

Métodos no paramétricos para la comparación de dos muestras

Pértega Díaz, S.

Unidad de Epidemiología Clínica y Bioestadística. Complejo Hospitalario Universitario Juan Canalejo.

Pita Fernández, S.

Unidad de Epidemiología Clínica y Bioestadística. Complejo Hospitalario Universitario Juan Canalejo

Cad Aten Primaria
Año 2006
Volumen 13
Pág. 109-113

A la hora de analizar los datos recogidos para una investigación, la elección de un método de análisis adecuado es crucial para evitar llegar a conclusiones erróneas. La selección de la técnica de análisis más apropiada ha de hacerse tomando en cuenta distintos aspectos relativos al diseño del estudio y a la naturaleza de los datos que se quieren cuantificar. El número de grupos de observaciones a comparar, la naturaleza de las mismas (según se trate de muestras independientes u observaciones repetidas sobre los mismos individuos), el tipo de datos (variables continuas / cualitativas) o su distribución de probabilidad son elementos determinantes a la hora de conocer las técnicas estadísticas que se pueden utilizar¹.

En el análisis de datos cuantitativos, los métodos estadísticos más conocidos y utilizados en la práctica, como el test t de Student o el análisis de la varianza, se basan en asunciones que no siempre son verificadas por los datos de los que se dispone. Así, es frecuente tener que asumir que la variable objeto de interés sigue por ejemplo una distribución gaussiana. Cuando la ausencia de normalidad es obvia, o no puede ser totalmente asumida por un tamaño muestral reducido, suele recurrirse a una transformación de la variable de interés (por ejemplo, la transformación logarítmica) para simetrizar su distribución o bien justificar el uso de las técnicas habituales recurriendo a su robustez (esto es, su escasa sensibilidad a la ausencia de normalidad). Existen a su vez otros métodos, usualmente llamados no paramétricos, que no requieren de este tipo de hipótesis sobre la distribución de los datos, resultan fáciles de implementar y pueden calcularse incluso con tamaños de muestra reducidos. En el presente trabajo se describirán algunos de los métodos no paramétricos más utilizados en la práctica¹⁻³.

Dos muestras independientes: la prueba U de Mann-Whitney y la prueba de la suma de rangos de Wilcoxon.

En muchas situaciones se desea contrastar si la distribución de una variable X es igual en dos poblaciones, o bien si dicha variable tiende a ser mayor (o menor) en alguno de los dos grupos,

basándose en los datos muestrales. Por ejemplo, puede resultar interesante comparar el descenso de peso en pacientes sometidos a dos dietas alimenticias distintas, o el nivel de dolor en sujetos con artrosis que reciben un tratamiento frente a placebo. En la teoría estadística "tradicional" la prueba que se aplicaría para realizar este tipo de comparaciones sería el test t de Student para dos muestras independientes⁴, siendo la U de Mann-Whitney o la prueba de la suma de rangos de Wilcoxon pruebas de carácter no paramétrico equivalentes que podrían emplearse también en esta situación.

De un modo más formal, supongamos que se dispone de observaciones de una misma variable X (pérdida de peso, puntuación de dolor, etc.) en dos poblaciones distintas sobre muestras de tamaño n_1 y n_2 , respectivamente:

Población 1: $x_1^1, x_2^1, \dots, x_{n_1}^1$

Población 2: $x_1^2, x_2^2, \dots, x_{n_2}^2$

Un modo intuitivo de proceder consiste en ordenar las observaciones obtenidas, independientemente de su población de origen, de menor a mayor valor y asignar rangos a los datos así ordenados. De esta forma, a la observación con un valor más pequeño se le asigna rango 1, a la siguiente rango 2, y así sucesivamente. En caso de empates (si dos o más observaciones coinciden en valor) se le asignará a cada una de estas observaciones el promedio de los rangos que les serían asignados si no hubiese empate.

