Test Data

Jamie Pantazi Esmond

April 19, 2023

Table of contents

List of Figures

List of Tables

```
library(tidyverse)
```

```
-- Attaching packages --------------------------------------- tidyverse 1.3.2 --
v ggplot2 3.4.0      v purrr   1.0.1
v tibble  3.1.8      v dplyr   1.0.10
v tidyr   1.2.1      v stringr 1.5.0
v readr   2.1.3      v forcats 0.5.2
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

```
library(broom)
library(modelsummary)
library(kableExtra)
```

```
Attaching package: 'kableExtra'
```

```
The following object is masked from 'package:dplyr':

    group_rows
```

```
library(ggdag)
```

```
Attaching package: 'ggdag'
The following object is masked from 'package:stats':

    filter
```

```
library(scales)
```

```
Attaching package: 'scales'
The following object is masked from 'package:purrr':

    discard
The following object is masked from 'package:readr':

    col_factor
```

```
library(truncnorm)
```

```
Warning: package 'truncnorm' was built under R version 4.2.3
```

```
library(lubridate)
```

```
Loading required package: timechange
Attaching package: 'lubridate'
The following objects are masked from 'package:base':

    date, intersect, setdiff, union
```

```r
library(randomNames)
```

```
Warning: package 'randomNames' was built under R version 4.2.3
```

```r
set.seed(80)
n_ppl <- 50

esteem <- tibble(id = 1:n_ppl) %>%
  # Dates
  mutate(year = sample(2017:2022, n_ppl, replace = TRUE)) %>%
  mutate(month = sample(1:12, n_ppl, replace = TRUE)) %>%
  mutate(day = sample(1:31, n_ppl, replace = TRUE),
         day = case_when(month == 2 & day > 28 ~ sample(1:28, n_ppl, replace = TRUE),
                         month %in% c(4, 6, 9, 11) & day > 30 ~ sample(1:30, n_ppl, replac
                         TRUE ~ day)) %>%
  mutate(date = make_date(year, month, day)) %>%
  # Demographics
  mutate(age = round(rtruncnorm(n_ppl, 16, 65, 18, 5), 0)) %>%
  mutate(race = sample(c("Black", "Latinx", "White", "Other"), n_ppl,
                       replace = TRUE, prob = c(.4, .2, .3, .1)),
         race_num = case_when(race == "Black" ~ 1,
                              race == "Latinx" ~ 2,
                              race == "White" ~ 3,
                              race == "Other" ~ 4)) %>%
  mutate(gender = sample(c("Man", "Woman"), n_ppl, replace = TRUE,
                         prob = c(.45, .55)),
         woman = case_when(gender == "Man" ~ 0,
                           gender == "Woman" ~ 1)) %>%
  # Social Media
  mutate(social_base = rtruncnorm(n_ppl, 1, 10, 5, 2.5),
         age_social = age * rtruncnorm(n_ppl, -.05, 0, -.025, .0025),
         gender_social = case_when(gender == "Man" ~ rnorm(n_ppl, 1, 2),
                                   gender == "Woman" ~ rnorm(n_ppl, 4, 1)),
         race_social = case_when(race == "Black" ~ rnorm(n_ppl, 3, 2),
                                 race == "Latinx" ~ rnorm(n_ppl, 1, 1.5),
                                 race == "White" ~ rnorm(n_ppl, 0, 2.5),
                                 race == "Other" ~ rnorm(n_ppl, 2, 2)),
         social = social_base + age_social + gender_social + race_social,
         social = case_when(social < 1 ~ 1,
                            social > 10 ~ 10,
                            TRUE ~ round(social))) %>%
```

```r
  # Self-esteem
  mutate(esteem_base = rtruncnorm(n_ppl, 1, 10, 5, 1.25),
         age_esteem = age * rtruncnorm(n_ppl, -.25, 0, -.15, .05),
         age_gender = ifelse(gender == "Woman", age * rtruncnorm(n_ppl, -.4, 0, -.2, .05),
         gender_esteem = case_when(gender == "Man" ~ rnorm(n_ppl, 2.5, 1),
                                   gender == "Woman" ~ rnorm(n_ppl, -2, .75)),
         race_esteem = case_when(race == "Black" ~ rnorm(n_ppl, -1.5, .75),
                                 race == "Latinx" ~ rnorm(n_ppl, -1, .5),
                                 race == "White" ~ rnorm(n_ppl, 2, .75),
                                 race == "Other" ~ rnorm(n_ppl, -1, .5)),
         social_esteem = social * rtruncnorm(n_ppl, -.5, 0, -.175, .05),
         esteem = esteem_base + age_social + gender_esteem + race_esteem + social_esteem,
         esteem = case_when(esteem < 1 ~ 1,
                            esteem > 10 ~ 10,
                            TRUE ~ round(esteem))) %>%
  select(id, date, age, gender, woman, race, race_num, social, esteem)

write.csv(esteem, "esteem.csv", row.names=FALSE)

m1 <- lm(esteem ~ social, data = esteem)
tidy(m1)
```

```
# A tibble: 2 x 5
  term        estimate std.error statistic       p.value
  <chr>          <dbl>     <dbl>     <dbl>         <dbl>
1 (Intercept)    6.35      0.865      7.35 0.00000000217
2 social        -0.389     0.107     -3.64 0.000676
```

