University of Cape Town
Department of Statistical Sciences
Masters in Data Science and STA Honours
Statistical Computing: Assignment 1

February 27, 2017

Globular Clusters and Galaxies

You have already encountered the globular clusters data. These data record the properties of 422 galaxies, and the number of globular clusters. The data are availabe at: http://www.physics.mcmaster.ca/~harris/GCS_table.txt.

The variables are described in this text file.

Assignment

Use these data to investigate the relationship between the number of globular clusters and the following galaxy properties: central black hole mass, dynamical bulge mass, bulge velocity dispersion, and absolute visual magnitude and type of galaxy.

Your assignment should include some exploratory data analysis (numerical and visual), a summary of some models fitted, all formulae in LaTeX, graphical illustration of your predictions, and a paragraph on what you can learn about the relationship between the number of clusters and galaxy properties from this data set.

The emphasis should be on a good report format, good code (check the R Style Guide), good informative graphics, good interpretation of what you have done, not so much on finding the best statistical model. However, you should be able to recognize the limitations of your model.

You should include at least one model with a categorical variable (type of galaxy) and give a clear interpretation of the regression coefficients.

? Any mentioning of 'accept' or 'reject' or '5% significance level' (or any other significance level) will be penalized.

All figures and tables should have a caption (tables above, figures below).

Include a reference list (which must include a reference to the R software, and the data source at the least).

You can discuss the problem amongst each other. However, the final code and report must be entirely your own.

Handin date: Sunday 5 March 2017, 10pm, on Vula under the Assignments tab. Please name your file according to your student number: e.g. `abnlix045.Rmd`.

Format: R Markdown file (.Rmd). No other format will be accepted. I need to be able to run this file from my computer. You have two options: either to download directly from the website, or to read in the .txt file, assuming it is in the current working directory. Additional required packages are fine.

References

- Catalog of Globular Cluster Systems in Galaxies: http://www.physics.mcmaster.ca/~harris/GCS_table.txt
- Harris, W.E., Harris, G.L. and Alessi, M., 2013. A Catalog of Globular Cluster Systems: What Determines the Size of a Galaxy's Globular Cluster Population? The Astrophysical Journal, 772(2), p.82.
- De Souza, R.S., Hilbe, J.M., Buelens, B., Riggs, J.D., Cameron, E., Ishida, E.E.D.O., Chies-Santos, A.L. and Killedar, M., 2015. The overlooked potential of generalized linear models in astronomyIII. Bayesian negative binomial regression and globular cluster populations. Monthly Notices of the Royal Astronomical Society, 453(2), pp.1928-1940.