

14

Estadística no paramétrica

En este capítulo se desarrollan modelos estadísticos para realizar validaciones de todo tipo de magnitud biológica, tanto cualitativa como cuantitativa. Se desarrollan modelos equivalentes a los del capítulo anterior, sin las restricciones o supuestos de los mismos. El primer modelo es el de la Binomial, aplicable al caso de una sola muestra, y que puede ser aproximado con el de Gauss cuando las muestras sean grandes. El segundo modelo de rachas en una muestra sirve para validar la aleatoriedad en el orden o secuencia de extracción de muestras de una población. A continuación se desarrollan modelos para comparar dos muestras entre sí. Cuando las muestras sean apareadas se tienen dos modelos: el del Signo y el de Wilcoxon. Si bien este último es más poderoso que el otro porque usa mayor cantidad de información, el del Signo se presenta por su uso difundido en Análisis Clínicos y Farmacia. Cuando las muestras son independientes, se presenta el modelo paramétrico más poderoso, el modelo de la U de Mann-Whitney, casi tanto como su equivalente paramétrico: el modelo Student para dos muestras independientes.

14.1 Conceptos básicos

En los capítulos anteriores se han analizado modelos estadísticos que implican distribuciones continuas con ciertos supuestos básicos para la aplicación de estas técnicas. El principal uso de esos modelos es la estimación de parámetros desconocidos de la población en estudio, para poder hacer pruebas de validación o ensayos de significación y testear así las hipótesis planteadas. Estos supuestos se plantean fundamentalmente sobre el valor que toman los parámetros poblacionales o sobre comparaciones de dos de ellos. Hasta ahora se ha trabajado con magnitudes biológicas de tipo cuantitativas y continuas. A las magnitudes discretas se las ha tratado como proporciones para poder usar los modelos vistos, y cuando se usó el modelo de Gauss se tuvo que hacer una corrección por continuidad. A esta metodología de trabajo se la denomina Estadística Paramétrica, por contraposición a otra donde lo que interesa es comparar distribuciones en lugar de parámetros. Mientras los supuestos usados en la paramétrica especifican la distribución original (generalmente la gaussiana), hay otros casos en la práctica donde no se puede hacer esto, donde no se puede especificar la forma de distribución original. Se requiere entonces otra metodología de trabajo, una estadística de *distribuciones libres*, donde no se necesitan hacer supuestos acerca de la distribución poblacional, donde se puede comparar distribuciones entre sí o verificar supuestos a cerca de la forma de la población. Por ejemplo, verificar el supuesto de normalidad necesario para usar el modelo Student. La solución para estos casos es el empleo de la *Estadística no paramétrica*. Hay ciertas ventajas en su uso, tales como:

- trabajar con magnitudes *cualitativas*, además de las cuantitativas;
- estudiar casos donde no es posible precisar la naturaleza de la distribución;
- ídem para los casos donde los supuestos de la forma poblacional son débiles;

- aplicar el mismo modelo a casi todas las distribuciones en lugar a una sola;
- es más fácil de entender para quienes no poseen base matemática adecuada.

Y también tiene algunas desventajas como:

- cálculos usualmente más engorrosos;
- no extraen tanta información como los paramétricos si se aplican al mismo caso;
- son menos eficientes si las muestras son grandes.

Los modelos paramétricos tienen mayor capacidad para detectar diferencias muestrales que los no paramétricos. Es decir, son capaces de ver una diferencia significativa en casos donde los otros no pueden. Como su poder discriminador es mejor, *siempre que se pueda, conviene usar modelos paramétricos antes que los no paramétricos*, por su mayor sensibilidad para detectar diferencias significativas. A menos que las diferencias sean tan grandes que con cualquier modelo pueden detectarse. Pero como los no paramétricos se aplican casi todos los casos, son más fáciles de entender y no tienen tanta “complicación matemática”; se están poniendo de moda en Bioquímica y Farmacia cada vez más.

El independizarse de la forma de la población llevó a estos modelos a otras aplicaciones no clásicas, como en las ciencias de la conducta, marketing, ciencias sociales, etc. En algunas técnicas, como las pruebas de *rango* o de *orden*, se trabaja con puntajes, que no son verdaderamente numéricos, lo cual ocasiona deformaciones en los datos si se empleasen técnicas paramétricas y el valor de las conclusiones de la validación estadística quedaría menoscabado. Por ejemplo, se pueden asignar rangos por textura, coloración, sabor, olor (magnitudes organolépticas), clasificar por infección con cierto tipo de virus, y otros casos donde no se cumpla el supuesto de homogeneidad de varianzas; acá el modelo de rangos puede ser la salida. Cuando se comparan dos muestras, los modelos paramétricos hacen hincapié en la comparación de las medias, mientras que los no paramétricos fijan su atención en comparar medianas. A continuación se presenta el Modelo de la Binomial para trabajar en una sola muestra.

14.2 La prueba de la binomial

Esta prueba se emplea cuando los resultados del experimento se expresan en forma dicotómica. Los dos tipos de resultados se pueden clasificar en “éxito” o “fracaso” y aplicar la fórmula de la probabilidad binomial para poder calcular la probabilidad de ocurrencia de la hipótesis nula formulada. Comparando esta probabilidad B_x con el nivel de significación adoptado α por el investigador para el experimento, se procede a tomar la decisión de acuerdo con uno de los dos casos posibles siguientes:

- Si $B_x < \alpha$ se rechaza la H_0 , porque significa que la probabilidad de que ocurra esta hipótesis es muy baja, comparada con el nivel de significación adoptado para la validación.
- Si ocurre lo contrario y resulta $B_x > \alpha$, entonces no hay evidencia significativa como para rechazarla. Esto es especialmente útil para muestras pequeñas porque da la probabilidad exacta de la ocurrencia de los sucesos.

Cuando las muestras sean grandes es muy práctico usar la aproximación de la binomial con la normal. Una muestra se considera grande si $n > 25$. Hay tres maneras de resolver un problema de acuerdo con su tamaño muestral. El procedimiento del modelo Binomial se puede resumir en los pasos siguientes:

Paso 1) definir la cantidad de muestras o experimentos a realizar: n ;

Paso 2) medir las frecuencias de ocurrencia en ambos casos posibles (éxito para la menor);

Paso 3) plantear la hipótesis nula y su alternativa;

Paso 4) de acuerdo con tamaño muestral, seguir uno de estos tres caminos:

(I) Si $n \leq 25$ y si $p \cong q$ se usa la Tabla 8 del Anexo de Tablas Estadísticas para determinar el valor crítico en un ensayo de una sola cola $B_x = B_\alpha$. Para un ensayo de dos colas se duplica el valor de tablas $B_x = 2 B_\alpha$. El estadígrafo comparativo es la probabilidad para la frecuencia menor obtenida; o sea, con el valor x (frecuencia menor) se saca de tablas el valor B_α y se saca B_x según si es de una o dos colas. Luego si $B_x < \alpha$ se rechaza H_0 .

(II) Si $n \leq 25$ y si $p \neq q$ habrá que calcular la probabilidad con la fórmula de la binomial de acuerdo con la hipótesis nula. Esto es, se debe obtener la probabilidad acumulativa desde 0 hasta el valor predicho x (o su complementaria si es más corta). De esa forma se calcula B_x . Luego es:

$$P_B = \sum_0^x C(n, x) \cdot p^x (1-p)^{n-x}$$

Entonces se compara P_B con $(1 - P_B)$ y al menor de ellos se lo llama B_x . Este valor se compara con el nivel de significación α adoptado por el investigador, y se toma la decisión de aceptar o rechazar H_0 como se vio más arriba.

(III) Si $n > 25$ se trata de una muestra "grande" y se usa la aproximación normal a la binomial. Es decir, se transforma el valor x de frecuencia observada a unidades tipificadas (Z). Cuidando de efectuar la corrección por continuidad y de tablas se obtiene su probabilidad B_x para compararlo con el nivel de significación α adoptado. O sea:

$$Z = \frac{(X \pm 0,5) - \mu_r}{\sigma_r} = \frac{(X \pm 0,5) - n \cdot p}{\sqrt{n \cdot p \cdot q}}$$

El signo \pm se usa para mostrar las dos alternativas de corrección por continuidad. Si el ensayo es de una sola cola y la H_0 es $x \leq n p$ se usa $(X + 0,5)$. Por su parte, si la H_0 es: $x \geq n p$ se usa $(X - 0,5)$. Una vez calculado Z , se busca en la tabla de la normal su probabilidad B_x la que se compara con el nivel α para poder decidir. En el cuadro 14.1 se presentan ejemplos ilustrativos de cada caso.

CUADRO 14.1: La prueba de la Binomial para una sola muestra

Caso 1 - Muestras pequeñas con equiprobabilidad:

Ejemplo 1) En una maternidad ocurrieron 18 nacimientos en una semana, 11 de los cuales fueron varones. La hipótesis de trabajo es que la proporción de sexos es la habitual.

Si bien este problema se puede resolver usando la fórmula de la binomial, como n es menor que 25 y las probabilidades son aproximadamente iguales, entonces es más práctico usar la Tabla 8 del Anexo. Para un tamaño muestral $n = 18$ y para la frecuencia menor $x = 7$, resulta un valor de probabilidad $B_x = 0,24 > \alpha$. Por lo tanto, no se puede rechazar la hipótesis nula. No hay evidencia significativa que pruebe que la proporción de sexos no es la habitual.

Ejemplo 2) Un profesor le enseñó a un grupo de 20 alumnos dos modelos estadísticos: el A y el B para resolver un mismo problema en la clase de coloquios de la siguiente manera: Seleccionó 10 al azar y les enseñó primero el A y más tarde el B; con los demás procedió al revés: primero el B y luego el A. Finalizado el examen parcial, les tomó a los mismos 20 alumnos el problema explicado en coloquios. Se supuso que por el estrés del examen habría una tendencia a usar el método aprendido primero. Los resultados mostraron 17 con este y 3 con el método aprendido segundo.

