STA610 Lab05

Hun Kang

2024-09-27

- Write down your answers in any blank sheet and submit your work in paper during the lab.
- Your work will not be graded. As long as you submit, you will get a full credit.
- For those who missed the lab today, you can submit it via email to me for half credit.

Random vector and matrix

A random vector is a stack of random variables $X_1, \cdots, X_n$ put together in a vector

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

whose distribution is specified by a joint cdf

$$\begin{aligned} F(X) &= F_{X_1,\cdots,X_n}(x_1, \cdots, x_n) \\ &= \mathbb{P}(X_1 \leq x_1, \cdots, X_n \leq x_n) \end{aligned}$$

A random matrix is defined in a similar way.

Expectation

The expectation of a random vector and matrix is defined element-wise. That is,

$$(\mathbb{E}[M])_{ij} = \mathbb{E}[M_{ij}]$$

Since $\mathbb{E}$ is a linear operator, it is clear that

$$\mathbb{E}(AX + b) = A\mathbb{E}(X) + b$$

Covariance

The covariane matrix of a random vector X is defined as

$$\Sigma = \mathbb{E}\left[(X - \mathbb{E}X)(X - \mathbb{E}X)^T\right]$$

Q1-1: Show that $\Sigma_{ij} = Cov(X_i, X_j)$

Q1-2: Show that $\Sigma = \mathbb{E}[XX^T] - \mathbb{E}X(\mathbb{E}X)^T$

Q1-3: Write the covariance matrix of $Y = AX + b$

A symmetric matrix A is **positive semi-definite** if and only if

$$u^T Au \geq 0 \quad {}^\forall u \in \mathbb{R}^n$$

and **positive definite** if the inequality is strict. To put it intuitively, it means the angle between any vector u and its linear transformation Au is acute, below 90 degree.

Q1-4: Explain why. Note that $u^T Au = \|u\|_2 \|Au\|_2 \cos\theta$.

An important property of a covariance matrix is that it is postivie semidefinite.

Q1-5: Explain why. Start from a scalar random variable $k = u^T(X - \mathbb{E}X)$ where $u \in \mathbb{R}^n$ is arbitrary. Obviously $\mathbb{E}k^2 \geq 0$.

Using `lmer` to fit a linear mixed effects model

```
library(tidyverse)
library(ggplot2)
library(cowplot)
library(datasets)
library(lme4)
head(ChickWeight)
```

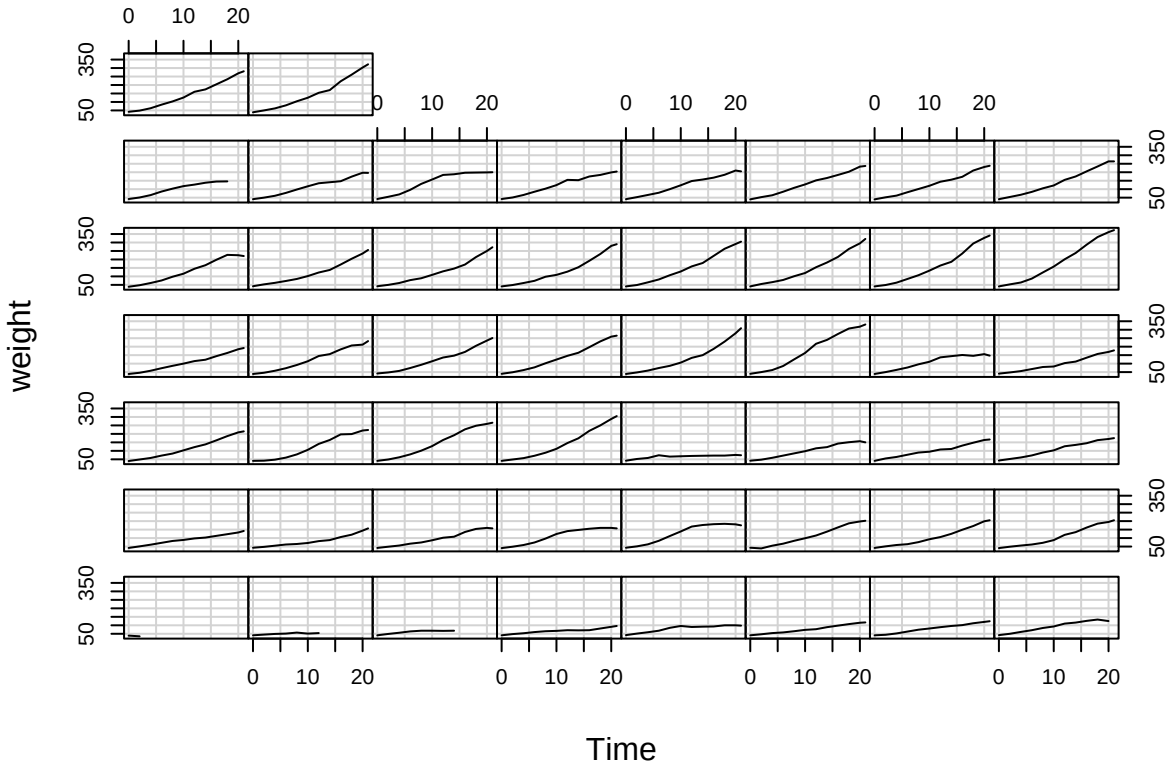
```
## Grouped Data: weight ~ Time | Chick
##   weight Time Chick Diet
## 1     42    0     1    1
## 2     51    2     1    1
## 3     59    4     1    1
## 4     64    6     1    1
## 5     76    8     1    1
## 6     93   10     1    1
```

```
table(ChickWeight$Chick, ChickWeight$Diet)
```

