

Longitudinal Analysis for final project

Shih-Ni Prim

11/12/2020

Contents

1	Data Analysis	1
1.1	Longitudinal Data Analysis	1
1.2	With Drugs	5
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.	

1 Data Analysis

1.1 Longitudinal Data Analysis

As seen in Figure 1 and Figure 2, the mean trend is not linear, and the different time points have different variances. This information suggests that we should use piecewise linear models and set variances as unequal over time.

[Figure 1 about here.]

[Figure 2 about here.]

We first consider a model with time point as the only covariate:

$$Y_{ij} = \beta_0 + \beta_1 Time_{ij} + e_{ij}$$

Thus we will use a piecewise linear model, in which each segment has different intercepts and slopes. We use three indicator variables: $S1, S2, S3$ as the indicator variables, where

$$S1 = \begin{cases} 1 & \text{if } 0 \leq \text{Timepoint} < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$S2 = \begin{cases} 1 & \text{if } 1 \leq \text{Timepoint} < 2 \\ 0 & \text{otherwise} \end{cases}$$

$$S3 = \begin{cases} 1 & \text{if Timepoint} \geq 2 \\ 0 & \text{otherwise} \end{cases}$$

The new model is thus

$$Y_{ij} = S1(\beta_0 + \beta_1 Time_{ij}) + S2(\beta_2 + \beta_3 Time_{ij}) + S3(\beta_4 + \beta_5 Time_{ij}) + e_{ij}$$

We also want to make sure that the trend is continuous at timepoint = 1 and 2.

Our final model is $Y_{ij} = \beta_0(S1 + 2S2 - S2Time_{ij}) + \beta_1(S1Time_{ij} + 2S2 - S2Time_{ij}) + \beta_4(-S2 + S2Time_{ij} + S3) + \beta_5(-2S2 + 2S2Time_{ij} + S3Time_{ij}) + e_{ij}$ where

$$\mathbf{e}_i \sim N(0, \sigma^2 I)$$

```
## Generalized least squares fit by REML
## Model: meanform
## Data: dataLDA1
##      AIC      BIC    logLik
## 3323.05 3375.322 -1652.525
##
## Correlation Structure: Compound symmetry
## Formula: ~1 | id
## Parameter estimate(s):
##      Rho
## 0.05863157
## Variance function:
## Structure: Different standard deviations per stratum
## Formula: ~1 | time
## Parameter estimates:
##      1      0      2      3
## 1.0000000 0.3974633 6.4383596 6.2783005
##
## Coefficients:
##                                     Value Std.Error   t-value p-
value
## I(S1 + 2 * S2 - S2:time)          -0.2221651 0.01817456 -12.223964    0
## I(S1:time + 2 * S2 - S2:time)       0.2432183 0.02461193   9.882128    0
## I(-S2 + S2:time + S3)              0.7699600 0.07779948   9.896725    0
## I(-2 * S2 + 2 * S2:time + S3:time) -0.2432756 0.02462066  -9.880955    0
##
## Correlation:
##                                     I(S1+2*S2-S I(S1:+2*S-S I(+S+S
## I(S1:time + 2 * S2 - S2:time)      -0.801
## I(-S2 + S2:time + S3)              -0.676      0.961
## I(-2 * S2 + 2 * S2:time + S3:time)  0.656      -0.946      -0.995
##
## Standardized residuals:
```

```
##           Min           Q1           Med           Q3           Max
## -0.75325057 -0.69183273 -0.09644641  0.10970274 15.19092093
##
## Residual standard error: 0.2182886
## Degrees of freedom: 2464 total; 2460 residual
```

```
S1: -0.2221651 + 0.2432183 * time
S2: (2*-0.2221651+2*0.2432183-0.7699600+2*0.2432756)+(0.2221651-0.2432183+
0.7699600 - 2 * 0.2432756) * time = -0.2413024 + 0.2623556 * time
S3: 0.7699600 - 0.2432756 * time
```

[Figure 3 about here.]

