

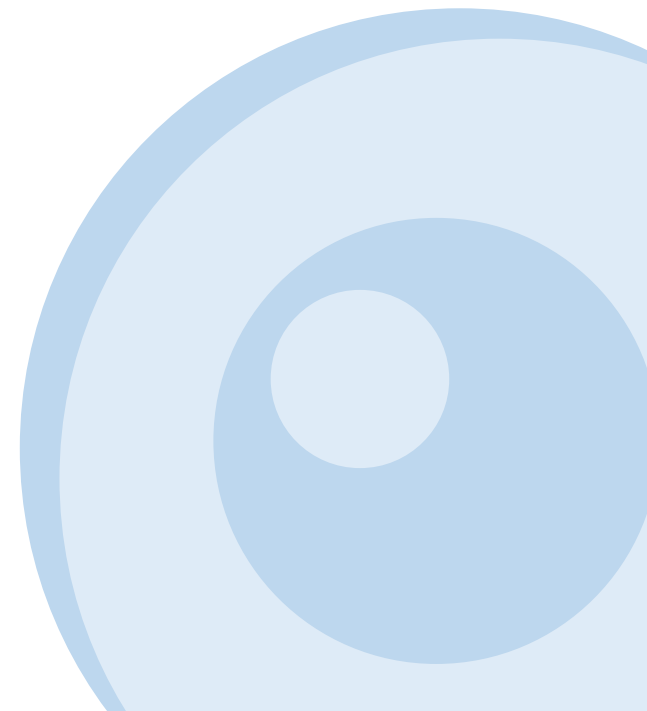


PSYC 193R:

Data Analysis & Visualization with R

Lecture 4: Hypothesis Testing

Jonas S.-H. Lau [silau -et- ucsd.edu]
psyc193r.ucsd.edu



Agenda

- Hypothesis Testing
- Confidence Interval
- Effect Size

Probability / Quantile Conversion

- `pnorm()`
- `qnorm()`

Law of Large Numbers:

- Get multiple estimates for your parameter

Central Limit Theorem:

- Sample size $\uparrow$ Dispersion $\downarrow$
- Get sample size >30 if population is not normal

A typical research question

- We know some properties of the population (e.g., IQ distribution)
- We would like to know if a manipulation would change (raise / decrease) people's scores *reliably*.

Above and beyond
reasonable
fluctuations

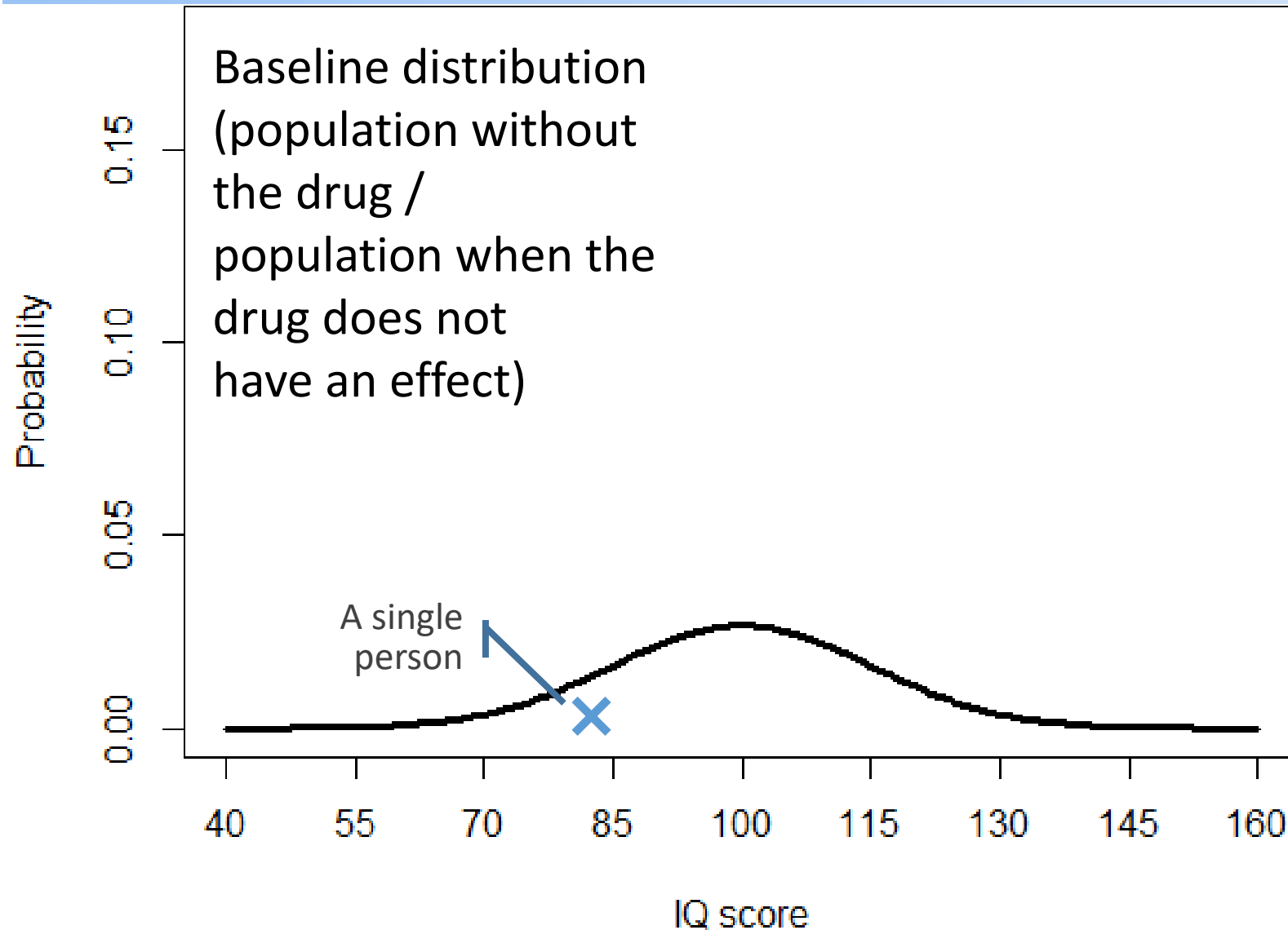


The Basic Logic

- You assume that the treatment does not have an effect
- Construct a *sampling distribution of the baseline population*
- Compare the *mean of the treatment group* with the baseline sampling distribution
- Case 1: If the mean is not different from the majority of the baseline sampling distribution, retain the H_0
 - You cannot conclude that there is a treatment effect
(The population with treatment is not different from the baseline population above and beyond reasonable fluctuations)
- Case 2: If the mean lies in the rejection region (extreme ends), reject the H_0
 - You conclude that there is a treatment effect

