Are You Sure ?

On testing your Hypothesis

Outline

0. The Question
1. Burgers
2. Graphics
3. A/B Test
4. Summary

The Question

red button

blue button

Will the conversion rate be different?

I have a **hypothesis**

No difference

Burgers

Are the burgers being sold at the same price?

My Hypothesis:

The burgers being sold at Hans im Glück and Freddy's have the same price.

Go buy a burger

Hans im Glück		Freddy's
€8	<	€9

Is my hypo. wrong?

Need more samples.

	Hans im Glück	Freddy's	Δ
Köln	8	9	−
Berlin	7	8	−
Freiburg	9	9.5	−
Manheim	7.5	7	+
München	9	9.5	−
Offenburg	7	6.5	+
Hamburg	8	9	−
Frankfurt	9	9.5	−
Bonn	8	9	−
Siegen	7	8	−

Out of 10

+ : 2

− : 8

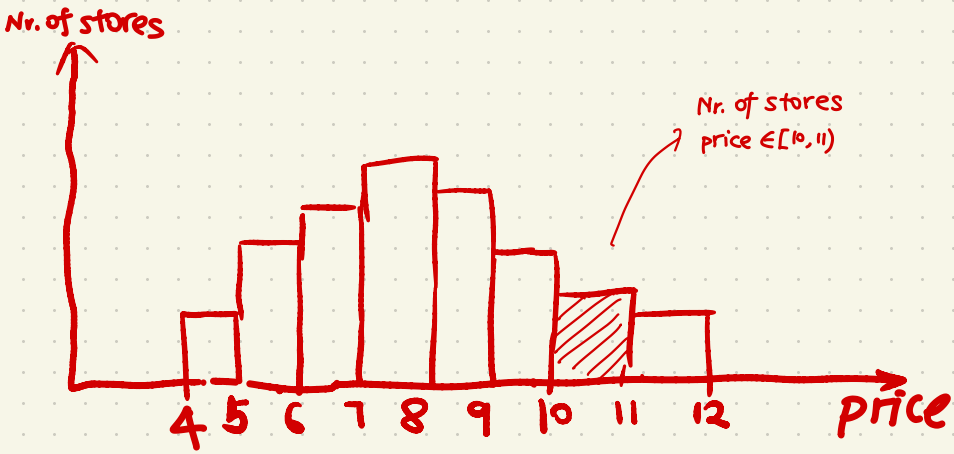
I am wrong! With more confidence!

From ~~records~~ tables

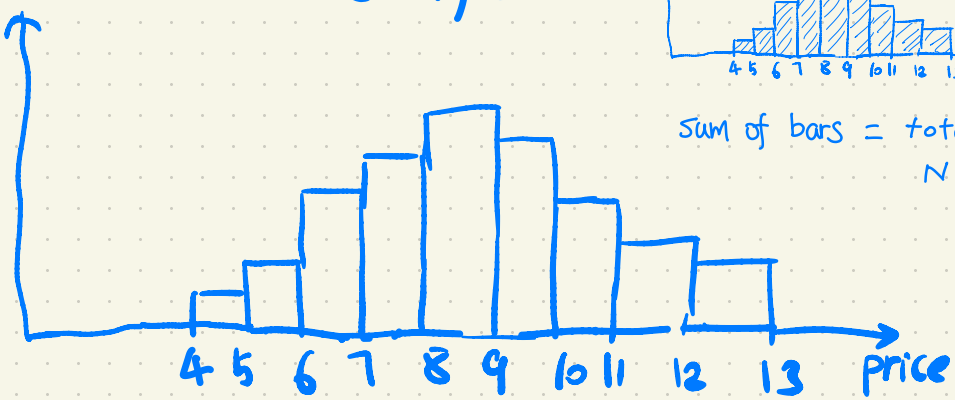
to ~~distributions~~ graphics

Histograms

Hans im Glück



Freddy's

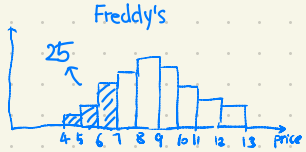


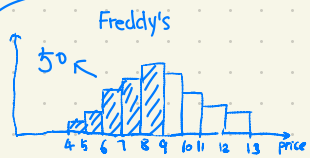
sum of bars = total exp.

