

# MARKETINGKUTATÁS II.

Oktatási segédanyag

Budapest, 2004. február

# Tartalomjegyzék

<b>ELŐSZÓ .....</b>	<b>2</b>
<b>1 AZ SPSS-RŐL ÁLTALÁBAN .....</b>	<b>3</b>
1.1 DATA EDITOR.....	3
1.2 VIEWER .....	4
1.3 CHART EDITOR.....	4
<b>2 ADATBEVITEL.....</b>	<b>5</b>
2.1 ELSŐDLEGES ADATBEVITEL.....	5
2.2 MÁSODLAGOS ADATBEVITEL .....	5
<b>3 ADATELŐKÉSZÍTÉS .....</b>	<b>6</b>
3.1 SELECT (DATA > SELECT CASES).....	6
3.2 COMPUTE (TRANSFORM > COMPUTE) .....	6
3.3 COUNT (TRANSFORM > COUNT).....	7
3.4 RECODE (TRANSFORM > RECODE) .....	7
3.5 CATEGORIZE (TRANFORM > CATEGORIZE VARIABLES) .....	7
<b>4 EGY- ÉS KÉTVÁLTOZÓS ELEMZÉSEK .....</b>	<b>9</b>
4.1 GYAKORISÁGI ELOSZLÁSOK .....	9
4.2 KERESZTTÁBLÁK.....	14
4.3 PARAMÉTERES PRÓBÁK.....	19
<b>5 VARIANCIAELEMZÉS .....</b>	<b>23</b>
5.1 EGYSZEMPONTOS VARIANCIAANALÍZIS .....	23
5.2 TÖBBSZEMPONTOS VARIANCIAANALÍZIS .....	28
<b>6 KORRELÁCIÓELEMZÉS.....</b>	<b>31</b>
6.1 PEARSON-FÉLE (SZORZAT-MOMENTUM) KORRELÁCIÓ.....	31
6.2 PARCIÁLIS KORRELÁCIÓ .....	32
<b>7 REGRESSZIÓELEMZÉS.....</b>	<b>35</b>
<b>8 FAKTORELEMZÉS .....</b>	<b>45</b>
<b>9 KLASZTERELEMZÉS.....</b>	<b>59</b>
9.1 A KLASZTEREK JELLEMZÉSE.....	63
<b>FÜGGELÉK .....</b>	<b>66</b>

# Előszó

Kedves Marketing Szakirányos Hallgatók!

A segédanyag a Marketingkutatás kvantitatív módszerei tantárgy tanulása során a Malhotra könyv gyakorlati kiegészítéseként használható. A gyakorlatokon az egyes alapszereknek az SPSS programcsomag segítségével történő elsajátítását segíti. A segédanyag az azonos helyről letölthető, nyaralási szokásokkal és magatartással foglalkozó kérdőív megkérdezése alapján elkészített adatbázison elvégzett számításokat és az eredmények értelmezését segítő magyarázatokat tartalmazza. Nem tartalmazza az egyes módszerek használhatóságára vonatkozó feltételeket, a módszerek statisztikai bemutatását, vagyis mindazt, amit az egyes módszerekről a Malhotra könyv tartalmaz.

A segédanyag eredményes használatához ajánljuk, hogy a gyakorlatok előtt olvassák el a vonatkozó fejezeteket, ez könnyíti a szemináriumi anyag megértését. Ajánljuk, hogy a kinyomtatott segédanyagot mind az előadásra, mind a gyakorlatokra vigyék magukkal. Így lehetőség nyílik az egyéni jegyzetek beleírására anélkül, hogy a számításokat még egyszer le kellene írni.

A számítások során, ha minden beállítás a segédanyagban leírtak szerint történik az SPSS 11.0 programcsomaggal történő munka során, akkor az eredménynek is meg kell egyeznie az itt leírtakkal. Ha mindezek ellenére nem ugyanez az eredmény, kérjük, jelezzék számunkra.

Mivel a segédanyag új, a korábbi évek segédanyagával nem egyezik meg, vagyis azzal nem pótolható a használata. A korábbi segédanyagokhoz képest nem csak a terjedelme bővült, hanem új adatbázist és az SPSS programnak újabb verzióját használja.

Minden hiba feltárását és egyéb észrevételeiket köszönettel fogadjuk!

A segédanyag összeállításáért, a végzett nagyon hasznos és értékes munkáért köszönetünket fejezzük ki Benke András és Csanda Gergely hallgatóinknak, valamint Pusztai Tamásnak, aki a Marketing szakirányon végzett és a Hoffmann Research International piackutató intézet kutatója.

Sok örömet és jó munkát kívánunk a segédanyag használatához:

Dr. Simon Judit és a tárgy oktatói

# 1 Az SPSS-ről általában

Az SPSS használatakor a jól megszokott Windows-os környezettel találkozhatunk, tehát például az Office termékcsaládnál megismert műveletek az SPSS-ben is rendelkezésre állnak (pl.: Copy/Cut – Paste parancsok), így kezelése nem okozhat gondot. A gyors és hatékony munkát a gyorsbillentyűk, eszközgombok és a „jobb-klikk-pop-up” menük is segítik. Ugyanakkor zavaró lehet, hogy az Office alkalmazásokkal ellentétben az SPSS csak a legutolsó műveletet tudja visszavonni függetlenül a mentési fázistól. Nagy segítséget jelent viszont, hogy bármely eljárás, művelet elvégzése során kitöltött párbeszédablakok mindaddig megőrzik tartalmukat, amíg újra nem indítjuk az SPSS-t. Így egy-egy rosszul vagy hiányosan elvégzett elemzés után gyorsan javíthatjuk a hiányosságot, mindemellett, ha mégis az üres párbeszédablakra van szükség, a Reset gombbal üríthetjük a mezőket.

Az SPSS különféle nézeteivel fogunk találkozni a munkánk során:

## 1.1 Data Editor

Leginkább egy Excel táblázatra (ún. spreadsheet formátum) hasonlít, amely két munkalapból áll, úgymint Data View és Variable View, amelyek a bal alsó sarokban levő fülekre kattintva váltogathatóak.

### 1.1.1 Data View

Minden sor egy-egy megkérdezett lekódolt válaszait tartalmazza (**case**), minden oszlop pedig egy-egy változót (**variable**).

*P: 4. sor 2. oszlopban lévő „1” jelentése: Az egyes sorszámú kérdőív kitöltője volt nyaralni az elmúlt nyáron.*

A táblázat mezőiben a kódszámokat látjuk alapértelmezésben, de a menüsor View > Value Labels kipipálásával a változók kódértékeit is megjeleníthetjük (feltéve, hogy vannak ilyenek). Ez hasznos lehet mindaddig, amíg ismerkedünk az adatbázissal.

