# Statistics Review I

## EC 320: Introduction to Econometrics

Amna Javed
Winter 2020

# Prologue

# Housekeeping

I'll post Problem Set 1 by Friday.

- Due on January 17th (next Friday) on Canvas.
- Includes both analytical and computational components.

Issues with R?

- Come to office hours.

# Motivation

The focus of our course is **regression analysis**, a useful toolkit for learning from data.

To understand regression, its mechanics, and its pitfalls, **we need to understand the underlying statistical theory.**

- Insights from theory can help us become better practitioners and savvier consumers of science.

Today, we will review important concepts you learned in Math 243.

- Maybe some you missed, too.

# A Brief Math Review

# Notation

**Data** on a variable $X$ **are*** a sequence of $n$ observations, indexed by $i$ :

$$\{x_i : 1, \ldots, n\}.$$

Example: $n = 5$

| $i$ | $x_i$ |
|---|---|
| 1 | 4 |
| 2 | 2 |
| 3 | 8 |
| 4 | 10 |
| 5 | 6 |

- $i$ indicates the row number.
- $n$ is the number of rows.
- $x_i$ is the value of $X$ for row $i$ .

* *Data* = **plural** of *datum*.

# Summation

The **summation operator** adds a sequence of numbers over an index:

$$\sum_{i=1}^{n} x_i \equiv x_1 + x_2 + \cdots + x_n.$$

- "The sum of $x_i$ from 1 to $n$."

Example: $n = 4$

| $\boldsymbol{i}$ | $\boldsymbol{x_i}$ |
|---|---|
| 1 | 7 |
| 2 | 4 |
| 3 | 10 |
| 4 | 2 |

$$\begin{aligned}\sum_{i=1}^{4} x_i &= 7 + 4 + 10 + 2 \\ &= 23\end{aligned}$$

## Rule 1

For any constant $c$ ,

$$\sum_{i=1}^{n} c = nc.$$

Example: $n = 4$

| $\boldsymbol{i}$ | $\boldsymbol{c}$ |
|---|---|
| 1 | 2 |
| 2 | 2 |
| 3 | 2 |
| 4 | 2 |

$$\sum_{i=1}^{4} 2 = 4 \times 2$$
$$= 8$$

## Rule 2

For any constant $c$ ,

$$\sum_{i=1}^{n} cx_i = c\sum_{i=1}^{n} x_i.$$

Example: $n = 3$

| $i$ | $c$ | $x_i$ |
|---|---|---|
| 1 | 2 | 7 |
| 2 | 2 | 4 |
| 3 | 2 | 10 |

$$\begin{aligned}\sum_{i=1}^{3} 2x_i &= 2 \times 7 + 2 \times 4 + 2 \times 10 \\ &= 14 + 8 + 20 \\ &= 42\end{aligned}$$

$$\begin{aligned}2\sum_{i=1}^{3} x_i &= 2(7 + 4 + 10) \\ &= 42\end{aligned}$$

## Rule 3

If $\{(x_i, y_i) : 1, \ldots, n\}$ is a set of $n$ pairs, and $a$ and $b$ are constants, then

$$\sum_{i=1}^{n} (ax_i + by_i) = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} y_i.$$

Example: $n = 2$

| $i$ | $a$ | $x_i$ | $b$ | $y_i$ |
|---|---|---|---|---|
| 1 | 2 | 7 | 1 | 4 |
| 2 | 2 | 4 | 1 | 2 |

$$\sum_{i=1}^{2} (2x_i + y_i) = 18 + 10$$
$$= 28$$

$$2\sum_{i=1}^{2} x_i + \sum_{i=1}^{2} y_i = 2 \times 11 + 6$$
$$= 28$$

## Caution

The **sum of the ratios** **is not** the **ratio of the sums**:

$$\sum_{i=1}^{n} x_i / y_i \neq \left(\sum_{i=1}^{n} x_i\right) \Big/ \left(\sum_{i=1}^{n} y_i\right).$$

- If $n = 2$ , then $\frac{x_1}{y_1} + \frac{x_2}{y_2} \neq \frac{x_1+x_2}{y_1+y_2}$ .

The **sum of squares** **is not** the **square of the sums**:

$$\sum_{i=1}^{n} x_i^2 \neq \left(\sum_{i=1}^{n} x_i\right)^2.$$

- If $n = 2$ , then $x_1^2 + x_2^2 \neq (x_1 + x_2)^2 = x_1^2 + 2x_1x_2 + x_2^2$ .

# Probability Review

# Random Variables

**Experiment:** Any procedure that is *infinitely repeatable* and has a *well-defined set of outcomes*.

- Flip a coin 10 times and record the number of heads.
- Roll two six-sided dice and record the sum.

**Random Variable:** A variable with *numerical values determined by an experiment or a random phenomenon*.

- Describes the sample space of an experiment.
- **Sample space:** The set of potential outcomes an experiment could generate, *e.g.*, the sum of two dice is an integer from 2 to 12.
- **Event:** A subset of the sample space or a combination of outcomes, *e.g.*, rolling a two or a four.

# Random Variables

**Notation:** capital letters for random variables (*e.g.*, $X$, $Y$, or $Z$) and lowercase letters for particular outcomes (*e.g.*, $x$, $y$, or $z$).

**Example 1:** Flipping a coin.

- Two outcomes: heads or tails.
- Quantify the outcomes: Define a random variable $\text{Heads}$ such that $\text{Heads} = 1$ if heads and $\text{Heads} = 0$ if tails.

**Example 2:** Flipping a coin 10 times.

- Several outcomes: 10 heads and 0 tails, 9 heads and 1 tails, 8 heads and 2 tails, *etc*.
- The number of heads is a random variable: $\{\text{Heads}: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$.

# Discrete Random Variables

**Discrete Random Variable:** A random variable that takes a countable set of values.

A **Bernoulli** (or binary) random variable takes values of either 1 or 0.

- Characterized by $\mathrm{P}(X = 1)$ , "the probability of success."
- Probabilities sum to 1: $\mathrm{P}(X = 1) + \mathrm{P}(X = 0) = 1$ .
  - For a "fair" coin, $\mathrm{P}(\text{Heads} = 1) = \frac{1}{2} \implies \mathrm{P}(\text{Heads} = 0) = \frac{1}{2}$ .
- More generally, if $\mathrm{P}(X = 1) = \theta$ for some $\theta \in [0, 1]$ , then $\mathrm{P}(X = 0) = 1 - \theta$ .
  - If the probability of passing this class is 75%, then the probability of not passing is 25%.

## Probabilities

We describe a discrete random variable by listing its possible values with associated probabilities.

