

Mivel foglalkozik a statisztika?

A kvalitatív technikák mellett szükséges a kvantitatív technikák, információk alkalmazása. A számszerűsíthető, mérhető (kvantitatív) információk kinyerésében, rendszerezésében, elemzésében nyújthat segítséget a statisztika. A statisztika **tömegesen előforduló jelenségek** számszerű leírásával foglalkozó gyakorlati tevékenység és tudományos módszertan, mely magában foglalja az információk gyűjtését, feldolgozását, közzétételét.

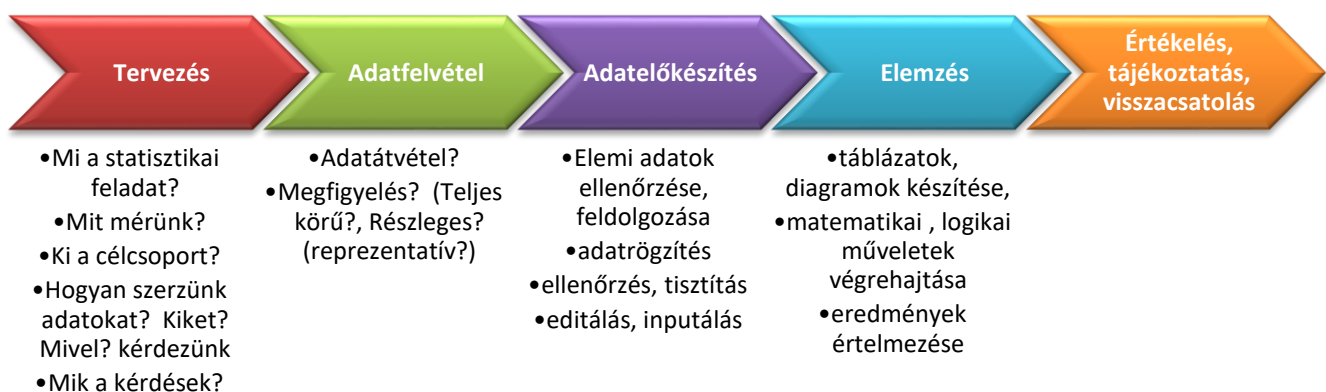
A statisztikát csoportosíthatjuk úgyis, mint általános statisztika, illetve szakstatisztika. Az **általános statisztika**, ahogyan a neve is sugallja általános elméleti, módszertani kérdésekkel foglalkozik. Ezen módszerek alkalmazása általában tudományágakhoz kötődő szakmákon belül, úgynevezett szakstatisztikák keretein belül történik. A **szakstatisztikák** (például: társadalomstatisztika, gazdaságstatisztika, oktatásstatisztika, idegenforgalmi statisztika, egészségügyi statisztika) egy adott témakör problémáinak, folyamatainak elemzésével foglalkoznak. Ehhez szükséges mind az általános statisztika, mind az adott téma terület fogalmainak ismerete. Például, hogyan tudnánk jellemezni a munkanélküliséget, ha nem is ismernénk a mérhető jegyeit, illetve nem tudnánk, hogy az adott terület mit ért munkanélküliség alatt.

A statisztikai számítások eredményei – ellentétben a matematikával – jelentéssel bírnak. Pontosan ebben rejlik a statisztika elsajátításának nehézsége. Hiszen, **nem elég, hogy egy probléma vizsgálatakor megfelelő elemzési módszert kell választanunk, a kapott eredményeket megfelelően kell tudnunk értelmezni is.** Kérdés, hogy hogyan juthatunk elemezhető adatokhoz.

A statisztikai munka fázisai

A statisztikai tevékenység kezdete egy probléma, egy kutatási kérdés, melyet megszeretnénk vizsgálni.

