Distribuciones de probabilidad para variables aleatorias discretas

Héctor de la Torre Gutiérrez
hdelatorreg@up.edu.mx

Variables aleatorias discretas (v.a.d)

X es una v.a.d si su soporte (valores posibles) es finito o a lo más infinito contable (numerable).

Experimento	*v.a. X*	*Valores posibles*
Llamar a cinco clientes	Cantidad de clientes que han pedido	0,1,2,3,4,5
Inspeccionar un embarque de 50 radios	Cantidad de radios defectuosos	0,1,2,3,...,50
Funcionamiento de un restaurante durante un dia	Cantidad de clientes	0,1,2,...
Vender un automovil	Sexo del cliente	0 si es hombre,1 si es mujer
Inspeccionar artículos de una línea de producción	número de artículos inspeccionados hasta encontrar 5 defectuosos	5, 6, . . .

fmp (fdp) y fpa (FD)

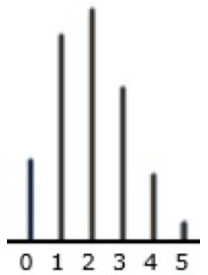
La fmp o fdp caracteriza una v.a.d "*X*" que toma un valor particular "$x \in S_X$", es decir $P(X=x)$, se denota como: $P_X(x)$ o $f_X(x)$. Y debe de cumplir tres condiciones:

- $P(X = x) = f(x) > 0,$
- $\sum_{x \in S} f(x) = 1$
- $P(X \in A) = \sum_{x \in A} f(x)$

Dado que fmp es una fn se puede representar:

- En forma tabular
- En forma gráfica
- En forma matemática

x	$f(x)$
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$



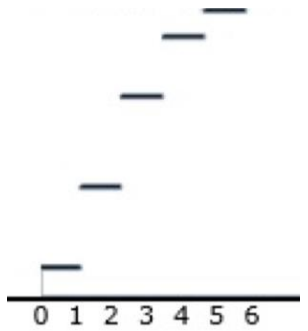
Masa de Probabilidad

$$f(x;p) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \end{cases}$$

La fpa o FD de una v.a.d "*X*" es la probabilidad de que "*X*" sea menor o igual a un valor particular t ϵ S_X , está dada por: $F_X(t) = P_X(x \leq t) = \sum_{x \leq t} f_X(t)$ y de cumplir:

- $\lim_{x \to -\infty} F(x) = 0$ y $\lim_{x \to -\infty} F(x) = 1$
- F(x) es una funcion no decreciente
- F(x) es continua por la derecha:

$$\forall h > 0, \lim_{h \to -\infty} F(x+h) = F(x)$$



Valor esperado de una v.a

Representa el punto de equilibrio de una distribución de datos (i.e es una medida de tendencia central que coincide con el promedio cuando se trabaja a nivel muestral).

$$E_X(x) = \sum_{x \in S_X} x f_X(x) = \mu_X$$

Ejemplo. Definamos la v.a X como el número que aparece en la cara superior cuando lanzamos un dado, entonces la fmp=⅙ si x=1,2,3,4,5,6 y, 0 de otra forma.
Su valor esperado es:

$$E(X) = 1\left(\frac{1}{6}\right) + 2\left(\frac{1}{6}\right) + 3\left(\frac{1}{6}\right) + 4\left(\frac{1}{6}\right) + 5\left(\frac{1}{6}\right) + 6\left(\frac{1}{6}\right) = 3.5$$

La varianza indica qué tanto se dispersan los datos respecto a su media.
Los puntos con mayor varianza tendrán más dispersión (incertidumbre)

$$Var(X) = E(X - E(X))^2 \qquad Var(X) = \sigma^2 = \sum (x - \mu)^2 p(x)$$

Modelos de Probabilidad Discretos (MPD)

- Ahora es importante revisar algunos MDP básicos:
  - Uniforme
  - Bernoulli
  - Binomial
  - Poisson

- Otras distribuciones discretas importantes son: Geométrica, Binomial Negativa, por mencionar algunas.

MPD- uniforme

Dfn. Cada uno de sus valores tiene la misma probabilidad de ocurrir.

