

Test para dos muestras independientes Test de Mann-Whitney-Wilcoxon

Consideremos una situación en la cual un investigador ha obtenido dos muestras aleatorias independientes de poblaciones posiblemente diferentes y desea testear la hipótesis nula de que ambas poblaciones son idénticas. Es decir, el investigador desea detectar diferencias entre las dos poblaciones en base a muestras aleatorias de ambas.

Intuitivamente se podría pensar en combinar ambas muestras, asignar rangos a las observaciones y definir el estadístico del test como la suma de los rangos de las observaciones en una de las dos poblaciones. Si la suma es demasiado pequeña o demasiado grande, esto sería evidencia de que los valores de una de las poblaciones tienden a ser más pequeños o más grandes que los de la otra y se rechazaría la hipótesis nula.

Tal como lo presentamos se trataría de un test para diferencia entre distribuciones, es decir que la hipótesis nula sería $H_0: F(x) = G(x) \quad \forall x$, siendo F la distribución de la m.a. X_1, \dots, X_n y G la de la m.a. Y_1, \dots, Y_m . Sin embargo, nosotros lo trataremos como un test para el parámetro de posición, para lo cual supondremos que $G(x) = F(x - \Delta)$, para algún Δ .

Datos: Sea $F \in \Omega_0$ y sean X_1, \dots, X_n una m.a. de una distribución $F(x)$ e Y_1, \dots, Y_m una m.a. de una distribución $F(x - \Delta)$, independiente de la primera, es decir que sólo suponemos que F es absolutamente continua con única mediana y que las dos distribuciones tienen la misma forma (en particular, igual varianza).

Hipótesis a testear:

A. $H_0: \Delta = 0$ vs $H_1: \Delta \neq 0$

B. $H_0: \Delta = 0$ vs $H_1: \Delta < 0$

C. $H_0: \Delta = 0$ vs $H_1: \Delta > 0$

Estadístico del test y zona de rechazo: Consideremos el caso **C**. Wilcoxon (1945) propuso el siguiente procedimiento. Se ordenan los datos de las dos muestras combinadas de menor a mayor y se define el estadístico U como la suma de los rangos de la m.a. Y_1, \dots, Y_m . Se rechaza H_0 si U es grande.

Si hubiésemos considerado las hipótesis **B** rechazaríamos H_0 para valores pequeños de U , o equivalentemente para valores grandes del estadístico T definido como la suma de los rangos de la m.a. X_1, \dots, X_n en la muestra combinada.

Sea $N = n + m$, y llamemos $R(Y_i)$ y $R(X_i)$ a los rangos de Y_i y X_i respectivamente, en la muestra ordenada de los datos combinados. Si hay empates se asigna a los valores empatados el promedio de los rangos que les corresponderían si no hubiese empates. Se definen los estadísticos

$$U = \sum_{i=1}^m R(Y_i) \qquad T = \sum_{i=1}^n R(X_i)$$

que verifican $U + T = \frac{N(N+1)}{2}$.

Mann y Whitney (1947) consideraron, por analogía con el test del signo, el siguiente estadístico:

$$W = \sum_{i,j} s(Y_i - X_j)$$

donde ahora las variables que intervienen en la suma no son independientes.

Si no hay empates,

$$R(Y_i) = \#\{X_j < Y_i \mid 1 \leq j \leq n\} + \#\{Y_k \leq Y_i \mid 1 \leq k \leq m\}$$

entonces,

$$U = W + \frac{m(m+1)}{2}$$

Teorema: Bajo $H_0: \Delta = 0, F \in \Omega_0$,

a)
$$P(R(Y_i) = k) = \frac{1}{N} \quad 1 \leq k \leq N$$

$$P(R(Y_i) = k, R(Y_j) = l) = \begin{cases} \frac{1}{N(N-1)} & \text{si } k \neq l, i \neq j \\ 0 & \text{en caso contrario} \end{cases}$$

b)
$$E(R(Y_i)) = \frac{N+1}{2} \quad V(R(Y_i)) = \frac{N^2-1}{12} \quad \text{cov}(R(Y_i), R(Y_j)) = -\frac{(N+1)}{12} \text{ si } i \neq j$$