If $X$ takes on $k$ possible values $\{x_1, \ldots, x_k\}$ , then the probabilities $p_1, p_2, \ldots, p_k$ are defined by

$$p_j = \mathrm{P}(X = x_j), \quad j = 1, 2, \ldots, k,$$

where

$$p_j \in [0, 1]$$

and

$$p_1 + p_2 + \cdots + p_k = 1.$$

## Probability density function

The **probability density function** (**pdf**) of $X$ summarizes possible outcomes and associated probabilities:

$$f(x_j) = p_j, \quad j = 1, 2, \ldots, k.$$

## Example

2020 Presidential election: 538 electoral votes at stake.

- $\{X: 0, 1, \ldots, 538\}$ is the number of electoral votes won by the Democratic candidate.
- Extremely unlikely that she will win 0 votes or all 538 votes: $f(0) \approx 0$ and $f(538) \approx 0$ .
- Nonzero probability of winning an exact majority: $f(270) > 0$ .

## Example

Basketball player goes to the foul line to shoot two free throws.

- $X$ is the number of shots made (either 0, 1, or 2).
- The pdf of $X$ is $f(0) = 0.3$ , $f(1) = 0.4$ , $f(2) = 0.3$ .
- **Note:** the probabilities sum to 1.

Use the pdf to calculate the probability of the **event** that the player makes *at least one shot*, *i.e.*, $\mathrm{P}(X \geq 1)$ .

- $\mathrm{P}(X \geq 1) = \mathrm{P}(X = 1) + \mathrm{P}(X = 2) = 0.4 + 0.3 = 0.7$ .

# Continuous Random Variables

**Continuous Random Variable:** A random variable that takes any real value with *zero* probability.

- **Wait, what?!** The variable takes so many values that we can't count all possibilities, so the probability of any one particular value is zero.

Measurement is discrete (*e.g.*, dollars and cents), but variables with many possible values are best treated as continuous.

- *e.g.*, electoral votes, height, wages, temperature, *etc.*

# Continuous Random Variables

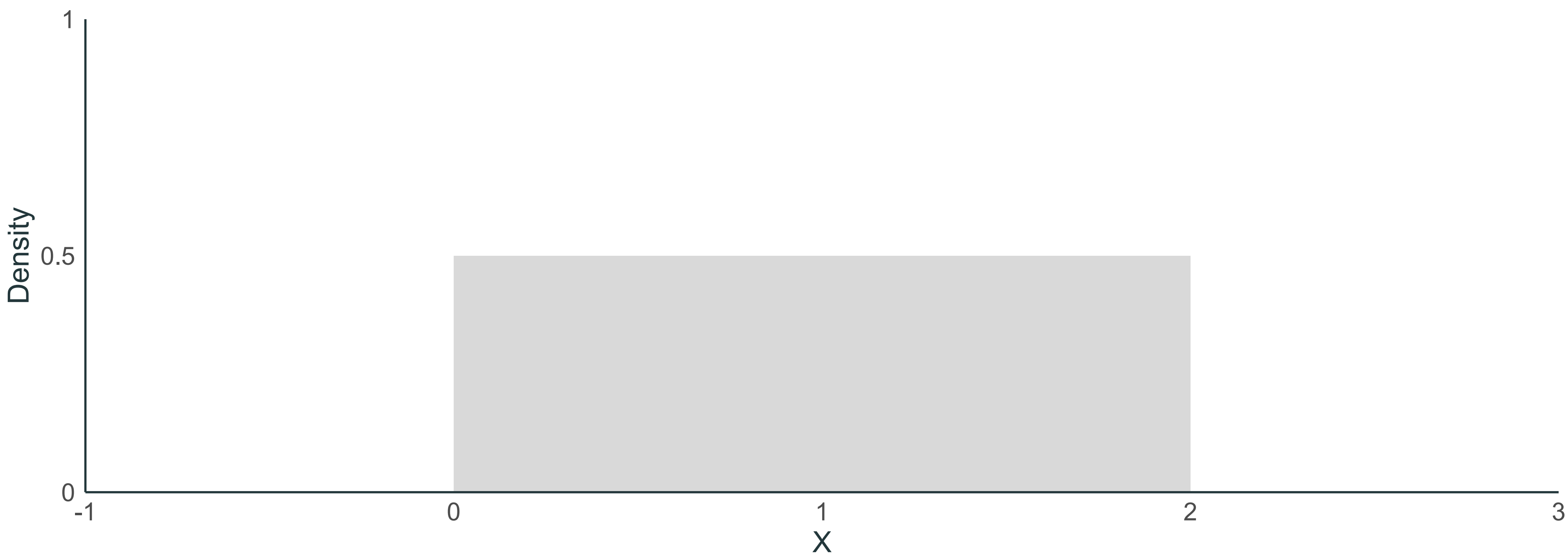
Probability density functions also describe continuous random variables.

- Difference: Interested in the probability of events within a *range* of values.
- *e.g.* What is the probability of more than 1 inch of rain tomorrow?

## Uniform Distribution

The probability density function of a variable uniformly distributed between 0 and 2 is

$$f(x) = \begin{cases} \frac{1}{2} & \text{if } 0 \leq x \leq 2 \\ 0 & \text{if } x < 0 \text{ or } x > 2 \end{cases}$$

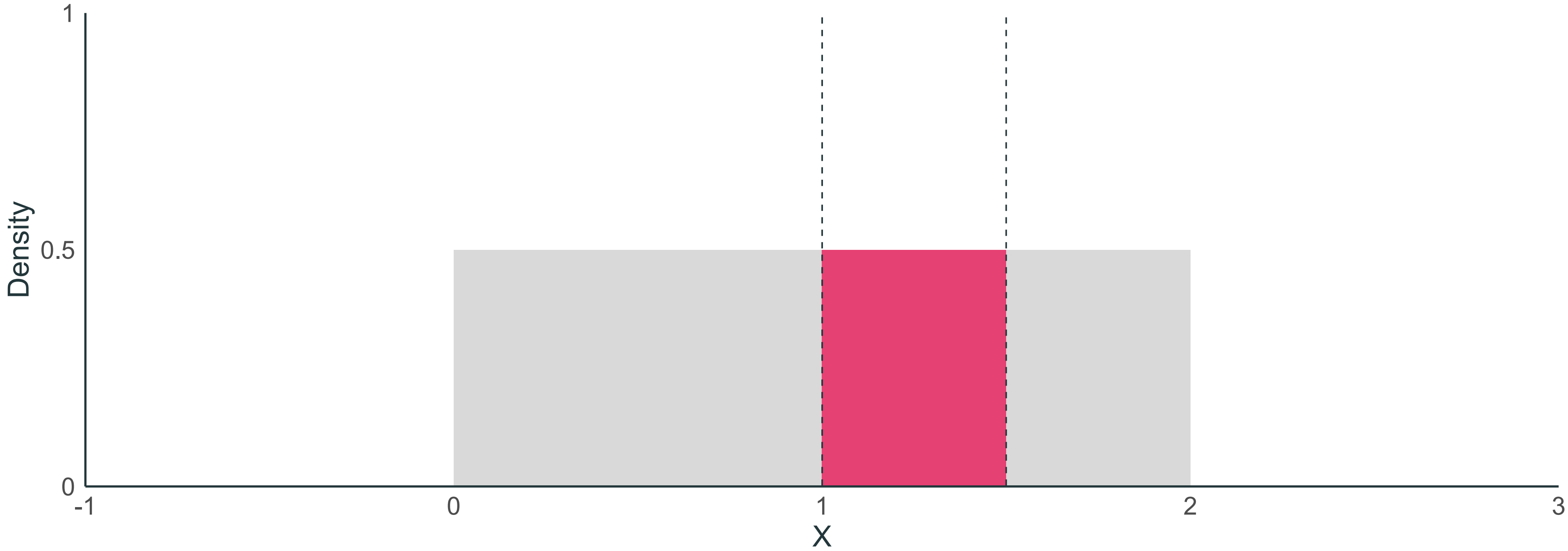


## Uniform Distribution

By definition, the area under $f(x)$ is equal to 1.

