What’s an R\(^2\) and what’s a t-value?

Two Important Statistics
The R\(^2\) and the t-value

Notes

By
Edward Leamer
January 26, 1999

What is an R\(^2\)? ................................................................. 2
What is a t-value? ................................................................. 2
Examples of the effect of sample size ........................................ 3
Questions about these numbers that you might be asking yourself: ........................................ 4
   Why is the difference between the adjusted R\(^2\) and the R\(^2\) hardly noticeable, except when the sample size is very small? ........................................ 4
   Why is the adjusted R\(^2\) a U-shaped function of sample size? ........................................ 4
   Why does the standard error fall with sample size and the t-value increase? ........................ 4
How big should be my R\(^2\)? ................................................................ 4
Example: Two Multiple Regressions that Explain GDP .................................................. 5
Which of these equations is actually better? ................................................................. 6
How big should my t-value be? I’m a t=1 kind of guy. .................................................. 6
Summary Table ........................................................................ 7
What’s an \( R^2 \) and what’s a t-value?

What is an \( R^2 \)?

The \( R^2 \) measures the percent of variation in the “dependent” variable that can be accounted for or “explained” by your “independent” variables. For example, you can account for 97% of the variation in adult heights using wingspan as a predictor. Using the same wingspan variable, you can account for only 40% of the variability in their weights.

What is an estimate, a standard error and a t-value?

I am glad that you asked me about all three of these items at the same time since it reveals that you understand that they are all related. Indeed they come together in a neat little package. When you run a regression, for each predictor (independent/explanatory) variable you will get all three: an estimated coefficient, a standard error, and a t-value.

The estimate represents the computer’s attempt to find the equation that best summarizes the data you are studying. The standard error measures how sure the computer feels about its choice of estimate. The standard error for one variable cannot be compared effortlessly with the standard error of another variable. One reason is that units in which the variables are measured affect the standard errors. The t-value is the ratio of the estimate divided by the standard error. It is another way of measuring how sure the computer feels. Unlike the standard error, the t-value doesn’t depend on the units. Unlike the standard error, the t-value can be compared across all variables.

Remember that all those numbers that come jumping out of the computer when you ask for a regression are only answering four questions:

1. What are the data that you are examining?
2. What is the best line for summarizing these data?
3. How much noise is there around that line?
4. How well is that line determined?

Items 1, 2 and 3 are entirely straightforward, and you should be able to see them in this scatter diagram which compares growth of US real GDP in one quarter against the growth in the previous quarter. You can see that

1. We are working with the growth of real GDP.
2. The best line summarizing this scatter has a slight upward slope.
3. There is a lot of noise around that line.

But what about item 4: “How well is that line determined?” Maybe you are thinking that you can see the answer to that question also in this picture. You could put a line with a much steeper slope into that picture and not distort the relationship very much. Must be that line is not well determined by the data. That’s treating questions 3 and 4 as if they were identical: if there is a lot of noise, you cannot tell where the line is.

\[ \text{Quarterly Growth at Annualized Rates, 1947-1998} \]

Is there much momentum in US real GDP?

1 How far they can stretch their arms horizontally.
What’s an $R^2$ and what’s a $t$-value?

But there is something else that affects the answer to question 4: the sample size. Noisy experiments don’t provide the best information, but if you run a lot of them, you can average out the noise to find the truth. To illustrate that point I have successively trimmed more and more data from the sample and computed the corresponding regression. The numbers are reported in Table 1 below. If all the quarterly data are included there are 204 observations. If a little less than a decade of data are trimmed, the number drops to 170; then to 130, to 90, to 50 and last to only 10. Take a look at the numbers in the tables and the figures below that illustrate the numbers. Think of some questions that these numbers suggest. We’ll answer those questions next.

