Hypothesis: Republicans and Democrats both experienced the same amount of difficulty voting in the 2020 election; H0 muR = muD (sorry I don’t know how to do latex)

Dataset information: n = 8280, variables beginning with 201 are pre-election, 202 post-election, 200 and 203 don’t look applicable for us

Potential primary independent variable

Potential secondary independent variable

Potential primary dependent variable

Variables of interest:

* Quick disclaimer – the variable list is like 700 pages so did not go through everything, but did a good amount of ctrl+f searching key words. So there could certainly be other variables of interest that I missed
* V201018: PARTY OF REGISTRATION:
  + 3197 R/D
  + 1029 Independent/None
* V201231 Party ID:
  + 8245 valid results (R/D/Ind)
  + So this is the variable I think would be good to use – I think I mixed this one up with party registration when we spoke earlier. This is also a Likert variable (strong D, not strong D, independent D and vice versa with republican plus a single independent option). I think we could make an argument that your self-identification is more applicable to voting behavior than party registration (not sure how hard of an argument that is to make…or if it is actually valid, would need to do some research). But that is an option.
* V201228 DOES R THINK OF SELF AS DEMOCRAT, REPUBLICAN, OR INDEPENDENT:
  + 5428 R/D
  + 2527 Independent
* V201230 NO PARTY IDENTIFICATION - CLOSER TO DEMOCRATIC PARTY OR REPUBLICAN PARTY:
  + 1855 R/D
* V202119: HOW DIFFICULT WAS IT FOR R TO VOTE:
  + Likert variable (not difficult = 1 -> extremely difficult = 5)
  + 6401 valid results
  + I think this is the “dependent” variable we should use. We have a high amount of responses so no issues with sample size
* V202443: WHICH PARTY DOES R FEEL CLOSEST TO:
  + 5810 valid results (R/D)

Approach:

* One potential approach is grouping R/D by defining R/D as what someone self-identifies as. For independents in 201231 we can look at 201230 or maybe 202443 (uncertain here as that is a post-election variable and party id is pre-election).
* Another idea is grouping R/D by party registration. So we could primarily base R/D on V201018 but grouping independents would be interesting – we’d probably have to result to grouping via self-identification (so 201230 or 202443 again)
* My initial thought on testing was running a Wilcoxin rank sum since we are looking at a Likert variable (ordinal) as our dependent variable and our samples (R/D) are not paired. 2 sided test since we are looking for any difference in the means.