
6. AZ ADATBÁZISOK FELÉPÍTÉSE

MURINKÓ LÍVIA

Az *Életünk fordulópontjai (ÉF)* adatgyűjtési program keretében eddig négy adatfelvételi hullámra került sor. A panelvizsgálat négy hullámából létrehozható egy akár tizenegy évet átfogó longitudinális adatbázis, de az egyes adatfelvételi hullámok önállóan, keresztmetszeti szemléletben is elemezhetőek. Az adatgyűjtéshez két kiegészítő minta (a harmadik hullám gyermekmintája és a negyedik hullám kiegészítő fiatal mintája) csatlakozott, és lekérdezésre került két speciális blokk is (a második hullám élettörténeti kalendáriuma és a fogamzásgátlási szokásokról szóló önkitöltős kérdőív a harmadik hullámban). (Részletesen lásd a 2. fejezet adatfelvételt bemutató részét.) Az adatgyűjtési program összetettsége már sejteti, hogy a létrejött adatbázisok is igen összetettek és nagyméretűek (összességében több ezer változót tartalmaznak).

A következőkben ismertetjük az adatrendszer felépítését, a változónevek logikáját, a képzett változókat és az adatbázisok tematikus áttekintését. A fejezet segítséget kíván nyújtani a felhasználóknak a kutatási kérdésekhez legmegfelelőbb adatbázis kiválasztásában, a változók közötti eligazodásban és az elemzések elkészítésében.

AZ ADATBÁZISOK

Az *Életünk fordulópontjai* adatrendszeréből a következő adatbázisok érhetőek el a Népeségtudományi Kutatóintézet adatbankján keresztül (lásd az adatokhoz való hozzáférésről szóló 8. fejezetet és az *1. táblázatot*).

- 1. hullám keresztmetszeti adatbázisa;
- 4. hullám keresztmetszeti adatbázisa (az eredeti panelminta és a kiegészítő fiatal al minta válaszadóit egyaránt tartalmazza);
- 1-4. hullám longitudinális adatbázisa (a mind a négy kérdezési hullámban válaszolókat tartalmazó összefűzött adatbázis, amely a 2001 és 2012 közötti időszakot fogja át, és természetesen csak az eredeti panelminta tagjaiból áll).

1. táblázat: Az *Életünk fordulópontjai* adatbankon keresztül elérhető adatbázisai

Az adatbázis neve	Az adatfelvétel éve	Esetszám	A válaszadók életkora
1. Az 1. hullám keresztmetszeti adatbázisa	2001/2002	16 363	18–75
2. A 4. hullám keresztmetszeti adatbázisa (az eredeti panelminta és a kiegészítő fiatal minta egyesített adatbázisa)	2012/2013	12 437	18–86
3. Az 1–4. hullám longitudinális adatbázisa	2001–2013	8 103	18–75 (2001-ben)

Elsősorban ezt a három adatbázist ajánljuk a felhasználók számára, mert rendkívül gazdag és széleskörű elemzési lehetőségeket kínálnak, és képzett változók is segítik a használatukat. Az érdeklődő kutatók részére a többi adatbázist egyedi megállapodás alapján, külön kérésre adjuk ki. Ezek a következők:

- a 2. hullám keresztmetszeti adatai (n = 13 540),
- a 2. hullám élettörténeti kalendáriuma (n = 13 540),
- a 3. hullám keresztmetszeti adatai (amelyek tartalmazzák a fogamzásgátlási szokásokról szóló, kényes kérdéseket tartalmazó önkitöltős blokkot) (n = 10 641),
- a 3. hullám kiegészítő gyermekmintája (n = 1 126).

Az *Életünk fordulópontjai* adatrendszer tartalmilag széles skálát ölel át. Egyes témakörök és kérdések csak egy-két hullámban fordultak elő, míg mások végigkövették a kérdezés tizenegy évét. Az adatbázis tényleges struktúrája nem feltétlenül követi a témakörök logikáját, így nehézséget jelenthet egy-egy adott témakör vagy kérdés megtalálása. A Mellékletben található tartalmi összefoglaló táblázat célja, hogy áttekintse az egyes adatfelvételek témaköreit, és segítse a hullámok közötti összehasonlító elemzéseket.

