Expectation and Variance

Christopher Lucas

Expectation of Discrete Random Variables

What is the average value shown on a six sided die?

Expectation of Discrete Random Variables

What is the average value shown on a six sided die?

y	p(y)
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

What is $\mathbb{E}(Y)$?

Expectation of Discrete Random Variables

In general, for discrete random variables,

$$\mathbb{E}(Y) = \sum yp(y)$$

For this RV,

Expectation of Discrete Random Variables

In general, for discrete random variables,

$$\mathbb{E}(Y) = \sum yp(y)$$

For this RV,

$$= 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = 3.5$$

Expectation of Continuous Random Variables

What about a continuous random variable?

$$Y \sim U(0, 2)$$

Expectation of Continuous Random Variables

What about a continuous random variable?

$$Y \sim U(0, 2)$$

$$f(y) =$$

Expectation of Continuous Random Variables

What about a continuous random variable?

$$Y \sim U(0, 2)$$

$$f(y) = \frac{1}{2 - 0}$$

Expectation of Continuous Random Variables

What about a continuous random variable?

$$Y \sim U(0,2)$$

$$f(y) = \frac{1}{2-0}$$

In general,

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} y \cdot f(y)dy$$

For this RV, this becomes

$$\mathbb{E}(Y) = \int_{0}^{2} y \cdot f(y)dy =$$

Expectation of Continuous Random Variables

What about a continuous random variable?

$$Y \sim U(0,2)$$

$$f(y) = \frac{1}{2-0}$$

In general,

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} y \cdot f(y) dy$$

For this RV, this becomes

$$\mathbb{E}(Y) = \int_{0}^{2} y \cdot f(y) dy = \int_{0}^{2} y \cdot \frac{1}{2} dy =$$

Expectation of Continuous Random Variables

What about a continuous random variable?

$$Y \sim U(0,2)$$

$$f(y) = \frac{1}{2-0}$$

In general,

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} y \cdot f(y) dy$$

For this RV, this becomes

$$\mathbb{E}(Y) = \int_0^2 y \cdot f(y) dy = \int_0^2 y \cdot \frac{1}{2} dy = \frac{1}{2} \cdot \frac{1}{2} y^2 |_0^2 =$$

Expectation of Continuous Random Variables

What about a continuous random variable?

$$Y \sim U(0, 2)$$

$$f(y) = \frac{1}{2 - 0}$$

In general,

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} y \cdot f(y) dy$$

For this RV, this becomes

$$\mathbb{E}(Y) = \int_0^2 y \cdot f(y) dy = \int_0^2 y \cdot \frac{1}{2} dy = \frac{1}{2} \cdot \frac{1}{2} y^2 |_0^2 = \frac{1}{4}(2^2 - 0^2) = 1$$

Linearity of Expectation

Simple rules for expectations:

- If c is a constant, then $\mathbb{E}(c) =$

Linearity of Expectation

Simple rules for expectations:

- If c is a constant, then $\mathbb{E}(c) = c$

Linearity of Expectation

Simple rules for expectations:

- If c is a constant, then $\mathbb{E}(c) = c$
- $\mathbb{E}(cY) =$

Linearity of Expectation

Simple rules for expectations:

- If c is a constant, then $\mathbb{E}(c) = c$
- $\mathbb{E}(cY) = c \cdot \mathbb{E}(y)$

Linearity of Expectation

Simple rules for expectations:

- If c is a constant, then $\mathbb{E}(c) = c$
- $\mathbb{E}(cY) = c \cdot \mathbb{E}(y)$
- $\mathbb{E}(Y_1 + Y_2 + Y_3) =$

Linearity of Expectation

Simple rules for expectations:

- If c is a constant, then $\mathbb{E}(c) = c$
- $\mathbb{E}(cY) = c \cdot \mathbb{E}(y)$
- $\mathbb{E}(Y_1 + Y_2 + Y_3) = \mathbb{E}(Y_1) + \mathbb{E}(Y_2) + \mathbb{E}(Y_3)$

Linearity of Expectation

Simple rules for expectations:

- If c is a constant, then $\mathbb{E}(c) = c$
- $\mathbb{E}(cY) = c \cdot \mathbb{E}(y)$
- $\mathbb{E}(Y_1 + Y_2 + Y_3) = \mathbb{E}(Y_1) + \mathbb{E}(Y_2) + \mathbb{E}(Y_3)$

In addition, if Y_1 and Y_2 are independent, then $\mathbb{E}(Y_1 \cdot Y_2) = \mathbb{E}(Y_1) \cdot \mathbb{E}(Y_2)$

Expectation of Functions of Random Variables

Suppose Y is discrete with PDF $p(y)$, and $g(Y)$ is a function of Y. The expectation is:

$$\mathbb{E}[g(Y)] = \sum g(y)p(y)$$

Next, suppose Y is continuous with PDF $f(y)$, and $g(Y)$ is a function of Y. The expectation is:

$$\mathbb{E}[g(Y)] = \int_{-\infty}^{\infty} g(y) \times f(y)dy$$

Expectation of Functions of Random Variables

Suppose Y is discrete with PDF $p(y)$, and $g(Y)$ is a function of Y. The expectation is:

$$\mathbb{E}[g(Y)] = \sum g(y)p(y)$$

Next, suppose Y is continuous with PDF $f(y)$, and $g(Y)$ is a function of Y. The expectation is:

$$\mathbb{E}[g(Y)] = \int_{-\infty}^{\infty} g(y) \times f(y)dy$$

Is this consistent with the previous rules? What if $g(y) = y$? If $g(y) = cy$?

For the following discrete PDF:

y	p(y)
0	1/8
1	1/4
2	3/8
3	1/4

Let's define $g(Y) = y^2$.

