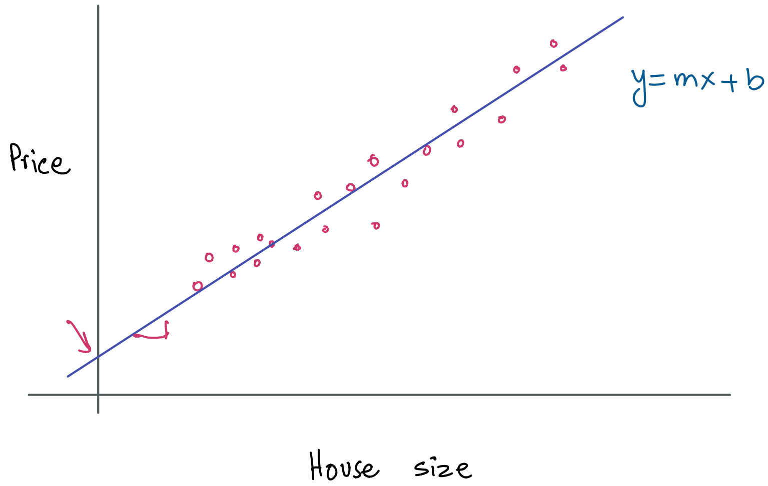
Predicting house prices



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

Formulae:

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

Regression

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

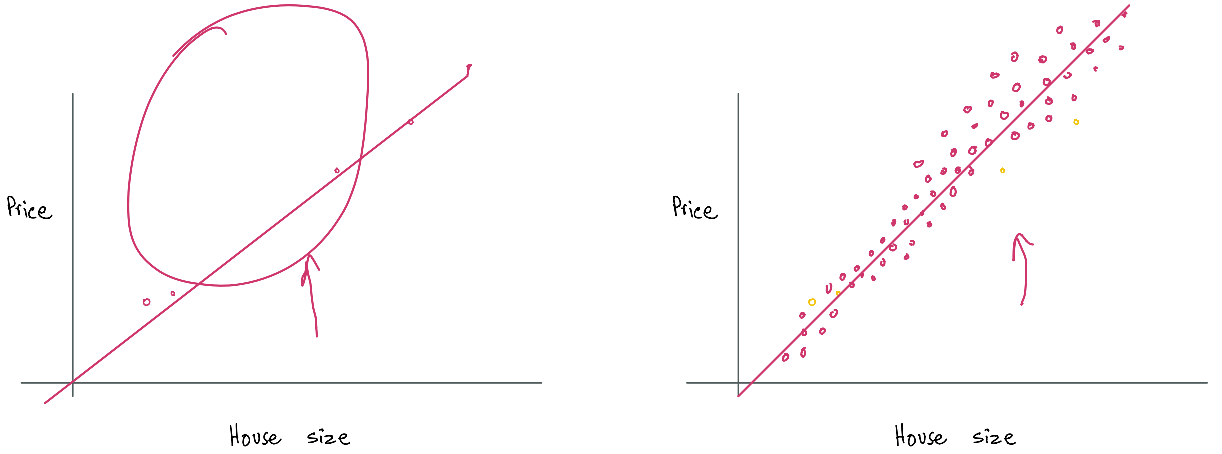
x	y	xy	x^2	y^2
⋮	⋮	⋮	⋮	⋮
$\Sigma =$ ...	...	...	...	...

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!

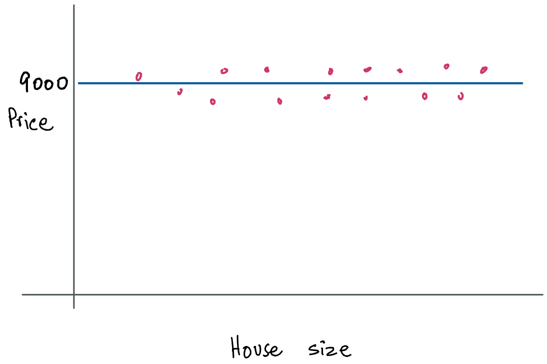
(multiverse shenanigans)
p-value

So, what's the solution?

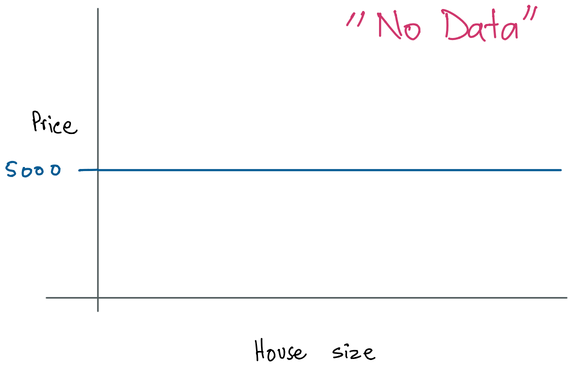
Bayesian inference!

