Medidas de Dispersión

- Miden que tanto se dispersan o desvían los datos en torno a la media.

- El *rango* es la medida de dispersión más simple (y menos útil). El rango es simplemente la diferencia entre la observación más alta y la más baja.

- La desventaja es que sólo considera dos de los (posiblemente) cientos de observaciones, ignorando el resto de los datos.

Medidas de Dispersión - Varianza

- La varianza es el promedio de las desviaciones de las observaciones con respecto a su media al cuadrado.

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

- La desviación estándar es la raíz de la varianza. Es una medida muy útil de dispersión ya que tiene las mismas unidades que la variable estudiada.

$$\sigma = \sqrt{\sigma^2}$$

Medidas de Dispersión – Varianza(2)

- La *varianza muestral* sigue la misma lógica:

$$s^2 = \frac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{n-1} \qquad s = \sqrt{s^2}$$

- Llama la atención que se divida por *n-1*, lo que se debe a que este estadístico tiene *n-1 grados de libertad*.

- Los grados de libertad equivalen al número de observaciones menos el número de *restricciones* impuesta en tales observaciones.

Medidas de Dispersión – Varianza(3)

- Una restricción es cualquier valor que deba calcularse de dichas observaciones. En este caso la restricción es el cálculo de la media muestral.
- Ej: Se tienen n=4 obs. que producen una media de 10, la media de 10 sirve como restricción y hay n-1=3 g.l.
- Se pueden escoger 3 obs. Cualquiera, por ej. 8, 9 y 11.
- Después ya no hay libertad para escoger la última obs., que debe ser 12 si se quiere tener un promedio de 10.

Medidas de Dispersión

- Otras medidas de dispersión son los *cuartiles*, los *deciles* y los *percentiles*.

- Cada conjunto de datos tiene tres cuartiles que lo dividen en cuatro partes iguales. El primer cuartil (inferior) cuenta con el 25% de las observaciones. El segundo es justo la mitad (50% de los datos) y el tercero el 25% superior.

- Los deciles separan el conjunto de datos en 10 subconjuntos iguales, y los percentiles en 100 partes.

Medidas de Dispersión(2)

- Un percentil y su ubicación en un arreglo ordenado se identifica mediante subíndices. Por ejemplo, el 15vo percentil es indica como P_{15}, y su ubicación en la serie es L_{15}

- El lugar del P-ésimo percentil es: $L_p = (n+1)\frac{P}{100}$

- Ejemplo: número de acciones transadas en la Bolsa de Valores de Nueva York:

Medidas de Dispersión(3)

3	10	19	27	34	38	48	56	67	74
4	12	**20**	29	34	39	48	59	67	74
7	14	**21**	31	36	43	52	62	69	76
9	15	25	31	37	45	53	63	72	79
10	17	27	34	38	47	56	64	73	80

- Se desea calcular el percentil 25, P_{25}, para estas acciones. Primero debe hallarse su ubicación en la serie:

$$L_{25} = (50+1)\frac{25}{100} = 12.75$$

Medidas de Dispersión(3)

- El valor resultante (12.75) dice que el percentil 25 está ubicado al 75% del trayecto comprendido entre la doceava (20) y la treceava (21) observación, i.e.

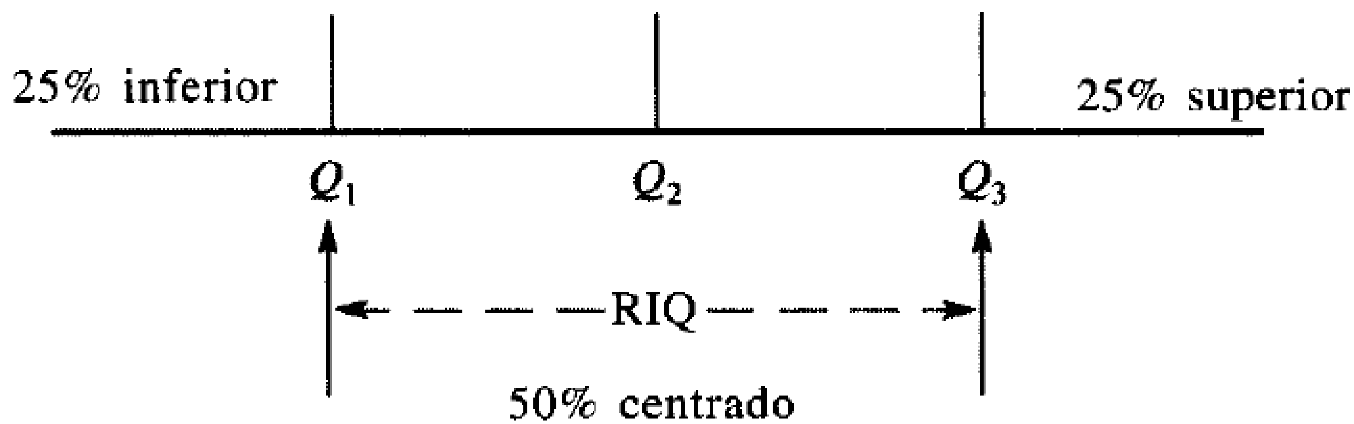
$$P_{25} = 20 + 0.75(21 - 20) = 20.75$$

- Notar que el primer decil es equivalente a P_{10}, el segundo P_{20} y así sucesivamente.

