Hypergeometric Distributions

Lesson objectives

- I can recognise conditions that give rise to a hypergeometric distribution
- I can calculate the probability associated with each random variable of a hypergeometric distribution
- I can represent the hypergeometric distribution using a table and a probability histogram
- I can solve problems involving hypergeometric distributions

1.1

Lesson objectives | Teachers' notes | Lesson notes

MHR Page 178 #s 1 - 7, 9 & 10

Warm up

The binomial distribution involves independent trials. This section develops a distribution involving dependent trials. Cutting cards with a standard deck provides independent trials, whereas dealing the cards involves dependent trials. Similarly, selecting a jury pool and catching and tagging animals for scientific research involve dependent trials. Brainstorm other examples of dependent trials.

Definitions

Hypergeometric Probability Distributions

- A distribution with dependent trials whose outcomes are either success or failure
- The random variable is the number of successes in a given number of trials

Example 1

Hypergeometric Probability

A committee of six people is to be formed from a pool of six grade 11 students and seven grade 12 students. Determine the probability that the committee will have two grade 11 students.

We want ${}_6C_2$ Grade 11s
and also want ${}_7C_4$ Grade 12s
Total combinations is ${}_{13}C_6$

$$P(2 \text{ Grade 11s}) = \frac{{}_6C_2 \times {}_7C_4}{{}_{13}C_6}$$

$$= \frac{15 \times 35}{1716}$$

$$= 0.3059$$

The probability of having exactly two Grade 11s on the committee is about 30.59%

Your Turn

On a team of 15 astronauts, six are women and nine are men. If four astronauts are selected at random for a flight simulation, what is the probability that two men and two women are selected?

We want ${}_9C_2$ men
and we also want ${}_6C_2$ women
Total combinations is ${}_{15}C_4$

$$P(2 \text{ men and } 2 \text{ women}) = \frac{{}_9C_2 \times {}_6C_2}{{}_{15}C_4}$$

$$= \frac{36 \times 15}{1365}$$

$$\approx 0.3956$$

The probability of having 2 men and 2 women for the simulation is about 39.56%

Probability in a Hypergeometric Distribution

The probability of x successful outcomes in r dependent trials is

$$P(x) = \frac{{}_aC_x \cdot {}_{n-a}C_{r-x}}{{}_nC_r}$$

where a is the number of successful outcomes available in a population of size n.

${}_aC_x \longrightarrow$ successes

${}_{n-a}C_{r-x} \longrightarrow$ what is left over

${}_nC_r \longrightarrow$ all possible combinations

Expectation for a Hypergeometric Distribution

The ratio of the expectation to the number of trials is proportional to the ratio of the number of available successes to the size of the population.

$$\frac{E(x)}{r} = \frac{a}{n}$$

So, $E(X) = \frac{ra}{n}$.

Example 2

Hypergeometric Distribution

A five-card hand is dealt from a standard deck of cards.

a) Show the probability distribution for the number of hearts in the hand.

b) Illustrate the distribution with a probability histogram.

c) Describe the shape of the graph.

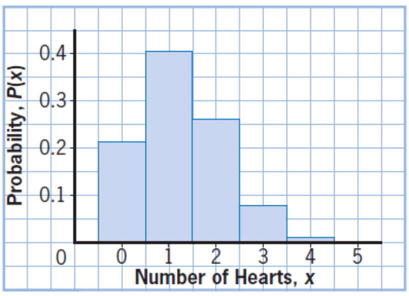
d) What does $P(5)$ tell you?

e) Calculate the expectation and explain its meaning.

a)

