

MRP_analysis.R

n8than

2024-06-19

```
# multilevel regression and post-stratification analysis

library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(glue)
library(gtsummary)

## Warning: package 'gtsummary' was built under R version 4.3.3

# # load regression data
# {
#   data <-
#     haven::read_dta(
#       file = "C:/Users/Nathan/Documents/Newham Fellowship/data/Skills for Life Survey 2011/UKDA-7240-
#
#   save(data, file = "data/skills_for_life_data.RData")
# }

load(here::here("data/skills_for_life_data.RData"))

# select variables

data <-
  data |>
  select(
    WORKINGSTATUS2,
    GROSS_ANNUAL_INCOME_OLDBANDS,
    BUK,
    QxTenu1,
    Sex1,
    AGE1NET,
    Sesol, # is English first language
    ETHNICSIMPLE,
```

```

HIQUAL,
CLITSPEAK,          # ENFL everyday English skills (literacy and speaking)
IMDScoreB4,         # Index of Multiple Deprivation banded into deciles
NSSEC7,
# outcomes
SUMMARYCOMP,        # self-assessed computer skills (summary)
TSKILLA,            # self-assessed computer skills (summary 2)
COMBLIT,            # self-assessed reading a writing (summary)
LiteracyScoreA_1,    # literacy level
starts_with("LiteracyThreshold"), # literacy threshold
NumeracyScoreA_1,    # numeracy level
starts_with("NumeracyThreshold"), # numeracy threshold,
MultipleChoiceLevelA_1, # ICT level
MultipleChoiceLevelA_1Thres, # ICT threshold
# weights
rimweight2003,
rimweightLIT2003,
rimweightNUM2003,
rimweightICT2003,
rimweightNUMICT2003,
rimweightLITICT2003,
rimweightLITNUM2003
)

# these are the S4L variables used in
# Rowlands (2015) British Journal of General Practice
#
# job status: National Statistics Socioeconomic Classification 3 bands (Managerial/professional, Intern
# employment status: employed, not employed
# gross income: >=10000, <10000
# place of birth: UK, non UK
# home ownership: Owns or part-owns home, does not own home
# sex: male, female
# age: 16-44, >=45
# first language: English, other
# ethnicity: white, black and minority ethnic
# qualification level: NQF >= level at age 16 (level 2), below level 2
# area deprivation: IMD quintiles

# matching with survey

# WORKINGSTATUS2: 0-No, 1-Yes
# GROSS_ANNUAL_INCOME_OLDBANDS: {<£5,000, £5,000 - £9,999}, {£10,000 - £14,999, £15,000 - £19,999, £20,
# BUK: 1-Yes, 2-No
# QxTenu1: 1-Own home outright or with a mortgage or loan
# Sex1: 1-Male, 2-Female
# AGE1NET: {16-24, 25-44}, 45-65
# Sesol: 1-Yes, 2-No
# ETHNICSIMPLE: 1-White, 2-BME
# HIQUAL: {1-4}, {5-Level 1 qualification or below}
# IMDScoreB4: 1,...,9
# NSSEC7: 1 Higher managerial and professional

