**Minőségjelentés**

**Erdei fatermékek termelése és készletváltozása (OSAP 1259/1)**

1. **Általános információk:**

Adatszolgáltatók meghatározása: állami erdőgazdaságok[[1]](#footnote-1) és egyéb erdőgazdálkodók

Kijelölt adatszolgáltatók száma: 546

Tényleges adatszolgáltatók száma: 366

Elektronikus beérkezés száma/aránya: 366 / 100 %

Nullás adatszolgáltatók száma: 265

A kérdőív adathelyeinek száma: 432

Adatgyűjtés módja: kombinált (az állami erdőgazdaságokra teljes körű, az egyéb erdőgazdálkodók esetében mintavételes)

Gyakorisága: éves

Tárgyidőszak: 2020

Beérkezési határidő: 2021. március 17.

Tervezett publikálás időpontja: 2021.06.30.

1. **Az adatgyűjtés tárgyidőszaki minősége**
	1. **Megbízhatóság**

A korábbi évekhez képest a következő módon módosítottuk a mintavételt: az állami szektort (állami erdőgazdaságok, nemzeti parkok, vízügyi szervek) teljes körűen mintavételeztük, a magánszektorban (beleértve a közösségi szektort is) pedig három réteget különítettünk el az erdőterv alapján várhatóan (egyeztetett sürgősség = 1) fahasználattal érintett terület nagysága szerint (5 hektár ill. azalatti – „kis”; 5 hektárnál nagyobb, de 50 hektárnál kisebb – „közepes”; 50 hektárnál nagyobb – „nagy”). Erre azért volt szükség, mert a korábbiakkal ellentétben a mintavételt súlyozatlan módon végeztük annak érdekében, hogy a magánszektorra vonatkozóan is közvetlenül a mintából lehessen becsülni az országos szintű nettó fakitermelést.

A módosítás eredményeképpen az előző évekhez képest jelentősen megnőtt a nullás adatszolgáltatások száma. Ez elsősorban a kis fahasználati területű erdőgazdálkodókra vezethető vissza. Emellett csökkent a beérkezett kérdőívek száma is, ami feltehetően részben a pandémiás helyzet következménye. A nullás adatszolgáltatások arányának növekedése a csökkent válaszadási hajlandósággal együttesen a statisztikák bizonytalanságát jelentősen növeli. Ez szükségessé teszi a későbbiekben a mintanagyság növelését, amit már jövőre tervezünk végrehajtani.

A beérkezett adatokat azok jóváhagyása előtt tisztítjuk. Ennek során változótípusonként megvizsgáljuk az adatok eloszlását, és statisztikai módszerekkel azonosítjuk az előforduló kiugró értékeket. Ezek leggyakoribb oka a nem megfelelő mértékegység használata, amit az adatszolgáltató megkeresése nélkül is javítani lehet. A kérdéses eseteket az adatszolgáltató megkeresésével tisztázzuk. A 1259-es számú adatgyűjtés során a beérkezett adatokat számos esetben korrigálnunk kellett (ennek számszerűsített értékét ld. lejjebb).

Statisztikai értelemben az adatok megbízhatóságának két, egymástól független összetevője van: a pontosság (*accuracy*), ami az adatokban lévő torzítottságot, vagyis a szisztematikus hibát fejezi ki, valamint a precizitás (*precision*), ami az adatokban rejlő véletlenszerű hiba számszerűsítése. Mindkét típusú hibának számos oka lehet. Jelen esetben az adatok torzítottsága elsősorban akkor lenne feltételezhető, ha az adatszolgáltatónak anyagi érdeke fűződne ahhoz, hogy a ténylegesen kitermelt fatérfogathoz ill. árbevételhez képest tudatosan többet vagy kevesebbet jelentsen be. Ilyen érdek azonban nincs, hiszen az adatok csak statisztikai célra használhatók fel, hatósági eljárás azok alapján nem indítható.

 A véletlenszerű hibák főként a mintavételi hibára vezethetők vissza. A véletlenszerű mintavételi hibát konfidencia intervallum számításával lehet számszerűsíteni. (Állami erdőgazdaságoknál a mintavétel teljes körű, így mintavételezési hiba csak abban az esetben léphet fel, ha adott erdőgazdaság adott évre valamilyen oknál fogva nem közöl adott értéket). A konfidencia intervallum a megbízhatósági szintnek megfelelő valószínűséggel fed át[[2]](#footnote-2) a kérdéses statisztika tényleges értékével. A mintavételi hibát a mintanagyság növelésével lehet csökkenteni. A statisztikák megbízhatóságának szempontjából ezért kulcsfontosságú az optimális mintanagyság meghatározása. Bonyolítja a helyzetet, hogy a 1259-es nyilvántartási számú adatgyűjtés sokváltozós.