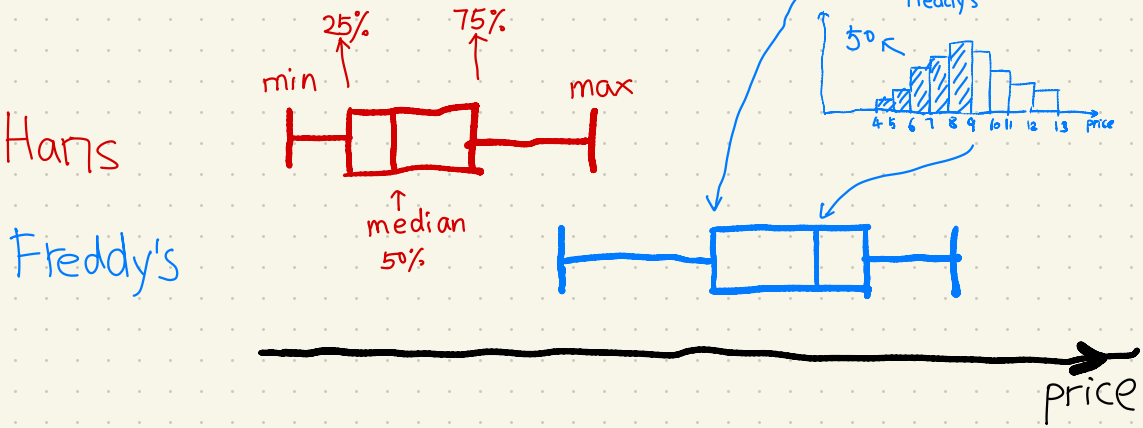
N

Box

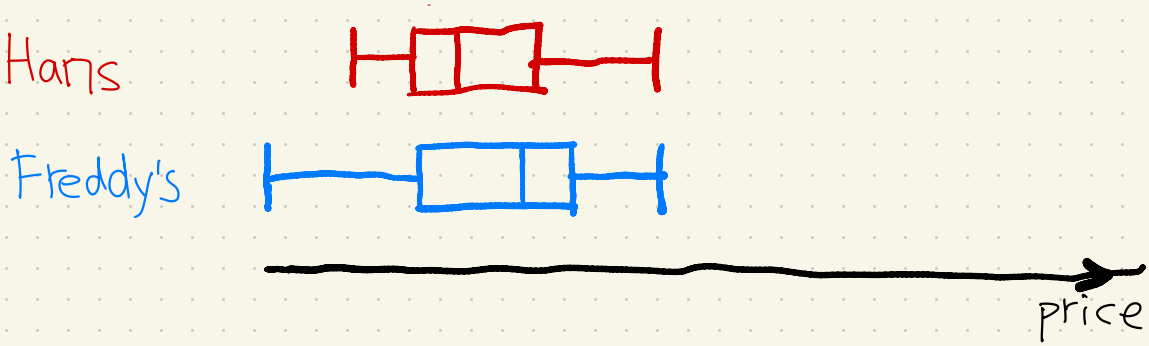
If in total 100 stores







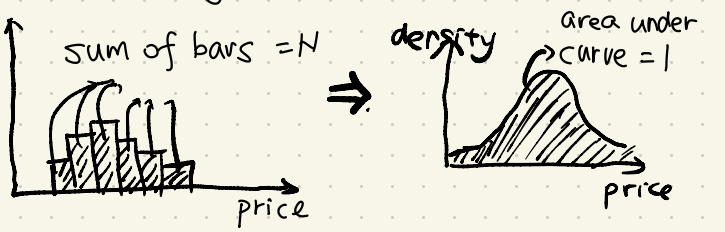
Freddy's is more expensive

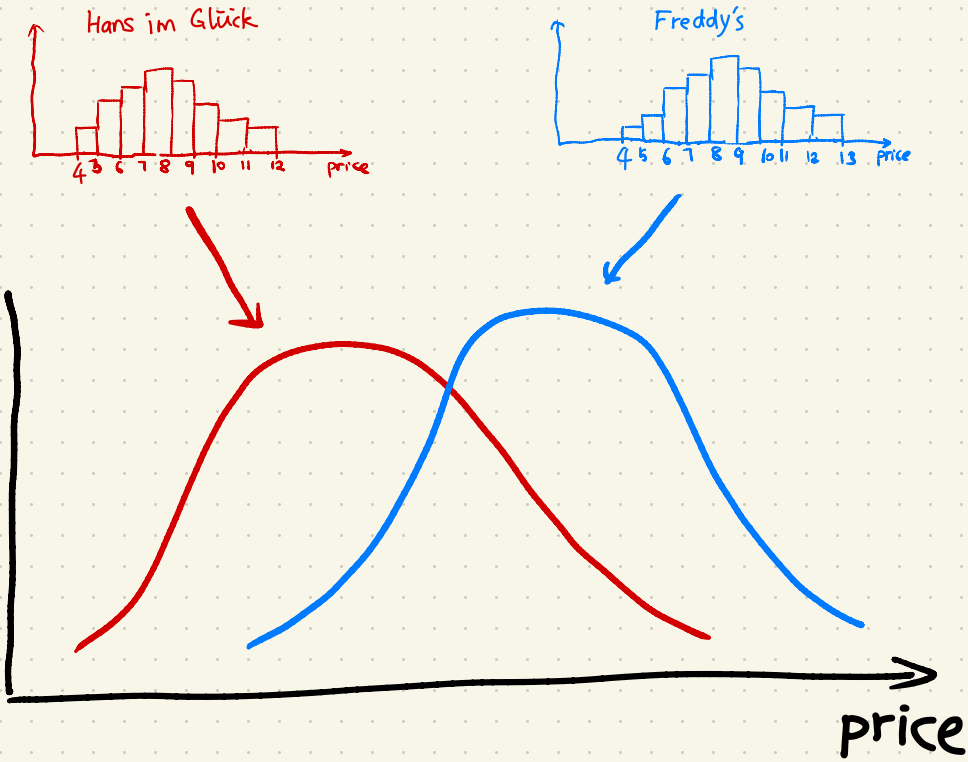


not sure

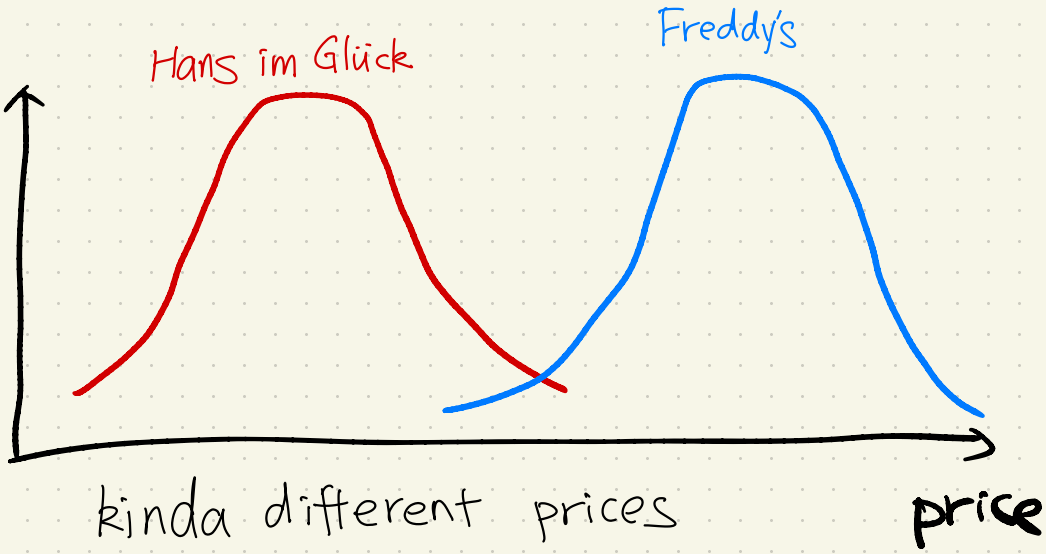
Density

