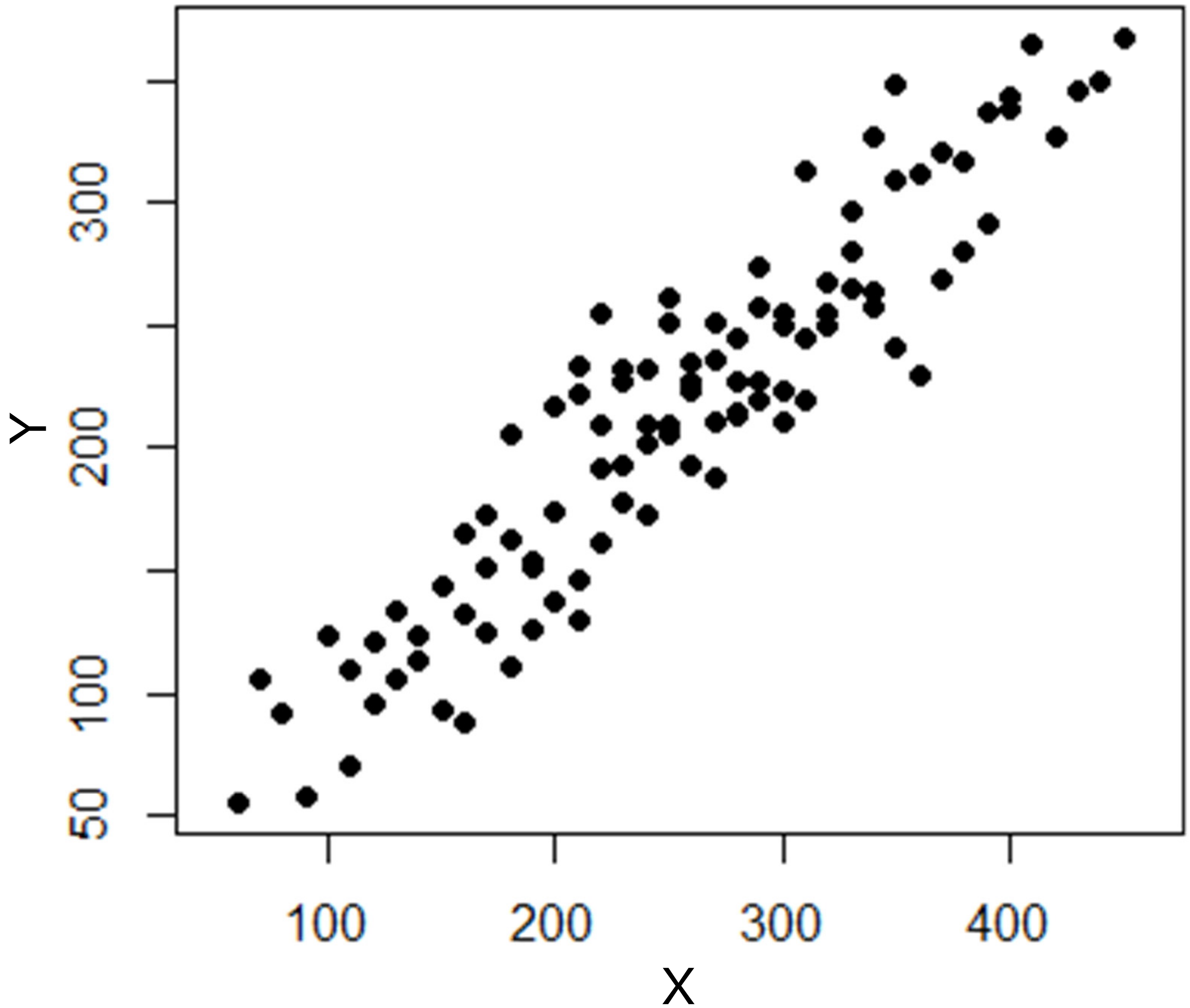
DO NOT TRUST MY ENGLISH!

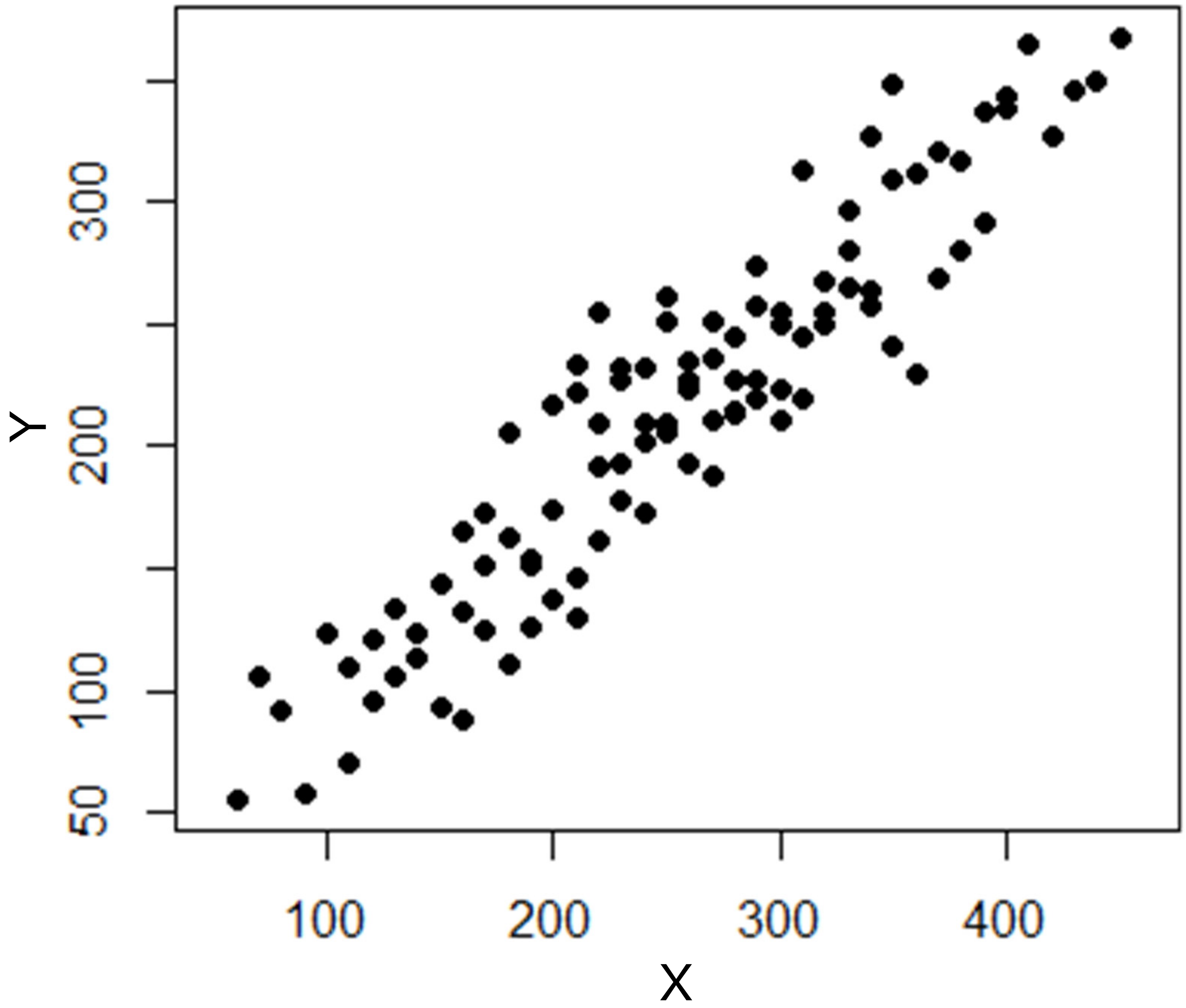
A short explanation of Linear Mixed Models (LMM) (updated for “lmerTest”)



