



Clase 5: Distribuciones Binomial e Hipergeométrica.

Aproximación Binomial de la Hipergeométrica

Profesora: Olga Alexandra Bustos Giraldo

Escuela de Estadística

Universidad Nacional de Colombia, Sede Medellín

oabustos@unal.edu.co



Combinatoria

Factorial

El factorial de un número entero positivo se define como el producto de todos los números naturales anteriores o iguales a él, se escribe $n!$, y se lee "n factorial". (Por definición $0! = 1$)

Por ejemplo, $5! = 5 * 4 * 3 * 2 * 1 = 120$

Combinaciones

Sean n y r números naturales, el símbolo $\binom{n}{r}$ se lee " n tomados de a r " y se define como:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

Es el número de combinaciones de r elementos escogidos de un conjunto con n elementos. Este símbolo se conoce como *Coeficiente Binomial*.



Distribución Discreta: Bernoulli

Si se realiza un único experimento (un ensayo) con dos posibles resultados: éxito o fracaso, se dice que la v.a discreta se distribuye Bernoulli con parámetro p (probabilidad de éxito), se escribe $X \sim \text{Bernoulli}(p)$ y la f.m.p es:

$$P(X = x) = p^x(1 - p)^{1-x} \quad ; \quad x = 0, 1$$



Distribución Discreta: Binomial

Un experimento aleatorio es llamado **Binomial** si cumple las siguientes características:

1. El experimento consiste de una sucesión de n ensayos donde n es fijo (se fija antes del experimento).
2. Los ensayos son idénticos y cada uno puede resultar en 'Éxito' o 'Fracaso'. Aplica a muestras de las que se toman elementos al azar con reemplazo o bien en poblaciones grandes.
3. Los ensayos son independientes, lo que implica que la probabilidad de éxito es constante de un ensayo a otro y la denotaremos como p .
4. La v.a de interés será llamada # éxitos o # de personas, animales, unidades que cumplan con cierta característica de interés en los n ensayos.

La distribución Binomial son n ensayos Bernoulli.



Distribución Binomial

Suponga que la variable aleatoria X es el número de éxitos de los n ensayos. El rango de X está dado por $A_X = \{0, 1, \dots, n\}$. La f.m.p. de X está dada por:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad ; \quad x = 0, 1, 2, \dots, n .$$

Por notación se escribe $X \sim \text{bin}(n, p)$. Donde n, p son los parámetros de la distribución.

El valor esperado y la varianza son:

$$E[X] = n p \quad , \quad \text{Var}[X] = n p (1 - p).$$



Ejemplo Binomial

Un examen de opción múltiple contiene 10 preguntas. Cada pregunta tiene cuatro opciones de las cuáles sólo una es la correcta. El examen se aprueba si se responden correctamente al menos seis preguntas. El estudiante adivina las respuestas. Conteste las siguientes preguntas:

- a) ¿Cuál es la probabilidad de aprobar el examen?
- b) Si el estudiante adivina al menos tres de las preguntas, ¿Cuál es la probabilidad de reprobado el examen?
- c) Halle el valor esperado y la varianza de la variable aleatoria.



Ejemplo Binomial

Solución

Sea X : # preguntas con respuesta correcta de las 10 contestadas. Entonces $X \sim \text{bin}(10, p)$, con $p = \frac{1}{4}$.

Así,

$$p(x) = \binom{10}{x} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{10-x}; \quad x = 0, 1, \dots, 10.$$

a)

$$P(X \geq 6) = 1 - P(X \leq 5) = 1 - \sum_{x=0}^5 \binom{10}{x} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{10-x} = 0.01973.$$

En R

Forma 1: `1 - pbinom(q=5, size=10, prob=1/4)`

Forma 2: `sum(dbinom(x=6:10, size=10, prob=1/4))`

(Clase 5).



