a Use la prueba de signos para determinar si existe suficiente evidencia para indicar una diferencia en la respuesta media para los dos estímulos. Use una región de rechazo para la cual $\alpha \leq .05$.

b Pruebe la hipótesis de que no haya diferencia en la respuesta media, usando una prueba t de Student.

15.8 Consulte el Ejercicio 12.15. Usando la prueba de signos, ¿encuentra suficiente evidencia para apoyar la conclusión de que los tiempos de terminación difieren para las dos poblaciones? Use $\alpha = .10$.

15.9 El conjunto de datos de la siguiente tabla representa el número de accidentes industriales en 12 plantas de manufactura, durante periodos de una semana antes y después de intensa promoción sobre seguridad.

Planta	Antes	Después	Planta	Antes	Después
1	3	2	7	5	3
2	4	1	8	3	3
3	6	3	9	2	0
4	3	5	10	4	3
5	4	4	11	4	1
6	5	2	12	5	2

a ¿Los datos apoyan la afirmación de que la campaña fue un éxito? ¿Cuál es el nivel de significancia alcanzado? ¿Qué concluiría usted con $\alpha = .01$?

b Analice los problemas asociados con un análisis paramétrico diseñado para contestar la pregunta del inciso a.

15.4 Prueba de rangos con signo de Wilcoxon para un experimento de observaciones pareadas

Al igual que en la Sección 15.3 suponga que tenemos n observaciones pareadas de la forma (X_i, Y_i) y que $D_i = X_i - Y_i$. De nuevo suponemos que estamos interesados en probar la hipótesis de que las X y las Y tienen la misma distribución contra la alternativa de que las distribuciones difieren en localización. De acuerdo con la hipótesis nula de que no hay diferencia en las distribuciones de las X y las Y, se esperaría (en promedio) que la mitad de las diferencias en los pares sean negativas y la mitad positivas. Esto es, el número esperado de diferencias negativas entre pares es $n/2$ (donde n es el número de pares). Además, se podría inferir que las diferencias positivas y negativas de igual magnitud absoluta deberían presentarse con la misma probabilidad. Si ordenáramos las diferencias de acuerdo con sus valores absolutos y las clasificáramos de menor a mayor, las sumas de los rangos esperados para las diferencias negativas y positivas serían iguales. Las diferencias grandes en las sumas de los rangos asignados a las diferencias positivas y negativas darían evidencia para indicar un desplazamiento de localización entre las dos distribuciones.

Para llevar a cabo la prueba de Wilcoxon calculamos las diferencias (D_i) para cada uno de los n pares. Las diferencias iguales a cero se eliminan y el número de pares, n, se reduce de conformidad. Entonces clasificamos los *valores absolutos* de las diferencias asignando un 1 al más pequeño, un 2 al segundo más pequeño y así sucesivamente. Si dos o más diferencias absolutas están empatadas para el mismo rango, entonces el promedio de las clasificaciones que se hubieran asignado a estas diferencias se asigna a cada miembro del grupo empatado. Por ejemplo, si dos diferencias absolutas están empatadas para los rangos 3 y 4, entonces cada una recibe clasificaciones 3.5 y a la siguiente diferencia absoluta más alta se le asigna el rango 5. Entonces calculamos la suma de los rangos (suma de rango) para las diferencias negativas y también calculamos la suma de rango para las diferencias positivas. Para una prueba de dos colas usamos T, la *más* *pequeña* de estas dos cantidades, como estadístico de prueba para probar la hipótesis nula de que los dos histogramas de frecuencia relativa poblacionales son idénticos. Cuanto menor sea el valor de T, mayor será el valor de evidencia a favor del rechazo de la hipótesis nula. En consecuencia, rechazaremos la hipótesis nula si T es menor o igual a algún valor, por ejemplo, T_0.

