Question Appendix.

Unless noted in parentheses, all questions are from GSS

Section A: Question Wordings

Illegal version

1. What is your gender? A) male, B) female?
2. Are you of Hispanic, Latino, or Spanish origin? A) No, not of Hispanic, Latino or Spanish origin. B) Yes, Mexican, Mexican American, Chicano; B) Yes, Puerto Rican, C) Yes, Cuban; D) Yes, other;
3. What is your race? A) white, B) black/African American, C) American Indian, D) Asian Indian, E) Japanese, F) Native Hawaiian, G) Chinese, H) Korean, Guanmanian or Chamomo, I) Filipino, J) Vietnamese, K) Samoan, L) Other Asian (fill in blank: ),
4. What is your date of birth? Give month and year only.
5. What is your annual income (as you would report it in the Census)? Fill in blank.
6. What is the highest degree or level of school you have completed?
7. No schooling completed,
8. Nursery or preschool through grade 12: Fill in blank.
9. Regular high school diploma
10. GED or alternative credential,
11. Some college credit, but less than 1 year of college credit,
12. 1 or more years of college credit, no degree,
13. Associate’s degree (for example: AA, AS)
14. Bachelor’s degree (for example: BA, BS)
15. Master’s Degree (for example MA, MS, Meng, Med, MSW, MBA),
16. Professional degree beyond a bachelor’s degree (for example MD, DDS, DVM, LLB, JD)
17. Doctorate degree (for example PhD, EdD)
18. In which US State do you live? Fill in blank:
19. Were you born in the US? A) Yes, B) No.
20. With which party do you identify with?
21. Democrat, B) Republican, C) Independent
22. This survey is halfway over, so if you are really paying attention, click strongly agree.  
    A) strongly disagree, B) disagree, C) agree, D) strongly agree
23. Do you think the number of illegal immigrants to America nowadays should be... a) increased a lot, b) increased a little, c) remain the same as it is, d) reduced a little, e) reduced a lot, f) don’t know
24. In the long run, do you think that illegal immigrants who are immigrating to the United States today will make American society better, will make American society worse, or do you think that today's illegal immigrants won't affect American society one way or another?

Answers: a) better, b) worse, c) won’t affect American society one way or another, d) don’t know

Questions 20-25 start with the following opener:

|  |
| --- |
| Now I'm going to read you some statements and would like to get your reaction to them. After I read each statement, please tell me if you strongly agree, |
| agree, neither agree nor disagree, disagree, or strongly disagree with the statement. How strongly would you agree with the following statements on a scale of 1 to 5 (where 1 is highly disagree and 5 is highly agree):   1. “Illegal immigrants are getting too demanding in their push for equal rights.” |

1. “The Irish, Italians, Jews, and many other minorities overcame prejudice and worked their way up. Today's illegal immigrants should do the same without any special favors.”
2. “English will be threatened if other languages are frequently used in large illegal immigrant communities in the U.S.”

For 26-30: (the sequence of the following options will be randomized for each participant who will respond on a 1-5 scale (ranging from very unlikely (1) to very likely (5) ). Each question will be prefaced with the following:

|  |
| --- |
| “What do you think will happen as a result of more illegal immigrants coming to this country?  Is the following result very likely, somewhat likely, |
| not too likely, or not at all likely?” where very likely is 5 and not at all likely is 1. |

|  |
| --- |
| 1. Making the country more open to new ideas and cultures 2. Higher crime rates 3. People born in the US losing their jobs, 4. Higher economic growth 5. Making it harder to keep the country united. |
|  |

1. Vignette: “Recently, there have been several investigative reports regarding the condition of illegal immigrants being held in detention centers around the United States. The reports indicate that some detainees are held for long periods of time under inhumane conditions, lacking sufficient nutrition, proper sanitary facilities, access to legal counseling, or even contact with family members. The reports also discuss cases of medical negligence concerning the treatment of detainees with serious health problems. In one case, a detainee told authorities he was suffering from severe chest pain, profound shortness of breath, and dizziness, and asked for immediate medical attention. However, these situations require supervised transportation to a hospital several miles away from the detention center. In response, detention center officials denied the detainee this medical service because they considered him to be a flight risk. The officials said they had a reasonable cause to deny medical attention outside the facility. The detainee, on the other hand, said that the denial of medical attention was unwarranted and that his health was continuing to deteriorate.”

Do you think the government should transport this illegal migrant (SPLIT) to the hospital, at the risk of enabling him to escape?

A) Yes,

B) No.

1. Which U.S. Political Party do you most Identify with?  
   A) Democrat, B) Republican, C) Independent
2. Would you say that you tend to consume what most people would call “conservative” media sources or “liberal” media sources?

Seven point scale from very liberal (1) to very conservative (5)

1. How much do you identify as conservative, moderate, or liberal?

extremely liberal (1)

liberal (2)

slightly liberal (3)

moderate (4)

slightly conservative (5)

conservative (6)

extremely conservative (7)

1. Do you agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, or disagree strongly with the following statement? “It is better to live in an orderly in which the laws are vigorously enforced than to give people too much freedom.” (from Sniderman, When Ways of Life Collide)
2. Strongly Disagree,
3. Disagree
4. Agree
5. Strongly Agree

I’d like to get your feelings toward some groups that have been in the news these days. I’d like you to rate each group using something we call the feeling thermometer. Ratings between 5 degrees and 10 degrees mean that you feel favorable and warm toward that group. Ratings between 0 and 5 degrees mean that you don’t feel favorable toward that group and that you don’t care too much for that group. You would rate the person at 5 degree mark if you don’t feel particularly warm or cold toward that group. Please enter a number between 0 and 10 for:

1. "white people"
2. "Latino people"
3. Is there anything else you would like to tell us about this survey?

Section B: Justification of Variables Based on the Literature

I include the following variables because they are on the Census 2010, whereas the 2020 Census content still being in dispute: 1) gender, 2) Hispanic/Latino/ Spanish Origin, 3) race, 4) date of birth, 5) annual income, 6) highest degree of education, and 7) US state of residence. Some of these questions are also relevant to the current study as around 50 years of prior studies suggest that views toward immigration might be significantly associated with ethnicity, race, age, annual income, education, and state of residence and therefore one should control for these variables (Kinder and Kam 2010; Burnham 1974). Although some of them the structure of the questions (only providing “male” or “female” for gender) may be a bit archaic and confusing, I maintain the exact structure of the question so I can later reweight my responses with census data to make my sample more representative. In addition, years of education is a relevant control variable because Sniderman et al. (2004) finds that it is significantly related to stereotypes and social distance toward outgroups.