Az Excelhez hasonlóan szúrhatunk be sorokat és oszlopokat; amire vigyázni kell, hogy ellentétben a táblázatkezelővel, az SPSS a Paste parancsnál nem tolja automatikusan arrébb az oszlopokat/sorokat, hanem felülírja a kijelölt oszlopot/sort. Tehát a változóink rendezésénél csak üres oszlopnál alkalmazzuk a beillesztést, különben könnyen elveszthetünk adatokat.

### 1.1.2 Variable View

A korábbi SPSS verziókhoz képest a 11.0 lehetővé teszi a változók kényelmes áttekintését és változtatását egy önálló munkalapon. A táblázat sorai az egyes változókat (**variables**), az

oszlopok ezek tulajdonságait tartalmazzák. Egy-egy adott cellára kattintva definiálhatjuk az aktuális változó nevét, típusát, méretét, kódértékeit, stb. a felajánlott beállítások segítségével (legördülő menük, nyilak, egyéb opciók). Egy egész változósor menedzseléséhez (törlés, beszúrás, másolás, stb.) a sor szélén történő jobb-klikkre felnyíló menüben férhetünk hozzá.

*F: Próbaképpen definiáljuk még egyszer az első kérdés változóját!*

*M: Jobb klikk > Insert Variable > Name: k1a, Width: 9, Values: {1=Igen, 2=Nem}*

Vigyázat, ahogy az előbbi példán is láthatjuk, egy új változó definiálása vagy egy régi törlése azonnali változást okoz az adattáblában is!

## **1.2 Viewer**

Funkcionalitását tekintve a Viewer az output táblák megjelenítője. Automatikusan felnyílik, amikor lefuttatunk valamilyen vizsgálatot. Kezelésében kissé hasonlít a Windows Intézőre, hiszen az ablakmező két részre osztott, és a baloldalon fa szerkezetben kezelhetők az outputok. A kutatás tulajdonképpen elemzés része a Viewer segítségével történik.

## **1.3 Chart Editor**

Különböző diagrammok és ábrák megjelenítését teszi lehetővé, a Viewerben lévő diagrammon dupla kattintásra nyílik fel. Alkalmas a diagrammok finomhangolására, és a chartok könnyen exportálhatók más, népszerű formátumba is (pl.: JPEG, TIF, BMP, WMF), amely lehetőség előnyt jelent egy látványos prezentáció elkészítésénél.

Az előzőekben említett nézeteken kívül létezik még Draft Viewer, Pivot Table Editor, Script Editor, stb nézet, de ezekkel a félév során nem foglalkozunk behatóbban.

## 2 Adatbevitel

Az adatbevitelnek alapvetően két módja van, az elsődleges és a másodlagos adatbevitel. Elsődleges adatbevitelnek nevezzük, amikor az adatokat közvetlenül az SPSS programban rögzítjük, másodlagosnak, amikor egy másik alkalmazásban (pl.: Excel, dBase, MS Access) rögzített adatokat importálunk az SPSS-be.

### 2.1 Elsődleges adatbevitel

Az adatbevitelhez szükség van a kódolási útmutatóra és az adatbevivő formra, vagyis hogy milyen változókat definiálunk. Változókat legegyszerűbben a Variable View nézetben „jobb klikk”- **Insert Variable** paranccsal adhatunk meg. Definiáláskor meghatározhatjuk a változó rövid nevét (**Name**), és megadhatjuk az értékeit (**Values**), jelentését (**Labels**); ez hasznos, mert ezután az output táblákon is ezt használja majd az SPSS, ami javítja az áttekinthetőséget. A változó típusának megadását az SPSS nem használja semmire, ez csak a kutatónak nyújthat segítséget (**Measurement**). A **Missing** oszlop lehetővé teszi a hiányzó válaszok kezelését. Az SPSS korábbi verzióival ellentétben a változók egyszerű **Copy – Paste** parancsok segítségével másolhatók Variable View nézetben a gyorsabb változódefiniálás érdekében.

### 2.2 Másodlagos adatbevitel

Az SPSS más alkalmazások adatbázisait is tudja kezelni, ehhez a funkcióhoz a **File** menü **Open Database > New Query** parancsával lehet hozzáférni. Ekkor egy szokásos Windows varázsló segíti az adatbázis SPSS által feldolgozhatóvá alakítását (pl.: otthon xls formátumban bevitt adatbázisunkat könnyen importálhatjuk SPSS-be).

A másodlagos adatbevitelhez sorolhatjuk az adatbázisok egyesítését is. A menüsor **Data > Merge Files** sorával érhetjük el az egyesítési lehetőségeket (**Add Cases/Variables**). Az egyesítésnél ügyelni kell arra, hogy azonos sorokat/oszlopokat egyesítsünk, ugyanis az SPSS logikai hiba esetén sem ad hibajelzést. (Két nyilvánvalóan különböző változó szinkronizálása esetén egyszerűen az eredeti adatbázis változóbeállításait hagyja meg, a hozzáadott változó beállításai eltűnnek.) Az egyesítés elkerülhetetlen, ha többen dolgoznak a lekérdezett kérdőívek kódolásán, ekkor egy mesterséges változóban az egyesítésnél rögzíthetjük, hogy eredetileg melyik adatbázisból származik (lásd: **source01**).

### 3 Adatelőkészítés

Az elemzések elvégzése előtt sok esetben szükség van az eredeti adatok átalakítására, adattisztításra, skálatranszformációra illetve egyes válaszok kizárására. Az SPSS lehetőséget ad mindezekre, a Select és Transform műveletek alkalmazásával.

#### 3.1 Select (Data > Select Cases)

A válaszok közül választhatjuk ki azokat, amelyeket szeretnénk bevonni az adott elemzésbe (illetve zárhatjuk ki a nem megfelelőeket). Leggyakrabban az If lehetőséget szoktuk alkalmazni. Itt beállítható, hogy milyen feltétel (akár függvény is alkalmazható) mellett válogasson az SPSS a mintában. A ki nem választott eseteket törölhetjük, vagy kizárhatjuk az elemzésből (Unselect Cases Are Deleted/Filtered). Amennyiben a Filtered opciót jelöljük meg, a ki nem választott elemek sorszámát áthúzással fogja jelölni az SPSS, illetve egy mesterséges változóban (filter\_\$, utolsó oszlop) is rögzíti az állapotukat. Ez a változó mindig a legutolsó kiválasztási állapotot tükrözi.