Corolario:
$$E(U) = m \frac{N+1}{2} \quad E(T) = n \frac{N+1}{2} \quad E(W) = \frac{nm}{2}$$

$$V(U) = V(T) = V(W) = nm \frac{N+1}{12} \qquad (2)$$

Si hay empates, la esperanza de U no cambia pero sí la varianza. Si s es el número de grupos de empates y d_j es el número de valores empatados en el grupo j, entonces:

$$V(U / d_1, \dots, d_s) = mn \frac{N+1}{12} - \frac{mn \sum_{i=1}^s (d_i^3 - d_i)}{12N(N-1)}$$

donde el segundo sumando se denomina "corrección por empates". La varianza de U se puede estimar en este caso mediante

$$\hat{V}(U) = \frac{mn}{N(N-1)} \sum_{i=1}^N R_i^2 - \frac{mn(N+1)^2}{4(N-1)} \quad (3)$$

siendo

$$R_i = \begin{cases} R(X_i) & 1 \leq i \leq n \\ R(Y_{i-n}) & n+1 \leq i \leq n+m \end{cases}$$

Volviendo a los tests planteados inicialmente, las zonas de rechazo serán las siguientes:

$$\mathbf{A.} \quad H_0: \Delta = 0 \quad \text{vs} \quad H_1: \Delta \neq 0$$

Se rechaza H_0 si $U \leq w_{\alpha/2}^{mn}$ ó $U > w_{1-\alpha/2}^{mn}$, siendo w_{β}^{mn} el cuantil β de la distribución exacta del estadístico de Wilcoxon para tamaños de muestra m y n . La Tabla A7 de Conover da los cuantiles inferiores de la distribución exacta para $2 \leq m, n \leq 20$.

Como $U + T = \frac{N(N+1)}{2}$, los cuantiles superiores pueden obtenerse mediante la relación:

$w_{1-\beta}^{mn} = \frac{N(N+1)}{2} - w_{\beta}^{mn}$. Si n y m son suficientemente grandes o hay empates, se utiliza la aproximación Normal. En este caso, el cuantil aproximado es

$$w_p \cong \frac{m(N+1)}{2} + z_p \sqrt{\hat{V}(U)}$$

La expresión del estimador de la varianza de U se reemplaza por (2) si no hay empates o por (3) si hay empates.

B. $H_0: \Delta = 0$ vs $H_1: \Delta < 0$

Se rechaza H_0 si $U \leq w_{\alpha}^{mm}$ o, si las muestras son grandes o hay empates si $U \leq \frac{m(N+1)}{2} - z_{\alpha} \sqrt{V(U)}$.

C. $H_0: \Delta = 0$ vs $H_1: \Delta > 0$

Se rechaza H_0 si $U > w_{1-\alpha}^{mm}$ o, si las muestras son grandes o hay empates si $U > \frac{m(N+1)}{2} + z_{\alpha} \sqrt{V(U)}$.

Distribución exacta bajo H_0 : Consideremos, a modo de ejemplo, el caso $n=m=2$ sin empates, con lo cual los rangos toman valores 1, 2, 3 y 4. En la siguiente tabla se presentan todas las configuraciones posibles y los correspondientes valores de los estadísticos:

Rangos				W	U	T
1	2	3	4			
y	y	x	x	0	3	7
y	x	y	x	1	4	6
y	x	x	y	2	5	5
x	y	y	x	2	5	5
x	y	x	y	3	6	4
x	x	y	y	4	7	3

La distribución exacta del estadístico U, por ejemplo, será:

U	3	4	5	6	7
p_U	1/6	1/6	2/6	1/6	1/6

Hay una fórmula recursiva que permite obtener la distribución exacta. (Hettmansperger, pag. 137)

Distribución asintótica bajo $H_0: \Delta = 0$: Supongamos que n y m tienden a infinito, de modo tal que

$$\frac{m}{n+m} \rightarrow \lambda \quad 0 < \lambda < 1$$

entonces, bajo H_0 ,

$$\frac{W - E(W)}{\sqrt{V(W)}} \xrightarrow{d} N(0,1) \qquad \bar{W} = \frac{W}{nm} \xrightarrow{p} 1/2$$

$$\frac{U - E(U)}{\sqrt{V(U)}} \xrightarrow{d} N(0,1) \qquad \frac{T - E(T)}{\sqrt{V(T)}} \xrightarrow{d} N(0,1)$$

