A LEHMANN PRÓBÁRÓL

ZAJTA AURÉL[1]

A LEHMANN-próba két minta összehasonlítására szolgál. A próba a nemparaméteres módszerek közé, pontosabban: a rendezett minták elméletébe tartozik, és a WILCOXON-próba továbbfejlesztésének tekinthető.

Definíciók és jelölések

A két összehasonlítandó minta közül az egyiknek elemeit ξ_i-vel ($i = 1, 2, \ldots m$), a másiknak az elemeit η_k-val ($k = 1, 2, \ldots n$) jelöljük. A minták elemszámai, m és n nem szükségképpen egyenlők. A ξ_i, ill. η_k mintavételi változók eloszlásfüggvényeit $F(x)$-szel, ill. $G(x)$-szel jelöljük; ezekről feltesszük, hogy folytonosak. Így zérussal egyenlő annak a valószínűsége, hogy a ξ_i és η_k változók együttesében két változó értéke megegyezzék, s ezért jogosan tételezhetjük fel, hogy valamennyi ξ_i és η_k különböző. Ebből viszont tovább következik, hogy a ξ_i és η_k változók, külön-külön és együttesen is, szigorúan növekvő sorozatba rendezhetők.

A nagyság szerint rendezett ξ-változók közül az i-ediket ξ_i^*-gal, s hasonlóképp az η-változók rendezett sorában a k-adikat η_k^*-gal jelöljük. Ha a két mintát egyesítjük és ismét nagyság szerint növekvő sorozatba rendezzük, a ξ_i^*, ill. η_k^* változók pozíciója általában megváltozik, rangszámaik általában nagyobbak, mint az eredeti i, ill. k rangszámok (sorszámok). A ξ_i^* új rangszámát r_i-vel, az η_k^* új rangszámát pedig s_k-val jelölve, fennáll, hogy

$$(1) \qquad i + k = \begin{cases} r_i, & \text{ha} \quad \eta_k^* < \xi_i^* < \eta_{k+1}^*, \\ s_k, & \text{ha} \quad \xi_i^* < \eta_k^* < \xi_{i+1}^*. \end{cases}$$

A minták eloszlásfüggvényeit (az ún. empirikus eloszlásfüggvényeket) $F_m(x)$-szel, ill. $G_n(x)$-szel jelöljük. Ezek definíciója, mint ismeretes, a következő:

$$(2) \qquad F_m(x) = \begin{cases} 0\,, & \text{ha} \quad x \leqq \xi_1^*, \\ \dfrac{i}{m}\,, & \text{ha} \quad \xi_i^* < x \leqq \xi_{i+1}^*, \\ 1\,, & \text{ha} \quad \xi_m^* < x, \end{cases}$$

[1] Agrártudományi Egyetem.

$$G_n(x) = \begin{cases} 0\,, & \text{ha} \quad x \leqq \eta_1^*\,, \\ \dfrac{k}{n}\,, & \text{ha} \quad \eta_k^* < x \leqq \eta_{k+1}^*\,, \\ 1\,, & \text{ha} \quad \eta_n^* < x\,. \end{cases} \tag{2a}$$

Szükségünk lesz még a z_{ik} és $\bar{z}_{ik}$ karakterisztikus változókra. Ezeket az alábbi módon értelmezzük:

$$z_{ik} = \begin{cases} 0\,, & \text{ha} \quad \eta_k < \xi_i\,, \\ 1\,, & \text{ha} \quad \eta_k > \xi_i\,. \end{cases} \tag{3}$$

$$\bar{z}_{ik} = 1 - z_{ik} = \begin{cases} 0\,, & \text{ha} \quad \eta_k < \xi_i\,, \\ 1\,, & \text{ha} \quad \eta_k > \xi_i\,. \end{cases} \tag{4}$$

Figyelemre méltó, hogy míg az empírikus eloszlásfüggvények definíciói a rendezett mintaelemek alapján történtek, addig a z_{ik} definíciójához az eredeti mintaelemeket használtuk fel.

Bevezetés

A Lehmann-próbánál, akárcsak a Wilcoxon próbánál, nullhipotézisnek az $F(x) = G(x)$ hipotézist, alternatív hipotézisnek pedig az ettől különböző hipotéziseket nevezzük, tehát minden alternatív hipotézisnél legalább egy mérhető halmazon $F(x) \neq G(x)$. Amíg azonban a Wilcoxon próbánál az

$$I_1 = \int_{-\infty}^{\infty} \big(G(x) - F(x)\big)\, dF(x)\,. \tag{5}$$

Stieltjes-integrált becsüljük meg, melynek értéke alternatív hipotézis esetén is zérussá válhat, a Lehmann-próbánál lényegében az

$$I_2 = \int_{-\infty}^{\infty} \big(G(x) - F(x)\big)^2\, dF(x) \tag{6}$$

integrált becsüljük, amely csak a két eloszlásfüggvény egyezése esetén tűnhet el. A két eloszlásfüggvény különböző volta viszont mindig egy irányban (pozitív irányban) tolja el az I_2 értékét, amiből következik, hogy az I_2 becslésére szolgáló mennyiségnek is csak a pozitív irányú eltéréséből következtethetünk a két eloszlásfüggvény különböző voltára, más szóval: a próba egyoldalú.

