Distribution Theory & Mathematical Models

STAT 240 - Fall 2025

Robert Sholl

Distribution Theory

Last stop out of theory-land

Motivation

$S_T = [0, \infty)$

Describe the support of the following random variables:

r.v.	Description
T	Time it takes for a drug to enter the blood stream
U	Probability of landing on 3 on a fair 6-sided dice
V	Volume of the water in a swimming pool
W	Proportion of people who get sick at an Aggieville bar
X	The result of 50 coin flips
Y	Number of wild boar observed in a 10 acre plot

$S_u = \{1/6\}$

$S_v = [0, \infty)$

$S_w = [0, 1]$

$S_x = \{0, 1, 2, \dots, 50\}$

$S_y = \{0, 1, 2, \dots\}$

Normality

Given:

Assumptions behind Normality

$$X \sim N(\mu, \sigma^2)$$

- X is continuous
- The support of X is $S_X = (-\infty, \infty)$
- X has a centered mean as well as constant variance
  - Both are finite (a.k.a., they exist)

Design vs. Modeling

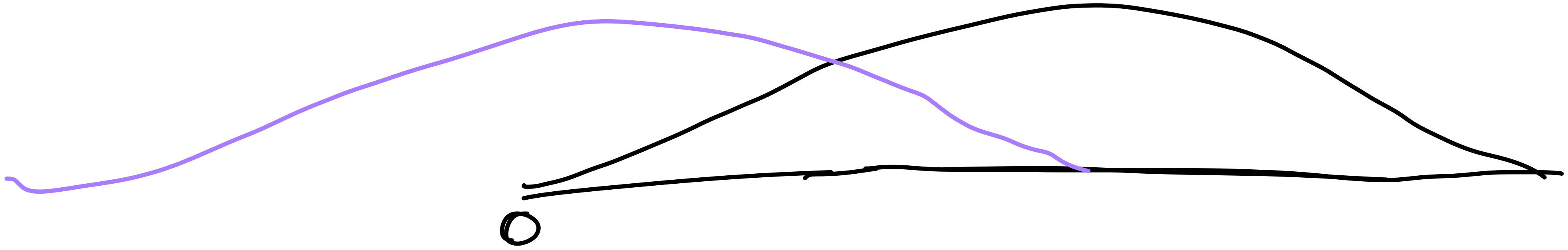
There are *two ways* to approach the application of statistics: Design or modeling

- We can use simplistic modeling if we have a very good study/experimental design
- If our design is flawed we can make up for this with better modeling
- Both can be improved with increased sample sizes

Modeling

For modeling we have *two options* for getting stronger inference:

- Increase sample size
- Load on assumptions
  - Is the assumption of normality asking much?
  - If μ is large and σ is small we don't care about the lower bounds



Tranformation of the response

Continuous

$$Y = \{\text{The number of wild boar observed in a 10 acre plot}\}$$
$$S_Y = \{0, 1, 2, \ldots\}$$

Let $\ell = \ln(Y)$

$$S_\ell = [0, \infty)$$

CONTINUITY IS SATISFIED

Negative support

$$T = \{\text{The time it takes for a drug to enter the blood stream}\}$$

Let $\tilde{T} = T + 100$

- Now no matter how small T is, $\tilde{T}$ won't cross 0
- If $\tilde{T}$ looks bell shaped we can use the normal distribution!

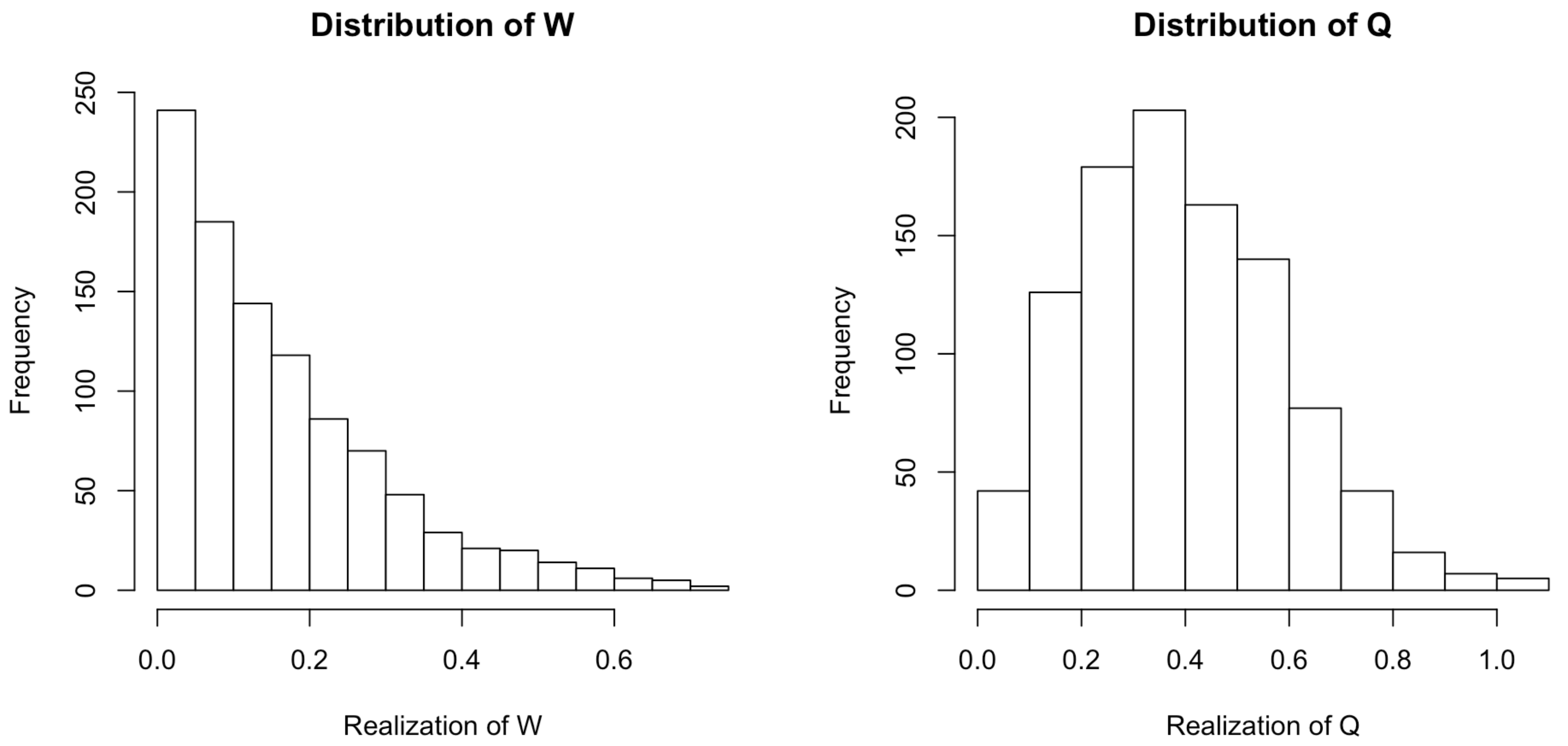
Symmetry

W = {The proportion of people who get sick at an Aggieville bar}

Say that W was right-skewed:

- Let $Q = \arcsin \sqrt{W}$
- the arcsine is asinine

Symmetry


Distribution of W
Frequency
0
50
100
150
200
250
0.0
0.2
0.4
0.6
Realization of W
Distribution of Q
Frequency
0
50
100
150
200
0.0
0.2
0.4
0.6
0.8
1.0
Realization of Q

Other Named Distributions

Assumptions

- Hand-wavey transformations are useful
  - But they can be uncomfortable
- Let's look at what we dealt with last class

$$\mu_{\bar{x}} = \mu \quad , \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Bias vs. Variance

$\mu_{\bar{x}} = \mu \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

Let $\mu = 10$ and $\sigma = 5$. Show the distribution of $\bar{x}$ at $n = 1$, $n = 10$, $n = 100$, and $n = 1000$.

LET $\bar{X}_1$ BE $\bar{X}$ AT $n=1$, $\bar{X}_3$ $n=100$

$\bar{X}_2$ $n=10$, $\bar{X}_4$ $n=1000$

STATISTICAL BIAS:
THE TENDENCY TOWARDS AN INCORRECT RESULT.

$\bar{X}_1 \sim N(10, (5/\sqrt{1})^2) \Rightarrow \bar{X}_1 \sim N(5, 25)$ $\quad 5/3.16 = 1.58^2 = 2.5$

$\bar{X}_2 \sim N(10, (5/\sqrt{10})^2) \Rightarrow \bar{X}_2 \sim N(5, 2.5)$

$\bar{X}_3 \sim N(10, (5/\sqrt{100})^2) \Rightarrow \bar{X}_3 \sim N(5, 0.25)$

$\bar{X}_4 \sim N(10, (5/\sqrt{1000})^2) \Rightarrow \bar{X}_4 \sim N(5, 0.025)$

$\mu = \sum_i x_i / N$

Assumptions = Samples?

As we increase assumptions we see the same result:

- More assumptions $\rightarrow$ more bias
- Less assumptions $\rightarrow$ less variance

> What's a possible explanation for this? Does the opposite relationship make sense?

Distribution Theory

- In distribution theory we seek to:
  - Formally describe distributions of random variables
  - Determine their moments and other features
  - Put them into practice
  - Locate any useful transformations or cases of the distributions

Discrete

Bernoulli

Let X be the result of a **fair** coin toss.

$$X = \begin{cases} 1 & \text{if heads} \\ 0 & \text{if tails} \end{cases}$$

$$P(X = x) = \begin{cases} 0.5 & \text{if } \; x = 1 \\ 0.5 & \text{if } \; x = 0 \end{cases}$$

Bernoulli

Let X be the result of an **unfair** coin toss. BINARY OUTCOME

- We don't know the probability of heads, let's label it p

$$P(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

$$X \sim \text{Bern}(p)$$

IS DISTRIBUTED

$p \equiv$ PROBABILITY OF $X = 1$

Binomial

Let X be the result of 2 fair coin tosses.

- There are multiple combinations of $X = x$:

Flip 1	Flip 2	
1	1	2
1	0	1
0	1	1
0	0	0

Binomial

Thinking a little differently:

- We have 3 possible outcomes (if order doesn't matter)

	0	1	2
0	1	0	0
1	1	1	0
2	1	2	1

PASCAL'S TRIANGLE

Binomial

We can represent this with the "choose" function:

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

$3! = 3 * 2 * 1$

where $n! = n \times (n-1) \times (n-2) \times \ldots \times 3 \times 2 \times 1$

If the flips are *independent* we can express the probability as:

$$\text{Success} = p^x$$
$$\text{Failure} = (1-p)^{n-x}$$

Binomial

Smashing those together:

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

$$X \sim \text{Binom}(n, p)$$

n = # of TRIALS

p = P(Success)

Binomial

$\text{Bern}(p) = \text{Binom}(1, p)$

$$X \sim \text{Binom}(n, p)$$

$$\mathbb{E}X = np$$] *

$$\mathbb{V}X = np(1-p)$$ *

- What about Bernoulli? $\mathbb{E}X = p$

  $\mathbb{V}X = p(1-p)$

Poisson

Poisson: the French word for "fish"

- Think about counting fish in a pond
  - Discrete and strictly positive

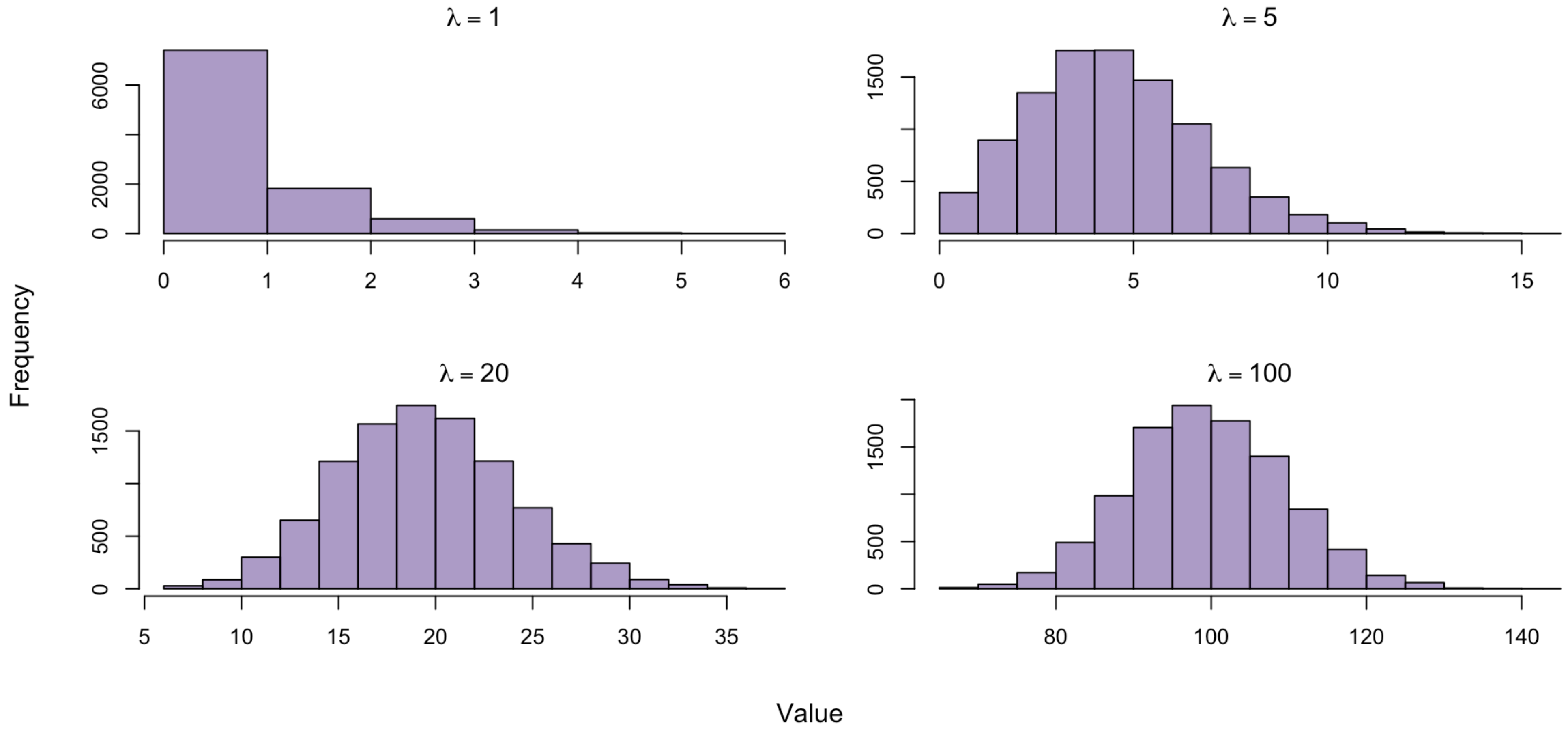
DISCRETE COUNTS

$$X \sim \text{Pois}(\lambda)$$

Pois(5)

$$\mathbb{E}X = \mathbb{V}X = \lambda$$

Poisson


λ = 1
λ = 5
λ = 20
λ = 100
Frequency
Value
0
2000
6000
0
1
2
3
4
5
6
0
500
1500
0
5
10
15
0
500
1500
5
10
15
20
25
30
35
0
500
1500
80
100
120
140