From histogram to prob. distributions

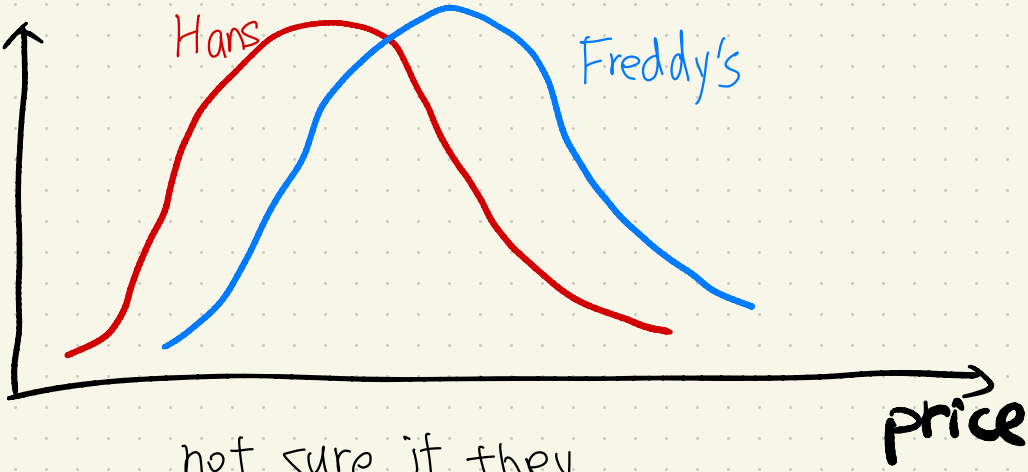




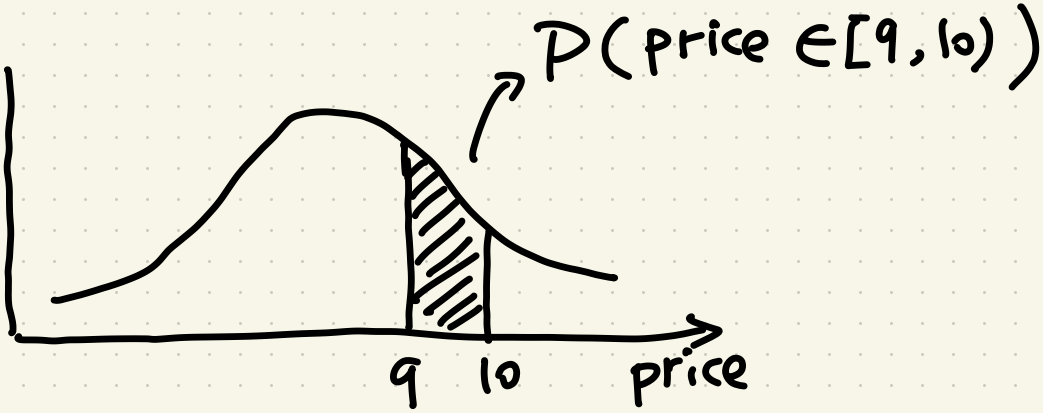
Naive ideas

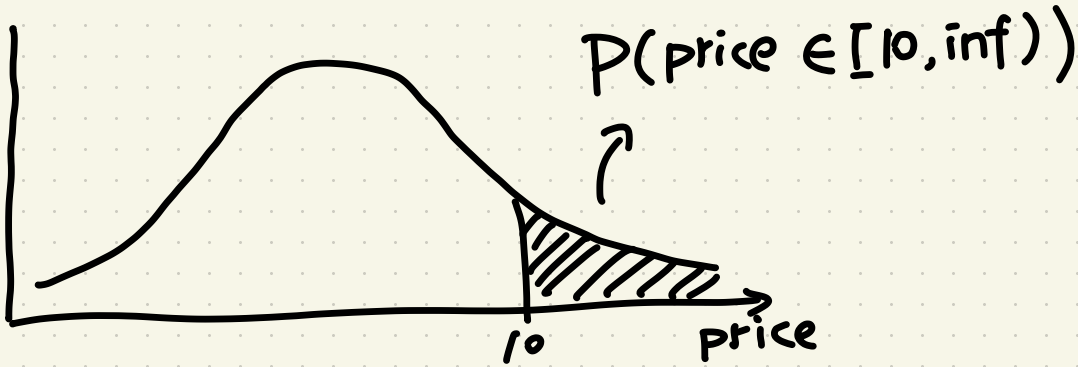


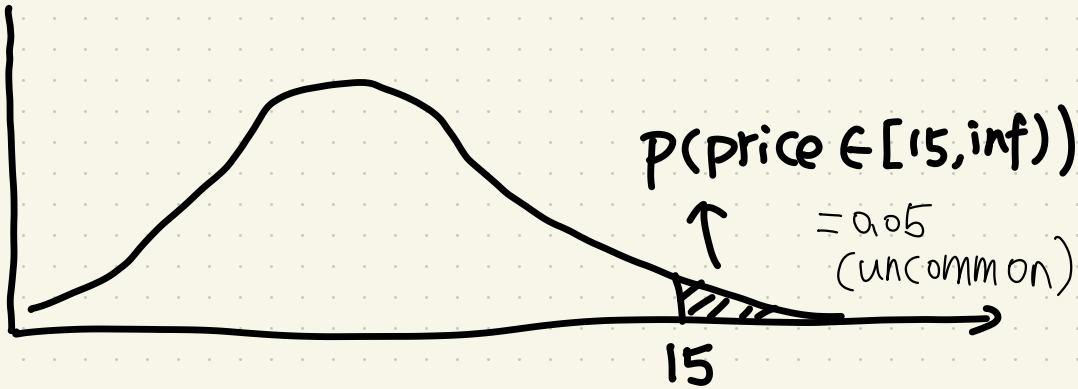
kinda different prices

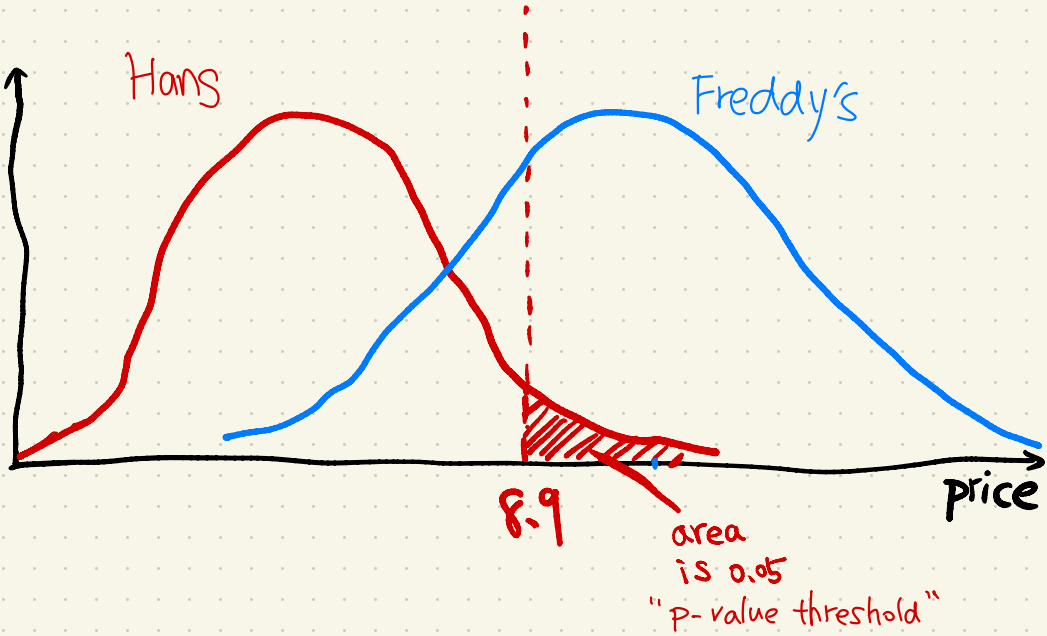


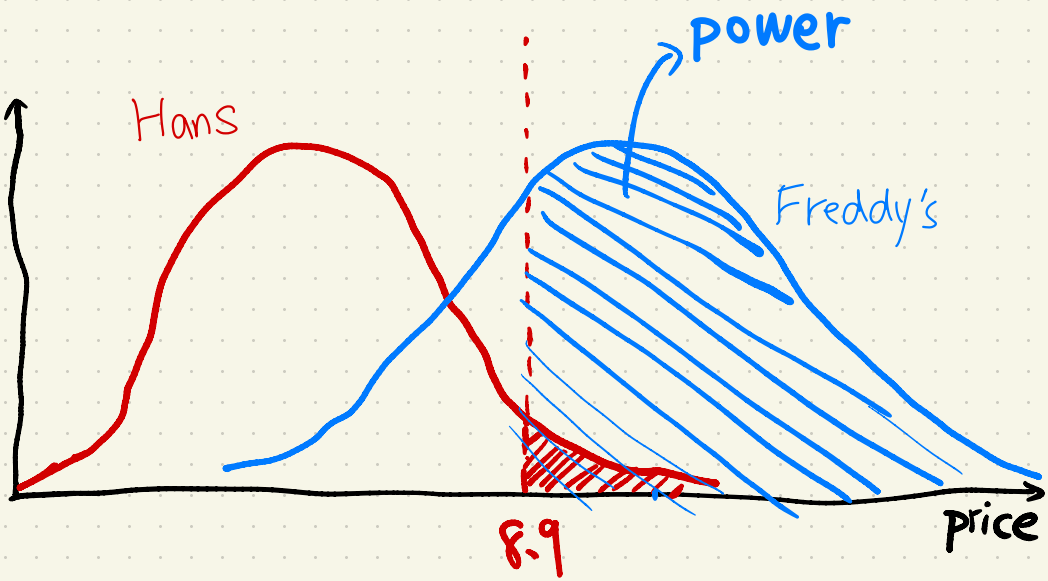
