

## How to derive variance-covariance matrix of coefficients in linear regression

Asked 6 years, 4 months ago   Active 3 months ago   Viewed 89k times



I am reading a book on linear regression and have some trouble understanding the variance-covariance matrix of  $\mathbf{b}$ :

36

### Regression coefficients



The variance-covariance matrix of  $\mathbf{b}$ :

★  
49

$$(6.77) \quad \sigma^2(\mathbf{b}) = \begin{bmatrix} \sigma^2(b_0) & \sigma(b_0, b_1) \\ \sigma(b_1, b_0) & \sigma^2(b_1) \end{bmatrix}$$

is:

$$(6.78) \quad \underset{2 \times 2}{\sigma^2(\mathbf{b})} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$$

or, using (6.27):

$$(6.78a) \quad \sigma^2(\mathbf{b}) = \begin{bmatrix} \frac{\sigma^2 \sum X_i^2}{n \sum (X_i - \bar{X})^2} & \frac{-\bar{X} \sigma^2}{\sum (X_i - \bar{X})^2} \\ \frac{-\bar{X} \sigma^2}{\sum (X_i - \bar{X})^2} & \frac{\sigma^2}{\sum (X_i - \bar{X})^2} \end{bmatrix}$$

The diagonal items are easy enough, but the off-diagonal ones are a bit more difficult, what puzzles me is that

$$\sigma(b_0, b_1) = E(b_0 b_1) - E(b_0)E(b_1) = E(b_0 b_1) - \beta_0 \beta_1$$

but there is no trace of  $\beta_0$  and  $\beta_1$  here.

regression

asked Aug 23 '13 at 9:13



qed

1,898   3   17   29

3 ▲ Related question: [stats.stackexchange.com/questions/44838/...](https://stats.stackexchange.com/questions/44838/...) – ocram Nov 21 '13 at 11:14

2 ▲ Which is the book? – Konstantinos May 29 '15 at 23:34

▲ Neter et al., Applied Linear Regression Models, 1983, page 216. You can find the same material in Applied Linear Statistical Models, 5th Edition, page 207. – akavalor Nov 19 at 4:59

### 3 Answers



This is actually a cool question that challenges your basic understanding of a regression.

53

First take out any initial confusion about notation. We are looking at the regression:

$$y = b_0 + b_1x + \hat{u}$$

where  $b_0$  and  $b_1$  are the estimators of the true  $\beta_0$  and  $\beta_1$ , and  $\hat{u}$  are the residuals of the regression. Note that the underlying true and unobserved regression is thus denoted as:

$$y = \beta_0 + \beta_1x + u$$

With the expectation of  $E[u] = 0$  and variance  $E[u^2] = \sigma^2$ . Some books denote  $b$  as  $\hat{\beta}$  and we adapt this convention here. We also make use the matrix notation, where  $\mathbf{b}$  is the  $2 \times 1$  vector that holds the estimators of  $\beta = [\beta_0, \beta_1]'$ , namely  $b = [b_0, b_1]'$ . (Also for the sake of clarity I treat  $X$  as fixed in the following calculations.)

Now to your question. Your formula for the covariance is indeed correct, that is:

$$\sigma(b_0, b_1) = E(b_0b_1) - E(b_0)E(b_1) = E(b_0b_1) - \beta_0\beta_1$$

I think you want to know how comes we have the true unobserved coefficients  $\beta_0, \beta_1$  in this formula? They actually **get cancelled out** if we take it a step further by expanding the formula. To see this, note that the population variance of the estimator is given by:

$$\text{Var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

This matrix holds the variances in the diagonal elements and covariances in the off-diagonal elements.

To arrive to the above formula, let's generalize your claim by using matrix notation. Let us therefore denote variance with  $\text{Var}[\cdot]$  and expectation with  $E[\cdot]$ .

$$\text{Var}[b] = E[b^2] - E[b]E[b']$$

Essentially we have the general variance formula, just using matrix notation. The equation resolves when substituting in the standard expression for the estimator  $b = (X'X)^{-1}X'y$ . Also assume  $E[b] = \beta$  being an unbiased estimator. Hence, we obtain:

$$E\left[\left((X'X)^{-1}X'y\right)^2\right] - \beta^2_{2 \times 2}$$

Note that we have on the right hand side  $\beta^2$  - 2x2 matrix, namely  $bb'$ , but you may at this point already guess what will happen with this term shortly.