*F: Válasszuk ki azokat a válaszadókat, akik legalább 8 napot töltöttek nyaralással, és repülőgéppel utaztak!*

*M: Data>Select Cases>If k9>=8 & k7=1*

*21 ilyen válaszadó van, az első a 9-es.*

Ha valamilyen szelektálást alkalmazunk, azt egy Filter On felirat jelzi az alsó státuszszáv jobb oldalán, ezt érdemes figyelemmel kísérni, mert egy elfelejtett filter gondokat okozhat a további elemzés során. (Az eredeti minta a Select Cases > All Cases kijelöléssel állítható vissza, ekkor eltűnik a felirat.)

#### 3.2 Compute (Transform > Compute)

A művelet segítségével a meglévő változókból különböző számításokkal új változókat hozhatunk létre. A számítások között a számtani és logikai alapl műveleteken kívül rendelkezésre állnak a fontosabb függvények is. Az új változó típusát és címkéit a Target Variable mezőben, a számítás képletét a Numeric Expression mezőben adhatjuk meg.

*F: Számítsuk ki, hogy összesen mennyit költöttek az egyes válaszadók a nyaralás folyamán (13. kérdés)!*

*M: Tranform>Compute> k13\_s = sum(k13\_1, k13\_2, k13\_3, k13\_4, k13\_5, k13\_6)*

### 3.3 Count (Transform > Count)

Az eljárás hasonlít a Selecthez, de a Count nem filterezi a kijelölt eseteket, csak egy mesterséges változóban jelöli meg azokat. Az értékek és feltételek definiálása a fentiekhez hasonlóan történik.

### 3.4 Recode (Transform > Recode)

A Recode segítségével könnyen és egyszerűen alakíthatjuk át változóink kódolását. Ha az átalakítás végleges, és az eredeti változóra már nincs többé szükségünk (ez elég ritka eset – tehát legyünk óvatosak), válasszuk az Into Same Variables opciót, bármilyen más esetben pedig az Into Different Variables lehetőséget (ekkor új változóba kerülnek az újrakódolt értékek).

Ebben az esetben meg kell adnunk az új változó nevét és címkéjét (Output Variable mezők), majd definiálnunk kell az újrakódolási eljárást (Old and New Values gomb).

*F: Tegyük fel, hogy a 7. kérdésben a kutató valamilyen oknál fogva össze akarja vonni a vonat és autóbusz válaszokat a további elemzések előtt. Végezzük el az újrakódolást!*

*M: Tranform>Recode>Into Different Variables>Output Variable: k7uj*

*Old and New Values>*            *2 thru 3 -->2*  
   *7 -->2*  
   *ELSE -->Copy*

Érdemes még megjegyezni, hogy az új változónk értékeinek címkéi nem öröklődnek, hanem újra meg kell adnunk (Variable View > Values).

### 3.5 Categorize (Tranform > Categorize Variables)

Akkor van szükség erre a műveletre, ha valamilyen alacsonyabb rangú elemzés elvégzéséhez az eredetileg metrikus változónkat ordinálissá akarjuk alakítani. Mindehhez csak a kategorizálandó változó(k) nevét és a kategóriák számát kell megadnunk, és az SPSS elkészíti az általa optimálisnak vélt kategorizálást. Ezzel kapcsolatban két probléma merülhet fel:

- Dönteni kell a kategóriák számáról
- Az SPSS által felajánlott kategorizálás nem felel meg a kutatónak.

Mindkét probléma azonos törőlr fakad: az SPSS úgy határozza meg a kategóriákat, hogy mindegyikbe megközelítőleg ugyanannyi eset tartozzon (kategóriák elemszáma megegyezzen). Ez könnyen megoldható, ha az értékek viszonylag homogén eloszlást mutatnak, ellenkező esetben azonban (ha egyes válaszok sok válaszdónál szerepelnek) a kategóriák elem-egyensúlya eltolódhat. Ekkor a kutató döntése oldja meg a problémát: más



kategóriaszámmal próbálkozik, hogy javítsa az elemgyensúlyt (ezt a megoldást a lenti példán lehet tanulmányozni), vagy a **Recode** művelet segítségével egyedi kategória meghatározást végez (ún. „kézi vezérlés”).

*P egyenletes eloszlásra: Kategorizáljuk a havi megélhetésre költött összesen összeget 4 osztályba (23. kérdés), és vizsgáljuk meg a kategóriákat!*

*M: Transform>Categorize Variables>Create Categorize for k23\_6; Number of categories: 4 (Eredmény: nk23\_6 kategorizált változó)*

*Analyze>Descriptive Statistics>Crosstabs>Rows: k23\_6; Columns: nk23\_6*

A kereszttáblában (1. táblázat) megfigyelhető, hogy az SPSS a következő kategóriákat alakította ki:

1. táblázat

<i>Kategória száma</i>	<i>Értékek</i>	<i>Kategória elemszáma</i>
<i>1</i>	<i>-24 999</i>	<i>86</i>
<i>2</i>	<i>25 000-34 999</i>	<i>85</i>
<i>3</i>	<i>35 000-47 999</i>	<i>88</i>
<i>4</i>	<i>48 000-</i>	<i>87</i>

*P kategóriaszám döntésre: Kategorizáljuk a nyaralási költségeket (13. kérdés)!*

*M: Az előbb bemutatott eljárás segítségével kialakítjuk, és vizsgáljuk a kategóriákat! Például a szállás esetén (k13\_2) érdemes 3-ra csökkenteni a kategóriák számát, hiszen a kereszttáblát vizsgálva láthatjuk, hogy a kevesebb kategória egyenletesebb elosztást eredményez.*

*A szórakozási költségeknél (k13\_5) épp fordított a helyzet, itt még egy kategória felvétele eredményezi az egyenletesebb elosztást.*

## 4 Egy- és kétváltozós elemzések

Ebben a fejezetben a gyakorisági eloszlások, a kereszttáblák és a hipotézisvizsgálat gyakorlati megoldásaival fogunk foglalkozni; ezek a legegyszerűbb, ugyanakkor nélkülözhetetlen adatfeldolgozási módszerek (többek között a minta összetételét is ezekkel kell megvizsgálni).

### 4.1 Gyakorisági eloszlások

Az SPSS-ben a gyakorisági eloszlások vizsgálatát az **Analyze > Descriptive Statistics > Frequencies** menüpontban érhetjük el. A megnyíló párbeszédablakban csak ki kell választanunk a vizsgálandó változó(ka)t, és már indíthatjuk is az elemzést. A **Statistics** gombra klikkeléssel módunkban áll a kapcsolatos statisztikai mutatókat is lekérni (helyzetmutatók: medián, módusz, átlag, kvantilisok; szóródási mutatók: szórás, variancia, terjedelem; összeg). A **Charts** lehetőséget ad diagrammok megjelenítésére (hisztogram, illetve oszlop és torta diagramm formák), a **Format** alatt pedig beállítható a megjelenés módja.

