Let's Guess!

Dollar or not with equal probability?

Guess how much you get!

Guess a 1/2!

The expected value.

Win X, 100 times.

How much will you win the 101st.

Guess average!

Let's Guess!

How much does random person weigh?

Guess the expected value!

How much does professor Rao weigh?

Remember: I am pretty tall!

Knowing that I am tall should you guess he is heavier than expected!
Let’s Guess!

Dollar or not with equal probability?

Guess how much you get!

Guess a 1/2!

The expected value.

Win $X, 100 times.

How much will you win the 101st.

Guess average!

Let’s Guess!

How much does random person weigh?

Guess the expected value!

How much does professor Rao weigh?

Remember: I am pretty tall!

Knowing that I am tall should you guess he is heavier than expected!
Let’s Guess!
Dollar or not with equal probability?
Let’s Guess!
   Dollar or not with equal probability?
   Guess how much you get!
Let’s Guess!
Dollar or not with equal probability?
   Guess how much you get!
Guess a 1/2!
Let’s Guess!
Dollar or not with equal probability?
Guess how much you get!
Guess a 1/2! The expected value.
Let’s Guess!
Dollar or not with equal probability?
Guess how much you get!
Guess a 1/2! The expected value.
Win X, 100 times.
Let’s Guess!
Dollar or not with equal probability?
   Guess how much you get!
Guess a 1/2! The expected value.
Win X, 100 times.
   How much will you win the 101st.
Let’s Guess!
Dollar or not with equal probability?
  Guess how much you get!
Guess a 1/2! The expected value.
Win X, 100 times.
  How much will you win the 101st.
Guess average!
Let’s Guess!
   Dollar or not with equal probability?
   Guess how much you get!
Guess a 1/2! The expected value.
Win X, 100 times.
   How much will you win the 101st.
Guess average!

Let’s Guess!
Let’s Guess!
Dollar or not with equal probability?
   Guess how much you get!
Guess a 1/2! The expected value.
Win X, 100 times.
   How much will you win the 101st.
Guess average!

Let’s Guess!
How much does random person weigh?
Let’s Guess!
Dollar or not with equal probability?
    Guess how much you get!
Guess a 1/2! The expected value.
Win X, 100 times.
    How much will you win the 101st.
Guess average!

Let’s Guess!
How much does random person weigh?
    Guess the expected value!
Let’s Guess!
   Dollar or not with equal probability?
       Guess how much you get!
   Guess a 1/2! The expected value.
   Win X, 100 times.
       How much will you win the 101st.
   Guess average!

Let’s Guess!
   How much does random person weigh?
       Guess the expected value!
   How much does professor Rao weigh?
Let’s Guess!
Dollar or not with equal probability?
  Guess how much you get!
Guess a 1/2! The expected value.
Win X, 100 times.
  How much will you win the 101st.
Guess average!

Let’s Guess!
How much does random person weigh?
  Guess the expected value!
How much does professor Rao weigh?
  Remember: I am pretty tall!
Let's Guess!
  Dollar or not with equal probability?
  Guess how much you get!
Guess a 1/2! The expected value.
Win X, 100 times.
  How much will you win the 101st.
Guess average!

Let's Guess!
  How much does random person weigh?
  Guess the expected value!
How much does professor Rao weigh?
  Remember: I am pretty tall!

Knowing that I am tall should you guess he is heavier than expected!
Previously: Single variable.

When do you get an accurate measure of a random variable.

Predictor: Expectation.

Accuracy: Variance.

Want to find expectation? Poll.

Sampling: Many trials and average.

Accuracy: Chebyshev.

Chernoff.

Today:

What does the value of one variable tell you about another?

Exact: Conditional probability among all events.

Summary: Covariance.

Predictor: Linear function.

Bayesian: Best linear estimator from covariance, and expectations.

Sampling: Linear regression from set of samples.
Previously: Single variable. When do you get an accurate measure
Previously: Single variable.
When do you get an accurate measure of a random variable.
Previously: Single variable.
When do you get an accurate measure of a random variable.
Predictor:
Previously: Single variable.
When do you get an accurate measure of a random variable.
Predictor: Expectation.
Previously: Single variable.
When do you get an accurate measure of a random variable.
Predictor: Expectation.
Accuracy:
Previously: Single variable.
   When do you get an accurate measure of a random variable.
   Predictor: Expectation.
   Accuracy: Variance.

Today:
   What does the value of one variable tell you about another?
   Exact: Conditional probability among all events.
   Summary: Covariance.
   Predictor: Linear function.
   Bayesian: Best linear estimator from covariance, and expectations.
   Sampling: Linear regression from set of samples.
Previously: Single variable.
When do you get an accurate measure of a random variable.
Predictor: Expectation.
Accuracy: Variance.
Want to find expectation?
Previously: Single variable.
When do you get an accurate measure of a random variable.
Predictor: Expectation.
Accuracy: Variance.
Want to find expectation? Poll.
Previously: Single variable.
When do you get an accurate measure of a random variable.
Predictor: Expectation.
Accuracy: Variance.
Want to find expectation? Poll.
Sampling: Many trials and average.
Previously: Single variable.
When do you get an accurate measure of a random variable.
Predictor: Expectation.
Accuracy: Variance.
Want to find expectation? Poll.
Sampling: Many trials and average.
Accuracy: Chebyshev.
Previously: Single variable.
When do you get an accurate measure of a random variable.
Predictor: Expectation.
Accuracy: Variance.
Want to find expectation? Poll.
Sampling: Many trials and average.
Accuracy: Chebyshev. Chernoff.
Previously: Single variable.
When do you get an accurate measure of a random variable.
Predictor: Expectation.
Accuracy: Variance.
Want to find expectation? Poll.
Sampling: Many trials and average.
Accuracy: Chebyshev. Chernoff.
Previously: Single variable.
When do you get an accurate measure of a random variable.
Predictor: Expectation.
Accuracy: Variance.
Want to find expectation? Poll.
Sampling: Many trials and average.
Accuracy: Chebyshev. Chernoff.

Today:
Previously: Single variable.
When do you get an accurate measure of a random variable.
Predictor: Expectation.
Accuracy: Variance.
Want to find expectation? Poll.
Sampling: Many trials and average.
Accuracy: Chebyshev. Chernoff.

Today:
Previously: Single variable.
When do you get an accurate measure of a random variable.
Predictor: Expectation.
Accuracy: Variance.
Want to find expectation? Poll.
Sampling: Many trials and average.
Accuracy: Chebyshev. Chernoff.

Today:
What does the value of one variable tell you about another?
Previously: Single variable.
  When do you get an accurate measure of a random variable.
  Predictor: Expectation.
  Accuracy: Variance.
  Want to find expectation? Poll.
  Sampling: Many trials and average.
  Accuracy: Chebyshev. Chernoff.

Today:

What does the value of one variable tell you about another?
  Exact: Conditional probability among all events.
Previously: Single variable.
   When do you get an accurate measure of a random variable.
   Predictor: Expectation.
   Accuracy: Variance.
   Want to find expectation? Poll.
   Sampling: Many trials and average.
   Accuracy: Chebyshev. Chernoff.

Today:

What does the value of one variable tell you about another?
   Exact: Conditional probability among all events.
   Summary: Covariance.
Previously: Single variable.
    When do you get an accurate measure of a random variable.
Predictor: Expectation.
Accuracy: Variance.
Want to find expectation? Poll.
Sampling: Many trials and average.
Accuracy: Chebyshev. Chernoff.

Today:

What does the value of one variable tell you about another?
    Exact: Conditional probability among all events.
Summary: Covariance.
Predictor: Linear function.
Previously: Single variable.
When do you get an accurate measure of a random variable.
Predictor: Expectation.
Accuracy: Variance.
Want to find expectation? Poll.
Sampling: Many trials and average.
Accuracy: Chebyshev. Chernoff.

Today:
What does the value of one variable tell you about another?
Exact: Conditional probability among all events.
Summary: Covariance.
Predictor: Linear function.
Bayesian: Best linear estimator from covariance, and expectations.
Previously: Single variable.
  When do you get an accurate measure of a random variable.
  Predictor: Expectation.
  Accuracy: Variance.
  Want to find expectation? Poll.
  Sampling: Many trials and average.
  Accuracy: Chebyshev. Chernoff.

Today:

What does the value of one variable tell you about another?
  Exact: Conditional probability among all events.
  Summary: Covariance.
  Predictor: Linear function.
    Bayesion: Best linear estimator from covariance, and expectations.
    Sampling: Linear regression from set of samples.
Previously: Single variable.
When do you get an accurate measure of a random variable.
Predictor: Expectation.
Accuracy: Variance.
Want to find expectation? Poll.
Sampling: Many trials and average.
Accuracy: Chebyshev. Chernoff.

Today:
What does the value of one variable tell you about another?
Exact: Conditional probability among all events.
Summary: Covariance.
Predictor: Linear function.
Bayesian: Best linear estimator from covariance, and expectations.
Sampling: Linear regression from set of samples.
Previously: Single variable.
When do you get an accurate measure of a random variable.
Predictor: Expectation.
Accuracy: Variance.
Want to find expectation? Poll.
Sampling: Many trials and average.
Accuracy: Chebyshev. Chernoff.

Today:

What does the value of one variable tell you about another?
   Exact: Conditional probability among all events.
   Summary: Covariance.
   Predictor: Linear function.
      Bayesion: Best linear estimator from covariance, and expectations.
      Sampling: Linear regression from set of samples.
Linear Regression
1. Examples
2. History
3. Multiple Random variables
4. Linear Regression
5. Derivation
6. More examples
Illustrative Example

Example 1: 100 people.

Let \((X_n, Y_n)\) = (height, weight) of person \(n\), for \(n = 1, \ldots, 100\):

The blue line is \(Y = -114.3 + 106.5X\). (\(X\) in meters, \(Y\) in kg.)

Best linear fit: Linear Regression.

Should you really use a linear function? Cubic, maybe. Then logHeight and logWeight is linear.
Illustrative Example

Example 1: 100 people.

Let \((X_n, Y_n) = (\text{height, weight})\) of person \(n\), for \(n = 1, \ldots, 100\):
Illustrative Example

Example 1: 100 people.

Let \((X_n, Y_n) = \text{(height, weight)}\) of person \(n\), for \(n = 1, \ldots, 100:\)

The blue line is \(Y = -114.3 + 106.5 \cdot X\). (\(X\) in meters, \(Y\) in kg.)

Best linear fit: Linear Regression.

Should you really use a linear function? Cubic, maybe. Then \(\log(\text{Height})\) and \(\log(\text{Weight})\) is linear.
Illustrative Example

Example 1: 100 people.
Let \((X_n, Y_n) = \text{(height, weight)}\) of person \(n\), for \(n = 1, \ldots, 100\):

The blue line is \(Y = -114.3 + 106.5X\). (\(X\) in meters, \(Y\) in kg.)
Illustrative Example

Example 1: 100 people.

Let \((X_n, Y_n) = \text{(height, weight)}\) of person \(n\), for \(n = 1, \ldots, 100\):

The blue line is \(Y = -114.3 + 106.5X\). (\(X\) in meters, \(Y\) in kg.)

Best linear fit: **Linear Regression**.
Illustrative Example

Example 1: 100 people.

Let $(X_n, Y_n) = \text{(height, weight)}$ of person $n$, for $n = 1, \ldots, 100$:

The blue line is $Y = -114.3 + 106.5X$. ($X$ in meters, $Y$ in kg.)

Best linear fit: **Linear Regression**.

Should you really use a linear function?
Illustrative Example

Example 1: 100 people.

Let \((X_n, Y_n) = (\text{height, weight})\) of person \(n\), for \(n = 1, \ldots, 100\):

The blue line is \(Y = -114.3 + 106.5X\). (\(X\) in meters, \(Y\) in kg.)

Best linear fit: Linear Regression.

Should you really use a linear function? Cubic, maybe.
Illustrative Example

Example 1: 100 people.

Let \((X_n, Y_n) = \text{(height, weight)}\) of person \(n\), for \(n = 1, \ldots, 100\):

The blue line is \(Y = -114.3 + 106.5X\). (\(X\) in meters, \(Y\) in kg.)

