

Cross Validated is a question and answer site for people interested in statistics, machine learning, data analysis, data mining, and data visualization. Join them; it only takes a minute:

Sign up

Here's how it works:

Anybody can ask a question

Anybody can answer

The best answers are voted up and rise to the top

How to use anova for two models comparison?

How should I understand the `anova` result when comparing two models?

Example:

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	9	54.032				
2	7	4.632	2	49.4	37.329	0.0001844 ***

The manpage states: "Compute analysis of variance (or deviance) tables for one or more fitted model objects." However, our professor mentioned that it may be employed for model comparison - that's what I intend to do.

Hence I assume I could use `anova(model1, model2)` and obtain a p-value which tells me whether I should reject the null hypothesis: "the models are the same".

May I state that if the p-value is less than (let's say) 0.05, the models differ significantly?

r regression anova

asked May 15 '15 at 15:13



petrbel

211 2 6

In your example, are model1 and model2 nested? That is, do both models have a shared set of predictor variables and the same outcome variable, but one model has one or more additional predictor variables? – EdM May 15 '15 at 15:26

One is like $Y \sim X + X^2$ and the second one $Y \sim X + X^2 + X^3$ – petrbel May 15 '15 at 15:30

1 Answer

Assuming your models are nested (i.e. same outcome variable and model 2 contains all the variables of model 1 plus 2 additional variables), then the ANOVA results state that the 2 additional variables jointly account for enough variance that you can reject the null hypothesis that the coefficients for both variables equal 0. This is effectively what you said. If both coefficients equal 0 then the models are the same.

Just as an additional note, in case you weren't aware, ANOVA is always equivalent to doing model comparisons. When you are looking at the ANOVA for a single model it gives you the effects for each predictor variable. That is equivalent to doing a model comparison between your full model and a model removing one of the variables. i.e.

Model1 : $y = a + bx_1 + cx_2 + dx_3$; *Model2* : $y = a + bx_1 + cx_2$ will give you the sum of squares (type III) and test statistic for x_3 . Just note that R gives you type I sum of squares. If you need type III, use `car::Anova` or use `anova` and keep changing the order of the variables in the model and only take the sum of squares for the last variable.

edited May 15 '15 at 15:40

answered May 15 '15 at 15:34



le_andrew

1,034 1 8

If I understood you properly, the p-value less than 0.05 proves that the models differ, tight? – petrbel May 15 '15 at 15:48

- 3 I wouldn't use those words (i.e. "prove" and "models differ"), but we mean the same thing. I would say that your data does not support the null hypothesis that the coefficients are both 0 or that the data supports the alternative hypothesis that the coefficients are not both 0. – le_andrew May 15 '15 at 16:53

Before a claim is made that models are proved to be different or that the null hypothesis is not supported, be sure that the data reasonably meet the **assumptions of ANOVA** that underlie the calculation and interpretation of the p-values. – EdM May 15 '15 at 21:37

