Mengxhi was asking in class why the SE's for the terms generated by poly() were identical.

For background on exactly what the poly() function does, see the long answer starting with the bold faced heading "Raw Polynomials" at

`https://stackoverflow.com/questions/19484053/what-does-the-r-function-poly-really-do`

The key point is that `poly(x,p)` generates a basis (new set of columns of the X matrix) for the same set of models as $x, x^2, x^3, ...x^p$, with the following properties:

1. Each new column in the basis has length 1
2. The columns are orthogonal to each other
3. All the columns are orthogonal to the column of 1's that is needed for the intercept in the model

This means we take the original matrix X for the polynomial terms

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^p \\ 1 & x_2 & x_2^2 & \cdots & x_2^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^p \end{bmatrix} = \begin{bmatrix} 1 & | & | & \cdots & | \\ \vdots & x^1 & x^2 & \cdots & x^p \\ 1 & | & | & \cdots & | \end{bmatrix}$$

and replace it with

$$X_{poly} = \begin{bmatrix} 1 & | & | & \cdots & | \\ \vdots & o_1 & o_2 & \cdots & o_p \\ 1 & | & | & \cdots & | \end{bmatrix}$$

where the columns $o_1, o_2, \ldots, o_p$ satisfy

1. $||o_j|| = \sqrt{o_j^T o_j} = 1$ for all $j = 1, \ldots, p$
2. $o_j^T o_k = 0$ whenever $j \neq k$
3. $o_j^T 1 = 1^T o_j = 0$ for all j

(for details on where the columns o_j's come from, see the stackoverflow answer above).

Now when we fit `lm(y ~ X`$_{poly}$`)` [which is the same as `lm(y ~ poly(x, p))`], the standard errors for the $\hat{\beta}$'s will be the square roots of the diagonal elements of our old friend from the matrix algebra of multiple regression, $(X_{poly}^T X_{poly})^{-1} s^2$

Let's see what this is:

$$X_{poly}^T X_{poly} = \begin{bmatrix} 1 & \cdots & 1 \\ - & o_1^T & - \\ \vdots & \vdots & \vdots \\ - & o_p^T & - \end{bmatrix} \begin{bmatrix} 1 & | & | & \cdots & | \\ \vdots & o_1 & o_2 & \cdots & o_p \\ 1 & | & | & \cdots & | \end{bmatrix}$$

$$= \begin{bmatrix} n & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix},$$

using the three properties of the columns o_j that we listed above.

So $(X_{poly}^T X_{poly})^{-1}$ will be

$$(X_{poly}^T X_{poly})^{-1} = \begin{bmatrix} \frac{1}{n} & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Therefore, picking off the diagonal elements of the formula $Var(\hat{\beta}) = (X_{poly}^T X_{poly})^{-1} s^2$ you can see that

$Var(\hat{\beta}_0) = \frac{1}{n} s^2$

$Var(\hat{\beta}_1) = s^2$

$Var(\hat{\beta}_2) = s^2$

$\vdots$

$Var(\hat{\beta}_p) = s^2$

Since all the $\hat{\beta}_j$'s (except for $\hat{\beta}_0$) have the same variances, they will also have the same SE's (square roots of the variances).

(and so... matrix algebra helps us understand a curious result when we fit models with poly(x,p)...)

hope this helps,

-BJ