Cross Validated

How do I calculate the variance of the OLS estimator β_0, conditional on $x_1, \ldots , x_n$?

Asked 10 years, 2 months ago Modified 5 years, 10 months ago Viewed 206k times

25

I know that

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

and this is how far I got when I calculated the variance:

$$\begin{aligned} Var(\hat{\beta}_0) &= Var(\bar{y} - \hat{\beta}_1 \bar{x}) \\ &= Var((-\bar{x})\hat{\beta}_1 + \bar{y}) \\ &= Var((-\bar{x})\hat{\beta}_1) + Var(\bar{y}) \\ &= (-\bar{x})^2 Var(\hat{\beta}_1) + 0 \\ &= (\bar{x})^2 Var(\hat{\beta}_1) + 0 \\ &= \frac{\sigma^2 (\bar{x})^2}{\displaystyle\sum_{i=1}^{n} (x_i - \bar{x})^2} \end{aligned}$$

but that's far as I got. The final formula I'm trying to calculate is

$$Var(\hat{\beta}_0) = \frac{\sigma^2 n^{-1} \displaystyle\sum_{i=1}^{n} x_i^2}{\displaystyle\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

I'm not sure how to get

$$(\bar{x})^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2$$

assuming my math is correct up to there.

Is this the right path?

$$\begin{aligned} (\bar{x})^2 &= \left(\frac{1}{n} \sum_{i=1}^{n} x_i \right)^2 \\ &= \frac{1}{n^2} \left(\sum_{i=1}^{n} x_i \right)^2 \end{aligned}$$

I'm sure it's simple, so the answer can wait for a bit if someone has a hint to push me in the right direction.

regression self-study

Share Cite Edit Follow Flag

edited Jul 13, 2013 at 0:26
QuantIbex
4,036 1 29 44

asked Jul 12, 2013 at 22:14
M T
611 1 8 15

2 This is not the right path. The 4th equation doesn't hold. For example, with $x_1 = -1$, $x_2 = 0$, and $x_3 = 1$, the left term is zero, whilst the right term is $2/3$. The problem comes from the step where you split the variance (3rd line of second equation). See why? – QuantIbex Jul 12, 2013 at 23:55

Hint towards QuantIbex point: variance is not a linear function. It violates both additivity and scalar multiplication. – David Marx Jul 13, 2013 at 1:53

@DavidMarx That step should be

$$= Var((-\bar{x})\hat{\beta}_1 + \bar{y}) = (\bar{x})^2 Var(\hat{\beta}_1) + \bar{y}$$

, I think, and then once I substitute in for $\hat{\beta}_1$ and $\bar{y}$ (not sure what to do for this but I'll think about it more), *that* should put me on the right path I hope. – M T Jul 13, 2013 at 3:53

This is not correct. Think about the condition required for the variance of a sum to be equal to the sum of the variances. – QuantIbex Jul 13, 2013 at 10:29

2 No, $\bar{y}$ is random since $y_i = \beta_0 + \beta_1 x_i + \epsilon$, where ϵ denotes the (random) noise. But OK, my previous comment was maybe misleading. Also, $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$, if a and b denote constants. – QuantIbex Jul 13, 2013 at 19:38

2 Answers

Sorted by: Highest score (default)

28

This is a self-study question, so I provide hints that will hopefully help to find the solution, and I'll edit the answer based on your feedbacks/progress.

The parameter estimates that minimize the sum of squares are

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1\bar{x}, \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x})y_i}{\sum_{i=1}^n (x_i - \bar{x})^2}. \end{aligned}$$

To get the variance of $\hat{\beta}_0$, start from its expression and substitute the expression of $\hat{\beta}_1$, and do the algebra

$$\mathrm{Var}(\hat{\beta}_0) = \mathrm{Var}(\bar{Y} - \hat{\beta}_1\bar{x}) = \ldots$$

Edit:

We have

$$\begin{aligned}\operatorname{Var}(\hat{\beta}_0) &= \operatorname{Var}(\bar{Y} - \hat{\beta}_1\bar{x}) \\ &= \operatorname{Var}(\bar{Y}) + (\bar{x})^2\operatorname{Var}(\hat{\beta}_1) - 2\bar{x}\operatorname{Cov}(\bar{Y}, \hat{\beta}_1).\end{aligned}$$

The two variance terms are

$$\operatorname{Var}(\bar{Y}) = \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\operatorname{Var}(Y_i) = \frac{\sigma^2}{n},$$

and

$$\begin{aligned}\operatorname{Var}(\hat{\beta}_1) &= \frac{1}{\left[\sum_{i=1}^n (x_i - \bar{x})^2\right]^2}\sum_{i=1}^{n}(x_i - \bar{x})^2\operatorname{Var}(Y_i) \\ &= \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2},\end{aligned}$$