The **shaded area** illustrates the probability of the event $1 \leq X \leq 1.5$ .

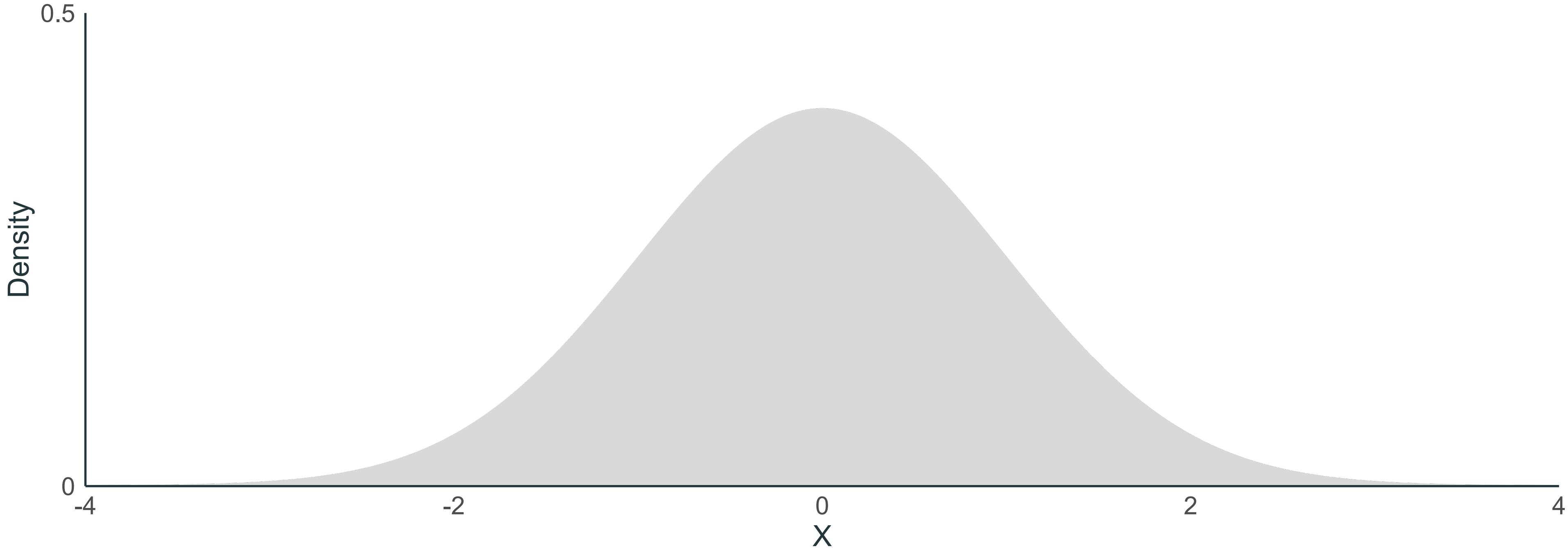
- $\mathrm{P}(1 \leq X \leq 1.5) = (1.5 - 1) \times 0.5 = 0.25$ .



## Normal Distribution

**The "bell curve."**

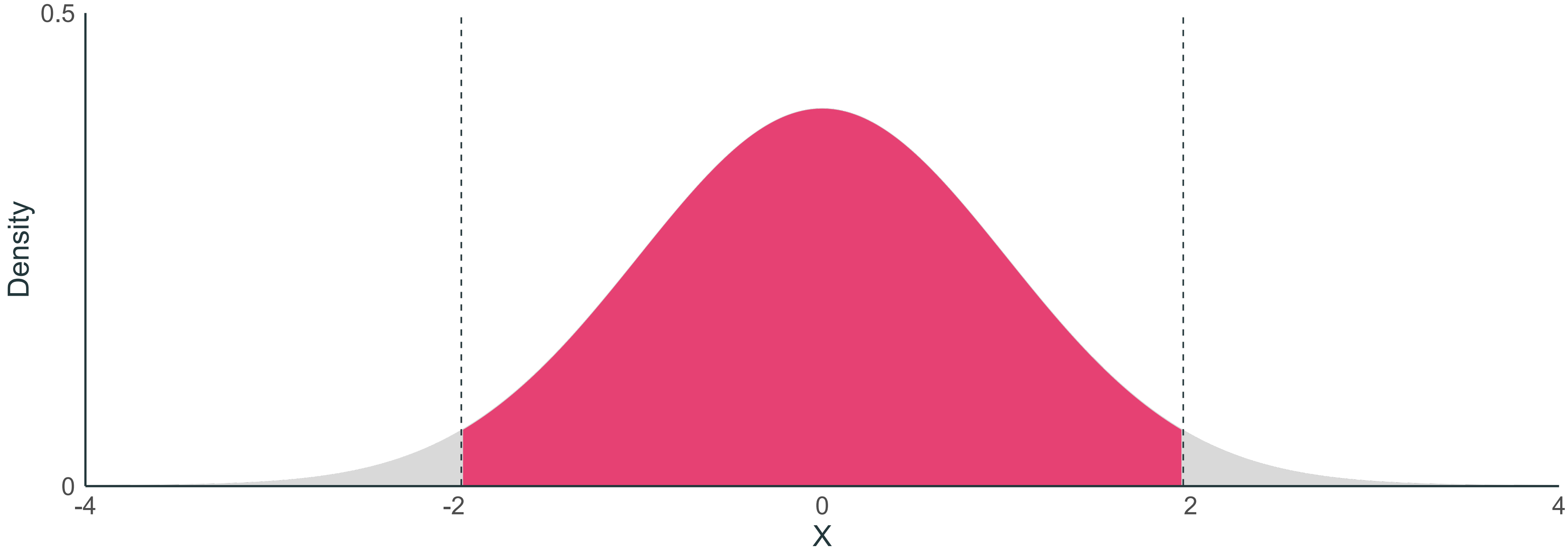
- Symmetric: mean and median occur at the same point (*i.e.*, no skew).
- Low-probability events in tails; high-probability events near center.



