Draft Quarto document

Nicole

## Header 1

## Results

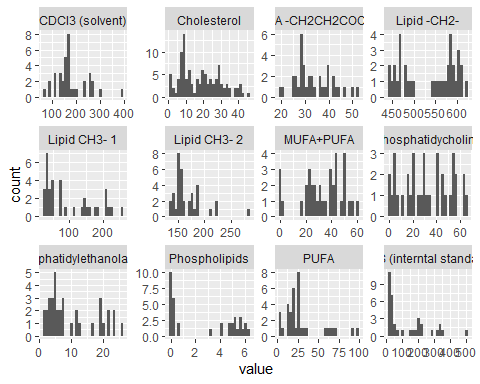
tar\_read(df\_stats\_by\_metabolite) %>%  
 mutate(MeanSD = glue::glue("{value\_mean} ({value\_sd})")) %>%  
 select(Metabolite = metabolite, `Mean SD` = MeanSD) %>%  
 knitr::kable(caption = "Descriptive statistics of the metabolites.")

Descriptive statistics of the metabolites.

| Metabolite | Mean SD |
| --- | --- |
| CDCl3 (solvent) | 180 (67) |
| Cholesterol | 18.6 (11.4) |
| FA -CH2CH2COO- | 33.6 (7.8) |
| Lipid -CH2- | 536.6 (61.9) |
| Lipid CH3- 1 | 98.3 (73.8) |
| Lipid CH3- 2 | 168.2 (29.2) |
| MUFA+PUFA | 32.9 (16.1) |
| PUFA | 30 (24.1) |
| Phosphatidycholine | 31.7 (20.5) |
| Phosphatidylethanolamine | 10 (7.6) |
| Phospholipids | 2.7 (2.6) |
| TMS (interntal standard) | 123 (130.4) |

tar\_read(fig\_metabolite\_distribution)

`stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



## Building the model

log\_reg\_specs <- logistic\_reg() %>%  
 set\_engine("glm")  
log\_reg\_specs

lipidomics\_wide <- lipidomics %>%  
 mutate(metabolite = snakecase::to\_snake\_case(metabolite)) %>%  
 pivot\_wider(names\_from = metabolite, values\_from = value, values\_fn = mean, names\_prefix = "metabolite\_")  
lipidomics\_wide

lipidomics %>%  
 column\_values\_to\_snake\_case()

lipidomics %>%  
 column\_values\_to\_snake\_case(metabolite) %>%  
 metabolites\_to\_wider()

recipe(class ~ metabolite\_lipid\_ch\_3\_1 + age + gender,  
 data = lipidomics\_wide  
)

recipe(lipidomics\_wide) %>%  
 update\_role(metabolite\_lipid\_ch\_3\_1, age, gender, new\_role = "predictor") %>%  
 update\_role(class, new\_role = "outcome") %>%  
 step\_normalize(starts\_with("metabolite\_"))

recipe\_specs <- lipidomics\_wide %>%  
 create\_recipe\_spec(metabolite\_lipid\_ch\_3\_1)  
recipe\_specs

workflow() %>%  
 add\_model(log\_reg\_specs) %>%  
 add\_recipe(recipe\_specs)

model\_workflow <- create\_model\_workflow(  
 logistic\_reg() %>%  
 set\_engine("glm"),  
 lipidomics\_wide %>%  
 create\_recipe\_spec(metabolite\_lipid\_ch\_3\_1)  
)  
model\_workflow

fitted\_model <- model\_workflow %>%  
 fit(lipidomics\_wide)  
fitted\_model

fitted\_model %>%  
 extract\_fit\_parsnip() %>%  
 tidy(exponentiate = TRUE)

fitted\_model %>%  
 tidy\_model\_output()

create\_model\_workflow(  
 logistic\_reg() %>%  
 set\_engine("glm"),  
 lipidomics\_wide %>%  
 create\_recipe\_spec(metabolite\_lipid\_ch\_3\_1)  
) %>%  
 fit(lipidomics\_wide) %>%  
 tidy\_model\_output()

