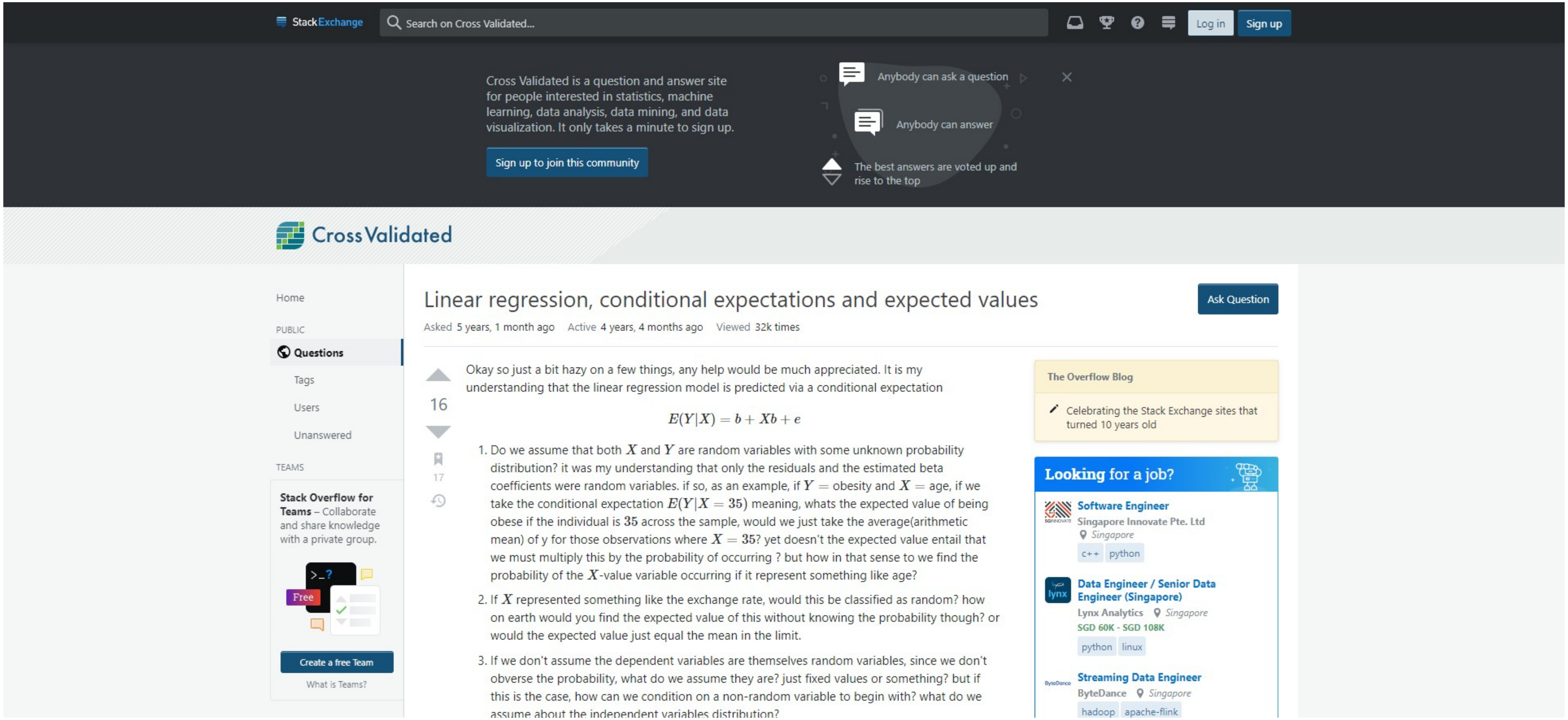


Linear regression, conditional expectations and expected values

Ask Question

Asked 5 years, 1 month ago Active 4 years, 4 months ago Viewed 32k times

16

17

Okay so just a bit hazy on a few things, any help would be much appreciated. It is my understanding that the linear regression model is predicted via a conditional expectation

