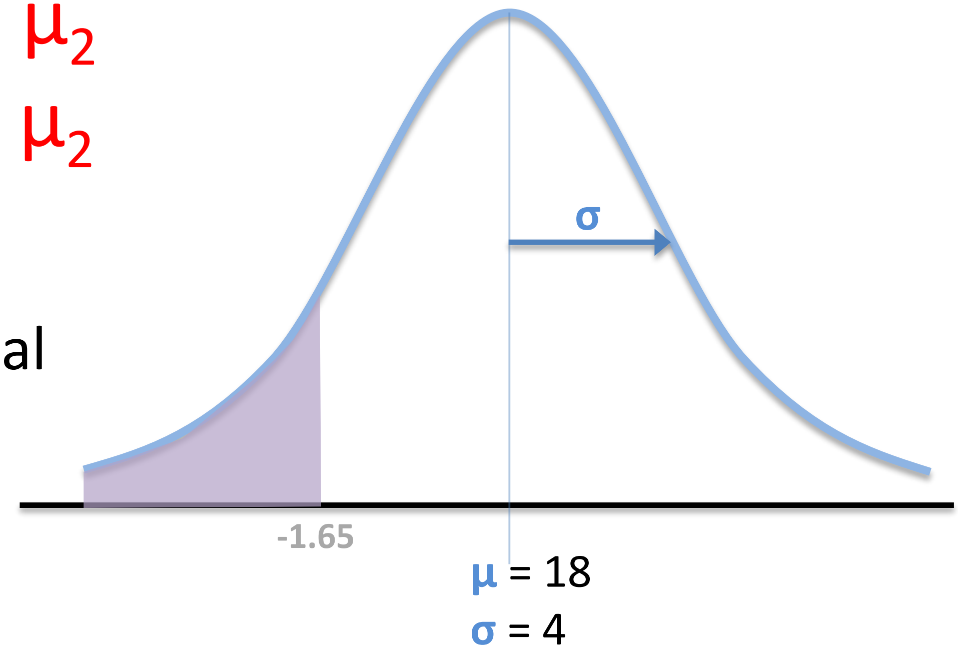
H_A – Alternative Hypothesis (some effect)

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_A : \mu_1 \neq \mu_2 \end{cases}$$

