

# Minőségjelentés

## Erdészeti szaporítóanyagok és munkák vállalkozói kivitelezési díjai

(OSAP 1259/3)

### 1. Általános információk:

Adatszolgáltatók meghatározása:	állami erdőgazdaságok <sup>1</sup> és egyéb erdőgazdálkodók
Kijelölt adatszolgáltatók száma:	522
Tényleges adatszolgáltatók száma:	345
Elektronikus beérkezés száma/aránya:	345 / 100 %
Nullás adatszolgáltatók száma:	182
A kérdőív adathelyeinek száma:	67
Adatgyűjtés módja:	kombinált (az állami erdőgazdaságokra teljes körű, az egyéb erdőgazdálkodók esetében mintavételes)
Gyakorisága:	éves
Tárgyidőszak:	2019
Beérkezési határidő:	2020. március 17.
Tervezett publikálás időpontja:	2020.06.30.

### 2. Az adatgyűjtés tárgyidőszaki minősége

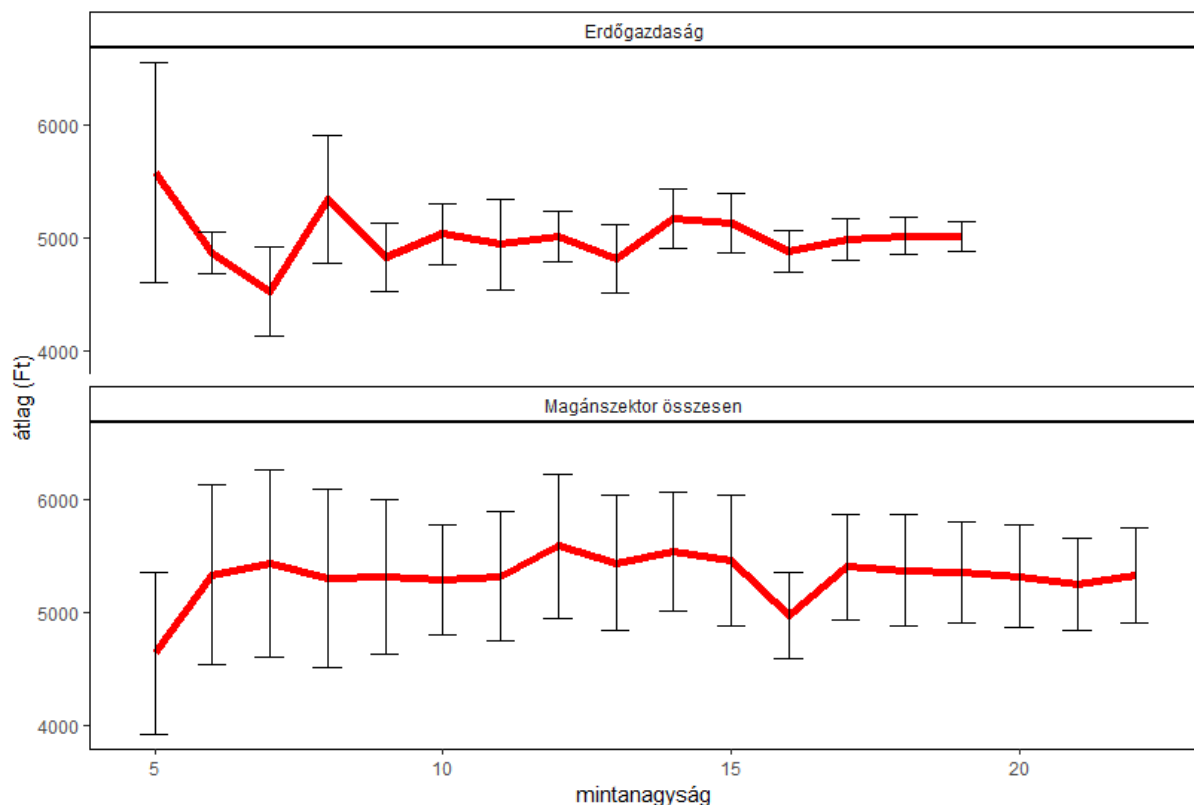
#### 2.1. Megbízhatóság

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.

---

<sup>1</sup> A jelentésben közölt statisztikákban – a védett adatokkal kapcsolatosak kivételével – az állami erdőgazdaságok magukban foglalják a HM vagyongazdálkodó által művelt erdőket is.

