

Re: r-Squared Question

Source: <http://sci.tech-archive.net/Archive/sci.stat.math/2005-07/msg00341.html>

- *From:* Jerry Dallal <gdallal@xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx>
 - *Date:* Thu, 14 Jul 2005 23:26:53 -0300
-

Reef Fish wrote:

Jerry Dallal wrote:

Reef Fish wrote:

Jerry Dallal wrote:

Reef Fish wrote:

Jerry Dallal wrote:

< Snip >

Also, p 241:

"The coefficient of multiple determination, denoted R^2 , is defined as follows:

$$(7.35) R^2 = SSR/SSTO = 1 - SSE/SSTO$$

Re: r-Squared Question

That's better, as the definition.

Ah, this came FIRST, didn't it? (7.35). You were putting (7.71) first in this post as if it were the definition when Neter et al were just relating some of the ANOVA table entries to little r^2 , in the SIMPLE regression chapter, I presume, because the relation applies ONLY to simple regression.

No, not a typo. The page numbers and equation numbers are correct. r^2 is defined for simple linear regression; R^2 for multiple regression.

Never said THAT was a typo. Read what I wrote again. I said you chose to show (7.71) FIRST, instead of the definition (7.35).

But it's not 7.71. It's 3.71!

"It measures the proportion of total variation fitted by the regression".

I've been using that for DECADES in my Lecture Notes.

That's why I like your suggestion of "variation fitted". No text that I've read has an equally suitable replacement for "explained by". It's all mumbo-jumbo.

I am quite sure others have used much less misleading terms than "percent variation explained". My co-author Harry Roberts did use the word "explain" but immediately explained at length that it must NOT be taken to mean causal or other meaning of "explain". In retrospect, I should have suggested the simple, unambiguous wording of "variation fitted" because that's all it is, no more,

Re: r-Squared Question

no less.

Less misleading, yes. Concise, no. The language is often so tortured as to be unintelligible to a naive audience, hence my descriptor "mumbo-jumbo".

So, what happened to this:

```
JD> Kleinbaum et al,, latest: (RegSS-ResSS)/TotSS
```

```
RF> IMPOSSIBLE! It's WRONG. That's not R^2 at all. I assume it's  
RF> your copying error.
```

or how YOU and the others got the $R^2 = -.03$?

I assume it's typo and carelessness respectively, but wanted to know if otherwise.

-- Bob.

Typo, yes; but not completely careless

Sorry, the "respectively" did not make it clear that the typo was referring to ONLY

Re: r-Squared Question

JD> Kleinbaum et al,, latest: (RegSS-ResSS)/TotSS

RF> IMPOSSIBLE! It's WRONG. That's not R^2 at all. I assume it's
RF> your copying error.

which you posted for the first time. So, what was ACTUALLY in
Kleinbaum's book?

As I posted in my earlier correction, $(TotSS-ResSS)/TotSS$

The "careless" was referring to

RF> > or how YOU and the others got the $R^2 = -.03$?

In Google, you made THREE consecutive posts, at 8:08 pm, 8:16 pm and
8:27 om of July 12.

Your correction of your own post (8:27 pm) was this:

JD> I've canceled my earlier post, but given the way cancels
JD> propagate, some copies of the original will survive. So, for the
JD> record, keep this post and the one with $R^2 = -0.03$, and ignore the
one
JD> with $R^2 = 0$.

You KEPT the $R^2 = -.03$,

which certainly did not follow from any of the definitions you
cited.

Re: r-Squared Question

I gave the data!

```
> X   Y     YHat Y-Yhat
> 1  101    97    4
> 2  102    99    3
> 3  103   101    2
> 4  104   103    1
> 5  105   105    0
> 6  106   107   -1
> 7  107   109   -2
> 8  108   111   -3
> 9  109   113   -4
> 10 110   115   -5
```

X and Y are the data. Yhat is the fit of the model proposed by the poster. The values were given by him. They have a correlation coefficient of 1 with Y. Hence, the square of the correlation between observed and expected values is 1, even though the fit is far from perfect. This is why he was asking whether it was a "defect" in R^2 .

I calculated the residuals: Y-Yhat

ResSS is the sum of their (residuals) squares = 85. However, TSS = $\text{Sum}[(Y-105.5)^2]$ is only 82.5.

I plug those numbers into $1-\text{ResSS}/\text{TotSS}$ and get -0.03. Do you get something different?

One might also "argue" that since the model does worse than no model at all, that the RegSS is negative (the net amount it accounts for is negative) and get at it that way.

Given these Ys and Yhats, -0.03 is what you get when you plug the numbers into the formula! It's like assigning code numbers to subjects' ethnicity and calculating the mean. It's *worse* than meaningless (because the result is ennobled by having gone through a statistics program), but a number pops out nonetheless.

Hey, this *is* Alice in Wonderland! The whole point is that the result is nonsensical. But it *is* -0.03. :-)

--Jerry

.