## Running multiple models

# lipidomics %>%  
# column\_values\_to\_snake\_case(metabolite) %>%  
# group\_split(metabolite) %>% #one metabolite in each df  
# map(metabolites\_to\_wider)  
  
# convert to a function  
  
#' Convert the long form dataset into a list of wide form data frames.  
#'  
#' @param data lipidomics  
#'  
#' @return A list of data frames.  
  
split\_by\_metaoblite <- function(data) {  
 data %>%  
 column\_values\_to\_snake\_case(metabolite) %>%  
 dplyr::group\_split(metabolite) %>% # one metabolite in each df  
 purrr::map(metabolites\_to\_wider)  
}

lipidomics %>%  
 split\_by\_metaoblite()

[[1]]  
# A tibble: 36 × 5  
 code gender age class metabolite\_cd\_cl\_3\_solvent  
 <chr> <chr> <dbl> <chr> <dbl>  
 1 ERI109 M 25 CT 166.  
 2 ERI111 M 39 CT 171.  
 3 ERI163 W 58 CT 262.  
 4 ERI375 M 24 CT 172.  
 5 ERI376 M 26 CT 300.  
 6 ERI391 M 31 CT 241.  
 7 ERI392 M 24 CT 172.  
 8 ERI79 W 26 CT 148.  
 9 ERI81 M 52 CT 168.  
10 ERI83 M 25 CT 253.  
# ℹ 26 more rows  
  
[[2]]  
# A tibble: 36 × 5  
 code gender age class metabolite\_cholesterol  
 <chr> <chr> <dbl> <chr> <dbl>  
 1 ERI109 M 25 CT 18.6   
 2 ERI111 M 39 CT 20.8   
 3 ERI163 W 58 CT 15.5   
 4 ERI375 M 24 CT 10.2   
 5 ERI376 M 26 CT 13.5   
 6 ERI391 M 31 CT 9.53  
 7 ERI392 M 24 CT 9.87  
 8 ERI79 W 26 CT 17.6   
 9 ERI81 M 52 CT 17.0   
10 ERI83 M 25 CT 19.7   
# ℹ 26 more rows  
  
[[3]]  
# A tibble: 36 × 5  
 code gender age class metabolite\_fa\_ch\_2\_ch\_2\_coo  
 <chr> <chr> <dbl> <chr> <dbl>  
 1 ERI109 M 25 CT 31.6  
 2 ERI111 M 39 CT 28.9  
 3 ERI163 W 58 CT 36.6  
 4 ERI375 M 24 CT 39.4  
 5 ERI376 M 26 CT 52.1  
 6 ERI391 M 31 CT 42.8  
 7 ERI392 M 24 CT 39.9  
 8 ERI79 W 26 CT 32.7  
 9 ERI81 M 52 CT 28.4  
10 ERI83 M 25 CT 26.5  
# ℹ 26 more rows  
  
[[4]]  
# A tibble: 36 × 5  
 code gender age class metabolite\_lipid\_ch\_2  
 <chr> <chr> <dbl> <chr> <dbl>  
 1 ERI109 M 25 CT 587.  
 2 ERI111 M 39 CT 585.  
 3 ERI163 W 58 CT 558.  
 4 ERI375 M 24 CT 606.  
 5 ERI376 M 26 CT 554.  
 6 ERI391 M 31 CT 597.  
 7 ERI392 M 24 CT 607.  
 8 ERI79 W 26 CT 546.  
 9 ERI81 M 52 CT 593.  
10 ERI83 M 25 CT 606.  
# ℹ 26 more rows  
  
