NHANES Report

Your Name

2024-05-31

Table of contents

# Load packages  
library(tidyverse)

Warning: package 'ggplot2' was built under R version 4.3.3

Warning: package 'tidyr' was built under R version 4.3.3

Warning: package 'readr' was built under R version 4.3.3

Warning: package 'dplyr' was built under R version 4.3.3

Warning: package 'stringr' was built under R version 4.3.3

── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
✔ dplyr 1.1.4 ✔ readr 2.1.5  
✔ forcats 1.0.0 ✔ stringr 1.5.1  
✔ ggplot2 3.5.0 ✔ tibble 3.2.1  
✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
✔ purrr 1.0.2   
── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
✖ dplyr::filter() masks stats::filter()  
✖ dplyr::lag() masks stats::lag()  
ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(here)

here() starts at C:/Users/csz400/Documents/1 - Projects/2024\_Workshop-reproducible-research

# Loading the functions,   
# here() gets the location of the project  
source(here("R/descriptive.R"))  
source(here("R/models.R"))  
source(here("R/gt\_models.R"))  
  
# Alternative is to use .. to refer to the project root  
# source("../R/descriptive.R")  
# source("../R/models.R")  
# source("../R/gt\_models.R")

# 1. 2007

## 1.1 Descriptive stats

load(here("data/nh2007.RData"))  
  
# Descriptive stats  
purrr::map(.x = nh2007, .f = compute\_descriptive\_stats) |>  
 dplyr::bind\_rows(.id = "column")

column mean sd 0% 25%  
1 id 4.656261e+04 2966.3258794 41477.00 44060.75000  
2 gender NA NA NA NA  
3 gender NA NA NA NA  
4 gender NA NA NA NA  
5 age\_screening 4.741593e+01 19.2214951 16.00 31.00000  
6 education NA NA NA NA  
7 education NA NA NA NA  
8 education NA NA NA NA  
9 education NA NA NA NA  
10 education NA NA NA NA  
11 education NA NA NA NA  
12 education\_child 1.283616e+01 9.2602059 8.00 10.00000  
13 marital\_status 2.418417e+00 1.8024603 1.00 1.00000  
14 creatinine 1.264135e+02 79.7312853 8.00 65.00000  
15 lead 8.322468e-01 1.5173860 0.07 0.32000  
16 barium 2.193648e+00 3.9465255 0.08 0.69000  
17 cadmium 4.006332e-01 0.4691938 0.03 0.12925  
18 asthma 1.866765e+00 0.3782668 1.00 2.00000  
19 heart\_failure NA NA NA NA  
20 heart\_failure NA NA NA NA  
21 heart\_failure NA NA NA NA  
22 coronary\_heart\_disease NA NA NA NA  
23 coronary\_heart\_disease NA NA NA NA  
24 coronary\_heart\_disease NA NA NA NA  
25 heart\_attack NA NA NA NA  
26 heart\_attack NA NA NA NA  
27 heart\_attack NA NA NA NA  
28 stroke NA NA NA NA  
29 stroke NA NA NA NA  
30 stroke NA NA NA NA  
31 chronic\_bronchitis NA NA NA NA  
32 chronic\_bronchitis NA NA NA NA  
33 chronic\_bronchitis NA NA NA NA  
34 cancer NA NA NA NA  
35 cancer NA NA NA NA  
36 cancer NA NA NA NA  
 50% 75% 100% level Freq  
1 46513.0000 49178.75000 51622.00 <NA> NA  
2 NA NA NA 1 1022  
3 NA NA NA 2 1012  
4 NA NA NA <NA> 0  
5 47.0000 63.00000 80.00 <NA> NA  
6 NA NA NA 1 266  
7 NA NA NA 2 327  
8 NA NA NA 3 453  
9 NA NA NA 4 473  
10 NA NA NA 5 338  
11 NA NA NA <NA> 177  
12 11.0000 13.00000 66.00 <NA> NA  
13 1.0000 4.00000 6.00 <NA> NA  
14 114.0000 171.00000 528.00 <NA> NA  
15 0.5700 0.95750 52.30 <NA> NA  
16 1.3600 2.50000 101.00 <NA> NA  
17 0.2545 0.49275 4.72 <NA> NA  
18 2.0000 2.00000 9.00 <NA> NA  
19 NA NA NA FALSE 1799  
20 NA NA NA TRUE 58  
21 NA NA NA <NA> 177  
22 NA NA NA FALSE 1794  
23 NA NA NA TRUE 63  
24 NA NA NA <NA> 177  
25 NA NA NA FALSE 1781  
26 NA NA NA TRUE 76  
27 NA NA NA <NA> 177  
28 NA NA NA FALSE 1782  
29 NA NA NA TRUE 75  
30 NA NA NA <NA> 177  
31 NA NA NA FALSE 1751  
32 NA NA NA TRUE 106  
33 NA NA NA <NA> 177  
34 NA NA NA FALSE 1684  
35 NA NA NA TRUE 173  
36 NA NA NA <NA> 177