Si no existiesen diferencias en la distribución entre ambas poblaciones, los rangos deberían mezclarse aleatoriamente entre las dos muestras. En cambio, si la suma de los rangos asignados a las observaciones de una de las poblaciones resulta mucho mayor que la suma de los rangos asignados a las observaciones de la otra población, esto indicaría una diferencia en la distribución de la variable X entre ambas.

Denotemos por $R(x_j^i)$ $i=1,2, j=1,\dots,n_i$ el rango asignado a cada una de las observaciones disponibles. Consideraremos como estadístico de contraste para la prueba de la suma de rangos de Wilcoxon la suma de los rangos en una de las poblaciones:

$$T = T_1 = \sum_{j=1}^{n_1} R(x_j^1) \quad \text{ó} \quad T = T_2 = \sum_{j=1}^{n_2} R(x_j^2)$$

La distribución de probabilidad de los estadísticos anteriores ha sido tabulada para tamaños de muestra pequeños y en el caso de no existir empates (Tabla 1). Así, la Tabla 1 sirve para conocer si el resultado es significativo a nivel bilateral si se trabaja con una seguridad del 95% y tamaños muestrales ≤ 15 .

Para tamaños muestrales mayores (>15), es adecuado utilizar la aproximación normal, obteniendo a partir de T la variable:

$$z = \frac{T - m_T}{\sigma_T}$$

donde m_T y σ_T son la media y desviación estándar de T si la hipótesis nula es cierta, y vienen dadas por las siguientes fórmulas:

$$m_T = \frac{1}{2}n_1(n_1 + n_2 + 1) \quad \text{y} \quad \sigma_T^2 = \frac{n_1n_2(n_1 + n_2 + 1)}{12}$$

El número de empates debe ser además pequeño en relación con el número total de observaciones. En el caso de empates, la varianza del estadístico T debe modificarse, de modo que la expresión anterior quedaría como sigue:

$$z = \frac{T - \frac{n_1}{2}(n_1 + n_2 + 1)}{\sqrt{\frac{n_1n_2}{(n_1 + n_2)(n_1 + n_2 + 1)} \sum_{j=1}^2 \sum_{j=1}^{n_j} R(x_j^j) - \frac{n_1n_2(n_1 + n_2 + 1)^2}{4(n_1 + n_2 - 1)}}$$

Una vez obtenido el valor de z éste se debe referir a las tablas de la distribución normal para obtener el valor de significación asociado⁵.

TABLA 1

Percentiles 95 para la prueba de la suma de rangos de 2 muestras de Wilcoxon.

		n_1														
		4	5	6	7	8	9	10	11	12	13	14	15			
n_2																
4		10														
5		11	17													
6		12	18	26												
7		13	20	27	36											
8		14	21	29	38	49										
9		14	22	31	40	51	62									
10		15	23	32	42	53	65	78								
11		16	24	34	44	55	68	81	96							
12		17	26	35	46	58	71	84	99	115						
13		18	27	37	48	60	73	88	103	119	136					
14		19	28	38	50	62	76	91	106	123	141	160				
15		20	29	40	52	65	79	94	110	127	145	164	184			

Se define n_1 como el tamaño muestral menor, $n_1 \leq n_2$. Se calcula T como la suma de los rangos asignados a las observaciones de la muestra 1. Si $T > \frac{1}{2}n_1(n_1 + n_2 + 1)$, se toma $T = n_1(n_1 + n_2 + 1) - T$. El resultado es significativo al nivel bilateral del 5% si T es menor o igual al valor tabulado.

Para ilustrar el uso de esta prueba, consideraremos los datos de la Tabla 2, correspondiente a los valores de medición del dolor (en una escala de 0 a 10) en dos grupos de 11 pacientes sometidos a dos tratamientos analgésicos diferentes. En este caso $n_1=n_2=11$. La suma de los rangos asignados a las observaciones del primer grupo es

$$T=171, \text{ y su media } \frac{1}{2}n_1(n_1 + n_2 + 1) = \frac{1}{2} \times 11 \times (11 + 11 + 1) = \frac{1}{2} \times 11 \times 23 = 126,5$$

Puesto que la suma de rangos obtenida supera a la esperada, consideraremos como estadístico final $T=171-126,5=44,5$, y lo referiremos a los valores en la Tabla 1. Trabajando con un planteamiento bilateral y una seguridad del 95%, la región de rechazo corresponde a valores de T menores o iguales a 96, por lo cual se rechazaría la hipótesis nula de igual nivel de dolor en ambos grupos de tratamiento con un nivel de significación $p<0.05$.