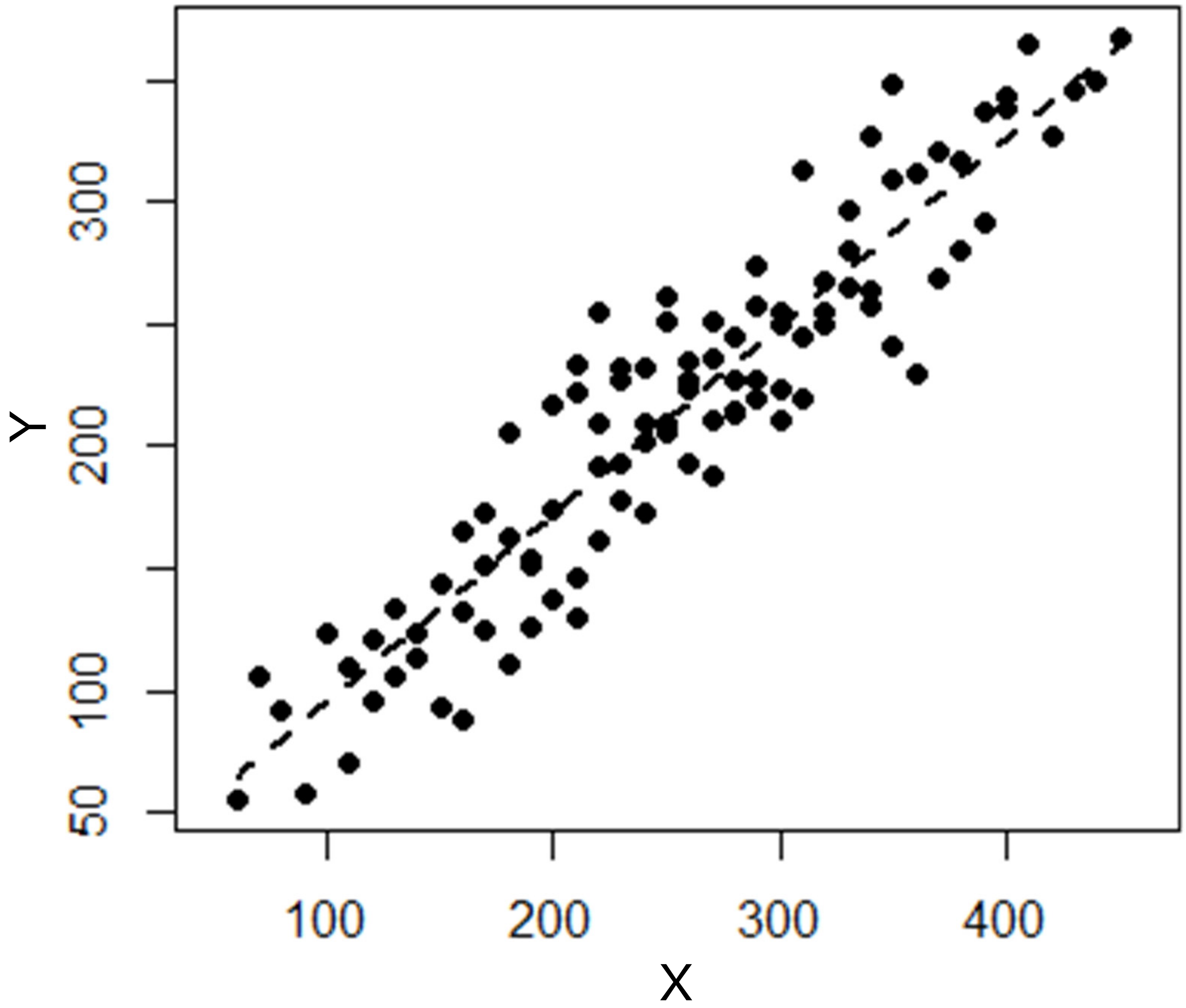
This PDF is downloadable at "My learning page" of
https://www2.kpu.ac.jp/for_ecol/sumida-index.html

ver 20230212e

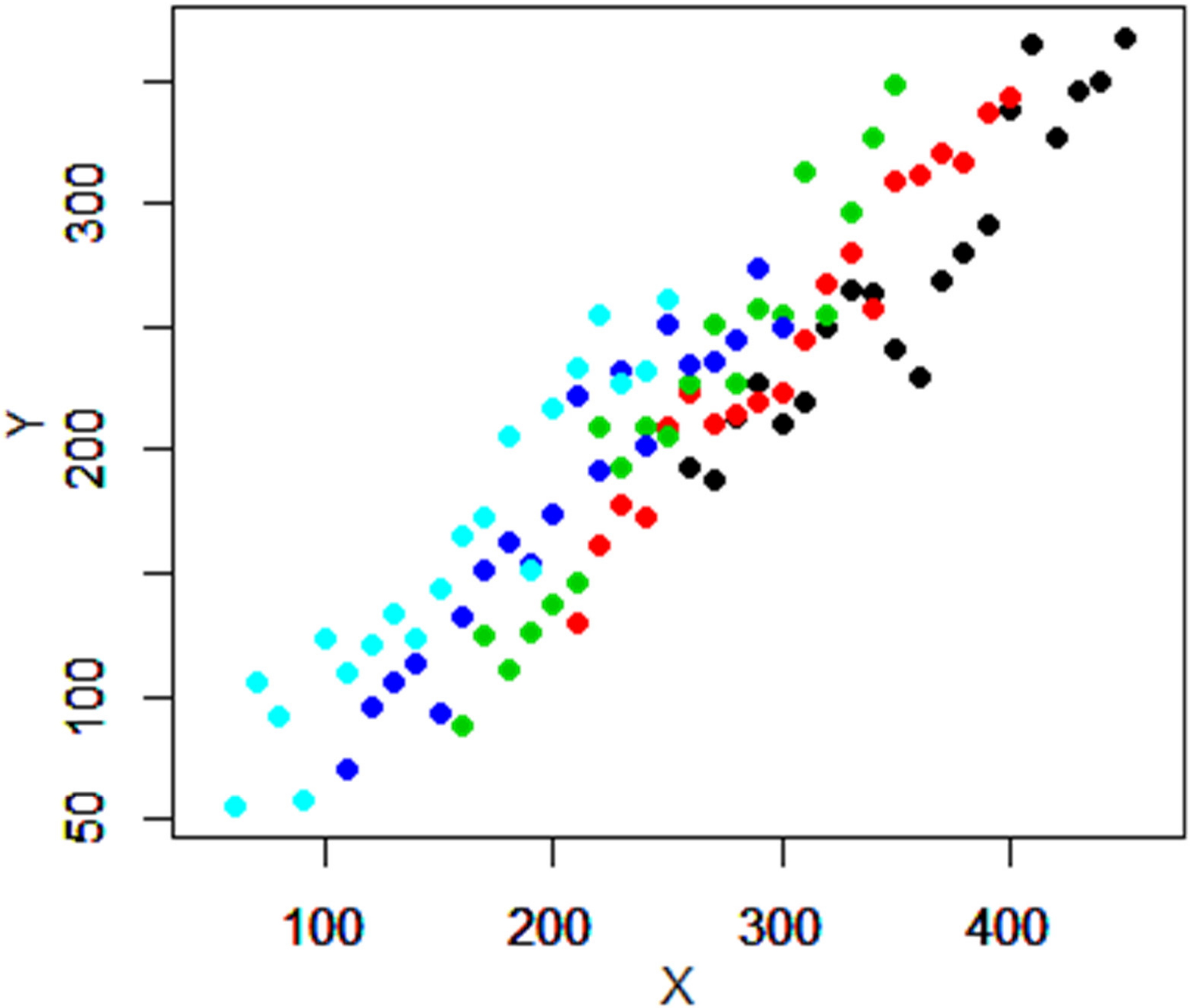
Imagine that you've found a relationship between X and Y as below