```

```

# 2 Lower managerial and professional
# 3 Intermediate
# 4 Small employers and own account workers
# 5 Lower supervisory and technical
# 6 Semi-routine occupations
# 7 Routine occupations
# 8 Never worked/ long term unemployed
# 9 Full-time student
# 10 Not classifiable

#####
# data cleaning

model_dat <-
  data |>
  mutate(
    WORKINGSTATUS2 = unclass(WORKINGSTATUS2),
    GROSS_ANNUAL_INCOME_OLDBANDS = unclass(GROSS_ANNUAL_INCOME_OLDBANDS),
    BUK = unclass(BUK),
    QxTenu1 = unclass(QxTenu1),
    Sex1 = unclass(Sex1),
    AGE1NET = unclass(AGE1NET),
    Sesol = unclass(Sesol),
    ETHNICSIMPLE = unclass(ETHNICSIMPLE),
    HIQUAL = unclass(HIQUAL),
    IMDSCOREB4 = unclass(IMDSCOREB4),
    NSSEC7 = unclass(NSSEC7),
    LiteracyThresholdA_1 = unclass(LiteracyThresholdA_1),
    NumeracyThresholdA_1 = unclass(NumeracyThresholdA_1),
    MultipleChoiceLevelA_1Thres = unclass(MultipleChoiceLevelA_1Thres),
    LiteracyScoreA_1 = unclass(LiteracyScoreA_1),
    NumeracyScoreA_1 = unclass(NumeracyScoreA_1)) |>
  transmute(
    workingstatus = factor(WORKINGSTATUS2, levels = 0:1, labels = c("No", "Yes")),
    gross_income =
      ifelse(GROSS_ANNUAL_INCOME_OLDBANDS %in% 1:2,
        "<10000",
        ifelse(GROSS_ANNUAL_INCOME_OLDBANDS %in% 3:6,
          ">=10000", "other")) |>
      as.factor(),
    uk_born = factor(BUK, levels = 1:2, labels = c("Yes", "No")),
    sex = factor(Sex1, levels = c(2,1), c("Female", "Male")),
    own_home = ifelse(QxTenu1 == 1, "Yes", "No") |> as.factor(),
    age = ifelse(AGE1NET %in% 1:2, "16-44",
      ifelse(AGE1NET == 3, ">=45", "other")) |>
      as.factor(),
    english_lang = factor(Sesol, levels = 1:2, labels = c("Yes", "No")),
    ethnicity = factor(ETHNICSIMPLE, levels = 1:2, labels = c("White", "BME")),
    qualification = ifelse(HIQUAL %in% 1:4, ">=level 2", "<=Level 1") |>
      as.factor(),
    imd = factor(IMDSCOREB4),
    job_status = ifelse(NSSEC7 %in% 1:2, "higher",
      ifelse(NSSEC7 == 3, "intermediate",

```

```

        ifelse(NSSEC7 %in% 4:10, "lower", "other")))) |>
  as.factor(),
  lit_thresholdL1 =
    ifelse(LiteracyThresholdA_1 == 1, "below",
           ifelse(LiteracyThresholdA_1 == 2, "above", "other")),
  lit_thresholdL2 = ifelse(LiteracyScoreA_1 == 5, "above",
                           ifelse(LiteracyScoreA_1 %in% 1:4, "below", "other")), # >= L2
  num_thresholdEL3 =
    ifelse(NumeracyThresholdA_1 == 1, "below",
           ifelse(NumeracyThresholdA_1 == 2, "above", "other")),
  num_thresholdL1 = ifelse(NumeracyScoreA_1 == 4:5, "above",
                           ifelse(NumeracyScoreA_1 %in% 1:3, "below", "other")), # >= L1
  ict_thresholdEL3 =
    ifelse(MultipleChoiceLevelA_1Thres == 1, "below",
           ifelse(MultipleChoiceLevelA_1Thres == 2, "above", "other")),
  weights = unclass(rimweight2003),
  lit_weightsL1 = unclass(rimweightLIT2003),
  num_weightsEL3 = unclass(rimweightNUM2003),
  ict_weightsEL3 = unclass(rimweightICT2003)
)

summary(model_dat)