Ejemplo Binomial

$$b) P(X < 6 | X \geq 3) = \frac{P(3 \leq X < 6)}{P(X \geq 3)} = \frac{P(3 \leq X \leq 5)}{1 - P(X < 3)}$$

$$P(X < 6 | X \geq 3) = \frac{\sum_{x=3}^5 \binom{10}{x} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{10-x}}{1 - P(X \leq 2)} = \frac{0.4547}{1 - 0.5256} = 0.9585 .$$

En R:

- numerador: `sum(dbinom(x=3:5, size=10, prob=1/4))`
- denominador: `pbinom(q=2, size=10, prob=1/4)`

$$c) E[X] = 10 \left(\frac{1}{4}\right) = 2.5 ; \quad Var[X] = 10 \left(\frac{1}{4}\right) \left(\frac{3}{4}\right) = \frac{15}{8} .$$

Se espera en promedio 2.5 preguntas con respuesta correcta. La desviación estándar es 1.36 ($\sqrt{15/8}$), mide la dispersión alrededor de la media, (1.14 , 3.86).

En R: `sum(0:10 * dbinom(x=0:10, size=10, prob=1/4))`



Ejemplo Binomial

Suponga que la probabilidad de tener una unidad defectuosa en una línea de ensamble es de 0.05. Se seleccionan al azar 20 unidades con reemplazo. Halle:

- a) ¿Cuál es la probabilidad de que entre 20 unidades seleccionadas al azar, dos sean defectuosas?
- b) ¿Cuál es la probabilidad de que a lo más (a lo sumo, como máximo) dos de las 20 unidades estén defectuosas?
- c) ¿Cuál es la probabilidad de que por lo menos 2 unidades estén defectuosas?



Ejemplo Binomial

Solución

Sea X : el # de unidades defectuosas de las 20 unidades seleccionadas. Este experimento cumple con todas las condiciones para ser un ensayo Binomial (se tiene un n fijo antes del experimento, los resultados son éxito o fracaso, son independientes, la probabilidad es constante en cada ensayo).

Así, $X \sim \text{bin}(20, 0.05)$, donde

$$p(x) = \binom{20}{x} (0.05)^x (0.95)^{20-x} \quad ; \quad x = 0, 1, \dots, 20 .$$

a) $P(X = 2) = p(2) = \binom{20}{2} (0.05)^2 (0.95)^{18} = 0.1887 .$

En R: `dbinom(x=2, size=20, prob=0.05)`



Ejemplo Binomial

b) $P(X \leq 2) = p(0) + p(1) + p(2) = 0.3585 + 0.3774 + 0.1887 = 0.9246$.

En R: `pbinom(q=2, size=20, prob=0.05)`

c) $P(X \geq 2) = 1 - P(X < 2) = 1 - P(X \leq 1) = 0.2641$.

En R

Forma 1: `1 - pbinom(q=1, size=20, prob=0.05)`

Forma 2: `sum(dbinom(x=2:20, size=20, prob=0.05))`

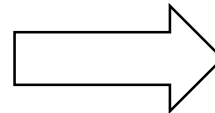
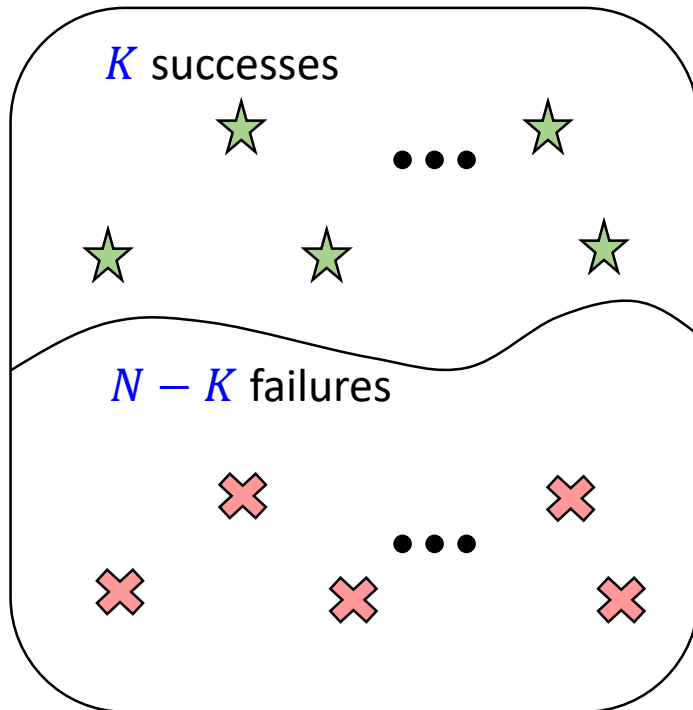