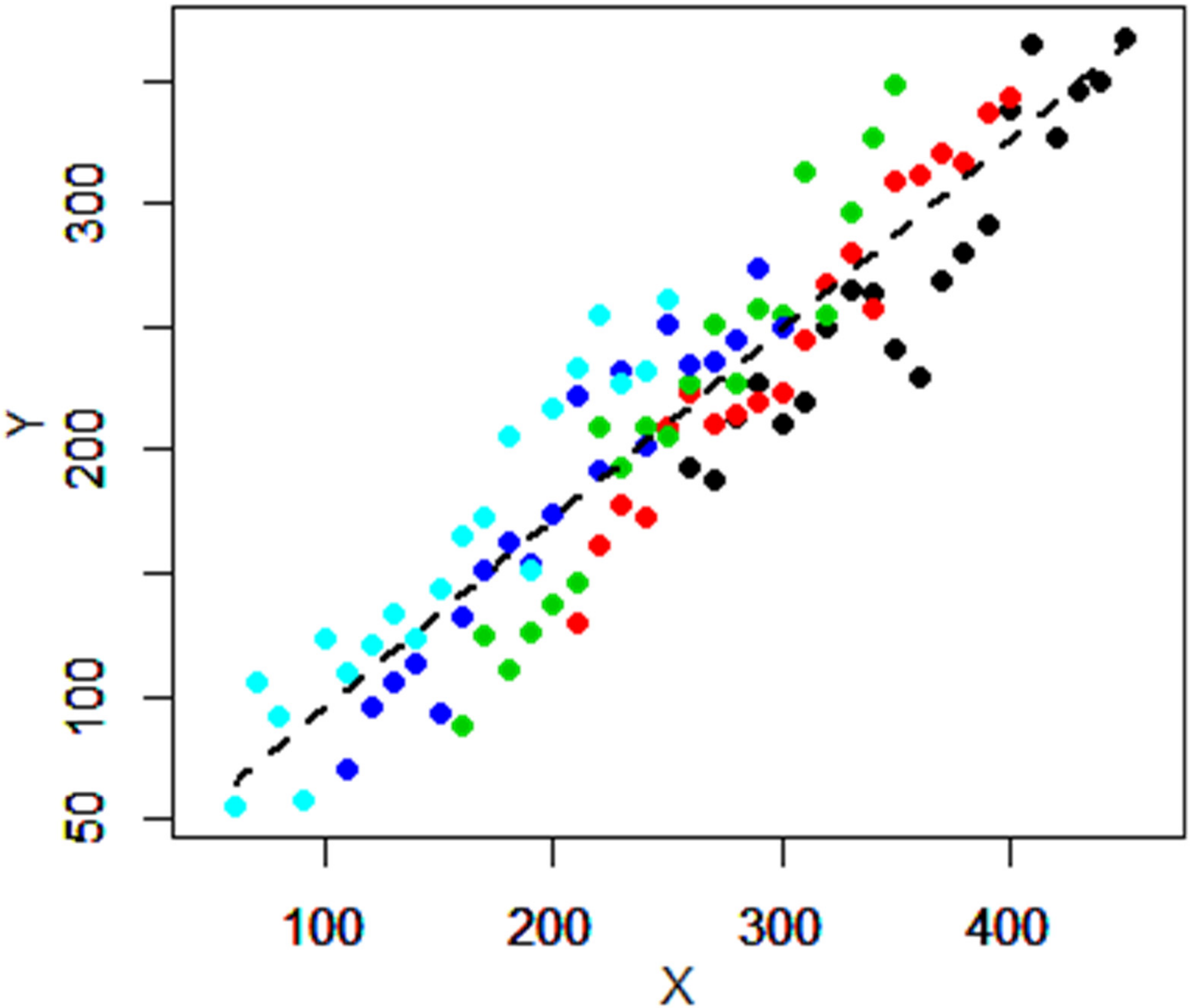
The regression line between X and Y is shown by the dashed line.



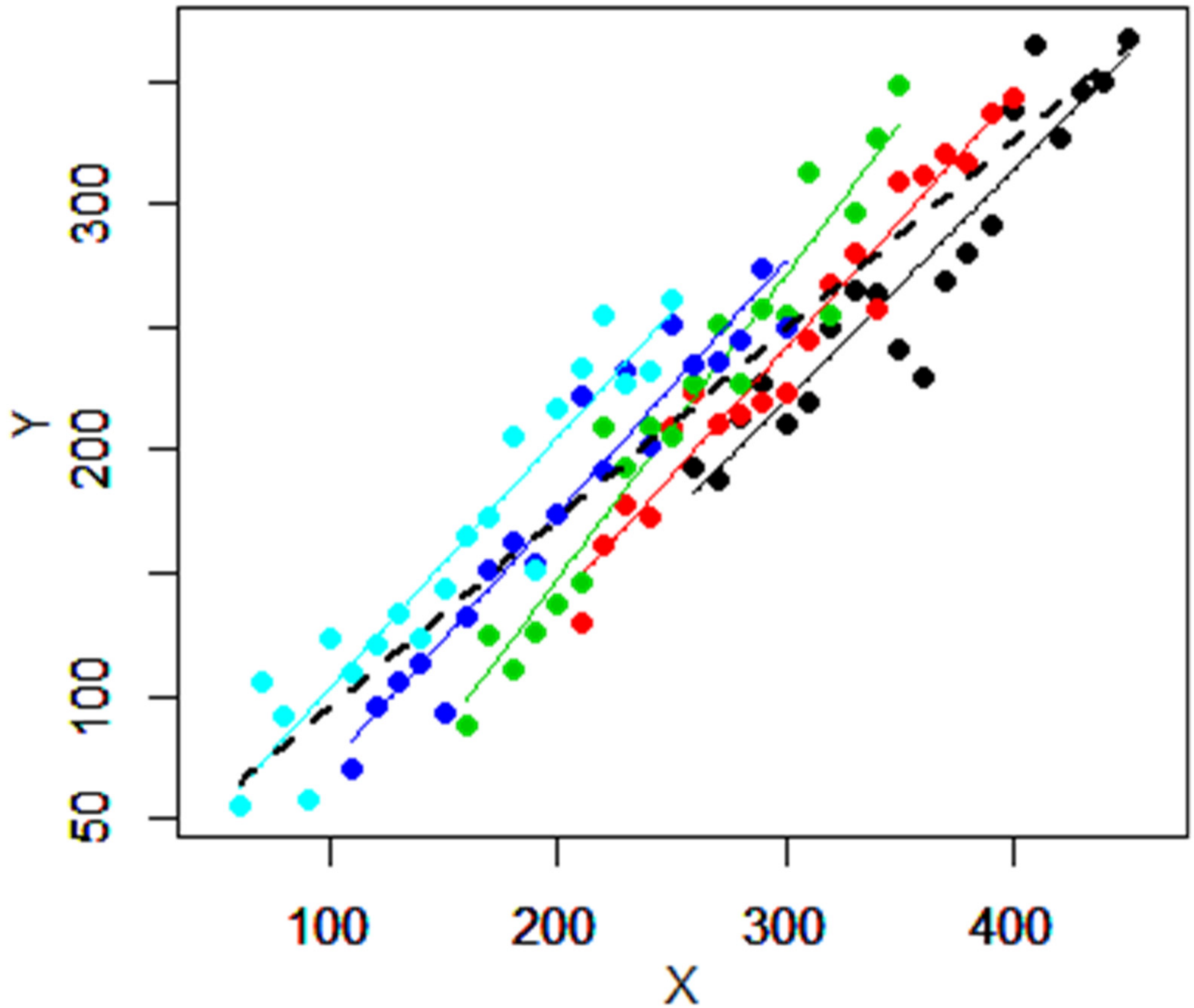
In fact, these data were taken from five different sites;
Site = {Site 1, Site 2, Site 3, Site 4, Site 5}



The regression line for all the data (the dashed line) does not appear to represent a trend within a given site.

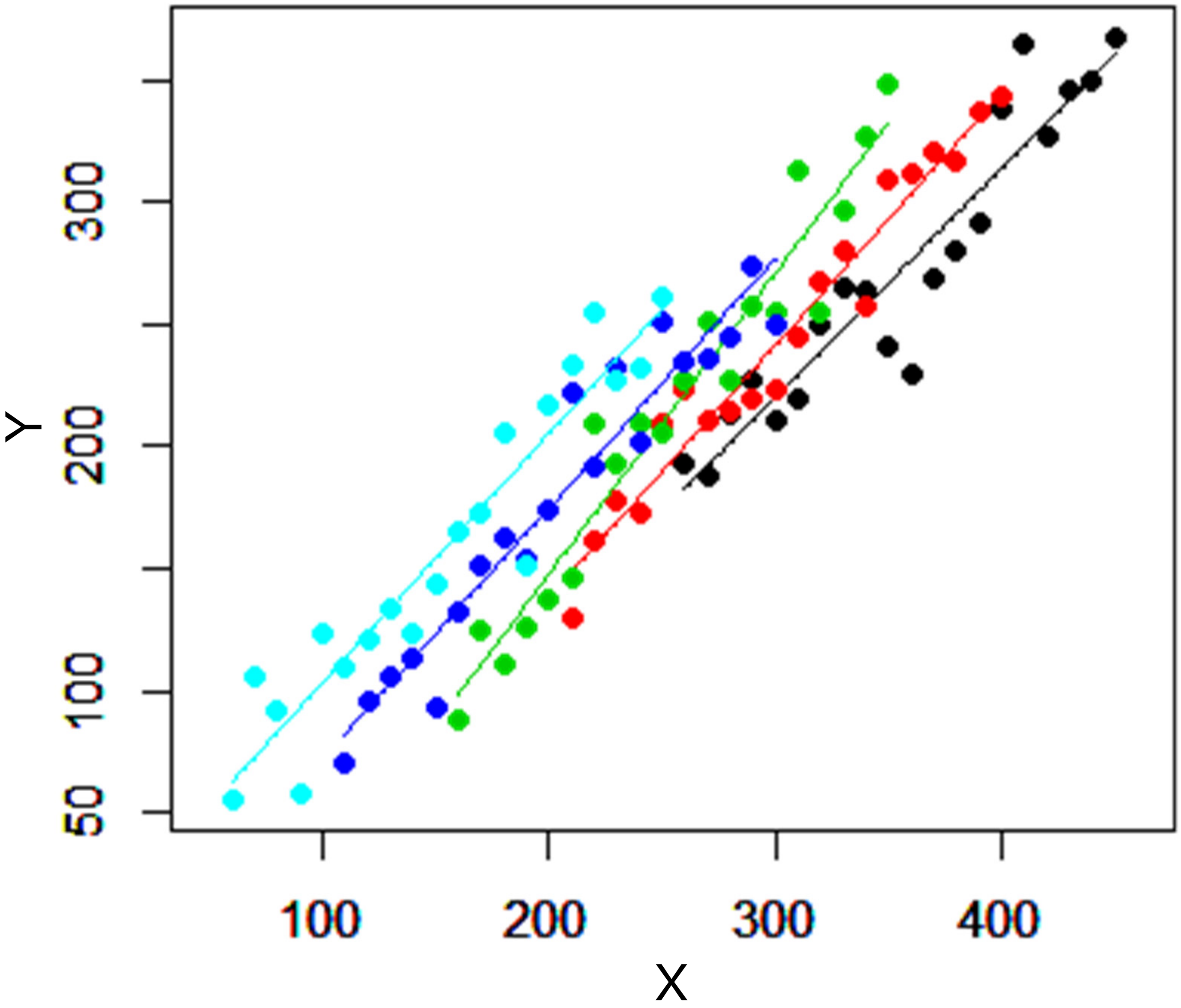


Actually, the slopes of the regression lines (colored solid lines) drawn using the data of the respective sites were different from that of the whole data (the dashed line). That is, the regression by pooling the data of all the sites does not show the trend within each site.



