

# Nonresponse-problémák és kezelésük hazai survey példáján

*Nagy Ilona*

*ELTE-TÁTK Statisztika Tanszék*

*SZTE-BTK Szociológia Tanszék*

## **Bevezetés**

Az empirikus kutatások során alapvető kritérium, hogy az adatok megfelelő minőségben álljanak rendelkezésre az elemzéshez, amely eredményei alapján a következtetéseket levonjuk. A feltétel teljesülése tudományterülettől függetlenül fontos, legyen szó kvantitatív vagy kvalitatív kutatásról.

Amennyiben az empirikus vizsgálat másodelemzésen alapul, az előzetesen előállított adatbázisok, vizsgálati anyagok elemzése során figyelemmel kell lenni a metaadatokra. Vagyis fontos tisztában lenni azzal a konkrét módszertannal, amely alapján az adatok gyűjtése és maga a kiválasztási folyamat zajlott. Amennyiben pedig elsődleges adatgyűjtést tervezünk, úgy fontos, hogy már az előkészítés fázisában is számoljunk azokkal a problémákkal, melyek a felmérés során jelentkezhetnek, és tervezzük meg azok kiküszöbölését, illetve javítási lehetőségeit a felmérés különböző szakaszaiban.

Különösen a kvantitatív kutatások módszertana esetében találunk olyan kidolgozott keretrendszert, mely segítségével a survey-problémák szempontjából jellemezhető egy kutatás. Ezt a módszertani áttekintést érdemes elvégezni akár a tervezés-megvalósítás fázisában, akár az elemzés előtti adatkezelési szakaszban is.

A jelen tanulmányban arra kerestük a választ, hogy a survey-kutatások nemzetközi tapasztalatai alapján az egyik legfontosabb probléma, a csökkenő válaszadási ráta kimutatható-e egy általunk vizsgált hazai survey példáján is. Majd a kétféle *nonresponse*-probléma kezelési módszereinek alkalmazhatóságát hasonlítottuk össze.

## **Survey-problémák**

A kvantitatív módszertanon alapuló kutatások leggyakrabban kérdőíves felmérések formájában biztosítják az adatfelvételt a vizsgálatokhoz. Ezek a survey-kutatások azonban számos problémával találják szembe magukat, melyek

mentén érdemes is megvizsgálni a felméréseket a kutatás tervezésekor, a design kialakításakor, és akkor sem szabad erről az áttekintésről elfeledkezni, amikor másodelemzés céljából egy kész adatbázissal van dolgunk.

A két talán legismertebb survey-probléma az érvényesség és a megbízhatóság kérdése. Érvényességről kétféle értelemben is beszélhetünk. Egyrészt szó lehet külső érvényességről, amikor a mintánk eredményei alapján a sokaságra is levonhatunk következtetéseket. Másrészt lehet szó belső érvényességről, amikor a mérőeszközökkel, a kérdőív kérdéseivel azt a mérni kívánt jelenséget, fogalmat mérjük, amit mérni terveztünk. A megbízhatóság alatt azt értjük, hogy ha közel azonos körülmények között elvégezzük ismételten a felmérést, akkor az eredmény hasonló lesz (Biemer 2010; Groves/Lyberg 2010).

Kevésbé esik szó a módszertani vizsgálatok során a szórás és torzítás problémájáról. A szórás különösen kismintás kutatás esetén jelent problémát, amikor a mérni kívánt változó mentén is nagy a szórás. A torzítás pedig tulajdonképpen a szórás azon esete, amikor egy bizonyos irányba szórnak el szisztematikusan az értékek. Ez a két probléma pedig azt eredményezi, hogy a következtetések esetlegessé válnak, vagy éppen torzítják a valóságot (Groves/Lyberg 2010).

A survey-kutatások egyik sarkalatos kérdése, hogy számolni kell-e a mintavételi hibával, és ha igen, milyen mértékben. Ezt két lépcsőben érdemes vizsgálni, méghozzá abból a szempontból, hogy a tervezett minta illeszkedik-e a sokasághoz, valamint abból, hogy a megvalósult minta illeszkedik-e a tervezett mintához (Meyer/Mittag 2020). Számolnunk kell természetesen a nem mintavételi hibával is, mely a mintavételi eljárástól függetlenül is fellép. Ide sorolható például a kérdezőbiztosok nem megfelelő felkészítéséből eredő hiba (Schreiber/Asner-Self 2011; Meyer/Mittag 2020).