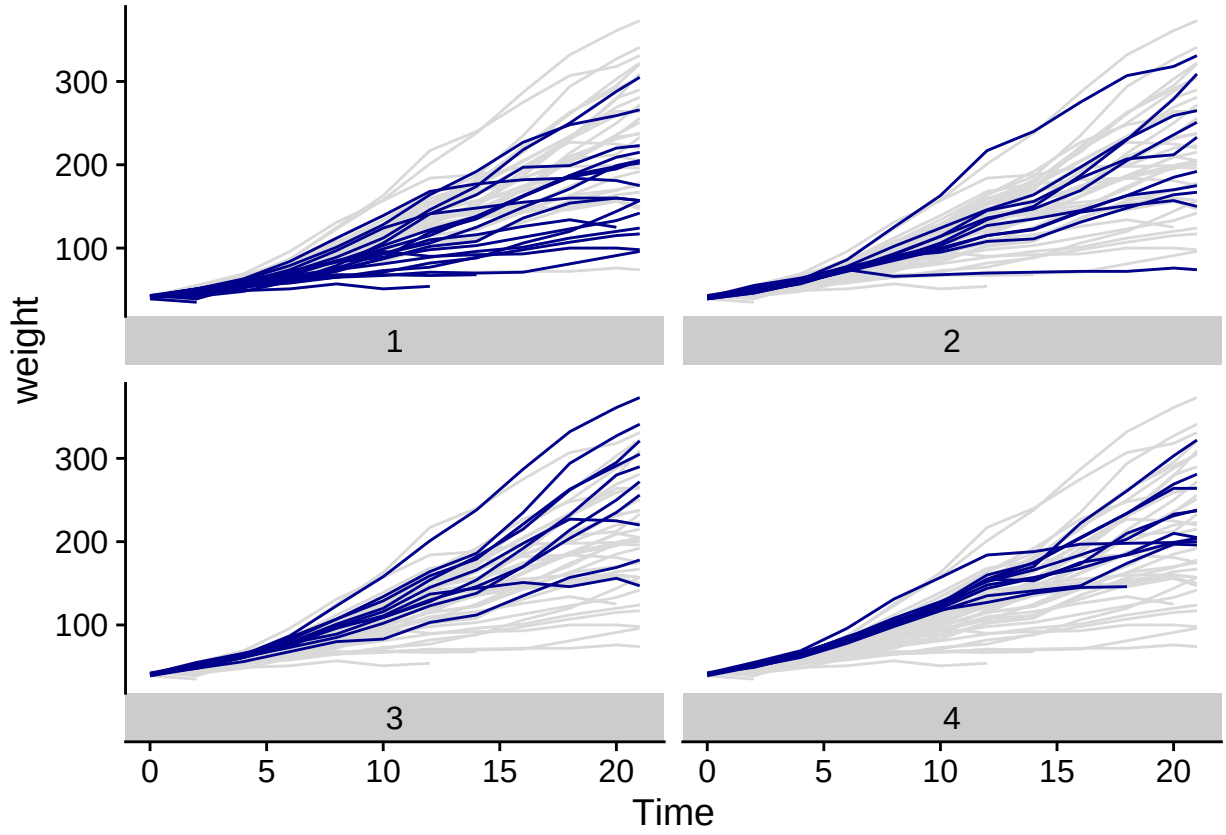
```
##
##       1  2  3  4
##   18  2  0  0  0
##   16  7  0  0  0
##   15  8  0  0  0
##   13 12  0  0  0
##   9  12  0  0  0
##   20 12  0  0  0
##   10 12  0  0  0
##   8  11  0  0  0
##   17 12  0  0  0
##   19 12  0  0  0
##   4  12  0  0  0
##   6  12  0  0  0
```

```
##    11 12  0  0  0
##    3  12  0  0  0
##    1  12  0  0  0
##    12 12  0  0  0
##    2  12  0  0  0
##    5  12  0  0  0
##    14 12  0  0  0
##    7  12  0  0  0
##    24  0 12  0  0
##    30  0 12  0  0
##    22  0 12  0  0
##    23  0 12  0  0
##    27  0 12  0  0
##    28  0 12  0  0
##    26  0 12  0  0
##    25  0 12  0  0
##    29  0 12  0  0
##    21  0 12  0  0
##    33  0  0 12  0
##    37  0  0 12  0
##    36  0  0 12  0
##    31  0  0 12  0
##    39  0  0 12  0
##    38  0  0 12  0
##    32  0  0 12  0
##    40  0  0 12  0
##    34  0  0 12  0
##    35  0  0 12  0
##    44  0  0  0 10
##    45  0  0  0 12
##    43  0  0  0 12
##    41  0  0  0 12
##    47  0  0  0 12
##    49  0  0  0 12
##    46  0  0  0 12
##    50  0  0  0 12
##    42  0  0  0 12
##    48  0  0  0 12
```

```r
coplot(weight ~ Time | Chick, data = ChickWeight, type = "l", show.given = F)
```




```
ChickWeight %>%
  ggplot(aes(x = Time, y = weight, group = Chick)) +
  geom_line(data = transform(ChickWeight, Diet = NULL), color = "grey85") +
  geom_line(show.legend = F, color="darkblue") +
  facet_wrap(~ Diet, strip.position = "bottom") +
  theme_cowplot()
```



