

**Minőségi irányelvek a Központi Statisztikai Hivatal  
statisztikai munkafolyamatainak egyes  
szakaszaira**

# Tartalomjegyzék

A dokumentumban használt rövidítések jegyzéke .....	4
Regiszterek kiválasztása .....	5
1. Leírás: .....	5
2. Alapelvek: .....	5
3. Minőségi irányelvek: .....	5
Felvételi keret meghatározása.....	7
1. Leírás: .....	7
2. Alapelvek: .....	7
3. Minőségi irányelvek: .....	7
Célok, felhasználás és felhasználók meghatározása .....	9
1. Leírás: .....	9
2. Alapelvek: .....	9
3. Minőségi irányelvek: .....	9
Fogalmak, definíciók és osztályozások meghatározása.....	11
1. Leírás: .....	11
2. Alapelvek: .....	11
3. Minőségi irányelvek: .....	11
Felhasználható igazgatási adatok számbavétele, statisztikai hasznosításuk .....	15
1. Leírás: .....	15
2. Alapelvek: .....	15
3. Minőségi irányelvek: .....	16
A mintavételi terv kialakítása .....	19
1. Leírás: .....	19
2. Alapelvek: .....	19
3. Minőségi irányelvek: .....	19
Kérdőív és segédanyagainak tervezése .....	22
1. Leírás: .....	22
2. Alapelvek: .....	22
3. Minőségi irányelvek: .....	22
Adatgyűjtés szervezése és adatgyűjtés.....	25
1. Leírás: .....	25
2. Alapelvek: .....	25
3. Minőségi irányelvek: .....	25
Adatok előkészítése (rögzítés, editálás) .....	28
1. Leírás: .....	28
2. Alapelvek: .....	28
3. Minőségi irányelvek: .....	28
Imputálás (pótlás) .....	31
1. Leírás: .....	31
2. Alapelvek: .....	31
3. Minőségi irányelvek: .....	32
Súlyozás, becslés és a mintavételi hiba számítása.....	35
1. Leírás: .....	35
2. Alapelvek: .....	35

3. Minőségi irányelvek: .....	35
<b>Indexszámok képzése .....</b>	<b>37</b>
1. Leírás: .....	37
2. Alapelvek: .....	37
3. Minőségi irányelvek: .....	37
<b>Makrovalidálás.....</b>	<b>39</b>
1. Leírás: .....	39
2. Alapelvek: .....	39
3. Minőségi irányelvek: .....	39
<b>A szezonális kiigazítás.....</b>	<b>41</b>
1. Leírás: .....	41
2. Alapelvek: .....	41
3. Minőségi irányelvek: .....	41
Irodalom: .....	42
<b>Elemzések készítése.....</b>	<b>44</b>
1. Leírás: .....	44
2. Alapelvek: .....	44
3. Minőségi irányelvek: .....	44
<b>Az adatok bizalmas kezelése és a felfedhetőség elleni védelem .....</b>	<b>46</b>
1. Leírás: .....	46
2. Alapelvek: .....	46
3. Minőségi irányelvek: .....	46
<b>Tájékoztatás.....</b>	<b>49</b>
1. Leírás: .....	49
2. Alapelvek: .....	49
3. Minőségi irányelvek: .....	49
<b>Archiválás .....</b>	<b>52</b>
1. Leírás: .....	52
2. Alapelvek: .....	52
3. Minőségi irányelvek: .....	52
<b>Értékelés, felülvizsgálat és visszacsatolás.....</b>	<b>54</b>
1. Leírás: .....	54
2. Alapelvek: .....	55
3. Minőségi irányelvek: .....	55
<b>További tanácsok .....</b>	<b>57</b>

### A dokumentumban használt rövidítések jegyzéke

<i>Rövidítés</i>	<i>Magyarázat</i>
CAPI	Computer Assisted Personal Interview – Számítógéppel támogatott személyes kérdezés
CASI	Computer Assisted Self Interviewing – Elektronikus önkitöltős adatgyűjtés
CATI	Computer Assisted Telephone Interviewing – Számítógéppel támogatott telefonos adatgyűjtés
GÉSA	Gazdálkodó szervezetek és statisztikai adatszolgáltatásaik – Egységes adatgyűjtés-szervezési rendszer
GSZR	Gazdasági szervezetek regisztere
SFT	Stratégiai Fejlesztési Tanács
TRAMO/SEATS	Time series Regression with ARIMA noise, Missing values and Outliers/ Signal Extraction in ARIMA Time Series (Szezonális kiigazítási eljárás)

# Regiszterek kiválasztása

## 1. Leírás:

A regiszter valamely sokaság egyedeinek azonosítására, elérhetőségére és az adatfelvételi folyamatot támogató jellemzőkre vonatkozó rendszeresen karbantartott nyilvántartás. A regiszterből elérhetők a sokaság aktuális és történeti állapotai, a változások oka, hatálya, forrása. A regiszter a sokaság egyedeire (egységeire) vonatkozó adatokat strukturáltan, adatbázisban tárolja, kezeli.

A KSH-ban vállalkozások és intézmények felvételére a Gazdasági Szervezetek Regisztere (GSZR) használatos. A kiskereskedelmi üzlethálózat regisztere (KISREG), a kereskedelmi szálláshelyek nyilvántartása (KERSZA), a Non-profit Szervezetek Regisztere (NPREG), a Mezőgazdasági tevékenységet végző gazdaságok (gazdasági szervezetek és egyéni gazdaságok) regisztere (MEZŐREG), az Európai Unión belül külkereskedelmi tevékenységet végző szervezetek regisztere (INTRAREG), a Szociális szolgáltató intézmények regisztere (SZOCCIM) speciális, az adott területre vonatkozó statisztikákhoz használatos, de GSZR-rel való kapcsolatuk vagy megoldott, vagy a közeljövő fejlesztéseibe tartozik.

A KSH-ban vezetett Településregiszter Magyarország településeit és azok jellemzőit tartja nyilván. A KSH-ban a lakossági felvételek korábban a népszámlálási címállományt használták felvételi keretként, 2009-ben megkezdődött az áttérés a címregiszter használatára. Ritkább esetben a Közigazgatási és Elektronikus Közszolgáltatások Központi Hivatala által vezetett népesség-nyilvántartást használják.

## 2. Alapelvek:

Regisztereket kell létrehozni és működtetni az azonos, vagy egymással kapcsolatban álló célsokaságú felvételek koordinálása, standard osztályozások alkalmazása, a karbantartás hatékony megvalósítása érdekében.

## 3. Minőségi irányelvek:

- Az adatfelvétel tervezése során **vegyük számba**, és **értékeljük** a szóba jöhető **regiszterek minőségét, alkalmasságát**.
- A regiszterek kialakításánál a regiszterek közötti **átjárhatóságot a lehető legnagyobb mértékben** biztosítani kell.
- A regiszterek megfelelő színvonalának biztosítása érdekében alkalmazzunk **standard eljárásokat** a változások átvezetésére (egységek születése, halála/megszűnése, körből kikerülő egységek, jellemzők változása). Biztosítani kell a változások okának, hatályának, forrásának tárolását, valamint a történeti adatok visszavezethetőségét.
- A regiszter karbantartásánál a lehető legtöbb rendelkezésre álló **adminisztratív adatot** használjuk fel. Emellett a **statisztikai adatgyűjtésekből** nyert információk **visszacsatolását**, esetlegesen kereskedelmi forrásokat is használjuk fel a regiszterek minőségének javítására.
- A regiszter **aktualizálását** a lehető leggyakrabban végezzük el, különös tekintettel a felhasználók (keretek) igényeire.
- A regiszterben szereplő egységeknek lehetőleg teljes mértékben le kell fedniük a sokaságot, valamint a sokasághoz nem tartozó egységek nem szerepelhetnek benne. Rendszeresen végezzük el a regiszter **lefedettség ellenőrzését**.

- A regiszter ellenőrzése során a **többszörös előfordulásokat ki kell szűrni**, biztosítani kell a rekordonkénti **teljességet**.

## **Irodalom:**

### **Egyes regiszterekről:**

REGULATION (EC) No. 177/2008 of the European Parliament and of the Council of 20 February 2008 establishing a common framework for business registers for statistical purposes and repealing Council Regulation (EEC) No. 2186/93

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:061:0006:0016:EN:PDF>

Az Európai Parlament és a Tanács 177/2008/EK rendelete (2008. február 20.) a vállalkozások statisztikai célú nyilvántartása közös keretének létrehozásáról és a 2186/93/EGK tanácsi rendelet hatályon kívül helyezéséről

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:061:0006:0016:HU:PDF>

LAKATOS MIKLÓS: Stratégiai főirány S103: A statisztikai nómenklatúrák minőségfejlesztése. Stratégiai terv 2005. május (KSHáló, KSH-Stratégia)

1993. évi XLVI. törvény a statisztikáról

[http://net.jogtar.hu/jr/gen/hjegy\\_doc.cgi?docid=99300046.TV](http://net.jogtar.hu/jr/gen/hjegy_doc.cgi?docid=99300046.TV)

GSZR dokumentáció (Informatikai doktár – 3.1.1), kiemelten:

A Gazdasági Szervezetek Regiszterének általános ismertetése 2007 (3.1.1.1)

APEH-KSH adatkapcsolati rendszer 2009 (3.1.1.2)

Gazdasági Szervezetek Regisztere, rendszerterv 2009 (3.1.1.2)

Vállalatcsoportok és összetett gazdasági egységek a Gazdasági Szervezetek Regiszterében 2005 (3.1.1.2)

GSZR – Felhasználó kézikönyv 2.0 2009 (3.1.1.3)

Egyablakos rendszer 2003 (3.1.1.5)

TELEPÜLÉSREGISZTER dokumentáció (Informatikai doktár – 3.1.2), kiemelten:

A településregiszter (F009) általános leírása 2006 (3.1.2.1)

A településregiszter (F009) szerkezete 2009 (3.1.2.2)

CÍMREGISZTER dokumentáció (Informatikai doktár – 3.1.3), kiemelten:

Címregiszter, magyarországi címek statisztikai regisztere – Felhasználó kézikönyv 2008 (3.1.3.3)

NYITRAI FERENCNÉ (1996): Gazdaságstatisztika. KSH, Budapest.

## Felvételi keret meghatározása

### 1. Leírás:

Az alapsokaság azon egységek összessége, melyek a vizsgálat tárgyát képezik.

A célsokaság azon – az alapsokasággal jellemzően megegyező, vagy szűkebb, – egységek összessége, amelyekre az adatgyűjtés irányul.

A felvételi keret lista, térkép vagy adathordozó, ami behatárolja, azonosítja, és hozzáférést biztosít az adatszolgáltató és/vagy megfigyelési egységekhez. Általában két típusa van: területi és listás. Felvételi kerethez leggyakrabban statisztikai regisztereknek a célsokaság elemeit tartalmazó, befagyasztott állapotát használjuk.

A felvételi keretet a KSH-ban, más néven az adatgyűjtés vonatkozási körének is nevezzük.

A felvételi keret szolgál az adatszolgáltatók és/vagy megfigyelési egységek meghatározására.

Teljeskörű felvétel esetén a felvételi keret minden eleme kiválasztásra kerül, míg reprezentatív adatgyűjtés esetén ez szolgál a mintaválasztás alapjául. A mintába bekerülő egységeket nevezzük a minta elemeinek.

Vannak olyan felvételek, amelyeknél egyszerre több keretet is használhatunk (adatszolgáltatók kerete; beszámolási, számbavételi egységek kerete; megfigyelési egységek kerete).

A lefedettség a felvételi keret és a célsokaság közötti eltérést jellemzi. Lefedettségi hibát jelent, ha a keret nem fedi le a teljes célsokaságot (lefedettségi hiány), ha olyan egységeket is tartalmaz, ami nem tartozik a célsokasághoz (lefedettségi többlet), vagy ha a felvételi keretben előfordul duplikáció vagy téves besorolás.

### 2. Alapelvek:

A felvételhez használt keret feleljen meg a célsokaságnak, a lefedettségi hiány illetve többlet a lehető legkisebb legyen. A felvételi keret a legfrissebb információkat tartalmazza.

A felvételi keret egységeihez tartozó információk (pl. azonosítók, elérhetőség, a besoroláshoz szükséges jellemzők) pontosak legyenek, mert ezek segítségével végezzük a rétegezést, minta kijelölését, adatfeldolgozást, imputálást, becslést, adatkapcsolást, minőségértékelést és elemzést. Fordítsunk figyelmet a felvételi keret hibáira (lefedettségi problémák és avult információk), mivel azok torzíthatják a felvétel eredményeit, ronthatják minőségét és növelhetik a költségeket.

### 3. Minőségi irányelvek:

- Azonos célsokaságra vonatkozó felvételeknél **ugyanazt a felvételi keretet** kell használni a konzisztencia érdekében.
- A felvételi keretben szereplő információk **aktualizálását** a lehető leggyakrabban végezzük el.
- Rendszeresen végezzük el a **felvételi keret lefedettségi ellenőrzését**. Alkalmazható módszerek például: a felvételi keret (vagy egy része) megfeleltetése más, az adott célsokaságra (vagy annak részére) vonatkozó összehasonlítható forrásokkal; a felvételi keretet használó különböző felvételek megvalósulásának, eredményeinek (halál/megszűnés, körből kikerülő egységek, jellemzők változása) elemzése; a

kérdőíveken a felvételi keret karbantartását szolgáló kérdések szerepeltetése; helyi szervekkel (regionális igazgatóság, összeírók, önkormányzatok) végzett ellenőrzés.

- Egyes esetekben figyelembe kell venni a **felvételi keretben bekövetkező változásokat**, az adott felvétel mintakiválasztásának ideje és a felvétel referenciaideje között.
- A **felvételi keret és a célsokaság közötti eltérések** kezelésére keressünk és használjunk kiegészítő információkat.
- A **felvétel dokumentációjában** szerepeljen a célsokaság, a felvételi keret és a lefedettség leírása.
- A statisztikusok általános és minőségbiztosítással kapcsolatos **képzésében** megfelelő hangsúllyal szerepeljen a lefedettség kérdése.

### **Irodalom:**

VONATKOZÁSI KÖR KIJELÖLÉS LEÍRÁSA –adatgyűjtés leíró metaadatok. GÉSA Felhasználói Kézikönyv 2007. (Informatikai doktár)

MARTON Á. (1991): A reprezentatív felvételek megbízhatósága. KSH Könyvtár és dokumentációs szolgálat, Budapest.

HUNYADI L. – VITA L.(2002): Statisztika közgazdászoknak. KSH, Budapest.

COLLEDGE, M. (2005): Coverage Procedures for Statistical Surveys and Methods for Assessment of their Quality. Mintavételi és módszertani osztály honlapja: „Minőség a statisztikában” Pilot Projectbeszámoló anyagai)

LESSLER, J. T. – KALSBECK, W. D. (1992): Nonsampling Errors in Surveys. Wiley, New York.

LUNDSTRÖM, S. – SÄRNDAL, C.-E. (2001): Estimation in the presence of nonresponse and frame imperfections. Statistics Sweden, SCB-tryck, Örebro.



## Célok, felhasználás és felhasználók meghatározása

### 1. Leírás:

Egy új vagy egy módosítás alatt álló felvétel tervezési folyamata tartalmazza a tágran értelmezett célkitűzések meghatározását, a megcélzott felhasználói kört és azokat a kulcskérdéseket vagy problémaköröket, amelyek vizsgálatára az elemzés irányulni fog. Amikor ez a kezdeti tervezés átalakul aktuális adatgyűjtési tevékenységgé, akkor a célokat és a felhasználás módját már pontosabban kell megfogalmaznunk annak érdekében, hogy biztosíthassuk, hogy az új vagy a módosított felmérés megfelel a speciális felhasználói igényeknek is.

A statisztikai felvétel célkitűzései azok az információ szükségletek, amelyek társadalmi és gazdasági programok, kutatási problémák vagy hipotézisek keretében információszükségletként fogalmazódnak meg. A felvétel eredményeinek felhasználása (hasznosítása) pontosítja és részletezi, hogy milyen döntések megalapozását szolgálják a begyűjtendő adatok. A rendszeres adatgyűjtések esetén idővel más felhasználási módok is felmerülhetnek. A felhasználók azok a szervezetek, hivatalok, csoportok vagy magánszemélyek, amelyek feltehetően hasznosítani fogják az összegyűjtött információkat.

### 2. Alapelvek:

A speciális felvételi célokról és a felhasználásról kialakított konszenzus elősegíti a racionális döntéseket az adatgyűjtés tervezése során.

A célok meghatározásánál a felhasználók bevonása elősegítheti, hogy valóban a számukra releváns információk kerüljenek előállításra a felvétel során, ezért minden esetben célszerű őket bevonni a folyamat ezen szakaszába.

Konkrétan fogalmazzuk meg, és világosan értelmezzük a célkitűzéseket arra vonatkozóan, hogy milyen hipotéziseket akarunk tesztelni, és milyen adatigényeink vannak, figyelembe véve az elvárt minőséget, a költségvetési lehetőségeket és az elvárt határidőket.

A célkitűzéseket úgy kell meghatározni, hogy segítse a (potenciális) felhasználók számára érthetővé tenni, hogy mit várhatnak el ezektől a statisztikai termékektől, és hogy milyen szempontokat kell figyelembe venni az adatok felhasználása során.

### 3. Minőségi irányelvek:

- A felhasználói igények elemzése során törekedni kell a **legköltséghatékonyabb megoldások** megtalálására, rövid és hosszú távra egyaránt. Mielőtt egy új statisztikai tevékenységbe (vagy egy meglévő újjáalakításába) kezdenénk, elemezni kell az aktuálisan hozzáférhető statisztikákat a források, gyakoriságok, minőség, időtényezők stb. szempontjából. Szamba kell venni, hogy a hozzáférhető adatok mennyire felelnek meg a megrendelők igényeinek, egy új felvétel végrehajtásához milyen idő-költség keretekkel lehet számolni.
- A felvétel célkitűzéseit a fontos **felhasználókkal és résztvevőkkel való együttműködésben** alakítsuk ki. Hozzuk létre és ápoljuk a kapcsolatot az információk felhasználóival a magán- és a közsférában, a tudományos világban és a közvéleménnyel is. Ez a hivatali termék- és szolgáltatási marketing részeként is felfogható. A fontos felhasználók között található a potenciális piac képviselői, döntéshozók és hivatalok,

akik az információkat hivatalos felhasználásra igénylik. A tervezés, illetve újratervezés előtt rutinszerűen hajtsunk végre átfogó, alapos és fókuszcsoport jellegű konzultációkat a felhasználókkal, annak érdekében, hogy azonosítsuk a tartalmi szempontokat, és hogy nyilvános támogatást szerezzünk az adatfelvételi program számára (adatszolgáltatók megnyerése). A releváns és pontos statisztikák csak a meglévő bizalommal együtt érnek bármit is, ezért, nagyon fontos a nyitott megközelítés egy adatfelvétel kialakításakor, felújításakor.

- Miközben meghatározzuk, hogy az adatfelvétel milyen mértékig elégíti ki a felhasználók igényeit, ezeket az igényeket **a költségekkel, a válaszadói terhekkel és az adatvédelmi szempontokkal** is össze kell hangolni. Lehet, hogy az adatfelvételt végrehajtó szervezetnek csak kis beleszólása van, ha jogszabályban elrendelt adatfelvételtől van szó, más esetekben viszont érdemes alternatív módszertani megközelítéseket (adatgyűjtési módszereket, gyakoriságokat, földrajzi részleteket stb.) is kidolgozni ahhoz, hogy egy optimális megoldáshoz lehessen eljutni.
- Rendszeres időközönként **vizsgáljuk felül a folyamatban lévő statisztikai tevékenységeket**. A statisztikai programokat a felhasználói igényeknek megfelelően kell fejleszteni, átdolgozni és megújítani. A tevékenység céljait vagy a célkitűzések szempontjait időnként felül kell vizsgálni, hogy növelhessük a statisztikai termék relevanciáját a felhasználói igények szerint, amelyek idővel bővíthetnek, megváltozhatnak.
- Ha vannak kifejezetten megcélzott adatminőségi elvárások, emeljük be ezeket a kutatási célkitűzések mérhető minőségi szempontjainak sorába. A célokat csoportosíthatjuk a minőség összetevők – úgymint például pontosság vagy időbeli pontosság – szerint. Az adminisztratív adatok és a statisztikai adatok másodlagos felhasználása esetében az output minősége közvetlenül függ az input adatforrások minőségétől.
- A speciális statisztikai tevékenység céljainak és felhasználásainak meghatározásakor vegyük figyelembe a statisztikai adatok másodlagos felhasználása vagy statisztikai keretrendszerek (pl. a Nemzeti Számlák Rendszere) célkitűzéseit is.

