

Homework 10

Alex Soupir

November 23, 2019

Packages: HSAUR3, gee, lme4, Matrix, multcomp, ggplot2, gridextra, tidyr, dplyr

Collaborators:

Chapter 14

Please do the following problems from the text book R Handbook and stated.

1. Consider the **respiratory** data from the **HSAUR3** package.
 - a. Investigate the use of other correlational structures than the independence and exchangeable structures used in the text for the respiratory data.

```
## (Intercept)      centre2      trttrt  gendermale baselinegood      age
## -0.90017133    0.67160098  1.29921589  0.11924365   1.88202860 -0.01816588

## (Intercept)      centre2      trttrt  gendermale baselinegood      age
## -0.90017133    0.67160098  1.29921589  0.11924365   1.88202860 -0.01816588

## (Intercept)      centre2      trttrt  gendermale baselinegood      age
## -0.90017133    0.67160098  1.29921589  0.11924365   1.88202860 -0.01816588

## (Intercept)      centre2      trttrt  gendermale baselinegood      age
## -0.90017133    0.67160098  1.29921589  0.11924365   1.88202860 -0.01816588

##
## Coefficients for correlation structure 'stat_M_dep':

## $coefficients
##           Estimate Naive S.E.      Naive z Robust S.E.      Robust z
## (Intercept) -0.96294479 0.42071989 -2.2888026  0.46116071 -2.0880894
## centre2      0.74270146 0.29713257  2.4995626  0.35622997  2.0848932
## trttrt       1.24728240 0.29394223  4.2432909  0.35189738  3.5444492
## gendermale   0.11323226 0.36675166  0.3087437  0.44945059  0.2519348
## baselinegood 1.91139525 0.29916032  6.3892005  0.35018727  5.4582088
## age         -0.01691639 0.01101599 -1.5356209  0.01292727 -1.3085816

##
## Coefficients for correlation structure 'non_stat_M_dep':

## $coefficients
##           Estimate Naive S.E.      Naive z Robust S.E.      Robust z
## (Intercept) -0.98263481 0.41928627 -2.3435893  0.4633336 -2.1207932
## centre2      0.74848974 0.29571189  2.5311452  0.3573102  2.0947897
## trttrt       1.24711331 0.29268905  4.2608814  0.3520803  3.5421273
## gendermale   0.11671467 0.36524841  0.3195488  0.4510766  0.2587469
## baselinegood 1.90029171 0.29758567  6.3856963  0.3505178  5.4213839
## age         -0.01646838 0.01096429 -1.5020019  0.0129291 -1.2737454

##
## Coefficients for correlation structure 'AR-M':

## $coefficients
##           Estimate Naive S.E.      Naive z Robust S.E.      Robust z
```

```
## (Intercept) -0.96294479 0.44551609 -2.1614142 0.46116071 -2.0880894
## centre2 0.74270146 0.31464483 2.3604438 0.35622997 2.0848932
## trttrt 1.24728240 0.31126647 4.0071210 0.35189738 3.5444492
## gendermale 0.11323226 0.38836711 0.2915599 0.44945059 0.2519348
## baselinegood 1.91139525 0.31679209 6.0335952 0.35018727 5.4582088
## age -0.01691639 0.01166525 -1.4501525 0.01292727 -1.3085816

##
## Coefficients for correlation structure 'unstructured':

## $coefficients
## Estimate Naive S.E. Naive z Robust S.E. Robust z
## (Intercept) -0.93127982 0.47918516 -1.9434655 0.46124991 -2.0190352
## centre2 0.67279473 0.33907786 1.9841895 0.35482020 1.8961568
## trttrt 1.27891545 0.33544088 3.8126404 0.34944996 3.6597956
## gendermale 0.09467354 0.41729637 0.2268736 0.44362952 0.2134068
## baselinegood 1.93462516 0.34281837 5.6432949 0.34804681 5.5585200
## age -0.01688917 0.01255736 -1.3449620 0.01290535 -1.3086948

## (Intercept) centre2 trttrt gendermale baselinegood age
## 4.348356e-02 5.793932e-02 2.524165e-04 8.310097e-01 2.720717e-08 1.906378e-01
```

Looking at the other correlation structures, minus the “fixed” correlation structure because the command never seemed to finish for some reason, is shown above with the coefficients for each covariate. The starting point for each variable is the same between models, as mentioned in the lecture video. However, the differences between the naive standard error and the robust standard error does differ between the correlation structures `stat_M_dep`, `non_stat_M_dep`, `AR_M`, and `unstructured`. The correlation structure that resulted in the greatest differences overall between the the naive s.e. and the robust s.e. was the `non_stat_M_dep`, but it wasn’t great as the difference seen in the lecture model with the correlation structure of independence. The least difference between naive s.e. and robust s.e. was with the model using `unstructured` correlation structure.