Continuous

Uniform

- Before we defined this as a shape
- Imagine now that you can control the *interval* of that uniform bar

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

$$X \sim \text{Unif}(a, b)$$

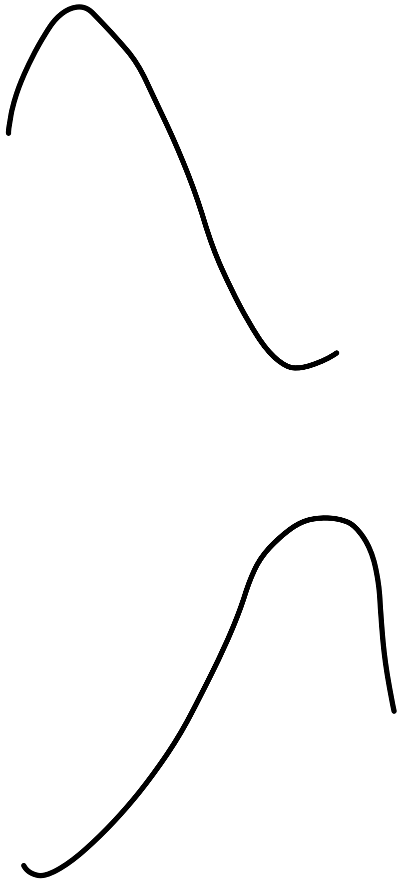
a = MIN

b = MAX

Beta

PROPORTIONS & PROBABILITIES

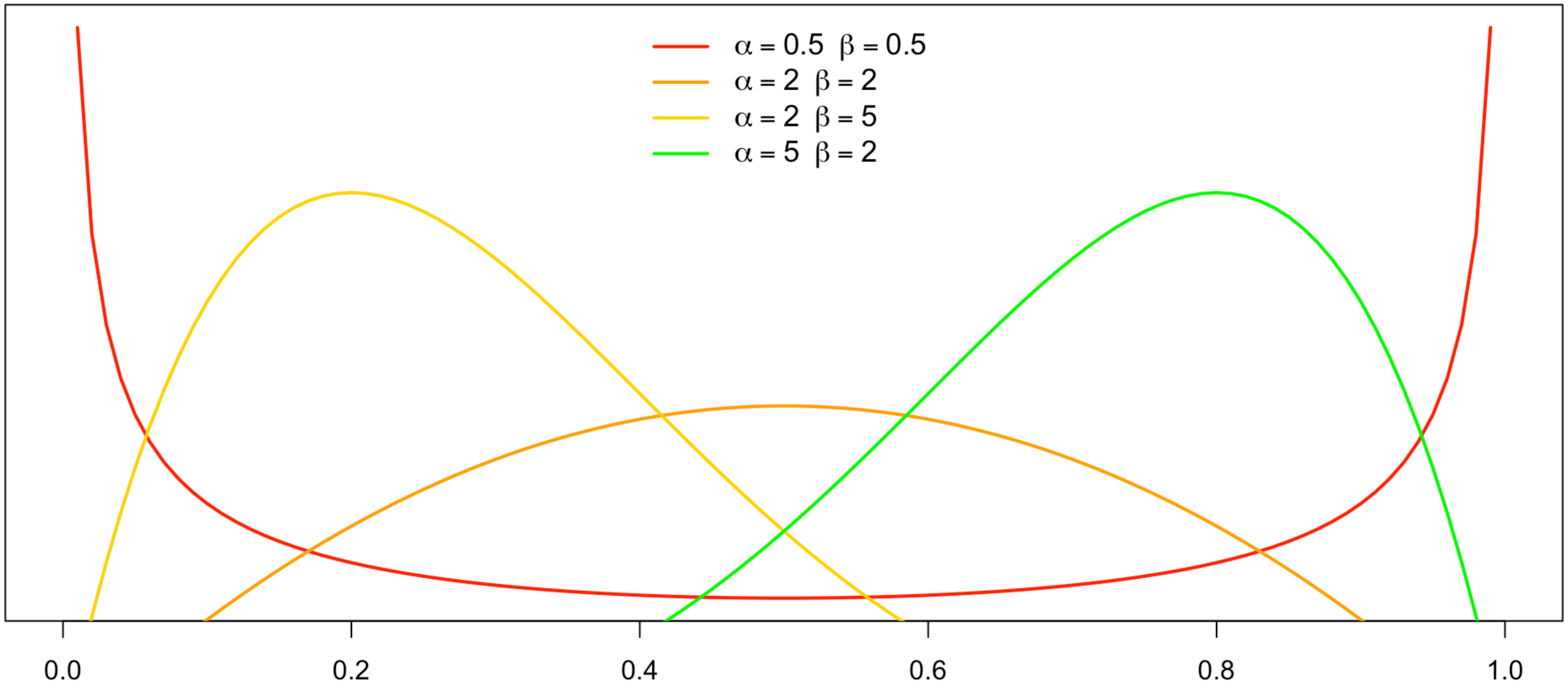
- For data that takes on values between 0 and 1
  - And those values *are not* constant



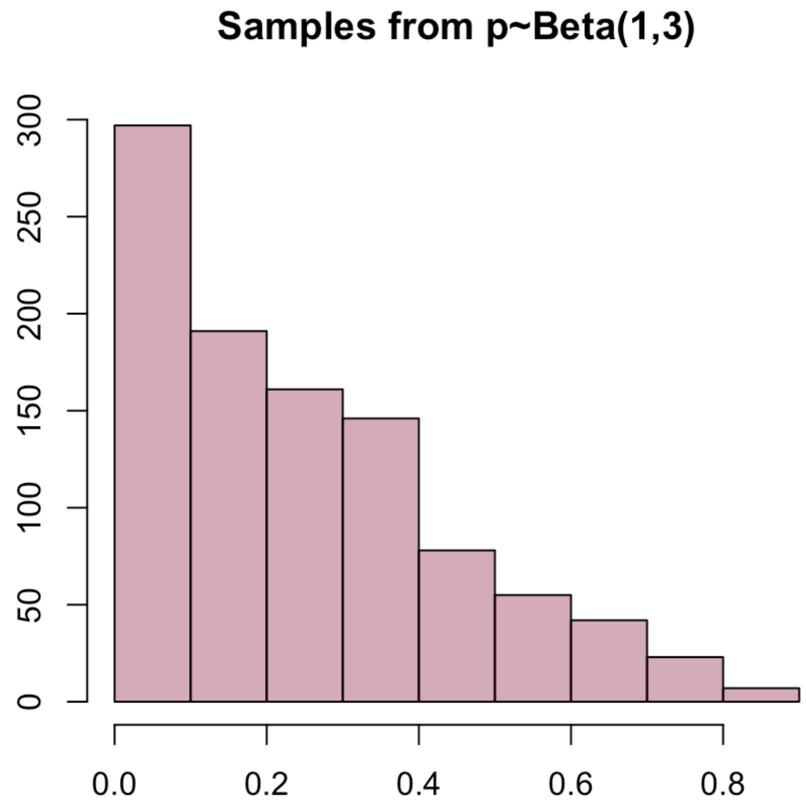
$$X \sim \text{Beta}(\alpha, \beta)$$

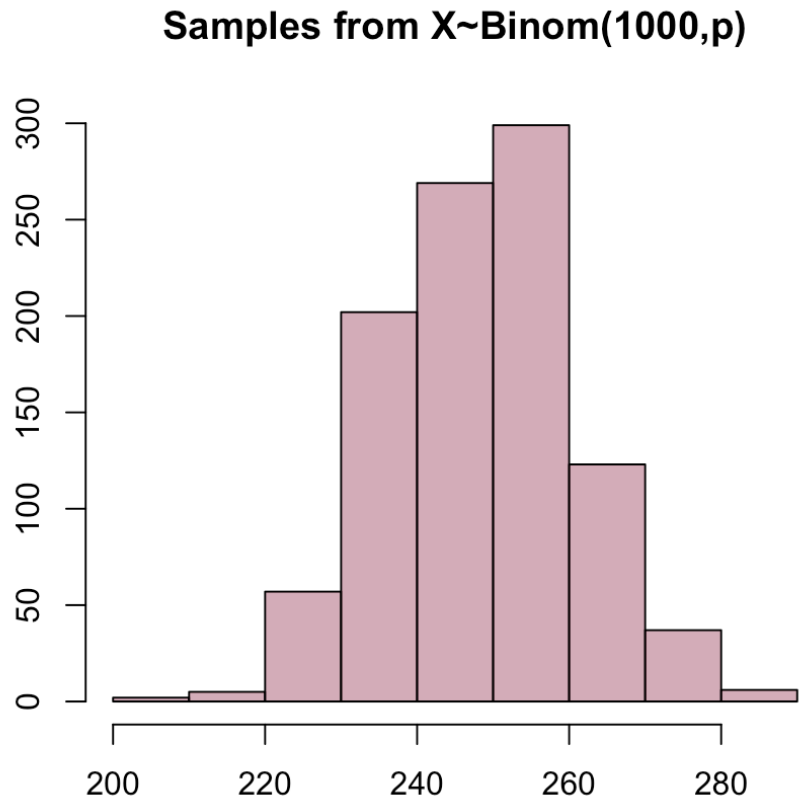
$$\mathbb{E}X = \frac{\alpha}{\alpha + \beta} \qquad \mathbb{V}X = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Beta



Mixing Distributions





Time

Is time **discrete** or **continuous**?

- Imagine how you would describe both!
- Now imagine the constraints for both
  - Can time be negative?
  - Can time stretch to infinities?

Gamma

For **continuous** and **strictly positive** data:

CONT

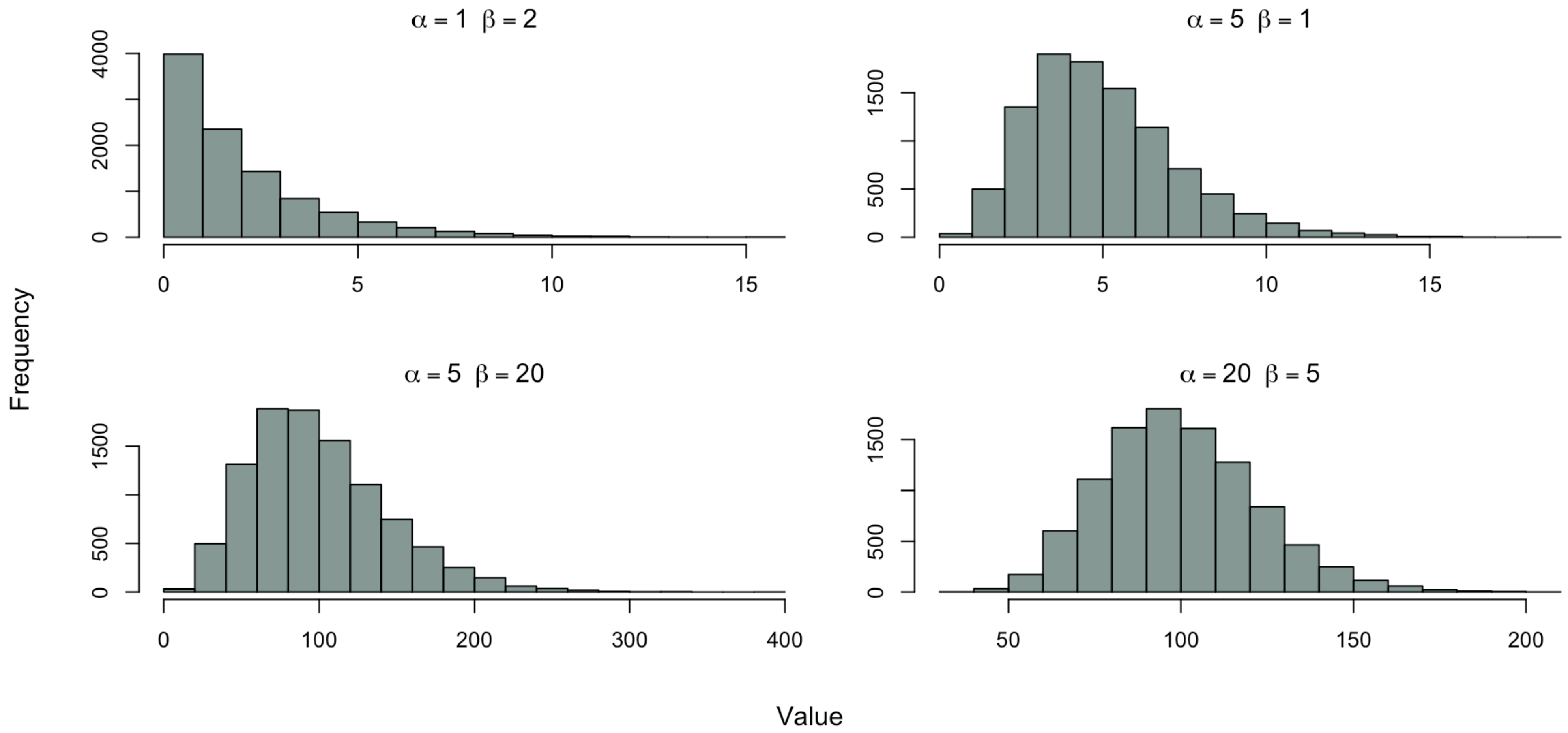
$0 \rightarrow \infty$

$$X \sim \text{Gamma}(\alpha, \beta)$$

- The expectation and variance of this differ
  - Depends how we parameterize it
  - As such, we won't discuss those