For the following discrete PDF:

y	p(y)
0	1/8
1	1/4
2	3/8
3	1/4

Let's define $g(Y) = y^2$. We can take the expectation of $g(Y)$, because functions of random variables are also random variables.

For the following discrete PDF:

y	g(x)	p(y)
0	0	1/8
1	1	1/4
2	4	3/8
3	9	1/4

Let's define $g(Y) = y^2$. We can take the expectation of $g(Y)$, because functions of random variables are also random variables.

What is $\mathbb{E}(Y^2)$?

For the following discrete PDF:

y	g(x)	p(y)
0	0	1/8
1	1	1/4
2	4	3/8
3	9	1/4

Let's define $g(Y) = y^2$. We can take the expectation of $g(Y)$, because functions of random variables are also random variables.

What is $\mathbb{E}(Y^2)$?

$$\mathbb{E}(Y^2) = \sum y^2 p(y) = 0^2 \times \frac{1}{8} + 1^2 \times \frac{1}{4} + 2^2 \times \frac{3}{8} + 3^2 \times \frac{1}{4} = 4$$

Discrete Random Variables

Probability distribution for six-sided die:

y	p(y)
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

What is $\mathbb{E}(Y^2)$?

Discrete Random Variables

The expected value is:

$$\begin{aligned}\mathbb{E}(Y^2) &= \sum y^2 p(y) \\ &= 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + 3^2 \times \frac{1}{6} + 4^2 \times \frac{1}{6} + 5^2 \times \frac{1}{6} + 6^2 \times \frac{1}{6} \\ &= \frac{91}{6} \approx 15.1667\end{aligned}$$

Now let's try one for a continuous random variable.

$$f(y) = \{ \; y/2 \quad 0 \leq y \leq 2$$

What is $\mathbb{E}(Y^2)$?

$$f(y) = \{ \; y/2 \quad 0 \leq y \leq 2$$

$$\mathbb{E}(Y^2) = \int_0^2 y^2 \frac{y}{2} dy = \int_0^2 \frac{y^3}{2} dy = \frac{y^4}{8} |_0^2 = \frac{2^4}{8} = 2$$

Variance

- One method for describing a PDF is the expectation.
- A second feature of a distribution that we are often interested in is the dispersion, or the extent to which the values of the distribution are spread out.

Variance

One measure of dispersion that we could potentially use:

Variance

One measure of dispersion that we could potentially use:

$$\mathbb{E}[Y - \mathbb{E}(Y)] = \mathbb{E}(Y - \mu)$$

where we define $\mu = \mathbb{E}(Y)$ for convenience.

Variance

One measure of dispersion that we could potentially use:

$$\mathbb{E}[Y - \mathbb{E}(Y)] = \mathbb{E}(Y - \mu)$$

where we define $\mu = \mathbb{E}(Y)$ for convenience. But this measure has a problem.

Variance

One measure of dispersion that we could potentially use:

$$\mathbb{E}[Y - \mathbb{E}(Y)] = \mathbb{E}(Y - \mu)$$

where we define $\mu = \mathbb{E}(Y)$ for convenience. But this measure has a problem.

$$\mathbb{E}(Y - \mu) = \mathbb{E}(Y) - \mathbb{E}(\mu) = \mu - \mu = 0$$

One way to get around this is to use squares, which are always nonnegative.

$$\mathbb{E}[(Y - \mathbb{E}(Y))^2] = \mathbb{E}[(Y - \mu)^2]$$

Variance

One measure of dispersion that we could potentially use:

$$\mathbb{E}[Y - \mathbb{E}(Y)] = \mathbb{E}(Y - \mu)$$

where we define $\mu = \mathbb{E}(Y)$ for convenience. But this measure has a problem.

$$\mathbb{E}(Y - \mu) = \mathbb{E}(Y) - \mathbb{E}(\mu) = \mu - \mu = 0$$

One way to get around this is to use squares, which are always nonnegative.

$$\mathbb{E}[(Y - \mathbb{E}(Y))^2] = \mathbb{E}[(Y - \mu)^2] = \mathrm{Var}(Y) = \sigma^2$$

which we call the variance of Y.

Variance

The variance for a discrete variable is:

$$\mathrm{Var}(Y) = \sigma^2 = \sum (y - \mu)^2 \cdot p(y)$$

and for a continuous variable is:

$$\mathrm{Var}(Y) = \sigma^2 = \int_{-\infty}^{\infty} (y - \mu)^2 f(y) dy$$

$\sigma = \sqrt{\mathrm{Var}(Y)}$ is called the "standard deviation" and is often easier to interpret, since it is on the same scale (measured in the same units) as Y

Variance

There is also an equivalent but easier to use formula for the variance:

$$\mathrm{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = \mathbb{E}(Y^2) - \mu^2$$

Variance

There is also an equivalent but easier to use formula for the variance:

$$\mathrm{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = \mathbb{E}(Y^2) - \mu^2$$

Why?

Variance

There is also an equivalent but easier to use formula for the variance:

$$\mathrm{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = \mathbb{E}(Y^2) - \mu^2$$

Why?

$$\begin{aligned}
\mathbb{E}[(Y-\mu)^2] &= \mathbb{E}(Y^2 - 2\mu Y + \mu^2) \\
&= \mathbb{E}(Y^2) - \mathbb{E}(2\mu Y) + \mathbb{E}(\mu^2) \\
&= \mathbb{E}(Y^2) - 2\mu\,\mathbb{E}(Y) + \mu^2 \\
&= \mathbb{E}(Y^2) - 2\mu^2 + \mu^2 \\
&= \mathbb{E}(Y^2) - \mu^2
\end{aligned}$$

Note that μ is a constant and $\mathbb{E}(Y) = \mu$.

y	p(y)
0	1/8
1	1/4
2	3/8
3	1/4

The expected value is:

$$\mathbb{E}(Y) = \sum yp(y) = 0 \times \frac{1}{8} + 1 \times \frac{1}{4} + 2 \times \frac{3}{8} + 3 \times \frac{1}{4} = 1.75$$

