

The questions of interest in Exercise 28 from Chapter 10 are to explore the effects of El Niño temperature and rain in West Africa on the number of tropical storms, hurricanes, and a storm index for the Atlantic Basin. For the lecture, I will examine the effects of these variables on the number of tropical storms. Your homework assignment asks you to do a similar analysis on each of the response variables. There is no single correct analysis of this data. (For example, my analysis will differ substantially from that in the solution manual for instructors.)

This document shows the R commands but none of the output or graphs to carry out a rather thorough analysis of the data from this exercise.

El Niño temperature is categorized as cold, neutral, and warm and West African seasons are classified as wet or dry. The data set in the textbook creates an artificial numerical coding of El Niño temperature as -1 , 0 , and 1 and also has the West African variable coded as 0 or 1 . Our first task will be to create a new data set that has these variables as proper categorical variables.

```
> ex1028 <- read.table("sleuth/ex1028.csv", header = T, sep = ",")
> attach(ex1028)
> africa <- rep("A", ncol(ex1028))
> africa[west.africa == 0] <- "dry"
> africa[west.africa == 1] <- "wet"
> x <- data.frame(year, el.nino, africa, storms, hurricanes, index = storm.index)
> rm(africa)
> detach()
> attach(x)
```

Next, we should make some plots of the response *storm* versus the explanatory variables *year*, *el.nino*, and *africa*. I will make two scatterplots of storms versus year, one showing *elnino* with different symbols and one showing *africa* with different symbols.

```
> par(mfrow = c(1, 2))
> levelsELNino <- levels(as.factor(el.nino))
> levelsAfrica <- levels(as.factor(africa))
> plot(year, storms, type = "n")
> for (i in 1:length(levelsELNino)) {
+   set <- el.nino == levelsELNino[i]
+   points(year[set], storms[set], pch = i)
+ }
> legend(1950, 19.5, levelsELNino, pch = 1:length(levelsELNino))
> plot(year, storms, type = "n")
> for (i in 1:length(levelsAfrica)) {
+   set <- africa == levelsAfrica[i]
+   points(year[set], storms[set], pch = i)
+ }
> legend(1950, 19.5, levelsAfrica, pch = 1:length(levelsAfrica))
> par(mfrow = c(1, 1))
```

It is clear from the plots that there is a relationship of *storms* with both *el.nino* and *africa*. There is no obvious time trend. We can also consider similar plots for the log transformed variable.

```
> par(mfrow = c(1, 2))
> levelsELNino <- levels(as.factor(el.nino))
> levelsAfrica <- levels(as.factor(africa))
> plot(year, log(storms), type = "n")
> for (i in 1:length(levelsELNino)) {
+   set <- el.nino == levelsELNino[i]
+   points(year[set], log(storms[set]), pch = i)
+ }
> legend(1950, 19.5, levelsELNino, pch = 1:length(levelsELNino))
```

```

> plot(year, log(storms), type = "n")
> for (i in 1:length(levelsAfrica)) {
+   set <- africa == levelsAfrica[i]
+   points(year[set], log(storms[set]), pch = i)
+ }
> legend(1950, 19.5, levelsAfrica, pch = 1:length(levelsAfrica))
> par(mfrow = c(1, 1))

```

We can consider a linear model to predict *storms* based on all three variables.

```

> fit1 <- lm(storms ~ year + el.nino + africa)
> summary(fit1)
> plot(fitted(fit1), residuals(fit1))
> abline(h = 0, lty = 2)
> lines(lowess(fitted(fit1), residuals(fit1)))

```

The summary indicates that there is at most marginal evidence of a time effect, but that both *el.nino* and *africa* have significant effects. The residual plot does not indicate any great need for a transformation, although we could try a log transformation for the fun of it. I have added a local regression fit to the residual plots to make spotting nonlinear trends easier.

```

> fit2 <- lm(log(storms) ~ year + el.nino + africa)
> summary(fit2)
> plot(fitted(fit2), residuals(fit2))
> abline(h = 0, lty = 2)

```

To compare the two fits, we could examine normal probability plots of the residuals.

```

> par(mfrow = c(1, 2))
> qqnorm(residuals(fit1))
> qqnorm(residuals(fit2))