and the covariance term is

$$\begin{aligned}\operatorname{Cov}(\bar{Y}, \hat{\beta}_1) &= \operatorname{Cov}\left\{\frac{1}{n}\sum_{i=1}^{n} Y_i, \frac{\sum_{j=1}^n (x_j - \bar{x})Y_j}{\sum_{i=1}^n (x_i - \bar{x})^2}\right\} \\ &= \frac{1}{n}\frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2}\operatorname{Cov}\left\{\sum_{i=1}^{n} Y_i, \sum_{j=1}^{n}(x_j - \bar{x})Y_j\right\} \\ &= \frac{1}{n\sum_{i=1}^n (x_i - \bar{x})^2}\sum_{i=1}^{n}(x_j - \bar{x})\sum_{j=1}^{n}\operatorname{Cov}(Y_i, Y_j) \\ &= \frac{1}{n\sum_{i=1}^n (x_i - \bar{x})^2}\sum_{i=1}^{n}(x_j - \bar{x})\sigma^2 \\ &= 0\end{aligned}$$

since $\sum_{i=1}^n (x_j - \bar{x}) = 0.$

And since

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n}\bar{x}^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2,$$

we have

$$\begin{aligned}\operatorname{Var}(\hat{\beta}_0) &= \frac{\sigma^2}{n} + \frac{\sigma^2\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \frac{\sigma^2}{n\sum_{i=1}^n (x_i - \bar{x})^2}\left\{\sum_{i=1}^{n}(x_i - \bar{x})^2 + n\bar{x}^2\right\} \\ &= \frac{\sigma^2\sum_{i=1}^n x_i^2}{n\sum_{i=1}^n (x_i - \bar{x})^2}.\end{aligned}$$

Edit 2

Why do we have $\operatorname{var}(\sum_{i=1}^n Y_i) = \sum_{i=1}^n \operatorname{Var}(Y_i)$?

The assumed model is $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$, where the ϵ_i are independant and identically distributed random variables with $\mathrm{E}(\epsilon_i) = 0$ and $\operatorname{var}(\epsilon_i) = \sigma^2$.

Once we have a sample, the X_i are known, the only random terms are the ϵ_i. Recalling that for a random variable Z and a constant a, we have $\operatorname{var}(a + Z) = \operatorname{var}(Z)$. Thus,

$$\begin{aligned}
\operatorname{var}\left(\sum_{i=1}^n Y_i\right) &= \operatorname{var}\left(\sum_{i=1}^n \beta_0 + \beta_1 X_i + \epsilon_i\right) \\
&= \operatorname{var}\left(\sum_{i=1}^n \epsilon_i\right) = \sum_{i=1}^n \sum_{j=1}^n \operatorname{cov}(\epsilon_i, \epsilon_j) \\
&= \sum_{i=1}^n \operatorname{cov}(\epsilon_i, \epsilon_i) = \sum_{i=1}^n \operatorname{var}(\epsilon_i) \\
&= \sum_{i=1}^n \operatorname{var}(\beta_0 + \beta_1 X_i + \epsilon_i) = \sum_{i=1}^n \operatorname{var}(Y_i).
\end{aligned}$$

The 4th equality holds as $\operatorname{cov}(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$ by the independence of the ϵ_i.

Share Cite Edit Follow Flag edited Nov 27, 2017 at 9:53 answered Jul 13, 2013 at 9:31

QuantIbex
4,036 1 29 44

I think I got it! The book has suggested steps, and I was able to prove each step separately (I think). It's not as satisfying as just sitting down and grinding it out from this step, since I had to prove intermediate conclusions for it to help, but I think everything looks good. – M T Jul 14, 2013 at 0:25

See edit for the development of the suggested approach. – QuantIbex Jul 14, 2013 at 6:19

The variance of the sum equals the sum of the variances in this step:

$$\operatorname{Var}(\bar{Y}) = \operatorname{Var}\left(\frac{1}{n}\sum_{i=1}^n Y_i\right) = \frac{1}{n^2}\sum_{i=1}^n \operatorname{Var}(Y_i)$$

because since the X_i are independent, this implies that the Y_i are independent as well, right? – M T Jul 14, 2013 at 18:40

Also, you can factor out a constant from the covariance in this step:

$$\frac{1}{n}\frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \operatorname{Cov}\left\{\sum_{i=1}^n Y_i, \sum_{j=1}^n (x_j - \bar{x}) Y_j\right\}$$

even though it's not in both elements because the formula for covariance is multiplicative, right? – M T Jul 14, 2013 at 18:42

1 @oort, in the numerator you have the sum of n terms that are identical (and equal to σ^2), so the numerator is $n\sigma^2$. – QuantIbex Apr 7, 2016 at 14:40

1

I got it! Well, with help. I found the part of the book that gives steps to work through when proving the $Var\left(\hat{\beta}_0\right)$ formula (thankfully it doesn't actually work them out, otherwise I'd be tempted to not actually do the proof). I proved each separate step, and I think it worked.

I'm using the book's notation, which is:

$$SST_x = \sum_{i=1}^{n}(x_i - \bar{x})^2,$$

and u_i is the error term.

1) Show that $\hat{\beta}_1$ can be written as $\hat{\beta}_1 = \beta_1 + \sum_{i=1}^{n} w_i u_i$ where $w_i = \frac{d_i}{SST_x}$ and $d_i = x_i - \bar{x}$.

