Unidad 5: Procesos Estocásticos
Estadística Descriptiva

La **estadística** se fundamenta en **métodos científicos** para la **recolección**, **organización**, **resumen**, **presentación** y **análisis de datos**. Dicho análisis permitirá obtener **conclusiones** y efectuar **decisiones** al respecto.

- El grupo a ser analizado se denomina **población**, que puede ser finito o infinito. Debido a que generalmente resulta impráctico examinarlo completamente, se toma una **muestra representativa**
- Las muestras representativas permiten **inferir conclusiones** acerca de la población (inferencia estadística). Si se puede trabajar con la población en su conjunto, la estadística resulta descriptiva
- Una variable es un **símbolo** que puede asumir un conjunto de valores preestablecidos (dominio). Puede ser **continua** (ej. Altura) o **discreta** (ej. Número de elementos)

Caracterización de Poblaciones

¿Qué es una distribución de Frecuencia?

La ***primera evaluación*** que puede llevarse a cabo sobre un conjunto de datos "crudo" es ***agruparlos en clases o categorías***, cuantificando la ***cantidad*** de elementos que ***pertenecen*** a cada una de ellas. Dicha cantidad se denomina ***"frecuencia".*** Por ejemplo, sea la ***altura*** de los estudiantes del curso:

Rango de Clase

Intevalo=10
Punto medio=5

ALTURA [cm]	NÚMERO DE ESTUDIANTES
145 - 155	2
155 – 165	5
165 – 175	13
175 – 185	34
185 – 195	14
196 – 205	1
TOTAL	**69**

Tabla de Distribución de frecuencias
Proporciona una vision ***GLOBAL*** de los datos

¿Cómo se conforma una distribución de frecuencia?

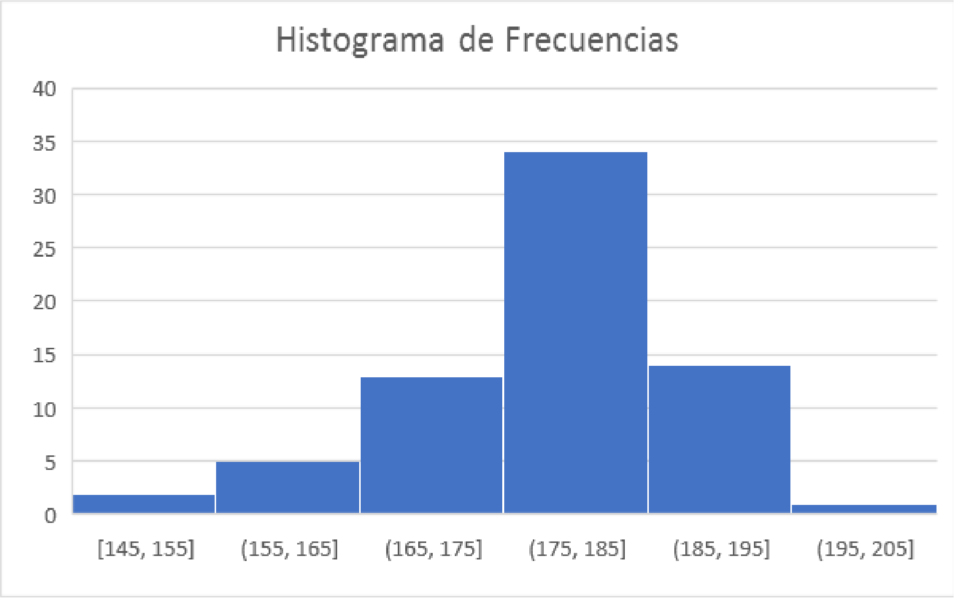
1. Determinar el ***mayor*** y el ***menor*** valor del conjunto de datos
2. Dividir el ***rango resultante*** en intervalos de clase del mismo tamaño (de ser posible). Por lo general se utilizan de 5 a 20 intervalos
3. Determinar el numero de observaciones que caen en cada clase

*La distribución de frecuencias se puede representar **gráficamente** a través de **Histogramas**. Un histograma es un gráfico de barras rectangulares, donde **cada barra representa una clase**.*

- La base de cada barra está centrada en el punto medio de la clase y su ancho es el del intervalo correspondiente
- La altura de cada barra corresponde a la cantidad de elementos (frecuencia) de dicha clase

¿Cómo se conforma una distribución de frecuencia?

