INTRODUCTION TO STATISTICS

LECTURE 13

LAST TIME

- Statistical tests
  - Parametric tests
  - Non-parametric tests
  - Practice in Python

- Two random variables:
  - Covariance
  - Correlation

TODAY

- Linear regression
- Recap

LOGISTICS

- Assignment 4 (part 2) was due yesterday
  - You can still submit

LOGISTICS

- Assignment 4 (part 2) was due yesterday
  - You can still submit
- Assignment 5
  - Part 1 – published yesterday, due Friday, December 17, 23:59.
  - Part 1 – will be published today, due Saturday, December 18, 23:59.

LOGISTICS

- Assignment 4 (part 2) was due yesterday
  - You can still submit

- Assignment 5
  - Part 1 – published yesterday, due Friday, December 17, 23:59.
  - Part 1 – will be published today, due Saturday, December 18, 23:59.

- Final exam
  - Tomorrow, Friday, December 18, 09:00 – 12:30
  - Available on Google Classroom (same as the mid-term)

LINEAR REGRESSION

REGRESSION

- Bivariate data $(x_i, y_i), i = 1, \ldots, n$.

REGRESSION

- Bivariate data $(x_i, y_i), i = 1, \ldots, n$.

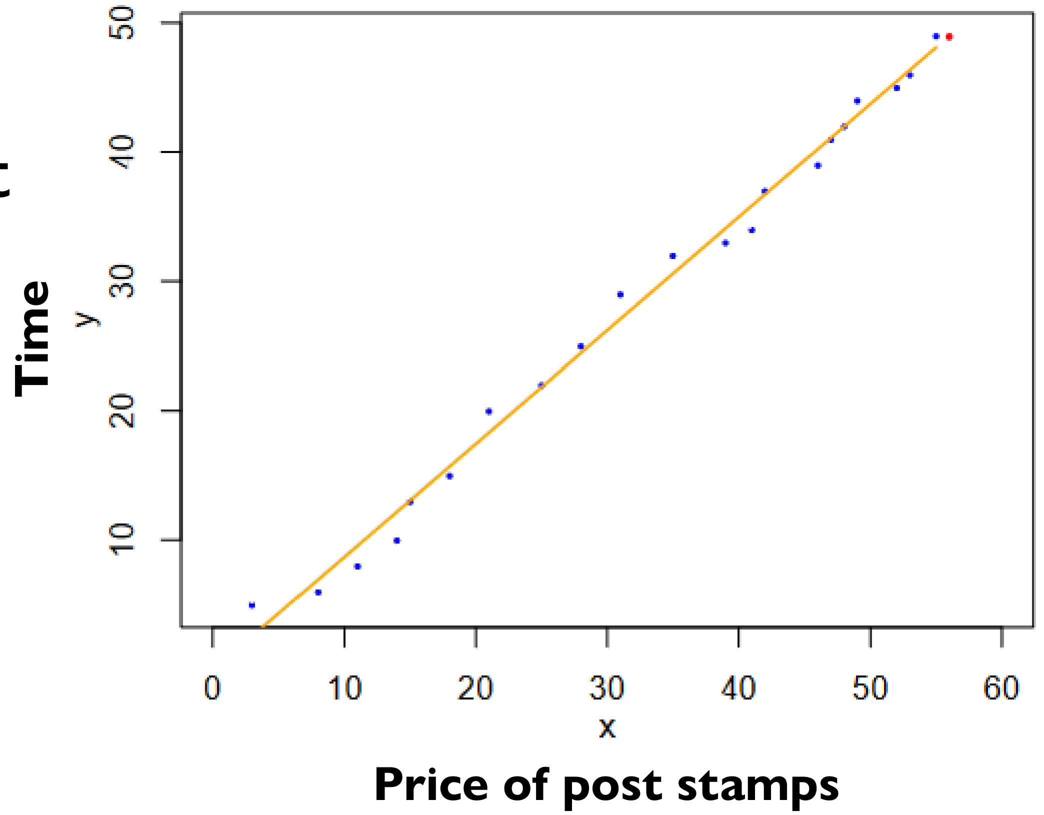
- The goal: model the relationship between x and y as $y = f(x)$ that is a close fit to the data.

REGRESSION

- Bivariate data $(x_i, y_i), i = 1, \ldots, n$.
- The goal: model the relationship between x and y as $y = f(x)$ that is a close fit to the data.
- **Assumptions**:
  - x_i is not random - **predictor**
  - y_i, is a function of xi plus some random noise - **response**

REGRESSION

- Bivariate data $(x_i, y_i), i = 1, \ldots, n$.
- The goal: model the relationship between x and y as $y = f(x)$ that is a close fit to the data.
- **Assumptions**:
  - x_i is not random - **predictor**
  - y_i, is a function of xi plus some random noise - **response**



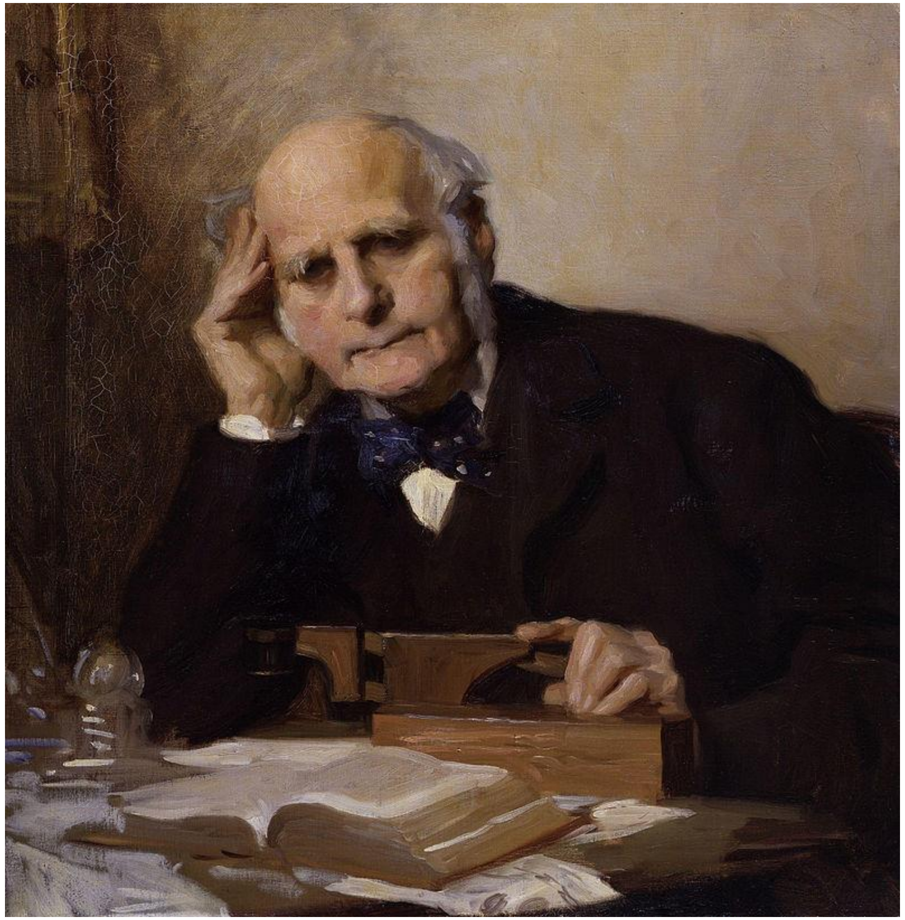
EXAMPLE

- Francis Galton, *second half of the 19th century*:

Suppose we have n pairs of fathers and adult sons.

Let x_i and y_i be the heights of the ith father and son, respectively.

Predict the adult height of a young boy from that of his father.



SIMPLE LINEAR REGRESSION

- Bivariate data $(x_i, y_i), i = 1, \ldots, n$
- $y = f(x)$

SIMPLE LINEAR REGRESSION

- Bivariate data $(x_i, y_i), i = 1, \ldots, n$
- $y = f(x)$
- Linear $f : y = ax + b$

SIMPLE LINEAR REGRESSION

- Bivariate data $(x_i, y_i), i = 1, \ldots, n$
- $y = f(x)$
- Linear f : $y = ax + b$
- Our model will predict y_i up till some error ε_i:

$$y_i = ax_i + b + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2)$$

SIMPLE LINEAR REGRESSION

- Bivariate data $(x_i, y_i), i = 1, \dots, n$
- $y = f(x)$
- Linear f : $y = ax + b$
- Our model will predict y_i up till some error ε_i:

$$y_i = ax_i + b + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2)$$

How to chose a and b?

FINDING THE BEST FIT

$$y_i = ax_i + b + \varepsilon_i$$

How to chose a and b?

FINDING THE BEST FIT

$$y_i = ax_i + b + \varepsilon_i$$

How to chose a and b?

$$\varepsilon_i = y_i - ax_i - b$$

FINDING THE BEST FIT

$$y_i = ax_i + b + \varepsilon_i$$

How to chose a and b?

$$\varepsilon_i = y_i - ax_i - b$$

$$\varepsilon_i^2 = (y_i - ax_i - b)^2$$

FINDING THE BEST FIT

$$y_i = ax_i + b + \varepsilon_i$$

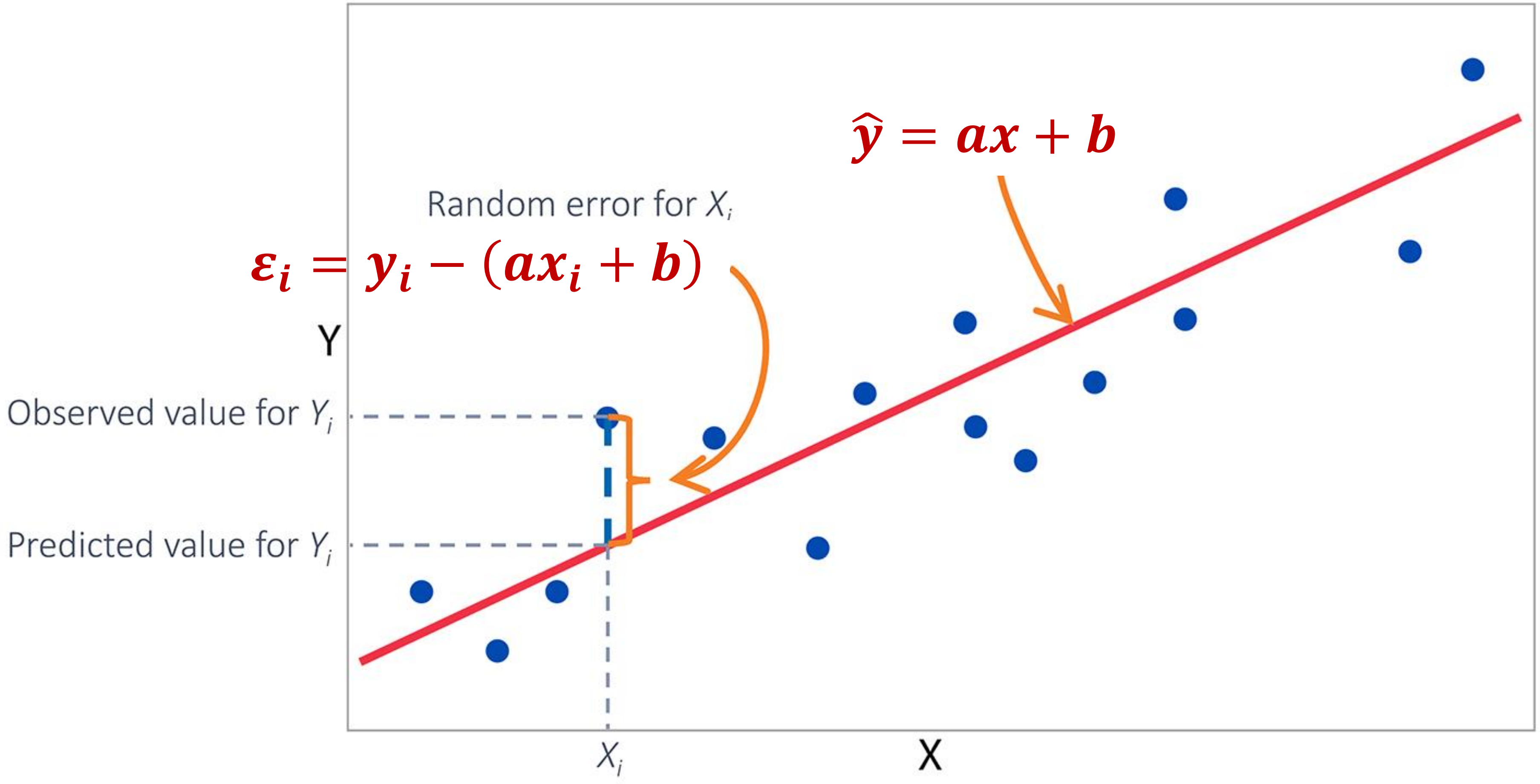
How to chose a and b?

