Tests no paramétricos para una muestra

1. Se midió el tiempo (en segundos) que demora la concentración de un compuesto en reducirse a la mitad durante una reacción. Se realizaron 28 repeticiones de la reacción en condiciones independientes e idénticas. Los datos se exhiben ordenados.

617.2	617.2	617.3	617.4
617.4	617.5	617.6	617.6
617.6	617.7	617.7	617.7
617.7	617.8	617.8	617.8
618.0	618.0	618.2	618.5
619.9	621.9	623.7	626.7
628.1	632.6	648.0	652.7

   (a) Graficar el boxplot y el qqplot (o gráfico de probabilidad normal) de los datos.

   (b) Nos interesa testear las hipótesis $H_0 : \widetilde{\mu} = 620$ versus la alternativa $H_0 : \widetilde{\mu} < 620$ donde $\widetilde{\mu}$ es la mediana de la distribución del tiempo (en segundos) que demora la concentración de un compuesto en reducirse a la mitad durante una reacción. Hacerlo usando el test apropiado. ¿Se puede usar el test de t?

2. Los datos siguientes corresponden al contenido de cierto compuesto en el suero de pacientes que padecen una enfermedad. Interesa testear la hipótesis de que la mediana de la población de la cual provienen los pacientes es $\widetilde{\mu} = 40$. Los datos se presentan ordenados

Suero	35	36	37	37	39	44	48	56	60	76	81	83	83	84	85
$D_i = X_i - m_0$															
$\|D_i\|$															
Rango															

   (a) Realice el histograma de estos datos. ¿Le parece que los datos tienen una distribución normal?

   (b) Aplique el test de rangos signados de Wilcoxon para la mediana de una población a estos datos y dé la conclusión a nivel 0.05. ¿Son válidos los supuestos para realizar el test? Hágalo con el R y también a mano, completando la tabla anterior.

Tests para dos muestras

3. Se consideran dos fórmulas químicas A y B para un nuevo producto que se utilizará para teñir telas. La empresa está interesada en telas especialmente resistentes a perder color tras la exposición al sol. Diez piezas de diferentes tejidos se cortan en dos mitades y a cada una se le aplica uno de los dos tintes. Los 20 trozos de tela se exponen al sol durante un periodo de tiempo, al cabo del cual se mide la intensidad del color, obteniéndose:

Fórmula	1	2	3	4	5	6	7	8	9	10
A	7.2	4.3	5.8	6.5	4.9	6.8	6.3	7.0	6.5	6.2
B	5.1	4.1	5.5	4.1	5.0	5.1	5.3	7.3	4.8	5.8

   Se supone que la diferencia entre la intensidad de color con la fórmula A y la intensidad de color con la fórmula B sigue una distribución $N(\mu_A - \mu_B, \sigma^2)$. Se desea saber si existen diferencias entre las medias de ambas fórmulas.

   (a) Plantee las hipótesis correspondientes y proponga un test de nivel 0.05 para las hipótesis planteadas.

   (b) Construir un intervalo de confianza de nivel 0.95 para la diferencia de intensidad media entre las fórmulas. ¿Tiene razones la empresa para sospechar que la fórmula A es mejor que la fórmula B?

   (c) Suponiendo que la varianza poblacional es $\sigma^2 = 1$ calcular la probabilidad de cometer error de tipo II cuando la diferencia de intensidad de color entre A y B es de 0.97.

4. Se mide el consumo diario de energa (MJ/da) en dos grupos de mujeres elegidas al azar: delgadas y obesas. Los resultados se encuentran el archivo obesas y delgadas.tex. Se desea saber si las medias de consumo de energa de las poblaciones de las que provienen ambos conjuntos de datos coinciden o no. Reponder mediante un test de hiptesis apropiado de nivel 0.05. Realizar y validar los supuestos necesarios.