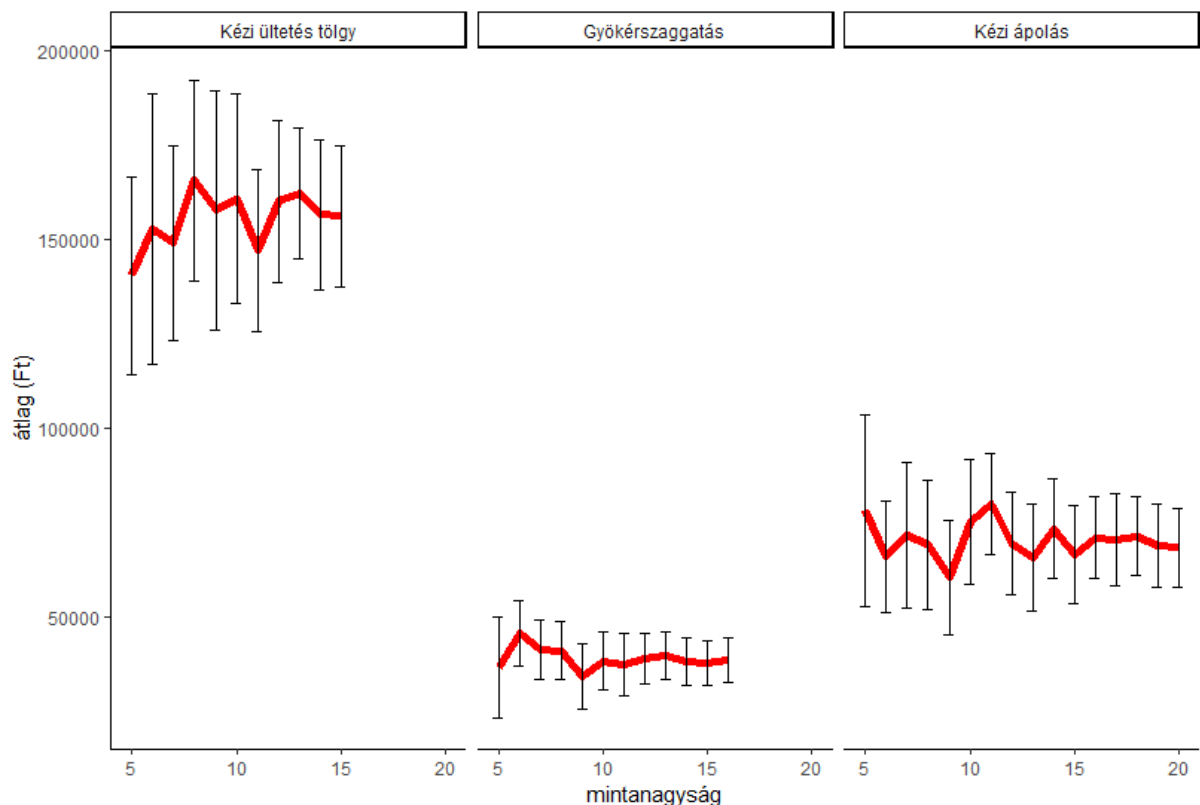
Statisztikai értelemben az adatok megbízhatóságának két, egymástól független összetevője van: a pontosság (*accuracy*), ami az adatokban lévő torzítottságot, vagyis a szisztematikus hibát fejezi ki, valamint a precizitás (*precision*), ami az adatokban rejlő véletlenszerű hiba számszerűsítése. Mindkét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



**1. ábra** Az átlagos akác fakitermelés díj (köbméterenként, sík terepen tarvágás esetén rakodóra közelítve) megbízhatósága a mintanagyság függvényében az erdőgazdaságok ill. a magánszektor esetében. A konfidencia intervallumok 95 %-os megbízhatósági szintnek felelnek meg.

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<sup>2</sup> 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, egyidejűleg több szolgáltatásról gyűjtünk adatokat.

<sup>2</sup> 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.



2. ábra Egyes szolgáltatások hektáronkénti átlagos árának a megbízhatósága a mintanagyság függvényében a magánszektor esetében. A konfidencia intervallumok 95 %-os megbízhatósági szintnek felelnek meg.

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). Az 1. és a 2. ábraábrán példaként bemutatott tételek esetében látható, hogy a beérkezett adatok mennyisége igen csekély, legfeljebb 20-25 körüli, ugyanakkor a konfidencia intervallum szélessége elfogadható, ami azt jelenti, hogy a mintanagyság növelésével lényegesen nem változik. Ennek oka az adatok viszonylag kismértékű szórása.

