# 5 puntos: Dataset del Titanic (dificultad baja)

|  |  |
| --- | --- |
|  | El dataset train.csv contiene datos del accidente que sufrió el Titanic en la noche del 14 de abril de 1912.  Las variables que contiene son las siguientes:   * Passengerid: Nº identificador de cada pasajero. * Survived: 1 o 0 para indicar si sobrevivió o no. * Pclass: A qué categoría pertenecía el pasajero 1ª a 3ª * Name * Sex * Age * Fare: Precio del billete en dólares. * Cabin: Cabina en la que se alojaba el pasajero. * Embarked: En qué puerto embarcó.   Utiliza estos datos para contestar a las siguientes preguntas: |

Ejercicio 1

* Carga el set de datos en una variable llamada "titanic"
* Obtén el número de filas y columnas incluidas en el set de datos.
* Obtén las primeras 10 observaciones para ver qué pinta tienen.
* Obtén el listado de variables incluidas en el set de datos y sus tipos.
* Obtén la distribución de las variables categóricas (sex, embarked, survived, pclass) del set de datos. Es decir, valores únicos / valores totales.
* Obtén los estadísticos básicos de las variables del set de datos.

Ejercicio 2

* Indica en qué variables hay valores con NA, NULL o vacíos "" y cuál es su proporción dentro de la variable.
* Imputa la media del valor de la variable en el caso de las numéricas y "No disponible" en el caso de string.
* Vuelve a comprobar la existencia de NAs y NULL para verificar que los has eliminado.

Ejercicio 3

* ¿Fallecieron más mujeres u hombres? (Cuánto en porcentaje sobre el total de su género).
* ¿Qué clase de pasajeros (primera, segunda o tercera) sobrevivió más? Quiero que el resultado de, únicamente, el porcentaje y la clase que más sobrevivió (no el porcentaje de las 3 clases).
* ¿Cuál fue la edad media y máxima de los supervivientes en cada una de las clases?

Ejercicio 4

* ¿Cuál de los puertos de embarque es el que tiene la media de precio de billete más barata? Indica solo ese puerto, no una lista con todos.
* ¿Qué correlación hay entre la longitud del nombre de un pasajero y el importe de su billete? No modifiques el nombre del pasajero para hacer el cálculo (úsalo tal y como venga).

Ejercicio 5

* Obtén los nombres de los pasajeros que no sobrevivieron y el precio de su billete está en el decil superior.
* ¿En qué cabina deberías alojarte para tener una mayor probabilidad de sobrevivir siendo hombre de entre 30 y 40 años? Indica una única cabina.

# 3 puntos: Mi primer algoritmo de inversión (dificultad moderada)

|  |  |
| --- | --- |
|  | Ibex\_data.csv contiene las cotizaciones de las empresas del Ibex 35 desde el 2004, hasta el 2019, en un dataset limpio y libre de sesgos.  Ten presente que la composición del índice no siempre ha sido la misma. Este dataset refleja la composición y cotización del Ibex, a lo largo del tiempo. Si una empresa ha entrado, salido, y vuelto a entrar en el índice, se le añade un sufijo. Por ejemplo: NHH\_1 |
|  | price\_departure.csv contiene un indicador que utilizo en diversos algoritmos de inversión. Lo necesitarás, como referencia en los ejercicios 7 y 8. |

Ejercicio 6

* Desarrolla un algoritmo de mechas

Para cada activo, cada día, compra a precio de apertura y vende cuando ocurra el primero de los siguientes eventos:

El activo sube 3 céntimos (stop profit)

El activo cae 10 céntimos (stop loss)

Si no ocurre ninguno de los anteriores, vende a precio de cierre

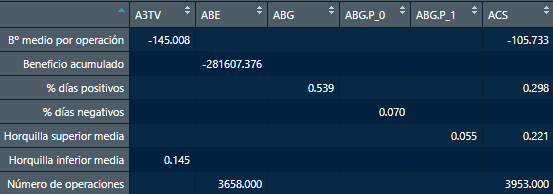
Ojo, habrá días positivos y negativos a la vez, en estos casos, supón que toca primero el stop loss

El capital que invertimos en cada activo, cada día, debe ser 30.000 €

La comisión de compra será de 0.0003 \* capital. Lo mismo para la venta.

