Terminologies

H_0 Null Hypothesis

H_a Alternate Hypothesis

$p-$value

Test Statistic

Significance Level α

Reject if p-value $< \alpha$

		Decision	
		Accept	Reject
H_0	True	True negative	False positive
	False	False negative	True positive

Burger company

A company selling burgers claims that its burger weighs 200 grams.

An unsatisfied customer, still feeling hungry after eating, wants to disprove the claim

Which is the right setup?

✓

H_0: weight = 200

H_a: weight < 200

H_0: weight = 200

H_a: weight $\neq$ 200

H_0: weight = 200

H_a: weight > 200

H_0: weight >= 200

H_a: weight < 200

H_0: weight <= 200

H_a: weight > 200

AI Chip startup

An AI chip startup claims that it beats the GPU in computer vision tasks

The training time for ResNet is 15 minutes on the GPU

What is the right way for the startup to claim that it is better?

Which is the right setup?

✔

H_0: training time = 15

H_a: training time < 15

H_0: training time = 15

H_a: training time > 15

H_0: training time = 15

H_a: training time $\neq$ 15

H_0: training time >= 15

H_a: training time < 15

H_0: training time <= 15

H_a: training time > 15

Height from your state

The average height of Indians is 65 inches

You want to verify whether this is true for people from your state

Let μ be the average height of people from your state

Which is the right setup?

H_0: $\mu = 65$
H_a: $\mu < 65$

H_0: $\mu = 65$
H_a: $\mu \neq 65$

H_0: $\mu = 65$
H_a: $\mu > 65$

H_0: $\mu >= 65$
H_a: $\mu < 65$

H_0: $\mu <= 65$
H_a: $\mu > 65$

Left Vs right tailed

Burger company

H_0: weight = 200

H_a: weight < 200

Left-tailed

AI Chip startup

H_0: training time = 15

H_a: training time < 15

Left-tailed

Height from your state

H_0: $\mu = 65$

H_a: $\mu \neq 65$

Two-tailed

Hypothesis Testing Framework

1) Setup the Null and Alternate Hypothesis
2) Choose the right test statistic
3) Left-tailed Vs Right-tailed Vs Two-tailed
4) Compute p-value
5) If p-value is less than alpha, then reject the null hypothesis

Recap central limit theorem

- Average height is 65 inches with std dev 2.5

We take a sample of 50 people

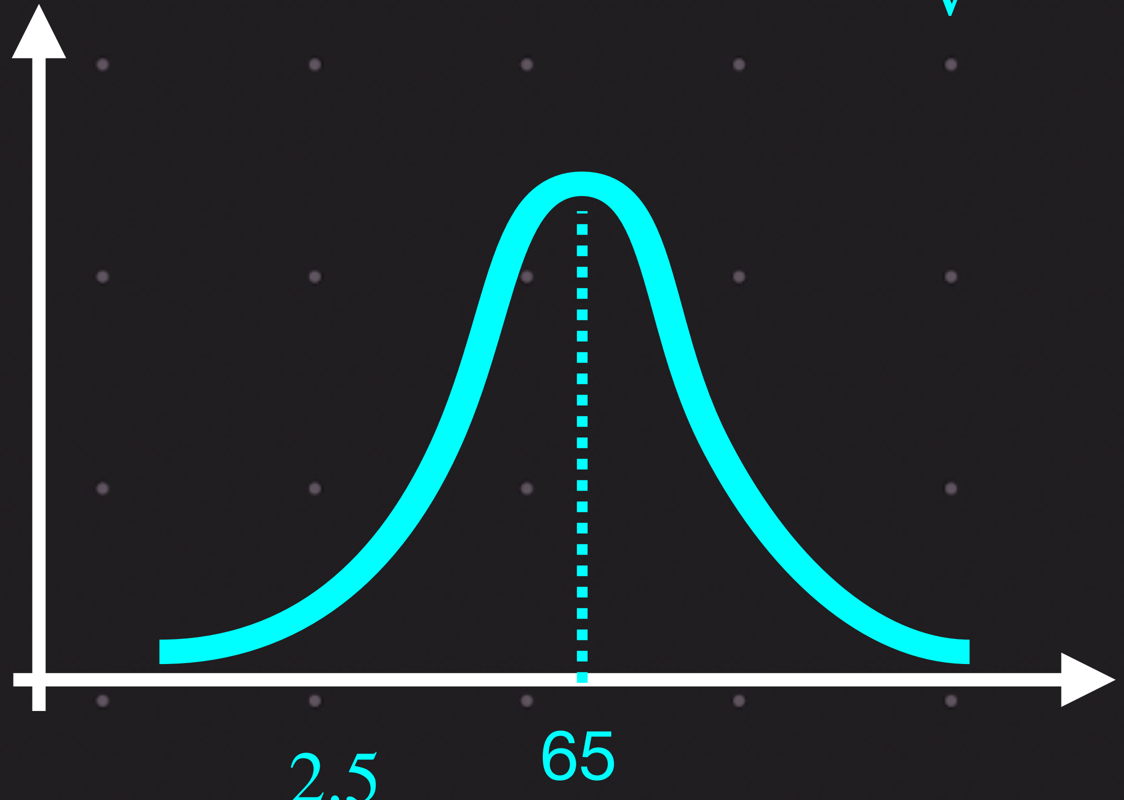
Let m represent the sample mean

Is m a random variable? Yes

What is its distribution? Gaussian (From CLT)

