Chapter 10 in R Book

OVERVIEW OF EXPERIMENT

-Part Two of Analyzing Longitudinal Data
-Will present some theory and covert only first example (treating respiratory illness)
-Basic point: generalized estimation equation approach essentially extends generalized linear models to longitudinal data, and allows for analysis of longitudinal data when normality isn't assumed for the response variable
-*p.175 Table 11.1* data for trial comparing two trts. For respiratory illness (only first four patients shown)
-2 centers, at each eligible patients randomly assigned either active treatment or placebo
-respiratory status of poor or good determined at four monthly visits
- 111 participants (54 active, 57 placebo)
-evaluate effect of treatment and whether it's effective

GENERALIZED ESTIMATING EQUATIONS

-in respiratory example, repeatedly measured response variable has non-normal distribution
-cannot assume independence of repeated measurements, our assuming of there being more independent data points than are justified causes us to estimate standard errors too small for between-subject covariates
-sandwich estimate of variance gives good estimates of standard errors we need
- best to use approach that fully uses the information about the data structure-generalized estimating equations (GEE)
- idea behind GEE: structure of correlation matrix not very useful in terms of practicality, want simple structures for within-subject correlations-use working correlation matrix
- estimates of parameters of most interest valid even in cases where we have the wrong idea about the correlation structure
-if the standard error estimates are off, we can use the sandwich estimator
-Useful Working Correlation Matrices:
1. Identity Matrix: good for independent working model, repeated responses assumed to be independent, can use generalized linear modeling software, simple to use
2. Exchangeable Correlation Matrix: with single parameter, gives correlation of each pair of repeated measures
3. Autoregressive Correlation Matrix: single parameter, formula for corr *p.178 next to last paragraph*, variable *p. 178* less than one gives pattern in which repeated measures further apart in time less correlated than closer ones
4. Unstructured Correlation Matrix: # of parameters and corr formula *p. 178 last paragraph*, q is the # of repeated measures, not good because too many parameters

R ANALYSIS

-gee function similar to lme function of Ch. 10, with added features of flm function giving the appropriate error distribution for response and implied link function and argument to specify structure of working correlation matrix
-need to install package gee to use gee function R prompt R>library(“gee”)
- we’ll fit an independence structure and then an exchangeable one
-using R code to fit generalized estimation equations to BtheB_long data of ch. 10 with identity working correlation matrix *p.179 starts with* R>osub,-order(as.integer(BtheB-long$subject))
- with exchangeable correlation *p. 179 starts with* R>btb_gee1<-gee(bdi~bdi.pre+treatment+length….etc
-summary methods to inspect fitted models of identity correlation matrix (*p.180 Figure 11.1 also includes R commands)* and exchangeable correlation matrix (*p.181 Figure 11.2 also includes R commands)*
-naïve and sandwich estimates of standard errors very different in Figure 11. and 11.2
-shows that exchangeable matrix better for these data, standard errors good even without using sandwich method
-random intercept model of Figure 10.2 also implies exchangeable structure for correlation of repeated measurements, results almost the same
-now we’ll apply GEE to respiratory data of Table 11.1
-response variable is binary, choose binomial error distribution and logistic link function
-fix scale parameter, theta (of Ch.6) as 1
-just like before, do procedure 2x, first w/ indep. structure and then w/exchangeable one
-then fit logistic regression model using glm to compare results
-baseline status enters models as explanatory variable, have to enter certain R commands to create this new variable *commands on p. 182, starts with* R>data (“respiratory”, package=HSAUR”)
-nstat is a new variable which codes for poor respiratory status
-now, we’ll use the data resp to fit logistic regresstion model and GEE model w/ indep. And exchangeable structure *R commands on p.182 starts w/* R>resp_gilm,-glm(status~centre+treatment+sex+baseline+age,data=resp,family=”binomial”)…..etc.
-use summary method to get details of fitted models *Figures 11.3 on p. 183, 11.4 on 184, 11.5 on 185*
- results of using regression with glm gives same results as those from gee with independence correlation structure
-robust standard errors for between subject covariateslot bigger than ones you get assuming independence-not realistic to assume independence foe these data
-using GEE with exchangeable correlation structure gives identical naïve and robust standard errors, the standard errors are also similar to those from the independence structure
-exchangeable structure better reflects correlational structure of observed repeated measurements
-estimated treatment effect from exchangeable/GEE model: 1.299, using robust standard errors: 95% CI *enter R commands on top of p. 186*

-values we get reflect effects on log-odds scale, better to exponentiate values to get effects in terms of odds
-treatment effect: 3.666 *enter R command on p. 186 starting w/* R>exp (coef(resp_gee2)….etc. *to get 95 % CI*
-#s we get for CI tell us that odds of getting good respiratory status with active treatment b/w 2x and 7x corresponding odds for placebo