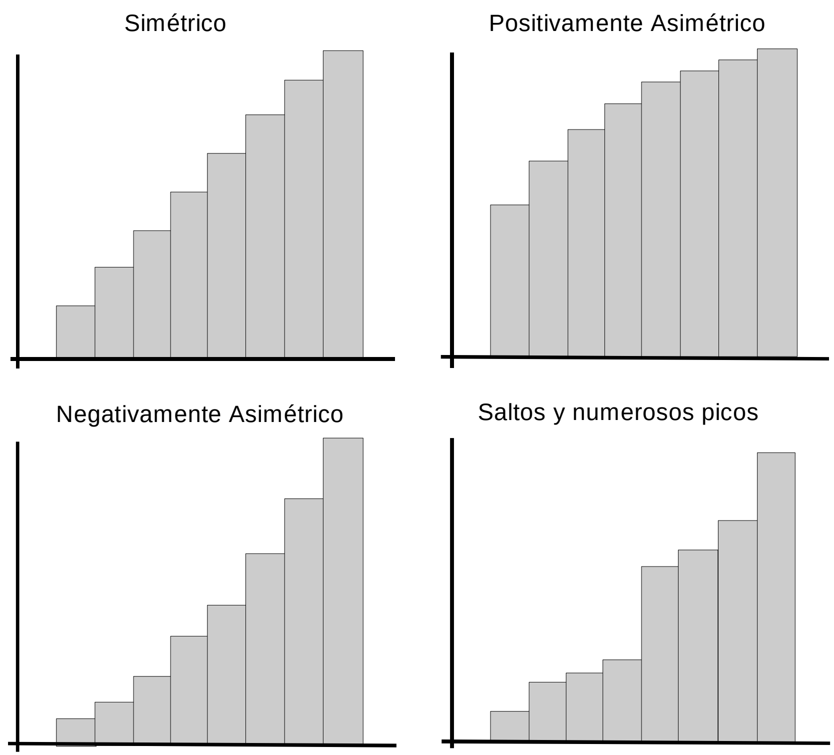
ALTURA [cm]	NÚMERO DE ESTUDIANTES
145 - 155	2
155 – 165	5
165 – 175	13
175 – 185	34
185 – 195	14
196 – 205	1
TOTAL	69



Si ***se divide cada frecuencia por el número total de elementos*** se obtiene un ***Histograma de frecuencias relativas***. En dicho caso la suma de todas las frecuencias equivale a 1

¿A qué se denomina distribución de frecuencia acumulada?

La distribución de frecuencia acumulada indica el ***número de elementos*** cuyos valores son **iguales** o **menores** a un determinado valor.



Tendencia Central y Dispersión

¿Qué valoraciones numéricas se pueden aplicar a las poblaciones?

Las ***medidas de TENDENCIA CENTRAL*** posibilitan resumir ***en un solo valor*** el comportamiento de los valores de la muestra a analizar. Representan ***el centro*** en torno al cual ***se ubica*** el conjunto de datos:

- **MEDIA ARITMÉTICA:** Constituye el ***promedio*** de todos los datos evaluados ($X_1, X_2, \ldots X_N$).

  Asimismo se la puede calcular a partir de las ***frecuencias de ocurrencia*** de dichos datos ($f_1, f_2, \ldots f_K$), considerando X_i como ***los valores a los que ocurren*** dichas frecuencias

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_N}{N} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$\bar{X} = \frac{f_1 X_1 + f_2 X_2 + \ldots + f_K X_K}{f_1 + f_1 + \ldots + f_K} = \frac{\sum_{i=1}^{K} f_i X_i}{\sum_{i=1}^{K} f_K}$$

Tendencia Central y Dispersión

- **MEDIANA:** Constituye el valor que ***separa*** un conjunto de datos ***ordenados*** por magnitud ***en dos conjuntos de igual cantidad de elementos***. Si el total de datos es ***par*** se considerará el ***promedio*** de los dos centrales

$$3,4,4,5,6,8,8,8,10 \rightarrow Me = 6$$

$$5,5,7,9,11,12,15,18 \rightarrow Me = \frac{(9+11)}{2} = 10$$

- **MODA:** Representa ***el valor mayor ocurrencia (la frecuencia más alta)*** de un conjunto de datos. Puede no existir o existir y no ser única.