Replacing  $y$  with our expression for the true underlying data generating process above, we have:

$$\begin{aligned} E\left[\left((X'X)^{-1}X'y\right)^2\right] - \beta^2 &= E\left[\left((X'X)^{-1}X'(X\beta + u)\right)^2\right] - \beta^2 \\ &= E\left[\underbrace{\left((X'X)^{-1}X'X\beta\right)}_{=I} + (X'X)^{-1}X'u\right]^2 - \beta^2 \\ &= E\left[\left(\beta + (X'X)^{-1}X'u\right)^2\right] - \beta^2 \\ &= \beta^2 + E\left[\left(X'X\right)^{-1}X'u\right]^2 - \beta^2 \end{aligned}$$

since  $E[u] = 0$ . Furthermore, the quadratic  $\beta^2$  term cancels out as anticipated.

Thus we have:

$$\text{Var}[b] = ((X'X)^{-1}X')^2 E[u^2]$$

By linearity of expectations. Note that by assumption  $E[u^2] = \sigma^2$  and  $((X'X)^{-1}X')^2 = (X'X)^{-1}X'X(X'X)^{-1} = (X'X)^{-1}$  since  $X'X$  is a  $K \times K$  symmetric matrix and thus the same as its transpose. Finally we arrive at

$$\text{Var}[b] = \sigma^2(X'X)^{-1}$$

Now that we got rid of all  $\beta$  terms. Intuitively, the variance of the estimator is independent of the value of true underlying coefficient, as this is not a random variable per se. The result is valid for all individual elements in the variance covariance matrix as shown in the book thus also valid for the off diagonal elements as well with  $\beta_0\beta_1$  to cancel out respectively. The only problem was that you had applied the general formula for the variance which does not reflect this cancellation at first.

Ultimately, the variance of the coefficients reduces to  $\sigma^2(X'X)^{-1}$  and independent of  $\beta$ . But what does this mean? (I believe you asked also for a more general understanding of the general covariance matrix)


Look at the formula in the book. It simply asserts that the variance of the estimator increases for when the true underlying error term is more noisy ( $\sigma^2$  increases), but **decreases** for when the spread of  $X$  increases. Because having more observations spread around the true value, lets you in general build an estimator that is more accurate and thus closer to the true  $\beta$ . On the other hand, the covariance terms on the off-diagonal become practically relevant in hypothesis testing of joint hypotheses such as  $b_0 = b_1 = 0$ . Other than that they are a bit of a fudge, really. Hope this clarifies all questions.

edited Jan 8 at 19:20

 **zyxue**  
583 3 15

answered Feb 10 '14 at 17:43

 **Majte**  
1,577 11 22

 and when keep the spread constant and decrease the x's, the standard error of the intercept becomes smaller, which makes sense. – **Theta30** Feb 27 '14 at 23:21

▲ I don't follow the expansion of the square. Why is not simplified to  $((X'X)^{-1}X')^2 = ((X'X)^{-1}X')((X'X)^{-1}X') = X^{-2}$ ? – David Sep 28 '18 at 22:33  
 ▼

▲ In your case we have

2

$$X'X = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix}$$

▼ Invert this matrix and you will get the desired result.

answered Nov 21 '13 at 11:10



**mpiktas**

**31.1k** 4 71 134

▲ It appears that  $\beta_0\beta_1$  are the predicted values (expected values). They make the switch between  $E(b_0) = \beta_0$  and  $E(b_1) = \beta_1$ .

1

answered Aug 23 '13 at 9:44



**Drew75**

**1,015** 6 11

▲  $\beta_0$  and  $\beta_1$  are generally unknown, what can they switch to? – qed Aug 23 '13 at 9:49  
 ▼

▲ I think I understand the confusion, and I think they perhaps should have written  $\beta_0^*$  rather than  $\beta_0$ . Here's another post that goes through the calculation: [link](#) – Drew75  
 ▼ Aug 23 '13 at 10:16

2 ▲ @qed: to sample estimates of the unknown quantities. – Glen\_b -Reinstate Monica Aug 23 '13 at 10:20  
 ▼