Paul Johnson
August 18, 2004

What about b's and betas and R-square?

Start with a model like this:

$$y_i = b_0 + b_1 X1_i + b_2 X2_i + ... + b_k Xk_i + e_i \quad i = 1, ..., N \tag{1}$$

and through ______________ procedure, you make estimates with which to calculate predicted values:

$$\widehat{y}_i = \widehat{b}_0 + \widehat{b}_1 X1_i + \widehat{b}_2 X2_i + ... + \widehat{b}_k Xk_i \quad i = 1, ..., N \tag{2}$$

1 How are you supposed to interpret the $\hat{b}'s$?

These are "partial regression coefficients".

"other things equal, a 1 unit increase in X2 causes an estimated $\hat{b}_2$ unit increase in the predicted value of y".

Maybe the calculus says it best:

$$\frac{d\hat{y}}{dX2} = \hat{b}_2$$

This foreshadows the problem with multicollinearity, by the way. We prefaced with "other things equal," but if there is multicollinearity, then other things don't remain equal when you change $X2$.

2 Betas. AKA

3 Standardized Regression Coefficients

If you do a regression in which you replace y_i and $X1_i$ and $X2_i$ and $X3_i$ by standardized variables, what you get is called a standardized regresison equation, and the estimated coefficients are "standardized regression coefficients" and, in the slang of statistics, they are called "Betas.".

Recall a standardized variable is calculated like so:

$$standardized\, y_i = \frac{y_i - \bar{y}}{s.d.(y_i)}$$

Or, if you like to use s_y for standard deviation of y, write it that way instead:

$$standardized\, y_i = \frac{y_i - \bar{y}}{s_y}$$

By definition, all standardized variables have a mean of 0 and a standard deviation of 1. See why?

So if you standardize the variables in a regression model, you have a model like so:

$$\left(\frac{y_i - \bar{y}}{s_y}\right) = \beta_1 \left(\frac{X1_i - \bar{X}1}{s_{X1}}\right) + \beta_2 \left(\frac{X2_i - \overline{X2}}{s_{X2}}\right) + \beta_3 \left(\frac{X3_i - \overline{X3}}{s_{X3}}\right) + u_i \tag{3}$$

The error term has a new name because it gets "automagically" rescaled when you rescale the variables.

1. Why do they do this?

They seek an easy comparison, like "a one standard deviation rise in X1 causes a $\widehat{\beta}_1$-standard-deviation-increase in y." So, if X1 is measured in "dollars" and y is measured in some grossly different unit, like "bushels of wheat per year", the standardization intends to make them comparable.

2. Guess where the y-intercept term went?

3. How does the beta, say β_1 differ from the unstandardized coefficient, b_1?

To find out, let's just make some scale translations in the orignal regression equation. Let's change X1 and see what effect they have.

A. Suppose we change X1 by a constant, so we replace X1 by $X1_i - \overline{X1}$.

Does that have an effect?
Not really.
Start here:

$$y_i = b_0 + b_1X1_i + b_2X2_i + ... + b_kXk_i + e_i$$

change X1:

$$y_i^* = b_0^* + b_1^*(X1_i - \overline{X1}) + b_2^*X2_i + ... + b_k^*Xk_i + v_i$$

If you estimated that model with data, your estimate $\widehat{b}_1^*$ would be equal to the original estimate $\hat{b}_1$.
The estimate of the constant would be changed, however. It would be $b_1^*\overline{X1}$ units different.
B. Suppose we change X1 by dividing it by a constant, which we could call s_{X1} for fun. Then the model changes to:

$$y_i^* = b_0^* + b_1^*(\frac{X1_i}{s_{X1}}) + b_2^*X2_i + ... + b_k^*Xk_i + v_i$$

How would the estimate of b_1^* differ from b_1 in the original model? Obviously, the two are proportionally related, as you can clearly see that

$$\frac{b_1^*}{s_{X1}} = b_1$$

or

$$b_1^* = b_1 * s_{X1}$$

So, if you take any independent variable and divide it by a constant, the only impact is that you end up re-scaling the parameter estimate for that variable. Humpf!
So look at 1 and then compare it against 3. After a while it becomes apparent:

$$\beta_1 = \frac{s_{X1}}{s_y} b_1$$

You can prove this to yourself by multiplying 3 by s_y

$$(y_i - \bar{y}) = \beta_1 \left[\frac{s_y}{s_{X1}}\right] \left(X1_i - \overline{X1}\right) + \beta_2 s_y \left(\frac{X2_i - \overline{X2}}{s_{X2}}\right) + \beta_3 s_y \left(\frac{X3_i - \overline{X3}}{s_{X3}}\right) + u_i \tag{4}$$

4 Betas are no good.

King says, flat out, that the people who want betas because they can compare the “impact” of variables are misguided. They want to say “a 1 standard deviation increase in X1 causes a β_1 standard deviations in crease in y_i.”
There are big problems, however.
1. If we knew for sure the “true standard deviation” σ_{X1} the above procedure might not be a total disaster. However, we don’t. We estimate that by the sample standard deviation, s_{X1}. That means that, even if the underlying regression relationship is the same, then different samples will have different standardized coefficients. There was never a regression assumption that the variance of X1 is fixed, and it should not matter. But betas make it matter.
This means that betas in a single sample are not so meaningful as we originally thought, and furthermore
2. One must not compare betas across samples, because differences in the distribution of X in two cases will affect standardized coefficients.
3. Does standardization have any meaning for dichotomous variables? A “one standard deviation increase in the variable ’male’ causes....”

5 What about the R^2.

King's "how not to lie with statistics" essay digs into this pretty well.

First start by noting that R^2 is not a proper parameter estimate of an underlying parameter of the model. Some (Luskin, Lewis-Beck) will contend otherwise.

Here are some bad things about R^2

1. R^2 depends on the variance of the X's
2. adding variables always makes the R^2 get bigger. The "adjusted-R^2" statistic is an ad hoc adjustment to penalize the addition of variables.
3. No absolute standard exists to answer the question "is my R^2 good enough".
4. Emphasis on R^2 undercuts the main objective of understanding the relationship between X and Y and finding the coefficients of that relationship.

As far as I know, the best use of the R^2 is this. Given a dependent variable, suppose we estimate several models with different functional forms. The R^2 is a good criterion for selecting the "best" one.

As far as I know, there is another pretty good use for R^2. That is in the diagnosis of multi collinearity.