## 1.2 Models

# List outcomes  
outcomes <- c("asthma", "heart\_failure", "coronary\_heart\_disease", "heart\_attack")  
exposures <- c("creatinine", "lead", "barium", "cadmium")  
  
  
models\_parameters\_nh2007 <- tidyr::expand\_grid(outcomes, exposures)  
  
models\_nh2007 <- map2(  
 .x = models\_parameters\_nh2007$outcomes,  
 .y = models\_parameters\_nh2007$exposures,  
 .f = \(x, y) build\_model(x, y, dataset = nh2007) # we need to change the dataset  
)  
  
# Extract model results  
models\_results\_nh2007 <- map(models\_nh2007, extract\_model\_result)

Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...

results\_nh2007 <- models\_parameters\_nh2007 |>  
 dplyr::mutate(  
 models = models\_nh2007,  
 models\_results = models\_results\_nh2007  
 )  
  
results\_model\_clean\_nh2007 <- results\_nh2007 |>  
 unnest(models\_results) |>  
 dplyr::filter(exposures == term) |>  
 select(-models)  
  
gt\_models(results\_model\_clean\_nh2007)

Warning: package 'gt' was built under R version 4.3.3

Table 1: Models results

| exposures | Estimate | Pr(>|t|) | 95% CI | aic |
| --- | --- | --- | --- | --- |
| asthma | | | | |
| creatinine | 0.000 | 0.13505 | 0.000 - 0.000 | 1802.2351 |
| lead | 0.002 | 0.66801 | -0.008 - 0.013 | 1804.2892 |
| barium | -0.003 | 0.15283 | -0.007 - 0.001 | 1802.4252 |
| cadmium | 0.001 | 0.94783 | -0.035 - 0.037 | 1804.4692 |
| heart\_failure | | | | |
| creatinine | 0.000 | 0.60937 | 0.000 - 0.000 | -1270.1983 |
| lead | -0.003 | 0.29314 | -0.008 - 0.002 | -1271.0444 |
| barium | 0.000 | 0.61605 | -0.002 - 0.001 | -1270.1887 |
| cadmium | 0.002 | 0.76814 | -0.014 - 0.019 | -1270.0237 |
| coronary\_heart\_disease | | | | |
| creatinine | 0.000 | 0.41578 | 0.000 - 0.000 | -1147.8785 |
| lead | -0.003 | 0.24778 | -0.008 - 0.002 | -1148.5536 |
| barium | 0.000 | 0.69466 | -0.002 - 0.002 | -1147.3691 |
| cadmium | 0.014 | 0.09495 | -0.003 - 0.031 | -1150.0099 |
| heart\_attack | | | | |
| creatinine | 0.000 | 0.92662 | 0.000 - 0.000 | -809.6718 |
| lead | -0.003 | 0.29655 | -0.009 - 0.003 | -810.7556 |
| barium | -0.001 | 0.35293 | -0.003 - 0.001 | -810.5283 |
| cadmium | 0.026 | 0.00538 | 0.008 - 0.045 | -817.4306 |

# 2. 2009

## 2.1 Descriptive stats

load(here("data/nh2009.RData"))  
  