- El primer cuartil es igual a P_{25}, el segundo a P_{50} y el tercero a P_{75}

Medidas de Dispersión(4)

- Una medida única de dispersión es el rango intercuartílico (RIQ), la diferencia entre el tercer cuartil y el primer cuartil (50% de las obs.).



- Esta medida no está muy influenciada por observaciones extremas

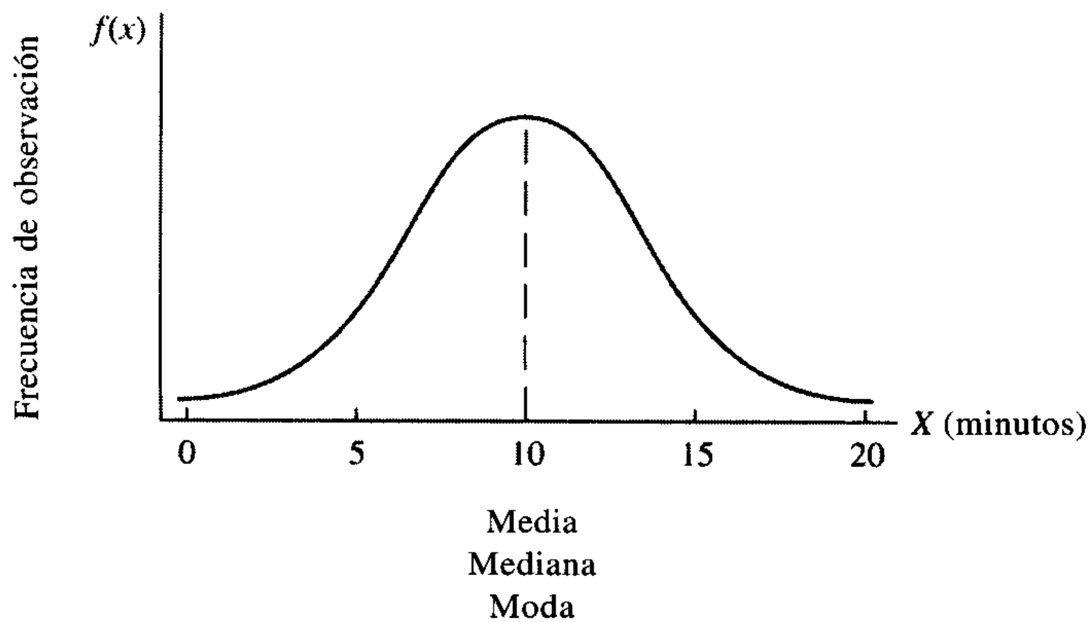
Aplicaciones de la desviación estándar

- **Teorema de Chebyshev:** Establece que para todo conjunto de datos, por lo menos un $1-1/K^2\%$ $(K>1)$ de las observaciones están dentro de K desviaciones estándar de la media.

- Ej: $K=3$ (3 desviaciones estándar por debajo y 3 por encima de la media), entonces por el $(1-1/3^2)=88.9\%$ de los datos debería caer en dicho intervalo.

Aplicaciones de la desviación estándar(2)

- **Distribución Normal y Regla Empírica:** En forma simple, la distribución normal (datos continuos) produce una curva simétrica en forma de campana:



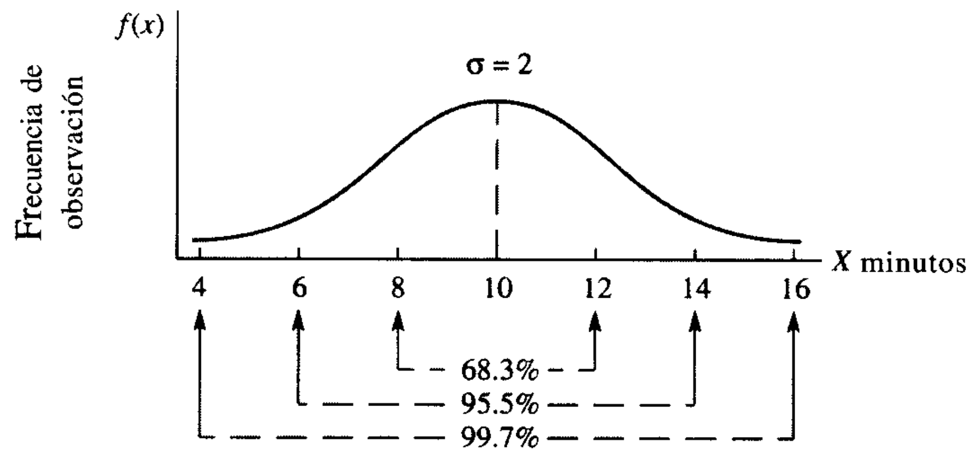
Aplicaciones de la desviación estándar(3)