Para detectar la alternativa unilateral, que afirma que la distribución de las X se desplaza a la derecha de la de las Y, usamos la suma de rango T^- de las diferencias negativas y rechazamos la hipótesis nula para valores pequeños de T^-, por ejemplo, $T^- \leq T_0$. Si deseamos detectar un desplazamiento de la distribución de las Y a la derecha de las X, usamos la suma de rango T^+ de las diferencias positivas como estadístico de prueba y rechazamos valores pequeños de T^+, por ejemplo, $T^+ \leq T_0$.

La probabilidad de que T sea menor o igual a algún valor T_0 se ha calculado para una combinación de tamaños muestrales y valores de T_0. Estas probabilidades, dadas en la Tabla 9, Apéndice 3, se pueden usar para hallar la región de rechazo para la prueba basada en T.

Por ejemplo, supongamos que usted tiene $n = 7$ pares y desea realizar una prueba de dos colas de la hipótesis nula de que las dos distribuciones de frecuencia relativa poblacionales son idénticas. Entonces, con $\alpha = .05$, rechazaría la hipótesis nula para todos los valores de T menores o iguales a 2. La región de rechazo para la prueba Wilcoxon de suma de rango para un experimento pareado es siempre de esta forma: rechazar la hipótesis nula si $T \leq T_0$ donde T_0 es el valor crítico para T. Los límites para el nivel de significancia alcanzado (valor p) se determinan como sigue. Para una prueba de dos colas, si $T = 3$ se observa cuando $n = 7$, la Tabla 9, Apéndice 3, indica que H_0 sería rechazada si $\alpha = .1$, pero no si $\alpha = .05$. Entonces, $.05 <$ valor $p < .1$. Para una alternativa unilateral de que las X están desplazadas a la derecha de las Y con $n = 7$ y $\alpha = .05$, H_0 es rechazada si $T = T^- \leq 4$. En este caso, si $T = T^- = 1$, entonces $.01 <$ valor $p < .025$. La prueba basada en T, llamada *prueba de de rangos con signo Wilcoxon*, se resume como sigue.

Prueba de rangos con signo de Wilcoxon para un experimento de observaciones pareadas

H_0 : las distribuciones poblacionales para las X y las Y son idénticas.

H_a : (1) las dos distribuciones poblacionales difieren en localización (dos colas), *o bien* (2) la distribución de frecuencia relativa poblacional para las X se desplaza a la derecha de la de las Y (una cola).

Estadístico de prueba:

1. Para una prueba de dos colas, use $T = \text{mín}(T^+, T^-)$, donde $T^+ =$ suma de los rangos de las diferencias positivas y $T^- =$ suma de los rangos de las diferencias negativas.
2. Para una prueba de una cola (para detectar la alternativa de una cola que acabamos de dar), use la suma de rango T^- de las diferencias negativas.[2]

Región de rechazo:

1. Para una prueba de dos colas, rechace H_0 si $T \leq T_0$, donde T_0 es el valor crítico para la prueba de dos lados dada en la Tabla 9, Apéndice 3.
2. Para una prueba de una cola (como se describe líneas antes), rechace H_0 si $T^- \leq T_0$, donde T_0 es el valor crítico para la prueba unilateral.

2. Para detectar un desplazamiento de la distribución de las Y a la derecha de la distribución de las X, use la suma de rango T^+, la suma de los rangos de las diferencias positivas, y rechace H_0 si $T^+ \leq I_0$.

EJEMPLO **15.3** Debido a la variación de un horno a otro, se utilizó un experimento de observaciones pareadas para probar diferencias en pasteles elaborados usando la mezcla A y la mezcla B. Dos pasteles elaborados usando cada mezcla, se hornearon en cada uno de seis hornos diferentes (un total de 12 pasteles). Pruebe la hipótesis de que no hay diferencia en las distribuciones poblacionales de las densidades de pastel usando las dos mezclas. ¿Qué se puede decir acerca del nivel de significancia alcanzado?

Solución Los datos originales y las diferencias en densidades (en onzas por pulgada cúbica) para los seis pares de pasteles se muestran en la Tabla 15.2.