Best linear fit: **Linear Regression**.

Should you really use a linear function? Cubic, maybe.

Then \(\log\text{Height}\) and \(\log\text{Weight}\) is linear.
Painful Example

Midterm 1 v Midterm 2.

\[ Y = 0.97X - 1.54 \]
Painful Example

Midterm 1 v Midterm 2.
$Y = 0.97X - 1.54$

Midterm 2 v Midterm 3
Painful Example

Midterm 1 v Midterm 2.
\[ Y = 0.97X - 1.54 \]

Midterm 2 v Midterm 3
\[ Y = 0.67X + 6.08 \]
Illustrative Example: sample space.

Example 3: 15 people.
Illustrative Example: sample space.

Example 3: 15 people.
We look at two attributes: $(X_n, Y_n)$ of person $n$, for $n = 1, \ldots, 15$: 

The line $Y = a + bX$ is the linear regression.
Illustrative Example: sample space.

Example 3: 15 people.

We look at two attributes: \((X_n, Y_n)\) of person \(n\), for \(n = 1, \ldots, 15\):
Illustrative Example: sample space.

Example 3: 15 people.

We look at two attributes: \((X_n, Y_n)\) of person \(n\), for \(n = 1, \ldots, 15\):

The line \(Y = a + bX\) is the linear regression.
Galton produced over 340 papers and books. He created the statistical concept of correlation. In an effort to reach a wider audience, Galton worked on a novel entitled *Kantsaywhere*. The novel described a utopia organized by a eugenic religion, designed to breed fitter and smarter humans. The lesson is that smart people can also be stupid.

**Sir Francis Galton**

- **Born**: 16 February 1822
  - Birmingham, England
- **Died**: 17 January 1911 (aged 88)
  - Haslemere, Surrey, England
Galton produced over 340 papers and books. He created the statistical concept of correlation.
Galton produced over 340 papers and books. He created the statistical concept of correlation.

In an effort to reach a wider audience, Galton worked on a novel entitled Kantsaywhere.
Galton produced over 340 papers and books. He created the statistical concept of correlation.

In an effort to reach a wider audience, Galton worked on a novel entitled Kantsaywhere. The novel described a utopia organized by a eugenic religion, designed to breed fitter and smarter humans.
Galton produced over 340 papers and books. He created the statistical concept of correlation.

In an effort to reach a wider audience, Galton worked on a novel entitled Kantsaywhere. The novel described a utopia organized by a eugenic religion, designed to breed fitter and smarter humans.

The lesson is that smart people can also be stupid.
Multiple Random Variables

The pair \((X, Y)\) takes 6 different values with the probabilities shown. This figure specifies the joint distribution of \(X\) and \(Y\).

Questions: Where is \(\Omega\)? What are \(X(\omega)\) and \(Y(\omega)\)?

Answer: For instance, let \(\Omega\) be the set of values of \((X, Y)\) and assign them the corresponding probabilities. This is the "canonical" probability space.
Multiple Random Variables

The pair \((X, Y)\) takes 6 different values with the probabilities shown.

The figure specifies the joint distribution of \(X\) and \(Y\).

Questions: Where is \(\Omega\)? What are \(X(\omega)\) and \(Y(\omega)\)?

Answer: For instance, let \(\Omega\) be the set of values of \((X, Y)\) and assign them the corresponding probabilities. This is the "canonical" probability space.
Multiple Random Variables

The pair \((X, Y)\) takes 6 different values with the probabilities shown. This figure specifies the joint distribution of \(X\) and \(Y\).

**Questions**: Where is \(\Omega\)? What are \(X(\omega)\) and \(Y(\omega)\)?

**Answer**: For instance, let \(\Omega\) be the set of values of \((X, Y)\) and assign them the corresponding probabilities. This is the “canonical” probability space.
Multiple Random Variables

The pair \((X, Y)\) takes 6 different values with the probabilities shown. This figure specifies the joint distribution of \(X\) and \(Y\).

Questions: Where is \(\Omega\)? What are \(X(\omega)\) and \(Y(\omega)\)?
Multiple Random Variables

The pair \((X, Y)\) takes 6 different values with the probabilities shown. This figure specifies the joint distribution of \(X\) and \(Y\).

Questions: Where is \(\Omega\)? What are \(X(\omega)\) and \(Y(\omega)\)?

Answer: For instance, let \(\Omega\) be the set of values of \((X, Y)\) and assign them the corresponding probabilities.
Multiple Random Variables

The pair \((X, Y)\) takes 6 different values with the probabilities shown. This figure specifies the joint distribution of \(X\) and \(Y\).

Questions: Where is \(\Omega\)? What are \(X(\omega)\) and \(Y(\omega)\)?

Answer: For instance, let \(\Omega\) be the set of values of \((X, Y)\) and assign them the corresponding probabilities. This is the “canonical” probability space.
Definitions Let $X$ and $Y$ be RVs on $\Omega$. 
Definitions

Let $X$ and $Y$ be RVs on $\Omega$.

- **Joint Distribution:** $Pr[X = x, Y = y]$
**Definitions** Let $X$ and $Y$ be RVs on $\Omega$.

- **Joint Distribution:** $Pr[X = x, Y = y]$
- **Marginal Distribution:** $Pr[X = x] = \sum_y Pr[X = x, Y = y]$
Definitions

Let $X$ and $Y$ be RVs on $\Omega$.

- **Joint Distribution:** $Pr[X = x, Y = y]$
- **Marginal Distribution:** $Pr[X = x] = \sum_y Pr[X = x, Y = y]$
- **Conditional Distribution:** $Pr[Y = y | X = x] = \frac{Pr[X = x, Y = y]}{Pr[X = x]}$
**Definitions**

Let $X$ and $Y$ be RVs on $\Omega$.

- **Joint Distribution**: $Pr[X = x, Y = y]$
- **Marginal Distribution**: $Pr[X = x] = \sum_y Pr[X = x, Y = y]$
- **Conditional Distribution**: $Pr[Y = y | X = x] = \frac{Pr[X = x, Y = y]}{Pr[X = x]}$
Marginal and Conditional

\[ Pr[X = 1] = 0.05 + 0.15 = 0.2 \]

\[ Pr[X = 2] = 0.4 \]

\[ Pr[X = 3] = 0.45 \]

This is the marginal distribution of \( X \):

\[ Pr[X = x] = \sum y Pr[X = x, Y = y] \]

\[ Pr[Y = 1 | X = 1] = \frac{Pr[X = 1, Y = 1]}{Pr[X = 1]} = \frac{0.05}{0.2} = 0.25 \]

This is the conditional distribution of \( Y \) given \( X = 1 \):

\[ Pr[Y = y | X = x] = \frac{Pr[X = x, Y = y]}{Pr[X = x]} \]

Quick question: Are \( X \) and \( Y \) independent?
Marginal and Conditional

\[ \Pr[X = 1] = 0.05 + 0.15 = 0.20 \]

Quick question: Are \( X \) and \( Y \) independent?
Marginal and Conditional

$Pr[X = 1] = 0.05 + 0.1 = 0.15$;
Marginal and Conditional

$Pr[X = 1] = 0.05 + 0.1 = 0.15; Pr[X = 2] =$
Marginal and Conditional

$Pr[X = 1] = 0.05 + 0.1 = 0.15; Pr[X = 2] = 0.4;$
Marginal and Conditional

\[ \Pr[X = 1] = 0.05 + 0.1 = 0.15; \Pr[X = 2] = 0.4; \Pr[X = 3] = \]

- \[ \Pr[Y = 1 | X = 1] = \frac{\Pr[X = 1, Y = 1]}{\Pr[X = 1]} = \frac{0.05}{0.15} = \frac{1}{3} \]

This is the marginal distribution of \( X \):

\[ \Pr[X = x] = \sum_y \Pr[X = x, Y = y] \]

This is the conditional distribution of \( Y \) given \( X = 1 \):

\[ \Pr[Y = y | X = x] = \frac{\Pr[X = x, Y = y]}{\Pr[X = x]} \]

Quick question: Are \( X \) and \( Y \) independent?
Marginal and Conditional

- $Pr[X = 1] = 0.05 + 0.1 = 0.15$; $Pr[X = 2] = 0.4$; $Pr[X = 3] = 0.45$. 

Quick question: Are $X$ and $Y$ independent?
Marginal and Conditional

\[ \text{Pr}[X = 1] = 0.05 + 0.1 = 0.15; \text{Pr}[X = 2] = 0.4; \text{Pr}[X = 3] = 0.45. \]

\[ \text{This is the marginal distribution of } X: \]
Marginal and Conditional

- $Pr[X = 1] = 0.05 + 0.1 = 0.15; Pr[X = 2] = 0.4; Pr[X = 3] = 0.45$.

- This is the **marginal distribution** of $X$: 
  $Pr[X = x] = \sum_y Pr[X = x, Y = y]$. 
Marginal and Conditional

- \( \text{Pr}[X = 1] = 0.05 + 0.1 = 0.15; \text{Pr}[X = 2] = 0.4; \text{Pr}[X = 3] = 0.45. \)
- This is the marginal distribution of \( X \): 
  \[ \text{Pr}[X = x] = \sum_{y} \text{Pr}[X = x, Y = y]. \]
- \( \text{Pr}[Y = 1|X = 1] = \)
Marginal and Conditional

- $Pr[X = 1] = 0.05 + 0.1 = 0.15$; $Pr[X = 2] = 0.4$; $Pr[X = 3] = 0.45$.
- This is the marginal distribution of $X$:
  $Pr[X = x] = \sum_y Pr[X = x, Y = y]$.
- $Pr[Y = 1|X = 1] = Pr[X = 1, Y = 1]/Pr[X = 1]$
Marginal and Conditional

- $Pr[X = 1] = 0.05 + 0.1 = 0.15$; $Pr[X = 2] = 0.4$; $Pr[X = 3] = 0.45$.
- This is the **marginal distribution** of $X$: $Pr[X = x] = \sum_y Pr[X = x, Y = y]$.
- $Pr[Y = 1|X = 1] = Pr[X = 1, Y = 1]/Pr[X = 1] = 0.05/0.15 = 1/3$. 

Quick question: Are $X$ and $Y$ independent?
Marginal and Conditional

- $\Pr[X = 1] = 0.05 + 0.1 = 0.15; \Pr[X = 2] = 0.4; \Pr[X = 3] = 0.45.$
- This is the marginal distribution of $X$: $\Pr[X = x] = \sum_y \Pr[X = x, Y = y].$
- $\Pr[Y = 1|X = 1] = \Pr[X = 1, Y = 1]/\Pr[X = 1] = 0.05/0.15 = 1/3.$
- This is the conditional distribution of $Y$ given $X = 1$: 
Marginal and Conditional

- \( Pr[X = 1] = 0.05 + 0.1 = 0.15; \ Pr[X = 2] = 0.4; \ Pr[X = 3] = 0.45. \)

- This is the marginal distribution of \( X \):
\[
Pr[X = x] = \sum_y Pr[X = x, Y = y].
\]

- \( Pr[Y = 1|X = 1] = Pr[X = 1, Y = 1]/Pr[X = 1] = 0.05/0.15 = 1/3. \)

- This is the conditional distribution of \( Y \) given \( X = 1 \):
\[
Pr[Y = y|X = x] =
\]

Quick question: Are \( X \) and \( Y \) independent?
Marginal and Conditional

- $\Pr[X = 1] = 0.05 + 0.1 = 0.15$; $\Pr[X = 2] = 0.4$; $\Pr[X = 3] = 0.45$.

- This is the marginal distribution of $X$: $\Pr[X = x] = \sum_y \Pr[X = x, Y = y]$.

- $\Pr[Y = 1|X = 1] = \Pr[X = 1, Y = 1]/\Pr[X = 1] = 0.05/0.15 = 1/3$.

- This is the conditional distribution of $Y$ given $X = 1$: $\Pr[Y = y|X = x] = \Pr[X = x, Y = y]/\Pr[X = x]$. 
Marginal and Conditional

- $Pr[X = 1] = 0.05 + 0.1 = 0.15; Pr[X = 2] = 0.4; Pr[X = 3] = 0.45$.
- This is the **marginal distribution** of $X$: 
  $Pr[X = x] = \sum_y Pr[X = x, Y = y]$.