$$\varepsilon_i = y_i - ax_i - b$$

$$\varepsilon_i^2 = (y_i - ax_i - b)^2$$

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - ax_i - b)^2 \to \min \quad w.r.t.\ a, b$$

FINDING THE BEST FIT



FINDING THE BEST FIT

$$y_i = ax_i + b + \varepsilon_i$$

How to chose a and b?

$$\varepsilon_i = y_i - ax_i - b$$

$$\varepsilon_i^2 = (y_i - ax_i - b)^2$$

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - ax_i - b)^2 \rightarrow \min \quad w.r.t.\ \ a, b$$

METHOD OF LEAST SQUARES

FINDING THE BEST FIT

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - ax_i - b)^2 \to \min \quad w.r.t.\ \ a, b$$

Partial derivatives:

FINDING THE BEST FIT

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - ax_i - b)^2 \to \min \quad w.r.t.\ \ a, b$$

Partial derivatives:

$$-2\sum_{i=1}^{n} x_i(y_i - ax_i - b) = 0$$

FINDING THE BEST FIT

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - a x_i - b)^2 \to \min \quad w.r.t.\ a, b$$

Partial derivatives:

$$-2 \sum_{i=1}^{n} x_i (y_i - a x_i - b) = 0$$

$$-2 \sum_{i=1}^{n} (y_i - a x_i - b) = 0$$

FINDING THE BEST FIT

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - ax_i - b)^2 \to \min \qquad w.r.t.\ \ a, b$$

Solution:

$$\boldsymbol{a = \frac{s_{xy}}{s_{xx}}}, \qquad \boldsymbol{b = \bar{y} - a\bar{x}}$$

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})\,, \qquad s_{xx} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

MAKE IT FIT

- Imagine that you've got the following data points:

$$(0, 1) \;\; (2, 3) \;\; (1, -1)$$

- Fit a simple linear regression model.

MAKE IT FIT

- Imagine that you've got the following data points:

$$(0,1)\ \ (2,3)\ \ (1,-1)$$

- Fit a simple linear regression model.

$$\boldsymbol{a} = \frac{\boldsymbol{s_{xy}}}{\boldsymbol{s_{xx}}}, \qquad \boldsymbol{b} = \boldsymbol{\bar{y}} - \boldsymbol{a\bar{x}}$$

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$s_{xx} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

MAKE IT FIT

- Imagine that you've got the following data points:

$$(0, 1)\ \ (2, 3)\ \ (1, -1)$$

- Fit a simple linear regression model.

$$\boldsymbol{a = \frac{s_{xy}}{s_{xx}},} \qquad \boldsymbol{b = \bar{y} - a\bar{x}}$$

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$s_{xx} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\bar{x} = \frac{1+2+1}{3} = 1$$

$$\bar{y} = \frac{1+3-1}{3} = 1$$

MAKE IT FIT

- Imagine that you've got the following data points:

$$(0, 1)\ \ (2, 3)\ \ (1, -1)$$

- Fit a simple linear regression model.

$$\boldsymbol{a} = \frac{\boldsymbol{s_{xy}}}{\boldsymbol{s_{xx}}}, \qquad \boldsymbol{b} = \boldsymbol{\bar{y}} - \boldsymbol{a\bar{x}}$$

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$s_{xx} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\bar{x} = \frac{1+2+1}{3} = 1$$

$$\bar{y} = \frac{1+3-1}{3} = 1$$

$$s_{xx} = \frac{(0-1)^2 + (2-1)^2 + (1-1)^2}{3-1} = 1$$

$$s_{xy} = \frac{(-1)\cdot 0 + 1\cdot 2 + 0\cdot(-2)}{3-1} = 1$$

MAKE IT FIT

- Imagine that you've got the following data points:

$$(0, 1)\ \ (2, 3)\ \ (1, -1)$$

- Fit a simple linear regression model.

$$\boldsymbol{a = \frac{s_{xy}}{s_{xx}}}, \qquad \boldsymbol{b = \bar{y} - a\bar{x}}$$

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$s_{xx} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\bar{x} = \frac{1 + 2 + 1}{3} = 1$$

$$\bar{y} = \frac{1 + 3 - 1}{3} = 1$$

$$s_{xx} = \frac{(0-1)^2 + (2-1)^2 + (1-1)^2}{3-1} = 1$$

$$s_{xy} = \frac{(-1)\cdot 0 + 1 \cdot 2 + 0 \cdot (-2)}{3-1} = 1$$

$$\boldsymbol{a = \frac{1}{1} = 1}, \qquad \boldsymbol{b = 1 - 1 \cdot 1 = 0}$$

MAKE IT FIT

- Imagine that you've got the following data points:

$$(0,1)\ \ (2,3)\ \ (1,-1)$$

- Fit a simple linear regression model.

$$\boldsymbol{a = \frac{s_{xy}}{s_{xx}}}, \qquad \boldsymbol{b = \bar{y} - a\bar{x}}$$

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$s_{xx} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$\bar{x} = \frac{1+2+1}{3} = 1$$

$$\bar{y} = \frac{1+3-1}{3} = 1$$

$$s_{xx} = \frac{(0-1)^2 + (2-1)^2 + (1-1)^2}{3-1} = 1$$

$$s_{xy} = \frac{(-1)\cdot 0 + 1 \cdot 2 + 0 \cdot (-2)}{3-1} = 1$$

$$\boldsymbol{a = \frac{1}{1} = 1}, \qquad \boldsymbol{b = 1 - 1 \cdot 1 = 0}$$

$$\boldsymbol{y = x}$$

QUALITY OF THE FIT

- How good is our model? Some definitions:

QUALITY OF THE FIT

- How good is our model? Some definitions:
  - **TOTAL** Sum of Squares: how much variation in there in y?

$$SS_{tot} = \sum (y_i - \bar{y})^2$$

QUALITY OF THE FIT

- How good is our model? Some definitions:
  - **TOTAL** Sum of Squares: how much variation in there in y?

$$SS_{tot} = \sum (y_i - \bar{y})^2$$

  - **MODEL** Sum of Squares: how much of it the model explains?

$$SS_{mod} = \sum (\hat{y}_i - \bar{y})^2$$

QUALITY OF THE FIT

- How good is our model? Some definitions:
  - **TOTAL** Sum of Squares: how much variation in there in y?

$$SS_{tot} = \sum (y_i - \bar{y})^2$$

  - **MODEL** Sum of Squares: how much of it the model explains?

$$SS_{mod} = \sum (\hat{y}_i - \bar{y})^2$$

  - **RESIDUAL** Sum of Squares: how much the model doesn't explain?

$$SS_{res} = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - ax_i - b)^2$$

QUALITY OF THE FIT

$$SS_{tot} = \sum (y_i - \bar{y})^2 , \qquad SS_{mod} = \sum (\hat{y}_i - \bar{y})^2 , \qquad SS_{res} = \sum (y_i - \hat{y}_i)^2$$

$$SS_{tot} = SS_{mod} + SS_{res}$$

QUALITY OF THE FIT

$$SS_{tot} = \sum (y_i - \bar{y})^2 , \qquad SS_{mod} = \sum (\hat{y}_i - \bar{y})^2 , \qquad SS_{res} = \sum (y_i - \hat{y}_i)^2$$

$$SS_{tot} = SS_{mod} + SS_{res}$$

- **Coefficient of determination**: explained variation

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = \frac{SS_{mod}}{SS_{tot}}$$

QUALITY OF THE FIT - EXAMPLE

- Datapoints $(0, 1)$ $(2, 3)$ $(1, -1)$
- Estimated regression line: $y = x$
- How good is the fit?

QUALITY OF THE FIT - EXAMPLE

- Datapoints $(0, 1)$ $(2, 3)$ $(1, -1)$
- Estimated regression line: $y = x$
- How good is the fit?

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} =$$

QUALITY OF THE FIT - EXAMPLE

- Datapoints $(0, 1)$ $(2, 3)$ $(1, -1)$
- Estimated regression line: $y = x$
- How good is the fit?

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} =$$

$$SS_{tot} =$$

$$SS_{res} =$$

QUALITY OF THE FIT - EXAMPLE

- Datapoints $(0, 1)$ $(2, 3)$ $(1, -1)$
- Estimated regression line: $y = x$
- How good is the fit?

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} =$$

$$SS_{tot} = \frac{(1-1)^2 + (3-1)^2 + (-1-1)^2}{3} = \frac{8}{3}$$

$$SS_{res} =$$

QUALITY OF THE FIT - EXAMPLE

- Datapoints $(0, 1)$ $(2, 3)$ $(1, -1)$
- Estimated regression line: $y = x$
- How good is the fit?

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} =$$

$$SS_{tot} = \frac{(1-1)^2 + (3-1)^2 + (-1-1)^2}{3} = \frac{8}{3}$$

$$SS_{res} = \frac{(0-1)^2 + (2-3)^2 + (1+1)^2}{3} = \frac{6}{3}$$

QUALITY OF THE FIT - EXAMPLE

- Datapoints $(0, 1)$ $(2, 3)$ $(1, -1)$
- Estimated regression line: $y = x$
- How good is the fit?