What you want to know is the X-Y relationship WITHIN A SITE if a study site is given.

Of course, the regression line of each site may be enough to show the relationship within a sites. However, you are not interested in the relationship of a specific site. You want to infer a *general* trend within a given site using the data of multiple sites.



What you want to know is the X-Y relationship WITHIN A SITE if a study site is given.

So, let's try (general) linear mixed models analyses using a package "lmerTest" of R

(If you are not interested in R, just read notations in the boxes)

```
# Below is an example of how to conduct a linear mixed model
calculation on the "console" of R (how to understand its meaning)

mixedM <- lmer( y ~ x + (x | Site), XYdata )
```

Saving the calculation results as the variable named 'mixedM'

Designating to process the linear relationship between x and y by the 'Site' to which x and y belong.

'XYdata' is the name of the data file in which x, y, and Site data are saved

In this example, the model designates that the slope and intercept of the regression line are ***fixed*** by the pooled data of the whole sites, **while the model also assumes that intercepts and slopes of different sites may vary randomly depending on "Site"**. Consideration of this "***random effects***" is why the model is called a "***mixed model***".

Try a mixed model with a "Site" as a "random effect"

What are "fixed effects" and "random effects"?

$$y = (a_{\text{Fixed}} + a_{\text{Randome\_by\_site}}) + (b_{\text{Fixed}} + b_{\text{Random\_by\_site}})\, x$$

↑The above equation assumes the regression model as follows:

"The intercept and the slope determined as the **fixed effects** are common to all Sites. Meanwhile, the intercept and the slope of each Site are determined by adding the intercept and the slope of the fixed effects to randomly varying values determined for each Site. "

Variables assigned as random effect ('Site' in this case) must be qualitative/categorical variable.

y = (fixed-effect intercept + by-Site random variation in the intercept)

+(fixed-effect slope + by-Site random variation in the slope) × X

i.e., a mixed model includes both fixed-effect coefficients and random-effect coefficients.

Ex.) a regression showing only fixed-effect coeffs after a mixed-model analysis.

y = -48.82 + 1.04 × X

↑fixed-effect intercept ↑fixed-effect slope

For a given Site, the slope and the intercept are determined by using both fixed- and random-effects, as follows...

y=(-48.82+by-Site random variation in intercept)+(1.04 + by-Site random variation in slope)×X

Ex.) In the case of Site 3 (green data points and regression line),

y = (-48.82 +(-28.4))+(1.04 + 0.11) × x = −77.2 + 1.15 × x

random variation for Site3 random variation for Site 3 mixed-model regression for Site 3

You may skip this slide

Try a mixed model with a "Site" as a "random effect"

```
mixedM <- lmer( y ~ x + (x | Site), XYdata )# 'XYdata' is the name of the data file
  #                         ↑ Meaning that "processing data x by Site"
```

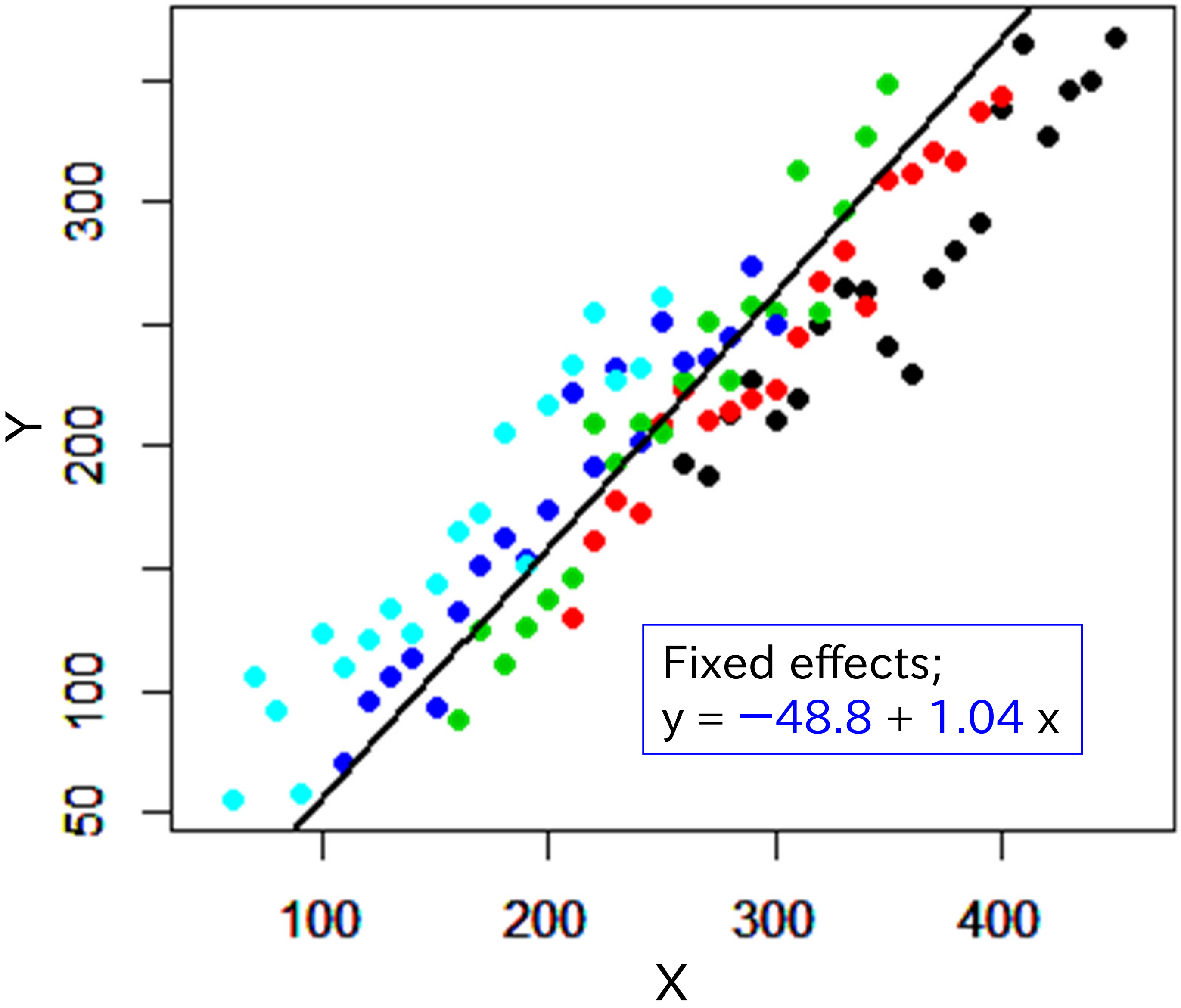
This model designates that the intercept and the slope can vary randomly among Sites.

```
> summary( mixedM )      # ↓[R] outputs:
Linear mixed model fit by REML
Formula: y ~ x + (x | Site)
   Data: XYdata
REML criterion at convergence: 883.0652

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 Site     (Intercept) 9.581e+02 30.95307
          x           7.439e-03  0.08625 -0.34
 Residual             3.283e+02 18.12008
Number of obs: 100, groups:  Site, 5

Fixed effects:
             Estimate Std. Error       df t value Pr(>|t|)
(Intercept) -48.82200   16.09966  6.21986  -3.032    0.022 *
x             1.03949    0.04949  4.54485  21.003 1.06e-05 ***   #←See next slides
                                                                     for explanation

Correlation of Fixed Effects:
  (Intr)
x -0.536
```

Random effects; Here only the information of how intercepts and slopes varied among Sites appears

