RESUMEN LAMINE YAMAL. NEW GEN.

Formulas Teorías de conjuntos.

Regla aditiva unión de eventos:

$$\boldsymbol{P(C) = P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$

$$P(D) = P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

Eventos mutuamente excluyentes:

$$P(A \cap B) = (\emptyset) = 0$$

$$P(C) = P(A \cup B) = P(A) + P(B)$$

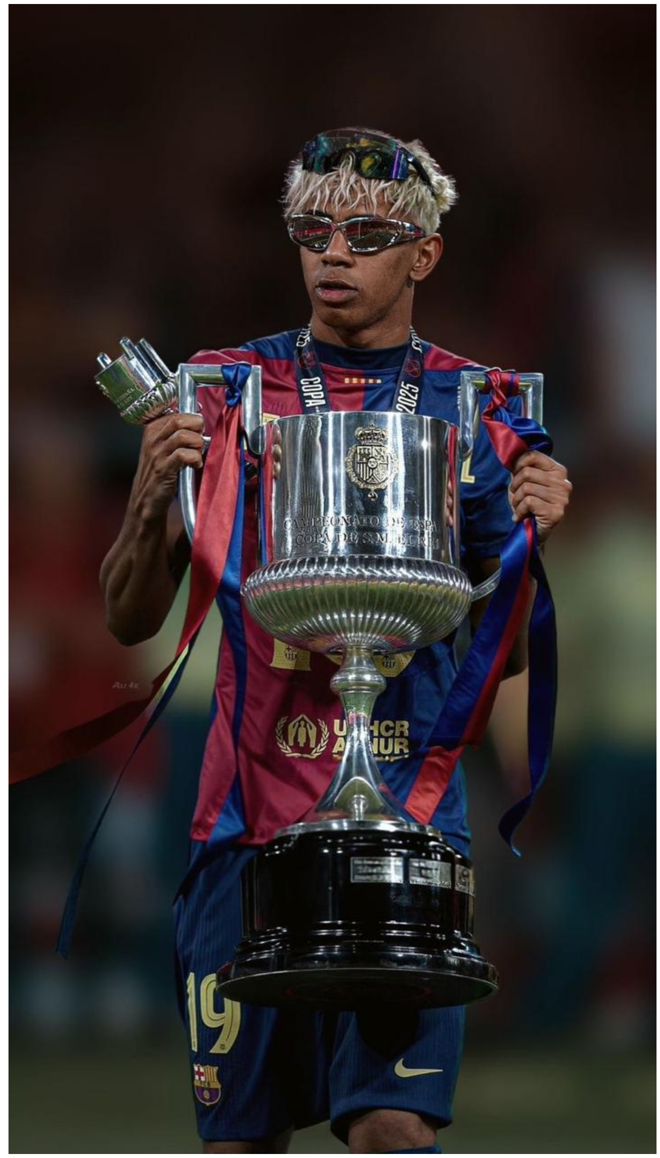
$$P(D) = P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

Eventos Complementarios:

$$P(B) = P(\overline{A})$$

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

$$P(A) = 1 - P(\overline{A}) = 1 - P(B)$$



Eventos Condicionales:

$$P(B/A) = \frac{P(A \cap B)}{P(A)}$$

Eventos Independientes:

$$P(B/A) = P(B); P(A/B) = P(A)$$

Probabilidad Total:

$$P(A) = \sum_{i=1}^{k} P(B_i \cap A) = \sum_{i=1}^{k} [P(B_i).P(A/B_i)]$$

Resumen de fórmulas de Tema 2

Datos sin agrupar	$NIC = log_{10} n$ *o* $NIC = \sqrt{n}$, $5 \leq NIC \leq 15$ ***Datos agrupados*** *Ancho del IC:* $A = \frac{R}{NIC}$
Medidas de posición o tendencia central	
Media: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ *Se puede obtener con Alcula.	$\bar{X} = \sum_{i=1}^{k}(X_{PMi}.\, fr_i)$
Mediana/cuartiles: n impar: $Q_k = X_{\frac{k}{4}(n+1)}$ con k=1,2,3. n par: $Q_k = \frac{1}{2}\left[X_{\left(\frac{k}{4}n\right)} + X_{\left(\frac{k}{4}n+1\right)}\right]$ con k=1,2,3	Se determina a cuál intervalo de clase pertenece el cuartil. Es el primero que tenga Fr $\geq \frac{k}{4}$. $Q_k = L_i + \frac{\frac{k}{4} - Fr_{(i-1)}}{Fr_i - Fr_{(i-1)}}.A$ con k=1,2,3
Moda: Valor de mayor frecuencia.	Se determina a cuál intervalo de clase pertenece la moda. Es aquél con mayor fr. $Mo = L_i + \frac{A}{\frac{f_i - f_{(i+1)}}{f_i - f_{(i-1)}} + 1}$
Percentiles: $P_k = X_{\left(\frac{k}{100}n\right)}$ con k=1,…,99	Se determina a cuál intervalo de clase pertenece el percentil. Es el primero que tenga Fr $\geq \frac{k}{100}$. $P_k = L_i + \frac{\frac{k}{100} - Fr_{(i-1)}}{Fr_i - Fr_{(i-1)}}.A$ con k=1,…,99
Medidas de dispersión	
Varianza: $S^2 = \frac{n\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2}{n(n-1)}$ **Desvío estándar:** $S = \sqrt{S^2}$ *Se puede obtener con Alcula.	$S^2 = \frac{n\sum_{i=1}^{k} X_{PM_i}^2 .\, fa_i - \left[\sum_{i=1}^{k}(X_{PM_i}.\, fa_i)\right]^2}{n(n-1)}$ $S = \sqrt{S^2}$
Rango: $R = X_{máx} - X_{mín}$	$R = X_{máx} - X_{mín}$
Rango intercuartil: $IQ = Q_3 - Q_1$	$IQ = Q_3 - Q_1$
Coeficiente de variación: $CV = S/\bar{X}$	$CV = S/\bar{X}$

Medidas de forma	
Asimetría: $$SK = \frac{3(\ X - X)}{S}$$	$$SK = \frac{3(\ X - X)}{S}$$
Curtosis: $$Cu = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_i - X\ )^4}{S^4} - 3$$	

<u>Teorema de Chebyshev</u>: en relación a un conjunto de datos cualquiera (poblacional o muestral) y una constante k > 1 cuando menos (1 - $1/k^2$) de los datos debe estar dentro de k desvíos estándar a uno y otro lado de la media para que la dispersión se considere pequeña.

Ejemplo: si elegimos $k = 2$ entonces $1 - \frac{1}{k^2} = \frac{3}{4} = 0{,}75$. El 75% de los datos debe estar a $\overline{X} + 2S$ y $\overline{X} - 2S$ para que la desviación se considere pequeña.

Estadistica Descriptiva Formulas:
Medidas de Posición o de Tendencia Central

Media Aritmética:

- de una muestra: $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
- de una población: $\mu = \frac{\sum_{i=1}^{N} X_i}{N}$

Media Ponderada:

- de una muestra: $\overline{X_w} = \frac{\sum_{i=1}^{n}(w.x)}{\sum_{i=1}^{n}(w)}$
- de una población: $\mu_w = \frac{\sum_{i=1}^{N}(w.x)}{\sum_{i=1}^{N}(w)}$

Mediana (centro de los datos):
Divide al conjunto de datos en dos partes iguales (*se deben ordenar los datos*).

$\tilde{X} = X_{\frac{n+1}{2}}$ n impar

$\tilde{X} = \frac{1}{2}\,[X_{\frac{n}{2}} + X_{\frac{n}{2}+1}]$ n par

Cuartiles, Deciles y Percentiles (con datos ordenados):

$Q_1 = X_{\frac{n+1}{4}}$; $Q_2 = X_{\frac{n+1}{2}} = \tilde{X}$; $Q_3 = X_{\frac{3.(n+1)}{4}}$ con n impar

$Q_1 = \frac{1}{2}.(X_{\frac{n}{4}} + X_{\frac{n}{4}+1})$; $Q_2 = \frac{1}{2}.(X_{\frac{n}{2}} + X_{\frac{n}{2}+1})$; $Q_3 = \frac{1}{2}.(X_{\frac{3.n}{4}} + X_{\frac{3}{4}.(n+1)})$ con n par