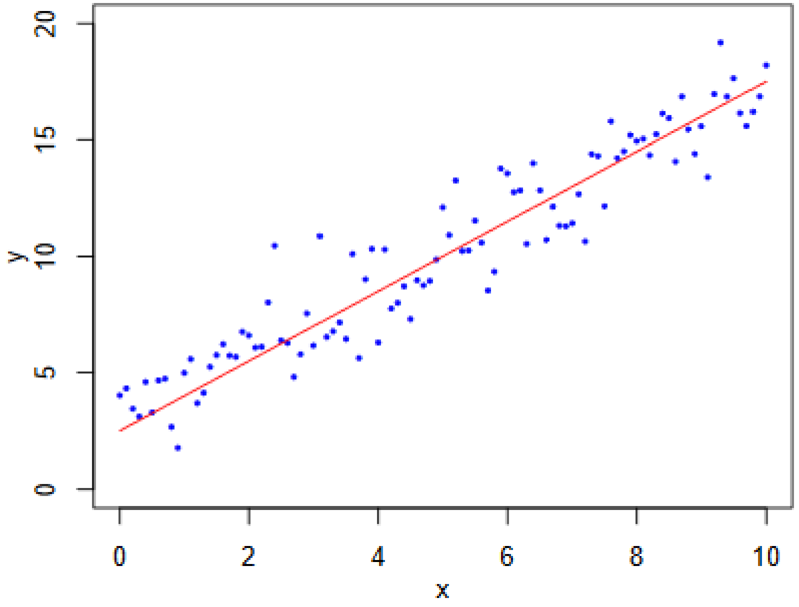
$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{6}{8} = 0.25$$

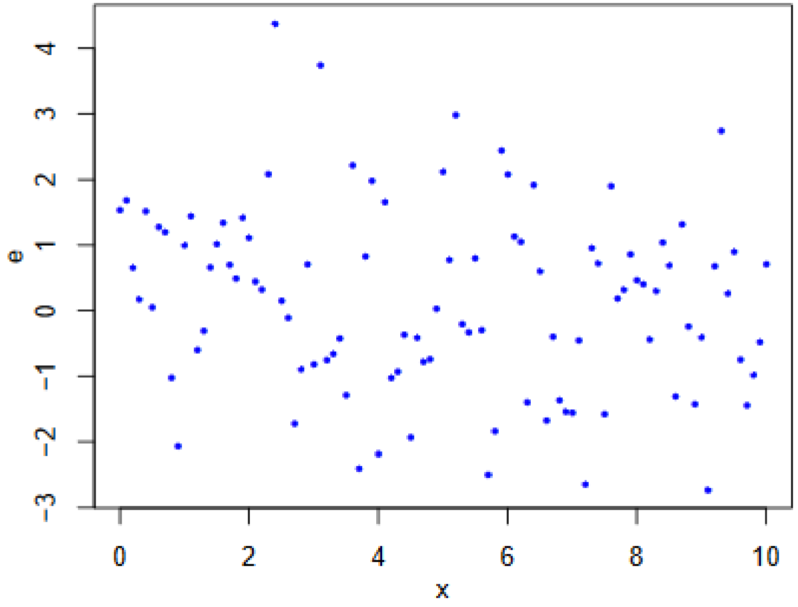
$$SS_{tot} = \frac{(1-1)^2 + (3-1)^2 + (-1-1)^2}{3} = \frac{8}{3}$$

$$SS_{res} = \frac{(0-1)^2 + (2-3)^2 + (1+1)^2}{3} = \frac{6}{3}$$

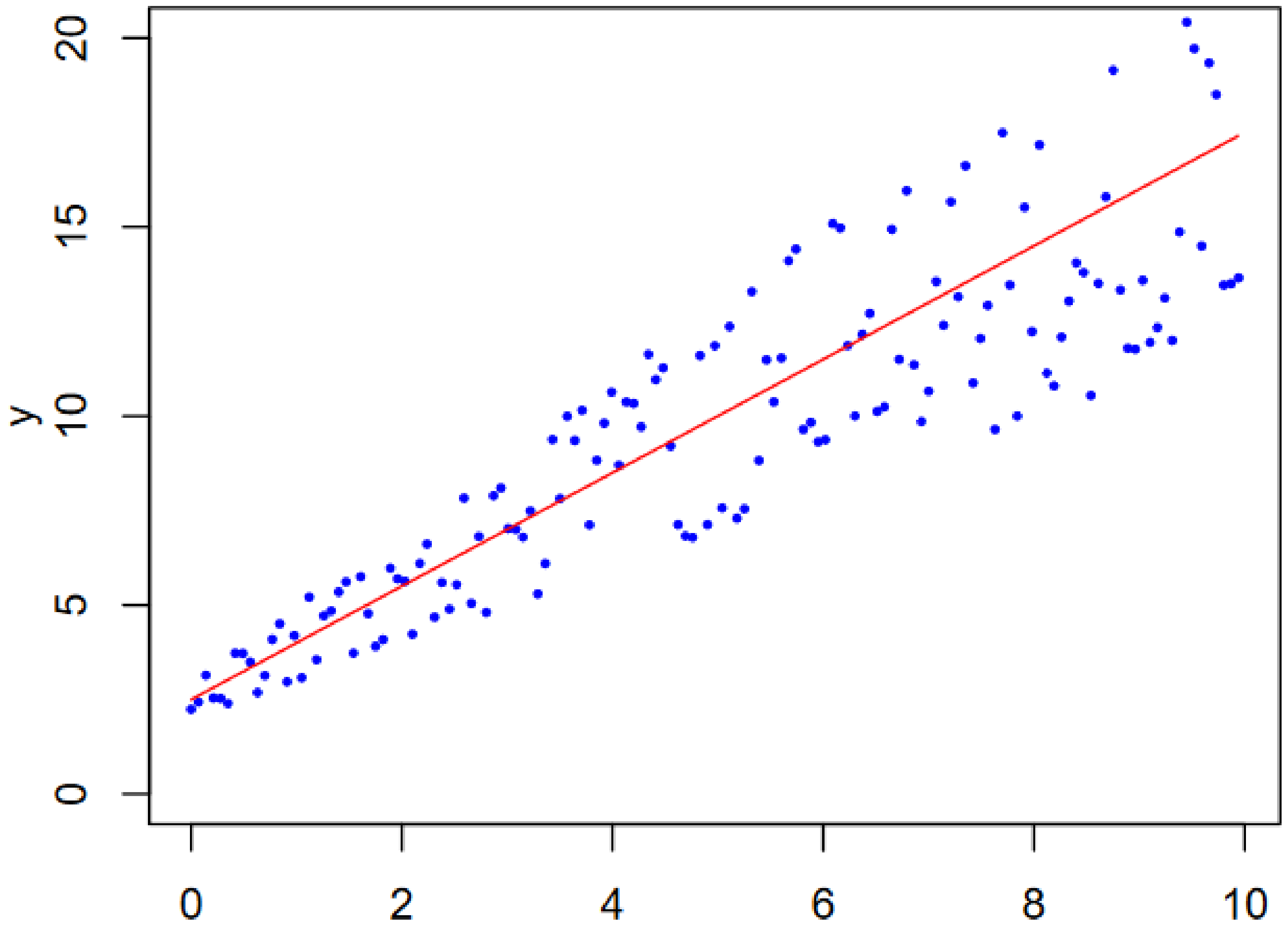
ASSUMPTIONS

- Simple linear regression: $y_i = ax_i + b + \varepsilon_i$
- Assumption: $\varepsilon_i \sim N(0, \sigma^2)$
- **Homoscedasticity**: errors are uniformly distributed around the regression line



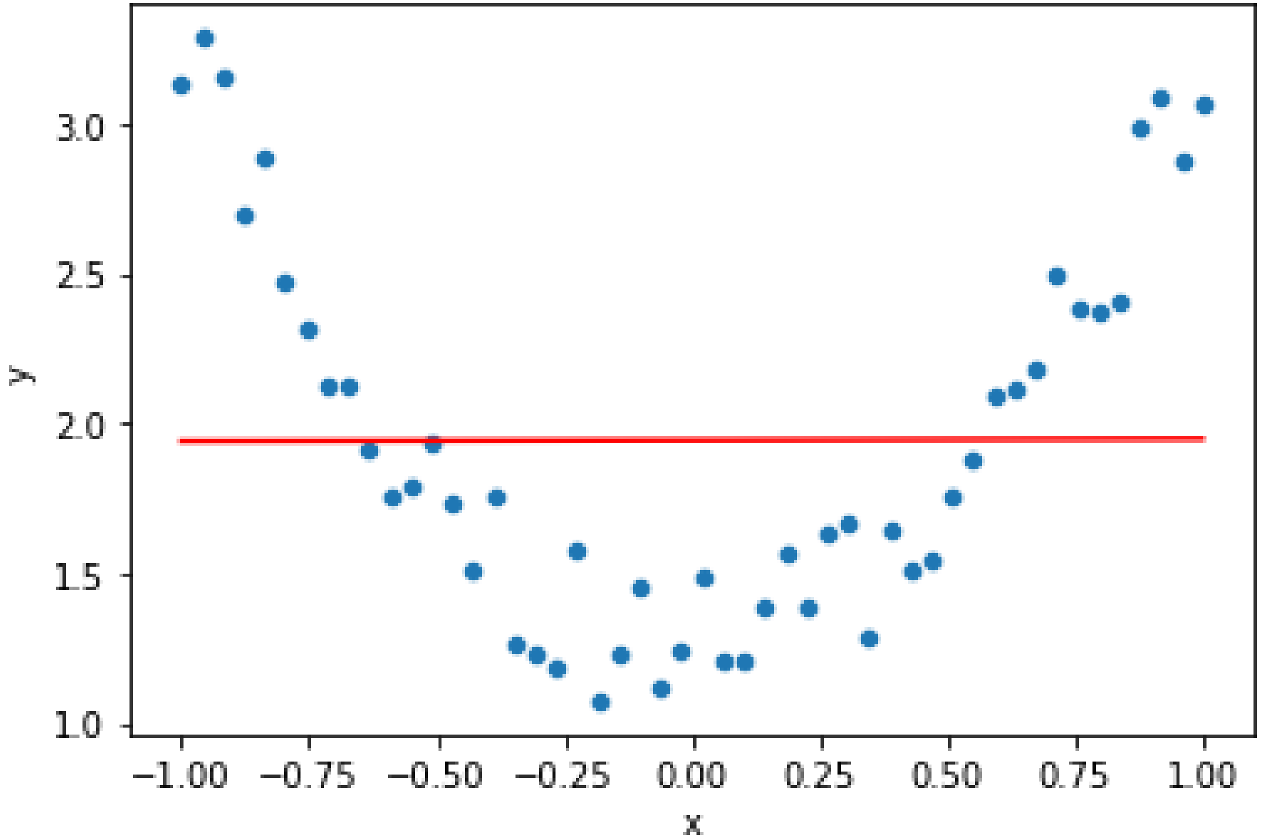


HETEROSCEDASTIC



WHAT IS LINEAR THERE

- Not all the data is linear



WHAT IS LINEAR THERE

Given the data $(x_1, y_1), \ldots, (x_n, y_n)$, is the following a simple linear regression model?

$$y_i = a x_i^2 + b + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2)$$

WHAT IS LINEAR THERE

Given the data $(x_1, y_1), \dots, (x_n, y_n)$, is the following a simple linear regression model?

