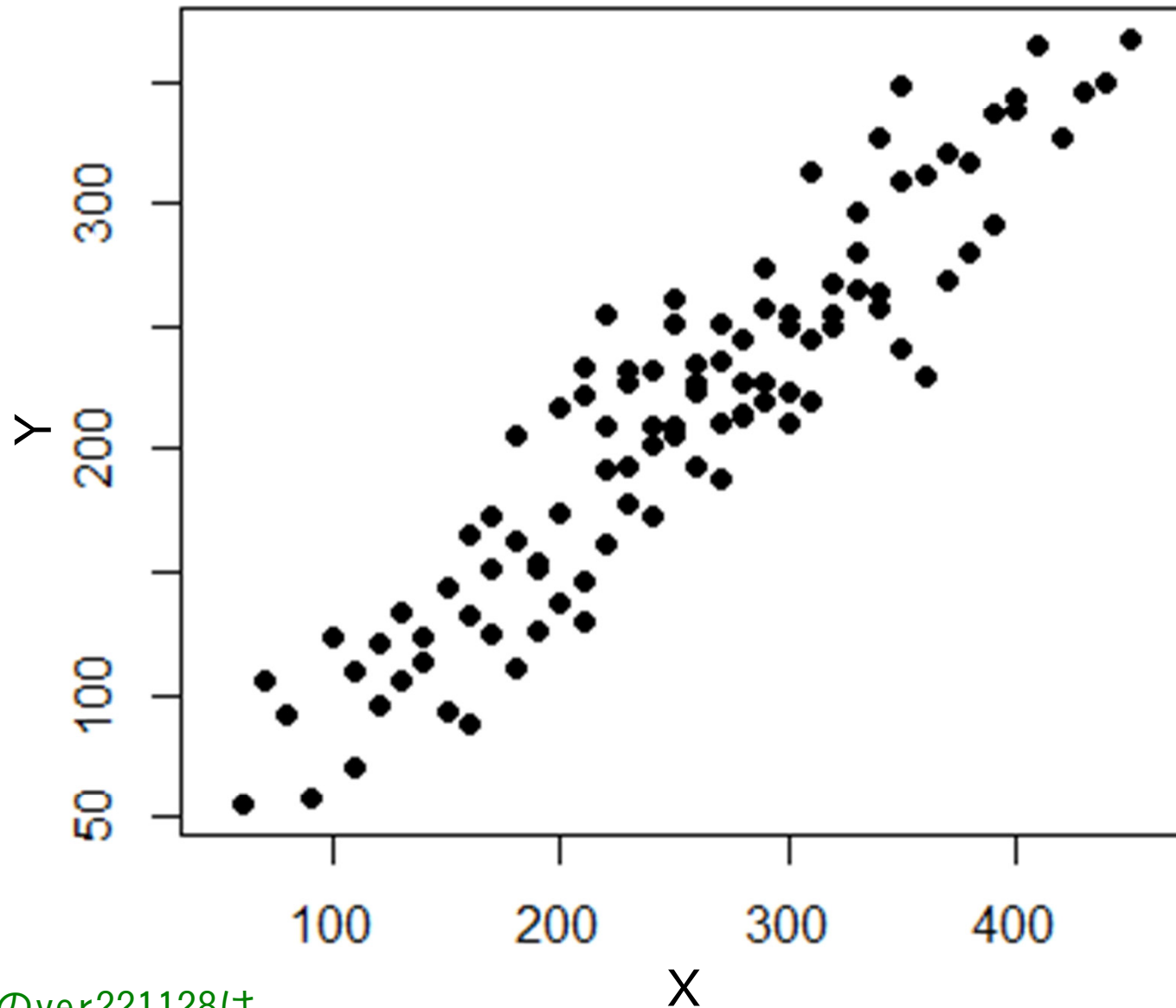
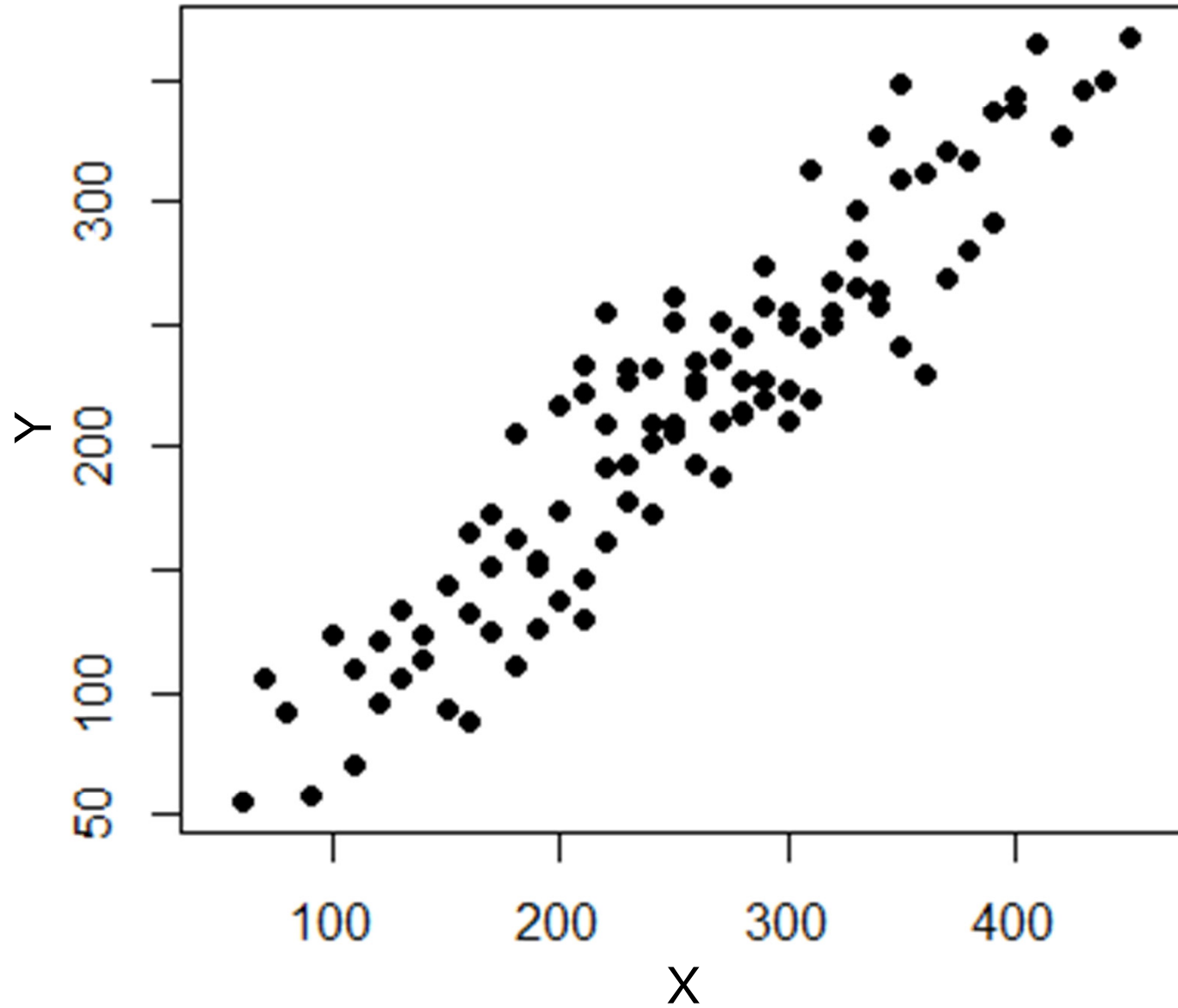