$$E(Y|X) = b + Xb + e$$

1. Do we assume that both X and Y are random variables with some unknown probability distribution? it was my understanding that only the residuals and the estimated beta coefficients were random variables. if so, as an example, if Y = obesity and X = age, if we take the conditional expectation $E(Y|X = 35)$ meaning, whats the expected value of being obese if the individual is 35 across the sample, would we just take the average(arithmetic mean) of y for those observations where $X = 35$? yet doesn't the expected value entail that we must multiply this by the probability of occurring ? but how in that sense to we find the probability of the X-value variable occurring if it represent something like age?
2. If X represented something like the exchange rate, would this be classified as random? how on earth would you find the expected value of this without knowing the probability though? or would the expected value just equal the mean in the limit.
3. If we don't assume the dependent variables are themselves random variables, since we don't obverse the probability, what do we assume they are? just fixed values or something? but if this is the case, how can we condition on a non-random variable to begin with? what do we assume about the independent variables distribution?

What is Teams?

assume about the independent variables distribution?

Sorry if anything doesn't make sense or is obvious to anyone.

regression

Share Cite Improve this question Follow

edited Apr 17 '17 at 18:36
David C.
123 7

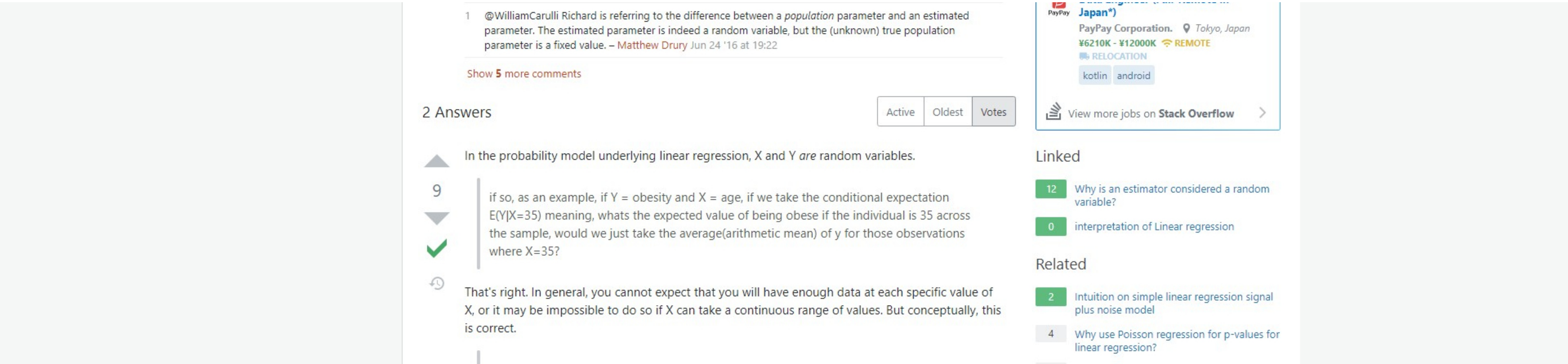
asked Jun 24 '16 at 17:18
William Carulli
163 1 1 6

1 The regression coefficient β is an unknown constant, not a random variable (in a frequentist world at least). – Richard Hardy Jun 24 '16 at 17:22

what you mean by conditional expectations? E(Y|X) simply means Y given X, that is, expected value of Y at X. Say, y = 5 + x, then you E(Y|X = 5) is 10. I did not get your point with conditional expectation – Zamir Akimbekov Jun 24 '16 at 17:25

@RichardHardy, it was my understanding that since B is the mean of the of the sampling distribution of of the beta's, that it is a random variable characterised by a normal distribution. are you referring to the population model? – William Carulli Jun 24 '16 at 17:51

Yes, population model. – Richard Hardy Jun 24 '16 at 18:53



1 @WilliamCarulli Richard is referring to the difference between a *population* parameter and an estimated parameter. The estimated parameter is indeed a random variable, but the (unknown) true population parameter is a fixed value. – Matthew Drury Jun 24 '16 at 19:22

Show **5** more comments

2 Answers

Active Oldest Votes

9

In the probability model underlying linear regression, X and Y *are* random variables.