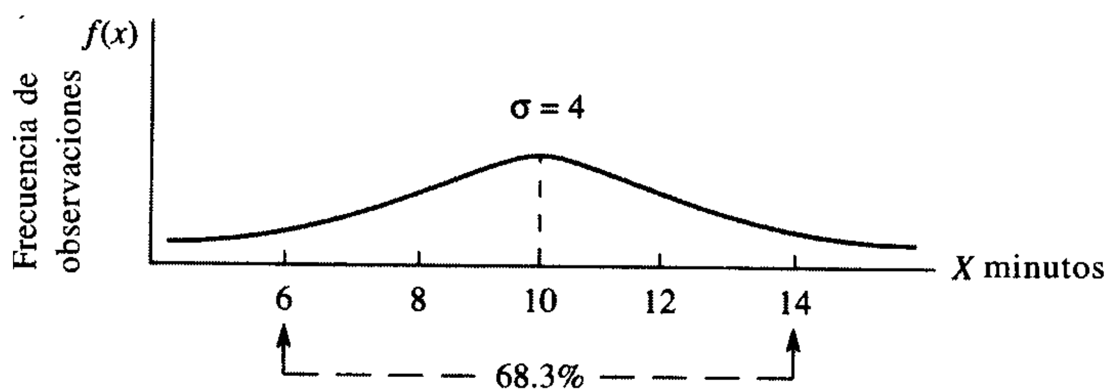
- **Distribución Normal y Regla Empírica(2):** La regla empírica dice que si se incluyen todas las observaciones que están a una desviación estándar de la media, éstas serán el 68.3% de todas las observaciones.

- De la misma manera: 95.5% de los datos están dentro de dos desv. estándar y 99.7% de las obs. están dentro de tres desv. estándar.

Aplicaciones de la desviación estándar(4)

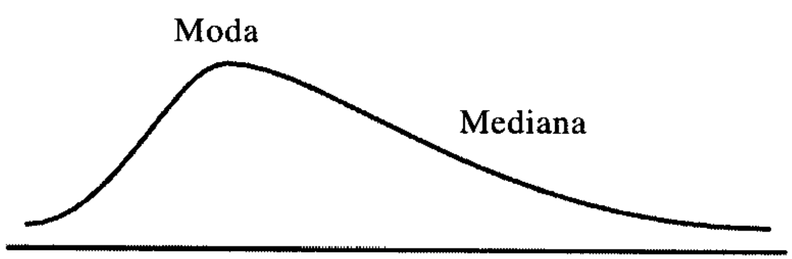
- Ejemplo:

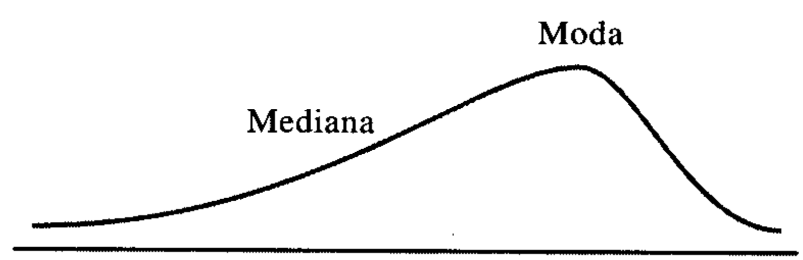




Asimetría

- **Asimetría o Sesgo (*skewness*):** No todas las distribuciones son normales, algunas están sesgadas a la izquierda o derecha:





- Se define el coeficiente de Asimetría o Sesgo (Pearson):

$$A_p = \frac{\mu - moda}{\sigma},$$

Asimetría (2)

Interpretación:

- Si $P<0$, los datos están sesgados a la izquierda (asimetría negativa)
- si $P>0$ entonces están sesgados a la derecha (asimetría positiva)
- $P=0$ implica que los datos se distribuyen normalmente.

Curtosis

- Se basa en el promedio de las desviaciones típicas a la cuarta potencia y representa el apuntalamiento de la distribución:

$$\sum_{i=1}^{n}\left[\frac{(x_i - \bar{X})^4 / S^4}{n}\right] - 3$$

- Se corrige en 3 que corresponde a la curtosis de la normal.
- Si vale cero es *mesocúrtica*
- Si es positiva es más apuntalada que la normal y se llama leptocúrtica.
- Si es negativa es más achatada que la normal y se llama platicúrtica.

Curtosis(2)

- La curtosis es uno de los conceptos peor comprendidos en la estadística. Se suele confundir con la varianza:

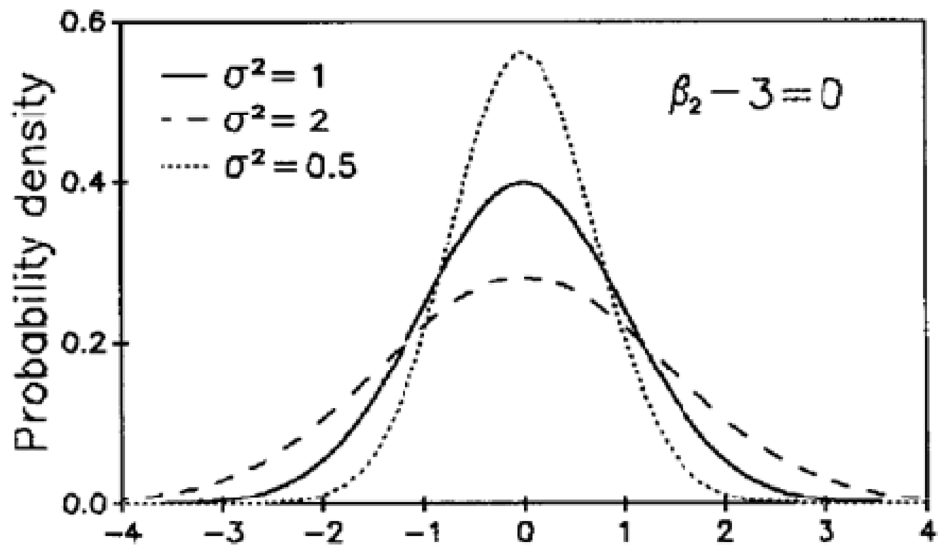
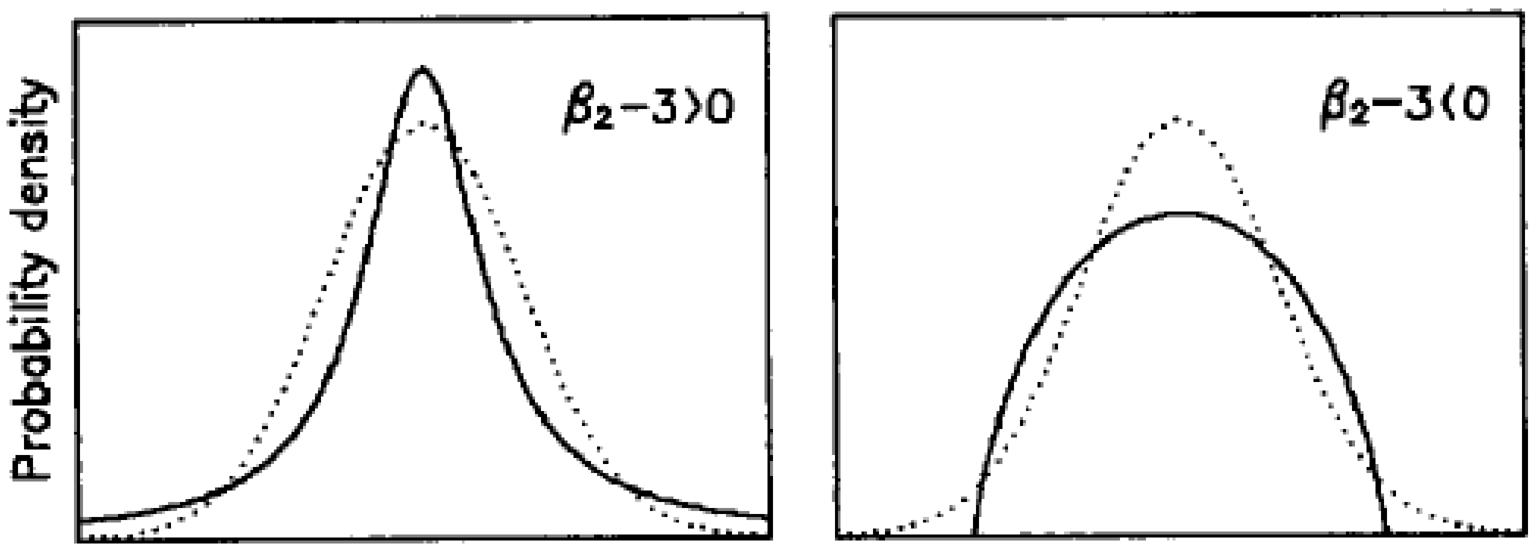


Figure 4. Three normal distributions with variances (σ^2) of 0.5, 1, and 2.

Curtosis(3)

- La curtosis representa una medida adimensional que representa un movimiento de masa que no afecta la varianza.



Probabilidades

- La *probabilidad* es la posibilidad numérica de que ocurra un evento.

- La probabilidad de un evento es medida por valores comprendidos entre 0 (imposibilidad) y 1 (certeza).

- Un *experimento* (aleatorio) es una acción que puede tener distintos resultados posibles.

- El *espacio muestral* es el conjunto de resultados posibles de un experimento. Se suele representar por EM o **Ω**.

Probabilidades(2)

- Ejemplos:
  - **Experimento 1: “tirar una moneda y ver que sale”**

  $\mathbf{\Omega}$ ={C,S}

  - **Experimento 2: “tirar dos monedas y ver que sale”**

  $\mathbf{\Omega}$ ={CC,SS,CS}

- Un *suceso* es un subconjunto del espacio muestral

  S_1={C}, S_2={S}, S_3={CC,SS}

Probabilidades(3)

- Modelo clásico (Laplace):

$$P(E) = \frac{\text{Número de formas que puede ocurrir el evento E}}{\text{Número total de posibles resultados}}$$

- Ejemplo: P(cara)=1/2

Probabilidades(4)

- Modelo empírico:

$$P(E) = \frac{\text{Número de veces que ha ocurrido el evento E}}{\text{Número total de experimentos}}$$

- Ejemplo: P(cara)