Fixed effects; y = -48.82 + 1.04 x

The regression line using only the coefficients of the Fixed Effects of the mixed model outputs



about outputs of Random effects :

You may skip this slide

```
mixedM <- lmer(y ~ x + (x | Site), XYdata )   #  an example of linear mixed model (continued)
```

how to extract the coefficients of Fixed Effects parameters of a linear mixed model

```
ALLA <- fixef(mixedM)[1]   # saving the intercept with the name ' ALLA'; ALLA= −48.8

ALLB <- fixef(mixedM)[2]   # saving the slope with the name 'ALLB'; ALLB = 1.04
```

Fixed effects are y = −48.8 + 1.04 x

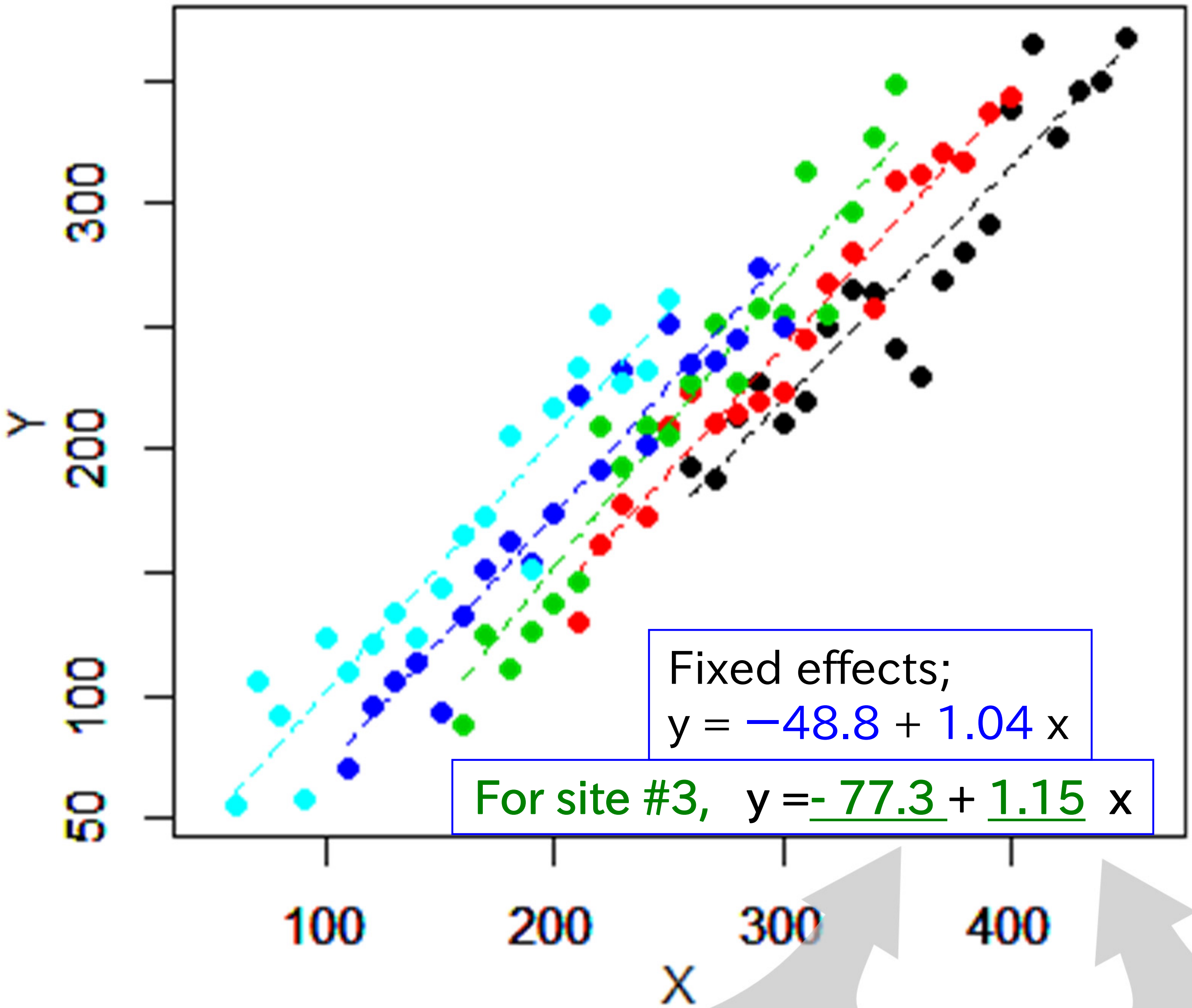
```
> ranef(mixedM)    # ↓How to output random effects coefficients by Site:
  $Site            #  Adding these to the Fixed effects coefficients gives
1   -18.07045 -0.083976266
2   -16.63463 -0.015141819
3    -28.44816  0.109892274
4    15.98362 -0.004122772
5    47.16962 -0.00651416
 ↑Site number
```

Coefficients of the regressions of each Site =
Fixed effects coefs (blue)+ Random effects coefs (brown)

```
# How to interpret  for Site 3, as an example,
AA <-  ALLA + ranef(mixedM)$Site[3, 1]  #   intercept 'AA' = −48.8 +(-28.4) = -77.3
BB <-  ALLB + ranef(mixedM)$Site[3, 2]  #       slope 'BB' =   1.04  + 0.11 =  1.15
```

Hence, for Site 3, y = - 77.3+ 1.15 x

Regression lines of the five Sites calculated from both the fixed effects coefficients and random effects coefficients of respective five Sites; the green lines is for Site #3