A beérkezett adatok alacsony számának az oka többértű. Egyfelől a válaszadási arány is viszonylag alacsony volt, ami részben a koronavírus járvánnyal is összefüggésben lehet. Másfelől az adatlap sokféle és jól körülhatárolt tevékenységet tartalmaz, amivel kapcsolatban az adatszolgáltatók nem feltétlenül bírnak információval. A jövőben a válaszadási arány ill. a beérkezett adatok mennyiségének növelése elengedhetetlenül szükséges feladat.

### Indikátorok

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

66 %

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

66 %

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: 37/2090

Adatszolgáltatók megkeresésével javított hibák aránya: 0,1 %

Adatfeldolgozó által javított hibák aránya: 2 %

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

6 hónap.

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

Tartható.

**2.3. Hozzáférhetőség:**

**Indikátorok:**

Védett adatok száma és aránya:

Állami erdőgazdaságok: 1 db, 2 %

HM vagyonkezelő: 28 db, 42 %

Állami szektor: 30 db, 45 %

Közösségi szektor: 5 db, 8 %

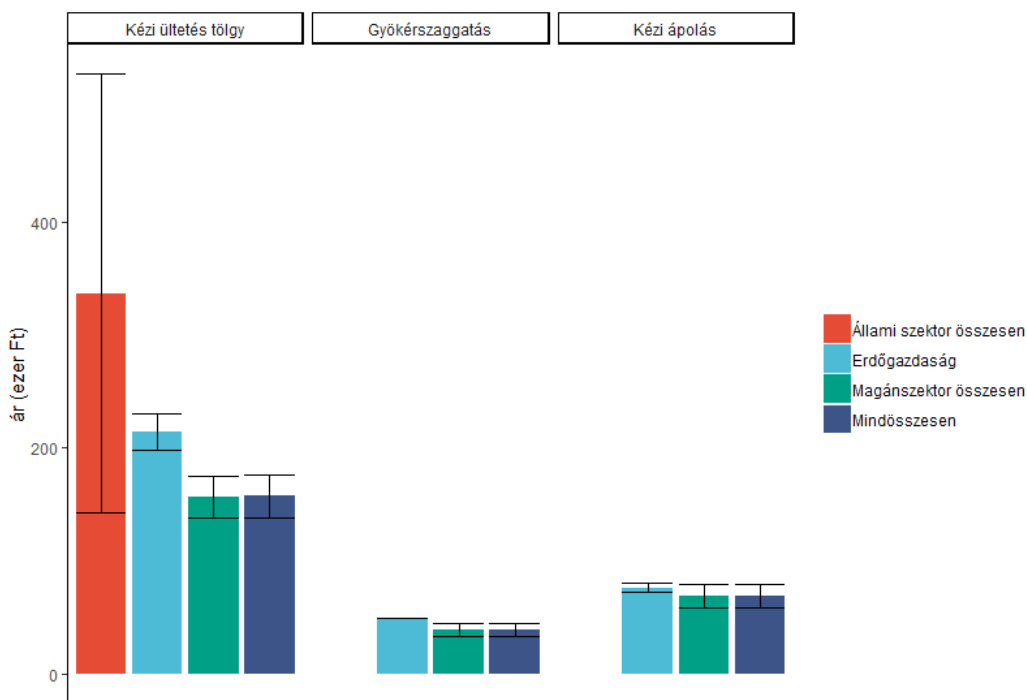
Magánszektor: 5 db, 8 %

## 2.4. Összehasonlíthatóság és koherencia

Az erdészeti tevékenységgel kapcsolatos díjakról gyűjtött statisztikák országos szinten egyedülállók, így a más nyilvántartásból származó való összehasonlítás nem lehetséges. Az adatlap alapján ugyanakkor szektorközi vizsgálatokat lehet végezni. Ha a szektorok közötti eltérések kiugróan magasak, az utalhat az adatok torzítottságára. Ilyen jelenséget ugyanakkor nem tapasztaltunk (példaként lásd az 1. ábra a 3. ábra ábrát).

### Időbeli összehasonlítás

Az erdészeti tevékenységgel összefüggő díjakról az idei év során gyűjtöttük először adatokat, így időbeli összehasonlítás nem tehető.



3. ábra Egyes tevékenységek hektáronkénti ára különböző szektorokban. A konfidencia intervallumok 95 %-os megbízhatósági szintnek felelnek meg. A háromnál kisebb mintanagysághoz tartozó értékeket nem jelenítettük meg.