Modelos no paramétricos y de regresión | Estadística II

Tarea 1

Fecha de entrega: 9 de febrero

Dada una muestra aleatoria $X_1, \ldots, X_n$ de una distribución $F(\cdot)$, se define la función de distribución empírica como

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I\left\{X_i \leq x\right\}, \qquad x \in \mathbb{R}.$$

1. Para $x \in \mathbb{R}$ fijo, mostrar que $F_n(x)$ es un estimador insesgado y consistente de $F(x)$.
2. Para cada $x \in \mathbb{R}$, se definen los estadísticos $L(x)$ y $U(x)$ como sigue

   $$L(x) = \max\left\{F_n(x) - \epsilon_n, 0\right\} \qquad \text{y} \qquad U(x) = \min\left\{F_n(x) + \epsilon_n, 1\right\},$$

   donde

   $$\epsilon_n = \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}, \qquad \alpha \in (0,1).$$

   Mostrar que

   $$P\left\{L(x) \leq F(x) \leq U(x), \text{ para todo } x \in \mathbb{R}\right\} \geq 1 - \alpha.$$

   Hint: utilizar la desigualdad de Dvoretsky-Keifer-Wolfowitz.
3. Generar 100 observaciones de la distribución $N(0,1)$.
   a. Calcular y gráficar la función de distribución empírica de las observaciones generadas.
   b. Agregar al gráfico del inciso anterior la verdadera *fda* $N(0,1)$.
   c. A partir de las gráficas responder, ¿la función de distribución empírica es similar a la *fda* teoríca?
   d. Para $x \in \{0.0, 0.5, 1.0, \ldots, 2.5, 3.0\}$ comparar los valores de $F(x)$ (la distribución teórica) y los de $F_n(x)$ (la distribución empírica).
   e. A partir de los valores calculados responder, ¿la función de distribución empírica es similar a la *fda* teoríca?
   f. Calcular una banda de confianza 90% para la distribución teórica y graficarla junto con la distribución empírica y la teórica, ¿la distribución teórica está completamente contenida en la banda de confianza?
4. Se sabe que el 20% de los individuos de cierta especie de insectos exhibe cierta característica particular X. Se seleccionan 18 insectos de un ambiente inusual y se observa que cinco tienen la característica X. ¿Es razonable suponer que los insectos de este ambiente exhiben la característica con la misma probabilidad 0.2 que tiene la especie en general?
   a. Establecer H_0 y H_1.
   b. Encontrar la región de rechazo de la prueba binomial de dos colas de tamaño $\alpha = 0.1$.
   c. Calcular el *p-value* de la prueba anterior.
   d. Calcular un intervalo de confianza 90% para la proporción de insectos que tienen la característica X en este ambiente inusual.
5. Un grupo cívico reportó al consejo de la ciudad que al menos el 60% de los residentes están a favor de una emisión de bonos. El consejo de la ciudad seleccionó una muestra aleatoria de 100 miembros y se les preguntó si estaban a favor de la emisión de bonos. 48 de los encuestados respondió que sí.
   a. ¿Es razonable el reporte del grupo cívico? Establecer las hipótesis pertinentes y contrastarlas. Usar $\alpha = 0.1$.
   b. Calcular el *p-value* de la prueba anterior.

c. Calcular un intervalo de confianza 90% para la proporción de residentes a favor de la emisión de bonos.

6. Se seleccionó una muestra de estudiantes de preparatoria, resultando los siguientes pesos (en libras) observados:

$$142, 134, 98, 119, 131, 103, 154, 122, 93, 137, 86, 119, 161, 144, 158, 165, 41, 117, 128, 193$$

Utilizar los datos anteriores para resolver lo siguiente.

a. Probar la hipótesis de que la media de los pesos es 103 lbs.
b. Probar la hipótesis de que el cuartil superior es al menos 150 lbs.
c. Probar la hipótesis de que el tercer decil no es mayor que 100.

7. Se tiene interés en diseñar un automóvil que tenga espacios suficientes para acomodar confortablemente a todos los conductores menos el 5% de mayor estatura. Estudios previos indican que el porcentil 95 era 70.3 pulgadas. Con el propósito de averiguar si los estudios siguen siendo validos se selecciono una muestra aleatoria de conductores de tamaño 100. Se encontró que las 12 personas más altas de las muestras tienen las siguientes estaturas:

$$72.6, 70.0, 71.3, 70.5, 70.8, 76.0, 70.1, 72.5, 71.1, 70.6, 71.9, 72.8$$

a.Es razonable utilizar 70.3 como el porcentil 95.

b. Construir un intervalo de confianza al 95% para el porcentil 95 de la estatura de los conductores de los cuales fue seleccionada la muestra.

8. Sea $\mathbf{X} \sim \text{Multinomial}(n, p_1, \ldots, p_c)$, con $p_k \in (0, 1)$ y $\sum_{k=1}^{c} p_k = 1$. Mostrar lo siguiente

a. $E(X_k) = n\theta_k$, $k = 1, \ldots, c$.
b. $V(X_k) = n\theta_k(1 - \theta_k)$, $k = 1, \ldots, c$.
c. $Cov(X_k, X_j) = -n\theta_k\theta_j$, $k \neq j = 1, \ldots, c$.