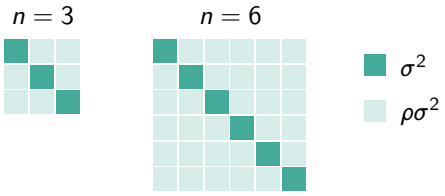
How Much Can a Larger Poll Reduce Uncertainty?

Due: Wednesday, September 9 at 11:59pm

Instructions: Please submit your *handwritten* solutions as a single PDF on Gradescope.

A poll asks n people a yes/no question and reports $\bar{X} = n^{-1}\sum_{i=1}^{n} X_i$. Every response has variance $\mathrm{Var}(X_i) = \sigma^2 > 0$. People exposed to the same news or recruited from the same social group may respond alike, so suppose every distinct pair has common correlation ρ, where $0 \le \rho \le 1$.

An analyst claims that surveying enough people will make $\mathrm{Var}(\bar{X})$ arbitrarily small, even if the recruitment process is unchanged. The diagrams show the covariance of each pair of responses, with one row and one column per respondent.



1. The poll averages n correlated responses. What is the exact variance of $\bar{X}$?
2. The recruitment process stays fixed as the poll grows. Find $\lim_{n\to\infty} \mathrm{Var}(\bar{X})$ for fixed ρ.
3. The analyst claims that enough responses will make the variance arbitrarily small. For which values of ρ is this claim true?

Hint: You may use $\mathrm{Var}(aZ) = a^2\,\mathrm{Var}(Z)$ and

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} \mathrm{Cov}(X_i, X_j).$$

Takeaway: *With a fixed positive correlation between responses, the uncertainty approaches a positive floor as the poll grows. Reducing that floor requires a sampling process with less dependence between responses.*