Why do we use conditional expectation vs regular expectation in regression?

Ask Question

Asked 7 years, 9 months ago Active 2 years, 7 months ago Viewed 4k times

4

What is the significance of using conditional expectation $E(Y \mid X = x)$ in regression and not just use the regular expectation $E(Y)$. What will be the consequence if we use $E(Y)$ instead? Can we do this, since we know one of the assumptions of regression is that X is a known quantity and it is fixed?

Given the following

$$Y = a + bX + e$$

and that $E(e) = 0$, why is the following wrong?

$$E(Y) = a + bX$$

I may not have fully understood conditional expectation. So, any comments on the difference between expectation and conditional expectation will definitely help.

regression probability mathematical-statistics expected-value

Share Cite Improve this question Follow

edited Jan 20 '19 at 12:59 Sven Hohenstein 6,187 25 28 37

asked Nov 11 '13 at 5:09 leviathan 765 4 9 19

What is Teams?

1 The data X is itself a random variable in general. $E(Y|X)$ is the Hilbert space projection of Y onto "functions of X", thus is the best one can do in the L^2-sense. – Michael Nov 11 '13 at 5:18

Add a comment

3 Answers

Active Oldest Votes

6

$E(Y)$ is just the mean of your responses: it's the same thing as a regression where all you have is an intercept. For all values of X, you are predicting the same value for the response. We use conditional expectation because we expect there to be a relationship between a predictor variable and the response variable, such that we want our predictions to be made in the context of a specific value of the predictor(s).

Consider the following prediction task: you have logged the distance your car has traveled between fillups and the amount of gas you put in the tank each time, so you have data for the MPG of your car. Sometimes you fill up the tank when it's nearly empty, sometimes you just top it off. Let's say you want to use your data to predict how much gasoline you will expend on a particular trip. You regress your response variable of gasoline volume against the predictor variable of distance and get a model of the form $gas = b_0 + b_1 distance$.

If you were to use $E[y]$ for your predictions here, you would calculate the mean gas volume for all trips and say "regardless of the distance we're going to travel, I anticipate that we will use $E[gas]$ gallons of fuel." But, because we have the regression model, we can say "given that we will travel d miles, we expect to use $E[gas|distance = d] = b_0 + b_1 d$ gallons of fuel."

If most of your trips are about the same distance, just using $E[gas]$ might actually give you decent results. But given that we know there is a relationship between distance traveled and gas used, we anticipate that the conditional expectation will give us much better predictions than the blanket prediction of the unconditioned expectation.

There are two basic versions of a regression specification, regarding the assumptions about the nature of the variables involved:

In the first, the regressors are assumed *deterministic*, and so the actual matrix contains elements that are not realizations of random variables. This comes from an "experimental design" tradition (and hence sometimes the regressor matrix is called the "design matrix"). In such a set up the regressors are not random variables and so it is not meaningful to "condition" on them. So we have to formulate the model in terms of unconditional expected values, like the ones OP wrote.

The second version mainly relates to a non-experimental setting, like economics for example, we treat the regressors as random variables, while the specific sample/regressor matrix as being comprised of realizations of random variables. Here, we condition on X to be humble: we do not claim that we know the unconditional behavior of Y or of the error, but we make "narrower" assumptions, conditional assumptions. Hence $E(u \mid X) = 0$ etc. Intuitively we restrict ourselves to "what can I say about the relation between Y and X, given the specific data that I have in hand?" Apart for humbleness, it also cleverly permits us to derive all sorts of useful results (which we could *not* if we assumed stochastic regressors *and* considered at the same time assumptions on unconditional expectations).

For example, to stick to basics, in proving unbiasedness of the OLS estimator we write

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}'\beta + \mathbf{u}) = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$$

Assume that we treat the regressors as stochastic. If you want to check unbiasedness, you will examine

$$E(\hat{\beta}) = \beta + E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\right)$$

Where does that leave you? Nowhere - it is difficult to justify an a priori assumption like $E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\right) = 0$. But note that, by the properties of expected values

$$E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\right) = E\left[E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\right) \mid X\right]$$

$$= E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E\left(\mathbf{u} \mid X\right)\right]$$

Now, to have unbiasedness, we only need to a priori assume that $E\left(\mathbf{u} \mid X\right) = 0$, which can at least be logically argued upon.

