

# Lecture 26 Error, Power And Sample Size Estimation

BIO210 Biostatistics

---

Xi Chen

Spring, 2022

School of Life Sciences

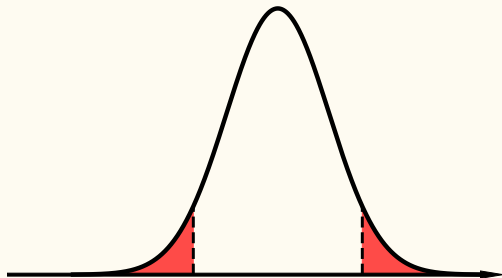
Southern University of Science and Technology



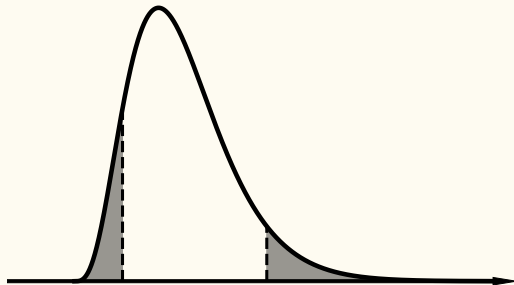
南方科技大学生命科学学院  
SUSTech · SCHOOL OF  
**LIFE SCIENCES**

# One-sample Hypothesis Testing For Variance

Sampling distribution of the sample mean/proportion



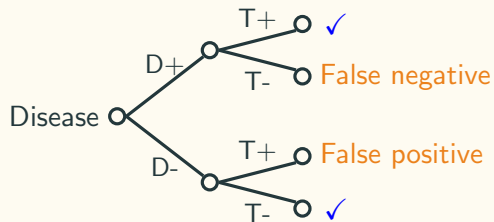
Sampling distribution of the sample variance



Significance level  $\alpha$ : never 0 !

# Types of Errors

## Diagnostic testing



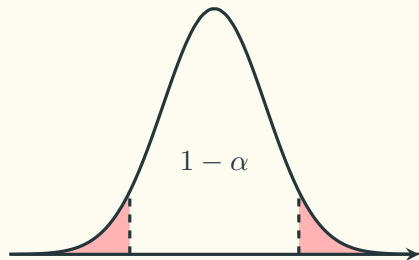
## Hypothesis testing

Decision \ Truth	Population	
	$H_0$ is true	$H_0$ is false
Reject $H_0$	Type I Error	✓
Do not reject $H_0$	✓	Type II Error

# Probability of Making Different Errors

- $P(\text{Type I Error}) = ?$
- $P(\text{reject } H_0 \mid H_0 \text{ is true}) = ?$

When  $H_0$  is true:



Type I error is also called the **rejection error** or  **$\alpha$  error**.

- $P(\text{Type II Error}) = ?$
- $P(\text{Do not reject } H_0 \mid H_0 \text{ is false}) = \beta$
- $1 - \beta$  is more useful in reality

## Definition

The **Power** of the test is defined as  
 $P(\text{reject } H_0 \mid H_0 \text{ is false}) = 1 - \beta$

## Calculating Power - Roughly

- Serum cholesterol level for 20- to 74-year-old males
- Truth about the whole population: normally distributed, mean 200 mg/100ml ( $\mu = 200$ ), standard deviation 46 mg/100ml ( $\sigma = 46$ ): we don't know this!
- A subpopulation (20- to 24-year-old males): normally distributed, mean 180 mg/100ml ( $\mu = 180$ ), standard deviation 46 mg/100ml ( $\sigma = 46$ ): we do know this from a previous study!
- Now we are interested in serum cholesterol level for 20- to 74-year-old males, and a random sample of size ( $n = 25$ ) is drawn from the 20- to 74-year-old male population. We conduct a hypothesis testing about the mean.

## Calculating Power - Roughly

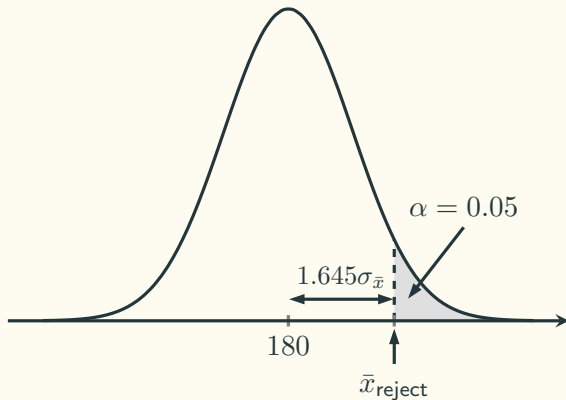
- We have reasons to believe that the mean serum cholesterol level of 20- to 74-year-old male should be higher than 180 mg/100ml.

