Sampling Design:

Surveying High School Educators

Lucas Anthony

Aroona Atmaram

Ruchi Kumar

Professional Studies, Northwestern University

MSDS 402: Introduction to Data Science

Dr. John Derwent

May 24, 2020

Introduction:

We are a consulting firm assisting the state of New Jersey in determining how to restart schools in September 2020 while observing social distancing. Our goal is to determine which subjects matters are best suited for online learning. Data will be collected from the New Jersey Department of Education (DOE).

Target Population:

As we are supporting the state of New Jersey, the target population is the estimated 25,556 public high school teachers. We decided to draw a sample rather than attempt a census due to resource availability and the size of the sampling frame. The population is broken down by gender to verify random sampling is representative of overall population.

Region	Gender	Total	Region %	% of Total Pop
North	Female	6782	59%	
	Male	4668	41%	
	Total	**11450**		45%
Central	Female	5619	62%	
	Male	3467	38%	
	Total	**9086**		36%
South	Female	2950	59%	
	Male	2070	41%	
	Total	**5020**		19%

Sampling Frame:

Our sample frame is the list of all New Jersey public high school teachers from the 2018-2019 New Jersey DOE Certified Staff Record available on the New Jersey DOE website. We are not sampling independent or parochial school teachers as this is not in the scope of the engagement. The sample frame will be supplemented with relevant contact information such as

email addresses, phone numbers, and statistics on the number of educators by subject matter, which would be provided by the New Jersey DOE.

Sampling Methodology:

For this survey, we intend to use probability sampling to minimize bias. More specifically, the proportionate random stratified sampling method is chosen to have a representation from even the smallest subgroup of the population. Region served and subject area are the key stratification variables to maintain a balanced representation. This stratification was chosen to ensure schools of all sizes and all socio-economic levels are represented in the sample. Additionally, representation from educators of all subjects is necessary for survey success. Once the stratums are selected, a systematic random number generator will be used to select the sample itself. For the population of 25,556, sample size has been deduced as 2,996 given the margin of error to be 2% with a confidence level of 95%.

Calculating Sample Size:	
Total Sample size: $n = \frac{z^2 * p * (1-p)}{e^2}$	Where: $N =$ Size of the entire population $n =$ Size of the entire sample $e =$ Margin of error $z = z - \text{score}$
Proportionate Stratified Random Sampling: $n_h = \frac{N_h}{N} * n$	$N_h =$ Population size of the h^{th} stratum $n_h =$ Sample size of the h^{th} stratum $p =$ Population propotion

Sample:

The table below is the stratification for each region.

TOTALS:	Region	Subject	Teacher Count	% of Region	Invitees
Region Pop.	**North**	Math	2112	0.184343	996
11450		English	1889	0.164978	891
		Science	1842	0.160873	869
Region Sample Size		Language Arts	1003	0.085858	465
5392		History	1854	0.161919	874
		Technology (Programming)	1012	0.088383	477
Response Target		Arts	987	0.086201	466
1348		PE	751	0.065656	354

Totals:	Region	Subject	Teacher Count	% of Region	Invitees
Region Pop.	**Central**	Math	1245	0.137024	591
9086		English	1206	0.132732	573
		Science	1277	0.140546	606
Region Sample Size		Language Arts	952	0.104777	452
4315		History	1073	0.118094	510
		Technology (Programming)	1135	0.124918	539
Response Target		Arts	1409	0.155074	669
1079		PE	789	0.086837	375

Totals:	Region	Subject	Teacher Count	% of Region	Invitees
Region Pop.	South	Math	700	0.139442	318
5020		English	740	0.14741	335
		Science	653	0.130079	296
Region Sample Size		Language Arts	599	0.119322	272
2277		History	633	0.126096	287
		Technology (Programming)	601	0.119721	273
Response Target		Arts	567	0.112948	257
569		PE	527	0.10498	239

Survey Implementation:

The plan is to send an email with a link to a web survey, and one reminder email the next week. According to Madriaga, if we send an email on Monday morning, with a follow up reminder on the following Monday morning, we can reasonably expect a response rate of 25% (2017). Respondents are given the option to opt-out of this survey; however, as an additional incentive, a $5 Amazon gift card will be offered to all respondents who complete the survey. To achieve our break off point 2,996 respondents, 11,984 invites will be distributed. In this survey, a panel is not being utilized.

References:

Madariaga, L., Nussbaum, M., Burq, I., Marañón, F., Salazar, D., Maldonado, L., . . . Naranjo, M. (2017, April 03). Online survey: A national study with school principals. *Computers in Human Behavior*, vol. 74, 2017, pp. 35–44., doi:10.1016/j.chb.2017.03.067.