

1 Introduction

A tropical cyclone is a rotating, organized system of clouds and thunderstorms that originates over tropical or subtropical waters and has a closed low-level circulation. Tropical cyclones rotate counterclockwise in the Northern Hemisphere. In ascending order of severity, they are classified as tropical depression, tropical storm, hurricane, major hurricane¹. Global areas of tropical cyclone formation are divided into seven basins².

Meteorologists discover that tropical cyclone's occurrence is positively correlated to El Niño-Southern Oscillation (ENSO). ENSO is a naturally occurring phenomenon that involves fluctuating ocean temperatures in the equatorial Pacific. ENSO is measured by Oceanic El Niño Index (ONI). A higher ONI indicates higher level of ENSO, which affects cyclone's frequency and severity. Since August is the middle of the cyclone season, ONI taken in August can be used as a reliable parameter to gauge the intensity of ENSO in the whole season.

2 Data

The dataset we have contains cyclone data in Eastern Pacific Basin and North Atlantic Basin from 1995 to 2016, inclusively. Each observation is a data entry on cyclones of one basin in one year (therefore, there are $2 \times 22 = 44$ observations.) Each observation reports the number of tropical storms, hurricanes, and major hurricanes. The dataset contains also includes August-ONI of the respective year. Since ONI is a global parameter, two observations that belong to the same year share the August-ONI value.

Marginal totals for each type of cyclones are shown below:

	Tropical Storm	Hurricane	Major Hurricane	Total
Eastern Pacific	142	88	82	312
North Atlantic	152	89	74	315

The table shows that there is not much difference of cyclone counts between the two basins.

The plot of total number of cyclones in Eastern Pacific basin against ONI shows a positive correlation, while the plot of total number of cyclones in Northern Atlantic basin shows a negative correlation between cyclones and ONI. Plots are attached at the end of the report.

¹National Oceanic and Atmospheric Administration, <http://www.nhc.noaa.gov/climo/>

²namely Atlantic, Northeast Pacific, Northwest Pacific, North Indian, Southwest Indian, Southeast Indian, and Australian. National Oceanic and Atmospheric Administration, <http://www.aoml.noaa.gov/phod/cyclone/data/seven.html>

3 Analysis of Total Storms

I fit a loglinear regression model for the total number of named storms, as is instructed by the exam question. The formula for the regression model is given by

$$\hat{y} = e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_2 x_3},$$

where \hat{y} denotes fitted value of total number of cyclones, x_1 denotes Season, x_2 denotes Basin, and x_3 denotes ONI. α denotes intercept, and four β s denote regression coefficient of the respective term.

Maximum likelihood estimators produced by the regression model in R are given below:

	Estimate	Std.err	P-value
α	-16.0180	12.6393	0.2050
β_1	0.0093	0.0063	0.1396
β_2	-0.0139	0.0809	0.8641
β_3	0.1695	0.0754	0.0246
β_4	-0.3694	0.1116	0.0009

Goodness-of-fit test shows that the model deviance is 35.65, with p-value being 0.6233. So we conclude the model holds.

For Eastern Pacific cyclones, since $x_{2i} = x_{2i}x_{3i} = 0$, the model becomes

$$\hat{y} = e^{\alpha + \beta_1 x_{1i} + \beta_3 x_{3i}},$$

Therefore, one-unit increase in ONI in Eastern Pacific causes average number of storms to be $e^{0.1695} = 1.1847$ times bigger.

Similarly, one-unit increase in ONI in Northern Atlantic causes average number of storms to be $e^{-0.0139 + 0.1695 - 0.3694} = 0.8075$ times smaller.

4 Analysis of Proportion of Major Hurricanes

In compliance with the data report instructions, first I fitted a model with three-way full interaction terms, and then performed backward selection from it. The final model produced is

$$\hat{y} = e^{-52.60 + 0.02569x_1 + 132.1x_2 + 0.3758x_3 + -0.06596x_1x_2 - 0.5037x_2x_3}$$

where x_1 is Season, $x_2 = 0$ for Eastern Pacific and 1 for North Atlantic, x_3 is ONI.

