Outcome analysis

Jake Bowers

January 11, 2021

This file assesses the relationship between treatment assignment and the outcome. The hypothesis is that differences in framing of the message should not change behavior among these people.

```
## 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)
datw4$vaccine_timing_interestN <- as.numeric(datw4$vaccine_timing_interest == "Clicked on list")
datw4$vaccine_timing_interestF <- factor(datw4$vaccine_timing_interestN)
```

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

Analysis as a block-randomized experiment: Treatment Effects in the Sample

First, assuming that treatment was assigned by coin flip within block. We present three analyses that use asymptotic assumptions and one that approximates the distribution of the test statistic directly using permutations of treatment assignment. All of these analyses might be called "randomization based statistical inference" — they differ in whether they assume that the randomization-based distribution is well approximated by a Normal distribution or not.

Since we are acting like randomization occurred within block, we have to represent that in the analysis. This means we are using the Cochran-Mantel-Haenszel test rather than the Chi-square test (equivalently we use a block-specific weighted analysis using OLS in the `difference_of_means` line).

```
## Two asymptotic analyses using tests like the CMH test:
res1 <- xBalance(q115N ~ vaccine_timing_interestN, strata = list(pid3 = ~pid3), data = datw4, report = "all")

res1
```

```
                       strata    pid3
                       stat   q115N=0 q115N=1 adj.diff adj.diff.null.sd std.diff    z
vars
vaccine_timing_interestN         0.291   0.216    -0.075           0.039    -0.172  -1.922 .
---Overall Test---
     chisquare df p.value
pid3       3.7  1   0.055
---
Signif. codes:  0 '***' 0.001 '** ' 0.01 '*  ' 0.05 '.  ' 0.1 '   ' 1
```

```
res2 <- cmh_test(vaccine_timing_interestF ~ q115F | pid3, data = datw4, distribution = asymptotic())

res2
```

```
    Asymptotic Generalized Cochran-Mantel-Haenszel Test

data:  vaccine_timing_interestF by q115F (0, 1)
     stratified by pid3
chi-squared = 3.7, df = 1, p-value = 0.05
```

```
## Least Squares Asymptotic Approximation using Neyman standard errors ("estimate" below is the same as "adj.diff" in xBalance)
res3 <- difference_in_means(vaccine_timing_interestN ~ q115N, blocks = pid3, data = datw4)

res3
```

```
Design:  Blocked
      Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
q115N -0.07474    0.03874  -1.929  0.05426  -0.1509 0.001373 490
```

```
set.seed(12345)
## Using the permutation distribution to check on the asymptotic approximation
res4 <- cmh_test(vaccine_timing_interestF ~ q115F | pid3, data = datw4, distribution = approximate(nresample = 10000))

res4
```

```
    Approximative Generalized Cochran-Mantel-Haenszel Test

data:  vaccine_timing_interestF by q115F (0, 1)
     stratified by pid3
chi-squared = 3.7, p-value = 0.06
```

Treatment Effects in subgroups

What is the effect of the treatment among those who are not definitely sure they will be vaccinated: these are the people whose behavior we would aim to change with different messaging strategies.

```
## First, we look for any chance imbalance (with prior vaccination intention): none found
table(datw4$q1, datw4$q115_treat, exclude = c())
```

```
                 Family Community
  Definitely won't     34        27
  Maybe won't          19        21
  Not sure             41        47
  Maybe will           40        44
  Definitely will     115       112
```

```
q1test <- cmh_test(q1 ~ q115F | pid3, data = datw4)
q1test
```

```
    Asymptotic Generalized Cochran-Mantel-Haenszel Test

data:  q1 by q115F (0, 1)
     stratified by pid3
chi-squared = 1.6, df = 4, p-value = 0.8
```

```
## Next we repreat the previous analysis, but excluding those who report that they were certain about vaccination.

datw4a <- droplevels(datw4[datw4$q1 != "Definitely will", ])

res1a <- xBalance(q115N ~ vaccine_timing_interestN,
  strata = list(pid3 = ~pid3),
```

```
  data = datw4a,
  report = "all"
)

res1a
```

```
                        strata    pid3
                        stat   q115N=0 q115N=1 adj.diff adj.diff.null.sd std.diff    z
vars
vaccine_timing_interestN        0.2577  0.1260   -0.1316          0.0481   -0.3363 -2.7396 **
---Overall Test---
     chisquare df p.value
pid3       7.5  1  0.0062
---
Signif. codes:  0 '***' 0.001 '** ' 0.01 '*  ' 0.05 '.  ' 0.1 '   ' 1
```

```
res2a <- cmh_test(vaccine_timing_interestF ~ q115F | pid3, data = datw4a, distribution = asymptotic())

res2a
```

```
    Asymptotic Generalized Cochran-Mantel-Haenszel Test

data:  vaccine_timing_interestF by q115F (0, 1)
     stratified by pid3
chi-squared = 7.5, df = 1, p-value = 0.006
```

```
## Least Squares Asymptotic Approximation using Neyman standard errors ("estimate" below is the same as "adj.diff" in xBalance)
res3a <- difference_in_means(vaccine_timing_interestN ~ q115N, blocks = pid3, data = datw4a)

res3a
```

```
Design:  Blocked
      Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
q115N  -0.1318    0.04752  -2.773 0.005956  -0.2253 -0.03819 263
```

```
## Using the permutation distribution to check on the asymptotic approximation
set.seed(12345)
res4a <- cmh_test(vaccine_timing_interestF ~ q115F | pid3, data = datw4a, distribution = approximate(nresample = 10000))

res4a
```

```
    Approximative Generalized Cochran-Mantel-Haenszel Test

data:  vaccine_timing_interestF by q115F (0, 1)
     stratified by pid3
chi-squared = 7.5, p-value = 0.009
```

