

# Statistics Review I

## EC 320: Introduction to Econometrics

---

Philip Economides

Winter 2022

# Prologue

# Housekeeping

Problem set 1 posted. Will post next one upon deadline of the previous. This will allow folks more time/flexibility to get homework(s) organized.

- Due next Monday on Canvas by 1pm.
- Analytical and computational components (turned in separately).
- Each homework due every 2 weeks, Monday 1pm, last homework graded within week 10.

Issues with R?

- Office hours times included here.

# 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**<sup>\*</sup> a sequence of  $n$  observations, indexed by  $i$ :

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

Example:  $n = 5$

$i$	$x_i$
1	8
2	9
3	4
4	7
5	2

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

<sup>\*</sup> 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:  $n = 4$

$i$	$x_i$
1	7
2	4
3	10
4	3

$$\sum_{i=1}^4 x_i = 7 + 4 + 10 + 3 \\ = 23$$

$$\frac{1}{4} \sum_{i=1}^4 x_i = 6 \text{ (sample average)}$$

# Summation

## Rule 1

For any constant  $c$ ,

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

Example:  $n = 4$

$i$	$c$
1	2
2	2
3	2
4	2

$$\begin{aligned}\sum_{i=1}^4 2 &= 4 \times 2 \\ &= 8\end{aligned}$$



# Summation

## 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

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

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

# Summation

## Rule 3

If  $\{(x_i, y_i) : 1, \dots, n\}$  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

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

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

# Summation

## Caution

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

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

- 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  $x_1^2 + x_2^2 \neq (x_1 + x_2)^2 = x_1^2 + 2x_1x_2 + x_2^2$ .

# Cartesian coordinate system

# Cartesian coordinate system

A Cartesian plane (named after French mathematician Rene Descartes, who formalized its use in mathematics) is defined by two perpendicular number lines: the x-axis, which is horizontal, and the y-axis, which is vertical. Using these axes, we can describe any point in the plane using an ordered pair of numbers (x,y).

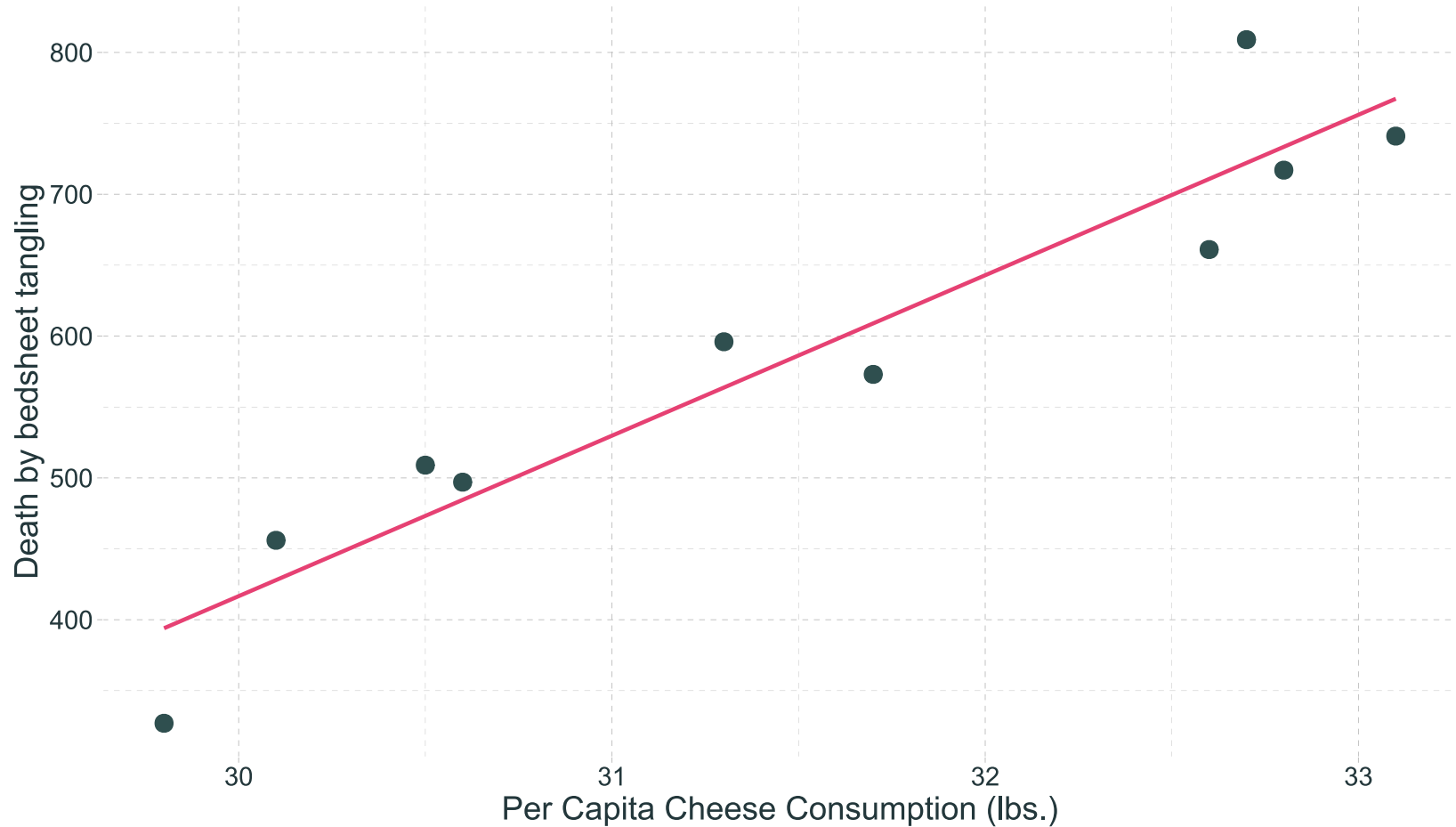
A particular line on this plane takes the form

$$y = a + bx$$

where  $a$  is known as the intercept and  $b$  is the slope.

