Machine Learning (V): Conditional Expectation and Linear Regression

Le Wang

CEF: Motivation, Definition, Properties, and Use

Linear Regression as a Computational Tool for CEF

When is linear regression not working? And what to do?

CEF: Motivation, Definition, Properties, and Use

Motivation

The use of the entire conditional distribution to characterize the relationship between two (or even more) variables is purely nonparametric and hence model free.

However, what if one or more of the variables are **continuous**?

1. Dependent Variable, Y
2. Independent Variable, X

Motivation

For continuous **dependent** variables (X is till kept as **discrete**)

Instead, we can focus on **parts of the distribution**

1. **Moments**: $\mathbb{E}[Y|X]$ (Conditional Expectation (Mean))
2. **Quantiles**: $Q_{Y|X}(y|x)$ is defined as usual, but for the conditional distribution (Conditional Quantiles)

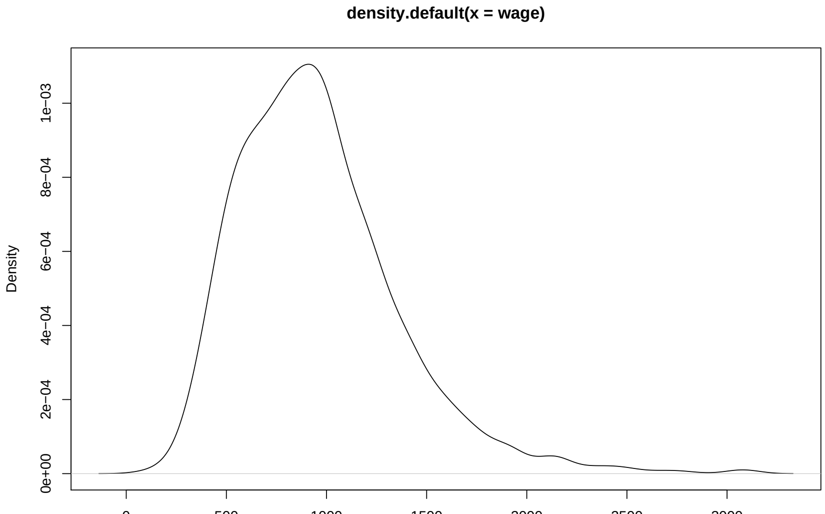
$$\inf\{y : F_{Y|X=x}(y) \geq \tau\}$$

In other words, we will obtain the mean or quantiles of Y for the subpopulation when $X = x$. Business as usual.

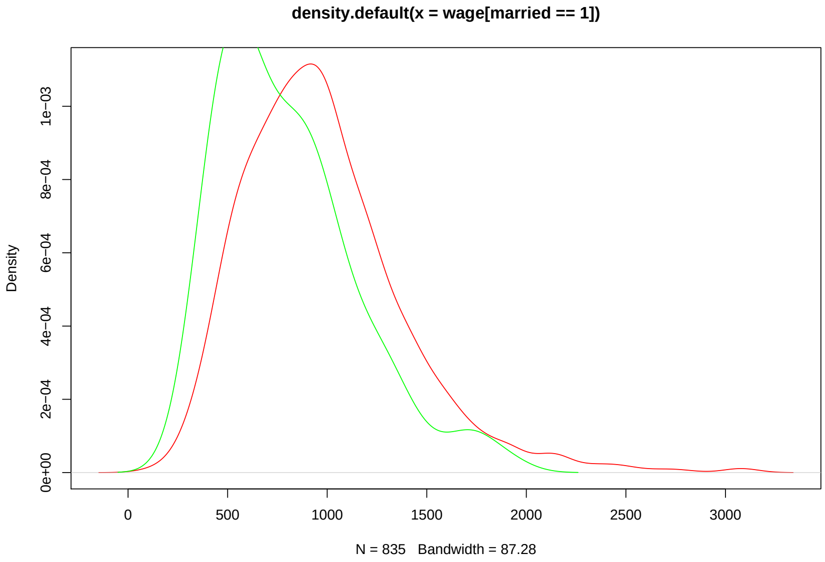
```r
summary(wage)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
##   115.0   669.0   905.0   957.9  1160.0  3078.0 
```

```r
plot(density(wage))
```



```r
plot(density(wage[married==1]),col="red")
lines(density(wage[married==0]),col="green")
```



Motivation

Question: Which feature of the conditional distribution to focus on?

Motivation

Question: Which feature of the conditional distribution to focus on?

Answer: It depends! But here we will illustrate several reasons why we can focus on the CEF.

Motivation

Contrary to the **unconditional expectation** $\mathbb{Y}$, the conditional expectation is a **random variable** (since $f_{Y|X}(y|X)$ is a random function)!

We can think of Y as a function of X:

$$\mathbb{E}[Y|X] = g(X) \quad \text{for some function } g(\cdot)$$

where $g(\cdot)$ is a function determined by the joint (and hence conditional) distribution of (X, Y).

Our goal: To figure out what is $g(X)$!

Definition

Conditional Expectation

$$\mathbb{E}[Y \mid X]$$

has a value for every value of X. If X is multi-dimensional, $X_1, X_2, \ldots, X_k$ just for **every** combination of the values of $X_1, X_2, \ldots, X_k$.

