SZEGEDI TUDOMÁNYEGYETEM MEZŐGAZDASÁGI KAR

> Kutatásmódszertani alapismeretek Bevezetés az SPSS használatába

> > Oktatási segédlet

Összeállította: Hódiné Szél Margit Mikó Józsefné Jónás Edit

Összeállította:

Hódiné Szél Margit (Gazdálkodási és Vidékfejlesztési Intézet) 1., 2., 3., 4., 5., 7. fejezet Mikó Józsefné Jónás Edit (Állattudományi és Vadgazdálkodási Intézet) 6., 7. fejezet

Jelen tananyag a Szegedi Tudományegyetemen készült az Európai Unió támogatásával. Projekt azonosító: EFOP-3.4.3-16-2016-00014



Tartalomjegyzék

Előszó.		4
1. Tuc	dományos kutatás alapfogalmai	6
1.1	Alapfogalmak	6
1.2	A kutatás folyamata	7
1.3	A kutatás módszere	9
1.4	Mérés	. 12
1.5	Mintavétel	. 15
2 SPS	SS alapismeretek	. 25
2.1	SPSS felületei	. 25
2.2	Adatbevitel	. 32
3 SPS	SS menüpontok	. 34
3.1	File menü	. 35
3.2	EDIT menüpont	. 37
3.3	View menü	. 39
3.4	DATA menüpont	. 40
3.5	Transform menü	. 43
3.6	Graphs menü	. 48
4 Leí	ró statisztika	. 49
4.1	Leíró statisztika (Descriptive Statistics):	. 51
5 Leí	ró statisztika az SPSS-ben	. 60
5.1	Analyze/Descriptive Statistcs /Frequencies	. 62
5.2	Analyze/Descriptive Statistcs/ Descriptive	. 64
5.3	Normális eloszlás, a normalitás tesztelése	. 65
5.4	Analyze/Descriptive Statistcs /Explore.	. 67
Varianc	iaanalízis	. 70

	5.5	Egytényezős varianciaanalízis (/One Way ANOVA)	74
6	Ko	rreláció- és regressziószámítás	79
	6.1	Lineáris korreláció	79
	6.2	Kétváltozós lineáris regresszióanalízis	81
7	Gya	akorló feladatok	87
8	Fel	használt irodalom	97

Előszó

A korábbi évek tapasztalatai felhívták a figyelmet arra, hogy a kutatások egyre jobban megkövetelik a matematikai statisztikai ismereteken alapuló mérés-értékelés kreatív ismeretét. Ezen tudáselemek birtokában a kutatás során kapott adatok elemzése végezhető el, melyekkel feltárhatók a vizsgált paraméterek mélyebb összefüggései. A numerikusan kapott eredmények értelmezése és a helyes következtetések levonása meghatározója a további kutatás menetének.

A jegyzet szorosan kapcsolódik az előző félévekben hallgatott statisztika és informatika tantárgyakhoz. Az összeállított tananyag segítséget nyújt a hallgatóknak a statisztikai módszerek alapos és érthető megismerésében, a korábban manuálisan megoldott feladatok statisztikai programcsomagokkal történő könnyebb megoldásában. Részletes útmutatásokkal nyílik lehetőség a hallgató számára a kutatás adatainak szakszerű, korszerű feldolgozására, majd az alkalmazandó statisztikai próbák kiválasztására és elvégzésére, paraméterek közötti összefüggések keresésére, az eredmények kiértékelésére és a következtetések levonására.

Az oktatási segédlet egy rövid kutatásmódszertani összefoglaló után az SPSS programcsomag felépítését, alkalmazási lehetőségét mutatja növénytermesztési és állattenyésztési példákon keresztül. A kapott eredményekből történő megfelelő következtetések levonásához elengedhetetlen a szakmai ismeret.

Így az SPSS statisztikai programcsomag egy részének megismerésével egy hatékony eszközt kapnak a hallgatók a kutatásaik során összegyűjtött, illetve mért adataik statisztikai módszerekkel történő elemzéséhez.

Az oktatási segédletet a SZTE Mezőgazdasági Kar hallgatói számára állítottuk össze segítséget nyújtva a szakdolgozat elkészítéséhez. A Karon oktatott tárgyak között szerepel a Mezőgazdasági kísérletek tervezése és értékelése fakultatív kurzus, melynek az elméleti és főleg gyakorlati részéhez kapcsolódik a jegyzet. Az utolsó fejezet a hallgatók részére a gyakorlatok keretében történő kisebb adathalmazon megoldandó feladatokat tartalmazza, melyeket az órán közösen kívánunk megoldani.

Tudás	Képesség	Attitűd	Autonómia/felelősség
Rendelkezik alapvető statisztikai és informatikai (Excel) ismeretekkel. Ismeri a mintavételezési módokat és a mérési eredmények rögzítésének módját Excel és SPSS programokkal.	Képes a növénytermesztés, állattenyésztési kiséletek elvégzésére, és a mérési eredményekből adatbázist elkészíteni Excel és SPSS programokkal.	Motivált az mezőgazdasági kutatások területén keletkezett új tudományos eredmények	Saját kutatásait, a kutatás során alkalmazott statisztikai próbákat felelősséggel végzi és a kapott eredményeket kritikusan szemléli.
Ismeri a leíró és matematikai statisztikai módszereket, azok alkalmazásának kritériumait és menetét és meghatározását az SPSS programcsomaggal.	Képes leíró és matematikai statisztikai módszereket megválasztani és alkalmazni Excel és SPSS programokkal.	megismerésére, saját kutatásának statisztikai vizsgálatára. Kritikusan szemléli a kapott eredményeket. Folyamatosan törekszik az önképzésre, tudása, ismeretei	A döntéseit tudományosam megalapozottan hozza, figyelembe veszi az eredményeit, lehetőségeit.
Ismeri a hipotézisvizsgálat menetét, az egyes próbák alkalmazhatóságát. Ismeri a kétváltozós lineáris regresszióanalízist	Képes a hipotéziseit a megfelelő statisztikai módszerekkel megvizsgálni, az eredményeket értelmezni. Képes ok-okozati összefüggéseket feltárni a vizsgált paraméterek között.	aktualizálására, munkáját a minőségi munkavégzés iránti igény jellemzi	Szakmai munkájában is használja az ellenőrzést, mint módszert. Véleményét önállóan, szakmailag, ha szükséges matematikailag is megalapozottan és felelőssége tudatában fogalmazza meg.

A tantárggyal a következő konkrét *tanulási eredmények* alakíthatók ki:

Hódmezővásárhely, 2018.09.15

1. Tudományos kutatás alapfogalmai

1.1 Alapfogalmak

A **kutatás** során új ismeretek (összefüggések, törvényszerűségek) feltárása a cél. Egy általunk kiválasztott populáció vizsgálata az általunk előre meghatározott kritériumok szerint, amely megvalósulhat átfogóbb és szűkebb populációban is. Így például különböző vegyszerezéses kezelésekre az egyes napraforgó fajták hogyan reagálnak.

Kutatás tárgya: a fejlesztés során érvényesülő törvényszerűségek feltárása. Kutatás metodikája: a tudomány előírásainak megfelelő megismerési folyamat (technikák,

eljárások).

A kutatás alatt értendő valamilyen tudatosult igény, probléma megoldására irányuló tevékenység, melynek során a jelenséget komplex módon előre átgondolt hipotézis alapján kell tanulmányozni.

Kutatások típusai (Falus):

- Alapkutatások (új ismeretek gyakorlati cél nélkül),
- alkalmazott kutatások (új ismeretek gyakorlati céllal, a tudományos ismeretek alkalmazása),
- Akciókutatások (a létező gyakorlat kritikája alapján egy új elmélet bemutatása, a felvetett probléma megoldása).

Kutatási stratégiák

- *Deduktív (analitikus) kutatási stratégia*: a forrásokat, dokumentumokat és eddigi tapasztalatokat elemezve fogalmazza meg az elveket, törvényszerűségeket.
 - Dedukció: a gondolkodás a megfigyelés felé halad, azaz sejtéseket, hipotéziseket vezetünk le az elméletből.
- *Induktív (empirikus) kutatási stratégia:* a következtetéseket a tapasztalati mérésekre és azok elemzésére alapozva kell levonni.
 - Indukció: a gondolkodás a megfigyelések felől indul, azaz bizonyos megfigyelések általánosítása történik

1.2 A kutatás folyamata

A kutatás kritériuma megköveteli a kutatótól, hogy új ismeret feltárását célzó probléma megoldására a javaslatait megtegye. A következő 1.1. ábrán a kutatás folyamatát kísérhetjük végig.



Forrás: Tóthné: Kutatásmódszertan matematikai alapjai

1.1. ábra. A kutatás folyamata

Kutatási probléma meghatározása: az elméleti ismeretek, tételek gyakorlati szituációkban lévő létjogosultságát bizonyítása. Gyakorlat, melynek során pl. több módszer közül választjuk ki a leghatékonyabbat.

Elemzési egységek és időfaktor kiválasztása: elemzési egységnek azt tekintjük, hogy kit vagy mit kívánunk tanulmányozni. Az időfaktor azt jelenti, hogy adott jelenséget egy időpontban, vagy időintervallumban kívánjuk mérni, megfigyelni.

Korábbi eredmények áttekintése: a témához kapcsolódó szakirodalom összegyűjtése, áttekintése.

- A szakirodalom ismerete fontos, ha egy elméleti modellt, tézist tesztelünk (deduktív logikájú kutatásoknál).
- A témához kapcsolódó bibliográfia áttekintése/elemzése segítséget ad abban, hogy
 - o ne ismételjünk meg már elvégzett kutatást,
 - o felállítsuk a hipotéziseket,
 - o a korábbi kutatásoknál már bevált módszereket/eszközöket ismerjünk meg.

• Eredményeinket összehasonlítjuk a szakirodalmi eredményekkel, hogy megítéljük a saját eredmények újszerűségét és érvényességét.

Hipotézis megfogalmazása

Hipotézis nem más, mint a kutatási problémára adott feltételezett válasz, azaz a kutató feltételezéseit megfogalmazó kijelentés, a problémában szereplő változókra, azok kapcsolatára, és magyarázatot ajánl a probléma megoldásához.

A hipotézis:

- a kutatás megkezdése előtt kell kialakítani,
- előzetes tapasztalatok, vagy korábbi tudás alapján állítjuk fel,
- alapulhat
 - o sejtésre,
 - o korábbi vizsgálatra,
 - o következtetésre
- funkciója az, hogy a kutatási célok eléréséhez vezessen, a kutatás vezérfonalát alkotja,
- a hipotézisben a vizsgálat eredményével kapcsolatos következtetések elfogadhatóságát, illetve tarthatatlanságát fogalmazzuk meg

Hipotézissel szemben támasztott követelmények:

- rendelkezzen magyarázó erővel;
- jelölje a változók kapcsolatát ítélet formájában;
- legyen egyértelműen igazolható vagy elvethető;
- a hipotézis igazolása vagy elvetése megvalósítható módszereket vagy technikákat igényeljen;
- legyen világos, egyértelmű, operatív terminusokban megfogalmazva;
- támaszkodjon a meglévő ismeretekre;
- a legegyszerűbben és a legtömörebben kell megfogalmazni;
- a hipotézisnek összességének választ kell adnia a kiinduló problémára.

Hipotézisek modellezése választott kutatási témánkra:

arra kell törekedni, hogy azok leírják a probléma teljes körét (a kutatási terv szerves része)

A hipotézis megfogalmazásának módja lehet:

- Induktív: feltételezéseink a tapasztalatainkból erednek
- Deduktív: általános elvekből, törvényszerűségekből indulunk ki

Hipotézisek fajtái megfogalmazásuk szerint lehetnek:

- Null-hipotézis: feltételezzük, hogy nincs összefüggés a változók között
- Alternatív, irány nélküli hipotézis: feltételezünk összefüggést, de annak irányát nem adjuk meg
- Alternatív, irányt is jelző hipotézis: megjelöljük a változók közötti feltételezett kapcsolatot is

Konceptualizálás, operacionalizálás: azaz a vizsgálandó fogalmak és változók jelentésének pontos meghatározása a kutatásoknak, az empirikus adatgyűjtésnek nélkülözhetetlen és lényegbevágó kritériuma, amely a vizsgálat változójának, mérési eljárásának (technikai megközelítés) megfogalmazása, a fogalmak mérésére szolgáló technikák meghatározását jelenti. Az indikátor, a fogalmak, a hipotézisek mérhető leírása.

A konceptualizálás folyamatának a folytatása az operacionalizálás, amely az adott fogalom méréséhez vezető konkrét eljárások, lépések megadását jelenti: Miként mérjük majd a vizsgált változót?

Módszer kiválasztása: válasz a hipotézis, az adott vizsgálati eljárás megválasztása.

Mintavétel: a populáció és a reprezentativitást biztosító mintavételi technika meghatá-rozása.

Adatgyűjtés: az információk gyűjtése. az adatok elemezhető formába rendezése (gyakran kódolással)

Eredmények közlése: az adatok statisztikai feldolgozását követően az eredményeket értelmezve, tanulmányban összefoglalva közli a kutató.

1.3 A kutatás módszere

A kutatás általában már előzően, pl. előfelmérések, tapasztalatok alapján valósul meg. A kutatás során szöveges, vagy numerikus formában kapott az információ halmaz, alkalmas azok **kvantitatív** és **kvalitatív** feldolgozására (Babbie, 2003). A kutatás során azt tapasztalhatjuk, hogy nem lehet éles határt húzni a két módszer között, mivel mindkettőt komplex módon

alkalmazva kell értékelni az eredményeket. A **kvantitatív** mérési eredmények számadatait értelmezni kell didaktikai szempontból is.

Kvalitatív kutatás

- A kutatás során minőségi kérdésekre, "Mi? Miért? adunk választ.
- A kvalitatív módszerek általában mélyebb, árnyaltabb ismeretek megszerzésére irányulnak és viszonylag kis elemszámú mintán történik az adatfelvétel. A kapott eredmények nem számszerűsíthetők, nem mérhetők. A kvalitatív vizsgálatok abban az esetben alkalmazhatók sikeresen, amikor a különböző viselkedésformák, magatartásbeli sajátosságok mozgatórugóit igyekeznek feltárni.

Kvantitatív kutatás

A kutatómunka során a legismertebb kvalitatív eljárások:

- Tipizálás a kutatómunka során az adatok rendezése útján történik (pl. időigény, tipikus hibák), melynek alapja a megoldás logikai menete, megtervezettsége.
- Táblázatba foglalás a modulrendszerek formai lehetősége.
- Összehasonlítás során a vizsgált csoportok közötti kategóriák, típusok, táblázatok közötti összehasonlítását végezzük el.
- Elemzés a kommunikáció szempontjából igen fontos része a kutatásnak, mely a kiegészítő következmények leírását tartalmazza.
- Forráselemzés a kritikai érzékkel feltárt írásbeli, szóbeli anyag, mely a kutatásunk részét képezi
- Következtetés, mely a nyert tapasztalatok alapján a problémák, tapasztalatok megállapítása.
- Eredményrögzítés a kvalitatív kutatási eljárás szövegesen megfogalmazott leírása.

A kvalitatív eljárás előnyei:

- Nyitott, dinamikus, rugalmas.
- Mélyreható megértés lehetőségét kínálja.
- Felhasználja a kutató kreativitását.
- Gazdagabb ötletforrásokat szolgáltat.
- Áthatol az egyszerűsített vagy felületes válaszokon

Kvantitatív kutatás

A kutatás során mennyiségi kérdésekre – "Mennyi?" – adunk választ. Azokat az eljárásokat nevezzük kvantitatíveknek, melynek során numerikus adatokból, statisztikai eljárásokkal vonjuk le a populációra vonatkozó következtetéseket.

Kvantitatív eljárás előnyei

- Statisztikai és számszerű mérés
- Alcsoport mintavétel vagy összehasonlítások lehetősége
- Felmérés a jövőben megismételhető és az eredmények összevethetők
- Egyéni válaszokra épít
- Kevésbé függ a kutató szemléletétől



1.2. ábra. A kutatási módszerek csoportosítása a kutatás jellege alapján

Kutatási módszerek a kutatás jellege alapján (1.2. ábra):

- *Feltáró kutatás* akkor használható, ha a kutató kevés előzetes ismerettel rendelkezik a problémáról, illetve új ötletekre kíváncsi.
 - o Rugalmas, strukturálatlan technikák alkalmazása
 - o Kvalitatív technikák használata

- *Következtető kutatások* abban az esetben alkalmazhatók, ha a kutató tisztában van a probléma jellegével, struktúrájával.
 - Leíró kutatás célja, hogy beszámoljon egy adott helyzetről, pl. piaci részesedésről.
 - Kérdőíves felmérés
 - Kérdőíves megfigyelés
 - Ok-okozati kutatás célja függő és független változók felhasználásával logikai összefüggések feltárása.

1.4 Mérés

Mérésnek azt a folyamatot nevezzük, amikor a dolgok kijelölt tulajdonságához adott szabály alapján számot, úgynevezett adatot rendelünk. Több egymással összefüggő döntés eredményeképpen meghatározzuk a kutatási céloknak megfelelő pontosságot, a mérés terjedelmét, kiemeljük a főbb dimenziókat, határozunk a változó attribútumairól (vagy értékeiről), valamint a mérési szintről.

Mérési skálák

- Nem metrikus skála
 - o Névleges (nominális) skála
 - o Sorrendi (ordinális) skála
- Metrikus skála
 - o Intervallumskála,
 - o Arányskála.

Skálák jellemzői

- Névleges (nominális) skála (mérési szint) a számok kötetlen hozzárendelését jelenti.
 - o Számok (kódszámok) sokság egységeinek azonosítása,
 - o Területi és minőségi ismérvek szerinti megfigyelésnél alkalmazzuk,
 - o Számok közötti relációk, számtani műveletek nem értelmezhetők.
 - pl.: rendszám, irányítószám, TAJ szám, adószám stb.

- Sorrendi (ordinális) mérési skála a sokaság egyedeinek egy közös tulajdonság alapján való sorba rendezése.
 - A skálán az egyes egyedek nem feltétlenül egyenlő távolságra helyezkednek le egymástól.
 - A mérésből származó adatokkal (sorszámokkal) csak azok a műveletek végezhetők, amelyek során kizárólag a skálát képező számértékek sorrendisége kerül kihasználásra.

pl.: hallgatók osztályzata, sportolók helyezései, országok hitelképességi sorrendje, termékek minőségi osztályai

- Intervallumskála (különbségi skála) a skálaértékek különbségei is valós információt adnak a sokaság egységeiről.
 - Mértékegység és a nullapont meghatározása önkényes, és e nulla érték nem tükrözi a tulajdonság hiányát (0 °C).
 - A társadalomtudományokban alkalmazott skálák, mint például a Likert-, a szemantikus differenciál és a Stapel-skála.
 - A skálán két érték összege vagy aránya nem értelmezhető.
 - Kék-két adat különbsége, két különbség összege, aránya már értelmezhető:
 - pl. 5 °C és 10 °C közötti különbség azonos a 15 °C és 20 °C közötti különbséggel.
- Arányskála -a legmagasabb mérési szint- nyújtja a legtöbb információt.
 - A skálának valódi nullpontja van, amely nullpont a tulajdonság hiányát jelzi.
 - o A skála bármely két értékének aránya független a mértékegységtől.
 - E skálán mért számokkal a statisztikai elemzésekhez szükséges összes műveletek elvégezhetők.
 - pl.: hosszúság, tömeg, jövedelem, költség, termelés mennyisége, stb.

	Skálatípus	Alapvető tulajdonsága	Példák	A legmegfelelőbb számítási műveletek és leíró statisztikák
Nom	Névleges skála	A számok azonosításra és csoportosításra szolgálnak.	Nem, vallás, lakhely, reklámeszköz fajtája	Gyakorisági eloszlások Százalék, módusz
Nem metrikus	Sorrendi skála	A számok relatív pozíciót jelölnek, de különbségük nem értelmezhető.	Preferencia- sorrend, iskolai végzettség, piaci pozíció	Medián, kvartilis
	Intervallum skála	Egyenlő szakaszokra osztott skála, természetes nullapont nélkül. A különbségek összehasonlíthatók.	Hőmérséklet (°C, °F), attitűd, vélemény	Kivonás, összeadás. Terjedelem, átlag, szórás
Metrikus	Arányskála	Egyenlő szakaszokra osztott skála, természetes nullaponttal. Az arányok is összehasonlíthatók.	Hőmérséklet (KJ, magasság, súly, születési év, jövedelem, piaci pozíció	Össszeadás, kivonás, osztás, szorzás. Mértani átlag, harmonikus átlag

1.4.1. táblázat. A metrikus és nem metrikus mérésekhez tartozó elsődleges skálák

Forrás: Sajtos, Mitev: SPSS kutatási és adatelemzési kézikönyv

A társadalomtudományokban alkalmazott skálák a következők:

- Likert-skála: öt válaszkategóriával rendelkező skála. Az állítások mellett a kitöltőknek az "egyáltalán nem ért egyet" és a "teljesen egyetért" közti tartományból kell számokat kell megjelölniük. Könnyen alkalmazható, ezért közkedvelt mérési skála.
- Szemantikus differenciál: Hétfokozatú értékelőskála. A végpontokon ellentétes jelentésű melléknevek szerepelnek. A köztes szakasz egyforma szegmensekre van osztva és a válaszadó ezen jelöli meg a véleményét.
- Stapel-skála: egypólusú skála, amely 10 kategóriából áll, -5-től +5-ig. Ezt általában függőlegesen szokták ábrázolni úgy, hogy középen található a kifejezés, amelyet jellemezni akarnak.

A skálák között létezik átmenet, ami azt jelenti, hogy a skálák transzformálhatók a fejlettebbtől a kevésbé fejlett irányába, ami mindig információveszteséggel jár. Az arányskálából -mivel ez a legfejlettebb típus -bármilyen más skálatípus (intervallum, ordinális, nominális) előállítható, míg az ordinális skála csak nominálissá transzformálható. Megjegyzendő az is, hogy egyanazon

változó eltérő skálákon is mérhető, például a jövedelem alapjában véve arányskála, de ha kategorizáljuk, akkor sorrendi skálaként értelmezhető.

A skálák kialakításakor a következő választási lehetőségekkel találkozunk:

- A skála fokozatainak száma. Általánosan elfogadott az 5 és 9 közötti kategóriaszám.
- A kiegyensúlyozott vagy kiegyensúlyozatlan skálák használata. Általánosan elfogadott a kiegyensúlyozott skálák alkalmazása.
- A páros vagy páratlan számú kategória alkalmazása, azaz tartalmaz-e a skála semleges középső) fokozatot.
- Kényszerítő vagy nem kényszerítő skálák alkalmazása, azaz van-e lehetőség nem tudom/ nincs véleményem válaszra vagy mindenképp választani kell.
- A kategóriaértékek szöveges megfogalmazása.

1.5 Mintavétel



A mintavétel során az alapvető lépések a következők:

- a sokaság meghatározása;
- mintavételi keret meghatározása;
- mintavételi technikák kiválasztása;
- mintanagyság meghatározása;
- mintavétel kivitelezése.

A sokaság megfigyelése

Egy adatfelvételi folyamatban két lehetőség van: egyrészt a teljes sokaság megkérdezése (cenzus), másrészt a mintavétel (1.3. ábra).



Forrás Dr. Illyésné dr. Molnár, 2008

1.3. ábra. Adatszerzési módok

A mintavétel során a sokaság a vizsgálat tárgyát képező, a kutatás szempontjából valamilyen közös jellemzővel rendelkező egységek összessége.