- $Pr[Y = 1|X = 1] = Pr[X = 1, Y = 1]/Pr[X = 1] = 0.05/0.15 = 1/3$.
- This is the **conditional distribution** of $Y$ given $X = 1$: 
  $Pr[Y = y|X = x] = Pr[X = x, Y = y]/Pr[X = x]$.

Quick question:
Marginal and Conditional

- \( Pr[X = 1] = 0.05 + 0.1 = 0.15; \) \( Pr[X = 2] = 0.4; \) \( Pr[X = 3] = 0.45. \)

- This is the marginal distribution of \( X \):
  \[ Pr[X = x] = \sum_y Pr[X = x, Y = y]. \]

- \( Pr[Y = 1|X = 1] = Pr[X = 1, Y = 1]/Pr[X = 1] = 0.05/0.15 = 1/3. \)

- This is the conditional distribution of \( Y \) given \( X = 1 \):
  \[ Pr[Y = y|X = x] = Pr[X = x, Y = y]/Pr[X = x]. \]

Quick question: Are \( X \) and \( Y \) independent?
Covariance

**Definition** The covariance of $X$ and $Y$ is

$$\text{cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

Quick Question: For independent $X$ and $Y$, $\text{cov}(X, Y) =$

1. $0$

Proof:

$$\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY - \mathbb{E}[X]Y - X\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y]] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$
Covariance

**Definition** The covariance of $X$ and $Y$ is

$$\text{cov}(X, Y) := E[(X - E[X])(Y - E[Y])].$$

**Fact**

$$\text{cov}(X, Y) = E[XY] - E[X]E[Y].$$
Covariance

Definition The covariance of $X$ and $Y$ is

$$cov(X, Y) := E[(X - E[X])(Y - E[Y])].$$

Fact

$$cov(X, Y) = E[XY] - E[X]E[Y].$$

Quick Question: For independent $X$ and $Y$,

$$cov(X, Y) =$$

?
Covariance

Definition The covariance of $X$ and $Y$ is

$$\text{cov}(X, Y) := E[(X - E[X])(Y - E[Y])].$$

Fact

$$\text{cov}(X, Y) = E[XY] - E[X]E[Y].$$

Quick Question: For independent $X$ and $Y$,

$$\text{cov}(X, Y) = \ ? \ ?$$
Covariance

**Definition** The covariance of $X$ and $Y$ is

$$cov(X, Y) := E[(X - E[X])(Y - E[Y])].$$

**Fact**

$$cov(X, Y) = E[XY] - E[X]E[Y].$$

Quick Question: For independent $X$ and $Y$,

$$cov(X, Y) =$$

? 1 ? 0?
Covariance

**Definition** The covariance of $X$ and $Y$ is

$$\text{cov}(X, Y) := E[(X - E[X])(Y - E[Y])].$$

**Fact**

$$\text{cov}(X, Y) = E[XY] - E[X]E[Y].$$

Quick Question: For independent $X$ and $Y$,

$$\text{cov}(X, Y) = \ ? 1 \ ? 0?$$

**Proof:**

Covariance

Definition The covariance of $X$ and $Y$ is

$$\text{cov}(X, Y) := E[(X - E[X])(Y - E[Y])].$$

Fact

$$\text{cov}(X, Y) = E[XY] - E[X]E[Y].$$

Quick Question: For independent $X$ and $Y$,

$$\text{cov}(X, Y) =$$

1 or 0?

Proof:


Covariance

**Definition** The covariance of $X$ and $Y$ is

$$\text{cov}(X, Y) := E[(X - E[X])(Y - E[Y])].$$

**Fact**

$$\text{cov}(X, Y) = E[XY] - E[X]E[Y].$$

Quick Question: For independent $X$ and $Y$,

$$\text{cov}(X, Y) = \ ? \ ? \ 0?$$

**Proof:**


$$= E[XY] - E[X]E[Y].$$
Examples of Covariance

Note that \( E[X] = 0 \) and \( E[Y] = 0 \) in these examples. Then \( \text{cov}(X, Y) = E[XY] \).

When \( \text{cov}(X, Y) > 0 \), the RVs \( X \) and \( Y \) tend to be large or small together.

When \( \text{cov}(X, Y) < 0 \), when \( X \) is larger, \( Y \) tends to be smaller.

\[
\begin{align*}
\text{Four equally likely pairs of values} \\
\text{cov}(X, Y) = 1/2 \\
\text{cov}(X, Y) = -1/2 \\
\text{cov}(X, Y) = 0
\end{align*}
\]
Examples of Covariance

Note that $E[X] = 0$ and $E[Y] = 0$ in these examples. Then $cov(X, Y) = E[XY]$. 
Examples of Covariance

Note that $E[X] = 0$ and $E[Y] = 0$ in these examples. Then $\text{cov}(X, Y) = E[XY]$.

When $\text{cov}(X, Y) > 0$, the RVs $X$ and $Y$ tend to be large or small together.
Examples of Covariance

Note that $E[X] = 0$ and $E[Y] = 0$ in these examples. Then $\text{cov}(X, Y) = E[XY]$.

When $\text{cov}(X, Y) > 0$, the RVs $X$ and $Y$ tend to be large or small together.

When $\text{cov}(X, Y) < 0$, when $X$ is larger, $Y$ tends to be smaller.
Examples of Covariance

\[ E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 1.9 \]

\[ E[X^2] = 1^2 \times 0.15 + 2^2 \times 0.4 + 3^2 \times 0.45 = 5.8 \]

\[ E[Y] = 1 \times 0.2 + 2 \times 0.6 + 3 \times 0.2 = 2 \]

\[ E[XY] = 1 \times 0.05 + 2 \times 0.15 + \cdots + 3 \times 0.25 = 4.85 \]

\[ \text{cov}(X, Y) = E[XY] - E[X]E[Y] = 4.85 - 1.9 \times 2 = 0.95 \]

\[ \text{var}[X] = E[X^2] - E[X]^2 = 5.8 - 1.9^2 = 2.19 \]
Examples of Covariance

\[ E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 1.9 \]
Examples of Covariance

\[ E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 1.9 \]

\[ E[X^2] = 1^2 \times 0.15 + 2^2 \times 0.4 + 3^2 \times 0.45 = 5.8 \]
Examples of Covariance

\[ E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 1.9 \]
\[ E[X^2] = 1^2 \times 0.15 + 2^2 \times 0.4 + 3^2 \times 0.45 = 5.8 \]
\[ E[Y] = 1 \times 0.2 + 2 \times 0.6 + 3 \times 0.2 = 2 \]
Examples of Covariance

\[ E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 1.9 \]
\[ E[X^2] = 1^2 \times 0.15 + 2^2 \times 0.4 + 3^2 \times 0.45 = 5.8 \]
\[ E[Y] = 1 \times 0.2 + 2 \times 0.6 + 3 \times 0.2 = 2 \]
\[ E[XY] = 1 \times 0.05 + 1 \times 2 \times 0.1 + \cdots + 3 \times 3 \times 0.2 = 4.85 \]
Examples of Covariance

\[ E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 1.9 \]
\[ E[X^2] = 1^2 \times 0.15 + 2^2 \times 0.4 + 3^2 \times 0.45 = 5.8 \]
\[ E[Y] = 1 \times 0.2 + 2 \times 0.6 + 3 \times 0.2 = 2 \]
\[ E[XY] = 1 \times 0.05 + 1 \times 2 \times 0.1 + \cdots + 3 \times 3 \times 0.2 = 4.85 \]
\[ \text{cov}(X, Y) = E[XY] - E[X]E[Y] = 1.05 \]
Examples of Covariance

\[ E[X] = 1 \times 0.15 + 2 \times 0.4 + 3 \times 0.45 = 1.9 \]
\[ E[X^2] = 1^2 \times 0.15 + 2^2 \times 0.4 + 3^2 \times 0.45 = 5.8 \]
\[ E[Y] = 1 \times 0.2 + 2 \times 0.6 + 3 \times 0.2 = 2 \]
\[ E[XY] = 1 \times 0.05 + 1 \times 2 \times 0.1 + \cdots + 3 \times 3 \times 0.2 = 4.85 \]
\[ \text{cov}(X, Y) = E[XY] - E[X]E[Y] = 1.05 \]
\[ \text{var}[X] = E[X^2] - E[X]^2 = 2.19. \]
Properties of Covariance

Properties of Covariance


Fact
(a) \( \text{var}[X] = \text{cov}(X, X) \)
Properties of Covariance


**Fact**
(a) \( \text{var}[X] = \text{cov}(X, X) \)
(b) \( X, Y \) independent \( \Rightarrow \) \( \text{cov}(X, Y) = \)
Properties of Covariance

\[
\]

Fact
(a) \( \text{var}[X] = \text{cov}(X, X) \)
(b) \( X, Y \) independent \( \Rightarrow \) \( \text{cov}(X, Y) = 0 \)
Properties of Covariance

\[
\]

**Fact**
(a) \( \text{var}[X] = \text{cov}(X, X) \)
(b) \( X, Y \) independent \( \Rightarrow \) \( \text{cov}(X, Y) = 0 \)
(c) \( \text{cov}(a + X, b + Y) = \text{cov}(X, Y) \)
Properties of Covariance


**Fact**

(a) \( \text{var}[X] = \text{cov}(X, X) \)

(b) \( X, Y \) independent \( \Rightarrow \text{cov}(X, Y) = 0 \)

(c) \( \text{cov}(a + X, b + Y) = \text{cov}(X, Y) \)

(d) \( \text{cov}(aX + bY, cU + dV) = ac \cdot \text{cov}(X, U) + ad \cdot \text{cov}(X, V) \)
\[ + \quad bc \cdot \text{cov}(Y, U) + bd \cdot \text{cov}(Y, V). \]
Properties of Covariance

\[
\]

**Fact**
(a) \( \text{var}[X] = \text{cov}(X, X) \)
(b) \( X, Y \) independent \( \Rightarrow \) \( \text{cov}(X, Y) = 0 \)
(c) \( \text{cov}(a + X, b + Y) = \text{cov}(X, Y) \)
(d) \( \text{cov}(aX + bY, cU + dV) = ac \cdot \text{cov}(X, U) + ad \cdot \text{cov}(X, V) \\
+ bc \cdot \text{cov}(Y, U) + bd \cdot \text{cov}(Y, V). \)

**Proof:**
(a)-(b)-(c) are obvious.
Properties of Covariance


Fact
(a) \( \text{var}[X] = \text{cov}(X, X) \)
(b) \( X, Y \) independent \( \Rightarrow \) \( \text{cov}(X, Y) = 0 \)
(c) \( \text{cov}(a + X, b + Y) = \text{cov}(X, Y) \)
(d) \( \text{cov}(aX + bY, cU + dV) = ac \cdot \text{cov}(X, U) + ad \cdot \text{cov}(X, V) \)
\[ + bc \cdot \text{cov}(Y, U) + bd \cdot \text{cov}(Y, V). \]

Proof:
(a)-(b)-(c) are obvious.
(d) In view of (c), one can subtract the means and assume that the RVs are zero-mean.
Properties of Covariance


**Fact**

(a) \( \text{var}[X] = \text{cov}(X, X) \)
(b) \( X, Y \) independent \( \Rightarrow \) \( \text{cov}(X, Y) = 0 \)
(c) \( \text{cov}(a + X, b + Y) = \text{cov}(X, Y) \)
(d) \( \text{cov}(aX + bY, cU + dV) = ac \cdot \text{cov}(X, U) + ad \cdot \text{cov}(X, V) \)
\[ + bc \cdot \text{cov}(Y, U) + bd \cdot \text{cov}(Y, V). \]

**Proof:**

(a)-(b)-(c) are obvious.
(d) In view of (c), one can subtract the means and assume that the RVs are zero-mean. Then,

\[ \text{cov}(aX + bY, cU + dV) = E[(aX + bY)(cU + dV)] \]
Properties of Covariance