Estimador puntual e intervalo de confianza: El test para $H_0: \Delta = \Delta_0$ vs $H_1: \Delta \neq \Delta_0$ se puede obtener trabajando con $\tilde{Y}_i = Y_i - \Delta_0$ que tiene mediana nula bajo H_0 , y el estadístico del test se basará en

$$W(\Delta_0) = \sum_{i,j} s(Y_j - X_i - \Delta_0) = \#\{(i, j) / 1 \leq i \leq n, 1 \leq j \leq m, Y_j - X_i > \Delta_0\}$$

De esta forma del estadístico del test se deduce el estimador de Hodges-Lehmann de Δ ,

$$\hat{\Delta} = \underset{i,j}{med}(Y_j - X_i)$$

Como bajo H_0 , la distribución de W es simétrica alrededor de su media $nm/2$, si k satisface $P_{H_0}(W \leq k) = \alpha/2$, entonces

$$[D^{(k+1)}, D^{(nm-k)}]$$

será un intervalo de confianza de nivel $1 - \alpha$ para Δ , siendo $D^{(1)} \leq \dots \leq D^{(mn)}$ los estadísticos de orden de las diferencias $D_{ij} = Y_j - X_i$.

Como $U = W + \frac{m(m+1)}{2}$, el valor de k se puede obtener, a partir de la tabla del libro de Conover, como

$$k = w_{\alpha/2}^{mn} - \frac{m(m+1)}{2}$$

o bien, en forma aproximada,

$$k \cong \frac{mn}{2} - \frac{1}{2} - z_{\alpha/2} \sqrt{V(\hat{U})}$$

con $V(\hat{U})$ reemplazada por (3) o (2) según haya o no empates.

Ejemplo: En una clase de una escuela de enseñanza superior hay 48 alumnos varones, de los cuáles 12 viven en el campo y 36 en la ciudad. Se desarrolló un test para determinar la condición física de los alumnos. Se aplicó este test a los 48 alumnos y se asignó a cada uno una puntuación. Una puntuación baja indica mala condición física. Los resultados obtenidos son los siguientes:

Campo (X _i)		Ciudad (Y _i)					
14.8	10.6	12.7	16.9	7.6	2.4	6.2	9.9
7.3	12.5	14.2	7.9	11.3	6.4	6.1	10.6
5.6	12.9	12.6	16.0	8.3	9.1	15.3	14.8
6.3	16.1	2.1	10.6	6.7	6.7	10.6	5.0
9.0	11.4	17.7	5.6	3.6	18.6	1.8	2.6
4.2	2.7	11.8	5.6	1.0	3.2	5.9	4.0

Se desea testear

H₀: los alumnos que viven en el campo y los que viven en la ciudad tienen la misma condición física

H₁: los alumnos que viven en el campo tiene mejor condición física que los que viven en la ciudad.

O, equivalentemente, si se supone que ambas poblaciones tienen igual distribución, excepto quizás en la posición, que

$$H_0: \Delta = 0 \quad \text{vs} \quad H_1: \Delta < 0$$

donde $\Delta = \text{med}(Y_i) - \text{med}(X_i)$.

Se ordenan las puntuaciones de ambos grupos y se les asigna el correspondiente rango:

X	Y	Rango	X	Y	Rango	X	Y	Rango
	1.0	1		6.2	17		11.3	33
	1.8	2	6.3		18	11.4		34
	2.1	3		6.4	19		11.8	35
	2.4	4		6.7	20.5	12.5		36
	2.6	5		6.7	20.5		12.6	37
2.7		6	7.3		22		12.7	38
	3.2	7		7.6	23	12.9		39
	3.6	8		7.9	24		14.2	40
	4.0	9		8.3	25		14.8	41.5
4.2		10	9.0		26	14.8		41.5
	5.0	11		9.1	27		15.3	43
	5.6	13		9.9	28		16.0	44
	5.6	13		10.6	30.5	16.1		45
5.6		13		10.6	30.5		16.9	46
	5.9	15	10.6		30.5		17.7	47
	6.1	16		10.6	30.5		18.6	48

Calculamos el valor del estadístico $T = \sum_{i=1}^n R(X_i) = 321$. Para utilizar la aproximación normal, necesitamos calcular $\sum_{i=1}^N R_i^2 = 38016$ y por lo tanto el valor del estadístico standarizado será:

$$\frac{T - n \frac{N+1}{2}}{\sqrt{\frac{mn}{N(N-1)} \sum_{i=1}^N R_i^2 - \frac{mn(N+1)^2}{4(N-1)}}} = \frac{321 - 12 \frac{49}{2}}{\sqrt{\frac{12 \cdot 36}{48 \cdot 47} 38016 - \frac{12 \cdot 36 \cdot 49^2}{4 \cdot 47}}} = 0.6431$$

Rechazaríamos H_0 a nivel 0.05 si el valor del estadístico fuese mayor que 1.65, entonces, no rechazamos H_0 . El p-valor es 0.26.

