STATS 506 Final Project

Calder Moore

Cleaning

Interested in the RURALURBAN and the FUNDING catgeoricals. Need to deal with the weights too. I think the survey pacakage can help with this.

This Pew Research page in the report has an example doing some regression in R with weighted survey data.

```
non <- read.csv("nonprofit-survey-spring-2021-puf.csv")

# Create variable for proportion of funding from each source for each nonprofit

# Only want funding for year 2021

# Use FINANCES3_1_1 - FINANCES3-8-1 to create variables for proportion of funding from each s

# Need to deal with the -99 entries for non response. Turn to NA? I think more justifiable th

non$Finances_3_1_1[non$Finances_3_1_1 == -99] <- NA
non$Finances_3_2_1[non$Finances_3_2_1 == -99] <- NA
non$Finances_3_3_1[non$Finances_3_3_1 == -99] <- NA
non$Finances_3_4_1[non$Finances_3_4_1 == -99] <- NA
non$Finances_3_5_1[non$Finances_3_5_1 == -99] <- NA
non$Finances_3_6_1[non$Finances_3_6_1 == -99] <- NA
non$Finances_3_7_1[non$Finances_3_7_1 == -99] <- NA
non$Finances_3_8_1[non$Finances_3_8_1 == -99] <- NA

# To clean RuralUrban_1, does it makes sense to turn -99 to 0 even though I didn't do it for

non$RuralUrban_1[non$RuralUrban_1 == -99] <- 0
non$RuralUrban_2[non$RuralUrban_2 == -99] <- 0
non$RuralUrban_3[non$RuralUrban_3 == -99] <- 0
```

```r
non$RuralUrban_4[non$RuralUrban_4 == -99] <- 0

# Create the funding variables as proportions of total funding

non$FundPropGovAg <- non$Finances_3_1_1 / rowSums(non[, 188:195], na.rm = TRUE)
non$FundPropSelfPay <- non$Finances_3_2_1 / rowSums(non[, 188:195], na.rm = TRUE)
non$FundPropGov3Party <- non$Finances_3_3_1 / rowSums(non[, 188:195], na.rm = TRUE)
non$FundPropIndDon <- non$Finances_3_4_1 / rowSums(non[, 188:195], na.rm = TRUE)
non$FundPropGift <- non$Finances_3_5_1 / rowSums(non[, 188:195], na.rm = TRUE)
non$FundPropFound <- non$Finances_3_6_1 / rowSums(non[, 188:195], na.rm = TRUE)
non$FundPropCorp <- non$Finances_3_7_1 / rowSums(non[, 188:195], na.rm = TRUE)
non$FundPropOther <- non$Finances_3_8_1 / rowSums(non[, 188:195], na.rm = TRUE)

# Create some other variables to be covariates. Number of staff (STAFF_3_1_1), include the du

# The PROGDEM variables have a 0, 1, 2 levels for don't serve group (0), this is a primary gr

# Clean staff

non$Staff_3_1_1[non$Staff_3_1_1 == -99] <- NA
non$Staff_3_2_1[non$Staff_3_2_1 == -99] <- NA
non$Staff_3_3_1[non$Staff_3_3_1 == -99] <- NA
non$Staff_3_4_1[non$Staff_3_4_1 == -99] <- NA
non$Staff_3_5_1[non$Staff_3_5_1 == -99] <- NA
non$Staff_3_6_1[non$Staff_3_6_1 == -99] <- NA
non$Staff_3_7_1[non$Staff_3_7_1 == -99] <- NA

# Combine all forms of staff into one Total Staff variable just to approximate organization s

non$TotalStaff <- rowSums(non[, 94:100], na.rm = TRUE)

# Clean groups served

for (i in 1:21) {
  varname <- paste0("ProgDem_", i)

  non[[varname]][non[[varname]] == -99 | non[[varname]] == 99] <- NA

  non[[varname]][non[[varname]] == 2] <- 1
}
```

Modeling

```
# Quasibinomial since we have [0,1]. Will need to use svyglm() from survey package with famil

library(survey)
```

```
Warning: package 'survey' was built under R version 4.5.2
```

```
Loading required package: grid
```

```
Loading required package: Matrix
```

```
Loading required package: survival
```

```

Attaching package: 'survey'
```

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

    dotchart
```

```
# Survey design
design <- svydesign(
  ids = ~1,
  weights = ~weight_complete_partials,   # Use complete partials to include every nonprofit
  data = non
)

