

**"Few things are harder to put up with than the annoyance of a good (statistics) example"<sup>1</sup>**

Bruce Ratner, Ph.D.

EXAMPLE: Different Data, Identical Regression Models: Which Model is Better?

As a data miner, I can say with 100% confidence that I have discovered two patterns when submitting a Regression Model Proposal to a prospective client, and handing out the final presentation deck to the client:

- 1) The prospective client always looks at the last page of the proposal for the cost of the project.
- 2) If I get the project, the client always looks for the page, seeking the value of R-squared for the regression model built.

As a consulting statistician presenting the realized regression model, I have an effortful job explaining (rather, re-teaching) to otherwise bright clients, who inform me of the statistics courses under their ("black") belts, some basics concepts that they never fully understood at the outset. At this point, I know that building the model is not the hard part of the project, but presenting the results is. Invariably, the first obstacle is explaining why their *pet* variables are not in the model. The second hurdle is to absorb their shock when they discover their misunderstanding of R-squared.

Accordingly, I always scrounge for statistical tidbits as aids in helping me explain (by show-and-tell) statistical concepts, which the clients have unwittingly misunderstood for too long. My latest statistical tidbit: I build two OLS regression models, regressing Y1 on X1, and regressing Y2 on X2, using the data in Table 1, below. Both OLS regression models are identical! Which Model is Better? Which pair of variables is better explained by this one model? Which model produces the more accurate predictions?

$$Y = 134.94743 + 0.10005 * X$$

If you like, please [email](#) me your answer, or [email](#) me for the answer.

Thanks.



1- Mark Twain

**Table 1. The Data**

<b>ID</b>	<b>X1</b>	<b>Y1</b>	<b>X2</b>	<b>Y2</b>
1	158.2	157.8	140.4	153.2
2	214.9	146.6	211.9	157.4
3	153.2	147.5	152.4	149.6
4	196.0	153.1	124.7	154.9
5	88.5	143.7	103.9	145.2
6	55.5	132.3	128.5	141.7
7	86.4	144.3	187.1	159.7
8	223.6	169.1	168.5	145.3
9	256.9	160.9	138.3	151.7
10	252.4	157.1	137.9	141.7
11	20.9	141.6	203.3	153.3
12	92.9	145.4	102.5	145.8

To download The Data in "txt" format, click [here](#).