

Multilevel models ("mixed effects models")

Dr Marju Kaps ([she/her](#))



m.kaps@bbk.ac.uk

Where we are

- ❖ This week (1 March), next week (8 March):
Multilevel Models
- ❖ **15 March** – new project deadline
- ❖ 15 March: Revising for the exam
 - ❖ Practice problems for Multilevel Models posted under next week
- ❖ The exam schedule should be announced soon,
Intermediate Quants exam likely takes place mid-May

This week

- Introducing linear mixed effects models – an extension of linear models
- Random effects (random slopes, random intercepts)
- Tutorial: Fitting models to longitudinal data – individuals as a grouping factor
- Next week: Model diagnostics for multilevel models, more practice

On terminology

You will often see the following terms used interchangeably:

- Multilevel models
- Mixed effects models
- Hierarchical linear models

All of these get at the idea that the data involves grouping, which can contribute to the outcome in addition to the model predictors.

The Linear Model Revisited

Linear model (regression)

```
lm(outcome ~ predictors, data = dat)
```

Fitting a slope to estimate the relationship between the predictor(s) and the outcome variable

- Each data point has a residual – its distance from the regression line, the error that isn't explained by the regression line

Linear regression assumptions

(Gelman & Hill, 2007, pp.45-46)

1. **Validity.** The data map to the research question

2. **Additivity** and **linearity**

$$y = B_0 + B_1x_1 + B_2x_2 + \dots$$

(Transforming the x s and y might help, if not)

3. **Independence** of errors

4. **Equal variance** of errors (“Homoscedasticity”)

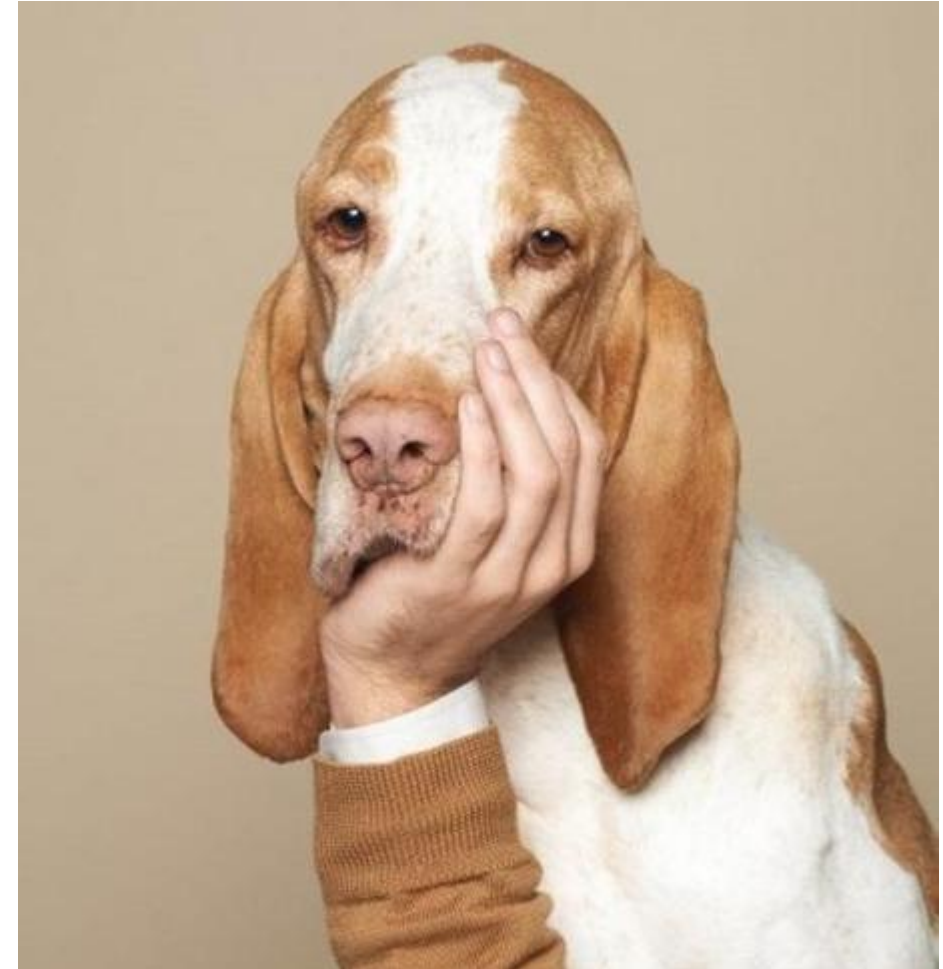
5. **Normal distribution** of errors

Some “problem cases” for simple (linear) models

- Data that have a hierarchical structure, e.g. students in different classes, in different schools, in different parts of the country
 - Students in the same class are likely to be more similar to each other than to students in a different class (or even at a different school)
- Data obtained from the same source (e.g. participant) on multiple occasions
 - Observations likely highly correlated

Some “problem cases” for simple (linear) models

Any examples from articles you have read or datasets you have looked at (e.g. for your project)?



Linear regression assumptions

(Gelman & Hill, 2007, pp.45-46)

1. **Validity.** The data map to the research question

2. **Additivity** and **linearity**

$$y = B_0 + B_1x_1 + B_2x_2 + \dots$$

(Transforming the x s and y might help, if not)

3. **Independence** of errors

4. **Equal variance** of errors (“Homoscedasticity”)

5. **Normal distribution** of errors

Linear regression assumptions

(Gelman & Hill, 2007, pp.45-46)

1. **Validity.** The data map to the research question

2. **Additivity** and **linearity**

$$y = B_0 + B_1x_1 + B_2x_2 + \dots$$

(Transforming the x s and y might help, if not)

3. **Independence** of errors

4. **Equal variance** of errors (“Homoscedasticity”)

5. **Normal distribution** of errors

When observations are not independent, residuals start to behave in funny ways. Multilevel models introduce random effects, which structure the residuals.

Linear mixed effects model

```
library(lme4)
```

```
lmer(outcome ~ predictors +  
      (predictors | group), data = dat)
```

Linear mixed effects model

```
library(lme4)
```

```
lmer(outcome ~ predictors +  
      (predictors | group), data = dat)
```

A random intercept:

(1 | group) – the intercept is allowed to vary for different values of the **group** variable

Linear mixed effects model

```
library(lme4)
```

```
lmer(outcome ~ predictors +  
      (predictors | group), data = dat)
```

A random slope:

(Var1 | group) – the slope for the **Var1** variable is allowed to vary for different values of the **group** variable

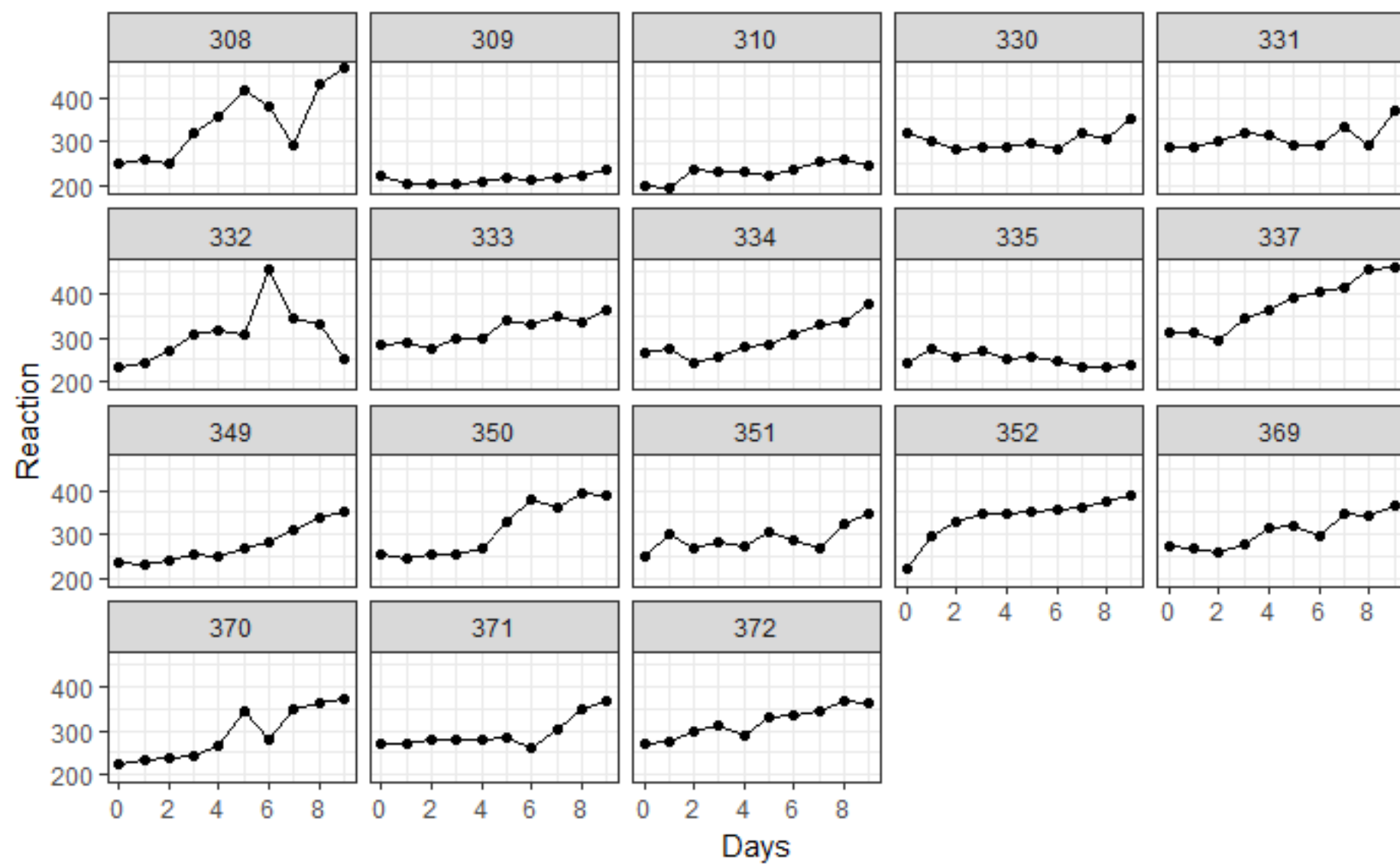
Let's get visual with an example!