## Normal Distribution

The **shaded area** illustrates the probability of the event $-2 \leq X \leq 2$ .

- "Find area under curve" = use integral calculus (or, in practice, R).
- $\mathrm{P}(-2 \leq X \leq 2) \approx 0.95$ .



# Expected Value

A density function describes an entire distribution, but sometimes we just want a summary.

The **expected value** describes the *central tendency* of distribution in a single number.

- *Central tendency* = typical value.

## Definition (Discrete)

The expected value of a discrete random variable $X$ is the weighted average of its $k$ values $\{x_1, \ldots, x_k\}$ and their associated probabilities:

$$\begin{aligned} \mathrm{E}(X) &= x_1 \mathrm{P}(x_1) + x_2 \mathrm{P}(x_2) + \cdots + x_k \mathrm{P}(x_k) \\ &= \sum_{j=1}^{k} x_j \mathrm{P}(x_j). \end{aligned}$$

- Also known as the **population mean**.

## Example

Rolling a six-sided die once can take values $\{1, 2, 3, 4, 5, 6\}$ , each with equal probability.
**What is the expected value of a roll?**

$$\mathrm{E}(\mathrm{Roll}) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3.5 .$$

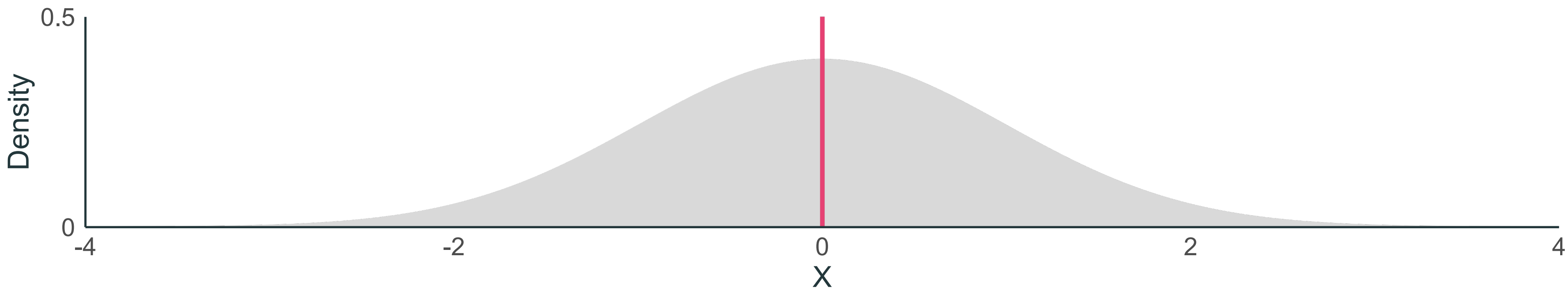
- **Note:** The expected value can be a number that isn't a possible outcome of $X$ .

## Definition (Continuous)

If $X$ is a continuous random variable and $f(x)$ is its probability density function, then the expected value of $X$ is

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

- **Note:** $x$ represents the particular values of $X$ .
- Same idea as the discrete definition: describes the **population mean**.



# Expected Value

## Rule 1

For any constant $c$ , $\mathrm{E}(c) = c$ .

## Not-so-exciting examples

$\mathrm{E}(5) = 5$ .

$\mathrm{E}(1) = 1$ .

$\mathrm{E}(4700) = 4700$ .

## Rule 2

For any constants $a$ and $b$, $\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$.

## Example

Suppose $X$ is the high temperature in degrees Celsius in Eugene during August. The long-run average is $\mathrm{E}(X) = 28$. If $Y$ is the temperature in degrees Fahrenheit, then $Y = 32 + \frac{9}{5}X$.

**What is $\mathrm{E}(Y)$ ?**

- $\mathrm{E}(Y) = 32 + \frac{9}{5}\mathrm{E}(X) = 32 + \frac{9}{5} \times 28 = 82.4$.

## Rule 3

If $\{a_1, a_2, \ldots, a_n\}$ are constants and $\{X_1, X_2, \ldots, X_n\}$ are random variables, then

$$\mathrm{E}(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1\mathrm{E}(X_1) + a_2\mathrm{E}(X_2) + \cdots + a_n\mathrm{E}(X_n).$$

In English, **the expected value of the sum** = **the sum of expected values**.

## Rule 3

**The expected value of the sum** = **the sum of expected values**.

## Example

Suppose that a coffee shop sells $X_1$ small, $X_2$ medium, and $X_3$ large caffeinated beverages in a day. The quantities sold are random with expected values $\mathrm{E}(X_1) = 43$ , $\mathrm{E}(X_2) = 56$ , and $\mathrm{E}(X_3) = 21$ . The prices of small, medium, and large beverages are $1.75$ , $2.50$ , and $3.25$ dollars. **What is expected revenue?**

$$\begin{aligned}
\mathrm{E}(1.75X_1 + 2.50X_2 + 3.35X_n) &= 1.75\mathrm{E}(X_1) + 2.50\mathrm{E}(X_2) + 3.25\mathrm{E}(X_3) \\
&= 1.75(43) + 2.50(56) + 3.25(21) \\
&= 283.5
\end{aligned}$$

## Caution

Previously, we found that the expected value of rolling a six-sided die is $\mathrm{E}(\text{Roll}) = 3.5$ .

- If we square this number, we get $[\mathrm{E}(\text{Roll})]^2 = 12.25$ .

**Is** $[\mathrm{E}(\text{Roll})]^2$ **the same as** $\mathrm{E}\left(\text{Roll}^2\right)$ **?**

**No!**

$$\begin{aligned}\mathrm{E}\left(\text{Roll}^2\right) &= 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + 3^2 \times \frac{1}{6} + 4^2 \times \frac{1}{6} + 5^2 \times \frac{1}{6} + 6^2 \times \frac{1}{6} \\ &\approx 15.167 \\ &\neq 12.25.\end{aligned}$$