```r
m2 <- lm(esteem ~ social + age, data = esteem)
tidy(m2)
```

```
# A tibble: 3 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)  6.17        2.29     2.69   0.00976
2 social      -0.390       0.109   -3.59   0.000784
3 age          0.00882     0.104    0.0849 0.933
```

```r
m3 <- lm(esteem ~ social + gender, data = esteem)
tidy(m3)
```

```
# A tibble: 3 x 5
  term        estimate std.error statistic  p.value
  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)    6.90     0.699      9.87 4.85e-13
2 social        -0.273    0.0883    -3.09 3.31e- 3
3 genderWoman   -2.73     0.514     -5.31 2.95e- 6
```

```r
m4 <- lm(esteem ~ social + age + gender, data = esteem)
tidy(m4)
```

```
# A tibble: 4 x 5
  term        estimate std.error statistic    p.value
  <chr>          <dbl>     <dbl>     <dbl>      <dbl>
1 (Intercept)   5.19      1.82       2.85 0.00650
2 social       -0.277     0.0883    -3.14 0.00298
3 age           0.0844    0.0833     1.01 0.316
4 genderWoman  -2.81      0.521     -5.40 0.00000224
```

```r
m5 <- lm(esteem ~ social + age + gender + race, data = esteem)
tidy(m5)
```

```
# A tibble: 7 x 5
  term        estimate std.error statistic       p.value
  <chr>          <dbl>     <dbl>     <dbl>         <dbl>
1 (Intercept)   4.24      1.59       2.67  0.0106
2 social       -0.188     0.0878    -2.14  0.0381
3 age           0.0774    0.0681     1.14  0.262
4 genderWoman  -3.39      0.437     -7.75  0.00000000108
5 raceLatinx   -0.195     0.610     -0.321 0.750
6 raceOther    -1.12      1.23      -0.912 0.367
7 raceWhite     2.11      0.513      4.12  0.000171
```

```r
notes <- c("t statistics in parentheses",
           "confidence intervals in brackets",
           "(+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001)")
```

```r

together <- modelsummary(list("Social Media" = m1,
                              "+ Age" = m2,
                              "+ Gender" = m3,
                              "+ Age & Gender" = m4,
                              "+ Age, Gender, & Race" = m5),
             coef_rename = c(social = "Social Media",
                             age = "Age",
                             genderWoman = "Woman",
                             raceLatinx = "Latinx",
                             raceOther = "Other",
                             raceWhite = "White"),
             output = "kableExtra",
             estimate = "{estimate}{stars}",
             statistic = "({statistic}) <br> p = {p.value} <br> [{conf.low}, {conf.high}]"
             fmt =  3,
             gof_omit = "IC|Log|Adj|p\\.value|statistic|se_type|F|RMSE|R2") %>%
  row_spec(c(1,3,5,7,9,11,13), background = "#AC9EE8") %>%
  column_spec(2:5, width = "9em") %>%
  footnote(general = notes, general_title = "", footnote_as_chunk = FALSE)

together %>%
  kable_styling(font_size = 12)
```

	Social Media	+ Age	+ Gender	+ Age & Gende
(Intercept)	6.353***	6.173**	6.897***	5.192**
	(7.346) p = <0.001 [4.615, 8.092]	(2.694) p = 0.010 [1.563, 10.784]	(9.872) p = <0.001 [5.491, 8.302]	(2.851) p 0.006 [1.5 8.857]
Social Media	−0.389***	−0.390***	−0.273**	−0.277**
	(−3.635) p = <0.001 [−0.604, −0.174]	(−3.591) p = <0.001 [−0.608, −0.171]	(−3.095) p = 0.003 [−0.451, −0.096]	(−3.136) p 0.003 [−0.4 −0.099]
Age		0.009		0.084
		(0.085) p = 0.933 [−0.200, 0.218]		(1.014) p 0.316 [−0.0 0.252]
Woman			−2.725***	−2.814***
			(−5.307) p = <0.001 [−3.759, −1.692]	(−5.404) p <0.001 [−3.862, −1.76
Latinx				
Other				
White				
Num.Obs.	50	50	50	50

t statistics in parentheses
confidence intervals in brackets
(+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001)