Definición Axiomática

- Axioma 1: la probabilidad no puede ser negativa:

$$P(A) \geq 0$$

- Axioma 2: la probabilidad del espacio muestral es uno

$$P(EM) = 1$$

- Axioma 3: dos conjuntos son disjuntos ssi la probabilidad de su unión es la suma de sus probabilidades:

$$A \cap B = \emptyset \leftrightarrow P(A \cup B) = P(A) + P(B)$$

Consecuencias

- Consecuencia 1: $P(A) \leq 1$
- Consecuencia 2: $P(A) + P(\bar{A}) = 1$
- Consecuencia 3: $P(\emptyset) = 0$
- Consecuencia 4: $A \subseteq B \rightarrow P(A) \leq P(B)$
- Consecuencia 5: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Probabilidad Condicional

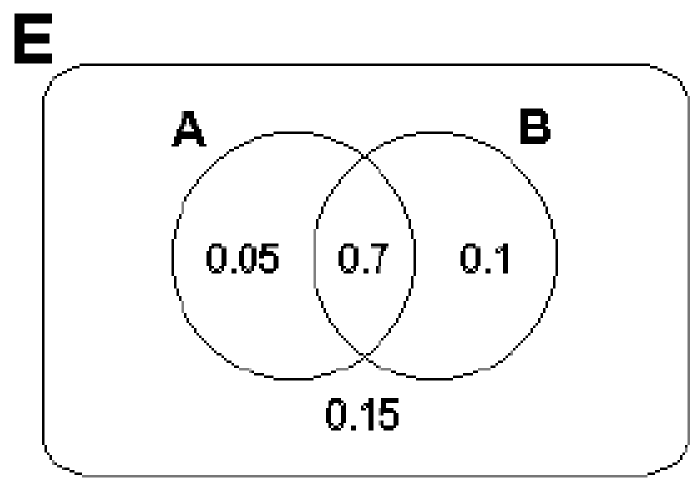
- *Probabilidad condicional* P(A|B) es la probabilidad de que ocurra el evento A, dado que el evento B ya haya ocurrido.
- Ejemplo:
  - **el 80% de los alumnos estudió para el examen**
  - **el 75% de los alumnos aprobó el examen**
  - **el 15% de los alumnos no estudió para el examen y no aprobó.**

- Sea A el suceso "alumno aprobó examen" y B el suceso "el alumno estudió". Se tiene que P(A)=0.75, P(B)=0.8 y

$$P(\bar{A} \cap \bar{B}) = 0.15$$

Probabilidad Condicional(2)

- Gráficamente:



- **Cual es la probabilidad de que un alumno que estudió haya aprobado el examen?**

Probabilidad Condicional(3)

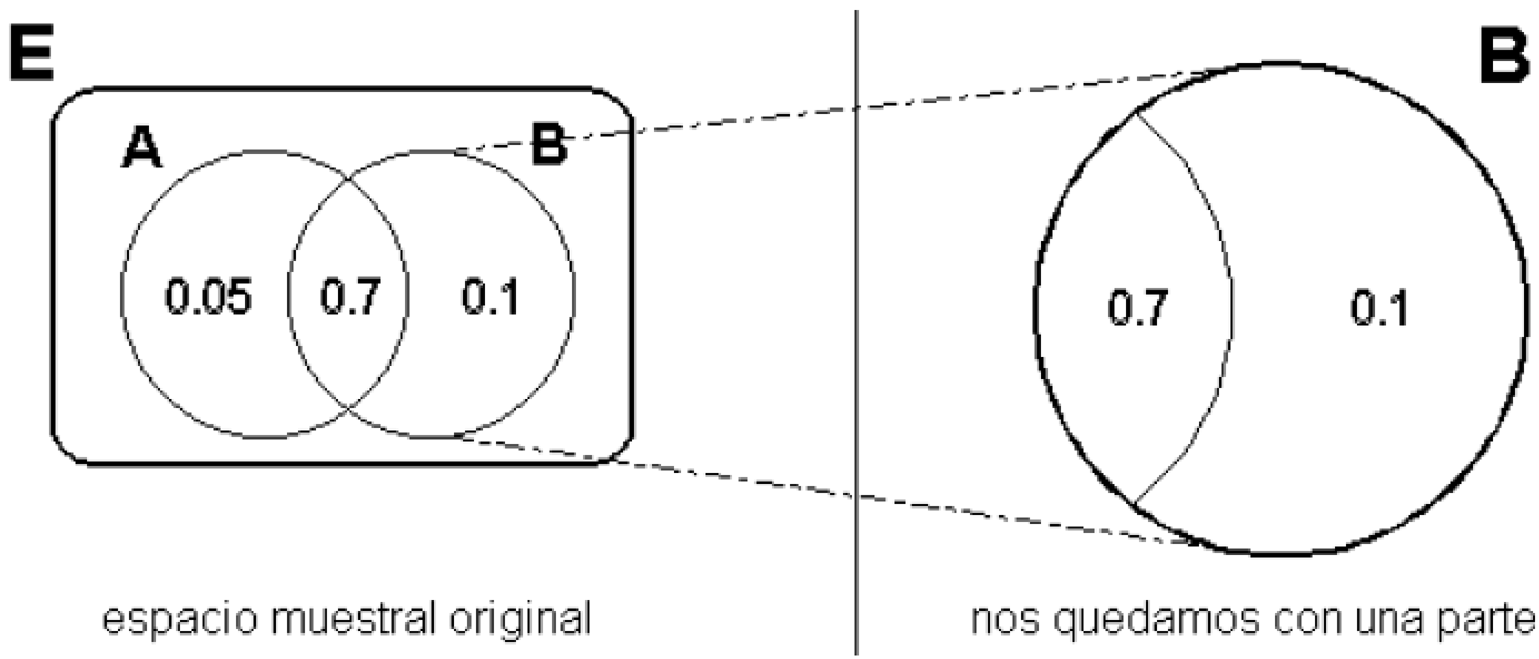
- Intuitivamente, los alumnos que estudiaron fueron el 80%