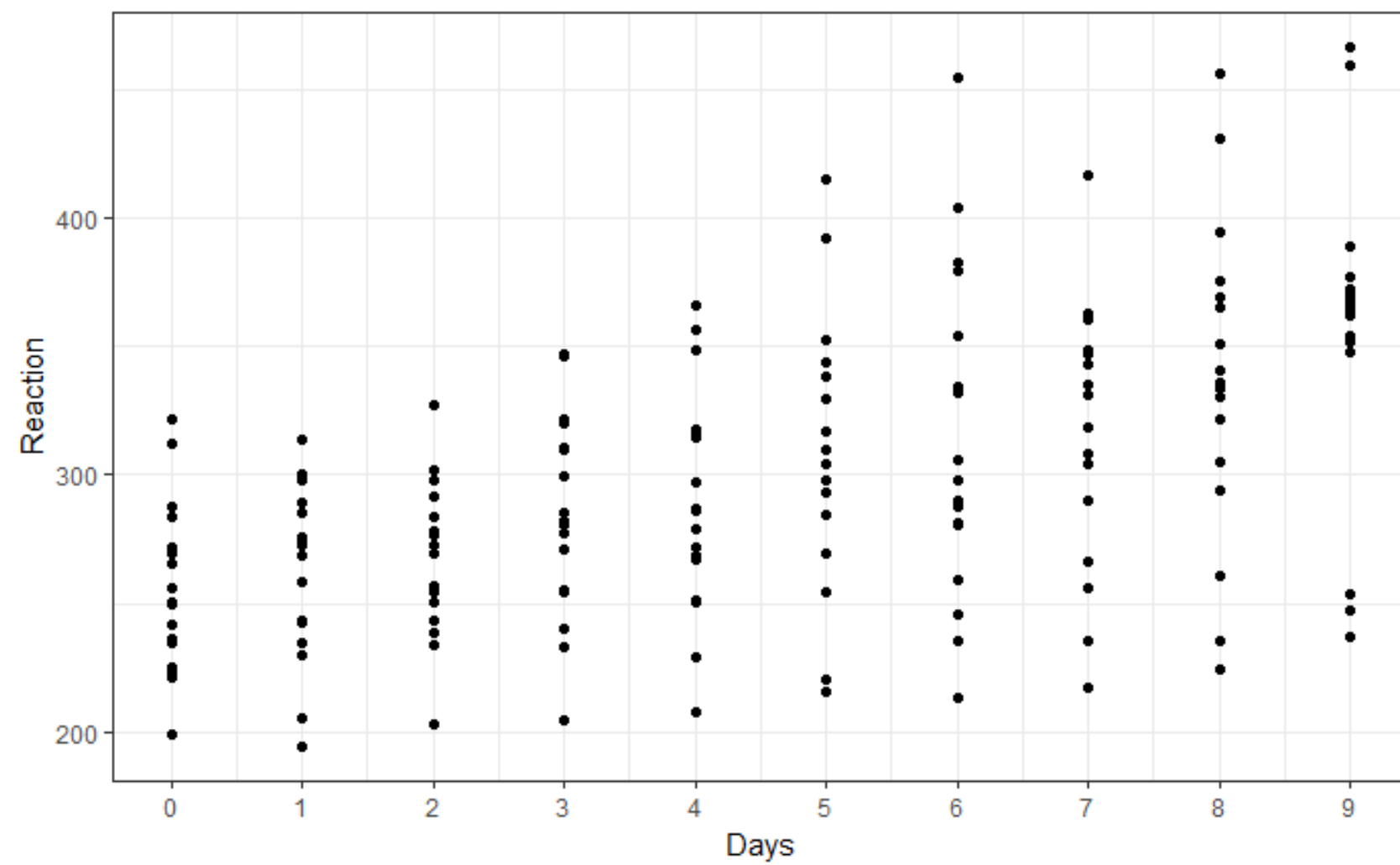
Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: a sleep dose-response study

GREGORY BELENKY, NANCY J. WESENSTEN, DAVID R. THORNE,
MARIA L. THOMAS, HELEN C. SING, DANIEL P. REDMOND,
MICHAEL B. RUSSO and THOMAS J. BALKIN

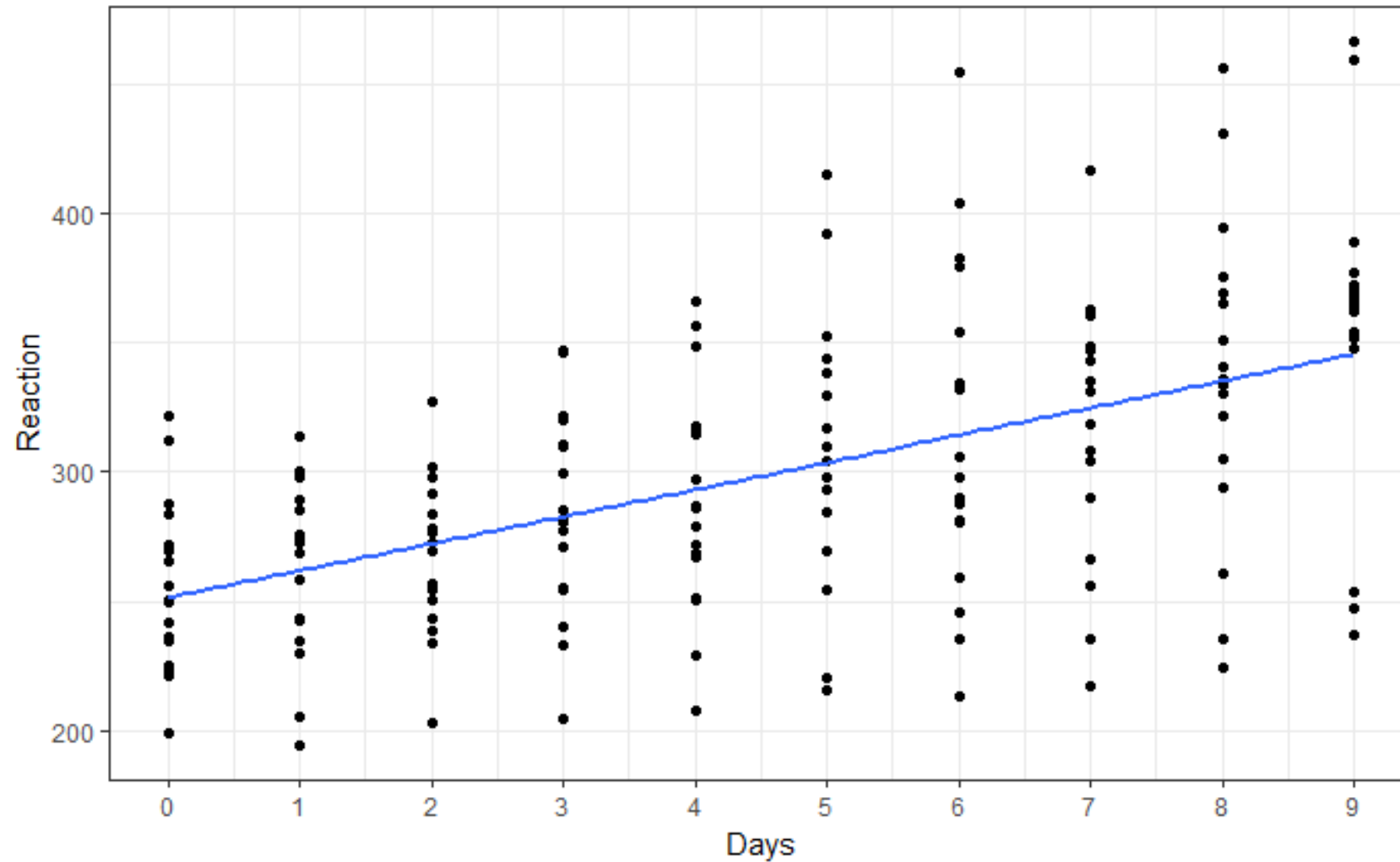
Division of Neuropsychiatry, Walter Reed Army Institute of Research, Silver Spring, MD, USA