Definition

1. Wages for different marital status (married vs non-married):

Definition

1. Wages for different marital status (married vs non-married):

$$\mathbb{E}[\text{wages} \mid \text{married}] = \begin{cases} 798.4400 & \text{if married } = 0 \\ 977.0479 & \text{if married } = 1 \end{cases}$$

Definition

1. Wages for different marital status (married vs non-married):

$$\mathbb{E}[\text{wages} \mid \text{married}] = \begin{cases} 798.4400 & \text{if married } = 0 \\ 977.0479 & \text{if married } = 1 \end{cases}$$

2. Wages for different education levels (below high school, high school, above high school):

Definition

1. Wages for different marital status (married vs non-married):

$$\mathbb{E}[\text{wages} \mid \text{married}] = \begin{cases} 798.4400 & \text{if married } = 0 \\ 977.0479 & \text{if married } = 1 \end{cases}$$

2. Wages for different education levels (below high school, high school, above high school):

$$\mathbb{E}[\text{wage} \mid \text{education}] = \begin{cases} 774.2500 & \text{if educ} = \text{below high school} \\ 862.6718 & \text{if educ} = \text{high school} \\ 1076.0242 & \text{if educ} = \text{above high school} \end{cases}$$

3. Wage for different race (black vs non-blacks) and marital status (married vs non-married)

3. Wage for different race (black vs non-blacks) and marital status (married vs non-married)

$$\mathbb{E}[\text{wages}|\text{married,black}] = \begin{cases} 841.9756 & \text{if unmarried non-blacks} \\ 1007.2797 & \text{if married non-blacks} \\ 600.1111 & \text{if unmarried blacks} \\ 759.7941 & \text{if married blacks} \end{cases}$$

Use of CEF

We focus on the CEF for the following two purposes

1. Prediction
2. Marginal Effects

Use of CEF (I): Prediction

Question: Why We focus on the conditional mean (mean for sub-populations): $\mathbb{E}[Y|X]$?

1. Prediction Property (best prediction under certain assumptiosn)
2. Decomposition Property (prediction errors on average are zero!)

Use of CEF (I): Prediction – Reason 1: Prediction Property

The CEF is a good summary of the relationship between x and y for a number of reasons.

Use of CEF (I): Prediction – Reason 1: Prediction Property

The CEF is a good summary of the relationship between x and y for a number of reasons.

Just like unconditional mean, conditional mean provides a representative value (or “best” prediction) for a random variable, y, for a sub-population with characteristics, x.

This property is called **prediction property**

More formally, Let $m(x)$ be any function of x. Then, $\mathbb{E}[y|x]$ solves the following minimization problem.

$$\min \quad \mathbb{E}[(y - m(x))^2]$$

More formally, Let $m(x)$ be any function of x. Then, $\mathbb{E}[y|x]$ solves the following minimization problem.

$$\min \quad \mathbb{E}[(y - m(x))^2]$$

Intuition: The conditional expectation is the best prediction in the sense that it minimizes mean squared error.

Use of CEF (I): Prediction – Reason 2: Decomposition Property

Indeed, you can think of any random variable, y, as follows

$$y = \mathbb{E}[y|x] + \epsilon$$

1. A part that can be "explained by x"
2. A part that can **NOT** be explained by x at all!

Question What do you mean by “a part that cannot be explained by x”?

$$\mathbb{E}[\epsilon|x] = \mathbb{E}[\epsilon] = 0!!$$

Intuition: If you use $\mathbb{E}[y|x]$ as your prediction, you will make some erorrs, but on average, the errors are equal to zero!!

Use of CEF (II): Partial Effects

Question:

1. How to predict wages taking into account married status?

$$\mathbb{E}[\text{wages}|\text{married}] \qquad \mathbb{E}[\text{wages}|\text{non-married}]$$

Use of CEF (II): Partial Effects

Question:

1. How to predict wages taking into account married status?

$$\mathbb{E}[\text{wages}|\text{married}] \qquad \mathbb{E}[\text{wages}|\text{non-married}]$$

2. What is the marriage premium (for average individuals)? (unlike before we are going to look at this question for the continuous measure of income)

Use of CEF (II): Partial Effects

Question:

1. How to predict wages taking into account married status?

$$\mathbb{E}[\text{wages}|\text{married}] \qquad \mathbb{E}[\text{wages}|\text{non-married}]$$

2. What is the marriage premium (for average individuals)? (unlike before we are going to look at this question for the continuous measure of income)

$$\mathbb{E}[\text{wages}|\text{married}] - \mathbb{E}[\text{wages}|\text{non-married}]$$

Computation of Predictions and Partial Effects using CEF

Lets look at our example: `mv06_cond_expectation01.R`

Linear Regression as a Computational Tool for CEF

Review of Linear Regression

Suppose (with no good reason) that

$$y = g(x) + \epsilon = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$$

There are many possible $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$. How to choose?

Review of Linear Regression

Suppose (with no good reason) that

$$y = g(x) + \epsilon = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$$

There are many possible $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$. How to choose?

Note that when you choose $\beta_1, \beta_1, \beta_2, \ldots, \beta_k$, you are choosing a function, $g(x)$ but a linear one!

Regression coefficients are those, $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$, that solve the following minimization problem.

$$\min \quad \mathbb{E}[(y - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k))^2]$$

Once you pick the optimal $\beta_1, \beta_1, \beta_2, \ldots, \beta_k$, your prediction of y for a given set of $x_1, x_2, \ldots, x_k$ will be

$$\widehat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

Some necessary notations:

1. β_0: Intercept. The value of y when all xs are set to zero.

$$\widehat{y} = \beta_0 + \beta_1 \cdot 0 + \beta_2 \cdot + \cdots + \beta_k \cdot 0$$

2. $\beta_j, j = 1, \ldots, k$: slope coefficient for x_j. (We will see how we can interpret them in a second).