NEVER SAY:
There is no treatment effect

Population distribution of IQ scores

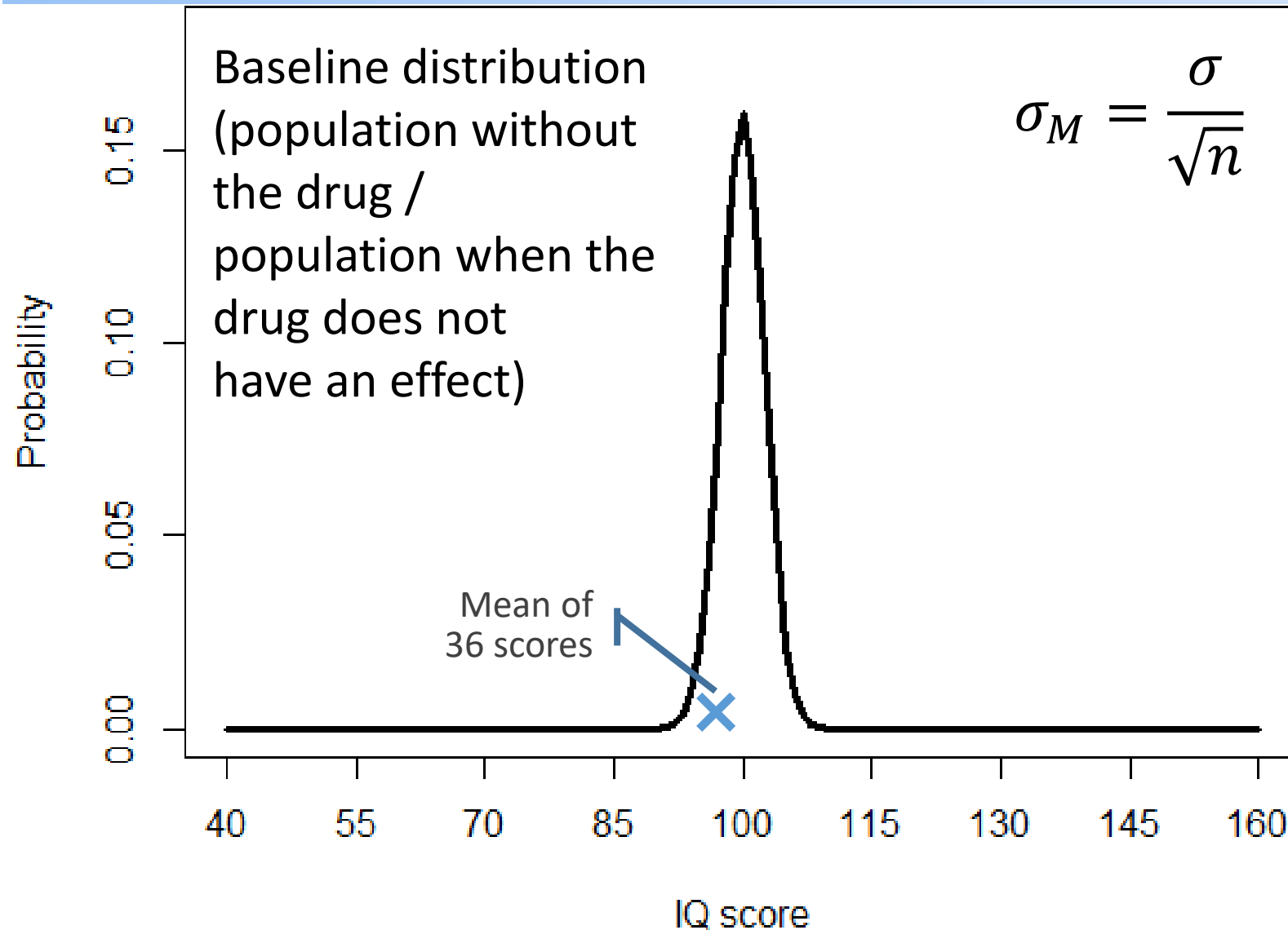


A Sample
Representing Population
with Treatment (a drug)

VS

Test score
 $n = 36$

Sampling distribution ($n = 36$) of IQ scores

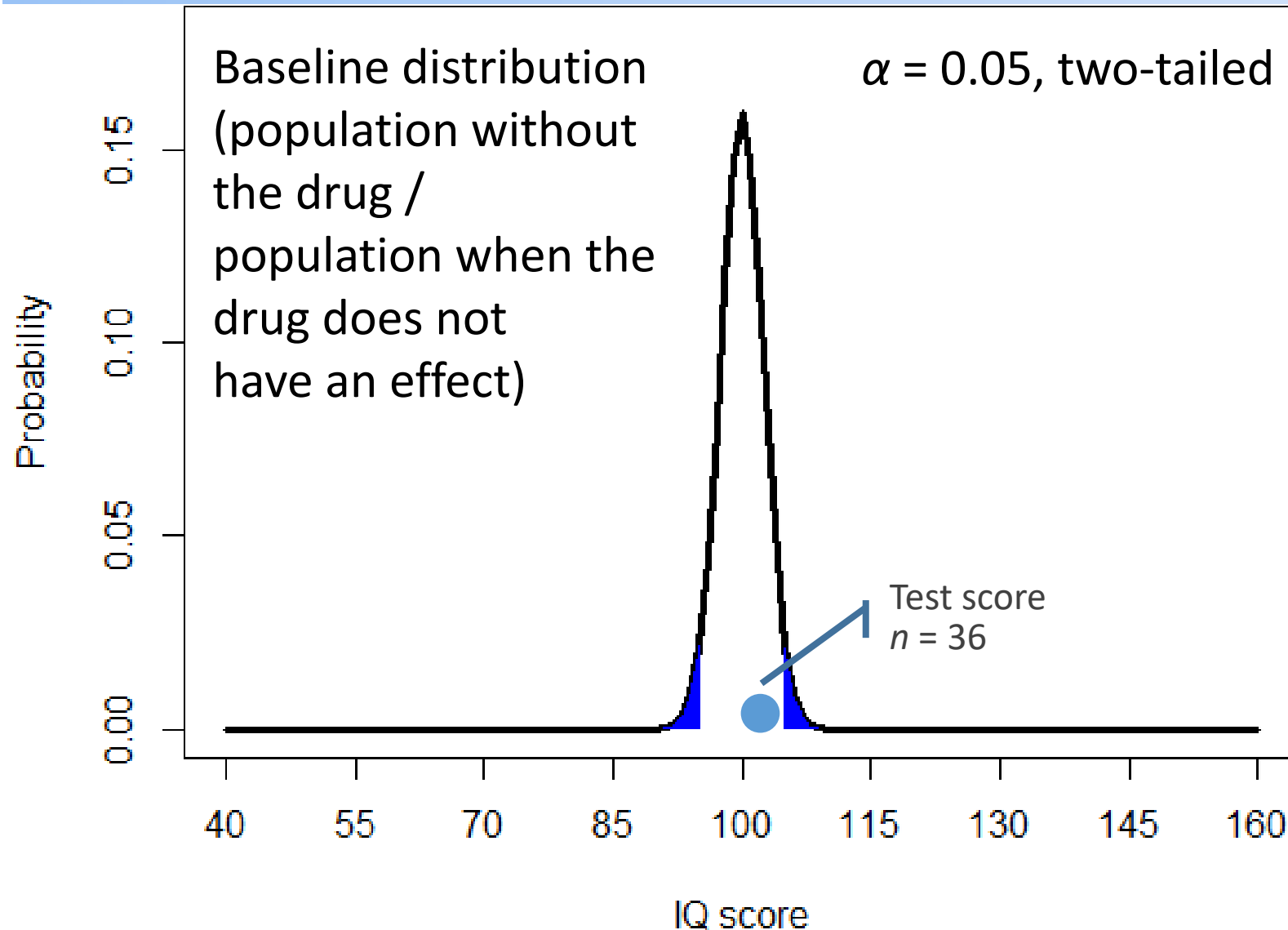


A Sample
Representing Population
with Treatment (a drug)