## **Irodalom:**

BRACKSTONE, G. J. (1991). Shaping Statistical Services to Satisfy User Needs. Statistical Journal of the United Nations, ECE 8, 243-257. (A statisztikai szolgáltatások átalakítása a felhasználói igények kielégítése érdekében)

BRACKSTONE, G. J. (1993). Data Relevance: Keeping Pace with User Needs. Journal of Official Statistics, 9, 49-56. (Adat-relevancia: Lépéstartás a felhasználói igényekkel)

## **Referencia:**

A felhasználói véleménykutatás rendszerének kidolgozása

[http://kshsps/PMA/Shared%20Documents/SFT%20előterjesztések/S20302\\_Zarojelentes\\_sft080627.doc](http://kshsps/PMA/Shared%20Documents/SFT%20előterjesztések/S20302_Zarojelentes_sft080627.doc)

## Fogalmak, definíciók és osztályozások meghatározása

### 1. Leírás:

A fogalmak a felvételben szereplő statisztikai egység vagy célsokaság általános jellemzőire vagy tulajdonságaira vonatkoznak. A definíciók nem csak a fogalmak leírását tartalmazzák, hanem azokét a változókét, mutatókét, amelyeket a fogalmak által leírt jelenségek méréséhez használunk. Az **osztályozás** olyan csoportosítás, ahol a csoportok teljesen lefedik a sokaságot, és egymást kizárják, vagyis a sokaság minden eleme egy és csakis egy csoportba sorolható.

### 2. Alapelvek:

A fogalmak, változók, mutatók definícióinak egyértelműnek és világosnak kell lenniük az adatgyűjtés céljának megfelelően. Mivel a statisztikai adatokat valamilyen ismérv szerint csoportosítani kell a tájékoztatási célú elemzés érdekében, ezért a kiválasztott ismérv az elemzés céljához kell, hogy igazodjon.

Annak érdekében, hogy egy adatállományból következtetéseket lehessen levonni, rendkívül fontos a felhasználók számára, hogy megismerhessék ezeket a fogalmakat, változókat, mutatókat, vagyis azt, hogy mi a felvétel tárgya. Bár az összehangolt fogalmak, változók, mutatók és ezek definíciói, valamint az osztályozások segítik a felhasználókat az adatok összehasonlításában és az adatintegrációban, de helyettesítő definíciók használata előfordulhat az eltérő igények (célok) miatt. A fogalmakat, változókat, mutatókat és osztályozásokat pontosan dokumentálni kell, és fel kell hívni a figyelmet mindenfajta eltérésre a standardoktól vagy azoktól, amelyeket a vonatkozó adatok előállításához használtak. Ez a dokumentálás különösen fontos azoknak a felhasználóknak, akik az adatokat pl.: továbbszámításhoz használnák.

A változók, mutatók és osztályozások kiválasztásánál figyelembe kell venni további tényezőket is, mint például a szükséges információ megszerzésének nehézsége, az okozott válaszadói teher, az adatgyűjtés módszere, a kérdések szövegkörnyezete, az adatok feldolgozásánál használt módszerek, a felhasználható adminisztratív nyilvántartásoknál alkalmazott definíciók, osztályozások, valamint az adatgyűjtés és feldolgozás költsége. Egy adott időpontban választott definíció, később elavulttá válhat, ezért lehet, hogy módosítani vagy változtatni kell ezeket.

### 3. Minőségi irányelvek:

- A fogalmakat, változókat és mutatókat a felhasználási céljukkal együtt világosan határozzuk meg.
- Törekedjünk a **nemzetközileg elfogadott módszertan (fogalmak, változók, mutatók, osztályozások)** használatára. Az elnevezések megválasztásakor vegyük figyelembe a meglévő standardokkal és felhasználói célokkal való azonosságokat vagy eltéréseket. (Ld. a KSH Metainformációs rendszerét.) A meglévő standard fogalmakat csak olyan célra használjuk, amely a standardban meghatározásra került. A KSH-ban a fogalmak, változók, mutatók tekintetében a metainformációs rendszer a mérvadó.
- A fogalmak, változók, mutatók könnyebb értelmezhetősége érdekében a tájékoztatásban jelezzük a más fogalmakkal, változókkal, mutatókkal való kapcsolatot is (pl.: szűkebb értelmű, tágabb értelmű, kapcsolódó, szinonim fogalom).

- A rugalmas és maximális felhasználhatóság érdekében kódoljuk a mikroadatokat, és az adatállományokat a megfelelő osztályozás lehető legalacsonyabb szintjén kezeljük. A magasabb szintű aggregációra különleges elemzési célokhoz lehet szükség vagy az adatvédelem vagy megbízhatósági követelmények kielégítéséhez. Egységes aggregálási eljárásokat használunk és pontosan határozzuk meg, hogy a standard osztályozás melyik szintjéből, kategóriájából állítjuk elő az aggregátumokat. Dokumentáljuk a standard osztályozásoktól való eltéréseket. Használunk **olyan osztályozásokat, amelyek mind részletesebb, mind pedig aggregált szinteket is tartalmaznak**. Tegyük világossá a felhasználók számára, hogyan illeszkednek ezek a magasabb szintű (vagyis kevésbé részletes) osztályozásokhoz.
- A standard definíciók és osztályozások használata lehetővé teszi a különböző forrásból származó **adatok összehasonlítását és az adatintegrációt**. A KSH-ban a nemzetközileg elfogadott, az EU-ban, ENSZ-ben, illetve más nemzetközi intézményekben is alkalmazott standard osztályozásokat kell használni.
- Az adatok összehasonlíthatósága érdekében fordítókulcsot kell készíteni az azonos tárgyú, de különböző, illetve a teljesen azonos, de időben eltérő nemzetközi és magyar osztályozások között. Nomenklatúra váltásnál illetve a nemzetközi adatszolgáltatásokhoz a hivatalos fordítókulcsokat kell használni.
- Az adatok összehasonlítása érdekében **standard megfigyelési egységeket** kell használni.
- A statisztikai adatok **másodlagos használata vagy statisztikai keretrendszerek** (pl. a Nemzeti Számlák Rendszere) összeállítása során alkalmazott fogalmakra, változókra és mutatókra külön figyelmet kell fordítani, mivel jelentős hatással lehetnek az egyes adatgyűjtésekre.
- Hivatalos statisztikai standard hiányában meg kell vizsgálni a **kapcsolódó statisztikai programokban** használt fogalmakat, változókat, mutatókat és osztályozásokat.

## Irodalom:

GAZDASÁGI TEVÉKENYSÉGEK egységes ágazati osztályozási rendszere és a tevékenységek tartalmi meghatározása (TEÁOR'08)

[http://portal.ksh.hu/pls/portal/docs/PAGE/KSHPORTAL/SZOLGALTATASOK/OSZTALY\\_OZASOK/HATALYOS\\_MAGYAR\\_OSZTALYOZASOK/TEAOR08TARTALOM\\_AKT\\_2\\_009\\_05\\_13.PDF](http://portal.ksh.hu/pls/portal/docs/PAGE/KSHPORTAL/SZOLGALTATASOK/OSZTALY_OZASOK/HATALYOS_MAGYAR_OSZTALYOZASOK/TEAOR08TARTALOM_AKT_2_009_05_13.PDF)

A STATISZTIKAI EGYSÉGEK gazdálkodási forma szerinti osztályozása, 2006 (GFO)

[http://portal.ksh.hu/pls/portal/docs/PAGE/KSHPORTAL/SZOLGALTATASOK/OSZTALY\\_OZASOK/HATALYOS\\_MAGYAR\\_OSZTALYOZASOK/GFO\\_06.PDF](http://portal.ksh.hu/pls/portal/docs/PAGE/KSHPORTAL/SZOLGALTATASOK/OSZTALY_OZASOK/HATALYOS_MAGYAR_OSZTALYOZASOK/GFO_06.PDF)

SZOLGÁLTATÁSOK JEGYZÉKE (SZJ'03)

[http://portal.ksh.hu/pls/portal/docs/PAGE/KSHPORTAL/SZOLGALTATASOK/OSZTALY\\_OZASOK/HATALYOS\\_MAGYAR\\_OSZTALYOZASOK/SZJ03.PDF](http://portal.ksh.hu/pls/portal/docs/PAGE/KSHPORTAL/SZOLGALTATASOK/OSZTALY_OZASOK/HATALYOS_MAGYAR_OSZTALYOZASOK/SZJ03.PDF)

SZAKMAKÓDOK JEGYZÉKE, 2008

[http://portal.ksh.hu/pls/portal/docs/PAGE/KSHPORTAL/SZOLGALTATASOK/OSZTALY\\_OZASOK/HATALYOS\\_MAGYAR\\_OSZTALYOZASOK/SZAKMAKOD08.PDF](http://portal.ksh.hu/pls/portal/docs/PAGE/KSHPORTAL/SZOLGALTATASOK/OSZTALY_OZASOK/HATALYOS_MAGYAR_OSZTALYOZASOK/SZAKMAKOD08.PDF)

NACE Rev 2.

[http://portal.ksh.hu/pls/portal/docs/PAGE/KSHPORTAL/SZOLGALTATASOK/OSZTALY\\_OZASOK/NEMZETKOZI\\_OSZTALYOZASOK/NACE%20REV2%20PUBLICATION\\_2008-09-02.PDF](http://portal.ksh.hu/pls/portal/docs/PAGE/KSHPORTAL/SZOLGALTATASOK/OSZTALY_OZASOK/NEMZETKOZI_OSZTALYOZASOK/NACE%20REV2%20PUBLICATION_2008-09-02.PDF)

CPA 2008

[http://portal.ksh.hu/pls/portal/docs/PAGE/KSHPORTAL/SZOLGALTATASOK/OSZTALY\\_OZASOK/NEMZETKOZI\\_OSZTALYOZASOK/CPA%202008%20STRUCTURE%20AND%20EXPLANATORY%20NOTES\\_2008-09-02.PDF](http://portal.ksh.hu/pls/portal/docs/PAGE/KSHPORTAL/SZOLGALTATASOK/OSZTALY_OZASOK/NEMZETKOZI_OSZTALYOZASOK/CPA%202008%20STRUCTURE%20AND%20EXPLANATORY%20NOTES_2008-09-02.PDF)

ÉPÍTMÉNYJEGYZÉK, 2000

[http://portal.ksh.hu/pls/portal/docs/PAGE/KSHPORTAL/SZOLGALTATASOK/OSZTALY\\_OZASOK/HATALYOS\\_MAGYAR\\_OSZTALYOZASOK/EPITMENYJEGYZEK.PDF](http://portal.ksh.hu/pls/portal/docs/PAGE/KSHPORTAL/SZOLGALTATASOK/OSZTALY_OZASOK/HATALYOS_MAGYAR_OSZTALYOZASOK/EPITMENYJEGYZEK.PDF)

TERÜLETI SZÁMJELRENDSZER, 2008

[http://portal.ksh.hu/pls/portal/docs/PAGE/KSHPORTAL/SZOLGALTATASOK/OSZTALY\\_OZASOK/HATALYOS\\_MAGYAR\\_OSZTALYOZASOK/TERSZJEL2008.PDF](http://portal.ksh.hu/pls/portal/docs/PAGE/KSHPORTAL/SZOLGALTATASOK/OSZTALY_OZASOK/HATALYOS_MAGYAR_OSZTALYOZASOK/TERSZJEL2008.PDF)

NUTS – Nomenclature of Territorial Units for Statistics, 2003

ORSZÁG REGISZTER, 2009

[http://portal.ksh.hu/pls/ksh/docs/hirek/kshxml/Geo2081\\_vKSHXML\\_1.xls](http://portal.ksh.hu/pls/ksh/docs/hirek/kshxml/Geo2081_vKSHXML_1.xls)

<http://portal.ksh.hu/pls/ksh/docs/hun/info/02osap/nomen/Orszagregiszter.xls>

CN (Combine Nomenclature), azaz Kombinált Nómenklatúra a külkereskedelmi termékforgalom megfigyelésére alkalmazott, az EU által előírt termékosztályozás (8 számjegyű). Megegyezik az adójogszabályokban hivatkozott Kereskedelmi Vámtarifával (VTSZ).

[http://portal.ksh.hu/pls/portal/docs/PAGE/INTRASTAT/INTRASTAT\\_KOMB\\_NOM/32008\\_R1031.PDF](http://portal.ksh.hu/pls/portal/docs/PAGE/INTRASTAT/INTRASTAT_KOMB_NOM/32008_R1031.PDF)

PRODCOM Jegyzék

A FEOR foglalkozásainak tartalmi leírása

<http://kshintra/intra/tajanyag/nomenklaturak/Feor93/index.htm>

<http://portal.ksh.hu/pls/ksh/docs/szolgalatasok/hun/feor/index.html>

FEOR FELÜLVIZSGÁLATA, FEOR 08 bevezetése (S10302) Stratégiai terv

HÁZTARTÁSSAL KAPCSOLATOS MÓDSZERTANI FÜZETEK, leírások

[http://portal.ksh.hu/portal/page?\\_pageid=37,184441&\\_dad=portal&\\_schema=PORTAL](http://portal.ksh.hu/portal/page?_pageid=37,184441&_dad=portal&_schema=PORTAL)

KÜLKERESKEDELEM-STATISZTIKA ORSZÁGBESOROLÁS, Geonómenklatúra

[http://eur-lex.europa.eu/LexUriServ/site/en/oj/2006/l\\_354/l\\_35420061214en00190028.pdf](http://eur-lex.europa.eu/LexUriServ/site/en/oj/2006/l_354/l_35420061214en00190028.pdf)

A FOGLALKOZTATÁSI VISZONY OSZTÁLYOZÁSI RENDSZER

ISCED 1997 (International Standard Classification of Education ), UNESCO

Termékek és Szolgáltatások Osztályozási Rendszere (TESZOR08)

[http://portal.ksh.hu/pls/portal/docs/PAGE/KSHPORTAL/SZOLGALTATASOK/OSZTALY\\_OZASOK/HATALYOS\\_MAGYAR\\_OSZTALYOZASOK/TEZSOR08\\_TARTALOM2009\\_05\\_12.PDF](http://portal.ksh.hu/pls/portal/docs/PAGE/KSHPORTAL/SZOLGALTATASOK/OSZTALY_OZASOK/HATALYOS_MAGYAR_OSZTALYOZASOK/TEZSOR08_TARTALOM2009_05_12.PDF)

Ipari Termékosztályozás (ITO'08)

[http://portal.ksh.hu/pls/portal/docs/PAGE/KSHPORTAL/SZOLGALTATASOK/OSZTALY\\_OZASOK/HATALYOS\\_MAGYAR\\_OSZTALYOZASOK/EVES\\_TERMEKJEGYZEK\\_2008.PDF](http://portal.ksh.hu/pls/portal/docs/PAGE/KSHPORTAL/SZOLGALTATASOK/OSZTALY_OZASOK/HATALYOS_MAGYAR_OSZTALYOZASOK/EVES_TERMEKJEGYZEK_2008.PDF)

# Felhasználható igazgatási adatok számbavétele, statisztikai hasznosításuk

## 1. Leírás:

A statisztikáról szóló 1993. évi XLVI. Törvény 21.§ 4. (b) pontja értelmében a KSH jogosult az igazgatási szerveknél rendelkezésre álló nyilvántartásokból<sup>1</sup> – statisztikai célra – adatokat átvenni. Az adminisztratív nyilvántartások olyan adatokat tartalmaznak, melyeket nem statisztikai programok végrehajtása céljából gyűjtöttek<sup>2</sup>, pl. adóadatok, bizonyos események (anyakönyvezés), bizonyos juttatások (nyugdíjak, szociális támogatások) regisztrálása.

Az adminisztratív nyilvántartások adatainak statisztikai célú felhasználása közérdek, mert úgy növeli a KSH adatokhoz való hozzáférési lehetőségeit, hogy közben csökkenti az adatszolgáltatói terheket, és ezzel megtakarítja egy további adatgyűjtés költségeit.

Az adminisztratív nyilvántartások statisztikai célú felhasználása magában foglalja:

- a felvételek keretként való felhasználását, közvetlenül keretként vagy a meglévő keretek kiegészítéseként;
- adatgyűjtések helyettesítését (pl. a kisvállalkozások esetében az adóadatok használata a statisztikai felvételek helyett);
- a hivatali mintavételes felvételek editálásában és imputálásában való felhasználásukat;
- közvetett használatukat a becslésekben (pl. kiegészítő információként a kalibrálásban, összehasonlító elemzésben);
- végül, a statisztikai adatfelvételek értékelését, beleértve a különböző forrásokból származó adatok ütköztetését (pl. a felvételekből származó becslések összehasonlítását az igazgatási adatokból történő becslésekkel).

## 2. Alapelvek:

A KSH mindig törekedjen az igazgatási adatok felhasználására, ha azzal kiválthat egy közvetlen adatgyűjtést, azaz csökkentheti az adatszolgáltatói terheket.

Az adminisztratív nyilvántartás adatainak felhasználásáról szóló döntéshez az illető nyilvántartást értékelni, tesztelni, az értékelést pedig dokumentálni kell. Ez az értékelés ki kell, hogy terjedjen az adott adatgyűjtés céljára, a megfigyelési egységeire, lefedettségére, tartalmára, fogalmaira és definícióira, az alkalmazott osztályozási rendszerére, valamint az adminisztratív program végrehajtása során alkalmazott minőségbiztosításra és folyamatellenőrzésre, az adatok gyakoriságára és a statisztikai hivatalnak átadott adatok határidőire.

Az adminisztratív nyilvántartások statisztikai célú használata növeli a nyilvánosságra kerülő információk adatvédelmével kapcsolatos kockázatokat, főként, ha az adminisztratív információkat más adatforrásokkal kapcsolják össze. Ezért ezt az adatvédelem hivatali megvalósításakor figyelembe kell venni.

---

<sup>1</sup> A KSH-ban 2004-ben kezdődött az igazgatási adatok hivatalon belüli számbavétele az S105 „Adminisztratív nyilvántartások adatainak statisztikai hasznosítása” program keretében.

<http://kshintra:8080/mport/index.php?option=displaypage&Itemid=90&op=page&SubMenu=>

<sup>2</sup> A Kormány 2000-ben döntött a 1113/2000. (XII. 27.) Kormány Határozatában a közigazgatási adatvagyon-nyilvántartás kialakításáról.