From the General Social Survey I borrow the questions about 1) whether a person was born in the U.S., 2) all the questions about immigration, 3) party identification, because literature suggests that native-born/Republicans/liberals are respectively more opposed to immigration than foreign-born/Democrats/conservatives (Kinder and Kam 2010), and even non-prominent individuals may be constrained by needing to be consistent with their past opinions to avoid cognitive dissonance —even if perhaps less so than politicians or public figures and online (Helbling 2014). The question about the ideological valence of a participant’s media diet I designed because of evidence from the a National Hispanic Media Coalition study’s finding that young people were more likely to have xenophobic sentiments than those that after consuming more “conservative” news (e.g. Fox News) than those that consumed more “liberal” news (e.g. MSNBC). Finally, the question about whether the participant agrees that ““It is better to live in an orderly in which the laws are vigorously enforced than to give people too much freedom” is a conventional measure of Theodore Adorno-conceptualized “authoritarian personality.” I use it because Sniderman et al. ( 2004) finds this is significantly associated with both stereotypes and prejudice toward outgroups. I use the feeling thermometers for whites and Hispanics to construct a measure of ethnocentrism because these Kinder and Kam (2010) suggest this is strongly associated with values toward immigration and Sniderman et al. (2004) finds so as well.

Section C.

Below is the R code script for my reweighted results based on three different poststratification procedures: 1) post stratification by simple means, 2) means with post stratification, 3) Means with post-stratification (160 groups) and missing group imputation, 4) Model-based estimation with post-stratification, 5) Approach 5: Multilevel-Model-based estimation with post-stratification (MRP).

No matter which approach I took, my coefficients were not that substantively or significantly different from what I obtained without any weights.I can supply more detailed results upon request though I exclude them

Table of Contents for Code:

1) Approach 1: Simple means

* 1. Calculate means
  2. Plot estimated means against benchmarks
  3. Plot distribution of estimation-benchmark differences2

2) Approach 2: Means with post-stratification (8 groups)

2.1) Calculate group means, group weights, and weighted means

2.2) Plot estimated means (which as the sum of the means for each stratified group) against benchmarks by first left joining this data set to the pew data set qid

2.3) Plot distribution of estimation-benchmark difference

3) Approach 3: Means with post-stratification (160 groups) and missing group imputation 3.1) Calculate group means, group weights, and weighted means

3.1.1) Dealing with missing groups: imputing with sample means

3.2) Plot estimated means against benchmarks

3.3) Plot distribution of estimation-benchmark differences

4) Approach 4: Model-based estimation with post-stratification

4.1) Predict group means with simple regression model; combine with group weights to create weighted means

4.2) Plot estimated means against benchmarks

4.3) Plot distribution of estimation-benchmark differences

5) Approach 5: Multilevel-Model-based estimation with post-stratification (MRP)

5.1) Predict group means with multi-level regression model; combine with group weights to create weighted means

5.2) Plot estimated means against benchmarks

5.3) Plot distribution of estimation-benchmark differences

5.4) Compare differences from MRP with other methods

R-Script

title: "new\_file"

output: html\_document

---

```{r setup, include=FALSE}

knitr::opts\_chunk$set(echo = TRUE)

```

#if you are having problems with some code and cannot figure out what it is, then copy and paste your code into a new R file line by line andd run it, searching fo the errors.

# Load packages and data

```{r, message = F, warning = F, echo = T }

# load packages

library(tidyverse)

library(lme4)

# set your working directory

# setwd("~/user/working\_directory")

# load cleaned data file for survey results

ds <- read.csv("https://github.com/compsocialscience/summer-institute/raw/master/2019/materials/day4-surveys/activity/2019-06-13\_mturk\_data\_clean.csv")

## not using education or political attention check, so drop these vars

## (you can use these if you want!)

#data <- data %>% select(-attention1, -educ)

# load external information -- in this case, population info

census <- read.csv("https://github.com/compsocialscience/summer-institute/raw/master/2019/materials/day4-surveys/activity/2017\_acs\_data\_clean.csv")

# load pew benchmarks

pew <- read.csv("https://github.com/compsocialscience/summer-institute/raw/master/2019/materials/day4-surveys/activity/2019\_pew\_benchmark\_data.csv",

col.names = c("qid", "label", "pew\_estimate", "source"))

pew <- pew %>% select(qid, pew\_estimate)

class(pew$qid)

class(pew$pew\_estimate)

#everything is a factor except the nuemric colummn

```

\newpage

# Approach 1: Simple means

First, we'll just take the mean of the whole sample for each question. This approach doesn't use any post-stratification.

## 1.1) Calculate means

```{r}

# take the mean of survey responses in mturk data

## remove demographic variables (factor vars)

## get column means

mturk\_means <- ds %>% select(-sex, -race, -age\_cat, -region, -educ, -attention1) %>%

summarise\_all(~mean(., na.rm = T))

# reshape from wide to long

## with columns for questions (call this qid) and for mean

mturk\_means <- mturk\_means %>% gather(qid, mean)

# preview

head(mturk\_means)

#this gives the mean of each variable in the whole data set ds

```

## 1.2) Plot estimated means against benchmarks

\*\*Tip\*\*: You will be making this type of plot each time you generate a new set of estimates, so it would be helpful to write a function for this.

```{r}

# merge mturk mean estimates with pew benchmark by quetion ID ("qid")

mean\_est <- inner\_join(pew, mturk\_means, by = c("qid"))

head(mean\_est)

# make function for plot

plot\_comparison <- function(est\_table, method, caption){

graph <- ggplot(est\_table,

aes(x = pew\_estimate, y = method)) +

geom\_point() +

labs(x = "Estimates from Pew", y = caption) +

scale\_x\_continuous(limits = c(0,1)) +

scale\_y\_continuous(limits = c(0,1)) +

geom\_abline(intercept = 0, slope = 1, linetype = "dotted") +

coord\_fixed()

return(graph)

}

# plotthe estimates of means(mean\_est) by assigning it to est\_table, plot the means of the mean\_est data frame as y against the pew\_estimates as x by assigning mean\_est$mean to method, and assing "Non-weighted estimates from MTurk" to caption

plot\_comparison(est\_table = mean\_est,

method = mean\_est$mean,

caption = "Non-weighted estimates from MTurk")

#why is this not plotting anything?)