Al igual que con otras pruebas no paramétricas, la hipótesis nula a probar es que las dos distribuciones de frecuencia poblacionales de densidades de pastel son idénticas. La hipótesis alternativa es que las distribuciones difieren en localización, lo cual implica que se requiere una prueba de dos colas.

Debido a que la cantidad de datos es pequeña, realizaremos nuestra prueba usando $\alpha = .10$. De la Tabla 9, Apéndice 3, el valor crítico de T para una prueba de dos colas, $\alpha = .10$, es $T_0 = 2$. En consecuencia, rechazaremos H_0 si $T \leq 2$.

Hay sólo una diferencia positiva y tiene rango 3; por tanto, $T^+ = 3$. Como $T^+ + T^- = n(n + 1)/2$ (¿por qué?), $T^- = 21 - 3 = 18$ y el valor observado de T es mín(3, 18) = 3. Observe que 3 excede el valor crítico de T, lo cual implica que no hay suficiente evidencia para indicar una diferencia en las dos distribuciones de frecuencia poblacional en las densidades de los pasteles. Como no podemos rechazar H_0 para $\alpha = .10$, sólo podemos decir que valor $p > .10$.

Tabla 15.2 Datos pareados y sus diferencias para el Ejemplo 15.3

A	B	Diferencia A − B	Diferencia absoluta	Rango de diferencia absoluta
.135	.129	.006	.006	3
.102	.120	−.018	.018	5
.108	.112	−.004	.004	1.5
.141	.152	−.011	.011	4
.131	.135	−.004	.004	1.5
.144	.163	−.019	.019	6

■

Aunque la Tabla 9, Apéndice 3, es aplicable para valores de n (el número de pares de datos) de hasta $n = 50$, es conveniente observar que T^+ (o T^-) estará distribuida normalmente en forma aproximada cuando la hipótesis nula sea verdadera y n sea grande (25 o más, por ejemplo). Esto hace posible que construyamos una prueba Z de muestra grande, donde si $T = T^+$,

$$E(T^+) = \frac{n(n+1)}{4} \quad \text{y} \quad V(T^+) = \frac{n(n+1)(2n+1)}{24}.$$

Entonces el estadístico Z

$$Z = \frac{T^+ - E(T^+)}{\sqrt{V(T^+)}} = \frac{T^+ - [n(n+1)/4]}{\sqrt{n(n+1)(2n+1)/24}}$$

se puede usar como estadístico de prueba. Entonces, para una prueba de dos colas y $\alpha = .05$, rechazaríamos la hipótesis de distribuciones poblacionales idénticas cuando $|z| \geq 1.96$. Para una prueba de una cola de que la distribución de las X está desplazada a la derecha (izquierda) de la distribución de las Y, rechace H_0 cuando $z > z_\alpha$ ($z < -z_\alpha$).

Una prueba de rangos con signo de Wilcoxon con muestras grandes para un experimento de observaciones pareadas: $n > 25$

Hipótesis nula: H_0: las distribuciones de frecuencia relativa poblacionales para las X y las Y son idénticas.

Hipótesis alternativa: (1) H_a: las dos distribuciones de frecuencia relativa poblacionales difieren en localización (una prueba de dos colas),

o bien, (2) la distribución de frecuencia relativa poblacional para las X está desplazada a la derecha (o izquierda) de la distribución de frecuencia relativa de las Y (pruebas de una cola).

Estadístico de prueba $Z = \dfrac{T^+ - [n(n+1)/4]}{\sqrt{n(n+1)(2n+1)/24}}$.

Región de rechazo: rechazar H_0 si $z \geq z_{\alpha/2}$ o $z \leq -z_{\alpha/2}$ para una prueba de dos colas. Para detectar un desplazamiento en las distribuciones de las X a la derecha de las Y, rechazar H_0 cuando $z \geq z_\alpha$. Para detectar un desplazamiento en la dirección opuesta, rechazar H_0 si $z \leq -z_\alpha$.

Ejercicios

15.10 Se lleva a cabo un experimento de observaciones pareadas que usa n pares de observaciones, si $T^+ =$ la suma de las columnas de los valores absolutos de las diferencias positivas y $T^- =$ la suma de las columnas de los valores absolutos de las diferencias negativas, ¿por qué es $T^+ + T^- = n(n+1)/2$?