A mintanagyságnak az adatok megbízhatóságára gyakorolt hatását szimulációs eljárással becsülhetjük. Ennek során a beérkezett adatokból veszünk véletlenszerűen különböző nagyságú mintákat, és ezekből konfidencia intervallumokat számítunk. Ritkább fafajok és fatermékek esetében az adatok kevésbé megbízhatók (vagyis a konfidencia intervallumok szélesebbek). A magánszektorból beérkezett adatok alacsony száma miatt a konfidencia intervallumok viszonylag szélesek (1. ábra).



1. ábra Rönk árbevételből számított átlagos egységárának a megbízhatósága a mintanagyság függvényében a magánszektor esetében. A konfidencia intervallumok 95 %-os megbízhatósági szintnek felelnek meg.

 **Indikátorok:**

Egység szintű súlyozatlan beérkezési arány: beérkezett kérőívek száma/összes adatszolgáltató

67,3 %

Tétel szintű súlyozatlan beérkezési arány: beérkezett válaszok száma/releváns adatszolgáltatók száma adathelyenként:

67,3 %

Egység szintű imputálási arány: imputált adatszolgáltatók száma/összes adatszolgáltató

Nem volt szükség imputálásra.

Tétel szintű imputálási arány: imputált értékek száma/összes érték adathelyenként

Nem volt szükség imputálásra.

Hibajavítások száma és aránya: javított rekordok száma/összes rekord: 79/1436

Adatszolgáltatók megkeresésével javított hibák aránya: 0 %

Adatfeldolgozó által javított hibák aránya: 6 %

**Időszerűség:** a tervezett közzététel és a leírt esemény referenciaideje közötti idő hossza

6 hónap.

* 1. **Időbeli pontosság:** a publikálás tervezett időpontja tartható-e, ha nem mi ennek az oka (pl. minőség-ellenőrzés, javítás)

Tartható.

* 1. **Hozzáférhetőség:**

**Indikátorok:**

Védett adatok száma és aránya:

Állami erdőgazdaságok: 67 db, 16 %

Nemzeti parkok: 85 db, 20 %

Vízügyi szervek: 39 db, 9 %

Magánszektor „kis” réteg: 66 db, 15 %

Magánszektor „közepes” réteg: 52 db, 12 %

Magánszektor „nagy” réteg: 58 db, 13 %

* 1. **Összehasonlíthatóság és koherencia**

Az előállított statisztikai adatok és velük kapcsolatban álló más adatok összehasonlítása:

Az árbevételi adatokból kiszámolható egységár a 1259-es számú adatgyűjtés *Erdei faválasztékok értékesítési* árai c. résztáblázatának statisztikáival vethető egybe. Az utóbbi táblázat az előbbitől abban különbözik, hogy az adott fatermék minősége, feldolgozottsága és az értékesítés helyszíne jobban körülhatárolt, valamint nem feltétlenül az adott évi könyvelésen alapul (specifikált egységár). Ennek ellenére a két adatlap árstatisztikái között kiugróan magas eltérésnek nem szabadna lennie. A 2. ábrán példaként bemutatott árstatisztikák között az adatok megbízhatóságát megkérdőjelező különbség nem látható.

**Időbeli összehasonlítás**: a gazdasági, szakpolitikai és egyéb körülmények alapján az adatok időbeli változása elfogadható, magyarázható-e:

Az adott évi nettó fakitermelést, választékszerkezetet, azok értékesítését és a kapott árbevételt egyrészt a fahasználati (fakitermelési) lehetőségek, másrészt a piac határozza meg. Egyik évről a másikra hirtelen változást legfeljebb egy váratlan esemény (pl. természeti katasztrófa) okozhat. Ilyen változást az idei évben nem tapasztaltunk. A tavalyi adatokkal való összehasonlítást a statisztikai közleményben adjuk közre.

 

2. ábra A1259-es számú adatgyűjtés két résztáblázatából számolt rönk árak összehasonlítása. A konfidencia intervallumok 95 %-os megbízhatósági szintnek felelnek meg. A háromnál kisebb mintanagysághoz tartozó értékeket nem jelenítettük meg.

1. A jelentésben közölt statisztikákban az állami erdőgazdaságok magukban foglalják a HM vagyonkezelő által művelt erdőket is. [↑](#footnote-ref-1)
2. A becsülni kívánt statisztika fix értékű, a számított konfidencia intervallum az, ami ahhoz képest eltérhet. Helytelen ezért azt mondani, hogy az intervallum adott valószínűséggel „tartalmazza” a statisztika tényleges értékét. [↑](#footnote-ref-2)