Q2-1: Write down the model formula below, differentiating parameters to be estimated from random variables with distributions.

```
mod <- lmer(weight ~ Diet + Time + (Time | Chick), data=ChickWeight, REML=F)
summary(mod)
```

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: weight ~ Diet + Time + (Time | Chick)
##    Data: ChickWeight
##
##      AIC      BIC   logLik deviance df.resid
##   4834.1   4873.3  -2408.0   4816.1      569
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7927 -0.5834 -0.0388  0.4759  3.5200
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  Chick    (Intercept) 147.71   12.154
##           Time         13.85    3.721   -0.99
##  Residual             163.44   12.784
## Number of obs: 578, groups:  Chick, 50
##
## Fixed effects:
```

```
##             Estimate Std. Error t value
## (Intercept)  26.3563     2.2435  11.748
## Diet2         2.8383     2.2655   1.253
## Diet3         2.0076     2.2655   0.886
## Diet4         9.2548     2.2687   4.079
## Time          8.4439     0.5349  15.786
##
## Correlation of Fixed Effects:
##       (Intr) Diet2  Diet3  Diet4
## Diet2 -0.344
## Diet3 -0.344  0.344
## Diet4 -0.343  0.344  0.344
## Time  -0.806 -0.004 -0.004 -0.005
```

```
coef(summary(mod))
```

```
##              Estimate Std. Error    t value
## (Intercept) 26.356292  2.2434749 11.7479774
## Diet2        2.838296  2.2655272  1.2528192
## Diet3        2.007621  2.2655272  0.8861608
## Diet4        9.254783  2.2686722  4.0793830
## Time         8.443895  0.5348869 15.7863183
```