# Quasibinomial regression. Do one for each funding proportion. Baseline categorical predicto
modGovAg <- svyglm(
  FundPropGovAg ~ RuralUrban_1 + RuralUrban_2 + RuralUrban_3 +
                  TotalStaff + ProgDem_1 + ProgDem_2 + ProgDem_3 + ProgDem_4 + ProgDem_5 + Pr
  design = design,
  family = quasibinomial(link = "logit")
)

summary(modGovAg)
```

```
Call:
svyglm(formula = FundPropGovAg ~ RuralUrban_1 + RuralUrban_2 +
    RuralUrban_3 + TotalStaff + ProgDem_1 + ProgDem_2 + ProgDem_3 +
    ProgDem_4 + ProgDem_5 + ProgDem_6 + ProgDem_7 + ProgDem_8 +
    ProgDem_9 + ProgDem_10 + ProgDem_11 + ProgDem_12 + ProgDem_13 +
    ProgDem_14 + ProgDem_15 + ProgDem_16 + ProgDem_17 + ProgDem_18 +
    ProgDem_19 + ProgDem_20 + ProgDem_21, design = design, family = quasibinomial(link = "log

Survey design:
svydesign(ids = ~1, weights = ~weight_complete_partials, data = non)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.431783   0.641585   2.232  0.03277 *
RuralUrban_1 -1.442238   0.513760  -2.807  0.00844 **
RuralUrban_2 -0.618713   0.431069  -1.435  0.16090
RuralUrban_3  0.496407   0.487384   1.019  0.31607
TotalStaff   -0.001873   0.001563  -1.198  0.23962
ProgDem_1    -0.545380   0.979068  -0.557  0.58138
ProgDem_2     0.701384   1.380414   0.508  0.61487
ProgDem_3     0.357697   1.613138   0.222  0.82593
ProgDem_4    -0.259933   0.560952  -0.463  0.64623
ProgDem_5     2.093588   0.705893   2.966  0.00567 **
ProgDem_6    -0.057938   0.620914  -0.093  0.92624
ProgDem_7     0.774489   0.997303   0.777  0.44311
ProgDem_8    -2.090257   0.862668  -2.423  0.02123 *
ProgDem_9     0.207696   0.956205   0.217  0.82942
ProgDem_10   -1.094128   1.243898  -0.880  0.38564
ProgDem_11   -0.456243   1.231568  -0.370  0.71348
ProgDem_12   -0.956564   1.700495  -0.563  0.57768
ProgDem_13    0.280213   1.969481   0.142  0.88775
ProgDem_14    2.136148   1.888414   1.131  0.26638
ProgDem_15   -0.262754   0.620730  -0.423  0.67491
ProgDem_16    1.765895   1.775148   0.995  0.32730
ProgDem_17   -0.132687   1.490091  -0.089  0.92960
ProgDem_18   -0.294960   2.381728  -0.124  0.90221
ProgDem_19   -1.465317   1.593187  -0.920  0.36459
ProgDem_20   -0.710734   0.581491  -1.222  0.23054
ProgDem_21   -0.029099   0.838908  -0.035  0.97255
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for quasibinomial family taken to be 0.3698369)

Number of Fisher Scoring iterations: 5
```

```r
library(car)
```

```
Loading required package: carData
```

```r
vif(modGovAg)
```

```
RuralUrban_1 RuralUrban_2 RuralUrban_3   TotalStaff    ProgDem_1    ProgDem_2
    4.706117     2.741772     4.839018    13.198758    14.623949    25.149730
   ProgDem_3    ProgDem_4    ProgDem_5    ProgDem_6    ProgDem_7    ProgDem_8
   36.248537     5.329208     8.764444     5.480469    14.074252    10.090709
   ProgDem_9   ProgDem_10   ProgDem_11   ProgDem_12   ProgDem_13   ProgDem_14
   12.610530    24.266510    27.864131    25.133795    36.236908    30.862573
  ProgDem_15   ProgDem_16   ProgDem_17   ProgDem_18   ProgDem_19   ProgDem_20
    8.322270    45.319983    33.705241    79.720087    35.961005     3.152434
  ProgDem_21
    2.350171
```

```r
# High VIF for many of the variables. Removing ProgDem variables in order of highest VIF brir

modGovAg2 <- svyglm(
  FundPropGovAg ~ RuralUrban_1 +
                  TotalStaff + ProgDem_1 + ProgDem_2 + ProgDem_4 + ProgDem_5 + ProgDem_6 + Pr
  design = design,
  family = quasibinomial(link = "logit")
)