# Descriptive stats  
purrr::map(.x = nh2009, .f = compute\_descriptive\_stats) |>  
 dplyr::bind\_rows(.id = "column")

column mean sd 0% 25%  
1 id 5.685692e+04 3025.145830 51630.00 54224.750  
2 gender NA NA NA NA  
3 gender NA NA NA NA  
4 gender NA NA NA NA  
5 age\_screening 4.623440e+01 19.448558 16.00 30.000  
6 education NA NA NA NA  
7 education NA NA NA NA  
8 education NA NA NA NA  
9 education NA NA NA NA  
10 education NA NA NA NA  
11 education NA NA NA NA  
12 education NA NA NA NA  
13 education NA NA NA NA  
14 education\_child 1.191556e+01 5.468006 8.00 10.000  
15 marital\_status 2.491828e+00 1.829802 1.00 1.000  
16 creatinine 1.227959e+02 79.409829 7.00 62.750  
17 lead 7.707041e-01 1.550689 0.07 0.290  
18 barium 2.329029e+00 9.417887 0.08 0.680  
19 cadmium 3.791720e-01 0.494888 0.03 0.117  
20 asthma 1.870766e+00 0.432992 1.00 2.000  
21 heart\_failure NA NA NA NA  
22 heart\_failure NA NA NA NA  
23 heart\_failure NA NA NA NA  
24 coronary\_heart\_disease NA NA NA NA  
25 coronary\_heart\_disease NA NA NA NA  
26 coronary\_heart\_disease NA NA NA NA  
27 heart\_attack NA NA NA NA  
28 heart\_attack NA NA NA NA  
29 heart\_attack NA NA NA NA  
30 stroke NA NA NA NA  
31 stroke NA NA NA NA  
32 stroke NA NA NA NA  
33 chronic\_bronchitis NA NA NA NA  
34 chronic\_bronchitis NA NA NA NA  
35 chronic\_bronchitis NA NA NA NA  
36 cancer NA NA NA NA  
37 cancer NA NA NA NA  
38 cancer NA NA NA NA  
 50% 75% 100% level Freq  
1 56840.5000 5.9436e+04 62158.00 <NA> NA  
2 NA NA NA 1 1097  
3 NA NA NA 2 1147  
4 NA NA NA <NA> 0  
5 45.0000 6.2000e+01 80.00 <NA> NA  
6 NA NA NA 1 239  
7 NA NA NA 2 333  
8 NA NA NA 3 469  
9 NA NA NA 4 592  
10 NA NA NA 5 384  
11 NA NA NA 7 1  
12 NA NA NA 9 1  
13 NA NA NA <NA> 225  
14 11.0000 1.3000e+01 66.00 <NA> NA  
15 1.0000 4.0000e+00 6.00 <NA> NA  
16 108.0000 1.6600e+02 489.00 <NA> NA  
17 0.5000 8.8000e-01 49.60 <NA> NA  
18 1.3350 2.5300e+00 419.00 <NA> NA  
19 0.2355 4.5225e-01 8.35 <NA> NA  
20 2.0000 2.0000e+00 9.00 <NA> NA  
21 NA NA NA FALSE 1970  
22 NA NA NA TRUE 49  
23 NA NA NA <NA> 225  
24 NA NA NA FALSE 1956  
25 NA NA NA TRUE 63  
26 NA NA NA <NA> 225  
27 NA NA NA FALSE 1945  
28 NA NA NA TRUE 74  
29 NA NA NA <NA> 225  
30 NA NA NA FALSE 1954  
31 NA NA NA TRUE 65  
32 NA NA NA <NA> 225  
33 NA NA NA FALSE 1918  
34 NA NA NA TRUE 101  
35 NA NA NA <NA> 225  
36 NA NA NA FALSE 1797  
37 NA NA NA TRUE 222  
38 NA NA NA <NA> 225

## 2.2 Models

# List outcomes  
outcomes <- c("asthma", "heart\_failure", "coronary\_heart\_disease", "heart\_attack")  
exposures <- c("creatinine", "lead", "barium", "cadmium")  
  
  
models\_parameters\_2009 <- tidyr::expand\_grid(outcomes, exposures)  
  
models\_2009 <- map2(  
 .x = models\_parameters\_2009$outcomes,  
 .y = models\_parameters\_2009$exposures,  
 .f = \(x, y) build\_model(x, y, dataset = nh2009) # we need to change the dataset  
)  
  
# Extract model results  
models\_results\_2009 <- map(models\_2009, extract\_model\_result)

Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...  
Waiting for profiling to be done...

results\_2009 <- models\_parameters\_2009 |>  
 dplyr::mutate(  
 models = models\_2009,  
 models\_results = models\_results\_2009  
 )  
  
results\_model\_clean\_2009 <- results\_2009 |>  
 unnest(models\_results) |>  
 dplyr::filter(exposures == term) |>  
 select(-models)  
  
gt\_models(results\_model\_clean\_2009)

Table 1: Models results

| exposures | Estimate | Pr(>|t|) | 95% CI | aic |
| --- | --- | --- | --- | --- |
| asthma | | | | |
| creatinine | 0.000 | 0.2371 | 0.000 - 0.000 | 2611.982 |
| lead | 0.004 | 0.5313 | -0.008 - 0.015 | 2612.990 |
| barium | -0.002 | 0.0443 | -0.004 - 0.000 | 2609.328 |
| cadmium | -0.028 | 0.1436 | -0.066 - 0.010 | 2611.239 |
| heart\_failure | | | | |
| creatinine | 0.000 | 0.0639 | 0.000 - 0.000 | -1891.887 |
| lead | -0.004 | 0.0466 | -0.008 - 0.000 | -1892.416 |
| barium | 0.000 | 0.4765 | -0.001 - 0.000 | -1888.955 |
| cadmium | -0.008 | 0.2510 | -0.021 - 0.005 | -1889.768 |
| coronary\_heart\_disease | | | | |
| creatinine | 0.000 | 0.1801 | 0.000 - 0.000 | -1410.245 |
| lead | -0.003 | 0.2730 | -0.007 - 0.002 | -1409.649 |
| barium | 0.000 | 0.5532 | -0.001 - 0.001 | -1408.797 |
| cadmium | -0.008 | 0.3203 | -0.022 - 0.007 | -1409.435 |
| heart\_attack | | | | |
| creatinine | 0.000 | 0.1923 | 0.000 - 0.000 | -1091.589 |
| lead | -0.002 | 0.4105 | -0.007 - 0.003 | -1090.564 |
| barium | 0.000 | 0.4657 | -0.001 - 0.001 | -1090.418 |
| cadmium | -0.002 | 0.8184 | -0.018 - 0.014 | -1089.938 |