VS

Test score
 $n = 36$

Sampling distribution ($n = 36$) of IQ scores (case 1)

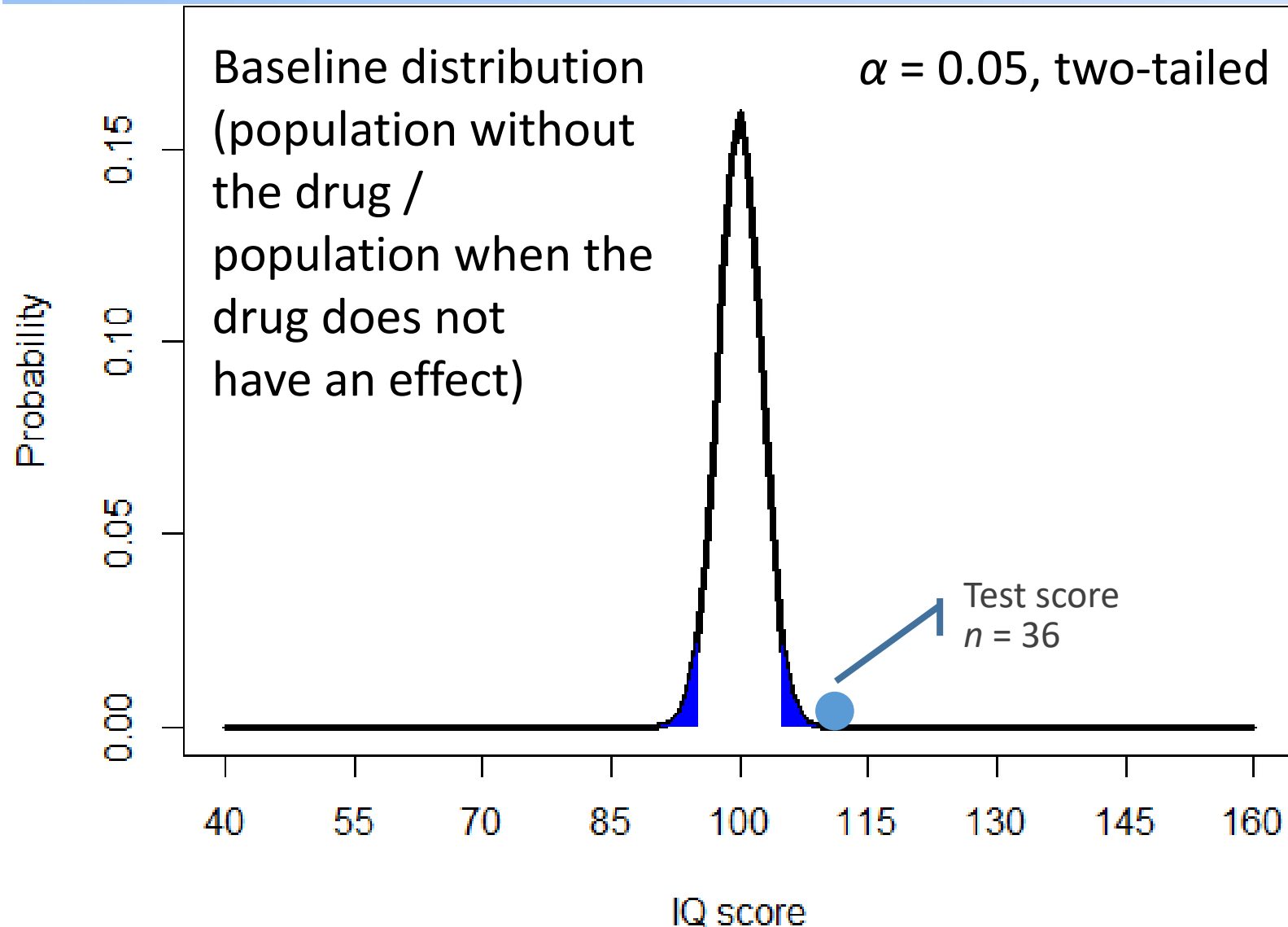


Assuming that the drug has no effects, this is how the distribution of people taking the drug is like (null distribution).

People taking the drug are consistent with the distribution that the drug has no effects.

We **cannot conclude** that the drug has an effect on IQ.

Sampling distribution ($n = 36$) of IQ scores (case 2)



Assuming that the drug has no effects, this is how the distribution of people taking the drug is like (null distribution).

People taking the drug score higher than the majority of the population assuming the drug has no effects.

It is unlikely that the drug has no effects.

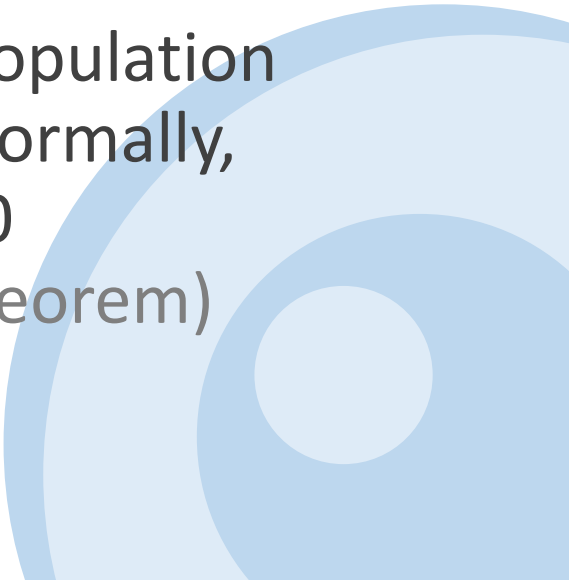
We conclude that the drug improves IQ.

Information we need: Z-test

The Baseline Distribution

- Population Mean μ
- Population Standard Deviation σ
- Kurtosis: Whether the scores in the population are normally distributed

The Test Scores

- A representative sample / reasonable sample size (of the treated population)
 - Sample mean with the treatment
 - If the scores in the population are not distributed normally, get a sample size >30 (ref: Central Limit Theorem)
- 

The Mozart Effect -- Music experience on IQ

We know that the general population has normally distributed IQ scores: The mean is 100 and the standard deviation is 15.

Some researchers wanted to test whether the “Mozart effect” is reliable.

They recruited 25 newborns and sent their parents a Mozart CD as a gift.

These parents played the CD to the babies for an hour a day for 3 years before the babies were tested.

The mean IQ score of the treated group was 105.



