How can I perform seemingly unrelated regression in R?

Authored by
stats writer

June 30, 2024

ARABPSYCHOLOGY.COM

RECOMMENDED CITATION

stats writer (2024). *How can I perform seemingly unrelated regression in R?*. PSYCHOLOGICAL SCALES. Retrieved from https://scales.arabpsychology.com/?p=161839

Seemingly unrelated regression (SUR) is a statistical technique used to analyze multiple dependent variables that may appear unrelated, but are related through a common set of independent variables. In R, this can be achieved by using the "sur" package, which allows for the estimation of the correlations among the dependent variables and the simultaneous estimation of the regression coefficients. This approach is particularly useful in situations where the dependent variables are correlated and traditional regression methods may lead to biased results. By incorporating the correlations between the dependent variables, SUR in R can provide more accurate and reliable estimates. With the "sur" package, users can easily specify the model and obtain the necessary output for further analysis and interpretation. This makes R a powerful tool for performing seemingly unrelated regression and gaining deeper insights into complex data sets.

How can I perform seemingly unrelated regression in R? | R FAQ

A single model may contain a number of linear equations. In such a model, it is often unrealistic to expect that the equation errors would be uncorrelated. A set of equations that has contemporaneous cross-equation error correlation (i.e. the error terms in the regression equations are corrlated) is called a seemingly unrelated regression (SUR) system. At first look, the equations seem unrelated, but the equations are related through the correlation in the errors.
The systemfit R package allows a user to specify multiple equations and fit them
in an SUR. After doing so, one can perform tests on coefficients across the
equations.

ARABPSYCHOLOGY.COM

We will illustrate SUR using the hsb2 dataset, predicting read and math with the overlapping sets of coefficients and then comparing some coefficients across the two equations. We will first define our model equations as R formulas.

First let's install the systemfit R package which requires installation of three other packages: nnet, mgcv, and quantreg.

```
install.packages("nnet")
install.packages("mgcv")
install.packages("quantreg")
install.packages("systemfit")
install.packages("foreign")
install.packages("car")
install.packages("Rcpp")
library(foreign)
library(systemfit)

hsb2 <-
read.dta("https://stats.idre.ucla.edu/stat/stata/notes/hsb2.dta")
```

ARABPSYCHOLOGY.COM

```
r1 <- read~female + as.numeric(ses) + socst
r2 <- math~female + as.numeric(ses) + science
```

Once the equations have been defined, they can be passed in a list to the
systemfit command. A summary of the systemfit first shows a summary
of the system (where N = 400), then the separate equations, and then how the
residuals of the two equations are related. These are followed by the OLS fits
of the separate equations.

```
fitsur <- systemfit(list(readreg = r1, mathreg = r2), data=hsb2)
summary(fitsur)
```

systemfit results
method: OLS

N DF SSR detRCov OLS-R2 McElroy-R2
system 400 392 22835.2 3227.86 0.405103 0.342707

N DF SSR MSE RMSE R2 Adj R2
readreg 200 196 12550.9 64.0351 8.0022 0.400037

ARABPSYCHOLOGY.COM

0.390854
mathreg 200 196 10284.4 52.4712 7.2437 0.411171 0.402158

The covariance matrix of the residuals
readreg mathreg
readreg 64.0351 11.4952
mathreg 11.4952 52.4712

The correlations of the residuals
readreg mathreg
readreg 1.00000 0.19831
mathreg 0.19831 1.00000

OLS estimates for 'readreg' (equation 1)
Model Formula: read ~ female + as.numeric(ses) + socst

Estimate Std. Error t value Pr(>|t|)
(Intercept) 20.6824980 2.9789550 6.94287 5.5019e-11 ***
femalefemale -1.5111280 1.1510793 -1.31279 0.19079
as.numeric(ses) 1.2183658 0.8399004 1.45061 0.14849
socst 0.5699327 0.0562967 10.12373 < 2.22e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 8.002195 on 196 degrees of

ARABPSYCHOLOGY.COM

freedom
Number of observations: 200 Degrees of Freedom: 196
SSR: 12550.883066 MSE: 64.035118 Root MSE: 8.002195
Multiple R-Squared: 0.400037 Adjusted R-Squared: 0.390854

OLS estimates for 'mathreg' (equation 2)
Model Formula: math ~ female + as.numeric(ses) + science

Estimate Std. Error t value Pr(>|t|)
(Intercept) 19.305181 2.998047 6.43925 9.0557e-10 *
femalefemale 1.160903 1.041641 1.11449 0.266432
as.numeric(ses) 1.399639 0.742390 1.88531 0.060867 .
science 0.575330 0.054328 10.58993 < 2.22e-16 *****

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Residual standard error: 7.243704 on 196 degrees of freedom
Number of observations: 200 Degrees of Freedom: 196
SSR: 10284.364144 MSE: 52.471246 Root MSE: 7.243704
Multiple R-Squared: 0.411171 Adjusted R-Squared: 0.402158

ARABPSYCHOLOGY.COM

We may be interested in comparing the effect of female on read,
controlling for ses and socst, to the effect of female on math, controlling for ses and science. For this, we will use the linear.hypothesis command from the car package. To do
this, we create a "restriction" on the model system. We will force the
coefficient of female to be the same in both equations and then compare
such a system fit to the one seen when the coefficients are not equal.

```
library(car)
restriction <- "readreg_femalefemale- mathreg_femalefemale"
linearHypothesis(fitsur, restriction, test = "Chisq")
```

Linear hypothesis test (Chi^2 statistic of a Wald test)

Hypothesis:
readreg_femalefemale - mathreg_femalefemale = 0

Model 1: restricted model
Model 2: fitsur

ARABPSYCHOLOGY.COM

Res.Df Df Chisq Pr(>Chisq)

1 393

2 392 1 2.9626 0.08521 .

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

ARABPSYCHOLOGY.COM