### 3. Minőségi irányelvek:

- Miután az igazgatási adatok bekerülnek a KSH rendszerébe, akkor az abból származó mutatókért a KSH-nak felelősséget kell vállalnia függetlenül attól, hogy az statisztikai vagy adminisztratív adatgyűjtésből származik.
- A további fejezetek több irányelve alkalmazandó az adminisztratív adatokra is. Pl. a mintavételi és az adatrögzítési irányelvek relevánsak lehetnek, ha az adminisztratív adatok csak papíron vannak meg és azokat kódolni kell, és számítógépre kell vinni.
- **Együtt kell működni** az új vagy fejlesztés alatt levő adminisztratív nyilvántartási rendszerek tervezőivel. Ez már a kezdetektől segíti a statisztikai igények adminisztratív rendszerekbe való beépítését.
- Folyamatos kapcsolatban kell lenni az adminisztratív adatok szolgáltatóival. Különösen szükséges ez a kapcsolat az adminisztratív adatok felhasználásának kezdetekor. A statisztikai információk, és az **adatokban talált hibák visszacsatolása** értékes lehet az adatszolgáltatónak, mert az adminisztratív források javítását segíti elő.
- Meg kell ismerni azt a környezetet, melyben az **adminisztratív szervezet** végrehajtotta az adatgyűjtési programot (pl. jogi alapok, célok, igények).
- A statisztikai célból begyűjtött adatokhoz hasonlóan az igazgatási adatok esetében is előfordul a részleges vagy teljes nemválaszolás. Ha a nem-válaszolók nem lehet megkeresni és így megszerezni az adatokat, akkor ki kell dolgozni egy imputálási vagy súlykorrekciós eljárást, és így kell kezelni a nem-válaszolást. Az adminisztratív források néha elavultak, nem aktuálisak. Ezért, az imputálási folyamat részeként külön figyelmet kell fordítani a létező (aktív) és a megszűnt (nem aktív) egységek azonosítására. Valamilyen imputálásra vagy korrekcióra szükség lehet azokban az esetekben, amikor néhány egység eltérő gyakoriságú adatot jelent (pl. heti vagy negyedéves), mint amit elvárnak (pl. havi).
- A dokumentáció segít a statisztikusoknak abban, hogy eldöntsék, melyik felhasználás a legjobb az adott igazgatási adat esetében. Az igazgatási adatok statisztikai feldolgozási programjához **megfelelő módszertant** kell választani, és a felhasználókat tájékoztatni kell a módszertanról és az adatminőségről.
- Figyelembe kell venni, hogy az **igazgatási adatforrás múltbeli adatai**, és az azok feletti hatáskör mindig teljes mértékben az adminisztratív szervezet kezében van. Azok az igazgatási szempontok, melyek eredetileg meghatározták a programban használt fogalmakat, módszereket idővel megváltozhatnak, s ez torzíthatja az adminisztratív állományokból nyert idősorokat. Óvatosan kell kezelni az ilyen jellegű változásokat, és ezek hatását korrigálni kell a statisztikai célú feldolgozás keretében.
- Az igazgatási adatokból származó információk publikálásakor az **adattvédelmi következményekre** is figyelemmel kell lenni.
- **Adattvédelmi kérdések** merülhetnek fel egyetlen adminisztratív adatforrás használatakor is, melyek megsokszorozódnak, amikor más forrásokat is összekapcsolnak. A KSH alkossa meg az „Adatintegrációs szabályzatát”, mely az adatok összehangolásából származó közhaszon biztosítása mellett garantálná a megfelelő adattvédelmet.
- Az adatintegrációs eljárások két típusát különböztetjük meg. Amennyiben az integrálni kívánt adatforrások egységei egy részében átfedik egymást és rendelkeznek egyedi, korrekt azonosítókkal, akkor a Record Linkage módszerrel a kulcsváltozókon keresztül történik az adatintegráció. Segítséget nyújt a közös (egységes) kódlisták, fogalmak, változók, mutatók, osztályozások használata. Bonyolultabb a helyzet, amennyiben nincs ismeretünk arról, hogy van-e átfedés az adatforrások egységei között. Vannak változók,



mutatók, melyek nem lettek megfigyelve mindkét adatforrásban. Ezeket hiányzó adatoknak tekintjük. A közös változókon, mutatókon keresztül Statistical Matching módszerrel szükséges a hiányzó adatokról információt szerezni, keressük a leginkább hasonló egységekre vonatkozó rekordokat.

- Néhány **adminisztratív adat hosszmetzeti (longitudinális) jellegű** (pl. jövedelemadó, termék- és szolgáltatási adó). A különböző időpontra vonatkozó állományok integrációjával a kutatók számára sokoldalúan felhasználható adatállományok hozhatók létre. Különös gonddal kell eljárni ilyen (longitudinális) jellegű és személyekre vonatkozó adatbázisok esetében, mert használatuk komoly adatvédelmi problémákkal jár. Az azonosító használatára nagy figyelmet kell fordítani, mivel idővel változhat az egység azonosítója. Nyomon kell követni az ilyen változásokat, hogy megfelelő időbeli adatelemzés készüljön. Néhány esetben, ugyanannak az egységnek kettő vagy több azonosítója is lehet, ugyanabban a referencia időszakban, így duplázások fordulhatnak elő az adminisztratív állományban. Ha ez előfordul, akkor ki kell dolgozni egy duplázódásokat kizáró eljárást.
- Az adminisztratív állományok gyakran nagyon nagy méretűek, és felhasználásuk néha jelentős feldolgozási költséggel és időigénnyel járhat. A **költségcsökkentés érdekében** - igény szerint - **véletlen mintát lehet venni** a nagy adminisztratív állományokból.

## Irodalom:

1993. évi XLVI. Törvény a statisztikáról

[http://net.jogtar.hu/jr/gen/hjegy\\_doc.cgi?docid=99300046.TV](http://net.jogtar.hu/jr/gen/hjegy_doc.cgi?docid=99300046.TV)

Az Európai Parlament és Tanács 223/2009/EK rendelete (2009. március 11.) az európai statisztikákról...

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:087:0164:0173:HU:PDF>

1992. évi LXIII. törvény a személyes adatok védelméről és a közérdekű adatok nyilvánosságáról

<http://abiweb.obh.hu/adatved/magyar/avtv.htm>

2005. évi XC. törvény az elektronikus információszabadságról

[http://net.jogtar.hu/jr/gen/hjegy\\_doc.cgi?docid=a0500090.tv&timeshift=1](http://net.jogtar.hu/jr/gen/hjegy_doc.cgi?docid=a0500090.tv&timeshift=1)

2007. évi CI. törvény a döntés

[http://www.nfu.hu/download/2074/CI\\_2007\\_tv.pdf](http://www.nfu.hu/download/2074/CI_2007_tv.pdf)

A Kormány 2212/2007. (XI. 14.) kormányhatározata az államigazgatási nyilvántartások adatainak statisztikai célra történő felhasználásáról

<http://www.kozlonykiado.hu/kozlonyok/Kozlonyok/10/PDF/2007/56.pdf>

1113/2000. (XII. 27.) Kormány Határozat a közigazgatási adatvagyon-nyilvántartásról

ELŐREHALADÁSI JELENTÉS az SFT részére - Az Adminisztratív nyilvántartások adatainak statisztikai célú hasznosítása (S105) főirány keretében végzett 2006. évi tevékenységekről

<http://kshintra:8080/mport/index.php?option=displaypage&Itemid=90&op=page&SubMenu>

HANDBOOK ON IMPROVING quality by analysis of process variables

[http://epp.eurostat.ec.europa.eu/cache/ITY\\_PUBLIC/G0-LEG-20040510/EN/G0-LEG-20040510-EN.PDF](http://epp.eurostat.ec.europa.eu/cache/ITY_PUBLIC/G0-LEG-20040510/EN/G0-LEG-20040510-EN.PDF)

QUALITY ASSESSMENT of Administrative Data for Statistical Purposes (2003). Eurostat.

BRACKSTONE, G.J. (1987). Issues in the use of administrative records for statistical purposes. *Survey Methodology*, 13, 29–43.

COX, L.H.. – BORUCH, R.F. (1988). Record linkage, privacy and statistical policy. *Journal of Official Statistics*, 4, 3–16.

HIDIROGLOU, M.A., LATOUCHE, M.J., ARMSTRONG, B., and GOSSEN, M. (1995). Improving survey information using administrative records: the case of the Canadian Employment Survey. *Proceedings of the Annual Research Conference, U.S. Bureau of the Census*, 171–197.

KOVACEVIC, M. (1999). Record linkage and statistical matching – they aren't the same! *SSC Liaison*, Vol. 13, No. 3, 24–29.

MICHAUD, S., DOLSON, D., ADAMS, D., and RENAUD, M. (1995). Combining administrative and survey data to reduce respondent burden in longitudinal surveys. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 11–20.

MONTY, A.. – FINLAY, H. (1994). Strengths and weaknesses of administrative data sources: experiences of the Canadian Business Register. *Statistical Journal of the United Nations*, ECE 11, 205–210.

GÁRDOS É.: A 2212/2007. (XI.14.) Korm. hat. 2-3. pontjának végrehajtása (2008.)

<http://kshsps/C6/Adminisztratív%20Adatforrások%20és/default.aspx?RootFolder=%2fC6%2fAdminisztrat%c3%adv%20Adatforr%c3%a1sok%20%c3%a9s%2fDocument%20Library%2f2008%2f2212%5f2007%5fkorm%5fhat&View=%7bE1088579%2d2B11%2d4758%2d9EA6%2d3D4E9EF7D166%7d>

## A mintavételi terv kialakítása

### 1. Leírás:

Mintavétel esetén a célsokaságból úgy választunk ki egy részt, amelyet mintának nevezünk, hogy az tükrözze, reprezentálja a célsokaságot. Ilyenkor reprezentatív megfigyelésről beszélünk. A mintavételi módszer megválasztása közvetlen hatással van az adatminőségre. A választást sok tényező befolyásolja, többek között az előállítandó információ pontosságának kívánt szintje, alkalmas mintavételi keretek hozzáférhetősége, megfelelő rétegzéző változók elérhetősége, a felhasználásra kerülő becslési módszerek és a rendelkezésre álló költségvetési források.

A mintavétel célja olyan hasznos információk nyerése a mintába kiválasztott egységekből, amelyek segítségével a célsokaságra vonatkozó következtetéseket tudunk levonni. Ezeknek a következtetéseknek a helyessége jelentős mértékben a minta reprezentativitásától függ. Ahhoz, hogy reprezentatív mintát válasszunk ki a célsokaságból, valószínűségi, véletlen mintavételt használunk. A valószínűségi mintavétel a keretből történő olyan valószínűségi kiválasztás, amelynek során a keret minden elemének ismert és pozitív bekerülési valószínűsége van. A mintanagyság meghatározása a kívánt pontosság és a kiválasztott egységek megfigyelésére felhasználható költségvetési források nagysága alapján történik. A minta kiválasztásának alapjául szolgáló valószínűség-eloszlást mintavételi tervnek nevezzük. Egyes értelmezésekben a mintavételi terv fogalma ennél bővebb, beletartozik a kiválasztási terv is. A mintavételi terv ekkor tartalmazza a rétegzésre, a kiválasztási lépcsőkre és a mintavételi egységek kiválasztásának módjára vonatkozó információkat is.

### 2. Alapelvek:

A mintavételi terv és a becslési módszer (lásd a *Súlyozás, becslés és hibaszámítás* c. fejezetet) együttes kialakítását úgy kell végeznünk, hogy az eredményül kapott becslések az adott költségvetési források mellett a lehető legpontosabbak legyenek, vagy a legkisebb költségráfordítással érjünk el egy rögzített pontosságot. A becslés során a mintába kiválasztott egységekre gyűjtött adatok sok esetben kiegészíthetők a felvételen kívüli más forrásokból, például adminisztratív nyilvántartásokból vagy censusokból származó információval (ún. **segédinformációval**). A segédinformációk elérhetősége befolyásolja a mintavételi terv kialakítását.

### 3. Minőségi irányelvek:

- **Rétegzésről** beszélünk, ha a sokaságot diszjunkt részhalmazokra (ún. rétegekre) bontjuk, és a minta kiválasztását rétegenként végezzük. A kiválasztás egy adott rétegben a többi rétegtől függetlenül történik. Rétegek képzését gyakorlati szempontból igényelheti a felvétel célja, például az, hogy szükség van területi becslésekre. A statisztikai hatásosság szempontjából a rétegzést úgy kell végeznünk, hogy az egyes rétegek a felvételtől várt információ szempontjából viszonylag homogén egységeket tartalmazzanak.
- A rétegzéshez vagy a későbbi becslésekhez szükséges információkat néha nem tartalmazza a keret. Ezekben az esetekben egy első, nagy minta kiválasztásával juthatunk a kívánt információhoz. Az első nagy mintából aztán újabb mintát veszünk, ahol már a kiválasztásnál figyelembe vehetjük az első minta elemeiről rendelkezésre álló információt, vagy a későbbi becslés során használjuk őket. Az ilyen mintavételt **kétfázisú mintavételnek** nevezzük. Ennek tervezéséhez fontos figyelembe vennünk a mintavétel

egyes fázisainak költségét, az egyes fázisokban szükséges információ elérhetőségét és azt, hogy az első minta rétegzése mennyire javítja a pontosságot.

- A gyakorlatban néha előfordul, hogy nehéz kiválasztani vagy kényelmetlen közvetlenül felkeresni azokat az egységeket, amelyektől a kívánt adatokat várjuk. Az ilyen esetekben először olyan egységeket (ún. **elsődleges mintavételi egységeket**) választunk ki, amelyek az adatszolgáltató egységek **csoportjai (klaszterei)**, majd részminták csoportonkénti kiválasztásával jutunk a tényleges mintához. Az ilyen mintavételt **kétlépcsős mintavételnek** nevezzük. Költségvetési vagy egyéb korlátok szükségessé tehetnek kettőnél több lépcsőt. Fontos meghatározni a szükséges lépcsők számát és azt, hogy milyen mintavételi egységek alkalmasak az egyes lépcsőkben. Ennek során figyelembe kell vennünk az adatszolgáltató elérhetőségének és az adatgyűjtésnek a nehézségét, az egység által szolgáltatott adatok minőségét, valamint az egységre vonatkozó, illetve tőle kért adatok gyűjtésének költségét.
- A **mintanagyság** meghatározása során figyelembe kell vennünk a felvételtől előállítandó becslések pontosságának kívánt szintjét, a mintavételi terv típusát, az alkalmazott becslési módszert, segédinformáció elérhetőségét és a várható nem mintavételi tényezőket (pl. nemválaszolás). Periodikus felvételek esetén szem előtt kell tartanunk, hogy a célsokaság változik, egységek jöhetnek létre és szűnhetnek meg.
- A legtöbb felvétel sok változóra ad becslést. Ezért előfordulhat, hogy az a minta, amely optimális egy változó szempontjából, más, fontos változókat illetően hátrányos. Ezt a problémát általában úgy kezeljük, hogy először meghatározzuk a **legfontosabb változókat**, majd e változók részhalmaza alapján állapítjuk meg a mintanagyságot és allokáljuk a mintát.
- Nagyon ferde eloszlású sokaságok esetén, a felvétel során a nagy egységekből létre kell hozni egy **teljes körűen megfigyelt réteget** oly módon, hogy a réteg a sokasági értékösszegek nem elhanyagolható részét kiadja.
- A legtöbb felvétel során a sokaság különböző részeire (angolul ezeknek **domain** a neve, ilyenek lehetnek például a megyék) is adunk becslést. A rétegzés megfelelő kialakításánál ezt figyelembe kell venni.
- A mintavétel megvalósítása során össze kell hasonlítani a tényleges és a várt minta nagyságát és egyéb jellemzőit. Ugyancsak fontos, hogy összehasonlítsuk a becslések tényleges és tervezett pontosságát.
- Az olyan periodikus felvételek esetén, amelyeknél a sokaság nő, mérlegelnünk kell, hogy érdemes-e növelni a mintát. Egy felvétel **pontosságát** általában jobban befolyásolja a teljes mintanagyság, mint a **mintavételi arány** (a minta és a sokaság nagyságának hányadosa).
- Periodikus felvételek mintavételi tervét a lehető **legrugalmasabban** kell kialakítani a tekintetben, hogy kezelni tudjon olyan jövőbeli változásokat, mint a mintanagyság növelése vagy csökkentése, új rétegzés kialakítása, a kiválasztási valószínűségek aktualizálása és új minta ez alapján történő kiválasztása.
- Mind a sokaság változásának megfelelő nyomon követése, mind az adatszolgáltatói terhek figyelembe vétele megköveteli, hogy folyamatos vagy periodikus felvételek mintáinál **rotációt** alkalmazzunk, és időszakonként kicseréljük a mintaelemek egy részét. A rotáció mértékének meghatározása kompromisszumot kíván a változás becslésének pontossága és az egyes adatszolgáltatók terhei között. Kismértékű rotáció azzal az előnnyel is jár, hogy csökkenti az adatszolgáltatókkal történő kapcsolatfelvétel esetleges többletköltségeit.

- A mintavételi terv alapvető jellemzőinek rögzítésére szolgáló dokumentációs sémát minden mintavételes felvétel indulásakor ki kell tölteni, illetve az alkalmazott mintavételi terv változásakor frissíteni kell.

### **Irodalom:**

A Központi Statisztikai Hivatal elnökének III/2005. (SK 2.) KSH előírása a Központi Statisztikai Hivatal rendszeres reprezentatív gazdaságstatisztikai felvételei során alkalmazott rotációról.

A Központi Statisztikai Hivatal elnökének VIII/2008.. (SK 2.) KSH előírása és annak I/2009. (SK 2.) módosítása a statisztikai adatfelvételek tervezéséről.

COCHRAN, W. G. (1977): Sampling Techniques. 3rd ed. John Wiley & Sons. New York.

ÉLTETŐ Ö. – MARTON, Á. – MIHÁLYFFY, L. – TELEGDI, L. (1997): sampling surveys in Hungary. Statistics in Transition. 3 (2): 267–279.

ÉLTETŐ Ö. – MESZÉNA GY – ZIERMANN M. (1982): Sztochasztikus módszerek és modellek. Közgazdasági és Jogi Könyvkiadó. Budapest.

HUNYADI L. (2001): A mintavétel alapjai. Budapesti Közgazdaságtudományi és Államigazgatási Egyetem.

KISH, L. (1989): Kutatások statisztikai tervezése. Angolból fordítva. Statisztikai Kiadó Vállalat. Budapest.

KISH, L. (1965): Survey Sampling. John Wiley & Sons. New York.

SARNDAL, C. E. – SWENSSON, B. – WRETMAN, J. (1992): Model Assisted Survey Sampling. Springer-Verlag. New York.

TELEGDI L. (2004): A kisszervezetek integrált reprezentatív évközi megfigyelése a 2000-es években. Statisztikai Szemle. 82 (8): 668–690.

TELEGDI L. (1999): A nem válaszolás megelőzése és kezelése a gazdaságstatisztikában. I–II. Gazdaság és Statisztika (GÉS). 11 (50/4): 43–64, 11 (50/5): 28–56.

LAKATOS J. – MIHÁLYFFY L.(2003): Az új népszámlálási módszerek hatása a munkaerő-felmérésre. Statisztikai Szemle. 81 (12): 1045-1053

ÉLTETŐ Ö. (2004): Az új HKF-minta kiválasztási eljárása és a 2003. évi tapasztalatok. Statisztikai Szemle.82(8): 648-667

## Kérdőív és segédanyagainak tervezése

### 1. Leírás:

A kérdőív a válaszadótól való információszerzés strukturált eszköze. A kérdőív lehet kérdező segítségével vagy a válaszadó saját maga által kitöltött, papír alapú vagy számítógéppel támogatott adatgyűjtési módszer része. A kérdőívnek központi szerepe van az adatgyűjtési folyamatban. Döntő hatást gyakorol az adatok minőségére, a válaszadói magatartásra, a kérdező teljesítményére, valamint a válaszadóval való kapcsolatra. A kérdőívtervezés az előre meghatározott elérendő célokból indul, majd a dimenzionálás, a dimenziók konkrét mérhetővé tétele, a nyers kérdések, kérdésvázlatok elkészítése következik, végül pedig a konkrét kérdések, információkra irányuló kérések megfogalmazásával zárul. A kérdőívtesztelés elengedhetetlen kelléke a kérdőívtervezésnek. A tesztelés során a kérdőív kitöltésével járó terheket mérjük, valamint azt vizsgáljuk, hogy adatszolgáltató-barát-e a kérdőív és valóban a célkitűzéseknek megfelelő információk nyerhetők belőle.

### 2. Alapelvek:

A kérdőívtervezésnek figyelembe kell vennie az adatfelhasználók statisztikai igényeit, az adatfelvételt végrehajtó szervezet hivatalos követelményeit, továbbá az adatfeldolgozás szempontjait éppúgy, mint a válaszadói kör jellegzetességeit. A jó kérdőív kitöltő-, kérdező-, válaszadó-, feldolgozó- és elemző-barát. A jó kérdőívvel hatékonyan és minimális hibázási kockázattal lehet adatokat gyűjteni, az adatszolgáltatók lehető legkisebb terhelése mellett. A kódolást, adatrögzítést segítő kérdőív csökkenti az editálási és imputálási szükségletet. A kérdőív-szerkesztésnek segíteni kell mind a kitöltést, mind a feldolgozást. (Kérdőív állandó elemei, kérdőív egységes külalakja, stb.)

A folyamatos, rendszeres adatfelvételek kérdőiveit időnként értékelni kell. Minden új és módosított kérdőivet tesztelni kell, mielőtt végleges alkalmazásra kerülne.

