APÉNDICE 6 A: LA DISTRIBUCIÓN BINOMIAL

Estadística - Lic. en Nutrición – UCEL
Marta Ruggieri, Julia Fernández, M. Eugenia Tesser

DISTRIBUCIÓN DE VARIABLES ALEATORIAS

Asociadas a cada variable aleatoria hay un modelo :

Función de probabilidad	• Variable aleatoria discreta • En este curso veremos el modelo **Binomial**
Función de densidad de probabilidad	• Variable aleatoria continua • En este curso veremos el modelo **Normal**

DISTRIBUCIÓN DE VARIABLE ALEATORIA DISCRETA

Una **distribución de probabilidad** está dada por todos los valores que puede tomar la variable aleatoria discreta y las probabilidades asociadas a cada uno de ellos.

Puede representarse mediante una tabla, un gráfico o una fórmula.

La probabilidad de que la variable aleatoria asume un valor, se simboliza: $P(X=x) = p(x)$

DISTRIBUCIÓN DE VARIABLE ALEATORIA DISCRETA

La función de probabilidad de una variable aleatoria discreta X es una función, a menudo indicada con $p(x)= P(X=x)$, que satisface:

- $p(x) \geq 0$ para todo x
- La suma de todos los $p(x)$ es 1.

DISTRIBUCIÓN BINOMIAL

- Permite calcular la probabilidad asociada a cada uno de los valores que puede tomar la variable X, bajo el cumplimiento de ciertas condiciones.
- Para su empleo es necesario definir previamente, un tipo particular de experimento, llamado **Ensayo de Bernoulli .**

Es un experimento en el cual sólo hay dos posibles resultados que llamaremos Éxito (E) y Fracaso (F), **con P(éxito)=p y P(fracaso)=q=1-p**

ENSAYO DE BERNOULLI

Ejemplos:

- Un tratamiento médico para cáncer de pulmón está relacionado con quimioterapia. Para cualquier paciente que reciba este tratamiento, existe un 70% de probabilidad de sobrevivir por lo menos 10 años. El éxito se define como “un paciente que sobreviva por lo menos 10 años después del tratamiento”. La probabilidad de sobrevida de al menos 10 años paciente está dada por p=0,70.
- Un grupo de pacientes con sobrepeso están realizando una dieta sin gluten, y existe un 90% de probabilidad de que bajen de peso. El éxito se define como “ un paciente que baje de peso”. La probabilidad de éxito p=0,90.

VARIABLE ALEATORIA BINOMIAL

Una variable aleatoria binomial X es el número de éxitos en n repeticiones independientes de Bernoulli donde cada prueba tiene una probabilidad de éxito p.

DISTRIBUCIÓN BINOMIAL

Propiedades básicas de un modelo Binomial:

- El experimento consiste en n repeticiones idénticas e independientes
- Cada prueba tiene dos posibles resultados (éxito, fracaso).
- La probabilidad de éxito p, permanece constante para cada prueba.
- La variable binomial X es el número de éxitos en n pruebas y se simboliza $X \sim B(n,p)$, donde:

X puede tomar valores 0,1,2….n.

n: es el número de repeticiones independientes

p: es la probabilidad de éxito

DISTRIBUCIÓN BINOMIAL

La función de probabilidad de la distribución binomial:

$$p(x) = P(X = x) = \binom{n}{x} p^x q^{n-x}, \qquad x = 0,1,2, \ldots n$$

Donde :

- p= la probabilidad de éxito
- n= número de repeticiones independientes
- x= número de éxitos en n pruebas

PARÁMETROS ESTADÍSTICOS DISTRIBUCIÓN BINOMIAL

- Esperanza matemática

$$E(X) = n.p$$

- Variancia

$$Var(X) = n.p.(1-p) = n.p.q$$

- Desvío estándar

$$\sigma = \sqrt{n.p.q}$$

DISTRIBUCIÓN BINOMIAL

- Ejemplo1:

De un grupo de 46 pacientes con fiebre 14 tienen coronavirus y 32 no. Se desea seleccionar una muestra aleatoria con reposición de 3 pacientes y registrar el número de personas con coronavirus. ¿Cuál es la probabilidad de que 1 paciente tenga coronavirus?

- El experimento aleatorio es seleccionar a un paciente y registrar si tiene COVID-19 o no.
- Si denominamos "éxito" al hecho de que una persona tenga COVID-19, entonces la situación anterior constituye a un ensayo de Bernoulli con p=P(éxito)=14/46=0,3.
- El experimento aleatorio se repite 3 veces (n=3) y la probabilidad *p* de éxito se mantiene constante en cada uno de los 3 ensayos: p=P(éxito)=0,3.
- Dado que la muestra de personas se selecciona con reposición, los 3 ensayos son independientes entre sí, o sea que el resultado de un ensayo no afecta el resultado de los demás.

DISTRIBUCIÓN BINOMIAL

- En este caso X: es el número de personas que presentan COVID-19. Los valores posibles de X son: {0,1,2,3}
- Esta variable tiene una distribución binomial con la siguiente función de probabilidad:

$$p(x) = P(X = x) = \binom{3}{x}.0{,}3^{x}.(1-0{,}3)^{3-x}$$

A la cual escribimos : $X \sim B(n = 3, p = 0{,}3)$

- Para calcular la probabilidad de que una persona seleccionada tenga coronavirus, se reemplaza x por 1:

$$P(X = 1) = \binom{3}{1}.0{,}3^{1}.0{,}7^{3-1} = 3.0{,}3.0{,}49 = 0{,}441$$

DISTRIBUCIÓN BINOMIAL

Ejemplo 2

Suponga que se tira una moneda al aire 5 veces. Calcular la probabilidad de que salga a lo sumo 2 veces cara.