$D_1 = X_{\frac{n+1}{10}}$; $D_2 = X_{\frac{2}{10}.(n+1)}$ … n impar

$D_1 = \frac{1}{2}.(X_{\frac{n}{10}} + X_{\frac{n}{10}+1})$; $D_2 = \frac{1}{2}.(X_{\frac{2}{100}.n} + X_{\frac{2}{100}.(n+1)})$ … n par

$P_1 = X_{\frac{n+1}{100}}$; $P_2 = X_{\frac{2}{100}.(n+1)}$ … n impar

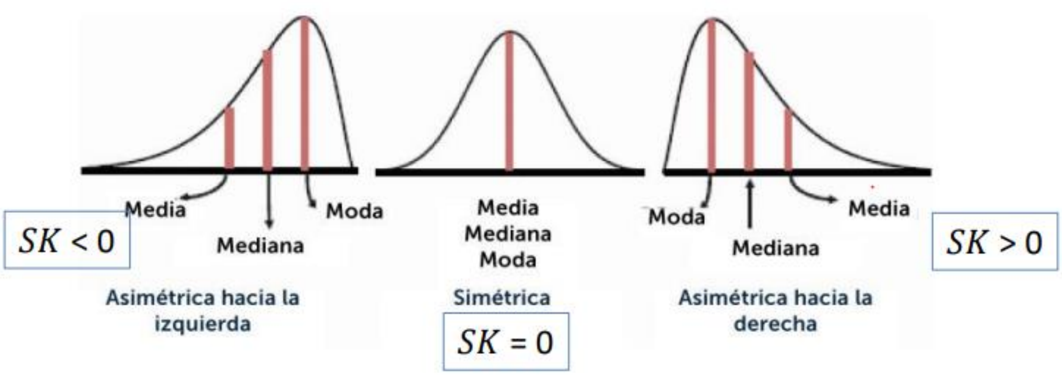
$P_1 = \frac{1}{2}.\left(X_{\frac{n}{100}} + X_{\frac{n}{100}+1}\right)$; $P_2 = \frac{1}{2}.\left(X_{\frac{2}{100}.n} + X_{\frac{2}{100}.(n+1)}\right)$ n par

Medidas de Dispersión

- Varianza
  - Muestra: $S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$ (n-1 grados de libertad de los datos de la muestra)
  - Población: $\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N}$

- Desvío estándar:
  - Muestra: $S = \sqrt{S^2}$
  - Población: $\sigma = \sqrt{\sigma^2}$
- Rango: $R = X_{max} - X_{min}$
- Coeficiente de Variación
  - Muestra: $CV = \frac{S}{\overline{X}}$
  - Población: $CV = \frac{\sigma}{\mu}$

Un índice de Asimetría muy utilizado es el de Pearson. $\boxed{SK = \frac{3(\overline{X} - \tilde{X})}{S}}$



Curtosis: Miden la mayor o menor concentración de datos alrededor de la media.

El grado de Curtosis es:

$$Cu = \frac{1}{n}\frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^4}{S^4} - 3$$

CU=0 (Mesocúrtica).
CU>0 (leptocúrtica).
CU<0(Platicurtica).

Teorema de Chebyshev

$1 - \frac{1}{k^2}$

$\overline{X} + 2S$ y $\overline{X} - 2S$

Cuartiles Datos agrupados:

- **Cuartiles:** $Q_k = Li + \frac{\frac{n}{4}.k - Fa_{(k-1)}}{fa_k} \, .A_{IC}$

Moda Datos Agrupados:

- **Moda:** $\hat{x} = Li + \frac{d_1}{d_1 + d_2} \, .A_{IC}$

con $d_1 = fa_k - fa_{k-1}$ y $d_2 = fa_k - fa_{k+1}$

Buscamos la clase con mayor frecuencia absoluta o relativa.

