

Statisztikai programrendszerek

Kis-Tóth Lajos – Lengyelné Molnár Tünde –
Tóthné Parázsó Lenke

MÉDLAINFORMATIKAI KIADVÁNYOK

Statisztikai programrendszerek

Kis-Tóth Lajos – Lengyelné Molnár Tünde –
Tóthné Parázsó Lenke



Eger, 2013



Korszerű információtechnológiai szakok magyarországi adaptációja

TÁMOP-4.1.2-A/1-11/1-2011-0021

Nemzeti Fejlesztési Ügynökség
www.ujszecenyterv.gov.hu
06 40 638 638



A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Lektorálta:

Nyugat-magyarországi Egyetem Regionális Pedagógiai Szolgáltató és
Kutató Központ

Felelős kiadó: dr. Kis-Tóth Lajos

Készült: az Eszterházy Károly Főiskola nyomdájában, Egerben

Vezető: Kérészy László

Műszaki szerkesztő: Nagy Sándorné

Tartalom

1.	<i>Bevezetés</i>	<i>9</i>
1.1	Célkitűzések, kompetenciák a tantárgy teljesítésének feltételei..	9
1.1.1	Célkitűzés.....	9
1.1.2	Kompetenciák:.....	9
1.1.3	A tantárgy teljesítésének feltételei	10
1.2	A kurzus tartalma	10
1.3	Tanulási tanácsok, tudnivalók	12
1.4	tananyag	12
2.	<i>Adatbevitel a gyakorlatban SPSS szoftverrel</i>	<i>15</i>
2.1	Célkitűzések és kompetenciák	15
2.2	Tananyag	15
2.2.1	Sablonszerű gondolkodás	16
2.2.2	Adattípusok	17
2.2.3	A válaszok felvitele	18
2.2.4	Az adatok kiértékelésének csoportosítása	20
2.2.5	Nominális adatok értékelése	22
2.2.6	Nominális adatok kiértékelése kereszttáblával	25
2.2.7	Khi-négyzet-próba	27
2.3	Összefoglalás, kérdések	31
2.3.1	Összefoglalás	31
2.3.2	Önellenőrző kérdések.....	31
3.	<i>Leíró statisztikai értékelés SPSS táblázatkezelőkkel.....</i>	<i>33</i>
3.1	Célkitűzések és kompetenciák	33
3.2	Tananyag	33
3.2.1	Leíró statisztika	34
3.2.2	Számított középértékek és helyzeti középértékek	34
3.2.3	Gyakoriság	40
3.2.4	Gyakorisági poligon és a középérték-mutatók	45
3.2.5	A középértékek egymáshoz viszonyított kapcsolata	46
3.2.6	Szóródási mérőszámok.....	47
3.3	Összefoglalás, kérdések	53
3.3.1	Összefoglalás	53

3.3.2	Önellenőrző kérdések.....	53
4.	<i>Matematikai statisztikai lehetőségek az SPSS táblázatkezelőkben</i>	55
4.1	Célkitűzések és kompetenciák.....	55
4.2	Tananyag.....	55
4.2.1	Matematikai statisztika	56
4.2.2	Korreláció	56
4.2.3	Korrelációanalízis.....	62
4.2.4	Regressziószámítás.....	66
4.2.5	Faktoranalízis.....	69
4.2.6	Parciális korreláció	74
4.2.7	A Spearman-féle rangkorreláció.....	77
4.2.8	Klaszteranalízis	77
4.3	Összefoglalás, kérdések	78
4.3.1	Összefoglalás	78
4.3.2	Önellenőrző kérdések.....	79
5.	<i>Magasabb szintű értékelési módszerek a gyakorlatban</i>	81
5.1	Célkitűzések és kompetenciák.....	81
5.2	Tananyag.....	81
5.2.1	Hipotézisvizsgálatok	82
5.2.2	Null- és alternatív hipotézisek, döntési szituációk	83
5.2.3	t-próba.....	84
5.2.4	Egymintás t-próba	84
5.2.5	Kétmintás t-próba	86
5.2.6	Varianciaanalízis	89
5.2.7	A Mann–Whitney-próba, Wilcoxon-próba, Kruskal–Wallis-próba értelmezése	93
5.3	Összefoglalás, kérdések	94
5.3.1	Összefoglalás	94
5.3.2	Önellenőrző kérdések.....	94
6.	<i>Értékelési eredmények szemléltetésének lehetőségei a táblázatkezelő szoftverekben</i>	97
6.1	Célkitűzések és kompetenciák.....	97
6.2	Tananyag.....	97
6.2.1	A diagramok szerepe	98

6.2.2	Diagramtípusok	98
6.3	Összefoglalás, kérdések	102
6.3.1	Összefoglalás	102
6.3.2	Önellenőrző kérdések.....	102
7.	<i>Nemparaméteres eljárások.....</i>	103
7.1	Célkitűzések és kompetenciák	103
7.2	Tananyag	103
7.2.1	Paraméteres és nem paraméteres próbák.....	104
7.2.2	Kolmogorov-Szmirnov Test.....	105
7.2.3	Két független minta Mann-Whittney próba.....	108
7.2.4	Két összetartozó minta Wilcoxon-féle előjeles rangpróba	112
7.2.5	Több független minta egy szempont szerint Kruskal- Wallis próba.....	115
7.3	Összefoglalás, kérdések	119
7.3.1	Összefoglalás	119
7.3.2	Önellenőrző kérdések.....	119
8.	<i>ANOVA</i>	121
8.1	Célkitűzések és kompetenciák	121
8.2	Tananyag	121
8.2.1	ANOVA elmélet kérdései	121
8.2.2	One way ANOVA egyutas variacionális	122
8.3	Összefoglalás, kérdések	133
8.3.1	Összefoglalás	133
8.3.2	Önellenőrző kérdések.....	133
9.	<i>Multiple Responses.....</i>	135
9.1	Célkitűzések és kompetenciák	135
9.2	Tananyag	135
9.2.1	Multiple Responses alapértelmezése.....	135
9.3	Összefoglalás, kérdések	143
9.3.1	Összefoglalás	143
9.3.2	Önellenőrző kérdések.....	143
10.	<i>Klaszteranalízis.....</i>	145
10.1	Célkitűzések és kompetenciák	145

10.2	Tananyag.....	145
10.3	Klaszteranalízis alapértelmezése	145
10.3.1	A klaszteranalízis csoportosítása	147
10.3.2	Klaszterosítási módszerek	148
10.3.3	a klaszteranalízis gyakorlati alkalmazása	149
10.4	Összefoglalás, kérdések	158
10.4.1	Összefoglalás	158
10.4.2	Önellenőrző kérdések.....	158
11.	Faktorelemzés	159
11.1	Célkitűzések és kompetenciák.....	159
11.2	Tananyag.....	159
11.2.1	Faktorelemzés alapjai.....	160
11.2.2	A faktorelemzés alkalmazási területei:	162
11.2.3	A faktorelemzés lépései	164
11.2.4	Feladat megoldás értelmezése.....	167
11.3	Összefoglalás	180
11.4	Önellenőrző kérdések	180
12.	Összefoglalás	181
12.1	Tartalmi összefoglalás.....	181
13.	Kiegészítések.....	184
13.1	Irodalomjegyzék	184
13.1.1	Hivatkozások.....	184
13.2	Médiaelemek összesítése.....	185
13.2.1	Táblázatjegyzék	185
13.2.2	Ábrajegyzék	185
13.3	Glosszárium, kulcsfogalmak értelmezése	190

1. BEVEZETÉS

1.1 CÉLKITŰZÉSEK, KOMPETENCIÁK A TANTÁRGY TELJESÍTÉSÉNEK FELTÉTELEI

1.1.1 Célkitűzés

A tantárgy célja, hogy a hallgató legyen képes gyakorlatcentrikusan használni az SPSS programrendszert. Ismerje meg a program működtetésének interaktív eszközeit. Azonosítsa a statisztikai ismereteit a program funkcionális lehetőségeivel. Legyen képes alkalmazni az SPSS-t a leíró statisztikai alkalmazásokra, hipotézis vizsgálatokra, korreláció és regresszió számításra, faktoranalízisre és klaszteranalízisre.

1.1.2 Kompetenciák:

- A tanulók műveltségének, készségeinek, és képességeinek fejlesztése, ennek alapján az adott tudományterületen a számonkérési eljárások megismertetése
- A pedagógiai értékelés változatos eszközeinek alkalmazása
- Neveléstudományi kutatások fontosabb módszereinek, elemzési eljárásainak alkalmazása,
- A pedagógiai mérés, értékelés változatos eszközeinek alkalmazása.
- Releváns ismeretekkel rendelkezik az elektronikus adatkezelést és a hálózat pedagógiai szolgáltatásait illetően.

Tudás:

- Ismeri az SPSS kezelőfelületét, a változók formai és tartalmi szerepét.
- Feleleveníti és alkalmazza korábbi statisztikai tudását.

Attitűdök:

- A hallgató munkájával a kvantitatív kutatások szükségességét és jelentőségét fogja igazolni, különös tekintettel a számítógépes támogatásra
- Az információs társadalom oktatási alapproblémái ismeretében, a kihívások tudatában legyen képes számonkérési stratégiákat kialakítani és megvalósítani.
- A tananyag elsajátítása során képes a tudás tartalmi és értelmi szintjeit on-line tesztekkel mérni, értékelni.

- Alakuljanak ki azok a nézetek, kompetenciák, amelyek az önellenőrzés, számonkérés, értékelés működtetéséhez és továbbfejlesztéséhez szükségesek.
- Nyitott az új kutatási eredményekre.

Képességek:

- A hallgató képessé válik különböző kutatási eredmények feldolgozására, ill. a kutatói teamben ellátja a számítógépes inkubáció feladatát
- Képessé válnak egy önállóan lefolytatott online teszt empirikus mérési folyamat megtervezésére, kivitelezésére, az eredmények értékelésére, a következtetések levonására.
- Rendelkezik a tanulási folyamatok önellenőrzés, számonkérés pedagógiai módszertani ismereteivel, folyamatszervező és irányító képességekkel.
- Képes on-line számonkérési formákat kezelni, forrásokat felkutatni, és az önálló tanulási folyamatba illeszteni
- Képes a tudásszintet, a tananyag elsajátítását, a hatékonyságot értékelni. Forrásanyagokból – a tananyag tartalmi és értelmi műveletek birtokában – tudjon on-line tesztet megjelenítésre alkalmas formában összeállítani.

1.1.3 A tantárgy teljesítésének feltételei

Önálló kutatási forrásból származó adatok feldolgozása SPSS segítségével, output nézet elkészítése és leadása a kurzus végén

1.2 A KURZUS TARTALMA

1. Bevezetés
2. Adatbevitel a gyakorlatban SPSS szoftverrel
 - Adattípusok
 - A válaszok felvitele
 - Az adatok kiértékelésének csoportosítása
 - Nominális adatok értékelése
 - Nominális adatok kiértékelése kereszt táblával
 - Khi-négyzet-próba
3. Leíró statisztikai értékelés SPSS táblázatkezelővel
 - Leíró statisztika
 - Számított középértékek és helyzeti középértékek
 - Gyakoriság

- Gyakorisági poligon és a középérték-mutatók
 - A középértékek egymáshoz viszonyított kapcsolata
 - Szóródási mérőszámok
4. Matematikai statisztikai lehetőségek SPSS táblázatkezelőkben
 - Korreláció
 - Korrelációanalízis
 - Regressziószámítás
 - Faktoranalízis
 - Parciális korreláció
 - A Spearman-féle rangkorreláció
 - Klaszteranalízis
 5. Magasabb szintű értékelési módszerek a gyakorlatban
 - Hipotézisvizsgálatok
 - Null- és alternatív hipotézisek, döntési szituációk
 - t-próba (Egymintás t-próba, Kétmintás t-próba)
 - Varianciaanalízis
 - A Mann–Whitney-próba, Wilcoxon-próba, Kruskal–Wallis-próba értelmezése
 6. Értékelési eredmények szemléltetésének lehetőségei a táblázatkezelő szoftverekben
 - A diagramok szerepe
 - Diagramtípusok
 - Gyakorisági poligon, hisztogram
 7. Nemparaméteres eljárások
 - Paraméteres és nem paraméteres próbák
 - Kolmogorov-Szmirnov Test
 - Két független minta Mann-Whitney próba
 - Két összetartozó minta Wilcoxon-féle előjeles rangpróba
 - Több független minta egy szempont szerint Kruskal-Wallis próba
 8. ANOVA elmélet kérdései
 - One way ANOVA egyutas variacionális megoldásmenetét.
 9. Multiple Responses
 10. Klaszteranalízis
 - A klaszteranalízis csoportosítása
 - Klaszterosítási módszerek
 - a klaszteranalízis gyakorlati alkalmazása
 11. Faktorelemzés
 - Faktorelemzés alapjai
 - A faktorelemzés alkalmazási területei
 - A faktorelemzés lépései
 - Feladat megoldás értelmezése

1.3 TANULÁSI TANÁCSOK, TUDNIVALÓK

A leckékben – ahol a lecke jellege azt indokolja – talál feladatokat, és önellenőrző kérdéseket. Ezeket a feladatokat és az önellenőrző kérdésekre adott válaszokat *nem kell beküldenie*, viszont azok megválaszolása jelentős mértékben növeli a jobb vizsgaeredmény elérésének esélyeit.

A online-mérés tantárgy tanulásának végső célja a képessé váljon önállóan készített kérdőív kiértékelésére.

A sikeres munkához feltétlenül szükséges, hogy

- Először a tananyag egyes leckéinek elméletét sajátítsa el, mert e nélkül nem fogja érteni a következő lecek anyagát, és nem lesz képes az önálló ismeretszerzésre más kutatásmethodikai irodalomban.
- Olvassa el a jegyzetben található példákat is, melyek segítik a megértést, és próbáljon minden esetben a témához kapcsolódó példát kitalálni, vagy felidézni. Gondolja végig Ön is hasonlóképpen oldotta volna meg a problémát, hasonló kiértékelő módszert választott volna, hasonló következtetést vont volna le?
- Ha a leckéhez tartoznak feladatok, vagy önellenőrző kérdések, oldja, vagy válaszolja meg őket!

1.4 TANANYAG

A 3. fejezetből megismerhetjük az adatok fajtáit, melyek alapvetően meghatározzák, mely kiértékelési statisztikai lehetőségek körét, ezért ez az elméleti alapozás kulcsfontosságú része a tankönyvnek. A fejezet bemutatja az adattípusokhoz kapcsolódó elemzési lehetőségek összefoglalását, külön táblázatban a leíró statisztikai lehetőségek, valamint a matematikai statisztikai elemzési lehetőségeket. Fontos különbséget tenni azon eljárások között melyek célja az adatok közt megtalálni a legmeghatározóbbakat, elkülöníteni az eredményeket befolyásoló és kevésbé befolyásoló értékeket, illetve az összefüggések kimutatására szolgáló eljárások közt. A nominális adatok kiértékelésére viszonylag kevesebb statisztikai mutató áll a rendelkezésünkre, ezek közül a leggyakrabban használtak a fejezetben bemutatásra kerülnek.

A következő fejezet leíró statisztikai mutatók matematikai és statisztikai értelmezéseit veszi számba, majd bemutatja az SPSS táblázatkezelővel történő megvalósításának módszerét. A fejezetben a leíró statisztika mutatók ismertetése a középérték mutatók (Átlag, Medián, Módusz) meghatározásával indul, majd megismerhetjük a gyakorisági mutatókat, melyek közül az abszolút-, relatív-, kumulált-, és kumulált százalékos gyakoriság értelmezését és kivitelezését

tanulhatjuk meg a fejezetből. Érdeemes megnézni a gyakorisági poligon és a középérték-mutatók közti összefüggéseket, valamint a középértékek egymáshoz viszonyított kapcsolatát is. A leíró statisztikai elemzések a szóródási mérőszámok meghatározásával vállnak teljessé, ezért a fejezet bemutatja a szóródási terjedelem, az átlagos eltérés, négyzetes összeg, variancia, szórás, relatív szórás mutatók képletét és használatát is.

Ha nem teljeskörű mintán történt a felmérés, akkor a reprezentatív minta adatai alapján szükséges a matematikai statisztikai műveleteinek végrehajtása, ahhoz hogy választ kapjunk arra a kérdésre, hogy a kapott eredmények a teljes mintára is érvényesek-e, vagy csak a mintát jellemzik. Ennek érdekében a fejezet bemutatja a korrelációs számítás folyamatát, és értelmezését, a korrelációanalízis lehetőségét, valamint a regressziószámítás elméleti és gyakorlati értelmezését. Egy mester képzésen a hallgatóknak ismerni kell a számítások finomítási lehetőségeit, ezért a fejezet kitér a Parciális korreláció, a Spearman-féle rangkorreláció értelmezésére is. A fejezet további részében megismerkedhetünk a faktoranalízis és a klaszteranalízis témakörével.

Külön fejezet foglalkozik a hipotézisvizsgálatok témakörével, mely még mindig a matematikai statisztikai mutatók közé tartoznak. Az olvasó megismeri a null- és alternatív hipotézisek, döntési szituációk fogalmait, lényegét, valamint az egymintás és kétmintás t-próba és a varianciaanalízis értelmezését és feladatokon keresztül történő megvalósítását. A fejezet a Mann–Whitney-próba, Wilcoxon-próba, Kruskal–Wallis-próba értelmezésével zárul.

A kapott eredmények szemléltetése külön témakör, hiszen hasznosságuk miatt érdemes ismerni a lehetőségeket. A fejezet bemutatja a diagramok fontosságát, és az egyes diagramtípusok alkalmazási lehetőségeit. A gyakorisági poligon és hisztogram külön témakörként is feldolgozásra.

2. ADATBEVITEL A GYAKORLATBAN SPSS SZOFTVERREL

2.1 CÉLKITŰZÉSEK ÉS KOMPETENCIÁK

A fejezetben a kérdőívkészítés kiértékelésének kezdő lépéseit ismerheti meg az olvasó. Az egyik legfontosabb elem, hogy hogyan alakítjuk ki a válasz-adatok felvitelére szolgáló adattáblát. Végig kell gondolni, hogy ahol több választ is adhattak a kitöltők, ott minden választ külön kérdésként kell kezelni, és annyit kell csak eltárolni, hogy megjelölte-e vagy sem a kitöltő, esetleg hogy hányadik helyre rangsorolta. Azt is végig kell gondolni, hogy kódokat vigyünk-e fel, vagy a szöveges válaszokat, és mi történjen, ha nem válaszolt a kitöltő.

A fejezet bemutatja azt is, mire kell figyelni, hogy a helyes adatfelvitellel megalapozzuk a további kiértékeléseket.

2.2 TANANYAG

Adattípusok

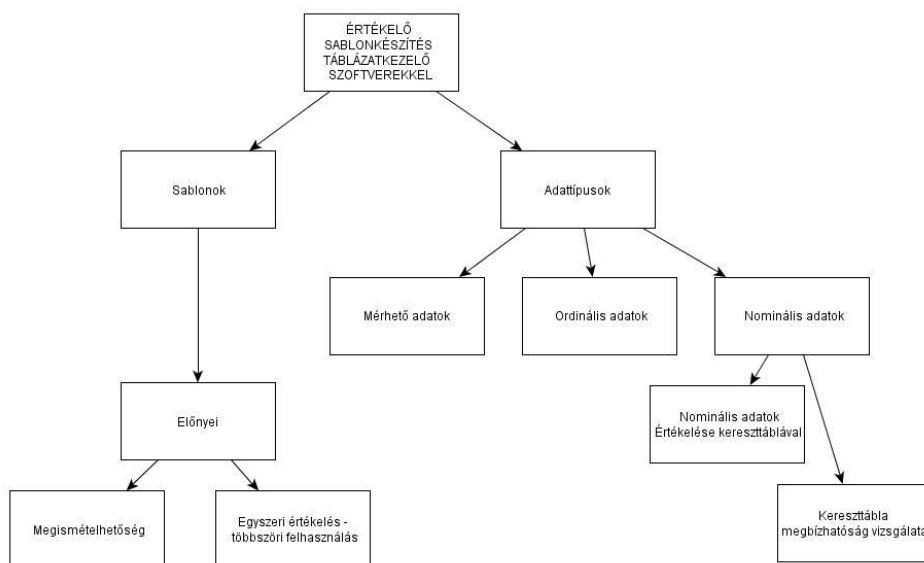
A válaszok felvitele

Az adatok kiértékelésének csoportosítása

Nominális adatok értékelése

Nominális adatok kiértékelése kereszttáblával

Khi-négyzet-próba



1. ábra: Fogalomtérkép

2.2.1 Sablonszerű gondolkodás

Bármely területen dolgozzunk is, a XXI. századra jelentősen megnőtt a mérések szerepe. Az ipar világában vevői elégedettségméréseket kell végezni, a közsférában a fenntartó elégedettségét mérjük, növekszik a minőségi díjak szerepe (az év könyvtára, Minőségi díj a felsőoktatásban), melyekhez a folyamatos mérések elengedhetetlenek, és gyakoriak a közvélemény-kutatási felmérések is.

Mégis az oktatás területén a leglátványosabb a mérés fejlődése! Az évenként megállapításra kerülő kompetenciamérések feladatsorokkal mérik, hogy képesek-e a tanulók a tudásukat az életben alkalmazni és minden ilyen felméréshez kapcsolódik háttérkérdőív. A felsőoktatásban terjed a kurzusértékelés, a diplomás pályakövetés, melyek mindegyike ugyanazon kérdőív kitöltését jelenti más-más években, valamint egyre jelentősebb a nemzetközi mérésekben való folyamatos részvételünk is.

Ezért érdemes a kérdőívek kiértékelését úgy elkészíteni, hogy a következő években csupán az adatok munkalapon kelljen cserélni a nyers adatokat. Míg erre az EXCEL táblázatkezelő szoftvernél tudatosan nekünk kell odafigyelni, addig egy professzionális táblázatkezelő, mint például az SPSS alapvetően így gondolkodik, és külön tárolja az adatokat, valamint külön objektumként a kiértékeléseket.

2.2.2 Adattípusok

 **Az adatoknak három típusát különböztetjük meg:**

 **Mért adatoknak vagy intervallumadatoknak nevezzük a mennyiségi tartalommal bíró számokat.**



Például az alábbi kérdésekre választ adó adatokat: Hány cm magas? Melyik évben született? Hány könyvet olvasott el a múlt hónapban?

 **Nominális adatokról beszélünk, amikor a válaszokat számokkal, jelekkel helyettesítjük.**

Ez esetben a számoknak nincs mennyiség tartalmuk, nem lehet őket összeadni, átlagolni, helyette megszámlálhatjuk, hányan adták az adott választ. Nagyon fontos, hogy ez esetben a számok sorrendiséget nem jelentenek, azonban a kérdés megszerkesztése során törekedni kell az elfogadott normák/szokások betartására.



Például a neme kérdésnél az 1-es jelölje a férfit és a 2-es a nőt, mivel a személyi számunk használata során ezt szoktuk meg.

 **Ordinális adatok a sorrendiséget jelölő számok.**



Például: Hányadik lett Magyarország a PISA-felmérésen; „Rangsorolja, hogy a következő tulajdonságok közül melyek a legjellemzőbbek Önre...” jelleget kérdésekre adott válaszok.



Elemezzük ki az alábbi kérdőívet, megismerve a kiértékelési módszereket!



Kérdőív

1. Neme:
 1. Férfi
 2. Nő
2. Legmagasabb iskolai végzettsége?
 1. 8 általános vagy kevesebb
 2. Szakmunkás, szakiskolai végzettség
 3. Érettségi
 4. Főiskolai diploma
 5. Egyetemi diploma

6. Magasabb végzettség
3. Életkora?.....
4. Az Ön lakhelye?
 - Főváros
 - Vidéki város
 - Falu
5. Családjában lévő eltartott gyermekek száma?.....
6. A család nettó jövedelme?.....Ft
7. A család megtakarítása?.....Ft

A kérdőív kiértékelését a következő fejezetekben az elméleti anyaggal párhuzamosan találjuk.

2.2.3 A válaszok felvitele

Míg Excel szoftver alkalmazása során a kiértékelés első lépéseként begéptük a válaszokat, addig a SPSS szoftverek használata során az első lépés a változók megadása.

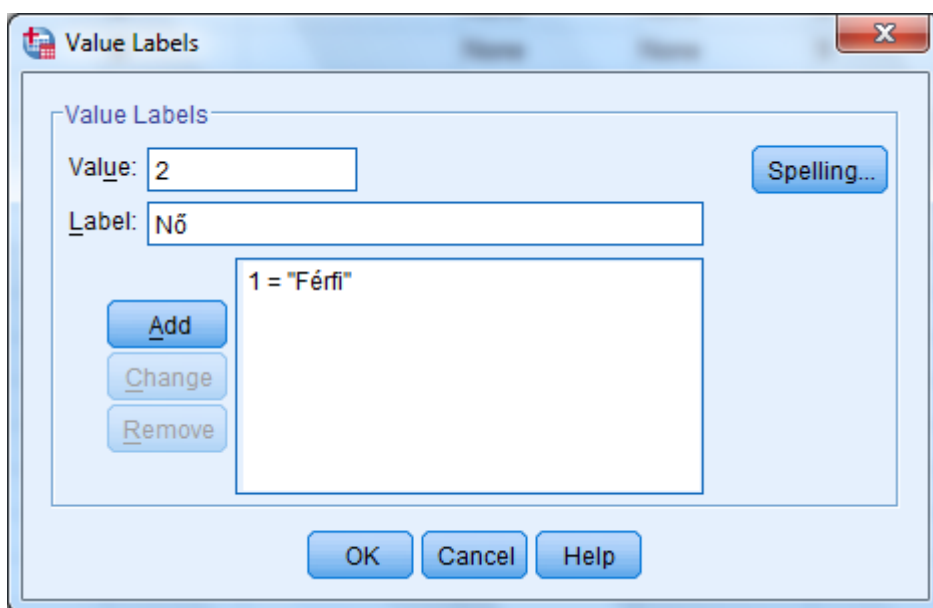
Természetesen, ha online kérdőívkészítő programmal dolgoztunk, vagy szkennelő programmal vesszük gépre az adatokat, akkor az adatokat készen kapjuk, ha van SPSS kimenete a szoftvernek. Azt azonban érdemes ez esetben is ellenőrizni, hogy a számadatokat tartalmazó mezőket tényleg számnak tekinti-e a táblázatkezelő, vagy pedig szöveggként kerültek be az adatok, illetve általánosságban érdemes átellenőrizni, hogy megfelelő változóknak azonosította-e be a szoftver a változókat, mert sajnos ez gyakran nem valósul meg.

A mintaként megadott kérdőív esetén a kérdéseket a következőképpen deklarálni:

Az első kérdésre két válasz adható a férfiak 1-s, míg a nők 2-st jelölnek meg. Az SPSS-ben elérhető változótípusok közül a legmegfelelőbb a NUMERIC, ahol lehetőségünk van a változó értékeinek deklarálására is. A változó megadásakor:

1. adjunk meg egy változó nevet (Pl. Neme), majd
2. beállíthatjuk a változó típusát a RESTRICTED NUMERIC-re.
3. Ezt követően a változó szélességét állítsuk 1-ra,
4. 0 tizedes jeggyel.
5. Label alkalmazása nem szükséges, mert a változó név most reprezentálja a mezőt, de vannak esetek, amikor érdemes a bővebb magyarázat elhelyezése, mint például egy ÉV változó

- zónévnél, nem mindegy hogy a születési évre, vagy az életkorra gondolunk.
6. A VALEU cellájában hozzuk elő a paraméterek beállítására szolgáló ablakot, és a VALEU sorába írjuk a nominális kódot, LABEL sorába pedig a hozzátartozó megnevezést, mint ahogy a képen is látható.



2. ábra: Címkezés

A kérdőív következő nominális kérdése esetén is járunk el hasonlóan.

A számadatokat tartalmazó kérdések esetén pl. „Hány éves?” a NUMERIC változót alkalmazzuk, melynél a szélességet és a tizedes jegyek számát kell megadni. Ez utóbbit állítsuk nullára, a tizedesjegyek számát pedig 3-ra, mert ma már nem ritka a 100 év feletti életkor.

A különbség a negyedik kérdés esetén lesz Excelhez viszonyítva: ahhoz, hogy további elemzést tudjunk végezni, az eddig alkalmazott REGSTICTED NUMERIC változó típusban rendeljünk nominális kódokat a válaszokhoz, még akkor is, ha a kérdőívben ez nem szerepelt.

A kérdőív többi kérdése a fenti módszerekkel felvihető.

Speciális kérdések válaszainak rögzítése

- Több válasz esetén minden egyes alkategória külön „kérdésként” fog szerepelni, ahol a válasz annyi, hogy megjelölésre került-e az az alpont, vagy sem.
- A „Tegye sorrendbe az alábbi szempontokat!” jellegű kérdések minden egyes szempontja külön kérdésként kerüljön felvitelre, ahol a válasz a rangsor száma, a változó típusa NUMERIC.

A változókeret elkészítése után rögzíthetjük a kérdőív kitöltők válaszait, mely a mintában 30 ember válaszát jelenti.

	Sorszam	Neme	Iskolai_vegzetseg	Eletkor	Lakhely	Eltartott_gyerek	Csalad_jovedelme	Megtakarítás	var	var	var
8	8	2	4	23	2	1	233000	23000			
9	9	1	6	55	1	0	131000	30000			
10	10	1	3	57	3	0	151000	30200			
11	11	1	6	43	2	2	128000	25600			
12	12	1	6	40	3	0	95000	19000			
13	13	2	3	51	2	1	136000	27200			
14	14	1	3	56	2	0	125000	25000			
15	15	2	5	58	1	1	128100	25620			
16	16	2	4	29	1	2	256900	0			
17	17	2	6	29	2	3	321000	10000			
18	18	1	3	48	3	3	148700	29740			
19	19	2	3	53	3	1	163400	32680			
20	20	2	4	18	3	2	152600	0			
21	21	1	2	57	3	0	131000	32000			
22	22	1	6	36	1	1	242900	12000			
23	23	1	4	51	1	1	156900	40000			
24	24	2	2	40	1	3	163500	0			
25	25	1	5	47	3	1	198700	39740			
26	26	1	6	39	1	2	250000	50000			
27	27	2	4	19	2	4	189000	0			
28	28	1	2	19	2	3	242500	0			
29	29	2	3	35	3	0	220000	44000			

3. ábra: Kérdőíves válaszok

2.2.4 Az adatok kiértékelésének csoportosítása

Nézzünk meg a kiértékelés megkezdése előtt egy összefoglaló táblázatot arról, hogy milyen típusú adatokat milyen statisztikai módszerekkel értékelhetünk!

Az alábbi táblázatot Falus Iván – Ollé János klasszikus statisztikai módszereket bemutató könyvének új kiadásában¹ találjuk, amely logikusan, áttekinthető módon ismerteti az eljárásokat, melyek kiértékelését a következőkben megismerhetünk:

Leíró statisztikai elemzéseknek a minta elemeinek elemzésére szolgáló módszerek összességét nevezzük.

1. Leíró statisztikai mutatók:

Gyakoriságok	Középértékek	Szóródások	Korreláció
abszolút	átlag	szóródási terjedelem	korrelációs számítás
relatív (százalékos)	módusz	interkvartilis félterjedelem	
kumulatív	medián	átlagos eltérés	
		variancia	
		szórás	
		relatív szórás	

A matematikai statisztika választ ad arra, hogy a reprezentatív mintából vonható-e le következtetés az alapsokaságra.

A matematikai statisztikai vizsgálatokat két csoportba sorolhatjuk:

1. **Különbözőségvizsgálatok**, melyek célja az adatsorok közti különbségek kimutatása.
2. **Összefüggés-vizsgálatok**, melyek célja az adatsorok közti kapcsolatok feltárása.

2. Különbözőség vizsgálatok

Adatfajták Minták száma	intervallum	ordinális	nominális
egy	egymintás t-próba	Wilcoxon-próba	Keresztábra-elemzés, khi-négyzet-próba
kettő	kétmintás t-próba F-próba	Mann-Whitney-próba	Keresztábra-elemzés, khi-négyzet-próba
három	varianciaanalízis	Kruskal-Wallis-próba	Keresztábra-elemzés, khi-négyzet-próba

¹ FALUS Iván – Ollé János (2008): Empirikus kutatások gyakorlata. Nemzeti Tankönyvkiadó, Budapest, p. 138.

3. Összefüggésvizsgálatok

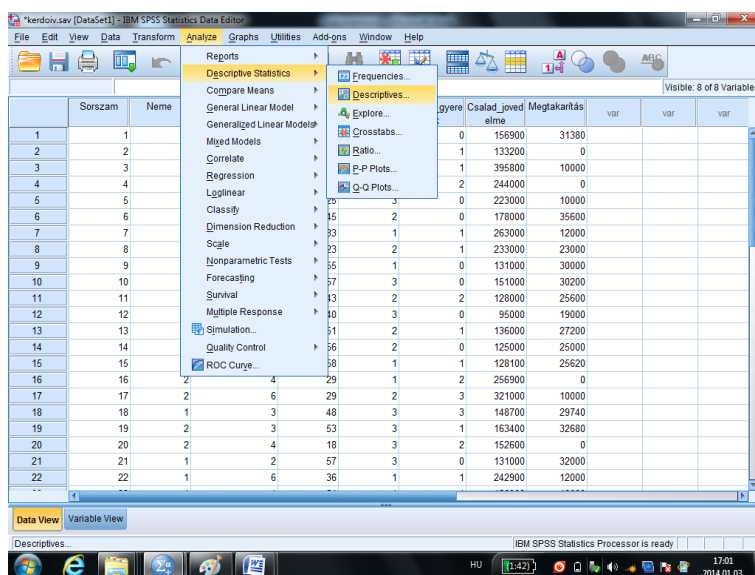
Adatfajták Minták száma	intervallum	ordinális	nominális
kettő	korrelációs számítás	Spearman-féle rangkorreláció	Keresztábla-elemzés, khi-négyzet-próba
Kettő vagy több mint kettő	regresszióanalízis		
Több mint kettő	Parciális korreláció- számítás Faktoranalízis Klaszteranalízis		

2.2.5 Nominális adatok értékelése

Nominális adatok kiértékelése során van lehetőségünk az adott választ adók megszámlálására, illetve a válaszolók arányának meghatározására az „összesen”-hez képest.

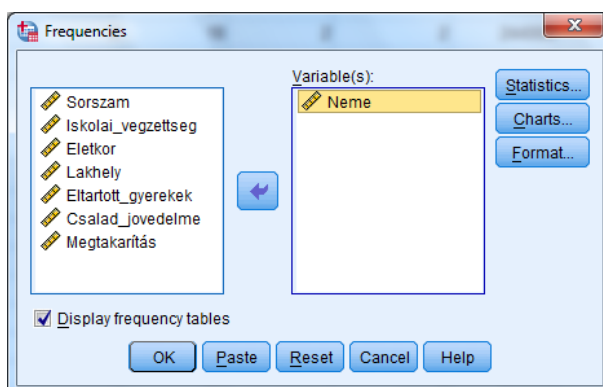
Első lépésként kezdjük a klasszikus „Kérem, adja meg a nemét!” kérdés kiértékelésének áttekintésével.

Az SPSS szoftverben az ANALYZE/DESCRIPTIVE menüpont első parancsát a FREQUENCY-t választjuk ki:



4. ábra: Gyakoriság parancs helye

A megjelenő panelen felkínálja az összes változót, tegyük az elemzendő adatsort a Variable(s) panelre.



5. ábra: Gyakoriság adatpanel

A megjelenő eredmény az SPSS output1 nevű külön ablakában kerül megjelenítésre, ahol az egyes kategóriákba tartozó gyakorisági értékeken kívül láthatjuk a válaszok megoszlását a PERCENT oszlopban (mely automatikusan előállításra kerül, míg excelben képlet alkalmazásával kellett meghatározni). A megjelenő eredménytábla a kumulatív gyakorisági értéket is megadja százalékos formában a COMULATIVE PERCENT oszlopban.

FREQUENCIES VARIABLES=Neme
/ORDER=ANALYSIS.

[DataSet1] C:\Tunde_SPSS\kerdoiv.sav

Statistics

Neme		
N	Valid	30
	Missing	0

Neme				
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Férfi	16	53,3	53,3	53,3
Nő	14	46,7	46,7	100,0
Total	30	100,0	100,0	

6. ábra: Gyakorisági elemzés nominális adatok esetén

Mint látható az eredmény ablak először megadja, hogy összesen hányan vettek részt a felmérésben, és a hiányzó adatok számát, majd külön táblázatban láthatjuk az előfordulási adatokat.

✿ Értékeljük ki a kérdőív második kérdését!

A kérdőív második kérdése a családi állapotra vonatkozik:

Az előző feladathoz hasonlóan ANALYZE/DESCRIPTIVE menüpont első parancsát a FREQUENCY-t parancsát kell használnunk.

Ehhez váltsuk vissza az ablakot az adattáblához, és a megjelenő panelen vegyük ki az előbb használt változót a Variables(s) ablakrészről, és tegyük a helyére a második kérdésünk mezőnevét.

Az eredményt tartalmazó Output1 oldal kiegészül az új elemzési táblázattal, de közben az előző kimutatás sem vesz el.

The screenshot shows the IBM SPSS Statistics Viewer window with the following content:

Output tree on the left:

- Log
- Frequencies
 - Title
 - Notes
 - Active Dataset
 - Statistics
 - Name
- Log
- Frequencies
 - Title
 - Notes
 - Active Dataset
 - Statistics
 - Iskolai_vegzettség

Frequency Table:

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Férfi	16	53,3	53,3	53,3
Nő	14	46,7	46,7	100,0
Total	30	100,0	100,0	

Command:

```
FREQUENCIES VARIABLES=Iskolai_vegzettség
/ORDER=ANALYSIS.
```

Statistics

Iskolai_vegzettség

N	Valid	Missing
	30	0

Iskolai_vegzettség

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 8 általánosnál kevesebb	4	13,3	13,3	13,3
8 általános	6	20,0	20,0	33,3
Szakmunkás vagy szakiskolai	8	26,7	26,7	60,0
Érettségi	5	16,7	16,7	76,7
Technikum	2	6,7	6,7	83,3
Felsőfokú végzettség	5	16,7	16,7	100,0
Total	30	100,0	100,0	

7. ábra: Nominális adatok gyakoriság elemzése több táblázat esetén

2.2.6 Nominális adatok kiértékelése kereszttáblával

A nominális adatok kiértékelésére elég kevés statisztikai eszközünk van. A kereszttábla segítségével jól használható összefüggéseket tudunk kimutatni, és alkalmas a nominális adatok kiértékelésére.

A kereszttábla készítésére az ANALYZE/DESCRIPTIVE menüpont CROSSTABS parancsát-t válasszuk ki:

Végig kell gondolni, hogy mit szeretnénk látni oszlopfeliratként, például ha a „Lakhely” kérdésre adott válaszokat, akkor annak feliratát húzzuk az ROW(s) részre. Ennek hatására fognak megjelenni a válaszok a sorokban (az adattábla nem szöveges adatokat tartalmaz, de az eredmény tábla a helyettesítő szövegeket fogja megjeleníteni).

Hasonlóan adjuk meg a Column(s) oszlop mezőit is.

Az eredményként megjelenő Output tartalmaz egy Case Processing Summary ablakot, mely a két adatsor érvényes és hiányzó adatainak számát és arányát, majd a kereszttáblát.

Crosstabs

[DataSet1] C:\Tunde_SPSS\kerdoiv.sav

Case Processing Summary


	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Lakhely * Neme	30	100,0%	0	0,0%	30	100,0%

Lakhely * Neme Crosstabulation

Count		Neme		Total
		Férfi	Nő	
Lakhely	Főváros	5	5	10
	Város	5	5	10
	Falu	6	4	10
Total		16	14	30

8. ábra: Kereszttábla

Keresztábra-készítés során az egyik kérdésre adott válaszokat összevetjük a másik kérdésre adott válaszokkal.

 **Keresztábrának vagy kotigenciatábrának nevezzük azt a táblázatot, melynek oszlopait és sorait két nominális változó határozza meg.**

A keresztábrák szimmetrikusak, így az oszlopokat és sorokat inkább áttekinthetőség szempontjából válasszuk, mint az elemzés tartalma szerint.



Például: hány vidéken lakó nő töltötte ki a kérdőívünket?

➔ Crosstabs

[DataSet1] C:\Tunde_SPSS\kerdoiv.sav

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Neme * Lakhely	30	100,0%	0	0,0%	30	100,0%

Neme * Lakhely Crosstabulation

		Lakhely			Total
		Főváros	Város	Falu	
Neme	Férfi	5	5	6	16
	Nő	5	5	4	14
Total		10	10	10	30

9. ábra: Keresztábra

Ha megvizsgáljuk az eredményként létrejövő táblázatot, látható, hogy a felmérésben részt vevő férfiak és nők megoszlása nem tér el a településtípusra vonatkozóan.

Vizsgáljuk meg az iskolai végzettség és nemek viszonyát lakhelytípusonként. Az látható, hogy a felmérésben részt vevő férfiak magasabban kvalifikáltak, mint a nők, hiszen a magasabb iskolai végzettségnél magasabb létszámmal jelennek meg a férfiak, mint a nők.

➔ Crosstabs

[DataSet1] C:\Tunde_SPSS\kerdoiv.sav

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Neme * Lakhely	30	100,0%	0	0,0%	30	100,0%

Neme * Lakhely Crosstabulation

		Lakhely			Total
		Főváros	Város	Falu	
Neme	Férfi	5	5	6	16
	Nő	5	5	4	14
Total		10	10	10	30

10. ábra: Keresztábra

2.2.7 Khi-négyzet-próba

Keresztábra szignifikanciavizsgálata

Nominális adatok esetén használandó statisztikai módszer, amely választ ad arra a kérdésre, hogy az adatok megoszlása a véletlen következménye-e vagy a populációra jellemző tulajdonság.

Ha a példában megvizsgáljuk, hogy milyen az iskolai végzettség lakóhelyek szerinti megoszlása, akkor azt láthatjuk, hogy minél nagyobb a település típusa, annál több a kvalifikált személy.

➔ Crosstabs

[DataSet1] C:\Tunde_SPSS\kerdoiv.sav

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Iskolai_vegzettség * Lakhely	30	100,0%	0	0,0%	30	100,0%

Iskolai_vegzettség * Lakhely Crosstabulation

Count		Lakhely			Total
		Főváros	Város	Falu	
Iskolai_vegzettség	8 általánosnál kevesebb	0	1	0	1
	8 általános	1	1	2	4
	Szakmunkás vagy szakiskolai	1	2	5	8
	Érettségi	2	2	1	5
	Technikum	1	1	1	3
	Felsőfokú végzettség	5	3	1	9
Total		10	10	10	30

11. ábra: Keresztábra- Iskolai végzettség és lakhely

Vajon ez csak a megkérdezett emberek, azaz a minta esetén igaz, vagy általánosíthatunk a teljes populációra?

A választ a kérdésre a Khi-négyzet-próba adja.

Ahhoz, hogy eldönthessük, csak a felmérésben résztvevőkre igaz-e ez az állítás, vagy pedig általánosíthatunk, és elmondhatjuk, hogy a kvalifikált személyek „magasabb” településtípusokon laknak, végezzük el a szignifikancia-vizsgálatot!

Khi-négyzet-próba

Első lépésként létre kell hozni a tapasztalt értékek táblázata alapján a „várt táblázatot”. **A szignifikanciavizsgált nullhipotézise** az, hogy a tapasztalt és a várt érték táblázat adatai között nincs jelentős különbség. Ha mégis jelentős különbséget tapasztalunk, annak oka, hogy a keresztábra két nominális adata között van valamilyen kapcsolat.

Ahol a tapaszt értékek táblázata adott, a várt értékek táblázata alatt a következőt értjük:

A várt érték táblázat meghatározásban a sorösszegekkel és az oszlopösszegekkel kell számolni. Minden cella érték helyett az adott sorösszeget szorozzuk az adott oszlopösszeggel, melynek eredményét osszuk a végösszeggel.

A folyamat folytatásként a várt tábla adatainak meghatározás után cellánként venni kell a kapott és a várt érték közti különbség négyzetét, és osztani kell a cellában lévő várt értékkel.

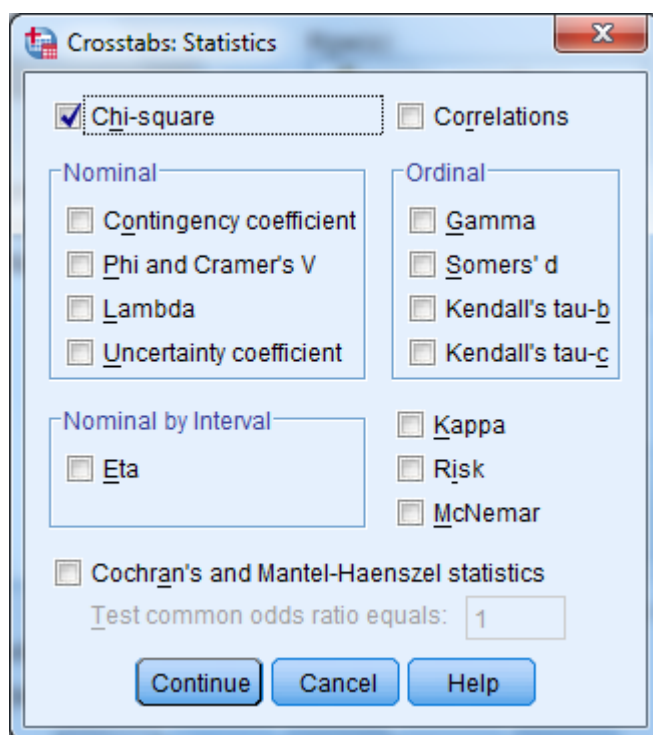
Ezeket a hányadosokat összeadva kapjuk meg a **khi-négyzet-értéket**, melyet össze kell vetni a khi-négyzet-próba szignifikanciátáblázatának megfelelő elemével.

A táblázatban a 95%-os valószínűségi szint oszlopát és a táblázat szabadságfokához tartozó sor metszeténél lévő értéket keressük meg.

A szabadságfok az alapul szolgáló táblázatban lévő sorok számától eggyel kisebb és az oszlop számától eggyel kisebb érték szorzata. (Példánkban $5 \cdot 2$).

- Ha a táblázatban lévő érték kisebb, mint az általunk számolt khi-négyzet-érték, abban az esetben a két táblázat közti különbség nem a véletlennek köszönhető.
- Ha a táblázatban lévő érték nagyobb, mint az általunk számolt khi-négyzet-érték, abban az esetben a két táblázat közti különbség csupán a véletlen műve, nem tudunk kapcsolatot felfedezni benne.

A SPSS szoftver használatával ezt a folyamatot nem kell végig vezetnünk, helyette néhány kattintással megkapjuk a szignifikancia táblázatot. Előállításához adjuk ki az ANALYZE/DESCRIPTIVE/CROSSTABS parancsot, ahol az oszlopok és sorok –ba kerülő mezők megadása után kattintsunk a STATISTICS gombra, és kapcsoljuk be, a Chi-square választógombot.



12. ábra: Khi-négyzet próba beállítása

Az eredményként előálló 8,817 –es Khi-négyzet értéket, ha Khi-négyzet táblázatot használnánk, akkor a táblázatban a szabadság fok (3 lakhely kategória-1* 6 iskolai végzettség kategóriája-1), azaz a 10. sorának és a 95%-os valószínűségi szint találkozási pontjánál lévő érték a khi-négyzet-próba táblázatában: 18,307-es értékhez kell viszonyítanunk.

Az általunk kapott érték 8,817 kisebb, mint a táblázatban szereplő érték, tehát az, hogy a városban lakók iskolai végzettsége magasabb a falusiaknál, a fővárosiaké pedig a legmagasabb, nem került igazolásra a felmérésben résztvevők adatai alapján.

