Randomization assessment

Jake Bowers

January 11, 2021

This file assesses the randomization. We have no reason to doubt the YouGov randomizer. However, this is our standard practice to check to learn about possible chance covariate imbalances. Notice that we do not use survey weights here: our goal is to ask whether the randomization done *among the sampled respondents* worked as it should — i.e had no systematic relationship with a collection of covariates.

```
## here() starts at /Users/jwbowers/Documents/PROJECTS/COVID-YouGovSurveyAnalysis

## Loading required package: survival

## Loading required package: SparseM

##
## Attaching package: 'SparseM'

## The following object is masked from 'package:base':
##
##     backsolve
```

```
dat <- read.spss(here("data/Fourth_Wave", "TPL_Testing_Survey_FourthWave_YouGov_MERGEDWITHTHIRDWAVEFORMESSING.sav"), to.data.frame = TRUE)
```

```
## Treatment assignment q115_treat

# Looks like not everyone was included in the
## experiment in the fourth wave, that 500 people were excluded (perhaps not
## included in this wave)

table(dat$q115_treat, exclude = c())
```

```
   Family Community      <NA>
      249       251       500
```

```
## So, just focus on the valid respondents.

datw4 <- droplevels(dat[!is.na(dat$q115_treat), ])

## Some of the code below wants either treatment assignment or outcome to be either a factor variable (with labels) or a binary variable.
table(datw4$q115_treat, exclude = c())
```

```
   Family Community
      249       251
```

```
datw4$q115N <- as.numeric(datw4$q115_treat == "Community")
datw4$q115F <- factor(datw4$q115N)
```

We think that treatment assignment was done within strata of pid3 but are not sure. For now, presenting the analysis **both** ways.

Since I am not 100% sure I understand all of the recoding, I present a couple of different approaches here. First, I try to let R handle the categorical vs numeric variable issue. I think these are the relevant covariates where a chance imbalance could change our interpretation of the results.

