# Probabilidad

1. **Conceptos Básicos de Probabilidad**
2. **Espacio Muestral**
3. **Tipos de Eventos**
4. **Eventos Dependientes e Independientes**
5. **Probabilidad Condicional**
6. **Reglas de la Probabilidad**
7. **Combinatoria Básica**
8. **Distribuciones de Probabilidad**
9. **Aplicación de la Probabilidad en la Ciencia Política**

# Introducción a la probabilidad

## Distintos tipos de probabilidad

- PROBABILIDAD CLÁSICA (equiprobable) →

  P(A) = casos favorables a A / casos totales

- PROBABILIDAD COMO FRECUENCIA RELATIVA →

  P(A) = cant de veces que ocurre A / nro total de experimentos

- PROBABILIDAD SUBJETIVA (BAYESIANA) →

  la probabilidad se actualiza a medida que se obtiene nueva información

- DEFINICIÓN AXIOMÁTICA (MATEMÁTICA) DE PROBABILIDAD →

  reglas de la probabilidad

# ¿Qué es la probabilidad?

¿Qué es la probabilidad?

¿Qué la diferencia de la inferencia?

- Teoría de probabilidad: provee modelos para entender (ordenar) los experimentos o situaciones inciertas o azarosas, sobre las cuales no conocemos el resultado hasta que ocurren

- Inferencia estadística: provee herramientas para, a partir de una muestra, afirmar con fundamentos respecto de algún elemento o característica de la población. Posibilidad de tomar decisiones.

- Probabilidad asociada al azar ⇒ Probabilidad frecuentista
- Probabilidad asociada a la incertidumbre/ignorancia/desconocimiento ⇒ Probabilidad bayesiana

# Ejemplo: Tirar una moneda 5 veces

**Resultado = {Cara, Cara, Cruz, Cara, Cruz}**

**¿Cuál es la probabilidad de que salga cara en un tiro?**

**Prob (cara) = ?**

¿ 3 / 5 ?

~~¿ 3 / 5 ?~~

~~¿ 3 / 5 ?~~

**Frecuencia relativa (de cara)**

3 / 5 $\Rightarrow$ Dato que viene de la observación

$\neq$

Probabilidad

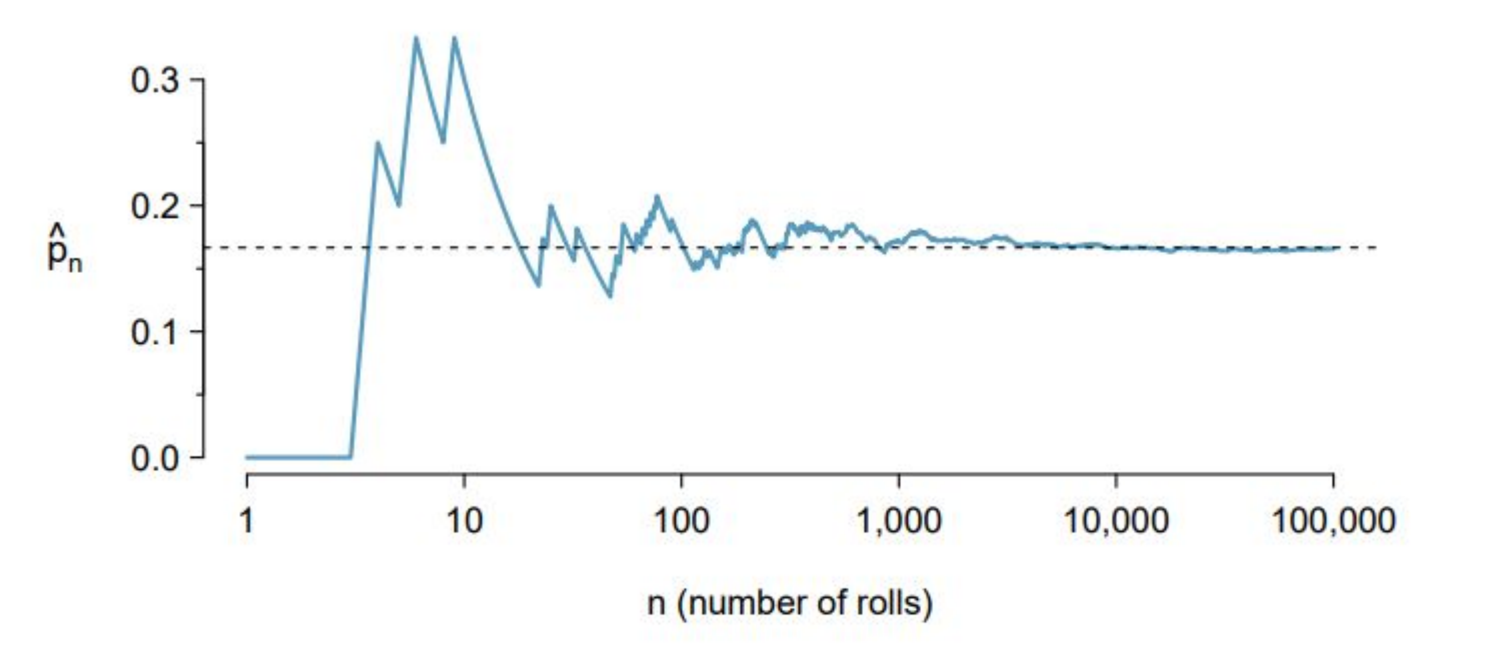
Teoría frecuentista: relación entre teoría y observación:

Cuando
Lim tiros $\rightarrow \infty$

Freq. relativa (cara) $\approx$ Prob(cara)

# Vínculo entre frecuencia relativa y probabilidad

- **pn** representa la proporción de resultados = 1 a medida que se tira el dado (eje x)
- Vemos que a medida que se incrementa n la proporción de 1 sobre el total converge a **p= ⅙**, que es la probabilidad de obtener un 1.
- Esta convergencia se da por la ley de los grandes números



# Definiciones de probabilidad

- **EXPERIMENTO ALEATORIO: proceso con resultado incierto. Puede repetirse en iguales condiciones para la obtención de “datos”**

# Definiciones de probabilidad

- EXPERIMENTO ALEATORIO: proceso con resultado azaroso. Puede repetirse en iguales condiciones para la obtención de “datos”
- **RESULTADOS: realizaciones del experimento aleatorio. No conocemos con seguridad el resultado, pero sí podemos conocer la probabilidad de ocurrencia de cada resultado**

# Definiciones de probabilidad

- EXPERIMENTO ALEATORIO: proceso con resultado azaroso. Puede repetirse en iguales condiciones para la obtención de “datos”
- RESULTADOS: realizaciones del experimento aleatorio. No conocemos con seguridad el resultado, pero sí podemos conocer la probabilidad de ocurrencia de cada resultado
- **ESPACIO MUESTRAL: conjunto de todos los posibles resultados de un experimento aleatorio**

# Definiciones de probabilidad

- EXPERIMENTO ALEATORIO: proceso con resultado azaroso. Puede repetirse en iguales condiciones para la obtención de "datos"
- RESULTADOS: realizaciones del experimento aleatorio. No conocemos con seguridad el resultado, pero sí podemos conocer la probabilidad de ocurrencia de cada resultado
- ESPACIO MUESTRAL: conjunto de todos los posibles resultados de un experimento aleatorio
- **EVENTO: algún subconjunto del espacio muestral**

# Experimento aleatorio

Por ejemplo: tirar un dado, una moneda, sacar una carta de un mazo.

Todos estos experimentos pueden repetirse bajo las mismas condiciones (diferencia con los experimentos o eventos sociales).

Si bien conocemos cuáles son sus posibles salidas o resultados (sacar alguna de las cartas del mazo, sacar números del 1-6 en el dado, etc), no es posible predecir con exactitud cuál será el resultado al ejecutarlo.

## Experimento aleatorio $\neq$ experimento social

Evaluar la probabilidad de ocurrencia de cierto evento social no es independiente del contexto y de lo social. Por eso no es un experimento que pueda repetirse siempre bajo las mismas condiciones. Probabilidad **condicional**.

Por ejemplo: ¿qué probabilidad existe de que durante este mes se produzca una huelga laboral en Argentina?

# Experimento aleatorio ≠ experimento social

La frecuencia relativa se basa en **observar qué tan seguido ocurre un evento en la práctica**, a través de repeticiones del experimento.

- Para la huelga, podemos considerar los datos históricos:
  - Supongamos que en los últimos 120 meses hubo 12 huelgas.
  - La **frecuencia relativa** sería: 12/120=0,11 → “probabilidad histórica” de que haya huelga en un mes cualquiera = 10%.
- **Limitación:** Esto **asume que las condiciones son comparables**, pero sabemos que factores económicos y sociales varían mes a mes, por lo que la probabilidad real es condicional a esas circunstancias.

## Resultados

Lo que se obtiene al “tirar” la moneda o el dado o sacar la carta es un resultado.

Resultados posibles de tirar un dado = {1, 2, 3, 4, 5, 6}

Resultados posibles de tirar una moneda = {cara, cruz}

Existe una probabilidad asociada a cada resultado:

Prob(cara)

Prob(cruz)

## Espacio muestral $\Omega$

Conjunto de todas las posibles salidas, resultados, eventos que pueden obtenerse

Ejemplo:

$\Omega = \{\varnothing, \{1, 2, 3, 4, 5, 6\}, \{1, 2\}, \{2, 4, 6\}, \{1, 3, 5\} \ldots \}$

## Eventos

A partir de la combinación de resultados obtenemos eventos posibles.

$$\boldsymbol{\Omega} = \{\varnothing, \{1, 2, 3, 4, 5, 6\}, \{1, 2\}, \{2, 4, 6\}, \{1, 3, 5\}$$

En algunos casos puede interesarnos estimar la probabilidad de un evento más que la de un resultado.

Por ejemplo: una apuesta a partir de si sale un dado par o impar

## Definición de la probabilidad: teoría de conjuntos

A partir de dos eventos del espacio muestral **Ω** (A y B):

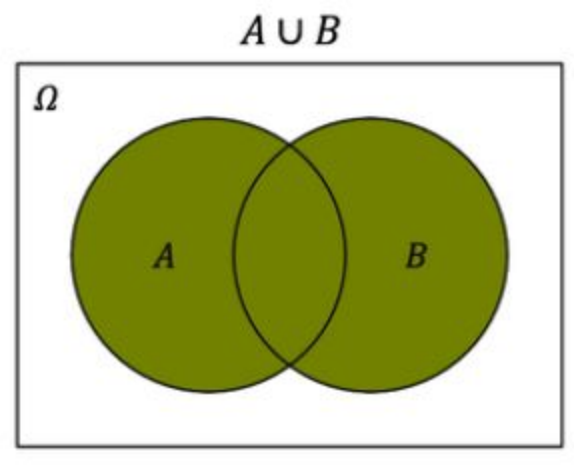
Evento A = {1, 3, 5} Evento B = {menor a 3}

# Definición de la probabilidad: teoría de conjuntos

Evento A = {1, 3, 5} Evento B = {menor a 3}

A U B (unión de A y B): significa A o B o ambos

A U B = {1, 2, 3, 5}

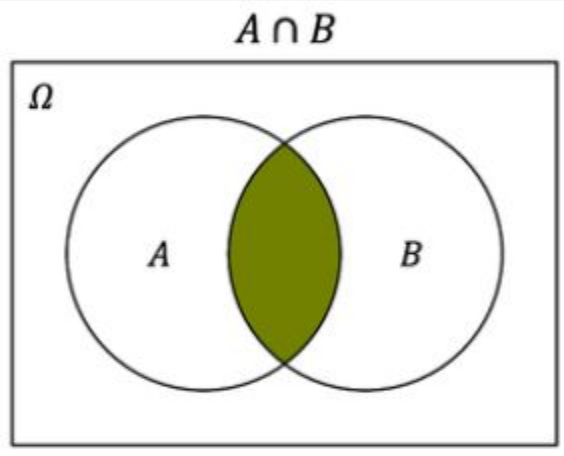


## Definición de la probabilidad: teoría de conjuntos

Evento A = {1, 3, 5} Evento B = {menor a 3}

$A \cap B$ (intersección de A y B): sólo los resultados que ocurren en ambos eventos

$A \cap B = \{1\}$

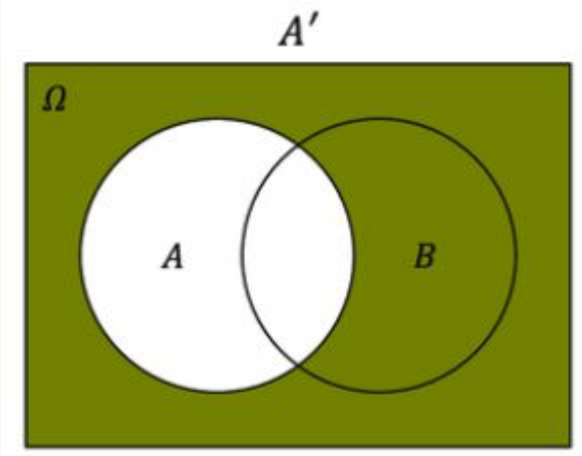


# Definición de la probabilidad: teoría de conjuntos

Evento A = {1, 3, 5} Evento B = {menor a 3}

$A^c$ (Complemento de A): todos los resultados que no están en A

$A^c$ = {2, 4, 6}



# Definición de la probabilidad: teoría de conjuntos

Si $A \cap B = \varnothing$ $\longrightarrow$ A y B son mutuamente excluyentes

Ejemplo: ¿puede un dado salir par e impar a la vez?

# Definición de la probabilidad: teoría de conjuntos

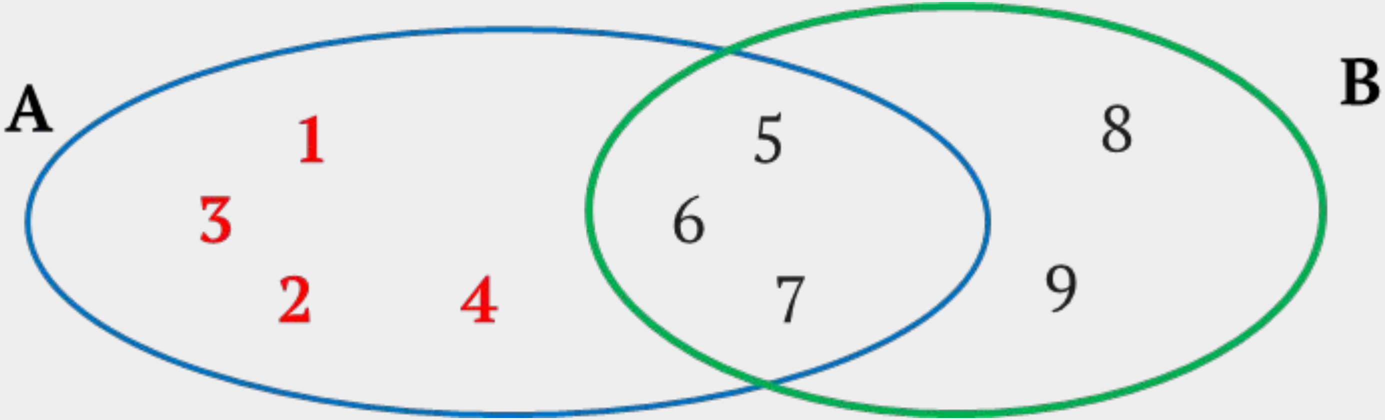
## Diferencia de conjuntos

Dado que:

- A={1, 2, 3, 4, 5, 6, 7}
- B={5, 6, 7, 8, 9}

Entonces:

- A-B= {1, 2, 3, 4}



# Ejemplo de unión e intersección de conjuntos

Una bolsa tiene 10 canicas numeradas del 0 al 9. Describir los sucesos: A="Mayor que 6“; B="No obtener 6“; C="Menor que 6“, definiendo por extensión; hallar los sucesos AUB, A∩B y B'∩A' (complemento).

# Ejemplo de unión e intersección de conjuntos

Una bolsa tiene 10 bolas numeradas del 0 al 9. Describir los sucesos: A="Mayor que 6“; B="No obtener 6“; C="Menor que 6“, definiendo por extensión; hallar los sucesos AUB, A∩B y B'∩A' (complemento).

E={ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 } (espacio muestral)

A={7, 8, 9}; B={0, 1, 2, 3, 4, 5, 7, 8, 9}; C={0, 1, 2, 3, 4, 5}

# Ejemplo de unión e intersección de conjuntos

Una bolsa tiene 10 bolas numeradas del 0 al 9. Describir los sucesos: A="Mayor que 6“; B="No obtener 6“; C="Menor que 6“, definiendo por extensión; hallar los sucesos $A \cup B$, $A \cap B$ y $B' \cap A'$ (complemento).

E={ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 } (espacio muestral)

A={7, 8, 9}; B={0, 1, 2, 3, 4, 5, 7, 8, 9}; C={0, 1, 2, 3, 4, 5}

A U B={0, 1, 2, 3, 4, 5, 7, 8, 9}

$A \cap B$={7, 8, 9} que es =A;   $B' \cap A'$={6} que es =B’

# Definición axiomática de la probabilidad: algunas reglas

- No negatividad: $P(A) \geq 0, \forall A \subseteq \Omega$
- Las probabilidades suman 1: $P(\Omega) = 1$
- Las probabilidades están entre 0 y 1: $0 \leq P(A) \leq 1$
- Probabilidad de un subconjunto: Siendo A y B eventos en $\Omega$ $(A, B \subseteq \Omega)$, si $B \subseteq A \rightarrow P(B) \leq P(A)$
- Regla de la suma para eventos mutuamente excluyentes: Si $P(A) = 0.20$ y $P(B) = 0.35$

  $P(A \text{ o } B) = 0.55$

- Regla de la suma para eventos mutuamente excluyentes:

Si P(A) = 0.20 y P(B) = 0.35 P(A o B) = 0.55

- Si los sucesos no son mutuamente excluyentes, la regla de la adición se formula del siguiente modo:

$$P\ (A \text{ o } B) = P\ (A) + P(B) - P\ (A \text{ y } B)$$

Donde P (A y B) denota la probabilidad de que A y B ocurran al mismo tiempo. -INTERSECCIÓN

## Regla de Suma para eventos no mutuamente excluyentes

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

# Eventos equiprobables (probabilidad clásica)

$$1 = P(\Omega) = P(\text{cara} \cup \text{cruz})$$ -eventos mutuamente exluyentes-

$$= P(\text{cara}) + P(\text{cruz}) = P(\text{cara}) + P(\text{cara})$$

$$= 2 * P(\text{cara})$$

$$= P(\text{cara}) = P(\text{cruz}) = 1/2$$

Si hay 2 resultados posibles → c/u ocurre con probabilidad 1/2

Si hay 6 resultados posibles (ejemplo dado) → c/u ocurre con probabilidad 1/6

# Siguiendo las reglas de la probabilidad...

¿Cuál de las siguientes distribuciones de probabilidad de ingreso familiar es correcta? ¿por qué?

| Income Range | \$0-25k | \$25k-50k | \$50k-100k | \$100k+ |
|---|---|---|---|---|
| (a) | 0.18 | 0.39 | 0.33 | 0.16 |
| (b) | 0.38 | -0.27 | 0.52 | 0.37 |
| (c) | 0.28 | 0.27 | 0.29 | 0.16 |

# Siguiendo las reglas de la probabilidad…

¿Cuál de las siguientes distribuciones de probabilidad de ingreso familiar es correcta? ¿por qué?

| Income Range | \$0-25k | \$25k-50k | \$50k-100k | \$100k+ |
|---|---|---|---|---|
| (a) | 0.18 | 0.39 | 0.33 | 0.16 |
| (b) | 0.38 | -0.27 | 0.52 | 0.37 |
| (c) | 0.28 | 0.27 | 0.29 | 0.16 |

- Cada probabilidad individual está entre 0 y 1 (no hay probabilidades negativas)
- Las probabilidades totales suman 1

## Pasemos a un ejemplo…

¿Qué tan probable es que dos personas, elegidas al azar, compartan el día de cumpleaños?

¿Qué les parece?

**Pasemos a un ejemplo...**

Punto clave: personas elegidas al azar supone considerar todas las combinaciones posibles (de personas que existan)

**Supongamos equiprobabilidad e invirtamos el problema:**

$P(P2 \neq P1) = 364/365$

**¿Y qué ocurre si quiero pensar el problema para estimar la probabilidad de que 3 personas compartan cumpleaños?**

$$P(P2 \neq P1 \cap P3 \neq P1 \text{ y } P2) =$$

$$= 364/365 * 363/365 \ldots.\ 365 - (n-1)/365$$

# Si nos preguntamos por la probabilidad de que en un grupo de 30, al menos 2 de ellas compartan cumpleaños…

P(30 personas cumplan en días distintos) =

= 364/365 * 363/365…. 365 - 29/365

= 0.3

P(30 personas cumplan en días distintos)$^c$ = P(al menos dos personas entre las 30 compartan cumpleaños)

O

1 - P(30 personas cumplan en días distintos) = P(al menos dos personas entre las 30 compartan cumpleaños)

1 - 0.3 = 0.7

La probabilidad es del 70%

# Probabilidad condicional

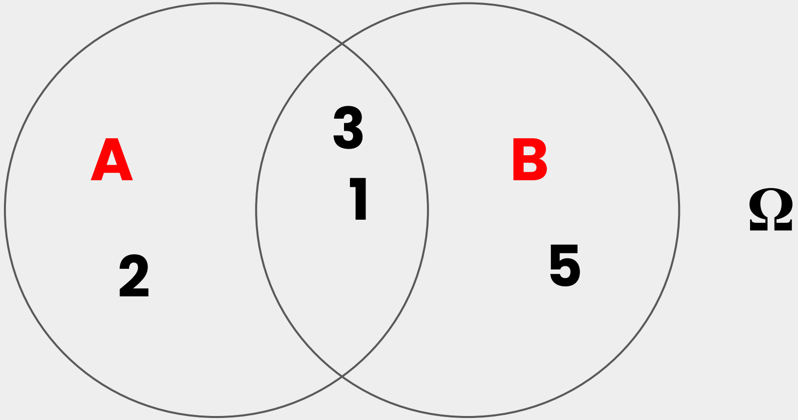
Supongamos que dado $\Omega$, conocemos que ocurrió el evento B

$B = \{1, 3, 5\}$

$A = \{1, 2, 3\}$

¿Cuál es la probabilidad de ocurrencia del evento A?