Linear Regression in R

Let's first look at how to implement it in R.

```
lm(y ~ x1 + x2,data=data)
```

As we discussed before,

1. The "$\sim$" symbol separates response (outcome) variables from the explanatory variables (or predictors)
2. The "$+$" sybmol separates predictor variables.

As emphasized earlier, we focus our attention on discrete X first. It is treated as a **factor variable**. Every value of X should be considered as a separate category and coresponds to a new variable indicating whether or not the observation belonging to this category.

As emphasized earlier, we focus our attention on discrete X first. It is treated as a **factor variable**. Every value of X should be considered as a separate category and coresponds to a new variable indicating whether or not the observation belonging to this category.

In a model, if you type `factor(x)`, R will automatically create the dummy variables for each possible value of x and include them properly in estimations.

Recall from the previous lecture on factor variable: If X can take only four different values (say, $1, 2, 3, 4$ or first, second, third, fourth seasons), then we need to create four **additional** dummy variables, denoted by x_1, x_2, x_3, x_4:

$$\begin{aligned} x_1 &= \mathbb{I}[X = 1] \\ x_2 &= \mathbb{I}[X = 2] \\ x_3 &= \mathbb{I}[X = 3] \\ x_4 &= \mathbb{I}[X = 4] \end{aligned}$$

In other words, we practically have 4 different variables from each discrete variable. There are two **equivalent** ways to use these newly constructed variables (the reason is something called **perfect collinearity issue**, to be explained later)

In other words, we practically have 4 different variables from each discrete variable. There are two **equivalent** ways to use these newly constructed variables (the reason is something called **perfect collinearity issue**, to be explained later)

Two Equivalent forms of Linear Regressions with a factor variable with 4 different values

In other words, we practically have 4 different variables from each discrete variable. There are two **equivalent** ways to use these newly constructed variables (the reason is something called **perfect collinearity issue**, to be explained later)

Two Equivalent forms of Linear Regressions with a factor variable with 4 different values

1. Without intercept

In other words, we practically have 4 different variables from each discrete variable. There are two **equivalent** ways to use these newly constructed variables (the reason is something called **perfect collinearity issue**, to be explained later)

Two Equivalent forms of Linear Regressions with a factor variable with 4 different values

1. Without intercept

$$y = \quad \beta_1 x_1 + \beta_2 x_2 + \beta_3 \cdot x_3 + \beta_4 \cdot x_4 + \epsilon$$

In other words, we practically have 4 different variables from each discrete variable. There are two **equivalent** ways to use these newly constructed variables (the reason is something called **perfect collinearity issue**, to be explained later)

Two Equivalent forms of Linear Regressions with a factor variable with 4 different values

1. Without intercept

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 \cdot x_3 + \beta_4 \cdot x_4 + \epsilon$$

2. With intercept

In other words, we practically have 4 different variables from each discrete variable. There are two **equivalent** ways to use these newly constructed variables (the reason is something called **perfect collinearity issue**, to be explained later)

Two Equivalent forms of Linear Regressions with a factor variable with 4 different values

1. Without intercept

$$y = \qquad \beta_1 x_1 + \beta_2 x_2 + \beta_3 \cdot x_3 + \beta_4 \cdot x_4 + \epsilon$$

2. With intercept

$$y = \beta_0 \qquad + \beta_2 x_2 + \beta_3 \cdot x_3 + \beta_4 \cdot x_4 + \epsilon$$

These two equivalent forms of linear regression will produce **identical** predictions, but the coefficients will have different interpretations.

Let's now examine it using the example of wages and marital status (married vs non-married) as an example

Exampe: `mv06_cond_expectation02.Rmd`

Lessons Learned

1. Use of a discrete predictor, two approaches deliver the same predictions
   (a) An indicator for all possible values: directly informing **predictions**
   (b) Omitting one of them but include an intercept: directly informing **partial effects**
2. Linear regressions are equilvalent to find conditional means!

Why does it work?

Suppose that

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \epsilon$$

Regression coefficients are those, $\beta_1, \beta_2, \ldots, \beta_k$, that solve the following minimization problem.

$$\min \quad \mathbb{E}[(y - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k))^2]$$

$

Intuition: The linear regression coefficient is the best **linear** prediction in the sense that it minimizes mean squared error!

Note that when you choose $\beta_1, \beta_2, \ldots, \beta_k$, you are choosing a function, $m(x)$, but a linear one!

Note that when you choose $\beta_1, \beta_2, \dots, \beta_k$, you are choosing a function, $m(x)$, but a linear one!

Regression coefficients are those, $\beta_1, \beta_2, \dots, \beta_k$, that solve the following minimization problem.

$$\min \quad \mathbb{E}[(y - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k))^2]$$

Conditional Expectation, $\mathbb{E}[y|x]$ solves the following minimization problem.

$$\min \quad \mathbb{E}[(y - m(x))^2]$$

If the conditional expectation function, $\mathbb{E}[y|x]$,is indeed linear, and of course, the regression coefficient is just the CEF!!

In other words, **if the CEF is linear**, then you can just use regression to obtain the CEF!

