Ricard Garcia Isern
2023

MIXED EFFECT MODELS

1. INTRODUCTION

- Standard regression assumes → **uncorrelated errors.**

- **Correlated errors:**
    - Student grades when students are grouped in schools.
    - Biochemical markers of individuals inside families.
    - Treatment variables of patients inside hospitals.
    - Repeated measurements of individuals over time.

- **MIXED EFFECT MODELS → CORRELATION AMONG OBSERVATION IN CLUSTERS.**

2. RANDOM INTERCEPT MODEL

- Let y_{ij} represent observation j in cluster i
- Random intercept model:

  change β_0

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \underline{u_i} + \varepsilon_{ij} \qquad i = 1, \ldots, M \quad j = 1, \ldots, n_i$$

- with

  RANDOM EFFECT

$$u_i \sim N(0, \sigma_u^2) \qquad \varepsilon_{ij} \sim N(0, \sigma^2) \qquad u_i, \varepsilon_{ij} \text{ independent}$$

  → centered at 0!

- β_1 is a fixed effect, u_i is a random effect.

  β_0
  β_-

  → residuals

$$V(y_{ij}) = V(u_i + \varepsilon_{ij}) = \sigma_u^2 + \sigma^2$$

- The correlation between two error terms of the same individual, the intraclass correlation, is

$$Cor(u_i + \varepsilon_{ij}, u_i + \varepsilon_{ik}) = \frac{\sigma_u^2}{\sigma_u^2 + \sigma^2}$$
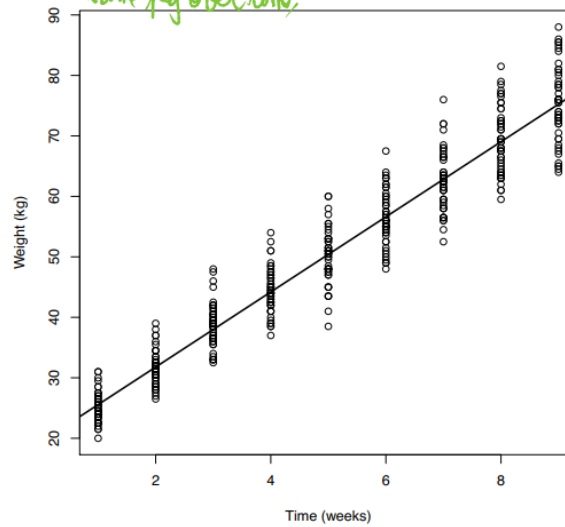
- Alternatively,

$$y_{ij} = \alpha_i + \beta_1 x_{ij} + \varepsilon_{ij} \quad \Rightarrow \quad \alpha_i = \beta_0 + u_i \qquad \alpha_i \sim N(\beta_0, \sigma_u^2)$$

- HOW TO FIND **B:**
    - **Maximum likelihood estimation → ML**
    - **Restricted maximum likelihood estimator → REML.**
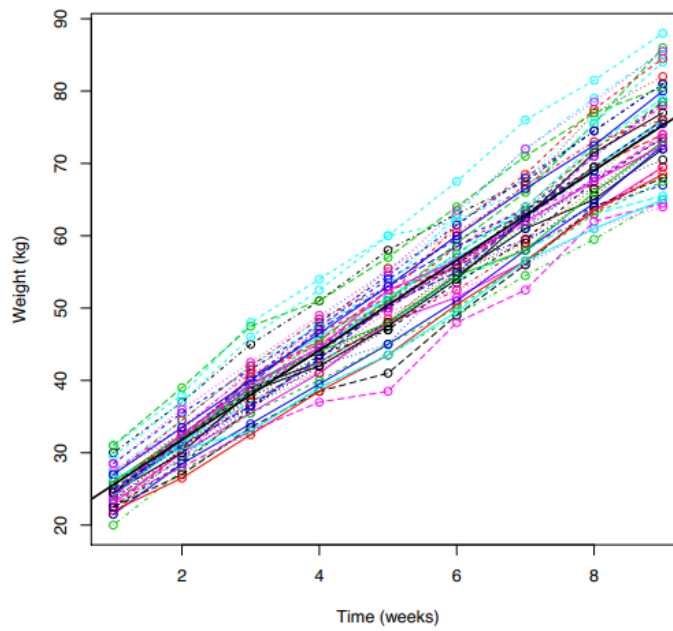   - We need a variance.

