STAT 426 — Sections 1GR, 1UG — Spring 2017

Data Report

DUE: April 5

You will submit a data analysis report based on the scenario described below. Your report must be accurate, complete, clear, concise, clearly structured, well-written, grammatical, and professional. It must be on paper, typewritten, with at least 1 inch margins, and limited to 5 pages, or 7 pages for the graduate section. Please note that some parts should be submitted by the graduate section only.

Scenario: The data file `cyclone.txt` contains seasonal counts of named tropical cyclones for 1995–2016, for the Eastern Pacific and North Atlantic basins. The counts are categorized according to *maximum* status attained: major hurricane, hurricane, or tropical storm. The file also contains the August-centered Oceanic Niño Index (ONI)[1], a measure of the El Niño-Southern Oscillation (ENSO). (August is typically the middle of the hurricane season.)

Use R software, and use only the data in `cyclone.txt`. All R code and any direct R output you choose to list should be in the Appendix only.

Your report must follow this outline:

1. **Introduction** Provide some background information about tropical cyclones, basins, cyclone categories, seasons, ENSO, and ONI. (Use footnotes to acknowledge any sources you consult, including web sites.) Formulate at least two research questions your analyses will answer.

2. **Data** Describe and summarize the variables. Include a table of marginal totals (summed over all seasons) of each category for each basin, and comment. Also, for each basin separately, produce a plot of total number of storms versus August ONI.

3. **Analysis of Total Named Storms** Fit a loglinear regression model for the total number of named storms, with a term for season (linear in year — *not* categorical), basin (indicator variable), August ONI, and interaction between basin and August ONI. Give the form of the equation for the linear predictor, explaining each symbol. Provide maximum likelihood estimates, standard errors, and (when appropriate) P-values for presence of terms. Check goodness of fit. Use the model to estimate the multiplicative effect of a one-unit increase in August ONI on the *mean number of named storms*, separately for each basin. Compare.

4. **Analysis of Proportion of Major Hurricanes** Consider logistic regression for the probability that a named tropical cyclone becomes a *major hurricane*. (Note that, for this

[1] `http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml`

analysis, the data are grouped.) Starting with the model that has full three-way interaction between season, basin, and August ONI, perform *backward elimination*, using R function `step`. (This will stop when AIC can't be further reduced.) Summarize the resulting fitted model. Using this model, estimate the probability that a named storm becomes a major hurricane for the current year (2017) under six different scenarios: for each basin for three different August ONI values: ONI = -1.5 (strong La Niña), ONI = 0 (neutral), and ONI = 1.5 (strong El Niño). Summarize your results in a table.

5. GRADUATE SECTION ONLY **Further Analyses** (a) For the loglinear regression: For each basin, form a (transformed) Wald 95% confidence interval for the multiplicative effect of a one-unit increase in August ONI on the mean number of named storms. Interpret. (b) For the logistic regression: Assess the fit of your final model (after backward elimination) using appropriate residuals. (Note that many expected counts are small, so the usual goodness-of-fit tests may be inaccurate.) Also assess the influence of observations. Are there any especially poorly-fitting or influential observations?

6. **Conclusions** Briefly summarize your results in a non-technical manner.

7. **Appendix** Provide the R code you used to conduct your analysis. Include comments that label the purpose of each block of code.

NOTES FOR GRADUATE SECTION: You may need to use the `vcov` function in R to obtain the estimated asymptotic covariance matrix needed to compute the standard errors for the Wald intervals.