```
confint(mod, method = "boot")
```

```
## Computing bootstrap confidence intervals ...
```

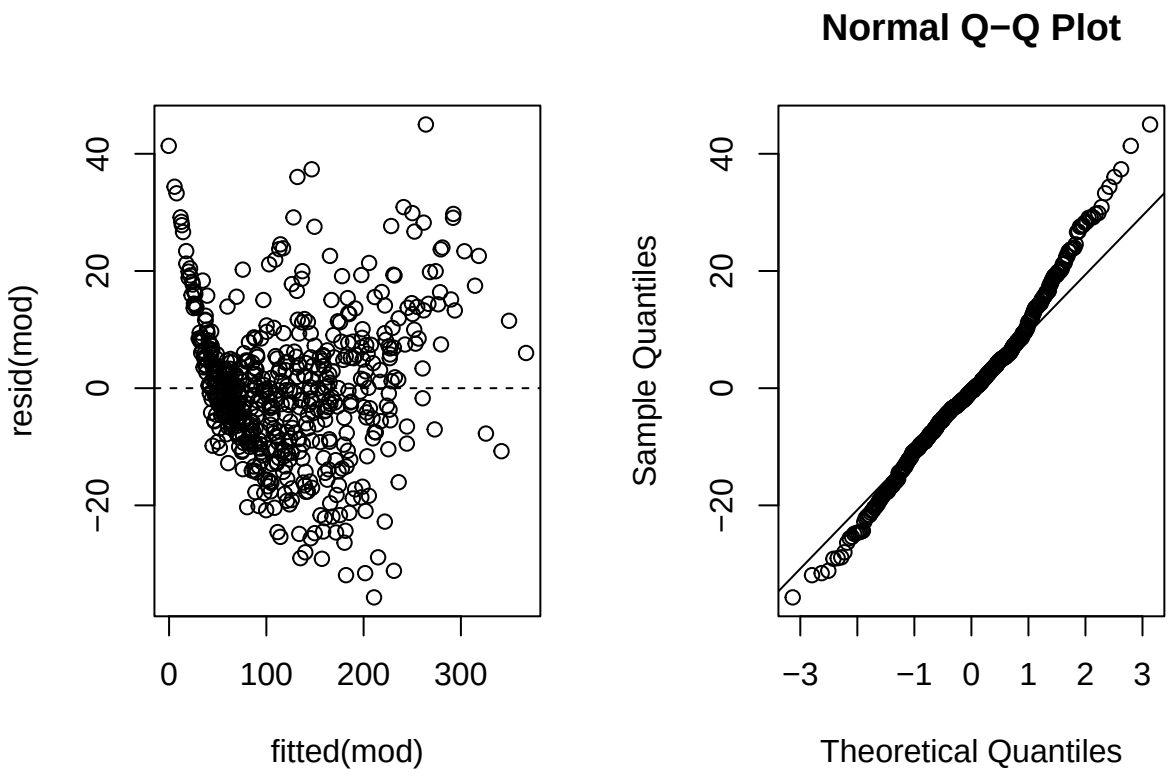
```
##
## 252 message(s): boundary (singular) fit: see help('isSingular')
## 56 warning(s): Model failed to converge with max|grad| = 0.00200963 (tol = 0.002, component 1) (and
```