Ejemplo. El experimento consiste en lanzar un dado. El resultado de interés es el número de la cara superior del dado, la cual se denota con la variable aleatoria *X*, *¿Cúal es el MPD asociados con X?*, *respuesta:*

x	$P(X=x)$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

x	$F(X \leq x)$
1	$\frac{1}{6}$
2	$\frac{2}{6}$
3	$\frac{3}{6}$
4	$\frac{4}{6}$
5	$\frac{5}{6}$
6	$\frac{6}{6}$

$$P(X=x)=\frac{1}{6};x=1,2,3,\dots,6$$

$$P(X=x)=\frac{1}{n},X=x_1,x_2,x_3,\dots,x_n$$

MPD - Bernoulli

- El modelo de probabilidad más simple (quizá) es el modelo Bernoulli.
- Este modelo tiene como base los famosos eventos de éxito y fracaso (variable binaria), que bien pueden referirse a varias aplicaciones básicas, como:
  - Volados (águila o sol)
  - Salud (sano, enfermo)
  - Género (hombre, o mujer)
  - Deportes (gana, o pierde un equipo favorito)
- Los valores que se asocian a una distribución Bernoulli suelen ser 0 y 1. El 1 usualmente se refiere al éxito (evento que se desea medir)

MPD - Bernoulli

- Para las variables mencionadas, por ejemplo, se pueden asociar:
  - Águila **1**, Sol **0** (Probabilidad de Águila)
  - Sano **0**, Enfermo **1** (probabilidad de estar enfermo)
  - Hombre **0**, Mujer **1** (probabilidad de ser mujer)
  - Ganar **1**, Perder **0** (probabilidad de que un equipo gane un partido)
- Los valores 0,1 se asocian a una variable X de interés; es decir, la "variable aleatoria".
- El modelo asociado a este tipo de evento está dado por:

$$P(X = 1) = p^x(1 - p)^{1-x} = p^x q^{1-x}$$

donde $x = 0,1$; $p = probabilidad\ de\ éxito$.

MPD - Bernoulli

- Es interesante notar que los modelos de probabilidad también tienen valor esperado y varianza. En este caso, si se aplica la definición, en la distribución Bernoulli, se tiene que:

$$E(X) = p$$
$$Var(X) = p(1-p)$$

- La parte más interesante de la distribución ocurre cuando se usa como base en otros métodos relacionados con la distribución como "regresión logística".

•Por ejemplo, determinar los factores de riesgo asociados a la presencia o no de diabetes mellitus gestacional dadas condiciones de salud y demográficas de la madre. https://authors.elsevier.com/sd/article/S2605073023000093

•En demografía, la probabilidad de que un habitante con condiciones, por ejemplo, marginación, de una etnia específica, en edad escolar, etc., reciba una beca de apoyo estudiantil.

MPD - Binomial

- Esta distribución tiene como origen la distribución Bernoulli, a través de eventos repetidos y se desea conocer cuantos eventos favorables hay en n eventos realizados.
- Esta distribución se genera como una suma de **variables aleatorias independientes Bernoulli con el mismo valor de** p.
- Por ejemplo:
  - De 20 volados cual es la probabilidad de que se den más de 10 águilas.
  - De 10 partidos de futbol entre México y Brasil, cual es la probabilidad de que México gane al menos 3.
  - De 50 pacientes revisados cual es la probabilidad de que 5 o menos estén infectados del “estomago”.
  - De 100 nuevos nacimientos de vacas, cual es la probabilidad de que nazcan más 70 hembras.

•**Ejercicio:**

¿De 4 lanzamientos de monedas, de cuántas maneras ocurren 2 águilas?

