

Mintavétel alapjai

Az első félévben vizsgálataink során azt feltételeztük, hogy egy sokaság minden egyedét megfigyeljük, azaz **Teljes körű megfigyelések** során vizsgált sokaság egészét megfigyelik. A teljes körű megfigyelések általában nagyon költségesek, sokszor lehetetlen megvalósítani. A teljes körű adatgyűjtés tipikus példája a népszámlálás.

A teljes körű megfigyelés helyett általában a részleges megfigyeléseket használjuk a gyakorlatban. **Részleges megfigyelések** során csak a sokaság egy részét, néhány egyedét figyeljük meg.

A részleges megfigyelések például a **monográfia**, melynek során a sokaság néhány kiemelt, fontos egyedének a vizsgálatát jelenti. **Ennek eredménye nem általánosítható a sokaság egészére nézve.**

Statistikai szempontból az lenne a szerencsés, ha a sokaság egy olyan részhalmazát (**minta**) tudnánk kiválasztani, amely alapján a sokaságra vonatkozóan is tudnánk állításokat megfogalmazni. Ehhez az kell, hogy a megfigyelt minta tulajdonságai tükrözzék az alapsokaság tulajdonságait (**reprezentativitás**)

Egy másik lehetőség a **véletlen mintavétel**, ami azt jelenti, hogy az alapsokaság **mindegyik** egyede **valamilyen valószínűséggel**, eséllyel kerülhet a mintába. A **reprezentatív** megfigyelés során a megfigyelt egyedek kiválasztása különböző kritériumok alapján történik. A reprezentatív mintavételhez szükséges, hogy a sokaság minden egyedének legyen esélye a mintába kerülni. Ennek megvalósításának módja a **véletlen mintavétel**. Ez azt jelenti, hogy az alapsokaság mindegyik egyede **valamilyen valószínűséggel**, eséllyel bekerülhet a mintába. A valóságban léteznek **nem véletlen mintavételi technikák** is, ilyen például az önkényes mintavétel, a hólabda, a kvótázás. Fontos megjegyezni, hogy ezek a minták **nem** tekinthetők **reprezentatív mintavételnek**, így ezek az alapsokaságra vonatkozóan **torzító** eredményeket hozhatnak.

A továbbiakban a véletlen mintákkal foglalkozunk.

Ahhoz, hogy mintát vegyünk szükséges az alapsokaság meghatározása, majd a mintavételi technika (**mintavételi eljárás/terv**) kiválasztása. Nézzünk néhány eljárást.

1. **Független, azonos eloszlású (FAE) minta** esetén az alapsokaság egyedeiről van egy egyedlistánk, ahonnan a mintába kerülő elemeket **viSSzatevéssel** választják ki. Ekkor az alapsokaság mindegyik egyede ugyanakkora valószínűséggel kerülhet be a mintába. Ennél a technikánál az alapsokaság egyedei akár többször is bekerülhetnek a mintába. Ez problémát okozhat akkor, ha valamilyen szélsőséges elem többször bekerül a mintába.
2. **Egyszerű véletlen (EV) minta** esetén az alapsokaság egyedeiről van egy egyedlistánk, ahonnan a mintába kerülő elemeket **viSSzatevés nélkül** választják ki. Ekkor az alapsokaság egyedei legfeljebb egyszer kerülhetnek a mintába.
3. **Rétegzett (R) minta** esetén az alapsokaság egyedeiről szintén van egy egyedlistánk. Amennyiben a sokaság heterogén és ezt valamilyen szempont szerint megközelítőleg homogén részcsoporthoz, rétegekbe tudnak bontani, akkor alkalmazzuk ezt a technikát. Például, ha a vizsgálat szempontjából fontos tényező a válaszadók iskolai végzettsége, akkor a mintát úgy rakjuk össze, hogy mind alacsony, mind közepes, mind magas végzettségűeket tartalmazzon.

mind felsőfokú végzettségű kerüljön a mintába. A rétegzett mintát úgy kapjuk meg, hogy minden rétegből (részsokaságból) EV, vagy FAE-mintát vesznek. Ekkor csak az a kérdés, hogy az egyes részcsoporthoz (rétegekhez) hány elem kerüljön a mintába. Erre megoldásként gyakran vagy **egyenletes elosztású**, vagy **arányos elosztású rétegzett mintát** alkalmaznak.

