

This problem set covers material from Week 9, dates 4/14 – 4/17.

Instructions: Write or type complete solutions to the following problems and submit answers to the corresponding Canvas assignment. Your solutions should be neatly-written, show all work and computations, include figures or graphs where appropriate, and include some written explanation of your method or process (enough that I can understand your reasoning without having to guess or make assumptions). A general rubric for homework problems appears on the final page of this assignment.

In some the following, you may need to use R. If you do, please write down the corresponding code to “show your work”. Drawing and labelling curves are also good examples of “showing your work”.

Monday 4/14

1. The standard deviation for students at particular Ivy League college is 250 points. Two students, Raina and Luke, want to estimate the average SAT score of students at this college based on a sample of some students. They want their margin of error to be no more than 25 points. For this part, assume that conditions for CLT-based inference are satisfied.
 - (a) Raina wants to use a 90% confidence level. How large a sample does Raina need to collect?
 - (b) Luke wants to use a 95% confidence level. Without calculations, determine whether Luke’s sample should be larger or smaller than Raina’s. Explain your reasoning.
2. Researchers interested in lead exposure due to car exhaust sampled the blood of 28 traffic controllers subjected to constant inhalation of automobile exhaust fumes while working traffic enforcement in a primarily urban environment. The blood samples of these controllers had an average lead concentration of 54.32 microgramm/l, and a standard deviation of 37.74 microgramm/l. A previous study of individuals from nearby suburbs, with no history of exposure, found an average blood level concentration of 35 microgramm/l.
 - (a) Write down the hypotheses that would be appropriate for testing if the traffic controllers appear to have been exposed to a higher concentration of lead than expected.
 - (b) Explicitly state and check all conditions necessary for CLT-based inference on these data. If you do not have enough information, state what would need to be true in order for the conditions to be met and if you believe that the conditions would be met given the context.
 - (c) Regardless of your answer in part (b), finish conducting the test of the hypotheses you defined in (a). Don’t forget to make a conclusion in context.

- (d) Regardless of your answer in part (b), obtain a 90% CI for the true average blood levels of these traffic controllers (no need to interpret).

Wednesday 4/16

3. *Lymphatic filariasis* is a disease caused by a parasitic worm. Complications of the disease can lead to extreme swelling and other complications. Here we consider results from a randomized experiment that compared three different drug treatment options to clear people of the parasite. The results for the second year of the study are given below:

group	Outcome		Total
	Clear at Year 2	Not Clear at Year 2	
Three drugs	52	2	54
Two drugs	31	24	55
Two drugs annually	42	14	56
Total	125	40	165

Conduct a hypothesis test at the 0.05 level to evaluate whether there is any difference in the performance of the treatments. Be sure to make a conclusion in context.

Thursday 4/17

4. Each year the US Environmental Protection Agency (EPA) releases fuel economy data on cars manufactured in that year. Below are summary statistics on fuel efficiency (in miles/gallon) from random samples of cars with manual and automatic transmissions, as well as the summary statistics of the differences. The larger the miles per gallon, the more fuel efficient the vehicle. Do these data provide strong evidence that manual transmission cars have a better average fuel efficiency than automatic transmission cars? Assume that conditions for CLT-based inference are satisfied.

	City miles per gallon	
	Automatic	Manual
Mean	15.26	19.64
SD	4.38	4.70
n	26	23

5. According to a report on sleep deprivation by the Centers for Disease Control and Prevention, the proportion of California residents who reported insufficient rest or sleep during each of the preceding 30 days is 7.7%, while this proportion is 9% for Oregon

residents. These data are based on simple random samples of 11,545 California and 4,691 Oregon residents.

Conduct a CLT-based hypothesis test at the 0.01-level to determine if these data provide strong evidence that the rate of sleep deprivation is different for the two states.

6. Using the same data as in Problem 5, obtain a CLT-based 95% confidence interval for the difference in the rate of sleep deprivation for the two states. Based on your interval, does it appear to be that one state has a higher rate of sleep deprivation than the other? If so, which one?

General rubric

Points	Criteria
5	The solution is correct <i>and</i> well-written. The author leaves no doubt as to why the solution is valid.
4.5	The solution is well-written, and is correct except for some minor arithmetic or calculation mistake.
4	The solution is technically correct, but author has omitted some key justification for why the solution is valid. Alternatively, the solution is well-written, but is missing a small, but essential component.
3	The solution is well-written, but either overlooks a significant component of the problem or makes a significant mistake. Alternatively, in a multi-part problem, a majority of the solutions are correct and well-written, but one part is missing or is significantly incorrect.
2	The solution is either correct but not adequately written, or it is adequately written but overlooks a significant component of the problem or makes a significant mistake.
1	The solution is rudimentary, but contains some relevant ideas. Alternatively, the solution briefly indicates the correct answer, but provides no further justification.
0	Either the solution is missing entirely, or the author makes no non-trivial progress toward a solution (i.e. just writes the statement of the problem and/or restates given information).
Notes:	For problems with multiple parts, the score represents a holistic review of the entire problem. Additionally, half-points may be used if the solution falls between two point values above.
Notes:	For problems with code, well-written means only having lines of code that are necessary to solving the problem, as well as presenting the solution for the reader to easily see. It might also be worth adding comments to your code.