What is $E[m]$? 65

What is the std dev of m ? $\dfrac{2.5}{\sqrt{50}}$



std dev = $\dfrac{2.5}{\sqrt{50}}$

We take a sample of 5 people

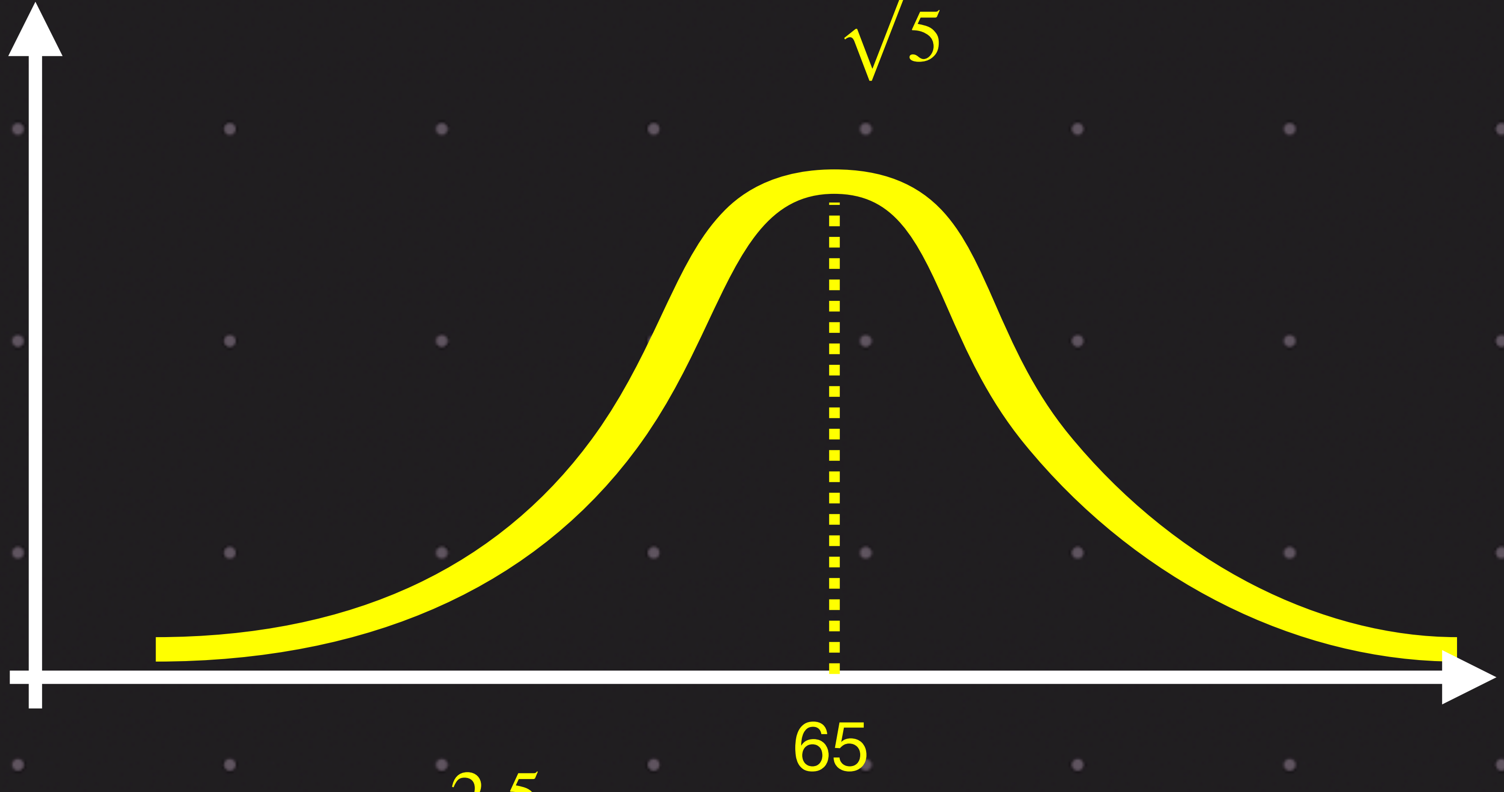
Let m represent the sample mean

Is m a random variable? Yes

What is its distribution? Gaussian (From CLT)

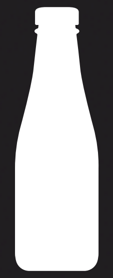
What is $E[m]$? 65

What is the std dev of m ? $\dfrac{2.5}{\sqrt{5}}$



std dev = $\dfrac{2.5}{\sqrt{5}}$

Supply chain example



$\mu = 1800$
$\sigma = 100$

A retailer has 2000 stores in the country

Historical data tells us that weekly sales of shampoo bottles has an average of 1800, with a standard deviation of 100

Sales team wants to improve sales by hiring a marketing team

Hiring a marketing team can be expensive, so we need to be <u>very sure</u> that they will improve sales

Before deploying their strategy for all 2000 stores, they are tested in 50 stores

On the 50 stores, their average sales for that week was 1850

You are the data scientist who should tell your sales team whether this is statistically significant

Sales team has said that we will hire only if we are <u>$99\,\%$ confident</u> $\alpha = 0.01$

Another marketing team is also being considered

They are tested on 5 stores

On the 5 stores, their average sales for that week was 1900

Would you say this team is better than the first one?