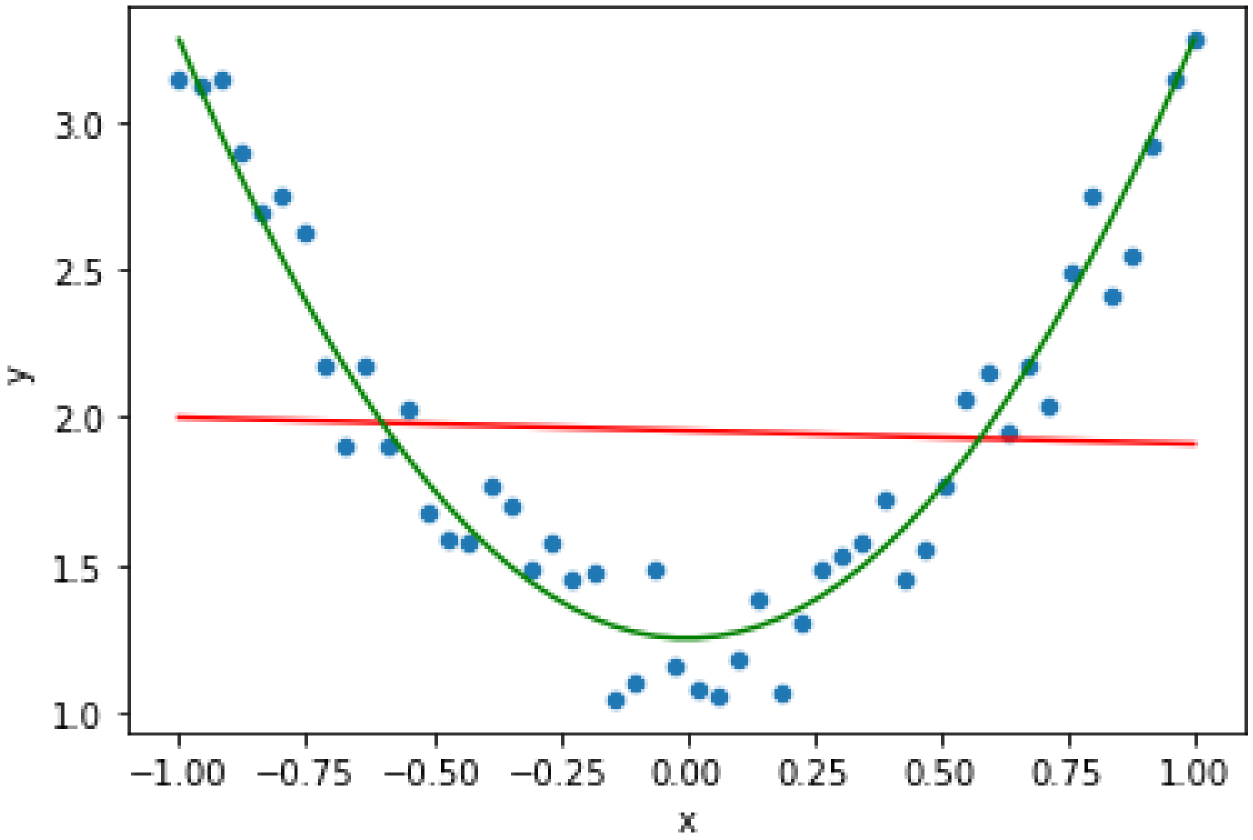
$$y_i = ax_i^2 + b + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2)$$

YES!

Linear in terms of <u>parameters</u> a, b, not in terms of the data.

WHAT IS LINEAR THERE

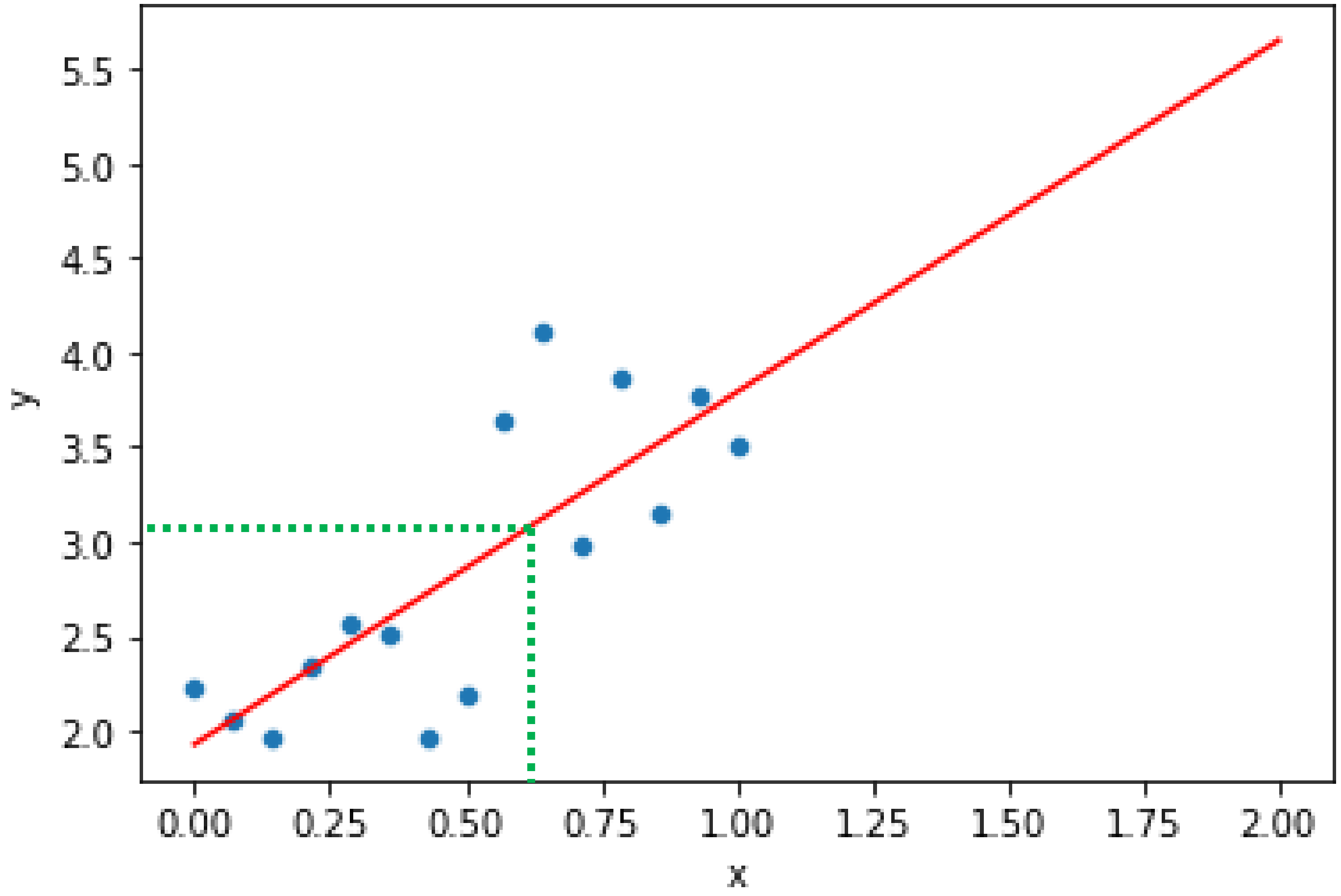
$$y_i = a x_i^2 + b + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2)$$



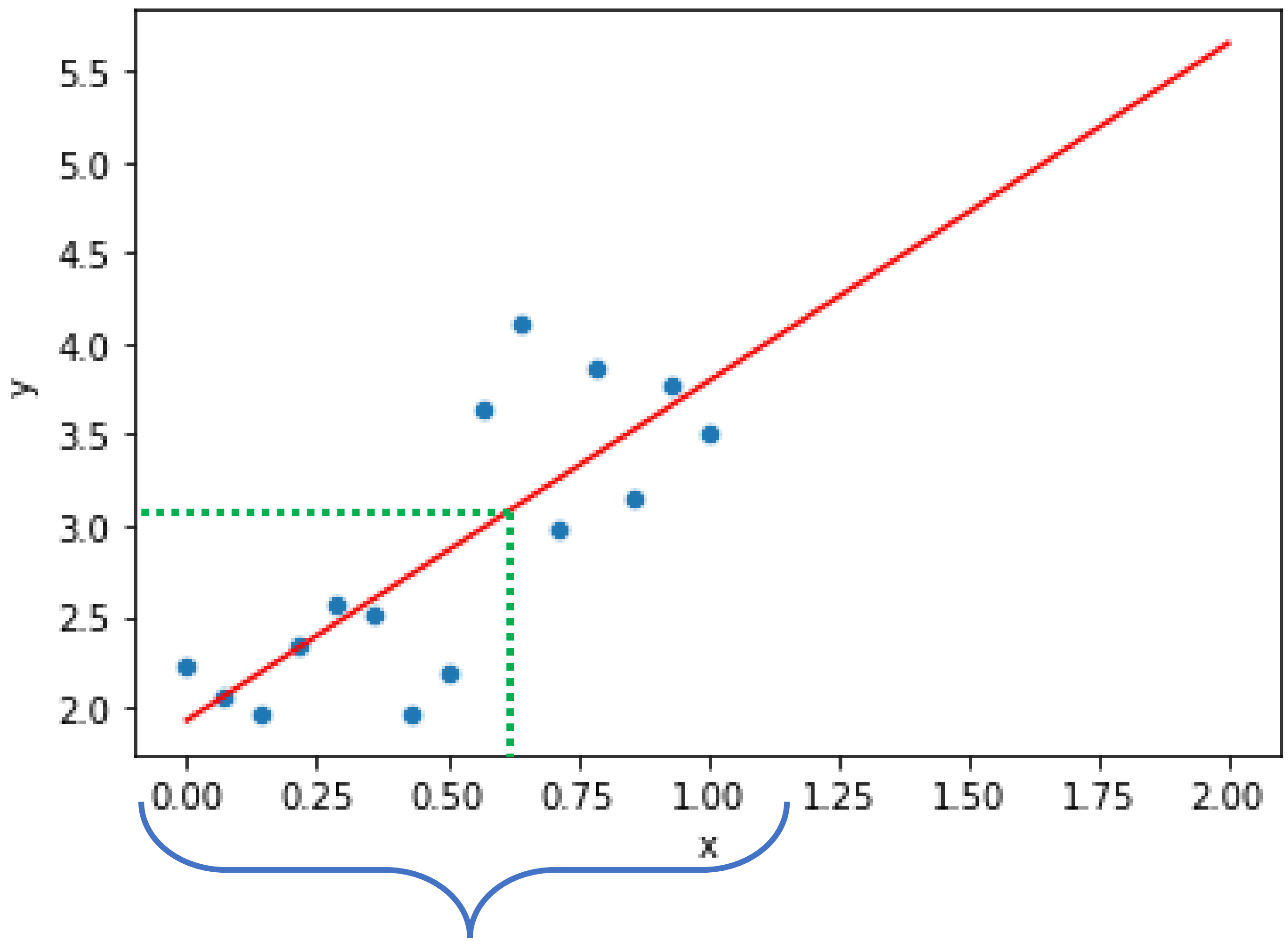
PRACTICE!