A kérdőíves felmérések problémáinak egy további átfogó keretét adja a TSE, vagyis a *Total Survey Error* módszertani kerete. Ebbe tartozik a *coverage*, vagyis a lefedettség kérdése, amikor abból a szempontból kell az adott felmérést vizsgálni, hogy az adatgyűjtés mennyire találta el a tervezett mintánkat. A mérési vagy válaszhiba akkor lép fel, amikor a válaszadók nem értették, félreértették a kérdést, esetleg nem tudták a válasz, de mégis válaszoltak, vagy éppen szándékosan hazudtak. A mintavétel hibája pedig magából a mintavételi eljárásból fakad (Schreiber/Asner-Self 2011).

A TSE magába foglalja az egyik legizgalmasabb vizsgálati szempontot is, vagyis azt, hogy van-e nemválaszolásból eredő hiba. Azaz fennáll-e a nonresponse, tehát a válassztagadás esete. Tulajdonképpen minden survey

esetén számolni kell azzal, hogy a válaszadási ráta nem lesz 100%-os. Ráadásul a nemválaszolás kétféle is lehet (Smith 2011).

A *unit nonresponse* esetén teljes válaszmegtagadásról beszélünk, amikor a mintába került személy nem hajlandó részt venni a felmérésben. Az *item nonresponse* pedig bizonyos kérdések megválaszolása esetén jelent csak válaszmegtagadást (Meyer/Mittag 2020). Míg a *unit nonresponse* csökkentését leggyakrabban a terepi munka fejlesztésével próbálják megoldani, az *item nonresponse* csökkentésénél viszont már hangsúlyosabb az adatbázis utólagos kezelése (Meyer/Mittag 2020).

### **Unit nonresponse és kezelése**

A nemzetközi közvéleménykutatások eredményei és módszertani tapasztalatai alapján a fokozatosan csökkenő válaszadási ráta jellemző a kérdőíves felmérésekre. Ezek a kutatások azért is alkalmasak a tendenciák vizsgálatára, mert szinte változatlan módszertannal követik egymást az évenkénti, kétévenkénti felmérések. Ugyanakkor nemcsak egy-egy felmérés csökkenő válaszadási rátái figyelhetők meg, az egyes nemzetközi felmérések tendenciája egymást is alátámasztja, hiszen általános a növekvő teljes válaszmegtagadás, vagyis a *unit nonresponse* mértéke (Smith 2010).

A teljes válaszmegtagadásnak több olyan oka van, mely a társadalom megváltozásából következik. Az emberekben az elmúlt évtizedekhez képest fokozatosan nőtt a bizalmatlanság érzése általában véve is. Érthető tehát, ha nem szeretnének kérdezőbiztosokat fogadni otthonukban. Ugyanakkor vannak kulturális-történelmi hagyományokra visszavezethető okok is. Ilyen például, amikor hosszútávon megmarad az a felfogás, hogy a válaszokból kellemetlensége származhat valakinek, így inkább megtagadják a választ. Ami pedig a legszembetűnőbb ok, az az, a felgyorsult élet, így már nincs idő kérdezők fogadására és hosszú kérdőívekre válaszolásra (Lee 2012).

Nemzetközi tapasztalatok alapján viszont megállapítható, hogy a *unit nonresponse* jelentkezésének mértéke csökkenthető a kérdezőbiztosok megfelelő felkészítésével, képzésével a kérdezői oldalról. A válaszadói oldalról viszont megfelelő tájékoztatással és motiválással lehet elérni javulást a válaszadási arányokban. Meg kell jegyezni, hogy a motiválást leginkább a megkérdezettekkel kapcsolatban alkalmazzák, viszont ugyanolyan fontos az is, hogy a kérdezők megfelelően motiváltak legyenek az adatfelvétel során.

### Item nonresponse és kezelése

Ideális esetben, amikor egy kérdés esetén válaszhiány van, de tudjuk, hogy milyen értékkel lehet visszapótolni, akkor egyszerű *data editinggel* van dolgunk. Egyszerűen megoldható az értékek visszapótlása, ha például egy számított értékről van szó, mely alapváltozóinak az adatai rendelkezésre állnak. Akkor is használhatjuk a *data editinget*, ha egy kérdést kicsit más megfogalmazásban többször is feltettünk a kérdőívben, és ekkor a meglévő válasz alapján viszonylag egyértelműen visszapótolható a hiányzó is (Lee et al. 2012).