Una regla para determinar si un dato es **anómalo** (outlier) es:

- Si un dato es < Q1 – 1.5(Q3-Q1)
- Si un dato es > Q3 + 1.5(Q3-Q1)

- Mínimo: $x_{mín} = 31$
- Primer cuartil: $Q_1 = 61,8$
- Mediana: $\tilde{x} = 71,6$
- Tercer cuartil: $Q_3 = 83,25$
- Máximo: $x_{máx} = 97$

Intervalos de clase	Punto medio PM_i	Frecuencia fa_i	Frecuencia relativa $fr_i = \frac{fa_i}{n}$	Frecuencia acumulada Fa_i	Frec. relativa acumulada $Fr_i = \frac{Fa_i}{n}$	Frec. acumulada porcentual $Fr_{\%} = Fr_i . 100\%$
[31,42)	36.5	3	0.075	3	0.075	7.5

Medidas para datos AGRUPADOS

Nro. de intervalos de clase: $k = \sqrt{n}$ **Amplitud de intervalos de clase:** $A_{IC} = \frac{R}{k}$

Medidas de **tendencia central**

Media aritmética ($\bar{x}$)

$$\bar{x} = \sum_{i=1}^{k} \frac{x_{PM_i} fa_i}{n}$$

Mediana ($\tilde{x}$ o Me)

$$\tilde{x} = Li + \frac{\frac{n}{2} - Fa_{(k-1)}}{fa_k} . A_{IC}$$

Moda ($\hat{x}$ o Mo)

$$\hat{x} = Li + \frac{d_1}{d_1 + d_2} . A_{IC}$$

con $d_1 = fa_k - fa_{k-1}$ y $d_2 = fa_k - fa_{k+1}$

Medidas de **dispersión**

Rango (R)

$$R = x_{máx} - x_{mín}$$

Varianza (s^2)

$$s^2 = \sum_{i=1}^{n} \frac{(x_{PM_i} - \bar{x})^2 fa_i}{n-1}$$

Desvío estándar (s)

$$s = \sqrt{\sum_{i=1}^{n} \frac{(x_{PM_i} - \bar{x})^2 fa_i}{n-1}}$$

Rango intercuartil (RI)

$$RI = Q_3 - Q_1$$

Coeficiente de variación (CV)

$$CV = \frac{s}{\bar{x}}$$

Medidas de **posición**

Cuartiles (Q_k)

Números que dividen la distribución de lo datos en cuatro partes iguales.

$$Q_k = Li + \frac{\frac{n}{4}.k - Fa_{(k-1)}}{fa_k} . A_{IC}$$

con $k = 1,2,3$

Deciles (D_k)

Números que dividen la distribución de lo datos en diez partes iguales.

$$D_k = Li + \frac{\frac{n}{10}.k - Fa_{(k-1)}}{fa_k} . A_{IC}$$

con $k = 1, \dots, 9$

Percentiles (P_k)

Números que dividen la distribución de lo datos en cien partes iguales.

$$p_k = Li + \frac{\frac{n}{100}.k - Fa_{(k-1)}}{fa_k} . A_{IC}$$

con $k = 1, \dots, 9$

Medida de **forma**

Coeficiente de asimetría de Pearson (A_P)

$$A_P = \frac{3(\bar{x} - \tilde{x})}{s}$$

Medidas para datos SIN agrupar

Medidas de **tendencia central**

Media aritmética ($\bar{x}$)

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

Mediana ($\tilde{x}$ o Me)

$$x_{\left(\frac{n+1}{2}\right)}$$

Ordenando los datos de menor a mayor.

Moda ($\hat{x}$ o Mo)

Observación más frecuente.

Medidas de **dispersión**

Rango (R)

$$R = x_{máx} - x_{mín}$$

Varianza (S^2)

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}$$

Desvío estándar (s)

$$s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}}$$

Rango intercuartil (RI)

$$RI = Q_3 - Q_1$$

Coeficiente de variación (CV)

$$CV = \frac{s}{\bar{x}}$$

Medidas de **posición**

Ordenando los datos de menor a mayor.

Cuartiles (Q_k)

Números que dividen la distribución de lo datos en cuatro partes iguales.

Posición: $(n+1).\frac{k}{4}$

con $k = 1,2,3$

Deciles (D_k)

Números que dividen la distribución de lo datos en diez partes iguales.