Next we check whether adding random effects improve the model. We assume that random effects exist in the intercept and slope. Our linear mixed model is then: $Y_{ij} = \beta_0(S1 + 2S2 - S2Time_{ij}) + \beta_1(S1Time_{ij} + 2S2 - S2Time_{ij}) + \beta_4(-S2 + S2Time_{ij} + S3) + \beta_5(-2S2 + 2S2Time_{ij} + S3Time_{ij}) + b_{0i} + b1i + e_{ij}$ where

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

and

$$\mathbf{e}_i \sim N(0, \sigma^2 I)$$

```
## Linear mixed-effects model fit by REML
## Data: dataLDA1
##      AIC      BIC    logLik
## 3243.345 3313.04 -1609.673
##
## Random effects:
## Formula: ~time | id
## Structure: General positive-definite, Log-Cholesky parametrization
##           StdDev      Corr
## (Intercept) 0.05079595 (Intr)
## time         0.08412997 0.63
## Residual    0.23801862
##
## Correlation Structure: Compound symmetry
## Formula: ~1 | id
## Parameter estimate(s):
##      Rho
## 0.2269978
## Variance function:
## Structure: Different standard deviations per stratum
## Formula: ~1 | time
## Parameter estimates:
```

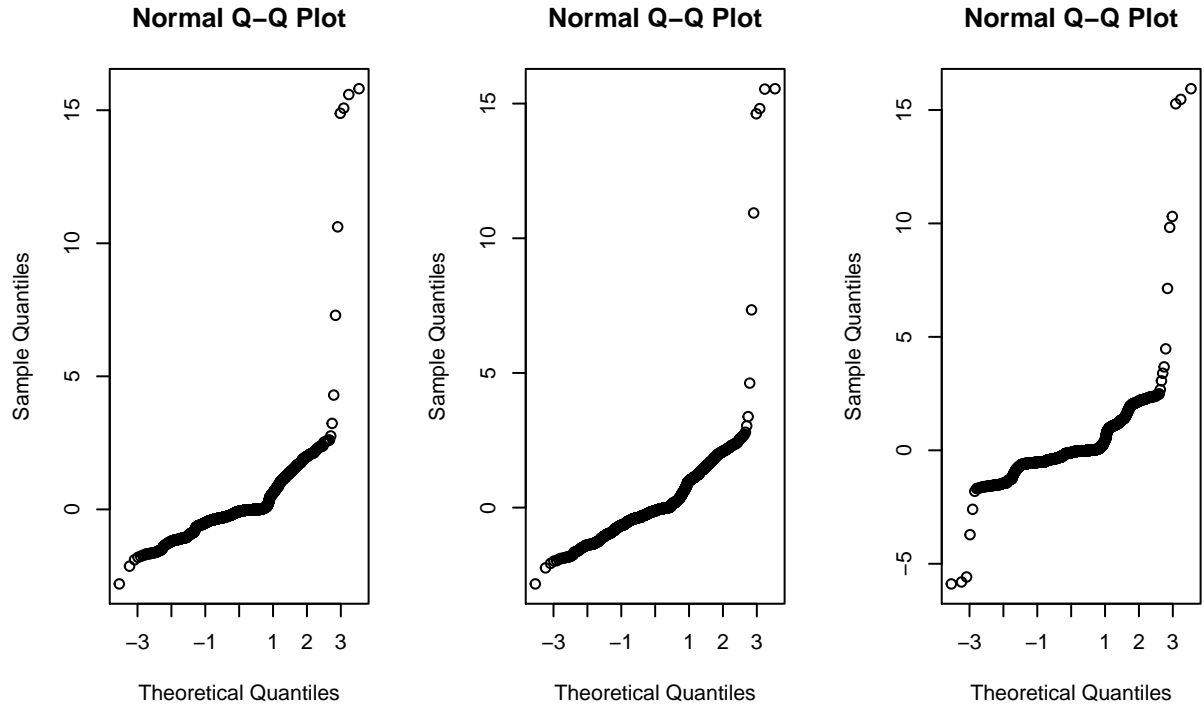
```

##          1          0          2          3
## 1.000000 0.384423 6.213855 6.203545
## Fixed effects: list(meanform)
##
##                               Value   Std.Error   DF   t-
value p-value
## I(S1 + 2 * S2 - S2:time)      -0.2181056 0.03230299 2441 -6.751869    0
## I(S1:time + 2 * S2 - S2:time)  0.2372496 0.04979896 2441  4.764148    0
## I(-S2 + S2:time + S3)         0.7376267 0.15968906 2441  4.619143    0
## I(-2 * S2 + 2 * S2:time + S3:time) -0.2378041 0.04980253 2441 -4.774940    0
## Correlation:
##                               I(S1+2*S2-S I(S1:+2*S-S I(+S+S
## I(S1:time + 2 * S2 - S2:time)      -0.633
## I(-S2 + S2:time + S3)              -0.558          0.987
## I(-2 * S2 + 2 * S2:time + S3:time)  0.596          -0.988          -0.995
##
## Standardized Within-Group Residuals:
##          Min          Q1          Med          Q3          Max
## -1.0881729 -0.6684333 -0.1941391  0.6437099 13.3804416
##
## Number of Observations: 2464
## Number of Groups: 20

```

	df	AIC
fit.gls	9	3323.050
fit.a1	12	3243.345
fit.a2	12	3072.122
	df	BIC
fit.gls	9	3375.322
fit.a1	12	3313.040
fit.a2	12	3141.817

We now check residuals for both models.