Lo procesamos en R usando la función wilcox.test:

```
campo<-c(14.8,10.6,7.3,12.5,5.6,12.9,6.3,16.1,9.0,11.4,4.2,2.7)
ciudad<-c(12.7,16.9,7.6,2.4,6.2,9.9,14.2,7.9,11.3,6.4,6.1,10.6,12.6,16.0,
8.3,9.1,15.3,14.8,2.1,10.6,6.7,6.7,10.6,5.0,17.7,5.6,3.6,18.6,1.8,2.6,11.8,
5.6,1.0,3.2,5.9,4.0)
wilcox.test(campo,ciudad,alternative="greater",paired=FALSE)
```

y obtenemos la siguiente salida:

```
Wilcoxon rank sum test with continuity correction

data: campo and ciudad
W = 243, p-value = 0.2639
alternative hypothesis: true location shift is greater than 0

Mensajes de aviso perdidos
In wilcox.test.default(campo, ciudad, alternative = "greater", paired
= FALSE) :
cannot compute exact p-value with ties
```

Distribución del estadístico bajo la alternativa: Sean X_1, \dots, X_n e Y_1, \dots, Y_m m.a. de dos distribuciones arbitrarias continuas $G(x)$ y $H(y)$ respectivamente. Recordemos que el estadístico del test de Mann Whitney es equivalente a

$$W = \sum_{i,j} s(Y_i - X_j)$$

que podemos escribir en la forma

$$W = \sum_i \sum_j T_{ij} \quad \text{donde} \quad T_{ij} = 1 \quad \text{si} \quad Y_i - X_j > 0$$

y definamos, análogamente a lo que hicimos en el caso del test de Wilcoxon,

$$p_1 = P(Y > X) = \int_{-\infty}^{\infty} G(y) h(y) dy = \int_{-\infty}^{\infty} [1 - H(x)] g(x) dx$$

$$p_2 = P(Y_1 > X_1, Y_2 > X_1) = \int_{-\infty}^{\infty} [1 - H(x)]^2 g(x) dx$$

$$p_3 = P(Y_1 > X_1, Y_1 > X_2) = \int_{-\infty}^{\infty} G^2(y) h(y) dy$$

Teorema:

$$E(W) = m n p_1$$

$$V(W) = m n (p_1 - p_1^2) + m n (m - 1) (p_2 - p_1^2) + m n (n - 1) (p_3 - p_1^2)$$

Demostración: $E(W) = m n E(T_{11}) = m n P(Y > X) = m n p_1$. Por otro lado, expresando

$$E(W^2) = E \left[\sum_i \sum_j T_{ij} \right]^2 = E \left[\sum_i \sum_j \sum_k \sum_l T_{ij} T_{kl} \right]$$

se obtiene

$$E(W^2) = mnE(T_{11}^2) + mn(m-1)E(T_{11}T_{12}) + mn(n-1)E(T_{12}T_{22}) + mn(m-1)(n-1)E(T_{11}T_{22})$$

Pero, por ejemplo, $E(T_{11}T_{12}) = P(Y_1 > X_1, Y_2 > X_1) = p_2$ y, operando con los otros términos en la misma forma, y calculando $V(W) = E(W^2) - [E(W)]^2$ se obtiene la expresión dada.

Cuando $m, n \rightarrow \infty$, de manera que $n/N \rightarrow \lambda, 0 < \lambda < 1$, usando el método de las proyecciones puede probarse que, si $0 < p_1 < 1$,

$$\frac{W - E(W)}{\sqrt{V(W)}} \xrightarrow{d} N(0,1)$$

Más precisamente, si $\bar{W} = W / mn$,

$$\sqrt{N}(\bar{W} - p_1) \xrightarrow{d} N(0, (p_2 - p_1^2) / \lambda + (p_3 - p_1^2) / (1 - \lambda))$$

Consistencia: Sea $\mu(G, H) = E(\bar{W}) = p_1$ y sea Ω_M la clase de funciones de distribución continuas definida como

$$\Omega_M = \{(G, H) \text{ tales que } G(x) \geq H(x) \forall x \text{ con desigualdad estricta en al menos un } x\}$$

Entonces,

$$\mu(G, H) = \begin{cases} 1/2 & \text{si } G(x) = H(x) \\ > 1/2 & \text{si } (G, H) \in \Omega_M \end{cases}$$

y, por lo tanto \bar{W} provee un test consistente para la clase de alternativas Ω_M , en particular para alternativas de posición, como las planteadas por nosotros.

