Longitudinal Data: Repeated Measures

More GEE

Alan Hubbard

Oct 23, 2018

Intro

- We will examine 3 different analyses all from the CD4/HIV data.
- For each, we will compare
  - standard linear regression ignoring repeated measures
  - standard linear regression using GEE to adjust inference robustly to account for repeated measures (GEE with independence working correlation, robust variance)
  - GEE with exchangeable working correlation, naive variance
  - GEE with exchangealbe working correlation, robust variance
- Purpose is to compare
  - coefficient estimates comparing independence and different correlation model, and
  - inference comparing robust versus naive (assuming model is correct) versus robust (agnostic correlation model)

Binary time-independent variable (balanced data)

Read in data

```
dat1 = read_csv("deeks_ex1.csv")
```

```
## Parsed with column specification:
## cols(
##   id = col_integer(),
##   cd4 = col_integer(),
##   binage = col_integer()
## )
```

```
tbl_df(dat1)
```

```
## # A tibble: 594 x 3
##       id   cd4 binage
##    <int> <int>  <int>
##  1     1    45      0
##  2     1   119      0
##  3     2   196      1
##  4     2   369      1
##  5     4    29      0
##  6     4   137      0
##  7     5    84      1
##  8     5    93      1
##  9     7   246      0
## 10     7   439      0
## # ... with 584 more rows
```

Independence Working Correlation

```r
# independence
gee_ind = gee_fit <- gee(cd4 ~ binage, id, data = dat1, family = gaussian,
    corstr = "independence")
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
```

```
## running glm to get initial regression estimate
```

```
## (Intercept)      binage
##    225.9020     24.2404
```

```r
# Make easier to read summary
ss1 <- data.frame(summary(gee_fit)$coefficients)
ss1 = data.frame(ss1, pvalue = 2 * (1 - pnorm(abs(ss1[, 5]))))
round(ss1, 4)
```

```
##             Estimate Naive.S.E. Naive.z Robust.S.E. Robust.z pvalue
## (Intercept) 225.9020     9.8839 22.8555     12.6001  17.9285 0.0000
## binage       24.2404    14.1947  1.7077     19.2294   1.2606 0.2075
```

Exchangeable

```
gee_exc <- gee(cd4 ~ binage, id, data = dat1, family = gaussian,
    corstr = "exchangeable")
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
```

```
## running glm to get initial regression estimate
```

```
## (Intercept)      binage
##    225.9020     24.2404
```

```
# Make easier to read summary
ss2 <- data.frame(summary(gee_exc)$coefficients)
ss2 = data.frame(ss2, pvalue = 2 * (1 - pnorm(abs(ss2[, 5]))))
round(ss2, 4)
```

```
##             Estimate Naive.S.E. Naive.z Robust.S.E. Robust.z pvalue
## (Intercept) 225.9020    13.3772 16.8870     12.6001  17.9285 0.0000
## binage       24.2404    19.2116  1.2618     19.2294   1.2606 0.2075
```

Binary time-dependent variable (balanced data)

Read in data

```
dat2 = read_csv("deeks_ex2.csv")
```

```
## Parsed with column specification:
## cols(
##   id = col_integer(),
##   cd4 = col_integer(),
##   medvl = col_integer()
## )
```

```
tbl_df(dat2)
```

```
## # A tibble: 142 x 3
##       id   cd4 medvl
##    <int> <int> <int>
##  1    16   449     0
##  2    16   226     1
##  3    18   294     0
##  4    18   138     1
##  5    21   132     1
##  6    21   132     0
##  7    26   324     1
##  8    26   500     0
##  9    30   216     1
## 10    30   254     0
## # ... with 132 more rows
```

Independence Working Correlation

```r
# independence
gee_ind <- gee(cd4 ~ medvl, id, data = dat2, family = gaussian,
    corstr = "independence")
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
```

```
## running glm to get initial regression estimate
```

```
## (Intercept)       medvl
##   310.66197   -48.28169
```

```r
# Make easier to read summary
ss1 <- data.frame(summary(gee_ind)$coefficients)
ss1 = data.frame(ss1, pvalue = 2 * (1 - pnorm(abs(ss1[, 5]))))
round(ss1, 4)
```

```
##             Estimate Naive.S.E. Naive.z Robust.S.E. Robust.z pvalue
## (Intercept) 310.6620    21.4871 14.4581     23.5065  13.2160 0.0000
## medvl       -48.2817    30.3873 -1.5889     15.7747  -3.0607 0.0022
```