Las posibles combinaciones son:

AAAA	AAAS	AASA	ASAA
SAAA	AASS	SSAA	ASAS
SASA	ASSA	SAAS	ASSS
SASS	SSAS	SSSA	SSSS

- Las combinaciones de k tomadas de n eventos realizados – es decir, de n eventos realizados de cuantas maneras salen k favorables – están dadas por:

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

Donde: $n! = n(n-1)(n-2)\ldots(2)(1)$ n factorial

- La fórmula de cálculo de las probabilidades binomiales esta dada por:

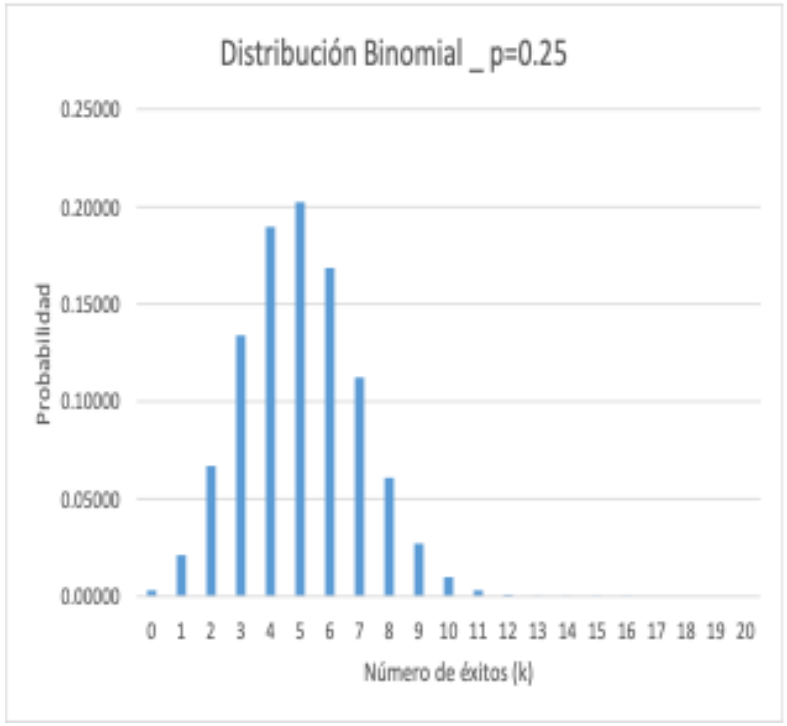
$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

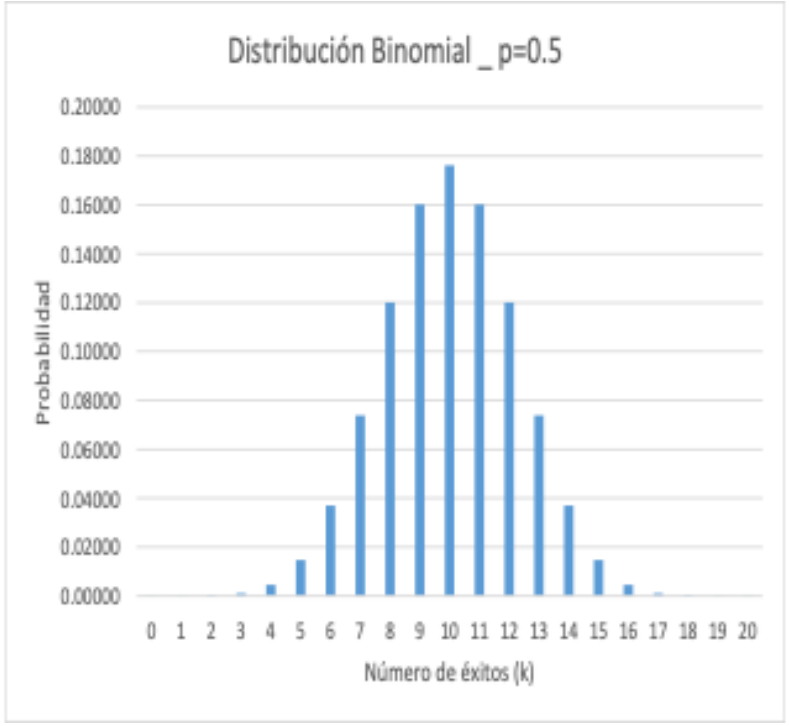
donde $k = número\ de\ éxitos$, el cual toma valores desde 0 hasta n; además el valor de p **debe ser mayor** a 0 **y menor** a 1 (intervalo abierto).