Eficacia: Puede hacerse una deducción rigurosa de la eficacia, verificando las condiciones de Pitman, pero daremos sólo su expresión. Consideremos $G(x) = F(x)$ y $H(x) = F(x - \Delta)$. Sabemos que la varianza asintótica es

$$\sigma^2 = (p_2 - p_1^2) / \lambda + (p_3 - p_1^2) / (1 - \lambda)$$

Bajo H_0 , $\sigma^2(0) = \frac{1}{12\lambda(1-\lambda)}$. Además,

$$\mu(\Delta) = p_1 = \int_{-\infty}^{\infty} [1 - F(x - \Delta)] f(x) dx \quad \Rightarrow \quad \mu'(0) = \int_{-\infty}^{\infty} f^2(x) dx$$

y por lo tanto la eficacia del test es

$$c = \frac{\mu'(0)}{\sigma(0)} = \sqrt{12\lambda(1-\lambda)} \int f^2(x) dx$$

Observemos que la eficacia es $\sqrt{\lambda(1-\lambda)}$ veces la eficacia del test de Wilcoxon para muestras apareadas y es máxima cuando $\lambda = 1/2$ o sea cuando las dos muestras son de igual tamaño.

De aquí podemos deducir cuál es la distribución asintótica del estimador de Hodges-Lehmann de Δ , ya que

$$\sqrt{N}(\hat{\Delta} - \Delta) \xrightarrow{d} N(0, 1/c^2)$$

Respecto a la **eficiencia asintótica relativa** del test de Mann-Whitney respecto al test de t, fue estudiada bajo la hipótesis de que las distribuciones de ambas poblaciones difieren sólo en su media. Si ambas poblaciones son normales la eficiencia es 0.955, si son uniformes es 1.0 y si tienen distribución doble exponencial es 1.5. Bajo la hipótesis mencionada sobre las distribuciones, la eficiencia nunca es inferior a 0.864.

Witting (1960) estudió la eficiencia del test de Mann-Whitney relativa al test de t para dos muestras independientes pequeñas. Usando aproximaciones numéricas, Witting mostró que aún para tamaños de muestra pequeños ($m=n=10$, por ejemplo), la eficiencia, bajo normalidad nunca era menor que 0.94.

Problema de Fisher-Behrens: Hasta ahora supusimos que ambas distribuciones tienen la misma forma y difieren a lo sumo en la posición. Ahora supongamos que X_1, \dots, X_n es una m.a. de una distribución $G(x)$, $G \in \Omega_0$ e Y_1, \dots, Y_m es una m.a. de una distribución $H(y-\Delta)$, $H \in \Omega_0$ y que deseamos testear

$$H_0: \Delta = 0 \quad \text{vs} \quad H_1: \Delta > 0$$

Cuando G y H son normales con diferentes varianzas usamos el test de Welch.

La idea es usar la distribución límite de W para G y H arbitrarias e introducir un estimador consistente de la varianza de W. Sabemos que, $E(W) = nmp_1$ y

$$V(W) = mn(p_1 - p_1^2) + mn(m-1)(p_2 - p_1^2) + mn(n-1)(p_3 - p_1^2)$$

Buscamos un estimador consistente de la varianza y lo haremos a partir de las siguientes expresiones

$$p_1 - p_1^2 = \left(\int G(t)h(t)dt \right) \left(\int H(t)g(t)dt \right)$$

$$p_2 - p_1^2 = \int H^2(t)g(t)dt - \left(\int G(t)h(t)d(t) \right)^2$$

$$p_3 - p_1^2 = \int G^2(t)h(t)dt - \left(\int G(t)h(t)d(t) \right)^2$$

La idea es reemplazar en estas ecuaciones las funciones de distribución por las correspondientes funciones de distribución empíricas.

Sin embargo, estos estimadores se expresan en forma más simple usando las denominadas ubicaciones, que definimos a continuación.

