CSCI 145 Problem 1: The Correlation Floor

August 27, 2026

Submission Instructions

Please upload your work to Gradescope by **11:59pm on DATE**.

- Turn in a single PDF of *handwritten* solutions. Typed (including LaTeX) solutions will not be accepted.
- Problems are grouped into three-week **cycles**. Once per cycle, you will meet with the grader and present the solution to *one* problem from that cycle, selected at random when you arrive — so bring your work and be ready to explain any of them.

Problem 1: The Correlation Floor 2 points

A poll asks $n = 10{,}000$ people a yes/no question and reports $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, where every response has variance $\mathrm{Var}(X_i) = \sigma^2$. The usual margin-of-error calculation treats the responses as independent, but suppose instead that every distinct pair has the same small correlation ρ, so $\mathrm{Cov}(X_i, X_j) = \rho\sigma^2$ for $i \neq j$.

Give a mathematical audit of the claim that "a large enough poll can make sampling error arbitrarily small." Your audit should derive the exact variance of $\bar{X}$ from covariance bilinearity, identify what remains as $n \to \infty$, and quantify the effect for $\sigma = \frac{1}{2}$: determine the value of ρ for which a poll of 10,000 has standard deviation 0.03. Finally, compute the size n_{ind} of an independent poll having the same variance as an equicorrelated poll with parameters (n, ρ), and use your formula to explain whether collecting more responses from the same population can remove the three-point uncertainty.

Results you may use. Covariance is bilinear, $\mathrm{Var}(Z) = \mathrm{Cov}(Z, Z)$, $\mathrm{Var}(aZ) = a^2\,\mathrm{Var}(Z)$, and correlation is covariance divided by the two standard deviations.