0, 1, 2

Gamma



α = 1 β = 2
4000
2000
0
0
5
10
15
α = 5 β = 1
1500
500
0
0
5
10
15
Frequency
α = 5 β = 20
1500
500
0
0
100
200
300
400
α = 20 β = 5
1500
500
0
50
100
150
200
Value

Mathematical Models

Time to draw lines

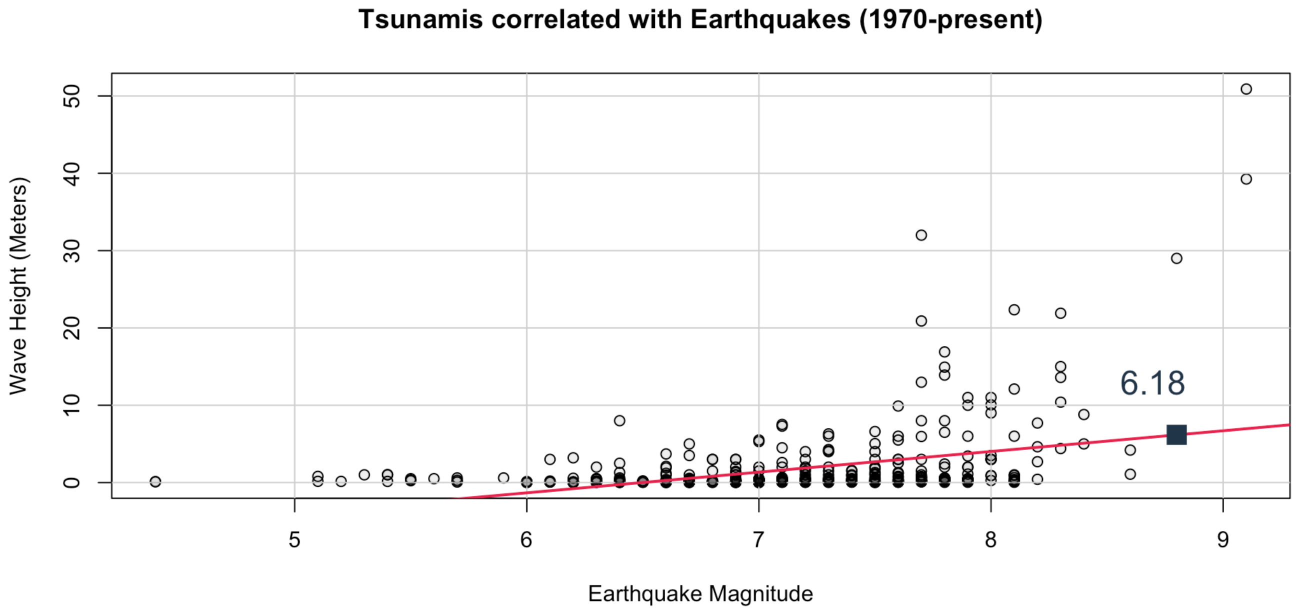
Models

- Any mathematical representation of a process
- Think about model planes:
  - Can we make a very simple one?
  - What about a very complex one?
  - Can they both fly?
  - Is either as good as a real plane?

Phenomenological Models

Building a model, blind of a process, to better understand a process.

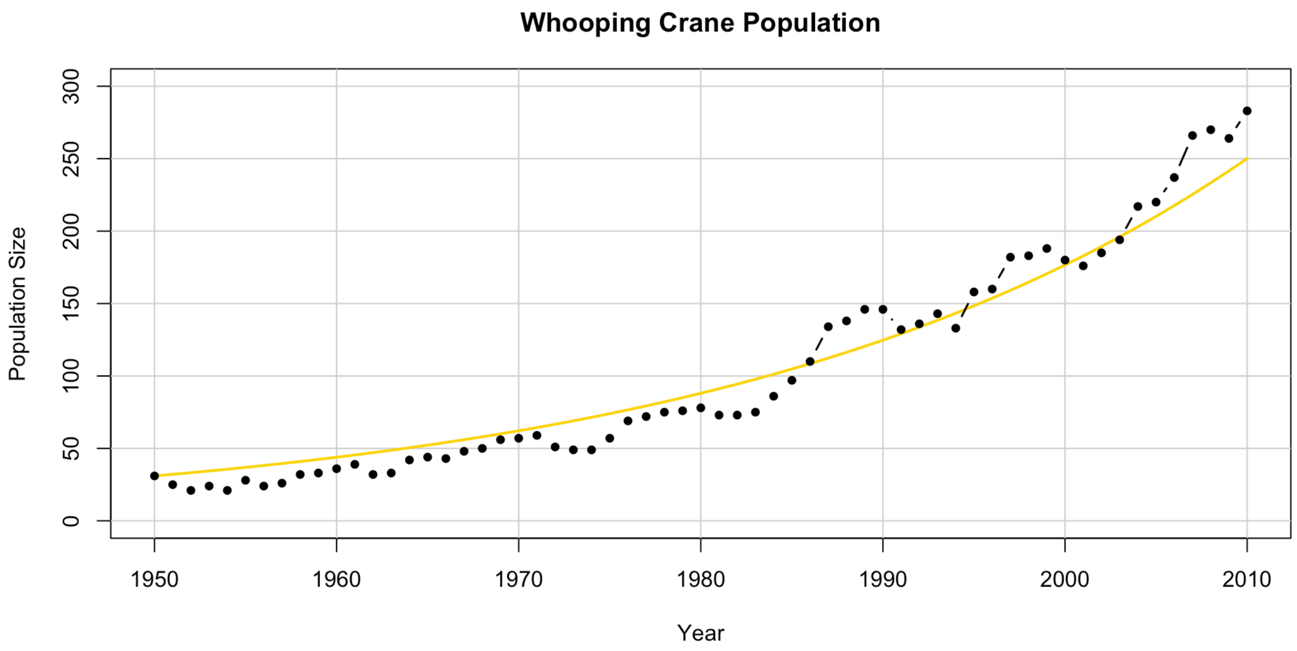
$$y = mx + b$$



Mechanistic Models

Building a model using knowledge of a process, to better understand the data.

$$\lambda(t) = \lambda_0 e^{\gamma(t-t_0)}$$



Linear

Linear *in the parameters*:

$$Y = aX + b$$

$$w = \beta_0 + \beta_1 x + \beta_2 x^2$$

$$P = \mu + \alpha \log(r \times t)$$

$$\boldsymbol{\Delta} = \boldsymbol{\beta} \cos(\boldsymbol{X}) + 10$$

Nonlinear

We did something *nonlinear* to those parameters:

$$Y = a^2X + b$$

$$w = \beta_0 + \log(\beta_1)x$$

$$\lambda(t) = \lambda_0 e^{\gamma(t-t_0)}$$

Deterministic

The same inputs **always** result in the same outputs:

$$y = 2x$$

- Let $x = 4$
- Let $x = 4$ again
- Did you get the same result?

Probabilistic

There's some **random element** to the model:

$$y = mx + b + \text{Random Error}$$

- The same input won't always get the same output
  - This is where we can slap distributions into the mix

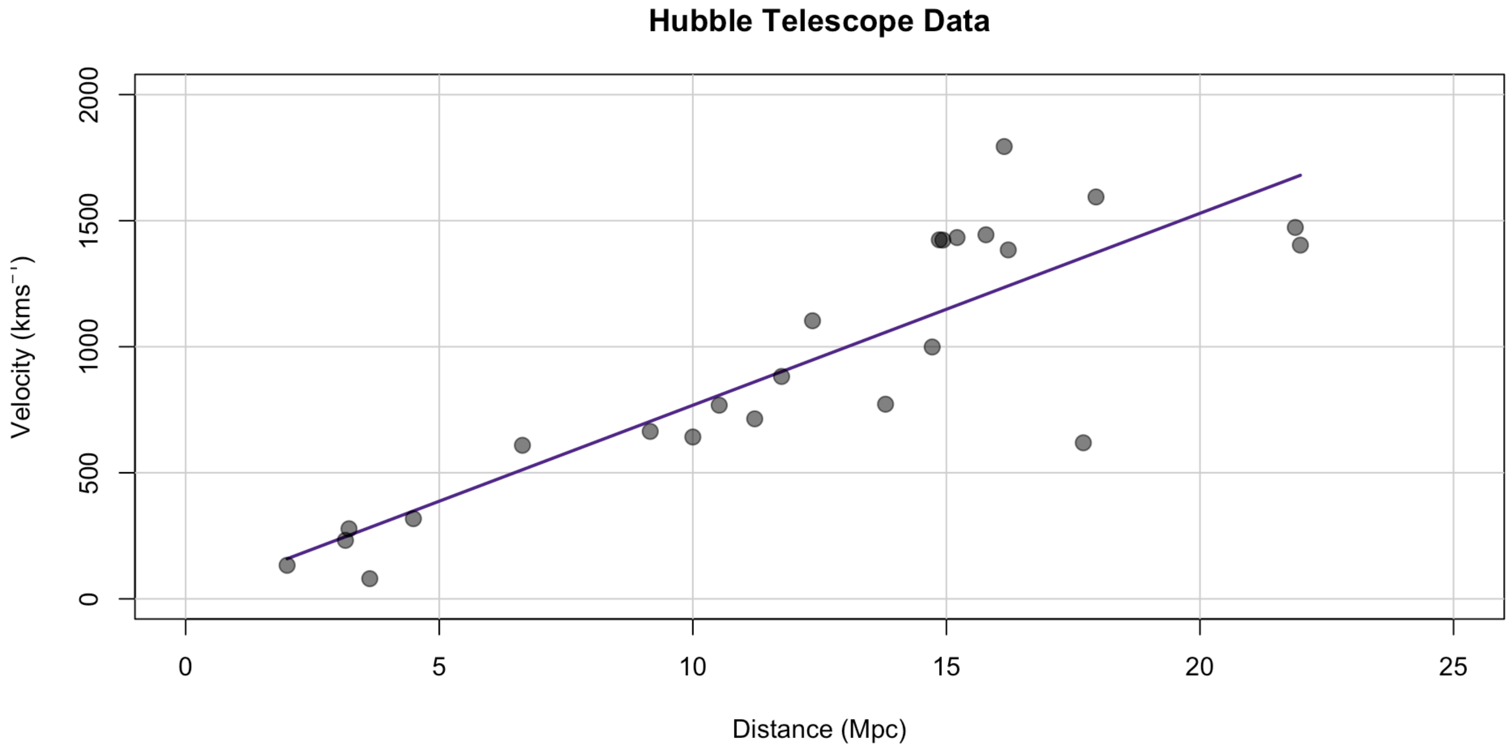
“Solving” models

- **Explicit models:** Input values for the parameters and solve for the output (a.k.a. response)
- **Implicit models:** Solve for the parameters to understand the structure of the model
- We do both in statistics (usually back to back)

$$P(t) = P_0 e^{rt}$$

$$y = mx + b$$

Derived Quantities



Final Note

There's two more "distinctions" of models, specific to statistics:

- Frequentist
  - This is *every* method we learn in this class
- Bayesian
  - This exists, we'll talk about it a little, not too much
  - STAT 341, STAT 610/611, STAT 768