OLS regression

↳ ordinary

here we ignore the fact that they are related, they are the same pig over time.

Assessing fit graphically

* We observe linear regression over all the pigs.
* We don't see a lot of difference amongst all the pigs, small ones keep smaller.
* There's correlation between the pigs

```
> library(nlme)
> model.1 <- lme(weight~time,data=Pigs,random=~1|subject)
> summary(model.1)
Linear mixed-effects model fit by REML
 Data: Pigs
        AIC       BIC    logLik
   2041.797 2058.052 -1016.898

Random effects:
 Formula: ~1 | subject
        (Intercept) Residual
StdDev:  3.891253   2.096356

Fixed effects: weight ~ time
             Value Std.Error  DF   t-value p-value
(Intercept) 19.355613 0.6031390 383  32.09146       0
time         6.209896 0.0390633 383 158.97012       0
 Correlation:
      (Intr)
time -0.324

Standardized Within-Group Residuals:
     Min         Q1         Med         Q3        Max
-3.73902210 -0.54562381  0.01835208  0.51221200  3.93133783

Number of Observations: 432
Number of Groups: 48
>
```

Handwritten annotations: intercept; how data is grouped; variance; $u \sim N(0, 3.89^2)$; random effect; residuals; when $t=0$, mean kg is 19.35; increase 6.2 kg for week

```
> intervals(model.1)
Approximate 95% confidence intervals

 Fixed effects:
                lower       est.      upper
(Intercept) 18.16974  19.355613  20.541492
time         6.13309   6.209896   6.286701
attr(,"label")
[1] "Fixed effects:"

 Random Effects:
  Level: subject
                     lower       est.     upper
sd((Intercept))   3.158269  3.891253  4.79435

 Within-group standard error:
    lower      est.     upper
 1.953029  2.096356  2.250202
```

Handwritten annotation: Interval

$$\hat{\rho} = \frac{(3.89)^2}{(3.89)^2 + (2.096)^2} = 0.775$$

Handwritten annotations: only affect β_0 each pig; u, when entered at 0, can modify the initial β_0 if the pig initial weight is not 19.35

