Advanced Mixed Effects Modelling

Mariel Finucane, PhD

Gladstone Bioinformatics Core

November 5, 2013

Goals

1. How to use *indicator variables* to describe a mixed effects model with multiple groups?
   (a) Write down the regression equation
   (b) Fit and plot the regression in `R`.
   (c) Interpret the regression output.
2. What to do when time trends are not linear?
3. What to do if your reviewer doesn't like the default method of calculating p-values?

Mixed = fixed + random

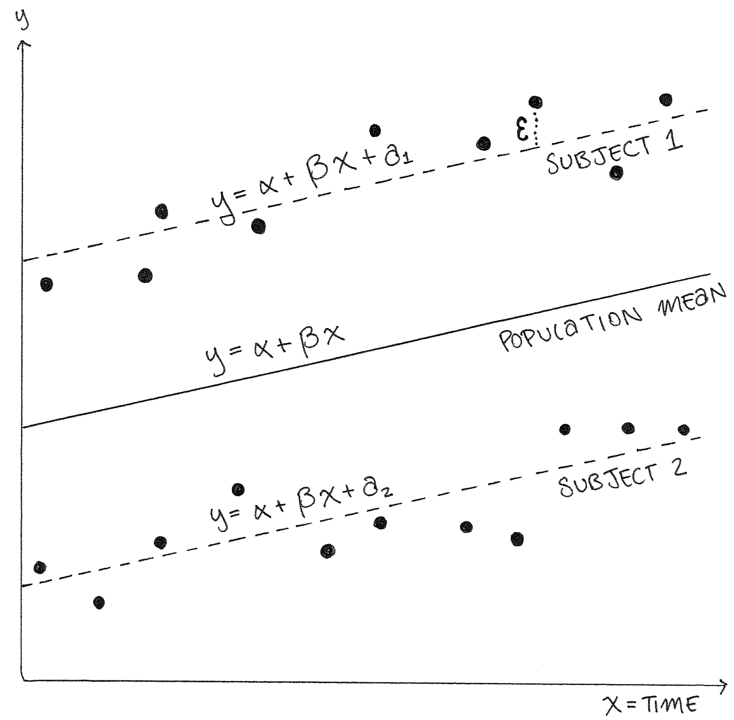
- Fixed effects (α, β)
  - Population characteristics
  - Shared by all individuals
  - Describe the mean response trajectory in the population
  - Useful to epidemiologists
- Random effects (a, b)
  - Subject-specific effects
  - Vary from one individual to another
  - Describe individuals' response trajectories
  - Useful to clinicians

The simplest case: a random intercept model

$$
\begin{aligned}
y &= \alpha + \beta x + a + \varepsilon \\
&= (\alpha + a) + \beta x + \varepsilon
\end{aligned}
$$

- α is the fixed, population-level intercept
- β is the fixed, population-level time slope
- a is the random, subject-specific intercept:
  By how much does each individual deviate from the population average?
- ε is the error term

Random intercept model

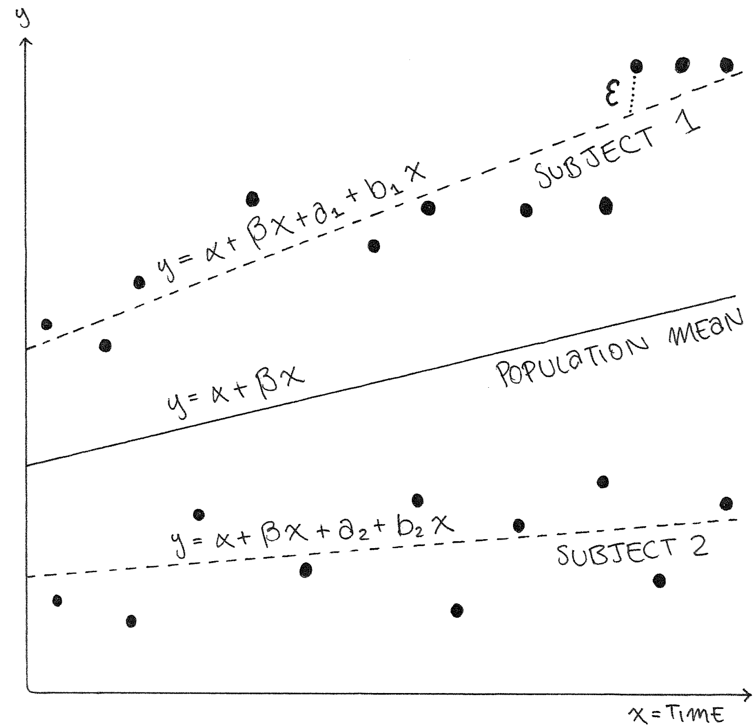


The more common scenario: a random intercept and slope model

$$\begin{aligned} y &= \alpha + \beta x + a + b x + \varepsilon \\ &= (\alpha + a) + (\beta + b) x + \varepsilon \end{aligned}$$