Non-Directional

$$\begin{cases} H_0 : \mu_1 \leq \mu_2 \\ H_A : \mu_1 > \mu_2 \end{cases}$$

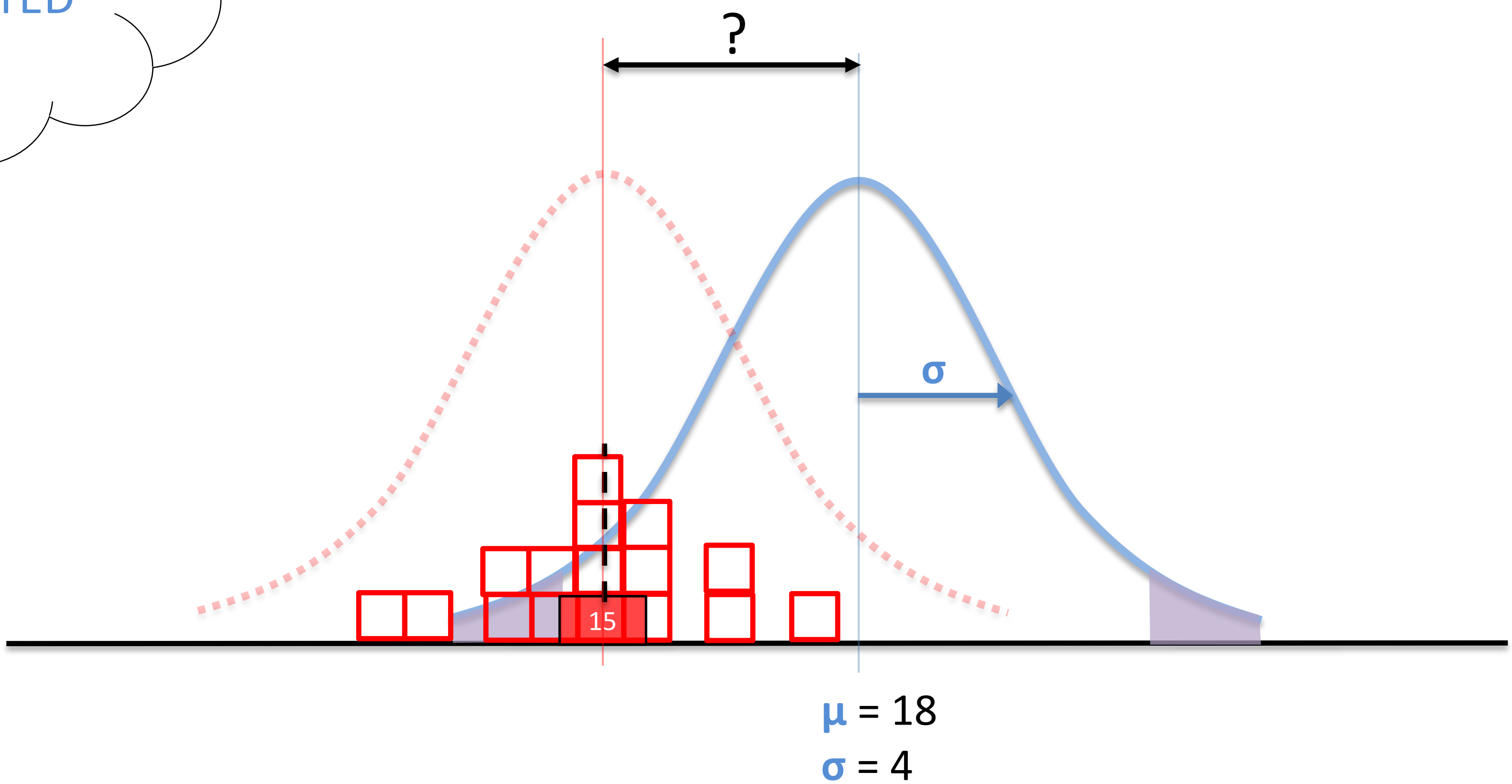
Directional



Testing Hypotheses

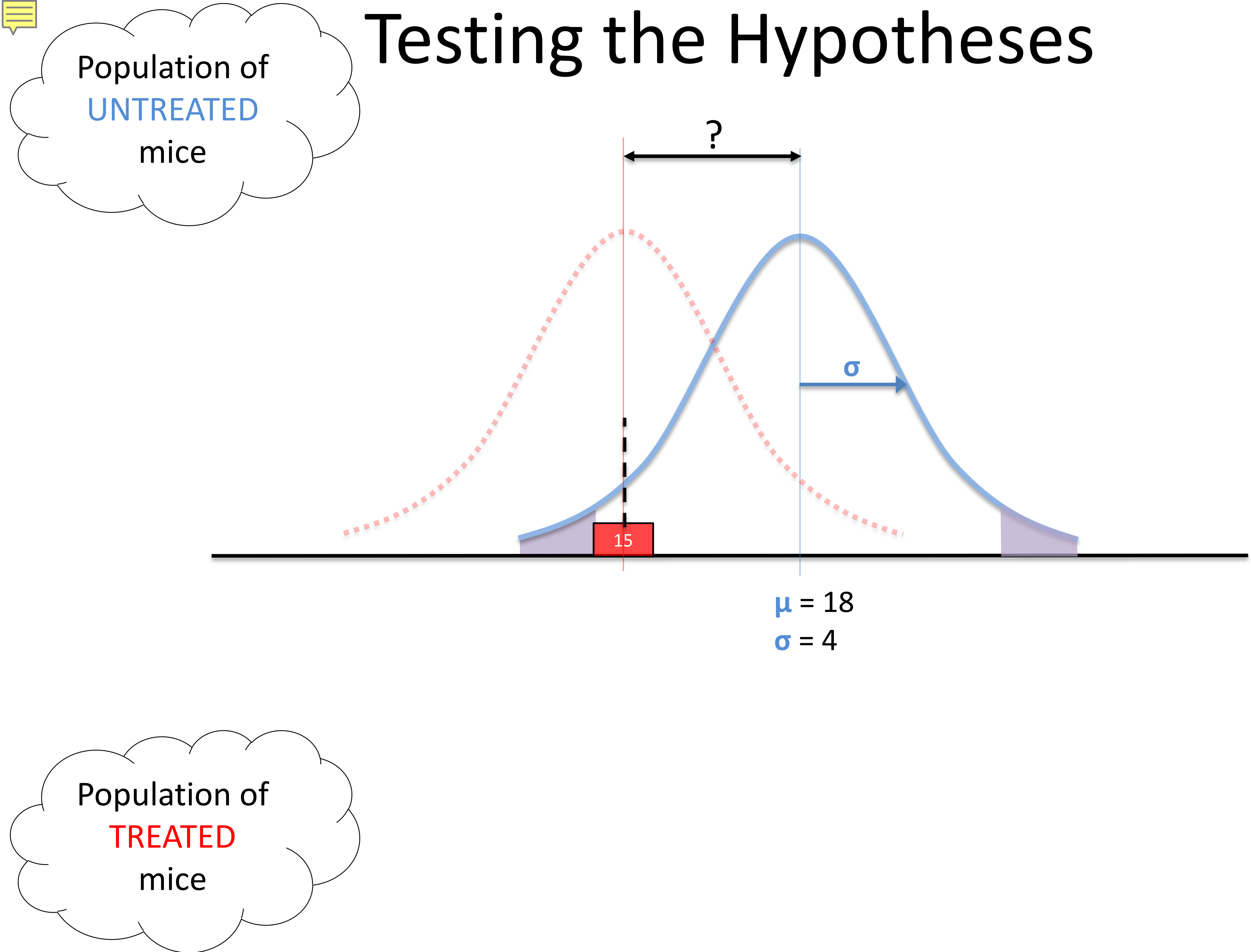
Testing the Hypotheses

Population of
UNTREATED
mice



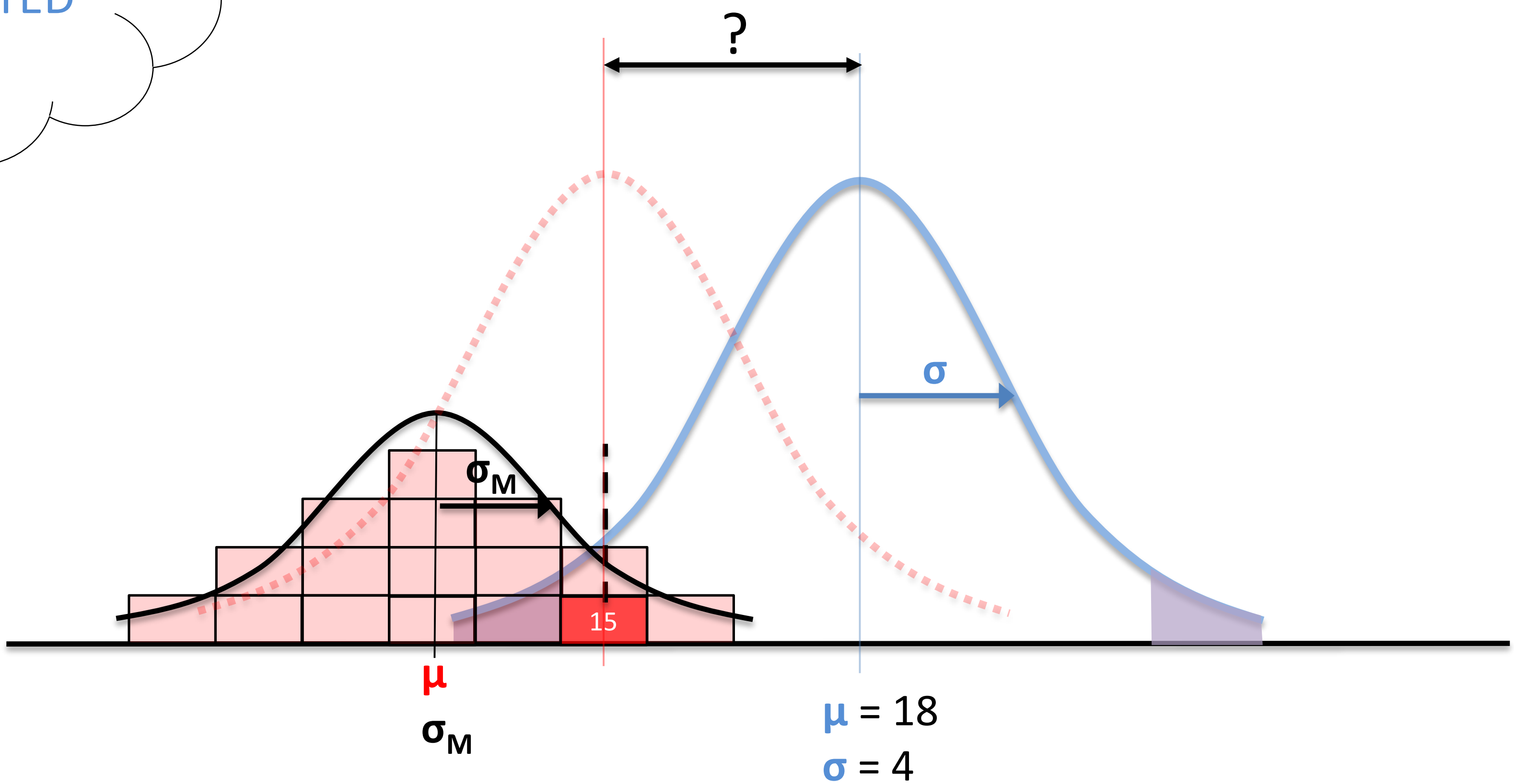