A sokaság típusai:

- Diszkrét: ha a sokaság valóságos és jól elkülönülő egységekből áll, pl.: magyar tejfogyasztók 2016-ban.
- Folytonos: ha a sokaság valóságos, de csak önkényesen elkülöníthető egységekből áll, (például tejfogyasztás Magyarországon 2016-ban).
- Fiktív: ha a sokaság csak elképzelt elemekből ál, pl.: jövő évi várható tejfogyasztás).

Egy sokaság megadása az egységeinek tételes felsorolásával vagy az azt alkotó egységek összes közös tulajdonságának megadásával definiálható. A közös tulajdonságok megadása gyakran jelenti az időben, térben vagy mindkét tekintetben való lehatárolást. Hiába rendelkezik két sokaság pontosan ugyanazokkal a közös tulajdonságokkal, ha az egységek időbeli és/vagy térbeli helyzete eltér, akkor időben és vagy térben különböző sokaságokról beszélünk (Hunyadi-Mundruczó-Vita)

Mintavételi keret

A véges elemszámú sokaságból történő mintavételnél alapvető fontosságú, hogy rendelkezésre álljon egy ún. mintavételi keret, amely egyenként tartalmazza a vizsgálni kívánt sokaság elemeit, mégpedig mindegyiket, és mindegyiket csak egyszer.

Egy ilyen teljes keret biztosítása sokszor nem könnyű feladat, mert vannak olyan sokaságok, amelyeknél az elemek száma és összetétele napról napra változik, s bármilyen jó is a megszűnő és az újonnan létrejövő egységek nyilvántartása, ez szükségszerűen különbözik a mintavételi keret összeállításakor létező sokaságtól pl. Magyarország népessége.

Mintavételi eljárások

Alapvetően két mintavételi eljárást különböztetünk meg:

- Nem véletlen mintavételi eljárások
- Véletlen mintavételi eljárások

I. Nem véletlen mintavételi eljárások

A mintavételi eljárások hátránya – nincs biztosítva, hogy a minta a sokaságra valóban jellemző legyen, ennek eredményeképpen félrevezető következtetéseket lehet levonni.

A nem véletlen minták esetén:

- Nem lehetséges a mintából számított jellemzők hibájának meghatározása,
- Nem becsülhető a bizonytalanság, a tévedés várható hibája.
- Előnye végrehajtása egyszerűbb, olcsóbb, mint a véletlen mintavétel.

Fajtái:

- Szisztematikus kiválasztás
- Kvóta szerinti kiválasztás
- Önkényes kiválasztás
- Hólabda

Szisztematikus kiválasztás (véletlen mintavételi eljárásnál is említett)

- Ha a listaképző ismérv és a megfigyelt ismérv között nincs sztochasztikus kapcsolat akkor ez az eljárás véletlen mintát eredményez.
- Ellenkező esetben a kapott mintaelemek nem lesznek függetlenek egymástól a következtetések levonása során figyelembe kell venni a mintaelemek függőségéből adódó torzítást is.
- Nem célszerű alkalmazni időbeli megfigyeléseknél a periodicitás veszélye miatt.

Kvóta szerinti kiválasztás

- A felvételt végző személyek (kérdezőbiztosok) előre megkapják, hogy milyen összetételű mintához kell jutniuk, de az előre adott kereteken belül rájuk van bízva a véletlenszerű kitöltés.
- Hátránya a kapott minta a kérdezőbiztosok szimpátiája szerint áll össze, és ez befolyásolja a kapott eredményeket.

Önkényes kiválasztás

- A felvételt végző személy szakmai ismereteire támaszkodva a véletlent figyelmen kívül hagyva – választja ki a sokaságra jellemző (vagy legalábbis általa jellemzőnek tartott) mintát.
- Az ilyen kiválasztáson alapuló megfigyelés sokszor erősen torzított eredményt ad.

Exit pool eljárás

- Elsősorban a választási eredmények előrejelzésére alkalmazzák.
- Lényege hogy a szavazóhelyiségből kijövő választót megkérdezik arról, hogy kire adta a voksát, és az így kapott minta alapján következtetnek a választási eredményekre

Hólabda

 A nehezen hozzáférhető populációk esetében alkalmazható az u.n. "hólabda" mintavétel: ekkor egy vizsgált személyen keresztül jutunk el a következőhöz, azon keresztül a következőhöz, és így tovább.

II. Véletlen mintavételi eljárások

Véletlen (valószínűségi) mintavételnél a sokaság valamennyi eleme ismert valószínűséggel kerülhet a mintába. Az eredményeket kivetítjük az alapsokaságra.

Típusai:

Független, azonos eloszlású (FAE) minta kiválasztása:

- ha homogén és végtelen (vagy nagyon nagy) sokaságból veszünk véletlen (visszatevéses vagy visszatevés nélküli) mintát,
- Véges sokaságból visszatevéssel választjuk ki a mintát.
- Alkalmazása: tömegtermelés minőségellenőrzésénél.
- Pl.: az 1 literes tej töltőtömegének ellenőrzésénél.

Egyszerű véletlen (EV) mintavétel:

- homogén, véges elemszámú sokaság,
- a mintát visszatevés nélkül választjuk ki,
- elemenként egyenlő valószínűségek.

Végrehajtása: a mintavételi keretből a mintaelemek kiválasztása:

- o sorsolással, ún. véletlenszám-táblázattal,
- o véletlenszám-generálással

Rétegezett mintavétel:

- A vizsgált ismérv szempontjából heterogén sokaságokat több homogén részsokaságra (rétegekre) bontjuk úgy, hogy a csoportok kiadják az egész sokaságot,
- egyetlen sokasági elem sem tartozhat két vagy több csoportba,

Az egyes rétegeken belül a minta elemeinek kiválasztása egyszerű véletlen mintavétellel történik.

A rétegzett kiválasztási technikát két alcsoportra bonthatjuk. Az arányosan rétegzett, amelyben minden réteg ugyanolyan arányt képvisel. A nem arányos (diszproporcionális) rétegzés azt jelenti, hogy a "kis arányú" rétegeket nagyobb részben szerepeltetik a mintában, mint azt részarányuk biztosítaná.

Csoportos mintavétel

- A csoportos mintavétel során homogén sokaság elemeinek (természetes vagy mesterséges) csoportjai közül egyszerű véletlen mintát veszünk, majd a kiválasztott csoportokon belül minden egyes egyedet megfigyelünk.
- Elsődleges szempont a költségtakarékosság, a megfigyelés megbízhatósága háttérbe szorul.
- Csoportos mintavétel során kétféle egység különül el:
 - Elsődleges mintavételi egység amelyre a felvétel közvetlenül irányul. (pl: helyi iskolák)
 - Végső mintavételi egység amelyre vonatkozóan következtetéseket akarunk levonni a kapott mintából. (pl.: tanulók)

Többlépcsős mintavétel

- A többlépcsős mintavételt hasonló esetekben alkalmazzuk, mint a csoportos mintavételt.
- Különbség többször ismételjük meg egymás után az egyszerű véletlen mintavételt. A mintaelemek kiválasztása több fokozatban történik.
- A mintavétel végrehajtása során kiválasztjuk az elsődleges mintavételi egységet.
- Attól függően, hogy hányszor ismételjük meg egymás után az egyszerű véletlen kiválasztást, két-, három- vagy többlépcsős mintavételről beszélhetünk.
 - Kétlépcsős mintavétel az elsődleges mintavételi egységeken belül rögtön a megfigyelni kívánt elemeket választjuk ki.
 - Három- vagy többlépcsős mintavétel az elsődleges mintavételi egységeken belül először újabb nagyobb csoportokat választunk ki, majd az így képzett csoportokból választjuk ki a mintaelemeket.

Kombinált eljárások

- Egy lépésben alkalmaznak több, eddig ismertetett mintavételi módszert.
- A kombinált eljárások külön csoportját képezik az ismétlődő felvételek, ill. a panelfelvételek.
- Alkalmazásuk ha a vizsgált sokaság szerkezetét vagy az egyes egyedek jellemzőinek időbeni változását akarjuk vizsgálni.
 - o Ismétlődő felvételek
 - Nem szükséges, hogy a mintában szereplő egyedek azonosak legyenek.
 - Egy egy időpontban a vizsgált sokaság keresztmetszetéről megbízható képet ad.
 - Végrehajtása a minta elemei néhány egymás után következő megkérdezéskor azonosak, majd előírt rend szerint cserélődnek. Pl. munkaerőfelvételek, KSH munkapiaci kutatások
 - o Panelfelvételek
 - A minta elemeinek a lehetőségek keretei között azonosaknak kell lenniük.
 - Előnye pontosabb információt ad, mint a szerkezeti változásokból levonható következtetések.

 Hátránya – a mintába került egyedek nyomon követése nehéz és a válasz megtagadása miatti torzítás gyorsan növekszik. Pl. médiakutatások

Hibák

Az adatgyűjtések, megfigyelések hibákkal járnak. Két fajta hibát különböztetünk meg:

- A **nemmintavételi hibák** azok a hibák, amelyek mind a teljes, mind a részleges megfigyeléseknél felléphetnek.
 - o Ezek matematikai eszközökkel nem kezelhetők.
 - o Ilyenek például
 - a definíciós hiba
 - a válaszadási hiba
 - a végrehajtási hiba
 - az adatrögzítési hiba.
- A **mintavételi** hiba a részleges megfigyelésből fakadó hiba. Ez a típus matematikailag kezelhető. A sokaság minden egyes egységének megfigyeléséről való lemondás ára

Megbízhatósági szint: a minta alapján számolt becslések milyen valószínűséggel lesznek igazak az alapsokaság tagjaira

Mintavételi hiba: a minta alapján becsült paraméter milyen mértékben ingadozik a valós érték körül (konfidencia intervallum)

A mintanagyság meghatározása

A minta szükséges nagyságának meghatározásában különbséget kell tennünk aszerint, hogy milyen mintavételi technikát alkalmazunk. Amig a valószínűségi minta esetén a szabályok jól körülhatároltak, addig a nem véletlen mintavételnél (ahol nem kalkulálható a mintavételi hiba) csupán hüvelykujjszabályokat tudunk megfogalmazni.

Valószínűségi mintavételnél a mintanagyság meghatározása előtt tisztázni kell:

- Mire fogjuk felhasználni az eredményeket?
- Milyen részletességű elemzéseket akarunk végezni az adatokkal?
- Milyen pontosságra van szükség az mintáknál és az összeredményeknél?

- Össze kell gyűjteni mindazokat a releváns statisztikai információkat, amelyek az alapsokaságról rendelkezésre állnak.
- Meg kell határozni, hogy a vizsgálni kívánt populáció mennyire különböző (heterogén).
- A rendelkezésre álló költségkeretet figyelembe véve lehet csak meghatározni a szükséges minta- nagyságot.

A következő táblázatban összefoglalva láthatók a mintavétellel kapcsolatban előforduló kérdések.

Szempont	Magyarázat	Példa
Mekkora az alapsokaság?	Ha az alapsokaság viszonylag nagy, akkor a minta standard hibája, azaz az eredmények megbízhatósága egyedül a minta abszolút nagyságától függ. Ez azt jelenti, hogy kisebb populációhoz relatíve nagyobb mintát kell venni, nagy populációknál pedig meglepően kis minták is elégségesek.	Egy százezer fős városnál, egy 10 milliós vagy egy 100 milliós országnál egyaránt elégséges lehet 1000 fős minta használata - lényegében ugyanolyan megbízhatóság mellett. Ez azonban fordítva is igaz: a né- hány ezer vagy néhány száz fős populációknál nem használhatunk sokkal kisebb mintákat, mint a nagy populációkon végzett vizsgálatoknál.
Mekkora az alap- sokaságban a vizsgálni kívánt jellemzők szerinti szóródás [alap- sokaság heterogenitása]?	Minél heterogénebb a sokaság a releváns jellemzőket tekintve, annál nagyobb mintanagyságra van szükség. Ilyenkor rétegzett mintavételt érdemes alkalmazni egyszerű véletlen minta helyett.	Egy termék fogyasztóinak heterogenitása a vásárlási szokásaikat tekintve azt vonja maga után, hogy minél nagyobb mintát kell vennünk ahhoz, hogy minden egyes különböző vásárlási szokással rendelkező vásárló reprezentálva legyen a mintában.
Mekkora standard mintahibát tartunk elfogadhatónak?	A közvélemény-kutatási gyakorlatban általában a maximum ±3 százalékos standard mintahibát tartják elfogadhatónak. Ez azt jelenti, hogy a kutatás révén nyert adatok a teljes mintára vonatkoztatva legfeljebb ennyivel térnek el attól a hipotetikus eredménytől, amelyet akkor kapnánk, ha a kutatás során a populáció minden egyes tagját megvizsgálnánk. Piackutatásoknál ez a hibahatár nagyobb lehet.	Egy 20 fős mintánál vagy részmintánál a mintavételi hiba akár +20 százalék, egy 100 fős mintánál pedig +10 százalék is lehet. Hogy ezek a lehetséges hibák még elviselhetők-e vagy már elviselhetetlenül megbízhatatlanná teszik-e az adatokat, azt a vizsgálat vezetőjének kell eldöntenie. Egy 1000 fős mintánál a mintahiba már csak +3,2 százalék.
Milyen mélységű elemzést akarunk végezni?	A mintán belül hány elemezni kívánt részmintát kívánunk létrehozni?	Egy 1000 fős minta is lehet túl kicsi, ha a mintát az elemzés során sok kicsi különböző részmintára akarjuk bontani, hiszen ekkor az egyes részminták megbízhatósága már kérdésessé válhat (például 16 alminta esetén az átlagos elemszám 62).
Mekkora költségvetés áll rendelkezésre?	A minta nagyságának növelése egy ponton túl már nem növeli érzékelhetően a minta pontosságát, egyenes arányban növeli viszont a vizsgálat költségét.	Ezer vagy kétezer főnél nagyobb minta kiválasztása csak akkor indokolt, ha az elemzés során a mintát sok almintára szeretnénk bontani.

5.1. táblázat Mintavételekkel kapcsolatos kérdések

Forrás: Sajtos, Mitev: SPSS kutatási és adatelemzési kézikönyv

2 SPSS alapismeretek

A menürendszer sokban hasonlít a Microsoft Office programcsomagnál megszokottakhoz, vannak olyan műveletek, melyek itt is ugyanúgy alkalmazhatóak – másolás, kivágás, beillesztés, törlés –, de találunk eltérőket is pl.: a visszavonás csak az utolsóra terjed ki, a beillesztés (PASTE) pedig nem szúr be oszlopokat és sorokat, ekkor az adatvesztés lehetősége nagyobb a figyelmetlen használat esetén.

Az SPSS 22 program megnyitásakor az alábbi ablak jelenik meg.



Az SPSS kezdő felülete

2.1 SPSS felületei

A program alapvetően két felületet használ:

- Data Editor
- Output Viewer

Data Editor (Adat szerkesztő felület/ablak) ahol adatokat tudunk bevinni, illetve módosítani/szerkeszteni.



2.1.1. ábra. Data editor ablak

Output Viewer(Eredmény Kijelző felület/ablak) minek segítségével a táblák, grafikonok, és a feldolgozott statisztikai eredmények jelennek meg nyomtatható formában.



2.1.2. ábra. Eredmény kijelző ablak (Output Viewer)

A képernyő bal alsó sarkában található kis fülekkel tudjuk váltogatni e két ablakot. Az adatszerkesztő ablakban a statisztikai adatokat meghatározzuk, bevisszük és szerkesztjük. Az eredmény kijelző ablak még nincs megjelenítve, mivel az a statisztikai feldolgozás végbemenetelét követően automatikusan megjelenik. Itt a kiszámított statisztikai mutatók és grafikus számítások kerülnek megjelenítésre. A két ablak menüje ugyanaz, viszont az ikonok különböznek. Az SPSS menüje tartalmaz néhány alapvető almenüt a Windows programokhoz, és néhány különleges funkciót is (2.1.3. ábra).



2.1.3. ábra. Az SPSS menüsora

Adatbázisunk feltöltött sorokból, úgynevezett rekordokból (vagy esetekből - case) és változókból, azaz oszlopokból áll. Az adatszerkesztő (Data Editor) két lapból áll: egyik Data View és a Variable View, amelyeket az ablak bal alsó sarkában található fülekre kattintva, illetve a CTRL+T billentyű- kombinációval lehet váltogatni (2.1.4. ábra).



2.1.4. ábra. Data view és Variable view ablakok

Az adatbázisunk primer adatokkal való feltöltése az Excel programhoz hasonlóan történik az adatnézet (*Data View*) ablakban. Az adatbázisunkat esetekkel (rekordokkal) sorokkal (Case), illetve oszlopokkal, vagyis változókkal (*Variable*) jellemezhetjük. Ez jelentheti, hogy a sokaság egy egyedét a sorok (*Cases*), míg az egyes változókat (*Variable*) az oszlopok jelentik.

🔄 Napraforgó tányérátmérő és magasság adatok.sav [DataSet7] - IBM SPSS Statistics Data Editor

<u>File</u> Edit	<u>View</u> <u>D</u> ata	Transform	<u>A</u> nalyze D	Direct <u>M</u> arket	ing <u>G</u> raphs <u>U</u>	tilities Add- <u>o</u> r	ns <u>W</u> indow	Help			
				*						ABG	
	Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	fajta	Numeric	8	0	Napraforgó fajták	{1, POMAR}	None	8	를 Right	🙈 Nominal	S Input
2	kezeles	Numeric	2	0	Kezelések	{1, Pledge +	None	8	疆 Right	💑 Nominal	🔪 Input
3	tanyer_a	Numeric	8	0	tányér átmérő	None	None	8	署 Right	Scale Scale	🔪 Input
4	magassag	Numeric	8	0	magasság	None	None	8	遍 Right	I Scale	> Input
5											
6											
7											
8											
9											
10											
11											
12											
13											

2.1.5. ábra. Variable view ablak

A **Variable View** ablakban a sorok tartalmazzák a változókat, az oszlopok pedig ezek tulajdonságait (2.1.5. ábra):

- Name: a változó rövid nevét kell megadni ékezetek nélkül. Ha a mezőt üresen hagyjuk, az automatikusan generált elnevezés a VAR00001, ahol a sorszám értéke folyamatosan nő. A változók bővebb kifejtését a Label (címke) teszi lehetővé.
- *Type:* a változó típusát és formáját kell meghatározni, minden sor egy változó, így a műveletet soronként el kell végezni a rendszer következő választási lehetőségeket adja (2.1.6. ábra):



2.1.6. ábra. Változók típusának definiálása

 Numeric (numerikus): legegyszerűbb formában jeleníti meg a számokat, ez a leggyakrabban használt forma például: 3697,40

- Comma (vessző): tizedesvesszőt ponttal (.) és ezres helyiértéket vesszővel (,) jelöli pl.: 42,711.540
- Dot (pont): tizedesvessző ponttal (,) és ezres helyiértéket ponttal (.) jelöli. pl.:
 43.711,54
- Scientific notation (tudományos alak): a számok normálalakban vannak: 1-10 közötti szám és a tíz megfelelő hatványának szorzatai. Például: 587 = 5,87E2 = 5,87*100.
- o Date (dátum): év/hónap/nap sorrendjének beállítása.
- o Dollar (dollár): pénzben mért érték jelölésére alkalmas.
- Custom currency (speciális pénzformátum): azok közül a pénzformátumok közül lehet választani, melyeket ezelőtt az OPTIONS menüben beállítottunk.
- String (szöveges változó): szöveges adatok tárolása nyílt kérdés egyéni válasz estén. Például: Miért...?
- *Width* (szélesség): A Data View ablak rekordjai mennyi karaktert tartalmaznak cellánként.
- Decimals (tizedes jegyek száma): mennyi karakter található a tizedesvessző után.
- Label (címke): a változó jelentését, vagy magát a változót lehet magyarázni itt a későbbi output táblák (Viewer és a Chart Editor) is ezt jeleníti meg, valamint a Data View is, amennyiben a változó nevére irányítjuk az egeret.
- Values (érték): a változó értékeinek definiálása itt lehetséges pl. hőmérsékleti értékek.
 Ez a Value Labels ablakban lehetséges az "Add" gombra kattintva, majd a folyamat befejeztével az "OK" gombot választani, és bezárni az ablakot (2.1.7. ábra).

File	Edit	View	Data	Transform	Analyze	Direct Market	ing Graphs L	Itilities Add-on	s Window	Help	
8	6		Ū,		× .	- -	r w				- 0
		Nam	e	Туре	Width	Decimals	Label	Values	Missing	Columns	Align
	1	fajta		Numeric	8	0	Napraforgó fajták	{1, POMAR}	None	8	Right
	2	kezeles		Numeric	2	0	Kezelések	{1, Pledge +	None	8	🚟 Right
	3	tanyer_a	i i	Numeric	8	0	tányér átmérő	None	None	8	I Right
	4	magassa	ag	Numeric	8	0	magasság	None	None	8	遭 Right
	5					- F					
	6						Value Labels				×
	7						-Value Labels				
	8						Volue				0
	9						vai <u>u</u> e. 1				Spennig
	10						Label: POMAR				
	11							2 = "ALSAN"		-	
	12						Add	3 = "EGH 527"			
	13						Change	4 = "PACIFIC"			
	14						Remove	5 = "JOANNA"			
	15						Mentove	T = "LUTA"			
	16							- LOLITA			
	17							OK	Cancel He	ID	
	18									_	
	19										
	20										

2.1.7. ábra. Változók értékcímkéinek megadása

Missing (hiányzó érték): itt olyan értéket rendelünk hozzá, amely az adat hiányát, a nem kielégítő választ mutatja pl.: magyarországi átlaghőmérsékletek esetén 50°C-vagy ennél nagyobb adat. Ha ez a hozzárendelés nem történik meg, számos hibát eredményezhet. A hiányzó értéket általában 9-esekből álló olyan számmal jelöljük, ami érték nem fordul elő az adott változóban (2.1.8. ábra).

*N	apraforg	gó tányéri	itmérő e	és magasság a	datok.sav [Da	ataSet7] - IBM S	SPSS Statistics Data	Editor						
File	Edit	View	Data	Transform	Analyze	Direct Marketi	ng <u>G</u> raphs <u>L</u>	tilities Ad	dd- <u>o</u> n:	s <u>W</u> indow	Help			
8			Ū,		¥ 📕	1 📥 🗐	r H		4			- 0	ABC	
		Na	ne	Туре	Width	Decimals	Label	Values	s	Missing	Columns	Align	Me	asur
	1	fajta		Numeric	8	0	Napraforgó fajták	{1, POMA	R}	None	8	Right	💑 Non	ninal
	2	kezeles	1	Numeric	2	0	Kezelések	{1, Pledge	e +	None	8	温 Right	💑 Non	ninal
	3	tanyer	а	Numeric	8	0	tányér átmérő	None		None	8	Right Right	Sca	le
	4	magas	sag	Numeric	8	0	magasság	None		None	8	Right	Sca	le
	5								_		_			
	6									Missing Values			×	
	7				_									
	8								Q	No missing va	lues			
	9								0	Discrete missi	ng values			
	10													
	11								0	Range plus on	e optional di	screte missing va	lue	
	12									Low:	Hig	iti:		
	13									Discrate value				
	14									L/ISciete value.				
	15									OK	Cancel	Help		
	16							l	_					
	17													
	18													

2.1.8. ábra. Missing value (hiányzó értékek) párbeszédablak

Missing value ablakban van mind erre lehetőség, ahol három változat közül választhatunk:

- No missing values: ha nem adunk meg hiányzó értéket, azt a program egy ponttal
 (.) jelöli.
- Discrete missing values: egyedi kódot adhatunk meg a hiányzó értékekre (maximum 3 darabot).
- Range plus one optional discrete missing value: itt megadhatunk egy számtartomány vagy egy tartományt és egy különálló értéket.
- *Columns* (oszlopok): A Data View oszlopszélességének a mértéke, amely nem lehet kisebb, mint amilyen hosszú a változó neve.
- Align: A Data View cellatartalmainak igazítása: jobbra, balra, középre
- *Measure* (mérési skála): a skálatípust kell megadni:
 - Scale: metrikus intervallum- vagy arányskála.
 - o Ordinal: sorrendi skála.
 - o Nominal: névleges, nominális skála



Output Viewer (Eredmény Kijelző felület/ablak)



2.1.9. ábra. Output Viewer (Eredmény kijelzőablak)

2.2 Adatbevitel

Az adatok bevitelének legegyszerűbb módja, ha beírjuk az egyes cellákba az előre definiált változónak megfelelően, míg másodlagos rögzítésnél már létező adatbázisok (Excel, dBase) importálásával történik.