Az SPSS-ben megkapott keresztábra adatai alapján azonban nem kell elővenni a Khi-négyzet táblát. A táblázat első oszlopa (*Value*) megadja a konkrét Khi-négyzet értéket, a *df* jelenti szabadságfokot, az utolsó (*Asymp.Sig.*) oszlop a szignifikancia szintet. A táblázatban az látható, hogy az adott Khi-négyzet érték, 550-s szignifikanciaszinten értelmezhető. Ez a jelölés mód a 0,550-os értéket jelenti (a tizedes vessző előtti nulla nem kerül az SPSS eredmény táblázataiban kiírásra) azt, hogy 55,0%-os szignifikanciaszinten igaz az állításunk. A pedagógiai kutatások során a 95%-os vagy annál magasabb szignifikanciaszintet fogadjuk el

érvényesnek, így a kapott értéket úgy kell értelmezni, hogy a táblázatunk tapasztalható eltérést, szabályszerűséget a fővárosi-vidéki városi-falu és az iskolai végzettség között csak a vizsgált minta esetén tapasztalható, a teljes populációra nem terjeszthetjük ki érvényességét.

2.3 ÖSSZEFOGLALÁS, KÉRDÉSEK

2.3.1 Összefoglalás

A fejezetben áttekintésre került a kérdőív kitöltetése utáni folyamat kezdete. Az adatok felvitele után az adattípusoknak megfelelő kiértékelési módszert ki kell választani.

A fejezetben a válaszok felvitelének specialitásain túl a megismerhettük az adatok validálásának fontosságát, valamint a nominális adatok kiértékelésének technikáját.

Nominális adatok esetén is van lehetőség használni olyan próbákat, melyek segítenek annak eldöntésében, hogy a kereszttáblában megkapott adatok csak a mintára érvényesek-e, vagy a teljes populációt jellemzik. Ez a khi-négyzet-próba.

2.3.2 Önellenőrző kérdések

1. Sorolja fel az adattípusokat!
2. Hogyan határozzuk meg a változókat? Mondjon példákat!
3. Mikor használunk kereszttáblát?
4. Gondolja végig a nominális adatok szignifikanciavizsgálatának lépéseit!
5. Mikor tekinthető a khi-négyzet-értéke szignifikánsnak?

3. LEÍRÓ STATISZTIKAI ÉRTÉKELES SPSS TÁBLÁZATKEZELŐKKEL

3.1 CÉLKITŰZÉSEK ÉS KOMPETENCIÁK

A fejezet célkitűzése: az online (vagy hagyományos papíralapú) kérdőív mérhető adatainak statisztikai, SPSS táblázatkezelő szoftverrel történő kiértékelését megismertetni a hallgatókkal.

A fejezetben a leíró statisztikai elemzésekkel ismerkedhetünk meg. Az elsajátítás hatékonyságának növelése érdekében konkrét kérdőív kérdéseinek kiértékelésén keresztül ismerhetjük meg a statisztikai módszereket.

3.2 TANANYAG

Leíró statisztika

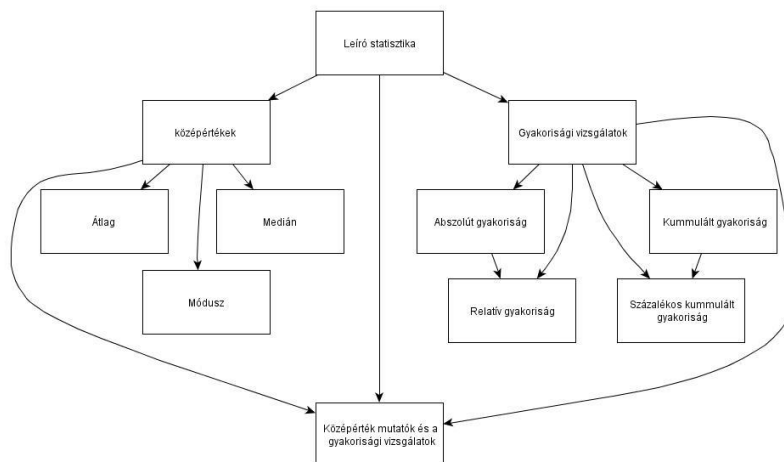
Számított középértékek és helyzeti középértékek

Gyakoriság

Gyakorisági poligon és a középérték-mutatók

A középértékek egymáshoz viszonyított kapcsolata

Szóródási mérőszámok



13. ábra: Fogalomtérkép

3.2.1 Leíró statisztika

A leíró statisztikai elemzéseket minta elemeinek elemzése esetén végezzük. Ha teljeskörű mintavételt alkalmaztunk, azaz a populáció megegyezik a mintával, akkor a leíró statisztikai elemzések szintjén meg is állhatunk. Ha azonban a minta a populáció egy része, akkor szükség lesz a további elemzésekre. A matematikai statisztikai vizsgálatok esetén is elvégezhetjük előbb a leíró statisztikai elemzéseket, melynek mutatói segítik a minta jellemzését, viszont a kapott értékek általánosítására nem adnak információt.

Az előző fejezetben láthattuk a statisztikai mutatókat összefoglaló táblázatot, most közelebbről ismerkedjünk meg a

- középérték-mutatókkal,
- szóródási mutatókkal,
- gyakorisági mutatókkal.

A leíró statisztikai mutatók közül több van egymással kapcsolatban. A fejezetből megismerhetjük, az egyes mutatókból hogyan lehet következtetni a többi statisztikai mérőszámra.

3.2.2 Számított középértékek és helyzeti középértékek

Számtani átlag

A leggyakrabban használt középérték-vizsgálat a számtani közép meghatározása.

☞ **Számtani átlagnak, más néven számtani középnek nevezzük a minta elemeinek összeadásából és a minta elemszámával történő osztásából származó értéket.**

$$\bar{x} = x_1 + x_2 + \dots + x_n = \frac{\sum_{i=1}^n x_i}{n}$$

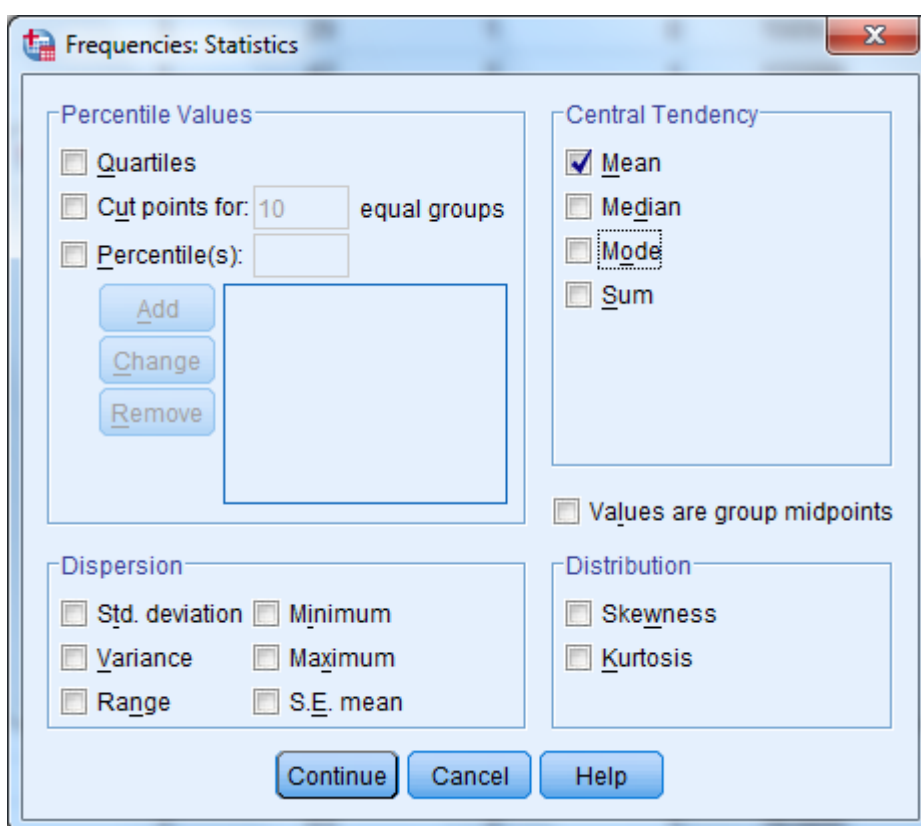
14. ábra: Átlag képlete

A definícióból adódik, ha vesszük az egyes elemek átlagtól való eltéréseinek összegét, az eredmény nulla lesz.

A táblázatkezelők használata természetesen leegyszerűsíti a művelet végrehajtás, az SPSS szoftvernél az átlag kiszámítására a MEAN függvény fog szolgálni, melyet több parancsnál is beállíthatunk, hogy megjelenítésre kerüljön.

Feladat: Elemezzük az előző fejezetben bemutatott kérdőív harmadik kérdését: azaz a „Hány éves?” kérdésre adott válaszokat. Határozzuk meg, milyen átlagos életkorú emberek töltötték ki kérdőívünket!

ANALYZE/DESCRIPTIVE STATISTIC/FREQUENCIES parancsnál tegyük az életkor oszlopot az elemzendő Variables(s) ablakba, és az OPTIONS nyomógombot használva megjelenő panelen kapcsoljuk be a MEAN (Átlag) választógombot.



15. ábra: Átlag meghatározása

Az ÁTLAG meghatározása önmagában soha nem elegendő egy minta jellemzésére. Az ÁTLAG egy felületes mérőszám, mely elfedi a minta összetételéből eredő eltéréseket, ezért meghatározása után mindig tovább kell folytatni az elemzést a többi középérték-mutató, illetve a szóródási mutatók meghatározásával!



Nézzünk meg egy szélsőséges példát! Gondoljunk arra, ha két osztályban eltérő matematikatanár oktatja a gyerekeket, akik közül az egyik folyamatosan versenyre viszi a gyerekeket, ahol országos eredményeket érnek el, de csak az osztály egy részével, míg a többi tanuló éppen csakelkerüli a bukást matematikából. Az osztályban született jegyek a két végletnek felelnek meg, az osztályátlag közepes körüli.

A másik osztályban azt az elvet érvényesíti az oktató, hogy inkább lassabban haladjanak, de amit megtanulnak, azt az utolsó ember is megértse. A többség teljesítménye közepes környékén van. Az osztályátlag hasonló az előző csoportéhoz! Mégsem lehet azonosan értékelni a két osztály teljesítményét, de az átlagot tekintve ez a különbség nem látható.

Módusz



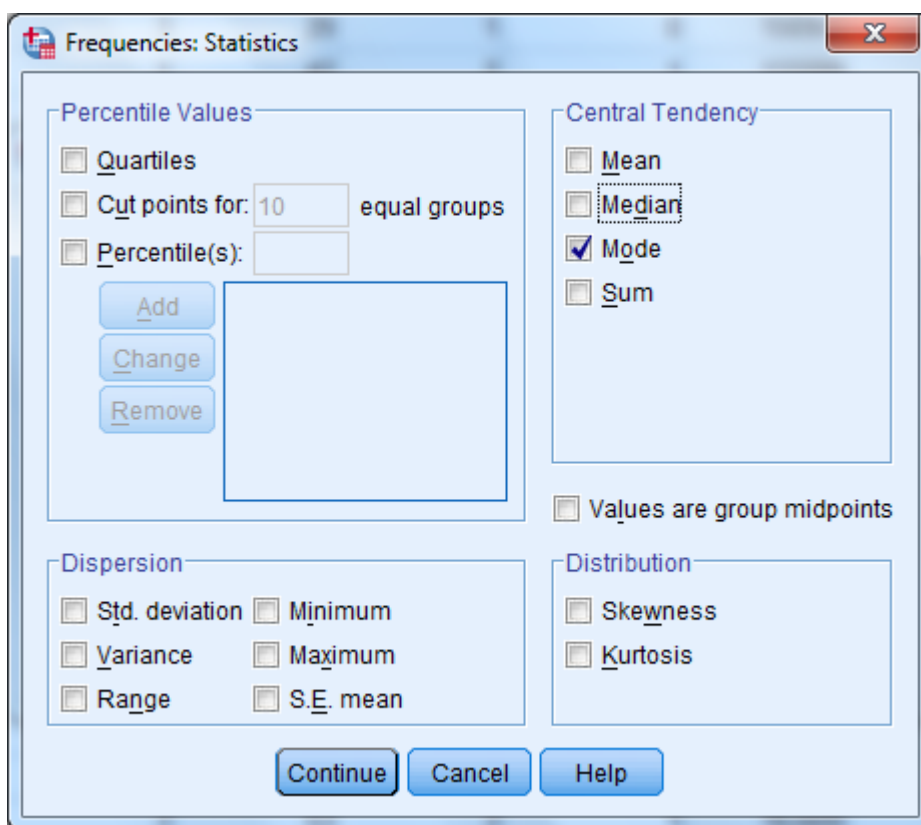
A módusz a minta adatai között a leggyakrabban előforduló elem.

Könnyen előfordulhat azonban, hogy nincs olyan eleme a mintának, amely gyakrabban fordul elő, mint a többi, vagy több elemnek is egyforma az előfordulása. Ezen esetekben nem rendelkezik módusszal a minta, hisz nem tudunk olyan elemet kiválasztani, melynek a gyakorisága nagyobb, mint a többié. Míg Excelben a „#Hiányzik” érték kerül a cellába, addig SPSS szoftver a többször előforduló módusz közül a kisebbet adja meg egy kiegészítéssel, hogy több módusz van, és a legkisebb került megjelenítésre. („Multiple modes exist. The smallest value is shown.”)

Megjegyzés: Abban az esetben is ezt a megoldást látjuk, ha minden elem például csak egyszer fordul elő.

Határozza meg a kérdőívet kitöltők életkorának móduszát!

SPSS parancsa a MODE, melyet az átlaghoz hasonló módon az ANALYZE/DESCRIPTIVE STATISTIC/FREQUENCIES parancsnál adhatunk meg. Tegyük az életkor oszlopot az elemzendő Variables(s) ablakba, és az OPTIONS nyomógombot használva megjelenő panelen kapcsoljuk be a MODE (Módusz) választógombot.



16. ábra: Módusz megadása

Medián

A medián a minta közepe, azaz ugyanannyi elem nagyobb nála, mint ahány elem kisebb. Ha ábrázoljuk a minta gyakorisági eloszlását, akkor a medián értékéhez húzott függőleges vonal felezi a gyakorisági görbe területét.

 **A medián a minta elemeinek sorba rendezése után a középső elem.**

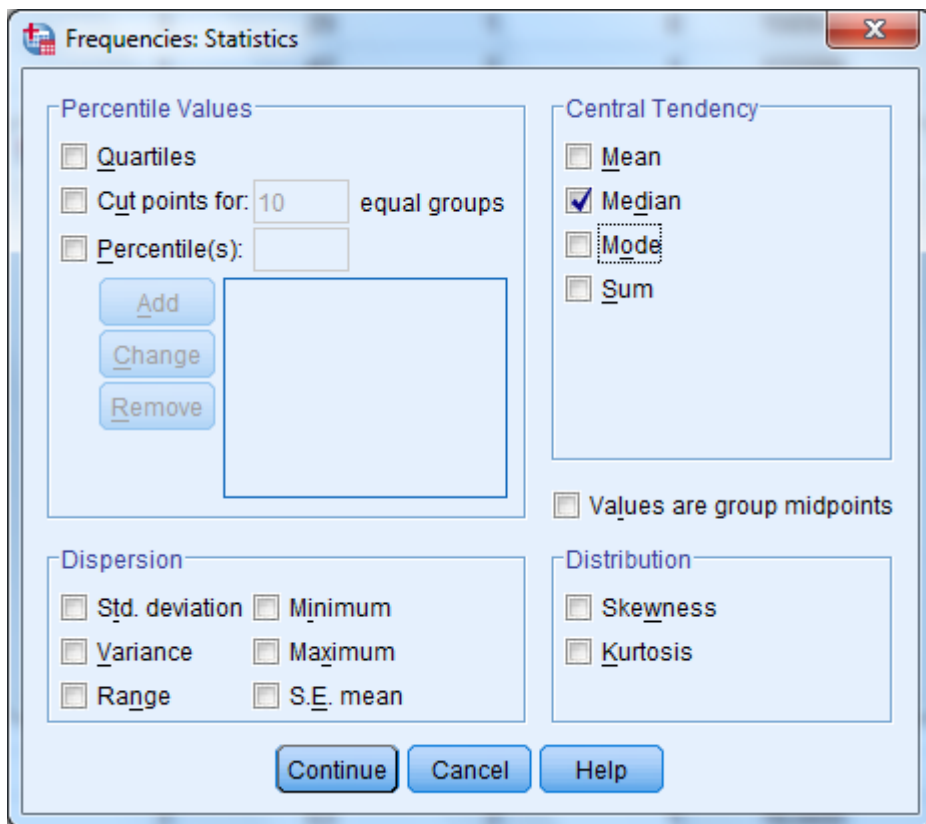
Páratlan számú adat esetén azonnal adódik a medián, míg páros számú adat esetén a két középső értéknek kell a számtani közepét vennünk.

A medián meghatározása akkor is változatlan, ha a rendezett adatok középső eleméből több létezik. Például: 2, 2, 3, 5, 6, 6, 6, 9, 10. A medián tehát: 6.

A táblázatkezelő szoftverek használatával eltekinthetünk a sorba rendezéstől, a **MEDIÁN** függvény alkalmazásával megkapjuk az adatokhoz tartozó medián értékét:

- ✱ Határozza meg a kérdőívet kitöltők életkorának mediánját!

SPSS parancsa a MEDIAN, melyet az átlaghoz hasonló módon az ANALYZE/DESCRIPTIVE STATISTIC/FREQUENCIES parancsnál adhatunk meg. Tegyük az életkor oszlopot az elemzendő Variables(s) ablakba, és az OPTIONS nyomógombot használva megjelenő panelen kapcsoljuk be a MEDIAN választógombot.



17. ábra: Medián megadása

Leggyakrabban a három középértéket együttesen határozzuk meg, melynek eredménye a mintafeladatunkban, az életkorokat alapul véve:

➔ Frequencies

[DataSet1] C:\Tunde_SPSS\kerdoiv.sav

Statistics

Eletkor

N	Valid	30
	Missing	0
Mean		38,10
Median		39,50
Mode		29

18. ábra: Medián megadása

Feladat középérték-vizsgálatokra

- ✿ Tekintsük meg az alábbi feladatot, melynek nem a megoldása, hanem értelmezése érdekes! Láthatjuk, hogy egy osztály tanulójának egy tantárgyból elért dolgozati eredményeit tartalmazza a táblázat. A félév során 4 évfolyamdolgozatot írtak a tanulók.

*meresek.sav [DataSet3] - IBM SPSS Statistics Data Editor						
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help						
17:						
	tanuló	dolgozat_1	dolgozat_2	dolgozat_3	dolgozat_4	var
1	Bartók Renáta	3	5	5	5	
2	Budai Anna	4	2	2	5	
3	Buckai László	5	3	3	4	
4	D. Tóth Katalin	2	4	4	4	
5	Görbe Iván	1	5	4	3	
6	Kiss Kálmán	4	4	4	5	
7	Kovács Péter	5	5	5	5	
8	Papp József	1	3	2	4	
9	Piroska Attila	2	4	3	4	
10	Tóth Alíz	3	5	3	4	
11	Varga Letti	4	2	4	4	
12	Vértes Ottó	3	3	5	4	
13						

19. ábra: Feladat középérték-vizsgálatra

A mérés-értékelés területén a legfontosabb, hogy a kapott adatok adatokat hogyan értelmezzük.

Határozzuk meg a dolgozatokra vonatkozó átlag, medián, módusz értéket. A kapott értékek értelmezése: megkaptuk az első évfolyamdolgozat átlagát, leggyakoribb jegyét, és mediánját, és a táblázat ezt a második, harmadik és negyedik évfolyamdolgozatra is megmutatja.

Értelmezésekor megállapítható, melyik volt a leggyengébb évfolyamdolgozat, mi volt a maximum értékük, és ezáltal a tanár tevékenységét tudjuk jellemezni: a leggyengébb évfolyamátlagot az első dolgozatnál produkálta a csoport, míg a legjobb dolgozat az utolsó dolgozat eredménye, tehát a pedagógus fejlesztő munkája jól sikerült, hisz fejlődést ért el a félév során.

3.2.3 Gyakoriság

Mint láttuk, a középértékek a teljes mintát három számértékkel jellemzik. Még alacsony elemszámú minta esetén sem vonhatunk le következtetéseket csupán a középérték-mutatók alapján, azonban a minta összes elemének figyelembevételét sem tudjuk megvalósítani.

Ezt a problémát oldják meg a gyakoriságvizsgálatok. A gyakorisági vizsgálat az adatokat kategóriákba sorolja, és meghatározza az egyes kategóriákba tartozó elemeket. Az adatok kategóriákba sorolása, csoportosítása sok esetben az általános összefüggések felismerését jobban szolgálja.

Abszolút gyakoriság

Az abszolút gyakoriság megmutatja, hogy a mintha hány eleme tartozik az adott kategóriákba.

A kategória létrehozásának az szabályai vannak!

3. A kategóriák számának meghatározása.

Első lépésként döntenünk kell a kategóriák számáról, amit az adatok számának függvényében tehetünk meg. Ha túl sok kategóriát hozunk létre, nem tölti be funkcióját a csoportosítás (továbbra is nehezen áttekinthetők lesznek az adatok), míg ha alacsony a kategóriák száma, túl nagy intervallumok jönnek létre, és ez pontatlanná teszi a munkánkat. A csoportok számának célszerű 10 és 20 közötti páratlan számot választani, de alacsony számú (50 körüli elmeszámú) minta esetén kevesebb (7-9 kategória) is lehet.

4. A lépésköz meghatározása.

Ha döntöttünk a kategóriák számáról, hozzuk is létre azokat! Ehhez ismernünk kell az adataink értéktartományát, amit megkapunk, ha meg-


határozzuk a legkisebb és legnagyobb elemét a mintánknak, majd vesszük a kettő különbségének abszolút értékét.

A csoportintervallumok nagyságának meghatározásakor rendszerint 1, 2, 3, 5, 10 vagy ennek többszöröseit használjuk, mint intervallumhosszokat. A konkrét döntést a kategóriaszám befolyásolja.


5. **Diszjunkttság:** A csoportok meghatározásánál ügyelni kell arra, hogy a minta minden eleme pontosan egy kategóriába legyen besorolható, ezért a csoportok nem fedhetik át egymást.
6. A csoporthatárok elkészítése után meg kell határozni, hány adat tartozik az egyes kategóriákba.

Az egyes kategóriákba tartozó értékeket az adott **csoport gyakoriságának** nevezzük, a létrejövő értékeket együttesen pedig **a minta abszolút gyakorisági eloszlásának**.

Az abszolút gyakorisági adatokkal az a probléma, hogy önmagukban nem értelmezhetőek. Csak akkor van értelme annak, hogy pl. 12 fő van a kitöltők között, aki 41 és 50 év közötti, ha hozzátesszük, hogy összesen 30 kitöltő volt, hiszen ez esetben ez a legjellemzőbb korosztály, míg ha 100 kitöltőnk lenne, akkor elenyészőnek kell értékelni a 40 éveseket. Ezért célszerű a relatív gyakoriságot is feltüntetni az adatok mellett!

 **Relatív gyakoriságnak nevezzük az abszolút gyakorisági értékek és az elemszám hányadosát, azaz az egyes kategóriába tartozók összes kitöltőhöz viszonyított százalékos arányát.**

Kumulatív gyakoriság

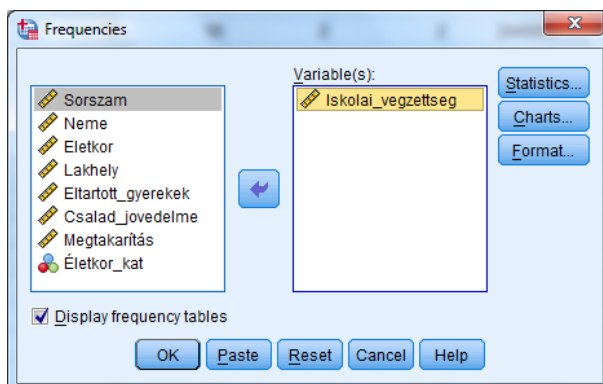
 **A kumulatív gyakoriság megmutatja, a minta hány eleme található a kategória felső határa alatt.**

Százalékos kumulált gyakoriság

A halmozott %-os gyakoriságot, vagy más néven százalékos kumulált gyakoriságot a relatív gyakoriság értékeiből képezzük, melynek előnye, hogy százalékosan mutatja, hogy a kategória felső határa alatt a minta hány százaléka található.

A SPSS szoftver a gyakoriság vizsgálatot nagyon egyszerűen kezeli.

Használjuk a **ANALYZE/DESCRIPTIVE STATISTIC/FREQUENCIES** parancsot. A Variable(s) ablakba tegyük be az elemzendő mezőt (pl. iskolai végzettség), majd kapcsoljuk be a Display frequency tables választó kapcsolót.



20. ábra: Gyakoriság vizsgálat

A kapott táblázat egy lépésben visszaadja az abszolút gyakoriság értékét (Frequency), a relatív gyakoriság értékét (Percent), mely az teljes minta elemszámához viszonyít, illetve találunk egy Valid Percent értéket, mely szintén a relatív gyakoriságot mutatja, de csak az érvényes (Missing adatok elhagyásán kívüli) adatokhoz viszonyít. Az utolsó Cumulative Percent a kumulált gyakoriság értékét adja meg, mely értékek százalékos formában értendők.

➔ Frequencies

[DataSet1] C:\Tunde_SPSS\kerdoiv.sav

Statistics

Iskolai_vegzettseg

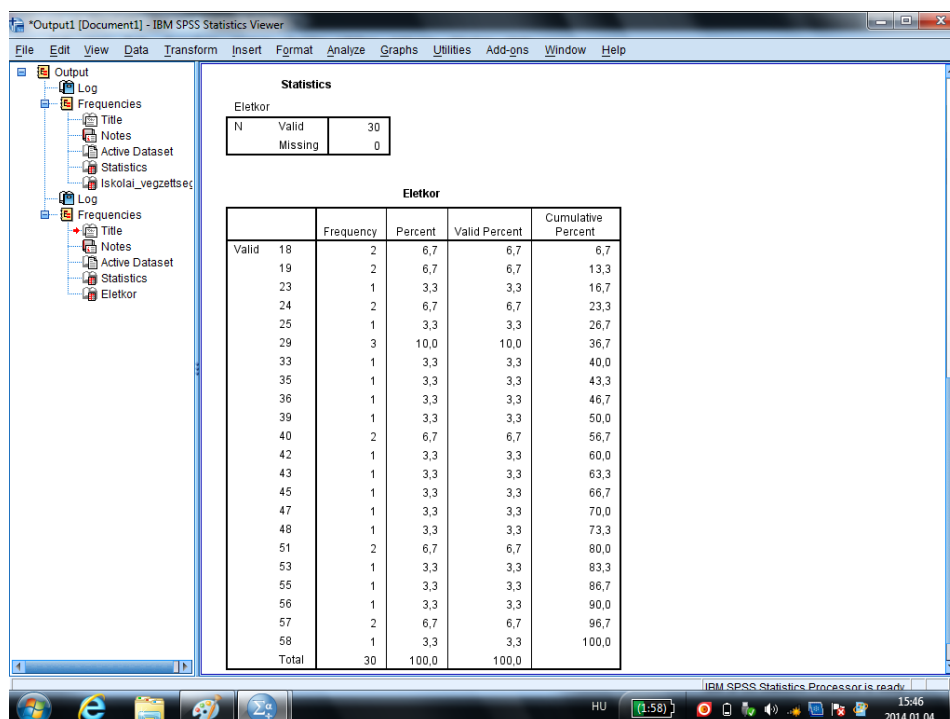
N	Valid	30
	Missing	0

Iskolai_vegzettseg

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid 8 általánosnál kevesebb	1	3,3	3,3	3,3
8 általános	4	13,3	13,3	16,7
Szakmunkás vagy szakiskolai	8	26,7	26,7	43,3
Érettségi	5	16,7	16,7	60,0
Technikum	3	10,0	10,0	70,0
Felsőfokú végzettség	9	30,0	30,0	100,0
Total	30	100,0	100,0	

21. ábra: Gyakoriság elemzés eredménye

Az előbbi eredménytáblát abban az esetben kapjuk, ha a mezőben előforduló kódokhoz készítettünk szöveges feloldást. Ha tisztán számértékeket tartalmaz a mező, akkor a gyakorisági tábla nem kategóriákat hoz létre, hanem az egyes elemek előfordulását, gyakoriságát adja meg.



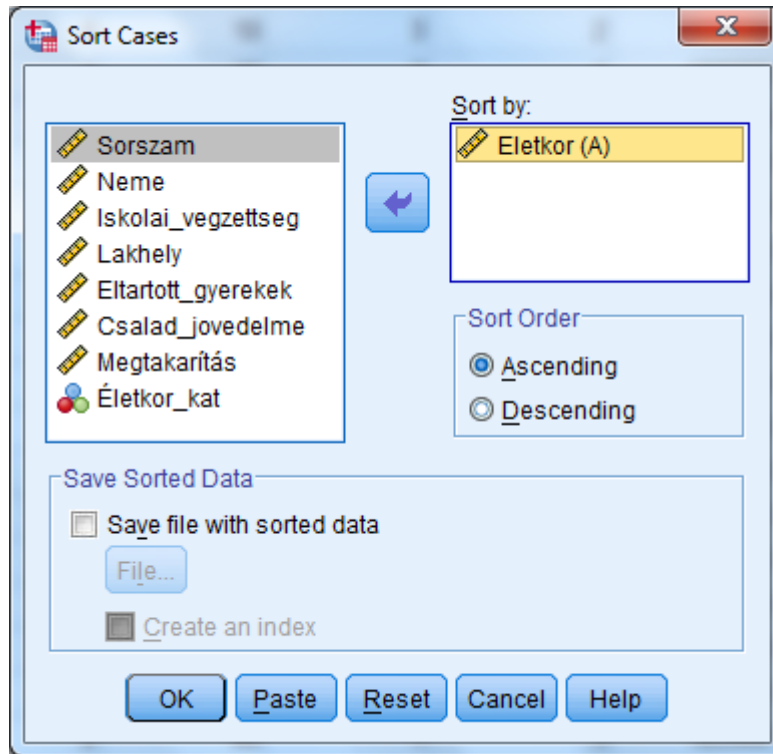
22. ábra: Gyakoriság elemzés eredménye

Ha szeretnénk a számadatokat tartalmazó adatsorhoz is kategóriákat létrehozni, arról a fenti szabályokat betartva nekünk kell gondoskodni, egy új mezőt létrehozva. Legyen ez az életkor_kategória mező, melynek value értékeihez vegyük fel az alábbi kategóriákat:

4. Gyakorisági kategóriák

Value	Label
1	20 év alatti
2	20-29 éves
3	30-39 éves
4	40-49 éves
5	50 éves vagy idősebb

Ha létrehoztuk a mezőt, akkor fel kell tölteni értékekkel, melyet kézzel kell megtenni. Előbb rendezzük az életkorok szerint a teljes táblázatot. A rendezésre a DATA/SORT CASES parancs szolgál. Adjuk meg a SORT BY ablakba a mezőnevet a rendezési elvnek, és a rendezés iránya legyen növekvő (Ascending).



23. ábra: Rendezés

MÁSİK MEGOLDÁS

Az eredményként elő álló táblázatban gyorsan ki lehet tölteni az Életkor_kategória mezőt 1 és 5 közötti értékekkel, a kategóriák alapján.

(Érdemes az adatbázist visszarendezni, melyhez jó szolgálatot tesz a Sorszám mező → végezzük el a rendezést sorszám szerint növekvő sorrendbe.)

Ezután végezzük el a gyakoriság vizsgálatot, de ne az Életkor mezőre, hanem az Életkor_kategóriája mezőre, akkor az alábbi eredményt látjuk.

➔ Frequencies

[DataSet1] C:\Tunde_SPSS\kerdoiv.sav

Statistics

Életkor_kat

N	Valid	30
	Missing	0

Életkor_kat

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	20 év alatti	4	13,3	13,3	13,3
	20-29 éves	3	10,0	10,0	23,3
	30-39 éves	9	30,0	30,0	53,3
	40-49 éves	8	26,7	26,7	80,0
	50 éves vagy idősebb	6	20,0	20,0	100,0
	Total	30	100,0	100,0	

24. ábra: Gyakoriság elemzés eredménye

Vizsgáljuk meg, milyen adatokat kaptunk eredményül!

Az abszolút gyakoriság értékeit vizsgálva láthatjuk, hogy a legmagasabb létszám az 41-50 és a 31-50-es életkor-kategóriában található. Ez azt jelenti, hogy a vizsgálatban résztvevők 56,76%-a (ezt a relatív gyakoriság mutatja meg nekünk) középkorú.

A kumulatív gyakoriság abban az esetben használható, ha például a legfiatalabb 15 ember véleményére vagyunk kíváncsiak.

A megoldáshoz nézzük meg a kumulatív gyakoriság alakulását! Látható, hogy 16 fő 40 év alatti, így az ő válaszíveiket kell részletesebben tanulmányozni.

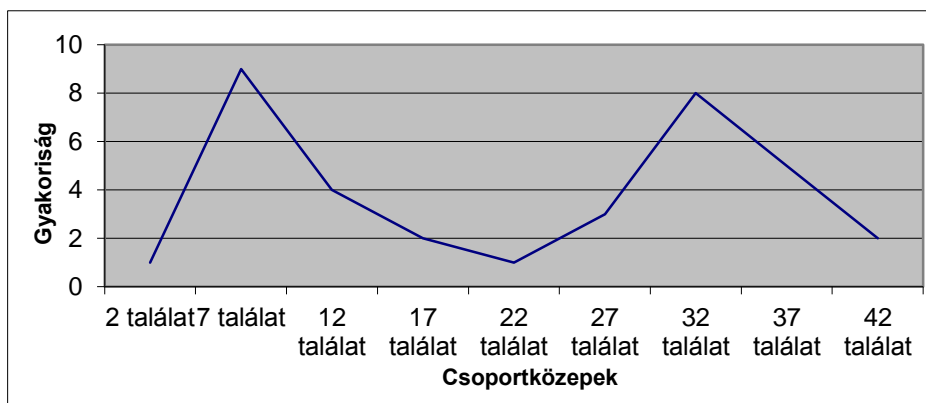
3.2.4 Gyakorisági poligon és a középérték-mutatók

Gyakorisági poligon és a középérték-mutatók

A középértékek és a gyakorisági eloszlás közötti viszony leolvasható a gyakorisági poligonról.

Ha ábrázoljuk a gyakorisági eloszlást egy poligonnal, akkor a poligon csúcának a helye a módusz.

Ha a gyakorisági poligonnak két, jól meghatározható „csúcossága” van, akkor a mintában két elemnek nagyobb a gyakorisága, mint a többinek. Ekkor **bimodális eloszlásról** beszélünk.



25. ábra: Bimodális eloszlás

A bimodális eloszlás oka többféle lehet. Ha az ábra azt mutatja, hogy egy megyei könyvtárban 50 önálló keresésből hány esetben voltak sikeresek a tanulók, az eltérés oka származhat abból, hogy a felmérésben résztvevők egy része gyakran látogatja a könyvtárat, a másik része pedig ritkán jut el egy nagyobb könyvtárba.



Az előbbi feladatban a két osztály matematikai jegyeinek móduszát megtekintve már látható a különbség. Ha az egyik osztálynál a leggyakoribb jegy a jeles, az osztályátlag pedig közepes, akkor ez tükrözi, hogy nem egyenletes az osztály tanulóiinak teljesítménye.

3.2.5 A középértékek egymáshoz viszonyított kapcsolata

A három középérték megvizsgálásából következtethetünk a minta gyakorisági ábrázolásának poligonalakjára!

Ha teljesen szimmetrikus a gyakorisági eloszlást szemléltető poligon, akkor a három középérték egybeesik. (És fordítva: ha már meghatároztuk a középértékeket, és a három érték azonos, akkor várhatóan a gyakorisági értékek kiszámítása, majd ábrázolása után előálló gyakorisági poligon szimmetrikus lesz.)

Ekkor ugyanis a mintát nem jellemzik a szélsőséges értékek kiugróan magas számában.

Egy „**balra ferdült**” gyakorisági poligonnal rendelkező minta esetében gyakoriak a szélsőségesen magas értékek, ezért a módusz lesz a legmagasabb a három középtérték közül, majd a medián következik, és a számtani középtértéknek lesz a legalacsonyabb értéke az ilyen jellegű eloszlás esetén.

Egy „**jobbra ferdült**” gyakorisági eloszlással rendelkező mintában az alacsony értékek előfordulása a legnagyobb, ezért a középtértékek a következő sorrendben követik egymást:

Módusz < Medián < Számtani közép



Mintánk számtani középtértéke legyen 38, módusza 51, mediánja pedig 40. Ez esetben a minta balra ferdült eloszlással rendelkezik, a kiszámított értékek alapján a következő sorrend állt elő: Módusz < Medián < Számtani közép.

3.2.6 Szóródási mérőszámok

A középtérték-mutatók önmagukban nem elegendők a minta jellemzésére. Amikor a minta elemeinek az átlagtól való eltéréseit elemezzük, akkor a szóródási mutatókat határozzuk meg.

A gyakoriság- és középtérték-vizsgálatok elkerülhetetlen lépési a minta elemzésének. Azonban vannak esetek, amikor a középtértéket jellemző mérőszámok egybeesnek.

✿ Nézzük meg kérdőívünk középtértégeit!

Az életkorok elemzése után az alábbi eredményt kaptuk!

➔ Frequencies

[DataSet1] C:\Tunde_SPSS\kerdoiv.sav

Statistics

Eletkor

N	Valid	30
	Missing	0
Mean		38,10
Median		39,50
Mode		29

26. ábra: Középtértékek

Láthatjuk, hogy az adatok különbözőek, de a középérték-vizsgálatok eredményének elemzése után nem sikerült a mintát részletesen megismerni.

A középérték-vizsgálatok önmagukban nem elegendőek a minta jellemzésére, meg kell határozni az adatoknak a szóródási mutatóit is.

Szóródási terjedelem

- ☞ **A szóródási terjedelem megegyezik a minta értéktartományával, tehát a minta legnagyobb és a legkisebb elemének a különbsége.**

Átlagos eltérés

Ha vesszük minden elemnek az átlagtól való eltérését, és összeadjuk, akkor az eredmény nulla. Ezért önmagában az átlagtól való eltérések összege nem lesz mérőszám. Azonban ha az ezen eltérések abszolút értékét adjuk össze, már használható értéket kapunk! Küszöböljük ki az elemszámból adódó eltéréseket, azaz osszuk el az összeget a minta elemszámával, és megkaptuk az első szóródási mutatónkat: az átlagos eltérést.

- ☞ **Átlagos eltérésnek nevezzük a minta elemeinek az átlagtól való átlagos távolságát.**

$$AE = \frac{\sum_{i=1}^n |x - x_i|}{n}$$

27. ábra: Az átlagos eltérés képlete

Figyeljük meg a definícióban szereplő távolság szót! A távolság mindig pozitív szám, ezért használhatjuk az abszolút érték kifejezésére.

A következő mérőszámmal még mindig az átlagtól való eltérést elemezzük, de ne abszolút értékkel küszöböljük ki az átlagtól való eltérések összegének nulla értékét, hanem négyzetre emeléssel.

A négyzetre emelés jobban tükrözi a minta szóródását, hiszen a „kisebb eltérések” is négyzetesen jelennek meg.

Négyzetes összeg

- ☞ **A minta elemeinek az átlagtól való eltéréseinek négyzete összegezve a minta minden eleme esetén a négyzetes összeg.**

$$N\ddot{O} = \sum (x_i - \bar{x})^2$$

28. ábra: Négyzetes összeg

A négyzetes összeg nem küszöböli ki a minta elemszámából adódó eltéréseket.

Variancia

- ☞ A variancia a négyzetes összeg osztva a minta szabadságfokával.
- ☞ Szabadságfoknak nevezzük a minta független elemeinek számát.

$$s^2 = \frac{\sum_{i=1}^n (\bar{x} - x_i)^2}{n - 1}$$

29. ábra: Variancia

Egyváltozós minta esetén a minta szabadságfoka $n-1$.

Ha a matematikában tekintjük meg a variancia képletét, azt láthatjuk, hogy a négyzetes összeget nem a szabadságfokkal, hanem a minta elemszámával osztják.

30-nál kisebb elemű minta esetén a szabadságfokkal történő osztás jobb közelítést ad a variancia értékére, 30 fölötti elemszám esetén ez a különbség elhanyagolható.

Statisztikában a szabadságfokkal történő osztást használjuk.

A variancia jól tükrözi az átlag körüli ingadozást, ezért több olyan statisztikai mutatóval fogunk találkozni, ami használja a variancia értékét (főleg azok, melyek érzékenyek a nagyon heterogén adatösszetételű csoportokra).

A varianciát **szórásnégyzetnek is nevezik**, illetve ez a jelölésében is megmutatkozik.

Szórás

- ☞ A szórás a variancia pozitív előjelű négyzetgyöke.

$$s = \sqrt{\frac{\sum_{i=1}^n (\bar{x} - x_i)^2}{n-1}}$$

30. ábra: A szórás képlete


A szórás mérőszáma az átlagértékkel együtt megadva számos információt szolgáltat a mintáról. Ennek oka az alábbi tételekben rejlik:

- A mintától 1 szórásnyi terjedelemben tartozik az adatok több mint 2/3-a.
- A mintától 2 szórásnyi terjedelemben tartozik az adatok több mint 95%-a.
- A mintától 3 szórásnyi terjedelemben tartozik az adatok több mint 99 %-a.

Ebből következik, hogy az átlag és szórás értékének ismeretében jól össze lehet hasonlítani az eltérő összetételű mintákat.

Ha kicsi a szórás értéke, akkor a csoport tagjai az átlag körül mozognak, míg magas szórás értéke esetén sokkal nagyobb a változatosság az adatokban.

Relatív szórás

 **A relatív szórás a szórás átlaghoz viszonyított mérőszáma, azaz a szórás és az átlag hányadosának eredménye.**

A relatív szórás értékének kiszámításával megoldhatjuk azt a problémát, hogy a szórás értéke csak azonos értéktartományú minták összehasonlítására alkalmas. A relatív szórás (vagy más néven **variációs együttható**) a szórás és annak számtani középértékéből képzett százalékos viszonyszám.

Kvartilisek, percentilisek


A medián számításakor megadtuk, melyik elem a minta közepe. Nem csak a középső elem meghatározására van lehetőség, ha sorba rendezzük a minta elemeit, meghatározhatjuk a minta negyedelési pontjait, azaz a kvartilisek értékeit.

 **A kvartilis a minta negyedelő pontja.**

1. kvartilis, az a szám, amelytől a minta elemeinek egynegyede kisebb, háromnegyede pedig nagyobb sorba rendezés esetén.

Hasonló elven adhatjuk meg a minta 2. kvartilisét (mely megegyezik a minta mediánjával), és a minta 3. kvartilisét is, mely az az érték, amitől a minta eleminek háromnegyede kisebb, egynegyede pedig nagyobb sorba rendezés esetén.

A 0. kvartilist nem más, mint a minimum, valamint a 4. kvartilis megegyezik a maximum értékével.

 **Az n-edik percentilis az az érték, melytől a minta n%-a kisebb egyenlő, n-n%-a pedig nagyobb egyenlő.**

A definícióból adódóan a mediánt 50. percentilisnek (vagy 50%-os percentilisnek) is szokták nevezni, a kvartilisek pedig a 25., 50. és 75. percentilisek.

Leggyakrabban 10., 20....90. percentiliseket szoktunk meghatározni, melyek a minta tizedelési pontjai.



Példa: A gyermekgyógyászatban növekedési görbék értékeit veszik alapul a gyermekek súlyára és magasságára vonatkozóan. A percentilis kalkulátor segítségével megadják a gyermekre vonatkozó percentilis értékeket.

Például ha a gyermek magassága 80 percentilis, akkor az azt jelenti, hogy a hasonló korú gyermekek 80%-a alacsonyabb, és 20%-a magasabb a szóban forgó gyermektől.

A percentilis táblázat folyamatos nyomon követése képes felhívni a figyelmet betegségekre: „Mivel a gyermekek növekedési üteme általában azonos, ezért a percentilis görbéken többnyire tartják azt a percentilist, amelyikbe korábban tartoztak. A percentilis értékekben bekövetkező jelentős csökkenés növekedésleállásra hívhatja fel a figyelmet, ezért ilyen esetekben mindenképpen gyermekgyógyász felkeresése javasolt.”²

A fenti elemzéseket az SPSS szoftver ANALYZE/DESCRIPTIVE STATISTICS/FREQUENCIES parancsával végeztethetjük el, ahol a STATISTICS nyomógombra megjelenő panelen adjuk meg az alábbiakat:

Quariles (Kvartilisek)

Std.deviation (Szórás)

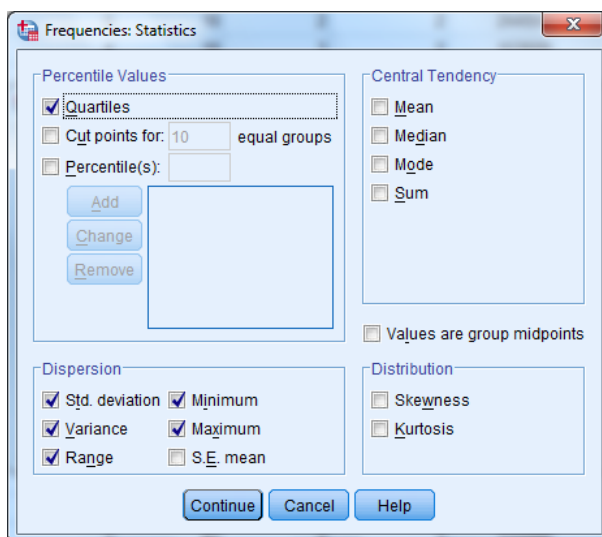
Variance (Variancia)

² Diagnózisok közérthetően. <online> <<http://www.medstart.hu/gyerme-percentilis-kalkulator.html>>

Range (Szóródási terjedelem)

Minimum

Maximum



31. ábra: Szóródási mérőszámok megadása

Az eredményként megjelenő táblázatban a kvartilisek helyett a 25,50, és 75 %-s perszentilisek kerülnek feltüntetésre (ami megegyezik a kvartilisekkel).

➔ Frequencies

[DataSet1] C:\Tunde_SPSS\kerdoiv.sav

Statistics		
Életkor_kat		
N	Valid	30
	Missing	0
Std. Deviation		1,291
Variance		1,666
Range		4
Minimum		1
Maximum		5
Percentiles	25	2,75
	50	3,00
	75	4,00

32. ábra: Szóródási mérőszámok eredménytáblája

Interkvartilis félterjedelem

- ☞ **Interkvartilis félterjedelem a harmadik és az első kvartilis különbsége.**

A minta nagyon érzékeny a kiugró értékekre. Például a pontversenyeken sem veszik figyelembe a legmagasabb és a legalacsonyabb pontot. Az interkvartilis félterjedelem kiküszöböli a minta alacsony és magas elemeit, még pedig pont minden irányban egyenyednyi adatot hagy el.

