

# LING82100: homework 6

(Due 4/29)

In this assignment you will model data from Labov's 1966 department store study.<sup>1</sup> This is part 1; we will continue to model and interpret this data the following week.

## 1 Logistic regression

Fit a logistic regression model in which the use or non-use of (r) is the dependent variable and the independent variables are word (*fourth* or *floor*), emphasis ("normal" or "emphatic"), and department store (S. Klein's, Macy's, or Saks 5th Ave.). Use sum-coding for all variables.

### What to turn in

In the form of a table, provide

- all model coefficients,
- their standard errors,
- the  $\chi^2$  likelihood ratio test statistics for the independent variables, and
- the associated  $p$ -values.

### Hints

- An sample table is shown in Table 1; simply fill in the elided values.
- For a categorical variable with  $n$  levels R computes and prints  $n - 1$  coefficient; However, it is straightforward to estimate the  $n$ th coefficient. Under sum coding, the missing coefficient is simply  $-1$  times the sum of the other coefficients. Or, equivalently, the  $n$  coefficients for a given categorical variable must sum to 0.
- However, there's no standard error for the  $n$ th coefficient.
- You can use the `lrtest` function from the `lmtest` package (install it if you don't already have it) to perform the likelihood ratio test; it takes two models in a nesting relationship and computes the necessary numbers.

---

<sup>1</sup><http://wellformedness.com/courses/LING82100/Data/NYC.csv>

	Coef.	S.E.	$\chi^2$	$p(\chi^2)$
(Intercept)	...	...		
Store:			...	...
S. Klein's	...	...		
Macy's	...	...		
Saks 5th Ave.	...			
Word:			...	...
<i>floor</i>	...	...		
<i>fourth</i>	...			
Emphasis:			...	...
emphatic	...	...		
normal	...			

Table 1: Sample results table for §1.

## Stretch goals

- Also add an interaction of target word and emphasis and interpret the results. Note that `drop1` does not fully understand interaction terms, so you will have to use `lrtest` or compute the likelihood ratio test manually.
- The data is currently in “long” format, meaning that there is a single observation or trial per row. R also permits you to use data in “wide” format, in which the dependent variable is represented as a table of successes and failures for each possible value of the independent variables.<sup>2</sup> Map the data onto “wide” format and show that you obtain the same results.

## References

Labov, William. 1966. *The Social Stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.

<sup>2</sup>For an example, see <http://stackoverflow.com/questions/9111628/>.