## Linear Mixed Model (以下、混合モデル)の短い解説

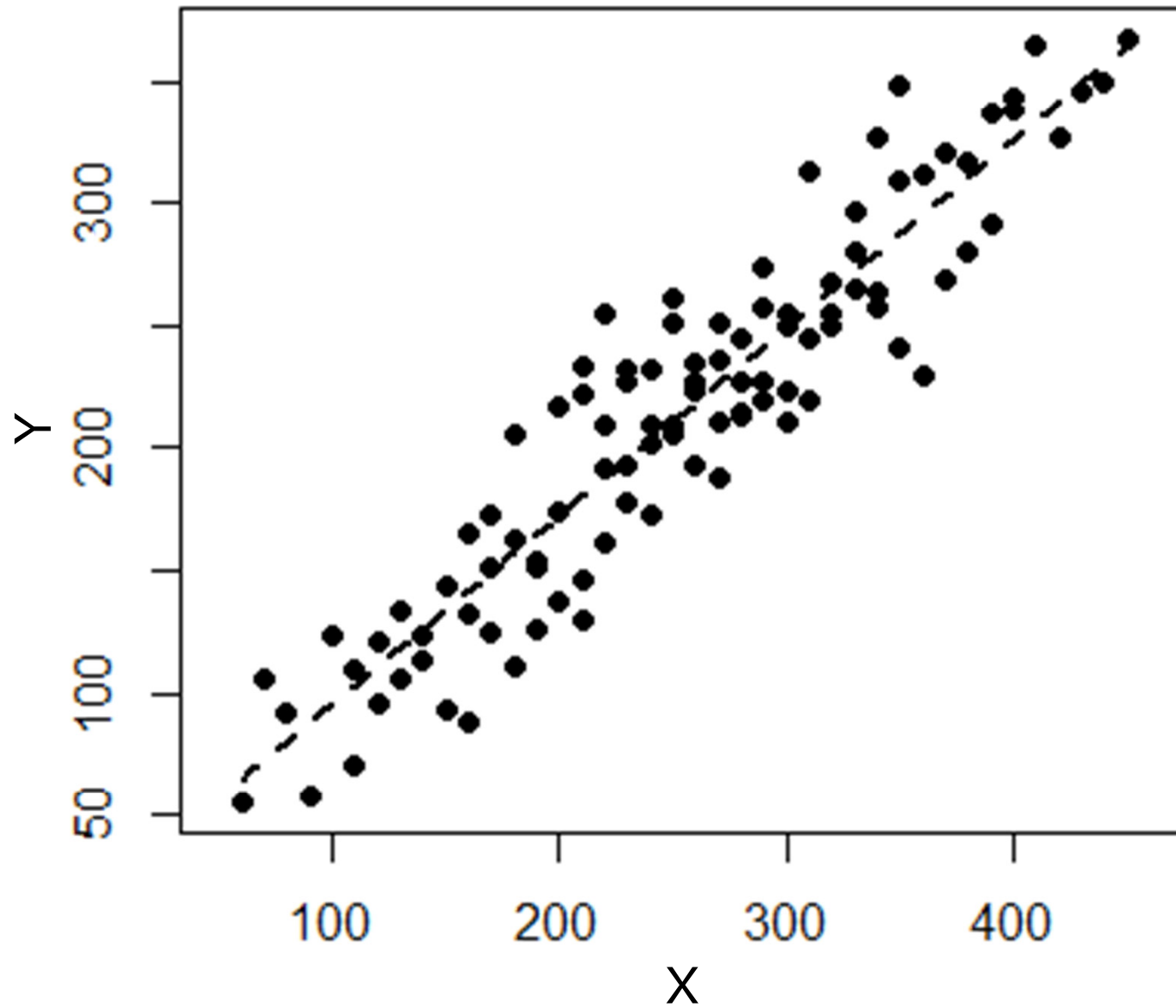


この解説のPDFのver221128は  
[https://www2.kpu.ac.jp/for\\_ecol/sumida-index.html](https://www2.kpu.ac.jp/for_ecol/sumida-index.html)  
の「お勉強」のページにあります。