3.3 ÖSSZEFOGLALÁS, KÉRDÉSEK

3.3.1 Összefoglalás

A fejezetben a leíró statisztika alábbi mutatóit ismerhettük meg:

- A középérték-mutatók közül:
 - Számtani közép (átlag)
 - Medián, a középső elem
 - Módusz, a leggyakoribb elem
- Szóródási mérőszámok:
 - A szóródás terjedelme
 - Átlagos eltérés
 - Négyzetes összeg
 - Variancia
 - Szórás
 - Relatív szórás
 - Kvartilisek
 - Percentilisek
- Gyakorisági mutatók:
 - Abszolút gyakoriság
 - Relatív gyakoriság
 - Kumulatív gyakoriság
 - Halmazott százalékos kumulált gyakoriság

3.3.2 Önellenőrző kérdések

1. Mi az előnye és mi a hátránya a középérték-mutatóknak?
2. Miért van szükség a szóródási mérőszámok elemzésére?
3. Miért határozzuk meg a kvartiliseket?
4. Mondjon példát, hol használják a percentiliseket!
5. Sorolja fel a gyakorisági elemzések kategóriaképzésének lépéseit!
6. Mi a gyakorlati különbség az abszolút és a relatív szórás között?

4. MATEMATIKAI STATISZTIKAI LEHETŐSÉGEK AZ SPSS TÁBLÁZATKEZELŐKBEN

4.1 CÉLKITŰZÉSEK ÉS KOMPETENCIÁK

A fejezetben a matematikai statisztika mutatóit ismerhetjük meg. Akkor használunk matematikai statisztikai elemzési módszereket, ha nem az összes személy adatából szeretnénk következtetéseket levonni a teljes populációra vonatkozóan. A fejezetben a leggyakoribb korrelációs számítás elméletét és eseteit ismerhetjük meg, melyek két adatsor közti kapcsolat kimutatására szolgálnak. Hasonló módon összefüggést vizsgál a regresszióanalízis is, ahol két összefüggő adatsor hiányzó elemére adunk becslést. A fejezet foglalkozik a speciálisabb problémák megoldására szolgáló parciális korrelációs számítás elméletével, valamint a rangsorolt adatok összefüggés-vizsgálatára szolgáló Spearmann-féle rangkorreláció-számítással. A fejezet kitér a változók számának minőségi módon történő csökkentésére szolgáló faktoranalízisre és a rendezetlen adatok átláthatóbbá tételét szolgáló klaszteranalízis mutatóinak ismertetésére.

4.2 TANANYAG

Matematikai statisztika

Korreláció

Korrelációanalízis

Regressziószámítás

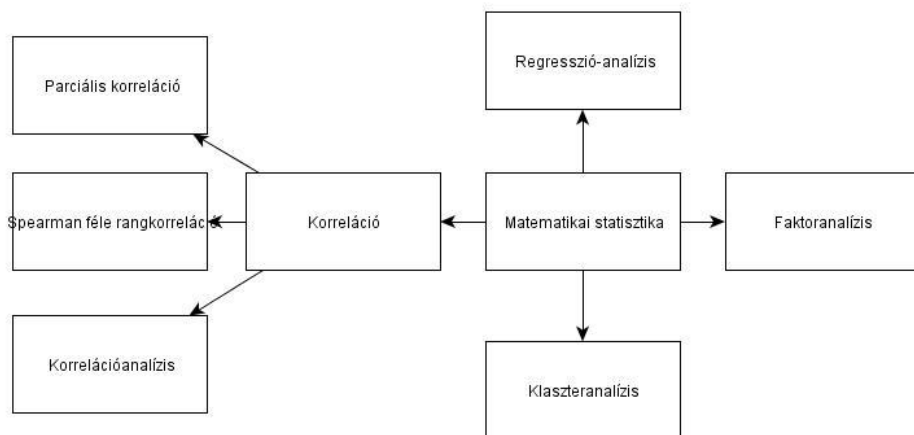
Regresszióanalízis

Faktoranalízis

Parciális korreláció

Spearman-féle rangkorreláció

Klaszteranalízis



33. ábra: Fogalomtérkép

4.2.1 Matematikai statisztika

Matematikai statisztikára azokban az esetekben van szükségünk, amikor a minta nem egyenlő a populációval. A probléma ezen esetekben, hogy rendelkezésünkre állnak 50, 500, 5000 stb. ember adatai, válaszai, melyet leíró statisztikai elemzésekkel értékelünk. Kialakul egy kép, hogy mi jellemzi a mintát. De jó lenne tudni, hogy a teljes populációra is igazak-e ezek a megállapítások, vagy sem!

☞ **A matematikai statisztika választ ad arra a kérdésre, hogy a reprezentatív mintából vonható-e le következtetés az alapkasságra.**

A fejezetben azokat a matematikai statisztikai módszereket ismerhetjük meg, melyek az adatsorok közti összefüggések feltárására törekednek.

4.2.2 Korreláció

Korrelációszámítást végzünk minden olyan esetben, amikor két vagy több adathalmaz közötti kapcsolat szorosságát (meglétét) mérjük.

☞ **A korrelációs együttható két mennyiségi adatsor közti kapcsolat erősségét és irányát megadó együttható.**

A korrelációs együttható a két adatsor szórásán alapszik:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \cdot \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}}}$$

34. ábra: Lineáris korreláció képlete

A korreláció értéke -1 és 1 közé eső szám. Minél közelebb van az érték az 1 egészhez, annál szorosabb kapcsolat van a két adatsor között. A nulla körüli korrelációs érték pedig korrelálatlanságot fejez ki, azaz a két adatsor között nincs összefüggés.

A korrelációs együttható előjele a kapcsolat irányát mutatja meg. Negatív előjelű korreláció esetén az egyik adatsor magas értékéhez a másik adatsor magas értéke tartozik, és fordítva. Pozitív előjel esetén pedig az egyik adatsor magas értékéhez a másik adatsor magas értéke tartozik.

A korrelációs együttható értelmezéséhez használhatunk korrelációs táblát, ahol a minta elemszámának figyelembe vételével kapjuk meg, hogy milyen korrelációs értéktől beszélhetünk kimutatható kapcsolatról. Precízen korrelációs táblázat alkalmazásával határozhatjuk meg az értéket, de ennek hiányában használhatjuk a következő általánosan elfogadott intervallumot, amelybe, ha belesik a korrelációs együttható, akkor kimutathatónak tekinthetjük a kapcsolatot:

5. A korrelációs együttható értelmezése

A korrelációs együttható értéke	Változók közötti kapcsolat
0,9 – 1	rendkívül szoros
0,75 – 0,9	szoros
0,5 – 0,75	érzékelhető
0,25 – 0,5	laza
0,0 – 0,25	nincs kapcsolat

A korrelációs együttható megmutatja, hogy van-e kapcsolat a két adatsor között, azonban nem ad választ arra, hogy melyik adatsor hat a másikra, miért van köztük összefüggés.

Például ha kimutatjuk, hogy egy osztály tanulóinak jegyeit megvizsgálva a matematika és fizika jegyek között pozitív korrelációs viszony van, akkor elmondható, hogy aki jó fizikából, az jó matematikából is, és fordítva, a gyenge fizikaeredményű tanuló matematikából sem brillírozik. Azt azonban nem lehet tudni, hogy ennek oka az-e, hogy pl. aki mindig szerette a matematikát, azt is pozitívan fogadta-e, amikor bevezették a fizikaoktatást, vagy amikor elkezdtek fizikát tanulni, annak pozitív hozadéka volt-e a matematika iránt megnövekvő érdeklődés.



Vizsgáljuk meg Magyarország infokommunikációs ellátottságát.

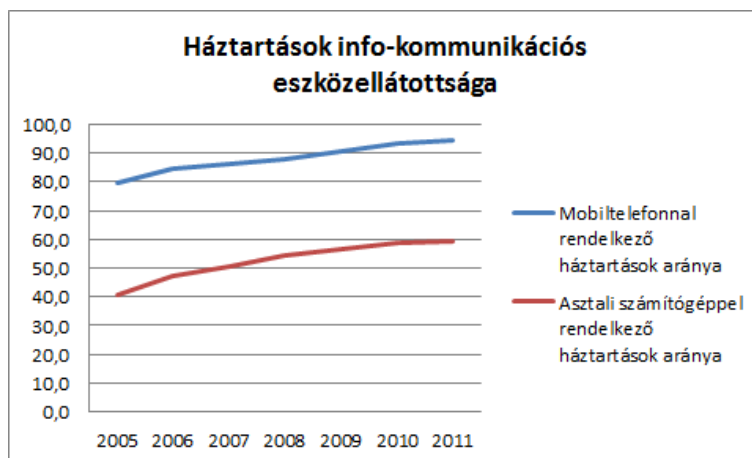
4.7.14. Háztartások infokommunikációs eszközellátottsága (2005–) [%] ³							
Mutató megnevezése	2005	2006	2007	2008	2009	2010	2011
Mobiltelefonnal rendelkező háztartások aránya	79,9	84,4	86,4	88,0	90,4	93,2	94,7
Asztali számítógéppel rendelkező háztartások aránya	40,7	47,1	50,6	54,6	56,8	58,6	59,5
Laptoppal rendelkező háztartások aránya	6,3	9,3	11,4	15,7	21,0	26	31
Palmtoppal rendelkező háztartások aránya	1,6	1,8	1,8	2,8	3,6	3,9	4,7
Internetkapcsolattal rendelkező háztartások aránya	22,1	32,3	38,4	48,4	55,1	60,5	65,2
Szélessávú internetkapcsolattal rendelkező háztartások aránya	10,9	22,0	33,0	42,3	50,9	52,2	60,8

SPSS-ben előállítva:

	Év	Mobiltelefon	Asztali_szg	Laptop	Palmtop	Internet	Szélessávú_internet	var	var	var	var	var
1	2005	80	41	6	2	22	11					
2	2006	84	47	9	2	32	22					
3	2007	86	51	11	2	38	33					
4	2008	88	55	16	3	48	42					
5	2009	90	57	21	4	55	51					
6	2010	93	59	26	4	61	52					
7	2011	95	60	31	5	65	61					
8												
9												
10												
11												
12												
13												
14												
15												
16												
17												
18												
19												
20												
21												
22												

35. ábra: Háztartások IKT ellátottsága

- ✿ Vessük össze a mobiltelefonnal rendelkező háztartások számát az asztali számítógéppel ellátott háztartásokéval! Ha diagramon ábrázoljuk az adatsorokat, akkor azt láthatjuk, hogy „együtt haladnak”.

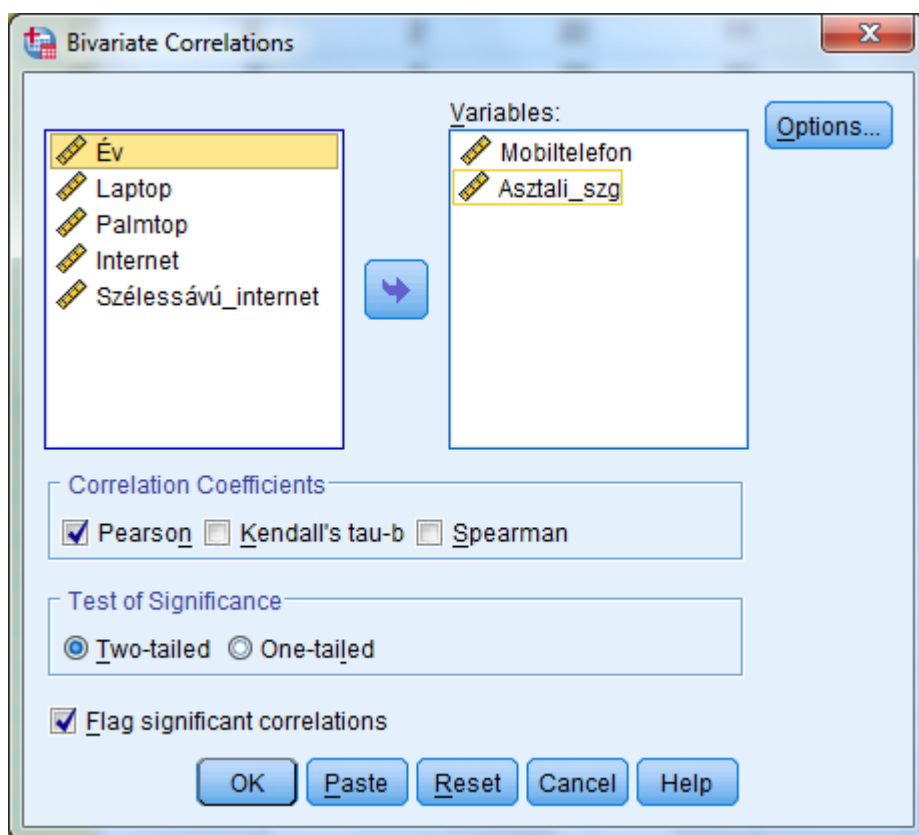


36. ábra: Háztartások infokommunikációs eszközellátottsága

Azt láthatjuk, hogy a két adatsor együtt halad, azaz magas mobiltelefonos ellátottsághoz magas asztali számítógép-ellátottság társul.

Határozzuk meg a korrelációs együtthatót!

Ehhez használjuk az ANALYZE/ CORRELATE /BIVARIATE CORRELATIONS parancsot.



37. ábra: Korrelációszámítás

Az eredményként előálló táblában (ahol a nullák nem kerülnek feltüntetésre) látható, hogy a két adatsor közti korrelációs együttható alátámasztja a diagramot: 0,980, pozitív korrelációs viszony. Érdeemes figyelembe venni az éveket is: határozzuk meg a mobilhasználat és az évek közti korrelációs viszonyt! A 0,990-es pozitív korrelációjával támasztjuk alá azt, hogy kimutathatóan évről évre magasabb a mobiltelefon-ellátottság. Ezt mindenki sejtette, de pontosan ez a lényeg: az, hogy érezzük/sejtjük, hogy évről évre nő a mobiltelefonok száma, az egy hipotézis. Adatsorok megvizsgálásával, korreláció kimutatásával bizonyítani is tudjuk ezt. A különbség a ránézésre megállapíthatóságtól az, hogy

lehetnek stagnáló évek, kisebb növekedés, nagyobb ugrások stb., míg a korrelációs együttható meghatározásával bizonyíthatjuk hipotézisünket.

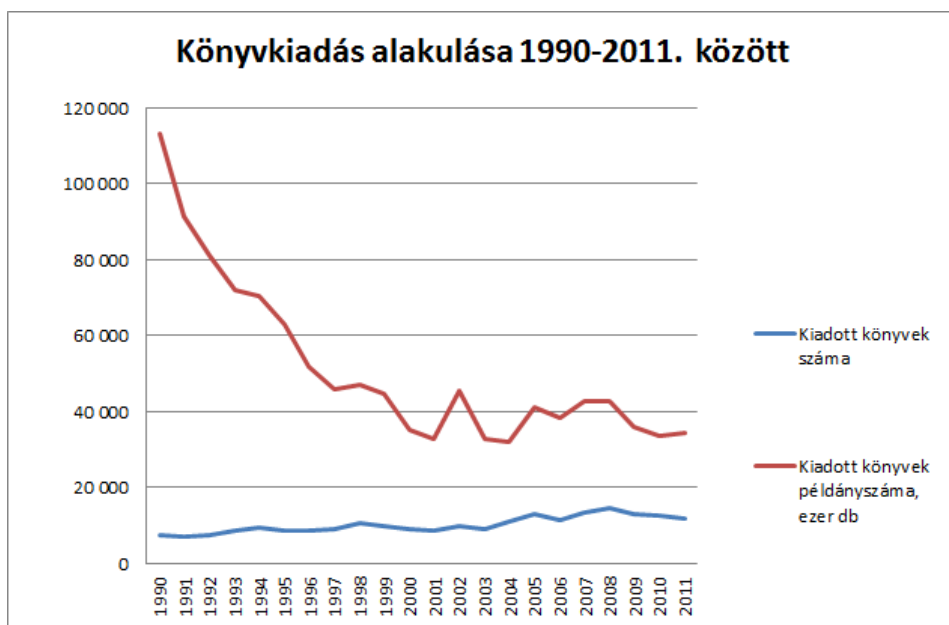


Vizsgáljuk meg, hogy Magyarországon hogyan alakultak a könyvkiadási adatok az elmúlt 20 évben!

6. Könyvkiadási adatok 1990-2011.

Év	Kiadott könyvek száma	Kiadott könyvek példányszáma, ezer db
1990	7 464	113 100
1991	7 210	91 400
1992	7 629	81 000
1993	8 458	72 100
1994	9 383	70 300
1995	8 749	63 000
1996	8 835	51 900
1997	8 911	45 700
1998	10 626	47 000
1999	9 731	44 700
2000	8 986	35 200
2001	8 837	32 600
2002	9 990	45 500
2003	9 204	32 600
2004	11 211	32 000
2005	12 898	41 000
2006	11 377	38 300
2007	13 239	42 600
2008	14 447	42 500
2009	12 841	36 000
2010	12 480	33 600
2011	11 821	34 200

Látható, hogy a megjelenő könyvek száma emelkedik, bár vannak évek, amikor visszaesés van, a kiadott példányszám esetén pedig jóval kevesebb a megjelenő példányszám, mint 1990-ben, de az adatok évről évre vizsgálva már nem ennyire egyértelműek. Nézzük meg, van-e kapcsolat a két adatsor között!



38. ábra: A könyvkiadási adatok alakulása 1990-2011 között

A kapcsolat nem állapítható meg egyértelműen! Ezért nézzük a korrelációs együtthatót, ami $-0,6336$ -os értékével negatív korrelációs viszonyt tükröz. Az elemzés kimutatja, hogy az évről évre növekvő számú szerzői kedv, könyvmegjelenés (év-kiadott könyvek száma korrelációs viszonya: $0,8826$) csökkenő példányszámmal társul. Azaz az embereknek nincs pénze, ideje több könyvre (kötet megvásárlására), mint korábban, ezért a magasabb könyvkiadáshoz alacsonyabb példányszámban történő megjelentetés társul.

4.2.3 Korrelációanalízis

A korrelációs együttható két adatsor közti kapcsolat erősségét tudja kimutatni, azonban vannak komplex kutatások, amikor nemcsak két adatsor viszonyát kell feltárnunk, hanem több adatsor kapcsolatát kell elemezni. A korrelációanalízis is páronként képi a korrelációs értékeket, de alkalmazása során egy lépésben határozzuk meg mindegyik adatsor mindegyik adatsorral való korrelációs viszonyát. Az eredményt mátrix formájában kerül ábrázolásra, ahol a mátrix átlóján minden adatsor önmagával való szorosságát, azaz 1 egészes korrelációs együtthatót láthatunk.

- ✿ Vizsgáljuk meg, milyen hatással van egymásra a könyvtárak, illetve szolgáltató helyek számának alakulása, valamint az állományfejlesztésre

fordított összeg, az állománygyarapítás és a beiratkozott olvasói létszám!

7. A könyvtárak alakulása 1990-2000-ig

Év	Könyvtárak száma	Szolgáltató helyek száma	Állományfejlesztésre fordított összeg	Év folyamán leltárba vett állományok száma	Beiratkozott olvasói létszám
1990	4262	4262	170680	1449897	644590
1991	3987	3987	205425	1753216	588037
1992	4012	3870	322350	1441556	539476
1993	3888	3792	343274	1297120	527809
1994	3822	3723	466858	1309491	517337
1995	3765	3635	470235	1174892	491450
1996	2854	3566	635025	1076750	1359667
1997	2883	3518	768347	1148884	1343508
1998	2775	3315	1139856	1201687	1350225
1999	2586	3273	1368162	1141647	1364488
2000	2573	3132	1213279	1164266	1357449

Korrelációanalízis készítése nem különbözik az két adatsor közti korreláció meghatározásától az SPSS szoftver használata során, egyszerűen a két adatsor helyett a BIVARIETE CORRELATIONS parancsablakba adjuk meg az összes változót!

➔ Correlations

[DataSet4] C:\Tunde_SPSS\Könyvtarak_alakulasa.sav

Correlations							
		Év	Könyvtárak_száma	Szolgáltatóhelyek_száma	Állományfejlesztésre_fordított_összeg	Leltárba_vett_állomány	Beiratkozott_olvasó
Év	Pearson Correlation	1	-.952**	-.985**	.953**	-.785**	.828**
	Sig. (2-tailed)		.000	.000	.000	.004	.002
	N	11	11	11	11	11	11
Könyvtárak_száma	Pearson Correlation	-.952**	1	.921**	-.929**	.728*	-.951**
	Sig. (2-tailed)	.000		.000	.000	.011	.000
	N	11	11	11	11	11	11
Szolgáltatóhelyek_száma	Pearson Correlation	-.985**	.921**	1	-.933**	.730*	-.765**
	Sig. (2-tailed)	.000	.000		.000	.011	.006
	N	11	11	11	11	11	11
Állományfejlesztésre_fordított_összeg	Pearson Correlation	.953**	-.929**	-.933**	1	-.664*	.840**
	Sig. (2-tailed)	.000	.000	.000		.026	.001
	N	11	11	11	11	11	11
Leltárba_vett_állomány	Pearson Correlation	-.785**	.728*	.730*	-.664*	1	-.638*
	Sig. (2-tailed)	.004	.011	.011	.026		.035
	N	11	11	11	11	11	11
Beiratkozott_olvasó	Pearson Correlation	.828**	-.951**	-.765**	.840**	-.638*	1
	Sig. (2-tailed)	.002	.000	.006	.001	.035	
	N	11	11	11	11	11	11

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

39. ábra: Korrelációs mátrix

A mátrix átlóján 1-eseket találunk, aminek az oka, hogy önmagával minden korrelál.

Ha megvizsgáljuk a korrelációs együttható értékeit, azt tapasztaljuk, hogy a példafeladatnál minden változó összefüggésben áll az összes többivel, ugyanis a korrelációk abszolút értéke minden esetben 0,65-nél nagyobb (0,5 fölött korrelálnak az adatok egymással). Érdekes összefüggéseket kapunk, ha az évek számát is bevonjuk a korreláció mátrixunkba, hiszen jól leolvasható, hogy az évek múlásával egyre csökken a könyvtárak és a szolgáltató helyek száma, valamint kevesebb az állománygyarapítás mértéke is. Viszont az év pozitív korrelációban áll az állományfejlesztésre fordított összeggel (0,9525) és a beiratkozott olvasói létszámmal (0,8280).

Leolvasható, hogy a szolgáltató helyek, a könyvtárak (0,9206) és az állománygyarapítás száma áll egymással pozitív korrelációban (0,7280), és ennek a három tényezőnek a beiratkozott olvasói létszámmal és az állományfejlesztésre fordított összeggel negatív korrelációja mutatható ki.

Mi lehet az oka ennek? Hogyan értelmezhetjük a kapott értékeket?

Az évek száma és a fejlesztésre fordított összeg pozitív korrelációban van egymással (0,9525), azaz évről évre több pénzt ad az állam a könyvtáraknak. De a fejlesztésre fordított összeg és a leltárba vett kötetek száma negatív korrelációban áll egymással (-0,6636). Ennek oka lehet:

- Évről évre több pénzt kapnak a könyvtárak, de nem annyival többet, mint amit az infláció megkövetelne.
- Évről évre több pénz kapnak a könyvtárak, de a könyvek ára erőteljesebb mértékben emelkedik, drágul, mint amennyivel nő a könyvtárak költségvetése.

Az évek számának és a leltárba vet kötetek számának negatív korrelációját az előbb említett érveken kívül azzal is magyarázhatjuk, hogy:

- A könyvtárak szép számmal vásárolnak könyveket, de nem olyan mértékben, mint ahogy a régi könyveket selejtezik. (A selejtezés oka nemcsak a fizikai megrongálódottság/elhasználtság lehet, hanem a kötelező irodalmak átalakulása, pl. napjainkban kevesebb szocializmussal kapcsolatos kötetre van szükség, mint az 1980-as években).

Azért hasznos a korrelációs mátrix, mert sok esetben két adatsor értelmezése nem reális értékeket tükröz, azonban együtt szemlélve a befolyásoló tényezőket, értelmet nyernek a folyamatok:

- A beiratkozott olvasói létszám és a kötetek száma között $-0,6377$ -es korrelációs együtthatóval kimutatható negatív korrelációt tapasztalunk. Ebből azt a – téves – következtetést lehetne levonni, hogy
 - ha kevesebb könyv van a könyvtárakban, akkor magasabb a beiratkozott olvasói létszám;
 - magas kötetszámhoz alacsonyabb olvasói létszám tartozik, ezért ne vásároljunk könyveket.

Egyértelmű, hogy az állítások nem felelnek meg a tudományos megállapításokkal szemben támasztott követelményeknek, azaz:

- legyen logikus
- legyen mögötte empiria.

Ha megnézzük a tényeket befolyásoló egyéb adatokat, akkor rögtön kiküszöböljük az állítások helytelenségét!

- Láthatjuk, hogy évről évre több a beiratkozott olvasó ($0,8280$).
- Miközben évente több pénzt fordítanak állományfejlesztésre ($0,9525$),
- mégis kimutathatóan csökken a kötetek száma ($-0,7845$).

Konklúzió: Mivel az évek száma folyamatosan emelkedik, ezért a folyamatot úgy kell értelmezni, hogy ugyan az állam évről évre egyre több pénzt fordít állományfejlesztésre, de gyorsabb ütemben drágulnak a könyvek, mint ahogy nő a támogatás, vagy annyira magas a selejtezés száma a könyvtárakban, hogy új könyvekkel nem tudják pótolni a kieső példányszámot. Azonban az emberek sem tudják megvásárolni a dráguló könyveket, ezért évről évre többen fordulnak a könyvtárak felé, és próbálják kölcsönzéssel megoldani a könyvigényüket. Viszont a két folyamat együttes eredménye, hogy egyre kevesebb kötetet tudják a könyvtárak a növekvő olvasói réteg igényeit kielégíteni.

Az 1990-es évek könyvtári állományát nagymértékben befolyásolta a rendszerváltást követő selejtezések száma, ezért érdemes megnézni a következő 10 évben hogyan alakultak a könyvtárak mutatói. (Sajnos a Központi Statisztikai Hivatal adatszolgáltatása ezekre az évekre vonatkozóan már nem tartalmaz pénzügyi értékeket).

8. Magyarország könyvtárai és állományuk, 2000-2011

Év	intézmények száma	könyvállomány, ezer egység	beiratkozott olvasó, ezer	kölcsönzött könyvtári egység, ezer
2000	8640	142545	11756	41434
2001	9363	148251	12956	42574
2002	9086	152194	11477	42159
2003	9328	156159	12055	41612
2004	9311	158333	11764	40009
2005	9185	162033	12191	38785
2006	8166	146897	12350	38165
2007	7846	149404	11172	34175
2008	7700	146923	11414	33188
2009	7780	151120	11248	33897
2010	7712	152453	11441	33149
2011	7101	150189	11116	31748

Látható, hogy a második évezred első tíz évében a magyarországi könyvtárak száma folyamatosan csökken (folytatva az 1990-es években tapasztalt tendenciát). Az olvasói létszám és a kötetek száma pedig stagnál.

4.2.4 Regressziószámítás

A regressziószámítás segítségével lehetőségünk van meglévő adataink alapján előre jelezni, megbecsülni a következő adatot. A becslést első lépésként grafikusán ábrázoljuk az adatsort.

A GRAPHS / CHART BUILDER... menüpontnál adjuk meg az elemzendő mezőket ábrázolásra, ehhez előbb válasszuk a Line típus első fajtáját, és az elemzendő mezőket húzzuk az Y-Axis területre, az évszámokat (ha van) húzzuk az X-axis? területre. A kész diagramot vizsgáljuk meg, és becsljük meg melyik trendvonal fog illeszkedni leginkább:

- lineáris;
- logaritmikus;
- polinomiális;
- hatvány;
- exponenciális;
- mozgó átlag.

☙ Nézzünk meg néhány példát!

Lineáris trend



A magyarországi népesség számának 2001 és 2011 közötti alakulását mutatja az alábbi táblázat. Határozzuk meg a szolgáltató helyek számának várható értékét 2012-ben!

9. A népesség száma, 2001-2011

Év	A népesség száma, január 1.
2001	10 200 298
2002	10 174 853
2003	10 142 362
2004	10 116 742
2005	10 097 549
2006	10 076 581
2007	10 066 158
2008	10 045 401
2009	10 030 975
2010	10 014 324
2011	9 985 722

Első lépéseként ábrázoljuk grafikusan az adatokat! (Vonaldiagram esetén könnyebb felismerni a legjobban illeszkedő trendvonalat.)

A diagramvonal leginkább egyenesre hasonlít, kicsiny kilengésekkel.

Adjuk ki az ANALYZE / REGRESSION / LINEAR utasítást, ahol az x adatsor, azaz az évszámok képviselik a független (Independent(s)) változót, a tényleges adatok az y adatsor a Függő (Dependent) változóhoz kerül.

A Statistics nyomógomb használatával beállíthatjuk, hogy mutassa az R-négyzet értékét (R squared change). Az R-négyzet értéke a trendvonal megbízhatóságát jellemzi. Ha a trendvonal tökéletesen illeszkedik a grafikonunkhoz, az R-négyzet értéke 1. Ezért törekedni kell olyan trendvonal választására, ahol az R-négyzet értéke közel van 1 egészhez. (Minél közelebbi az érték, annál megfelelőbb a trendvonal.)

Eredményként az alábbi táblázatot kapjuk:

Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	VAR00002 ^b	.	Enter

a. Dependent Variable: VAR00001

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,993 ^a	,986	,985	,41084

a. Predictors: (Constant), VAR00002

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	108,481	1	108,481	642,686	,000 ^b
	Residual	1,519	9	,169		
	Total	110,000	10			

a. Dependent Variable: VAR00001

b. Predictors: (Constant), VAR00002

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2498,161	19,414		128,678	,000
	VAR00002	-4,879E-005	,000	-,993	-25,351	,000

a. Dependent Variable: VAR00001

40. ábra: Lineáris regresszió

A lineáris trend egy regressziós egyenes, amelyet egyszerű lineáris adathalmazokhoz használhatunk. A lineáris trendvonal rendszerint valamilyen érték egyenletes növekedését vagy csökkenését mutatja, ami jelen esetben csökkenés lesz.

A lineáris trendvonal a legkisebb négyzetek módszere alapján határozza meg a legjobban illeszkedő görbét az $y=mx+b$ függvényvel, ahol m a meredekség és b a tengelymetszet.

Visszatérve az eredménypanelhez, látható, hogy a külön táblázat mutatja az R-négyzet értékét. Ez azt tükrözi, hogy az eredményként létrejött trendvonal jól fedi az eredeti függvényünket, mivel az R-négyzet értéke 0,9862, ami egészen megközelíti az 1 egységet.

Az ANOVA táblázat azt mutatja, hogy a két adatsor között van-e összefüggés (lásd. varianciaanalízis).

A feladat szempontjából a leglényegesebb táblázat a COEFFICIENTS táblázat! Itt találhatjuk a lineáris trendvonal $y=mx+b$ függvényének paramétereit. Az első sorban található CONSTANS érték mutatja meg a B értékét, az alatta található szám pedig az m meredeksége a függvénynek. Az előjel a meredekség irányát adja meg.

Ha ezt behelyettesítjük a képletbe, és 2012-s x értékhez szeretnénk meghatározni a várható népességet, akkor a következő képletet kapjuk:

$$y = 4879 \cdot 2012 + 2498,161 = 9.816.572,98161$$

Kerekítve 2012-ben 9.816.572 lakos várható Magyarországon.

<<
<
<
<
<
<
<
<
<
<

Megjegyzés: a tankönyv írásakor már rendelkezésre állt az adat a 2012. január 1-jei népesség számáról: 9.957.731 fő. Tehát a rendszer a trend függvényvel és grafikus ábrázolással 141.159 fő eltéréssel határozta meg a várható értéket.


4.2.5 Faktoranalízis

Az eddig bemutatott elemzések kettő változót vettek figyelembe. „Az elemzések során gyakran kettőnél több változót kell figyelembe venni az adott probléma megoldása során. Több változónak nagy elemszámú mintán történő mérése során óriási adathalmazt egy egységként kezelni bonyolult feladat. A kapcsolatok feltárásánál több, egymástól is függő változó kapcsolat lehetőségét elemezve kell a feladatot megoldani, melynek elemzése és az eredmények értelmezése a faktoranalízis segítségével történhet.”⁴

Adott: egy sokváltozós mintaállomány, ahol a faktorok korrelálatlanok és a vizsgálat kezdetén még nem ismertek. A faktoranalízist a regresszióanalízistől az

⁴ Székelyi Mária, Barna Ildikó: Túlélőkészlet az SPSS-hez. Többváltozós elemzési technikákról társadalomkutatók számára. Typotex, Budapest, 2002.

különbözteti meg, hogy a független változók ismertek. Egy adatállományon a **faktoranalízis** csak akkor végezhető el, ha az adatok összefüggnek, más szóval korreláltak, minek értelmében a változók redundáns információkat hordoznak.

 **A faktoranalízis a változók száma csökkentésének a legelterjedtebb módszere. A jelenség feltárását szolgáló vizsgálati módszerek, amelyek a mért változók háttérében lehetnek, egymástól függenek, és a jelenségekre magyarázatot adnak.**

A változók számának csökkentése a statisztikai mintában lévő információ-lehetőség csökkentésével ugyanazt a jelenséget írja le kevesebb változóval. A feladat a sokváltozós adatállomány jellemzése a változónál kisebb számú, célszerűen választott, ún. faktorra oly módon, hogy a faktorok az eredeti lehetőség szerinti legtöbb információt tartalmazzák, és az így azonosított faktorokat célszerű értelmezni és elnevezni, hiszen ezek az eljárás kezdetén még ismeretlenek. Másik fontos célkitűzés, hogy a nagyszámú változó közötti korrelációs struktúrát írjunk le kevés számú látens változó, ún. faktor segítségével. A faktoroknak fizikai jelentésük nincs, közvetlenül nem figyelhetőek meg, nem mérhetőek, és létezésük csak elképzelhető az eredeti változók alapján.

A faktoranalízis alapfeltevése, hogy ezek látens változók. A faktoranalízis során a faktorok meghatározása a vizsgált változók korrelációs mátrixából kiindulva:

- Ha a változó nem korrelál más változókkal, nagy valószínűséggel önálló faktorra rendelkezik.
- Ha két vagy több változó között szoros a korreláció, akkor feltételezhető, hogy egy vagy több közös faktorra rendelkeznek.

A faktoranalízis alkalmazási feltételei:

- ha a korrelációs mátrix alapján a változók úgy csoportosíthatóak, hogy az egy csoporton belüli változók között viszonylag magas a korreláció, ezzel szemben a csoportok között alacsony. (Egy ilyen csoport mögött egy faktor áll.)
- a parciális korrelációk kicsik,
- a **Kaiser-féle mutatószám** (0 és 1 közé eső érték) az adatok összefüggő voltának, korrelációs vizsgálatának módszere, amelyet Kaiser–Meyer–Olkin-statisztikának is neveznek. Ha ez a mutatószám 0,8-nél nagyobb, akkor ajánlott, ha ez a mutatószám viszont 0,5-nél kisebb, akkor nem ajánlott faktoranalízis végrehajtása. A faktoranalízis egyaránt támaszkodhat a kovariancia, illetve a korrelációs mátrix elemzésére. A Kaiser–Meyer–Olkin-mérték az alábbi képlet alapján határozható meg:

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p \rho_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}$$

41. ábra: A Kaiser-féle mutatószám

Ahol:

r_{ij} – az i-edik és a j-edik változók korrelációs együtthatója,

$\rho_{ij} = \frac{R_{ij}}{\sqrt{R_{ii} \cdot R_{jj}}}$ – az i-edik és a j-edik változó parciális korrelációs

együtthatója,

- ha a KMO értéke $\geq 0,5$ abban az esetben az adatok alkalmasak a faktoranalízisre,
- ha a KMO értéke $< 0,5$ abban az esetben az adatok nem alkalmasak a faktoranalízisre.

A faktoranalízis alkalmazási területei:

- A nagyszámú és egymással korreláló változó között tanulmányozhatjuk a kapcsolatokat úgy, hogy a változókat kisebb számú, ún. faktorokba rendezzük, amelyekben belül a korrelációk nagyobbak, mint ezeken kívül.
- A faktorok a hozzájuk tartozó változók alapján értelmezhetők.
- A faktoranalízis segítségével a nagyszámú populáció a kisebb számú faktorok, a faktorpontok segítségével mennyiségileg áttekinthetőbbé válik.

A faktormodell fogalma, felépítése

Meghatározza, hogyan függnek az egyes változók a faktoroktól, mely lineáris kombinációval állíthatók elő. Tehát a főkomponens-analízissel szemben, ahol az egyes főkomponenseket állítottuk elő az eredeti változók lineáris kombinációjaként, itt az egyes változók fejezhetők ki a faktorok lineáris függvényeként. Fontos tudni, hogy faktoranalízist többféle módszerrel hajthatunk végre, a legfontosabbak ezek közül a főkomponenses módszer, a főfaktor-analízis és a maximum likelihood faktoranalízis.

A faktort számának megválasztása

A faktoranalízis az adatrendszer belső struktúráját, az adatrendszer egészét látva egyenrangúnak tekinti a változókat. A faktoranalízis célja a jelenséget leíró változók „mögött” megkeresni olyan rejtett változókat, amelyek a vizsgált jelenséget megmagyarázzák, számuk kisebb, mint az eredeti változóké, és egymástól függetlenek.

A faktoranalízis során a faktorok meghatározásakor a vizsgált változók korrelációs mátrixából kell kiindulni. Amelyik változó nem korrelál más változókkal, nagy valószínűséggel önálló faktorral rendelkezik. Ha viszont két vagy több változó között szoros korreláció van, akkor feltételezhető, hogy egy vagy néhány közös faktorral rendelkeznek.

A faktoranalízis modelljében a következő faktorok különböztethetők meg:

- közös faktor (több változót befolyásol),
- általános faktor (az összes változóra hatással van),
- csoportfaktor (nem az összes változót befolyásolja),
 - egyedi faktor (csak egyetlen változót befolyásol),
 - hibafaktor (mérési, becslési hiba hatása).

Egy-egy változót eltérő súllyal befolyásolhatnak a különböző faktorok, másrészt egy faktor eltérő súllyal befolyásolja az egyes változók értékét.

Az eredeti változók helyett meghatározott, hipotetikus változók, ún. faktorok tartalmazzák a rendszerről ismert információink nagy részét annak ellenére, hogy számuk kisebb. A faktoroknak nincs semmilyen fizikai jelentésük, közvetlenül nem figyelhetők meg, nem mérhetők, létezésüket csak feltételezhetjük az eredeti változók kapcsolatai alapján. A változók számának csökkentése azt jelenti, hogy a statisztikai mintában lévő információ lehetőleg kismértékű csökkentésével ugyanazt a jelenséget kevesebb változóval írjuk le.

A különböző faktorok hatásainak figyelembevételével az X változó az alábbiak szerint írható fel:

$$X_i = a_{i1} \cdot F_{i1} + a_{i2} \cdot F_{i2} + \dots + a_{iq} \cdot F_{iq} + b_{im} \cdot F_{im} + e_i \cdot F_i$$

42. ábra: X változó

ahol:

- a: a közös faktorok súlya
- b: az egyedi faktorok súlya

c: a hibafaktorok súlya

A feltételezés alapján a hibakomponens korrelálatlan a közös, illetve az egyedi faktorokkal, valamint a hibakomponensek függetlenek.

A standardizált változó szórásnégyzete:

$$s = \sum_{j=1}^q a_{ij}^2 + b_{im}^2 + e_i^2 = 1$$

43. ábra: Standardizált változó szórásnégyzete

A megfigyelt értékek mátrixa, mely a faktoranalízis bemeneti (input) adathalmazának tekintendő:

$$X = \begin{bmatrix} x_{11}, x_{12}, \dots, x_{1p} \\ \dots, \dots, \dots, \dots, \\ x_{n1}, x_{n2}, \dots, x_{np} \end{bmatrix}$$

44. ábra: Megfigyelt értékek mátrixa

ahol:

p: a változók száma

n: a mintaelemek száma

A faktoranalízis lépéseinek fázisai

- Minden változóra meg kell határozni az átlagot és a korrigált tapasztalati szórást.
- Minden adatból ki kell vonni a változókhoz tartozó adatok átlagát.
- Az eredményt el kell osztani a korrigált tapasztalati szórással.
- A feladat megoldása során olyan új F_1, F_2, \dots, F_k valószínűségi változókat kell keresni, ahol az F_k faktorok közös jellemzői:
 - számuk maximum p,
 - normális eloszlásúak
 - korrelálatlanok (bármely kettő korrekciós együtthatója zérus).

A fenti mátrixból az X_i valószínűségi változók és a faktorok közötti kapcsolatok az alábbiak alapján képezhetők:

$$\begin{aligned}
 X_1 &= a_{11} \cdot F_1 + a_{12} \cdot F_2 + \dots + a_{1k} \cdot F_k + W_1 \\
 X_2 &= a_{21} \cdot F_1 + a_{22} \cdot F_2 + \dots + a_{2k} \cdot F_k + W_2 \\
 &\dots\dots \\
 X_p &= a_{p1} \cdot F_1 + a_{p2} \cdot F_2 + \dots + a_{pk} \cdot F_k + W_p
 \end{aligned}$$

45. ábra: Kapcsolat

ahol:

W_1, W_2, \dots, W_p : egyedi faktorok, mivel egyenként csak egy változó kifejezésében szerepelnek

F_1, F_2, \dots, F_k : közös faktorok

A W -k és az F -ek korrelálatlanok egymással. A W értékétől függően, ha W értéke nagy, a faktoranalízis nem sikeres, ha W értéke kicsi, abban az esetben jó eredményt kaptunk.

a_{1j} – a faktorsúly, amely azt fejezi ki, hogy az F_1 faktor milyen súllyal szerepel az X_1 meghatározásában.

Tekintsük át a faktoregyütthatók és a faktorsúlyok közötti különbséget:

- A faktoregyütthatók a faktorok együtthatói a faktormodellben, melyek a megfelelő változó és faktor közötti korreláció nagyságát mérik.
- A faktorsúlyok ezzel szemben azt mondják meg, hogy mennyi a bevezetett új, közös faktorok értéke az egyes megfigyeléseknél.

4.2.6 Parciális korreláció

Parciális korrelációt akkor alkalmazunk, ha két adatsor között sejtjük a kapcsolatot, de nem tudjuk kimutatni, mert egy harmadik adatsor eltakarja az összefüggést.

- ☞ **A parciális korreláció két változó kapcsolatának erősségét és irányát adja meg, a többi változó hatásának figyelembe vétele nélkül.**
- ☞ **Parciális korreláció számítása során a függő és egy meghatározott független változó közötti korreláció mérése valósul meg, úgy, hogy minden más változó konstansként szerepel.**
- ☞ **A parciális korreláció képlete az alábbi, ahol x és y változó kapcsolatából szeretnénk kiküszöbölni a z változó hatását:**

$$r_{XYZ} = \frac{r_{XY} - r_{XZ}r_{YZ}}{\sqrt{(1-r_{XZ}^2)(1-r_{YZ}^2)}}$$

46. ábra: A parciális korreláció képlete

☼ Nézzük meg a könyvtárak adatait 2000 és 2011 között.

Készítsünk a táblázat adataiból korrelációs mátrixot!

➔ Correlations

[DataSet6] C:\Tunde_SPSS\Konyvtarak_alakulasa_2011.sav

		Correlations				
		Év	Könyvtárak_száma	Könyvállomány	Olvasok_száma	Kölcsönzések_száma
Év	Pearson Correlation	1	-,852**	,061	-,617*	-,957**
	Sig. (2-tailed)		,000	,850	,033	,000
	N	12	12	12	12	12
Könyvtárak_száma	Pearson Correlation	-,852**	1	,435	,689*	,927**
	Sig. (2-tailed)	,000		,158	,013	,000
	N	12	12	12	12	12
Könyvállomány	Pearson Correlation	,061	,435	1	,082	,124
	Sig. (2-tailed)	,850	,158		,799	,702
	N	12	12	12	12	12
Olvasok_száma	Pearson Correlation	-,617*	,689*	,082	1	,697*
	Sig. (2-tailed)	,033	,013	,799		,012
	N	12	12	12	12	12
Kölcsönzések_száma	Pearson Correlation	-,957**	,927**	,124	,697*	1
	Sig. (2-tailed)	,000	,000	,702	,012	
	N	12	12	12	12	12

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

47. ábra: Korrelációs mátrix (Könyvtárak 2001-2011)

Leolvasható, hogy az intézmények száma erős korrelációban van az évekkel, tehát kimutathatóan csökken a könyvtárak száma.

Viszont a beiratkozott olvasói létszám és az év kapcsolata megfordult az előző 10 évhez képest, és a -0,617-es korrelációs együttható jelzi, hogy a növekvő évszámhoz csökkenő olvasószám tartozik.

Ami érdekes adat, hogy a könyvállomány és a kölcsönzött állomány között nem látunk összefüggést. A 0,124-os korrelációs együttható azt mutatja, hogy nincs semmi összefüggés a két adatsor között.

Pedig valami azért elvárható lenne!

Ilyen esetekben kell végiggondolni, hogy mi takarhatja el ezt az összefüggést!

Az adatokat megnézve két változó jöhet szóba:

- Az olvasók számának alakulása
- Az évek hatása

☼ Kiszöböljük ki a két változót parciális korreláció alkalmazásával!

- Olvasók számának kikiszöbölése:

X változó: könyvállomány

Y változó: kölcsönzött könyvtári egység

Z változó olvasók száma

korreláció $r_{x,y}$ (könyvállomány; kölcsönzött könyvtári egység): 0,1236

korreláció $r_{x,z}$ (könyvállomány; olvasók száma): 0,0825

korreláció $r_{y,z}$ (kölcsönzött könyvtári egység; olvasók száma): 0,6968

Parciális korreláció $r_{xy.z} = 0,0925$

A folyamatot SPSS szoftverrel a ANALYZE/ CORRELATE /PARTIAL... menüponttal kell kivitelezni, ahol a

Variable(s) ablakba húzzuk az X és Y változót, azaz a könyvállomány és a kölcsönzött könyvtári egység mezőit

A Controlling for: ablakba pedig a kikiszöbölendő Z változót, azaz az Olvasók számát, ahogy az alábbi képen látható:

➔ Correlations

[DataSet6] C:\Tunde_SFSS\Könyvtarak_alakulasa_2011.sav

Correlations						
		Év	Könyvtarak_száma	Könyvállomány	Olvasók_száma	Kölcsönzések_száma
Év	Pearson Correlation	1	,852 ^{**}	,061	-,617 [*]	-,957 ^{**}
	Sig. (2-tailed)		,000	,850	,033	,000
	N	12	12	12	12	12
Könyvtarak_száma	Pearson Correlation	-,852 ^{**}	1	,435	,689 [*]	,927 ^{**}
	Sig. (2-tailed)	,000		,158	,013	,000
	N	12	12	12	12	12
Könyvállomány	Pearson Correlation	,061	,435	1	,082	,124
	Sig. (2-tailed)	,850	,158		,799	,702
	N	12	12	12	12	12
Olvasók_száma	Pearson Correlation	-,617 [*]	,689 [*]	,082	1	,697 [*]
	Sig. (2-tailed)	,033	,013	,799		,012
	N	12	12	12	12	12
Kölcsönzések_száma	Pearson Correlation	-,957 ^{**}	,927 ^{**}	,124	,697 [*]	1
	Sig. (2-tailed)	,000	,000	,702	,012	
	N	12	12	12	12	12

** . Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

48. ábra: Parciális korreláció beállítása

Az összefüggés nem mutat korreláltságot, tehát az olvasói létszám nem felel az összefüggést.

- Nézzük meg az évi csökkenés takarja-e el:

X változó: könyvállomány

Y változó: kölcsönzött könyvtári egység

Z változó évek száma

korreláció x,y (könyvállomány; kölcsönzött könyvtári egység): 0,1236

korreláció x,z (könyvállomány; évek száma): 0,0615

korreláció y,z (kölcsönzött könyvtári egység; évek száma): -0,9570

Parciális korreláció xy.z= 0,623

Értelmezés: Ha kiküszöböljük az évek hatását a két adatsorra, akkor kimutathatóvá válik a két adatsor kapcsolata: 0,6296-os korreláció van a két adatsor között, ami azt jelenti, hogy a könyvek számának alakulása kapcsolatban áll a kölcsönzött példányok számával, pozitív összefüggéssel, azaz a könyvek számának emelkedése a kölcsönzött példányok emelkedésével jár.