Hypothesis Testing – Step-by-step

- Step 0: Gather all the information you need, check the assumptions

$\mu_{\text{baseline}} =$
 $\sigma =$
 $M_{\text{treatment}} =$; $n =$
Kurtosis of population

- Step 1: Set up the hypotheses

- H_0 (null hypothesis)
- H_1 (Research hypothesis)

$H_0: \mu_{\text{baseline}} = \mu_{\text{treatment}}$
 $H_1: \mu_{\text{baseline}} \neq \mu_{\text{treatment}}$

- Step 2: Find the rejection region

$\alpha = 0.05$, two-tailed
 $Z_{\text{critical}} =$ quantiles corresponding to 2.5% on the each tail

- Step 3: Calculate the test statistic

$Z_{\text{sample}} = \frac{M_{\text{treatment}} - \mu_{\text{baseline}}}{\sigma_M}$

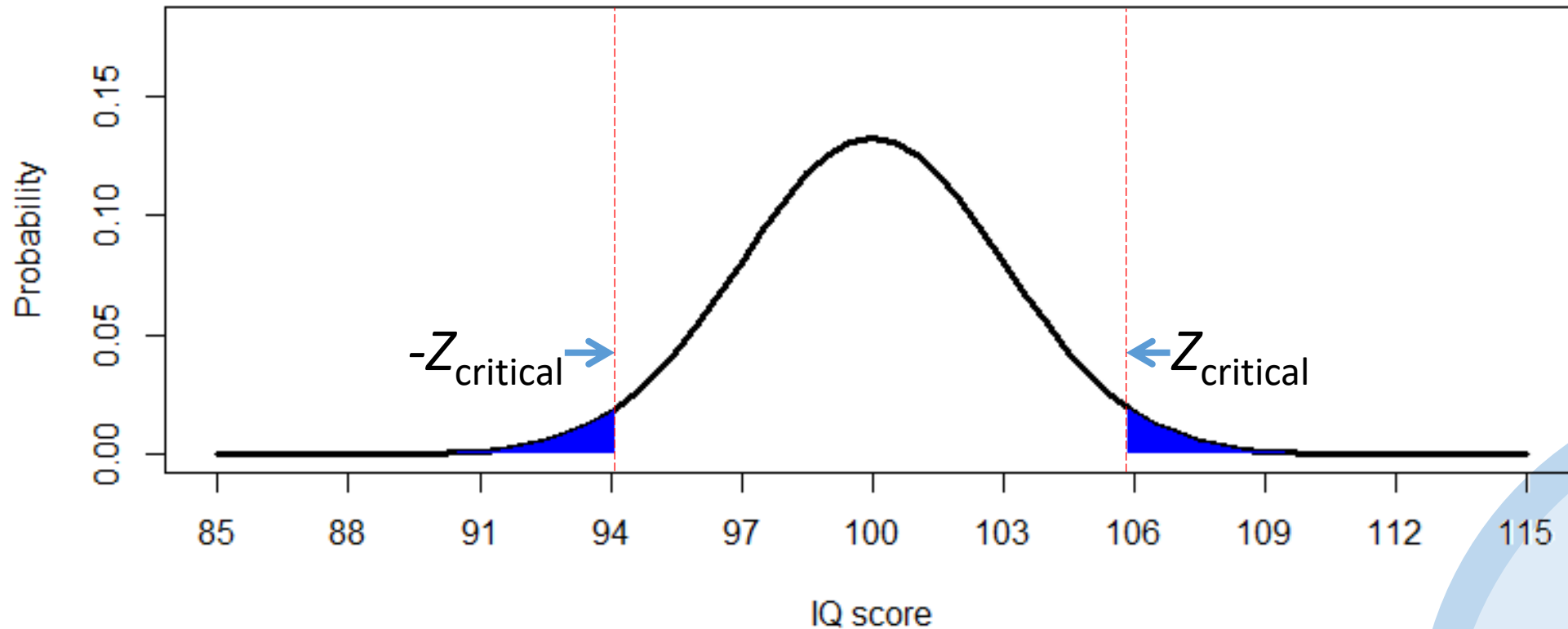
- Step 4: Decision & Conclusion

Reject or not reject null hypothesis, conclusion

Hypothesis Testing – Step-by-step

- Step 0: Gather all the information you need, check the assumptions
 - $\mu_{\text{baseline}} = 100$
 $\sigma = 15$
 $M_{\text{treatment}} = 105$; $n = 25$
Kurtosis of population ✓
- Step 1: Set up the hypotheses
 - H_0 (null hypothesis)
 - H_1 (Research hypothesis)
- Step 2: Find the rejection region
 - $H_0: \mu_{\text{baseline}} = \mu_{\text{treatment}}$
 $H_1: \mu_{\text{baseline}} \neq \mu_{\text{treatment}}$
 - $\alpha = 0.05$, two-tailed
 $Z_{\text{critical}} =$
- Step 3: Calculate the test statistic
 - $Z_{\text{sample}} = \frac{M_{\text{treatment}} - \mu_{\text{baseline}}}{\sigma_M}$
- Step 4: Decision & Conclusion

Sampling distribution ($\mu_{\text{baseline}} = 100, \sigma_M = 3$)



If $Z_{\text{sample}} > Z_{\text{critical}}$ or $Z_{\text{sample}} < -Z_{\text{critical}} \rightarrow \text{Reject } H_0$

Hypothesis Testing – Step-by-step

- Step 0: Gather all the information you need, check the assumptions

$\mu_{\text{baseline}} = 100$
 $\sigma = 15$
 $M_{\text{treatment}} = 105$; $n = 25$
Kurtosis of population ✓

- Step 1: Set up the hypotheses

- H_0 (null hypothesis)
- H_1 (Research hypothesis)

$H_0: \mu_{\text{baseline}} = \mu_{\text{treatment}}$
 $H_1: \mu_{\text{baseline}} \neq \mu_{\text{treatment}}$

- Step 2: Find the rejection region

$\alpha = 0.05$, two-tailed
 $Z_{\text{critical}} = \text{qnorm}(c(0.05/2, 1-0.05/2),$
 $\text{mean} = 0, \text{sd} = 1) = [-1.96, 1.96]$

- Step 3: Calculate the test statistic

$Z_{\text{sample}} = \frac{M_{\text{treatment}} - \mu_{\text{baseline}}}{\sigma_M} = +1.67$

- Step 4: Decision & Conclusion

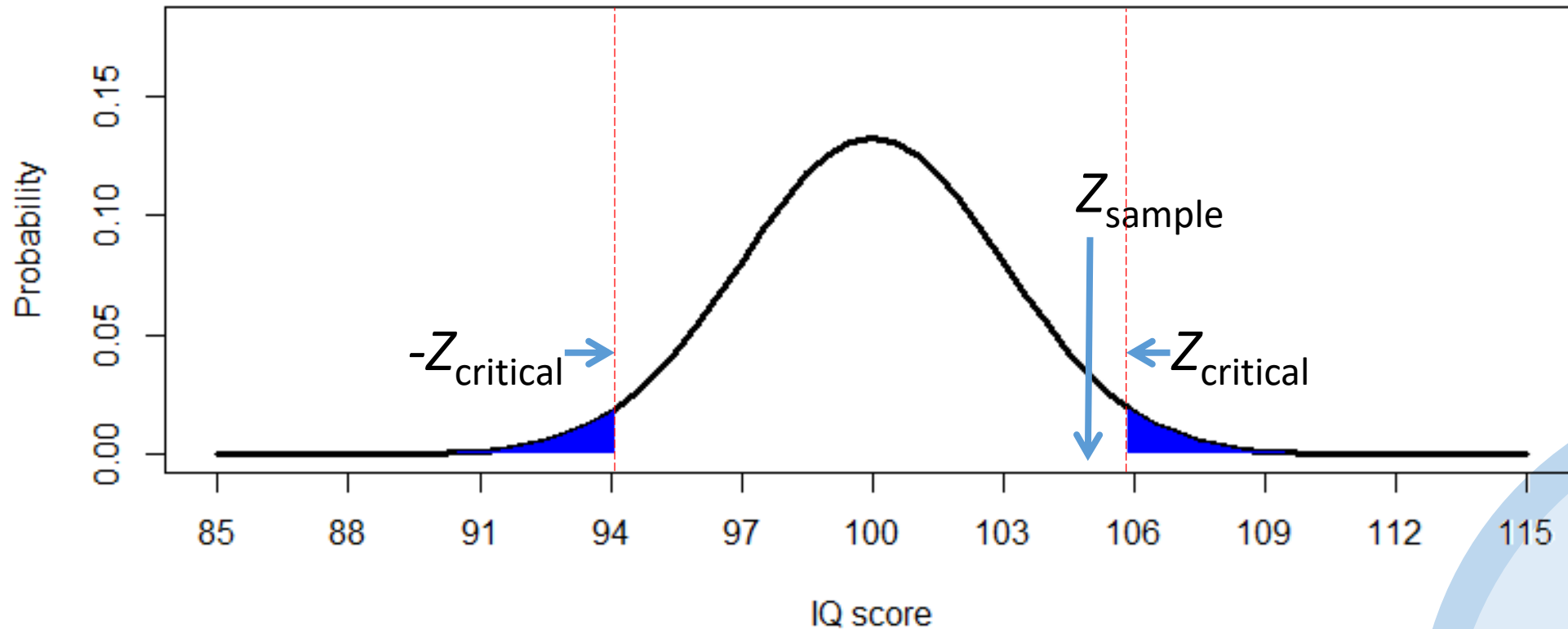
- Do not reject null hypothesis.

