What does 'proportion of variance explained' mean?

Jeremy Miles, PhD in Psychology, Data Scientist at Google   10 Jun, 2013

You have a set of scores on an outcome variable [for example, a set of research participants' scores on an Internet use scale].

You can calculate the variance of those scores. **The variance is a measure of how much people vary [for example, how much people vary on an Internet use scale; some people might use the Internet a lot, some people might use the Internet only a little, and all possibilities could exist in between. That's what we mean by how much people can vary on an Internet use scale, and when you calculate the variance you get a measure of this variation].**

Then you [use a predictor variable, for example, the same research participants' scores on a Neuroticism scale] to calculate a predicted score for each person [on your outcome variable, for example, Internet use].

[**You use your *predictor* variable, *Neuroticism*, to predict the variation in the actual scores on your *outcome* variable, *Internet* use.**]

You calculate the difference between your predicted scores and the actual scores [to answer the question] **how much variance [in the outcome variable, Internet use] did you explain [with your predictor variable, Neuroticism]?**

If your predicted scores exactly match the participants' [actual] outcome scores, then you've perfectly predicted the [the participants' actual outcome] scores, **and you've explained all of the variance [in those outcome scores]**.

[**But if you've explained only a small proportion of the variance, for instance, only 2% or 3% of the variance, then your predictor variable, e.g., Neuroticism, isn't very good at predicting the variation in your outcome measure, e.g., Internet use.**]