A LONGITUDINÁLIS ADATOK JAVÍTÁSA

A longitudinális adatbázis összeállítása során alapvetően arra törekedtünk, hogy az eredeti változókat abban az esetben is megőrizzük, amikor az egyes kérdezési hullámokban adott válaszok összehasonlítása inkonzisztens, logikátlan vagy valószínűtlen élettörténeteket tárt fel. A „valóság” ismerete nélkül gyakran nem dönthető el, valójában mi történt, így a kutatókra bízunk, hogy a saját kutatási kérdésük alapján döntsenek ezen esetek sorsáról. Néhány kulcsváltozó esetén azonban mégis szükségessé vált az utólagos adatjavítás, amely azonban egyetlen esetben sem nem jelentette a hiányzó adatok pótlását (logikai alapon történő imputálás stb.). A nyilvánosan elérhető ÉF-adatbázisok már a javított adatokat tartalmazzák.

Az első komolyabb javításra értelemszerűen a második adatfelvételi hullámot követően kerülhetett sor. Ekkor a válaszadók korára, nemére és családi állapotára vonatkozó inkonzisztenciákat próbáltuk kezelni. Ez a javítás 1474 válaszadót érintett, de az esetek nagy részében ez csak kisebb pontosítást jelentett (rögzítési hibák, elért hónapok javítása stb.).

A következő jelentősebb adatjavításra a negyedik hullámot követően került sor. Ezek közül is kiemelt jelentőségű a válaszadók gyermekszámának „rendbetétele”. Ugyanis a négy hullám adatait egybevetve fény derült arra, hogy sok esetben nem konzekvens a gyermekek számbavétele, vagyis ugyanazon válaszadó az egyes adatfelvételi hullámokban eltérő számú, eltérő születési idejű gyermekekről számolt be. Az egyik legszembetűnőbb dolog az volt, hogy az első három hullámban nagyon alacsony volt az elhunyt gyermekek száma, majd a negyedik hullámban hirtelen megugrott. A másik pedig, hogy sok válaszadónak csökkent a gyermekszáma: egy későbbi hullámban kevesebb gyermekről számolt be, mint egy korábbiiban. Emiatt a negyedik hullám lekérdezése után mintegy 1640 válaszadó adataiban történt a gyermekekkel kapcsolatos javítás (részletesen lásd Makay 2016).

A VÁLTOZÓK

A VÁLTOZÓNEVEK LOGIKÁJA

Az adatbázisokban hét különböző típusú változó fordul elő. A változók elnevezése utal a változó típusára, a kérdezési hullámra és esetenként az almintára, ezzel is segítve az adatbázisban való tájékozódást (lásd a 2. táblázatot). A változótípusok az adatbázisokban a megadott sorrend szerint követik egymást.

A változónevek első karaktere egy betű, amely a változó típusára utal, a második karaktere pedig általában egy, a kérdezési hullámot jelölő szám. A szám értéke 1 és 4 között lehet. Szám hiányában a változó az összes kérdezési hullámban azonos (pl. az egyéni azonosító vagy a válaszadó neve) vagy nem vonatkozik konkrét kérdésre (pl. melyik mintába vagy almintába tartozik a válaszadó). Ha ugyanazt a kérdést több hullámban is azonos formában tették fel, akkor a változók neve is jelzi az azonosságot, és csak a kérdezési hullámra utaló szám tér el.

Az a betűvel kezdődő ún. azonosító változók a kérdésre vonatkozó információkat tartalmazzák. Ide tartozik a válaszadók egyéni azonosítója (*aid*), amely a követett válaszadók esetében hullámról hullámra megegyezik. Az *a1hullam*, *a2hullam* stb. elnevezésű változók azt mutatják meg, hogy a válaszadó válaszolt-e az adott kérdezési hullámban. Az *aminta* változó jelzi, hogy a válaszadó az eredeti panelminta, a kiegészítő gyermekminta (3. hullám) vagy a fiatalok kiegészítő mintájának (4. hullám) a tagja-e. Az *a4web* változó azonosítja azokat a válaszadókat, akik a negyedik hullám során a kérdőív egy részét online töltötték ki. A kérdés évét és hónapját az *a1kerdev*, *a1kerdho* stb. változók tartalmazzák.¹

¹ Az 1. hullám esetén a kérdéses 2001. decemberi eszmei időponttal történt, ezért az *a1kerdev* és az *a1kerdho* változók nem a kérdéses tényleges dátumát tartalmazzák.