[[5]]  
# A tibble: 36 × 5  
 code gender age class metabolite\_lipid\_ch\_3\_1  
 <chr> <chr> <dbl> <chr> <dbl>  
 1 ERI109 M 25 CT 44.1  
 2 ERI111 M 39 CT 28.1  
 3 ERI163 W 58 CT 75.1  
 4 ERI375 M 24 CT 22.0  
 5 ERI376 M 26 CT 29.5  
 6 ERI391 M 31 CT 38.0  
 7 ERI392 M 24 CT 34.8  
 8 ERI79 W 26 CT 109.   
 9 ERI81 M 52 CT 49.6  
10 ERI83 M 25 CT 29.9  
# ℹ 26 more rows  
  
[[6]]  
# A tibble: 36 × 5  
 code gender age class metabolite\_lipid\_ch\_3\_2  
 <chr> <chr> <dbl> <chr> <dbl>  
 1 ERI109 M 25 CT 147.  
 2 ERI111 M 39 CT 153.  
 3 ERI163 W 58 CT 144.  
 4 ERI375 M 24 CT 220.  
 5 ERI376 M 26 CT 282.  
 6 ERI391 M 31 CT 220.  
 7 ERI392 M 24 CT 215.  
 8 ERI79 W 26 CT 153.  
 9 ERI81 M 52 CT 150.  
10 ERI83 M 25 CT 153.  
# ℹ 26 more rows  
  
[[7]]  
# A tibble: 36 × 5  
 code gender age class metabolite\_mufa\_pufa  
 <chr> <chr> <dbl> <chr> <dbl>  
 1 ERI109 M 25 CT 50.6   
 2 ERI111 M 39 CT 53.2   
 3 ERI163 W 58 CT 60.7   
 4 ERI375 M 24 CT 0.532  
 5 ERI376 M 26 CT 1.15   
 6 ERI391 M 31 CT 0.602  
 7 ERI392 M 24 CT 0.422  
 8 ERI79 W 26 CT 36.3   
 9 ERI81 M 52 CT 40.1   
10 ERI83 M 25 CT 39.3   
# ℹ 26 more rows  
  
[[8]]  
# A tibble: 36 × 5  
 code gender age class metabolite\_phosphatidycholine  
 <chr> <chr> <dbl> <chr> <dbl>  
 1 ERI109 M 25 CT 41.7  
 2 ERI111 M 39 CT 52.9  
 3 ERI163 W 58 CT 35.3  
 4 ERI375 M 24 CT 66.9  
 5 ERI376 M 26 CT 32.7  
 6 ERI391 M 31 CT 62.9  
 7 ERI392 M 24 CT 64.3  
 8 ERI79 W 26 CT 41.0  
 9 ERI81 M 52 CT 56.1  
10 ERI83 M 25 CT 57.8  
# ℹ 26 more rows  
  
[[9]]  
# A tibble: 36 × 5  
 code gender age class metabolite\_phosphatidylethanolamine  
 <chr> <chr> <dbl> <chr> <dbl>  
 1 ERI109 M 25 CT 6.78  
 2 ERI111 M 39 CT 3.66  
 3 ERI163 W 58 CT 3.59  
 4 ERI375 M 24 CT 3.59  
 5 ERI376 M 26 CT 2.33  
 6 ERI391 M 31 CT 1.46  
 7 ERI392 M 24 CT 2.00  
 8 ERI79 W 26 CT 4.93  
 9 ERI81 M 52 CT 5.20  
10 ERI83 M 25 CT 5.01  
# ℹ 26 more rows  
  
[[10]]  
# A tibble: 36 × 5  
 code gender age class metabolite\_phospholipids  
 <chr> <chr> <dbl> <chr> <dbl>  
 1 ERI109 M 25 CT 5.58  
 2 ERI111 M 39 CT 6.16  
 3 ERI163 W 58 CT 5.19  
 4 ERI375 M 24 CT 4.20  
 5 ERI376 M 26 CT 3.27  
 6 ERI391 M 31 CT 4.71  
 7 ERI392 M 24 CT 4.14  
 8 ERI79 W 26 CT 5.70  
 9 ERI81 M 52 CT 5.46  
10 ERI83 M 25 CT 4.89  
# ℹ 26 more rows  
  