To predict the probability of a storm becoming a major hurricane in 2017, we use this model to produce the following results:

	Eastern Pacific	Northern Atlantic
ONI=-1.5	0.2581	0.2255
ONI= 0	0.4536	0.1861
ONI= 1.5	0.7969	0.1537

5 Further Analysis

5.1 Loglinear Model

In this part, I will find the 95% confidence interval for the coefficient of the interaction term in the loglinear model fitted in Part 3, for each basin. Recall that the model is given by:

$$\hat{y} = e^{\hat{\alpha} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_2 x_3},$$

where \hat{y} denotes fitted value of total number of cyclones, x_1 denotes Season, $x_2 = 0$ for Eastern Pacific, an 1 for Northern Atlantic, and x_3 denotes ONI. α denotes intercept, and four β s denote regression coefficient of the respective term. Since we are interested in the effect on the fitted value from the change of x_3 , i.e. ONI, I define \hat{y} as a function of x_3 , which is:

$$\hat{y}(x_3) = e^{\hat{\alpha} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_2 x_3},$$

The exam asks for the multiplicative effect of a one-unit change of x_3 , which is essentially

$$\frac{\hat{y}(x_3 + 1)}{\hat{y}(x_3)} = \frac{e^{\hat{\alpha} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 (x_3 + 1) + \hat{\beta}_4 x_2 (x_3 + 1)}}{e^{\hat{\alpha} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_2 x_3}} = e^{\hat{\beta}_3 + \hat{\beta}_4 x_2}$$

In the case of Eastern Pacific basin, since $x_2 = 0$, the multiplicative effect is caused by $\hat{\beta}_3$ only. We know $E[\hat{\beta}_3] = 0.1695$ and $sd.err[\hat{\beta}_3] = 0.07539$, thus a 95% confidence interval for $\hat{\beta}_3$ is $0.1695 \pm 1.96 \times 0.07539 = [0.02172, 0.372]$. Exponentiating the values above, we get

$$[1.022, 1.373]$$

In the case of Northern Atlantic basin, $x_2 = 1$, and the multiplicative effect is represented by $e^{\hat{\beta}_3 + \hat{\beta}_4}$. By simple mathematical statistics:

$$E[\hat{\beta}_3 + \hat{\beta}_4] = E[\hat{\beta}_3] + E[\hat{\beta}_4] = 0.1695 - 0.3694 = -0.1999, \text{ and}$$

$$Var[\hat{\beta}_3 + \hat{\beta}_4] = Var[\hat{\beta}_3] + Var[\hat{\beta}_4] + 2Cov[\hat{\beta}_3, \hat{\beta}_4] = 0.06551 + 0.005683 + 2 \times 0.0001987 = 0.07159$$

The 95% confidence for $\hat{\beta}_3 + \hat{\beta}_4$ is given by

$$-0.1999 \pm 1.96 \times \sqrt{0.07159} = [-0.7243, 0.3245]$$

. Exponentiating the values above, we get

$$[0.4847, 1.3833]$$

which is the 95% confidence interval for the multiplicative effect of a one-unit increase in ONI on the count of named storms in North Atlantic basin.

5.2 Logistic Model

Recall the logistic model after backward selection is

$$\hat{y} = e^{-52.60 + 0.02569x_1 + 132.1x_2 + 0.3758x_3 + -0.06596x_1x_2 - 0.5037x_2x_3}$$

where x_1 is Season, $x_2 = 0$ for Eastern Pacific and 1 for North Atlantic, x_3 is ONI.

First I assess the goodness-of-fit of the model. I am tempted to use the deviance for the assessment. However, after inspecting fitted values, it turns out some of the fitted values are too small to render the test to be optimal (in which case χ^2 does not well approximate the distribution of the test statistic.)

As is suggested by the exam instructions, I choose to use standardized residuals to assess goodness-of-fit. Testing result shows 14 observations have a standardized residual with absolute value being greater than 1, which is alarming given that we have only 44 observations.