What is the variance?

$$\begin{aligned}
\mathrm{Var}(Y) &= \sum (y-\mu)^2 p(y) \\
&= (0-1.75)^2\frac{1}{8} + (1-1.75)^2\frac{1}{4} + (2-1.75)^2\frac{3}{8} + (3-1.75)^2\frac{1}{4} \\
&= .9375
\end{aligned}$$

Or

$$\mathbb{E}(Y^2) = \sum y^2 p(y) = 0^2\frac{1}{8} + 1^2\frac{1}{4} + 2^2\frac{3}{8} + 3^2\frac{1}{4} = 4$$

$$\mathrm{Var}(Y) =$$

$$\begin{aligned}\mathrm{Var}(Y) &= \sum (y-\mu)^2 p(y) \\ &= (0-1.75)^2\frac{1}{8} + (1-1.75)^2\frac{1}{4} + (2-1.75)^2\frac{3}{8} + (3-1.75)^2\frac{1}{4} \\ &= .9375\end{aligned}$$

Or

$$\mathbb{E}(Y^2) = \sum y^2 p(y) = 0^2\frac{1}{8} + 1^2\frac{1}{4} + 2^2\frac{3}{8} + 3^2\frac{1}{4} = 4$$

$$\mathrm{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 =$$

$$\begin{aligned}\mathrm{Var}(Y) &= \sum (y-\mu)^2 p(y) \\ &= (0-1.75)^2\frac{1}{8} + (1-1.75)^2\frac{1}{4} + (2-1.75)^2\frac{3}{8} + (3-1.75)^2\frac{1}{4} \\ &= .9375\end{aligned}$$

Or

$$\mathbb{E}(Y^2) = \sum y^2 p(y) = 0^2\frac{1}{8} + 1^2\frac{1}{4} + 2^2\frac{3}{8} + 3^2\frac{1}{4} = 4$$

$$\mathrm{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = 4 - 1.75^2 = 0.9375$$

What is the variance for a six-sided die?

What is the variance for a six-sided die?

$\mathbb{E}(Y) = 3.5$

$\mathbb{E}(Y^2) = \frac{91}{6} \approx 15.167$

$$\mathrm{Var}(Y) = \mathbb{E}(Y^2) - \mathbb{E}(Y)^2 = \frac{91}{6} - 3.5^2 \approx 2.917$$

What is average squared distance between two rolls on a six-sided die?

What is average squared distance between two rolls on a six-sided die?

$$\mathbb{E}[(Y_1 - Y_2)^2] =$$

What is average squared distance between two rolls on a six-sided die?

$$\mathbb{E}[(Y_1 - Y_2)^2] = \mathbb{E}[Y_1^2 - 2Y_1Y_2 + Y_2^2] =$$

What is average squared distance between two rolls on a six-sided die?

$$\mathbb{E}[(Y_1 - Y_2)^2] = \mathbb{E}[Y_1^2 - 2Y_1Y_2 + Y_2^2] = \mathbb{E}[Y_1^2] - \mathbb{E}[2Y_1Y_2] + \mathbb{E}[Y_2^2]$$

$$=$$

What is average squared distance between two rolls on a six-sided die?

$$\mathbb{E}[(Y_1 - Y_2)^2] = \mathbb{E}[Y_1^2 - 2Y_1Y_2 + Y_2^2] = \mathbb{E}[Y_1^2] - \mathbb{E}[2Y_1Y_2] + \mathbb{E}[Y_2^2]$$

$$= \mathbb{E}[Y_1^2] - 2\,\mathbb{E}[Y_1]\,\mathbb{E}[Y_2] + \mathbb{E}[Y_2^2] \text{ by independence}$$

$$=$$

What is average squared distance between two rolls on a six-sided die?

$$\mathbb{E}[(Y_1 - Y_2)^2] = \mathbb{E}[Y_1^2 - 2Y_1Y_2 + Y_2^2] = \mathbb{E}[Y_1^2] - \mathbb{E}[2Y_1Y_2] + \mathbb{E}[Y_2^2]$$

$$= \mathbb{E}[Y_1^2] - 2\,\mathbb{E}[Y_1]\,\mathbb{E}[Y_2] + \mathbb{E}[Y_2^2] \text{ by independence}$$

$$= \mathbb{E}[Y^2] - 2\,\mathbb{E}[Y]\,\mathbb{E}[Y] + \mathbb{E}[Y^2] =$$

What is average squared distance between two rolls on a six-sided die?

$$\mathbb{E}[(Y_1 - Y_2)^2] = \mathbb{E}[Y_1^2 - 2Y_1Y_2 + Y_2^2] = \mathbb{E}[Y_1^2] - \mathbb{E}[2Y_1Y_2] + \mathbb{E}[Y_2^2]$$

$$= \mathbb{E}[Y_1^2] - 2\,\mathbb{E}[Y_1]\,\mathbb{E}[Y_2] + \mathbb{E}[Y_2^2] \text{ by independence}$$

$$= \mathbb{E}[Y^2] - 2\,\mathbb{E}[Y]\,\mathbb{E}[Y] + \mathbb{E}[Y^2] = 2\,\mathbb{E}[Y^2] - 2\,\mathbb{E}[Y]^2 = 2\mathrm{Var}(Y)$$

Rules for Variance

- $\text{Var}(c) = 0$
- $\text{Var}(cY) = c^2\text{Var}(Y)$
- $\text{Var}(Y + c) = \text{Var}(Y)$

Rules for Variance

- $\mathrm{Var}(c) = 0$
- $\mathrm{Var}(cY) = c^2\mathrm{Var}(Y)$
- $\mathrm{Var}(Y + c) = \mathrm{Var}(Y)$