- $H_0 : \mu \leq \mu_0 = 180 \text{ mg/100ml}$

- $H_1 : \mu > \mu_0 = 180 \text{ mg/100ml}$

- Truth:

$$H_1 : \mu = \mu_1 = 200 \text{ mg/100ml}$$

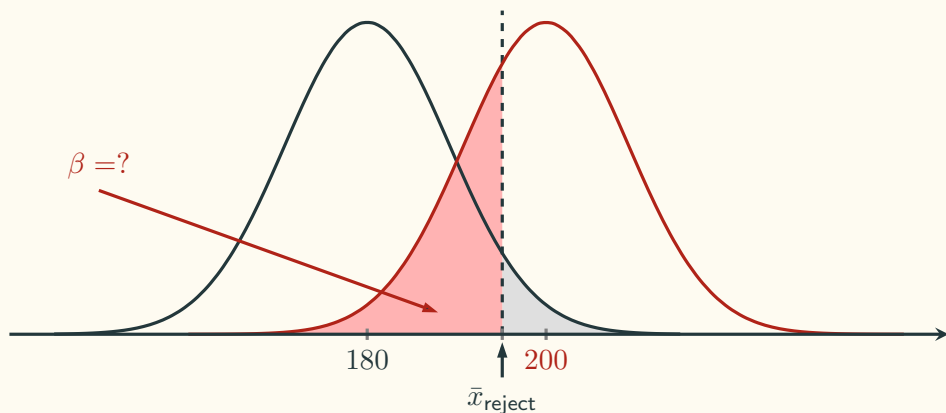


$$\bar{x}_{\text{reject}} = 180 + 1.645 \times \frac{46}{\sqrt{25}} = 195.1$$

## Calculating Power - Roughly

What we want to calculate: **Power** ( $1 - \beta$ ):  $P(\text{reject } H_0 \mid H_0 \text{ is false})$

Truth:  $H_1 : \mu = \mu_1 = 200 \text{ mg}/100\text{ml}$



## Calculating Power - Roughly

$$\beta = P(\text{do not reject } H_0 \mid H_0 \text{ is false})$$

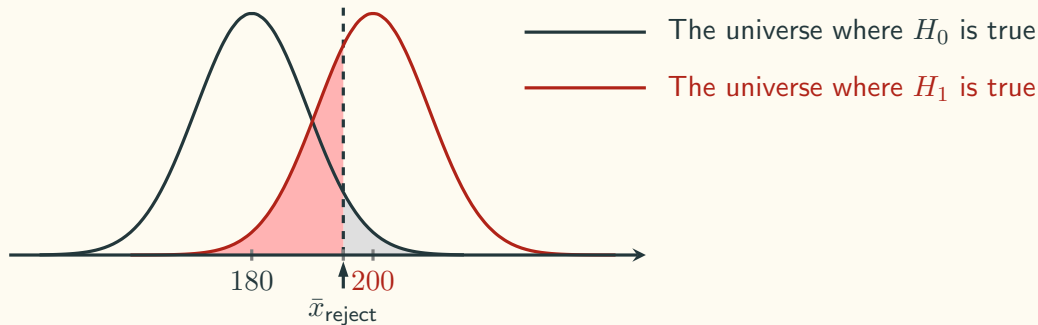
$$= P\left(\bar{x} \leq 195.1 \mid \mu_{\bar{x}} = \mu_1 = 200, \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{25}} = \frac{46}{5} = 9.2\right)$$

$$P\left(z \leq \frac{195.1 - 200}{9.2}\right) = P(z \leq -0.53) = 0.298$$

Power:  $1 - \beta = 1 - 0.298 = 0.702$



# How To Increase Power



## Increasing the power of the test

1. Shift  $\bar{x}_{\text{reject}}$  to the left
2.  $\mu_0$  shifts to the left, or  $\mu_1$  shifts to the right.
3. Make the sampling distribution narrower

## Ways of achieving the goal

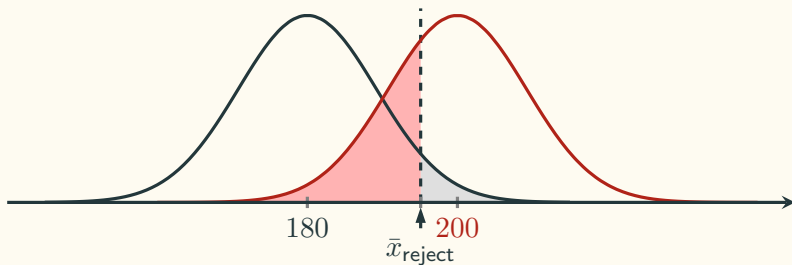
1. Increase  $\alpha$
2. Increase  $|\mu_0 - \mu_1|$
3. Decrease  $\sigma_{\bar{x}} \Leftrightarrow$  Increase  $n$

## Sample Size Estimation

In the previous example about the serum cholesterol level, suppose we want a significance level of 0.01 and a power of 0.95. What is the minimum sample size needed for the test?