H_0 : No hay preferencias por usar el método aprendido antes o después ($p = q = 1/2$).

H_1 : Prefieren el aprendido primero ($p > q$).

Los datos a usar son $n = 20$; $p = q = 0,5$; la menor frecuencia es $x = 3$ entonces usando la Tabla de la Binomial se obtiene una probabilidad $B_x = 0,001$ que es mucho menor que $\alpha = 0,05$. Por lo tanto se rechaza la H_0 y se concluye que hay una especie de regresión hacia lo aprendido primero.

Caso 2 - Muestras pequeñas sin equiprobabilidad: en un laboratorio que fabrica medicamento en ampollas, históricamente 10% de la producción presenta fallas de cierre hermético. Para probar la calidad de un lote fabricado se extraen 5 muestras al azar y se verifican. Si encuentran por lo menos una fallada se devuelve todo el lote, en caso contrario sale a la venta. Se desea saber si con esta política se tiene un nivel aceptable (1%) de rechazos.

Este problema se resuelve con la aplicación de la fórmula binomial, con $p = 0,1$ y $q = 0,9$

$$B_x (x < 1) = P(x = 0) = C(n, x) \cdot p^x (1 - p)^{n-x} = C(5, 0) (0,1)^0 (0,9)^5 = 0,59$$

Hay una probabilidad del 59% de no encontrar ninguna fallada. O sea, en el 41% de los casos habría devolución, lo cual no es conveniente. Si se cambia la política y se rechazan lotes con dos o más falladas. Eso significa que se aceptarían lotes con una o ninguna fallada. Esto es:

$$B_x (x < 2) = P(x = 0) + P(x = 1) = 0,59 + C(5, 1) (0,1)^1 (0,9)^4 = 0,59 + 0,328 = 0,918$$

Ahora, con esta nueva política se rechazarían lotes en el 8% de los casos lo cual es más aceptable. Si se la hace más amplia y se rechazan lotes con tres o más falladas, se tiene:

$$B_x (x < 3) = P(x = 0) + P(x = 1) + P(x = 2) = 0,59 + 0,328 + C(5, 2) (0,1)^2 (0,9)^3 = 0,9909$$

Entonces esta última política es la buscada.

Caso 3 - Muestras grandes: en 90 pacientes se probó un nuevo medicamento y 10 de ellos no se curaron en el plazo previsto. La idea es aceptar la droga si logra mejoría en el 75% de los casos. ¿Qué decisión se debe tomar a la luz de estos resultados experimentales?

Como la muestra es grande, conviene usar la aproximación normal a la binomial. Entonces,

$H_0 : x \leq n \cdot p = 67,5$ casos curados y la droga no sirve.

$H_1 : x > 67,5$ Y La droga sirve.

$$\mu_r = n \cdot p = 90 \cdot 0,75 = 67,5 ; SE(r) = \sigma_r = \sqrt{n \cdot p \cdot q} = \sqrt{90(0,75)0,25} = 4,11 ; X = (90-10) = 80$$

$$Z = \frac{(X \pm 0,5) - \mu_r}{\sigma_r} = \frac{(X + 0,5) - n \cdot p}{\sqrt{n \cdot p \cdot q}} = \frac{(80,5) - 67,5}{4,11} = 3,163^{***}$$

Se tiene evidencia altamente significativa de la bondad del medicamento probado. Por otro lado $x = 67,55$ no cae dentro del 99,9% CI (70,4 ; 89,6).

NOTA: si se hubiese calculado con el programa Excel resultarían:

- 1) la aproximación normal con $Z = 3,163$, da una probabilidad de 0,9997028;
- 2) la probabilidad binomial acumulada, que es la exacta da un valor de 0,9992192.

Esto es, $Bx = 0,00078 \lll \alpha = 0,001$ para el 99,9% de confianza y se rechaza H_0 .

14.3 La prueba de rachas de una muestra

Esta prueba es para aleatoriedad. En los casos anteriores vistos se trabajó bajo el supuesto de extracción de muestras aleatorias de la población. De acuerdo con el diseño del experimento se puede lograrlo, sin embargo, hay casos donde no es tan sencillo hacerlo. Otras veces puede que sea necesario probar la aleatoriedad. Para ello se han venido desarrollando una serie de modelos estadísticos que estudian el *orden o secuencia* en que las muestras individuales fueron obtenidas, para probar que la muestra es aleatoria. Este modelo se basa en la cantidad de *rachas* que una muestra exhibe. Por rachas se entiende a una sucesión de símbolos idénticos que pueden estar separados o no por otro tipo de símbolos. Por ejemplo, sea una serie de mediciones de magnitudes dicotómicas identificadas con los símbolos de resultado positivo (+) o negativo (-) a juicio del investigador, de acuerdo con cierto criterio profesional empleado:

Resultados: $\frac{++}{1} \frac{---}{2} \frac{+}{3} \frac{---}{4} \frac{++}{5} \frac{-}{6} \frac{+}{7}$
 N° de rachas:

La primera racha empieza con una serie de 2 símbolos positivos, la segunda racha con 3 negativos, la tercera con un positivo, y así sucesivamente hasta la séptima racha con un positivo. El número de rachas es $r = 7$. El número total de rachas indica si una muestra es o no aleatoria. Si se da un número pequeño de rachas puede deberse a una falta de independencia o a una tendencia

temporal. Mientras que si por el contrario hay un número muy grande de rachas, las fluctuaciones cíclicas sistemáticas, en un período corto de tiempo, pueden causar influencia en los valores asignados por el investigador. Por ejemplo, si se lanza al aire una moneda 30 veces y se obtienen 30 rachas es razonable dudar de esa moneda, son demasiadas. Lo mismo ocurriría si se obtienen únicamente dos rachas, como primero todas caras y luego secas.

El número de rachas no depende de la frecuencia de los sucesos. En efecto, dos situaciones bien diferentes pueden tener las mismas frecuencias. Como las del ejemplo anterior de las monedas donde se tiene una frecuencia relativa de $\frac{1}{2}$, en ambos casos. Sin embargo, en el primer caso sale una cara, luego una seca, y así sucesivamente una y una, mientras que en el segundo primero se dan 15 caras seguidas y luego las 15 secas finales. Generalmente se conoce la distribución muestral en muestreos repetidos, lo que permite obtener la probabilidad asociada para poder probar hipótesis. El *procedimiento* para aplicar este modelo es como sigue:

Paso 1) Se calcula el número n_1 de elementos de una clase identificadas por un símbolo y n_2 la cantidad de elementos de la otra.

Paso 2) Se ordenan los $n = n_1 + n_2$ sucesos en el orden en que ocurrieron.

Paso 3) Se cuenta el número r de rachas.

Paso 4) Se postula una hipótesis nula de trabajo que permita calcular la probabilidad asociada.

Paso 5) Se determina la probabilidad que ocurran r rachas, usando H_0 , y se compara con el nivel de significación α adoptado para aceptar o rechazar la H_0 .

Para determinar el valor de la probabilidad se presentan dos casos según el tamaño de la muestra:

(I) *Muestras pequeñas*: si los tamaños muestrales de ambos resultados son menores que 20 se emplea la Tabla 9 del Anexo. En la primera de las tablas, ubicada en la parte superior, se dan valores de r para una (H_0) de menor o igual, tal que su probabilidad asociada es $p = 0,05$ para dos colas (para una cola es la mitad: $p = 0,025$). En la parte inferior se presenta una segunda tabla, es para valores de r con (H_0) de mayor o igual, tal que su probabilidad asociada para dos colas es $p = 0,05$. A la de arriba se la denota con $R_{\text{mín}}$. Y a la de abajo con $R_{\text{máx}}$.

(II) *Muestras grandes*: si alguno de los tamaños muestrales es mayor a 20, se determina el valor de estadígrafo comparación Z con una distribución aproximadamente normal:

$$Z = \frac{r - \mu_r}{\sigma_r} \quad \text{el cual se compara como siempre con } z_\alpha \text{ de Gauss.}$$

Con:

$$\mu_r = 1 + [(2 n_1 n_2) / n]$$

$$\text{VAR}(r) = (\sigma_r)^2 = \frac{(2 n_1 n_2) (2 n_1 n_2 - n)}{n^2 (n - 1)}$$

Para ilustrar mejor estas ideas, se presentan dos problemas resueltos en el Cuadro 14.2 siguiente, uno para muestras pequeñas y el otro para las grandes.

CUADRO 14.2: La Prueba de rachas para una sola prueba

Caso 1 - Muestras pequeñas: en un Laboratorio de investigación se prueba un antiinflamatorio nuevo. Los resultados son aceptables si al segundo día de aplicado al paciente se observa una reducción del 90% en la inflamación; se le asigna (+) a ese caso. Se quiere testear la hipótesis que la sucesión de signos positivos y negativos se produce al azar. La sucesión de los 24 casos analizados fue:

Sucesión: + - + + + + - + + + - - - - + - - + + + - - - -
 Rachas: 1 2 3 4 5 6 7 8 9 10

Se usa un test dos colas porque no se predice la dirección de la desviación que supone. El tamaño de cada muestra es 12. Usando ambas tablas del apéndice se determina la zona de rechazo de la hipótesis nula cuando r es menor o igual a 7 o cuando r es mayor o igual a 19; con esto se define:

Zona de aceptación: $7 < r < 19$ y como $r = 10$ cae dentro de esta zona y no se rechaza (H_0).

Se concluye que se deben suponer aleatorias a las muestras tomadas.