Finally, note that there is a drawback in assuming deterministic, non-stochastic regressors, which is sometimes overlooked: If our model is

$$y_i = a + bx_i + u_i, \quad i = 1, .., n$$

then under non-stochastic regressors we have

$$E(y_i) = a + bx_i$$

But this means that each y_i has in principle a *different expected value* : so the y_i's here do not come from an identically distributed population. If they don't, then our sample $\{Y, X\}$, that contains as a

But this means that each y_i has in principle a *different expected value* : so the y_i's here do not come from an identically distributed population. If they don't, then our sample $\{Y, X\}$, that contains as a random variable only the Y is not "random" (i.e. it is not i.i.d), due to the assumption that the X's are deterministic.

Share Cite Improve this answer Follow

answered Nov 11 '13 at 17:04

Alecos Papadopoulos
52k 4 126 235

Add a comment

1

Let's say you observe a RV X and you want to predict a second RV Y where your predictor of Y is $g(X)$.

Then you can prove that the "best" predictor of Y is $g(X) = E[Y \mid X]$, where "best" is the function g that minimizes $E[(Y - g(X))^2]$.

So $E[Y \mid X]$ will always be at least as good as, if not better than $E[Y]$.

To prove that $E[Y|X]$ is the "best" predictor possible we just need to show that

$$E[(Y - g(X))^2] \geq E[(Y - E[Y \mid X])^2]$$

Let's start off by looking at $E[(Y - g(X))^2 \mid X]$ which is not quite the same as the LHS above but not to worry, at the end we can leverage the property that $E[E[X|Y]] = E[X]$. So,

$$\begin{aligned} E[(Y - g(X))^2 \mid X] &= E[(Y - E[Y \mid X] + E[Y \mid X] - g(X))^2 \mid X] \\ &= E[(Y - E[Y \mid X]))^2 \mid X] \\ &+ E[(E[Y \mid X] - g(X))^2 \mid X] \\ &+ 2E[(Y - E[Y \mid X])(E[Y \mid X] - g(X)) \mid X] \end{aligned}$$

Where the above is basically doing

$$(x + y)^2 = (x - c + c + y)^2 = (x - c)^2 + (x + c)^2 + 2(x - c)(x + c)$$

and using the linearity of expectations.

Focusing on the third term of the expansion only, we can show that it equals zero.

$$2E[(Y - E[Y \mid X])(E[Y \mid X] - g(X)) \mid X] = 0$$

This is because given X, the term $E[Y \mid X] - g(X)$ is like a constant and can be pulled out of the expectation using the property that $E[aX] = aE[X]$. Letting the constant we pulled out equal c we have

$$\begin{aligned} c \cdot E[Y - E[Y \mid X] \mid X] &= c \cdot (E[Y \mid X] - E[E[Y \mid X] \mid X]) \\ &= c \cdot (E[Y \mid X] - E[Y \mid X]) = 0 \end{aligned}$$

From the property $E[g(X) \mid X] = g(X)$ we know that $E[E[Y \mid X] \mid X] = E[Y \mid X]$ since $E[Y \mid X]$ is a function of X.

Therefore going back to the expansion, we have

$$E[(Y - g(X))^2 \mid X] = E[(Y - E[Y \mid X]))^2 \mid X] + E[(E[Y \mid X] - g(X))^2 \mid X]$$

Noting that the second term on the RHS must always be positive, we can say

$$E[(Y - g(X))^2 \mid X] \geq E[(Y - E[Y \mid X]))^2 \mid X]$$

and then taking the expected value of both sides, leveraging the property that $E[X] = E[E[X \mid Y]]$, we get

What should I do during an interview when it feels like it is going downhill?
Is publishing in a journal considered "commercial
What should I do during an interview when it feels like it is going downhill?
Is publishing in a journal considered "commercial use"?
What is a "Safe Space"?
Would being a co-applicant on a loan have any impact on minimum payment each month?
Question feed

$$E[(Y-g(X))^2] \geq E[(Y-E[Y \mid X])^2]$$

Which shows that the "best" predictor is $E[Y \mid X]$. Note that it will be as good as $E[Y]$ if Y and X are independent.

Share Cite Improve this answer Follow

answered Sep 7 '18 at 23:17
HJ_beginner
161 1 4

Add a comment

Your Answer

Sign up or log in

Sign up using Google

Sign up using Facebook

Sign up using Email and Password

Post as a guest

Name

Email
Required, but never shown

Post Your Answer *By clicking "Post Your Answer", you agree to our terms of service, privacy policy and cookie policy*

Not the answer you're looking for? Browse other questions tagged regression probability mathematical-statistics expected-value or ask your own question.