> if so, as an example, if Y = obesity and X = age, if we take the conditional expectation E(Y|X=35) meaning, whats the expected value of being obese if the individual is 35 across the sample, would we just take the average(arithmetic mean) of y for those observations where X=35?

That's right. In general, you cannot expect that you will have enough data at each specific value of X, or it may be impossible to do so if X can take a continuous range of values. But conceptually, this is correct.

Linked

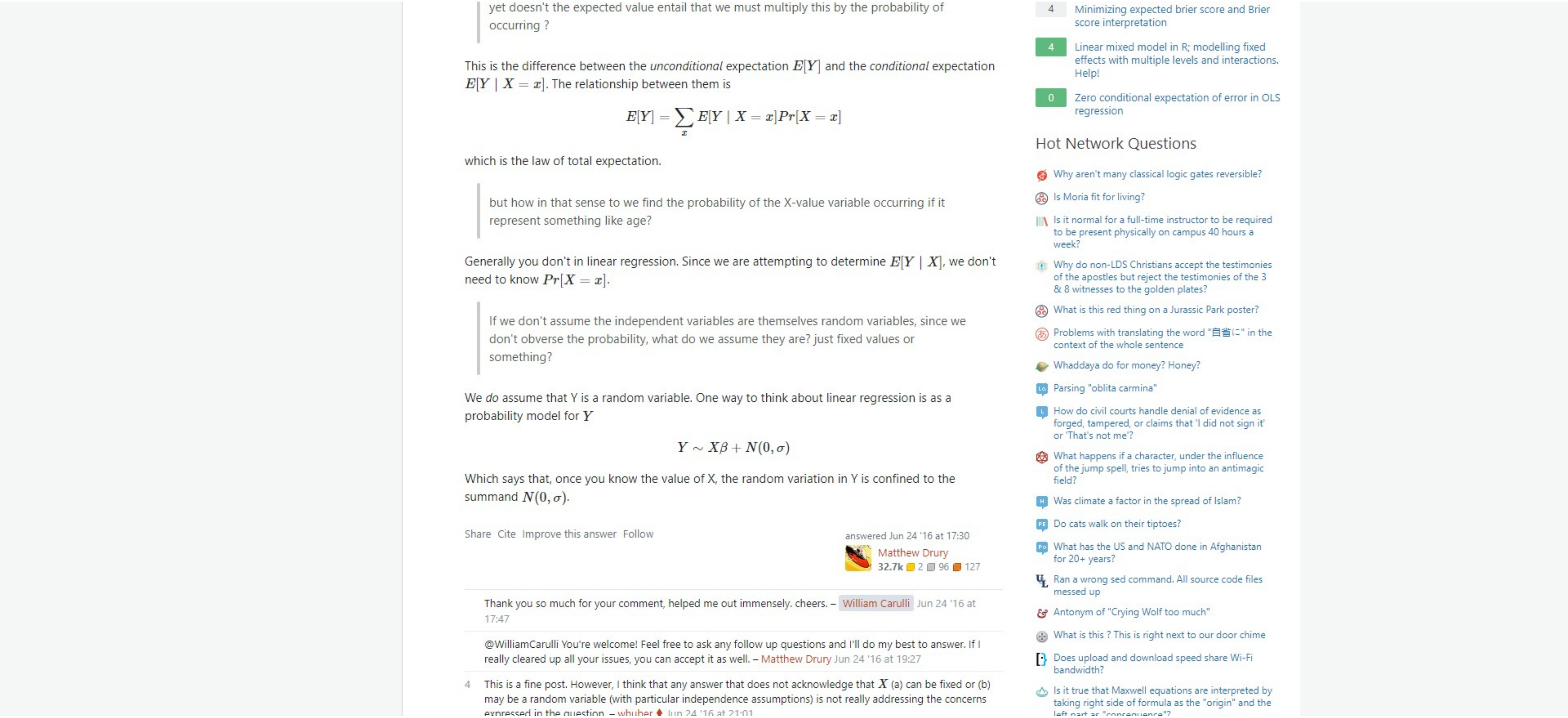
12 Why is an estimator considered a random variable?