Any incremental unit increase in  $x$  results in  $y$  increasing by  $b$ .

# Cartesian coordinate system



# 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.

- **Events:** heads or tails.
- Random Variable  $X$ : Win 1USD if heads ( $x_i = 1$ ), pay 1USD if tails ( $x_i = -1$ )
- **Sample Space:**  $\{-1, 1\}$

**Example 2:** Wages in a Census Study.

- **Events:** people make a certain wage.
- Random Variable  $W$ : the number that the study records as wage.
- **Sample Space:**  $[0, \max(\text{wage})]$

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

# Discrete Random Variables

## Probabilities

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

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

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

where

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

and

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

# Discrete Random Variables

## Probability density function

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

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

## Example

2020 Presidential election: 538 electoral votes at stake.

- $\{X : 0, 1, \dots, 538\}$  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:  $f(270) > 0$ .

# Discrete Random Variables

## 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.,  $\mathbb{P}(X \geq 1)$ .

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

# 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.*

In some cases, beneficial to switch to **bins** of measurement values.

- **B., Hansen (2015)** "Punishment and Deterrence: Evidence from Drunk Driving", AER
- Presents blood alcohol content within bins of 0.05 units

# Continuous Random Variables

Probability density functions also describe continuous random variables.

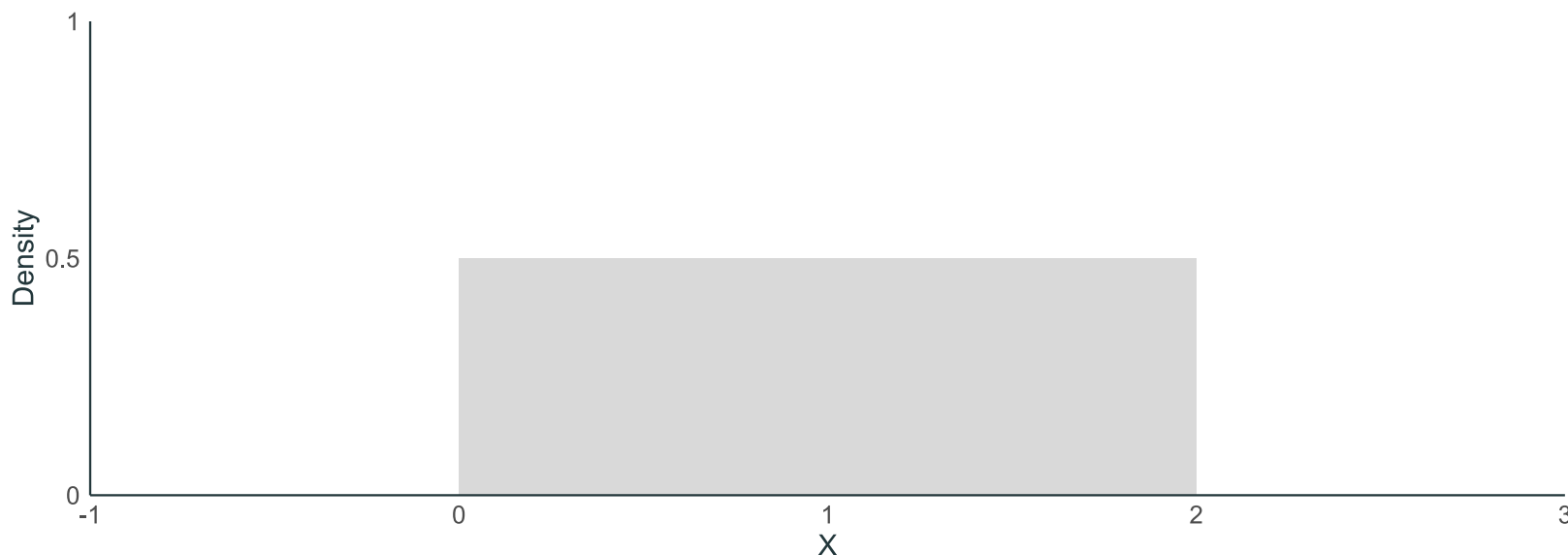
- Difference: Interested in the probability of events within a *range* of values.
- *e.g.* What is the probability of more than 1 inch of rain tomorrow?