Definición: Dados X_1, \dots, X_n e Y_1, \dots, Y_m , la ubicación de X_i entre Y_1, \dots, Y_m es

$$\rho_i(x) = \#\{j / Y_j \leq X_i\}$$

y la ubicación de Y_j entre X_1, \dots, X_n es

$$\rho_j(y) = \#\{i / X_i \leq Y_j\}$$

Entonces, un estimador de $1 - p_1 = \int H(t)g(t)dt$ es

$$\int H_m(t)dG_n(t) = \frac{1}{n} \sum_{i=1}^n H_m(X_i) = \frac{1}{nm} \sum_{i=1}^n \rho_i(x) = \frac{1}{m} \bar{\rho}(x)$$

De la misma forma se expresan todos los otros estimadores, y llamando

$$s^2(x) = \sum_{i=1}^n (\rho_i(x) - \bar{\rho}(x))^2$$
$$s^2(y) = \sum_{j=1}^m (\rho_j(y) - \bar{\rho}(y))^2$$

el estimador de la varianza de W será

$$V(\hat{W}) = \bar{\rho}(x) \bar{\rho}(y) + s^2(x) + s^2(y)$$

y, como las ubicaciones son funciones de los rangos, el estadístico

$$\hat{W} = \frac{W - mn/2}{\sqrt{V(\hat{W})}}$$

es un estadístico de rangos. El test rechazará $H_0: \Delta = 0$ a favor de $H_1: \Delta > 0$ a nivel α cuando $\hat{W} > z_\alpha$.

Si $\Delta = 0$ y $G = H$, \hat{W} es distribución libre y su distribución ha sido tabulada para algunos tamaños de muestra por Fligner y Policello (1981). Sin embargo, para asegurar que \hat{W} siga siendo al menos asintóticamente distribución libre cuando $\Delta = 0$ y $G \neq H$, debe suponerse que $G, H \in \Omega_s$.

En la tabla que sigue (Fligner y Policello (1981)) se presentan los niveles empíricos, para un nivel nominal 0.05, basados en 10000 simulaciones para W, \hat{W}, t y t_w , siendo este último el estadístico de Welch. Los tamaños de muestra son $n=11$ y $m=10$ y las distribuciones consideradas son la distribución Normal y la Normal contaminada con $\varepsilon = 0.1$, con H de la forma $H(t)=G(t/\sigma)$ para diferentes valores de σ .

Distribución	σ	W	\hat{W}	t	t_w
Normal	0.1	0.081	0.048	0.048	0.048
	0.25	0.069	0.054	0.050	0.052
	1	0.050	0.048	0.048	0.047
	4	0.071	0.054	0.060	0.047
	10	0.082	0.062	0.069	0.052
Normal contaminada	0.1	0.076	0.051	0.033	0.034
	0.25	0.065	0.052	0.033	0.033
	1	0.048	0.046	0.035	0.033
	4	0.068	0.052	0.043	0.032
	10	0.083	0.063	0.050	0.035

Se observa que el test basado en \hat{W} es el más estable, superando al test de Welch en el caso contaminado. Los mismos autores simularon otras distribuciones subyacentes e incluyeron un estudio de la potencia de los tests, concluyendo que el test basado en \hat{W} es superior en el sentido de que conserva su nivel y al mismo tiempo alcanza alta potencia.

Scores generales:

Definición: Sean $N=n+m$, $a(1) \leq \dots \leq a(N)$ una sucesión no constante y definamos

$$V = \sum_{j=1}^m a(R_j)$$

donde $R_j = R(Y_j)$ es el rango de Y_j en la muestra combinada. V se denomina estadístico de scores generales con scores $a(i)$.

Teorema: Bajo $H_0: \Delta = 0$,

$$E(V) = \frac{m}{N} \sum_{i=1}^N a(i) = m \bar{a}$$

$$Var(V) = \frac{nm}{N(N-1)} \sum_{i=1}^N (a(i) - \bar{a})^2$$

Demostración: $E(V) = \sum_{i=1}^m E(a(R_j)) = mE(a(R_1)) = m \sum_{s=1}^N a(s) \frac{1}{N} = m \bar{a}$.

$$E(a^2(R_1)) = \sum_{s=1}^N a^2(s) \frac{1}{N}$$

entonces, $Var(a(R_1)) = \sum_{s=1}^N a^2(s) \frac{1}{N} - \bar{a}^2 = \frac{\sum_{s=1}^N (a(s) - \bar{a})^2}{N}$