Az elsődleges adatbevitel

Az elsődleges adatbevitelnél a változók definiálása az első lépés (Insert Variable), majd a rekordok (adatsorok) begépelése következik (Insert Cases). A Data menüpont alatt megtalálható az Insert Variable és az Insert Cases, amelyeket a sorok vagy oszlopok elején történő jobb egérrel történő kattintással is elő lehet hívni. Ha a Variable View ablakban alkalmazzuk az Insert Variable menüpontot, akkor már csak a változó paramétereit kell beállítani, ha azonban a Data View ablakban tesszük mindezt, akkor az új változót a legutolsó oszlopba illeszti a program, s a változó paramétereinek definiálásához az oszlop tetején dupla klikkelést kell alkalmazni (2.2.1. ábra).

A változókat célszerű előbb Variable View nézetben definiálni, és csak ezután átlépni Data View nézetbe.

ta Na	praforgo	ó tányérátmérő	és magasság ad	atok.sav [Data:	Set1] - IBM S	PSS Statistics Data E	ditor					
File	Edit	View Data	Transform	<u>A</u> nalyze C)irect <u>M</u> arketi	ing <u>G</u> raphs <u>U</u>	tilities Add- <u>o</u> n	s <u>W</u> indow	Help			
				~		P K			2		46	
		Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
	1	fajta	Numeric	8	0	Napraforgó fajták	{1, POMAR}	None	8	I Right	\delta Nominal	🔪 Input
1	2	kezeles	Numeric	2	0	Kezelések	{1, Pledge +	None	8	疆 Right	\delta Nominal	🔪 Input
	3	tanyer_a	Numeric	8	0	tányér átmérő	None	None	8	■ Right	Scale Scale	🔪 Input
4	4	magassag	Numeric	8	0	magasság	None	None	8	■ Right	Scale Scale	🔪 Input
1	5											
(6		-									
1	7											
1	8											

2.2.1. ábra. Változók definiálása

A paraméterek meghatározása egyesével, vagyis cellánként történik: name és label begépeléssel, a type, values, missing cella jobb oldalán lévő gombra kattintva a megjelenő panelt töltjük ki, a többi esetén a legördülő sáv lehetőségeit használjuk. A szöveges adatokat szöveges változókban tároljuk (string), a szám jellegű adatokat pedig numerikusban. Ha numerikus adatoknál nem szeretnénk tizedesjegyeket megjeleníteni, akkor a decimals értékét 0-ra kell megadni.

Ezt követően Data View nézetre váltunk, ahol a már meghatározott változókat töltjük ki az adatokkal.

Ha a View menüpontnál a Value Labels opciónál pipát látunk, akkor nem a változók nevei, hanem a megfelelő változó label mezőjének értéke látszik. Utóbbi pedig sokkal egyértelműbb lesz nem csak a táblázatban, de a kimeneti táblázatokban és grafikonokon is.

Másodlagos adatbevitel

Már létező adatbázisból történő adatbevitel kétféle módon valósítható meg pl. Excel fájl esetén:

- File / Open Database / New Query/ Excel
- Open/ File / Data .xls, .xlsx kiterjesztésű fájl kiválasztása

Ha az első utat (File / Open Database / New Query/ Excel) választjuk, akkor a "Next" gomb használatával egy új ablak kerül elénk – ahol a "Browse" gomb megnyomásával – az importálni kívánt (jelen esetben .xls) fájl elérési útját kell megadni. A forrásból a megfelelő adatokat tartalmazó munkalapot áthúzzuk a jobb oldali ablakba – mindezt úgy tehetjük meg, hogy az egérrel rákattintunk a mozgatni kívánt névre, majd áthúzzuk arra a helyre, ahova szeretnénk. Az SPSS program felismeri a tartalmat, így a változókat és az eseteket is helyesen értékeli. A "Next" gomb kétszeri megnyomásával megkapjuk a végeredményt. A köztes állapotban az újrakódolásra, illetve a változók meghatározására van lehetőség. Az így kapott adatbázis tartalmilag megegyezik az elsődleges adatbevitel során kapott eredménnyel, csupán a változók néhány paraméterében van különbség. Ami a .xls fájlban oszlopcím, az itt a név (name) lesz.



2.2.2. ábra. Az importálandó fájltípus kiválasztása

Ha a második módszert (Open/ File / Data - .xls, .xlsx) választjuk (2.2.2. ábra), ugyan erre az eredményre jutunk. Hátránya, hogy átalakítás előtt nem változtathatunk a változókon. Itt a program automatikusan beolvassa és a forrás fájl első sora alapján definiálja az SPSS változóit (2.2.3. ábra).

Untitled1 [DataSet0] - IBM SPS	S Statistics Data Editor						to manufacture		-C-101 ID14 CDCC Charles				
File Edit View Data T	ransform Analyze	Direct Marketing	Graphs	Utilities	Add-ons	Window	Eile Ed	it ⊻iev	ew Data Transfor	rm Analyze Direct Marketing	Graphs Utilitie	s Add- <u>o</u> ns <u>W</u>	indow <u>H</u> elp
			-	-	-		(2)		💧 🛄 r	~ ~ 🖹 🛓 🗐	TP H		
New	,	<u> 2</u>	14	6		2 📟			Name Type	e Width Decimals	Label	Values Mi	ssing Colu
<u>O</u> pen		Data					1		to Opene Data				~
Onen Database	,		Label	Va	alues	Missing	2	_	Lashier and	11411		2a 14 1-	~
open balagase		Syntax			P		4		Look III. pe	erdak	· •	9 II: II	_
Read Text Data		Cutout		_			5		Eg Fizu_regrb	(Janosa Andras)xis			
Read Cooper Data		- zaput					6	_					
Read Cognos Data		Script					8	-					
Close	Ctrl+F4						9						
-	011.0			_			10	-	-				
Save	Ctri+S						12						
Save As							13						
-				_			14	_					
Save All Data							16		File name:				Open
Event to Database							17		Files of type:	Excel (*.xls, *.xlsx, *.xlsm)		*	Paste
Experie to Database				_			18		Encoding:	SPSS Statistics Compressed (*.zsa	1)	×	Cancel
Mark File Read Only				_			20			Systat (*.syd, *.sys)			Help
Deserve Debeed		-					21			Portable (*.por) Excel (*.xis. *.xism)			
Rename Dataset							22	_		Lotus (*.w*)			
Display Data File Information	>			_			24	_		Sylik (*.SIK) dBase (*.dbf)		*	
							25						

2.2.3. ábra. Importált fájl megnyitás az Open menüponttal

3 SPSS menüpontok

🔒 Na	praforgo	ó tányérá	tmérő és	magasság a	datok.sav [[ataSet1] - IBM SPSS	Statistics Da	ata Editor							
File	Edit	View	Data	Transform	Analyze	Direct Marketing	Graphs	Utilities	Add-ons	Window	Help				
			0,	F	7	š 🛓 🗐	μ	H I		4	2			5	
		Na	me	Type	Widt	h Decimals	Label	Va	dues	Missing	Columns	Alian	Mea	sure	Role

Az SPSS programban a következő menüpontok találhatók:

- 1. File
- 2. Edit
- 3. View
- 4. Data
- 5. Transform
- 6. Analyze
- 7. Graphs
- 8. Utilities
- 9. Window
- 10. Help

3.1 File menü

A File menüben találjuk a fájlkezelő műveleteket.

1	Napratorg	to tanyera	atmero e	s magassag ad	atoksav [D	atase	(1) - IBINI S	PSS Statistics Dat	a Edi	tor		
File	Edit	View	Data	Transform	Analyze	Dir	ect <u>M</u> arket	ing <u>G</u> raphs	Util	ities Add- <u>o</u> r	is <u>Window</u>	Help
	New Open					р р	* =		ĥ			2
	Open Da	tabase				•	Decimals	Label		Values	Missing	Colum
•	Read Tex	t Data						Napraforgó fajt	ák {	1, POMAR}	None	8
	Read Co	onos Da	ta					Kezelések	{	1, Pledge +	None	8
60	Close				Ctrl+E4			tányér átmérő	1	Vone	None	8
	Caus				Del + O			magasság	1	Vone	None	8
	Dave								-			
	Save As											
-	Save All C	Data							-			
Ro	Export to	Databas	e									
0	Mark File	Read Or	nly						-			
×.	Rename	Dataset										
	Display D	Data File	Informat	ion		۶.						
۲	Cache Da	ata										
	Collect V:	ariable In	formatio	n								
0	Stop Prod	cessor			Ctrl+Period							
æ	Switch Se	erver										
	Reposito	огу				۶.						
	Set Viewe	er Output	Options	(Syntax)								
8	Print Pres	/iew				-						
A	Print			,	Ttrl+P							
0		Distan			541-1	-						
	welcome	Dialog.										
	Recently	Used Da	stal									
	Recently	Used Fil	es									
	Exit											

3.1.1. ábra. File menü

- New: új adatfájlt (Data) vagy output fájlt (Output) hozhatunk létre.
- Open: már létező adatfájlt (Data) vagy output állományt (Output) tudunk megnyitni.
- Open Database: új vagy létező SQL szervezésű adatfájlt megnyitása
- Read Text Data: szövegformátumú állomány megnyitása.
- Save: az aktív állományt arra a helyre menti, ahonnan azt megnyitottuk. A fájl előszöri mentésekor ebből a menüpontból a program automatikusan átlép a "SAVE As" menüpontba.
- Save As: az aktív állomány mentése általunk megadott helyre, néven és fájltípusban.
- Save All Dala: az összes nyitott állomány mentése.
- Mark File Read Only: az adatfájl csak olvasható fájlként történő megjelölése (későbbiekben a fájlban javítani nem lehet).
- Rename Dataset: az adatbázis elnevezése, illetve annak megváltoztatása. A fájlnév mellett az adatbázisnak adhatunk egy külön nevet, amely a fájl- név után jelenik meg. Előnyös lehet, ha egy fájlból (azonos név alatt) számos verzió létezik.
- Display Data File Information: a .sav kiterjesztésű adatfájlokról ad összesített információt egy output ablakban.
- Cache Data: ha ezt a funkciót futtatása esetén az adatokon addig senki nem tud változtatni, amíg be ne fejeztük a munkát. Nagy adatbázisok esetén az adatoknak az adatszerkesztőben való áttekintése gyorsabbá válik.
- Print: az aktív ablak nyomtatási beállításainak megadása.
- Print Preview: Nyomtatási kép.
- Swith Server: szervergépre történő csatlakozás.
- Stop Processor: a program számolási műveleteket végző egységének leállítása.
- Recently Used Data: alegutóbb használt .sav kiterjesztésű adatfájlok elérése.
- Recently Used Files: legutóbb használt nem .sav kiterjesztésű fájlok elérése.
- Exit: a program bezárása.

3.2 EDIT menüpont

🔒 Na	praforg	ó tányéra	átmérő é:	s magasság ad	atok.sav [Dat	taSet1] - IBM S	PSS Statistics Data	Editor	
Eile	Edit	View	Data	Transform	Analyze	Direct Market	ing <u>G</u> raphs	Utilities Add-or	ns <u>W</u> ir
		ndo edo		Ctrl+Z	1	i 📥 =		a 📰 🖬	
	X	ut		Ctrl+X	Width	Decimals	Label	Values	Mis
				04.0	8	0	Napraforgó fajtál	{ {1, POMAR}	None
	2	opy		CIII+C	2	0	Kezelések	{1, Pledge +	None
	3 🖬 P	aste		Ctrl+V	8	0	tányér átmérő	None	None
	4 P	aste <u>V</u> ari	ables		8	0	magasság	None	None
	5 / C	lear		Delete					
	🛙 🔛 In	isert V <u>a</u> ri	able						
-] 📰 In	isert Cas	es			-			
	AN E	ind		Ctrl+F		-			
1	4 F	ind Ne <u>x</u> t		F3					
	R	eplace		Ctrl+H					
1	S	earch Da	ata Files						
1	G	o to Ca <u>s</u>	e						
1	L G	o to Varia	able						
1	G	o to Impi	station						
1	=10	otions			-				
1	18								
1	19								

3.2.1. ábra. Edit menüpont

- Undo: az utoljára kiadott utasítás visszavonása.
- Redo: a visszavont utasítást lehet újra érvényessé tenni.
- Cut: az aktív ablakban valamely adat- vagy szövegrészt kivágása, majd más helyre történő beillesztése a "PASTE" paranccsal.
- Copy: az aktív ablakban valamely adat- vagy szövegrészt másolása, majd más helyre beilleszteni a "PASTE" paranccsal.
- Paste: beilleszti, azaz bemásolja a vágóasztalra helyezett adat- vagy szövegrészt.
- Paste Variables: előzőleg kiválasztott változók bemásolása.
- Clear: adat- vagy szövegrészek törlése. Sorok vagy oszlopok törlésekor nem keletkeznek helyükön üres mezők.
- Insert Variable: új változót (oszlopot) illeszt be attól az oszloptól balra, amelyen állunk. Ikonján oszlopok közti piros ék látható.
- Insert Cases: új eset (sor) beillesztése azon sor fölé, ahol állunk. Sorok közti piros ék az ikonja.
- Find: a DATA View ablakban aktív, változókra lehet alkalmazni, esetekre nem.
- Go to Case: a megadott esethez (sorhoz) viszi a kurzort. Ikonján egy sor fölött álló piros nyíl látható.

 Go to Variable: a megadott változóhoz viszi a kurzort. Ikonján egy sor fölött álló piros nyíl látható (3.2.2. ábra).



3.2.2. ábra. Go to Variable

 Options: E menüpont alatt adatelemek, változók, ablakok beállítására és az SPSS működésének szabályozására használható parancsokat találunk. Az egyik legfontosabb ezen belül a "General" fül, amelyen eldönthetjük, hogy az elemzések során a változók nevét (Display Name) vagy a változók jelentését (Display Labels) kívánjuk látni (3.2.3. ábra).

e <u>E</u> dit	View Data	Transform	Analyze Di	rect <u>M</u> arketing	g Graphs Utilities Add-ons Window Help	
111			-	<u> </u>	Options	×
12 : fajta	4					
	fajta	kezeles	tanyer_a	magassag	Charts Pivot Tables File Locations Scripts Multiple Imputations Syntax Editor	
166	3	6	25			
167	3	6	22		General Language Viewer Data Currency Output	
168	3	6	23		-Variable Lists	
169	3	6	22		Display labels No scientific notation for small numbers in tables	
170	3	6	23		© Display names	
171	3	7	21	18	Apply locale's digit grouping format to numeric values	
172	3	7	25	1	Hangurament sustam:	
173	3	7	23	11	O Alphabetical	
174	3	7	23	18	<u>File</u>	_
175	3	7	24	18	O Measurement level Notification:	
176	3	7	22		Raise viewer window	
177	3	7	24		Roles Scroll to new output	
178	3	7	24		To save time, some dialogs allow the use of predefined field	
179	3	7	24		dialogs.	
180	3	7	23		Use predefined roles Look and feel: SPSS Standard	
181	4	1	25	11	O Use custom assignments	
182	4	1	25	11	Open syntax window at stattap	
183	4	1	24	11	Open only one dataset at a time	
184	4	1	23	17		
185	4	1	22	11		
186	4	1	22			
187	4	1	22			
188	4	1	22			
189	4	1	24			
190	4	1	26			
191	4	3	27	11		
192	4	3	23	11	Our Councel Tobbia Lieb	

3.2.3. ábra. A változók nevének vagy jelentésének beállítása

• Viewer: az output ablakok beállítása (betűméret, betűstílus, szín).

- Az Output Labels: segítségével lehet beállítani, hogy az output ablakban megjelenő táblázatokban, grafikonokon a változó neve (Names), jelentése (Labels) vagy mindkettő (Names and Labels) szerepeljen.
- o Pivot Tables: az output ablakban megjelenő táblázatok formai beállításai.

3.3 View menü

A View menü segítségével az aktív ablak szemmel látható tulajdonságait állíthatjuk be. Igény szerint eldönthetjük, milyen ikonok, feliratok jelenjenek meg.

- **Status Bar:** az állapotsor beállítására szolgál. Megmutatja, hogy az SPSS ma- tematikai műveleteket végző egysége (processor) dolgozik-e. Ezt az ablak alsó részének közepén lehet ellenőrizni, amennyiben az állapotsor aktív állapotban van.
- **Toolbars:** a különböző ablakok eszköztárainak megjelenítése. Beállítható, hogy milyen parancsok és ikonok jelenjenek meg az ablakok felső soraiban.
- Fonts: az aktív ablakban alkalmazott karakterek betűtípusának, stílusának, méretének beállítása.
- Grid Lines: aktív állapotban szemmel látható az ablak rácsozata, ha kikapcsoljuk, akkor eltűnik.
- Value Labels: Ha aktiváljuk, a program a Variable View nézetben meghatározott változó jelentését mutatja a Data View ablakban, egyébként a változók értékeit (3.3.1. ábra).

🔚 Na	praforgo	b tányérá	itmérő e	és magasság ad	atok.sav [Da	taSet1	- IBM SPSS :	Statistics D	ata Editor	
File	Edit	View	Data	Transform	Analyze	Direc	t <u>M</u> arketing	Graphs	Utilities	Add-on
		~	<u>S</u> tatus I Toolba Menu E	Bar rs ditor		•		K	#	
			Fonts				agassag	var	var	
	1	2	- Grid Lie				172			
	2		Gild Ell	103			175			
	3	💌 1में	Value L	abels			163			
	4		Mar <u>k</u> Im	puted Data			161			
	5	ľ	Custon	nize Variable Vi	9W		173			
	6	35	Variable	es	Ctrl	+T				
	7	Р	UMAR	Pledge + F		21				
	8	P	OMAR	Pledge + F		21				
	9	P	OMAR	Pledge + F		20				
1	10	P	OMAR	Pledge + F		25				
	11	P	LSAN	Pledge + F		21	165			
. 1	12	P	LSAN	Pledge + F		23	174			
1	13	F	LSAN	Pledge + F		23	165			

3.3.1. ábra. Value Labels

• Variables / Data: a két ablak között vált.

3.4 DATA menüpont

🔄 Na	praforgo	ó tányérát	tmérő és	magasság ac	latok.sav [Da	ataSet1] - IBM SPSS	Statistics Data I	ditor					
File	Edit	View	Data	Transform	Analyze	Direct Marketing	<u>G</u> raphs <u>L</u>	tilities Add- <u>o</u>	ns <u>W</u> indow	Help			
			😡 Defi Z Set	ne <u>V</u> ariable I Measuremei	Properties nt Level for U	Jnknown			- S			A#6	
		Nar	Con	v Data Prope	erties		Label	Values	Missing	Columns	Align	Measure	Role
	1	fajta	Nev	Custom Att	ribute		aforgó fajták	{1, POMAR}	None	8	疆 Right	🗞 Nominal	S Input
	2	kezeles		no Dotoc	<u>_</u>		lések	{1, Pledge +	None	8	疆 Right	& Nominal	> Input
	3	tanyer_		ne Dates			ér átmérő	None	None	8	看 Right	Scale	> Input
-	4	magass	Den Den	ne <u>M</u> ultiple F	cesponse s	ets	asság	None	None	8	疆 Right	Scale Scale	> Input
	5		Vali	dation		•							_
	6		ider 🔡	ntify Duplicate	Cases								
	7		ider 🔝	ntify Unusual	Cases								_
	8		📆 Con	n <u>p</u> are Datas	ets				-	_			
	9		Sort	Cases									
	4		Sort	Variables			-			-			-
	2		Trar	nspose									
1	3		Mer	de Files		•						-	
1	4		Res	tructure									
1	5		Rak	e Weights									
1	6		Brou	annaity Coor	Hotobing								-
1	7		Fiu	Density Score	e matching								
1	8		Cas	e Control Ma	aching								_
1	9		Agg	regate									
2	0		Spli	t into Files									
2	1		Ort	ogonal Des	ign	• •							
2	2		📆 Cop	iy <u>D</u> ataset									
2	3		📰 Spli	t <u>F</u> ile									
2	4		E Sele	ect Cases									
2	5		4 Wei	ght Cases			1						
2	6											12	
2	7												

3.4.1. ábra. Data menü

- Copy data properties: az adattulajdonságok másolása:
 - egy külső SPSS fájlban található adatok tulajdonságainak átmásolása az aktív adatkészletbe,
 - o a jelenlegi adatkészlet tulajdonságai alapján definiálhatók további változók.
- Define Dates: Dátumformátumú változók meghatározása. A program külön változókat illeszt be az év, a hónap, a nap, valamint a másod- perc pontossággal is meghatározható időpont számára.
- Define Multiple Response Sets: többválaszos változók definiálása. Ez a funkció az ANALYZE menüpontban is megtalálható Multiple response név alatt. A kettő között az a különbség, hogy a létrehozott változókat máshol tudjuk felhasználni. Az itt definiált változók az Analyze/Tables menüponton belül használhatók fel különböző táblázatok részeként, míg az Analyze/Multiple Response/Define Sets alatt létrehozott változók elemzésére külön menüpont van (Frequence, Crosstabs), amely csak akkor aktiválódik, ha a változót már definiáltuk.
- Split File: a parancs az adatbázist egy meghatározott változó szerint "gondolatban" részekre, csoportokra bontja, s a további a statisztikai elemzéseket ezen elkülönített csoportokon végezzük (3.4.2. ábra).

ile	Edit	View	Data	Transform	Analyze	Direct Marketing	Graph	s Uti	ities	Add-ons	Wine	dow	Help			
8					a [1 📥 🗐	ų	H	*,			4		(1ର୍ଶ	0	ABG
3:																
		fajt	a	kezeles	tanyer_	a magassag	var		var	1	ar	vai		var	var	var
	1		1	1		25 172										
	2		1	1		25 175										
	3		1	🥼 Split File						×						
	4		1	-												
	5		1	💑 Kezelé:	sek [kezeles] O Analyze a	III cases, o	to not cr	eate gro	ups						
	6		1	🔗 tányér á	itmérő (tany	e Ompari	groups									
	7		1	🛷 magas	sag (magas	© Organize	output by	groups								
	8		1			Gro	ups Base	d on:								
	9		1				Napraforg	ó fajták	[fajta]							
	0		1													
	1		1													
	2		1			<u>S</u> ort the f	ile by grou	ping var	ables							
3	3		1	-		© <u>F</u> ile is al	eady sorte	d								
	4		1	Current Stat	us: Analysis	by groups is off.										
	5		1													
	6		1													
	7		1	l	OK	Paste Reset	Cancel	Help								
	8		1	3		27										
	9		1	3		26										
1	0		1	3		26										
:	1		1	4		20 114										
:	2		1	4		21 128										
	-															

3.4.2. ábra. Split fájl

Ha az adatbázisunkra ezt a menüpontot aktiváljuk, akkor a "Data Editor" ablak jobb alsó sarkán megjelenik egy felirat: pl.: "Split by fajta".

