FXN.Variable.Features.R

ChristianRummey

2025-08-04

# ---  
# output:  
# word\_document:  
# reference\_docx: "template.docx"  
# ---  
  
# intro -------------------------------------------------------------------  
# reference\_docx: !expr here::here("template.docx")  
  
rm(list=ls())  
  
require(skimr)

## Loading required package: skimr

require(moments)

## Loading required package: moments

require(here)

## Loading required package: here

## here() starts at C:/Users/ChristianRummey/OneDrive - CDS/Projects/Ataxia/R.FXN.Slopes

library(corrr)

##   
## Attaching package: 'corrr'

## The following object is masked from 'package:skimr':  
##   
## focus

options(digits = 5)  
  
fxn. <- read\_rds(here("DATA derived/fxn.rds"))  
  
  
# . -----------------------------------------------------------------------  
  
fxn.tmp <- fxn. %>%   
 filter( type != 'FXN-T' ) %>%   
 # filter( grepl( 'TQ', analysis.group )) %>%   
 select( -analysis.group ) %>%   
 mutate( type = paste (type, unit) ) %>%   
 select(-unit) %>%   
 spread( type, value) %>%  
 rename(  
 FXN.M.LF = `FXN-M %`,  
 FXN.M = `FXN-M ng/ml`,  
 FXN.E = `FXN-E ng/ml`,  
 ) %>%  
 mutate( ratio = `FXN.M`/`FXN.E` ) %>%  
 ungroup  
  
fxn.tmp %<>%   
 filter( status == 'patient' ) %>%   
 filter( pm == 0) %>%   
 select( -study, -pm, -status ) %>%   
 mutate\_at(vars('sex'), as.factor)  
  
fxn.tmp %<>%   
 mutate(gaa1.log = log10(gaa1)) %>%   
 mutate(gaa2.log = log10(gaa2)) %>%   
 mutate(FXN.M.log = log10(FXN.M)) %>%   
 mutate(FXN.E.log = log10(FXN.E)) %>%   
 mutate(FXN.M.LF.log = log10(FXN.M.LF))  
   
my\_skim <- skim\_with(  
 numeric = sfl(  
 mean = ~ mean(.x, na.rm=TRUE),  
 sd = ~ sd(.x, na.rm=TRUE),  
 median = ~ median(.x, na.rm=TRUE),  
 IQR = ~ IQR(.x, na.rm=TRUE),  
 skewness = ~ skewness(.x, na.rm=TRUE)  
 ),  
 append = FALSE  
)  
  
# 2. Pipe your data into it  
fxn.tmp %>%  
 filter(is.na(FXN.M.LF.log)) %>%   
 select(-contains('FXN.M.LF')) %>%   
 .purge.df() %>%   
 my\_skim()

Data summary

|  |  |
| --- | --- |
| Name | Piped data |
| Number of rows | 145 |
| Number of columns | 13 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 1 |
| factor | 2 |
| numeric | 10 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| sjid | 0 | 1 | 2 | 4 | 0 | 145 | 0 |

**Variable type: factor**

| skim\_variable | n\_missing | complete\_rate | ordered | n\_unique | top\_counts |
| --- | --- | --- | --- | --- | --- |
| sev.o | 0 | 1 | FALSE | 4 | 8-1: 70, 0-7: 45, 15-: 24, >24: 6 |
| sex | 0 | 1 | FALSE | 2 | f: 74, m: 71 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | median | IQR | skewness |
| --- | --- | --- | --- | --- | --- | --- | --- |
| aoo | 0 | 1.00 | 11.15 | 6.81 | 10.00 | 7.00 | 1.66 |
| gaa1 | 11 | 0.92 | 688.32 | 231.05 | 700.00 | 284.75 | -0.42 |
| gaa2 | 11 | 0.92 | 974.89 | 250.79 | 989.50 | 257.25 | -0.59 |
| FXN.E | 0 | 1.00 | 4.32 | 3.34 | 3.57 | 3.68 | 2.12 |
| FXN.M | 0 | 1.00 | 3.71 | 2.38 | 3.17 | 2.18 | 1.63 |
| ratio | 0 | 1.00 | 1.11 | 0.64 | 1.00 | 0.72 | 1.22 |
| gaa1.log | 11 | 0.92 | 2.80 | 0.21 | 2.85 | 0.18 | -2.24 |
| gaa2.log | 11 | 0.92 | 2.97 | 0.16 | 3.00 | 0.11 | -3.14 |
| FXN.M.log | 0 | 1.00 | 0.49 | 0.29 | 0.50 | 0.30 | -0.64 |
| FXN.E.log | 0 | 1.00 | 0.52 | 0.34 | 0.55 | 0.43 | -0.48 |

fxn.tmp %>%  
 filter(!is.na(FXN.M.LF.log)) %>%   
 select(-c(FXN.E, FXN.M, FXN.E.log, FXN.M.log)) %>%   
 .purge.df() %>%   
 my\_skim()

Data summary

|  |  |
| --- | --- |
| Name | Piped data |
| Number of rows | 417 |
| Number of columns | 11 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| character | 1 |
| factor | 2 |
| numeric | 8 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: character**

| skim\_variable | n\_missing | complete\_rate | min | max | empty | n\_unique | whitespace |
| --- | --- | --- | --- | --- | --- | --- | --- |
| sjid | 0 | 1 | 1 | 4 | 0 | 417 | 0 |