**Examples of the effect of sample size**

Table 1
The Effect of Sample Size
R-squared and Slope of Regression of Growth on Growth(-1)

<table>
<thead>
<tr>
<th>Period</th>
<th>N. of Obs</th>
<th>R-squared</th>
<th>Adj. R-sq</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1947:3 1998:2</td>
<td>204</td>
<td>0.12</td>
<td>0.11</td>
<td>0.34</td>
<td>0.07</td>
<td>5.16</td>
</tr>
<tr>
<td>1956:1 1998:2</td>
<td>170</td>
<td>0.09</td>
<td>0.08</td>
<td>0.30</td>
<td>0.07</td>
<td>4.07</td>
</tr>
<tr>
<td>1966:1 1998:2</td>
<td>130</td>
<td>0.08</td>
<td>0.07</td>
<td>0.27</td>
<td>0.08</td>
<td>3.25</td>
</tr>
<tr>
<td>1976:1 1998:2</td>
<td>90</td>
<td>0.08</td>
<td>0.07</td>
<td>0.29</td>
<td>0.10</td>
<td>2.83</td>
</tr>
<tr>
<td>1986:1 1998:2</td>
<td>50</td>
<td>0.12</td>
<td>0.10</td>
<td>0.35</td>
<td>0.14</td>
<td>2.56</td>
</tr>
<tr>
<td>1996:1 1998:2</td>
<td>10</td>
<td>0.47</td>
<td>0.41</td>
<td>-0.78</td>
<td>0.29</td>
<td>-2.69</td>
</tr>
</tbody>
</table>
What’s an $R^2$ and what’s a t-value?

Questions about these numbers that you might be asking yourself:

Why is the difference between the adjusted $R^2$ and the $R^2$ hardly noticeable, except when the sample size is very small?

Keep in mind that the $R^2$ goes up each time you add another variable to the equation. If the number of variables is a significant fraction of the number of observations, you are probably “over-fitting” – asking too much of the data. The adjusted $R^2$ is a measure of the noise around the regression line, correcting for this overfitting problem. This correction doesn’t matter much if the number of observations is a large multiple of the number of variables in the equation.

Why is the adjusted $R^2$ a U-shaped function of sample size?

This is not a general property but instead is a function of some special features of the GDP growth series. If the relationship were stable over time, that adjusted $R^2$ should not depend on the sample size. It is not difficult to form some conjectures regarding the U-shaped function that we see here. The older data, prior to 1985, is more variable and the intertemporal correlation at that time was larger. But the most recent data is anomalous – suggesting a negative relationship – you can see the negative estimate. It doesn’t work well to combine that most recent data with any of the less anomalous data.

Why does the standard error fall with sample size and the t-value increase?

As you add more and more data, the choice of the best fitting regression line becomes more and more clear. This means that the standard error declines. Technically, the standard error of a coefficient declines like $1/\sqrt{n}$. The t-value is the ratio of the estimate divided by the standard error. If the estimate is stable, the t-value thus grows like $\sqrt{n}$. Can you see that the t-value does indeed grow like the square root of the sample size?

How big should be my $R^2$?

A highly trained professional statistician can with confidence tell you that an $R^2$ of one is great and an $R^2$ of zero is terrible, but can’t really say anything more. What is big or small between those extremes is a matter of judgement and depends on the context and the variable that is being explained.

It’s best to think of it as a horse race. There is no absolute standard. It depends on how fast the other horses run. But make sure the horses are on the same track. If one equation is predicting the growth of GDP, don’t compare that $R^2$ with another equation that is predicting some other variable or the same variable for a different sample period.

To see how the context matters, it will help to take a look at some examples. Below are two regressions estimated from US real GDP data from 1946 second quarter to 1998 second quarter. Each is meant to help to answer similar questions: Do unemployment and/or federal defense expenditures have a positive or a negative effect on real GDP?

The first equation explains the rate of growth of GDP. The second explains the level of GDP. In the first equation, the predictor variables are the growth rate in the previous quarter, the percentage of the workforce that was unemployed in the previous quarter, the change in the percent unemployed, and the growth in real defense expenditures. In the second equation, the predictor variables are the GDP in the previous quarter, the number of unemployed in the previous quarter, and the change in the number of unemployed, and the level of real defense expenditures in the previous quarter.

The second equation that explains the level of real GDP looks a lot better, doesn’t it. It has an $R^2$ nearly equal to one while the other equation has that miserable little $R^2$ of only 0.145. The second equation has that great predictor variable, the lagged value of real GDP, with a t-value of 394.0. The best t-value the first equation can find is only –2.799. Equation 2 is the best choice, that’s for sure.

