

CURSO:

# **METODOS ESTADÍSTICOS BÁSICOS Y TÉCNICAS DE MUESTREO**

PROFESOR: LUIS E. NIETO BARAJAS

EMAIL: lnieto@itam.mx

URL: <http://allman.rhon.itam.mx/~lnieto>

*Instituto Nacional de Ecología, INE*

## **Curso: Métodos estadísticos básicos y técnicas de muestreo**

- **OBJETIVO:** Brindar al alumno los conocimientos básicos de Estadística tanto de carácter descriptivo como de carácter inferencial, así como mostrar las principales técnicas de muestreo.

- **TEMARIO:**

### **Parte 0. Introducción**

### **Parte 1. Probabilidad y Estadística descriptiva**

- 1.1 Análisis exploratorio de datos
- 1.2 Conceptos básicos de probabilidad
- 1.3 Variables aleatorias y distribuciones de probabilidad
- 1.4 Algunas distribuciones específicas de probabilidad

### **Parte 2. Estadística inferencial**

- 2.1 Distribuciones muestrales
- 2.2 Estimación puntual
- 2.3 Estimación por intervalos
- 2.4 Pruebas de hipótesis

### **Parte 3. Técnicas de muestreo**

- 3.1 Introducción al muestreo
- 3.2 Muestreo aleatorio simple (MAS)
- 3.3 Muestreo aleatorio estratificado (MAE)
- 3.4 Muestreo aleatorio por conglomerados (MAC)
- 3.5 Otras técnicas de muestreo

➤ REFERENCIAS:

- ❑ Aguirre, V., et al. (2003). Fundamentos de probabilidad y estadística. Jit Press: México.
- ❑ Bart, J., Fligner, M. A. & Notz, W.I. (1998). Sampling and statistical methods for behavioral ecologists. Cambridge University Press: Cambridge.
- ❑ Scheaffer, R. L., Mendenhall, W. & Ott, L. (1987). Elementos de muestreo. Grupo editorial Iberoamérica: México.

➤ PAQUETES ESTADÍSTICOS: Los conceptos y las técnicas se ejemplificarán mediante el uso de los siguientes paquetes (estadísticos):

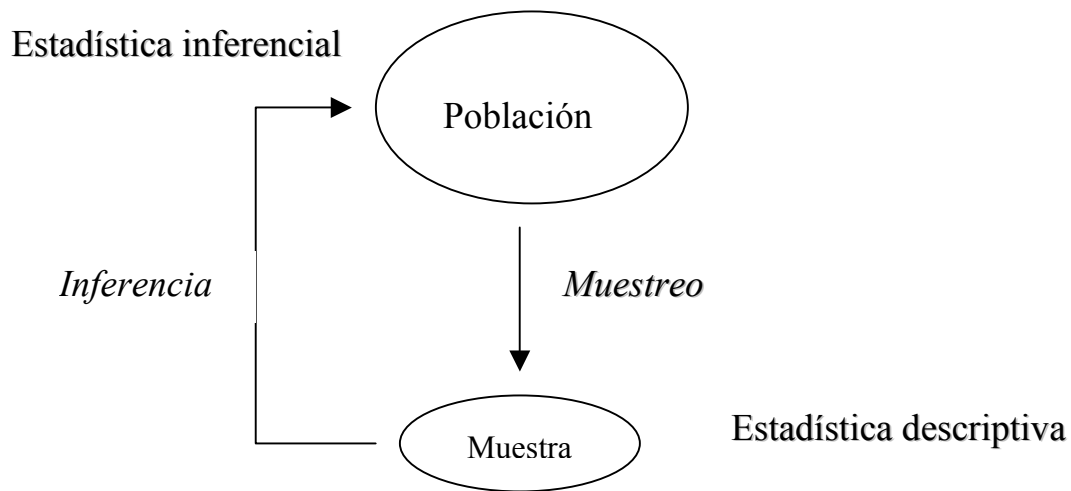
- ❑ Excel
- ❑ SPSS

➤ EVALUACIÓN: Al final del curso se aplicará un examen en el que el alumno tendrá que demostrar su comprensión de los conceptos enseñados y su capacidad de aplicarlos para la solución de problemas reales.

## Parte 0. Introducción

- ¿Qué es la ESTADÍSTICA? La estadística es una ciencia que tiene que ver con la recolección, presentación, análisis e interpretación de datos, con el fin de ayudar en el proceso de toma de decisiones.
- Los procesos estadísticos se dividen en dos grandes ramas: Estadística descriptiva y Estadística inferencial.
  - ❑ *Estadística descriptiva*: Proceso de recolectar, organizar y presentar datos de manera tal que describa fácil y rápidamente dichos datos para facilitar su entendimiento.
  - ❑ *Estadística inferencial*: Involucra el análisis y la interpretación de los datos para ayudar en el proceso de la toma de decisiones.
- ¿Dónde entra el muestreo?
  - ❑ *Técnicas de muestreo*: Los procesos de recolección de datos son parte de la estadística descriptiva y son tan variados e importantes que se les ha dado su propio lugar en la estadística.
- En resumen, el OBJETIVO de la estadística, es proporcionar una metodología para recolectar información (*muestreo*), analizar adecuadamente esa información con la que se cuenta (*estadística descriptiva*) y decidir de manera razonable sobre la mejor forma de actuar (*estadística inferencial*).

➤ **DIAGRAMA de la Estadística:**



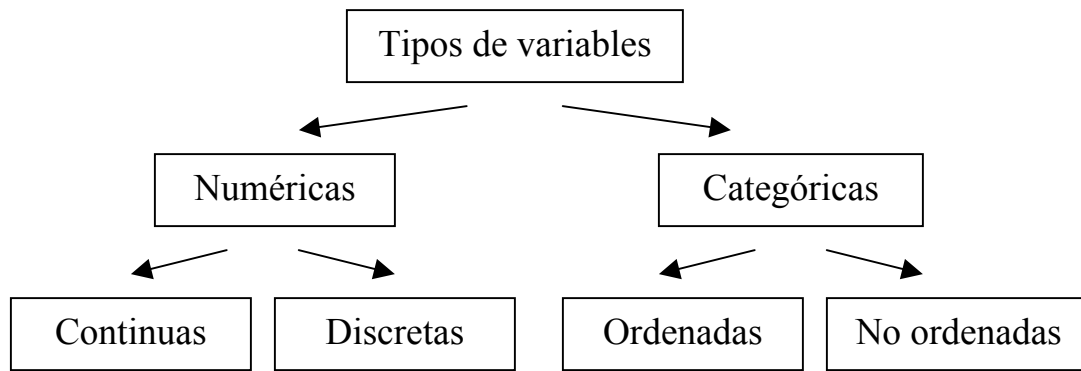
➤ **¿Qué es INFERENCIA?**

Es el proceso de tratar de conocer cualquier característica poblacional a través de una muestra. Para poder entender propiamente lo qué es inferencia es necesario definir lo siguiente,

➤ **Algunas definiciones básicas:**

- ❑ *Elemento o individuo*: Objeto (persona, empresa, animal, planta, etc.) sobre el cuál se toma una medición de cierta característica de interés.
- ❑ *Población*: Conjunto de elementos de interés para el investigador.
- ❑ *Muestra*: Subconjunto representativo de elementos de la población, es decir, que la muestra refleje las características esenciales de la población.
- ❑ *Parámetro*: Característica numérica de la población (medida resumen).
- ❑ *Estadística(o)*: Característica numérica de la muestra.

➤ **VARIABLE**: Característica o atributo que se le mide a un individuo.



- EJEMPLO 1: Supongamos que se desea conocer el consumo de agua promedio por vivienda a nivel nacional.

*Individuos:* viviendas.

*Población:* conjunto de todas las viviendas del país.

*Muestra:* 1) todas las viviendas del D.F. (¿será representativa?)

2) diez viviendas de cada estado de la república seleccionadas de manera aleatoria. (¿será representativa?).

*Parámetro:* consumo de agua promedio nacional (por vivienda).

*Estadística:* promedio de los consumos de agua de las viviendas en la muestra.

*Variable:* Consumo de agua (de tipo numérica continua).

- ¿Cómo obtener una muestra representativa?