Ezzel a menüponttal tudunk az adatbázisunkban valamelyik változó szerint csoportokat képezni, az Analyze menüben alkalmazható statisztikai számítások eredményei csoportokra bontottan jelenek meg. (3.4.3. ábra)

ta *Output2 [Docu	ment2] - IBM	A SPSS Stati	istics Viewer									
<u>File Edit Vie</u>	w Data	Transform	m <u>I</u> nsert	Format	Analyze	Direct Mar	keting Gr	raphs <u>U</u> tiliti	es Add-g	ns <u>Window</u>	Help	
😑 H 🤅) 🗟	Ø	II. K					0	£	PB		+
	riptives itle lotes .ctive Datas lescriptive S	et itatist ✦	SET DIGI SORT CAS SPLIT FI DESCRIPT /STATI Descrip [DataSet	TGROUPIN ES BY : LE LAYEN IVES VAN STICS=MB tives 1] D:\An	NG=No T fajta. RED BY RIABLES EAN STD Aya\Doc	Look=None fajta. =tanyer_a DEV MIN M uments\An	TABLEREN AX. ya\Kutata	NDER=light ásmódszert	SUMMARY an excel	=None ROWSB SPSS\példá	REAK=100 SmaJ k\Napraforgó	11=0. tány
		Description Statistics										
			Napraforg	ó fajták		N	Minimum	Maximum	Mean	Std. Deviation	ר	
			Napraforg POMAR	ó fajták tányér á Valid N i	tmérő (listwise)	N 60 60	Minimum 17	Maximum 27	Mean 22,13	Std. Deviation 2,296]	
			Napraforge POMAR ALSAN	ó fajták tányér á Valid N tányér á Valid N	tmérő (listwise) tmérő (listwise)	N 60 60 60 60	Minimum 17 18	Maximum 27 25	Mean 22,13 21,22	Std. Deviation 2,296 1,617	-	
		***	Napraforge POMAR ALSAN EGH 527	ó fajták tányér á Valid N tányér á Valid N tányér á Valid N	tmérő (listwise) tmérő (listwise) tmérő (listwise)	N 60 60 60 60 60 60	Minimum 17 18 14	Maximum 27 25 25	Mean 22,13 21,22 22,18	Std. Deviation 2,296 1,617 2,259	-	
		***	Napraforge POMAR ALSAN EGH 527 PACIFIC	ó fajták tányér ál Valid N tányér ál Valid N tányér ál Valid N tányér ál Valid N	tmérő (listwise) tmérő (listwise) tmérő (listwise) tmérő (listwise)	N 60 60 60 60 60 60 60 60	Minimum 17 18 14 18	Maximum 27 25 25 25 27	Mean 22,13 21,22 22,18 23,47	Std. Deviation 2,296 1,617 2,259 1,891	-	
		111	Napraforgi POMAR ALSAN EGH 527 PACIFIC JOANNA	ó fajták tányér ár Valid N tányér ár Valid N tányér ár Valid N tányér ár Valid N tányér ár Valid N	tmérő (listwise) tmérő (listwise) tmérő (listwise) tmérő (listwise) tmérő (listwise)	N 60 60 60 60 60 60 60 60 60 50	Minimum 17 18 14 14 18 14	Maximum 27 25 25 25 27 25	Mean 22,13 21,22 22,18 23,47 20,44	Std. Deviation 2,296 1,617 2,259 1,891 2,837	-	
		**	Napraforgi POMAR ALSAN EGH 527 PACIFIC JOANNA ALBATRE	5 fajták tányér ál Valid N I tányér ál Valid N I tányér ál Valid N I tányér ál Valid N I tányér ál Valid N I	tmérő (listwise) tmérő (listwise) tmérő (listwise) tmérő (listwise) tmérő (listwise)	N 60 60 60 60 60 60 60 60 60 50 50 50 50	Minimum 17 18 14 14 18 14 12	Maximum 27 25 25 27 25 27 25 25 24	Mean 22,13 21,22 22,18 23,47 20,44 19,02	Std. Deviation 2,296 1,617 2,259 1,891 2,837 2,519	-	
		11	Napraforgi POMAR ALSAN EGH 527 PACIFIC JOANNA ALBATRE LOLITA	5 fajták tányér ár Valid N I tányér ár Valid N I	Imérő (listwise) Imérő (listwise) Imérő (listwise) Imérő (listwise) Imérő (listwise) Imérő (listwise)	N 60 60 60 60 60 60 60 60 50 50 50 50 50 50 50	Minimum 17 18 14 14 14 14 14 12 16	Maximum 27 25 25 27 27 27 25 25 24 24 28	Mean 22,13 21,22 22,18 23,47 20,44 19,02 22,36	Std. Deviation 2,296 1,617 2,259 1,891 2,837 2,519 2,625		

3.4.3. ábra. Csoportosítás utáni az eredmények megjelenítése

• Select Cases: az adatbázisból eseteket lehet véglegesen vagy ideiglenesen kizárni. Az általunk megadott feltételeknek megfelelő rekordokkal dolgozunk tovább a statisztikai elemzések során. Ekkor a feltételeknek nem megfelelő rekordok sorszáma fekete vonallal áthúzásra kerül és a jobb alsó sarokban megjelenik egy felirat: "Filter On" (amennyiben csak ideiglenesen zártuk ki a rekordjainkat az adatbázisból (3.4.4. ábra).

Napraforg	ó tányérátmérő	és magasság a	datok.sav [DataSet1]	- IBM SPSS Statistics Data Edi	tor	ana Mind	law blate									۵	×
Eus Eou		Transform		- Graphs Com			VP Telb	A	ABC								
			<u> </u>		19 🛎			1.4									
23 :															Visible: 5	of 5 Va	riables
	fajta	kezeles	tanyer_a ma	agassag filter_\$	var	var	var	Va	ar var v	ar	var var	var	var	var var	Va	r [
-	1	1	24	the Select Cases						1	r i i i i i i i i i i i i i i i i i i i	1	1				4
-8-	1	1	24						Select Cases: IT							^	
6	1	1	26		Selec	t											
1	1	1	21	Napraforgó fajták [f	- O <u>A</u>	l cases			💑 Napraforgó fajták (fa		tanyer_a >= 25						
-	1	1	21	Kezelesek [kezeles kézelesek [kezeles	0 If	condition is a	satisfied		💑 Kezelések [kezeles]								
-0-	1	1	20	magasság (magas		If tar	nyer_a >= 25		tányér átmérő [tanye					Function group:			
10	1	1	25	align tanyer_a >= 25 (FIL	OR	andom sam	ple of cases		atanver a >= 25 (FIL					All		-	
11	1	3	21		Sa	mple						7 8	9	Arithmetic			
12	1	3	20		OB	ased on time	e or case range				· <= >=	4 5	6	CDF & Noncen	tral CDF		
13	1	3	21		R	ange								Conversion	ima		
14	1	3	19	1	00	se filter varia	ble:					1 2	3	Date Arithmetic			
15	1	3	25		6							0		Date Creation			
-16-	1	3	23										<u> </u>	Contractions	Des sist 1 (ssist		
17	1	3	22		0.1						··· ~ ()	Delete	+	Eunctions and a	special variat	Dies:	
18	1	3	27		Outpu	n.											
19	1	3	26		• Eit	ter out unsel	ected cases										
20	1	3	26		O Co	py selected	cases to a new d	lata									
21	1	4	20		L			_									
22	1	4	21		ODe	lete unselec	ded cases										
23	1	4	20	Current Status: Filter ca	es by value	s of filter_\$											
-24	1	4	18														
-25-	1	4	20		Basta	Beent	Canaal										
-26	1	4	17	UK	Paste	reset	Cancer Help				Continue	Cancel H	elp				
27	1	4	18	. 0						-		1			_		-
-28	1	4	19	. 0													
29	1	4	17	. 0													_
	1	4	21	. 0													-
74	1	· · · · · · · · · · · · · · · · · · ·	24	404 0											-		Þ
Data View	variable View									_							
Cara View	Id AlfA																
											IBM SP:	SS Statistics Pr	ocessor is ready	y Unicode	ON Filter C)n	

3.4.4. ábra. Select case menüpont

- All Cases: Minden esetet bevonunk az elemzésbe, vagyis nem szűrünk (alapbeállítás).
- If condition is satisfied: Logikai feltétel alapján, egyszerűbb műveletek, relációs jelek, illetve függvények segítségével választunk ki az adatbázisból eseteket. Szűrés során a program szűrőváltozót kreál, amely újra felhasználható.
- Random sample of cases: Véletlenszerűen választ ki eseteket:
 - Approximately: az összes esetnek körülbelül hány százalékát vonjon be az elemzésbe.
 - Exactly: Pontosan megadható, hogy hány esetet vonjon be az elemzésbe az első "x" darab esetből.
- Based on time or case range: Szűrés időrend vagy sorrendiség szerint.
- Use filter variable: Szűrőfeltételként egy megadott változót használunk.
- Output: A szűrés eredményének sorsáról rendelkezhetünk itt.
 - Filter out unselected cases: A ki nem választott esetek az adatbázisban maradnak, de az elemzés során nem használjuk őket.
 - Copy selected cases to a new dataset: A kiválasztott eseteket egy új adatbázisba másoljuk.
 - Delete unselected cases: A ki nem választott eseteket töröljük az adatbázisból.
 Ennek használata a legkevésbé ajánlott.

3.5 Transform menü

E menüpont alatt különböző adatkezelési lehetőségek találhatók. Így új változókat lehet előállítani régi változók segítségével, kategorizálni, illetve az esetek ismérvértékeit lehet újrakódolni.

Compute: új változó számítása
 E menüponttal új változókat tudunk létrehozni a régi változók felhasználásával. A régi változók felhasználása jelenthet köztük lévő függvényszerű kapcsolatot, egyszerű vagy bonyolultabb logikai viszonyt.

Feladat: Nézzük meg a szarvasmarhák laktációs tejtermelését tartalmazó adatbázisban hány olyan egyed van, amelynek az átlagos zsír kg értéke nagyobb, mint 4,2 és az átlagos fehérje százalék nagyobb mint 3,8!

🍓 szarvasmar	ha laktációs adato	sav [DataSet3] - IBM SPSS Statistics Data Editor			- a ×
<u>File</u> Edit	View Data	Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help	6		
			2		
1	Name	Compute Variable	eru -	Role	
2	ellesssz	Taroet Variable: Numeric Expression:	al	> Input	
3	tejelonap	zsirfehattagzsirkg >= 4.2 & attagfeh >= 3.8		S Input	
4	szaraznap	Type & Label		> Input	
5	osszestejkg	a tuiszam		S Input	
7	atlagzsirkg	elles ssz [ellesssz]		> Input	
8	osszesfeher I			🔪 Input	
9	atlagfeh I	Posszes tej kg (ossze		> Input	
10	@305napos	O SSZES ZSIF Kg [OSSZ - <= >= 4 5 6 Search Significance		> Input	
12	@305napos	🖉 osszes feherie ka la 🔹 🔹 = 🖂 1 2 3 Statistical		> Input	
13	@305napos I	Compute Variable: Type and X		> Input	
14	@305napos I	Label Functions and Special Variables		S Input	
15	legnagyobb	Ctvar	1	S Input	
16		O Use expression as label Max Mean			
18		Type			
19		Min Min Sd			
20		Sum			
21		Variance			
22		Continue Cancel Help n)			
24			1		
25		OK Paste Reset Cancel Help			
26					
27					
29					
	4		_		Þ
Data View	Variable View				
					IBM SPSS Statistics Processor is ready Unicode:ON
م 🗄		🐵 🗾 🧔 🏟 🔇 🖊			へ (〕 (虎 中)) 18:23 口

3.5.1. ábra. Új változó számítása a Compute menüponttal

Megoldás lépései (3.5.1. ábra):

- Alkalmazzuk a "Transform/Compute Variable" menüpontot és adjuk meg a "Numeric Expresson" panelben azt a képletet, amely segítségével a régi változókból az új kiszámolható.
- 2. A "Target Variable" dobozba pedig az új változó nevét írjuk.
- 3. A "Type&Label" gombra klikkelve a változó bővebb jelentését adhatjuk meg a "Label"-ben.
- 4. A szűrőfeltételek megadása az alapablakban lehetséges, ha az "If" gombra kattintunk.
- 5. A "Continue" majd az "Ok" egymást követő lenyomásával a végeredmény megjelenik
- Visual Bander: változók kategorizálása

Az adatok (ismérvértékek) tömörítésére van szükség, hogy a sokaság összetételéről szerkezetéről, belső arányairól megfelelő képet kapjunk. Ennek legelterjedtebb módja a sokaság egységeinek mennyiségi ismérv szerinti osztályozása, csoportosítása.

Az eljárás segítségével egy sokaságot jellemző folytonos mennyiségi ismérv értékei egyedi kategóriákba, osztályokba sorolhatók.

Ennek létrehozásában segít a Visual Bander parancs, amely lehetővé teszi, hogy a folytonos változót gyorsan és egyszerűen kategóriákra bontsuk. Ehhez egy hisztogramot rajzol, ahol megadhatók a kategóriák alsó és felső értékei.

Feladat: Kategorizáljuk a napraforgó tányérétmérő adatbázisunkban a tányérátmérő változót. A megoldás lépései az alábbi ábrákon láthatók:

- Transform/Visual Binning menü kiválasztása
- Tegyük át a kategorizálandó a tányérátmérő változót a Variable to Bin ablakba. A Variables ablakban, csak azok a változók vannak feltüntetve, amelyek ordinálisak vagy metrikusak, hiszen a nominális változóknál nincs értelme a kategorizálásnak) és nyomjuk meg a Continue gombot. (3.5.2. ábra)

Eile Edit	View Data	Transform Analyze	Direct Marketing	Graphs	Utilities	A	<u>File</u> Edit	View	Data	Transform	Analyze	Direct Marke	eting <u>G</u> raphs	; <u>U</u> til	ities A
		<u>Compute Variable</u> . Programmability Tr	ansformation		M 🕈	1 13	2			F	a	🖥 📥 🗄		М	*
6:		Count Values withi	n Casas				6:						10		
	fajta		1000000		var			fajt	a	kezeles	tanyer_a	a magass	ag var	_	var
1	1	Snin values			Î		1	🛃 Vis	sual Binni	ng				×	
2	1	Recode into Same	Variables				2	-	Palact th	a variables v	those walves	s will be groun	and into hing		
3	1	Recode into Differe	nt Variables				3	. 🛡	Data will	be scanned	when you d	lick Continue.	ped into bins.		
4	1	Automatic Recode.					4	-	The Varia	ables list bel	ow contains	all numeric o	rdinal and scale	e –	
5	1	Create Dummy Var	ishlee				6		Variable	s. S:		Variables	to Bin:		
6		B. Butternet Directory					7	1	🖉 mag	asság (mag	as	🖋 tányé	ir átmérő (tanye		
7	1	P: visual Binning					8								
9	1	🔀 Optimal Binning					9								
0		Prepare Data for M	odeling	•			10				_				
9		Rank Cases			-		11	1			-				
10		🗎 Date and Time Wiz	ard				12								
11	1	Create Time Parier					13								
12	1	Create Time Series	2		-		14								
13	1	Replace Missing V	alues				15								
14	1	🛞 Random Number 🤇	Senerators				16	-	📃 Limit	number of c	ases scann	ed to:			
15	1	Run Pending Trans	sforms	Ctrl+G			17	-		Continue	Cancel	Help			
16	1	5	23		1		18	-			Jesuncer	ac		_	
17	1	3	22 .				19		1	3		20			
18	1	3	27 .				20	-	1	3		20	114		
19	1	3	26 .				21	-	1	4		21	128		
20			00				22			-			120		

3.5.2. ábra. Tányérátmérő változó kategorizálásának lépései

 Ezután a Scanned Variable List ablakban klikkeljünk a "tányérátmérő" változóra, s megjelenik a változót ábrázoló hisztogram, amelynek segítségével eldönthetjük, hogy hány kategóriát hozzunk létre. A hisztogram megmutatja az eloszlás képét, vagyis azt, mely érték(ek) körül sűrűsödnek az adatok. (3.5.3. ábra)



3.5.3. ábra. Eloszlás hisztogramja

• Ezután meg kell határoznunk az osztópontokat (Make Cutpoints).

Az osztópontok meghatározása háromféleképpen történhet:

 Azonos szélességű intervallumok (Equal Widht Intervals), ahol megadhatjuk az első osztópontot, az osztópontok számát, valamint a szélességet is (hisztogram segít).

- Egyenlő percentilisek alapján (Equal Ppercentiles). Itt nem feltétlenül lesznek azonos szélességűek az intervallumok, ellenben ugyanannyi esetet tartalmaznak.
 Ebben az esetben 2 osztópontra (Number of Cutpoints) van szükségünk.
- Átlag és szóras alapján (Cutpoints at Mean and Selected Stadard Deviations Based on Scanned Cases). Itt meghatározhatjuk, hogy az átlag körül hány szórásnyira (1, 2, 3) legyenek az osztópontok.

Az első módon kategorizáljuk a változónkat.



3.5.4. ábra. Osztópontok meghatározása

Az Apply gomb megnyomása után visszatérünk a kiinduló ábrához, ahol láthatjuk, hogy az értékeknél (Value) megjelent a két osztópont értéke, amelyeket a hisztogramon pedig kék vonal jelöl. (3.5.5. ábra)

🔚 *Nap	praforge	ó tányérátmér	5 és magassi	ág adatok	.sav [Data]	Set1] - IBM SPS	S Statistics Data Edit	litor	
Eile	Edit	View Data	Transfo	rm <u>A</u> na	alyze D	irect Marketing	Graphs Utiliti	ities Add- <u>o</u> ns <u>Window H</u> elp	
				2			P H	👪 🔛 🚍 🖧 🧮 🛃 ⊘ 🌑 🦇	
5:									
		fajta	kezele	s ta	nyer_a	magassag	tanyeratm	var var var var var var var var	va
1				1	25	172	8	8 🚱 Visual Binning X	
2				1	25	175	8	8	
3				1	21	163	6	6 Scanned Variable List: Name: Label:	
4				1	24	161	7	7 Vanyer atmerő (tanyer Current Variable: tanyer_a tányér átmérő	
5				1	24	173	7	7 Binned Variable: tanyeratm tányér átmérő (Binned)	
6				1	26		8	8 Minimum: 12 Nonmissing Values Maximum: 28	
7				1	21		6	6	
8				1	21		6		
9				1	20		6		
10)			1	25		8	8	
11	1			3	21	174	6	6 13,33 16,00 18,67 21,33 24,00 26,67	
12	2			3	20	160	5	5 Enter interval cutpoints or click Make Cutpoints for automatic intervals. A cutpoint value	
13	3			3	21	160	6	6 Grid: Grid	
14	L.			3	19	170	5	5 Cases Scanned: Japa Value Label Upper Endpoints	
15	5			3	25	166	8	8 13,0 <= 13 0 Included (<=)	
16	5	-		3	23		7	7 Missing Values: 0 2 14,914 - 15 0 Excluded (<)	
17				3	22		6	6 Copy Bins 4 18,6 18 - 19	
18	3			3	27		9	9 From Another Variable. 6 20,5 20 - 21 Make Cutpoints	
19)			3	26		8	8 7 22,422-22 Make Labels	
20)			3	26		8	8 26,125-26 Reverse scale	
21	1	1		4	20	114	5	5 OK Paste Reset Cancel Help	
22	2			4	21	128	6		
23	3			4	20	126	5	5	
24					10	424			

3.5.5. ábra. Kategóriák mutatása kék osztópontok segítségével

A Make Labels paranccsal a program automatikusan hozzárendeli az értékekhez (Value) a címkét (Label). Majd nevezzük el a kategorizált változót (Banded Variable) "tanyeratm", az ok gomb megnyomásával a ábrán látható módon az utolsó oszlopban létrejött egy új változó 8 kategóriával.

Ha a Analyze/Frequencies menüpont alkalmazásával megmutatható, hogy a mennyiségi ismérv szerint képzett egy-egy osztályközökbe a mintának hány egysége tartozik. (3.5.6 ábra).



3.5.6. ábra. Gyakorisági eloszlás meghatározása

A gyakorisági eloszlás a 3.5.7. ábrán látható.



3.5.7. ábra. A kategórizálás eredménye grafikonon ábrázolva

3.6 Graphs menü

A statisztikai elemzésekből nyert adatok gyors, szemléletes megjelenítését segítik a menüpont alatt található különböző grafikonok, ábrák, diagramok (3.6.1 ábra).

ta Na	apraforg	ó tányéra	átmérő	és magasság ad	datok.sav [Da	taSet1] - IBM S	PSS St	atistics Da	ta Editor					
File	Edit	View	Data	Transform	Analyze	Direct Market	ing	Graphs	Utilities	Add-ons	Window	Help		
8					~	ä 📥 =		💼 <u>C</u> har	t Builder hboard Ten	nnlate Cho	oser			ARG
		Na	ame	Type	Width	Decimals		Com	nore Subar	oune		Columns	Align	Measu
	1	fajta		Numeric	8	0	Napr	Com	pare ouvyr	oups		8	Right	\delta Nominal
	2	kezele	s	Numeric	2	0	Keze	Regr	ession vari	able Plots		8	I Right	🔥 Nominal
	3	tanyer	а	Numeric	8	0	tány	Lega	cy Dialogs			🚹 <u>B</u> ar		Scale 8
1	4	magas	sag	Numeric	8	0	mag	asság	None	1	lone	11 <u>3</u> -D Ba	r	Scale 8
	5											Line		
	6											Area		
	7											Pio		
	8													
	9											High-L	ow	
	10											Boxplot	t	
	11											Error B	ar	
	12											Popula	tion Pyramid	
	13											Scatter	Dot	_
	14											Histor	am	
	15]										. ijstogi		
	16													
	17													

3.6.1. ábra. Graphs menü

4 Leíró statisztika

Statisztikában a megfigyelés és mérés tárgyát képező megfigyelési egységeket *egyedek*nek nevezzük. A statisztikai megfigyelés tárgyát képző egyedek összességét statisztikai *sokaság*nak is nevezzük. A statisztikai sokaságra használatos még a populáció elnevezés is. Kiemelt fontossággal bír a statisztikában a sokaság egyedeire vonatkozó tulajdonságok és jellemzők, melyeket *változók*nak vagy *ismérvek*nek nevezünk. Az ismérvek vagy változók lehetséges kimeneteit *ismérvváltozatoknak* nevezzük. Az ismérvek/változók csoportosítása többfélképpen lehetséges.

Változók (ismérvek) csoportosítása tulajdonság alapján:

- alternatív ismérvek
- időbeli ismérvek
- területi ismérvek
- minőségi ismérvek
- mennyiségi ismérvek

Léteznek statisztikai módszertanok pl. a regresszióanalízis, amelyek megkövetelik a változók más jellegű definiálását is. Ennek megfelelően megkülönböztetünk *függő* és *független* változókat. A függő változót minden esetben a független változó határozza meg, ok és okozat kapcsolat áll fenn közöttük.

- Független változók: azok a változók, amelyekről úgy véljük, meghatározó szerepet játszanak a minket érdeklő problémában. Ezen változók értékeit (attribútumait) mi változtatjuk (illetve választjuk) meg.
- Függő változók: azok a változók, amelyek "viselkedésére" (eloszlására) kíváncsiak vagyunk.

A változók csoportosítása felvett értékek alapján:

- *diszkrét*: csak bizonyos (általában véges számú) értékeket vehetnek fel pl. osztályzatok, célbaérés sorrendje, gólok száma.
- *folytonos:* bizonyos intervallumon belül és a mérési pontosságon belül bármilyen értéket felvehetnek pl.: testmagasság, testsúly, futási sebesség.

A gyakorlati adatelemzéshez, főleg a számítógéppel támogatott feldolgozások esetén a változók mérési szintjét, illetve a mérési skálákat pontosan meg kell határozni. (4.1. ábra)

A mérési skálák típusai:

- nominális (névleges) skála,
- ordinális (sorrendi) skála,
- intervallum skála,
- arányskála.