In addition, if Y_1 and Y_2 are independent, then
$\mathrm{Var}(Y_1 + Y_2) = \mathrm{Var}(Y_1) + \mathrm{Var}(Y_2)$

Suppose Y is a continuous random variable with mean μ and variance σ^2, and a and b are constants. Solve for the expectation and variance of $(aY + b)$.

$$\mathbb{E}(aY + b) =$$

Suppose Y is a continuous random variable with mean μ and variance σ^2, and a and b are constants. Solve for the expectation and variance of $(aY + b)$.

$$\mathbb{E}(aY + b) = \mathbb{E}(aY) + \mathbb{E}(b) =$$

Suppose Y is a continuous random variable with mean μ and variance σ^2, and a and b are constants. Solve for the expectation and variance of $(aY + b)$.

$$\mathbb{E}(aY + b) = \mathbb{E}(aY) + \mathbb{E}(b) = a \cdot \mathbb{E}(y) + \mathbb{E}(b) = a\mu + b$$

Suppose Y is a continuous random variable with mean μ and variance σ^2, and a and b are constants. Solve for the expectation and variance of $(aY + b)$.

$$\mathbb{E}(aY + b) = \mathbb{E}(aY) + \mathbb{E}(b) = a \cdot \mathbb{E}(y) + \mathbb{E}(b) = a\mu + b$$

$$\begin{aligned} \mathrm{Var}(aY + b) &= \mathrm{Var}(aY) \\ &= a^2 \mathrm{Var}(Y) \\ &= a^2\sigma^2 \end{aligned}$$

Variance of the Binomial Mean

Suppose that $p = 0.51$ of likely voters support a candidate. You randomly survey N people, and each time the chances of getting a supporter are

Variance of the Binomial Mean

Suppose that $p = 0.51$ of likely voters support a candidate. You randomly survey N people, and each time the chances of getting a supporter are $Y_i \sim \text{Bernoulli}(p)$.

- What is the expectation of support for a single respondent?

Variance of the Binomial Mean

Suppose that $p = 0.51$ of likely voters support a candidate. You randomly survey N people, and each time the chances of getting a supporter are $Y_i \sim \text{Bernoulli}(p)$.

- What is the expectation of support for a single respondent? $\mathbb{E}(Y_i) = 0 \times (1 - p) + 1 \times p = p = .51$

Variance of the Binomial Mean

Suppose that $p = 0.51$ of likely voters support a candidate. You randomly survey N people, and each time the chances of getting a supporter are $Y_i \sim \text{Bernoulli}(p)$.

- What is the expectation of support for a single respondent?
  $\mathbb{E}(Y_i) = 0 \times (1 - p) + 1 \times p = p = .51$
- The expected proportion of respondents supporting?

Variance of the Binomial Mean

Suppose that $p = 0.51$ of likely voters support a candidate. You randomly survey N people, and each time the chances of getting a supporter are $Y_i \sim \text{Bernoulli}(p)$.

- What is the expectation of support for a single respondent?
  $\mathbb{E}(Y_i) = 0 \times (1-p) + 1 \times p = p = .51$
- The expected proportion of respondents supporting?
  $\mathbb{E}(\frac{1}{N}\sum_{i=1}^{N} Y_i) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}(Y_i) = \frac{N}{N} \cdot \mathbb{E}(Y_i) = p = .51$
- The standard deviation of support for a single respondent?

Variance of the Binomial Mean

Suppose that $p = 0.51$ of likely voters support a candidate. You randomly survey N people, and each time the chances of getting a supporter are $Y_i \sim \text{Bernoulli}(p)$.

- What is the expectation of support for a single respondent?
  $\mathbb{E}(Y_i) = 0 \times (1 - p) + 1 \times p = p = .51$
- The expected proportion of respondents supporting?
  $\mathbb{E}(\frac{1}{N}\sum_{i=1}^{N} Y_i) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}(Y_i) = \frac{N}{N} \cdot \mathbb{E}(Y_i) = p = .51$
- The standard deviation of support for a single respondent?
  First, note that $0^2 = 0$ and $1^2 = 1$, so $Y_i^2 = Y_i$.

Variance of the Binomial Mean

Suppose that $p = 0.51$ of likely voters support a candidate. You randomly survey N people, and each time the chances of getting a supporter are $Y_i \sim \text{Bernoulli}(p)$.

- What is the expectation of support for a single respondent?
  $\mathbb{E}(Y_i) = 0 \times (1 - p) + 1 \times p = p = .51$
- The expected proportion of respondents supporting?
  $\mathbb{E}(\frac{1}{N}\sum_{i=1}^{N} Y_i) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(Y_i) = \frac{N}{N} \cdot \mathbb{E}(Y_i) = p = .51$
- The standard deviation of support for a single respondent?
  First, note that $0^2 = 0$ and $1^2 = 1$, so $Y_i^2 = Y_i$. Then
  $\mathrm{Var}(Y_i) =$

Variance of the Binomial Mean

Suppose that $p = 0.51$ of likely voters support a candidate. You randomly survey N people, and each time the chances of getting a supporter are $Y_i \sim \text{Bernoulli}(p)$.

- What is the expectation of support for a single respondent?
  $\mathbb{E}(Y_i) = 0 \times (1 - p) + 1 \times p = p = .51$
- The expected proportion of respondents supporting?
  $\mathbb{E}(\frac{1}{N} \sum_{i=1}^{N} Y_i) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}(Y_i) = \frac{N}{N} \cdot \mathbb{E}(Y_i) = p = .51$
- The standard deviation of support for a single respondent?
  First, note that $0^2 = 0$ and $1^2 = 1$, so $Y_i^2 = Y_i$. Then
  $\mathrm{Var}(Y_i) = \mathbb{E}(Y_i^2) - \mathbb{E}(Y_i)^2 =$

Variance of the Binomial Mean

Suppose that $p = 0.51$ of likely voters support a candidate. You randomly survey N people, and each time the chances of getting a supporter are $Y_i \sim \text{Bernoulli}(p)$.