Ha csak az összesítő statisztikákra van szükségünk, akkor érdemes az **Analyze > Descriptive Statistics > Descriptives** menüpontot használni, mert sokkal hatékonyabb, hiszen nem készít gyakorisági táblát. (Megjegyezzük: a **Descriptives**-ben elérhető statisztikák mindegyike elkészíthető **Frequencies**-ben is (Malhotra, (2002).)

*F: Készítsük el a minta összetételét jellemző gyakorisági eloszlásokat és alapstatisztikákat!*

*M: Csak a legfontosabbakat mutatjuk be, mint lakhely, tanulmányi eredmények, nem.*

*Analyze>Descriptive Statistics>Frequencies>Variables: k20, k22, k28>Charts: Pie*

2. táblázat

Állandó lakhelyed...

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Budapesten van	126	34,2	35,5	35,5
	Nem Budapesten van	229	62,2	64,5	100,0
	Total	355	96,5	100,0	
Missing	System	13	3,5		
Total		368	100,0		

A 2. táblázatból kiolvasható, hogy összesen 368 válaszadót (Total) vizsgálunk, és 13 válaszadó nem válaszolt erre a kérdésre (Missing). Az oszlopok rendre az érték, gyakoriság, a relatív gyakoriság, az érvényes relatív gyakoriság és a kumulált relatív gyakoriság adatokat

mutatják. Egy példa az értelmezésre: a válaszadóink 35,5%-ának állandó lakhelye Budapesten van (ha a nem válaszolók között is a válaszolók eloszlását feltételezzük).

3. táblázat

**Milyen volt a tanulmányi eredményed az elmúlt félévben az évfolyamátlaghoz viszonyítva?**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Az átlag felett volt (*)	212	57,6	57,9	57,9
	Körülbelül az évfolyam átlaga volt	115	31,3	31,4	89,3
	Az átlag alatt volt	39	10,6	10,7	100,0
	Total	366	99,5	100,0	
Missing	System	2	,5		
Total		368	100,0		

A 3. táblázatban „Az átlag felett volt” eset után található (\*) jelre hívnánk fel a figyelmet. Ez azt hivatott jelezni, hogy az output táblák tetszőlegesen alakíthatóak, formázhatóak Viewer nézetben, az objektumon egy dupla kattintással lehet ehhez hozzáférni; természetesen így került oda a (\*) szimbólum is.

4. táblázat

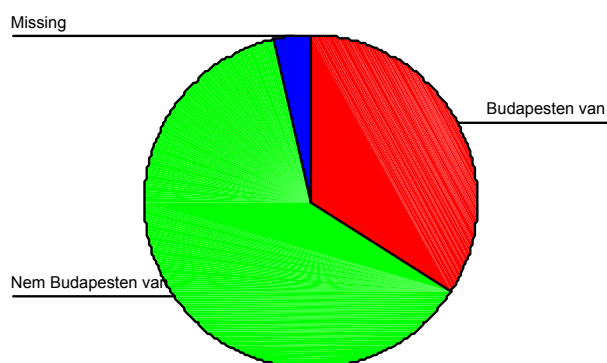
**Nemed?**

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Fiú	147	39,9	39,9	39,9
	Lány	221	60,1	60,1	100,0
	Total	368	100,0	100,0	

A 4. táblázat figyelmeztet minket arra, hogy a mintában felülreprezentáltak a lányok.

I. ábra

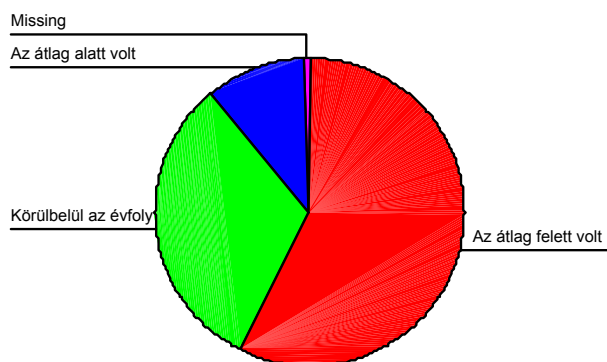
### Állandó lakhelyed...



Az I. ábrán megtekinthetjük torta diagrammon is a lakhely szerinti eloszlást. A kutatási beszámoló prezentálásakor látványos és hasznos megoldás ezeket alkalmazni (a szokásos jobb klikk **Copy objects > Paste** paranccsal illeszthetőek be Power Point-ba).

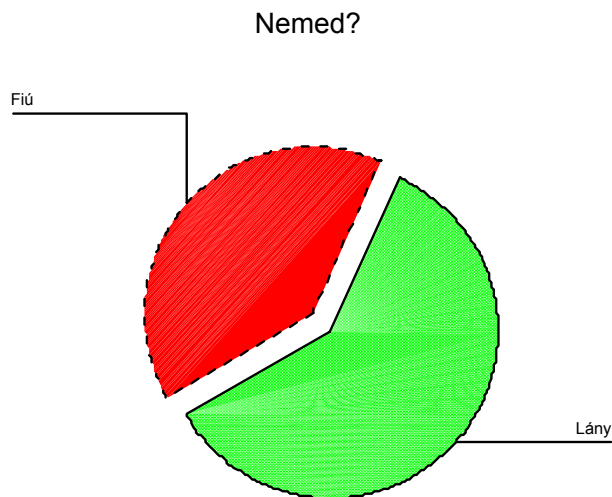
II. ábra

Milyen volt a tanulmányi eredményed az elmúlt félévben  
az évfolyamátlaghoz képest?



A II. ábrán láthatjuk, hogy a válaszadók többségének átlag felett volt az eredménye, kérdés, hogy a marketing tantárgy kutatási részvételén voltak-e felülreprezentálva a jó tanulók, vagy a hallgatók szeretik egy kicsit jobb színben feltűntetni magukat. 😊

III. ábra



A III. ábrán bemutatunk egy-két ábrafinomítási lehetőséget, ami a Chart Editor-ral elvégezhető (objektumon dupla klikkre nyílik fel): tortaszeletek széthúzása, szeletek színének, mintázatának változtatása, szegélyvonalak stílusa, szelet elrendezés változtatása, stb. Szintén a prezentáció és az írásos beszámoló színvonalát, élvezhetőségét növeli. (Ne feledjük, fenn kell tartani a hallgatók, olvasók figyelmét!)