**Fact**
(a) \( \text{var}[X] = \text{cov}(X, X) \)
(b) \( X, Y \) independent \( \Rightarrow \) \( \text{cov}(X, Y) = 0 \)
(c) \( \text{cov}(a + X, b + Y) = \text{cov}(X, Y) \)
(d) \( \text{cov}(aX + bY, cU + dV) = ac \cdot \text{cov}(X, U) + ad \cdot \text{cov}(X, V) \)
\[ + bc \cdot \text{cov}(Y, U) + bd \cdot \text{cov}(Y, V). \]

**Proof:**
(a)-(b)-(c) are obvious.
(d) In view of (c), one can subtract the means and assume that the RVs are zero-mean. Then,
\[ \text{cov}(aX + bY, cU + dV) = E[(aX + bY)(cU + dV)] \]
\[ = ac \cdot E[XU] + ad \cdot E[XV] + bc \cdot E[YU] + bd \cdot E[YV]. \]
Properties of Covariance


Fact
(a) \( \text{var}[X] = \text{cov}(X, X) \)
(b) \( X, Y \) independent \( \Rightarrow \) \( \text{cov}(X, Y) = 0 \)
(c) \( \text{cov}(a + X, b + Y) = \text{cov}(X, Y) \)
(d) \( \text{cov}(aX + bY, cU + dV) = ac \cdot \text{cov}(X, U) + ad \cdot \text{cov}(X, V) \)
\[ + bc \cdot \text{cov}(Y, U) + bd \cdot \text{cov}(Y, V). \]

Proof:
(a)-(b)-(c) are obvious.
(d) In view of (c), one can subtract the means and assume that the RVs are zero-mean. Then,

\[ \text{cov}(aX + bY, cU + dV) = E[(aX + bY)(cU + dV)] \]
\[ = ac \cdot E[XU] + ad \cdot E[XV] + bc \cdot E[YU] + bd \cdot E[YV] \]
\[ = ac \cdot \text{cov}(X, U) + ad \cdot \text{cov}(X, V) + bc \cdot \text{cov}(Y, U) + bd \cdot \text{cov}(Y, V). \]
Properties of Covariance


Fact
(a) var[X] = cov(X, X)
(b) X, Y independent ⇒ cov(X, Y) = 0
(c) cov(a + X, b + Y) = cov(X, Y)
(d) cov(aX + bY, cU + dV) = ac · cov(X, U) + ad · cov(X, V)
   + bc · cov(Y, U) + bd · cov(Y, V).

Proof:
(a)-(b)-(c) are obvious.
(d) In view of (c), one can subtract the means and assume that the RVs are zero-mean. Then,

\[
cov(aX + bY, cU + dV) = E[(aX + bY)(cU + dV)]
= ac \cdot E[XU] + ad \cdot E[XV] + bc \cdot E[YU] + bd \cdot E[YV]
= ac \cdot cov(X, U) + ad \cdot cov(X, V) + bc \cdot cov(Y, U) + bd \cdot cov(Y, V).
\]
Linear Regression: Non-Bayesian

Definition
Given the samples \( \{(X_n, Y_n), n = 1, \ldots, N\} \),
Linear Regression: Non-Bayesian

Definition
Given the samples \( \{(X_n, Y_n), n = 1, \ldots, N\} \), the Linear Regression of \( Y \) over \( X \) is

\[
\hat{Y} = a + bX
\]
Linear Regression: Non-Bayesian

Definition
Given the samples \( \{(X_n, Y_n), n = 1, \ldots, N\} \), the Linear Regression of \( Y \) over \( X \) is

\[
\hat{Y} = a + bX
\]

where \((a, b)\) minimize

\[
\sum_{n=1}^{N} (Y_n - a - bX_n)^2 .
\]
Linear Regression: Non-Bayesian

**Definition**

Given the samples $\{(X_n, Y_n), n = 1, \ldots, N\}$, the Linear Regression of $Y$ over $X$ is

$$
\hat{Y} = a + bX
$$

where $(a, b)$ minimize

$$
\sum_{n=1}^{N} (Y_n - a - bX_n)^2.
$$

Thus, $\hat{Y}_n = a + bX_n$ is our guess about $Y_n$ given $X_n$. 

The squared error is $(Y_n - \hat{Y}_n)^2$. The LR minimizes the sum of the squared errors.

Why the squares and not the absolute values? Main justification: much easier!

Note: This is a non-Bayesian formulation: there is no prior.

Single Variable: Average minimizes squared distance to sample points.
Linear Regression: Non-Bayesian

Definition
Given the samples \( \{(X_n, Y_n), n = 1, \ldots, N\} \), the Linear Regression of \( Y \) over \( X \) is

\[
\hat{Y} = a + bX
\]

where \((a, b)\) minimize

\[
\sum_{n=1}^{N} (Y_n - a - bX_n)^2.
\]

Thus, \( \hat{Y}_n = a + bX_n \) is our guess about \( Y_n \) given \( X_n \). The squared error is \((Y_n - \hat{Y}_n)^2\).
Linear Regression: Non-Bayesian

Definition
Given the samples \( \{(X_n, Y_n), n = 1, \ldots, N\} \), the Linear Regression of \( Y \) over \( X \) is

\[ \hat{Y} = a + bX \]

where \((a, b)\) minimize

\[ \sum_{n=1}^{N} (Y_n - a - bX_n)^2. \]

Thus, \( \hat{Y}_n = a + bX_n \) is our guess about \( Y_n \) given \( X_n \). The squared error is \( (Y_n - \hat{Y}_n)^2 \). The LR minimizes the sum of the squared errors.
Linear Regression: Non-Bayesian

Definition
Given the samples \( \{(X_n, Y_n), n = 1, \ldots, N\} \), the Linear Regression of \( Y \) over \( X \) is

\[
\hat{Y} = a + bX
\]

where \((a, b)\) minimize

\[
\sum_{n=1}^{N} (Y_n - a - bX_n)^2.
\]

Thus, \( \hat{Y}_n = a + bX_n \) is our guess about \( Y_n \) given \( X_n \). The squared error is \((Y_n - \hat{Y}_n)^2\). The LR minimizes the sum of the squared errors. Why the squares and not the absolute values?
Linear Regression: Non-Bayesian

Definition
Given the samples \( \{(X_n, Y_n), n = 1, \ldots, N\} \), the Linear Regression of \( Y \) over \( X \) is

\[
\hat{Y} = a + bX
\]

where \((a, b)\) minimize

\[
\sum_{n=1}^{N} (Y_n - a - bX_n)^2.
\]

Thus, \( \hat{Y}_n = a + bX_n \) is our guess about \( Y_n \) given \( X_n \). The squared error is \( (Y_n - \hat{Y}_n)^2 \). The LR minimizes the sum of the squared errors.

Why the squares and not the absolute values?
Main justification: much easier!
Linear Regression: Non-Bayesian

Definition
Given the samples \( \{(X_n, Y_n), n = 1, \ldots, N\} \), the Linear Regression of \( Y \) over \( X \) is

\[
\hat{Y} = a + bX
\]

where \( (a, b) \) minimize

\[
\sum_{n=1}^{N} (Y_n - a - bX_n)^2.
\]

Thus, \( \hat{Y}_n = a + bX_n \) is our guess about \( Y_n \) given \( X_n \). The squared error is \( (Y_n - \hat{Y}_n)^2 \). The LR minimizes the sum of the squared errors.

Why the squares and not the absolute values?
Main justification: much easier!

Note: This is a non-Bayesian formulation: there is no prior.
Linear Regression: Non-Bayesian

Definition
Given the samples \( \{(X_n, Y_n), n = 1, \ldots, N\} \), the Linear Regression of \( Y \) over \( X \) is

\[
\hat{Y} = a + bX
\]

where \((a, b)\) minimize

\[
\sum_{n=1}^{N} (Y_n - a - bX_n)^2.
\]

Thus, \( \hat{Y}_n = a + bX_n \) is our guess about \( Y_n \) given \( X_n \). The squared error is \((Y_n - \hat{Y}_n)^2\). The LR minimizes the sum of the squared errors.

Why the squares and not the absolute values?
Main justification: much easier!

Note: This is a non-Bayesian formulation: there is no prior.

Single Variable: Average minimizes squared distance to sample points.
Linear Least Squares Estimate

**Definition**
Given two RVs $X$ and $Y$ with known distribution $Pr[X = x, Y = y]$,

$\hat{Y} = a + bX =: L[Y|X]$ where $(a, b)$ minimize $E[(Y - a - bX)^2]$.

Thus, $\hat{Y} = a + bX$ is our guess about $Y$ given $X$.

The squared error is $(Y - \hat{Y})^2$.

The LLSE minimizes the expected value of the squared error.

Why the squares and not the absolute values? Main justification: much easier!

Note: This is a Bayesian formulation: there is a prior.

Single Variable: $E(X)$ minimizes expected squared error.
Linear Least Squares Estimate

**Definition**
Given two RVs $X$ and $Y$ with known distribution $Pr[X = x, Y = y]$, the **Linear Least Squares Estimate** of $Y$ given $X$ is

$$\hat{Y} = a + bX =: L[Y|X]$$

Thus, $\hat{Y} = a + bX$ is our guess about $Y$ given $X$.

The squared error is $(Y - \hat{Y})^2$.

The LLSE minimizes the expected value of the squared error.

Why the squares and not the absolute values?
Main justification: much easier!

Note: This is a Bayesian formulation: there is a prior.

Single Variable: $E(X)$ minimizes expected squared error.
Linear Least Squares Estimate

Definition
Given two RVs $X$ and $Y$ with known distribution $Pr[X = x, Y = y]$, the Linear Least Squares Estimate of $Y$ given $X$ is

$$\hat{Y} = a + bX =: L[Y|X]$$

where $(a, b)$ minimize

$$E[(Y - a - bX)^2].$$
Linear Least Squares Estimate

Definition
Given two RVs $X$ and $Y$ with known distribution $Pr[X = x, Y = y]$, the Linear Least Squares Estimate of $Y$ given $X$ is

$$\hat{Y} = a + bX =: L[Y|X]$$

where $(a, b)$ minimize

$$E[(Y - a - bX)^2].$$

Thus, $\hat{Y} = a + bX$ is our guess about $Y$ given $X$. 
Linear Least Squares Estimate

**Definition**
Given two RVs $X$ and $Y$ with known distribution $Pr[X = x, Y = y]$, the Linear Least Squares Estimate of $Y$ given $X$ is

$$\hat{Y} = a + bX =: L[Y|X]$$

where $(a, b)$ minimize

$$E[(Y - a - bX)^2].$$

Thus, $\hat{Y} = a + bX$ is our guess about $Y$ given $X$. The squared error is $(Y - \hat{Y})^2$. 

Why the squares and not the absolute values?
Main justification: much easier!

Note: This is a Bayesian formulation: there is a prior.

Single Variable: $E(X)$ minimizes expected squared error.
Linear Least Squares Estimate

Definition
Given two RVs $X$ and $Y$ with known distribution $Pr[X = x, Y = y]$, the Linear Least Squares Estimate of $Y$ given $X$ is

$$\hat{Y} = a + bX =: L[Y|X]$$

where $(a, b)$ minimize

$$E[(Y - a - bX)^2].$$

Thus, $\hat{Y} = a + bX$ is our guess about $Y$ given $X$.

The squared error is $(Y - \hat{Y})^2$.
The LLSE minimizes the expected value of the squared error.
**Linear Least Squares Estimate**

**Definition**
Given two RVs $X$ and $Y$ with known distribution $Pr[X = x, Y = y]$, the **Linear Least Squares Estimate** of $Y$ given $X$ is

$$
\hat{Y} = a + bX =: L[Y|X]
$$

where $(a, b)$ minimize

$$
E[(Y - a - bX)^2].
$$

Thus, $\hat{Y} = a + bX$ is our guess about $Y$ given $X$.
The squared error is $(Y - \hat{Y})^2$.
The LLSE minimizes the expected value of the squared error.

