Tutorial 4:

Introduction to Data Science

Sampling and Probability

Ernst C. Wit

October 7, 2024

Exercises

1. **Mark and Recapture Sampling (Equal Probabilities).**
   Context: In ecology, the *mark and recapture* method is used to estimate the population size of animals.

   Procedure:

   - A sample of animals is captured, marked, and then released back into the wild.
   - After some time, another sample is captured, and the number of marked animals in the recaptured group is counted.

   Example: Consider a population of $N = 500$ animals in a forest. A researcher captures and marks $M = 100$ animals and then releases them. After a period, the researcher recaptures a random sample of $n = 50$ animals.

   Assumptions:

   - All animals have an equal probability of being recaptured.
   - The population remains closed (no animals are added or removed during the period).

   Let X be the number of marked animals recaptured. This can be modeled as a **binomial distribution**:

   $$X \sim \text{Binomial}(n = 50, p = \frac{M}{N})$$

   where $p = \frac{M}{N}$ is the probability of capturing a marked animal in the recapture phase.

   Tasks:

   (a) What is the expected number of marked animals in the recaptured sample, $E(X)$?

   (b) Calculate the probability that exactly 10 marked animals are recaptured.
   Use the following binomial probability formula for your calculations:

   $$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

   where $p = \frac{M}{N}$.

   (c) Use Python or R to calculate the probability that k marked animals are recaptured and represent this in a histogram.

   (d) Use Python or R to create the population with 500 individuals of which 100 hundred are marked. Then

- sample 1000 times 50 animals at random;
- for each sample store the number of marked animals x_i that are recaptured ($i = 1, \ldots, 1000$).
- make a histogram of all values $x = (x_1, \ldots, x_{1000})$.

Compare your plot to the plot of the previous exercise.

(e) The point of mark-recapture experiments is to estimate the total number of animals N. If you recaptured 12 marked animals in a sample of 50, then how would you estimate N?

(f) For your simulation in (e) take all the sampled values x_i of recaptured marked animals ($i = 1, \ldots, 1000$) and use the answer to your previous question to transform this into estimates of M. Show your results in a histogram.

2. **Mark and Recapture Sampling (Unequal Probabilities)**
Context: The mark and recapture method can be extended to account for different probabilities of recapturing marked animals. In fact, it has been argued (Tilling K, Sterne JAC. Capture-recapture models including covariate effects. Am J Epidemiology, 1999;149:392–400.) that originally marked animals tend to be animals that in general are easier to capture (e.g. because they are more familiar with humans, they are not as fast to get away, etc.).

Modified Procedure:

- As in Question 1, $M = 100$ animals are marked and released into a population of $N = 500$.
- In the recapture phase, the probability of capturing a (specific) marked animal is **50% higher** than the probability of capturing an unmarked individual.

Modeling the Recapture:

- Let p_m be the probability of capturing a marked animal.
- The sample size in the recapture phase is $n = 50$, and X is the number of marked animals recaptured.

Tasks:

(a) What is the probability p_m of capturing a marked animal?

(b) What is the new distribution of the number of recaptured marked animals X?

(c) What is the new expected number of marked animals recaptured, $E(X)$?

(d) If you did not know N, then how would you write your answer to question (c)?

(e) If you observe $X = 12$ marked animals in your recapture experiment, then use the formula in (d) to estimate N.

(f) Use Python or R to simulate 1000 recapture experiments with $N = 500$, and use the sampled values to calculate 1000 estimates of N. Show the results in a histogram.