Number of Hearts, x	Probability, $P(x)$	$x \cdot P(x)$
0	$\frac{{}_{13}C_0 \times {}_{39}C_5}{{}_{52}C_5} \approx 0.2215$	0
1	$\frac{{}_{13}C_1 \times {}_{39}C_4}{{}_{52}C_5} \approx 0.4114$	0.4114
2	$\frac{{}_{13}C_2 \times {}_{39}C_3}{{}_{52}C_5} \approx 0.2743$	0.5486
3	$\frac{{}_{13}C_3 \times {}_{39}C_2}{{}_{52}C_5} \approx 0.0815$	0.2445
4	$\frac{{}_{13}C_3 \times {}_{39}C_1}{{}_{52}C_5} \approx 0.0107$	0.0428
5	$\frac{{}_{13}C_5 \times {}_{39}C_0}{{}_{52}C_5} \approx 0.0005$	0.0025

b)



c) It has a bell like shape, with a mode of $x = 1$ and it is skewed to the right.

d) $P(5) = 0.05\%$ which tells us that it is very unlikely to happen (1 in 2000 chance).

e) $E(x) = \frac{ra}{n}$

$= \frac{5(13)}{52} = 1.25$

We expect to get 1.25 hearts in our hand of 5 cards.

Your Turn

A bag contains 10 jellybeans. Four are blue and six are green. Four jellybeans are selected at random.

a) Show the probability distribution for the number of green jellybeans selected.

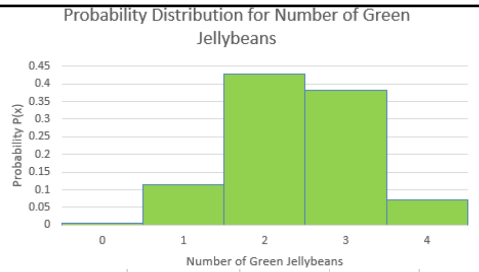
b) Illustrate the distribution with a probability histogram.

c) Compare the shape of the graph to the one in Example 2. Explain any differences.

d) What does $P(0)$ tell you?

e) Calculate the expectation and explain its meaning.

b)



a)

x	n	r	a	n-a	r-x	aCx	n-aCr-x	nCr	P(x)
0	10	4	6	4	4	1	1	210	0.004762
1	10	4	6	4	3	6	4	210	0.114286
2	10	4	6	4	2	15	6	210	0.428571
3	10	4	6	4	1	20	4	210	0.380952
4	10	4	6	4	0	15	1	210	0.071429

c) This distribution is less bell like than before in that it has two taller bars and the others are much shorter. It has a mode of $x = 2$ and the data is skewed to the left.

d) $P(0) = 0.476\%$ which means it is very unlikely to occur (less than 1 in 200)

e) $E(x) = \frac{ra}{n}$

$= \frac{4(6)}{10}$

$= 2.4$

We would expect to get 2.4 green jellybeans from the 4 chosen.

$n = 10$

$r = 4$

$a = 6$ (green)

$$P(0) = \frac{{}_6C_0 \times {}_4C_4}{{}_{10}C_4} = 0.0048$$

$$P(1) = \frac{{}_6C_1 \times {}_4C_3}{{}_{10}C_4} = 0.1143$$

$$P(2) = \frac{{}_6C_2 \times {}_4C_2}{{}_{10}C_4} = 0.4286$$

$$P(3) = \frac{{}_6C_3 \times {}_4C_1}{{}_{10}C_4} = 0.3810$$

$$P(4) = \frac{{}_6C_4 \times {}_4C_0}{{}_{10}C_4} = 0.0714$$

Example 3

Apply the Hypergeometric Distribution to Selections

In a class of 30 students, 18 have a driver's licence. Ten students are selected at random.

a) What is the probability that at least four have their driver's licence?

b) What is the expected number of students with their driver's licence?

$n = 30$
$r = 10$
$a = 18$

a) $P(x \geq 4) = 1 - P(0) - P(1) - P(2) - P(3)$

$= 1 - \frac{{}_{18}C_0 \times {}_{12}C_{10}}{{}_{30}C_{10}} - \frac{{}_{18}C_1 \times {}_{12}C_9}{{}_{30}C_{10}} - \frac{{}_{18}C_2 \times {}_{12}C_8}{{}_{30}C_{10}} - \frac{{}_{18}C_3 \times {}_{12}C_7}{{}_{30}C_{10}}$

$\approx 1 - 0.0000022 - 0.000013 - 0.0025 - 0.0215$

≈ 0.9760 The probability that at least 4 of the students have their licence is 97.60%

b) $E(x) = \frac{ra}{n}$
$= \frac{10(18)}{30}$
$= 6$

On average, we can expect 6 of the 10 students selected to have their licence.

Your Turn

Twenty-four students have signed up to attend a workshop. Fourteen are female and ten are male. Seven are randomly chosen to attend.

a) What is the probability that at least three are male?

b) What is the expected number of male and female students chosen?

$n = 24$
$r = 7$
$a = 10$

a) $P(x \geq 3) = 1 - P(0) - P(1) - P(2)$

$= 1 - \frac{{}_{10}C_0 \times {}_{14}C_7}{{}_{24}C_7} - \frac{{}_{10}C_1 \times {}_{14}C_6}{{}_{24}C_7} - \frac{{}_{10}C_2 \times {}_{14}C_5}{{}_{24}C_7}$

$\approx 1 - 0.0099 - 0.0868 - 0.2603$

≈ 0.6430

b) $E(x) = \frac{ra}{n}$
$= \frac{7(10)}{24}$
$= 2.9167$ males

Expected females
$= 7 -$ males
$= 7 - 2.9167$
$= 4.0833$

On average, we would expect 2.9167 males and 4.0833 females to attend the workshop.

Example 4

Apply the Expectation Formula

Wildlife officials tagged 350 seals from a particular colony. Forty seals were caught later, and 17 of them had been tagged. What is the approximate size of the seal population in this colony?

$n = ?$
$r = 40$
$a = 350$
$E(x) = 17$

Set up a proportion

$$\frac{E(x)}{r} = \frac{a}{n}$$

$$\frac{17}{40} = \frac{350}{n}$$

$$17n = 350(40)$$

$$\frac{17n}{17} = \frac{14{,}000}{17}$$

$$n = 823.53$$

$\Rightarrow$ The seal population is about 824 seals

Your Turn

During one summer, 500 foxes were caught and vaccinated against rabies. At that time, they were also tagged. Eighty foxes were later caught to estimate the size of the fox population, and 34 of them had been tagged. Estimate the size of the fox population.

$n = ?$
$r = 80$
$a = 500$
$E(x) = 34$

Set up a proportion

$$\frac{E(x)}{r} = \frac{a}{n}$$

$$\frac{34}{80} = \frac{500}{n}$$

$$34n = 500(80)$$

$$\frac{34n}{34} = \frac{40{,}000}{34}$$

$$n = 1176.47$$

$\Rightarrow$ The fox population is about 1176 foxes.

Key Concepts

- A hypergeometric probability distribution occurs when there are two outcomes, success and failure, and all trials are dependent. The random variable is the number of successes in a given number of trials.
- You can represent a hypergeometric distribution using a table, a probability histogram, or a formula.
- The probability of x successes in r dependent trials is $P(x) = \frac{{}_aC_x \cdot {}_{n-a}C_{r-x}}{{}_nC_r}$, where a is the number of successful outcomes available in a population of size n.
- Expectation $E(X) = \frac{ra}{n}$.

R1. For each example of a hypergeometric distribution, identify the random variable, the size of the sample space, the size of the population, and the range of the random variable.

a) A bag contains six red and four green marbles. Five marbles are randomly selected from the bag. The number of red marbles is recorded.

b) A seven-card hand is dealt from a standard deck. The number of hearts is recorded.

a) Random variable: # of reds selected. Size of sample space: 5 marbles. Size of population: 10 marbles. Range of random variable: 0 to 5 marbles.

b) Random variable: # of hearts selected. Size of sample space: 7 cards. Size of population: 52 cards. Range of random variable: 0 to 7 cards.

R2. A standard die is rolled five times and the number of 3s is noted. Explain why this would or would not be a valid hypergeometric probability situation.

Since the trials are independent, this is not a hypergeometric probability situation. It would actually be binomial.