Now, let's go through the previous examples and see why the CEFs are linear.

Prior to continuing, it is worth pointing out that linearity refers to linearity in coefficients.

Prior to continuing, it is worth pointing out that linearity refers to linearity in coefficients.

Examples of Linear Models:

1. $y = \beta_0 + \beta_1 x_1 + \epsilon$

Prior to continuing, it is worth pointing out that linearity refers to linearity in coefficients.

Examples of Linear Models:

1. $y = \beta_0 + \beta_1 x_1 + \epsilon$
2. $y = \beta_1 x_1 + \beta_2 x_2 + \epsilon$

Prior to continuing, it is worth pointing out that linearity refers to linearity in coefficients.

Examples of Linear Models:

1. $y = \beta_0 + \beta_1 x_1 + \epsilon$
2. $y = \beta_1 x_1 + \beta_2 x_2 + \epsilon$
3. $y = \beta_1 x_1 + \beta_2 \exp(x_2) + \beta_3 x_1^{10} + \epsilon$

As long as we can write it in the following form

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \cdots + \beta_k \cdot x_k + \epsilon$$

Examples of **non-linear** regression models

1. $y = \beta_0 + \beta_1 \cdot \beta_0 \cdot x_1 + \epsilon$

Examples of **non-linear** regression models

1. $y = \beta_0 + \beta_1 \cdot \beta_0 \cdot x_1 + \epsilon$
2. $y = \exp(\beta_0 + \beta_1 \cdot x_1 + \epsilon)$

Now, lets see why the previous examples indeed have a linear CEF

$$\mathbb{E}[\textit{wages}|\text{married}] = \begin{cases} 798.4400 & \text{if married } = 0 \\ 977.0479 & \text{if married } = 1 \end{cases}$$

$\Longrightarrow$

Now, lets see why the previous examples indeed have a linear CEF

$$\mathbb{E}[\textit{wages}|\text{married}] = \begin{cases} 798.4400 & \text{if married } = 0 \\ 977.0479 & \text{if married } = 1 \end{cases}$$

$\Longrightarrow$

$$\mathbb{E}[\text{wages}|\text{married}] = 798.4400 \cdot (1 - \text{married}) + 977.0479 \cdot \text{married}$$

Now, lets see why the previous examples indeed have a linear CEF

$$\mathbb{E}[\textit{wages}|\text{married}] = \begin{cases} 798.4400 & \text{if married } =0 \\ 977.0479 & \text{if married } =1 \end{cases}$$

$\Longrightarrow$

$$\begin{aligned} \mathbb{E}[\text{wages}|\text{married}] &= 798.4400 \cdot (1 - \text{married}) + 977.0479 \cdot \text{married} \\ &= \gamma_1 x_0 + \gamma_1 x_1 \end{aligned}$$

Now, lets see why the previous examples indeed have a linear CEF

$$\mathbb{E}[\textit{wages}|\text{married}] = \begin{cases} 798.4400 & \text{if married } =0 \\ 977.0479 & \text{if married } =1 \end{cases}$$

$\Longrightarrow$

$$\begin{aligned} \mathbb{E}[\text{wages}|\text{married}] &= 798.4400 \cdot (1 - \text{married}) + 977.0479 \cdot \text{married} \\ &= \gamma_1 x_0 + \gamma_1 x_1 \end{aligned}$$

$$\mathbb{E}[\text{wages}|\text{married}] = 798.4400 + (977.0479 - 798.4400) \cdot \text{married}$$

Now, lets see why the previous examples indeed have a linear CEF

$$\mathbb{E}[\mathit{wages}|\text{married}] = \begin{cases} 798.4400 & \text{if married } = 0 \\ 977.0479 & \text{if married } = 1 \end{cases}$$

$$\Longrightarrow$$

$$\begin{aligned} \mathbb{E}[\text{wages}|\text{married}] &= 798.4400 \cdot (1 - \text{married}) + 977.0479 \cdot \text{married} \\ &= \gamma_1 x_0 + \gamma_1 x_1 \end{aligned}$$

$$\begin{aligned} \mathbb{E}[\text{wages}|\text{married}] &= 798.4400 + (977.0479 - 798.4400) \cdot \text{married} \\ &= 798.4400 + 178.6079 \cdot \text{married} \end{aligned}$$

Now, lets see why the previous examples indeed have a linear CEF

$$\mathbb{E}[\textit{wages}|\text{married}] = \begin{cases} 798.4400 & \text{if married } =0 \\ 977.0479 & \text{if married } =1 \end{cases}$$

$$\Longrightarrow$$

$$\begin{aligned}\mathbb{E}[\text{wages}|\text{married}] &= 798.4400 \cdot (1 - \text{married}) + 977.0479 \cdot \text{married} \\ &= \gamma_1 x_0 + \gamma_1 x_1\end{aligned}$$

$$\begin{aligned}\mathbb{E}[\text{wages}|\text{married}] &= 798.4400 + (977.0479 - 798.4400) \cdot \text{married} \\ &= 798.4400 + 178.6079 \cdot \text{married} \\ &= \beta_0 + \beta_1 \cdot x_1\end{aligned}$$

These two models are equilvalent but the interpretation of the coefficients has changed.

1. Group averages from each sub-group
2. Partial Effects: Everything relative to the base group. Group averages for married people are the sum of group averages for non-married people and the difference in group averages (i.e., marriage premium).