not sure if they
have different prices

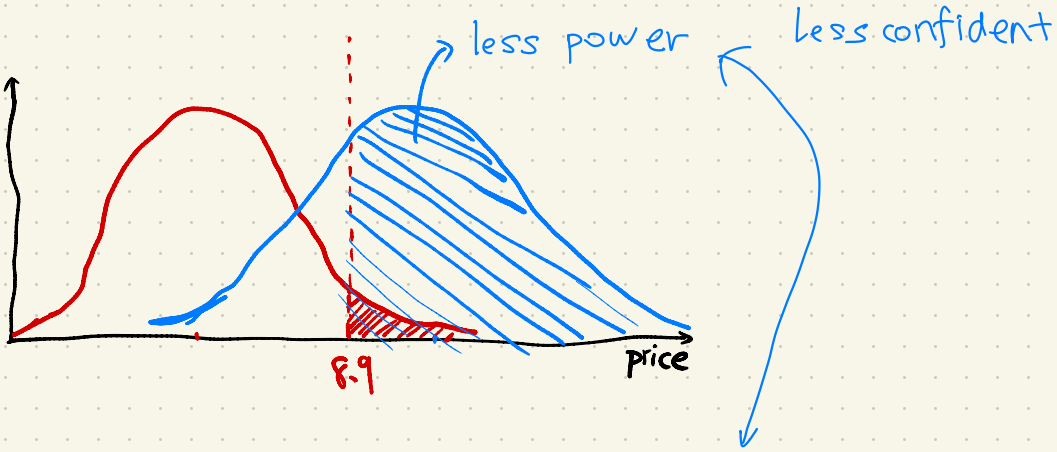

P(price ∈ [9, 10))
9
10
price

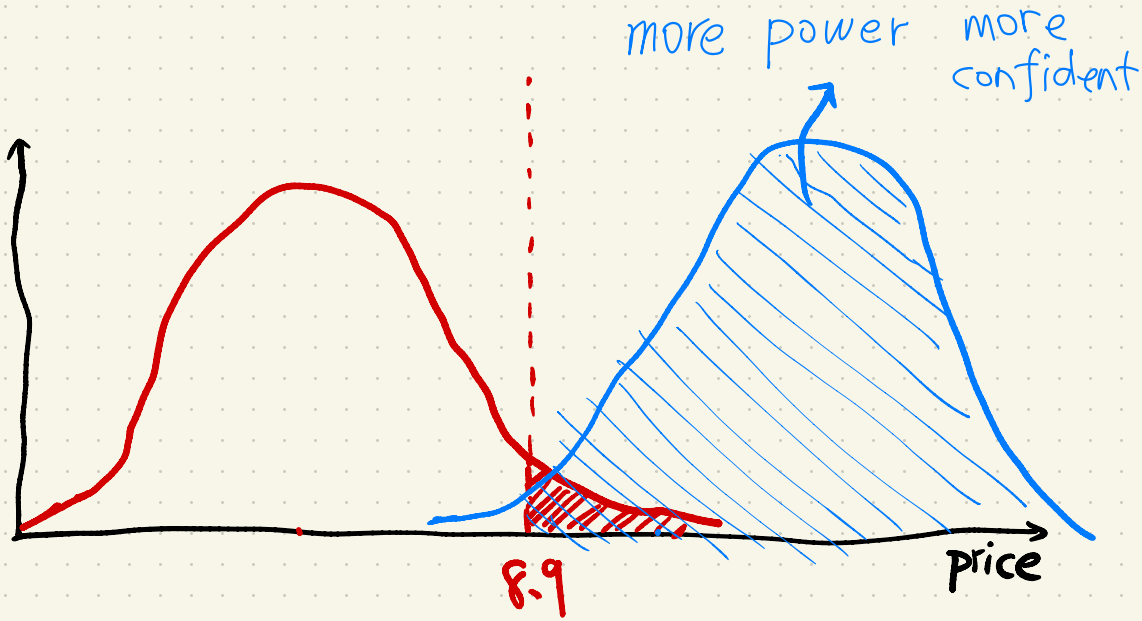

P(price ∈ [10, inf))
10
price


P(price ∈ [15, inf))
= 0.05
(uncommon)
15


Hans
Freddy's
8.9
price
area
is 0.05
"p-value threshold"

