DISTRIBUCIÓN DE PROBABILIDAD DISCRETA HIPERGEOMÉTRICA

Marlly Olave cód. 1125614

Anderson Hoyos cód.30602

Presentado a: Julián Ignacio López

Asignatura: Inferencia Estadística

Programa: Ing. de Industrial

               Ing. de Sistemas

Facultad de Ingeniería

UNIVERSIDAD DE SAN BUENAVENTURA

MARZO 21 DE 2019

CALI

**DISTRIBUCIÓN HIPERGEOMÉTRICA**

1. **Crear perfil github [www.github.com/](http://www.github.com/)**
2. **Explicar el Concepto**

La distribución hipergeométrica es una distribución discreta en el cual se observa el número de eventos en una muestra cuando se conoce el número total de elementos en la población de la cual proviene la muestra.

Cada elemento de la muestra tiene dos resultados posibles (**es un evento o un no evento**). Si se elige un elemento de la población, no se vuelve a elegir.

A diferencia de Binomial, los elementos los tomamos uno a uno, de manera que un elemento no aparecerá dos veces en una muestra (**muestreo sin reemplazo**).

La distribución hipergeométrica se define por **3** parámetros:

1. **tamaño de la población**
2. **conteo de eventos en la población**
3. **tamaño de la muestra.**
4. **Explicar para qué aplicaciones es útil**

La distribución Hipergeométrica posee grandes aplicaciones especialmente en **control de calidad** para procesos experimentales en los que no es posible retornar a la situación de partida. Facilitando el cálculo de probabilidades para **productos defectuosos** y le da la posibilidad al ingeniero de tener una idea de la cantidad del **producto útil**.

**Ejemplo:**

Una pieza de equipo electrónico contiene seis chips de computadora, dos de

que son defectuosos Se seleccionan aleatoriamente tres chips de computadora para la inspección, y se registra el número de chips defectuosos. Encuentra la distribución de probabilidad para x, el número de chips de computadora defectuosos.

1. **Explicar el proceso para su uso**

Es una distribución para el estudio de **muestras pequeñas** de **poblaciones pequeñas** y para calcular probabilidades de **juegos de azar**, en pocas palabras se utiliza para calcular **la probabilidad** de una **selección aleatoria** de un objeto que no tiene reemplazo.

1. **Explicar la generación de datos en la distribución**

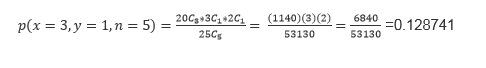
Los datos en una distribución hipergeométrica se calcula:

1. Las probabilidades requeridas de que un objeto tengan defectos mayores o menores.
2. Se colocan el número de combinaciones según el problema, es decir dependiendo del número total de productos con mayor o menor defectos y que cantidad se requiere de cada uno y se divide sobre la combinación entre la cantidad de la población y el tamaño de la muestra.

**Ejemplo de generación de datos:**

En un lote de productos se tienen 20 productos sin defectos, 3 con defectos menores y 2 con defectos mayores, se seleccionan al azar 5 productos de este lote, determine la probabilidad de que:

1. 3 de los productos seleccionados no tengan defectos
2. 1 tenga defectos menores.



1. **Explicar el ajuste de datos a la función**

La función de probabilidad es:

**Donde:**

Tamaño de población **(N)**

N° de individuos que…(**k)**

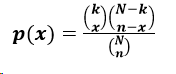
Tamaño de la muestra **(n)**

Valor que toma la variable **(x)**

**7.**    **ejercicios aplicados con su respectiva aplicación en R**

**ejemplo 1.** usted recibe un envío de pedido especial de 500 etiquetas. Supongamos que el 2% de las etiquetas es defectuoso. El conteo de eventos en la población es de 10 (0.02 \* 500). Usted toma una muestra de 40 etiquetas y desea determinar la probabilidad de que haya 3 o más etiquetas defectuosas en esa muestra.

**Desarrollo:**

****

**Datos:**

|  |  |
| --- | --- |
| **N** | **500** |
| **k** | **10** |
| **n** | **40** |
| **x** | **3** |

4

**Formulando en R:**



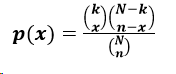
**Solución:**

La probabilidad de que haya 3 o más etiquetas defectuosas en la muestra es de 0.0384

**Ejemplo 2:** Supóngase que se tienen 30 representantes de cierto estado, a una convención política nacional, de los cuales 20 apoyan al candidato del PRI y 10 al candidato del PAN. Si se selecciona aleatoriamente 6 representantes.

* 1. ¿Cuál es la probabilidad de que exactamente 5 apoyen al candidato del PRI?
  2. ¿Cuál es la probabilidad de que 3 o más apoyen al candidato del PAN?
  3. ¿Cuál es la probabilidad de que cuando mucho 2 apoyen al candidato del PAN?