Accepted in revised form 11 December 2002; received 28 June 2002

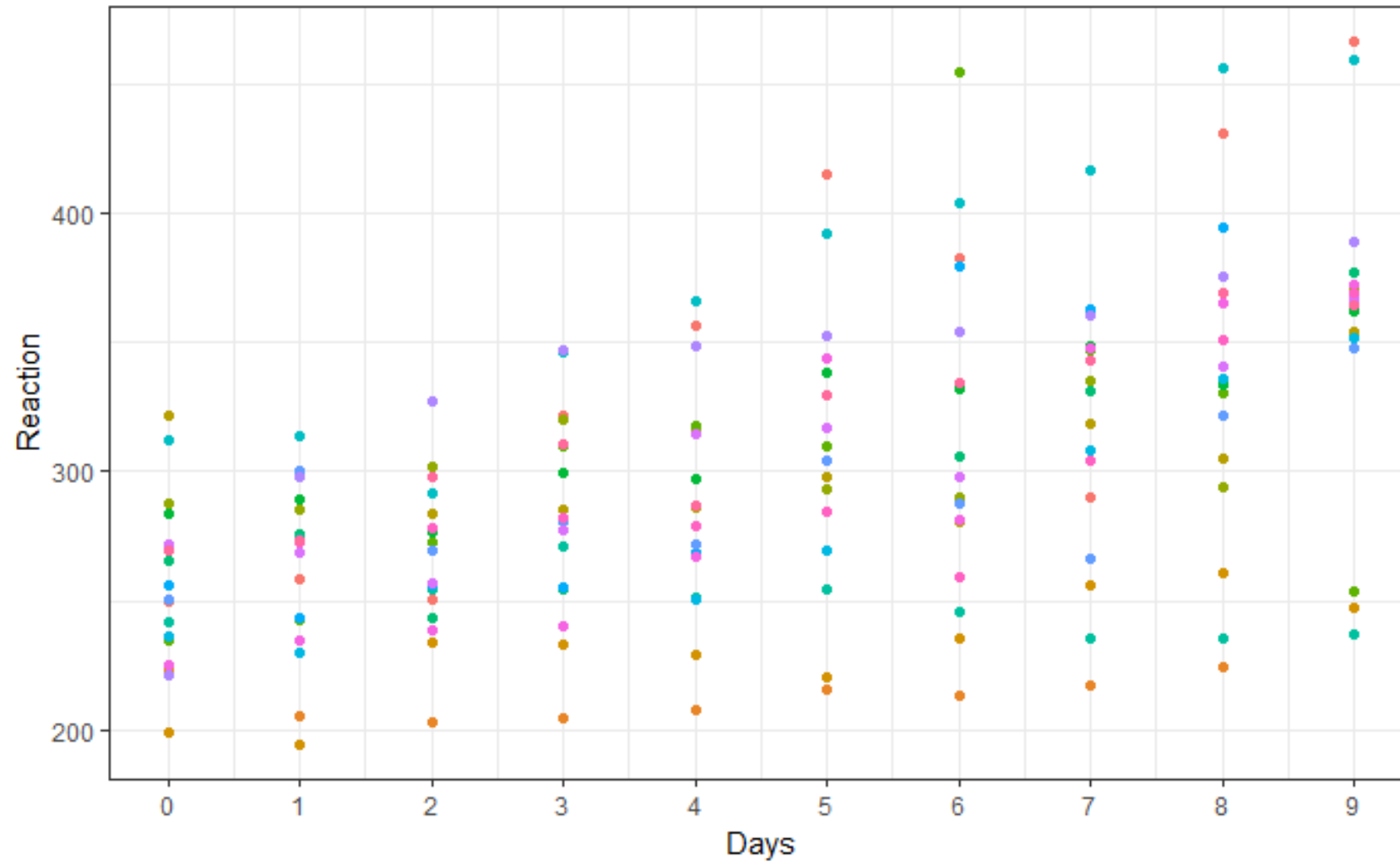


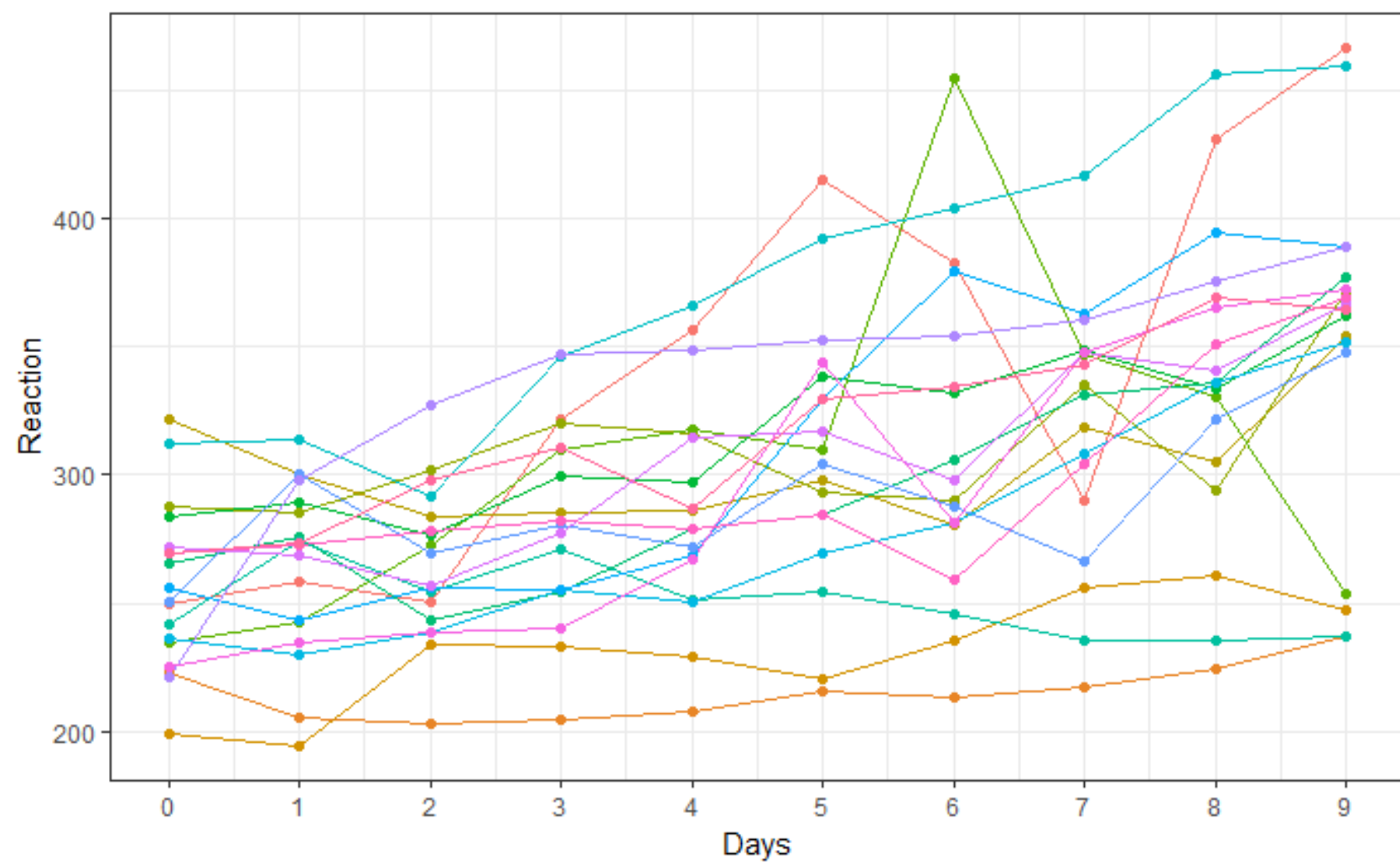


You could just use linear regression...