$\alpha = 0.01, \beta = 0.05, \text{power} = 0.95$ : when  $H_0$  is true, we want to risk a 1% chance of rejecting it; when  $H_0$  is false, we want to risk a 5% chance of failing to reject it. The power of the test is 95%.

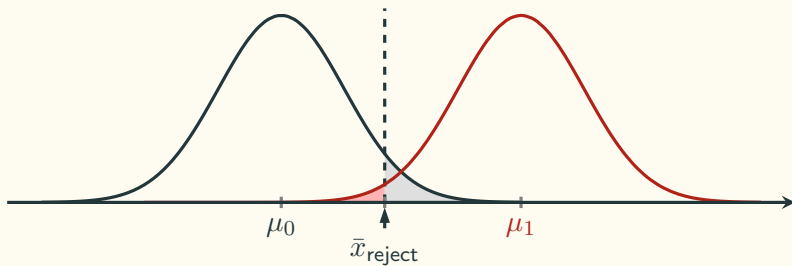
## Sample Size Estimation



1. The minimum sample mean to reject  $H_0$ :  $\bar{x}_{\text{reject}} = 180 + 2.32 \times \frac{46}{\sqrt{n}}$
2. Calculate  $\beta$ :

$$\begin{aligned}\beta &= P(\bar{x} \leq \bar{x}_{\text{reject}} \mid H_0 \text{ is false}) = P(\bar{x} \leq \bar{x}_{\text{reject}} \mid \mu_{\bar{x}} = 200, \sigma_{\bar{x}} = \frac{46}{\sqrt{n}}) \\ &= P\left(z \leq \frac{180 + 2.32 \times \frac{46}{\sqrt{n}} - 200}{\frac{46}{\sqrt{n}}}\right) = 0.05 \Rightarrow n = 83.165\end{aligned}$$

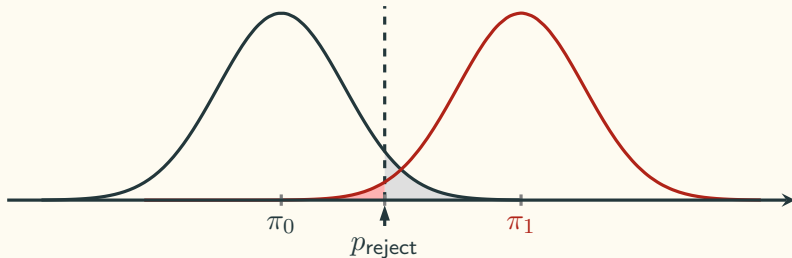
## Sample Size Estimation For One Sided Test For The Mean



1. The minimum sample mean to reject  $H_0$ :  $\bar{x}_{\text{reject}} = \mu_0 + Z_\alpha \times \frac{\sigma}{\sqrt{n}}$
2. Calculate  $\beta$ :

$$\begin{aligned}\beta &= P(\bar{x} \leq \bar{x}_{\text{reject}} \mid H_0 \text{ is false}) = P(\bar{x} \leq \bar{x}_{\text{reject}} \mid \mu_{\bar{x}} = \mu_1, \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}) \\ &= P\left(z \leq \frac{\mu_0 + Z_\alpha \times \frac{\sigma}{\sqrt{n}} - \mu_1}{\frac{\sigma}{\sqrt{n}}}\right) \Rightarrow n = \left[ \frac{(Z_\alpha + Z_\beta)\sigma}{\mu_1 - \mu_0} \right]^2\end{aligned}$$

# Sample Size Estimation For One Sided Test For The Proportion



1. The minimum sample mean to reject  $H_0$ :  $p_{\text{reject}} = \pi_0 + Z_\alpha \times \sqrt{\frac{\pi_0(1 - \pi_0)}{n}}$
2. Calculate  $\beta$ :

$$\beta = P(p \leq p_{\text{reject}} \mid H_0 \text{ is false}) = P\left(p \leq p_{\text{reject}} \mid \mu_p = \pi_1, \sigma_p = \sqrt{\frac{\pi_1(1 - \pi_1)}{n}}\right)$$

$$\Rightarrow n = \left[ \frac{Z_\alpha \sqrt{\pi_0(1 - \pi_0)} + Z_\beta \sqrt{\pi_1(1 - \pi_1)}}{\pi_1 - \pi_0} \right]^2$$