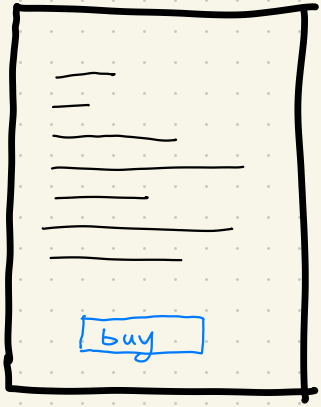

power
Hans
Freddy's
8.9
price


less power
less confident
8.9
price


more power
more confident
8.9
price

A/B Test

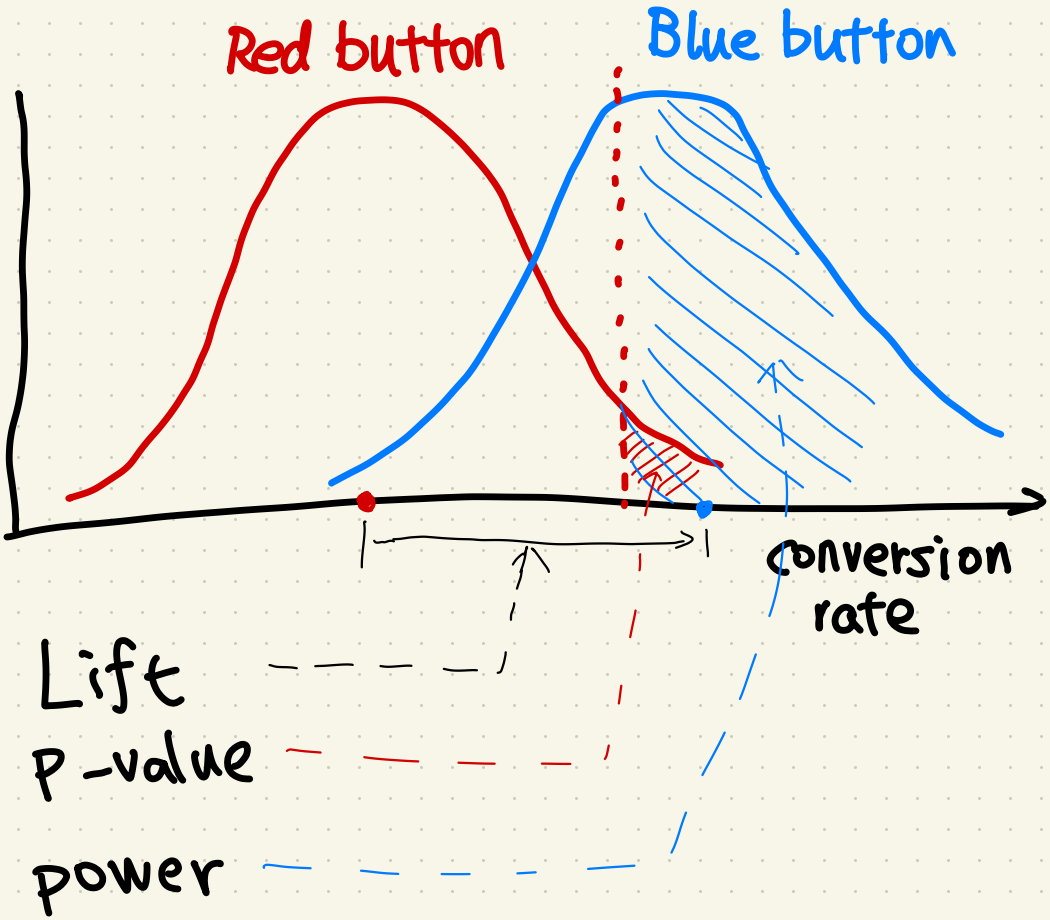
red button

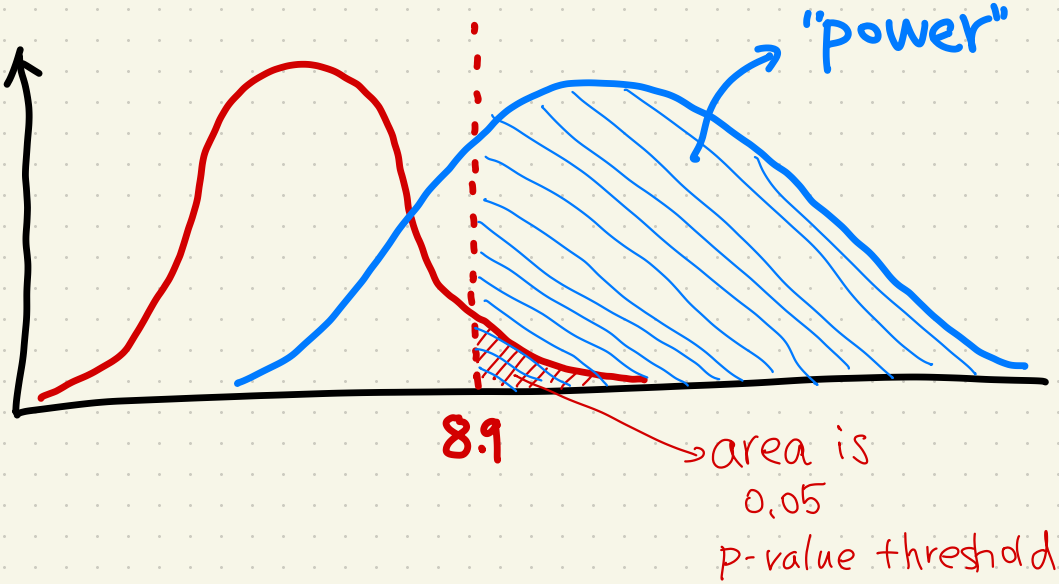


blue button

1. Split users into 2 groups, RANDOMLY

2. Assign each group to different designs.


Red button
Blue button
conversion rate
Lift
p-value
power



p-value → P(price ≥ 8.9)

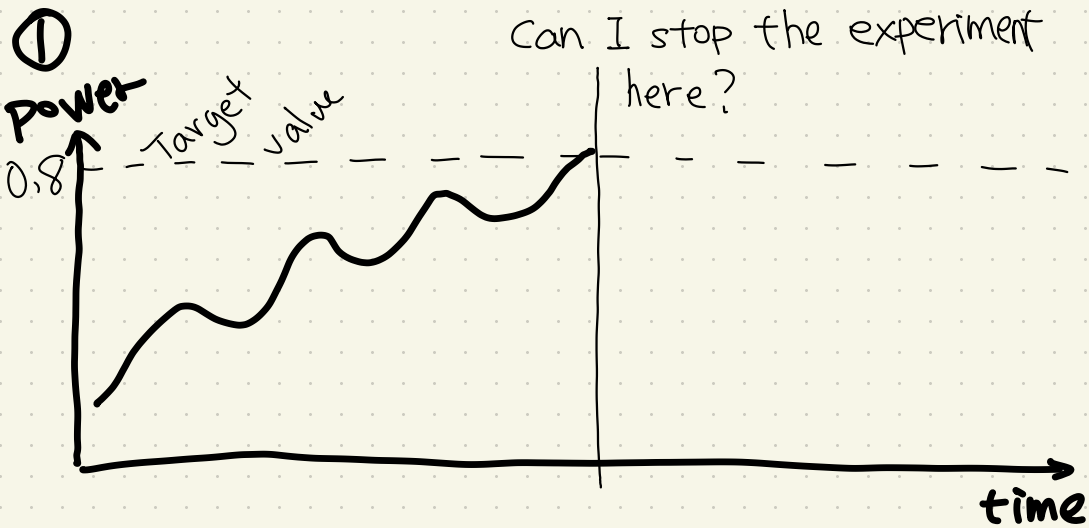
in Hans im Glück price distribution

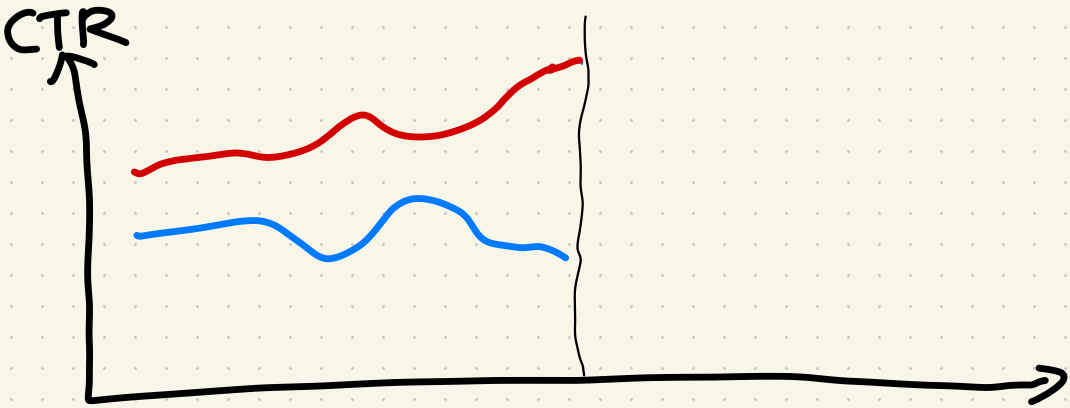
power = P(price ≥ 8.9)