- α is the fixed, population-level intercept
- β is the fixed, population-level time slope
- a is the random, subject-specific intercept:
  By how much does each individual deviate from the population average?
- b is the random, subject-specific time slope:
  By how much does the effect of time on each individual deviate from the population-average effect?
- ε is the error term

Random intercept and slope model



To include multiple groups, we need indicator variables

Consider an experiment with **two groups** of interest:

- Some subjects receive a **placebo**,
- Others receive a **drug**.

To describe this experiment statistically, we use an **indicator** (a.k.a. **dummy**) variable:

$$z = \begin{cases} 0 & \text{if the subject received the placebo} \\ 1 & \text{if the subject received the drug} \end{cases}$$

An extension: two groups

Model	$y = \alpha + \beta x + \delta z + \gamma xz + a + bx + \varepsilon$ $= (\alpha + \delta z + a) + (\beta + \gamma z + b)x + \varepsilon$
Placebo	$y = (\alpha + \delta 0 + a) + (\beta + \gamma 0 + b)x + \varepsilon$ $= (\alpha + a) + (\beta + b)x + \varepsilon$
Drug	$y = (\alpha + \delta 1 + a) + (\beta + \gamma 1 + b)x + \varepsilon$ $= (\alpha + \delta + a) + (\beta + \gamma + b)x + \varepsilon$

- α is the fixed, population-level intercept
- β is the fixed, population-level time slope
- δ is the fixed, population-level effect of the drug
- γ is the fixed, population-level effect of the drug *on the time slope*
- a is the random, subject-specific intercept
- b is the random, subject-specific time slope effect?
- ε is the error term

A further extension: 3+ groups

Consider an experiment with **four groups** of interest (e.g. diets 1, 2, 3, and 4). To describe this experiment statistically, we use **3** indicator variables:

$$z_2 = \begin{cases} 1 & \text{if the subject received diet 2} \\ 0 & \text{otherwise} \end{cases}$$

$$z_3 = \begin{cases} 1 & \text{if the subject received diet 3} \\ 0 & \text{otherwise} \end{cases}$$

$$z_4 = \begin{cases} 1 & \text{if the subject received diet 4} \\ 0 & \text{otherwise} \end{cases}$$

Let's look at some data

```
> # install.packages('arm')
> library(arm)
> head(ChickWeight)
```

INSTALL AND LOAD A PACKAGE CALLED "arm" THAT CONTAINS SOME OF THE FUNCTIONS WE'LL BE USING.

```
Grouped Data: weight ~ Time | Chick
  weight Time Chick Diet
1     42    0     1    1
2     51    2     1    1
3     59    4     1    1
4     64    6     1    1
5     76    8     1    1
6     93   10     1    1
```

SHOW ME THE "head" OF THE DATASET CALLED "ChickWeight".

```
> attach(ChickWeight)
```

GIVE ME ACCESS TO THE VARIABLES IN THE DATASET "ChickWeight".

