VALIDACIÓN DEL MODELO

La validación del modelo matemático es una etapa crucial dentro del diseño experimental debido a que garantiza tanto la viabilidad como el dominio de operación del modelo que describe el fenómeno físico-químico de estudio.

El planteamiento de la Relación Cuantitativa Estructura - Propiedad (QSPR, por sus siglas en inglés) conlleva las siguientes prácticas:

Prácticas comunes:

1. **Preparación de los datos experimentales:** Incluye (i) la recolección y *limpieza* de los datos; (ii) cálculo de los decriptores químicos que contengan las propiedades objetivo; y (iii) unión de la propiedad y la respuesta en una base de datos de tipo SPR.
2. **Generación del modelo:** Implica establecer relaciones estadísticas entre las propiedades objetivo y la respuesta experimental.

   *Prácticas **no** comunes:*

3. **Validación del modelo:** Implica la evaluación cuantitativa de la solidez del modelo y su capacidad predictiva.
4. **Definición del dominio de aplicación:** Dominio matemático donde es viable emplear el modelo propuesto.

In [1]:

```
%store -r Modelo
%store -r Data_Val
%store -r Data_Plant
%store -r nombre_bd
%store -r Estab
%store -r Datos
%store -r archivo
```

Metodologías de validación

Existen diversas metodologías de validación, entre ellas:

- **LMO:** La premisa de esta metodología está en que si un modelo QSPR tiene una media alta q^2 durante el proceso de validación, se puede concluir que el modelo obtenido es robusto. Los n datos se dividen en G grupos de igual tamaño, $m_j(= n/G)$. Dependiendo del valor de n, G generalmente se selecciona entre 2 y 10. Un gran grupo de modelos son desarrollados con cada $n - m_j$ objetos en el conjunto de entrenamiento y m_j objetos en el conjunto de validación. Para cada modelo correspondiente, se predicen los objetos m_j y se computa q^2. Son deseables altos valores de q^2.
- **Bootstrapping:** Metodología de remuestreo que funciona con una muestra representativa de la población. Como sólo hay un grupo de datos, el bootstrapping simula qué ocurriría si las muestras se seleccionaran de manera aleatoria. En un procedimiento de validación típico, se generan k grupos aleatorios de tamaño n del grupo de datos original. Al igual que la validación LMO, es deseable un alto q^2 que demuestre la robustez del modelo.
- **Prueba de aleatoriedad - Y:** Técnica ampliamente usada para garantizar la robustez del modelo QSPR. En esta prueba, el vector de variables dependiente, Y, se cambia de manera aleatoria para la generación de un nuevo modelo QSPR usando la matriz original de variables independientes. El proceso se repite en varias ocasiones. Se espera que los modelos QSPR resultantes tengan bajos valores de R^2 y LOO (Leave-One-Out) q^2. Es probable, aunque infrecuente, que se obtengan altos valores q^2 debido a una correlación fortuita o una redundancia estructural del conjunto de datos. Si todos los modelos QSPR obtenidos por aleatoriedad - Y presentan altos valores de R^2 y q^2, implica que un modelo QSPR aceptable **no** puede ser obtenido por el grupo de datos dado.
- **Validación externa:**
  - *Selección de datos de planteamiento y validación:* En situaciones típicas, es difícil encontrar nuevos compuestos probados experimentalmente para este propósito. El recurso empleado es, entonces, dividir el conjunto de datos experimentales en datos para el plantemiento del modelo QSPR y datos para validación externa. El objetivo en esta etapa es garantizar que tanto los datos de validación como los de planteamiento ocupen el mismo dominio del fenómeno físico-químico de estudio. La partición de datos entre planteamiento y validación es un área de investigación activa. Es recomendable que el conjunto de validación externa contenga al menos 5 compuestos que describan el rango de actividad de los compuestos incluidos en el conjunto de planteamiento. Está demostrado que modelos QSPR desarrollados y validados mediante esta metodología tiene un alcance predictivo mayor que los mencionados anteriormente.
  - *Evaluación del poder predictivo de modelos QSPR:* En orden de estimar la verdadera capacidad predictiva del modelo QSPR, es necesario comparar las actividades predecibles y observadas de una base de datos lo suficientemente grande que no hubiese sido empleada para el desarrollo del modelo matemático. El poder predictivo del modelo QSPR se puede estimar mediante un q^2 externo definido mediante la Ecuación 1.

$$q^2_{ext} = 1 - \frac{\sum_{i=1}^{Pr}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{Pr}\left(y_i - \bar{y}_{tr}\right)^2}$$

Dónde y_i y $\hat{y}_i$ son los valores medido y predecido, respectivamente, de la variable dependiente $\bar{y}_{tr}$, que es el valor medio del conjunto de datos.

Se considera que el modelo es **predictivo** cuando se satisfacen las siguientes condiciones: $$ \begin{equation} \begin{array}

```
q q^2 > 0.5\\ \\
R ^2 > 0.6 \\ \\
\frac{\left(R^2 - R ^2 _0 \right)}{R^2} < 0.1 \\ \\
0.85 \leq k \leq 1.15
```

\end{array} \end{equation} $$

La metodología de validación empleada a continuación es la **prueba de aleatoriedad - Y**.

In [15]:

```
from App.Validation.AlY import AleatoriedadY
from App.Validation.Results import TablaAlY
AlY = AleatoriedadY(Modelo, Datos['Y'], Estab, nombre_bd)
ValData = AlY.Respuesta
TablaAlY(ValData, Modelo['Ecuación'])
```

R^2_{cal}	R^2_{val}	$RMSEC$	$RMSEP$	LOF	β_0	β_1	β_2	β_3	β_4	
0.981	0.932	0.542	2.936	9.337	2.794	1.083	1.075	2.344	4.047	2.0
0.983	0.685	0.544	12.208	41.483	2.786	1.101	1.045	2.306	4.031	2.0
0.968	0.997	0.802	0.198	4.054	2.742	1.039	1.046	2.292	3.972	2.0

Out[15]:

```
<App.Validation.Results.TablaAlY at 0x210a55b2080>
```

Definición del dominio de aplicación

No importa cuán robusto sea el modelo QSPR validado: *no es confiable pensar que los resultados del modelo apliquen para todo el universo químico.*

In [ ]:

```

```

Guardar

Los resultados, tanto los del desarrollo del modelo matemático como los de validación, se guardarán y los podrás encontrar en la dirección: App/DataBase/Resultados

In [16]:

```
from App.DataBase.Guardar import Save
Save(Modelo, ValData, Datos['Y'], archivo)
```

Out[16]:

```
<App.DataBase.Guardar.Save at 0x210a55bcac8>
```