*P: Leíró statisztikák és gyakoriság eloszlás segítségével vizsgáljuk meg, hogy mennyire voltak alanyaink elégedettek a nyaralás során az ellátással (11. kérdés, 8. pont)!*

*M: Analyze>Descriptive Statistics>Frequencies>Variable: k11\_8>*

*Statistics: Percentile Values: Quartiles*

*Central Tendency: minden*

*Dispersion: minden*

*Charts: Bar és Percentages*

5. táblázat

**Statistics**

Mennyire voltál elégedett... - Az ellátással (étkezés, stb.)

N	Valid	344
	Missing	24
Mean		5,3547
Std. Error of Mean		,08708
Median		6,0000
Mode		7,00
Std. Deviation		1,61510
Variance		2,60855
Range		6,00
Minimum		1,00
Maximum		7,00
Sum		1842,00
Percentiles	25	4,0000
	50	6,0000
	75	7,0000

Az 5. táblázat tartalmazza a leíró statisztikákat, és természetesen az érvényes és hiányzó válaszok számát is. Megtalálhatjuk az átlagot, az átlag standard hibáját, a mediánt, módooszt, szórást, varianciát, a terjedelmet, a legkisebb és legnagyobb értékeket, az összeget, és a 25, 50, 75-ös percentiliseket, vagyis az alsó és felső kvartilist és a mediánt még egyszer.

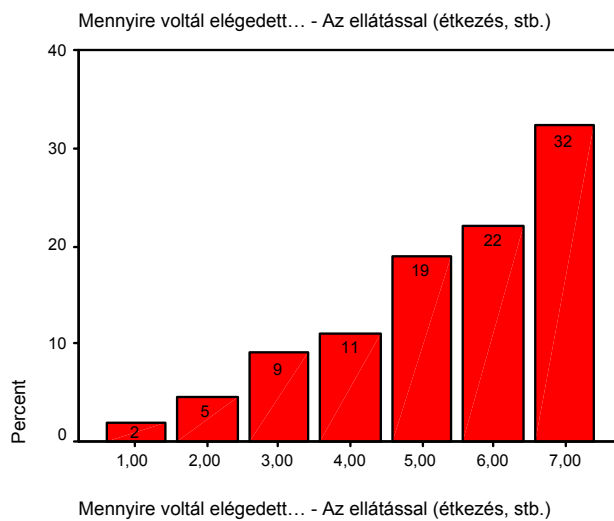
6. táblázat

Mennyire voltál elégedett... - Az ellátással (étkezés, stb.)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1,00	7	1,9	2,0	2,0
	2,00	16	4,3	4,7	6,7
	3,00	31	8,4	9,0	15,7
	4,00	38	10,3	11,0	26,7
	5,00	65	17,7	18,9	45,6
	6,00	76	20,7	22,1	67,7
	7,00	111	30,2	32,3	100,0
	Total	344	93,5	100,0	
Missing	System	24	6,5		
Total		368	100,0		

A gyakorisági eloszlás táblázata a már megismert képet mutatja, azonban az avatott szem máris látja benne az eloszlást, amit a következő oszlopdiagram is látványosan megmutat nekünk.

IV. ábra



Az ábrából kitűnik, hogy a válaszadók igencsak elégedettek voltak az ellátással.

Az oszlopdiagramok is tetszőlegesen csinosíthatók a Chart Editor segítségével, érdemes használni például az oszlop címkéket (Bar Label Styles), és ne feledkezzünk meg a fent említett exportálási lehetőségről sem.

## 4.2 Kereszt táblák

A kereszt tábla elemzés egyike a leggyakrabban használt vizsgálati módszereknek, SPSS-ben az **Analyze > Descriptive Statistics > Crosstabs** útvonalon érhető el. Bár a program bármilyen típusú adatra elvégzi az elemzést, a kereszt táblák input változóit nominális vagy ordinális adatok kell, hogy legyenek.

A **Crosstabs** párbeszédablakában választhatjuk ki az elemzendő változókat, értelemszerűen az egyes változók a sorokban, illetve oszlopokban lesznek feltüntetve. Mivel a kereszt tábla elemzés csak az összefüggés vizsgálatára alkalmas, az irányára nem mond semmit, így a függő és független változó kiválasztása, és ezek elhelyezése a sorokban vagy oszlopokban a kutató ízlésére van bízva, nincs rá általános érvényű szabály.

A szokásos gombok közül a **Statistics**-re és a **Cells**-re érdemes nagyobb figyelmet fordítani (a **Format** ugyanazt kínálja, mint mindenhol máshol). A **Statistics**-ben érhetőek el a kereszt táblák statisztikái: a  $\chi^2$  mutató, a  $\phi$  és a kontingencia együttható, a Cramer-féle V, a  $\lambda$  együttható, és egyéb statisztikák ( $\tau_b$ ,  $\tau_c$ ,  $\gamma$ ,  $\eta$ ), illetve a nem metrikus korreláció mutatói (azokról később, a korrelációnál lesz szó).

A **Cells** gomb segítségével állíthatjuk be a táblázat celláiban megjelenítendő értékeket, egyrészt a gyakoriságra (**Counts**) vonatkozóan (mért (**Observed**) vagy becsült (**Expected**) cellagyakoriság), másrészt a relatív gyakoriságra (**Percentages**) vonatkozóan (százalékos mutatók sor (**Row**), oszlop (**Column**) vagy teljes minta (**Total**) szerint), harmadrészt a mért és becsült adat közti különbségre (**Residuals**) vonatkozóan. Mivel túl sok érték esetén a táblázat áttekinthetetlen lesz, érdemes átgondolni először, hogy melyik változót választjuk függőnek, és melyiket függetlennek. Az általános szabály az, hogy a független változó szerint számítjuk a százalékokat a függő változóra (Malhotra, 2002).

A legtöbb esetben kétváltozós kereszt táblákat szoktunk készíteni, de van lehetőség arra, hogy több változót is bevonjunk az elemzésbe, hiszen ez többlet információval láthat el minket (pl.: kezdeti összefüggés hamis, összefüggés finomítása, stb.) Többváltozós kereszt tábla esetén a harmadik, negyedik, stb. változót egyszerűen bevonjuk a sorokba, oszlopokba (... > **Crosstabs > Rows/Columns**), de vigyázni kell, mert ez minden esetben az áttekinthetőség rovására történik.