## Caution

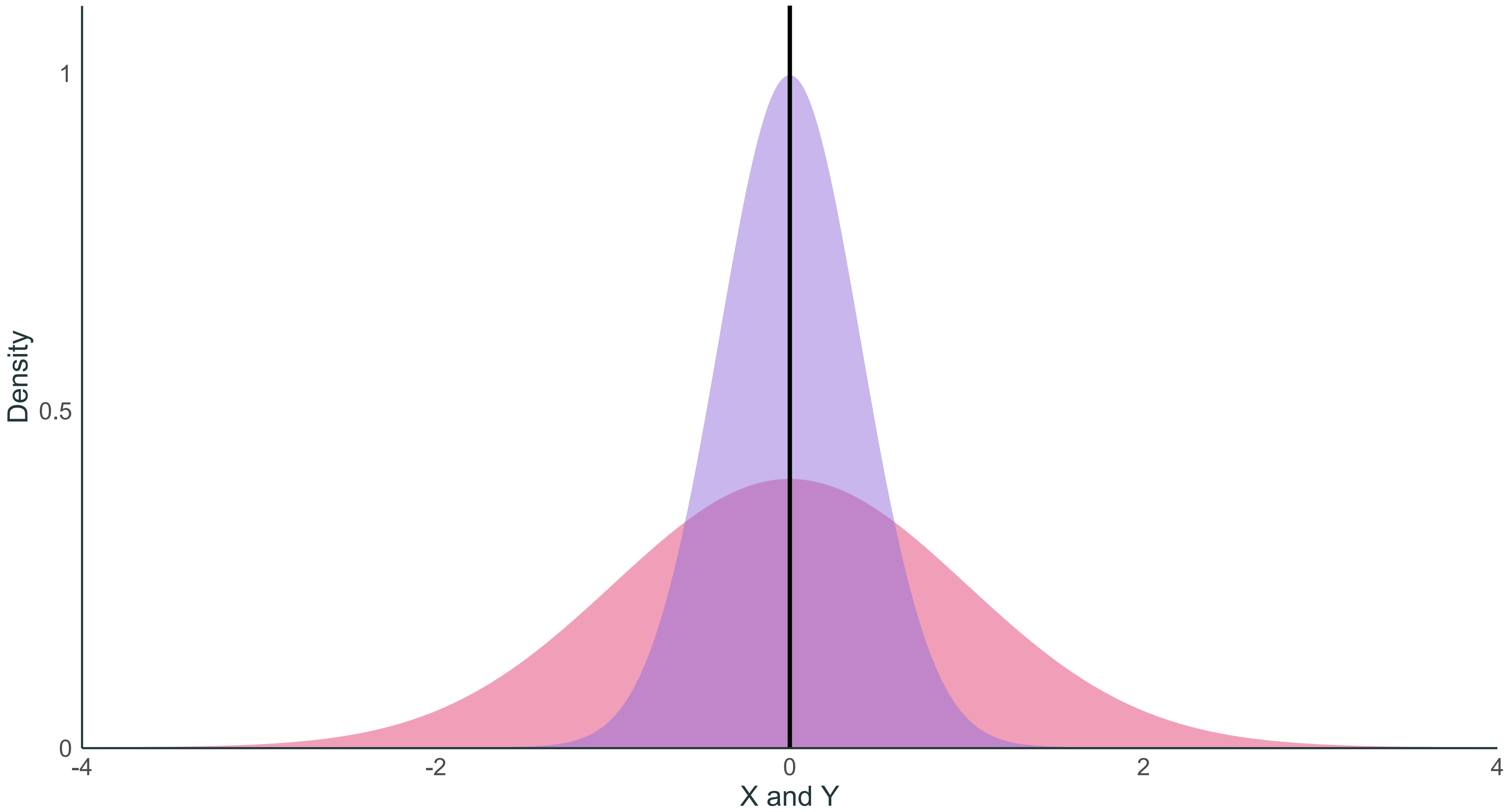
Except in special cases, **the transformation of an expected value** **is not** **the expected value of a transformed random variable**.

For some function $g(\,\cdot\,)$, it is typically the case that

$$g(\mathrm{E}(X)) \neq \mathrm{E}(g(X)).$$

# Variance

Random variables $X$ and $Y$ share the same population mean, but are distributed differently.



# Variance

How tightly is a random variable distributed about its mean?

- Let $\mu = \mathrm{E}(X)$ .
- Describe the distance of $X$ from its population mean $\mu$ as the squared difference: $(X - \mu)^2$ .

**Variance** tells us how far $X$ deviates from $\mu$ , *on average*:

$$\mathrm{Var}(X) \equiv \mathrm{E}\left((X - \mu)^2\right) = \sigma^2$$

- $\sigma^2$ is shorthand for variance.

## Rule 1

$\text{Var}(X) = 0 \iff X$ is a constant.

- If a random variable never deviates from its mean, then it has zero variance.
- If a random variable is always equal to its mean, then it's a (not-so-random) constant.

## Rule 2

For any constants $a$ and $b$ , $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$ .

## Example

Suppose $X$ is the high temperature in degrees Celsius in Eugene during August. If $Y$ is the temperature in degrees Fahrenheit, then $Y = 32 + \frac{9}{5}X$ . **What is $\mathrm{Var}(Y)$ ?**

- $\mathrm{Var}(Y) = (\frac{9}{5})^2\mathrm{Var}(X) = \frac{81}{25}\mathrm{Var}(X)$ .

# Standard Deviation

**Standard deviation** is the positive square root of the variance:

$$\mathrm{sd}(X) = +\sqrt{\mathrm{Var}(X)} = \sigma$$

- $\sigma$ is shorthand for standard deviation.

# Standard Deviation

## Rule 1

For any constant $c$ , $\mathrm{sd}(c) = 0$ .

## Rule 2

For any constants $a$ and $b$ , $\mathrm{sd}(aX + b) = |a|\mathrm{sd}(X)$ .

# Standardizing a Random Variable

When we're working with a random variable $X$ with an unfamiliar scale, it is useful to **standardize** it by defining a new variable $Z$ :

$$Z \equiv \frac{X - \mu}{\sigma}.$$

$Z$ has mean $0$ and standard deviation $1$ . How?

- First, some simple trickery: $Z = aX + b$ , where $a \equiv \frac{1}{\sigma}$ and $b \equiv -\frac{\mu}{\sigma}$ .
- $\mathrm{E}(Z) = a\mathrm{E}(X) + b = \mu\frac{1}{\sigma} - \frac{\mu}{\sigma} = 0$ .
- $\mathrm{Var}(Z) = a^2\mathrm{Var}(X) = \frac{1}{\sigma^2}\sigma^2 = 1$ .

# Covariance

**Idea:** Characterize the relationship between two random variables $X$ and $Y$.

**Definition:** $\text{Cov}(X, Y) \equiv \text{E}\left[(X - \mu_X)(Y - \mu_Y)\right] = \sigma_{xy}$.

- **Positive correlation:** When $\sigma_{xy} > 0$, then $X$ is above its mean when $Y$ is above its mean, *on average*.
- **Negative correlation:** When $\sigma_{xy} < 0$, then $X$ is below its mean when $Y$ is above its mean, *on average*.

# Covariance

## Rule 1

If $X$ and $Y$ are independent, then $\mathrm{Cov}(X, Y) = 0$ .

- **Statistical independence:** If $X$ and $Y$ are independent, then $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$ .
- $\mathrm{Cov}(X, Y) = 0$ means that $X$ and $Y$ are *uncorrelated*.

**Caution:** $\mathrm{Cov}(X, Y) = 0$ **does not imply** that $X$ and $Y$ are independent.

## Rule 2

For any constants $a$ , $b$ , $c$ , and $d$ , $\mathrm{Cov}(aX + b, cY + d) = ac\mathrm{Cov}(X, Y)$

# Correlation Coefficient

A problem with covariance is that it is sensitive to units of measurement.