By Decomposition Property:

$$y = \mathbb{E}[y|x] + \epsilon$$

We can write our wage equation as follows

$$\text{wage} = \gamma_1 \cdot \text{non-married} + \gamma_2 \cdot \text{married} + \epsilon$$

$$\text{wage} = \beta_0 + \beta_1 \cdot \text{married} + \epsilon$$

Lets look at the education example

$$\mathbb{E}[\text{wage}|\text{education}] = \begin{cases} 774.2500 & \text{if educ } = \text{below high school} \\ 862.6718 & \text{if educ } = \text{high school} \\ 1076.0242 & \text{if educ } = \text{above high school} \end{cases}$$

Lets look at the education example

$$\mathbb{E}[\text{wage}|\text{education}] = \begin{cases} 774.2500 & \text{if educ } = \text{below high school} \\ 862.6718 & \text{if educ } = \text{high school} \\ 1076.0242 & \text{if educ } = \text{above high school} \end{cases}$$

We can write our model as follows:

$$\begin{aligned} \text{wage} &= 774.2500 \cdot \text{below high school} + 862.6718 \cdot \text{high school} \\ &\quad + 1076.0242 \cdot \text{above high school} + \epsilon \\ &= \beta_1 \cdot \text{below high school} + \beta_2 \cdot \text{high school} \\ &\quad + \beta_3 \cdot \text{above high school} + \epsilon \end{aligned}$$

Lets look at the education example

$$\mathbb{E}[\text{wage}|\text{education}] = \begin{cases} 774.2500 & \text{if educ} = \text{below high school} \\ 862.6718 & \text{if educ} = \text{high school} \\ 1076.0242 & \text{if educ} = \text{above high school} \end{cases}$$

We can write our model as follows:

$$\begin{aligned} \text{wage} &= 774.2500 \cdot \text{below high school} + 862.6718 \cdot \text{high school} \\ &\quad + 1076.0242 \cdot \text{above high school} + \epsilon \\ &= \beta_1 \cdot \text{below high school} + \beta_2 \cdot \text{high school} \\ &\quad + \beta_3 \cdot \text{above high school} + \epsilon \end{aligned}$$

Or, you can similarly manipulate it into another form of conditional expectation function:

$$\text{wage} = 774.2500 + 88.42 \cdot \text{high school} + 301.77 \cdot \text{above high school} + \epsilon$$

key result:

The conditional mean function is linear whenever all regressors are discrete (i.e., take on only a finite number of possible values).

Example: What is the relationship between marital status, races and wages?

When is linear regression not working? And what to do?

Motivating Example

Let's look at our example code first:
`mv06_cond_expectation04.Rmd`

Two Lessons:

1. The conditional mean is better. *Prediction Property*
2. The linear regression result is not too far off (We will have a name for this!).

Now, lets try to understand why our linear regression is not delivering the best predictions or our conditional means.

$$\mathbb{E}[\text{wages}|\text{married,black}] = \begin{cases} 841.9756 & \text{if unmarried non-blacks} \\ 1007.2797 & \text{if married non-blacks} \\ 600.1111 & \text{if unmarried blacks} \\ 759.7941 & \text{if married blacks} \end{cases}$$

From last semester, If Q is a statement, such as "Last name is Wang", then $\mathbb{I}[Q]$ is set to 1 when the statement Q is true, and zero otherwise. For an event

$$A \in \mathcal{F}\mathbb{I}[\omega \in A] = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases}$$

Properties of Indicator Function: If $A_1, A_2, \ldots A_N$, then

1. $\mathbb{I}[\cap A_N] = \Pi_{n=1}^{N} \mathbb{I}[A_n]$
2. $\mathbb{I}[\cup A_N] = \sum_{n=1}^{N} \mathbb{I}[A_n]$ whenever the sets are disjoint

$$\text{unmarried, non-blacks:} \quad \mathbb{I}[\text{married} = 0 \text{ and black} = 0]$$
$$= (1 - \text{married}) \cdot (1 - \text{black})$$

$$\begin{aligned} \text{unmarried, non-blacks:} \quad & \mathbb{I}[\text{married } = 0 \text{ and black} = 0] \\ & = (1 - \text{married}) \cdot (1 - \text{black}) \end{aligned}$$

To see why,

1. If married=0 and black=0, $(1 - \text{married}) \cdot (1 - \text{black}) = 1$

$$\text{unmarried, non-blacks:} \quad \mathbb{I}[\text{married} = 0 \text{ and black} = 0]$$
$$= (1 - \text{married}) \cdot (1 - \text{black})$$

To see why,

1. If married=0 and black=0, $(1 - \text{married}) \cdot (1 - \text{black}) = 1$
2. If either married $\neq 0$ or black $\neq 0$,
   $(1 - \text{married}) \cdot (1 - \text{black}) = 0$

Similarly, we can write all the possible values in the following forms

$$\begin{aligned}\text{married,non-blacks} &= \mathbb{I}[\text{married}{=}1,\text{black}{=}0] \\ &= \text{married} \cdot (1 - \text{black})\end{aligned}$$

Similarly, we can write all the possible values in the following forms

$$\begin{aligned}\text{married,non-blacks} &= \mathbb{I}[\text{married=1,black=0}] \\ &= \text{married} \cdot (1 - \text{black})\end{aligned}$$

$$\begin{aligned}\text{non-married,blacks} &= \mathbb{I}[\text{married=0,black=1}] \\ &= (1 - \text{married}) \cdot \text{black}\end{aligned}$$