in Freddy's price distribution

Comments

①

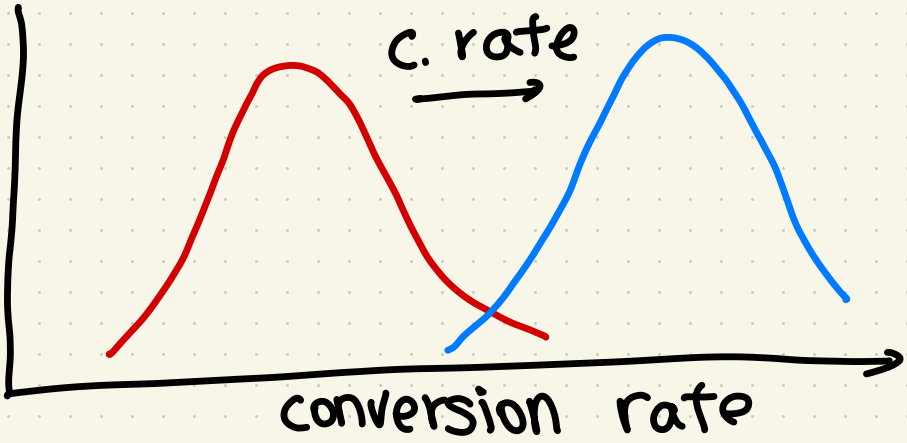


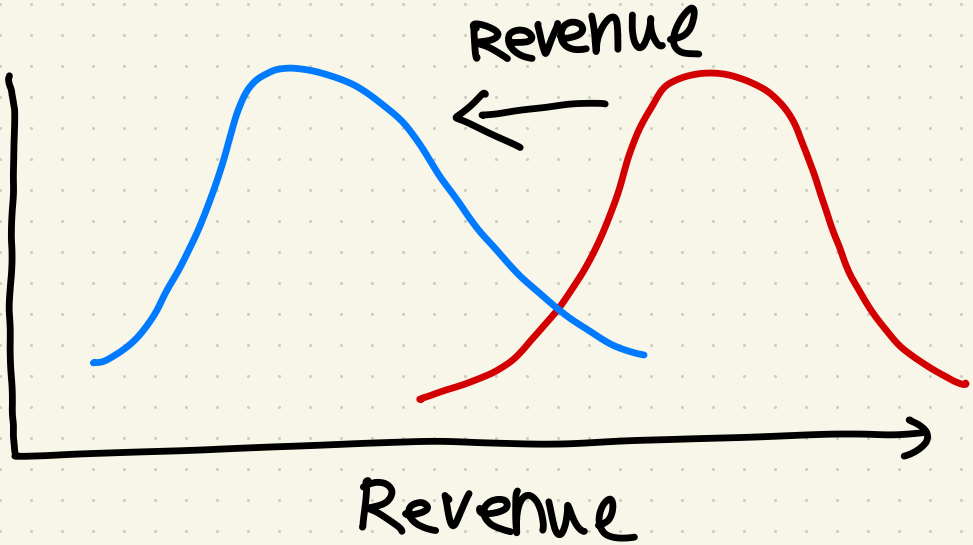


Calculate sample size
before
starting the experiment

μ,	p-value threshold
Lift	, power

② Which KPI to watch



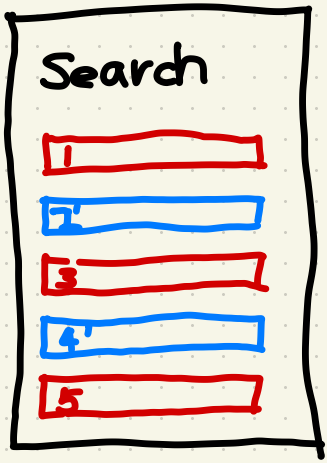


③ Ranking Problem

two ranking algorithms

Algorithm 1

Algorithm 2



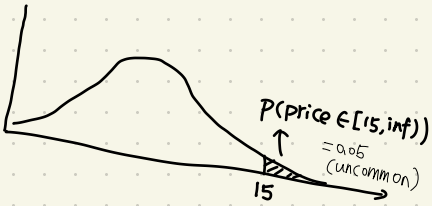
interleaving test

Sumary

① We work on samples.

② It's probabilistic.

③ Rare events.



More
math
More
graphics

and
making
mistakes

Null Hypothesis H_0

Burger price are the same at Hans im Glück and Freddy's

errors

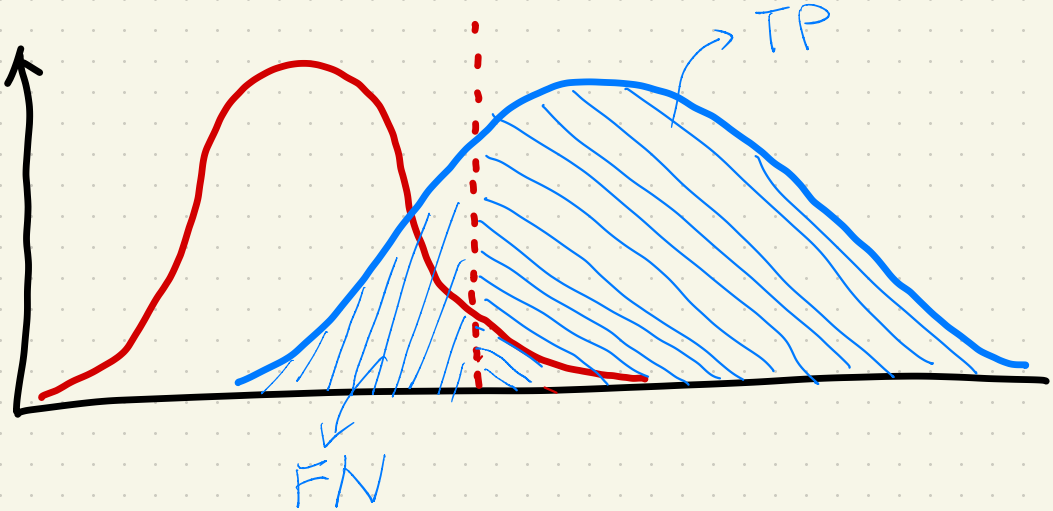
	H_0 is true	H_0 is false
accept H_0	✓	× (Type II)
reject H_0	× (Type I)	✓

