Reminder No. 2: Propagation of Error, and Standard Errors for Derived Quantities

36-402, Advanced Data Analysis

Last update: 27 February 2013

A reminder about how we get approximate standard errors for functions of quantities which are themselves estimated with error.

Suppose we are trying to estimate some quantity $\theta$. We compute an estimate $\hat{\theta}$, based on our data. Since our data is more or less random, so is $\hat{\theta}$. One convenient way of measuring the purely statistical noise or uncertainty in $\hat{\theta}$ is its standard deviation. This is the **standard error** of our estimate of $\theta$.\(^1\) Standard errors are not the only way of summarizing this noise, nor a completely sufficient way, but they are often useful.

Suppose that our estimate $\hat{\theta}$ is a function of some intermediate quantities $\hat{\psi}_1, \hat{\psi}_2, \ldots, \hat{\psi}_p$, which are also estimated:

$$\hat{\theta} = f(\hat{\psi}_1, \hat{\psi}_2, \ldots, \hat{\psi}_p)$$

For instance, $\theta$ might be the difference in expected values between two groups, with $\psi_1$ and $\psi_2$ the expected values in the two groups, and $f(\psi_1, \psi_2) = \psi_1 - \psi_2$. If we have a standard error for each of the original quantities $\psi_i$, it would seem like we should be able to get a standard error for the **derived quantity** $\hat{\theta}$. There is in fact a simple if approximate way of doing so, which is called **propagation of error**\(^2\).

We start with (what else?) a Taylor expansion. We’ll write $\psi_i^*$ for the true (ensemble or population) value which is estimated by $\hat{\psi}_i$.

$$f(\psi_1^*, \psi_2^*, \ldots, \psi_p^*) \approx f(\hat{\psi}_1, \hat{\psi}_2, \ldots, \hat{\psi}_p) + \sum_{i=1}^{p} (\psi_i^* - \hat{\psi}_i) \left. \frac{\partial f}{\partial \psi_i} \right|_{\psi_i^* = \hat{\psi}_i}$$

$$f(\hat{\psi}_1, \hat{\psi}_2, \ldots, \hat{\psi}_p) \approx f(\psi_1^*, \psi_2^*, \ldots, \psi_p^*) + \sum_{i=1}^{p} (\hat{\psi}_i - \psi_i^*) \left. \frac{\partial f}{\partial \psi_i} \right|_{\psi_i = \hat{\psi}_i}$$

\(^1\)Thanks to Prof. Howard Seltman for corrections

\(^2\)It is not, of course, to be confused with the standard deviation of the data. It is not even to be confused with the standard error of the mean, unless $\theta$ is the expected value of the data and $\hat{\theta}$ is the sample mean.

\(^2\)Or, sometimes, the delta method.
\[
\hat{\theta} \approx \theta^* + \sum_{i=1}^{p} (\hat{\psi}_i - \psi_i^*) f'_i(\hat{\phi})
\]  

(4)

Introducing \( f'_i \) as an abbreviation for \( \frac{\partial f}{\partial \psi_i} \). The left-hand side is now the quantity whose standard error we want. I have done this manipulation because now \( \hat{\theta} \) is a linear function (approximately!) of some random quantities whose variances we know, and some derivatives which we can calculate.

Remember the rules for arithmetic with variances: if \( X \) and \( Y \) are random variables, and \( a, b \) and \( c \) are constants,

\[
\text{Var} [a] = 0 \quad (5)
\]

\[
\text{Var} [a + b X] = b^2 \text{Var} [X] \quad (6)
\]

\[
\text{Var} [a + b X + c Y] = b^2 \text{Var} [X] + c^2 \text{Var} [Y] + 2 bc \text{Cov} [X, Y] \quad (7)
\]

While we don’t know \( f(\psi^*_1, \psi^*_2, \ldots, \psi^*_p) \), it’s constant, so it has variance 0. Similarly, \( \text{Var} [\hat{\psi}_i - \psi_i^*] = \text{Var} [\hat{\psi}_i] \). Repeatedly applying these rules to Eq. 4,

\[
\text{Var} [\hat{\theta}] \approx \sum_{i=1}^{p} (f'_i(\hat{\phi})^2 \text{Var} [\hat{\psi}_i] + 2 \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} f'_i(\hat{\phi})f'_j(\hat{\phi}) \text{Cov} [\hat{\psi}_i, \hat{\psi}_j] \quad (8)
\]

The standard error for \( \hat{\theta} \) would then be the square root of this.

If we follow this rule for the simple case of group differences, \( f(\psi_1, \psi_2) = \psi_1 - \psi_2 \), we find that

\[
\text{Var} [\hat{\theta}] = \text{Var} [\hat{\psi}_1] + \text{Var} [\hat{\psi}_2] - 2 \text{Cov} [\hat{\psi}_1, \hat{\psi}_2] \quad (9)
\]

just as we would find from the basic rules for arithmetic with variances. The approximation in Eq. 8 comes from the nonlinearities in \( f \).

If the estimates of the initial quantities are uncorrelated, Eq. 8 simplifies to

\[
\text{Var} [\hat{\theta}] \approx \sum_{i=1}^{p} (f'_i(\hat{\phi}))^2 \text{Var} [\hat{\psi}_i] \quad (10)
\]

and, again, the standard error of \( \hat{\theta} \) would be the square root of this. The special case of Eq. 10 is sometimes called the propagation of error formula, but I think it’s better to use that name for the more general Eq. 8.