But that would miss the structure of the data

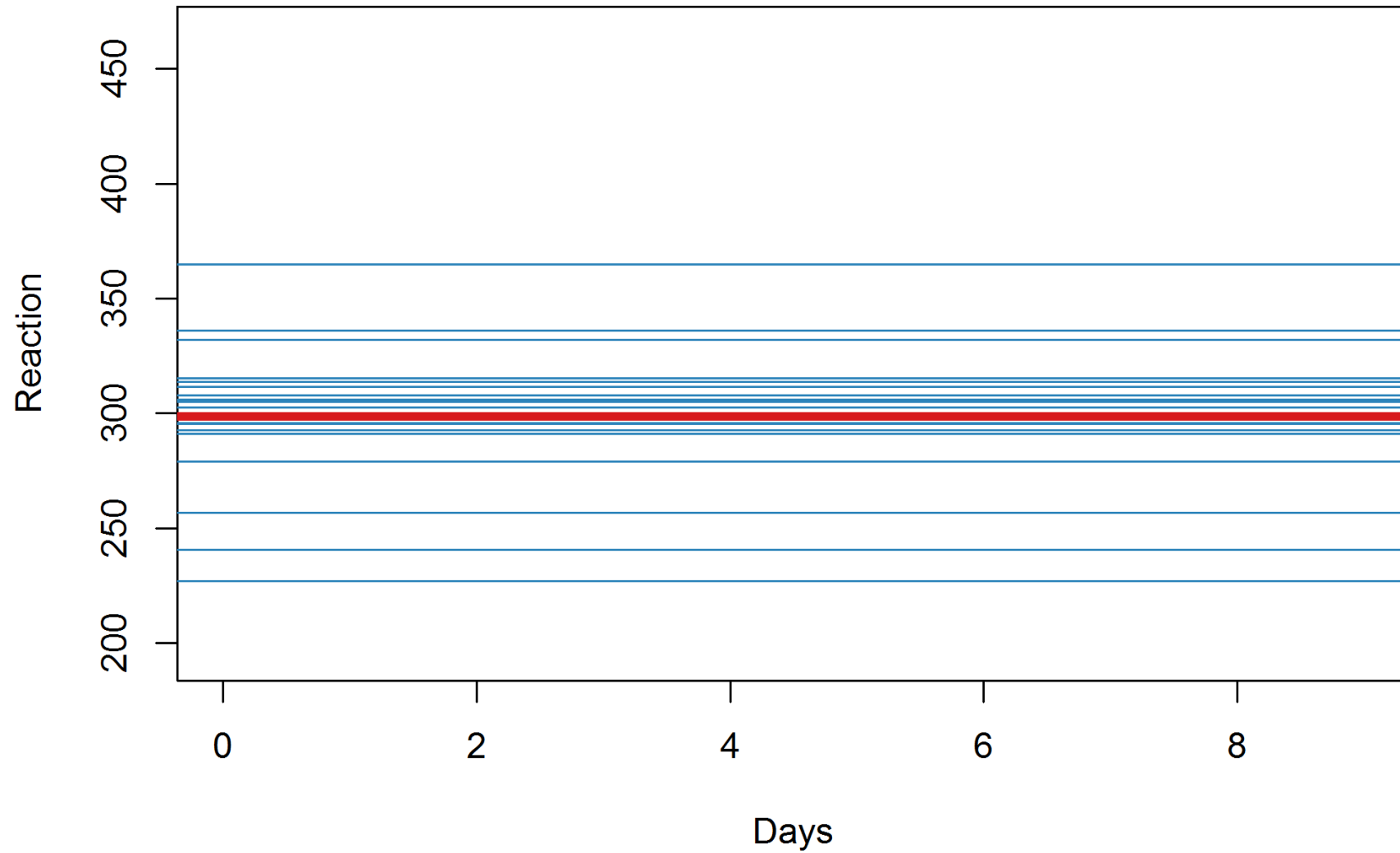


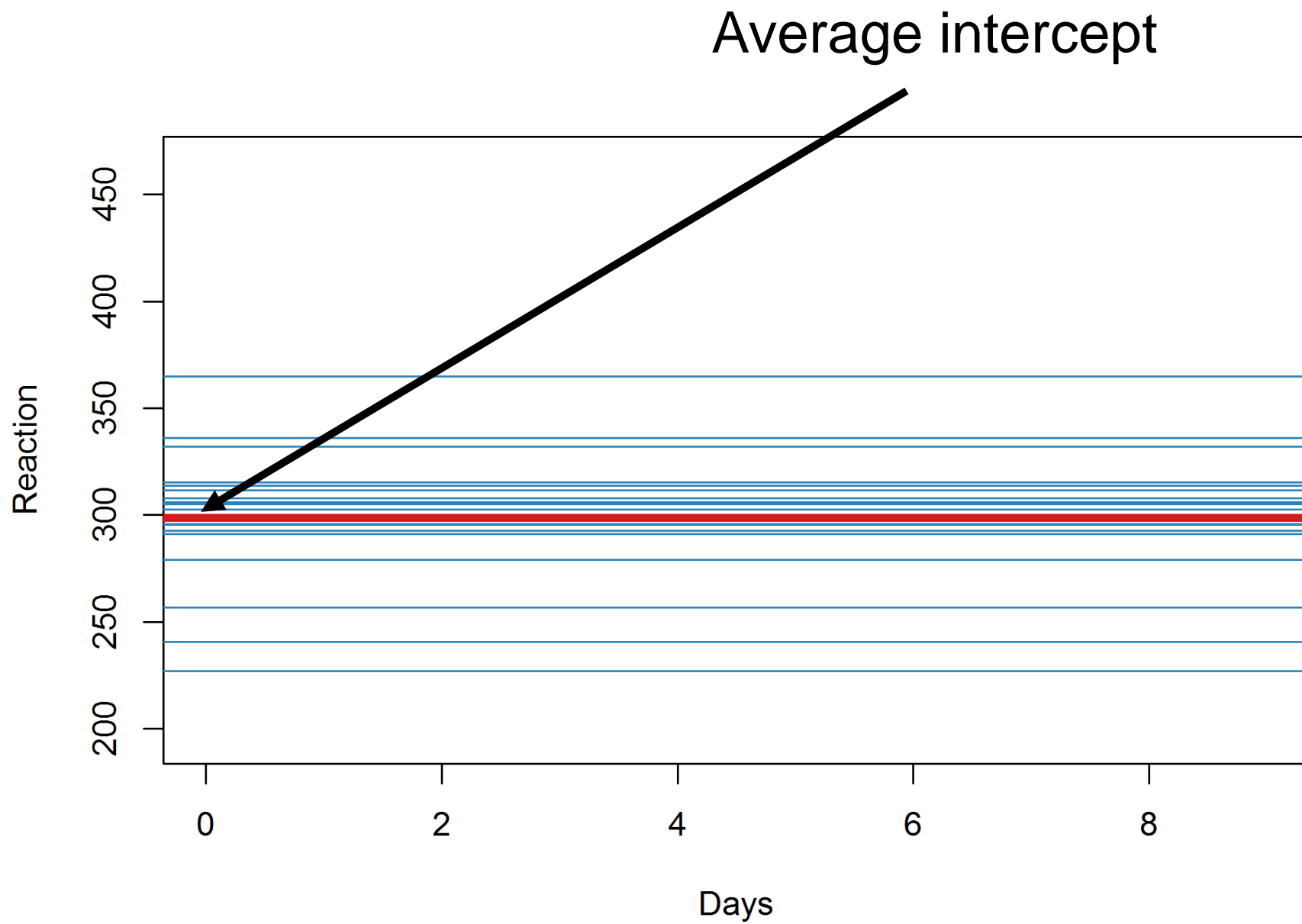


Thinking question

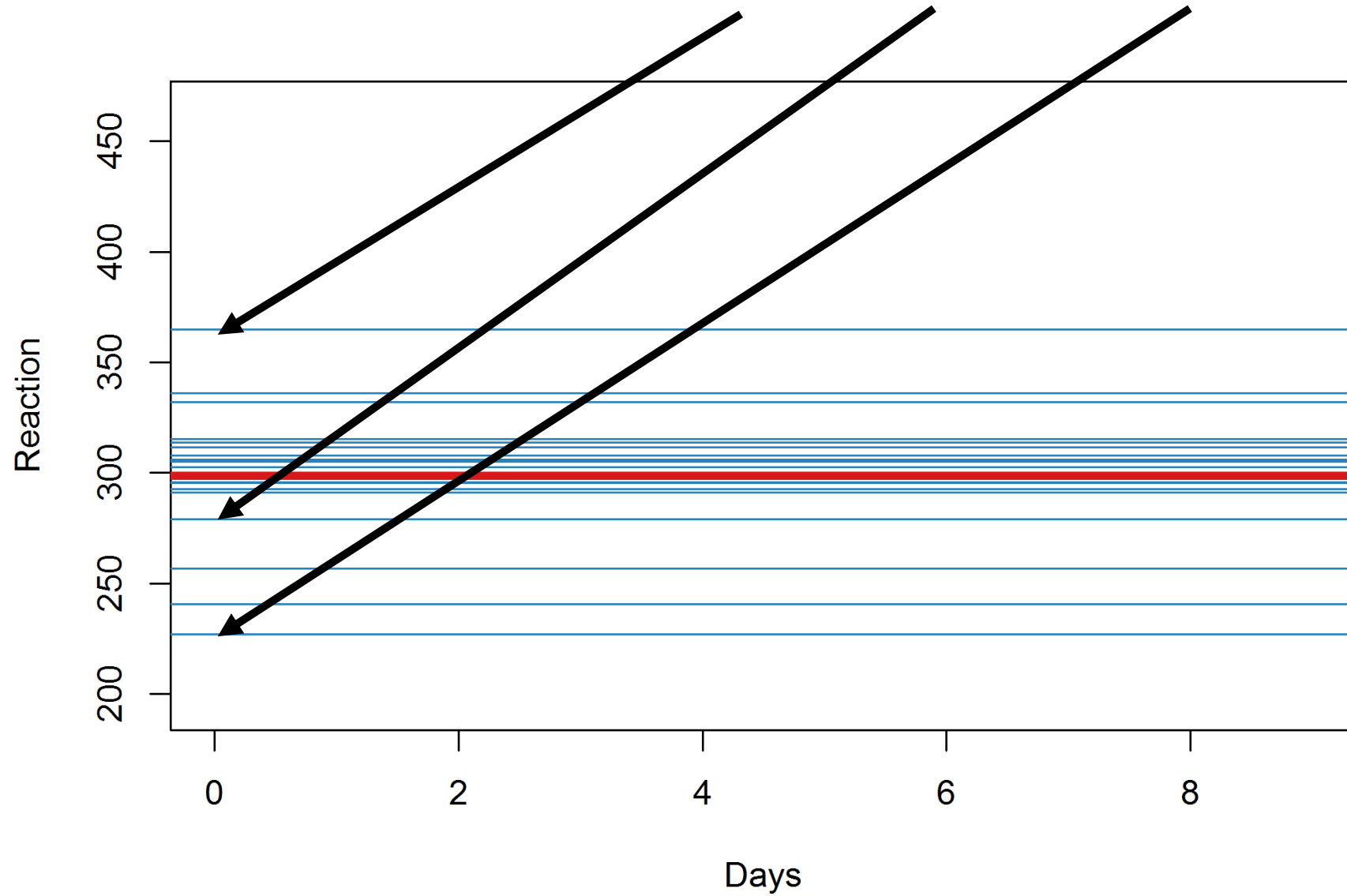
Why not just fit a separate regression model for each participant?