Caso 2 - Muestras Grandes: en un hospital se forma todas las mañanas temprano, una cola de pacientes esperando su turno para la extracción de sangre. La bioquímica a cargo decide verificar si la colocación de hombres y mujeres es al azar. Anota el sexo de cada uno de los primeros 50 pacientes que entraron al laboratorio. Los resultados fueron:

Sucesos : HH M H M HHH MM H MM H M HH MMM HH MM HH
 Rachas : 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Sucesos : MM H M H M H MM H M HH M HH M H M H M H MM
 Rachas : 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34

Total de rachas: $r = 34$. El total de hombres y mujeres fue: $25 = n_1 = n_2$

$$\mu_r = 1 + [(2 n_1 n_2) / n] = 1 + [2 (25) 25] / 50 = 26$$

$$(\sigma_r)^2 = \frac{(2 n_1 n_2) (2 n_1 n_2 - n)}{n^2 (n - 1)} = \frac{2 (25) 25 \{ [2 (25) 25] - 50 \}}{(50)^2 (50 - 1)} = 12,245 \quad \text{y} \quad \sigma_r = 3,5$$

$$Z = (r - \mu_r) / \sigma_r = (r - 26) / 3.5 = (34 - 26) / 3.5 = 2,29^* \quad (0,01 < P_N(Z_{2,29}) = 0,011 < 0,05)$$

Se rechaza la hipótesis de que los sexos guardan un orden aleatorio al formar la fila. Se encontró evidencia significativa para rechazarla. Si se calcula la probabilidad gaussiana para $Z = 2,29$ resulta $p = 0,011$; esto es, más cerca del rechazo al 99% que al 95% de nivel de confianza.

La prueba de aleatoriedad algunas veces es necesaria para comprobar los supuestos específicos de otros modelos estadísticos. Debería realizarse primero para decidir si el modelo es aplicable porque se verifica su supuesto de aleatoriedad, como en el caso Gauss, Student, Fisher, etc. Para evitar en la prueba de aleatoriedad, lo aconsejable es usar sorteos al azar para la selección de muestras.

14.4 Modelos para dos muestras apareadas

Los modelos no paramétricos para realizar comparaciones entre muestras apareadas son dos: el modelo del Signo se usa cuando la magnitud biológica medida en el experimento es de tipo dicotómica. El modelo de Wilcoxon es más potente porque usa mayor información: el signo y el rango de las diferencias halladas. Ambos modelos usan el *signo* de las diferencias entre los pares de muestras testeados.

Cuando el investigador asigna puntajes a las mediciones que realiza, basado en algún criterio profesional, el problema es traducir eso a una dicotomía. La manera de hacerlo es calcular la mediana de los puntajes y asignar un signo positivo a los resultados mayores y negativo a los menores. Entonces, se ha “dicotomizado” la variable “puntajes” medida y se pueden aplicar estos modelos. Los puntajes se suelen usar mucho en ciencias sociales y humanas, donde el comportamiento es juzgado de acuerdo con una cierta escala. También en las notas de los exámenes en el sistema educativo. Puede hacerse lo mismo con magnitudes cualitativas y cuantitativas, o sea que el campo de aplicación se extiende a cualquier magnitud medida. En el caso dicotómico es directo. En el caso politómico hay que separar en dos grandes grupos, de acuerdo con algún criterio profesional adecuado. En el caso ordinal se procede en forma análoga. Por ejemplo, en los antibiogramas donde lo que se mide es el diámetro de acción del antibiótico en una caja de Petri, los resultados se expresan con tres valores: resistente, poco sensible y muy sensible. Como lo que interesa son los antibióticos potentes, la clasificación es (+) para los calificados como muy sensibles y (-) para los otros dos, donde aplicar tal antibiótico produce poco o ningún resultado en el paciente. Las magnitudes organolépticas (turbidez de orina, sangre en heces, olor color, etc.) son fácilmente adaptables a este modelo. Finalmente, en el caso de magnitudes continuas se debe definir un punto de corte en la variable con la que se está trabajando. Por ejemplo, si se está midiendo el peso de los pacientes, se puede definir: (+) hasta 60 kg (Flacos), y (-) más de 60 kg (Gordos); con la temperatura corporal puede ser (+) hasta 37 °C (normal), y (-) más de 37 °C (afiebrado). Lo cual muestra que *toda magnitud biológica puede ser dicotomizada*.

Las muestras apareadas se diseñan para el caso clásico en que se usa un mismo individuo como su propio control, para hacer comparaciones entre dos casos de un cierto factor en estudio. Por ejemplo, las del tipo “antes-después” donde un factor cualquiera es el tratamiento que se le aplica al individuo, el caso de medirle el peso antes del régimen de adelgazar y después del mismo, o su velocidad para correr antes del entrenamiento y un régimen de comida especial y después, o las medicaciones en Medicina, etc. Otra forma muy común en las investigaciones farmacológicas es la del tipo “droga *versus* placebo”. En Bioquímica se solía recomendar para comparar dos técnicas de laboratorio; ver si la nueva técnica es equiparable a la vieja. Usando a los mismos pacientes que vienen a hacerse determinaciones diariamente, se los mide con ambas técnicas para comparar. El costo de esta manera es más bajo porque una de las dos determinaciones que se le hacen a cada paciente es la de rutina. En cambio, el modelo más potente de Student pa-

ra muestras independientes requiere usar más reactivos, porque el mismo paciente no puede aparecer dos veces en los datos para no perder independencia. Como regla general, se debe recordar que la mejor forma de comparar dos métodos o rutinas de laboratorio es usando mediciones repetidas del mismo valor patrón. Por ejemplo, fraccionando en alícuotas a un mismo suero control para realizar el experimento de comparación. Si la homogeneización se hace correctamente, no hay que preocuparse por los supuestos gaussianos, recordando el Teorema Central del Límite.

También sirve para testear alguna parte de un método con dos variantes; como diferentes marcas comerciales, personal, equipos, instrumentos, etc. Por ejemplo, probar el resultado de dos métodos de fabricación de medicamentos, dos vendedores de una farmacia, dos marcas comerciales de analgésicos, etc. De lo antedicho, se puede inferir el amplio campo de aplicación en la práctica de estos modelos. Su gran limitación es que no se pueden usar para comparar más de dos muestras a la vez.

14.4.1 Modelo del Signo

Esta prueba se basa en los signos de las diferencias observadas entre valores apareados. La independencia se resguarda si la unidad de muestreo es el resultado de comparar a las parejas entre sí, en lugar de usar las observaciones individuales. Cada pareja de datos debe ser independiente de las otras, y la forma de lograrlo fácilmente es seleccionado al azar al elegir los componentes de las muestras. Una ventaja es que puede usarse cuando las observaciones pareadas están simplemente ordenada por rangos. No es necesaria la homogeneidad de las varianzas, ni que las muestras sean extraídas de la misma población. Su desventaja es que elimina mucha información pues la reduce a una dicotomía. Otras ventajas son: la facilidad de aplicación, la reducción de tiempo y costo para el investigador. Los supuestos básicos para aplicar este modelo son:

- La variable subyacente en las determinaciones es continua;
- las pares de muestras son independientes;
- la hipótesis nula es: los signos se distribuyen al azar alrededor de la mediana.

Este último implica que las diferencias deben ser tales que aproximadamente la mitad de ellas sean positivas y la otra mitad negativas. Por supuesto que puede haber *ligas*, esto es, parejas de muestras cuya diferencia sea nula. Para el modelo solo interesan las otras. Se toma en cuenta el número n de diferencias halladas (n_+ : cantidad de signos positivos y n_- : número de negativos). Donde $n_+ + n_- = n$. Este valor puede ser a lo sumo igual al número total N de parejas testeadas. O sea, $N = n + e$ (donde e es la cantidad de empates o ligas encontrados). En resumen, el procedimiento para aplicar este modelo sigue los pasos:

Paso 1) Se efectúan las parejas de mediciones y se determina el signo de diferencia.

Paso 2) Se cuenta el número de (+) y el de (-) hallados y así se calculan: $n_+ + n_- = n$.

Paso 3) Se elige al menor de ellos como $x = n_{\text{mínimo}}$ para efectuar el test.

Paso 4) Dependiendo del tamaño de la muestra se siguen dos caminos:

(I) *Muestras pequeñas* ($n \leq 25$): se usa la Tabla 8 del Anexo de Tablas Estadísticas para determinar el valor crítico en un ensayo de una sola cola $B_x = B_{\alpha}$. Para un ensayo de dos colas se duplica el valor de tablas $B_x = 2 B_{\alpha}$. El estadígrafo comparativo es la probabilidad para la frecuencia menor obtenida, o sea con el valor x (frecuencia menor) se saca de tablas el valor B_{α} y se calcula B_x según si es de una o dos colas. Luego si $B_x < \alpha$ se rechaza H_0 .

(I) *Muestras grandes* ($n > 25$): se usa la aproximación normal a la binomial, para determinar Z en forma análoga a las vistas antes, con la debida corrección por continuidad. Como la (H_0) supone una igualdad entre la cantidad de positivos y negativos, se toma $p = q = 0,5$ y así:

$$Z = \frac{(r \pm 0,5) - \mu_r}{\sigma_r} = \frac{(r \pm 0,5) - n p}{\sqrt{n \cdot p \cdot q}} = [(2r \pm 1) - n] / \sqrt{n} \text{ versus } Z_{\alpha}$$

Ejemplo 1) Muestras pequeñas: en un laboratorio se desea testear una nueva técnica. Para ello se toman las determinaciones hechas en el día con la manera usual (método viejo) y se las replica con el nuevo procedimiento, obteniéndose los valores:

Paciente	Nuevo	Viejo	Signo
1	2,3	2,2	+
2	2,6	2,5	+
3	3,5	3,4	+
4	2,9	2,9	0
5	2,5	2,4	+
6	3,3	3,1	+
7	1,7	1,8	-
8	2,8	2,8	0
9	2,9	2,9	0
10	3,1	3,2	-
11	2,4	2,3	+
12	3,2	3,3	-
13	2,5	2,4	+
14	2,6	2,5	+
15	3,3	3,2	+
16	2,9	2,8	+
17	2,2	2,1	+

$N = 17$
 $(+) = 11 = n_+$
 $(-) = 3 = n_-$
 $(0) = 3$
 Entonces
 $n = 14$
 $x = 3$

Del total de pruebas $N = 17$ solo interesan los $n = 14$ que no fueron empates ($e = 3$). Con los 14 casos restantes se tienen : $n_+ = 11$ signos positivos y $n_- = 3$ negativos; por lo tanto, se toma el valor mínimo como $x \leq 3$. Entrando a la Tabla 8 con $n = 14$ y $x = 3$, resulta un valor de probabilidad binomial: $B_x = 0,029$ que es menor que el nivel de significación $\alpha = 0,05$, por lo que se debe rechazar la H_0 . Esto es, se tiene evidencia científica que ambos métodos no son equivalentes para las mediciones realizadas. Por la cantidad de signos + parece que el método nuevo arroja valores más grandes que el viejo. Para continuar, es conveniente obtener un suero control y revisar la calibración de los métodos.