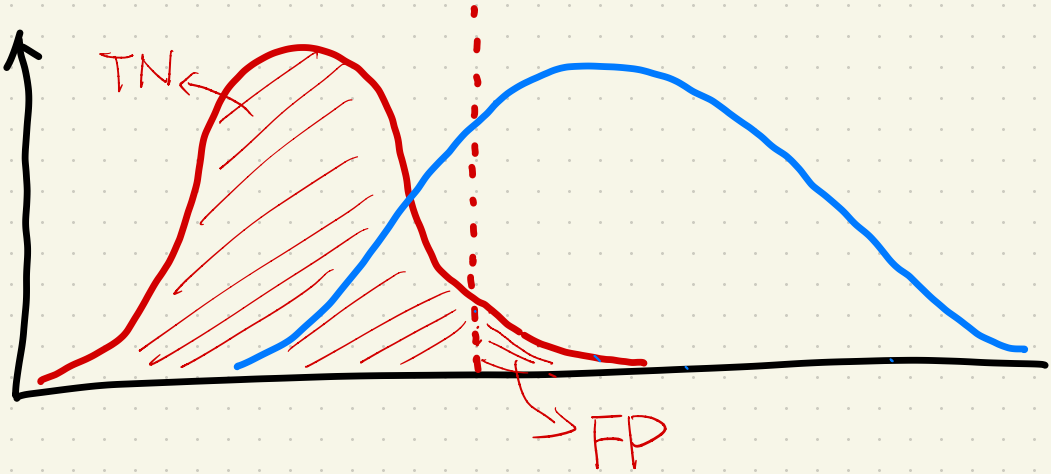
× (Type II)

I conclude that there is no difference while there is actually a difference

× (Type I)

I conclude that there is a difference while there is actually no difference


TP
FN

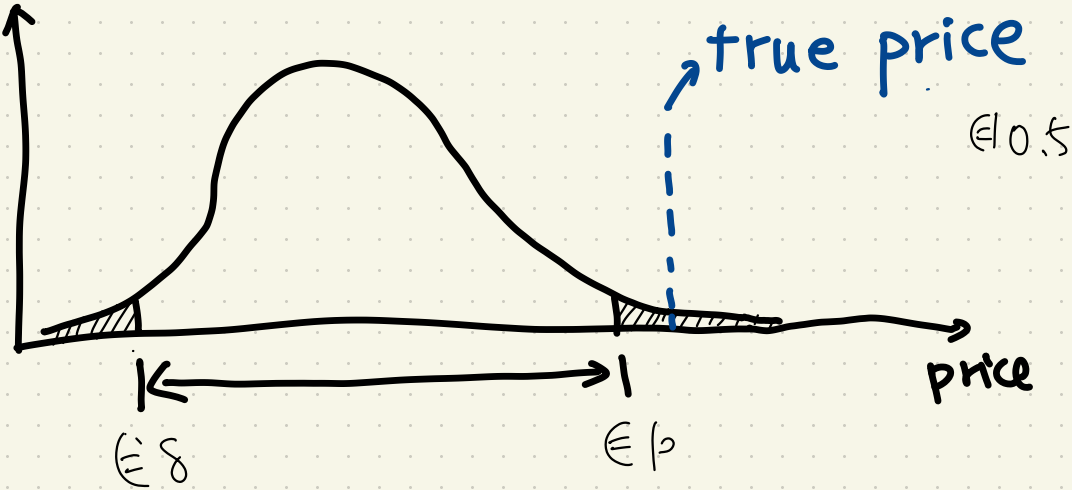

TN
FP

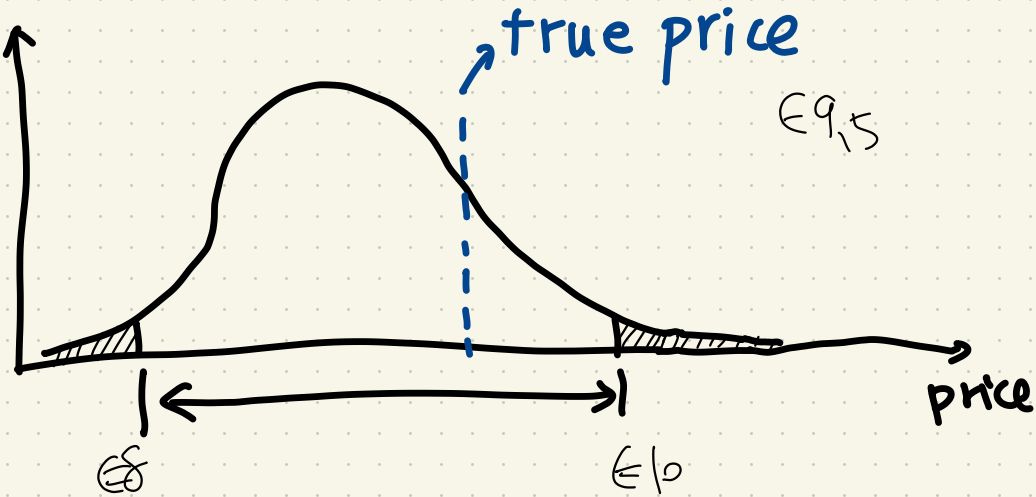
why is it called power?

$\beta \rightarrow FN$

$\text{Power} = 1 - \beta \rightarrow TP$

confidence interval





$$\text{p-value} \propto \overset{0.5}{P(+)}^{\overset{2}{N_+}}\, \overset{0.5}{P(-)}^{\overset{8}{N_-}}$$

$$\text{p-value} = 45 \cdot \frac{1}{2^{10}}$$

$$= 0.044$$

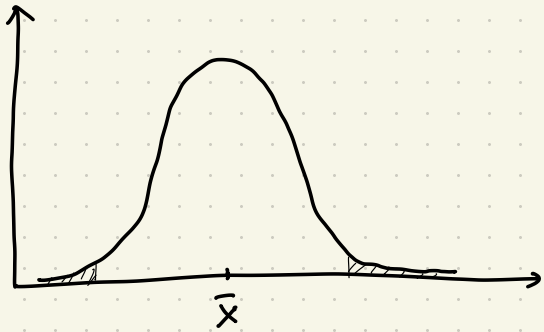
$$< 0.05$$

	Hans im Glück	Freddy's	Δ
Köln	8	9	−
Berlin	7	8	−
Freiburg	9	9.5	−
Mannheim	7.5	7	+
München	9	9.5	−
Offenburg	7	6.5	+
Hamburg	8	9	−
Frankfurt	9	9.5	−
Bonn	8	9	−
Siegen	7	8	−

