

Comparing two models using anova() function in R

From the documentation for `anova()` :

When given a sequence of objects, 'anova' tests the models against one another in the order specified...

What does it mean to test the models against one another? And why does the order matter?

Here is an example from the [GenABEL tutorial](#):

```
> modelAdd = lm(qt~as.numeric(snp1))
> modelDom = lm(qt~I(as.numeric(snp1)>=2))
> modelRec = lm(qt~I(as.numeric(snp1)>=3))
anova(modelAdd, modelGen, test="Chisq")
Analysis of Variance Table

Model 1: qt ~ as.numeric(snp1)
Model 2: qt ~ snp1
  Res.Df  RSS Df Sum of Sq  Pr(>Chi)
1    2372 2320
2    2371 2320   1    0.0489    0.82
anova(modelDom, modelGen, test="Chisq")
Analysis of Variance Table

Model 1: qt ~ I(as.numeric(snp1) >= 2)
Model 2: qt ~ snp1
  Res.Df  RSS Df Sum of Sq  Pr(>Chi)
1    2372 2322
2    2371 2320   1     1.77    0.18
anova(modelRec, modelGen, test="Chisq")
Analysis of Variance Table


Model 1: qt ~ I(as.numeric(snp1) >= 3)
Model 2: qt ~ snp1
  Res.Df  RSS Df Sum of Sq  Pr(>Chi)
1    2372 2324
2    2371 2320   1     3.53    0.057 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


How do I interpret this output?

`r` `anova`

edited Sep 19 '16 at 17:52

asked Mar 26 '13 at 10:01

 **amoeba**  
50.6k 12 172 232

 **qed**  
858 2 12 27

1 Answer


When you use `anova(lm.1,lm.2,test="Chisq")` , it performs the Chi-square test to compare `lm.1` and `lm.2` (i.e. it tests whether reduction in the residual sum of squares are statistically significant or not). Note that this makes sense only if `lm.1` and `lm.2` are nested models.


For example, in the 1st anova that you used, the p-value of the test is 0.82. It means that the fitted model "modelAdd" is not significantly different from modelGen at the level of  $\alpha = 0.05$ . However, using the p-value in the 3rd anova, the model "modelRec" is significantly different form model "modelGen" at  $\alpha = 0.1$ .

Check out [ANOVA for Linear Model Fits](#) as well.

edited Sep 19 '16 at 17:53

answered Mar 27 '13 at 21:29

 **amoeba**  
50.6k 12 172 232

 **Stat**  
5,296 1 14 39

- 1 But does that imply anything about whether one of the two is better? Thanks! – qed Mar 29 '13 at 17:22
- 4 It depends on how you define the term "better". If you define it as the model that provides less residual sum of squares, then the answer is yes. This is because, this test compares the reduction in the residual sum of squares. – Stat Mar 29 '13 at 21:32
- 1 On the other hand, if the two models are not significantly different, could one argue that the simpler model is "better"? I am thinking about parcimony here. – Sininho May 31 '16 at 15:00
- what if the anova(mod1, mod2, test = "LRT") what's the difference does this make? – ElleryL Aug 13 '17 at 21:36

protected by amoeba Sep 19 '16 at 17:45

Thank you for your interest in this question. Because it has attracted low-quality or spam answers that had to be removed, posting an answer now requires 10 reputation on this site (the association bonus does not count).