Posición: $(n+1).\frac{k}{10}$

con $k = 1, \ldots, 9$

Percentiles (P_k)

Números que dividen la distribución de lo datos en cien partes iguales.

Posición: $(n+1).\frac{k}{100}$

con $k = 1, \ldots, 99$

Todas las medidas se expresan en la misma unidad que la variable estudiada, excepto la varianza (unidades al cuadrado) y el coeficiente de variación (carece de unidades).

Variables Aleatorias Unidimensionales.

Función de distribución: Tabla con los pares $(x, f(x))$

X	*f(x)*
0	1/8
1	3/8
2	3/8
3	1/8

Propiedades de una fdp de VAD

1. $f(x) \geq 0$

2. $\sum_{\forall x} f(x) = 1$

3. $P(X = x) = f(x)$

función de probabilidad VAD

Distribución acumulada $F(X)$

La $F(X)$ de una VAD con distribución de probabilidad $f(x)$ es:

$$\boxed{F(X) = P(X \leq x) = \sum_{t \leq x} f(t) \quad para \quad -\infty < X < +\infty}$$

Para el ejemplo anterior:

¿Qué probabilidad existe de comprar a lo sumo 1 defectuosa?

$$F(X \leq 1) = \sum_{t \leq 1} f(t) = f(0) + f(1) = \frac{10}{28} + \frac{15}{28} = \frac{25}{28}$$

¿Qué probabilidad existe de comprar al menos 1 defectuosa?

$$F(X \geq 1) = 1 - \sum_{t \leq 0} f(t) = 1 - [f(0)] = \frac{28}{28} - \frac{10}{28} = \frac{18}{28}$$

Distribución acumulada $F(X)$

La $F(X)$ de una VAD con distribución de probabilidad $f(x)$ es:

$$\boxed{F(X) = P(X \leq x) = \sum_{t \leq x} f(t) \quad para \quad -\infty < X < +\infty}$$

Para el ejemplo anterior:
¿Qué probabilidad existe de comprar a lo sumo 1 defectuosa?

$$F(X \leq 1) = \sum_{t \leq 1} f(t) = f(0) + f(1) = \frac{10}{28} + \frac{15}{28} = \frac{25}{28}$$

¿Qué probabilidad existe de comprar al menos 1 defectuosa?

$$F(X \geq 1) = 1 - \sum_{t \leq 0} f(t) = 1 - [\,f(0)] = \frac{28}{28} - \frac{10}{28} = \frac{18}{28}$$

Una función $f(x)$ es una función de densidad de probabilidad para una $VAC \in \mathbb{R}$ si:

Propiedades de una fdp de VAC

1. $f(x) \geq 0 \quad \forall x \in \mathbb{R}$

2. $\int_{-\infty}^{+\infty} f(x)\,dx = 1$

3. $P(a < x < b) = \int_a^b f(x)\,dx$

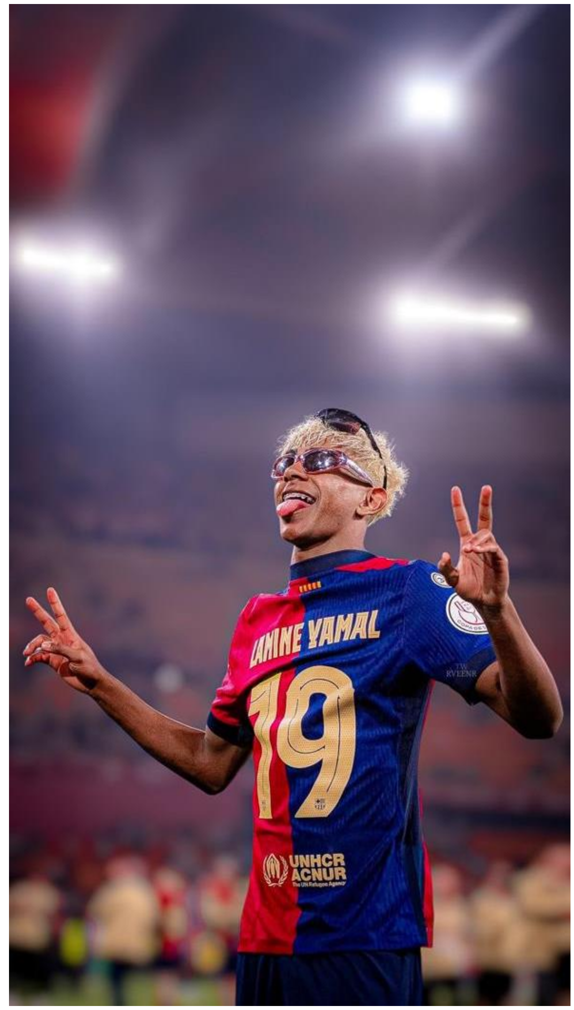
Función de distribución Acumulada de una VAC

