

Homework 7

Problems 1-3 due Fri, Nov 6 by midnight. Grad problem 4 due Mon, Nov. 9 by midnight. Show your reasoning on all problems. Neatly handwritten is fine.

1. This problem relies on results in Chapters 14-15. An insurance company has sold 10,000 life insurance policies on 45-year old women. Each of the policies pays \$100,000 if the woman dies in the next year. The yearly premium charged by the company for each woman is \$250. Use the table of mortality rates by sex and age for 2017 at <http://www.ssa.gov/OACT/STATS/table4c6.html>. If you write a script in R to carry out the computations (using n to denote the number of policies sold), then you can simply change n and rerun the script to answer part (e). (You could also use variables to denote the premium amount and the payout amount. Then you'd be an actuary.)
 - (a) Let X be the number of these women who die in the next year. Assuming that whether these women live or die in the next year are like independent Bernoulli trials, what is the probability distribution of X ? What are $E(X)$ and $SD(X)$?
 - (b) Write an expression for the profit, say P , that the company will earn on these women in terms of X . Compute the expected value and standard deviation of P . Remember that all the women, even those who die, have paid the premium.
 - (c) Argue that the distribution of P is approximately normal.
 - (d) There's approximately a 95% probability that the company's profit on these policies in the next year will be between what two values?
 - (e) What would your answer to part (d) be if the company sold 100,000 policies like this (you can just show the answer, without showing all the intermediate calculations)? How does show why insurance companies are usually really big?
2. Problem R.5.4, p. 598.
3. Problem R.5.14, p. 599.
4. Graduate credit problem. Please submit separately. The purpose of this exercise is to compare estimation of a population proportion using an SRS to using a stratified random sample and to show how you can use the results you learned in Chapters 15-17 to derive a confidence interval for a population proportion from a stratified random sample.

Suppose that support for marijuana legalization among U.S. adults varies by age group as shown in the table below. The second column shows the proportion of all U.S. adults in each age group. The third column represents the assumed true level of support among all adults in the three age groups.

Age group (years)	Proportion of adult population	Proportion who support legalization
18-34	0.2	0.8
35-64	0.5	0.6
65+	0.3	0.3

- (a) Given this information, what proportion of all U.S. adults favor legalization? You can also think of it this way: if I randomly choose a person from this population, what's the probability that they support legalization?
- (b) Suppose I plan to take a simple random sample of 1000 U.S. adults to compute the sample proportion \hat{p} of those who favor legalization. What is the approximate sampling distribution of \hat{p} ? In particular, what are $E(\hat{p})$ and $SD(\hat{p})$?
- (c) An alternative sampling plan is a stratified random sample of 1000 voters with what is called "proportional allocation" where the sample size in each age group reflects the proportion of the population in that age group, that is, sample sizes 200,500,300. Let \hat{p}_1 , \hat{p}_2 , and \hat{p}_3 be the sample proportions of legalization supporters for the three groups. A natural estimate of the proportion of all adults who favor legalization from the stratified random sample is $\hat{p}_c = 0.2\hat{p}_1 + 0.5\hat{p}_2 + 0.3\hat{p}_3$ (the subscript "c" stands for "combined"). What is the sampling distribution of \hat{p}_c ? In particular, what are $E(\hat{p}_c)$ and $SD(\hat{p}_c)$? Hint: first calculate the expected value and variance of \hat{p}_1 , \hat{p}_2 , and \hat{p}_3 , then use the fact that \hat{p}_c is a linear combination of these random variables and use the methods of Chapter 14 to calculate its expected value and variance.
- (d) Which sampling plan gives a more accurate estimate of the population proportion?
- (e) The above results assumed you knew the population proportions. Suppose you don't know the population proportions and you carry out the stratified sampling plan. You find that in your sample, 84% of the adults age 18-34, 62% of the adults age 35-64, and 38% of the adults 65 and over support legalization. Use your result in part (c) to calculate $SE(\hat{p}_c)$. Then calculate a 95% confidence interval for the proportion of all adults who favor legalization.