Why the squares and not the absolute values?
Linear Least Squares Estimate

Definition
Given two RVs $X$ and $Y$ with known distribution $Pr[X = x, Y = y]$, the Linear Least Squares Estimate of $Y$ given $X$ is

$$\hat{Y} = a + bX =: L[Y|X]$$

where $(a, b)$ minimize

$$E[(Y - a - bX)^2].$$

Thus, $\hat{Y} = a + bX$ is our guess about $Y$ given $X$. The squared error is $(Y - \hat{Y})^2$. The LLSE minimizes the expected value of the squared error.

Why the squares and not the absolute values? Main justification: much easier!
**Definition**
Given two RVs $X$ and $Y$ with known distribution $Pr[X = x, Y = y]$, the **Linear Least Squares Estimate** of $Y$ given $X$ is

$$
\hat{Y} = a + bX =: L[Y|X]
$$

where $(a, b)$ minimize

$$
E[(Y - a - bX)^2].
$$

Thus, $\hat{Y} = a + bX$ is our guess about $Y$ given $X$. The squared error is $(Y - \hat{Y})^2$. The LLSE minimizes the expected value of the squared error.

Why the squares and not the absolute values? Main justification: much easier!

Note: This is a **Bayesian** formulation: there is a prior.
Linear Least Squares Estimate

**Definition**
Given two RVs $X$ and $Y$ with known distribution $Pr[X = x, Y = y]$, the Linear Least Squares Estimate of $Y$ given $X$ is

$$
\hat{Y} = a + bX =: L[Y|X]
$$

where $(a, b)$ minimize

$$
E[(Y - a - bX)^2].
$$

Thus, $\hat{Y} = a + bX$ is our guess about $Y$ given $X$.

The squared error is $(Y - \hat{Y})^2$.

The LLSE minimizes the expected value of the squared error.

Why the squares and not the absolute values?
Main justification: much easier!

Note: This is a **Bayesian** formulation: there is a prior.

Single Variable: $E(X)$ minimizes expected squared error.
LR: Non-Bayesian or Uniform?

Observe that

\[
\frac{1}{N} \sum_{n=1}^{N} (Y_n - a - bX_n)^2 = E[(Y - a - bX)^2]
\]

where one assumes that

\[(X, Y) = (X_n, Y_n), \ \text{w.p.} \ \frac{1}{N} \ \text{for} \ n = 1, \ldots, N.\]
LR: Non-Bayesian or Uniform?

Observe that

$$\frac{1}{N} \sum_{n=1}^{N} (Y_n - a - bX_n)^2 = E[(Y - a - bX)^2]$$

where one assumes that

$$(X, Y) = (X_n, Y_n), \text{ w.p. } \frac{1}{N} \text{ for } n = 1, \ldots, N.$$  

That is, the non-Bayesian LR is equivalent to the Bayesian LLSE that assumes that $(X, Y)$ is uniform on the set of observed samples.
Observe that

$$\frac{1}{N} \sum_{n=1}^{N} (Y_n - a - bX_n)^2 = E[(Y - a - bX)^2]$$

where one assumes that

$$(X, Y) = (X_n, Y_n), \text{ w.p. } \frac{1}{N} \text{ for } n = 1, \ldots, N.$$  

That is, the non-Bayesian LR is equivalent to the Bayesian LLSE that assumes that $(X, Y)$ is uniform on the set of observed samples. Thus, we can study the two cases LR and LLSE in one shot.
LR: Non-Bayesian or Uniform?

Observe that

\[
\frac{1}{N} \sum_{n=1}^{N} (Y_n - a - bX_n)^2 = E[(Y - a - bX)^2]
\]

where one assumes that

\[(X, Y) = (X_n, Y_n), \text{ w.p. } \frac{1}{N} \text{ for } n = 1, \ldots, N.\]

That is, the non-Bayesian LR is equivalent to the Bayesian LLSE that assumes that \((X, Y)\) is uniform on the set of observed samples.

Thus, we can study the two cases LR and LLSE in one shot.

However, the interpretations are different!
Theorem
Consider two RVs \( X, Y \) with a given distribution \( Pr[X = x, Y = y] \). Then,

\[
L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)}(X - E[X]).
\]
**Theorem**
Consider two RVs $X$, $Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)}(X - E[X]).$$
**Theorem**
Consider two RVs $X$, $Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)}(X - E[X]).$$

If $cov(X, Y) = 0$, what do you predict for $Y$?
Theorem
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)}(X - E[X]).$$

If $cov(X, Y) = 0$, what do you predict for $Y$? $E(Y)$!
Theorem
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)}(X - E[X]).$$

If $cov(X, Y) = 0$, what do you predict for $Y$? $E(Y)$!
Make sense?
**Theorem**

Consider two RVs $X$, $Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)} (X - E[X]).$$

If $cov(X, Y) = 0$, what do you predict for $Y$? $E(Y)$! Make sense? Sure.
Theorem
Consider two RVs \( X, Y \) with a given distribution \( Pr[X = x, Y = y] \). Then,

\[
L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{\text{var}(X)} (X - E[X]).
\]

If \( cov(X, Y) = 0 \), what do you predict for \( Y? \) \( E(Y) \)!
Make sense? Sure.
Independent!
Theorem
Consider two RVs $X$, $Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)}(X - E[X]).$$

If $cov(X, Y) = 0$, what do you predict for $Y$? $E(Y)$!

Make sense? Sure.
Independent!

If $cov(X, Y)$ is positive, and $X > E(X)$, is $\hat{Y} \geq E(Y)$?
Theorem
Consider two RVs $X$, $Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)}(X - E[X]).$$

If $cov(X, Y) = 0$, what do you predict for $Y$? $E(Y)$!
Make sense? Sure.
Independent!

If $cov(X, Y)$ is positive, and $X > E(X)$, is $\hat{Y} \geq E(Y)$? Sure.
Theorem
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)}(X - E[X]).$$

If $cov(X, Y) = 0$, what do you predict for $Y$? $E(Y)$!
Make sense? Sure.
Independent!

If $cov(X, Y)$ is positive, and $X > E(X)$, is $\hat{Y} \geq E(Y)$? Sure.
Make sense?
Theorem
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)} (X - E[X]).$$

If $cov(X, Y) = 0$, what do you predict for $Y$? $E(Y)$!
Make sense? Sure.
Independent!

If $cov(X, Y)$ is positive, and $X > E(X)$, is $\hat{Y} \geq E(Y)$? Sure.
Make sense? Sure.
Theorem
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)}(X - E[X]).$$

If $cov(X, Y) = 0$, what do you predict for $Y$? $E(Y)$!
Make sense? Sure.
Independent!

If $cov(X, Y)$ is positive, and $X > E(X)$, is $\hat{Y} \geq E(Y)$? Sure.
Make sense? Sure.
Taller
Theorem
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{\text{cov}(X, Y)}{\text{var}(X)} (X - E[X]).$$

If $\text{cov}(X, Y) = 0$, what do you predict for $Y$? $E(Y)$!
Make sense? Sure.
Independent!

If $\text{cov}(X, Y)$ is positive, and $X > E(X)$, is $\hat{Y} \geq E(Y)$? Sure.
Make sense? Sure.
Taller $\rightarrow$ Heavier!
**Theorem**
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)} (X - E[X]).$$

If $cov(X, Y) = 0$, what do you predict for $Y$? $E(Y)$!
Make sense? Sure.
Independent!

If $cov(X, Y)$ is positive, and $X > E(X)$, is $\hat{Y} \geq E(Y)$? Sure.
Make sense? Sure.
Taller $\rightarrow$ Heavier!

If $cov(X, Y)$ is negative, and $X > E(X)$, is $\hat{Y} \geq E(Y)$?
Theorem
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)}(X - E[X]).$$

If $cov(X, Y) = 0$, what do you predict for $Y$? $E(Y)$!
Make sense? Sure.
Independent!

If $cov(X, Y)$ is positive, and $X > E(X)$, is $\hat{Y} \geq E(Y)$? Sure.
Make sense? Sure.
Taller $\rightarrow$ Heavier!

If $cov(X, Y)$ is negative, and $X > E(X)$, is $\hat{Y} \geq E(Y)$? No!
Theorem
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{\text{var}(X)} (X - E[X]).$$

If $cov(X, Y) = 0$, what do you predict for $Y$? $E(Y)$!
Make sense? Sure.
Independent!

If $cov(X, Y)$ is positive, and $X > E(X)$, is $\hat{Y} \geq E(Y)$? Sure.
Make sense? Sure.
Taller $\rightarrow$ Heavier!

If $cov(X, Y)$ is negative, and $X > E(X)$, is $\hat{Y} \geq E(Y)$? No! $\hat{Y} \leq E(Y)$
**Theorem**
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{\text{var}(X)}(X - E[X]).$$

If $cov(X, Y) = 0$, what do you predict for $Y$? $E(Y)$!
Make sense? Sure.
Independent!

If $cov(X, Y)$ is positive, and $X > E(X)$, is $\hat{Y} \geq E(Y)$? Sure.
Make sense? Sure.
Taller $\rightarrow$ Heavier!

If $cov(X, Y)$ is negative, and $X > E(X)$, is $\hat{Y} \geq E(Y)$? No! $\hat{Y} \leq E(Y)$
Make sense?
**Theorem**
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{\text{cov}(X, Y)}{\text{var}(X)}(X - E[X]).$$

If $\text{cov}(X, Y) = 0$, what do you predict for $Y$? $E(Y)$!
Make sense? Sure.
Independent!

If $\text{cov}(X, Y)$ is positive, and $X > E(X)$, is $\hat{Y} \geq E(Y)$? Sure.
Make sense? Sure.
Taller $\rightarrow$ Heavier!

If $\text{cov}(X, Y)$ is negative, and $X > E(X)$, is $\hat{Y} \geq E(Y)$? **No!** $\hat{Y} \leq E(Y)$
Make sense? Sure.
Theorem
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)}(X - E[X]).$$

If $cov(X, Y) = 0$, what do you predict for $Y$? $E(Y)$!
Make sense? Sure.
Independent!

If $cov(X, Y)$ is positive, and $X > E(X)$, is $\hat{Y} \geq E(Y)$? Sure.
Make sense? Sure.
Taller $\rightarrow$ Heavier!

If $cov(X, Y)$ is negative, and $X > E(X)$, is $\hat{Y} \geq E(Y)$? No! $\hat{Y} \leq E(Y)$
Make sense? Sure.
Heavier
Theorem
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)}(X - E[X]).$$

If $cov(X, Y) = 0$, what do you predict for $Y$? $E(Y)$!
Make sense? Sure.
Independent!

If $cov(X, Y)$ is positive, and $X > E(X)$, is $\hat{Y} \geq E(Y)$? Sure.
Make sense? Sure.
Taller $\rightarrow$ Heavier!

If $cov(X, Y)$ is negative, and $X > E(X)$, is $\hat{Y} \geq E(Y)$? No! $\hat{Y} \leq E(Y)$
Make sense? Sure.
Heavier $\rightarrow$ Slower!
Theorem
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)}(X - E[X]).$$
LLSE

Theorem
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)}(X - E[X]).$$

Proof:

$$Y - \hat{Y} = (Y - E[Y]) - \frac{cov(X, Y)}{var[X]}(X - E[X]).$$
Theorem
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{\text{cov}(X, Y)}{\text{var}(X)}(X - E[X]).$$

Proof:
$Y - \hat{Y} = (Y - E[Y]) - \frac{\text{cov}(X,Y)}{\text{var}(X)}(X - E[X])$. Hence, $E[Y - \hat{Y}] = 0.$
Theorem
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)}(X - E[X]).$$

Proof:

$$Y - \hat{Y} = (Y - E[Y]) - \frac{cov(X, Y)}{var[X]}(X - E[X]).$$

Hence, $E[Y - \hat{Y}] = 0$.

Also, $E[(Y - \hat{Y})X] = 0,$
Theorem
Consider two RVs $X$, $Y$ with a given distribution $Pr[X = x, Y = y]$. Then,
$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)}(X - E[X]).$$

Proof:
$$Y - \hat{Y} = (Y - E[Y]) - \frac{cov(X, Y)}{var[X]}(X - E[X]).$$ Hence, $E[Y - \hat{Y}] = 0.$

Also, $E[(Y - \hat{Y})X] = 0$, after a bit of algebra.
LLSE

**Theorem**
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)}(X - E[X]).$$

**Proof:**

$$Y - \hat{Y} = (Y - E[Y]) - \frac{cov(X, Y)}{var[X]}(X - E[X]).$$

Hence, $E[Y - \hat{Y}] = 0$.