Similarly, we can write all the possible values in the following forms

$$\begin{aligned}
\text{married,non-blacks} &= \mathbb{I}[\text{married=1,black=0}] \\
&= \text{married} \cdot (1 - \text{black})
\end{aligned}$$

$$\begin{aligned}
\text{non-married,blacks} &= \mathbb{I}[\text{married=0,black=1}] \\
&= (1 - \text{married}) \cdot \text{black}
\end{aligned}$$

$$\begin{aligned}
\text{married,blacks} &= \mathbb{I}[\text{married=1,black=1}] \\
&= \text{married}\cdot\text{black}
\end{aligned}$$

$$\mathbb{E}\left[\text{wages}|\text{married,black}\right] = \begin{cases} \mu_{00} = 841.9756 & \text{if married } = 0 \text{ and black } = 0 \\ \mu_{10} = 1007.2797 & \text{if married } = 1 \text{ and black } = 0 \\ \mu_{01} = 600.1111 & \text{if married } = 0 \text{ and black } = 1 \\ \mu_{11} = 759.7941 & \text{if married } = 0 \text{ and black } = 1 \end{cases}$$

$$\mathbb{E}\Big[\text{wages}|\text{married,black}\Big] = \begin{cases} \mu_{00} = 841.9756 & \text{if married} = 0 \text{ and black} = 0 \\ \mu_{10} = 1007.2797 & \text{if married} = 1 \text{ and black} = 0 \\ \mu_{01} = 600.1111 & \text{if married} = 0 \text{ and black} = 1 \\ \mu_{11} = 759.7941 & \text{if married} = 0 \text{ and black} = 1 \end{cases}$$

$\Longrightarrow$

$$\begin{aligned} \mathbb{E}[\text{wages}|\text{married,black}] &= \mu_{00} \cdot (1 - \text{married}) \cdot (1 - \text{black}) \\ &+ \mu_{10} \cdot \text{married} \cdot \text{(1-black)} \\ &+ \mu_{01} \cdot (1 - \text{married}) \cdot \text{black} \\ &+ \mu_{11} \cdot \text{married} \cdot \text{black} \\ \\ &= \mu_{00} \cdot x_1 + \mu_{10} \cdot x_2 \\ &+ \mu_{01} \cdot x_3 + \mu_{11} \cdot x_4 \end{aligned}$$

where x_1 can be thought of as a new variable $(1 - \text{married})(1 - \text{black})$, other xs are simiarly defined based on the variables in the equation.

Let's use this insight to implement the correct linear regression

Lets look at our example: `mv06_cond_expectation04.Rmd`

Why do I want to go through all the troubles to manipulate my four binary variables into the products between married and black?

We can show the missing part in the original regression model when you include only two variables (married and black).

Let's manipulate this in more natural form that you guys are used to:

$$
\begin{aligned}
\mathbb{E}[\text{wages}|\text{married,black}] &= \mu_{00} \cdot (1 - \text{married}) \cdot (1 - \text{black}) \\
&+ \mu_{10} \cdot \text{married} \cdot (\text{1-black}) \\
&+ \mu_{01} \cdot (1 - \text{married}) \cdot \text{black} \\
&+ \mu_{11} \cdot \text{married} \cdot \text{black}
\end{aligned}
$$

Let's manipulate this in more natural form that you guys are used to:

$$
\begin{aligned}
\mathbb{E}[\text{wages}|\text{married,black}] &= \mu_{00} \cdot (1 - \text{married}) \cdot (1 - \text{black}) \\
&+ \mu_{10} \cdot \text{married} \cdot (\text{1-black}) \\
&+ \mu_{01} \cdot (1 - \text{married}) \cdot \text{black} \\
&+ \mu_{11} \cdot \text{married} \cdot \text{black} \\
&= \mu_{00}[1 - \text{black} - \text{married}+\text{married}\cdot\text{black}] \\
&+ \mu_{10}[\text{married} - \text{married} \cdot \text{black}] \\
&+ \mu_{01}[\text{black} - \text{married} \cdot \text{black}] + \mu_{11} \cdot \text{married} \cdot \text{black}
\end{aligned}
$$

Let's manipulate this in more natural form that you guys are used to:

$$
\begin{aligned}
\mathbb{E}[\text{wages}|\text{married,black}] &= \mu_{00} \cdot (1 - \text{married}) \cdot (1 - \text{black}) \\
&+ \mu_{10} \cdot \text{married} \cdot (\text{1-black}) \\
&+ \mu_{01} \cdot (1 - \text{married}) \cdot \text{black} \\
&+ \mu_{11} \cdot \text{married} \cdot \text{black} \\
&= \mu_{00}[1 - \text{black} - \text{married} + \text{married} \cdot \text{black}] \\
&+ \mu_{10}[\text{married} - \text{married} \cdot \text{black}] \\
&+ \mu_{01}[\text{black} - \text{married} \cdot \text{black}] + \mu_{11} \cdot \text{married} \cdot \text{black} \\
&= \mu_{00} + (-\mu_{00} + \mu_{01}) \cdot \text{black} + (-\mu_{00} + \mu_{10}) \cdot \text{married} \\
&+ (\mu_{00} - \mu_{10} - \mu_{01} + \mu_{11})\text{married} \cdot \text{black}
\end{aligned}
$$