Ejemplo 2) Muestras grandes: el centro de estudiantes de cierta Facultad piensa que una de sus autoridades debe ser removida de su cargo. Para ello planifica una campaña de pegatinas y una asamblea estudiantil ad-hoc, donde sus seguidores harán acotaciones significativas para convencer a los concurrentes. Uno del grupo, que cursó Bioestadística, duda de la efectividad del plan y decide buscar evidencia científica para decidir si tuvo efecto la asamblea. Prepara un cuestionario a los asistentes para detectar los cambios "antes y después" de la reunión, y selecciona al azar a 100 de ellos para encuestarlos.

Los casos que interesan son los $30 + 42 = 72$ casos de cambios de opinión. Si la asamblea no afectara a sus concurrentes, cerca de la mitad de los que cambiaron de opinión hubieran pasado de SÍ a NO y la otra mitad de NO a SÍ. Esto es: H_0 : Se esperan 36 cambios porque la asamblea no tuvo efecto. Los casos que interesan son los que cambiaron de opinión $n = 30 + 42 = 72$.

		Antes		
		SI	NO	
	SI	9	42	$\mu_r = n p = 72 (0,5) = 36$ $\sigma_r = \sqrt{n.p.q} = \sqrt{72 (0,5)0,5} = 4,243$ $r = 42$ cambiaron de opinión de Si a NO
Después	NO	30	19	$Z = [(42 - 0,5) - 36] / \sqrt{72} = 1,3$

Notar que si se hubiera tomado $r = 30$, el resultado sería el mismo: $Z = 1,3$ (no significativo). Como cae en la zona de aceptación con un nivel del 95% ($Z_\alpha = 1,96$), no se puede rechazar la hipótesis nula. La conclusión es que a pesar de los 42 casos que pasaron de NO a SÍ no hay evidencia significativa como para pensar que la asamblea fue útil a los fines estudiantiles.

Ejemplo 3) Muestras grandes: un farmacéutico que ha realizado transformaciones en el aspecto de su negocio, realiza una encuesta entre sus clientes habituales para conocer su opinión respecto a los cambios. Descarta a los que no emitieron opinión y a los que no tenían un conocimiento previo de su empresa. Elige al azar 60 entre los remanentes. Encontró que 48 de ellos emitieron una opinión que considera positiva y 12 negativas. Ver si obtuvo pruebas de que la imagen de su empresa ha mejorado.

(H_0) $p = q = 0,5$ las opiniones emitidas se deben al azar y no hay mejora de imagen.

De los datos surgen: $\mu_r = n p = 60 (0,5) = 30$; $\sigma_r = \sqrt{n.p.q} = 3,873$ y $r = 48$

$$Z = [(r \pm 0,5) - \mu_r] / \sigma_r$$

Reemplazando con los de opinión positiva es: $Z = [(48 - 0,5) - 30] / 3,873 = 4,52***$

Reemplazando con los de opinión negativa es: $Z = [(12 + 0,5) - 30] / 3,873 = - 4,52***$

Y también se puede calcular con la otra fórmula: $Z = [2 (48 - 0,5) - 60] / \sqrt{60} = 4,52***$

Por lo tanto, se rechaza la H_0 de que el cambio de opinión se debe al azar, pues tiene evidencia altamente significativa que justifica su esfuerzo por mejorar la imagen de su farmacia.

14.4.2 La prueba de rangos de Wilcoxon

En la prueba del Signo se utiliza la información acerca de la dirección de las diferencias encontradas en la pareja de muestras. Pero no se considera la *magnitud* relativa de tales diferencias. En cambio, en el modelo de Wilcoxon se toma en cuenta ambas cosas y por eso es más poderoso que el del Signo. Acá se le da un peso a cada signo, relativo a la magnitud de la diferencia encontrada. Se la denomina también: *Prueba de rangos señalados y pares encontrados*. El investigador debe hacer dos cosas básicas al examinar una pareja de datos:

- 1) *determinar en la pareja cual es el "mayor" de ambos.*
- 2) *ordenar por rango las diferencias halladas* (rango acá significa "ranking" u orden)

Entonces puede usar este test para validar la hipótesis nula de que no hay diferencias entre los pares debido al tratamiento aplicado. En casos donde la valoración se hace en forma subjetiva, usando puntajes, también se puede aplicar el modelo de Wilcoxon (Psicología, Sociología, etc.). En la Bioquímica se la puede emplear cuando se valora el tamaño de una reacción ante una droga, vacuna, coloración, etc., con algo similar a un puntaje. En Farmacia ocurre otro tanto cuando se valoran con puntos la eficacia de medicamentos, o con escalas las encuestas de opinión en técnicas de mercadeo, propaganda, etc. *Esta prueba también puede usarse en una muestra única donde deseen hacer inferencias acerca de la mediana:* aquí al valor supuesto en la hipótesis nula se le resta a cada observación realizada y se tienen las diferencias buscadas para hacer el test.

El supuesto básico para poder usarlo es:

las magnitudes con las que se trabaja provienen de una distribución simétrica.

No importa si cada muestra proviene de una población distinta, lo importante es que ambas deben provenir de poblaciones con distribuciones simétricas. El método puede ser resumido en:

Paso 1) Se determina las diferencias D_i entre los pares de observaciones realizadas.

Paso 2) Sin tomar en cuenta el signo, se ordenan en forma creciente. Las ligas se descartan pues en este modelo no se toman en cuenta los empates únicamente se consideran los rangos de las diferencias encontradas (notar que ahora *rango* no significa lo mismo que lo visto en el capítulo 4).

Paso 3) Se coloca el signo a cada uno de los rangos R_i hallados. Se suman entre sí los rangos de las diferencias positivas calculando su total T_+ , y el de las negativas obteniendo T_-

Paso 4) Se elige la menor de ambas sumas y se la define como el estadígrafo T .

Paso 5) Se compara el valor T obteniendo con el valor crítico T_α para tomar decisiones.

La hipótesis nula es que los tratamientos aplicados son equivalentes. Esto es, la suma de los rangos positivos y negativos son aproximadamente iguales. Habrá diferencias de ambos signos pero con valoraciones que se equiparan. En cambio, si las sumas de rangos son muy diferentes, se puede deducir que el efecto del factor analizado no es despreciable ni producto del azar.

En este modelo puede haber dos tipos de ligas. La primera ocurre cuando no se aprecia diferencia entre la pareja analizada ($D_i = 0$) y se deja de lado en los cálculos, al igual que en el caso del modelo anterior. La segunda ocurre cuando se produce un empate en el valor de las diferencias ($D_i = D_j = \dots = D_k$). O sea, la diferencia de rangos no es nula sino que tiene el mismo módulo que otra diferencia. Es posible hallar 1, 2, 3, ..., k empates. En este caso, la solución es promediar el valor de los rangos empatados hallados E_i y asignárselo a cada uno de los empates. Esto es, a cada empate se le da un valor $E_i = (R_1 + R_2 + \dots + R_k) / k$.

De acuerdo con el tamaño muestral se pueden presentar dos casos:

(I) *Muestras pequeñas* ($6 \leq n \leq 25$): en estos casos se usa la Tabla 10 del Anexo de rangos señalados de Wilcoxon. Allí se entra con el tamaño muestral y el nivel α , según sea de una o dos colas, para encontrar el valor crítico T_α , el cual se compara con el valor T calculado como la menor, de la suma de los dos rangos positivo y negativo. Si $T \leq T_\alpha$ se puede rechazar la H_0 al nivel de significación α adoptado, caso contrario se acepta. Por ejemplo, si $n = 15$ y $T = 21$ entonces $T_\alpha = 25$ es mayor y se rechaza H_0 al nivel de 0,05 en dos colas. En este caso se rechaza la hipótesis de que ambos grupos no difieren entre sí y se concluye que hay evidencia significativa que prueba que hay diferencia entre ambos. No se puede rechazar al nivel 0,01.

(II) *Muestras grandes* ($n > 25$): cuando la muestra es grande se usa la aproximación normal pues se ha demostrado que para $n > 25$ el estadígrafo T se aproxima a la función de Gauss en forma aceptable. Para tipificar el estadígrafo se usa:

$$Z_T = \frac{T - \mu_T}{\sigma_T} \quad \text{y se lo compara con } Z_\alpha \text{ como es habitual, con:}$$

$$\mu_T = (n/4)(n+1) \quad \text{y} \quad \sigma_T^2 = (n/24)(n+1)(2n+1) = \text{VAR}(T)$$

Donde n es el total de parejas de datos medidas útiles, esto es aquellas parejas cuyas diferencias no son nulas. Entonces, $n = N - D(0)$. Los siguientes ejemplos ilustran el uso de este modelo:

Caso 1) Muestras pequeñas: una determinación clínica se realizó en 8 pacientes que concurren al laboratorio usando el método A; se repitió la medición pero usando otra marca de espectrofotómetro (método B). Los resultados obtenidos fueron los del cuadro siguiente. Averiguar si es lo mismo usar uno u otro espectro.

