

Emlékeztető a STAB 2016. június 16-ai üléséről

A MTA Statisztikai és Jövőkutatói Tudományos Bizottsága (SJT) **Statisztikai Tudományos Albizottsága (STAB) 2016. június 16-án** ülésezett. Az ülés első napirendi pontja a Statisztikai Szemle 2016. januári számában megjelent *Bárdits Anna – Németh Renáta – Terplán Győző: Egy régi probléma újra előtérben: a nullhipotézis szignifikanciateszt téves gyakorlata* cikk apropóján szervezett beszélgetés volt, amelynek levezetésére Rappai Gábort kértük fel. Felvezetésként a szerzők röviden összegezték cikkük fő üzeneteit, illetve az Amerikai Statisztikai Társaság (ASA) 2016. júniusban megjelent állásfoglalását.

A szerzők szerint a fő problémák között megjelennek matematikai hibák (a tesztek erejének figyelmen kívül hagyása, a szignifikanciatesztek használata kifejezetten nagy mintánál, a p-érték azonosítása a nullhipotézis valószínűségével, az ötszázalékos küszöbértékhez való ragaszkodás, a tesztek feltételeinek figyelmen kívül hagyása), interpretációs hibák (hatásnagyság vizsgálatának hiánya, kizárólag a p-érték alkalmazása, a szakmai fontosság és a statisztikai szignifikancia összevetése) és tudományszociológiai háttérű problémák (üzleti döntések, tudományos konklúziók kizárólag az ötszázalékos küszöbértékre alapozása, elemzési torzítás, publikációs torzítás) is. Rávilágítottak arra a körforgásra, hogy azt mondjuk, azokat a módszereket tanítjuk, amiket a gyakorlatban alkalmaznak, de a gyakorlat azért azt alkalmazza azokat, mert azokat tanulta.

Az ASA állásfoglalása nem-statisztikus kutatóknak, szakíróknak, alkalmazóknak szól. Az állásfoglalás az alkalmazás kultúráját kívánja növelni, téves gyakorlatokra mutat rá, praktikus használati tanácsokat ad (<http://amstat.tandfonline.com/doi/pdf/10.1080/00031305.2016.1154108> és <https://www.amstat.org/newsroom/pressreleases/P-ValueStatement.pdf>).

Az előadás után felkért hozzászólások közül elsőként Singer Júlia, a Klinikai Biostatistikai Társaság elnöke kapott szót. Meglátása szerint a probléma nem új keletű, betiltani semmit sem kell, minden eszökről elmondható, hogy a hibás alkalmazás veszélyes. A kutatásoknak előre tervezettnek kell lennie, a nullhipotézisek sok cikkel ellentétben előre kell, hogy keletkezzenek, amit több vizsgálattal kellene alátámasztani.

Vargha András – aki a Statisztikai Szemle áprilisi számában is reflektált a cikkekre – hozzászólásában azt boncolgatta, hogy mit csinálunk, mit tanítunk rosszul, ha már a szakirodalomban, cikkekben, doktori értekezésekben, szakdolgozatokban rosszul alkalmazott eljárásokat, következtetéseket látni. Az elméleti statisztika megfogalmazza, milyen legyen a minta, ugyanakkor ezek az elvárások a gyakorlatban nem igazán teljesülnek, tehát a levont következtetések nem lesznek hibátlanok. A kérdés, hogy mit engedjünk meg. Véleménye szerint ne mondjunk statisztikai véleményt ott, ahol nem látjuk a valós viszonyokat. Emellett ritkák az igazi alternatív módszerek: ezeket kevesen ismerik, kevesen alkalmazzák, a szoftverekbe nincsenek beépítve. Többen azt mondják, hogy ha nem teljesül a normalitás, akkor alternatív megoldásként alkalmazzunk nemparametrikus teszteket, ugyanakkor azt nem veszik figyelembe, hogy ezeknek a teszteknek pont a normalitás sérülése esetén többnyire más a nullhipotézise. Vargha András véleménye szerint a mi hibánk az, hogy amit tanítunk az nem teljes, illetve nem fér bele a hallgatók fejébe. Amit könyveinkben leírunk, nem feltétlen értik meg, így egyéni utat választanak. Ennek az is az oka, hogy a

statisztikát sok alapképzésben az első évben tanítják, amikor a hallgatók még a szakmai ismereteknek is jócskán híján vannak.