La respuesta a esta pregunta se verá con detalle en la Parte 3 del curso. Por el momento diremos lo siguiente:

- ❑ Una muestra representativa se obtiene seleccionando de manera aleatoria (al azar) un subconjunto de individuos de la población de interés.
- ❑ Si la población es pequeña o de tamaño regular (*población finita*) la selección se hará con reemplazo, mientras que

- Si la población es muy grande (*población infinita*) la selección se hará con o sin reemplazo.
- La forma de seleccionar a los individuos (con o sin reemplazo), de acuerdo al tamaño de la población, es para garantizar la independencia entre ellos.
- INDIVIDUOS INDEPENDIENTES: El hecho de haber seleccionado a un individuo en la muestra no afecta la probabilidad de selección del otro individuo en la muestra.
- Azar  $\Rightarrow$  Representatividad
- Misma probabilidad de selección de cada individuo  $\Rightarrow$  Independencia

## Parte 1. Probabilidad y Estadística descriptiva

### 1.1 Análisis exploratorio de datos

- Supongamos que la labor de recolección de datos ya ha sido realizada y que los datos (la muestra) ya están disponibles. Denotemos por  $X_1, X_2, \dots, X_n$  una muestra de observaciones de la variable de interés, donde cada  $X_i$  representa la característica de interés para el individuo  $i$ .
- Las técnicas exploratorias se dividen en técnicas gráficas y medidas descriptivas.

### Descripción gráfica

- VARIABLES CATEGÓRICAS (O CUALITATIVAS):

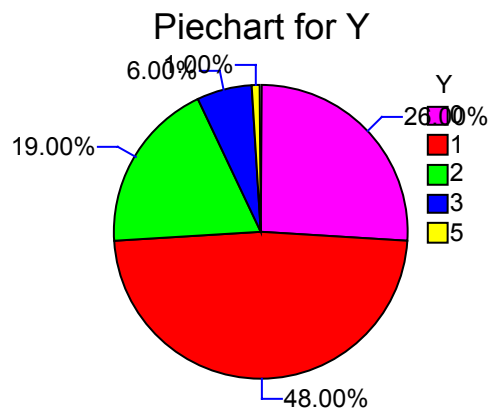
- 1) *Distribución de frecuencias*: es una tabla que muestra todas las posibles categorías que puede tomar la variable junto con el número de veces que cada una de ellas es observada en la muestra. Las frecuencias (número de ocurrencias) pueden ser absolutas ( $f$ ) o relativas ( $fr$ ).

Frequency Table for Y

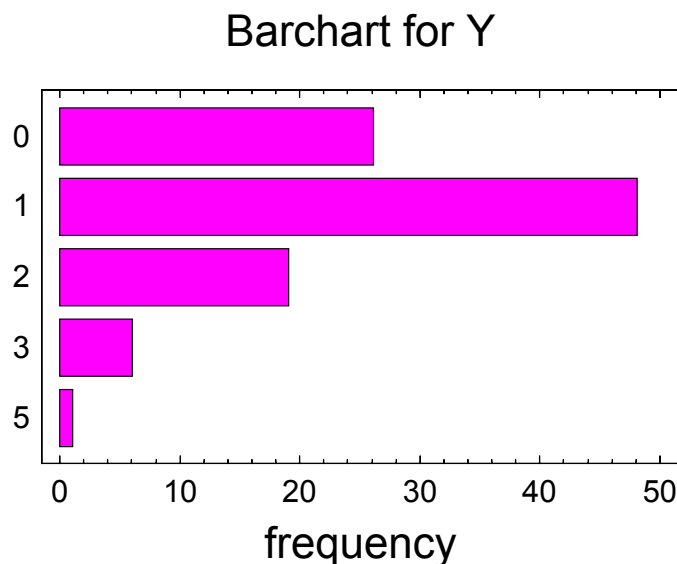
Class	Value	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
1	0	26	0.2600	26	0.2600
2	1	48	0.4800	74	0.7400
3	2	19	0.1900	93	0.9300
4	3	6	0.0600	99	0.9900
5	5	1	0.0100	100	1.0000



- 2) *Diagrama circular (o de pastel)*: Es una representación gráfica de las frecuencias relativas de las categorías como un todo. Se construye dividiendo al círculo en el número de categorías de la variable y cada parte tiene el tamaño correspondiente a su frecuencia relativa.



- 3) *Diagrama de barras*: es una representación gráfica de la distribución de frecuencias absolutas o relativas, en donde cada frecuencia es representada mediante una barra (puede ser horizontal o vertical).



➤ VARIABLES NUMÉRICAS (O CUANTITATIVAS):

- 1) *Diagrama de tallo y hojas*: Es una gráfica que nos da una idea de la distribución (dispersión y acumulación) de los datos. Se construye dividiendo convenientemente a cada observación en dos partes, un tallo y una hoja y se colocan de forma ordenada de menor a mayor separando el tallo de la hoja mediante una recta vertical.

Stem-and-Leaf Display for X: unit = 0.1    1|2 represents 1.2

```

1      -2|5
2      -2|2
6      -1|9755
21     -1|443333332222100
39     -0|999888877766666555
(16)   -0|4444433322222111
45      0|0001112333444
32      0|55556667777888899
15      1|0123444
8       1|56666789

```

- 2) *Distribución de frecuencias*: Es una tabla que muestra los posibles valores que puede tomar la variable junto con sus frecuencias.
  - Si la variable es discreta, la distribución de frecuencias se construye de la misma manera que la distribución de frecuencias para variables categóricas.
  - Si la variable es continua se crean intervalos de clase disjuntos (no se intersecten) y exhaustivos (cubran todos los posibles valores de la variable). Finalmente se calculan las frecuencias absolutas (o relativas) contando el número de observaciones en la muestra que pertenecen a cada uno de los intervalos de clase.

Frequency Tabulation for X

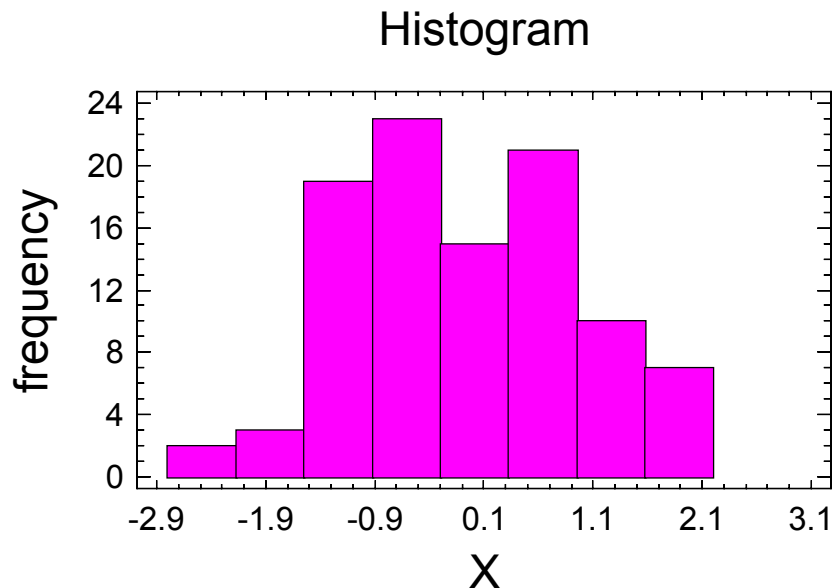
Class	Lower Limit	Upper Limit	Midpoint	Frequency	Relative Frequency	Cumulative Frequency	Cum. Rel. Frequency
at or below		-2.8		0	0.0000	0	0.0000
1	-2.8	-2.175	-2.4875	2	0.0200	2	0.0200
2	-2.175	-1.55	-1.8625	3	0.0300	5	0.0500
3	-1.55	-0.925	-1.2375	19	0.1900	24	0.2400
4	-0.925	-0.3	-0.6125	23	0.2300	47	0.4700
5	-0.3	0.325	0.0125	15	0.1500	62	0.6200
6	0.325	0.95	0.6375	21	0.2100	83	0.8300
7	0.95	1.575	1.2625	10	0.1000	93	0.9300
8	1.575	2.2	1.8875	7	0.0700	100	1.0000
above	2.2			0	0.0000	100	1.0000

Mean = -0.104884    Standard deviation = 1.03239

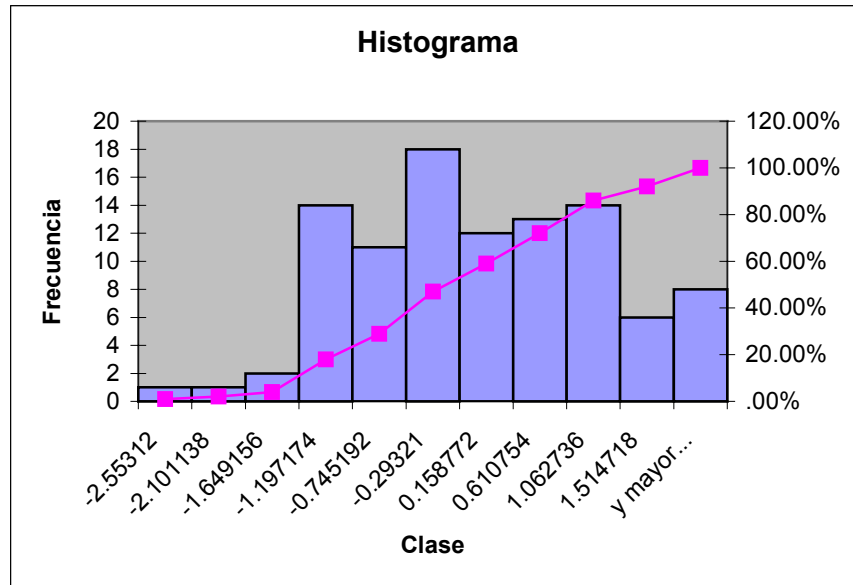
3) *Histogramas*: Es la representación gráfica de una distribución de frecuencias. Un histograma es una forma especial de una gráfica de barras en la cual los intervalos de clase son representados por el ancho de las barras y las frecuencias (relativas) son proporcionales al área de los rectángulos.

