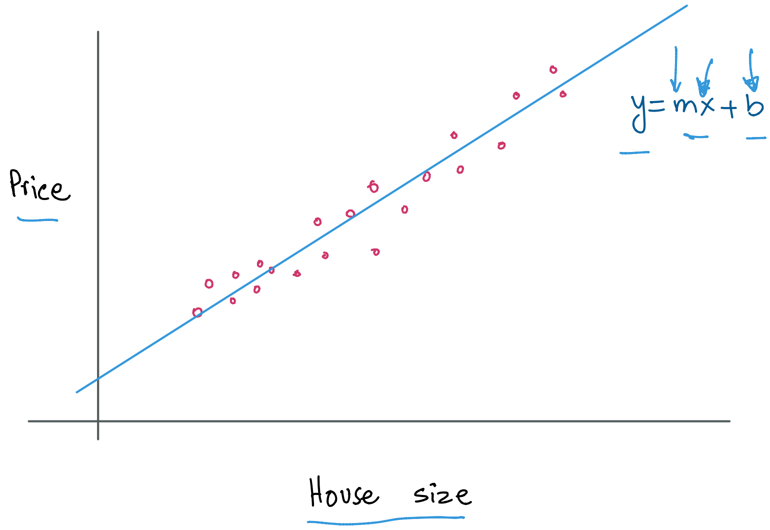
Predicting house prices



To find the relation between X and Y, we need m and b

"Regression"

25

"point estimates"

1.9

Formulae:

$$b = \frac{\Sigma y \Sigma x^2 - \Sigma x \Sigma xy}{n\Sigma x^2 - (\Sigma x)^2}$$

$$m = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

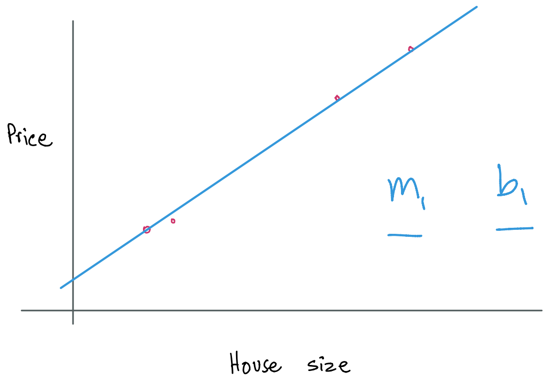
	x	y	xy	x^2	y^2
	⋮	⋮	⋮	⋮	⋮
$\Sigma =$	...	...	...	...	...

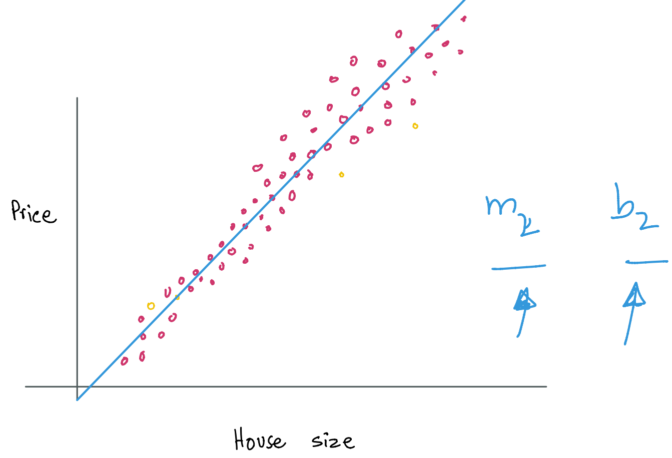
We do something very similar in machine learning!

But!

m and b are both "point estimates"

They say nothing about the uncertainty!





The problem is $\{m_1, b_1\}$ here and $\{m_2, b_2\}$ here are equally "confident".

This is a major problem in classical statistics!

Model parameters are learned through observations alone and they are point estimates!

