Assignment 9

Shinjon Ghosh

2025-04-04

```
exer <- read.csv("https://stats.idre.ucla.edu/stat/data/exer.csv")
 ## Convert variables to factor
exer <- within(exer, {
   diet <- factor(diet)
   exertype <- factor(exertype)
    time <- factor(time)
    id <- factor(id)
})

# Load Pacakage
library (ggplot2)
library(lme4)
```

```
## Warning: package 'lme4' was built under R version 4.4.3
```

```
## Loading required package: Matrix
```

```
library(nlme)
```

```
## Warning: package 'nlme' was built under R version 4.4.3
```

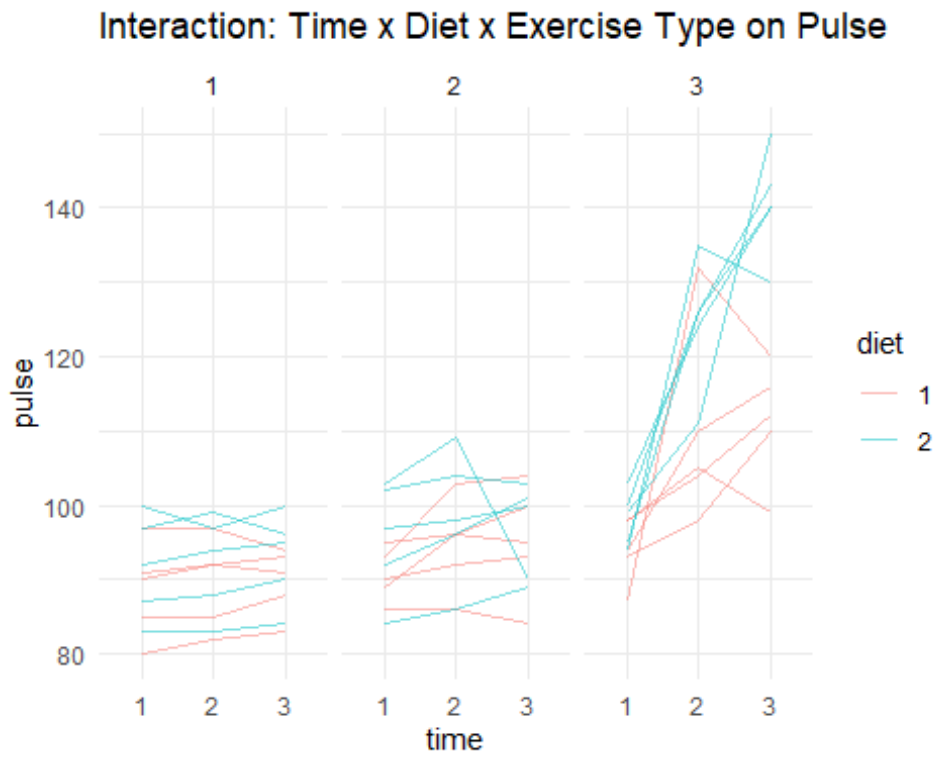
```
##
## Attaching package: 'nlme'
```

```
## The following object is masked from 'package:lme4':
##
##     lmList
```

```
library(Matrix)
```

Question 1: Spaghetti plot with interaction of time, diet, and exertype

```
ggplot(exer, aes(x = time, y = pulse, group = id, color = diet)) +
  geom_line(alpha = 0.5) +
  facet_wrap(~exertype) +
  labs(title = "Interaction: Time x Diet x Exercise Type on Pulse") +
  theme_minimal()
```



Question 2: Linear Model (lm)

```
# Fit the linear model
lm1 <- lm(pulse ~ time + diet + exertype + time:diet + time:exertype, data =
exer)

# Summary of the model
summary(lm1)

##
## Call:
## lm(formula = pulse ~ time + diet + exertype + time:diet + time:exertype,
##     data = exer)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -21.367  -4.867  -0.150   4.708  18.433
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      88.1333     2.9463  29.913  < 2e-16 ***
## time2            -0.7667     4.1667  -0.184 0.854492
## time3            -2.3667     4.1667  -0.568 0.571669
## diet2             4.1333     2.9463   1.403 0.164618
## exertype2         2.9000     3.6085   0.804 0.424033
## exertype3         5.9000     3.6085   1.635 0.106069
## time2:diet2       2.9333     4.1667   0.704 0.483532
```

```
## time3:diet2       7.1333     4.1667   1.712 0.090872 .
## time2:exertype2   2.8000     5.1031   0.549 0.584791
## time3:exertype2   1.6000     5.1031   0.314 0.754713
## time2:exertype3  20.3000     5.1031   3.978 0.000154 ***
## time3:exertype3  28.7000     5.1031   5.624  2.8e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.069 on 78 degrees of freedom
## Multiple R-squared:  0.7416, Adjusted R-squared:  0.7051
## F-statistic: 20.35 on 11 and 78 DF,  p-value: < 2.2e-16
```

time2:exertype3 and time3:exertype3 are significant predictors when $\alpha = 0.05$. Because p-value is less than 0.05.

