M^3 Challenge First Runner Up, Magna Cum Laude Team Prize of $15,000

High Technology High School, Team #249
Lincroft, NJ

Coach: Raymond Eng

Students: Sidney Douglas Buchbinder, Christian Paul Gennaro, Joshua Ma, Alexander Saso Pavincic, Matthew Ethan Warshauer

Summary

A group of researchers from a think tank in Washington, D.C. have released a scathing indictment of the United State Census Bureau. At the request of Congressional leaders, the Team 249 Foundation investigated the practices of the Bureau, in preparation for the upcoming counting of the population in the Constitutionally mandated decennial Census of the United States. The Foundation was asked to study whether the census has historically had any significant error. It was also asked to determine a method for Congress to use along with the Census results in order to apportion seats in the House of Representatives to the various states of the United States. Finally, the Foundation was tasked with investigating the way that Congressional districts are drawn nationwide.

While looking into potential census error, the Foundation discovered that the Census Bureau has historically undercounted people in every state. Certain states suffered more than others. In 2000, the Bureau missed an estimated 2.67% of Alaska's population. Nationwide, the Bureau missed an estimated 1.18% of the total population.

Because the Federal government uses census data to determine how to distribute Congressional seats in the House of Representatives and how to allocate much of its spending, these errors can result in serious consequences for the average citizen. The Foundation found that the errors in the 2000 census resulted in 48 billion dollars being given to the wrong states over the last ten years. The Foundation proposed an elegant solution to this problem of undercounting. It suggests to Congress that they implement a follow-up "micro-census" which takes a sample of regions across the country to correct for errors in the Census data. It has shown that this efficient technique would improve the quality of the Census data significantly.

The Foundation also claims that the current method of apportioning seats in the House of Representatives is inherently unjust and unfair. Under the present system, different people in different states have a different amount of representation in the House. People in sparse states like Wyoming get more than their fair share, while people in populous states like California get less than theirs. In response to this problem, the Foundation suggested that the size of the House of Representatives could be adjusted after each census to optimize the fairness to each citizen. It proposes a mathematical calculation where each person would have the representation they deserve in Congress. One conclusion that the Foundation noted is that the size of the House of Representatives is most definitely too small and must be increased in order to make the system fair. It is suggested the House of Representatives would be fairer if immediately increased to 489 members, up from the present total of 435.

The final conclusions of the Team 249 Foundation concerned how states draw Congressional districts. The Foundation found that State Legislatures often abuse this power to manipulate electoral politics and secure the reelection of incumbents. Seeing how harmful this is to American ideals of democracy, the Foundation suggested a series of steps to standardize this process and make it fair for all citizens. The recommendations are manifold. First, states should be prohibited from redistricting mid-decade in order to prevent scandals such as those of Texas and California around the time of the 2004 elections. Second, all states should appoint Independent Commissions with expressly apolitical purposes to redistrict. Third, states should follow a procedure designed by the Foundation to ensure that each and every district is logically and fairly drawn.

Congressional leaders have received the Foundation's report. They have issued statements praising it as an insightful analysis and a set of creative solutions, all based on well analyzed mathematical models.

Table of Contents

Introduction

Background

"Representatives and direct Taxes shall be apportioned among the several States which may be included within this Union, according to their respective Numbers, which shall be determined by adding to the whole Number of free Persons, including those bound to Service for a Term of Years, and excluding Indians not taxed, three fifths of all other Persons. The actual Enumeration shall be made within three Years after the first Meeting of the Congress of the United States, and within every subsequent Term of ten Years, in such Manner as they shall by Law direct."

The preceding paragraph on the formation of a nationwide census is part of the second section of Article I of the United States Constitution. The nation's first census, conducted in 1790, asked for the number of free white males 16 years and older, free white males less than 16 years old, free white females, other free people by color and gender, and slaves. It found that there were approximately 3.9 million people in the country, about six-hundred thousand of which were slaves (Census of Population and Housing: 1790 Census, 2010). In the following decades' censuses, questions on occupation as well as more meticulous ones on age were added to the survey. By 1850 the name and exact age of every member of a household were required on the census form. Since then, censuses have remained largely the same except that they now show the dates of birth, years of immigration, and relationships (father, brother, husband) of the people in a household (Genealogy - The Evolution of the Census).

The biggest problem with the current census is that it severely undercounts the population. Dr. Eugene Ericksen of PricewaterhouseCooper has estimated that 1.18% of the total population was undercounted in the 2000 Census and that the total amount of people in the country who were not taken into account was around 3.4 million, a massive number considering that the population determined by that census was approximately 281 million people (2001). The places where the undercount is greatest are areas with high concentrations of impoverished and/or illegal inhabitants. Poverty-stricken regions are the most likely to have a large amount of homeless residents; because the census is sent through the mail these people are often not counted. In addition, there are many housing units where the government is unsure of their vacancy status. Even if they are occupied, a certain amount of them have household counts that vary drastically over a short interval of time. Additionally some of these houses do not have telephone numbers, so surveyors have a very difficult time determining if a building that does not respond to the census form is inhabited or not. Immigrants also are commonly undercounted. Many of them live in fear of a government that they do not trust, especially ones who are in this country illegally. As a result, they often try to conceal themselves and avoid filling out the census form and revealing themselves to the government (Sung, 1991). As later evidenced by the astronomical rates of undercounting in states such as Alaska and Hawaii, the United States government has also had a difficult time counting Native Americans. Despite small amounts of overcounting from duplicated forms, the overall census still drastically undercounts the population. This report concludes that all of these factors contribute to approximately 48 billion dollars being incorrectly allocated to various parts of the nation.