Population of
TREATED
mice

Testing the Hypotheses



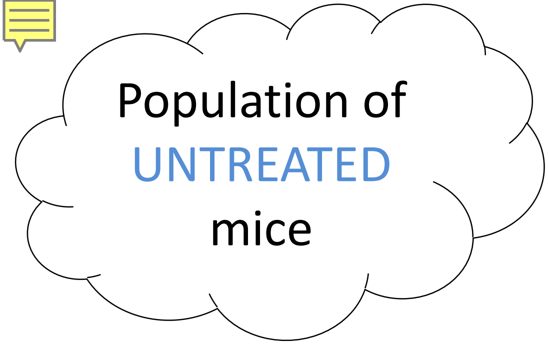
Testing the Hypotheses

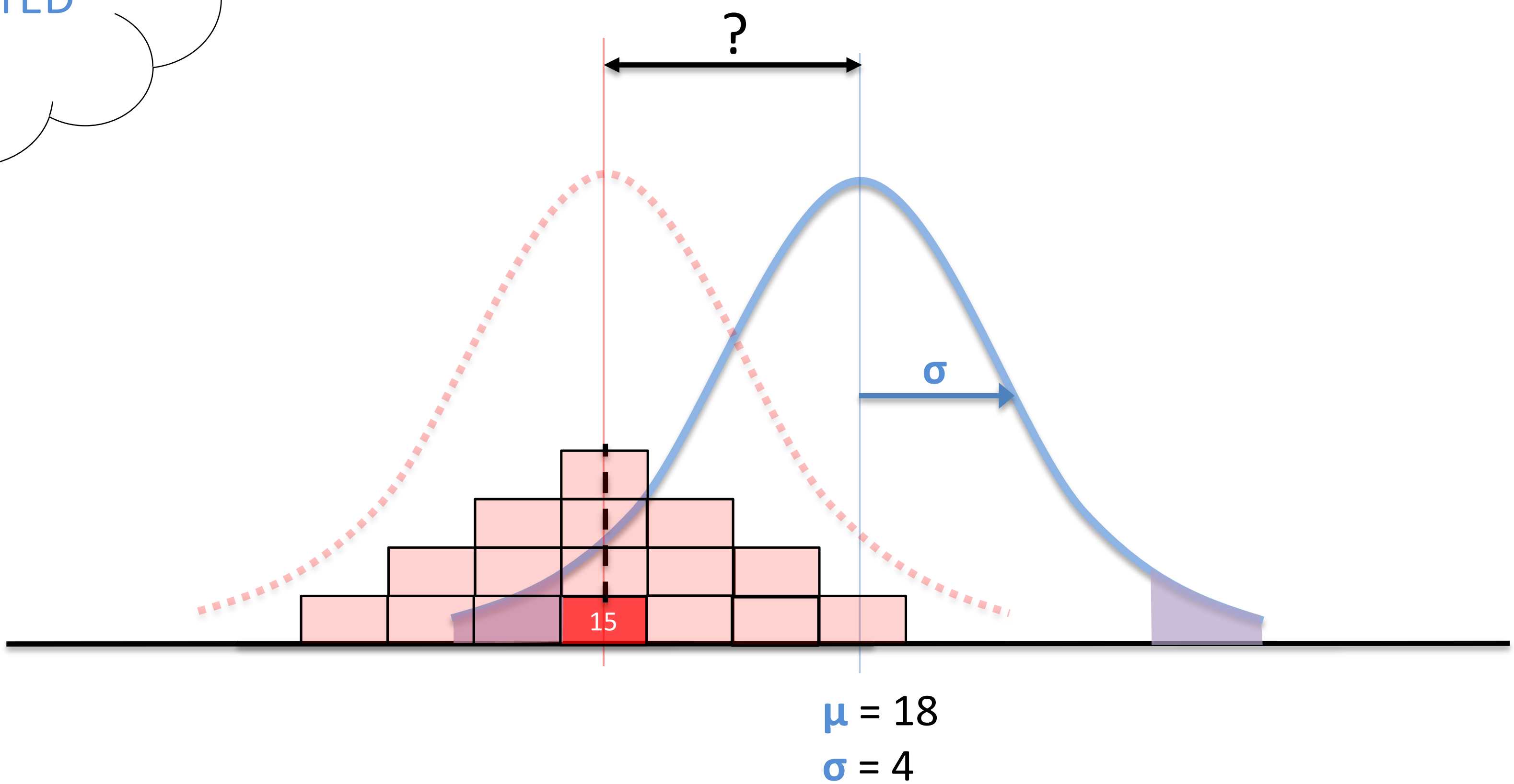
Population of
UNTREATED
mice

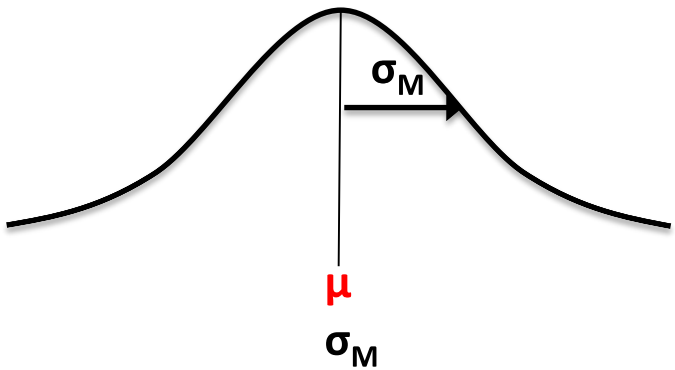


Population of
TREATED
mice

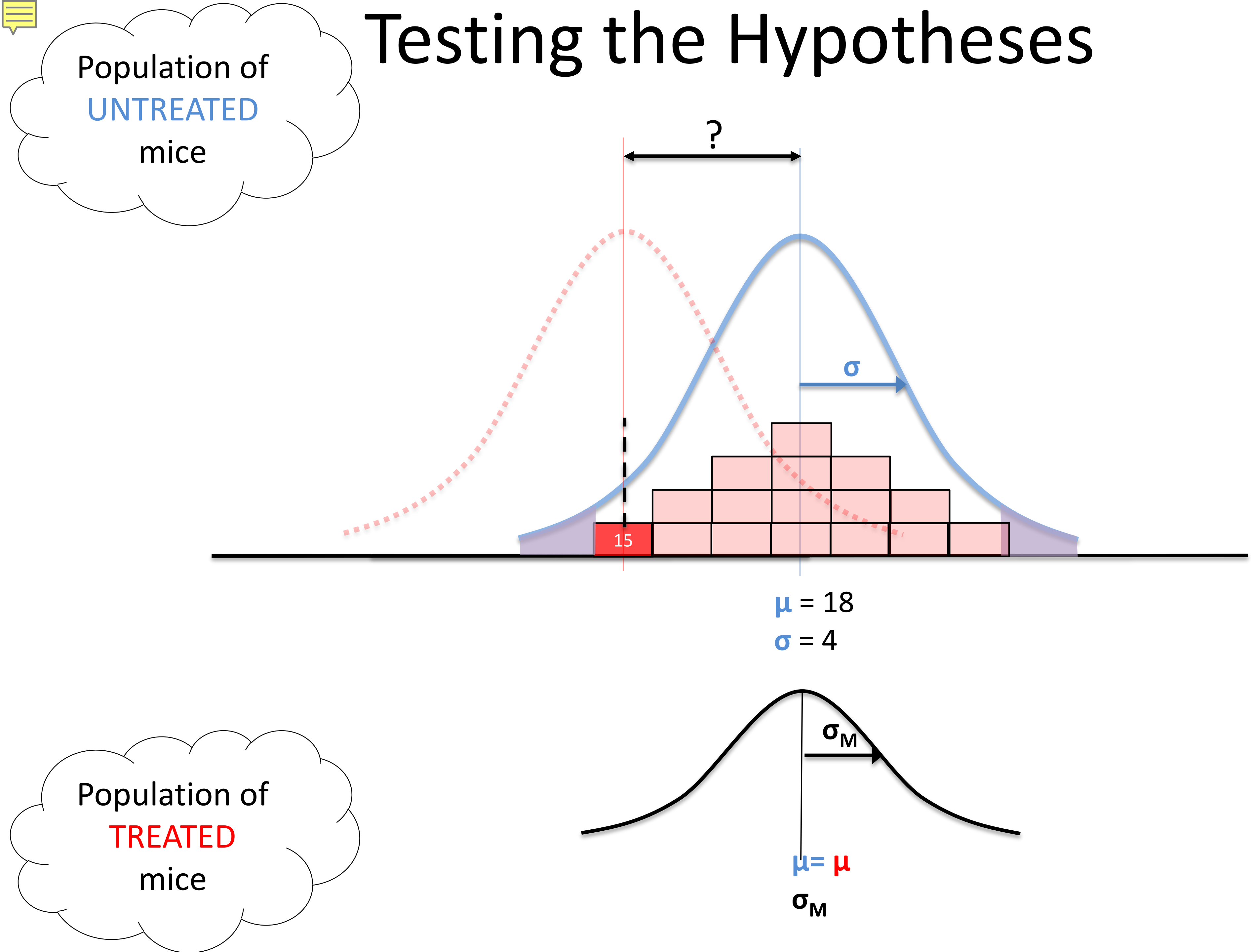