- **Egyenletes elosztás** esetén mindegyik rétegből ugyanannyi elemet válogatnak a mintába. Például iskolai végzettség esetén a mintában mind az alacsony, mind a közép, mind a felsőfokú végzettségűek száma megegyezik.
 - **Arányos elosztás** esetén a rétegek sokaságbeli arányát figyelembe véve történik a kiválasztás. Ekkor a minta és az alapsokaság rétegszerinti összetétele megegyezik, például ha az alapsokaság egyedeinek 30 százaléka diplomás, akkor a mintába került egyedei esetén is 30 százalék a diplomások aránya.
4. **Csoportos (CS) mintavétel** esetén nem kell egyedlista a teljes sokaságról. Ekkor az alapsokaságot heterogén csoportokra bontjuk szét. Ezután a csoportok közül veszünk mintát. A kiválasztott csoportokat pedig teljes körűen megfigyeljük.
5. A **többlépcsős (TL) mintavétel** az előző eljárások kombinálását jelenti. Például egy kétlépcsős mintavétel esetén először csoportos mintavételt alkalmazunk, majd a kiválasztott csoportokat nem teljes körűen figyeljük meg, hanem ezekből EV-mintákat veszünk.

Példa

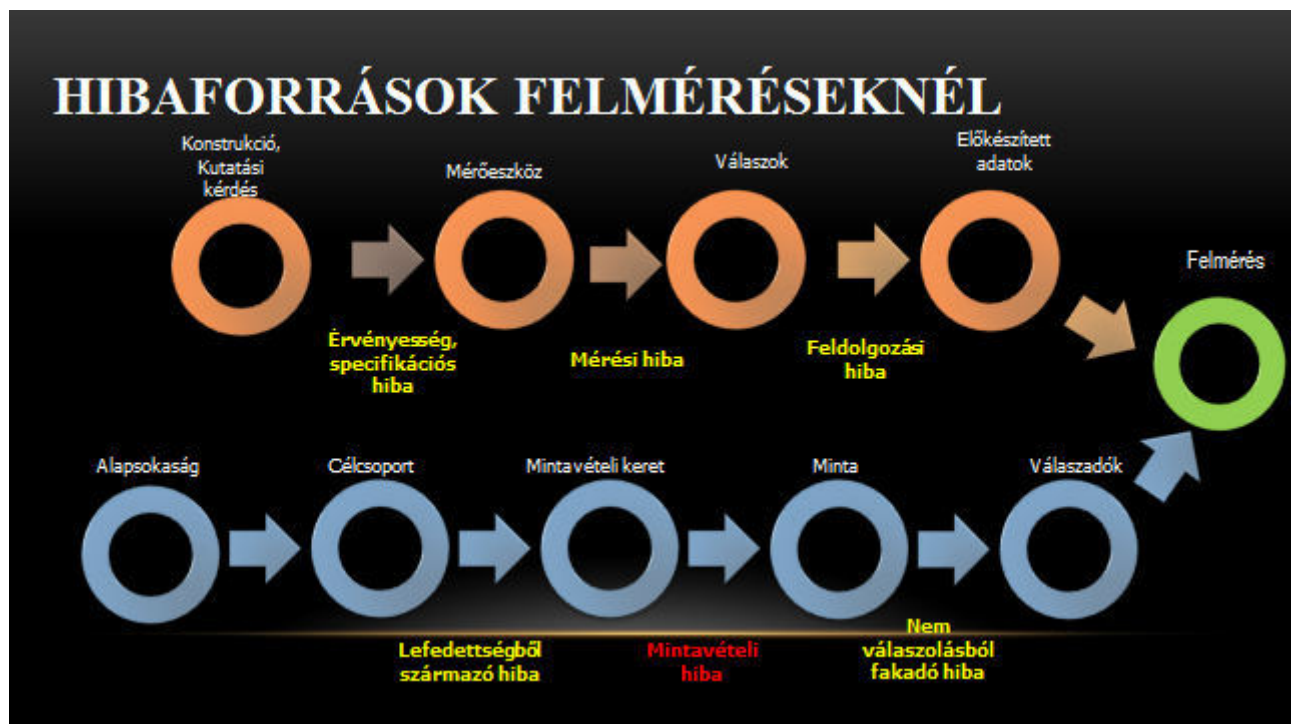
Egy áruházlánc közérzetjavító intézkedéseket szeretne végrehajtani. Ehhez meg szeretné kérdezni a dolgozók véleményét is.