The aforementioned facts and figures clearly show that a different protocol has to be put into place to assuage the enormous undercounting. This new procedure must significantly reduce the U.S. Census' undercount. Any new error that it introduces cannot be nearly as consequential as that of the undercounts.

Additional debate has arisen over whether the system by which the House of Representatives is apportioned should be changed because of the inherent inflexibility due to the legislated number of Representatives in the House. By creating elasticity in the number of seats, a fairer system can be established. Empowering each citizen as equal to their fellow voter is a quintessential American ideal.

A final matter of discussion is the issue of how Congressional districts are drawn in each state. In the majority of states they are created by the State Legislatures and are drawn in a very partisan way to manipulate the electoral system. They are also frequently redrawn with the goal of preserving incumbent elected officials. Notorious examples of this redistricting, known as gerrymandering, were apparent around the 2004 elections in California and Texas. Like the misapportionment of Congressional seats to the states, the conniving of Legislatures to direct Congressional districts is also counter to the ideas of American democracy.

Restatement

In essence this paper does the following:

1. Determine whether the census figures should be adjusted to counteract the undercount and how this should be done. Estimate the errors that this new procedure introduces and compare them to that of the undercounts.
2. Devise a new method for Congress to follow to apportion the House of Representatives. Explain why it is superior to other solutions that deal with this problem.
3. Recommend how states should fairly draw Congressional districts. Justify your proposal.

Global Assumptions

In order to create statistical models to solve the previously mentioned tasks, a few global assumptions needed to be made:

I. All gathered information from external sources is correct and reliable.
II. Undercounting will inevitably continue to exist in the U.S. Census, as evidenced by historical error.
III. Impoverished neighborhoods and illegal immigration will continue to exist.

Undercount Adjustment

Rationale

The massive nature of the national census is the primary cause of undercounting, as it is difficult to ensure that everyone is properly counted on such a large scale. To better assess the population of the United States, this report details a sampling method to measure the population of randomly chosen areas throughout the nation with much greater scrutiny. It is believed that trained teams will be able to sample smaller areas with much greater precision, and comparisons of these new accurate populations with the census populations can provide an idea of the inaccuracies of the census data. Through a statistical analysis, these comparisons are used to establish overall correction factors for each state, factors that can be multiplied by the census state populations to obtain more accurate, corrected populations.

The factors behind undercounting are numerous and vague. Due to the nature of the problem, the definition of undercounting entails missing information that could not be obtained. Estimation of variables such as the uncounted homeless is difficult, and studies to find whether or not an individual was counted suffer from the exact same nonresponse bias as the original census. The best method is to use another round of measurement, this time using smaller, intensive population sampling.

Assumptions

I. Due to difficult-to-predict variations in birth- and death-rates between the time that the census is taken and the time that the following sample is taken, it is assumed that the true population has not changed.

II. Different types of demographics (cities, suburbs, rural areas) typically have different amounts of error in their census calculations (Bock, Velleman, & De Veaux, 2004, p. 230).

III. Within the types of demographics the error is relatively constant because of similar population densities across each demographic.

IV. A team of trained surveyors will be able to count virtually all of the people in a town due to their more meticulous and precise methods of counting.

Undercounted Population by State

To justify that the census undercounts population unequally between states because of poverty, illegal immigration, and presence of Native Americans, it needs to be established that there is a significant difference in the rates of undercounting between states that have more or less of them. A Chi-Square (χ^2) test of "goodness-of-fit" was used for this. Data from auditing firm PricewaterhouseCoopers that estimated the 2000 Census undercount by state was used as the observed range of values. If the rates of undercounts were constant for all states, then the amount of undercounts per state would be directly proportional to the states' populations. As a result, the populations by state according to the 2000 Census multiplied by the national average undercount value of 0.012 (as determined by PricewaterhouseCoopers) were treated as the expected range of values. The goodness-of-fit test squares the individual residuals and divides them by their expected values ($(x_{obs} - x_{exp})^2 / x_{exp}$). The hypotheses used in the test were H_o: $\rho_1 = \rho_2$ and H_A: $\rho_1 \neq \rho_2$, where ρ_1 is the observed amount of undercounts in a state and ρ_2 is the expected amount of undercounts based on the national average. Before the test could be conducted, the following conditions had to be met:

I. Counted Data Condition – All of the data are counts for their categories

II. Randomization Condition – Individuals who have been counted must be a random sample from a population

III. Expected Frequency Condition – At least five individuals must be counted in each category (Bock, Velleman, & De Veaux, 2004).

All three conditions were met, so a goodness-of-fit test was conducted. A level of significance (α) of 0.05 was used. The test yielded a χ^2 of 432,532 and a P-value of 0.0000. Because the P-value was less than the level of significance ($P < \alpha$), the data rejects the null hypothesis and supports the alternative hypothesis. There is statistically significant evidence that suggests there is an uneven distribution in the rates of undercounting in the 2000 Census between states.

Because of this difference in the rates of undercounting throughout the United States, different states have a larger percentage of uncounted people. This uneven spread causes an unfair distribution when it comes to the general functions of the census. One of its main functions is to help the federal government allocate necessary funds to regions that are in the most need and that contain the most people. These funds help communities build hospitals, schools, senior centers, public works, and emergency services. In 2010, nearly 400 billion dollars will be distributed throughout the nation on the basis of census data (How It Affects the Nation). Because of this misallocation, nearly 48 billion dollars has been given to the wrong people over the last decade. If the census continues to undercount 1.18 percent of the total population of the United States as it did in 2000, many regions, and the people that live in them, will not acquire the funds that they need (PricewaterhouseCoopers, 2001). This is especially problematic because the regions that have the largest error in their undercounts often have the most homeless and destitute people and need the funds the most.