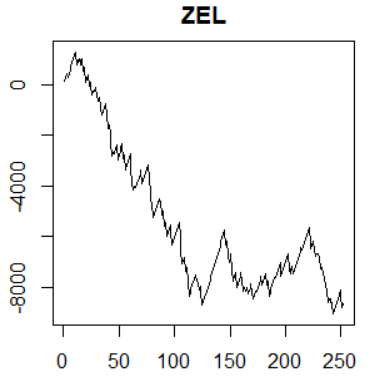
Comprueba que tienes al menos 30 datos para hacer los cálculos (si no es así descarta el activo)

Entregable: Código que genere un dataframe con la siguiente estructura (para todos los activos):



Horquilla superior media: (max – open) y horquilla inferior media (open – low)

Y grafique el beneficio acumulado por activo. Por ejemplo:



Ejercicio 7

* Partiendo del algoritmo de mechas anterior, añade el parámetro price\_departure.

Para cada activo, cada día, si el price\_departure es >= 0.75, compra a precio de apertura y vende cuando ocurra el primer de los siguientes eventos:

El activo sube 3 céntimos (stop profit)

El activo cae 10 céntimos (stop loss)

Si no ocurre ninguno de los anteriores, vende a precio de cierre

Ojo, habrá días positivos y negativos a la vez, en estos casos, supón que toca primero el stop loss

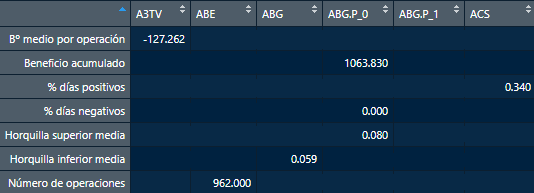
El capital que invertimos en cada activo, cada día, debe ser 30.000 €

La comisión de cada compra y venta será de 0.0003 \* capital

Homogeneiza los datos Ibex\_data y price\_departure (utiliza solo las fechas que existan en ambos DF)

Comprueba que tienes al menos 30 datos para hacer los cálculos, antes de aplicar el filtro del price\_departure (si no es así descarta el activo)

Entregable: Código que genere el mismo dataframe que en el ejercicio anterior y los mismos gráficos.



Ejercicio 8

* Optimización de la mecha y el capital por activo

Utilizar un stop profit, o un stop loss estático, no parece lo más eficiente.

Tampoco lo parece el utilizar el mismo capital para todos los activos.

Objetivo: Partiendo del algoritmo de mechas anterior

* + Modifica el capital asignado a cada activo: usa la media de datos de cierre y el 0,5% del volumen.
  + Modifica el stop profit de cada activo: utiliza el cuantil 30 de la mecha superior (max – open)
  + Modifica el stop loss de cada activo: utiliza el cuantil 80 de la mecha inferior (open – low)

Entregable: Código que genere un dataframe con la siguiente estructura (para todos los activos), y los mismos gráficos.



# 2 puntos: Más allá de lo visto en clase (dificultad elevada)

Ejercicio 9

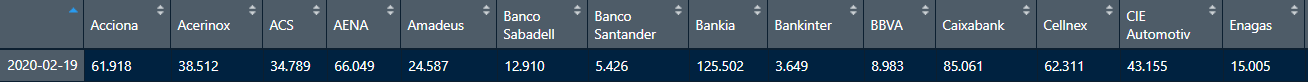
* Obtén de la CNMV la suma de las posiciones largas, para cada empresa, del listado de empresas que se indica más abajo.

Ten cuidado, la url cambia cada día, por lo que deberás solventar este problema.



lista\_empresas = empresas pertenecientes al Ibex35 a la fecha de entrega del enunciado.

