

## Estadística (Q)

Ejercicios para las clases de la quinta semana (E), más tests de hipótesis

---

1. Se mide el consumo diario de energía (MJ/día) en dos grupos de mujeres elegidas al azar: delgadas y obesas. Los resultados son

Delgadas	6.02	7.4	7.88	8.39	8.7	8.76	9.09	9.27
	9.3	9.8	9.84	10.03	10.27			
Obesas	8.42	9.16	9.69	10.21	10.4	10.48	10.93	11.14
	11.14	11.81						

Se desea saber si las medias de consumo de energía de las poblaciones de las que provienen ambos conjuntos de datos coinciden o no. Responder mediante un test de hipótesis apropiado de nivel 0.05. Asuma que los consumos diarios de cada mujer tienen distribución normal, y que las varianzas poblacionales de ambos grupos de mediciones son iguales.

¿Puede validar esta última hipótesis?

Datos:  $qf(0.90, 9, 12) = 2.2135$  y  $qf(0.90, 12, 9) = 2.3789$

2. El jefe de un Centro de gestión y participación (CGP) afirma que al menos el 92% de los documentos allí tramitados se entregan en un plazo estipulado. El defensor del pueblo, dudando de ese porcentaje, decide realizar un test de hipótesis en base a una muestra aleatoria de 130 documentos tramitados en ese CGP.
- Plantear el test de hipótesis, si el defensor del pueblo quiere tener una probabilidad aproximada de 0.05 de contradecir a jefe del CGP cuando su afirmación es correcta. Dar las hipótesis nula y alternativa, indicar el estadístico utilizado, su distribución aproximada bajo  $H_0$  y dar la región de rechazo para el nivel propuesto.
  - De los 130 documentos analizados 112 se entregaron a tiempo. ¿Cuál es su conclusión al nivel propuesto? ¿Es cierta la afirmación del jefe del CGP?
  - Si en realidad se entregan a tiempo el 89% de los documentos ¿qué probabilidad tiene el defensor del pueblo de descubrir que la afirmación del jefe del CGP es falsa?
  - ¿Cuántos de los 130 documentos deben ser entregados a tiempo para rechazar la hipótesis nula a nivel 0.0246?
3. Se consideran dos fórmulas químicas A y B para un nuevo producto que se utilizará para teñir telas. La empresa productora 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 lo siguiente (bajos resultados indican menos intensidad, es decir, mayor pérdida de color):

Fórmulas	Telas									
	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 normal. Se desea saber si existen diferencias entre las medias de ambas fórmulas. Plantee las hipótesis correspondientes y proponga un test de nivel 0.05 para las hipótesis planteadas.

4. 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. Ver el archivo `conc.txt`, donde están los datos.

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  $H_0 : \tilde{\mu} = 620$  versus la alternativa  $H_0 : \tilde{\mu} < 620$  donde  $\tilde{\mu}$  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?

5. 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  $\tilde{\mu} = 40$ . 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) Realizar el histograma de estos datos. ¿Le parece que los datos tienen una distribución normal?
  - (b) Aplicar 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. Comparar con el test de t y el del signo. ¿Se pueden usar ambos? ¿Cuál es preferible?
6. 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.
7. Los datos que siguen corresponden a la ingesta diaria promedio en los diez días previos a la menstruación (pre) y en los diez días posteriores (post) para la misma persona. Se quiere testear si el consumo calórico medio de ambos períodos es el mismo. Hallar un intervalo de confianza de nivel 0.95 para la diferencia media en el consumo previo respecto del posterior a la fecha de la menstruación, basado en esta muestra, validando primero los supuestos necesarios. Testea la hipótesis de que no existe diferencia entre ambos consumos medios.

Sujeto	Pre ( $X_i$ )	Post ( $Y_i$ )	$D_i = X_i - Y_i$	Sujeto	Pre ( $X_i$ )	Post ( $Y_i$ )	$D_i = X_i - Y_i$
1	5260	3910	1350	7	6805	5265	1540
2	5470	4220	1250	8	7515	5975	1540
3	5640	3885	1755	9	7515	6790	725
4	6180	5160	1020	10	8230	6900	1330
5	6390	5645	745	11	8770	7335	1435
6	6515	4680	1835				

8. En un culto suplemento de cultura se afirma que el precio mediano de los libros vendidos en Ciudad Gótica es de 120 pesos. Para validar dicha afirmación, se eligen al azar 20 facturas correspondientes a las ventas de un libro en distintas librerías de la ciudad elegidas al azar. Los precios de dichos libros aparecen listados más abajo. ¿A nivel 0.05 permiten sostener la afirmación hecha en el suplemento? ¿Qué tests podría aplicar? ¿Qué test conviene aplicar?