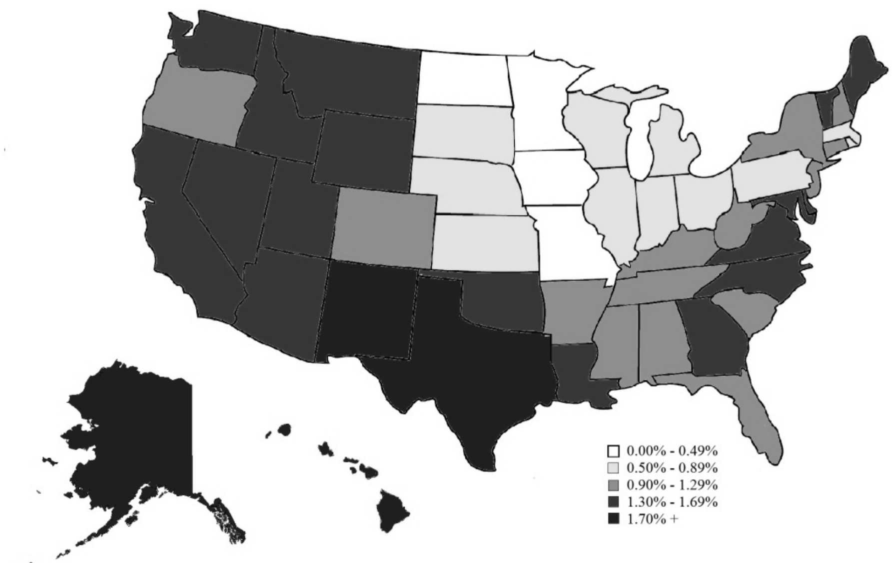

Figure 1 – Percentage of Undercounted Population in 2000 Census by State (PricewaterhouseCoopers, 2001)

Procedure

It is important to begin with the Census as currently conducted in the United States. This methodology considers that Census, conducted with the same techniques as have been used in the past, as both a necessary and a valuable starting off point for the purpose of producing more accurate totals for the population. After the Census has been conducted, this methodology then proposes a second, multistage sample stratified by state. Within each state, the populations of zip code regions (ZCRs) are sampled, but the ZCRs are stratified by demographics (urban, suburban, and rural). By dividing these ZCRs into urban, suburban, and rural demographics, we can more accurately account for the fact that ZCRs of differing demographics are likely to have had a differing error in the original Census. This is because urban areas have more uncounted people than suburban and rural areas, because a large portion of the uncounted consists of the homeless, the poor, and the indigent living in the cities (Bock, Velleman, & De Veaux, 2004, p. 230). To determine the demographic of a given ZCR, its population density based on the Census data was calculated. As defined by the Census Bureau, ZCRs with a population density greater than 1,000 people per square mile are considered urban. Similarly, areas with less than 400 people per square mile are considered rural, leaving areas with between 400 and 1000 people per square mile as suburban (United States Census Bureau, 2009).

In order to maintain statistical independence, the follow-up sample of ZCRs can only contain up to 10% of the total population of ZCRs. These ZCRs will be randomly selected within each demographic. In other words, 10% of urban ZCRs in each state will be selected, 10% of suburban ZCRs in each state will be selected, and 10% of rural ZCRs in each state will be selected. After randomly selecting which districts will be resurveyed, the Census Bureau should compile their work force to take a population count of these ZCRs by sending trained professionals into the chosen ZCRs to precisely and accurately measure the population. By clustering the follow-up sample within ZCRs, the Census Bureau can more efficiently use its resources and still maintain statistical significance. After the sample is complete, the data that was collected can be compared to the data collected by the whole census for those ZCRs. Then, by dividing the census population p_s by the sample population p_c for a ZCR, the ratio k is calculated that stands for the correction ratio for the population of the census in that ZCR,

$$\frac{p_s}{p_c} = k$$

This ratio can be found for each sampled ZCR under a common demographic (urban, suburban, or rural) in a state. In each demographic, the average ratios are denoted k_u, k_s, and k_r, respectively. Furthermore, for each area the standard deviation, s (s_u, s_s, and s_r for each area, respectively), can be calculated as

$$s = \sqrt{\frac{\Sigma(k - \bar{k})^2}{n - 1}},$$

where the summation is over all values of k for ZCRs in that area and n is the number of ZCRs sampled from that area. This variance will be used later for error analysis.

A weighted arithmetic average of the average ratio of each demographic is used to calculate the average for the entire state, where n_u, n_s, and n_r represent the weighting for the urban, suburban, and rural regions, respectively. Weighting for each demographic is the number of ZCRs of that demographic in the state.

$$k_{state} = \frac{(n_u k_u + n_s k_s + n_r k_r)}{n_u + n_s + n_r}$$

This is the overall correction ratio for the state, and we can multiply the state population from the census, $p_{c,state}$ by this ratio to obtain an adjusted population estimate for that state:

$$p_{c,state} \times k_{state} = p_{adjusted}$$

Because the main purpose of the census is to determine the distribution of the population amongst states for apportionment and the distribution of funds, it is not necessary to sum the adjusted populations for a national total.