Also, $E[(Y - \hat{Y})X] = 0$, after a bit of algebra. (See next slide.)
Theorem
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{\text{cov}(X, Y)}{\text{var}(X)}(X - E[X]).$$

Proof:

$$Y - \hat{Y} = (Y - E[Y]) - \frac{\text{cov}(X, Y)}{\text{var}[X]}(X - E[X]).$$

Hence, $E[Y - \hat{Y}] = 0$.

Also, $E[(Y - \hat{Y})X] = 0$, after a bit of algebra. (See next slide.)

Hence, by combining the two brown equalities,

$$E[(Y - \hat{Y})(c + dX)] = 0.$$
**Theorem**

Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)} (X - E[X]).$$

**Proof:**

$$Y - \hat{Y} = (Y - E[Y]) - \frac{cov(X, Y)}{var[X]} (X - E[X]).$$

Hence, $E[Y - \hat{Y}] = 0$.

Also, $E[(Y - \hat{Y})X] = 0$, after a bit of algebra. (See next slide.)

Hence, by combining the two brown equalities,

$$E[(Y - \hat{Y})(c + dX)] = 0.$$  Then, $E[(Y - \hat{Y})(\hat{Y} - a - bX)] = 0, \forall a, b.$
Theorem
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{\text{cov}(X, Y)}{\text{var}(X)}(X - E[X]).$$

Proof:
$Y - \hat{Y} = (Y - E[Y]) - \frac{\text{cov}(X, Y)}{\text{var}[X]}(X - E[X])$. Hence, $E[Y - \hat{Y}] = 0$.

Also, $E[(Y - \hat{Y})X] = 0$, after a bit of algebra. (See next slide.)

Hence, by combining the two brown equalities,
$E[(Y - \hat{Y})(c + dX)] = 0$. Then, $E[(Y - \hat{Y})(\hat{Y} - a - bX)] = 0, \forall a, b$.

Indeed: $\hat{Y} = \alpha + \beta X$ for some $\alpha, \beta$,
**Theorem**

Consider two RVs \(X, Y\) with a given distribution \(Pr[X = x, Y = y]\). Then,

\[
L[Y|X] = \hat{Y} = E[Y] + \frac{\text{cov}(X, Y)}{\text{var}(X)} (X - E[X]).
\]

**Proof:**

\[
Y - \hat{Y} = (Y - E[Y]) - \frac{\text{cov}(X, Y)}{\text{var}[X]} (X - E[X]).
\]

Hence, \(E[Y - \hat{Y}] = 0\).

Also, \(E[(Y - \hat{Y})X] = 0\), after a bit of algebra. (See next slide.)

Hence, by combining the two brown equalities,

\[
E[(Y - \hat{Y})(c + dX)] = 0.
\]

Then, \(E[(Y - \hat{Y})(\hat{Y} - a - bX)] = 0, \forall a, b\).

Indeed: \(\hat{Y} = \alpha + \beta X\) for some \(\alpha, \beta\), so that \(\hat{Y} - a - bX = c + dX\) for some \(c, d\).
Theorem
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X,Y)}{var(X)}(X - E[X]).$$

Proof:

$Y - \hat{Y} = (Y - E[Y]) - \frac{cov(X,Y)}{var[X]}(X - E[X])$. Hence, $E[Y - \hat{Y}] = 0$.

Also, $E[(Y - \hat{Y})X] = 0$, after a bit of algebra. (See next slide.)

Hence, by combining the two brown equalities, $E[(Y - \hat{Y})(c + dX)] = 0$. Then, $E[(Y - \hat{Y})(\hat{Y} - a - bX)] = 0, \forall a, b$.

Indeed: $\hat{Y} = \alpha + \beta X$ for some $\alpha, \beta$, so that $\hat{Y} - a - bX = c + dX$ for some $c, d$. Now,

$$E[(Y - a - bX)^2] = E[(Y - \hat{Y} + \hat{Y} - a - bX)^2]$$
**Theorem**
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{\text{var}(X)} (X - E[X]).$$

**Proof:**

$$Y - \hat{Y} = (Y - E[Y]) - \frac{cov(X, Y)}{\text{var}[X]} (X - E[X]).$$

Hence, $E[Y - \hat{Y}] = 0$.

Also, $E[(Y - \hat{Y})X] = 0$, after a bit of algebra. (See next slide.)

Hence, by combining the two brown equalities, $E[(Y - \hat{Y})(c + dX)] = 0$. Then, $E[(Y - \hat{Y})(\hat{Y} - a - bX)] = 0, \forall a, b$.

Indeed: $\hat{Y} = \alpha + \beta X$ for some $\alpha, \beta$, so that $\hat{Y} - a - bX = c + dX$ for some $c, d$. Now,

$$E[(Y - a - bX)^2] = E[(Y - \hat{Y} + \hat{Y} - a - bX)^2]$$

$$= E[(Y - \hat{Y})^2] + E[(\hat{Y} - a - bX)^2] + 0$$
Theorem
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)} (X - E[X]).$$

Proof:

$$Y - \hat{Y} = (Y - E[Y]) - \frac{cov(X, Y)}{var(X)} (X - E[X]).$$

Hence, $E[Y - \hat{Y}] = 0$.

Also, $E[(Y - \hat{Y})X] = 0$, after a bit of algebra. (See next slide.)

Hence, by combining the two brown equalities,

$$E[(Y - \hat{Y})(c + dX)] = 0.$$ Then, $E[(Y - \hat{Y})(\hat{Y} - a - bX)] = 0, \forall a, b$.

Indeed: $\hat{Y} = \alpha + \beta X$ for some $\alpha, \beta$, so that $\hat{Y} - a - bX = c + dX$ for some $c, d$. Now,

$$E[(Y - a - bX)^2] = E[(Y - \hat{Y} + \hat{Y} - a - bX)^2]$$

$$= E[(Y - \hat{Y})^2] + E[(\hat{Y} - a - bX)^2] + 0 \geq E[(Y - \hat{Y})^2].$$
Theorem
Consider two RVs $X$, $Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)}(X - E[X]).$$

Proof:
$Y - \hat{Y} = (Y - E[Y]) - \frac{cov(X, Y)}{var[X]}(X - E[X])$. Hence, $E[Y - \hat{Y}] = 0$.

Also, $E[(Y - \hat{Y})X] = 0$, after a bit of algebra. (See next slide.)

Hence, by combining the two brown equalities,
$E[(Y - \hat{Y})(c + dX)] = 0$. Then, $E[(Y - \hat{Y})(\hat{Y} - a - bX)] = 0, \forall a, b$.

Indeed: $\hat{Y} = \alpha + \beta X$ for some $\alpha, \beta$, so that $\hat{Y} - a - bX = c + dX$ for some $c, d$. Now,

$$E[(Y - a - bX)^2] = E[(Y - \hat{Y} + \hat{Y} - a - bX)^2]$$

$$= E[(Y - \hat{Y})^2] + E[(\hat{Y} - a - bX)^2] + 0 \geq E[(Y - \hat{Y})^2].$$

This shows that $E[(Y - \hat{Y})^2] \leq E[(Y - a - bX)^2]$, for all $(a, b)$. 

LLSE
Theorem
Consider two RVs $X, Y$ with a given distribution $Pr[X = x, Y = y]$. Then,

$$L[Y|X] = \hat{Y} = E[Y] + \frac{cov(X, Y)}{var(X)} (X - E[X]).$$

Proof:

$$Y - \hat{Y} = (Y - E[Y]) - \frac{cov(X, Y)}{var(X)} (X - E[X]).$$

Hence, $E[Y - \hat{Y}] = 0$.

Also, $E[(Y - \hat{Y})X] = 0$, after a bit of algebra. (See next slide.)

Hence, by combining the two brown equalities,

$$E[(Y - \hat{Y})(c + dX)] = 0.$$

Then, $E[(Y - \hat{Y})(\hat{Y} - a - bX)] = 0$, $\forall a, b$.

Indeed: $\hat{Y} = \alpha + \beta X$ for some $\alpha, \beta$, so that $\hat{Y} - a - bX = c + dX$ for some $c, d$. Now,

$$E[(Y - a - bX)^2] = E[(Y - \hat{Y} + \hat{Y} - a - bX)^2]$$

$$= E[(Y - \hat{Y})^2] + E[(\hat{Y} - a - bX)^2] + 0 \geq E[(Y - \hat{Y})^2].$$

This shows that $E[(Y - \hat{Y})^2] \leq E[(Y - a - bX)^2]$, for all $(a, b)$.

Thus $\hat{Y}$ is the LLSE.
A Bit of Algebra

\[ Y - \hat{Y} = (Y - E[Y]) - \frac{\text{cov}(X,Y)}{\text{var}[X]} (X - E[X]). \]
A Bit of Algebra

\[ Y - \hat{Y} = (Y - E[Y]) - \frac{\text{cov}(X,Y)}{\text{var}[X]}(X - E[X]). \]

Hence, \( E[Y - \hat{Y}] = 0. \)
A Bit of Algebra

\[ Y - \hat{Y} = (Y - E[Y]) - \frac{\text{cov}(X,Y)}{\text{var}[X]}(X - E[X]). \]

Hence, \( E[Y - \hat{Y}] = 0. \) We want to show that \( E[(Y - \hat{Y})X] = 0. \)
A Bit of Algebra

\[ Y - \hat{Y} = (Y - E[Y]) - \frac{\text{cov}(X, Y)}{\text{var}[X]} (X - E[X]). \]

Hence, \( E[Y - \hat{Y}] = 0. \) We want to show that \( E[(Y - \hat{Y})X] = 0. \)

Note that

\[ E[(Y - \hat{Y})X] = E[(Y - \hat{Y})(X - E[X])], \]
A Bit of Algebra

\[ Y - \hat{Y} = (Y - E[Y]) - \frac{\text{cov}(X,Y)}{\text{var}[X]} (X - E[X]). \]

Hence, \( E[Y - \hat{Y}] = 0 \). We want to show that \( E[(Y - \hat{Y})X] = 0 \).

Note that

\[ E[(Y - \hat{Y})X] = E[(Y - \hat{Y})(X - E[X])], \]

because \( E[(Y - \hat{Y})E[X]] = 0 \).
A Bit of Algebra

\[ Y - \hat{Y} = (Y - E[Y]) - \frac{\text{cov}(X,Y)}{\text{var}[X]}(X - E[X]). \]

Hence, \( E[Y - \hat{Y}] = 0. \) We want to show that \( E[(Y - \hat{Y})X] = 0. \)

Note that

\[ E[(Y - \hat{Y})X] = E[(Y - \hat{Y})(X - E[X])], \]

because \( E[(Y - \hat{Y})E[X]] = 0. \)

Now,

\[
E[(Y - \hat{Y})(X - E[X])]
= E[(Y - E[Y])(X - E[X])] - \frac{\text{cov}(X,Y)}{\text{var}[X]}E[(X - E[X])(X - E[X])]
\]
A Bit of Algebra

\[ Y - \hat{Y} = (Y - E[Y]) - \frac{\text{cov}(X,Y)}{\text{var}[X]} (X - E[X]). \]

Hence, \( E[Y - \hat{Y}] = 0. \) We want to show that \( E[(Y - \hat{Y})X] = 0. \)

Note that

\[ E[(Y - \hat{Y})X] = E[(Y - \hat{Y})(X - E[X])], \]

because \( E[(Y - \hat{Y})E[X]] = 0. \)

Now,

\[
E[(Y - \hat{Y})(X - E[X])] \\
= E[(Y - E[Y])(X - E[X])] - \frac{\text{cov}(X,Y)}{\text{var}[X]} E[(X - E[X])(X - E[X])] \\
= (*) \text{cov}(X, Y) - \frac{\text{cov}(X,Y)}{\text{var}[X]} \text{var}[X] = 0. \quad \square
\]

\( (*) \) Recall that \( \text{cov}(X, Y) = E[(X - E[X])(Y - E[Y])] \) and \( \text{var}[X] = E[(X - E[X])^2]. \)
A picture

The following picture explains the algebra:

\[
X, Y \text{ vectors where } X_i, Y_i \text{ is outcome. }
\]

\[
c \text{ is a constant vector.}
\]

We saw that
\[
E\left[ Y - \hat{Y} \right] = 0.
\]

In the picture, this says that
\[
Y - \hat{Y} \perp c, \text{ for any } c.
\]

We also saw that
\[
E\left[ (Y - \hat{Y})X \right] = 0.
\]

In the picture, this says that
\[
Y - \hat{Y} \perp X.
\]

Hence,
\[
Y - \hat{Y} \text{ is orthogonal to the plane } \{c + dX, c, d \in \mathbb{R}\}.
\]

Consequently,
\[
Y - \hat{Y} \perp \hat{Y} - a - bX.
\]

Pythagoras then says that \( \hat{Y} \) is closer to \( Y \) than \( a + bX \).