Here I present the results for **both** the simply or completely randomized case ("unblocked") and the block-randomized case (where randomization occurred within strata/blocks of party id) ("pid").

```
## Trying to let R handle the dummy recoding
xb1 <- xBalance(q115N ~ q1 + age + gender_client + educ + dependents_dummy_coded + race + core_city,
  strata = list(notblocked = NULL, pid3 = ~pid3), data = datw4, report = "all"
)
```

```
## Overall no detectable departure from what we'd expect from a well executed random assignment
xb1$overall
```

```
           chisquare df p.value
notblocked     25.55 22  0.2716
pid3           25.20 22  0.2874
```

```
## And here just comparing the one-by-one p-values (without adjustment)

theps <- xb1$results[, "p", ]
theps_adj <- apply(theps, 2, function(x) {
  p.adjust(x, method = "holm")
})
pres <- data.frame(cbind(unadj = theps, adj = theps_adj))
names(pres)[3:4] <- paste("Adj:", names(pres)[3:4])
pres
```

```
                              notblocked     pid3 Adj: notblocked.1 Adj: pid3.1
q1Definitely won't              0.322731 0.412253            1.0000      1.0000
q1Maybe won't                   0.761879 0.698271            1.0000      1.0000
q1Not sure                      0.507580 0.431736            1.0000      1.0000
q1Maybe will                    0.661497 0.622476            1.0000      1.0000
q1Definitely will               0.725827 0.505230            1.0000      1.0000
age                             0.682188 0.786337            1.0000      1.0000
gender_clientMale               0.056803 0.057295            1.0000      1.0000
gender_clientFemale             0.057513 0.057043            1.0000      1.0000
gender_clientOther              0.319246 0.309629            1.0000      1.0000
gender_clientPrefer not to say  0.558240 0.517973            1.0000      1.0000
educNo HS                       0.006362 0.008773            0.1718      0.2369
educHigh school graduate        0.047555 0.025165            1.0000      0.6543
educSome college                0.705877 0.754355            1.0000      1.0000
educ2-year                      0.174243 0.199644            1.0000      1.0000
educ4-year                      0.644771 0.795427            1.0000      1.0000
educPost-grad                   0.437057 0.387745            1.0000      1.0000
dependents_dummy_coded          0.067179 0.066982            1.0000      1.0000
raceWhite                       0.971241 0.969594            1.0000      1.0000
raceBlack                       0.623111 0.575248            1.0000      1.0000
raceHispanic                    0.605798 0.551242            1.0000      1.0000
raceAsian                       0.406272 0.488105            1.0000      1.0000
raceNative American             0.995482 0.998567            1.0000      1.0000
raceTwo or more races           0.469361 0.397263            1.0000      1.0000
raceOther                       0.995482 0.971729            1.0000      1.0000
raceMiddle Eastern              0.558240 0.538924            1.0000      1.0000
core_cityNot core city          0.210747 0.221087            1.0000      1.0000
core_cityCore city              0.210747 0.221087            1.0000      1.0000
```

```
## More information (including unadjusted p-values)
xb1$results[, c("q115N=0", "q115N=1", "adj.diff", "std.diff", "p"), ]
```

```
, , strata = notblocked

                              stat
vars                             q115N=0   q115N=1  adj.diff  std.diff        p
  q1Definitely won't            0.136546  0.107570 -0.028976 -0.088445 0.322731
  q1Maybe won't                 0.076305  0.083665  0.007360  0.027078 0.761879
  q1Not sure                    0.164659  0.187251  0.022592  0.059233 0.507580
  q1Maybe will                  0.160643  0.175299  0.014656  0.039131 0.661497
  q1Definitely will             0.461847  0.446215 -0.015632 -0.031339 0.725827
  age                          47.088353 47.752988  0.664635  0.036595 0.682188
  gender_clientMale             0.429719  0.346614 -0.083105 -0.170825 0.056803
  gender_clientFemale           0.562249  0.645418  0.083169  0.170334 0.057513
  gender_clientOther            0.000000  0.003984  0.003984  0.089086 0.319246
  gender_clientPrefer not to say 0.008032 0.003984 -0.004048 -0.052331 0.558240
  educNo HS                     0.056225  0.011952 -0.044273 -0.245642 0.006362
  educHigh school graduate      0.188755  0.262948  0.074193  0.177740 0.047555
  educSome college              0.220884  0.235060  0.014176  0.033727 0.705877
  educ2-year                    0.120482  0.083665 -0.036817 -0.121630 0.174243
  educ4-year                    0.244980  0.262948  0.017968  0.041204 0.644771
  educPost-grad                 0.168675  0.143426 -0.025248 -0.069485 0.437057
  dependents_dummy_coded        0.244980  0.318725  0.073745  0.164112 0.067179
  raceWhite                     0.859438  0.860558  0.001120  0.003221 0.971241
  raceBlack                     0.028112  0.035857  0.007744  0.043923 0.623111
  raceHispanic                  0.060241  0.071713  0.011472  0.046126 0.605798
  raceAsian                     0.016064  0.007968 -0.008096 -0.074257 0.406272
  raceNative American           0.004016  0.003984 -0.000032 -0.000506 0.995482
  raceTwo or more races         0.020080  0.011952 -0.008128 -0.064683 0.469361
  raceOther                     0.004016  0.003984 -0.000032 -0.000506 0.995482
  raceMiddle Eastern            0.008032  0.003984 -0.004048 -0.052331 0.558240
  core_cityNot core city        0.546185  0.601594  0.055409  0.112003 0.210747
  core_cityCore city            0.453815  0.398406 -0.055409 -0.112003 0.210747
```

```
, , strata = pid3

                                stat
vars                               q115N=0   q115N=1   adj.diff   std.diff        p
  q1Definitely won't              0.132133  0.108609 -0.0235236 -0.0718014 0.412253
  q1Maybe won't                   0.076112  0.085478  0.0093659  0.0344574 0.698271
  q1Not sure                      0.163254  0.189964  0.0267106  0.0700300 0.431736
  q1Maybe will                    0.159799  0.176340  0.0165414  0.0441641 0.622476
  q1Definitely will               0.468703  0.439609 -0.0290943 -0.0583266 0.505230
  age                            47.218418 47.651181  0.4327630  0.0238278 0.786337
  gender_clientMale               0.431975  0.349147 -0.0828281 -0.1702553 0.057295
  gender_clientFemale             0.559791  0.643021  0.0832303  0.1704593 0.057043
  gender_clientOther              0.000000  0.004080  0.0040804  0.0912403 0.309629
  gender_clientPrefer not to say  0.008234  0.003752 -0.0044826 -0.0579479 0.517973
  educNo HS                       0.053901  0.011795 -0.0421059 -0.2336194 0.008773
  educHigh school graduate        0.183520  0.266500  0.0829794  0.1987880 0.025165
  educSome college                0.221742  0.233505  0.0117632  0.0279860 0.754355
  educ2-year                      0.118871  0.084342 -0.0345292 -0.1140729 0.199644
  educ4-year                      0.250841  0.260885  0.0100433  0.0230310 0.795427
  educPost-grad                   0.171124  0.142974 -0.0281508 -0.0774729 0.387745
  dependents_dummy_coded          0.244407  0.318485  0.0740779  0.1648521 0.066982
  raceWhite                       0.862975  0.861804 -0.0011716 -0.0033699 0.969594
  raceBlack                       0.027668  0.036503  0.0088356  0.0501138 0.575248
  raceHispanic                    0.057854  0.070927  0.0130729  0.0525622 0.551242
  raceAsian                       0.014823  0.008044 -0.0067799 -0.0621846 0.488105
  raceNative American             0.003953  0.003963  0.0000102  0.0001613 0.998567
  raceTwo or more races           0.020750  0.011255 -0.0094948 -0.0755591 0.397263
  raceOther                       0.003953  0.003752 -0.0002011 -0.0031793 0.971729
  raceMiddle Eastern              0.008023  0.003752 -0.0042713 -0.0552166 0.538924
  core_cityNot core city          0.549437  0.603380  0.0539426  0.1090394 0.221087
  core_cityCore city              0.450563  0.396620 -0.0539426 -0.1090394 0.221087
```