- Existe una relación directa entre los histogramas y las curvas poblacionales, en general esperaríamos que la distribución de frecuencias de una muestra nos sugiera un perfil similar al de la población de la cual se tomó. Por lo tanto la curva poblacional se puede ver como una versión suavizada de un histograma pero considerando a toda la población: Las curvas poblacionales pueden tener distintos tipos de formas: Simétrica, sesgada o asimétrica, unimodal, bimodal.
- PRECAUCIÓN: Vale la pena hacer notar que tanto la distribución de frecuencias para variables numéricas continuas como su representación gráfica (histograma) pueden ser engañosos debido a la libertad de decidir el número de clases y los anchos de los

intervalos de clase a conveniencia del investigador, pudiéndose producir distintos tipos de comportamiento.



- 4) *Polígonos de frecuencias*: Es un perfil de la distribución de frecuencias. Se construye suavizando el histograma uniendo los puntos medios de la parte superior de las barras y cerrando los extremos con el eje horizontal. El polígono de frecuencias nos da una buena aproximación a la distribución poblacional de los datos.
- 5) *Ojiva*: Es la curva que resulta de graficar las frecuencias relativas acumuladas contra las fronteras superiores de cada intervalo de clase. La ojiva sirve para calcular frecuencias cumuladas de una manera rápida.



## Medidas descriptivas

- Las medidas descriptivas se clasifican en varios tipos: de tendencia central, de posición o localización y de dispersión.
- MEDIDAS DE TENDENCIA CENTRAL: Las medidas de tendencia central son valores numéricos, que tienden a localizar la parte central de un conjunto de datos o de la distribución de frecuencias. Las tres más importantes son:
  - *Media*: es un promedio aritmético de las observaciones.

$$\mu = \frac{1}{N} \sum_{i=1}^N X_i = \text{Media poblacional}$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \text{Media muestral}$$

La media **no es un buen indicador** de tendencia central cuando la distribución de los datos es sesgada (asimétrica).

- *Mediana*: es la observación que se encuentra justo a la mitad de un conjunto de datos una vez que se han ordenado. La mediana poblacional se denota como “M” y la mediana muestral como “m”.

$$l = n \times 0.5 + 0.5 = \text{posición de la mediana}$$

$$m = X_{(l)} = \text{mediana (observación } X \text{ que se encuentra en la posición } l \text{ después de ordenar los datos).}$$

La mediana **es un buen indicador** de la tendencia central cuando la distribución de los datos es sesgada.

- *Moda*: La moda u observación modal es aquella observación que ocurre con mayor frecuencia en un conjunto de datos. Si este valor es único diremos que la distribución de frecuencias es unimodal.

➤ **MEDIDAS DE POSICIÓN O DE LOCALIZACIÓN**: Las medidas de posición son llamadas cuantiles o percentiles. Para  $p \in (0,1)$ , el p-ésimo percentil es la observación que divide al conjunto de datos de tal manera que  $p \times 100\%$  de las observaciones son menores y  $(1-p) \times 100\%$  son mayores. Los percentiles más comunes son:

- *Cuartiles*: son observaciones que dividen al conjunto de datos ordenados en 4 partes iguales.

$$Q_1 = X_{(n \times 0.25 + 0.5)} = \text{Primer cuartil o cuartil inferior}$$

$$Q_2 = X_{(n \times 0.50 + 0.5)} = \text{Segundo cuartil o mediana}$$

$$Q_3 = X_{(n \times 0.75 + 0.5)} = \text{Tercer cuartil o cuartil superior}$$

- **MEDIDAS DE DISPERSIÓN:** Las medidas de dispersión, como su nombre lo indica, son medidas de la dispersión (variabilidad, concentración, separación) de un conjunto de datos. Las medidas más comunes son:

- *Amplitud o rango:* es la medida más simple e indica la separación entre la observación más pequeña y las más grande.

$$R = \text{Máximo} - \text{mínimo} = \text{rango}$$

- *Amplitud intercuartílica:* es la distancia que existe entre el primer y tercer cuartiles.

$$AIC = Q_3 - Q_1 = \text{amplitud intercuartílica}$$

- *Varianza:* es un promedio de las desviaciones al cuadrado de cada una de las observaciones con respecto a su media.

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2 = \text{varianza poblacional}$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left( \sum_{i=1}^n X_i^2 - n\bar{X}^2 \right) = \text{varianza muestral}$$

A la raíz cuadrada de la varianza se le conoce como desviación estándar, i.e.,

$$\sigma = \sqrt{\sigma^2} = \text{desviación estándar poblacional}$$

$$S = \sqrt{S^2} = \text{desviación estándar muestral}$$

- *Coefficiente de variación:* mide la dispersión relativa de un conjunto de datos con respecto a su localización.

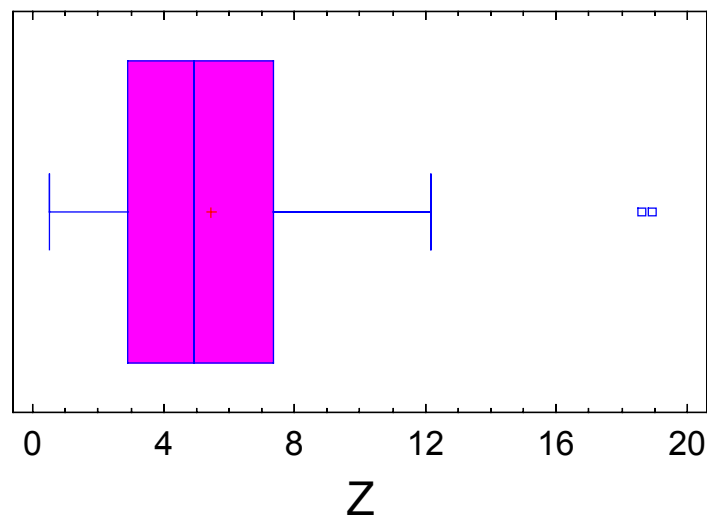
$$CV = \frac{\sigma}{\mu} = \text{coeficiente de variación poblacional}$$

$$cv = \frac{S}{\bar{X}} = \text{coeficiente de variación muestral}$$

Este coeficiente es muy útil para comparar la variabilidad entre dos conjuntos de datos, debido a que no depende de las unidades de medición.

- **DIAGRAMA DE CAJA Y BRAZOS:** Los diagramas de caja y brazos resumen mediante una gráfica las características más importantes de un conjunto de datos, como son la localización y dispersión. Además permiten observar características de la distribución como simetría o asimetría y presencia de observaciones atípicas.

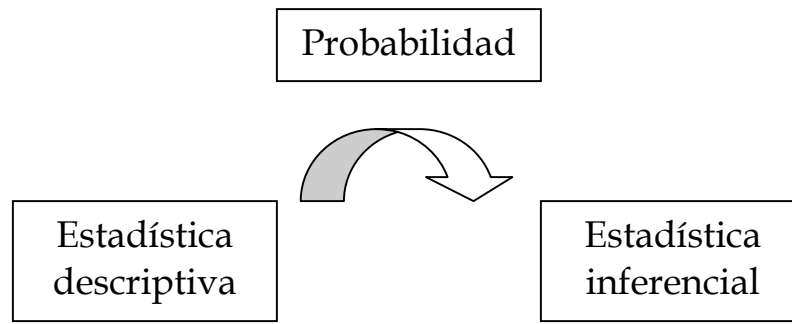
Box-and-Whisker Plot





## 1.2 Conceptos básicos de probabilidad

- Los conceptos básicos de probabilidad nos ayudan a establecer un puente de enlace entre la Estadística descriptiva y la Estadística inferencial.