15.11 Consulte el Ejercicio 15.10. Si T^+ se ha calculado, ¿cuál es la forma más fácil de determinar el valor de T^-? Si $T^+ > n(n+1)/4$, ¿es $T = T^+$ o T^-? ¿Por qué?

15.12 La siguiente tabla muestra las calificaciones de un grupo de 15 estudiantes en matemáticas y artes.

Estudiante	Matemáticas	Artes	Estudiante	Matemáticas	Artes
1	22	53	9	62	55
2	37	68	10	65	74
3	36	42	11	66	68
4	38	49	12	56	64
5	42	51	13	66	67
6	58	65	14	67	73
7	58	51	15	62	65
8	60	71			

a Use la prueba de rangos con signo de Wilcoxon para determinar si las localizaciones de las distribuciones de calificaciones para estos estudiantes difieren de manera importante para las dos materias. Establezca límites para el valor p e indique la conclusión apropiada cuando $\alpha = .05$.

b Exprese las hipótesis nula y alternativa para la prueba que realizó en el inciso a.

15.13 Consulte el Ejercicio 15.4. ¿Qué respuestas se obtienen si la prueba de rangos con signo de Wilcoxon se usa en el análisis de los datos? Compare estas respuestas con las obtenidas en el Ejercicio 15.4.

15.14 Consulte el Ejercicio 15.6(a). Conteste la pregunta usando la prueba de rangos con signo de Wilcoxon.

15.15 A ocho personas se les pidió ejecutaran un sencillo trabajo de ensamblar un rompecabezas en condiciones normales y en condiciones de estrés. Durante la condición de estrés, a las personas se les indicó que se les aplicaría una ligera descarga eléctrica 3 minutos después de empezar el experimento y cada 30 segundos de ahí en adelante hasta terminar el trabajo. Las lecturas de presión sanguínea se tomaron en ambas condiciones. Los datos de la siguiente tabla representan la lectura más alta durante el experimento.

Sujeto	Normal	Estrés
1	126	130
2	117	118
3	115	125
4	118	120
5	118	121
6	128	125
7	125	130
8	120	120

¿Los datos presentan suficiente evidencia para indicar lecturas de presión sanguínea más alta durante los condiciones de estrés? Analice los datos usando la prueba de rangos con signo de Wilcoxon para un experimento de observaciones pareadas. Proporcione el valor p apropiado.

15.16 Se emplearon dos métodos, A y B, para controlar el tránsito en cada uno de $n = 12$ cruceros durante una semana. Los números de accidentes que ocurrieron durante este tiempo se registraron en la siguiente tabla. El orden de uso (cuál método se empleó para la primera semana) se eligió aleatoriamente para cada crucero.

Crucero	Método A	Método B	Crucero	Método A	Método B
1	5	4	7	2	3
2	6	4	8	4	1
3	8	9	9	7	9
4	3	2	10	5	2
5	6	3	11	6	5
6	1	0	12	1	1

a Analice estos datos usando la prueba de signos.

b Analice estos datos usando la prueba de rangos con signo de Wilcoxon para un experimento de observaciones pareadas.

15.17 Investigadores en odontología han desarrollado un nuevo material para prevenir la caries, un sellador plástico que se aplica a las superficies de masticación de los dientes. Para determinar si el sellador es eficaz, se aplicó a la mitad de los dientes de cada uno de 12 niños en edad escolar. Después de 2 años se hizo un recuento del número de caries en los dientes con sellador y en los dientes no tratados. Los resultados se muestran en la siguiente tabla. ¿Hay suficiente evidencia para indicar que los dientes con sellador son menos propensos a la caries que los dientes no tratados? Pruebe usando $\alpha = 0.05$.