```
##                  2.5 %     97.5 %
## .sig01       8.369076 14.8372173
## .sig02      -1.000000 -0.9503468
## .sig03       2.889505  4.3954718
## .sigma      11.908765 13.5777293
## (Intercept) 21.510512 30.6027627
## Diet2       -1.498149  7.6344686
## Diet3       -2.639103  6.2490098
## Diet4        4.724288 13.6257428
## Time         7.441146  9.5406790
```

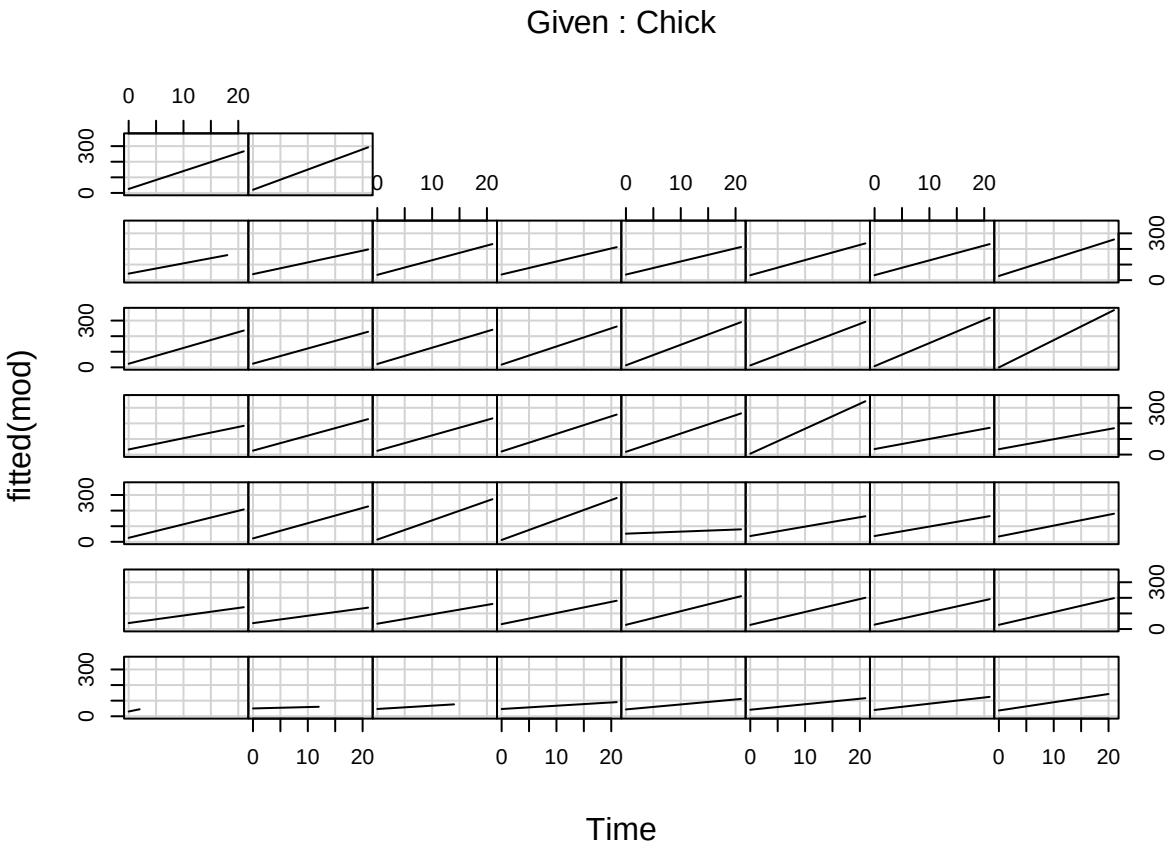
```
VarCorr(mod)
```