Testing the Hypotheses



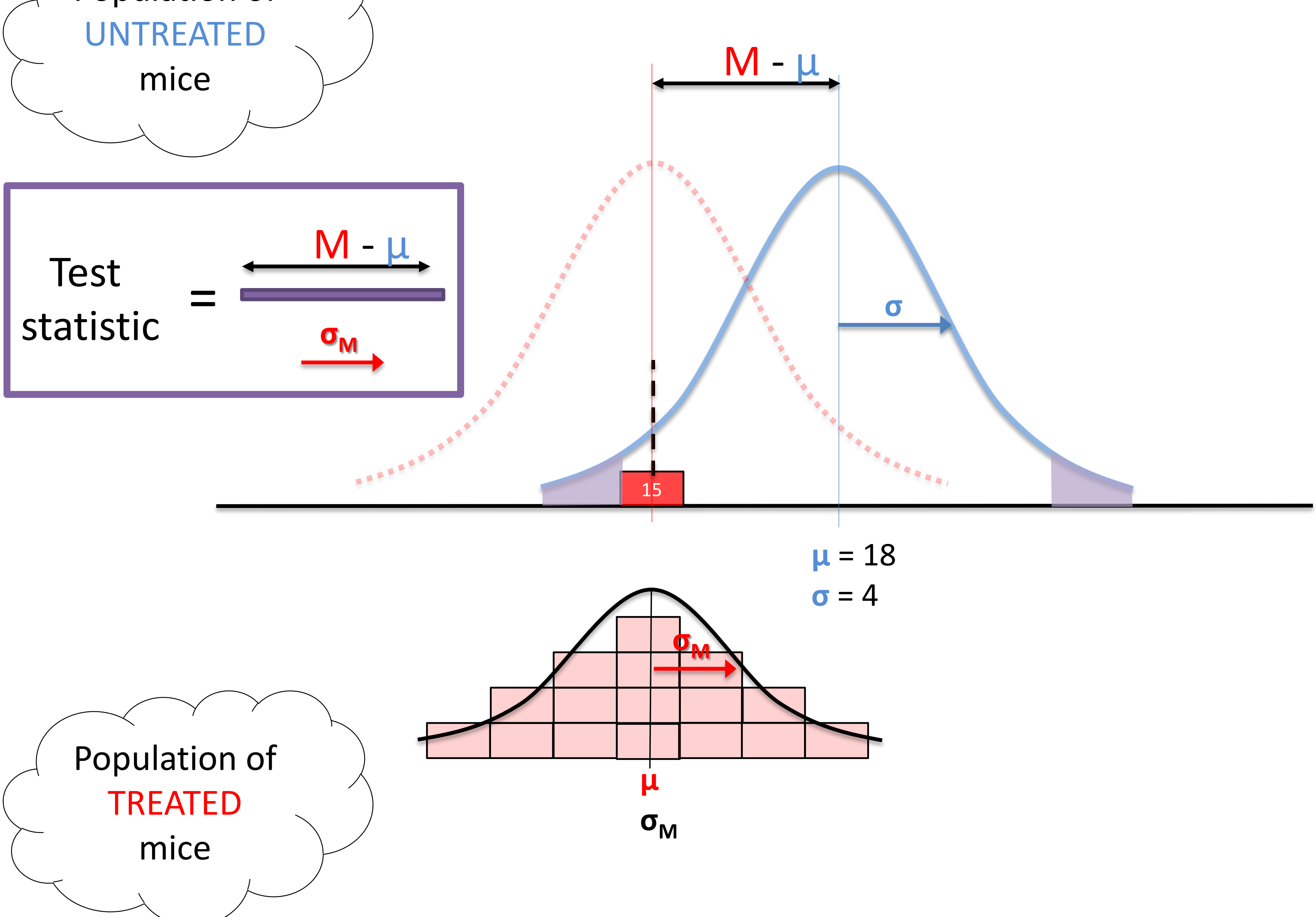




Testing the Hypotheses



Testing the Hypotheses



Test Statistics

$$\frac{M - \mu}{\sigma_M} = \text{Test statistic} = \frac{\text{Observed Difference}}{\text{Difference due to chance}}$$

Conducting Z-test