We cannot conclude that listening to Mozart improves IQ

($Z = +1.67, p > 0.05$)

Probability of getting such a score, or a more extreme score

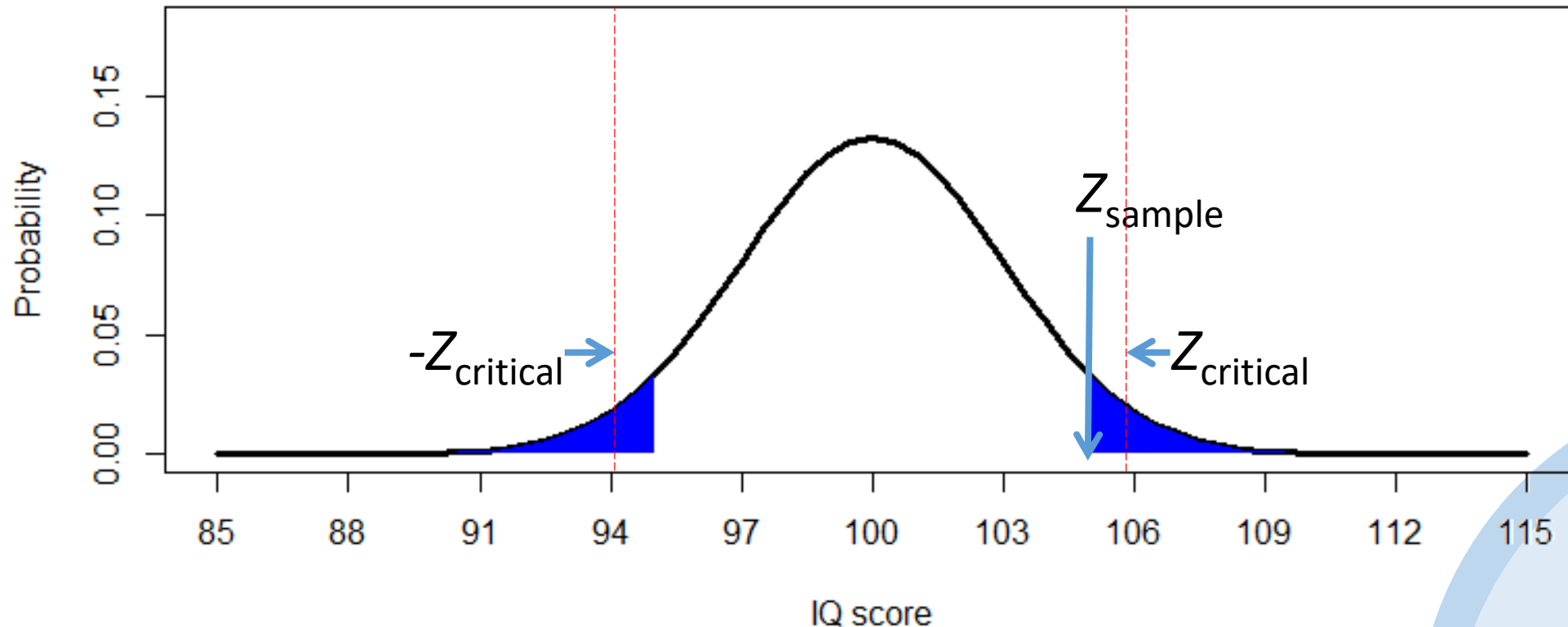
Sampling distribution ($\mu_{\text{baseline}} = 100, \sigma_M = 3$)



If $Z_{\text{sample}} > Z_{\text{critical}}$ or $Z_{\text{sample}} < -Z_{\text{critical}} \rightarrow \text{Reject } H_0$

Sampling distribution ($\mu_{\text{baseline}} = 100, \sigma_M = 3$)

(Alternative Method: finding probability associated with Z_{sample})



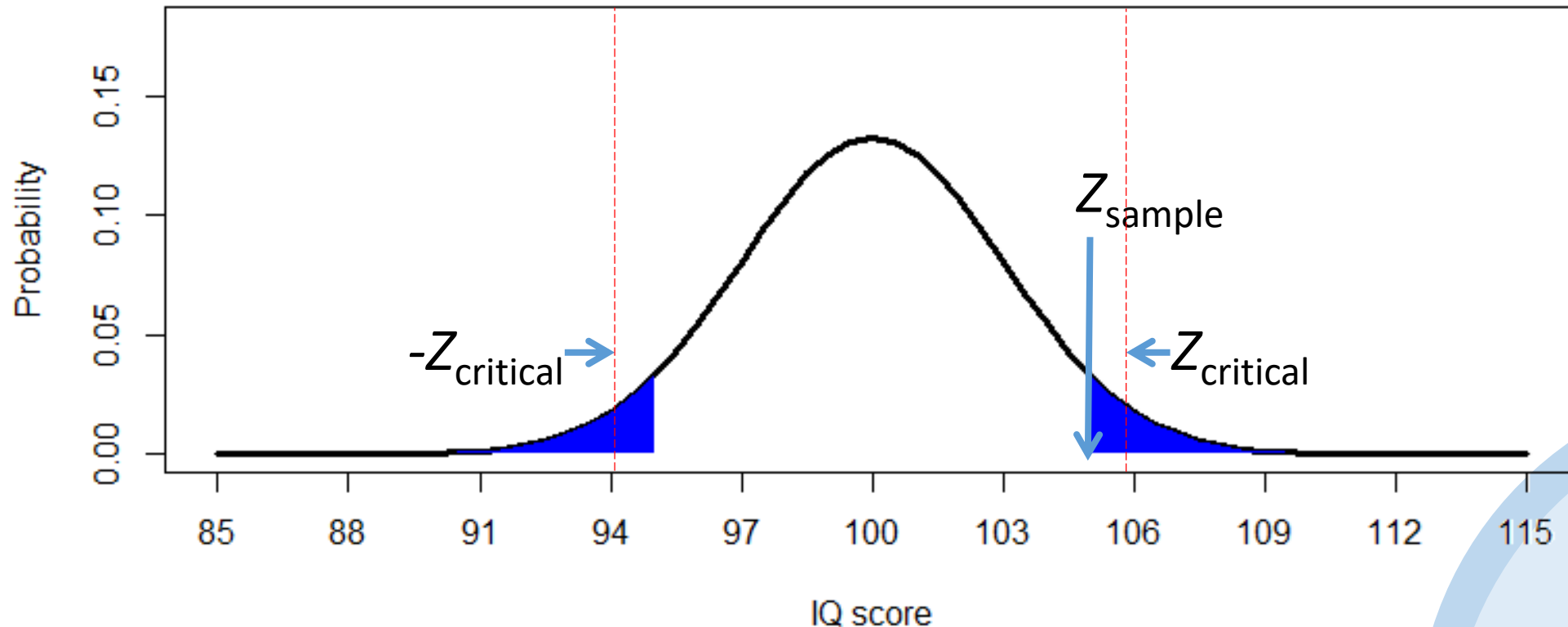
If $p < \alpha$, reject H_0

Hypothesis Testing – Step-by-step (alternative method)