Between the “blue team” and the “yellow team”, whom will you choose?

Supply chain example

$\alpha = 0.01$

$\mu = 1800$
$\sigma = 100$

50 stores with average of 1850

Let $x_1, x_2, \cdots, x_{50}$ denote the sales

$H_0 : \mu_b = 1800$

$H_a : \mu_b > 1800$

$$m_b = \frac{x_1 + x_2 + \cdots + x_{50}}{50}$$

m_b is the sample mean

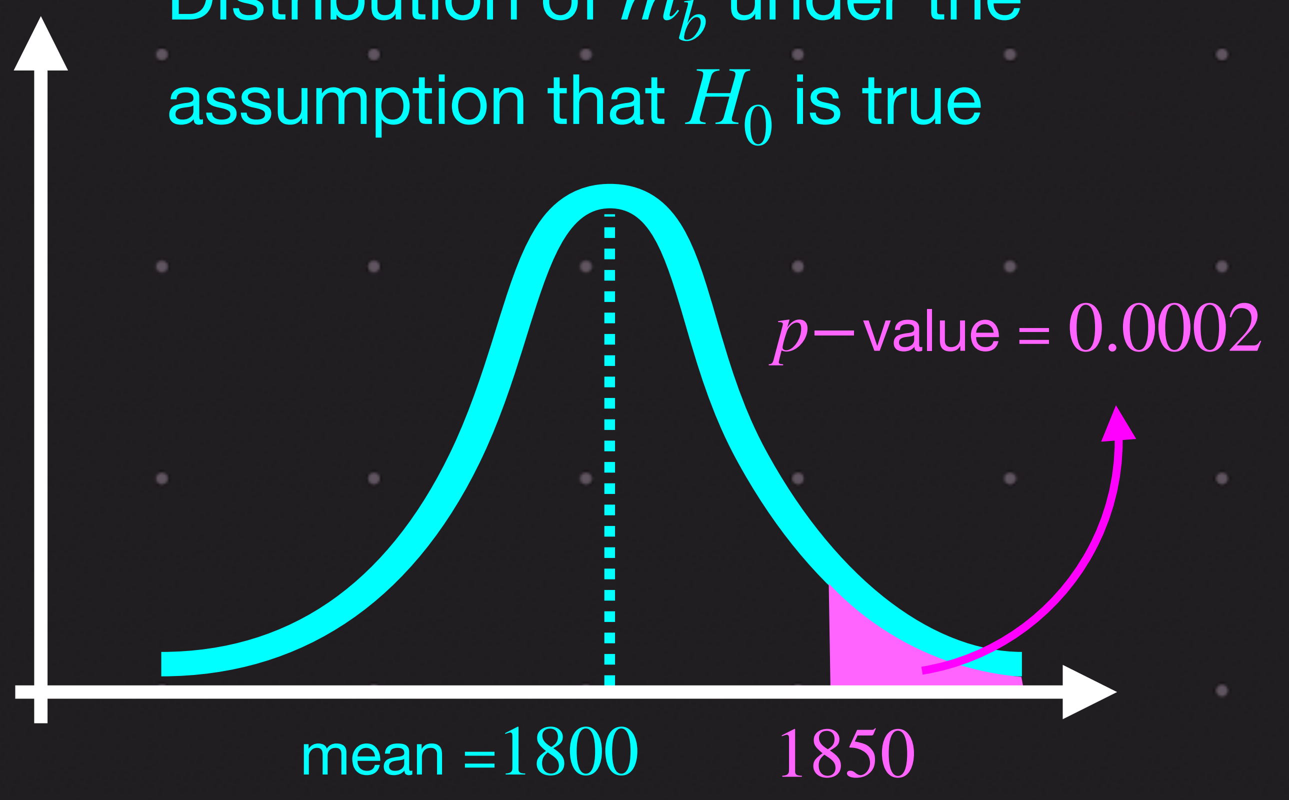
Is m_b a random variable? Yes

What is its distribution? Gaussian (From CLT)

What is $E[m_b]$? 1800

What is the std dev of m_b ? $\frac{100}{\sqrt{50}}$

Distribution of m_b under the assumption that H_0 is true



std dev = $\frac{100}{\sqrt{50}}$

What is the observed value of m_b? 1850

$$z = \frac{1850 - 1800}{100/\sqrt{50}} = 3.53$$

Right or Left tailed? Right tailed

H_a says “greater than”

How to compute $p-$value? $P\left[m_b \geq 1850 \;\middle|\; H_0 \text{ is true}\right] =$ `1 - norm.cdf(3.53)` $= 0.0002$