# Continuous Random Variables

## Uniform Distribution

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

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





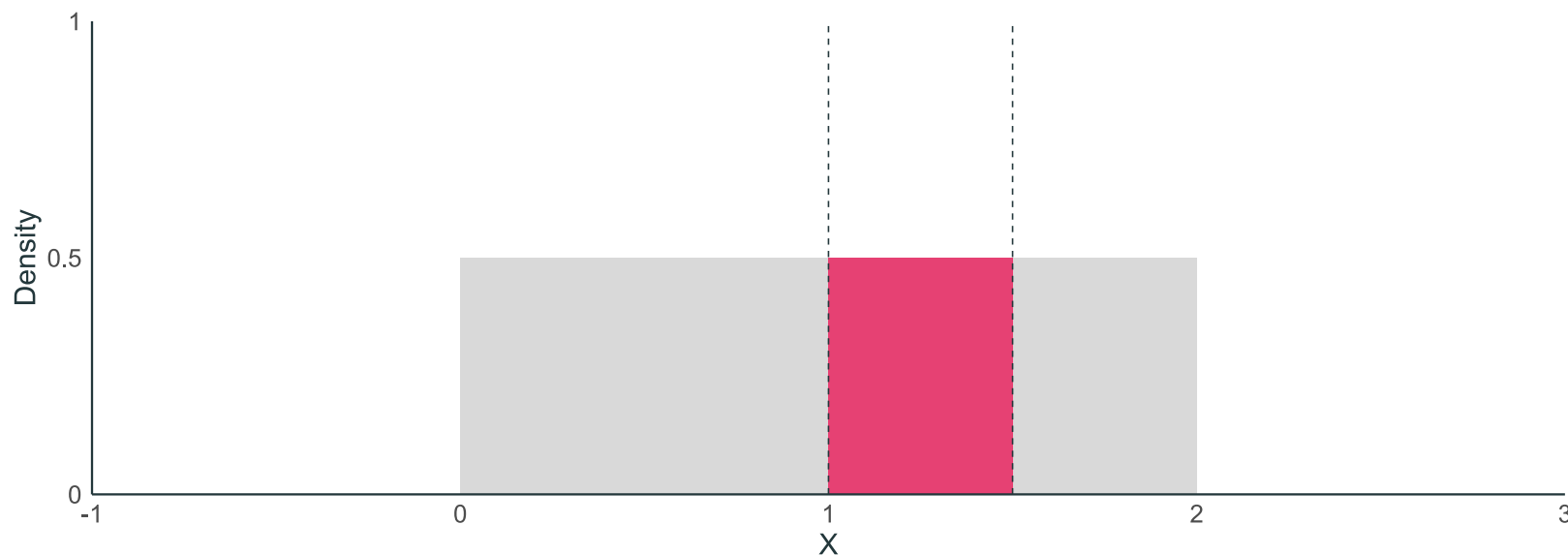
# Continuous Random Variables

## Uniform Distribution

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

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

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

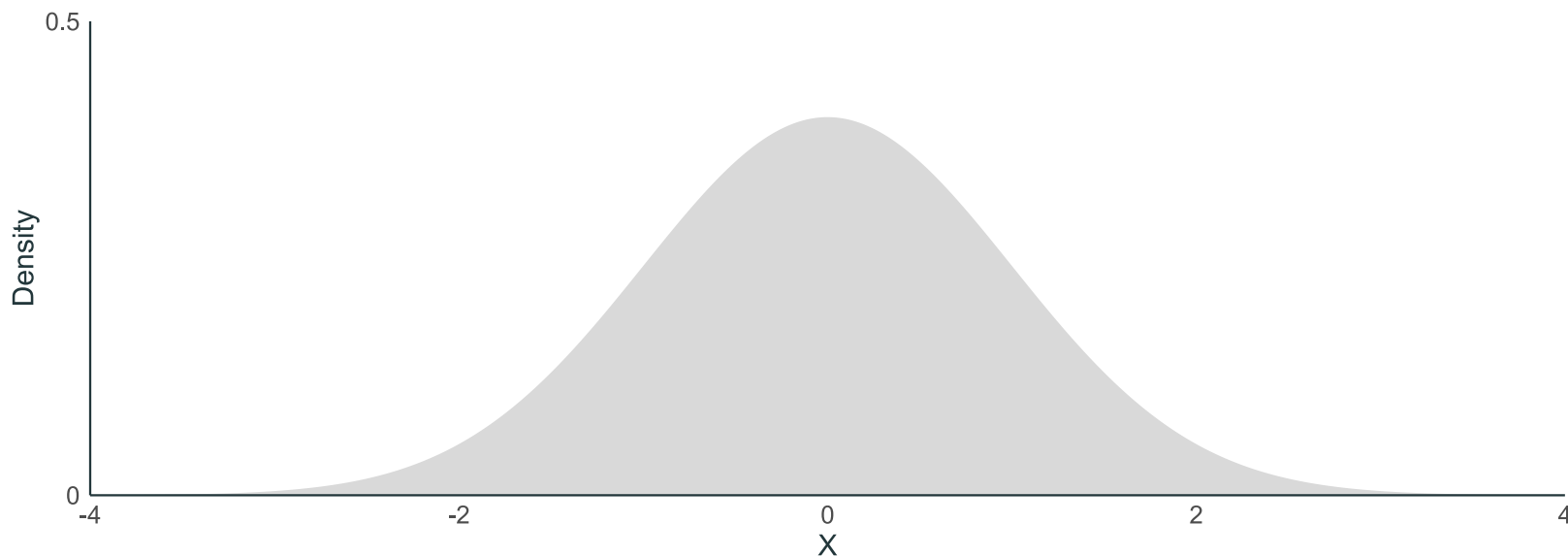


# Continuous Random Variables

## 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.

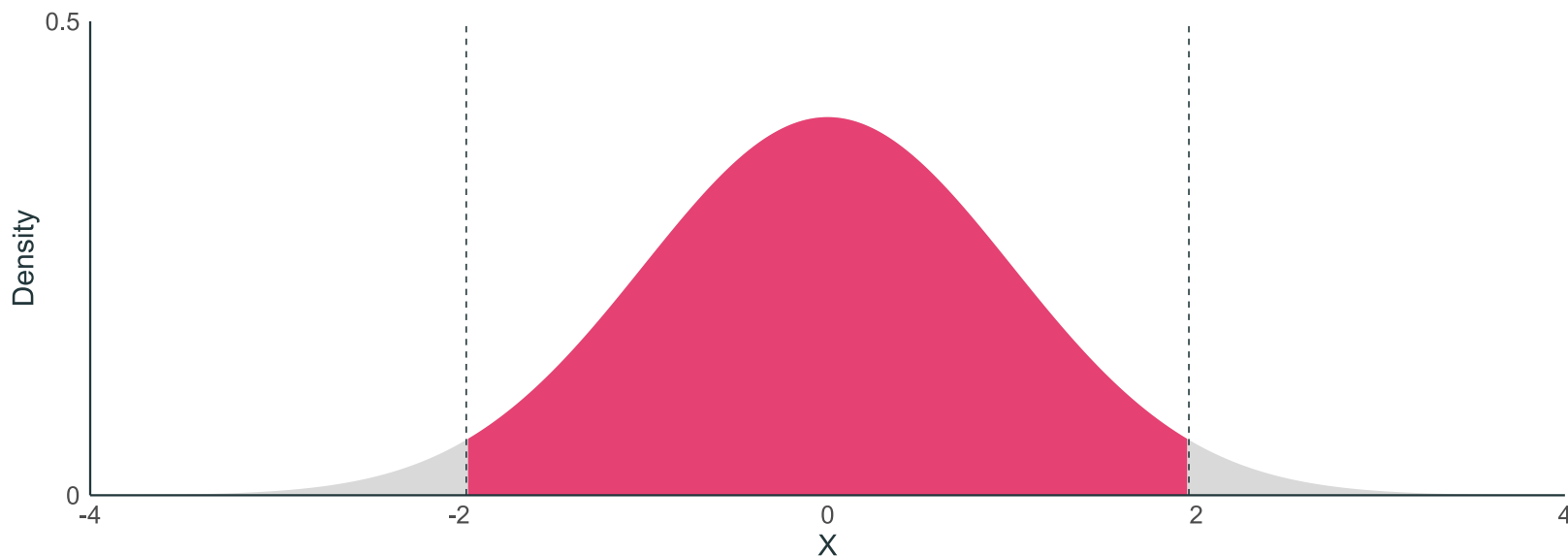


# Continuous Random Variables

## 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).
- $\mathbb{P}(-2 \leq X \leq 2) \approx 0.95$ .



# 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 to expect upon drawing from the distribution.

Other summary statistics we may be interested in include

- Median
- Standard deviation
- 25th percentile
- 75th percentile

# Expected Value

## Definition (Discrete)

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

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

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

# Expected Value

## 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?**

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