A Lehmann- és a Wilcoxon-próba rendstatisztikai jellege abból a tényből fakad, hogy az I_2 és az I_1 Stieltjes-integrálok becsléséhez nincs szükségünk a mintaelemek tényleges értékeire, csupán egymáshoz viszonyított helyzetükre, ezt pedig a rangszámok egyértelműen meghatározzák. Az I_2-vel kapcsolatos becslés céljára Lehmann [1] és Rényi [2] más-más kifejezést adtak; a Rényitől származó W lényegesen egyszerűbb, mint a Lehmann-féle L-lel jelölt kifejezés. Az L és W között egyszerű lineáris kapcsolat áll fenn:

$$L = 2W - 1\,. \tag{7}$$

Ezt CSÁKI E. ismerte fel először [3], s ezzel bebizonyította, hogy a RÉNYI által javasolt próba azonos a LEHMANN-próbával.

E cikkben új bizonyítását adom a (7) összefüggésnek, s ezenkívül megadok egy harmadik statisztikai függvényt is (ezt V-vel fogjuk jelölni) a próba végrehajtására, mely bizonyos szempontból még a RÉNYI-féle W-nél is egyszerűbb. Végül kiszámítom a W szórásnégyzetét az általános esetben (tehát alternatív hipotézis esetén), ahonnan $G(x) = F(x)$ helyettesítésre könnyen adódik a szórásnégyzetnek nullhipotézis teljesülése esetén érvényes képlete.

1. §. A próba alapjául szolgáló kifejezések

A LEHMANN-féle L-függvény eredeti definíciója:

$$\frac{L_1 + L_2}{\binom{m}{2}\binom{n}{2}}, \tag{8}$$

ahol L_1 jelenti azon $(\xi_i, \xi_j, \eta_k, \eta_l)$ elemnégyek számát, melyekre az

$$\eta_k < \xi_i, \ \eta_k < \xi_i, \ \eta_l < \xi_i, \ \eta_l < \xi_j$$

egyenlőtlenségek egyidejűleg teljesülnek, s hasonlóképpen L_2 jelenti azon $(\xi_i, \xi_j, \eta_k, \eta_l)$ elemnégyesek számát, melyekre az

$$\eta_k > \xi_i, \ \eta_k > \xi_j, \ \eta_l > \xi_i, \ \eta_l > \xi_j$$

egyenlőtlenségek egyszerre teljesülnek. A bevezetett karakterisztikus változókkal L_1 és L_2 a következőképp írható:

$$\left\{\begin{aligned} L_1 &= \sum_{i,j,k,l} z_{ik} z_{il} z_{jk} z_{jl}, \\ L_2 &= \sum_{i,j,k,l} \bar{z}_{ik} \bar{z}_{il} \bar{z}_{jk} \bar{z}_{jl}, \end{aligned}\right. \tag{9}$$

megjegyezve, hogy itt (és a továbbiakban mindenütt) az egyező helyen álló de eltérő módon jelölt indexek egyúttal eltérő értékű indexeket is jelentenek

A képletekből látható, hogy az $\dfrac{L_1}{\binom{m}{2}\binom{n}{2}}$ kifejezéssel a $z_{ik}z_{il}z_{jk}z_{jl}$ szorzat várható értékét, az $\dfrac{L_2}{\binom{m}{2}\binom{n}{2}}$ kifejezéssel pedig a $\bar{z}_{ik}\bar{z}_{il}\bar{z}_{jk}\bar{z}_{jl}$ várható értékét becsüljük meg. Mivel pedig

$$\left\{\begin{aligned} \mathbf{M}\,(z_{ik} z_{il} z_{jk} z_{jl}) &= 2 \int_{-\infty}^{\infty} G^2(x)\big(1 - F(x)\big)\, dF(x), \\ \mathbf{M}\,(\bar{z}_{ik} \bar{z}_{il} \bar{z}_{jk} \bar{z}_{jl}) &= 2 \int_{-\infty}^{\infty} F^2(x)\big(1 - G(x)\big)\, dG(x), \end{aligned}\right. \tag{10}$$

kis számítással azonnal adódik, hogy

$$\mathbf{M}(L) = \frac{1}{3} + 2\int_{-\infty}^{\infty} (G(x) - F(x))^2\, dF(x) = \frac{1}{3} + 2I_2. \tag{11}$$