0 interpretation of Linear regression

Related

2 Intuition on simple linear regression signal plus noise model

4 Why use Poisson regression for p-values for linear regression?



> yet doesn't the expected value entail that we must multiply this by the probability of occurring ?

This is the difference between the *unconditional* expectation $E[Y]$ and the *conditional* expectation $E[Y \mid X = x]$. The relationship between them is

$$E[Y] = \sum_x E[Y \mid X = x] Pr[X = x]$$

which is the law of total expectation.

> but how in that sense to we find the probability of the X-value variable occurring if it represent something like age?

Generally you don't in linear regression. Since we are attempting to determine $E[Y \mid X]$, we don't need to know $Pr[X = x]$.

> If we don't assume the independent variables are themselves random variables, since we don't obverse the probability, what do we assume they are? just fixed values or something?

We *do* assume that Y is a random variable. One way to think about linear regression is as a probability model for Y

$$Y \sim X\beta + N(0, \sigma)$$

Which says that, once you know the value of X, the random variation in Y is confined to the summand $N(0, \sigma)$.

Share Cite Improve this answer Follow

answered Jun 24 '16 at 17:30
Matthew Drury
32.7k 2 96 127

Thank you so much for your comment, helped me out immensely. cheers. – William Carulli Jun 24 '16 at 17:47

@WilliamCarulli You're welcome! Feel free to ask any follow up questions and I'll do my best to answer. If I really cleared up all your issues, you can accept it as well. – Matthew Drury Jun 24 '16 at 19:27

4 This is a fine post. However, I think that any answer that does not acknowledge that X (a) can be fixed or (b) may be a random variable (with particular independence assumptions) is not really addressing the concerns expressed in the question. – whuber ♦ Jun 24 '16 at 21:01

4 Minimizing expected brier score and Brier score interpretation

4 Linear mixed model in R; modelling fixed effects with multiple levels and interactions. Help!

0 Zero conditional expectation of error in OLS regression

Hot Network Questions

Why aren't many classical logic gates reversible?

Is Moria fit for living?

Is it normal for a full-time instructor to be required to be present physically on campus 40 hours a week?

Why do non-LDS Christians accept the testimonies of the apostles but reject the testimonies of the 3 & 8 witnesses to the golden plates?

What is this red thing on a Jurassic Park poster?

Problems with translating the word "自省に" in the context of the whole sentence

Whaddaya do for money? Honey?

Parsing "oblita carmina"

How do civil courts handle denial of evidence as forged, tampered, or claims that 'I did not sign it' or 'That's not me'?

What happens if a character, under the influence of the jump spell, tries to jump into an antimagic field?

Was climate a factor in the spread of Islam?

Do cats walk on their tiptoes?

What has the US and NATO done in Afghanistan for 20+ years?