Is the $p-$value less than α ? Yes

We reject the null hypothesis

This means the marketing team had a positive effect on the sales

Supply chain example

$\alpha = 0.01$

$\mu = 1800$
$\sigma = 100$

5 stores with average of 1900

Let x_1, x_2, x_3, x_4, x_5 denote the sales

$H_0 : \mu_y = 1800$

$H_a : \mu_y > 1800$

$$m_y = \frac{x_1 + x_2 + \cdots + x_5}{5}$$

m_y is the sample mean

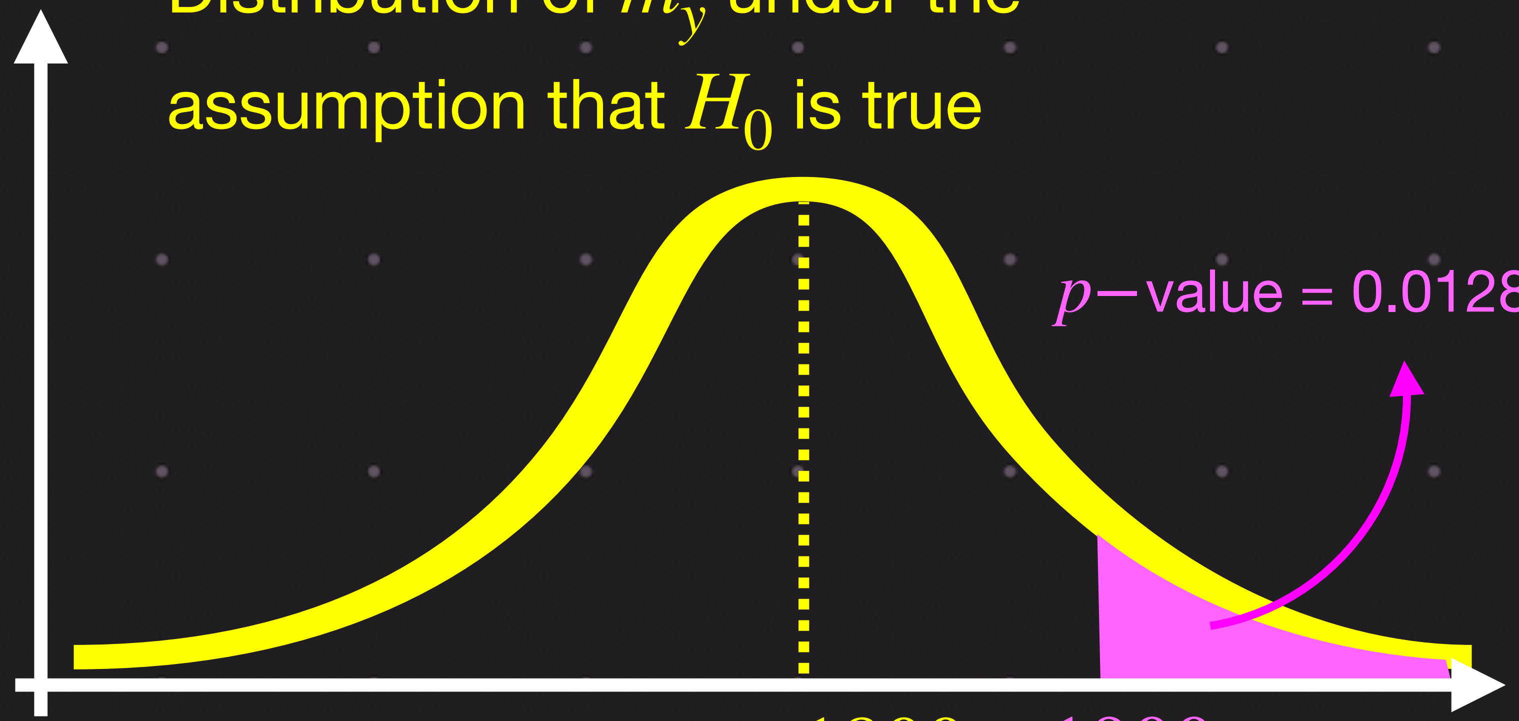
Is m_y a random variable? Yes

What is its distribution? Gaussian (From CLT)

What is $E[m_y]$? 1800

What is the std dev of m_y ? $\frac{100}{\sqrt{5}}$

Distribution of m_y under the assumption that H_0 is true



std dev = $\frac{100}{\sqrt{5}}$ mean $= 1800$ 1900

What is the observed value of m_y? 1900

$$z = \frac{1900 - 1800}{100/\sqrt{5}} = 2.23$$

Right or Left tailed? Right tailed

H_a says "greater than"

How to compute $p-$value? $P\left[m_y \geq 1900 \;\middle|\; H_0 \text{ is true}\right] =$ `1 - norm.cdf(2.23)` $= 0.0128$

Is the $p-$value less than α ? No

We fail to reject the null hypothesis The effect of marketing was not statistically significant