A Rényi-féle W-kifejezés definíciója egyszerűbb:

$$W = \frac{W_1}{m\binom{n}{2}} + \frac{W_2}{n\binom{m}{2}}. \tag{12}$$

Itt W_1 jelenti azon (ξ_i, η_k, η_l) elemhármasok számát, amelyekre

$$\eta_k < \xi_i \text{ és } \eta_l < \xi_i$$

egyidejűleg teljesül, s hasonlóképp W_2 jelenti azon (ξ_i, ξ_j, η_k) elemhármasok számát, amelyekre

$$\eta_k > \xi_i \text{ és } \eta_k > \xi_j$$

egyszerre teljesül. A W_1 és W_2 karakterisztikus változókkal így fejezhetők ki:

$$\begin{cases} W_1 = \sum\limits_{i,k,l} z_{ik} z_{il}, \\ W_2 = \sum\limits_{i,j,k} \bar{z}_{ik} \bar{z}_{jk}. \end{cases} \tag{13}$$

Ezeknek megfelelően a $\dfrac{W_1}{m\binom{n}{2}}$-vel a $z_{ik}z_{il}$ szorzat várható értékét, a $\dfrac{W_2}{n\binom{m}{2}}$-vel pedig a $\bar{z}_{ik}\bar{z}_{jk}$ szorzat várható értékét becsüljük meg. Minthogy pedig

$$\begin{cases} \mathbf{M}(z_{ik} z_{il}) = \int\limits_{-\infty}^{\infty} G^2(x)\, dF(x), \\ \mathbf{M}(\bar{z}_{ik} \bar{z}_{jk}) = \int\limits_{-\infty}^{\infty} F^2(x)\, dG(x), \end{cases} \tag{14}$$

kis számítással most azt kapjuk, hogy

$$\mathbf{M}(W) = \frac{2}{3} + \int_{-\infty}^{\infty} (G(x) - F(x))^2\, dF(x) = \frac{2}{3} + I_2. \tag{15}$$

A (11)-ből és (15)-ből következik az

$$\mathbf{M}(L) = 2\mathbf{M}(W) - 1,$$

ill.

$$\mathbf{M}(2W - L - 1) = 0$$

összefüggés. Ebből már sejteni lehet, hogy a $(2W - L - 1)$ mennyiségnek nemcsak a várható értéke, hanem az értéke is zérus, azaz a (7) is fennáll.[2] Ennek bizonyítását a 2. pontban végezzük el. A bizonyítás a W és L explicit alakjainak összehasonlításával történik, ezért előbb az explicit alakokat adjuk közre:

$$
(16) \qquad \begin{cases} L_1 = \dfrac{1}{2} \cdot \sum\limits_{i=1}^{m} (r_i - i)(r_i - i - 1)(m - i)\,, \\[2ex] L_2 = \dfrac{1}{2} \cdot \sum\limits_{k=1}^{m} (s_k - k)(s_k - k - 1)(n - k)\,, \end{cases}
$$

$$
(17) \qquad \begin{cases} W_1 = \dfrac{1}{2} \cdot \sum\limits_{i=1}^{n} (r_i - i)(r_i - i - 1)\,, \\[2ex] W_2 = \dfrac{1}{2} \cdot \sum\limits_{k=1}^{n} (s_k - k)(s_k - k - 1)\,. \end{cases}
$$

E rangszámok függvényeiként megadott formulák könnyen beláthatók a fentebb közölt definíciók alapján.

2. §. Identikus összefüggések levezetése

Az (1) összefüggést felhasználva, rendre kiszámítjuk az empirikus eloszlásfüggvényekből képezett alábbi Stieltjes-integrálokat:

$$
(18) \qquad \int_{-\infty}^{\infty} G_n(x)\, dF_m(x) = \sum_{i=1}^{m} \frac{k}{n} \cdot \frac{1}{m} = \frac{1}{nm} \cdot \sum_{i=1}^{m} (r_i - i)\,,
$$

$$
(19) \qquad \int_{-\infty}^{\infty} F_m(x)\, dG_n(x) = \sum_{k=1}^{n} \frac{i}{m} \cdot \frac{1}{n} = \frac{1}{mn} \cdot \sum_{k=1}^{n} (s_k - k)\,,
$$

$$
(20) \qquad \int_{-\infty}^{\infty} G_n^2(x)\, dF_m(x) = \sum_{i=1}^{m} \left(\frac{k}{n}\right)^2 \cdot \frac{1}{m} = \frac{1}{n^2 m} \cdot \sum_{i=1}^{m} (r_i - i)^2\,,
$$

$$
(21) \qquad \int_{-\infty}^{\infty} F_m^2(x)\, dG_n(x) = \sum_{k=1}^{n} \left(\frac{i}{m}\right)^2 \cdot \frac{1}{n} = \frac{1}{m^2 n} \cdot \sum_{k=1}^{n} (s_k - k)^2\,,
$$