Error Analysis

When sampling the populations of the ZCRs, the variance for each demographic was calculated using

$$s = \sqrt{\frac{\Sigma(k \quad \bar{k})^2}{n - 1}}.$$

Dividing each value of a random variable by a constant divides the average by that constant and the variance by the constant squared. The variance of the sum of random variables, assumed to be independent of each other, is simply equal to the sum of the variances of each random variable. Therefore, the standard error of k_{state}, SE_{state}, is

$$SE_{state} = \sqrt{\frac{1}{(n_u + n_s + n_r)^2}\left(\frac{n_u^2 s_u^2}{n_u} + \frac{n_s^2 s_s^2}{n_s} + \frac{n_r^2 s_r^2}{n_r}\right)}$$

$$SE_{state} = \frac{1}{n_u + n_s + n_r}\sqrt{(n_u s_u^2 + n_s s_s^2 + n_r s_r^2)}.$$

The standard deviation, *s*, is a measure of the spread of the distribution. Because only the sample's standard deviation is known, a Student's *t* distribution is used with a *t* statistic that can be found such that it includes 95% of the distribution with $n_u + n_s + n_r - 1$ degrees of freedom. These *t* statistics are available using a table of values or a computing utility such as a graphing calculator with a statistics package.

Having calculated the appropriate 95% *t* statistic, t^*, the upper and lower bounds are

$$k_{state} \pm t^* SE_{state}$$

Using this equation, the agency can be 95% confident that the true population of the United States is in the interval $(k_{state} - t^* SE_{state}, k_{state} - t^* SE_{state})$. This interval provides a sense of the magnitude of the error in our estimate.

Testing

Testing this plan entails adjusting census population totals by implementing this plan by conducting it on a smaller scale. For instance, count the population of the state of Wyoming as a part of the traditional census count, and then use this sampling technique to adjust the population total of Wyoming and compare that to values calculated by the Census Bureau as part of its regular operations.

Apportionment

Rationale

The House of Representatives is currently appointed using the Huntington–Hill Method. This was established in 1911 by Joseph A. Hill, the Chief Statistician of the Bureau of the Census, and Edward V. Huntington, Professor of Mechanics and Mathematics at Harvard University (University of Alabama, 2001). In the same year, Public Law 62-5 established the number of members of the United States House of Representatives to be four hundred thirty five, which, apart from a temporary increase to 437 members with the statehood of Alaska and Hawaii, has remained constant (Office of the Clerk). The U.S. Constitution sets a size limit for the size of the House that reflects both the number of states in the United States as well as the United States population, i.e., that every state have at least one Representative and that there be no more than one Representative for every thirty thousand (The Constitution of the United States of America, 1788).

The method of apportionment described in this section is developed to create a political environment that adheres to the founding principles of the House of Representatives, particularly equal representation proportional to population. The criteria used to calculate apportionment are compared to criteria used in other methods, and they are subsequently shown to be better measurements of the aforementioned equal representation.

Assumptions/Definitions of Terms

I. Ideal representation takes the form of equal representation closest to that of a true democracy. It occurs when each Representative votes in the name of an equal number of people, regardless of state and voting district.

II. The goal of apportionment is to create this ideal representation and fairness, without any political bias and under the restrictions of state boundaries.

III. The drawing of Congressional districts is assumed to fairly distribute voter representation, as clarified in part three of this report.

IV. Fairness of apportioning is defined in terms of the portion of people in each state that are under- or overrepresented as compared to the "ideal representation" of every United States Citizen. Specifically, fairness is calculated by determining the standard deviation in the percent of the population of each state that is either not represented or that is overrepresented (in smaller states) by the proposed apportioning of Representatives.

Design

Apportioning is meant to distribute the Representatives of the United States evenly so that each Representative votes in the name of the same number of citizens. This ratio is defined as voting power, *v*. In ideal representation, *v* should be the same for every Representative and should equal the ratio of the population of the United States to the number of Representatives.

The number of seats in the House, *N*, is identified as the independent variable, a value to be adjusted in order to achieve as close to idealized representation as possible. Akin to its definition in other apportionment methods, the standard divisor *D* is the ratio of the U.S. population, *P*, to the number of seats. This method of apportionment attempts to give each representative a voting power as close as possible to *D*, and it begins by assigning each representative this voting power:

$$D = \frac{P}{N}.$$

The index *i* is defined such that $1 \le i \le 50$ and each state has a unique value of *i*. The population of the state with index *i* is p_i, and

$$\sum_{i=1}^{50} p_i = P \; .$$

For every state, the largest number of seats that the state can have is defined to be n_i, and it is limited by its population:

$$n_i D \le p_i$$

since the number represented per Representative times the number of Representatives in the state cannot exceed the population of the state. With a set value of *D* and the restriction that n_i must be an integer, there will be a maximum number of Representatives available to each state, such that the total number of people represented $r_i = n_i D$ cannot exceed p_i. A difference between the population of each state and the number of people represented, $p_i - n_i D$, must exist; this difference is defined to be u_i, the number of "unrepresented" people per state. Thus, $u_i + r_i = p_i$, where r_i is the number of represented people in the state. In addition, u_i can be calculated by subtracting r_i from p_i:

$$u_i = p_i - \left\lfloor \frac{p_i}{D} \right\rfloor D$$

and so

$$\left\lfloor \frac{p_i}{D} \right\rfloor D = r_i$$

is the product of D and the floor function of the population in state i divided by D (the voting power). It should be noted that because the United States Constitution requires at least one Representative per state, for all small states with $p_i < D$ the state is instead over-represented rather than underrepresented. Later in calculations, u_i will reflect the "unfairness" in representation for state i, and in this situation, u_i for the overrepresented state is better represented with $D - p_i$, the extra represented people assigned to that state.

As the value of N was varied, the u_i value was calculated for every state, based on U.S. Census 2000 population per state data (United States Census Bureau). It is important to note that because the number of Representatives must be odd, N was restricted to only odd numbers.