- Ese 80% está formado por un 70% que aprobó y un 10% que no aprobó. La probabilidad de aprobar es 70/80=0,88

- Formalmente: $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

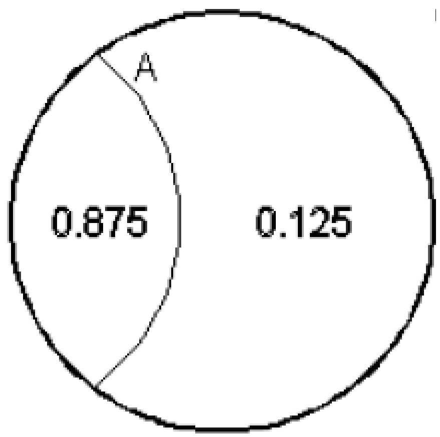
Probabilidad Condicional(4)

- Intuitivamente, P(A|B) es la probabilidad de “estar parados en A, sabiendo que estamos parados en B".



Probabilidad Condicional(5)

- Sin embargo, B no está listo para ser espacio muestral (probabilidades no suman 1)
- Es necesario dividir las probabilidades de B por un factor para que sea EM manteniendo la proporción relativa.
- Como las probabilidades contenidas en B suman P(B), dividiendo por este factor se cumple lo anterior



el nuevo espacio muestral

Propiedades

- Conmutatividad intersección:

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) = P(B \cap A)$$

- Intersección 3 eventos:

$$P(A \cap B \cap C) = P(A)P(A|B)P(C|A \cap B)$$

- Principio Bayes:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Independencia

Dos sucesos A,B son independientes ssi:

- $P(A|B) = P(A)$

$$\Leftrightarrow$$

- $P(B|A) = P(B)$

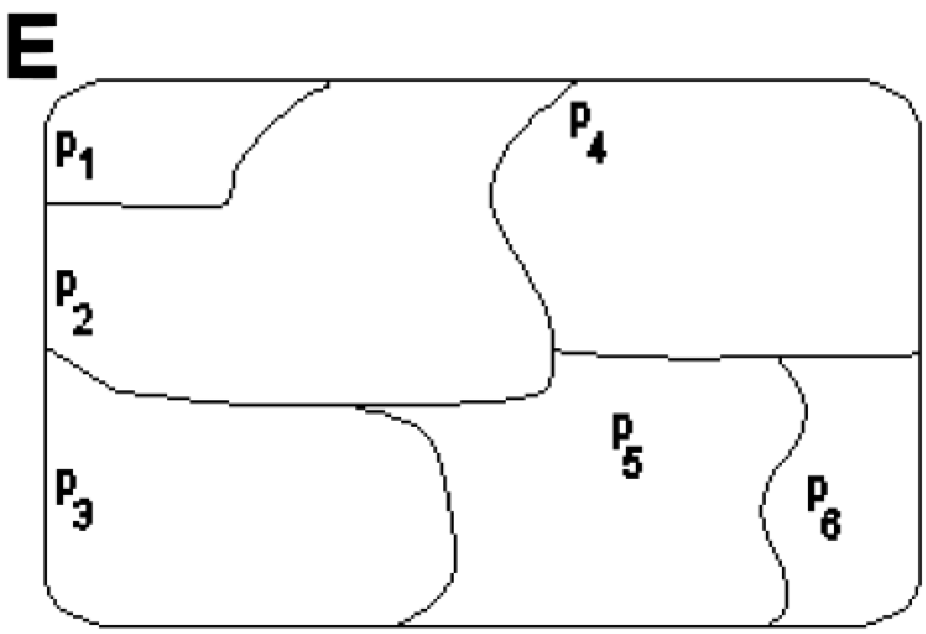
$$\Leftrightarrow$$

- $$P(A \cap B) = P(A)P(B)$$

- Advertencia: La independencia de dos sucesos no tiene nada que ver con que dos sucesos sean disjuntos. De hecho, si dos sucesos, con probabilidades no nulas, son independientes, entonces no pueden ser disjuntos, ya que $p(A\cap B)=p(A).p(B) \neq 0$.