That is,
\[
\hat{Y} \text{ is the projection of } Y \text{ onto the plane.}
\]

Note: this picture corresponds to uniform probability space.
The following picture explains the algebra:

\[ \hat{Y} = L[Y|X] \]

- \( \hat{Y} \) is the projection of \( Y \) onto the plane.
- \( \hat{Y} \) is closer to \( Y \) than \( a + bX \).
- \( \{c + dX, c, d \in \mathbb{R}\} \) represents the plane.
- \( c \) is a constant vector.
- \( X, Y \) are vectors where \( X_i, Y_i \) is outcome.
- We saw that \( E[Y - \hat{Y}] = 0 \).
- In the picture, this says that \( Y - \hat{Y} \perp c \), for any \( c \).
- We also saw that \( E[(Y - \hat{Y})X] = 0 \).
- In the picture, this says that \( Y - \hat{Y} \perp X \).
- Hence, \( Y - \hat{Y} \) is orthogonal to the plane \( \{c + dX, c, d \in \mathbb{R}\} \).
- Consequently, \( Y - \hat{Y} \perp \hat{Y} - a - bX \).
- Pythagoras then says that \( \hat{Y} \) is closer to \( Y \) than \( a + bX \).
The following picture explains the algebra:

X, Y vectors where \( X_i, Y_i \) is outcome.
The following picture explains the algebra:

\[ \hat{Y} = L[Y|X] \]

\[ X, Y \text{ vectors where } X_i, Y_i \text{ is outcome.} \]
\[ c \text{ is a constant vector.} \]

We saw that \( E[Y - \hat{Y}] = 0 \).
The following picture explains the algebra:

$X, Y$ vectors where $X_i, Y_i$ is outcome.
$c$ is a constant vector.

We saw that $E[Y - \hat{Y}] = 0$. In the picture, this says that $Y - \hat{Y} \perp c$, for any $c$. 

Note: this picture corresponds to uniform probability space.
A picture

The following picture explains the algebra:

\[ \hat{Y} = L[Y|X] \]

\( X, Y \) vectors where \( X_i, Y_i \) is outcome. 
\( c \) is a constant vector.

We saw that \( E[Y - \hat{Y}] = 0 \). In the picture, this says that \( Y - \hat{Y} \perp c \), for any \( c \).

We also saw that \( E[(Y - \hat{Y})X] = 0 \). In the picture, this says that \( Y - \hat{Y} \perp X \).
A picture

The following picture explains the algebra:

$X, Y$ vectors where $X_i, Y_i$ is outcome.
$c$ is a constant vector.

We saw that $E[Y - \hat{Y}] = 0$. In the picture, this says that $Y - \hat{Y} \perp c$, for any $c$.

We also saw that $E[(Y - \hat{Y})X] = 0$. In the picture, this says that $Y - \hat{Y} \perp X$.

Hence, $Y - \hat{Y}$ is orthogonal to the plane $\{c + dX, c, d \in \mathbb{R}\}$. 

Note: this picture corresponds to uniform probability space.
The following picture explains the algebra:

We saw that \( E[Y - \hat{Y}] = 0 \). In the picture, this says that \( Y - \hat{Y} \perp c \), for any \( c \).

We also saw that \( E[(Y - \hat{Y})X] = 0 \). In the picture, this says that \( Y - \hat{Y} \perp X \).

Hence, \( Y - \hat{Y} \) is orthogonal to the plane \( \{c + dX, c, d \in \mathbb{R}\} \).

Consequently, \( Y - \hat{Y} \perp \hat{Y} - a - bX \). Pythagoras then says that \( \hat{Y} \) is closer to \( Y \) than \( a + bX \).
The following picture explains the algebra:

We saw that $E[Y - \hat{Y}] = 0$. In the picture, this says that $Y - \hat{Y} \perp c$, for any $c$.

We also saw that $E[(Y - \hat{Y})X] = 0$. In the picture, this says that $Y - \hat{Y} \perp X$.

Hence, $Y - \hat{Y}$ is orthogonal to the plane $\{c + dX, c, d \in \mathbb{R}\}$.

Consequently, $Y - \hat{Y} \perp \hat{Y} - a - bX$. Pythagoras then says that $\hat{Y}$ is closer to $Y$ than $a + bX$.

That is, $\hat{Y}$ is the projection of $Y$ onto the plane.
The following picture explains the algebra:

We saw that $E[Y - \hat{Y}] = 0$. In the picture, this says that $Y - \hat{Y} \perp c$, for any $c$.

We also saw that $E[(Y - \hat{Y})X] = 0$. In the picture, this says that $Y - \hat{Y} \perp X$.

Hence, $Y - \hat{Y}$ is orthogonal to the plane $\{c + dX, c, d \in \mathbb{R}\}$.

Consequently, $Y - \hat{Y} \perp \hat{Y} - a - bX$. Pythagoras then says that $\hat{Y}$ is closer to $Y$ than $a + bX$.

That is, $\hat{Y}$ is the projection of $Y$ onto the plane.

Note: this picture corresponds to uniform probability space.
Example 1:
Example 1:
Example 2:

\[ E[X] = 0; \]
\[ E[Y] = 0; \]
\[ E[X^2] = \frac{1}{2}; \]
\[ E[XY] = \frac{1}{2}; \]
\[ \text{var}[X] = E[X^2] - E[X]^2 = \frac{1}{2}; \]
\[ \text{cov}(X,Y) = E[XY] - E[X]E[Y] = \frac{1}{2}; \]

LR:
\[ \hat{Y} = E[Y] + \frac{\text{cov}(X,Y)}{\text{var}[X]} (X - E[X]) = X. \]
Example 2:

We find:

\[ E[X] = 0; \]
\[ E[Y] = 0; \]
\[ E[X^2] = \frac{1}{2}; \]
\[ E[XY] = \frac{1}{2}; \]
\[ \text{var}[X] = E[X^2] - E[X]^2 = \frac{1}{2}; \]
\[ \text{cov}(X, Y) = E[XY] - E[X]E[Y] = \frac{1}{2}; \]

LR:

\[ \hat{Y} = E[Y] + \frac{\text{cov}(X, Y)}{\text{var}[X]} (X - E[X]) = X. \]
Linear Regression Examples

Example 2:

We find:

\[ E[X] = \frac{1}{2}; \quad E[Y] = \frac{1}{2}; \quad E[X^2] = \frac{1}{2}; \quad E[XY] = \frac{1}{2}; \]

\[ \text{var}[X] = E[X^2] - (E[X])^2 = \frac{1}{2}; \]

\[ \text{cov}(X, Y) = E[XY] - E[X]E[Y] = \frac{1}{2}; \]

\[ \hat{Y} = E[Y] + \frac{\text{cov}(X, Y)}{\text{var}[X]} (X - E[X]); \]

\[ X \]

\[ Y \]
Linear Regression Examples

Example 2:

We find:

\[ E[X] = 0; \]

\[ E[Y] = 0; \]

\[ E[X^2] = \frac{1}{2}; \]

\[ E[XY] = \frac{1}{2}; \]

\[ \text{var}[X] = E[X^2] - E[X]^2 = \frac{1}{2}; \]

\[ \text{cov}(X, Y) = E[XY] - E[X]E[Y] = \frac{1}{2}; \]

LR:

\[ \hat{Y} = E[Y] + \frac{\text{cov}(X, Y)}{\text{var}[X]} (X - E[X]) = X. \]
Linear Regression Examples

Example 2:

We find:

\[ E[X] = 0; \quad E[Y] = \]

\[ \text{var}[X] = E[X^2] - E[X]^2 = \frac{1}{2}; \quad \text{cov}(X, Y) = E[XY] - E[X]E[Y] = \frac{1}{2}; \]

\[ \hat{Y} = E[Y] + \frac{\text{cov}(X, Y)}{\text{var}[X]} (X - E[X]) = X. \]
Linear Regression Examples

Example 2:

We find:

\[ E[X] = 0; \ E[Y] = 0; \]
Linear Regression Examples

Example 2:

We find:

\[ E[X] = 0; \ E[Y] = 0; \ E[X^2] = \]

\[ \text{var}(X) = E[X^2] - (E[X])^2 \]

\[ \text{cov}(X,Y) = E[XY] - E[X]E[Y] \]

\[ \hat{Y} = E[Y] + \frac{\text{cov}(X,Y)}{\text{var}(X)}(X - E[X]) \]
Linear Regression Examples

Example 2:

We find:

\[ E[X] = 0; \quad E[Y] = 0; \quad E[X^2] = 1/2; \]
Linear Regression Examples

Example 2:

We find:

\[ E[X] = 0; \ E[Y] = 0; \ E[X^2] = 1/2; \ E[XY] = \]
Linear Regression Examples

Example 2:

\[ E[X] = 0; E[Y] = 0; E[X^2] = 1/2; E[XY] = 1/2; \]

We find:
Linear Regression Examples

Example 2:

We find:

\[ E[X] = 0; \ E[Y] = 0; \ E[X^2] = 1/2; \ E[XY] = 1/2; \]
\[ var[X] = E[X^2] - E[X]^2 = \]
Linear Regression Examples

Example 2:

We find:

\[ E[X] = 0; \ E[Y] = 0; \ E[X^2] = 1/2; \ E[XY] = 1/2; \]

\[ \text{var}[X] = E[X^2] - E[X]^2 = 1/2; \]
Example 2:

We find:

\[ E[X] = 0; \ E[Y] = 0; \ E[X^2] = 1/2; \ E[XY] = 1/2; \]
\[ \text{var}[X] = E[X^2] - E[X]^2 = 1/2; \ \text{cov}(X, Y) = E[XY] - E[X]E[Y] = \]
Example 2:

We find:

\[ E[X] = 0; \quad E[Y] = 0; \quad E[X^2] = 1/2; \quad E[XY] = 1/2; \]

\[ \text{var}[X] = E[X^2] - E[X]^2 = 1/2; \quad \text{cov}(X, Y) = E[XY] - E[X]E[Y] = 1/2; \]
Linear Regression Examples

Example 2:

We find:

\[ E[X] = 0; E[Y] = 0; E[X^2] = 1/2; E[XY] = 1/2; \]
\[ \text{var}[X] = E[X^2] - E[X]^2 = 1/2; \text{cov}(X, Y) = E[XY] - E[X]E[Y] = 1/2; \]
\[ \text{LR: } \hat{Y} = E[Y] + \frac{\text{cov}(X, Y)}{\text{var}[X]}(X - E[X]) = \]
Linear Regression Examples

Example 2:

We find:

\[
E[X] = 0; \ E[Y] = 0; \ E[X^2] = 1/2; \ E[XY] = 1/2; \\
var[X] = E[X^2] - E[X]^2 = 1/2; \ \text{cov}(X, Y) = E[XY] - E[X]E[Y] = 1/2; \\
\text{LR:} \ \hat{Y} = E[Y] + \frac{\text{cov}(X, Y)}{\text{var}[X]}(X - E[X]) = X.
\]
Example 3:

\[
E[X] = 0; \quad E[Y] = 0; \quad E[X^2] = \frac{1}{2}; \quad E[XY] = -\frac{1}{2};
\]

\[
\text{var}[X] = E[X^2] - (E[X])^2 = \frac{1}{2};
\]

\[
\text{cov}(X, Y) = E[XY] - E[X]E[Y] = -\frac{1}{2};
\]

\[
\hat{Y} = E[Y] + \frac{\text{cov}(X, Y)}{\text{var}[X]} (X - E[X]) = X.
\]
Linear Regression Examples