**Desarrollo:**

****



|  |  |
| --- | --- |
| **N(población)** | **30** |
| **k(éxitos incluidos en la población)** | **20** |
| **n(muestra)** | **6** |
| **x(valor de la variable)** | **5** |

La probabilidad , es decir, de que 5 apoyen al candidato del PRI es 0.261109

**Formulado en R:**

**>dhyper(x, n, N-n, k)**

https://lh3.googleusercontent.com/P79qa8CjkW5PS89D4vqxhta25N7RmVODgt9b3AwTN5zm_2-NL9Mnl1pc-a7wfdhuJ2z51OYJjEx8WYMjYrTvoAa4u0RZGFL-hjf1r6YmoOZFb2f2EFm4mL0OK6dyAN6DNmKbauSN

La probabilidad , es decir, de que 5 apoyen al candidato del PRI es 0.261109

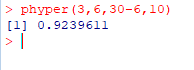
**b).**

|  |  |
| --- | --- |
| **N(población)** | **30** |
| **k(éxitos incluidos en la población)** | **10** |
| **n(muestra)** | **6** |
| **x(valor de la variable)** | **3, 4,5…** |

**)**

La probabilidad, es decir, de que 3 o más apoyen al candidato del PAN es 0.9

**Formulado en R: >phyper (x,n,N-n,k)**



La probabilidad, es decir, de que 3 o más apoyen al candidato del PAN es 0.9239611

**c)**

|  |  |
| --- | --- |
| **N(población)** | **30** |
| **k(éxitos incluidos en la población)** | **10** |
| **n(muestra)** | **6** |
| **x(valor de la variable)** | **0,1,2** |

**Formulado en R:**

**>phyper(x, n, N-n, k)**



La probabilidad, es decir, de que cuando mucho 2 o menos apoyen al candidato del PAN es 0.6935708

**Ejemplo 3**

Suponiendo la extracción aleatoria de 10 elementos de un conjunto formado por 50 elementos totales (**cartas baraja española**) de los cuales 20 son del tipo A (**salir oro**) y 30 son del tipo complementario (**no salir oro**).  
Realizando las extracciones sin devolver los elementos extraídos y llamamos X al número de elementos del tipo A (**oros obtenidos**) que extraemos en las 10 cartas; X seguirá una distribución hipergeométrica de parámetros (50, 10, 20/50).   
Para calcular la probabilidad de obtener **5 oros**:

**Variables**

**N** = 50 elementos

**k** = 20 elementos si llega a **salir oro**

**n =** 10 cartas de muestra

**x =** 5 cartas seleccionadas

La probabilidad que salga 5 cartas de oro de las 10 cartas que tenemos como muestra es **P= 0.215**

**Formulando en R:**

> dhyper(x, n,N-n,k)

> dhyper(5,10,50-10,20)

[1] 0.215085

**Ejemplo 4.** Para evitar que lo descubran en la aduana, un viajero ha colocado **6 tabletas**de narcótico en una botella que contiene **9 píldoras** de vitamina que son similares en apariencia. Si el oficial de la aduana selecciona **3 tabletas aleatoriamente** para analizarlas. **a)** ¿Cuál es la probabilidad de que el viajero sea arrestado por posesión de narcóticos?, **b)** ¿Cuál es la probabilidad de que no sea arrestado por posesión de narcóticos? **Existen dos formas de resolverlo.**

**Solución:**

**N** = 9+6 =15 total de tabletas

**k** = 6 tabletas de narcótico

**n =** 3 tabletas seleccionadas

**x =** 0, 1, 2, o 3 tabletas de narcótico

**Primera forma:**

http://www.itchihuahua.edu.mx/academic/industrial/sabaticorita/_private/03Ddistr%20Hipergeometrica_archivos/image018.gif

http://www.itchihuahua.edu.mx/academic/industrial/sabaticorita/_private/03Ddistr%20Hipergeometrica_archivos/image020.gif

**Segunda forma:**

**Es lo mismo**

=**0.8154**

La probabilidad que la aduana lo arreste es P=**0.8154** yP=**0.1846** que no sea arrestado.

**Presentándolo en R:**

**>** dhyper(x, n, N-n,k)

> dhyper (5, 10,50-10,20)

[1] 0.215085

> dhyper (1, 3, 15-3,6)

[1] 0.4747253

> dhyper (2, 3,15-3,6)

[1] 0.2967033

> dhyper (3, 3,15-3,6)

[1] 0.04395604

> dhyper(3,3,15-3,6)+dhyper(2,3,15-3,6)+dhyper(1,3,15-3,6)

[1] 0.8153846

> dhyper(0,3,15-3,6)

[1] 0.1846154

**Ejemplo 5.**

De un lote de 10 proyectiles, 4 se seleccionan al azar y se disparan. Si el lote contiene 3 proyectiles defectuosos que no explotarán, ¿cuál es la probabilidad de que, **a)** los 4 exploten?, **b)** al menos 2 no exploten?