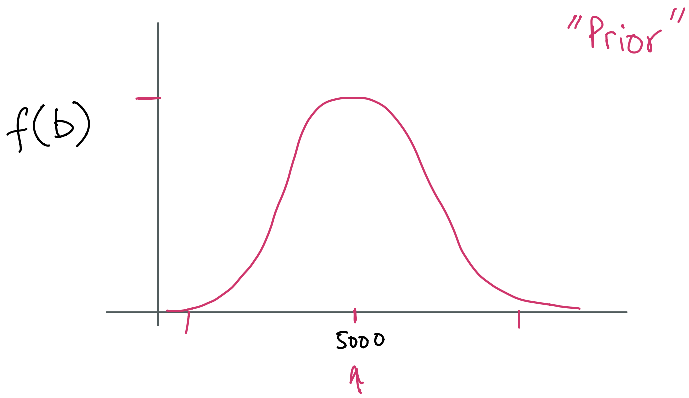
- Start with a prior distribution
- Update prior to posterior distribution based on evidence

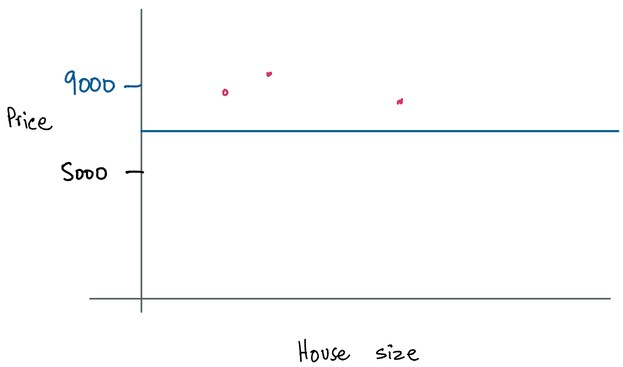
Let's ignore the slope for now and focus on "b"