$$\widehat{W} = 19.36 + u + 6.2 \cdot time$$

The fitted model

Handwritten annotation: GOOD MODEL LINEAR REGRESSION
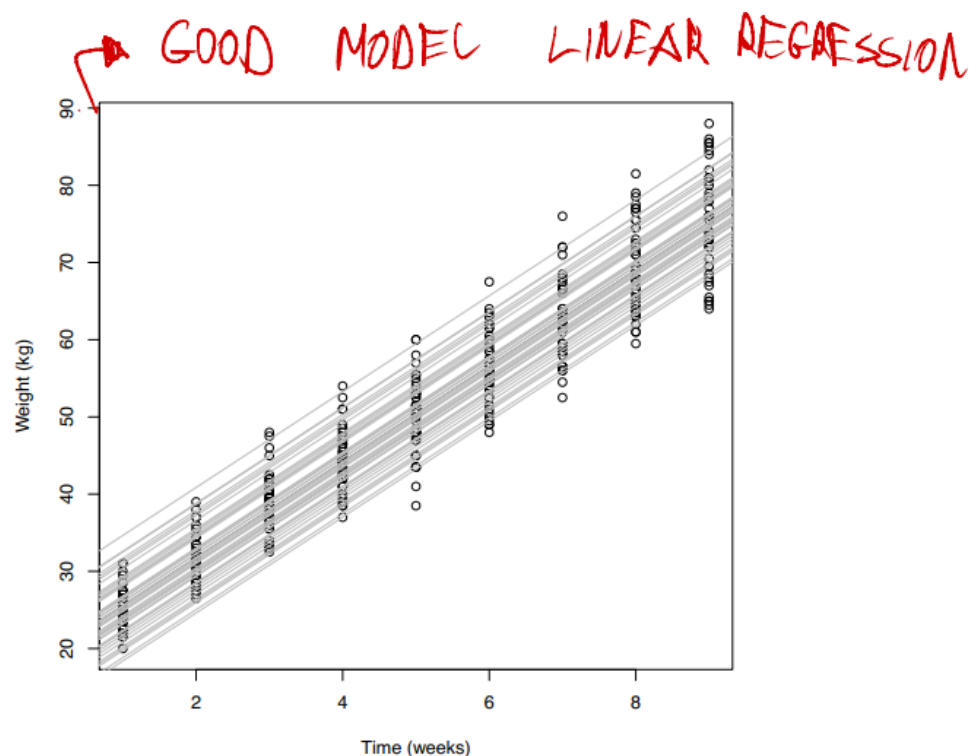
Residuals random intercept model
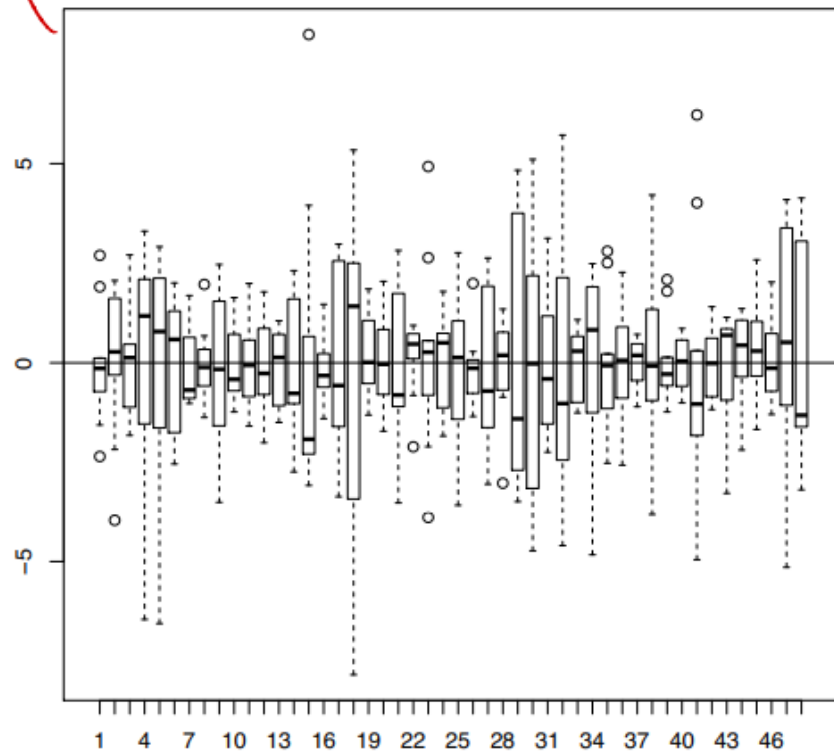
- **Residuals** →values observed vs values predicted.

COMPARING MODELS

- Let k be the difference in number of parameters between two models.
- Difference in deviance (likelihood ratio test) between general model L_2 and restricted model L_1

$$G^2 = 2\ln\left(\frac{L_2}{L_1}\right) = 2\ln(L_2) - 2\ln(L_1) = D_1 - D_2 \sim \chi_k^2 \text{ under } H_0$$