**Solución:**

**N =** 10 proyectiles en total

**a =** 7 proyectiles que explotan

**n =** 4 proyectiles seleccionados

**x =** 0, 1, 2, 3 o 4 proyectiles que explotan = variable que nos define el número de proyectiles que explotan entre la muestra que se dispara.

http://www.itchihuahua.edu.mx/academic/industrial/sabaticorita/_private/03Ddistr%20Hipergeometrica_archivos/image030.gif

**b)  N =** 10 proyectiles en total

**a =** 3 proyectiles que no explotan

**n =** 4 proyectiles seleccionados

**x =** 0, 1, 2 o 3 proyectiles que no explotan

**p(**al menos 2 no exploten**) = p (**2 o más proyectiles no exploten**) = p(x = 2 o 3; n=4)**

3http://www.itchihuahua.edu.mx/academic/industrial/sabaticorita/_private/03Ddistr%20Hipergeometrica_archivos/image032.gif

**Formulando en R:**

**>** dhyper(x,n,N-n,k)

> dhyper(4,4,10-4,7)

[1] 0.1666667

> dhyper(1,4,10-4,7)+dhyper(2,4,10-4,7)

[1] 0.3333333

**Ejemplo 6.**

**a)¿**Cuál es la probabilidad de que una mesera se rehúse a servir bebidas alcohólicas únicamente a dos menores de edad si verifica aleatoriamente solo 5 identificaciones de entre 9 estudiantes, de los cuales 4 no tienen la edad suficiente**?, b) ¿**Cuál es la probabilidad de que como máximo 2 de las identificaciones pertenezcan a menores de edad?

**Solución:**

**a) N** = 9  total de estudiantes

**a =** 4 estudiantes menores de edad

**n =** 5 identificaciones seleccionadas

**x =** 0, 1, 2, 3 o 4 identificaciones de personas menores de edad**,** variable que nos define el número de identificaciones que pertenecen a personas menores de edad.

http://www.itchihuahua.edu.mx/academic/industrial/sabaticorita/_private/03Ddistr%20Hipergeometrica_archivos/image034.gif

**b) N =** 9 total de estudiantes

**a** = 4 estudiantes menores de edad

**n =** 5 identificaciones seleccionadas

**x =** 0, 1, 2,  3 o 4 identificaciones de personas menores de edad, variable que nos define el número de identificaciones que pertenecen a personas menores de edad

http://www.itchihuahua.edu.mx/academic/industrial/sabaticorita/_private/03Ddistr%20Hipergeometrica_archivos/image036.gif

http://www.itchihuahua.edu.mx/academic/industrial/sabaticorita/_private/03Ddistr%20Hipergeometrica_archivos/image038.gif

**Formular en R:**

con el commando dhyper

**>** dhyper(x,n,N-n,k)

> dhyper(2,5,9-5,4)

> dhyper(0,5,9-5,4)+dhyper(1,5,9-5,4)+dhyper(2,5,9-5,4)

[1] 0.6428571

**Ejemplo 7:**

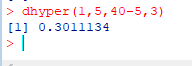
De un lote de 40 microcomponentes, cada uno se denomina aceptable si no tiene más de tres defectuosos. El procedimiento para muestrear el lote es la selección de 5 componentes al azar y rechazar el lote si se encuentra un componente defectuoso.

¿Cuál es la probabilidad de que se encuentre exactamente un defectuoso en la muestra si hay tres defectuosos en todo el lote?

Datos:

**Formulando en R:**

>dhyper(x, n, N-n, k)

:

**Solución:** La probabilidad de que se encuentre un microcomponente defectuoso habiendo 3 defectuosos en el lote es de:

**Glosario**

para la distribución Hipergeométrica en R, se dispone de las funciones:

|  |  |
| --- | --- |
| **R: Distribución Hipergeométrica.** | |
| **dhyper(**x, m, n, k, log = F**)** | Devuelve resultados de la función de densidad. |
| **phyper(**q, m, n, k, lower.tail = T, log.p = F**)** | Devuelve resultados de la función de distribución acumulada. |
| **qhyper(**p, m, n, k, lower.tail = T, log.p = F**)** | Devuelve resultados de los cuantiles de la Hipergeométrica. |
| **rhyper(**nn, m, n, k**)** | Devuelve un vector de valores de la Hipergeométrica aleatorios. |

Los argumentos que podemos pasar a las funciones expuestas en la anterior son:

**x, q:** Vector de cuantiles. Corresponde al número de particulares en la muestra.

**m:** Selección aleatoria particular.

**n:** El número total de la población menos la selección aleatoria particular. n = N - m.

**n:** El número de la selección a evaluar.

**prob:** Probabilidad.

**nn:** Número de observaciones.

**log, log. p:** Parámetro booleano, si es TRUE, las probabilidades p son devueltas como log (p).

**lower.tail**: Parámetro booleano, si es TRUE (por defecto), las probabilidades son P[X ≤ x], de lo contrario, P [X > x].