Google Classroom -> Lecture 12 -> Simple Linear Regression

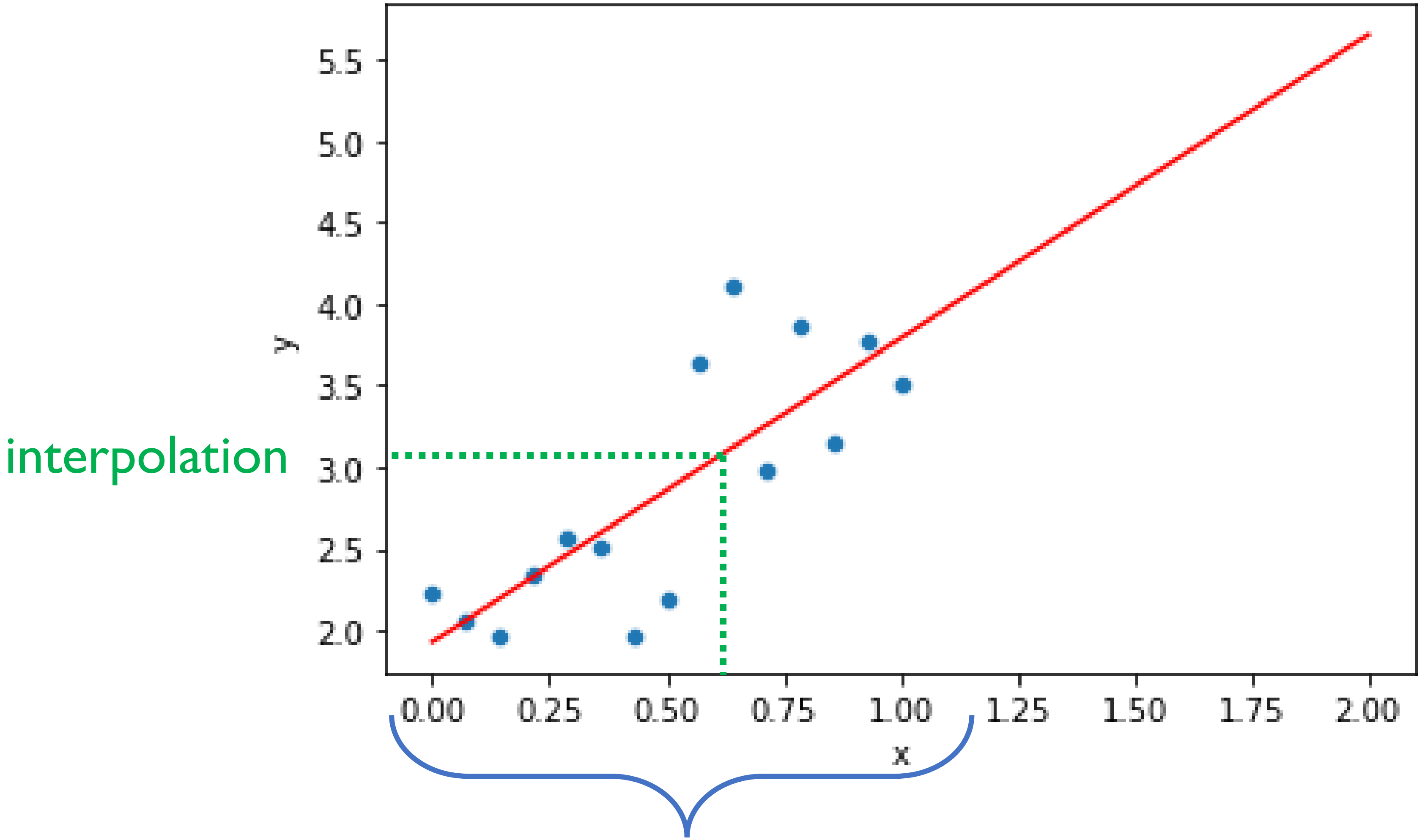
INTERPOLATION VS EXTRAPOLATION



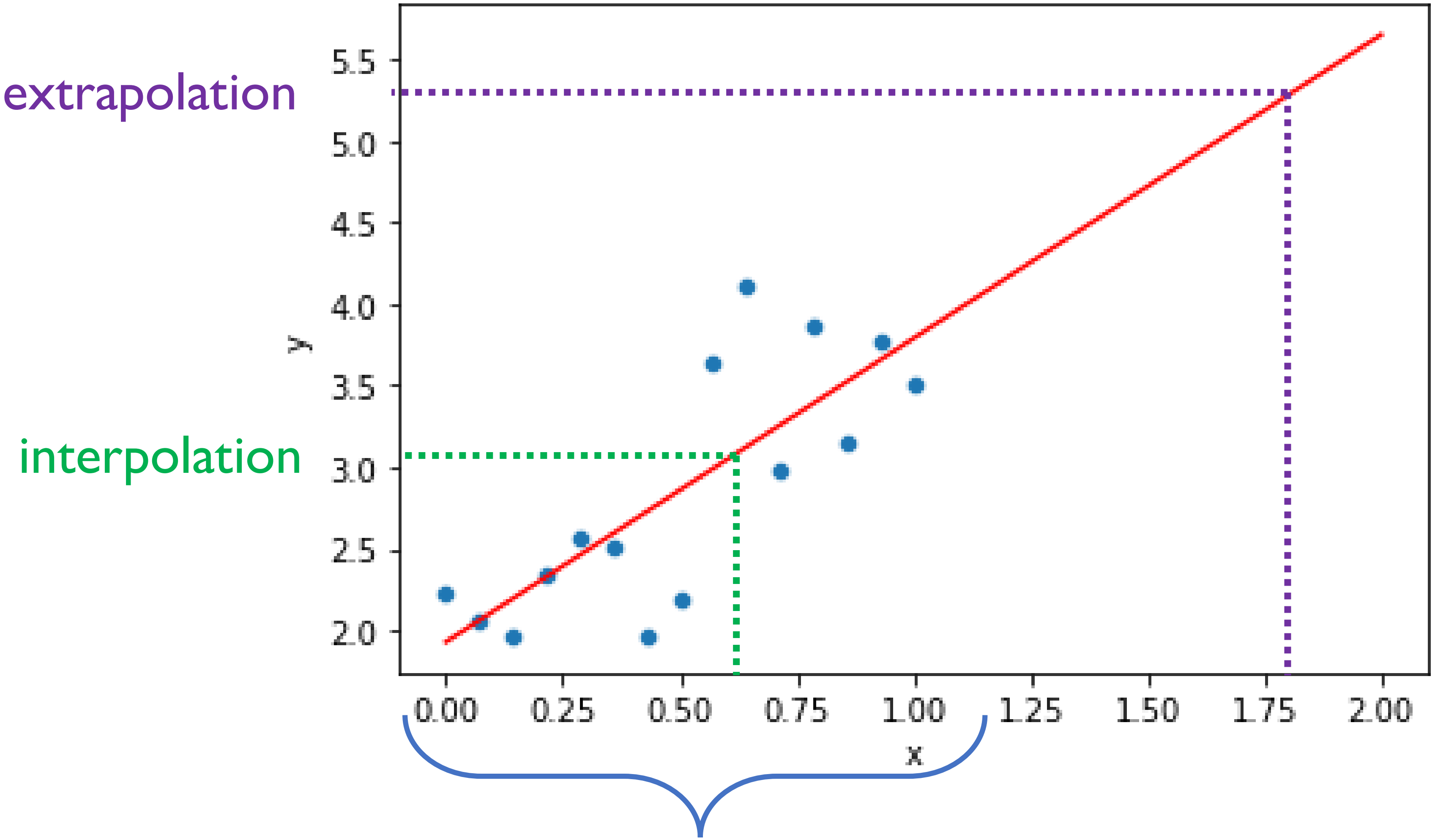
INTERPOLATION VS EXTRAPOLATION



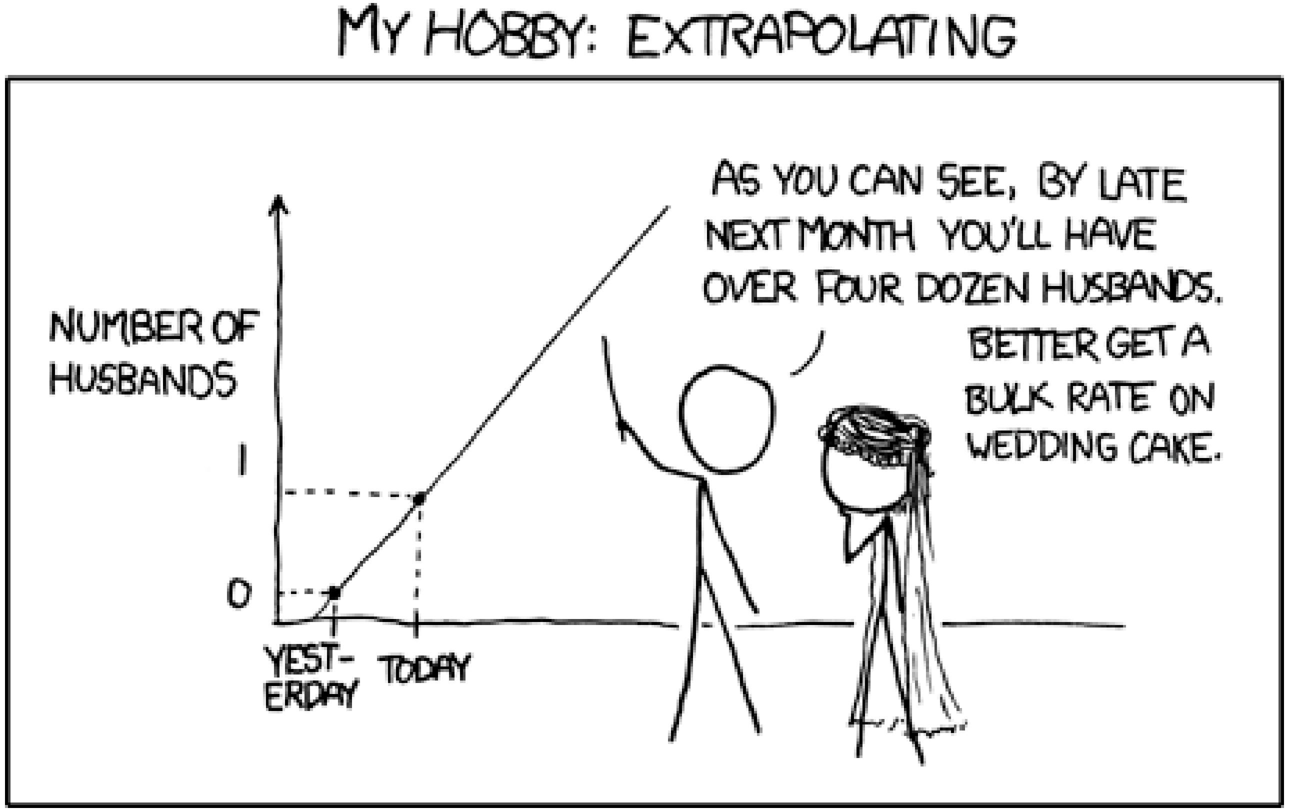
INTERPOLATION VS EXTRAPOLATION



INTERPOLATION VS EXTRAPOLATION



INTERPOLATION VS EXTRAPOLATION



MULTIPLE LINEAR REGRESSION

- **Simple linear regression**: *bivariate data*

$$y = ax + b$$

$$y_i = ax_i + b + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2)$$

MULTIPLE LINEAR REGRESSION

- **Simple linear regression**: *bivariate data*

$$y = ax + b$$

$$y_i = ax_i + b + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2)$$

- **Multiple linear regression**: *multivariate data*

$$y = a_1 x_1 + a_2 x_2 + \cdots + a_m x_m + b$$

$$y^{(i)} = a_1 x_1^{(i)} + \cdots + a_m x_m^{(i)} + b + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2)$$