Let's manipulate this in more natural form that you guys are used to:

$$
\begin{aligned}
\mathbb{E}[\text{wages}|\text{married,black}] &= \mu_{00} \cdot (1 - \text{married}) \cdot (1 - \text{black}) \\
&+ \mu_{10} \cdot \text{married} \cdot \text{(1-black)} \\
&+ \mu_{01} \cdot (1 - \text{married}) \cdot \text{black} \\
&+ \mu_{11} \cdot \text{married} \cdot \text{black} \\
&= \mu_{00}[1 - \text{black} - \text{married} + \text{married} \cdot \text{black}] \\
&+ \mu_{10}[\text{married} - \text{married} \cdot \text{black}] \\
&+ \mu_{01}[\text{black} - \text{married} \cdot \text{black}] + \mu_{11} \cdot \text{married} \cdot \text{black} \\
&= \mu_{00} + (-\mu_{00} + \mu_{01}) \cdot \text{black} + (-\mu_{00} + \mu_{10}) \cdot \text{married} \\
&+ (\mu_{00} - \mu_{10} - \mu_{01} + \mu_{11})\text{married} \cdot \text{black} \\
&= \beta_0 + \beta_1 \cdot \text{black} + \beta_2 \cdot \text{married} + \beta_3 \cdot \text{married} \cdot \text{black}
\end{aligned}
$$

By Decomposition Property of the CEF, the outcome variable should be written as follows

$$\text{wage} == \beta_0 + \beta_1 \cdot \text{black} + \beta_2 \cdot \text{married} + \beta_3 \cdot \text{married} \cdot \text{black} + \epsilon$$

Let's use this insight to implement the correct linear regression

Lets look at our example: `mv06_cond_expectation04.Rmd`

Economic intuition behind what is implied the **initial linear regression model:**

$$\text{wages} = \gamma_0 + \gamma_1 \cdot \text{married} + \gamma_2 \cdot \text{black} + \epsilon$$

Economic intuition behind what is implied the **initial linear regression model:**

$$\text{wages} = \gamma_0 + \gamma_1 \cdot \text{married} + \gamma_2 \cdot \text{black}{+}\epsilon$$

1. For blacks: wage difference between married and non-married

workers:

$$(\gamma_0 + \gamma_1 \cdot 1 + \gamma_2 \cdot 1 + \epsilon) - (\gamma_0 + \gamma_1 \cdot 0 + \gamma_2 \cdot 1 + \epsilon) = \gamma_1$$

Economic intuition behind what is implied the **initial linear regression model:**

$$\text{wages} = \gamma_0 + \gamma_1 \cdot \text{married} + \gamma_2 \cdot \text{black} + \epsilon$$

1. For blacks: wage difference between married and non-married workers:

$$(\gamma_0 + \gamma_1 \cdot 1 + \gamma_2 \cdot 1 + \epsilon) - (\gamma_0 + \gamma_1 \cdot 0 + \gamma_2 \cdot 1 + \epsilon) = \gamma_1$$

2. For non-blacks: wage difference between married and non-married workers:

$$(\gamma_0 + \gamma_1 \cdot 1 + \gamma_2 \cdot 0 + \epsilon) - (\gamma_0 + \gamma_1 \cdot 0 + \gamma_2 \cdot 0 + \epsilon) = \gamma_1$$

Economic intuition behind what is implied the **initial linear regression model:**

$$\text{wages} = \gamma_0 + \gamma_1 \cdot \text{married} + \gamma_2 \cdot \text{black} + \epsilon$$

1. For blacks: wage difference between married and non-married

workers:

$$(\gamma_0 + \gamma_1 \cdot 1 + \gamma_2 \cdot 1 + \epsilon) - (\gamma_0 + \gamma_1 \cdot 0 + \gamma_2 \cdot 1 + \epsilon) = \gamma_1$$

2. For non-blacks: wage difference between married and non-married workers:

$$(\gamma_0 + \gamma_1 \cdot 1 + \gamma_2 \cdot 0 + \epsilon) - (\gamma_0 + \gamma_1 \cdot 0 + \gamma_2 \cdot 0 + \epsilon) = \gamma_1$$

Why does marriage premium have to be the same across races?

Economic intuition behind what is implied by **the CEF model:**

$$\text{wages} = \beta_0 + \beta_1 \cdot \text{married} + \beta_2 \cdot \text{black} + \beta_3 \cdot \text{married}{\cdot}\text{black} + \epsilon$$

Economic intuition behind what is implied by **the CEF model:**

$$\text{wages} = \beta_0 + \beta_1 \cdot \text{married} + \beta_2 \cdot \text{black} + \beta_3 \cdot \text{married}{\cdot}\text{black} + \epsilon$$

For blacks: wage difference between married and non-married workers:

Economic intuition behind what is implied by **the CEF model:**

$$\text{wages} = \beta_0 + \beta_1 \cdot \text{married} + \beta_2 \cdot \text{black} + \beta_3 \cdot \text{married}{\cdot}\text{black} + \epsilon$$

For blacks: wage difference between married and non-married workers:

$$(\beta_0{+}\beta_1{\cdot}1{+}\beta_2{\cdot}1{+}\beta_3{\cdot}1{\cdot}1{+}\epsilon){-}(\beta_0{+}\beta_1{\cdot}0{+}\beta_2{\cdot}1{+}\beta_3{\cdot}0{\cdot}1{+}\epsilon) = \beta_1{+}\beta_3$$