Forrás: Ács Pongrácz: Gyakorlati adatelemzés

4.1. ábra. Az ismérvek és mérési skálák összefüggése és azok jelölése az SPSS programban

4.1 Leíró statisztika (Descriptive Statistics):

A leíró statisztika, vagy a leíró statisztikai elemzés azt jelenti, hogy az adathalmazunkból minden egyes változót egyenként megvizsgálunk a neki megfelelő statisztikai mutatókkal. A leíró statisztikai elemzést egyváltozós elemzésnek is szokták nevezni.

Célja lehet egy változó jellemzőinek a bemutatása, de ugyanakkor egy nagyobb adatstruktúrába való elsődleges betekintés is.

Úgyis fogalmazhatunk, hogy a leíró statisztika az információk összegyűjtését, tömör, számszerű jellemzését szolgáló módszereket foglalja magában:

- adatgyűjtés, számszerű információk gyűjtése,
- adatok rögzítése, jellemzése,
- csoportosítása, osztályozása,
- adatok ábrázolása,
- egyszerűbb aritmetikai műveletek elvégzése,
- eredmények megjelenítése.

Gyakoriság (Frequence): A gyakoriság azt mutatja meg, hogy a mennyiségi ismérv szerint képzett egy-egy osztályba (osztályközbe) a sokaságnak hány egysége (f_i) tartozik.



4.1.1. ábra. Gyakoriság táblázatban és grafikonon ábrázolva

Gyakorisági táblázat: A statisztikai adatsokaságban előforduló lehetséges értékeket a gyakoriságukkal együtt egy táblázatba rendezzük. (4.1.1. ábra)

Egy statisztikai vizsgálat során egy kísérletet mindig többször végeznek el. Az egyes események bekövetkezési számát az esemény *gyakoriságának* nevezzük.

Relatív gyakoriság (Percent): A relatív gyakoriság azt mutatja meg, hogy a mennyiségi ismérv szerint képzett egy-egy osztályba (osztályközbe) a sokaságnak hányad része (hány százaléka) tartozik.

Kumulatív relatív gyakoriság (Comulative Percent): A relatív gyakoriságok halmozott összeadása, kumulálása.

A kumulált gyakoriságok (jele: f_i), ill. kumulált relatív gyakoriságok (jele: g_i), adatai azt mutatják meg, hogy az adott osztályköz felső határának megfelelő és annál kisebb ismérvértékek hányszor (f_i), ill. milyen arányban (g_i) fordulnak elő.

Gyakorisági eloszlás: a csoportok és a hozzájuk tartozó gyakoriságok összessége. A statisztikai adatsokaságban előforduló lehetséges értékeket a gyakoriságukkal együtt gyakorisági eloszlásnak nevezzük.

Egyváltozós elemzések

Az egyváltozós elemzés minden adatbázis-elemzés első lépései között szerepel. Ennek során a változókat egyenként, egymástól függetlenül elemezzük. *Változó* alatt a vizsgált jelenség valamely kiválasztott számszerű tulajdonságát értjük. A statisztikai sokaság mérete általában nagy, ezért fontos, hogy néhány számmal tömören tudjuk jellemezni az összegyűjtött adatokat. Ezeket a számokat *statisztikai mutató*knak nevezzük.

Az egyváltozós elemzések során a leggyakrabban alkalmazott mutatókat négy csoportba sorolhatjuk:

- helyzetmutatók,
- szóródási mutatók,
- alakmutatók,
- egyéb mutatószámok.

A következő táblázat mutatószámok csoportosítását tartalmazza (4.1.1).

Helyzetmutatók	Szóródási mutatók	Alakmutatók	Egyéb mutatók
Középértékek: átlag módusz medián Kvantilisek	Terjedelem Variancia Szórás Relatív szórás	Csúcsosság Ferdeség	Elemek száma Össszeg Minimum Maximum

4.1.1. táblázat. Mutatószámok csoportosítása

Az SPSS-ben a leíró statisztikai mutatókat az alábbi táblázat tartalmazza.

		Statistic	Std. Error
magasság	Mean	174,27	1,044
	95% Confidence IntervalLower Bound	172,21	
	for Mean Upper Bound	176,32	
	5% Trimmed Mean	174,49	
	Median	174,00	
	Variance	212,382	
	Std. Deviation	14,573	
	Minimum	114	
	Maximum	227	
	Range	113	
	Interquartile Range	11	
	Skewness	-,367	,174
	Kurtosis	3,327	,346

Descriptives

Átlagok

Számtani átlag (Mean)

A számtani átlag az a szám, amellyel az egyes átlagolandó értékeket helyettesítve azok összege változatlan marad.



A számtani átlag a leggyakrabban alkalmazott középérték, de nem minden esetben alkalmazható, a kiugró értékekre való érzékenysége miatt. Alkalmazása leginkább az arány és intervallum skála típusú adatoknál a legrelevánsabb.

Medián (Median)

A medián a sorba rendezett mennyiségi ismérvnek (rangsor) az az értéke, amelynél ugyanannyi kisebb, mint nagyobb érték fordul elő.

Meghatározása:

Rangsorból:

- ha a megfigyelt sokaság elemszáma páratlan: akkor a medián a rangsor ⁿ⁺¹/₂ -edik ismérvértékkel azonos.
- Ha a sokaság elemszáma páros, akkor a medián a két középső ismérvértékek számtani átlaga.

Módusz (Mode)

A módusz az ismérvértékek tipikus, leginkább jellemző értékét jelöli.

Meghatározása:

- Diszkrét mennyiségi ismérv módusza a sokaságban leggyakrabban előforduló ismérvérték.
- Folytonos mennyiségi ismérv módusza ott található, ahol az előforduló értékek legjobban sűrűsödnek, ahol a gyakorisági görbének maximuma van.

Alkalmazása leginkább a diszkrét, kategorizált változók esetén megfelelő, de alkalmazható a skála típusú változók esetében is.

Látható, hogy a skálatípusoknak megvan a legmegfelelőbb középértéke, vagyis a nominális skáláknál a módusz, az ordinális skáláknál a medián, az arányskálák esetében az összes középérték alkalmazható.

Kvantilisek:

Speciális helyzetmutatók, a medián általánosításai. Osztópontok segítségével a növekvő sorrendbe állított adataink egyenlő gyakoriságú osztályokra bonthatók.

Legyen 0 < q < 1. Ha a rangsorba rendezett sokaságot egy X ismérvérték q:(1-q) arányban osztja ketté, akkor ezt az ismérvértéket q-adik kvantilisnek nevezzük, és Q_q -val jelöljük. Fajtái:

- Tercilisek: Q_{1/3}=T₁, Q_{2/3}=T₂,
- Kvartilisek: $Q_{1/4}=Q_1$, $Q_{1/2}=Q_2=Me$, $Q_{3/4}=Q_3$,
- Kvintilisek: $Q_{1/5}=K_1$, $Q_{2/5}=K_2$, $Q_{3/5}=K_3$, $Q_{4/5}=K_4$,
- Decilisek: Q_{1/10}=D₁, Q_{2/10}=D₂, ..., Q_{9/10}=D₉,
- Percentilicisek: Q_{1/100}=P₁, Q_{2/100}=P₂, ..., Q_{99/100}=P₉₉.

azaz:

- A medián 2 egyenlő részre osztja a nagyság szerint sorba rendezett sokaságot 1 osztópont segítségével.
- A tercilis 3 egyenlő részre osztja a nagyság szerint sorba rendezett sokaságot 2 osztópont segítségével.
- A kvartilis (quartilis) 4 egyenlő részre osztja a nagyság szerint sorba rendezett sokaságot 3 osztópont segítségével.
- A kvintilis (quintilis): 5 egyenlő részre osztja a nagyság szerint sorba rendezett sokaságot 4 osztópont segítségével.
- A decilis: 10 egyenlő részre osztja a nagyság szerint sorba rendezett sokaságot 9 osztópont segítségével.
- A percentilis: 100 egyenlő részre osztja a nagyság szerint sorba rendezett sokaságot 99 osztópont segítségével.

Szóródásmutatók

Szóródáson azonos fajta számszerű adatok különbözőségét értjük, melyek azt mutatják meg, hogy az adatok az átlagtól kevésbé vagy jobban térnek el, azaz hogy az átlag körül mennyire szóródnak az adatok.

Szóródási mérőszámok:

- szóródás terjedelme,
- átlagos eltérés,
- variancia,
- szórás,
- átlagos különbség,
- relatív szórás.

Terjedelem (*Range*): a legnagyobb és legkisebb érték különbsége, azaz annak az intervallumnak a teljes hossza, amelyen belül a tényleges ismérvértékek elhelyezkednek.

$$R = x_{max} - x_{min}$$

Kizárólag arány és intervallumskálán alkalmazzuk.

A szórás terjedelménél szokás még az interkvartilis terjedelmet is említeni. Az *interkvartilis terjedelem* azt az intervallumot jelöli, ahol az összes érték középső 50 %-a helyezkedik el.

$$TQ = Q_3 - Q_1$$

Korrigált terjedelem: A terjedelem egy-egy adatra nagyon érzékeny, tehát nagyon nagy lehet ez az érték, ha van egy kiugró adat a többi között (nagyon nagy vagy nagyon kicsi), valójában pedig az adatok egy szám környékén tömörülhetnek. Ez kiküszöbölhető úgy, hogy a legnagyobb és legkisebb adatot. Amennyiben az alsó és felső 1-1%-ot hagyjuk el, a kapott eredményt alsó illetve felső centilisnek, míg az alsó és felső 10-10% elhagyása esetén alsó illetve felső decilisnek nevezzük. Tehát alsó decilis esetében a rangsorba rendezett adatok egytizede kisebb és kilenctizede nagyobb. Felső decilis esetén pedig fordítva. A minta nagysága határozza meg, hogy melyiknek van értelme.

Átlagos eltérés: Azt mutatja meg, hogy az egyes ismérvértékek átlagosan mennyivel térnek el a számtani átlagtól. Hátránya, hogy az eltérések iránya, azaz előjele befolyásolja az értékét.

Szórás (*Standard Deviation*): Azt mutatja meg, hogy az adatok mennyire szóródnak az átlag körül, mennyire heterogén a minta. Valójában az egyes értékek átlagtól való eltérésének négyzetes átlaga. A szórás mindig nemnegatív szám (pozitív vagy nulla).

$$s = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \overline{X})^2}{n - 1}} = \sqrt{\frac{\sum_{i=1}^{N} d_i^2}{n - 1}}$$
$$s = \sqrt{\frac{\frac{\sum_{i=1}^{k} f_i (X_i - \overline{X})^2}{(\sum_{i=1}^{k} f_i) - 1}} = \sqrt{\frac{\frac{\sum_{i=1}^{k} f_i d_i^2}{(\sum_{i=1}^{k} f_i) - 1}},$$

Variancia (Variance): A szórás négyzete, szokás szórásnégyzetnek is nevezni. A négyzetfüggvény miatt hangsúlyosabban emeli ki az eltéréseket. A variancia számlálója az eltérés-négyzetösszeg (jele: SS, SQ)

$$SS = \sum_{i=1}^{n} \left(x_{i} - \overline{x}\right)^{2} = \sum_{i=1}^{N} x_{i}^{2} - \frac{\left(\sum_{i=1}^{N} x_{i}\right)^{2}}{n},$$

$$SS = \sum_{i=1}^{k} f_{i} \left(x_{i} - \overline{x}\right)^{2} = \sum_{i=1}^{k} f_{i} x_{i}^{2} - \frac{\left(\sum_{i=1}^{k} f_{i} x_{i}\right)^{2}}{\sum_{i=1}^{k} f_{i}}$$

Alakmutatók

Ferdeség (*Skewness*): Az eloszlás alakját leíró mutatószám. Az eloszlásnak az a tulajdonsága, hogy milyen irányban és mennyire tér el a szimmetrikus eloszlástól. A szimmetrikus eloszlás ferdesége 0. Ha a gyakorisági eloszlásnak az oszlopdiagramon történő (hisztogram) ábrázolása alapján az eloszlás jobbra, azaz pozitív értékek irányába elnyúltabb, jobbra ferdének (skewed to right), ha balra, azaz a negatív értékek irányába torzított, akkor balra ferdének nevezzük (skewed to left).

A következő ábráról (4.1.2. ábra) mutatja a ferdeséget.



Forrás: Korpás Attiláné: Általános statisztika I.

4.1.2. ábra. Szimmetrikus és aszimmetrikus eloszlások jellemzői

Csúcsosság (*Kurtosis*): Az eloszlás alakját vertikálisan leíró mutatószám. Relatív fogalom, azt jelzi, hogy az eloszlás az azonos középértékű és szórású normális eloszláshoz viszonyítva az eloszlás csúcsos (jobban tömörül) vagy lapos (kevésbé tömörül). A pozitív értékek viszonylag csúcsos, míg a negatív értékek viszonylag lapos elosztást jeleznek.

Egyéb mutatószámok

Összeg (Sum): A mintában lévő elemek összege.
Esetek száma (Number of Cases): A megfigyelt esetek száma, a minta nagysága.
Minimum (Minimum): A mintában lévő elemek közül a legkisebb elem.
Maximum (Maximum): A mintában lévő elemek közül a legnagyobb elem.

Az alábbi táblázatból láthatjuk, hogy mely mutatókat alkalmazhatunk a különböző skálákon mért vizsgálatkor.

Változók Mutatók	Intervallum	Ordinális	Nominális
Középértékek	Átlag, (Módusz, Medián)	Medián (Módusz)	Módusz
Szóródási mutatók	Szóródás, Variancia	Terjedelem	Gyakoriság, relatív gyakoriság
Alakmutatók	Ferdeség, Csúcsosság	-	-
Egyéb mutatók	Minimum, Maximum	Minimum, Maximum	-

4.1.2. táblázat: Statisztikai mutatók alkalmazhatósága különböző skálákon mért változók esetén

Forrás: Csallner: Bevezetés az SPSS statisztikai programcsomag használatába

5 Leíró statisztika az SPSS-ben

Az SPSS programban három elérési útvonalon keresztül van lehetőségünk a leíró statisztika elkészítésére (5.1. ábra):

- Analyze/Descriptive Statistcs /Frequencies,
- Analyze/Descriptive Statistcs/ Descriptive,
- Analyze/Descriptive Statistcs /Explore.

ta */	🦆 •Analizishez Napraforgó tányérátmérő és magasság adatoksav [DataSet1] - IBM SPSS Statistics Data Editor															
File	Edit	View	Data	Transform	<u>A</u> nalyze	Direct Marketing	Graphs	Utilities	Add-ons	Window	Help					
8			00,		Repo	orts criptive Statistics	۲ ۲	Freque	encies	-			ABS			
		Nan	ne	Type	Table	es		Deser	intius e	Aissing	Columns	Align	Measure	Role		
	1	fajta		Numeric	Com	pare Means		Desci	puves	e	8	疆 Right	\delta Nominal	S Input		
	2	kezeles		Numeric	Gene	General Linear Model		A Explore		9	8	I Right	\delta Nominal	> Input		
	3	tanyer_a	1	Numeric	Gene	Generalized Linear Models	Generalized Linear Models		Cross	tabs	9	8	疆 Right	Scale 8	S Input	
1	4	magass	ag	Numeric	Mixed Models		•	TURF	Analysis	e	8	疆 Right	I Scale	> Input		
	5				Corre	alata		Ratio	24							
	6				Bearcassion		Begraggion		<i></i>	P-P PI	ots					
	7]			Regi	ession		0-0 P	lots							
	8]			Logi	near										
1	9				Neur	al Networks										
	10				Clas	sijy										
	11				Dime	ension Reduction	*									
	12				Scale	9	*									
	13				Nonp	parametric Tests										

5.1. ábra. Analyze menü

Annak eldöntésére, hogy melyik módszerrel számítjuk a leíró statisztikai módszereket kötelezően alkalmazandó szabály nincsen.

- A *Descriptive* modult leggyakrabban az intervallum vagy arányskálán (SPSS-ben skála) mért változók esetében használjuk, abban az esetben, ha gyakorisági táblára nincsen szükségünk.
- A Frequenciens modul leginkább nominális vagy ordinális skálán mért változók esetében használandó, amikor gyakorisági táblára és grafikus ábrázolásra is szükségünk van. Természetesen ebben az opcióban is vizsgálhatunk arány vagy intervallumskálán mért változókat, azonban talán az eredményközlés során készített összefoglaló táblázat nem olyan látványos, mint a *Descriptive* modulban.
- Az *Explore* modulban a már megismert módszertanokon túl, leíró statisztikai mutatókat is számoltathatunk, ahol a mintát részsokaságokra bonthatjuk.

A következő táblázat összefoglalva mutatja az SPSS programban található utasítások alapján, a leíró statisztika eszköztárát a megfelelő mérési skálák vonatkozásában (a táblázatban a D jelöli a *Descriptive* és az F a *Frequenciens* modult).

5.1.	. táblázat: leíró statisztikában használt mutatók és skálatípusok összefüggése, a	az SPSS-
	ben használt modulok	

Marta 4/1	Normán (linolo (lo		Intervallum vagy			
Mutatok	Nominalis skala	Ordinalis Skala	arányskála			
Középértékek	Módusz	Medián, (Módusz)	Átlagok (MEDIÁN,			
	FREQUENCIES	FREQUENCIES	MÓDUSZ)			
Szóródási		Minimum, Maximum	Minimum, Maximum			
mutatószámok		DECRIPTIVES,	DECRIPTIVES,			
		FREQUENCIES	FREQUENCIES			
Eloszlási	Gyakoriság, relatív	Terjedelem	Szórás, variancia			
mutatószámok	gyakoriság		DECRIPTIVES,			
	FREQUENCIES		FREQUENCIES			
Egyéb mutatók			Ferdeség, csúcsosság			
Grafikus	Oszlop és		Hisztogram			
megjelenítés	kördiagram		FREQUENCIES			
	(gyakoriságra)					
	FREQUENCIES					

Forrás: Ács Pongrácz: Gyakorlati adatelemzés

5.1 Analyze/Descriptive Statistcs /Frequencies

Analizish	ez Napraforgó tá	nyérátmérő és m	lagasság adal	tok.sav [DataS	et1] - IBM SPSS Stat	tistics Da	ta Editor					
le <u>E</u> dit	View Data	Transform	Analyze	Direct Market	ing <u>G</u> raphs <u>L</u>	Itilities	Add-on	s <u>Window</u>	Help			
2		, in 1	~	1 📥 =				4	2		● 4 6	
	Name	Туре	Width	Decimals	Label	Va	lues	Missing	Columns	Align	Measure	Role
1	fajta	Numeric	8	0	Napraforgó fajták	{1, P0	MAR}	None	8	3 Right	🙈 Nominal	🔪 Input
2	kezeles	Numeric	2	0	Kezelések	{1, 1.	kezel	None	8	疆 Right	💑 Nominal	🔪 Input
3	tanyer_a	Numeric	8	0	tányér átmérő	None		None	8	3 Right	Scale 8	🔪 Input
4	magassag	Numeric	8	0	magasság	None	-	None	8	a Right	Scale 🔗	🔪 Input
5	🚺 🤹 Frequen	cies				×	to rea	europeines Ctatie	tion :			
6			11-	2-11-1-1-1			rie rie	quencies, statis	ucs			^
7	A Manra	foraó foilák lfo		tápyár átmi	Stati	stics	Perc	entile Values -		Ce	entral Tendency	
8	Kezelések [kezeles]							1	Mean			
9	Format Cut points for: 10 equal groups 🗹 Median											
10	Percentile(s)											
11						110		Add			Sum	
12					Boot	strap		Mun				
13								Change				
14			JL					Remove				
15	Display	frequency table	s									
16			Pacta Pa	cat Cance	Halp						Values are group m	idpoints
17			Laste I IVe	Cance	Treip		Disp	ersion		- Di	stribution	
18							Vs	td deviation 🔽	Minimum		Skewness	
19							V	ariance	Maximum		Kurtosis	
20								lange	S.E. mean			
21								-23° [.	modil			
22									Continue	Cancel H	lelp	
23												_
24												
25												
26												

5.1.1. ábra. Analyze/Descriptive Statistcs /Frequencies menü



5.1.2. ábra. Eredmények

Az 5.1.2. eredmények ábrán látható táblázat első oszlopában (*Valid*) a folytonos változó ténylegesen előforduló ismérvértékei találhatók. A második oszlop az ismérvértékhez előforduló abszolút gyakoriságokat mutatja (*Frequency*). A harmadik oszlop (*Percent*) az

ismérvváltozók százalékos megoszlását tartalmazza. A negyedik oszlop (*Valid Percent*) az érvényes, felhasználható értékek százalékban kifejezett megoszlását (relatív gyakoriságát) mutatja. Az ötödik (*Cumulative Percent*) oszlop a kumulált gyakoriságot mutatja.

A *Frequency* menüpont (*Charts*) modulja beépített grafikus ábrázolási lehetőséget kínál fel. Az egyszerű ábratípusok közül (oszlop, kördiagram, hisztogram) van lehetőségünk választani, legyen ez most a Hisztogramot, melyre egy Gauss-görbét is illesszünk (5.1.3. ábra). Az eredmény a 5.1.4. ábrán látható.