A baseline model

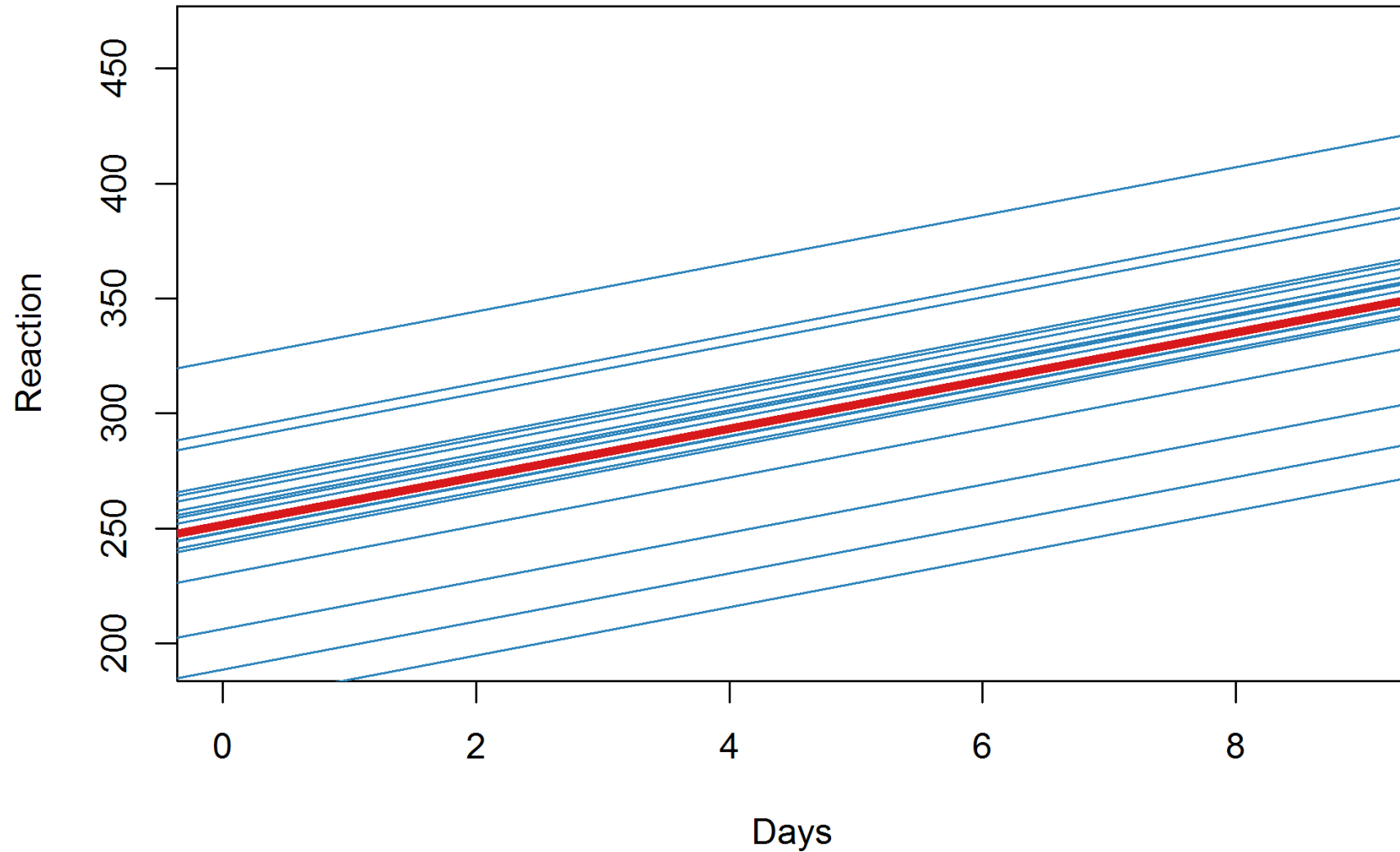




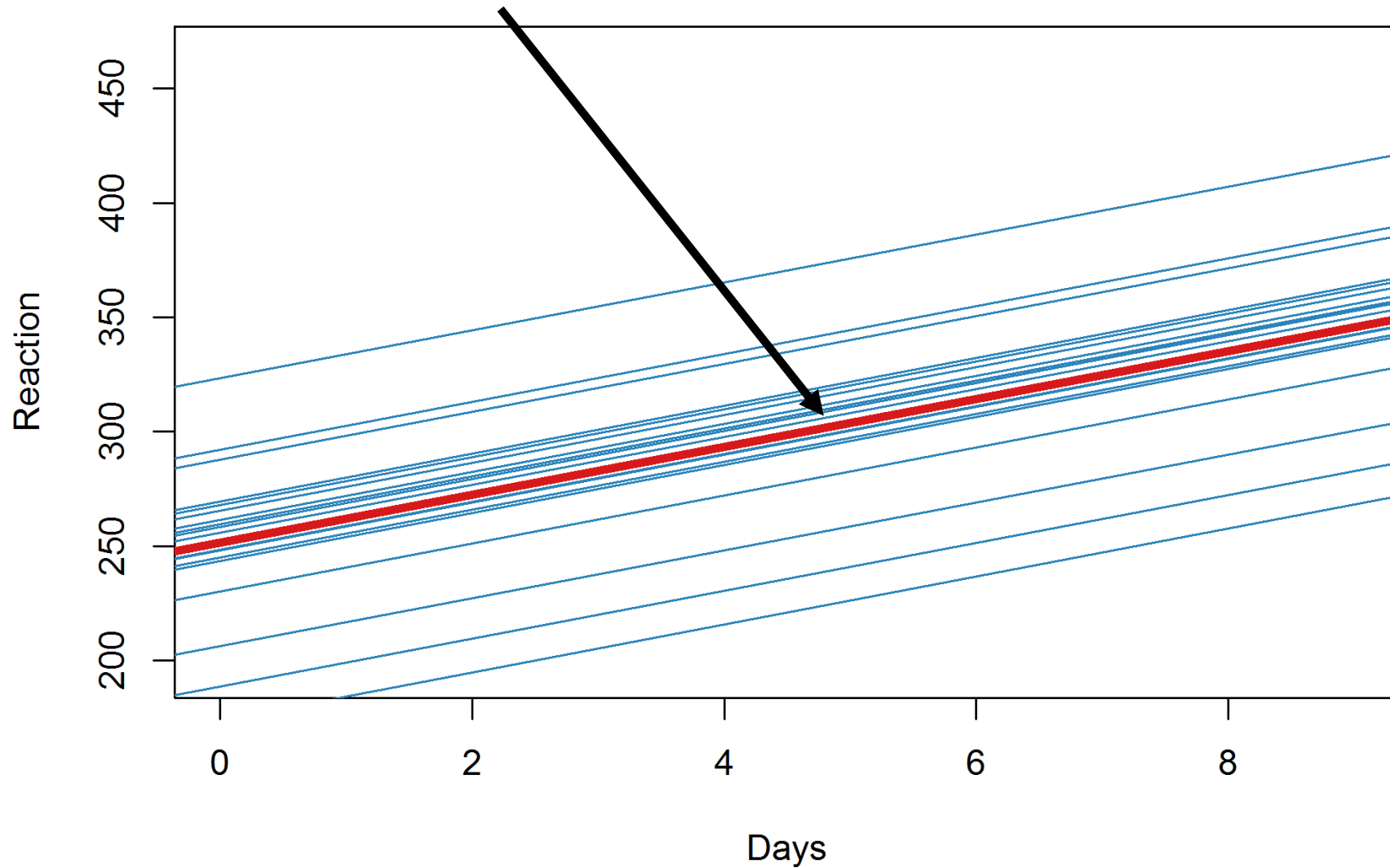
Intercept varies by participant