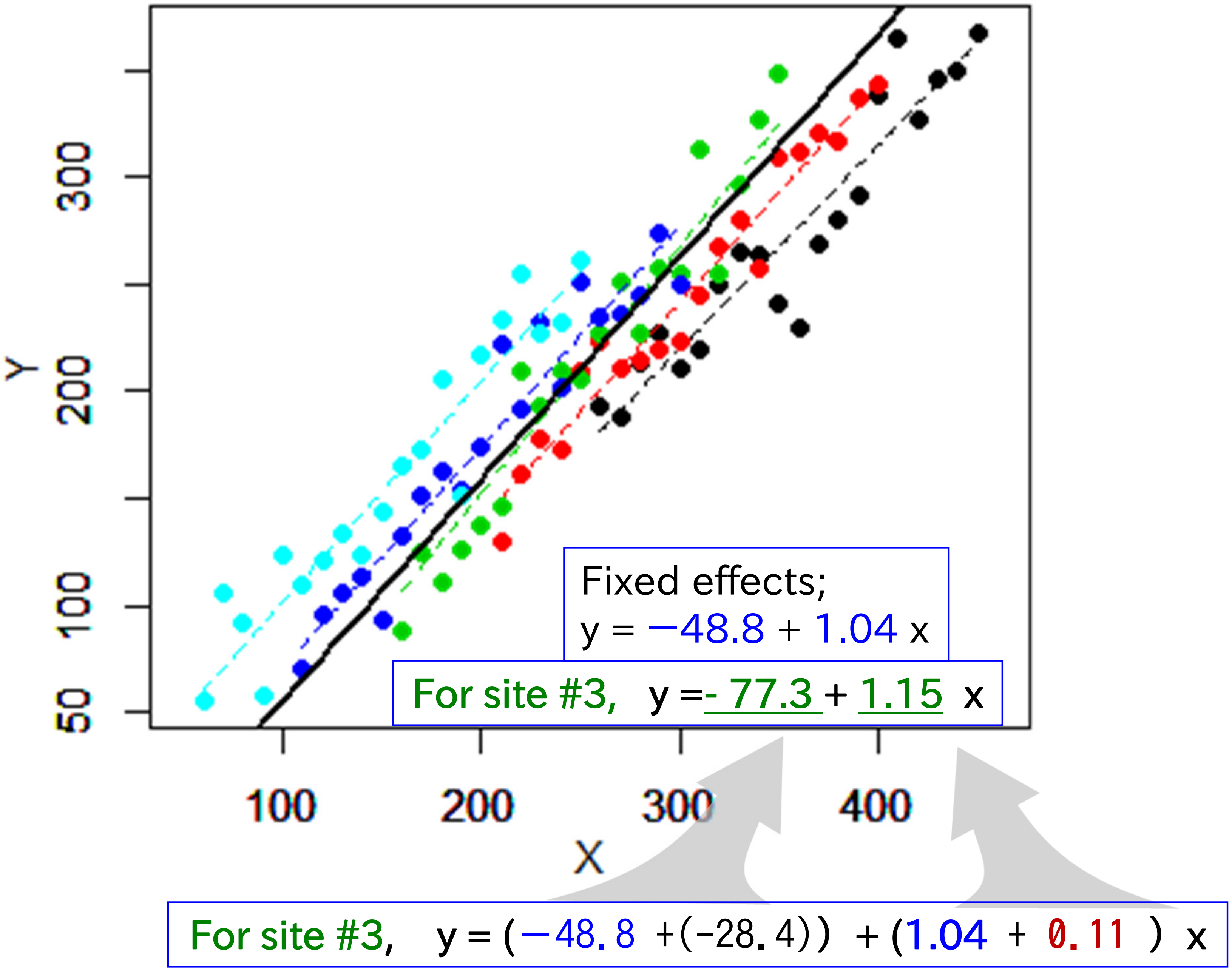


Fixed effects;
$y = -48.8 + 1.04\,x$

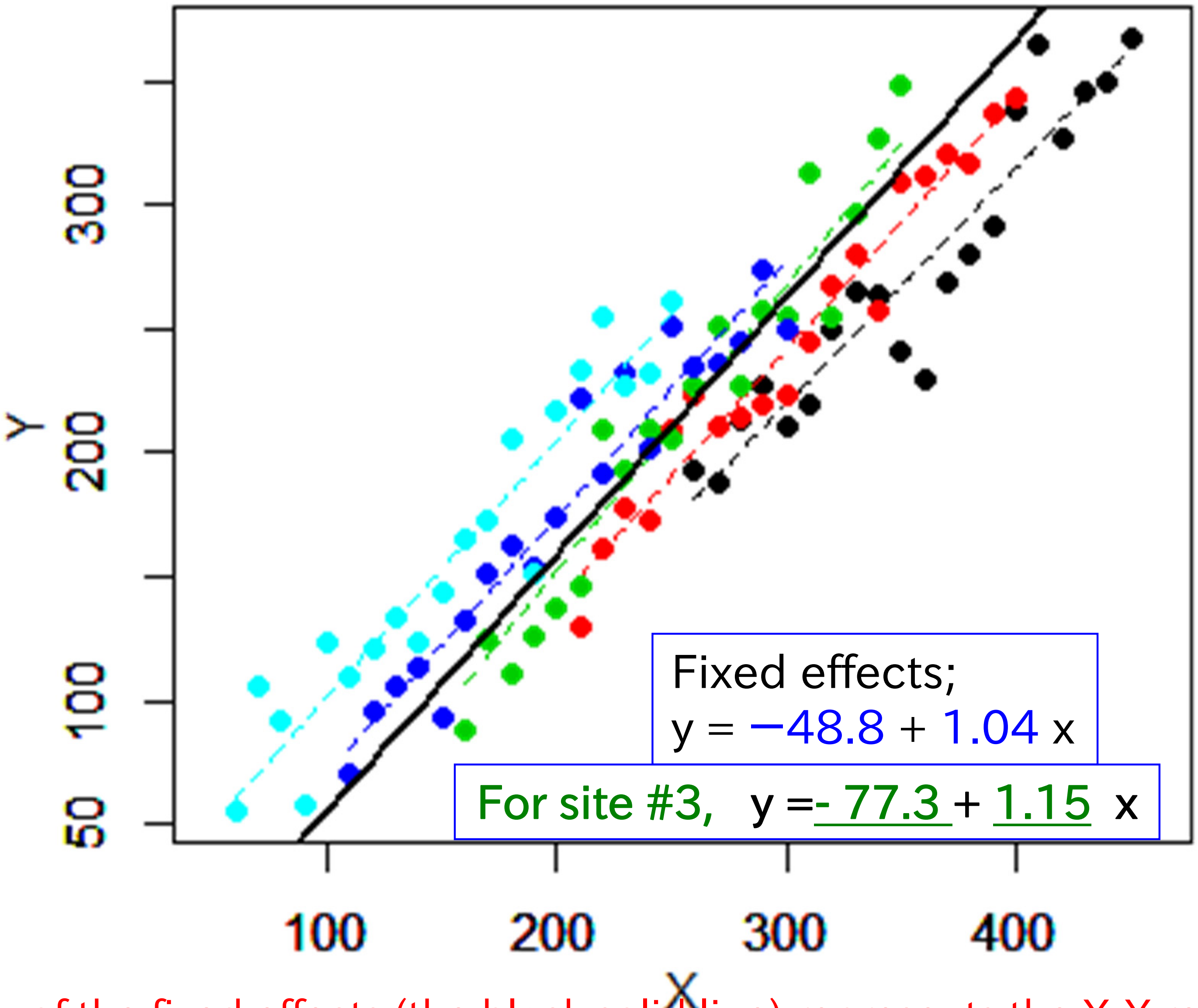
For site #3, $y = -77.3 + 1.15\,x$

For site #3, $y = (-48.8 + (-28.4)) + (1.04 + 0.11)\,x$

Comparison of the Fixed effects regression line (the black solid line) and those of the respective Sites considering the random effects among the five Sites (colored dashed lines).

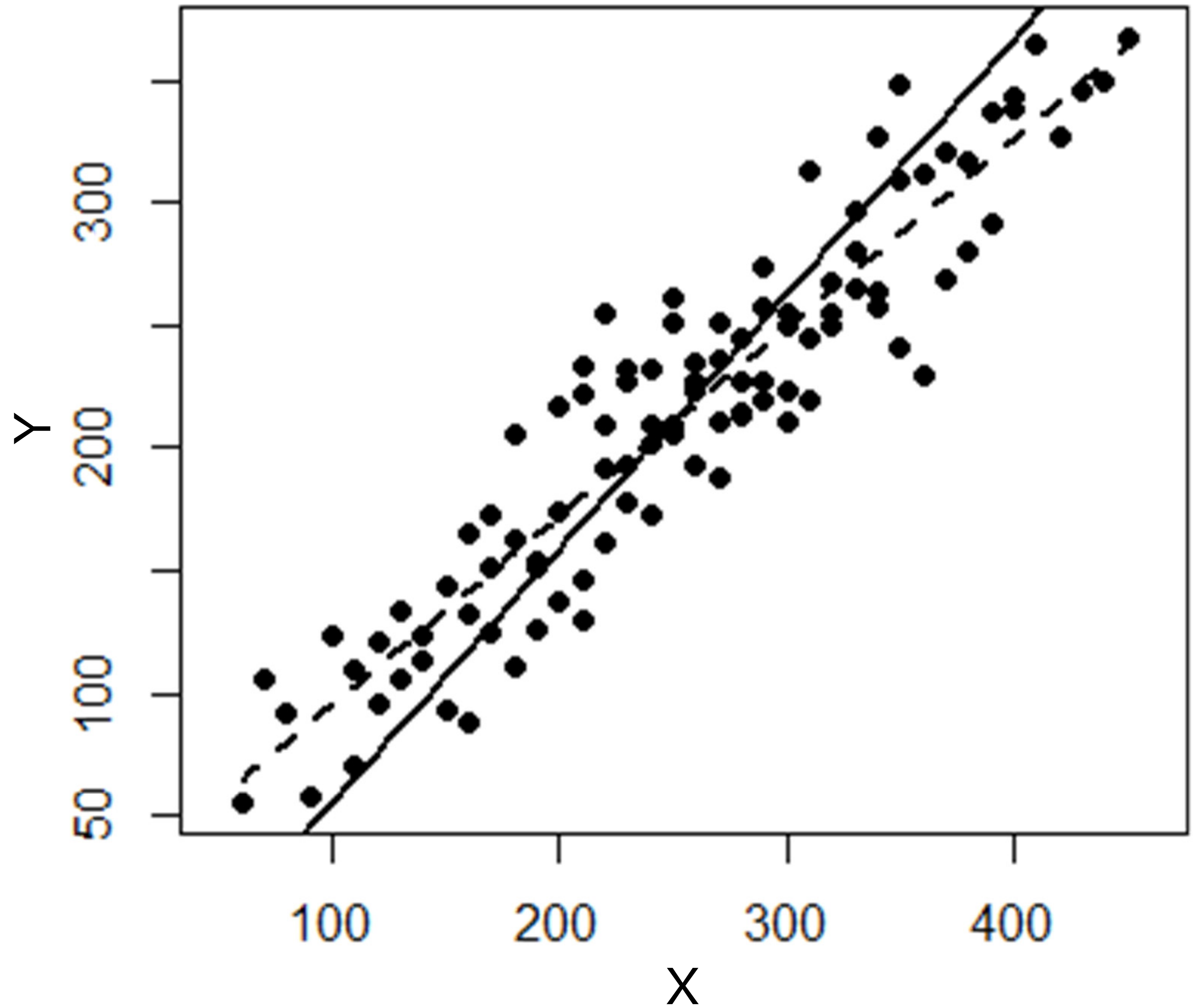


Comparison of the Fixed effects regression line (the black solid line) and those of the respective Sites considering the random effects among the five Sites (colored dashed lines).