Amennyiben

- minden dolgozó véleményét megkérdezzük, akkor teljes körű megfigyelésről van szó.
- Amennyiben a dolgozók közül reprezentatív véletlen mintát vesznek akkor az előbbi eljárásokra az alábbi példa adható.
 - EV-minta: véletlenszerűen választunk ki néhány dolgozót.
 - FAE-minta: véletlenszerűen választunk ki néhány dolgozót. (ismétlődés lehet, valakit többször is megkérdezhetünk).
 - R-minta: az alkalmazottakat beosztásuk szerinti csoportosítását tekintve (pl. pénztáros, eladó, osztályvezető, stb.) minden egyes csoportból választunk elemeket a mintába.
 - Egyenletes elosztás esetén minden csoportból ugyanannyi főt kérdezzük meg,
 - míg arányos elosztás esetén a mintában ugyanannyi lesz minden csoport aránya, mint az alapsokaságban.
 - CS-minta: véletlenszerűen kiválasztunk néhány áruházat, majd az ott dolgozók mindegyikét megkérdezzük.
 - TL-minta: véletlenszerűen kiválasztunk néhány áruházat, majd az ott dolgozókból EV-mintát veszünk.

A mintavételi módszer megválasztása közvetlen hatással van az adatminőségre. A választást sok tényező befolyásolja, többek között az előállítandó adatok pontosságának kívánt szintje, megfelelő rétegtípusú változók elérhetősége és a rendelkezésre álló költségvetési források.

Az adatgyűjtések, megfigyelések hibákkal járnak.



A **nemmintavételi hibák** azok a hibák, amelyek mind a teljes, mind a részleges megfigyeléseknél felléphetnek. Ezek **matematikai eszközökkel nem kezelhetők**. Fellépésük torzítást okozhat. Ilyenek például a specifikációs hiba, definíciós hiba, a mérési hiba, a feldolgozási hiba, a lefedettségből származó hiba, illetve a nem válaszolásból (például a kiküldött kérdőívek nem érkeznek vissza) fakadó hiba, a válaszadási hiba (téves adat közlése).

A **mintavételi hiba** a részleges megfigyelésből fakadó hiba. Ez a hibatípus **matematikailag kezelhető**.

A mintavétel másik nagykérdése a minta nagyságának meghatározása. Erre vonatkozóan csupán ajánlásokat lehet megfogalmazni. Általában a sokaság elemszámát N , míg a mintanagyságot n jelöli. Ekkor n/N értéket **kiválasztási aránynak** nevezzük. Olyan sokaságok esetén, melyek elemszáma többszázezer, vagy több millió, megfelelő mintavételi tervvel az 1-2 százalékos kiválasztási aránnyal rendelkező minták is megfelelően alkalmazhatóak. A kiválasztási arányt és a sokaság elemszámát egyszerre érdemes vizsgálni. A minta elemszámának meghatározásához támpont lehet, hogy az eredményeket milyen hibahatárral (maximális hibával) szeretnénk megkapni. Ezzel a következő leckében fogunk foglalkozni.

Önellenőrző kérdések

1. Milyen megfigyelés típusokat ismer?
2. Mit jelent a reprezentativitás?
3. Milyen mintavételi terveket ismer? Miért fontos a mintavételi terv?
4. Milyen hibatípusok léphetnek fel megfigyelések során?
5. Egy város a háztartások vízfogyasztásának átlagát szeretné megbecsülni 300 kiválasztott háztartás adatai alapján. Mire kell figyelni a háztartások kiválasztásánál?

SZEGEDI TUDOMÁNYEGYETEM
GAZDASÁGTUDOMÁNYI KAR
KÖZGAZDÁSZ KÉPZÉS
TÁVOKTATÁSI TAGOZAT
LECKESOROZAT
COPYRIGHT © SZTE GTK 2017/2018

A LECKE TARTALMA, ILLETVE ALKOTÓ ELEMEI ELŐZETES,
ÍRÁSBELI ENGEDÉLY MELLETT HASZNÁLHATÓK FEL.

JELLEN TARTALOM
A SZEGEDI TUDOMÁNYEGYETEMEN KÉSZÜLT
AZ EURÓPAI UNIÓ TÁMOGATÁSÁVAL.
PROJEKT AZONOSÍTÓ: EFOP-3.4.3-16-2016-00014

SZÉCHENYI 2020



MAGYARORSZÁG
KORMÁNYA

Európai Unió
Európai Szociális
Alap



BEFEKTETÉS A JÖVŐBE