El resultado del ejercicio deberá ser similar al siguiente (es un ejemplo, alguna empresa puede no estar ya en el Ibex)



* Algunas empresas suman más del 100% en posiciones largas. Explica cómo es posible

Ejercicio 10

**Diseño experimental – poniendo a prueba el euromillón**

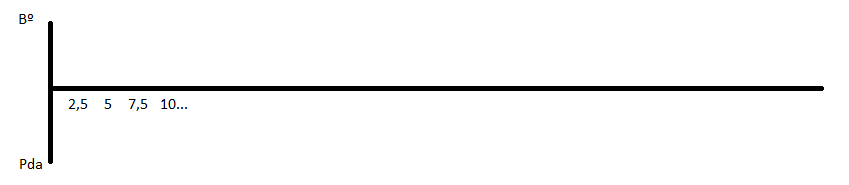
Evidentemente no vamos a ganar el euromillón con este programa, dado que es imposible predecir los números. Pero sí aprenderéis mucho acerca de la programación y el diseño experimental.

El objetivo del ejercicio es comprobar si existe algún punto óptimo en el que, invirtiendo una cantidad determinada, la esperanza matemática es mayor al resto de inversiones. Es decir, ¿es lo mismo invertir 2,5€ que 250€?, ¿qué esperanza matemática tiene cada uno?, ¿existe una inversión de dinero “óptima”?

El objetivo es que diseñes un experimento para poner a prueba la hipótesis anterior. Deberás detallar con todo lujo de detalles el experimento, ya sea en un Word, o en la parte inicial del script.

Como mínimo, el experimento deberá detallar las siguientes cuestiones:

* ¿Cómo generas los números aleatorios? Vas a generar miles de millones de números. ¿Cómo solventas el problema de la pseudo aleatoriedad vs tiempo de ejecución? Detalla el planteamiento tanto como puedas.
* ¿Cuántas combinaciones hay que sacar para cada importe? Este punto es importante. Para cada importe, por ejemplo 10€, ¿cuántas apuestas vas a realizar para calcular la esperanza matemática?, ¿porqué ese número de apuestas y no otro?, ¿cuántas veces repites el experimento para cada importe?, ¿por qué?
* El experimento empieza en 2,5 €. ¿cómo vas a plantearte seguir y hasta qué importe? Ir de 2,5€ en 2,5€ no parece la mejor aproximación. Sin embargo, intervalos relativamente grandes pueden hacer que no encuentres la inversión óptima. Detalla tanto como puedas la aproximación que vas a realizar, y por qué.
* ¿Cómo tienes en cuenta el reparto de premios, la evolución del bote o si hay más de 1 acertante? El importe de los premios no es estático. Para calcular la esperanza matemática debes tener en cuenta todos los premios que te pueden tocar, no solo el de mayor importe. Debes intentar ser lo más realista posible, especialmente en este apartado. Detalla qué aproximación vas a seguir y por qué.
* ¿Estás teniendo en cuenta las apuestas combinadas? <https://www.euromillones.com.es/tipos-de-apuestas-euromillones.html> no solo es interesante, también parece importante.
* ¿Cuánto tarda el programa? No me vale una aproximación teórica. El código debe estar desarrollado y ejecutado por tu parte. Todo lo anterior no vale de nada si el diseño no está llevado a término.
* ¿Dónde está el punto óptimo?, ¿es positivo o negativo? Dibuja el gráfico de la esperanza matemática para cada importe (según la evolución y límite máximo que hayas determinado).



# Entregables y normas aplicables

* La práctica se realizará de manera individual.
* Se deberá entregar el script de R. Código suelto o en cualquier formato que no sea un script de R no será evaluado.
* La entrega se realizará en la plataforma de Instituto BME antes de que acabe el plazo establecido. Cualquier entrega fuera de la plataforma no será evaluada. Recomendación: no esperes a la última hora para entregar la práctica (si internet falla en el último momento echarás a perder 3 semanas de trabajo).
* El nombre del fichero deberá ser nombre\_1erapellido.R (no hacerlo restará un punto en la evaluación).
* Los ejercicios deberán estar claramente diferenciados por su numeración y se incluirá el enunciado en la solución entregada. No hacerlo restará 1 punto en la evaluación.

Recomendación: Aunque disponéis de 3 semanas de plazo, no os pongáis a trabajar dentro de 2 semanas porque no os dará tiempo.

Advertencia: Si observo dos códigos muy similares, ambos alumnos defenderán su trabajo en un examen oral, para determinar la autoría de los mismos.