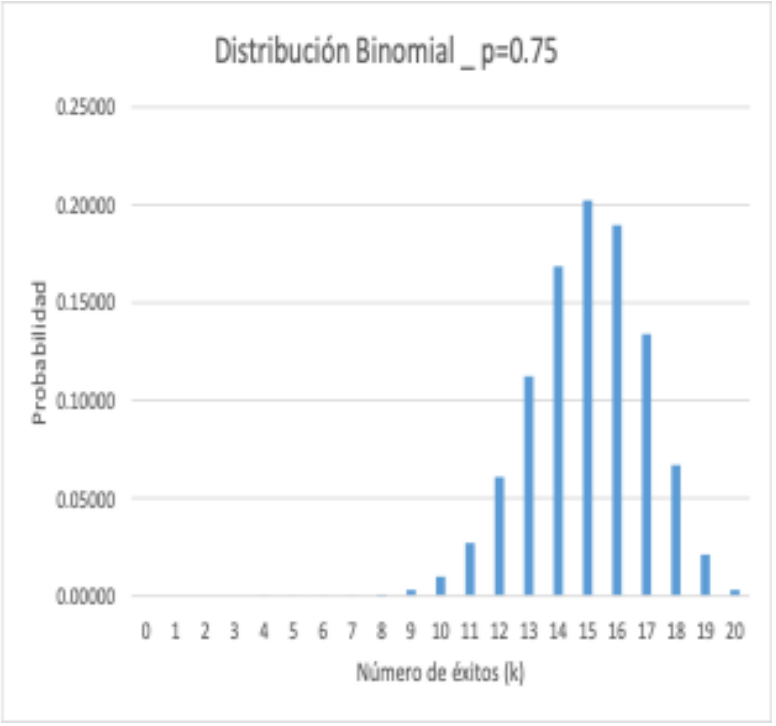
- Este es un detalle muy importante ya que, si p llega a valer 0 o 1, la distribución no tiene sentido (degenerada).
- Los parámetros de la distribución son n y p.
- Respecto al valor esperado y la varianza son:

$$E(Y) = np,$$
$$Var(Y) = np(1-p)$$

Ejemplos:


Distribución Binomial _ p=0.25
Probabilidad
0.25000
0.20000
0.15000
0.10000
0.05000
0.00000
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Número de éxitos (k)


Distribución Binomial _ p=0.5
Probabilidad
0.20000
0.18000
0.16000
0.14000
0.12000
0.10000
0.08000
0.06000
0.04000
0.02000
0.00000
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Número de éxitos (k)


Distribución Binomial _ p=0.75
Probabilidad
0.25000
0.20000
0.15000
0.10000
0.05000
0.00000
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
Número de éxitos (k)

Ejercicio:

En una oficina de gobierno, se presume que el 30% es fumador; ¿cuál es la probabilidad de que en una muestra de 50 personas se identifiquen al menos 10 fumadores y a lo más 40?

Solución:

Se pregunta encontrar desde 10 hasta 40 fumadores; Entonces la solución es encontrar las probabilidades desde 10 hasta 40.

Con p=0.3 y n=50, se estiman las probabilidades desde p(X=10) hasta p(X=40) y se suman; es decir, P(X=10)+P(X=11)+…+P(X=40) lo que da como resultado 0.959768366≃96%.

MPD - Poisson

- La distribución Poisson, se usa comúnmente para estimar probabilidades relacionadas con eventos poco frecuentes; de ahí que también se le conoce como distribución de fallas o de errores.
- Su definición aplica a datos por unidad de tiempo o espacio, cuando el número de caso es pequeño (raro).
- Surge también como aproximación a la distribución binomial cuando n es grande, p muy pequeña y $np = \theta$ se mantiene, al menos de manera aproximada, constante.
- Una diferencia notable respecto a la distribución binomial es que la variable aleatoria puede tomar valores desde 0 hasta infinito.