Az *s* betűvel kezdődő súlyváltozók tartalmazzák az adatbázisok súlyait – külön-külön az egyes keresztmetszeti adatbázisokhoz és a követéses paneladatokhoz. A súlyváltozók kialakításáról és használatáról a 8. fejezetben tájékozódhat.

A területi változók a lekérdezés helyszínének (a válaszadó lakhelyének) területi jellemzőit tartalmazzák (*t.*): megye, régió, a település jogállása és lakosság száma. A megye vagy a régió ismerete lehetővé teszi a makroszintű adatok adatbázishoz való kapcsolását és a többszintű elemzések elkészítését.

A változók első karaktere a legtöbb esetben *k*, amely a tőzskérdőívek eredeti változóit jelöli. Ezek a változók egy az egyben megfeleltethetők a kérdőívekben található kérdéseknek vagy itemeknek, és változtatás nélkül kerülnek be az adatbázisba. A konkrét változók megtalálását segíti, hogy az adatbázisban a változó címkék nemcsak a kérdés (szükség esetén egyszerűsített) szövegét, hanem a kérdésnek a kérdőívben található sorszámát is tartalmazzák. A szó szerint idézett kérdőív kérdések idézőjelek között szerepelnek. A nem idézőjeles címkék egyszerűsített formában utalnak a kérdés tartalmára. A válaszlehetőségek felsorolása során is törekedtünk arra, hogy kövessük a kérdőívek eredeti megfogalmazását. Külön technikai kód jelöli a „nem tudja” (9, 99, 999, vagy 9999) és a „nem kíván válaszolni” (8, 88, 888 vagy 8888) válaszlehetőségeket.

Az adatfelvétel során lekérdezett kiegészítő minták és speciális kérdőívek változóinak első karaktere eltér. A 2. hullám élettörténeti kalendáriumának változó *c* betűvel, a negyedik hullámban csak a fiatal almintára vonatkozó kérdések pedig *kk* betűvel kezdődnek.

Az utólag kialakított képzett változókat (*x.*) a fejezet következő részében ismertetjük. Ezekre az adatokra a kérdőívek nem kérdeztek rá direkt módon, de a törzsváltozókból előállíthatók, és az elemzéshez elengedhetetlenek (pl. gyermekszám, háztartási jövedelem).

2. táblázat: A különböző típusú változók elnevezésének logikája az ÉF adatbázisaiban

Változó típusa	Változónevek
1. Azonosító változók, a kérdésre vonatkozó információk (1–4. hullám)	a1..., a2..., a3..., a4..., a...
2. Súlyváltozók (1–4. hullám)	s1..., s2..., s3..., s4...
3. Területi változók (1–4. hullám)	t1..., t2..., t3..., t4...
4. Tőzskérdőívek eredeti változó (1–4. hullám)	k1..., k2..., k3..., k4...
5. Élettörténeti kalendárium változó (2. hullám)	c...
6. A csak a fiatal almintának feltett kérdések (4. hullám)	kk...
7. Képzett változók (1–4. hullám)	x1..., x2..., x3..., x4..., x...

A KÉPZETT VÁLTOZÓK

Az adatbázis minden kérdezési hullámához ún. képzett változók is kapcsolódnak, amelyek megkönnyítik az adatok elemzését.

A képzett változóknak két típusát különböztetjük meg:

1. A minden egyes hullámban elérhető információból képzett változók: olyan adaton alapulnak, amelyet minden hullámban megkérdeztünk, mert gyakran változnak; ilyen például a gyermekszám vagy a jövedelem (3. táblázat).
2. Az egy-egy hullámban elérhető információból képzett változók: olyan adaton alapulnak, amelyre nem kérdeztünk rá minden egyes kérdés során, például a születési hely vagy a szülők iskolai végzettsége. Ezeket a változókat azon kérdezési hullámok adataihoz is hozzákapcsoltuk, amelyekben nem kérdeztük le őket, mert a múltra vonatkoznak (pl. a gyermekkori család jellemzői, első párcapcsolat), vagy mert feltételezhető, hogy ritkán változnak (pl. nemzetiség) (4. táblázat). Ezekben az esetekben jelezzük, hogy a képzett változó egy korábbi kérdezési hullámból származó információn alapul.

