Supply chain example

$\alpha = 0.01$

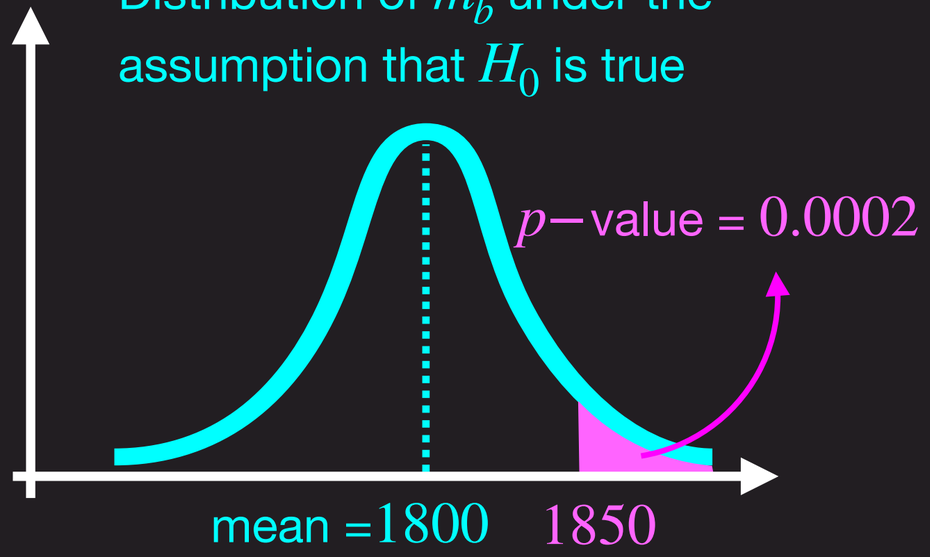
$\mu = 1800$
$\sigma = 100$

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



Test statistic = $\dfrac{m - \mu}{\sigma/\sqrt{n}}$

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

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

Reject H_0

Supply chain example

50 stores

$H_0 : \mu_b = 1800$

$H_a : \mu_b > 1800$

Suppose we have access to the sales data after one teams marketing strategy

In practice, we use the data to infer the sample mean and standard deviation to conduct the hypothesis test

This is an example of the Z-Test

Let us see this implementation in Python

```
from statsmodels.stats.weightstats import ztest
```

Drug Recovery

Suppose two companies develop a drug for a disease.

Drug 1 was tested on 100 people, and the recovery days look like this

$[8, 5, 9, 10, \cdots, 16]$ The mean recovery time was 7.1 days

Drug 2 was tested on 120 people, and the recovery days look like this

$[12, 4, 7, 13, \cdots, 8]$ The mean recovery time was 8.07 days

Can we say one drug was better than the other?

Or was this small difference a coincidence?

For such cases we use the two-sample Z-test

```
from statsmodels.stats.weightstats import ztest
```

IQ across two schools

Suppose there are two schools competing against each other

Each school says their students have higher IQ

So we conduct a test

Say the first school had numbers like this

$[101,\ 115,\ 90,\ \cdots,\ 112,\ 97]$ Let us say average was 103.7

And yellow team has these numbers

$[108,\ 105,\ 99,\ \cdots,\ 111,\ 98]$ Let us say average was 102.9

Is there a statistical significance in this difference?

Or was it just chance?

For such cases we use the two-sample Z-test

One sample Test

Supply chain: We compared average sales in 50 stores against a fixed number: 1800

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

Premature children IQ We compared average IQ of 50 prematurely born children fixed number: 1800

$H_0 : \mu = 100$ $H_a : \mu \neq 100$

Drug Vs No drug Patients recover on average 15 days without drug. Will it reduce after taking the drug?

$H_0 : \mu = 15$ $H_a : \mu < 15$

We are comparing one set of samples against a fixed number (population mean)

Two sample test

Supply chain: We directly compare two marketing teams

$H_0 : \mu_b = \mu_y$ $H_a : \mu_b \neq \mu_y$

Two states IQ Two competing states, where each state says their students have higher IQ

Let μ_1 be the average IQ of first state, and μ_2 be the average IQ of another state

$H_0 : \mu_1 = \mu_2$ $H_a : \mu_1 \neq \mu_2$

Two Drugs Let μ_1 be the average recovery for drug 1, and μ_2 be the average for drug 2

$H_0 : \mu_1 = \mu_2$ $H_a : \mu_1 \neq \mu_2$

We are comparing two sets of samples against each other

Youtube ads

Youtube wants to increase its ads revenue. They decide to put up 2 ads instead of one

Is this a good move? This could lead to a decrease in the average watch time

Can we test this effect? We need two groups: control and treatment

Control group sees one ad

Watch times of control group are like this

$[3.4,\ 2.4,\ 1.7,\ 0.4,\ \cdots,\ 4.2]$ Mean: $m_1 = 3.6$

Treatment group sees two ads

Watch times of the treatment group are like this

$[3.5,\ 3.2,\ 2.5,\ 0.1,\ \cdots,\ 3.1]$ Mean: $m_2 = 3.05$

1) Setup the hypothesis $H_0 : \mu_1 = \mu_2$

$H_a : \mu_1 \neq \mu_2$

2) Choose the right test statistic

$m_1 - m_2$? ✘

$$\frac{m_1 - m_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$ ✓

3) Right/Left/Two-tailed? Two-tailed

4) Compute p-value

5) Compare p-value with α

If we are sure that the effect will only reduce the time, we can go for “greater” alternative

$H_0 : \mu_1 = \mu_2$

$H_a : \mu_1 > \mu_2$

Right-tailed