1. *Multistage Cluster Sampling* is a sample design technique where individuals are first sorted into groups (i.e. clusters) and sample is obtained first by randomly selecting a cluster, and then selecting a certain number of individuals from the chosen cluster.

   Suppose we are testing for the prevalence of a disease by first choosing a city in the U.S. and then sampling n individuals from that city. Let Q be the proportion of people in the chosen city with the disease, and let X be the number of individuals in the sample with the disease. Since difference cities may have very different disease prevalence, then both Q and X are random variables. Suppose $Q \sim \text{Unif}(0, 1)$, and that each individual in the sample independently has probability Q of having the disease.

   (a) Briefly explain why $Q \sim \text{Unif}(0, 1)$ is a reasonable, conservative guess for the for disease prevalence among a randomly selected city, assuming that you have no prior information before collecting a sample.

   (b) Briefly explain why it is reasonable to say that each individual in the sample independently has probability Q of having the disease, assuming sampling is performed with replacement, or if the city population is very large relative to the sample size.

   (c) What is the conditional distribution of $X|Q = q$?

   (d) Find $E[X|Q]$.

   (e) Find $\text{Var}(X|Q)$.

   (f) Compute $E[X]$ and $\text{Var}(X)$. *Hint: Use Law of Total Expectation, and Eve's Law.*