En el ejemplo propuesto podemos comprobar el resultado que se obtendría al utilizar la aproximación normal. Tal y como vimos:

$$m_T = \frac{1}{2}n_1(n_1 + n_2 + 1) = 126,5$$

$$\sigma_T^2 = \frac{n_1n_2(n_1 + n_2 + 1)}{12} = \frac{11 \times 11 \times (11 + 11 + 1)}{12} = 231,17 \Rightarrow \sigma_T = \sqrt{231,17} = 14,6$$

Con lo cual se utilizaría el estadístico:

$$z = \frac{T_1 - m_T}{\sigma_T} = \frac{171 - 126,5}{14,6} = 3,05$$

que debe referirse a los valores de una distribución normal estándar⁵. Así, se obtiene $p=0,002$, concluyéndose igualmente que el nivel de dolor es diferente según la terapia analgésica utilizada.

TABLA 2

Valores de medición del dolor (en una escala de 0 a 10) en dos grupos de pacientes sometidos a dos tratamientos analgésicos diferentes.

	Tratamiento A	Rango asignado	Tratamiento B	Rango asignado
1	100	22	85	18
2	85	18	62	8
3	72	15	57	6,5
4	64	9,5	32	1
5	70	12,5	70	12,5
6	56	5	54	4
7	87	20	45	2
8	71	14	50	3
9	95	21	64	9,5
10	80	16	66	11
11	85	18	57	6,5
Suma de rangos		171		82

Por otra parte, es frecuente referirse a la prueba de la suma de rangos de Wilcoxon con el nombre de prueba U de Mann-Whitney. En realidad, son dos pruebas diferentes, aunque esencialmente equivalentes entre sí. Para el cálculo de la prueba U de Mann-Whitney, en lugar de la suma de rangos se calcularán los valores:

U_{12} : el número de pares para los cuales una observación de la primera población es inferior a una observación de la segunda población, $x_i^1 < x_j^2$.

U_{21} : el número de pares para los cuales una observación de la primera población es superior a una observación de la segunda población, $x_1^1, x_2^1, \dots, x_n^1$.

En caso de empate se contabilizarán 0,5 unidades a mayores en cada una de las cantidades anteriores. De forma análoga a como ocurría con la prueba anterior, valores bajos de U_{12} indicarán una diferencia hacia valores más altos de la variable en la primera población, mientras que valores altos indicarán que estos tienden a ser más altos en la segunda población.

Los parámetros anteriores se relacionan con el estadístico T descrito anteriormente mediante la siguiente ecuación:

$$T = n_1 n_2 + \frac{1}{2} n_1 (n_1 + 1) - U_{12}$$

De forma que a partir del estadístico U puede obtenerse inmediatamente el valor del estadístico de Wilcoxon y utilizar la metodología anterior para obtener el valor de significación asociado. De hecho, la mayor parte de programas estadísticos, como el SPSS, muestran en sus salidas los valores de ambos estadísticos, junto con un p-valor común, bien calculado a partir de la aproximación asintótica mediante una distribución normal o a partir de las tabulaciones correspondientes, corrigiendo la posibilidad de empates. Otra prueba equivalente aunque menos conocida es la S de Kendall, calculada según $S = U_{12} - U_{21}$.