Checking the asymptotic assumption of `xBalance` (compare to `xb1$overall` above)

```
coin1_asympt <- independence_test(q1 + age + gender_client + educ + dependents_dummy_coded + race + core_city ~ q115F | pid3,
  data = datw4,
  teststat = "quadratic", distribution = asymptotic()
)
coin1_asympt
```

```
    Asymptotic General Independence Test

data:  q1, age, gender_client, educ, dependents_dummy_coded, race, core_city by q115F (0, 1)
     stratified by pid3
chi-squared = 25, df = 22, p-value = 0.3
```

```
pvalue(coin1_asympt)
```

```
[1] 0.2874
```

```
coin1_perm <- independence_test(q1 + age + gender_client + educ + dependents_dummy_coded + race + core_city ~ q115F | pid3,
  data = datw4,
  teststat = "quadratic", distribution = approximate(nresample = 5000)
)
coin1_perm
```

```
    Approximative General Independence Test

data:  q1, age, gender_client, educ, dependents_dummy_coded, race, core_city by q115F (0, 1)
     stratified by pid3
chi-squared = 25, p-value = 0.3
```

```
## This next shows a "confidence interval" because another 5000 sims might change the p-value
pvalue(coin1_perm)
```

```
[1] 0.2682
99 percent confidence interval:
 0.2522 0.2847
```

