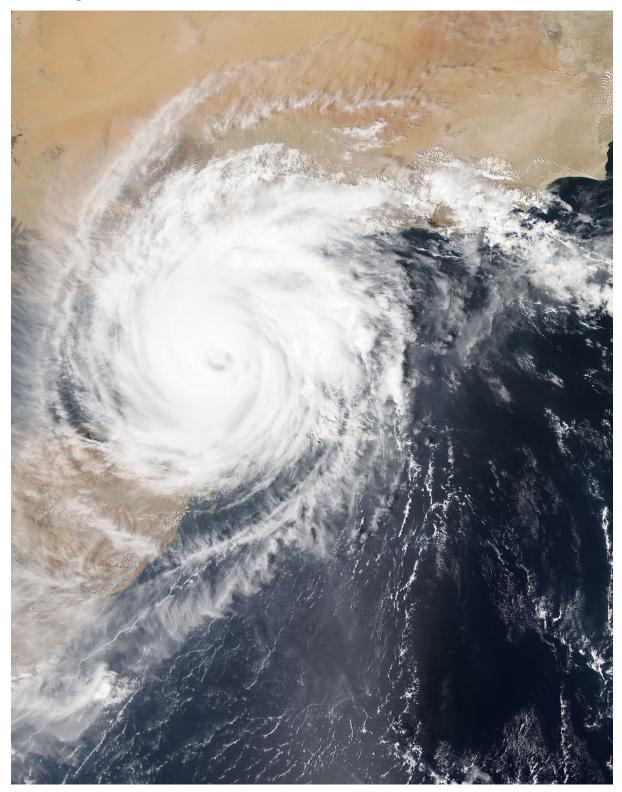# Lecture 2.3 – Developing regression wisdom

Student

2025-04-03

Table of contents

Interpretation steps

Interpretation goals

Interpreting the output of regression tables is an essential skill for a statistical analyst. Being able to tell if your regression coefficients indicate a meaningful relationship as well as reading the diagnostic information found in a regression table is key to model building and statistical literacy. We will develop these skills in three exercises.

- Learning objectives:

    - Practice interpreting slopes
    - Practice interpreting intercepts
    - Develop skills to understand statistical magnitude
    - Understand under what domains models are valid

Interpretation process

These steps are steps that every statistical modeler should take when analyzing data.

1. Before starting, write a sentence or two on your expectation of the relationship – what direction will it be, how strongly will the two be related, and does the predictor variable have a large or small impact on the response variable.
2. What are the units of the predictor and response variable?
3. Draw a scatterplot of the relationship between the two variables. Does it appear that one variable needs to be reexpressed? If so, find the best reexpression possible (the data may not look very pretty even after reinterpretation, your goal is to make the relationship as linear as possible)
4. Conduct and view the results of your regression. Store it in a nice table using the `modelsummary()` command.
5. Interpret the slope of the regression line – does a one unit change in the predictor variable lead to a large or small change in the response variable?
6. What are reasonable values for the predictor variables? For what range of x would our best fit line have meaning and when would it not?
7. Interpret the intercept. Does it have meaning in this case?
8. Solve for the predicted value of the response variable at Q1 and Q3 of the predictor variable. Does moving from Q1 to Q3 of the predictor variable result in a large change (in your opinion) in the response variable?
9. Draw a box plot by hand of the residuals based on the regression results you can see from `summary(<your model name>)`. Does the boxplot suggest any problems? What are the units?
10. Interpret the residual standard error from these same results – how big or small is it?

11. Make a scatterplot of the fitted values ($\hat{y}$) vs. the residuals with a line at the 0 value of the y axis. Interpret this plot.
12. Interpret the R^2 Specifically, how much of the variance in the response variable is this model explaining?
13. What is some more information you'd like to know before making a final conclusion about the relationship between the two variables? Do you think there are any lurking variables?

Real data

Hurricane Data

Variable definitions:

- `Name` - Hurricane name

- `Year` - Numeric

- `LF.WindsMPH` - Maximum sustained windspeed ($>= 1$ minute) to occur along the US coast. Prior to 1980, this is estimated from the maximum windspeed associated with the Saffir-Simpson index at landfall. If 2 or more landfalls, the maximum is taken

- `LF.PressureMB` - Atmospheric pressure at landfall in millibars. If 2 or more landfalls, the minimum is taken

- `LF.times` - Number of times the hurricane made landfall

- `BaseDamage` - Property damage (in millions of dollars for that year)

- `NDAM2014` - Damage, had hurricane appeared in 2014

- `AffectedStates` - Affected states (2-digit abbreviations), pasted together

- `firstLF` - Date of first landfall

- `deaths` - Number of continental US direct and indirect deaths

- `mf` - Gender of name

- `LF.WindsKPH` - Maximum sustained windspeed expressed in kilometers/hr

Investigate

Using the 13 steps described above, investigate the relationship between two variables that you think should be related. Write up your response to each of the fourteen points listed above. If you finish early, pick another predictor variable that you think may be related to the response variable and repeat the process.