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

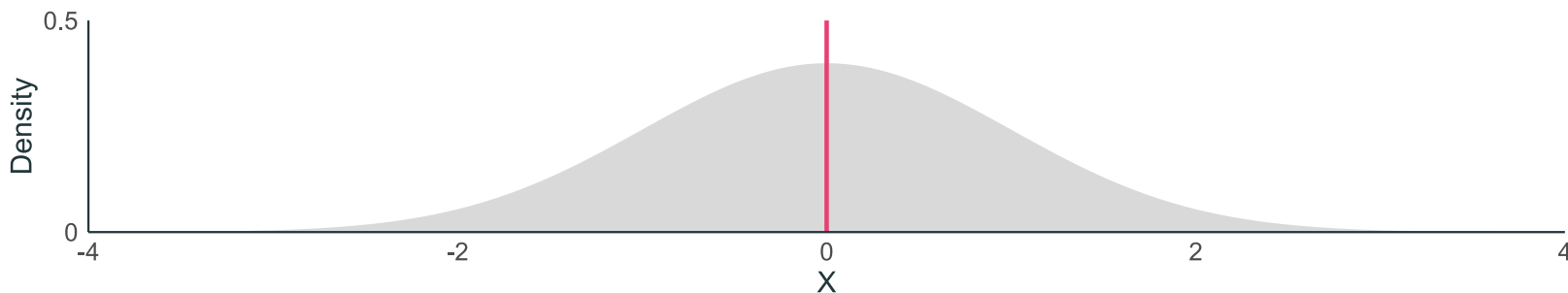
# Expected Value

## Definition (Continuous)

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

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

- **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  $c$ ,  $\mathbb{E}(c) = c$ .

## Not-so-exciting examples

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

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

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



# Expected Value

## Rule 2

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

## Example

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

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

# Expected Value

## Rule 3

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

$$\mathbb{E}(a_1 X_1 + a_2 X_2 + \dots + a_n X_n) = a_1 \mathbb{E}(X_1) + a_2 \mathbb{E}(X_2) + \dots + a_n \mathbb{E}(X_n).$$

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

# Expected Value

## 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  $\mathbb{E}(X_1) = 43$ ,  $\mathbb{E}(X_2) = 56$ , and  $\mathbb{E}(X_3) = 21$ . The prices of small, medium, and large beverages are 1.75, 2.50, and 3.25 dollars. **What is expected revenue?**