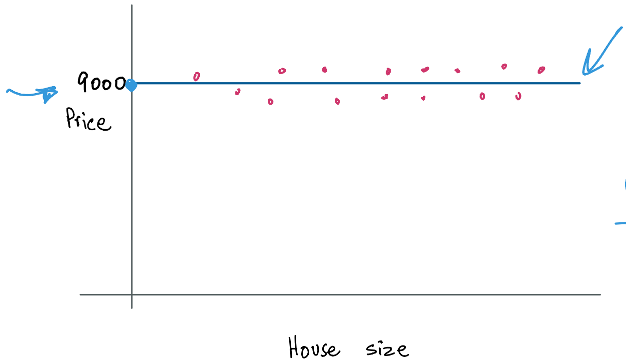
(multiverse shenanigans) χ-square, t-testing

So, what's the solution?

Bayesian inference!

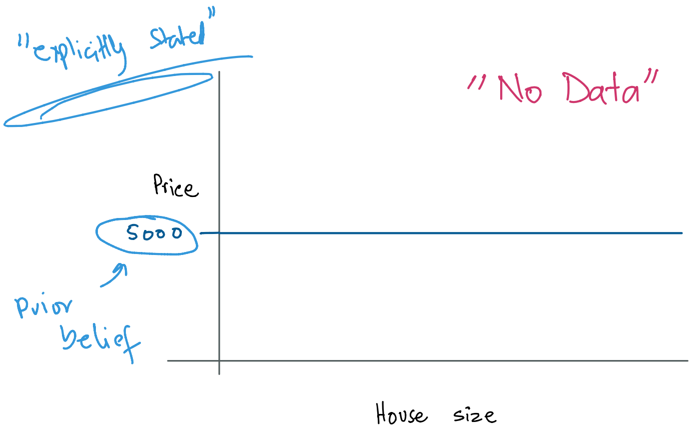
- Start with a prior distribution
- Update prior to posterior distribution based on evidence

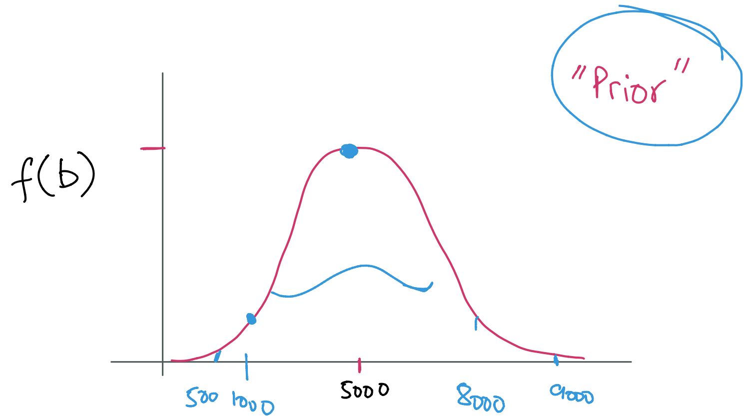
Let's ignore the slope for now and focus on "b"