3. táblázat: A mind a négy hullámban elérhető információból képzett változók

Változó leírása	Változónév	Válaszkategóriák	Megjegyzés
<i>A) Alapadatok</i>			
1. Életkor	x1kor, x2kor, x3kor, x4kor	–	A kérdés időpontjáig betöltött életévek száma
2. Korcsoportok (10 és 5 éves)	x1korcsoport, x1korcsoport2, x2korcsoport, x2korcsoport2, x3korcsoport, x3korcsoport2, x4korcsoport, x4korcsoport2	x-29, 30-39, 40-49, 50-59, 60-69, 70-x; x-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-x	A minta „öregedése” miatt a korcsoportok legalsó és legfelső kategóriái hullámonként eltérnek
3. Szülővel él-e	x1szulovel, x2szulovel, x3szulovel, x4szulovel	igen, nem	Közös háztartásban él-e egyik vagy mindkét vér szerinti szülővel
4. Iskolai végzettség	x1isk, x2isk, x3isk, x4isk	kevesebb mint nyolc osztály, nyolc általános, szakmunkásképző, érettségi, felsőfokú diploma	Legmagasabb befejezett iskolai végzettség

Változó leírása	Változónév	Válaszkategóriák	Megjegyzés
<i>B) Gyermek</i>			
5. Élveszületett gyermekek száma	x1gyszam, x2gyszam, x3gyszam, x4gyszam	-	Vér szerinti gyermekek
6. A válaszadóval közös háztartásban élő gyermekek száma	x1eelogysz, x2eelogysz, x3eelogysz, x4eelogysz	-	Vér szerinti gyermekek
7. A válaszadó külön élő gyermekeinek száma	x1kelogysz, x2kelogysz, x3kelogysz, x4kelogysz	-	Vér szerinti gyermekek
8. Meghalt gyermekek száma	x1meghgysz, x2meghgysz, x3meghgysz, x4meghgysz	-	Vér szerinti gyermekek
9. Legidősebb és legfiatalabb gyermek betöltött életkora	x1legfigyk, x2legfigyk, x3legfigyk, x4legfigyk; x1legidgyk, x2legidgyk, x3legidgyk, x4legidgyk	-	Vér szerinti gyermekek
10. Két hullám között született-e gyermeke	x2szul, x3szul, x4szul	igen, nem	
<i>C) Párkapcsolatok, házas- vagy élettárs jellemzői</i>			
11. Tényleges partnerkapcsolati helyzet	x1partner, x2partner, x3partner, x4partner	háztartásával él, élettársával él, külön élő partnere van, nincs partnere	
12. Elvált-e valaha	x1valas, x2valas, x3valas, x4valas	igen, nem	
13. Partner iskolai végzettsége	x1pisk, x2pisk, x3pisk, x4pisk	kevesebb mint nyolc osztály, nyolc általános, szakmunkásképző, érettségi, felsőfokú diploma	Legmagasabb befejezett iskolai végzettség
14. Partner munkaerő-piaci helyzete	x1pmunka, x2pmunka, x3pmunka, x4pmunka	alkalmazott, önálló, öregségi/ özvegyi nyugdíjas, rokkantnyugdíjas, gyese/gyed/gyet-en van, munkanélküli, tanuló, sorkaton, háztartásbeli, egyéb inaktív	
15. Partner foglalkozásának jellege (Andorka-séma)	x1pfogl, x2pfogl, x3pfogl, x4pfogl	önálló gazdálkodó, iparos, kereskedő, szolgáltató, szellemi szabadfoglalkozású; közvetlen termelésirányító, szakmunkás, betanított munkás, segédmunkás; felsővezető, középvezető, alsóvezető, beosztott szellemi; alkalmi szellemi, alkalmi fizikai	