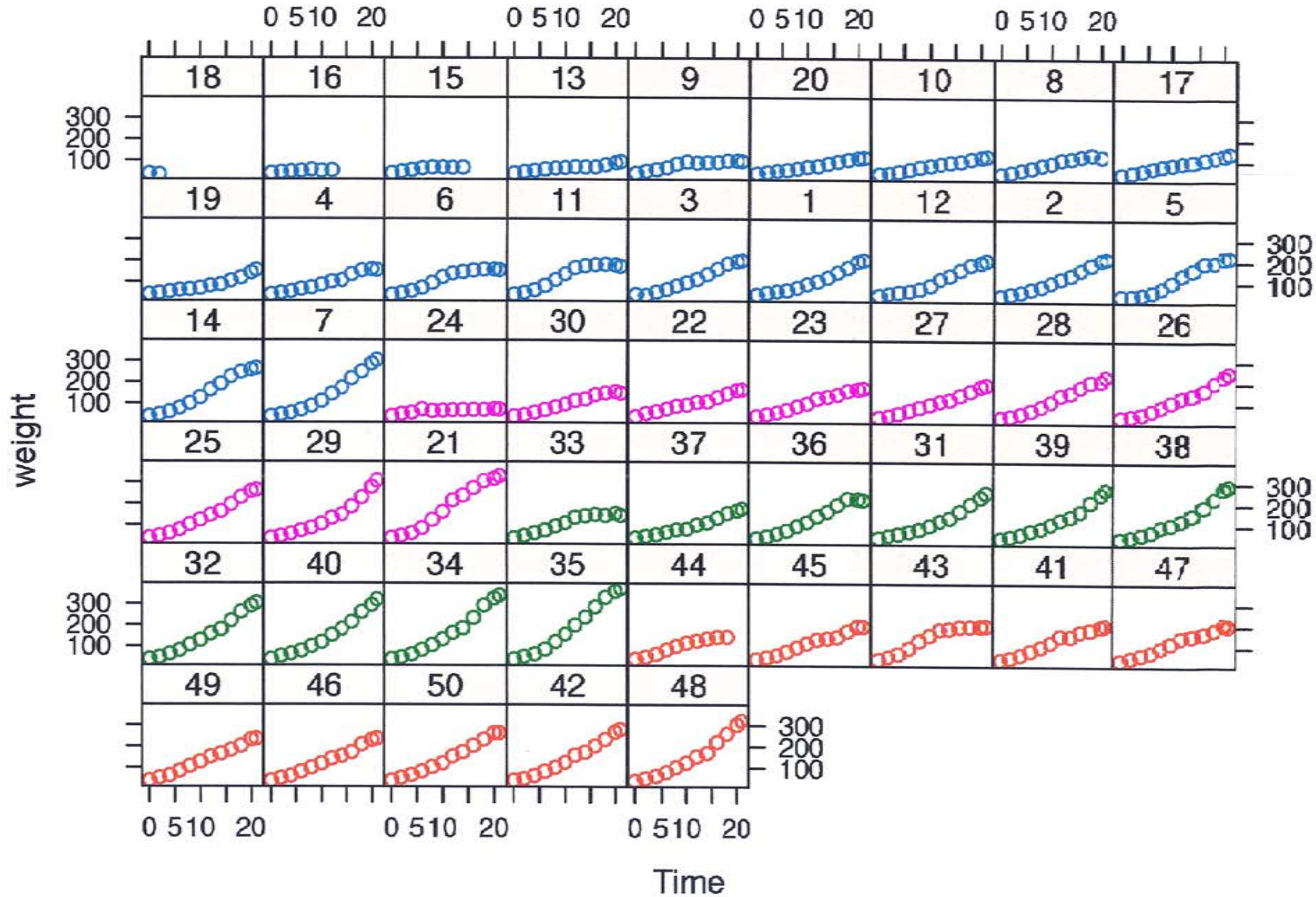
Let's look at some data

["FOR EACH"]

["VS."]

```
> xyplot(weight ~ Time | Chick, group = Diet, as.table = T)
```

[DRAW A "LATTICE" PLOT OF y VS x FOR EACH CHICK]



Fitting a mixed effects model

[CREATE AN OBJECT CALLED "model".]

```
> model <- lmer(weight ~ Time + Diet + Time*Diet + (1+Time|Chick))
```

$$y = \alpha + \beta x + \delta z + \gamma x z + a + bx + \varepsilon$$

[AND PUT INTO IT]

[A "L"INEAR MIXED EFFECTS REGRESSION.]

SHOW ME THE "FIXED" "EFFECTS" FROM "model".