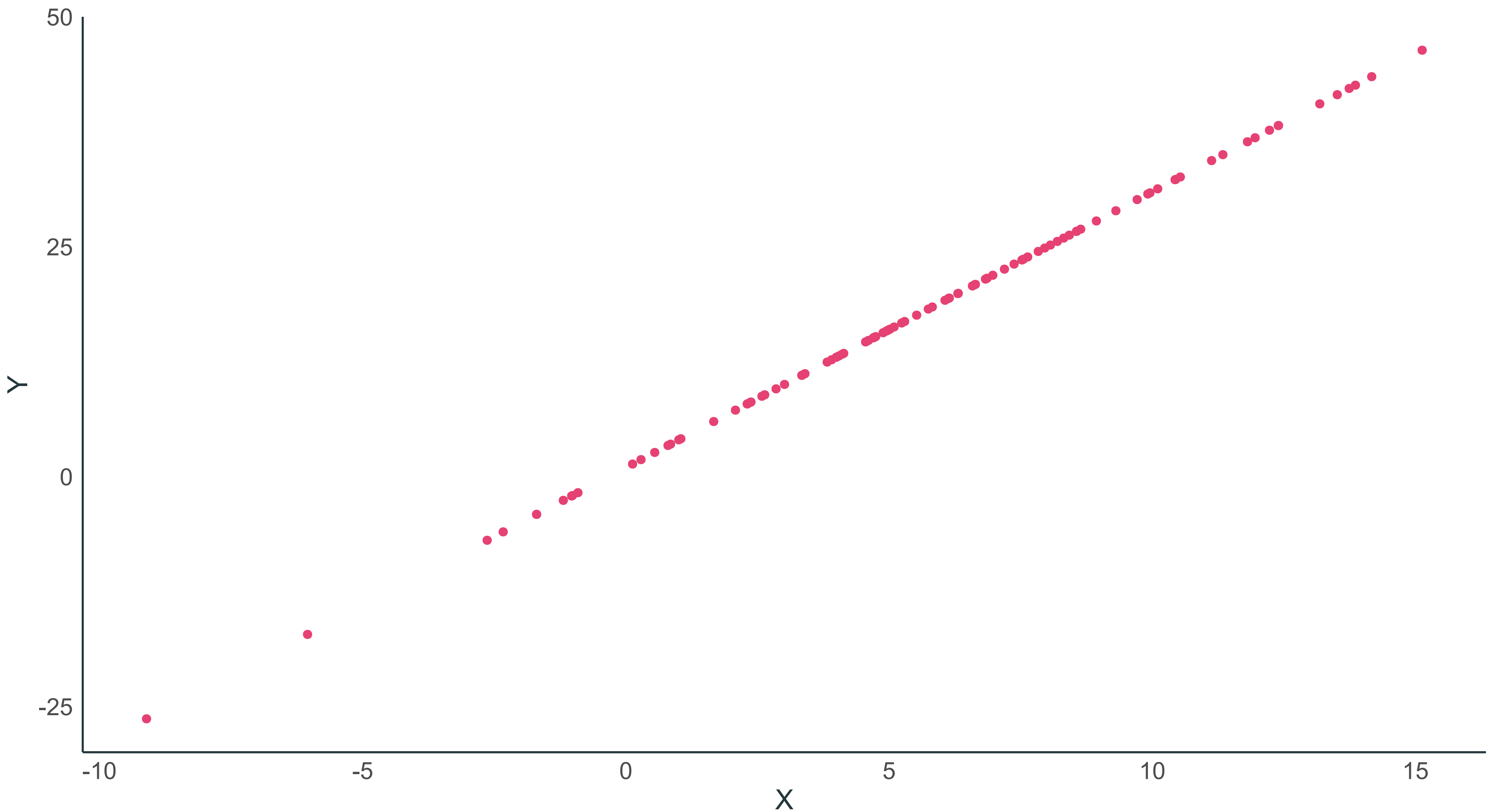
The **correlation coefficient** solves this problem by rescaling the covariance:

$$\text{Corr}(X, Y) \equiv \frac{\text{Cov}(X, Y)}{\text{sd}(X) \times \text{sd}(Y)} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

- Also denoted as $\rho_{XY}$.
- $-1 \leq \text{Corr}(X, Y) \leq 1$
- Invariant to scale: if I double $Y$, $\text{Corr}(X, Y)$ will not change.