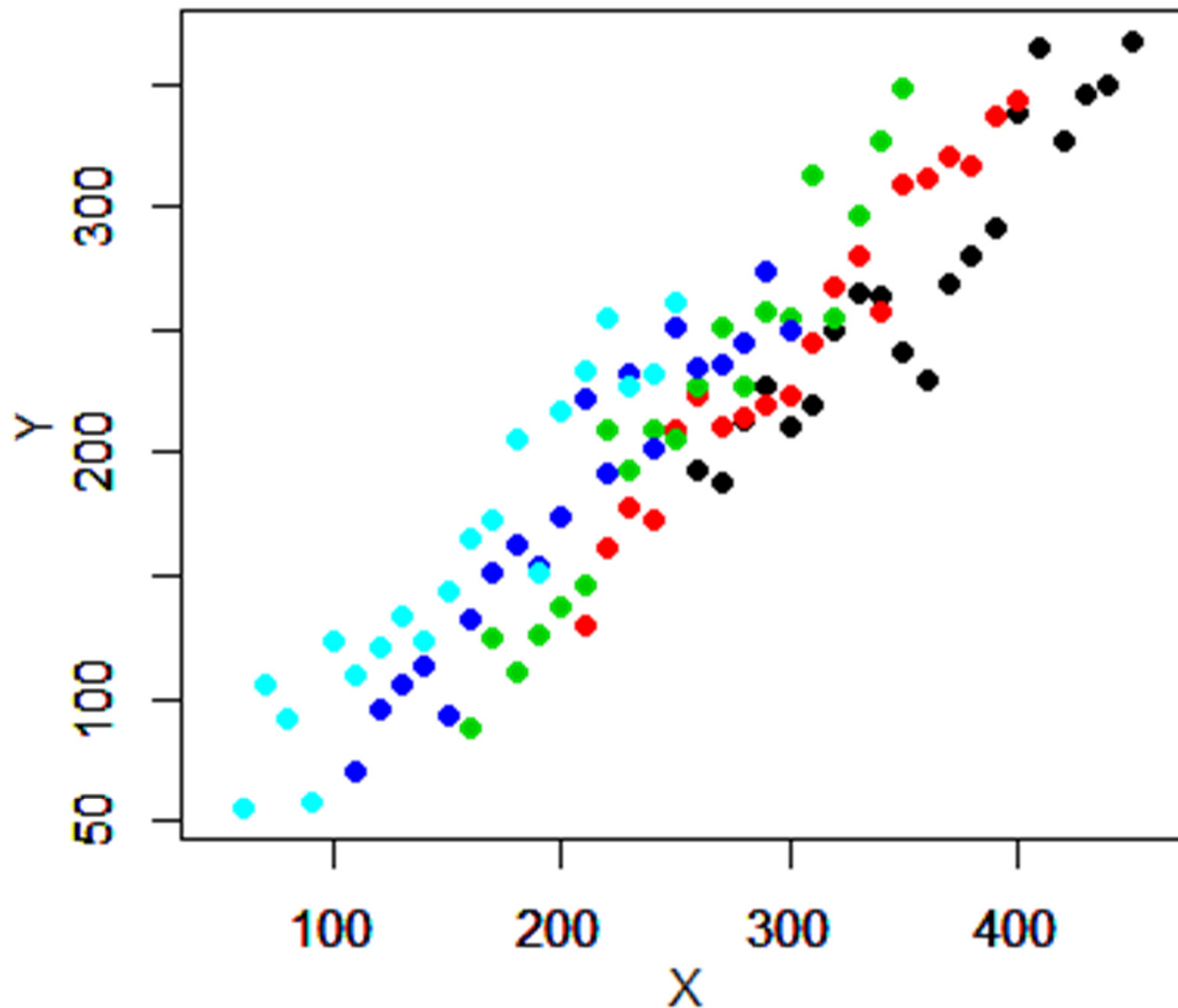
XとYとの間に次のような関係が見つかったとしよう。



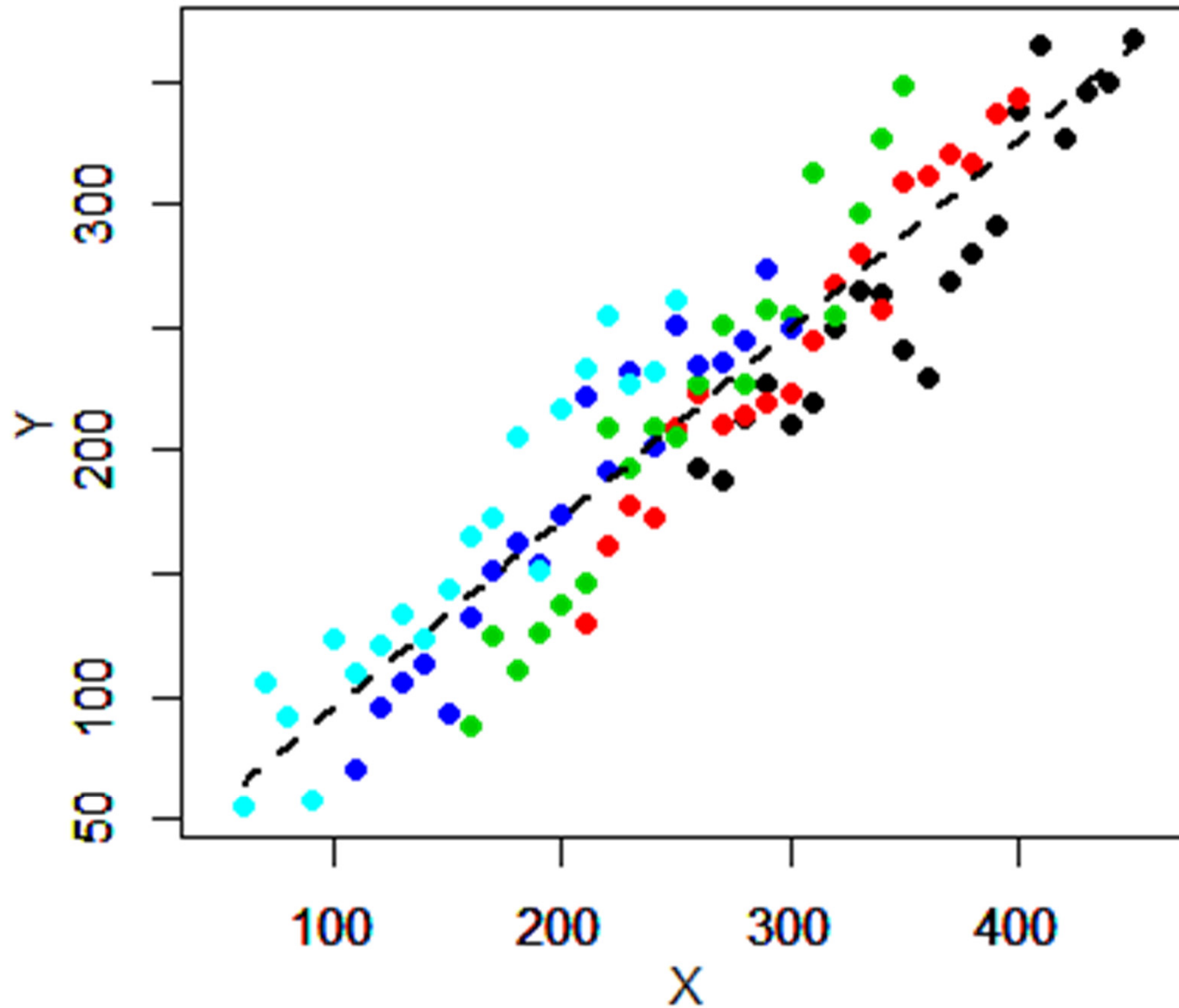
全体的な傾向に対する回帰直線を点線で示した。



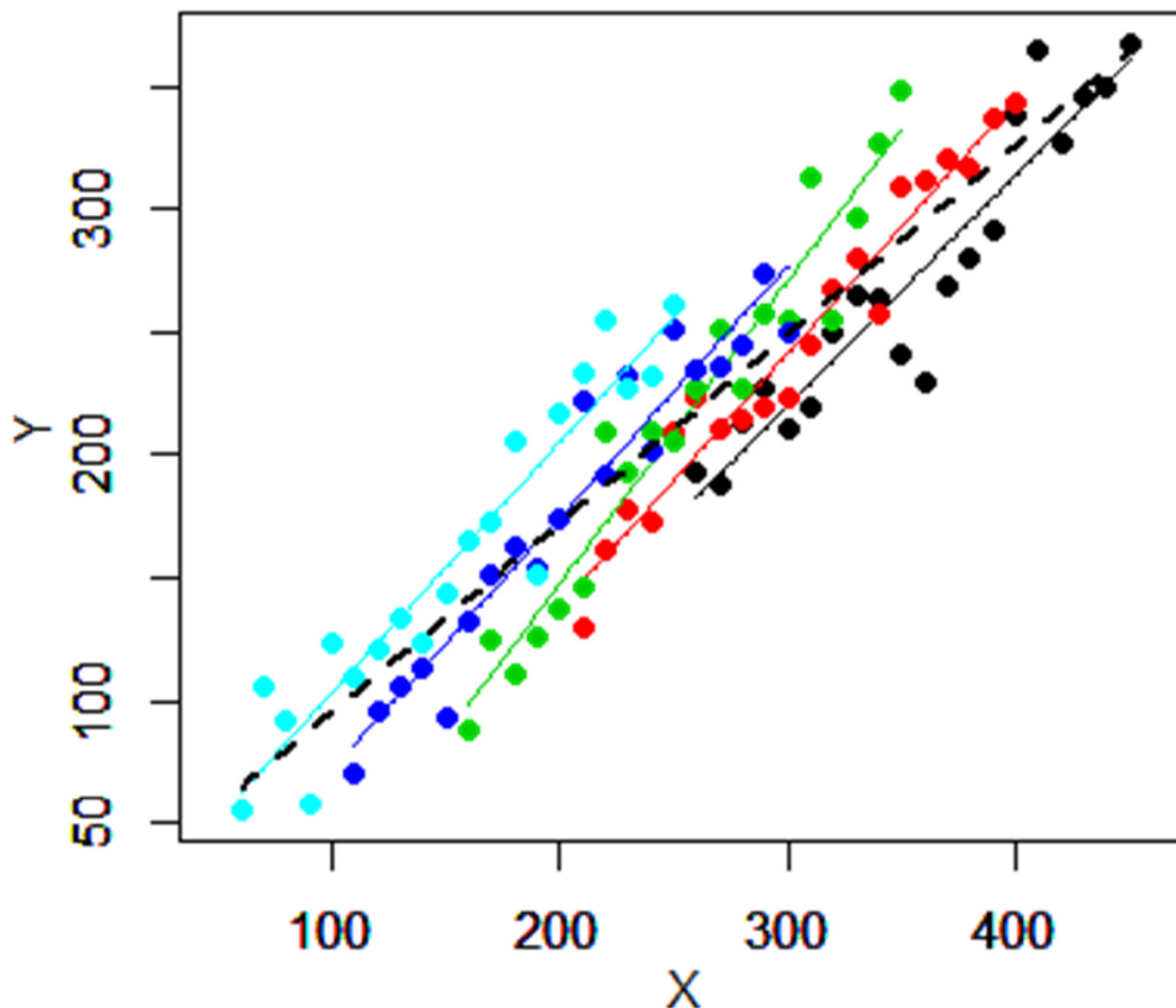
ところが、これらのデータは実は異なる5つの調査地(Site)からとったものだった。  
Site = {Site1, Site 2, Site 3, Site 4, Site 5}



