Longitudinal Analysis for final project

Shih-Ni Prim

11/11/2020

Contents

3. **How does the binding strength of the antibodies develop in response to the number of vaccine dosages by treatment?** This will be evaluated with a longitudinal analysis by test subject.

Data Analysis

Longitudinal Data Analysis

[This section probably needs some more revisions, since we still have three more lectures.]

First we don't consider treatments but only plot the mean trend over time. The plot shows that binding does vary over time. The red line shows the mean trend over time, and the blue line shows the variance over time. The variance does not seem equal over time, so we use unequal variance over time for the covariance structure.

$$Y_{ij} = \beta_0 + \beta_1 Time_{ij} + e_{ij}$$
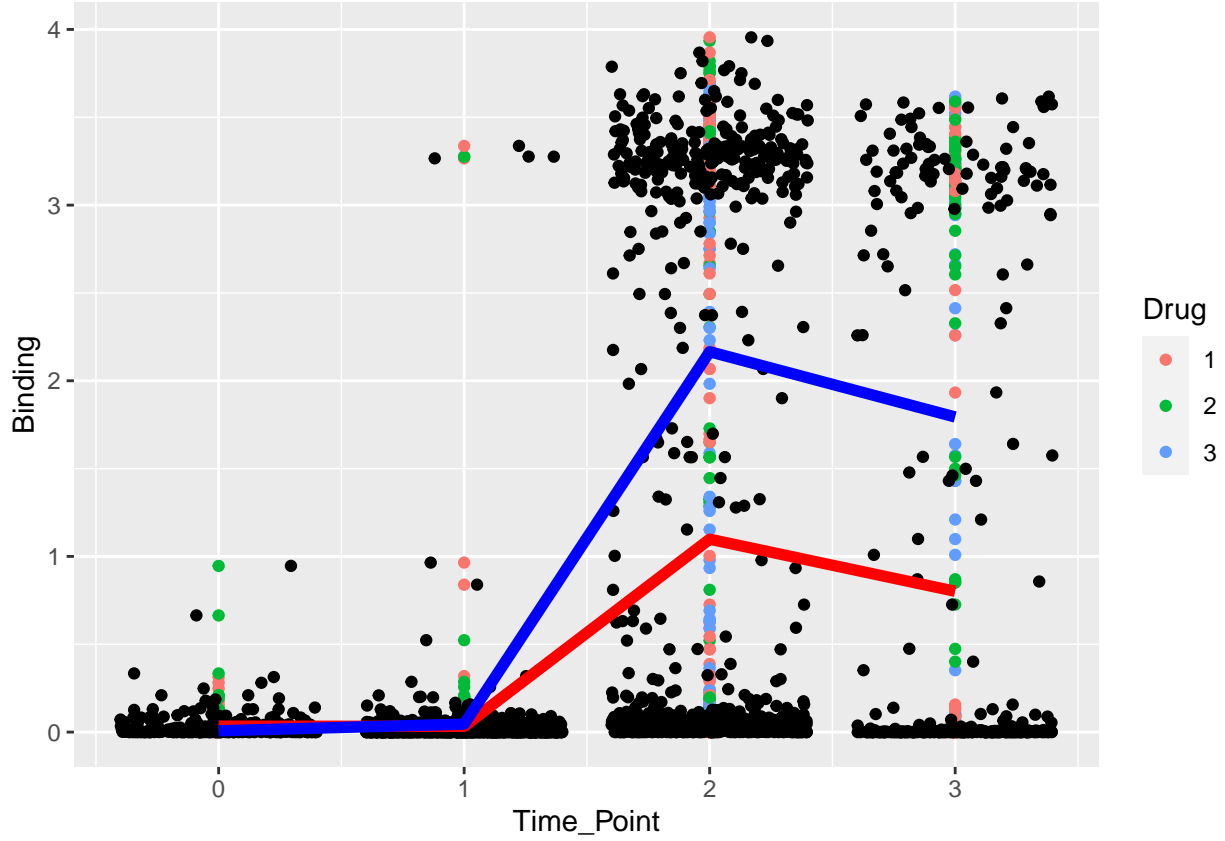
Here we use `Binding` as the response, `Time_Point` as the time factor, and `Drug` as the covariates. Random effect for both intercept and slope. Now we want to add one covariate: `Drug`. We use two indicator variables: `D1` and `D2`, where

$$D1 = \begin{cases} 1 & \text{if Drug} = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$D2 = \begin{cases} 1 & \text{if Drug} = 2 \\ 0 & \text{otherwise} \end{cases}$$

Assuming that the random effects are the same for each drug, our full model is:

$$Y_{ij} = \beta_0 + \beta_1 Time_{ij} + D1_i(\beta_2 + \beta_3 Time_{ij}) + D2_i(\beta_4 + \beta_5 Time_{ij}) + b_{0i} + b_{1i} Time_{ij} + e_{ij}$$

$$\mathbf{b_i} \sim N\left(0, \mathbf{D} = \begin{bmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{bmatrix}\right)$$