- Para hallar la probabilidad primero necesitamos identificar el experimento aleatorio

E: tirar una moneda al aire y registrar qué sale.

- El mismo experimento se repite 5 veces, por lo cual n=5.
- En este caso el éxito es que salga cara, por lo cual p=P(C)=0,50.

DISTRIBUCIÓN BINOMIAL

La variable aleatoria en estudio es:

X: número de caras en 5 tiradas de una moneda.

Los posibles valores que puede asumir X son: {0,1,2,3,4,5}

Esta variable tiene una distribución binomial con la siguiente función de probabilidad:

$$p(x) = P(X = x) = \binom{5}{x} . 0{,}5^{x} . (1 - 0{,}5)^{5-x}$$

A la cual escribimos : $X \sim B(n = 5, p = 0{,}50)$

DISTRIBUCIÓN BINOMIAL

La probabilidad de que salga a lo sumo 2 veces salga cara:

$$P(X \leq 2) = \binom{5}{0}.0{,}5^{0}.0{,}5^{5-0} + \binom{5}{1}.0{,}5^{1}.0{,}5^{5-1} + \binom{5}{2}.0{,}5^{2}.0{,}5^{5-2}$$
$$= 0{,}03125 + 0{,}1562 + 0{,}3125 = 0{,}500$$

Otra forma es buscar la probabilidad acumulada en la tabla o usar la aplicación Probability distributions.

TABLA DE DISTRIBUCIÓN BINOMIAL ACUMULADA

n=5

p / x	0,01	0,05	0,10	0,20	0,25	0,30	0,40	0,50	0,60	0,70	0,75	0,80	0,90	0,95	0,99
0	0,9510	0,7738	0,5905	0,3277	0,2373	0,1681	0,0778	0,0313	0,0102	0,0024	0,0010	0,0003	0,0000	0,0000	0,0000
1	0,9990	0,9774	0,9185	0,7373	0,6328	0,5282	0,3370	0,1875	0,0870	0,0308	0,0156	0,0067	0,0005	0,0000	0,0000
2	1,0000	0,9988	0,9914	0,9421	0,8965	0,8369	0,6826	0,5000	0,3174	0,1631	0,1035	0,0579	0,0086	0,0012	0,0000
3		1,0000	0,9995	0,9933	0,9844	0,9692	0,9130	0,8125	0,6630	0,4718	0,3672	0,2627	0,0815	0,0226	0,0010
4			1,0000	0,9997	0,9990	0,9976	0,9898	0,9688	0,9222	0,8319	0,7627	0,6723	0,4095	0,2262	0,0490
5				1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

DISTRIBUCIÓN BINOMIAL

¿Cuál es la probabilidad de que salga más de 3 veces cara?

$$P(X > 3) = P(X \geq 4) = 1 - P(X \leq 3) = 1 - 0{,}8125 = 0{,}1875$$

Busco en la tabla

TABLA DE DISTRIBUCIÓN BINOMIAL ACUMULADA

n=5

p x	0,01	0,05	0,10	0,20	0,25	0,30	0,40	0,50	0,60	0,70	0,75	0,80	0,90	0,95	0,99
0	0,9510	0,7738	0,5905	0,3277	0,2373	0,1681	0,0778	0,0313	0,0102	0,0024	0,0010	0,0003	0,0000	0,0000	0,0000
1	0,9990	0,9774	0,9185	0,7373	0,6328	0,5282	0,3370	0,1875	0,0870	0,0308	0,0156	0,0067	0,0005	0,0000	0,0000
2	1,0000	0,9988	0,9914	0,9421	0,8965	0,8369	0,6826	0,5000	0,3174	0,1631	0,1035	0,0579	0,0086	0,0012	0,0000
3		1,0000	0,9995	0,9933	0,9844	0,9692	0,9130	0,8125	0,6630	0,4718	0,3672	0,2627	0,0815	0,0226	0,0010
4			1,0000	0,9997	0,9990	0,9976	0,9898	0,9688	0,9222	0,8319	0,7627	0,6723	0,4095	0,2262	0,0490
5				1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

DISTRIBUCIÓN BINOMIAL

¿Cuál es la probabilidad de que salga más de una vez cara pero menos de 4?

$$P(1 < X < 4) = P(X \leq 3) - P(X \leq 0) = 0{,}8125 - 0{,}0313 = 0{,}7812$$

Busco en la tabla

PRUEBA DE HIPÓTESIS

Se sabe que la proporción de personas que padecen cáncer de pulmón con supervivencia de cinco años es 0,40. Sin embargo, con un nuevo tratamiento se cree que esta proporción aumentó.

$$H_0)p = 0{,}40 \qquad H_1)p > 0{,}40$$

En una muestra elegida al azar de 15 pacientes bajo este nuevo tratamiento, 7 sobreviven cinco años.

¿Cuál es la probabilidad de observar este valor o uno más extremo?

PRUEBA DE HIPÓTESIS

X: número de personas que sobreviven bajo tratamiento

$$X \sim B(n = 15, p = 0{,}40)$$

$$P(X \geq 7) = 1 - P(X \leq 6) = 1 - 0{,}6098 = 0{,}3902 \longrightarrow \boxed{\text{P-value}}$$

Si el investigador decide trabajar con un nivel de significación del 5%

<u>Regla de decisión</u>: Rechazo H_0 si p-value ≤ 0,05

PRUEBA DE HIPÓTESIS

P-value=0,3902>0,05 No rechazo la hipótesis nula

En base a la evidencia muestral y con un nivel de significación del 5% es de esperar que la proporción de personas enfermas de cáncer de pulmón con supervivencia de cinco años no ha incrementado bajo el nuevo tratamiento.