Por último, decir que al igual que el análisis de la varianza en el abordaje estadístico "tradicional" extiende la prueba t de Student al caso en el que se quieren comparar más de dos grupos, el test de Kruskal-Wallis es una extensión natural de la prueba de Mann-Whitney a esta situación. Para su cálculo se ordenarán las N observaciones obtenidas, independientemente de su población de origen, de menor a mayor valor y se asignarán los rangos correspondientes. El estadístico de contraste para la prueba de Kruskal-Wallis vendrá dado por:

$$H = \frac{12 \sum_{i=1}^k (\bar{R}_i - \bar{R})^2}{N(N+1)}$$

donde N denota al número total de observaciones en los k grupos que se comparan, $\bar{R}_i, i = 1, \dots, k$ es el promedio de los rangos de las observaciones del i-ésimo grupo y el promedio de todos los rangos. Así definido, el estadístico H sigue una distribución χ^2 con k-1 grados de libertad.

Dos muestras relacionadas: la prueba del signo y la prueba de la suma de rangos con signo de Wilcoxon.

Otra situación muy frecuente es aquella en la que se desea comparar la distribución de una variable X en dos muestras de casos apareados, usualmente sobre los mismos individuos en dos momentos diferentes de tiempo. Por ejemplo, puede quererse comparar el nivel de dolor en una articulación antes y después de un tratamiento con infiltraciones, o el peso antes y después de someterse a algún programa de adelgazamiento. En estas situaciones, es lógico trabajar con la diferencia de las observaciones entre ambos momentos (pérdida de peso, disminución del nivel de dolor, etc.):

$$\begin{matrix} x_1^1 - x_1^2 \\ x_2^1 - x_2^2 \\ \dots \\ \dots \\ x_n^1 - x_n^2 \end{matrix}$$

donde aquí $x_j^1, j = 1, 2, \dots, n$ denotan los valores observados de la variable X en n individuos en el primer instante y $x_j^2, j = 1, 2, \dots, n$ los valores observados en un instante posterior.

Una forma sencilla de proceder consiste en contabilizar el número r de diferencias positivas y el número s de diferencias negativas (sin contar los valores 0). Bajo la hipótesis nula de que no existen diferencias, será igualmente probable obtener una diferencia positiva o negativa, por lo que ambos valores se distribuirán según una distribución

TABLA 3

Tabla de la distribución binomial acumulada Bi (n,p) con p=1/2.

$$X \sim Bi(n, p) \quad P(X \leq k) = \sum_{j=0}^k \binom{n}{j} p^j (1-p)^{n-j}, p = \frac{1}{2}$$

n	k															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
5	0,031	0,188	0,500	0,812	0,969											
6	0,016	0,109	0,344	0,656	0,891	0,984										
7	0,008	0,062	0,227	0,500	0,773	0,938	0,992									
8	0,004	0,035	0,145	0,363	0,637	0,855	0,965	0,996								
9	0,002	0,020	0,090	0,254	0,500	0,746	0,910	0,980	0,998							
10	0,001	0,011	0,055	0,172	0,377	0,623	0,828	0,945	0,989	0,999						
11		0,006	0,033	0,113	0,274	0,500	0,726	0,887	0,967	0,994						
12		0,003	0,019	0,073	0,194	0,387	0,613	0,806	0,927	0,981	0,997					
13		0,002	0,011	0,046	0,133	0,291	0,500	0,709	0,867	0,954	0,989	0,998				
14		0,001	0,006	0,029	0,090	0,212	0,395	0,605	0,788	0,910	0,971	0,994	0,999			
15			0,004	0,018	0,059	0,151	0,304	0,500	0,696	0,849	0,941	0,982	0,996			
16			0,002	0,011	0,038	0,105	0,227	0,402	0,598	0,773	0,895	0,962	0,989	0,998		
17			0,001	0,006	0,025	0,072	0,166	0,315	0,500	0,685	0,834	0,928	0,975	0,994	0,999	
18			0,001	0,004	0,015	0,048	0,119	0,240	0,407	0,593	0,760	0,881	0,952	0,985	0,996	0,999
19				0,002	0,010	0,032	0,084	0,180	0,324	0,500	0,676	0,820	0,916	0,968	0,990	0,998
20				0,001	0,006	0,021	0,058	0,132	0,252	0,412	0,588	0,748	0,868	0,942	0,979	0,994
21				0,001	0,004	0,013	0,039	0,095	0,192	0,332	0,500	0,668	0,808	0,905	0,963	0,987
22					0,002	0,008	0,026	0,067	0,143	0,262	0,416	0,584	0,738	0,867	0,933	0,974
23					0,001	0,005	0,017	0,047	0,105	0,202	0,339	0,500	0,661	0,798	0,895	0,953
24					0,001	0,003	0,011	0,032	0,076	0,154	0,271	0,419	0,581	0,729	0,846	0,924
25						0,002	0,007	0,022	0,054	0,115	0,212	0,345	0,500	0,655	0,788	0,885