全データを使った回帰直線(点線)は、  
特定の調査地内の傾向を表していないように見える。

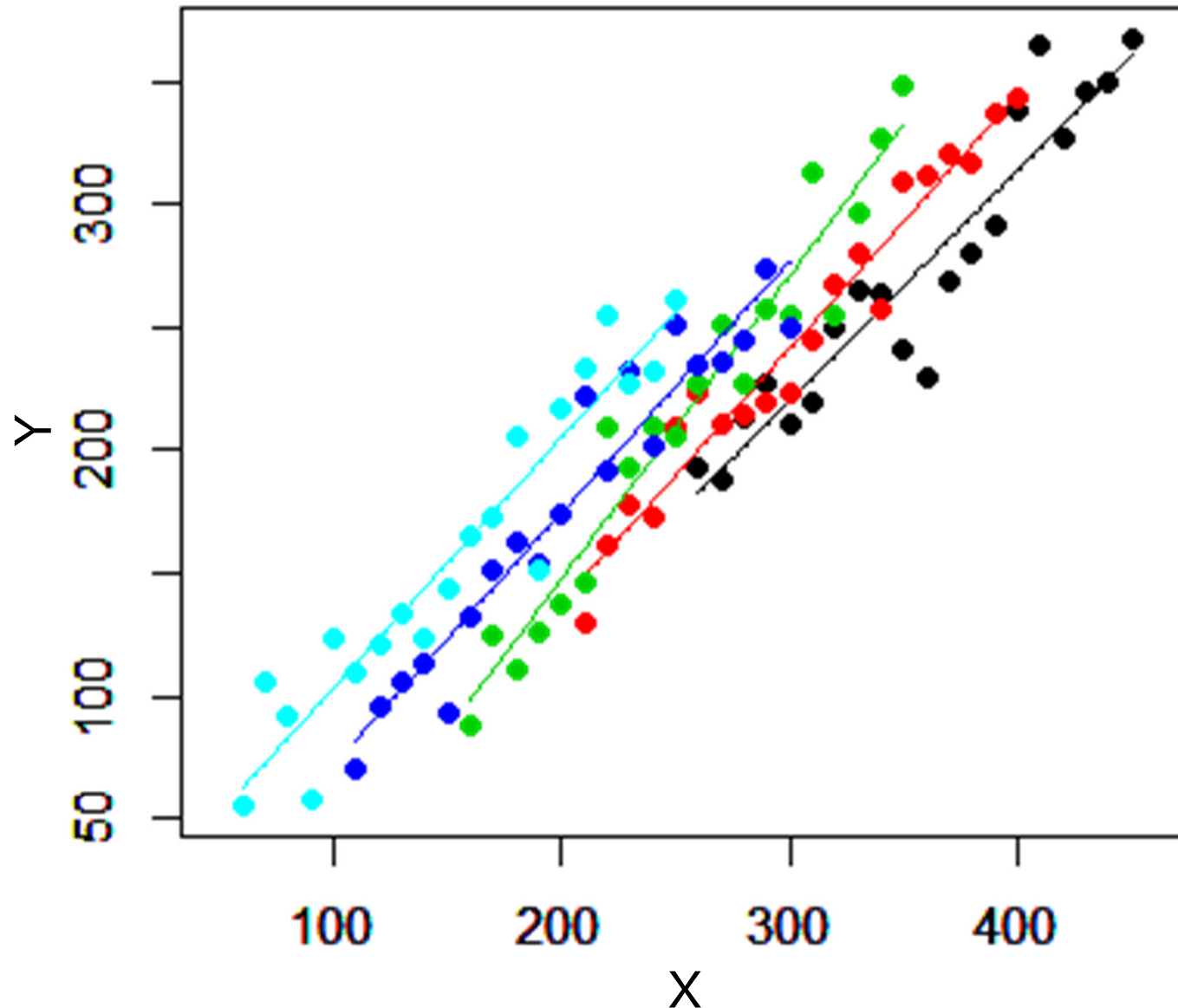


実際、個々の調査地ごとのデータを使って計算した調査地別の回帰直線(色つき実線)は、全調査地データの回帰直線(点線)と傾きが異なる。すなわち、全調査地のデータをプールした回帰直線は、個々の調査地内の傾向を表現していない。



ここで知りたい事は、任意の調査地が与えられたときにその内部でx-y関係がどうなるか、という事である。

もちろん、個々の調査地内の傾向は、調査地ごとに回帰関係を調べて表現すればすむことである。しかし、複数の調査地のデータ全体から、各々の調査地内の傾向がどのようなかについて、何か一般的な傾向を導きたい。個々の調査地内の関係にはあまり興味ない。



ここで知りたい事は、任意の調査地が与えられたときにその内部でx-y関係がどうなるか、という事である。

そこで R のパッケージ “lmerTest” の線形混合モデル用関数  
“lmer” をつかった混合モデル解析をやってみよう  
(R に興味のない人は説明文だけ読んでください)

# R の画面上での 混合モデルの計算命令の例 (式の見方)

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

計算結果を  
'mixedM' という  
変数名で保存する、と  
いう意味

「データyとデータ x の  
線形関係を、データ x, y  
が属するSiteごとに処理  
する」という指定

'XYdata' は  
データをいれた  
ファイル名

この例の場合、「全サイトのデータから回帰の切片や傾き  
が決まるが、**それで決められた切片や傾きは調査地 (Site)  
ごとにランダムに変動する**」と設定したモデルにしている。  
このランダム効果が入っているところが「混合モデル」



## 調査地"Site"をランダム効果にして混合モデルをやってみる

### 固定効果 (fixed effects) ・ ランダム効果 (random effects) とは？

(ランダム効果に指定する変数(この例ではSite)は質的変数!)

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

↑上の式は、回帰式を次のようなモデルに設定している

「固定効果で決められる回帰の切片や傾きは、すべての調査地で共通している。  
一方、各調査地の切片や傾きは、調査ごとに決まるランダムな変動が  
固定効果で決められる切片や傾きに加わることによって決まる」

$$y = (\text{固定効果の切片} + \text{調査地ごとのランダムな切片の変動}) \\ + (\text{固定効果の傾き} + \text{調査地ごとのランダムな傾きの変動}) \times X \\ \rightarrow \text{すなわち、各係数に固定効果とランダム効果が入るのが混合モデル}$$

例) 混合モデル解析の結果、固定効果 Fixed effects だけを表記した例

$$y = \underline{-48.82} + \underline{1.04} \times X \\ \uparrow \text{固定効果の切片} \quad \uparrow \text{固定効果の傾き}$$

これに対し、各調査地の切片と傾きは、固定効果+ランダム効果の両方を使って、..

$$y = (-48.82 + \text{調査地ごとのランダムな切片の変動}) + (1.04 + \text{調査地ごとのランダムな傾きの変動}) \times X$$

例) 調査地3(緑の点と回帰直線)の場合

$$y = (-48.82 + (-28.4)) + (1.04 + 0.11) \times X = \underline{-77.2} + \underline{1.15} \times X \\ \text{調査地3の切片の変動} \quad \text{調査地3の傾きの変動} \quad \text{調査地3の回帰直線}$$

## # R の画面上での 混合モデルの計算命令の例 (式の見方)

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

→ 回帰式の切片や傾きに, Site=1, Site=2,,, の違いが反映される。

### 追加解説：ランダム効果 (random effects) とは？

ランダム効果に指定する変数は、グルーピング(グループの違い)を表す **質的** 変数(上の例ではSite)である。

したがって, Site に 1, 2, ... という変数が入っていても, 質的変数として扱われる。



量的変数(下の測定値Zなど)をランダム効果に指定してはいけない(ひとつひとつの値が一つのグループとして計算されてしまうため)

#### データの例

X	Y	Site	Z
2.3	4.1	A	6.5
4.2	2.0	A	2.9
...			
6.4	8.2	B	1.6

## 調査地"Site"をランダム効果にして混合モデルをやってみる

```
mixedM <- lmer(y ~ x + (x | Site), XYdata) # 'XYdata' はデータをいれたファイル名  
# ↑ Rで「データ x をSiteごとに処理する」という意味
```

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

このモデルは、「回帰の切片や傾きは調査地ごとにランダムに変動する」と設定したモデルにしている。

```
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
```