Least Squares

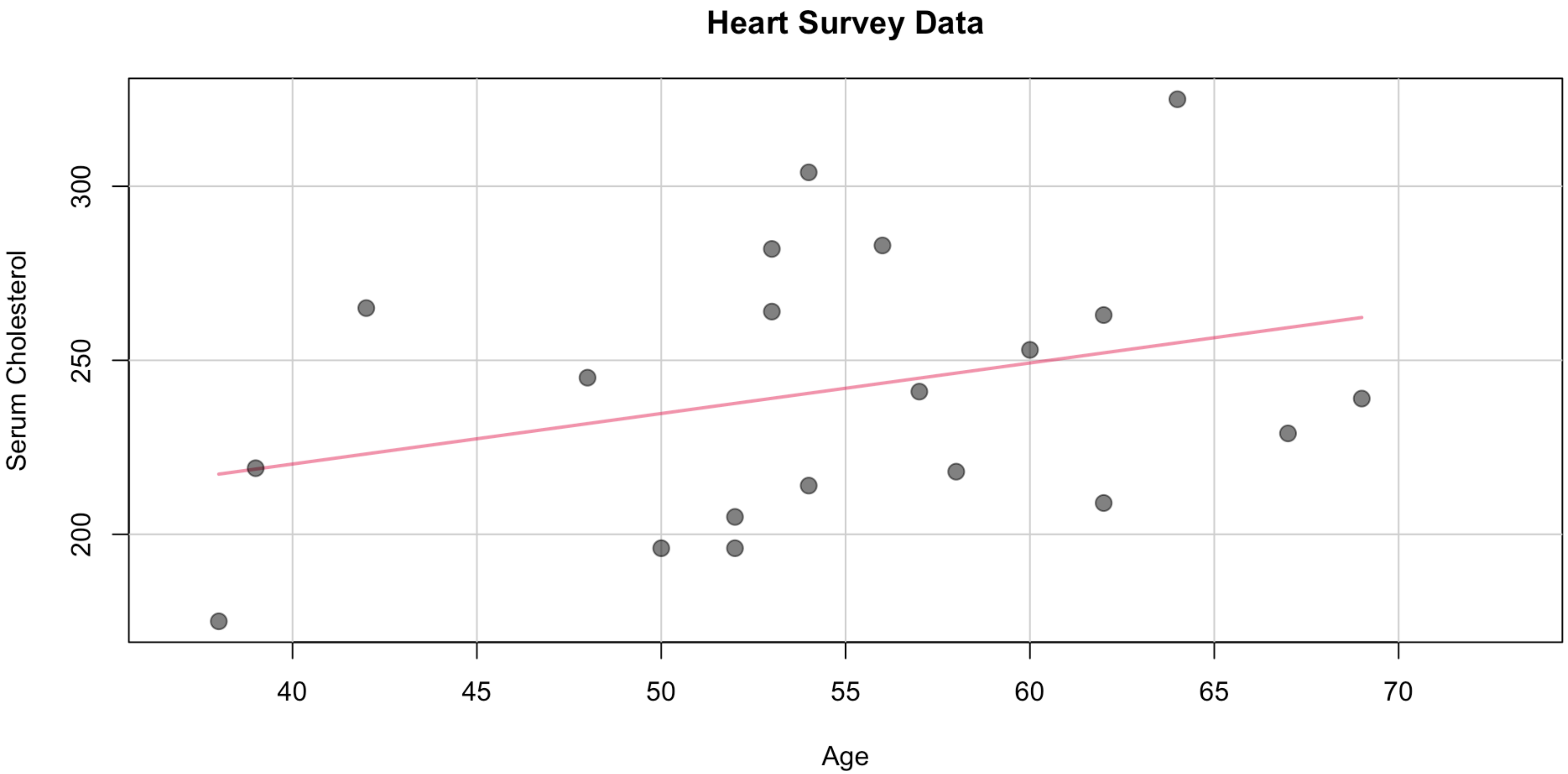
STAT 240 - Fall 2025

Robert Sholl

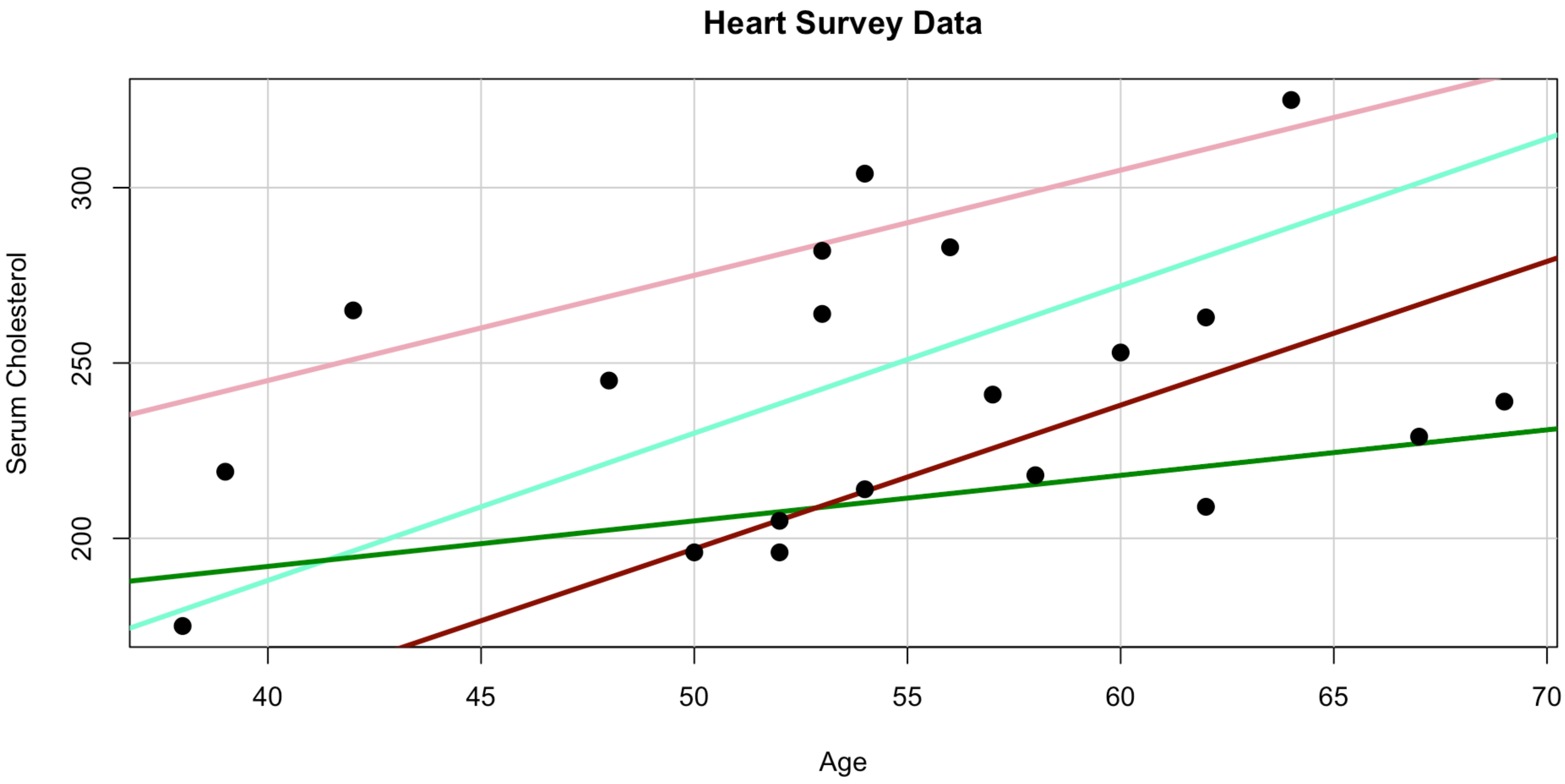
Motivation

- **Phenomena**: If I give three groups of patients three different blood pressure medications in an experimental trial, what effect do those drugs have? Are they different?
- **Mechanisms**: We know that speed estimates of stars increase linearly with their observed distance from Earth, how can we use this to predict where the edges of the Universe are?
- **Machine learning**: How can we train a machine to recognize whether a recipe is for a dinner or a dessert?

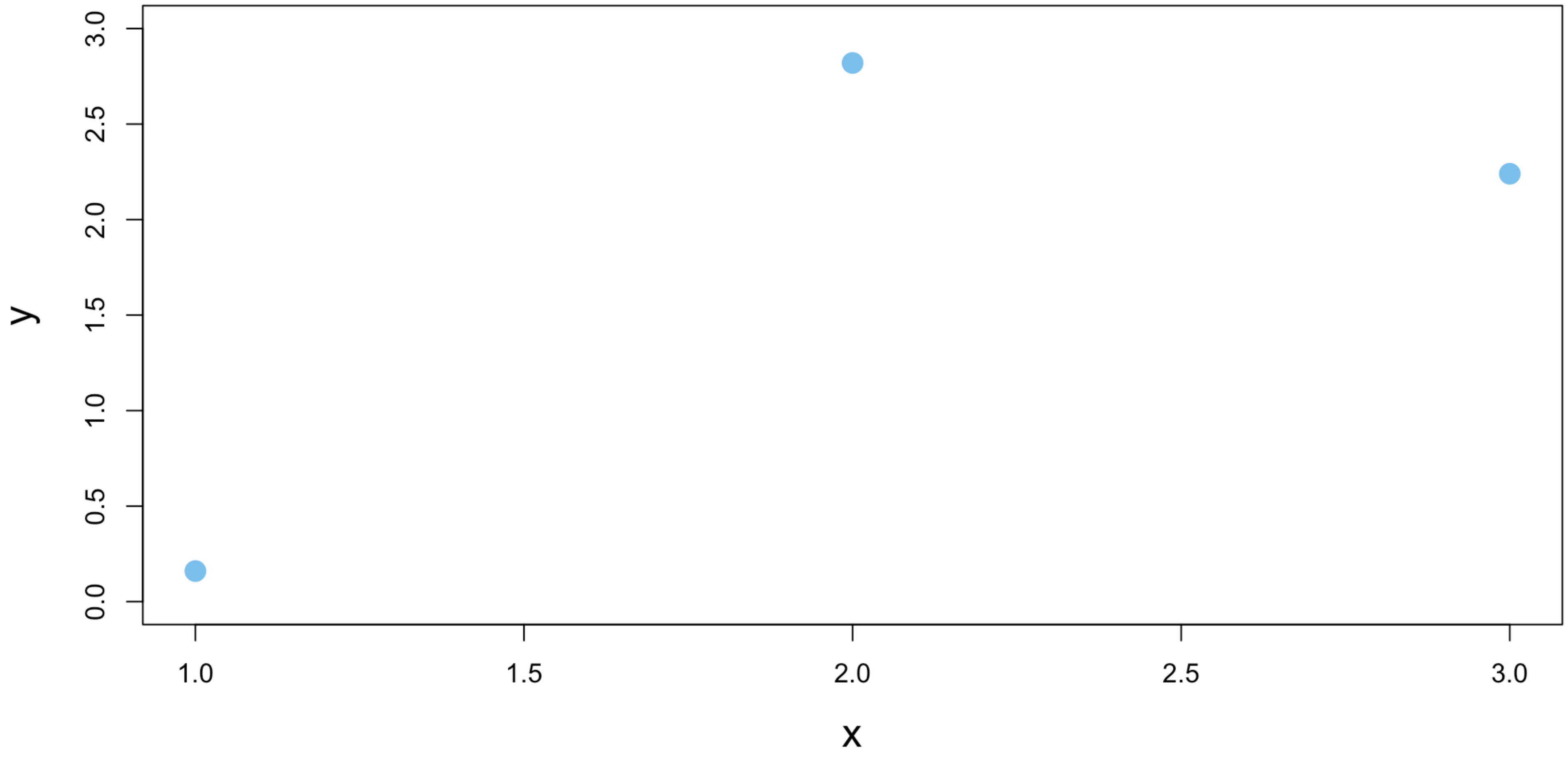
Line of Best Fit



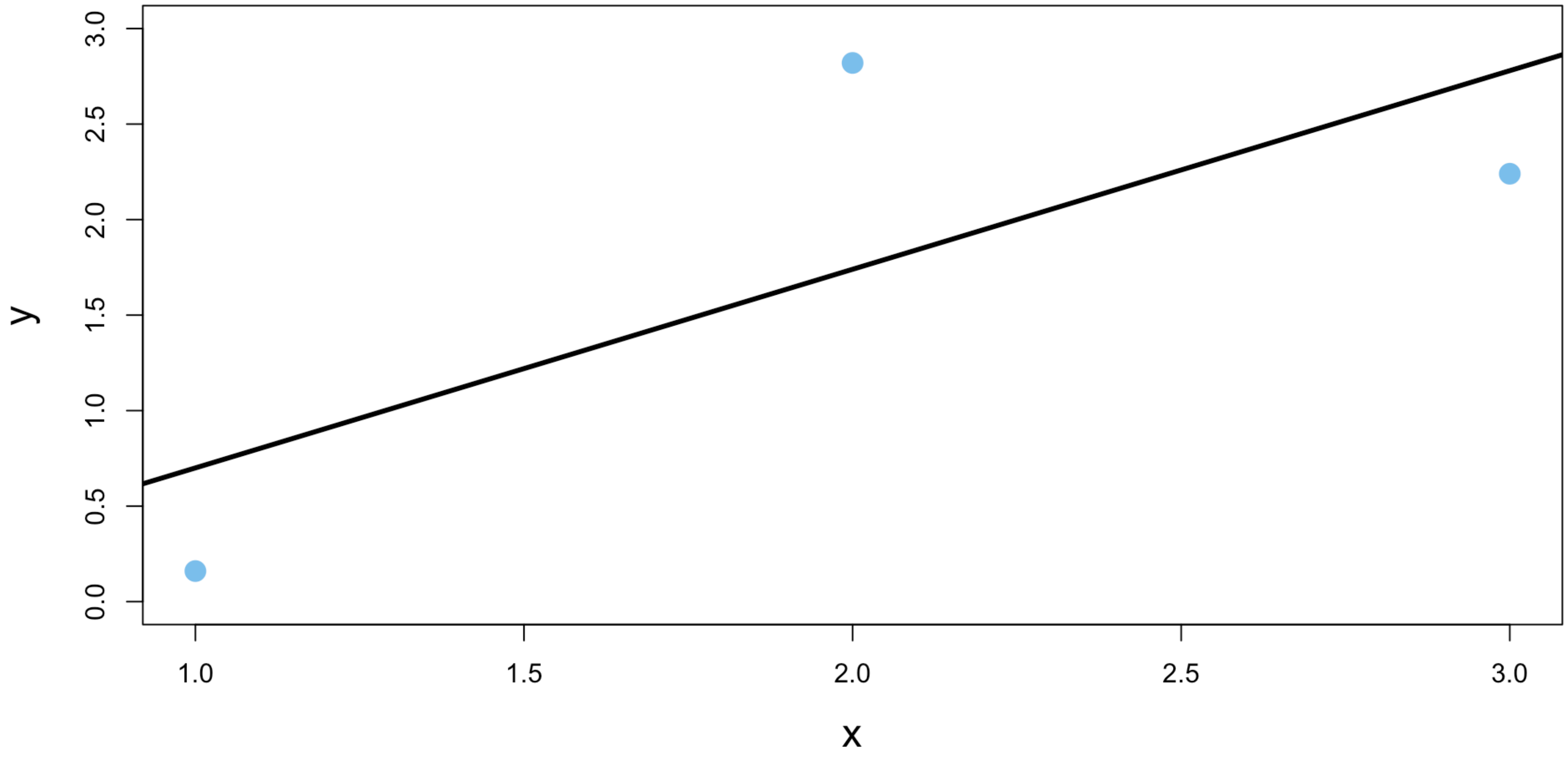
Which is “best”?