```

## 1.3) Plot distribution of estimation-benchmark differences

\*\*Tip\*\*: You will also be making this type of plot each time you generate a new set of estimates, so it would be helpful to write a function for this as well.

```{r}

# calculate difference by assigning a diff column to mean\_est which equals the absolute difference of the mean\_est$mean minus the pew\_estimate column of mean\_est

mean\_est$diff <- abs(mean\_est$mean - mean\_est$pew\_estimate)

# function for plotting difference

plot\_diff <- function(est\_table){

diff\_graph <- ggplot(est\_table, aes(x = diff)) +

geom\_histogram(aes(y = (..count..)/sum(..count..)), binwidth = .025,

colour = "black", fill = "white") +

theme\_bw() +

geom\_vline(aes(xintercept = median(diff)), linetype = "longdash") + #draw a vertical line at the x-intercept that is median of the difference and make the line a long dash

labs(x = "absolute difference", y = "density") +

scale\_y\_continuous(limits = c(0, 0.45)) #make the y-axis scale run from 0 to 0.45

return(diff\_graph)

}

# plot

plot\_diff(mean\_est)

```

\newpage

# Approach 2: Means with post-stratification (8 groups)

## 2.1) Calculate group means, group weights, and weighted means

To start, group by sex and region only. This should give you 8 groups (2 sexes by 4 regions).

Group weights can be calculated as $\frac{N\_{h}}{N}$. They should sum to 1. You will need to calculate these group weights for the other approaches as well.

```{r}

# get total census population, create a new vector that is the total number of people in the US census

N <- sum(census$POP)

N

# calculate group weights

## group population data by sex and region,

## get the sum for each cell and divide by total pop

#NOTE: You need to detach the plyr package or it will transform population\_counts into 1 by 1 dataframe.

library(plyr)

detach(package:plyr)

#but you have to load it again later when you do the match\_df command and then detach it.

population\_counts <- census %>%

group\_by(sex, region) %>%

summarise(group\_weight = sum(POP)/N)

#why is this deleting everything but one obs and 1 var?)

#this creates a column called the group\_weight that is the number of people in each sex-region group divided by the total population

# check that weights sum to one

if (sum(population\_counts$group\_weight) != 1) {

print("weights don't sum to one")

}

#this will say weight don't sum to 1 if when you add all the population weights they do not add to 1.

head(population\_counts)

#here you calculate the average response for each dependent varaible within each group

# calculate group means for each question response

## group data by sex and region

## remove non-numeric variables (demographic vars)

## calculate group means for each column

sample\_counts <- ds %>%

group\_by(sex, region) %>%

select\_if(is.numeric) %>%

summarise\_all(list(~mean(.,na.rm = T)))

# preview -- scroll for more columns

head(sample\_counts)

```

```{r}

# check that there are no empty cells

if (nrow(sample\_counts) < nrow(population\_counts)) {

print("GROUPS MISSING:")

print(nrow(population\_counts) - nrow(sample\_counts))

}

# merge population counts with sample counts

# left join and retain all groups in population

cell\_based <- left\_join(population\_counts,

sample\_counts,

by = c("sex", "region"))

# reshape wide to long

cell\_based\_long <- cell\_based %>% gather(qid, mean,

-c(sex, region, group\_weight),

na.rm = F)

head(cell\_based\_long)

# with mutate create a new column in cell\_based\_long called weighted mean which is the group means times group weights in the cell\_based\_long dataframe

cell\_based\_long <- mutate(cell\_based\_long, weighted\_mean = group\_weight\*mean)

head(cell\_based\_long)

# sum weighted means, grouping by question

#this creates a mturk\_cell\_est data frame which is the cell\_based\_long data but with a single column that lists the sum of each group's weighted means for each question ID (removing missing cases)

mturk\_cell\_est <- cell\_based\_long %>%

group\_by(qid) %>%

summarise(mturk\_cell\_estimate = sum(weighted\_mean, na.rm = T))

mturk\_cell\_est

head(mturk\_cell\_est)

```

## 2.2) Plot estimated means (which as the sum of the means for each stratified group) against benchmarks by first left joining this data set to the pew data set qid

```{r}

# merge mturk cell-based weighted estimates with benchmark

simple\_cell\_est <- inner\_join(pew, mturk\_cell\_est, by = c("qid"))

head(simple\_cell\_est)

# plot (you can use the function we created above, except now the est\_table is the simple\_cell\_est object you just created, the method (thing on the y axis is the mturk\_cell\_estimate within the simple\_cell\_est data frame, and the caption is weighted estimates from MTurk))

plot\_comparison(est\_table = simple\_cell\_est,

method = simple\_cell\_est$mturk\_cell\_estimate,

caption = "weighted estimates from MTurk")

```

## 2.3) Plot distribution of estimation-benchmark differences

```{r}

#calculate difference

#here we do the same thing as before, except now we are creating a column diff which is the absolute difference between the mturk\_cell\_estimate in the simple\_cell\_est data frame and the pew\_estimate of the simple\_cell\_est

simple\_cell\_est$diff <- abs(simple\_cell\_est$mturk\_cell\_estimate - simple\_cell\_est$pew\_estimate)

#plot

plot\_diff(simple\_cell\_est)

```

\newpage

# Approach 3: Means with post-stratification (160 groups) and missing group imputation

## 3.1) Calculate group means, group weights, and weighted means

Can you get better estimates grouping by more variables? Try grouping on sex, region, age group, and race.

You will now have 160 groups (2 x 4 x 5 x 4). Some of groups may be missing from your sample (e.g. 50-64 year old black women in the midwest). If a group is missing, their answers will automatically be treated as "zero" when computing weighted means. As a result, some question responses may be underestimated. One way to deal with this is to impute the missing values with the sample average for that variable (aka the simple means we calculated in the first step). You will do this in the next step.

First, calculate the new group means, group weights, and weighted means as you did above in Approach 2.

```{r}

# get total population

N <- sum(census$POP)

# calculate group means, group weights, and weighted means

#I will call this population\_counts\_160 to indicate that these are population counts in the census within each of 160 groups. We calculate a group\_weight which is the proportion of the total population that is in that group.

population\_counts\_160 <- census %>%

group\_by(sex, region, age\_cat, race) %>%

summarise(group\_weight = sum(POP)/N)

# check that weights sum to one

if (sum(population\_counts\_160$group\_weight) != 1) {

print("weights don't sum to one")

}

head(population\_counts\_160)

tail(population\_counts\_160)

#calculate group means for each question response

## remove non-numeric variables (demographic vars)

## calculate group means for each column

sample\_counts\_160 <- ds %>%

group\_by(sex, region, age\_cat, race) %>%

select\_if(is.numeric) %>%

summarise\_all(list(~mean(.,na.rm = T)))

# preview -- scroll for more columns

head(sample\_counts\_160)

