

4.5: Hypergeometric Distribution

The hypergeometric distribution arises when one samples from a finite population, thus making the trials dependent on each other. There are five characteristics of a hypergeometric experiment.

Characteristics of a hypergeometric experiment

1. You take samples from **two** groups.
2. You are concerned with a group of interest, called the first group.
3. You sample **without replacement** from the combined groups. For example, you want to choose a softball team from a combined group of 11 men and 13 women. The team consists of ten players.
4. Each pick is **not** independent, since sampling is without replacement. In the softball example, the probability of picking a woman first is $\frac{13}{24}$. The probability of picking a man second is $\frac{11}{23}$ if a woman was picked first. It is $\frac{10}{23}$ if a man was picked first. The probability of the second pick depends on what happened in the first pick.
5. You are **not** dealing with [Bernoulli Trials](#).

The outcomes of a hypergeometric experiment fit a *hypergeometric probability* distribution. The random variable X = the number of items from the group of interest.

Example 4.5.1

A candy dish contains 100 jelly beans and 80 gumdrops. Fifty candies are picked at random. What is the probability that 35 of the 50 are gumdrops? The two groups are jelly beans and gumdrops. Since the probability question asks for the probability of picking gumdrops, the group of interest (first group) is gumdrops. The size of the group of interest (first group) is 80. The size of the second group is 100. The size of the sample is 50 (jelly beans or gumdrops). Let X = the number of gumdrops in the sample of 50. X takes on the values $x = 0, 1, 2, \dots, 50$. What is the probability statement written mathematically?

Answer

$$P(x = 35)$$

Exercise 4.5.1

A bag contains letter tiles. Forty-four of the tiles are vowels, and 56 are consonants. Seven tiles are picked at random. You want to know the probability that four of the seven tiles are vowels. What is the group of interest, the size of the group of interest, and the size of the sample?

Answer

The group of interest is the vowel letter tiles. The size of the group of interest is 44. The size of the sample is seven.

Example 4.5.2

Suppose a shipment of 100 DVD players is known to have ten defective players. An inspector randomly chooses 12 for inspection. He is interested in determining the probability that, among the 12 players, at most two are defective. The two groups are the 90 non-defective DVD players and the 10 defective DVD players. The group of interest (first group) is the defective group because the probability question asks for the probability of at most two defective DVD players. The size of the sample is 12 DVD players. (They may be non-defective or defective.) Let X = the number of defective DVD players in the sample of 12. X takes on the values $0, 1, 2, \dots, 10$. X may not take on the values 11 or 12. The sample size is 12, but there are only 10 defective DVD players. Write the probability statement mathematically.

Answer

$$P(x \leq 2)$$

? Exercise 4.5.2

A gross of eggs contains 144 eggs. A particular gross is known to have 12 cracked eggs. An inspector randomly chooses 15 for inspection. She wants to know the probability that, among the 15, at most three are cracked. What is X , and what values does it take on?

Answer

Let X = the number of cracked eggs in the sample of 15. X takes on the values $0, 1, 2, \dots, 12$

✓ Example 4.5.3

You are president of an on-campus special events organization. You need a committee of seven students to plan a special birthday party for the president of the college. Your organization consists of 18 women and 15 men. You are interested in the number of men on your committee. If the members of the committee are randomly selected, what is the probability that your committee has more than four men?

This is a hypergeometric problem because you are choosing your committee from two groups (men and women).

- Are you choosing with or without replacement?
- What is the group of interest?
- How many are in the group of interest?
- How many are in the other group?
- Let X = _____ on the committee. What values does X take on?
- The probability question is $P(\text{_____})$.

Solution

- without
- the men
- 15 men
- 18 women
- Let X = the number of men on the committee. $x = 0, 1, 2, \dots, 7$.
- $P(x > 4)$

? Exercise 4.5.3

A palette has 200 milk cartons. Of the 200 cartons, it is known that ten of them have leaked and cannot be sold. A stock clerk randomly chooses 18 for inspection. He wants to know the probability that among the 18, no more than two are leaking. Give five reasons why this is a hypergeometric problem.

Answer

- There are two groups.
- You are concerned with a group of interest.
- You sample without replacement.
- Each pick is not independent.
- You are not dealing with Bernoulli Trials.

📌 Notation for the Hypergeometric: H = Hypergeometric Probability Distribution Function

$$X \sim H(r, b, n) \quad (4.5.1)$$

Read this as " X is a random variable with a hypergeometric distribution." The parameters are r, b , and n ; r = the size of the group of interest (first group), b = the size of the second group, n = the size of the chosen sample.

✓ Example 4.5.4

A school site committee is to be chosen randomly from six men and five women. If the committee consists of four members chosen randomly, what is the probability that two of them are men? How many men do you expect to be on the committee?

Let X = the number of men on the committee of four. The men are the group of interest (first group).