Niño	Con sellador	No tratado	Niño	Con sellador	No tratado
1	3	3	7	1	5
2	1	3	8	2	0
3	0	2	9	1	6
4	4	5	10	0	0
5	1	0	11	0	3
6	0	1	12	4	3

15.18 Consulte el Ejercicio 12.16. Con $\alpha = .01$, use la prueba de rangos con signo de Wilcoxon para ver si hubo una pérdida importante en la profundidad del humus entre el principio y el fin del estudio.

15.19 Suponga que $Y_1, Y_2, \ldots, Y_n$ es una muestra aleatoria de una función de distribución continua $F(y)$. Se desea probar una hipótesis respecto a la mediana ξ de $F(y)$. Construya una prueba de $H_0 : \xi = \xi_0$ contra $H_a : \xi \neq \xi_0$, donde ξ_0 es una constante especificada.

a Use la prueba de signos.

b Use la prueba de rangos con signo de Wilcoxon.

15.20 El vocero de una organización que apoya reducciones en impuesto a la propiedad, de cierta sección de una ciudad, expresó que el ingreso medio anual de los jefes de familia en esa sección era de $15 000. Una muestra aleatoria de diez jefes de familia de esa sección reveló los siguientes ingresos anuales:

14 800	16 900	18 000	19 100	13 200
18 500	20 000	19 200	15 100	16 500

Con $\alpha = .10$, pruebe la hipótesis de que el ingreso medio para la población de esa sección es $15 000 contra la alternativa de que sea mayor que $15 000.

a Use la prueba de signos.

b Use la prueba de rangos con signo de Wilcoxon.

15.5 Uso de rangos para comparar dos distribuciones poblacionales: muestras aleatorias independientes

Una prueba estadística para comparar dos poblaciones basada en muestras aleatorias independientes, la prueba de *suma de rangos*, fue propuesta por Frank Wilcoxon en 1945. De nuevo suponemos que estamos interesados en probar si las dos poblaciones tienen la misma distribución contra el desplazamiento (o localización) alternativo (véase la Sección 15.2). Supongamos que usted debe seleccionar muestras aleatorias independientes de n_1 y n_2 observaciones de las poblaciones I y II, respectivamente. La idea de Wilcoxon era combinar las $n_1 + n_2 = n$ observaciones y clasificarlas, en orden de magnitud, de 1 (la más pequeña) a n (la más grande). Los empates se tratan como en la Sección 15.4. Esto es, si dos o más observaciones están empatadas para el mismo rango, el promedio de los rangos que se hubiera asignado a estas observaciones se asigna a cada miembro del grupo empatado. Si las observaciones se seleccionaron de poblaciones idénticas, las *sumas de rango* para las muestras deben ser más o menos proporcionales a los tamaños muestrales n_1 y n_2. Por ejemplo, si n_1 y n_2 fueran iguales, se esperaría que las sumas de rango fueran casi iguales. En contraste, si las observaciones en una población, la I por ejemplo, tendían a ser mayores que las de la población II, las observaciones de la muestra I tenderían a recibir las clasificaciones más altas y la muestra I tendría una suma de rango mayor que lo esperado. Entonces (con los tamaños muestrales siendo iguales), si una suma de rango es muy grande (y, de modo correspondiente, la otra es muy pequeña), esto puede indicar una diferencia importante, desde el punto de vista estadístico, entre las localizaciones de las dos poblaciones.

Mann y Whitney propusieron en 1947 una prueba estadística equivalente que también empleaba sumas de rango para dos muestras. Como la prueba U de Mann–Whitney y las tablas de valores críticos de U se presentan con tanta frecuencia en la literatura, explicaremos su uso en la Sección 15.6 y daremos varios ejemplos de sus aplicaciones. En esta sección ilustramos la lógica de la prueba de la suma de rango y demostramos cómo determinar la región de rechazo para la prueba y el valor de α.

EJEMPLO **15.4** La cantidad de bacterias por unidad de volumen se muestran en la Tabla 15.3 para dos tipos de cultivos, I y II; se hicieron cuatro observaciones para cada cultivo. Con n_1 y n_2 represente el número de observaciones en las muestras I y II, respectivamente.