File	Edit	View	Data	Transform	Analyze	Direct Market	ing <u>G</u> raphs <u>L</u>	Itilities Add-o	ns <u>W</u> indow	Help		
2					×		P A					ARG
		Na	ime	Туре	Width	Decimals	Label	Values	Missing	Columns	Align	Mea
	1	fajta		Numeric	8	0	Napraforgó fajták	{1, POMAR}	None	8	Right	🗞 Nomir
	2	kezele	S	Numeric	2	0	Kezelések	{1, 1. kezel	None	8	疆 Right	🙈 Nomir
	3	tanyer	a	Numeric	8	0	tányérátmérő	None	None	8	疆 Right	Scale &
	4	magas	sag	Numeric	8	0	magasság	None	None	8	i Right	Scale 🔗
	5	💽 F	requenc	ies				×	Frequencies: Ch	arts	×	
	6						_				100012	
	7		Manraf	oraó faiták lfa	1	Ariable(s):	Statis	stics	Chart Type			
	8	1 🙎	Kezelé	sek (kezeles)	1	lanyeraune	<u>Cha</u>	arts	© N <u>o</u> ne			
	9		magas	ság (magas			For	mat	Bar charts			
	10						St	de la	Pie charts			
	11								Histograms:			
	12						Boots	strap	Show nor	mal curve on	n histogram	
	13								Chart Values			
	14								Chart values	8.0		
	15		Display	requency table	5				Frequencies	@ Percenta	iges	
	16				Dente D				Continue	Cancel	Help	
	17			UK [aste	Cancel	Help					
	18				1							
	19	ĺ										
					-						-	_

5.1.3. ábra. Hisztogram és Gauss görbe illesztése



5.1.4. ábra. Eredmények ábrázolása

5.2 Analyze/Descriptive Statistcs/ Descriptive

A bal oldali dobozból jelöljük ki a célváltozónkat, mely most a tányérátmérő. Lehetőségünk van egyszerre több változót is vizsgálni, de ha többet kívánunk egyszerre elemezni, akkor törekedjünk az azonos mérési skálán mért változók alkalmazására. (5.2.1. ábra)



5.2.1. ábra. Analyze/Descriptive Statistcs/ Descriptive

Az 5.2.2. ábra kezelésenként mutatja be a leíró statisztika adatait.

heiró statisztika. Napraforgó tányé	rátmérő és magassa	ig adatok.spv [Docur	nent2] - IBM	SPSS Statistic	s Viewer											-	٥	×
Eile Edit View Data Transform Insert Format Analyze Direct Marketing Graphs Utilities Add-ons Window Help																		
😑 🖩 🖨 👌 🥹) 🛄 🖛				0		PP		•	•	+ -	- 6		7 🗗				
Cutput Cog Cutput Cog Cutput	Descript	tives																*
Active Dataset			Danas	Minimum	Manianum	0	escriptive State	sucs	Deviation	Variance	01		Kee		1			
G Statistics		Statistic	Statistic	Statistic	Statistic	Statistic	Statistic Std	Error	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error				
tânyêr atmêrô	tányér átmé	rő 390	16	12	28	8431	21.62	.134	2.637	6.951	506	.124	.363	.247				
E Descriptives	Valid N (list	wise) 390				0000000	0.00040.000											
Construction of the second secon	SPLIT FII DESCRIPTI /STATIS	E LAYERED BY) VES VARIABLES TICS=MEAN STDI	kezeles. =tanyer_a DEV VARIA	NCE MIN >	AX KURTOS	SIS SKEW	IESS.											
			N	Minimum	Maximum	Mean	Std Deviation	Variance	Ske	whase	Ku	toeie	1					
	Kezelések		Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error	1					
	1. kezelés	tányér átmérő	70	16	28	22,17	2,519	6,347	-,222	,287	-,527	,566	1					
		Valid N (listwise)	70					-										
	2. kezeles	tanyer atmero	70	15	27	22,04	2,726	7,433	-,568	,287	,242	,566						
	3. kezelés	tányér átmérő	70	12	27	19.51	3.124	9,761	.045	.287	243	.566	1					
		Valid N (listwise)	70							1								
	4. kezelés	tányér átmérő	70	17	26	21,60	1,764	3,113	-,260	,287	,462	,566	1					
	6 hanalán	Valid N (listwise)	70	47		24.04	0.400	4.000	205	207		500	-					
	5. Rezeles	Valid N (listwise)	70	1/	21	21,91	2,198	4,833	,205	,287	-,034	996,						
	6. kezelés	tányér átmérő	40	20	26	23,10	1,549	2,400	-,393	,374	-,493	,733	1					
		Mallel M. (linksing)											1					
		valid iv (listwise)	40															
		valid N (listwise)	40			1							1					
		valid N (IStWISE)	40										J					-
		valid N (itstwise)	40		1	1	1			.1			IBM	SPSS Statist	tics Processor is ready	l	Unicode:C	T NC

5.2.2. ábra. Csoportosított (kezelések) adatok eredménye

5.3 Normális eloszlás, a normalitás tesztelése

- Az intervallum és arányskálán mért változók eloszlására általában jellemző, hogy az értékek az átlag körül csoportosulnak, és attól távolodva egyre kevesebb érték található ("haranggörbe").
- Nem minden ilyen eloszlás normál, de sok biológia adat nagy mintaszám esetén közelít a normál eloszláshoz.
- A centrális határeloszlás tétele szerint, ha egy nem normál eloszlású populációból véletlen mintákat veszünk, akkor ezen minták átlagainak eloszlása a normál eloszláshoz közelít. A biológiai változók értéke általában sok különböző ható tényező eredménye, így közelítenek a normál eloszláshoz.
- A parametrikus tesztek alkalmazásakor meg kell győződnünk arról, hogy a normalitás feltétele teljesül-e. A normalitás ellenőrzésére leggyakrabban használt teszt a Kolmogorov-Smirnov teszt, amelynek az ereje azonban nagyon alacsony, ezért használata nem ajánlott.
- A másik gyakran használt teszt a normalitás ellenőrzésére a **Shapiro-Wilk teszt**, amelynek az ereje jó, ha a minta nem tartalmaz azonos értékeket.

- Az azonos értékek (ties) előfordulása azonban a biológiai adatokban gyakori, így ennek a tesztnek az alkalmazhatósága is korlátozott.
- Ezért a normalitást gyakran grafikusan vizsgálják **kvantilis-kvantilis ábra (Q-Q plot)** segítségével. Ha a minta eloszlása nem tér el jelentősen a normál eloszlástól, akkor az elméleti és a megfigyelt kvantilisek közel egyenes vonalat adnak. (5.3.1 ábra)





Parametrikus és nem parametrikus statisztikai tesztek

A statisztikai tesztek két nagy csoportba oszthatók a vizsgálni kívánt változók eloszlása alapján.

- A parametrikus tesztek, mint nevük is mutatja a vizsgált populáció valamilyen paraméterét becsülik. Feltételezik, hogy a vizsgált változó (illetve a hiba) eloszlása normális. Ezeknek a teszteknek az ereje nagy (kis különbségek is kimutathatók velük), de sok feltételük van, és általában csak arány vagy intervallum skálán mért változókra alkalmazhatók.
- Ezzel szemben a **nem parametrikus tesztek** nem becsülnek paramétert. Nem követelik meg a normalitást, de egy részüknél feltétel, hogy az eloszlásnak egy bizonyos alakja

legyen, pl. szimmetrikus. Kevesebb feltételük van és nominális és ordinális változókon is használhatók. Erejük sokszor kisebb, mint a parametrikus megfelelőiknek, és sok, főleg a bonyolultabb parametrikus tesztnek nincs meg a nem parametrikus megfelelője.

5.4 Analyze/Descriptive Statistcs /Explore.

A hisztogramhoz hasonlóan két másik diagramtípus is alkalmas a normalitás vizsgálatára. Az egyik az úgynevezett stem-and-leaf (szár-levél) diagram, a másik pedig a normál eloszlás ábra (normal probablity plot). Mindkét az *Analyze/Descriptive Statistcs* /Explore menüponttal érhető el. A napraforgó szármagasság adatainak normalitás vizsgálat a Plots menüpontban a STEM-AND-LEAF és a NORMAL PROBABILITY PLOTS WITH TESTS parancsok kiválasztásával végezzük el. (5.4.1. ábra)



5.4.1. ábra. Normalitásvizsgálat

Stem-and-leaf grafikon

A változó értékeit "stem=tő" és "leaf=level" bontjuk számjegyeik alapján, általában az első vagy első két helyiértéket választva tőnek (figyelem: **első helyiértéket** és nem első számjegyet!).

Ezután növekvő sorrendbe rendezzük a töveket, majd az azonos tőhöz tartozó leveleket soronként ismét rendezzük. Az így kapott ábra kissé hasonlít egy elfordított hisztogramra, azzal a különbséggel, hogy attól eltérően a tényleges értékeket ábrázolja. (5.4.2. ábra)

Frequency Stem & Leaf

7,00 E	xtremes (=<152)
1,00	15.4
8,00	15.55566679
16,00	16.0000011111222334
18,00	16.555555555677778899
49,00	17.0000000000000000011111112222222233333333
40,00	17.555555555555556666777777788888888888999
22,00	18.0000000000001133344
9,00	18. 555566889
9,00	19.00000002
1,00	19.5
15,00 E	Extremes (>=197)
Stem wid	lth: 10

Each leaf: 1 case(s)

5.4.2. ábra Napraforgó szármagasság Stem-and-leaf grafikonja

Az ábráról leolvasható, hogy 155 cm-es szármagasság 3 db, 156 cm-s 3 db, 157 cm 1 db és 159 cm-s szintén 1 db fordul elő az adathalmazban.

A **normalitás vizsgálatára** alkalmas tesztek közül a két leggyakrabban alkalmazott próba a Kolmogorov-Smirnov- és a Shapiro-Wilk-próba, melyet az Explore menüpont kiszámol. A próbák szignifikanciaszintje kis (kevesebb mint 30 megfigyelésből álló) mintáknál nem használható, míg nagy minták (1000 fő felett) esetében pedig igen érzékeny.

A próba nullhipotézise az, hogy a változó nem normál eloszlású és a szignifikanciaszint alapján az adatok eloszlása szignifikánsan különbözik a normáleloszlástól, tehát ez azt jelenti, hogy a változó nem normális eloszlású.

A normalitás tesztelésekor mind a grafikai, mind pedig a statisztikai vizsgálatok elengedhetetlenek. A normalitás feltételének megsértése esetén adattranszformáció végrehajtása szükséges, amely segít normális eloszlásúvá alakítani az adatokat.

Tests of Normality

	Kolmogor	ov-Smirno	v ^a	Shapiro-Wilk				
	Statistic	df	Sig.	Statistic	df	Sig.		
magasság	,137	195	,000	,926	195	,000		

a. Lilliefors Significance Correction

A normál eloszlás 5.4.3 ábrája adataink tényleges eloszlásának és a normál eloszlás kumulatív görbéjét veti össze egymással. A normál eloszlás ábra megbízhatóbb képet ad a normalitásról, mint a hisztogram: itt a normál eloszlást képviselő átlós vonalat hasonlítjuk az adatainkat jelképező kis körökhöz. Minél nagyobb a kettő közötti eltérés, annál kevésbé felel meg változónk a normalitás feltevésének. Ez esetben egy bal oldali aszimmetrikus, csúcsos eloszlást találtunk, néhány kiugró értékkel.



5.4.3. ábra. Normál eloszlás grafikus vizsgálata

Varianciaanalízis

A varianciaelemzés két vagy több sokaság várható értékének összehasonlítására szolgál, és azt vizsgálja, hogy van-e különbség két vagy több csoport átlaga között. A kérdés tehát, hogy egy független változó hogyan befolyásolja egy függő változó alakulását. A példánkban majd arra a kérdésre szeretnénk választ kapni, hogy az egyes növényvédőszerek hogyan befolyásolják a tányérátmérő alakulását, amely szorosan összefügg a termésmennyiséggel.

A varianciaanalízis alkalmazásakor a független változó egyes kimenetei (különböző növényvédő szerek) milyen hatással vannak a függő változó értékeire (tányérátmérő), átlagaira. Ha ezek az átlagok szignifikánsan különböznek, akkor az azt jelenti, hogy a független változó érdemi hatással van a függő változóra. Az analízis fontos feltétele, hogy míg a független változók nominális, kategorizált változók (tehát a kimeneteik nem számok), a függő változók metrikusak legyenek.

A varianciaanalízis-modell alkalmazásának feltétele:

- a minták függetlenek legyenek,
- normális eloszlású sokaságokból származzanak, valamint
- azonos legyen a sokaságok varianciája.

A minták függetlenségét leginkább a megfelelő kísérleti elrendezéssel biztosíthatjuk. A normális eloszlással kapcsolatban vagy szakmai ismereteink lehetnek, vagy a minta alapján ellenőrizhetjük a normalitás teljesülését. A szóráshomogenitás vizuális megjelenítésére

- a pontfelhődiagramon az átlag és a szórás együttes ábrázolásával,
- statisztikai vizsgálata a *Levene-teszttel* történik.

Ha ez a feltétel nem teljesül, akkor ezt heteroszkedaszticitásnak hívjuk

 A harmadik feltétellel kapcsolatban a következő megkötést szokták tenni: ha közel azonosak a minták elemszámai, akkor a legnagyobb varianciának kisebbnek kell lennie, mint a legkisebb variancia kétszerese.

Egytényezős kísérletek (Más néven: egyutas osztályozás, egyszempontos varianciaanalízis ANOVA)

Egytényezős a kísérlet, ha

- k számú független mintánk van (ezek a kezelések),
- és minden mintában r számú mérés vagy megfigyelés található. Szokás az r-et ismétlésnek is nevezni.

Hipotézis

- A nullhipotézis az, hogy az összes kezelés átlaga egyenlő.
- Az alternatív hipotézis pedig azt jelenti, hogy legalább egy olyan középérték pár van, ahol nem tekinthetők a középértékek azonosnak.

$$H_0: \mu_1 = \mu_2 = \ldots = \mu_k$$

 $H_1: \text{legalabb egyszer } \mu_i \neq \mu_j, \, i=1,\,2,\,\ldots\,,\,k,\,j=1,\,2,\,\ldots,\,k.$

A varianciaanalízis-modell felállítása

- Az adatokra úgynevezett modellegyenletek állíthatók fel.
- Ez egy lineáris egyenletrendszer lesz, ahol a kísérletben megfigyelt értékeket egy olyan összegre bontjuk fel, melynek egyik tagja a mesterséges hatást (kezelést), a másik tagja pedig a véletlen hatást (hibát) tartalmazza.
- Kezelés hatása alatt értjük azt a szisztematikus hatást, ami a valószínűségi változónkat befolyásolhatja (Baráthné, 1996).

A modellegyenlet más formája:

$$X_{ij} = \mu + \alpha_i + e_{ij}$$

ahol

μ a sokaságok közös várható értéke
α_i az i-edik sokaság várható értékének és a μ-nek a különbsége. *e_{ij}* a véletlen (hiba) hatása.

Mivel $\mu_i = \mu + \alpha_i$,

H₀ hipotézis megfogalmazható a következő alakban is:

 $H_{o}: \alpha_{i} = 0;$ i = 1, 2,...,k.

A modellegyenletet átrendezve: $X_{ij} - \mu = \alpha_i + e_{ij}$.
A bal oldalon látható X_{ij} - μ az egyes mintaelemek eltérése a közös várható értéktől. Ez két részre bontható:

- α_i a szisztematikus hatás okozta eltérésre,
- e_{ij} a véletlen okozta eltérésre.
- A modellben a kétféle eltérés várható értéke E(∑ a_i) = 0 és E(∑ e_{ij}) = 0 (Baráthné, 1996).

Varianciaanalízis táblázat

- Azt vizsgáljuk, hogy a minták középértékei közötti kezelés hatása okozta variancia nagyobb-e a mintavételezésből származó véletlen hatás okozta hibavariancia értékénél.
- A döntést F-próba segítségével hozzuk meg.

Tényező (Source)	Eltérés négyzetösszeg (Sum of squares)	Szabadságfok (Degrees of freedom)	Szórásnégyzet (Mean square)	F	Sig.
Csoportok között (Betwen groups) kezelés hatása	$SS_B = \sum_{i=1}^k n_i (\overline{x}_i - \overline{X})^2$	k-1	$MS_B = \frac{\sum_{i=1}^k n_i (\bar{x}_i - \bar{X})^2}{k - 1}$	$F = \frac{MS_B}{MS_W}$	
Csoportokon belül (Within groups) véletlen hiba	$SS_W = \sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \overline{x}_i)^2$	n-k	$MS_{W} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \overline{x}_{i})^{2}}{k(n-1)}$		
Teljes (Total)	$SS_T = \sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \overline{X})^2$	n-1	$MS_T = \frac{\sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \overline{X})^2}{k(n-1)}$		

ANOVA táblázat

F-próba a varianciák összehasonlítására

- Ha a sokaság normális eloszlású, akkor a szórásnégyzet fenti két becslése független egymástól,
- és ha a nullhipotézis igaz, akkor a kezelés okozta variancia (*MS_B*) nem lehet nagyobb a véletlen hatása okozta varianciánál (*MS_W*).
- Ha mégis nagyobb, vagyis a nullhipotézist elvetjük, akkor az azt jelenti, hogy a vizsgált ismérv szempontjából a minta nem homogén.

$$F = \frac{MS_B}{MS_W} \qquad DF = (k-1), (n-k)$$

- Az F próbában mindig a kezelés okozta variancia áll a számlálóban, mert azt vizsgáljuk, hogy ez a variancia nagyobb-e a véletlennek tulajdonítható hibavarianciánál.
- Egyoldali próbát kell végezni. A H₀-t elfogadjuk, ha $F_{számított érték} \leq F_{táblázatbeli érték}$.
- Ellenkező esetben elvetjük a nullhipotézist.

Középértékek többszörös összehasonlítása

- Amennyiben szignifikáns különbségek mutathatók ki a kezelés hatására, vagyis az alternatív hipotézis bizonyul igaznak, tovább kereshetjük, hogy melyik sokaság átlaga tér el jelentősen melyiktől.
- Azt a legnagyobb különbséget, amely még véletlenszerűen jelentkezik szignifikáns differenciának nevezünk és SzD_α vagy SzD_{P%}-kal jelöljük.
 - Ha két kezelés átlagagának különbsége $(\overline{x_i} \overline{x_j})$ kisebb a szignifikáns differenciánál, akkor a különbség még a véletlennek,
 - ha nagyobb, akkor pedig a szisztematikus hatásnak (kezeléseknek) tulajdonítható (Baráthné 1996).
- A szignifikáns differencia tetszőleges megbízhatósági szintre megadható, de leginkább a 95%-os megbízhatósági szint használatos.

$$SzD_{\alpha} = t_{1-\alpha/2} \cdot \sqrt{MS_w \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

Különböző sokaságok/ csoportok átlagainak összehasonlítására az ANOVA-n kívül más eljárás is alkalmas, mint például a *t-próba*.

A t-próba azonban *kis minták* (n<30) esetén használható, és típusait tekintve megkülönböztetünk egy-, két- és többmintás t-próbát.

- Az egymintás t-próbákkal (Analyze/Compare Means/ One Sample T Test) egy adott minta valamely jellemzőjére vonatkozó feltevések helyességének ellenőrzésére történik.
- A *két- és többmintás próbák* két vagy több sokaság egymással való összehasonlítására alkalmas.

5.5 Egytényezős varianciaanalízis (/One Way ANOVA)

A következő feladat megoldása során arra a kérdésre szeretnénk választ kapni, hogy az egyes kezelések hogyan befolyásolják a tányérátmérő alakulását, amely szorosan összefügg a termésmennyiséggel.

Az SPSS programcsomag Analyse/Compera Means/One Way ANOVA menüjét alkalmazzuk, amely a 6.1.1.ábrán látható. Függő változónak (Dependent List) válasszuk a tányérátmérőt, míg független változónak (Factor) az egyes növényvédőszeres kezeléseket. Options menüpontot választva pipáljuk be a leíró statisztikát (Descriptive) és a homogenitás vizsgálatot (Homogenety of variance test).

- Eo	t <u>V</u> iew <u>D</u> ata	Transform	Analyze	Direct Market	ing <u>G</u> raphs <u>U</u>	tilities Add- <u>o</u> r	s <u>W</u> indow	Help			
a k			7 🖁		P K			3		 ▲ ▲	
	Name	Туре	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Rol
1	fajta	Numeric	8	0	Napraforgó fajták	{1, POMAR}	None	8	I Right	\delta Nominal	🔪 Input
2	kezeles	Numeric	2	0	Kezelések	{1, 1. kezel	None	8	疆 Right	\delta Nominal	🔪 Input
3	tanyer_a	Numeric	8	0	tányér átmérő	None	None	8	疆 Right	Scale 8	🔪 Input
4	magassag	Numeric	8	0	magasság	None	None	8	遭 Right	🔗 Scale	🔪 Input
5											
6		ta One-	Way ANOVA				X				
7								Cone-W	lay ANOVA: Opti	ons X	
8					Dependent List	Cor	trasts	Statistic	۹		
9		💦 Naj	praforgó fajtá)	< [fa	🔗 tányér átmér	5 [tanye	tHoc	Des	minting		
10		🛷 ma	gassag (mag	as				Eiva	and random of	Tacte	
11				\$		Ob	uons	Hom	oneneity of varia	ance test	
12				_		Boo	tstrap	Brow	m-Foreuthe	100 1001	
13								E Welc	h		
14					Factor			E How	**		
15				\$	Kezelések [k	ezeles]		🛅 Mear	ns plot		
16			Car					Missing	Values		
			UK	Paste	Reset Cancel	нер		@ Eucle	de conce anolu	ais by applying	
17								O Evel	iue cases analy:	sis by analysis	
17 18										no.	
17 18 19								O Excit	ide cases listwis	se	
17 18 19 20								Conti	nue Cases listivis	Help	
17 18 19 20 21								Conti	nue Cases IIstwis	Help	

6.1.1. ábra. One Way ANOVA Options

Az eredményeket a 6.1.2. ábrán látjuk.

variancia	analízis Na	praforgó	tányérá	tmérő és mag	asság adatok	cspv [Docu	ment1] - IBM SP	SS Statistics Vie	wer				
Edit	View	Data	Transt	form Insert	Format	Analyze	Direct Marke	ting <u>G</u> raphs	Utilities	Add-ons	Window	Help	
		a	Ð	1	n 1		E		•	F 17	B	b	🔶 י
Cut	Dut Log Cneway	6	•	Oneway				Descripti	Wes				
	- Dese	criptives		tányér átmé	rő								
	Test	of Homo	9						95% Co	nfidence Inte Mean	erval for		
6	Post	Hoc Tes	ts		N	Mean	Std. Deviatio	n Std. Error	Lower Bo	und Upp	er Bound	Minimum	Maximum
	-0	Title		1. kezelés	70	22,17	2,51	9 .301	2	1,57	22,77	16	28
		Multiple C		2. kezelés	70	22,04	2,72	.326	2	1,39	22,69	15	27
		Title		3. kezelés	70	19,51	3,12	.373	1	8,77	20,26	12	27
		E tány	é i	4. kezelés	70	21,60	1,76	.211	2	1,18	22,02	17	26
		一節	т	5. kezelés	70	21,91	2,19	.263	2	1,39	22,44	17	27
			1	6. kezelés	40	23,10	1,54	9 ,245	2	2,60	23,60	20	26
			<u>.</u>	Total	390	21,62	2,63	17 ,134	2	1,36	21,88	12	28
	Log Oneway			Tes tányér átmé Levene Statistic	t of Homoger rő df1	neity of Var	riances ! Sig.						
1	• Title			7,1	77	5	384 ,001	0					
	Desc Test	of Homo VA	g	tányér átmé	rõ		ANOVA						
	- Post	Hoc les Title			S	Sum of quares	df	Mean Square	F	Sig.			
		and the second states of the s				107.000	6	97 679	14 840	000			
	-6	muluple (1	Between Gr	oups	437,889	0	07,070	14,040	,000			
	-6	muluple (Between Gr Within Grou	ps	437,889	384	5,902	14,040	,000			

6.1.2. ábra. Varianciaanalízis eredménye

Test of Homogeneity of Variances

tányér átmérő

Levene			
Statistic	df1	df2	Sig.
7,177	5	384	,000

A fenti táblázat alapján megállapítható, hogy a szóráshomogenitás feltétele nem teljesül, amely abból látható, hogy az első táblázat (Test of Homogenity of Variances) szignifikáns (0,000). A Levene-teszt nullhipotézise ugyanis azt mondja ki, hogy a szórások nem egyenlők, amelynek elvetése azt jelenti, hogy a szóráshomogenitás teljesül.

Az ANOVA táblázatot nézve megállapítható, hogy a kezelések szignifikánsan különböznek, mivel az F próbához tartozó valószínűség szignifikanciaszintje 0,000<0,05 ezért a nullhipotézist elutasítjuk, azaz létezik legalább két olyan kezelés, amely szignifikánsan különbözik.

