

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^tX 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 cide the first component is ∇ as Equation them you got the result $\theta = \bar{a}$ orthogonal to the data column. This means that the first row of X^tX 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 $\bar{\eta}$. 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

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

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



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 quessing 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

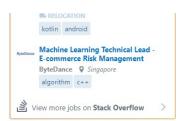


To get the unconditional mean (or marginal mean) of Y, the distribution of X is needed when the mean of Y depends on X as in your question. If you do not know and cannot estimate the distribution of X, it is impossible to derive the unconditional mean of Y.



Share Cite Improve this answer Follow





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
- Law of Iterated Expectations in Linear Regression
- 5 Conditional variance in OLS regression
- 2 Wis normal conditional on a normal
- Conditional variance in OLS regression
- 2 W is normal conditional on a normal variable Z. Does it necessarily follow that W is normal unconditionally?
- 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
- How would I say "twin sister"?
- Exactly how smart are Demon Lords? Demogorgon
- Why are UFOs disk shaped?
- Mo Is there any version of the Banach-Tarski paradox
- 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.
- Re-number and rename
- Mo Squares in a triquadratic field
- Converting string to float using QgsExpression()
- What is the meaning of 再 before adjectives?