binomial de parámetros $Bi(r+s, 1/2)$. Recurriendo a las tablas de la distribución binomial, podemos obtener a partir de r (o, equivalentemente, de s) el valor exacto de significación asociado (Tabla 3).

Como ejemplo, utilizaremos los datos de la Tabla 4 en la que se muestra la pérdida de peso alcanzada por 20 sujetos sometidos a un programa de adelgazamiento. El número de observaciones positivas (pacientes que realmente perdieron peso) es $r=14$, mientras que el número de observaciones negativas (pacientes que ganaron peso) es $s=6$. Refiriendo estos valores a los de una distribución binomial de parámetros $Bi(20, 1/2)$ se obtiene un valor de $p=2 \times 0,058=0,116$, por lo que no puede concluirse que exista una pérdida de peso significativa en los pacientes estudiados.

TABLA 4

Valores del peso en una muestra de 20 pacientes antes y después de someterse a un programa de adelgazamiento.

Peso inicial	Peso final	Pérdida de peso	Signo	Rango asignado
110	89	21	+	20
98,5	90,5	8	+	14
87	89	-2	-	6
95,5	98	-2,5	-	7
79	72,5	6,5	+	13
84,5	80	4,5	+	11,5
96	85	11	+	16,5
99	98	1	+	1,5
120	110	10	+	15
104	105,5	-1,5	-	4
76,5	75	1,5	+	4
83	80	3	+	9
83	84	-1	-	1,5
92	80	12	+	18
90	79	11	+	16,5
96,5	92	4,5	+	11,5
101	98	3	+	9
106	90	16	+	19
96	97,5	-1,5	-	4
87	90	-3	-	9

Número de valores positivos: $r=14$
 Número de valores negativos: $s=6$
 Suma de los rangos de los valores positivos: $T_+=178,5$
 Suma de los rangos de los valores negativos: $T_-=31,5$

Para tamaños muestrales grandes ($n \geq 20$) puede utilizarse como estadístico de contraste:

$$t = \frac{r - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}}$$

que seguirá aproximadamente una distribución normal estándar $N(0,1)$.

En el ejemplo anterior:

$$t = \frac{r - \frac{1}{2}n}{\frac{1}{2}\sqrt{n}} = \frac{14 - \frac{1}{2} \times 20}{\frac{1}{2}\sqrt{20}} = \frac{4}{2,24} = 1,78$$

Si referimos el valor obtenido a la función de probabilidad de una distribución $N(0,1)$ se obtiene $p=0,075$, no resultando en un valor significativo tal y como ocurría con la aproximación por la binomial.

La prueba del signo, tal y como se denomina a la prueba que se acaba de describir, presenta como mayor limitación el hecho de que no tiene en cuenta la magnitud (positiva o negativa) de las observaciones. Así, puede ocurrir que existan muchas diferencias positivas pero de escasa magnitud (pacientes que pierden peso pero en poco volumen) y pocas diferencias negativas pero de mucha mayor importancia (pacientes que ganan mucho peso). Este tipo de situaciones deberían reducir la posibilidad de encontrar diferencias significativas entre las observaciones.