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

# Expected Value

## Caution

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

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

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

**No!**

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

# Expected Value

## Caution

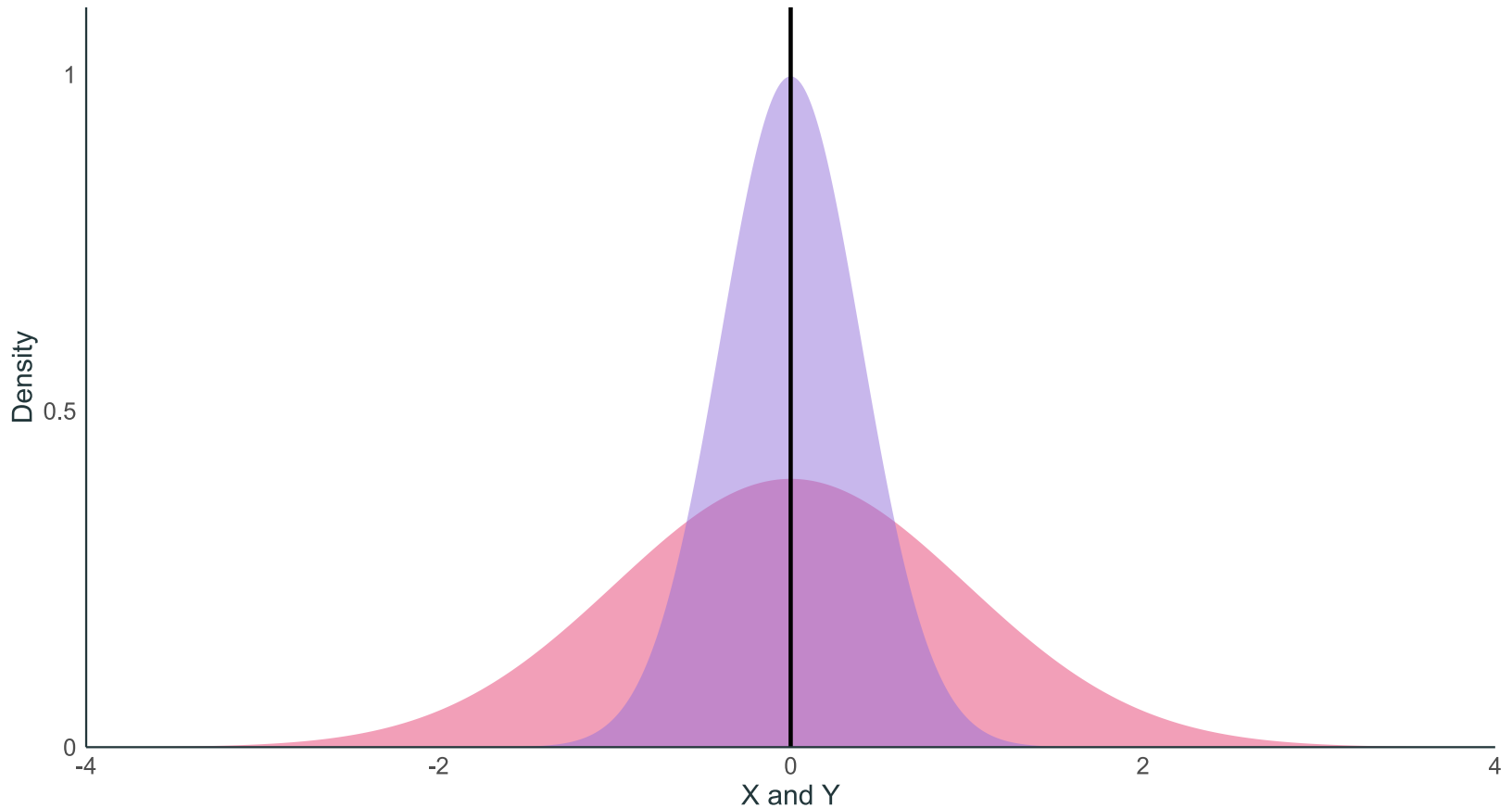
Except in special cases, **the transformation of an expected value is not the expected value of a transformed random variable.**

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

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

# 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  $\mu = \mathbb{E}(X)$ .
- 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*:

$$\text{Var}(X) \equiv \mathbb{E}((X - \mu)^2) = \sigma^2$$

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

# Variance

## Rule 1

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

- 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.



# Variance

## Rule 2

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

## 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  $\text{Var}(Y)$ ?**

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

# Standard Deviation

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

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

- $\sigma$  is shorthand for standard deviation.

# Standard Deviation

## Rule 1

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

## Rule 2

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

# 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  $b \equiv -\frac{\mu}{\sigma}$ .
- $\mathbb{E}(Z) = a \mathbb{E}(X) + b = \mu \frac{1}{\sigma} - \frac{\mu}{\sigma} = 0$ .
- $\text{Var}(Z) = a^2 \text{Var}(X) = \frac{1}{\sigma^2} \sigma^2 = 1$ .

# Covariance

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

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

- **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  $\text{Cov}(X, Y) = 0$ .

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

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

# Covariance

## Rule 2

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

# 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:

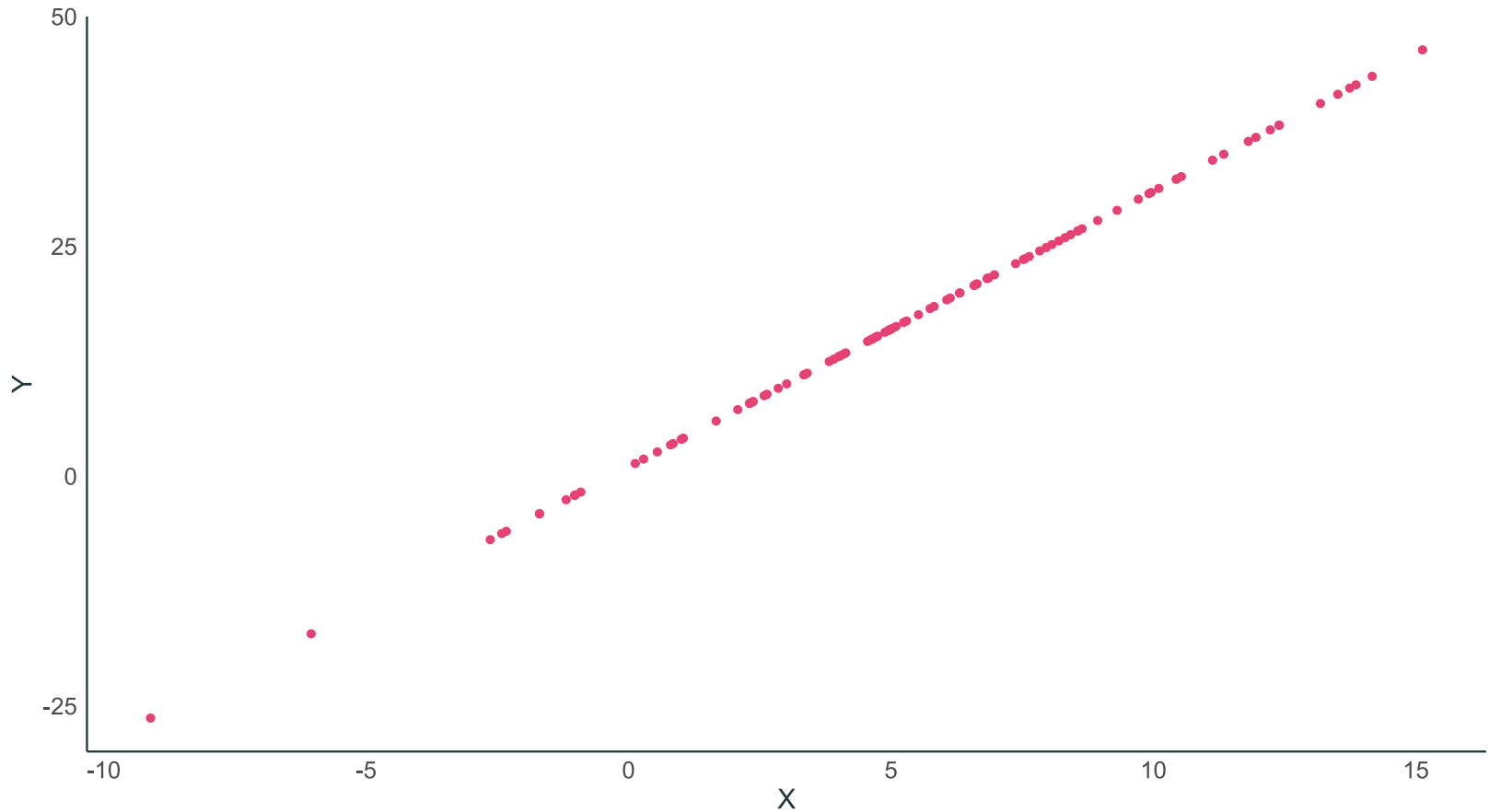
$$\text{Corr}(X, Y) \equiv \frac{\text{Cov}(X, Y)}{\text{sd}(X) \times \text{sd}(Y)} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