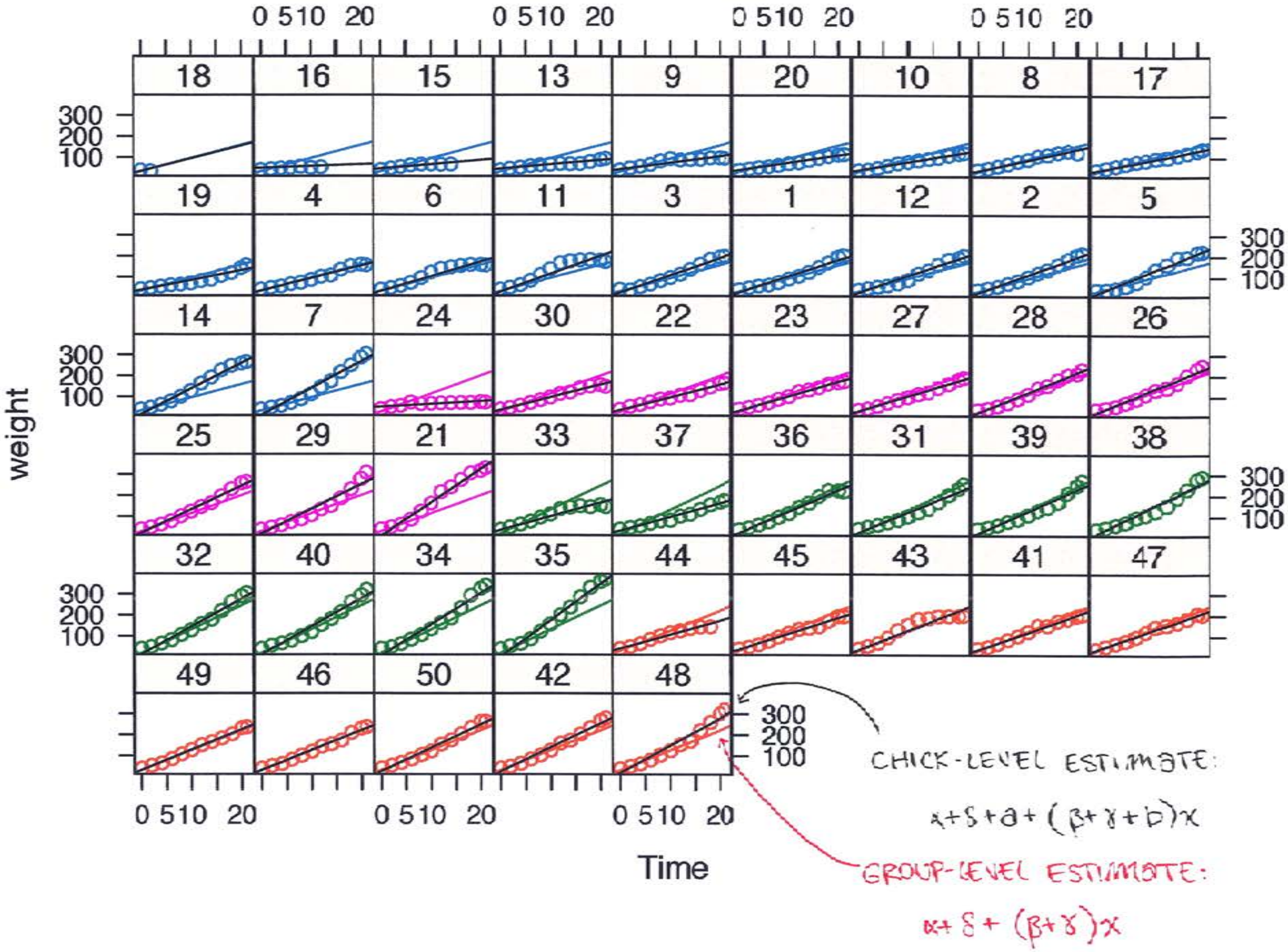
Extracting fixed/random effects

```
> fixef(model)

(Intercept)        Time       Diet2       Diet3       Diet4
     33.661       6.277      -5.028     -15.411      -1.750
 Time:Diet2  Time:Diet3  Time:Diet4
      2.332       5.146       3.255
```

SHOW ME THE "head" OF THE "CHICK-SPECIFIC" "RANDOM" "EFFECTS" FROM MODEL.

```
> head(ranef(model)$Chick)

   (Intercept)   Time
18      0.3389 -0.175
16     15.9678 -5.405
15     13.1031 -4.207
13     12.6833 -4.173
9      10.9972 -3.142
20      8.0702 -2.759
```

Plotting the fitted model



Summarizing the fitted model

```
> display(model)

lmer(formula = weight ~ Time + Diet + Time * Diet + (1 + Time |
    Chick))
            coef.est coef.se
(Intercept)  33.66 α    2.92
Time          6.28 β    0.76
Diet2        -5.03 δ2   5.01
Diet3       -15.41 δ3   5.01
Diet4        -1.75 δ4   5.02
Time:Diet2    2.33 γ2   1.30
Time:Diet3    5.15 γ3   1.30
Time:Diet4    3.25 γ4   1.31

Error terms:
 Groups   Name        Std.Dev. Corr
 Chick    (Intercept) 10.81
          Time         3.30    -0.97
 Residual             12.78
---
number of obs: 578, groups: Chick, 50
AIC = 4805.5, DIC = 4820
deviance = 4800.5
```