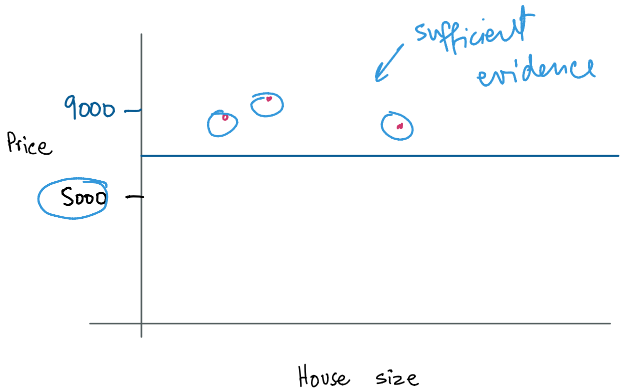


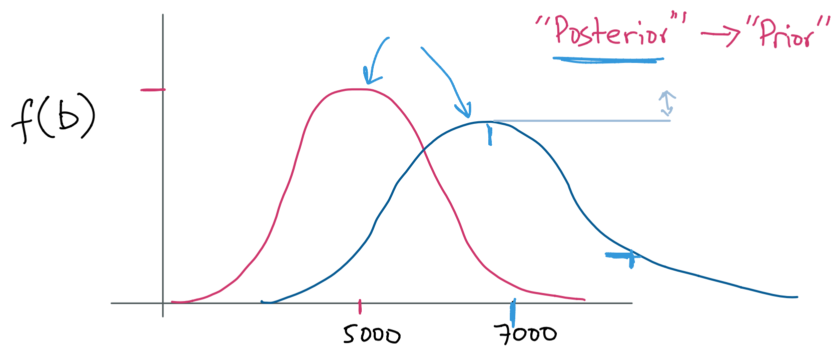
Our data will eventually look like this

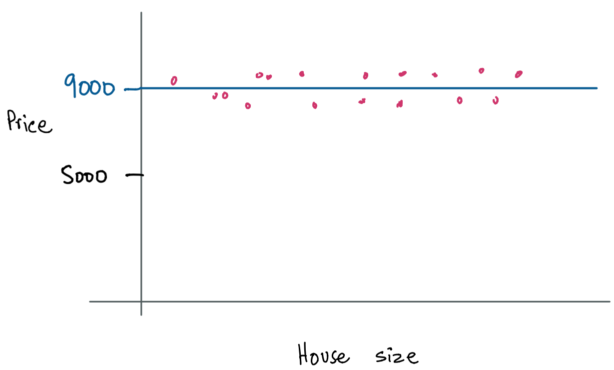
Ground truth

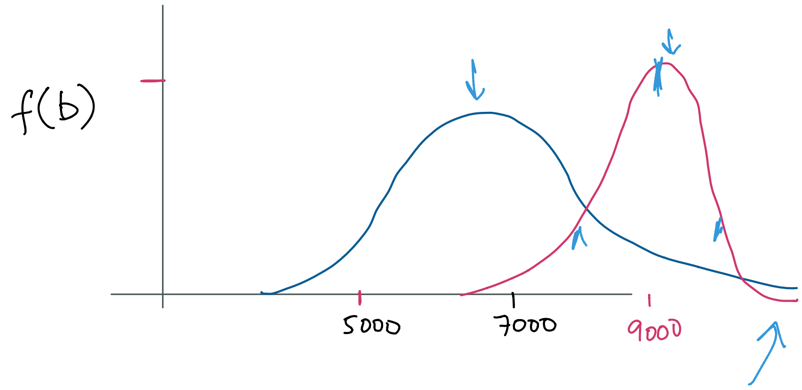


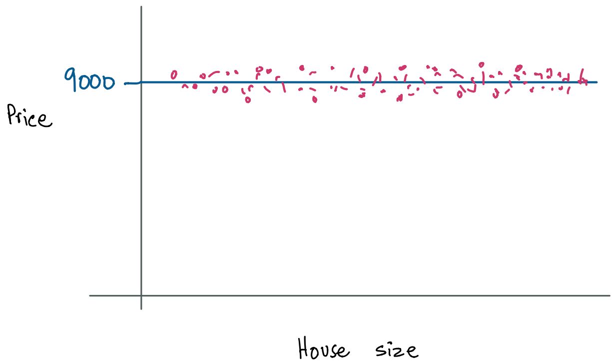


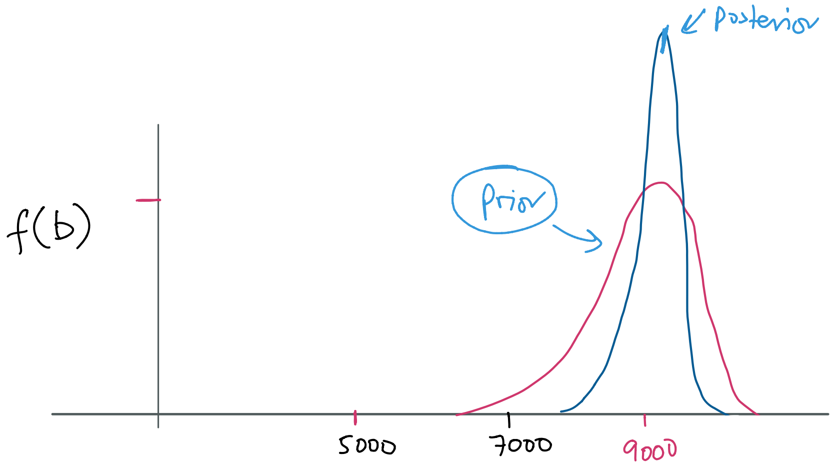












Notice that :

"b" still has a distribution. We have the uncertainty quantified — at all times!