Supply chain example

$\alpha = 0.01$

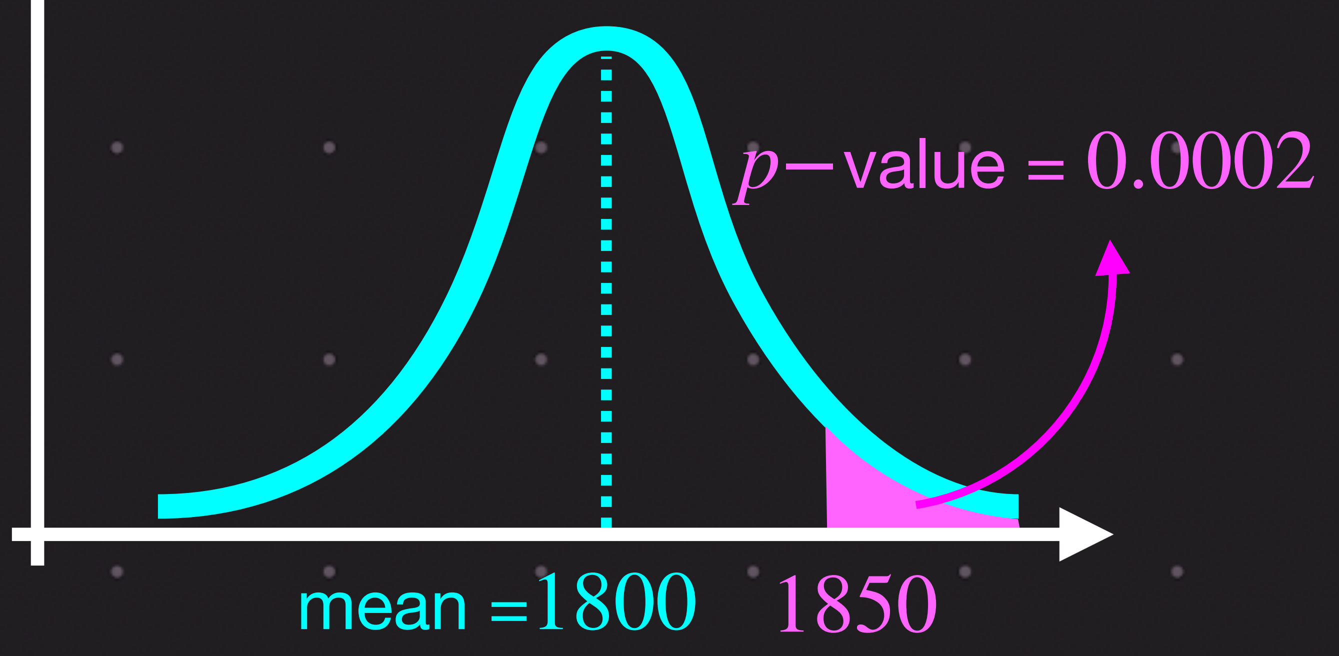
$\mu = 1800$
$\sigma = 100$

50 stores with average of 1850

$H_0 : \mu_b = 1800$

$H_a : \mu_b > 1800$

Distribution of m_b under the assumption that H_0 is true



std dev = $\dfrac{100}{\sqrt{50}}$

$$z = \frac{1850 - 1800}{100/\sqrt{50}} = 3.53$$

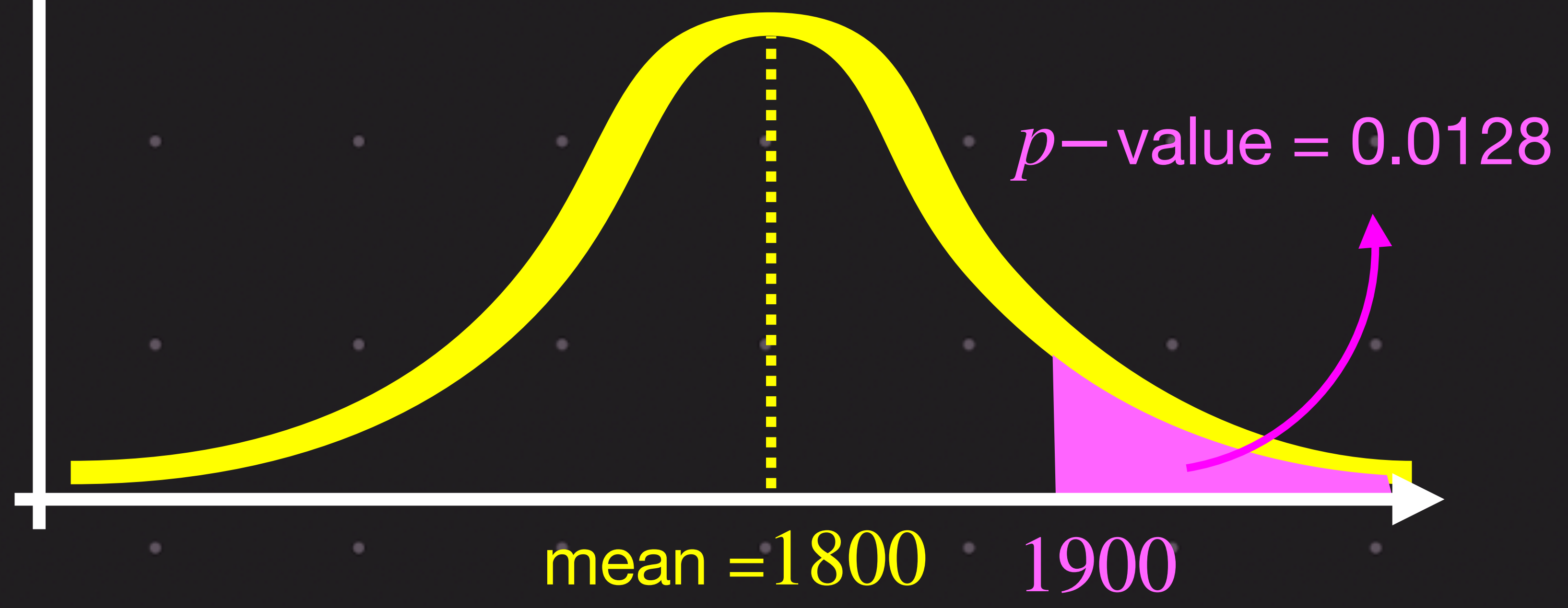
Reject H_0

5 stores with average of 1900

$H_0 : \mu_y = 1800$

$H_a : \mu_y > 1800$

Distribution of m_y under the assumption that H_0 is true



std dev = $\dfrac{100}{\sqrt{5}}$

$$z = \frac{1900 - 1800}{100/\sqrt{5}} = 2.23$$

Fail to reject H_0

Supply chain example

$\alpha = 0.01$

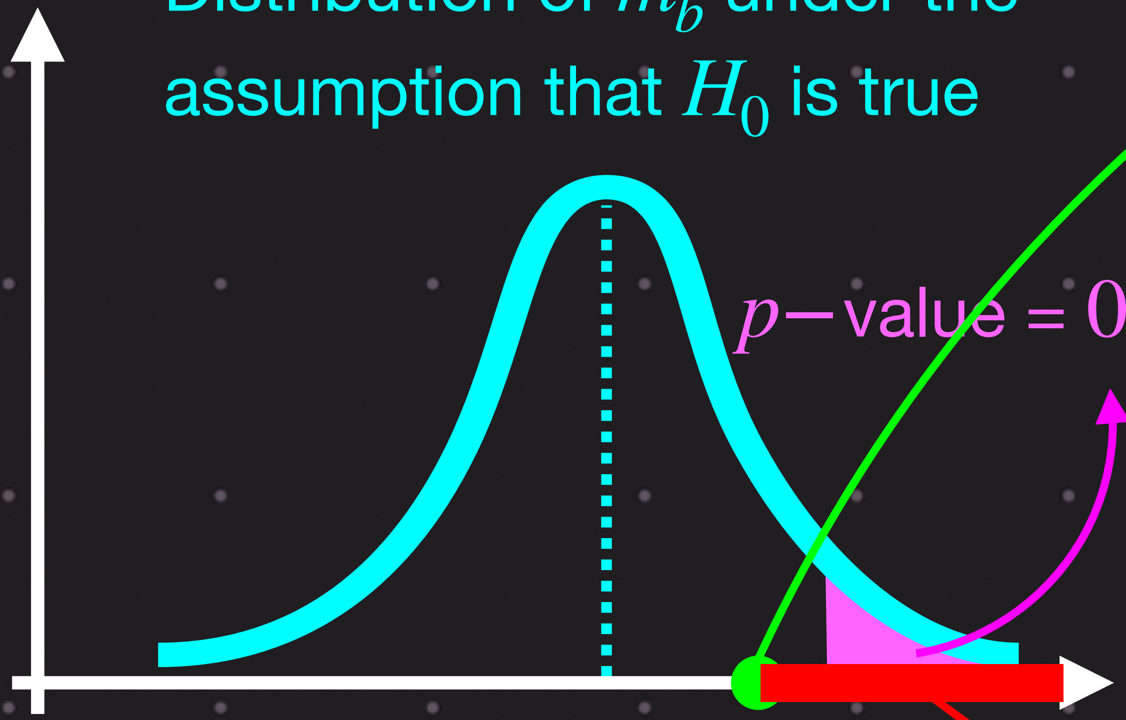
$\mu = 1800$
$\sigma = 100$

50 stores with average of 1850

$H_0 : \mu_b = 1800$

$H_a : \mu_b > 1800$

Distribution of m_b under the assumption that H_0 is true



$p-$value $= 0.0002$

mean $= 1800$

1850

1832.8

std dev $= \dfrac{100}{\sqrt{50}}$

$z = \dfrac{1850 - 1800}{100/\sqrt{50}} = 3.53$

What should be the z-score such that we can reject if mean is larger, and accept if mean is lesser?

We want only 1% area to the right

Z-score of upper critical value = `norm.ppf(0.99)` $= 2.32$

Upper critical value = $1800 + 2.32 * \dfrac{100}{\sqrt{50}} = 1832.8$

To summarise, if we are testing for 50 samples, we can reject the null hypothesis only if the average sales is greater than 1832.8

This region is called the “critical region”