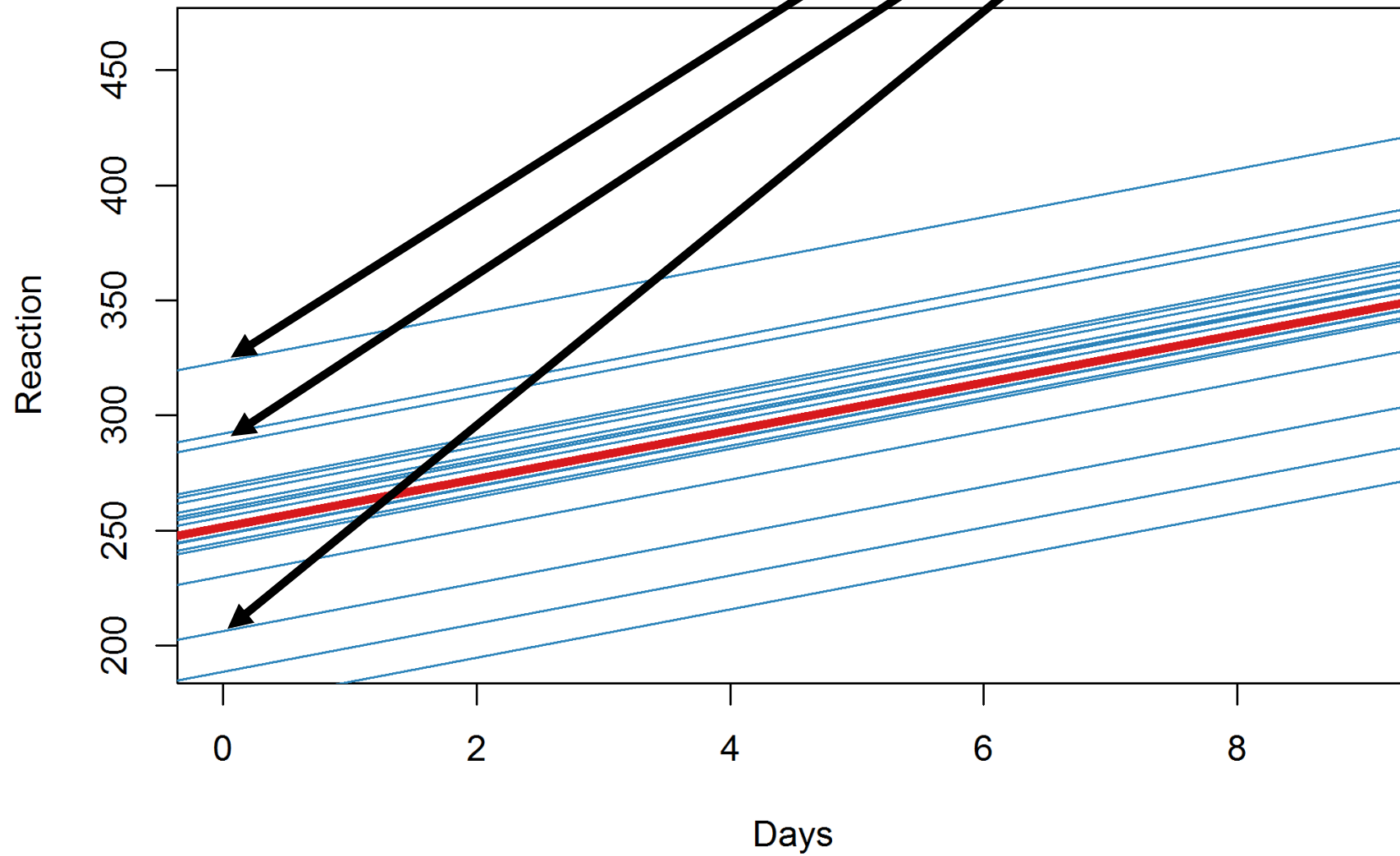
Adding a slope for Days



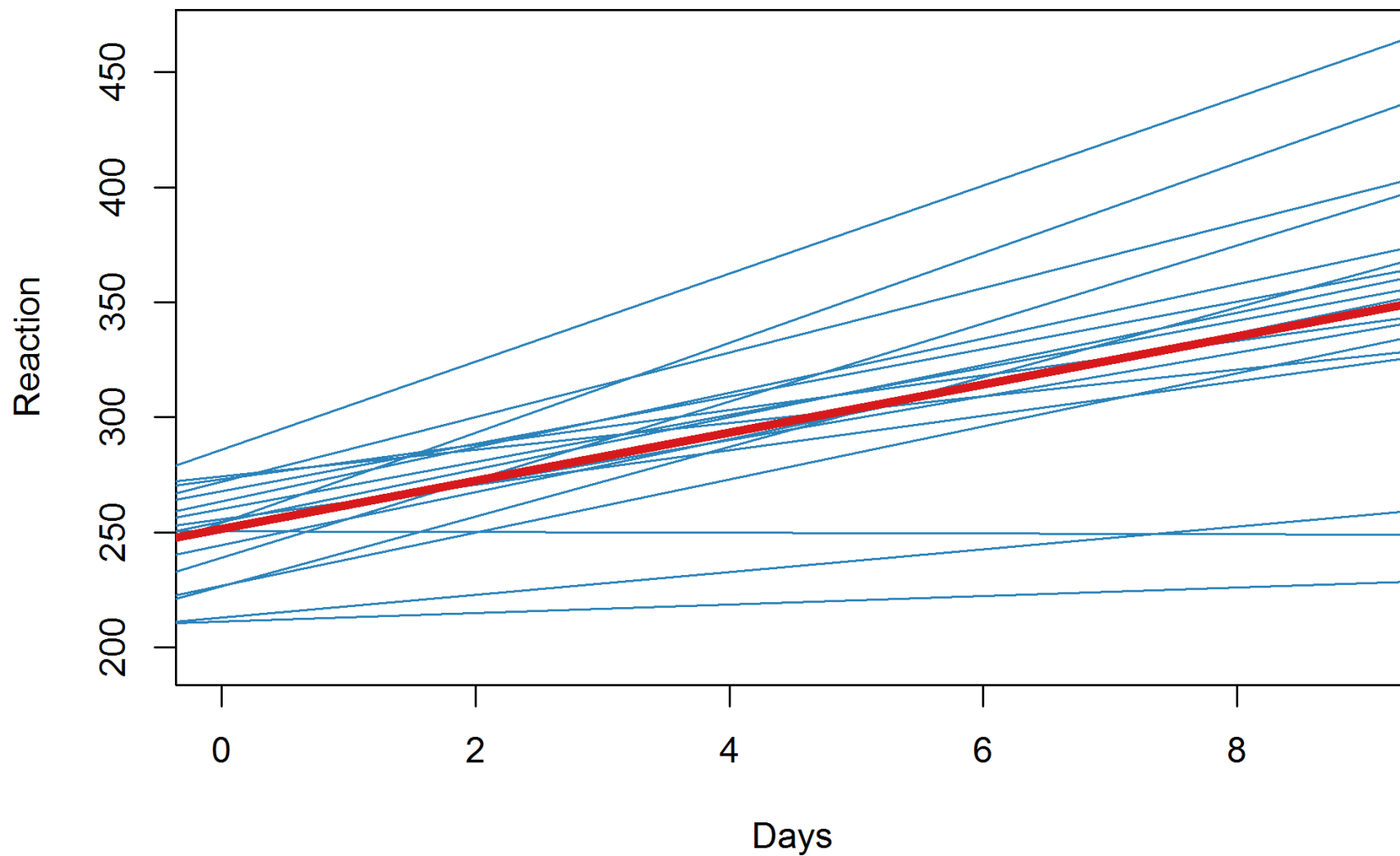
Average slope (slower reactions over time)