In fact, these two equations are saying pretty much the same thing:

1. GDP growth is relatively high following periods with high but falling unemployment.
2. The direction that unemployment is moving is ten times more important than the level of unemployment.
3. It is very difficult to detect any effect of defense expenditures on growth.
Example: Two Multiple Regressions that Explain GDP

Equation 1

Dependent Variable: Growth of Real GDP
Method: Least Squares
Date: 01/16/99   Time: 09:50
Included observations: 150 after adjusting endpoints

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>1.121</td>
<td>1.265</td>
<td>0.886</td>
<td>0.377</td>
</tr>
<tr>
<td>Growth(-1)</td>
<td>0.112</td>
<td>0.103</td>
<td>1.094</td>
<td>0.276</td>
</tr>
<tr>
<td>Unemploy %(-1)</td>
<td>0.301</td>
<td>0.194</td>
<td>1.549</td>
<td>0.123</td>
</tr>
<tr>
<td>Change in Unemploy %(-1))</td>
<td>-3.233</td>
<td>1.155</td>
<td>-2.799</td>
<td>0.006</td>
</tr>
<tr>
<td>Growth in Real Defense(-1)</td>
<td>-0.018</td>
<td>0.034</td>
<td>-0.534</td>
<td>0.594</td>
</tr>
</tbody>
</table>

R-squared 0.145     Mean dependent var 3.341
Adjusted R-squared 0.122     S.D. dependent var 3.727
S.E. of regression 3.494     Akaike info criterion 5.372
Sum squared resid 1769.7     Schwarz criterion 5.473
Log likelihood -397.9     F-statistic 6.154
Durbin-Watson stat 1.892     Prob(F-statistic) 0.0001

Equation 2

Dependent Variable: Real GDP ($b 1992)
Method: Least Squares
Date: 01/24/99   Time: 17:37
Sample(adjusted): 1948:3 1998:2
Included observations: 200 after adjusting endpoints

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>7.20</td>
<td>13.25</td>
<td>0.54</td>
<td>0.59</td>
</tr>
<tr>
<td>Real GDP(-1)</td>
<td>1.00</td>
<td>0.003</td>
<td>394</td>
<td>0.000</td>
</tr>
<tr>
<td>Unemployed (-1) (billion)</td>
<td>4810</td>
<td>1637</td>
<td>2.94</td>
<td>0.004</td>
</tr>
<tr>
<td>Change in unemployed(-1))</td>
<td>-43174</td>
<td>6875</td>
<td>-6.28</td>
<td>0.000</td>
</tr>
<tr>
<td>Fed Defense $b 1992</td>
<td>0.00</td>
<td>0.04</td>
<td>0.01</td>
<td>0.99</td>
</tr>
</tbody>
</table>

R-squared 0.9996     Mean dependent var 3957.83
Adjusted R-squared 0.9996     S.D. dependent var 1720.77
S.E. of regression 33     Akaike info criterion 9.86
Sum squared resid 213529     Schwarz criterion 9.94
Log likelihood -981     F-statistic 134481
Durbin-Watson stat 1.77     Prob(F-statistic) 0.00000
What’s an $R^2$ and what’s a t-value?

Take another look at Equation 1. The previous quarters’ growth of GDP, the unemployment rate and the growth in defense expenditures together account for 14.5% of the variability of the rate of growth of real GDP across quarters. That looks pretty bad. Equation 2 looks a lot better with an $R^2$ of 0.9996. But think again. The $R^2$ of the first equation is an answer to a different question than the $R^2$ of the second regression. The first is asking what percentage of the historical variation in the growth rate can be explained by these variables. The second is asking what percentage of the historical variability of real GDP can be explained. Most of the historical variability of real GDP is just the time trend. Any variable with a strong time trend can account for a lot of the so-called variability in the level of real GDP. The one that is doing the job here is the previous quarter’s GDP. It gets that astronomical t-value of 394. But to say that a good prediction of next quarter’s GDP is this quarter’s GDP plus $4810 for each unemployed person may or may not give you a very accurate predictor of the growth of GDP. To decide, you’d have to compute the $R^2$ differently.