ランダム効果 random effects;  
ここでは切片および傾きがサイト間でどれだけばらついてたか、という情報だけが出力される

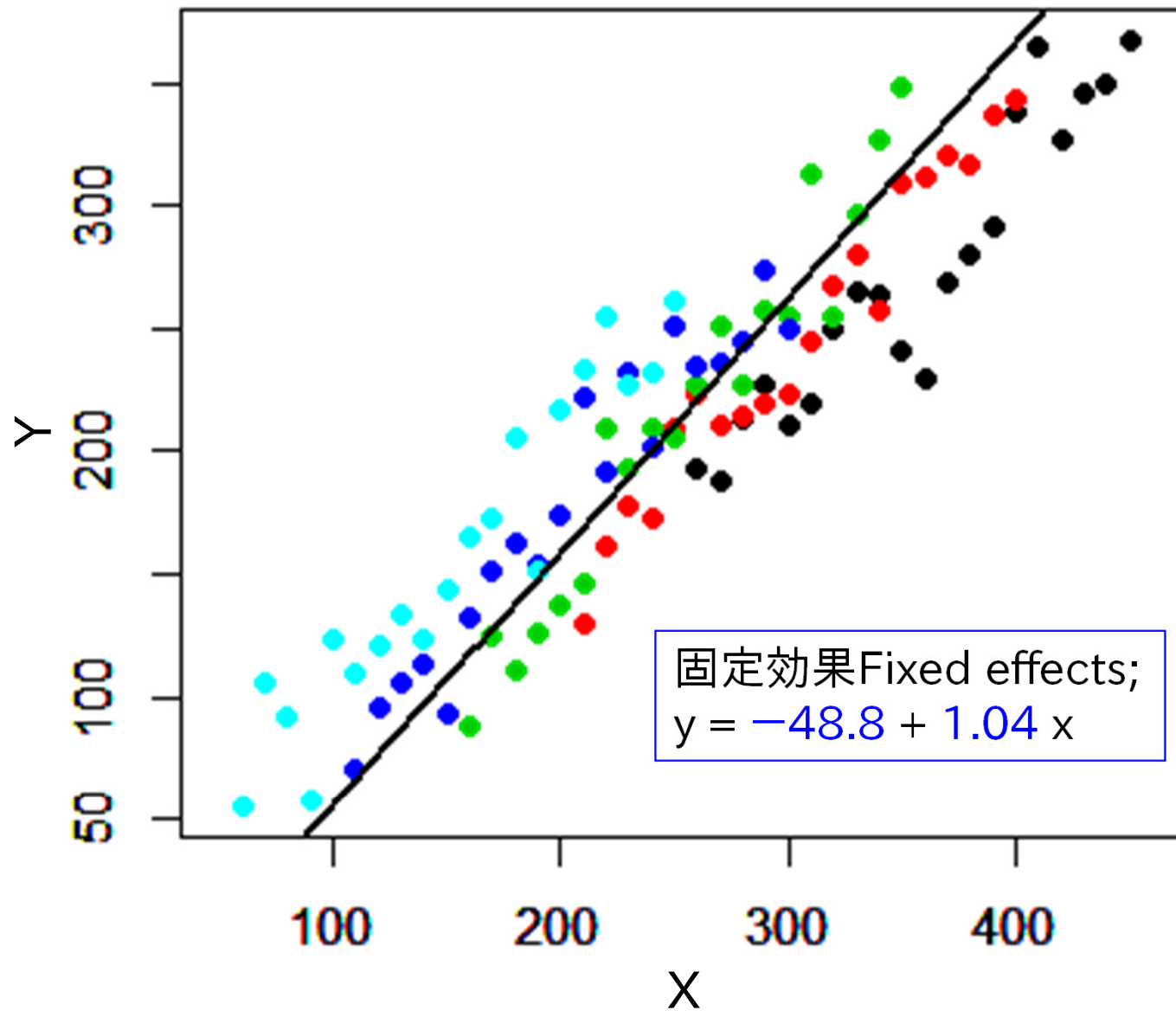
```
Fixed effects: ←-----  
Estimate Std. Error df t value Pr(>|t|)  
(Intercept) -48.82200 16.09966 6.21986 -3.032 0.022 *  
x 1.03949 0.04949 4.54485 21.003 1.06e-05 ***
```

固定効果 Fixed effects;  
 $y = -48.8 + 1.04 x$

} #←解説は以下のスライドに

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

混合モデルの結果の固定効果 Fixed Effectsのパラメタだけを使った回帰直線



## # ランダム効果Random effect の出力について:

ここは読み飛ばしてもかまいません

```
mixedM <- lmer(y ~ x + (x | Site), XYdata) # 線形混合モデルの例, つづき
```

# 線形混合モデルの結果から固定効果の回帰パラメータを取り出す方法

```
ALLA <- fixef(mixedM)[1] # 回帰の切片をALLAという名で保存;          ALLA = -48.8  
                                # 前の "GLMM の出力のスライド" 参照      ↑ ↓  
ALLB <- fixef(mixedM)[2] # 回帰の傾きをALLBという名で保存;      ALLB = 1.04
```

固定効果Fixed effectsは  $y = -48.8 + 1.04x$

```
> ranef(mixedM) # ↓各サイトごとのランダム効果パラメータの出力のさせかた:  
$Site # これらに固定効果の切片や傾きを足すと各調査地の値になる  
(Intercept) x(=傾きの変動)  
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  
↑ 調査地番号
```

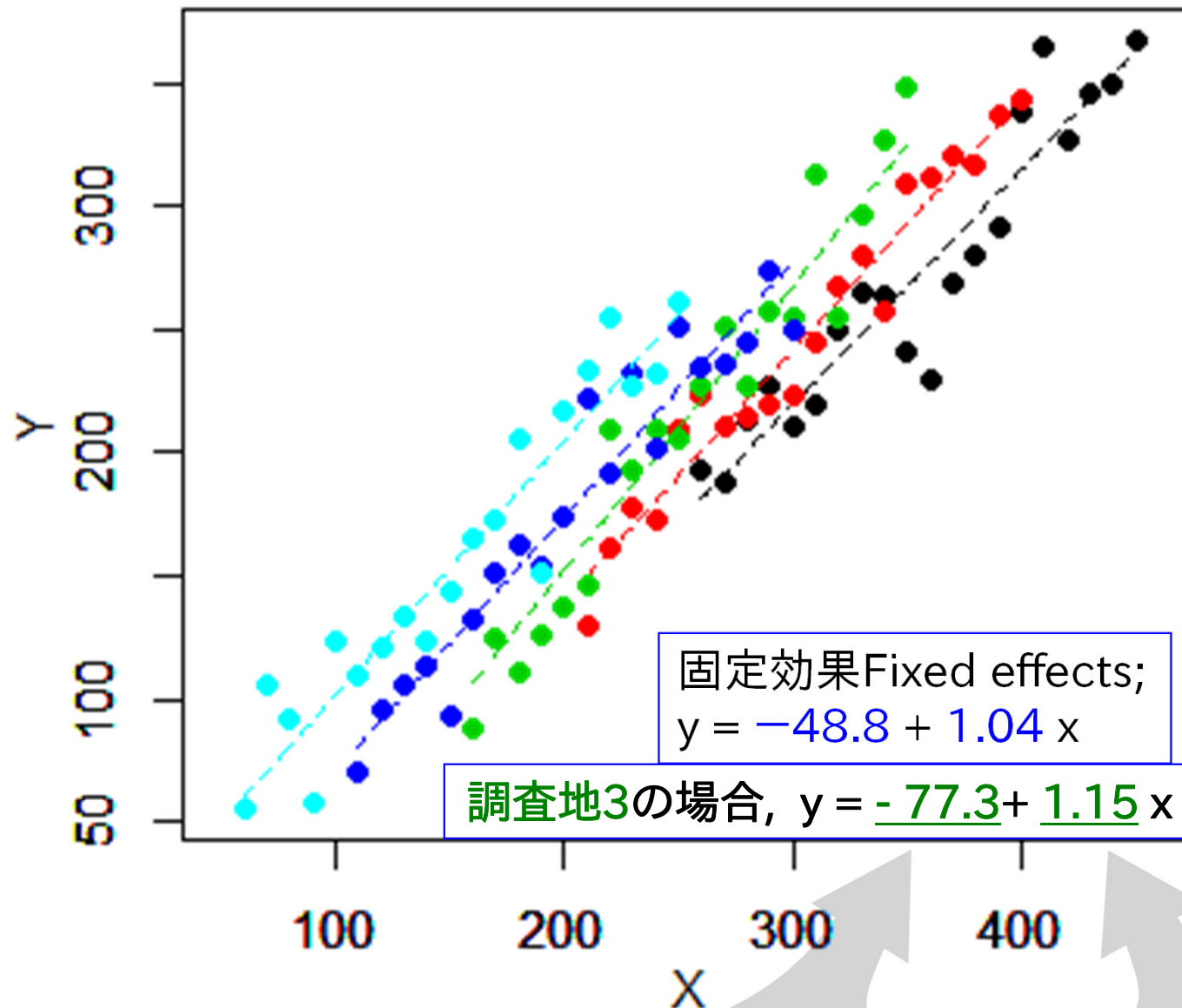
各調査地の回帰直線パラメータ =  
固定効果のパラメータ(青) + ランダム効果のパラメータ(茶)