$$2,2,5,7,9,9,9,10,10,11 \rightarrow Mo = 9$$

$$5,5,7,9,11,11,15,18 \rightarrow \begin{cases} Mo = 5 \\ Mo = 11 \end{cases} (\text{bimodal})$$

$$5,6,7,9,11,13,15,18 \rightarrow \textit{No existe}$$

Tendencia Central y Dispersión

Por otra parte, ***las medidas de DISPERSIÓN*** posibilitan determinar cómo estan ***distribuidos*** los valores ***en torno a la media***:

- **DESVÍO ESTÁNDAR:** Representa las desviaciones de los valores respecto de la media. Puede calcularse a partir de un conjunto de ***N*** datos o ***K*** frecuencias de ocurrencia (recordar que en este ultimo caso los X_i son los valores a los que ocurren dichas frecuencias)

$$\boxed{\sigma = \sqrt{\frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2}{N}} = \sqrt{\frac{\sum_{i=1}^{K} f_K\left(X_i - \overline{X}\right)^2}{N}}}$$

Unidad 5: Procesos Estocásticos

Tendencia Central y Dispersión

- **VARIANZA:** Cuadrado de la desviación estándar

$$\sigma^2 = \frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2}{N} = \frac{\sum_{i=1}^{K} f_K\left(X_i - \overline{X}\right)^2}{N}$$

- **COEFICIENTE DE VARIACIÓN:** Es indicativo de la ***dispersion relativa de los datos*** ya que relaciona el desvío estándar respecto del valor de la media

$$CV = \frac{\sigma}{\overline{X}}$$

Tendencia Central y Dispersión

Ejemplo: Caracterizar estadísticamente el peso de la siguiente población:

INDIVIDUO	PESO [kg]
1	56
2	57
3	63
4	68
5	71
6	77
7	80
8	87

$$Media : \overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i = 69{,}875$$

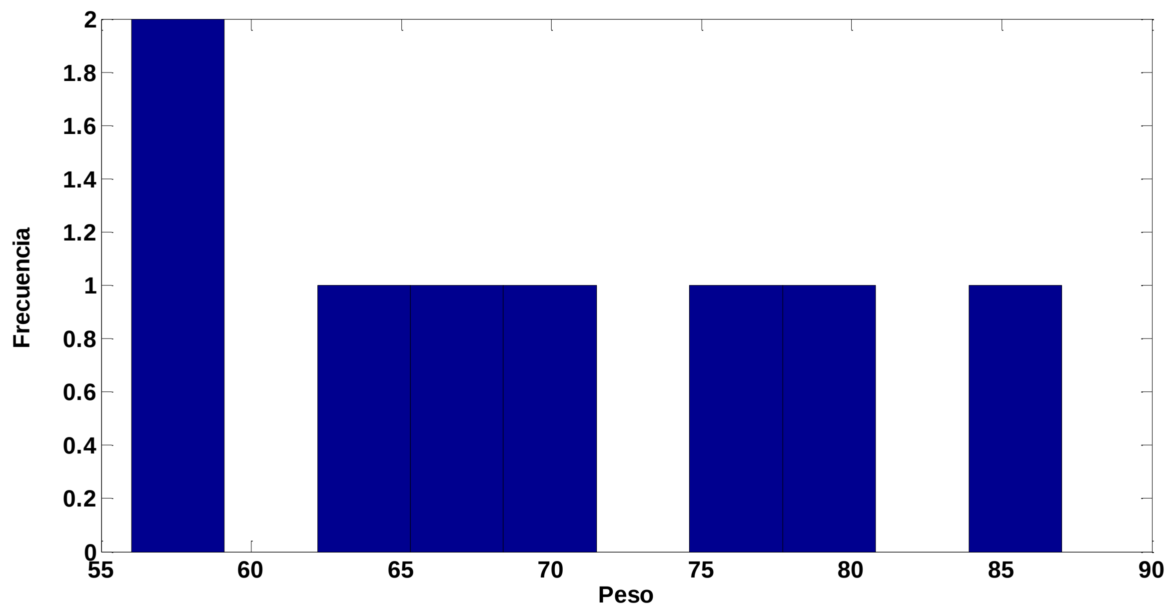
$$Mediana : Me = (68 + 71)/2 = 69{,}5$$

$$Moda : Mo = NE$$

$$Desvío\ \ Estándar : SD = \sqrt{\frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2}{N}} = 11{,}0639$$