```
##  Groups   Name        Std.Dev. Corr
##  Chick    (Intercept) 12.1538
##           Time         3.7212  -0.990
##  Residual             12.7843
```

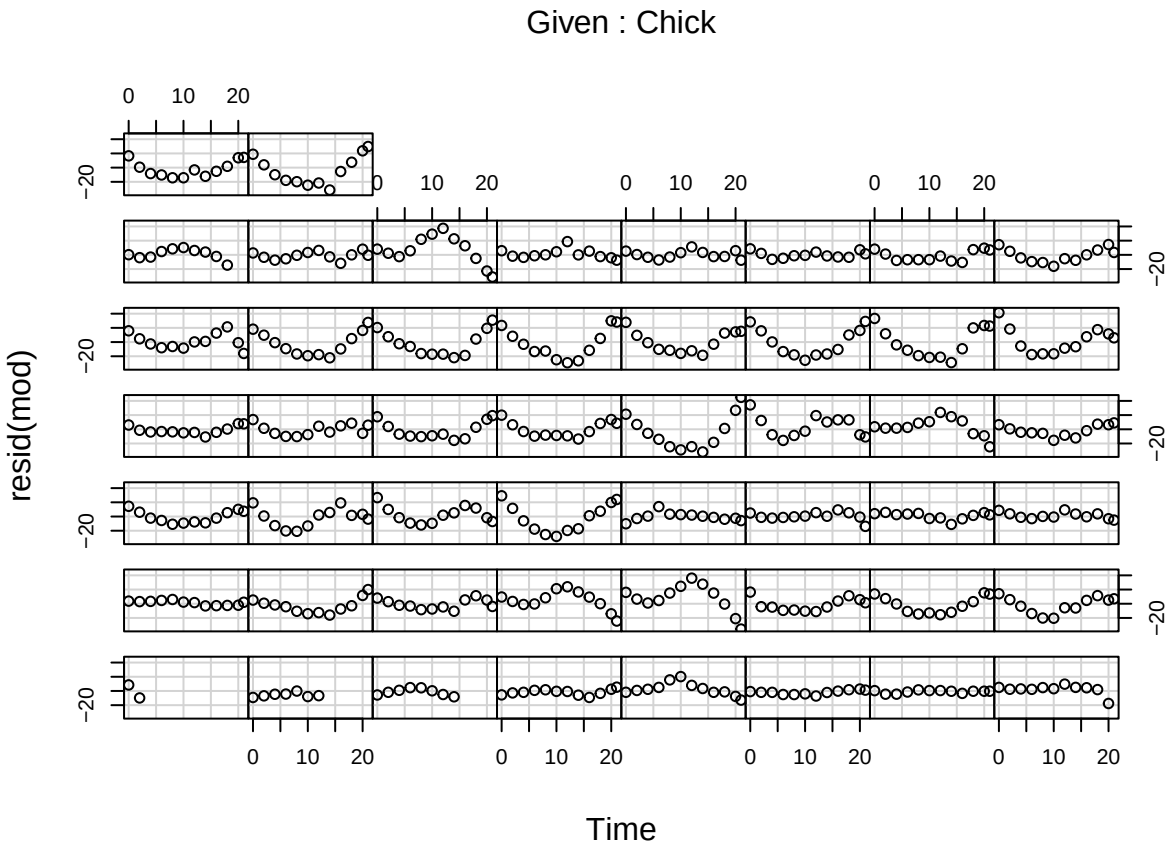
```r
par(mfrow=c(1,2))
plot(fitted(mod), resid(mod))
abline(h=0, lty=2)
qqnorm(resid(mod))
qqline(resid(mod))
```



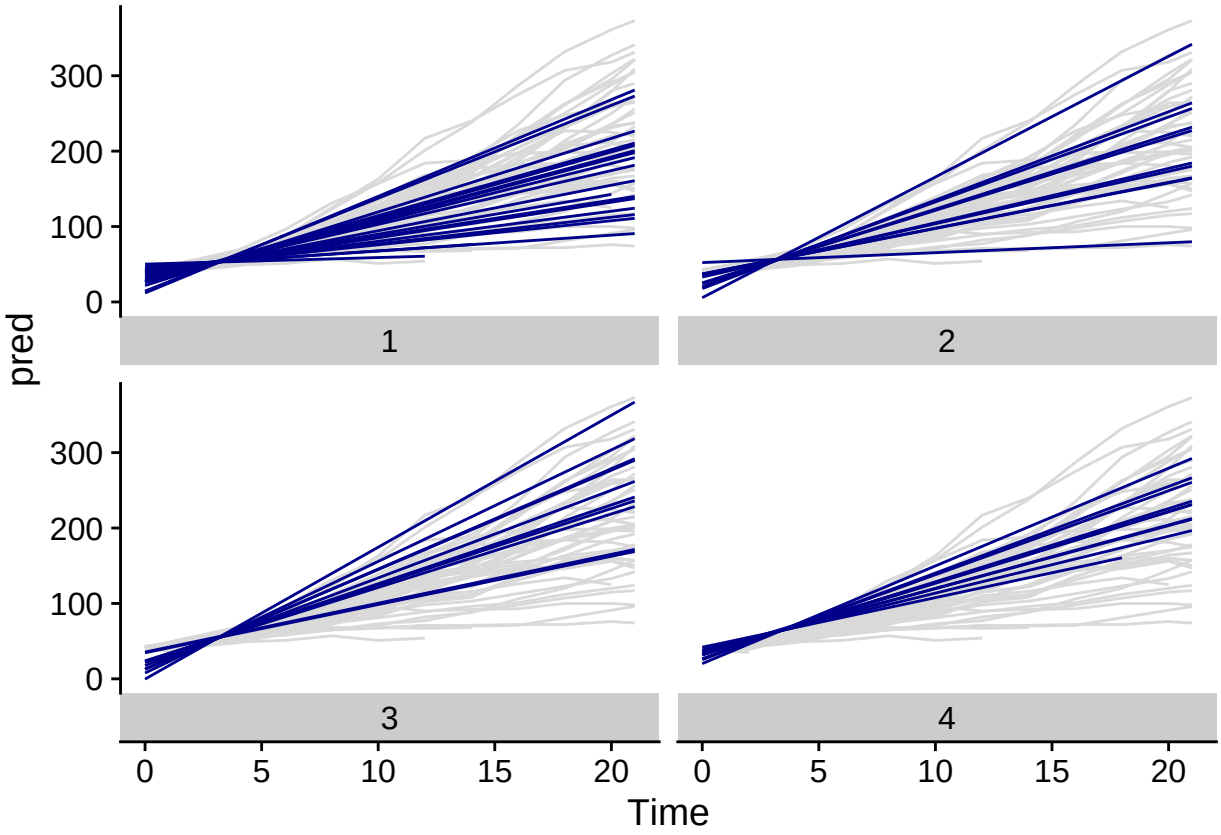

