Stat 610 Homework 4

Due Wednesday, October 1, 11:59pm

Assignment

We are going to be interested in the relationship between tornado magnitude and tornado damage, and whether the relationships is different from state to state. As we discussed in class, please make a first attempt without using generative AI. If you are happy with the code you wrote on your own, you can turn in just what you wrote. If you felt like you needed to use chatGPT or something similar, compare the code you wrote to what chatGPT gives you in response to the questions. Note whether the code works and whether there are any stylistic or substantive differences with what you wrote.

You should do the following:

1. Get the data using `td <- readRDS('tornadoes.RDS')`
2. Use `lapply` or `sapply` and the `typeof` function to get the classes of all of the columns of `td`. What happens when you use `lapply` or `sapply` on a data frame? Why?
3. Using the split/apply/combine strategy discussed in class, compute the fraction of tornadoes with magnitude (`mag`) greater than or equal to 3 for each state. Remember that the previous part showed that magnitude is of type character and you will need to make it into a number.
4. Write a function that takes a data frame as its argument and returns the slope coefficient in a linear model with loss (`loss`) as a linear function of magnitude (`mag`, coded as a number). (Hint: the `coef` function will extract the coefficients from the output of the `lm` function.)
5. Using the split/apply/combine strategy discussed in class and the function you wrote in the previous section, compute the slope in a linear model of loss as a function of magnitude for each state.
6. Notice that the coding for loss changes after 1996 (https://www.spc.noaa.gov/wcm/data/SPC_severe_database_description.pdf), that loss = 0 means unknown, and that mag = -9 also means unknown. Split the analysis by pre/post-1996 in addition to state (so now there should be two values for each state, one for pre-1996 and one for post-1996) and modify either the dataset or your function so that it does not include loss = 0 or mag = -9 in the calculations.
7. One might expect states that see strong tornadoes more often to build more defensively against hurricanes. To check this, merge the results in parts (3) and (6), and see if there is a correlation between the fraction of hurricanes with magnitude at least 3 and the increase in loss due to stronger hurricanes. Since you have pre- and post-1996 slopes for each state, you should do the analysis separately for each time period.

Submission parameters

You should submit an Rmd file and the corresponding pdf or html on canvas. The files should contain both the code you ran and the answers to the problems.