21.92	25.75	28.41	33.00	45.00	49.79	51.47
53.55	62.91	63.90	75.01	96.99	102.52	124.33
129.44	135.76	143.42	144.93	156.99	159.99	

9. 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). Están en el archivo `zapatos.txt`.

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.
10. La respiración del suelo es una medida de la actividad microbiana en el suelo, que afecta el crecimiento de las plantas. En un estudio se extrajeron muestras de tierra de dos ubicaciones geográficas en un bosque: en un claro del follaje y en una ubicación cercana bajo una densa arboleda. Se midió la cantidad de dióxido de carbono emitida por cada muestra de suelo (mol  $\text{CO}_2/g$  suelo/hora). Los resultados figuran a continuación

claro	22	29	13	16	15	18	14	6
denso	17	20	170	315	22	190	64	

El objetivo del investigador es comparar los niveles de respiración de ambos suelos, y concluir a nivel 0.05. Realice el test apropiado para dar una conclusión. Justifique la utilización del test.

11. Hagamos un test sobre datos generados. Genere 50 datos con distribución Exponencial, con  $\lambda = 1/4$ . Luego aplique un test apropiado para chequear si la esperanza de la población de la que proviene la muestra es igual a 4. De paso, chequee si los datos generados superan un test de normalidad.
12. Un diseñador de productos está interesado en reducir el tiempo de secado de una pintura. Se prueban dos fórmulas de pintura; la fórmula 1 tiene el contenido químico estándar y la fórmula 2 tiene un nuevo ingrediente secante que tiende a reducir el tiempo de secado. Se sabe que el tiempo de secado es una variable aleatoria con distribución normal. Se pintan 16 placas con la fórmula 1 y otras 16 con la fórmula 2. Los dos tiempos promedio de secado obtenidos fueron 118 y 112 minutos respectivamente, mientras que los desvíos estándar fueron de 10.5 y 7.
- (a) Defina cuáles son las hipótesis a testear y qué test se debe utilizar.
- (b) Realice el test de hipótesis correspondiente a nivel 0.05 e indique claramente la conclusión.
- (c) Calcule el p-valor. ¿Qué decisión se hubiera tomado a nivel 0.01?
13. Queremos comparar los datos de la ingesta media diaria (durante 10 días) de un grupo de individuos (kJ) que siguen una determinada dieta alimentaria, con la ingesta recomendada de 7752 kJ. Es decir, cada dato corresponde al promedio de ingesta de energía durante 10 días para un individuo en particular.

5260      5470      5640      6180      6390      6515      6805      7515      7515      8230      8770

- (a) Se busca responder con un test de hipótesis de nivel del 5%, definiendo claramente las variables aleatorias, los parámetros de interés y las hipótesis en cuestión, así como el estadístico utilizado y su distribución bajo la hipótesis nula, y los supuestos realizados y verificados. Escriba su conclusión en los términos del problema. Los datos se presentan ordenados por comodidad. ( $\bar{x} = 6753.6$   $s = 1142.1$ )
- (b) Construya intervalos de confianza de nivel 0.90, 0.95 y 0.99 para la ingesta media diaria (en kJ) para los que siguen esta dieta específica. ¿Cuál es más largo?
14. Se mide el grado de impurezas de un producto químico. El método de medición está afectado por un error que se supone  $N(0, \sigma^2)$ ,  $\sigma^2$  desconocido. Además los errores correspondientes a diferentes mediciones son independientes entre sí. Se hicieron 12 observaciones obteniendo que el promedio es 0.85 con un desvío estándar muestral de 0.05. A partir de estos datos, ¿hay evidencia significativa para decir que el grado de impurezas del producto es distinto de 0.7 a nivel 0.05?