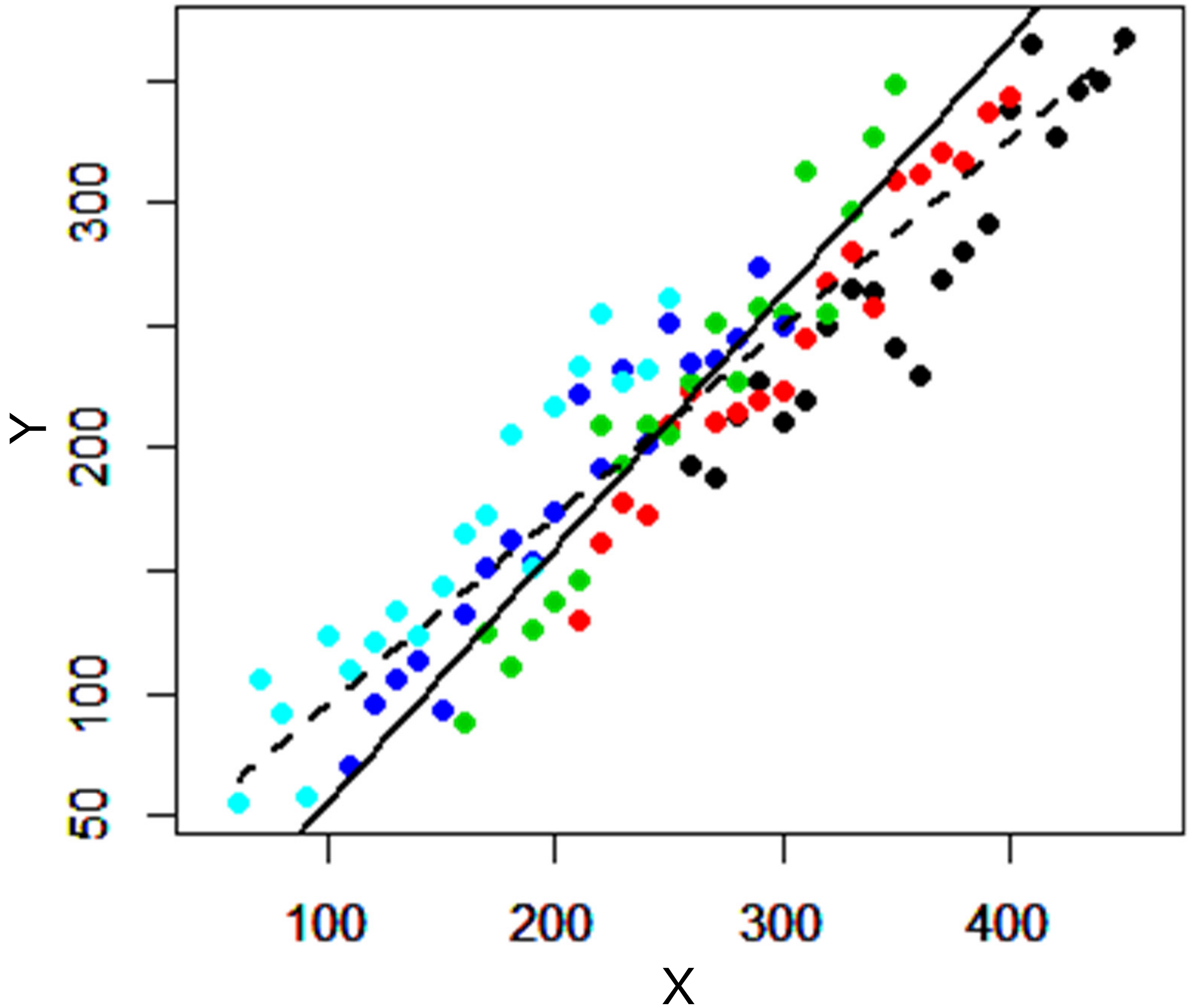
The regression of the fixed effects (the black solid line) represents the X-Y relationship within a given Site. Meanwhile, among Sites, intercepts and slopes of regression lines vary randomly (such that the sum of the variation is 0)).

To understand what a mixed models result mean, let's get back again to the first figure. Here the regression line by pooling all the data (dashed) is compared with the one using the fixed effects outputs of the mixed model (solid).



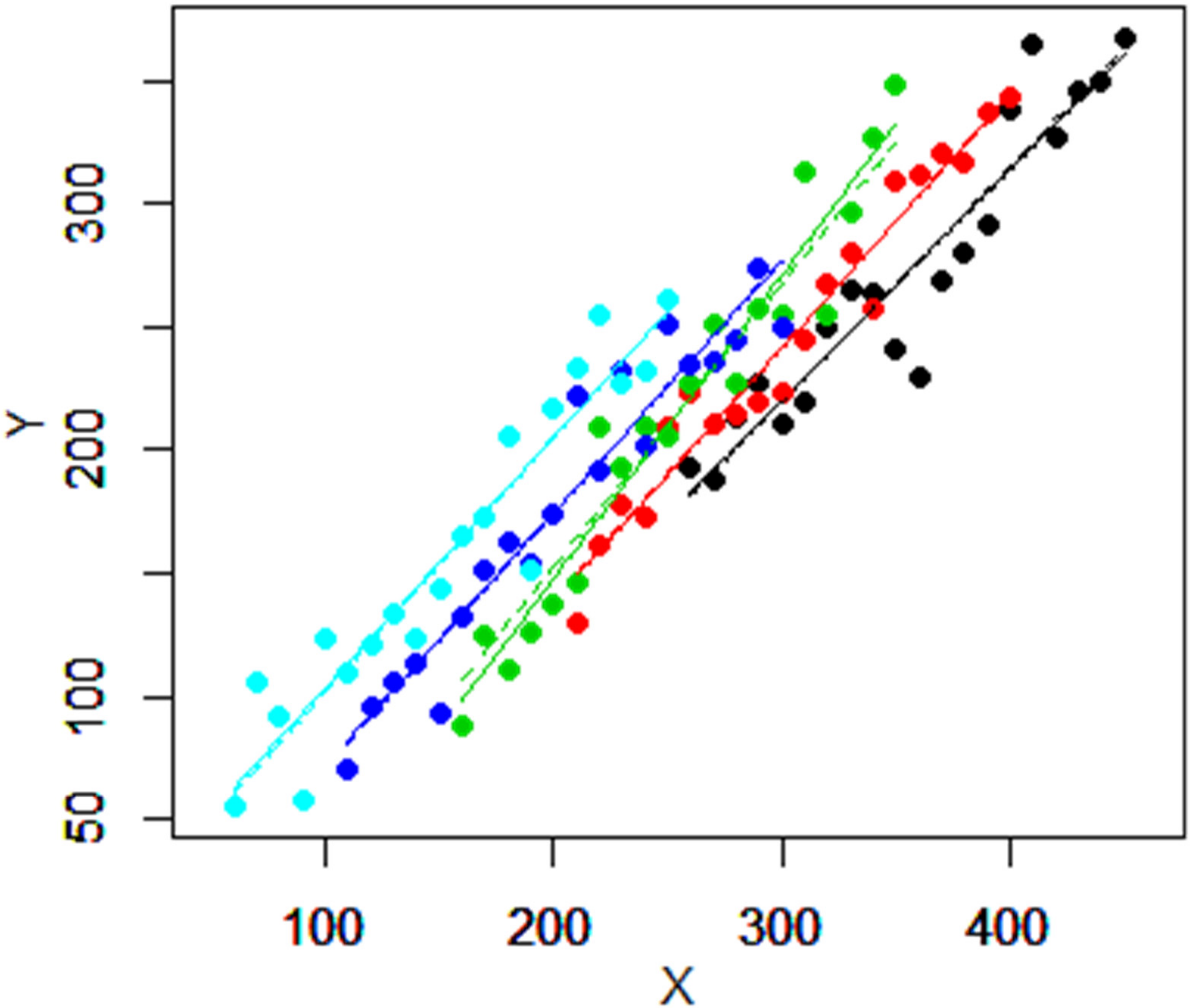
We found that the regression by pooling the data of all the sites (dashed) did not consider the trend within a given site. That is . . .

To understand what a mixed models result mean, let's get back again to the first figure. Here the regression line by pooling all the data (dashed) is compared with the one using the fixed effects outputs of the mixed model (solid).



The results of the FIXED effects only represent a trend within a kind of an 'average' Site, and do not represent the overall trend if the data of all the site are pooled.

Incidentally, in this example, the regression lines with the mixed model (dashed lines) showed good agreement with those obtained by those calculating independently by each site (solid lines).



SUMMARY

The regression by pooling the data of all the Sites without considering the random effects could not express the trend within a given site.

In the present example, Site was considered as a random effect of a mixed model. In this mixed model, it was assumed that the slope and the intercept of the regression of a given site vary randomly among Sites.

