"Optional" Assignment on Hierarchical Bayesian Analysis with BUGS or JAGS

Points for this assignment: 2 (+ 1 for presenting the results - on a voluntary basis - on November 29, 2019)
Due on November 28, 2019
To be sent to giancarlo.manzi@unimi.it
For presentation use ppt or pdf file

This assignment is not compulsory

Consider the dataset downloadable at the link on Ariel from which you downloaded this assignment, having the following columns:

Column name	**Description**
`Station_ID`	Bike-sharing docking station (ds) ID
`Name`	Name of the ds
`Railway`	No. of railway stations within 1km from the ds
`Streetcar`	No. of streetcar lines passing within 1km from the ds
`Bus`	No. of bus lines passing within 1km from the ds
`Underground`	No. of underground lines passing within 1km from the ds
`7-8, 8-9, etc.`	No. of bike checking out within time slots

1. Present some explorative statistics based on this table with plots and graphs. Use `R` or `Python` for doing this.
2. Fill the following table with the number of bike check-outs according to the time slot and the number of underground stations within 1km radius around the docking station:

	No. of underground lines		
Time slots	0	1	2
7-10	....	....	....
10-13	....	....	....
13-16	....	....	....
16-19	....	....	....

   Use `R` or `Python` for doing this.
3. Implement a frequentist (i.e. non-Bayesian) Poisson regression model having the number x_i of underground lines as independent variable, with the following distributional assumption:

$$y_{ij} \sim Pois(\mu_i),$$

where y_{ij} is the dependent variable representing the number of check-outs for docking stations having i underground lines within 1km, observed within time slot j. Use `BUGS` or `JAGS` for doing this. Compare the results from this analysis with those obtained in `R` or `Python` using `glm()` or the library `pymc3`, respectively. Limit your analysis to the intercept and the regression coefficient only.

4. Implement a hierarchical Bayesian Poisson regression model having the number x_i of underground lines as independent variable, with the following distributional assumption:

$$y_{ij} \sim Pois(\mu_i),$$

where y_{ij} is the dependent variable representing the number of check-outs for docking stations having i underground lines within 1km, observed within time slot j. Use `BUGS` or `JAGS` for doing this. Limit your analysis to the intercept and the regression coefficient only.

(*Hint.* For constructing the hierarchical structure in the function expressing the link function you should *additively* introduce the parameter λ_{ij} allowing for variability across observations and exchangeability. λ_{ij} should be such that $\lambda_{ij} \sim N(0, \omega_\lambda)$ and you should choose the right hyperprior for ω_λ).

5. Present and comment the results obtained in 1., 2., 3. and 4.

Notes on the Poisson regression.
Suppose that the data y_i, $i = 1, \ldots, n$ come from a specific distribution in the exponential family (Normal, Bernoulli, Binomial, Multinomial, Exponential, Poisson, Dirichelet, etc.), and we have the following relationships:

$$E[y_i] = \mu_i = g^{-1}(\eta_i),$$

$$\eta_i = \beta_0 + \sum_{k=1}^{p} \beta_k x_{ki},$$

for independent variables x_{ki}, $k = 1, \ldots, p$. g is called the *link function* and could be the *identity* function, the *log*, the *logit*, etc. For data from the Poisson distribution, $y_{ij} \sim Pois(\mu_i)$ and the link function is the natural *log* function, so that we have:

$$log(\mu_i) = \beta_0 + \beta_1 x.$$

The Poisson regression model is fit for situations where counts are to be modeled according to some covariates.

Notes on the structure of the data list from table created in point 2. to be passed to `BUGS` or `JAGS`.
The list to be used in `BUGS` or `JAGS` can have the following format:

```
Example of a list() structure in BUGS or JAGS
list(y = structure(.Data = c(..,..,..,etc.),
.Dim = c(., .)),
x = c(0, 1, 2))
```

where inside `.Data` the values in the table should be entered and `.Dim` indicates the dimension of the data set.