# 結果の読み取り方の例;たとえば調査地3の場合,

```
AA <- ALLA + ranef(mixedM)$Site[3, 1] # 切片 AA = -48.8 + (-28.4) = -77.3  
BB <- ALLB + ranef(mixedM)$Site[3, 2] # 傾き BB = 1.04 + 0.11 = 1.15
```

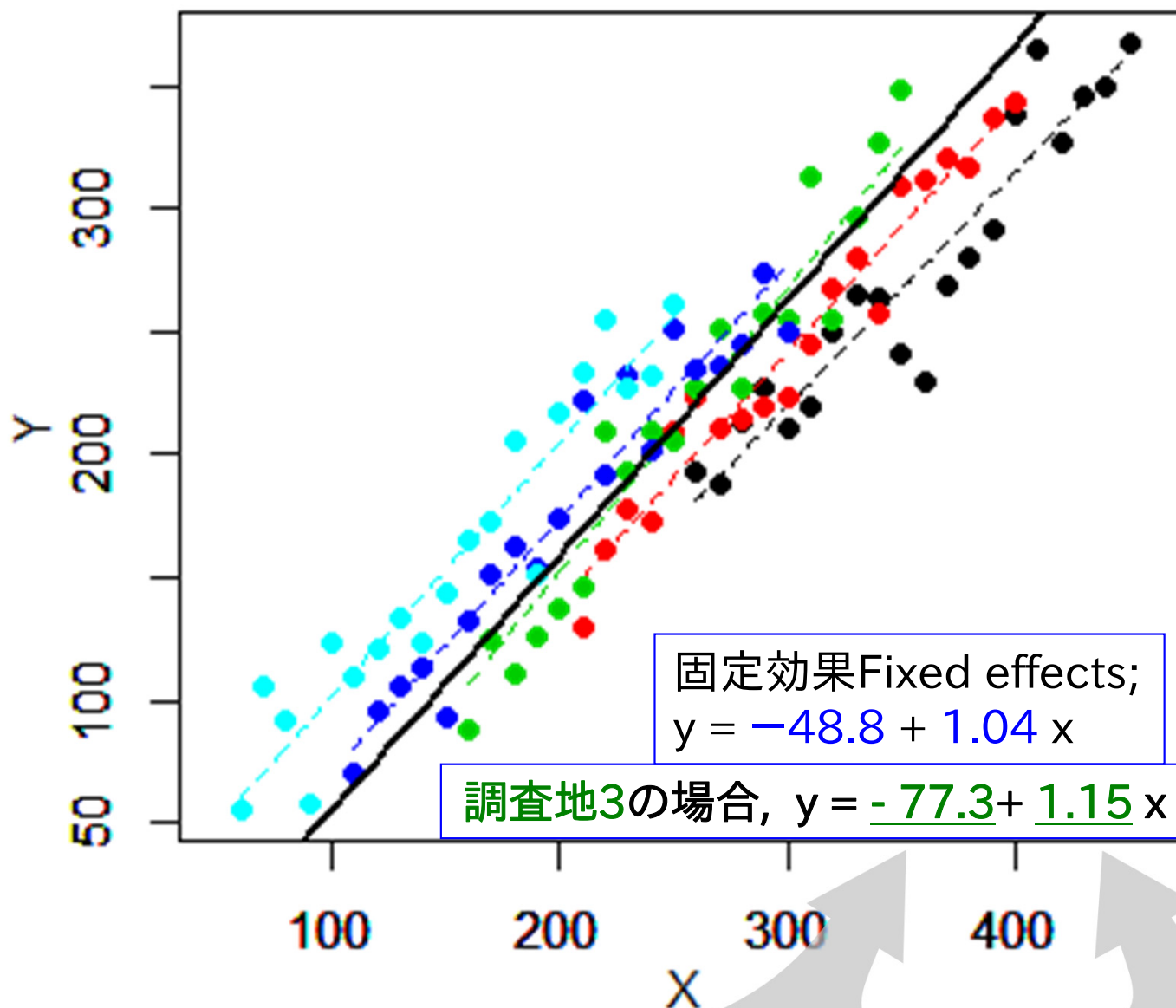
よって調査地3の場合,  $y = -77.3 + 1.15x$

固定効果のパラメタと、調査地ごとのランダム効果のパラメタの両方から計算した  
5つの調査地の回帰直線; **緑の線が調査地 3**



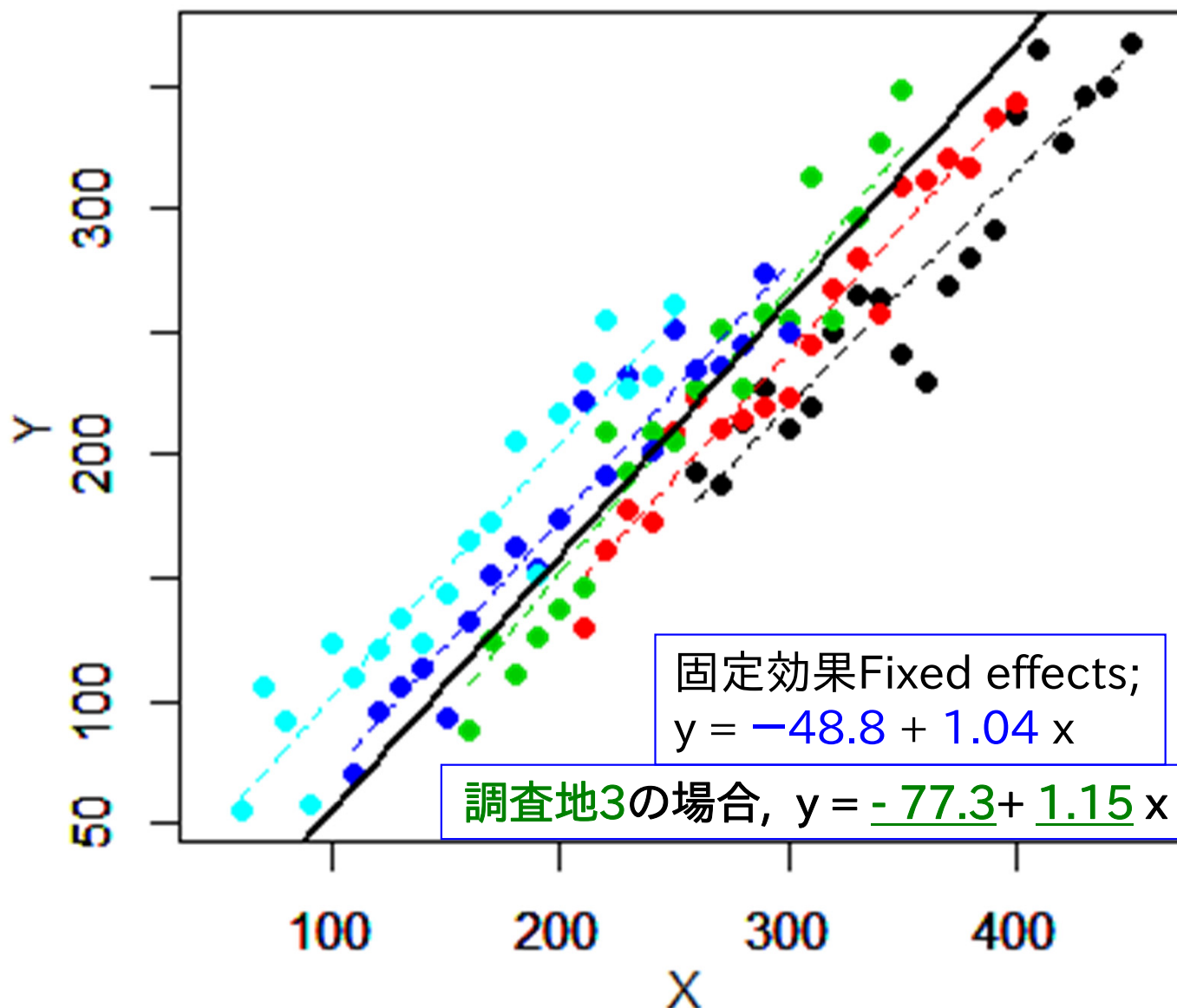
調査地 3の場合,  $y = (-48.8 + (-28.4)) + (1.04 + 0.11)x$