[2] A várható értékek között fennálló összefüggésből tulajdonképpen szintén következik a megfelelő statisztikai függvények közötti hasonló összefüggés, ha alkalmazzuk Lehmann és Scheffé egy lemmáját (vö. [1] vagy [3], 314. o.).

$$(22)\qquad \int_{-\infty}^{\infty} F_m(x)\, dG_n^2(x) = \sum_{k=1}^{n} \frac{i}{m}\cdot\left\{\left(\frac{k}{n}\right)^2 - \left(\frac{k-1}{n}\right)^2\right\} =$$

$$= \frac{1}{mn^2}\cdot \sum_{k=1}^{n} (s_k - k)\,(2\,k-1)\,,$$

$$(23)\qquad \int_{-\infty}^{\infty} G_n(x)\, dF_m^2(x) = \sum_{i=1}^{m} \frac{k}{n}\cdot\left\{\left(\frac{i}{m}\right)^2 - \left(\frac{i-1}{m}\right)^2\right\} =$$

$$= \frac{1}{nm^2}\cdot \sum_{i=1}^{m} (r_i - i)\,(2\,i-1)\,,$$

$$(24)\qquad \int_{-\infty}^{\infty} G_n^2(x)\, dF_m^2(x) = \sum_{i=1}^{m} \left(\frac{k}{n}\right)^2 \cdot \frac{2i-1}{m^2} = \frac{1}{n^2 m^2}\cdot \sum_{i=1}^{m} (r_i - i)^2\cdot (2i-1)\,,$$

$$(25)\qquad \int_{-\infty}^{\infty} F_m^2(x)\, dG_n^2(x) = \sum_{k=1}^{n} \left(\frac{i}{m}\right)^2 \cdot \frac{2\,k-1}{n^2} = \frac{1}{m^2 n^2}\cdot \sum_{k=1}^{m} (s_k - k)^2\cdot (2\,k-1)\,.$$

A parciális integrálás képletét alkalmazva, a (18)- és (19)-ből az alábbi eredményt nyerjük:

$$\int_{-\infty}^{\infty} G_n(x)\, dF_m(x) + \int_{-\infty}^{\infty} F_m(x)\, dG_n(x) = \Big[G_n(x)\cdot F_m(x)\Big]_{-\infty}^{\infty} =$$

$$= 1 = \frac{1}{nm}\cdot \sum_{i=1}^{m} (r_i - i) + \frac{1}{mn}\cdot \sum_{k=1}^{n} (s_k - k)\,,$$

azaz

$$(26)\qquad \sum_{i=1}^{m} (r_i - i) + \sum_{k=1}^{n} (s_k - k) = mn\,.$$

Hasonló összefüggéseket kapunk a (20)—(22), (21)—(23), ill. a (24) és (25) összevetésével:

$$(27)\qquad \sum_{i=1}^{m} (r_i - i)^2 + \sum_{k=1}^{n} (s_k - k)\,(2\,k-1) = n^2\, m\,,$$

$$(28)\qquad \sum_{k=1}^{n} (s_k - k)^2 + \sum_{i=1}^{m} (r_i - i)\,(2\,i-1) = m^2\, n\,,$$

$$(29)\qquad \sum_{i=1}^{m} (r_i - i)^2\,(2\,i-1) + \sum_{k=1}^{n} (s_k - k)^2\,(2\,k-1) = m^2\, n^2\,.$$

Megjegyezzük, hogy a (26) közvetlenül is levezethető:

$$\sum_{i=1}^{m}(r_i - i) + \sum_{k=1}^{n}(s_k - k) = \sum_{t=1}^{m+n} t - \sum_{i=1}^{m} i - \sum_{k=1}^{n} k =$$

$$= \frac{1}{2}(m+n)(m+n+1) - \frac{1}{2}m(m+1) - \frac{1}{2}n(n+1) = mn\,.$$

Fentebb mégis azért alkalmaztuk az integrálokkal történő levezetést, hogy ezzel a módszert bemutassuk.

