**R Exercises**

*Learning to use R for the evaluation of informative hypotheses with bain and restriktor*

**Intro to R (if needed)**

See the file ‘R hands-on mini course.pdf’ or search online for an intro course for free.

**Exercise 1A. Bayesian hypothesis evaluation**

1. Open R and Rstudio
2. Check out some of the files in the ‘R - bain’ subfolder.
3. Install (if needed) and load the bain package.
4. Read the description at the end of this document to obtain an impression regarding what is contained in sesamesim.txt.
5. Come up with one or more informative hypotheses, after considering the meaning of postnumb and site (in this case, don’t hesitate to be inspired by the descriptives which can be obtain in Step 6; but in practice, one should never do this of course).
6. Run an ANOVA:

Use site as the factor and postnumb as the dependent variable.

Use sesamesim.txt as your data file (which is contained in the ‘JASP meeting’ folder, that is, the same directory as this exercise).

1. Specify informative hypotheses.
2. Evaluate the informative hypotheses by running the bain function.
3. Interpret the analysis results that are presented in the right hand screen.

**Exercise 1B. GORIC hypothesis evaluation**

Go through the following steps to evaluate informative hypotheses in an ANOVA model using the GORIC.

1. Open R and Rstudio
2. Check out the R file with example code in the subfolder ‘R - GORIC slides’.
3. Install (if needed) and load the restriktor package.
4. If not done in Exercise 1A: Read the description at the end of this document to obtain an impression regarding what is contained in sesamesim.txt (this file is available in the ‘R meeting’ folder).
5. If not done in Exercise 1A: Come up with one or more informative hypotheses, after considering the meaning of postnumb and site (in this case, don’t hesitate to be inspired by the descriptives which can be obtain in Step 6; but in practice, one should never do this of course).
6. Run an ANOVA:

Use site as the factor and postnumb as the dependent variable.

Use sesamesim.txt as your data file (which is contained in the ‘JASP meeting’ folder, that is, the same directory as this exercise).

1. Specify informative hypotheses.
2. Decide on the choice of failsafe.

In case of one hypothesis of interest, one can use its complement as failsafe (set as default when specifying only one hypothesis). The complement of a hypothesis consists of all hypotheses / theories except the one of interest.

In case you have multiple informative, competing hypotheses that do not cover all possible theories / hypotheses, you need a failsafe hypothesis. Currently, the complement of a set is not available; thus, you need the unconstrained hypothesis as failsafe (which is included by default). The unconstrained does not restrict any parameters and, therefore, represents all possible hypotheses/theories, including the ones in the set. Bear in mind that the unconstrained will always obtain support, so it is best to not see it as a competing hypothesis (only a failsafe).

1. Evaluate the informative hypotheses by running the goric function.
2. Interpret the analysis results that are presented in the right hand screen.

*Optional*:

It is probably better to take into account the pre-test knowledge as well. Try of you can do the above for

* either an ANCOVA model correcting for the variable prenumb
* or an ANOVA with the difference score postnumb-prenumb as outcome

**Exercise 2A. Linear regression and Bayesian updating**

1. Run a regression on the data file sesamesim30.txt.

By doing this you pretend that you are a researcher that has so far collected only a sample containing 30 cases.

1. Use the null-hypotheses that all regression coefficients are equal to zero and an informative hypotheses that all regression coefficients are larger than 0.
2. Look at the BFs and the PMPs. Do you want to stop, or do you want the evidence to be stronger and continue to update? Write down the most important BFs and PMPs.
3. Repeat the above steps for the data sets containing 60-90-and-120 cases.

By doing this you pretend that you added 30 persons to your sample, recomputed the BFs and PMPs, and repeated that two more times.

1. You have executed a Bayesian updating. What is your conclusion, how many children do you need to obtain convincing conclusions?

**Exercise 2B. GORIC updating**

1. Run a regression or ANOVA on the data file sesamesim30.txt.

By doing this you pretend that you are a researcher that has so far collected only a sample containing 30 cases.

1. Specify hypotheses and run goric

Note: you could follow the steps from Exercise 1B.

1. Look at the BFs and the PMPs. Do you want to stop, or do you want the evidence to be stronger and continue to update? Write down the most important BFs and PMPs.
2. One can now update these results by doing the same for the data sets where more observations are collected. That is, re-run the same analysis for the data sets containing 60, 90, and 120 cases (files are available in the ‘R meeting’ folder).

By doing this, you pretend that you added 30 persons to your sample, recomputed the GORIC (weights), and repeated that two more times.

1. You have now executed a GORIC updating. What is your conclusion? How many children did you need to obtain convincing conclusions?

**Exercise 3. NHST vs evaluating informative hypotheses**

Use classical statistics (NHST) to answer the same research questions you addressed in Exercises 1 and 2. There is some example code in the R file in the subfolder ‘R - GORIC slides’.

What can and can’t you achieve using classical statistics? What do you think are the advantages and disadvantages of classical and Bayesian statistics?

**Exercise 4. The sky is the limit**

You can try out evaluating informative hypotheses using other data sets or your own data.

**Optional: Exercise 5. GORIC evidence synthesis (via Shiny app)**

Download (one of the two):

<https://github.com/rebeccakuiper/Tutorials/blob/main/Tutorial_GORIC_restriktor_AggrSupport.html>

<https://github.com/rebeccakuiper/Tutorials/blob/main/Hands-on%20files/Hands-on_5_GORICA_CombEv_restriktor.R>

Try one example to aggregate evidence from multiple studies or create your own one.

What is the aggregated support for your hypothesis/-es of interest?

**Description of the data**

**Description Sesamesim, a Simulated Data Set Inspired by the Sesame Data**

The exercises use a simulated data set inspired by the Sesame Street data set from: Stevens, J. P. (1996). Applied Multivariate Statistics for the Social Sciences. Mahwah NJ: Lawrence Erlbaum. This data set is included in the bain package and in the older containing these exercises. The data are obtained from children before and after watching Sesame street (an educative program for children) for a year.

The variables contained in sesamesim are subsequently:

* sex (1 = boy, 2 = girl) of the child
* site (1 = disadvantaged inner city, 2 = advantaged suburban , 3 = advantaged rural, 4 = disadvantaged rural, 5 = disadvantaged Spanish speaking) from which the child originates
* setting (1 = at home, 2 = at school) in which the child watches sesame street
* age (in months) of the child
* viewenc (0 = no, 1 = yes), whether or not the child is encouraged to watch Sesame Street
* peabody (mental age) score of the child (higher score is higher mental age)
* prenumb (score on a numbers test before watching Sesame Street for a year)
* postnumb (score on a numbers test after watching Sesame Street for a year)
* funumb (follow up numbers test score measured one year after postnumb)
* Bb Knowledge of body parts before
* Bl Knowledge of letters before
* Bf Knowledge of forms before
* Bn Knowledge of numbers before
* Br Knowledge of relations before
* Bc Knowledge of classifications before
* Ab Knowledge of body parts after
* Al Knowledge of letters after
* Af Knowledge of forms after
* An Knowledge of numbers after
* Ar Knowledge of relations after
* Ac Knowledge of classifications after