Ennél a példánál is fontos hangsúlyozni, hogy nem lehet tudni, melyik változó hat a másikra:

- Ha növekszik a könyvtárak kötetszáma, akkor jobban találnak maguknak köteteket az olvasók, és többet kölcsönöznek. (Vagy fordítva: a csökkenő kötetszám hatására a kölcsönzés is csökken).
- Nagyobb a kölcsönzési igény az olvasók részéről, amit jeleznek a könyvtárak felé (azok előjegyzéseket vesznek fel), és ennek hatására növelik a kötetek számát. (Vagy csökkenő kereslet hatására kevesebb kötetet vásárolnak a könyvtárak).

4.2.7 A Spearman-féle rangkorreláció


A Spearman-féle rangkorreláció két ordinális adatsorra vonatkozva adja meg, van-e összefüggés a két adatsor között. A korrelációs együtttható mértéke, iránya és képlete megegyezik az előbb bemutatottakkal.

4.2.8 Klaszteranalízis

Ha együttes előfordulásokat, kapcsolatokat szeretnénk vizsgálni, viszont nagy az adatmennyiség, akkor a klaszterezéshez kell folyamodnunk.

A klaszteranalízist nagy mennyiségű adat redukciójához használjuk.

A megvalósítás során törekszünk olyan csoportok, klaszterek létrehozására, melyek elemei a lehető legszorosabban kapcsolódnak egymáshoz, és minél jobban eltérnek a többi klaszter elemeitől.⁵

 **A klaszteranalízis során egy rendezetlen adathalmazból igyekszünk egy strukturált rendszert létrehozni, melyben az egymásra hasonlító elemek klasztereket alkotnak.**

Ezt kétféle technika alkalmazásával érhetjük el:

- Az egyik lehetőség, ha adottak a klaszterek, és az adatokat ezekbe soroljuk be.
- **A hierarchikus módszer** során nem rendelkezünk előre létrehozott osztályokkal, hanem az adatelemzés során alkotjuk meg azokat.
 - A klaszteralkotás történhet alulról is: minden elemet külön osztálynak tekintünk, és az egymáshoz közel állókat összevonjuk. Ezt **összevonó eljárásnak** hívják.
 - A technikai kivitelezés során több módszert használhatunk. Leggyakrabban a szomszédos elemek összevonása. Ennek alkalmazása során először minden elemet önálló klaszternek tekintünk, majd megkeressük a két legközelebbi klasztert, és összevonjuk azokat. Ezt addig folytatjuk, amíg van két különálló klaszterünk.
 - A klaszteralkotás történhet felülről, amikor is az egységes egészet osztjuk részekre. Ezt **felosztó eljárásnak** nevezzük.

4.3 ÖSSZEFOGLALÁS, KÉRDÉSEK

4.3.1 Összefoglalás

A fejezetben a matematikai statisztika azon eszközeivel ismerkedhettünk meg, melyek alkalmasak arra, hogy kimutassák, van-e összefüggés a vizsgált adatsorok közt úgy, hogy ennek alapja nem a populáció összes eleme, hanem a reprezentatív minta.

A fejezetben bemutatásra kerültek az intervallum vagy mérhető adatok esetén használható statisztikai módszerek.

Két adatsor esetén korrelációs számítással mutathatjuk ki az összefüggéseket, még pedig az összefüggés irányát, illetve erősségét.

⁵ CASTANO, S. – ANTONELLIS, V. De – FUGINI, M.G. – PERNICI: Conceptual schema analysis: techniques and application. – In: AMC Transactions on Database Systems, Vol. 23. no. 3. September 1998. pp. 297–298.

Kettő vagy több adatsor esetén alkalmazhatjuk

- a korrelációanalízist, mely továbbra is két adatsor között határozza meg a korrelációs összefüggést, de mindezt egy mátrixban kapjuk meg, ahonnan bármely két adatsor korrelációs együtthatóját azonnal le tudjuk olvasni.
- a regresszióanalízist, mely két összefüggő (egymással korreláló) adatsor esetén lehetőséget ad az egyik adatsor hiányzó értékének becslésére a másik adatsor értékének birtokában. A regresszió típusa lehet:
 - lineáris;
 - logaritmikus;
 - polinomiális;
 - hatvány;
 - exponenciális;
 - mozgó átlag.

Kettőnél több adatsor esetén használható matematikai statisztikai lehetőségek:

- Parciális korrelációszámítás, melynek alkalmazása akkor ajánlott, ha két adatsor között sejtünk kapcsolatot, de a korrelációs érték ezt nem támasztja alá, viszont van olyan adatsor, ami elfedheti ezt az összefüggést. A parciális korreláció kiküszöböli a harmadik változó hatását, így tisztán a két adatsor kapcsolatára kapunk együtthatót.
- Faktoranalízis, melynek célja a jelenséget leíró változók „mögött” megkeresni olyan rejtett változókat, amelyek a vizsgált jelenséget megmagyarázzák, számuk kisebb, mint az eredeti változóké, és egymástól függetlenek.
- A klaszteranalízis célja rendezettség létrehozása, melynek alkalmazásakor egy rendezetlen adathalmazból igyekszünk egy strukturált rendszert létrehozni, ahol az egymásra hasonlító elemek klasztereket alkotnak.

Az előzőekben az intervallumadatok matematikai statisztikai mutatóit foglaltuk össze, melyek a kapcsolatok kimutatására szolgálnak.

Ha nem mért adatok, hanem rangsorolt, ordinális adatsorok állnak a rendelkezésünkre, akkor a Spearmann-féle rangkorrelációs együttható mutatja meg a kapcsolat erősségét és irányát.

4.3.2 Önellenőrző kérdések

1. Foglalja össze, mely esetekben használhatunk korrelációszámítást!
2. Mekkora intervallumon mozog a korrelációs együttható?

3. Miben más a korrelációanalízis a korrelációs számítástól?
4. Miben más a korrelációs számítás a parciális korrelációs számítástól?
5. Mi a faktoranalízis célja?
6. Milyen elemzést használna ordinális adatok összefüggésének kimutatásához?

5. MAGASABB SZINTŰ ÉRTÉKELÉSI MÓDSZEREK A GYAKORLATBAN

5.1 CÉLKITŰZÉSEK ÉS KOMPETENCIÁK

A fejezetben a matematikai statisztikai vizsgálatok azon csoportját tekintjük át, melyek célja az adatsorok közti különbségek kimutatása.

Szignifikancia-vizsgálatokként is szokták emlegetni az alábbi módszereket, melyek célja annak eldöntése, hogy az adatsorok között van-e szignifikáns/kimutatható különbség.

A fejezetben megismerhetjük a hipotézisvizsgálatok alapját képző elméletet, a nullhipotézis fogalmát, valamint a szignifikáns különbség jelentőségét.

A konkrét módszerek közül a fejezetben előbb hipotézisvizsgálatokkal ismerkedhetünk meg, melyek intervallumadatok esetén használhatóak.

- Egy minta esetén egymintás t-próbát használhatunk,
- két minta esetén kétmintás t-próbát alkalmazunk, melynek előpróbája az F-próba
- kettőnél több adatsor esetén használjuk a varianciaanalízist.

Ordinális adatok esetén a következő próbákat ismerhetjük meg a fejezetben:

- Wilcoxon-próba
- Mann–Whitney-próba
- Kruskal–Wallis-próba

5.2 TANANYAG

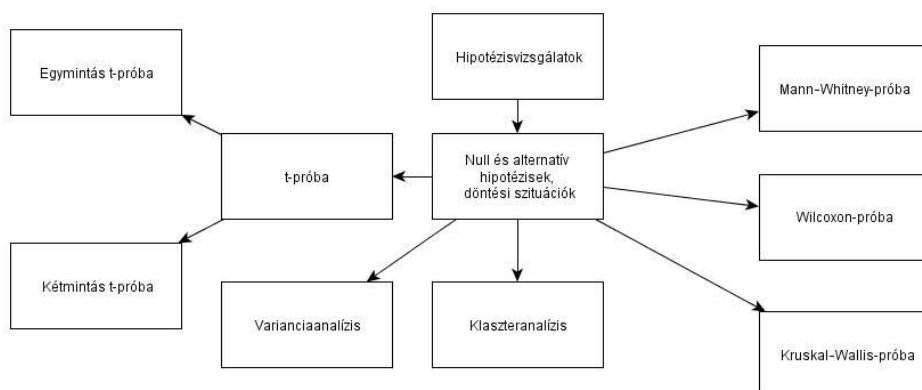
Hipotézisvizsgálatok

Null- és alternatív hipotézisek, döntési szituációk

t-próba

Varianciaanalízis

A Mann–Whitney-próba, Wilcoxon-próba, Kruskal–Wallis-próba értelmezése



49. ábra: Fogalomtérkép

5.2.1 Hipotézisvizsgálatok

A hipotézisvizsgálatok nélkülözhetetlenek a kutatási folyamat utolsó lépéséhez. Egy kutatás elvégzése nem ér véget a felmérés lebonyolításával, sőt még az adatok elemzésével sem. Az egész folyamat lényege a következtetések levonásakor, az általánosításkor valósul meg. Ekkor tudjuk meg, sikeres volt-e a munkánk, választ kaptunk-e a kutatás kezdetekor kitűzött célokra, és eredményeink felhasználhatók-e a problémák megfogalmazásához, esetleg magyarázatok, ajánlások készítéséhez.

Ha a mintavételi eljárásunk során a teljes felmérés mellett döntöttünk, tehát annak a populációnak, melyre vonatkozóan következtetéseinket le szeretnénk vonni, minden tagja részt vett a felmérésben, akkor nincs szükségünk a hipotézisvizsgálatokra.

Viszont a kutatások többségében nincs lehetőségünk a teljes populáció felmérésére, ilyenkor az egyik mintavételi eljárást alkalmazva kiválasztunk egy mintát, melynek tagjait megvizsgáljuk, és a nyert adatok alapján általánosítunk. Az általánosítást csak akkor tehetjük meg, ha tudjuk igazolni, hogy a felmérés adataiból származó eredmények nem a véletlen következtében jöttek létre, hanem, ha a felmérésünket más mintán is elvégeznénk (a populáció más tagjaival), ugyanezeket az eredményeket kapnánk.

A matematikai statisztikai vizsgálatok teszik lehetővé, hogy megállapítsuk, az eredményeink alapján kapott különbségek a véletlen következtében állnak elő, vagy sem. Ha kimutathatóan nem a véletlennek köszönhető az értékekben tapasztalt eltérések, akkor, szignifikáns különbségről beszélünk, és az általánosítás elvégezhető.

5.2.2 Null- és alternatív hipotézisek, döntési szituációk

A kutatások nagy részét képzik az egy- vagy többcsoportos kísérletek, melyek lezárásakor rendelkezésünkre áll:

- önkontrollos kísérlet esetén a csoport teljesítményének eredménye a kísérlet előtt és után,
- illetve többcsoportos kísérlet esetén a kísérleti és a kontrollcsoport eredménye.

A hipotézisvizsgálatok során egy null hipotézissel indulunk, azaz feltételezzük, hogy a rendelkezésünkre álló adatok közti különbségek a véletlennek köszönhetőek, tehát nincs szignifikáns különbség az adatok között. Statisztikai vizsgálatokkal kell bizonyítanunk, hogy nullhipotézisünk igaz-e vagy sem.

Ha nullhipotézisünk igaznak bizonyul, akkor nem vonhatunk le semmilyen következtetést, hiszen az eltérések, melyeket a minta adatai alapján látunk, lehetnek a véletlen művei is, azaz a minták közti különbségek nem elég jelentősek az általánosításhoz.

Az általánosítás csak abban az esetben tehető meg, ha nullhipotézisünk hamisnak bizonyul, tehát a minták közti különbség olyan nagy, hogy az már nem a véletlen műve, azaz a különbség szignifikáns. Ekkor minden állítás, melyet a mintákra vonatkozóan teszünk, igaz arra a populációra, melyet a minta reprezentál.

A szignifikáns különbségek diagrammal is szemléltethetők.



50. ábra: Minták közti átfedés

Ha a két mintát jellemző Gauss-görbének kicsi az átfedése, akkor a minták közti különbség szignifikáns.

Ha a mintákat jellemző átlagok közti különbség sokkal kisebb, mint a szórási, a mintákat jellemző Gauss-görbék között nagy lesz az átfedés. Ez esetben nincs szignifikáns különbség a minták között.

Fontos kérdés, hogy mikor tekinthetjük az eltérést szignifikánsnak?

Abban az esetben, ha a két Gauss-görbe teljesen fedi egymást, biztosak lehetünk benne, hogy nincs eltérés a két minta eredményei között.

De mi a helyzet a többi esettel? Mikor tekinthetjük a különbségeket elég nagyoknak ahhoz, hogy azokat ne lehessen a véletlennek tulajdonítani?

100%-os bizonyossággal ritkán jelenthetjük ki, hogy a minták közti különbségek nem a véletlen következtében jöttek létre, viszont 95%-os valószínűségi szint felett már szignifikánsnak tekintjük a különbségeket. Azaz, ha a kutatási eredményekben a tévedés lehetősége nem nagyobb, mint 5%, szignifikáns a különbség.

A szignifikáns különbségek meghatározására szolgálnak a t-próbák.

5.2.3 t-próba

A t-próba két minta megállapítható tulajdonságai közötti különbség szignifikanciájának számszerűsítését adja meg. A t-próba arra az összefüggésre alapoz, hogy a számtani középértéktől számított 2 szórásonyi terjedelemben tartozik az adatok több mint 95%-a. A t-próba az átlagok, a szórák és a minta elemszámának figyelembevételével határozza meg, van-e szignifikáns különbség a két adatsor között. Ha a vizsgált minták számtani középértékének különbsége nagyobb, mint azok eloszlási szóráinak kétszerese, akkor a vizsgált minták számtani középértéke közötti különbség szignifikáns.

5.2.4 Egymintás t-próba

Ha önkontrollos kísérletet végeztünk, tehát egy mintával dolgoztunk, akkor a két adatsor a kísérlet előtti és a kísérleti változó hatására létrejött eredményeket mutatja. Ez esetben az egymintás t-próba elvégzésével tudjuk meghatározni a különbség szignifikanciájának a szintjét.

Egymintás t-próba esetén a következő képletet használhatjuk:

$$t = \frac{\bar{x} - m}{s/\sqrt{n}}$$

51. ábra: A t-próba képlete

A képletben alkalmazott jelölések:

- \bar{x} a vizsgált valószínűségi változó átlaga a mintában,
- s a vizsgált valószínűségi változó szórása,

- m a nullhipotézisben feltételezett átlagérték,
- n a minta elemszáma.

Nézzük meg a következő példát!

- ✿ Egy kollégiumban felmérést készítettek, mennyire elégedettek a diákok a kollégium felszereltségével. A 10 diák véleményét tartalmazza a táblázat.

Majd bevezettek néhány változtatást, és felmérték hogyan változott az elégedettség a diákok körében. Ajánlhatók-e hasonló változások a többi kollégiumban is?

10. Példa egymintás t -próba alkalmazására

	Elégedettség (%)	
	Változtatások előtt	Változtatások után
1. tanuló	82	80
2. tanuló	91	91
3. tanuló	67	66
4. tanuló	55	73
5. tanuló	25	42
6. tanuló	30	35
7. tanuló	38	44
8. tanuló	37	42
9. tanuló	43	51
10. tanuló	48	71

A felmérés alapján vannak olyan személyek, akik elégedettebbek, vannak, aki kevésbé. Átlagosan az elégedettség a felmérés előtt 51,6 volt, míg a felmérés után 59,5. Van tehát fejlődés, de ennek mértéke elég nagy ahhoz, hogy ajánljuk a bevezetést másoknak is, általánosítsuk az eredményeket?

Mivel itt ugyanazon személyek elégedettségét mérték a régi és az új rendszerben, önkontrollos kísérletről beszélhetünk, ezért a megbízhatóság eldöntéséhez egymintás t -próbát kell végeznünk. A művelet elvégzésére a T.PRÓBA függvény szolgál.

Az SPSS szoftver ANALYZE / COMPARE MEANS / PAIRED-SAMPLE T TEST parancsával tudjuk kivitelezni.

A függvénypanelen adjuk meg a két változót, és az OPTIONS nyomógombnál ellenőrizzük, hogy 95%-os valószínűséggel dolgozik-e a program.

→ T-Test

[DataSet7]

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Változások_előtt	51,60	10	22,072	6,980
	Változások_után	59,50	10	19,133	6,050

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Változások_előtt & Változások_után	10	,922	,000

Paired Samples Test

		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	Változások_előtt - Változások_után	-7,900	8,647	2,734	-14,086	-1,714	-2,889	9	,018

52. ábra: A t-próba eredménye

Ha számítógépet használunk a t-próba kiszámításához, akkor tehető meg az általánosítás, ha a kapott érték 5% alatti, ugyanis ekkor lesz a felmérés megbízhatósága 95% feletti. Az eredményül kapott t értéket kell figyelni az eredménypanelen, melynek, 0179-es értéke 1,79%-nak felel meg, tehát szintén levonhatjuk azt a következtetést, hogy a hasonló változtatások bevezetését minden kollégiumban ajánljuk.

A eredmény táblázat a t-próbán kívül megadja az átlag és szórás értékét is, valamint a két adatsor közti korrelációs viszonyszámot.

5.2.5 Kétmintás t-próba

Azoknál a kísérleteknél, ahol két csoportot vizsgálunk, és míg az egyiknél megváltoztatunk bizonyos tényezőket (a független változókat), és vizsgáljuk, hogy ezek milyen hatásokat váltanak ki (függő változó), addig a másik csoport a hagyományos módon, beavatkozás nélkül éli életét. A kísérlet végén elvégezzük felmérésünket, és az eredmények vizsgálatával kell döntenünk arról, hogy a kísérleti tényezők hatására bekövetkezett változások általános érvényűnek tekinthetők-e.

A kontrollcsoportos kísérleteknél kétmintás t-próbát kell alkalmazni a szignifikanciaszint meghatározásához.

Egyenlő szórásnégyzetű csoportok esetén a

- ☞ **homoscedasztikus t-próba** abból indul ki, hogy a két adathalmaz szórásnégyzete egyenlő.

Ez a t-próba csak akkor végezhető el, ha a vizsgálat alapjául szolgáló adatok varianciája nem tér el jelentős mértékben egymástól. Az F-próba meghatározásával adhatjuk meg a választ erre a kérdésre.

- ☞ **Az F-próba (Fisher–Snedecor-eljárás) értéke a két minta szórásnégyzetének hányadosa. (Minden esetben a nagyobb szórásnégyzetet kell osztani a kisebbel).**

A szignifikancia szintjét 95%-os valószínűség esetén fogadjuk el, mely akkor áll elő, ha a számítógéppel meghatározott F-próba értékének a fele 5% feletti.

Kézzel történő meghatározás esetén szükségünk van egy F-próba táblázatra, ahol megnézzük a két minta szabadságfokához tartozó sor (a kisebb minta elemszáma -1) és oszlopsor (a nagyobb minta elemszáma -1) találkozásánál lévő elvárt értéket. Ha a táblázatban szereplő érték nagyobb, mint az általunk kapott F-próba értékének fele, akkor elvégezhetjük a t-próbát, különben nem.

Ha $F_{\text{táblázat}} > F/2$, akkor H_0 : hamis

Ha $F_{\text{táblázat}} < F/2$, akkor H_0 : p valószínűséggel igaz

Ha az F-próba lehetővé teszi, meg kell határoznunk a t-próba értékét. Ha az érték 95 % feletti szignifikanciát mutat, általánosíthatjuk a kísérlet eredményeit. (Ne felejtjük, ha számítógép segítségével határozzuk meg a t-próba értékét, az 5 % alatti érték jelenti a szignifikanciát!)

A kétmintás t-próba meghatározása az alábbi képlettel lehetséges:

$$t'' = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sum_{i=1}^n (\bar{x} - x_i)^2 + \sum_{i=1}^m (\bar{y} - y)^2}{n + m - 2} \cdot \frac{n + m}{n \cdot m}}}$$

53. ábra: A kétmintás t-próba képlete

ahol

- \bar{x} az egyik minta átlaga,
- \bar{y} a másik minta átlaga,
- n az egyik minta elemszáma és
- m a másik minta elemszáma.

Nézzük meg ezt egy konkrét feladaton keresztül!



Az általános iskolákban a természetismeret tantárgy tanítása a hagyományos tanteremben zajlik a szokásos elméleti anyagok megtanításával és gyakorlati feladatok elvégzésével. Egy osztályban kísérleti jelleggel interaktív táblára készült tananyag oktatásával, a tanulók egyéni laptopokkal való ellátottságának biztosítása mellett, tanulhatják a tanulók a természetismeret anyagát. A félév végén ugyanazt a tantárgytesztet írták a hagyományos módon tanuló és az interaktív tananyagot használó osztály tagjaival. Az eredményeket az alábbi táblázat mutatja (maximálisan 25 pontot lehetett szerezni). Hasznosnak mondható-e az oktatóprogrammal megvalósított tanítás?

11. Példa két mintás t-próba meghatározására

Interaktív tábla használatával tanuló csoport eredményei (pont)	Hagyományos módon tanuló csoport eredményei (pont)
19	15
25	25
21	22
25	14
16	19
18	21
14	20
23	16
12	6
25	19
25	20
19	21
22	10
23	15
18	17
25	13
24	21
12	
23	
20	

Az adatok felvitele után határozzuk meg az F-próba és t-próba értékét, mely az SPSS szoftver használatánál egy lépésben előáll.

Ehhez használjuk, a `ANALYZE / COMPARE MEANS/ INDEPENDENT-SAMPLES T TEST` parancsot.

Ezután kapjuk meg a helyes eredményt: f-próba=34,2763%. Mivel az érték 5% feletti, elvégezhető a t-próba.

Ha megvizsgáljuk a felmérések átlagos eredményét, látható, hogy az interaktív táblával tanuló csoport 20,45-ös átlagpontoszámot ért el, míg a hagyományos módszerekkel tanuló csoport 17,29-es átlagpontoszámot ért el. A t-próba 4,4828%-os eredményének következtében kijelenthetjük, hogy a jobb eredmény az új módszernek tulajdonítható, és a tévedés lehetősége kisebb, mint 5%, azaz a két adathalmaz különbsége 95,5172%-os valószínűséggel nem a véletlennek tudható be.

Nem egyenlő szórásnégyzetű csoportok esetén

A t-próbákat akkor használhatjuk, ha meg szeretnénk állapítani, hogy két minta várható értéke egyenlő-e. Abban az esetben, ha a vizsgált csoportok szórásnégyzete különböző, akkor nem használhatunk két-mintás t-próbát, hanem a Welch-féle d-próbát kell használnunk.

5.2.6 Varianciaanalízis

A t-próbák – mint láthattuk – csak két adatsor közti különbség szignifikanciájának megállapítására alkalmasak, ha több adatsor áll rendelkezésünkre, például többcsoportos kísérletet végeztünk, és minden csoportnál más és más független változót vezetünk be, akkor a t-próba vizsgálatokkal csak párokban tudnánk összehasonlítani az eredményeket.

Egy többcsoportos kísérlet során úgy történik az új eljárás, az új módszer hatékonyságának vizsgálata, hogy minden csoportban ugyanazt a tényezőt változtatjuk meg, de másképp (ezek lesznek a független változók), és vizsgáljuk, hogy milyen hatással van a változás a függő változókra. Varianciaanalízis segítségével meghatározhatjuk a többdimenziós minta ugyanazon változója közötti különbség szignifikanciaszintjét.

A varianciaanalízist szórásanalízisnek is szokták nevezni, melynek lényege, hogy megmutassa, van-e szignifikáns eltérés a mintaátlagok között, miközben feltételeztük, hogy azonos varianciából vettük a mintákat.

- ☞ **Varianciaanalízisnek** nevezzük azt a statisztikai eljárást, mely több egydimenziós minta ugyanazon változója közötti különbség szignifikanciaszintjének összehasonlítását teszi lehetővé.

Mivel több mintát elemzünk, ezért különbséget teszünk belső és külső variancia között.

A belső variancia a minták elmei közötti különbségeket vizsgálja.

- ☞ **A csoporton belüli variancia** a mintaelemek csoportátlaguktól való eltéréseinek négyzetösszege osztva a minták szabadságfokával.

$$s_b^2 = \frac{\sum_{j=1}^h \sum_{i=1}^{n_j} (\bar{x}_j - x_{ij})^2}{N - h}$$

54. ábra: A belső variancia képlete

ahol x_{ij} a csoport minden elemének figyelembevételével számolt átlag,

N az összes mintacsoport egyedszámának összege,

h a minták száma.

A külső variancia a csoportok/minták közötti különbözőséget vizsgálja.

- ☞ **A külső, vagy más szóval: a minták közötti variancia** a minták egymáshoz viszonyított eltérései alapján meghatározott érték.

$$s_k^2 = \frac{\sum_{j=1}^h n_j (\bar{x} - \bar{x}_j)^2}{h - 1}$$

55. ábra: A külső variancia képlete

ahol:

\bar{x} - az egyes minták súlyozott számtani középértéke, mely az összes minta minden elemének átlaga. (A felmérés adatcsoportjainak minden egyes elemét összeadjuk, majd osztjuk a felmérésben részt vevő személyek számával).

Azaz:

7. Meghatározzuk az egyes minták átlagát.
8. Meghatározzuk az összes elemhez tartozó átlagot.
9. Vegyük a kettő különbségét, és emeljük négyzetre, majd szorozzuk meg az adott minta elemszámával.
10. Összegezzük minden mintára vonatkozóan.
11. Osszuk el a minták számának eggyel csökkentett értékével.

Az F értékének meghatározása

A varianciaanalízis során az előbb kiszámított külső és belső varianciák alapján meg kell határozni az F-próba értékét, mely a két variancia négyzetének hányadosa.

Annak eldöntésére, hogy az adatsorok közti különbségek szignifikánsak-e, a kapott értéket össze kell hasonlítani az F-eloszlás táblázatában lévő megfelelő értékkel.

Keressük ki a külső szabadságfoknak (h-1) megfelelő sor és belső szabadságfoknak (N-h) megfelelő oszlop találkozási pontján lévő F-értéket.

A szignifikanciaszint eldöntése:

- Ha F számolt értéke kisebb, mint a táblázat értéke, azaz $F < F_{\text{táblázat}}$, akkor a két variancia (belső és külső) nem különbözik egymástól, nincs szignifikáns különbség a csoportok között.
- Ha F számolt értéke nagyobb, mint a táblázat értéke, azaz $F > F_{\text{táblázat}}$, akkor a két variancia (belső és külső) különbözik egymástól. Ebben az esetben az eredmények szignifikánsan különböznek, a kapott adatokat általánosíthatjuk.

Egytényezős varianciaanalízis

Egytényezős varianciaanalízis esetén egy szempontra összpontosítva elemezzük a felmérés hatékonyságát.

A gyakorlatban feltételezünk egy nullhipotézist, majd a bizonyításához végezzük el a varianciaanalízist!

Varianciaanalízis készítéséhez válasszuk az ANALYZE/COMPARE MEANS/ONE-WAY ANOVA vagy egyszerűbben az ANALYZE → COMPARE MEANS → MEANS parancsot.

Az adatsorokat adjuk meg a Dependent List változóablakba, majd a Faktor oszlopot tegyük a Factor(s) változóhoz.

A MEANS parancs használatánál az Options nyomógombon az ANOVA tabla and ETA kapcsológombját be kell kapcsolni, és ekkor kapjuk meg az eredménytáblát.

ANOVA Table							
			Sum of Squares	df	Mean Square	F	Sig.
A * Faktor	Between Groups	(Combined)	288,011	1	288,011	2,917	,102
	Within Groups		2172,489	22	98,749		
	Total		2460,500	23			
B * Faktor	Between Groups	(Combined)	124,844	1	124,844	1,424	,246
	Within Groups		1929,156	22	87,689		
	Total		2054,000	23			
C * Faktor	Between Groups	(Combined)	6,501	1	6,501	,084	,775
	Within Groups		1866,461	24	77,769		
	Total		1872,962	25			

56. ábra: Varianciaanalízis

Az ANOVA table-n kívül kapunk egy a REPORT táblát is melyről leolvasható melyik csoportban beállított kutatási módszer a leghatásosabb (az átlagos eredményből), majd be kell bizonyítanunk, hogy nullhipotézisünk hamis.

Ehhez nézzük meg a varianciaanalízis táblázatát. A táblázat értelmezése

- a külső és belső eltérések négyzetösszegét (SS, azaz Sum of Squares),
- ezek szabadságfokát (df)
- és hányadosukat (Mean Square, $MS=SS/df$).
- Az F-érték a becsült külső és belső szórásnégyzetek hányadosa. (F)
- A p-érték megmutatja, hogy a nullhipotézis elvetése esetén mekkora a tévedésünk lehetősége. (Sig.)

A kiszámolt F-értéket össze kell hasonlítanunk a 95%-os valószínűségi szinthez tartozó F-eloszlás táblázatában a külső (csoportok közötti) variancia szabadságfokának megfelelő oszlop és a belső variancia (csoporton belüli) szabadságfokának megfelelő sor metszetében található értékkel. Ha az összehasonlításakor azt kapjuk, hogy

1. az F értéke kisebb, mint az F -eloszlás táblázat megfelelő értéke, akkor a két variancia nem különbözik lényegesen egymástól, tehát a képzelt (a nullhipotézisben feltételezett) populációnk létezhet.
2. az F értéke nagyobb, mint az F -eloszlás táblázat megfelelő értéke, akkor a két variancia lényegesen különbözik egymástól, tehát a képzelt (a nullhipotézisben feltételezett) populációnk nem létezik.

Az ANOVA tábla használatára láthatunk példát a 8.2.2. fejezetben levezetve.

Két- vagy többtényezős varianciaanalízis

Független csoportok esetén, ha többféle szempont szerint végezzük az elemzést, akkor két- vagy több szempontos varianciaanalízissel hasonlítsuk össze az átlagokat.

5.2.7 A Mann–Whitney-próba, Wilcoxon-próba, Kruskal–Wallis-próba értelmezése

„A **Mann–Whitney-próba** a független minták összehasonlítását szolgáló eljárás. A két mintát együtt rangsorolva, a két rangszámösszeg közel azonos értéke a nullhipotézis beigazolását jelenti.

Wilcoxon előjeles rangpróba: két, összetartozó minta vizsgálata során alkalmazott előjelpróba, ha a nullhipotézis a két minta eloszlásának megegyezését feltételezi. Az egyszerű eljárást a gyors tájékozódásra használják a vizsgálat során. Az eljárás a két minta negatív és pozitív különbségeinek eloszlását vizsgálja. A nullhipotézis igazolása esetén a különbség eloszlás szimmetrikus.

Kruskal–Wallis-próba az eljárás 3 vagy több minta elemzésére alkalmas módszer. A vizsgálat feltételei: a mintavétel véletlen volta, a minták függetlensége és legalább ordinális változók megléte. Rangtranszformációs eljárásnak is nevezik, mivel a minták egyesítését követően a rangszámok meghatározását kell elvégezni, majd azokat az eredeti csoportok alapján csoportosítani. A transzformált értékek átlagrangjából vonható le a hipotézisre vonatkozó következtetés.”⁶

⁶ Tóthné Parázsó Lenke: A kutatómódszertan matematikai alapjai. E-learning tananyag. Eger, 2011, p. 55.

5.3 ÖSSZEFOGLALÁS, KÉRDÉSEK

5.3.1 Összefoglalás

A fejezetben a hipotézisvizsálatok alapfogalmait, a nullhipotézis, az alternatív hipotézis fogalmait ismerhettük meg, valamint a bizonyításukra, illetve elvetésükre szolgáló próbákat. Az önkontrollós kísérletek esetén erre az egy-mintás t-próba szolgál, míg a kontrollcsoportos kísérlet általánosítható eredményeit a kétmintás t-próbával bizonyíthatjuk. Ne feledjük, előbb szükséges az F-próba elvégzése, amely segít meghozni a döntést, hogy a kétmintás t-próba vagy a Welch-, illetve d-próba a megfelelő módszer. A többcsoportos kísérletek eredményeinek megbízhatóságát varianciaanalízissel bizonyíthatjuk, míg ha három vagy annál több mintánk is van, akkor a Kruskal–Wallis-próbát kell használni, ha teljesül a minták függetlensége.

5.3.2 Önellenőrző kérdések

Önellenőrzésként oldjuk meg az alábbi feladatot!



Egy könyvtár felmérte, hogy az olvasók milyen hatékonysággal találják meg a szükséges információt a könyvtárakban. A kísérletbe bevont személyeknek előbb katalóguscédula alkalmazásával kell megkeresniük az adatokat, majd újabb feladatot kellett elvégezniük elektronikus keresőrendszer használatával. A kísérlet azt vizsgálta, hatékonyabb-e az információkeresés elektronikus keresőrendszerek használatával? (Az adatok azt mutatják, hogy az adott feladattípust a bevont személyek hány százaléka oldotta meg sikerrel.)

12. Feladat egymintás t-próba alkalmazására

Művelet típusa	Katalóguscédula esetén	Elektronikus keresőrendszer esetén
Faktografikus tájékoztatás	80	82
Bibliográfiai ajánlás	91	91
Olvasótermi kötetek keresése	66	67
Kölcsönözhető művek keresése	73	55
Audiotermékek keresése	42	25
Videokazetták keresése	35	30
Multimédiák keresése	44	38
Folyóiratok keresése	42	37
Folyóiratcikkek keresése	51	43

A felmérés alapján van olyan művelet, amelyben hatékonyabbak voltak az olvasók a katalóguscédula használata során, van olyan rész, ahol nincs változás, és több területen figyelhető meg fejlődés. A kérdés az, hogy ez a felmérés elég megbízható-e ahhoz, hogy általánosítsunk?

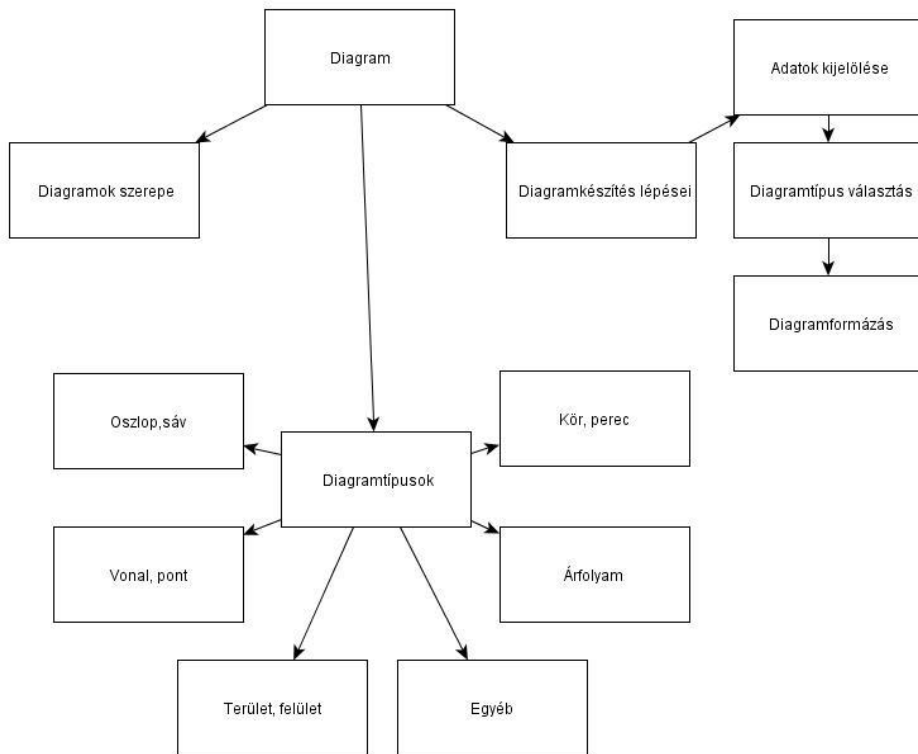
6. ÉRTÉKELÉSI EREDMÉNYEK SZEMLÉLTETÉSÉNEK LEHETŐSÉGEI A TÁBLÁZATKEZELŐ SZOFTVEREKBEN

6.1 CÉLKITŰZÉSEK ÉS KOMPETENCIÁK

A fejezetben az eredmények szemléltetésére szolgáló diagramokkal ismerkedhetünk meg. Természetesen a táblázatkezelők diagramjait mindenki ismeri, de mégis: megfelelő használatukhoz nem árt áttanulmányozni a fejezetet. Munkánk sikerét tudja befolyásolni egy-egy jól megválasztott diagram, melyekkel irányítani tudjuk a figyelmet. Gondoljunk csak arra, hogy egy vaskos pályázati anyag, egy éves jelentés, egy fenntartói elemzés stb. esetén az első dolog, amin a szakértők/elemzők átfutnak, azok a diagramok. Ha jól tudjuk használni a diagramkészítés apró lehetőségeit, akkor könnyedén tudunk olyan anyagokat összeállítani, mellyel segítünk a sorok között olvasni, és felhívni a figyelmet az általunk legfontosabbnak tartott adatokra.

6.2 TANANYAG

1. A diagramok szerepe
2. Diagramtípusok
3. Diagramkészítés a gyakorlatban
4. Speciális feladatok standard diagramtípusai



57. ábra: Fogalomtérkép

6.2.1 A diagramok szerepe

A grafikus ábrázolással szemléletesebbé tehetjük kinyert adatainkat, segíthetjük a következtetések levonását. Ezen funkció eléréséhez viszont lényeges, hogy milyen diagramtípust alkalmazunk.

A diagramoknál nem az a lényeg, hogy színesek legyenek, hanem hogy irányítsák a figyelmet, ezért a diagram megválasztásával lehetőséget kapunk, hogy arra hívjuk fel a figyelmet, ami az elemzés szempontjából fontos.

Nézzük meg, hogy mely diagramtípusokat mikor érdemes használni!

6.2.2 Diagramtípusok

Oszlopdiagram (Bar)

Az oszlopdiagram alapvetően a mennyiség szemléltetésére szolgál. Például amikor a hangsúlyt arra szeretnénk helyezni, hogy például iskolánk tanulói lét-

száma elérte az 1000 főt, a benzin árfolyama átlépte a 400 forintot, az ország kukoricaexportja 10.000 tonna fölött volt stb.

Illetve a mennyiség olyan jellegének hangsúlyozására, ami hatalmas értékű emelkedést vagy esést mutat.

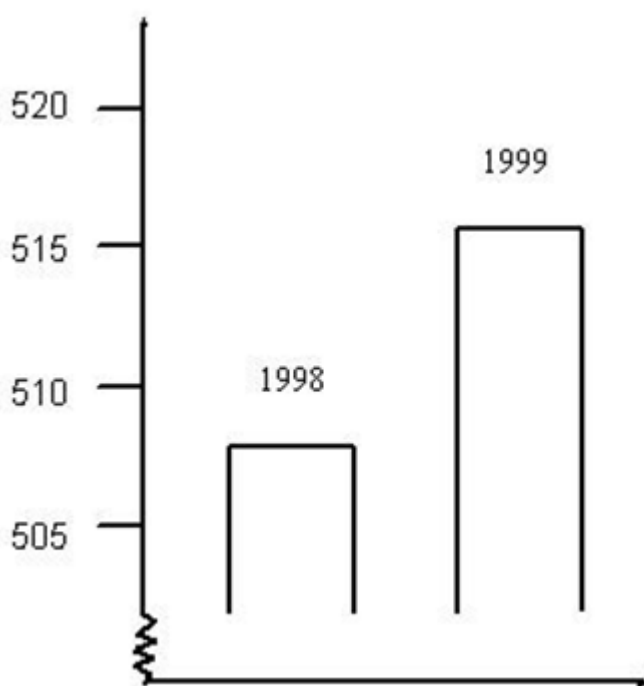
Természetesen a többi diagramról is leolvasható a mennyiség, de a lélektani hatás ezen érzékeltethető a legjobban.

A médiában nagyon jól használják ezeket a lélektani hatásokat, ezért már az iskolásoknak is felhívják a figyelmét a diagramokkal elérhető hatásokra. Például a PISA-mérésben találkozunk ilyen feladattal:



Egy tévériporter az alábbi diagramot mutatva a következőket mondta: „A diagram szerint a betörések száma óriásit nőtt 1999-ben 1998-hoz képest.”

Majd azt a kérdést kapták, hogy „Mit gondolsz, helyesen értelmezte a riporter a diagramot? Válaszodat indokold is meg!”



58. ábra: PISA-feladat

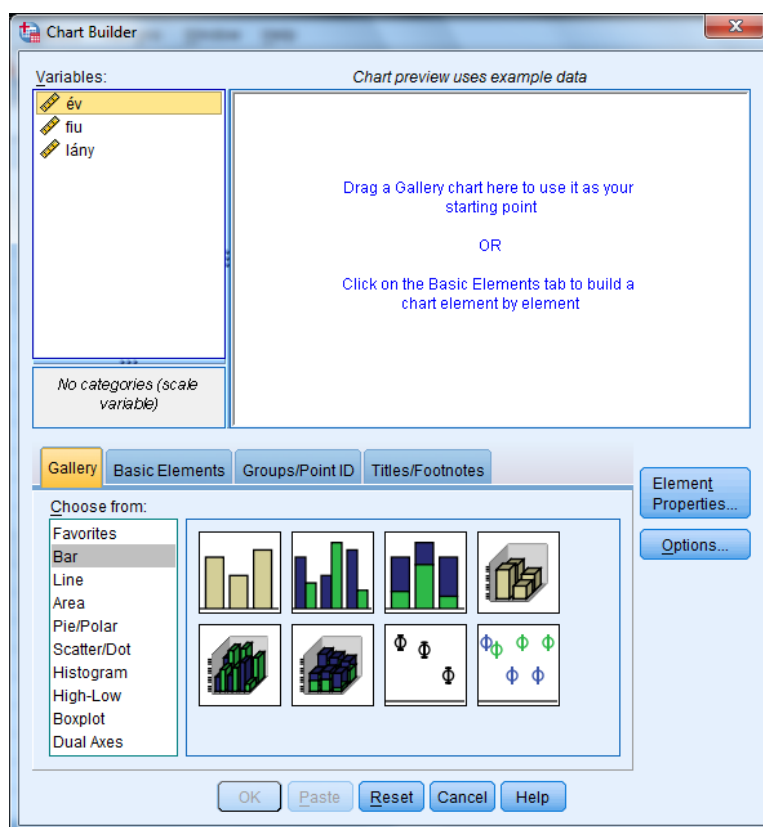


Érdeemes megnézni, mire adtak pontot az elemzés során: „Értékelés: 2 pont – „Nem, nem értelmezte helyesen”, Rámutat, hogy a diagramnak csak egy kis része látható, VAGY hogy az arányos, ill. százalékos növekedés nem nagy, VAGY hogy tendenciára vonatkozó adatokra volna szükség.

1 pont – „Nem, nem értelmezte helyesen”, de hiányoznak a magyarázat részletei. (Pl. a betörések száma közti különbséggel foglalkozik, és nem veti ezt össze a betörések teljes számával)⁷

Az oszlopdiagram altípusait is érdemes átnézni, mert szintén teljesen más hatásokat érhetünk el vele!

A GRAPHS menüpont CHART BUILDER parancsával tudunk diagramot készíteni.



59. ábra: Oszlopdiagram

⁷ PISA-mérés – Nyílt feladatok. < <http://www.oecd-pisa.hu/> >

Az első csoportosított oszlopdiagram használata esetén a hangsúly a kategóriákon van.



Például: 2011-ben felvettünk 400 fiút és 600 lányt intézményünkben, 2012-ben 500 fiút és 500 lányt. Itt az elemzésnél azon kell elgondolkodni, mi változott, hogy a fiúk száma emelkedett, és miért csökkent a lányok száma.

A halmozott oszlop esetén az összesített adatokon van a hangsúly, az pedig egy pluszinformáció (de mellékesebb), hogy milyen a csoportok összetétele.

Az előző példát halmozott diagrammal szemléltetve azt látjuk, hogy a felvett diákok száma változatlan. Nincs mit elemezni! Bár az összetételben meg fog jelenni a fiúk/lányok változása, ami még két adat esetén látható is, de ha több alcsoportunk van, és az „összesen”-ek is változnak, akkor nem szemléletes az alcsoportok összetétele. De vannak esetek, amikor nem is kell annak lennie!

Vonaldiagram (Line)

A vonaldiagram a folyamatok szemléltetésére a legmegfelelőbb.

Jól használható demográfiai folyamatok szemléltetésére, benzin árfolyam-változásának bemutatására stb. (Vonaldiagrammal azt tudjuk szemléltetni, hogy hónapok óta emelkedik az árfolyam, a hatás a folyamaton van, annak ellenére, hogy itt is látható, a benzinár átlépte pl. a 400 forintot).



Például ha a fenntartó felé hangsúlyozni szeretnénk, hogy folyamatosan emelkedik a gyermeklétszám intézményünkben (ezért szükség lenne pl. bővítésre), akkor használjuk a vonaldiagramot.

Kördiagram (Pie/Polar)

A kördiagram az egész megoszlásának szemléltetésére szolgál.

Fontos viszont, hogy a teljes egészről álljon rendelkezésre információ, vagy legalábbis vegyünk fel olyan kategóriát, hogy „nem válaszolt”, „nincs adat”.

A kördiagram képes arra, hogy százalékszámítás nélkül százalékosan ábrázolja az adatok megoszlását.

Területdiagram (Area)

Statisztikai vizsgálatok során a terület-/felületdiagramok használata

Pontdiagram (Scatter/Dot)

A vonaldiagrammhoz hasonló esetekben használjuk, ha csak a pontok kerülnek ábrázolásra, nem annyira szemléletes, viszont jól használható egyedi eredmények/esetek szemléltetésére. (Pl. Kompetenciamérés egyéni tanulói teljesítmények szemléltetésére)

Sugárdiagram

Sugárdiagram esetén mindegyik kategóriához tartozik egy-egy külön érték-tengely. Ezek a tengelyek egy közös pontból indulnak ki, és az ábrázolásra kerülő adatokat egy-egy vonal köti össze.

Például partneri elégedettségmérés szemléltetésére szokták használni a minőségbiztosítás területén.

Összefoglalva: a diagramtípusok alkalmazási lehetőségeit érdemes ismerni, és ennek tükrében tudatosan válasszunk a rendelkezésre álló diagramok közül! Ha elrontottuk a diagram típusát, lehetőségünk van az utólagos módosítása is a diagramtípus megváltoztatásával.

6.3 ÖSSZEFOGLALÁS, KÉRDÉSEK

6.3.1 Összefoglalás

A fejezetben megismerhettük, milyen jellegű adatoknál melyik a legmegfelelőbb diagramtípus, illetve altípusaik hogyan segítenek a diagram mondani-
valójának megváltoztatásában.

Érdemes a gyakorlatban is tesztelni az eltérő típusok okozta értelmezési változásokat, és kitapasztalni, mely esetekben mivel tudjuk leginkább alátámasztani mondanivalónkat.

6.3.2 Önellenőrző kérdések

1. Milyen diagramtípust használna, ha célja a folyamatok szemléltetése?
2. Mely esetben használhatunk kördiagramot?

7. NEMPARAMÉTERES ELJÁRÁSOK

7.1 CÉLKITŰZÉSEK ÉS KOMPETENCIÁK

Ez a fejezet a **nemparaméteres** nominális- és ordinális változók statisztikai elemzésének lehetőségeivel foglalkozik. Az előző fejezetek taglalták azokat az eljárásokat, melyekben a változók középértékét elemezve vonható le a következtetés: egymintás-, páros-, független mintás t-próba, egy- és többszemponos varianciaanalízis, melyek mindegyike a paraméteres eljárások közé sorolhatók.