Paciente	Método A	Método B	Diferencias D_i	Rangos R_i	T
1	87	76	11	8	T = 4
2	66	56	10	7	
3	75	76	-1	-1	
4	95	86	9	6	
5	73	76	-3	-3	
6	88	86	2	2	
7	69	64	5	4	
8	77	70	7	5	

La menor de las sumas de rangos señalados es $T = 4$. Como no hay diferencias nulas encontradas, resulta $N = n = 8$, de la Tabla 10 del Anexo se obtiene el valor $T\alpha = 4$. Como este estadígrafo no es menor que el valor crítico, se rechaza la hipótesis nula de igualdad entre ambos espectros con un nivel significación de 0,05 para un ensayo de dos colas.

Este mismo problema se puede resolver con el modelo del signo. Para ello, se calcula un valor de $x = 2$ y de la tabla respectiva se obtiene una posibilidad asociada $Bx = 2p = 2(0,145) = 0,029$ para una prueba de dos colas. Acá no se puede rechazar la hipótesis nula (H_0) $x = 2$ para un valor de $\alpha = 0,05$. A primera vista se tienen entonces dos resultados diferentes usando estos dos modelos. Es importante destacar que no son contradictorios entre sí. El modelo del Signo tiene menor sensibilidad y no puede detectar diferencia alguna entre ambos métodos clínicos. No puede discriminar la pequeña diferencia que implica cambiar el espectrofotómetro en la técnica realizada. En cambio, el modelo de Wilcoxon, más sensible, ya detecta diferencias entre ambos. Si bien en el límite, pero suficiente para tener una prueba científica. Esto no es extraño pues este modelo, además de emplear la información del signo, agrega más información con los rangos. Es una especie de moderación o prorrateo de los signos: no todos pesan igual a la hora de contarlos.

Si se pudiese aplicar un modelo paramétrico porque se cumplen los supuestos básicos, el apropiado sería el modelo de Student para comparar dos muestras apareadas, visto en el capítulo anterior. En tal caso resultaría:

$$t = \frac{\bar{d} - 0}{DS_d / \sqrt{n}} = \frac{5 - 0}{5,21 / \sqrt{8}} = 2,7^* \text{ que está claramente en la zona de rechazo del nivel } 0,05.$$

Entonces, si este caso cumpliera los requisitos adecuados se podría resolver con los tres modelos: el de Student, que es el más poderoso de los tres, que detecta claramente las diferencias entre ambos espectrofotómetros; el no paramétrico de Wilcoxon que las detecta, pero en el borde de la zona de rechazo; y el más débil del Signo que no llega a detectar las diferencias. Con lo cual se puede plantear una regla general:

Se debe usar el modelo más sensible disponible para cada caso.

Caso 2) Muestras grandes: se desea analizar si el efecto de dos antibióticos de amplio espectro, el A y el B, es equivalente en el tratamiento de infección bacteriana. Para ello se eligen 30 muestras al azar de otros tantos pacientes de bacteriología. Se siembran y desarrollan en cajas de Petri durante 2 días con el mismo caldo de cultivo. Luego se colocan las dos drogas en cada uno, y a la 24 hs se mide el desarrollo de las aureolas, sus diámetros aproximados, como índice del poder de ambas drogas. Interesa destacar las diferencias entre ellas como se muestra en el cuadro siguiente, con $D_i = A_i - B_i$:

Paciente	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Di	-2	0	0	1	0	0	4	4	1	1	5	3	5	3	-1
Rangos	-11,5			4,5			20	20	4,5	4,5	23	16,5	23	16,5	-4,5
Paciente	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Di	1	-1	5	8	2	2	2	-3	-2	1	4	8	2	3	-1
Rangos	4,5	-4,5	23	25,5	11,5	11,5	11,5	-16,5	-11,5	4,5	20	25,5	11,5	16,5	-4,5

El rango menos frecuente es el de signo negativo: $T = 11,5 + 4,5 + 4,5 + 16,5 + 11,5 + 4,5 = 53$
Por su parte, hay 4 casos de diferencias nulas que no se toman en cuenta, entonces resulta el valor $n = 30 - 4 = 26$ y reemplazando en las fórmulas para el cálculo del estadígrafo T queda:

$$\mu_T = (n/4)(n+1) = (26/4)(26+1) = 175,5$$

$$\sigma_T^2 = (n/24)(n+1)(2n+1) = (26/24)(26+1)(52+1) = 1550,25 \text{ o sea, } \sigma_T = 39,37 = SE(T)$$

$$Z_T = \frac{T - \mu_T}{\sigma_T} = (53 - 175,5) / 39,37 = -3,11^{**} \quad (-3,29 < -3,11 < -2,58)$$

Se ha encontrado evidencia muy significativa que prueba la diferencia entre ambas drogas testeadas en el experimento. Por lo que se rechaza la (H_0) de igualdad entre ambas (dos colas).

14.5 Modelos para dos muestras independientes

Existe una variedad de modelos no paramétricos para comparar dos muestras entre sí. El grupo de modelos que usan magnitudes discretas (datos de recuentos) se deja para el capítulo dedicado al Análisis de Frecuencias que se verá más adelante. Entre las demás cabe mencionar:

La *prueba de aleatoriedad* para dos muestras independientes es un modelo que emplea valores numéricos como puntajes y se requiere que la magnitud considerada tenga una medida de intervalo. Sirve para comparar a las dos muestras en medidas de posición cuando ambos tamaños muestrales son pequeños. Cuando los tamaños son grandes, los cálculos se tornan muy engorrosos y es preferible usar la aproximación con el Modelo Student. Sin embargo, el modelo de la U puede considerarse una prueba de aleatoriedad aplicada a los rangos de las observaciones realizadas. Whitney, en 1948, mostró que hay ocasiones en que el modelo de la U es más poderoso que el de Student y es la mejor alternativa.

La *prueba de Moses para reacciones extremas* es un modelo con muchas aplicaciones en las ciencias de la conducta. Está diseñado para estudiar las reacciones de los individuos ante una condición experimental que les provoque diferencias en la dirección a conductas extremas. Por ejemplo, la inestabilidad política o una depresión económica, provoque una radicalización en las opiniones políticas o sociales de algunas personas. Esta prueba fue concebida específicamente para ser usada con datos (por lo menos ordinales) organizados para estudiar tales hipótesis. Sin embargo, también es aplicable cuando el investigador espera calificaciones altas para un grupo y bajas para el otro, dentro del área educativa. Pero el mismo autor señala que la prueba de la U es más eficiente en esos casos y por lo tanto preferible a este modelo.

La *prueba de rachas de Wald-Wolfowitz* es aplicable a casos donde la (H_0) supone que las dos muestras fueron extraídas de la misma población, frente a la alternativa que difieren totalmente entre sí. Si las muestras son lo suficientemente grandes esta prueba detecta diferencias en la tendencia central de las poblaciones, en variabilidad, en oblicuidad, casi en cualquier cosa. Sin embargo, exige el supuesto que la magnitud considerada provenga de una distribución continua y que las medidas efectuadas sean por lo menos ordinales. Ambos autores consideran que el

modelo de la U es más potente que el de las rachas, especialmente en el caso de dos muestras extraídas de poblaciones con la misma ubicación. La razón es que la de rachas está diseñada para detectar diferencias de cualquier tipo; en cambio, el modelo de la U es específico para descubrir diferencias de este tipo y no pierde potencia en buscar cualquier desvío.

Repasando los tres modelos mencionados antes, se ve que todos son superados en potencia por el modelo de la U, el cual se constituye en la mejor alternativa frente al modelo paramétrico equivalente para comparar dos muestras independientes: el de Student. Por esa razón no se desarrollan en la presente obra y se deja a la inquietud del lector que puede encontrarlos en los trabajos de Siegel, Hollander, Mc Collough, Moses y otros mencionados en la bibliografía. Otro motivo para desarrollar el modelo de la U es su uso cada vez más extendido en Medicina, Bioquímica y Farmacia.

14.5.1 Modelo U de Mann-Whitney

Este modelo U sirve para testear si dos muestras independientes han sido tomadas de la misma población. Se tiene, por lo menos, una magnitud ordinal de la misma. Este es el modelo no-paramétrico más poderoso para comparar dos muestras cuando no son apareadas. Es para el caso donde se tiene dudas acerca de la verificación de los supuestos que piden el modelo Student, o cuando las medidas son ordinales. La hipótesis de trabajo (H_0), siempre es que ambas muestras provienen de la misma población.

El procedimiento a seguir para usar este método es como sigue:

Paso 1) Se ordenan todos los datos, de menor a mayor, de ambas muestras y en un solo conjunto, cuidando de identificar a cada uno con su muestra respectiva.

Paso 2) Se determina la muestra de referencia. Conviene que sea la de menor tamaño, caso contrario la muestra de control, placebo o blanco.

Paso 3) Comenzando con el menor valor, se cuenta el número de muestras que preceden al primero de la muestra control (U_1); luego se busca el segundo de la muestra control y se cuenta el número de muestras precedente del otro grupo (U_2); después se ubica al tercer valor y se procede en forma análoga para determinar (U_3), y así sucesivamente hasta recorrer toda la muestra de control o referencia. En caso de empate, se le asigna medio punto a cada uno.

Paso 4) Se obtiene el estadígrafo $U = U_1 + U_2 + U_3 + \dots$

Paso 5) Se procede a comparar este valor contra el valor crítico de tablas.

Hay dos tipos de tablas para el modelo U. La Tabla 11 del Anexo corresponde a muestras pequeñas (ninguna de las muestras es mayor que 9), y arroja el valor de la probabilidad del estadígrafo U calculado en el Paso 4. La Tabla 12 corresponde a un tamaño mediano de las muestras (entre 9 y 20 cada una), y da un valor crítico U_α que se debe comparar con el obtenido experimentalmente. Para $n > 20$ se usa la aproximación con la función de Gauss. En la práctica se pueden presentar tres casos posibles:

(I) *Muestras muy pequeñas* ($n_1 < 9$ y $n_2 < 9$): cuando ninguna de las muestras es mayor que 8, se emplea la Tabla 11 del Anexo. El procedimiento para usar esta tabla es como sigue.