La prueba de la suma de rangos con signo de Wilcoxon toma en consideración la deficiencia anterior. Las observaciones se ordenan de menor a mayor valor absoluto y se les asignan rangos (ignorando los valores nulos y actuando igual que en el caso de la prueba de suma de rangos ante empates). Se utilizará como estadístico de contraste la suma T_+ de los rangos asignados a valores positivos o bien la suma T_- de los rangos asignados a valores negativos. Para valores pequeños de n , la distribución de T_+ y T_- está completamente tabulada, y puede utilizarse para obtener los valores críticos del test (Tabla 5). En el caso de muestras grandes ($n \geq 20$), la distribución de T_+ y T_- puede aproximarse por la de una variable normal. Así, realizando la transformación:

TABLA 5

Valores críticos para la prueba de la suma de rangos con signo de Wilcoxon según el tamaño de muestra y el nivel de significación asociado

Tamaño de la muestra excluidas las diferencias nulas	Nivel de significación bilateral	
	0,05	0,01
6	0	---
7	2	---
8	3	0
9	5	1
10	8	3
11	10	5
12	13	7
13	17	9
14	21	12
15	25	15
16	29	19
17	34	23
18	40	27
19	46	32
20	52	37
21	58	42
22	66	48
23	73	54
24	81	61
25	89	68

Se muestra el valor crítico para la menor de las sumas de rangos T_+ ó T_- . Un valor es significativo al nivel bilateral señalado si la suma de rangos es menor o igual al valor tabulado.

$$\frac{\left| T_+ - \frac{1}{4}n'(n'+1) \right|}{\sqrt{\{n'(n'+1)(2n'+1)/24\}}}$$

la distribución asociada es la de una normal estándar (donde n' es el número de observaciones no nulas).

Al igual que ocurría en la prueba de la suma de rangos para muestras independientes, en caso de empate la varianza del estadístico varía y debería efectuarse alguna corrección en la expresión anterior. Así mismo, los valores críticos en la Tabla 5 para el caso de empates suelen resultar algo conservadores, es decir, con empates se tenderá a aceptar la hipótesis nula de no diferencias cuando en realidad ésta debería ser rechazada.

Volviendo al ejemplo anterior, y tras ordenar las observaciones en la Tabla 4 de menor a mayor valor absoluto, se obtiene las sumas de rangos $T_+=178,5$ y $T_-=31,5$ correspondientes a las observaciones con pérdida o ganancia de peso, respectivamente. Refiriendo estos valores a los que se muestran en la Tabla 5 con $n=20$, se obtiene un valor de $p<0.01$.

Obviamente, el tipo de procedimientos aquí expuestos permiten cubrir sólo una pequeña parte de las situaciones que se nos pueden plantear en la práctica. En la mayoría de las ocasiones, se dispone de información en una gran cantidad de variables, lo cual requiere recurrir a otros métodos de análisis que no sólo permitan estudiar las relaciones entre un par de variables, sino estudiar el efecto conjunto de todas ellas.

BIBLIOGRAFÍA

1. Altman D.G. Practical Statistics for Medical Research. London: Chapman & Hall, 1991.
2. Armitage P., Berry G. Estadística para la investigación biomédica. Barcelona: Doyma, 1992.
3. Milton J.S., Tsokos J.O.. Estadística para biología y ciencias de la salud. Madrid: Interamericana_McGraw Hill; 1989.
4. Pértega Díaz S, Pita Fernández S. Métodos paramétricos para la comparación de dos medias. T de Student. Cad Aten Primaria 2001; 8: 37-41
5. Pértega Díaz S, Pita Fernández S. La distribución normal. Cad Aten Primaria 2001; 8: 268-274.

Información

Cadernos de Atención Primaria

Material adicional en Internet

A revista Cadernos de Atención Primaria é accesible por internet (www.agamfec.com). Os autores que o desexen poden presentar material complementario na páxina web (enquisas utilizadas, anexos, aspectos metodolóxicos máis detallados, fotos,...). Este tipo de material enviarase de modo diferenciado ao resto do texto. Cando algún traballo inclúa material adicional na páxina web, éste identificarase na revista con esta imaxe.