- Step 0: Gather all the information you need, check the assumptions
 - $\mu_{\text{baseline}} = 100$
 $\sigma = 15$
 $M_{\text{treatment}} = 105$; $n = 25$
Kurtosis of population ✓
 - Step 1: Set up the hypotheses
 - H_0 (null hypothesis)
 - H_1 (Research hypothesis)
 - Step 2: Calculate the test statistic
 - Step 3: Find the probability associated with the test statistic
(i.e., the probability of getting such a score or a more extreme score x 2, given H_0 is true)
 - Step 4: Decision & Conclusion
(compare p against α)
- $H_0: \mu_{\text{baseline}} = \mu_{\text{treatment}}$
 $H_1: \mu_{\text{baseline}} \neq \mu_{\text{treatment}}$
 - $Z_{\text{sample}} = \frac{M_{\text{treatment}} - \mu_{\text{baseline}}}{\sigma_M} = +1.67$
 - $p(|Z_{\text{sample}}| | H_0) =$

Sampling distribution ($\mu_{\text{baseline}} = 100, \sigma_M = 3$)

(Alternative Method: finding probability associated with Z_{sample})



If $p < \alpha$, reject H_0

Hypothesis Testing – Step-by-step (alternative method)

- Step 0: Gather all the information you need, check the assumptions

- $\mu_{\text{baseline}} = 100$
 $\sigma = 15$
 $M_{\text{treatment}} = 105$; $n = 25$
Kurtosis of population ✓

- Step 1: Set up the hypotheses

- H_0 (null hypothesis)
- H_1 (Research hypothesis)

- $H_0: \mu_{\text{baseline}} = \mu_{\text{treatment}}$
 $H_1: \mu_{\text{baseline}} \neq \mu_{\text{treatment}}$

- Step 2: Calculate the test statistic

- $Z_{\text{sample}} = \frac{M_{\text{treatment}} - \mu_{\text{baseline}}}{\sigma_M} = +1.67$

- Step 3: Find the probability associated with the test statistic
(i.e., the probability of getting such a score or a more extreme score x 2, given H_0 is true)

- $p(|Z_{\text{sample}}| | H_0) = 2 * (1 - \text{pnorm}(z_{\text{sample}}, \text{mean} = 0, \text{sd} = 1))$

For a two-tailed test and Z_{sample} on the right

- Step 4: Decision & Conclusion
(compare p against α)

- Do not reject null hypothesis.
We cannot conclude that listening to Mozart improves IQ
($Z = +1.67, p = 0.096$)

Class Challenge

- SAT is a standardized college entrance exam. It has a maximum score of 1600, a mean of 990 and standard deviation of 220. To test the effectiveness of “The Official SAT Practice”, The College Board (which administers SAT) teamed up with Khan Academy for a study. They recruited 49 high school students, providing them with 20 hours of free SAT practice. The group attained an average score of 1045 points. Test The College Board’s claims that “The Official SAT Practice” helps students for the SAT.
(with $\alpha = 0.05$, two-tailed test)

Class Challenge: Hypothesis Testing

- Step 0: Gather all the information you need, check the assumptions

- $\mu_{\text{baseline}} =$
 $\sigma =$
 $M_{\text{treatment}} =$; $n =$
Kurtosis of population

- Step 1: Set up the hypotheses
 - H_0 (null hypothesis)
 - H_1 (Research hypothesis)

- $H_0:$
 $H_1:$

- $Z_{\text{sample}} = \frac{M_{\text{treatment}} - \mu_{\text{baseline}}}{\sigma_M}$

- Step 2: Calculate the test statistic
- Step 3: Find the probability associated with the test statistic
(i.e., the probability of getting such a score or a more extreme score $\times 2$, given H_0 is true)
- Step 4: Decision & Conclusion
(compare p against α)

- $p(|Z_{\text{sample}}| | H_0) =$

Class Challenge: Hypothesis Testing

- Step 0: Gather all the information you need, check the assumptions
 - $\mu_{\text{baseline}} = 990$
 $\sigma = 220$
 $M_{\text{treatment}} = 1045$; $n = 49$
Kurtosis of population ✓
 - Step 1: Set up the hypotheses
 - H_0 (null hypothesis)
 - H_1 (Research hypothesis)
 - Step 2: Calculate the test statistic
 - Step 3: Find the probability associated with the test statistic
(i.e., the probability of getting such a score or a more extreme score x 2, given H_0 is true)
 - Step 4: Decision & Conclusion
(compare p against α)
- $H_0: \mu_{\text{baseline}} = \mu_{\text{treatment}}$
 $H_1: \mu_{\text{baseline}} \neq \mu_{\text{treatment}}$
 - $Z_{\text{sample}} = \frac{M_{\text{treatment}} - \mu_{\text{baseline}}}{\sigma_M} = +1.75$
 - $p(|Z_{\text{sample}}| | H_0) = 2 * (1 - \text{pnorm}(1.75, \text{mean} = 0, \text{sd} = 1))$
 - Do not reject null hypothesis. We cannot conclude that spending 20 hours with The Official SAT Practice improves SAT scores ($Z = +1.75, p = 0.080$)

Confidence Interval (C.I.): Reliability Measure

In an ideal world

- When we devise a study --
We are interested in the *population parameters* of the treated population (usually the mean)
- All high school students with “The Official SAT Practice”

$\mu_{treatment}$



Can never be known

In reality

- What we have is a *sample mean*

$M_{treatment}$



Not entirely accurate

Confidence Interval (C.I.): Reliability Measure

In an ideal world

- When we devise a study --
We are interested in the *population parameters* of the treated population (usually the mean)

$$\mu_{treatment}$$

In reality

- What we have is a *sample mean*

$$M_{treatment} \pm Z_{(1-\alpha/2)} \times \sigma_M$$

If the experiment is repeated over and over (same n and condition), for $(1 - \alpha)$ of the time, the sample means will fall within the interval