Population of
UNTREATED
mice

$$\frac{M - \mu}{\sigma_M} = \text{Test statistic}$$

$H_0 : \mu_1 = \mu_2$

$H_A : \mu_1 \neq \mu_2$

Population of
TREATED
mice

Conducting Z-test

Population of UNTREATED mice

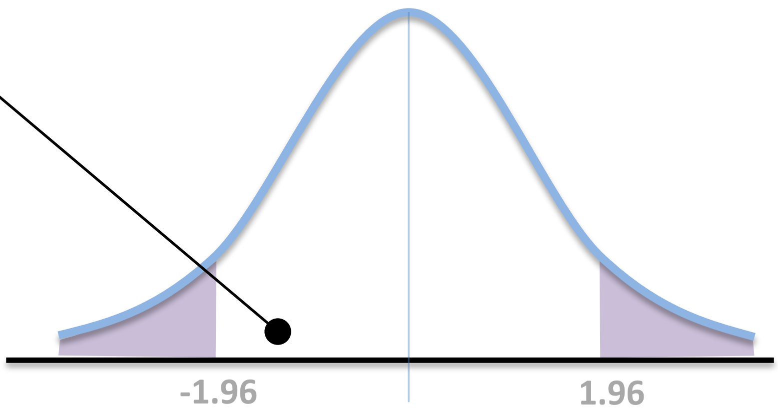
$$\frac{M - \mu}{\sigma_M} = \text{Test statistic}$$