- La probabilidad surge por la necesidad de cuantificar la incertidumbre sobre la ocurrencia de ciertos eventos que no se pueden controlar ni saber con certeza sus resultados (*Juegos de azar*).
- Cotidianamente escuchamos comentarios en donde se expresa incertidumbre:
  - “yo creo que hoy lloverá por la tarde”
  - “yo creo que la tasa de interés va a subir”
- La probabilidad va más allá, cuantifica la creencia de la ocurrencia de eventos cuyos resultados son inciertos. Por ejemplo:
  - “existe una probabilidad del 80% de que hoy lloverá por la tarde”
  - “hay una probabilidad del 70% de que la tasa de interés suba, 10% de que se mantenga igual y 20% de que baje”

**La probabilidad es una medida de la incertidumbre**

- Algunos conceptos y definiciones importantes para entender la probabilidad son:
  - *Experimento*: es un proceso mediante el cuál se obtiene una observación o un dato.
  - *Experimento aleatorio*: es un experimento cuyos resultados no pueden predecirse (son inciertos). Al menos debe de tener dos posibles resultados.
  - *Espacio muestral* ( $\Omega$ ): es el conjunto de todos los resultados posibles (resultados elementales) de un experimento aleatorio.
  - *Evento*: es un subconjunto de un espacio muestral. Un evento está formado por uno o varios resultados elementales. Se denotan por letras latinas mayúsculas.
  
- EJEMPLO 2: Un experimento aleatorio consiste en lanzar una moneda honesta.  
 $\Omega = \{\text{sol}, \text{águila}\}$   
 $A = \{\text{sol}\}, B = \{\text{águila}\}$
  
- EJEMPLO 3: Un experimento aleatorio consiste en lanzar un dado con 6 caras.  
 $\Omega = \{1, 2, 3, 4, 5, 6\}$   
 $A = \{1, 3, 5\}$  = el valor observado en un número non  
 $B = \{3, 4, 5, 6\}$  = el valor observado es mayor a 2
  
- EJEMPLO 4: Un experimento aleatorio consiste en revolver una baraja de 52 cartas y repartir 5 cartas.

$$\Omega = \{(w_1, w_2, w_3, w_4, w_5) \in \{A, 2, \dots, 10, J, Q, K \text{ para los cuatro palos } \clubsuit \diamondsuit \heartsuit \spadesuit\}\}$$

$$B = \{A \spadesuit, A \diamondsuit\} = \text{par de ases}$$

$$C = \{10 \clubsuit, J \clubsuit, Q \clubsuit, K \clubsuit, A \clubsuit\} = \text{flor imperial}$$

➤ Aunque la definición de la probabilidad es única (medida de la incertidumbre), existen varias INTERPRETACIONES o formas de calcular la probabilidad:

- PROBABILIDAD CLÁSICA: Supone que todos los posibles resultados de un experimento aleatorio son igualmente verosímiles (posibles) y propone como medida de probabilidad el cociente entre casos favorables y casos totales,


$$P(A) = \frac{N_A}{N}$$

Nota: No es necesario realizar el experimento para conocer las probabilidades de eventos.

- PROBABILIDAD FRECUENTISTA: Supone que un experimento aleatorio puede ser repetido un número infinito de veces bajo condiciones similares y propone como medida de probabilidad la proporción de veces que ocurrió el evento de interés,

$$P(A) = \lim_{N \rightarrow \infty} \frac{N_A}{N}$$

- **PROBABILIDAD SUBJETIVA:** Es simplemente una medida de la incertidumbre asociada a un evento, asignada subjetivamente. En otras palabras, es un juicio personal sobre la verosimilitud de que ocurra un resultado,

$$P(A) =$$


➤ **EJEMPLO 2. (Continuación...)**

$N = 2$  resultados posibles

$$\Rightarrow P(A) = 1/2 \text{ y } P(B) = 1/2$$

➤ **EJEMPLO 3. (Continuación...)**

$N = 6$  resultados posibles

$$\Rightarrow P(A) = 3/6 = 1/2$$

$$P(B) = 4/6 = 2/3$$

➤ **EJEMPLO 4. (Continuación...)**

$$N = \binom{52}{5} = 2\,598\,960$$

$$\Rightarrow P(\text{un par}) = 0.4225$$

$$P(\text{dos pares}) = 0.0950$$

$$P(\text{tercia}) = 0.0211$$

$$P(\text{corrida}) = 0.0039$$

$$P(\text{color}) = 0.0019$$

$$P(\text{full}) = 0.0014$$

$$P(\text{pocker}) = 0.00024$$

$$P(\text{flor}) = 0.000015$$

- EJEMPLO 5. Casos en donde no es posible usar la probabilidad clásica para calcular una probabilidad:

$P(\text{caiga sol al lanzar una moneda cargada}) = ?$

$P(\text{una persona viva más de 50 años}) = ?$

$P(\text{sacar un 10 en un examen de probabilidad}) = ?$

$P(\text{el valor de una acción suba mañana}) = ?$

- Sin importar la forma en la que se calcule la probabilidad (clásica, frecuentista o subjetiva) debe de cumplir con los siguientes AXIOMAS DE PROBABILIDAD:

a)  $P(A) \geq 0$ , para cualquier  $A \subset \Omega$ .

b)  $P(\Omega) = 1$ .

c) Si  $A_1, A_2, \dots, A_k$  son  $k$  eventos disjuntos (su intersección es vacía),

$$\text{entonces } P\left(\bigcup_{i=1}^k A_i\right) = \sum_{i=1}^k P(A_i).$$

- A partir de los tres axiomas anteriores, se pueden demostrar las siguientes PROPIEDADES DE LA PROBABILIDAD:

1.  $P(A^c) = 1 - P(A)$ , donde  $A^c$  es el evento complemento de  $A$ .

2.  $P(\emptyset) = 0$ . La probabilidad del evento imposible es cero.

3.  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .

4. Si  $B \subset A$ , entonces  $P(B) \leq P(A)$ .

- EJEMPLO 6. En un estanque conviven dos tipos de peces, el pez ángel y el pez japonés, ambos peces pueden ser de dos colores, naranja y gris. Estudios recientes del ecosistema del estanque muestran que la

probabilidad de que un biólogo capture un pez ángel o un pez naranja es de 0.40. La probabilidad de que capture un pez ángel es de 0.20 y la probabilidad de que capture un pez de color naranja es de 0.30. Si un pez es capturado, ¿cuál es la probabilidad de que el pez:

- a) sea un pez japonés?
- b) sea un pez ángel de color naranja?
- c) sea un pez ángel de color gris?
- d) sea un pez japonés de color naranja?
- e) sea un pez japonés de color gris?

Sea  $A$  = pez ángel y  $B$  = pez de color naranja,  $\Rightarrow P(A) = 0.20$ ,  $P(B) = 0.30$  y  $P(A \cup B) = 0.40$ .

- a)  $P(A^c) = 1 - 0.20 = 0.80$
- b)  $P(A \cap B) = 0.20 + 0.30 - 0.40 = 0.10$
- c)  $P(A \cap B^c) = P(A) - P(A \cap B) = 0.20 - 0.10 = 0.10$
- d)  $P(A^c \cap B) = P(B) - P(A \cap B) = 0.30 - 0.10 = 0.20$
- e)  $P(A^c \cap B^c) = 1 - P(A \cup B) = 1 - 0.40 = 0.60$

- **PROBABILIDAD CONDICIONAL:** Una vez que sabemos que el evento  $B$  ya ocurrió nos interesa conocer la probabilidad de que el evento  $A$  ocurra (la probabilidad de  $A$  puede verse afectada por la ocurrencia de  $B$ ),

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{si } P(B) \neq 0$$

□ **EJEMPLO 6. (Continuación...)**

Si ya sabemos que el biólogo capturó un pez de color naranja, ¿cuál es la probabilidad de que el pez sea un pez ángel?

$$P(A|B) = 0.10/0.30 = 0.33 \neq P(A) = 0.20$$

- REGLA DE LA MULTIPLICACIÓN: La probabilidad condicional puede usarse para calcular probabilidades de intersección de eventos,

$$P(A \cap B) = P(B)P(A|B) \quad \text{ó} \quad P(A \cap B) = P(A)P(B|A)$$

□ EJEMPLO 6. (Continuación...)

¿Cuál es la probabilidad de capturar un pez ángel de color naranja?

$$P(A \cap B) = (0.30)(0.33) = 0.10$$

- INDEPENDENCIA DE EVENTOS: Dos eventos son independientes si la probabilidad de ocurrencia de uno de ellos no se ve afectada por la ocurrencia ni por la no ocurrencia del otro. Los eventos A y B son independientes si

$$P(A|B) = P(A) \quad \text{ó} \quad P(B|A) = P(B) \quad \text{ó} \quad P(A \cap B) = P(A)P(B)$$

□ EJEMPLO 6. (Continuación...)