### 3. Minőségi irányelvek:

- Olyan **szavakat és fogalmakat használjunk a kérdőívben**, amelyek ugyanazt jelentik a válaszadó és a kérdőív tervezője számára. Intézményi adatfelvételeknél lehetőség szerint olyan kérdéseket, referencia időszakot és válaszkategóriákat jelöljünk meg, amelyek kompatibilisek az intézmény saját adat-nyilvántartási gyakorlatával. Ahol erre lehetőség van, harmonizáljuk a fogalmakat és a kérdések megfogalmazását a már kipróbált, működésben lévőkkel. Ha megfelelnek a felvétel céljainak, hasznosítsuk más adatfelvételek kérdéseit.
- A **kérdések megformázása és nyelvezete** bátorítsa a válaszadókat a kérdőív kitöltésére. Ennek érdekében a kérdőívnek az adatfelvétel témájára kell irányulnia, a kérdések gördülékenyen, logikai sorrendet követve haladjanak előre, segítve a válaszadót a szükséges információk feltárásában. Ügyeljünk arra, hogy mindig a legadekvátabb kérdés típust válasszuk az adott kérdés felmérésére. A nyitott kérdések minden válaszadó számára érthetők legyenek, biztosítsuk, hogy zárt kérdések esetén a válaszkategóriák kölcsönösen kizárják egymást, illetve lefedjék az adott témakört.
- A felvétel célját tegyük világossá a válaszadó számára, hangsúlyozzuk válaszadása, illetve az adott kérdőív kitöltésének fontosságát.
- A kérdőív szerkesztésének **a kitöltést, a feldolgozást, valamint az elemzést** egyszerre kell segítenie.

- A kérdőíven szerepeljen az adatfelvétel címe vagy témája, a referencia idő, valamint az adatszolgáltató és a kitöltő adatai. Tüntessük fel az adatgyűjtést elrendelő szerv- és/vagy adatgyűjtést végző szerv nevét, elérhetőségét, kerüljön fel a kérdőívre a jogi háttér, valamint az adatszolgáltatás kötelező vagy önkéntes jellege. Utaljunk arra, hogy milyen adatvédelmi előírások, kísérik a felmérést.
- Az önkitöltős kérdőíveket könnyen kitölthetővé tervezzük. Ennek érdekében adjunk egy kedvező első benyomást az **Elnöki levél és a felkérő levél** révén. Ha személyes vagy telefonos lekérdezéssel kell kitölteni, akkor kérdező-baráttá tervezzük a kérdőívet.
- A kérdőív legyen áttekinthető, az eltérő témakörök kérdésblokkjait érzékelhetően különítsük el, és a kérdésre vonatkozó magyarázat, kitöltési útmutató lehetőleg már a kérdéshez kapcsolódva, szükség szerint a kérdőív mellékleteként elérhető legyen. A kérdőíven belül törekedjünk a **könnyen értelmezhető magyarázatokra**. Ha a kérdések egy része nem vonatkozik minden adatszolgáltatóra, legyen egyértelmű, hogy mikor mely kérdéskör megválaszolása következik soron. Az első oldalon hagyjunk helyet a válaszadó egyéb megjegyzéseinek, és köszönjük meg az együttműködést.
- A **kérdőív segédanyagainak** (kitöltési útmutató, egyéb segédletek) elsődleges célja a válaszadói hiba minimalizálása, a kitöltéssel (kérdéssel, válaszadással) járó terhek csökkentése. A kitöltési útmutatónak tartalmaznia kell a kitöltést segítő információkat, a fogalmak meghatározásait, magyarázatait, valamint az ellenőrzést segítő algoritmusokat. Az útmutató elkészítésénél a kérdőív szerkezetét, jelöléseit, sor- és oszlophivatkozásait pontosan kell követni, a korábbihoz képest a változásokat jelölni kell. A fontos részek hangsúlyozására kiemelések alkalmazandóak.
- **Elektronikus önkitöltős kérdőívek** (CASI) esetén törekedjünk, hogy a kérdőív tartalmában azonos legyen a papír alapú kérdőívvel, ezzel biztosítva a válaszadók számára az állandóságot. Az e-kérdőíveknek segíteni kell a kitöltést az automatikus ellenőrzések, kitöltés-vezetések útján, és a meghatározásokhoz, magyarázatokhoz könnyű hozzáférést kell biztosítani.
- Elektronikus kérdőívek esetén (CASI, CAPI, CATI, stb.) a tervezésnél figyelembe kell venni különböző **szerkesztési előírásokat**, mint: képernyőn való megjelenítése, kérdőíven belüli ugrás, stb.
- Sokféle módszer alkalmazható a **kérdőív tesztelésére és értékelésére**. Hogy melyiket használjuk, sokféle tényezőtől függhet. Általában figyelembe kell venni az adatfelvétel típusát és méretét, tartalmát, korábbi adatfelvételekből átvett kérdések számát, azt, hogy egy folyamatban lévő adatfelvételtől van-e szó vagy sem, az adatgyűjtés módszerét, az ütemezést, a költségvetést és az erőforrások hozzáférhetőségét.
- Alkalmazzunk **kvalitatív tesztelési módszereket**, ha azt akarjuk megvizsgálni, hogyan reagálnak a válaszadók a kérdésekre, a kérdőívre. Ilyen módszer a fókuszcsoport, a mélyinterjú, a kognitív módszerek, pl., mint a „hangosan gondolkodó interjú”, a „parafázis” és a „magatartáskódolás”. A **fókuszcsoport és a személyes mélyinterjúk** célja a kérdések megfogalmazásának, nyelvezetének, sorrendjének és formájának tesztelése. A kognitív módszereket arra használják, hogy megvizsgálják a válaszadó gondolatmenetét a válaszadás közben, és megállapítsák, vajon megértették-e a kérdések értelmét, képesek-e rá megfelelően válaszolni. A magatartáskódolás rendszerezett, objektív eszköz a kérdőív hatásának vizsgálatára a kérdező-válaszadó interakciójának megfigyelése és elemzése révén. A kvalitatív tesztelést használhatjuk még a kérdőív tartalmi kialakításánál a kulcsfontosságú fogalmak feltárásához és értékeléséhez.
- A **kérdőív előtesztelését** is szükség szerint el kell végezni, ami olyan informális vizsgálat, amely segíti a kérdések pontos megfogalmazását vagy sorrendi kialakítását, a kérdőív

külső megjelenésével, instrukcióival, vagy a válaszadó képtelenségével, vonakodásával kapcsolatos problémák feltárását. Az előtesztelést arra is használják, hogy javaslatokat tegyenek kiegészítő előre kódolható válaszkategóriákra, megbecsülik az interjú hosszát és jelezzék a válaszmegtagadási problémákat. Kérdezőkkel folytatott visszacsatoló beszélgetések követhetik az előtesztelést azzal a céllal, hogy az ő javaslataik is beépüljenek az adatfelvétel át/megtervezési folyamatába.

- **Osztott mintás tesztelés** javasolható abban az esetben, ha el kell dönteni, hogy melyik a „legjobb” alternatíva a kérdőív, illetve az adatfelvételi módszer két vagy több változata közül. Ez a módszer olyan kísérlet, amely az adatgyűjtési folyamat, valamint a kérdések megfogalmazása és sorrendje tesztelésére irányul.
- Az alapos (kvalitatív) tesztelés és előtesztelés után következhet a **próba felvétel a terepen**, amely azt vizsgálja, hogy miként működnek az adatfelvétel különböző elemei, beleértve a kérdőívvel kapcsolatos műveleteket is. A próba felvétel „kosztümös főpróbának” számít a bemutató előadás előtt, amikor a teljes adatfelvételt eljuttatjuk kicsinyített változatban a kezdetektől a végéig, egészen az adatok feldolgozásáig, elemzéséig. A próba felvétel egy utolsó lehetőség a kérdőív finomítására a végső adatfelvétel előtt.

### **Irodalom:**

KSH (2006): Kérdőív-szerkesztési ajánlás (Informatikai doktár)

1993. évi XLVI törvény végrehajtásáról szóló 170/1993.(XII. 3.) Kormányrendelet 6.§.

A SZOCIOLÓGIAI FELVÉTEL módszerei (1975). (Szerkesztette Cseh-Szombathy László, Ferge Zsuzsa), *KJK, Budapest, 1975.*

BIEMER, P. P. – LYBERG, E. L. (2003): Introduction to Survey Quality, *John Wiley&Sons Publication.*

Statistics Canada (2002): Policy on the Review and Testing of Questionnaires. *Statistics Canada Methods and Standards Committee, Statistics Canada, Ottawa.*

BABBIE, E. (1994): Társadalomkutatások módszertana. *Balassi Kiadó, Budapest.*

HOFFMAN M., KOZÁK A. és VERES Z. (szerk.) (2000): Piackutatás. *Műszaki Könyvkiadó, Budapest.*

WERDY, G. – LANGMAID, R. (1997): Kvalitatív piackutatás. *HVG Rt. Budapest.*

HANDBOOK OF RECOMMENDED Practices for Questionnaire Design and Testing in the European Statistical System. (2006). *Eurostat*

KSHXML adatgyűjtési rendszer, Kérdőívszerkesztési (KSZ) komponens, Felhasználói kézikönyv

<http://3r.ksh.hu:7778/pls/kshdok/docs/1008.DOC>



# Adatgyűjtés szervezése és adatgyűjtés

## 1. Leírás:

Az adatgyűjtésnek nevezzük azokat a folyamatokat, amelyeknek célja a statisztikai adatok létrehozása vagy a létrehozásuk elősegítése. Az adatokat egyének vagy szervezetek szolgáltatják, gyűjtésük történhet összeírás (interjú, kérdezés) és adatszolgáltatás (bevallás), vagy megfigyelés révén. Az adatgyűjtés egy megfelelő hordozó eszköz segítségével történik (lásd: a kérdőívtervezésről). Összeírás típusú adatfelvételek esetén a folyamat a terepen történő összeírást, valamint az érkeztetést foglalja magában. Az önkitöltős adatfelvételek esetén az adatszolgáltatókkal való kapcsolattartást, az érkeztetést és sűrgetést tekintjük a folyamatszakasz részének. Az adatgyűjtés szervezési és adatgyűjtési műveletek végrehajtásának mikéntje közvetlen hatással van a begyűjtött adatok minőségére és ezen keresztül a végső statisztikai termék minőségére.

Az adatgyűjtés szervezés végrehajtásakor (a felvételi folyamat tervezésének megfelelően) megtörténik a kérdőív és segédanyagainak előállítása, mintakiválasztás, expediálás, összeírók képzése, végrehajtási útmutató előállítása, stb.

## 2. Alapelvek:

Az adatgyűjtés végrehajtása során a legfontosabb, hogy megtaláljuk azokat és csak azokat az adatszolgáltatókat, akik relevánsak az adott adatgyűjtés sokasága szempontjából, valamint választ kapjunk tőlük.

Az adatgyűjtési folyamat legfontosabb szereplője az adatszolgáltató, azok a személyek és szervezetek, akik ellentételezés (fizetés) nélkül kitöltik a kérdőíveket. A folyamatos együttműködés biztosításához lényeges, hogy csökkentjük a válaszadói terheket.

Az adatgyűjtési folyamat színvonalának megőrzése és az adatminőség biztosítása érdekében használjunk megfelelő minőség és teljesítmény-mérési eszközöket. Az adatgyűjtési folyamat során az adatszolgáltatót megilletik mindazon jogok, amelyek az adatvédelmi törvényben vannak lefektetve, ezért tegyünk megfelelő lépéseket az egyedi adatok védelmében.

Az adatokban lévő hiányokat vagy inkonzisztenciákat legjobb a válaszadókkal való egyeztetés során kijavítani még az adatgyűjtés folyamán vagy az adatok rögzítése közben.

## 3. Minőségi irányelvek:

- Az **adatgyűjtés szervezés** végrehajtása során az adatszolgáltatók listájának összeállításakor (teljeskörű felvétel, reprezentatív felvétel, kombinált felvétel), a kérdőívek nyomtatásakor, a megszemélyesített kérdőívek összeállításakor, valamint az expediálás végrehajtásánál mind a határidők betartására, mind a lefedettségi hibák kezelésére, mind pedig a maximális pontosság elérésére kell törekedni.
- Az adatgyűjtési munka szervezése során **egyenletes leterheltséget kell biztosítani**. Az érkeztetés, sűrgetés során nyert fontos információk visszacsatolására kellő figyelmet kell fordítani. (pl.: az adathiányok különböző okainak visszacsatolása a következő időszakra, elérhetőség megváltozása, stb.).
- Mind az expediálás, mind a beérkezés során a kérdőívek és segédanyagainak továbbítására és kezelésére alakítsunk ki **hatékony ellenőrzési rendszert**. Biztosítsuk, hogy jogosulatlanok ne férjenek hozzá az információkhoz. Előzzük meg a rendszerhibákból vagy a humánerő hibájából eredő információvesztést.

- Alkalmazzunk **megfelelő technológiát** az adatgyűjtési folyamat hatékonyságának és minőségének biztosításához. Az adatgyűjtés gyakran az adatfelvételi folyamat legköltségesebb része, amelynek jelentős befolyása van az adatminőségre. A kommunikációs technológia és az informatikai rendszerek gyors fejlődése új lehetőségeket nyit az adatgyűjtések költségeinek csökkentésére, ugyanakkor hozzájárulnak az adatbiztonság, megbízhatóság javulásához, gyorsítják az adatokhoz való hozzáférést. A számítógéppel támogatott adatgyűjtési technikák (CASI, CAPI, CATI, stb.) jó példák azokra az új megközelítésekre, amelyek a meglévő technológiák előnyeit hasznosítják.
- Az **elektronikus adatgyűjtés** (CASI) az önkitöltős adatgyűjtések esetén az egyik leghatékonyabb (gyorsaság, pontosság, biztonság) adatgyűjtési formát jelenti. Törekedjünk az elektronikus adatgyűjtés minél szélesebb körű alkalmazására.
- Elektronikus adatgyűjtés esetén az adatszolgáltatónak adjuk meg azt a lehetőséget, hogy saját belső rendszereiből (könyvelés, számlázás, készletnyilvántartás, stb.) elektronikus formában exportált adatokkal is feltölthesse a kérdőívet.
- Az adatszolgáltatóknak minden lehetőséget meg kell adni, hogy az elektronikus úton történő adatszolgáltatást jó minőségben, biztonságosan teljesíteni tudja. Ennek eléréséhez honlapról elérhető információra, telefonos forró drótra és e-mailen igénybe vehető segítségre is szükség van.
- Az interjú típusú felvételeknél a sikeres adatgyűjtés kulcsa **az összeírók szakmai munkája**. Ezért biztosítani kell számukra az adatgyűjtéshez szükséges eszközöket, és megfelelő képzésben kell részesíteni őket.
- Az adatgyűjtési folyamat kézbentartására és ellenőrzésére állítsunk fel egy monitoring rendszert. Ez lehetővé teszi, hogy az adatgyűjtési irányítói minden egyes címről, kérdőívről tudják, hol, milyen minőségi szinten tart az adatgyűjtés folyamatában.
- Az adatgyűjtés folyamán az irányítók a kiadási, a teljesítmény és minőségi mutatók alapján áttervezhetik, módosíthatják a folyamatot. Kövessük **nyomon az aktuális kiadásokat** az expedálás, telefonálás, és ember-nap felhasználás tekintetében.
- A **válaszadási készség optimalizálása** és a gazdasági szervezetektől és intézményektől kapott információk minősége érdekében, a szervezetben a megfelelő emberrel kell felvenni a kapcsolatot, abban az időpontban, amikor az információ könnyen megadható, és engedjünk meg olyan módszereket és formátumot, amely kényelmes a válaszadó számára.
- Kerüljenek kidolgozásra olyan, az adatgyűjtés során alkalmazható **jó gyakorlatok**, amelyek segítségével az adatszolgáltatói terhek minimalizálhatók, a begyűjtött adatok minősége pedig javul.
- Az összeírók eredményes szakmai munkájához az eszközök és a képzés biztosítása mellett el kell érni, hogy az összeírók kötődjenek a hivatalhoz, motiváltak legyenek.

## Irodalom:

BIEMER, P. P. – LYBERG, E. L. (2003): Introduction to Survey Quality. *John Wiley&Sons Publication*.

KSH (2008): Központi adatgyűjtés szervezés, Informatikai Doktár

KSH (2005): A központosított expedálás minőségbiztosítása (Informatikai Doktár)

Központosan irányított adatgyűjtés-szervezési rendszerek (KIRASZ) S10201 projekt

[http://kshintra/intra/kiemtema/kshmpi/databank/informat/s102\\_haladasi\\_jelentes\\_sft1125.doc](http://kshintra/intra/kiemtema/kshmpi/databank/informat/s102_haladasi_jelentes_sft1125.doc)

[http://kshsps/PMA/Strategiai%20foiranyok/Az%20adatfelvelelek%20megtervezese/KIRASZ/Shared%20Documents/s10201\\_kirasz\\_elorehaladas\\_sft0518.doc](http://kshsps/PMA/Strategiai%20foiranyok/Az%20adatfelvelelek%20megtervezese/KIRASZ/Shared%20Documents/s10201_kirasz_elorehaladas_sft0518.doc)

Tájékoztató a nemzetközi turisztikai felvételek modernizációs programjának (TURA) végrehajtásáról

[http://kshsps/PMA/Shared%20Documents/SFT%20előterjesztések/tura\\_projektzaro\\_sft1215.doc](http://kshsps/PMA/Shared%20Documents/SFT%20előterjesztések/tura_projektzaro_sft1215.doc)

KSH: Adatgyűjtés szervezés – GÉSA (Informatikai Doktár)

KSH: Adatgyűjtés szervezés – LAKOS (Informatikai Doktár)

KSH: Adatelőkészítés - BLUMEN (Informatikai Doktár)

<http://3r.ksh.hu:7778/pls/kshdok/docs/index.html>

## Adatok előkészítése (rögzítés, editálás)

### 1. Leírás:

Az adatelőkészítési folyamat szorosan kapcsolódik az adatgyűjtési folyamathoz, magában foglalja a manuális editálást, az adatrögzítést, editálást. A folyamatszakasz akkor zárul le, amikor a feldolgozásra kész mikroadatállomány rendelkezésre áll. Ha az adatgyűjtés során nem kapunk megfelelő adatokat, vagy ha az adatok az adatelőkészítés első lépésekor, a manuális editálás során alkalmatlannak bizonyulnak, akkor még az adatgyűjtés részeként, adat-kiegészítés céljából másodszori kapcsolatot lehet kezdeményezni az adatszolgáltatóval. Az előzetesen validált információt ezután olyan elektronikus formába alakítják át (adatrögzítés), amely alkalmassá teszi a következő folyamatokban való felhasználáshoz. Ez az átalakítás gyakran kódolást is foglal magában, és olykor az adatok más helyszínre való eljuttatását is jelenti. Az adatrögzítés és az adatok editálásának bizonyos lépései általában párhuzamosan, egy rendszerbe építve kerülnek megvalósításra.

Az editálás ellenőrzések alkalmazása, amelyek azonosítják a hiányzó, érvénytelen vagy inkonzisztens információkat, vagy amelyek rámutatnak olyan adatrekordokra, amelyek potenciálisan hibásak. Egyes ellenőrzési eljárások logikai kapcsolatokra épülnek, amelyek a fogalmak és a változók definícióiból adódnak. Mások sokkal gyakorlatiasabb természetűek vagy statisztikai tesztek vagy folyamatok alkalmazásának eredményeként jönnek létre (pl. outlier elemzési technikák). Az editálás eszköze lehet az azonos adatfelvétel korábbi adatgyűjtésének adatai, de más külső források is használhatók.

Az editálás változatos tevékenységeket jelent, kezdve az összeírók területi ellenőrzésével, az adatgyűjtés vagy rögzítés során alkalmazott, az előzetes specifikációból eredő számítógépes jelzéseken át az újból felkeresett adatszolgáltató azonosításáig, a komplex összefüggések különböző módozatáig, az imputálási célú hibaaazonosításig és adat validálásig.