Supply chain example

$\alpha = 0.01$

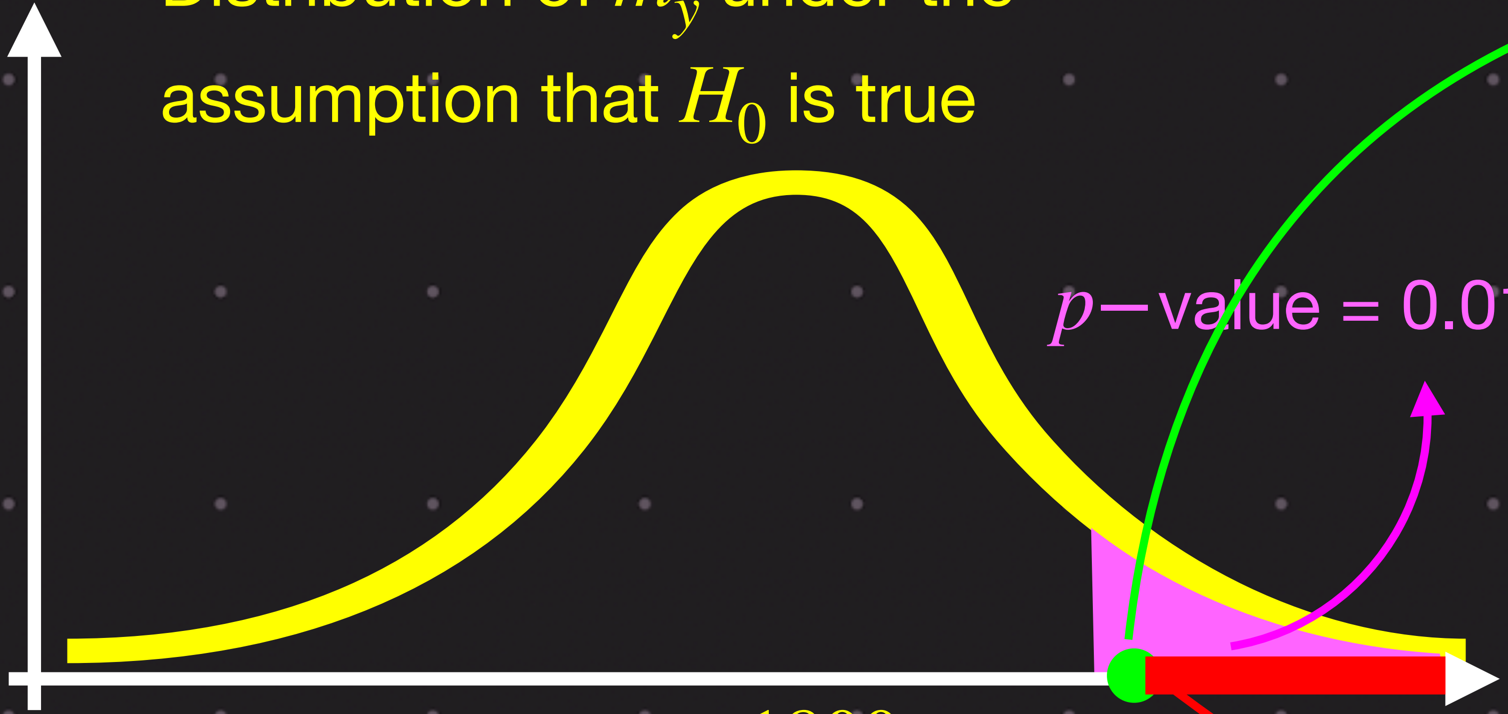
$\mu = 1800$
$\sigma = 100$

5 stores with average of 1900

$H_0 : \mu_y = 1800$

$H_a : \mu_y > 1800$

Distribution of m_y under the assumption that H_0 is true



std dev = $\dfrac{100}{\sqrt{5}}$

$z = \dfrac{1900 - 1800}{100/\sqrt{5}} = 2.23$

What should be the z-score such that we can reject if mean is larger, and accept if mean is lesser?

We want only 1% area to the right

Z-score of upper critical value = `norm.ppf(0.99)` $= 2.32$

Upper critical value = $1800 + 2.32 * \dfrac{100}{\sqrt{5}} = 1903.7$

To summarise, if we are testing for 5 samples, we can reject the null hypothesis only if the average sales is greater than 1903.7

This region is called the “critical region”