En Matlab…

```
%Ejercicio Estadística Básica
Peso=[56,57,63,68,71,77,80,87]
%Histograma
hist(Peso)
xlabel('Peso'),ylabel('Frecuencia');
%Valor medio
mean(Peso)
%Mediana
median(Peso)
%Moda
mode(Peso)
%Desviación estándar
std(Peso)
```



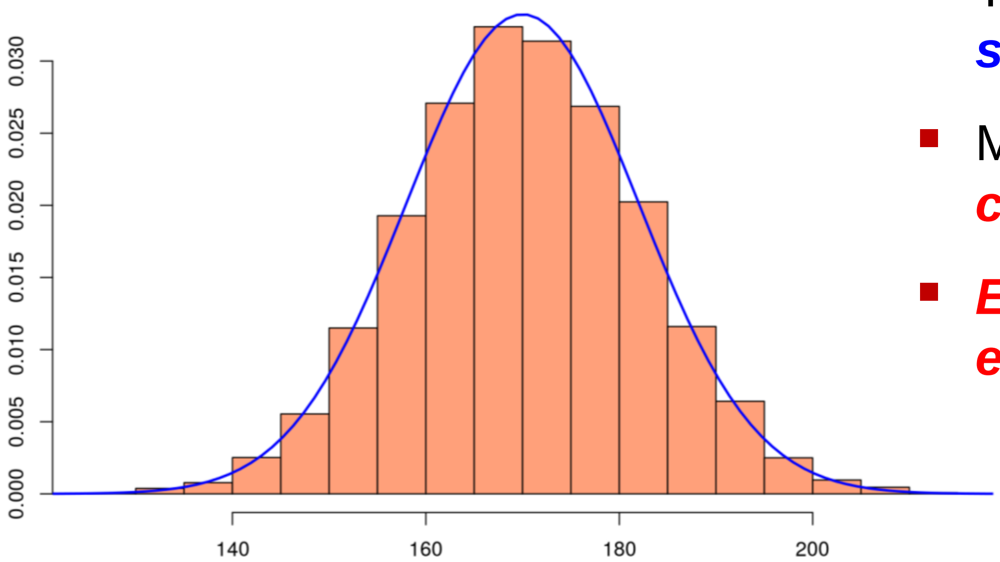
Caracterización de Poblaciones

¿Hay distribuciones de frecuencia continuas?

Si existe la posibilidad de considerar ***pequeños intervalos de clase*** (en muestras o poblaciones de grandes cantidades de datos), las ***distribuciones de frecuencia*** pueden aproximarse a ***"curvas de frecuencia"***

Tipos de distribución: La Distribución Normal

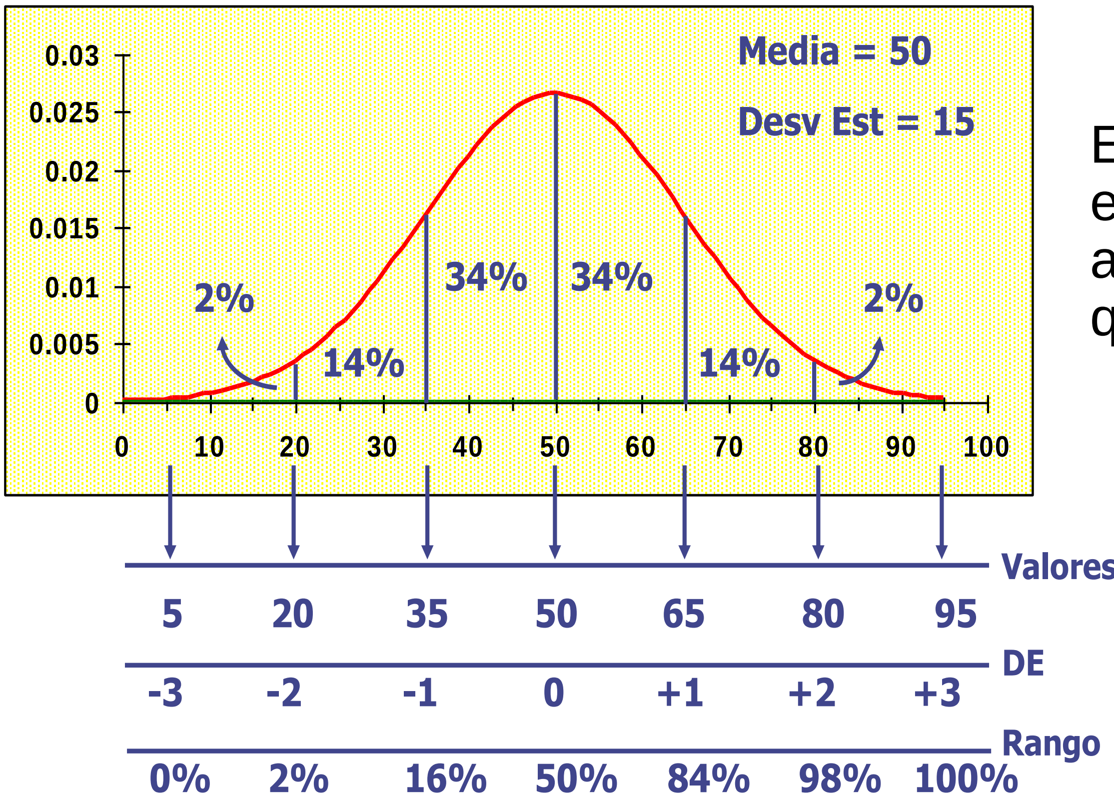
Existen ***numerosas variables de la vida diaria*** (la Altura, el peso, las notas de los exámenes) que tienden a comportarse como una ***distribución NORMAL***:



- Tiene forma ***acampanada*** y es ***simétrica*** respecto a la media
- Media, moda y mediana ***coinciden***
- ***Es una función de la media y el desvío estándar***

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\overline{X})^2}{2\sigma^2}}}$$

Tipos de distribución: La Distribución Normal



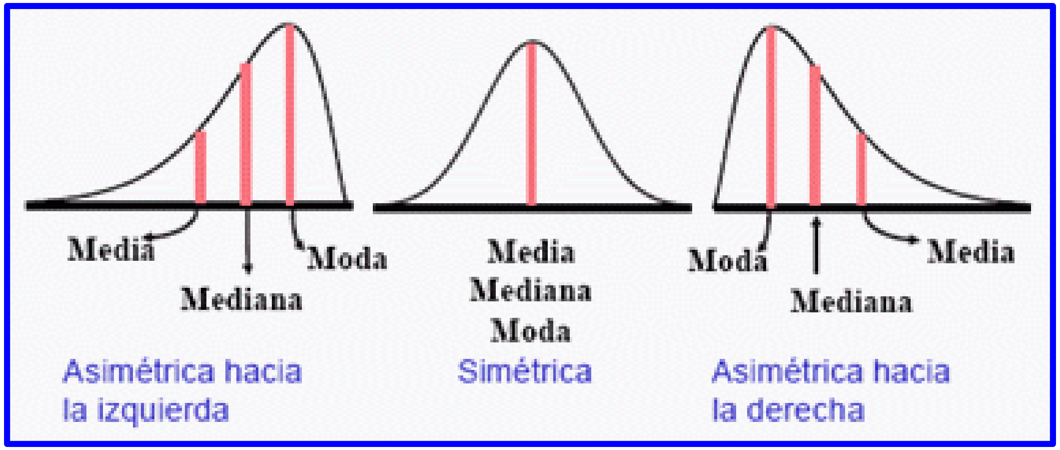
En la ***Distribución Normal***, el ***68%*** de la población se agrupa en ***±1DE***, mientras que el ***96%*** en ***±2DE***

Medidas de Asimetría

La asimetría constituye una medida ***de forma*** de la distribución :

Puede ser:

- **Negativa** (a la izquierda), donde ***media<mediana<moda (AS<0)***
- **Simétrica** (centrada), donde ***media=mediana=moda (AS=0)***
- ***Positiva*** (a la derecha), donde ***media>mediana>moda*** (***AS>0***)



$$AS = \frac{3(\overline{X} - Mediana)}{SD}$$

Coeficiente de asimetría de Pearson

Consigna de la clase (20 minutos)

1. Utilizando la tabla de valores proporcionada en el campus virtual, ***utilizar Matlab*** para determinar:

- Histograma
- Media
- Mediana
- Moda
- Desvío Estándar
- Coeficiente de Variación
- Asimetría