*P: Kereszt tábla elemzéssel vizsgáljuk meg, van-e összefüggés a hallgatók neme (28. kérdés) és az egyetemen kívüli Internet hozzáférés (26. kérdés, 2. pont) között!*

*M: Analyze>Descriptive Statistics>Crosstabs>Rows: k28, Columns: k26\_2*

*Statistics:      Chi-square  
                    Phi and Cramer's V*

*Cells:           Counts: Observed  
                    Percentages: Row, Column*

7. táblázat

Case Processing Summary						
	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Nemed? * A következők közül melyekkel rendelkezel? - az egyetemen kívül Internet-hozzáférés	364	98,9%	4	1,1%	368	100,0%

A 7. táblázat a vizsgálat alapadatait tartalmazza, láthatjuk például, hogy négyen nem adtak választ valamelyik kérdésre (gyanítható, hogy ez az Internetre vonatkozó kérdés).



**Nemed? \* A következők közül melyekkel rendelkezel? - az egyetemen kívül  
Internet-hozzáférés Crosstabulation**

			A következők közül melyekkel rendelkezel? - az egyetemen kívül Internet-hozzáférés		Total
			Igen	Nem	
Nemed?	Fiú	Count	121	26	147
		% within Nemed?	82,3%	17,7%	100,0%
		% within A következők közül melyekkel rendelkezel? - az egyetemen kívül Internet-hozzáférés	43,7%	29,9%	40,4%
	Lány	Count	156	61	217
		% within Nemed?	71,9%	28,1%	100,0%
		% within A következők közül melyekkel rendelkezel? - az egyetemen kívül Internet-hozzáférés	56,3%	70,1%	59,6%
Total		Count	277	87	364
		% within Nemed?	76,1%	23,9%	100,0%
		% within A következők közül melyekkel rendelkezel? - az egyetemen kívül Internet-hozzáférés	100,0%	100,0%	100,0%

A kereszt táblát soronként elemezve kiderül, hogy a fiúk nagyobb arányban (82,3%) rendelkeznek egyetemen kívüli Internet hozzáféréssel, mint a lányok (71,9%), de az összesített 76,1%-os arány is elég magasnak mondható. Ha oszloponként elemzünk, akkor elmondhatjuk, hogy az otthoni Internet kapcsolattal rendelkezők 56,3%-a lány, míg az Internettel nem rendelkezők 70,1%-a lány. Tekintve, hogy a mintában 59,6% a lányok aránya, így ugyanarra a következtetésre jutunk, mint az előbb.

### Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	5,235 <sup>b</sup>	1	,022	,024	,014
Continuity Correction <sup>a</sup>	4,678	1	,031		
Likelihood Ratio	5,376	1	,020		
Fisher's Exact Test					
Linear-by-Linear Association	5,221	1	,022		
N of Valid Cases	364				

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expected count is 35,13.

A  $\chi^2$  teszt előrevezet a hipotézisvizsgálathoz, hiszen a 9. táblázatban nem csak az értékét láthatjuk, hanem az ehhez tartozó szignifikanciaszintet is. Mielőtt azonban ezt megvizsgálánk, meg kell bizonyosodni arról, hogy a kereszttáblánk megfelel-e az elemzési követelményeknek (bár ez inkább nagyobb táblák esetén fontos).

(Emlékeztetőül a két feltétel: 1. a várható érték egy cellában sem lehet kisebb, mint 1 és 2. azon cellák aránya, ahol a várható érték kisebb, mint 5, nem lehet több mint 20%.)

A tábla alatt a *b* pontban olvasható megjegyzés szerint a feltételeknek megfelel a kereszttábla (ahogy azt egy 2×2-es táblától el is várjuk).

A  $\chi^2$  teszt nullhipotézise szerint a becült és mért adatok megegyeznek, vagyis nincs összefüggés a két változó között. A  $\chi^2$  értéke 5,235, az ehhez tartozó szignifikanciaszint 0,022, ami azt jelenti, hogy a szokásos 95%-os biztonság mellett (5%-os hibával) a nullhipotézist elvetjük, tehát a két változó között összefüggés van.

10. táblázat

**Symmetric Measures**

		Value	Approx. Sig.
Nominal by	Phi	,120	,022
Nominal	Cramer's V	,120	,022
N of Valid Cases		364	

- a. Not assuming the null hypothesis.
- b. Using the asymptotic standard error assuming the null hypothesis.

A  $\phi$  együtttható nagyon erős statisztika 2×2-es táblák esetén. Ennek szignifikanciája jelen esetben 0,022, ami megerősít minket az összefüggés létét illetően.

*P: Vizsgáljuk meg, hogy van-e összefüggés a lakhely (20. kérdés) és a nyaralási összes kiadás (13. kérdés) között!*

*M: Analyze>Descriptive Statistics>Crosstabs>Rows: k20, Columns: **k13\_skat**(!!!)*

*Statistics: Chi-square  
Phi and Cramer's V*

*Cells: Counts: Observed & Expected  
Percentages: Row*

11. táblázat

**Case Processing Summary**

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Állandó lakhelyed... * Összes költség kat.	318	86,4%	50	13,6%	368	100,0%

Az 50 hiányzó adat mellett is elég nagy marad a mintánk.

12. táblázat

**Állandó lakhelyed... \* Összes költség kat. Crosstabulation**

			Összes költség kat.				Total
			12000 alatt	12001-26000	26001-73800	73800 felett	
Állandó lakhelyed... Budapest	van	Count	27	21	28	32	108
		Expected Count	27,2	27,5	26,2	27,2	108,0
		% within Állandó lakhelyed...	25,0%	19,4%	25,9%	29,6%	100,0%
Nem Budapest	van	Count	53	60	49	48	210
		Expected Count	52,8	53,5	50,8	52,8	210,0
		% within Állandó lakhelyed...	25,2%	28,6%	23,3%	22,9%	100,0%
Total		Count	80	81	77	80	318
		Expected Count	80,0	81,0	77,0	80,0	318,0
		% within Állandó lakhelyed...	25,2%	25,5%	24,2%	25,2%	100,0%

A várható és mért cellagyakoriságok majdnem megegyeznek minden cellában (tehát a  $\chi^2$  értéke kicsi lesz), így már sejthetjük, hogy nem találunk összefüggést.

13. táblázat

**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	3,832 <sup>a</sup>	3	,280
Likelihood Ratio	3,906	3	,272
Linear-by-Linear Association	1,518	1	,218
N of Valid Cases	318		

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 26,15.

Az *a* megjegyzés alapján elemezhetjük a táblát; a  $\chi^2$  értéke (3,832) és szignifikanciája (0,28) mellett a nullhipotézis nem vethető el, nem tudjuk biztosra venni az összefüggés létét.

