

Name: _____

Date: _____

Stratified sampling

1. The following represents survey results in a large school:

Grade	$\hat{p}$	Demographic weight
Freshman	0.67	522
Sophomores	0.56	478
Juniors	0.48	382
Seniors	0.39	303

Figure out the overall survey $\hat{p}$ for the entire school.

2. A survey of Californians oversampled all minority groups on purpose to calculate decent margins of error for subpopulations. Here are the results:

Ethnicity	n	Yes votes	$\hat{p}$	Confidence interval
White (non-Latino)	1087	702		
Latino	989	506		
African American	893	517		
Asian American	1112	855		
Native American	482	316		

- A. Calculate the $\hat{p}$ and the confidence interval for each subpopulation. Fill those results into the table above.
- B. What is the overall survey $\hat{p}$ that should be reported out if the true demographic weights in California are White, non-Latino (37%); African American (6%); Asian American (15%); Native Americans (1%); Latinos (41%)?

3. A survey asked people about the number of pets they owned. A stratified sampling method was used to account for age:

Age	n	$\bar{x}$	s	S.E.	Weight
18-30	180	0.4	0.05		25%
30-45	225	0.9	0.10		30%
45-60	241	1.1	0.09		30%
60-75	209	1.4	0.07		10%
75 and up	148	0.2	0.05		5%

- A. Calculate the standard error for each subgroup. Put the results in the table above.
- B. What is the overall $\bar{x}$ that the survey should report out?
- C. What is the confidence interval for the whole sample they should report out?

The formula for standard error for the stratified sampling requires knowing the subpopulation's (or strata's) demographic weights and standard errors. In this case, the demographic weight w for 18–30-year-olds is 0.25, the weight w for 30–45-year-olds is 0.30.... And we already can figure out their standard errors. For example,

$$SE_{18-30} = \frac{0.05}{\sqrt{180}} \approx 0.0037 \text{ and } SE_{30-45} = \frac{0.10}{\sqrt{225}} \approx 0.0067$$

The general formula is¹

$$SE_{strat} = \sqrt{(w_1 \cdot SE_1)^2 + (w_2 \cdot SE_2)^2 + \dots}$$

for as many strata as there are. So for our example,

$$SE_{strat} = \sqrt{(0.25 \cdot 0.0037)^2 + (0.30 \cdot 0.0067)^2 + \dots}$$

Figure out the standard error for the stratified sample.

- D. Figure out the confidence interval for the stratified sample's $\bar{x}$.

¹ <http://www.stat.cmu.edu/~fienberg/Stat36-303-03/Handouts/StratificationNotes-03.pdf>