- Akaike information criterion (AIC)

$$AIC = 2k - 2\ln\left(L(\hat{\theta})\right)$$

- Bayesian information criterion (BIC)

$$BIC = k\ln(N) - 2\ln\left(L(\hat{\theta})\right)$$

- Smaller AIC and BIC indicate better fit

- Two model comparison:

```
> anova(model.1,model.0)
        Model df      AIC      BIC    logLik   Test L.Ratio p-value
model.1     1  4 2041.797 2058.052 -1016.898
model.0     2  3 2512.945 2525.136 -1253.472 1 vs 2 473.148  <.0001
>
```

3. RANDOM SLOPE & INTERCEPT MODEL

we have random effect also in B_1

Random slope and intercept model

randome effect in slope (time in pigs)

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_i + \boxed{v_i x_{ij}} + \varepsilon_{ij} \qquad i = 1, \ldots, M \quad j = 1, \ldots, n_i$$

random effect in slope

$$u_i \sim N(0, \sigma_u^2), \quad \boxed{v_i \sim N(0, \sigma_v^2),} \quad \boxed{Cov(u, v) = \sigma_{u,v},} \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

Alternatively,

how $v_i \, v_i$ varies together (between random effects)

$$y_{ij} = \alpha_i + \gamma_i x_{ij} + \varepsilon_{ij}$$

with

$$\boxed{\alpha_i = \beta_0 + u_i} \quad \text{and} \quad \boxed{\gamma_i = \beta_1 + v_i}$$

$$\boxed{\alpha_i \sim N(\beta_0, \sigma_u^2) \quad \gamma_i \sim N(\beta_1, \sigma_v^2)}$$

Fitting the random slope and intercept model

?? what to put ??

```
* > model.2 <- lme(weight~time,data=Pigs,random=~time|subject)
* > summary(model.2)
  Linear mixed-effects model fit by REML
   Data: Pigs
        AIC      BIC    logLik
     1752.871 1777.254 -870.4356

> Random effects:
    Formula: ~time | subject
    Structure: General positive-definite, Log-Cholesky parametrization
               StdDev    Corr
  (Intercept) 2.6431920 (Intr)
  time        0.6164379 -0.063
  Residual    1.2636572

  Fixed effects: weight ~ time
                 Value Std.Error  DF  t-value p-value
  (Intercept) 19.355613 0.4038676 383 47.92564       0
  time         6.209896 0.0920382 383 67.47085       0
   Correlation:
       (Intr)
  time -0.133

  Standardized Within-Group Residuals:
         Min         Q1        Med         Q3        Max
  -3.62018844 -0.54735954  0.01503617  0.54855117  2.99391406

  Number of Observations: 432
  Number of Groups: 48
```

intercept between random effects

if close to 0, we don't see any pattern for pigs under mean.

- Different example → **PENGUINS.**

```
model.3 <- lme(logHeart~Depth+Duration,data=X,random=~Depth|Bird)
> summary(model.3)
Linear mixed-effects model fit by REML
  Data: X
      AIC       BIC   logLik
  -52.7974 -33.16925 33.3987

Random effects:
 Formula: ~Depth | Bird
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev       Corr
(Intercept) 1.480502e-01 (Intr)
Depth       7.838200e-08 0
Residual    1.531288e-01

Fixed effects:  logHeart ~ Depth + Duration
              Value   Std.Error  DF  t-value  p-value
(Intercept)  4.560688 0.05872885 114  77.65669  0e+00
Depth        0.001657 0.00043084 114   3.84527  2e-04
Duration    -0.100821 0.00395385 114 -25.49932  0e+00
 Correlation:
          (Intr) Depth
Depth    -0.183
Duration -0.310 -0.467

Standardized Within-Group Residuals:
      Min         Q1         Med         Q3         Max
-2.06274654 -0.63446745 -0.03841057  0.50174882  2.29256217

Number of Observations: 125
Number of Groups: 9
> anova(model.2,model.3)
       Model df    AIC       BIC   logLik  Test      L.Ratio p-value
model.2     1  5 -56.7974 -42.77729 33.3987
model.3     2  7 -52.7974 -33.16925 33.3987 1 vs 2 1.468885e-08       1
>
```

no correlation between random effects (handwritten annotation)