**Symmetric Measures**

		Value	Approx. Sig.
Nominal by Nominal	Phi	,110	,280
	Cramer's V	,110	,280
N of Valid Cases		318	

- a. Not assuming the null hypothesis.  
 b. Using the asymptotic standard error assuming the null hypothesis.

A Cramer-féle V mutató is megerősíti ezt (0,28-as szignifikancia), tehát a lakhely és a nyaralás alatti költség között nem mutatható ki összefüggés. (A  $\phi$  együtthatót most nem elemezhetjük, hiszen a tábla nem 2×2-es, a számszerű megegyezés a képletből adódik (lásd Malhotra (2002) 548-549.o.).

### 4.3 Paraméteres próbák

Egyszerű hipotézisvizsgálatokat végezhetünk el az **Analyze > Compare Means** menüpont segítségével. A menü lehetőségei közül itt csak a t próbákat mutatjuk be, mert a **Means** és a **One-Way ANOVA** előrevet a varianciaelemzéshez.

Értelemszerűen alkalmazhatjuk a **One-Sample**, az **Independent-Samples** és a **Paired-Samples T test** pontokat egymintás, két független mintás és páros mintás t próbáknál; mindössze az elemzendő metrikus(!) változókat (**Test Variables**) és a konfidencia intervallumot (**Options**) kell beállítanunk, illetve az egyszerű t próbánál tesztstatisztika értékét (**Test Value**).

*F (egymintás t próba): Tegyük, hogy szekunder kutatásból rendelkezésünkre állnak adatok az egyetemisták nyaralási szokásairól a 2000-es évben. Ezek szerint átlagosan 11 napig tartott a leghosszabb nyaralás. Ezt figyelembe véve vizsgáljuk meg, hogyan változtak a nyaralási szokások az eltöltött napok számát tekintve!*

*M: Analyze>Compare Means>One-Sample T Test>Test Variable: k9, Test Value: 11*

**One-Sample Statistics**

	N	Mean	Std. Deviation	Std. Error Mean
Hány napig tartott ez a nyaralás?	347	9,7522	8,37915	,44982

Mint minden próbánál, így itt is az alapadatokat megjelenítő táblázattal találkozhatunk először. Megfigyelhetjük, hogy a mintaátlag (9,7522) jelentősen eltér a tesztértéktől (11).

16. táblázat

### One-Sample Test

	Test Value = 11					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Hány napig tartott ez a nyaralás?	-2,774	346	,006	-1,2478	-2,1326	-,3631

A várakozásnak megfelelően a t próba alapján elvethetjük a nullhipotézist, hiszen a szignifikanciaszint 5% alatt van (0,006), tehát a mintaátlag kisebb mint a tesztérték, vagyis 2002-re az egyetemisták leghosszabb nyaralásának átlagos időtartama csökkent 2000-hez képest.

*F (két független mintás t próba): Tekintsük független részmintának a válaszadóinkat aszerint, hogy A vagy B változatú kérdőívet (source01) töltötték-e ki! Vizsgáljuk meg, hogy van-e különbség a két minta között a nyaralás iránti attitűdök tekintetében, legyen a vizsgált állítás a kulturális úticélra vonatkozó (k18\_17)!*

*M: Analyze > Compare Means > Independent-Samples T Test > Test Variable: k18\_17, Grouping Variable: source01(0 1) (A csoportosítás beállítása: Define Groups gomb)*

17. táblázat

### Group Statistics

	Melyik kérdőívet töltötte ki?	N	Mean	Std. Deviation	Std. Error Mean
Mennyire értesz egyet vele... - Nem érdekel az olyan úticél, amely nem szolgál kulturális érdekességekkel.	B-vált	176	3,8011	1,77287	,13364
	A-vált	187	3,5989	1,74276	,12744

Az induló adatokból látható, hogy független részmintáink számossága hasonló, a szórások és átlagok is majdnem megegyeznek.

18. táblázat

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Mennyire értesz egyet vele... - Nem érdekel az olyan úticél, amely nem szolgál kulturális érdekességekkel.	Equal variances assumed	,020	,888	1,096	361	,274	,2022	,18457	-,16075	,56517
	Equal variances not assumed			1,095	358,822	,274	,2022	,18466	-,16095	,56536

A fenti táblázat nem csak egyszerűen a t próbát tartalmazza, hanem számos egyéb fontos dolgot is megtudhatunk belőle. Mivel a t próba csak akkor végezhető el tiszta lelkiismerettel, ha a független minták szórása megegyezik, így adódik, hogy ezt a Levene teszt F próbájával vizsgáljuk. A Levene teszt az egyetlen vizsgálat, amelynél a szignifikanciaszintet fordítva kell értelmezni, hiszen a  $H_0$  a kedvező alternatíva, vagyis a magas érték (minimum 0,1) a megfelelő számunkra. Ez egészen egyszerűen a hipotézisek felállításából adódik, hiszen a Levene teszt F próbájánál a nullhipotézisben a szórásnégyzetek egyenlőségét fogalmazzuk meg, amit jelen esetben természetesen nem szándékozunk elvetni, hiszen számunkra ez jelenti, hogy a minták alkalmasak a t próbára.

A 18. táblázatban az F értéke kicsi (0,020), a szignifikanciaszint magas (0,888), tehát vizsgálhatjuk a t statisztikákat (vagyis az első sor tartalmazza a releváns értékeket, hiszen teljesül a varianciák egyenlősége feltétel). A t próba szignifikanciája az elfogadott 5%-os határ fölé esik (0,274), így nem vethetjük el a nullhipotézist, tehát a két mintában nincs eltérés a kulturális érdekességek megítélésének tekintetében. (A többi adat is erről tanúskodik: átlagok különbsége, szórások különbsége, stb.)

Az előbb megvizsgált független mintákon kívül érdekes lehet még megvizsgálni a nemek szerinti részmintákat, hiszen ez jellemzően előfordul a gyakorlatban. (Gyakorlásképp meg lehet nézni a különböző nyaralásköltségek eltérését a nemek szerinti bontásban, mind t próbával, mind varianciaelemzéssel.)

*F (páros mintás t próba): Vizsgáljuk meg, hogy válaszadóink szerint szignifikánsan különbözik-e Norvégia és Magyarország árszínvonala!*

*M: Analyze>Compare Means>Paired-Samples T Test>Paired Variables: k16\_2\_1 – k16\_2\_5*

19. táblázat

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Árszínvonal - Norvégia	3,4241	316	3,28980	,18507
	Árszínvonal - Magyarország	5,1551	316	1,37252	,07721

A válaszadók megítélése szerint Magyarország árszínvonala kedvezőbb (5,1551-es átlag 7 fokozatú Likert skálán), mint Norvégiáé (3,4241), ugyanakkor Norvégia megítélésénél sokkal nagyobb szórást fedezhetünk fel, ami a válaszadók véleményének sokszínűségét jelzi.