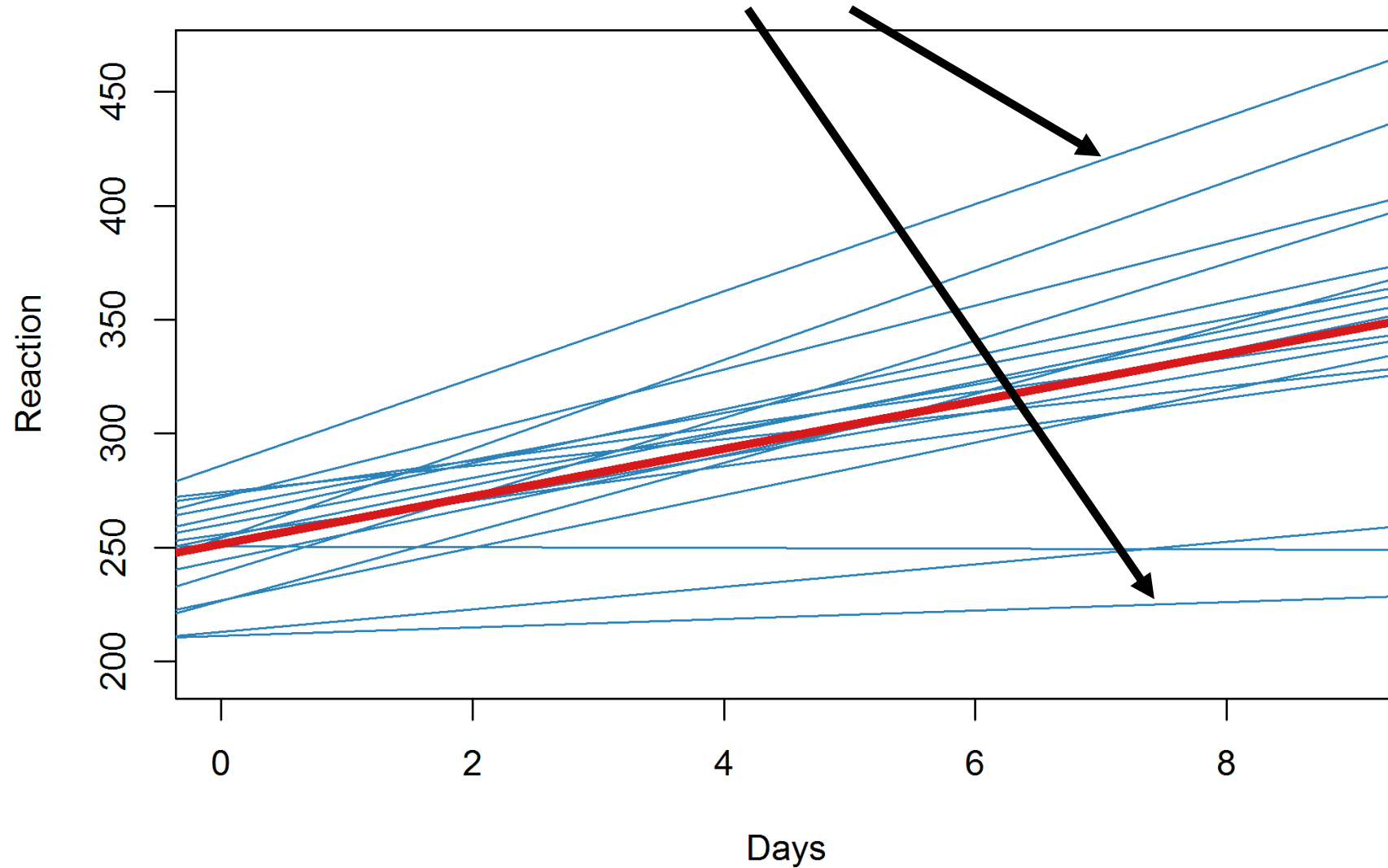
Intercept (still) varies by
participant

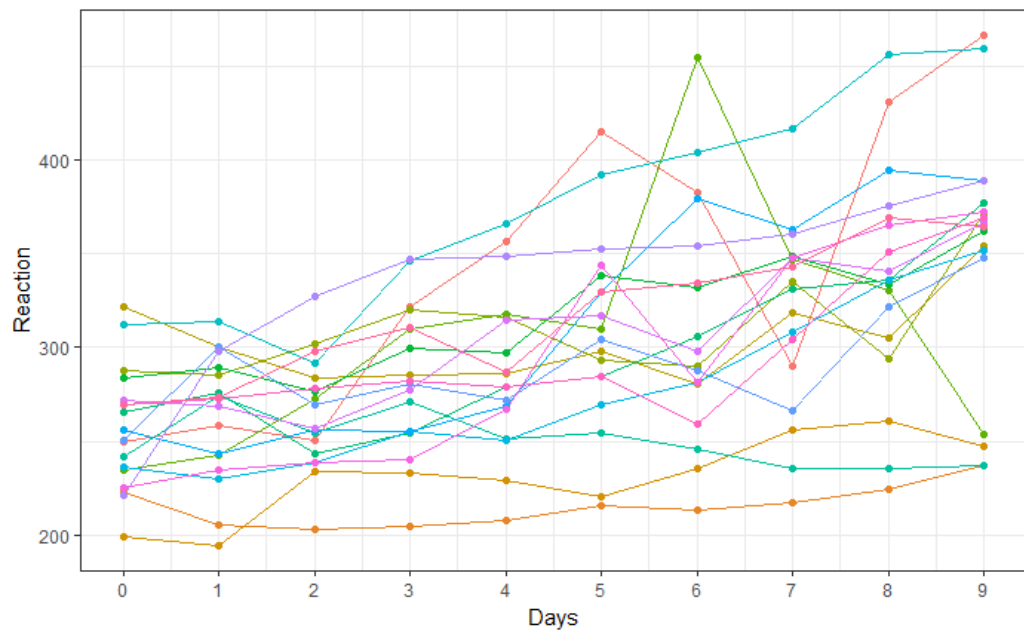


Adding random slopes

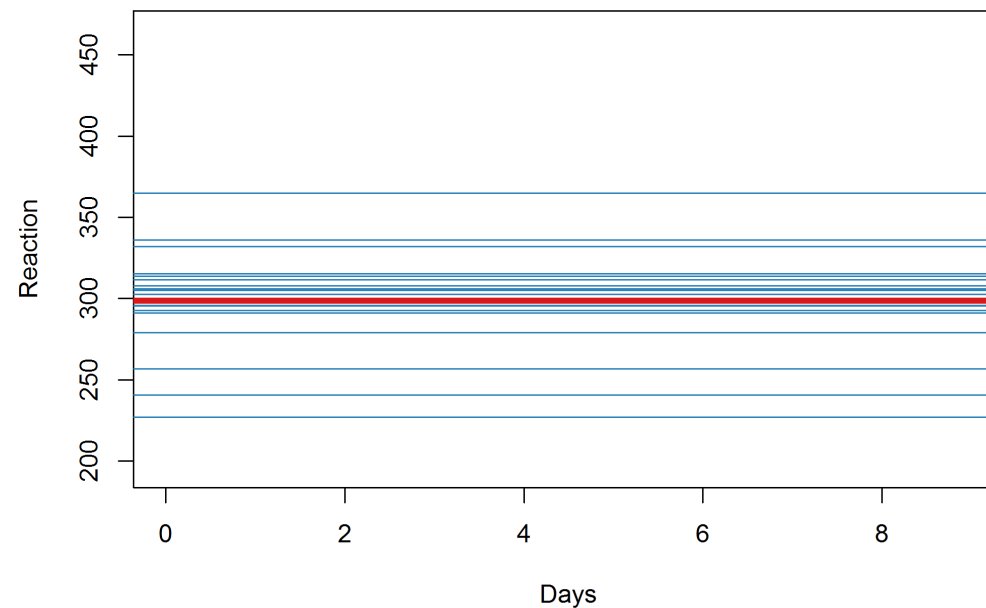


Now the slopes vary by participant too

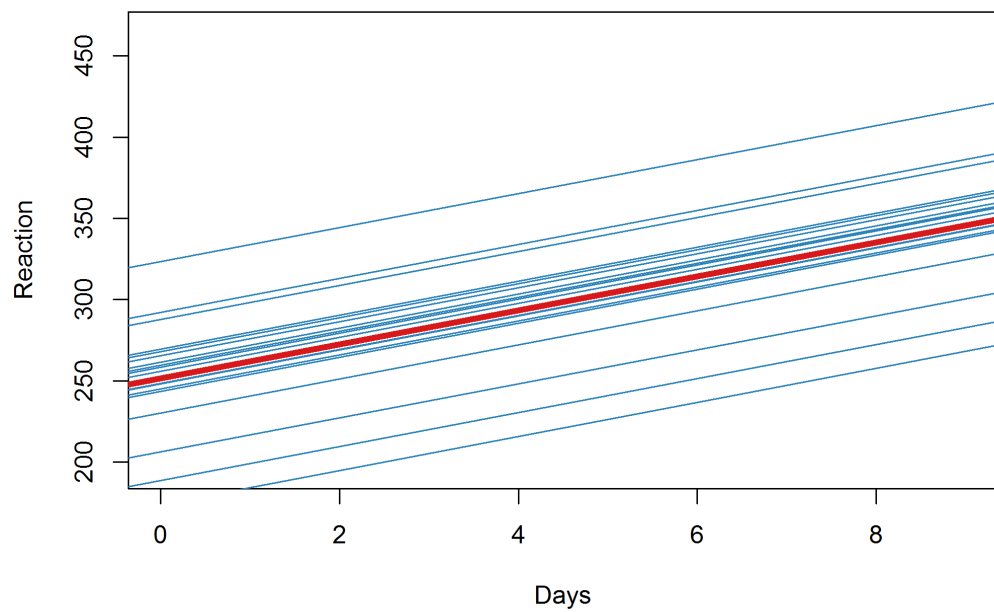




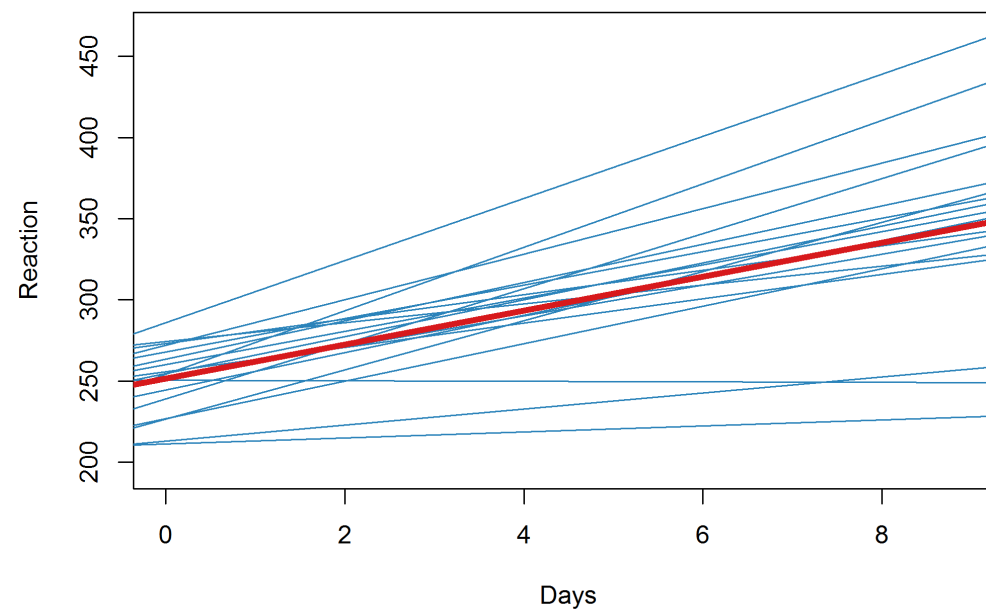
Reaction $\sim 1 + (1|\text{Subject})$



Reaction $\sim \text{Days} + (1|\text{Subject})$



Reaction $\sim \text{Days} + (\text{Days}|\text{Subject})$



Let's try this out in R

PRE-SESSIONAL ACTIVITIES



Tutorial: Multilevel Models (Part 1)



sleepstudy.csv

This is the dataset for this week's Tutorial.

What's in the lmer output?

Formula: Reaction ~ 1 + (1 | Subject)

Data: dat

REML criterion at convergence: 1904.3

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.4983	-0.5501	-0.1476	0.5123	3.3446

Random effects:

Groups	Name	Variance	Std.Dev.
Subject	(Intercept)	1278	35.75
Residual		1959	44.26

Number of obs: 180, groups: Subject, 18

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	298.51	9.05	17.00	32.98	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Formula: Reaction ~ 1 + Days + (1 | Subject)

Data: dat

REML criterion at convergence: 1786.5

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.2257	-0.5529	0.0109	0.5188	4.2506

Random effects:

Groups	Name	Variance	Std.Dev.
Subject	(Intercept)	1378.2	37.12
Residual		960.5	30.99

Number of obs: 180, groups: Subject, 18

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	251.4051	9.7467	22.8102	25.79	<2e-16 ***
Days	10.4673	0.8042	161.0000	13.02	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