[STANDARD DEVIATION OF a's]
[STANDARD DEVIATION OF b's]

Summarizing the fitted model

```
> # install.packages('car')
> library(car)
> Anova(model)

Analysis of Deviance Table (Type II tests)

Response: weight
           Chisq Df Pr(>Chisq)
Time       314.1  1    < 2e-16 ***
Diet        15.6  3    0.00136 **
Time:Diet   17.1  3    0.00068 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

CALCULATED ACCORDING TO THE PRINCIPLE OF MARGINALITY, TESTING EACH TERM AFTER ALL OTHERS, EXCEPT IGNORING THE TERM'S HIGHER-ORDER RELATIVES

Interpreting the R output

- For your Methods section:
  - *"We fit a linear mixed effects model of chick weight on time, including fixed intercepts and slopes for each of the four diets, and random intercepts and slopes for each chick."*
  - *"We assessed differences in the time effect across diets using a Wald chi-square test."*
- For your Results section:
  - *"We found a significant difference in the rates of weight gain across diets ($p<0.001$), with chicks on diets 1, 2, 3, and 4 gaining 6.3, 8.6, 11.4, and 9.5 g/day, respectively."*
  - *"Chicks varied in their baseline weight (SD=10.8) and in their rates of weight gain over time (SD=3.3)."*

Postscript 1 – Nonlinear time effects

CREATE AN OBJECT CALLED model2.

```
> model2 <- lmer(weight ~ Time + I(Time^2) + Diet + Time*Diet +
+                    I(Time^2)*Diet + (1+Time+I(Time^2)|Chick))
```

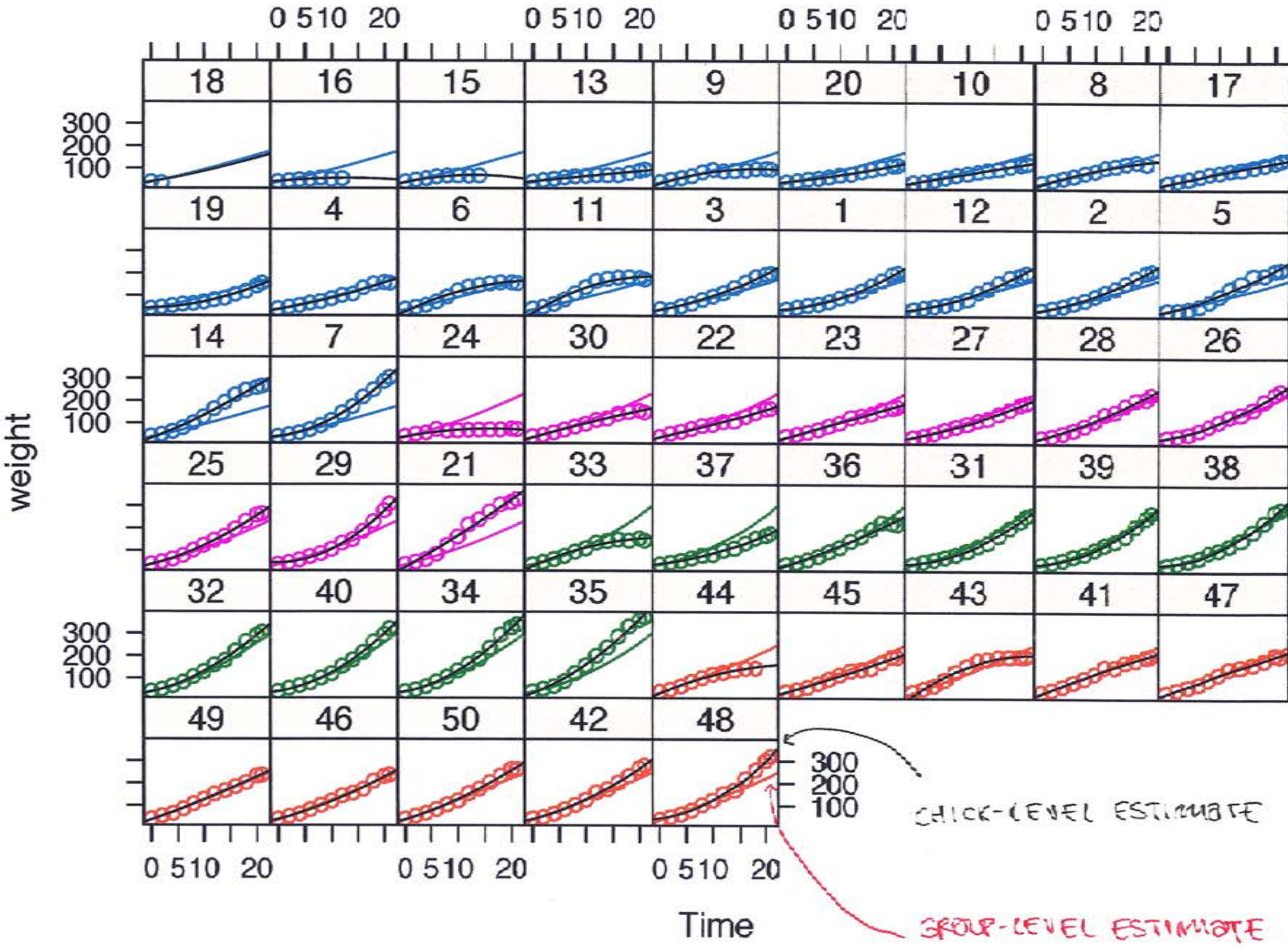
$$y = \alpha + \beta x + \varphi x^2 + \delta z + \gamma xz + \zeta x^2 z + a + bx + cx^2 + \varepsilon$$

Extracting fixed/random effects