Analysis also shows that Observations 5 and 41 have overly large Cook's Distance.

Dfbetas are inspected for the model. Results show that removal of any of the following observations will result in drastic change of at least two parameters: 5 and 41.

6 Conclusion

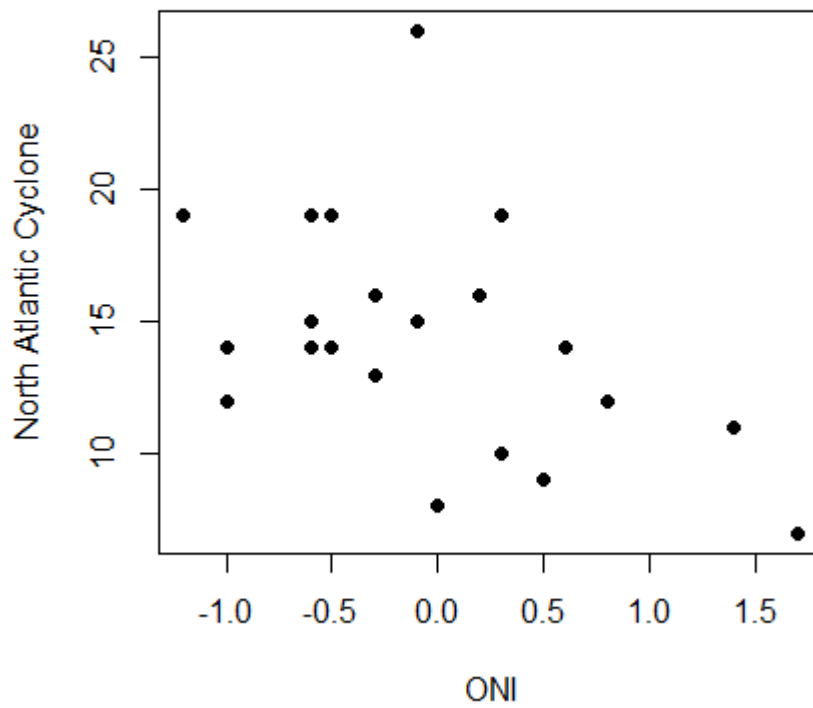
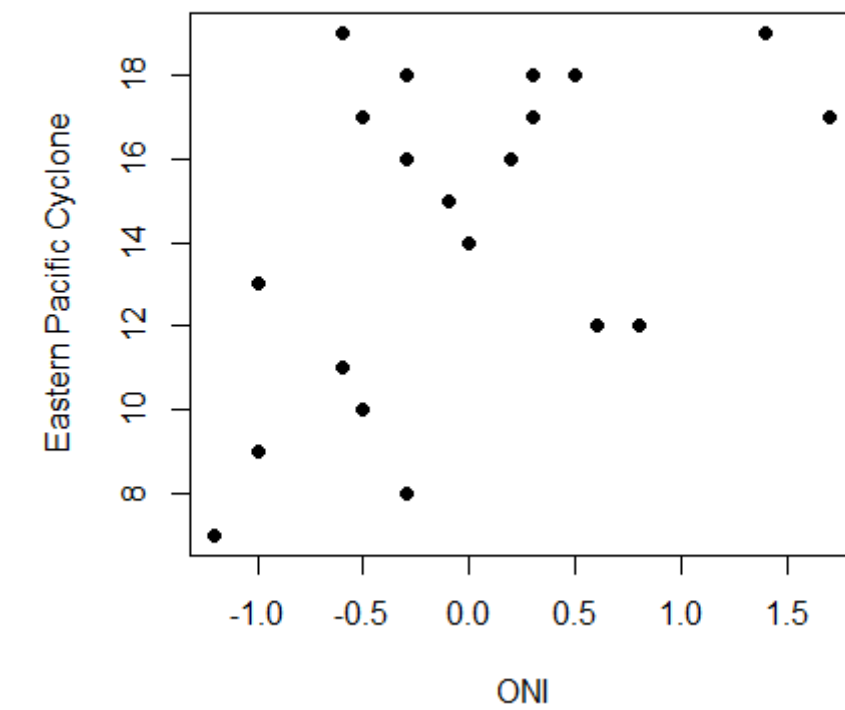
The analysis above shows that cyclones are associated with many predictor variables. Specifically, the loglinear regression model shows that the total number of cyclones in each basin is correlated to ONI at a 95% significant level. In Eastern Pacific basin, this correlation is positive; in Northern Atlantic basin, this correlation is negative.