```

Next, let me try a fit without year, but with an interaction between *elnino* and *africa* for both the untransformed and transformed variable.

```

> fit3 <- lm(storms ~ el.nino * africa)
> summary(fit3)
> plot(fitted(fit3), residuals(fit3))
> abline(h = 0, lty = 2)
> lines(lowess(fitted(fit3), residuals(fit3)))
> fit4 <- lm(log(storms) ~ el.nino * africa)
> summary(fit4)
> plot(fitted(fit4), residuals(fit4))
> abline(h = 0, lty = 2)
> lines(lowess(fitted(fit4), residuals(fit4)))

```

In both cases, the interaction term is not significant. So, lets make fits without the interaction terms.

```

> fit5 <- lm(storms ~ el.nino + africa)
> summary(fit5)
> plot(fitted(fit5), residuals(fit5))
> abline(h = 0, lty = 2)
> lines(lowess(fitted(fit5), residuals(fit5)))
> fit6 <- lm(log(storms) ~ el.nino + africa)
> summary(fit6)
> plot(fitted(fit6), residuals(fit6))
> abline(h = 0, lty = 2)
> lines(lowess(fitted(fit6), residuals(fit6)))

```

Now, there is only marginal evidence that the variable *africa* is significant, when *storms* is transformed or not. Finally, I will fit a model with $\log(\text{storms})$ as the response using *el.nino* as the sole explanatory variable.

```
> fit7 <- lm(log(storms) ~ el.nino)
> summary(fit7)
> plot(fitted(fit7), residuals(fit7))
> abline(h = 0, lty = 2)
> lines(lowess(fitted(fit7), residuals(fit7)))
```

Again, only one of the levels of *el.nino* is significant. There is little evidence for treating cold and neutral values separately. Here is an eighth fit with a single indicator variable that *el.nino* is warm.

```
> warm <- as.factor(el.nino == "warm")
> notWarm <- as.factor(el.nino != "warm")
> fit8 <- lm(log(storms) ~ warm)
> summary(fit8)
> plot(fitted(fit8), residuals(fit8))
> abline(h = 0, lty = 2)
```

Finally, here is a ninth fit with the log transformed variable and *notWarm* as the only explanatory variable to make find 95% confidence intervals easier.

```
> fit9 <- lm(log(storms) ~ notWarm)
> summary(fit9)
```

Here are some calculations useful for the summary below.

```
> fit8.coef <- summary(fit8)$coefficients
> fit9.coef <- summary(fit9)$coefficients
> tcrit <- qt(0.975, 46)
> est8 <- round(exp(fit8.coef[1, 1]))
> est9 <- round(exp(fit9.coef[1, 1]))
> lo8 <- round(exp(fit8.coef[1, 1] - tcrit * fit8.coef[1, 2]))
> hi8 <- round(exp(fit8.coef[1, 1] + tcrit * fit8.coef[1, 2]))
> lo9 <- round(exp(fit9.coef[1, 1] - tcrit * fit9.coef[1, 2]))
> hi9 <- round(exp(fit9.coef[1, 1] + tcrit * fit9.coef[1, 2]))
```

After all of this fitting and plotting and analysis, I would conclude that the simplest model with an indicator only for warm El Niño temperatures is sufficient. (Such a simple model may not be sufficient for the other response variables you are considering.)

My solution to the problem would be this.

Exercise 10.28, part (a) I considered several models for estimating the number of tropical storms in the Atlantic Basin based on El Niño temperature, whether or not West Africa was rainy or dry, and time. The variables were *el.nino* temperature (levels cool, neutral, and warm), *africa* (levels rainy and dry), and *year*. Models for the logarithm of the number of storms were more nearly linear (as judged by residual plots) and had residuals with more similar variability across different values of the explanatory variables, so we selected $\log(\text{storms})$ to be the response variable.

In a model with explanatory variables *year*, *el.nino*, and *africa*, *year* was not significant ($p = 0.17$), so we dropped *year* from further consideration. In a model with *el.nino* and *africa* and an interaction term, the interaction term did not significantly improve the fit. In a model with *el.nino* and *africa* and no interaction, *africa* has little significance ($p = 0.11$). A model with *el.nino* alone as an explanatory variable is consistent with other models in that there is strong evidence for a different mean when the El Niño temperature is warm, but the neutral and cold temperatures are not significantly different ($p = 0.24$). Our preferred model has as its single explanatory variable an indicator whether the El Niño temperature is warm. For years in which the El Niño temperature is warm, we estimate a median number of 7 tropical storms (95% confidence interval 6 to 8) and a median number of 10 tropical storms (95% confidence interval 9 to 11) when the temperature is cold or neutral.