Re: A multiple regression stumper

Source: http://sci.tech−archive.net/Archive/sci.stat.math/2005−09/msg00430.html

- *From*: "Reef Fish" <Large_Nassau_Grouper@xxxxxxxxx>

- *Date*: 29 Sep 2005 11:59:24 −0700

rick.deshon@xxxxxxxxx wrote:
> Hi All.
>
> I can't figure out the solution to what should be a fairly
> straightforward regression problem.
>
> Assume you have a set of variables (X) that you use to predict a single
> variable (Y) in a standard multiple regression model. X is nxp and Y
> is nx1.
>
> In this model, Y = Xb + e, where e is a nx1 vector of residuals.
>
> The OLS estimate of b is inv(X`X)*X'Y. Consider b to be a (px1) vector
> of optimal weights that minimize the variance of e.

So far, standard, as you say.

>
> One way to examine the quality of the fitted regression is to compute
> R^2 (the coefficient of variation or determination). R2 = (b'*
> cov_XY)/var(Y) where cov_XY is a px1 vector of covariances between the
> columns of X with Y (cov_XY = (X'Y)/(n−1)) and var(Y) is the variance
> of the vector Y. Conceptually, R2 is the ratio of predictable variance
> in Y to total variance in Y.
>
> I would like to compute R2 for non−optimal sets of weights. What
> happens to R2 as you use less and less optimal weights?

By non−optimal sets of weights, I think you mean the estimates b
that is not "least squares", and so the SSE will be larger, and
your R−square will be smaller.

> This would be simple under normal circumstances but I'd like to do it
> for a special case where you don't know Y. In other words, you have X,
> b, and you know the R2 for the optimal model. Further, using knowledge
> of X, b, and the optimal R2 you can compute the variance of Y so you
> know that quantity also.

>
> Is it possible to estimate R2 for non−optimal weights if you know b, X,
> and var(y)? The only missing quantity is cov_XY but b clearly has
> information on these missing covariances of X and Y. I have not been
> able to determine a unique solution to this apparently simple problem.
> Perhaps an orthogonal projection?

A most unusual question for an OLS problem. I couldn't help but wonder WHY you ask such a question, and what is the practical significance of your inquiry.

If you don't know Y, then in what sense do you mean by "predicting Y" which is unknown.
>
> Thanks for any insights you can provide!
>
> Rick

One insight I can provide is that the R of your R2 is the simple correlation between the observed Y and the fitted Y in your "usual" OLS regression.

Given that you don't know Y, but some Witches of the West gave you the regression estimates b and the variance of Y, if I were you, I would just forget about the problem and enjoy a few sights of the Boston Harbor and have lobster tail for dinner as I'll be doing tonight.

−− Bob.

.

- ***References***:
  - ♦ ***A multiple regression stumper***
    - ◊ *From:* rick . deshon

- Prev by Date: ***Handling survey option "I don't know"***
- Next by Date: ***Re: Standard Deviation of PISA***
- Previous by thread: ***A multiple regression stumper***
- Next by thread: ***Re: A multiple regression stumper***
- Index(es):
  - ♦ ***Date***
  - ♦ ***Thread***