Drug 1: $Y_{ij} = \beta_0 + \beta_1 Time_{ij} + \beta_2 + \beta_3 Time_{ij} + b_{0i} + b_{1i} Time_{ij} + e_{ij}$
Drug 2: $Y_{ij} = \beta_0 + \beta_1 Time_{ij} + \beta_4 + \beta_5 Time_{ij} + b_{0i} + b_{1i} Time_{ij} + e_{ij}$
Drug 3: $Y_{ij} = \beta_0 + \beta_1 Time_{ij} + b_{0i} + b_{1i} Time_{ij} + e_{ij}$

```
## Linear mixed-effects model fit by REML
##  Data: dataLDA
##        AIC      BIC    logLik
```

```
##    3673.651 3749.144 -1823.826
##
## Random effects:
##  Formula: ~Time_Point | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 0.6893734 (Intr)
## Time_Point  0.6524155 -0.999
## Residual    0.2169511
##
## Variance function:
##  Structure: Different standard deviations per stratum
##  Formula: ~1 | Time_Point
##  Parameter estimates:
##         1         0         2         3
## 1.0000000 0.3827704 7.1563614 6.5754859
## Fixed effects: binding ~ Time_Point + D1 + D1:Time_Point + D2 + D2:Time_Point
##                   Value Std.Error   DF    t-value p-value
## (Intercept)   -0.0432994 0.3982221 2441 -0.1087318  0.9134
## Time_Point     0.1772970 0.3772249 2441  0.4700034  0.6384
## D1            -0.3162408 0.5043030   17 -0.6270850  0.5389
## D2            -0.8725346 0.5123761   17 -1.7029181  0.1068
## Time_Point:D1  0.2407466 0.4811398 2441  0.5003672  0.6169
## Time_Point:D2  0.7867662 0.4891206 2441  1.6085323  0.1078
##  Correlation:
##               (Intr) Tm_Pnt D1     D2     T_P:D1
## Time_Point    -0.998
## D1            -0.790  0.788
## D2            -0.777  0.775  0.614
## Time_Point:D1  0.782 -0.784 -0.998 -0.608
## Time_Point:D2  0.769 -0.771 -0.608 -0.998  0.605
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -1.18636862 -0.29361355 -0.10681754  0.02325697 15.04537744
##
## Number of Observations: 2464
## Number of Groups: 20
```

The p-values for `Drug` and the interaction of `Drug` and `Time_Point` are large. So we try another model with `Time_Point` as the only predictor. [This is skipping the part where we fit only main effect (not interaction) with Drug]

$$Y_{ij} = \beta_0 + \beta_1 Time_{ij} + b_{0i} + b_{1i} Time_{ij} + e_{ij}$$

$$\underbrace{\begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{im_i} \end{bmatrix}}_{\mathbf{Y_i}} = \underbrace{\begin{bmatrix} 1 & Time_{i1} \\ \vdots & \vdots & \vdots \\ 1 & Time_{im_i} \end{bmatrix}}_{\mathbf{X_i}} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} 1 & Time_{i1} \\ \vdots & \vdots \\ 1 & Time_{im_i} \end{bmatrix}}_{\mathbf{Z_i}} \underbrace{\begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix}}_{\mathbf{b_i}} + \underbrace{\begin{bmatrix} e_{i1} \\ \vdots \\ e_{im_i} \end{bmatrix}}_{\mathbf{e_i}}$$

$$\mathbf{b_i} \sim N\left(0, \mathbf{D} = \begin{bmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{bmatrix}\right)$$

$$\mathbf{e}_{ij} \sim N(0, \mathbf{R}_i = \sigma^2 I_{mi})$$

[need to consider whether time point 2 is the optimal point]

```
## Linear mixed-effects model fit by REML
##  Data: dataLDA
##       AIC      BIC    logLik
##   3661.551 3713.83 -1821.776
##
## Random effects:
##  Formula: ~Time_Point | id
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev    Corr
## (Intercept) 0.6628601 (Intr)
## Time_Point  0.6255252 -0.998
## Residual    0.2163048
##
## Variance function:
##  Structure: Different standard deviations per stratum
##  Formula: ~1 | Time_Point
##  Parameter estimates:
##        1          0          2          3
## 1.0000000 0.3842252 7.1971216 6.6106513
## Fixed effects: binding ~ Time_Point
##                 Value Std.Error   DF   t-value p-value
## (Intercept) -0.5031390 0.1871486 2443 -2.688447  0.0072
## Time_Point   0.5695081 0.1798267 2443  3.166983  0.0016
##  Correlation:
##            (Intr)
## Time_Point -0.998
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -1.15655984 -0.26620653 -0.11153392  0.02881313 15.06729096
##
## Number of Observations: 2464
## Number of Groups: 20
```

This simpler model has lower AIC and BIC, as shown below. So we prefer the model with `Time_Point` as the predictor and, with the low p-values of the slope of `Time_Point`, conclude that the binding rates vary over time. In other words, the number of HIV vaccines given do affect the binding rate, but the drugs given do not have significant effects.

```
##      df      AIC df.1      BIC
## lda  13 3673.651   13 3749.144
## lda2  9 3661.551    9 3713.830
```