

---

# Trabajo Práctico N° 2

Análisis Exploratorio de Datos - Maestría en Estadística Aplicada

Año 2020

---

## Problema 1

La Cinética Química explora las leyes que rigen el cambio de la composición de un sistema en el tiempo y su relación con las variables que definen su estado, en particular, con la presión, la temperatura y la composición. En un estudio sobre un proceso de degradación química se determinó durante 25 días la concentración del principio activo de una solución medicamentosa en cuya etiqueta se informa que la misma es de 1.5 %. No obstante, el experimento comenzó algún tiempo después de la elaboración del medicamento, obteniéndose en la primera medición (día 0) un valor igual a 1.240 %. Los datos se encuentran en el archivo *cc.txt*.

1. Explorar este conjunto de datos con las técnicas bivariadas discutidas en la unidad, intentando caracterizar el comportamiento de la concentración del principio activo en función del tiempo.
2. En el ámbito de la cinética química existe una clasificación para los procesos de degradación, que determina el orden de la reacción. Se dice que el proceso de degradación será de:
  - **orden 0**, si la ecuación que describe dicho proceso es  $C_t = C_0 - kt$ . En este caso, la representación gráfica de la concentración ( $C$ ) en función del tiempo ( $t$ ) sería una recta en la que el opuesto de la *constante de velocidad*  $k$  es la pendiente.
  - **orden 1**, si la ecuación que describe dicho proceso es  $C_t = C_0 e^{-kt}$ , es decir la concentración decrece exponencialmente con el tiempo. En este caso, la representación gráfica del  $\ln(C)$  en función de  $t$  sería una recta de pendiente  $-k$ .
  - **orden 2**, si la ecuación que describe el proceso es  $C_t = \frac{C_0}{1+C_0 kt}$ . En este caso, la representación gráfica de  $1/C$  en función de  $t$  sería una recta.
- a. Para cada orden de reacción, transformar correspondientemente las observaciones, representar gráficamente y ajustar el modelo lineal implícito en cada caso. Determinar, mediante el análisis de los residuos, el orden de este proceso de degradación.
- b. A partir del modelo elegido en el punto anterior, estimar de manera puntual:
  - I. El valor de la constante de velocidad  $k$ .
  - II. El tiempo transcurrido desde la fecha de elaboración del medicamento al día de inicio del experimento, suponiendo que la reacción de degradación sufre el mismo proceso todo el tiempo.
  - III. El tiempo de vida media del medicamento, definido como el tiempo al que la concentración del principio activo es la mitad de lo declarado en la etiqueta.

## Problema 2

Jalali-Heravi y Knouz (2002) midieron en 32 compuestos químicos una propiedad física llamada *punto de Krafft* o *temperatura Krafft*, que es la temperatura a la cual la solubilidad de los agentes de superficie iónicos aumenta bruscamente, junto con algunas variables que cumplen el rol de ser descriptores moleculares, con el fin de encontrar una ecuación predictiva del punto de Krafft. En este problema usaremos solamente el predictor *heat* (calor o entalpía de formación del compuesto). Los datos se encuentran en el archivo *punto\_krafft.txt*.

1. Explorar este conjunto de datos con las técnicas bivariadas discutidas en la unidad, intentando caracterizar el comportamiento del punto de Krafft en función de la entalpía de formación.
2. Como ya se ha mencionado, existen diversos métodos de estimación robusta. Uno es el bicuadrado. Otro que se puede emplear en problemas de regresión es el método MM (una extensión del método de estimación robusta M). Buscar literatura y hacer un muy breve resumen de qué se trata este método. Tener en cuenta que esto corresponde a un tema con amplio desarrollo teórico y con este punto sólo se persigue el objetivo de tener una noción acerca de lo que propone el método. Pueden utilizar cualquier material que encuentren; las primeras secciones de las siguientes publicaciones pueden ser útiles: [link 1](#), [link 2](#).
3. Ajustar una regresión lineal simple mediante el método de mínimos cuadrados y dos versiones robustas empleando el método bicuadrado y el estimador MM. Describir los resultados obtenidos y comparar gráficamente los tres ajustes.
4. Este ejemplo es presentado por Maronna (Robust Statistics: Theory and Methods, 1era Edición, 2006, página 146) con el objetivo de comparar el ajuste clásico con el ajuste robusto con estimador MM. Resumir el argumento del autor en favor del resultado logrado por el ajuste MM y relacionarlo con la *exact fit property* mencionada en la misma sección. ¿Por qué el autor prefiere un ajuste en el cual los gráficos de residuos se ven “peor”? ¿A cuál de estos dos ajustes se asemeja más el del método bicuadrado?

Para el ajuste del estimador MM en R, emplear el paquete `robust` y su función `lmRob`, con todas sus opciones por default, para poder reproducir el ejemplo del libro. Es decir, ajustar el modelo con: `lmRob(y ~ x, data = datos)`. Tener en cuenta hay una gran variedad de parámetros y funciones que se pueden elegir para controlar el comportamiento de este estimador robusto, probar con ellas en este mismo paquete o en otros puede arrojar ajustes que no necesariamente reproduzcan el ejemplo mencionado.