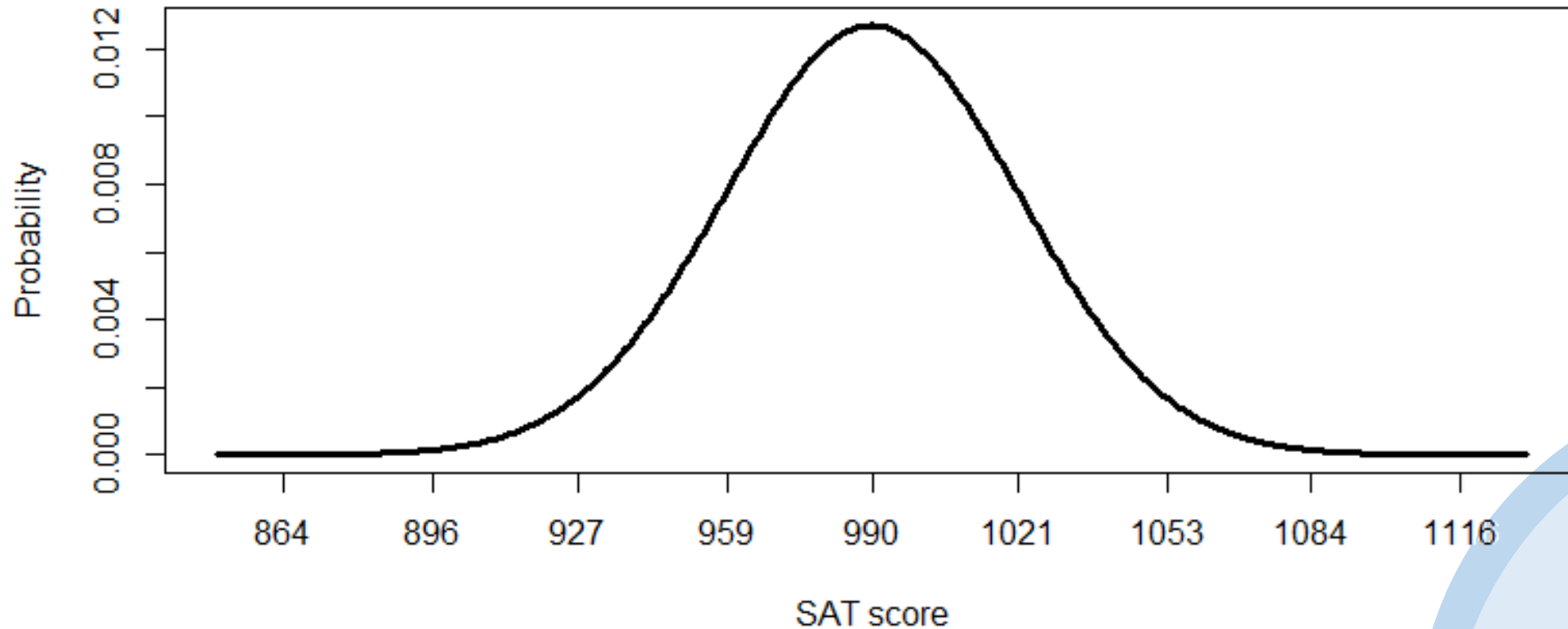
Class Challenge: 95% Confidence Interval

- SAT is a standardized college entrance exam. It has a maximum score of 1600, a mean of 990 and standard deviation of 220. To test the effectiveness of “The Official SAT Practice”, The College Board (which administers SAT) teamed up with Khan Academy for a study. They recruited 49 high school students, providing them with 20 hours of free SAT practice. The group attained an average score of 1045 points. Test The College Board’s claims that “The Official SAT Practice” helps students for the SAT.
(with $\alpha = 0.05$, two-tailed test)

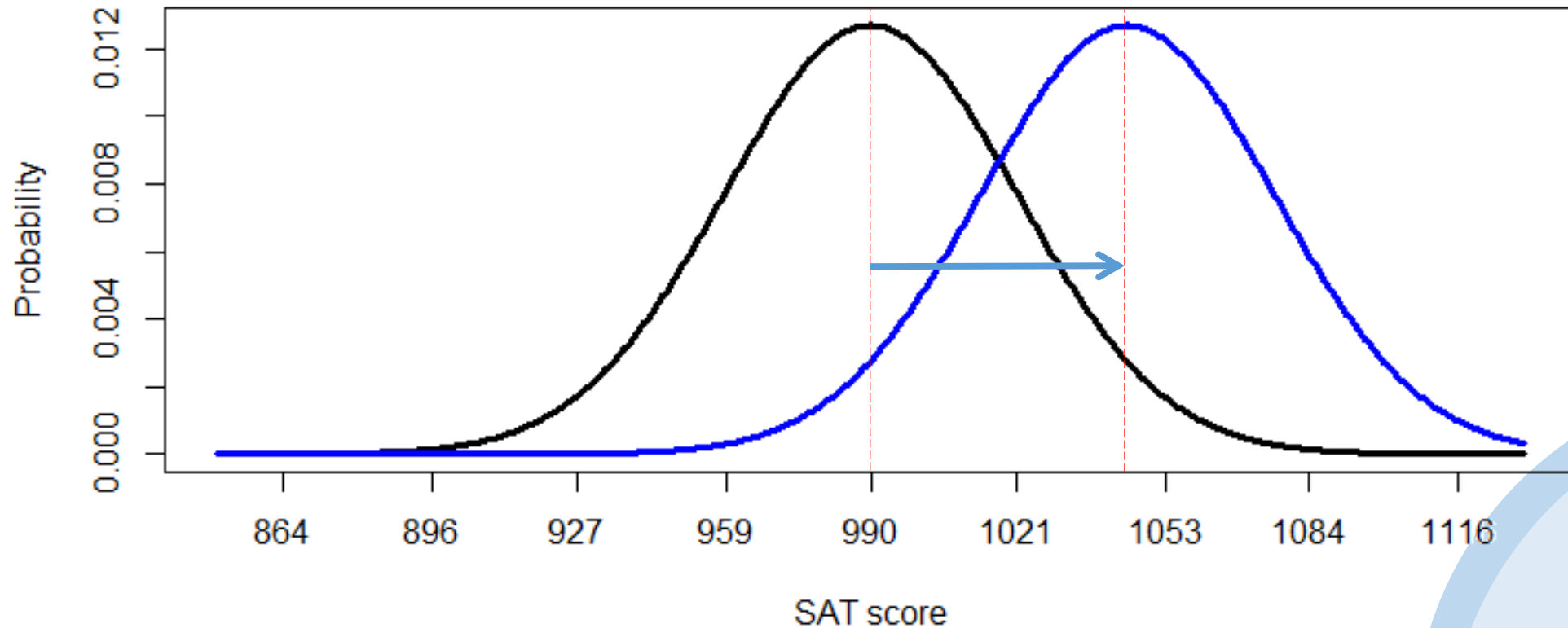
$$95\% \text{ C.I.} = M_{treatment} \pm Z_{(1-0.05/2)} \times \sigma_M$$

If the experiment is repeated over and over (same n and condition), for $(1 - \alpha)$ of the time, the sample means will fall within the interval