X takes on the values 0, 1, 2, 3, 4 where $r = 6$, $b = 5$, and $n = 4$. $X \sim H(6, 5, 4)$

Find $P(x = 2)$. $P(x = 2) = 0.4545$ (calculator or computer)

Currently, the TI-83+ and TI-84 do not have hypergeometric probability functions. There are a number of computer packages, including Microsoft Excel, that do.

The probability that there are two men on the committee is about 0.45.

The graph of $X \sim H(6, 5, 4)$ is:

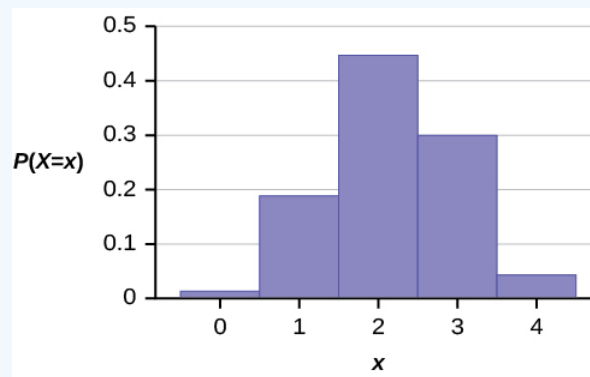


Figure 4.5.1.

The y-axis contains the probability of X , where X = the number of men on the committee.

You would expect $m = 2.18$ (about two) men on the committee.

The formula for the mean is

$$\mu = \frac{nr}{r+b} = \frac{(4)(6)}{6+5} = 2.18 \quad (4.5.2)$$

? Exercise 4.5.4

An intramural basketball team is to be chosen randomly from 15 boys and 12 girls. The team has ten slots. You want to know the probability that eight of the players will be boys. What is the group of interest and the sample?

Answer

The group of interest is the 15 boys. The sample consists of the ten slots on the intramural basketball team.

Summary

A hypergeometric experiment is a statistical experiment with the following properties:

- You take samples from two groups.
- You are concerned with a group of interest, called the first group.
- You sample without replacement from the combined groups.
- Each pick is not independent, since sampling is without replacement.
- You are not dealing with Bernoulli Trials.

The outcomes of a hypergeometric experiment fit a hypergeometric probability distribution. The random variable X = the number of items from the group of interest. The distribution of X is denoted $X \sim H(r, b, n)$, where r = the size of the group of interest

(first group), b = the size of the second group, and n = the size of the chosen sample. It follows that $n \leq r + b$. The mean of X is $\mu = \frac{nr}{r+b}$ and the standard deviation is $\sigma = \sqrt{\frac{rbn(r+b-n)}{(r+b)^2(r+b-1)}}$.

Formula Review

$X \sim H(r, b, n)$ means that the discrete random variable X has a hypergeometric probability distribution with r = the size of the group of interest (first group), b = the size of the second group, and n = the size of the chosen sample.

X = the number of items from the group of interest that are in the chosen sample, and X may take on the values $x = 0, 1, \dots$, up to the size of the group of interest. (The minimum value for X may be larger than zero in some instances.)

$$n \leq r + b$$

The mean of X is given by the formula $\mu = \frac{nr}{r+b}$ and the standard deviation is $= \sqrt{\frac{rbn(r+b-n)}{(r+b)^2(r+b-1)}}$.

Use the following information to answer the next five exercises: Suppose that a group of statistics students is divided into two groups: business majors and non-business majors. There are 16 business majors in the group and seven non-business majors in the group. A random sample of nine students is taken. We are interested in the number of business majors in the sample.

? Exercise 4.5.5

In words, define the random variable X .

Answer

X = the number of business majors in the sample.

? Exercise 4.5.6

$X \sim \underline{\hspace{1cm}}(\underline{\hspace{1cm}}, \underline{\hspace{1cm}})$

? Exercise 4.5.7

What values does X take on?

Answer

2, 3, 4, 5, 6, 7, 8, 9

? Exercise 4.5.8

Find the standard deviation.

? Exercise 4.5.9

On average (μ), how many would you expect to be business majors?

Answer

6.26

Glossary

Hypergeometric Experiment

a statistical experiment with the following properties:

1. You take samples from two groups.
2. You are concerned with a group of interest, called the first group.
3. You sample without replacement from the combined groups.

4. Each pick is not independent, since sampling is without replacement.
5. You are not dealing with Bernoulli Trials.

Hypergeometric Probability

a discrete random variable (RV) that is characterized by:

1. A fixed number of trials.
2. The probability of success is not the same from trial to trial.

We sample from two groups of items when we are interested in only one group. X is defined as the number of successes out of the total number of items chosen. Notation: $X \sim H(r, b, n)$, where r = the number of items in the group of interest, b = the number of items in the group not of interest, and n = the number of items chosen.

This page titled [4.5: Hypergeometric Distribution](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [OpenStax](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.