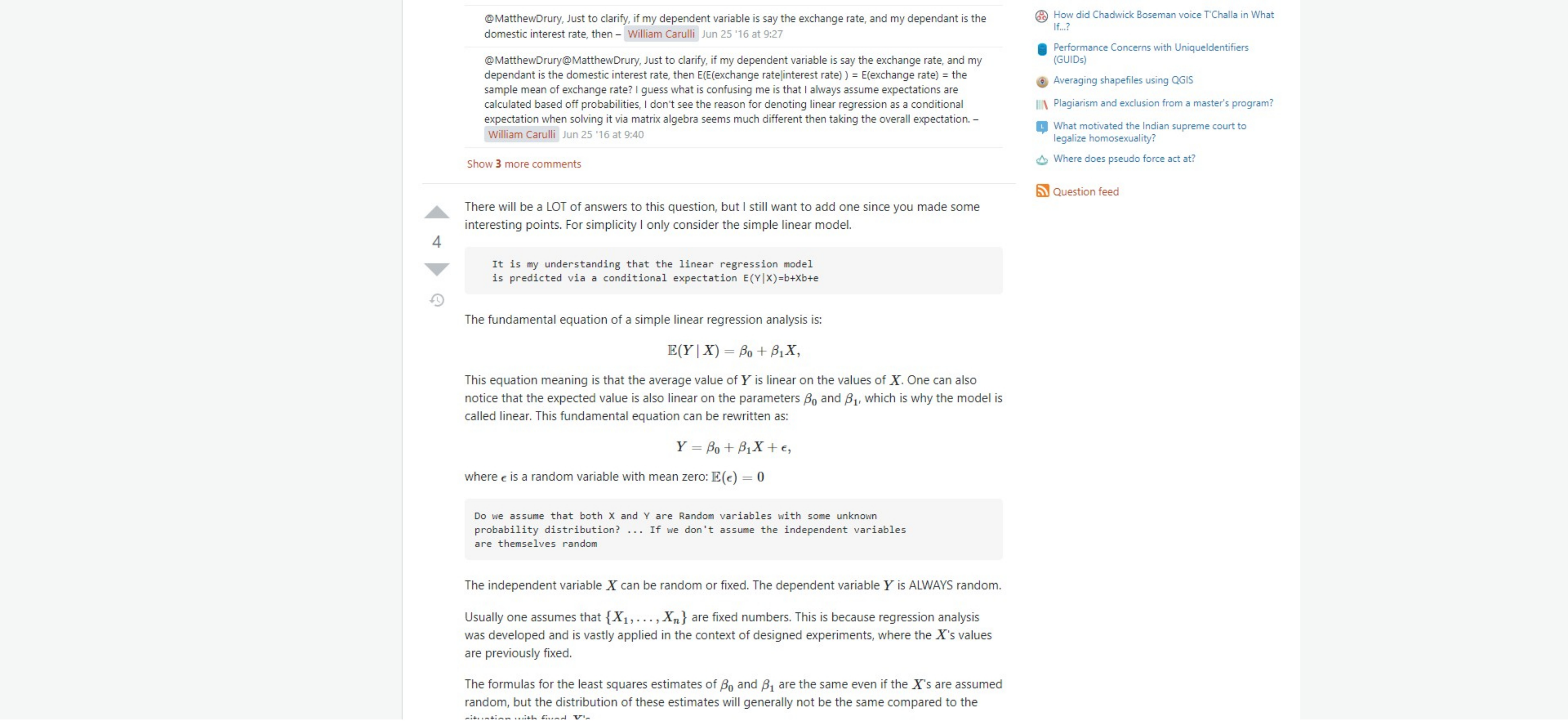
Ran a wrong sed command. All source code files messed up

Antonym of "Crying Wolf too much"

What is this ? This is right next to our door chime

Does upload and download speed share Wi-Fi bandwidth?

Is it true that Maxwell equations are interpreted by taking right side of formula as the "origin" and the left part as "consequence"?



expressed in the question. – whuber ♦ Jun 24 '16 at 21:07

left part as "consequence"?

@MatthewDrury, Just to clarify, if my dependent variable is say the exchange rate, and my dependant is the domestic interest rate, then – William Carulli Jun 25 '16 at 9:27

@MatthewDrury@MatthewDrury, Just to clarify, if my dependent variable is say the exchange rate, and my dependant is the domestic interest rate, then E(E(exchange rate|interest rate)) = E(exchange rate) = the sample mean of exchange rate? I guess what is confusing me is that I always assume expectations are calculated based off probabilities, I don't see the reason for denoting linear regression as a conditional expectation when solving it via matrix algebra seems much different then taking the overall expectation. – William Carulli Jun 25 '16 at 9:40

Show **3** more comments

4

There will be a LOT of answers to this question, but I still want to add one since you made some interesting points. For simplicity I only consider the simple linear model.

```
It is my understanding that the linear regression model
is predicted via a conditional expectation E(Y|X)=b+Xb+e
```

The fundamental equation of a simple linear regression analysis is:

$$\mathbb{E}(Y\,|\,X) = \beta_0 + \beta_1 X,$$

This equation meaning is that the average value of Y is linear on the values of X. One can also notice that the expected value is also linear on the parameters β_0 and β_1, which is why the model is called linear. This fundamental equation can be rewritten as:

$$Y = \beta_0 + \beta_1 X + \epsilon,$$

where ϵ is a random variable with mean zero: $\mathbb{E}(\epsilon) = 0$

```
Do we assume that both X and Y are Random variables with some unknown
probability distribution? ... If we don't assume the independent variables
are themselves random
```

The independent variable X can be random or fixed. The dependent variable Y is ALWAYS random.