Bonyolultabb esetben, amikor nem állnak rendelkezésre alapváltozók, vagy nincsenek ellenőrző kérdések, az adatimputálási eljárásokat kell használnunk. Az adatok pótlásának kérdése elemzői döntésen alapul, hogy mekkora mértékű adathiányt kell már kezelni, de talán általánosnak mondható az a nézőpont, hogy 20%-os adathiányt már mindenképpen kezelni kell. Azonban az adatimputálási eljárás kiválasztása előtt meg kell vizsgálni az adathiány jelentkezését a változók kapcsolatának függvényében (Danis 2012).

Az adathiány jelentkezésének három tipikus esetét különböztethetjük meg aszerint, hogy mennyire véletlenszerű a felbukkanása. Az MCAR, vagyis *Missing Completely at Random* esetén az adathiány egy változó mentén teljesen véletlenszerűnek tekinthető más változóktól függetlenül. Ez a típus a legszerencsésebb, mivel torzításmentesen elhagyhatók az adathiányos esetek a mintából. A MAR, vagyis *Missing at Random* esetén az adathiány véletlenszerűnek tekinthető, azonban egy, a változóval kapcsolatban lévő változó értékeitől függ a jelentkezése. A MAR-típust már kezelni kell valamilyen adatimputálási eljárással, azonban a különböző módszerek különböző mértékben ugyan, de jó megoldást jelenthetnek. Az MNAR, vagyis a *Missing not at Random* esetén nem tekinthető véletlenszerűnek az adathiány, a hiányos és a vele kapcsolatban lévő változó értékeitől is függ a megjelenése. Ez a legrosszabb típus, ami előfordulhat, hiszen ennek a kezelésére jelenleg nincs kidolgozott módszertan (Kalton/ Kasprzyk 1986; Shoemaker et al. 2002; Danis 2012; Máder 2012).

### Unit nonresponse a hazai surveyben

A vizsgálat alapja a *Szeged Studies survey* kutatássorozat, melyet a Szegedi Tudományegyetem Szociológia Tanszéke 2000 óta évente lefolytat. Eredményeképpen évente rendelkezésre állnak információk a szegediek életével és a lakosok közszolgáltatásokról alkotott véleményével kapcsolatban. A felmérések során valószínűségi mintavételt alkalmaznak. A központi nyilvántartásból megkért, a szegedi felnőtt lakosság 10000 fős mintájából

szisztematikus mintavétellel egy 1000 fős főcímlistát vesznek, amely a kérdőíves vizsgálat elemzési egységeit jelenti. A felmérés kérdezőbiztosok segítségével történik, melynek során személyes adatfelvételt végeznek. A személyes adatfelvétel előnyeiket használják ki, mivel így sokkal nagyobb a válaszadási hajlandóság a lakosok részéről, elkerülhetők a nem megfelelő értelmezéséből adódó adatvesztési hibák, kiküszöbölhető a környezeti befolyásoló hatás.

Az elmúlt évek felméréseit vizsgálva a szegedi felméréssorozatban is a csökkenő válaszadási ráta mutatható ki, összességében egészen 40%-os válaszadási arányig. Ugyanakkor, mivel rendelkezésre állnak a területi adatok is az adatfelvételezésről, a területi különbségek is vizsgálhatóak. A Belváros és a további belső városi választóközrzetekben a *unit nonresponse* jóval nagyobb arányban jelentkezik, mint a külső választóközrzetek esetén. Előbbiek esetében a válaszadási ráta lefelé az 5%-ot is eléri, utóbbiak esetén felfelé az 50%-ot is meghaladja.

A Szeged Studies felmérések során a teljes válaszmegtagadókat különböző módszerekkel pótolják. A legtöbb évben a pótcímezt alkalmazták, amikor is a tervezett mintából kieső címeteket, vagyis válaszadókat helyettesítették alap demográfiai jellemzőik alapján illesztett válaszadókkal egy pótcímlistából. Ez a módszer természetesen nem csökkenti a válaszmegtagadók arányát, azonban a keletkező adatbázis javítható általa. Ugyanakkor a módszer meglehetősen költséges.

A válaszadási ráta növelésére és a terepi munka könnyítésére próbálták ki a jobbszomszéd-módszert, ami azt jelenti, hogy a válaszadó el nem érése esetén a jobb szomszédját választhatták a kérdezőbiztosok. Ez persze egyfajta szabadságot is biztosít a kérdezőknek, hogy a számukra szimpatikus címeteket kérdezzék csak le. Előzővel szemben viszont költségmegtakarítást tesz lehetővé, és látszólag a válaszadási arányok is javulnak.