Example 3:

We find:

\[ E[X] = 0; \]  \[ E[Y] = 0; \]  \[ E[X^2] = \frac{1}{2}; \]  \[ E[XY] = -\frac{1}{2}; \]

\[ \text{var}[X] = E[X^2] - E[X]^2 = \frac{1}{2}; \]

\[ \text{cov}(X, Y) = E[XY] - E[X]E[Y] = -\frac{1}{2}; \]

\[ \text{LR}: \hat{Y} = E[Y] + \frac{E[XY] - E[X]E[Y]}{\text{var}[X]} (X - E[X]) = -X. \]
Linear Regression Examples

Example 3:

We find:

\[ E[X] = \]

\[ E[Y] = \]

\[ E[X^2] = \frac{1}{2} \]

\[ E[XY] = -\frac{1}{2} \]

\[ \text{var}[X] = E[X^2] - (E[X])^2 = \frac{1}{2} \]

\[ \text{cov}(X, Y) = E[XY] - E[X]E[Y] = -\frac{1}{2} \]

\[ \hat{Y} = E[Y] + \frac{\text{cov}(X, Y)}{\text{var}[X]} (X - E[X]) = -X \]
Example 3:

We find:

\[ E[X] = 0; \]
Example 3:

We find:

\[ E[X] = 0; E[Y] = \]
Linear Regression Examples

Example 3:

We find:

\[ E[X] = 0; \ E[Y] = 0; \]
Linear Regression Examples

Example 3:

We find:

\[ E[X] = 0; E[Y] = 0; E[X^2] = \]

\[\text{var}[X] = E[X^2] - E[X]^2 = \]

\[\text{cov}(X, Y) = E[XY] - E[X]E[Y] = - \frac{1}{2} \]

\[\hat{Y} = E[Y] + \frac{\text{cov}(X, Y)}{\text{var}[X]} (X - E[X]) = \]

\[- \frac{1}{2} X.\]
Linear Regression Examples

Example 3:

We find:

\[ E[X] = 0; \ E[Y] = 0; \ E[X^2] = 1/2; \]
Example 3:

We find:

\[ E[X] = 0; E[Y] = 0; E[X^2] = 1/2; E[XY] = \]
Example 3:

We find:

\[ E[X] = 0; E[Y] = 0; E[X^2] = 1/2; E[XY] = -1/2; \]
Example 3:

We find:

\[ E[X] = 0; \ E[Y] = 0; \ E[X^2] = 1/2; \ E[XY] = -1/2; \]
\[ var[X] = E[X^2] - E[X]^2 = \]
Linear Regression Examples

Example 3:

We find:

\[
E[X] = 0; \ E[Y] = 0; \ E[X^2] = 1/2; \ E[XY] = -1/2;
\]

\[
var[X] = E[X^2] - E[X]^2 = 1/2;
\]
Linear Regression Examples

Example 3:

We find:

\[ E[X] = 0; E[Y] = 0; E[X^2] = 1/2; E[XY] = -1/2; \]
Linear Regression Examples

Example 3:

We find:

\[ E[X] = 0; \ E[Y] = 0; \ E[X^2] = 1/2; \ E[XY] = -1/2; \]
\[ var[X] = E[X^2] - E[X]^2 = 1/2; \ cov(X, Y) = E[XY] - E[X]E[Y] = -1/2; \]
Linear Regression Examples

Example 3:

We find:

\[ E[X] = 0; \ E[Y] = 0; \ E[X^2] = 1/2; \ E[XY] = -1/2; \]
\[ var[X] = E[X^2] - E[X]^2 = 1/2; \ cov(X, Y) = E[XY] - E[X]E[Y] = -1/2; \]
\[ LR: \ \hat{Y} = E[Y] + \frac{cov(X, Y)}{var[X]}(X - E[X]) = \]
We find:

\[
E[X] = 0; E[Y] = 0; E[X^2] = 1/2; E[XY] = -1/2;
\]
\[
\text{var}[X] = E[X^2] - E[X]^2 = 1/2; \text{cov}(X, Y) = E[XY] - E[X]E[Y] = -1/2;
\]
\[
\text{LR: } \hat{Y} = E[Y] + \frac{\text{cov}(X, Y)}{\text{var}[X]} (X - E[X]) = -X.
\]
Linear Regression Examples

Example 4:

\[ E[X] = 3; \quad E[Y] = 2.5; \quad E[X^2] = \left(\frac{3}{15}\right)(1 + 2^2 + 3^2 + 4^2 + 5^2) = 11; \quad E[XY] = \left(\frac{1}{15}\right)(1 \times 1 + 1 \times 2 + \cdots + 5 \times 4) = 8.4; \quad \text{var}[X] = 11 - 9 = 2; \quad \text{cov}(X,Y) = 8.4 - 3 \times 2.5 = 0.9; \]

LR: \[ \hat{Y} = 2.5 + 0.92(X - 3) = 1.15 + 0.45X. \]
Example 4:

\[ E[X] = 3; \quad E[Y] = 2.5; \quad E[X^2] = \frac{3}{15}(1 + 2^2 + 3^2 + 4^2 + 5^2) = 11; \quad E[XY] = \frac{1}{15}(1 \times 1 + 1 \times 2 + \ldots + 5 \times 4) = 8.4; \]

\[ \text{var}[X] = 11 - 9 = 2; \quad \text{cov}(X, Y) = 8.4 - 3 \times 2.5 = 0.9; \]

LR: \[ \hat{Y} = 2.5 + 0.9X. \]
Linear Regression Examples

Example 4:

We find:

\[
E[X] = \frac{3}{15}; \quad E[Y] = 2.5;
\]

\[
E[X^2] = \frac{1}{15}(1^2 + 2^2 + \ldots + 5^2) = 11;
\]

\[
E[XY] = \frac{1}{15}(1 \times 1 + 1 \times 2 + \ldots + 5 \times 4) = 8.4;
\]

\[
\text{var}[X] = E[X^2] - (E[X])^2 = 11 - \left(\frac{3}{15}\right)^2 = 2;
\]

\[
\text{cov}(X,Y) = E[XY] - E[X]E[Y] = 8.4 - 3 \times 2.5 = 0.9;
\]

LR:

\[
\hat{Y} = 2.5 + 0.9 \cdot (X - 3) = 1.15 + 0.45X.
\]
Linear Regression Examples

Example 4:

We find:

\[ E[X] = 3; \]

\[ E[Y] = 2.5; \]

\[ E[X^2] = \frac{3}{15}(1 + 2^2 + 3^2 + 4^2 + 5^2) = 11; \]

\[ E[XY] = \frac{1}{15}(1 \times 1 + 1 \times 2 + \cdots + 5 \times 4) = 8.4; \]

\[ \text{var}[X] = 11 - 9 = 2; \]

\[ \text{cov}(X,Y) = 8.4 - 3 \times 2.5 = 0.9; \]

\[ \hat{Y} = 2.5 + 0.92(X - 3) = 1.15 + 0.45X. \]
Linear Regression Examples

Example 4:

We find:

\[ E[X] = 3; E[Y] = \]

\[ E[X^2] = \frac{3}{15}(1^2 + 2^2 + 3^2 + 4^2 + 5^2) = 11; \]

\[ E[XY] = \frac{1}{15}(1 \times 1 + 1 \times 2 + \cdots + 5 \times 4) = 8.4; \]

\[ \text{var}[X] = 11 - 9 = 2 ; \]

\[ \text{cov}(X, Y) = 8.4 - 3 \times 2.5 = 0.9; \]

\[ \hat{Y} = 2.5 + 0.9X. \]
Linear Regression Examples

Example 4:

We find:

\[ E[X] = 3; \ E[Y] = 2.5; \]
We find:

\[ E[X] = 3; \quad E[Y] = 2.5; \quad E[X^2] = \left(\frac{3}{15}\right)(1 + 2^2 + 3^2 + 4^2 + 5^2) = 11; \]
Linear Regression Examples

Example 4:

We find:

\[ E[X] = 3; \ E[Y] = 2.5; \ E[X^2] = (3/15)(1 + 2^2 + 3^2 + 4^2 + 5^2) = 11; \]
\[ E[XY] = (1/15)(1 \times 1 + 1 \times 2 + \cdots + 5 \times 4) = 8.4; \]
Linear Regression Examples

Example 4:

We find:

\[ E[X] = 3; \ E[Y] = 2.5; \ E[X^2] = (3/15)(1 + 2^2 + 3^2 + 4^2 + 5^2) = 11; \]
\[ E[XY] = (1/15)(1 \times 1 + 1 \times 2 + \cdots + 5 \times 4) = 8.4; \]
\[ var[X] = 11 - 9 = 2; \]
We find:

\[ E[X] = 3; \ E[Y] = 2.5; \ E[X^2] = (3/15)(1 + 2^2 + 3^2 + 4^2 + 5^2) = 11; \]
\[ E[XY] = (1/15)(1 \times 1 + 1 \times 2 + \cdots + 5 \times 4) = 8.4; \]
\[ var[X] = 11 - 9 = 2; \ \text{cov}(X, Y) = 8.4 - 3 \times 2.5 = 0.9; \]
We find:

\[ E[X] = 3; \ E[Y] = 2.5; \ E[X^2] = (3/15)(1 + 2^2 + 3^2 + 4^2 + 5^2) = 11; \]
\[ E[XY] = (1/15)(1 \times 1 + 1 \times 2 + \cdots + 5 \times 4) = 8.4; \]
\[ var[X] = 11 - 9 = 2; \ cov(X, Y) = 8.4 - 3 \times 2.5 = 0.9; \]
\[ LR: \hat{Y} = 2.5 + \frac{0.9}{2}(X - 3) = 1.15 + 0.45X. \]
Note that the LR line goes through \((X_n, Y_n)\). Its slope is \(\frac{\text{cov}(X,Y)}{\text{var}[X]}\).
Note that

- the LR line goes through \((E[X], E[Y])\)
Note that

- the LR line goes through \((E[X], E[Y])\)
- its slope is \(\frac{cov(X,Y)}{var[X]}\).
Summary

Linear Regression

1. Multiple Random variables: $X, Y$ with $P[X=x, Y=y]$.
2. Marginal & conditional probabilities
3. Linear Regression: $L[Y|X] = E[Y] + \text{cov}(X,Y) \text{var}(X)(X - E[X])$
4. Non-Bayesian: minimize $\sum_n (Y_n - a - bX_n)^2$
5. Bayesian: minimize $E[(Y - a - bX)^2]$
Summary

1. Multiple Random variables: $X, Y$ with $Pr[X = x, Y = y]$.
Summary

1. Multiple Random variables: $X, Y$ with $Pr[X = x, Y = y]$.
2. Marginal & conditional probabilities
3. Linear Regression: $L[Y|X] = E[Y] + \frac{\text{cov}(X,Y)}{\text{var}(X)}(X - E[X])$
Summary

1. Multiple Random variables: $X, Y$ with $Pr[X = x, Y = y]$.
2. Marginal & Conditional probabilities
3. Linear Regression: $L[Y|X] = E[Y] + \frac{cov(X,Y)}{var(X)}(X - E[X])$
4. Non-Bayesian: minimize $\sum_n (Y_n - a - bX_n)^2$
Summary

Linear Regression

1. Multiple Random variables: $X, Y$ with $Pr[X = x, Y = y]$.
2. Marginal & conditional probabilities
3. Linear Regression: $L[Y|X] = E[Y] + \frac{cov(X, Y)}{var(X)}(X - E[X])$
4. Non-Bayesian: minimize $\sum_n(Y_n - a - bX_n)^2$
5. Bayesian: minimize $E[(Y - a - bX)^2]$