Usually one assumes that $\{X_1, \ldots, X_n\}$ are fixed numbers. This is because regression analysis was developed and is vastly applied in the context of designed experiments, where the X's values are previously fixed.

The formulas for the least squares estimates of β_0 and β_1 are the same even if the X's are assumed random, but the distribution of these estimates will generally not be the same compared to the

How did Chadwick Boseman voice T'Challa in What If...?

Performance Concerns with UniqueIdentifiers (GUIDs)

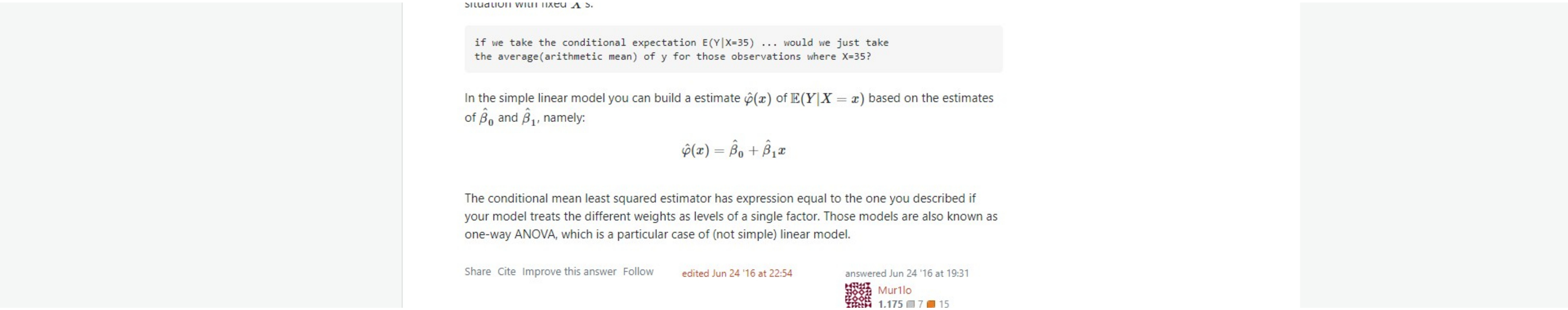
Averaging shapefiles using QGIS

Plagiarism and exclusion from a master's program?

What motivated the Indian supreme court to legalize homosexuality?

Where does pseudo force act at?

Question feed



situation with fixed X's.

```
if we take the conditional expectation E(Y|X=35) ... would we just take
the average(arithmetic mean) of y for those observations where X=35?
```

In the simple linear model you can build a estimate $\hat{\varphi}(x)$ of $\mathbb{E}(Y|X=x)$ based on the estimates of $\hat{\beta}_0$ and $\hat{\beta}_1$, namely:

$$\hat{\varphi}(x) = \hat{\beta}_0 + \hat{\beta}_1 x$$

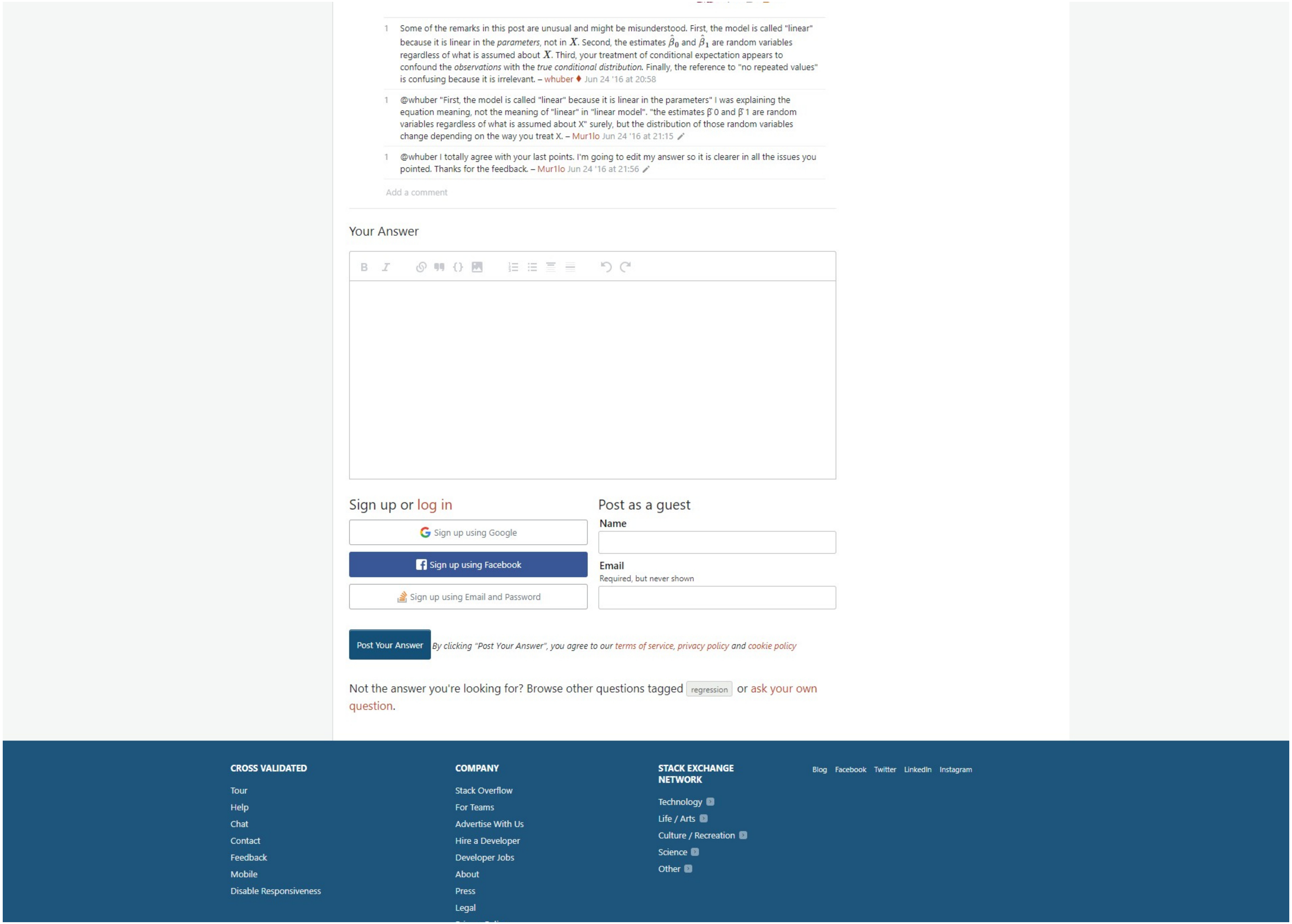
The conditional mean least squared estimator has expression equal to the one you described if your model treats the different weights as levels of a single factor. Those models are also known as one-way ANOVA, which is a particular case of (not simple) linear model.

Share Cite Improve this answer Follow

edited Jun 24 '16 at 22:54

answered Jun 24 '16 at 19:31

Mur1lo
1.175 ● 7 ● 15



1 Some of the remarks in this post are unusual and might be misunderstood. First, the model is called "linear" because it is linear in the *parameters*, not in X. Second, the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ are random variables regardless of what is assumed about X. Third, your treatment of conditional expectation appears to confound the *observations* with the *true conditional distribution*. Finally, the reference to "no repeated values" is confusing because it is irrelevant. – whuber ♦ Jun 24 '16 at 20:58

1 @whuber "First, the model is called "linear" because it is linear in the parameters" I was explaining the equation meaning, not the meaning of "linear" in "linear model". "the estimates β̂0 and β̂1 are random variables regardless of what is assumed about X" surely, but the distribution of those random variables change depending on the way you treat X. – Mur1lo Jun 24 '16 at 21:15

1 @whuber I totally agree with your last points. I'm going to edit my answer so it is clearer in all the issues you pointed. Thanks for the feedback. – Mur1lo Jun 24 '16 at 21:56

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 or ask your own question.