

322_Project1

2023-10-21

```
merge_final_v3 <- read_csv("merge_final_v4.csv")
```

```
## New names:
## Rows: 3143 Columns: 16
## -- Column specification
## ----- Delimiter: "," chr
## (2): county, state dbl (14): ...1, pop_county, num_county, pop_state, rep_vote,
## rep_vote_pct, d...
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

```
#stratification based on state
```

```
H = 51 # number of strata = state
N = 3143 # total number of counties in the US
n = 300 # total number of sample size
merge_final_v3$num_sample_county = (merge_final_v3$num_county / N) * 300 # this is n_h
merge_final_v3$num_sample_county = round(merge_final_v3$num_sample_county)
```

```
#create unique identifier for county
```

```
merge_final_v3 <- merge_final_v3 %>%
  group_by(state) %>%
  mutate(state_id = cur_group_id())

merge_final_v3 <- merge_final_v3 %>% group_by(state) %>% mutate(count = sequence(n()))

merge_final_v3$county_id = 1000 * merge_final_v3$state_id + merge_final_v3$count
```

```
#Proportional allocation based on size
```

```
merge_final_v3$weights = merge_final_v3$pop_state / (merge_final_v3$num_sample_county * merge_final_v3$pop_state)
# weights = t_x / (n*x_i)
```

```
set.seed(0)
```

```
merge_final_v3 <- merge_final_v3 %>% filter(num_sample_county != 0) #drop states that have a sample size of 0
```

```
sort_str <- merge_final_v3[order(merge_final_v3$state),] # sort by stratum
```

```
num_sample_county_vector <- c(6,3,1,7,6,6,1,6,15,4,10,9,9,10,11,6,2,2,1,8,8,8,11,5,9,2,1,2,3,6,10,5,8,7)
```

```
sample_data <- sort_str[ppssstrat(sort_str$pop_county, sort_str$state, num_sample_county_vector),]
```

```
svydes = svydesign(id =~ county_id, strata =~ state, weights =~ weights, fpc =~ num_county, data = sample_data)
options(survey.lonely.psu="adjust")
```

Questions

1. What is an estimate of the average population density per county in the U.S. in 2020?

```
svymean(~ pop_county, svydes)
```

```
##              mean      SE
## pop_county 110838 10852
```

2. What is an estimate of the total number of people in the U.S. in 2020 who identify as Hispanic or Latino, any race?

```
svytotal(~ total_hispanic_2020, svydes)
```

```
##              total      SE
## total_hispanic_2020 61067748 5780359
```

3. What is an estimate of the total change in the number of people in the U.S. who identify as Hispanic or Latino, any race, between the 2010 and 2020 censuses?

```
sample_data2 <- sample_data %>% mutate(hispanic_diff = total_hispanic_2020 - total_hispanic_2010)
sample_data2$hispanic_diff[is.na(sample_data2$hispanic_diff)] = 0
svydes2 <- svydesign(id =~ county_id, strata =~ state, weights =~ weights, fpc =~ num_county, data = sample_data2)
svytotal(~ hispanic_diff, svydes2)
```

```
##              total      SE
## hispanic_diff 11471893 870947
```

4. What are estimates of the percentages of people in 2020 in the U.S. who voted Republican? How about Democrat? How about a third party?

```
total_vote_pop <- sum(merge_final_v3$vote_pop) # total number of people who voted
```

```
svytotal(~rep_vote, svydes)
```

```
##              total      SE
## rep_vote 88186122 4130536
```

```
88186122 / total_vote_pop
```

```
## [1] 0.5642405
```

```
svytotal(~dem_vote, svydes)
```

```
##              total      SE
## dem_vote 64853308 3070263
```

```
64853308 / total_vote_pop
```

```
## [1] 0.4149504
```

```
third_vote <- total_vote_pop - 88186122 - 64853308
third_vote / total_vote_pop
```

```
## [1] 0.02080912
```

5. Answer one other question that interests you from the data that you could collect from the website.

We will estimate percentage of population who did not vote in 2020 election.

```
sample_data3 <- sample_data2 %>% mutate(no_vote = pop_county - vote_pop,
                                         no_vote_pct = no_vote / pop_county)
svydes3 <- svydesign(id =~ county_id, strata =~ state, weights =~ weights, fpc =~ num_county, data = sample_data3)
```

```
svytotal(~no_vote, svydes3)

##           total      SE
## no_vote 171531886 7651459

total_pop <- sum(merge_final_v3$pop_county)
171531886 / total_pop

## [1] 0.524215
```