This was easy because we know that

$$\begin{aligned}\hat{\beta}_1 &= \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{SST_x} \\ &= \beta_1 + \sum_{i=1}^{n}\frac{d_i}{SST_x}u_i \\ &= \beta_1 + \sum_{i=1}^{n} w_i u_i\end{aligned}$$

2) Use part 1, along with $\sum_{i=1}^{n} w_i = 0$ to show that $\hat{\beta}_1$ and $\bar{u}$ are uncorrelated, i.e. show that $E[(\hat{\beta}_1 - \beta_1)\bar{u}] = 0$.

$$\begin{aligned}E[(\hat{\beta}_1 - \beta_1)\bar{u}] &= E[\bar{u}\sum_{i=1}^{n} w_i u_i] \\ &= \sum_{i=1}^{n} E[w_i \bar{u} u_i] \\ &= \sum_{i=1}^{n} w_i E[\bar{u} u_i] \\ &= \frac{1}{n}\sum_{i=1}^{n} w_i E\left(u_i \sum_{j=1}^{n} u_j\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} w_i \left[E\left(u_i u_1\right) + \cdots + E(u_i u_j) + \cdots + E\left(u_i u_n\right)\right]\end{aligned}$$

and because the u are i.i.d., $E(u_i u_j) = E(u_i)E(u_j)$ when $j \neq i$.

When $j = i$, $E(u_i u_j) = E(u_i^2)$, so we have:

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n} w_i \left[E(u_i)E(u_1) + \cdots + E(u_i^2) + \cdots + E(u_i)E(u_n)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} w_i E(u_i^2) \\
&= \frac{1}{n}\sum_{i=1}^{n} w_i \left[Var(u_i) + E(u_i)E(u_i)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} w_i \sigma^2 \\
&= \frac{\sigma^2}{n}\sum_{i=1}^{n} w_i \\
&= \frac{\sigma^2}{n \cdot SST_x}\sum_{i=1}^{n}(x_i - \bar{x}) \\
&= \frac{\sigma^2}{n \cdot SST_x}(0) \qquad\qquad = 0
\end{aligned}
$$

3) Show that $\hat{\beta}_0$ can be written as $\hat{\beta}_0 = \beta_0 + \bar{u} - \bar{x}(\hat{\beta}_1 - \beta_1)$. This seemed pretty easy too:

$$
\begin{aligned}
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1\bar{x} \\
&= (\beta_0 + \beta_1\bar{x} + \bar{u}) - \hat{\beta}_1\bar{x} \\
&= \beta_0 + \bar{u} - \bar{x}(\hat{\beta}_1 - \beta_1).
\end{aligned}
$$

4) Use parts 2 and 3 to show that $Var(\hat{\beta}_0) = \frac{\sigma^2}{n} + \frac{\sigma^2(\bar{x})^2}{SST_x}$:

$$
\begin{aligned}
Var(\hat{\beta}_0) &= Var(\beta_0 + \bar{u} - \bar{x}(\hat{\beta}_1 - \beta_1)) \\
&= Var(\bar{u}) + (-\bar{x})^2 Var(\hat{\beta}_1 - \beta_1) \\
&= \frac{\sigma^2}{n} + (\bar{x})^2 Var(\hat{\beta}_1) \\
&= \frac{\sigma^2}{n} + \frac{\sigma^2(\bar{x})^2}{SST_x}.
\end{aligned}
$$

I believe this all works because since we provided that $\bar{u}$ and $\hat{\beta}_1 - \beta_1$ are uncorrelated, the covariance between them is zero, so the variance of the sum is the sum of the variance. β_0 is just a constant, so it drops out, as does β_1 later in the calculations.

5) Use algebra and the fact that $\frac{SST_x}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2$:

$$\begin{aligned}
Var(\hat{\beta}_0) &= \frac{\sigma^2}{n} + \frac{\sigma^2(\bar{x})^2}{SST_x} \\
&= \frac{\sigma^2 SST_x}{SST_x n} + \frac{\sigma^2(\bar{x})^2}{SST_x} \\
&= \frac{\sigma^2}{SST_x}\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2\right) + \frac{\sigma^2(\bar{x})^2}{SST_x} \\
&= \frac{\sigma^2 n^{-1}\sum_{i=1}^{n} x_i^2}{SST_x}
\end{aligned}$$

Share Cite Edit Follow Flag

edited Jul 18, 2013 at 17:40

answered Jul 14, 2013 at 0:23

M T

611 1 8 15

There might be a typo in point 1; I think $\mathrm{var}(\hat{\beta})$ should read $\hat{\beta}$. – QuantIbex Jul 15, 2013 at 22:10

You might want to clarify notations, and specify what u_i and SST_x are. – QuantIbex Jul 15, 2013 at 22:13

u_i is the error term and SST_x is the total sum of squares for x (defined in the edit). – M T Jul 15, 2013 at 22:37

1 In point 1, the term β_1 is missing in the last two lines. – QuantIbex Jul 16, 2013 at 6:06

1 In point 2, you can't take $\bar{u}$ out of the expectation, it's not a constant. – QuantIbex Jul 16, 2013 at 6:07