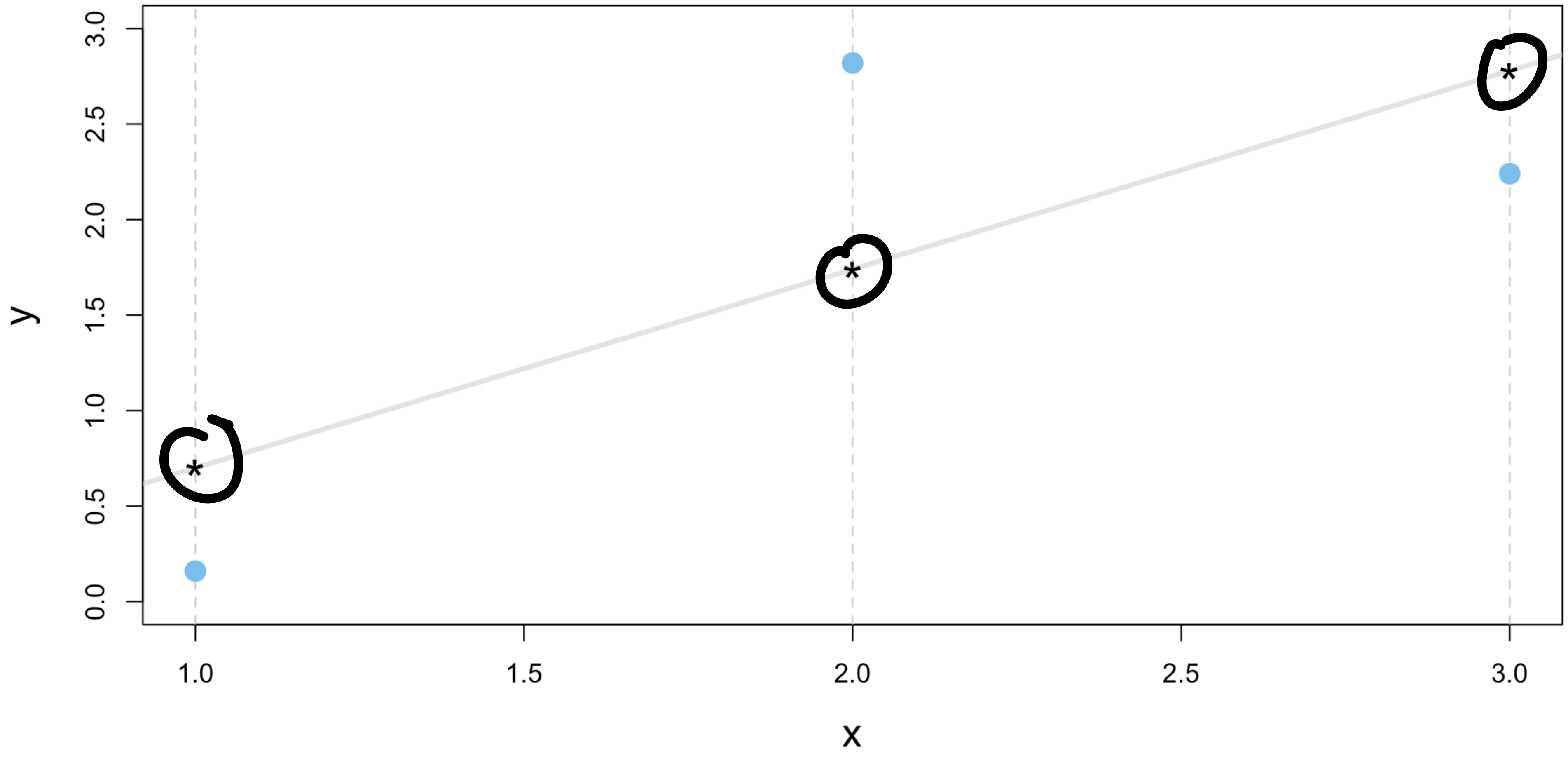
Least Squares



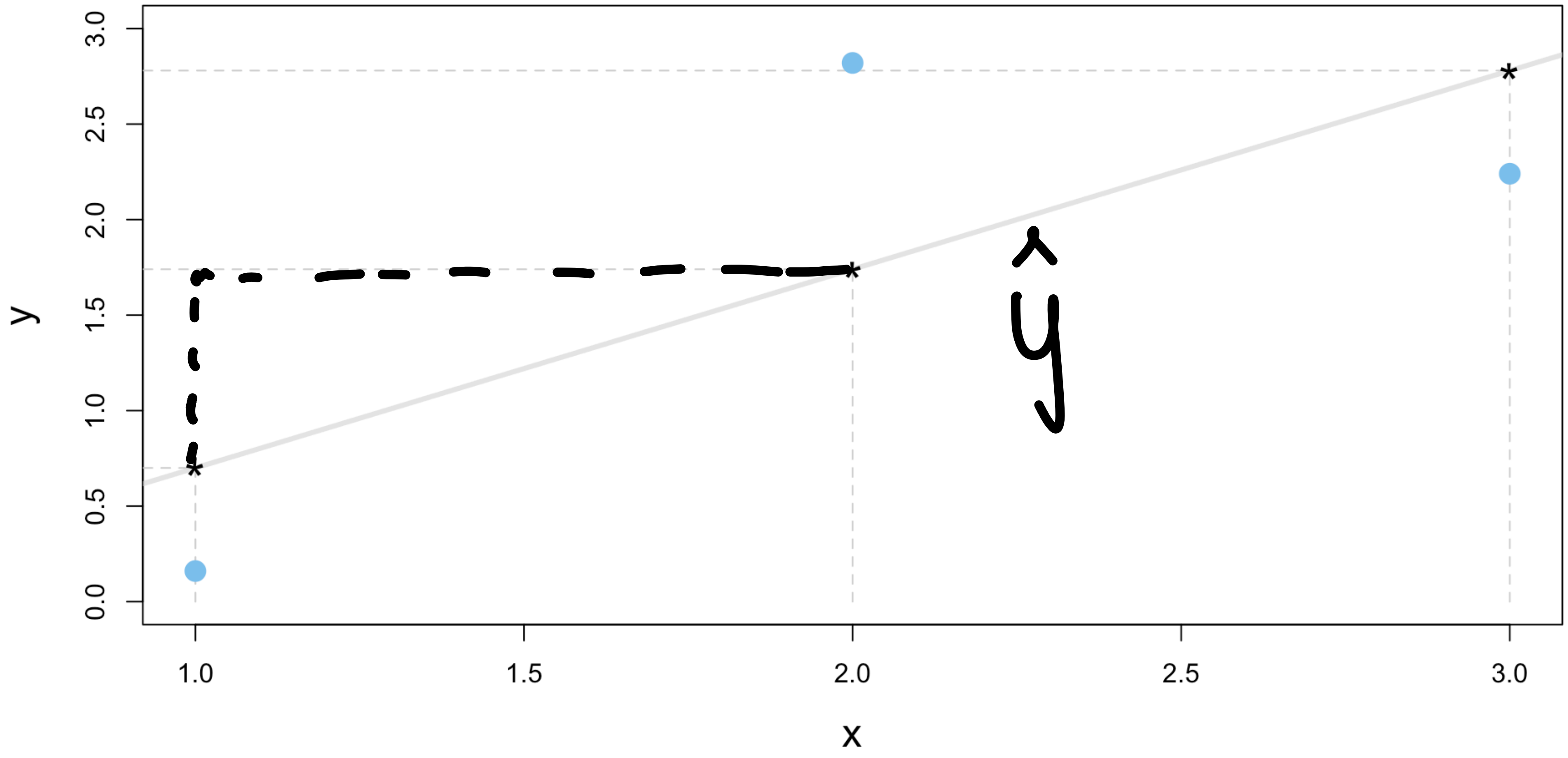
Least Squares



Least Squares



Least Squares



Least Squares

x	y	$\hat{y}$
1	0.16	0.70
2	2.82	1.74
3	2.24	2.78

$$\hat{y} = mx + b$$

Least Squares

$$m = \frac{1.74 - 0.70}{1} = 1.04$$

$$0.70 = 1.04(1) + b$$

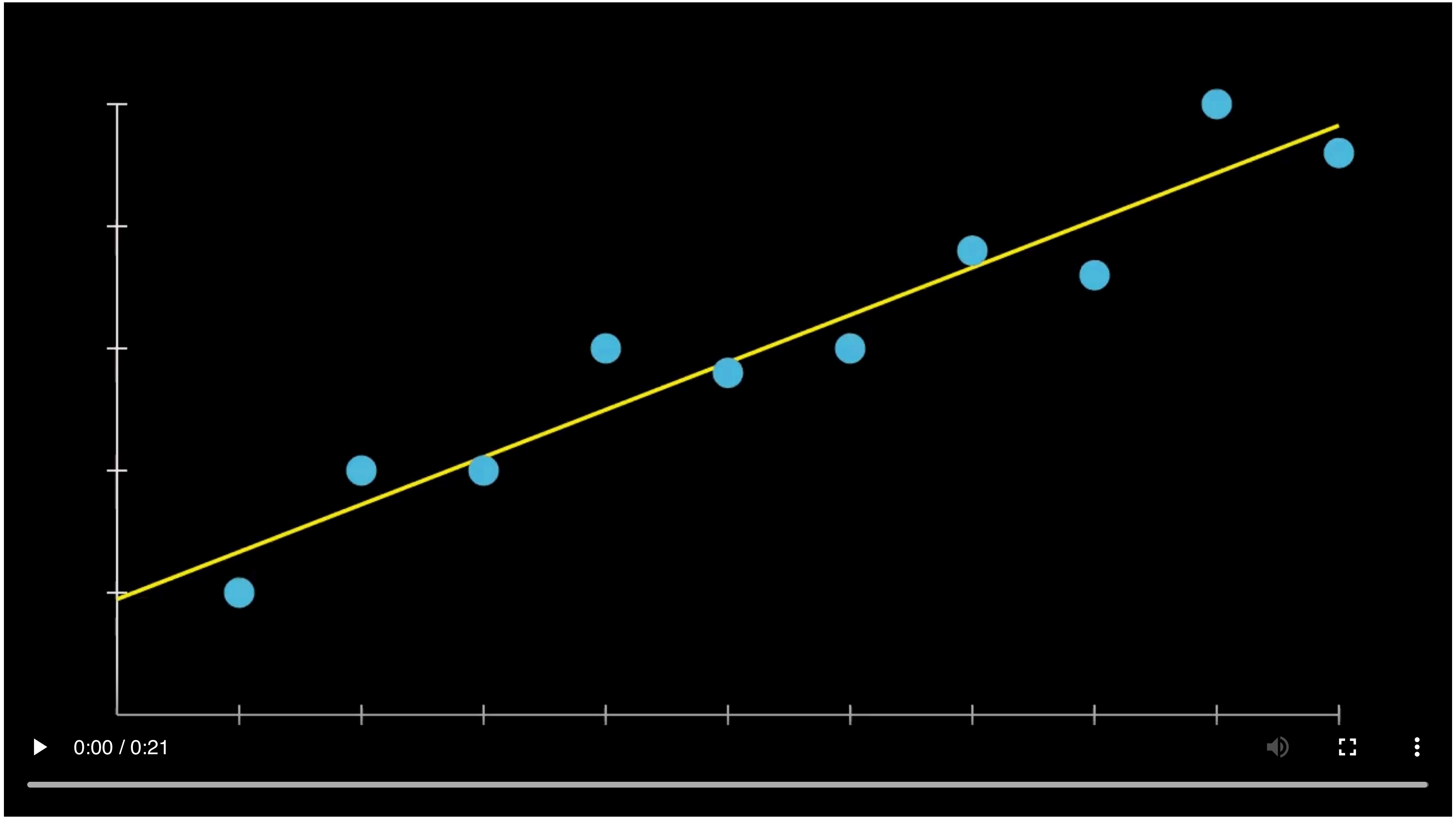
$$\hat{y} = 1.04x - 0.34$$

Least Squares

What if this is all we have?

x	y
1	0.16
2	2.82
3	2.24

Least Squares



Least Squares

$$m = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^n (x_i - \bar{x})^2}$$

$$b = \bar{y} - m\bar{x}$$

$$\hat{y} = mx + b$$

Least Squares

$$y = b + mx$$

$$y = y$$
$$m = \beta_1$$
$$x = x$$
$$b = \beta_0$$

$$y = \beta_0 + \beta_1 x + \epsilon$$

$$\hat{\epsilon}_i = y_i - \hat{y}_i$$

Least Squares

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

$$\hat{\beta}_1 = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^n (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\epsilon}_i = y_i - \hat{y}_i$$

Interpretations

Variable	Name	Interpretation
y	Response	Variable being predicted
x	Predictor	Variable predicting the response
β_0	Intercept	Baseline level of response
β_1	Slope	Effect of predictor on response
ϵ	Residuals	Difference between prediction and data

The Method of Least Squares

$\bar{x} = \sum_i x_i / n$

$\bar{y} = \sum_i y_i / n$

$\hat{\beta}_1 = \dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

PROD(a, b)	x	y	a: $(x_i - \bar{x})$	b: $(y_i - \bar{y})$	$(x_i - \bar{x})^2$
1.58	1	0.16	-1	-1.58	1
0	2	2.82	0	1.08	0
0.5	3	2.24	1	0.5	1
Sum 2.08	6	5.22	X	X	2
MEAN	2	1.74			

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

$\hat{y}_i = -0.34 + 1.04 x_i$

$\hat{\beta}_1 = \dfrac{2.08}{2} = 1.04$

$\hat{\beta}_0 = 1.74 - (1.04)2$

$1.74 - 2.08 = -0.34$

$$\hat{y}_i = -0.34 + 1.04x_i$$

$$\hat{y}_1 = -0.34 + 1.04(1) = 0.7 \qquad y_1 = 0.16 \qquad E_1 = y_1 - \hat{y}_1$$

$$\hat{y}_2 = -0.34 + 1.04(2) = 1.74 \qquad y_2 = 2.82 \qquad E_2 = y_2 - \hat{y}_2$$

$$\hat{y}_3 = -0.34 + 1.04(3) = 2.78 \qquad y_3 = 2.24 \qquad E_3 = y_3 - \hat{y}_3$$

Assumptions

Besides linearity, that's an *assumed* assumption

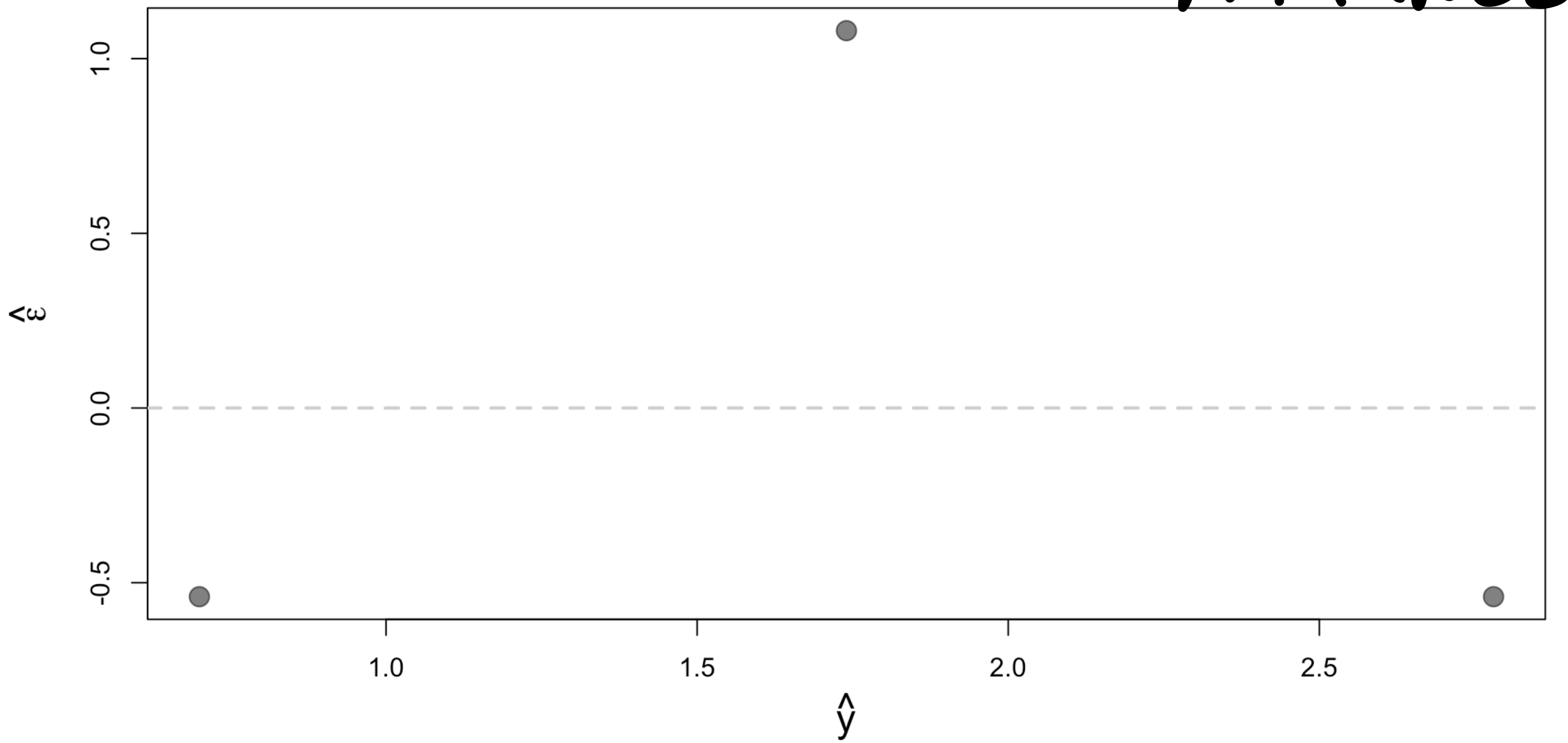
Independence ✓

x	y	$\hat{y}$	$y - \hat{y}$
1	0.16	0.7	–0.54
2	2.82	1.74	1.08
3	2.24	2.78	–0.54

$$\bar{\epsilon} = \frac{\hat{\epsilon}_1 + \hat{\epsilon}_2 + \hat{\epsilon}_3}{3} = \frac{-0.54 + 1.08 - 0.54}{3} = \frac{0}{3} = 0$$