Al mayor tamaño muestral se lo denomina n_2 y sirve para ubicar la correspondiente celda dentro de la Tabla 11 (hay 6 celdas una para cada tamaño entre 3 y 8). Al menor tamaño muestral se lo denomina n_1 y se usa para ubicar la columna correspondiente dentro de la celda. El valor de probabilidad (p) se saca cruzando esta columna con la fila que corresponde al estadígrafo U calculado en el Paso 4. La tabla está construida para ensayos de una sola cola, en cuyo caso se hace $U\alpha = p$. Cuando el ensayo sea de dos colas será $U\alpha = 2 p$. Finalmente este valor $U\alpha$ debe compararse con el valor U calculado en el Paso 4. Si el valor de U es tan grande que no aparece en la tabla, se usa la transformación $U' = n_1 n_2 - U$ como se muestra en el ejemplo siguiente:

Ejemplo 1) A 3 alumnos elegidos al azar se les enseñó una nueva técnica de preparar un examen final, un día antes de la fecha, otros 4 alumnos; sin preparación especial, se eligieron para control. Los resultados del examen en el grupo entrenado fueron: (10, 8 y 6), los del grupo control obtuvieron (9, 7, 5 y 4). Ordenado los datos en forma ascendente es:

Pareja:	4	5	6	7	8	9	10
Grupo:	C	C	E	C	E	C	E

La nota más chica es del grupo control y no da puntos al igual que la siguiente. La tercer nota solo (7) tiene un valor del otro grupo, otorga un punto, la cuarta nota del control (9) tiene dos notas menores del otro grupo y entrega dos puntos; luego $U = 0 + 0 + 1 + 2 = 3$.

De tablas con los tamaños muestrales de $n_1 = 3$ y $n_2 = 4$, se entra por la fila de $U = 3$ y se encuentra una probabilidad $p = 0,2$ para una cola. Como este ensayo es para ver si hay diferencias, es de dos colas, por lo que debe duplicarse este valor $U\alpha = 0,4$. Comparando con el nivel habitual de significación es $U\alpha = 0,4 > \alpha = 0,05$. Luego, no se puede rechazar la hipótesis nula pues entra en la zona de aceptación. La preparación especial no tuvo efecto en el rendimiento de los alumnos en el examen final.

Cabe destacar que si en vez del grupo control se hubiese tomado el otro como referencia, el puntaje hubiese sido $U = 2 + 3 + 4 = 9$: la tabla no llega al valor 9 en la celda de $n_2 = 4$. Por lo tanto, se debe usar el valor $U' = n_1 n_2 - U = 12 - 9 = 3$, que resulta ser igual al valor hallado anteriormente.

(II) *Muestras medianas* (n entre 9 y 20): cuando el tamaño muestral de la muestra más grande de las dos es mayor que 9 ($n_2 \geq 9$), la tabla anterior no puede usarse. Se debe emplear en cambio la Tabla 12 del Anexo. No interesa si el tamaño de la otra muestra es menor que 9. Esta tabla arroja valores críticos de U , llamados U_c , que se deberán comparar con el calculado en el Paso 4. No da probabilidades como la Tabla 11 por lo que se hace el ensayo comparando los valores de U para decidir si se acepta o rechaza la H_0 .

Como el trabajo de calcular U puede ser engorroso cuando la muestra es algo grande, se muestra una alternativa de cálculo más sencilla, como sigue:

Pasos 1 y 2) Son iguales a los detallados más arriba.

Paso 3) Se coloca el rango correspondiente debajo de la sucesión ordenada. Así, corresponde el rango 1 al menor de los valores, el 2 al siguiente y así sucesivamente. En casos de empates se coloca el rango promedio entre los empatados.

Paso 4) Se calcula $R1 = r_i + r_j + \dots$ como la sumatoria de todos los rangos hallados correspondientes a la muestra de tamaño n_1 . Y $R2$ a su correspondiente de la muestra n_2 .

Paso 5) Se calculan dos estadígrafos con:

$$U1 = (n_1 \cdot n_2) + (n_1/2) (n_1 + 1) - R1$$

$$U2 = (n_1 \cdot n_2) + (n_2/2) (n_2 + 1) - R2$$

Y se elige al menor de ellos para adoptarlo como el estadígrafo muestral U . Luego si este valor es menor o igual que el valor crítico de tablas, se puede rechazar la H_0 . Si $U > U_c$ es lo contrario.

Ejemplo) En un ensayo de valoración se escogieron 16 jueces al azar. A 10 de ellos se les pidió que calificaran de 0 a 100 el sabor de un nuevo jarabe que se está por lanzar al mercado. A los 6 restantes se les hizo probar el sabor tradicional para usarlos como control. La prueba se hizo a ciegas; ni los jueces ni el organizador sabían cómo serían repartidos. Se obtuvieron los resultados siguientes ordenados de menor a mayor:

Valores	48	52	53	67	73	79	81	85	86	88	90	92	93	95	97	99
Grupo	E	C	C	E	C	E	E	E	C	E	C	E	E	E	C	E
Rango	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

Por lo tanto se calculan:

$$R1 = 2 + 3 + 5 + 9 + 11 + 15 = 45$$

$$R2 = 1 + 4 + 6 + 7 + 8 + 10 + 12 + 13 + 14 + 16 = 91$$

$$U1 = (10)(6) + (6/2)(6 + 1) - 45 = 36$$

$$U2 = (10)(6) + (10/2)(10 + 1) - 91 = 24$$

Se adopta $U = 24$ - Notar que $U' = (10)(6) - 24 = 36$ (llega al mismo menor valor) -

De tablas surgen para $n_1 = 6$ y $n_2 = 10$ (una sola cola): $U_{0.05} = 14$; $U_{0.01} = 8$ y $U_{0.001} = 3$

Como el valor muestral $U = 24$ es mayor que los valores críticos, no se puede rechazar la H_0 .

(III) *Muestras grandes* ($n > 20$): cuando el tamaño muestral de una de las muestras es grande, o sea $n > 20$, la distribución de U se acerca a una gaussiana $N(0,1)$, con los parámetros siguientes:

$$Z_U = \frac{U - \mu_u}{\sigma_u}$$

y se lo compara con $Z\alpha$ como es habitual, donde:

$$\mu_u = (n_1 \cdot n_2) / 2$$

$$\sigma_u^2 = (n_1 \cdot n_2) (1 / 12) (n_1 + n_2 + 1) = \text{VAR}(u)$$

Para el cálculo del estadígrafo U se procede como se explicó para el caso de muestras de tamaño mediano, en el punto anterior.

Ejemplo) Para comparar dos grupos de alumnos luego de un examen parcial, la cátedra eligió al azar a 22 de la carrera de Farmacia y 16 de Bioquímica, a los cuales se les tomaron los mismos temas, el mismo día y a la misma hora. Los resultados ordenados fueron:

P	30	35	40	40	40	40	48	50	55	58	60	60	60	65	65	65	70	70	75
C	B	B	B	B	F	F	B	B	F	F	B	B	F	B	F	F	F	B	F
R	1	2	4,5	4,5	4,5	4,5	7	8	9	10	12	12	12	15	15	15	17,5	17,5	19

P	78	78	78	79	80	80	80	83	84	85	85	85	90	90	90	92	95	95	95
C	B	B	F	F	B	F	F	B	F	B	F	F	B	F	F	F	F	F	F
R	21	21	21	23	25	25	25	27	28	30	30	30	33	33	33	35	37	37	37

$$R_B = 1 + 2 + 4,5 + 4,5 + 7 + 8 + 12 + 12 + 15 + 17,5 + 21 + 21 + 25 + 27 + 30 + 33 = 240,5$$

$$R_F = 4,5 + 4,5 + 9 + 10 + 12 + 15 + 15 + 17,5 + 19 + 21 + 23 + 25 + 25 + 28 + 30 + 30 + 33 + \dots + 33 + 35 + 37 + 37 + 37 = 500,5$$

$$\text{Luego es: } U_B = (n_B \cdot n_F) + (n_B / 2) (n_B + 1) - R_B = (16 \cdot 22) + (16/2)(16+1) - 240,5 = 247,5$$

$$U_F = (n_B \cdot n_F) + (n_F / 2) (n_F + 1) - R_F = (16 \cdot 22) + (22/2)(22+1) - 500,5 = 104,5$$

Se elige al menor como $U = 104,5$ (Farmacia) y $U' = 22(16) - 104,5 = 247,5$

$$\mu_u = (n_1 \cdot n_2) / 2 = (22 \cdot 16) / 2 = 176$$

$$\sigma_u^2 = (n_1 \cdot n_2) (1 / 12) (n_1 + n_2 + 1) = 1144. \text{ Luego es } \sigma_u = 33,8 \text{ y se calcula Z con:}$$

$Z_U = (U - \mu_u) / \sigma_u = (104,5 - 176) / 33,8 = -2,12^*$: Hay una diferencia significativa en el rendimiento de ambos grupos siendo los de Farmacia mejor que los de Bioquímica.