```

# check that there are no empty cells

if (nrow(sample\_counts\_160) < nrow(population\_counts\_160)) {

print("GROUPS MISSING:")

print(nrow(population\_counts\_160) - nrow(sample\_counts\_160))

}

#this tells us that 76 groups are missing. Now we must find out which groups those are.

```

```{r}

#calculate the means of the non imputed groups

#calculate the weighted means:

sample\_counts\_160\_means <- sample\_counts\_160 %>%

group\_by(sex, region, race, age\_cat) %>%

select\_if(is.numeric) %>%#this gets all the variables that are the means

summarise\_all(list(~mean(.,na.rm = T)))#give the means of all sample\_counts as a list removing the NAs

head(sample\_counts\_160\_means)

# preview -- scroll for more columns

head(sample\_counts\_160\_means)

# merge population counts with sample counts of means for 160 stratified groups

# left join and retain all groups in population

cell\_based\_160 <- left\_join(population\_counts\_160,

sample\_counts\_160\_means,

by = c("sex", "region", "race", "age\_cat"))

#I am noticin that PartyLn still has some NaNs. Is that a problem? No, they just did not answer

#now we calculate weighted means by putting hte data in long form:

#put race back as a factor

cell\_based\_160$race<-as.factor(cell\_based\_160$race)

# reshape wide to long

cell\_based\_long <- cell\_based\_160 %>% gather(qid, mean,

-c(sex, age\_cat, race, region, group\_weight),

na.rm = F)

#you need to remember to remove all varaibles

head(cell\_based\_long)

# with mutate create a new column in cell\_based\_long called weighted mean which is the group means times group weights in the cell\_based\_long dataframe

weighted\_means\_long <- mutate(cell\_based\_long, weighted\_mean = group\_weight\*mean)

head(weighted\_means\_long)

#we need to convert qid to a factor because we need to group by something and to group by something you need to have that thing be a factor

weighted\_means\_long$qid<-as.factor(weighted\_means\_long$qid)

weighted\_means\_long\_by\_group\_nonimputed<-weighted\_means\_long%>%

group\_by(qid)%>%

summarise(weighted\_means\_by\_group=sum(weighted\_mean, na.rm=T))

head(weighted\_means\_long\_by\_group\_nonimputed)

#I did an inner\_join as it is in the original code. This takes everything that is in pew and #weighted\_means\_long\_by\_group\_nonimputed and excludes everything else

weighted\_means\_long\_by\_group\_nonimputed<-inner\_join(weighted\_means\_long\_by\_group\_nonimputed, pew, by=c("qid"))

weighted\_means\_long\_by\_group\_nonimputed

weighted\_means\_long\_by\_group\_nonimputed$qid<-as.factor(weighted\_means\_long\_by\_group\_nonimputed$qid)

names(weighted\_means\_long\_by\_group\_nonimputed)

```

### 3.1.1) Dealing with missing groups: imputing with sample means

Now, replace the missing groups with the sample means you computed in 1.1.

```{r}

#here we will anti-join on the common groups

# replace missing group means with sample means

## Find the missing rows (groups)

library(plyr) #Load plyr here so it doesn't mess with dplyr earlier

##Missing rows should be the non-matched rows

missing\_rows <- anti\_join(population\_counts\_160, match\_df(population\_counts\_160, sample\_counts\_160, on = c("sex", "race", "region", "age\_cat")), by = c("sex", "race", "region", "age\_cat"))

head(missing\_rows)

#notice that unsurprisingly many of the empty rows are elderly asian, black and other ethnic minority women in the midwest

detach(package:plyr) #Avoid later problems between plyr and dplyr (which often mask each other)

#now we make a new dataframe called missing\_groups\_test which is identical to missing\_rows.

#except then we run an anonymous for loop that goes through each nrow in mturk\_means and assigns a value of NA to the mturk\_means column of missing\_groups\_test, and then assigns the vlaue (which is the overall mean for a question within each stratified group ) that is in the second column of mturk\_means?) Draw a diagram

missing\_groups\_test = missing\_rows

for (i in 1:nrow(mturk\_means)) {

missing\_groups\_test[ , mturk\_means[i, 1]] = NA

missing\_groups\_test[, mturk\_means[i, 1]] = mturk\_means[i, 2]

}

#you need to use brackets to build for loops, cannot use $, so this is why it is hard to create new variables in a loop.

#crate new dataframe called missing rows that consists of all rows that are not in both

#population\_counts\_160 and the rows that are in both in sample\_counts\_160 and population\_counts\_160

#as matched by "sex", "race", "region", "age\_cat"?) help me draw this.

#notice how each group has the same mean for each question

#you need to use brackets to build for loops, cannot use $, so this is why it is hard to create new variables in a loop.

#Add missing groups (that are stored in the missing\_groups\_test and contain the

#overall mean answer for a given question of the whole population) to the sample\_counts\_160 dataframe

sample\_counts\_160<-rbind(sample\_counts\_160, missing\_groups\_test)

head(sample\_counts\_160)

#I am getting NAs for group\_weight--is that a problem?)

ds$age\_cat <- as.factor(ds$age\_cat) #Put age\_cat back into factor form (why?)

#an alternative way to do the above is to just create a list of the missing cases with is.na

#then do an inner join on the population data to get the missing cases

<!-- missing\_groups <- cell\_based\_long %>% filter(is.na(mean)) -->

<!-- # merge sample means vector created in 1.1 (mturk\_means) with this new dataframe -->

<!-- missing\_groups\_imputed <- inner\_join(missing\_groups, mturk\_means, by = c("qid")) %>% -->

<!-- select(-mean.x, -weighted\_mean) %>% -->

<!-- rename(mean = mean.y) -->

<!-- # now merge back with all non-missing groups (stored in cell\_based\_long) -->

<!-- cell\_based\_long\_imputed <- right\_join(missing\_groups\_imputed, cell\_based\_long, -->

<!-- by = c("sex", "age\_cat", "region", "race", -->

<!-- "group\_weight" , "qid")) %>% -->

<!-- mutate(mean = ifelse(is.na(mean.x), mean.y, mean.x)) %>% -->

<!-- select(-mean.x, -mean.y, -weighted\_mean) %>% -->

<!-- # and recalculate weighted means -->

<!-- mutate(weighted\_mean\_imputed = group\_weight\*mean) -->

#starting our from the first menthod:

#calculate the weighted means:

sample\_counts\_160\_means <- sample\_counts\_160 %>%

group\_by(sex, region, race, age\_cat) %>%

select(-group\_weight) %>% #Don't take a mean of the group weight, but of the questions

select\_if(is.numeric) %>%#this gets all the variables that are the means

summarise\_all(list(~mean(.,na.rm = T)))#give the means of all sample\_counts as a list removing the NAs

head(sample\_counts\_160\_means)

# preview -- scroll for more columns

head(sample\_counts\_160\_means)

# merge population counts with sample counts of means for 160 stratified groups

# left join and retain all groups in population

cell\_based\_160 <- left\_join(population\_counts\_160,

sample\_counts\_160\_means,

by = c("sex", "region", "race", "age\_cat"))

View(cell\_based\_160)

#I am noticin that PartyLn still has some NaNs. Is that a problem? No, they just did not answer

#now we calculate weighted means by putting hte data in long form:

# reshape wide to long

weighted\_means\_long <- cell\_based\_160 %>% gather(qid, mean,

-c(sex, age\_cat, race, region, group\_weight),

na.rm = F)

#you need to remember to remove all varaibles

head(weighted\_means\_long)

# with mutate create a new column in cell\_based\_long called weighted mean which is the group means times group weights in the cell\_based\_long dataframe

weighted\_means\_long <- mutate(weighted\_means\_long, weighted\_mean = group\_weight\*mean)

head(weighted\_means\_long)

#we need to convert qid to a factor because we need to group by something and to group by something you need to have that thing be a factor

weighted\_means\_long$qid<-as.factor(weighted\_means\_long$qid)

weighted\_means\_long\_by\_group<-weighted\_means\_long%>%

group\_by(qid)%>%

summarise(weighted\_means\_by\_group=sum(weighted\_mean, na.rm=T))

head(weighted\_means\_long\_by\_group)

```

## 3.2) Plot estimated means against benchmarks

```{r}

#Plot both your new group means and the estimated means against the Pew benchmarks.

################################## WITH NO IMPUTATION ###################################

#first join the pew estimates to

weighted\_means\_long\_by\_group<-weighted\_means\_long%>%

group\_by(qid)%>%

summarise(weighted\_means\_by\_group=sum(weighted\_mean, na.rm=T))

head(weighted\_means\_long\_by\_group)

weighted\_means\_long\_by\_group<-inner\_join(pew, weighted\_means\_long\_by\_group, by=c("qid"))

#this coerces qid to be a character vector. It is best to immediately change it to a factor varaible otherwise it may present problems later when you group by again.

weighted\_means\_long\_by\_group

```

weighted\_means\_long\_by\_group$qid<-as.factor(weighted\_means\_long\_by\_group$qid)

#reminder of code for plot\_comparison function

plot\_comparison <- function(est\_table, method, caption){

graph <- ggplot(est\_table,

aes(x = pew\_estimate, y = method)) +

geom\_point() +

labs(x = "Estimates from Pew", y = caption) +

scale\_x\_continuous(limits = c(0,1)) +

scale\_y\_continuous(limits = c(0,1)) +

geom\_abline(intercept = 0, slope = 1, linetype = "dotted") +

coord\_fixed()

return(graph)

}

#when you plot the method, you need to specify which column you want to plot.

plot\_comparison\_nonimputed<-plot\_comparison(est\_table=weighted\_means\_long\_by\_group\_nonimputed, method=weighted\_means\_long\_by\_group\_nonimputed$weighted\_means\_by\_group, caption= "Weighted Means With 160 Group and Imputation")

plot\_comparison\_nonimputed

```

################################## WITH IMPUTATION ######################################

#first join the pew estimates to

weighted\_means\_long\_by\_group<-left\_join(pew, weighted\_means\_long\_by\_group, by=c("qid"))

#this coerces qid to be a character vector. It is best to immediately change it to a factor varaible otherwise it may present problems later when you group by again.

weighted\_means\_long\_by\_group$qid<-as.factor(weighted\_means\_long\_by\_group$qid)

#reminder of code for plot\_comparison function

plot\_comparison <- function(est\_table, method, caption){

graph <- ggplot(est\_table,

aes(x = pew\_estimate, y = method)) +

geom\_point() +

labs(x = "Estimates from Pew", y = caption) +

scale\_x\_continuous(limits = c(0,1)) +

scale\_y\_continuous(limits = c(0,1)) +

geom\_abline(intercept = 0, slope = 1, linetype = "dotted") +

coord\_fixed()

return(graph)

}

#when you plot the method, you need to specify which column you want to plot.

plot\_comparison\_imputation<-plot\_comparison(est\_table=weighted\_means\_long\_by\_group, method=weighted\_means\_long\_by\_group$weighted\_means\_by\_group, caption= "Weighted Means With 160 Group and Imputation")

plot\_comparison\_imputation

## 3.3) Plot distribution of estimation-benchmark differences

```{r}

#################################### WITH NO IMPUTATION #################################

plot\_diff <- function(est\_table){

diff\_graph <- ggplot(est\_table, aes(x = diff)) +

geom\_histogram(aes(y = (..count..)/sum(..count..)), binwidth = .025,

colour = "black", fill = "white") +

theme\_bw() +

geom\_vline(aes(xintercept = median(diff)), linetype = "longdash") + #draw a vertical line at the x-intercept that is median of the difference and make the line a long dash

labs(x = "absolute difference", y = "density") +

scale\_y\_continuous(limits = c(0, 0.45)) #make the y-axis scale run from 0 to 0.45

return(diff\_graph)

}

#first add the diff column to weighted\_means\_long\_by\_group\_nonimputed

weighted\_means\_long\_by\_group\_nonimputed$diff<-abs(weighted\_means\_long\_by\_group\_nonimputed$weighted\_means\_by\_group-weighted\_means\_long\_by\_group\_nonimputed$pew\_estimate)

#first add the diff column to weighted\_means\_long\_by\_group\_imputed

weighted\_means\_long\_by\_group$diff<-abs(weighted\_means\_long\_by\_group$weighted\_means\_by\_group-weighted\_means\_long\_by\_group$pew\_estimate)

# plot

#non\_imputed

plot\_diff(weighted\_means\_long\_by\_group\_nonimputed)

#################################### IMPUTATION #######################################

#imputed

plot\_diff(weighted\_means\_long\_by\_group)

