

# Stat 215B (Spring 2005): Lab 3

GSI: Victor Panaretos  
victor@stat.berkeley.edu

Due March 8 at the Lab Section

## **Part 1: Simultaneous Inference**

In this section we revisit the data from the previous lab. Recall that the scenario involved the study of linear relationships between Bond Sheer Strength (S) and four independent variables of interest: Force (F), Ultrasonic Power (P), Temperature (T) and Time (D). In particular, we had considered the model

$$S_i = \beta_0 + \beta_1 F_i + \beta_2 P_i + \beta_3 T_i + \beta_4 D_i + \epsilon_i, \quad (1)$$

and carried out certain hypothesis tests regarding the coefficient vector, say  $\tilde{\beta}$ . The aim of this part of the lab is to construct simultaneous confidence regions for certain linear combinations of the components of  $\tilde{\beta}$ . In particular you are asked to:

- Give a 95% confidence interval for  $2\beta_1 + \beta_2 + 4\beta_3$ . What distributional assumptions are implicit in the procedure you use?
- Find an appropriate 95% simultaneous confidence region for  $\beta_4$  and  $2\beta_1 + \beta_3$ . Make sure to clearly explain how you proceed, and justify your choice of method.

## **Part 2: ANOVA/Indicator Variable Regression**

This section is dedicated to modelling the foraging strategies of ants. Foraging is a dangerous activity for the ants, as for most species. Each forager is supplied with some amount of food before leaving for his search. If the forager is lost, there are two costs for the colony: the loss of a worker, and the loss of the food he was carrying. In this context, there seem to exist two strategies:

- “Worker-conservative” where ants foraging at greater distances consume relatively more food in order to minimise the risk of starvation (effort to conserve workers).
- “Energy Conservative” where long distance foragers, who are more likely to die, will consume less food so that their deaths will not be as much of a strain on the colony’s food supply (effort to conserve food/energy).

In this report, we study the *Pogonomyrmex Salinus* ants, and wish to shed light on their foraging strategies using linear models. The data for this analysis are drawn from a study by Dr. Peter

Nonacs of UCLA [1], conducted at the Sierra Nevada Aquatic Research Laboratory (SNARL), in the Great Basin Desert Province<sup>1</sup>. In particular, we have measurements on 363 ants, picked at random from the vicinities of 5 different colonies. For each ant, the following characteristics were recorded:

- Which colony the ant belongs to (the colonies are labelled A-E). The corresponding variable is `colony`.
- At what distance from the mound's entrance was the ant captured (a variable with 3 categories: close to 0, 5 and 10 meters). The corresponding variable is `distance`.
- The mass of the ant, taken to be indicative of the food/energy it is carrying (measured in milligrams). The corresponding variable is `mass`.
- The size of the ant. The corresponding variable is `class`, and the classification is with respect to the headwidth of the ant.

Begin by conducting an appropriate exploratory graphical analysis using boxplots and coplots. You are then asked to regress `mass` on `colony`, `distance` and `class`. Comment on the duality between analysis of variance and indicator variable regression<sup>2</sup>. Give an interpretation of the resulting coefficients. Can this model be of use in suggesting answers to the following questions:

- Does the average mass of an ant differ according to colonies (i.e. are some colonies more successful?).
- Is there an overall foraging strategy (worker-conservative or energy-conservative), or does the strategy vary from colony to colony? (hint: consider including an appropriate interaction in the model - explore this possibility using a coplot).
- Depending on the answer to the previous question, which is the strategy(ies)?

Remember to always clearly state your modelling hypotheses, and check them with the appropriate plots.

**Note:** The data are contained in the file `lab3.txt` on the lab section webpage.

## References

[1] P. Nonacs (2003). Personal Webpage:

<http://research.mednet.ucla.edu/cfm/lifesci/obeeindivfaculty.cfm?FacultyKey=1130>

---

<sup>1</sup>The full data set may be accessed through the data library of the Department of Statistics, UCLA: [www.stat.ucla.edu](http://www.stat.ucla.edu)

<sup>2</sup>also known as *dummy* variable regression