5. Una compañía pretende decidir si basado solamente en el aspecto del envase será posible modificar el precio de un perfume. Para ello selecciona 15 clientes al azar les hace "probar" el perfume presentado en un envase de aspecto tradicional y les solicita que indiquen cual es el máximo valor que estarían dispuestos a pagar por el mismo. Selecciona otros 15 clientes al azar y repite la prueba pero usando el envase moderno. El precio máximo reportado por cada uno de los 30 clientes se muestra en la tabla siguiente.

Envase	Máximo precio que pagaría														
MODERNO	20	25	40	44	43	13	32	34	35	11	12	46	13	17	47
TRADICION	5	7	10	11	12	17	21	28	33	35	40	40	41	44	45

(a) Realice gráficos descriptivos de ambos conjuntos de datos: histogramas, boxplot, qqplots.

(b) Decida qué test realizar para definir si el precio máximo que están dispuestos a pagar los clientes depende del envase. Indique en base a qué gráficos hace su elección.

(c) Realice el test de hipótesis correspondiente a nivel 0.05 e indique claramente la conclusión.

6. Para evaluar el efecto de la aspirina en la reducción del riesgo de ataques al corazón se llevó a cabo un ensayo clínico en el que participaron 2000 personas. Los individuos fueron divididos en dos grupos de forma aleatoria. A los individuos del grupo control se les suministró un placebo y los del grupo tratamiento recibieron aspirina durante todo el experimento, en dosis controladas. 189 de los 1000 individuos del grupo control sufrieron posteriormente ataques al corazón, en tanto que lo mismo le sucedió a 154 de los 1000 que ingirieron aspirina.

(a) Sea p_1 la proporción poblacional de personas que padecen un ataque cardíaco cuando toman placebo y p_2 la proporción poblacional cuando toman aspirina (en términos médicos, a estas proporciones poblacionales se las denomina tasas de incidencia). Hallar un intervalo de confianza de nivel asintótico 0.95 para $p_1 - p_2$. Defina las variables aleatorias con las que trabaja, el/los estadísticos involucrados y su/s distribución/es.

(b) Escriba en palabras la interpretación del intervalo construido en a).

(c) Diga si la siguiente afirmación es verdadera o falsa, justificando brevemente. "Si se repitiera el estudio asignando aleatoriamente otras 1000 personas a placebo y 1000 a aspirina, la diferencia de proporciones obtenida en este nuevo estudio estará contenida en el intervalo construido en (a) con probabilidad 95%".

(d) Testear $H_0 : p_1 = p_2$ versus la alternativa de que la proporción poblacional de episodios de ataques cardíacos es menor para la población que consume aspirina con nivel 0.05. Hallar el $p-$valor del test para los datos dados.

7. Se quiere estudiar si existen diferencias en la resistencia de dos tipos de materiales utilizados para la fabricación de calzado, a los que denominaremos A y B. Para ello, se arman pares de zapatos en los cuales uno de los zapatos –elegido al azar- se realiza enteramente con el material A y el otro con el material B. Se eligieron 12 chicos al azar, los que usaron durante 2 meses el calzado y al final de este período se midió el estado de deterioro de los zapatos. La tabla muestra los valores de una medida de deterioro (mayor deterioro implica menor resistencia).

Chico	A	B
1	8.14	18.05
2	22.84	21.79
3	6.17	10.16
4	11.88	23.23
5	22.93	33.39
6	14.79	13.35
7	42.84	37.09
8	31.72	42.05
9	7.42	12.50
10	9.52	16.31
11	5.32	15.43
12	3.92	18.52

(a) Realice boxplots y qqplots de los datos de A, B y de la diferencia ($DIF = A - B$), los boxplots realícelos en la misma escala.

(b) Decida qué test realizar para decidir si la resistencia del calzado depende del material. Indique en base a qué gráficos hace su elección.

(c) Defina claramente las variables aleatorias involucradas y los parámetros en cuestión. Realice el test de hipótesis correspondiente a nivel 0.01, escriba las hipótesis e indique claramente la conclusión.