Estadística I
Grado en Matemáticas, UAM, 2014-2015
Trabajo computacional

- Para realizar este trabajo debéis organizaros en grupos (como máximo de tres personas). Se comunicará a `pablo.fernandez@uam.es` la composición de los grupos antes del **10 de diciembre de 2014**.
- **Entregables**. Se deberán enviar, electrónicamente,
  - las hojas de cálculo creadas en el trabajo;
  - y una (breve) memoria explicativa (en pdf). En la memoria se recogerán los *resultados* obtenidos en cada ejercicio, gráficas ilustrativas, y algunos comentarios y conclusiones.
- La **fecha límite de entrega** de los trabajos es el **12 de enero de 2015**. Aunque, por supuesto, se puede enviar en cualquier momento anterior.
- Recordad que las dos últimas sesiones del curso (17 y 18 de diciembre) estarán dedicadas a tutorías. Podéis usarlas para consultar cualquier duda que tengáis.
- Véanse algunos comentarios más al final del examen.

Ejercicio 1. En la hoja de cálculo adjunta encontrarás las series históricas de los últimos 10 años de cotizaciones diarias de las acciones de tres compañías (Telefónica, Banco de Santander y Repsol), además de las del índice IBEX 35. Junto a ellas se han calculado los correspondientes rendimientos (variaciones porcentuales) diarias.

En el ejercicio solo trabajaremos con estas series de rendimientos diarios.

1a) Toma las series de rendimientos y

- calcula las respectivas medias y cuasidesviaciones típicas muestrales, además de los rendimientos mínimo y máximo en el periodo de estudio;
- calcula las correlaciones muestrales entre cada pareja de series;
- construye un histograma para cada una de las series. Por ejemplo, que contenga 30 clases (equiespaciadas) entre el mínimo y el máximo rendimiento. Observa el aspecto que tienen esos histogramas.
- Toma la pareja de series de Repsol e IBEX, y calcula la recta de regresión, indicando la bondad del ajuste. Acompáñalo con un gráfico de la nube de puntos (gráfico de dispersión en excel).

1b) Vamos a suponer que los rendimientos de TFN son muestras independientes de normales. Calcula, a partir de la muestra dada,

- un intervalo de confianza con $\alpha = 5\,\%$ para la media;
- un intervalo de confianza con $\alpha = 5\,\%$ para la varianza;

1c) De nuevo suponiendo normalidad, para la serie de SAN,

- contrasta la hipótesis $H_0 : \mu = 0$ con nivel de significación $\alpha = 1\,\%$;
- halla el p-valor.

1d) Suponiendo normalidad una vez más, contrasta la hipótesis

$$H_0 : \mu_{\text{TFN}} = \mu_{\text{SAN}}$$

con nivel de significación del $5\,\%$. Halla el p-valor.

1e) Dados unos datos $(x_1, \ldots, x_n)$, ordenados de menor a mayor, llamamos "percentil-5 %" al dato que deja, aproximadamente, un 5 % de los datos a su izquierda, y el 95 % restante a la derecha.

Toma la serie de rendimientos de TFN y determina el percentil-5 %.

Planteamos ahora un esquema de simulación. Vamos a suponer que los rendimientos de TFN siguen una normal con los parámetros que hayas ajustado en el apartado b). Sortea ahora, por ejemplo, 1000 muestras independientes de esta normal y calcula el percentil-5 %. Repite esto un buen número de veces, para obtener muchas muestras de esos percentiles. A la vista de esos datos, calcula (numéricamente) un intervalo de confianza (al 1 %) para el percentil-5 %.

Ejercicio 2. Consideramos una variable aleatoria X que sigue una distribución de Rayleigh dada por

$$f(x;\theta) = \frac{x}{\theta^2}\, e^{-x^2/(2\theta^2)} \quad \text{para } x \geq 0,$$

donde θ es un parámetro positivo. Recuerda todos los resultados sobre esta distribución que hemos ido planteando en las hojas de ejercicios del curso.

Como estimador del parámetro θ usaremos $T(X_1,\ldots,X_n) = \sqrt{\frac{1}{2}\,\overline{X^2_{(n)}}}$.

1a) Escribe explícitamente las fórmulas de la función de distribución $F(x;\theta)$ de la variable, y de su inversa $F^{-1}(u;\theta)$. Dibuja las gráficas de la función de densidad y de la función de distribución para algunos valores del parámetro θ (por ejemplo, $\theta = 1/2, 1, 2, 10$).

1b) En la hoja de cálculo adjunta encontrarás 3000 muestras de una distribución de Rayleigh para un cierto valor (desconocido, al menos para ti) del parámetro θ. Halla una estimación $\widehat{\theta}$ de ese valor y calcula un intervalo de confianza al 5 %. Usa esa estimación para calcular $\mathbf{P}(3 \leq X \leq 5)$, donde X sigue una Rayleigh con parámetro $\widehat{\theta}$.

1c) Este apartado pretende "comprobar", por simulación, el resultado de normalidad asintótica

$$\sqrt{n}\left(\sqrt{\tfrac{1}{2}\,\overline{X^2_{(n)}}} - \theta\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0, \theta^2/4).$$

Fijamos $\theta = 3/2$. Usa la expresión explícita de F^{-1} obtenida en el apartado a) para obtener, por ejemplo, 500 muestras de la variable. Calcula, sobre esa muestra,

$$\left(\frac{\sqrt{\frac{1}{2}\,\overline{x^2}} - \theta}{\frac{\theta/2}{\sqrt{n}}}\right)$$

Repite el experimento un buen número de veces, dibuja el correspondiente histograma y comprueba (gráficamente) si se ajusta a una normal estándar.

Notas y comentarios:

- Se colgará en la red un pequeño manual de excel, por si fuera de utilidad. El propio excel cuenta con una ayuda para cada función.
- Podéis crear varias hojas de cálculo para los ejercicios. Parte del examen es que estas hojas de cálculo estén bien organizadas y sea sencillo seguir la información contenida en ellas.
- Algunas funciones de excel que pueden resultar útiles:
  - La función `promedio` calcula la media aritmética de un rango de datos;
  - La función `desvest` calcula la cuasidesviación típica de un rango de datos;
  - La función `coef.de.correl` calcula la correlación muestral entre dos rangos de datos;
  - Funciones `min` y `max`.
  - La función `contar` y `contar.si` cuentan cuántos elementos incluye un cierto rango de datos; la segunda, solo aquellos que cumplan un cierto criterio (ver manual).
  - Las funciones `k.esimo.mayor` o `k.esimo.menor` pueden ser útiles para ordenaciones automáticas de datos.
  - `aleatorio()` sortea una muestra de la uniforme en $[0, 1]$;