Además,

$$\begin{aligned} cov(a(R_1), a(R_2)) &= E((a(R_1) - \bar{a})(a(R_2) - \bar{a})) = \sum_{i \neq j} \frac{1}{N(N-1)} (a(i) - \bar{a})(a(j) - \bar{a}) = \\ &= \frac{1}{N(N-1)} \sum_{i=1}^N (a(i) - \bar{a}) \sum_{j \neq i} (a(j) - \bar{a}) = -\frac{1}{N(N-1)} \sum_{i=1}^N (a(i) - \bar{a})^2 \end{aligned}$$

Entonces,

$$\begin{aligned} Var(V) &= Var \sum_{j=1}^m a(R_j) = m Var(a(R_1)) + 2 \frac{m(m-1)}{2} cov(a(R_1), a(R_2)) = \\ &= \frac{m}{N} \sum_{s=1}^N (a(s) - \bar{a})^2 - \frac{m(m-1)}{N(N-1)} \sum_{s=1}^N (a(s) - \bar{a})^2 = \frac{m}{N(N-1)} \sum_{s=1}^N (a(s) - \bar{a})^2 (N-1-m+1) = \\ &= \frac{mn}{N(N-1)} \sum_{s=1}^N (a(s) - \bar{a})^2 \end{aligned}$$

Más aún, puede probarse que si $a(j) + a(N + 1 - j) = K$ (constante) para todo j , entonces bajo H_0 , V tiene distribución simétrica alrededor de $m\bar{a}$.

Definición: Sea $\psi(u)$, $0 < u < 1$, una función no decreciente. Supongamos además que

$$0 < \int_0^1 (\psi(u) - \bar{\psi})^2 du < \infty \quad \text{con} \quad \bar{\psi} = \int_0^1 \psi(u) du$$

Si $a(i) = \psi\left(\frac{i}{N+1}\right)$, definimos el estadístico generado por la función generadora de scores ψ , como:

$$\bar{V} = \frac{1}{N} \sum_{j=1}^m a(R(Y_j))$$

Distribución asintótica bajo H_0 : En el caso de una muestra, el estadístico \bar{V} era asintóticamente normal y el resultado se obtenía aplicando el TCL. En este caso, \bar{V} no es suma de v.a. independientes y es necesario usar el teorema de proyecciones para obtener su distribución asintótica. Hajek y Sidak (1967) prueban que si $\min(n, m) \rightarrow \infty$, entonces bajo H_0 ,

$$\frac{\bar{V} - E(\bar{V})}{\sqrt{\text{Var}(\bar{V})}} \xrightarrow{d} N(0,1)$$

Supongamos que $\frac{n}{N} \rightarrow \lambda$, $0 < \lambda < 1$, es decir que ninguno de los tamaños de muestra domina asintóticamente, entonces

$$E(\bar{V}) = \frac{m}{N} \bar{a} = \frac{m}{N} \frac{1}{N} \sum_{i=1}^N \psi\left(\frac{i}{N+1}\right) \rightarrow (1 - \lambda) \bar{\psi}$$

$$N \text{Var}(\bar{V}) = \frac{mn}{N^2(N-1)} \sum_{i=1}^N (a(i) - \bar{a})^2 \rightarrow \lambda(1 - \lambda) \int_0^1 (\psi(u) - \bar{\psi})^2 du$$

y, bajo H_0

$$\sqrt{N} \frac{\bar{V} - (1 - \lambda) \bar{\psi}}{\left[\lambda(1 - \lambda) \int_0^1 (\psi(u) - \bar{\psi})^2 du \right]^{1/2}} \xrightarrow{d} N(0,1)$$

Algunos ejemplos de funciones de scores:

1) Scores normales de Van der Waerden: Sea $\psi(u) = \Phi^{-1}(u)$, entonces el estadístico de scores normales se define como

$$\bar{V} = \frac{1}{N} \sum_{j=1}^m \Phi^{-1}\left(\frac{R_j}{N+1}\right)$$

En este caso, $\bar{\psi} = \int_0^1 \Phi^{-1}(u) du = 0$ y $\int_0^1 (\psi(u) - \bar{\psi})^2 du = 1$. Para probarlo basta hacer el cambio de variable $\Phi^{-1}(u) = t$.