¿Será independiente el evento de que el pez capturado sea un pez ángel y el evento de que sea de color naranja?

Como  $P(A|B) = 0.10/0.30 = 0.33 \neq P(A) = 0.20$  entonces A y B no son independientes.

- *Nota:* El concepto de independencia es distinto al concepto de mutuamente excluyentes o disjuntos.

- **TEOREMA DE LA PROBABILIDAD TOTAL:** Suponga que  $\Omega$  está formado por la unión de  $k$  eventos mutuamente excluyentes  $B_1, B_2, \dots, B_k$ . Considere un evento  $A \subset \Omega$ , entonces

$$P(A) = \sum_{j=1}^k P(A|B_j)P(B_j)$$

Este resultado nos sirve para calcular de una manera fácil la probabilidad de un evento cuando el espacio muestral  $\Omega$  se puede obtener como la unión de eventos mutuamente excluyentes (disjuntos).

- **TEOREMA DE BAYES:** Suponga que  $\Omega$  está formado por la unión de  $k$  eventos mutuamente excluyentes  $B_1, B_2, \dots, B_k$ . Considere un evento  $A \subset \Omega$ , entonces,

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^k P(A|B_j)P(B_j)}, \quad i=1, \dots, k$$

Este teorema nos sirva para calcular probabilidades condicionales de una forma sencilla. Además nos sirve para actualizar información en el sentido de que nuestro conocimiento se vuelve más preciso,

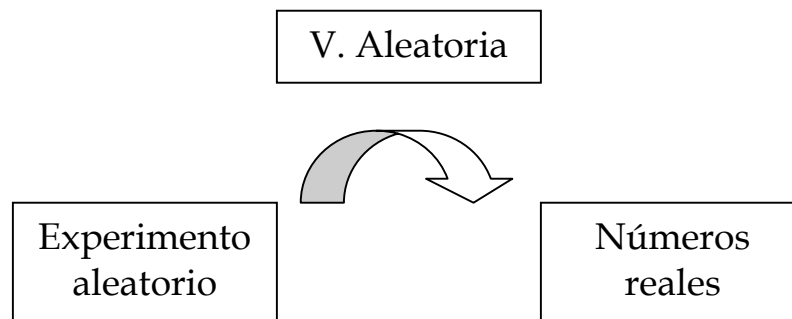
$$P(B_i) = \text{probabilidad inicial}$$

$$P(B_i|A) = \text{probabilidad final}$$



### 1.3 Variables aleatorias y distribuciones de probabilidad

- Algunos experimento aleatorios producen resultados u observaciones no necesariamente numéricas, como la opinión a favor o en contra de una propuesta o el nivel de satisfacción de un servicio.
- Las variables aleatorias son un medio para relacionar cualquier resultado de un experimento aleatorio con una medida cuantitativa (numérica).



- EJEMPLO 7: Un experimento aleatorio consiste en lanzar una moneda 3 veces y observar la cara obtenida. El resultado de cada lanzamiento puede ser águila (A) o sol (S). El espacio muestral es

$$\Omega = \{SSS, SSA, SAS, ASS, AAS, ASA, SAA, AAA\}$$

Si el interés radica en el total de águilas obtenidas en los 3 lanzamientos, los resultados pueden representarse por los números 0, 1, 2 y 3. A cada resultado del espacio muestral se le asocia un número, i.e.,

SSS→0,

SSA→1, SAS→1, ASS→1

AAS→2, ASA→2, SAA→2

AAA→3

Si definimos a la variable  $X$  = total de águilas observadas, entonces los números 0, 1, 2 y 3 son los valores que toma la variable aleatoria.

Como el resultado del experimento es incierto (aleatorio), el valor de la variable  $X$  también lo es, por esta razón la variable  $X$  es llamada *variable aleatoria*.

- Para un mismo espacio muestral es posible definir más de una variable aleatoria dependiendo del interés que se tenga.

- EJEMPLO 7. (Continuación...)

$Y$  = total de soles observados

$Z$  = número de soles en el segundo lanzamiento de la moneda

- CLASIFICACIÓN de variables aleatorias (v.a.): Las v.a. se clasifican, de acuerdo a los valores que pueden tomar, en continuas y discretas.
  - *v.a. Discretas*: si puede tomar un número finito o infinito numerable (que se puede contar) de valores.
  - *v.a. Continuas*: si puede tomar cualquier valor en un intervalo dado (sus valores son no numerables).

- EJEMPLO 8.

Algunos ejemplos de v.a. discretas son:

$X$  = número de alumnos, en un grupo de 30, que obtienen calificación superior a 8.5 en un examen. Los posibles valores de  $X$  serían 0,1,2,...,30.

$Y$  = número de peces sin escamas que habitan en un lago. Los posibles valores de  $Y$  son 0,1,2,....

Algunos ejemplos de v.a. continuas son:

$Z$  = tiempo requerido por los trabajadores de INE para trasladarse desde su casa al trabajo. Los posibles valores de  $Z$  son  $z \in [0, \infty)$ .

$W$  = altura en centímetros de un cierto tipo de plantas. Los valores que toma  $W$  son  $w \in (0, 100]$ .

- El grado de incertidumbre de los valores que toma una variable aleatoria no es el mismo, existen algunos valores que ocurren más frecuentemente que otros. Es por esto que es necesario cuantificar la incertidumbre mediante una *función de probabilidad*.

- FUNCIÓN DE PROBABILIDAD DE UNA V.A. DISCRETA. Es una función que asigna a cada posible valor  $x$  de  $X$  su probabilidad de ocurrencia,

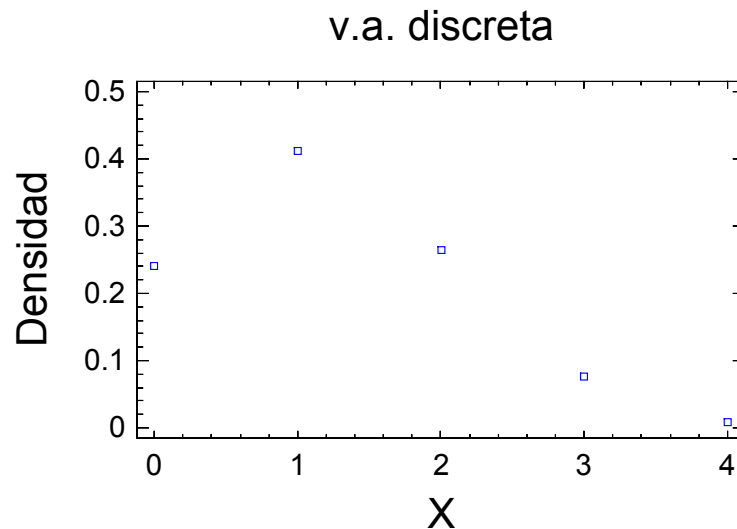
$$f_X(x) = P(X = x)$$

La función de probabilidad también es conocida como función de densidad o función de masa. Esta función tiene las siguientes propiedades:

1.  $f_X(x) \geq 0$  para todo valor  $x$ .
2.  $\sum_x f_X(x) = 1$  (la suma es sobre todos los valores que toma la v.a.)

- A continuación se presenta una ilustración de una función de probabilidad junto con su representación gráfica.

Variable	Probability Mass
0	Dist. 1 0.2401
1	0.4116
2	0.264601
3	0.0756
4	0.0081



La representación gráfica de una función de probabilidad discreta puede ser mediante un histograma, un diagrama de líneas verticales o mediante puntos.

➤ EJEMPLO 7. (Continuación...)

$$f_X(0) = P(X=0) = P(\{SSS\}) = 1/8 = 0.125$$

$$f_X(1) = P(X=1) = P(\{SSA\} \cup \{SAS\} \cup \{ASS\}) = 0.125 + 0.125 + 0.125 = 0.375$$

$$f_X(2) = P(X=2) = P(\{AAS\} \cup \{ASA\} \cup \{SAA\}) = 0.125 + 0.125 + 0.125 = 0.375$$

$$f_X(3) = P(X=3) = P(\{AAA\}) = 0.125$$

- FUNCIÓN DE DISTRIBUCIÓN DE UNA V.A. DISCRETA. Esta función se define como la probabilidad de que X tome valores menores o iguales a x,

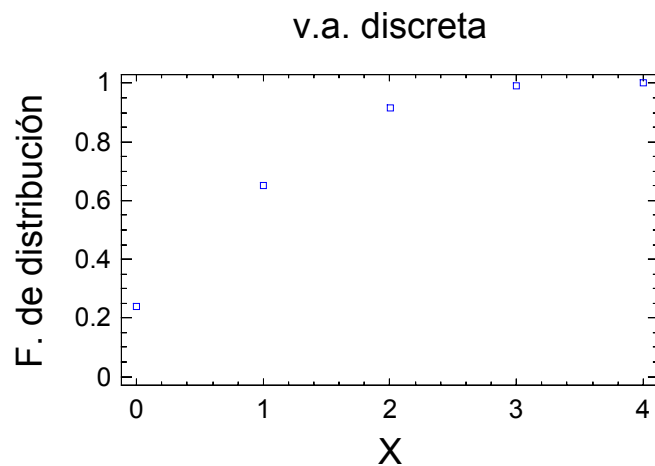
$$F_X(x) = P(X \leq x) = \sum_{u \leq x} f_X(u)$$

La función de distribución también es conocida como función de distribución acumulada (de probabilidad). Esta función tiene las siguientes propiedades:

1.  $0 \leq F_X(x) \leq 1$  para todo valor  $x$ .
2.  $F_X(x)$  es no decreciente
3.  $F_X(x)$  es continua por la derecha

➤ A continuación se presenta una ilustración de una función de distribución junto con su representación gráfica.