Here is a version using the recoding from SPSS. I'm not sure why we have a difference.

```
           chisquare df p.value
notblocked     22.22 14 0.07425
pid3           21.75 13 0.05939

, , strata = notblocked

                                        stat
vars                                      q115N=0  q115N=1  adj.diff  std.diff        p
  age                                    47.08835 47.75299  0.664635  0.036595 0.682188
  gender_dummy_coded_female               0.56225  0.64542  0.083169  0.170334 0.057513
  party_dummy_coded_republican            0.14859  0.15139  0.002800  0.007826 0.930207
```

```
  race_dummy_coded_black                          0.02811  0.03586  0.007744  0.043923 0.623111
  race_dummy_coded_latino                         0.06024  0.07171  0.011472  0.046126 0.605798
  core_city_excluding_Newport_and_noncentral_PVD  0.36948  0.31474 -0.054737 -0.115347 0.197483
  eighteen_to_twentynine_dummy_coded              0.22088  0.20717 -0.013712 -0.033372 0.708830
  sixty_and_over_dummy_coded                      0.29317  0.32669  0.033521  0.072380 0.418222
  low_SES_dummy_coded                             0.27309  0.24303 -0.030064 -0.068616 0.442805
  No_HS_dummy_coded                               0.05622  0.01195 -0.044273 -0.245642 0.006362
  HS_dummy_coded                                  0.18876  0.26295  0.074193  0.177740 0.047555
  Some_college_dummy_coded                        0.22088  0.23506  0.014176  0.033727 0.705877
  two_year_degree_dummy_coded                     0.12048  0.08367 -0.036817 -0.121630 0.174243
  dependents_dummy_coded                          0.24498  0.31873  0.073745  0.164112 0.067179

, , strata = pid3

                                                stat
vars                                              q115N=0  q115N=1   adj.diff   std.diff        p
  age                                            47.21842 47.65118  4.328e-01  2.383e-02 0.786337
  gender_dummy_coded_female                       0.55979  0.64302  8.323e-02  1.705e-01 0.057043
  party_dummy_coded_republican                    0.15060  0.15060 -2.676e-17 -7.478e-17 1.000000
  race_dummy_coded_black                          0.02767  0.03650  8.836e-03  5.011e-02 0.575248
  race_dummy_coded_latino                         0.05785  0.07093  1.307e-02  5.256e-02 0.551242
  core_city_excluding_Newport_and_noncentral_PVD  0.36609  0.31528 -5.081e-02 -1.071e-01 0.229367
  eighteen_to_twentynine_dummy_coded              0.21681  0.20838 -8.432e-03 -2.052e-02 0.817474
  sixty_and_over_dummy_coded                      0.29437  0.32460  3.023e-02  6.527e-02 0.461717
  low_SES_dummy_coded                             0.26981  0.24519 -2.462e-02 -5.620e-02 0.526827
  No_HS_dummy_coded                               0.05390  0.01180 -4.211e-02 -2.336e-01 0.008773
  HS_dummy_coded                                  0.18352  0.26650  8.298e-02  1.988e-01 0.025165
  Some_college_dummy_coded                        0.22174  0.23351  1.176e-02  2.799e-02 0.754355
  two_year_degree_dummy_coded                     0.11887  0.08434 -3.453e-02 -1.141e-01 0.199644
  dependents_dummy_coded                          0.24441  0.31849  7.408e-02  1.649e-01 0.066982

                                                strata
vars                                             notblocked     pid3
  age                                              0.682188 0.786337
  gender_dummy_coded_female                        0.057513 0.057043
  party_dummy_coded_republican                     0.930207 1.000000
  race_dummy_coded_black                           0.623111 0.575248
  race_dummy_coded_latino                          0.605798 0.551242
  core_city_excluding_Newport_and_noncentral_PVD   0.197483 0.229367
  eighteen_to_twentynine_dummy_coded               0.708830 0.817474
  sixty_and_over_dummy_coded                       0.418222 0.461717
  low_SES_dummy_coded                              0.442805 0.526827
  No_HS_dummy_coded                                0.006362 0.008773
  HS_dummy_coded                                   0.047555 0.025165
  Some_college_dummy_coded                         0.705877 0.754355
  two_year_degree_dummy_coded                      0.174243 0.199644
  dependents_dummy_coded                           0.067179 0.066982
```