A nemparaméteres eljárásokban a változó folytonos vagy finom beosztású, melyben a mediánnak kiemelt szerepe van. A fentiek következtében a nemparaméteres eljárásokban nem tényleges értékekkel történik az elemzés, hanem az értékek sorrendjével, melyeket **“rang”**-nak nevez a szakirodalom. Az érték rangja azt jelenti, hogy a nagysága alapján hányadik a mintában. A fejezetben az alábbi próbák értelmezésére és gyakorlati feladatokon keresztül kerülnek bemutatásra az SPSS-ben történő elemzés során.

- Kolmogorov-Szmirnov
- Wilcoxon-próba
- Mann–Whitney-próba
- Kruskal–Wallis-próba

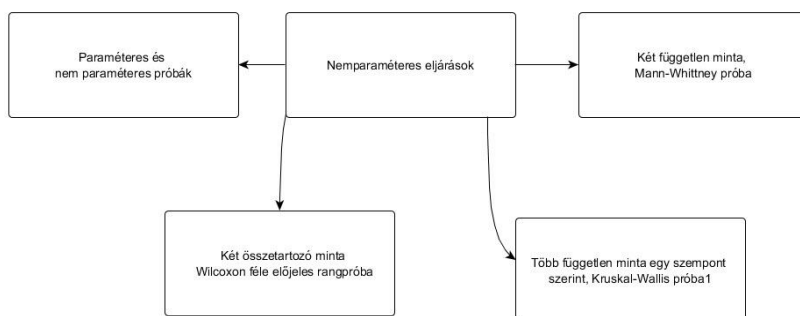
7.2 TANANYAG

Paraméteres és nem paraméteres próbák

Két független minta Mann-Whittney próba

Két összetartozó minta Wilcoxon-féle előjeles rangpróba

Több független minta egy szempont szerint Kruskal-Wallispróba



60. ábra: Fogalomtérkép

7.2.1 Paraméteres és nem paraméteres próbák

A **paraméteres** módszerek feltétele, hogy az adatok sokasága normális eloszlású, mely az átlagaik és a szórásaik összehasonlításával jellemezhető. Abban az esetben, ha a szórások azonosak az összehasonlítás az átlagok alapján (nullhipotézis) történik.

Nemparaméteres módszerek, előjelpróba, a változó adatsora folytonos vagy elég finom beosztású, azaz közel folytonos, szórása véges. Az adatpárok különbségének eloszlására nem jellemző az eloszlás normalitás, melynek ellenőrzése vagy nem valósítható meg, vagy ellenőrzésének nincs értelme. Ebben az esetben lehet a minta elemszáma kicsi, az adatok nominálisak valamint a mediánnak van fontos szerepe. A nemparaméteres eljárások során nem a tényleges értékekkel számolunk, hanem a rangokkal, melynek jelentése a nagyság szerinti helyzete a mintában.

13. Táblázat: Paraméteres és a nemparaméteres eljárás párhuzamba állítása

	Paraméteres eljárások	Nemparaméteres eljárások
Egy minta	egymintás t-próba	Kolmogorov-Szmirnov
Két független minta	független mintás (kétmintás t-próba)	Mann-Whitney Kolmogorov-Szmirnov
Két összetartozó minta	páros t-próba	Wilcoxon
Több független minta egy szempont szerint	Egyszempontos varianciaanalízis	Kruskal-Wallis
Több összetartozó minta egy szempont szerint	Egyszempontos varianciaanalízis ismételt mérés	Friedman

7.2.2 Kolmogorov-Szmirnov Test

A minta eloszlásának tesztelése (normál, Poisson, egyenletes exponenciális) tesztelése. A művelet elvégzésének feltétele, hogy a változó normál eloszlású legyen. Összehasonlítja a megfigyelt kumulatív eloszlást a teoretikus kumulatív eloszlás függvényével és a közöttük lévő legnagyobb abszolút különbségből számítja a Kolmogorov-Szmirnov Z értéket, melyet a megfigyelések négyzetgyökével szorozva képezi.

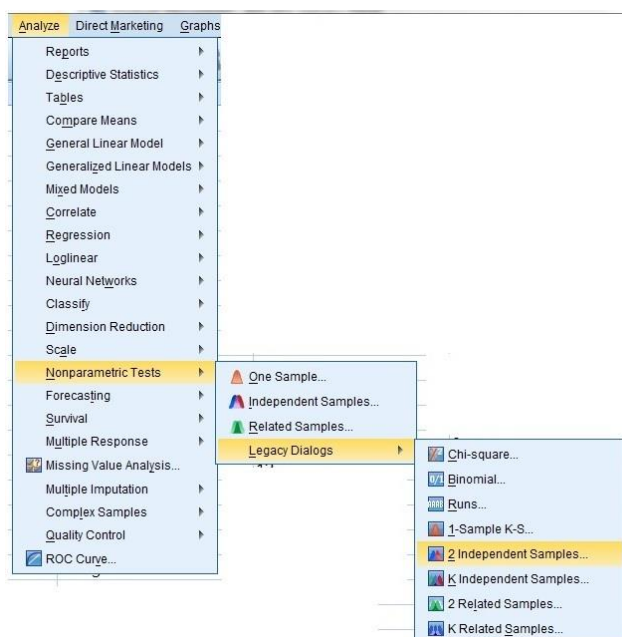
- ✿ A testőmeg index a mozgással töltött idő kapcsolatát vizsgáljuk meg a nemek függvényében. Feltételezzük:

H_0 A két mért változó eloszlása megegyezik a férfiak és a nők esetében

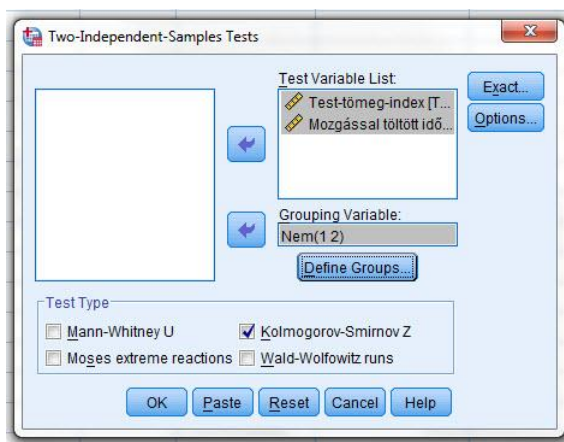
H_1 A két eloszlás nem egyezik meg

Az eljárás az SPSS-sel az alábbi parancsokkal kiválasztásával és lefuttatásával történik:

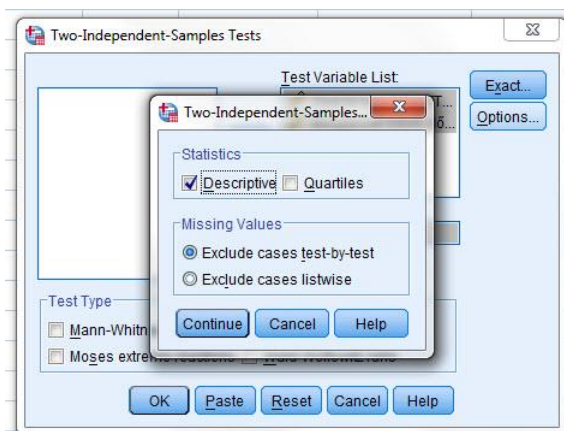
**Analyze/Nonparametric Test/Legacy Dialogs/2 Independent Samples/
Test Variable Test-tömeg index és a Mozgással töltött idő
Grouping Variable Nem**



61. ábra: Kolmogorov-Szmirnov lehulló menü



62. ábra: Kolmogorov-Szmirnov lehulló menü beállítása



63. ábra: Lehulló párbeszédablak beállítása

A Test Variable ablakba behúzzuk a két nemparaméteres változót és kipi-páljuk a Descriptive rádiógombot.

Test Statistics ^a			
		Test-tömeg-index	Mozgással töltött idő (perc)
Most Extreme Differences	Absolute	,400	,600
	Positive	,400	,067
	Negative	,000	-,600
Kolmogorov-Smirnov Z		,775	1,162
Asymp. Sig. (2-tailed)		,586	,134

a. Grouping Variable: Nem

64. ábra: Kolmogorov-Szmirnov lehulló menü

Test Statistics^a

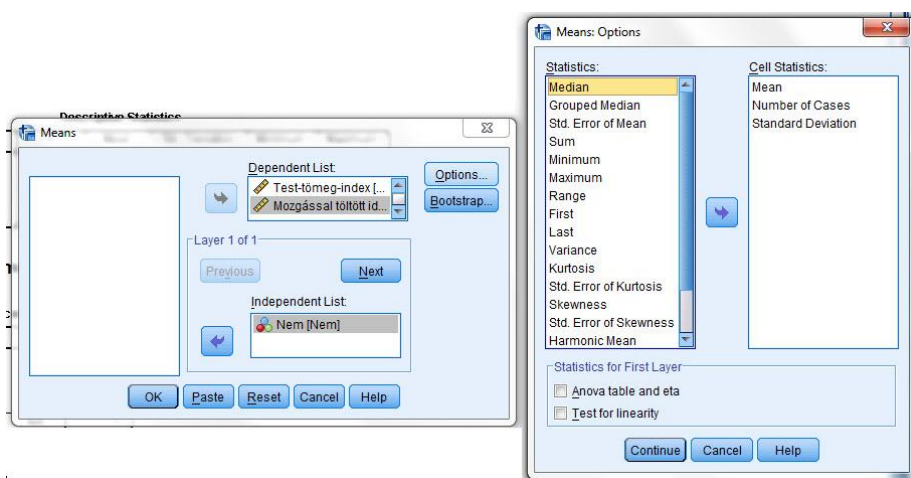
		Test-tömeg-index	Mozgással töltött idő (perc)
Most Extreme Differences	Absolute	,400	,600
	Positive	,400	,067
	Negative	,000	-,600
Kolmogorov-Smirnov Z		,775	1,162
Asymp. Sig. (2-tailed)		,586	,134

a. Grouping Variable: Nem

65. ábra: A Kolmogorov-Szmirnov próba output felülete

Az Asymp.Sign (2 tailed) szignifikancia szint $p=0,586$ alacsony volta arra mutat, hogy a testtömeg index nem változott jelentősen, de a mozgással töltött idő szignifikancia szintje magasabb, ezért a nullhipotézis nem vethető el. Az eredmény tehát nem szignifikáns $p=0,134$.

Vizsgáljuk meg az átlagok mellett a szórásokat mindkét csoportban. Kattintsunk az **Analyze/Compare Means/Means** parancsra és a lehulló ablakba húzzuk be Dependent List Test-tömeg index és a Mozgással töltött idő, az Independent Nem adatokat. Az Options menüre kattintva a Cell Statistics ablakba ellenőrizzük a Mean, Number of Cases és a Standard Deviation beírását. Majd az OK-ra kattintva el kell indítani a műveletet.



66. ábra: A Compare Means Option ablaka

Az alábbi táblázat alapján elemezhető a kapott eredmény. Megállapítható, hogy az átlagok mellett a szórások is lényegesen különböznek.

Case Processing Summary						
	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Test-tömeg-index * Nem	20	100,0%	0	,0%	20	100,0%
Mozgással töltött idő (perc) * Nem	20	100,0%	0	,0%	20	100,0%

Report			
Nem		Test-tömeg-index	Mozgással töltött idő (perc)
férfi	Mean	20,5287	76,33
	N	15	15
	Std. Deviation	4,91891	37,248
nő	Mean	23,0500	51,00
	N	5	5
	Std. Deviation	7,43587	13,416
Total	Mean	21,1590	70,00
	N	20	20
	Std. Deviation	5,54288	34,451

67. ábra: Compare Means output ablaka

7.2.3 Két független minta Mann-Whitney próba

Két független csoport, ahol az egyik változó ordinális és eloszlásuk eltér a normálistól. Ordinális és skála típusú adatoknál alkalmazható, ahol nem feltétel a normáloszlás. A kétmintás t-próba nemparametrikus alternatívája, ahol nem teljesül a kétmintás t-próba előfeltétele a szórás egyezés és/vagy a normális eloszlás. Alkalmazásának feltételei az eloszlások a mintában hasonló alakúak és függetlenek. Elsősorban kontrollcsoportos kísérletek esetén alkalmazzuk, hogy összehasonlítsuk a két csoport által elért eredmények közötti különbséget és választ kapjunk, hogy szignifikáns-e.

$H_0: M(x)=M(y)$, az alternatív hipotézis $M(x) \neq M(y)$.

14. Példa a Mann-Whitney alkalmazására

- ✿ A normál normál és gyógytestneveléses foglalkozásokon résztvevő tanulók intenzív foglalkozás eredményességét mérve az alábbi tevékenységet végezték a gyerekek:

- Cooper-teszt (12 perces futás)
- Fekvőtámaszban karhajlítás és nyújtás (max. 4 perc vagy maximális kifáradás)

- Hasonfekvésben végzett törzsemelés (max. 4 perc vagy maximális kifáradás)

	NEME	ÁLLAPOT	TESTSÚLY	COOPER	TÖRZSEMELES	FEKVŐTÁMASZ
1	1	2	23	1080	104	21
2	1	1	33	1905	141	37
3	1	1	28	1365	100	8
4	1	2	33	1352	110	32
5	1	1	20	1760	113	41
6	2	1	30	1202	79	9
7	1	1	31	1880	102	43
8	2	1	25	1435	100	35
9	2	1	29	1260	110	14
10	1	1	28	1260	108	25
11	1	1	36	1620	150	26
12	1	2	24	1270	91	22
13	1	1	24	1680	140	51
14	2	1	36	1620	101	32
15	2	1	18	1140	98	31
16	2	1	28	1321	100	42
17	1	1	27	1770	148	49
18	2	2	46	1122	85	17
19	1	2	28	1620	96	35
20	1	2	23	1440	79	36
21	1	2	31	1452	109	21

	NEME	ÁLLAPOT	TESTSÚLY	COOPER	TÖRZSEMELES	FEKVŐTÁMASZ
1	1	2	25	1295	125	30
2	1	1	36	2100	170	47
3	1	1	31	1820	141	19
4	1	2	36	1490	140	33
5	1	1	23	2340	144	57
6	2	1	33	1620	150	15
7	1	1	33	2260	160	38
8	2	1	28	1620	140	36
9	2	1	32	1520	165	33
10	1	1	30	1480	123	24
11	1	1	39	2010	152	32
12	1	2	25	1950	92	22
13	1	1	25	2070	144	42
14	2	1	37	1800	160	40
15	2	1	19	1950	150	33
16	2	1	30	1800	160	50
17	1	1	28	2070	170	46
18	2	2	47	1440	149	18
19	1	2	29	1800	130	38
20	1	2	25	1980	108	37
21	1	2	32	1890	150	40

68. ábra: A normál és a gyógytestnevelés őszi és tavaszi adatai

Feltételezés:

H_0 : A feltételezett és a kiinduló állapot közötti teljesítmény különbség értéke „0”.

H_1 : A feltételezett és a kiinduló állapot közötti teljesítmény különbség értéke nem „0”.

A fenti két táblázat adataiból kiindulva a vizsgálat azt célozza, hogy milyen teljesítménybeli eltérések vannak az egészséges és elváltozásban szenvedő tanulók között. A további vizsgálat azt célozza, hogy a két félétet összehasonlítva a teljesítményük mennyit javult a torna és a speciális gyógytestnevelés hatására.

Mindhárom esetünkben 5% ($p=0,05$) határozzuk meg, akkor száz esetből ötször fordul elő, hogy a próbafüggvény minta alapján kiszámított értéke a kritikus tartományba esik. Amennyiben az **Asymp. Sig. értéke magasabb, mint 0,05**, akkor nullhipotézist megtartható.

Annak bizonyítása, hogy a tanulók sportteljesítményét befolyásolja az egészségi állapotuk, a Mann-Whitney U próbát alkalmazva az alábbi megállapítás tehető.

A művelet elvégzése az az SPSS-sel az alábbi parancsokkal kiválasztásával és lefuttatásával történik:

Analyze/Nonparametric Test/Legacy Dialogs/2 Independend Samples/Mann Whitney U

Test Variable mezőbe be kell húzni gyakorlatot (cooper, törzsemelés, fekvőtámasz).

Gruping Variable (csoport változó) az egészségi állapotot

Eredmény az output ablakban értelmezhető:

Mann-Whitney

Ranks				
	ALLAPOT	N	Mean Rank	Sum of Ranks
COOPER	EGÉSZSÉGES	14	12,36	173,00
	BETEG	7	8,29	58,00
	Total	21		
TÖRZSEMELEÉS	EGÉSZSÉGES	14	12,57	176,00
	BETEG	7	7,86	55,00
	Total	21		
FEKVŐTÁMASZ	EGÉSZSÉGES	14	12,07	169,00
	BETEG	7	8,86	62,00
	Total	21		

Test Statistics ^b			
	COOPER	TÖRZSEMELEÉS	FEKVŐTÁMASZ
Mann-Whitney U	30,000	27,000	34,000
Wilcoxon W	58,000	55,000	62,000
Z	-1,420	-1,645	-1,120
Asymp. Sig. (2-tailed)	,156	,100	,263
Exact Sig. [2*(1-tailed Sig.)]	,172 ^a	,110 ^a	,287 ^a

a. Not corrected for ties.

b. Grouping Variable: ALLAPOT

69. ábra: Output eredmények

A próba eredménye, is mutatja az asymp. sig. értéke mindhárom esetben nagyobb, mint a megállapított szignifikációs szintünk, ezért nullhipotézisünket megtartjuk, vagyis a feltételezett és a kiinduló állapot közötti teljesítmény különbség értéke „0”.

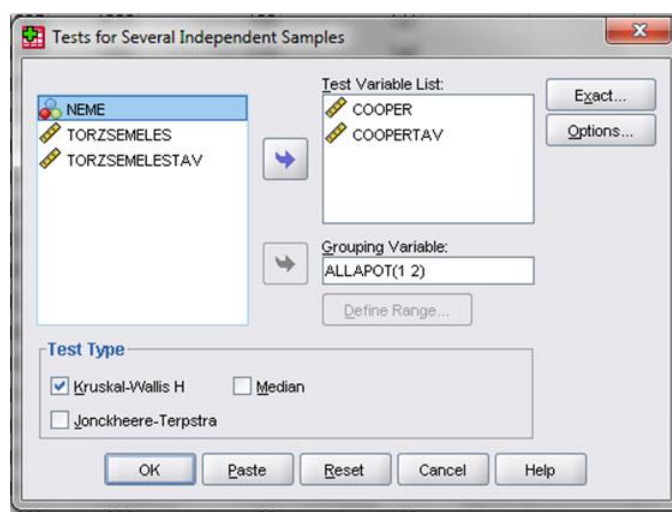
A „Ranks” táblázat pedig tételeken igazolja, hogy az egészséges tanulók magasabb teljesítményre képesek, mint a betegek, tehát egészségi állapotuk befolyásolja a teljesítőképességüket.

További vizsgálatokat folytatva annak feltárására, hogy milyen hatással van a tanulóra az intenzív és speciális testnevelés.

H₂ Az intenzív és speciális gyógytestnevelés növeli a tanulók sportteljesítményét

- egészséges gyerekek
- beteg gyerekek

Az őszi és tavaszi félév eredményeit összesítő táblázatba helyezve véghezük az elemzést tovább a **Kruskal-Wallis H** próbával.



70. ábra: Mann-Whitney H beállítása

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
COOPER	21	1458,29	247,359	1080	1905
COOPERTAV	21	1824,52	280,749	1295	2340
ALLAPOT	21	1,33	0,483	1	2

Kruskal-Wallis

Ranks

	ALLAPOT	N	Mean Rank
COOPER	EGÉSZSÉGES	14	12,36
	BETEG	7	8,29
	Total	21	
COOPERTAV	EGÉSZSÉGES	14	12,50
	BETEG	7	8,00
	Total	21	

Test Statistics^{a,b}

	COOPER	COOPERTAV
Chi-Square	2,016	2,467
df	1	1
Asymp. Sig.	,156	,116

a. Kruskal Wallis Test

b. Grouping Variable: ALLAPOT

71. ábra: Kruskal Wallis H eredmény az outputban

A statisztikai eredményeket alapján a fenti táblázatból kiolvassa az Asymp.Sig értékét (0,116), megállapítható, hogy nagyobb, mint a mérés tervezésekor meghatározott $p=0,05$. Ennek alapján a nullhipotézis nem vethető el. A Kruskal-Wallis „Ranks” táblázata alapján megállapítható, hogy az egészséges tanulók teljesítményében javulás következett be. A beteg tanulók esetében az eredményesség elmarad az egészségesekéhez viszonyítva.

7.2.4 Két összetartozó minta Wilcoxon-féle előjeles rangpróba

Az egymintás t-próba nemparametrikus alternatívája. Az eljárás tipikus módszere a Wilcoxon-féle előjeles rangpróba, mely az előjelek mellett a különbségek közötti nagyságrendeket is figyelembe veszi, ezáltal nagyobb erejű rangpróba. A vizsgálat feltétele, hogy a különbség eloszlás szimmetrikus legyen. Végrehajtása a következő lépésekből áll: a mintaelemek közötti különbségek rangsorolása, az előjelektől függetlenül, a nullák kimaradhatnak. A rangsorolás menete a következő lépésekből áll.

az adatsor nagyságszerinti sorba rendezése (első 1, második 2....-utosó n), egyenlő számok esetén a z adott rangszámok átlagával számolunk, ez a korrigált rangszám. A rangsorolás helyességének ellenőrzése

kapott rangszám összege = $\frac{n \times (n+1)}{2}$ ahol n az első "n" egész szám összege

Összegezve a rangszámokat, a pozitív és a negatív különbségekhez tartozó különbségekhez tartozó rangszámösszegek azonosak vagy közel azonosak abban az esetben két sokaság azonos eloszlású tehát **igaz a nullhipotézis**.

Minél nagyobb a két sokaság között, annál nagyobb a különbség a rangszámok között.

A normális eloszlás során a szignifikancia:

$$z = \frac{\sum_{i=1}^n R_i}{\sqrt{\sum_{i=1}^n R_i^2}}$$

ahol: számlálóban az összes előjeles rangszám összege, a nevezőben pedig az összes előjeles rangszám összegének négyzetgyöke szerepel. Ez az érték a programok által kapott „p” értékei.

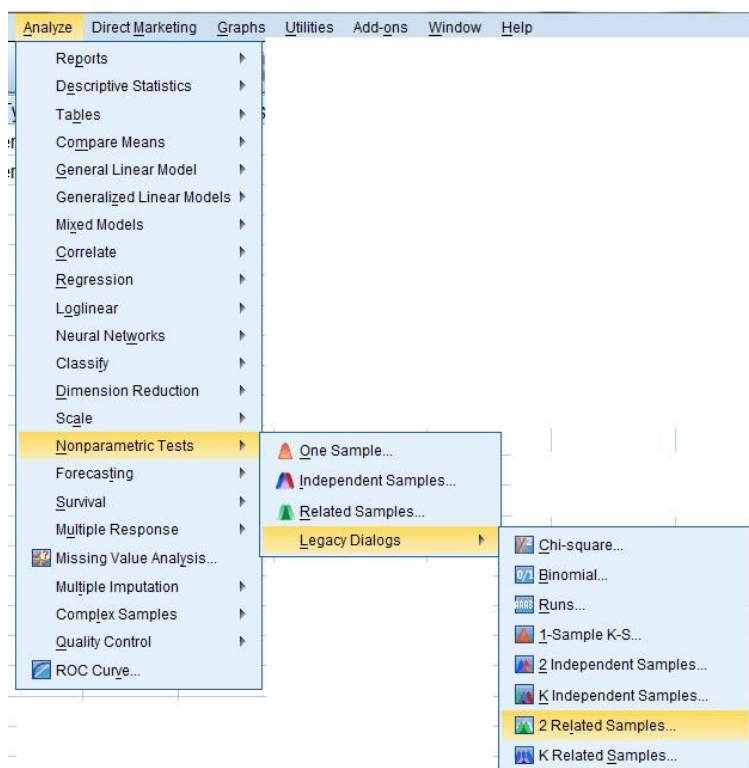
- ✿ Az 5. osztályban a tanulókat matematikai teljesítményük alapján év elején rangsorolták, majd év végén.

H_0 az évelejei és év végi eredmények azonos eloszlásúak

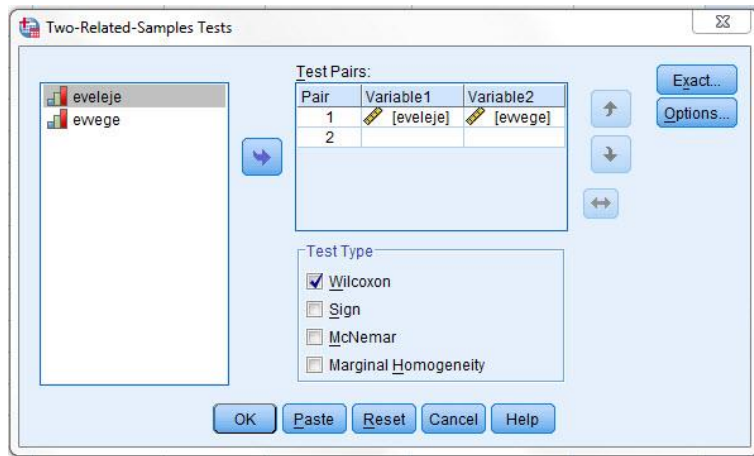
H_1 Az év végi eredmények eloszlása nagyobb

A művelet elvégzése az az SPSS-sel az alábbi parancsokkal kiválasztásával és lefuttatásával történik:

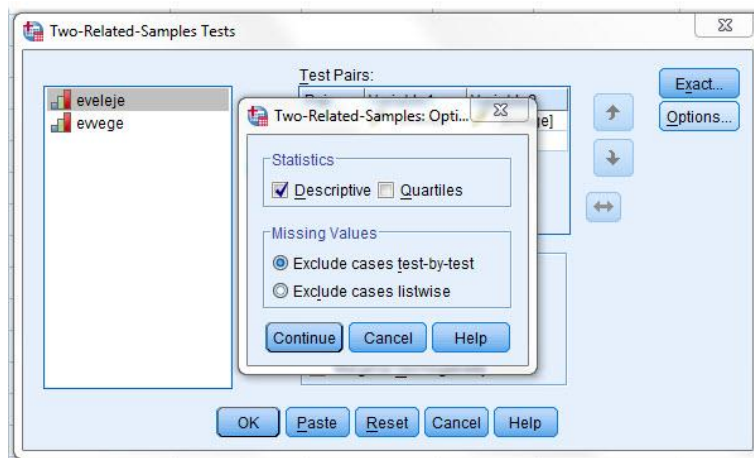
Analyze/Nonparametric Test/Two Related Samples
Test/Options → *Descriptives*



72. ábra: Wilcoxon lehulló menü



73. ábra: Wilcoxon párbeszédpanl beállítása



74. ábra: Wilcoxon Rank Options ablak beállítása

Az alábbi output táblázat eredményei alapján megállapítható, hogy az év eleji és év végi átlaga és a szórások azonosak, azaz -0,070, azaz az év elejei teljesítmény magasabb volt, bár közel azonosak az értékek. Az eredmény nem szignifikáns, a nullhippótezis nem vethető el (az évelejei és év végi eredmények azonosak).

NPar Tests

[DataSet1] H:\SAJAT_fontos\konyv_LMT\TPL\Wilcoxon.sav

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
eveleje	10	5,50	3,028	1	10
ewege	10	5,50	3,028	1	10

Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
ewege - eveleje	Negative Ranks	4 ^a	4,63	18,50
	Positive Ranks	4 ^b	4,38	17,50
	Ties	2 ^c		
	Total	10		

a. ewege < eveleje

b. ewege > eveleje

c. ewege = eveleje

Test Statistics^b

	ewege - eveleje
Z	-,070 ^a
Asymp. Sig. (2-tailed)	,944

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test

75. ábra: Wilcoxon Rank test eredménytáblázata

7.2.5 Több független minta egy szempont szerint Kruskal-Wallis próba

A Kruskal-Wallis próba független csoportok mediánját hasonlítja, melyek közül legalább az egyik változó ordinális. A próbát rangsorolt adatokkal végezzük.

- ✿ Három tanulócsoporthoz tartozó kollokvium jegyeit hasonlítjuk össze. A jegyek alapján rangsoroltuk az adatokat. Feladat a rangsorok átlagát. Vizsgáljuk meg van-e a részminták eredményei között szignifikáns különbség.

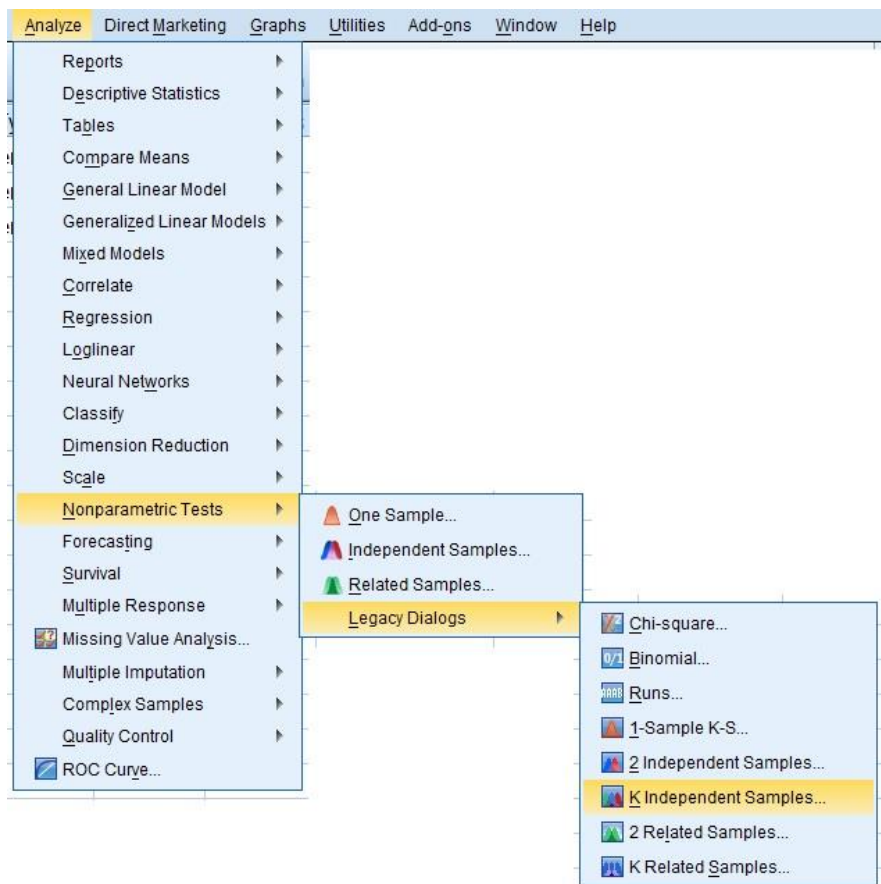
A művelet elvégzése az az SPSS-sel az alábbi parancsokkal kiválasztásával és lefuttatásával történik

Analyze/Nonparametric Test/K Independent/Test Samples

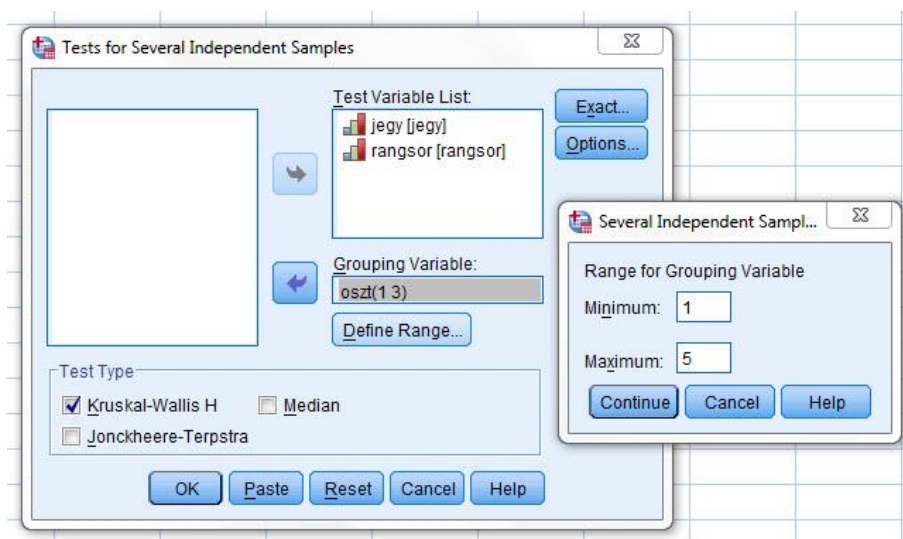
Test Variable List jegy, rangsor

Gruoping Variables 1, 5

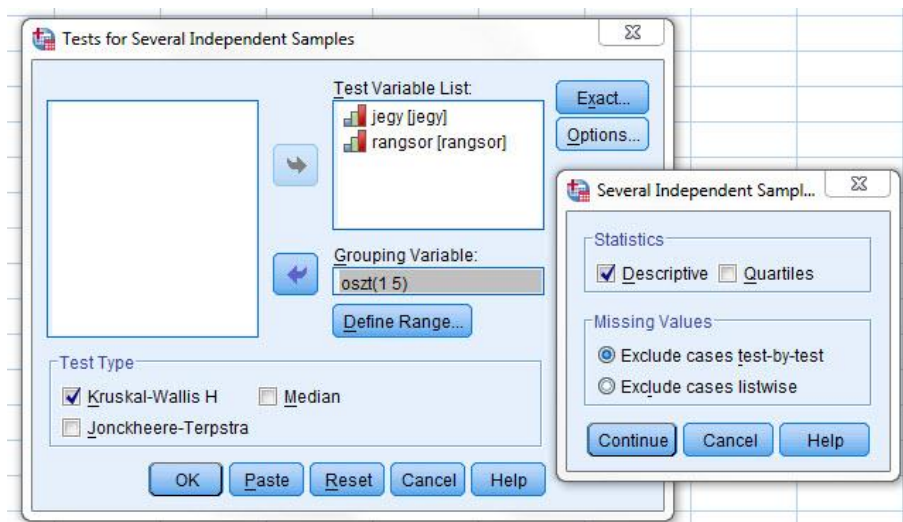
Options Descriptive



76. ábra: Kruskal-Wallis próba lehulló menü



77. ábra: Kruskal-Wallis próba lehulló ablak beállítása



78. ábra: Kruskal-Wallis próba lehulló Statistics menü

Az Output ablakban a táblázat alapján megállapítható, hogy nincs szignifikáns különbség a csoportok között. A három csoport kollokvium jegyei hasonlóak. A csoportok rangszámainak átlagait a középső táblázat alapján leolvasha-

tó, hogy az első csoporté a legmagasabb, vagyis összehasonlítva a rangsorban szereplő eredményeket a legjobb a harmadik csoport.

NPar Tests

[DataSet0] H:\SAJAT_fontos\konyv_LMT\TPL\Kr_W\KW.sav

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
jegy	50	3,36	1,191	1	5
rangsor	50	25,50	14,113	3	46
osztály	50	2,02	,869	1	3

Kruskal-Wallis Test

Ranks

	osztály	N	Mean Rank
jegy	1	18	29,83
	2	13	27,54
	3	19	20,00
	Total	50	
rangsor	1	18	29,28
	2	13	27,42
	3	19	20,61
	Total	50	

Test Statistics^{a,b}

	jegy	rangsor
Chi-Square	4,846	3,817
df	2	2
Asymp. Sig.	,089	,148

a. Kruskal Wallis Test

b. Grouping Variable: osztály

79. ábra: Kruskal-Wallis próba output ablaka

7.3 ÖSSZEFOGLALÁS, KÉRDÉSEK

7.3.1 Összefoglalás

Ebben a tananyagrészben az olvasó megismerkedhetett a nemparaméteres próba jellemzőivel. Bemutatásra került a nemparaméteres Kolmogorov-Szmirnov, Mann–Whitney-próba, Wilcoxon-próba, Kruskal–Wallis-próba alkalmazása. Az eljárások feltételeit mintákon keresztül és az SPSS-ben való analízis menetét és az eredmények értelmezését tanulmányozható.

7.3.2 Önellenőrző kérdések

1. Elemezze Kolmogorov-Szmirnov, Mann–Whitney-próba, Wilcoxon-próba, Kruskal–Wallis-próba a nemparaméteres próbák jellemzőit.
2. Ismertesse Kolmogorov-Szmirnov, Mann–Whitney-próba, Wilcoxon-próba, Kruskal–Wallis-próba a nemparaméteres próbák eljárás lényegét az SPSS-ben.

8. ANOVA

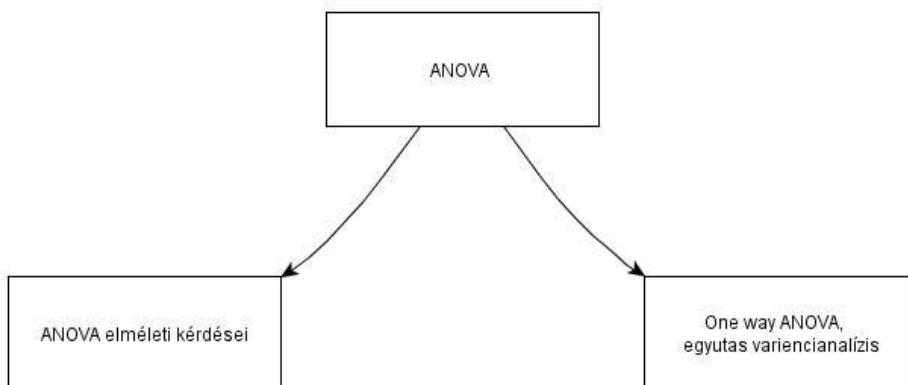
8.1 CÉLKITŰZÉSEK ÉS KOMPETENCIÁK

Az ANOVA több egyenlő szórású, normál eloszlású sokaság összevetésére alkalmas módszer, mely az Analysis Of Variance kezdőbetűiből kapta a nevét. A vizsgálandó sokaság legalább egy ismérv szerint részekre bontható. A vizsgálat során arra keresünk választ, hogy a mért ismérv szerinti csoportokban az összehasonlítható csoportok szignifikánsan különböznek-e egymástól.

8.2 TANANYAG

ANOVA elméleti kérdései

One way ANOVA egyutas variacionális



80. ábra: Fogalomtérkép

8.2.1 ANOVA elmélet kérdései

Az ANOVA-ban beszélhetünk **főátlagról**, mely az a minták egyben való ömlesztése során kapott adatsorátlag. A minták átlagai kiszámítva a mintaátlagok képezhetőek. Fisher⁸ a kutatásai során vizsgálta az adatok főátlagtól való eltéréseinek forrását és megállapította, hogy a minták átlaga és a minták közötti átlagok különbözősége az eltérések forrása.. Az általa kidolgozott eljárás során az átlagok helyett a szórásnégyzettel javasolta az elemzés elvégzését.

⁸ <http://irh.inf.unideb.hu/~jsztrik/education/11/valseg-matstat-elo.pdf>

Az eljárás során a minták saját belső variancia, a minták közötti variancia kerül meghatározásra.

H0: Az eljárás során a csoporton belüli variancia egyenlő a csoportok közötti varianciával, vagyis $F \cong 1$.

H1: a H0 tagadása, vagyis a csoporton belüli variancia nem egyenlő a csoportok közötti varianciával, vagyis $F > 1$.

Minél nagyobb az F értéke, annál erősebb bizonyítékot kapunk a H_0 ellen. A vizsgálat alapfeltétele, hogy az mérések egymástól függetlenek legyenek, eloszlásuk normál, a belső varianciák nem különböznek egymástól szignifikánsan.

Több információ képzésére a Post Host tesztet kell elindítani és az eredményeket értékelni.

A Levene teszt nullhipotézise azt mondja ki, hogy a szórások nem egyenlők, melynek következményeként a nullhipotézist elvetjük és a szóráshomogenitás teljesül. A szóráshomogenitás feltétele a Levene statisztika nem szignifikáns ($p < 0,05$).

A varianciaanalízis annak függvényében ahány ismerv hatását kell elemezni beszélhetünk egyutas/egyszempontos, kétszempontos/kétutas és több szempontos/többutas varianciaanalízisről.

A továbbiakban a egyutas/egyszempontosanalízis kerül bemutatásra.

8.2.2 One way ANOVA egyutas varianciaanalízis

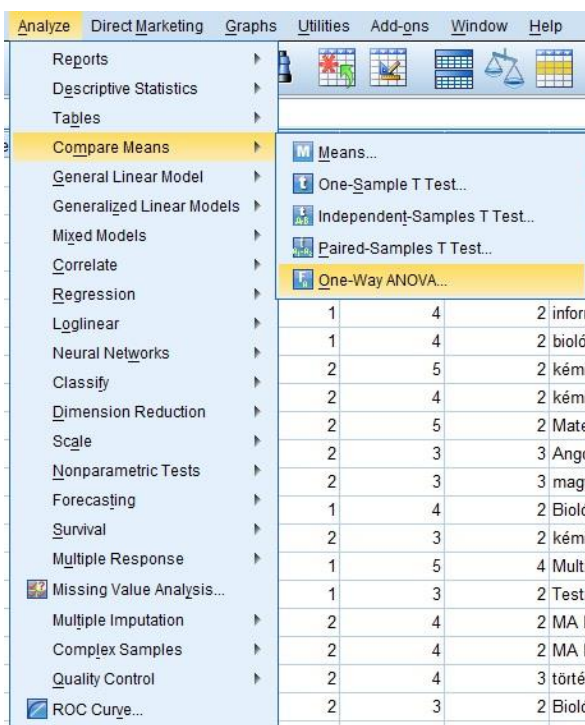
One way ANOVA analízist kettőnél több független minta összevetésére alkalmas eljárás.

✿ IKT eszközök használata a korosztály függvényében

A mérést **ANOVA elemzéssel** végezhetjük, melynek során a független változó az életkor és a függő változó az informatikai eszközök alkalmazása (K1 informatikai eszköz használata, K2 korosztály)

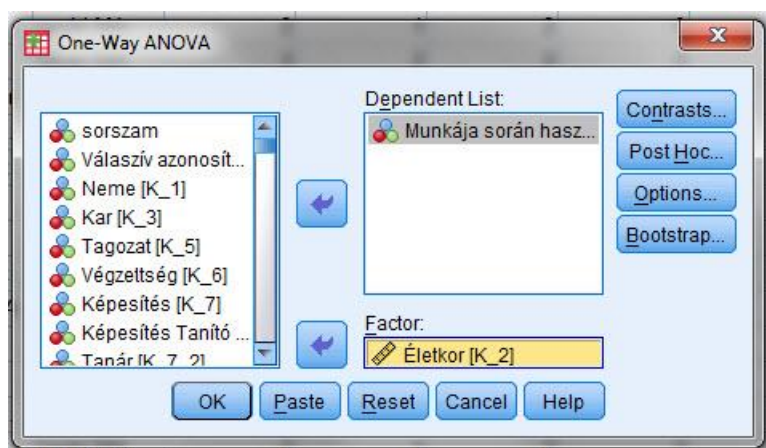
A varianciaanalízis rámutatott a varianciák megegyezésének kérdésére.

Analyze/Compare Means/One-Way ANOVA menüre kattintunk



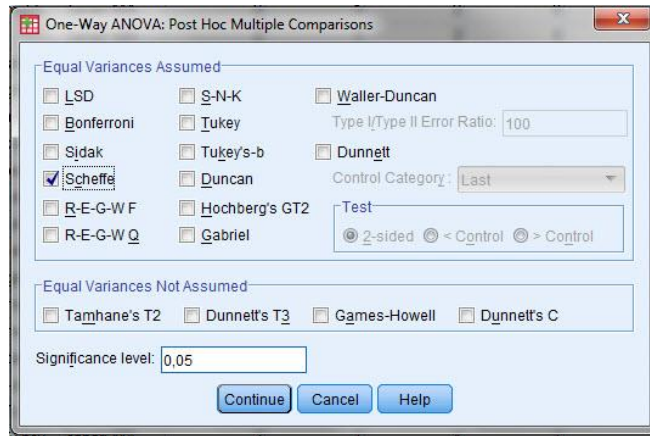
81. ábra: One Way ANOVA lehulló menü

Behúzzuk a függő változót (Dependent List), mely a munka során alkalmazott IKT eszköz, és a Faktort, ez az életkor.



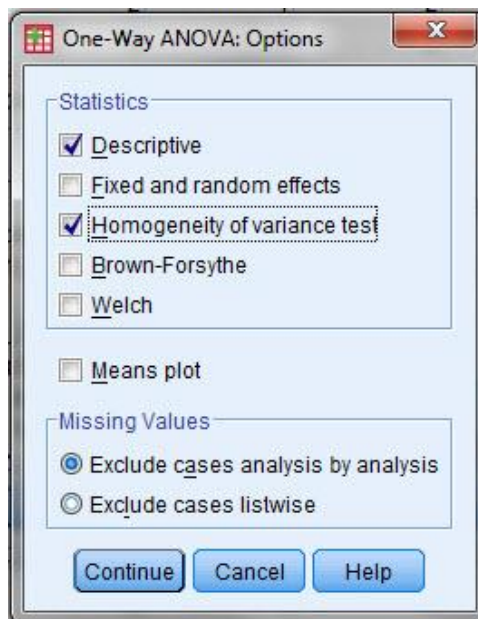
82. ábra: A független változó és a faktor kiválasztása

A **Post Hocra** kattintva a lehulló ablakban kipipáljuk a SCHEFFE-t. Majd a Continue nyomjuk meg.



83. ábra: Post Hoc Multiple Comparisons menüpont

Az Options rádiógomra kattintva a lehulló ablakba kipipáljuk a Descriptive és a Homogeneity of variance test rádiógombokat.



84. ábra: Descriptive és a Homogeneity of variance test rádiógombok kijelölése

A kapott válaszok (853 fő) alapján az eredmények kiértékelése:

Az alábbi táblázat alapján megállapítható, hogy a szóráshomogenitás feltétele teljesül, melynek bizonyítéka, hogy a **Levene** statisztika nem szignifikáns (0,228). A **Levene teszt** nullhipotézise azt mondja ki, hogy a szórások nem egyenlők, melynek következménye hogy a nullhipotézist elvetjük. de a szóráshomogenitás teljesül.

Test of Homogeneity of Variances

Munkája során használ-e informatikai eszközöket

Levene Statistic	df1	df2	Sig.
1,411	4	847	,228

A továbbiakban tekintsük át, hogy a csoportok közötti és a csoporton belüli négyzetösszegek vizsgálatával mit kapunk.

ANOVA

Munkája során használ-e informatikai eszközöket

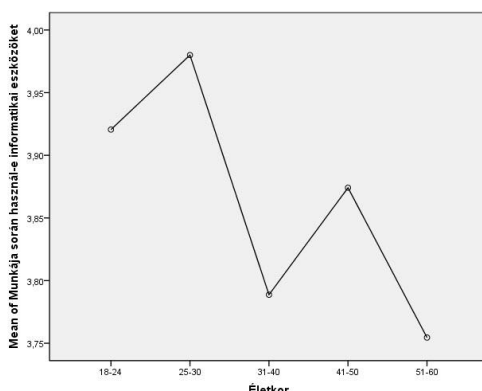
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4,350	4	1,088	,935	,443
Within Groups	985,466	847	1,163		
Total	989,816	851			

ANOVA táblázat

A táblázat alapján megállapítható, hogy a csoportok közötti (Between Groups) és a csoporton belüli (Within Groups) eltérésnégyzetek arány ($F=1,088/1,163=0,935$). Az F próbához tartozó szignifikanciaszint 0,443, vagyis nagyobb, mint 0,05.

Következtetésként megállapítható, hogy a kategóriaátlagok nem különböznek egymástól szignifikánsan, az informatikai eszközök használata közel azonos mértékben hat a különböző korosztályú hallgatókra.