A kapott identitások kombinációjából újabbakat nyerhetünk. Adjuk össze a (29) és (26) megfelelő oldalait és vonjuk ki ezekből a (27) és (28) megfelelő oldalait. Az összevonások elvégzése és 4-gyel való osztás után az alábbi összefüggés származik:

$$\frac{1}{2}\cdot\sum_{i=1}^{m}(r_i - i)(r_i - i - 1)(i-1) + \frac{1}{2}\cdot\sum_{k=1}^{n}(s_k - k)(s_k - k - 1)(k-1) =$$

$$= \binom{m}{2}\cdot\binom{n}{2}\,.$$

Ennek felhasználásával és a (17)-re való tekintettel a Lehmann-féle L-függvény a következő alakra hozható:

$$L = \frac{\frac{1}{2}\cdot\sum_{i=1}^{m}(r_i-i)(r_i-i-1)(m-1) + \frac{1}{2}\cdot\sum_{k=1}^{n}(s_k-k)(s_k-k-1)(n-1) - \binom{m}{2}\binom{n}{2}}{\binom{m}{2}\binom{n}{2}} =$$

$$= \frac{\sum_{i=1}^{m}(r_i - i)(r_i - i - 1)}{m\binom{n}{2}} + \frac{\sum_{k=1}^{n}(s_k - k)(s_k - k - 1)}{n\binom{m}{2}} - 1\,,$$

amiről már nem nehéz felismerni, hogy $(2W - 1)$-gyel egyenlő (vö. (12) és (17)). Ezzel a (7) azonosságot bebizonyítottuk.

További újabb identitásokhoz jutunk, ha a (27), ill. (28) megfelelő oldalaiból kivonjuk a (26) megfelelő oldalait:

$$\sum_{i=1}^{m}(r_i - i)(r_i - i - 1) + 2\sum_{k=1}^{n}(s_k - k)(k-1) = nm(n-1)\,, \tag{30}$$

$$\sum_{k=1}^{n}(s_k - k)(s_k - k - 1) + 2\sum_{i=1}^{m}(r_i - i)(i-1) = nm(m-1)\,. \tag{31}$$

Ha bevezetjük a következő jelöléseket:

$$(32)\qquad \begin{aligned} V_1 &= \sum_{i=1}^{m} (r_i - i)(i-1), \\ V_2 &= \sum_{k=}^{n} (s_k - k)(k-1), \end{aligned}$$

ezeknek és a (17) alatti összefüggéseknek figyelembevételével a (30) és (31). így is írható:

$$(33)\qquad \begin{aligned} W_1 + V_2 &= m\binom{n}{2}, \\ W_2 + V_1 &= n\binom{m}{2}. \end{aligned}$$

Amennyiben tehát a

$$(34)\qquad V = \frac{V_1}{nm(m-1)} + \frac{V_2}{mn(n-1)}$$

képlettel definiálunk új statisztikai függvényt, ez szintén felhasználható a próba céljára, mivel a W-vel a következő kapcsolatban áll:

$$(35)\qquad V = 1 - \frac{W}{2}.$$

A (35)-ös összefüggés a (12), (33) és (34) összevetéséből adódik.

Néhány megjegyzés az új kifejezéssel kapcsolatban. A (34)-gyel definiált V azért tekinthető egyszerűbbnek a W-nél, mivel a rangszámoknak csak lineáris kifejezéseit tartalmazza. Ez mindenek előtt az elméleti számításoknál jelenthet előnyt, mivel a próba gyakorlati alkalmazásánál felmerülő numerikus számításoknál a W könnyebben számítható ki, mint a V, annak ellenére, hogy a rangszámokban kvadratikus. A V-nek a gyakorlatban való felhasználásánál ügyelni kell arra is, hogy a nullhipotézis nem-teljesülésére most a negatív irányú eltérésből következtetünk, tekintettel a (35)-re.

Az L-hez és W-hez hasonlóan a V is definiálható a mintaváltozók nagyságrendi eseményei segítségével. V_1 jelenti azon (η_k, ξ_i, ξ_j) elemhármasok számát, melyekre

$$\eta_k < \xi_j, \quad \xi_i < \xi_j$$

és V_2 azon (ξ_i, η_k, η_l) elemhármasok számát, melyekre

$$\xi_i < \eta_l, \quad \eta_k < \eta_l$$

egyidejűleg teljesül. Tekintettel azonban arra, hogy

$$\mathbf{P}(\eta_k < \xi_j,\ \xi_i < \xi_j) = \int_{-\infty}^{\infty} G(x)\,F(x)\,dF(x),$$

és

$$\mathbf{P}(\xi_i < \eta_l,\ \eta_k < \eta_l) = \int_{-\infty}^{\infty} F(x)\,G(x)\,dG(x),$$

a V-vel tulajdonképpen a jobboldalakon álló két integrál összegét becsüljük meg, azaz

$$\mathbf{M}(V) = \int_{-\infty}^{\infty} G(x)\, F(x)\, dF(x) + \int_{-\infty}^{\infty} F(x)\, G(x)\, dG(x)\,,$$

vagy a parciális integrálások megfelelő alkalmazásával:

$$\mathbf{M}(V) = \frac{2}{3} - \frac{1}{2} \cdot \int_{-\infty}^{\infty} \big(G(x) - F(x)\big)^2 \cdot dF(x) = \frac{2}{3} - \frac{1}{2} \cdot I_2\,,$$

teljes összhangban a (35)-tel és (15)-tel.