```

\newpage

# Approach 4: Model-based estimation with post-stratification

## 4.1) Predict group means with simple regression model; combine with group weights to create weighted means

```{r}

#make a copy of everything because we want to put te means in for each group

ds\_copy<-ds

library(dplyr)

#you need to take out group\_weight or it will another group\_weight and get confused

sample\_counts\_160\_copy<-sample\_counts\_160%>%select(-group\_weight)

# for this, we will need convert everything into factors

ds\_copy <- ds\_copy %>% mutate\_all(funs(as.factor))

# Now we will regress each survey answer on demographic characteristics and

# use those model parameters to generate predicted probabilities for each group

# loop through each survey answer and store each vector of pred.probs

# in a 160 x 44 matrix

# but first, write a warning function for later to make sure

# that all estimates are 0 to 1 inclusive

prob\_range\_warning <- function(predictions){

if (any(predictions < 0)) {

warning("some predictions less than zero")

}

if (any(predictions > 1)) {

warning("some predictions more than one")

}

}

#below we build our models based on each response

#we need to take our data, and our variables, for each question, they create a logistic regression model, adn they still use the group based weights, but the mean is coming from the model, you run each group for the model. We take whole data set, create a model,

#P(Y) means the probability that they give answer 1 to a given option in a multiple choice question. #P(Y=1)=logit (beta\_0+beta\_male\*male...beta\_30-40\*(30-40)+beta\_Asian\*(Asian)

#the model is a logistic regression model that predicts whether they will answer 1 to a given question based on whether they belong to a given demographic group

#they run a model for each question

#they use the coefficients from the model to calculate the mean

#eventually we want to write a function with a for loop, but for now we are just going to write one model to test it

model1<-glm(ds\_copy$MILITARY.1~ds\_copy$sex + ds\_copy$race +ds\_copy$region+ds\_copy$age\_cat, family="binomial", data=ds\_copy)

summary(model1)

names(ds\_copy)

#write a nested loop where you run teh covariates ds\_copy$sex + ds\_copy$race +ds\_copy$region+ds\_copy$age\_cat on each column from 1 to 44 in data frame ds\_copy, and within each iteration, calculate the mean for that column.

#how to start a lesson on for loops

for (i in 1:ncol(ds\_copy[,1:44])) {

print(i)

}

#this for loop is making a prediction for every observation in all the data, but what we want is to run separate models on each of 160 stratified groups

#note when you run a regression, you should not specify taht your columns(variables) are coming from a particula data set in your regression with the $--if you do that, R will look for the column within your data et withoin your data set and do too much.

for (i in 1:ncol(ds\_copy[,1:44])) {

print(i)

model<-glm(ds\_copy[, i]~sex + race +region+age\_cat, family="binomial", data=ds\_copy)

print(model$coefficients)

for (j in 1:nrow(sample\_counts\_160\_copy)) {

prediction<-predict(model, sample\_counts\_160\_copy[j, 1:4], type="response")#we write 1:4 here becuase there are only four pre-treatment covariates)

print.data.frame(sample\_counts\_160\_copy[j, 1:4]) #so we know what group is being printed , we type print.data.frame

print(prediction)

prediction<-sample\_counts\_160\_copy[j, i+4]#I type i+4 because in sample copy they are in column 5 to 5 to 49 so we need to shift over 4 columns

}

}

#predictions will give you the means and then you reweight it with the weights.

cell\_based\_model\_stratification <- left\_join(population\_counts\_160,

sample\_counts\_160\_copy,

by = c("sex", "region", "age\_cat", "race"))

names(cell\_based\_model\_stratification)

# reshape wide to long

cell\_based\_model\_stratification\_long <- cell\_based\_model\_stratification %>% gather(qid, mean,

-c(sex, region, age\_cat, race, group\_weight),

na.rm = F)

head(cell\_based\_model\_stratification\_long)

# with mutate create a new column in cell\_based\_long called weighted mean which is the group means times group weights in the cell\_based\_long dataframe

cell\_based\_model\_stratification\_long <- mutate(cell\_based\_model\_stratification\_long, weighted\_mean = group\_weight\*mean)

head(cell\_based\_model\_stratification\_long)

```

```{r}

#here is a slightly more efficient way to do this:

#first we are going to create a new data frame called data\_factor which is the ds\_copy data frame with all variables mutated into factors with the functions as.factor

#funs() provides a flexible way to generate a named list of functions for input to other functions like summarise\_at(). funs() is like apply series of vunctions, you could apply multiple functions

data\_factor <- ds\_copy %>% mutate\_all(funs(as.factor))

# create a character vector of the 44 question names

# these question names can be found in the column names of the data

relevant\_questions <- colnames(ds)[!colnames(ds\_copy) %in% c("sex", "age\_cat", "region", "race")]

# create container dataframe called model\_predictions that is a matrix with the rows of

#populations\_counts and columns with relevant\_questions: Here the columns are options to the #questions and the rows are the groups (does not tell you which groups they are)

#you need to make sure that you your population\_counts\_160 dataframe to have thesame categorys by depulicating with the name population\_counts. Remember that this is going to write out our just age and region population\_counts

population\_counts<-population\_counts\_160

model\_predictions <- as.data.frame(matrix(nrow = nrow(population\_counts),

ncol = length(relevant\_questions), NA))

colnames(model\_predictions) <- relevant\_questions

#you could put in row names but you would have to

#1)put the factorized values of the demographic variables into characters

#2) permute each value of all the varaibles with each other and form a single string with all 166 the possible values from the four variables,

#3) assign that to population\_count

#3)

#rownames(model\_predictions)<-population\_counts

#actually relevant\_questions is a string with each option for all questions

#they did everything not ina nested loop but all at once

# loop through

for (i in relevant\_questions) {

# get outcome (option to a question), put into outcome the ith column

outcome <- data\_factor[ , i]

# fit model

model <- glm(outcome ~ sex + age\_cat + region + race,

data = data\_factor,

family = binomial(link = "logit"))

# create predicted probabilities

reg\_predicted\_values <- predict(model, newdata = population\_counts, type = "response")

# check for errors

prob\_range\_warning(reg\_predicted\_values)

# store in container

model\_predictions[ , i] <- reg\_predicted\_values

}

# bind demographic categories to predictions--binding the columns of the population\_counts and then model\_predictions

model\_wide <- bind\_cols(population\_counts, model\_predictions)

head(model\_wide)

```

```{r}

# reshape wide to long, gather on qid and predicted value but not everything inside -c()

model\_long <- model\_wide %>% gather(qid, predicted\_value,

-c(sex, age\_cat, region, race, group\_weight),

na.rm = F)

head(model\_long)

# weight predictions and sum by qid

model\_est <- model\_long %>%

mutate(weighted\_prediction = group\_weight\*predicted\_value) %>%

group\_by(qid) %>%

summarise(model\_prediction = sum(weighted\_prediction, na.rm = T))

head(model\_est)

# merge with pew benchmarks

pew\_model\_est <- inner\_join(pew, model\_est, by = c("qid"))

head(pew)

head(model\_est)

head(pew\_model\_est)

