322 Study Design Project

Kathleen Zhang, Yura Heo

2025-10-20

Table 1: Stratified Sampling Plan Based on School Type and Size

type	size	Nh	nh
Public	Small	824	9
Public	Medium	861	10
Public	Large	176	2
Private	Small	3130	35
Private	Medium	381	4
Private	Large	23	1

We defined six strata by crossing school type (Public vs. Private, where Private combines non-profit and for-profit) with size (Small, Medium, Large). For each stratum, we constructed a sampling frame directly from the Scorecard search pages, then randomly sampled institutions within each stratum using proportional allocation. We recorded variables relevant to each research question—such as undergraduate enrollment, presence of a statistics major, median alumni earnings, and percentage of students receiving Pell Grants—and used appropriate Horvitz–Thompson estimators with stratum weights $w_h = N_h/n_h$ and finite population corrections N_h. Because one of the strata (Private–Large) contained only a single sampled institution, a lonely PSU adjustment was applied using the "adjust" method in the survey package. This method inflates the estimated variance slightly to account for strata with only one primary sampling unit (PSU), providing a more conservative and reliable confidence interval.

Stratifying by both sector and size is reasonable because these dimensions capture key structural differences among U.S. institutions that influence all measured outcomes—enrollment, program offerings, alumni earnings, and financial aid distribution. Public and private colleges differ substantially in student body size, tuition levels, and available programs, while grouping schools by size ensures that institutions with similar characteristics are compared within the same category. This stratification reduces variability within strata and captures meaningful differences across them, improving the precision and representativeness of all estimates compared to a simple random sample or clster sample.

1.

What is an estimate of the total number of undergraduate students enrolled in colleges? Design your survey with the goal of answering this question most accurately.

```
                        total       SE
undergrad_enrollment 15598354 1786576


                         2.5 %    97.5 %
undergrad_enrollment 12096730 19099978
```

The Horvitz–Thompson estimator produced an estimated total undergraduate enrollment of 15,598,354 students, with a standard error (SE) of approximately 1,786,576. The 95% confidence interval for the total enrollment ranges from 12,096,730 to 19,099,978 students.

2.

What fraction of colleges have a major in statistical science? Use your sample to estimate this fraction; do not get the answer by filtering the schools.

```
                    mean      SE
statistics_major 0.10411 0.0295


                      2.5 %     97.5 %
statistics_major 0.04621349 0.1620163
```

To estimate the fraction of colleges in the United States that offer a major in statistical science, we used the same stratified simple random sample, believing that public vs. private colleges, and small vs. large colleges may offer very different likelihoods of offering a statistics major while having similar academic offerings within strata.

After confirming that the N_h for each stratum were about the same as Question 1, we used the same random sample of schools from each stratum, recording whether each sampled school offered a major in statistical science (1 if yes, 0 if no). Again, we used a Horvitz–Thompson estimator for the population mean with stratum weights $w_h = N_h/n_h$ and finite population corrections N_h, also adjusting for the Private–Large strata that contained only a single sampled institution.

The Horvitz–Thompson estimator produced an estimated 10.4% of colleges in the United States that offer a statistics major, with a standard error (SE) of approximately 2.95%. The 95% confidence interval for the total number of colleges offering a statistics major ranges from 4.62% to 16.2%. This makes sense, as statistics is still more of a specialized field compared to broader majors such as mathematics or computer science. Many community colleges or smaller schools may not have the demand to support a statistics department, but larger universities are more likely to offer one.

3.

What is the average of the median earnings among alumni from public schools? From private schools (combine for profit and not for profit private schools)?

```
          type  mean    SE  CI_l  CI_u
Private Private 44489 3438 37750 51228
Public   Public 45650 2387 40971 50329
```

To estimate the average median earnings for public and private schools, we used a stratified simple random sample with proportional allocation across six strata formed by crossing school type (public vs. private) with institution size (small, medium, large). Specifically, public institutions were divided into three strata—Public–Small, Public–Medium, and Public–Large—and private institutions were divided into Private–Small, Private–Medium, and Private–Large. Each stratum contributed to its respective type's overall estimate in proportion to its population size.

The Horvitz–Thompson estimator produced an estimated average median earning of $45,650 among alumni from public schools, with a standard error (SE) of approximately $2,387. The corresponding 95% confidence interval ranges from $40,971 to $50,329.

For private schools, the estimated average median earning was $44,489, with a standard error (SE) of approximately $3,438. The 95% confidence interval for private-school alumni median earnings ranges from $37,750 to $51,228.

Given their overlapping confidence intervals between public and private schools, there is no statistically significant difference in typical alumni earnings between the two groups. While public-school alumni appear to earn slightly more on average, the uncertainty in the estimates suggests that overall earnings outcomes are comparable across institution types.

4.

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

```
      mean     SE
pell 0.417 0.0209


       2.5 %  97.5 %
pell 37.60892 45.7915


    pell
41.70021
```

To answer an additional question with our stratified sample, we were curious about estimating the mean percentage of undergraduate students who receive Pell grants across U.S. institutions. This is because we believe that Pell grant rates differ systematically across size and type of school- public schools typically enroll more lower-income students, leading to higher Pell grant recipients within said school, while private large universities and elite institutions may have fewer recipients. Our stratified design could give a more precise estimate of Pell recipients.

The data for the percentage of Pell grant recipients for each school in our sample was pulled from the csv on the website titled "Most Recent Institution-Level Data" and was already coded as a decimal. We used the same stratum weights and fpc as in previous questions.

The Horvitz–Thompson estimator produced an estimated mean Pell recipient percentage of 41.7%, with a standard error (SE) of approximately 2.09%. The 95% confidence interval is 37.61% to 45.79%, indicating that, on average, about two in five undergraduates at U.S. colleges in our frame receive Pell grants under this stratified design. This result reflects the national importance of federal aid in college access, and our design was able to capture the differences in socioeconomic composition across school types.