Az alábbi görbe az átlagok vizuális képét szemlélteti. A legerősebb válaszreakció a 25-30 és a 41-50 éves korosztály esetében.



85. ábra: Internet eszközök használat átlagértékei

- ✿ Feladat
- ✿ Az otthoni IKT támogató tanulási feltételei értékelje (saját laptop, táblaszoftver elektronikus tananyaggal, internet)?
- ✿ PC használata

A mérést ANOVA elemzéssel végeztük, melynek során a független változó az életkor és a függő változó a PC alkalmazása. A varianciaanalízis rámutatott a varianciák megegyezésének kérdésére.

Az alábbi táblázat alapján megállapítható, hogy a szóráshomogenitás feltétele teljesül, melynek bizonyítéka, hogy a Levene statisztika nem szignifikáns (0,233). A Levene teszt nullhipotézise azt mondja ki, hogy a szórások nem egyenlők, melynek következménye hogy a nullhipotézist elvetjük és a szóráshomogenitás teljesül.

Test of Homogeneity of Variances

Milyenek tartja otthoni IKT támogatótanulási feltételeit? PC

Levene Statistic	df1	df2	Sig.
1,396	4	847	,233

ANOVA

Milyenek tartja otthoni IKT támogatótanulási feltételeit? PC

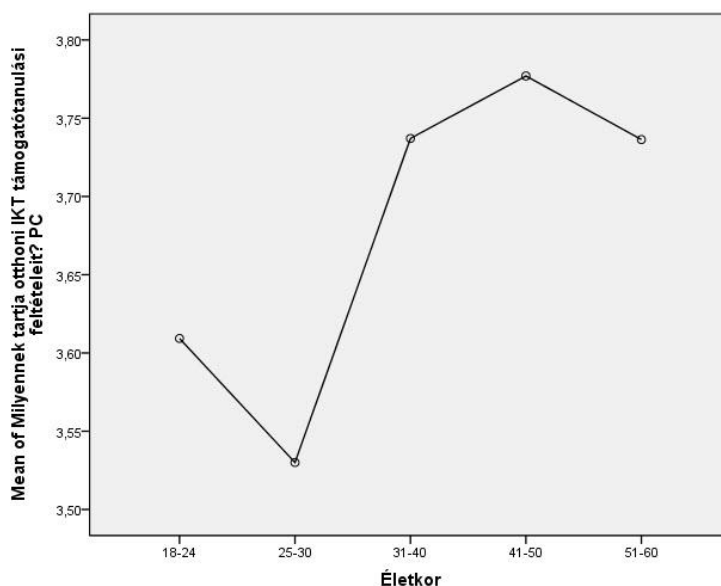
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	6,211	4	1,553	,784	,536
Within Groups	1677,661	847	1,981		
Total	1683,872	851			

86. ábra: Anova elemzés

A továbbiakban tekintsük át, hogy a csoportok közötti és a csoporton belüli négyzetösszegek vizsgálatával mit kapunk.

A táblázat alapján megállapítható, hogy a csoportok közötti (Between Groups) és a csoporton belüli (Within Groups) eltérésnégyzetek arány $F=1,553/1,981=0,784$. Az F próbához tartozó szignifikanciaszint 0,536, vagyis nagyobb, mint 0,05. A kategóriaátlagok nem különböznek egymástól szignifikánsan, a **PC használata közel azonos mértékben hat a különböző korosztályú hallgatókra.**

Az alábbi görbe az átlagok vizuális képét szemlélteti. A legerősebb válaszreakció a 31-40; 41-50 és az 51-60 éves korosztályra jellemző.



87. ábra: PC használat átlagértékei

Laptop alkalmazása

A mérést ANOVA elemzéssel végeztük, melynek során a független változó az életkor és a függő változó a laptop alkalmazása. A varianciaanalízis rámutatott a varianciák megegyezésének kérdésére.

Az alábbi táblázat alapján megállapítható, hogy a szóráshomogenitás feltétele teljesül, melynek bizonyítéka, hogy a Levene statisztika nem szignifikáns (0,996). A Levene teszt nullhipotézise azt mondja ki, hogy a szórások nem egyenlők, melynek következménye hogy a nullhipotézist elvetjük és a szóráshomogenitás teljesül.

Test of Homogeneity of Variances			
Laptop			
Levene Statistic	df1	df2	Sig.
,048	4	847	,996

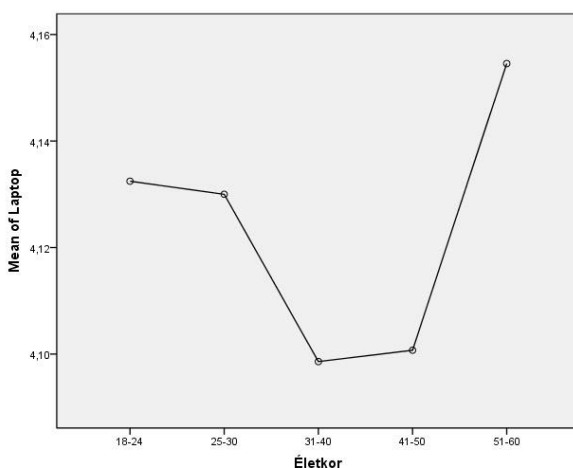
ANOVA					
Laptop					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	,353	4	,088	,055	,994
Within Groups	1361,143	847	1,607		
Total	1361,496	851			

88. ábra: Levene és ANOVA statisztika

A továbbiakban tekintsük át, hogy a csoportok közötti és a csoporton belüli négyzetösszegek vizsgálatát.

A táblázat alapján megállapítható, hogy a csoportok közötti (Between Groups) és a csoporton belüli (Within Groups) eltérésnégyzetek arány $F=0,088/1,607=0,055$. Az F próbához tartozó szignifikanciaszint 0,994, vagyis nagyobb, mint 0,05. A kategóriaátlagok nem különböznek egymástól szignifikánsan, az informatikai eszközök használata közel azonos mértékben hat a különböző korosztályú hallgatókra.

Az alábbi görbe az átlagok vizuális képét szemlélteti. A legerősebb válaszreakció a 18-24; 25-30; 51-60 éves korosztályra jellemző.



89. ábra: Laptop használat átlagértékei

A továbbiakban vizsgáljuk meg összehasonlítva az IKT eszközök alkalmazását

Test of Homogeneity of Variances

	Levene Statistic	df1	df2	Sig.
Milyennek tartja otthoni IKT támogatótanulási feltételeit? PC	1,396	4	847	,233
Laptop	,048	4	847	,996
Szkenner	1,293	4	847	,271
Internet	6,140	4	847	,000
webkamera	,239	4	847	,916
ebook olvasó	7,429	4	847	,000
táblagép	,997	4	847	,408
okostelefon	1,163	4	847	,326
interaktív tévé	2,138	4	847	,074

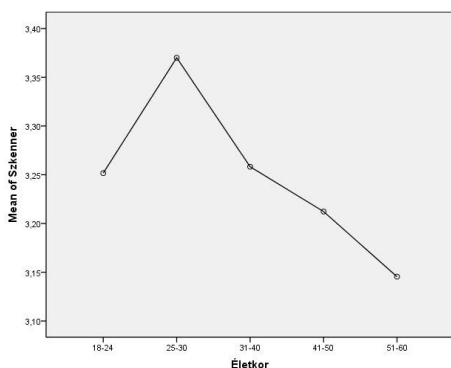
90. ábra: Homogenitás vizsgálat

ANOVA

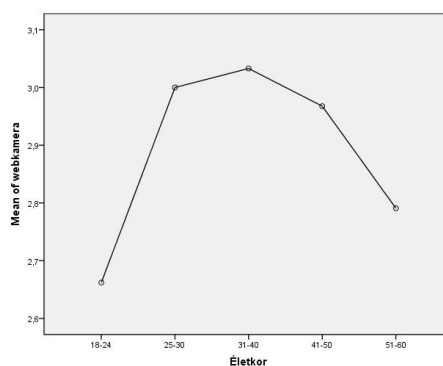
		Sum of Squares	df	Mean Square	F	Sig.
Milyennek tartja otthoni IKT támogatótanulási feltételeit? PC	Between Groups	6,211	4	1,553	,784	,536
	Within Groups	1677,661	847	1,981		
	Total	1683,872	851			
Laptop	Between Groups	,353	4	,088	,055	,994
	Within Groups	1361,143	847	1,607		
	Total	1361,496	851			
Szkenner	Between Groups	2,979	4	,745	,325	,861
	Within Groups	1938,696	847	2,289		
	Total	1941,675	851			
Internet	Between Groups	5,500	4	1,375	1,937	,102
	Within Groups	601,120	847	,710		
	Total	606,620	851			
webkamera	Between Groups	15,776	4	3,944	1,779	,131
	Within Groups	1877,444	847	2,217		
	Total	1893,221	851			
ebook olvasó	Between Groups	36,618	4	9,154	4,430	,002
	Within Groups	1750,195	847	2,066		
	Total	1786,812	851			
táblagép	Between Groups	15,078	4	3,769	1,461	,212
	Within Groups	2185,115	847	2,580		
	Total	2200,192	851			
okostelefon	Between Groups	33,185	4	8,296	3,105	,015
	Within Groups	2263,213	847	2,672		
	Total	2296,398	851			
interaktív tévé	Between Groups	8,302	4	2,076	1,007	,403
	Within Groups	1746,236	847	2,062		
	Total	1754,539	851			

91. ábra: ANOVA elemzés

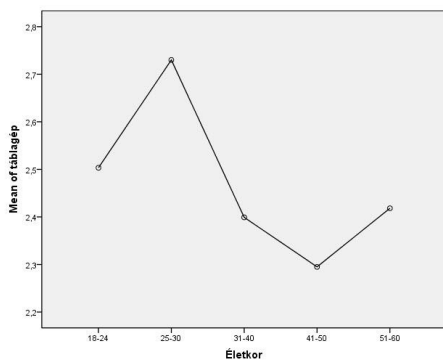
A fenti táblázat alapján megállapítható, hogy a szóráshomogenitás feltétele teljesül, melynek bizonyítéka, hogy a Levene statisztika nem szignifikáns ($\text{Sig} > 0,05$). A Levene teszt nullhipotézise azt mondja ki, hogy a szórások nem egyenlők, melynek következménye hogy a nullhipotézist elvetjük és a szóráshomogenitás teljesül. A kategóriaátlagok nem különböznek egymástól szignifikánsan, az informatikai eszközök közül a PC, laptop, szkennер, webkamera, táblagép, okostelefon és az interaktív TV használata közel azonos mértékben hat a különböző korosztályú hallgatókra. (25 és 27 ábra).



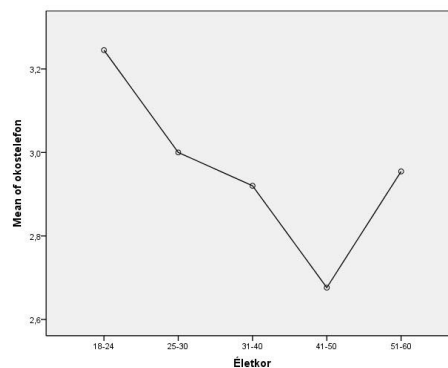
92. ábra: Szkenner használat átlagértékei



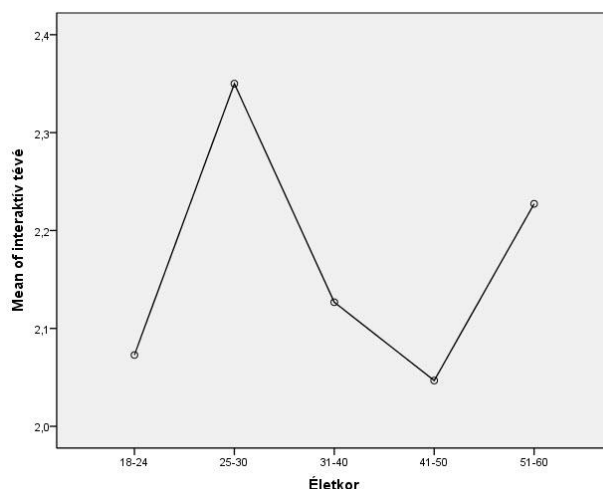
93. ábra: Webkamera használat átlagértékei



94. ábra: Táblagép használat átlagértékei

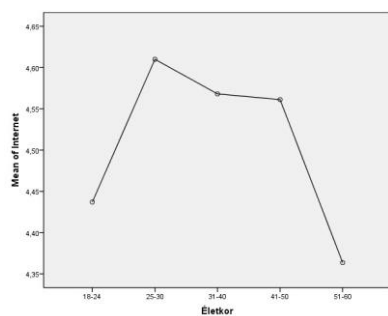


95. ábra: Okostelefon használat átlagértékei

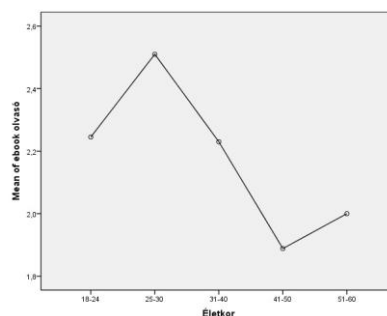


96. ábra: Interaktív TV használat átlagértékei

Az **Internet** és az **e-book** olvasó esetében a Levene statisztika szignifikáns ($\text{Sig} < 0,05$), a szórások egyenlők, melynek következménye hogy a nullhipotézist nem vetjük el és a szóráshomogenitás nem teljesül. Az Internet és az e-book olvasó használata nem azonos mértékben hat a különböző korosztályú hallgatókra.



97. ábra: Internet használat átlagértékei



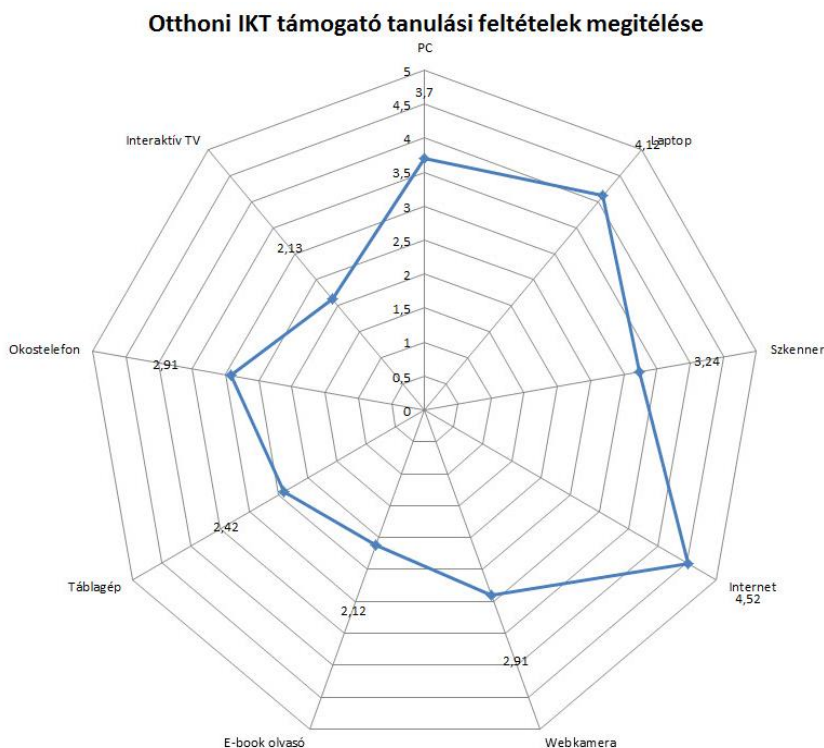
98. ábra: Ebook olvasó használat átlagértékei

Összefoglalva az alábbi pókdiagram alapján megállapítható, hogy a hallgatók válaszai alapján az átlagot képezve, az első három legnépszerűbb IKT eszköz az internet, (4,52), laptop (4,12) és a PC. (3,52)

A PC, laptop, szkennel, webkamera, táblagép, okostelefon és az interaktív TV használata és a különböző korosztályú hallgatók között varianciaanalízis eredménye nem szignifikáns, tehát az otthon alkalmazott IKT eszközök alkalmazása nem függ korosztálytól.

Az **Internet** és az **e-book** használata és a hallgatók korosztálya között szignifikáns ($\text{Sig} < 0,05$), kapcsolat mutatható ki az otthon alkalmazott IKT eszközök és a korosztály között.

A pókdiagram alapján megállapítható, hogy a hallgatók válaszai alapján az átlagot képezve, az első három legnépszerűbb IKT eszköz az internet, (4,52), laptop (4,12) és a PC. (3,52)



99. ábra: IKT eszközök otthoni tanulási feltételeinek megítélése, pók diagram

8.3 ÖSSZEFOGLALÁS, KÉRDÉSEK

8.3.1 Összefoglalás

Ebben a tananyagrészen az olvasó megismerkedhetett a több egyenlő szórású, normál eloszlású sokaság /minta várható értékének összehasonlítását szolgáló statisztikai módszerrel, mely ANOVA néven ismert. Az ismérvek alapján csoportosításra került, melynek jelentőségét az egyutas variaanalízisre mutatott példa szemlélteti.

A feladat során arra kapott választ, hogy egy nem metrikus függő változóra milyen hatással van a független változó értékei. A független változó szerint létrehozott csoportok átlagait képezve, azok egymástól eltérnek. Az eltérés szignifikáns jellegét a külső szórásnégyzet bizonyítja. A szignifikancia teszt az átlagos eltérés okozta varianciát vizsgálja. Szignifikáns eredményt kapva a H_0 elvethető és az alternatív hipotézisben megfogalmazott tény, a csoportátlagok különbözősége lesz bizonyított.

8.3.2 Önellenőrző kérdések

1. Mutassa be az ANOVA jellemzőit.
2. Ismertesse One way ANOVA egyutas variacionális eljárás lényegét az SPSS-ben.

9. MULTIPLE RESPONSES

9.1 CÉLKITŰZÉSEK ÉS KOMPETENCIÁK

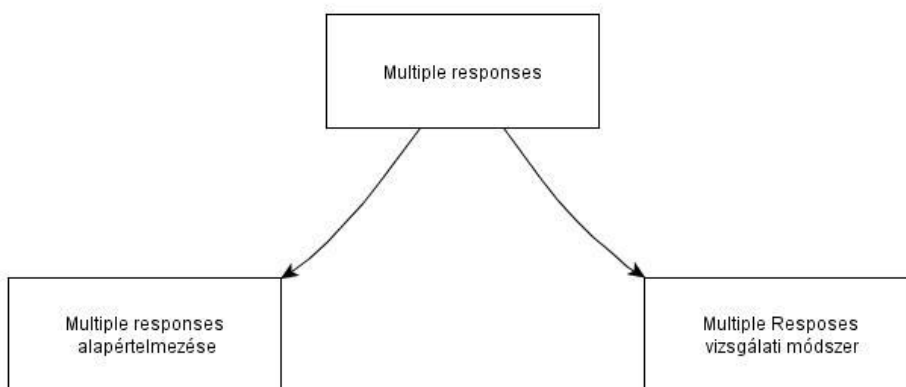
A statisztikai számítások során gyakori feladat, hogy a többszöri válaszadások alapján elemzés megvalósítása válik szükségessé. A kérdésben a válaszok logikailag összetartozó változók együttese. A változók dichotóm vagyis kétértékűek. A feldolgozás során arra vagyunk kíváncsiak, hogy a megkérdezettek válaszai milyen gyakorisággal kerültek kiválasztásra.

Az olvasó képessé válik a kutatásokban gyakran alkalmazott kérdéscsoportok elemzését.

9.2 TANANYAG

Multiple responses alapértelmezése

Multiple Responses vizsgálati módszer



100. ábra: Fogalomtérkép

9.2.1 Multiple Responses alapértelmezése

Az adott változókat dichotóm (igaz/hamis; igen/, által nem; választotta/nem választotta). A menüpont által maximum 20 csoport elemezhető, gyakorisági és keresztkorrelációs kapcsolatok vizsgálata hozható létre

Statisztikai elemzés végrehajtása során kattintsunk az **Analyze** parancsra és a lehulló menüből válasszuk ki a **Multiple response, Define Sets...** parancsot.

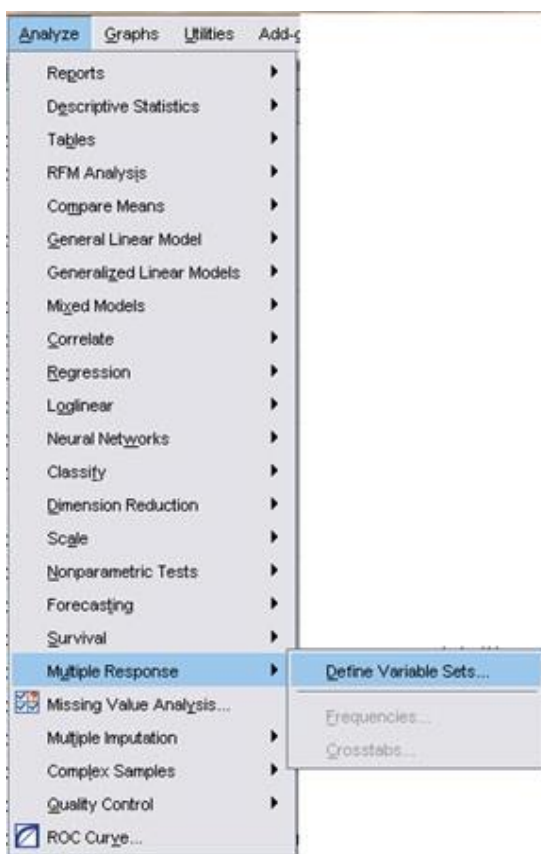
Az adatokból a logikailag összetartozókat és húzzuk be a Variables in Set ablakba (max. 20 lehet), mindegyik kétértékű.

Az összeszámlálандó érték 1, ezt rögzítjük a Counted Value cellájába. Az adatok két értéket vehetnek fel, ezért kattintsunk a Dichotomies rádiógombra.

A gyakoriság elemzése az Analyze, Multiple Response, Frequencies... paranccsal

Van-e összefüggés a tantárgy szeretete és a hallgató kora között, választ a többszörös választású keresztábra (Multiple Response Crosstabs) elemzésével kapunk., A válaszadók a biológiát kedveli.

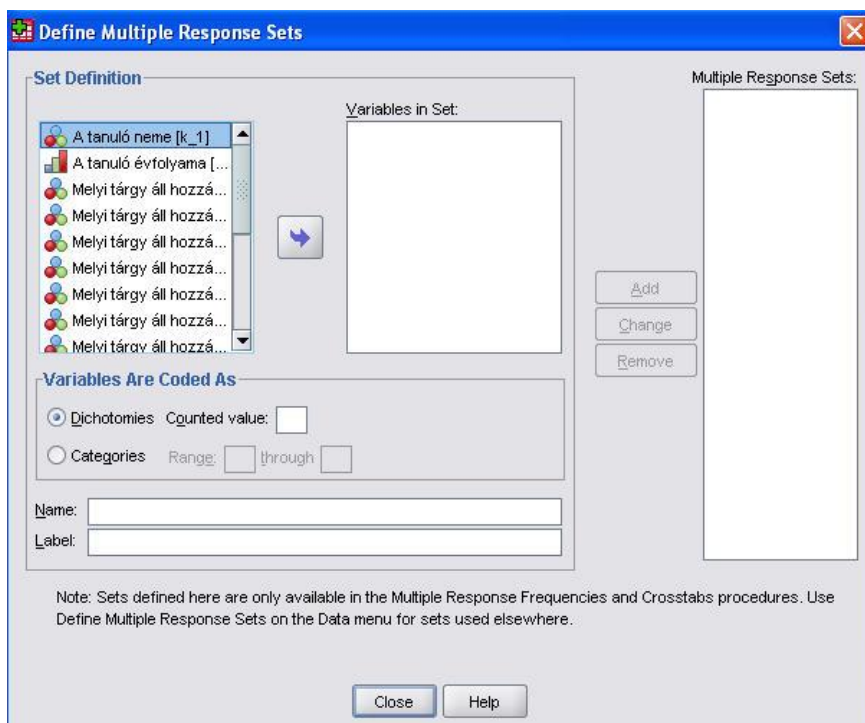
A Percent of Cases oszlopban az érvényes válaszok százalékban megadott értékeit kapjuk, mely nem tartalmazza az adott kérdésre választ nem adók arányát.



101. ábra: A Multiple Response, Define Variable Sets... parancs

Az összeszámlálандó érték 1, ezt rögzítjük a Counted Value cellájába. Az adatok két értéket vehetnek fel, ezért kattintsunk a Dichotomies rádiógombra.

- ✿ Van-e összefüggés az iskolai tantárgyak iránti motiváltság és a tanulók osztálya között? Választ a többszörös választású keresztábla (Multiple Response Crosstabs) elemzésével kapunk. Az elemzést a **Analyze/Multiple responses/Define Variable Sets** parancssor indításával érhető el.

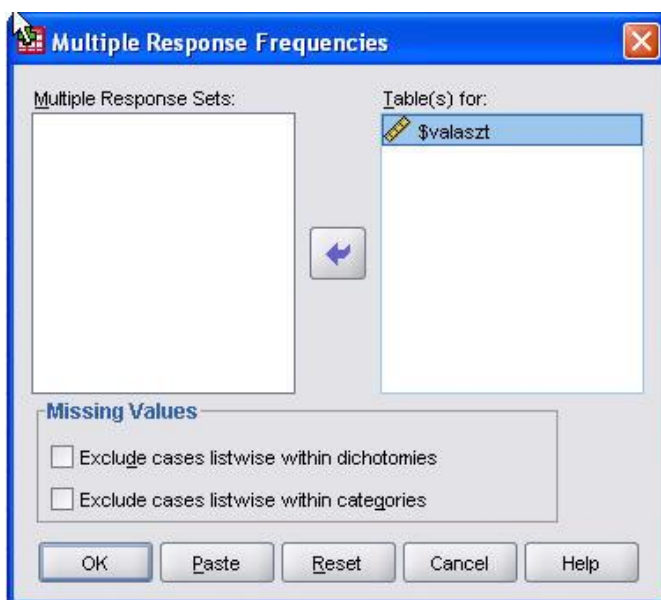


102. ábra: A Multiple Response, Define Variable Sets lehulló menü

A gyakoriság elemzése az **Analyze, Multiple Response, Frequencies...** parancssal valósítható meg.

Van-e összefüggés a tantárgy szeretete és a hallgató kora között, választ a többszörös választású keresztábla (Multiple Response Crosstabs) elemzésével kapunk. A válaszadók a biológiát kedvelik legnagyobb százalékban.

A Percent of Cases oszlopban az érvényes válaszok százalékban megadott értékeit kapjuk, mely nem tartalmazza az adott kérdésre választ nem adók arányát.

Az eljárás menete és a válaszpanelek beállításai:

103. ábra: A többszörös választás keresztábla beállítása

Analyze/Multiple responses/Crosstabs

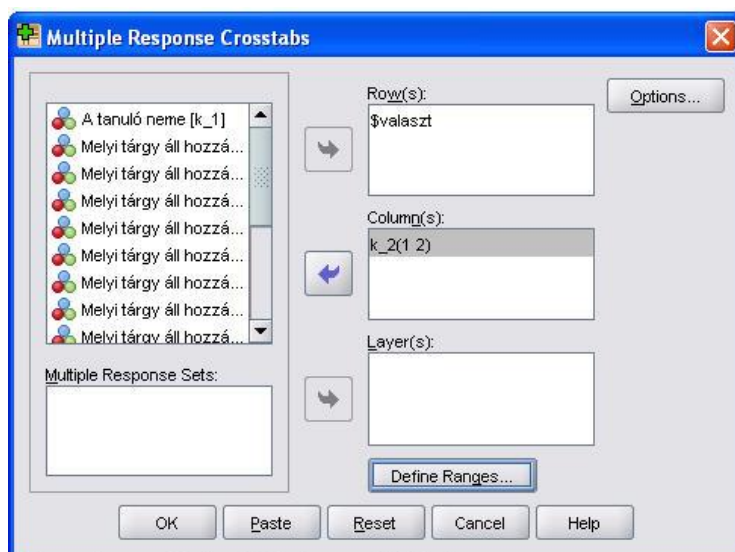
A kérdésben a válaszok logikailag összetartozó változók együttese. A változók dichotóm vagyis kétértékűek. A feldolgozás során arra vagyunk kíváncsiak, hogy a megkérdezettek válaszaik milyen gyakorisággal kerültek kiválasztásra.

Statisztikai elemzés: kattintsunk az Analyze parancsra és a lehulló menüből válasszuk ki a Multiple response, Define Sets...parancsot.

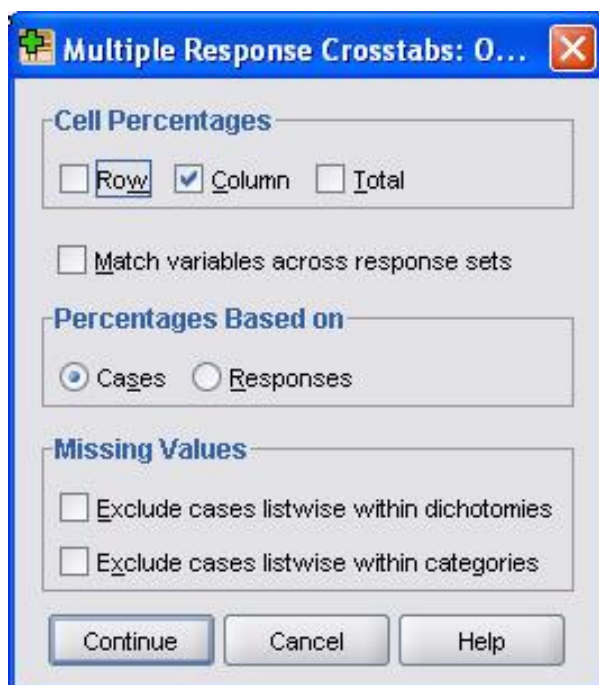
Az adatokból a logikailag összetartozókat és húzzuk be a Variables in Set ablakba (max. 20 lehet), mindegyik kétértékű.

Az összeszámlálандó érték 1, ezt rögzítjük a Counted Value cellájába. Az adatok két értéket vehetnek fel, ezért kattintsunk a Dichotomies rádiógombra.

Az alábbi keresztábla beállításánál a Row(s) ablakba behúzzuk a virtuális függő változót, a Colum(s)ba pedig a független változót, melynek a határait meghatározzuk a Define Ranges által.



104. ábra: A többszörös választás keresztábla beállítása



105. ábra: A Multiple Responses lehulló sor és oszlop beállítása

Eredmény:

A Percent of Cases oszlopban az érvényes válaszok százalékban megadott értékeit kapjuk, mely nem tartalmazza az adott kérdésre választ nem adók arányát.

Kutatásunk során felmerül, hogy a válaszok hogyan függnek a korosztálytól. A választ a többválasztásos kereszttábla, a Multiple Response Crosstabs elemzéssel kapjuk. Az Analyze, Multiple Response, Crosstabs...parancssorra kattintva, a lehulló válaszpanelen a sorba Row, húzzuk be a \$vizsga és az oszlopba pedig a hallgatók korát.

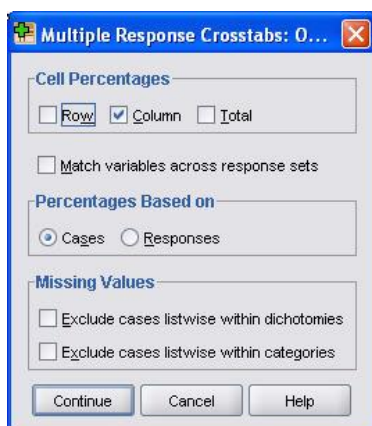
A Define Rangerre kattintva meg kell adni az elemzésbe bevont hallgatók korának tartományát, melyet 5 kategóriába soroltunk. Tehát a minimum értéke 1, a maximum 5.

Majd az Opcions-ra kattintva beállítjuk a cellákat, hogy az oszlopok értékeit százalékba kapjuk.

Ezt követően az Options gombra kattintva a lehulló ablakba beállítjuk, hogy a cellákban kapott értékek oszloponként kiszámítva, százalékban kérjük és a korosztályok alapján.

A beállítások végrehajtását követően az OK gomb aktív lesz, melyre kattintva az output felületén kapott táblázatos eredményből kiolvashatóak a következtetések.

Az iskolai felmérés során feltett kérdés alapján a tanulók az adott osztály tantárgyi listájában megjelölték az általuk kedvelt tantárgyakat (többet is megjelölhettek)



106. ábra: A Multiple Responses lehulló Kereszttábla beállítása

Az OK gombra kattintva az Output ablakban a kapott keresztábra eredményei értelmezhető.

Case Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
\$valaszt*k_2	46	100,0%	0	,0%	46	100,0%

\$valaszt*k_2 Crosstabulation

			A tanuló évfolyama		Total
			9. évfolyam	12. évfolyam	
Valaszt ^a	Melyi tárgy áll hozzád közelebb? Egyik sem	Count	24	16	40
		% within k_2	100,0%	72,7%	
	Melyi tárgy áll hozzád közelebb? matematika	Count	13	18	31
		% within k_2	54,2%	81,8%	
	Melyi tárgy áll hozzád közelebb?	Count	14	20	34
		% within k_2	58,3%	90,9%	
	Melyi tárgy áll hozzád közelebb? kémia	Count	17	16	33
		% within k_2	70,8%	72,7%	
	Melyi tárgy áll hozzád közelebb? magyar	Count	22	20	42
		% within k_2	91,7%	90,9%	
	Melyi tárgy áll hozzád közelebb? történelem	Count	22	17	39
		% within k_2	91,7%	77,3%	
	Melyi tárgy áll hozzád közelebb? angol	Count	22	20	42
		% within k_2	91,7%	90,9%	
	Melyi tárgy áll hozzád közelebb? testnevelés	Count	23	21	44
		% within k_2	95,8%	95,5%	
	Melyi tárgy áll hozzád közelebb? biológia	Count	21	22	43
		% within k_2	87,5%	100,0%	
Total		Count	24	22	46

Percentages and totals are based on respondents.

a. Dichotomy group tabulated at value 1.

107. ábra: Multiple Responses keresztábra eredmény táblázata

A fenti táblázat alapján megállapítható, hogy az osztályban a legkedveltebb tantárgy a testnevelés. A 8. osztályosok 95,8%-a, a 12. osztályosok 95,5%-a jelölte meg a legnépszerűbb testnevelést.

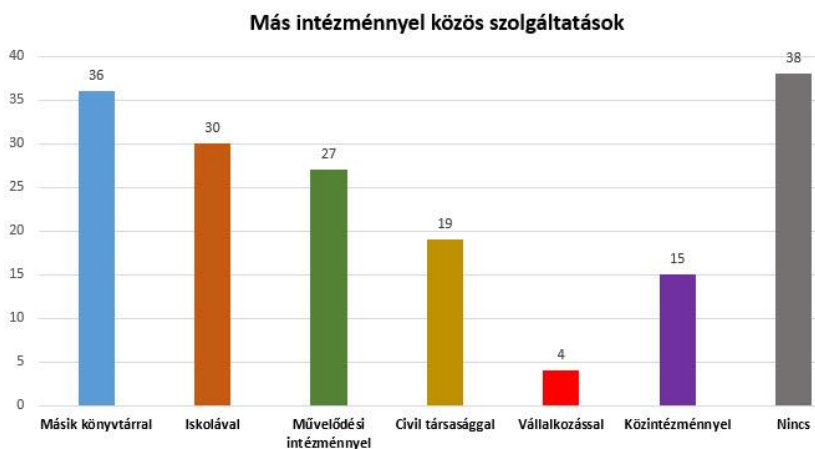
- ✿ A kutatás során arra keresünk választ hogy egy megnevezett könyvtár milyen kapcsolatot ápol más könyvtárakkal, melynek listája felsorolt. A válaszok során több választ is megjelölhettek.

- a gyakorisági táblázata a megkérdezettek összes válaszait tartalmazza, egy válaszadó több szolgáltatást is választhatott, a TOTAL N az összes válaszadás száma esetenként.
- Értelmezhető a tevékenység során milyen arányban használják az adott szolgáltatást
- Percent of cases oszlop az érvényes válaszok %-ban kifejezett aránya. Vannak-e más intézménnyel közös szolgáltatások a könyvtárban?

		Responses		Percent of Cases
		N	Percent	
SZOLG ^a	Vannak-e más intézménnyel közös szolgáltatásaik? Másik könyvtárral	36	21,3%	37,9%
	Iskolával	30	17,8%	31,6%
	Művelődési intézménnyel	27	16,0%	28,4%
	Civil társasággal	19	11,2%	20,0%
	Vállalkozással	4	2,4%	4,2%
	Közüntézménnyel	15	8,9%	15,8%
	Nincs	38	22,5%	40,0%
Total		169	100,0%	177,9%

a. Dichotomy group tabulated at value 1.

108. ábra: Multiple Responses gyakorisági táblázata



109. ábra: Multiple Responses eredménydiagram

			neme		Total
			férfi	nő	
SZOLG* Vannak-e más intézményei közös szolgáltatásaik? Másik könyvtárral	Count		6	30	36
	% within szolg		16,7%	83,3%	
	% of Total		6,4%	31,9%	38,3%
Iskolával	Count		6	24	30
	% within szolg		20,0%	80,0%	
	% of Total		6,4%	25,5%	31,9%
Művelődési intézménnyel	Count		7	20	27
	% within szolg		25,9%	74,1%	
	% of Total		7,4%	21,3%	28,7%
Civil társasággal	Count		4	15	19
	% within szolg		21,1%	78,9%	
	% of Total		4,3%	16,0%	20,2%
Vállalkozással	Count		1	3	4
	% within szolg		25,0%	75,0%	
	% of Total		1,1%	3,2%	4,3%
Közműintézménnyel	Count		2	13	15
	% within szolg		13,3%	86,7%	
	% of Total		2,1%	13,8%	16,0%
Nincs	Count		1	36	37
	% within szolg		2,7%	97,3%	
	% of Total		1,1%	38,3%	39,4%
Total		Count	12	82	94
		% of Total	12,8%	87,2%	100,0%

Percentages and totals are based on respondents.

a. Dichotomy group tabulated at value 1.

110. ábra: Multiple Responses keresztábra

Legmagasabb az eredmény a „Másik könyvtárral” közös szolgáltatás meg-
léte de megjegyzendő, hogy a minta 39,4%-a nem adott választ a kérdésre.

9.3 ÖSSZEFOGLALÁS, KÉRDÉSEK

9.3.1 Összefoglalás

A kérdőívek eredményeit kiértékelő eljárásoknak kell alá vetni. A fejezetben az olvasó a többválasztásos kérdéstípus SPSS statisztikai szoftverrel dolgoz-
za fel és megismerkedik mely lépései alkalmasak a értékelésre.

A fejezet tanulmányozása során a hallgató képessé válik a teszt empirikus
adatainak értékelésére és a következtetések levonására.

9.3.2 Önellenőrző kérdések

1. A több választ adó kérdések kiértékelése milyen lépésekben tör-
ténik az SPSS szoftverrel?
2. Mutassa be a Multiple Responses keresztábra elemzés folyama-
tát és értelmezze a kapott paramétereket.

10. KLASZTERANALÍZIS

10.1 CÉLKITŰZÉSEK ÉS KOMPETENCIÁK

A többváltozós statisztikai vizsgálatok jellegzetes feladata az objektumok elemzése, a struktúrát egészében vizsgáló módszer. Alkalmazásakor az osztályozandó objektumokat számkomponensű vektorokkal kell megadni. Általában törekedni kell arra, hogy a vektor dimenziója ne legyen túl nagy.

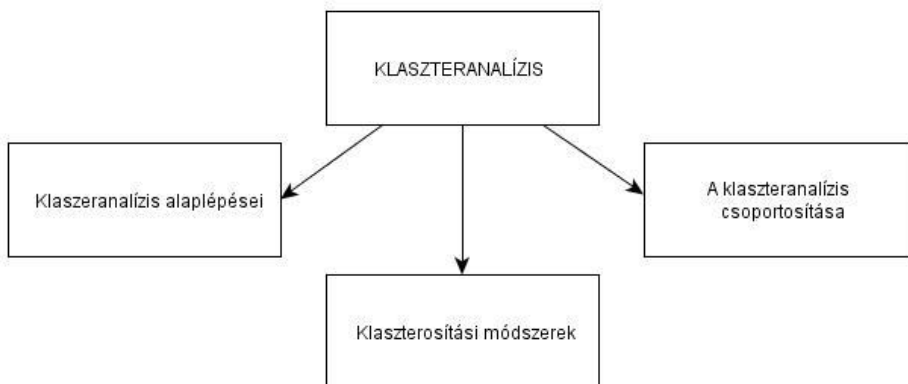
Egy ilyen igen gyakran alkalmazott osztályozási módszer a klaszteranalízis. Feladata az, hogy csoportokba soroljuk a különböző objektumokat azok hasonlósága alapján, közös tulajdonságaik figyelembe vételével.

10.2 TANANYAG

Klaszteranalízis alaplépései

A klaszteranalízis csoportosítása

Klaszterosítási módszerek



111. ábra: Fogalomtérkép

10.3 KLASZTERANALÍZIS ALAPÉRTELMEZÉSE

A kutatások során gyakori feladat, hogy a változókat homogén csoportba rendezze az elemzés alapját képező változók alapján. Egyik legismertebb osztályozási módszer a **klaszteranalízis**. Feladata az, hogy csoportokba soroljuk a különböző objektumokat azok hasonlósága alapján, közös tulajdonságaik figye-

lembe vételével. Olyan osztályokat definiál, amelyekben az objektumok egy vagy több változóban nem feltétlenül ekvivalensek, de hasonlóak. Osztályozást hajt végre az eljárás, feltétele stabil és optimális megoldás biztosítása.

A klaszteranalízis a megfigyelések (vagy a változók) osztályozásának dimenziócsökkentő módszere. A klasszikus logika első lépésben definiálja a típusokat és utána sorolja az egyedeket osztályokat.

Mi különbözteti meg a diszkriminancia analízistől klasteranalízist? A diszkriminancia analízissel szemben a klaszteranalízisben nincsenek előre megadott osztályok, a feladatunk éppen ezeknek a létrehozása.

Mi a különbség a faktorelemzés és a klaszterelemzés között? A klaszterelemzés feladata nem a változók számának csökkentése, hanem a megfigyelési egységek számát határozza meg. A diszkriminancia elemzéshez hasonlóan csoportosítással foglalkozik, de klaszterelemzés során nem ismert a csoportba tartozás előzetes ismeretei, nincs előzetes ismeret a klasztertagságról. Természetes elvárás az, hogy **azok** a megfigyelések kerüljenek egy osztályba (klaszterbe), amelyek a legközelebb vannak egymáshoz, illetve a leginkább hasonlóak egymáshoz. Az elemzés kezdetén meg kell határozni a megfigyelések közötti távolság módját vagy az ezzel ellentétesen viselkedő hasonlóságot, de megfelelő a standard euklideszi távolság. A diszkrét vagy bináris adatok esetén más távolság alkalmazása is megengedett.

A klaszteranalízist leggyakrabban alkalmazott területei:

- Típusalkotás,
- Modellillesztés,
- csoportokon alapuló becslés,
- hipotézis-tesztelés,
- adatstruktúrák felderítése,
- hipotézis felállítása,
- adatredukció,

Elméletileg a klaszteranalízis a következő lépések sorozatából áll⁹:

A **klasztertendencia vizsgálat** célja annak eldöntése, hogy az adatok mutatnak-e hajlamosságot a természetes csoportosulásra. Ha az adatok hasonlóságot mérő mátrix elemei ordinális értékek, akkor a véletlen gráfelmélet az alkalmas módszer

⁹ Tóthné Parázso Lenke: A kutatómódszertan matematikai alapjai Eger 2011. ISBN 978-615-5221-25-5 pp.67-69

Amennyiben az adatok intervallum skálátípusúak, akkor az ún. véletlen gráf hipotézist kell alkalmazni. Az elméleti eredményeket nem könnyű a gyakorlatban megvalósítani, ezért alkalmazásokban még elég ritkán lehet találni klasztertendencia vizsgálatot.

A klaszterezés az objektumok osztályba sorolását jelenti, vagyis az objektumok halmazának (X) C_1, C_2, \dots, C_M részhalmazokra való felbontását.

A csoportoknak diszjunktaknak kell lenniük és együttesen a teljes halmazt kell adniuk. A klaszterezés során az objektumok a hasonlóak egy klaszterbe, a különbözőek külön klaszterbe kerülnek. Az osztályok kialakítása a rendelkezésre álló mintából, valamilyen döntési kritérium alapján történik. A klaszteranalízis nem egy módszer, mint a faktoranalízis, hanem **módszerek együttese**. Ezek sokfélesége miatt a klaszteranalízisnek igen sokféle eljárása létezik.

10.3.1A klaszteranalízis csoportosítása

Az osztályokba való sorolás két módszerét különböztethetjük meg: a hierarchikus módszereket és a nemhierarchikus módszereket.

A hierarchikus módszereken belül megkülönböztetünk összevonó és felosztó eljárásokat.

Az összevonó eljárások általános menete:

- n db egyelemű csoportból történő kiindulás
- A két leghasonlóbb klaszter megkeresése.
- A két klasztert összevonása, így a klaszterek számát eggyel való csökkentése. Az új klaszter és a régi klaszterek közti hasonlósági mérőszámok újra meghatározása.
- A második és a harmadik lépést $n-1$ -szer elvégezve minden egyed egy osztályba kerül.

A módszerek a csoportok hasonlóságának definiálásának módszerében különböznek. Ilyen módszerek pl.: egyszerű lánc-, teljes lánc-, centroid módszer stb. Az összevonó eljárások eredményét megjeleníthetjük az ún. dendrogramon mely a klaszterek hierarchikus elrendeződését ábrázolja. A vízszintes tengelyen az egyedek sorszámain, a függőleges tengelyen a klaszterek összevonási szintjeit jelöljük.

A vízszintes tengelyen az egyedek sorszámain, a függőleges tengelyen a klaszterek összevonási szintjeit jelöljük. A különböző klaszterezési módszerek általában különböző dendrogramokat eredményeznek, melyek jellemzői:

- Szimmetrikusak,

- eggyel egyenlő, ha a két dendrogram azonos,
- nullával egyenlő, ha a két dendrogram teljesen különböző.

A felosztó eljárások közé tartozik pl. az asszociációs elemzés, ahol a csoportokat dichotómia szerint bontjuk egymás után kisebb elemszámú csoportokra.

A nemhierarchikus módszerek általános felépítése

- a kezdő klaszterek kialakítása,
- az egyedek elhelyezése a kezdő klaszterekbe,
- az egyedek átrendezése a klaszterek között valamilyen optimalizálókritérium szerint.

10.3.2 Klaszterosítási módszerek

A különbség a hierarchikus módszerrel, mely átlagos kapcsolású, legközelebbi társ vagy centroid módszere és a dinamikus módszerrel valósítható meg, melyek közötti különbség:

- A hierarchikus módszereknél nem kell előzetesen ismernünk a létrehozandó klaszterek számát, ebben különféle grafikonok segítenek, csak kis mintaelemszámú populáció esetén ajánlott. A dinamikus módszernél, ezzel szemben már kiinduláskor adott a klaszterek száma, a feladat csak a megfigyelések besorolása. A klasztereket iterációval kell számolni.
- A másik fontos különbség, hogy egy hierarchikus módszer általában időigényesebb mint egy dinamikus klaszterezés, amelyet emiatt gyakran gyors klaszterezésnek is nevezik.