A statisztikai munka fázisai



Az adatok begyűjtése előtt e kutatási kérdés, az adatgyűjtés, illetve a statisztikai elemzés céljának függvényében egy **tervezési** szakaszt kell végrehajtánunk, melyben letisztázzuk többek között

- a vizsgált jelenség fogalmát, fogalmait, illetve megpróbáljuk ezeket indikátorokkal, mutatókkal leképezni, azaz meghatározzuk az alkalmazandó indikátorok körét (ehhez szükséges más szakirodalmak, elemzések tanulmányozása). A felhasználói igények sok esetben nem közvetlenül megfigyelhető jelenségekre, folyamatokra vonatkoznak (például ügyfelek lojalitása, országok versenyképessége, stb.). A statisztikusnak az igényeket le kell fordítani a statisztikai fogalmak nyelvére, hogy információkat tudjon gyűjteni róluk. E munkafolyamat része az elemzési munkát segítő munkatábláknak és a közlési tábláknak a megtervezése, a jelenséget leíró mutatók definiálása. A mutatók csak abban az esetben jellemzik megfelelően a vizsgált jelenséget, ha sikerül megragadni az adott jelenség mérhető jegyeit. A statisztikusok általában a vizsgált szakterület fogalmaiból és kategóriáiból indulnak ki a statisztikai fogalomalkotás során. Azonban, a tudomány fogalmai, kategóriái azonban nem mindig alkalmasak számszerűsítésre, egyes esetekben azért, mert több tényezőtől összetevődő jelenségre vonatkoznak, azaz túl bonyolultak, esetenként pedig azért, mert a megfigyelés feltételei nem adóttak.
- E fázisban megtervezzük az adatszerzés **módját is**: más adatforrásból veszünk át adatokat, vagy saját megfigyeléseinkre támaszkodunk. Az egyik legnagyobb hazai, statisztikai adatforrást a **Központi Statisztikai Hivatal** (KSH) adatbázisai jelentik. A KSH webcíme: <http://www.ksh.hu>. Emellett számos szerv tesz közzé statisztikai adatokat a saját honlapján is. Nemzetközi adatok keresésekor érdemes az Eurostat vagy az OECD statisztikai adatforrásait felhasználni, illetve a World Development Indicator (<http://databank.worldbank.org>) forrásokat ajánlom.
- Amennyiben saját megfigyeléseinkre, saját adatgyűjtésünkre támaszkodunk, meg kell terveznünk, hogy
 - **kiket** kérdezzük meg, mikor, milyen gyakorisággal kérdezzük. A megfigyelt **egyedek (egységek)** összességét statisztikai **sokaságnak** nevezzük. Statisztikai sokaságot bármi alkothat. Például egy ország lakossága egy adott időpontban, a vállalat termelése egy adott időszakban, egy bíróság ügyforgalma, a gazdálkodó szervezetek egy időpontban.
A megfigyelések során meg kell különböztetnünk megfigyelési és számbavételi egységet. A **megfigyelési egységek** a sokaság azon egyedei, amelyekre az adatgyűjtés vonatkozik, míg a **számbavételi egységek** az adatszolgáltató egységek. Például egy kutyaösszeírás során az adatgyűjtés a kutyákra vonatkozik, azaz a megfigyelési egység 1 kutya, míg az adatszolgáltatók a háztartások, azaz a számbavételi egység 1 háztartás. Egy sokaság az egyedeinek száma szerint lehet véges (megszámlálható), vagy végtelen (például egy kábelben áramló jelsorozatot vizsgálva). A gazdasági, társadalmi jelenségek vizsgálata véges sokaságokkal dolgozik.
 - Tisztázni kell, hogy teljes körű, vagy részleges a megfigyelést alkalmazunk. **Teljes körű megfigyelések** (cenzusok) során az egész vizsgált sokaságot megfigyelik. Ezek általában nagyon költségesek, és néha lehetetlen megvalósítani. A teljes körű adatgyűjtés tipikus példája a népszámlálás. A nemzetközi gyakorlat szerint általában 10 évenként tartanak népszámlálást. Legutóbb 2011-ben volt hazánkban ilyen összeírás.
Részleges megfigyelések során csak a sokaság egy részét, néhány egyedet figyeljük meg. A részleges megfigyelések főbb típusa a monográfia, a reprezentatív megfigyelés, egyéb részleges megfigyelések. A **monográfia** a

sokaság néhány kiemelt, fontos egyedének a vizsgálatát jelenti (például mélyinterjúval). Ezek statisztikai eszközökkel közvetlenül nem kezelhetők, azonban léteznek olyan technikák, amelyek a kvalitatív információkat kvantitatívvá alakítják. Például fókuszcsoporthoz vizsgálatoknál adott reakciót milyen időközönként, milyen hosszan mutat egy megfigyelt.