```

## 4.2) Plot estimated means against benchmarks

```{r}

plot\_comparison <- function(est\_table, method, caption){

graph <- ggplot(est\_table,

aes(x = pew\_estimate, y = method)) +

geom\_point() +

labs(x = "Estimates from Pew", y = caption) +

scale\_x\_continuous(limits = c(0,1)) +

scale\_y\_continuous(limits = c(0,1)) +

geom\_abline(intercept = 0, slope = 1, linetype = "dotted") +

coord\_fixed()

return(graph)

}

#when you plot the method, you need to specify which column you want to plot.

plot\_comparison\_imputation<-plot\_comparison(est\_table=weighted\_means\_long\_by\_group, method=weighted\_means\_long\_by\_group$weighted\_means\_by\_group, caption= "Weighted Means With 160 Group and Imputation")

plot\_comparison\_imputation

#their way of doing this

plot\_comparison(est\_table = pew\_model\_est,

method = pew\_model\_est$model\_prediction,

caption = "Model-based predicted values")

```

## 4.3) Plot distribution of estimation-benchmark differences

```{r}

#calculate difference

pew\_model\_est$diff <- abs(pew\_model\_est$model\_prediction - pew\_model\_est$pew\_estimate)

#plot

plot\_diff(pew\_model\_est)

```

```

\newpage

# Compare distribution of differences across methods and questions

Which questions worked well and which didn't? Which methods worked well for which questions?

```{r}

# put all differences into one table . Take all the differences out of each table and join them into an object called all\_diff

#need to change their data frame names to the corresponding data frames

all\_diff <- inner\_join(mean\_est, simple\_cell\_est, by = "qid") %>%

select(qid, diff\_mean = diff.x, diff\_simple\_cell = diff.y) %>%

inner\_join(., cell\_based\_est, by = "qid") %>%

select(qid, diff\_mean, diff\_simple\_cell, diff\_cell = diff) %>%

inner\_join(., cell\_est\_imputed, by = "qid") %>%

select(qid, diff\_mean, diff\_simple\_cell, diff\_cell, diff\_cell\_imputed = diff) %>%

inner\_join(., pew\_model\_est, by = "qid") %>%

select(qid, diff\_mean, diff\_simple\_cell, diff\_cell, diff\_cell\_imputed, diff\_model = diff)

# summarize

summary(all\_diff, digits = 2)

# calculate MSE, -1 one refers tot he last column which is the average difference. Give the column means froma pplying all dif to the last column, with two decimal places the function of x^2. In sum, this gives you the square of diff\_model (the modal based post stratification), which is the mean square error:

colMeans(apply(all\_diff[ ,-1], 2, FUN = function(x){x^2}))

# calculate average difference across all methods for each question. Create a new column called avg\_diff which you create by applying the mean Function to all but the last column (which is the model differences)

all\_diff$avg\_diff <- apply(all\_diff[ ,-1], 1, FUN = mean)

#subset to only the qid adn avg-diff

all\_diff[,c("qid", "avg\_diff")]

```

\newpage

# Optional Extension -- Approach 5: Multilevel-Model-based estimation with post-stratification (MRP)

### 5.1) Predict group means with multi-level regression model; combine with group weights to create weighted means

```{r}

## if using Bayesian estimation for multi-level model, you will need to load rstanarm

## note that Bayesian estimation is more computationally intensive/takes longer

#install.packages("rstanarm")

library(rstanarm)

# create container, an empty vector of all teh stratified group combinations (row) by relevant questions (columns)

mrp\_model\_predictions <- as.data.frame(matrix(nrow = nrow(population\_counts),

ncol = length(relevant\_questions), NA))

colnames(mrp\_model\_predictions) <- relevant\_questions

# loop through model fitting and prediction

#you need to redefine data as ds because tht is what you prefer to call it.

for (i in relevant\_questions) {

outcome <- data\_factor[ , i]

# fit -- note that this is using default priors

# nested the model name in "capture.out" to silently fit

#instead of printing a bunch of stuff to the terminal capture.output does not show it. 1 is the random effects term and whatever is after | is the age\_cat and race and region, but they don't do it for sex because that is binary (ask Cambria?)

output <- capture.output(multilevel\_model <-

glmer(outcome ~ sex + (1|age\_cat) + (1|race) +

(1|region), data = ds, family = binomial(link = "logit")))

# # predict using the multilevel model for each 160 stratified combination of the sex, race, region and age

mrp\_predictions <- predict(multilevel\_model,

newdata = population\_counts, type = "response")

# # errors? Are my probabilities within 0 and 1

prob\_range\_warning(mrp\_predictions)

# # feed into dataframe

mrp\_model\_predictions[ , i] <- mrp\_predictions

}

mrp\_model\_predictions

##################### Bayesian version with STAN #################################################

#STAN is used to specify a Bayesian statistical model. R package for Bayesian model.

library(rstanarm)

#This model is the same except htey add a , adapt\_delta=0.99 which changes the acceptable threshold for statistical signficiance to 0.01

for (i in relevant\_questions[1:2]) {

outcome <- data\_factor[ , i]

# fit -- note that this is using default priors

# nested the model name in "capture.out" to silently fit

output <- capture.output(multilevel\_model <- stan\_glmer(outcome ~ sex + (1|age\_cat) + (1|race) +

(1|region), data = ds, family = binomial(link = "logit"), adapt\_delta = 0.99))

# predict: instead of using predict you have to use posterior-linpred because this Extract the posterior draws of the linear predictor, possibly transformed by the inverse-link function

mrp\_predictions <- posterior\_linpred(multilevel\_model,

newdata = population\_counts, type = "response")

mrp\_predictions\_invlog <- exp(mrp\_predictions)/(1 + exp(mrp\_predictions))

mrp\_pred2 <- unname(apply(mrp\_predictions\_invlog, 2, mean))

#apply the mean funtion to mrp\_predictions\_invlog and then remove the name for something and store it in mrp\_pred2

# errors? See if the value is not between 0 and 1

prob\_range\_warning(mrp\_pred2)

# feed into dataframe

mrp\_model\_predictions[ , i] <- mrp\_pred2

}

#now they return to non Bayesian version

# bind to demographic categories and group weights--they creating a new dataframe that binds hte population counts adn the mrp\_model\_predictions (which can either be from the ordinary mrp or the bayesian mrp)

mrp\_wide <- bind\_cols(population\_counts, mrp\_model\_predictions)

head(mrp\_wide)

