Today

- Application 2: Intelligent Searches and Sampling
- Application 3: Random Sums

Intelligent Searching and Sampling

a) Group Testing: A large number $n$ of blood samples are to be tested for a relatively rare disease. Can we find all the infected samples in fewer than $n$ tests?
Consider the case of splitting each of \( n \) samples in half. Combine half of each one is placed into a large combined pool.

Should this work better?

Now consider that we divide the \( m \) samples into \( m \) groups of size \( k \) each...

b) Stratified Sampling

Imagine that a population is naturally divided into \( n \) groups or strata.

What happens if you randomly sample from each stratum separately than it is to take a single random sampling?
How can we get an unbiased estimate of the population mean based on the separate strata means?

What is the variance of $\bar{y}_{strata}$?

When is stratified sampling better?
Random Sums
An insurance company receives \( N \) independent claims \( X_1, \ldots, X_N \) in a given time period. Where \( N \) is also a random variable (independent of the \( X_i \)).

What are the mean and variance of
\[
T = \sum_{i=1}^{N} X_i
\]

This would be much easier to work with if we could condition on \( N \) and consider \( T | N \).

\[
E(T | N = n) = E\left( \sum_{i=1}^{N} X_i | N = n \right)
\]
\[
= E\left( \sum_{i=1}^{n} X_i \right)
\]
\[
= \sum_{i=1}^{n} E(X_i) = nE(X)
\]

But we somehow need to take the expectation over \( N \) as well.
The general result is

\[ E(Y) = E_X[E_{Y|X}(Y|X)] \]

A similar result is

\[ \text{Var}(Y) = \text{Var}[E(Y|X)] + E[\text{Var}(Y|X)] \]