Nociones de Probabilidad

La probabilidad puede concebirse como algo similar a una ***"proporción"*** de ocurrencia de un evento. Se define como el ***cociente*** entre los casos favorables ***(h)*** y los casos posibles ***(n)*** en relación a un evento específico (**Ley de Laplace**):

$$\boxed{p = \frac{h}{n}}$$ es la ***"probabilidad de ocurrencia"*** (*éxito*)

$$\boxed{q = 1 - \frac{h}{n} = 1 - p}$$ es la ***"probabilidad de no ocurrencia"*** (*falla*)

La probabilidad de ocurrencia de un evento ***es un número entre 0 y 1***, donde si el evento no puede ocurrir ***p=0*** y si debe ocurrir (suceso seguro) ***p=1***.

Estadísticamente se define a la probabilidad como la frecuencia relativa de ocurrencia del evento cuando el número de observaciones es elevado. Ej: de una tirada de n=1000 veces una moneda y 529 caras, p=529/100=0,529 (p → 0,5 si n →∞)

Nociones de Probabilidad

Ejemplo: ¿Cuál es la probabilidad de obtener un número ***par*** al arrojar un dado? ***(Suceso A=Obtención Nro par)***

Casos favorables ***(A)***: ***2,4 y 6 (3)***
Casos Desfavorables ***(No A)***: ***1,3,5 (3)***
Casos posibles ***(A y No A)***: ***1,2,3,4,5 y 6 (6)***

$$\boxed{P(A)=\frac{3}{6}=0{,}5}$$

$$\boxed{P(A)+P(\overline{A})=1}$$

Sean dos sucesos **A** y **B**:

$$\boxed{P(A\cap B)=P(A)P(B)}$$

Probabilidad de que A y B ocurran de manera independiente *(P que salgan dos "6" seguidos)*

$$\boxed{P(AoB)=P(A)+P(B)}$$

Probabilidad de que ocurra A o B sin que ocurran a la vez (excluyentes) *(P que salga el "2" o el "4")*

$$\boxed{P(A|_B)=P(A\cap B)/P(B)}$$

Probabilidad de que A ocurra condicionada por la ocurrencia previa de B

$$\boxed{P(AUB)=P(A)+P(B)-P(A\cap B)}$$

Probabilidad de que ocurra A o B, pudiendo ocurrir ambos a la vez (no excluyentes)

Distribuciones de probabilidad discretas:

Si una variable ***X*** puede asumir un conjunto discreto de valores X_1, X_2,…,X_n con distinta probabilidad de ocurrencia p_1,p_2,…,p_n, puede definirse entonces una ***distribución de probabilidad discreta***:

Ejemplo: Se tiran un par de dados y se calcula la probabilidad de obtener la suma de sus caras:

X	2	3	4	5	6	7	8	9	10	11	12
p(X)	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

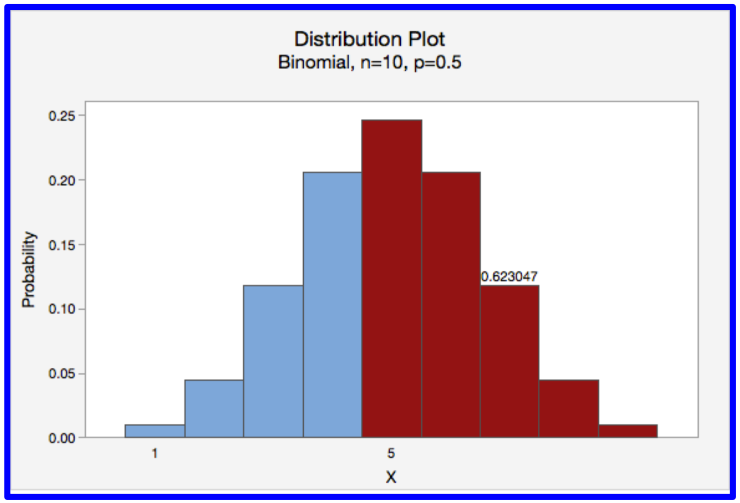
*La probabilidad de obtener suma **5** es **4/36=1/9***