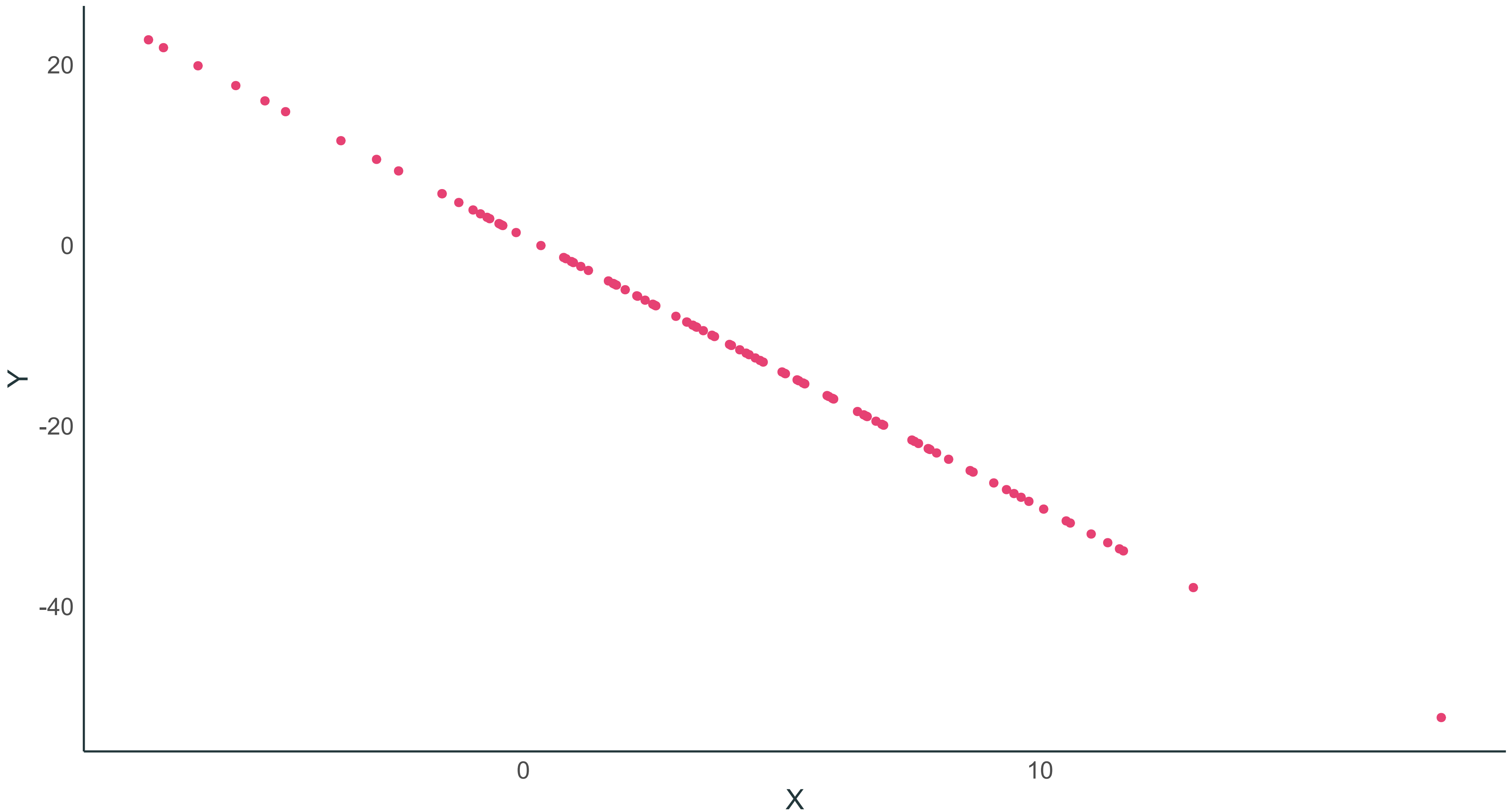
# Correlation Coefficient

Perfect positive correlation: $\mathrm{Corr}(X, Y) = 1$ .



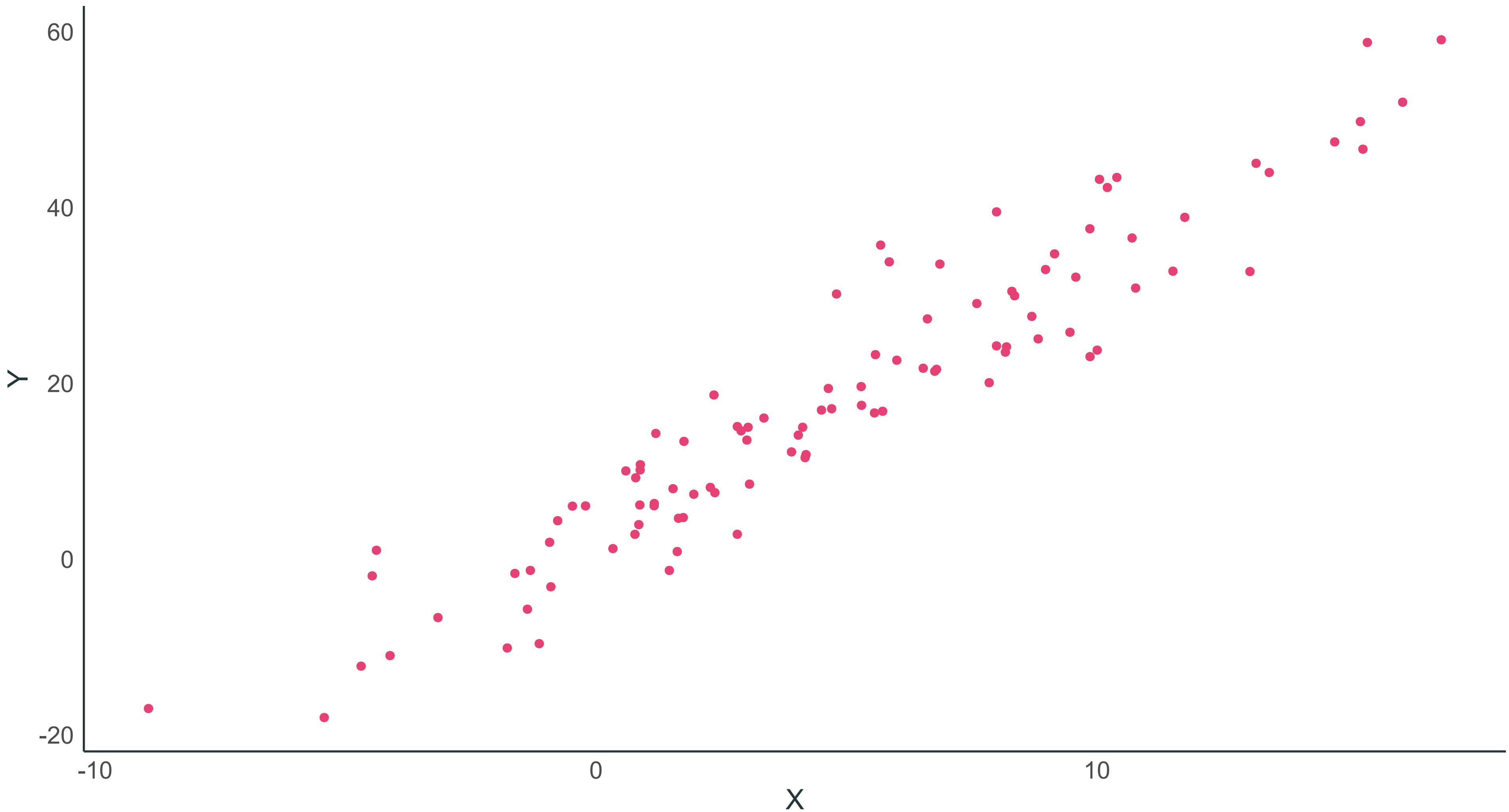
# Correlation Coefficient

Perfect negative correlation: $\text{Corr}(X, Y) = -1$ .