With the u_i calculated for every state over various N values, the ratio $\frac{u_i}{p_i}$ was calculated, representing the percentage of people in each state's population that remain "unrepresented". It was then shown that if the $\frac{u_i}{p_i}$ ratio is equal for all states, the $\frac{p_i}{n_i}$ ratio, the voting power of each representative in each state, is constant as well, no matter which state is being considered.

Proof:

$$u_i + r_i = p_i$$
$$u_i = p_i - r_i$$

$$u_i = p_i - n_i D$$
$$n_i D = p_i - u_i$$
$$c = \frac{u_i}{p_i} \Rightarrow u_i = cp_i$$

$$c = \frac{u_i}{p_i}$$

$$n_i D = p_i - cp_i = p_i(1 - c)$$
$$\frac{p_i}{n_i} = \frac{D}{1 - c}$$

D and k are independent of the state, i

$$\therefore \frac{p_i}{n_i} \text{ is independent of the state, } i$$

The ideal case, where the voting power of each Representative in each state is equal, is satisfied when $\frac{p_i}{n_i}$ is independent of the state. This requires that $\frac{u_i}{p_i}$ is independent of the state. To strive for this equality between states, the distribution of $\frac{u_i}{p_i}$ for all valid values of i and a set value of N is analyzed. The standard deviation is used as a measure of the spread of the distribution, and the distribution in which the N produces the lowest spread is determined to be the optimal apportionment in which the voting power of the representatives varies the least between states. This in turn results in a more equal congressional representation of each person in each state.

Every ten years, the data from the latest census will change, and the appropriate decisions will vary. Therefore, this report establishes guidelines to use to best decide what value of N to utilize. Below is a graph that displays the standard deviation for each value of N, as based on the 2000 Census population data.

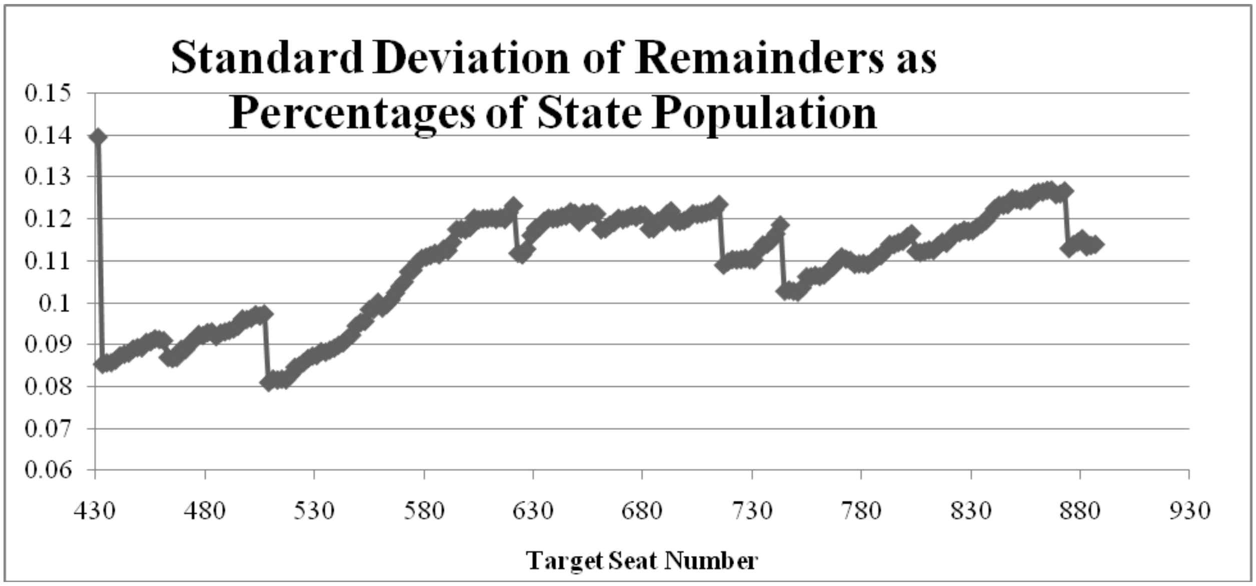

Figure 2 - Standard Deviation of Remainders as Percentages of State Population

Examining the above graph, for $N = 509$ the standard deviation appears to be the smallest. Unfortunately, there is no mathematically solid way of finding the minimum over all values of N — in the above graph, the maximum is limited to 887. After a certain point, increasing N, the number of seats in Congress, does not appear to decrease the standard deviation as significantly. A subjective decision is made to utilize the minimum that occurs at $N = 509$.

This value of N is actually a *target* number of seats. In reality, the resulting number of seats is less because some seats are spread amongst the population that remains "unrepresented." Therefore, when determining the actual number of seats the D value for the corresponding N, in this case $D = 552896$ (for the 2000 population and 509 seats) must be used. Again, because the Representatives' voting power is equal to $D = \frac{P}{N}$, the number of congressmen per state must reflect the p_i of that state, and not just the represented population r_i of the state. The number of congressmen per state, n_i, is calculated:

$$n_i = \left\lfloor \frac{p_i}{P} \right\rfloor * D \ ,$$

and in this case specifically for this data,

$$n_i = \left\lfloor \frac{p_i}{281424177} \right\rfloor 552896$$

Using this formula, the proposed number of seats per state was calculated. The table below shows each state and its current and proposed congressmen.

