Práctica 1

---

1. Utilizar la función `runif` para simular números al azar entre 0 y 1.

   a) Simular una muestra de 10 números al azar y obtener su suma.

   b) Simular otra muestra de 10 números al azar y obtener su suma.

   c) Comparando los ítems anteriores ¿Qué se observa?

2. El experimento consiste en tirar una moneda equilibrada y observar si sale cara.

   a) Guardar en un vector los resultados de $m = 100$ simulaciones del tiro de una moneda.

   b) Estimar la probabilidad de que salga cara al arrojar la moneda, calculando la frecuencia relativa sobre los 100 resultados simulados en el ítem anterior.

   c) Repetir los ítems a y b $n = 1000$ veces, guardando en un vector r las 1000 estimaciones para la probabilidad de que salga cara.

   d) Graficar los resultados guardados en el vector r, junto con una recta horizontal sobre el verdadero valor de la probabilidad que se busca estimar. ¿Qué observa?

3. El archivo `Debernardi.cvs` contiene los datos referentes a un estudio acerca del cancer de páncreas (más información en el archivo *Acerca de los datos*, en el campus).

   a) Cargar los datos

   b) Cosntruir una tabla con los valores observados para la variable *diagnosis* y su frecuencia relativa.

   c) Realizar un gráfico de barras (estimación de la función de probabilidad) y graficar la función empírica.

4. El archivo `Islander_data.cvs` contiene los datos referentes a un estudio acerca de un experimento en los efectos de un medicamento contra la ansiedad en un test de memoria cuando se expone a la persona ante recuerdos felices y tristes. Se quiere estudiar la diferencia (en segundos) entre el tiempo logrado en un test de memoria antes y después de tomar el medicamento (más información en el archivo *Acerca de los datos*, en el campus).

   a) Cargar los datos

   b) Realizar un histograma para la variable X :*diferencia de tiempos.*

   c) Graficar la función empírica para X.

   d) Estimar la densidad de X usando estimadores basados en núcleos, utilizando diferentes tamaños de ventana. ¿Qué observa?

5. *Simulación.* Para las diferentes distribuciones correspondientes a la variable aleatoria X, se pide simular una muestra de tamaño $n = 1000$, realizar un gráfico de barras o estimación de la densidad según corresponda, graficar la función empírica y realizar un QQ-norm.

   a) $X \sim \mathcal{N}(0, 1)$

   b) $X \sim \mathcal{B}(10, 0{,}4)$

   c) $X \sim \chi^2_5 0$

   d) $X \sim \Gamma(10, 0{,}7)$

6. A partir de los datos del archivo `datos1.txt`, se pide efectuar un histograma con los siguiente criterios:

   a) Con el esquema por defecto de R.

   b) Con 1 corte.

   c) Con 10 cortes.

   d) Con 100 cortes.

   e) Repetir los incisos anteriores cambiando el punto de inicio del gráfico.

   ¿Qué se puede observar de este análisis? Comparar con la información en `datos2.txt`. Realizar un histograma y un boxplot, diagnosticar si presenta algún problema y tratar de solucionarlo.

7. En el archivo `ciudades.txt` se observan los valores de población, en cientos de miles, de las 10 ciudades más pobladas de 4 países en 1967.

   a) Construir un boxplot paralelo con los cuatro grupos de datos y determinar puntos extremos en cada uno de ellos.

   b) Comparar sus centros, sus dispersiones y su simetría. ¿Cuál es el país más homogéneamente poblado?

8. El archivo `cpu.txt` contiene los tiempos de CPU (en segundos) correspondientes a 1000 trabajos enviados por una consultora. Para este conjunto de de datos:

   a) Calcular la media muestral, la mediana muestral y la media $\alpha-$podada con $\alpha = 0{,}1$.

   b) Calcular el desvío estándar muestral y la distancia intercuartil.

   c) Realizar un histograma y un boxplot, determinar características sobresalientes y puntos atípicos.

d) ¿Se puede suponer que los datos tienen una distribución normal? ¿Que gráfico corresponde analizar en este caso?

e) ¿Qué medida de posición se considera más apropiada para describir el centro de los datos en este caso?

9. En un famoso casino, la ruleta está formada por 18 números negros, 18 rojos y 2 verdes. Un jugador tiene un capital de \$100 y apuesta \$1 al rojo cada vez que juega. Si cuando apuesta al rojo sale rojo, entonces recupera su inversión duplicada. Si pierde no recupera nada.

   a) Simular una realización de juego de ruleta y a partir de ella calcular el capital del jugador al finalizar un juego.

   b) Crear una función `cantidad_juegos` que dada una cantidad $Nrep$ de simulaciones del experimento calcule la cantidad de veces que puede jugar el jugador hasta quedarse sin dinero.

   c) Simular 1000 veces el experimento del ítem b y calcular el promedio de los resultados.

   d) Repetir el ítem anterior $n = 100$ veces. Graficar e interpretar el resultado.

10. En una urna hay 4 bolas verdes, 3 amarillas y 3 rojas. Se extraen tres bolas sin reposición. Sean X la cantidad de bolas verdes extraídas e Y la cantidad de rojas.

   a) Simular 100 realizaciones del experimento que consiste en extraer 3 bolas y observar el color, guardando como resultado x la cantidad de verdes e y cantidad de rojas observadas.

   b) Realizar un gráfico de puntos de x vs. y.

   c) Para cada valor observado x, calcular el promedio de y

   d) Graficar los promedios de y en función de los valores observados x

   e) ✍Calcular la función de regresión $\phi(x) = \mathbf{E}[Y|X = x]$.

   f) Superponer en el gráfico antes realizado la verdadera función de regresión $\phi(x)$

11. Sea (X, Y) un vector aleatorio con densidad conjunta

$$f(x, y) = \frac{1}{2x+1} e^{-2x-\frac{y}{4x+2}} \mathbf{1}\{x > 0, y > 0\}$$

   a) ✍Hallar la desndidad de $Y|X = x$ y la densidad de X

   b) Simular una muestra de tamaño 100 para el vector (X, Y)

   c) Realizar un gráfico de puntos de x vs. y.

d) Para cada valor observado x, tomar una ventana de $(x - h, x + h)$, y calcular el promedio de los valores de y para todas las observaciones que caen dentro de dicho intervalo.Considerar $h = 0{,}2$.

e) Graficar los promedios de y en función de los valores observados x, superponiéndolo en el gráfico de puntos utilizando la función *lines*.

f) Repetir los dos ítems anteriores considerando $h = 0{,}6$.

g) Superponer en el gráfico anterior la verdadera función de regresión $\phi(x) = \mathbf{E}[Y|X = x]$

h) Hallar el estimador de Nadaraya Watson, superponer sobre el mismo gráfico y comparar.

12. La función de R *ksmooth* computa el estimador de Nadaraya-Watson a partir de un conjunto de datos, y lo evalúa en un conjunto de puntos intermedios. Para cada una de las tres bases de datos propuestas, se pide:

    a) Mediante la función de R ksmooth estimar la función de regresión que relaciona a las variables x e y a partir de los datos dados usando el núcleo normal con ventana h igual al rango de x dividido 20.

    b) Graficar la función de regresión estimada.

    c) Repetir para valores diferentes de h y superponer en el mismo plot los puntos correspondientes a las observaciones y el valor estimado de la función de regresión obtenida para las diferentes ventanas.

    Bases de datos

    i. Datos: *autos.txt*, X: precio, Y: calidad.

    ii. Datos: *cemento.txt*, X: X_3, Y: X_4.

    iii. Datos: *girasol.txt*, X: inversión, Y: rinde.