3. §. A W szórásnégyzetének kiszámítása

Alternatív hipotézis esetén a három statisztikai függvény (L, W és V) közül aránylag a W szórásnégyzete határozható meg legkönnyebben, az is csak akkor, ha a (12) és (13) alakokból indulunk ki. A (12)-ből következik:

$$\text{(36)} \qquad \mathbf{D}^2(W) = \frac{1}{m^2 \cdot \binom{n}{2}^2} \cdot \mathbf{D}^2(W_1) + \frac{1}{n^2 \binom{m}{2}^2} \cdot \mathbf{D}^2(W_2) + \frac{2\,\mathrm{Cov}\,(W_1, W_2)}{mn \binom{m}{2} \binom{n}{2}}\,.$$

Cov-val a zárójelbe tett két változó *kovarianciáját* jelöltük, ennek definíciója

$$\mathrm{Cov}(\zeta_1, \zeta_2) = \mathbf{M}\left\{\big(\zeta_1 - \mathbf{M}(\zeta_1)\big)\big(\zeta_2 - \mathbf{M}(\zeta_2)\big)\right\}.$$

A (36) szerint a W szórásnégyzetéhez a W_1 és W_2 kovarianciáját és szórásnégyzeteit kell előbb meghatároznunk.

A W_1 szórásnégyzetének kiszámítását az alábbi számításmenet mutatja:

$$W_1 = \sum_{i,k,l} z_{ik} z_{il}\,,$$

$$\text{(37)} \qquad \mathbf{D}^2(W_1) = \mathbf{M}\left[\Big(\sum_{i,k,l} z_{ik} z_{il}\Big)^2\right] - \left[\mathbf{M}\Big(\sum_{i,k,l} z_{ik} z_{il}\Big)\right]^2,$$

$$\mathbf{M}\Big(\sum_{i,k,l} z_{ik} z_{il}\Big) = \sum_{i,k,l} \mathbf{M}(z_{ik} z_{il}) = m \binom{n}{2} \cdot p_2\,,$$

$$\Big(\sum_{i,k,l} z_{ik} z_{il}\Big)^2 = \sum z_{ik}^2 z_{il}^2 + 2 \sum z_{ik} z_{il} z_{jk} z_{jl} + 2 \sum z_{ik_1} z_{il} z_{jk_2} z_{jl} +$$

$$+ 2 \sum z_{ik_1} z_{il_1} z_{jk_2} z_{jl_2} + 2 \sum z_{ik_1} z_{il} z_{ik_2} z_{il} + 2 \sum z_{ik_1} z_{il_1} z_{ik_2} z_{il_2}\,,$$

$$(38)\quad \begin{cases} \mathbf{M}(\sum z_{ik}^2 z_{il}^2) = m\binom{n}{2}\cdot p_2\,, \\ \mathbf{M}(2\sum z_{ik}\, z_{il}\, z_{jk}\, z_{jl}) = 2\binom{m}{2}\binom{n}{2}(2\,p_2 - 2\,p_{21})\,, \\ \mathbf{M}(2\sum z_{ik_1} z_{il}\, z_{jk_2} z_{jl}) = 2\binom{m}{2} n(n-1)(n-2)\cdot 2\,s\,, \\ \mathbf{M}(2\sum z_{ik_1} z_{il_1} z_{jk_2} z_{jl_2}) = 2\binom{m}{2}\binom{n}{2}\binom{n-2}{2}\cdot p_2^2\,, \\ \mathbf{M}(2\sum z_{ik_1} z_{il}\, z_{ik_2} z_{il}) = 2\,m\binom{n}{2}\cdot(n-2)\cdot p_3\,, \\ \mathbf{M}(2\sum z_{ik_1} z_{il_1} z_{ik_2} z_{il_2}) = 2\,m\cdot\frac{1}{2}\cdot\binom{n}{2}\binom{n-2}{2}\cdot p_4\,. \end{cases}$$