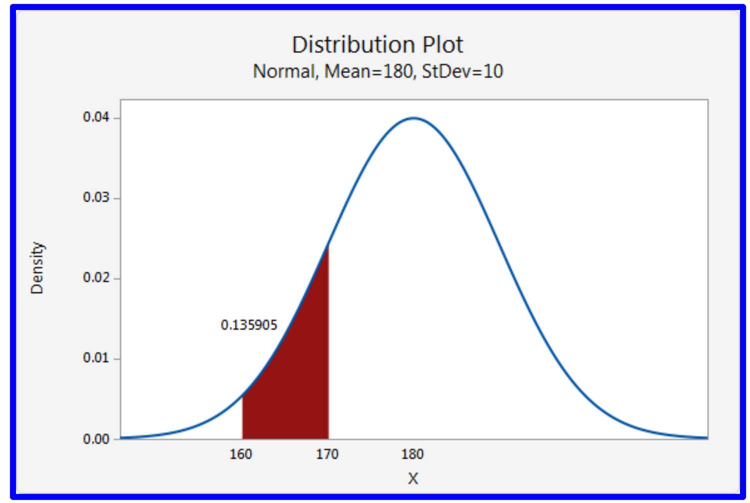
Las distribuciones de probabilidad pueden pensarse como distribuciones de frecuencia relativa "ideales", cuando el número de observaciones es muy elevado.

Dado que X puede asumir ciertos valores con probabilidad dada se la denomina "Variable ALEATORIA DISCRETA"

Distribuciones de probabilidad continuas:

Las mismas ideas pueden extenderse al caso continuo, considerando que ***X*** ***puede asumir un conjunto continuo de valores (envolvente)***.





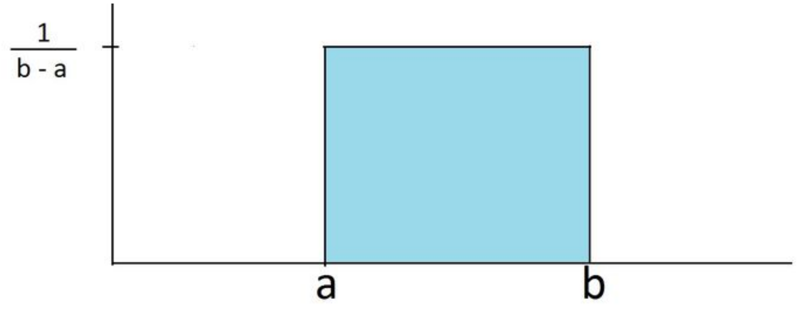
Al ser una función de distribución de probabilidad, ***el área bajo la curva resulta igual a 1***. Es por ello que el área bajo un ***intervalo [a,b]*** proporciona la ***probabilidad*** de que ***X se encuentre*** entre a y b. En este caso, ***la variable ALEATORIA*** pasa a ser ***CONTINUA***

¿En qué se convierten las distribuciones al ser continuas?

Las distribuciones continuas se describen a partir de la ***función densidad de probabilidad $f_X(x)$*** que determina la ***probabilidad*** que la variable aleatoria ***X*** asuma un valor específico en el intervalo ***[x, x+dx]*** *(no se dispone de un conjunto numerable de valores por lo que no se puede definir una probabilidad para cada uno de ellos, **se trabaja por intervalos**)*

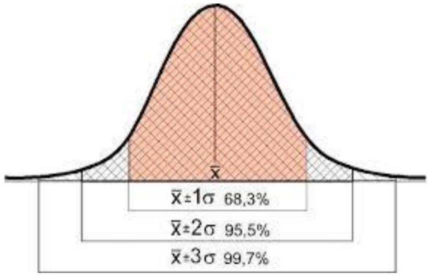
UNIFORME: Corresponde a una variable aleatoria continua X, en un intervalo [a,b], donde la ocurrencia es ***equiprobable***.

$$\boxed{f_X(x) = \begin{cases} \dfrac{1}{b-a} & a < x < b \\ 0 & \forall x \end{cases}}$$



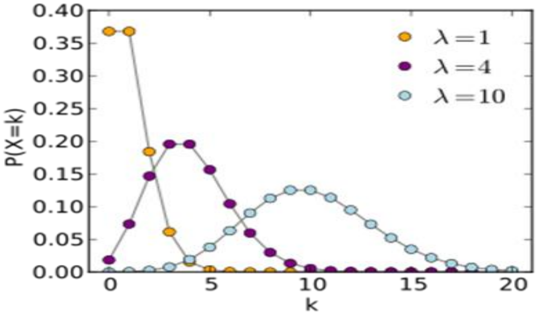
GAUSSIANA o NORMAL: Definida por los parámetros μ y σ (media y desvío estándar), permite modelar *estatura*, *consumo*, *ruido*…

