

Evaluate a set of hypotheses with GORIC or GORICA: Linear Regression Example

Rebecca M. Kuiper

07 September 2023

Contents

Example: Sesame	1
Data preparation	1
Example hypothesis	2
Standardize data	2
Preparation for GORIC(A)	2
GORIC & GORICA	3
GORIC values and weights for H1 and its complement	3
GORICA values and weights for H1 and its complement	3

Example: Sesame

Data preparation

First, load the required libraries (after they have been installed). These libraries contain functions, such as `goric`, that will be used in the R code below. Each time you reopen R, you will have to load the required libraries.

```
## First, install the packages, if you have not done this already:
if (!require("restriktor")) install.packages("restriktor")

## Then, load the packages:
library(restriktor) # for the goric function

# If you want to use restriktor from github:
# if (!require("devtools")) install.packages("devtools")
# library(devtools)
# install_github("LeonardV/restriktor")
# library(restriktor) # for goric function
```

Second, it is necessary to load the data.

Notably, it is only possible to load the data if you are using the correct working directory (with both your R script and data file). The command `getwd()` shows you your current working directory. You can change the working directory to the one you prefer using the function `setwd()` by specifying the correct location between parentheses. Alternatively, in Rstudio, you can use the “Session” tab (on top) or you can use the “Files”-pane (on top of probably the right lower box of your Rstudio-screen, this pane is located next to the panes for “Plots”, “Packages”, “Help” and “Viewer”).

If you open the data file `Data_Lucas.txt` in a text editor, you can see that the variable labels have been inserted (using quotes; i.e., "...") in the first line of the file, which is called a header. Therefore, you have to specify 'header = TRUE' when loading the data:

```
# Load the sesame data from a text file
Sesame <- read.table("data/sesamesimRegression.txt", header=TRUE)

# Inspect data
head(Sesame)
```

	prelet	prenumb	postnumb
1	30	35	42
2	24	29	50
3	26	32	48
4	26	42	39
5	7	19	34
6	19	30	21

Example hypothesis

The following hypothesis will be evaluated:

```
H1 <- 'prenumb > prelet'
# Note that the labeling of the estimates should be used
```

Standardize data

Since parameters of continuous variables (using different scales) are being compared, the data needs to be standardized (such that comparison of the parameters is meaningful / fair, since they are on the same, comparable scale then).

```
# Standardize all continuous variables (of interest) including the outcome:
Sesame_s <- as.data.frame(scale(Sesame))
#
# Or: Standardize solely the continuous predictors (of interest):
#Sesame_s <- Sesame
#Sesame_s[,2:3] <- as.data.frame(scale(Sesame[,2:3]))
```

Notes:

- Standardizing of ordinal and nominal variables is not helpful, since the estimate of the unstandardized ordinal and nominal variables reflects a group difference in the outcome (in that case, one probably does not want to standardize the outcome).
- In case there are multiple outcomes and estimates for multiple outcomes are compared, then one needs to standardize the outcome.

Preparation for GORIC(A)

```
# estimate the parameters of the statistical model at hand
fit_Sesame <- lm(postnumb ~ prenumb + prelet, data = Sesame_s)

# Check names used in model
names(coef(fit_Sesame))
```

```
[1] "(Intercept)" "prenumb"      "prelet"
```

```
# Specify restrictions should use these names (as done above)
#
# In case of the Intercept, one should use: .Intercept.
# E.g., when outcome is not also standardized:
# H1 <- '.Intercept. > 0'
```

GORIC & GORICA

GORIC values and weights for H1 and its complement

Here, we assume that the interest lies only in H1. In that case, it should be evaluated against its complement.

```
set.seed(123)
goric(fit_Sesame, hypotheses = list(H1 = H1), comparison = "complement")
```

restriktor (0.5-20): generalized order-restricted information criterion:

Results:

	model	loglik	penalty	goric	loglik.weights	penalty.weights	goric.weights
1	H1	-256.595	3.500	520.190	1.000	0.500	1.000
2	complement	-268.116	3.500	543.232	0.000	0.500	0.000

The order-restricted hypothesis 'H1' has 100793.255 times more support than its complement.

From the output, it is concluded that the support for H1 is many more (100793.255 times more) supported / likely than its complement. Hence, there is compelling evidence in favor of H1.

GORICA values and weights for H1 and its complement

```
set.seed(123)
goric(fit_Sesame, hypotheses = list(H1 = H1), type = "gorica", comparison = "complement")
```

restriktor (0.5-20): generalized order-restricted information criterion approximation:

Results:

	model	loglik	penalty	gorica	loglik.weights	penalty.weights	gorica.weights
1	H1	6.082	2.500	-7.164	1.000	0.500	1.000
2	complement	-5.859	2.500	16.718	0.000	0.500	0.000

The order-restricted hypothesis 'H1' has 153406.682 times more support than its complement.

From the output, it is concluded that the support for H1 is many more supported / likely than its complement. Hence, there is compelling evidence in favor of H1.