A **reprezentatív** megfigyelés során a megfigyelt egyedek (**minta**) kiválasztása különböző kritériumok alapján történik, úgy, hogy a kiválasztott minta tulajdonságai tükrözik az alapsokaság tulajdonságait. Ekkor a vizsgált sokaságot alapsokaságnak, míg a megfigyelt részsokaságot pedig mintának nevezzük.

- Fontos kérdés, **hogyan** kérdezzük meg a megfigyeltet: kérdőívvel vagy online módon, esetleg telefonos lekérdezés segítségével, kérdezőbiztossal, vagy a nélkül. A kérdőív a válaszadótól való információszerzés strukturált eszköze. A kérdőív lehet kérdező segítségével vagy a válaszadó saját maga által kitöltött. A kérdőívtervezés az előre meghatározott elérendő célokból indul ki. 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 valóban a célkitűzéseknek megfelelő információk nyerhetők-e belőle. Fény derül arra, hogy a kérdések egyértelműek-e a kitöltőknek, mindenre rákérdeztünk-e, amit szerettünk volna. Például, ha arra vagyunk kíváncsiak, hogy a felvételi jelentkezések során milyen tényezők befolyásolták az intézményválasztást, akkor megadhatunk általunk előre megadott szempontokat (például: család, barátok véleménye, hírnév, stb.), de előfordulhat, hogy valamilyen szempontra nem gondoltunk (például közelség). Ezek a szempontok a tesztelési fázisban nyitott kérdések segítségével előkerülhetnek.

Egy kérdőív leegyszerűsítve nyitott és zárt kérdéseket tartalmaz. **Zárt kérdés** esetén a válaszadó előre adott válaszalternatívákból (**ismérvváltozatok**) választhat, míg **nyitott kérdés** esetén a válaszadó bármit válaszolhat. A végső kérdőív változatban zárt kérdések alkalmazása a szerencsésebb, ugyanis nyitott kérdések esetén saját munkánk nehezítjük meg. Például, arra a kérdésre, hogy Elégedett-e Ön szolgáltatásunkkal., nyitott kérdés esetén az „Igen” és az „Igen!” válasz érdemileg nem különböző, azonban a szoftverek más alternatívának veszik. Ez esetben az elemzések előtt a változót **tisztítani** kell, azaz egységes formátumra kell hozni a válaszokat.

Nyitott kérdésként az is előfordulhat, hogy arra lennénk kíváncsiak, hogy „Mennyi az Ön havi jövedelme ezer forintban?”, és erre egy számértéket várunk el válaszként. Ez esetben kaphatunk olyan választ is: hogy „200”, „bruttó 300”, „kb. 120”, „120-130 ezer”. Láthatjuk, hogy ez esetben is a válaszok nem egészen felelnek meg a formai elvárásainak. Ekkor a kérdés pontosítása mellett (bruttó vagy nettó, stb. amire gondolunk) a végső kérdőívben vagy nyitott kérdésként tesszük fel a kérdést, de ekkor mindenféleképpen tisztítani kell az adatokat az elemzés előtt, vagy pedig a kérdést zárt formában tesszük fel, azaz kategóriákat, sávokat definiálunk. Ezek kialakításakor körültekintően kell eljárunk: hány darab és milyen széles tartományt kell készítenünk? Ezek meghatározásakor előfordulhat, hogy bizonyos kategóriákba nagyon kevesen fognak tartozni, viszont statisztikai szempontból kis elemszám alapján – kellő megbízhatóság mellett – nem

állíthatunk semmit, ilyenkor megint csak az elemzések megkezdése előtt ezt ellenőriznünk kell és a kevés elemet tartozó kategóriákat össze kell vonnunk. Zárt kérdések alkalmazásánál a válaszadó vagy egy (például mi az ön neve), vagy több (Mely témák érdeklik Önt? Egészség/oktatás/IT/szabadidő/egyéb) válaszalternatívát is megjelölhet. Előbbi esetben **egyszeres választásról**, utóbbi esetben **többszörös választásról** beszélünk.

Az **adattfelvétel** a statisztikai adatok beszerzését jelenti. Ez vagy más adatforrásokból való átvételt, vagy adatgyűjtést jelent. Az adatgyűjtés történhet kikérdezéssel (pld. kérdőívek alkalmazása), megfigyeléssel (pld. mérés), kísérlettel.