固定効果の回帰直線 (黒実線)、および、  
ランダム効果の結果も考慮した各調査地の回帰直線(点線)



調査地 3の場合,  $y = (-48.8 + (-28.4)) + (1.04 + 0.11) x$

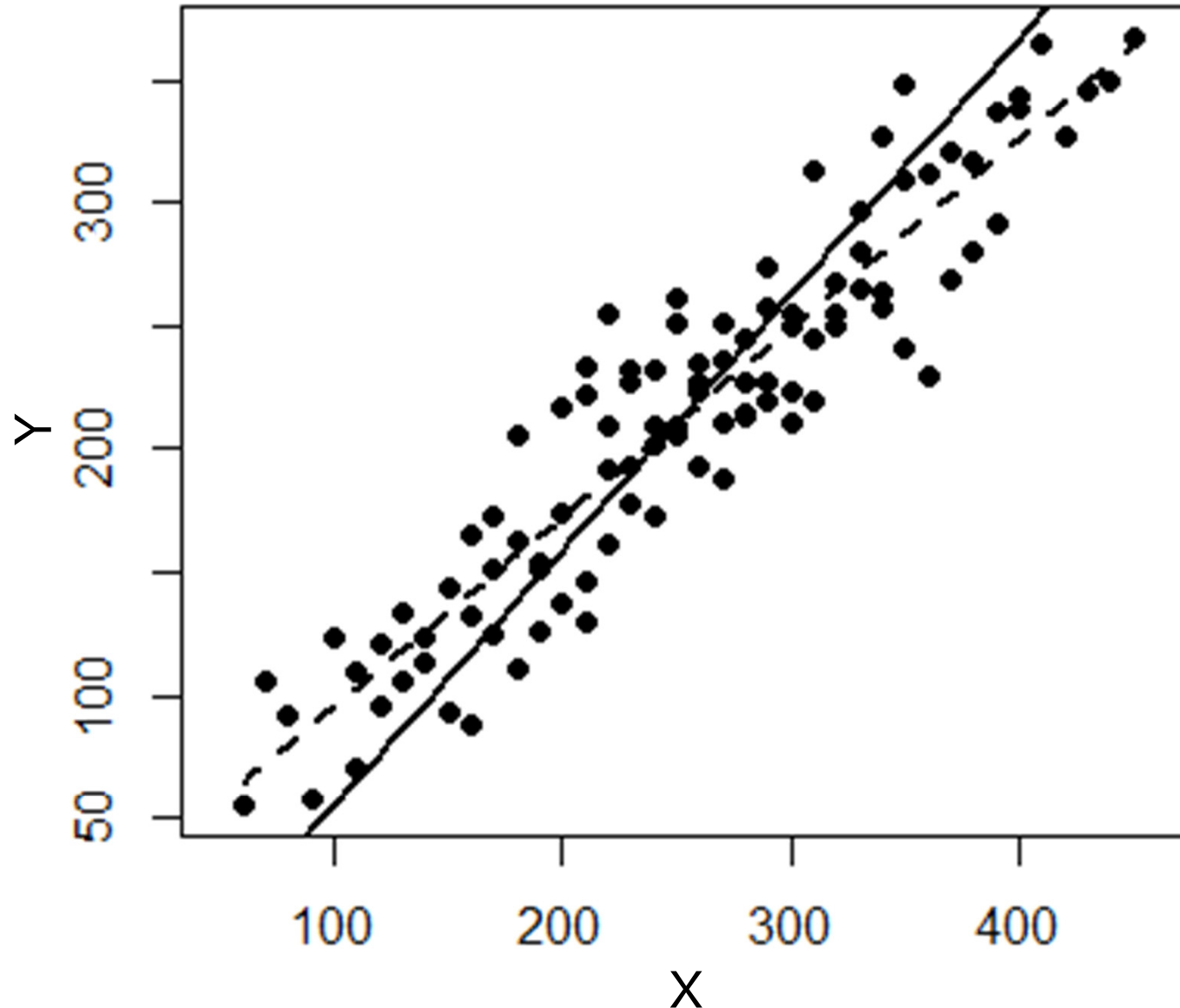
固定効果の回帰直線 (黒実線)、および、  
ランダム効果の結果も考慮した各調査地の回帰直線(点線)



固定効果による回帰直線 (黒実線)は、任意の調査地の中でのx-y関係を代表的に表している。  
ただし調査地間では回帰の切片や傾きがランダムにばらつく。

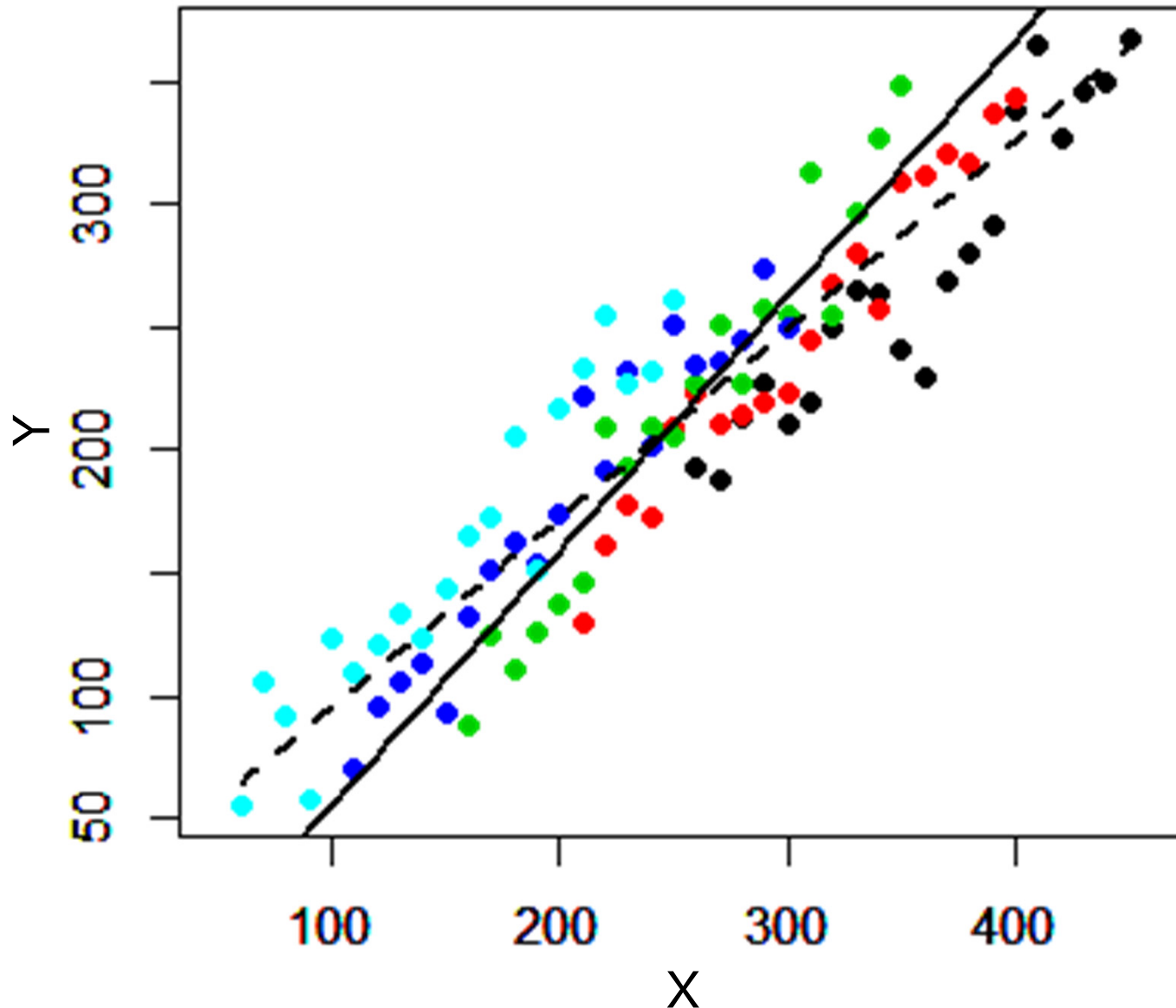


混合モデルの結果が何を意味するのかを理解するため、再度、最初の図にもどってみる。ここで、全調査地のデータをプールした回帰直線(点線)と混合モデルの固定効果による回帰直線(実線)を比べてみよう。