Minden év felméréséről elmondható, hogy nagy figyelmet fordítottak a tervezett mintába kerültek megfelelő tájékoztatására, és ezzel is szerették volna motiválni a felkeresetteket a válaszadásra. Ugyanakkor nem fordítottak kellő figyelmet a kérdezők motiválására, valamint a területi különbségeket sem vették figyelembe a terepi munka során végzett ösztönzéseknél.

### **Az item nonresponse jelentkezése a hazai surveyben**

Az elmúlt évek adatbázisainak áttekintése után megállapítottuk, hogy a legnagyobb arányban a háztartásjövdelemmel kapcsolatban jelentkezik adathiány minden évben, azonban a jelen tanulmány terjedelmi korlátai miatt csak

a 2017-es adatok vizsgálatát és kezelését mutatjuk be, melynek eredményei jól jellemzik az adathiányok kezelhetőségét a felmérésorozatra vonatkozóan. Arra a kérdésre, hogy „Önök (a háztartásban élők) egy átlagos hónapban összesen kb. hány forintból tudnak gazdálkodni (mennyit tudnak hazavinni)?” csak 56,1%-ban kaptunk érvényes választ, 43,9%-ban adathiánnyal kell számolunk. Ez tehát mindenképpen olyan alacsony válaszadási arányt jelent, amit már szükséges kezelni. Ezenkívül megvizsgáltunk más jövedelemmel, anyagi helyzettel kapcsolatos kérdéseket is, melyeket szintén feltettek a kérdezőbiztosok a megkérdezetteknek (1. táblázat).

1. táblázat: A kérdőív releváns kérdéseinek nonresponse arányai (forrás: Szeged Studies alapján saját szerkesztés)

A kérdőív kérdése	Item nonresponse aránya, %
A háztartás összjövedelméből összesen hány ember megélhetését kell biztosítani?	4,4
Hányan élnek Önök együtt közös háztartásban?	2,4
Tudnak-e Önök ebből a pénzből valamennyit félretenni, takarékoskodni, valamire gyűjteni?	7,0
Előfordult-e Önökkel az utóbbi 12 hónapban, hogy nem jutott elegendő pénz ennyivalóra?	3,1
Előfordult-e Önökkel az utóbbi 12 hónapban, hogy nem volt elég pénzük, lakbérre, fűtésre, villany- vagy gázzámlára?	3,1
Előfordult-e Önökkel az utóbbi 12 hónapban, hogy pénzhány miatt kevesebbet fűtöttek, vagy csak a lakás egy részét fűtötték?	7,0
Előfordult-e Önökkel az utóbbi 12 hónapban, hogy nem tudták befizetni a hitelük törlesztő részletét?	31,7
Előfordult-e Önökkel az utóbbi 12 hónapban, hogy valaki a családjukban munkanélkülivé vált?	3,6
Voltak-e Önök az elmúlt évben üdülni (nyaralni, telelni)?	1,3
Hogy érzi, a következők közül mi jellemzi leginkább az Önök jövedelmi helyzetét?	2,9

Egyedül a hitelrészlet fizetésével kapcsolatban magas a válaszhiány aránya, de ez abból is következik, hogy sok válaszadó esetén nem volt releváns a

kérdés. Minden más kérdés esetén a nonresponse aránya 10% alatt marad, tehát kezelést nem igényel.

Egy további változóként a vizsgálatba bevontuk az iskolai végzettség négykategóriájú változót, hiszen a jövedelmi helyzet gyakran mutat kapcsolatot ezzel a szociodemográfiai változóval. Egyszempontos varianciaanalízissel teszteltük, hogy szignifikáns kapcsolat van a háztartás összjövedelme és az iskolai végzettség között ( $F=7,978$ ,  $p<0,001$ ). Szintén ANOVA-t alkalmazva szignifikáns összefüggést találtunk az egy főre eső jövedelem és az iskolai végzettség között is ( $F=12,841$ ,  $p<0,001$ ). Ugyanakkor a válaszhiányt megvizsgálva az iskolai végzettség esetén kaptuk a legnagyobb érvényes válaszarányt, csupán 1,1%-os a nonresponse aránya.