- 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.



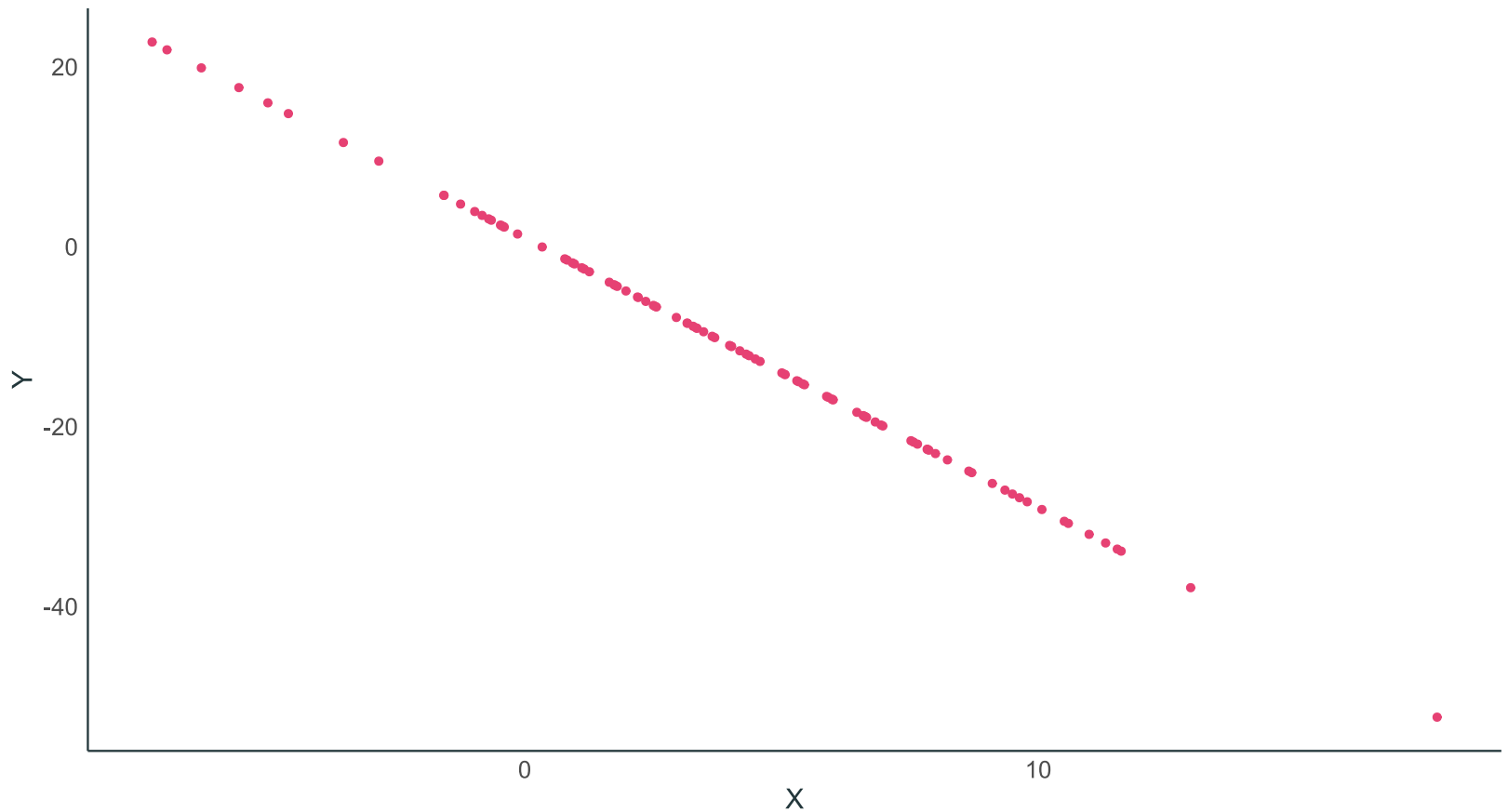
# Correlation Coefficient

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



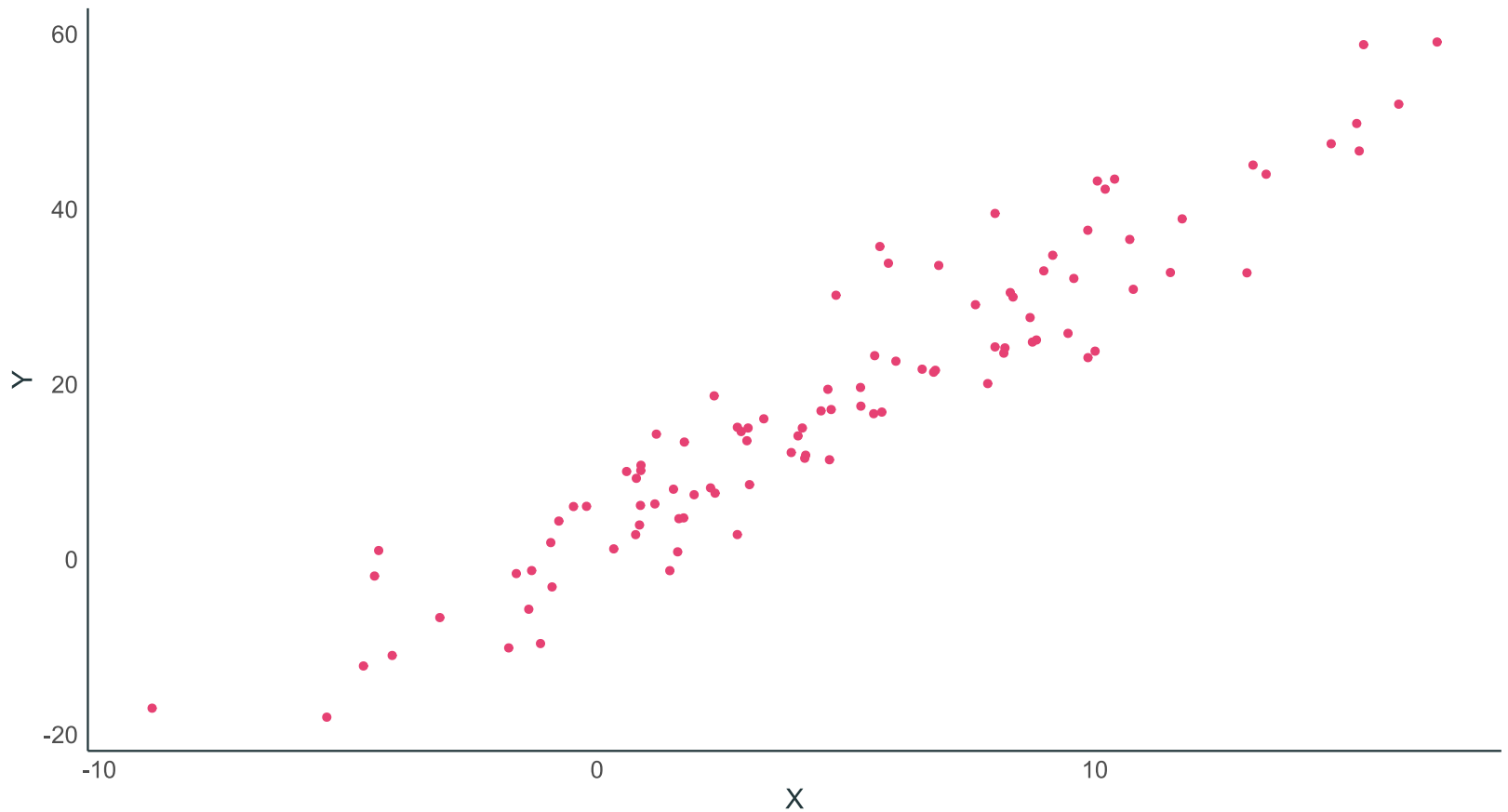