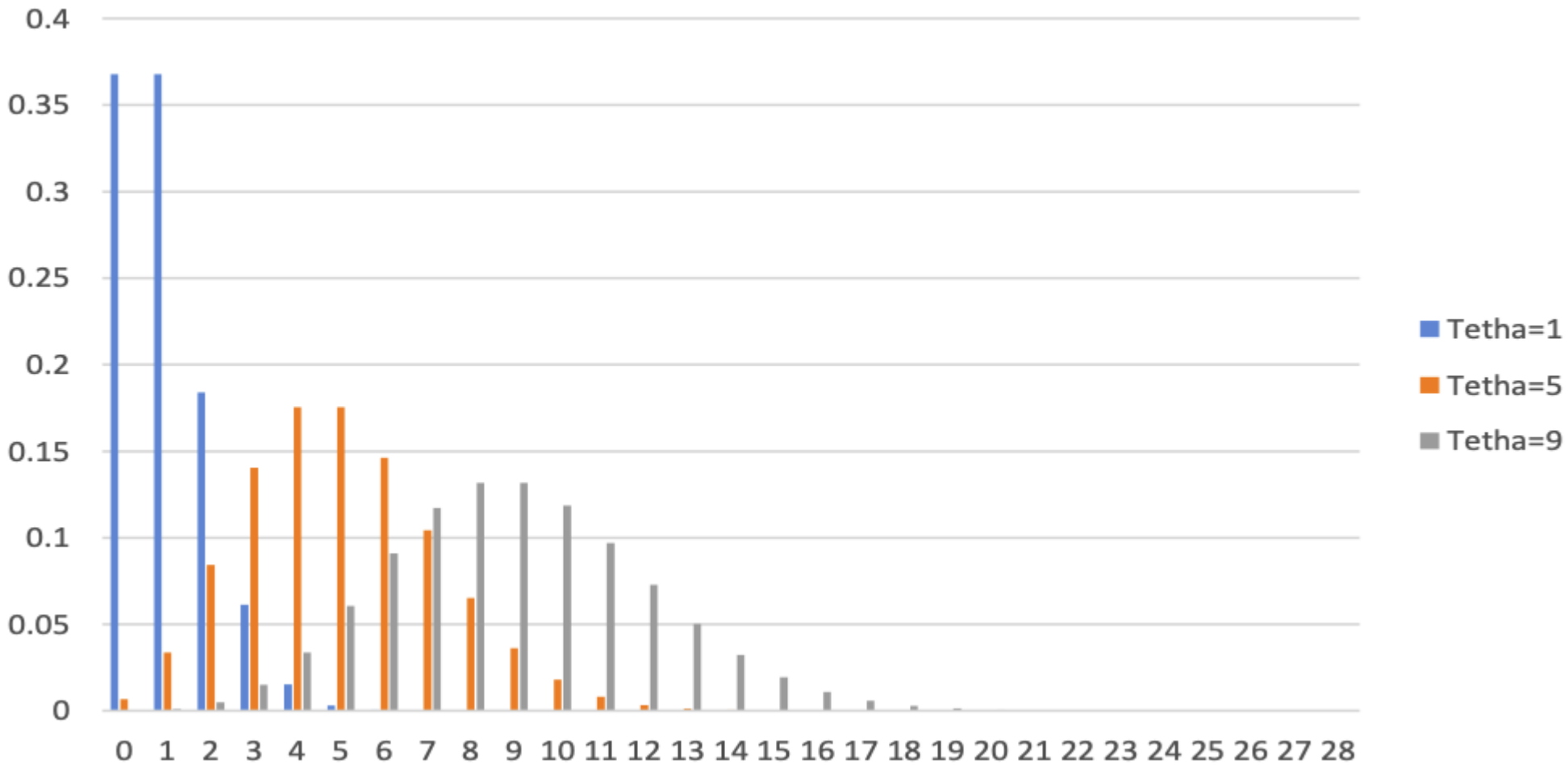
Aplicación: Pequeños productores de maíz en el Caribe colombiano
https://revistacta.agrosavia.co/index.php/revista/article/view/556/458

- La función de probabilidad Poisson esta dada por:

$$P(X = k) = \frac{e^{-\theta}\theta^{k}}{k!}$$

donde $k = 0,1,2, \ldots\ldots$ con $\theta > 0$; θ es el parámetro de la distribución.