Homoscedasticity

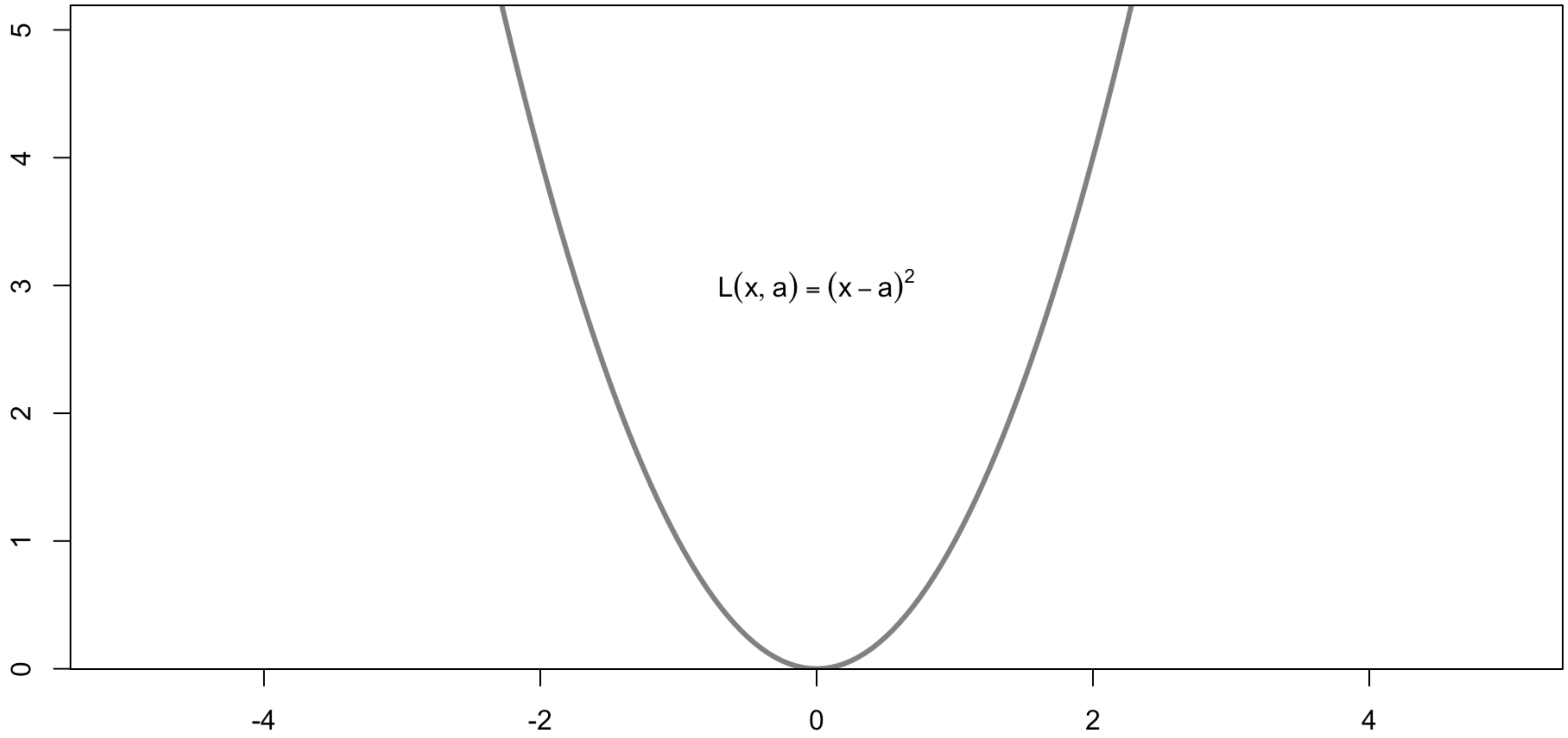
CONSTANT VARIANCE



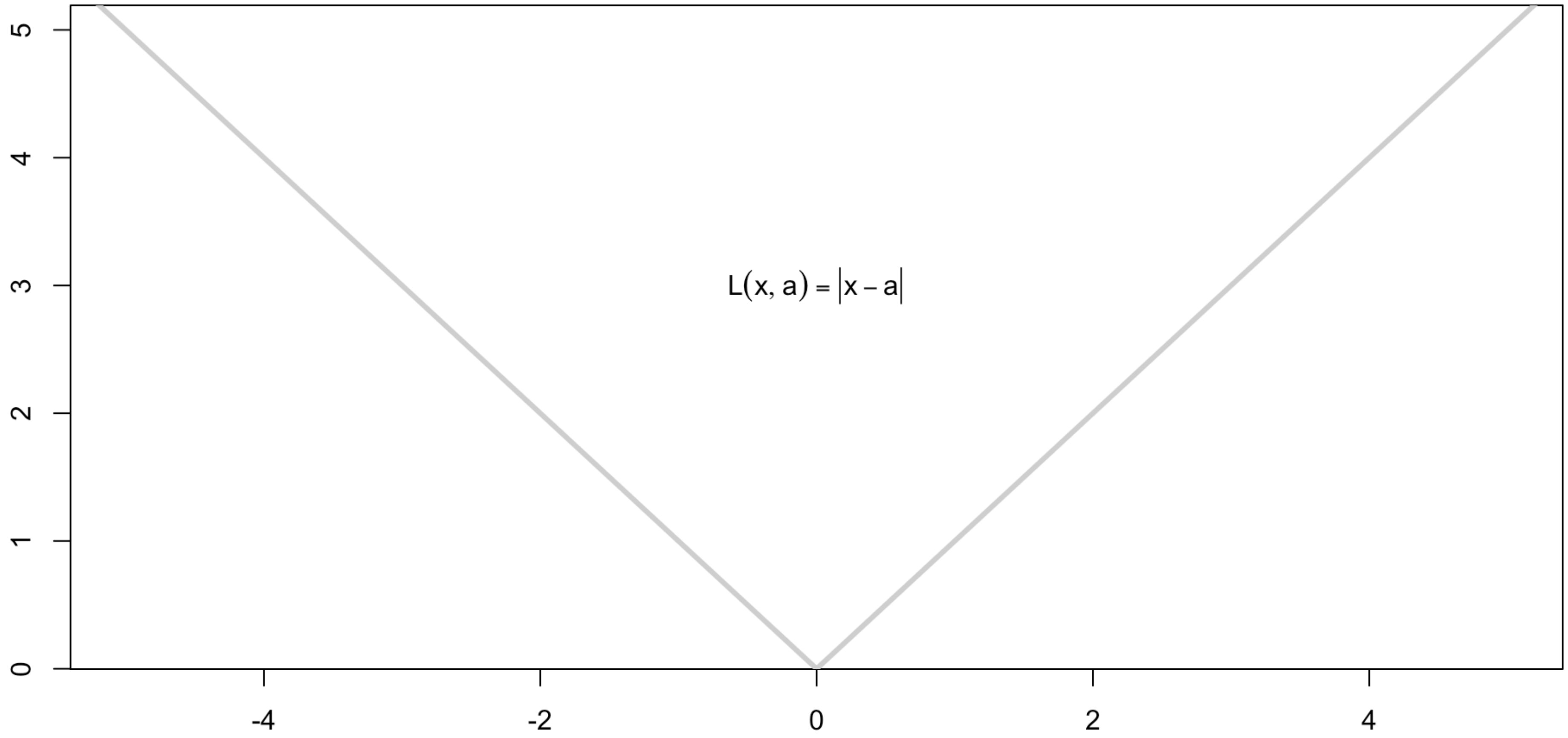
Loss Functions

Mathematical “rules” for optimization

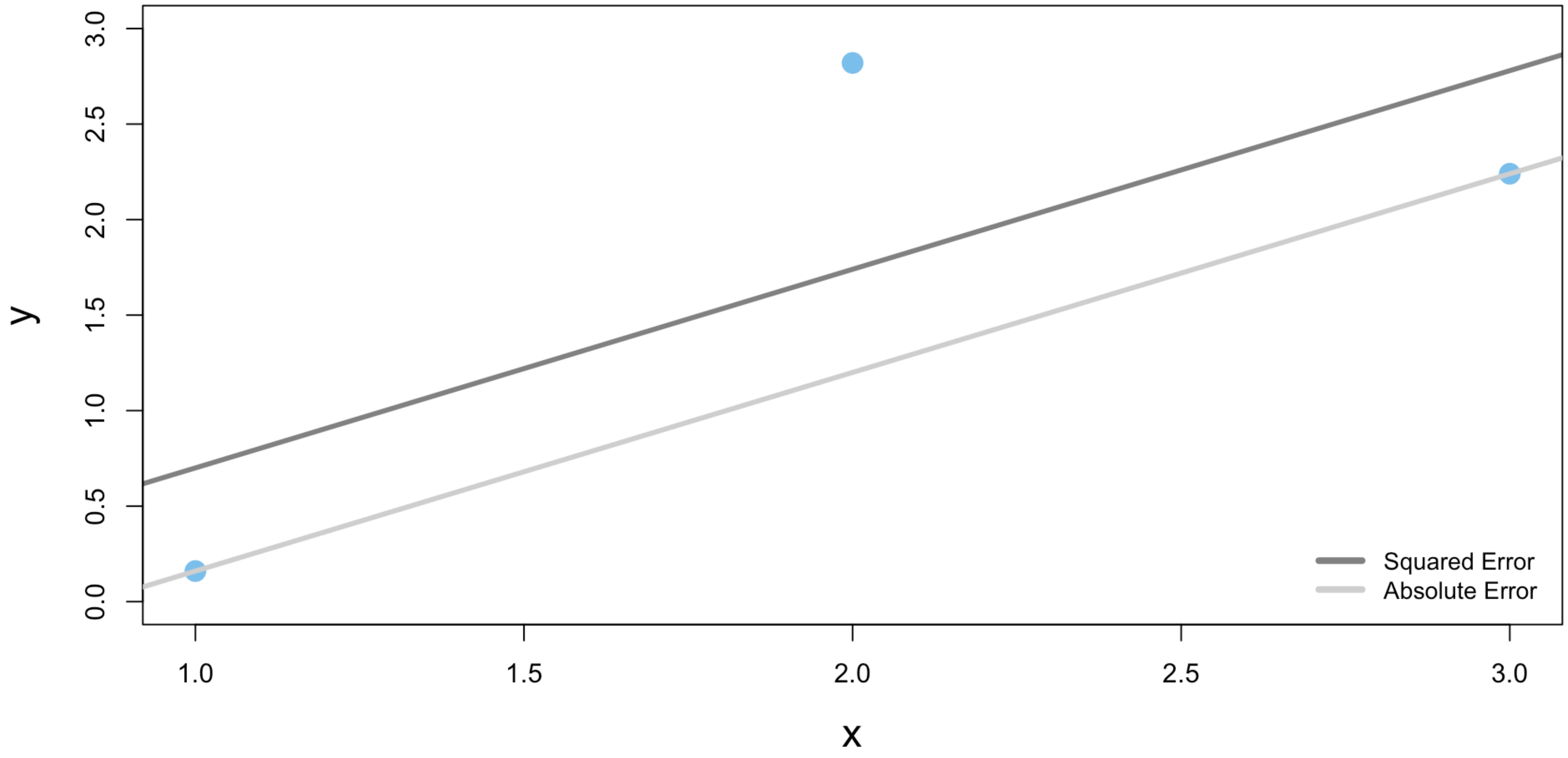
Squared Loss



Absolute Loss



Least Absolute Deviations

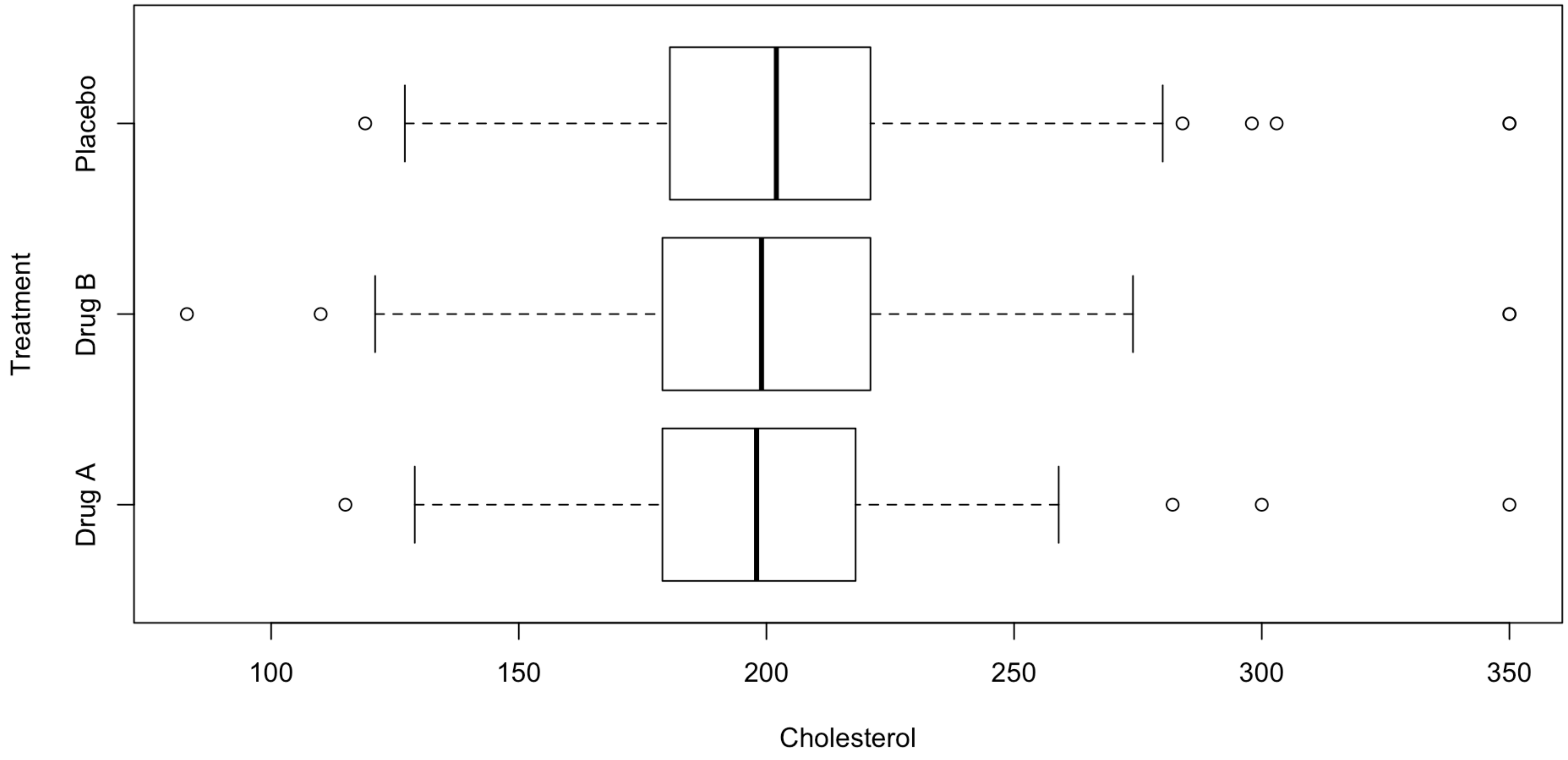


Least Absolute Deviations

x	y	Loss	$\hat{\beta}_0$	$\hat{\beta}_1$
1	0.16	Square	–0.34	1.04
2	2.82	Absolute	–0.88	1.04
3	2.24			

In-Practice

Which treatment?



Least Squares for Experiments

$$y_i = \beta_1 A_i + \beta_2 B_i + \beta_3 C_i + \epsilon_i$$

$$A_i = \begin{cases} 1 & \text{if Trt A} \\ 0 & \text{otherwise} \end{cases}$$

$$B_i = \begin{cases} 1 & \text{if Trt B} \\ 0 & \text{otherwise} \end{cases}$$

$$C_i = \begin{cases} 1 & \text{if Control} \\ 0 & \text{otherwise} \end{cases}$$

$$y_i = \beta_1 A_i$$

Results

$$\hat{y}_i = 199.2A_i + 199.6B_i + 202.0C_i$$

P-VALUE *

↳ t-TEST *

How old is the universe?