Economic intuition behind what is implied by **the CEF model:**

$$\text{wages} = \beta_0 + \beta_1 \cdot \text{married} + \beta_2 \cdot \text{black} + \beta_3 \cdot \text{married}{\cdot}\text{black} + \epsilon$$

For blacks: wage difference between married and non-married workers:

$$(\beta_0{+}\beta_1{\cdot}1{+}\beta_2{\cdot}1{+}\beta_3{\cdot}1{\cdot}1{+}\epsilon){-}(\beta_0{+}\beta_1{\cdot}0{+}\beta_2{\cdot}1{+}\beta_3{\cdot}0{\cdot}1{+}\epsilon) = \beta_1{+}\beta_3$$

For non-blacks: wage difference between married and non-married workers:

Economic intuition behind what is implied by **the CEF model:**

$$\text{wages} = \beta_0 + \beta_1 \cdot \text{married} + \beta_2 \cdot \text{black} + \beta_3 \cdot \text{married}{\cdot}\text{black} + \epsilon$$

For blacks: wage difference between married and non-married workers:

$$(\beta_0+\beta_1{\cdot}1+\beta_2{\cdot}1+\beta_3{\cdot}1{\cdot}1+\epsilon)-(\beta_0+\beta_1{\cdot}0+\beta_2{\cdot}1+\beta_3{\cdot}0{\cdot}1+\epsilon) = \beta_1+\beta_3$$

For non-blacks: wage difference between married and non-married workers:

$$(\beta_0+\beta_1\cdot 1+\beta_2\cdot 0+\beta_3\cdot 1{\cdot}0+\epsilon)-(\beta_0+\beta_1\cdot 0+\beta_2\cdot 0+\beta_3\cdot 0{\cdot}0+\epsilon) = \beta_1$$

Truth: (average) marriage premium varies across races!!

Heterogenous Partial Effects and Interactions

Including only married and black in your model artificially impose the assumption that the partial effect is the same across groups.

Heterogenous Partial Effects (across groups) are captured by the interactions between the main variable of interest and the group variable that you think the partial effects may vary with.

For example,

1. The effects of education on wages may depend on the education level itself (decreasing marginal returns)

$$\text{wages} = \beta_0 + \beta_1 \cdot \text{educ} + \beta_2 \cdot \text{educ} \cdot \text{educ} + \epsilon$$

2. Synergy effects in marketing: The effects of spending on TV advertising on sales may depend on the level of spending on radio advertising, too!

$$textsales = \beta_0 + \beta_1 \cdot \text{TV expenditures} + \beta_2 \cdot \text{Radio expenditures} + \beta_3 \cdot \text{TV} \cdot \text{Radi}$$

Fact:

In general, if there are p dummy variables, $x_{1,}x_2, \ldots, x_p$, then the CEF $\mathbb{E}[y|x_1, x_2, \ldots, x_p]$ takes at most 2^p distinct values, and can be written as a linear function of the 2^p regressors including $x_1, x_2, \ldots, x_p$ and all cross products. Cannot be practically!

$$\begin{aligned}\mathbb{E}[y|x_1, x_2, x_3] = &\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 \\ &+ \beta_5 x_1 x_3 + \beta_6 x_2 x_3 \\ &+ \beta_7 x_1 x_2 x_3\end{aligned}$$

Let's use this insight to implement the correct linear regression

Lets look at our example: `mv06_cond_expectation04.Rmd`

Summary

1. The conditional mean is the best (in terms of mean squared error)

Summary

1. The conditional mean is the best (in terms of mean squared error)
2. The conditional mean is linear (in parameters) when all the variables are discrete and take on finite values.

Summary

1. The conditional mean is the best (in terms of mean squared error)
2. The conditional mean is linear (in parameters) when all the variables are discrete and take on finite values.
   (a) In general, 2^p possible regressors, including $x_1, x_2, \ldots, x_p$ and all cross products

Summary

1. The conditional mean is the best (in terms of mean squared error)
2. The conditional mean is linear (in parameters) when all the variables are discrete and take on finite values.
   (a) In general, 2^p possible regressors, including $x_1, x_2, \ldots, x_p$ and all cross products
3. Two forms of the conditional mean

Summary

1. The conditional mean is the best (in terms of mean squared error)
2. The conditional mean is linear (in parameters) when all the variables are discrete and take on finite values.
   (a) In general, 2^p possible regressors, including $x_1, x_2, \ldots, x_p$ and all cross products
3. Two forms of the conditional mean

(a) Prediction form: Include dummy variables for all possible levels (exclusive groups), but omit an intercept
(b) Partial Effects form: Include an intercept and omit one of the dummy variables. The partial effects is the group difference.

Summary

1. The conditional mean is the best (in terms of mean squared error)
2. The conditional mean is linear (in parameters) when all the variables are discrete and take on finite values.
   (a) In general, 2^p possible regressors, including $x_1, x_2, \ldots, x_p$ and all cross products
3. Two forms of the conditional mean

(a) Prediction form: Include dummy variables for all possible levels (exclusive groups), but omit an intercept
(b) Partial Effects form: Include an intercept and omit one of the dummy variables. The partial effects is the group difference.

4. In the case of all discrete variables, linear regression delivers the Conditional Mean.
   (a) Straightforward case: binary or multi-valued cases
   (b) Less straightforward case: two or more than two binary variables (interactions)