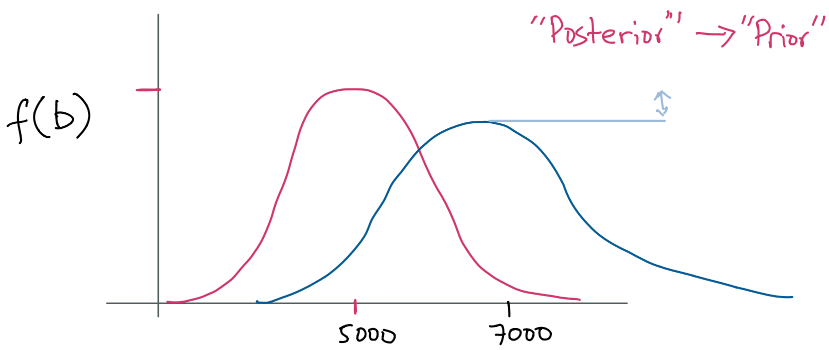


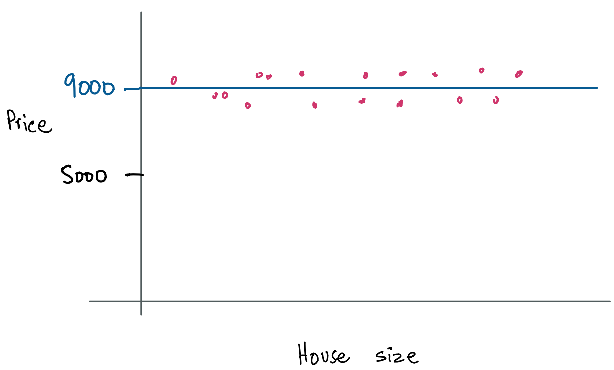
Our data will eventually look like this

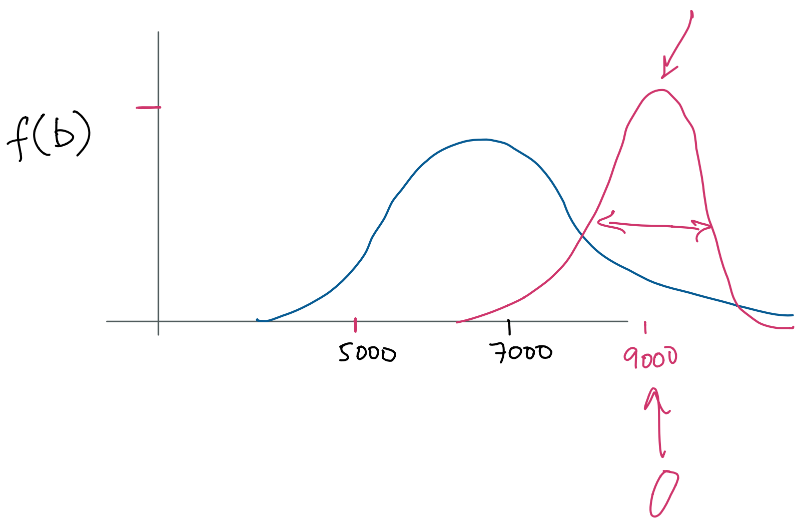


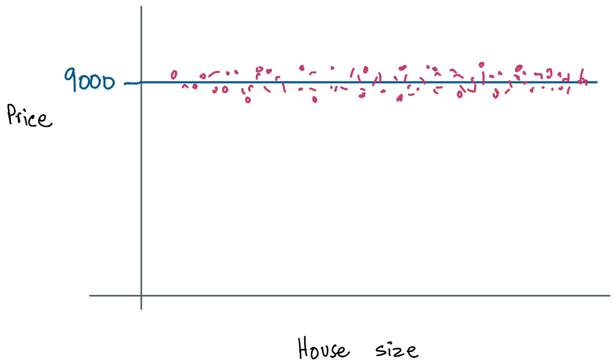


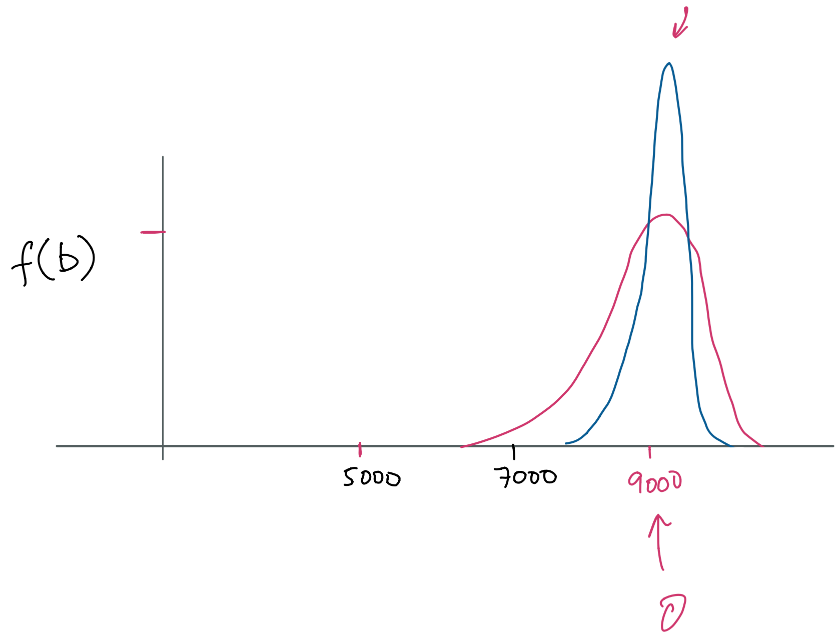






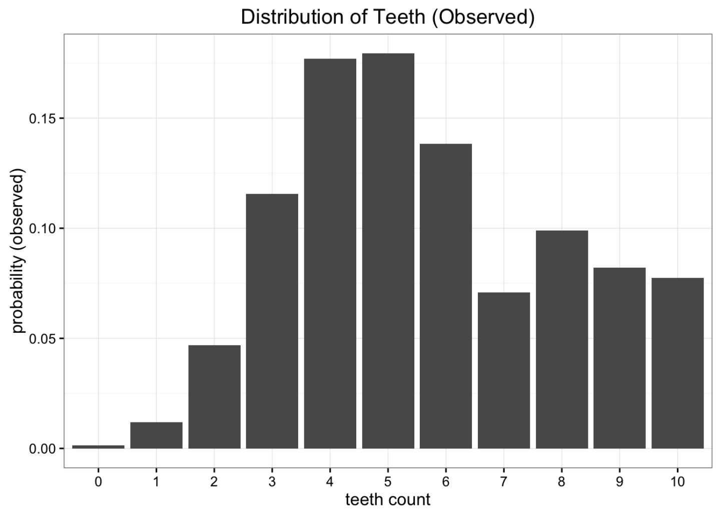


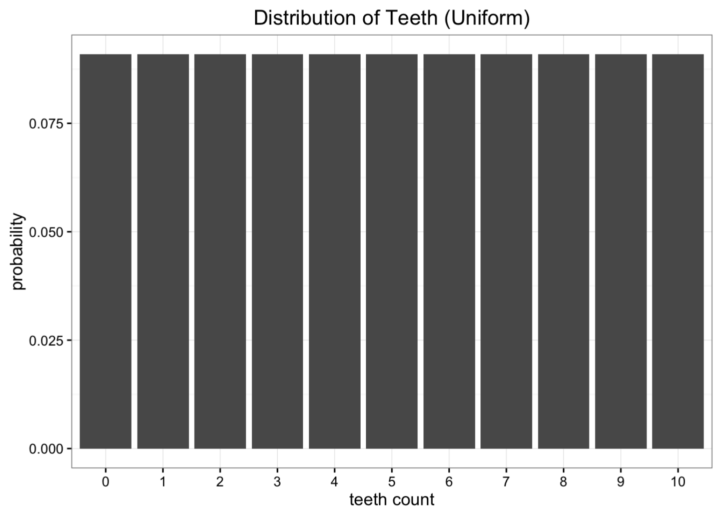


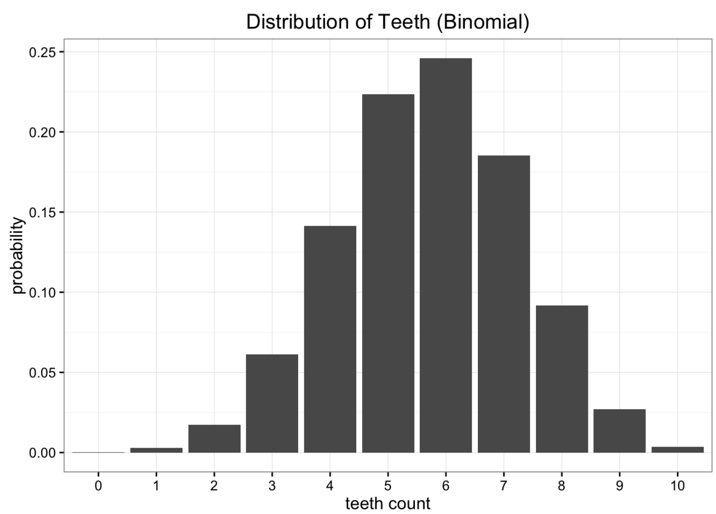


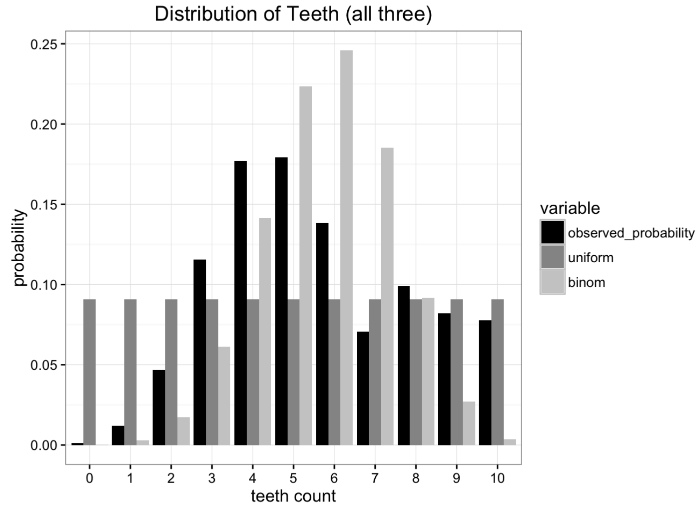
Notice that :

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


Distribution of Teeth (Observed)
probability (observed)
0.00
0.05
0.10
0.15
0
1
2
3
4
5
6
7
8
9
10
teeth count


Distribution of Teeth (Uniform)
probability
0.000
0.025
0.050
0.075
0
1
2
3
4
5
6
7
8
9
10
teeth count


Distribution of Teeth (Binomial)
probability
0.00
0.05
0.10
0.15
0.20
0.25
0
1
2
3
4
5
6
7
8
9
10
teeth count



$$H = -\sum_{i=1}^{N} p(x_i) \cdot \overbrace{\log p(x_i)}^{\text{neg. info}}$$

$-\lg p(x) \rightsquigarrow$ info in p

$-\lg q(x) \rightsquigarrow$ info in q

(Expected) information difference (loss)

$$\sum_{i=1}^{N} p(x_i) \left[-\lg q(x_i) - (-\lg p(x_i)) \right]$$

$$= \sum_{i} p(x_i) \left[\lg p(x_i) - \lg q(x_i) \right]$$

$$D_{KL}(p||q) = \sum_{i=1}^{N} p(x_i) \cdot log \frac{p(x_i)}{q(x_i)}$$

"Expected info difference".

$$D_{kl}(\text{Observed} \,||\, \text{Uniform}) = 0.338$$

$$D_{kl}(\text{Observed} \,||\, \text{Binomial}) = 0.477$$

(closer)