Kétmintás t-próba alapján megállapítottuk, hogy a háztartási összjövedelemben szignifikáns különbség mutatható ki a takarékoskodási probléma ( $t=-3,436$ ,  $p=0,001$ ) és a fűtési probléma megléte alapján ( $t=5,26$ ,  $p<0,001$ ). Az egy főre jutó jövedelemben szintén különbség mutatkozik a takarékoskodási lehetőség kapcsán ( $t=-5,317$ ,  $p<0,001$ ), a számlafizetési probléma megléte ( $t=5,653$ ,  $p<0,001$ ), a fűtési probléma jelentkezése ( $t=7,111$ ,  $p<0,001$ ) és a hitel problémák kapcsán ( $t=3,062$ ,  $p=0,002$ ). Az egyszempontos varianciaanalízis pedig kimutatta, hogy szignifikáns különbség van az összjövedelemben a nyaralás tekintetében ( $F=13,403$ ,  $p<0,001$ ) és az anyagi helyzet szubjektív megítélése alapján ( $F=3,058$ ,  $p=0,017$ ). Az egy főre eső jövedelemben szintén kimutatható szignifikáns különbség a nyaralási lehetőségek esetén ( $F=14,392$ ,  $p<0,001$ ) és az anyagi helyzet megítélése szerint ( $F=7,49$ ,  $p<0,001$ ).

A szignifikáns különbségeket okozó független változók szerint megvizsgáltuk, hogy az adathiány véletlenszerűen jelentkezik-e, vagy függ a függő változóktól. Mindegyik esetében azt találtuk, hogy a válaszhiány kisebb vagy nagyobb mértékben függ a független változó értékeitől. Mivel a független változók adathiányai minden esetben véletlenszerűnek tekinthetők, és csak a független változó kategóriái befolyásolják a függő változó adathiányát, az adatbázisban jelentkező nonresponse-típus a MAR-típus. Ez alapján az adathiányt kezelni tudjuk a különböző adatimputálási módszerekkel, és azok hatékonyságát is össze tudjuk hasonlítani.

### **Adatimputálási módszerek**

A háztartási összjövedelem és az egy főre eső jövedelem változók adathiányát egységes design alapján végeztük, mely tulajdonképpen egy adatvisszapótlási

szimulálásnak is tekinthető. Azért csak szimulálással végezhető el az imputálási eljárások összehasonlítása, mert a szimulált modell alapján van egy alaptábla, aminek adataihoz viszonyíthatjuk a visszapótlás utáni eredményeket. Az eredeti felvételezett adatbázis alapján ilyen értelemszerűen nem áll rendelkezésre, hiszen sokan nem adtak választ a kérdésekre.

Az átlaggal való visszapótlás során az ismert változóadatok átlagával pótoljuk vissza az adathiányos esetek értékeit. Jövedelemről lévén szó, bár jelentősen kiugró értékeket az adatbázis nem tartalmazott, célszerű lehet az átlag mellett a mediánnal történő visszapótlás is, így mindkét módszert alkalmaztuk.

A véletlen sorsolás módszerével történő imputálás során random kiválasztott esetek értékeivel történik a visszapótlás. A meghagyott, valid értékeket tartalmazó adatbázisból random 50%-os mintát vettünk, majd erre a mintára kiszámítottuk az iskolai végzettség szerinti átlagokat. A minta mediánját is kiszámítottuk a magyarázó változó értékei szerint.

A legközelebbi szomszéd módszere alapján történő adatimputálás úgy történik, hogy a magyarázó változók kategóriái szerint hasonló eseteket választunk ki. Mivel ez több paraméter alapján történik, ebben a módszerben felhasználtuk az anyagi helyzet szubjektív megítélése változót is. Ekképpen a valid értékeket tartalmazó adatbázis adataiból kategóriákat képeztünk az iskolai végzettség és az anyagi helyzet megítélése alapján. A létrejött csoportokban pedig kiszámítottuk a jövedelemátlagokat és a mediánokat is a magyarázó változók kombinált kategóriái szerint.

A regresszió módszere alapján történő adatimputáláskor a magyarázó változók segítségével regressziós becslést végzünk a függő, adathiányos változó átlagértékére. Ezen esetben kritériumként szerepel, hogy a lineáris regresszió előfeltételei teljesüljenek. A megfelelő magyarázóváltozók ennél a módszernél az intervallumskálának tekintett anyagi helyzet szubjektív megítélése változó, illetve az egyes anyagi problémák felmerüléseinek számából képzett arányskála mérési szintű változó.