Az editálás során a javíthatatlan hibákat, vagyis az érvénytelen vagy inkonzisztens adatokat (fatal error) törlik az adatállományból annak érdekében, hogy segítsék a további automatizált adatfeldolgozást és elemzést. Az editálásnak egy másik fajtája csak jelzi a potenciális hibákat (query editálás), illetve azokat, amelyek további információ kiegészítést igényelnek.

### 2. Alapelvek:

Az adatokban lévő hiányokat vagy inkonzisztenciákat legjobb a válaszadókkal való egyeztetés során kijavítani még az adatgyűjtés folyamán vagy az adatok rögzítése közben.

Az editálás célja megtisztítani az adatokat a hibáktól, valamint alapot teremteni a felvétel következő időszakban történő fejlesztéséhez. Általában az erőforrások java az adatok tisztítására koncentrálódik.

### 3. Minőségi irányelvek:

- Az adatelőkészítési folyamat legösszetettebb tevékenységeit személyek hajtják végre. Ezért a manuális editálásra, a kódolásra és az adatrögzítésre statisztikai **minőségkontrollt** kell kidolgozni.
- Az **adatrögzítés során előforduló hibák csökkentésére** jó megoldást jelent a számítógéppel támogatott adatfelvétel (CAPI, CATI, stb.), az elektronikus adatgyűjtés (CASI). Ebben az esetben a javítások jó része az adatgyűjtés folyamatán belül végrehajtható.

- Az adatgyűjtési folyamat megtervezése során, különösen az editálás és kódolás tekintetében, **tegyük a folyamatot egységessé** minden résztvevő számára és olyan hibamentessé, amennyire csak lehet (automatizálás). Egy kis szakértői csoport is segítheti a munkatársakat a bonyolult esetek megoldásában. Az adatelőkészítés centralizálása is előnyös amennyiben csökkenti a költségeket, és egyszerűbbé teszi a szakértői tudás hasznosítását.
- Biztosítsuk, hogy minden editálás **belsőleg konzisztens** legyen (tehát nem önellentmondó).
- A rendszerbe legyen beépítve, hogy egy hiba javításával ne okozhassunk további **közvetlen vagy közvetett hibát**.
- Az editálás során a valós hibákra utaló jelzések aránya általában alacsony. Továbbá a hibák hatása nagyon eltérő lehet, különösen azoknál az adatfelvételeknél, amelyek numerikus adatokat gyűjtenek. Más szóval: nem ritka, hogy csupán néhány hiba felelős a változások többségéért. Ha az editálás során ezekre a hibákra koncentrálunk, és lemondunk a partikuláris hibák javításáról, akkor nem fog romlani jelentősen az adatok minősége. A **hiba prioritásokat** a hiba típusa, gyakorisága valamint az adott változó fontossága alapján határozhatjuk meg.
- Az editálás hasznossága korlátozott, és a folyamat valójában ellen-termék lehet. Gyakran az adatváltozásokat hibásan adatkorrekcióknak tekintik. A **túlzott javítás** – az adatszolgáltatóval való egyeztetés nélküli korrekció – során eljön egy olyan pont, amikor majdnem ugyanannyi hiba kerülhet be a rendszerbe, mint amennyit kijavítottak. Ezért el kell kerülni a túlzott editálást.
- Elektronikus önkitöltős kérdőívek (CASI) esetén az adatszolgáltatói oldalon történjenek meg az alapvető, a kérdőívben belül végrehajtható ellenőrzések.
- Tanuljunk az editálási folyamatból, és a hibák csökkentése érdekében inkább a megelőzést alkalmazzuk, mint az utólagos tisztítást. Tehát **a hiba megelőzését gyakoroljuk inkább, mint a hiba javítását**. (kérdőívtervezés, kitöltési útmutató, stb.) Ennek érdekében az editálási lépést hozzuk előre az adatfelvételi folyamatban, leginkább akkor, amikor a válaszadó még elérhető, pl. CAPI, CATI, CASI módszerű adatgyűjtéseknél.
- Az editálás nem tud felfedni kicsi, szisztematikus hibákat, amelyek konzisztensen fordulnak elő ismételt adatfelvételekben, és olyan hibákat, amelyek komoly becslési torzításokhoz vezethetnek. Használjunk **más módszereket**, mint pl. a hagyományos minőségellenőrzési módszerek, a fogalmak és definíciók gondos elemzése és felülvizsgálata, utólagos kikérdezéses módszerek stb.
- Az **adatszolgáltató újbóli megkeresésekor** fontos, hogy rendelkezésére álljon az általa szolgáltatott információ, mely szükséges a hibás adatok korigálásához.
- Az editálási folyamat gyakran nagyon összetett. Ezért alkalmazzunk részletes és naprakész eljárásokat, megfelelő képzéssel egybekötve mindazok számára, akik részt vesznek a munka ellenőrzésében.

## Irodalom:

BIEMER, P. P.. – LYBERG, E. L. (2003): Introduction to Survey Quality, *John Wiley&Sons Publication*.

KSH (2003): Általános adatelőkészítő rendszer (*Informatikai Doktár*)

KSH (2003): Ellenőrzések leírása az ADEL-META rendszerben (*Informatikai Doktár*)

UNECE Statistical data editing

<http://www.unece.org/stats/archive/04.04.e.htm>

GLOSSARY OF TERMS ON STATISTICAL DATA EDITING, UNITED NATIONS  
Geneva, 2000

<http://www.unece.org/stats/publications/editingglossary.pdf>

EVALUATING EFFICIENCY OF STATISTICAL DATA EDITING: GENERAL  
FRAMEWORK, UNITED NATIONS Geneva, 2000

<http://www.unece.org/stats/publications/editingefficiency.pdf>

## Imputálás (pótlás)

### 1. Leírás:

Az imputálás az az eljárás, amely a rögzítés, editálás során azonosított hiányzó hibás, vagy inkonzisztens válaszokkal kapcsolatos problémák kezelésére szolgálhat. Ez azáltal érhető el, hogy a hiányzókat pótoljuk annak érdekében, hogy a belsőleg konzisztens rekordokhoz juthassunk. Némely probléma már korábban, a válaszadóval való kapcsolatfelvétel során kiküszöbölhető (nemválaszolás csökkentése), vagy pedig a kérdőív kézi feldolgozása során tisztázható. Általában viszont lehetetlen tisztázni minden kérdést ebben a korai fázisban, hiszen tekintettel kell lennünk a válaszadói teherre, a felmerülő költségekre, valamint az időbeliség követelményére. Szükség van ezért imputálásra a további feldolgozási lépések során, hogy kezelni tudjuk a megmaradó hibákat, hiszen kívánatos, hogy egy teljes és konzisztens állományt hozzunk létre, mely pótolta adatokat is tartalmaz.

Osztályozhatjuk az imputálási módszereket aszerint, hogy sztochasztikusak, vagy determinisztikusak, attól függően, hogy van-e valamilyen szintű véletlen hatás beépítve az imputált adatokba. A **determinisztikus módszerek** magukban foglalják a **logikai imputálást**<sup>3</sup>, a **történeti imputálást**<sup>4</sup>, az **átlaggal való pótlást**<sup>5</sup>, a **hányados**-<sup>6</sup> és **regressziós imputálást**<sup>7</sup>, valamint (bizonyos keretek között) a **legközelebbi szomszéd módszert**<sup>8</sup>. Ezek a módszerek tovább bonthatók azon eljárásokra, melyek az imputált értékre pusztán a nemválaszolókról rendelkezésre álló adatok, valamint bizonyos segédfinformációk alapján következtetnek (logikai, valamint történeti imputálás). A determinisztikus módszerek másik csoportjába tartoznak azon eljárások, melyek más válaszoló egységeknek az adott adatfelvétel körében megfigyelt adatait is felhasználják. Az aktuálisan megfigyelt adatok felhasználása történhet úgy, hogy közvetlenül átveszünk adatokat egy kiválasztott megfigyelési egység (ún. donor) rekordjából (**hot deck**<sup>9</sup>, illetve legközelebbi szomszéd módszerek); valamint úgy is, hogy bizonyos modellekre támaszkodunk (hányados és regressziós eljárások). A **sztochasztikus módszerek** közé tartozik a hot deck, illetve a regressziós módszer **véletlen reziduumokkal**<sup>10</sup>, valamint bármilyen más determinisztikus eljárás, melynél véletlen reziduumokat alkalmazunk, továbbá a legközelebbi szomszéd módszere, ahol véletlenszerűen választunk számos „legközelebbi” szomszéd közül.

### 2. Alapelvek:

Bár az imputálás javíthatja a végső adatok minőségét azáltal, hogy a hiányzó, hibás, vagy inkonzisztens adatokat korrigálja, nagyon oda kell figyelni arra, hogy milyen pótlási módszert

<sup>3</sup> Pl. ha tudjuk, hogy valaki 16 éves, akkor következtethetünk arra, hogy nincs jogosítványa.

<sup>4</sup> Korábbi felvételek adatait felhasználva következtetünk a hiányzó adatokra. Lásd pl. hányados-módszer!

<sup>5</sup> Legtöbbször segédfinformációk hiányában alkalmazott egyszerű eljárás. Hamis varianciacsökkentő hatása van!

<sup>6</sup> Pl. az előző időszakban megfigyelt adatok hányadosát vetítjük ki a tárgyidőszakra.

<sup>7</sup> Ha valamely változó értékei ismertek minden egyedre, valamely másik változó viszont némely esetben nem ismert, akkor ez utóbbi változó hiányzó értékeinek pótlására alkalmazhatunk egy (pl. lineáris) regressziós modellt. Sikeresége attól függ, hogy mennyire jól illeszkedik a regressziós modell!

<sup>8</sup> Valamilyen alkalmas távolságfüggvényt választva meghatározzuk a hiányzó adattal rendelkező egyed távolságát a többi egyedtől, majd a legközelebbi egyedet használjuk donorként. Ehhez találnunk kell, néhány olyan változót, melynek az értéke minden egyedre rendelkezésre áll. A távolságfüggvényt ezek felhasználásával számíthatjuk, ez lehet hagyományos, vagy súlyozott Euklideszi távolság.

<sup>9</sup> Valamilyen, pl. területi szempont szerint sorbarendezzük az egyedeket, ha hiányzó értékkel találkozunk, akkor a sorban legutóbbi kitöltött adathellyel rendelkező egyedet használjuk donorként. (Ez a legegyszerűbb séma, számos összetettebb változata van ennek a módszernek.)

<sup>10</sup> Egy (általában normális eloszlású) véletlen hibtagot adunk a használt imputálási módszer által kapott értékhez, ezáltal biztosítva, hogy az imputálás lehetőleg minimális mértékben torzítsa csak el az adott változóra jellemző eloszlást, valamint a majdan számított varianciát.

alkalmazunk. Némely imputálási módszer nem őrzi meg az egyes mutatók közötti összefüggéseket, vagy torzíthatja a mutatókra jellemző eloszlások képét. Ezért számításba kell vennünk az imputálás hatását, amikor becsléseket és ehhez kapcsolódóan variancia-becsléseket készítünk.

Az imputálás akkor végezhető a legjobb eredménnyel, ha teljes hozzáférésünk van a mikroadatokhoz, valamint jól használható segédinformációkkal rendelkezünk. Az imputálás lehet automatizált, kézi, vagy a kettő kombinációja. A jó pótlási eljárás korlátozza annak a torzításnak a mértékét, melyet az okoz, hogy nem sikerült megfigyelni minden kívánt értéket. Továbbá ellenőrizhető módon nyomon követhető a folyamata, mely által utólag kiértékelhető az eredményessége, és biztosítja az imputált rekordok belső konzisztenciáját. A jó pótlási eljárások automatizáltak, objektívek, reprodukálhatóak és hatékonyak. A Fellegi-Holt-féle alapelvek (1976) szerint arra kell törekedni, hogy a lehető legkevesebb mezőn módosítsunk úgy, hogy a kiegészített rekordok minden editálási kritériumnak megfeleljenek.

### 3. Minőségi irányelvek:

- **Értékelni kell a nemválaszolás típusát.** Ez abból áll, hogy próbáljuk meghatározni, milyen segédváltozók tudják megmagyarázni a nemválaszolási mechanizmust. Ez jól használható az imputálási módszer bővítéséhez. Ilyen segédváltozókat célszerű belefoglalni az imputálási módszerbe.
- Gondosan fejlesszük és **teszteljük az imputálási módszertant.** Tanulmányozzuk, hogy az elérhető mutatók milyen minőségűek és mennyire megfelelőek a vizsgálatainkhoz. Így állapíthatjuk meg, hogy melyeket lehet majd segédváltozóként, vagy illesztő változóként felhasználni, illetve melyek segítségével alkothatunk imputálási osztályokat. Ebből a célból érdemes a téma szakértőivel konzultálni, és **modellező technikákat** alkalmazni.
- Vegyük számításba azt is, hogy **milyen típusú becslést kell készítenünk** úgy, mint állapotbecslés, index-becslés, aggregátum-becslés, kisterületi becslés, keresztmetszeti, vagy longitudinális becslés.
- Az imputált rekordnak lehetőség szerint „**hasonlítania**” kell a sikertelenül editált rekordra. Ez úgy érhető el, ha bizonyos értelemben véve minimális számú mutatót pótlunk, ezáltal pedig a lehető legtöbb válaszadói információt megőrizzük. Az alapvető feltevés itt az, hogy valószínűleg csak egy vagy két hibás adatot ad a válaszadó és nem sokkal többet, bár ez a gyakorlatban nem mindig van így.
- Némely felvételnél szükséges lehet **több különböző típusú imputálási módszer alkalmazása.** Ezt rendszerint úgy végezhetjük el, hogy felállítunk egy automatizált hierarchiát a módszerek között. Igyekezzünk korlátozni az ilyen szintek számát, óvatosan fejlesszük és teszteljük az alkalmazott módszereket a hierarchia minden szintjén. Hasonlóképpen, amikor az imputálási osztályok összevonása szükséges, óvatosan fejlesszük és teszteljük az imputálási módszereket az új osztályokra.
- Nagy felmérések esetén, annak érdekében, hogy csökkenthessük a számítási igényt, szükséges lehet a mutatók **két, vagy több fázisban történő feldolgozása**, szemben az egyszeri futással. Ezáltal bonyolulttá válhat az irányelvek pontos követése: egynél több donort használhatunk, vagy a szükséges minimális számú változónál többet pótolhatunk. Amennyiben számos válaszolási hiba van egy rekordon belül, szükségessé válhat a minimálisnál nagyobb számú változó imputálása manuálisan, vagy automatizált módon.
- A pótolta rekordoknak **belsőleg konzisztensnek** kell lenniük.
- **A pótolta értékeket meg kell jelölni.** Az imputáláshoz választott módszert, és az általa használt forrásokat világosan azonosítani kell. Az egyes rekordok mezőjéhez tartozó



imputált, illetve imputálás nélküli értékeket meg kell őrizni, értékelési célból. Az imputálás mértékét és hatását értékelni kell. Mérlegelni kell olyan eljárások alkalmazásának a lehetőségét, melyek segítségével az imputálás által okozott variancia mérhető.

- Az adatok **elemzésekor figyelemmel kell lenni az imputálás mértékére és hatására**. A használt imputálási eljárásnak komoly hatása lehet az adatok eloszlására. Általánosságban: minél jelentősebb a mértéke és a hatása az imputálásnak, annál megfontoltabbnak kell lennie az elemzőnek, miközben az adatokkal dolgozik. Ilyen esetekben egyes elemzések félrevezetőek lehetnek, amennyiben a pótoltt értékeket úgy kezeljük, mintha megfigyelt értékek lennének.
- Az imputálási módszertan fejlesztése közben érdemes észben tartani, hogy számos általános rendszer létezik, melyek különféle algoritmusok használatával képesek kezelni folytonos, vagy diszkrét változókat. Ezek a rendszerek általában egyszerűen használhatók, amint specifikáltuk az editálási szempontokat, és tartalmazznak olyan algoritmusokat is, melyek segítségével meghatározzák, hogy mely mezőket kell imputálni. Ezek a rendszerek jól dokumentáltak, és lehetővé teszik az egész imputálási folyamat utólagos ellenőrzését, így könnyen kiértékelhető az eredmény.
- Az alkalmazott imputálási eljárásról dokumentációs sémát kell kitölteni, mely tartalmazza az imputálási eljárás leírását, a felhasznált információk körét, valamint az alkalmazás feltételeit a legfontosabb mutatókra.

## Irodalom:

[http://www.statistics.gov.uk/methods\\_quality/publications.asp](http://www.statistics.gov.uk/methods_quality/publications.asp)

NSMS03: Report of the Task Force on Imputation<sup>11</sup>

NSMS28: Evaluation Criteria for Statistical Editing and Imputation<sup>12</sup>

BANKIER, M., LACHANCE, M., POIRIER, P. (1999). A generic implementation of the New Imputation Methodology. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 548-553.

FELLEGI, I.P. és HOLT, D. (1976). A systematic approach to automatic edit and imputation. *Journal of the American Statistical Association*, 71, 17-35.

KALTON, G. - KASPRZYK, D. (1986). The treatment of missing survey data. *Survey Methodology*, 12, 1-16.

KOVAR, J.G., - WHITRIDGE, P. (1995). Imputation of business survey data. *In Business Survey Methods, B.G. Cox és mások szerkesztésében, Wiley, New York*, 403-423.

KOVAR, J.G., MACMILLAN, J., WHITRIDGE, P. (1988). Overview and strategy for the Generalized Edit and Imputation System. *Statistics Canada, Methodology Branch Working Paper No. BSMD 88-007 E/F*.

LEE, H., RANCOURT, E., SARNDAL, C.-E. (2002). Variance estimation from survey data under single imputation. *In Survey Nonresponse, R.M. Groves és mások szerkesztésében, Wiley, New York*, 315-328.

---

<sup>11</sup> Imputálási eljárások áttekintése, rövid leírással, hétköznapi nyelven, kevés képlettel.

<sup>12</sup>Technikaibb jellegű anyag, különféle imputálási módszerek minőségi jellemzéséről.

Statistics Canada (2000d). Policy on Informing Users of Data Quality and Methodology. *Policy Manual*, 2.3. (Reproduced in Appendix 2).

Statistics Canada (2000a). Functional description of the Generalized Edit and Imputation System. *Statistics Canada technical report*.

EDIMBUS:

<http://edibus.istat.it>

# Súlyozás, becslés és a mintavételi hiba számítása

## 1. Leírás:

Becslés során egy ismeretlen sokasági paraméter értékét közelítjük, a teljes sokaság egy mintájára vonatkozó információk alapján. A leggyakoribb becslendő paraméterek: értékösszeg, átlag, hányados, kvantilis, regressziós együttható.

A becslések minőségére a felvétel megelőző lépésein túl jelentős hatással lehet a választott becslőfüggvény (vagy becslési eljárás). A becslés pontosságának jellemzésére leggyakrabban használt mutató a becslőfüggvény (becsült) szórása. A mintára és a mintavételre vonatkozó adatokon kívül a becsléshez felhasználhatunk egyéb ismert sokasági paramétereket is, a szórás csökkentése érdekében.

A becslésekhez általában a mintaelemek megfigyelt jellemzőin túl a hozzájuk kötődő súlyváltozót is felhasználjuk. A súlyozás a minta súlyok (vagy design súlyok) kiszámításának és – szükség esetén – azok módosításának, átsúlyozásának folyamata. A design súly a mintaelem bekerülési valószínűségének reciproka. Meghiúsulás, mérési hibák, egyéb nem mintavételi hibák okozta torzítások csökkentése érdekében általában szükséges a minta súlyok módosítása. Becslések szórásának csökkentése érdekében is gyakori az átsúlyozás (pl. kalibrálás, outlier súlyok használata, utólagos rétegzés), ami az előzőtől nem feltétlenül elkülönülő lépés.

## 2. Alapelvek:

A becslőfüggvények (becslési eljárások) jósága jellemezhető azok torzításával és szórásával. A használni kívánt függvény (eljárás) kiválasztásánál a legfontosabb szempont, hogy az, lehetőség szerint **minimális torzítású és szórású** legyen, illetve az átlagos négyzetes hibája legyen minimális, ami a torzítás négyzetének és a szórásnégyzetnek az összege.

Design konzisztencia egy további kívánatos tulajdonsága a becsléseknek.

