P8110: Applied Regression II

Homework #8 [14 points]

NOTE: Please attach your SAS code. Use robust standard errors in GEE.

In a study of septic patients, each patient's temperature was measured at baseline, and 2, 4, and 8 hours after entry into study. Patients were randomly assigned to two treatment groups at baseline. Patients' APACHE scores were also measured. The data are saved in 'HW8.csv'. The columns of variables from left to right are:

ID	= patient ID
temp	= patient's temperature
treatment	= 1 - treatment B, 0 - treatment A
apache	= APACHE score at baseline
time	= 0, 2, 4, 8 hours after entry into study

1. **Fit a GEE model with temperature as outcome and time, treatment, and their interactions as covariates. Write the mean response of the GEE model and treat time as a categorical variable.** [**2 points**]

   The GEE model for temperature with time, treatment, and their interaction term is

$$E(Y_{ij}) = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij3} + \beta_4 X_{i4} + \beta_5 X_{ij1} X_{i4} + \beta_6 X_{ij2} X_{i4} + \beta_7 X_{ij3} X_{i4}$$

   Where,

$$Y_{ij} = \text{patient's temperature}, \quad j = 1 \cdots, 4$$

$$X_{ij1} = \begin{cases} 1, & \text{if time} = 2 \\ 0, & \text{otherwise} \end{cases}, \qquad X_{ij2} = \begin{cases} 1, & \text{if time} = 4 \\ 0, & \text{otherwise} \end{cases},$$

$$X_{ij3} = \begin{cases} 1, & \text{if time} = 8 \\ 0, & \text{otherwise} \end{cases}, \qquad X_{i4} = \begin{cases} 1, & \text{if treatment B} \\ 0, & \text{if treatment A} \end{cases}$$

2. **Try different working correlation structures (CS, AR(1), and UN) for the GEE model in (1). Which model yields the best QIC value? Show the SAS code and relevant SAS output.** [**5 points**]

   QIC values for these three working correlation structures.

   SAS output:

   GEE Fit Criteria (CS)

GEE Fit Criteria	
QIC	1702.3998
QICu	1703.0000

GEE Fit Criteria (AR(1))

GEE Fit Criteria	
QIC	1702.2653
QICu	1703.0000

GEE Fit Criteria (UN)

GEE Fit Criteria	
QIC	1702.2859
QICu	1703.0000

The QIC of AR(1) model is the smallest.

3. **Use the model selected in (2) to test whether the trajectory of temperature over time is different between the two treatments. Write down the hypothesis, test statistic, p-value, and conclusion.** [**3 points**]

   The hypothesis to investigate whether the trajectory of temperature over time is different between the two treatments is

$$H_0 : \beta_5 = \beta_6 = \beta_7 = 0 \text{ VS } H_\alpha : \text{ not } H_0$$

The following is the SAS output for the type 3 wald test result for these three working correlation structures.

SAS output:

Wald Statistics For Type 3 GEE Analysis			
Source	**DF**	**Chi-Square**	**Pr > ChiSq**
time	3	25.87	<.0001
treatment	1	0.46	0.4982
treatment*time	3	69.37	<.0001

Therefore the Wald statistic for AR(1) is

$$W_{AR(1)} = 69.37$$

with 3 degrees of freedom.

P-values and Conclusion: The p-value is less than .0001. We reject the null hypothesis and reach the conclusion that the trajectory of temperature over time is significantly different between the two treatments.

4. **Use the model selected in (2) to estimate the mean temperature change from baseline to two hours after entry into study for patients in treatment A group and those in treatment B group, respectively.** [**4 points**]

The maximum likelihood estimates for this model is

SAS output:

Analysis Of GEE Parameter Estimates							
Empirical Standard Error Estimates							
Parameter		**Estimate**	**Standard Error**	**95% Confidence Limits**		**Z**	**Pr > |Z|**
Intercept		100.4902	0.1256	100.2440	100.7364	800.07	<.0001
time	2	-0.3286	0.0722	-0.4700	-0.1871	-4.55	<.0001
time	4	-0.3752	0.0912	-0.5540	-0.1964	-4.11	<.0001
time	8	-0.4324	0.1113	-0.6506	-0.2142	-3.88	0.0001
treatment		-0.1287	0.1900	-0.5011	0.2437	-0.68	0.4982
treatment*time	2	-0.5395	0.1111	-0.7572	-0.3218	-4.86	<.0001
treatment*time	4	-1.0850	0.1453	-1.3698	-0.8003	-7.47	<.0001
treatment*time	8	-1.2962	0.1615	-1.6127	-0.9798	-8.03	<.0001

The mean temperature at baseline in treatment A is

$$\begin{aligned} E(Y_{i1}|A, time = 0) &= \beta_0 + \beta_1 \times 0 + \beta_2 \times 0 + \beta_3 \times 0 + \beta_4 \times 0 \\ &+ \beta_5 \times 0 + \beta_6 \times 0 + \beta_7 \times 0 = \beta_0 \end{aligned}$$