```r
coplot(fitted(mod) ~ Time | Chick, data = ChickWeight, type = "l", show.given = F)
```




```
coplot(resid(mod) ~ Time | Chick, data = ChickWeight, type = "p", show.given = F)
```



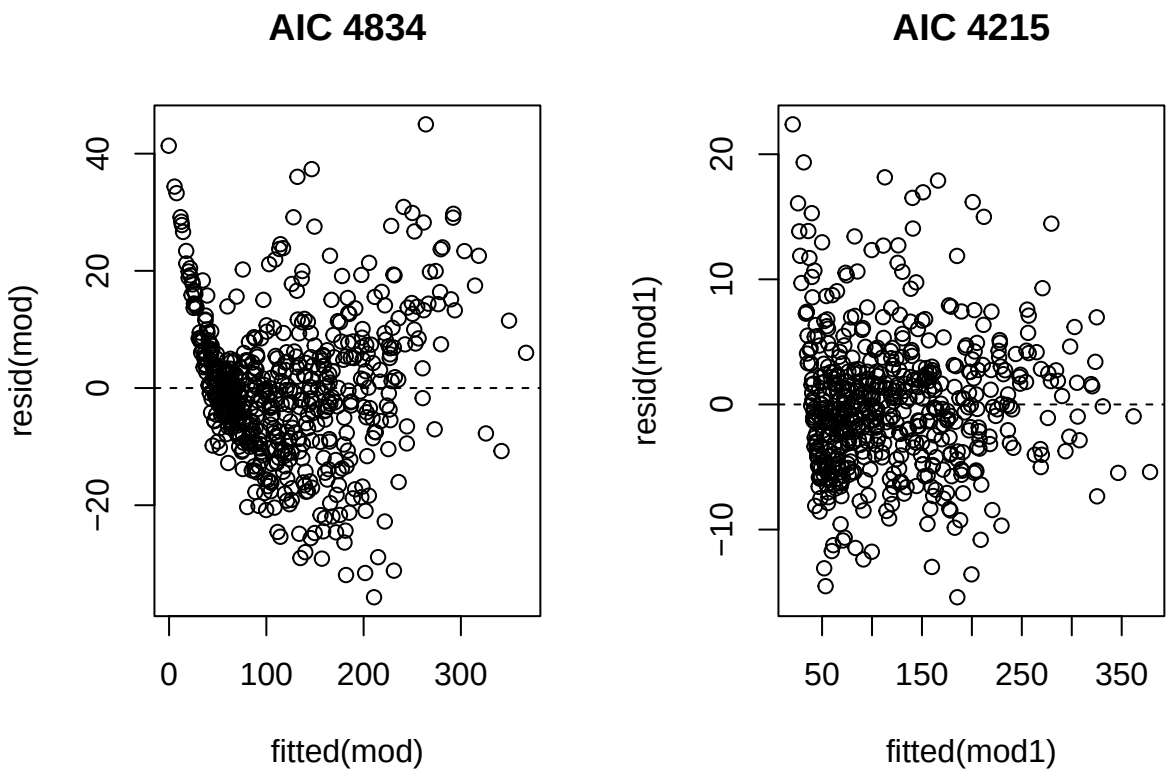

```
ChickWeight$pred = fitted(mod)
ChickWeight %>%
  ggplot(aes(x = Time, y = pred, group = Chick)) +
  geom_line(data = transform(ChickWeight, Diet = NULL, pred = weight), color = "grey85") +
  geom_line(show.legend = F, color="darkblue") +
  facet_wrap(~ Diet, strip.position = "bottom") +
  theme_cowplot()
```




```
mod1 <- lmer(weight ~ Diet + poly(Time,3) + (poly(Time,3) | Chick), data=ChickWeight, REML=F)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
par(mfrow=c(1,2))
plot(fitted(mod), resid(mod), main = paste0("AIC ", signif(AIC(mod),4)))
abline(h=0, lty=2)
plot(fitted(mod1), resid(mod1), main = paste0("AIC ", signif(AIC(mod1),4)))
abline(h=0, lty=2)
```



Q2-2: Reproduce similar plots for ‘mod1‘ to visualize model fits and residuals.