A cikkhez a Statisztikai Szemle áprilisi számában szintén hozzászóló Hunyadi László – Vita László szerzőpáros nevében Hunyadi László hozzászólásában azt vizsgálta, hogy mi lehet az oka a tesztek alkalmazását betiltó lépésnek: egyrészt az alkalmazók nem értik a következtető statisztika lényegét, hátterét, erre jobban rá kellene világítani az oktatás során. Másrészt, a klasszikus statisztikai eszköztár a múlt század elején mérésből alakult ki, ahol kulcs volt az ismételt mintavétel, a nagy minták megléte a maximum likelihood függvények használata, és a normalitás. Ezeket a módszereket átvették más területek, ugyanakkor a problémák nem ugyanazok, az alkalmazók ezt nem veszik figyelembe. Ugyan léteznek olyan módszerek, amik a nem klasszikus esetben is működnek, de ezek többnyire nem jutnak el a felhasználókhoz. Emellett problémának látják a fokozatszerzési kényszert, ami megköveteli a statisztika alkalmazását is, olyanoktól is, akiknek ebben nincs nagy gyakorlata. Negyedrészt a folyóiratokban megjelenő cikkek bírálati eljárása esetén a helyes és a hibás alkalmazások szétválasztása nem mindig sikerül.

Az ülés következő részében a jelenlévők közül bárki szabadon hozzászólhatott a beszélgetéshez. Pukli Péter megjegyezte, hogy a kritika az alkalmazóknak szól, foglalkozni kell minden jelzéssel, ami az alkalmazóktól jön. Kérdés, hogy a statisztikusoknak hogyan kell erre válaszolni. Véleménye szerint két kérdés merül fel: egyrészt meddig tart a statisztikus és honnantól kezdődik a felhasználó. Másrészt vannak-e itthon statisztikai kutatások, van-e statisztikai kutató. Javasolja, hogy erről készüljön egy hazai áttekintés.

Vargha Péter szerint a következtető statisztika elhagyása nem helyes. A p-érték helyett fontos lenne a mérték megadása. Ugyanakkor az alkalmazók nem tudják megmondani nekik mi a megfelelő mérték.

Katona Tamás szerint jó lenne, ha a vita nem csak statisztikus körben zajlana. Sok helyen bizonyos módszerek alkalmazását követelik meg. Fel kellene hívni az akadémiai testületek figyelmét arra, hogy a túlformalizált elvárások visszasak lehetnek.

Székely Ilona felvetette, hogy a statisztika oktatása során inkább tanítsunk kevesebbet, de azt mélyebben.

Szilágyi Roland javasolta, hogy az állásfoglalás a cikkek olvasóinak is szóljon, amelynek figyelembe vételével tudják mérlegelni, hogy a cikkben olvasottakat hogyan kezelik. Emellett javasolja, hogy a statisztikus közösségnek jobban kellene menedzselnie magát.

Kovács Péter elmondta, hogy őszi tudományos ülésre egy módszertani problémák szekcióval készülünk, ahol ez a téma nem csak statisztikus körben lesz megvitatva.

A napi rendi pont zárásaként egy állásfoglalás készült a tudományos élet szereplőinek a tesztek alkalmazásával kapcsolatosan.

Tudományos cikkek, eredmények közlésekor kívánatos az empiria alkalmazása. Minták alkalmazásakor nem elegendő csupán a leíró statisztikák közlése. A tesztek alkalmazási feltételeinek teljesülése megkövetelendő. Pusztán csak a p-értékek alapján történő merev küszöbértékek alapján a döntéshozatal nem elfogadható, az eredmények gazdagabb (például

konfidenciaintervallumok, hatásnagyság vizsgálatok) bemutatása szükséges. A kutatóktól elvárt, hogy a minták tervezésekor a próbák erejére is fordítsanak figyelmet. A folyóiratokban megjelenő cikkek esetében szerkesztői és lektori felelősség, hogy a cikkben mi jelenik meg.

Az ülés második napirendi pontjaként folytatódott tagtársaink bemutatkozó előadássorozata. Varga Beatrix egyik kedvenc területe a demográfia, így előadásának témája a termékenység alakulásának közgazdasági magyarázata volt.