```
> fixef(model2)

      (Intercept)                Time           I(Time^2)
        37.519476            5.123906            0.046931 φ
            Diet2               Diet3               Diet4
         0.160646            1.304480           -1.700774
       Time:Diet2          Time:Diet3          Time:Diet4
         0.686292           -0.066387            3.262591
I(Time^2):Diet2 I(Time^2):Diet3 I(Time^2):Diet4
         0.083358 φ2          0.249372 φ3          0.003115 φ4
```

```
> head(ranef(model2)$Chick)

    (Intercept)    Time  I(Time^2)
18       1.4061 -0.8994  0.0113795
16       4.8281 -3.0882 -0.1241312
15       0.3046 -0.1948 -0.2359140
13       4.2501 -2.7185 -0.0498864
9       -3.1273  2.0003 -0.2422971
20       3.7762 -2.4154  0.0001437
```

Plotting the fitted model



Postscript 2 – Bayesian Inference

- Most reviewers would be fine with the Wald chi-square test given here, but a small minority (myself included, to be honest) would argue that the Wald tests may be anti-conservative, especially for small datasets. (*Dempster, Rubin, & Tsutakawa (1981). Estimation in covariance components models. JASA, 76(374), 341-353.*)
- An alternative, which requires some extra statistical programming, is Bayesian inference.

Postscript 2 – Bayesian Inference

For example: "*Do chicks on diet 2 gain weight significantly faster than chicks on diet 1?*"

- For your Methods Section:
  - *"We obtained 5000 draws of the difference in time effects between diets 1 and 2, and we estimated a 95% confidence interval (CI) as the 2.5th and 97.5th quantiles of these draws. We calculated a p-value for this difference by inverting the simulated confidence interval."*
- For your Results Section:
  - *"Mice on diet 2 gain 2.3 more g/day than mice on diet 1 (95% CI -0.1, 4.9, p=0.07)."*

Postscript 2 – Bayesian Inference

```
> sims <- sim(model, 5000)
> fixefs <- fixef(sims)
> m <- fixef(model)['Time:Diet2']
> quantile(fixefs[,'Time:Diet2'], c(.025, .975))

   2.5%   97.5%
-0.1475  4.9553

> se <- sd(fixefs[,'Time:Diet2'])
> z  <- abs(m/se)
> p  <- (1-pnorm(z))*2
> p

Time:Diet2
   0.07363
```

[OBTAIN 5000 DRAWS, OR "SIMULATIONS".]

[ESTIMATE THE 95% CI.]

[INVERT THE CI TO OBTAIN A P-VALUE.]

Conclusions

If your outcome is continuous, but your observations are not independent:

1. Don't use repeated measures ANOVA!
2. Linear mixed effects models offer a flexible, powerful alternative.
3. You should now feel comfortable fitting a simple mixed effects model in R.
4. If your data are complex (e.g. multi-level hierarchy, extra-quadratic non-linearity) or your hypotheses are complex (e.g. comparisons at time x, predictions), please don't hesitate to drop me a line.