A mintavételi terv meghatározza a design súlyokat. A becslőfüggvény (eljárás) megválasztása az esetek egy részében egyben átsúlyozást is jelenthet, mint pl. kalibrálás esetén. Amennyiben a nemválaszolás kezelésére a kalibrálástól külön átsúlyozás történik, úgy célszerű minél hatékonyabb változókat felhasználó nem-válaszolási modell felállítása. Egy megfelelően megválasztott modell jelentősen csökkentheti a torzítást.

## 3. Minőségi irányelvek:

- Panel felvételek esetén az elemzési céloknak megfelelően súlyozásnál kétféle súlyokat lehet előállítani: panel súlyokat és keresztmetszeti súlyokat, melyek általában különböznek egymástól.
- Komplex mintavételi tervek és nem-lineáris becslőfüggvény esetén nem mindig adható analitikus módon **becslés a mintavételi hibára**. Ekkor közelítést kell alkalmazni: közelíthetjük a becslőfüggvényt (linearizálás), illetve kevésbé komplex mintavételi tervet feltételezve dolgozhatunk (figyelve, hogy az eredetitől minél kisebb mértékben térjünk el). Megoldás lehet az ismételt mintavételen alapuló becslések alkalmazása.
- A mintavételi hiba számításánál vegyük figyelembe **az imputálás és kalibrálás hatását**.
- A becslések megbízhatóságának növelése érdekében **segédadatok használata** javasolt, amikor csak lehet. (Hányados becslés, regressziós becslés)

- **Kalibrálással** egyrészt csökkenthető a nem-válaszolásból származó torzítás, másrészt megfelelően megválasztott változók esetén jelentős javulás várható a becslések pontosságában. Figyelemmel kell lenni arra, hogy a kalibrálás az eredeti súlyokat lehetőleg minél kisebb mértékben módosítsa.
- A kalibráláshoz olyan sarokszámokat használjunk, melyek meghatározásánál, előállításánál használt fogalmak megfelelnek a súlyozandó adatok előállításánál használt fogalmaknak.
- A kiugró értékek, **outlierek** jelenléte a megvalósult mintában jelentősen ronthatja a becslések pontosságát. Robusztus becslőfüggvények alkalmazásával ezek kezelhetők.
- Az alkalmazott becslésekről a dokumentációs sémát minden mintavételes felvétel indulásakor ki kell tölteni, illetve szükség esetén frissíteni kell.

### **Irodalom:**

COCHRAN, W. G. (1977). Sampling Techniques. Wiley, New York.

SARNDALI, C. E., SWENSSON, B., WRETMAN, J.H. (1992). Model Assisted Survey Sampling. Springer-Verlag, New York.

KISH, L. (1995). Survey Sampling. Wiley, New York

DAVISON, A.C. – HINKLEY D. V. (1997). Bootstrap Methods and Their Application. Cambridge University Press, Cambridge

SHAO, J., TU, D. (1995). The Jackknife and Bootstrap. Springer, New York

HUNYADI, L. (2001). A mintavétel alapjai. SZÁMALK Kiadó, Budapest

HORVÁTH B. – MIHÁLYFFY L. (2008.) Hibaszámítás jackknife módszerrel bonyolult felépítésű, kalibrált minták esetén. Statisztikai Szemle. 86 (6)

[http://portal.ksh.hu/pls/portal/docs/PAGE/STATSZEMLE/STATSZEMLE\\_ARCHIVUM/2008\\_ARCHIVUM/2008\\_06/MIHALYFFY.PDF](http://portal.ksh.hu/pls/portal/docs/PAGE/STATSZEMLE/STATSZEMLE_ARCHIVUM/2008_ARCHIVUM/2008_06/MIHALYFFY.PDF)

# Indexszámok képzése

## 1. Leírás:

Az indexszámok nem homogén statisztikai sokaságok (pl. a termelés, fogyasztás, termelékenység, árak, volumenek, stb.) időbeni változását, térbeni eltérését, vagy egyéb tulajdonságát, viszonyát mérik. Lényegüket tekintve átlagok, amik sokszor jelentős eltéréseket takarnak. Általában, összesíthető jelenségekről van szó. Utalni kell azonban arra, hogy amikor valaminek, vagy valamiknek a változásáról beszélünk, akkor más tényezőket változatlanok tekintünk.

Az indexek nagyon fontos szerepet játszanak a társadalmi és gazdasági folyamatok elemzésében, bemutatásában. Például az infláció, a gazdasági növekedés, a munkanélküliség vagy éppen tőzsdeindex alakulása a gazdasági folyamatok megítélésében meghatározó.

Egy-egy index kiszámításánál két állapotot hasonlítunk össze, amihez valamilyen „modellre” van szükség.

Sokféle változást jelző index van. A folyamatok leginkább értékük alakulásával írhatók le. Az érték azonban az árak és mennyiségek változásától függ. A leggyakoribb eset az, amikor árindexeket számítunk, és a reálfolyamatokat azok segítségével tárjuk fel.

Jól ismert az az összefüggés, hogy a V értékindex a P árindex és a Q volumenindex szorzata. Így az ismert értékindexből az árindex segítségével ki tudjuk a volumenindexet számítani, vagy fordítva.

## 2. Alapelvek:

Az indexeknek, mint reprezentatív felvételtől adódó mennyiségeknek kiszámítható a mintavételi hibája, amint erre a magyar és a nemzetközi szakirodalomban is található példák. Bemutatható a meghíúsulások, az imputálások száma, ill. azok változása. Nem jellemző azonban az, hogy ezeket az információkat a rendszeres publikálás során közölnék. Inkább a módszertani ismertetések, vizsgálódások, valamint a revíziók során, szakmai körökben kerülnek felhasználásra.

E mutatószámok minőségét tehát elsősorban a módszerek elfogadottsága, közismertsége, egyeztetettsége határozza meg. Ezt jelzi az is, hogy azok publikációjuk után hivatalos és széles körben használt adatokká válnak, amelyek utólagos javítása, korrigálása nem lehetséges, hacsak nem eleve előzetes adatokról volt szó. Ennek ellenére a folyamatos módszertani fejlesztéshez a szakértőknek sokrétű informáltságra van szükségük. Csak így biztosítható a mintavételi tervek, a minőségváltozás mérési módszereinek fejlesztése

## 3. Minőségi irányelvek:

- Az indexszámok a **felhasználói igények kielégítése** érdekében kötött kompromisszumok során kialakított bonyolult adatgyűjtési és becslési eljárások eredményeiként adódnak. Nagyon fontos mutatókról van szó, amelyek minőségét, elfogadottságát valójában felhasználhatóságuk határozza meg (fit for use). Nézzük át röviden a minőség összetevőit.
- Ki kell elégíteni a sokféle felhasználó többirányú igényeit. Biztosítani kell az infláció mérőszámát, a különböző gazdasági és társadalmi folyamatok vizsgálatát (GDP-deflátor, a reálfolyamatok mérhetősége, külső és belső hatások elkülönítése, társadalmi- gazdasági rétegek, stb.). A számítási módszereket az **érdekeltek bevonásával és a nyilvánosság biztosításával** rendszeresen felülvizsgálni, módosítani kell.

- Az árindexek képzésénél felmerülő **minőségváltozások kezelésére** számos módszer ismert: korrekció, helyettesítés, arányosítás, hedonikus – regressziós - technikák, stb.
- Az árindexek által vizsgált folyamatokat **időben nyomon kell követni**. (Térbeli eltérések vizsgálata – pl. országok közötti vásárlóerő-paritás – hasonló módon történhet.) Több évet átfogó árindexeket is számítani kell.
- Meg kell vizsgálni, hogy a kiszámított árindexek az árszínvonal változásának jelzése mellett, alkalmasak-e **értékindex deflálására**.
- Bár különböző mértékben, de minden index esetében fontos a **gyorsaság**. Annak kielégítése csak számos kompromisszum árán lehetséges.

### **Irodalom:**

HUNYADI L., VITA L. (2002): Statisztika közgazdászoknak. *KSH, Budapest*.

KÖVES P. (1981): Indexelmélet és közgazdasági valóság. *Akadémiai Kiadó, Budapest*.

KSH (2000): A fogyasztóiár-statisztika módszere. *Statisztikai Módszertani Füzetek 39.sz KSH, Budapest*.

MARTON Á. (2003): A fogyasztóiár-index koncepcionális és gyakorlati kérdéseiről. *Statisztikai Szemle 81. évf.12. sz. 1107-1113 old.*

MARTON Á. (2004): Az árindexek minőségét befolyásoló tényezőkről. *Statisztikai Szemle 82. évf. 9. sz. 859-865 old.*

ALLEN, R.G. D. (1975): Index numbers theory and practice. *London*.

BALK, B. (1995): Axiomatic price index theory: a survey. *Internatinal Statistical Review. 63. évf. 1. sz. 69-93 old.*

DIEWERT, W.E. – NAKAMURA, A. O. (1993): Essays in Index Number Theory. *Amsterdam: North-Holland*.

DIEWERT, W. E. (1998): Index Number Issues in the CPI, *Journal of Economic Perspectives, 12: Winter*

FORSYTH, F. G. – FOWLER, R. F. (1981): The Theory and Practice of Chain Price Index Numbers. *Journal of the Royal Statistical Society, sr. A, 144 224-246 old.*

GORDON, R. J. (1990): The Measurement of Durable Goods Prices. *University of Chicago Press*.

HANDBOOK ON PRICE and volume measures in national accounts (2001). Theme 2. General statistics. *European Comission*

MOULTON, B. R. (1996): Bias in the Consumer Price Index: What is the Evidence? *Journal of Economic Perspectives, 10: Fall*

# Makrovalidálás

## 1. Leírás:

A **makrovalidálás** aggregált adatok összehasonlítása, ellenőrzése és javítása más forrásokból (felvételekből, adminisztratív adatforrásokból, mérlegekből) származó adatok felhasználásával. Az ellenőrzések tipikusan grafikus vagy numerikus formulán alapuló modellekre épülnek.

Feldolgozáson belüli makrovalidálás: a makrovalidálás ezen típusa során még a teljeskörűsítést megelőzően kerülnek bizonyos aggregátumok ellenőrzésre. Az ellenőrzés során az adott felvétel korábbi adatai, adminisztratív adatok kerülnek felhasználásra. Ez a lépés szorosan beintegrálódik a feldolgozási folyamatba. A gyanúsak talált aggregátumoknál a hibák feltárása és javítása mikroszinten történik.

Feldolgozást követő makrovalidálás: a makrovalidálás ezen típusánál a teljeskörűsítést követően kerülnek bizonyos aggregátumok összehasonlításra. Az ellenőrzés során más adatgyűjtések hasonló mutatókörének aggregált adatai kerülnek felhasználásra. Az esetleges eltérések okainak feltárása ezek után az érintett felek közös feladatai. Az okok megkeresése során fontos az adatfelvétel módszertana (pl: imputálás, teljeskörűsítés, stb.) közötti különbségek vizsgálata, mivel az eltérések származhatnak ezekből. A feltárt hibák javítása általában mikroszinten történik.

Nemzeti számlák összeállítása során történő makrovalidálás: a makrovalidálás ezen típusa jelenti a legfelsőbb szintet, és ez van időben a legtávolabb az adatfelvétel időszakától. A validálásra kerülő aggregátumok száma bővebb, mint a korábbi szinten. Ez a típus elkülönül a feldolgozási folyamattól, az eredményeit azonban közvetlenül kell visszacsatolni az érintett folyamatokba. A makrovalidálás ezen a szinten az eltéréseken túl rámutathat az esetleges módszertani, kompetenciabeli hiányosságokra is. A feltárt hibák javítása a szakstatisztikákért felelősek feladata.

## 2. Alapelvek:

A makrovalidálás során az olyan típusú hibák kiszűrésére kell törekedni, amelyeket mikroszinten nem lehet azonosítani. Az összehasonlítás során mindig figyelembe kell venni, hogy használt, más forrásból származó adatok valóban összehasonlíthatóak a vizsgált adatokkal. Az összehasonlításra alkalmazott módszereket mindig a célnak megfelelően kell megválasztani. (pl.: modell alapú összehasonlítás, idősorok vizsgálata, stb.) Mindig olyan külső adatforrást kell használni, amely a vizsgált adatokhoz képest pontosabb.

## 3. Minőségi irányelvek:

- A makrovalidálás során a **legfontosabb aggregátumokat** kell elsődlegesen validálni, mivel ha túl sok ismérv mentén javítjuk az adatokat mikroszinten, az összefüggések elveszhetnek.
- A külső forrás megválasztásánál alaposan meg kell vizsgálni a fogalmak, osztályozások közötti különbségeket, mivel az eltérések adódhatnak abból, hogy az aggregált adatok nem **összehasonlíthatóak**.
- A validáláshoz mindig **olyan adatforrást használjunk** (más felvételek adatai, adminisztratív adatforrások), **melyek pontosabbak, megbízhatóbbak**. Ahhoz, hogy az

adatok minél inkább megbízhatóak és pontosak legyenek minél több külső adatforrást kell felhasználni a validálás során.

- A makrovalidálás során feltárt **hibák okát minden esetben keressük meg**, és szükség szerint orvosoljuk. A hibák oka lehet a mikroadatokat pontatlansága, vagy a felvételi folyamatban használt módszerek, eljárások nem megfelelősége.

### **Irodalom:**

LEOPOLD GRANQUIST (1991): Macro-editing

<http://www.eustat.es/prodserv/datos/vol0024.pdf>

GLOSSARY OF TERMS ON STATISTICAL DATA EDITING. (2000). *United Nations Statistical Commission and Economic Commission for Europe*.

<http://www.unece.org/stats/publications/editingglossary.pdf>

F. LAFLAMME, C. BARETT L. RAMSEY (1996): Experiences in Re-engineering the Approach to Editing and Imputing Canadian Imports Data

KATHERINE J. THOMPSON AND LAURA OZCOSKUN (2007): An Empirical Investigation into Macro Editing, Office of Statistical Methods and Research for Economic Programs, U.S. Census Bureau

[www.fcs.gov/07papers/Thompson.III-B.pdf](http://www.fcs.gov/07papers/Thompson.III-B.pdf)

TON DE WAAL (2008): An overview of statistical data editing. Statistics Netherlands, The Hague/Heerlen



## A szezonális kiigazítás

### 1. Leírás:

A szezonális kiigazítás célja a szezonális és naptári hatások, ingadozások kiszűrése az idősorokból, annak érdekében, hogy a vizsgálandó jellemzőkről tisztább képet kapjunk. A szezonális illetve naptári hatások olyan tényezők, melyek különböző évek azonos időszakában (hónap, negyedév) azonos irányban és közel azonos mértékben hatnak az idősor alakulására, mint például az időjárás, az adott időszakban ledolgozott munkanapok száma, társadalmi, vallási hagyományok (különböző ünnepnapok), stb. Az igazítás bázis évhez viszonyított indexekre, egy év adott időszakához viszonyított indexekre, illetve abszolút számokra történik. Az előállított legfontosabb adatsorok: a szezonálisan kiigazított adatok, a trend, és a naptári hatásokkal kiigazított adatok (ha szükséges).

### 2. Alapelvek:

A szezonális kiigazítás akkor megfelelő, ha a kiigazított idősorban már nem marad szezonális hatások. Ugyanakkor szezonális hatást nem tartalmazó idősort nem szabad szezonálisan illetve naptári hatással kiigazítani.

A szezonális kiigazítás során kapott adatok minden egyes alkalommal visszamenőleg módosulnak, ahogy az idősor újabb és újabb megfigyelésekkel egészül ki. Törekedni kell rá, hogy a közölt kiigazított adatok minél kevésbé módosuljanak visszamenőleg, azaz a revízió kicsi legyen, ugyanakkor minél kevesebb információt veszítsünk.

A trend, a szezonálisan kiigazított adatok és a naptári hatásokkal kiigazított adatok egymással konzisztens módon készüljenek. Például: nem képzelhető el, hogy a trend számítása más módszerrel történik, mint a szezonálisan kiigazított adatoké.

Az egymással szoros tartalmi kapcsolatban lévő idősorok igazításánál ügyeljünk arra, hogy a kapott eredmények egymással konzisztensek legyenek.

A szezonális kiigazítás során a megfelelő időben rendelkezésre álló, megalapozott szakértői információkat használjuk fel, a kiigazító program beállításainál ezeket vegyük figyelembe. Az eredmények értelmezhetőségét mindig tartsuk szem előtt.

Igény esetén a szezonális kiigazításhoz szükséges beállításokat a felhasználók rendelkezésére kell bocsátani.

Az időbeli konzisztencia biztosítása, nemzetközileg ajánlott módszer alkalmazásával, felhasználói nyomás esetén elégíthető ki.

Az aggregátumok szezonális kiigazítása direkt eljárással történik.

Az adatok nemzetközi összehasonlítása érdekében, az Európai Statisztikai Rendszer szezonális kiigazításra vonatkozó ajánlásait figyelembe kell venni.

### 3. Minőségi irányelvek:

- A szezonális kiigazításhoz a **TRAMO/SEATS** módszert használjuk, egységesen a **Demetra szoftver 2.04-es verziója** segítségével. A szoftver-frissítéseket nyomon kell követni.

- A revízió csökkentése érdekében az idősorokra a modellt és a **paramétereket évente egyszer rögzítjük, törekedve azok automatikus kezelésére.** Év közben ezekkel a beállításokkal történik a kiigazítás. Évközbeli újra rögzítés akkor lehetséges, ha vagy az alapadatok jelentősen módosulnak, vagy a kiigazítás minőségét statisztikai szempontból jellemző diagnosztikák romlása ezt indokolja. Ilyen esetben is, ha lehetséges, igyekszünk a paraméterek újrabecslésével megoldani a problémát, a modell változatlanul tartása mellett. A revíziók mértékét folyamatosan nyomon kell követni.
- A naptári hatásokat minden esetben tesztelni kell, kivéve, ha szakértői információ alapján nyilvánvaló, hogy ilyen hatással az idősor esetén nem kell számolni. A naptári hatásoknál meg kell vizsgálni a **húsvét-hatás, a szökőnap-hatás és a munkanap-hatás** lehetőségét. A munkanap-hatásnál rövid idősorok esetében csak a hétköznap-hétféve megkülönböztetést vesszük figyelembe, hosszabb idősorok esetében a hét különböző napjainak különböző hatást tulajdoníthatunk.
- Az **outliereket minden esetben tesztelni kell.** Három fajta outliert veszünk figyelembe: additív outlier, csillapodó törés és szint eltolódás. Szakértői információkat figyelembe veszünk az outlierek beállításánál. Különösen fontos a szakstatisztikus véleménye az idősorok végén jelentkező, statisztikailag rendszerint bizonytalan típusú outlierek esetében.
- A gazdasági válságból adódó helyzetet, valamint egy esetlegesen több területet érintő változást a hivatalon belül egységesen kell kezelni. Bizonytalan gazdasági, társadalmi helyzetben is alapvető cél, hogy a fenti elveknek is megfelelően mind a megalapozott szakmai érveket, mind a matematikai statisztikai elvárásokat figyelembe véve a program beállításait a lehető legkisebb mértékben módosítsuk.
- **Aggregátumokat a direkt igazítás módszerével igazítunk,** azaz az alágazatok mellett az aggregátumot is külön kiigazítjuk. Ilyenkor az alágazatok és az aggregátum közötti additivitás nem teljesül automatikusan, a különbség szétosztásával az additivitás igény szerint helyreállítható.
- Amennyiben az idősorban **nincs szezonális**, szezonálisan kiigazított adatsorként az eredeti idősort kell közölni. Amennyiben nem jelentkezik naptári hatás, akkor naptári hatással kiigazított idősorként az eredeti idősort kell közölni.
- A szezonális kiigazítás minősége érdekében valamennyi, a program által felkínált diagnosztikát, grafikus ellenőrzési eszközt figyelembe kell venni.
- A szezonális kiigazítás folyamatára vonatkozó dokumentációs sémát rendszeresen, a változásoknak megfelelően ki kell tölteni.

### **Irodalom:**

Szabályzat: XXXII/2005. (SK 2.)KSH előírás a KSH egységes szezonális kiigazítási gyakorlatáról.

BAUER P. – FÖLDESI E. (2003). Észrevételek az idősorelemzési módszerek alkalmazásával kapcsolatos kérdésekhez. *Statisztikai Szemle*, 81. évf. 9. szám, szeptember, 826-831. p.

