Cross Validated

What test can I use to compare slopes from two or more regression models?

Asked 7 years, 8 months ago   Active 4 months ago   Viewed 47k times

△

30

▽

★

16

↺

I would like to test the difference in response of two variables to one predictor. Here is a minimal reproducible example.

```
library(nlme)
## gls is used in the application; lm would suffice for this example
m.set <- gls(Sepal.Length ~ Petal.Width, data = iris,
             subset = Species == "setosa")
m.vir <- gls(Sepal.Length ~ Petal.Width, data = iris,
             subset = Species == "virginica")
m.ver <- gls(Sepal.Length ~ Petal.Width, data = iris,
             subset = Species == "versicolor")
```

I can see that the slope coefficients are different:

```
m.set$coefficients
(Intercept) Petal.Width
  4.7771775   0.9301727
m.vir$coefficients
(Intercept) Petal.Width
  5.2694172   0.6508306
m.ver$coefficients
(Intercept) Petal.Width
  4.044640   1.426365
```

2. How can I test the difference between residual variances?

3. What is a simple, effective way to present these comparisons?

A related question, Method to compare variable coefficient in two regression models, suggests re-running the model with a dummy variable to differentiate the slopes, are there options that would allow the use of independent data sets?

r　data-visualization　multivariate-analysis　hypothesis-testing

edited Apr 13 '17 at 12:44

Community ♦
1

asked Jul 25 '12 at 14:50

Abe
3,081 ● 5 ● 23 ● 40

In regards to the first question see stats.stackexchange.com/questions/55501/…. – russellpierce Sep 24 '14 at 19:57

4 Answers

Active　Oldest　Votes

How can I test the difference between slopes?

▲

23

▼

✓

↺

Include a dummy for species, let it interact with P_i, and see if this dummy is significant. Let L_i be the sepal length and P_i be the pedal width and S_1, S_2, S_3 be the dummy variables for the three species. The compare the model

$$E(L_i) = \beta_0 + \beta_1 P_i$$

with the model that allows the effect of P_i to be different for each species:

$$E(L_i) = \alpha_0 + \alpha_1 S_2 + \alpha_2 S_3 + \alpha_4 P_i + \alpha_5 P_i S_2 + \alpha_6 P_i S_3$$

The GLS estimators are MLEs and the first model is a submodel on the second, so you can use the likelihood ratio test here. The likelihoods can be extracted using the `logLik` function and the degrees of freedom for the test will be 4 since you've deleted 4 parameters to arrive at the submodel.

What is a simple, effective way to present the comparison?

I think the most appealing way would be to plot the regression lines for each species all on the same axes, maybe with error bars based on the standard errors. This would make the difference (or non-difference) between the species and their relationship to P_i very apparent.

Edit: I noticed another question has been added to the body. So, I'm adding an answer to that:

How can I test the difference between residual variances?

For this, you'll need to stratify the data set and fit separate models since, the interaction-based model I suggested will constraint the residual variance to be the same in every group. If you fit separate models, this constraint goes away. In that case, you can still use the likelihood ratio test (the likelihood for the larger model is now calculated by summing the likelihoods from the three separate models). The "null" model depends on what you want to compare it with

By using our site, you acknowledge that you have read and understand our Cookie Policy, Privacy Policy, and our Terms of Service.　　✕

freedom for the test are then 2.

- If you want to test the variance jointly with the coefficients, then the null model should be the first model I've written above. The degrees of freedom for the test is then 6.

edited Jul 25 '12 at 15:14   answered Jul 25 '12 at 15:08

Macro
35.8k 🟡 7 ⬜ 131 🟧 142

Why is there no S_1 in the second model? Is `gls(Sepal.Length ~ species:Petal.Width, data = iris)` the correct implementation of the model in R? – Abe Jul 25 '12 at 20:27

Hi @Abe. S_1 is the "reference" species - the regression line for that species is given by $\alpha_0 + \alpha_4 P_i$. If `species` is a categorical variable then I think `gls(Sepal.Length ~ species*Petal.Width, data=iris)` would be the syntax. – Macro Jul 25 '12 at 21:31

@Macro Nice answer (+1)! I wonder whether you could fit the `gls` model but allowing for different residual variances for each Species with the option `weights=varIdent(form=~1|Species)` (regarding the second question)? – COOLSerdash May 19 '13 at 14:06 ✏️

To answer these questions with R code, use the following:

1. How can I test the difference between slopes?

▲

16

▼

Answer: Examine the ANOVA p-value from the interaction of Petal.Width by Species, then compare the slopes using lsmeans::lstrends, as follows.

```
library(lsmeans)
m.interaction <- lm(Sepal.Length ~ Petal.Width*Species, data = iris)
anova(m.interaction)
 Analysis of Variance Table

 Response: Sepal.Length
                    Df Sum Sq Mean Sq  F value Pr(>F)
 Petal.Width         1 68.353  68.353 298.0784 <2e-16 ***
 Species             2  0.035   0.017   0.0754 0.9274
 Petal.Width:Species 2  0.759   0.380   1.6552 0.1947
 Residuals         144 33.021   0.229
 ---
 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Obtain slopes
m.interaction$coefficients
m.lst <- lstrends(m.interaction, "Species", var="Petal.Width")
 Species    Petal.Width.trend        SE  df   lower.CL upper.CL
 setosa             0.9301727 0.6491360 144 -0.3528933 2.213239
 versicolor         1.4263647 0.3459350 144  0.7425981 2.110131
 virginica          0.6508306 0.2490791 144  0.1585071 1.143154

# Compare slopes
pairs(m.lst)
 contrast                estimate        SE  df t.ratio p.value
 setosa - versicolor    -0.4961919 0.7355601 144  -0.675  0.7786
 setosa - virginica      0.2793421 0.6952826 144   0.402  0.9149
 versicolor - virginica  0.7755341 0.4262762 144   1.819  0.1669
```

2. How can I test the difference between residual variances?

If I understand the question, you can compare Pearson correlations with a Fisher transform, also called a "Fisher's r-to-z", as follows.

```r
iris <- as.data.table(iris)
# Calculate Pearson's R
m.correlations <- iris[, cor(Sepal.Length, Petal.Width), by = Species]
m.correlations
# Compare R values with Fisher's R to Z
paired.r(m.correlations[Species=="setosa", V1], m.correlations[Species=="versicolor",
V1],
         n = iris[Species %in% c("setosa", "versicolor"), .N])
paired.r(m.correlations[Species=="setosa", V1], m.correlations[Species=="virginica",
V1],
         n = iris[Species %in% c("setosa", "virginica"), .N])
paired.r(m.correlations[Species=="virginica", V1], m.correlations[Species=="versicolor",
V1],
         n = iris[Species %in% c("virginica", "versicolor"), .N])
```

3. What is a simple, effective way to present these comparisons?

"We used linear regression to compare the relationship of Sepal Length to Petal Width for each Species. We did not find a significant interaction in the relationships of Sepal Length to Petal Width for *I. Setosa* (B = 0.9), *I. Versicolor* (B = 1.4), nor *I. Virginica* (B = 0.6); $F_{(2, 144)}$ = 1.6, p = 0.19. A Fisher's r-to-z comparison indicated that the Pearson correlation for *I. Setosa* (r = 0.28) was significantly lower (p = 0.02) than *I. Versicolor* (r = 0.55). Similarly, the correlation for *I. Virginica* (r = 0.28) was significantly weaker (p = 0.02) than the one observed for *I. Versicolor*."

Finally, always visualize your results!

```r
plotly_interaction <- function(data, x, y, category, colors =
col2rgb(viridis(nlevels(as.factor(data[[category]]))))), ...) {
  # Create Plotly scatter plot of x vs y, with separate lines for each level of the
categorical variable.
  # In other words, create an interaction scatter plot.
  # The "colors" must be supplied in a RGB triplet, as produced by col2rgb().

  require(plotly)
  require(viridis)
  require(broom)

  groups <- unique(data[[category]])

  p <- plot_ly(...)

  for (i in 1:length(groups)) {
    groupData = data[which(data[[category]]==groups[[i]]), ]
    p <- add_lines(p, data = groupData,
                   y = fitted(lm(data = groupData, groupData[[y]] ~ groupData[[x]])),
                   x = groupData[[x]],
                   line = list(color = paste('rgb', '(', paste(colors[, i], collapse =
", "), ')')),
                   name = groups[[i]],
                   showlegend = FALSE)
    p <- add_ribbons(p, data = augment(lm(data = groupData, groupData[[y]] ~
groupData[[x]])),
                     y = groupData[[y]],
                     x = groupData[[x]],
                     ymin = ~.fitted - 1.96 * .se.fit,
                     ymax = ~.fitted + 1.96 * .se.fit,
                     line = list(color = paste('rgba','(', paste(colors[, i], collapse =
", "), ', 0.05)')),
                     fillcolor = paste('rgba', '(', paste(colors[, i], collapse = ", "),
', 0.1)'),
                     showlegend = FALSE)
    p <- add_markers(p, data = groupData,
```
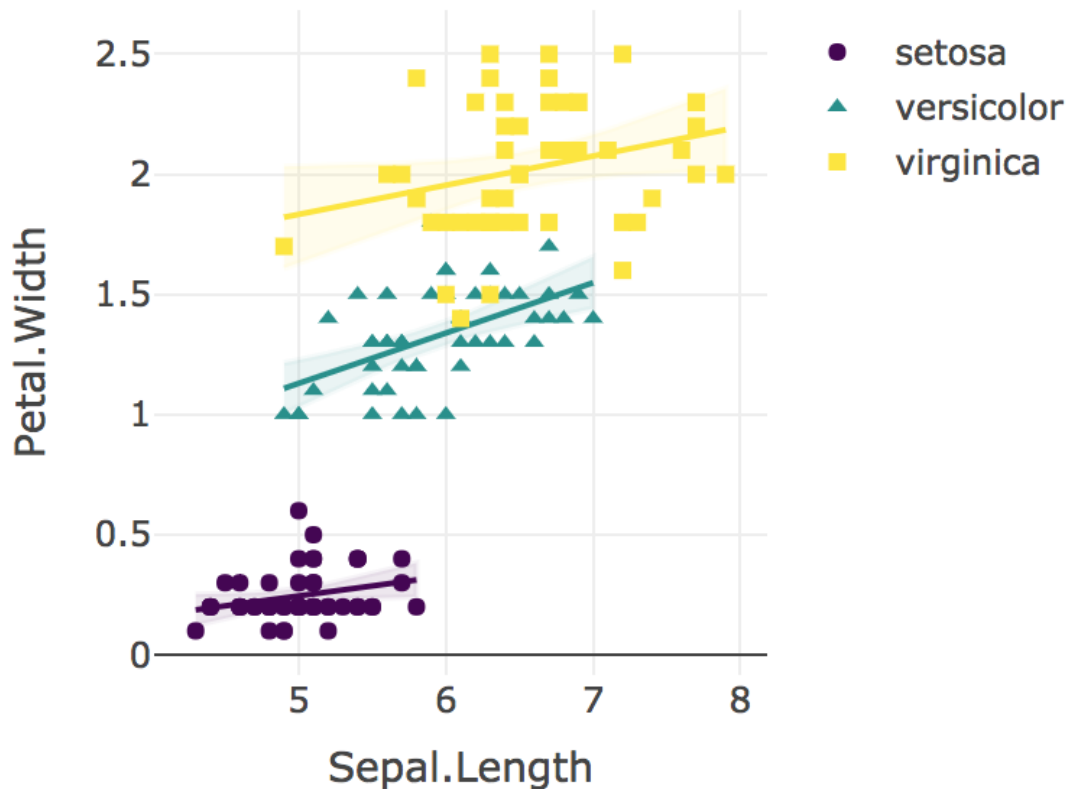
```
                      marker = list(color=paste('rgb','(', paste(colors[, i], collapse =
", "))))
    }
  p <- layout(p, xaxis = list(title = x), yaxis = list(title = y))
  return(p)
}

plotly_interaction(iris, "Sepal.Length", "Petal.Width", "Species")
```