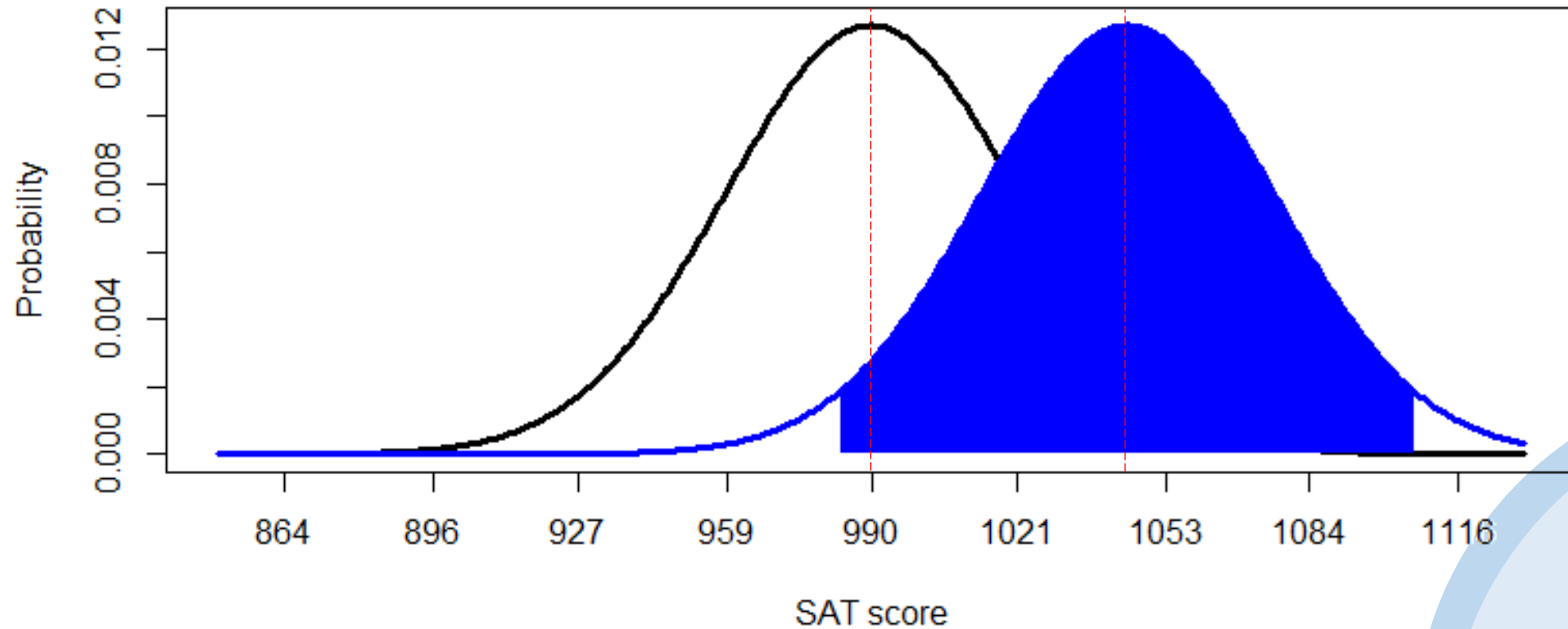
Sampling distribution ($\mu_{\text{baseline}} = 990$, $\sigma_M = 220/\sqrt{49}$)



Sampling distribution ($\mu_{\text{baseline}} = 990$, $\sigma_M = 220/\sqrt{49}$)



95% Confidence Interval



Class Challenge: 95% Confidence Interval

- SAT is a standardized college entrance exam. It has a maximum score of 1600, a mean of 990 and standard deviation of 220. To test the effectiveness of “The Official SAT Practice”, The College Board (which administers SAT) teamed up with Khan Academy for a study. They recruited 49 high school students, providing them with 20 hours of free SAT practice. The group attained an average score of 1045 points. Test The College Board’s claims that “The Official SAT Practice” helps students for the SAT.
(with $\alpha = 0.05$, two-tailed test)

$$95\% \text{ C.I.} = M_{treatment} \pm Z_{(1-0.05/2)} \times \sigma_M$$

If the experiment is repeated over and over (same n and condition), for $(1 - \alpha)$ of the time, the sample means will fall within the interval

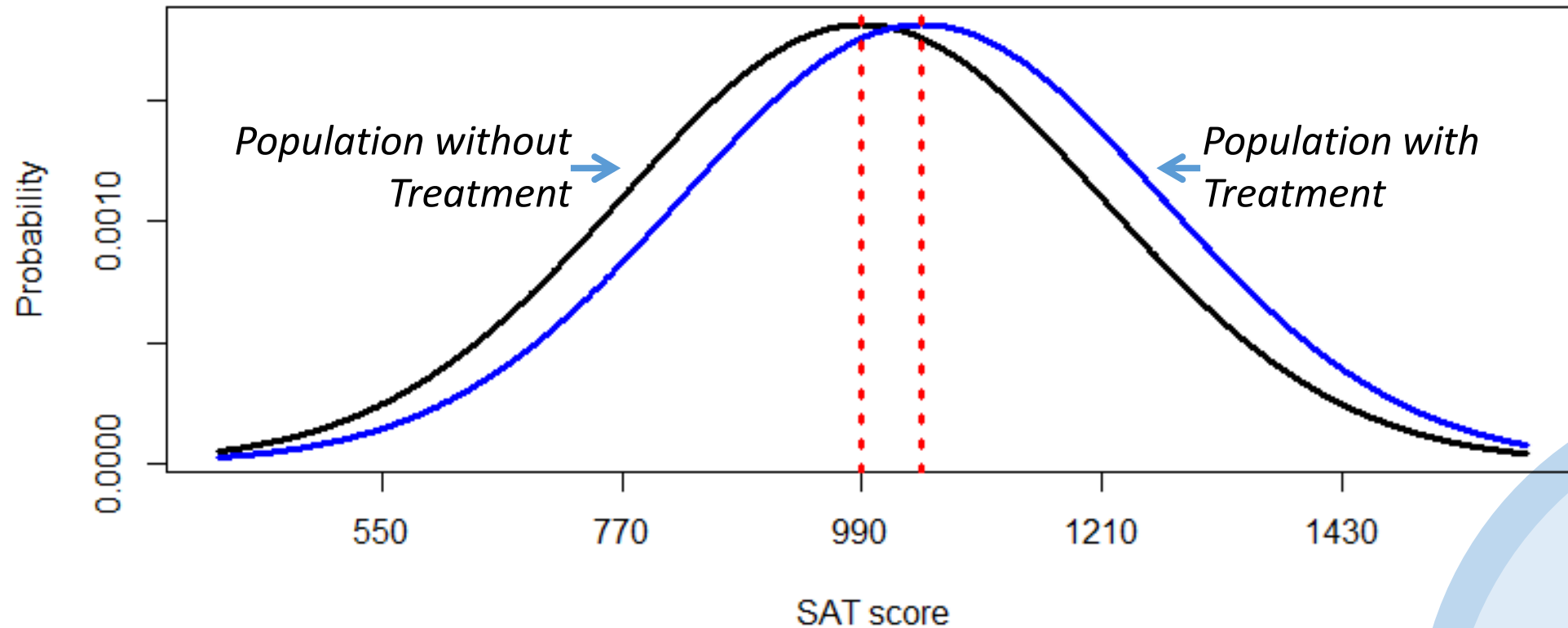
Effect size: Effect in Standard Deviations

- We want to describe the size of the effect with a unit-free measure
- So we don't have to know the scale of DV
- For example, what's the average improvement in SAT scores for an individual with 20 hours of the "The Official SAT Practice"

$$\text{Cohen's } d = \frac{\mu_{\text{treatment}} - \mu_{\text{baseline}}}{\sigma}$$

Cohen's d	Effect Size
0.2	Small
0.5	Medium
0.8	Large

Population distributions: Cohen's d



$$\text{cohens_d} = (1045 - 990) / 220$$

Wrap up!

- Hypothesis Testing: Reliability of the treatment effect
- Confidence Interval: Reliability of the estimated population parameter
- Effect Size: Average effect for individual in sd
- Next week -- Type I and II errors, Power

Reminder:

- First homework due tomorrow at 11:59pm
 - Second homework released by tomorrow night
 - Office hours: tomorrow 2 – 4 @ Mandler 1586
- 