Conditional Mean in Linear Regression

Ask Question

Asked 5 years, 6 months ago Active 4 years, 4 months ago Viewed 8k times

9

I have a question about linear regression in general. Suppose we have the following Data Generating Process:

$$y_i = x_i\beta + \epsilon_i$$

Now, the thing is that from my understanding, each observation is a random variable, i.e. each y_i has a conditional mean, that is:

$$E[y_i|x_i] = x_i\beta$$

under exogeneity assumptions. My question is as follows. What is the unconditional mean of y? Also, when we have a sample of n data points, what really does $E[y]$ mean? Is it the mean of the random vector of $y_i's$? How does the law of iterated expectations apply here?

regression conditional-expectation

Share Cite Improve this question Follow

edited Jan 27 '16 at 16:08
Gavin Simpson
35.6k 5 105 150

asked Jan 27 '16 at 15:09
ChinG
711 8 22

Add a comment

What is Teams?

2 Answers

Active Oldest Votes

8

It's important to be precise in these situations, and distinguish between the data model, and the data itself. One way to think about linear regression is that we hypothesize the following relationship on the unknowable statistical process that generated the data we do have

$$E[Y \mid X] = \beta_0 + X\beta$$

Beta is an unknown constant at this point, so we are just setting down a hypothesis on what we believe the *shape* of the relationship is like. Then, given the data, we use some method to determine what β should be so that the hypothesized relationship is likely to generate the data we do have (maximum likelihood being very popular).

Even without knowing β, we can manipulate the relationship to learn some things about the consequences of our assumptions

$$E[Y] = E[E[Y \mid X]] = \beta_0 + \beta E[X] = \beta_0 + \beta E[X]$$

Now, the distribution of X is generally *not* part of our structural assumptions in regression, so, in general, this is as far as we can go.

Oftentimes, we will center our data for X, which imposes the constraint $E[X] = 0$ on our model. In this case, we can derive

$$E[Y] = \beta_0$$

This is why, for example this book recommends centering predictors (in some situations) so that the model intercept is interpretable.

> Now, my question is how is this related to the sample average of y?

If you fit the model by least squares, and you have centered the predictor x, then the model intercept is the sample average.

Geometrically, the least squares line must pass through the center of mass of the data $(\bar{x}, \bar{y})$. When you have centered x, $\bar{x} = 0$, so the line passes through $(0, \bar{y})$. If you plug these values into the model equation, you get $\beta_0 = \bar{y}$.

Algebraically, the least squares equation is is $(X^t X)\vec{\beta} = X^t y$. If you think about the matrix X, the first column is all ones (the intercept column), and since x is centered, this intercept column is orthogonal to the data column. This means that the first row of $X^t X$ looks like $(N, 0)$ (where N is the number of data points). Then first component of the left hand side is $N\beta_0$. On the right hand
orthogonal to the data column. This means that the first row of $X^t X$ looks like $(N, 0)$ (where N is the number of data points). Then first component of the left hand side is $N\beta_0$. On the right hand
side, the first component is $\sum_i y_i$. Equating them, you get the result $\beta_0 = \bar{y}$.

It is also true that the mean of the *predictions* is equal to $\bar{y}$. As these are the estimated conditional means (by assumption), this gives you a relationship like the one you seek. To see this, just observe that the predictions are $X\vec{\beta}$, and group the least squares equation as

$$X^t(X\vec{\beta}) = X^t y$$

Now use a similar argument to what I did above.

Share Cite Improve this answer Follow   edited Jan 28 '16 at 17:13   answered Jan 27 '16 at 16:39
Matthew Drury
32.7k 2 96 127

Sorry @Matthew. I don't agree with your first equation; under exogeneity, the epsilon term should not be there. Another way to ask my question is can you make a statement about the conditional mean of the observation vs. the conditional mean of the dependent variable in the entire sample. Im guessing they are related! – ChinG Jan 27 '16 at 20:56

I believe you are absolutely correct on that first point, the ϵ should not be there, that's carelessness on my part, I'll fix it! I was trying to avoid the normality assumptions that are often stated here, but I got mixed up. Can you clarify your second point? – Matthew Drury Jan 27 '16 at 20:59

Thanks a lot for your quick answer. So my question is basically the following. Every observation has the conditional mean as you mentioned, which is basically x_i'b. Now, my question is how is this related to the sample average of y? I mean to compare the conditional mean of a specific observation with the unconditional mean across all observations in the sample. So basically, y_i would correspond to a specific observation, whereas Y would correspond to the vector of observations. My question is how is the conditional mean of each observation related to the mean across observations. Thanks! – ChinG Jan 27 '16 at 21:12

1 @ChinG I tried to address your question, hopefully that helps. – Matthew Drury Jan 28 '16 at 17:08

Add a comment

2

To get the unconditional mean (or marginal mean) of Y, the distribution of X is needed when the mean of Y depends on X as in your question. If you do not know and cannot estimate the distribution of X, it is impossible to derive the unconditional mean of Y.

Share Cite Improve this answer Follow   answered Apr 24 '17 at 4:49
user158565
6,947 2 8 18

Add a comment

Related

2 Independence assumptions in the classical regression model and higher moments

1 How do I derive this coefficient formula in a regression with non-linear conditional expectation function

4 Conditional vs. Unconditional Maximum Likelihood

1 Law of Iterated Expectations in Linear Regression

5 Conditional variance in OLS regression

5 Conditional variance in OLS regression

2 W is normal conditional on a normal variable Z. Does it necessarily follow that W is normal unconditionally?

4 Are the law of iterated expectation and the law of total expectations the same?

Hot Network Questions

- Technique for filling vacuum seal bags without getting any food on the rim
- Does the phrase "Return of the XXX" actually predate Return of the Jedi and Return of the King in common usage?
- Why is the "discount rate", charged to merchants in card transactions, called that?
- Why do X-ray telescopes have to be in space?
- Is it safe to screw into a single support beam holding up an entire house?
- Take me off the endless email threads
- Exactly why does Starship need to be this big for interplanetary travel?
- How would I say "twin sister"?
- Exactly how smart are Demon Lords? Demogorgon in particular
- Why are UFOs disk shaped?
- Is there any version of the Banach-Tarski paradox in ZF?
- Can you say "bald hill" to mean a hill that has no trees on it?
- Was climate a factor in the spread of Islam?
- Which scripture mentions Yuga length?
- Could you help me grammatically figure out this paragraph from All The Pretty Horses by C. McCarthy?
- Re-number and rename
- Squares in a triquadratic field
- Converting string to float using QgsExpression()
- What is the meaning of 再 before adjectives?

Add a comment

Your Answer

B I

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 conditional-expectation or ask your own question.

How do I contribute to the scientific community without publishing research in a scientific journal?

Contraction for negative short answers

Parsing "oblita carmina"

Is Moria fit for living?

Why does the Open Goldberg score have a G rather than A in bar 9 of variation 25?

Question feed