- What is the expectation of support for a single respondent?
  $\mathbb{E}(Y_i) = 0 \times (1-p) + 1 \times p = p = .51$
- The expected proportion of respondents supporting?
  $\mathbb{E}(\frac{1}{N}\sum_{i=1}^{N} Y_i) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(Y_i) = \frac{N}{N}\cdot\mathbb{E}(Y_i) = p = .51$
- The standard deviation of support for a single respondent?
  First, note that $0^2 = 0$ and $1^2 = 1$, so $Y_i^2 = Y_i$. Then
  $\mathrm{Var}(Y_i) = \mathbb{E}(Y_i^2) - \mathbb{E}(Y_i)^2 = \mathbb{E}(Y_i) - \mathbb{E}(Y_i)^2 = p - p^2 = p(1-p)$,
  so $\sigma = \sqrt{p(1-p)}$.

Variance of the Binomial Mean

Suppose that $p = 0.51$ of likely voters support a candidate. You randomly survey N people, and each time the chances of getting a supporter are $Y_i \sim \text{Bernoulli}(p)$.

- What is the expectation of support for a single respondent?
  $\mathbb{E}(Y_i) = 0 \times (1-p) + 1 \times p = p = .51$
- The expected proportion of respondents supporting?
  $\mathbb{E}(\frac{1}{N}\sum_{i=1}^N Y_i) = \frac{1}{N}\sum_{i=1}^N \mathbb{E}(Y_i) = \frac{N}{N} \cdot \mathbb{E}(Y_i) = p = .51$
- The standard deviation of support for a single respondent?
  First, note that $0^2 = 0$ and $1^2 = 1$, so $Y_i^2 = Y_i$. Then
  $\text{Var}(Y_i) = \mathbb{E}(Y_i^2) - \mathbb{E}(Y_i)^2 = \mathbb{E}(Y_i) - \mathbb{E}(Y_i)^2 = p - p^2 = p(1-p)$,
  so $\sigma = \sqrt{p(1-p)}$.
- The standard deviation for the proportion of respondents supporting?

Variance of the Binomial Mean

Suppose that $p = 0.51$ of likely voters support a candidate. You randomly survey N people, and each time the chances of getting a supporter are $Y_i \sim \text{Bernoulli}(p)$.

- What is the expectation of support for a single respondent?
  $\mathbb{E}(Y_i) = 0 \times (1-p) + 1 \times p = p = .51$
- The expected proportion of respondents supporting?
  $\mathbb{E}(\frac{1}{N}\sum_{i=1}^{N} Y_i) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(Y_i) = \frac{N}{N}\cdot\mathbb{E}(Y_i) = p = .51$
- The standard deviation of support for a single respondent?
  First, note that $0^2 = 0$ and $1^2 = 1$, so $Y_i^2 = Y_i$. Then
  $\text{Var}(Y_i) = \mathbb{E}(Y_i^2) - \mathbb{E}(Y_i)^2 = \mathbb{E}(Y_i) - \mathbb{E}(Y_i)^2 = p - p^2 = p(1-p)$,
  so $\sigma = \sqrt{p(1-p)}$.
- The standard deviation for the proportion of respondents supporting?
  $\text{Var}(\frac{1}{N}\sum_{i=1}^{N} Y_i) =$

Variance of the Binomial Mean

Suppose that $p = 0.51$ of likely voters support a candidate. You randomly survey N people, and each time the chances of getting a supporter are $Y_i \sim \text{Bernoulli}(p)$.

- What is the expectation of support for a single respondent? $\mathbb{E}(Y_i) = 0 \times (1-p) + 1 \times p = p = .51$
- The expected proportion of respondents supporting? $\mathbb{E}(\frac{1}{N}\sum_{i=1}^{N} Y_i) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(Y_i) = \frac{N}{N}\cdot\mathbb{E}(Y_i) = p = .51$
- The standard deviation of support for a single respondent? First, note that $0^2 = 0$ and $1^2 = 1$, so $Y_i^2 = Y_i$. Then $\text{Var}(Y_i) = \mathbb{E}(Y_i^2) - \mathbb{E}(Y_i)^2 = \mathbb{E}(Y_i) - \mathbb{E}(Y_i)^2 = p - p^2 = p(1-p)$, so $\sigma = \sqrt{p(1-p)}$.
- The standard deviation for the proportion of respondents supporting? $\text{Var}(\frac{1}{N}\sum_{i=1}^{N} Y_i) = \frac{1}{N^2}\text{Var}(\sum_{i=1}^{N} Y_i) =$

Variance of the Binomial Mean

Suppose that $p = 0.51$ of likely voters support a candidate. You randomly survey N people, and each time the chances of getting a supporter are $Y_i \sim \text{Bernoulli}(p)$.

- What is the expectation of support for a single respondent? $\mathbb{E}(Y_i) = 0 \times (1-p) + 1 \times p = p = .51$
- The expected proportion of respondents supporting? $\mathbb{E}(\frac{1}{N}\sum_{i=1}^{N} Y_i) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}(Y_i) = \frac{N}{N} \cdot \mathbb{E}(Y_i) = p = .51$
- The standard deviation of support for a single respondent? First, note that $0^2 = 0$ and $1^2 = 1$, so $Y_i^2 = Y_i$. Then $\text{Var}(Y_i) = \mathbb{E}(Y_i^2) - \mathbb{E}(Y_i)^2 = \mathbb{E}(Y_i) - \mathbb{E}(Y_i)^2 = p - p^2 = p(1-p)$, so $\sigma = \sqrt{p(1-p)}$.
- The standard deviation for the proportion of respondents supporting? $\text{Var}(\frac{1}{N}\sum_{i=1}^{N} Y_i) = \frac{1}{N^2}\text{Var}(\sum_{i=1}^{N} Y_i) = \frac{1}{N^2}\sum_{i=1}^{N} \text{Var}(Y_i) =$

Variance of the Binomial Mean

Suppose that $p = 0.51$ of likely voters support a candidate. You randomly survey N people, and each time the chances of getting a supporter are $Y_i \sim \text{Bernoulli}(p)$.