Distribución Discreta: Hipergeométrica

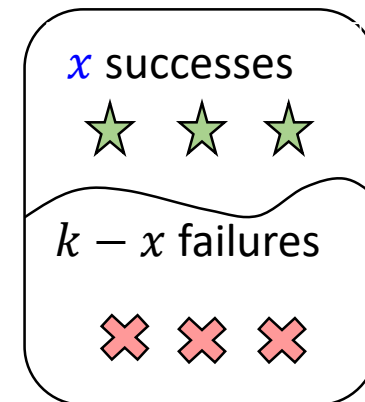
Suponga que una población finita tiene N elementos, cada uno de los cuales tiene una de dos características que se denominan éxitos o fracasos. Suponga que K de ellos se consideran éxitos o tienen la característica de interés y $N - K$ fracasos o no tienen la característica de interés. Se toman al azar y sin reemplazo n de estos elementos. Sea X : la variable aleatoria que representa el **número de éxitos o elementos que tienen la característica de interés** en los n seleccionados.

What is the probability to obtain x success in a sample with n elements?

Population with N elements



Sample with n elements





Distribución Hipergeométrica

$$p(x) = P(X = x) = \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}} \quad ; \quad \max\{0, n - (N - K)\} \leq x \leq \min\{n, K\}$$

Esta función de probabilidad se conoce como Distribución **Hipergeométrica**. Se denotará $X \sim \text{Hyper}(N, K, n)$. Donde N, K, n son los parámetros de la distribución.

$$\begin{aligned} E[X] &= np \\ V[X] &= \left(\frac{N-n}{N-1} \right) np(1-p). \quad \text{donde } p = K/N \end{aligned}$$

El término $\frac{N-n}{N-1}$ se conoce como *factor de corrección por población finita (muestra obtenida sin reemplazo)*. Observe que si n es muy pequeño comparado con N , entonces el factor de corrección tiende a uno, y la varianza de la hipergeométrica y la binomial son similares.



Distribución Hipergeométrica

Los experimentos que tienen este tipo de distribución tienen las siguientes características:

- Se esperan dos tipos de resultados: Éxito o fracaso
- Cada ensayo o repetición no es independiente de los demás.
- Las probabilidades asociadas a cada uno de los resultados no son constantes
- La distribución hipergeométrica es apropiada para modelar experimentos donde se hace muestras sin reemplazo o en poblaciones finitas.



Ejemplo Hipergeométrica

Un lote con 25 arandelas contiene tres en las que la variabilidad en el espesor alrededor de la circunferencia es inaceptable. Se toma una muestra al azar de tres arandelas sin reemplazo.

- a) ¿Cuál es la probabilidad de que ninguna de las arandelas inaceptables se encuentren en la muestra?
- b) ¿Cuál es la probabilidad de que al menos una de las arandelas inaceptables se encuentren en la muestra?
- c) ¿Cuál es la probabilidad de que exactamente una de las arandelas inaceptables se encuentre en la muestra?



Ejemplo Hipergeométrica

Solución

Sea X : # de arandelas inacceptables en la muestra.

La f.m.p. de X es de la forma:

$$p(x) = \frac{\binom{3}{x} \binom{22}{3-x}}{\binom{25}{3}} \quad ; \quad x = 0, 1, 2, 3 .$$

$$\text{a) } P(X = 0) = p(0) = \frac{\binom{22}{3}}{\binom{25}{3}} = \frac{1540}{2300} = 0.66957 .$$

En R: `dhyper(x=0, 3, 22, 3)`

$$\text{b) } P(X \geq 1) = 1 - P(X < 1) = 1 - p(0) = 1 - 0.66957 = 0.33043 .$$

Forma 1: `1 - dhyper(x=0, 3, 22, 3)`

Forma 2: `sum(dhyper(x=1:3, 3, 22, 3))`

c) $P(X = 1) = p(1) = \frac{\binom{3}{1} \binom{22}{2}}{\binom{25}{3}} = \frac{693}{2300} = 0.30130 .$

En R: `dhyper(x=1, 3, 22, 3)`

d) Calcular $P(X \leq 2)$ usando $F(x)$ en R.