Looking closer at the model model with the most similar naive s.e./robust s.e. we can see which of those variables are significant. Treatment is significant and so is the baseline, but gender and age do no show significance.

- b. Which model is the best? Compare the following models: independent, exchangeable, and what ever models you tried in part (a). Justify your answer. (Hint: use QIC (in **MESS**), MSE, misclassification rate, comparison of naive vs robust Z-score, or another method, be sure to state your method)

```
## (Intercept) centre2 trttrt gendermale baselinegood age
## -0.90017133 0.67160098 1.29921589 0.11924365 1.88202860 -0.01816588

## (Intercept) centre2 trttrt gendermale baselinegood age
## -0.90017133 0.67160098 1.29921589 0.11924365 1.88202860 -0.01816588

##
## Coefficients for correlation structure 'independence':

## $coefficients
## Estimate Naive S.E. Naive z Robust S.E. Robust z
## (Intercept) -0.90017133 0.337653052 -2.665965 0.46032700 -1.9555041
## centre2 0.67160098 0.239566599 2.803400 0.35681913 1.8821889
## trttrt 1.29921589 0.236841017 5.485603 0.35077797 3.7038127
## gendermale 0.11924365 0.294671045 0.404667 0.44320235 0.2690501
## baselinegood 1.88202860 0.241290221 7.799854 0.35005152 5.3764332
## age -0.01816588 0.008864403 -2.049306 0.01300426 -1.3969169
```

```

##
## Coefficients for correlation structure 'exchangeable':
## $coefficients
##           Estimate Naive S.E.      Naive z Robust S.E.      Robust z
## (Intercept) -0.90017133  0.4784634 -1.8813796  0.46032700 -1.9555041
## centre2      0.67160098  0.3394723  1.9783676  0.35681913  1.8821889
## trttrt       1.29921589  0.3356101  3.8712064  0.35077797  3.7038127
## gendermale   0.11924365  0.4175568  0.2855747  0.44320235  0.2690501
## baselinegood 1.88202860  0.3419147  5.5043802  0.35005152  5.3764332
## age         -0.01816588  0.0125611 -1.4462014  0.01300426 -1.3969169
##
## MSE of state_M_dep model:
## IF/ELSE 1 or 0 for the model fitted values from lecture
## [1] 0.259009
## Raw value MSE
## [1] 0.1826503
##
## MSE of non_stat_M_dep model:
## IF/ELSE 1 or 0 for the model fitted values from lecture
## [1] 0.259009
## Raw value MSE
## [1] 0.1826893
##
## MSE of AR-M model:
## IF/ELSE 1 or 0 for the model fitted values from lecture
## [1] 0.259009
## Raw value MSE
## [1] 0.1826503
##
## MSE of unstructured model:
## IF/ELSE 1 or 0 for the model fitted values from lecture
## [1] 0.259009
## Raw value MSE
## [1] 0.18265
##
## MSE of independence model:
## IF/ELSE 1 or 0 for the model fitted values from lecture
## [1] 0.259009
## Raw value MSE
## [1] 0.1825935

```

```
##
## MSE of exchangeable model:
## IF/ELSE 1 or 0 for the model fitted values from lecture
## [1] 0.259009
## Raw value MSE
## [1] 0.1825935
```

Adding to the coefficient naive and robust s.e. from Part A we can see that the independence correlation structure has a large difference. This means that we cannot trust the naive z values and want to use the robust, which show less significance due to them being more conservative (2.8 naive v 1.9 robust z). The MSE for the models using the code from lecture code results in the same MSE for all fo the models correcting to either 1 or 0 where 1 is good status and 0 is poor status, depending if the fitted value is >0.5 (1) or <0.5 (0). Not correcting to either 1 or 0 of the nstat and just keeping the raw value for MSE calculations resulted in some divergence but not a great deal (0.1825935 to 0.1826893). Because of this, sticking with the summary results comparing the naive s.e. to the robust s.e. makes more sense where the model with exchangeable, or even unstructured, correlation structures provided the least difference between them and allowing to trust the context of the model. If we were to go further, we could check for interaction terms between the insignificant or less significant variables as suggested in lecture.

2. The data set **schizophrenia2** from **HSAUR3** package were collected in a follow-up study of women patients with schizophrenia (Davis, 2002). The binary response recorded at 0, 2, 6, 8 and 10 months after hospitalization was “thought disorder” (absent or present). The single covariate is the factor indicating whether a patient had suffered early or late onset of her condition (age of onset less than 20 years or age of onset 20 years or above). The question of interest is whether the course of the illness differs between patients with early and late onset schizophrenia. (<https://www.rdocumentation.org/packages/HSAUR3/versions/1.0-9/topics/schizophrenia2>) Investigate this question using
 - a. plots and summary statistics
 - b. the GEE approach
 - c. mixed effects model (lmer) from previous chapter
 - d. Is there a difference? What model(s) work best? Describe your results.