> In 1929 Edwin Hubble investigated the relationship between distance and radial velocity of extragalactic nebulae (celestial objects). It was hoped that some knowledge of this relationship might give clues as to the way the universe was formed and what may happen later. His findings revolutionised astronomy and are the source of much research today on the 'Big Bang'.

Derived Quantities

- We can use the results of regression equations to form "extra" results
- In this case the model is:

$$y = \beta x + \epsilon$$

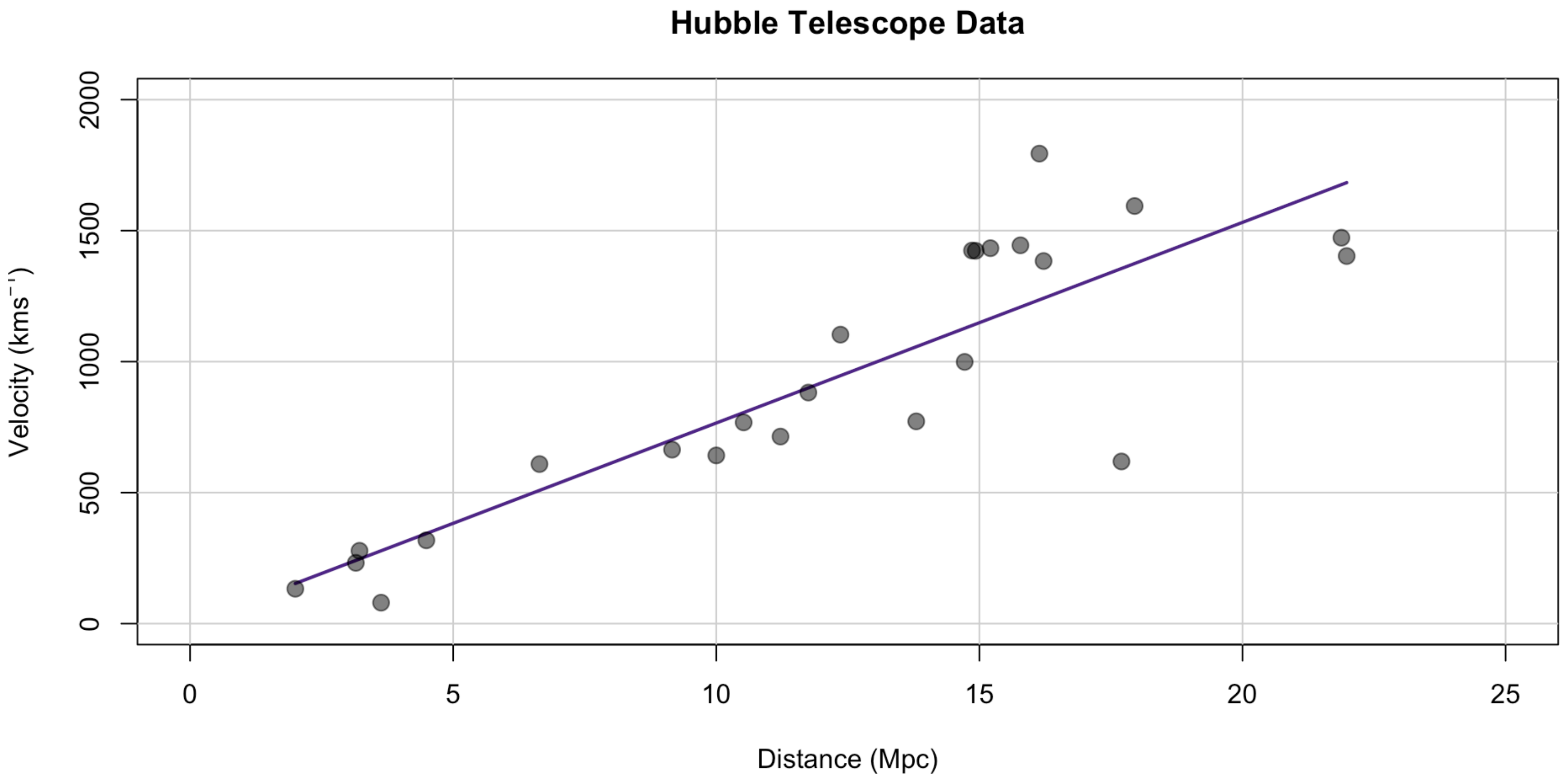
$\hat{\beta}_1$ ✓

$\hat{\beta}_0 = 0$

- Our derived quantity is β^{-1}

$x^{-1} = \frac{1}{x}$

The Data


Hubble Telescope Data
Velocity (kms⁻¹)
0
500
1000
1500
2000
0
5
10
15
20
25
Distance (Mpc)

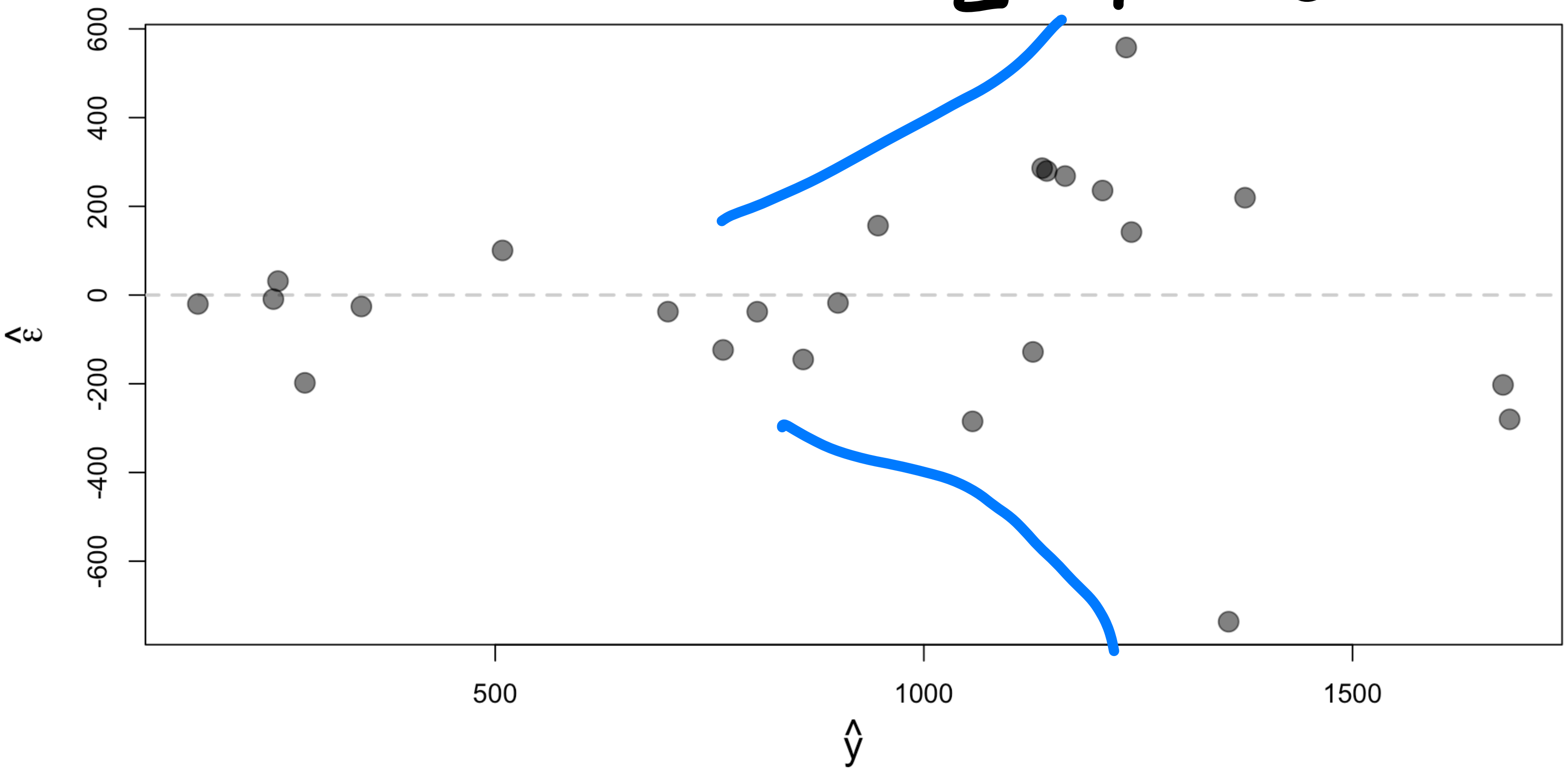
Assumption Checks

$E \, \epsilon_i = 0$

```r
# calculate the mean and round to 10 decimal places
round(mean(resids),10)
```

```
[1] 1.22088
```