[[11]]  
# A tibble: 36 × 5  
 code gender age class metabolite\_pufa  
 <chr> <chr> <dbl> <chr> <dbl>  
 1 ERI109 M 25 CT 29.0   
 2 ERI111 M 39 CT 27.4   
 3 ERI163 W 58 CT 35.5   
 4 ERI375 M 24 CT 6.92  
 5 ERI376 M 26 CT 3.22  
 6 ERI391 M 31 CT 3.43  
 7 ERI392 M 24 CT 3.52  
 8 ERI79 W 26 CT 18.7   
 9 ERI81 M 52 CT 20.7   
10 ERI83 M 25 CT 18.2   
# ℹ 26 more rows  
  
[[12]]  
# A tibble: 36 × 5  
 code gender age class metabolite\_tms\_interntal\_standard  
 <chr> <chr> <dbl> <chr> <dbl>  
 1 ERI109 M 25 CT 208.   
 2 ERI111 M 39 CT 219.   
 3 ERI163 W 58 CT 57.1  
 4 ERI375 M 24 CT 19.2  
 5 ERI376 M 26 CT 35.4  
 6 ERI391 M 31 CT 30.4  
 7 ERI392 M 24 CT 21.7  
 8 ERI79 W 26 CT 185.   
 9 ERI81 M 52 CT 207.   
10 ERI83 M 25 CT 322.   
# ℹ 26 more rows

model\_estimates <- lipidomics %>%  
 split\_by\_metaoblite() %>% # split  
 map(generate\_model\_results) %>% # apply  
 list\_rbind() %>% # combine  
 filter(str\_detect(term, "metabolite\_"))

Warning: glm.fit: algorithm did not converge

Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred  
  
Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

model\_estimates

# A tibble: 12 × 5  
 term estimate std.error statistic p.value  
 <chr> <dbl> <dbl> <dbl> <dbl>  
 1 metabolite\_cd\_cl\_3\_solvent 8.70e- 2 0.865 -2.82 0.00475  
 2 metabolite\_cholesterol 2.97e+ 0 0.458 2.38 0.0175   
 3 metabolite\_fa\_ch\_2\_ch\_2\_coo 1.52e+ 0 0.387 1.09 0.276   
 4 metabolite\_lipid\_ch\_2 2.59e- 3 3.14 -1.90 0.0578   
 5 metabolite\_lipid\_ch\_3\_1 4.45e+ 1 1.41 2.70 0.00697  
 6 metabolite\_lipid\_ch\_3\_2 8.85e- 1 0.361 -0.339 0.734   
 7 metabolite\_mufa\_pufa 4.56e- 1 0.449 -1.75 0.0798   
 8 metabolite\_phosphatidycholine 1.28e-120 116628. -0.00237 0.998   
 9 metabolite\_phosphatidylethanolamine 2.69e+ 1 1.32 2.49 0.0129   
10 metabolite\_phospholipids 2.39e- 19 68964. -0.000622 1.00   
11 metabolite\_pufa 3.27e+ 0 0.560 2.11 0.0345   
12 metabolite\_tms\_interntal\_standard 5.62e- 2 0.990 -2.91 0.00363

# lipidomics %>%  
# select(metabolite) %>%  
# mutate(term = metabolite) %>%  
# column\_values\_to\_snake\_case(term) %>%  
# mutate(term = str\_c("metabolite\_", term)) %>% # add "metabolite\_" infront of the metabolites  
# distinct(term, metabolite) %>%  
# right\_join(model\_estimates, by = "term")  
  