- What is the expectation of support for a single respondent?
  $\mathbb{E}(Y_i) = 0 \times (1 - p) + 1 \times p = p = .51$
- The expected proportion of respondents supporting?
  $\mathbb{E}(\frac{1}{N}\sum_{i=1}^N Y_i) = \frac{1}{N}\sum_{i=1}^N \mathbb{E}(Y_i) = \frac{N}{N} \cdot \mathbb{E}(Y_i) = p = .51$
- The standard deviation of support for a single respondent?
  First, note that $0^2 = 0$ and $1^2 = 1$, so $Y_i^2 = Y_i$. Then
  $\text{Var}(Y_i) = \mathbb{E}(Y_i^2) - \mathbb{E}(Y_i)^2 = \mathbb{E}(Y_i) - \mathbb{E}(Y_i)^2 = p - p^2 = p(1 - p)$,
  so $\sigma = \sqrt{p(1 - p)}$.
- The standard deviation for the proportion of respondents supporting?
  $\text{Var}(\frac{1}{N}\sum_{i=1}^N Y_i) = \frac{1}{N^2}\text{Var}(\sum_{i=1}^N Y_i) = \frac{1}{N^2}\sum_{i=1}^N \text{Var}(Y_i) =$
  $\frac{1}{N^2}\sum_{i=1}^N p(1 - p) =$

Variance of the Binomial Mean

Suppose that $p = 0.51$ of likely voters support a candidate. You randomly survey N people, and each time the chances of getting a supporter are $Y_i \sim \text{Bernoulli}(p)$.

- What is the expectation of support for a single respondent?
  $\mathbb{E}(Y_i) = 0 \times (1-p) + 1 \times p = p = .51$
- The expected proportion of respondents supporting?
  $\mathbb{E}(\frac{1}{N}\sum_{i=1}^{N} Y_i) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(Y_i) = \frac{N}{N}\cdot\mathbb{E}(Y_i) = p = .51$
- The standard deviation of support for a single respondent?
  First, note that $0^2 = 0$ and $1^2 = 1$, so $Y_i^2 = Y_i$. Then
  $\text{Var}(Y_i) = \mathbb{E}(Y_i^2) - \mathbb{E}(Y_i)^2 = \mathbb{E}(Y_i) - \mathbb{E}(Y_i)^2 = p - p^2 = p(1-p)$,
  so $\sigma = \sqrt{p(1-p)}$.
- The standard deviation for the proportion of respondents supporting?
  $\text{Var}(\frac{1}{N}\sum_{i=1}^{N} Y_i) = \frac{1}{N^2}\text{Var}(\sum_{i=1}^{N} Y_i) = \frac{1}{N^2}\sum_{i=1}^{N}\text{Var}(Y_i) =$
  $\frac{1}{N^2}\sum_{i=1}^{N} p(1-p) = \frac{1}{N}p(1-p)$, and the standard deviation is

Variance of the Binomial Mean

Suppose that $p = 0.51$ of likely voters support a candidate. You randomly survey N people, and each time the chances of getting a supporter are $Y_i \sim \text{Bernoulli}(p)$.

- What is the expectation of support for a single respondent? $\mathbb{E}(Y_i) = 0 \times (1-p) + 1 \times p = p = .51$
- The expected proportion of respondents supporting? $\mathbb{E}(\frac{1}{N}\sum_{i=1}^{N} Y_i) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}(Y_i) = \frac{N}{N}\cdot\mathbb{E}(Y_i) = p = .51$
- The standard deviation of support for a single respondent? First, note that $0^2 = 0$ and $1^2 = 1$, so $Y_i^2 = Y_i$. Then $\text{Var}(Y_i) = \mathbb{E}(Y_i^2) - \mathbb{E}(Y_i)^2 = \mathbb{E}(Y_i) - \mathbb{E}(Y_i)^2 = p - p^2 = p(1-p)$, so $\sigma = \sqrt{p(1-p)}$.
- The standard deviation for the proportion of respondents supporting? $\text{Var}(\frac{1}{N}\sum_{i=1}^{N} Y_i) = \frac{1}{N^2}\text{Var}(\sum_{i=1}^{N} Y_i) = \frac{1}{N^2}\sum_{i=1}^{N}\text{Var}(Y_i) = \frac{1}{N^2}\sum_{i=1}^{N} p(1-p) = \frac{1}{N}p(1-p)$, and the standard deviation is $\sqrt{\frac{p(1-p)}{N}}$.

Illustration: Causality with Potential Outcomes

Definition (Treatment)

D_i: Indicator of treatment intake for *unit* i, where $i = 1, ..., N$

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ received the treatment} \\ 0 & \text{otherwise} \end{cases}$$

Definition (Observed Outcome)

Y_i: Variable of interest whose value may be affected by the treatment

Definition (Potential Outcomes)

Y_{di}: Value of the outcome that *would* be realized if unit i received the treatment d, where $d = 0$ or 1

$$Y_{di} = \begin{cases} Y_{1i} & \text{Potential outcome for unit } i \text{ with treatment} \\ Y_{0i} & \text{Potential outcome for unit } i \text{ without treatment} \end{cases}$$

Causality with Potential Outcomes

Definition (Causal Effect, or Unit Treatment Effect)

Causal effect of the treatment on the outcome for unit i is the difference between its two potential outcomes:

$$\tau_i = Y_{1i} - Y_{0i}$$

Causality with Potential Outcomes

Definition (Causal Effect, or Unit Treatment Effect)