Table 1 - States and their current, proposed, and change in Representatives

State	Current Reps	Proposed Reps	Change in Number of Reps	**State (cont)**	Current Reps (cont)	Proposed Reps (cont)	Change in Number of Reps (cont)
Alabama	7	8	1	Nebraska	3	3	0
Alaska	1	1	0	Nevada	3	3	0
Arizona	8	9	1	New Hampshire	2	2	0
Arkansas	4	4	0	New Jersey	13	15	2
California	53	61	8	New Mexico	3	3	0
Colorado	7	7	0	New York	29	34	5
Connecticut	5	6	1	North Carolina	13	14	1
Delaware	1	1	0	North Dakota	1	1	0
Florida	25	28	3	Ohio	18	20	2
Georgia	13	14	1	Oklahoma	5	6	1
Hawaii	2	2	0	Oregon	5	6	1
Idaho	2	2	0	Pennsylvania	19	22	3
Illinois	19	22	3	Rhode Island	2	1	-1
Indiana	9	11	2	South Carolina	6	7	1
Iowa	5	5	0	South Dakota	1	1	0
Kansas	4	4	0	Tennessee	9	10	1
Kentucky	6	7	1	Texas	32	37	5
Louisiana	7	8	1	Utah	3	4	1
Maine	2	2	0	Vermont	1	1	0
Maryland	8	9	1	Virginia	11	12	1
Massachusetts	10	11	1	Washington	9	10	1
Michigan	15	18	3	West Virginia	3	3	0
Minnesota	8	8	0	Wisconsin	8	9	1
Mississippi	4	5	1	Wyoming	1	1	0
Missouri	9	10	1	**Total Apportionment**	435	489	54
Montana	1	1	0				

Comparison to Other Methods

Other methods such as the Hamilton, Jefferson, Adams, and Huntington–Hill methods of apportionment utilize similar concepts of Standard Divisor, *D* (University of Alabama, 2001). The way in which Representatives is calculated differs slightly amongst the models. In every method, however, the value of *N* is fixed. This causes error in representation as it does not allow for flexibility in the Standard Divisor. The Standard Divisor does not divide well into certain population sizes, and this inflexibility results in an uneven distribution of the $\frac{n_i}{p_i}$ ratio analyzed in this paper. Because this is a ratio of how many citizens Representatives stand for in their state, a widespread distribution of the ratio suggests uneven representation of people amongst the states.

This report recommends the method outlined above because it is able to adjust the number of seats in the House every ten years upon recorded changes in the population. Using the proposed method, legislature may restructure the House of Representatives so that the populations of each state are represented in a more equal and proportional way.

Redistricting

Rationale

To determine the recommendations that should be made to the various states in order to ensure that Congressional districts are all drawn as fairly as possible, the current state of affairs in the United States regarding the drawing of legislative districts was investigated. Under the U.S. system of federalism, each state is responsible for drawing its own districts. As of early 2010, 36 states allow the state legislature to draw districts. In 5 states, independent (or bipartisan) commissions are given the authority to draw districts. Two states grant independent authorities the power to propose district lines, but the state legislatures reserve the right to approve those plans. The final seven states have sufficiently small population that the entire state is a single Congressional district, so no districting is necessary (Purdue University, 2010). It is important to note that Congress does have the authority to regulate the redistricting process. Additionally, the Supreme Court decided in “Baker v. Carr” (1962) that matters of redistricting also qualify as judicial questions to be ruled upon by the court system. It is also important to note that this redistricting process usually, but not always, happens following the decennial census.

After this analysis, one thing was clear. The states that allowed the legislature complete control over the drawing of legislative districts suffered under an unmatched amount of gerrymandering – the act of “dividing (a territorial unit) into election districts to give one political party an electoral majority in a large number of districts while concentrating the voting strength of the opposition in as few districts as possible” (Merriam-Webster). This disenfranchises opposition voters in the district who are stripped of their political power.

The Tanner Fairness and Independence in Redistricting (FAIR) Act, which was never signed into law, outlines certain characteristics of fair redistricting that were adopted in this analysis as in line with traditional standards. These include a focus on independent commissions, restrictions of mid-decade redistricting, respect for regulations of the Voting Rights Act of 1965, acknowledgement of the principles of “one man, one vote” as noted by the Supreme Court, and the principle that districts should be as contiguous and compact as possible (Fair Vote, 2006).

Assumptions/Definitions of Terms

V. Fair is defined as “impartial and honest: free from self-interest, prejudice, or favoritism and conforming with the established rules” (Merriam-Webster, 2010). In practice, this means that the redistricting must be blind to party, race, or any other demographic characteristic.

VI. Districts drawn with the intention of either helping or hurting a specific race are equally discriminatory and are not the intention of the Voting Rights Act of 1965 (Miller v. Johnson, 1995). Racial issues are best served by creating districts without respect to race.

VII. It is optimal to maintain compactness and contiguity when constructing districts. Compactness is proportional to the ratio of Area to Perimeter (higher values are more compact). Contiguity is the principle that any district must be a self-contained, discrete geographic unit, unbroken by any other district.

VIII. It is optimal to draw districts that align with current political and geographic boundaries. This includes but is not limited to town borders, rivers, and county lines.

IX. Independent Commissioners will remain independent for as long as they serve and will not be swayed in any significant way while carrying out their duties.

X. The system used in this methodology to appoint members to the Independent Commissions is sufficient to ensure that the board is sufficiently apolitical.

Recommendations

The following instructions can be given to the state legislatures so that they can increase the fairness of their Congressional redistricting following the upcoming decennial census and for all subsequent censuses. It is suggested that Congress give these recommendations the force of law so that errant legislatures are not tempted to subvert them.