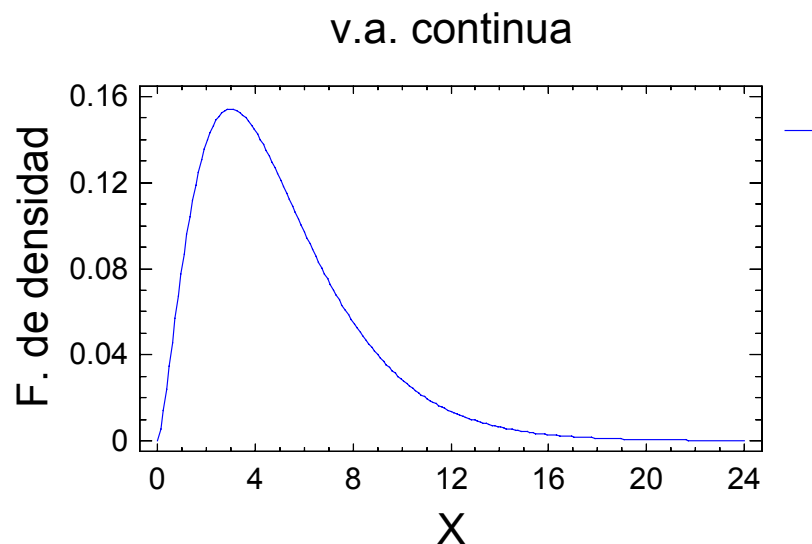
$x$	$F_X(x)$
0	0.2401
1	0.6517
2	0.9163
3	0.9919
4	1



➤ *Nota:* La función de distribución está definida para todo valor  $x$ . Además, es muy útil porque simplifica el cálculo de probabilidades.

➤ FUNCIÓN DE DENSIDAD DE UNA V.A. CONTINUA. La función de densidad para v.a.'s continuas no corresponde a la probabilidad de observar un valor  $x$ , simplemente nos da una idea de la distribución (repartición) de las probabilidades para los distintos valores  $x$ . Gráficamente la función de densidad corresponde a una forma límite del histograma de frecuencias relativas. Esta función tiene las siguientes propiedades:

1.  $f_X(x) \geq 0$  para todo valor  $x$ .
2.  $\int f_X(x)dx = 1$  (la integral es sobre todos los valores de la v.a.)



- A continuación se presenta una ilustración de una función de densidad.
- Nota: La forma de calcular probabilidades que involucren valores de una v.a. continua es integrando,

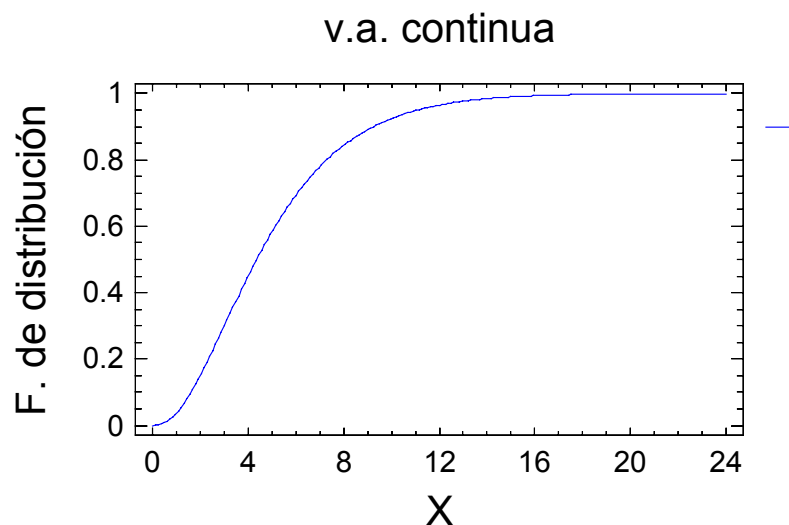
$$P(a < X < b) = \int_a^b f_X(x)dx$$

- FUNCIÓN DE DISTRIBUCIÓN DE UNA V.A. CONTINUA. Al igual que en el caso discreto, esta función se define como la probabilidad de que  $X$  tome valores menores o iguales a  $x$ ,

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(u) du$$

Esta función tiene las mismas propiedades que una función de distribución de una v.a. discreta.

- A continuación se presenta una ilustración de una función de distribución.



- CARACTERÍSTICAS DE UNA V.A. Existen cantidades numéricas (parámetros) que caracterizan una v.a. Las más comunes son:

- *Valor esperado o media*: es una medida de la localización central de una distribución,

$$\begin{aligned} \mu = E(X) &= \sum_x x f_X(x), \text{ si } X \text{ es discreta} \\ &= \int x f_X(x) dx, \text{ si } X \text{ es continua} \end{aligned}$$

- *Varianza*: es una medida de la dispersión de los valores de la v.a.  $X$ ,

$$\sigma^2 = \text{Var}(X) = E\{(X - \mu)^2\} = \sum_x (x - \mu)^2 f_X(x), \text{ si } X \text{ es discreta}$$

$$= \int (x - \mu)^2 f_X(x) dx, \text{ si } X \text{ es continua}$$

- NOTA: Todas las demás medidas descriptivas que se vieron en la sección de análisis exploratorio de datos se pueden calcular para cualquier v.a, por ejemplo: *Mediana, Moda, Cuartiles, Rango, Amplitud intercuartílica, desviación estándar y coeficiente de variación.*

- FUNCIONES BIVARIADAS: Algunas veces resulta de interés conocer el comportamiento conjunto de dos variables aleatorias definidas en el mismo espacio muestral  $\Omega$ . Sean  $X$  y  $Y$  son dos v.a.'s, su *función de densidad*  $f_{X,Y}(x,y)$  indica probabilidades si  $X$  y  $Y$  son discretas e indican la distribución de los valores de acuerdo a su probabilidad si  $X$  y  $Y$  son continuas. La *función de distribución*  $F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$ , en cualquier caso, indica probabilidades acumuladas.

- *Coeficiente de correlación*. Un parámetro importante que mide el grado de relación lineal entre dos v.a.'s es el coeficiente de correlación, que se define como:

$$\rho_{XY} = \text{Corr}(X, Y) = \frac{E\{(X - \mu_X)(Y - \mu_Y)\}}{\sqrt{\text{Var}(X)\text{Var}(Y)}} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y},$$

Propiedades:  $-1 \leq \rho_{XY} \leq 1$ .



## 1.4 Algunas distribuciones específicas de probabilidad

- Existen fenómenos aleatorios que se pueden representar por modelos (de probabilidad) conocidos y relativamente sencillos.
- Los MODELOS DE PROBABILIDAD son modelos que abstraen la realidad (experimentos aleatorios) mediante una ecuación matemática. Un modelo de probabilidad especifica la forma de la distribución de las probabilidades asociadas a una v.a.. Los modelos de probabilidad se describen mediante su *función de densidad*, su *función de distribución* o mediante sus *características resumen* (media, varianza, etc.).
- Algunos modelos más comunes de probabilidad son:

*Discretos:*

- 1) Uniforme discreta
- 2) Bernoulli
- 3) Binomial
- 4) Poisson

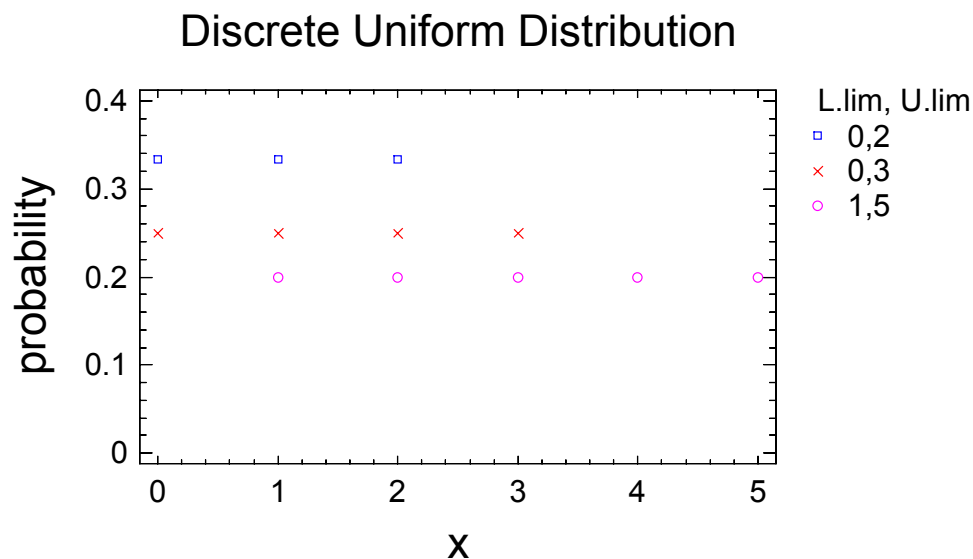
*Continuos:*

- 5) Uniforme continua
- 6) Normal
- 7) Exponencial

- 1) UNIFORME DISCRETA: Una v.a.  $X$  se dice que tiene una distribución uniforme discreta o rectangular si su función de densidad está dada por:

$$f_X(x) = \frac{1}{k} I_{\{x_1, x_2, \dots, x_k\}}(x)$$

- *Notación:*  $X \sim U\{x_1, x_2, \dots, x_k\}$  (la v.a.  $X$  se distribuye de manera uniforme discreta en el conjunto  $\{x_1, x_2, \dots, x_k\}$ ).



- *Aplicaciones:*

$X$  = asigna el valor de 1 si al lanzar una moneda honesta sale “sol”.

$Y$  = El número del boleto ganador en una rifa de 100 números.

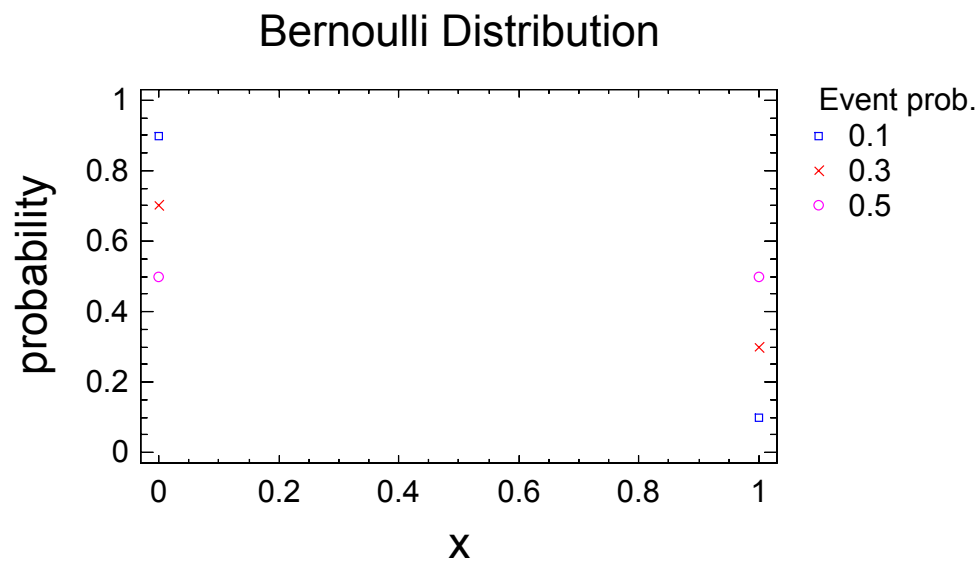
- *Propiedades:* Si  $X \sim U\{1, 2, \dots, N\}$ ,

$$\mu_X = \frac{N(N+1)}{2} \quad \text{y} \quad \sigma_X^2 = \frac{N^2 - 1}{12}$$

- 2) **BERNOULLI**: Una v.a.  $X$  se dice que tiene una distribución Bernoulli si su función de densidad está dada por:

$$f_X(x) = p^x (1-p)^{1-x} I_{\{0,1\}}(x), \text{ donde } 0 \leq p \leq 1$$

- *Notación*:  $X \sim \text{Ber}(p)$  (la v.a.  $X$  se distribuye como una Bernoulli con parámetro  $p$ ).



- *Aplicaciones*: Un ensayo Bernoulli es aquel experimento en el que sólo hay 2 posibles resultados.

$X$  = asigna el valor de 1 si al lanzar una moneda honesta sale “sol”.

$Y$  = asigna el valor de 1 si me saco el melate y 0 si no.

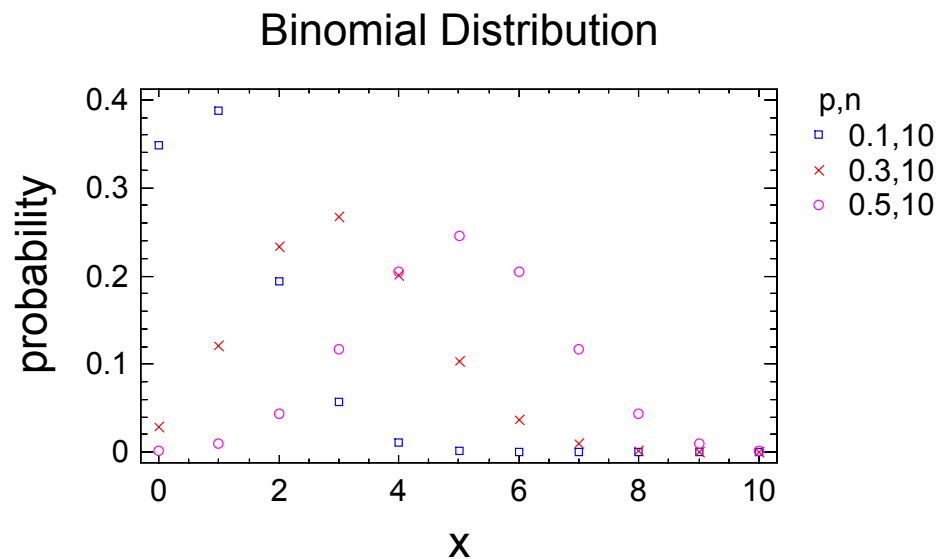
- *Propiedades*: Si  $X \sim \text{Ber}(p)$ ,

$$\mu_X = p \quad \text{y} \quad \sigma_X^2 = p(1-p)$$

- 3) BINOMIAL: Una v.a.  $X$  se dice que tiene una distribución Binomial si su función de densidad está dada por:

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x} I_{\{0,1,\dots,n\}}(x), \text{ donde } 0 \leq p \leq 1$$

- *Notación:*  $X \sim \text{Bi}(n,p)$  (la v.a.  $X$  se distribuye como una Binomial con parámetros  $n$  y  $p$ ).



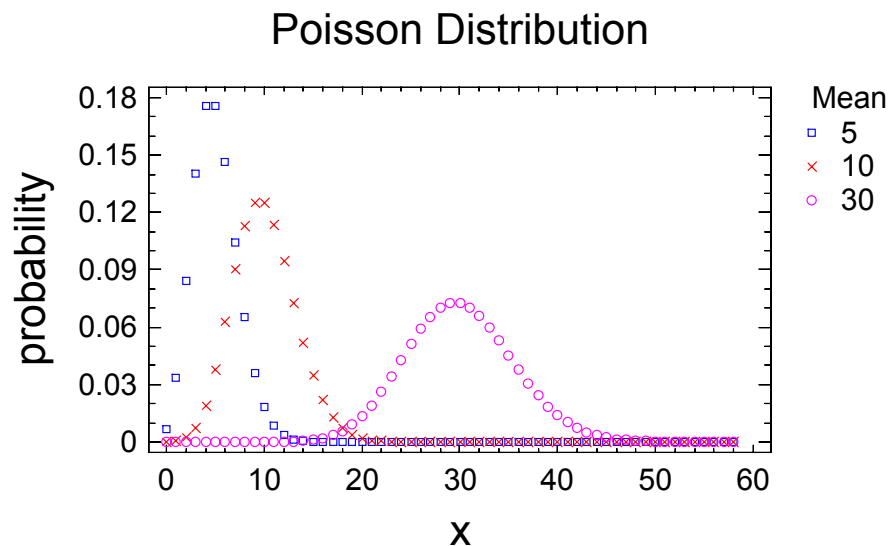
- *Aplicaciones:*  
 $X$  = número de éxitos en  $n$  ensayos Bernoulli.
- *Propiedades:* Si  $X \sim \text{Bi}(n,p)$ ,

$$\mu_X = np \quad \text{y} \quad \sigma_X^2 = np(1-p)$$

- 4) POISSON: Una v.a.  $X$  se dice que tiene una distribución Poisson si su función de densidad está dada por:

$$f_X(x) = e^{-\lambda} \frac{\lambda^x}{x!} I_{\{0,1,\dots\}}(x), \text{ donde } \lambda > 0$$

- *Notación:*  $X \sim \text{Po}(\lambda)$  (la v.a.  $X$  se distribuye como una Poisson con parámetro  $\lambda$ ).