RECAP

WHAT WE'VE SEEN IN THIS COURSE

- Probability theory
  - Discrete and continuous random variables
  - Expectation, (co-)variance, correlation
  - Basic distributions
    - CDFs and PDFs

WHAT WE'VE SEEN IN THIS COURSE

- Probability theory
  - Discrete and continuous random variables
  - Expectation, (co-)variance, correlation
  - Basic distributions
    - CDFs and PDFs

- Descriptive Statistics
  - Summary statistics (sample mean, sample variance, median, …)
  - Basic plots

WHAT WE'VE SEEN IN THIS COURSE

- Probability theory
  - Discrete and continuous random variables
  - Expectation, (co-)variance, correlation
  - Basic distributions
    - CDFs and PDFs

- Descriptive Statistics
  - Summary statistics (sample mean, sample variance, median, …)
  - Basic plots

- Inferential Statistics
  - Parameter estimation
    - point estimates (maximum likelihood);
    - confidence intervals.
  - Hypothesis testing

PROBABILITY THEORY

Discrete random variables

Discrete random variables can take only *countably* many values.

SOME DISCRETE DISTRIBUTIONS

- **Bernoulli**

X ~ Bernoulli(p) $\mathrm{P}(X = 1) = p, \quad P(X = 0) = 1 - p$

E(X) = p

- **Binomial**

X ~ Bi(n, p), $\mathrm{P}(X = k) = C_n^k p^k (1 - p)^{n-k}, \;\; 0 \leq k \leq 1$

E(X) = np

- **Poisson**

X ~ Po(λ), $\mathrm{P}(X = k) = \frac{\mathrm{e}^{-\lambda} \cdot \lambda^{\mathrm{k}}}{\mathrm{k}!}, \;\; k \geq 0$

E(X) = λ

SOME DISCRETE DISTRIBUTIONS

- **Bernoulli**

X ~ Bernoulli(p) $\mathrm{P}(X = 1) = p, \quad P(X = 0) = 1 - p$

E(X) = p Chance of success in a single trial with two outcomes

- **Binomial**

X ~ Bi(n, p), $\mathrm{P}(X = k) = C_n^k p^k (1 - p)^{n-k}, \;\; 0 \leq k \leq 1$

E(X) = np

- **Poisson**

X ~ Po(λ), $\mathrm{P}(X = k) = \frac{\mathrm{e}^{-\lambda} \cdot \lambda^{\mathrm{k}}}{\mathrm{k}!}, \;\; k \geq 0$

E(X) = λ

SOME DISCRETE DISTRIBUTIONS

- **Bernoulli**

X ~ Bernoulli(p) $\mathrm{P}(X = 1) = p, \quad P(X = 0) = 1 - p$

E(X) = p Chance of success in a single trial with two outcomes

- **Binomial**

X ~ Bi(n, p), $\mathrm{P}(X = k) = C_n^k p^k (1 - p)^{n-k}, \;\; 0 \leq k \leq 1$

E(X) = np # of successes in a series of n Bernoulli trials

- **Poisson**

X ~ Po(λ), $\mathrm{P}(X = k) = \frac{\mathrm{e}^{-\lambda} \cdot \lambda^{\mathrm{k}}}{\mathrm{k}!}, \;\; k \geq 0$

E(X) = λ

SOME DISCRETE DISTRIBUTIONS

- **Bernoulli**

X ~ Bernoulli(p) $\mathrm{P}(X=1)=p, \quad P(X=0)=1-p$

E(X) = p Chance of success in a single trial with two outcomes

- **Binomial**

X ~ Bi(n, p), $\mathrm{P}(X=k)=C_n^k p^k(1-p)^{n-k}, \;\; 0 \leq k \leq 1$

E(X) = np # of successes in a series of n Bernoulli trials

- **Poisson**

X ~ Po(λ), $\mathrm{P}(X=k)=\frac{\mathrm{e}^{-\lambda} \cdot \lambda^{\mathrm{k}}}{\mathrm{k}!}, \;\; k \geq 0$

E(X) = λ # events that occur within a fixed amount of time

Continuous random variables can take *uncountably* many values.

CDF & PDF

- Probability mass function *(discrete random variables):*

$$P(X = x)$$

- Cumulative distribution function (CDF):

$$F(x) = P(X \leq x)$$

- Probability density function (PDF) *(continuous random variables):*

$$F(x) = \int_{-\infty}^{x} p(t)dt$$

The probability that a continuous random variable takes a particular value is…

The probability that a continuous random variable takes a particular value is 0!

CDF

- Consider random variable X:

x	**1**	**2**	**3**
$P(X = x)$	*0.25*	*0.5*	*0.25*

- What's the CDF of X?

$$F(x) = P(X \leq x) =$$

CDF

- Consider random variable X:

x	**1**	**2**	**3**
$P(X = x)$	*0.25*	*0.5*	*0.25*

- What's the CDF of X?

$$F(x) = P(X \leq x) = \begin{cases} 0, & x < 1 \\ & \\ & \end{cases}$$

CDF

- Consider random variable X:

x	**1**	**2**	**3**
$P(X = x)$	*0.25*	*0.5*	*0.25*

- What's the CDF of X?

$$F(x) = P(X \leq x) = \begin{cases} 0, & x < 1 \\ 0.25, & 1 \leq x < 2 \\ & \\ & \end{cases}$$

CDF

- Consider random variable X:

x	**1**	**2**	**3**
$P(X = x)$	*0.25*	*0.5*	*0.25*

- What's the CDF of X?

$$F(x) = P(X \leq x) = \begin{cases} 0, & x < 1 \\ 0.25, & 1 \leq x < 2 \\ 0.75, & 2 \leq x < 3 \\ & \end{cases}$$

CDF

- Consider random variable X:

x	**1**	**2**	**3**
$P(X = x)$	*0.25*	*0.5*	*0.25*

- What's the CDF of X?

$$F(x) = P(X \leq x) = \begin{cases} 0, & x < 1 \\ 0.25, & 1 \leq x < 2 \\ 0.75, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$

WHAT IS DIFFERENT FROM THE REST

1. $P(X \leq 3)$

2. $\int_{3}^{+\infty} p(x)dx$

3. $F(3)$

4. $\int_{-\infty}^{3} p(x)dx$

WHAT IS DIFFERENT FROM THE REST

1. $P(X \leq 3)$

2. $\boldsymbol{\int_3^{+\infty} p(x)dx = P(X \geq 3)}$

3. $F(3) = P(X < 3)$

4. $\int_{-\infty}^{3} p(x)dx = F(3) = P(X < 3)$

WHAT IS DIFFERENT FROM THE REST

1. $P(X > 5)$

2. $\int_{5}^{+\infty} p(x)dx$

3. $\int_{-\infty}^{5} p(x)dx$

4. $1 - F(5)$

WHAT IS DIFFERENT FROM THE REST

1. $P(X > 5)$

2. $\int_{5}^{+\infty} p(x)dx = P(X \geq 5)$

3. $\boldsymbol{\int_{-\infty}^{5} p(x)dx = F(5) = P(X < 5)}$

4. $1 - F(5) = P(X \geq 5)$

WHAT IS DIFFERENT FROM THE REST

1. $\int_3^5 p(x)dx$

2. $P(3 < X \leq 5)$

3. $F(3) - F(5)$

4. $\int_{-\infty}^5 p(x)dx - \int_{-\infty}^3 p(x)dx$

WHAT IS DIFFERENT FROM THE REST

1. $\int_3^5 p(x)dx = F(5) - F(3) = P(3 < X \leq 5)$

2. $P(3 < X \leq 5)$

3. $\boldsymbol{F(3) - F(5) = -P}(3 < X \leq 5)$

4. $\int_{-\infty}^5 p(x)dx - \int_{-\infty}^3 p(x)dx = F(5) - F(3) = P(3 < X \leq 5)$

EXPECTED VALUE

DISCRETE RANDOM VARIABLE

- Sum up all the values a random variable can take, multiplying them by their probabilities:

$$E(X) = \sum_{X_i} X_i \cdot P(X = X_i)$$

CONTINUOUS RANDOM VARIABLE

- Same principle:

$$\boxed{\boldsymbol{E(X) = \int_{-\infty}^{+\infty} x \cdot p(x) dx}}$$

EXPECTED VALUE

DISCRETE RANDOM VARIABLE

- Sum up all the values a random variable can take, multiplying them by their probabilities:

$$E(X) = \sum_{X_i} X_i \cdot P(X = X_i)$$

$$E(f(X)) = \sum_{X_i} f(X_i) \cdot P(X = X_i)$$

CONTINUOUS RANDOM VARIABLE

- Same principle:

$$\boldsymbol{E(X) = \int_{-\infty}^{+\infty} x \cdot p(x)dx}$$

$$\boldsymbol{E\big(f(X)\big) = \int_{-\infty}^{+\infty} f(x) \cdot p(x)dx}$$

EXPECTED VALUE

- Consider a random variable X:

x	**1**	**2**	**3**
$P(X = x)$	*0.25*	*0.5*	*0.25*

- $E(X) =$

- $E(X^2) =$

EXPECTED VALUE

- Consider a random variable X:

x	**1**	**2**	**3**
$P(X = x)$	*0.25*	*0.5*	*0.25*

- $E(X) = \frac{1}{4} + \frac{2}{2} + \frac{3}{4} = 2$

- $E(X^2) =$

EXPECTED VALUE

- Consider a random variable X:

x	**1**	**2**	**3**
$P(X = x)$	*0.25*	*0.5*	*0.25*

- $E(X) = \frac{1}{4} + \frac{2}{2} + \frac{3}{4} = 2$

- $E(X^2) = \frac{1^2}{4} + \frac{2^2}{2} + \frac{3^2}{4} = 4.5$

EXPECTED VALUE

$$X \sim U(1,2)$$

$$E(X) =$$

$$E\left(\frac{1}{X}\right) =$$

EXPECTED VALUE

$$X \sim U(1,2)$$

$$E(X) = \frac{2-1}{2} = \frac{3}{2}$$

$$E\left(\frac{1}{X}\right) =$$

EXPECTED VALUE

$$X \sim U(1,2)$$

$$E(X) = \frac{2-1}{2} = \frac{3}{2}$$

$$E\left(\frac{1}{X}\right) = \int_{1}^{2} \frac{1}{x} \cdot 1 dx =$$

EXPECTED VALUE

$$X \sim U(1,2)$$

$$E(X) = \frac{2-1}{2} = \frac{3}{2}$$

$$E\left(\frac{1}{X}\right) = \int_{1}^{2} \frac{1}{x} \cdot 1 dx = \log 2 - \log 1 = \log 2$$

