Supplemental Computer Science and Statistical Handout

## Introduction

We can make compelling statements about the behavior of multi-agent simulations using traditional statistical tools (Lee et al. 2015). While making simplifications of the complexity of multi agent simulations, linear analysis can provide us a glimpse into their behavior and trends. This handout focuses on showing you how to draw inferences from this data using multiple linear regression. While making simplifications of the complexity of multi agent simulations, linear analysis can provide us a glimpse into their behavior and trends. Understanding the influence of the parameters (and how those influences were quantified) is useful to understand how adjusting parameters changes the simulation. Changing multiple parameters at the same time will lead to more complex interactions, but this section provides a general guideline for the impacts of parameters. Additionally, the nonlinear nature of complex adaptive systems means that this analysis provides us only a glimpse into the behavior of the simulation.

We can linearly model using the following equation:

$population =β\_{0}+β\_{1}⋅parameter value +β\_{2}⋅ max tick$ (1)

where, $β\_{0}$ is the y-intercept, $β\_{1} $is the coefficient of the parameter value, and $β\_{2}$ is the coefficient of the maximum number of ticks. Coefficients can be interpreted as the change in $population $resulting from a one-unit change in their corresponding variable.

A *p*-value is a statistical measure of the probability of achieving your results assuming that a relationship does not exist between the two variables (LaMorte 2021). A *p*-value is the level of confidence that there is a relationship between the two variables you’re comparing. In our case, we are comparing changes in the parameters with changes in simulation outcome. Scientists traditionally look for *p*-values less than 0.05 to establish a relationship between two variables. The smaller the *p*-value, the more confidence that two variables are related. *p-*values greater than 0.05 mean that we cannot say if two variables are related. In this simulation, we will use 0.05 as the p-value threshold for significance.

To run experiments on any individual variable, the simulation multiple times, varying the variable you would like to study each time you run it. Record the results after each experiment execution in Excel so that you can run statistics on it. To run linear regressions, use the LINEST function for Excel.

References

Lee, J. S., T. Filatova, A. Ligmann-Zielinska, B. Hassani-Mahmooei, F. Stonedahl, I Lorscheid, A. Voinov, J. G. Polhill, Z. Sun, and D. C. Parker. 2015. The complexities of agent-based modeling output analysis. Journal of Artificial Societies and Social Simulation 18. <https://doi.org/10.18564/jasss.2897>