  
#' Add the original metabolite names (not as snakecase) to the model results.  
#'  
#' @param model\_results The data frame with the model results.  
#' @param data The original, unprocessed lipidomics dataset.  
#'  
#' @return A data frame.  
#'  
add\_original\_metabolite\_names <- function(model\_results, data) {  
 data %>%  
 dplyr::select(metabolite) %>%  
 dplyr::mutate(term = metabolite) %>%  
 column\_values\_to\_snake\_case(term) %>%  
 dplyr::mutate(term = stringr::str\_c("metabolite\_", term)) %>% # add "metabolite\_" infront of the metabolites  
 dplyr::distinct(term, metabolite) %>%  
 dplyr::right\_join(model\_results, by = "term")  
}

model\_estimates %>%   
 add\_original\_metabolite\_names(lipidomics)

# A tibble: 12 × 6  
 term metabolite estimate std.error statistic p.value  
 <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
 1 metabolite\_tms\_interntal\_st… TMS (inte… 5.62e- 2 9.90e-1 -2.91 0.00363  
 2 metabolite\_cholesterol Cholester… 2.97e+ 0 4.58e-1 2.38 0.0175   
 3 metabolite\_lipid\_ch\_3\_1 Lipid CH3… 4.45e+ 1 1.41e+0 2.70 0.00697  
 4 metabolite\_lipid\_ch\_3\_2 Lipid CH3… 8.85e- 1 3.61e-1 -0.339 0.734   
 5 metabolite\_lipid\_ch\_2 Lipid -CH… 2.59e- 3 3.14e+0 -1.90 0.0578   
 6 metabolite\_fa\_ch\_2\_ch\_2\_coo FA -CH2CH… 1.52e+ 0 3.87e-1 1.09 0.276   
 7 metabolite\_pufa PUFA 3.27e+ 0 5.60e-1 2.11 0.0345   
 8 metabolite\_phosphatidyletha… Phosphati… 2.69e+ 1 1.32e+0 2.49 0.0129   
 9 metabolite\_phosphatidycholi… Phosphati… 1.28e-120 1.17e+5 -0.00237 0.998   
10 metabolite\_phospholipids Phospholi… 2.39e- 19 6.90e+4 -0.000622 1.00   
11 metabolite\_mufa\_pufa MUFA+PUFA 4.56e- 1 4.49e-1 -1.75 0.0798   
12 metabolite\_cd\_cl\_3\_solvent CDCl3 (so… 8.70e- 2 8.65e-1 -2.82 0.00475

## Figure of model estimates

model\_estimates <- tar\_read(df\_model\_estimates)  
model\_estimates

# A tibble: 12 × 6  
 term metabolite estimate std.error statistic p.value  
 <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
 1 metabolite\_tms\_interntal\_st… TMS (inte… 5.62e- 2 9.90e-1 -2.91 0.00363  
 2 metabolite\_cholesterol Cholester… 2.97e+ 0 4.58e-1 2.38 0.0175   
 3 metabolite\_lipid\_ch\_3\_1 Lipid CH3… 4.45e+ 1 1.41e+0 2.70 0.00697  
 4 metabolite\_lipid\_ch\_3\_2 Lipid CH3… 8.85e- 1 3.61e-1 -0.339 0.734   
 5 metabolite\_lipid\_ch\_2 Lipid -CH… 2.59e- 3 3.14e+0 -1.90 0.0578   
 6 metabolite\_fa\_ch\_2\_ch\_2\_coo FA -CH2CH… 1.52e+ 0 3.87e-1 1.09 0.276   
 7 metabolite\_pufa PUFA 3.27e+ 0 5.60e-1 2.11 0.0345   
 8 metabolite\_phosphatidyletha… Phosphati… 2.69e+ 1 1.32e+0 2.49 0.0129   
 9 metabolite\_phosphatidycholi… Phosphati… 1.28e-120 1.17e+5 -0.00237 0.998   
10 metabolite\_phospholipids Phospholi… 2.39e- 19 6.90e+4 -0.000622 1.00   
11 metabolite\_mufa\_pufa MUFA+PUFA 4.56e- 1 4.49e-1 -1.75 0.0798   
12 metabolite\_cd\_cl\_3\_solvent CDCl3 (so… 8.70e- 2 8.65e-1 -2.82 0.00475

tar\_read(fig\_model\_estimates)

