

Actividad 1: Entorno de análisis y programación estadística R

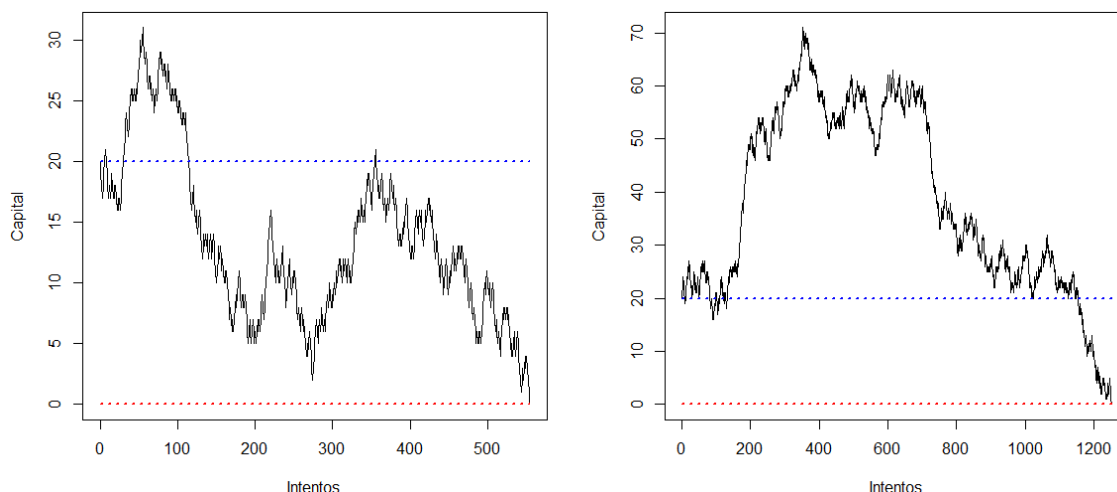
A continuación se proponen cinco ejercicios a resolver utilizando R. Los dos primeros son ejercicios de programación y el resto son aplicaciones con datos de técnicas estadísticas elementales. El estudiante deberá resolver las cuestiones planteadas en cada uno de ellos y con las soluciones, debidamente explicadas y detalladas, elaborar un documento o informe en formato pdf que someterá a través del enlace habilitado en PRADO.

No es necesario adjuntar ningún fichero adicional, el documento pdf debe recoger todos los detalles de la solución: planteamiento del problema, código en R, resultados obtenidos e interpretación y comentarios oportunos.

La actividad será evaluada atendiendo a los siguientes criterios: (i) soluciones correctas y completas, (ii) programación eficiente y bien estructurada, (iii) comentarios sobre el método de resolución seguido y descripción adecuada de los resultados y (iv) buena presentación del trabajo.

Ejercicio 1. Imaginemos la siguiente situación: un jugador tiene un capital de 20 euros y cada instante de tiempo apuesta un euro al lanzamiento de una moneda, ganando un euro si sale cara y perdiendo el euro apostado si sale cruz. Consideramos la variable aleatoria discreta T que nos indica el instante de tiempo en que el jugador se arruina por primera vez (obsérvese que esta variable en principio podría tomar un valor infinito si jamás se llegara a la ruina). Crear una función en R que permita simular y representar gráficamente la “ruina del jugador”. En la creación de la función tendremos en cuenta lo siguiente:

- Tendrá un único argumento que es el capital inicial. En el caso descrito antes es de 20 euros aunque admitiremos que este puede ser otro valor.
- La moneda se supone no trucada (probabilidad de cara y cruz es la misma, $\frac{1}{2}$).
- El juego termina en el instante en que el capital de que dispone el jugador es 0 (ruina).
- La función debe devolver los resultados en un gráfico cuyo formato y apariencia será similar al de los ejemplos mostrados a continuación:



Estos gráficos representan dos realizaciones del experimento. En el gráfico de la izquierda el jugador se arruina en el instante 554 y en de la derecha en el instante 1262.

Ejercicio 2. Crear una función con nombre “dif.eq” que devuelva el vector (x_1, x_2, \dots, x_n) donde

$$x_n = r * x_{n-1} * (1 - x_{n-1})$$

La función tendrá tres argumentos: “x1”, que corresponde al primer elemento x_1 , el coeficiente “r” y el número de iteraciones en la serie “n”.

- Ejecutar la función para $r = 2$ y $0 < x_1 < 1$. Se debería obtener que x_n tiende a 0.5 cuando n tiende a infinito.

- Representar gráficamente la serie obtenida con $x_1 = 0.95$, $r = 2.99$ y $n = 500$.
- Escribir una segunda función “dif.eq2” con argumentos “x1” y “r” que devuelva el número de iteraciones (n) necesarias para alcanzar el criterio de convergencia siguiente: $|x_n - x_{n-1}| < 0.02$. Para $x_1=0.95$ y $r=2.99$ la respuesta es 84.

Ejercicio 3. El conjunto de datos **cars**, contenido en el libro **datasets** de R, recoge información acerca de la velocidad de los coches (variable speed) y distancias necesarias para parar (variable dist). Utilizando dichos datos realizar un análisis descriptivo básico de las variables contenidas en él, ofreciendo la siguiente información:

- Para cada variable: medidas de posición central, dispersión, asimetría y curtosis; gráficos exploratorios adecuados para el tipo de datos.
- Considerando conjuntamente las dos variables speed y dist: construir un diagrama de dispersión, calcular la recta de regresión que permita explicar dist en función de speed y superponerla en el diagrama de dispersión. Cuantificar la bondad del ajuste lineal a través del coeficiente de determinación.

Ejercicio 4. Consideremos los siguientes datos del calor latente en la fusión del hielo expresados en cal/gm:

Muestra A: 79.98 80.04 80.02 80.04 80.03 80.03 80.04 79.97 80.05 80.03 80.02 80.00 80.02
Muestra B: 80.02 79.94 79.98 79.97 79.97 80.03 79.95 79.97

- Construir un gráfico de cajas y un gráfico cuantil-cuantil para comparar la distribución de cada muestra.
- Contrastar la igualdad de varianzas de las distribuciones de dichas muestras.
- Contrastar la igualdad de medias de las distribuciones de dichas muestras.

Ejercicio 5. Los siguientes datos corresponden a un estudio sobre el consumo de combustible de cuatro fabricantes diferentes de vehículos de motor 1600cc.

Fabricante	Consumo (litros/100km)								
A	9.3	8.9	8.7	9.1	8.6	9.1	9.6	9.7	9.0
B	9.1	8.3	8.2	9.0	9.1	8.3	8.8	8.0	9.0
C	9.8	10.4	9.2	9.7	10.4	8.9	8.8	9.7	9.5
D	8.6	8.9	8.1	7.8	9.1	8.0	8.4	8.1	8.8

En la tabla se recogen datos de consumo de combustible (litros/100km) para muestras independientes de 9 vehículos nuevos de cada fabricantes, después de recorrer la misma ruta en idénticas condiciones y con el mismo conductor.

- Construir un gráfico exploratorio de cajas que permita la comparación de los cuatro fabricantes.
- Realizar un contraste de hipótesis apropiado que permita concluir si hay o no diferencia significativa entre los fabricantes respecto al consumo de combustible.
- Justificar las hipótesis asumidas en el contraste de hipótesis realizado en el apartado b).