The mean temperature at 2 hours in treatment A is

$$\begin{aligned} E(Y_{i2}|A, time = 2) &= \beta_0 + \beta_1 \times 1 + \beta_2 \times 0 + \beta_3 \times 0 + \beta_4 \times 0 \\ &+ \beta_5 \times 0 + \beta_6 \times 0 + \beta_7 \times 0 = \beta_0 + \beta_1 \end{aligned}$$

Therefore mean temperature change from baseline to two hours after entry into study for patients in treatment A group is $\beta_0 + \beta_1 - \beta_0 = \beta_1$. Therefore its estimate is $\widehat{\beta_1} = -0.3286$

Similarly, the mean temperature at baseline in treatment B is

$$\begin{aligned} E(Y_{i1}|B, time = 0) &= \beta_0 + \beta_1 \times 0 + \beta_2 \times 0 + \beta_3 \times 0 + \beta_4 \times 1 \\ &+ \beta_5 \times 0 + \beta_6 \times 0 + \beta_7 \times 0 = \beta_0 + \beta_4 \end{aligned}$$

The mean temperature at 2 hours in treatment B is

$$\begin{aligned} E(Y_{i2}|B, time = 2) &= \beta_0 + \beta_1 \times 1 + \beta_2 \times 0 + \beta_3 \times 0 + \beta_4 \times 1 \\ &+ \beta_5 \times 1 + \beta_6 \times 0 + \beta_7 \times 0 = \beta_0 + \beta_1 + \beta_4 + \beta_5 \end{aligned}$$

Therefore mean temperature change from baseline to two hours after entry into study for patients in treatment B group is

$$\beta_0 + \beta_1 + \beta_4 + \beta_5 - (\beta_0 + \beta_4) = \beta_1 + \beta_5$$

Therefore its estimate is $\hat{\beta}_1 + \hat{\beta}_5 = -0.3286 - 0.5395 = -0.8680$

5. **Calculate the difference of the two estimates in (4). Denote the difference as DIFF. Which β coefficient does DIFF represent? Interpret this β coefficient. [3 points]**

The difference of the two estimates is $Diff = \beta_5 + \beta_1 - \beta_1 = \beta_5$. Its estimate is $\hat{\beta}_5 = -0.5395$. Therefore the coefficient of $X_{ij1}X_{i4}$ (β_5) represents $Diff$. It can be interpreted as the mean difference of temperature change from baseline to two hours after entry into study between two treatment groups.

SAS output:

Analysis Of GEE Parameter Estimates							
Empirical Standard Error Estimates							
Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		100.4902	0.1256	100.2440	100.7364	800.07	<.0001
time	2	-0.3286	0.0722	-0.4700	-0.1871	-4.55	<.0001
time	4	-0.3752	0.0912	-0.5540	-0.1964	-4.11	<.0001
time	8	-0.4324	0.1113	-0.6506	-0.2142	-3.88	0.0001
treatment		-0.1287	0.1900	-0.5011	0.2437	-0.68	0.4982
treatment*time	2	-0.5395	0.1111	-0.7572	-0.3218	-4.86	<.0001
treatment*time	4	-1.0850	0.1453	-1.3698	-0.8003	-7.47	<.0001
treatment*time	8	-1.2962	0.1615	-1.6127	-0.9798	-8.03	<.0001

Analysis Of GEE Parameter Estimates							
Empirical Standard Error Estimates							
Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept		100.6804	0.2196	100.2500	101.1107	458.53	<.0001
time	2	-0.3299	0.0725	-0.4720	-0.1878	-4.55	<.0001
time	4	-0.3762	0.0916	-0.5558	-0.1966	-4.11	<.0001
time	8	-0.4299	0.1116	-0.6487	-0.2111	-3.85	0.0001
treatment		-0.1358	0.1898	-0.5078	0.2361	-0.72	0.4742
treatment*time	2	-0.5385	0.1113	-0.7566	-0.3204	-4.84	<.0001
treatment*time	4	-1.0846	0.1455	-1.3698	-0.7994	-7.45	<.0001
treatment*time	8	-1.3001	0.1618	-1.6171	-0.9830	-8.04	<.0001
apache		-0.0118	0.0140	-0.0393	0.0156	-0.84	0.3982

SAS Code:

```
data sepsis;
infile "C:\HW4.csv" delimiter = ',' MISSOVER DSD;
input id temp treatment apache time;
run;
/* Q2*/
proc genmod data=sepsis;
class ID time (ref='0') /param=ref;
model temp = time treatment treatment*time/dist=nor link=identity type3 Wald;
repeated subject=ID /type=cs modelse;
run;
proc genmod data=sepsis;
class ID time (ref='0') /param=ref;
model temp = time treatment treatment*time/dist=nor link=identity type3 Wald;
repeated subject=ID /type=AR(1) modelse;
run;
proc genmod data=sepsis;
class ID time (ref='0') /param=ref;
model temp = time treatment treatment*time/dist=nor link=identity type3 Wald;
repeated subject=ID /type=UN modelse;
run;
```