BAUER P. – FÖLDESI E. (2004). A szezonális kiigazítás harmonizációja a Központi Statisztikai Hivatalban. *Statisztikai Szemle*, 82. évf. 8. szám, augusztus, 691-704. p.

Eurostat (1998a). Seasonal Adjustment Methods – A Comparison for Industry Statistics, revised version. *Luxembourg*.

Eurostat (1998b). Seasonal Adjustment Policy – Some Eurostat Proposals. *SAM 98 Seminar, 22-24 October, Bucharest*.

Eurostat (2007). Seasonal Adjustment Methods and Practices. [http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/SEASONAL\\_ADJUSTMENT\\_METHODS\\_PRACTICES.pdf](http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/SEASONAL_ADJUSTMENT_METHODS_PRACTICES.pdf)

Eurostat – ECB (2008). ESS Guidelines on Seasonal Adjustment.

FISCHER, B. (1995). Decomposition of Time Series – Comparing Different Methods in Theory and Practice. *Luxembourg*.

FRISS P. (2003). Kérdések az idősor-elemzési módszerek alkalmazásáról. *Statisztikai Szemle*, 81. évf. 7. szám, július, 588-595. p.

GÓMEZ, V. – MARAVALL, A. (1996). Programs TRAMO (Time series Regression with Arima noise, Missing observations, and Outliers) and SEATS (Signal Extraction in Arima Time Series). *Instructions for the User. Working Paper 9628, Servicio de Estudios, Banco de España*.

OECD (2002). Harmonising Seasonal Adjustment Methods in European Union and OECD Countries. *Short-term Economic Statistics Expert Group – Meeting, 24-25 June, Paris*.

SUGÁR A. (1999a). Szezonális kisimító eljárások összehasonlítása. *Gazdasági Minisztérium, Gazdaságelemző Intézet. Budapest*.

SUGÁR A. (1999b). Szezonális kiigazítási eljárások (I.). *Statisztikai Szemle*, 77. évf. 9. szám, szeptember, 705-721. p.

SUGÁR A. (1999c). Szezonális kiigazítási eljárások (II.). *Statisztikai Szemle*, 77. évf. 10-11. szám, október-november, 816-832. p.

Szezonális kiigazítás (2005). *Statisztikai módszertani füzetek*, 43, KSH, Budapest.

<http://www.ksh.hu/pls/ksh/docs/hun/xftp/idoszaki/pdf/szezonkiig.pdf>

FÖLDESI E. – BAUER P. – HORVÁTH B. – URR B. (2007). Seasonal adjustment methods and practices

[http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/SEASONAL\\_ADJUSTMENT\\_METHODS\\_PRACTICES.pdf](http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/SEASONAL_ADJUSTMENT_METHODS_PRACTICES.pdf)

## Elemzések készítése

### 1. Leírás:

Az adatelmezés a munkafolyamat azon szakasza, mely során az alapadatokból felhasználható információt állítunk elő. Az elemzések általában jelentések, cikkek formájában kerülnek megjelentetésre (elemző statisztikai kiadványok). Az elemzési folyamat legfőbb lépései a következők: megvizsgálni az adatokat, lényeges kérdéseket feltenni, a kérdéseknek megfelelő választ adni, és ezt a felhasználó felé érthető módon közvetíteni.

Az elemzés az egyik legkritikusabb szakasza a folyamatnak a minőségre gyakorlott hatás szempontjából. Az elemzés segítségével pontosabb képet kaphatunk a korábbi szakaszok minőségéről, biztosítva a megfelelő visszacsatolást. Egy relatíve rossz minőségű adat jó elemzése gyakran rosszabb, mint egy jó minőségű adat gyengébb elemzése.

Az elemzés ezek mellett segíthet a minőséggel kapcsolatos problémák feltárásában, valamint a felvételek továbbfejlesztésénél.

### 2. Alapelvek:

A hatékony adatelemzésnek a kérdésekre kell fókuszálnia, nem pedig az adatokra. Az elemzéseknek arra kell választ adniuk, hogy a nyers adatokból kiindulva milyen segítséget kaphatnak a felhasználók döntéseik meghozatalánál.

A felhasználók különböző szintű tudással rendelkezhetnek mind a társadalmi-gazdasági kérdésekkel kapcsolatban, mind pedig a statisztikai elemzési módszerek tekintetében. Ezért nemcsak a kapott eredmények részletes bemutatása, hanem az alkalmazott módszerek ismertetése is része lehet az elemzéseknek.

A felhasználótól függően, érdemes mérlegelni, hogy a kevésbé releváns részeket és az elemzés fő mondanivalóját milyen arányban közvetítjük.

### 3. Minőségi irányelvek:

- A következő kérdéseket kell feltenni: **Milyen kérdésre** szeretnénk választ kapni? **Milyen adatok** szükségesek, melyek állnak rendelkezésre? **Milyen matematikai statisztikai módszer** alkalmas a szükséges elemzés elkészítéséhez? A kapott **eredmények megfelelő módon** vannak-e prezentálva? Ezek alapján három fő részre lehet osztani az elemzési tevékenységet: érdekes/lényeges kérdések (bevezetés), a kérdések megválaszolása a megfelelő elemzési módszerek segítségével, a megfelelő adatokon (törzs), és válasz (konklúzió).
- A **kérdés-orientált** elemzéseket kell előnyben részesíteni az adat-orientált elemzésekkel szemben.
- Az elemzések során **matematikailag megalapozott módszereket** kell alkalmazni.
- Ha a változók összehasonlítása során a statisztikai tesztek szignifikáns különbséget mutatnak ki, a különbözőség okaira próbáljunk **magyarázatot** találni.
- Amennyiben az elemzésre használt adatok egy mintavételes felvételtől származnak, akkor **a mintavételi tervet figyelembe kell venni** az elemzésnél. Az elemzésnél a mintavételi (kalibrált) súlyokat, valamint a becslés variancia-kovariancia struktúráját is figyelembe kell venni.

- Az elemzés során nemcsak a mintavételi hibát, hanem a minőség különböző összetevőinek **minőség-indikátorait** is figyelembe kell venni. Például a pontosság (nem mintavételi hibák, imputálási arányok, stb.), vagy az összehasonlíthatóság (területi, időbeli, stb.) mutatóinak segítségével információt kaphatunk az adatokban lévő különböző torzításokról.
- Amennyiben az elemzéshez több felvételtől, illetve adatforrásból származó adatot is felhasználunk meg kell vizsgálni a különböző **adatok közötti koherenciát**. (referencia időszak, célsokaság, definíciók, stb.)
- A különböző forrásból származó adatok együttes alkalmazásakor a **megfelelő adatintegrációs módszert** (pl. statistical matching) kell megtalálni az elemzés céljainak megfelelően.
- Óvatosan kell bánni az **ok-okozati összefüggésekkel**, a lehetséges ok megállapításánál az összes lehetőséget vegyük számításba, ne csak egyet.

### **Irodalom:**

HUNYADI L. - VITA L. (2002): Statisztika közgazdászoknak. *KSH, Budapest.*

HAJDÚ O. (2003): Többváltozós statisztikai számítások, *Központi Statisztikai Hivatal, Budapest.*

# Az adatok bizalmas kezelése és a felfedhetőség elleni védelem

## 1. Leírás:

Adatvédelem alatt az adatszolgáltatók adatainak bizalmas kezelését értjük, biztosítva, hogy illetéktelenek ne férjenek hozzá érzékeny adatokhoz. Az adatvédelem jogszabályokon alapul<sup>13</sup>, amelyek egyrésztől meghatározzák a védelem kereteit, másrésztől szankciókat biztosítanak a visszaélések esetére. A jogszabályi kereten felül két szempont szerint kell eleget tenni az adatbiztonságnak. Az egyik az informatikai védelem, amely az adatok eltulajdonítása ellen biztosít. Ebbe beletartozik az Interneten keresztüli támadások kivédése, valamint az is, hogy illetéktelenek ne juthassanak hozzá az adatokhoz a hivatal épületeiben. A másik szempont az adatfelfedhetőség elleni védelem biztosítása. Ez magába foglalja a közzétett adatokból (adattáblák, adatbázisok, mikroadat-állományok, stb.) felfedhető egyedi adatok védelmét.

## 2. Alapelvek:

A statisztikai hivatalok alapfeladata a tájékoztatás. A tájékoztatás kereteit, határait a statisztikáról szóló jogszabályok határozzák meg, azonban 100 %-os adatbiztonságot csak akkor lehet elérni, ha nem közlünk semmit. A legfontosabb cél tehát, hogy megtaláljuk az optimális arányt a tájékoztatásra kerülő és a rejtve maradó adatok közt. A túlzott elrejtés a felhasználó igényeit, a korlátlan tájékoztatás pedig az adatszolgáltatók jogait sérti. A felhasználók részéről az adataink iránti érdeklődést, az adatszolgáltatók részéről pedig a hivatal iránti bizalmat kell fenntartanunk, megteremtünk.

Mivel a felfedhetőség szorosan kapcsolódik a tájékoztatáshoz, ezért a védekezés módját jelentősen meghatározza a tájékoztatás helye, módja, ideje. A módszerek alkalmazhatóságát is jelentősen befolyásolják ezek.

Az egymással tartalmi kapcsolatban lévő adatoknak az idősorba való illesztésénél lehetőség szerint ügyelni kell arra, hogy a védelemmel ellátott adatállományok egymással összehasonlíthatóak legyenek.

Ahol szükséges, a felhasználókat tájékoztatni kell arról a tényről, hogy a rejtve maradt adatoknak adatvédelmi okai vannak.

Amennyiben szerződés alapján végezzük az adatátadást, abba be kell építeni az adatok védelmére vonatkozó klauzulákat.

Az adatfelfedhetőség elleni védelemnek két fő területe van, amelyek alapjaiban hasonlítanak egymásra, mégis egész más szemléletmódot igényelnek. Az egyik a mikroadat-, a másik pedig a táblázatos adatvédelem.

## 3. Minőségi irányelvek:

- Az internetes **tájékoztatáshoz szükséges adatbázis** egy, a statisztikai adat-előállítás belső rendszereitől elkülönített szerveren található, és ebbe az adatbázisba védendő adatok nem kerülhetnek bele, annak biztosítékául, hogy az esetleges Interneten keresztüli támadás esetén se lehessen az érzékeny adatokat eltulajdonítani.

<sup>13</sup> 1992. évi LXIII. Törvény a személyes adatok védelméről és a közérdekű adatok nyilvánosságáról

- A **mikroadat védelemben** alkalmazható módszerek: átkódolás, kerekítés, cellaelnyomás, csonkolás, összekeverés (adatzavarás).
- A **táblázatos adatvédelemben** alkalmazható módszerek: aggregálás, kerekítés, cellaelnyomás (dimenziókorlátozás).
- Az **informatikai biztonságról** az informatikai rendszerek üzemeltetője gondoskodik, a felfedhetőség megakadályozásáról pedig az adatgazdák.
- Az érzékeny adatok körének meghatározása jogi és szakmai probléma, és ennek megfelelően **hazai és nemzetközi jogszabályok** rendelkeznek róla. A jogszabály megmondja ugyan, hogy mi az egyedi adat, de hogy a konkrét adat ide tartozik-e, azt szakmai szempontok alapján kell eldönteni. Fontos megemlíteni, hogy a magyar jogi szabályozás nem ismeri a mikroadat fogalmát, tehát minden esetben, amikor mikroadat kiadásáról van szó, meg kell vizsgálni, hogy a konkrét adat a Statisztikai törvény és a Végrehajtási rendelkezés értelmében minek minősül és kiadható-e. A táblázatos tájékoztatás során legegyszerűbb kritérium (amelyet jogszabály is rögzít) a védendő cellára vonatkozik, hogy az adott cella nem tartalmazhat kevesebb, mint 3 adatszolgáltatót. Az érzékenység másik fő szempontja, hogy egy adatszolgáltató adata domináns-e a cellában.
- Az adatokból mindig kiszámolható, hogy mekkora **felfedési kockázatot** hordoznak magukban. Ennek mérésére különböző szabályok alkalmazhatóak (küszöb szabály<sup>14</sup>, dominancia szabály<sup>15</sup>, p-szabály<sup>16</sup>, pq-szabály<sup>17</sup>).
- Külön figyelmet kell fordítani arra is, hogy az adatokat **kinek szolgáltadjuk** és milyen célra. Tudományos műhelyeknek statisztikai célú felhasználásra lehetőség van bővebb adatállományok (anonimizált mikroadat) átadására.
- **Elemezni kell a felhasználókat** és adatkéréseiket, hogy fel tudjuk mérni, mely csoportból kerülhetnek ki azok a felhasználók, amelyek potenciális támadónak tekintendők.
- Fel kell térképezni a támadók motivációit is, amely ugyancsak alapjául szolgál a **védelmi rendszer kialakításának**.
- Az adatvédelemhez dokumentációs sémát kell kitölteni, mely tartalmazza az adatkiadás és az adatvédelem legfontosabb paramétereit.

## Irodalom:

A KSH Informatikai Biztonsági Rendszeréről szóló VI/2006. (SK 2.) KSH előírás.

A Tanács 1990. június 11-i, 1588/90 sz. rendelete (EURATOM, EGK) a bizalmas statisztikai adatoknak az Európai Közösségek Statisztikai Hivatala részére történő átadásáról

A Tanács 1997. február 17-i. 322/97. (EK) számú rendelete a közösségi statisztikákról.

<sup>14</sup> Ha az adatszolgáltatók száma egy meghatározott  $M$  küszöbértéknél ( $M \geq I$ ) kevesebb, akkor a cella érzékeny.

<sup>15</sup> Érzékenynek tekintendő a cella, ha az értékét adó válaszadók közül egynek vagy többnek (az hogy mennyire vizsgáljuk, függ a védelem kialakításának mértékétől) az adata meghaladja a cella bizonyos százalékát (nemzetközi gyakorlatban ez 80-90%).

<sup>16</sup> Ez a szabály közvetlenül vizsgálja az egyes adatszolgáltatók adatainak a részvételét a teljes (cella) értékösszegben és feltételezi azt, hogy a támadó egyike a cellát alkotó válaszadóknak.

<sup>17</sup> A p szabály általánosított formája, ahol további q paraméter bevezetésével finomítja az érzékenység kritériumát.

1992. évi LXIII. Törvény a személyes adatok védelméről és a közérdekű adatok nyilvánosságáról.

1993. évi XLVI. Törvény a statisztikáról.

Bánszegi K. [1997]: Felfedést akadályozó módszerek a statisztikai tájékoztatásban, *Statisztikai Szemle*, 75. évf. 12. sz. 1039-1046. old.

Bánszegi K. - Lakatos M. [1994]: Információszabadság – adatvédelem – statisztika (III.), *Statisztikai Szemle*, 72. évf. 10. sz. 761-777. old.

Carlson, M. [2002]: Assessing Microdata Disclosure Risk Using the Poisson Inverse Gaussian Distribution., *Stockholm*.

<http://www.matstat.umu.se/personal/Gunnar/banocoss/papers/carlson.pdf>

Cox, L. H. [1981]: Linear sensitivity Measure in Statistical Disclosure Control, *Journal of Statistical Planning and Inference*, 5. volume , 2. issue, 153-164 p.

Duncan, G. T.- Keller-McNulty, S. A. – Stokes, S. L. [2001]: Disclosure Risk vs. Data Utility: The R-U Confidentiality Map, (manuscript)

ELLIOT, M. (1996): Attacks on census confidentiality using the Sample of Anonymised Records: an analysis, *3rd International Seminar on statistical confidentiality, Bled*.

Erdei V. – Horváth R. (2003). Adatfelfedés elleni védelem statisztikai eszközei. *Statisztikai Szemle*, 82. évf. 8. szám, augusztus, 705-728. p.

Erdei V. – Sánta J. [2000]: A statisztikai adatok védelmének nemzetközi szabályozása, módszertani kérdései. *Népszámlálások az ezredfordulón 3. (Tanulmányok) KSH Budapest*.

Eurostat [1996]: *Manual on disclosure control methods, Luxemburg*

Eurostat [1999]: *Statistical Data Confidentiality*

Lakatos M. [1994]: Információszabadság – adatvédelem – statisztika (I.), *Statisztikai Szemle*, 72. évf. 7. sz. 547-559. old.

Merola, G. [2003]: Generalized Risk measure for Tabular Data, (manuscript), *Roma*.

<http://neon.vb.cbs.nl/casc/ISIBerlin/merola.pdf>

Skinner, C.J. – Elliot, M. J. [2002]: A Measure of Disclosure Risk for Microdata, (manuscript)

<http://www.ccsr.ac.uk/publications/occasion/occ23.pdf>

DATA CONFIDENTIALITY (2001). *Statistical Journal of the United Nations ECE*, 285-407p.

Emlékeztető az Adatvédelmi Bizottság 2009. február 24-ei üléséről:

[http://kshsps/C11/Adatvédelmi%20bizottság/Document%20Library/2009/Emlekezteto\\_AVB\\_090224\\_vegso.pdf](http://kshsps/C11/Adatvédelmi%20bizottság/Document%20Library/2009/Emlekezteto_AVB_090224_vegso.pdf)

Eurostat [2008]: *Protection of Confidential data at Eurostat*

Hundepool, A. et al [2009]: *Handbook on Statistical Disclosure Control*

<http://neon.vb.cbs.nl/casc/handbook.htm>



# Tájékoztatás

## 1. Leírás:

A tájékoztatás az a tevékenység, amikor a felhasználók rendelkezésére bocsátják a statisztikai tevékenység során előállított információkat. A folyamat része nem csak az információk közzététele, hanem a felhasználók tájékoztatása a számukra hasznos információk forrásairól, továbbá igényeik felmérése is.

## 2. Alapelvek:

A magas színvonalú statisztikai szolgáltatások segítik az állami és az üzleti döntéshozókat, hogy tényekre támaszkodhassanak, az érdeklődőknek pedig széles körű adatokat nyújtanak az ország gazdasági, társadalmi, környezeti állapotáról. A tájékoztatás hivatali szerveinek feladata nem csak az adatok közlése, hanem a lehetséges felhasználók informálása is a létező adatokról, illetve az információs igények felmérése is.

Hagyományosan a statisztikai adatok közzététele elsősorban nyomtatott kiadványokat jelentett, de a technikai lehetőségek bővülésével egyre elterjedtebbek az elektronikus adatközlések is. Ez alatt nem csak a hagyományos kiadványok internetes terjesztését kell érteni, hanem akár ingyen vagy pénzért letölthető teljes, mikroszintű adatállományokat vagy lekérdezhető adatbázisokat. A tájékoztatás körébe tartoznak ez egyedi adatkérésekre adott válaszok (pl. telefonos tájékoztatás), és a tömegkommunikációs eszközök igénybevitelével történő adatközlés is (sajtótájékoztató, interjú). A Hivatal legújabb tájékoztatási csatornája a kutatószoba, ahol a kutatók térítésmentesen használhatnak jól dokumentált mikroadatbázisokat és elérhetők a hozzájuk kapcsolódó kiadványok, elemzések.

Nagyon lényeges szempont, hogy a felhasználókhöz kerülő információk relevánsak (a felhasználói igényeknek megfelelőek), megbízhatóak, pontosak, teljesek, elérhetőek, érthetőek, jól használhatóak legyenek, és időben rendelkezésre álljanak (pl.: ne legyen eltérés a tájékoztatási naptárban közölt dátum és a publikálás dátuma között), továbbá elérhető árért kínálja információit a hivatal.

Az időszerűség és a pontosság közötti mérlegelés alapján kell dönteni arról, hogy előzetes adatokat hozzáférhetővé tegyünk-e a felhasználók számára. Amennyiben az előzetes adatok közzététele mellett döntünk, minden esetben fel kell hívni a felhasználók figyelmét arra, hogy a közölt adatok csak előzetes adatok, így pontosság szempontjából elmaradhatnak a végleges adatoktól.

Az adatok nyilvánosságra hozatala közvetlenül érint adatvédelmi problémákat, és a tájékoztatás minőségét befolyásolják a dokumentálásra, illetve az elemzésre vonatkozó szabályok is.