```

```

##  workingstatus  gross_income  uk_born      sex      own_home      age
##  No :2319      <10000 : 979    Yes:6309   Female:4110  No :2972  >=45 :3315
##  Yes:4911     >=10000:2829   No : 921   Male :3120   Yes:4258  16-44:3912
##                other :3422                                     other: 3
##
##
##
##  english_lang ethnicity      qualification      imd      job_status
##  Yes:6620    White:6450  <=Level 1:2224  2      :2364    higher :2721
##  No : 610    BME : 776  >=level 2:5006  1      :1520    intermediate: 736
##                NA's : 4                                     3      :1290    lower :3773
##                4      : 892
##                5      : 619
##                6      : 306
##                (Other): 239
##  lit_thresholdL1  lit_thresholdL2  num_thresholdEL3  num_thresholdL1
##  Length:7230     Length:7230     Length:7230       Length:7230
##  Class :character Class :character   Class :character   Class :character
##  Mode :character  Mode :character   Mode :character     Mode :character
##
##
##
##  ict_thresholdEL3  weights      lit_weightsL1  num_weightsEL3
##  Length:7230      Min. : 0.3228  Min. : 0.3276  Min. :0.3085
##  Class :character  1st Qu.: 0.5511  1st Qu.: 0.5550  1st Qu.:0.5650
##  Mode :character   Median : 0.9245  Median : 0.9265  Median :0.9235
##                   Mean : 1.0000  Mean : 1.0000  Mean :1.0000
##                   3rd Qu.: 1.2172  3rd Qu.: 1.2141  3rd Qu.:1.2292

```

```
##               Max.      :11.0320   Max.      :10.6931   Max.      :8.8591
##               NA's      :1181      NA's      :1177
## ict_weightsEL3
## Min.      : 0.347
## 1st Qu.: 0.541
## Median : 0.903
## Mean      : 1.000
## 3rd Qu.: 1.200
## Max.      :11.744
## NA's      :4872
```

```
# test specific data sets
lit_dat <- model_dat |>
  filter(lit_thresholdL2 %in% c("above", "below")) |>
  mutate(lit_thresholdL2 = as.factor(lit_thresholdL2))

num_dat <- model_dat |>
  filter(num_thresholdL1 %in% c("above", "below")) |>
  mutate(num_thresholdL1 = as.factor(num_thresholdL1))

ict_dat <- model_dat |>
  filter(ict_thresholdEL3 %in% c("above", "below")) |>
  mutate(ict_thresholdEL3 = as.factor(ict_thresholdEL3))

#####
# summary stats

lit_dat$lit_thresholdL2 |> table() |> prop.table()
```

```
##
##      above      below
## 0.5693681 0.4306319
```

```
#####
# logistic regressions
```

```
rhs <- "1 + workingstatus + gross_income + uk_born + sex + own_home + age + english_lang + ethnicity +"
```

```
# unweighted
```

```
lit_glm <- glm(glue("lit_thresholdL2 ~ {rhs}"), data = lit_dat, family = binomial(), weights = weights)
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
```

```
# lit_glm
suppressWarnings({
  tbl_regression(lit_glm, exponentiate = TRUE)
})
```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danielsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	OR	95% CI	p-value
workingstatus			
No	—	—	
Yes	1.21	1.02, 1.44	0.029

Characteristic	OR	95% CI	p-value
gross_income			
<10000	—	—	
>=10000	0.90	0.74, 1.08	0.2
other	1.38	1.13, 1.69	0.002
uk_born			
Yes	—	—	
No	1.09	0.85, 1.39	0.5
sex			
Female	—	—	
Male	1.21	1.07, 1.36	0.002
own_home			
No	—	—	
Yes	0.70	0.62, 0.80	<0.001
age			
>=45	—	—	
16-44	0.66	0.58, 0.75	<0.001
other	0.67	0.06, 11.2	0.7
english_lang			
Yes	—	—	
No	2.32	1.74, 3.09	<0.001
ethnicity			
White	—	—	
BME	1.40	1.14, 1.72	0.001
qualification			
<=Level 1	—	—	
>=level 2	0.36	0.31, 0.41	<0.001
imd			
1	—	—	
2	1.26	1.07, 1.49	0.005
3	1.48	1.22, 1.78	<0.001
4	1.78	1.44, 2.19	<0.001
5	2.23	1.75, 2.85	<0.001
6	2.68	1.90, 3.80	<0.001
7	1.74	1.15, 2.66	0.009
8	1.51	0.78, 3.00	0.2
9	0.73	0.09, 6.39	0.8
job_status			
higher	—	—	
intermediate	1.16	0.94, 1.44	0.2
lower	2.04	1.78, 2.33	<0.001

```
num_glm <- glm(glue("num_thresholdL1 ~ {rhs}"), data = num_dat, family = binomial(), weights = weights)
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
```

```
# num_glm
suppressWarnings({
  tbl_regression(num_glm, exponentiate = TRUE)
})
```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danielsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	OR	95% CI	p-value
workingstatus			
No	—	—	
Yes	1.01	0.81, 1.25	>0.9
gross_income			
<10000	—	—	
>=10000	0.81	0.65, 1.01	0.059
other	1.30	1.02, 1.67	0.035
uk_born			
Yes	—	—	
No	1.02	0.76, 1.38	0.9
sex			
Female	—	—	
Male	0.52	0.45, 0.60	<0.001
own_home			
No	—	—	
Yes	0.61	0.52, 0.71	<0.001
age			
>=45	—	—	
16-44	0.83	0.71, 0.97	0.018
other	0.49	0.03, 10.1	0.6
english_lang			
Yes	—	—	
No	0.77	0.54, 1.10	0.2
ethnicity			
White	—	—	
BME	2.33	1.79, 3.06	<0.001
qualification			
<=Level 1	—	—	
>=level 2	0.34	0.28, 0.40	<0.001
imd			
1	—	—	
2	1.09	0.90, 1.31	0.4
3	1.61	1.29, 2.01	<0.001
4	1.95	1.50, 2.54	<0.001
5	2.24	1.66, 3.04	<0.001
6	1.84	1.21, 2.86	0.005
7	2.83	1.58, 5.40	<0.001
8	2.43	1.01, 7.02	0.068
9	128,413	0.00, NA	>0.9
job_status			
higher	—	—	
intermediate	1.32	1.04, 1.69	0.025
lower	2.27	1.93, 2.66	<0.001

```

ict_glm <- glm(glue("ict_thresholdEL3 ~ {rhs}"), data = ict_dat, family = binomial(), weights = weights)

## Warning in eval(family$initialize): non-integer #successes in a binomial glm!

# ict_glm
suppressWarnings({
  tbl_regression(num_glm, exponentiate = TRUE)
})

```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## https://www.danielsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	OR	95% CI	p-value
workingstatus			
No	—	—	
Yes	1.01	0.81, 1.25	>0.9
gross_income			
<10000	—	—	
>=10000	0.81	0.65, 1.01	0.059
other	1.30	1.02, 1.67	0.035
uk_born			
Yes	—	—	
No	1.02	0.76, 1.38	0.9
sex			
Female	—	—	
Male	0.52	0.45, 0.60	<0.001
own_home			
No	—	—	
Yes	0.61	0.52, 0.71	<0.001
age			
>=45	—	—	
16-44	0.83	0.71, 0.97	0.018
other	0.49	0.03, 10.1	0.6
english_lang			
Yes	—	—	
No	0.77	0.54, 1.10	0.2
ethnicity			
White	—	—	
BME	2.33	1.79, 3.06	<0.001
qualification			
<=Level 1	—	—	
>=level 2	0.34	0.28, 0.40	<0.001
imd			
1	—	—	
2	1.09	0.90, 1.31	0.4
3	1.61	1.29, 2.01	<0.001
4	1.95	1.50, 2.54	<0.001
5	2.24	1.66, 3.04	<0.001
6	1.84	1.21, 2.86	0.005
7	2.83	1.58, 5.40	<0.001
8	2.43	1.01, 7.02	0.068
9	128,413	0.00, NA	>0.9
job_status			
higher	—	—	
intermediate	1.32	1.04, 1.69	0.025
lower	2.27	1.93, 2.66	<0.001