VARIANCE

DISCRETE RANDOM VARIABLE **CONTINUOUS RANDOM VARIABLE**

Expected squared distance between a value and the mean:

$$Var(X) = E\left(\left(X - E(X)\right)^2\right) = E(X^2) - \left(E(X)\right)^2$$

LINEAR COMBINATION OF NORMALLY DISTRIBUTED VARIABLES

- A linear combination of independent random variables having a normal distribution also has a normal distribution:

$$X_1, X_2, \ldots, X_n \text{ - independent}$$

$$X_i \sim N\left(\mu_i, \sigma_i^2\right)$$

$$Y = a_1X_1 + a_2X_2 + \cdots + a_3X_3 \Rightarrow$$

$$Y \sim N(\mu_Y, \sigma_Y^2)$$

$$\mu_Y = \qquad\qquad , \qquad \sigma_Y^2 =$$

LINEAR COMBINATION OF NORMALLY DISTRIBUTED VARIABLES

- A linear combination of independent random variables having a normal distribution also has a normal distribution:

$$X_1, X_2, \dots, X_n \text{ - independent}$$

$$X_i \sim N\left(\mu_i, \sigma_i^2\right)$$

$$Y = a_1 X_1 + a_2 X_2 + \cdots + a_3 X_3 \Rightarrow$$

$$Y \sim N(\mu_Y, \sigma_Y^2)$$

$$\mu_Y = a_1 \mu_1 + a_2 \mu_2 + \cdots + a_n \mu_n, \qquad \sigma_Y^2 =$$

LINEAR COMBINATION OF NORMALLY DISTRIBUTED VARIABLES

- A linear combination of independent random variables having a normal distribution also has a normal distribution:

$$X_1, X_2, \ldots, X_n \text{ - independent}$$

$$X_i \sim N\left(\mu_i, \sigma_i^2\right)$$

$$Y = a_1 X_1 + a_2 X_2 + \cdots + a_3 X_3 \Rightarrow$$

$$Y \sim N(\mu_Y, \sigma_Y^2)$$

$$\mu_Y = a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n, \qquad \sigma_Y^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2$$

CENTRAL LIMIT THEOREM

Samples $X_1, X_2, \dots, X_n$:

- i.i.d.
- a *finite* mean μ and *finite* variance σ^2

Let

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Then

$$\bar{X}_n \approx$$

CENTRAL LIMIT THEOREM

Samples $X_1, X_2, \dots, X_n$:

- i.i.d.
- a *finite* mean μ and *finite* variance σ^2

Let

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Then

$$\bar{X}_n \approx N\left(\qquad \right)$$

CENTRAL LIMIT THEOREM

Samples $X_1, X_2, \ldots, X_n$:

- i.i.d.
- a *finite* mean μ and *finite* variance σ^2

Let

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Then

$$\bar{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

COVARIANCE AND CORRELATION

- **Covariance:**

$$\sigma_{XY}^2 = E[(X - \bar{X})(Y - \bar{Y})] = E(XY) - \bar{X}\bar{Y}$$

COVARIANCE AND CORRELATION

- **Covariance:**

$$\sigma_{XY}^2 = E[(X - \bar{X})(Y - \bar{Y})] = E(XY) - \bar{X}\bar{Y}$$

- **Correlation:**

$$\rho = \frac{\sigma_{XY}^2}{\sigma_X \sigma_Y} = \frac{E(X - \bar{X})(Y - \bar{Y})}{\sqrt{E(X - \bar{X})^2 E(Y - \bar{Y})^2}}$$

INFERENTIAL STATISTICS

TWO TYPES OF STATISTICS

Sample: $X_1, X_2, \ldots, X_n$

TWO TYPES OF STATISTICS

Sample: $X_1, X_2, \ldots, X_n$

Descriptive statistics: *describe your sample*

TWO TYPES OF STATISTICS

Sample: $X_1, X_2, \dots, X_n$

Descriptive statistics: *describe your sample*

$\bar{X} = \frac{1}{n}\sum X_i$ −sample mean

TWO TYPES OF STATISTICS

Sample: $X_1, X_2, \dots, X_n$

Descriptive statistics: *describe your sample*

$\bar{X} = \frac{1}{n} \sum X_i$ −sample mean

$s^2 = \frac{1}{n} \sum X_i^2 - \bar{X}^2$ − sample variance

TWO TYPES OF STATISTICS

Sample: $X_1, X_2, \ldots, X_n$

TWO TYPES OF STATISTICS

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

Sample: $X_1, X_2, \ldots, X_n$

Inferential Statistics:

reason about the values of the unknown parameters of the underlying distribution

TWO TYPES OF STATISTICS

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

Sample: $X_1, X_2, \dots, X_n$

Inferential Statistics:

reason about the values of the unknown parameters of the underlying distribution

$\hat{\mu}, \hat{\sigma}^2$ – point estimates

TWO TYPES OF STATISTICS

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

Sample: $X_1, X_2, \dots, X_n$

Inferential Statistics:

reason about the values of the unknown parameters of the underlying distribution

$\hat{\mu}, \hat{\sigma}^2$ − point estimates

$\mu \in \bar{X} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ − confidence intervals

TWO TYPES OF STATISTICS

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

Sample: $X_1, X_2, \ldots, X_n$

Inferential Statistics:

reason about the values of the unknown parameters of the underlying distribution

$\hat{\mu}, \hat{\sigma}^2$ – point estimates

$\mu \in \bar{X} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ – confidence intervals

$H_0: \mu = 5,\ H_1: \mu \neq 5$ – hypothesis testing

MAXIMUM LIKELIHOOD ESTIMATE

1. Write down the likelihood function:

Discrete: $L(\theta) = \prod_{i=1}^{n} P(X{=}X_i \mid \theta)$ **Continuous:** $L(\theta) = \prod_{i=1}^{n} p(X_i \mid \theta)$

2. Find its maximum w.r.t. the unknown parameter θ:

$$\widehat{\Theta} = \operatorname{argmax} L(\theta) \quad \text{w.r.t.} \quad \theta$$

*(!) In many cases, it's easier to maximize **log-likelihood**:*

Discrete: $\log L(\theta) = \sum_{i=1}^{n} \log P(X{=}X_i \mid \theta)$

Continuous: $\log L(\theta) = \sum_{i=1}^{n} \log p(X_i \mid \theta)$

$$\widehat{\Theta} = \operatorname{argmax} \log L(\theta)$$

MLE EXAMPLE

- Simple linear regression:

$$y_i = ax_i + b + \varepsilon_i, \qquad \varepsilon_i \sim N(o, \sigma^2)$$

- We obtain model parameters by least squares:

$$a, b\text{:} \;\; \sum (y_i - ax_i - b)^2 \rightarrow min$$

MLE EXAMPLE

- Simple linear regression:

$$y_i = ax_i + b + \varepsilon_i, \qquad \varepsilon_i \sim N(o, \sigma^2)$$

- We obtain model parameters by least squares:

$$a, b: \ \sum (y_i - ax_i - b)^2 \to min$$

Show that it's ML estimate of the parameters.

MLE EXAMPLE

$$y_i = ax_i + b + \varepsilon_i, \qquad \varepsilon_i \sim N(o, \sigma^2)$$

$$y_i \sim$$

MLE EXAMPLE

$$y_i = ax_i + b + \varepsilon_i, \qquad \varepsilon_i \sim N(o, \sigma^2)$$

$$y_i \sim N(\qquad\qquad\qquad)$$

MLE EXAMPLE

$$y_i = ax_i + b + \varepsilon_i, \qquad \varepsilon_i \sim N(o, \sigma^2)$$

$$y_i \sim N(ax_i + b, \sigma)$$

MLE EXAMPLE

$$y_i = ax_i + b + \varepsilon_i, \qquad \varepsilon_i \sim N(o, \sigma^2)$$

$$y_i \sim N(ax_i + b, \sigma)$$

$$L(a, b) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y_i - ax_i - b}{\sigma}\right)^2} \to max$$

MLE EXAMPLE

$$y_i = ax_i + b + \varepsilon_i, \qquad \varepsilon_i \sim N(o, \sigma^2)$$

$$y_i \sim N(ax_i + b, \sigma)$$

$$L(a, b) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y_i - ax_i - b}{\sigma}\right)^2} \rightarrow max$$

$$\Leftrightarrow (y_i - ax_i - b)^2 \rightarrow min$$

PROPERTIES OF ESTIMATORS

- **BIAS**

$$bias = E\big(T(X)\big) - \theta$$

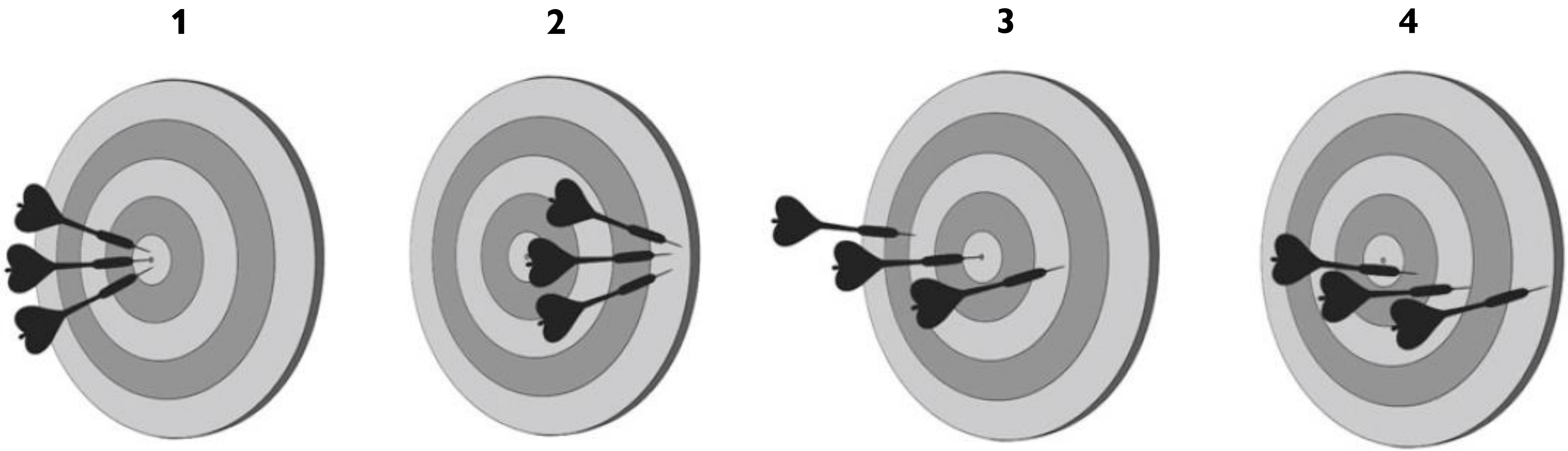
- **VARIANCE**

$$Var\big(T(X)\big)$$

- **CONSISTENCY**

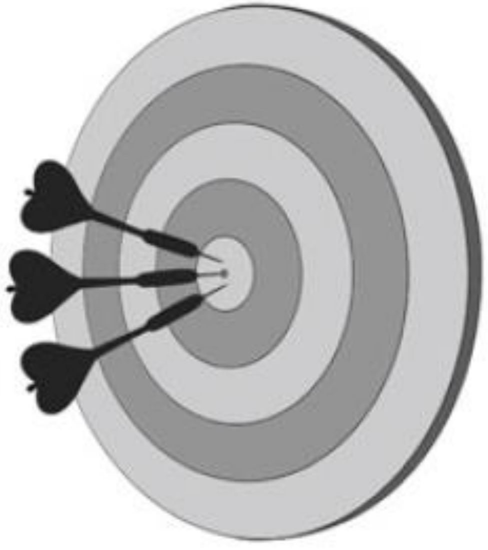
"The more data we have, the closer the estimate is to the true value of the parameter"