1. Limit redistricting to only once per decade, following the decennial census, for each state. This would eliminate such debacles as the 2003 redistricting of Texas by the Republican Texas State Legislature that resulted in the sudden loss of 5 Democratic seats (Greenhouse, 2005). It would also counter ploys such as those of California politicians who redistricted in 2004 such that not a single one of the 53 Representatives lost their seat in the election of that year (Nagourney, 2005).
2. In every other year between each census (i.e. 2001, 2003, 2005, 2007, and 2009), the Legislature of each state should nominate, and the Governor of each state should approve of two members to serve on the next Independent Redistricting Commission following the next decennial census (i.e., 2010). The two members appointed should be respected members of the community who have never run for public office or served as an officer in any organized political organization. They may be members of political parties, but should each be from a different political party. By spreading out the appointments over time, the state can ensure that changing populations and political makeups of the state are reflected in the Independent Commission by the multiple legislative and gubernatorial terms that are involved in the appointment process. At the end of the decade, each Commission can meet with its ten members and conduct the process of redistricting as described below.

The following steps are to be conducted by the Independent Commission, not the State Legislatures. It should be noted that the goal is to create n districts, where n is the number of Representatives apportioned to the state following the decennial census. The population of each district should be as close as possible to m/n where m is the population of the state. This fulfills the requirement of "one man, one vote." To be considered close enough, district populations should be within 1% of m/n.

3. Delineate boundaries across the State. Essentially, this means to create a map of the state with every boundary marked. These include but are not necessarily limited to (it is at the discretion of the Commission) town borders, geographic boundaries such as rivers, and county lines. Each of these boundaries should be assigned a level where higher levels are more important boundaries – the ones that should not be broken. For example, town borders might be Level 1, rivers might be Level 2, and county lines might be Level 3.

Obviously, the actual borders of the State are at the highest possible, as they cannot be broken by any Congressional district.

4. Begin at the Northwestern most corner of the state. This is Point A. This is an arbitrary point and could be any other corner at the discretion of the Commission.
5. Begin to draw a district starting at Point A and expanding rectangularly from that point. This means that the district should expand in an equal amount horizontally and vertically. This growth should be centered around point A. From here on, this technique shall be referred to as "growing" or "growth" – always from Point A. Any growth that crosses a boundary is ignored and not drawn.
6. As the district grows, it will encounter boundaries, including state boundaries. At any point where the growth encounters a boundary the growth should stop. However, the district should continue to grow up to the boundary, filling in all space. This has the effect of creating fully solid regions completely enclosed by boundaries. If a region were to be surrounded entirely by a growing district, that region should be subsumed into the district itself. This also holds if the region is surrounded only by the growing district and the outer boundaries of the state itself.
7. Once a closed region is created, the Commission should compare the population of that region to *m/n*. If it is within the range of 0.99**m/n* to 1.01**m/n*, then the region is a completed Congressional district, and the Commission should advance to step 11. If it is less than 0.99**m/n,* then the Commission should continue with step 8. If it is more than 1.01**m/n,* then the Commission should proceed to step 10.
8. The Commission should examine the boundaries currently enclosing the potential district and "break" any Level 1 boundaries. If there are no Level 1 boundaries, then any level 2 boundaries should be broken. This should continue until at least one boundary can be broken. A "broken" boundary is no longer an impediment to the growth of a district. For example, if a region was bounded on two sides by town borders (Level 1) and on the other two sides by two rivers (Level 2). The Commission would examine this region and "break" the two town borders. The two river borders would not be "broken." If, however, the region was surrounded on all sides by a river (Level 2), then all of the Level 2 borders would be "broken," because these were the lowest level boundaries available around the perimeter of the district.
9. The district should continue to expand past all the broken boundaries, filling past them until a new boundary is encountered. It should fill in the same manner as before, growing rectangularly away from Point A. If a path becomes available for the district to grow back towards Point A it should continue to fill at the same rate as the other directions of district growth. This boundary, like those before it, should stop the growth of the district. Once the district is again completely enclosed, the Commission should return to step 7.
10. At this point the Commission should return the district to its last state where the total enclosed population was less than 0.99**m/n*. Then it should examine every region enclosed by boundaries adjacent to the potential district that is not already a part of a district and add them in various combinations to the district in question until the population of the district is within the range of 0.99**m/n* to 1.01**m/n*. If this cannot be accomplished, as a last resort the Commission should carefully add or subtract small, unbounded regions that are along the boundaries of the district until the population of the district is within the range of 0.99**m/n* to 1.01**m/n*.

11. The Commission at this point should establish a new Point A at some location outside of any already established district. It is recommended that this point be at a corner to minimize the ease of application. With this new Point A selected, the Commission should return to step 5. When all areas of the state are included within a district, the Commission has completed its assigned function.

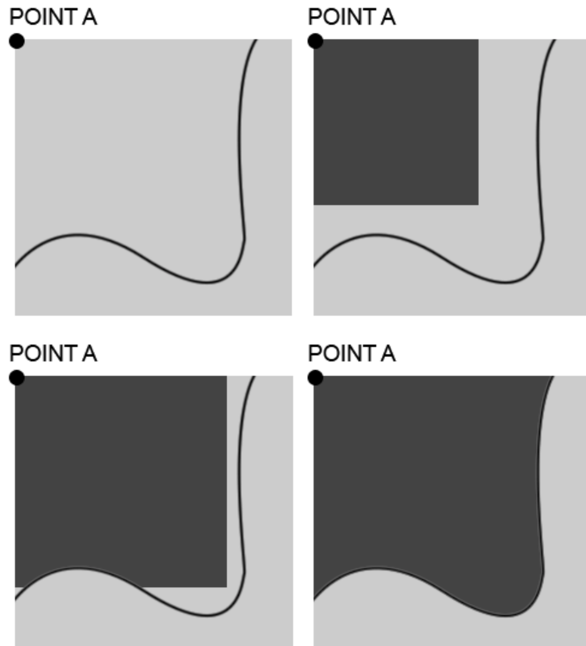

Figure 3. This diagram illustrates the procedure for filling in up to a boundary using this method of drawing districts. Imagine the line is a town border (Level 1), and note Point A in the corner. Imagine the Commission moving from the upper left diagram, to the upper right, to the lower left, and finally to the lower right. It is here that the boundaries would be "broken." Since all of these boundaries are Level 1, they would all be broken at the same time.