Treatment Effects in the Population

The preceding analyses did not try to estimate the effect of the messages **in the population** from which the sample was drawn. We can also estimate these, although we will pay a price in terms of precision (see Miratrix, et al 2018 doi: 10.1017/pan.2018.1). Following that paper, we can estimate the effects in the population using the following approximate approach:

```
pate1 <- difference_in_means(vaccine_timing_interestN ~ q115N, blocks = pid3, data = datw4, weights = weight)
pate1
```

```
Design:  Blocked (weighted)
      Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper  DF
q115N  -0.1139    0.05082  -2.242   0.0254  -0.2138 -0.01409 490
```

We could check that approach by using a two-step procedure for statistical inference: first, a weighted bootstrap sample (using `weight`) and then, second, permuting treatment assignment within block, to create the distribution of the relevant test statistic. We do not do this here because the question about the effect of the messaging is not about the effects in the population from which we have sampled — after all, this messaging strategy will be used in many different populations. Rather, it is to update our thinking about the theory behind the strategy — and for that purpose, the sample-specific effect is good enough (and helps us avoid the hassle of coding up this two-step procedure.)

Analysis as a simple or completely randomized experiment

If the treatment was assigned completely at random, i.e, not within block.

```
## Fisher/Chisq tests
res5 <- chisq_test(vaccine_timing_interestF ~ q115F, data = datw4, distribution = asymptotic())
res5
```

```
    Asymptotic Pearson Chi-Squared Test

data:  vaccine_timing_interestF by q115F (0, 1)
chi-squared = 3.2, df = 1, p-value = 0.07
```

```
set.seed(12345)
res6 <- chisq_test(vaccine_timing_interestF ~ q115F, data = datw4, distribution = approximate(nresample = 10000))
res6
```

```
    Approximative Pearson Chi-Squared Test

data:  vaccine_timing_interestF by q115F (0, 1)
chi-squared = 3.2, p-value = 0.09
```

```
res7 <- xBalance(q115N ~ vaccine_timing_interestN, data = datw4, report = "all")
res7
```

```
                       strata unstrat
                       stat   q115N=0 q115N=1 adj.diff adj.diff.null.sd std.diff    z
vars
vaccine_timing_interestN         0.289   0.219    -0.070           0.039    -0.161  -1.797 .
---Overall Test---
        chisquare df p.value
unstrat       3.2  1   0.072
---
Signif. codes:  0 '***' 0.001 '** ' 0.01 '*  ' 0.05 '.  ' 0.1 '   ' 1
```

```
res8 <- difference_in_means(vaccine_timing_interestN ~ q115N, data = datw4)
res8
```

```
Design:  Standard
      Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper    DF
q115N -0.07003     0.0389    -1.8  0.07242  -0.1465 0.006398 493.1
```

‘

```
    q115N              q1 age gender_client                educ dependents_dummy_coded            race     core_city weight
84      1  Definitely will  69          Male High school graduate                     0           Black     Core city  5.203
108     0 Definitely won't  55          Male         Some college                     0        Hispanic Not core city  5.722
166     1  Definitely will  58          Male High school graduate                     0           White Not core city  6.984
229     1       Maybe will  30          Male High school graduate                     1        Hispanic Not core city  5.110
257     0  Definitely will  31          Male               2-year                     1 Native American     Core city  7.004
258     1         Not sure  54          Male High school graduate                     0           White     Core city  7.004
418     1  Definitely will  67        Female               4-year                     0        Hispanic     Core city  5.253
470     1       Maybe will  20          Male         Some college                     1           White Not core city  6.090
482     1       Maybe will  59          Male High school graduate                     0           White Not core city  6.397
```

```
    q115N              q1 age gender_client         educ dependents_dummy_coded  race     core_city  weight
99      0 Definitely will  53          Male Some college                     0 White Not core city 0.09571
105     1 Definitely will  81          Male       4-year                     0 White Not core city 0.09304
491     1 Definitely will  46          Male       4-year                     1 White Not core city 0.08359
```