$\sum \epsilon_i / n = 0$



Results

$$y = 76.58x$$

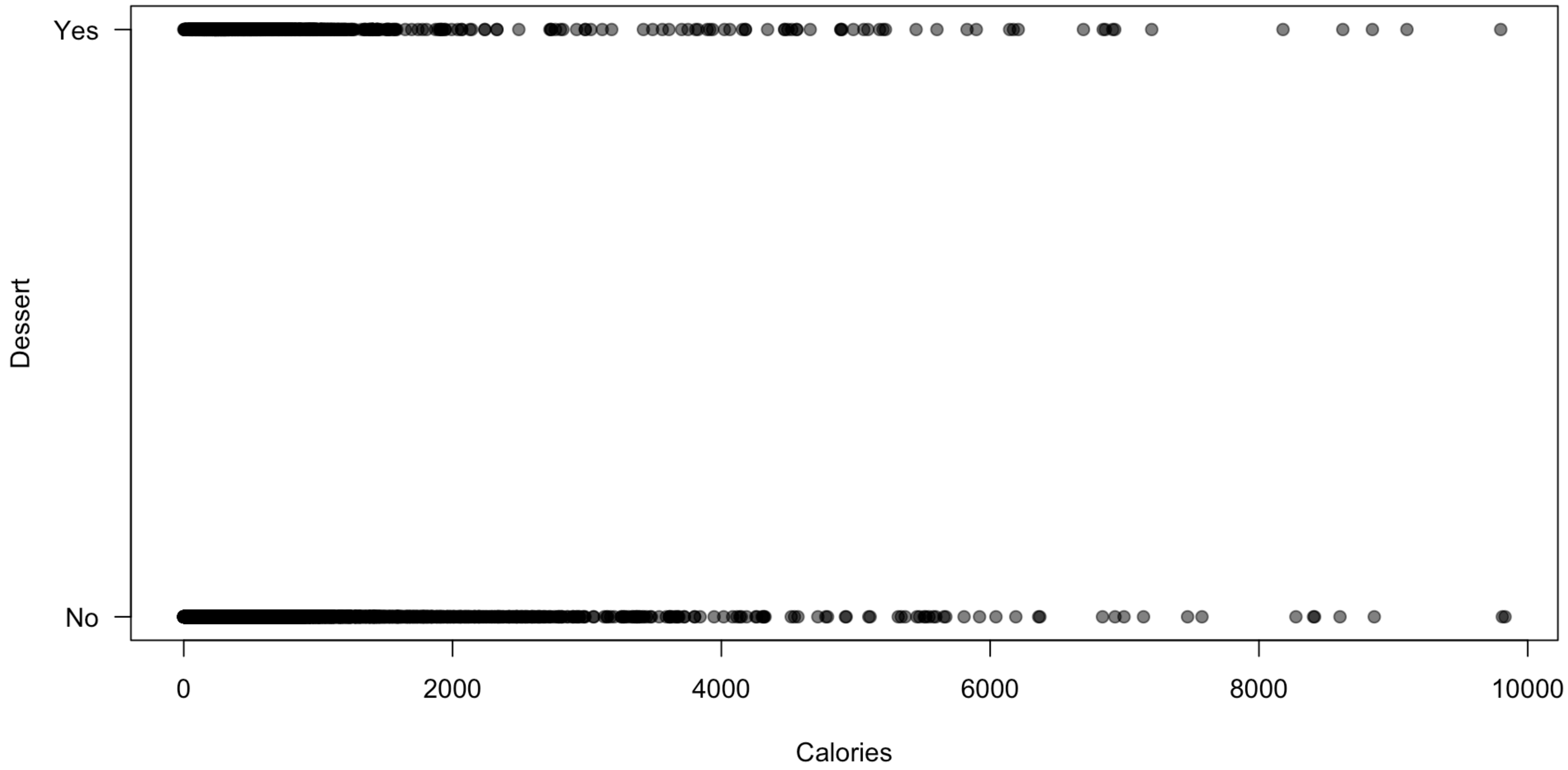
$$\text{Hubble Time} = \beta^{-1} \times 979.708$$

$$= \frac{1}{\beta} \times 979.708$$

$$= 979.708/\beta$$

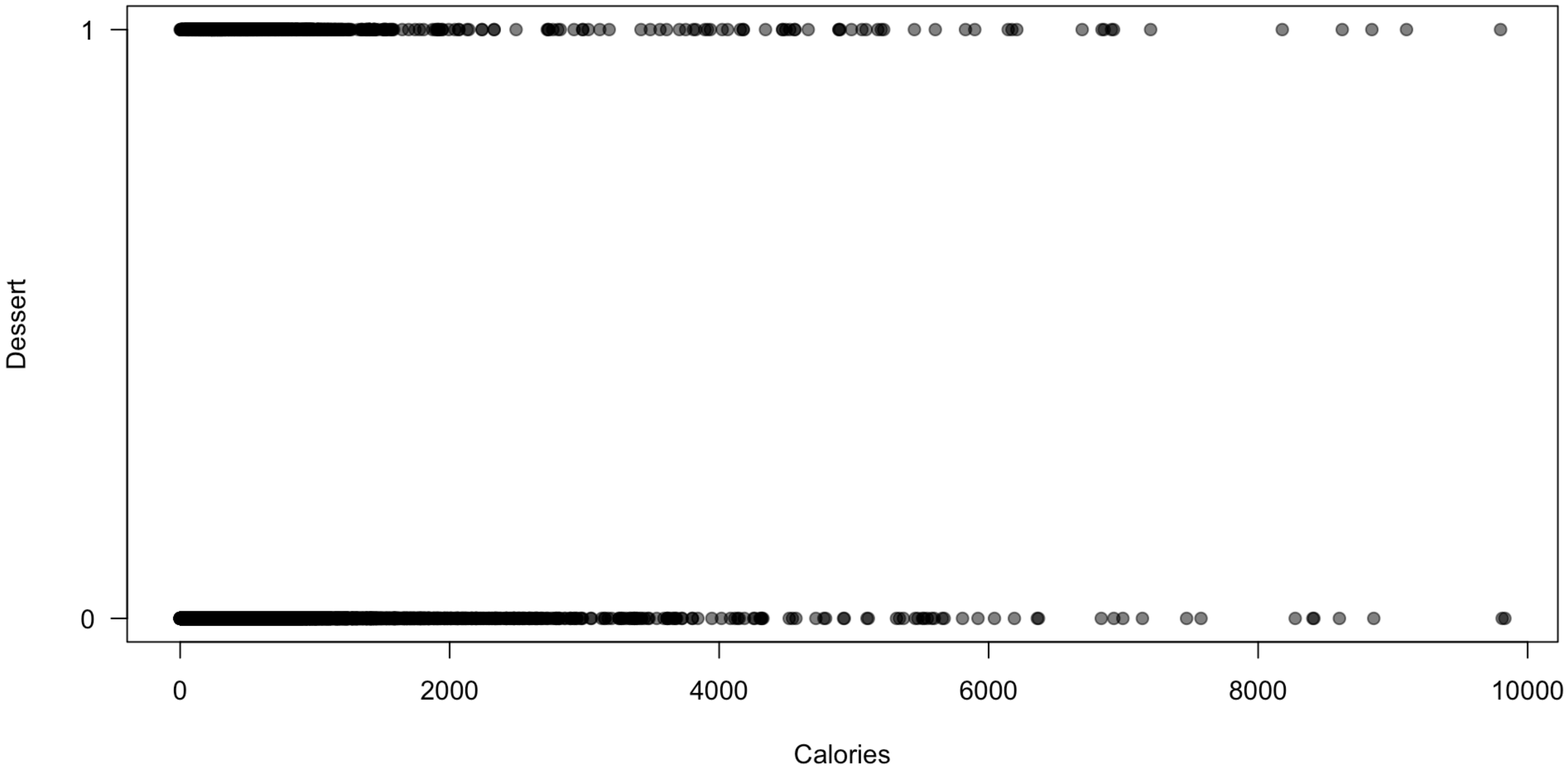
$$= 12.79 \text{ BILLION}$$

Is it cake?



Is it cake?

[0, 1] = Binary



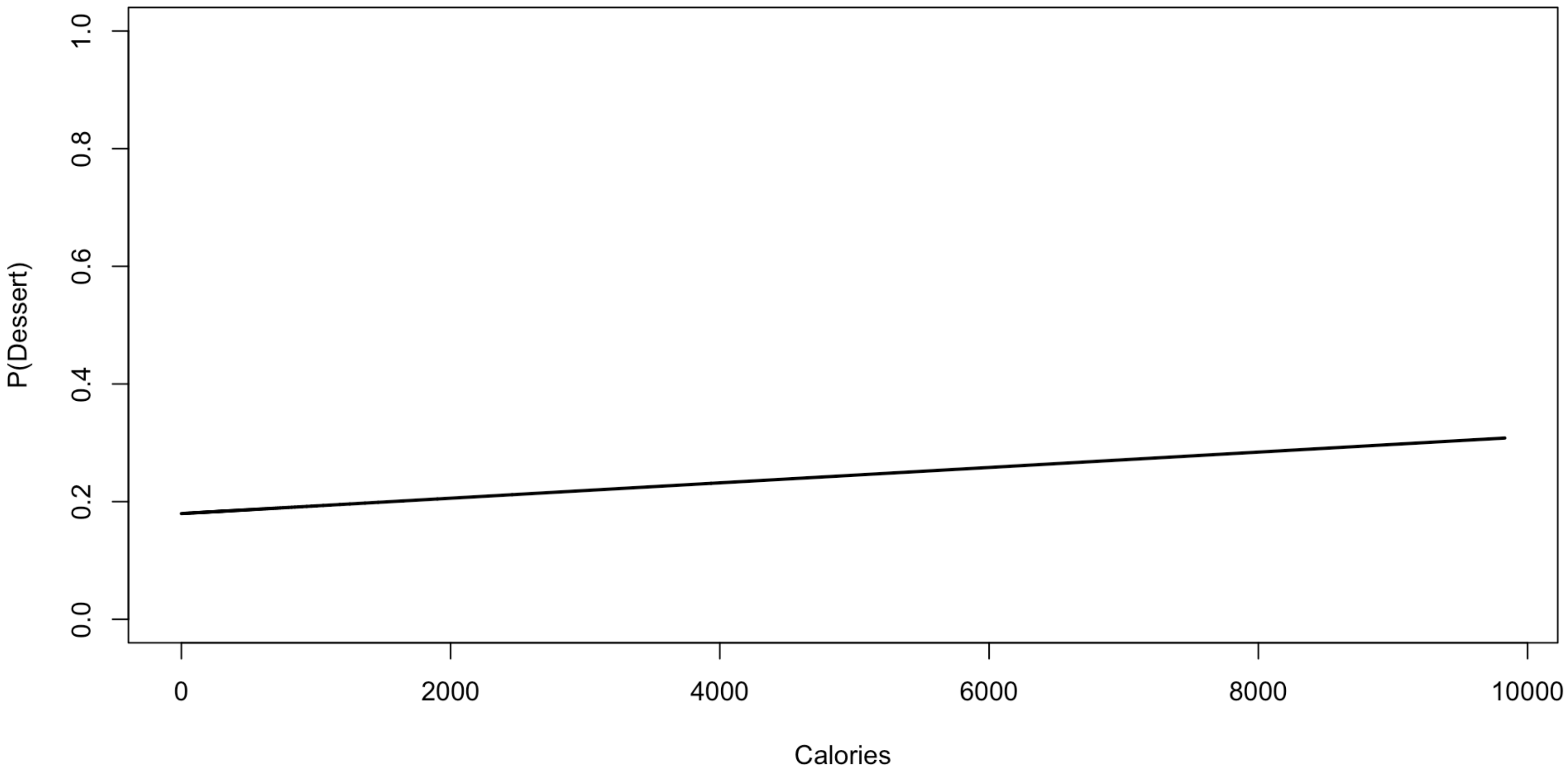
Is it cake?

$$y = \beta_0 + \beta_1 x + \epsilon_i$$

$$P_i = \beta_0 + \beta_1 C_i + \epsilon_i$$

- P_i will be the “probability of being a dessert” for the i-th entry.
- C_i will be the calories for the i-th entry.

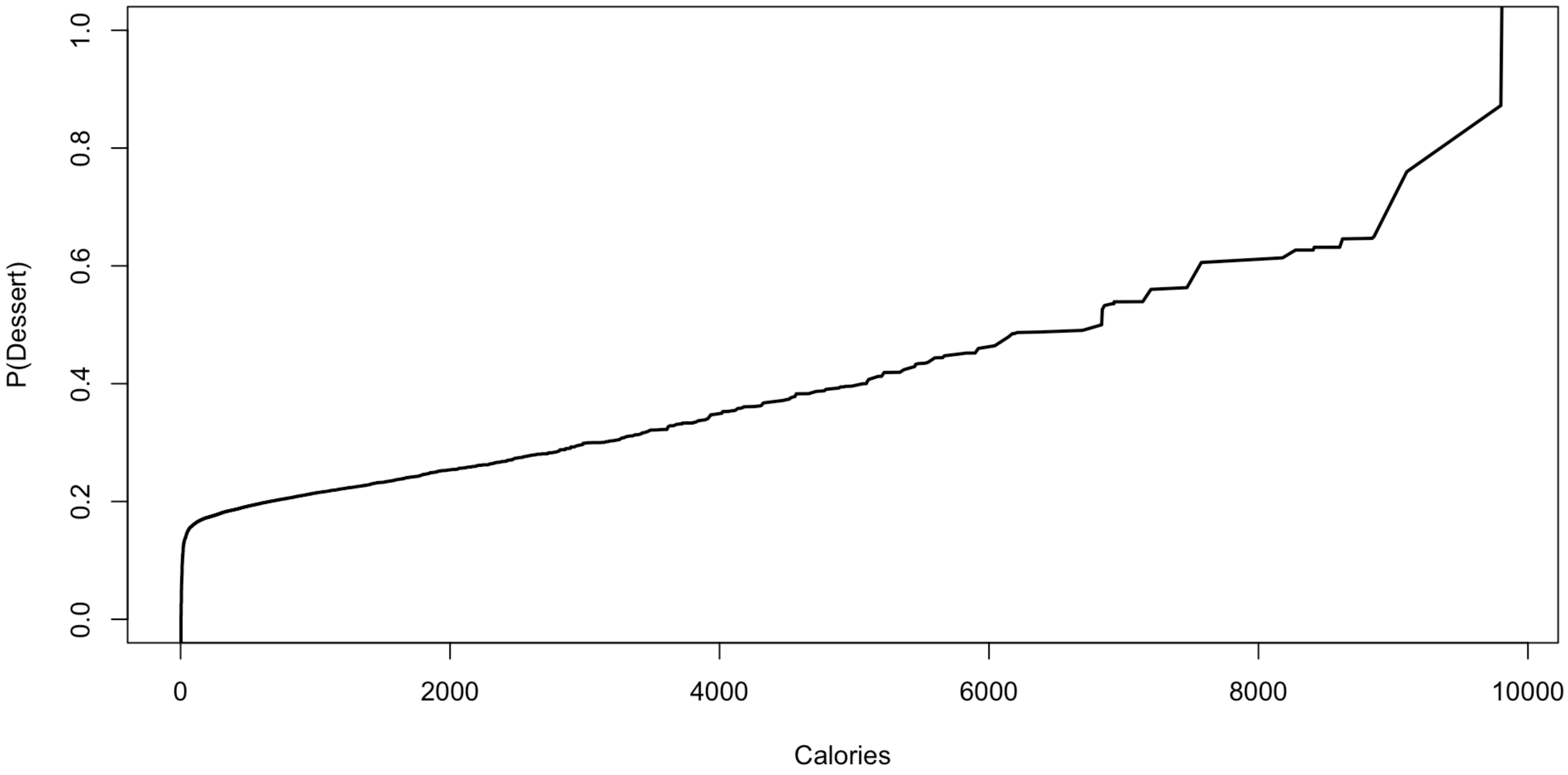
Is it cake?



$$\hat{P}_i = 0.18 + 0.000013C_i + \epsilon_i$$

More predictors

$$P_i = \beta_0 + \beta_1 C_i + \beta_2 F_i + \epsilon_i$$



DIRTY DAWGS

↳ COWBOY BOOTS → DIRTY

↳ FROSTED TIPS

↳ COWBOY HAT → PINK

↳ THURS

Go away

Go away