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



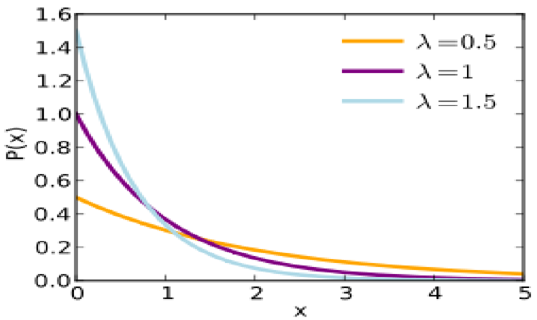
POISSON: Definida por un parámetro λ (valor medio de eventos por unidad de tiempo), permite modelar el número total de ocurrencias de un fenómeno durante un lapso específico:

$$f_X(x) = e^{-\lambda} \frac{\lambda^x}{x!} \; (con \;\; x \; entero \;\; positivo)$$



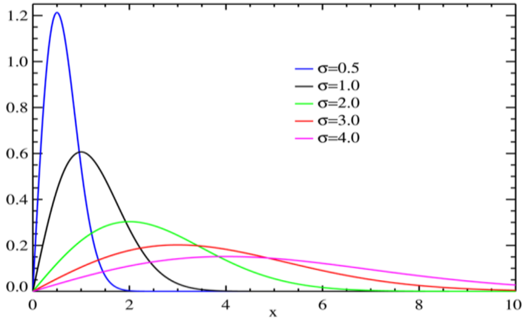
EXPONENCIAL: Definida a partir de un parámetro $\lambda>0$ (tasa de ocurrencia por unidad de tiempo), permite modelar *fallas en componentes*, *tiempos de espera* e *intervalos entre eventos*.

$$\boxed{f_X(x) = \lambda e^{-\lambda x}}$$



RAYLEIGH: Definida a partir de un parámetro σ permite modelar *la velocidad del viento, el esfuerzo al que se someten los materiales y tiempo-falla en componentes*

$$\boxed{f_X(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}}$$



¿Se pueden caracterizar numéricamente las funciones densidad?

Efectivamente, dichas funciones densidad de probabilidad se pueden ***caracterizar*** a partir su ***media*** y ***desvío estándar***:

Distribución	Media	D.E
UNIFORME	$(a+b)/2$	$(b-a)/\sqrt{12}$
POISSON	λ	$\sqrt{\lambda}$
GAUSSIANA	μ	σ
RAYLEIGH	$\sigma\sqrt{\pi/2}$	$\sigma\sqrt{(4-\pi)/2}$
EXPONENCIAL	$1/\lambda$	$1/\lambda$

¿Qué define la Distribución de Probabilidad Acumulada?

A partir de la función densidad $f_X(x)$, resulta factible calcular la ***probabilidad*** que la variable aleatoria ***X*** ***se encuentre dentro un intervalo [a,b]***:

$$\boxed{P(a \leq x \leq b) = \int_a^b f_X(x)dx}$$

Bajo dicha premisa, puede definirse la ***función de distribución acumulada $F_X(x)$***, que determina la probabilidad de que la variable aleatoria asuma ***valores inferiores o iguales a un valor x:***

$$\boxed{P(x \leq a) = F_X(x) = \int_{u=-\infty}^{x} f_X(u)du \Rightarrow f_X(x) = d\frac{F_X(x)}{dx}}$$

Nociones de Probabilidad

Consigna de la clase (20 minutos)

1. Se conectan en serie 2 resistencias R_1 y R_2 de 100Ω cada una, tomadas al azar de un lote, donde el valor está definido con un 10% de tolerancia. **Obtener:**

- *La densidad de probabilidad $f_{Ri}(x)$*
- *La densidad de probabilidad de la suma $R=R_1+R_2$ de las dos resistencias*
- *La probabilidad obtener una resistencia $\in [190,210]$*

NOTA: *Si X e Y son dos variables aleatorias independientes con funciones de densidad f y g, la* ***densidad de la suma*** *de ambas variables es $f*g$ (**convolución**)*