Változó leírása	Változónév	Válaszkategóriák	Megjegyzés
<i>Munka, jövedelem, anyagi helyzet</i>			
16. Munkaerő-piaci helyzet	x1munka, x2munka, x3munka, x4munka	alkalmazott, önálló, öregségi/ özüvegyi nyugdíjas, rokkantnyugdíjas, gyes/ gyed/gyet-en van, munkanélküli, tanuló, sorkatona, háztartásbeli, egyéb inaktív	
17. Foglalkozás jellege (Andorka-séma)	x1fogl, x2fogl, x3fogl, x4fogl	önálló gazdálkodó, iparos, kereskedő, szolgáltató, szellemi szabadfoglalkozású; közvetlen termelésirányító, szakmunkás, betanított munkás, segédmunkás; felsővezető, középvezető, alsóvezető, beosztott szellemi; alkalmi szellemi, alkalmi fizikai	
18. Jövedelmi változók: nettó személyes és háztartási jövedelem, házas- vagy élettárs jövedelme, ekvivalens háztartási jövedelem; személyi, háztartási és ekvivalens jövedelmi ötödök,	x1jov, x2jov, x3jov, x4jov; x1hjov, x2hjov, x3hjov, x4hjov; x2pjov, x3pjov, x4pjov; x1ekvj, x2ekvj, x3ekvj, x4ekvj; x1jov5, x2jov5, x3jov5, x4jov5; x1hjov5, x2hjov5, x3hjov5, x4hjov5; x1ekvj5, x2ekvj5, x3ekvj5, x4ekvj5	-	Kategoriális válasz esetén az osztásközök értéke szerepel; az ekvivalens háztartási jövedelem számítása 0,5-ös szorzóval történt
19. Szegénység	x1szegeny, x2szegeny, x3szegeny, x4szegeny	szegény, nem szegény	Ekvivalens háztartási jövedelem alapján a mediánjövedelem 60%-a alatt élők

4. táblázat: Az egy-egy hullámban elérhető információkból képzett változók

Változó leírása	Változónév	Válaszkategóriák	Megjegyzés
<i>A) Gyermekkori családi háttér</i>			
1. Édesanyja és édesapja iskolai végzettsége	xanyisk, xapisk	kevesebb mint nyolc osztály, nyolc általános, szakmunkásképző, érettségi, felsőfokú diploma	A 2. (a fiatal almintánál a 4.) hullámból származó adat
2. Édesanyja és édesapja halálának éve	xanyhal, xaphal	-	Ha már nincs(enek) életben; az 1. és a 4. hullámból származó adat
3. Szülők párkapcsolati helyzete a válaszadó születésekor	xszulkapcs	házasságban éltek, élettársi kapcsolatban éltek, nem éltek együtt	Az 1. (a fiatal almintánál a 4.) hullámból származó adat
4. Életkor a szülők (szét-) válásakor	xszulval	-	Ha a válaszadó 14 éves koráig megtörtént; az 1. (a fiatal almintánál a 4.) hullámból származó adat
<i>B) Születési hely, nemzetiség</i>			
5. Születési hely területi változói (ország, régió, megye, település neve és kódja, településtípus, lakosságszám)	xszország, xszregio, xszmegye, xsztelnev, xsztelkod, xszteljog, xsztelmer	-	Édesanyja lakhelye a válaszadó születésekor (nem a kórház szerinti település); a születés évében jellemző településtípus; a születési évhez legközelebb eső népszámlálás évében mért lakosságszám; az időközben átnevezett vagy összevont települések esetében a jelenlegi településnév és az összevonás utáni településre vonatkozó információk szerepelnek.
6. Mióta él Magyarországon (ha külföldön született)	xittelv, xittelho	-	
7. Nemzetiség	xnemzet	bolgár, cigány, cigány származású magyar, görög, horvát, lengyel, német, német származású magyar, örmény, román, ruszin, szerb, szlovák, szlovák származású magyar, szlovén, ukrán, magyar, egyéb, nincs válasz	Az 1. (a gyermekmin-tánál a 3., a fiatal almintánál a 4.) hullámból származó adat

Változó leírása	Változónév	Válaszkategóriák	Megjegyzés
<i>C) Élettörténeti változók</i>			
8. Első élettársi kapcsolat/ házasság kezdete	xpkezdev, xpkezdho	-	Ha az élettársi együttélés legalább 3 hónapig tartott
9. Első házasságkötés dátuma	xhazasev, xhazasho	-	
10. Első gyermek születésének dátuma	xgyerekev, xgyerekho	-	
11. Rendszeres pénzkereső munka kezdete	xmkezdev, xmkezdho	-	Az 1., a 2., a fiatal al minta esetén a 4. hullámból származó adat

HIVATKOZÁSOK

Makay Zsuzsanna (2016): Önnek hány gyermeke született? Gyermekszámra vonatkozó kérdésési technikák és adatjavítások az *Életünk fordulópontjai* panelkutatásban. *Demográfia*, 59(1), 121–143.