*Analizishe	ez Napraforgó tá	nyérátmérő és r	nagasság ada	itok.sav [Data	Set1] - IBM SPSS Sta	tistics Data Edito	r					
ile <u>E</u> dit	View Data	Transform	Analyze	Direct Market	ting <u>G</u> raphs <u>U</u>	tilities Add-or	ns <u>W</u> indow	Help				
26			N					2	A 🕢	All 1		
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role	
1	fajta	Numeric	8	0	Napraforgó fajták	{1, POMAR}	None	8	Right	🗞 Nominal	🔪 Input	
2	kezeles	Numeric	2	0	Kezelések	{1, 1. kezel	None	8	Right	🙈 Nominal	S Input	
3	tanyer_a	Numeric	8	0	tányér átmérő	None	None	8	酒 Right	Scale Scale	S Input	
4	magassag	Numeric	8	0	magasság	None	None	8	Right	Scale 🔗	🔪 Input	
5								_				
6		Cone-V	Nay ANOVA				×	👍 One-Way	ANOVA: Post H	oc Multiple Comparis	sons	×
7								-				
8					Dependent List:	Co	ntrasts	-Equal Van	ances Assume	d		
9		Map mar	ratorgo fajtal	(Ita	tanyer atmer	o (tanye Po:	st Hoc		<u>S</u>	N-K 🛄 <u>M</u>	Valler-Duncan	
10		•	,			Or	tions	Bonter	roni 🛄 <u>I</u> u	JKey Ty	pe vitype il Error Ra	
11						Rev	vietron	Sidak	E TI	ikey's-b 🔲 D	lunnett	
12							nau ap	Scheffe		uncan Co	ontrol Category :	ast
13	1						-	<u>R</u> -E-G	WF 🖪 H	ochberg's GT2	Fest	
14	1				Factor:		-	R-E-G-	w <u>o</u> 🖾 G	abriel	<u>2</u> -sided	ontrol 🔘 > Control
10					Kezelések [k	ezeles]	-	Familiar				
17			OK	Paste	Reset Cancel	Help		Equal van	ances Not Asst			0
18	1							Tamha	iners 12	unnetts 13 📗 G	ames-Howell	Dunnett's C
19								Significanc	e level: 0,05			
20												
20							-			Continue	Help	
22												
23	1											
24												

6.1.3. ábra. Post Hoc tesztek

Ha a varianciaanalízis eredménye nem szignifikáns, akkor várható értékek között nincs érdemi különbség. Ha viszont a kapott eredmény szignifikáns, akkor ez még csak azt jelzi, hogy a kategóriák átlagai között általában van eltérés azonban nem tudjuk, hogy pontosan mely átlagok között van eltérés. Ennek vizsgálatára szolgálnak az előzetes (a priori) és az utólagos (a posteriori), másnéven post-hoc tesztek (kontrasztok).

Így arra a kérdésre a választ, hogy mely kezelések között van szignifikáns különbség, a Post Hoc menüpont kiválasztásával ad hatjuk meg.

A 6.1.3. ábrán láthatóan csoportosítva találjuk az alkalmazható módszereket, amelyek egyik fele a vizsgált változók varianciájának homogenitása (Equal Variances Assumed) esetén, a másik pedig a változók varianciájának különbözősége (Equal Variances Not Assumed) esetén alkalmazható. Az első csoportban a próbák az erejük növekvő sorrendjében következnek egymás után.

Post-hoc tesztek

A változók varianciájának homogenitása (Equal Variances Assumed) esetén:

LSD (Least Significant Difference = legkisebb szignifikáns különbség): t-próbával ellenőrzi a kezelésátlagok (between groups) közötti különbségeket. Legengedékenyebb post-hoc teszt a feltételeket illetően, vagyis ennek az alkalmazásával lehet a kezelésátlagok közötti különbségeket a leghamarabb szignifikánsan kimutatni. Leginkább akkor célszerű az alkalmazása, ha a kezeléstípusok száma három.

Bonferroni-módszer: t-próbával vizsgálja a kezelésátlagok' különbségeket, azonban a próba p-értékét nem az általános alfa-szinttel (α = 0,05) veti össze, hanem egy módosított alfával számol, melyet a következőképpen határoz meg:

Módosított α = eredeti α /az összes lehetséges összehasonlítás száma

- Scheffé-próba: Egyidejű páros összehasonlításokat végez az átlagok minden lehetséges páros kombinációjára. A Scheffé-próba az F mintaelosztást használja. Az egyik legkonzervatívabb, és ezáltal a legbiztosabb próbának számít.
- Tukey-próba: Legszélesebb körben használt és legkevésbé ellentmondásos próba, amelyet akkor érdemes használni, ha a kategóriák száma nagyobb háromnál. (Sajos, Mityev)

A változók varianciájának inhomogenitása (Equal Variances Not Assumed) esetén:

- Tamhane's
- Dunnett's T3
- Games Howell
- Dunnett's

Válasszuk ki a Tamhane's-t, melynek eredménye a 6.1.4. ábrán látható. A táblázatban csillaggal megjelölésre került, amely kezelések között szignifikáns különbség adódott. A kezelések tulajdonságait fegyelembe véve vonhatunk következtetéseket a tányérátmérő alakulásáról, azaz melykezelés okoz szignifikáns különbséget.

€ *v	ariancia	analízis N	apraforgó	tányérá	átmérő és r	nagass	ág adato	k.spv [Do	ocument1] - IBM :	SPSS Sta	tistics	Viewer					
File	Edit	View	Data	Transf	orm <u>I</u> ns	ert	Format	Analyz	e Dire	ct <u>M</u> arke	eting	Graph	ns <u>U</u> tiliti	es Ad	d- <u>o</u> ns	Window	v <u>H</u> elp	
8			à	Ð		5	- 71					Q		F	12			
	Cutp	out	*	-	Total			2704.07	4	389								
	Log Log Doeway Coneway Coneway Coneway Coneway Conescriptives Conescrite Conescriptive Conescrite Conescriptive				Post H	oc T	ests	nyér átm	érő	Multiple	e Compa	arison:	s					
		E Post	Hoc Te		Tamhane	Э												
	-	-	Title						Mea	n		T		959	6 Confid	ence Inte	erval	
			Multiple		/I) Kozolá	nak	(1) 1/070	lának	Differen	ice (l-	Std E	rror	Sig	Lower	Bound	Upper	Bound	
			Homog () Titl		(I) Kezelé	sek	2 kezelé	iesek	0)	179	old. El	144	1 000		-1 19	abba.	1.45	
			E tár		1.11020101		3. kezelé	s	3	657		180	000		1.23		4 09	
			-6				4. kezelé	İs		.571		368	.860		- 53		1.67	
							5. kezelé	s		257		100	1,000		- 93		1.45	
			-4				6. kezelé	İs		- 929		388	.244		-2.09		.23	
					2. kezelé:	s	1. kezelé	śs		-,129	.4	144	1,000		-1,45	1	1,19	
							3. kezelé	śs	1	2,529		196	,000		1,05		4,01	
			-4	1			4. kezelé	is		.443		388	,988		-,72		1,60	
	00	- 10.0	G				5. kezelé	és		,129		119	1,000		-1,12		1,38	
		Log					6. kezelé	és	4	1,057		108	,151		-2,28		,16	
		- Title			3. kezelé:	s	1. kezelé	śs	-2	2,657	.4	180	,000		-4,09		-1,23	
	_	Note	s				2. kezelé	és	-2	2,529	.4	196	,000		-4,01		-1,05	
	-	- 🗿 Desc	criptives				4. kezelé	és	-2	2,086		129	,000		-3,37		-,80	
		Test	of Horr				5. kezelé	és	-2	2,400	,4	157	,000		-3,76		-1,04	
		ANO	VA Haa Ta				6. kezelé	İs	-3	8,586	.4	447	,000		-4,92		-2,25	
		Post	HOC TE		4. kezelé:	s	1. kezelé	śs		-,571		368	,860		-1,67		,53	
			Multiple				2. kezelé	és		-,443	3	388	,988		-1,60		,72	
		Log					3. kezelé	és	1	2,086	.4	129	,000		,80		3,37	
1	-	Oneway					5. kezelé	śs		-,314	1	337	,999		-1,32		,69	
	-	Title					6. kezelé	és	-1	,500	13	323	,000		-2,47		-,53	
		- Ca Desc	s		5. kezelé:	s	1. kezelé	és		-,257	,4	100	1,000		-1,45		,93	
		Test	of Hom				2. kezelé	śs		-,129	.4	19	1,000		-1,38		1,12	
		ANO	VA				3. kezelé	és	2	2,400	,4	\$57	,000		1,04		3,76	
	Ē.	E Post	Hoc Te				4. kezelé	śs		,314		337	,999		-,69		1,32	
			P.				a transit			100		100		1		1		

6.1.4. ábra. Post Hoc teszt eredményei

6 Korreláció- és regressziószámítás

A változók közötti kapcsolatok elemzésének egyik leggyakrabban alkalmazott módszere a *korreláció- és regressziószámítás*. A korrelációszámítás a metrikus változók kapcsolatainak vizsgálatával foglalkozik, a változók közötti kapcsolat meglétét, szorosságát és annak irányát értkeli. Egy paraméter, jelenség, esemény alakulását általában több tényező befolyásolja, ezért legtöbbször nem elegendő a két-változós modell elemzése. Amikor egyik tényezőre több másik tényező is hatással van, a kapcsolatok mennyiségi jellemzőinek, illetve szorosságának vizsgálatát *többszörös korreláció- és regressziószámítás*sal vizsgáljuk.

		Független változó								
		Nem metrikus	Metrikus							
	Nem	Kereszttábla	Diszkriminancia							
Fürer"	metrikus	elemzés	elemzés							
ruggo		Variancia	Korreláció,							
valtozo	Metrikus	variancia	regresszió							
		elemzes	elemzés							

Forrás: Sajtos, Mitev: SPSS kutatási és adatelemzési kézikönyv 7.1. ábra. A struktúravizsgáló módszerek összefoglalása

Jelen példáinkban azonban kizárólag a két-változós lineáris korreláció és regresszió számítás menetét mutatjuk be.

6.1 Lineáris korreláció

A korrelációs kapcsolat mérésének legelterjedtebb mutatószáma a *lineáris korrelációs (vagy Pearson-féle) együttható* (jele: r), amelynek alkalmazása során feltételezzük a változók közötti lineáris kapcsolatot, illetve ha a linearitás feltevése nem áll távol a vizsgált problémától. A korrelációs együttható kiszámítása a változók együttmozgását jellemző kovariancia mérőszáma és a változók szórása segítségével történik, az alábbi algoritmus segítségével:

$$r_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y}, ahol \ a \ ko \text{ var iancia} : C_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = \frac{\sum d_x d_y}{n} = \frac{\sum xy}{n} - \overline{xy}$$

ahol: σ_X és σ_V a változók szórásai.

Forrás: Ács Pongrácz: Gyakorlati adatelemzés

A korrelációs együttható abszolút értéke a a kapcsolat szorosságát, az előjele pedig a kapcsolat irányát fejezi ki. A korrelációs együttható (r) értéke -1 és +1 között lehet. Minél erősebb a kétváltozó között a kapcsolat, a korrelációs együttható értéke annál jobban megközelíti abszolút értékben az 1-et. Ha az r értéke nullával egyenlő, akkor a két változó között nincs lineáris kapcsolat (korrelálatlanok, de nem függetlenek).

r értéke	Kapcsolat erőssége
r =1	tökéletes kapcsolat (függvényszerű lineáris kapcsolat
0,81≤ r <0,99	a kapcsolat erős
$0,51 \le r < 0,8$	a kapcsolat közepesnél erősebb
r ≈0,5	a kapcsolat közepes
0,31≤ r <0,49	a kapcsolat közepesnél gyengébb
0,01≤ r <0,3	a kapcsolat gyenge
r =0	nincs kapcsolat (korrelálatlan)

Forrás: Jánosa: Adatelemzés SPSS használatával

7.1.1. ábra A kapcsolat erőssége

A korrelációs együttes szignifikancia vizsgálata megmutatja, hogy egy adott, többdimenziós minta esetén a változók között talált összefüggés mekkora valószínűséggel valódi és nem a véletlen műve. A mintához tartozó elemek szabadságfoka: szf=n-2. A szignifikanciavizsgálatot t-próbával végezzük el.

6.2 Kétváltozós lineáris regresszióanalízis

A mindennapi életben gyakran találkozunk olyan kapcsolatokkal, ahol az egyik jelenség nem határozza meg egyértelműen a másik jelenség lefolyását. A két véletlen jelenség közötti kapcsolatot sztochasztikus kapcsolatnak nevezzük akkor, ha az egyik jelenség lefolyását nem határozza meg egyértelműen a másik jelenség, csak befolyásolja.

A regressziós kapcsolat a kvantitatív változók sztochasztikus kapcsolatának matematikai leírását jelenti.

Így a regresszóanalízis két vagy több egymással sztochasztikus kapcsolatban lévő változók kapcsolatát legjobban leíró függvényt keres.

A változók közötti kapcsolat jellegét, minőségi jellemzőit vizsgálja a regressziószámítás (Huzsvai, 2012). A regressziószámítás a jelenségek tendenciáit, a kapcsolat természetét valamilyen jól megfogható és értelmezhető függvény formájában adja vissza. Ezeket a függvényeket regresszió-függvényeknek nevezzük.

Kétváltozós lineáris regresszió-analízis akkor használható, ha a függő változó skálatípusú, és lineáris kapcsolatban van egy magyarázó változóval.

A regressziós egyenes általános képlete a kétváltozós modellben:

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

ahol:

Y a függő vagy kritériumváltozó

X a független vagy magyarázó változó

 β_0 az az érték, ahol az egyenes az y tengelyt metszi

 β_1 a regressziós együttható, vagy az egyenes meredeksége, amely arra ad választ, hogy hogy az X változó egységnyi változás átlagosan mekkora változást okoz az Y eredményváltozóban.

ε hibatényező (reziduum).

A paraméterek becslése a legkisebb négyzetek módszerével történik úgy, hogy a tényleges és a becsült értékek eltérésösszege minimális legyen. Ez alapján a két paraméter az alábbi képletekkel határozható meg:

$$b_{1} = \frac{\sum (X_{i} - \overline{X})(Y_{i} - \overline{Y})}{\sum (X_{i} - \overline{X})^{2}} = \frac{\sum XY - n\overline{X}\overline{Y}}{\sum X^{2} - n\overline{X}^{2}}$$
$$b_{0} = \overline{Y} - b_{1}\overline{X} = \frac{\sum Y}{n} - \frac{b_{1}\sum X}{n}$$

A lineáris kapcsolat létének vizsgálatakor a nullhipotézis az, hogy a két változó között nem áll fenn lineáris kapcsolat, melyet kétoldali t-próbával ellenőrízünk.

A kapcsolat erősségének mérése a determinációs együttható (r^2) történik, amely a az Y teljes eltérés-négyzetösszegének X által magyarázott arányát mutatja.

A regresszióanalízis alkalmazásánál számos feltétel teljesülését kell ellenőrizni:

- A függő és független változó között lineáris kapcsolat van-e?
- A magyarázó változók egymástól függetlenek-e?
- A hibatényező normális eloszlást követ-e, melynek várható értéke 0, varianciája konstans?
- A hibatényezők korrelálatlanok-e, azaz egymástól nem függnek (nincs autokorreláció)?

A következő példában arra a kérdésre keressük a választ, hogy első laktációs tehenek napi tejtermelése és a megtermelt tej zsírtartalmának alakulása között milyen kapcsolat figyelhető meg. Először célszerű a mérési eredményeket grafikonon ábrázolni (6.2.1. ábra).



6.2.1. ábra. Napi tejtermelés és a zsírtartalom kapcsolata

Az SPSS programcsomag Analyse/Correlate/Bivariate menüjét alkalmazzuk (7.2.1. ábra), majd a Pearson-féle korrelációs együtthatót számítjuk ki.



6.2.2. ábra. Analyse/Correlate/Bivariate menü alkalmazása

A Variables ablakba áthelyezzük a vizsgálni kívánt változókat (Napi tejtermelés (kg) és Zsír tartalom (%)) és kiválasztjuk a Pearson féle korrelációs együtthatót. Az Options menüpontban az átlag és a szórás értékeit is kiírathatjuk (Means and standard deviations).

A kapott eredménytáblázat alapján megállapítható, hogy a tej mennyisége és zsírtartalma között laza, negatív kapcsolat áll fent, a korrelációs együttható értéke -0,346. A korrelációs együttható szignifikancia-vizsgálata (0,000<0,05) alapján megállapíthatjuk, hogy a kapott *r* érték valódi,

szignifikáns kapcsolatot jelent a két változó között, nem csupán a véletlen hatások eredőjeként keletkezett. Ezért a statisztikai próba nullhipotézisét, miszerint a két változó között nincs kapcsolat elutasítjuk

Correlations

Descriptive Statistics							
	Mean	Std. Deviation	Ν				
Napi tejtermelés (kg)	27,260	5,8327	95				
Zsír tartalom (%)	4,0476	,82288	95				

	Correlations										
		Napi tejtermelés (kg)	Zsír tartalom (%)								
Napi tejtermelés (kg)	Pearson Correlation	1	-,346**								
	Sig. (2-tailed)		,001								
	N	95	95								
Zsír tartalom (%)	Pearson Correlation	-,346**	1								
	Sig. (2-tailed)	,001									
	Ν	95	95								

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

6.2.3. ábra Korreláció számításakor kapott eredmények

Miután a korreláció számítás laza, szignifikáns kapcsolatot bizonyított a két változó között, ezért elvégezzük a regresszió analízist, melyben arra a kérdére keressük a választ, hogy az egyik paraméter ismeretében miként tudjuk a másik értékeit meghatározni.

Mivel szakmailag azt feltételezzük, hogy a tej mennyisége meghatározza annak zsírtartalmát, ezért arra a kérdésre keressük a választ, hogy a tejmennyiség ismeretében megtudjuk-e becsülni a zsír %-ot.

A regressziószámítás során elvégzéséhez az ANALYZE menüpont *REGRESSION* blokkjának *Linear* parancsát jelöljük ki. Mivel a tejmennyiség ismeretében szeretnénk az tejzsír tartalmat meghatározni, így a Napi tejmennyiség (kg) változót helyezzük a független (Independent(s)) ablakba, a Zsírtartalom (%) változót pedig függő (Dependent) változóként definiáljuk.

Courbent [Document] - IBM SPSS Statistics Viewer	-	- 0	×
Elle Edit View Data Transform Insert Format <mark>Analyze</mark> DirectMarkeling Graphs Utilities Add-ons <u>Window</u> Help			
Regols Dgscriptive Statistics Tagles Compare Mans Generalized Linear Model Miged Models Compare Mans Generalized Linear Model Compare Mans General			
Begression Impact And Advantate Linear Modeling. Lgglinear Impact And Advantate Linear Modeling. Neural Networks Impact Linear Modeling. Classify Impact Linear Modeling. Dimension Reduction Impact Linear Modeling. Scale Impact Linear Modeling. Nonparametic Tests Impact Linear Modeling. Scale Impact Linear Modeling. Multiple Response Impact Linear Modeling. Multiple Imputation. Impact Linear Scaling (CATREG). Optimal Scaling (CATREG). Impact Linear Scale	SZ]	Statistics Piots Save Options Style Bootstrap	
IBM SPSS Statistics Processor is ready Un	code:ON	H: 184, W: 1	095 pt.

6.2.4. ábra Lineáris regresszió alkalmazása

Az eredményül kapott **Model Summary** táblázat a már ismert korrelációs együttható (R=0,346), a determinációs együttható (R²=0,110) értékeit mutatja. A determinációs koefficiens értékéből arra következtethetünk, hogy a tejmennyiség alakulása 11%-osan befolyásolja a tejzsír tartalmat az általunk vizsgált adatok esetében.

Model Summary

			Adjusted R	Std. Error of
Model	R	R Square	Square	the Estimate
1	,346 ^a	,120	,110	,77610

a. Predictors: (Constant), Napi tejtermelés (kg)

Az ANOVA táblázatot nézve megállapítható, hogy a regresszió P<0,05 szinten szignifikáns. ANOVA^a

		Sum of				
Model		Squares	df	Mean Square	F	Sig.
1	Regression	7,633	1	7,633	12,673	,001 ^b
	Residual	56,017	93	,602		
	Total	63,650	94			

a. Dependent Variable: Zsír tartalom (%)

b. Predictors: (Constant), Napi tejtermelés (kg)

A regressziós egyenes becslő egyenletét ($y=b_0+b_1x$), **Coefficients**^a tálázat adatai alapján tudjuk felírni. A b_0 (Constant) értéke 5,379, a b_1 =-0,049, melynek előjele mutatja a korreláció negatív voltát.

Coefficients^a

		Unstandardized		Standardized		
		Coefficients		Coefficients		
Model		В	Std. Error	Beta	t	Sig.
1	(Constant)	5,379	,382		14,064	,000
	Napi tejtermelés (kg)	-,049	,014	-,346	-3,560	,001

a. Dependent Variable: Zsír tartalom (%)

Az egyenes egyenlete a következő:

Tejzsír%=5,379-0,049* Napi tejtermelés (kg)

Mivel a két változó között csupán laza kapcsolatot tapasztaltunk, így az egyenletbe a tejmennyiség értékeit behelyettesítve 12 %-os pontossággal tudjuk meghatározni a tej zsírtartalmát az adott mintában.

7 Gyakorló feladatok

1. Feladat

Az alábbi táblázatban egy ásványvizet forgalmazó cég két üzletében a havi forgalom (e Ft) látható.

A üzlet	B üzlet
(eFt)	(eFt)
790	640
600	710
560	830
710	820
810	760
720	690
840	720
720	790
690	710
610	740
570	610
550	570
	A üzlet (eFt) 790 600 560 710 810 720 840 720 840 720 690 610 570 550

Feladatok:

- Készítse el az adatbázist!
- Készítsen leíró statisztikát!
- Vizsgálja meg, hogy van-e szignifikáns különbség a két üzlet forgalmi adatai között!

2. Feladat:

Egy kísérletben különböző elemek élettartamát vizsgálták. Minden típusú elemből 5-5 darabot választottak ki. Az élettartam órában van megadva a következő táblázatban.

Elemek típusa	Élettartam (óra)
A típus	18,16,19,17,18
B típus	17,17,15,16,17
C típus	19,18,19,20,18
D típus	14,15,16,17,16

Feladatok:

- Készítse el az adatbázist!
- Készítsen leíró statisztikát!
- Vizsgálja meg, hogy van-e szignifikáns különbség a különböző típusú elemek élettartama között!

3. Feladat

Egy mezőgazdasági áruház három különböző városban lévő boltban a pénztárnál történő fizetéseket vizsgáltuk. Az adatok az alábbi táblázatban találthatók ezer Ft-ban megadva. Feltételezzük, hogy a fizetések normális eloszlásúak

Szentes	Csongrád	Hódmezővásárhelv
(eFt)	(eFt)	(eFt)
12,05	15,17	9,48
23,94	18,52	6,92
14,63	19,57	10,47
25,78	21,4	7,63
17,52	13,59	11,9
18,45	20,57	5,92
23,45	16,93	14,56
9,58	32,54	8,96
23,56	18,95	9,87
20,65	36,85	10,93
14,23	13,58	15,64

Feladatok:

- Készítse el az adatbázist!
- Készítsen leíró statisztikát!
- Vizsgálja meg, hogy van-e szignifikáns különbség a különböző telephelyek forgalmi adatai között.

4. Feladat

A megfigyelés során a Daphnia longispina (rasszokokat reprezentáló 7-7 klónnapban kifejezett átlagos élettartam adatokat vizsgáltuk.

Feltételeztük, hogy az átlagos élettartam normális eloszlású valószínűségi változó.

hím							
A rassz	7,2	7,1	9,1	7,2	7,3	7,2	7,5
B rassz	8,8	7,5	7,7	7,6	7,4	6,7	7,2
C rassz	7,3	7	8 <i>,</i> 9	7,4	7,3	7 <i>,</i> 5	7,1
nőstény							
A rassz	7,3	7	8,9	7,4	7,3	7 <i>,</i> 5	7,1
B rassz	7,2	7,1	9,1	7,2	7,3	7,2	7 <i>,</i> 5
C rassz	8,8	7 <i>,</i> 5	7,7	7,6	7,4	6,7	7,2

Feladatok:

- Készítse el az adatbázist!
- Készítsen leíró statisztikát összeségében és ivar szerinti kimutatásban!
- Első feltevésként vizsgálandó, hogy van-e szignifikáns különbség a rasszok várható élettartama között 5%-os szignifikancia szinten?
- Második felvetésként vizsgálandó, hogy a hímek átlagos élettartama rasszonként szignifikánsan eltér-e?
- Harmadik felvetésként vizsgálandó, hogy a nőstények átlagos élettartama rasszokként szignifikánsan eltér-e?
- Van-e eltérés bármely csoport élettartama között?