Para los datos dados en la Tabla 15.3, los rangos correspondientes se muestran en la Tabla 15.4. ¿Estos datos presentan suficiente evidencia para indicar una diferencia en la localización de las distribuciones poblacionales para los cultivos I y II?

Tabla 15.3 Datos para el Ejemplo 15.4

I	II
27	32
31	29
26	35
25	28

Solución Sea W igual a la suma de rango para la muestra I (para esta muestra, $W = 12$). Ciertamente, valores muy pequeños o muy grandes de W dan evidencia para indicar una diferencia entre las localizaciones de las dos distribuciones poblacionales; de aquí que W, la *suma de rango*, se pueda emplear como estadístico de prueba.

La región de rechazo para una prueba determinada se obtiene en la misma forma que para la prueba de signos. Empezamos por seleccionar los valores de W más contradictorios como la región de rechazo y sumamos éstos hasta que α sea de tamaño aceptable.

Tabla 15.4 Rangos

	I	II
	3	7
	6	5
	2	8
	1	4
Suma de rango	12	24

La suma de rangos mínima incluye los rangos 1, 2, 3, 4 o $W = 10$. Del mismo modo, la máxima incluye los rangos 5, 6, 7, 8, con $W = 26$. Por tanto, incluimos estos dos valores de W en la región de rechazo. ¿Cuál es el valor correspondiente de α?

Hallar el valor de α es un problema de probabilidad que se puede resolver usando los métodos del Capítulo 2. Si las poblaciones son idénticas, toda permutación de los 8 rangos representa un punto muestral y es igualmente probable. Entonces, α es la suma de las probabilidades de los mismos puntos (arreglos) que implican $W = 10$ o $W = 26$. El número total de permutaciones de los ocho rangos es 8! El número de arreglos diferentes de los rangos 1, 2, 3, 4 en la muestra I con los 5, 6, 7, 8 de la muestra II es $4! \times 4!$. De igual manera, el número de arreglos que ponen el valor máximo de W en la muestra I (rangos 5, 6, 7, 8) es $4! \times 4!$. Entonces, la probabilidad de que $W = 10$ o $W = 26$ es

$$p(10) + p(26) = \frac{(2)(4!)(4!)}{8!} = \frac{2}{\binom{8}{4}} = \frac{1}{35} = .029\,.$$

Si este valor de α es demasiado pequeño, la región de rechazo se puede agrandar para incluir las siguientes sumas de rangos más pequeña y más grande, $W = 11$ y $W = 25$. La suma de rango $W = 11$ incluye los rangos 1, 2, 3, 5, y

$$p(11) = \frac{4!\,4!}{8!} = \frac{1}{70}.$$

Del mismo modo,

$$p(25) = \frac{1}{70}\,.$$

Entonces,

$$\alpha = p(10) + p(11) + p(25) + p(26) = \frac{2}{35} = .057\,.$$

La expansión de la región de rechazo para incluir 12 y 24 aumenta considerablemente el valor de α. El conjunto de puntos muestrales que da un rango de 12 incluye todos los puntos muestrales asociados con clasificaciones de (1, 2, 3, 6) y (1, 2, 4, 5). Así,

$$p(12) = \frac{(2)(4!)(4!)}{8!} = \frac{1}{35},$$

y

$$\begin{aligned}\alpha &= p(10) + p(11) + p(12) + p(24) + p(25) + p(26)\\ &= \frac{1}{70} + \frac{1}{70} + \frac{1}{35} + \frac{1}{35} + \frac{1}{70} + \frac{1}{70} = \frac{4}{35} = .114.\end{aligned}$$

Este valor de α podría ser considerado demasiado grande para fines prácticos. Por tanto, estamos más satisfechos con la región de rechazo $W = 10, 11, 25$ y 26.

La suma de rangos para la muestra, $W = 12$, no cae en esta región de rechazo preferida, de modo que no tenemos evidencia suficiente para rechazar la hipótesis de que las distribuciones poblacionales de la cantidad de bacterias para los dos cultivos sean idénticas. ■