P8110: Applied Regression II Spring 2024

Homework #8 [14 points]

Due on Apr 8, 11:59**AM**

**NOTE: Use robust standard errors in GEE.**

The “hwdata4.csv” contains measurements on 79 spruce trees over the growing season. 54 trees were grown in an environment with introduced ozone at 70 ppb, while the other 25 were grown in an ozone-free environment. The variables included in the dataset are:

**size** size of the tree measured as log(height × diameter2)

**time** days after January 1st of the year

**tree** id number of the tree

**treat** ozone: grown under ozone environment; control: ozone free

1. Fit a GEE model with size of the tree as outcome and time, environment, andtheir interaction as covariates. Write down the mean response of the GEE model. [2 points]
2. Try different working correlation structures (CS and AR(1)) for the GEEmodel. Which model yields the better QIC value? Show the SAS/R code and relevant output. [2 points] (For R users, use geepack package and geeglm, geepack::QIC functions)
3. Use the model selected in (2) to test whether the trajectory of tree size overtime is different between the two environments. Write down the hypothesis, test statistic, p-value, and conclusion. [3 points]
4. Use the model selected in (2) to estimate the mean tree size change from day100 to day 200 after January 1st for trees grown in ozone environment and those grown in ozone-free environment, respectively. [4 points]
5. Calculate the difference of the two estimates in (4). Denote the difference asDIFF. Which *β* coefficient is DIFF related to? Interpret this *β* coefficient. [3 points]