A többszörös imputálás során pedig egy iteratív folyamattal végeztük el az adatvisszapótlást. Ehhez a már fent ismertetett véletlen esetillesztést alkalmaztuk. A visszapótlást 10 körben végeztük, így tíz pótoltt adatbázis keletkezett, majd ezen adatbázis jövedelem értékeit átlagoltuk ki, és ezzel a végső értékkel végeztük el a tényleges adatimputálást.

Az eljárásokat összehasonlítva a szimuláció alaptáblájával valamint egymással, azt tapasztaltuk, hogy ahogy haladunk a felsorolás szerint a egyszerűbb eljárásoktól a bonyolultabbak felé, egyre kevésbé eredményeznek



torzított és megbízhatatlan becsléseket. Ez pedig azért fontos, mert ha a mintából következtetéseket vonunk le a sokaságra, azt a lehető legtorzításmentesebben és legmegbízhatóbban kell, hogy tegyük.

### Összegzés

A hazai survey-kutatás vizsgálata alapján megállapítható, hogy a válaszmegtagadás növekvő mértéke, ami a nemzetközi felmérésekben már trendként azonosítható folyamat, a szegedi kérdőíves kutatássorozatban is kimutatható. Ezt a folyamatot viszont kezelni, vagy legalább lassítani kell annak érdekében, hogy az előállított adatok megfelelő minőségűek legyenek.

Bár a *unit nonresponse* összességében hasonló mértékű, mint a nemzetközi felmérésekben, azonban a teljes válaszmegtagadás arányában jelentős területi különbségek mutatkoznak a belső városrészek hátrányára. Ezt a különbséget kihasználva érdemes a térbeli eltérésekre specializálni a terepi munka során a válaszadás ösztönzését, mellyel a tényleges válaszadási arányok várhatóan javulnak. Ugyanakkor fontos szempont a *unit nonresponse*-esetek csökkentésére a kérdezőbiztosok megfelelő motiváltságának figyelembevétele is a terepi munka megvalósítása során.

Az *item nonresponse*-esetek legnagyobb arányban a jövedelemmel kapcsolatos kérdések során adódtak, azonban az adathiányok jelentkezése a vizsgált változókapcsolatok alapján a MAR-típusba sorolható, ekképpen pedig kezelhetők az adatimputálási eljárásokkal. Az imputálási eljárások összehasonlításával igazolódott, hogy az egyszerűbb matematikai statisztikai háttérű módszerek kevésbé hatékonyak a bonyolultabbakkal szemben, hiszen utóbbiak alkalmazásából torzításmentesebb és megbízhatóbb becslések származnak a sokaságra vonatkozóan.

### Irodalomjegyzék

- Biemer, P.*: Total Survey Error: Design, Implamantation, and Evaluation. Public Opinion Quarterly, vol. 74., issue 5., 2010, 817–848.
- Danis I.*: Az adathelyettesítés modern technikája – „multiple imputation (MI)”. Alkalmazott Pszichológia, 10. évf., 2. sz., 2012, 65–70.
- Groves, R./Lyberg, L.*: Total Survey Error: Past, Present, and Future. Public Opinion Quarterly, vol. 74., issue 5., 2010, 849–879.
- Kalton, G./Kasprzyk, D.*: The treatment of missing survey data. Survey Methodology, issue. 12., 1986, 1–16.
- Lee, G./Benoit-Bryan, J./Johnson, T.*: Survey Research in Public Administration: Assessing Mainstream Journals with a Total Survey Error Framework. Public Administration Review, vol. 72., issue 1., 2012, 87–97.

- Máder M. P.*: Adathiány kezelési eljárások és imputálások összehasonlítása. ELTE-TÁTK, Szociológia Doktori Iskola, Budapest, 2012.
- Meyer, B./Mittag, N.*: An empirical total survey error decomposition using data combination. *Journal of Econometrics*, vol. 219, issue 2., 2020, 102–113.
- Schreiber, J./Asner-Self, K.*: Education research: interrelationship of questions sampling design, and analysis. Hoboken NJ Willey, 2011.
- Shoemaker, P./Eichholz, M./Skewes, E.*: Item nonresponse: Distinguishing between don't Know and Refuse. *International Journal of Public Opinion Research*, vol. 14., issue 2., 2002, 193–201.
- Smith, T.*: Refining the Total Survey Error Perspective. *International Journal of Public Opinion Research*, vol. 23., issue 4., 2011, 464–484.