Luego si $\frac{n}{N} \cong \lambda$, $1 - \lambda \cong \frac{m}{N}$ y bajo H_0 , $\left(\frac{N^3}{mn}\right)^{1/2} \bar{V} \xrightarrow{d} N(0,1)$

2) Estadístico de Mood: Sea

$$\psi(u) = \begin{cases} 1 & 1/2 < u < 1 \\ 0 & 0 < u \leq 1/2 \end{cases}$$

entonces el estadístico de Mood, que es el número de Y_i 's que superan a la mediana de la muestra combinada (test de la mediana) se puede escribir como

$$V_+^* = \sum_{j=1}^m \psi\left(\frac{R_j}{N+1}\right)$$

Es fácil verificar que $\int_0^1 \psi(u) du = 1/2$ y $\int_0^1 (\psi(u) - \bar{\psi})^2 du = 1/4$.

Supongamos que $N = n+m = 2r$ y $m \leq n$, entonces

$$a(i) = \psi\left(\frac{i}{N+1}\right) = \begin{cases} 0 & 1 \leq i \leq r \\ 1 & r+1 \leq i \leq N \end{cases}$$

$$E(V_+^*) = m \bar{a} = m \frac{r}{N} = \frac{m}{2} = m \bar{\psi}$$

$$Var(V_+^*) = \frac{mn}{N(N-1)} \sum (a(s) - \bar{a})^2 = \frac{mn}{4(N-1)} = \frac{mn}{N-1} \int_0^1 (\psi(u) - \bar{\psi})^2 du$$

Por lo tanto, si denotamos $\bar{V}_+^* = \frac{V_+^*}{N}$,

$$\sqrt{N} \frac{\bar{V}_+^* - \frac{1-\lambda}{2}}{\sqrt{\frac{\lambda(1-\lambda)}{4}}} \xrightarrow{d} N(0,1)$$

3) **¿Cómo construir scores para el problema de dos muestras a partir de los de una muestra?:** Sea $\psi^+(u), 0 < u < 1$, una función generadora de scores para el problema de una muestra y el correspondiente estadístico

$$\bar{V}^+ = \frac{1}{n} \sum_{j=1}^n \psi^+ \left(\frac{R_j}{n+1} \right) s(X_j)$$

Consideremos ahora la siguiente extensión de ψ^+ al $(-1,1)$:

$$\psi^+(-u) = -\psi^+(u) \quad \text{si } u \in (0,1)$$

Sea

$$\psi(u) = \psi^+(2u-1) \quad \text{si } 0 < u < 1$$

$$\psi(u) = \begin{cases} -\psi^+(1-2u) & 0 < u < 1/2 \\ \psi^+(2u-1) & 1/2 < u < 1 \end{cases}$$

Se define el estadístico para el test de dos muestras en la forma $\bar{V} = \frac{1}{N} \sum_{j=1}^m \psi\left(\frac{R(Y_j)}{N+1}\right)$.

En este caso, $\bar{\psi} = 0$ y $\int_0^1 (\psi(u) - \bar{\psi})^2 du = \int_0^1 (\psi^+(v))^2 dv$ y se puede demostrar que la eficacia del test correspondiente a ψ es $\sqrt{\lambda(1-\lambda)}$ veces la del test para una muestra generado por ψ^+ .

Ejemplos: 3.1) Si $\psi^+(u) = u, 0 < u < 1$, el estadístico \bar{V}^+ es el del test de rangos signados de Wilcoxon. La función generadora $\psi(u) = 2u - 1, 0 < u < 1$ produce el test de Mann-Whitney para dos muestras.

3.2) Si $\psi^+(u) = 1, 0 < u < 1$, el estadístico \bar{V}^+ es el del test del signo. La función generadora $\psi(u) = \begin{cases} -1 & 0 < u < 1/2 \\ 1 & 1/2 < u < 1 \end{cases}$ produce el test de medianas de Mood para dos muestras.

3.3) Si $\psi^+(u) = \Phi^{-1}\left(\frac{u+1}{2}\right), 0 < u < 1$, el estadístico \bar{V}^+ es el correspondiente al test de scores normales. La función generadora $\psi(u) = \Phi^{-1}(u)$ produce el test de scores normales de Van der Waerden para dos muestras.

Eficiencia de scores generales: La eficacia del test de dos muestras generado por la función de scores Ψ es la del test de una muestra generado por $\psi^+(u) = \psi\left(\frac{u+1}{2}\right)$ multiplicada por $\sqrt{\lambda(1-\lambda)}$ y por lo tanto se preservan las eficiencias relativas.