Ejercicios para la clase 3

1. **Problema de simulación**

   a) Generar $n = 20$ datos de la siguiente manera. Tomar $X_i \sim U(0,1)$ y dejarlos fijos y definir
   $$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, i = 1, \ldots, n$$
   donde $\beta_0 = 5$, $\beta_1 = 3$ y $\varepsilon_i \sim N(0,1)$. Ajustar un modelo lineal a los datos. Guardar la pendiente y la ordenada al origen.

   b) Fijar la semilla para obtener reproducibilidad. Repetir el experimento (a) $N = 1000$ veces y guardar los valores de $\widehat{\beta}_1$ que se obtienen cada vez,
   $\widehat{\beta}_1^{(1)}, \widehat{\beta}_1^{(2)}, \ldots, \widehat{\beta}_1^{(N)}$. ¿Cuál es la media (muestral) de estos valores? ¿Cuál su varianza muestral? Realizar un histograma de $\widehat{\beta}_1^{(1)}, \widehat{\beta}_1^{(2)}, \ldots, \widehat{\beta}_1^{(N)}$.

   c) Grafique un estimador de la densidad a los $\{\widehat{\beta}_1^{(j)}\}_{1 \leq j \leq 1000}$. ¿Qué distribución parecen tener? ¿Se condice con el teorema visto en teórica?

   d) Hallar un intervalo de confianza de nivel 0.90 para β_1 para los datos del ítem (a). Chequear si el verdadero valor de β_0 pertenece al intervalo.

   e) Fijar la semilla para obtener reproducibilidad. Repetir el experimento (a) $N = 1000$ y en cada una de las repeticiones hallar un intervalo de confianza de nivel 0.90 para β_1 para los datos del ítem (a). Chequear si el verdadero valor de β_1 pertenece al intervalo cada vez y guardar el resultado (es decir, guardar un uno si el verdadero valor de β_1 está contenido en el intervalo de confianza y un cero sino). ¿Qué proporción de los IC construidos contiene al verdadero valor del parámetro? Esto se denomina la cobertura empírica del IC.

   f) Repetir (e) con β_0.

   g) Repetir (e) y (f). La pregunta ahora es ¿qué proporción de las simulaciones conduce a intervalos de confianza en los que ambos parámetros β_0 y β_1 están simultáneamente contenidos en sus respectivos intervalos de confianza? ¿Cuánto debería valer? ¿Cuánto debería valer si los IC construidos se basaran en pivotes independientes? Según lo calculado, le parece que los estimadores $\widehat{\beta}_0$ y $\widehat{\beta}_1$ calculados para una muestra dada son independientes entre sí? ¿Puede contestar a esta pregunta mirando las cuentas teóricas de la Práctica 2?

   h) Hacer un scatter-plot de $\left\{\left(\widehat{\beta}_0^{(j)}, \widehat{\beta}_1^{(j)}\right)\right\}_{1 \leq j \leq N}$. ¿Son independientes $\widehat{\beta}_0$ y $\widehat{\beta}_1$?

   i) Repetir lo anterior graficando histograma, boxplot y estimación de la densidad del estimador de la pendiente estandarizado, dividiendo por σ por un lado y dividiendo por el estimador del error estándar por el otro. ¿Debería tener distribución normal? ¿Qué dice la teoría al respecto? Hacer un boxplot de $\left\{\widehat{\beta}_1^{(j)}\right\}_{1 \leq j \leq N}$ y de los mismos valores estandarizados usando la estimación del standard error, en la misma escala.

   j) Repetir (b), (c), (e), (f), (g) y (h) en los siguientes escenarios:

      1) $n = 5$ y $\epsilon_i \sim N(0,1)$.
      2) $n = 80$ y $\epsilon_i \sim N(0,1)$.
      3) $n = 5$ y $\epsilon_i \sim Exp(1)$, pero acomodando la media de los errores para que sea 0.
      4) $n = 20$ y $\epsilon_i \sim Exp(1)$, pero acomodando la media de los errores para que sea 0.
      5) $n = 80$ y $\epsilon_i \sim Exp(1)$, pero acomodando la media de los errores para que sea 0.
      6) Repetir para estos 3 valores de n pero cambiando la distribución de los errores.

   ‘