P8110: Applied Regression II

Camille Okonkwo uni: co2554 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, and their 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)

R Output:

A screenshot of a computer

Description automatically generated



R Code:

```{r}

library(geepack)

# model with CS

fit\_cs =

geeglm(size ~ time \* treat,

data = hwdata4,

id = tree,

family = gaussian,

corstr = "exchangeable")

summary(fit\_cs)

# model with AR(1)

fit\_ar1 =

geeglm(size ~ time \* treat,

data = hwdata4,

id = tree,

family = gaussian,

corstr = "ar1")

summary(fit\_ar1)

# extract QIC

geepack::QIC(fit\_cs)

geepack::QIC(fit\_ar1)

```

**The CS correlation structure for the GEE model has a lower QIC value (289.8) compared to the AR(1) correlation structure (294.1), thus has a better fit since a smaller QIC value is preferred.**

1. 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]

**Hypothesis:**

**Test Statistic:**

**A screenshot of a computer

Description automatically generated**



**p-value:**

**A close-up of a number

Description automatically generated**



**Conclusion:**

1. 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]

**The mean tree size change from day 100 to day 200 after January 1st for trees grown in ozone environment is 0.363 log(). The mean tree size change from day100 to day 200 after January 1st for trees grown in ozone-free environment is 0.371 log().**

1. Calculate the difference of the two estimates in (4). Denote the difference as DIFF. Which *β* coefficient is DIFF related to? Interpret this *β* coefficient. [3 points]

**DIFF is related to the *β* coefficient interaction time\*treat (). Compared to the control, trees in an ozone environment of 70** **ppb have a 0.00803% reduction in the effect of time on tree size.**

R Code:

---

title: "hw8"

author: "Camille Okonkwo"

date: "2024-04-06"

output: pdf\_document

---

```{r setup, include=FALSE}

knitr::opts\_chunk$set(echo = TRUE)

library(tidyverse)

```

# Import Data and Preparation

```{r}

hwdata4 = read\_csv("data/hwdata4.csv")

# data prep

hwdata4$treat = as.factor(hwdata4$treat)

hwdata4$tree = as.factor(hwdata4$tree)

```

1. Fit a GEE model with size of the tree as outcome and time, environment, and their interaction as covariates. Write down the mean response of the GEE model.

```{r}

# fit the model

library(gee)

fit\_1 =

gee(size ~ time \* treat,

data = hwdata4,

id = tree,

family = gaussian)

summary(fit\_1)

```

E(y\_ij) = β\_0 + β\_1(X\_1) + β\_2(X\_2) + β\_3(X\_1)(X\_2)

where

y\_ij=size of the tree measured as log(height × diameter^2)

X\_1=treat{(0=control and 1=ozone)

X\_2=days after January 1st of the year

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)

```{r}

library(geepack)

# model with CS

fit\_cs =

geeglm(size ~ time \* treat,

data = hwdata4,

id = tree,

family = gaussian,

corstr = "exchangeable")

summary(fit\_cs)

# model with AR(1)

fit\_ar1 =

geeglm(size ~ time \* treat,

data = hwdata4,

id = tree,

family = gaussian,

corstr = "ar1")

summary(fit\_ar1)

# extract QIC

geepack::QIC(fit\_cs)

geepack::QIC(fit\_ar1)

```

The CS correlation structure for the GEE model has a lower QIC value (289.8) compared to the AR(1) correlation structure (294.1), thus has a better fit since a smaller QIC value is preferred.

3. Use the model selected in (2) to test whether the trajectory of tree size over time is different between the two environments. Write down the hypothesis, test statistic, p-value, and conclusion.

```{r}

fit\_cs\_2 =

gee(size ~ time \* treat,

data = hwdata4,

id = tree,

family = gaussian,

corstr = "exchangeable")

summary(fit\_cs\_2)

# Wald test p-values

1- pchisq(fit\_cs\_2$coefficients^2/

diag(fit\_cs\_2$robust.variance),df=1)

```

We fail to reject H\_0 at α=0.05. There is insufficient evidence to conclude that the trajectory of tree size over time is different between the ozone and control environments.