A klaszteranalízist alkalmazását eldöntő tényezők:

- A legfontosabb segítséget a megfigyelések grafikus ábrázolása adja. Ha az így kapott pontfelhőben jól elkülönülő csoportok alakulnak ki, akkor feltétlen érdemes klaszteranalízist alkalmazni. (Persze ez csak három változóra tehető meg, ennél több változó esetén előbb valamilyen dimenziócsökkentő eljárást, pl. főkomponensanalízist, kell alkalmaznunk).
- Egy másik lehetőség a bimodalitási együttható. Ha ez 0.555-nél (az egyenletes eloszlásnál ezt az értéket veszi fel) nagyobb, akkor az két vagy több csúcosságra utal, ami esetleg több klaszter jelenlétét jelenti. Ezen együttható maximális értéke 1, melyet a kétértékű **Bernoulli eloszlás** esetén vesz fel. A hierarchikus módszereknél a távolság definíciója mellett meg kell adni a klaszterösszevonási szabályát, melynek alapján ha több elemű, nagyobb klaszterünk van, akkor hogyan definiáljuk a közöttük lévő távolságot.

A hierarchikus módszereknél döntenünk kell arról, hogy hány klasztert érdemes választani. Ez a probléma máig sem teljesen megoldott.

Az elemzés lépései:

- *A megfigyelések grafikus ábrázolása a lehetséges klaszterek beazonosítása céljából.*
- *Leíró statisztikák: átlag, szórás, ferdeség, lapultság, bimodalitás.*
- *A klaszterezés történetét tartalmazó táblázat: az összevonások sorrendje és a kapcsolódó statisztikák.*
- *A klaszterezési szint megállapítását segítő grafikonok: pseudo F és t statisztikák.*
- *A klaszterezés végeredményének grafikus ábrázolása: a dendogram.*
- *A klaszterek számának megválasztása, az egyes klaszterek listázása.*

10.3.3a klaszteranalízis gyakorlati alkalmazása

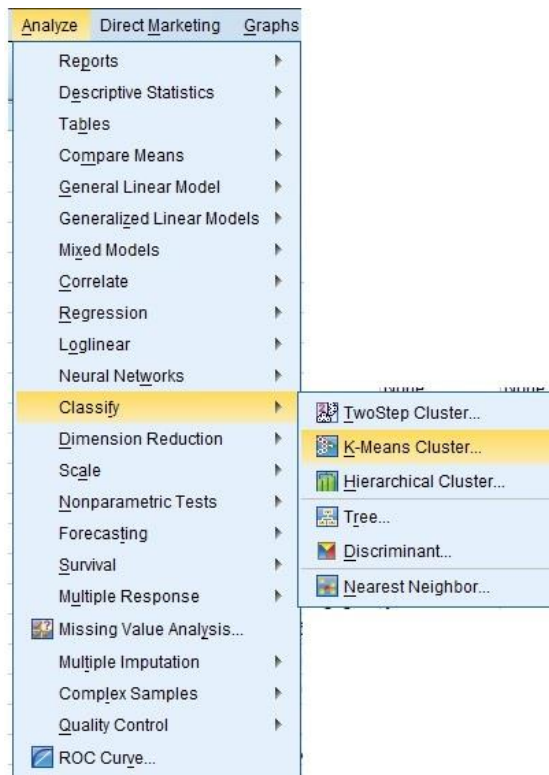
- ✿ Feladat a webkettes alkalmazásokat mennyire fontossága a pedagógiai munkában az életkor függvényében. A kérdőívet 852 fő töltötte ki.

A feladat során fontos tényező, hogy ne legyen sok kiugró adat, amelyet az adattisztítás során kell elvégezni. Feltételezzük az adatok tiszták.

Abban az esetben, ha változók különböző skálán felvett változók, standartizálni kell, amelynek során az átlagot ki kell vonni az egyes értékekből és különbség szórását kell alapul venni. Az SPSS ezt a műveletet az **Analyze/Descriptive Statistics/Descriptives** menűben a **Save standartized values as variables** parancs bejelölésével valósítja meg. Jelen esetben az életkor intervallumokban kódolt és a változóink is kódolt adatok, ezért értékük azonos skálában lettek felvéve, nem kell standartizálni.

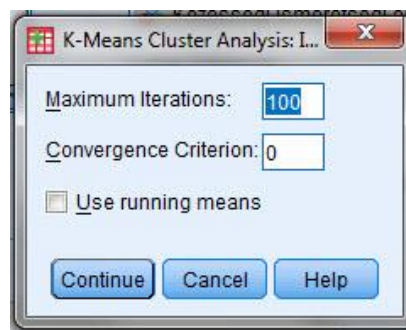
A klaszterelemzés folyamata

Első lépésünk 2 klaszter létrehozása a feladat, mely az alábbi parancssorral végezhető.

Analyze/Classify/K Means Cluster

112. ábra: *Analyze/Classify/K Means Cluster*

Beállítjuk a Number of Cluster mezőben a 2, az Iterations mezőben 10-t átállítjuk 100-ra, mivel feltételezés, hogy 10 ismétlés nem vezet eredményre.



113. ábra: *Iteration beállítása*

Lefuttatva az alábbi klaszterközpontokat tartalmazó táblázatot kapjuk.

Iteration History^a		
Iteration	Change in Cluster...	
	1	2
1	4,438	3,184
2	,761	,254
3	,476	,197
4	,203	,091
5	,144	,068
6	,082	,039
7	,071	,035
8	,040	,021
9	,031	,016
10	,029	,015
11	,010	,005
12	,000	,000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is ,000. The current iteration is 12. The minimum distance between initial centers is 10,630.

114. ábra: *Iteration history táblázat*

A fenti táblázatból kiolvasható, hogy az SPSS 12 iterálás után jutott el a végső klaszterstruktúra kialakításához

Final Cluster Centers		
	Cluster	
	1	2
RSS olvasó	2	2
linkmegosztó szolgáltatások hivatkozásmegosztót	2	2
fogalomtérkép készítő alkalmazást	2	2
Mikroblog szolgáltatást	2	2
Jegyzetelő szolgáltatás	2	2
Közösségi ismeretségi oldal	1	1
Szakmai közösségi oldal, referenciaoldal	1	2
Online tárhely megosztható tartalommal	1	2
Közösségi játékok	2	2
Az alábbi webkettes alkalmazásokat milyen gyakran használja? RSS olvasó	2	1
Linkmegosztó szolgáltatások hivatkozásmegosztót	3	1
fogalomtérkép készítő alkalmazást	2	1
Mikroblog szolgáltatás	2	1
Jegyzetelő alkalmazás	2	1
Közösségi ismertségi oldal	4	3
Szakmai közösségi oldal, referenciaoldal	3	2
Online tárhely megosztható tartalommal	3	1
Közösségi játékok	2	2

115. ábra: Végső klaszterközpontok

Végső klaszterközpontokban az esetek elhelyezkedése, az alábbi táblázatban a Number of Case arra mutat, hogy a klaszterekbe mennyi az előforduló esetek száma.

Number of Cases in each Cluster		
Cluster	1	293,000
	2	559,000
Valid		852,000
Missing		,000

116. ábra: Az esetek száma a klaszterekben

Következő lépésként meg kell vizsgálni, hogy a klaszterképző központok szignifikánsan különböznek. Ez a vizsgálat a *Analyze/Classify/K Means Cluster* menü *Option* almenüjében a klasszikus ANOVA rádiógomb kipipálásával végezhető el.

ANOVA						
	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
RSS olvasó	13,236	1	,151	850	87,787	,000
linkmegosztó szolgáltatások hivatkozásmegosztót	14,014	1	,181	850	77,401	,000
foglalomtérkép készítő alkalmazást	6,321	1	,104	850	61,064	,000
Mikroblog szolgáltatást	3,017	1	,088	850	34,142	,000
Jegyzetelő szolgáltatás	3,711	1	,060	850	61,607	,000
Közösségi ismeretségi oldal	,287	1	,133	850	2,152	,143
Szakmai közösségi oldal, referenciaoldal	18,238	1	,193	850	94,548	,000
Online tárhely megosztható tartalommal	33,965	1	,177	850	191,398	,000
Közösségi játékok	1,399	1	,204	850	6,853	,009
Az alábbi webkettes alkalmazásokat milyen gyakran használja? RSS olvasó	225,674	1	,915	850	246,516	,000
Linkmegosztó szolgáltatások hivatkozásmegosztót	294,623	1	,765	850	385,254	,000
foglalomtérkép készítő alkalmazást	132,030	1	,606	850	218,023	,000
Mikroblog szolgáltatás	121,429	1	,596	850	203,694	,000
Jegyzetelő alkalmazás	123,946	1	,487	850	254,437	,000
Közösségi ismertségi oldal	68,938	1	1,655	850	41,661	,000
Szakmai közösségi oldal, referenciaoldal	328,361	1	1,150	850	285,455	,000
Online tárhely megosztható tartalommal	595,254	1	,977	850	609,495	,000
Közösségi játékok	48,086	1	1,152	850	41,754	,000

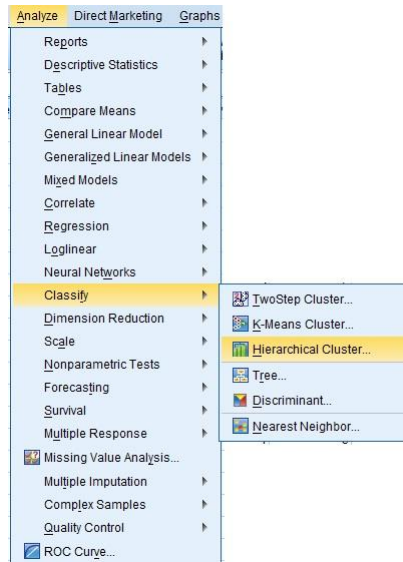
The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

117. ábra: Anova táblázat

A klaszterközpontok szignifikánsak. A táblázatban az F értéke azt jelenti, hogy mely változók körül alakulnak ki homogén csoportok. Minél nagyobb az F értéke, annál fontosabb szerepet játszik a klaszterstruktúra kialakításában. Mivel egy „a közösségi ismeretségi oldal nem szignifikáns, a fenti eljárás megismételhető háromklaszteres struktúrával.

A **nem hierarchikus struktúra** elvégzésével finomabb hangolás valósítható meg.

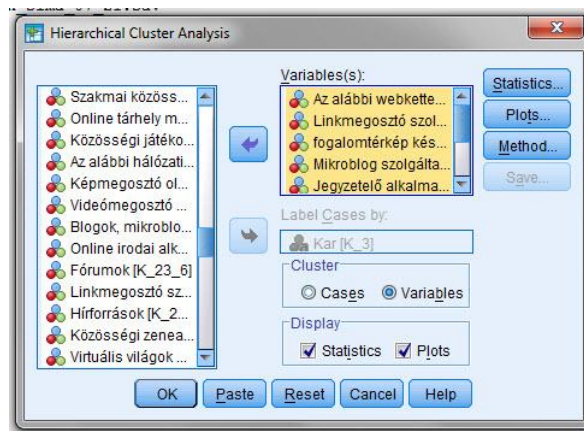
Indítsuk el az *Analyze/Classify/Hierarchical Cluster* menüsört.



118. ábra: *Analyze/Classify/Hierarchical Cluster*

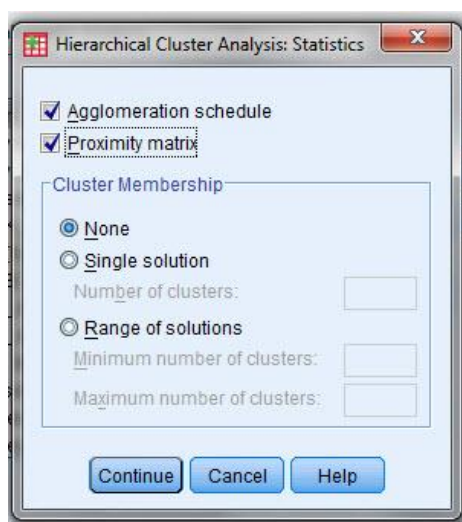
Ebben az esetben a Cases string alapú változó, jelen esetben a hallgatók karának megnevezése

A lehulló ablakban behúzzuk a változókat,



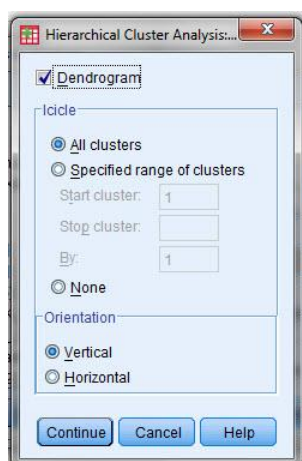
119. ábra: *Hierarchical Cluster menü*

Kipipáljuk a Variables rádiógombot., Majd a statistics gombra kattintunk.



120. ábra: A Statistics beállítása

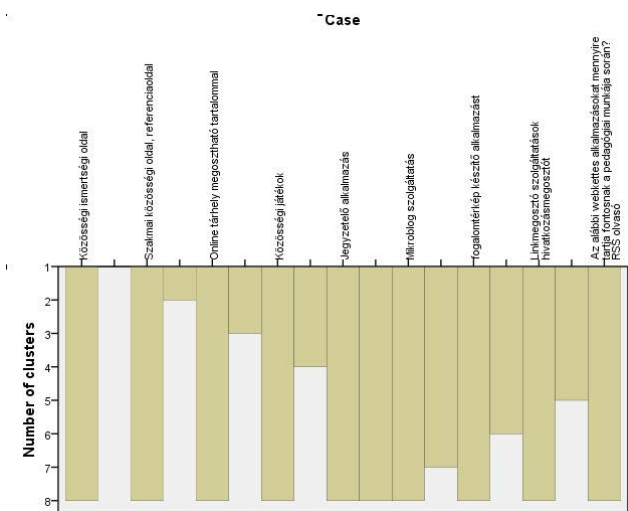
A Plots menüpontban kijelöljük a kívánt grafikai ábrázolásokat (dendrogram, icicle).



121. ábra: Plots menü beállítása

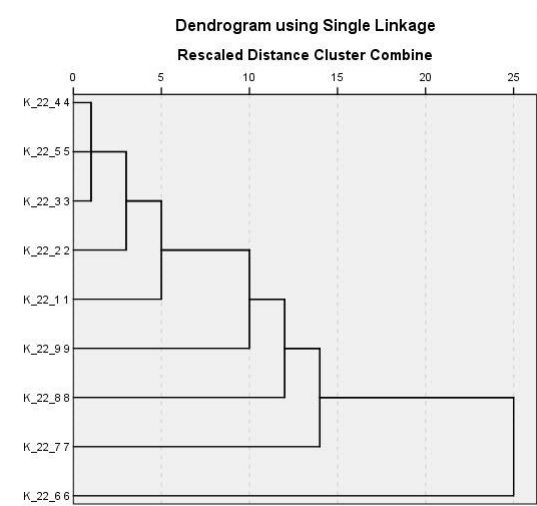
Ezt követően a Method menüben a Nearest Neighbort (legközelebbi szomszéd), mivel kiugró elemeket is vizsgálni kell. Az Interval esetében Squared Euclidean distance*t (négyzetes euklideszi távolság) célszerű jelölni.

Lefuttatva az eredménytábla fontosabb elemei az alábbi eredménytábla alapján értelmezhetőek.



122. ábra: Jégcsap diagram

A jégcsap diagramot alulról felfelé kell értelmezni, mely vizuálisan szemlélteti milyen sorrendbe vonta össze a program a klasztereket. Az első a „Az első „Közösségi ismertségi oldal (Facebook)”, mivel itt a köztük lévő leghosszabb vonal” Az összevonás következtében 7 vonal van.

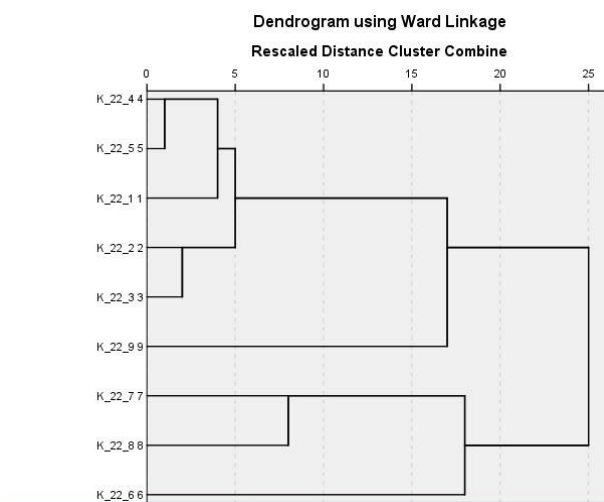


123. ábra: Dendrogram

A fenti Dendrogramot balról jobbra indulva kell értelmezni. Megjelenik az összevont egység és melyek különülnek el leginkább. Ebben az esetben elkülönül a „Az első a „Közösségi ismertségi oldal (Facebook)”, mely már az előző módszernél is jelentkezett.

22_1	RSS olvasó
22_2	Linkmegosztó szolgáltatások hivatkozásmegosztót (pl. delicious, diigo)
22_3	Fogalomtérkép készítő alkalmazást (MindMap, MindMeister)
22_4	Mikroblog szolgáltatás (twitter)
22_5	Jegyzetelő alkalmazás (Posterous)
22_6	Az első a „Közösségi ismertségi oldal (Facebook)
22_7	Szakmai közösségi oldal, referenciaoldal (LinkedIn)
22_8	Online tárhely megosztható tartalommal (Box.net, DropBox.net)
22_9	Közösségi játékok

A Wards módszerrel, mely a négyzetes euklédieszi távolság, a szórásnégyzet alapján kalkulál. Azokat a csoportokat vonja össze, amelyek a klaszteren belüli szórásnégyzetet a legkevésbé növelik és ezáltal a legkisebb távolságú csoportok egyesülnek.



124. ábra: Dendrogram Ward módszerrel

A klaszteranalízis **végeredményeként** megállapítható, hogy vizsgálatunkban kettő klasztert hozhatunk létre.

10.4 ÖSSZEFOGLALÁS, KÉRDÉSEK

10.4.1 Összefoglalás

A kérdőívek eredményeit kiértékelő eljárásoknak kell alá vetni. A fejezetben az olvasó a klaszteranalízissel és azon belül két módszer gyakorlati alkalmazásával, az eredmények értelmezésével ismerkedett meg. Az SPSS statisztikai szoftverrel lépései alapján képessé vált a módszer kiválasztására a vizsgálat megvalósítására.

A fejezet tanulmányozása során a hallgató képessé válik a teszt empirikus adatainak értékelésére és a következtetések levonására.

10.4.2 Önellenőrző kérdések

1. Értelmezze a klaszteranalízis fogalmát, alkalmazási területeit!
2. Sorolja fel a klaszteranalízis lépéseit

11. FAKTORELEMZÉS

11.1 CÉLKITÚZÉSEK ÉS KOMPETENCIÁK

A fejezet a többváltozós populációk statisztikai elemzési módszerével ismerteti meg négy alfejezetben az olvasót, az alábbiakban felsoroltak alapján:

A tananyag tanulmányozása segíti Önt abban, hogy a többváltozós populációk hogyan elemezhetők. Az olvasó megismeri háttérváltozók és a közöttük meglévő kapcsolatok feltárását és a jelenségek magyarázatát biztosító statisztikai módszer lényegét.

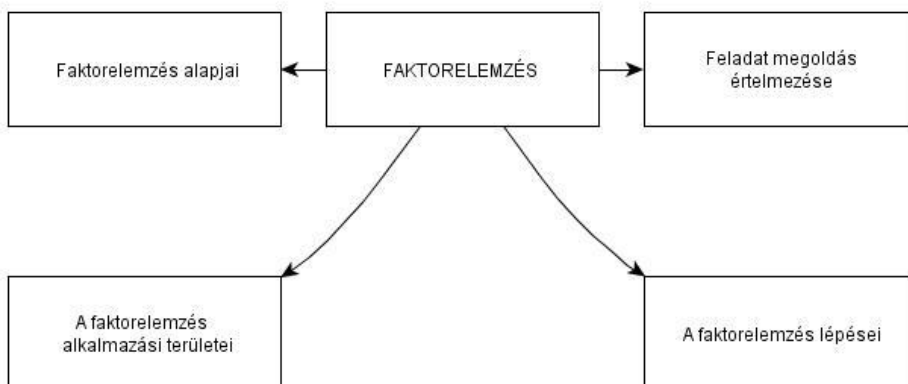
11.2 TANANYAG

Faktorelemzés alapjai

A faktorelemzés alkalmazási területei

A faktorelemzés lépései

Feladat megoldás értelmezése




125. ábra: Fogalomtérkép

11.2.1 Faktorelemzés alapjai¹⁰

Mint a Faktoranalízis és a Főkomponens-elemzés adatredukcióra és adatstruktúra feltárására szolgáló módszer. Az eljárás során cél az elemzésbe bevont változók számának csökkentése és a változók közt feltárható struktúrák azonosítása. Az eljárás a változók osztályozását eredményezi.

Az elemzések során gyakran kettőnél több változót kell figyelembe venni az adott probléma megoldása során. Több változónak nagy elemszámú mintán történő mérése során óriási adathalmazt egy egységként kezelni bonyolult feladat. A kapcsolatok feltárásánál több, egymástól is függő változó kapcsolat lehetőségét elemezve kell a feladatot megoldani, melynek elemzése és az eredmények értelmezése a faktoranalízis segítségével történhet (Székelyi-Barna, 2002). A faktorelemzés célja a struktúra feltárása és az adatok mennyiségének csökkentése

Adott: egy sokváltozós mintaállomány, ahol a faktorok korrelálatlanok és a vizsgálat kezdetén még nem ismertek. A faktoranalízist a regresszióanalízistől az különbözteti meg, hogy a független változók ismertek. Egy adatállományon a **faktoranalízis** csak akkor végezhető el, ha az adatok összefüggnek, más szóval korreláltak, melynek értelmében a változók redundáns információkat hordoznak.

 **A faktoranalízis a változók száma csökkentésének a legelterjedtebb módszere. A jelenség feltárását szolgáló vizsgálati módszer, amelyek a mért változók háttérében lehetnek, egymástól függenek és a jelenségekre magyarázatot adnak.**

A változók számának csökkentése a statisztikai mintában a lévő információ lehetőség csökkentésével ugyanazt a jelenséget írja kevesebb változóval. A **feladat** a sokváltozós adatállomány jellemzése a változónál kisebb számú célszerűen választott ún. faktorra oly módon, hogy a faktorok az eredeti lehetőség szerinti legtöbb információt tartalmazzák és az így azonosított faktorokat célszerű értelmezni és elnevezni, melyek az eljárás kezdetén ismeretlenek. **Másik fontos célkitűzés**, hogy a nagyszámú változó közötti korrelációs struktúrát írjunk le kevés számú látens változó, ún. faktor segítségével. A faktoroknak fizikai jelentésük nincs, közvetlenül nem megfigyelhetők, nem mérhetők és létezésük csak elképzelhető az eredeti változók alapján. A változók száma csökkentésének a legelterjedtebb módszere a **faktoranalízis**. Feladata az adatrendszer struktúrájának feltárása és a többváltozós adatrendszerek elemzése.

¹⁰ Tóthné Parázsó Lenke: Kutatásmódszertan matematikai alapjai: Eger, 2011. ISBN 978-615-5221-25-5 pp.70-75

A faktoranalízis alapfeltevése, hogy ezeket a látens változókat nem tudjuk megfigyelni, de a minta által adott változók révén kell azokra következtetni. A faktoranalízis során a faktorok meghatározása a vizsgált változók korrelációs mátrixából kiindulva:

- Ha a változó nem korrelál más változókkal, nagy valószínűséggel önálló faktorral rendelkezik.
- Ha két vagy több változó között szoros a korreláció, akkor feltételezhető, hogy egy vagy több közös faktorral rendelkeznek.

A faktoranalízist alkalmazási feltételei:

- ha a korrelációs mátrix alapján a változók úgy csoportosíthatóak, hogy az egy csoporton belüli változók között viszonylag magas a korreláció, ezzel szemben a csoportok között pedig alacsony. (Egy ilyen csoport olyan, mely mögött egy faktor áll.
- a parciális korrelációk kicsik,
- a **Kaiser-féle mutatószám** (0 és 1 közé eső érték) az adatok összefüggő voltának, korreláltság vizsgálatának módszere, amelyet **Kaiser-Meyer-Olkin statisztikának** is neveznek, Ha ez a mutatószám **0.5-nél nagyobb**, akkor **ajánlott**, ha ez a mutatószám viszont **0.5-nél kisebb**, akkor **nem ajánlott** faktoranalízis végrehajtása. A faktoranalízis egyaránt támaszkodhat a kovariancia illetve a korrelációs mátrix elemzésére. Kaiser-Meyer-Olkin mérték az alábbi képlet alapján határozható meg:

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p \rho_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}$$

Ahol:

r_{ij} - az i-edik és a j-edik változók korrelációs együtthatója,

$\rho_{ij} = \frac{R_{ij}}{\sqrt{R_{ii} \cdot R_{jj}}}$ - az i-edik és a j-edik változó parciális korreláció együt-

thatója,

- ha a KMO értéke $\geq 0,5$ abban az esetben az adatok alkalmasak a faktoranalízisre,

- ha a KMO értéke $< 0,5$ abban az esetben az adatok nem alkalmasak a faktoranalízisre.

11.2.2 A faktorelemzés alkalmazási területei:

- A nagyszámú és egymással korreláló változó között tanulmányozhatjuk a kapcsolatokat úgy, hogy a változókat kisebb számú un. faktorokba rendezzük, amelyekben belül a korrelációk nagyobbak, mint ezeken kívül.
- A faktorok a hozzájuk tartozó változók alapján értelmezhetőek.
- A faktoranalízis segítségével a nagyszámú populáció a kisebb számú faktorok a faktor-pontok segítségével mennyiségileg áttekinthetőbbé válik.

A faktormodell fogalma, felépítése

Meghatározza, hogyan függnek az egyes változók a faktoroktól, mely lineáris kombinációval állíthatók elő. Tehát a főkomponens analízissel szemben, ahol az egyes főkomponenseket állítottuk elő az eredeti változók lineáris kombinációjaként, itt az egyes változók fejezhetőek ki a faktorok lineáris függvényeként. Fontos tudni, hogy faktoranalízist többféle módszerrel hajthatunk végre, a legfontosabbak ezek közül a főkomponens módszer, a főfaktor analízis és a maximum likelihood faktoranalízis.

A faktor számának megválasztása

A faktoranalízis az adatrendszer belső struktúráját, az adatrendszer egészét látva egyenrangúnak tekinti a változókat. A faktoranalízis célja a jelenséget leíró változók „mögött” megkeresni olyan rejtett változókat, amelyek a vizsgált jelenséget megmagyarázzák, számuk kisebb, mint az eredeti változóké és egymástól függetlenek.

A faktoranalízis során a faktorok meghatározásakor a vizsgált változók korrelációs mátrixából kell kiindulni. Amelyik változó nem korrelál más változókkal, nagy valószínűséggel önálló faktorral rendelkezik. Ha viszont két vagy több változó között szoros korreláció van, akkor feltételezhető, hogy egy vagy néhány közös faktorral rendelkeznek.

A faktoranalízis modelljében a következő faktorokat különböztethetők meg:

- közös faktor (több változót befolyásol),
- általános faktor (az összes változóra hatással van),
- csoport faktor (nem az összes változót befolyásolja),
- egyedi faktor (csak egyetlen változót befolyásol),
- hiba faktor (mérési, becslési hiba hatása).

Egy-egy változót eltérő súllyal befolyásolhatják a különböző faktorok, másrészt egy faktor eltérő súllyal befolyásolja az egyes változók értékét.

Az eredeti változók helyett meghatározott hipotetikus változók, ún. faktorok tartalmazzák a rendszerről ismert információk nagy részét annak ellenére, hogy számuk kisebb. A faktoroknak nincs semmilyen fizikai jelentésük, közvetlenül nem figyelhetők meg, nem mérhetők, létezésüket csak feltételezhetjük az eredeti változók kapcsolatai alapján. A változók számának csökkentése azt jelenti, hogy a statisztikai mintában lévő információ lehetőleg kis csökkentésével ugyanazt a jelenséget kevesebb változóval írjuk le.

A különböző faktorok hatásainak figyelembevételével az X változó az alábbiak szerint írható fel:

$$X_i = a_{i1} \cdot F_{i1} + a_{i2} \cdot F_{i2} + \dots + a_{iq} \cdot F_{iq} + b_{im} \cdot F_{im} + e_i \cdot F_i$$

ahol:

- a: a közös faktorok súlya
- b: az egyedi faktorok súlya
- c: a hiba faktorok súlya

A feltételezés alapján a hibakomponens korrelálatlan a közös, illetve az egyedi faktorokkal, valamint, hogy a hibakomponensek függetlenek.

A standartizált változó szórásnégyzete:

$$s = \sum_{j=1}^q a_{ij}^2 + b_{im}^2 + e_i^2 = 1$$

A megfigyelt értékek mátrixa, mely a faktoranalízis bemeneti (input) adathalmazának tekintendő:

$$X = \begin{bmatrix} x_{11}, x_{12}, \dots, x_{1p} \\ \dots, \dots, \dots, \dots, \\ x_{n1}, x_{n2}, \dots, x_{np} \end{bmatrix}$$

ahol:

- p: a változók száma
- n: a mintaelemek száma

11.2.3A faktorelemzés lépései

- Minden változóra meg kell határozni az átlagot és a korrigált tapasztalati szórást.
- Minden adatból ki kell vonni a változókhoz tartozó adatok átlagát.
- Az eredményt el kell osztani a korrigált tapasztalati szórással.
- A feladat megoldása során olyan új F_1, F_2, \dots, F_k valószínűségi változókat kell keresni, ahol az F_k faktorok közös jellemzői:
- Számuk maximum p ,
- Normális eloszlásúak
- Korrelálatlanok (bármely kettő korrekciós együtthatója zérus)

A fenti mátrixból az X_i valószínűségi változók és a faktorok közötti kapcsolatot az alábbiak alapján képezhetők:

$$\begin{aligned} X_1 &= a_{11} \cdot F_1 + a_{12} \cdot F_2 + \dots + a_{1k} \cdot F_k + W_1 \\ X_2 &= a_{21} \cdot F_1 + a_{22} \cdot F_2 + \dots + a_{2k} \cdot F_k + W_2 \\ &\dots\dots\dots \\ X_p &= a_{p1} \cdot F_1 + a_{p2} \cdot F_2 + \dots + a_{pk} \cdot F_k + W_p \end{aligned}$$

ahol:

W_1, W_2, \dots, W_p : egyedi faktorok, mivel egyenként csak egy változó kifejezésében szerepelnek

F_1, F_2, \dots, F_k : közös faktorok

$W - k$ és $F - k$ korrelálatlanok egymással. A W értékétől függően, ha $W -$ értéke nagy, a faktoranalízis nem sikeres, ha $W -$ értéke kicsi, abban az esetben jó eredményt kaptunk.

a_{1j} – a faktorsúly, amely azt fejezi ki, hogy, az F_1 faktor milyen súllyal szerepel az X_1 meghatározásában.

Tekintsük át a faktoregyütthatók és a faktorsúlyok között a különbségét:


- A faktoregyütthatók a faktorok együtthatói a faktormodellben, melyek a megfelelő változó és faktor közötti korreláció nagyságát mérik.
- A faktorsúlyok ezzel szemben azt mondják meg, hogy mennyi a bevezetett új, közös faktorok értéke az egyes megfigyeléseknél.

A kommunalitás értelmezése

A kommunalitás a szórásnégyzetben a faktorok hatását mutató rész, melynek maximális értéke 1.

$$h_i^2 = \sum_{j=1}^q a_{ij}^2$$

$$h_i = a_{i1}^2 + a_{i2}^2 + \dots + a_{ik}^2$$

 **A kommunalitás alatt a közös faktorsúlyok hatása értendő. A bevezetett faktoroknak az eredeti változó szórásának százalékban megvalósított értékelését mutatja. Minél nagyobb a kommunalitás (maximum 1 lehet), annál jobb a választott faktormodell.**

Abban az esetben, ha a kommunalitás értéke közel van az 1-hez, a kommunalitás jól magyarázza és írja le az adott változót, vagyis arra ad választ, hogy a faktorok az adott változók varianciájának hány %-át értelmezi. Példaként említhető a faktorsúly +1-hez vagy -1-hez közeli értéke, melynek során x_i és F_i változók között erős pozitív vagy negatív korreláció áll fenn.

Fontos tényező a faktorok sajátértékére rámutatni, mely egy adott faktorthoz tartozik, mely matematikailag az adott faktor összes faktorsúlyának négyzetösszegével egyenlő. Az alábbi képlet rámutat arra, hogy a változórendszer teljes varianciájának magyarázatában az F_i milyen súllyal vesz részt.

$$F_i = a_{1i}^2 + a_{2i}^2 + \dots + a_{pi}^2$$

Matematikailag az un. „fontos” faktorok saját értéke nagy, míg a „kevésbé fontosaké” kicsi.

A faktorok rotációjának értelmezése

A faktorok rotációja során a nehezen értelmezhető faktorok egyszerűbbé tehetők. A rotációval kapott változók, melyek az új faktorokra nézve is az eredeti változók, még nagyobbaknak kell lenniük. Ezzel ellentétben a korábbi kis faktorsúlyú változók még kisebbekké válnak. Az eljárással kapott faktorszerkezet könnyebben értelmezhető és a legjellemzőbb változók alapján elnevezhető.

Egy ortogonális mátrixszal transzformálva mind a faktoregyüttható mátrixot, mind pedig a faktorokat, egy új, a régivel teljesen egyenértékű modellt eredményez. A forgatást a faktorok könnyebb interpretálhatóságára használják.

Ennek eredményeként a faktoregyütthatók értékei a 0-hoz vagy az 1-hez lesznek közel. Így könnyebben meghatározható, hogy az egyes faktorok mely változócsoporthoz tartoznak

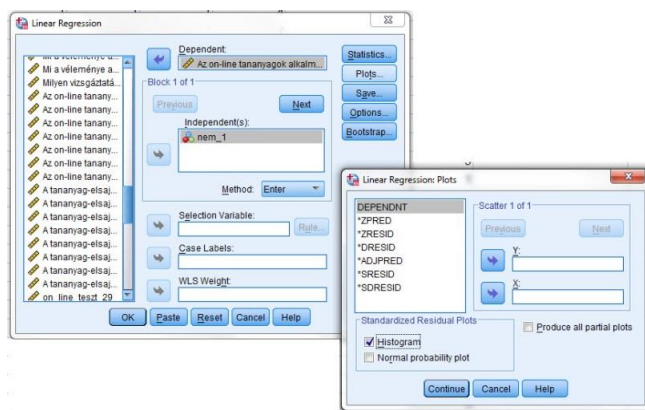
Az eredeti és a faktorváltozók közötti korrelációs együtthatók értéke rámutat, hogy az adott faktorok elsősorban mely változókkal állnak szorosabb kapcsolatban.

A faktoranalízis során a faktorsúlyok mátrixát kell előállítani. A mátrix és saját transzponáltjának szorzata egyenlő a korrelációs mátrix és hibák variancia-kovariancia mátrixának különbségével. Mivel a hibakomponensek függetlenek, ezért ez utóbbi mátrix diagonális, vagyis gyakorlatilag egy olyan módosított korrelációs mátrixot eredményeznek, ahol a főátlóban lévő elemek a kommunalításokkal lettek kicserélve, a mátrix többi elemeként pedig az r_{ij} korrelációs együttható maradt.

A faktoranalízis lépései

- A korrelációs mátrix meghatározása.
- A parciális korrelációs mátrix meghatározása.
- A minta faktoranalízisre való alkalmasságát mérő Kaiser statisztika kiszámítása.
- A kovariancia (korrelációs) mátrix sajátértékeinek és sajátvektorainak meghatározása.
- A sajátértékek szemléltetése törmelék grafikonnal.
- A faktoregyütthatók, mint az egyes változók és a faktorok közötti korrelációk meghatározása.
- Kommunalítások megadják, hogy az egyes faktorok a teljes szórásnak hány százalékát magyarázzák.
- A faktoregyütthatók grafikonja. A változók ábrázolása a faktortérben.
- A faktorok forgatása. A forgató mátrix és a forgatás utáni faktoregyütthatók meghatározása.
- Kommunalítások a forgatás után.
- A (standardizált) faktorsúlyok meghatározása.
- A forgatás utáni faktoregyütthatók grafikonja, a változók ábrázolása a forgatott faktortérben.

A lehulló ablakban beállítjuk

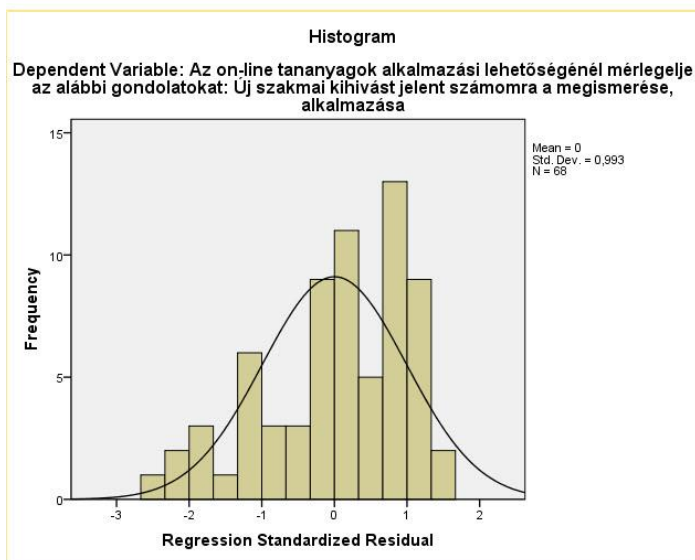


127. ábra: *Analyze/Regression/Linear histogram beállítása*

A kapott histogram az adatok normális eloszlását mutatja, melyet a többi változó esetében is szükséges megismételni.

Charts

A változó eloszlása normális

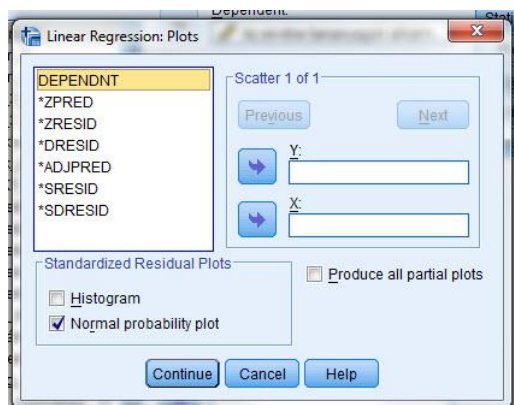


128. ábra: *Analyze/Regression/Linear histogramja*

Ezt követően vizsgáljuk meg az adatsor homoszkedaszticitását¹¹

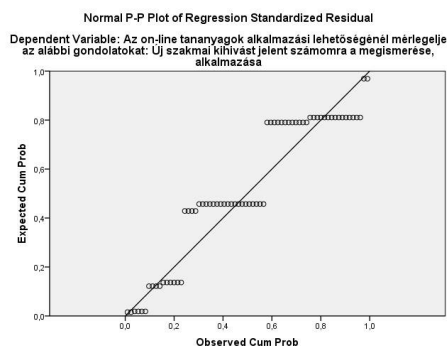
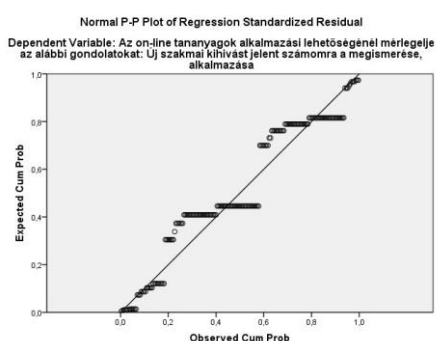
Analyze/regression/Linear/Plots..../Scatterdot...

A lehulló menüben pipáljuk ki a



129. ábra: Analyze/Regression/Linear Scatterdot beállítása

Eredményül az alábbi pontdiagramot kapjuk, melynek alapján megállapítható, hogy homoszkedaszticitásuk, azaz változók azonos szórásnégyzetűek.



130. ábra: Analyze/Regression/Linear Scatterdot

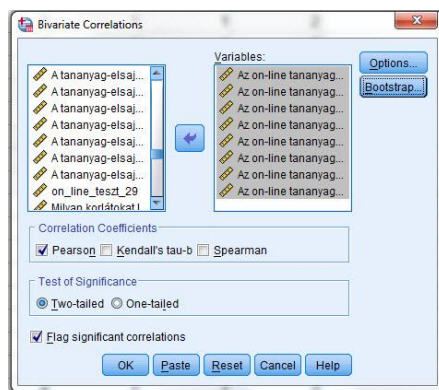
A Multikollinearitás (korrelációs mátrix) vizsgálata

Két vagy több változó között a korrelációs együttható meghaladja a 0,7-s értéket. Ebben az esetben az egyik változó értékeinek módosulása meghaladja

¹¹ homoszkedaszticitás változók azonos szórásnégyzete

értékeinek változását. Abban az esetben, ha a korrelációs együttható nagyobb, mint 0,7 abban az esetben érdemes változókat kihagynia modellből.

Analyze/Correlate/Bivariate



131. ábra: *Ábra. Analyze/Correlate/Bivariate lehulló ablak beállítása*

A fenti multikollinearitás vizsgálatból látható, hogy 0,7 korrelációs értéket nem haladnak meg a változók, tehát alkalmas a további elemzésre.

		Az on-line tananyagok alkalmazási lehetősége mérlegelje az alábbi gondolatokat: Uj szamaa kihivást jelent számomra a megismerése, alkalmazása	Az on-line tananyagok alkalmazási lehetősége mérlegelje az alábbi gondolatokat: A tananyaghoz és kapcsolódó anyagokhoz való elérés lehetősége a könnyebb	Az on-line tananyagok alkalmazási lehetősége mérlegelje az alábbi gondolatokat: A hallgatók zömea körül tájékoztatása a lehetőségeit biztosítja	Az on-line tananyagok alkalmazási lehetősége mérlegelje az alábbi gondolatokat: Követlen tananyag elérése lehetősége	Az on-line tananyagok alkalmazási lehetősége mérlegelje az alábbi gondolatokat: Tanári e-tananyag segítségével a felkészülésbe n	Az on-line tananyagok alkalmazási lehetősége mérlegelje az alábbi gondolatokat: Független az órádól és a tanár elemzésétől	Az on-line tananyagok alkalmazási lehetősége mérlegelje az alábbi gondolatokat: Lehetséget nyújt az oktatásban részvétel számára, hogy férfi és időséi függetlenül oldják meg a lespécifikusabb problémákat	Az on-line tananyagok alkalmazási lehetősége mérlegelje az alábbi gondolatokat: Lehetséget ad elektronikus n jelezni a tanulás problémáit
Az on-line tananyagok alkalmazási lehetősége mérlegelje az alábbi gondolatokat: Uj szamaa kihivást jelent számomra a megismerése, alkalmazása	Pearson Correlation	1	,333 ^{**}	,231 [*]	,292 [*]	,292 [*]	,407 ^{**}	,239	,237
	Sig. (2-tailed)		,006	,039	,016	,020	,001	,051	,056
N		68	68	68	68	68	68	67	66
Az on-line tananyagok alkalmazási lehetősége mérlegelje az alábbi gondolatokat: A tananyaghoz és kapcsolódó anyagokhoz való elérés lehetősége a könnyebb	Pearson Correlation	,333 ^{**}	1	,550 ^{**}	,592 ^{**}	,241 [*]	,472 ^{**}	,479 ^{**}	,340 [*]
	Sig. (2-tailed)	,006		,000	,000	,048	,000	,000	,005
N		68	68	68	68	68	68	67	66
Az on-line tananyagok alkalmazási lehetősége mérlegelje az alábbi gondolatokat: A hallgatók zömea körül tájékoztatása a lehetőségeit biztosítja	Pearson Correlation	,251 [*]	,550 ^{**}	1	,414 ^{**}	,176	,176	,412 ^{**}	,324 ^{**}
	Sig. (2-tailed)	,039	,000		,000	,152	,151	,001	,008
N		68	68	68	68	68	68	67	66
Az on-line tananyagok alkalmazási lehetősége mérlegelje az alábbi gondolatokat: Követlen tananyag elérése lehetősége	Pearson Correlation	,292 [*]	,592 ^{**}	,414 ^{**}	1	,419 ^{**}	,565 ^{**}	,571 ^{**}	,318 [*]
	Sig. (2-tailed)	,016	,000	,000		,000	,000	,000	,010
N		68	68	68	68	68	68	67	66
Az on-line tananyagok alkalmazási lehetősége mérlegelje az alábbi gondolatokat: Tanári e-tananyag segítségével a felkészülésbe n	Pearson Correlation	,292 [*]	,241 [*]	,176	,419 ^{**}	1	,373 ^{**}	,428 ^{**}	,464 ^{**}
	Sig. (2-tailed)	,020	,048	,152	,000		,002	,000	,000
N		68	68	68	68	68	68	67	66
Az on-line tananyagok alkalmazási lehetősége mérlegelje az alábbi gondolatokat: Független az órádól és a tanár elemzésétől	Pearson Correlation	,407 ^{**}	,472 ^{**}	,176	,565 ^{**}	,373 ^{**}	1	,519 ^{**}	,345 ^{**}
	Sig. (2-tailed)	,001	,000	,151	,000	,002		,000	,005
N		68	68	68	68	68	68	67	66
Az on-line tananyagok alkalmazási lehetősége mérlegelje az alábbi gondolatokat: Lehetséget nyújt az oktatásban részvétel számára, hogy férfi és időséi függetlenül oldják meg a lespécifikusabb problémákat	Pearson Correlation	,239	,479 ^{**}	,412 ^{**}	,571 ^{**}	,428 ^{**}	,519 ^{**}	1	,603 ^{**}
	Sig. (2-tailed)	,051	,000	,001	,000	,000	,000		,000
N		67	67	67	67	67	67	67	65

132. ábra: *Analyze/Correlate/Bivariate eredménytáblázat*

1. A minta elemszámának (68) és a változók számának (7) há-

$$\text{nyadosa: } \frac{190}{7} = 27$$

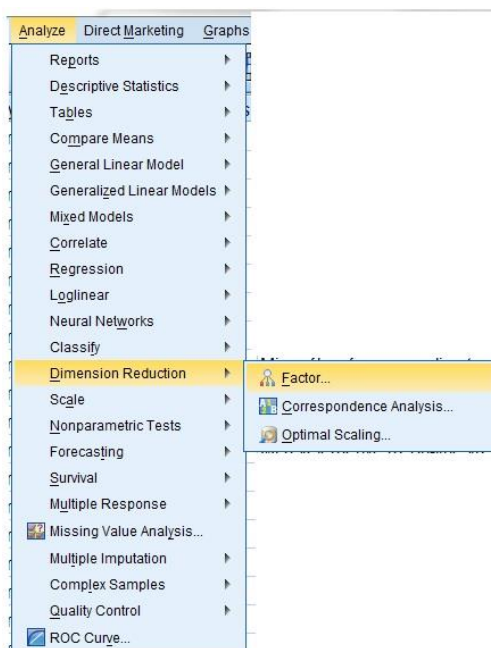
A válaszadók száma több mint 10 szeres, mely a feltételeknek megfelel.

2. Korrelációs mártix

A faktorelemzés feltétele a változók között kimutatható korreláció megléte, mivel ennek hiányában nem lehet faktorokba a változókat összevonni.

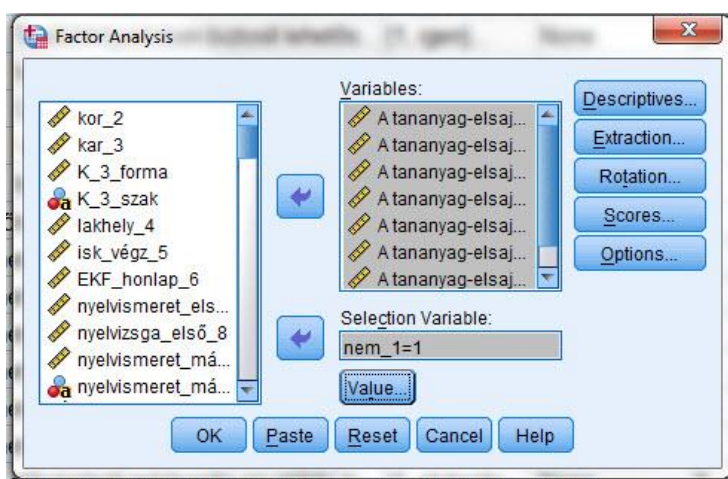
- ha a korreláció értéke alacsony, közel áll a nullához, abban az esetben nincs kapcsolat,
- ha az értéke magas, megközelíti az egyet, nem lesz megoldás, minden változó egy faktorba kerül.