**Variable type: factor**

| skim\_variable | n\_missing | complete\_rate | ordered | n\_unique | top\_counts |
| --- | --- | --- | --- | --- | --- |
| sev.o | 0 | 1 | FALSE | 4 | 8-1: 171, 0-7: 107, 15-: 96, >24: 43 |
| sex | 0 | 1 | FALSE | 2 | m: 212, f: 205 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | median | IQR | skewness |
| --- | --- | --- | --- | --- | --- | --- | --- |
| aoo | 0 | 1.00 | 13.43 | 9.09 | 11.00 | 9.00 | 1.87 |
| gaa1 | 12 | 0.97 | 621.74 | 231.86 | 660.00 | 304.00 | -0.47 |
| gaa2 | 13 | 0.97 | 906.66 | 212.68 | 906.00 | 250.00 | -0.24 |
| FXN.M.LF | 0 | 1.00 | 25.56 | 19.82 | 21.28 | 22.57 | 2.43 |
| ratio | 299 | 0.28 | 0.87 | 0.62 | 0.68 | 0.65 | 1.76 |
| gaa1.log | 12 | 0.97 | 2.75 | 0.23 | 2.82 | 0.21 | -1.88 |
| gaa2.log | 13 | 0.97 | 2.94 | 0.12 | 2.96 | 0.12 | -1.69 |
| FXN.M.LF.log | 0 | 1.00 | 1.28 | 0.36 | 1.33 | 0.47 | -0.71 |

# correlations ------------------------------------------------------------  
  
fxn.tmp %>%   
 select( aoo, gaa1, gaa2, starts\_with('FXN')) %>%  
 correlate(use = "pairwise.complete.obs", ) %>%  
 shave() %>% # hides lower triangle & diagonal  
 fashion()

## Correlation computed with  
## • Method: 'pearson'  
## • Missing treated using: 'pairwise.complete.obs'

## term aoo gaa1 gaa2 FXN.E FXN.M.LF FXN.M FXN.M.log FXN.E.log FXN.M.LF.log  
## 1 aoo   
## 2 gaa1 -.71   
## 3 gaa2 -.31 .43   
## 4 FXN.E .42 -.53 -.38   
## 5 FXN.M.LF .59 -.57 -.33 .59   
## 6 FXN.M .41 -.42 -.37 .62 .50   
## 7 FXN.M.log .38 -.43 -.32 .52 .49 .89   
## 8 FXN.E.log .40 -.52 -.32 .89 .54 .57 .56   
## 9 FXN.M.LF.log .53 -.58 -.31 .55 .85 .49 .56 .58

# regression models -----------------------------------------------------  
  
fxn.tmp %>%   
 select(-c('gaa1.log', 'gaa2.log', 'ratio')) %>%   
 gather(type, value, starts\_with('FXN.')) %>%   
 filter(!is.na(value)) %>%   
 group\_by(type) %>%   
 nest() %>%   
 mutate( model = map(data , ~ lm ( value ~ gaa1 + aoo + sex, data = . )) ) %>%   
 mutate( coefs = map(model, ~ broom::tidy(.))) %>%   
 unnest(coefs) %>%   
 select(-data, -model) %>%   
 .fixmod() %>%   
 flextable() %>%   
 theme\_booktabs()

| type | term | estimate | std.error | statistic | p.value |
| --- | --- | --- | --- | --- | --- |
| FXN.E | (Intercept) | 7.92461648 | 1.034958065 | 7.65694 | <0.0001 |
| FXN.E | gaa1 | -0.00641893 | 0.001045130 | -6.14175 | <0.0001 |
| FXN.E | aoo | 0.05117659 | 0.033401153 | 1.53218 | 0.1268 |
| FXN.E | sexm | 0.77149133 | 0.353267710 | 2.18387 | 0.0299 |
| FXN.M.LF | (Intercept) | 28.39144977 | 4.624246464 | 6.13969 | <0.0001 |
| FXN.M.LF | gaa1 | -0.02420346 | 0.004888921 | -4.95068 | <0.0001 |
| FXN.M.LF | aoo | 0.85764110 | 0.123446020 | 6.94750 | <0.0001 |
| FXN.M.LF | sexm | 1.81123833 | 1.544224712 | 1.17291 | 0.2415 |
| FXN.M | (Intercept) | 4.72104678 | 0.813862338 | 5.80079 | <0.0001 |
| FXN.M | gaa1 | -0.00290419 | 0.000821862 | -3.53367 | 0.0005 |
| FXN.M | aoo | 0.07326634 | 0.026265741 | 2.78943 | 0.0057 |
| FXN.M | sexm | -0.08673714 | 0.277799936 | -0.31223 | 0.7551 |
| FXN.M.log | (Intercept) | 0.72543810 | 0.104134616 | 6.96635 | <0.0001 |
| FXN.M.log | gaa1 | -0.00047192 | 0.000105158 | -4.48775 | <0.0001 |
| FXN.M.log | aoo | 0.00572711 | 0.003360732 | 1.70413 | 0.0896 |
| FXN.M.log | sexm | -0.04130732 | 0.035544819 | -1.16212 | 0.2463 |
| FXN.E.log | (Intercept) | 0.92674777 | 0.104481013 | 8.87001 | <0.0001 |
| FXN.E.log | gaa1 | -0.00066382 | 0.000105508 | -6.29163 | <0.0001 |
| FXN.E.log | aoo | 0.00377528 | 0.003371911 | 1.11963 | 0.2640 |
| FXN.E.log | sexm | 0.04987439 | 0.035663057 | 1.39849 | 0.1632 |
| FXN.M.LF.log | (Intercept) | 1.52938653 | 0.085351854 | 17.91861 | <0.0001 |
| FXN.M.LF.log | gaa1 | -0.00060672 | 0.000090237 | -6.72360 | <0.0001 |
| FXN.M.LF.log | aoo | 0.00992131 | 0.002278500 | 4.35432 | <0.0001 |
| FXN.M.LF.log | sexm | 0.00326947 | 0.028502469 | 0.11471 | 0.9087 |