By drawing districts primarily based on rectangular patterns, the ratio of Area to Perimeter can be maximized in order to maximize compactness. Obviously, circles have the maximum ratio of Area to Perimeter, but circles are inconvenient for the purposes of dividing a state into regions that include all area within the state.

The key advantage of this type of algorithmic system for the drawing of Congressional districts is that it removes the power of a few individuals to manipulate the electoral process by drawing lines based on political persuasions. Instead, this system ensures that districts will be drawn in the most eminently fair way possible. However, it is important to note that algorithms such as this one have their limitations and cannot account for every possible district drawing situation. For that reason the Independent Commission is tasked with overseeing the execution of this algorithm and using their (theoretically apolitical) human judgment to adjust the computed districts into more practical and cartographically appropriate real-world districts.

Error Analysis/Testing

This section does not propose a mathematical model so much as propose an algorithm for drawing Congressional districts, which are not precisely quantitatively better in a particular shape or size, but rather qualitatively better in the eyes of the population as nondisenfranchising and fair. In order to test these qualities of each district, each district should have a population within the range of 0.99*m/n to 1.01*m/n. Each district must be contiguous. Finally, each district must be as compact as possible. If all these conditions are met for all districts, then the algorithm and Independent Commission have succeeded at redistricting the given state.

Conclusion

This analysis found that the census figures are affected by a significant undercount and that the unequal distribution of this undercounting between the states results in a major error in Congressional appropriations of funds and apportionment of seats in the House of Representatives. It proposes a second, follow-up sample study after the completion of the census to determine a correction ratio to adjust the census values to counter the undercounting error. This sample of zip code regions was stratified for both states and demographics (urban, suburban, rural). The new error of these adjusted population totals was determined to be smaller than the error of the original census totals. Additionally, a methodology was created to apportion seats in the House of Representatives more fairly on the basis of the population. This method was found to be superior not only to the current system but also to many other proposed systems of apportioning seats in the House of Representatives. Finally, this paper crafted a procedure that states should use to draw fair Congressional districts. These districts follow natural geographic boundaries as well as traditional political ones. All in all, this paper accomplishes the goal of "Making Sense of the 2010 Census."

Works Cited

Baker v. Carr, 186 (United States Supreme Court, 1962).

Bock, D. E., Velleman, P. F., & De Veaux, R. D. (2004). *Stats: Modeling the World.* New York: Pearson Addison Wesley.

Census of Population and Housing: 1790 Census. (2010, March 4). Retrieved March 6, 2010, from U.S. Census Bureau: http://www.census.gov/prod/www/abs/decennial/1790.html

Fair Vote. (2006, March 1). *Congressman Tanner's Fairness and Independence in Redistricting Act*. Retrieved March 6, 2010, from Fair Vote Program for Representative Government: http://archive.fairvote.org/?page=1114

Genealogy—The Evolution of the Census. (n.d.). Retrieved March 6, 2010, from Family Education: http://life.familyeducation.com/genealogy/family-tree/45360.html?page=3&detoured=1

Greenhouse, L. (2005, December 13). *Supreme Court to Hear Dispute on Texas Redistricting*. Retrieved March 6, 2010, from The New York Times: http://www.nytimes.com/2005/12/13/politics/13scotus.html

How It Affects the Nation. (n.d.). Retrieved March 6, 2010, from 2010 Census: http://2010.census.gov/2010census/why/index.php

Merriam-Webster. (2010). *Fair*. Retrieved March 6, 2010, from Merriam-Webster OnLine: http://www.merriam-webster.com/dictionary/fair

Merriam-Webster. (n.d.). *Gerrymandering*. Retrieved March 6, 2010, from Merriam-Webster OnLine: http://www.merriam-webster.com/dictionary/gerrymandering

Miller v. Johnson, 94-631 (Supreme Court of the United States, 1995).

Nagourney, A. (2005, February 7). *States See Growing Campaign for New Redistricting Laws*. Retrieved March 6, 2010, from The New York Times: http%3A%2F%2Fquery.nytimes.com%252fgst%252ffullpage%252ehtml%253fres%253d9400E4DA1E3BF934A35751C0A9639C8B63%2526sec%253d%2526spon%253d

PricewaterhouseCoopers. (2001). Retrieved March 6, 2010, from Section 5: Effect of Census 2000 Undercount on Federal Funding to States and Selected Counties, 2002–2012: http://govinfo.library.unt.edu/cmb/cmbp/reports/final_report/fin_sec5_effect.pdf

Purdue University. (2010). *Government Redistricting Websites*. Retrieved March 6, 2010, from Purdue Libraries: http://www.lib.purdue.edu/govdocs/redistricting.html

Sung, B. L. (1991, August). *U.S. Census*. Retrieved March 6, 2010, from Behavioral Causes of Census Undercount: New York City's Chinatown: http://www.census.gov/srd/papers/pdf/ev91-06.pdf

United States Census Bureau. (2009, December 03). *Census 2000 Urban and Rural Classification*. Retrieved March 6, 2010, from U.S. Census 2000: http://www.census.gov/geo/www/ua/ua_2k.html

University of Alabama. (2001, February 14). *Apportionment Methods*. Retrieved March 6, 2010, from Center for Teaching and Learning: http://www.ctl.ua.edu/math103/apportionment/appmeth.htm#Apportionment