Az **adatelőkészítési folyamat** szorosan kapcsolódik az adatgyűjtési folyamathoz, magában foglalja az adatrögzítést, az ellenőrzést, és a javítást. A folyamat szakasz akkor zárul le, amikor a feldolgozásra kész elemi adatállomány rendelkezésre áll. Az ellenőrzések során azonosítják a hiányzó, érvénytelen vagy hibás információkat. Egyes ellenőrzési eljárások egyrészt logikai kapcsolatokra épülnek, amelyek a fogalmak és a változók definícióiból adódnak, másrészt statisztikai tesztek alkalmazásának eredményeként jönnek létre. Az **editálás** során a javíthatatlan hibákat, vagyis az érvénytelen vagy inkonzisztens adatokat (fatal error) törlik az adatállományból annak érdekében, hogy csökkentsék a becslés hibáját. Az **inputálás** az az eljárás, amellyel az editálás során azonosított hiányzó hibás, vagy inkonzisztens adatokat pótoljuk.

Az adatelőkészítés után az **elemzés** matematikai és logikai műveletek végzését jelenti. Ekkor különböző mutatók kiszámítása, értelmezése, szöveges elemzések készítése, grafikus reprezentáció készítése történik.

A **tájékoztatás** az a tevékenység, amikor a felhasználók rendelkezésére bocsátják a statisztikai tevékenység során előállított információkat. A folyamat része nemcsak az információk közzététele, hanem a felhasználók tájékoztatása az információk forrásairól, az alkalmazott módszerekről, az adatok minőségéről. Hagyományosan a statisztikai adatok közzététele elsősorban nyomtatott kiadványokat jelentett, de a technikai lehetőségek bővülésével egyre elterjedtebbek az elektronikus adatközlések. Ez alatt nem csak a hagyományos kiadványok internetes terjesztését kell érteni, hanem letölthető, mikroszintű adatállományokat vagy lekérdezhető adatbázisokat. A tájékoztatás körébe tartoznak ez egyedi adatkérésekre adott válaszok (pl. telefonos tájékoztatás), és a tömegkommunikációs eszközök igénybevételel történő adatközlés is (sajtótájékoztatás, interjú). Nagyon lényeges szempont, hogy a felhasználókhoz kerülő információk relevánsak (a felhasználói igényeknek megfelelnek), megbízhatóak, pontosak, érthetőek legyenek, és időben rendelkezésre álljanak.

Az adat-előállítás, mint termelési folyamat nélkülözhetetlen része a folyamatos, rendszeres **kiértékelés, visszacsatolás**. Ez vonatkozhat a teljes folyamatra, másrészt az előállított adat minőségére is. Az értékelés, felülvizsgálat célja, hogy képet kapjunk egyrészről az adat minőségéről, másrészt pedig feltárjuk az esetleges hiányosságokat, azonosítsuk azokat a pontokat, ahol további fejlesztésekre van szükség. Az értékelés, felülvizsgálat eredményeinek visszacsatolása nélkülözhetetlen a termékek minőségének folyamatos javításához. Ezek az információk nemcsak a statisztikusok, hanem a felhasználók számára is fontos információkat hordoznak az adatok használhatóságáról.

Hogyan lesz a kérdőívből adattábla?

Mivel egy sokaság az egyedein keresztül figyelhető meg, ezért fontosak az egyedek tulajdonságai, jellemzői, melyeket **ismérveknek (változóknak)** nevezünk. Az ismérvek lehetséges kimeneteleit **ismérvváltozatoknak (értékkészletnek)** nevezzük. Egy ember ismérve lehet például az életkora, a testmagassága, a súlya, a neme, az iskolai végzettsége, a lakóhelye, stb. Az ismérvváltozatok többfélék lehetnek. Például, életkorok esetében az ismérvváltozatok lehetnek a konkrét életkorok megadásai (1,2,3,... év), de képezhetünk ötéves hosszúságú korcsoportokat is (-4; 5-9; 10-14; ...), tízéves hosszúságú korcsoportokat is (-9; 10-19; 20- 29; stb.), vagy tetszőleges hosszúságú korcsoportokat is (-17; 18-35; 36-50; 51-). Lakóhely esetében ismérvváltozatként használhatjuk például a településeket, vagy a megyéket, vagy a régiókat is. Azt, hogy a kérdőíven mit szerepeltetünk ismérvváltozatként (például konkrét számokat, vagy csoportokat), az már a tervezési fázisban eldől, ennek pedig hatása lesz arra, hogy milyen módszertani eljárással vizsgálhatunk a későbbiekben egy-egy változót. A kérdőíveken a kérdés mellett a választható ismérvváltozatokat is feltüntetjük. Amikor a kitöltő kiválasztotta a válaszként adandó ismérvváltozatot, akkor ez a válasz a kitöltő **elemi adata** lesz.

