SAMPLING WITHOUT REPLACEMENT

Example:

An accounting population contains 52 line items of which 25% are in error. A simple random sample of 6 line items is drawn. What is the probability that the sample will contain 2 items in error?

Solution:

1. Sampling with Replacement $\Longrightarrow$ Binomial

   Binomial with $n = 6$ and $p = 0.25$;

$$P(\text{ 2 accounts in error}) = \begin{pmatrix} 6 \\ 2 \end{pmatrix} .25^2 .75^4 = 0.2966$$

2. Sampling without Replacement

13	39
In error	Correctly stated

$$P(\ 2\ \text{ accounts in error}) = \frac{\binom{13}{2}\binom{39}{4}}{\binom{52}{6}} = 0.31513$$

Hypergeometric Distribution

A finite population of size N consists of

M elements called successes

L elements called failures

A sample of n elements are selected at random

without replacement.

X = number of successes

$$P(X = x) = \frac{\begin{pmatrix} M \\ x \end{pmatrix} \begin{pmatrix} L \\ n - x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

X is said to have a **hypergeometric distribution**

Example:

Draw 6 cards from a deck without replacement.

What is the probability of getting two hearts?

Solution:

Here

$$\begin{aligned} M &= 13 \text{ number of hearts} \\ L &= 39 \text{ number of non-hearts} \\ N &= 52 \text{ total} \end{aligned}$$

$$P(2\ hearts) = \frac{\binom{13}{2}\binom{39}{4}}{\binom{52}{6}}$$

Example: Lotto

42 balls are numbered 1 - 42.

You select six numbers between 1 and 42. (The ones you write on your lotto card)

Number of possible ways to draw six mumbers in the range [1, 42] = $\binom{42}{6}$

What is the probability that they contain

(i) match 6?

(ii) match 5?

(ii) match 4?

(iii) match 3?

Solution:

Total = 42; Favourable = 6; Non-Favourable = 36.

Sample size n = 6.

$$P(\text{match } 4) = \frac{\binom{6}{4}\binom{36}{2}}{\binom{42}{6}} = .0018$$

ODDS OF ABOUT 1 in 500

Binomial or Hypergeometric?

An accounting population consists of 20 line items of which 10% are incorrectly state. Find the probability that no more than 2 incorrectly stated accounts will be found in a sample of size 10. Let X = no of misstated accounts.

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

With Replacement Sampling

$$\begin{aligned} P(X = 0) &= .9^{10} \quad = .3487 \\ P(X = 1) &= \begin{pmatrix} 10 \\ 1 \end{pmatrix} .1^1 .9^9 = .3874 \\ P(X = 2) &= \begin{pmatrix} 10 \\ 2 \end{pmatrix} .1^2 .9^8 = .1937 \end{aligned}$$

Without Replacement Sampling

$$P(X=0) = \frac{\binom{18}{10}}{\binom{20}{10}} = .2368$$

$$P(X=1) = \frac{\binom{2}{1}\binom{18}{9}}{\binom{20}{10}} = .5263$$

$$P(X=2) = \frac{\binom{2}{2}\binom{18}{8}}{\binom{20}{10}} = .2368$$

Binomial or Hypergeometric?

An accounting population consists of 200 line items of which 10% are incorrectly state. Find the probability that no more than 2 incorrectly stated accounts will be found in a sample of size 10. Let X = no of incorrectly stated accounts.

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

With Replacement Sampling

$$\begin{aligned} P(X = 0) &= .9^{10} \qquad = .3487 \\ P(X = 1) &= \begin{pmatrix} 10 \\ 1 \end{pmatrix} .1^1 .9^9 = .3874 \\ P(X = 2) &= \begin{pmatrix} 10 \\ 2 \end{pmatrix} .1^2 .9^8 = .1937 \end{aligned}$$

Without Replacement Sampling

$$P(X=0) \;=\; \frac{\binom{180}{10}}{\binom{200}{10}} = .3398$$

$$P(X=1) \;=\; \frac{\binom{20}{1}\binom{180}{9}}{\binom{200}{10}} = .3974$$

$$P(X=2) \;=\; \frac{\binom{20}{2}\binom{180}{8}}{\binom{200}{10}} = .1975$$

Binomial or Hypergeometric?

An accounting population consists of 2,000 line items of which 10% are incorrectly state. Find the probability that no more than 2 incorrectly stated accounts will be found in a sample of size 10. Let X = no of incorrectly stated accounts.

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

With Replacement Sampling

$$\begin{aligned} P(X = 0) &= .9^{10} \qquad = .3487 \\ P(X = 1) &= \begin{pmatrix} 10 \\ 1 \end{pmatrix} .1^1 .9^9 = .3874 \\ P(X = 2) &= \begin{pmatrix} 10 \\ 12 \end{pmatrix} .1^2 .9^8 = .1937 \end{aligned}$$

Without Replacement Sampling

$$P(X=0) = \frac{\binom{1800}{10}}{\binom{2000}{10}} = .3476$$

$$P(X=1) = \frac{\binom{200}{1}\binom{1800}{9}}{\binom{2000}{10}} = .3881$$

$$P(X=2) = \frac{\binom{200}{2}\binom{1800}{8}}{\binom{2000}{10}} = .1939$$

Theorem: As $N \to \infty$, the hypergeometric distribution converges to the binomial.

Proof:

$$\begin{aligned}
\text{Population Size} &= N \\
\text{Proportion of successes} &= p \\
\text{Number of successes in N} &= Np \\
\text{Number of failures} &= N(1-p)
\end{aligned}$$

Let X = number of successes in s sample of size n drawn without replacement from N

Np	N(1-p)
Successes	Failures

We will show that for the hypergeometric

$$P(X = x) = \frac{\begin{pmatrix} Np \\ x \end{pmatrix} \begin{pmatrix} N(1-p) \\ n-x \end{pmatrix}}{\begin{pmatrix} N \\ x \end{pmatrix}}$$

$$\rightarrow \begin{pmatrix} n \\ x \end{pmatrix} p^x q^{n-x} \quad \text{as} \quad N \rightarrow \infty$$

Proof:

$$P(X = x) \;=\; \frac{\begin{pmatrix} Np \\ x \end{pmatrix} \begin{pmatrix} Nq \\ n - x \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$$

$$= \frac{\left(\frac{(Np)!}{x!(Np-x)!}\right)\left(\frac{(Nq)!}{(n-x)!\,(Nq-(n-x))!}\right)}{\frac{N!}{n!(N-n)!}}$$

$$= \frac{n!}{x!(n-x)!}\left[\frac{(Np)!}{(Np-x)!}\frac{(Nq)!}{(Nq-(n-x))!}\frac{(N-n)!}{N!}\right]$$

$$\rightarrow \begin{pmatrix} n \\ x \end{pmatrix} p^x q^{n-x} \quad \text{as} \quad N \rightarrow \infty$$

$$
\begin{aligned}
\frac{(Np)!}{(Np-x)!} &= Np(Np-1)(Np-2)\cdots(Np-x+1) \\
&= N^x p(p-\frac{1}{N})(p-\frac{2}{N})\cdots(p-\frac{x-1}{N})
\end{aligned}
$$

$$
\begin{aligned}
\frac{Nq}{(Nq-(n-x))!} &= (Nq)(Nq-1)\cdots(Nq-(n-x)+1) \\
&= N^{n-x} q(q-\frac{1}{N})\cdots(q-\frac{n-x+1}{N})
\end{aligned}
$$

$$
\begin{aligned}
\frac{(N-n)!}{N!} &= \frac{1}{N(N-1)(N-2)\cdots(N-n+1)} \\
&= 1/\left(N^n(1-\frac{1}{N})(1-\frac{2}{N})\cdots(1-\frac{n-1}{N})\right)
\end{aligned}
$$