Exchangeable

```
gee_exc <- gee(cd4 ~ medvl, id, data = dat2, family = gaussian,
    corstr = "exchangeable")
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
```

```
## running glm to get initial regression estimate
```

```
## (Intercept)       medvl
##   310.66197   -48.28169
```

```
# Make easier to read summary
ss2 <- data.frame(summary(gee_exc)$coefficients)
ss2 = data.frame(ss2, pvalue = 2 * (1 - pnorm(abs(ss2[, 5]))))
round(ss2, 4)
```

```
##             Estimate Naive.S.E. Naive.z Robust.S.E. Robust.z pvalue
## (Intercept) 310.6620    21.4871 14.4581     23.5065  13.2160 0.0000
## medvl       -48.2817    15.5779 -3.0994     15.7747  -3.0607 0.0022
```

Continuous time-dependent variables (very unbalanced data)

Read in data

```
dat3 = read_csv("deeks_ex3.csv")
```

```
## Parsed with column specification:
## cols(
##   id = col_integer(),
##   cd4 = col_integer(),
##   logvlbase = col_double(),
##   logvlchange = col_double()
## )
```

```
tbl_df(dat3)
```

```
## # A tibble: 11,300 x 4
##       id   cd4 logvlbase logvlchange
##    <int> <int>     <dbl>       <dbl>
##  1     1    45      2.70        0
##  2     1   119      2.70        2.52
##  3     1   113      2.70       NA
##  4     1    74      2.70        2.50
##  5     1    95      2.70       NA
##  6     1   120      2.70       NA
##  7     1   209      2.70       NA
##  8     1   375      2.70       NA
##  9     1   137      2.70        1.17
## 10     1    NA      2.70        2.38
## # ... with 11,290 more rows
```

Independence Working Correlation

```r
# independence
gee_ind <- gee(cd4 ~ logvlbase + logvlchange, id, data = dat3,
    na.action = na.omit, family = gaussian, corstr = "independence")
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
```

```
## running glm to get initial regression estimate
```

```
## (Intercept)   logvlbase logvlchange
##   618.95554   -83.74371   -99.19400
```

```r
# Make easier to read summary
ss1 <- data.frame(summary(gee_ind)$coefficients)
ss1 = data.frame(ss1, pvalue = 2 * (1 - pnorm(abs(ss1[, 5]))))
round(ss1, 4)
```

```
##             Estimate Naive.S.E.  Naive.z Robust.S.E. Robust.z pvalue
## (Intercept) 618.9555    11.6185  53.2735     35.1552  17.6064      0
## logvlbase   -83.7437     2.9610 -28.2819      8.2867 -10.1058      0
## logvlchange -99.1940     2.4536 -40.4284      6.8227 -14.5389      0
```

Exchangeable

```
gee_exc <- gee(cd4 ~ logvlbase + logvlchange, id, data = dat3,
    na.action = na.omit, family = gaussian, corstr = "exchangeable")
```

```
## Beginning Cgee S-function, @(#) geeformula.q 4.13 98/01/27
```

```
## running glm to get initial regression estimate
```

```
## (Intercept)   logvlbase logvlchange
##   618.95554   -83.74371   -99.19400
```