Statisztikai elemzéseknél az adatok tárolása úgynevezett **adatmátrixokban, adattáblákban** történik. Ez gyakorlatilag egy olyan táblázatot jelent, ahol az egyes **rekordokban (sorokban)** az egyes megfigyelések, azaz az **egyedek** találhatóak. **Egyszeres választású kérdés** esetén az adattáblában egy oszlop egy kérdés lesz. Az adattábla **celláiban** pedig a válaszként adott elemi adatok kerülnek rögzítésre. Például tekintsük a következő kérdőív részletet.

Mi az Ön neme?

☒ 1-Férfi

☐ 2-Nő

Mely alapszak hallgatója?

☐ 1-Gazdálkodás és menedzsment

☐ 2-Kereskedelem és marketing

☐ 3-Pénzügy és számvitel

☐ 4-Egyéb, és pedig:

Osztályozza az előadót az előadások alapján! (9: nem tudom megítélni)

Felkészültség ☐ ☐ ☐ ☐ ☐ ☐

1 2 3 4 5 9

A kérdőív részleten 3 zárt, egyszeres választású kérdés található.

A kérdőív kitöltése történhet papír alapon, illetve elektronikus úton is például valamilyen interneten elérhető (például: Google Docs, kerdoivem.hu, vagy más) „lekérdező rendszer” segítségével is.

Mit kell tennünk, amennyiben a kitöltött kérdőívek alapján adattáblát szeretnénk építeni? Mivel három egyszeres választású kérdésünk van, így az adattábla szerkezete úgy néz ki, hogy 3 oszlopunk lesz (kérdésenként egy), illetve annyi sorunk ahányan kitöltötték a kérdőívet. Már csak az a kérdés, hogy mit kell „elemi adatként” rögzítenünk a táblázat celláiba.

A nem és az alapszak esetében az egyes válaszlehetőségek előtt egy-egy számkódot találunk, mely gyakorlatilag a válaszalternatíva sorszámát jelöli. Az adattáblába vagy a kiválasztott alternatívát, vagy ennek számkódját rögzítjük, tehát a nem esetében az adott kitöltő esetén vagy a „férfi” vagy az 1 kódot rögzítjük. Melyiket célszerű alkalmazni? Mivel szöveges

válaszalternatívák esetén egy betű elütés, illetve különbség is más alternatívának számít a szoftverek szerint, így papír alapú kérdőívek rögzítésénél célszerű a szövegek helyett az alternatívának megfelelő számkódokat tárolni. Ahol a változó alternatívái „számok” (például az oktató megítélése), ott egyértelmű, hogy a számértéket tároljuk.

Elektronikus kérdőívező rendszerek alkalmazásakor a rendszeren múlik, hogy a szöveg vagy a számkód kerül tárolásra. Például a Google nem a számkódokat, hanem a szövegeket tárolja, tehát például a nemnél a „Férfi”, illetve „Nő” alternatívákat.

Nagyon vigyázzunk, attól hogy a szövegek helyett számkódokat tárolunk, attól még matematikai műveleteket nem végezhetünk. Például számkódok tárolása esetén a szoftverek a nemből számolnának akár átlagot, akár szórást is, de ennek nincs értelme!

Fontos megjegyezni, hogy a mértékegységgel rendelkező adatok mértékegységeit, illetve a százalékjelet nem rögzítjük (azaz nem gépeljük be az adatokkal együtt), hanem az oszlop fejlécében rögzítjük, vagy cellaformátumként állítjuk be.