Supply chain example

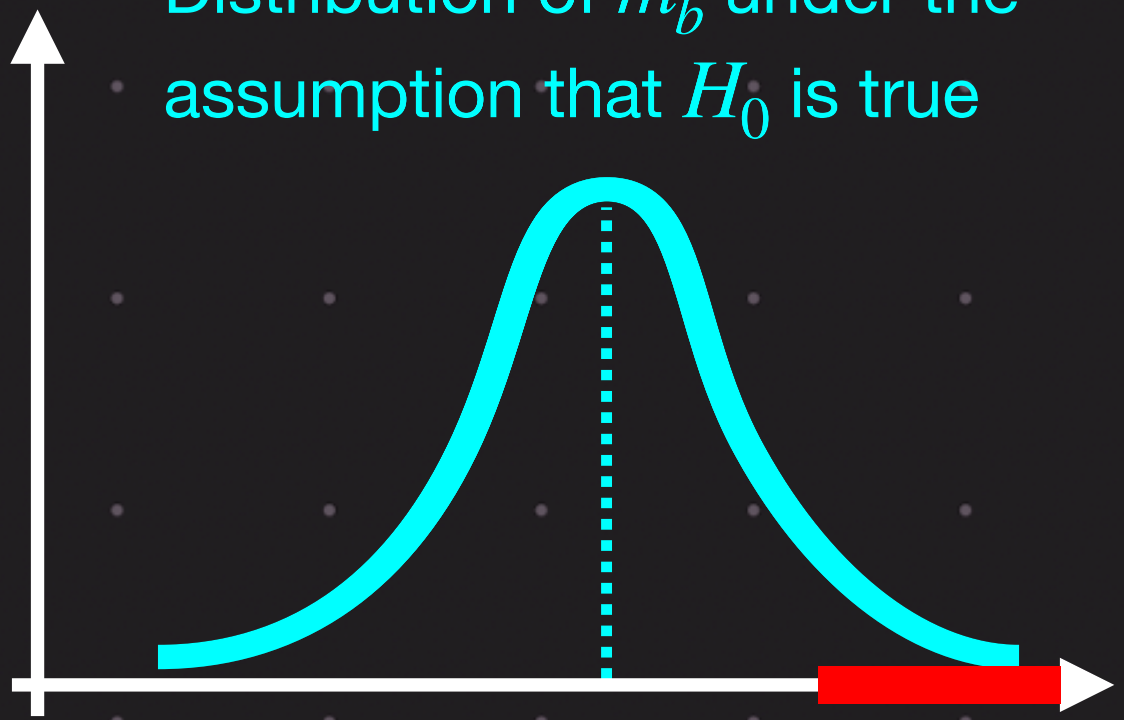
$\alpha = 0.01$

$\mu = 1800$
$\sigma = 100$

50 stores

$H_0 : \mu_b = 1800$
$H_a : \mu_b > 1800$

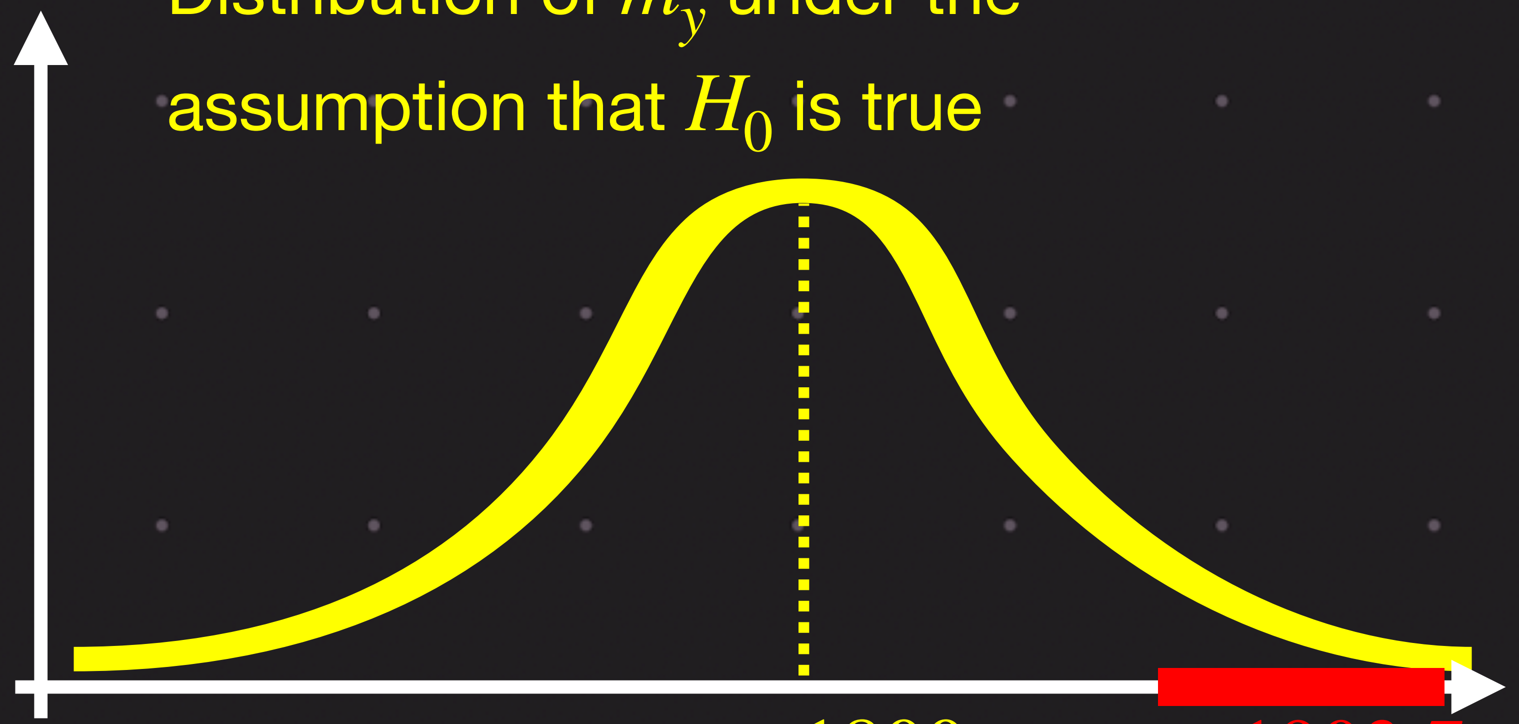
Distribution of m_b under the assumption that H_0 is true



5 stores

$H_0 : \mu_y = 1800$
$H_a : \mu_y > 1800$

Distribution of m_y under the assumption that H_0 is true



Note: For right-tailed test, the critical region is on the right
The probability associated with critical region is α

The rule to reject is very simple: If the observed test statistic is in the critical region, then reject the null hypothesis