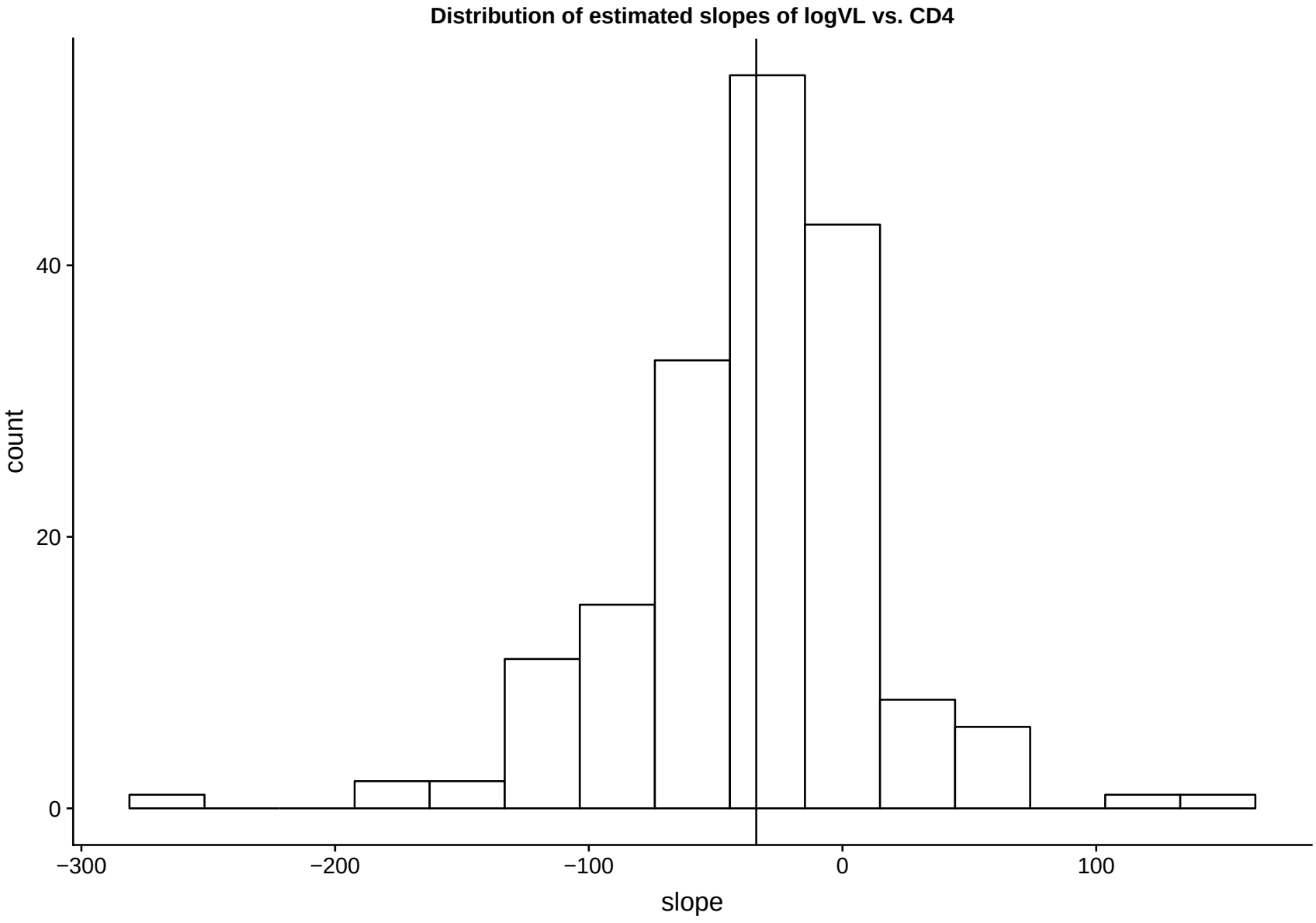
```
# Make easier to read summary
ss2 <- data.frame(summary(gee_exc)$coefficients)
ss2 = data.frame(ss2, pvalue = 2 * (1 - pnorm(abs(ss2[, 5]))))
round(ss2, 4)
```

```
##             Estimate Naive.S.E.  Naive.z Robust.S.E. Robust.z pvalue
## (Intercept) 509.1237    31.2343  16.3002     32.9122  15.4692      0
## logvlbase   -52.7572     7.4033  -7.1262      7.7138  -6.8393      0
## logvlchange -54.7519     2.1737 -25.1888      3.1545 -17.3566      0
```

Why large differences in estimates returned by same linear model, but different working correlation models

- Exchangeable will tends to give more influence to subjects with fewer observations relative to independence working correlation.
- Below, we use code from chapter 2 using tge data.table package to estimate the slope of CD4 versus logVL by id.
- First, make a variable that is the log(VL), which is simply baseline value plus change.
- We look at the subset of id's that have at least 10 repeated observations.
- We then plot the distribution of slopes.

Histogram of estimated slopes



Results

- See that the mean of the slopes is close to that found in the exchangeable gee analysis above.
- Implies that the exchangeable working correlation gives something closer to equal weights to individuals than the independence. -Independence gives more weight to individuals with more observations.