# Probabilidad condicional

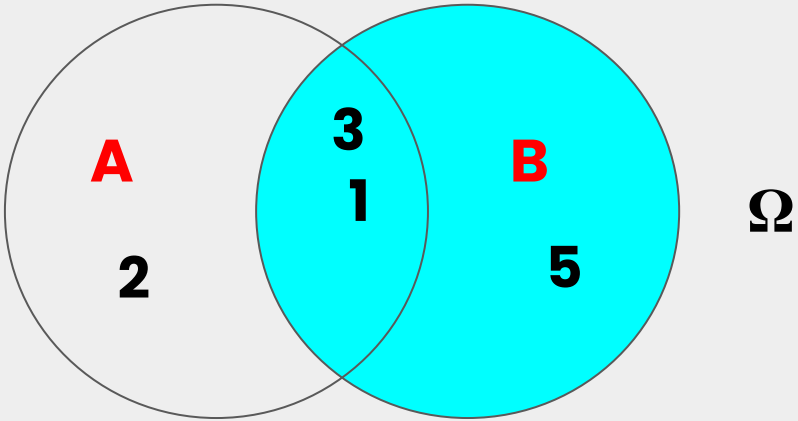



Inicialmente:

P(A) = 3/6 (o 1/2)

P(B) = 3/6 (o 1/2)

# Probabilidad condicional




**Sabiendo que ocurrió B:**

$P(A|B) = ?$

$P(B) = 1/2$

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

$= P\ (A \cap B)\ /\ P(B) = (2/6)\ /\ (3/6) = 2/3$

# Probabilidad condicional: ley de la multiplicación

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

$$\Downarrow$$

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

# Probabilidad condicional: ley de la probabilidad total

- Permite determinar la probabilidad incondicional de un evento a partir de probabilidades condicionales:

Dado un evento A y una partición del espacio muestral en distintas partes: B1, B2,... Bn, la probabilidad total de A es:

$$P(A) = P(A \mid B1) * P(B1) + P(A|B2) * P(B2) + P(A|Bn) * P(Bn)$$

Notemos que los Bi, 1=1..., n conforman la totalidad del espacio muestral y son mutuamente excluyentes (no presentan elementos en común)

# Probabilidad total: ejemplo 1

Aprobación de una política según el partido gobernante

**Contexto:**

- Un país elige el próximo partido gobernante.
- Cada partido tiene la siguiente probabilidad de ganar las elecciones:
  - Partido Verde: 40%
  - Partido Azul: 35%
  - Partido Rojo: 25%

**Probabilidad de aprobación de reforma educativa según partido:**

- Verde: 70%
- Azul: 50%
- Rojo: 20%

**Pregunta:** ¿Cuál es la probabilidad de que la reforma sea aprobada?.

# Probabilidad total: ejemplo 1

Aprobación de una política según el partido gobernante

**Cálculo (Regla de la probabilidad total):**

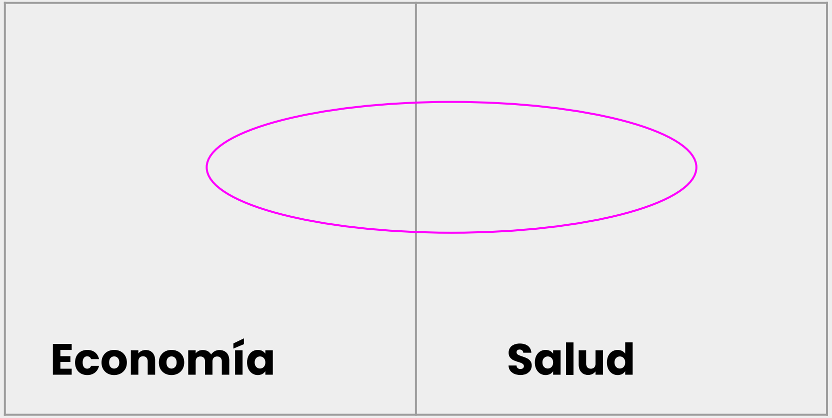
P(Reforma aprobada) = P(Verde)*P(Aprobada|Verde) + P(Azul)*P(Aprobada|Azul) + P(Rojo)*P(Aprobada|Rojo)

P(Reforma aprobada) = 0.4**0.7* + *0.35**0.5 + 0.25*0.2 = 0.28 + 0.175 + 0.05 = 0.505

**Resultado:** 50.5% de probabilidad de aprobación.

# Probabilidad total: ejemplo 2

Probabilidad de ser mujer en distintas áreas de la administración pública

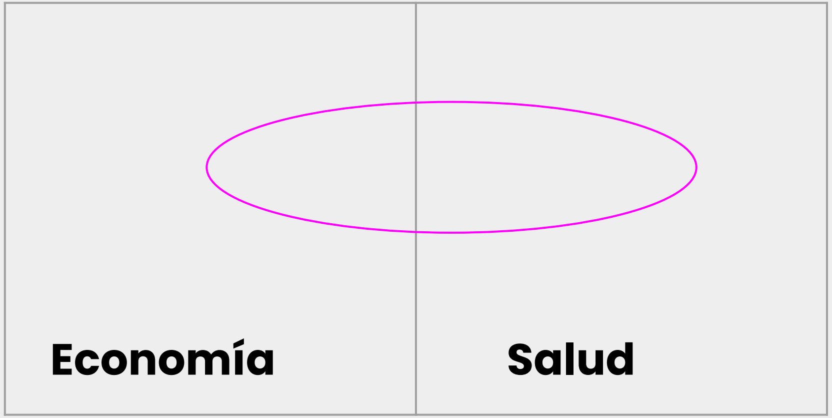



A = ser mujer

Nos preguntamos por la probabilidad del evento A dado que estoy en un “mundo” o el otro

# Probabilidad total: ejemplo 2

Probabilidad de ser mujer en distintas áreas de la administración pública




A = ser mujer

$P(A) = P(A|E) * P(E) + P(A|S) * P(S)$ – **en cada caso se pondera por la probabilidad de que estemos en uno u otro grupo**

**Y esto es igual a:**

$P(A) = P(A \cap E) + P(A \cap S)$ - porque ambos son eventos mutuamente excluyentes

## Un ejemplo interesante en ciencias sociales:
## Aplicación de la ley de probabilidad total para encuestas sensibles

**¿Cómo medir la incidencia de eventos no deseados socialmente?:**

**Por ejemplo:**

-Consumo de drogas

-ETS

-Religión

-Algunas posturas políticas

-Sueldo

## Un ejemplo interesante en ciencias sociales:
## Aplicación de la ley de probabilidad total para encuestas sensibles

Para realizar una encuesta se le pide a los participantes que saquen al azar (y observen en privado) una carta de un mazo.

40% pregunta inocua

60% pregunta de interés

Pregunta inocua: ¿Usted ha nacido en año par?

Pregunta de interés: ¿Alguna vez usted ha consumido drogas ilegales?

Cada encuestado responde **"sí"** o **"no"** pero no revela a qué pregunta contestó ➡ **¿Podemos conocer qué porcentaje de la muestra ha consumido drogas ilegales?**

# Un ejemplo interesante en ciencias sociales: Aplicación de la ley de probabilidad total para encuestas sensibles

| | P. inocua | P. interés |
|---|---|---|
| Sí | | |
| No | | |

$\Omega$

$P(\text{sí}) = P(\text{sí} \mid PI) * P(PI) + P(\text{sí} \mid Pint) * P(Pint)$

$P(\text{sí}) = \mathbf{P(\text{sí} \mid PI)} * 0.4 + P(\text{sí} \mid Pint) * 0.6$

Podemos aproximar la proporción de personas que nacen en año par e impar, y conocer **P(sí|PI) Supongamos que P(sí|PI) = 0.5**

$P(\text{sí}) = 0.5 * 0.4 + P(\text{sí} \mid Pint) * 0.6$

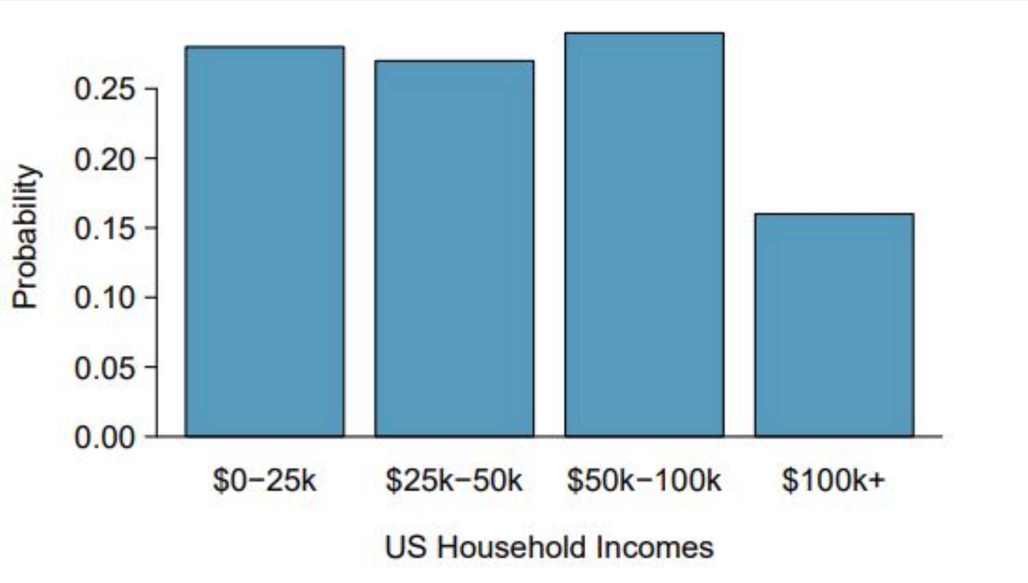
P(sí) = total de participantes que responden sí / sobre total de participantes

$\mathbf{P(\text{sí} \mid Pint)} * 0.6 = P(\text{sí}) - 0.5 * 0.4$

# Graficando distribuciones de probabilidad: un acercamiento a la estadística inferencial

Volvamos al ejemplo de la distribución del ingreso familiar…

Al igual que vimos en estadística descriptiva, las distribuciones de probabilidad también pueden graficarse para obtener información acerca de su comportamiento:



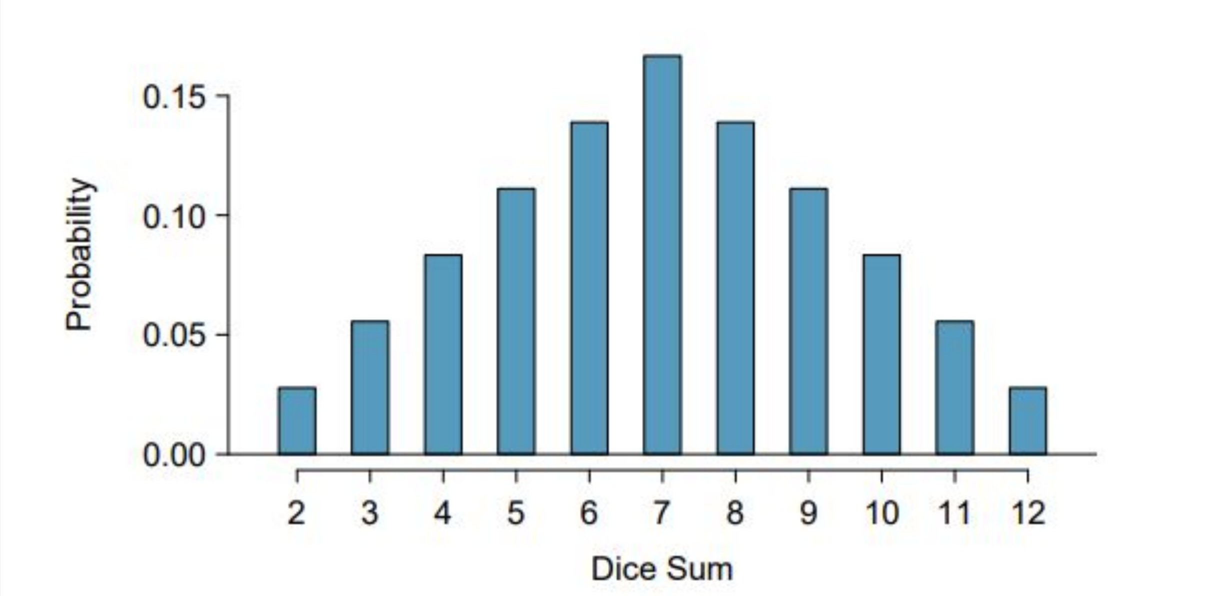

La frecuencia con la que se observa un determinado resultado o en este caso dato/rango de valores, nos da información acerca de la probabilidad de ocurrencia de los distintos niveles de ingreso en la muestra y más importantemente en la población.

# Graficando distribuciones de probabilidad: un acercamiento a la estadística inferencial

**Otro ejemplo: la distribución de probabilidad de la suma de 2 dados**

Se asemeja bastante a una distribución normal, donde los valores centrales ocurren con alta probabilidad y a medida que nos acercamos a los extremos caen las probabilidades de ocurrencia.

La lógica es que al tirar dos dados, su suma puede dar 7 de múltiples maneras (1 y 6, 2 y 5, 4 y 3, etc), la única forma de sumar 2 es obtener 1 y 1 (0,03 chance), y de sumar 12 es obtener 6 y 6 (0,03 chance). Hay 36 combinaciones posibles.



## Eventos independientes

Vimos un ejemplo en el que conocer la probabilidad de ocurrencia de un evento B nos brinda información acerca de la probabilidad de un evento A. Esto supone que A y B no son eventos independientes.

En contraste, 2 eventos son independientes entre sí cuando conocer información sobre la probabilidad de uno no es relevante para estimar la probabilidad del otro. Esto supone que:

$$\mathbf{P(A \mid B) = P(A)}$$

¿Qué ejemplos de eventos independientes se pueden pensar?

# Eventos independientes

Vimos un ejemplo en el que conocer la probabilidad de ocurrencia de un evento B nos brinda información acerca de la probabilidad de un evento A. Esto supone que A y B no son eventos independientes.

En contraste, 2 eventos son independientes entre sí cuando conocer información sobre la probabilidad de uno no es relevante para estimar la probabilidad del otro. Esto supone que:

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

$$P(A|Bc) = P(A)$$

Ejemplo:

A = Mañana va a aumentar el subte

B = Esta clase terminará a las 22 hs

**Podemos suponer que ambos eventos son independientes, en tanto que un aumento en el subte no aumenta ni disminuye las probabilidades de que la clase termine a las 22 hs y viceversa**

## Eventos independientes

Bajo eventos dependientes, partíamos de la siguiente ecuación:

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

Y al pasar **P(B)** multiplicando se obtenía la ley de la multiplicación

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

Con eventos independientes, siendo que $P(A|B) = P(A)$, hacemos un pasaje de términos análogo y obtenemos:

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

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

Si se cumple esta última igualdad entonces se trata de eventos independientes, de lo contrario, se trata de eventos dependientes.

# Eventos independientes

¿Qué sucede con la independencia y los eventos mutuamente excluyentes?

# Eventos (in)dependientes

Veamos un ejemplo en el que conocer la ocurrencia de un evento A nos da información sobre la probabilidad de ocurrencia de un evento B y cómo no se cumple dicha igualdad

**Evento A** = Una persona tiene un título universitario

**Evento B** = La persona gana un salario alto

Suponemos las siguientes probabilidades:

$P(A) = 0.3$

$P(B|A) = 0.6$

$P(B|Ac) = 0.2$

$P(B) = 0.4$

Y queremos ver si se cumple la igualdad de eventos independientes:

$\mathbf{P(A \cap B) = P(A) * P(B)}$

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

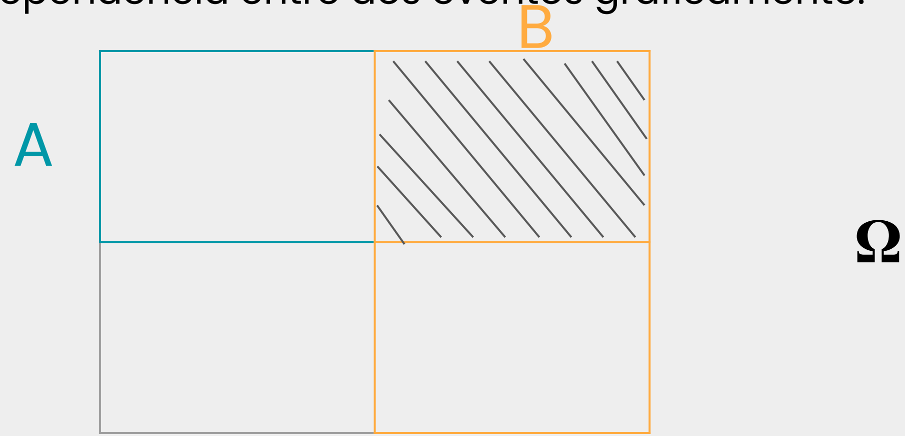
$\mathbf{P(A \cap B)=0.6*0.3 = 0.18}$

$\mathbf{P(A)*P(B) = 0.3*0.4 = 0.12}$

$P(A \cap B) \neq P(A) * P(B)$ □ A y B **no** son eventos independientes

# Eventos independientes

Ahora verifiquemos la independencia entre dos eventos gráficamente:




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

**Y bajo eventos independientes:**

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

Lo que equivale a decir que $P(A)$ tiene que ser a $\Omega$ como $P(A \cap B)$ a $P(B)$:

$$P(A)/\Omega = \frac{1}{2} / 1 = \frac{1}{2}$$

$$P(A \cap B) / P(B) = \frac{1}{4} / \frac{1}{2} = \frac{1}{2}$$

# Teorema de Bayes

## Probabilidad subjetiva:

Ya no hablamos de equiprobabilidad o de probabilidad frecuentista, sino que la probabilidad de un evento se relaciona con las creencias del observador. La probabilidad se actualiza a partir de la obtención de nueva información.

**Aplicaciones:**

-Medicina

-Juicios, criminología

-Economía y finanzas

-Marketing, diseño web (AB testing)

## Teorema de Bayes

Frecuentemente sucede que se desea calcular **$P(A|B)$**, cuando en realidad lo que se conoce es:

$P(B|A)$, o

$P(A)$

$P(Ac)$ …

## Teorema de Bayes: motivación

Ejemplo del uso de tests médicos para detectar una enfermedad rara

Dada la existencia de tests médicos poco específicos (alto % de falsos positivos), no se suele someter a la población a un testeo frecuente de todas las enfermedades. Más bien se utilizan los tests como confirmación a partir de que un paciente presenta síntomas.

**FALSOS POSITIVOS: ¿Cuál es la probabilidad de que una persona esté sana siendo que obtuvo un test positivo de una enfermedad X?**

$$P(A|B)$$

$$P(A) = \text{Sano},\ P(B) = +$$

# Teorema de Bayes: motivación

Ejemplo del uso de tests médicos para detectar una enfermedad rara

**Supuestos:**

- Una enfermedad rara afecta sólo al 1% de la población
- Si una persona está enferma, el test dará positivo con probabilidad 0.98
- Si la persona está sana, el test dará positivo con probabilidad 0.05 (0.99 de probabilidad de estar sano)

**¿Cuál es la probabilidad de que una persona esté sana siendo que obtuvo un test positivo de una enfermedad X? - $P(A|B)$ -**

**$P(A)$ = la persona está sana**

**$P(B)$ = la persona dió +**

**No conocemos $P(A|B)$** ➡ **¿Y conocemos $P(B|A)$?**

# Teorema de Bayes: motivación

Ejemplo del uso de tests médicos para detectar una enfermedad rara

**Supuestos:**

- Una enfermedad rara afecta sólo al 1% de la población
- Si una persona está enferma, el test dará positivo con probabilidad 0.98
- Si la persona está sana, el test dará positivo con probabilidad 0.05
  (0.99 de probabilidad de estar sano)

**¿Cuál es la probabilidad de que una persona esté sana siendo que obtuvo un test positivo de una enfermedad X? - $P(A|B)$ -**

**$P(A)$ = la persona está sana**

**$P(B)$ = la persona dió +**

→ **No conocemos $P(A|B)$**

→ **¿Y conocemos $P(B|A)$?**

**El teorema de bayes es una igualdad matemática resultante de aplicar:**

-Ley de la multiplicación
-Probabilidad conjunta
-Ley de probabilidad total

Repasemos un poco lo que estuvimos viendo para así llegar al teorema de Bayes y poder hallar **$P(A|B)$**...

# Probabilidad condicional

## Ley de la multiplicación

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

$$\Downarrow$$

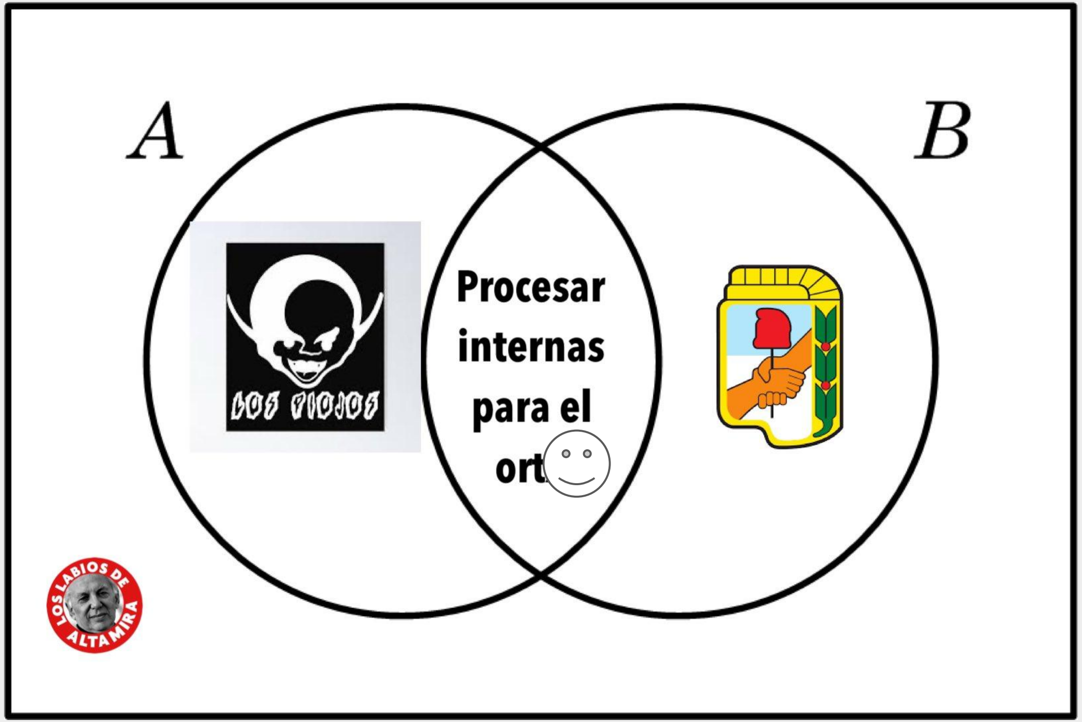
$$P(A \cap B) = P(A|B) * P(B)$$

## Simetría de la probabilidad conjunta

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

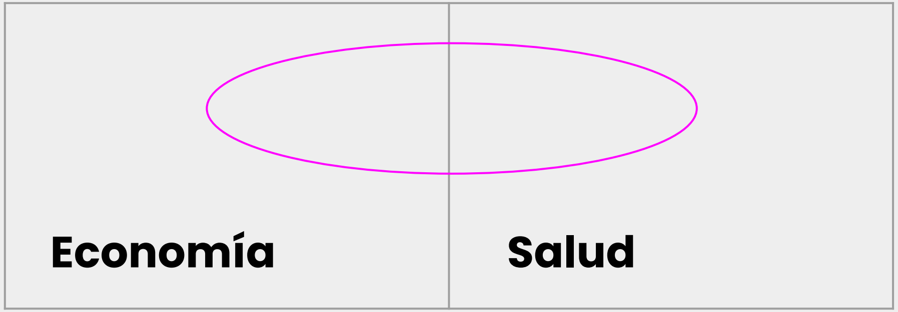
¿Recuerdan lo que era la intersección entre 2 eventos?

# ¿Recuerdan lo que era la intersección entre 2 eventos?



# Probabilidad condicional

## Ley de la probabilidad total




- Permite determinar la probabilidad incondicional de un evento a partir de probabilidades condicionales:

Dado un evento A y una partición del espacio muestral en distintas partes: B1, B2,... Bn, la probabilidad total de A es:

$$P(A) = P(A \mid B1) * P(B1) + P(A \mid B2) * P(B2) + P(A \mid Bn) * P(Bn)$$

Notemos que los Bi, 1=1..., n conforman la totalidad del espacio muestral y son mutuamente excluyentes (no presentan elementos en común)

# TEOREMA DE BAYES

Invertimos las probabilidades conjuntas (intersección) y usamos ley de la multiplicación en el numerador

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

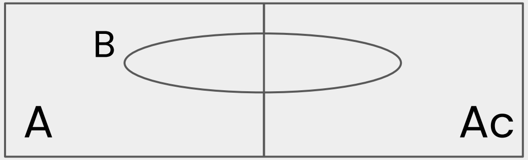
# TEOREMA DE BAYES

Invertimos las probabilidades conjuntas (intersección) y usamos ley de la multiplicación en el numerador

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

Ley de probabilidad total en el denominador

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




��

# TEOREMA DE BAYES: aplicación al estudio de los falsos positivos

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

**Datos:**

- Una enfermedad rara afecta sólo al 1% de la población
- Si una persona está enferma, el test dará positivo con probabilidad 0.98
- Si la persona está sana, el test dará positivo con probabilidad 0.05 (0.99 de probabilidad de estar sano)

**Definición de los eventos:**

$P(Sana) = 0.99 \longrightarrow P(A) = 0.99$

$P(Sanac) = 0.01 \longrightarrow P(Ac) = 0.01$

$P(+|Sana) = 0.05 \longrightarrow P(B|A) = 0.05$

$P(+|Sanac) = 0.98 \longrightarrow P(B|Ac) = 0.98$

# TEOREMA DE BAYES: aplicación al estudio de los falsos positivos

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

**Definición de los eventos:**

$P(Sana) = 0.99 \longrightarrow P(A) = 0.99$

$P(Sanac) = 0.01 \longrightarrow P(Ac) = 0.01$

$P(+|Sana) = 0.05 \longrightarrow P(B|A) = 0.05$

$P(+|Sanac) = 0.98 \longrightarrow P(B|Ac) = 0.98$

$$P(sana|+) = \frac{0.05 * 0.99}{0.05* 0.99 + 0.98 * 0.01} = 0.83$$

**Cuando un test da positivo, lo más probable es que la persona esté sana**

## TEOREMA DE BAYES: aplicación al estudio de los falsos positivos

¿Qué sucede si no se trata de una enfermedad rara?
Si se tratara de una enfermedad que afecta al 50% de la población (una gripe por ejemplo)...

$$P(sana|+) = \frac{0.05 * 0.5}{0.05 * 0.5 + 0.98 * 0.5}$$

$$P(sana|+) = \frac{0.05 * \cancel{0.5}}{\cancel{0.5} * (0.05 + 0.98)} = 0.048$$

**En este caso cuando el test da positivo lo más probable es que la persona esté enferma**

# TEOREMA DE BAYES: aplicación en juicios

¿De qué forma se podría aplicar el teorema de Bayes a la hora de resolver un caso judicial?

# **TEOREMA DE BAYES:** aplicación en juicios

- El uso de este teorema se relaciona con el estudio de la probabilidad de observar ciertas conductas o comportamientos sospechosos, dada la inocencia de una persona en un caso judicial

Estar cerca de la escena del crimen, portar ropa de la víctima, haber estado en contacto con ella poco antes (algunos ejemplos) $\longrightarrow$ **A** = tener comportamientos "sospechosos", tomar decisiones "dudosas"

Presunción de inocencia hasta que se demuestre lo contrario $\longrightarrow$ **B** = ser inocente

Al observar **$P(A|B)$** nos estamos preguntando por **¿qué tan probable es observar conductas sospechosas de una persona inocente?** Y el objetivo sería rechazar la H0 de que una persona es inocente únicamente si existe suficiente evidencia en su contra, es decir se busca que:
$P(A|B) \approx 0$ al condenar a una persona. (De lo contrario $\longrightarrow$ error de tipo I)

## El uso de la probabilidad y estadística en juicios: el caso de Sally Clarke

- En el año 1999 Sally Clarke, una abogada de Cheshire, Inglaterra fue acusada de haber asesinado a sus dos hijos. El primero de ellos murió a las 8 semanas de edad (1996), y el segundo a las 11 semanas (1998).

- Tres años después fue puesta en libertad. La sentencia de 1999 se considera uno de los mayores errores judiciales de la historia de Inglaterra.

El error se debió a lo que se conoce como prosecutor fallacy (el **error probabilístico** de hallar a alguien culpable a raíz de un falso razonamiento científico).
Veamos este error más de cerca...

# El uso de la probabilidad y estadística en juicios: el caso de Sally Clarke

El principal problema del razonamiento que dio lugar a la sentencia tiene que ver con haber considerado a los dos eventos como **independientes**. Esto llevó al siguiente cálculo por parte de un pediatra inglés:

$$P(A \cap B) = P(A) * P(B)$$
$$P(A \cap B) = 1/8543 * 1/8543 \approx 1 / 73.000.000$$

Por aquel entonces, nacían alrededor de 700.000 bebés por año, por lo que se trataría de un suceso con ocurrencia cada 100 años.

# El uso de la probabilidad y estadística en juicios: el caso de Sally Clarke

En este caso se debería haber utilizado el Teorema de Bayes, que da lugar a la ocurrencia de un evento altamente improbable al compararlo con otros también improbables eventos

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

P(H) = los dos hijos murieron de muerte súbita
P(H) = 1/1300 * 1/100 = 0.0000077
P(Hc) = 1 - P(H) = 0.9999923. Prob de que mueran de otra causa distinta
P(D) = ambos hijos murieron de forma repentina y sorpresiva
P(D|H) = 1
P(D|Hc) = 30/650.000*1/10 = 0.0000046 ⟶ la prosecución equipara esto a un crimen

Queremos estimar P(H|D) = probabilidad de que la causa haya sido muerte súbita dado lo inesperado y repentino de la misma

# El uso de la probabilidad y estadística en juicios: el caso de Sally Clarke

En este caso se debería haber utilizado el Teorema de Bayes, que da lugar a la ocurrencia de un evento altamente improbable al compararlo con otros también improbables eventos

$$P(H|D) = \frac{P(D|H) * P(H)}{P(D|H) * P(H) + P(D|Hc) * P(Hc)}$$

$P(H)$ = los dos hijos murieron de muerte súbita
$P(H) = 1/1300 * 1/100 = 0.0000077$
$P(Hc)$ = mueren de otra causa distinta
$P(D)$ = ambos hijos murieron de forma repentina y sorpresiva
$P(D|H) = 1$
$P(D|Hc) = 30/650.000*1/10 = 0.0000046$ → la prosecución equipara esto a un crimen

Queremos estimar $P(H|D)$ = probabilidad de que la causa haya sido muerte súbita dado lo inesperado y repentino de la misma

# El uso de la probabilidad y estadística en juicios: el caso de Sally Clarke

En este caso se debería haber utilizado el Teorema de Bayes, que da lugar a la ocurrencia de un evento altamente improbable al compararlo con otros también improbables eventos

$$P(H|D) = \frac{1*0.0000077}{1*0.0000077 + 0.0000046 * 0.999992} \approx 0.6$$

$P(H)$ = los dos hijos murieron de muerte súbita
$P(H) = 1/1300 * 1/100 = 0.0000077$
$P(Hc)$ = mueren de otra causa distinta
$P(D)$ = ambos hijos murieron de forma repentina y sorpresiva
$P(D|H) = 1$
$P(D|Hc) = 30/650.000*1/10 = 0.0000046$ $\longrightarrow$ la prosecución equipara esto a un crimen

Queremos estimar $P(H|D)$ = probabilidad de que la causa haya sido muerte súbita dado lo inesperado y repentino de la misma

# Variables Aleatorias - Elementos

1. ~~**Definición de Variable Aleatoria**~~
2. ~~**Variable Aleatoria Discreta (VAD)**~~
3. ~~**Distribución de Probabilidad de las VAD**~~
4. ~~**Función de Distribución Acumulada de las VAD**~~
5. ~~**Media y Varianza**~~
6. ~~**Variable Aleatoria Continua (VAC) (función de densidad, acumulado)**~~
7. ~~**Media y Varianza**~~
8. ~~**Transformación de una Variable Aleatoria: la estandarización**~~

# Variables Aleatorias - Definición y sus elementos

Una **variable aleatoria** es una **función** que asigna un **valor numérico a cada resultado** de un experimento aleatorio. Esto permite modelar la incertidumbre al mapear cada resultado a un número real definido por una función X cuyo valor dependerá de los resultados del experimento

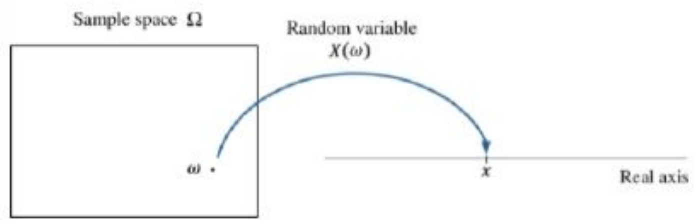


Figura 13.3 From the sample space Ω to the real axis.

El resultado de un experimento aleatorio **no puede saberse con certeza antes de hacerlo**.

Lo que sí podemos calcular es la **probabilidad** de que ocurran los diferentes valores.

## Repaso de los axiomas de Kolmogórov

- No negatividad: $P(A) \geq 0, \forall\ A \subseteq \Omega$
- Las probabilidades suman 1: $P(\Omega) = 1$
- Las probabilidades están entre 0 y 1: $0 \leq P(A) \leq 1$
- Probabilidad de un subconjunto: Siendo A y B eventos en $\Omega$ $(A, B \subseteq \Omega)$, si $B \subseteq A \rightarrow P(B) \leq P(A)$
- Regla de la suma para eventos mutuamente excluyentes: Si $P(A) = 0.20$ y $P(B) = 0.35$

  $P(A \text{ o } B) = 0.55$

# Ejemplos:

## Ejemplo de variable no aleatoria

**Variable:** *Edad del votante en un padrón electoral*

- Si tenemos acceso al padrón, la edad de cada persona **ya está determinada y es conocida**.
- No depende de un experimento ni del azar: no hay incertidumbre sobre el valor de esa variable.
- Por lo tanto, **no es una variable aleatoria**, sino una variable fija (determinística).

---

## Ejemplo de variable aleatoria

**Variable:** *Edad del votante en un distrito*

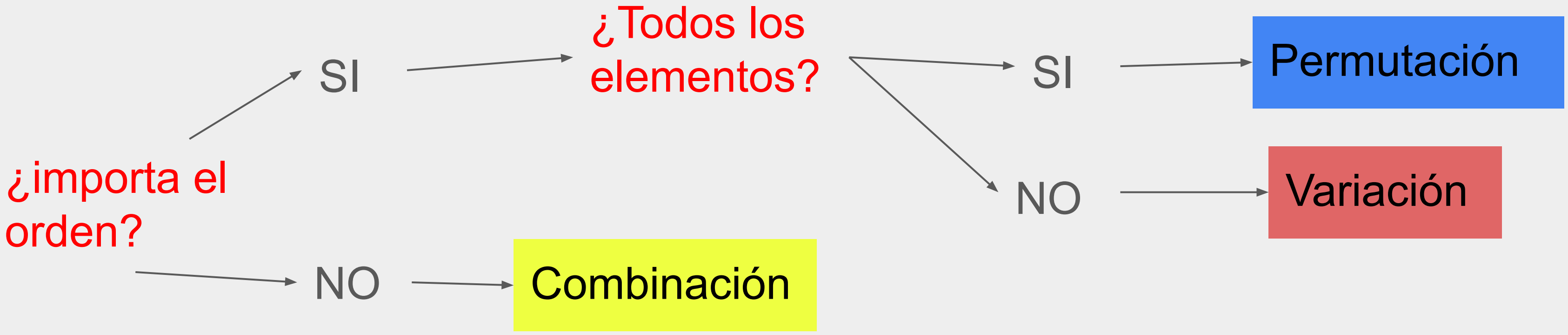
Si en lugar del padrón completo hacemos una **encuesta a una muestra de votantes**, entonces:

- Para el investigador, antes de preguntar, la edad del encuestado **es incierta**.
- En ese contexto, la edad **sí se modela como una variable aleatoria**, porque su valor depende del proceso de selección muestral (el azar de quién resulta encuestado).

# Por qué es importante tener una noción de combinatoria

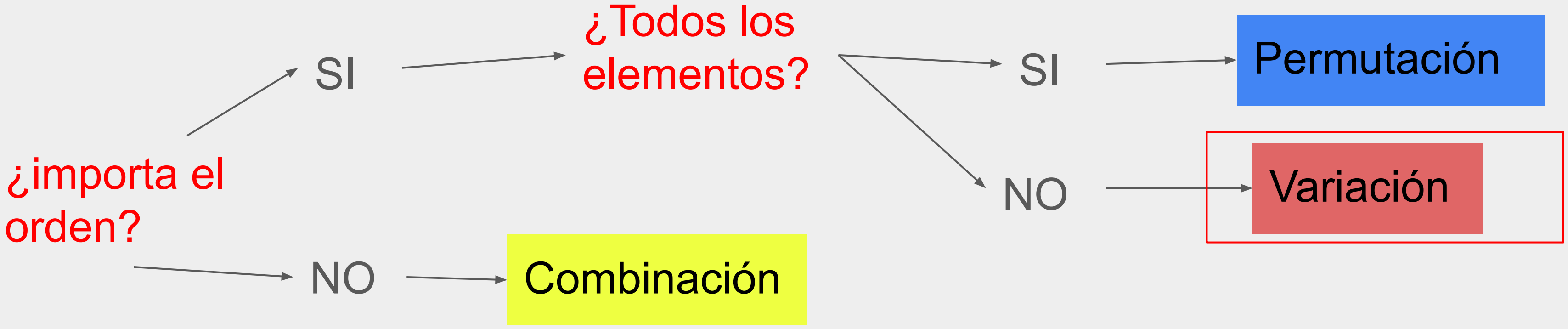
- Cálculo de probabilidades: Para determinar probabilidades, a menudo es necesario contar el número de resultados posibles. La combinatoria proporciona las herramientas para calcular de manera eficiente cuántos resultados diferentes se pueden obtener en un experimento (por ejemplo, lanzar dados, seleccionar cartas o formar grupos).
- Eventos complejos: Muchos problemas de probabilidad involucran eventos complejos donde se deben considerar diferentes formas de combinar elementos (por ejemplo, en cuántas maneras se pueden elegir equipos de un conjunto mayor o formar combinaciones ganadoras). La combinatoria te ayuda a descomponer estos problemas.
- Distribuciones probabilísticas: Algunas distribuciones de probabilidad, como la binomial, la hipergeométrica o la de Poisson, están directamente relacionadas con conceptos combinatorios. Entender cómo funcionan estos modelos implica conocer cómo se cuentan combinaciones y permutaciones.
- Espacio muestral: La combinatoria es útil para definir el espacio muestral de un experimento aleatorio, especialmente cuando hay muchas posibilidades. Esto es crucial para aplicar métodos estadísticos que se basan en todas las posibles combinaciones de eventos.

# Como saber de qué problema hablamos



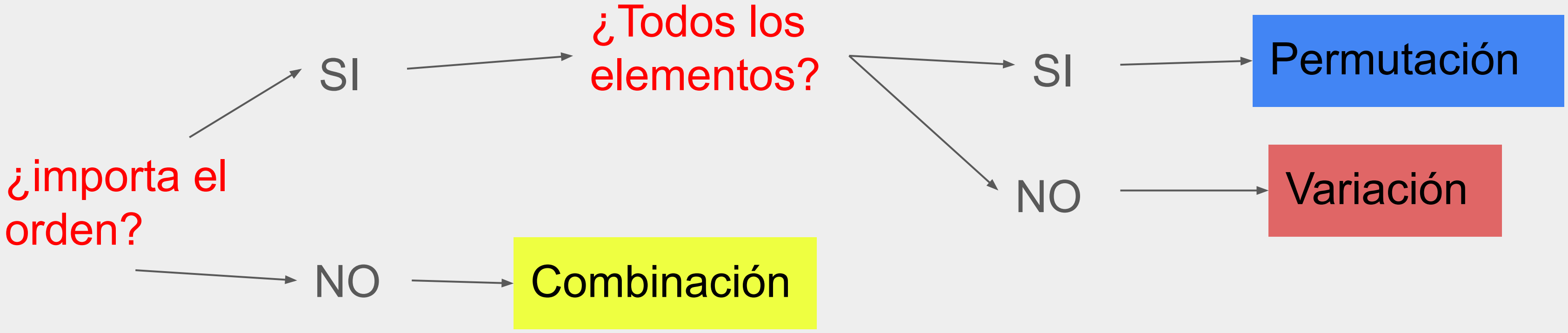

Un equipo de campaña tiene 10 miembros. Se necesitan elegir 3 para recorrer barrios de distinto tamaño (uno grande, uno mediano y uno pequeño).
¿De cuántas formas se pueden seleccionar y asignar esos miembros del equipo?

# Como saber de qué problema hablamos



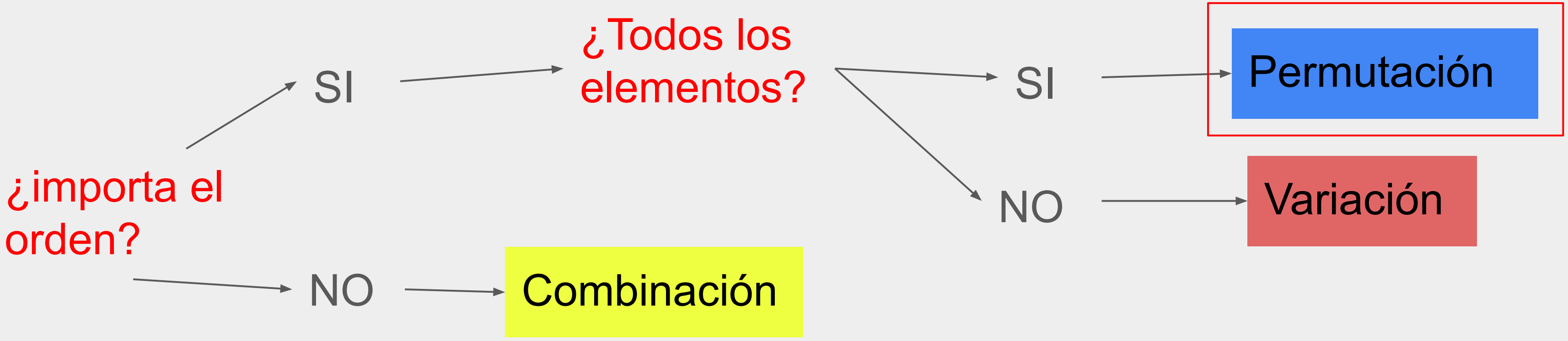

Un equipo de campaña tiene 10 miembros. Se necesitan elegir 3 para recorrer barrios de distinto tamaño (uno grande, uno mediano y uno pequeño).
¿De cuántas formas se pueden seleccionar y asignar esos miembros del equipo?

# Como saber de qué problema hablamos



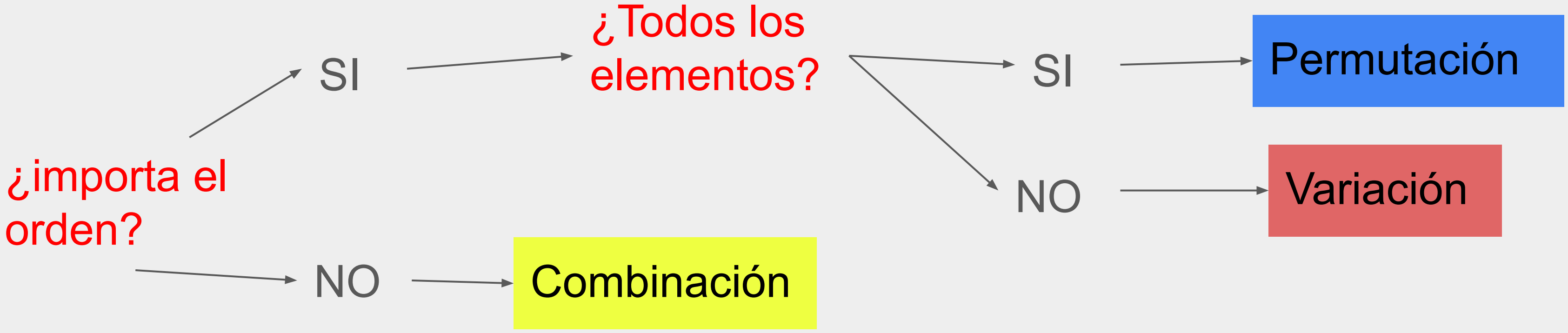

En un debate presidencial participan 4 candidatos.
¿De cuántas formas distintas pueden ordenarse en el escenario?

# Como saber de qué problema hablamos



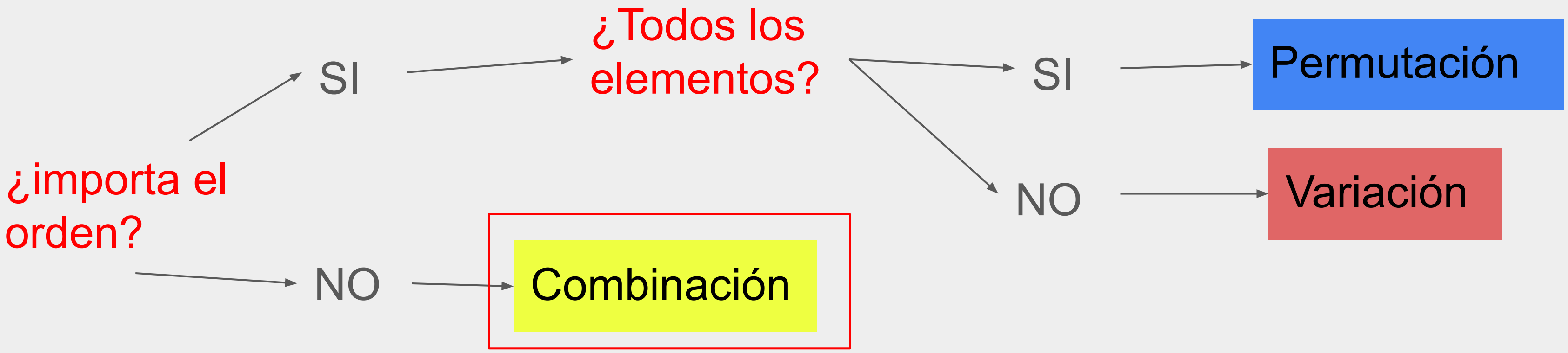

En un debate presidencial participan 4 candidatos.
¿De cuántas formas distintas pueden ordenarse en el escenario?

# Como saber de qué problema hablamos




En una comisión parlamentaria hay 8 legisladores.
¿De cuántas formas se pueden formar subcomisiones de 4 miembros (sin importar el orden)?

# Como saber de qué problema hablamos




En una comisión parlamentaria hay 8 legisladores.
¿De cuántas formas se pueden formar subcomisiones de 4 miembros (sin importar el orden)?

# Fórmulas para ejercicios **sin repetición**

**Importa el orden**

$$ {}_{n}V_{r} = \frac{n!}{(n-r)!} \quad \text{donde } r \leq n $$

$$ Pn = n! $$

**NO Importa el orden**

$$ {}_{n}C_{r} = \frac{n!}{r!(n-r)!} $$

## Importa el orden

$$_nV_r = \frac{n!}{(n-r)!} \quad \text{donde } r \leq n$$

$$P_n = n!$$

## NO Importa el orden

$$_nC_r = \frac{n!}{r!(n-r)!}$$

1. En una comisión parlamentaria hay 8 legisladores. ¿De cuántas formas se pueden formar **subcomisiones de 4** miembros

**Importa el orden**

$$_nV_r = \frac{n!}{(n-r)!} \quad \text{donde } r \leq n$$

$$P_n = n!$$

**NO Importa el orden**

$$_nC_r = \frac{n!}{r!(n-r)!}$$

**Combinación**

1. En una comisión parlamentaria hay 8 legisladores. ¿De cuántas formas se pueden formar **subcomisiones de 4** miembros

$n = 8 \qquad r = 4$

$$_8C_4 = \frac{8!}{4!\,.\,(8-4)!} = \frac{8!}{4!\,.\,4!}$$

Combinación

**Importa el orden**

$$ {}_{n}V_{r} = \frac{n!}{(n-r)!} \quad \text{donde } r \leq n $$

$$ P_n = n! $$

**NO Importa el orden**

$$ {}_{n}C_{r} = \frac{n!}{r!(n-r)!} $$

1. En una comisión parlamentaria hay 8 legisladores. ¿De cuántas formas se pueden formar **subcomisiones de 4** miembros

$n = 8 \qquad r = 4$

$$ {}_{8}C_{4} = \frac{8.\ 7.\ 6.\ 5.\ 4!}{(4.3.2).4!} = $$

**Importa el orden**

$${}_{n}V_{r} = \frac{n!}{(n-r)!} \quad \text{donde } r \leq n$$

$$P_n = n!$$

**NO Importa el orden**

$${}_{n}C_{r} = \frac{n!}{r!(n-r)!}$$

Combinación

1. En una comisión parlamentaria hay 8 legisladores. ¿De cuántas formas se pueden formar **subcomisiones de 4** miembros

$n = 8 \qquad r = 4$

$${}_{8}C_{4} = \frac{\cancel{8}^{\cancel{2}\,1} \cdot 7 \cdot \cancel{6}^{2} \cdot 5 \cdot \cancel{4!}}{(\cancel{4}_{1} \cdot \cancel{3}_{1} \cdot \cancel{2}_{1}) \cdot \cancel{4!}} = \frac{7 \cdot 2 \cdot 5}{1} = \mathbf{70}$$

## Importa el orden

$${}_nV_r = \frac{n!}{(n-r)!} \quad \text{donde } r \leq n$$

$$P_n = n!$$

## NO Importa el orden

$${}_nC_r = \frac{n!}{r!(n-r)!}$$

1. En un debate presidencial participan 4 candidatos. ¿De cuántas formas distintas pueden **ordenarse** en el escenario?

**Importa el orden**

$$_nV_r = \frac{n!}{(n-r)!} \quad \text{donde } r \leq n$$

$$P_n = n!$$

**NO Importa el orden**

$$_nC_r = \frac{n!}{r!(n-r)!}$$

## Permutación

1. En un debate presidencial participan 4 candidatos. ¿De cuántas formas distintas pueden **ordenarse** en el escenario?

$n = 4$

$P_4 = 4!$

## Importa el orden

$${}_{n}V_{r} = \frac{n!}{(n-r)!} \quad \text{donde } r \leq n$$

$$P_n = n!$$

## NO Importa el orden

$${}_{n}C_{r} = \frac{n!}{r!(n-r)!}$$

Permutación

- En un debate presidencial participan 4 candidatos. ¿De cuántas formas distintas pueden **ordenarse** en el escenario?

$n = \quad 4$

$P_4 \quad = \quad 4.\ 3.\ 2 \quad = \quad 24$

## Importa el orden

$${}_{n}V_{r} = \frac{n!}{(n-r)!} \quad \text{donde } r \leq n$$

$$P_n = n!$$

## NO Importa el orden

$${}_{n}C_{r} = \frac{n!}{r!(n-r)!}$$

- Un equipo de campaña tiene 10 miembros. Se necesitan elegir 3 para recorrer **barrios de distinto tamaño** (grande, mediano y pequeño). ¿De cuántas formas se pueden **seleccionar y asignar** esos miembros?

## Importa el orden

$${}_{n}V_{r} = \frac{n!}{(n-r)!} \quad \text{donde } r \leq n$$

$$P_n = n!$$

## NO Importa el orden

$${}_{n}C_{r} = \frac{n!}{r!(n-r)!}$$

Variación

- Un equipo de campaña tiene 10 miembros. Se necesitan elegir 3 para recorrer **barrios de distinto tamaño** (grande, mediano y pequeño). ¿De cuántas formas se pueden **seleccionar y asignar** esos miembros?

n = 10 r = 3

$${}_{10}V_{3} = \frac{10!}{(10-3)!} = \frac{10!}{7!}$$

**Variación**

## Importa el orden

$${}_nV_r = \frac{n!}{(n-r)!} \quad \text{donde } r \leq n$$

$$P_n = n!$$

## NO Importa el orden

$${}_nC_r = \frac{n!}{r!(n-r)!}$$

- Un equipo de campaña tiene 10 miembros. Se necesitan elegir 3 para recorrer **barrios de distinto tamaño** (grande, mediano y pequeño). ¿De cuántas formas se pueden **seleccionar y asignar** esos miembros?

$${}_{10}V_3 = \frac{10.\ 9.\ 8.\ 7!}{7!}$$

## Importa el orden

$$_{n}V_{r} = \frac{n!}{(n-r)!} \quad \text{donde } r \leq n$$

$$P_n = n!$$

## NO Importa el orden

$$_{n}C_{r} = \frac{n!}{r!(n-r)!}$$

Variación

- Un equipo de campaña tiene 10 miembros. Se necesitan elegir 3 para recorrer **barrios de distinto tamaño** (grande, mediano y pequeño). ¿De cuántas formas se pueden **seleccionar y asignar** esos miembros?

$$_{10}V_{3} = \frac{10.\ 9.\ 8.\ \cancel{7!}}{\cancel{7!}} = 720$$

## Importa el orden

$${}_nV_r = \frac{n!}{(n-r)!} \quad \text{donde } r \leq n$$

$$P_n = n!$$

## NO Importa el orden

$${}_nC_r = \frac{n!}{r!(n-r)!}$$

- En una comisión parlamentaria hay 8 legisladores. ¿De cuántas formas se pueden formar **subcomisiones de 4** miembros

**Importa el orden**

$${}_{n}V_{r} = \frac{n!}{(n-r)!} \quad \text{donde } r \leq n$$

$$P_n = n!$$

**NO Importa el orden**

$${}_{n}C_{r} = \frac{n!}{r!(n-r)!}$$

Combinación

- En una comisión parlamentaria hay 8 legisladores. ¿De cuántas formas se pueden formar **subcomisiones de 4** miembros

$n = 8 \qquad r = 4$

$${}_{8}C_{4} = \frac{8!}{4!\,.\,(8-4)!} = \frac{8!}{4!\,.\,4!}$$

**Importa el orden**

Combinación

$$_nV_r = \frac{n!}{(n-r)!} \quad \text{donde } r \leq n$$

$$P_n = n!$$

**NO Importa el orden**

$$_nC_r = \frac{n!}{r!(n-r)!}$$

- En una comisión parlamentaria hay 8 legisladores. ¿De cuántas formas se pueden formar **subcomisiones de 4** miembros

$n = 8$ $r = 4$

$$_8C_4 = \frac{8.\,7.\,6.\,5.\,4!}{(4.3.2).4!} =$$

**Combinación**

## Importa el orden

$$_{n}V_{r} = \frac{n!}{(n-r)!} \quad \text{donde } r \leq n$$

$$P_n = n!$$

## NO Importa el orden

$$_{n}C_{r} = \frac{n!}{r!(n-r)!}$$

- En una comisión parlamentaria hay 8 legisladores. ¿De cuántas formas se pueden formar **subcomisiones de 4** miembros

$$n = 8 \qquad r = 4$$

$$_{8}C_{4} = \frac{\overset{\overset{1}{\cancel{2}}}{\cancel{8}} \cdot 7 \cdot \overset{2}{\cancel{6}} \cdot 5 \cdot \cancel{4!}}{(\underset{1}{\cancel{4}} \cdot \underset{1}{\cancel{3}} \cdot \underset{1}{\cancel{2}}) \cdot \cancel{4!}} = \frac{7 \cdot 2 \cdot 5}{1} = \mathbf{70}$$

## ¿Que tiene de distinto este problema?

- **En un fuero federal hay 12 juzgados. Se registran 10 causas nuevas.** Queremos contar **de cuántas formas distintas** pueden **asignarse** las 12 juzgados a las 10 causas por azar.

## ¿ como se resuelve?

# Un ejemplo de combinatoria con repetición

- **En un fuero federal hay 12 juzgados. Se registran 10 causas nuevas.** Queremos contar **de cuántas formas distintas** pueden **asignarse** las 12 juzgados a las 10 causas por azar.

## ¿ como se resuelve?

$$CR_{n,r} = \binom{n+r-1}{r} = \frac{(n+r-1)!}{r!\,(n-1)!}$$

$$\cancel{{}_nC_r = \frac{n!}{r\,!(n\text{-}r)!}}$$

$$CR_{12,10} = \binom{12+10-1}{10} = \binom{21}{10} = \frac{21!}{10!\,11!}$$

$$\binom{21}{10} = \frac{21\cdot 20\cdot 19\cdot 18\cdot 17\cdot 16\cdot 15\cdot 14\cdot 13\cdot 12}{10\cdot 9\cdot 8\cdot 7\cdot 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1} = \mathbf{352{,}716}$$

# Un ejemplo de combinatoria con repetición

- **En un fuero federal hay 12 juzgados. Se registran 10 causas nuevas.** Queremos contar **de cuántas formas distintas** pueden **asignarse** las 12 juzgados a las 10 causas por azar.

$$CR_{n,r} = \binom{n+r-1}{r} = \frac{(n+r-1)!}{r!\,(n-1)!}$$

$$CR_{12,10} = \binom{12+10-1}{10} = \binom{21}{10} = \frac{21!}{10!\,11!}$$

$$\binom{21}{10} = \frac{21 \cdot 20 \cdot 19 \cdot 18 \cdot 17 \cdot 16 \cdot 15 \cdot 14 \cdot 13 \cdot 12}{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = \mathbf{352{,}716}$$

**¿Y QUE TIENE QUE VER ESTO CON LA PROBABILIDAD?**

1

# Aplicación en probabilidad

- **En un fuero federal hay 12 juzgados. Se registran 10 causas nuevas.** Queremos contar **de cuántas formas distintas** pueden **asignarse** las 12 juzgados a las 10 causas por azar. Rta: **352,716**

  **¿Cual es la probabilidad de que 9 de ellas sean asignadas a un mismo juzgado cualquiera? (suponiendo equiprobabilidad de las secuencias)**

# Aplicación en probabilidad

- **En un fuero federal hay 12 juzgados. Se registran 10 causas nuevas.** Queremos contar **de cuántas formas distintas** pueden **asignarse** las 12 juzgados a las 10 causas por azar. Rta: **352,716**

**¿Cual es la probabilidad de que 9 de ellas sean asignadas a un mismo juzgado cualquiera? (suponiendo equiprobabilidad de las secuencias)**

## Modelo de composiciones equiprobables

**Casos favorables (composiciones):**

- Elegir el juzgado que recibe **9** causas: $12$
- Elegir el juzgado (distinto) que recibe **1** causa: $11$
- Total composiciones favorables: $12 \times 11 = \mathbf{132}$

**Probabilidad en el modelo de composiciones uniformes:**

$$P_{\text{comp}}(\text{“9 y 1”}) = \frac{132}{352{,}716} \approx \mathbf{0.000374} = \mathbf{0.0374\%}$$

**Supone que 1 causa en cada juzgado tiene misma probabilidad que 5 causas en 1 y 5 en otro por dar un ejemplo**

**1 vez cada 2.673 sorteos**

# Aplicación en probabilidad

**¿Cual es la probabilidad de que 9 de ellas sean asignadas a un mismo juzgado cualquiera?** en un sorteo real

# Aplicación en probabilidad

**¿Cual es la probabilidad de que 9 de ellas sean asignadas a un mismo juzgado cualquiera?** en un sorteo real

## Modelo iid (independent and identical distribuited) para sorteo

**Espacio muestral (todas las secuencias):**

- Cada una de las 10 causas se asigna a 1 de 12 juzgados, **independiente** →

$$N_{\text{tot}} = 12^{10} = 61{,}917{,}364{,}224$$

**Casos favorables ("9 y 1", sin fijar juzgado):**

- Elegir el juzgado que recibe **9** causas: $12$
- Elegir el (distinto) que recibe **1** causa: $11$
- **Composiciones favorables:** $12 \times 11 = 132$
- Para cada composición, elegir **cuál** de las 10 causas "se escapa": $\binom{10}{9} = 10$
- **Secuencias favorables:** $132 \times 10 = \mathbf{1{,}320}$

**Probabilidad (modelo *iid*):**

$$P_{\text{iid}}(\text{“9 y 1”, sin juzgado fijo}) = \frac{1{,}320}{12^{10}} \approx 2.1319 \times 10^{-8} = \mathbf{0.00000213\%}$$

1 vez cada 47 millones de sorteos

3

# Aplicación en probabilidad

**¿Cual es la probabilidad de que 9 de ellas sean asignadas a un juzgado en particular ? en un sorteo real**

3

# Aplicación en probabilidad

**¿Cual es la probabilidad de que 9 de ellas sean asignadas a un juzgado en particular ? en un sorteo real**

## Modelo iid de Paenza

**Paso 3 — Secuencias *iid* (modelo de Paenza, juzgado fijo)**

**Espacio muestral (todas las secuencias):**

- 10 causas, cada una a 1 de 12 juzgados, **independiente** →

$$N_{\text{tot}} = 12^{10} = 61{,}917{,}364{,}224$$

**Casos favorables ("9 y 1", con juzgado fijo):**

- Elegir **cuál** de las 10 causas **no** cae en el juzgado fijo: $\binom{10}{1} = 10$
- Elegir **a cuál** de los **11** juzgados restantes va esa causa: $11$
- **Secuencias favorables:** $10 \times 11 = \mathbf{110}$

**Probabilidad (modelo *iid*, Paenza):**

$$P_{\text{iid}}(\text{"9 y 1", juzgado fijo}) = \frac{110}{12^{10}} \approx 1.7766 \times 10^{-9} = \mathbf{0.000000177\%}$$

**Lectura:** ~ 2 en mil millones (≈ 1 cada 563 millones).

**1 vez cada 563 millones de sorteos**

# Distribución de probabilidad:

Una distribución de probabilidad es una tabla de todos los resultados disjuntos posibles de un experimento, y su probabilidad asociada

| Dice sum | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Probability | $\frac{1}{36}$ | $\frac{2}{36}$ | $\frac{3}{36}$ | $\frac{4}{36}$ | $\frac{5}{36}$ | $\frac{6}{36}$ | $\frac{5}{36}$ | $\frac{4}{36}$ | $\frac{3}{36}$ | $\frac{2}{36}$ | $\frac{1}{36}$ |

Figure 3.5: Probability distribution for the sum of two dice.

REGLAS:

- Los resultados deben ser disjuntos
- Las probabilidades deben estar entre 0 y 1
- La suma de las probabilidades debe ser igual a 1

# Distribución de Bernoulli:

**Definición**

Una variable aleatoria Bernoulli es aquella que tiene **dos posibles resultados**:

- **Éxito (1)** con probabilidad $p$.
- **Fracaso (0)** con probabilidad $1-p$.

Se utiliza para modelar situaciones donde los dos eventos son mutuamente excluyentes.

En una encuesta preelectoral se pregunta:

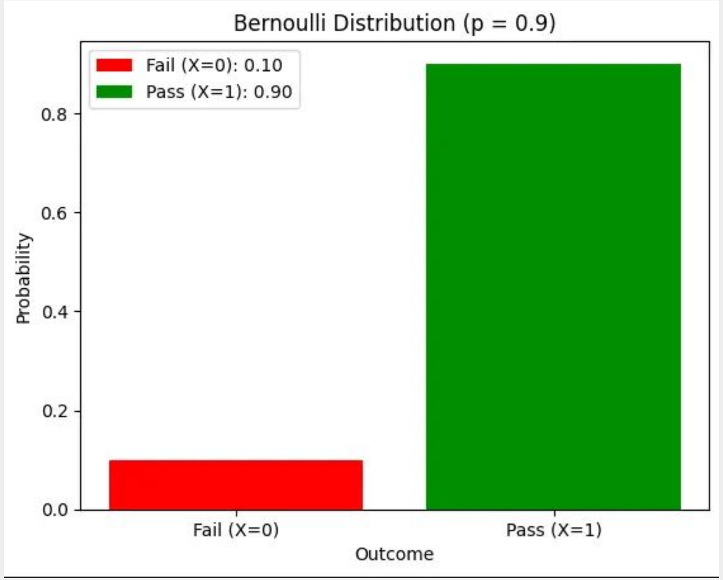
*"¿Votaría usted al candidato A en las próximas elecciones?"*

- Respuesta **Sí** → la consideramos un **éxito (1)**.
- Respuesta **No** → la consideramos un **fracaso (0)**.

Supongamos que el **70%** de los encuestados dice que sí:

$$p = 0.7 \qquad q = 1 - p = 0.3$$

* Ejemplo con p= 0.9



# Distribución de Bernoulli:

$$p = 0.7 \qquad q = 1 - p = 0.3$$

(Distribución teórica en la población)

Modelando entonces los resultados como 0 y 1 podemos calcular la media y desviación estándar de una muestra

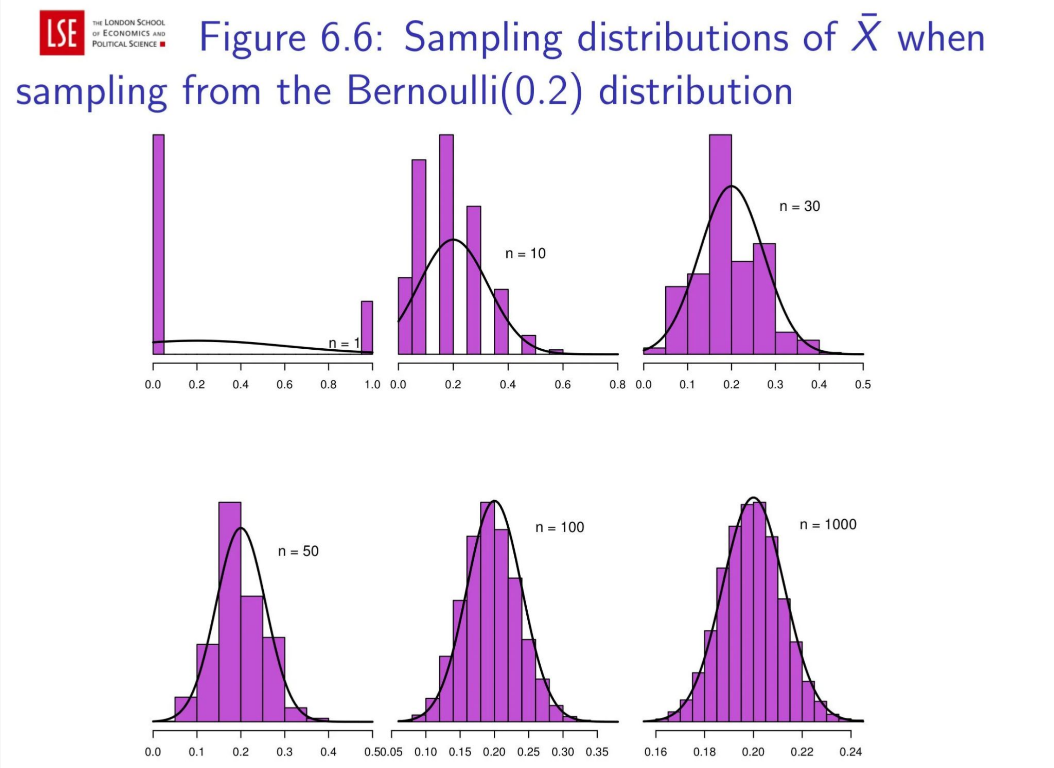
$$\hat{p} = \frac{\#\text{ of successes}}{\#\text{ of trials}} = \frac{1+1+1+0+1+0+0+1+1+0}{10} = 0.6$$

$$\mu = p \qquad \sigma = \sqrt{p(1-p)}$$

$$\sigma = \sqrt{p(1-p)} = \sqrt{0.6 \times 0.4} = \sqrt{0.24} \approx 0.490$$

# Distribución de Bernoulli:

Ejemplo de cómo la distribución de la media muestral de una variable aleatoria con distribución de Bernoulli tiende a una distribución normal a medida que n aumenta:



# Distribución geométrica:

**Definición**

- La distribución **Bernoulli** describe **un único ensayo** con dos posibles resultados (éxito o fracaso).
- La **distribución Geométrica** describe el **número de ensayos necesarios hasta obtener el primer éxito** en una secuencia de variables Bernoulli independientes e idénticamente distribuidas (*iid*).

**Problema**

Queremos encontrar al **primer votante** en una encuesta que declare que votará al **candidato A**.

- Definimos "éxito" = "Sí votará al candidato A" → probabilidad $p=0.7$.
- "Fracaso" = "No lo votará" → probabilidad $1-p=0.3$

**Pregunta:** ¿Cuál es la probabilidad de que el primer "Sí" aparezca en la 1ª, 2ª, 3ª,… respuesta?

# Distribución geométrica:

Si la probabilidad de éxito en el primer intento es ***p*** y la de fracaso es ***1-p***. Entonces la probabilidad del primer éxito en el ***i***esimo intento es:

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

- **Probabilidad de que el primer encuestado diga "Sí"**

$$P(X=1) = (0.3)^0 \cdot (0.7) = 0.7$$

- **Probabilidad de que el segundo encuestado sea el primero en decir "Sí"**

$$P(X=2) = P(1^{\text{o}} \text{ dice No, } 2^{\text{o}} \text{ dice Sí}) = (0.3)(0.7) = 0.21$$

- **Probabilidad de que el tercer encuestado sea el primero en decir "Sí"**

$$P(X=3) = P(1^{\text{o}} \text{ No, } 2^{\text{o}} \text{ No, } 3^{\text{o}} \text{ Sí}) = (0.3)(0.3)(0.7) = 0.063$$

- **Probabilidad general de que el primer "Sí" aparezca en el $n$-ésimo encuestado**

$$P(X=n) = (0.3)^{n-1}(0.7) = (1-0.7)^{n-1}(0.7)$$

# Distribución geométrica:

$$\mu = \frac{1}{p} \qquad \sigma^2 = \frac{1-p}{p^2} \qquad \sigma = \sqrt{\frac{1-p}{p^2}}$$

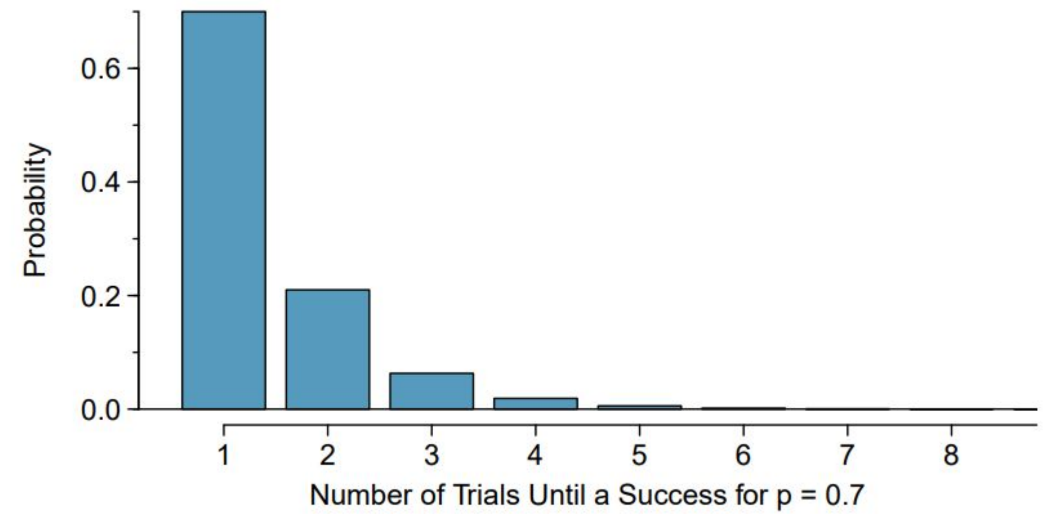



Figure 4.8: The geometric distribution when the probability of success is $p = 0.7$.

Siguiendo nuestro ejemplo deberíamos hablar en promedio con 1,42 personas hasta encontrar al primer votante del candidato A.

# Distribución geométrica:

Example problem: What is the chance that we would find the first success within the first 3 cases?

---

Solution to the example: This is the chance it is the first ($n = 1$), second ($n = 2$), or third ($n = 3$) case is the first success, which are three disjoint outcomes. Because the individuals in the sample are randomly sampled from a large population, they are independent. We compute the probability of each case and add the separate results:

$$\begin{aligned}
&P(n = 1, 2, \text{ or } 3) \\
&\quad = P(n = 1) + P(n = 2) + P(n = 3) \\
&\quad = (0.3)^{1-1}(0.7) + (0.3)^{2-1}(0.7) + (0.3)^{3-1}(0.7) \\
&\quad = 0.973
\end{aligned}$$

There is a probability of 0.973 that we would find a successful case within 3 cases.

# Distribución binomial:

**Definición**

- La distribución **Geometrica** describe el numero de ensayos que debemos esperar hasta observar el primer exito (o fracaso).

- La **distribución Binomial** se usa para describir el numero de exitos en un número fijo de intentos.
  - osea en *n ensayos Bernoulli iid (independientes e idénticamente distribuidos)*

**<u>¿Es binomial? 4 condiciones</u>**

1. Ensayos independientes.
2. ***n*** fijo
3. Cada ensayo es éxito/fracaso.
4. ***p*** constante.

**Parámetros:** n (tamaño muestral), k (cantidad de Exitos), p= probabilidad de k

$$\boxed{\mu = np \qquad \sigma^2 = np(1-p) \qquad \sigma = \sqrt{np(1-p)}}$$

# Distribución binomial:

**Problema 1: planteo**

- ¿ Cual es la probabilidad de encuestar 4 personas al azar y que 3 de ellas apoyen al candidato A y solo 1 no lo haga?

**Objetivo:** hallar P(X=3) cuando encuestamos a 4 personas

**Paso 1 - Calcular la probabilidad de un escenario en específico**

$$A = \text{No},\ B = \text{Sí},\ C = \text{Sí},\ D = \text{Sí}$$

Por independencia:

$$P(\text{escenario}) = (0.3)(0.7)(0.7)(0.7) = (0.7)^3(0.3)^1 = 0.103$$

**Fórmula para uno de los escenarios**

$$p^k(1-p)^{n-k}$$

**Paso 2 - Calcular el número de escenarios que satisfacen la condicíon**

$$\#\text{escenarios} = \binom{4}{1} = 4$$

**Fórmula de combinatoria**

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

**Paso 3 - Multiplicamos la cantidad de escenarios por su probabilidad**

$$P(X=3) = \binom{4}{1} \times (0.7)^3(0.3)^1 = 4 \times 0.103 = \mathbf{0.412}$$

**Fórmula binomial simplificada**

$$[\#\text{ of scenarios}] \times P(\text{single scenario})$$

# Distribución binomial:

**Problema 2: Ejercicio**

- ¿ Cual es la probabilidad de encuestar 8 personas al azar y que 3 personas voten al **candidato B**? (invertimos el problema para pensarlo en términos de 5 casos para el candidato A)

**Objetivo:** hallar P(k=5) cuando encuestamos a 8 personas.

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

**Fórmula general**

# Distribución binomial:

**Problema 2: Ejercicio**

- ¿ Cual es la probabilidad de encuestar 8 personas al azar y que 3 personas voten al **candidato B**? (invertimos el problema para pensarlo en términos de 5 casos para el candidato A)

**Objetivo:** hallar P(k=5) cuando encuestamos a 8 personas.

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

**Fórmula general**

$$\binom{8}{5} (0.7)^5 (1-0.7)^{8-5} = \frac{8!}{5!(8-5)!} (0.7)^5 (1-0.7)^{8-5}$$

$$= \boxed{\frac{8!}{5!3!}} \boxed{(0.7)^5 (0.3)^3}$$

**Planeteo**

# Distribución binomial:

**Problema 2: Ejercicio**

- ¿ Cual es la probabilidad de encuestar 8 personas al azar y que 3 personas voten al **candidato B**? (invertimos el problema para pensarlo en términos de 5 casos para el candidato A)

**Objetivo:** hallar P(k=5) cuando encuestamos a 8 personas.

$$\binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

**Fórmula general**

$$\binom{8}{5} (0.7)^5 (1-0.7)^{8-5} = \frac{8!}{5!(8-5)!} (0.7)^5 (1-0.7)^{8-5}$$
$$= \boxed{\frac{8!}{5!3!}} \boxed{(0.7)^5 (0.3)^3}$$

**Planeteo**

$$\boxed{\frac{8!}{5!3!} = \frac{8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{(5 \times 4 \times 3 \times 2 \times 1)(3 \times 2 \times 1)} = \frac{8 \times 7 \times 6}{3 \times 2 \times 1} = 56}$$

**Combinatoria**

Using $\boxed{(0.7)^5 (0.3)^3} \approx 0.00454$, the final probability is about $\boxed{56} \times 0.00454 \approx \boxed{0.254}$.

# Distribución de Poisson

## Definición:

La distribución de Poisson modela el **número de eventos** que ocurren en un intervalo de **tiempo o espacio**.

### Requisitos básicos

1. El evento se cuenta en una **población grande**.
2. Los eventos ocurren de manera **independiente**.
3. La tasa promedio de ocurrencia es **constante**.

👉 **Ejemplo:** La usamos para el **número de manifestaciones con corte** en el **AMBA** por **mes**.
⚠️ **Nota:** suponemos equiprobabilidad a lo largo de los meses.

# Distribución de Poisson

## Fórmula

Si X~Poisson(λ)X, la probabilidad de observar exactamente k eventos es:

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

λ=tasa promedio de eventos en el intervalo (Lambda)

k=0,1,2,
e≈2.718 (Euler, base de los logaritmos naturales)

Media y desviación estándar:

$$\mu = \lambda, \quad \sigma = \sqrt{\lambda}$$

- probabilidad de que se produzcan exactamente k sucesos en un intervalo fijo. Aquí, X es el número de sucesos, y k es el número concreto que nos interesa.
- **El numerador:** tiene dos partes. $\lambda^k$ muestra la probabilidad de que se produzcan k sucesos basándose en la tasa media λ. La $e^{-\lambda}$ tiene en cuenta la aleatoriedad de los sucesos, garantizando que la probabilidad disminuye a medida que el número de sucesos se desvía de la tasa esperada.
- **El denominador:** ¡k! ajusta de cuántas maneras pueden ocurrir los sucesos. La notación factorial calcula el número de disposiciones posibles, asegurándose de que la probabilidad refleja el hecho de que el orden de los acontecimientos no importa.

# Distribución de Poisson

## Problema 1: Ejercicio

- ¿cuál es la probabilidad de que en un mes haya **exactamente 5 manifestaciones**?

Supongamos que el promedio es **λ=3** manifestaciones por mes.

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

$$P(X = 5) = \frac{3^5 e^{-3}}{5!}$$

$$3^5 = 243, \quad 5! = 120, \quad e^{-3} \approx 0.0498$$

$$P(X = 5) = \frac{243 \times 0.0498}{120} = \frac{12.1}{120} \approx \mathbf{0.101}$$

- En Poisson:

$$\mu = \lambda, \quad \sigma = \sqrt{\lambda}$$

- Para $\lambda = 3$:

$$\mu = 3, \quad \sigma = \sqrt{3} \approx 1.73$$

- Comparando 5 manifestaciones con la media:

$$5 - 3 = 2 \text{ manifestaciones más que el promedio}$$

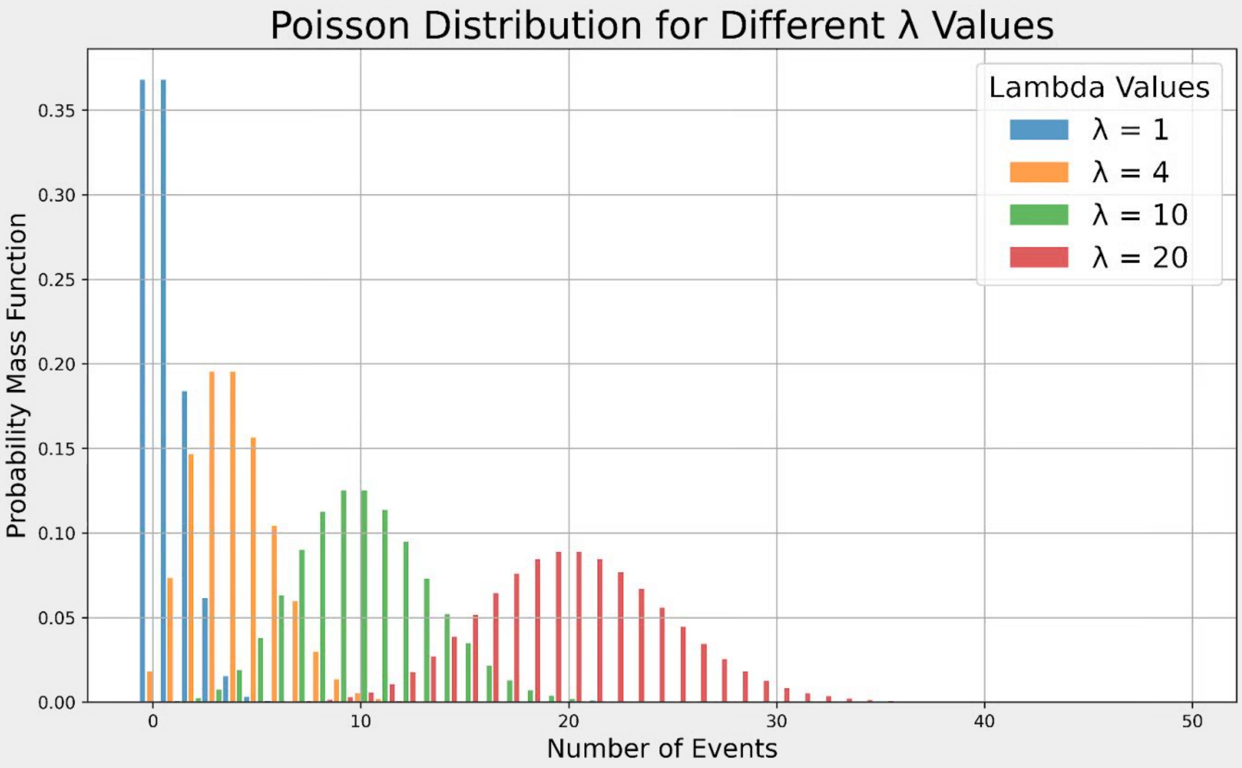
- Medido en **desvíos estándar**:

$$\frac{5 - 3}{1.73} \approx 1.15 \text{ desvíos}$$

Existe un 10% de probabilidades de que hayan 5 manifestaciones con cortes en el AMBA en un mismo mes

# Distribución de Poisson

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



# Variables Aleatorias - Discretas y su f(x) de probabilidad

Las variables aleatorias incorporan la idea del azar.

Al igual que en la estadística descriptiva, se dividen entre discretas y continuas.

Las variables aleatorias **discretas** son aquellas a las que se les asigna un número específico al resultado del experimento. Por ejemplo, si en una encuesta determinada se le pregunta a los votantes si apoyan o no a un determinado candidato, la variable aleatoria puede tomar valores como "1" si el votante apoya al candidato y "0" si no lo apoya. Esta sería una variable aleatoria discreta. Se entiende que **los valores que toma la variable son finitos o contables.**

A cada valor x de X se le asigna una probabilidad que es igual a la probabilidad de la unión de los resultados elementales para los que X = x. La probabilidad del evento X=x está dado por P (X=x), de aquí es posible hacer corresponder los valores de la variable aleatoria X con una **distribución de probabilidad** determinada. De ella surge la **función de probabilidad f(x) = P(X=x)**.  Por lo tanto, f(x) debe ser mayor a 0 y la suma de f(x) igual a 1.

# Variables Aleatorias - Discretas y f(x) de probabilidad

1. **Variable aleatoria X:** Imaginen que están analizando cuántos votantes apoyan a un candidato específico en una encuesta. La variable X podría representar **el número de personas que apoyan al candidato**, y x sería un **número específico** (por ejemplo, que X=5 significa que 5 personas apoyan al candidato).
2. **Evento asociado Ax:** El evento Ax representa el conjunto de condiciones o situaciones que hacen que la variable X tome el valor específico x. En este caso, Ax podría ser el conjunto de personas encuestadas que respondieron "sí" al apoyo del candidato, y el número de apoyos fue igual a 5.
3. **Probabilidad:** Entonces, la función nos dice que la probabilidad de observar exactamente 5 apoyos (es decir, X=5) es igual a la probabilidad de que se cumpla el evento A5, es decir, que cinco personas respondan "sí".
4. La función simplemente indica que la probabilidad de que la variable aleatoria X tome un valor específico x es igual a la probabilidad del evento Ax

# Variables Aleatorias - Discretas y su f(x) de probabilidad

**Ejemplo**

Supongamos que encuestamos a 10 personas y queremos saber cuál es la probabilidad de que exactamente 5 de ellas apoyen a un candidato. La probabilidad P(X=5) sería la probabilidad de que ocurran las condiciones $A_5$ (que consisten en obtener 5 respuestas "sí" y 5 respuestas "no" en la encuesta). Esta probabilidad se puede calcular usando distribuciones de probabilidad, como la **distribución binomial**.

# Variables Aleatorias - Discretas y su f(x) de probabilidad

A la función de distribución de probabilidad, también se le vincula **la función de distribución acumulada,** la cual nos dice la probabilidad de que una variable aleatoria X tome un valor **menor o igual** a x. En otras palabras, es la suma de todas **las probabilidades de que X sea hasta x.** La función se representa de la siguiente manera:

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

Esta es la suma (o integral) de las probabilidades individuales para todos los valores posibles de $X$ hasta $x$, lo que da lugar a la probabilidad acumulada. La función $f(t)$ es la función de probabilidad (si $X$ es discreta) o la función de densidad (si $X$ es continua), que nos indica la probabilidad de cada valor $t$ en el caso discreto, o la densidad de probabilidad en $t$ en el caso continuo.

Supongamos entonces el siguiente ejemplo:

<u>Tiramos una moneda 3 veces (experimento aleatorio)</u>

- Definimos una variable aleatoria como X = número de caras,
- X ∈ {0, 1, 2, 3} estos son los valores que puede asumir, es decir, x. Donde 0 es que nunca salga cara, 1 que sólo una vez salga, 2 que salga dos veces y 3 que las tres veces sea cara.

La **distribución de frecuencia acumulada** nos dice la probabilidad de obtener hasta un número determinado de caras (es decir, la probabilidad acumulada de obtener 0, 1, 2 o 3 caras).

# Variables Aleatorias - Discretas y su f(x) de probabilidad

$$F(x) = P(X \leq x)$$

Siguiendo la función, entonces, tenemos el siguiente procedimiento:

X = número de caras, $X \in \{0, 1, 2, 3\}$

- $F(0) = Pr(X \leq 0) = Pr(X = 0) = \frac{1}{8}$

La probabilidad de obtener 0 caras (es decir, ninguna cara) da $\frac{1}{8}$ ya que hay 2 posibles resultados por cada lanzamiento (cara o cruz) que se repiten 3 veces ($2^3 = 8$) y sólo es posible un resultado {cruz,cruz,cruz}

- $F(1) = Pr(X \leq 1) = Pr(X = 0) + Pr(X = 1) = 4/8 \rightarrow F(1)$: La probabilidad de obtener 1 o menos caras

Hay 3 maneras de obtener 1 cara: {cara, cruz, cruz}, {cruz, cara, cruz} y {cruz, cruz, cara} cada una con probabilidad $\frac{1}{8}$, por lo que se suman la probabilidad de F(0) y F(1) cuyo total es 4/8

- $F(2) = Pr(X \leq 2) = Pr(X = 0) + Pr(X = 1) + Pr(X = 2) = \frac{7}{8} \rightarrow F(2)$: La probabilidad de obtener 2 o menos caras

Para las al menos dos caras, hay 3 combinaciones posibles {cara, cara, cruz}, {cara, cruz, cara} y {cruz, cara, cara}. Por lo que la suma acumulada resulta $4/8 + \frac{3}{8} = \frac{7}{8}$

- $F(3) = Pr(X \leq 3) = 1 \rightarrow F(3)$: La probabilidad de obtener 3 o menos caras (es decir, cualquier resultado)

Se trata de la probabilidad de obtener **cualquier resultado posible** en los tres lanzamientos, lo que es **100%** o 1, porque todas las combinaciones posibles ya han sido cubiertas.

# Variables Aleatorias - Discretas y la esperanza y varianza

La media de una variable aleatoria discreta X se obtiene de la suma de los valores multiplicados por su probabilidad. En probabilidad se llama media, esperanza o valor esperado de una variable aleatoria X denotada por μ or E(X) y se calcula de la siguiente manera:

$$\mu = \sum_{x \in S_x} x\, f(x)$$

$S_X$ = el **conjunto de valores posibles** que puede tomar $X$.

$f(x)$ = la **función de probabilidad** de $X$, o sea, la probabilidad de que $X = x$.

La esperanza o media es el valor promedio que esperas de un fenómeno aleatorio si repites el experimento muchas veces. Nos da una idea de la escala o tamaño promedio del fenómeno que estamos observando y tiene la misma unidad que la variable aleatoria X. Ayuda a entender el comportamiento "normal" o "esperado" del fenómeno que estás analizando

Es un valor teórico que se obtiene de la distribución de probabilidades de X.

Un promedio puede ser 0.6 si hablamos de tiradas de monedas (6/10 veces cae cara por ej) pero la esperanza es 0.5.

# Variables Aleatorias - Discretas y la esperanza y varianza

La esperanza tiene una propiedad importante con respecto a su transformación lineal:

Dada una V.A. con media $\mu_X$, **la V.A. Y,** que se obtiene a través de una transformación lineal **Y = a +bX** , con a y b como constantes,

**tiene media** $\mu_Y = a + b\mu_X$

Esto significa que cuando aplicamos una transformación lineal a una variable aleatoria, la media de la nueva variable Y se calcula **ajustando la media de X** de acuerdo con los valores de a y b, siendo **a** una constante que **desplaza** los valores de X, mientras que **b** es una constante que **escala** los valores de X.

Por ejemplo:

Piensen que estamos estudiando el **nivel de apoyo a un candidato** en distintas encuestas. El número de personas que apoyan al candidato en cada encuesta es una variable aleatoria X con una **media $\mu_X$= 60** (es decir, en promedio, 60 personas apoyan al candidato en cada encuesta).

# Variables Aleatorias - Discretas y la esperanza y varianza

Ahora, supongamos que queremos crear una nueva variable Y que refleje un **ajuste en los resultados** por algún factor externo, como una campaña publicitaria que se espera que aumente el apoyo en **10 personas** en cada encuesta que además **duplica** el nivel de apoyo existente. En este caso:

- $a=10$ (el aumento constante de apoyo debido a la campaña),
- $b=2$ (el apoyo se duplica debido a la campaña).

La transformación sería:

$Y=10+2X$

Para calcular la **media de Y**, usamos la propiedad:

$$\mu_y = a + b\mu_X$$

$\mu_Y=10+2\mu_X = 10+2(60) = 10+120 =130$

Entonces, la media de la nueva variable Y, que incluye el aumento y la duplicación del apoyo, sería 130. Esto significa que, en promedio, después de la campaña, esperarías ver **130 personas apoyando al candidato** en cada encuesta, en lugar de las 60 iniciales.

# Variables Aleatorias - Discretas y la esperanza y varianza

La varianza de una V.A. está dada por la suma de las desviaciones de la media al cuadrado multiplicada por sus respectivas probabilidades.

Por lo tanto, siendo X una V.A. discreta con media μ, la varianza de X, denotada por $\sigma^2$ o Var(X) se calcula de la siguiente manera (para probabilidades teóricas):

$$\sigma^2 = \sum_{x \in S_X} (x - \mu)^2 f(x)$$

# Variables Aleatorias - Continua y su f(x) de densidad

Una V.A. es continua cuando no podemos calcular la probabilidad de que la variable tome un valor exacto (porque hay infinitos valores posibles). Sin embargo, podemos calcular la **probabilidad de que la variable esté dentro de un rango** de valores. Esto es lo que hace la **función de densidad de probabilidad**.

X es una V.A. continua si existe una función **f(x)** tal que la función de distribución acumulada $F(x) = P(X \leq x)$ está dada por el área debajo de f(x) hacia la izquierda de x. La **función de densidad de probabilidad** f(x) nos da una idea de **qué tan probable es que la variable esté alrededor de un valor x**, pero no exactamente en x. Para encontrar la probabilidad de que X esté en un cierto rango, calculamos el área bajo la curva de f(x) en ese rango.

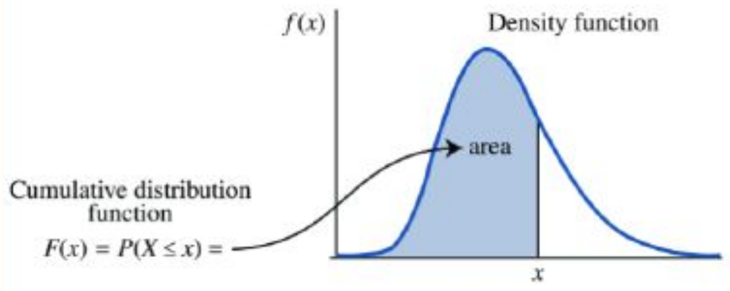



Figura 13.6 Relationship between cumulative distribution function and density function for a continuous random variable.

# Variables Aleatorias - Continua y su f(x) de densidad

La construcción de la función de densidad de una VA X resulta de tomar valores entre $l_1$ y $l_2$ y dividir el intervalo en N subintervalos con tamaño **$\delta = (l_2 - l_1) / N$**

dados por $(x_1-\delta/2, x_1-\delta/2), (x_2-\delta/2, x_2+\delta/2),...,(x_N-\delta/2, x_N+\delta/2)$, donde $x_1, x_2, \ldots, x_N$ son los puntos medios de los subintervalos. De ello, diremos que $p_1, p_2, \ldots p_N$ son las probabilidades de que X tome valores entre los subintervalos.

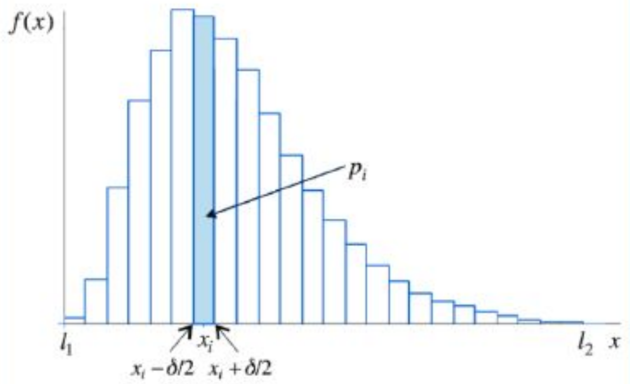



**Figura 13.7** Probability histogram of a continuous random variable.

# Variables Aleatorias - Continua y su f(x) de densidad

Un ejemplo:

Supongamos que estamos estudiando el porcentaje de votantes que apoyan a un candidato en una elección. El porcentaje puede ser cualquier valor entre 0% y 100%, por lo cual, tenemos una variable aleatoria continua. Ahora digamos que nos interesa ver el rango de porcentaje de apoyo que va desde 40% a 60%.

Nuestro intervalo de interés: $(I_1, I_2) = (40\%, 60\%)$

Definimos la longitud de nuestros intervalos en igual tamaño en 5 subintervalos (N=5)

$\delta = (I_2 - I_1) / N = (60\% - 40\%)/5 = 4\%$, por lo tanto cada intervalo tendrá una longitud de 4%

Lo siguiente es establecer los puntos medios tomando el valor inicial de cada subintervalo sumado por la mitad del subintervalo:

$x_1 = 40\% + 2/\delta = 40\% + 2\% = 42\% \rightarrow (x_1 - \delta/2, x_1 + \delta/2) = (42\% - 2\%, 42\% + 2\%) = (40\%, 44\%) = p_1$

$x_2 = 44\% + 2/\delta = 44\% + 2\% = 46\% \rightarrow (x_2 - \delta/2, x_2 + \delta/2) = (46\% - 2\%, 46\% + 2\%) = (44\%, 48\%) = p_2$

$x_3 = 48\% + 2/\delta = 48\% + 2\% = 50\% \rightarrow (x_3 - \delta/2, x_3 + \delta/2) = (50\% - 2\%, 50\% + 2\%) = (48\%, 52\%) = p_3$

$x_4 = 52\% + 2/\delta = 52\% + 2\% = 54\% \rightarrow (x_4 - \delta/2, x_4 + \delta/2) = (54\% - 2\%, 54\% + 2\%) = (52\%, 56\%) = p_4$

$x_5 = 56\% + 2/\delta = 56\% + 2\% = 58\% \rightarrow (x_5 - \delta/2, x_5 + \delta/2) = (58\% - 2\%, 58\% + 2\%) = (56\%, 60\%) = p_5$

# Variables Aleatorias - Continua y su f(x) de densidad

Volviendo al histograma presentado anteriormente, tenemos que la **altura** del rectángulo se calcula como la **densidad de probabilidad por unidad de longitud (**$p_i/\delta$), mientras que **el área está representado por** $p_i$

Ahora bien, el área total de los rectángulos del histograma es igual a 1 porque es la probabilidad de que X asuma cualquier valor dentro del intervalo ($l_1$, $l_2$). Si dividimos el intervalo en **más subintervalos** (es decir, aumentamos N y reducimos δ, el ancho de cada subintervalo), a medida que los subintervalos se hacen más pequeños, los rectángulos del histograma se vuelven más estrechos y la representación del histograma comienza a parecerse a una **curva suave**. Esta curva es la **función de densidad de probabilidad**.

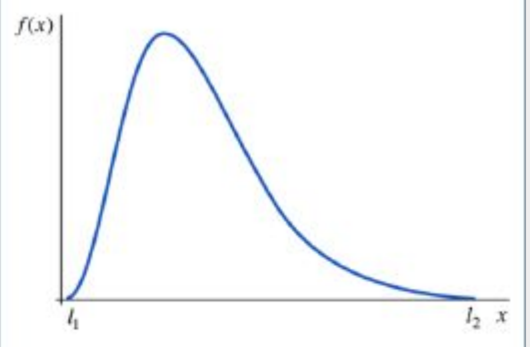



Figura 13.8 Density function of a continuous random variable.

# Variables Aleatorias - Continua y su f(x) de densidad

La propiedades de la función de densidad son las siguientes:

- es **no negativa** en todas partes, lo que significa que siempre tiene valores iguales o mayores a 0 (porque las probabilidades no pueden ser negativas).
- el **área bajo la curva** de la función de densidad es **igual a 1**, lo que asegura que la probabilidad total de que la variable aleatoria tome un valor en el intervalo completo es 100%.

# Variables Aleatorias - Continua y su f(x) de densidad

Por otro lado, la probabilidad de que una VA X tome valor entre *a* y *b* está representado por el área bajo la curva f(x) sobre el intervalo (*a,b*). Esto es posible de obtener como la diferencia entre los valores de la función de distribución acumulada en *b* y *a* donde

$$P(a < X < b) = P(X < b) - P(X < a) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$

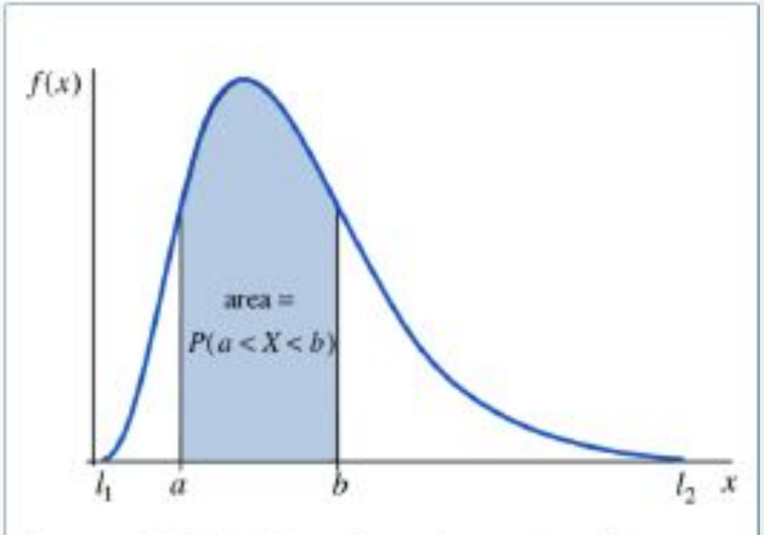



Figura 13.9 Probability of an interval for a continuous random variable.

# Variables Aleatorias - Continua y la esperanza y varianza

La media y la varianza de una VA continua se expresan a través de integrales

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx, \quad Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$$

De manera similar a una VA discreta, una VA continua Y se obtiene a partir de una transformación lineal

$Y = .a + b_X$ con $\mu_Y = a + b\mu_X$ y $\sigma^2_Y = b^2\sigma^2_Y$

# Variables Aleatorias - Continua y la estandarización

La **estandarización** es un tipo de **transformación lineal** que se aplica a una variable aleatoria X para convertirla en una nueva variable Y, de modo que tenga una **media (esperanza) de 0** y una **varianza de 1**.

Esto nos permite comparar variables que tienen diferentes escalas o unidades de medida de una manera más clara y objetiva. En el contexto de Ciencia Política, imaginemos que tenemos diferentes **encuestas** de apoyo a candidatos y queremos comparar cómo varían esos apoyos entre sí, aunque tengan **promedios y dispersiones diferentes**. Estandarizar los datos hace que las comparaciones sean más fáciles.

De aquí resulta la fórmula

$$Y = \frac{X - \mu}{\sigma}$$

# Variables Aleatorias - Continua y la estandarización

De la función anterior desagregamos lo siguiente:

$$Y = \frac{X - \mu}{\sigma}$$

X es la variable original, $\mu$ es la **media** de X y $\sigma$ es la **desviación estándar** de X.

**Restar la media $\mu$ centra** los datos alrededor de 0. Es decir, hace que la nueva variable Y tenga un promedio de 0. Si X es el porcentaje de apoyo a un candidato, al restar la media estamos midiendo **cuánto se desvía** cada encuesta del promedio de apoyo.

**Dividir por la desviación estándar $\sigma$** normaliza las diferencias respecto a la media en la nueva variable Y escalando los valores para que la dispersión sea más fácil de comparar entre distintas variables.

Por lo tanto $E(Y) = 0$ , $Var(Y) = 1$

# Variables Aleatorias - Estandarización vs transformación lineal

Una última aclaración para diferencias la estandarización y la transformación lineal

**Estandarización** es específica para resolver el problema de **comparar** variables o conjuntos de datos que tienen diferentes escalas, **normalizando** los valores.

**Transformación lineal** es más general y se utiliza cuando simplemente se quiere **reajustar** los valores de una variable a una escala distinta, sin necesariamente normalizarla.

# Variables Aleatorias - Elementos

**~~9. Definición de Distribuciones Probabilísticas Paramétricas~~**
**~~10. Distribución DIscreta Uniforme~~**
**~~11. Distribución de Bernoulli~~**
**~~12. Distribución Binomial~~**

**~~14. Distribución Continua Uniforme~~**
**~~16. Distribución Normal~~**
**~~17. Propiedades de la Distribución Normal~~**
**~~18. Regla Empírica de la Distribución Normal~~**
**19. Función de la Distribución Acumulativa de la Normal**
**20. Distribución Normal Estándar**
**21. Tablas (z y t)**

# Distribuciones Paramétricas de Probabilidad

Una **distribución paramétrica de probabilidad** es una herramienta que nos permite hacer predicciones sobre un conjunto de datos, basándonos en ciertos parámetros que describen cómo se comportan esos datos. Estos parámetros son **constantes numéricas** que afectan la forma en que se distribuyen los datos y que determinan las **probabilidades** de que ocurran ciertos valores.

En definitiva, las distribuciones paramétricas nos permiten modelar muchas situaciones diferentes con **una fórmula general** que cambia según los valores de sus parámetros.

Veamos algunas…

# Distribución Uniforme Discreta - V. A. Discretas

La **distribución uniforme discreta** es útil cuando analizamos situaciones donde **todos los resultados tienen la misma probabilidad**. Podemos aplicarla en situaciones como la elección aleatoria de candidatos, sorteos, o cualquier proceso donde haya un número igual de oportunidades para cada opción.

Una VA X tiene una distribución uniforme discreta en sus primeros n números naturales si la función de probabilidad de X es

$$f(x) = \frac{1}{n}, x = 1, 2, \ldots, n$$

Con $E(X) = \frac{n+1}{2},\ Var(X) = \frac{n^2-1}{12}.$

Esto significa que **cada número** del 1 al n tiene una probabilidad igual de 1/n, por lo tanto, el valor promedio estará aproximadamente en el medio y, a su vez, si el valor de n aumenta, la varianza también crece, lo que significa que los valores estarán más dispersos.

# Distribución Uniforme Discreta - V. A. Discretas

Veamos un ejemplo…

Supongamos un bolillero con 8 bolitas numeradas del 1 al 8. El experimento consiste en sacar una bolilla al azar, lo que significa que cada bolita tiene la misma probabilidad de ser seleccionada. Esta situación es un ejemplo de una **distribución uniforme discreta**, ya que todas las bolitas tienen la misma probabilidad de ser extraídas.

$$f(x) = \frac{1}{n}, x = 1, 2, \ldots, n$$

$$E(X) = \frac{n+1}{2},\ Var(X) = \frac{n^2-1}{12}.$$

La variable X representa el número de bolilla que se extrae, por lo tanto X puede asumir valores 1, 2, 3, 4, 5, 6, 7, 8. Así definimos entonces que la función de probabilidad es la siguiente:

$f(x) = 1/n$ , $X = 1, 2, .., n$ $\quad$ $f(x) = \frac{1}{8} = 0{,}125$, $X=1, 2, 3, 4, 5, 6, 7, 8$

$E(X) = n + 1/ 2 = 8+1/2 = 4{,}5$ $\quad$ y $\quad$ $Var(X) = n^2-1/12 = 8^2-1/12 = 5{,}25$

Según la media, es probable que el resultado de la bolita sea 4 o 5, pero la alta varianza, indica que el resultado pueda distar ampliamente, lo que indica equiprobabilidad (cada evento es igualmente probable)

# Distribución de Bernoulli - V.A. Discretas

Consideremos un experimento aleatorio cuyos resultados son dos eventos mutuamente excluyentes A y A′, convencionalmente llamados **éxito** y **fracaso**, codificados como 1 y 0, respectivamente. Denotemos por p la probabilidad de que ocurra A (éxito) y por 1−p la probabilidad de que ocurra A′ (fracaso). La ejecución única de este experimento se llama un **ensayo de Bernoulli**, denominado así en honor al matemático suizo Jacob Bernoulli.

Una **variable aleatoria discreta** X tiene una **distribución de Bernoulli** si toma los valores 0 y 1 con probabilidades p y 1−p, respectivamente, donde $0<p<1$. La **función de probabilidad** de esta variable aleatoria es:

$$f(x) = p^x(1-p)^{1-x}, x = 0, 1.$$

- La **media** es $E(X) = p$
- La **varianza** es $Var(X) = p(1-p)$

# Ensayo de Bernoulli y la Distribución Binomial

Veamos un ejemplo

Consideren un proceso industrial donde se producen bienes en el sector de electrónica. La experiencia sugiere que el 1% de los productos son defectuosos. La selección aleatoria de una pieza puede ser pensada como un ensayo de Bernoulli, donde el evento "éxito" es identificado con una pieza defectuosa. La probabilidad de éxito es de 0,01 (1%) con reposición. La reposición indica que la probabilidad no cambie (que se mantenga el n)

El ensayo de Bernoulli entonces tiene las siguientes características

**Éxito**: Obtener una pieza defectuosa.

**Fracaso**: Obtener una pieza no defectuosa.

**Probabilidad de éxito**: $p = 0.01$ (1%).

**Probabilidad de fracaso**: $1-p = 0.99$

# Ensayo de Bernoulli y la Distribución Binomial

Ahora supongamos que de un lote, 20 piezas son seleccionadas con reposición. El **número de piezas defectuosas** en las 20 extracciones es una **variable aleatoria**, cuyos posibles valores son los números enteros del 0 al 20.

Dado que estamos haciendo 20 ensayos de Bernoulli, esta situación sigue una **distribución binomial**, que es utilizada para contar cuántos éxitos (piezas defectuosas) ocurren en un número fijo de ensayos independientes (20 extracciones).

Las características principales son:

- Un ensayo de Bernoulli se repite un cierto número de veces.
- La probabilidad de éxito no cambia de un ensayo a otro.
- Los ensayos son independientes.

La probabilidad de éxito se mantiene sin cambios y los ensayos son independientes debido al hecho de que las selecciones se realizan con reemplazo.

# Distribución Binomial - V.A. Discretas

La distribución binomial es una distribución de probabilidad discreta que es, en definitiva, una generalización de la distribución de Bernoulli. Cuenta el número de éxitos en *n* intentos de Bernoulli.

Una variable aleatoria discreta X tiene distribución binomial si su función de probabilidad es

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, \ldots, n.$$

donde $p \in (0, 1)$ ;

n es un número entero positivo y $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ es el coeficiente binomial.

Con E(X) = np , Var(X) = np(1-p)

La distribución binomial depende de los parámetros n y p que pueden cambiar la forma de su distribución de probabilidad de manera tal que

- symmetric if $p = 0.5$
- positively skewed if $p < 0.5$
- negatively skewed if $p > 0.5$

# Distribución Binomial - V.A. Discreta

Repongamos el ejemplo de las piezas defectuosas. Dijimos que se seleccionaron 20 piezas al azar con reposición y establecimos que la Variable Aleatoria X es la probabilidad de sacar una pieza defectuosa. Según el ejemplo, la probabilidad de encontrar una pieza defectuosa es el 1%. Ya tenemos algunos parámetros establecidos.

p = 0,01

n = 20

por lo tanto, f (x) = $\binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, \ldots, n.$ con $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ como coeficiente,

si reemplazamos entonces tenemos la siguiente función de distribución

f(x) = (20/x)0,01$^x$(1-0,01), x= 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20

Con la función, lo que resta es establecer la probabilidad de encontrar x cantidad de piezas defectuosas.

# Distribución Binomial - V.A. Discreta

Reemplacemos entonces:

**Ejemplo 1: ¿Cuál es la probabilidad de que no haya piezas defectuosas?** Acá estamos buscando P(X=0)

$$P(X=0) = \binom{20}{0}(0.01)^0(0.99)^{20} = (0.99)^{20} \longrightarrow P(X=0) \approx 0.817$$

**Ejemplo 2: ¿Cuál es la probabilidad de que una pieza sea defectuosa?** Acá estamos buscando P(X=1)

$$P(X=1) = \binom{20}{1}(0.01)^1(0.99)^{19} \longrightarrow P(X=1) = 20 \cdot (0.01) \cdot (0.99)^{19} \approx 0.165$$

Con E(X) = np = 20.0,01 = 0,2 y Var(X) = np(1-p) = 20.0,01(1-0,01) = 0,198

$$E(X) = np, \ Var(X) = np(1-p)$$

# Distribución Binomial - V.A. Discreta

La distribución binomial depende de los parámetros n y p que pueden cambiar la forma de su distribución de probabilidad de manera tal que

- es **simétrica** si p = 0,5
- tiene **sesgo positivo** si p < 0,5
- tiene **sesgo negativo** si p > 0,5

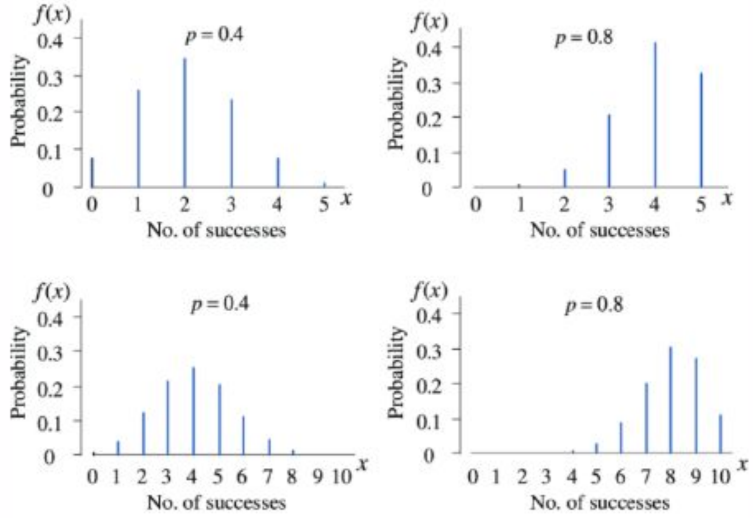



Figura 14.2 Binomial distribution for $p = 0.4, 0.8$ and $n = 5, 10$.

# Distribución Uniforme Continua - V.A. Continua

Nuevamente, estamos frente a una distribución donde todos los valores en un intervalo tienen la misma probabilidad de ocurrir. Una VA X tiene una distribución uniforme continua cuando su función de densidad es la siguiente

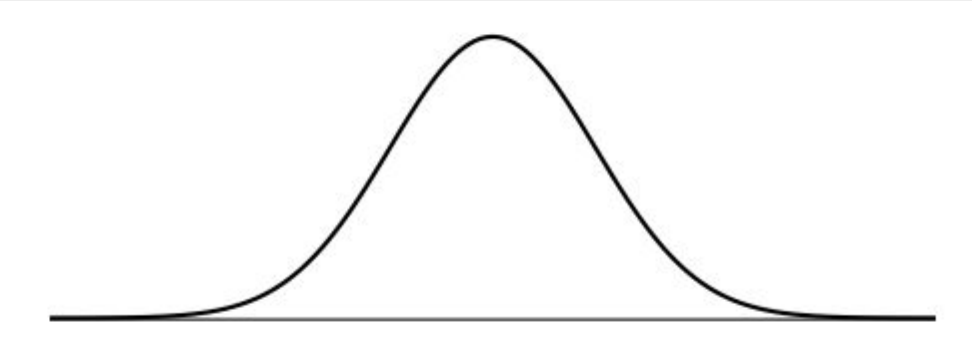
$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

Esto significa que la probabilidad de encontrar un valor en cualquier subintervalo del intervalo [a,b] es proporcional al tamaño de ese subintervalo.

A su vez, la media y la varianza se calculan como $E(X) = \frac{a+b}{2}, \quad Var(X) = \frac{(a-b)^2}{12}$

# Distribución normal

# Distribución normal



De entre todas las distribuciones, la normal es la más común. Es simétrica, unimodal, y tiene forma de campana de Gauss.

Es la forma de distribución que siguen variables como notas en un examen o alturas de mujeres o varones en un país dado.

Si bien muchas variables se acercan a tener una distribución normal, en generar ninguna lo es perfectamente. Aunque la distribución normal no sea perfecta para cada problema individual, es muy útil para aplicar a una multiplicidad de problemas.

# Distribución Normal o de Gauss - V.A. Continua

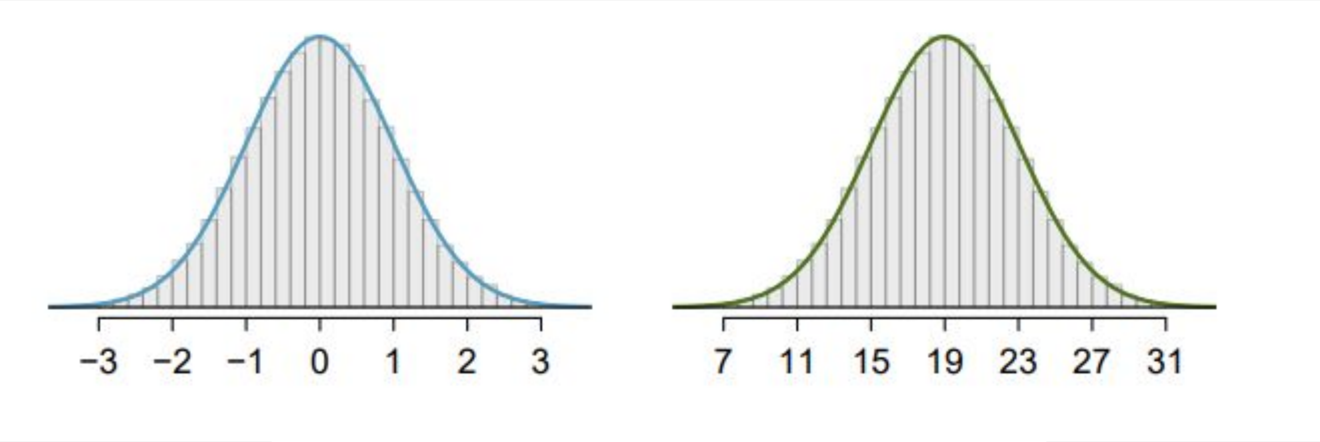
La distribución normal o de Gauss es una distribución de probabilidad continua que es simétrica y tiene forma de campana. Es una de las distribuciones más importantes en estadística debido a su presencia en muchos fenómenos naturales y sociales. Una VA continua X tiene distribución normal cuando su función de densidad está dada por

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad x \in \mathbb{R}$$

Donde la media y la varianza son $E(X) = \mu$ y $Var(X) = \sigma^2$

# Distribución Normal o de Gauss - V.A. Continua

Aunque todas las distribuciones normales son simétricas, unimodales y siguen una forma de campana de Gauss, no son todas iguales. La variabilidad en la forma de las distintas distribuciones normales está dada por dos parámetros de sus variables: μ (media) y σ (desviación estándar).

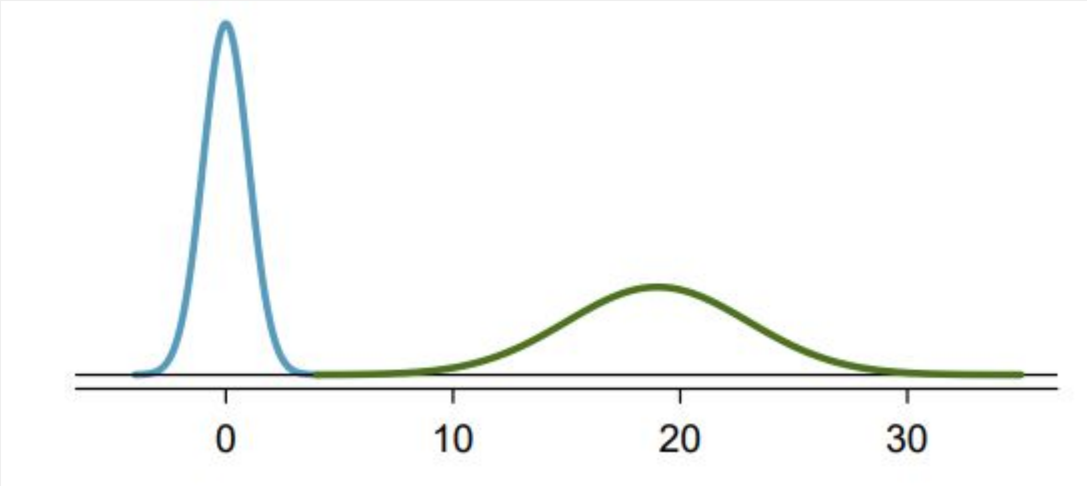



$$N(\mu = 0, \sigma = 1) \quad \text{and} \quad N(\mu = 19, \sigma = 4)$$

Se lee como: la azul sigue una distribución normal con media 0 y desvío estándar 1

# Distribución Normal o de Gauss - V.A. Continua

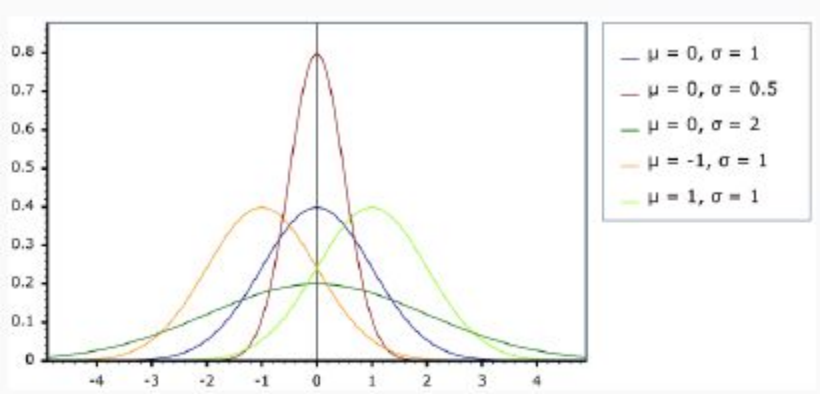
Parecen iguales, pero al graficarlas juntas notamos la diferencia:




La distribución verde presenta mayor dispersión alrededor de su media.

# Distribución Normal

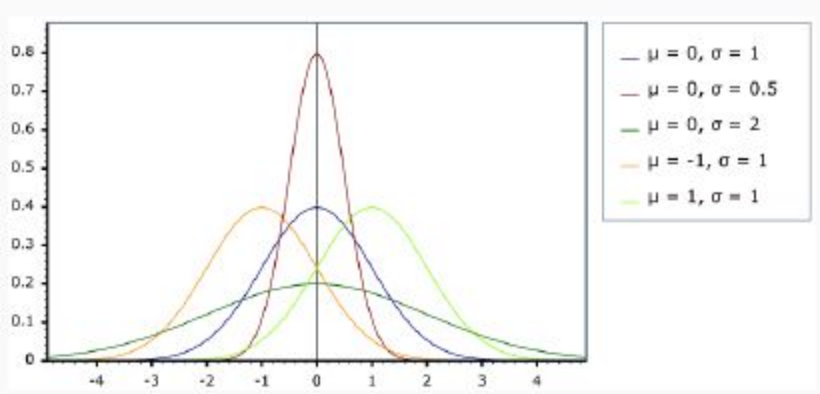
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad x \in \mathbb{R}$$




- La función **f(x)** nos indica qué tan probable es que la variable aleatoria tome **un valor cercano a x**. No es una probabilidad directa, sino que el área bajo la curva representa probabilidades, siendo la probabilidad total (el área total bajo la curva) igual a 1
- La media **μ** determina dónde está **centrada** la distribución. Es el punto donde la curva alcanza su valor más alto, es decir, donde está **el pico de la curva**.
- Una **desviación estándar σ** más grande significa que los datos están más dispersos, y la curva será **más ancha y plana**. Por el contrario, si σ es pequeña, la curva será **más estrecha y pronunciada**.

# Distribución Normal

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad x \in \mathbb{R}$$




- La constante **√2π** ajusta la curva de probabilidad para asegurarse de que **el área total** bajo ella **sea 1**. Es una expresión geométrica.
- En su conjunto, **σ√2π** es el factor de escalamiento que **ajusta la altura de la curva** según el valor de la desviación estándar σ para que **la suma de todas las probabilidades** (es decir, el área bajo la curva) sea **igual a 1**.
- El término exponencial **indica que la función cae rápidamente a medida que nos alejamos de** μ y que la caída es más pronunciada si la diferencia (x−μ) es grande. El **cuadrado** $(x-\mu)^2$ asegura que cualquier distancia de la media, ya sea positiva o negativa, afecte la altura de la curva de manera **simétrica**. Y, por último, la **división por $2\sigma^2$** ajusta la velocidad de esta caída. Cuanto mayor sea σ, más lentamente caerá la curva, ya que los valores están más dispersos.

# Distribución Normal

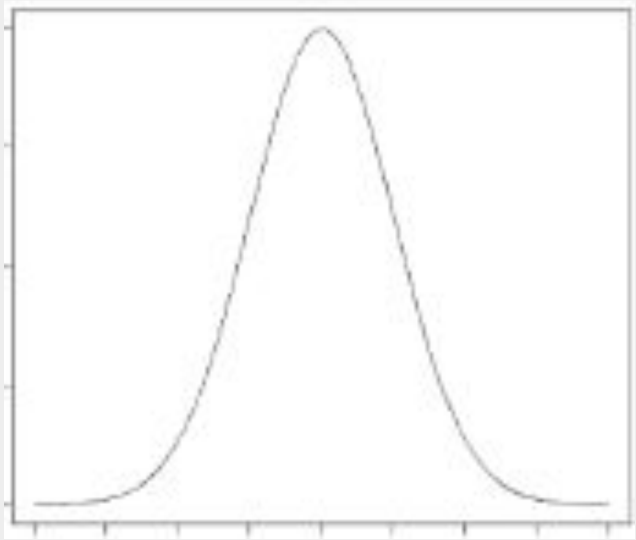
Entonces…

Imaginemos una curva de campana donde el **pico** está en la **μ** y

el **ancho de la curva** está determinado por la **desviación estándar σ**.

Cuanto uno más se aleja del centro, menor es la probabilidad de

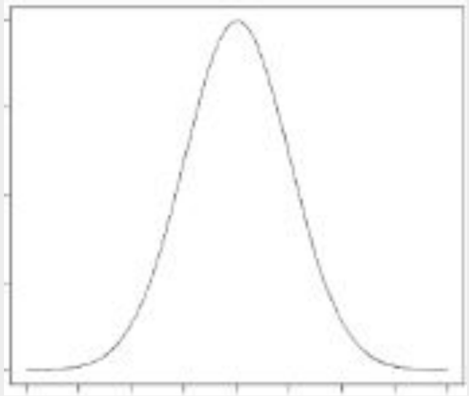
encontrar un valor allí



# Distribución Normal - Propiedades de la función

La curva normal cumple con las siguientes propiedades:

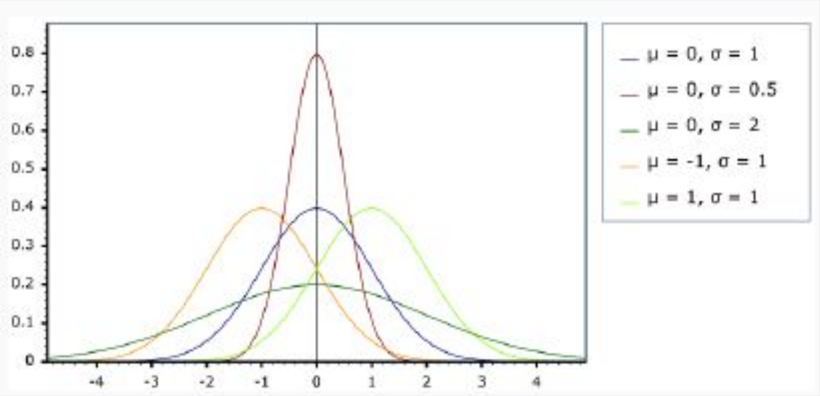
- Es **simétrica** alrededor de $\mu$ lo que significa que cualquier valor por encima de la media tiene una probabilidad equivalente a su contraparte por debajo de la media, siempre que ambas estén a la misma distancia de $\mu$
- Es **creciente** en el intervalo $(-\infty,\mu)$ y **decreciente** en el intervalo $(\mu,\infty)$; alcanzando su **máximo** en $x=\mu$.
- Tiene dos **puntos de inflexión** en $x=\mu-\sigma$ y $x=\mu+\sigma$; es **cóncava** (hacia abajo) en el intervalo $(\mu-\sigma,\mu+\sigma)$ y **convexa** en otros lugares (colas de la distribución).



# Distribución Normal - Propiedades de la función

Retomando el gráfico anterior…

- La **media** determina la **ubicación** de la función de densidad en el eje x: es el punto donde la curva normal alcanza su pico.
- La **varianza** determina la **altura** y el **ancho** de la curva normal. Una varianza grande produce una curva plana y ancha, mientras que una varianza pequeña produce una curva pronunciada y estrecha.



# Distribución Normal  - La regla empírica

En una distribución normal, la regla empírica nos indica dónde cae una observación o valor en la distribución. La misma establece que aproximadamente el 68% de los valores caen dentro de una desviación estándar de la media; alrededor del 95% dentro de dos desviaciones estándar; y el 99.7% dentro de tres desviaciones estándar.

Según la regla empírica las áreas bajo la curva normal en los intervalos $(\mu - \sigma, \mu + \sigma)$,$(\mu - 2\sigma, \mu + 2\sigma)$ y $(\mu - 3\sigma, \mu + 3\sigma)$ son aproximadamente iguales a 0.68, 0.95 y 0.997, respectivamente.

Esto significa que la probabilidad de observar un valor de X dentro de:

- una desviación estándar por encima o por debajo de la media es 0.68
- dos desviaciones estándar por encima o por debajo de la media es 0.95
- tres desviaciones estándar por encima o por debajo de la media es 0.997

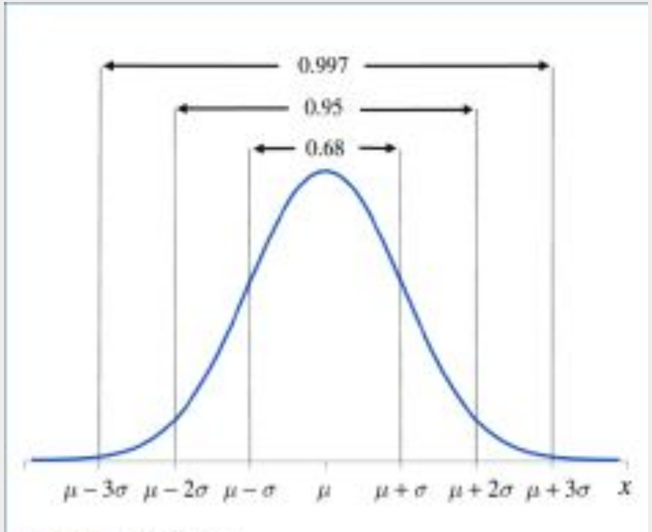


Figura 14.11 Empirical rule.

# Distribución Normal  - La regla Empírica

Por ejemplo…

Supongamos que se observa una muestra de n = 200 hospitales

con distribución normal,  media μ = 35 de paros cardíacos por mes, y

una desviación estándar de  σ = 6.

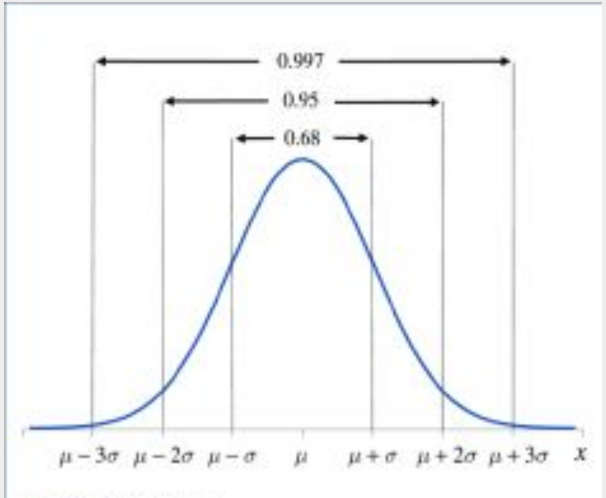


Figura 14.11 Empirical rule.

La regla empírica nos dirá que:

- Alrededor del 68% de los hospitales reportará paros cardíacos en el intervalo 35±6, es decir, entre 29 y 41
- Alrededor del 95% de los hospitales reporta paros cardíacos en el intervalo 35 ± (2×6), es decir, entre 23 y 47.
- Alrededor del  99.7%de los hospitales reporta paros cardíacos en el intervalo 35 ± (3×6), es decir, entre 17 y 53

# Distribución Normal - La función de acumulación

Volvamos nuevamente a la función de distribución acumulada.

Dijimos previamente que la función F(x) de acumulación de una variable aleatoria continua está dada por el área que se encuentra debajo de la curva normal hacia la izquierda de x; en el caso de la distribución normal, para encontrar la probabilidad de que un valor aleatorio sea menor o igual a **x** se integra la función de densidad desde **−∞** hasta **x**, lo que resulta en esta función:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt$$

La función de distribución acumulada nos dice qué proporción de los datos cae por **debajo** de un cierto valor **x**. La complejidad de la función sólo puede darnos aproximaciones numéricas utilizando softwares estadísticos o la tabla normal para acceder a la distribución acumulada.

# Estandarización

Dada la dificultad de utilizar la función de densidad de la normal y la integración para calcular probabilidades en la práctica se utiliza la estandarización de las variables. Esto se debe a que, cuando una variable sigue una distribución normal, la probabilidad de observar un valor determinado es más o menos conocida si se lo lleva a términos de desvío estándar. Es decir no nos quedamos con:

¿Cuál es la probabilidad de obtener un 5? o ¿Cuál es la probabilidad de que las personas midan entre 1.5 y 1.6 mts?

Sino con:

¿Cuán probable es encontrar un caso que está a x desvíos estándar de la media de la variable?

# Estandarización

Este proceso hace que se pase de una variable con distribución normal a una con distribución normal estándar.

Pero la estandarización tiene múltiples aplicaciones:

- Estandarizar un caso específico para ver su probabilidad de ocurrencia
- Llegar a una normal estándar (estandarización de una variable)
- Para evaluar cuánto se aleja un parámetro del de una distribución bajo la hipótesis nula (test de hipótesis)

Permite llevar datos, variables, a una misma escala y hacer comparaciones.

# Distribución Normal  - La función de acumulación

Sin embargo, se puede aproximar la probabilidad de que X caiga en un intervalo (a,b) calculado

como **P(a < X < b) = F(b) − F(a).**

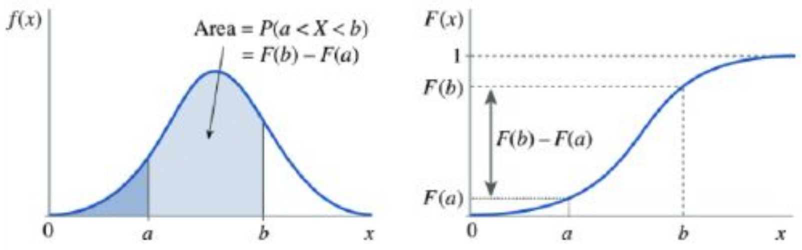



**Figura 14.13** Probability of the interval $(a, b)$ as difference between the values of $F(x)$ at $b$ and $a$.

# Distribución Normal - Ejemplo en Python

Definimos una distribución normal con una media $\mu = 50$ y una desviación estándar $\sigma = 10$.

Calculamos la probabilidad de que el valor de una variable aleatoria X sea menor o igual a 60 ($P(X \leq 60)$).

Vamos a calcularlo y verlo con algunas funciones de Python…

# Estandarización de una observación (puntaje z)

Si quisiéramos saber como fue el desempeño de dos alumnos/as con relación al resto, pero ambos alumnos/as rindieron exámenes distintos, con escalas distintas, ¿cómo se resuelve?

Por ejemplo Ann tomó el SAT y Tom el ACT:

| | SAT | ACT |
|---|---|---|
| Mean | 1100 | 21 |
| SD | 200 | 6 |

¿Quién tuvo mejor desempeño en relación al resto de sus compañeros/as?

Deberíamos poder establecer una medida de relación con la dispersión de cada grupo de notas

# Estandarización de una observación (puntaje z)

Para problemas de este estilo se suele usar una estandarización basada en el puntaje Z. Cuando llevamos a un caso a un puntaje Z, este nos indica el número de desviaciones estándar que se encuentra por encima o por debajo de la media.

Fórmula del puntaje Z:

$$Z = \frac{x - \mu}{\sigma}$$

Si μSAT = 1100 y σSAT = 200 y xAnn = 1300, podemos calcular el puntaje Z de Ann:

$$Z_{\text{Ann}} = \frac{x_{\text{Ann}} - \mu_{\text{SAT}}}{\sigma_{\text{SAT}}} = \frac{1300 - 1100}{200} = 1$$

Ann está un desvío estándar por encima de la media.

# Estandarización de una observación (puntaje z)

Por su parte, el Z-score de Tom está dado por:

$\mu ACT = 21$ y $\sigma ACT = 6$. Si $xTom = 24$, su puntaje Z es 0.5.

Tom está a 0,5 desvíos estándar por encima de la media.

¿Quién tuvo el mejor rendimiento en su examen?

# Estandarización de una observación (puntaje z)

Por su parte, el Z-score de Tom está dado por:

$\mu ACT = 21$ y $\sigma ACT = 6$. Si $xTom = 24$, su puntaje Z es 0.5.

Tom está a 0,5 desvíos estándar por encima de la media.

¿Quién tuvo el mejor rendimiento en su examen? → Ann tuvo un mejor rendimiento relativo.

# Estandarización de una normal: la Normal estándar

Si recuerdan, para las variables aleatorias, establecimos que hay transformaciones lineales que sirven para corregir valores dentro de la muestra observada.

A su vez, hay transformaciones que normalizan muestras distintas con medias y desviaciones propias para poder hacerlas comparables entre sí.

Mientras la distribución normal nos permite ver **cómo están distribuidos los datos en su forma original**, desde su centro hasta los extremos; la distribución normal estándar toma cualquier conjunto de datos que no son comparables entre sí y los **transforma** para que todos tengan una media de **0** y una desviación estándar de **1**.

En la primera los datos tienen una forma específica de distribución y queremos entender cómo se comportan en su escala original, en la segunda queremos **estandarizar** los datos para hacer comparaciones más fáciles o calcular probabilidades.

# Estandarización de una normal: la Normal estándar

El proceso de llevar una variable normal a una normal estándar supone la misma transformación que para un valor particular solo que para todos los valores de la variable y hace que la variable resultante tenga

$\mu = 0$

$\sigma = 1$

Así, podemos comparar en distintas distribuciones, distintas variables qué tan alejados están ciertos valores con respecto a la media permitiéndonos establecer probabilidades. En la práctica se estima:

$P(Z \leq z)$

$P(Z \leq 0.5) = 0.69$

Aproximadamente el 69% de los alumnos/as obtuvieron una nota igual o peor que la de Tom.

El 0.69 se obtiene a través de softwares estadísticos o de tablas z.

# Tabla z

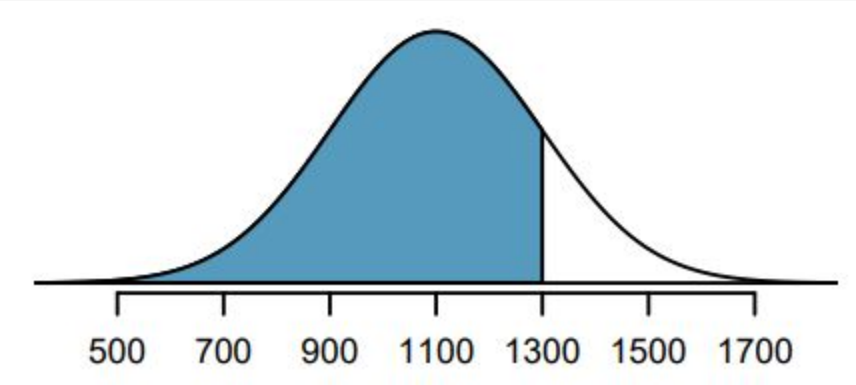
## Standard Normal Distribution
### Z-Score Table

| Z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| -1.4 | 0.08076 | 0.07927 | 0.07780 | 0.07636 | 0.07493 | 0.07353 | 0.07215 | 0.07078 | 0.06944 | 0.06811 |
| -1.3 | 0.09680 | 0.09510 | 0.09342 | 0.09176 | 0.09012 | 0.08851 | 0.08691 | 0.08534 | 0.08379 | 0.08226 |
| -1.2 | 0.11507 | 0.11314 | 0.11123 | 0.10935 | 0.10749 | 0.10565 | 0.10383 | 0.10202 | 0.10027 | 0.09853 |
| -1.1 | 0.13567 | 0.13350 | 0.13136 | 0.12924 | 0.12714 | 0.12507 | 0.12302 | 0.12100 | 0.11900 | 0.11702 |
| -1.0 | 0.15866 | 0.15625 | 0.15386 | 0.15151 | 0.14917 | 0.14686 | 0.14457 | 0.14231 | 0.14007 | 0.13786 |
| -0.9 | 0.18406 | 0.18141 | 0.17879 | 0.17619 | 0.17361 | 0.17106 | 0.16853 | 0.16602 | 0.16354 | 0.16109 |
| -0.8 | 0.21186 | 0.20897 | 0.20611 | 0.20327 | 0.20045 | 0.19766 | 0.19489 | 0.19215 | 0.18943 | 0.18673 |
| -0.7 | 0.24196 | 0.23885 | 0.23576 | 0.23270 | 0.22965 | 0.22663 | 0.22362 | 0.22062 | 0.21770 | 0.21476 |
| -0.6 | 0.27425 | 0.27093 | 0.26763 | 0.26435 | 0.26109 | 0.25785 | 0.25463 | 0.25143 | 0.24825 | 0.24510 |
| -0.5 | 0.30854 | 0.30503 | 0.30153 | 0.29806 | 0.29460 | 0.29116 | 0.28774 | 0.28434 | 0.28096 | 0.27760 |
| -0.4 | 0.34458 | 0.34090 | 0.33724 | 0.33360 | 0.32997 | 0.32636 | 0.32276 | 0.31918 | 0.31561 | 0.31207 |
| -0.3 | 0.38209 | 0.37828 | 0.37448 | 0.37070 | 0.36693 | 0.36317 | 0.35942 | 0.35569 | 0.35197 | 0.34827 |
| -0.2 | 0.42074 | 0.41683 | 0.41294 | 0.40905 | 0.40517 | 0.40129 | 0.39743 | 0.39358 | 0.38974 | 0.38591 |
| -0.1 | 0.46017 | 0.45620 | 0.45224 | 0.44828 | 0.44433 | 0.44038 | 0.43644 | 0.43251 | 0.42858 | 0.42465 |
| 0.0 | 0.50000 | 0.50399 | 0.50798 | 0.51197 | 0.51595 | 0.51994 | 0.52392 | 0.52790 | 0.53188 | 0.53586 |
| 0.1 | 0.53983 | 0.54380 | 0.54776 | 0.55172 | 0.55567 | 0.55962 | 0.56356 | 0.56749 | 0.57142 | 0.57535 |
| 0.2 | 0.57926 | 0.58317 | 0.58706 | 0.59095 | 0.59483 | 0.59871 | 0.60257 | 0.60642 | 0.61026 | 0.61409 |
| 0.3 | 0.61791 | 0.62172 | 0.62552 | 0.62930 | 0.63307 | 0.63683 | 0.64058 | 0.64431 | 0.64803 | 0.65173 |
| 0.4 | 0.65542 | 0.65910 | 0.66276 | 0.66640 | 0.67003 | 0.67364 | 0.67724 | 0.68082 | 0.68439 | 0.68793 |
| 0.5 | 0.69146 | 0.69497 | 0.69847 | 0.70194 | 0.70540 | 0.70884 | 0.71226 | 0.71566 | 0.71904 | 0.72240 |
| 0.6 | 0.72575 | 0.72907 | 0.73237 | 0.73565 | 0.73891 | 0.74215 | 0.74537 | 0.74857 | 0.75175 | 0.75490 |
| 0.7 | 0.75804 | 0.76115 | 0.76424 | 0.76730 | 0.77035 | 0.77337 | 0.77637 | 0.77935 | 0.78230 | 0.78524 |
| 0.8 | 0.78814 | 0.79103 | 0.79389 | 0.79673 | 0.79955 | 0.80234 | 0.80511 | 0.80785 | 0.81057 | 0.81327 |
| 0.9 | 0.81594 | 0.81859 | 0.82121 | 0.82381 | 0.82639 | 0.82894 | 0.83147 | 0.83398 | 0.83646 | 0.83891 |
| 1.0 | 0.84134 | 0.84375 | 0.84614 | 0.84849 | 0.85083 | 0.85314 | 0.85543 | 0.85769 | 0.85993 | 0.86214 |
| 1.1 | 0.86433 | 0.86650 | 0.86864 | 0.87076 | 0.87286 | 0.87493 | 0.87698 | 0.87900 | 0.88100 | 0.88298 |
| 1.2 | 0.88493 | 0.88686 | 0.88877 | 0.89065 | 0.89251 | 0.89435 | 0.89617 | 0.89796 | 0.89973 | 0.90147 |
| 1.3 | 0.90320 | 0.90490 | 0.90658 | 0.90824 | 0.90988 | 0.91149 | 0.91309 | 0.91466 | 0.91621 | 0.91774 |
| 1.4 | 0.91924 | 0.92073 | 0.92220 | 0.92364 | 0.92507 | 0.92647 | 0.92785 | 0.92922 | 0.93056 | 0.93189 |

# Estandarización de una normal: la Normal estándar

Y veamos gráficamente qué tan común es tener una nota igual o peor que la que obtuvo Ann?

(Recordemos que su puntaje Z era 1)




$P(Z \leq z)$

$P(Z \leq 1) = 0.84$

Alrededor del 84% de los que tomaron el SAT tuvieron una nota igual o peor que la de Ann.

# Distribución Normal Estándar

La distribución normal estándar es la distribución normal con media 0 y varianza 1 denotada por N(0,1). Su función de densidad es

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

Z deriva de una transformación estandarizada de X donde **Z = (X − μ)/σ.**

Esto significa que la nueva función de distribución acumulada de X resulta la siguiente que, aunque no parezca, es más simple y las probabilidades acumuladas se pueden buscar directamente en **tablas de valores Z**:

$$F(x) = P(X \leq x) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

# Distribución Normal Estándar

$\Phi(z)$ es la función de **distribución acumulada** de la **distribución normal estandarizada**.

Los valores de $\Phi(z)$, para muchos valores de z están tabulados en las tablas de estandarización normal, por ejemplo:

$$\Phi(0) = 0.5 \quad \Phi(0.1) = 0.5398$$
$$\Phi(1) = 0.8413 \quad \Phi(3) = 0.9986$$

La FDA de la normal estándar nos dice la probabilidad acumulada de que un valor estandarizado **Z** sea menor o igual a un valor determinado.

# Distribución Normal Estándar - Ejemplo

Digamos que hacemos una encuesta para medir el nivel de satisfacción con el desempeño de un presidente en una escala de 0 a 10. Supongamos que los resultados siguen una **distribución normal** con media **μ = 5** y desviación estándar **σ = 2**. Queremos saber:

- **¿Cuál es la probabilidad de que un encuestado otorgue una calificación de 7 o menos? P(X<7)**

Este problema puede resolverse utilizando la **función de distribución acumulativa**. Estandarizamos el valor **X = 7**:

$$Z = \frac{7-5}{2} = 1$$

Luego, buscamos el valor de **Z = 1** en una tabla de la distribución normal estándar, que nos da una probabilidad acumulada de **0.8413**. Esto significa que el **84.13%** de los encuestados otorgan una calificación de **7 o menos**.

# Distribución Normal Estándar  - Ejemplo

En una reciente **encuesta de opinión pública** sobre la **satisfacción con el gobierno** en una ciudad, se midió el grado de satisfacción de los encuestados en una escala continua del 0 al 100, donde:

- 0 representa "totalmente insatisfecho"
- 100 representa "totalmente satisfecho".

Supongamos que los resultados de la encuesta para una muestra de 1,000 personas se distribuyen aproximadamente de manera **normal** con una **media** $\mu$ de 65 y una **desviación estándar** $\sigma$ de 10.

Parámetros;

**Media**: $\mu=65$  (valor promedio de satisfacción de los encuestados)

**Desviación estándar**: $\sigma=10$ (variación respecto a la media)

**¿Cuál es la probabilidad de que un encuestado elegido al azar tenga una satisfacción mayor a 80?**

Para resolver esta pregunta, necesitamos calcular el área bajo la curva normal para $X>80$

# Distribución Normal Estándar - Ejemplo

Normalizamos el valor 80, es decir, lo transformamos en un valor Z estándar.

La fórmula de estandarización es: $Z = \frac{X - \mu}{\sigma}$

Por lo tanto, $Z = \frac{80 - 65}{10} = 1.5$

Luego, buscamos en la tabla de la **distribución normal estándar** la probabilidad de que Z > 1,5.

El valor de probabilidad asociado a Z=1,5 es aproximadamente 0,9319. Por lo tanto, la probabilidad de que Z > 1,5 es: P(Z>1,5) = 1- 0,9319 = 0,0681

Por lo tanto, hay un 6,68% de probabilidad de que un encuestado tenga una satisfacción mayor a 80.

# Distribución Normal Estándar

Acá tenemos la ejemplificación gráfica de cómo se

estandariza una distribución N(μ,σ2) hacia N(0,1) donde

el valor de x se transforma en un z que toma se obtiene a partir de

$Z= (x-\mu)/\sigma$

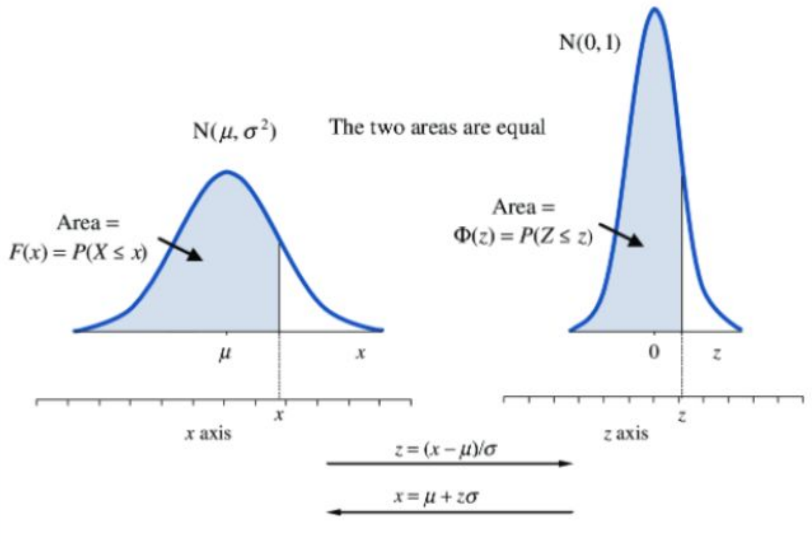


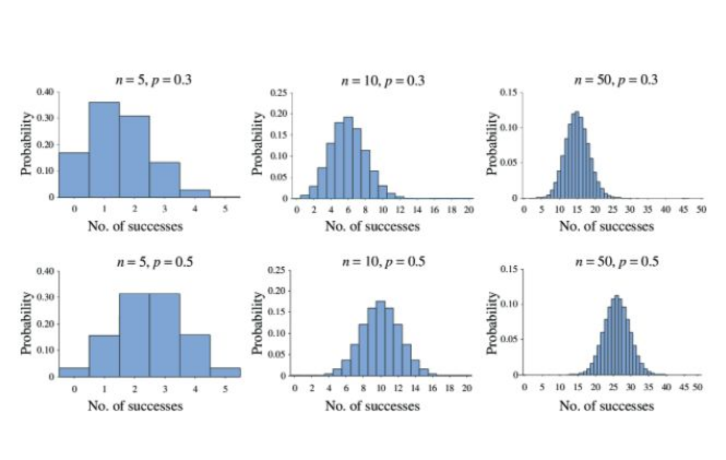

Figura 14.15 From N $(\mu, \sigma^2)$ to N(0, 1) and vice versa.

# Distribución normal - Teorema central del límite

Este teorema es uno de los pilares de la estadística porque permite utilizar la distribución normal para hacer inferencias sobre poblaciones, incluso si la población no sigue una distribución normal.

El Teorema central del límite establece que, dado un número suficientemente grande de muestras independientes extraídas de una población, la distribución de estadísticos de la muestra tiende a ser normal, independientemente de la forma de la distribución original de la población.

Nota: en este caso la variable cuya distribución nos interesa es la media muestral, o una proporción muestral, por ejemplo.

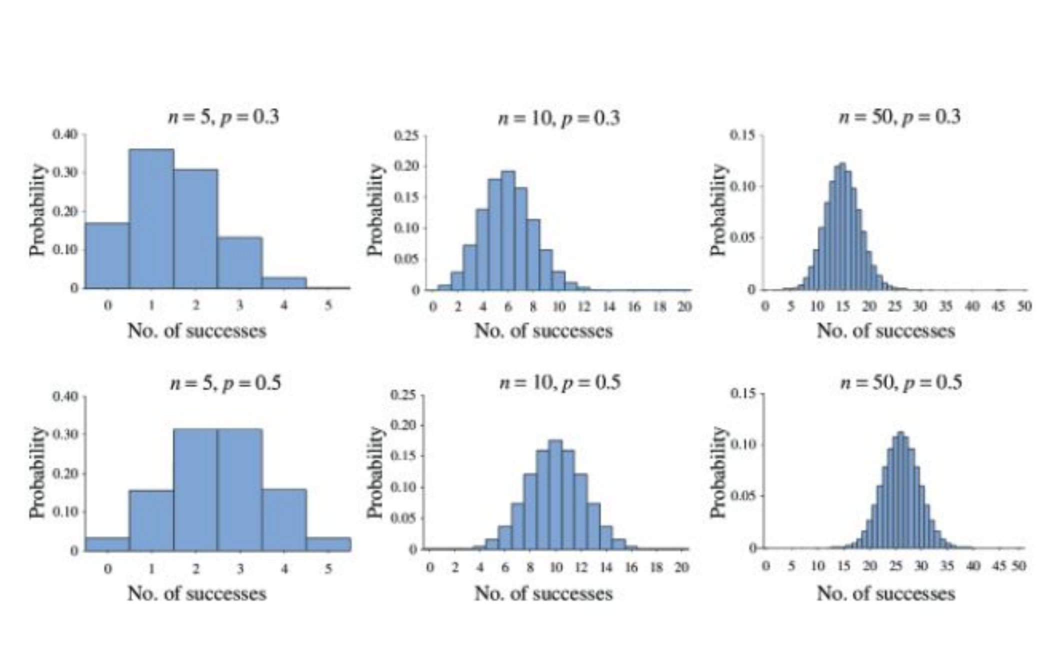


# Distribución Normal - Teorema central del límite

Vemos que, para el caso de distribuciones binomiales, la variable aleatoria proporción de votos a un candidato, por ejemplo (p=0.5), sigue una distribución normal cuando el n es lo suficientemente grande, es decir cuando se repite el experimento muchas veces muchas muestras). Cada barra indica el p obtenido de una muestra extraída al **azar.**

La lógica es similar para estadísticos como el promedio muestral. Se parte de la idea de que la distribución de la media muestral sigue una distribución normal.

.




**Para testear el teorema ustedes mismos:**
https://applied-science.com.ar/Estadistica/tcl.html

# Distribución Normal - Teorema central del límite

**Proporciones muestrales**

Vemos que, para el caso de distribuciones binomiales, la variable aleatoria **proporción** de votos a un candidato, por ejemplo (p=0.5), sigue una distribución normal cuando el n es lo suficientemente grande, es decir cuando se repite el experimento muchas veces muchas muestras). Cada barra indica el p obtenido de una muestra extraída al **azar.**

**La proporción muestral tenderá a seguir una distribución normal con:**

$$\mu_{\hat{p}} = p$$

$$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \longrightarrow$$ Desviación de la distribución muestral (error estándar)

.

# Distribución Normal - Teorema central del límite

**Promedios muestrales**

La lógica es similar para estadísticos como el promedio muestral. Se parte de la idea de que la distribución de la media muestral sigue una distribución normal. Si se repite el experimento múltiples veces, la distribución de las medias muestrales estará centrada en el parámetro real (la media poblacional), con una alta probabilidad de ocurrencia de valores cercanos a la misma.

La media muestral tiende a seguir una distribución normal con:

$$\mu_{\bar{X}} = \mu$$

. $$SE_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$ → Desviación de la distribución muestral (error estándar)

# Distribución Normal - Teorema central del límite

A partir de este teorema se puede:

-Construir intervalos de confianza

-Hacer tests de hipótesis

Veremos cómo hacer inferencia conjugando el TLC con lo que vimos sobre los puntajes Z (o como veremos, los puntajes T).

.

# Distribución Normal - Teorema central del límite

Requisitos para que se cumpla el TCL:

-**Muestreo aleatorio**: todas las muestras deben seleccionarse al azar para que tengan la misma posibilidad estadística de ser seleccionadas

-**Las muestras deben ser independientes**: las selecciones o los resultados de una muestra no deben tener influencia en muestras futuras ni en los resultados de otras muestras.

-**Tamaño de muestra grande**: a medida que aumenta el tamaño de la muestra, la distribución del muestreo debe acercarse cada vez más a la distribución normal. Si los promedios de cada muestra se alejan mucho del parámetro poblacional, también lo harán las distribuciones de estos estadísticos muestrales en su conjunto.

.