	A	B	C	D
1	Nem	Szak	Előadó szakmai felkészültsége	
2		2 pénzügy és számvitel		5
3		1 pénzügy és számvitel		5
4		2 gazdálkodás és menedzsment		5
5		1 kereskedelem és marketing		5
6		2 pénzügy és számvitel		5
7		2 gazdálkodás és menedzsment		5
8		1 gazdálkodás és menedzsment		5
9		2 pénzügy és számvitel		5
10		2 pénzügy és számvitel		4
11		1 pénzügy és számvitel		5
12		1 kereskedelem és marketing		5
13		2 pénzügy és számvitel		5
14		2 pénzügy és számvitel		4
15		2 gazdálkodás és menedzsment		5
16		2 gazdálkodás és menedzsment		5
17		2 gazdálkodás és menedzsment		5
18		1 kereskedelem és marketing		5
19		1 kereskedelem és marketing		5

A **többszörös választás** kezelésére többféle megoldás használatos. Ebből két lehetőséget említünk.

1. az adattáblában egy oszlop egy-egy alternatíva lesz. Például a mely témák érdeklik Önt? Egészség/oktatás/IT/szabadidő/egyéb kérdés esetén e kérdés öt (alternatívánként egy) változóval lenne leképezhető. A cellákba pedig 0-1 értékeket rögzíthetünk, annak megfelelően, hogy valaki az adott alternatívát nem jelölte meg, vagy megjelölte.

Egészség	Oktatás	IT	Szabadidő	Egyéb
0	0	1	1	0

0	0	0	1	0
1	1	1	1	1
0	0	0	0	0

A fenti táblázat szerint az első sorban szereplő válaszadót az IT és a szabadidő, a második sorban lévőt csak a szabadidő, harmadik sorban lévőt minden, a negyedik sorban lévőt semelyik lehetőség sem érdekli.

2. Az adattáblában a kérdést egy változóba képezzük le:

téma
IT, Szabadidő
Szabadidő
Oktatás, IT, Szabadidő, Egyéb

E formátum kezelése kicsit macerásabb, ha például arra gondolunk, hogy meg kellene határozni az IT iránt érdeklődők számát, akkor ennek meghatározása csak több függvény alkalmazásával lehetséges.

Nyitott kérdések esetén az adattáblában egy oszlop egy kérdés lesz.

Amennyiben új változót hozunk létre, meg kell adnunk, illetve meglévő változók esetében be kell állítanunk a **változó fontosabb tulajdonságait**, így például:

- a változó **nevét**, amivel azonosítani tudjuk. Ez egyfajta munkanévnek fogható fel, mely lehet rövidítés is.
- A változó **tárolási formáját**. Ez nagyon sokféle lehet, például szöveg, szám, pénznem, stb. A számítógépes szoftverek általában a változók lehetséges értékeit számkódokkal tárolják: például a nem esetében nem férfi és nő, hanem ezek helyett egy-egy számkód (pld. 1-2 kerül rögzítésre).
- A változó formai beállításait. Például hány tizedes jegy pontosságú értékekkel dolgozunk, vagy maximum hány karakter hosszúságú szöveget tárolunk, stb.
- A változó **mérési szintjét**. A változókat mérési szint szerint két nagyobb csoportja oszthatjuk: kategoriális és metrikus (skála) változókra.

A kategoriális változókra belül megkülönböztethetünk nominális és ordinális mérési szintű változókat. A **nominális** változók esetében a változó lehetséges alternatíváit csak megkülönböztetni tudjuk, de közöttük semmiféle rendezettségi viszonyról nem beszélhetünk, ilyen például a nem. **Ordinális** változók esetében a kategóriák között valamilyen rendezettséget is megállapíthatunk. Ennek tipikus példája az iskolai végzettség (alapfok, középfok, felsőfok).

A metrikus változók gyakorlatilag számok, melyekkel műveleteket végezhetünk. E változóknak is van ordinális jellege. **Intervallum (különbségi) skála** esetén a különbségképzés értelmezhető (például 2 dátum, vagy 2 hőmérsékleti érték különbségének van értelme). Abban az esetben, ha a különbségképzés mellett a hányados-képzésnek is van értelme, akkor **arányiskáláról** beszélünk.

A mérési szint határozza meg azt, hogy a változót milyen elemzésekbe, és ezeken belül milyen szerepkörbe lehet bevonni. A válaszalternatívák megadása a kérdőíven

nominális változók esetében ABC sorrendben, míg ordinális változók alapján a rendezettségnek megfelelően történik.