A jobboldalakon szereplő betűjelek (p_2, p_{21}, s, p_3, p_4) különféle integrálokat jelentenek; pontos jelentésüket a cikk végén levő táblázatban foglaltuk össze. A kapott eredményeknek a (37)-be való behelyettesítésével és az egyszerűsítések elvégzésével azt kapjuk, hogy

$$\mathbf{D}^2(W_1) = m\binom{n}{2}\cdot\begin{Bmatrix} p_2 + 2(m-1)(p_2 - p_{21}) + 4(m-1)(n-2)s - \\ - m(2\,n-3)\,p_2^2 + \frac{1}{2}(n-2)(n-3)(p_4 - p_2^2) + 2(n-2)p_3 \end{Bmatrix}. \tag{39}$$

Teljesen analóg módon nyerhető a W_2 szórásnégyzete is:

$$\mathbf{D}^2(W_2) = $$

$$= n\binom{m}{2}\cdot\begin{Bmatrix} \bar p_2 + 2(n-1)(\bar p_2 - \bar p_{21}) + 4(n-1)(m-2)\bar s - \\ - n(2\,m-3)\bar p_2^2 + \frac{1}{2}(m-2)(m-3)(\bar p_4 - \bar p_2^2) + 2(m-2)\,\bar p_3 \end{Bmatrix}. \tag{40}$$

A W_1 és W_2 kovarianciájának kiszámítása céljából a W_2-t a

$$W_2 = \sum_{i,j,k}(1 - z_{ik})(1 - z_{jk})$$

alakba írva vonjuk be a számításba. A számítás menete

$$\mathrm{Cov}(W_1, W_2) = \sum \mathrm{Cov}\{z_{ik}, z_{il}, (1 - z_{ik})(1 - z_{jk})\} +$$
$$+ \sum \mathrm{Cov}\{z_{ik}, z_{il}, (1 - z_{ik_2})(1 - z_{jk_2})\} + \sum \mathrm{Cov}\{z_{i_1k} z_{i_1l}, (1 - z_{i_2k})(1 - z_{jk})\},$$

$$
(41)\quad \begin{cases}
\sum \mathrm{Cov}\{z_{ik} z_{il}, (1 - z_{ik})(1 - z_{jk})\} = 4 \binom{m}{2} \binom{n}{2} \cdot (- p_2 \overline{p}_2), \\
\sum \mathrm{Cov}\{z_{ik_1} z_{il}, (1 - z_{ik_2})(1 - z_{jk_2})\} = \\
\qquad = 2 \binom{m}{2} \binom{n}{2} (n - 2)(- s - p_{31} - p_2 \overline{p}_2 + p_2), \\
\sum \mathrm{Cov}\{z_{i_1 k} z_{i_1 l}, (1 - z_{i_2 k})(1 - z_{jk})\} = \\
\qquad = \sum \mathrm{Cov}\{(1 - \overline{z}_{i_1 k})(1 - \overline{z}_{i_1 l}), \overline{z}_{i_2 k} \overline{z}_{jk}\} = \\
\qquad = 2 \binom{m}{2} \binom{n}{2} (m - 2)(- \overline{s} - \overline{p}_{31} - p_2 \overline{p}_2 + \overline{p}_2),
\end{cases}
$$

tehát végül is

$$
(42)\quad \mathrm{Cov}(W_1, W_2) = -2 \binom{m}{2} \binom{n}{2} \cdot \begin{Bmatrix} 2 p_2 \overline{p}_2 + (n - 2)(s + p_{31} + p_2 \overline{p}_2 - p_2) + \\ + (m - 2)(\overline{s} + \overline{p}_{31} + p_2 \overline{p}_2 - \overline{p}_2) \end{Bmatrix}.
$$

A W szórásnégyzete ezek után már könnyen kapható, mivel a nyert eredményeket ((39), (40) és (42)) csak be kell helyettesíteni a (36)-ba.

Nullhipotézis esetén a képletek nagyban leegyszerűsödnek. Ha az említett táblázatból a p_2, p_{21}, p_3, stb. integrálok nullhipotézis esetére kiszámított értékeit a képletekbe behelyettesítjük, az alábbi eredményeket kapjuk:

$$
(43.\mathrm{a})\qquad \mathbf{D}^2(W_1) = \frac{1}{90} m \binom{n}{2} (m + n + 1)(4n - 3),
$$

$$
(43.\mathrm{b})\qquad \mathbf{D}^2(W_2) = \frac{1}{90} n \binom{m}{2} (m + n + 1)(4m - 3),
$$

$$
(44)\qquad \mathrm{Cov}(W_1, W_2) = -\frac{4}{45} \binom{m}{2} \binom{n}{2} (m + n + 1),
$$

s ezeknek megfelelően:

$$
(45)\qquad \mathbf{D}^2(W) = \frac{1}{45} \cdot \frac{(m + n + 1)(m + n - 2)}{m(m - 1)\, n(n - 1)}.
$$

Ami a kész alakban megadott (38), ill. (41) alatti közbeeső eredményeket illeti, ezekkel kapcsolatban csak annyit jegyzünk meg, hogy a z_{ik} karakterisztikus változókból alkotható különféle szorzatok várható értékei a közölt táblázat integráljaival adhatók meg. A megfelelő számítások részletezése túlságosan hosszúra nyújtaná a dolgozatot, ezért itt azt elhagyjuk.