Causal effect of the treatment on the outcome for unit i is the difference between its two potential outcomes:

$$\tau_i = Y_{1i} - Y_{0i}$$

Note that observed outcomes are realized from potential outcomes as

$$Y_i = Y_{D_i i} = D_i Y_{1i} + (1 - D_i) Y_{0i} \quad \text{so} \quad Y_i = \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases}$$

Causality with Potential Outcomes

Definition (Causal Effect, or Unit Treatment Effect)

Causal effect of the treatment on the outcome for unit i is the difference between its two potential outcomes:

$$\tau_i = Y_{1i} - Y_{0i}$$

Note that observed outcomes are realized from potential outcomes as

$$Y_i = Y_{D_i i} = D_i Y_{1i} + (1 - D_i) Y_{0i} \quad \text{so} \quad Y_i = \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases}$$

Fundamental Problem of Causal Inference (Holland 1986):

We can never observe both Y_{1i} and Y_{0i} for the same i

This makes τ_i unidentifiable without further assumptions.

Causal Quantities of Interest, or Estimands

- Unit-level causal effects are fundamentally unobservable
- We instead focus on *averages* in most situations

Causal Quantities of Interest, or Estimands

- Unit-level causal effects are fundamentally unobservable
- We instead focus on *averages* in most situations

Definition (Average treatment effect, ATE)

$$\tau_{ATE} = \frac{1}{N}\sum_{i=1}^{N}\{Y_{1i} - Y_{0i}\}$$

or equivalently

$$\tau_{ATE} = \mathbb{E}[Y_{1i} - Y_{0i}]$$

- Note that τ_{ATE} is still unidentified
- In the rest of this course, we will consider various assumptions under which τ_{ATE} can be identified from observed information

Other Causal Estimands

Definition (Average treatment effect on the treated, ATT)

$$\tau_{ATT} = \frac{1}{N_1} \sum_{i=1}^{N} D_i \{Y_{1i} - Y_{0i}\} \quad \text{where} \quad N_1 = \sum_{i=1}^{N} D_i$$

or equivalently $\tau_{ATT} = \mathbb{E}[Y_{1i} - Y_{0i} | D_i = 1]$

- In words, N_1 equals

Other Causal Estimands

Definition (Average treatment effect on the treated, ATT)

$$\tau_{ATT} = \frac{1}{N_1} \sum_{i=1}^{N} D_i \left\{ Y_{1i} - Y_{0i} \right\} \quad \text{where} \quad N_1 = \sum_{i=1}^{N} D_i$$

or equivalently $\tau_{ATT} \;=\; \mathbb{E}[Y_{1i} - Y_{0i} | D_i = 1]$

- In words, N_1 equals the number of treated units.

Other Causal Estimands

Definition (Average treatment effect on the treated, ATT)

$$\tau_{ATT} = \frac{1}{N_1} \sum_{i=1}^{N} D_i \{Y_{1i} - Y_{0i}\} \quad \text{where} \quad N_1 = \sum_{i=1}^{N} D_i$$

or equivalently $\tau_{ATT} \;=\; \mathbb{E}[Y_{1i} - Y_{0i} | D_i = 1]$

- In words, N_1 equals the number of treated units.
- When would $\tau_{ATT} \neq \tau_{ATE}$?

Other Causal Estimands

Definition (Average treatment effect on the treated, ATT)

$$\tau_{ATT} = \frac{1}{N_1} \sum_{i=1}^{N} D_i \{Y_{1i} - Y_{0i}\} \quad \text{where} \quad N_1 = \sum_{i=1}^{N} D_i$$

or equivalently $\tau_{ATT} \; = \; \mathbb{E}[Y_{1i} - Y_{0i} | D_i = 1]$

- In words, N_1 equals the number of treated units.
- When would $\tau_{ATT} \neq \tau_{ATE}$? When D_i and Y_{di} are associated.
- Exercise: Define τ_{ATC}, ATE on the untreated (control) units.

Other Causal Estimands

Definition (Average treatment effect on the treated, ATT)

$$\tau_{ATT} = \frac{1}{N_1} \sum_{i=1}^{N} D_i \{Y_{1i} - Y_{0i}\} \quad \text{where} \quad N_1 = \sum_{i=1}^{N} D_i$$

or equivalently $\tau_{ATT} \; = \; \mathbb{E}[Y_{1i} - Y_{0i} | D_i = 1]$

- In words, N_1 equals the number of treated units.
- When would $\tau_{ATT} \neq \tau_{ATE}$? When D_i and Y_{di} are associated.
- Exercise: Define τ_{ATC}, ATE on the untreated (control) units.

Definition (Conditional average treatment effects)

$$\tau_{CATE}(x) = \mathbb{E}[Y_{1i} - Y_{0i} | X_i = x]$$

where X_i is a pre-treatment covariate for unit i

- In words, $\tau_{CATE}(x)$ is

Other Causal Estimands

Definition (Average treatment effect on the treated, ATT)

$$\tau_{ATT} = \frac{1}{N_1} \sum_{i=1}^{N} D_i \left\{ Y_{1i} - Y_{0i} \right\} \quad \text{where} \quad N_1 = \sum_{i=1}^{N} D_i$$

or equivalently $\tau_{ATT} = \mathbb{E}[Y_{1i} - Y_{0i} | D_i = 1]$

- In words, N_1 equals the number of treated units.
- When would $\tau_{ATT} \neq \tau_{ATE}$? When D_i and Y_{di} are associated.
- Exercise: Define τ_{ATC}, ATE on the untreated (control) units.

Definition (Conditional average treatment effects)

$$\tau_{CATE}(x) = \mathbb{E}[Y_{1i} - Y_{0i} | X_i = x]$$

where X_i is a pre-treatment covariate for unit i

- In words, $\tau_{CATE}(x)$ is a subgroup effect, treatment effect on units who have particular characteristics x.