Using the mixed models analyses, we can infer the representative trend if an arbitrary site is given. Further, we can also know how such a relationship may vary among different sites simultaneously.

Appendix

```
> summary( mixedM )     # ↓[R] outputs:
Linear mixed model fit by REML
Formula: y ~ x + (x | Site)
   Data: XYdata
REML criterion at convergence: 883.1

Random effects:
 Groups   Name        Variance  Std.Dev. Corr
 Site     (Intercept) 9.581e+02 30.95307
          x           7.439e-03  0.08625 -0.34
 Residual             3.283e+02 18.12008
Number of obs: 100, groups: Site, 5
   (etc.)
```

The output of this mixed model suggests that there was a weak correlation (Corr; r=0.34) between the intercepts (intercept) and the slopes (x) among Sites. That is, among Sites, intercepts and slopes were not completely independent. If a strong correlation is observed, it is "over-parameterized" (too much of parameters". In such cases, either intercept or slope is implemented as a random effect. We can also designate that intercepts and slopes are determined independently.

Variations of the model

(1) a model by Site.
mixedM <- lmer(y ~ x + (x | Site))

(2) a model with intercepts varying randomly by Site, but slopes being common.
mixedM <- lmer(y ~ x + (1 | Site))

(3) a model with slopes varying randomly by Site, but intercepts being common
mixedM <- lmer(y ~ x +(0 + x | Site)) # almost meaningless ··

(4) a model with both slopes and intercepts determined independently and varying randomly among Sites.
mixedM <- lmer(y ~ x + (1 | Site)+(0 + x | Site))

A comparison of results: In the present data, results of models (4) and (1) were almost the same.

```
Linear mixed model fit by REML
Formula: y ~ x + (1 | Site) + (0 + x | Site)
   Data: XYdata
REML criterion at convergence: 883.4
Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.61903 -0.62299 -0.00847  0.60479  2.17110
Random effects:
 Groups   Name        Variance  Std.Dev.
 Site     (Intercept) 1.229e+03 35.0521
 Site.1   x           5.806e-03  0.0762
 Residual             3.254e+02 18.0385
Number of obs: 100, groups:  Site, 5
Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept) -49.81758   17.68385  4.15330  -2.817   0.0459 *
x             1.04345    0.04608  5.00280  22.645 3.11e-06 ***
Correlation of Fixed Effects:
  (Intr)
x -0.298
> ranef(mixedMs)
$Site
  (Intercept)            x
1   -22.23224 -0.074135891
2   -16.67684 -0.016154450
3   -23.45800  0.090350351
4    15.98423 -0.002795307
5    46.38285  0.002735297
with conditional variances for  "Site"
> AIC(mixedMs)
[1] 893.4008
```

```
Linear mixed model fit by REML
Formula: y ~ x + (x | Site)
   Data: XYdata
 REML criterion at convergence: 883.1
Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.61768 -0.60644 -0.01338  0.62644  2.18897
Random effects:
 Groups   Name        Variance  Std.Dev. Corr
  Site     (Intercept) 9.581e+02 30.95307
           x           7.439e-03  0.08625 -0.34
  Residual             3.283e+02 18.12008
 Number of obs: 100, groups:  Site, 5
 Fixed effects:
             Estimate Std. Error       df t value Pr(>|t|)
 (Intercept) -48.82200   16.09966  6.21986  -3.032    0.022 *
 x             1.03949    0.04949  4.54485  21.003 1.06e-05 ***
  Correlation of Fixed Effects:
   (Intr)
 x -0.536
> ranef(mixedM)
$Site
  (Intercept)            x
1   -18.07045 -0.083976266
2   -16.63463 -0.015141819
3   -28.44816  0.109892274
4    15.98362 -0.004122772
5    47.16962 -0.006651416
> AIC(mixedM)
[1] 895.0652
```

A comparison of results: In the present data, results of models (2) and (1) were almost the same, too.

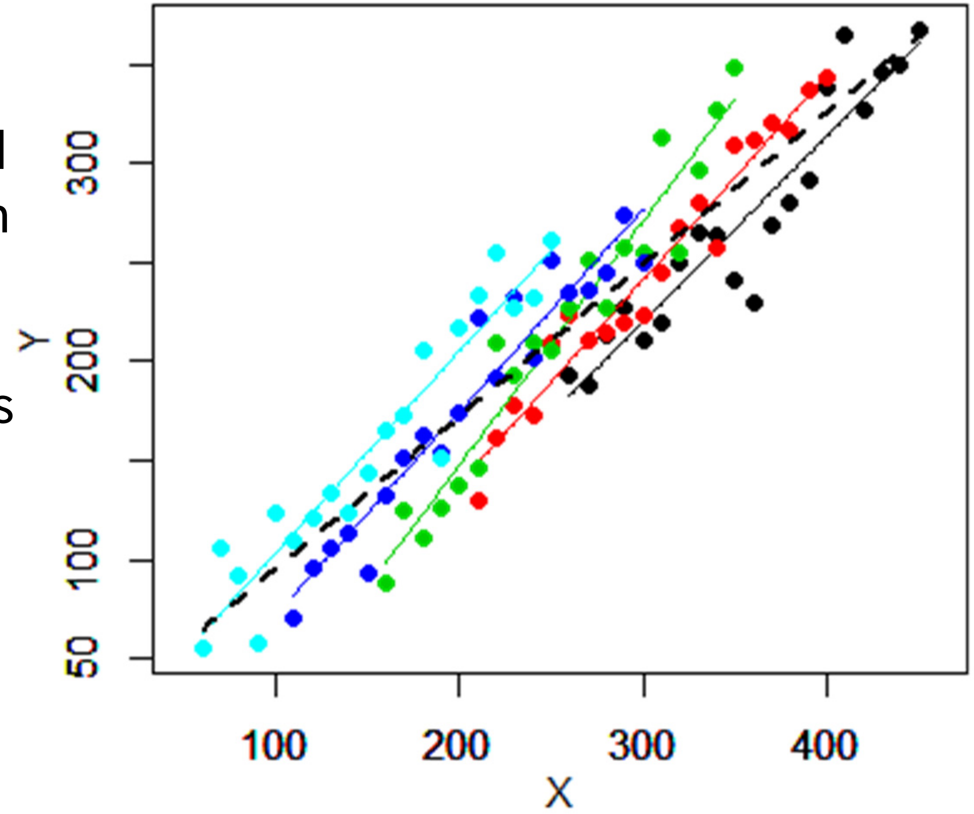
```
Linear mixed model fit by REML
Formula: y ~ x + (1 | Site)
 REML criterion at convergence: 885.8
Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.57606 -0.62083 -0.02495  0.55916  2.17551
Random effects:
 Groups   Name        Variance Std.Dev.
 Site     (Intercept)  1312     36.22
 Residual              346      18.60
Number of obs: 100, groups:  Site, 5
Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
(Intercept) -50.92900   18.22657  5.94197  -2.794   0.0317 *
x             1.04125    0.03195 96.81891  32.591   <2e-16 ***
Correlation of Fixed Effects:
  (Intr)
x -0.447
> ranef(mixedMs)
$Site
   (Intercept)
1 -46.3432089
2 -19.7831275
3   0.9294691
4  16.9604568
5  48.2364105
> AIC(mixedMs)
[1] 893.7776
```

```
Linear mixed model fit by REML
Formula: y ~ x + (x | Site)
   Data: XYdata
 REML criterion at convergence: 883.1
Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.61768 -0.60644 -0.01338  0.62644  2.18897
Random effects:
 Groups   Name        Variance  Std.Dev. Corr
  Site     (Intercept) 9.581e+02 30.95307
           x           7.439e-03  0.08625 -0.34
  Residual             3.283e+02 18.12008
 Number of obs: 100, groups:  Site, 5
 Fixed effects:
            Estimate Std. Error       df t value Pr(>|t|)
 (Intercept) -48.82200   16.09966  6.21986  -3.032    0.022 *
 x             1.03949    0.04949  4.54485  21.003 1.06e-05 ***
  Correlation of Fixed Effects:
   (Intr)
 x -0.536
> ranef(mixedM)
$Site
   (Intercept)            x
 1   -43.242669 -0.082590168
 2   -21.029597 -0.012001706
 3     5.121477  0.119677319
 4    14.748137 -0.008703391
 5    44.402652 -0.016382055
> AIC(mixedM)
[1] 895.0652
```

Further appendix

The data used in this explanation were artificially generated such that the mean values of X of the respective sites differed by 50, in order to differentiate slopes and intercepts of the five sites.

If you encounter the data of this kind, you would first consider a model that the intercept and the slope of each site would correlate with the mean value of X of each site (i.e., intercepts increase with decreasing mean value of X). In this case, a linear mixed model may not always be necessary.



In this explanation, I used a linear mixed model in order to show what the mixed model is about.