Probabilidades Totales

- Consideremos un espacio muestral E, con la siguiente partición:

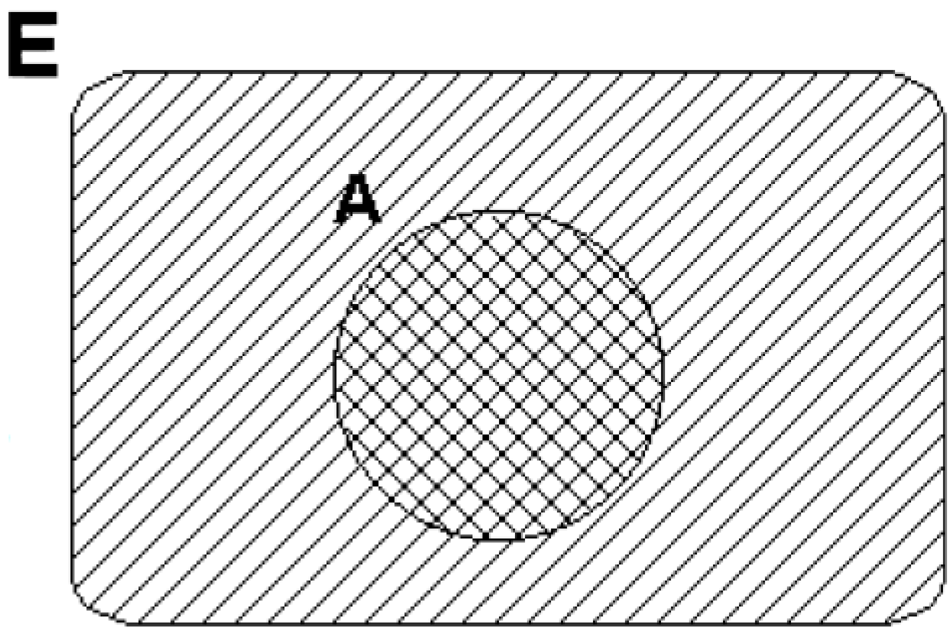


$$E = \sum_i p_i$$

$$p_i \cap p_j = \emptyset$$

Probabilidades Totales(2)

- Además se cuenta con el suceso A, que es subconjunto del espacio muestral:



$$P(A) = P(A \cap E)$$

Probabilidades Totales(3)

- Dado que E es la sumatoria de las probabilidades de la partición establecida:

$$P(A) = P(A \cap E) = P\left(A \cap \sum_i p_i\right)$$

- Aplicando la propiedad distributiva de conjuntos:

$$P\big(A \cap (p_1 \cup p_2 \cup \ldots \cup p_n)\big) = P\big((A \cap p_1) \cup (A \cap p_2) \cup \ldots (A \cap p_n)\big)$$

Probabilidades Totales(4)

- Utilizando el tercer axioma podemos escribir la probabili-dad de la suma (unión) como suma de probabilidades:

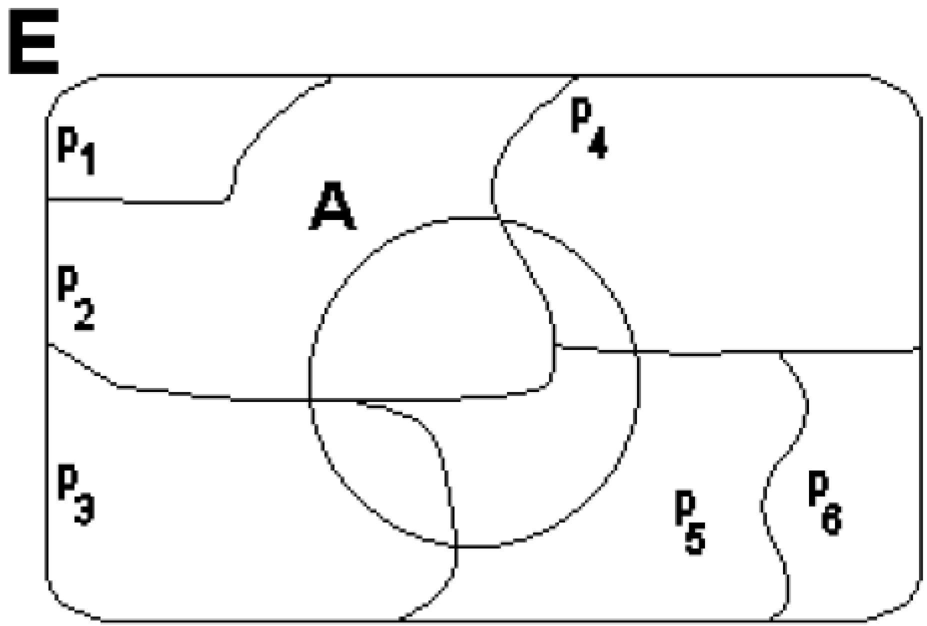
$$P\big((A \cap p_1) \cup \ldots \cup (A \cap p_n)\big)$$
$$= P\big((A \cap p_1)\big) + \cdots + P\big((A \cap p_n)\big)$$

- En resumen, llegamos a lo que se conoce como *probabilidad total:*

$$\sum_i P(A \cap p_i)$$

Probabilidades Totales(5)

- Gráficamente:



Probabilidades Totales(5)