En R: `phyper(q=2, 3, 22, 3)`



Ejemplo Hipergeométrica

Un geólogo ha recolectado 10 especímenes de roca basáltica y 10 de granito. Se instruye a un asistente de laboratorio para que seleccione al azar 6 de los especímenes sin reemplazo para analizarlos.

- a) ¿Cuál es la f.m.p. para el número de especímenes de basalto seleccionados para analizarlos?
- b) ¿Cuál es la probabilidad de que todos los especímenes de la muestra sean de uno de los dos tipos de roca seleccionados para el análisis?
- c) ¿Cuál es la probabilidad de que la cantidad de especímenes de basalto seleccionados para su análisis esté a menos de una desviación estándar de la media?



Ejemplo Hipergeométrica

Solución

Sea X : # de especímenes de basalto en la muestra para análisis.

a)

$$p(x) = \frac{\binom{10}{x} \binom{10}{6-x}}{\binom{20}{6}} ; \quad x = 0, 1, 2, 3, 4, 5, 6 .$$

b)

$$\begin{aligned} P(X = 6 \vee X = 0) &= P(X = 0) + P(X = 6) = p(0) + p(6) \\ &= \frac{\binom{10}{6}}{\binom{20}{6}} + \frac{\binom{10}{6}}{\binom{20}{6}} = \frac{210 + 210}{38760} = 0.01084 . \end{aligned}$$



Ejemplo Hipergeométrica

c)

$$E[X] = 6 \left(\frac{10}{20} \right) = 3 \quad \text{Var}[X] = \left(\frac{20-6}{20-1} \right) 6 \left(\frac{10}{20} \right) \left(1 - \frac{10}{20} \right)$$

$$\text{Var}[X] = \left(\frac{14}{19} \right) \left(\frac{6}{4} \right) = \frac{21}{19} = 1.10526 \quad \text{y} \quad \sigma = \sqrt{\frac{21}{19}} = 1.05.$$

$$P(|X - \mu_x| \leq \sigma) = P\left(-\sqrt{\frac{21}{19}} \leq X - 3 \leq \sqrt{\frac{21}{19}}\right) = P(1.95 \leq X \leq 4.05).$$

$$P(|X - \mu_x| \leq \sigma) = P(2 \leq X \leq 4) = p(2) + p(3) + p(4).$$

$$P(|X - \mu_x| \leq \sigma) = \frac{9450}{38760} + \frac{14400}{38760} + \frac{9450}{38760} = 0.85913.$$

phyper(4,10,10,6)-phyper(1,10,10,6)



Teorema de Aproximación

Sea X una variable aleatoria tal que $X \sim \text{Hiper}(N, K, n)$.

$\left(\frac{N-n}{N-1}\right) \approx 1$, si $n/N \leq 0.05$ entonces:

$$\frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}} \approx \binom{n}{x} p^x (1-p)^{n-x} ; \text{ donde } p = \frac{K}{N}.$$

Este resultado se conoce como *Aproximación Binomial de la Hipergeométrica*.

Por ejemplo: si $N=200$, $n=5$, $K=30$ la relación $n/N = 0.025 < 0.05$, entonces, calculemos la $P(X = 3)$, se tiene que:

$$P(X = 3) = \frac{\binom{30}{3} \binom{170}{2}}{\binom{200}{5}} = 0.023$$

Usando la aproximación se tiene que:

$$P(X = 3) \approx \binom{5}{3} \left(\frac{3}{20}\right)^3 \left(\frac{17}{20}\right)^2 = 0.024$$



Ejercicio

Un gran distribuidor tiene 800 equipos de computo. Se sabe que 8 de ellos están defectuosos, una compañía compra 20 de ellos y le son despachados al azar, ¿cuál es la probabilidad de que en la muestra se observen 2 o más defectuosos?

X es el # de defectuosos. Por esa razón $X \sim \text{hyper}(N=800, K=8, n=20)$