$$H_0 : \mu_1 = \mu_2$$

$$\cancel{H_A : \mu_1 \neq \mu_2}$$

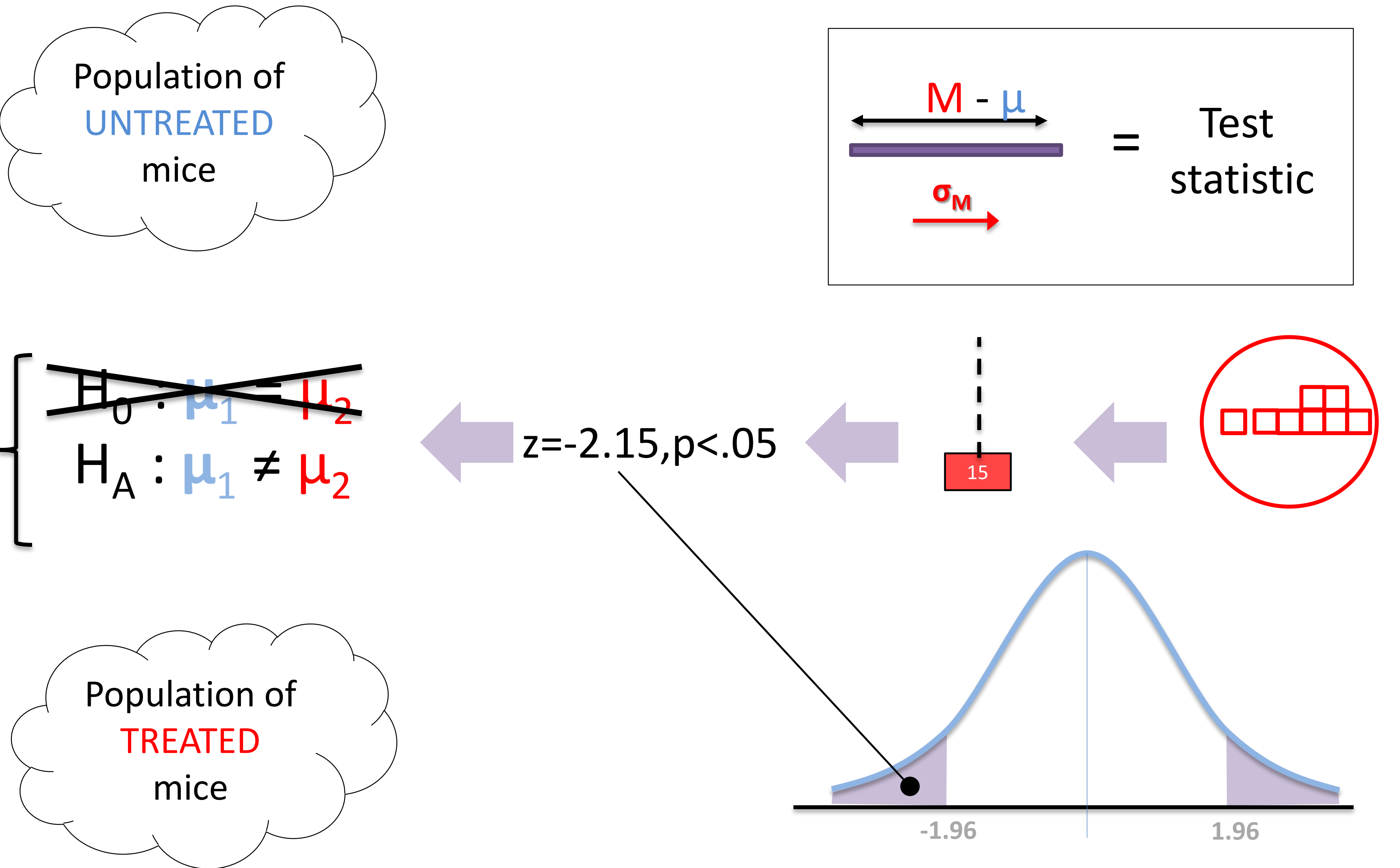
z=-1.75,p>.05

15



Population of TREATED mice

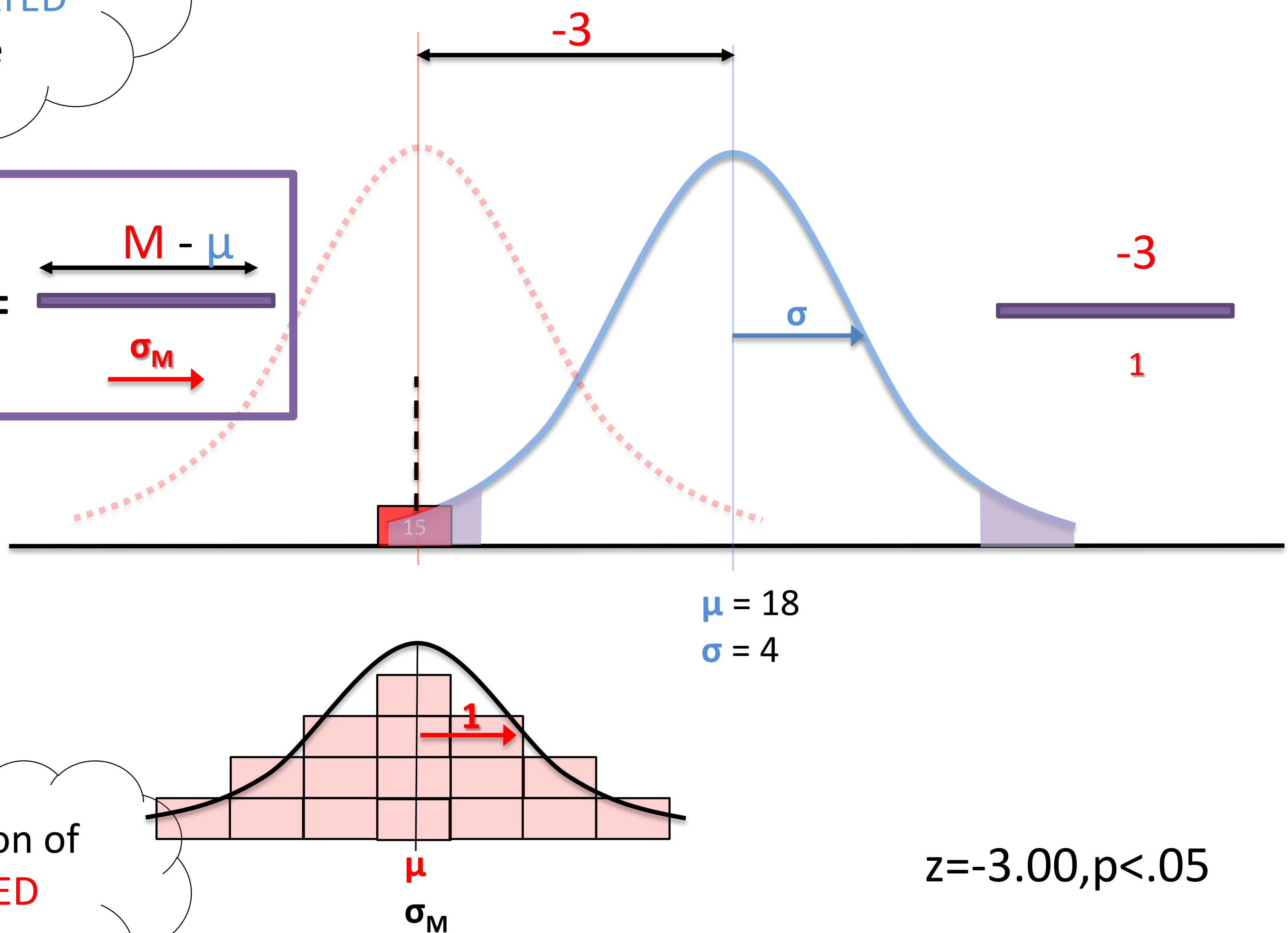
Conducting Z-test



Testing the Hypotheses

Population of
UNTREATED
mice

$$\text{Test statistic} = \frac{M - \mu}{\sigma_M}$$



$$\frac{-3}{1}$$

$\mu = 18$
$\sigma = 4$

Population of
TREATED
mice

$z=-3.00, p<.05$

Report your Results

You are not typically told *explicitly* what test was used, what alpha was used, or whether the null was rejected or retained.

