TAREA 2 ENP

1.- a) Probar a un nivel $\alpha \leq .10$ la hipótesis nula $H_0: M=2$ vs $H_a: M>2$, donde M es la mediana de una población simétrica continua y se extrae la m.a. => -3, -6, 1, 9, 4, 10, 12

i) Prueba del signo: Como $H_0: M=2 \Rightarrow Z_i = X_i - 2 \Rightarrow Z = \{-5, -8, -1, 7, 2, 8, 10\} \Rightarrow Z^+ = 4$ y $Z^- = 3$
Calculamos la $\mathbb{P}(X=4)$ con $X \sim Bin(7, 1/2)$ que es igual a .2734. Como $H_a: M>2$ (de una cola) entonces el p-value = .2734 > $\{\alpha \mid \alpha \leq .10\}$ ∴ No tenemos evidencia para rechazar $H_0: M=2$

ii) Prueba de rango Wilcoxon: Ahora observamos $|Z_i| = \{5, 8, 1, 7, 2, 8, 10\}$ y R_i denota el rango de $|Z_i| \Rightarrow \{R_i\} = \{3, 5.5, 1, 4, 2, 5.5, 6\}$ y def $\mathbb{1}(Z_i>0) = \begin{cases} 1 \text{ si } Z_i>0, \; 0 \text{ si } Z_i<0 \end{cases}$. Entonces el estadístico de Wilcoxon $T^+ = \sum_{i=1}^{7} R_i \mathbb{1}(Z_i>0) = 4+2+5+6 = 18.5$ Aplicando la prueba en R (wilcox.test) obtenemos un p-value = .2491 ∴ Tampoco rechazamos H_0 //

b) Dar la probabilidad exacta del error tipo I
El error tipo I es: rechazar H_0 dado que es verdadera. Buscamos $K\alpha$ t.q. $\mathbb{P}(X \geq K\alpha) \leq \alpha$. Para $\alpha = 0.05$
$\Rightarrow \sum_{i=[K\alpha]}^{n} \binom{n}{i} (1/2)^n \leq \alpha = .05 \Rightarrow K\alpha = 6$ ∴ la $\mathbb{P}$ exacta del error tipo I $\Rightarrow \alpha^* = 1 - .9921 = .0078$

Al tener empates la prueba de Wilcoxon no nos da el valor exacto.

c) Sobre la base de la m.a. de pares (X, Y). Probar con $\alpha \leq .10$, $H_0: M=2$ vs $H_a: M \neq 2$, donde M es la mediana de una población con distribución continua y simétrica de diferencias $D = X - Y$

i) Prueba de signo: Como $H_0: M=2 \Rightarrow Z_i = D_i - 2$. Entonces calculamos $Z^+ = \sum_{i=1}^{n} \mathbb{1}(Z_i>0)$, y obtenemos $Z^+ = 4$. Como $n=8$, nos interesa calcular $\mathbb{P}(X=4)$ con $X \sim Bin(8, 1/2)$ la cual es .2734 Como $H_a\; M \neq 2 \Rightarrow$ es una prueba de dos colas ∴ p-value $= 2 \times .2734 = .5468 \gg \{\alpha \mid \alpha \leq .10\}$
Entonces no tenemos evidencia para rechazar H_0.

ii) Prueba de Wilcoxon: Ahora aplicamos la prueba "wilcox.test" análoga al ejercicio anterior con el estadístico de Wilcoxon $T^+ = \sum_{i=1}^{8} R_i \mathbb{1}(Z_i>0) = 19.5$. Obtenemos un p-value = .8885
Por lo tanto tampoco rechazamos H_0 //

d) Dar la $\mathbb{P}$ exacta del error tipo I
En i) Buscamos $K\alpha$ con $\alpha = .05$ t.q. $\mathbb{P}(X \geq K\alpha) \leq \alpha \Rightarrow K\alpha = 6 \Rightarrow \sum_{i=6}^{n} \binom{n}{i} (1/2)^n = \alpha^* = .0351$ $\mathbb{P}$ exacta del error tipo 1
De la misma forma que en b), como tenemos empates Wilcoxon test no nos da el valor exacto.

e) Dar el intervalo de confianza a la prueba en c).
Para la prueba de Wilcox el IC del 95% en "R" nos arroja (-5.49, 11.49) para la mediana

3: Probar que la estadística de Wilcoxon $T^+ - T^-$ basada en un conjunto de obs ($\neq 0$) $X_1, \ldots, X_n$ puede ser escrita de la forma: $\sum_{1 \le i \le j \le n} sgn(X_i + X_j)$ donde $sgn(x) = I(x>0) - I(x<0)$

Como vimos en el ej. 1 el estadístico $T^+ = \sum_{i=1}^{n} R|X_i| \mathbb{1}(X_i > 0)$ y $T^- = \sum_{i=1}^{n} R|X_i| \mathbb{1}(X_i < 0)$

Sabemos que si $X_i > 0 \Rightarrow |X_i| = X_i$ y si $X_i < 0 \Rightarrow |X_i| = -X_i$. Como $R(X_i)$ es el rango de la obs. X_i y $X_i \neq 0 \ \forall i \Rightarrow T^+ - T^- = \sum_{i=1}^{n} R|X_i| \mathbb{1}(X_i > 0) - \sum_{i=1}^{n} R|X_i| \mathbb{1}(X_i < 0) = \sum_{i=1}^{n} R|X_i| (\mathbb{1}(X_i > 0) - \mathbb{1}(X_i < 0))$
$= \sum_{i=1}^{n} R|X_i| sgn(X_i)$ Lo cual equivale a la suma de los rangos con el signo original de X_i. Es decir, si $R|X_i| = m$ y $sgn(X_i) = -1 \Rightarrow$ el sumando sería $-m$. Por otro lado, $\sum_{1 \le i \le j \le n} sgn(X_i + X_j) = sgn(X_1 + X_1) + sgn(X_1 + X_2) + \ldots + sgn(X_1 + X_n) + sgn(X_2 + X_2) + \ldots + sgn(X_2 + X_n) + \ldots + sgn(X_n + X_n)$ cada sumando vale $\{-1, 0, 1\}$ y en total son $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$ que coincide con $\sum_{i=1}^{n} R|X_i|$ omitiendo el $sgn(X_i)$. Entonces para el caso que todas las observaciones tuvieran el mismo signo $\therefore T^+ - T^- = \sum_{1 \le i \le j \le n} sgn(X_i + X_j)$. Para el caso contrario, vemos que dentro de la suma $\sum sgn(X_i + X_j)$ podremos encontrar subconjuntos $\equiv R|X_i| sgn(X_i)$ pues si $R|X_i| = m$ habría m términos $sgn(X_i + X_j)$ con el mismo signo de $sgn(X_i)$ pues si $R|X_i| > R|X_j| \Rightarrow sgn(X_i + X_j) = sgn(X_i)$

Entonces $\sum sgn(X_i + X_j)$ equivale a n subconjuntos que forman cada elem. de $\sum R|X_i| sgn(X_i)$

Por lo tanto, se probó que $T^+ - T^- = \sum_{1 \le i \le j \le n} sgn(X_i + X_j)$ //