1.2 With Drugs

[Figure 4 about here.]

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:

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

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

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

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

$$\text{Drug 3: } Y_{ij} = \beta_0 + \beta_1 Time_{ij} + b_{0i} + b_{1i} Time_{ij} + e_{ij}$$

```

## 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]

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

$$\underbrace{\begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{im_i} \end{bmatrix}}_{\mathbf{Y}_i} = \underbrace{\begin{bmatrix} 1 & Time_{i1} & \\ \vdots & \vdots & \\ 1 & Time_{im_i} & \end{bmatrix}}_{\mathbf{X}_i} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}}_{\mathbf{\beta}} + \underbrace{\begin{bmatrix} 1 & Time_{i1} \\ \vdots & \vdots \\ 1 & Time_{im_i} \end{bmatrix}}_{\mathbf{Z}_i} \underbrace{\begin{bmatrix} b_{0i} \\ b_{1i} \end{bmatrix}}_{\mathbf{b}_i} + \underbrace{\begin{bmatrix} e_{i1} \\ \vdots \\ e_{im_i} \end{bmatrix}}_{\mathbf{e}_i}$$

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

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

[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. As seen in Table 1, blah blah...

[Table 1 about here.]

List of Figures

1	Mean trend by monkey	10
2	Variances over time by monkey	11
3	Piecewise Linear Function	12
4	Means and Variances over timepoints	13

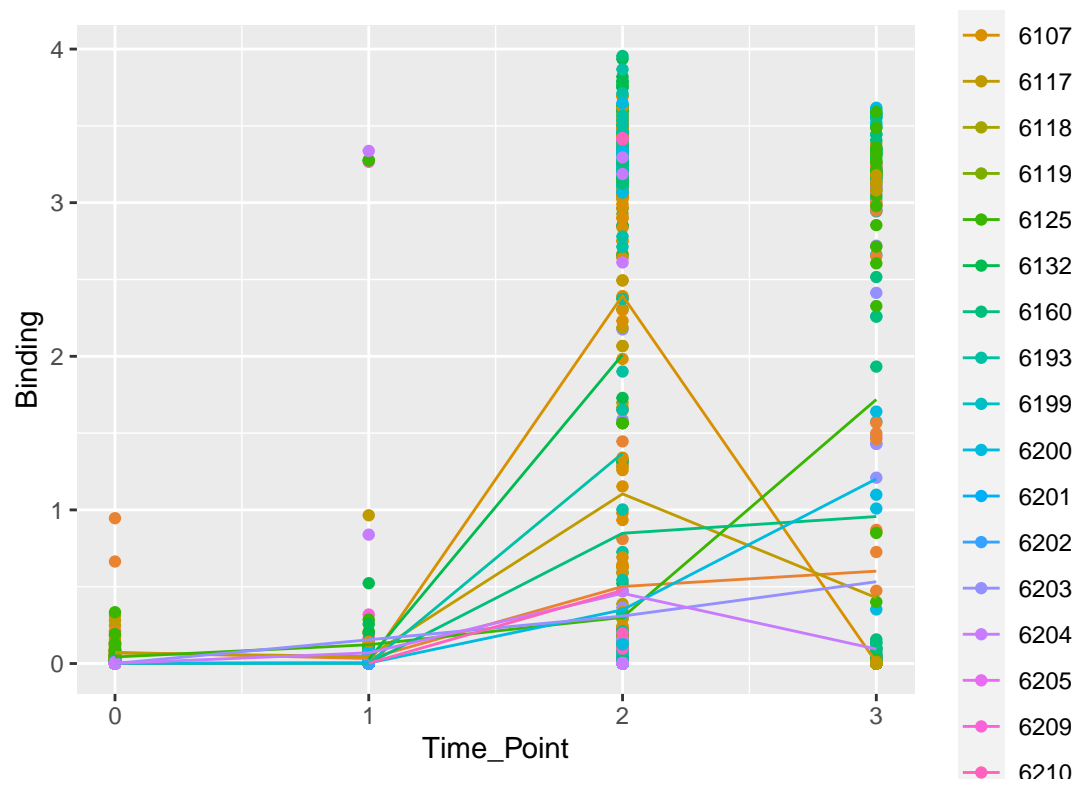


Figure 1: Mean trend by monkey

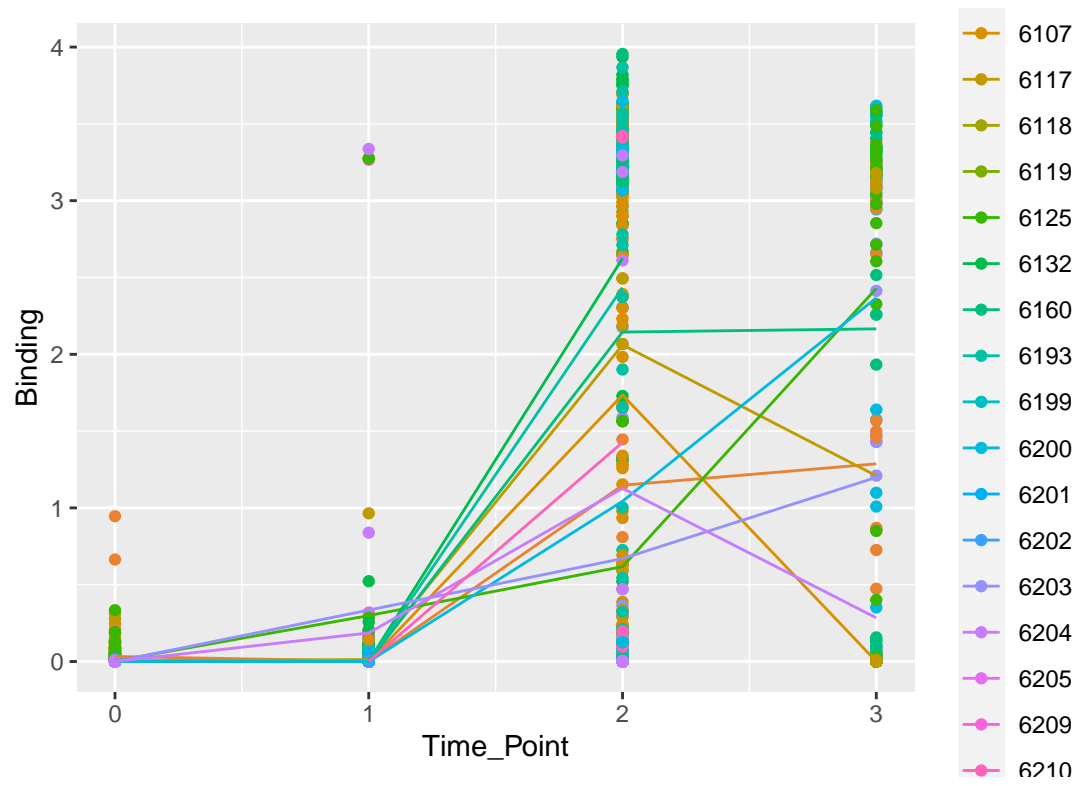


Figure 2: Variances over time by monkey

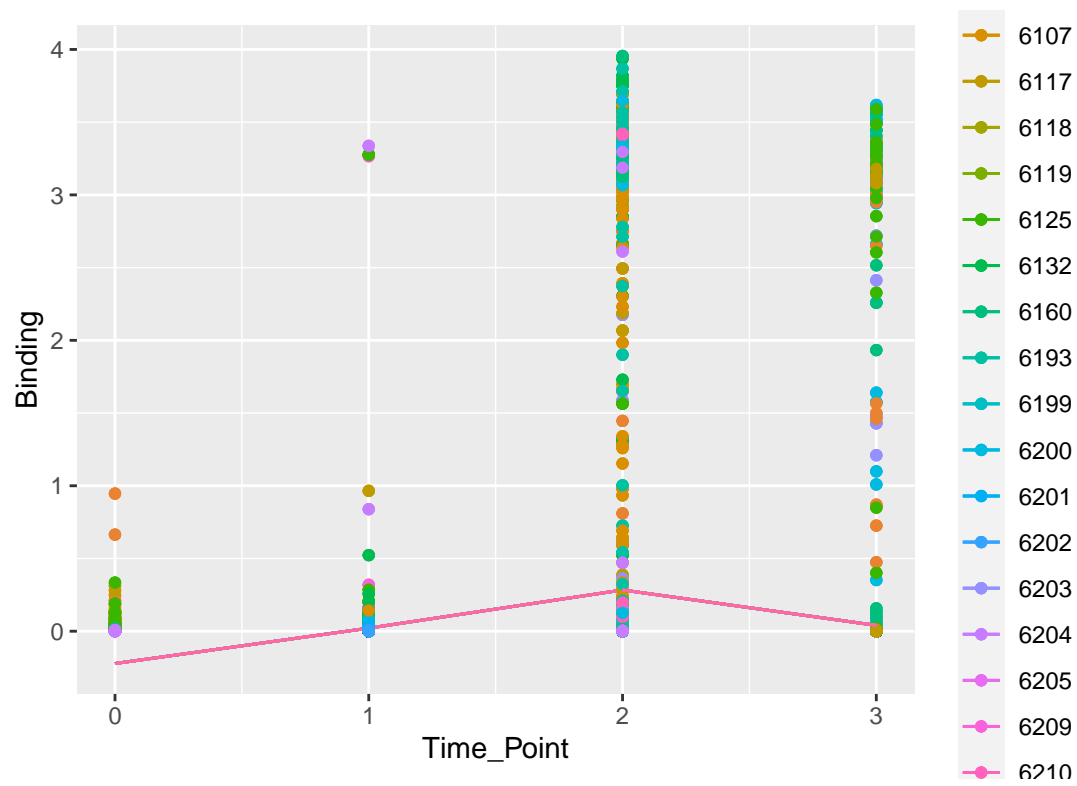


Figure 3: Piecewise Linear Function

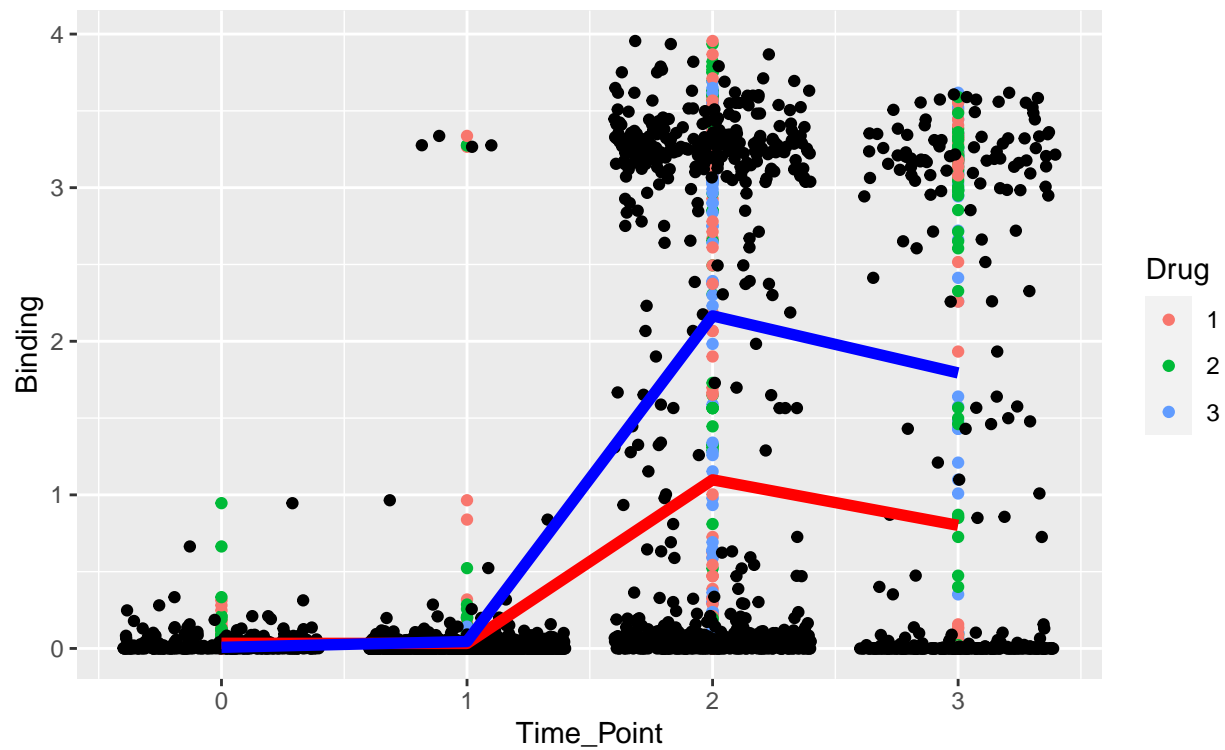


Figure 4: Means and Variances over timepoints

List of Tables

1	AIC and BIC for Longitudinal Models	15
---	---	----

Table 1: AIC and BIC for Longitudinal Models

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