- "The treatment showed a significant effect on IQ scores, $z = 2.5$, $p < .05$."
- "The treatment did not have a significant effect on IQ scores, $z = 1.5$, $p > 0.05$." OR "There was no evidence of an effect on IQ, $z = 1.5$, *ns*."

Null or *Alternative hypotheses do not need to be mentioned in formal reporting!*

If you are using software and given an *exact p*-value, report the EXACT value, your phrasing should indicate whether the results were significant.

More about Hypothesis Testing

Error and Uncertainty

- When we use a small sample to make judgments about an entire population, errors can be made
  - When the sample does not represent the population

- **Type I Errors:** occurs when a researcher rejects a null hypothesis that is true
  - Conclude there is a treatment effect when there is not
- **Type II Errors:** occurs when a researcher fails to reject a null (supports a null) that is actually false.
  - Concludes there is no treatment effect when there really is

Type I Errors

- **Type I Error:** Reject a null that is really true
  - I found support for my study guide improving test scores, but the study guide did not, in fact, change test scores.
- Type I errors occur when we select an extreme sample by chance (because of sampling error).
  - The probability of a Type I error is equal to the alpha level
  - We select an alpha level to reflect how much risk we're willing to take
- If the null is true and there was no treatment effect (i.e. no shift in the population mean post-treatment), we still *could* get an extreme statistical value, it's just rare.

Type II Errors

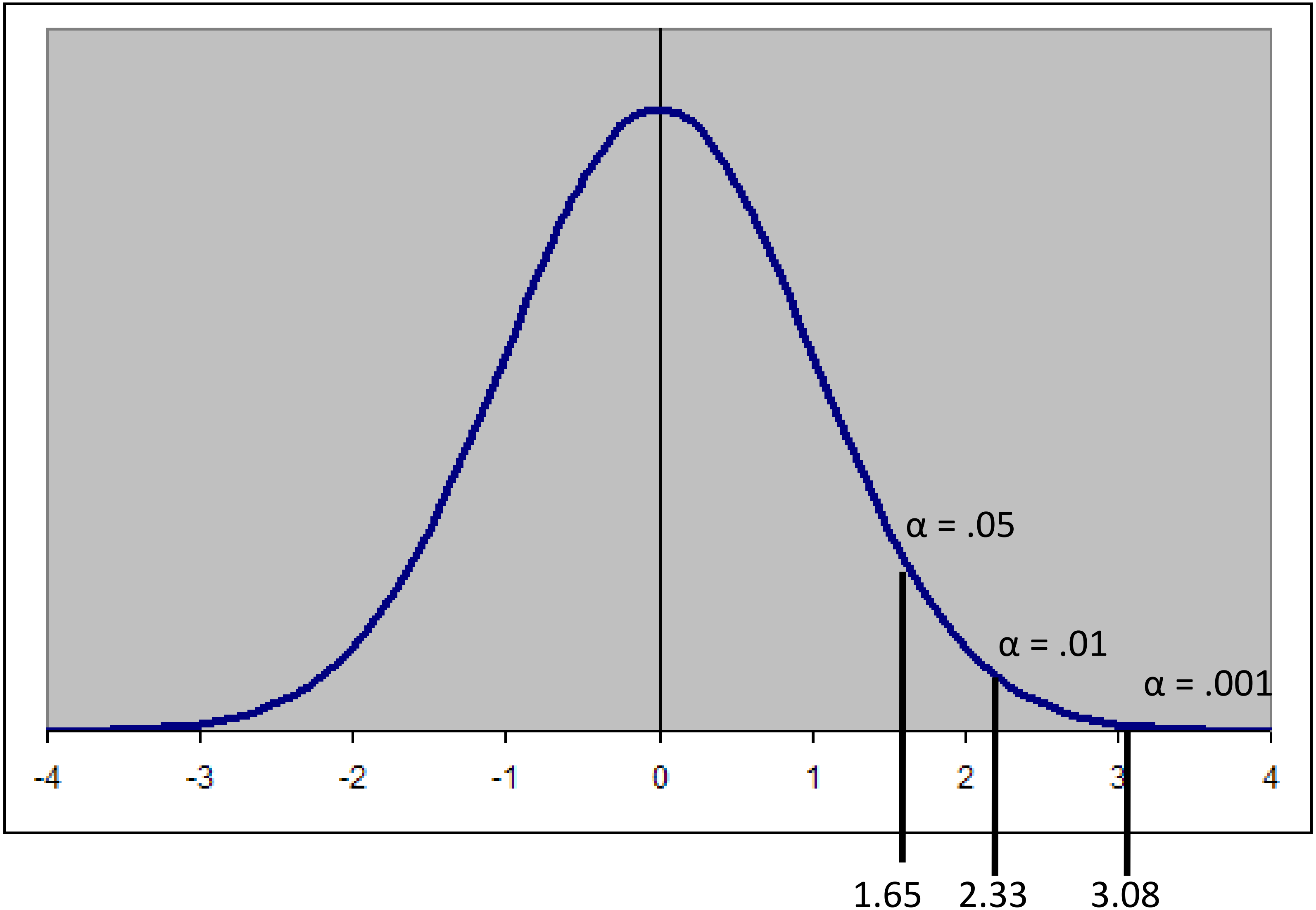
- **<u>Type II Errors</u>**: fail to reject a null that is, in fact, false.
  - I did not find support for my study guides improving test scores, but they really do work!