- Algunas aplicaciones de la distribución son:
  - Ecología: número de unidades inspeccionadas con un atributo específico (por unidad de área); por ejemplo, árboles enfermos por hectárea
  - Demografía: cantidad de migrantes de una nacionalidad por mes.
  - Industria: número de lotes de un producto con defecto, producidos por día (focos, transistores, etcétera).
  - Negocios: Clientes que compran automóvil último modelos (por mes).


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
Tetha=1
Tetha=5
Tetha=9

- Para la distribución Poisson, el valor esperado está dado por:

$$E(X) = \sum_{k=0}^{\infty} k \frac{e^{-\theta}\theta^k}{k!} = \theta$$

$$Var(X) = \sum_{k=0}^{\infty} (k-\theta)^2 \frac{e^{-\theta}\theta^k}{k!} = \theta$$

- ¡Es decir, la media y la varianza son iguales!; esta propiedad es particular de la Distribución Poisson.
- En general, para distribuciones discretas, la razón entre la varianza y el valor esperado se conoce como **índice de dispersión**; cuando es menor a 1, se dice que la distribución tiene un patrón agregado; cuando es mayor a 1 tiende a ser sistemático (0 disperso) y cuando es 1 se el patrón es aleatorio. Este es un tema muy discutido en estadística espacial (geo-estadística).

Aplicación: Modelación estadística

Consiste en la utilización de métodos estadísticos, computacionales y el conocimiento del experto de los fenómenos bajo estudio para describir, explicar y/o predecir el comportamiento de un fenómeno. Ayuda a la toma de decisiones o estrategias de atención del fenómeno.

“All models are wrong, but some are useful”... George Box

El 70% de los estudios “reales” se pueden abordar como problemas de clasificación y sólo el 30% como regresión.

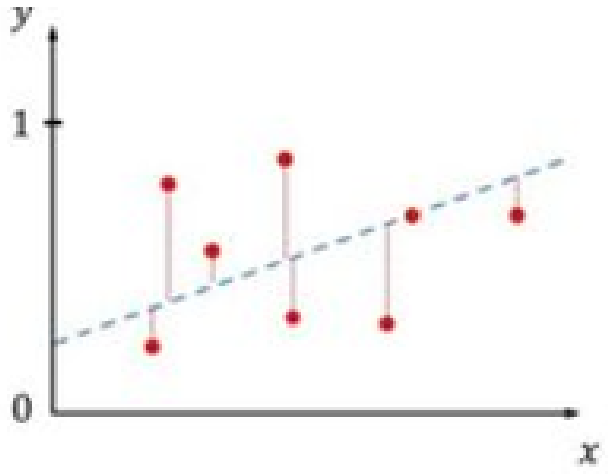
Modelo estadístico

Y= todos los factores que inciden en la explicación Y

Y= factores observados $(x\_1, x\_2, \ldots, x\_k)$ + factores no observados (ε).

$Y=E(Y/[x\_1,x\_2,\cdots,x\_k])+\varepsilon=\mu(Y/[x\_1,x\_2,\cdots,x\_k])+\varepsilon$

$Y=\beta\_0+\beta\_1 * x\_1+\ldots+\beta\_k * x\_k+\varepsilon$ Cuando tenemos RLM



- Sea $Y \sim Bernoulli(p), p \in [0,1]$; es decir, sólo puede tomar los valores 0 y 1, la cual es explicada por un conjunto de predictoras $(x_1, x_2, \ldots, x_k)$, entonces:

$$E(Y) = 0 \cdot P(Y=0) + 1 \cdot P(Y=1) = P(Y=1)$$

$$p = P(Y=1|x) = \pi(X) = \frac{e^{B_0+\beta_1 x}}{1+e^{B_0+\beta_1 x}} = \frac{1}{1-e^{B_0+\beta_1 x}} \quad \ln\left[\frac{P(Y|x)}{1-P(Y|x)}\right] = \beta_0 + \beta_1 x$$