BIAS VS VARIANCE

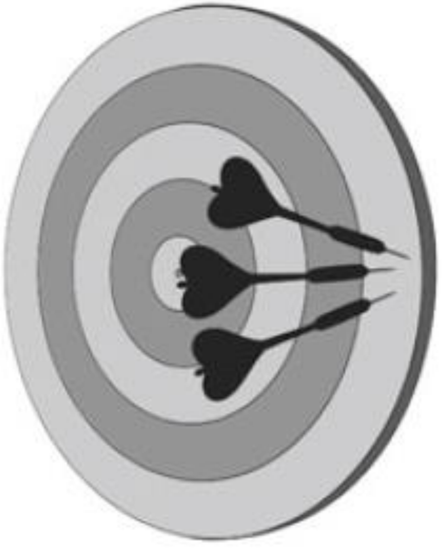


BIAS VS VARIANCE

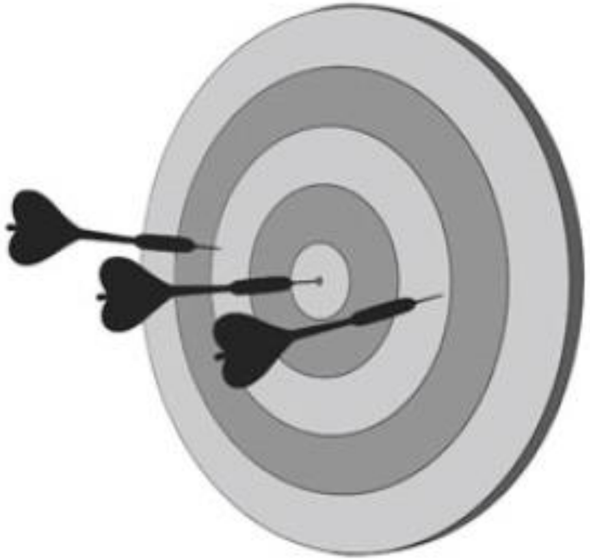
1



2



3



LOW BIAS
HIGH VARIANCE

4

CI: DEFINITION

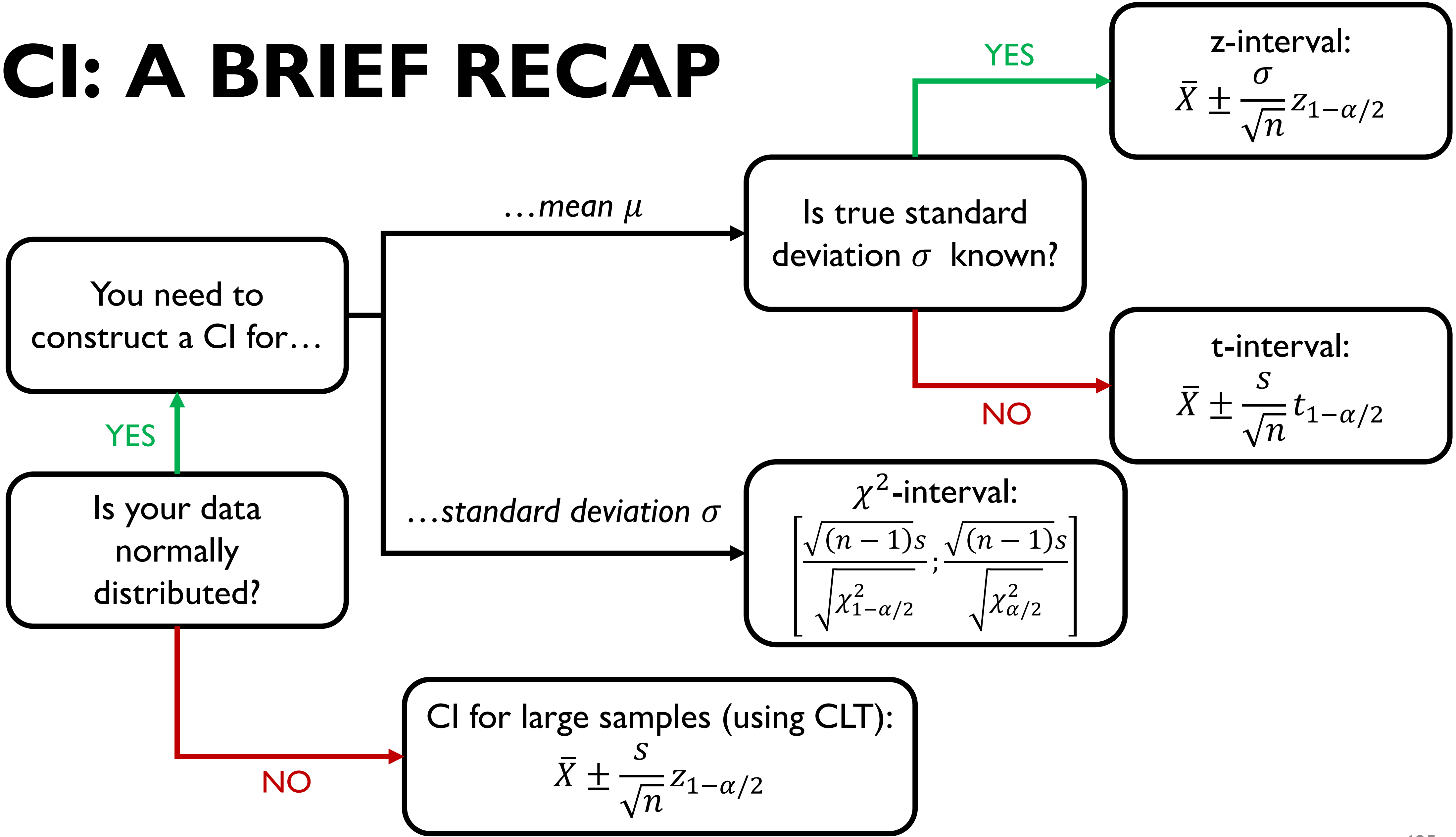
A $1-\alpha$ confidence interval for a parameter θ is an interval $C_n = (T_1, T_2)$ such that $\mathrm{T}_1 = t_1(X_1, \ldots, X_n),\ T_2 = t_2(X_1, \ldots, X_n)$ and

$$P(T_1 < \theta < T_2) \geq 1-\alpha$$

- **Random** intervals: T_1 and T_2 are functions of random samples.

- θ is unknown, but fixed
  T_1 and T_2 are random

CI: A BRIEF RECAP



EXAMPLE (ASSIGNMNET 4)

Suppose that the number of points a basketball team scores against a certain opponent is normally distributed with unknown mean μ and unknown variance σ^2.

Compute a 95% interval for μ

Now suppose that you learn that σ^2 = 25. Compute a 95% interval for μ.

EXAMPLE (ASSIGNMNET 4)

Suppose that the number of points a basketball team scores against a certain opponent is normally distributed with unknown mean μ and unknown variance σ^2.

Compute a 95% interval for μ

σ unknown -> t-interval

Now suppose that you learn that $\sigma^2 = 25$. Compute a 95% interval for μ.

σ known -> z-interval

EXAMPLE (ASSIGNMNET 4)

Suppose that the number of points a basketball team scores against a certain opponent is normally distributed with unknown mean μ and unknown variance σ^2.

Compute a 95% interval for μ

$$\sigma \text{ unknown -> t-interval: } \mu = \bar{X} + t_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Now suppose that you learn that σ^2 = 25. Compute a 95% interval for μ.

$$\sigma \text{ known -> z-interval: } \mu = \bar{X} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

CI: A BRIEF RECAP

A machine fills in wine bottles with a random amount of wine that follows normal distribution with unknown mean μ and unknown standard deviation σ.

You check the last 100 bottles and see that on average they were filled with $\bar{X} = 705$ ml of wine, with sample std $s = 3$ ml.

Which interval would you use to construct a 95%-CI for μ?

CI: A BRIEF RECAP

In a study on cholesterol levels a sample of 1000 patients was chosen.

The average plasma cholesterol levels subjects was $\bar{X} = 6$ mmol/L, with sample std $s = 0.4$ mmol/L.

Which interval would you use
to construct a 95%-CI for the true mean?

CI: A BRIEF RECAP

Height of a female student is a random variable following normal distribution with unknown mean μ and standard deviation $\sigma = 5$ cm.

The average height of 100 female students $\bar{X} = 165$ cm, and you sample std $s = 4$ cm

Which interval would you use to construct a 95%-CI for μ?

CI: A BRIEF RECAP

On a candy factory, a machine fills packs with random number of candies which follows normal distribution with unknown mean μ and unknown standard deviation σ.

In the last 50 packs, there were on average $\bar{X} = 90$ g of sweets, with sample std $s = 7$ g.

Which interval would you use to construct a 95%-CI for σ?

SAMPLE SIZE DETERMINATION

- Data collection is difficult.
- How much is 'just enough'?

- Example: estimating CI for the mean, σ is know.

$$\mu \in \bar{X} \pm \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}$$

Limit the width of the interval: $\frac{\sigma}{\sqrt{n}} z_{1-\alpha/2} \leq \epsilon$

$$n \geq \frac{\sigma^2 z_{1-\alpha/2}^2}{\epsilon^2}$$

HYPOTHESIS TESTING STEP-BY-STEP

- Collect data X
- Set up H_0 and H_1
  - two-sided or one-sided
- Chose test statistic $T(X)$
- Determine distribution of T assuming H_0
- Determine rejection area R
- Compute the value t of $T(X)$ from data
- Check if it falls into the rejection area R
  - YES $\Longrightarrow$ reject H_0
  - NO $\Longrightarrow$ don't reject H_0

You are checking a hypothesis H_0 against a two-sided alternative H_1 at the level of significance $\alpha = 0.01$.

After running a statistical test, you obtain a p-value of 0.1.

What is your conclusion?

You are checking a hypothesis H_0 against a two-sided alternative H_1 at the level of significance $\alpha = 0.05$.

Test statistic equals 0.5, and the $\alpha/2$ and $1 - \alpha/2$ – quantiles of the corresponding distribution are ± 1.96

What is your conclusion?

You are checking a hypothesis H_0 against a two-sided alternative H_1 at the level of significance $\alpha = 0.05$.

After running a statistical test, you obtain a p-value of 0.001.

What is your conclusion?

You are checking a hypothesis H_0 against a two-sided alternative H_1 at the level of significance $\alpha = 0.05$.

Test statistic equals 14.5, and the $\alpha/2$ and $1 - \alpha/2$ – quantiles of the corresponding distribution are ± 1.96

What is your conclusion?