simon-5502-12-solution

## File details

This program was written by Steve Simon on 2025-04-29 and is placed in the public domain. You can use this program any way you please.

library(lme4)  
library(tidyverse)  
  
R.version.string

[1] "R version 4.5.0 (2025-04-11 ucrt)"

Sys.Date()

[1] "2025-04-30"

## Read fasting-turtles

turtles <- read\_tsv(  
 file="../data/fasting-turtles.txt",  
 col\_names=TRUE,  
 col\_types="ncnnn")  
  
glimpse(turtles)

Rows: 8  
Columns: 5  
$ Subject <dbl> 1, 2, 3, 4, 5, 6, 7, 8  
$ Sex <chr> "Male", "Male", "Male", "Male", "Female", "Female", "Female",…  
$ Fed <dbl> 42.8, 43.1, 40.4, 46.6, 42.2, 38.7, 35.3, 40.5  
$ Fasted10 <dbl> 42.4, 42.2, 40.8, 45.9, 42.4, 38.1, 34.3, 40.1  
$ Fasted20 <dbl> 38.9, 40.3, 37.5, 42.9, 39.7, 35.8, 32.3, 37.3

## Question 1

You will not be graded on this, but get in the habit of drawing a few graphs and computing a few statistics that will help you better understand the dataset you are working with. Only examine variables that will be part of the further analyses.

It would also be fine to wait until after restructuring to calculate descriptive statistics.

turtles |>  
 summarize(  
 baseline\_mean=mean(Fed),  
 baseline\_sd=sd(Fed),  
 baseline\_min=min(Fed),  
 baseline\_max=max(Fed))

# A tibble: 1 × 4  
 baseline\_mean baseline\_sd baseline\_min baseline\_max  
 <dbl> <dbl> <dbl> <dbl>  
1 41.2 3.35 35.3 46.6

turtles |>  
 summarize(  
 time\_10\_mean=mean(Fasted10),  
 time\_10\_sd=sd(Fasted10),  
 time\_10\_min=min(Fasted10),  
 time\_10\_max=max(Fasted10))

# A tibble: 1 × 4  
 time\_10\_mean time\_10\_sd time\_10\_min time\_10\_max  
 <dbl> <dbl> <dbl> <dbl>  
1 40.8 3.45 34.3 45.9

turtles |>  
 summarize(  
 time\_20\_mean=mean(Fasted20),  
 time\_20\_sd=sd(Fasted20),  
 time\_20\_min=min(Fasted20),  
 time\_20\_max=max(Fasted20))

# A tibble: 1 × 4  
 time\_20\_mean time\_20\_sd time\_20\_min time\_20\_max  
 <dbl> <dbl> <dbl> <dbl>  
1 38.1 3.18 32.3 42.9

## Question 2

Restructure the data so Fed, Fasted10, and Fasted20 are in the same column. Create a new time variable from Fed, Fasted10, and Fasted20 that has values of 0, 10, and 20 respectively. The restructured dataset should have 24 rows. Show a glimpse of the new dataset.

## Restructure into long format

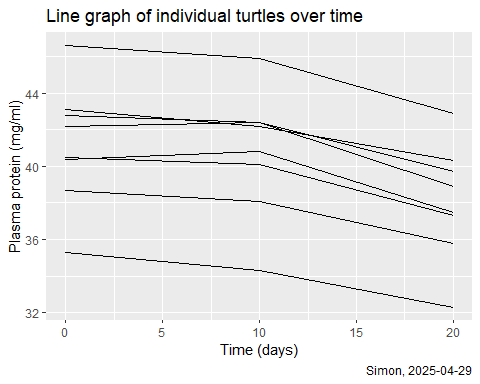
turtles |>  
 pivot\_longer(  
 cols=Fed:Fasted20,  
 names\_to="measure",  
 values\_to="plasma\_protein") |>  
 mutate(time=case\_when(  
 measure=="Fed" ~ 0,  
 measure=="Fasted10" ~ 10,  
 measure=="Fasted20" ~ 20)) -> turtles\_1  
  
glimpse(turtles\_1)

Rows: 24  
Columns: 5  
$ Subject <dbl> 1, 1, 1, 2, 2, 2, 3, 3, 3, 4, 4, 4, 5, 5, 5, 6, 6, 6, 7…  
$ Sex <chr> "Male", "Male", "Male", "Male", "Male", "Male", "Male",…  
$ measure <chr> "Fed", "Fasted10", "Fasted20", "Fed", "Fasted10", "Fast…  
$ plasma\_protein <dbl> 42.8, 42.4, 38.9, 43.1, 42.2, 40.3, 40.4, 40.8, 37.5, 4…  
$ time <dbl> 0, 10, 20, 0, 10, 20, 0, 10, 20, 0, 10, 20, 0, 10, 20, …

## Question 3

Draw a line graph, showing the change in plasma protein over time for each turtle. Ignore sex in this graph. Do the data show a consistent pattern (e.g., patients with large values at baseline tend to have large values after 10 and 20 days of fasting (relative to turtles that have small values at baseline)?

turtles\_1 |>  
 ggplot() +  
 aes(  
 x = time,   
 y = plasma\_protein,   
 group = Subject) +  
 geom\_line() +  
 labs(  
 x="Time (days)",  
 y="Plasma protein (mg/ml)",  
 title="Line graph of individual turtles over time",  
 caption="Simon, 2025-04-29")



## Question 4

Fit a random intercepts model with plasma protein as the dependent variable and time as the independent variable. Is there a statistically significant decline in plasma protein over time?

m1 <- lmer(  
 plasma\_protein ~ time + (1 | Subject),  
 data=turtles\_1)  
  
summary(m1)

Linear mixed model fit by REML ['lmerMod']  
Formula: plasma\_protein ~ time + (1 | Subject)  
 Data: turtles\_1  
  
REML criterion at convergence: 89.3  
  
Scaled residuals:   
 Min 1Q Median 3Q Max   
-1.1632 -0.5762 -0.2422 0.5314 1.6075   
  
Random effects:  
 Groups Name Variance Std.Dev.  
 Subject (Intercept) 10.7945 3.2855   
 Residual 0.5811 0.7623   
Number of obs: 24, groups: Subject, 8  
  
Fixed effects:  
 Estimate Std. Error t value  
(Intercept) 41.57708 1.18737 35.016  
time -0.15562 0.01906 -8.166  
  
Correlation of Fixed Effects:  
 (Intr)  
time -0.160

## Question 5

Calculate the intraclass correlation coefficient. Interpret this value.

icc <- 10.7945/(10.7945+0.5811)  
icc

[1] 0.948917