## 3. Minőségi irányelvek:

- A statisztikusoknak tisztában kell lenniük a felhasználók tájékoztatásra vonatkozó igényeivel, követniük kell a technológiai változásokat, hogy a felhasználók **minél egyszerűbben és költséghatékonyabban** érthessék el és használhassák a statisztikai információkat munkájuk során.
- A hivatali tájékoztatás nem nélkülözheti, hogy **felhasználói igényeit** megismerje. Ehhez érdemes használni a piackutatás és a marketing eszköztárát, elkülöníteni a felhasználói

csoportokat, hogy mindegyiknek a számára szükséges információt szolgáltatthassa, továbbá felhasználói elégedettség vizsgálatokat végezni.

- A statisztikai információ készítése során a forrásul szolgáló állományokat sokféle módon feldolgozzák. A közlés előtt ellenőrizték, hogy a **közölt adat és a forrás összhangban** legyen, azaz a forrásadatokból és a feldolgozás lépéseiből az az eredmény szülessen, amit közlétesznek.
- Amennyiben az adatminőség szempontjából szükséges lehet, hogy egy közlendő adat **validálását** külső szerv, vagy más főosztály végezze, akkor számukra még a közlés előtt elérhetővé kell tenni az adatot.
- Törekedni kell arra, hogy minél több szakstatisztika, adat hozzáférhető legyen a tájékoztatási adatbázison keresztül. A hozzáférés során biztosítsuk a felhasználók számára a különböző dimenziók szerinti lekérdezést, ügyelve az adatvédelmi szempontokra.
- Ha a közléshez valamilyen **elektronikus formát** választanak, akkor azt megfelelően tesztelni kell, hogy biztonságosan és megbízhatóan működjön.
- A közölt adatokkal együtt az alkalmazott **módszerek leírását és minőségmutatókat** is a felhasználók rendelkezésére kell bocsátani (Termékminőség mutatók összetevőnként: pl. pontosság: relatív szórás, mintavételi hiba, összehasonlíthatóság: idősorok hossza, a szakstatisztikák módszertani dokumentációján (Meta) keresztül).
- A tájékoztatás során különös figyelmet kell szentelni az **adatvédelemnek**.
- A tájékoztatók, kiadványok megjelenési formája legyen **egységes** (egységes arculat kialakítása).
- A közreadott információk mellett a felhasználók tájékoztatását segíti, ha a **kapcsolattartó személyt és elérhetőségeit** is megadják az adatok mellé, aki munkaidőben elérhető, és részletesebb magyarázattal, felvilágosítással tud szolgálni az érdeklődőknek.

## Irodalom:

1993. évi XLVI. törvény a statisztikáról

A Központi Statisztikai Hivatal elnökének III/2008. (SK 1.) KSH előírása a Központi Statisztikai Hivatal által összeállított egyes gyorstájékoztatók közzétételéről

IX/2009.KSH előírása a Statisztikai tükör és a Statistical reflections előkészítésének és közzétételének szabályairól

VII/2009. KSH előírása a gyorstájékoztatók stADAT-rendszerrel összekapcsolt előkészítésének, engedélyezésének és közzétételének szabályairól

11./2008. (HÉ 39.) MeHVM utasítás a Központi Statisztikai Hivatal Szervezeti és Működési Szabályzatáról

[http://kshsps/218%20SZMSZ/218%20SZMSZ/11%202008.%20\(HÉ%2039.\)%20MeHVM%20utasítás%20a%20Központi%20Statisztikai%20Hivatal%20Szervezeti%20és%20Működési%20Szabályzatáról/SZMSZ.doc](http://kshsps/218%20SZMSZ/218%20SZMSZ/11%202008.%20(HÉ%2039.)%20MeHVM%20utasítás%20a%20Központi%20Statisztikai%20Hivatal%20Szervezeti%20és%20Működési%20Szabályzatáról/SZMSZ.doc)

QUALITY GUIDELINES (2003). Statistics Canada, Fourth Edition.

QUALITY GUIDELINES for Official Statistics 2002. Helsinki.

A Központi Statisztikai Hivatal elnökének III/2009. KSH előírása a statisztikai adattárház és a tájékoztatási adatbázis működtetésének és fejlesztésének szabályairól

[A Központi Statisztikai Hivatal gazdasági elnökhelyettesének 5/2008. l e i r a t a a kutatószoba üzemeltetéséről](#)

A Központi Statisztikai Hivatal elnökének 1/2009. KSH statisztikai szabványa a szakstatisztikák rendszeréről, és a szakstatisztikákra vonatkozó módszertani dokumentációról

# Archiválás

## 1. Leírás:

Archiváláson az adatok, illetve metaadatok megőrzésének eljárásait értjük. Itt az eljárások valójában valamivel szélesebb körét tárgyaljuk, mert ez a tevékenységek összefügg az adatok megosztásával kapcsolatos feladatkörrel. Ugyanakkor le is szűkítjük a tárgyalt kérdéskört, amennyiben kizárólag digitális adattárolással foglalkozunk, és nem térünk ki az irattári tevékenység irányelveire, amelyre vonatkozólag elsősorban a vonatkozó jogszabályok adnak útmutatást.

## 2. Alapelvek:

### Törvényesség szem előtt tartása

A 2001. évi CVIII. Törvény 17. § (2) pont szerint felhatalmazást kap az informatikai és hírközlési miniszter, hogy a kormányzati informatika vonatkozásában a Miniszterelnöki Hivatal vezető miniszterrel együttesen rendeletben szabályozza:

- az informatikai biztonság követelményeit,
- az információs társadalommal összefüggő szolgáltatásokkal kapcsolatos elektronikus iratkezelés és a digitális archiválás szabályait,
- az információs társadalommal összefüggő szolgáltatás alapját képező tevékenység elektronikus úton való végzése miatt szükséges egyéb előírásokat.

Jelenleg a 3/2005. (III. 18.) IHM rendelet az elektronikus aláírással kapcsolatos szolgáltatásokra és ezek szolgáltatóira vonatkozó részletes követelményekről tartalmaz említést archiválásról.

A jogszabályok által nem szabályozott kérdésekben az adatkezelésre vonatkozó egyéb jogszabályok irányadóak.

Különös súlya van a személyi és egyéb nem nyilvános adatok védelmének.

Tartós és megbízható megőrzés

A titkos és szolgálati jellegű adatok titkosságának biztosítása

Egyéb okból védett adatok védettségének biztosítása (DC)

Archívum áttekinthető strukturálása, dokumentálása

Az archívumban rendelkezésre álló adatok tartalomjegyzékének karbantartása, listázása, esetleg nyilvánosságra hozatala, (esetleg) Interneten.

## 3. Minőségi irányelvek:

- **Kriptográfiai eljárások** alkalmazása esetén különös figyelmet kell fordítani arra, hogy az adat hozzáférhetővé tételét szolgáló információ: jelszó, kulcs, elektronikus aláírás tartósan, bizalmasan és biztonságosan megőrződjön. Valójában kriptográfiai eljárások alkalmazását lehetőség szerint kerülni javasoljuk, és hardver és rendszer szinten tartjuk kívánatosnak a megfelelő biztonság elérését.
- Az adathordozók minősége és az archiváló szoftver algoritmusai alapján garantálhatónak kell lennie, hogy bizonyos **elegendően hosszú időtartamra**

elegendően nagy valószínűséggel egyértelműen visszaállítható formában megőrződnek az archivált adatok. A teljesen egyértelmű visszaállíthatóságot követeljük meg. Független archiválásokkal egyszerűen növelhető a szükséges mértékűre a megbízhatóság. Archiváláskor a felelősnek meg kell győződnie róla, hogy az archívum valóban visszaállítható formában tartalmazza az archiválandó állományokat.

- **Megbízható tárolásról** beszélhetünk a következő esetekben:
  - Megőrzés
  - Bizalmas, illetve titkos kezelés, felfedés elleni védelem
- **Adatátadások, átvételek biztonsági és minőségi szempontjairól** az adatvédelemmel, valamint az igazgatási adatok felhasználásával foglalkozó fejezetek foglalkoznak részletesen.
- **Hozzáférhetőség és adatvédelem** szempontjából az alábbi tényezőket érdemes figyelembe venni:
  - Könyvtárstruktúrák kialakítása és karbantartása
  - Hardver védelem
  - Informatikai rendszer figyelemmel kísérése az archívum migrálása szükség esetén.

#### **Irodalom:**

PRESERVING & SHARING statistical material. Working Group on the Preservation and Sharing of Statistical Material: Information for Data Producers (2002). *The Royal Statistical Society & the UK Data Archive*.

A KSH Központ informatikai Üzemeltetési Szabályzat 4.1; 2008.03.31.

A Központi Statisztikai Hivatal központi informatikai rendszerének mentési rendje 1.3; 2008.03.29.

GYÖRFI L.- GYŐRI S. - VAJDA I. (2002): Információ és kódelmélet. Typotex, Budapest.

## Értékelés, felülvizsgálat és visszacsatolás

### 1. Leírás:

A statisztikai termelési folyamat nélkülözhetetlen része a folyamatos, rendszeres értékelés. Az értékelés egyrészt vonatkozhat a teljes folyamatra, másrészt az előállított termék minőségére. Az értékelés, felülvizsgálat célja, hogy képet kapjunk egyrészt az előállított termék minőségéről, másrészt pedig feltárjuk az esetleges hiányosságokat, azonosítsuk azokat a pontokat, ahol további fejlesztésekre van szükség. Az értékelés, felülvizsgálat eredményeinek visszacsatolása nélkülözhetetlen a statisztikai termékek minőségének folyamatos javításához. Az előállított termék minőségére vonatkozó információk nemcsak a statisztikusok, hanem a felhasználók számára is fontos információkat hordoznak az adatok használhatóságáról.

Az értékelés, felülvizsgálat és visszacsatolás leggyakrabban alkalmazott módszerei, eszközei:

- **Minőségjelentés:** egy statisztikai termék, vagy felvétel minőségével kapcsolatos információk összegyűjtésére és közzétételére szolgáló strukturált eszköz. Áttekintő képet ad az elkészült statisztikai termék minőségéről, mind az adat-előállítók, mind a döntéshozók, vagy a felhasználók számára.

A minőségjelentésben a következő minőség-összetevők mentén kell leírást, értékelést adni a minőségről:

- Relevanciára vonatkozóan: információ a felhasználókról, igényeikről, elégedettségükről,
  - Pontosságra vonatkozóan: mintavételi-, nem-mintavételi- (mérési és feldolgozási) hibákról,
  - Időszerűsége és időbeli pontosságra vonatkozóan: a tájékoztatáshoz kapcsolódó legfontosabb időpontokról
  - Hozzáférhetősége és érthetősége vonatkozóan: a tájékoztatási formákról, elérhető metaadatokról és dokumentációkról
  - Összehasonlíthatóságra és koherenciára vonatkozóan: a legfontosabb definíciós és időbeli eltérésekről.
- **Minőség-indikátorok:** a statisztikai termékek minőségét jellemző, mérhető indikátorok. (Pl.: mintavételi hiba, válaszadási arány)
  - **Folyamat indikátorok:** azok a tényezők, amelyek a folyamat ismétlése során folyamatosan változhatnak, jellemzik a folyamat minőségét, és legnagyobb hatással bírnak a végtermék minőségére. (Pl.: válaszadási arányok összeíronként)
  - **Folyamat-önértékelő kérdőív:** Az önértékelő kérdőív célja, hogy támogassa a statisztikai termelési folyamat minőségének vizsgálatát, és alkalmas eszköz legyen a fejlesztési intézkedések átgondolásához, a fejlesztési javaslatok megfogalmazásához. A kérdőív lehetőséget biztosít az adatfelvételek rendszeres, strukturált minőség-értékelésére, valamint részben a termékek minőségének értékelésére is.
  - **Felhasználói elégedettségi felmérés:** egy olyan felvétel, melynek az a célja, hogy képet kapjunk a felhasználók elégedettségéről a termékkel kapcsolatban, valamint információhoz jussunk a termék esetleges hiányosságairól, amely inputként szolgálhat a fejlesztésekhez.

- **Önértékelés és audit:** önértékelés során a felvétel felelőse, illetve az abban közreműködők tekintik végig a termelési folyamatot, illetve a végtermék minőségét és ez alapján hoznak döntést a szükséges fejlesztésekről (általában valamely modell, vagy standardok alapján történik). Ezzel szemben az audit során a felülvizsgálatot valamely független, harmadik, a folyamaton kívül álló (pl.. belső, külső szakértői csoport) fél végzi.
- **Címkézés, tanúsítvány kiállítása:** a statisztikai termék megfelel valamely (minőségi) követelményeknek, amelyek alapján jogosult az adott címke, tanúsítvány viselésére.

## 2. Alapelvek:

A folyamat és termékek minőségének értékelése minden esetben a folyamatos fejlődés, fejlesztés nélkülözhetetlen eleme, ezért minden esetben a termelési folyamat részévé kell tenni.

## 3. Minőségi irányelvek:

- Az értékelési folyamat során a termékek minőségét a minőség összetevők (relevancia, pontosság, időszerűség, időbeli pontosság; hozzáférhetőség, érthetőség; összehasonlíthatóság és koherencia) mentén kell áttekinteni, amely lehetőséget biztosít, hogy időben nyomon tudjuk követni a termékek minőségének a változását.
- Az értékelés során fontos, hogy ne csak a könnyen mérhető minőség összetevőkről (pl.: pontosságon belül a válaszadási arányok) próbáljunk információt összegyűjteni, hanem a minőség minden összetevőjéről gyűjtsük össze a rendelkezésre álló információkat, amelyek az értékelés alapjául szolgálhatnak.
- Az értékelési módszerek, eszközök megválasztásakor törekedjünk a fokozatosságra, ezért javasolt első lépésben a minőségjelentések, minőség-indikátorok, folyamat indikátorok használata, majd ezekre alapozva először az önértékelés, majd a belső audit, végül pedig a külső auditok alkalmazása.
- Az értékelési, felülvizsgálati folyamatokat a termelési folyamat részévé kell tenni. A feltárt hiányosságok visszacsatolásaként a javaslatokat figyelembe kell venni a következő időszak tervezésekor.
- Lehetőség szerint a felhasználókat (akár külső, akár belső felhasználókról legyen szó) be kell vonni az értékelési, felülvizsgálati folyamatokba, hogy a termék minőséggel szemben támasztott igényeiket valóban érvényre tudják juttatni.

## Irodalom:

HANDBOOK ON DATA QUALITY ASSESSMENT METHODS AND TOOLS, *Eurostat*, 2007.

STANDARD QUALITY REPORT, METHODOLOGICAL DOCUMENT, *Eurostat*, 2003.

STANDARD QUALITY INDICATORS, *Eurostat*, 2005.

JONES, N. - LEWIS D.: Handbook on improving quality by analysis of process variables, *Eurostat*, 2003

SZÉP KATALIN: A statisztikai termékek és folyamatok minőségbiztosítása (S106) SFT előterjesztés, 2005. (KSHáló, KSH-Stratégia)

FÖLDESI ERIKA: Termékminőség és mérési rendszere projekt - Projektzáró jelentés az SFT számára 2009. május 15. (KSHáló, KSH-Stratégia)

[http://kshsps/PMA/Shared%20Documents/SFT%20előterjesztések/Termékminőség%20Projektzaro\\_sft090529.doc](http://kshsps/PMA/Shared%20Documents/SFT%20előterjesztések/Termékminőség%20Projektzaro_sft090529.doc)

MAG KORNÉLIA: A termelési folyamat minőségbiztosítása . Projektzáró jelentés az SFT számára. 2005. (KSHáló, KSH-Stratégia)

Minőségügyi keretrendszer kialakítása, működése – Vezetői összefoglaló

[http://kshsps/C18/Módszertani%20Szakmai%20Kollégium/Mdszertani%20Szakmai%20Kollégium/ülések%20anyagai/2009.%20július%209/MSzK\\_0709\\_1%20naprend\\_Minosegugyi\\_keretrendszer\\_osszefoglalo\\_jav.doc](http://kshsps/C18/Módszertani%20Szakmai%20Kollégium/Mdszertani%20Szakmai%20Kollégium/ülések%20anyagai/2009.%20július%209/MSzK_0709_1%20naprend_Minosegugyi_keretrendszer_osszefoglalo_jav.doc)

Folyamatminőség indikátorok katalógusa

VI/2008. (SK 2.) KSH előírása a KSH termékminőség mérését szolgáló indikátorai számításának rendjéről

Önértékelő kérdőív az adatfelvétel felelősök számára

[http://kshsps/C18/Módszertani%20Szakmai%20Kollégium/Mdszertani%20Szakmai%20Kollégium/ülések%20anyagai/2008.%20október%201/KSH\\_ertekelo\\_v6\\_2.doc](http://kshsps/C18/Módszertani%20Szakmai%20Kollégium/Mdszertani%20Szakmai%20Kollégium/ülések%20anyagai/2008.%20október%201/KSH_ertekelo_v6_2.doc)

KSH Minőségjelentés

ESS Handbook for Quality Reports, Eurostat, 2009.

[http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/EHQR\\_FINAL.pdf](http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/EHQR_FINAL.pdf)

ESS Standard for Quality Reports, Eurostat, 2009

[http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/ESQR\\_FINAL.pdf](http://epp.eurostat.ec.europa.eu/portal/page/portal/ver-1/quality/documents/ESQR_FINAL.pdf)



## További tanácsok

### Kapcsolódási pontok

A statisztikai munkafolyamatokat nem lehet egymástól elkülönítve kezelni. Az egyes szakaszok legfontosabb minőségi jellemzői közvetlen hatással bírnak a kapcsolódó folyamatszakaszok minőségére.

Az egyes fejezetekben megfogalmazott minőségi irányelvek értelmezésénél minden esetben végig kell gondolni a folyamat egészére gyakorolt hatásukat is. Számptalan ponton kerülhetünk szembe döntési helyzetekkel, amikor is a felvételen résztvevők feladata a mérlegelés és az optimális megoldás kiválasztása. Ilyen lehet például, amikor pontosság és az adatszolgáltatási terhek csökkentése érdekében az optimális megoldást akarjuk megtalálni.

### Problémák forrása

Ahhoz, hogy a folyamat során tapasztalt problémákat a leghatékonyabb módon tudjuk kezelni, illetve a jövőben kiküszöbölni, mindig a probléma elsődleges forrását kell megtalálni, és azt megpróbálni orvosolni. Amennyiben erre már nincs lehetőség, a folyamaton belül minél korábban próbáljuk meg a hibákat feltárni, és a legkisebb torzítást okozó megoldást alkalmazni.

Tekintsünk erre egy példát.

- *Problémák:* nemválaszolás, válaszok gyenge minősége.
- *Okok:* kérdőív-szerkesztési hibák; túl sok kérdés; nehezen érthető, komoly felkészültséget feltételező kérdések; motiváltság, visszajelzések hiánya; túlzott leterheltség a sok kérdőív miatt.
- *A megelőzés fontossága:* Különböző technikákkal sokat tehetünk az adatok mind nagyobb mértékű beérkezéséért és mind jobb minőségéért. Az imputálás fázisában már egy meglévő adottsággal szembesülve nem várhatunk csodát.
- *Megoldási lehetőségek:* zöld szám üzemeltetése, kérdőívek előzetes tesztelése, próbafelvételek tervezése, apró ajándékok, visszajelzés az adatszolgáltatónak kis brosúra formájában, nemválaszolókból merített rész minta behajtása és elemzése.

Ahhoz, hogy a folyamat végrehajtása során szerzett tapasztalatokat a jövőben hasznosítani tudjuk, minden felvétel feldolgozását ki kell elemezni, az így levont tanulságok alapján kell felkészülni a következő felmérésre.

Tisztában kell lennünk az adatgyűjtés etikai normáival, s azokhoz mindenképpen igazodni szükséges.

### Erőforrások felhasználása

A statisztikai termelési folyamat végrehajtása során a minőség nem csupán az alkalmazott módszerektől, eljárásoktól függ, hanem a rendelkezésre álló erőforrások minősége és mennyisége is nagymértékben hatással lehet a végtermék, valamint a termelési folyamat minőségére. A rendelkezésre álló emberi erőforrást, valamint a felmerülő költségeket minden esetben figyelembe kell venni, és úgy kell megtervezni a termelési folyamatot, kiválasztani a

módszereket, meghatározni az eljárásokat, hogy a meglévő keretek között a lehető legjobb minőséget tudjuk elérni.