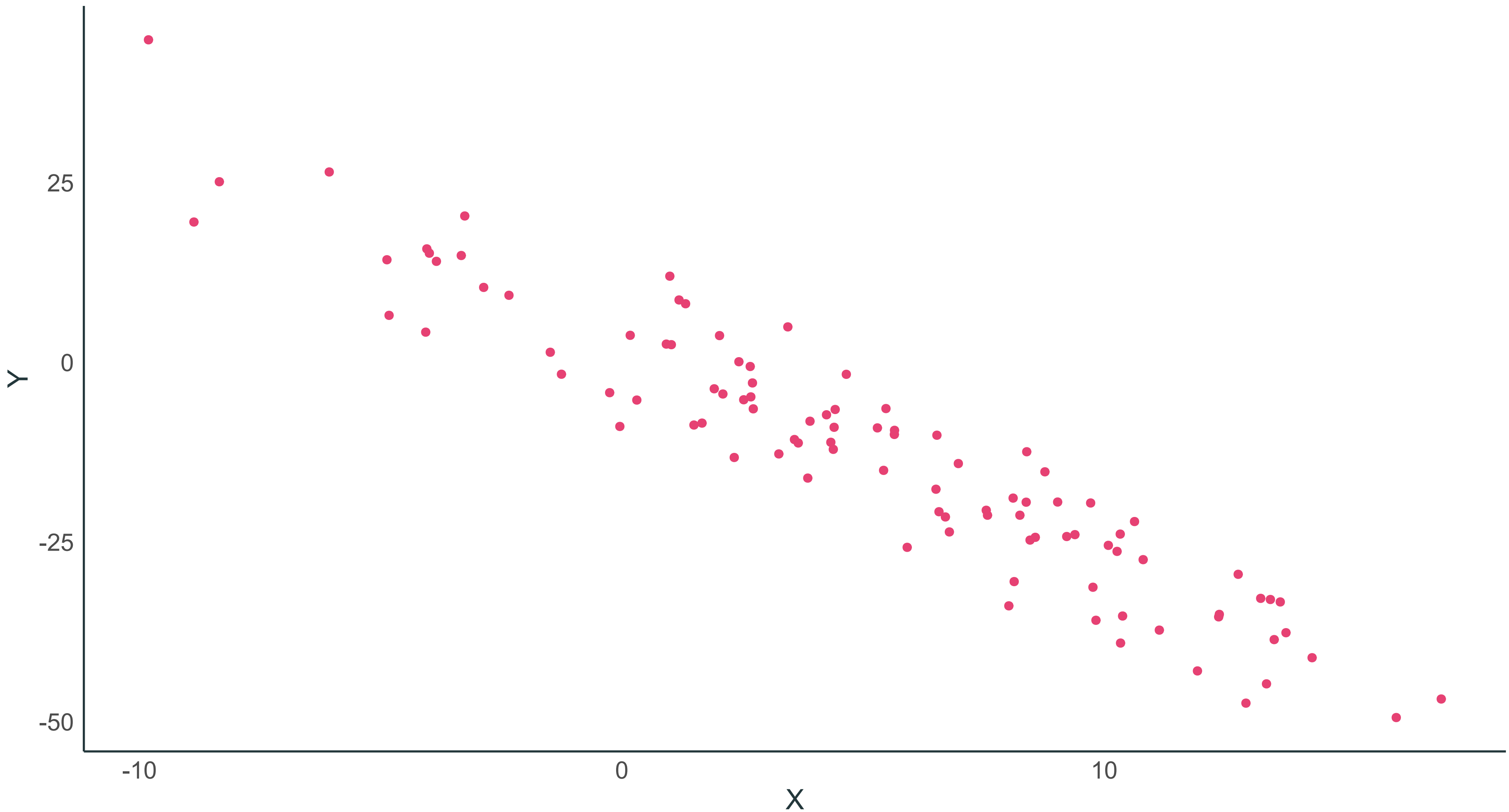
# Correlation Coefficient

Positive correlation: $\mathrm{Corr}(X, Y) > 0$ .



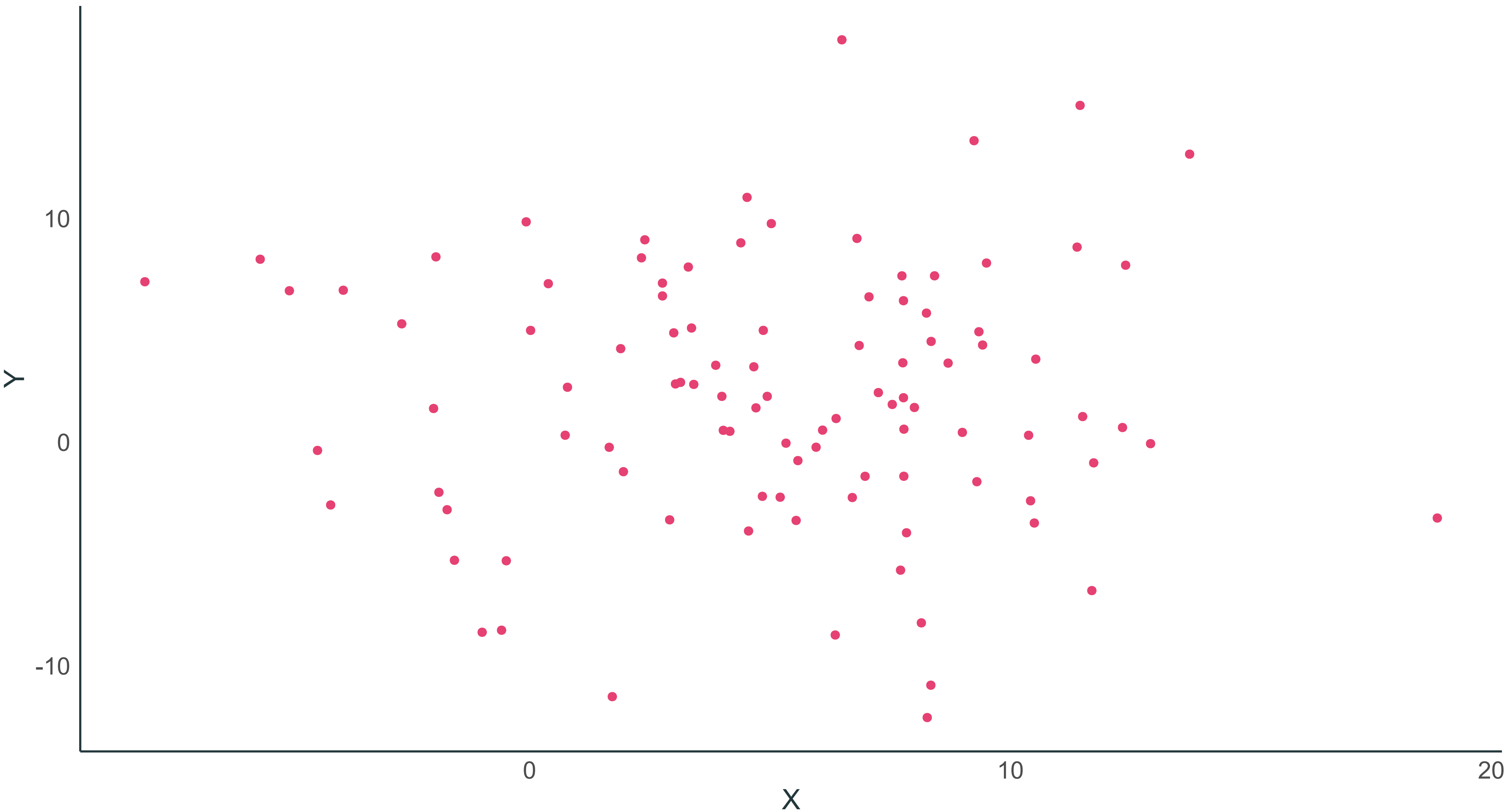
# Correlation Coefficient

Negative correlation: $\mathrm{Corr}(X, Y) < 0$ .



# Correlation Coefficient

No correlation: $\mathrm{Corr}(X, Y) = 0$ .



## Variance Rule 3

For constants $a$ and $b$ ,

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y).$$

- If $X$ and $Y$ are uncorrelated, then $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$
- If $X$ and $Y$ are uncorrelated, then $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$