A korrelációs mártix az Analyze/Dimension reduction/Factor/Descriptives menüben állítható elő.



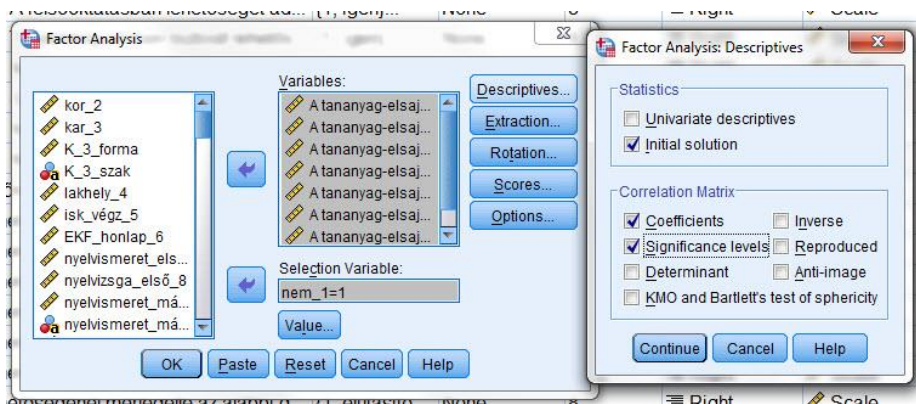
133. ábra: Analyze/Dimension reduction/Factor/Descriptives

A lehulló ablakban, a Descriptives menüben behúzzuk a változókat és kiválasztjuk a változót, amely alapján szelektálni kívánunk.



134. ábra: *Analyze/ Dimension reduction/Factor/Descriptives*

A Descriptives ablakban kipipáljuk az Initial solution, Coefficients, Significance levels rádiógombokat.



135. ábra: *Analyze/ Dimension reduction/Factor/Descriptives lehulló menü beállítása*

Az OK gomb lenyomásával a kapott táblázat eredményeiből megállapítható, hogy a a legmagasabb szignifikanciaszint 0,538.

Correlation Matrix^a

	Az on-line tananyagok alkalmazási lehetőségénél mérlegelje az alábbi gondolatokat: Új szakmai kihívást jelent számomra a megismerése, alkalmazása	Az on-line tananyagok alkalmazási lehetőségénél mérlegelje az alábbi gondolatokat: A tananyaghoz és kiegészítő anyagokhoz való elérési lehetősége a könnyebb	Az on-line tananyagok alkalmazási lehetőségénél mérlegelje az alábbi gondolatokat: A hallgatók széles körű tájékoztatásának lehetőségét biztosítja	Az on-line tananyagok alkalmazási lehetőségénél mérlegelje az alábbi gondolatokat: Kötetlen tananyag elérési lehetősége	Az on-line tananyagok alkalmazási lehetőségénél mérlegelje az alábbi gondolatokat: Tanárai e-tananyaga segít a felkészülésben	Az on-line tananyagok alkalmazási lehetőségénél mérlegelje az alábbi gondolatokat: Függetlenül óráktól és a tanár elérhetőségétől	Az on-line tananyagok alkalmazási lehetőségénél mérlegelje az alábbi gondolatokat: Lehetőséget nyújt az oktatásban résztvevők számára, hogy térítő és időtől függetlenül oldják meg a legspecifikusabb problémákat	Az on-line tananyagok alkalmazási lehetőségénél mérlegelje az alábbi gondolatokat: Lehetőséget ad elektronikusan jelezni a tanárnak a felmerülő problémákat
Correlation	1,000	,157	,331	,094	,238	,140	,163	,289
Az on-line tananyagok alkalmazási lehetőségénél mérlegelje az alábbi gondolatokat: Új szakmai kihívást jelent számomra a megismerése, alkalmazása		1,000	,442	,330	,338	,263	,385	,254
Az on-line tananyagok alkalmazási lehetőségénél mérlegelje az alábbi gondolatokat: A tananyaghoz és kiegészítő anyagokhoz való elérési lehetősége a könnyebb			1,000	,324	,397	,147	,375	,538
Az on-line tananyagok alkalmazási lehetőségénél mérlegelje az alábbi gondolatokat: A hallgatók széles körű tájékoztatásának lehetőségét biztosítja				1,000	,200	,298	,502	,450
Az on-line tananyagok alkalmazási lehetőségénél mérlegelje az alábbi gondolatokat: Kötetlen tananyag elérési lehetősége					1,000	,358	,295	,287
Az on-line tananyagok alkalmazási lehetőségénél mérlegelje az alábbi gondolatokat: Tanárai e-tananyaga segít a felkészülésben						1,000	,290	,296

136. ábra: *Analyze/ Dimension reduction/Factor/Descriptives eredménytábla*

Ez azt is jelenti, hogy az on-line tananyagok alkalmazási lehetőségei között „A hallgatók széleskörű tájékozódásának lehetőségét biztosítja” és a „Lehetőséget ad elektronikusan jelezni a tanárnak a felmerülő problémákat”, ez a két változó egy faktorban fog szerepelni.

3. Anti-Image matrix

Az anti-image matrix amely a változók szórásnégyzete. Két részből tevődik össze a magyarázott szórásnégyzet (image) és a nem magyarázott szórásnégyzet (anti-image)

Ez a felbontás a faktorelemzés során az anti-image kovariancia, más szóval a korrelációs mátrixok mutatják. Az átlóban szereplő értékek az magyarázott szórásnégyzet (image) alacsonyak, az átlóban szereplő értékek 1-hez közeli értékek (anti-image).

Az Anti -Image az Analyze/Dimension Reduction/Factor/Descriptives menüben állítható elő.

		Új szakmai kihívást jelent számomra a megismerése, alkalmazása	A tananyaghoz és kiegészítő anyagokhoz való elérési lehetősége a könnyebb	A hallgatók széles körű tájékoztatásának lehetőségét biztosítja	Kötetlen tananyag elérési lehetősége	Tanári e-tananyag segít a felkészülésben	Független az óráktól és a tanár elérhetőségétől	Magam oktatásban résztvevők számára, hogy tőlük és időtől függetlenül oldjak meg a legspecifikusabb problémákat	Lehetőséget ad elektronikus úton jelezni a tanárnak a felmerülő problémákat
Anti-image Covariance	Új szakmai kihívást jelent számomra a megismerése, alkalmazása	,750	,020	-,174	,041	-,053	-,169	-,053	-,063
	A tananyaghoz és kiegészítő anyagokhoz való elérési lehetősége a könnyebb	,020	,492	-,247	-,082	-,133	-,173	-,077	,010
	A hallgatók széles körű tájékoztatásának lehetőségét biztosítja	-,174	-,247	,673	-,112	,062	,132	,022	-,039
	Kötetlen tananyag elérési lehetősége	,041	-,082	-,112	,735	-,012	-,043	-,113	-,071
	Tanári e-tananyag segít a felkészülésben	-,053	-,133	,062	-,012	,735	-,021	-,091	,108
	Független az óráktól és a tanár elérhetőségétől	-,169	-,173	,132	-,043	-,021	,592	-,140	,001
	Magam oktatásban résztvevők számára, hogy tőlük és időtől függetlenül oldjak meg a legspecifikusabb problémákat	-,053	-,077	,022	-,113	-,091	-,140	,510	-,200
	Lehetőséget ad elektronikus úton jelezni a tanárnak a felmerülő problémákat	-,063	,010	-,039	-,071	-,108	,001	-,200	,680
Anti-image Correlation	Új szakmai kihívást jelent számomra a megismerése, alkalmazása	,814*	,033	-,245	,056	-,071	-,254	-,086	-,088
	A tananyaghoz és kiegészítő anyagokhoz való elérési lehetősége a könnyebb	,033	,772*	-,429	-,136	-,221	-,321	-,153	,018
	A hallgatók széles körű tájékoztatásának lehetőségét biztosítja	-,245	-,429	,638*	-,159	,088	,209	,038	-,058
	Kötetlen tananyag elérési lehetősége	,056	-,136	-,159	,887*	-,016	-,066	-,184	-,100
	Tanári e-tananyag segít a felkészülésben	-,071	-,221	,088	-,016	,872*	-,032	-,149	-,153
	Független az óráktól és a tanár elérhetőségétől	-,254	-,321	,209	-,066	-,032	,779*	,255	,001
	Magam oktatásban résztvevők számára, hogy tőlük és időtől függetlenül oldjak meg a legspecifikusabb problémákat	-,086	-,153	,038	-,184	-,149	-,255	,830*	-,340
	Lehetőséget ad elektronikus úton jelezni a tanárnak a felmerülő problémákat	-,088	,018	-,058	-,100	-,153	,001	-,340	,840*

137. ábra: *Analyze/Dimension Reduction/Factor/Descriptives anti-image eredménytábla*

A táblázat két részre bontható a felső része az anti-image kovariancia, az alsó fele az anti-image korrelációs mátrix.

A táblázat felső részén, az anti-image kovariancia átlón felüli elemei azok, amelyek a varianciák alapján függetlenek a többi változótól, értékei alacsonyak. Megfigyelhető, hogy az alacsony értékek (0,09 értékek) minimum az esetek 75%-ban jellemzőek, ezekben az esetekben a variancia értékek függetlenek a többi értékektől, tehát nincs szoros kapcsolat az adatok között.

Az MSA (Measuring of Sampling Adequacy) a változók közötti kapcsolat szorosságát jelenti. Az MSA értéke 0 és 1 közötti értéket vehet fel.

Az MSA 0,5 alatti értékeit ki kell zárni a további elemzésből. Ha az MSA értéke 1, hogy az adott változót hibától függetlenül becsli. Maradnak a közöttek

értékek, melyek a 0,638 és 0,830 értéket vesznek fel. Az átlóban lévő értékek a fontosak, ezek tartalmazzák az változókra érvényes MSA értékeket.

4. Bartlett teszt

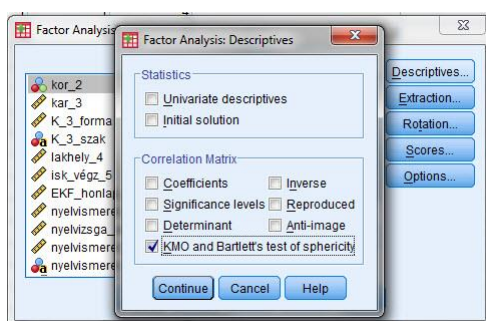
A Bartlett teszt vizsgálata arra irányul, hogy a mintában korrelációs mátrix főátlón kívüli elemei véletlenül, vagy igazoltan térnek el a nullától. A faktoranalízis feltétele, hogy a változók korreláljanak, vagyis a H_1 hipotézist kell igazolni, mely a KMO-val történik.

Hipotézisek:

H_0 : nincs korreláció

H_1 : van korreláció

Az Bartlett az Analyze/Dimension Reduction/Factor/Descriptives menüben állítható elő szintén. A lehulló ablakban



138. ábra: Analyze/Dimension Reduction/Factor/Descriptives KMO beállítása

KMO and Bartlett's Test ^a		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,802
Bartlett's Test of Sphericity	Approx. Chi-Square	213,936
	df	28
	Sig.	,000

a. Only cases for which nem_1 = nő are used in the analysis phase.

139. ábra: Analyze/Dimension Reduction/Factor/Descriptives eredménytábla

A KMO kritérium Kaiser-Meyer-Olkin kritérium, amely olyan mérőszám, amely megmutatja mennyire alkalmasak az adott változók a faktorelemzésre.

Mi a különbség KMO és a MSA között?

Az MSA érték az egyes változókra vonatkozik, a KMO az összes változóra. A KMO érték az anti-image átlaga.

KMO mutatószámai

$KMO \geq 0,9$	kiváló
$KMO \geq 0,8$	nagyon jó
$KMO \geq 0,7$	megfelelő
$KMO \geq 0,6$	közepes
$KMO \geq 0,5$	gyenge
$KMO \leq 0,5$	elfogadhatatlan

A fenti táblázatban a $KMO = 0,802$ nagyon jónak felel meg.

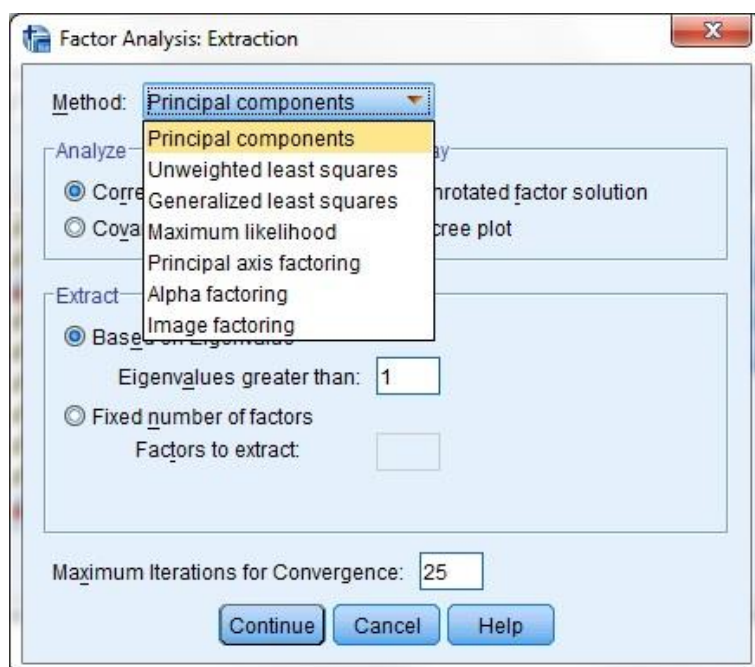
5. *A faktorok módszer kiválasztása*

A faktorok számának meghatározása nagy figyelmet igénylő feladat. Leggyakrabban a Kaiser-kritériumnak megfelelő faktorszámmal kezdődik a vizsgálat. Menetközben meg kell figyelni a kritériumok teljesülését. A helyes faktorszám megválasztásban elsődleges szerepe van a feladatban betöltött értelmezhetőségnek. A következő tényezőket kell figyelembe venni

- faktorszám kevés, fontos dimenziók kimaradhatnak az elemzésből
- sok faktorszám az értelmezésük bonyolulttá válhat

Nézzük meg a gyakorlatban:

Analyze/Dimension reduction/Factor/Extraction menüben a **Method** gombbal vizsgáljuk meg a lehetőséget.

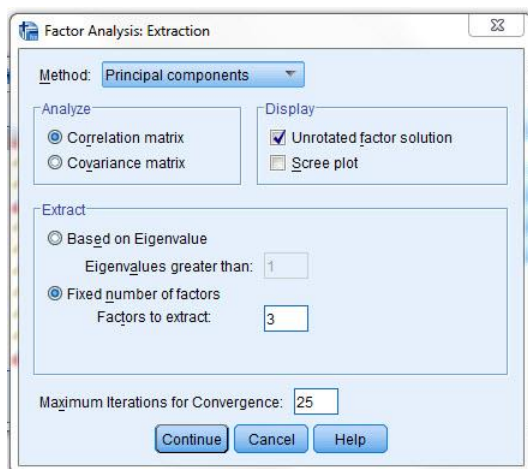


140. ábra: Faktormódszerek

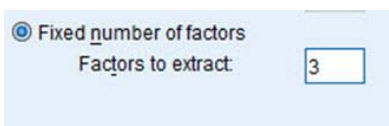
Kiemelve három módszert értelmezzük

- **Principal components** akkor alkalmazható, ha a változók száma magas
- **Unweighted least squares** alkalmazásakor nem kell ismerni a változók eloszlását
- **Image factoring** esetében az anti-image matrix átlójában lévő elemek 1-hez, az azon kívüliek 0-hoz közelítenek
- Ha változók száma magas ajánlott a maximum-lickehod, alfa, image módszer

Elsőként a Principal components által végezzük el a vizsgálatot.



141. ábra: *Principal components*



- Kaiser kritérium, abban az esetben ha 1 alatti a sajátérték, kevesebb információt hordoz a faktor, mint az 1 változó.
- Alkalmazása ajánlott 20-50 változó esetében
- Egy faktor által az összes változó varianciájából történik a magyarázott variancia létrehozása
- Minél több varianciahányados magyarázható, annál több információ marad az elemzés során (Varianciahányad módszer)
- priori információ alapján kerül beírásra

A faktor analízist lefuttatva a Kaiser –kritérium alapján (látens gyöknek is nevezik). az alábbi táblázaton kapott eredményeket elemezzük.

A táblázat két részből tevődik össze az

- Initial Eigenvalues, ez a kiinduló változók csoportja. Ha az értéke 1 alá csökken akkor kevesebb információt hordoz és nem érdemes a továbbiakban használni. Alkalmazása 20-50 változó között a legoptimálisabb.
- Extraction Sums of Squared Loadings a faktorelemzés varianciahányad módszere során kapott értéksorok, vagyis a faktorelemzés következtében

kapott értékek. Alapja a gyakorlati szignifikancián alapul (magas varianciahányad magyarázata maga után vonja a az információ jelentős hányadának megtartását is). A faktorok száma Cumulative% alapján is meghatározható, melyben annyi faktort hozunk létre, hogy egy adott szintet elérjünk. A természettudományos kutatásokban ajánlott a varianciahányad 95%, míg a társadalomtudományban 60% is elfogadott.

Total Variance Explained ^a						
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,244	40,555	40,555	3,244	40,555	40,555
2	1,042	13,022	53,577	1,042	13,022	53,577
3	,938	11,720	65,297	,938	11,720	65,297
4	,814	10,171	75,468			
5	,633	7,908	83,376			
6	,516	6,452	89,828			
7	,493	6,168	95,996			
8	,320	4,004	100,000			

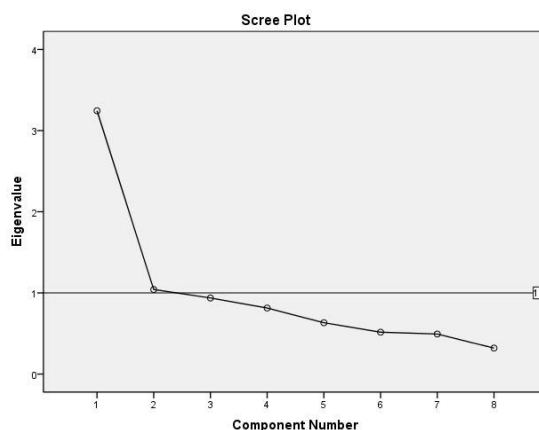
Extraction Method: Principal Component Analysis.

142. ábra: *Principal Component Analysis output*

A Cumulative oszlopban a három faktor összesített varianciáját tüntették fel, amely azt is jelenti, hogy a $100,000 - 65,297 = 34,703\%$ -t elvesztettük.

A Scree Plot ábra a sajátértékek ábrázolása a faktorok sorában.

- y-tengely a sajátérték
- x-tengely a faktorok száma
- 1-nél húzott vízszintes vonal a Kaiser kritérium



143. ábra: *Scree Plot*

A Sree Plot görbe alapján látható, hogy a kapott 8 faktor meredeksége az első háromnál a legnagyobb, majd a csökken. Itt érvényesül a könyökszabály.

Könyökszabály alapján a faktorok száma maximalizálható a görbe meredekségének gyors változásáig, vagyis addig a pontig, ahol a nagy meredekség egy lapos, szinte egyenes szakaszba vált át.

Ebben az esetben a 2 ill. a 3. pontnál következik be és meg kell gondolni, hogy 2 vagy 3 faktorról számolunk.

11.3 ÖSSZEFOGLALÁS

A faktoranalízissel „csökkenthető” a megfigyelt változók száma” vagyis helyesebben változók összevonása valósítható meg az eljárással. A kutatási koncepciók kidolgozásakor a klaszteranalízis ad lehetőséget átfedésmentes csoportosításra. A fentiekben az olvasó megvizsgálta az analízis előfeltételeit és elvégezte a faktoranalízist, értelmezte a kapott paramétereket.

11.4 ÖNELLENŐRZŐ KÉRDÉSEK

1. Ismertesse a faktoranalízis elméleti megfontolásait.
2. Ismertesse a faktoranalízis lépéseit.
3. Mely jelenségek statisztikai feltárását szolgálja a faktoranalízis?

12. ÖSSZEFOGLALÁS

12.1 TARTALMI ÖSSZEFOGLALÁS

Befejezésként elevenítsük fel a könyvben elsajátított fogalmak, mutatók rendszerét, melyek végig gondolása segítség rendszerben lássuk a mutatók körét!

Adattípusok

- Mért adatok
- Ordinális adatok
- Nominális adatok

Kiértékelések

- Nominális adatok kiértékelése
 - Khi-négyzet-próba

Leíró statisztika

- Számított középértékek és helyzeti középértékek
 - Átlag,
 - Medián,
 - Módusz
- Gyakoriság
 - Abszolút gyakoriság
 - Relatív gyakoriság
 - Kumulált gyakoriság
 - Százalékos kumulált gyakoriság
- Gyakorisági poligon és a középérték-mutatók
- A középértékek egymáshoz viszonyított kapcsolata
 - Szóródási mérőszámok
 - a szóródási terjedelem
 - az átlagos eltérés
 - négyzetes összeg
 - variancia, szórás
 - relatív szórás mutatók képletét és használatát is.

Matematikai statisztikai lehetőségek

- Korreláció
 - Korrelációs számítás
 - Korrelációanalízis
 - Parciális korreláció

- A Spearman-féle rangkorreláció
- Regressziószámítás
- Faktoranalízis
- Klaszteranalízis
- Hipotézisvizsgálatok
 - Null- és alternatív hipotézisek, döntési szituációk
 - t-próba (Egymintás t-próba, Kétmintás t-próba)
 - Varianciaanalízis
- A Mann–Whitney-próba, Wilcoxon-próba, Kruskal–Wallis-próba értelmezése

Szemléltetés

- A diagramok szerepe
- Diagramtípusok
- Gyakorisági poligon, hisztogram

Az előző fejezetek taglalták azokat az eljárásokat, melyekben a változók középértékét elemezve vonható le a következtetés: egymintás-, páros-, független mintás t-próba, egy- és többszemponos varianciaanalízis, melyek mindegyike a paraméteres eljárások közé sorolhatók.

A nemparaméteres eljárásokban a változó folytonos vagy finom beosztású, melyben a mediánnak kiemelt szerepe van.. A fejezetben az alábbi próbák értelmezésére és gyakorlati feladatokon keresztül kerülnek bemutatásra az SPSS-ben történő elemzés során.

Ebben a tananyagrészen az olvasó megismerkedhetett a nemparaméteres próba jellemzőivel. Bemutatásra került a nemparaméteres Kolmogorov-Szmirnov, Mann–Whitney-próba, Wilcoxon-próba, Kruskal–Wallis-próba alkalmazása. Az eljárások feltételeit mintákon keresztül és az SPSS-ben való analízis menetét és az eredmények értelmezését tanulmányozható.

Az ANOVA több egyenlő szórású, normál eloszlású sokaság összevetésére alkalmas módszer, mely az Analysis Of Variance kezdőbetűiből kapta a nevét. A vizsgálat során arra keresünk választ, hogy a mért ismerv szerinti csoportokban az összehasonlítandó csoportok szignifikánsan különböznek-e egymástól.

A statisztikai számítások során gyakori feladat, hogy a többszöri válaszadások alapján elemzés megvalósítása válik szükségessé. A kérdésben a válaszok logikailag összetartozó változók együttese. A feldolgozás során arra vagyunk kíváncsiak, hogy a megkérdezettek válaszai milyen gyakorisággal kerültek kiválasztásra.

A fejezetben az olvasó a többválasztásos kérdéstípus SPSS statisztikai szoftverrel dolgozza fel és megismerkedik mely lépései alkalmasak a értékelésre.

A fejezet tanulmányozása során a hallgató képessé válik a teszt empirikus adatainak értékelésére és a következtetések levonására.

A többváltozós statisztikai vizsgálatok jellegzetes feladata az objektumok elemzése, a struktúrát egészében vizsgáló módszer. Egy ilyen igen gyakran alkalmazott osztályozási módszer a klaszteranalízis. Feladata az, hogy csoportokba soroljuk a különböző objektumokat azok hasonlósága alapján, közös tulajdonságaik figyelembe vételével. Az SPSS statisztikai szoftverrel lépései alapján képessé vált a módszer kiválasztására a vizsgálat megvalósítására.

A faktoranalízissel „csökkenthető” a megfigyelt változók száma” vagyis helyesebben változók összevonása valósítható meg az eljárással. A kutatási koncepciók kidolgozásakor a klaszteranalízis ad lehetőséget átfedésmentes csoportosításra. A fentiekben az olvasó megvizsgálta az analízis előfeltételeit és elvégezte a faktoranalízist, értelmezte a kapott paramétereket.

12.2 ZÁRÁS

A tankönyv anyaga segítséget nyújt az empirikus kutatások kiértékelésének elméleti és gyakorlati elsajátításához. Fontos, hogy a kutatás módszeréhez és az adattípushoz illeszkedő statisztikai mutatókat használjunk, de ne feledjük a statisztika csak egy eszköz az adatok értékeléséhez, igazi eredménye a kapott adatok alapján levont következtetésekben rejlik!

13. KIEGÉSZÍTÉSEK

13.1 IRODALOMJEGYZÉK

13.1.1 Hivatkozások

- BABBIE, Earl: A társadalomtudományi kutatás gyakorlata. Balassi Kiadó. Budapest. 2003
- BARBIER, Frédéric – Lavenir, Catherine Bertho: A média története: Diderot-tól az internetig. Budapest, Osiris Kiadó, 2004.
- BENEDEK András: Digitális Pedagógia. tanulás IKT környezetben. TYPOTEX. Budapest, 2008.
- Bevezetés a pedagógiai kutatás módszereibe. szerk. Falus Iván. Keraban Kiadó Budapest. 1996.
- Dr ILLYÉSNÉ dr. Molnár Emese: Gondolatok a minőség mérhetőségéről az alkalmazható módszerekről. Tudományos Évkönyv 2007. Budapesti gazdasági Főiskola 2008.
- FALUS Iván és Ollé János: Az empirikus kutatások gyakorlata. Adatfeldolgozás és adatelemzés. Nemzeti Tankönyvkiadó. Budapest. 2008
- FALUS Iván és Ollé János: Statisztikai módszerek pedagógusok számára. Budapest: Okker, 2000.
- KETSKEMÉTI László – IZSÓ Lajos: Az SPSS for Windows programrendszer alapjai. -SPSS Partner Bt.
- KETSKEMÉTI László – Izsó Lajos: bevezetés az SPSS programrendszerbe. ELTE Eötvös Kiadó. Budapest, 2005
- KOVÁCS Zoltán: Termelésmenedzsment. Veszprémi Egyetem Kiadó. Veszprém. 2001. p. 235
- LENGYELNÉ Molnár Tünde, TÓVÁRY Judit: Kutatásmódszertan . –Eger: Líceum kiadó, 2001.
- Murray R. SPIEGEL: Statisztika. PANEM-McGRAW-HILL Inc.. Panem Kft. Budapest. 1995
- SZÉKELYI Mária - BARNA Ildikó: Túlélőkészlet az SPSS-hez. Többváltozós elemzési technikáról társadalomkutatók számára. Typotex Kiadó. 2002.
- TÓTHNÉ PARÁZSÓ Lenke: Kutatásmódszertan matematikai alapjai: Eger, 2011. ISBN 978-615-5221-25-5
- Ю.Н.Тюрин, А.А.Макаров Анализ данных на компьютере. Инфра-М. Москва. 2003.

Elektronikus dokumentumok / források

KOVÁCS László: *NIIFP hálózati multimédia pilot projekt*. Budapest, SZTAKI, 2008.
 [elektronikus dokumentum] [2010.február 1.] <URL:
<http://www.sztaki.hu>>

13.2 MÉDIAELEMENK ÖSSZESÍTÉSE

13.2.1 Táblázatjegyzék

1.	Leíró statisztikai mutatók:.....	21
2.	Különbözőség vizsgálatok	21
3.	Összefüggésvizsgálatok.....	22
4.	Gyakorisági kategóriák.....	43
5.	A korrelációs együttható értelmezése	57
6.	Könyvkiadási adatok 1990-2011.	61
7.	A könyvtárak alakulása 1990-2000-ig	63
8.	Magyarország könyvtárai és állományuk, 2000-2011	66
9.	A népesség száma, 2001-2011.....	67
10.	Példa egymintás t-próba alkalmazására	85
11.	Példa két mintás t-próba meghatározására	88
12.	Feladat egymintás t-próba alkalmazására	94
13.	Táblázat: Paraméteres és a nemparaméteres eljárás párhuzamba állítása	104
14.	Példa a Mann-Whitney alkalmazására	108

13.2.2 Ábrajegyzék

1. ábra:	Fogalomtérkép	16
2. ábra:	Címkézés	19
3. ábra:	Kérdőíves válaszok	20
4. ábra:	Gyakoriság parancs helye	22
5. ábra:	Gyakoriság adatpanel	23
6. ábra:	Gyakorisági elemzés nominális adatok esetén	23
7. ábra:	Nominális adatok gyakoriság elemzése több táblázat esetén	24
8. ábra:	Keresztábra	25
9. ábra:	Keresztábra	26
10. ábra:	Keresztábra	27
11. ábra:	Keresztábra- Iskolai végzettség és lakhely	28
12. ábra:	Khi-négyzet próba beállítása	30

13. ábra:	Fogalomtérkép.....	33
14. ábra:	Átlag képlete.....	34
15. ábra:	Átlag meghatározása	35
16. ábra:	Módusz megadása.....	37
17. ábra:	Medián megadása	38
18. ábra:	Medián megadása	39
19. ábra:	Feladat középérték-vizsgálatra	39
20. ábra:	Gyakoriság vizsgálat.....	42
21. ábra:	Gyakoriság elemzés eredménye	42
22. ábra:	Gyakoriság elemzés eredménye	43
23. ábra:	Rendezés.....	44
24. ábra:	Gyakoriság elemzés eredménye	45
25. ábra:	Bimodális eloszlás	46
26. ábra:	Középértékek	47
27. ábra:	Az átlagos eltérés képlete.....	48
28. ábra:	Négyzetes összeg.....	49
29. ábra:	Variancia	49
30. ábra:	A szórás képlete	50
31. ábra:	Szóródási mérőszámok megadása.....	52
32. ábra:	Szóródási mérőszámok eredménytáblája.....	52
33. ábra:	Fogalomtérkép.....	56
34. ábra:	Lineáris korreláció képlete.....	57
35. ábra:	Háztatások IKT ellátottsága	59
36. ábra:	Háztartások infokommunikációs eszközellátottsága.....	59
37. ábra:	Korrelációs számítás.....	60
38. ábra:	A könyvkiadási adatok alakulása 1990-2011 között.....	62
39. ábra:	Korrelációs mátrix.....	63
40. ábra:	Lineáris regresszió	68
41. ábra:	A Kaiser-féle mutatószám.....	71
42. ábra:	X változó	72
43. ábra:	Standardizált változó szórásnégyzete.....	73
44. ábra:	Megfigyelt értékek mátrixa	73
45. ábra:	Kapcsolat.....	74
46. ábra:	A parciális korreláció képlete.....	75
47. ábra:	Korrelációs mátrix (Könyvtárak 2001-2011).....	75
48. ábra:	Parciális korreláció beállítása	76
49. ábra:	Fogalomtérkép.....	82
50. ábra:	Minták közti átfedés	83
51. ábra:	A t-próba képlete.....	84
52. ábra:	A t-próba eredménye	86
53. ábra:	A kétmintás t-próba képlete.....	87

54. ábra:	A belső variancia képlete	90
55. ábra:	A külső variancia képlete	90
56. ábra:	Varianciaanalízis.....	92
57. ábra:	Fogalomtérkép.....	98
58. ábra:	PISA-feladat.....	99
59. ábra:	Oszlopdiagram	100
60. ábra:	Fogalomtérkép.....	104
61. ábra:	Kolmogorov-Szmirnov lehulló menű	105
62. ábra:	Kolmogorov-Szmirnov lehulló menű beállítása	106
63. ábra:	Lehulló párbeszédablak beállítái	106
64. ábra:	Kolmogorov-Szmirnov lehulló menü	106
65. ábra:	A Kolmogorov-Szmirnov próba output felülete.....	107
66. ábra:	A Compare Means Option ablaka	107
67. ábra:	Compare Means output ablaka	108
68. ábra:	A normál és a gyógytestnevelés őszi és tavaszi adatai	109
69. ábra:	Output eredmények.....	110
70. ábra:	Mann-Whitney H beállítása	111
71. ábra:	Kruskal Wallis H eredmény az outputban.....	111
72. ábra:	Wilcoxon lehulló menű	113
73. ábra:	Wilcoxon párbeszédpanrl beállítása	114
74. ábra:	Wikoxon Rank Options ablak beállítása.....	114
75. ábra:	Wilcoxon Rank test eredménytáblázata	115
76. ábra:	Kruskal-Wallis próba lehulló menű.....	116
77. ábra:	Kruskal-Wallis próba lehulló ablak beállítása	117
78. ábra:	Kruskal-Wallis próba lehulló Statisics menű	117
79. ábra:	Kruskal-Wallis próba output ablaka.....	118
80. ábra:	Fogalomtérkép.....	121
81. ábra:	One Way ANOVA lehulló menű	123
82. ábra:	A független változó és a faktor kiválasztása	123
83. ábra:	Post Hoc Multiple Comparisons menűpont.....	124
84. ábra:	Descriptive és a Homogeneity of variance test rádiógombok kijelölése	124
85. ábra:	Internet eszközök használat átlagértékei	126
86. ábra:	Anova elemzés	126
87. ábra:	PC használat átlagértékei.....	127
88. ábra:	Levene és ANOVA statisztika	128
89. ábra:	Laptop használat átlagértékei.....	128
90. ábra:	Homogenitás vizsgálat	129
91. ábra:	ANOVA elemzés	129
92. ábra:	Szkenner használat átlagértékei	130
93. ábra:	Webkamera használat átlagértékei	130

94. ábra:	Táblagép használat átlagértékei	130
95. ábra:	Okostelefon használat átlagértékei	130
96. ábra:	Interaktív TV használat átlagértékei	131
97. ábra:	Internet használat átlagértékei	131
98. ábra:	Ebook olvasó használat átlagértékei	131
99. ábra:	IKT eszközök otthoni tanulási feltételeinek megítélése, pók diagram.....	132
100. ábra:	Fogalomtérkép.....	135
101. ábra:	A Multiple Response, Define Variabl Sets... parancs.....	136
102. ábra:	A Multiple Response, Define Variabl Sets lehulló menü.....	137
103. ábra:	A többszörös választás keresztábla beállítása.....	138
104. ábra:	A többszörös választás keresztábla beállítása.....	139
105. ábra:	A Multiple Responses lehulló sor és oszlop beállítása	139
106. ábra:	A Multiple Responses lehulló Keresztábla beállítása	140
107. ábra:	Multiple Responses keresztábla eredmény táblázata.....	141
108. ábra:	Multiple Responses gyakorisági táblázata.....	142
109. ábra:	Multiple Responses eredménydiagram.....	142
110. ábra:	Multiple Responses keresztábla	143
111. ábra:	Fogalomtérkép.....	145
112. ábra:	Analyze/Classify/K Means Cluster	150
113. ábra:	Iteration beállítása.....	150
114. ábra:	Iteration history táblázat	151
115. ábra:	Végső klaszterközpontok	152
116. ábra:	Az estek száma a klaszterekben	152
117. ábra:	Anova táblázat	153
118. ábra:	Analyze/Classify/Hierarchical Cluster	154
119. ábra:	Hierarchical Cluster menü	154
120. ábra:	A Statistics beállítása	155
121. ábra:	Plots menü beállítása.....	155
122. ábra:	Jégcsap diagram.....	156
123. ábra:	Dendogram	156
124. ábra:	Dendogram Ward módszerrel	157
125. ábra:	Fogalomtérkép.....	159
126. ábra:	Analyze/Regression/Linear.....	167
127. ábra:	Analyze/Regression/Linear histogram beállítása	168
128. ábra:	Analyze/Regression/Linear histogramja.....	168
129. ábra:	Analyze/Regression/Linear Scatterdot beállítása.....	169
130. ábra:	Analyze/Regression/Linear Scatterdot	169
131. ábra:	ábra. Analyze/Correlate/Bivariate lehulló ablak beállítása	170
132. ábra:	Analyze/Correlate/Bivariate eredménytáblázat.....	170
133. ábra:	Analyze/Dimension reduction/Factor/Descriptives	171

134. ábra:	Analyze/ Dimension reduction/Factor/Descriptives	172
135. ábra:	Analyze/ Dimension reduction/Factor/Descriptives lehulló menü beállítása	172
136. ábra:	Analyze/ Dimension reduction/Factor/Descriptives eredménytábla	173
137. ábra:	Analyze/Dimension Reduction/Factor/Descriptives anti-image eredménytábla.....	174
138. ábra:	Analyze/Dimension Reduction/Factor/Descriptives KMO beállítása	175
139. ábra:	Analyze/Dimension Reduction/Factor/Descriptives eredménytábla	175
140. ábra:	Faktormódszerek	177
141. ábra:	Principal components	178
142. ábra:	Principal Component Analysis output	179
143. ábra:	Scree Plot	179

13.3 GLOSSZÁRIUM, KULCSFOGALMAK ÉRTELMEZÉSE

Determinisztikus	Azonos körülmények között mindig ugyanúgy játszódik le az esemény; a feltételek ismeretében a jelenség további jellemzői egyértelműen meghatározottak (pl. szabadesés stb.).
Abszolút gyakorisági elosztásnak	A mintára vonatkozó eredményt abszolút gyakorisági elosztásnak nevezzük.
Adat	Egy szimbólum, mely a hozzárendelt értékek bármelyikét felveheti.
Adatgyűjtés	Adatgyűjtés, amit a statisztikában megfigyelésnek, adatfelvételnek neveznek.
Adatok elemzése	<i>Adatok elemzése, aminek célja, hogy az adatso- rokat és a közöttük levő összefüggéseket tömö- ren, egy-egy számított értékkel jellemez- hessük, és ezeket értékeljük.</i>
Adatok feldolgozása	Adatok feldolgozása, ami az adatok csoportosítá- sát, rendszerezését, összesítést, valamint az ada- tok táblázatba foglalását jelenti.
Alpha if Item Deleted	Alpha if Item Deleted arra mutat, hogyan változ- na a feladat összalphája, ha az Itemet kivennénk a feladatsorból.
Arányskála	Az egyedek ismérveit numerikusan kifejező számérték. A változó értékei sorba rendezhetőek, különbségük és arányuk is értelmezhető (pl. testmagasság, súly...)
Átlagos eltérés	Átlagos eltérés a minta számtani közepétől való távolsága
Corrected Item-Total correlation - Diszkrimi- nációs érték	A vizsgázók milyen teljesítményt nyújtottak az adott Itemnél az egész feladatsorra kivetítve.
Crombach α	A Crombach α rámutat, hogy a teszt mennyire reábilis, azaz milyen megbízhatóan mér.
Cronbach-alfa (megbíz- hatósági koefficiens)	A teszt stabilitását méri. Minél magasabb a teszt alfa értéke, annál megbízhatóbb a teszt.
Diszkrétnek változó	Értéke véges.
Elkülönítési mutató	Az elkülönítési mutató azt mutatja meg, hogy az

	item azt méri-e, amit a teszt egésze, vagyis a különböző tudású tanulókat tesztel azonos módon különíti-e el egymástól.
Értéktartomány	A minta legnagyobb és legkisebb eleme által határolt intervallum.
Érvényesség– (Validitás)	Ennek a kritériumnak való megfelelés, hogy a kutatás valóban a vizsgálat tárgyára irányul-e.
Feleletválasztásos feladatok	Feleletválasztásos feladatoknak nevezzük a feladatok azon csoportját, amelyben a kérdéshez, feladathoz megadott válaszlehetőségek közül kell kiválasztani, megjelölni a jó vagy rossz válaszokat, párosítani kell adatsorokat, rangsorolni, időrendi vagy egyéb logikai feltétel szerint kell sorba állítani megadott válaszokat, továbbá ok-okozati összefüggéseket, kapcsolatokat kell felismerni.
Fogalmi háló	A fogalmi háló kapcsolódási pontjai alapján kell megtervezni a tanítási-tanulási folyamatot, az ismeretek logikai sorrendjét, valamint a számonkérés, a mérési értékelési folyamat feladatsorait.
Folytonos változó	Értéke végtelen.
Független változó	
Függő változó	Két változó együttes hatásának eredményeképp módosul.
Graf	A tananyag és az egyes részek (tények, fogalmak, összefüggések) kapcsolatainak vizuális megjelenítése.
Gyakoriság	Egy olyan mutató, amely jellemzi, hogy egy-egy csoportba hány adat tartozik.
Gyakoriság	A gyakoriság egy olyan mutató, amely jellemzi, hogy egy-egy csoportba hány adat tartozik.
Gyakorisági eloszlás	Egy olyan statisztikai mutató, mely arra mutat, hogy a minta elemei hogyan oszlanak meg a különböző csoportok között. A mintára vonatkozó eredményt abszolút gyakorisági eloszlásnak nevezzük.
Heteroszkedaszticitás	Az a szóráshomogenitás, melynek során a függő változó nem rendelkezik azonos szórással a független változó különböző szintjei mellett
Homoszkedaszticitás	Az a szóráshomogenitás, melynek során a függő

	változó azonos szórással rendelkezik a független változó különböző szintjei mellett
Intervallumskála	Az objektum kvantitatív mérése során a mérhető adatokat vizsgálva az egyedeket jellemző ún. méréssel kapott adatokat kapunk. Az intervallum nagyságát a két adat közötti eltérés adja, definiált mértékegységgel rendelkezik, tehát különbségük értelmezhető (születési dátum, életkor...).
Item	A tesztek legkisebb önállóan értékelhető egységét jellemző adat. Az alternatív elemek nehézségi foka, fontosságát kiegyenlítő elem.
Item determináció	Az item determináció az a mutatószám, amely jelzi, hogy az adott item milyen erősen determinálja a tesztben elért összpontszámot.
Item nehézsége	Az item nehézsége, nehézségi index: rámutat, hogy az itemet milyen valószínűséggel oldja meg a tanuló.
Itemdetermináció D_h	Arra mutat, hogy az egyes itemeknek mekkora befolyása van az összpontszámra, vagyis milyen a differenciáló ereje, mekkora a determinációs hatása. Itemrealibilitás: a teszt megbízhatóságáról ad információt.
Itemek	A tesztek legkisebb önállóan értékelhető egységét jellemző adat. Populációnak vagy más néven sokaságnak nevezzük azt a vizsgált csoportot, amely a vizsgált egyedek összességét foglalja magában. A populáció egyedei a statisztikai elemek.
Itemnehézség (p)	Rámutat, hogy az adott itemet az egyes tanulók milyen valószínűséggel oldották meg.
Kísérlet	Meghatározott hipotézisből kiindulva új, rejtett összefüggések, törvényszerűségek feltárására alkalmas módszer.
Korreláció szignifikanciája	Választ ad arra, hogy mennyire bízhatunk egy mintából számolt korrelációs együtthatóban.
Kumulatív gyakoriság	A kumulatív gyakoriság egy olyan statisztikai mutató, amely arra mutat, hogy a mintából mennyi azon elemek száma, amely egy előre meghatározott szintet ér el. Jele: cf
Kutatás	Valamilyen tudatosult igény, probléma megoldására irányuló megoldási folyamat, melynek során

	a jelenséget komplex módon előre átgondolt hipotézis alapján tanulmányozzuk.
Kutatások célja	A vizsgált minta által reprezentált vizsgálati eredmények populációra való általánosíthatóságának bizonyítása.
Médián	A nagyság szerint rendezett, vagyis rangsorba állított számhalmaz középső értéke. Páratlan számú minta esetén a sorba rendezést követő középső elem, páros számú minta esetén a sorba rendezést követő két középső elem átlaga.
Megbízhatóság (Reliability):	Ennek a kritériumnak való megfelelés azt jelenti, hogy a kutatás annak megisméltése, ismételt alkalmazása során is az eredetivel egyező, illetve kevéssé eltérő eredményt szolgáltat. Mérése a varianciák összehasonlításával történik.
Mérés-értékelés fogalma	Az értékelés megerősítési, visszacsatolási folyamat, amely során nemcsak a tanulók tevékenységét értékelhetjük, hanem az egész tanítási-tanulási folyamatot, annak hatékonyságát, beleértve a folyamat összes tényezőjét.
Minta	A populáció részhalmaza, amelyen a kísérletet végezzük.
Minta átlaga	A számhalmaz átlaga, más szóval – számtani közepe – az a szám, amelytől az adatok eltéréseinek összege zérus.
Módusz	Egy számhalmaz módusza a legnagyobb gyakorisággal rendelkező érték. A módusz nem feltétlenül létezik, és ha igen, nem biztos, hogy egyetlen érték képviseli.
Nominális skála	Olyan szimbólumok, számok, melyek csak az azonosítást szolgálják. A valós számok egy tulajdonsága sem jellemzi, vagyis még sorba sem rendezhetők (pl. nemek, beosztás, lakóhely, vallás...).
Objektivitás	Ennek a kritériumnak való megfelelés azt jelenti, hogy mennyire tárgyilagos, vagyis független a mérés során kapott eredmény az adott módszert alkalmazó, a felmérést végző személytől.
Ordinális skála	Olyan szimbólumok, számok, amelyek alkalmassá teszik a vizsgált egyedek közötti sorrendiség felállítását, mely lehet az egyenmű adatok rendezé-

	sének alapja is. A változó értékeinek különbsége nem értelmezhető. (pl. iskolai végzettség, attitűd skála értéke, a termékek minősítés értékei, osztályzatok...).
Populáció	Azon egyének (dolgok) összessége, akikről (amikről) információt szeretnénk kapni.
Relatív gyakoriság	A csoport abszolút gyakoriság értékének a minta elemszámához százalékosan viszonyított értéke.
relatív gyakoriság	A relatív gyakoriság a csoport abszolút gyakoriság értékének a minta elemszámához százalékosan viszonyított értéke.
Scale Mean if Item Deleted	Scale Mean if Item Deleted azt mutatja meg az Itemre kivetítve, hogyan változna meg az átlag, ha a feladatsorból az adott Itemet kivennénk.
<i>Szignifikanciaszintnek</i>	Az a valószínűség, amely esetén H_0 -t elvetjük, p -vel jelöljük és nevezzük. Értékei $p < 0,05$, $p < 0,01$ és $p < 0,001$. Ezekhez a szignifikancia szintekhez tartozó próbastatisztika értékek az ún. kritikus értékek.
Szignifikáns eltérés	Ha a próbastatisztika értéke nagyobb/egyenlő egy adott szignifikancia szinthez (pl. $p < 0,05$) tartozó kritikus értéknél, akkor H_0 -t elvetjük és azt mondjuk, hogy a $p < 0,05$ -ös szinten.
Szórás	Az adatok mintaátlagától vett négyzetes átlagát.
Szórás	Szórás alatt értjük az adatok mintaátlagától vett négyzetes átlagát (középértéke).
Sztokhasztikus	Más szóval véletlen a jelenségek kimenetele, azonos körülmények között sem egyértelműek (pl. pénzfeldobás, lottó stb.).
Tananyagelemzés	A tananyagelemzés az ismeret logikai struktúráját lépésekre bontja.
Rang	Az érték rangja azt jelenti, hogy a nagysága alapján hányadik a mintában.