# reshape wide to long

mrp\_long <- mrp\_wide %>% gather(qid, predicted\_value,

-c(sex, age\_cat, region, race, group\_weight),

na.rm = F)

head(mrp\_long)

# weigh, sum by qid, match with pew. Create a new weighted predictin based on group\_weight and predicted\_value

mrp\_est <- mrp\_long %>%

mutate(mrp\_weighted\_prediction = group\_weight\*predicted\_value) %>%

group\_by(qid) %>%

summarise(mrp\_prediction = sum(mrp\_weighted\_prediction, na.rm = T))

head(mrp\_est)

# merge with pew benchmarks

pew\_mrp\_est <- inner\_join(pew, mrp\_est, by = c("qid"))

pew\_mrp\_est

```

### 5.2) Plot estimated means against benchmarks

```{r}

plot\_comparison(est\_table = pew\_mrp\_est,

method = pew\_mrp\_est$mrp\_prediction,

caption = "MRP predicted values")

```

### 5.3) Plot distribution of estimation-benchmark differences

```{r}

#calculate difference

pew\_mrp\_est$diff <- abs(pew\_mrp\_est$mrp\_prediction - pew\_mrp\_est$pew\_estimate)

#plot

plot\_diff(pew\_mrp\_est)

```

### 5.4) Compare differences from MRP with other methods

```{r}

#need to change names of different means to what I have in my data analysis

all\_diff <- inner\_join(all\_diff, pew\_mrp\_est, by = "qid") %>%

select(qid, diff\_mean, diff\_simple\_cell, diff\_cell, diff\_cell\_imputed, diff\_model, diff\_mrp = diff)

# summarize

summary(all\_diff, digits = 2)

# calculate MSE

colMeans(apply(all\_diff[ ,-1], 2, FUN = function(x){x^2}))

```

Undocumented Version

Question Appendix.

Unless noted in parentheses, all questions are from GSS

1. What is your gender? A) male, B) female?
2. Are you of Hispanic, Latino, or Spanish origin? A) No, not of Hispanic, Latino or Spanish origin. B) Yes, Mexican, Mexican American, Chicano; B) Yes, Puerto Rican, C) Yes, Cuban; D) Yes, other;
3. What is your race? A) white, B) black/African American, C) American Indian, D) Asian Indian, E) Japanese, F) Native Hawaiian, G) Chinese, H) Korean, Guanmanian or Chamomo, I) Filipino, J) Vietnamese, K) Samoan, L) Other Asian (fill in blank: ),
4. What is your date of birth? Give month and year only.
5. What is your annual income (as you would report it in the Census)? Fill in blank.
6. What is the highest degree or level of school you have completed?
7. No schooling completed,
8. Nursery or preschool through grade 12: Fill in blank.
9. Regular high school diploma
10. GED or alternative credential,
11. Some college credit, but less than 1 year of college credit,
12. 1 or more years of college credit, no degree,
13. Associate’s degree (for example: AA, AS)
14. Bachelor’s degree (for example: BA, BS)
15. Master’s Degree (for example MA, MS, Meng, Med, MSW, MBA),
16. Professional degree beyond a bachelor’s degree (for example MD, DDS, DVM, LLB, JD)
17. Doctorate degree (for example PhD, EdD)
18. In which US State do you live? Fill in blank:
19. Were you born in the US? A) Yes, B) No.
20. With which party do you identify with?
21. Democrat, B) Republican, C) Independent
22. This survey is halfway over, so if you are really paying attention, click strongly agree.  
    A) strongly disagree, B) disagree, C) agree, D) strongly agree
23. Do you think the number of undocumented immigrants to America nowadays should be... a) increased a lot, b) increased a little, c) remain the same as it is, d) reduced a little, e) reduced a lot, f) don’t know
24. In the long run, do you think that undocumented immigrants who are immigrating to the United States today will make American society better, will make American society worse, or do you think that today's undocumented immigrants won't affect American society one way or another?

Answers: a) better, b) worse, c) won’t affect American society one way or another, d) don’t know

Questions 20-25 start with the following opener:

|  |
| --- |
| Now I'm going to read you some statements and would like to get your reaction to them. After I read each statement, please tell me if you strongly agree, |
| agree, neither agree nor disagree, disagree, or strongly disagree with the statement. How strongly would you agree with the following statements on a scale of 1 to 5 (where 1 is highly disagree and 5 is highly agree):   1. “Undocumented immigrants are getting too demanding in their push for equal rights.” |

1. “The Irish, Italians, Jews, and many other minorities overcame prejudice and worked their way up. Today's undocumented immigrants should do the same without any special favors.”
2. “English will be threatened if other languages are frequently used in large undocumented immigrant communities in the U.S.”

For 26-30: (the sequence of the following options will be randomized for each participant who will respond on a 1-5 scale (ranging from very unlikely (1) to very likely (5) ). Each question will be prefaced with the following:

|  |
| --- |
| “What do you think will happen as a result of more undocumented immigrants coming to this country?  Is the following result very likely, somewhat likely, |
| not too likely, or not at all likely?” where very likely is 5 and not at all likely is 1. |

|  |
| --- |
| 1. Making the country more open to new ideas and cultures 2. Higher crime rates 3. People born in the US losing their jobs, 4. Higher economic growth 5. Making it harder to keep the country united. |
|  |

1. Vignette: “Recently, there have been several investigative reports regarding the condition of undocumented immigrants being held in detention centers around the United States. The reports indicate that some detainees are held for long periods of time under inhumane conditions, lacking sufficient nutrition, proper sanitary facilities, access to legal counseling, or even contact with family members. The reports also discuss cases of medical negligence concerning the treatment of detainees with serious health problems. In one case, a detainee told authorities he was suffering from severe chest pain, profound shortness of breath, and dizziness, and asked for immediate medical attention. However, these situations require supervised transportation to a hospital several miles away from the detention center. In response, detention center officials denied the detainee this medical service because they considered him to be a flight risk. The officials said they had a reasonable cause to deny medical attention outside the facility. The detainee, on the other hand, said that the denial of medical attention was unwarranted and that his health was continuing to deteriorate.”

Do you think the government should transport this undocumented migrant (SPLIT) to the hospital, at the risk of enabling him to escape?

A) Yes,

B) No.

1. How would you describe the media sources you most frequently consume?

Seven point scale from very liberal (1) to very conservative (5)

1. How much do you identify as conservative, moderate, or liberal?

extremely liberal

liberal

slightly liberal

moderate

slightly conservative

conservative

extremely conservative

1. Is there anything else you would like to tell us about this survey?