Az integrálok táblázata

Az integrál jele	Az integrál kifejezése	Az integrál értéke nullhipotézis esetén
p_2	$\int_{-\infty}^{\infty} G^2(x)\, dF(x)$	$\frac{1}{3}$
$\bar{p}_2$	$\int_{-\infty}^{\infty} F^2(x)\, dG(x)$	$\frac{1}{3}$
p_3	$\int_{-\infty}^{\infty} G^3(x)\, dF(x)$	$\frac{1}{4}$
$\bar{p}_3$	$\int_{-\infty}^{\infty} F^3(x)\, dG(x)$	$\frac{1}{4}$
$p_{21} = \frac{1}{2} - \bar{p}_{21}$	$\int_{-\infty}^{\infty} G^2(x)\, F(x)\, dF(x) = \frac{1}{2} - \int_{-\infty}^{\infty} F^2(x)\, G(x)\, dG(x)$	$\frac{1}{4}$
p_4	$\int_{-\infty}^{\infty} G^4(x)\, dF(x)$	$\frac{1}{5}$
$\bar{p}_4$	$\int_{-\infty}^{\infty} F^4(x)\, dG(x)$	$\frac{1}{5}$
p_{31}	$\int_{-\infty}^{\infty} G^3(x)\, F(x)\, dF(x)$	$\frac{1}{5}$
$\bar{p}_{31}$	$\int_{-\infty}^{\infty} F^3(x)\, G(x)\, dG(x)$	$\frac{1}{5}$
s	$\int_{-\infty}^{\infty} G(x_2) \cdot \left\{ \int_{-\infty}^{x_2} G^2(x_1)\, dF(x_1) \right\} dF(x_2)$	$\frac{1}{15}$
$\bar{s}$	$\int_{-\infty}^{\infty} F(x_2) \cdot \left\{ \int_{-\infty}^{x_2} F^2(x_1)\, dG(x_1) \right\} dG(x_2)$	$\frac{1}{15}$

(Beérkezett: 1959. szeptember 29.)

IRODALOM

[1] LEHMANN, E. L.: „Consistency and unbiasedness of certain nonparametric tests." *Ann. Math. Stat.* **22** (1951) 165—180.

[2] RÉNYI A.: „Újabb kritériumok két minta összehasonlítására." *Alk. Mat. Int. közleményei* **2** (1953) 243—265.

[3] CSÁKI E.: „A Wilcoxon-statisztika két módosításáról." *MTA Mat. Kut. Intézet Közleményei* **4** (1959) 313—319.

ON THE LEHMANN TEST

A. ZAJTA

Abstract

In order to eliminate the deficiencies of the Wilcoxon test, LEHMANN [1] and RÉNYI [2] developed new tests which are consistent with any alternative hypothesis of the null hypothesis. Since the expressions L and W — proposed by LEHMANN and RÉNYI to perform the test — differ from each other, the two procedures have, for a long time, been considered as two separate tests. It was just recently that CSÁKI [3] realized their identity, the relationship between L and W being a simple linear one:

$$L = 2\ W - 1.$$

In the present paper, a new proof of this connection is given and a third expression (V) is put forward to perform the test. The new expression is, in certain respects, even simpler than W in that it contains only linear functions of the rank numbers (Paragraph 2.). In addition, the variance of W is calculated for the general case from which a formula (45), suitable for the substitution $G(x) = F(x)$ and valid for the case of the null hypothesis may be derived.

О ПРОБЕ LEHMANN-A

A. ZAJTA

Резюме

Чтобы освободиться от недостатков пробы WILCOXON-а, LEHMANN [1] и RÉNYI [2] построили новую пробу, консистентную относительно всех альтернативных гипотез гипотезы нуля. Так как выражения L и W, предложенные LEHMANN-ом и RÉNYI для проведения пробы, отличаются друг от друга, долгое время эти две пробы очитались различными, и лишь недавно E. CSÁKY [3] заметил, что они тождественны, так как между L и W имеет место простое линейное соотношение

$$L = 2W - 1.$$

В настоящей работе дается новое доказательство этого соотношения и дается третье выражение (V) для проведения пробы. Новое выражение с некоторой точки зрения проще чем W, так как содержит лишь линейные функции ранговых чисел (пункт 2). Кроме того вычисляется дисперсия W в общем случае, откуда в случае $G(x) = F(x)$ получается формула (45), справедливая в случае гипотезы нуля.