14.6 Modelo de Cochran para n muestras apareadas

Cuando se tiene el caso de N individuos, a cada uno de los cuales se le efectuó una medición de una magnitud clínica dicotómica (o binaria) mediante n técnicas diferentes ($n > 1$), se tiene el caso de mediciones repetidas al mismo paciente. Esto es un total de n valores apareados para cada paciente. Los resultados de cada test se pueden expresar como sano o enfermo, (+) o (-) etc. Lo mismo ocurre cuando n jueces opinan sobre N casos diferentes. Suponiendo que la magnitud clínica es X, entonces el valor X_{ij} corresponde al paciente i medido con el método j . Los resultados obtenidos se pueden expresar como se hace en la Tabla 14.1 siguiente:

Tabla 14.1. N pacientes medidos con n métodos clínicos

Pacientes	Test 1	Test 2	...	Test j	...	Test n	Total
1	X ₁₁	X ₁₂	...	X _{1j}	...	X _{1n}	T _{1.}
2	X ₂₁	X ₂₂	...	X _{2j}	...	X _{2n}	T _{2.}
3	X ₃₁	X ₃₂	...	X _{3j}	...	X _{3n}	T _{3.}
...
i	X _{i1}	X _{i2}	...	X _{ij}	...	X _{in}	T _{i.}
...
N	X _{N1}	X _{N2}	...	X _{Ni}	...	X _{Nn}	T _{N.}
Total	T _{.1}	T _{.2}	...	T _{.j}	...	T _{.n}	T

Donde como convención se adopta X = 0 cuando es (+) y X = 1 cuando es (-) entonces los valores observados son una serie de ceros y unos. Definiendo las sumatorias:

$$T_{.j} = \sum_{i=1}^N X_{ij} : \text{El total de la columna } j. \quad T = \sum_{i=1}^N \sum_{j=1}^n X_{ij} : \text{El total de todas las observaciones}$$

$$T_{i.} = \sum_{j=1}^n X_{ij} : \text{El total de la fila } i.$$

$$RS = \sum_{i=1}^N (T_{i.})^2 : \text{La suma de los cuadrados de las filas.}$$

$$CS = \sum_{j=1}^n (T_{.j})^2 : \text{La suma de los cuadrados de las columnas.}$$

Entonces se puede efectuar un test estadístico para ver si existe una diferencia significativa entre los tests clínicos llamado: *La prueba Q de Cochran*. El estadístico Q se calcula con:

$$Q = \{(n-1) [(n \cdot CS) - T^2]\} / [n T - RS]$$

El cual tiene aproximadamente una distribución Chi-cuadrado con n – 1 grados de libertad.

Ejemplo) Se han examinado 12 pacientes mediante tres diferentes técnicas para detectar toxoplasmosis. Si los resultados se informan como (+) y (-) detectar si hay diferencias significativas entre los tres métodos. Usando la convención de 1 para los (+) y 0 para los (-) resultó:

Pacientes	Test 1	Test 2	Test 3	Total
1	1	1	0	2
2	0	0	0	0
3	1	1	1	3
4	1	0	1	2
5	1	1	1	3
6	0	0	1	1
7	1	1	1	3
8	0	0	0	0
9	1	1	0	2
10	0	0	1	1
11	0	1	0	1
12	1	0	1	2
Total	7	6	7	20

$$T = \sum_{i=1}^N \sum_{j=1}^n X_{ij} = 20$$

$$CS = \sum_{i=1}^N (T_{i.})^2 = 49 + 36 + 49 = 134$$

$$RS = \sum_{j=1}^n (T_{.j})^2 = 4 + 0 + \dots + 4 = 46$$

$$Q = \{(2) [(402) - 400]\} / [60 - 46]$$

$$Q = 0.29^* < \chi^2_{0,95 ; 2} = 5,991$$

La conclusión es que no se han encontrado diferencias significativas entre los tres métodos.

Para el caso particular de tener que comparar dos tests clínicos es $n = 2$ y el problema se puede analizar como se muestra a continuación:

Tabla 14.2. N pacientes medidos con 2 métodos clínicos

Pacientes	Test 1	Test 2	Total
1	X_{11}	X_{12}	$T_{1.}$
2	X_{21}	X_{22}	$T_{2.}$
3	X_{31}	X_{32}	$T_{3.}$
...
i	X_{i1}	X_{i2}	$T_{i.}$
...
N	X_{N1}	X_{N2}	$T_{N.}$
Total	$T_{.1}$	$T_{.2}$	T

Esto resulta equivalente a presentar los datos en una tabla como la siguiente:

Tabla 14.3. Tabla de Concordancia entre dos tests clínicos

Test 2	Test 1		Total
	(+)	(-)	
(+)	r_{11}	r_{12}	$r_{11} + r_{12}$
(-)	r_{21}	r_{22}	$r_{21} + r_{22}$
Total	$r_{11} + r_{21}$	$r_{12} + r_{22}$	N

Donde el significado de las frecuencias es

- r_{11} : los pacientes que muestran ambos resultados (+) : concordancia positiva.
- r_{22} : los pacientes que muestran ambos resultados (-) : concordancia negativa.
- r_{21} : los pacientes que son (+) con el primer test y (-) con el segundo (discordancia).
- r_{12} : los pacientes que son (-) con el primer test y (+) con el segundo (discordancia).

Como los valores X_{ij} solo pueden ser 0 o 1, el total de cada fila de la Tabla 14.2 puede ser 0 cuando ambos resultados son (-), o 1 cuando un resultado es (+) y el otro (-) y 2 cuando ambos resultados son (+). Por su parte, el total de cada columna será igual al total de casos (+) en cada test. Por lo tanto el total de cada columna será:

$$T_{.1} = \sum_{i=1}^N X_{ij} = r_{11} + r_{21} \quad \text{y} \quad T_{.2} = \sum_{i=1}^N X_{ij} = r_{12} + r_{22} \quad \text{Luego el total general es:}$$

$$T = \sum_{i=1}^N \sum_{j=1}^n X_{ij} = T_{.1} + T_{.2} = 2r_{11} + r_{21} + r_{12} \quad \text{Pero el total ambas filas también es T:}$$

$$T = \sum_{i=1}^N \sum_{j=1}^n X_{ij} = \sum_{i=1}^N T_i = 0 (r_{22}) + 1 (r_{12}) + 1 (r_{21}) + 2 (r_{11}) \quad \text{Por lo tanto}$$

$$RS = 0^2 (r_{22}) + 1^2 (r_{12}) + 1^2 (r_{21}) + 2^2 (r_{11}) = (r_{12} + r_{21}) + 4 r_{11} \quad \text{Y así se puede calcular}$$

$$CS = (r_{11} + r_{21})^2 + (r_{11} + r_{12})^2 \quad \text{Con lo cual la fórmula del estadígrafo Q se reduce a:}$$

$$Q = [2 \cdot CS - T^2] / [2 T - RS] = (r_{12} - r_{21})^2 / (r_{12} + r_{21})$$

Notar que si se define como 1 a los casos de (-) y como 0 a los (+) el resultado es el mismo. Lo más importante para destacar es que si r_{11} y/o r_{22} cambia sus valores Q no varía.

Ejemplo 1) Se han efectuado dos mediciones de VDRL a cada uno de 16 pacientes con dos marcas diferentes de kits. Se desea saber si las marcas son equivalentes. Los resultados fueron

Pacientes	Test 1	Test 2	Total
1	1	1	2
2	0	0	0
3	1	1	2
4	1	0	1
5	1	1	2
6	0	0	0
7	1	1	2
8	0	0	0
9	1	1	2
10	0	0	0
11	0	1	1
12	1	0	1
13	0	0	0
14	0	1	1
15	0	0	0
16	1	0	1
Total	8	7	15

$$T = \sum_{i=1}^N \sum_{j=1}^n X_{ij} = 15$$

$$CS = \sum_{i=1}^N (T_i)^2 = 64 + 49 = 113$$

$$RS = \sum_{j=1}^n (T_j)^2 = 4 + 0 + \dots + 1 = 25$$

$$Q = \{(1) [(2 \cdot 113) - 225]\} / [2 \cdot 15 - 25] = 0,2$$

$$Q = 0,2 < \chi^2_{0,95; 1} = 3,871$$

Los resultados no fueron significativos. Esto indica que no hay diferencias estadísticamente significativas entre ambos tests.

También se puede resolver este problema con la Tabla de Concordancia:

Test 2	Test 1		Total
	(+)	(-)	
(+)	5	3	8
(-)	2	6	8
Total	7	9	16

$$RS = (r_{12} + r_{21}) + 4 r_{11} = (3 + 2) + 4 \cdot 5 = 25$$

$$CS = (r_{11} + r_{21})^2 + (r_{11} + r_{12})^2 = (5 + 2)^2 + (5 + 3)^2 = 49 + 64 = 113$$

$$Q = [2 \cdot CS - T^2] / [2 T - RS] = [226 - 225] / [30 - 25] = 0,2 = (3-2)^2 / (3+2) < \chi^2_{0,95; 1} = 3,871$$

14.7 Modelo de Cohen-Kappa

Este modelo se usa cuando se tienen dos métodos de diagnóstico diferentes A y B cuyos resultados se expresan en forma de dos o más categorías (dicotómica o polótómica). Cada muestra individual es juzgada por cada uno de los métodos en alguna de las clases posibles. El estadígrafo Kappa (K) da una medida del acuerdo que existe entre ambos métodos clínicos. Para ilustrar este procedimiento se presenta el siguiente ejemplo:

- Se desea ver si se puede reemplazar el antibiótico A que se está usando en el laboratorio, por uno nuevo B. El resultado de cada medición de acuerdo al diámetro de la aureola, se agrupa en tres categorías: R (resistente), S (sensible) y M (muy sensible). Se efectúan 100 pruebas con los resultados siguientes:

Frecuencias Observadas

		B			
		M	S	R	
A	M	44	5	1	50
	S	7	20	3	30
	R	9	5	6	20
		60	30	10	100

Paso 1) Se calcula el valor q como la sumatoria de los productos de los totales marginales de filas y columnas de cada clase, dividido el tamaño muestral:

$$q = q_M + q_S + q_R = [(50 \cdot 60) + (30 \cdot 30) + (20 \cdot 10)] / 100 = 41$$

Paso 2) Se calcula la sumatoria de los valores encontrados en la diagonal principal, o sea el número de las coincidencias encontradas:

$$c = 44 + 20 + 6 = 70$$

Paso 3) Se obtiene el estadígrafo Kappa (K) con:

$$K = (c - q) / (N - q) = (70 - 41) / (100 - 41) = 0,492$$

Esto significa que de todas las muestras que se esperaba que no sean concordantes, solo el 49,2% de ellas son de hecho concordantes. En general, cuanto más cerca este Kappa de 1, mejor será la concordancia encontrada. Pero para el ejemplo presente el máximo valor posible de Kappa será $K = 0,835$. Lo que se deduce del cuadro siguiente:

		B			
		M	S	R	
A	M	50	0	0	50
	S	0	30	0	30
	R	10	0	10	20
		60	30	10	100

Respetando los totales marginales, y haciendo máximas las coincidencias en la diagonal principal, se puede obtener el caso ideal de concordancia máxima posible dados los totales marginales encontrados. Para esa tabla ideal el valor $K = 0,835$. Luego el porcentaje entre el valor de K observado y su máximo posible es del 59,2%

Paso 4) Se calcula el error estándar de Kappa con la formula aproximada siguiente:

$$SE (K) = \sqrt{\frac{q}{N.(N-q)}} = \sqrt{\frac{41}{100.(59)}} = 0,08$$

Paso 5) El intervalo de confianza al 95% se calcula mediante la aproximación normal

$$\text{Limite inferior} = K - 1,96 SE (K) = 0,492 - 1,96 (0,08) = 0,335$$

$$\text{Límite superior} = K + 1,96 SE (K) = 0,492 + 1,96 (0,08) = 0,649$$

Luego es: K 95% CI (0,335 ; 0,649)

Paso 6) El test estadístico se puede hacer usando la aproximación normal con:

$$z = \frac{c - q}{\sqrt{\frac{q.(N-q)}{N}}} = \frac{70 - 41}{\sqrt{\frac{41.(59)}{100}}} = 5,9^{***} \quad \text{Resulta ser altamente significativo.}$$

Este resultado indica que se puede rechazar la H_0 y aceptar que existe una asociación estadística muy fuerte entre ambos antibióticos. Este tipo de asociación matemática es aceptado en Medicina como si fuese una asociación clínica. La concordancia entre ambas pruebas se toma como aceptable, clínicamente hablando. Sin embargo, no se debe olvidar que no es una asociación del tipo causa-efecto, sino que se trata del concepto de independencia estadística visto en el Tema 6.

Notar que en el caso poco frecuente de que los métodos A y B terminan con el mismo número de muestras en la categoría M, el mismo número en la S y así hasta la última categoría, en este caso particular R, entonces el valor máximo de Kappa sería uno. Pero cuando se termina eligiendo números diferentes en cada categoría, como en el ejemplo anterior, el valor máximo de Kappa será siempre menor a uno ($K = 0,835$ en el ejemplo).

También se puede analizar el grado de acuerdo dentro de cada categoría por separado. En el ejemplo visto para la categoría M hay 44 acuerdos en total entre A y B, pero el antibiótico A define 6 casos más como si fuesen M, mientras que el antibiótico B define 16 casos más como M. Entonces se tienen en total 66 (44 + 6 + 16) muestras definidas como M por ambos antibióticos, de los cuales en solo 44 hay concordancia. Por lo tanto, el porcentaje de acuerdo para la categoría M será de $44/66 = 0,6667$, o sea un 66,7% de concordancia en dicha categoría. Análogamente, la concordancia para S es del 50% y en R es del 25%. Por otro lado, observando la tabla con los máximos acuerdos posibles se ve que para la categoría M hay en total 50 acuerdos en ambos antibióticos, con un total de $50 + 0 + 10 = 60$ entre A y B. Por eso, la máxima proporción de acuerdo posible en esa categoría es $50/60 = 0,833$ (83,3%) para comparar con el 66,7% encontrado, y así sucesivamente.

Este método fue muy popular durante años, pero desde entonces arrecian cada vez más las críticas contra el mismo. Finalmente las conclusiones actuales permiten decir que:

- 1) Kappa no debería ser aceptada como el estándar de concordancia o el método de referencia para obtenerla, cuando se comparan dos métodos clínicos en forma apareada. Es decir, cuando se mide dos veces a la misma muestra.
- 2) No debería usarse un estadígrafo como K que genera mucha controversia.
- 3) Se debería considerar a K como una de las alternativas disponibles para la decisión final.

Se pueden identificar dos usos posibles de Kappa: a) como una manera de testear la independencia estadística de los métodos y b) como una manera de cuantificar el grado de concordancia entre ambos métodos. Para el primer caso (a) se trata de testear una H_0 sobre que no hay más concordancia de la que podría haber dadas las circunstancias, esto es dado los totales marginales hallados. Pero de esta forma la decisión a tomar sobre si existe una asociación clínica, es puramente *cualitativa*, en vez de cuantitativa como debería ser. Se puede concluir que para este exclusivo propósito el uso de Kappa en clínica es adecuado. Pero para el segundo caso (b) no se puede decir lo mismo, porque no puede cuantificar adecuadamente el grado de concordancia, lo que significa que no se puede tomar como un índice clínico del tipo Sensibilidad, etc. Kappa debe ser interpretada como una proporción de las veces que los métodos podrían ser concordantes debido únicamente al azar, pero como los métodos no se aplican a muestras independientes, sino a muestras apareadas, los métodos son intrínsecamente dependientes y por lo tanto su aplicación o la relevancia de este término, en la concordancia es muy discutible.

El estadígrafo Kappa es apropiado para testear cuando la concordancia excede el grado de acuerdo debido al azar. Y puede ser calculado fácilmente. Estas son las dos ventajas del mismo. La lista de desventajas es muy larga y puede ser encontrada en varios libros mencionados en la Bibliografía.

14.8 Problemas propuestos

1) Marcar la respuesta correcta a cada una de las afirmaciones siguientes, o completar la frase:

- | | | |
|--|-------|-------|
| 1) Los modelos no paramétricos (MNP) sirven para cualquier magnitud biológica. | V | F |
| 2) En los MNP no se necesita hacer supuestos acerca de la forma de la distribución original. | V | F |
| 3) Mencionar las ventajas y desventajas de trabajar con estadística no paramétrica: | | |
| 4) Siempre que se pueda conviene usar modelos no paramétricos. | V | F |
| 5) Los pasos a seguir para aplicar el modelo de la Binomial son: | | |
| 6) En el modelo Binomial se pueden dar 3 casos posibles de acuerdo con el tamaño muestral. | V | F |
| 7) La prueba de rachas de una muestra sirve para: | | |
| 8) Los pasos a seguir en una prueba de rachas son: | | |
| 9) En el modelo Rachas se pueden dar 3 casos posibles de acuerdo con el tamaño muestral. | V | F |
| 10) Si las muestras son grandes, el estadígrafo de Rachas tiene una distribución $N(0,1)$. | V | F |
| 11) Es lo mismo trabajar con dos muestras apareadas o dos muestras independientes. | V | F |
| 12) Los supuestos básicos para poder usar el Modelo del Signo son: | | |
| 13) Toda magnitud biológica puede ser dicotomizada. | V | F |
| 14) Para el modelo del signo una muestra es grande si supera los 30 valores. | V | F |
| 15) Los casos antes-después y droga-placebo se pueden estudiar con el modelo del Signo. | V | F |
| 16) Si se aplican dos tratamientos al mismo individuo en n casos, las muestras son: | | |
| 17) La prueba de Wilcoxon es más poderosa que su similar del Signo. | V | F |

- 18) La prueba de Wilcoxon usa un peso relativo para cada signo empleado. **V F**
 19) El supuesto para Wilcoxon es que las muestras provienen de una distribución simétrica. **V F**
 20) Los pasos a seguir en el modelo de Wilcoxon son:
 21) El tamaño muestral para Wilcoxon es grande si: $n > 20$ **V F**
 22) Se debe usar el modelo más sensible (robusto) en cada caso investigado. **V F**
 23) Para dos muestras independientes los modelos que se pueden usar son:
 24) De todos los modelos para comparar dos muestras independientes, la U es el mejor. **V F**
 25) El modelo de la U es el mejor no paramétrico para poder comparar a muestras
 26) Los pasos a seguir en el modelo de la U son:
 27) En el modelo de la U se pueden presentar 3 casos de acuerdo con el tamaño muestral. **V F**
 28) Para la U las muestras son muy pequeñas si ninguna es mayor de 9. **V F**
 29) Si las muestras son medianas o grandes, los pasos a seguir para la U se modifican en.....
 30) El modelo de Cochran se puede emplear para comparar 3 o más métodos clínicos. **V F**
 31) El tamaño de muestra total N tiene gran influencia en el cálculo de Q. **V F**
 32) El número de casos concordantes influencia el valor de Q. **V F**
 33) Kappa es la manera de cuantificar el concepto de concordancia entre 2 métodos clínicos **V F**
 34) Enumerar las ventajas del estadígrafo Kappa:.....
 35) Realizar un esquema sinóptico sobre cual modelo no paramétrico elegir, en los casos de la realidad.

2) Resolver (cuando se pueda) los problemas desarrollados en los capítulos anteriores y los problemas propuestos, con los modelos no paramétricos vistos en este capítulo.

3) Comparar el poder entre los modelos de punto anterior.

4) Se eligieron 17 días al azar de un semestre y se determinaron las ventas en las sucursales A y B de una misma farmacia. Determinar si hay diferencias significativas en las recaudaciones de cada una.

A	300	350	240	450	140	340	248	250	155	258	360	260	610	165	170	90	120
B	240	320	270	370	200	190	210	180	140	200	290	340	400	200	120	80	150

5) Resolver el mismo problema anterior si el procedimiento es diferente: se eligen 34 días al azar y se asignan a cada farmacia con un sorteo posterior. Explicar cómo haría esta asignación usando una tabla de números al azar.

6) Resolver si hay asociación entre ambos tests clínicos 1 y 2, en los dos casos siguientes:

	Test 1	
Test 2	(+)	(-)
(+)	5	3
(-)	2	5

	Test 1		
Test 2	A	B	C
A	55	2	1
B	2	25	2
C	1	5	7

7) Decidir si hay diferencias significativas entre los diámetros de aureola de 2 antibióticos A y B en cajas de Petri, con los datos siguientes expresados de mm:

Diá.	128	140	178	179	180	185	189	193	194	200	205	210	215	220	220	224	230
Antib.	A	B	B	A	B	A	A	A	B	B	A	B	B	A	B	A	B