- Often happens when treatment effect is small
  - Treatment did make a difference, but not enough to push the stat into the critical region.
- Difficult, if not impossible to determine exact probability – it depends on MANY things
- Signified by β (“beta”)

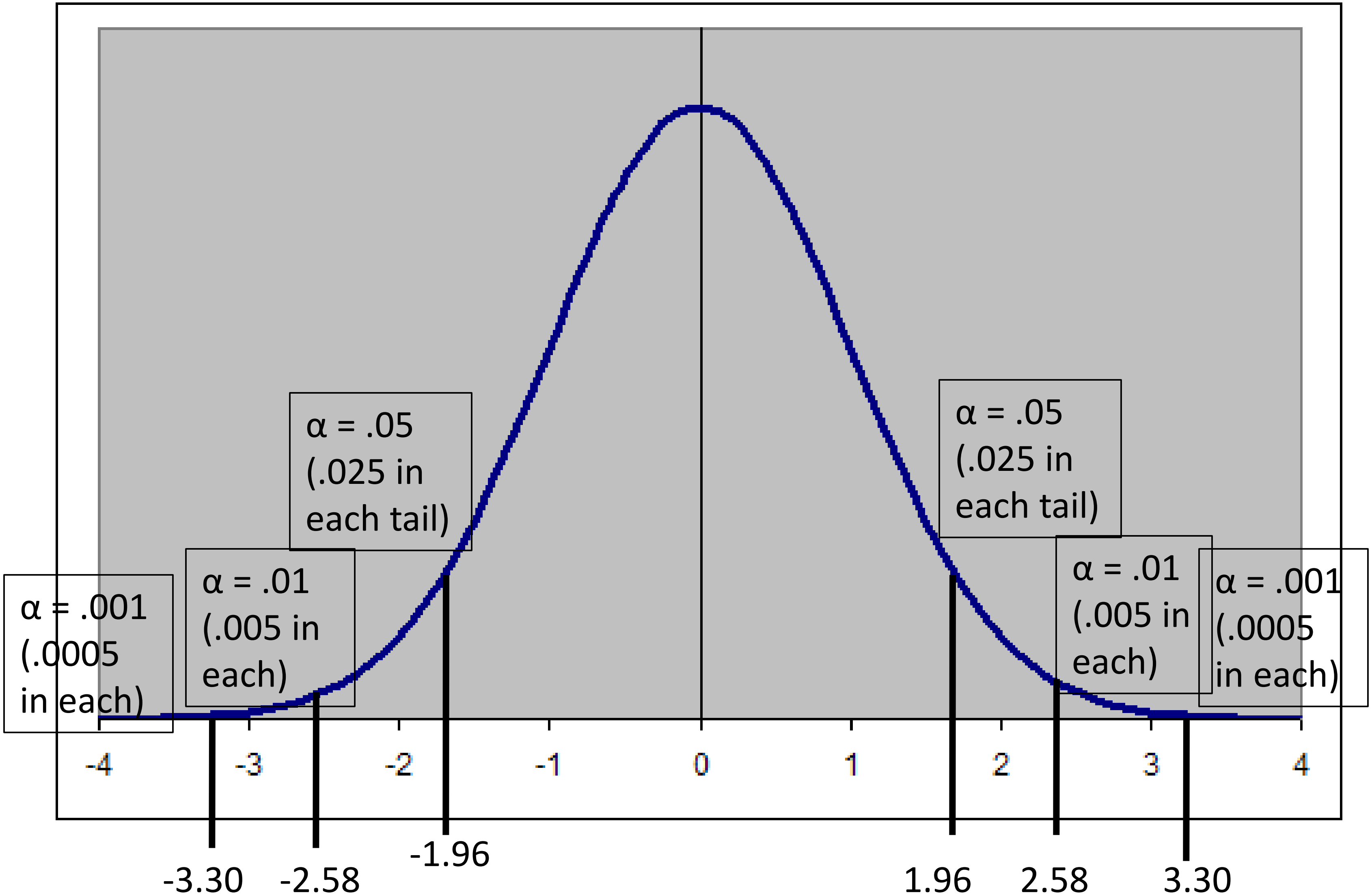
Errors Hypothesis Testing

		Reality	
Research Results		Treatment Effects DO NOT Exist **H_0 is TRUE**	Treatment Effects DO Exist **H_0 is FALSE**
	Treatment effects were not found **H_0 was *retained***	**Correct**	Type II Error p=β
	Treatment effects were found **H_0 was *rejected***	Type I Error p=α	**Correct** p=1 –β

Common Critical Values for One-Tailed *z*-tests



Common Critical Values for Two-Tailed *z*-tests



Assumptions of the *z*-test

- *Random Sampling*- the sample must be representative.

- *Independent observations* – one observation must have no effect on another, there must be no predictable relationship between them. Usually satisfied by random sampling.

- *Variability is unchanged by treatment*– Computation is based on standard error, calculated from the population's original variance. This must remain unchanged since *we cannot measure* the treated population's variance.

- *Sample means are normally distributed*– we use the Unit Normal Table to calculate probability—this only works on normal data!

Concerns with Hypothesis Testing

- Focus is on the data, not on the hypothesis
  - Significant results indicate that a particular *sample mean* is *unlikely if the null is true.*
  - *Does not tell us how likely it is that the null (or alternative) is true.*
  - ***Rejection of the null given alpha = .05 does <u>NOT</u> mean that there is a 5% chance that the null is true.***
  - ***<u>**It means there is a 5% chance of selecting a sample with this statistical value assuming the null is true. **</u>***
    - Our probabilities operate under assumption of the null (no shift in parameter values after treatment). This is where we test!

Effect Size

- A "significant" effect does NOT mean a substantial effect.
  - We are making a relative comparison: How great is the treatment effect **relative** to the standard error?

- With any significant effect, it is recommended that you report the **effect size:** intended to provide a measure of the absolute magnitude of a treatment effect, independent of sample size being used.

  - Remember, n is used in computation of z (and other later stats, too). Larger n → more likely to reject the null even with small effects.

Cohen's *d*

- Cohen's *d* is a relatively simple and direct effect size measure.
- Cohen's *d* = $\dfrac{\text{mean difference}}{\text{standard deviation}} = \dfrac{\mu_{Treatment} - \mu_{NoTreatment}}{\sigma}$

- We can't measure the population mean after treatment, so we *estimate* using the sample mean.

- Estimated Cohen's *d* = $\dfrac{M_{Treatment} - \mu_{NoTreatment}}{\sigma}$

- Size of Effects (these criteria are constant, even for other statitical effect size measures)
  - *d* = 0.2 = Small effect
  - *d* = 0.5 = Medium effect
  - *d* = 0.8 = Large effect

Power

- The **Power** of a statistical test is the probability that the test will correctly reject a false null hypothesis.
  - Related to the probability of a Type II error (β)
  - $1 - \beta$ is the measure of Power
  - I.E. if power is 75%, the probability of a Type II error is 25%
  - Power is calculated BEFORE a study is conducted.

  - Please read this section in your book carefully, it is difficult to explain through these slides but your book does a good job, and it's an important concept!