5. Feladat

Minősitett őszi búzafajták szemtermését, nyersfehérje tartalmát valamint nedves sikértartalmát vizsgálták. Az alábbi három táblázat a 8 különböző területen mért eredményeket tartalmazza. (Forrás: GOSZ-VSZT Őszi Búza Posztregisztrációs Fajtakísérlet 2014)

Feladatok:

- A három táblázat adatiból készítsen egy adatbázist!
- Készítsen fajtánként, termőhelyeként leíró statisztikát mindhárom paraméterre!
- Vizsgálja meg, hogy szemtermés, a nyersfehérje és a nedves sikértartalom paraméterek esetén van-e szignifikáns különbség az egyes fajták között!
- Vizsgálja meg, hogy a termőhelyek a három paramétert hogyan befolyásolják!

	Fajták	Szombat	Bábolna	Tordas	Iregszemc	Eszterág	Szarvas	Székkutas	Debrecen
1	Kalahari	9.80	7.68	7.96	10.11	10.00	8 37	8 53	9.40
2	Altigo	9.80	7.08	7,73	9.07	9.03	9.69	10.19	8 57
3	Genius	9.49	7,00	7.65	8 58	9.14	8.87	9.81	8 59
4	My Nádor	9.55	7 43	7 99	8 41	7 78	9.37	8 71	9.34
5	Babona	8.96	7,45	7.67	8.09	7,70	8.93	7 73	8 14
6	Amicus	8.80	7,30	7 33	8 30	8 28	8 56	6.98	8 76
7	Mulan	9.75	7.96	7 76	7.08	7 18	7 51	6.86	8.43
8	Ortegus	9.69	7 78	7 76	7 38	7,10	7 57	6,66	7 59
9	Balaton	8.80	7 34	8.02	7 36	6.52	8 54	6 79	8 60
10	Fidelius	8.62	7 48	7.61	7.06	6.07	7.66	6.73	8 48
11	My Pántlika	9.01	7,10	7.15	8 10	5 19	7,00	6.17	7 13
12	Hyland	9.11	7 45	8.03	5.03	6.81	6 99	7 11	7.17
13	My Karizma	9.48	7 11	7 27	7.89	673	6 64	5.05	6 44
14	My Kolompos	7.00	7.25	7.43	7.89	6 74	7.61	5 29	6.19
15	My Pengő	7.42	6.94	7,13	6.61	6.45	7.48	5 40	6 59
16	My Magdaléna	8.07	6.61	6.82	6.40	5 27	7.26	5,40	7.43
17	My Lucilla	8 71	6.80	7 20	7.09	4 74	6 77	5 41	6.82
18	Gallió	7.28	6.42	6.91	6.18	5 23	6.73	4.82	7.13
19	My Béres	7 94	6 34	6.67	6.43	4 43	7.05	3.97	6 33
20	My Suba	6.88	5 71	6.40	5 48	4 60	6.10	5 79	7 57
21	My Kikelet	6.88	5.82	6 49	5 39	5 94	6.04	5 77	6.08
22	My Ködmön	7 49	5.28	6 49	5 67	3 57	6.16	5 57	6.41
23	My Menijett	7.20	5.61	6.04	4.89	4.23	5.96	6.21	5.43
24	GK Körös	7.04	5.97	6.59	4.91	4.72	6.61	4.16	4.31
25	My Kolo	6.33	4.42	6.04	5.45	4.85	5.34	5.07	5.65
26	KG Kunhalom	7.89	5.48	5.83	5.25	2.95	5.97	4.13	5.20
27	My Toldi	6.09	3.89	6.36	5.22	5.20	5.47	5.58	4.77
28	Baletka	5.73	4.75	6.10	4.18	3.62	5.67	5.99	5.48
29	Antonius	6,63	6,25	5,53	4,70	4,07	4,70	4,31	5,25
30	KG Kunkapitány	7,48	4,17	5,50	3,76	3,53	6,05	4,57	5,51
31	Mv Karéj	8,12	4,39	6,72	4,34	3,98	5,44	3,18	4,36
32	Astrado	7,07	5,33	5,48	3,86	3,44	4,32	6,05	4,29
33	Lupus	6,81	4,89	6,28	5,18	3,45	5,18	3,48	4,38
34	Bitop	5,97	4,32	5,78	4,81	3,92	4,97	5,20	4,07
35	GK Csillag	6,79	5,24	5,79	3,88	3,01	4,86	3,95	4,59
36	Mv Lepény	6,18	5,13	5,73	3,17	3,68	4,30	4,38	4,14
37	Mv Marsall	4,94	3,41	4,92	3,81	4,69	4,38	3,92	3,66
38	GK Futár	6,06	3,40	5,35	2,23	1,48	3,48	3,29	4,90
39	Mv Kokárda	4,07	2,78	4,75	2,44	2,10	3,33	2,34	2,68
40	GK Berény	4,78	2,93	3,99	1,94	2,51	3,07	2,40	2,49
41	GK Kalász	4,08	2,39	4,05	1,82	0,93	2,76	2,45	2,79
42	GK Rozi	3,66	1,38	2,74	1,57	1,02	2,56	2,30	2,39
43	Mv Tallér	2,50	1,15	3,36	1,45	1,40	1,93	2,08	1,99
44	GK Békés	3,29	2,13	3,06	0,90	0.59	1,13	1,80	1,81

Minősitett őszi búzafajták szemtermese (t/ha) kisparcellás kiseretekben Fajtakísérleti Innovációs Tanács, 2014

Forrás: GOSZ-VSZT Őszi Búza Posztregisztrációs Fajtakísérlet 2014

	Faita	Szombat	Bábolna	Tordas	Iregszem	Eszterág	Szarvas	Székkutas	Debrecen
		hely			cse	puszta			
1	GK Békés	17,0	17,1	15,3	16,0	15,8	16,9	16,7	17,2
2	GK Rozi	16,4	18,1	15,1	16,1	16,7	15,7	15,9	17,3
3	GK Berény	16,6	18,1	14,7	16,5	15,7	15,7	15,7	17,2
4	Mv Tallér	16,3	17,3	13,9	15,5	15,0	14,2	15,2	16,3
5	GK Futár	14,2	16,5	13,1	15,8	16,2	14,0	16,2	16,2
6	Lupus	12,2	15,3	13,4	16,1	15,7	16,1	16,2	17,0
7	Antonius	14,4	16,0	13,4	15,8	14,5	15,0	15,9	16,4
8	Mv Béres	12,9	15,2	13,7	16,2	14,5	14,9	16,8	16,9
9	GK Kalász	15,5	16,7	12,7	15,5	16,0	14,0	14,3	15,8
10	Astardo	13,5	15,9	13,2	15,3	14,6	15,7	13,9	16,5
11	GK Csillag	13,5	15,8	12,6	15,9	15,5	14,2	14,7	16,0
12	Bitop	14,5	15,1	12,6	15,1	14,3	13,9	15,3	16,4
13	Mv Kolompos	12,5	15,9	12,3	14,9	13,6	15,0	16,2	16,7
14	Mv Toldi	14,2	15,9	12,3	15,6	13,5	14,4	14,9	16,0
15	Mv Karizma	13,0	15,8	12,5	15,5	13,4	14,6	15,5	16,1
16	Mv Suba	13,3	15,6	13,1	15,6	12,9	14,5	16,2	14,9
17	KG Kunkapitány	13,8	16,1	13,3	15,3	14,7	13,1	13,4	16,1
18	Mv Kolo	14,2	16,1	12,8	15,6	14,0	12,7	14,9	15,3
19	KG Kunhalom	13,3	16,0	13,2	15,0	13,9	13,9	14,3	16,1
20	Mv Magdaléna	12,9	14,8	13,4	15,3	13,8	13,8	15,6	15,4
21	Mv Marsall	13,9	15,8	11,9	15,1	13,9	14,2	13,9	14,8
22	Mv Karéj	12,7	15,2	12,7	15,2	13,7	13,8	14,6	15,6
23	Mv Pántlika	12,1	14,8	12,1	14,9	14,7	14,0	15,0	15,8
24	Gallió	12,0	15,2	12,7	15,7	13.3	13.6	14,3	16,2
25	Mv Menüett	13,0	15,1	12,7	14,6	13,1	14,2	14,4	15,5
26	Mv Ködmön	12.6	14.7	12.1	15.3	13.3	13.6	13.6	15.9
27	Mv Kikelet	13.6	14.9	12.6	14.3	12.5	13.5	13.9	15.5
28	Baletka	13.8	15.9	11.9	14.7	13.8	12.8	13.0	14.5
29	My Lepény	13.3	14.4	12.1	14.4	14.0	12.8	13.6	15.6
30	Babona	11.8	14.6	12.4	14.5	12.3	14.3	14 7	15.5
31	Genius	12.7	14.9	12.5	15.0	12.3	14.0	13.6	15.1
32	My Pengő	11.7	14.4	12.4	14 7	12.8	13.7	14.8	15.5
33	My Lucilla	11.8	13.9	11.9	13.9	12.9	12,8	14.1	16.0
34	GK Körös	12.0	14.0	11,7	15,5	13.3	12,8	13.5	15,0
35	My Nádor	11.6	14.1	11,7	13.8	13,3	12,1	14.5	15,1
36	Amicus	12.2	13.7	11,0	13,0	12.9	13.5	14.6	13,2
37	My Kokárda	12,2	14.5	11,2	14.3	12,9	11.8	12.4	15,5
38	Fidelius	12,0	14,0	11,5	14,5	12,0	13.1	12,4	14.3
30	Ortogus	11.0	12.6	10.8	14.2	11.5	13,1	13,5	14,5
40	Mulan	11,9	13,0	11.2	14,2	11,5	13,5	13,0	13.6
40		11.0	13,5	11,2	17.6	11,7	12.6	16.2	13,0
41	Ralaton	11,5	13,0	11,0	12,0	12.2	12,0	10,2	13,4
42	- Daiatoni Kalahami	11,/	13,1	11,0	12.2	12,2	12,0	12,2	14,0
43		10,8	13,5	11,0	13,2	11,5	13,1	13,5	13,/
44	путапа	11,2	13,2	11,0	13,3	11,5	12,7	13,3	15,8

Minősitett őszi búzafajták nyersfehérje tartalma (%) kisparcellás kísérletekben (gyorsvizsgálat eredményei) Fajtakísérleti Innovációs Tanács, 2014

Forrás: GOSZ-VSZT Őszi Búza Posztregisztrációs Fajtakísérlet 2014

	Faita	Szombat	Bábolna	Tordas	Iregszem	Eszterág	Szarvas	Szákkutas	Debracan
	Гајта	hely	Dabolila	Toruas	cse	puszta	SZalvas	SZEKKULAS	Debiecen
1	Lupus	27,5	36,4	31,3	36,9	36,7	38,0	38,6	39,6
2	GK Berény	36,2	40,2	32,2	33,0	34,0	35,5	35,1	38,1
3	GK Békés	34,3	38,5	33,8	33,7	34,7	35,7	35,2	38,0
4	Antonius	33,0	37,3	31,5	35,9	34,0	35,4	36,7	38,3
5	GK Rozi	34,2	37,1	30,9	32,0	35,9	35,2	34,7	38,0
6	Astardo	30,4	37,6	31,1	35,0	34,2	36,6	32,5	38,3
7	KG Kunhalom	30,5	38,0	31,1	34,7	32,2	32,7	33,9	38,1
8	GK Futár	31,5	36,2	29,9	30,8	33,6	31,5	36,2	37,9
9	Mv Béres	27,2	34,0	30,0	34,4	31,9	34,0	36,7	39,0
10	Bitop	32,3	33,9	28,5	32,8	33,0	31,7	34,4	37,4
11	GK Kalász	34,4	36,4	27,6	31,4	34,5	31,0	30,9	36,4
12	Mv Toldi	31,1	35,8	27,1	34,1	30,1	32,7	33,4	36,5
13	GK Csillag	30,4	35,4	28,5	31,6	33,6	32,0	31,4	37,4
14	Mv Karizma	27,7	35,9	27,7	34,2	29,3	33,7	34,4	36,8
15	Mv Kolo	31,3	36,4	28,5	34,4	31,2	27,8	33,5	35,7
16	Mv Kolompos	25,9	35,2	27,3	32,6	30,0	34,2	35,7	37,9
17	Mv Magdaléna	27,2	34,1	30,2	34,1	30,6	31,1	35,0	36,1
18	Mv Suba	28,5	35,4	29,1	33,8	27,7	32,8	36,5	34,7
19	Gallió	26,3	35,4	29,7	35,0	30,1	32,2	32,7	36,6
20	Mv Tallér	31,8	33,8	29,0	30,0	32,0	30,3	32,1	36,8
21	Genius	27,1	34,5	29,0	34,1	28,5	33,9	30,8	35,8
22	Mv Pántlika	25,7	33,3	25,9	32,2	32,8	31,5	33,2	36,5
23	Mv Karéj	26,1	33,2	27,9	31,2	29,7	31,7	32,5	36,8
24	Babona	24,8	33,0	28,5	31,3	27,4	33,0	33,3	35,5
25	Mv Menüett	27,5	33,6	27,6	31,0	28,4	32,1	31,4	34,9
26	Baletka	29,2	36,0	26,2	31,6	30,7	29,2	28,8	33,9
27	Mv Ködmön	26,3	33,4	25,9	32,8	28,3	30,8	29,9	37,0
28	Mv Pengő	23,5	31,8	26,9	31,5	27,2	31,6	32,5	35,7
29	KG Kunkapitány	29,2	32,6	29,2	28,9	30,2	28,2	26,7	35,6
30	Amicus	26,4	31,0	25,2	29,4	29,5	32,0	33,0	32,4
31	Mv Kikelet	27,9	32,8	26,9	29,5	26,8	29,7	29,9	34,8
32	Mv Lucilla	24,8	31,4	26,2	30,1	27,9	28,6	31,4	36,9
33	Mv Lepény	27,4	31,4	26,3	29,3	30,3	27,6	28,8	34,3
34	Mv Marsall	28,6	33,3	24,0	29,3	29,0	30,6	29,1	31,1
35	GK Körös	24,9	31,5	24,3	32,5	28,7	26,9	30,0	35,3
36	Fidelius	25,6	31,5	25,0	31,5	26,1	29,9	30,0	33,1
37	Mv Nádor	23,9	31,1	25,9	29,3	28,6	27,1	31,7	34,8
38	Mulan	24,0	30,3	24,4	30,9	25,4	31,2	31,1	31,6
39	Altigo	24,2	29,8	26,6	26,4	24,8	28,7	37.9	30,1
40	Ortegus	24,2	30,0	22,5	30,6	24,2	30,8	30,6	33,5
41	Kalahari	22.7	30.4	26.8	28,1	26.3	30.2	29.9	31.9
42	Balaton	24.8	29.2	25.5	29.2	26.2	28.9	26.9	33.3
43	My Kokárda	25.1	32.2	23.2	29.2	26.4	24.9	24.9	33.8
44	Hyland	22.9	28.9	23.3	27.8	24.5	28.3	29.4	31.8

Minősitett őszi búzafajták nedves sikértartalma (%) kisparcellás kísérletekben (gyorsvizsgálat eredményei) Fajtakísérleti Innovációs Tanács, 2014

Forrás: GOSZ-VSZT Őszi Búza Posztregisztrációs Fajtakísérlet 2014

6. Feladat

A Minta fájlok (SAMPLES) között található car_sales.sav fájlt töltse be.

Feladatok:

- Képezzen osztályközöket (Visual Binning) segítségével a Horsepow (LE) változóra!
- A létrehozott változó alapján készítsen kimutatást a gépkocsikról a modell (model) és az árak (price) szempontjából!
- Készítsen leíró statisztikát az értékesítésről (Sales in thousands) gyártók (manufact) és modell (model) alapján!
- Vizsgálja meg, hogy van-e szignifikáns különbség az egyes gyártók által eladott mennyiségek között!
- Vizsgája meg, hogyan befolyásolja a lóerő az eladott mennyiséget!

7. feladat

Nagyfehér malacokkal választást követően speciális takarmánykiegészítőt etettek. Értékelje a kísérlet eredményét. Befolyásolta-e a malacok növekedési erélyét az etetett takarmánykiegészítő?

	Kon	troll	Kísérleti		
Batéria	Választáskori átlag tömeg (kg)	Hízóba állításkori átlagtömeg (kg)	Választáskori átlag tömeg (kg)	Hízóba állításkori átlagtömeg (kg)	
1	6,75	29,90	7,42	33,67	
2	6,45	27,52	7,42	33,24	
3	6,45	27,52	7,10	32,29	
4	6,45	28,95	7,10	31,81	
5	6,45	28,95	7,42	33,67	
6	6,45	28,00	7,10	32,29	
7	7,10	31,33	6,13	25,62	
8	7,10	31,81	7,42	33,67	
9	5,83	27,52	6,77	30,86	
10	4,84	24,67	7,74	33,67	
11	7,42	32,29	6,77	31,33	
12	7,42	32,29	6,45	28,48	
13	7,10	31,81	6,45	29,43	
14	7,10	31,81	7,42	33,24	
15	6,77	29,90	7,10	31,81	
16	7,10	31,81	4,24	25,50	
17	6,77	29,90	6,77	30,86	
18	5,09	27,05	6,45	29,90	
19	7,10	31,81	6,77	30,38	
20	7,42	32,29	7,10	32,29	

Feladatok:

- Készítse el az adatbázist!
- Készítsen leíró statisztikát!
- Vizsgálja meg, hogy van-e szignifikáns különbség a kontroll és kísérleti csoportok súlygyarapodása között.

8. Feladat

Kisüzemi sertések hagyományos (kontroll) és speciális takarmánykiegészítővel etetett (kísérleti) színhús kihozatalát vetettük össze nagyüzemi csoportéval.

Állapítsuk meg a kisüzemi és a nagyüzemi csoportok közötti különbséget, valamint elemezzük a kiegészítés hatását a kisüzemi csoportok adatai között.

Színhús százalék						
Kísérleti	Kontroll	Neguine				
kisüzem	kisüzem	Nagyuzem				
54,7	54,4	58,4				
54,5	56	57,8				
53,5	55,3	57,2				
47,9	54,9	57,7				
50,6	54,1	58				
49,4	57	57,4				

Feladatok:

- Készítse el az adatbázist!
- Készítsen leíró statisztikát!
- Vizsgálja meg, hogy van-e szignifikáns különbség a kontroll és kísérleti csoportok, valamint a nagyüzemi és kisüzemi csoportok színhússzázaléka között.

9. Feladat

Vizsgáljuk meg Európa néhány országában az egy főre jutó GDP hogyan befolyásolja a gépkocsival rendelkező lakosok számát!

Ország	GDP/fő (EUR)	Gépkocsik száma (db/1000 fő)
Ausztria (AT)	28978	496
Belgium (BE)	30349	447
Csehország (CZ)	15216	362
Franciaország (FR)	26656	465
Görögország (GR)	17941	245
Hollandia (NL)	28669	388
Lengyelország (PL)	10135	259
Magyarország (H)	13767	235
Németország (DE)	28232	517
Svájc(CH)	31987	486

Feladat:

- Készítsen grafikont az elemzéshez!
- Regresszióanalízissel határozza meg a két változó kapcsolatát legjobban leíró függvényt! Értelmezze a kapott eredményeket!

• A korrelációs együttható vizsgálatával fogalmazza meg a kapcsolat szorosságát, és értelmezze a determinációs együtthatót!

10. Feladat

Holstein-fríz szarvasmarhák küllemi bírálati adataiból határozzuk meg, hogy a hátulsó láb oldalnézete és a hátulsó láb hátulnézete között milyen kapcsolat van.

Ellés	Hátulsó láb	Hátulsó láb
sorszáma	oldalnézet	hátulnézet
üsző	6,00	5,34
1	5,79	5,43
2	5,93	5,40
3	6,21	5,30
4	6,29	5,18
5	6,43	5,14
6	6,67	4,96
7	6,89	5,11

Feladat:

- Készítsen grafikont az elemzéshez!
- Regresszióanalízissel határozza meg a két változó kapcsolatát legjobban leíró függvényt! Értelmezze a kapott eredményeket!
- A korrelációs együttható vizsgálatával fogalmazza meg a kapcsolat szorosságát, és értelmezze a determinációs együtthatót!

8 Felhasznált irodalom

- Ács Pongrác Gyakorlati • (szerk.): adatelemzés Pécsi Tudományegyetem 2015. Egészségtudományi Kar. 295 ISBN: 978-963-642-723-8. p., http://www.etk.pte.hu/protected/OktatasiAnyagok/%21Palyazati/GyakorlatiAdatelemz es.pdf (2016.08.25.)
- Altbäcker Vilmos, Gácsi Márta, Kosztolányi András, Pogány Ákos, Lakatos Gabriella, Pongrácz Péter, Dóka Antal: Etológiaia gyakorlatok (2013) <u>https://www.tankonyvtar.hu/hu/tartalom/tamop412A/2011-</u> 0073_etologia_gyakorlatok/ch20s02.html (2017.05.06)
- Baráth Csabáné: Biometria Mezőgazda Kiadó 1996, 288 p., ISBN: 9637362312
- Dr. Csallner András Erik: Bevezetés az SPSS statisztikai programcsomag használatába Jegyzet, 2015, SZTE Juhász Gyula Pedagógus Kar Szeged.
- Dr. Illyésné dr. Molnár Emese: Gondolatok a minőség mérhetőségéről és az alkalmazható módszerekről. Tudományos Évkönyv 2007. Budapesti Gazdasági Főiskola 2008.
- Falus Iván, Ollé János: Az empirikus kutatások gyakorlata. Adatfeldolgozás és statisztikai elemzés, Nemzeti Tankönyvkiadó, Budapest, 2008, 341 p., ISBN 978-963-19-6011-2
- Huzsvai László Vincze Szilvia: SPSS könyv, Seneca Books, 2012, 325 p., ISBN: 978-963-08-5666-9 <u>http://seneca-books.hu/doc/spsskonyv.pdf</u>, (2017.04.20)
- Inger Persson & Daniela Capsa: SPSS Manual Quantitative methods (7.5hp) (Statistical Packages for the Social Sciences) Uppsala Universitet Statistiska institutionen, https://studentportalen.uu.se/portal/portal/uusp/student/filearea?uusp.portalpage=true &entityId=118316&toolAttachmentId=280387&toolMode=studentUse&mode=fileare
 a280387 (2017.03.16.)
- Hunyadi Mundroczó -Vita: Statisztika, Aula Kiadó, 2000, 883 p., ISBN:0689000216026
- Jánosa András: Adatelemzés SPSS használatával, Computerbooks, 2011, 376 p., ISBN: 9789636183684
- Korpás Attiláné (szerk.) Általános statisztika I. Nemzeti Tankönyvkiadó, 2012, 231 p., ISBN: 9789631955071

- Korpás Attiláné (szerk.) Általános statisztika II. Nemzeti Tankönyvkiadó, 2011, 299 p., ISBN: 9789631927818
- Kovács Erzsébet: Többváltozós adatelemzés, Typotex Kft, Budapest, 2014, 252 p., ISBN-13 978-963-2792-43-9
- Sajtos László, Mitev Ariel: SPSS kutatási és adatelemzési kézikönyv, Alinea Kiadó, Budapest, 2007, 402 p., ISBN 978-963-9659-08-7.
- Tóthné Parázsó Lenke: A kutatásmódszertan matematikai alapjai 2011. 134 p., TAMOP
 4.2.5 Pályázat könyvei. http://www.tankonyvtar.hu/hu/tartalom/tamop425/0005_31_kutatasmodszertan_scorm
 _02/index.html
- GOSZ-VSZT Őszi Búza Posztregisztrációs Fajtakísérlet 2014 <u>http://www.gabonatermesztok.hu/files/VSZT_14.11.3.pdf (</u>2018.08

Jelen tananyag a Szegedi Tudományegyetemen készült az Európai Unió támogatásával. Projekt azonosító: EFOP-3.4.3-16-2016-00014