$$\boxed{F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\,dx}$$

Variables Aleatorias bidimensionales

Propiedades:

Discreta	Continua
$f(x,y) \geq 0\,;\ \forall(x,y)$	$f(x,y)\ \geq 0\ \ ;\ f(x,y)$
$\sum_{\forall x}\sum_{\forall y} f(x,y)\ =\ 1;\ \forall(x,y)$	$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy\ =\ 1$
$P(X\ =\ x, Y =\ y)\ = f(x,y)$	$P(a < x < b, c < y < d\,)\ = \int_a^b \int_c^d f(x,y)dx\,dy$ $si\ \{a\ \leq x \leq b\,;\ c \leq y \leq d\}$
$P[(x,y)\ \in A]\ = \sum_x \sum_y f(x,y)$	



Distribuciones Marginales

$$g(x) = \sum_{\forall y} f(x,y) \qquad y \qquad h(y) = \sum_{\forall x} f(x,y)$$ VAD

$$g(x) = \int_{y=-\infty}^{+\infty} f(x,y)dy \qquad y \qquad h(y) = \int_{x=-\infty}^{+\infty} f(x,y)dx$$ VAC

Distribución Probabilidad Condicional

$$f({}^{y}/_{x}) = \frac{f(x,y)}{g(x)}$$

De la misma forma:

$$f({}^{x}/_{y}) = \frac{f(x,y)}{h(y)}$$

$$f(x,y) = g(x).h(y)$$

$$f(x,y) = h(y).g(x)$$

CONDICIÓN
DE
INDEPENDENCIA
ESTADÍSTICA

Regla de Bayes (Probabilidad de las causas)

Si los eventos $B_1, B_2, \ldots, B_k$ constituyen una partición del Espacio Muestral S y $P(B_i) \neq 0$ para $i = 1,2,\ldots,k$ entonces cualquier evento B_r en S tal que $P(A) \neq 0$ puede expresarse como $P(B_r/A) = P(B_r \cap A) / P(A)$ $[\,Probabilidad\ Condicional]$.
Como $P(A/B_r) = P(B_r \cap A) / P(B_r) \Rightarrow P(B_r \cap A) = P(B_r) \,.\, P(A/B_r)$

$$P(B_r/A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)\,.\,P(A/B_r)}{\sum_{i=1}^{k}[P(B_i)\,.\,P(A/B_i)]}$$

Ejemplo 21: (Con respecto al caso anterior del Ejemplo 20):
¿Qué probabilidad existe de que lo haya fabricado la máquina 2 (B_2) sabiendo que el producto elegido al azar es defectuoso?

$$P(B_2/A) = \frac{P(B_2 \cap A)}{\sum_{i=1}^{3} P(B_i \cap A)} = \frac{P(B_2)\,.\,P(A/B_2)}{\sum_{i=1}^{3}[P(B_i)\,.\,P(A/B_i)]}$$

$$P(B_2/A) = \frac{(0.45)\,.\,(0.03)}{(0.3)\,.\,(0.02) + (0.45)\,.\,(0.03) + (0.25)\,.\,(0.02)} = \frac{0.0135}{0.006 + 0.0135 + 0.005} = 0.55$$

Teorema de probabilidad total

Dado un evento ***A*** que puede ser descripto por suma de las intersecciones de éste con eventos mutuamente excluyentes:

$$B_i = \sum_{i=1}^{k} P(B_i) = 1\,,$$ formados como una partición del Espacio Muestral ***S***.

Entonces, según la fórmula de intersección de eventos: