Resumen Probabilidad

Probabilidades en eventos

★ Unión de dos eventos

- Si **A** y **B** son dos sucesos no excluyentes (pueden ocurrir simultáneamente), la probabilidad de que ocurran **A o B** está dada por:

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

- Si **A** y **B** son dos sucesos excluyentes (no ocurren simultáneamente), la probabilidad de que ocurran **A o B** está dada por:

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

★ Intersección de dos eventos

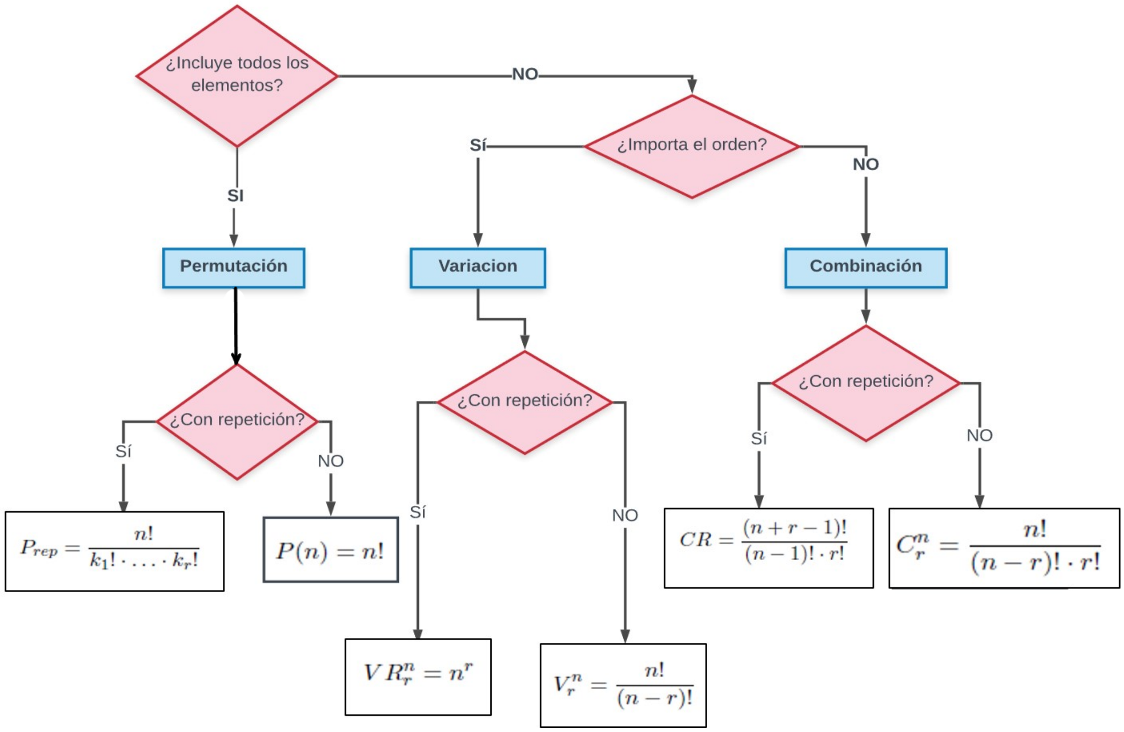
- Si **A** y **B** son sucesos **independientes** (la ocurrencia o no ocurrencia de uno **no influye** sobre la ocurrencia o no ocurrencia del otro). Luego la probabilidad de que ocurra **A y B** está dado por:

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

- Si **A** y **B** son sucesos **dependientes** (lo inverso de ser independiente), la probabilidad de que ocurra **A y B** está dado por:

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

Resumen de combinatoria



★ **Media:** $(\overline{x})$

$$\overline{x} = \frac{\sum_{i=0}^{n} x_i}{n} = \frac{\sum_{i=0}^{n} f_i \cdot x_i}{\sum_{i=0}^{n} f_i} = \frac{\sum_{i=0}^{n} f_i \cdot c_i}{\sum_{i=0}^{n} f_i}$$

★ **Esperanza:** $(E(x))$

$$E(x) = \sum_{i=1}^{n} x_i \cdot p_i$$

★ **Varianza:** (σ^2)

$$\sigma^2 = \sum_{i=1}^{n} (x_i - E(x))^2 \cdot p_i = \frac{\sum_{i=1}^{n} f_i \cdot (x_i - \overline{x}_i)^2}{\sum_{i=1}^{n} f_i} = E(x^2) - E(x)^2$$

★ **Desviación Estándar:** Raiz cuadrada de la varianza (σ).
Nota: σ y σ^2 son siempre positivos y además si a cada elemento del conjunto lo múltiplicamos por un natural K entonces σ y σ^2 aumentan K veces. ($Kx_i \Rightarrow K\sigma \Rightarrow K^2\sigma^2$)

- x_i: Datos ; c_i: Marca de clase (dato de al medio en un intervalo)
- f_i: Frecuencias ; p_i: Probabilidad de éxito

Distribuciones

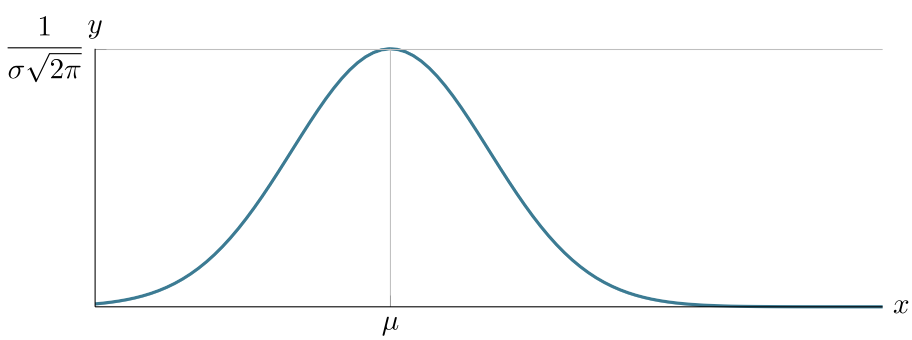
★ **Distribución binomial:** $X \sim B(n,p)$**:**

$$\boxed{P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}}$$

- n: Número de pruebas
- p_i: Probabilidad de éxito

★ **Distribución de probabilidad normal:** $X \sim N(\mu, \sigma) \sim N(\mu, \sigma)$**:**

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot exp\left(-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2\right)}$$



- *Dom*: $\mathbb{R}$, *Rec*: $\left]0, \frac{1}{\sigma\sqrt{2\pi}}\right]$
- Área igual al factor que múltiplica a $f(x)$. Media, mediana y moda coinciden
- Máximo valor del gráfico está dado por $\frac{1}{\sigma\sqrt{2\pi}}$
- σ más grande $\Rightarrow$ curva 'más ancha'
- Si $X \sim N(\mu, \sigma^2)$ y se extraen **n** observaciones de manera independiente en esta población, el promedio obtenido pertenece a una población normal de media μ y varianza $\frac{\sigma^2}{n}$

- Si $X \sim N(\mu, \sigma^2)$ no depende de $Y \sim N(\upsilon, \tau^2)$, entonces $X + Y \sim N(\mu + \upsilon, \sigma^2 + \tau^2)$

★ **Distribución normal estándar:** $Z \sim N(0, 1)$

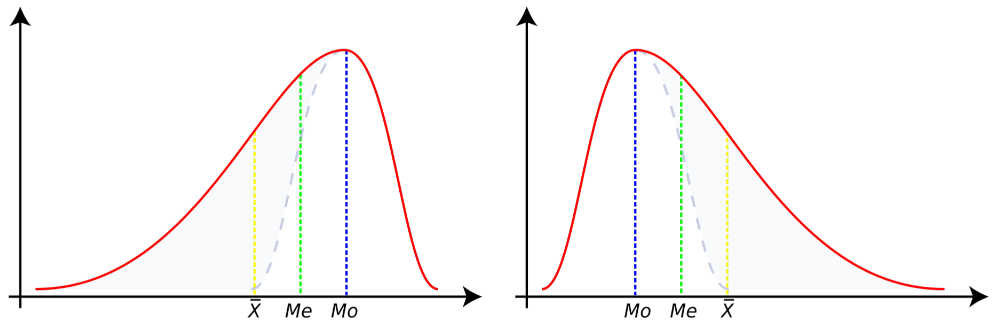
Para lograr una interpretación más simple de la distribución normal, se normaliza realizando el cambio de variable

$$\boxed{Z = \frac{X - \mu}{\sigma}}$$

Luego, la función probabilidad acumulada queda: $P(X \leq x) = P(Z \leq \frac{X-\mu}{\sigma})$

★ **Distribución normal asimétrica:**

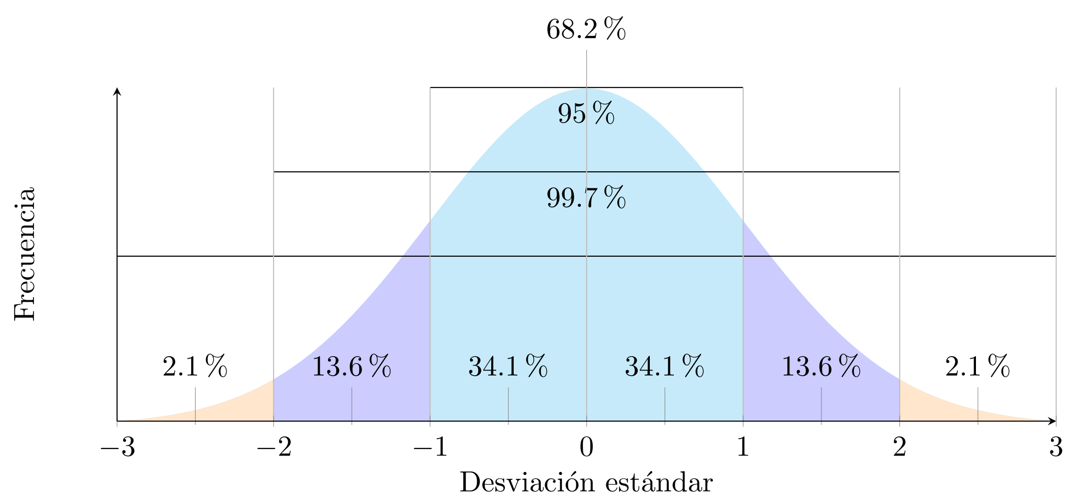
La simetría de la curva gaussiana depende de la distribución de los datos de la muestra.
Asimetria negativa: $\bar{x} < Me < Mo$
Asimetria positiva: $Mo < Me < \bar{x}$



Asimetría negativa　　Asimetría positiva

★ **Intervalos útiles en distribución normal:**

$$P(\mu - \sigma \leq x \leq \mu + \sigma) = 0,6826 = 68\,\%$$
$$P(\mu - 2\sigma \leq x \leq \mu + 2\sigma) = 0,9545 = 95\,\%$$
$$P(\mu - 3\sigma \leq x \leq \mu + 3\sigma) = 0,9973 = 99\,\%$$



★ **Distribución binomial aproximada a la Distribución normal**

- Para un N (cantidad de la población) lo suficientemente grande.
- $\boxed{\boxed{np > 5}} \wedge \boxed{\boxed{n \cdot (1-p) > 5}}$

$$\Rightarrow n \approx np \qquad \wedge \qquad \sigma \approx \sqrt{n \cdot p \cdot (p-1)}$$

$$\boxed{\Rightarrow B(n,p) \rightarrow N(np, \sqrt{n \cdot p \cdot (p-1)})}$$

★ **Media poblacional y media de muestras:**

$$\boxed{\mu = \frac{\displaystyle\sum_{i=0}^{n} \overline{x}_i}{k}}$$

- $\overline{x}_i$: medias muestrales de los subconjuntos
- N: tamaño de la población
- k: muestras de tamaño n, con n subconjuntos de N

★ **Error muestral:** $\boxed{e_i = \overline{x}_i - \mu}$

Propiedades:

- $\sum e_i = 0$
- $N \rightarrow \infty \Rightarrow e_i \rightarrow 0$
- $E(\overline{x}) = \mu$. "El valor de una muesra aleatoria corresponde al promedio poblacional"

★ **Teorema del limite central:**

Considerando $\overline{x}$ como una variable aleatoria y $N \rightarrow \infty \quad (N \geq 30)$ se tiene que *la distribución muestral se aproxima a una distribución normal*:

$$\boxed{\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)}$$

★ **Intervalo de confianza:**

Consideramos la media de la muestra como la media poblacional ($\overline{x} = \mu$). Recordamos que la distribución normal esta dado por $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ y con $z = \dfrac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$. Se le llamará intervalo de confianza de nivel $(1-\alpha)\cdot 100\,\%$ con $(\alpha \in \;]0,1[)$ a:

$$\boxed{\left[\overline{x} - z_{\left(1-\frac{\alpha}{2}\right)} \cdot \frac{\sigma}{\sqrt{n}} \leq x \leq \overline{x} + z_{\left(1-\frac{\alpha}{2}\right)} \cdot \frac{\sigma}{\sqrt{n}}\right]}$$

Propiedades:

- El nivel de confianza $(1-\alpha)$ corresponde a la probabilidad de que al tomar una muestra aleatoria, el promedio poblacional μ se encuentre dentro del intervalo de confianza.
- Los niveles de confianza $(1-\alpha)\cdot 100\,\%$ más usuales son: $90\,\%$, $95\,\%$ y $99\,\%$, que corresponden a niveles de significación (α) de $10\,\%$, $5\,\%$ y $1\,\%$, respectivamente. Con ello $z\left(1-\frac{\alpha}{2}\right)$ es $1,64$, $1,96$ y $2,98$ respectivamente.
- $\dfrac{\sigma}{\sqrt{n}}$ se denomina error estandar.
- $z_{\left(1-\frac{\alpha}{2}\right)} \cdot \dfrac{\sigma}{\sqrt{n}}$ es el margen de error.
- $2 \cdot \left(z_{\left(1-\frac{\alpha}{2}\right)} \cdot \dfrac{\sigma}{\sqrt{n}}\right)$ es la amplitud del intervalo.
- A mayor valor de **n** se da una amplitud menor, lo cual implica un intervalo más preciso.
- A mayor desviación estándar se da una mayor amplitud del intervalo de confianza. Esto significa que a una mayor variabilidad y manteniendo el mismo nivel de confianza, se pierde precisión en la estimación de la media poblacional μ.