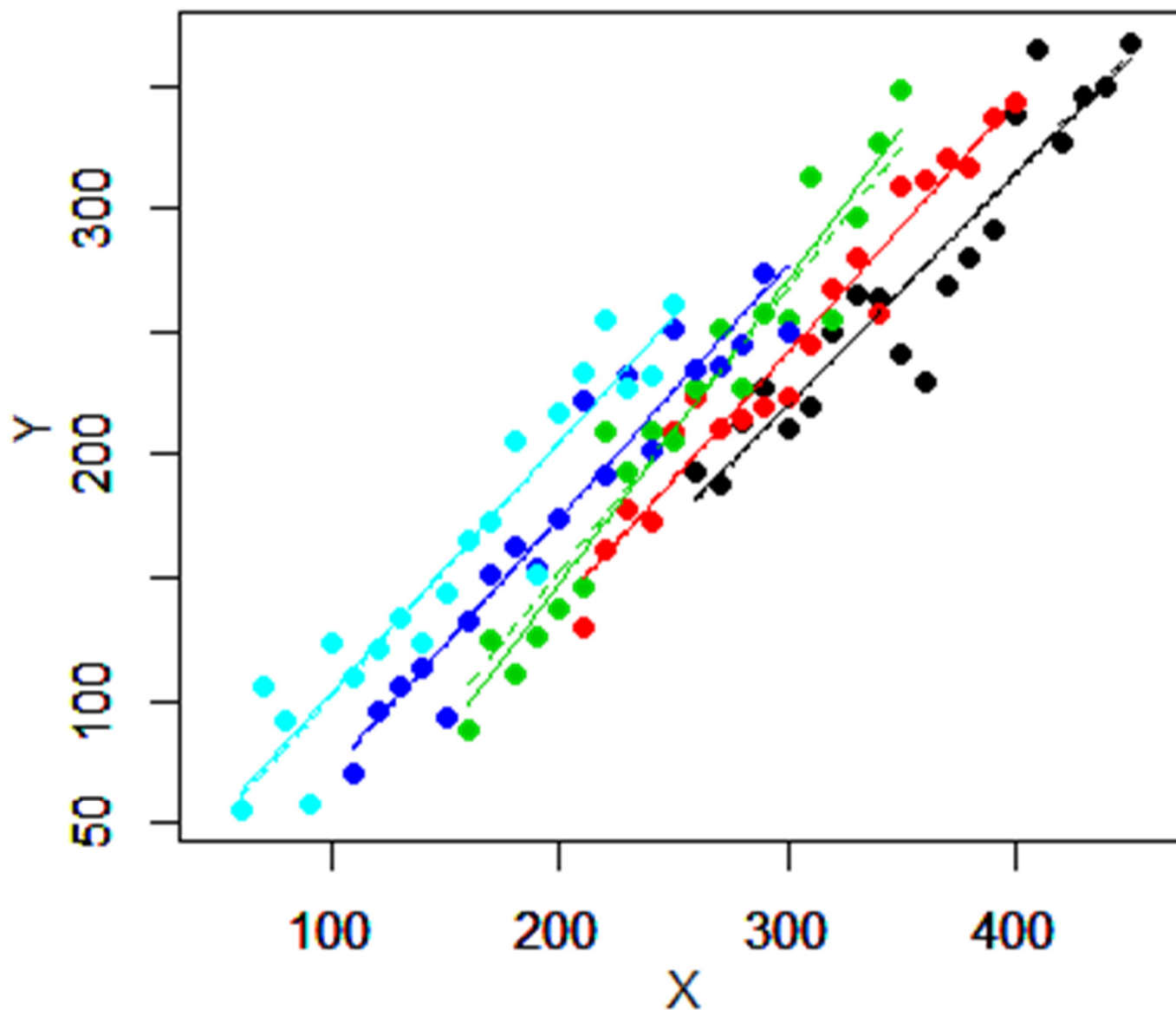
全調査地のデータをプールした回帰直線(点線)は、1つの調査地の中での傾向を考慮していないことを説明した。すなわち...

混合モデルの結果が何を意味するのかを理解するため、再度、最初の図にもどってみる。ここで、全調査地のデータをプールした回帰直線(点線)と混合モデルの固定効果による回帰直線(実線)を比べてみよう。



混合モデルの固定効果による結果は、1つの調査地が与えられたとき、その調査地内部での傾向を代表的に表しているのであって、全データをプールした時の傾向を表しているのではない。

ちなみに、この例では、混合モデルの結果から計算した各調査地の回帰直線(点線)は、個々の調査地ごとに計算した回帰直線(実線)と非常によく合っていた。



## まとめ

ランダム効果を考慮しないで全調査値のデータをプールした回帰は、  
個々の調査地の中での傾向をうまく表すことができなかった。

今回の例では、調査地をランダム効果とする混合モデルで取り扱った。この混合モデルでは、調査地間で回帰の切片や傾きがランダムに変動すると仮定したモデルにした。

混合モデルを使うと、1つの調査地の内部での代表的な傾向を知ることができるばかりでなく、ランダム効果の出力から、調査地間でその関係がどれくらいばらつくかを同時に知る事ができる。

## おまけ

```
> summary( mixedM ) # ↓[R] の出力:
```

```
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
```

(以下省略)

混合モデルの出力は、各調査地のランダム効果の切片 (**intercept**) と傾き (**x**) との間に弱い相関 (**Corr**; 相関係数-0.34) があることを示している。つまり、調査地の間で、切片と傾きとは完全に独立ではない。強い相関がある場合は over-parameterization (パラメタが多すぎ) の状態なので、切片あるいは傾きのどちらかのみをランダム効果に入れる。わざと切片と傾きとが独立に決まるように指定させたりもできる。

### いろいろな指定の仕方

(1) 調査地ごとに回帰させるモデル

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

(2) 調査地ごとに切片だけランダム効果にし、傾きは固定させるモデル

```
mixedM <- lmer(y ~ x + (1 | Site) )
```

(3) 調査地ごとに傾きだけランダム効果にし、切片は固定させるモデル

```
mixedM <- lmer(y ~ x + (0 + x | Site) ) # ほとんど無意味のような...
```

(4) 調査地ごとに傾きと切片とが独立に決まるようなランダム効果にするモデル

```
mixedM <- lmer(y ~ 1 + x + (1 | Site) + (0 + x | Site) )
```

# 方法の比較: 今回のデータの場合, (4)と(1)のモデルの結果は

ほとんど同じでした

```
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
 Site     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
```

# 方法の比較: 今回のデータの場合, (2)と(1)のモデルの結果も ほとんど同じでした

```
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(mixedM1)
$Site
 (Intercept)
1 -46.3432089
2 -19.7831275
3  0.9294691
4 16.9604568
5 48.2364105
> AIC(mixedM1)
[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 -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
```

## おまけのおまけのきしゃぽっぽ

この解説に使用したデータは、全調査地のデータをプールした回帰の傾きと各調査地の回帰の傾きが違うようにするため、各調査地のXの値の平均値を50ずつずらしてつくったものです。

実際にこのようなデータを得たら、各調査地ごとの回帰の切片とその調査地のXの平均値との間に相関がある(Xの平均値が小さいほど切片は大きくなる)、というモデルをまずやってみるのが普通でしょう。その場合には混合モデルは必要ありません。

この解説では、混合モデルがどんなものかを示すために、混合モデル関数を使ってみました。

ぽっぽー