edited Sep 7 '17 at 14:54   answered Aug 24 '17 at 21:09

                                                   Kayle Sawyer
                                                   446 ☐ 4 🟧 9

8

I agree with the previous suggestion. You should fit a multiple regression model with a dummy variable for each data set. This will allow you to test whether the intercepts differ. If you also want to know if the *slopes* differ, then you need to also include interactions between the dummies and the variable in question. There is no problem with the fact that the data are independent. Note that if they are *both* independent *and* (for example) different species, then you would not be able to tell whether the difference you find is due to the differing species or the differing data sets, as they are perfectly confounded. However, there is *no test* / get-out-of-jail-free card that will get you around that problem without gathering a new sample and running your study again.

answered Jul 25 '12 at 15:07

                                                   gung - Reinstate
                                                   Monica ♦

It looks like we've posted fairly similar answers at almost the exact same time. +1 – Macro Jul 25 '12 at 15:29

@Macro, yes, but yours is mostly better (+1 earlier); you addressed all 3 questions which I missed on my first (not thorough) reading of the question. My contribution here is the part about confounding. – gung - Reinstate Monica ♦ Jul 25 '12 at 15:31

yes that is a good point. I suppose if you were making this inquiry at all you'd have to be operating under the assumption that the data sets were measuring the same thing, etc... with the only difference being that the species were different. – Macro Jul 25 '12 at 15:44

3   From my way of thinking you both should get upvotes which is what I am doing. – Michael R. Chernick Jul 25 '12 at 15:58 ✎

1   The dummy variable suggestion is a good one *provided the error variance does not differ appreciably among models.* Otherwise you could apply a Satterthwaite-Welch t-test (which has the singular advantage of being available when only summary statistics are known, as is often the case when reading published papers) or use weighted least squares to fit the combined model. – whuber ♦ Oct 23 '16 at 16:00

This link provides a simple answer to the question:

https://stat.ethz.ch/pipermail/r-sig-teaching/2011q4/000387.html

0

answered Dec 6 '19 at 23:49

uoscar
1

1   Welcome to our site, uoscar. Could you turn this into an answer by adding an explanation? If not we can convert it to a comment to the question. Please visit our help center for more information. – whuber ♦ Dec 7 '19 at 0:03

@whuber The OP doesn't have enough reputation to make a comment. I think an answer explaining the link is a good suggestion. – Michael R. Chernick Dec 7 '19 at 0:55

@Michael "OP" refers to "original poster," but I think you must be referring to the respondent here. The OP can always comment on their own question. – whuber ♦ Dec 7 '19 at 18:23

Yes Bill Huber I know what OP stands for. You are right that I meant @uoscar who was providing an "answer" & is not the OP. Do we agree that the respondent with only 1 point representation may not make comments? – Michael R. Chernick Dec 7 '19 at 18:44