Illustration: Average Treatment Effect

Suppose we observe a population of 4 units:

i	D_i	Y_i
1	1	3
2	1	1
3	0	0
4	0	1

What is $\tau_{ATE} = \mathbb{E}[Y_{1i} - Y_{0i}]$?

Illustration: Average Treatment Effect

Suppose we observe a population of 4 units:

i	D_i	Y_i
1	1	3
2	1	1
3	0	0
4	0	1
$\mathbb{E}[Y_i \mid D_i = 1]$		2
$\mathbb{E}[Y_i \mid D_i = 0]$		0.5
$\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0]$		1.5

What is $\tau_{ATE} = \mathbb{E}[Y_{1i} - Y_{0i}]$?

Naïve estimator:

$$\begin{aligned} \tilde{\tau} &= \mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0] \quad \text{(observed difference in means)} \\ &= \frac{3+1}{2} - \frac{0+1}{2} = 1.5 \quad \text{Could this be wrong?} \end{aligned}$$

Illustration: Average Treatment Effect

Suppose we observe a population of 4 units:

i	D_i	Y_i	Y_{1i}	Y_{0i}	τ_i
1	1	3	3	?	?
2	1	1	1	?	?
3	0	0	?	0	?
4	0	1	?	1	?

What is $\tau_{ATE} = \mathbb{E}[Y_{1i} - Y_{0i}]$? We need potential outcomes that we do not observe!

Illustration: Average Treatment Effect

Suppose we observe a population of 4 units:

i	D_i	Y_i	Y_{1i}	Y_{0i}	τ_i
1	1	3	3	0	?
2	1	1	1	1	?
3	0	0	1	0	?
4	0	1	1	1	?

Suppose hypothetically: $Y_{01} = 0$, $Y_{02} = Y_{13} = Y_{14} = 1$.

Illustration: Average Treatment Effect

Suppose we observe a population of 4 units:

i	D_i	Y_i	Y_{1i}	Y_{0i}	τ_i
1	1	3	3	0	3
2	1	1	1	1	0
3	0	0	1	0	1
4	0	1	1	1	0
$\mathbb{E}[Y_{1i}]$			1.5		
$\mathbb{E}[Y_{0i}]$				0.5	
$\mathbb{E}[Y_{1i} - Y_{0i}]$					1

$$\tau_{ATE} = \mathbb{E}[Y_{1i} - Y_{0i}] = \mathbb{E}[\tau_i] = \frac{3+0+1+0}{4} = 1.$$

Illustration: Average Treatment Effect

Suppose we observe a population of 4 units:

i	D_i	Y_i	Y_{1i}	Y_{0i}	τ_i
1	1	3	3	0	3
2	1	1	1	1	0
3	0	0	1	0	1
4	0	1	1	1	0
$\mathbb{E}[Y_{1i}]$			1.5		
$\mathbb{E}[Y_{0i}]$				0.5	
$\mathbb{E}[Y_{1i} - Y_{0i}]$					1

$$\tau_{ATE} = \mathbb{E}[Y_{1i} - Y_{0i}] = \mathbb{E}[\tau_i] = \frac{3 + 0 + 1 + 0}{4} = 1.$$

Why $\tau_{ATE} \neq \tilde{\tau}$? When would they be equal?

Illustration: Average Treatment Effect on the Treated

Again suppose we observe a population of 4 units:

i	D_i	Y_i	Y_{1i}	Y_{0i}	τ_i
1	1	3	3	?	?
2	1	1	1	?	?
3	0	0	?	0	?
4	0	1	?	1	?

What is $\tau_{ATT} = \mathbb{E}[Y_{1i} - Y_{0i} \mid D_i = 1]$?

Illustration: Average Treatment Effect on the Treated

Again suppose we observe a population of 4 units:

i	D_i	Y_i	Y_{1i}	Y_{0i}	τ_i
1	1	3	3	0	3
2	1	1	1	1	0
3	0	0	1	0	1
4	0	1	1	1	0

What is $\tau_{ATT} = \mathbb{E}[Y_{1i} - Y_{0i} \mid D_i = 1]$?

Illustration: Average Treatment Effect on the Treated

Again suppose we observe a population of 4 units:

i	D_i	Y_i	Y_{1i}	Y_{0i}	τ_i
1	1	3	3	0	3
2	1	1	1	1	0
3	0	0	1	0	1
4	0	1	1	1	0
$\mathbb{E}[Y_{1i} \mid D_i = 1]$			2		
$\mathbb{E}[Y_{0i} \mid D_i = 1]$				0.5	
$\mathbb{E}[Y_{1i} - Y_{0i} \mid D_i = 1]$					1.5

$$\tau_{ATT} = \mathbb{E}[Y_{1i} - Y_{0i} \mid D_i = 1] = \mathbb{E}[\tau_i \mid D_i = 1] = \frac{3+0}{2} = 1.5.$$

Illustration: Average Treatment Effect on the Treated

Again suppose we observe a population of 4 units:

i	D_i	Y_i	Y_{1i}	Y_{0i}	τ_i
1	1	3	3	0	3
2	1	1	1	1	0
3	0	0	1	0	1
4	0	1	1	1	0
$\mathbb{E}[Y_{1i} \mid D_i = 1]$			2		
$\mathbb{E}[Y_{0i} \mid D_i = 1]$				0.5	
$\mathbb{E}[Y_{1i} - Y_{0i} \mid D_i = 1]$					1.5

$$\tau_{ATT} = \mathbb{E}[Y_{1i} - Y_{0i} \mid D_i = 1] = \mathbb{E}[\tau_i \mid D_i = 1] = \frac{3+0}{2} = 1.5.$$

Why does $\tau_{ATT} \neq \tau_{ATE}$?