- Az ismérvváltozatok kódolását, ha van. Például, ha az elemzésekben szerepel változóként a nem, akkor beállíthatjuk azt, hogy a 0 szám(jegy) a nőket, míg az 1 szám(jegy) a férfiakat jelentse.
- A hiányzó adatok helyettesítési körét. Ha egy megfigyelés során hiányzik egy adatunk és ezt jelölni szeretnénk, akkor ezt egy általunk meghatározható kóddal lehet helyettesíteni. Például, ha egy kérdőíven kérdésként az szerepel, hogy „Értékelje 1-4 között 4 fokú skálán főnökét!”, és a válaszok között meg akarjuk különböztetni azt, hogy ki miért nem válaszol, akkor lehetőségünk olyan alternatívák felvételére is: 5: Nem tudom megítélni; 6: Megtagadom a választ.

Tekintsük át az eddig leírtakat! Hogyan nézhet ki az adattábla néhány kérdőív rögzítése után, továbbá a változóknak milyen tulajdonságai vannak?

Az adattáblában a nem változó tárolási formáját tekintve szám (számkódokat rögzítettünk), ugyanakkor mérési szintjét tekintve nominális. A szak változó tárolási formáját tekintve szöveg, továbbá mérési szintjét tekintve nominális. Ez az adattábla részlet csupán demonstrációs célú abból a szempontból, hogy a valóságban mind a nem, mind a szak változó tárolási formája vagy szöveg, vagy szám lenne.

A nem változó esetében érték hozzárendelésről is kell beszélnünk: 1 a férfiakat, 2 pedig a nőket jelenti.

Az előadó szakmai felkészültsége változó tárolási formáját tekintve szám, mérési szintjét tekintve pedig skála (metrikus). Bár megjegyzendő, hogy az értékeljen 1-4, illetve 1-5 között kezdetű kérdések mérési szintjét tekintve megoszlanak a kutatók álláspontjai: van, aki ezeket skála változónak és van, aki pedig csak – statisztikailag ez lenne korrektebb – ordinális változónak tekinti.

E változóval kapcsolatosan van még egy fontos teendőnk, ugyanis e kérdés esetében volt egy nem tudom megítélni alternatíva is (a nem és a szak kérdésnél ilyen nem volt, ott ha valaki nem válaszol, üresen marad az adott cella). Ez a kérdés szempontjából hiányzó értéknek számít, melyet kezelniük kell. Ugyanis, ha e változóból netán átlagot számolnánk, akkor erre értéként 5,04 adódik, mely lehetetlen, mivel 1-5 között kellett értékelni, tehát az átlagnak is 1 és 5 közé kellene esni. Az ellentmondás a 9 értékekben keresendő. Ezeket, ha nem kezeljük, azaz nem tekintjük hiányzó értéknek, akkor a szoftverek beszámítják ezeket az értékeket is az átlagba, ami félrevezető eredményt adhat. Ezek kezelésére az Excelben szűrők segítségével van lehetőségünk.



Önellenőrző kérdések

1. Mivel foglalkozik a statisztika?
2. Mit nevezünk statisztikai sokaságnak? Mondjon rá példákat!
3. Milyen megfigyelési típusokat ismer?
4. Mit jelent a teljes körű megfigyelés?
5. Mi az alapsokaság?
6. Mi a minta?
7. Mi a különbség a teljes körű és a részleges megfigyelés között?
8. Mit jelent a részleges megfigyelés? Milyen típusait ismeri?
9. Mi a monográfia?

10. Mit jelent a reprezentatív mintavétel?
11. Mi a különbség megfigyelési és számbavételi egység között?
12. Mit nevezünk ismérveknek?
13. Mik az ismérvváltozatok?
14. Mi az elemi adat?
15. Mik a statisztikai munka főbb fázisai?
16. Mi a tervezési fázis szerepe és főbb feladatai?
17. Mi az adatelőkészítési fázis szerepe és főbb feladatai?
18. Általában hogyan néz ki egy statisztikai adatbázis, mit jelképeznek az oszlopok, mik a sorok?
19. Milyen fontosabb jellemzői vannak a változóknak?
20. Milyen mérési szinteket ismer? Mondjon példát rájuk!

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 TANYAG
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