

03_table1

randy

2022-03-22

```
demog <- here::here("data", "epic", "Demog.csv") %>%
  read.csv() %>%
  janitor::clean_names() %>%
  dplyr::select(id = cffidno, sex,
               ethnic, mutation1,
               mutation2, race)
epic <- here::here("data", "epic",
                  "registration_age_min_3_4.csv") %>%
  read.csv(row.name = 1) %>%
  janitor::clean_names()

data <- left_join(epic, demog, by = "id") %>%
  mutate(sex = as.factor(sex)) %>%
  mutate(genotype = case_when(mutation1 == "F508del" & mutation2 == "F508del" ~ "del/del",
                             mutation1 != "F508del" & mutation2 == "F508del" ~ "mut/del",
                             mutation2 != "F508del" & mutation1 == "F508del" ~ "mut/del",
                             mutation1 != "F508del" & mutation2 != "F508del" ~ "mut/mut"))

ID <- unique(data$id)
length(ID) / 3
```

```
## [1] 456.6667
```

```
test <- sample(ID, 457, replace = FALSE)
```

```
data0 <- data %>%
  mutate(group = case_when(id %in% test ~ "testing",
                           TRUE ~ "training"))
```

```
data1 <- data0 %>%
  group_by(id, group) %>%
  summarize(age_mean = mean(age),
            age_min = min(age),
            age_max = max(age),
            age_n = length(age),
            h_mean = mean(ht),
            h_max = max(ht),
            h_min = min(ht),
            w_mean = mean(wt),
            w_max = max(wt),
            w_min = min(wt),
```

```

      sex = sex,
      genotype = genotype,
      ethnic = ethnic,
      race = race) %>%
ungroup() %>%
unique()

```

'summarise()' has grouped output by 'id', 'group'. You can override using the
'.groups' argument.

```

# data0 <- data %>%
#   group_by(id, sex, ethnic, genotype) %>%
#   nest()

```

```

data2 <- full_join(data1, data) %>%
  as.data.frame() %>%
  mutate(time = age - age_min,
         age_diff = age_max - age_min,
         BMI = wt / (0.1 * ht)^2)

```

Joining, by = c("id", "sex", "genotype", "ethnic", "race")

```

# head(data2)
write.csv(data2, file = "data/epic_clean_randy.csv")

```

2 knots and bspline with cubic terms

- new table1 with two columns
 - include age_min & age_diff
 - time follow up time
 - training and testing
 - as different label for visit times
 - BMI, ht, wt
 - gender, r/eth, genotype
- split the data as 1/3 for testing and 2/3 for training
 - predicted value and the real obs for each subject in the testing set.
 - at least one observation for that individual
- cross validation GCV; as extra methodology for the model fitting
- use the predictive (dynamic prediction) as well as the marginal mean
- use the PML methods.

```

table1 <- data1 %>%
  unique() %>%
  select(-id) %>%
  mutate(
    ethnic = case_when(ethnic == 1 ~ "Hispanic",

```

```

        ethnic == 2 ~ "Non-Hispanic"),
  race = case_when(race == 1 ~ "White",
                  race != 1 ~ "Other"),
  sex = case_when(sex == "F" ~ "Female",
                 sex == "M" ~ "Male"),
  age_diff = age_max - age_min) %>%
select(group,
       Genotype = genotype,
       Gender = sex,
       Race = race,
       Ethnicity = ethnic,
       "Age mean" = age_mean,
       "Age min" = age_min,
       "Age diff" = age_diff,
       "Age max" = age_max,
       "Height mean" = h_mean,
       "Height baseline" = h_min,
       "Weight mean" = w_mean) %>%
## select all the variables for table1
tbl_summary(by = group) %>%
## just display all the variables in one column
modify_header(label = "**Variable**") %>%
# update the column header
bold_labels() %>%
italicize_labels() %>%
as_flex_table() %>%
flextable::bold(part = "header") %>%
## auto adjust the column widths
flextable::autofit()

```

table1

```

## Warning: Warning: fonts used in 'flextable' are ignored because the 'pdflatex'
## engine is used and not 'xelatex' or 'lualatex'. You can avoid this warning
## by using the 'set_flextable_defaults(fonts_ignore=TRUE)' command or use a
## compatible engine by defining 'latex_engine: xelatex' in the YAML header of the
## R Markdown document.

```

Variable	testing, N = 457 ¹	training, N = 913 ¹
<i>Genotype</i>		
del/del	240 (53%)	503 (55%)
mut/del	174 (38%)	340 (37%)
mut/mut	43 (9.4%)	70 (7.7%)
<i>Gender</i>		
Female	236 (52%)	457 (50%)
Male	221 (48%)	456 (50%)
<i>Race</i>		

¹n (%); Median (IQR)

Variable	testing, N = 457 ¹	training, N = 913 ¹
Other	22 (4.8%)	40 (4.4%)
White	435 (95%)	873 (96%)
<i>Ethnicity</i>		
Hispanic	16 (3.6%)	27 (3.1%)
Non-Hispanic	423 (96%)	851 (97%)
Unknown	18	35
<i>Age mean</i>	8.12 (6.58, 9.71)	8.11 (6.72, 10.23)
<i>Age min</i>	3.13 (3.05, 3.23)	3.13 (3.06, 3.22)
<i>Age diff</i>	9.4 (6.7, 12.2)	9.8 (7.2, 13.1)
<i>Age max</i>	12.5 (9.8, 15.5)	12.9 (10.4, 16.2)
<i>Height mean</i>	125 (116, 134)	126 (117, 135)
<i>Height baseline</i>	94.0 (91.4, 97.0)	94.0 (91.4, 96.9)
<i>Weight mean</i>	27 (22, 34)	27 (22, 34)

¹n (%); Median (IQR)

```
## save pptx -----
## flextable can be saved directly to powerpoints
flextable::save_as_pptx(
  table1,
  path = "figure/01_table1.pptx")
```