Which of these equations is actually better?

Although these two equations are saying about the same thing regarding the determinants of GDP growth, I prefer the one you don’t like – equation 1 that explains the growth in real GDP with the low $R^2$. I like this equation because I think it is capable of tracking the behavior of the economy over time better than the other one. I think that over time, with a growing economy, it doesn’t make sense to think of each unemployed worker making a constant contribution to the level of GDP. I think that it does make sense to think of there being a constant effect of the rate of unemployment on the rate of growth.

How big should my t-value be? I’m a t=1 kind of guy.

The t-value is an answer to the question: “Based on these data, can we tell what is the sign of the coefficient?” If the t-value is large, the data are suggesting that the sign conforms with the sign of the estimate. If the t-value is small, the data are saying: “I cannot tell the sign of the effect.”

The t-value is the estimate divided by the standard error. If the absolute t-value is one, then a one-standard-error interval around the estimate just includes the zero value. A statistician can give you reason to believe that there is a 68% chance that the true coefficient is within one standard error of its estimate, and a 95% chance that it is within 2 standard errors. Half of the remaining probability is on one side and half on the other. Thus if the t-value is one, there is only a (1.0 - 0.68)/2 = 16% chance that the true value is different in sign from the estimate. If the t-value is equal to two in absolute value, there is a (1.0 – 0.95)/2 = 2.5% chance of an opposite sign.

It is entirely arbitrary whether you choose one or two or some other number as the critical value of the t-statistic. It depends on the odds that you like to bet on. Is 1-0.16 = 84% virtual certainty, or do you need 1.0-.025 = 97.5% for virtual certainty?

The choice of a critical value equal to one is not entirely arbitrary, for another reason. The t-value is also an answer to the question: “What would happen to the adjusted $R^2$ if I omitted this variable from the equation?” If the t-value were greater than one in absolute value, then the adjusted $R^2$ would fall. If the t-value were less than one, then the adjusted $R^2$ would increase. If the adjusted $R^2$ is one target of your analysis, which it should be, variables with t-values less than one in absolute value are asking to be discarded.

There is another property of a t-value that I probably shouldn’t be revealing. If you drop a variable from your equation, the estimated coefficient of any variable with a bigger t-value than the one you omitted cannot change in sign. A big t-value thus means a sturdy inference. What is big? Big compared with the other variables in the equation. (You can do a lot of damage with this information, so don’t tell anyone else.)
### Summary Table

<table>
<thead>
<tr>
<th></th>
<th>Low R²</th>
<th>High R²</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low t-value</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Absolute Misery</td>
<td>- the effect of your favorite variable isn’t detectable using these data, and your equation overall sucks.</td>
<td>Forget that variable. Another variable is doing the work and giving you a high R². Maybe if you discard one of the other variables, then your favorite one will look better. The ones that are correlated with your favorite might be causing the problem.</td>
</tr>
<tr>
<td><strong>High t-value</strong></td>
<td>Oh Joy. You have found a way to measure the effect of your favorite variable using these data. The other variables in your equation are not crowding out your favorite. Maybe you are worried that your equation overall is not explaining much of the variability of the dependent variable. Don’t be upset. This is the normal state of affairs for variables defined in a way that makes them hard to predict, like growth rates or changes over time. Think about medicine as an example. Maybe they can find out that coffee drinking has a statistically significant and even important effect on the probability of heart disease, but still not be able to predict whether or not you will suffer from heart disease, regardless of your coffee-drinking habits.</td>
<td>Bliss – the best it can get. But be alert – your enemies are trying to find just the right variable to include to knock yours out of the equation. And think again how you defined your dependent variable. What it means to have a high R² depends on the variable. An R² of 0.9 in an equation explaining quarterly GDP is no big deal. You can do that using the previous quarter’s GDP. An R² of 0.5 for explaining the growth of GDP using only lagged variables is terrific. It is easier to predict your daughter’s height at age 16 than how much she will grow from birthday 16 to birthday 17.</td>
</tr>
</tbody>
</table>