(Intr)

Days -0.371

What's in the lmer output?

Formula: Reaction ~ 1 + Days + (1 | Subject)

Data: dat

REML criterion at convergence: 1786.5

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.2257	-0.5529	0.0109	0.5188	4.2506

Random effects:

Groups	Name	Variance	Std.Dev.
Subject	(Intercept)	1378.2	37.12
	Residual	960.5	30.99

Number of obs: 180, groups: Subject, 18

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	251.4051	9.7467	22.8102	25.79	<2e-16 ***
Days	10.4673	0.8042	161.0000	13.02	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

(Intr)

Days -0.371

Formula: Reaction ~ 1 + Days + (1 + Days | Subject)

Data: dat

REML criterion at convergence: 1743.6

Scaled residuals:

Min	1Q	Median	3Q	Max
-3.9536	-0.4634	0.0231	0.4634	5.1793

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Subject	(Intercept)	612.10	24.741	
	Days	35.07	5.922	0.07
	Residual	654.94	25.592	

Number of obs: 180, groups: Subject, 18

Fixed effects:

	Estimate	Std. Error	df	t value	Pr(> t)
(Intercept)	251.405	6.825	17.000	36.838	< 2e-16 ***
Days	10.467	1.546	17.000	6.771	3.26e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:


(Intr)

Days -0.138


To Do:

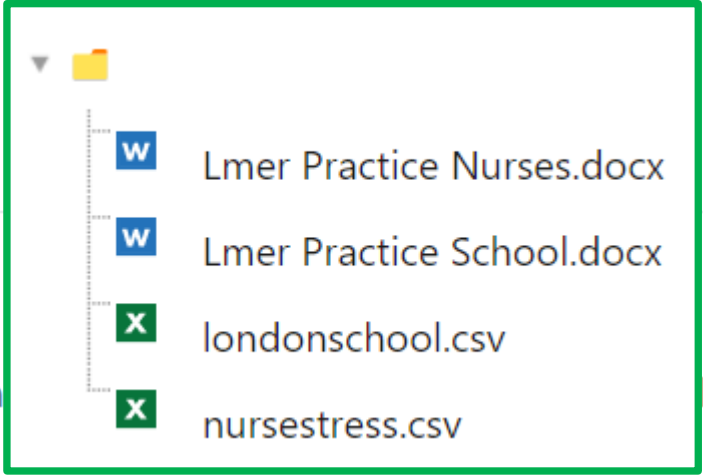
- Keep working on your **project**, set up a meeting if you run into issues or use the Q&A Forum (both links on Moodle)
- Practice exercises:

PRE-SESSIONAL ACTIVITIES

 Multilevel Model Practice

Here are two past exam assignments on m

 Optional: Submit your practice excercises



long with the corresponding datasets.

You can optionally post your answers (not marked)

Questions for me?



m.kaps@bbk.ac.uk