Question 3: Linear Mixed Model

```
library(lme4)

lm2 <- lmer(pulse ~ time + diet + exertype*diet + exertype + (1 | id), data =
exer)

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

summary(lm2)

## Linear mixed model fit by REML ['lmerMod']
## Formula: pulse ~ time + diet + exertype * diet + exertype + (1 | id)
##    Data: exer
##
## REML criterion at convergence: 621.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.19099 -0.63325  0.00713  0.60475  2.68263
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  id       (Intercept)  0.00    0.000
##  Residual             87.54    9.356
## Number of obs: 90, groups:  id, 30
##
## Fixed effects:
##                 Estimate Std. Error t value
## (Intercept)      82.7667     2.7896  29.670
## time2             8.4000     2.4158   3.477
## time3            11.3000     2.4158   4.677
## diet2             3.0000     3.4165   0.878
## exertype2         4.1333     3.4165   1.210
## exertype3        15.7333     3.4165   4.605
## diet2:exertype2   0.4667     4.8317   0.097
```

```
## diet2:exertype3  13.0000     4.8317   2.691
##
## Correlation of Fixed Effects:
##             (Intr) time2  time3  diet2  exrty2 exrty3 dt2:x2
## time2       -0.433
## time3       -0.433  0.500
## diet2       -0.612  0.000  0.000
## exertype2   -0.612  0.000  0.000  0.500
## exertype3   -0.612  0.000  0.000  0.500  0.500
## dit2:xrtyp2  0.433  0.000  0.000 -0.707 -0.707 -0.354
## dit2:xrtyp3  0.433  0.000  0.000 -0.707 -0.354 -0.707  0.500
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

Question 4: Compare the fitted model to the Spaghetti plots

```
exer$time_num <- as.numeric(as.character(exer$time))  # convert factor to
numeric

lm3 <- lmer(pulse ~ time + diet + exertype*diet + exertype + (time_num | id),
data = exer)
summary(lm3)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: pulse ~ time + diet + exertype * diet + exertype + (time_num |
##     id)
##    Data: exer
##
## REML criterion at convergence: 610
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8908 -0.4563 -0.0162  0.3700  3.2022
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  id       (Intercept) 150.72   12.277
##           time_num     48.98    6.998   -0.95
##  Residual              38.84    6.232
## Number of obs: 90, groups:  id, 30
##
## Fixed effects:
##                 Estimate Std. Error t value
## (Intercept)      83.9541     2.6853  31.264
## time2             8.4000     2.0547   4.088
## time3            11.3000     3.0199   3.742
## diet2             3.0000     3.3649   0.892
## exertype2         3.7336     3.3649   1.110
## exertype3        13.8287     3.3649   4.110
## diet2:exertype2   0.8899     4.7587   0.187
```

```
## diet2:exertype3  10.0608     4.7587   2.114
##
## Correlation of Fixed Effects:
##             (Intr) time2  time3  diet2  exrty2 exrty3 dt2:x2
## time2       -0.433
## time3       -0.430  0.735
## diet2       -0.627  0.000  0.000
## exertype2   -0.627  0.000  0.000  0.500
## exertype3   -0.627  0.000  0.000  0.500  0.500
## dit2:xrtyp2  0.443  0.000  0.000 -0.707 -0.707 -0.354
## dit2:xrtyp3  0.443  0.000  0.000 -0.707 -0.354 -0.707  0.500

anova(lm2, lm3)

## refitting model(s) with ML (instead of REML)

## Data: exer
## Models:
## lm2: pulse ~ time + diet + exertype * diet + exertype + (1 | id)
## lm3: pulse ~ time + diet + exertype * diet + exertype + (time_num | id)
##     npar    AIC    BIC  logLik -2*log(L)  Chisq Df Pr(>Chisq)
## lm2   10 669.52 694.52 -324.76    649.52
## lm3   12 661.27 691.27 -318.63    637.27 12.256  2   0.002181 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value is < 0.05, so adding random slopes for time significantly improves model fit.

The spaghetti plots illustrate that pulse trends over time differ not only by diet and exercise type but also by individual. In particular, individuals in Exercise Type 3 show steep and varied increases in pulse. The variability of slopes within each group suggests that modeling individual-specific slopes is appropriate.

The random slope model (lm3) fits significantly better than the random intercept model (lm2), as indicated by the likelihood ratio test (p = 0.0022). Thus, including a random slope for time captures meaningful individual differences in pulse response over time, improving model accuracy.