- *Aplicaciones:* En general cuenta número de ocurrencias en un intervalo de tiempo.

$X$  = número de reclamos que reciben al día en una compañía de seguros.

$Y$  = número de aviones que llegan al aeropuerto de la ciudad de México en una semana.

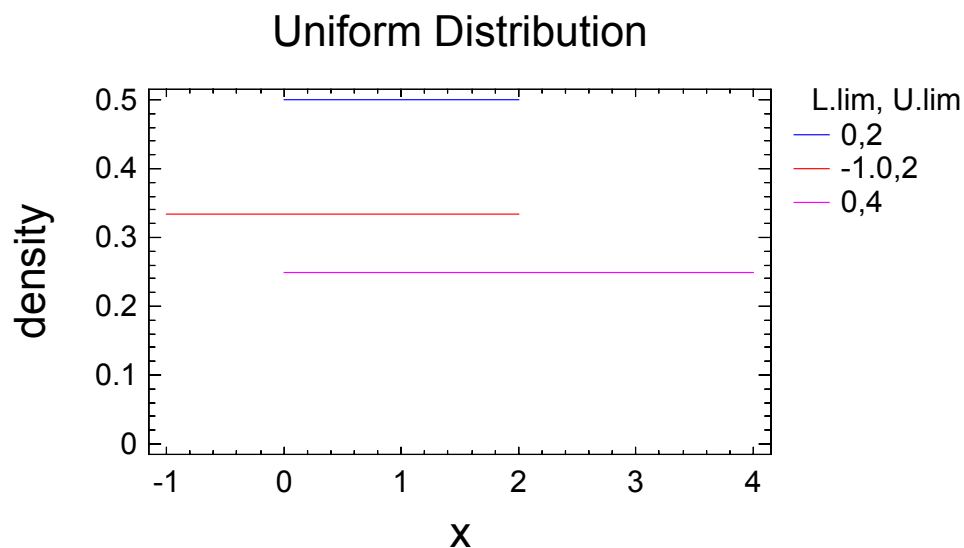
- *Propiedades:* Si  $X \sim \text{Po}(\lambda)$ ,

$$\mu_X = \lambda \quad \text{y} \quad \sigma_X^2 = \lambda$$

- 5) UNIFORME CONTINUA: Una v.a.  $X$  se dice que tiene una distribución Uniforme continua si su función de densidad está dada por:

$$f_X(x) = \frac{1}{b-a} I_{[a,b]}(x), \text{ donde } a < b$$

- *Notación:*  $X \sim U[a,b]$  (la v.a.  $X$  se distribuye como una Uniforme continua en el intervalo  $[a,b]$ ).



- *Aplicaciones:*

$X$  = tiempo en ser atendido por un servidor en la oficina de SAT.

$Y$  = tiempo en llegar desde mi casa al INE.

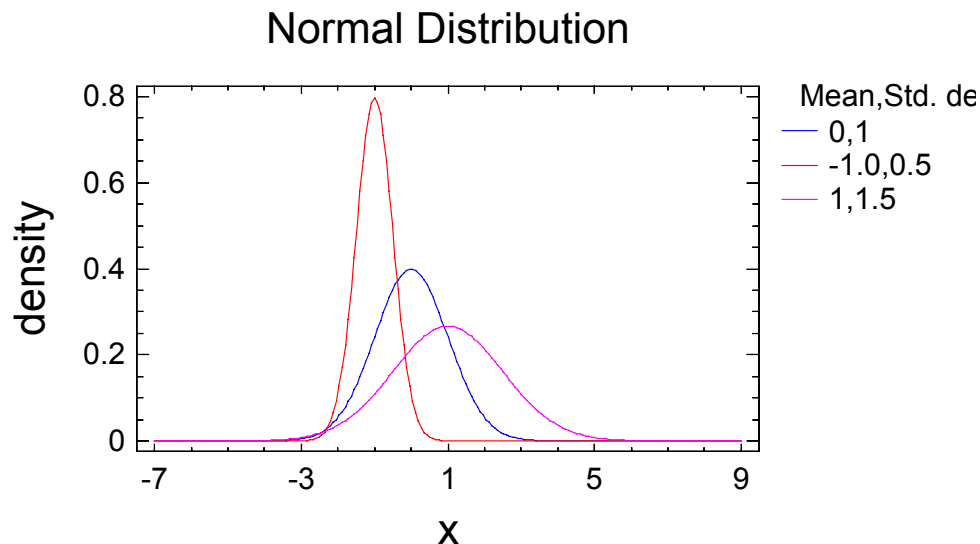
- *Propiedades:* Si  $X \sim U[a,b]$ ,

$$\mu_X = \frac{a+b}{2} \quad \text{y} \quad \sigma_X^2 = \frac{(b-a)^2}{12}$$

- 6) NORMAL: Una v.a.  $X$  se dice que tiene una distribución Normal si su función de densidad está dada por:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} I_{(-\infty, \infty)}(x), \text{ donde } \mu \in \mathbb{R}, \sigma^2 > 0$$

- *Notación:*  $X \sim N(\mu, \sigma^2)$  (la v.a.  $X$  se distribuye como una Normal con parámetros  $\mu$  y  $\sigma^2$ ).



- *Aplicaciones:*

$X$  = cantidad de calor en una barra metálica al ser calentada por el centro.

$Y$  = demanda mensual por un determinado producto.

- *Propiedades:* Si  $X \sim N(\mu, \sigma^2)$ ,

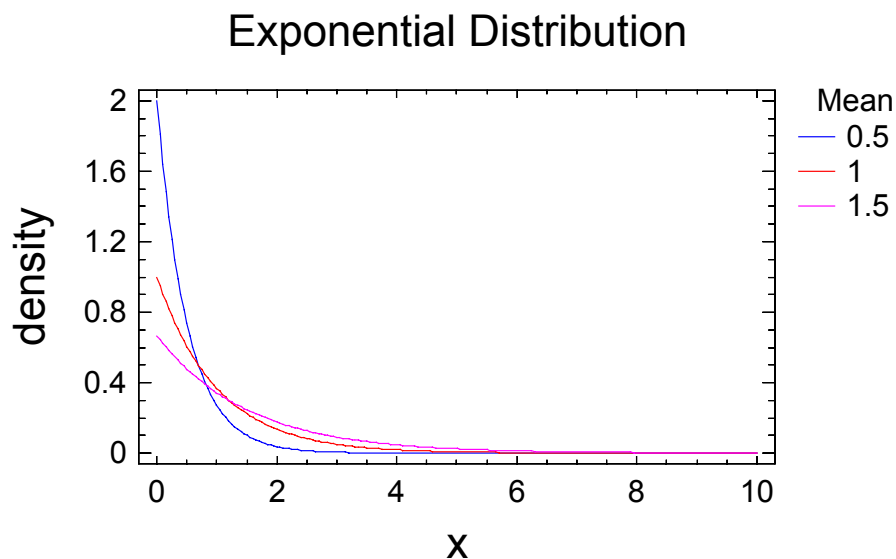
$$\mu_X = \mu \quad \text{y} \quad \sigma_X^2 = \sigma^2$$

Si  $X \sim N(0,1) \Rightarrow$  se dice que  $X$  es una v.a. normal estándar.

- 7) EXPONENCIAL: Una v.a.  $X$  se dice que tiene una distribución exponencial si su función de densidad está dada por:

$$f_X(x) = \frac{1}{\theta} e^{-x/\theta} I_{[0,\infty)}(x), \text{ donde } \theta > 0$$

- *Notación:*  $X \sim \text{Exp}(\theta)$  (la v.a.  $X$  se distribuye como una Exponencial con parámetro  $\theta$ ).



- *Aplicaciones:* En general mide tiempo entre la ocurrencia de dos eventos.  
 $X$  = tiempo entre la llegada de dos personas a la unifila de un banco.  
 $Y$  = tiempo entre la llegada de dos aviones al aeropuerto.

- *Propiedades:* Si  $X \sim \text{Exp}(\theta)$ ,

$$\mu_X = \theta \quad \text{y} \quad \sigma_X^2 = \theta^2$$