vif(modGovAg2)
```

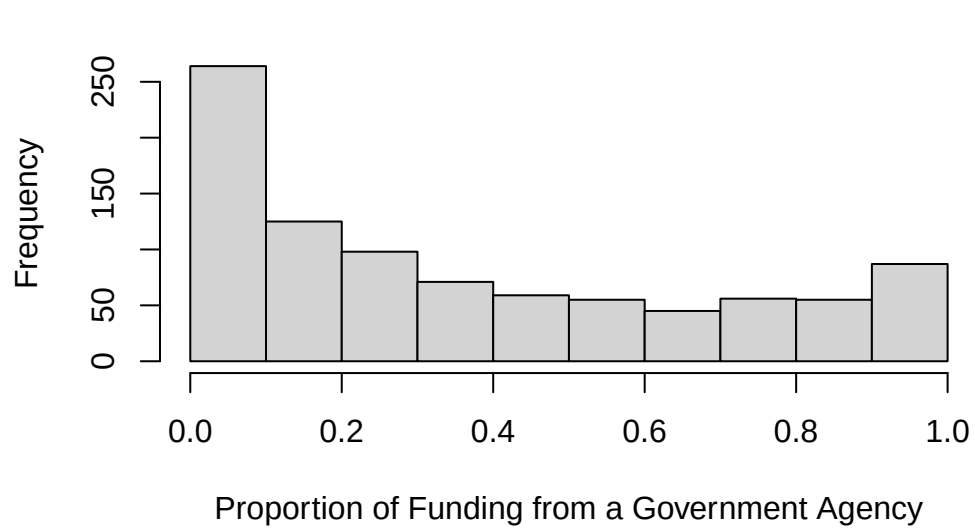
```
RuralUrban_1   TotalStaff    ProgDem_1    ProgDem_2    ProgDem_4    ProgDem_5
    2.413535     5.583503     6.161852     4.402303     2.073725     4.268490
   ProgDem_6    ProgDem_7    ProgDem_8    ProgDem_9   ProgDem_10   ProgDem_12
    3.428451     7.834004     5.720762     6.181232     6.365302     4.656543
  ProgDem_13   ProgDem_15   ProgDem_16   ProgDem_19   ProgDem_20   ProgDem_21
    6.951037     4.248827     7.035717     8.085225     2.349061     1.885404
```

```r
summary(modGovAg2)
```

```
Call:
svyglm(formula = FundPropGovAg ~ RuralUrban_1 + TotalStaff +
    ProgDem_1 + ProgDem_2 + ProgDem_4 + ProgDem_5 + ProgDem_6 +
    ProgDem_7 + ProgDem_8 + ProgDem_9 + ProgDem_10 + ProgDem_12 +
    ProgDem_13 + ProgDem_15 + ProgDem_16 + ProgDem_19 + ProgDem_20 +
    ProgDem_21, design = design, family = quasibinomial(link = "logit"))

Survey design:
svydesign(ids = ~1, weights = ~weight_complete_partials, data = non)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.204993   0.620650   1.942  0.05945 .
RuralUrban_1 -1.237323   0.478305  -2.587  0.01353 *
TotalStaff   -0.001705   0.001336  -1.277  0.20923
ProgDem_1    -0.272008   0.758797  -0.358  0.72192
ProgDem_2     0.673113   0.669163   1.006  0.32066
ProgDem_4    -0.306182   0.469405  -0.652  0.51805
ProgDem_5     2.122693   0.634430   3.346  0.00182 **
ProgDem_6     0.186946   0.596794   0.313  0.75576
ProgDem_7     0.689852   0.859727   0.802  0.42718
ProgDem_8    -2.010072   0.746720  -2.692  0.01041 *
ProgDem_9     0.202296   0.880787   0.230  0.81954
ProgDem_10   -1.187569   0.745179  -1.594  0.11908
ProgDem_12   -0.782671   0.820573  -0.954  0.34605
ProgDem_13    1.836616   0.963080   1.907  0.06390 .
ProgDem_15    0.023579   0.590663   0.040  0.96836
ProgDem_16    1.220093   0.819315   1.489  0.14449
ProgDem_19   -1.665765   0.907356  -1.836  0.07401 .
ProgDem_20   -0.837818   0.558583  -1.500  0.14169
ProgDem_21   -0.054329   0.806504  -0.067  0.94664
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 0.37745)

Number of Fisher Scoring iterations: 5
```

Export

```
# Export for use in report

# Want histogram to justify use of quasibinomial regression

hist(non$FundPropGovAg, main = "Histogram of Proportion of Funding from a Government Agency"
```




```
png("histogram.png", width=800, height=600)
hist(non$FundPropGovAg, main="Histogram of FundPropGovAg", xlab="FundPropGovAg")
dev.off()
```

```
pdf
  2
```

```
# Export

library(modelsummary)
```

```
Warning: package 'modelsummary' was built under R version 4.5.2
```

```
# Commenting out because the Tex compiler doesn't like the variable names with _

# modelsummary(
#   modGovAg2,
#   output = "model_table.tex",
#   stars = TRUE,
#   fmt = 3,
#   gof_omit = ".*",
#   escape = FALSE
# )

# Create a table of summary stats for our variables

vars <- c("FundPropGovAg", "RuralUrban_1", "TotalStaff",
          "ProgDem_1", "ProgDem_2", "ProgDem_3", "ProgDem_4",
          "ProgDem_5", "ProgDem_6", "ProgDem_7", "ProgDem_8",
          "ProgDem_9", "ProgDem_10", "ProgDem_11", "ProgDem_12",
          "ProgDem_13", "ProgDem_14", "ProgDem_15", "ProgDem_16")

sumstats <- non[, vars]




library(stargazer)
```

```
Warning: package 'stargazer' was built under R version 4.5.2
```

```
Please cite as:

 Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.

 R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```
# Commenting out because the Tex compiler doesn't like the variable names with _

# stargazer(sumstats,
#           type = "latex",
#           summary = TRUE,
#           out = "summary.tex")
```