Comments

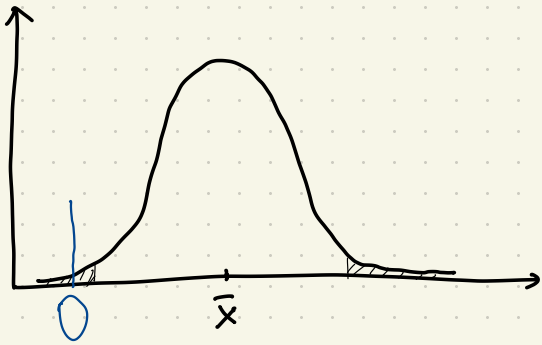
Alternative test method

Red button	Blue button	Δ
0.1	0.15	-0.05
0.3	0.2	0.1
0.15	0.25	-0.1
0.15	0.3	-0.15
0.4	0.25	0.15
0.2	0.3	-0.1

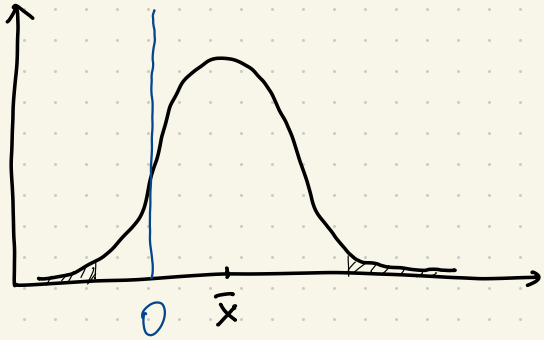


Hypothesis:

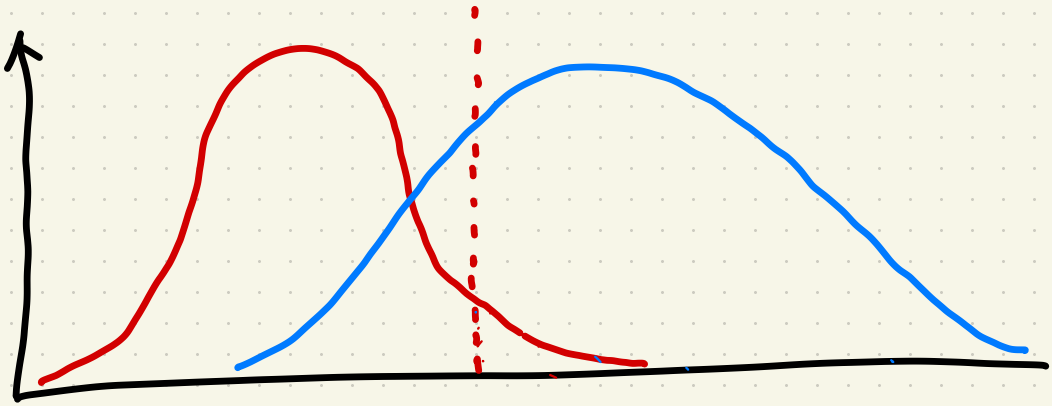
$\bar{x} = 0$

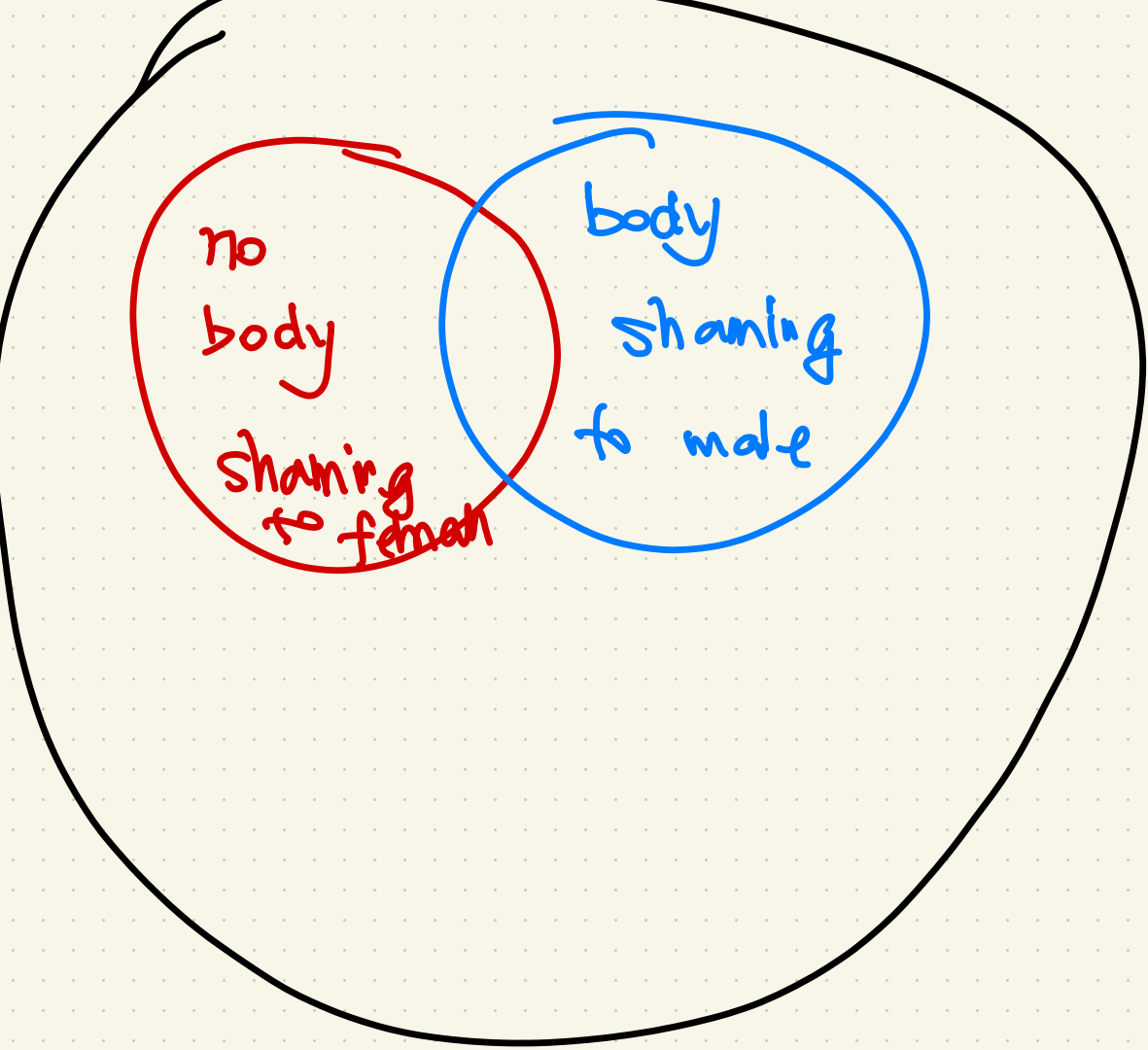


Hypothesis is wrong



Keep the hypothesis





	no bs → female	bs doesn't mal
bs male		
doesn't bs..		