- En particular, para una partición de un suceso D y su complemento:

$$P(A) = \mathrm{P}(A \cap D) + \mathrm{P}(A \cap \overline{D})$$

- Utilizando ahora la fórmula de probabilidad condicional:

$$P(A) = P(A|D)P(D) + P(A|\overline{D})P(\overline{D})$$

- En general:

$$P(A) = \sum_i P(A \cap p_i) = \sum_i P(A|p_i)\, P(p_i)$$

Teorema de Bayes

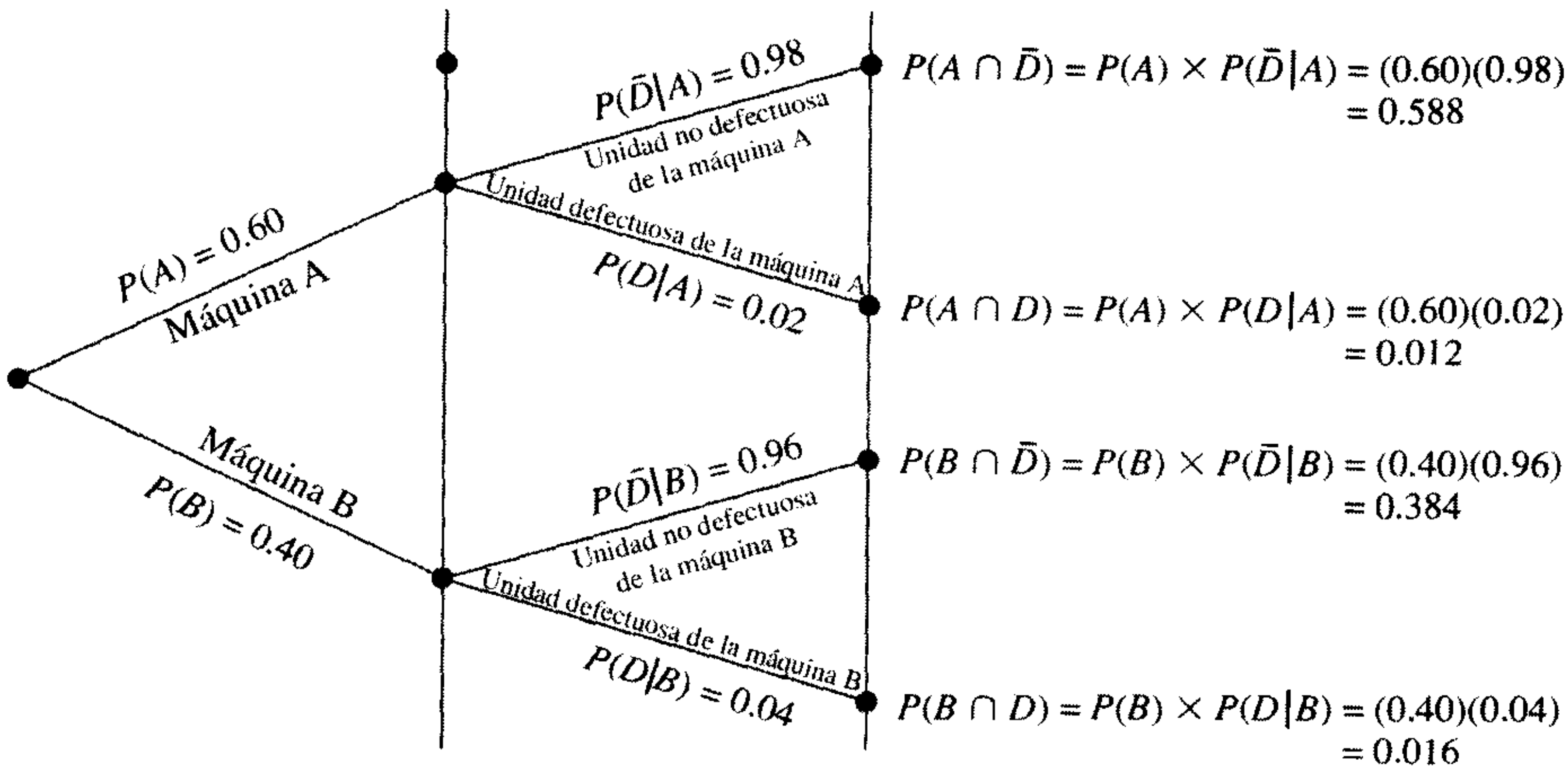
- Ejemplo: En una empresa manufacturera, una máquina A produce el 60% de la producción total, mientras que una máquina B el restante 40%.

- El 2% de las unidades producidas por A son defectuosas, mientras que B tiene una tasa de defectos del 4%.

- Se cuenta con una unidad defectuosa, se desea conocer la probabilidad de que venga de la máquina A.

Teorema de Bayes(2)

- Árbol:



Teorema de Bayes(3)

- La probabilidad P(A|D) se puede obtener utilizando la tercera propiedad obtenida por la probabilidad condicional.

- Sin embargo, se desconoce P(D). Necesitamos aplicar probabilidades totales:

$$P(D) = P(D|A)P(A) + P(D|B)P(B)$$

- Bayes: $$P(A|D) = \frac{P(D|A)P(A)}{P(D|A)P(A) + P(D|B)P(B)}$$

Teorema de Bayes(4)

- Volviendo al problema:

$$P(A|D) = \frac{0.012}{0.012 + 0.016} = 0.43$$

- Tiene sentido? $P(A|D)<0.5$? $P(A|D)<P(A)$?