

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)

# # 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)
IMDSOREB4,          # 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)
starts_with("LiteracyScore"),    # literacy level
starts_with("LiteracyThreshold"), # literacy threshold
starts_with("NumeracyScore"),    # literacy level
starts_with("NumeracyThreshold"), # literacy 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, Intermed
# 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}
# IMDSOREB4: 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)) |>
  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 = 1:2, labels = c("Male", "Female")),
    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(),
    literacy_threshold =

```

```

    ifelse(LiteracyThresholdA_1 == 1, "below",
           ifelse(LiteracyThresholdA_1 == 2, "above", "other")),
  numeracy_threshold =
    ifelse(NumeracyThresholdA_1 == 1, "below",
           ifelse(NumeracyThresholdA_1 == 2, "above", "other")),
  ict_threshold =
    ifelse(MultipleChoiceLevelA_1Thres == 1, "below",
           ifelse(MultipleChoiceLevelA_1Thres == 2, "above", "other")),
  lit_weights = unclass(rimweightLIT2003),
  num_weights = unclass(rimweightNUM2003),
  ict_weights = unclass(rimweightICT2003)
)

summary(model_dat)

## workingstatus gross_income uk_born sex own_home age
## No :2319 <10000 : 979 Yes:6309 Male :3120 No :2972 >=45 :3315
## Yes:4911 >=10000:2829 No : 921 Female:4110 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
## literacy_threshold numeracy_threshold ict_threshold lit_weights
## Length:7230 Length:7230 Length:7230 Min. : 0.3276
## Class :character Class :character Class :character 1st Qu.: 0.5550
## Mode :character Mode :character Mode :character Median : 0.9265
## Mean : 1.0000
## 3rd Qu.: 1.2141
## Max. :10.6931
## NA's :1181
## num_weights ict_weights
## Min. :0.3085 Min. : 0.347
## 1st Qu.:0.5650 1st Qu.: 0.541
## Median :0.9235 Median : 0.903
## Mean :1.0000 Mean : 1.000
## 3rd Qu.:1.2292 3rd Qu.: 1.200
## Max. :8.8591 Max. :11.744
## NA's :1177 NA's :4872

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

num_dat <- model_dat |>
  filter(numeracy_threshold %in% c("above", "below")) |>

```

```

mutate(numeracy_threshold = as.factor(numeracy_threshold))

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

#####
# logistic regressions

rhs <- "1 + workingstatus + gross_income + english_lang + ethnicity + qualification + imd + job_status"

# unweighted
lit_glm <- glm(glue("literacy_threshold ~ {rhs}"), data = lit_dat, family = binomial(), weights = lit_w

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

##
## Call:  glm(formula = glue("literacy_threshold ~ {rhs}"), family = binomial(),
##      data = lit_dat, weights = lit_weights)
##
## Coefficients:
##      (Intercept)      workingstatusYes      gross_income>=10000
##           -2.88009              0.05633              -0.08237
##      gross_incomeother      english_langNo      ethnicityBME
##           0.71302              1.29991              0.64666
##      qualification>=level 2              imd2              imd3
##          -1.36638              0.55541              0.88888
##           imd4              imd5              imd6
##           1.14594              0.95153              1.24784
##           imd7              imd8              imd9
##           1.06202              0.21174             -0.51934
##      job_statusintermediate      job_statuslower
##          -0.13855              0.75989
##
## Degrees of Freedom: 5820 Total (i.e. Null);  5804 Residual
## (3 observations deleted due to missingness)
## Null Deviance:      4879
## Residual Deviance: 3801  AIC: 3635

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

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

##
## Call:  glm(formula = glue("numeracy_threshold ~ {rhs}"), family = binomial(),
##      data = num_dat, weights = num_weights)
##
## Coefficients:
##      (Intercept)      workingstatusYes      gross_income>=10000
##          -1.4825              -0.1868              -0.2968
##      gross_incomeother      english_langNo      ethnicityBME
##           0.2093              0.2605              0.6130

```

```

## qualification>=level 2          imd2          imd3
##          -1.0237          0.2921          0.5162
##          imd4          imd5          imd6
##          0.6722          0.9401          0.8459
##          imd7          imd8          imd9
##          0.7648          0.2473          1.5389
## job_statusintermediate      job_statuslower
##          0.2846          0.8088
##
## Degrees of Freedom: 5818 Total (i.e. Null);  5802 Residual
## (4 observations deleted due to missingness)
## Null Deviance:          6374
## Residual Deviance: 5455  AIC: 5277

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

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

ict_glm

##
## Call:  glm(formula = glue("ict_threshold ~ {rhs}"), family = binomial(),
## data = ict_dat, weights = ict_weights)
##
## Coefficients:
##          (Intercept)      workingstatusYes      gross_income>=10000
##          -2.24851          -0.12154          -0.38160
##      gross_incomeother      english_langNo      ethnicityBME
##          0.28195          0.20944          0.41542
## qualification>=level 2          imd2          imd3
##          -1.75947          -0.06414          0.09380
##          imd4          imd5          imd6
##          0.47487          0.36204          0.18083
##          imd7          imd8          imd9
##          1.15077          -0.27724          -0.33285
## job_statusintermediate      job_statuslower
##          -0.02323          0.93603
##
## Degrees of Freedom: 2273 Total (i.e. Null);  2257 Residual
## Null Deviance:          1414
## Residual Deviance: 1129  AIC: 1078

```