# Correlation Coefficient

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



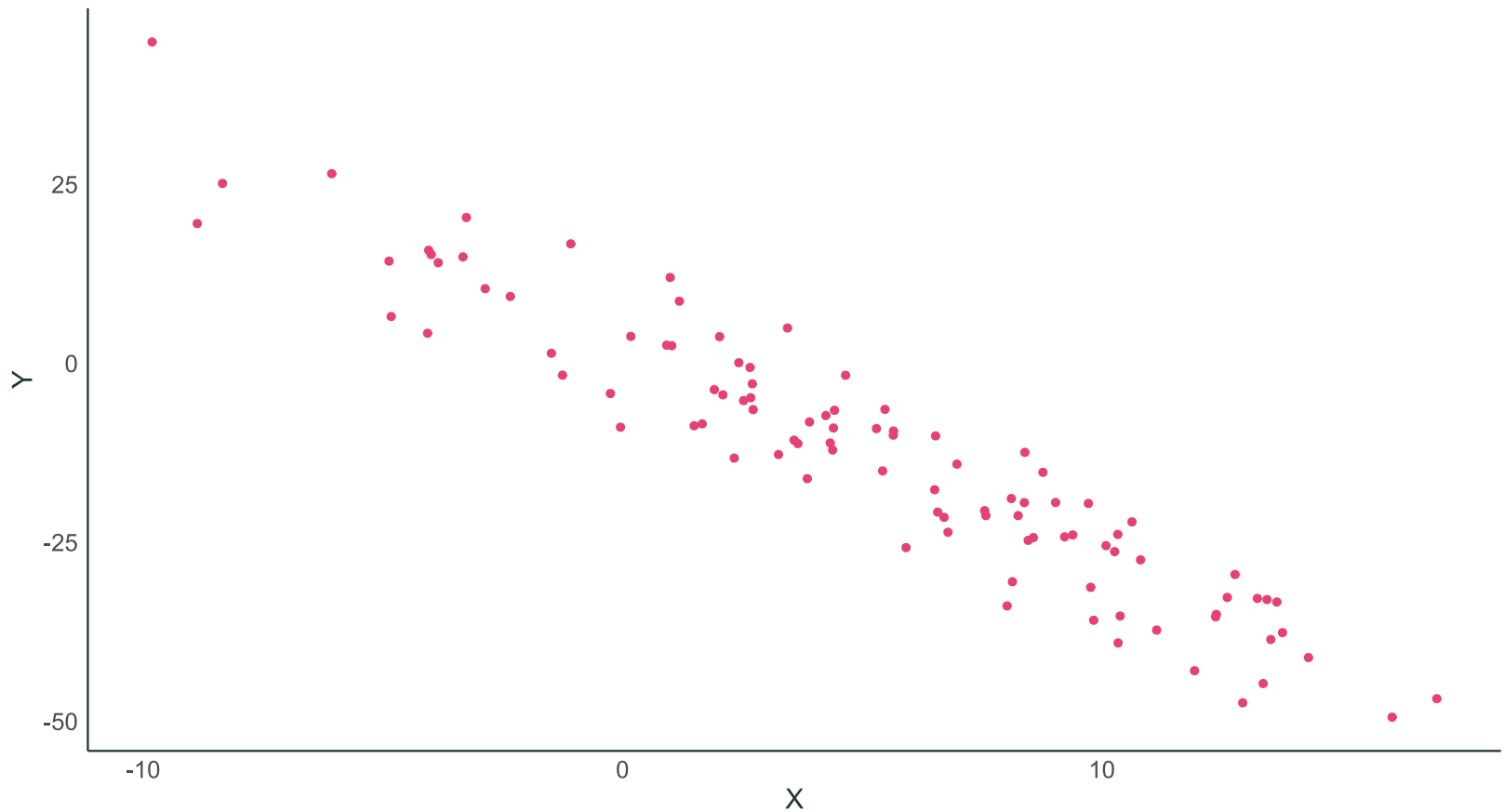
# Correlation Coefficient

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



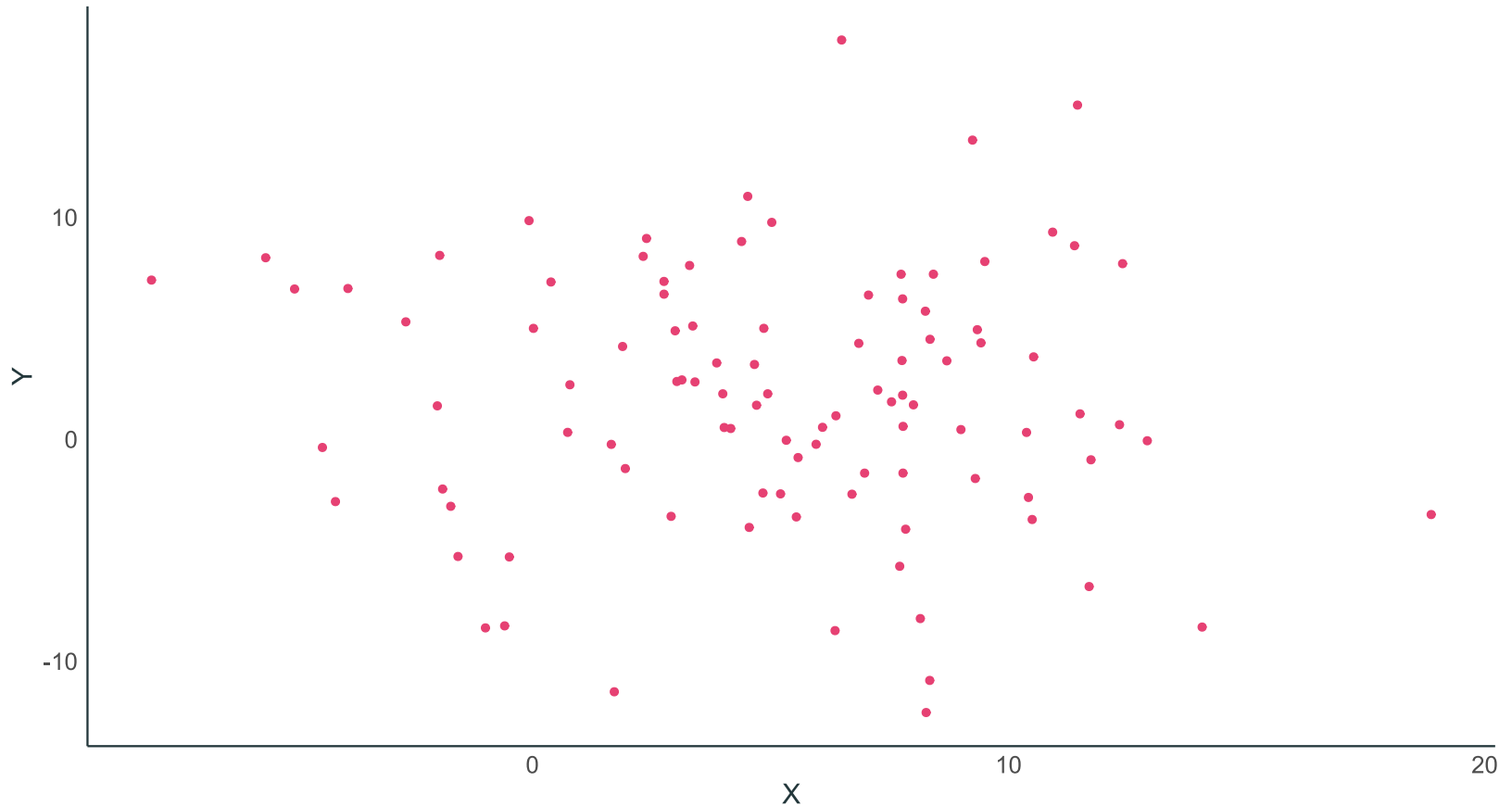
# Correlation Coefficient

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



# Correlation Coefficient

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



# Variance, Revisited

## 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  $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$