Logistic regression model shows the probability of a named storm developing into a Major Hurricane is correlated with Basin, ONI, and the interactive term of Season and Basic. Specifically, in Eastern Pacific basin, higher ONI tends to increase the probability of a storm developing into a Major Hurricane. In Northern Atlantic basin, however, the higher ONI tends to decrease the probability of a storm developing into a Major Hurricane.

Neither model provides enough evidence to conclude the number of storms increase or decrease over the years.

7 Appendix

Plots of the number of storms, by basin, are shown below.



All R codes associated with this report are given below, with annotations.

```

# Part1

# Reading data
cyclone <- read.table('cyclone.txt', header = TRUE)
head(cyclone)

# Part 2

# Marginal Totals
apply(cyclone[cyclone$Basin == 'EasternPacific',]
      [, c(3,4,5,6)], 2, sum)
apply(cyclone[cyclone$Basin == 'NorthAtlantic',]
      [, c(3,4,5,6)], 2, sum)

# Plot

EasternPacific <- subset(cyclone, Basin == 'EasternPacific')
NorthAtlantic <- subset(cyclone, Basin == 'NorthAtlantic')
plot(EasternPacific$ONIAug, EasternPacific$Total,
     pch = 16, xlab = "ONI", ylab = "Eastern_Pacific_Cyclone")
plot(NorthAtlantic$ONIAug, NorthAtlantic$Total,
     pch = 16, xlab = "ONI", ylab = "North_Atlantic_Cyclone")

# Part 3

## Fitting Loglinear model
cyclonefit1 <- glm(Total ~ Season + Basin + ONIAug + Basin : ONIAug,
                  family = 'poisson', data = cyclone)
summary(cyclonefit1)

### Goodness-of-Fit Test

deviance(cyclonefit1)

df.residual(cyclonefit1)

1 - pchisq(deviance(cyclonefit1), df.residual(cyclonefit1)) # P-value

# Part 4

# Logistic Regression

cyclonefit2 <- glm(cbind(MajorHurricane, Total - MajorHurricane) ~
                  Season * Basin * ONIAug, family = "binomial", data = cyclone)
backmod <- step(cyclonefit2)
summary(backmod)

```

```

exp(predict(backmod, data.frame(Season = 2017,
                                Basin = 'EasternPacific', ONIAug = -1.5)))
exp(predict(backmod, data.frame(Season = 2017,
                                Basin = 'EasternPacific', ONIAug = 0)))
exp(predict(backmod, data.frame(Season = 2017,
                                Basin = 'EasternPacific', ONIAug = 1.5)))
exp(predict(backmod, data.frame(Season = 2017,
                                Basin = 'NorthAtlantic', ONIAug = -1.5)))
exp(predict(backmod, data.frame(Season = 2017,
                                Basin = 'NorthAtlantic', ONIAug = 0)))
exp(predict(backmod, data.frame(Season = 2017,
                                Basin = 'NorthAtlantic', ONIAug = 1.5)))

# Part 5

vcov(cyclonefit1)

# Check if any fitted values are too small, which turns out to be
  the case. Therefore we cannot use  $\chi^2$  to approximate deviance.

fitted.values(cyclonefit2)*cyclone$Total

# check standardized residuals, for goodness-of-fit

stdres <- rstandard(cyclonefit2, type="pearson")
sum(abs(stdres)>=1)

# check Cook's distance
cooks.distance(cyclonefit2)
plot(cooks.distance(cyclonefit2), pch = 16, ylab = "Cook's Distance")

# Check dfbetas
dfbetas(cyclonefit2)
abs(dfbetas(cyclonefit2))>=1

```