20. táblázat

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Arszínvonal - Norvégia - Árszínvonal - Magyarország	-1,7310	3,76877	,21201	-2,1481	-1,3139	-8,165	315	,000

A páros mintás t próba kiszámításához ki kellett számítani egy új változót, amely a páronkénti különbségeket fejezi ki; ennek az átlagát (-1,7310) és szórását (3,73877) láthatjuk a táblázatban, és ezt teszteli a t próba. Jelen esetben a t próba szignifikanciaszintje nullához közeli, tehát elfogadhatjuk, hogy a magyar árszínvonal a válaszadók szerint kedvezőbb (értse: alacsonyabb), mint a norvég.

## 5 Varianciaelemzés

A varianciaelemzés az egyik legjobban használható vizsgálati módszer, mindössze egy dologra kell nagyon ügyelni a használatánál, mégpedig a függő és független változók megfelelő megválasztására. Az SPSS ugyanis bármilyen típusú változókra lefuttatja az elemzést, de a varianciaanalízisnek csak akkor van értelme, ha a függő változó metrikus, a független(ek) pedig kategorizáltak. Mivel a kereszttáblával ellentétben itt a kapcsolat irányának is van jelentősége, így a függő és független változók megválasztása már korántsem önkényes.

### 5.1 Egyszempontos varianciaanalízis

Az SPSS-ben az egyszempontos varianciaanalízis két módszerrel is elvégezhető, mindkettő más előnyöket kínál. Az egyik lehetőség az **Analyze > Compare Means > Means** úton érhető el, ekkor a függő és független változók bevonása után, az **Options** menüben kipipálva az **Anova table and eta** lehetőséget, az output tábla tartalmazni fogja az ANOVA-t is. Nagy előnye, hogy a **One-Way ANOVA**-ban nem elérhető  $\eta^2$  számítására is lehetőséget nyújt, mindemellett az alapstatisztikák és a linearitási teszt is rendelkezésünkre állnak.

Az **Analyze > Compare Means > One-Way ANOVA** menü alatt elérhető varianciaelemzés szofisztikáltabb elemzési lehetőségeket kínál (különösen a **Contrasts** és **Post Hoc** gombok alatt), amit mi ezek közül megnézünk, az a független mintás t próbánál megismert Levene teszt, hiszen ez az amiben számunkra előnyt nyújt a **Means**-hez képest. A Levene teszt az **Options** menüben érhető el a **Homogeneity-of-variance** négyzet kipipálásával (itt állíthatók be egyébként a leíró statisztikák is).

*P: Vizsgáljuk meg mindkét módszerrel, hogy van-e összefüggés a nem (k28) és az emberek vendégszeretetével való elégedettség (k11\_7) között!*

*MI: Analyze>Compare Means>Means>Dependent List: k11\_7, Independent List: k28*

*Options: Cell Statistics: Mean, Number of Cases, Standard Deviation  
Anova table and eta  
Test for linearity*



21. táblázat

**Case Processing Summary**

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Mennyire voltál elégedett... - Az emberek vendégszeretetével * Nemed?	347	94,3%	21	5,7%	368	100,0%

A 368 válaszadóból 347-et tudunk bevonni a vizsgálatba.

22. táblázat

**Report**

Mennyire voltál elégedett... - Az emberek vendégszeretetével

Nemed?	Mean	N	Std. Deviation
Fiú	5,2754	138	1,42331
Lány	5,6699	209	1,25234
Total	5,5130	347	1,33505

A mintában 138 fiú és 209 lány van, válaszaik szórása csaknem megegyezik (1,42331 és 1,25234), a lányok átlagosan elégedettebbek (5,6699) a vendéglátó ország embereinek vendégszeretetével, mint a fiúk (5,2754).

23. táblázat

**ANOVA Table<sup>a</sup>**

			Sum of Squares	df	Mean Square	F	Sig.
Mennyire voltál elégedett... - Az emberek vendégszeretetével * Nemed?	Between Groups	(Combined)	12,935	1	12,935	7,392	,007
	Within Groups		603,756	345	1,750		
	Total		616,692	346			

a. With fewer than three groups, linearity measures for Mennyire voltál elégedett... - Az emberek vendégszeretetével \*  
Nemed? cannot be computed.

Az ANOVA F próbája szerint a nullhipotézist el kell utasítani (szignifikancia 0,007), tehát a kategóriaátlagok nem egyeznek meg, vagyis összevetve a 20. táblázat adataival, beigazolódott, hogy a lányok elégedettebbek az emberek vendégszeretetével, mint a fiúk.

24. táblázat

**Measures of Association**

	Eta	Eta Squared
Mennyire voltál elégedett... - Az emberek vendégszeretetével * Nemed?	,145	,021

Ugyanakkor látható, hogy a kapcsolat nem túl erős, hiszen az  $\eta^2$  értéke közel van a nullához (0,021).

*M2: Analyze>Compare Means>One-Way ANOVA>Dependent List: k11\_7, Factor: k28*

*Options: Statistics: Descriptives, Homogeneity-of-variance*

25. táblázat

**Descriptives**

Mennyire voltál elégedett... - Az emberek vendégszeretetével

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Fiú	138	5,2754	1,42331	,12116	5,0358	5,5149	1,00	7,00
Lány	209	5,6699	1,25234	,08663	5,4991	5,8406	2,00	7,00
Total	347	5,5130	1,33505	,07167	5,3720	5,6539	1,00	7,00

A leíró tábla teljesen hasonló a Means Report táblájához, a különbség, hogy a Report tábla tartalma az Options menüben tetszőlegesen alakítható.

26. táblázat

**Test of Homogeneity of Variances**

Mennyire voltál elégedett... - Az emberek vendégszeretetével

Levene Statistic	df1	df2	Sig.
,309	1	345	,578

Az egyik előny a Means elemzéséhez képest, a variancia homogenitási teszt. Láthatjuk, hogy a Levene statisztika értéke alacsony (0,309), a szignifikancia magas (0,578), a nullhipotézist megtartjuk, tehát a csoporton belüli szórás megegyezik, vagyis az ANOVA eredménye nyugodt szívvel elemezhető.

27. táblázat

**ANOVA**

Mennyire voltál elégedett... - Az emberek vendégszeretetével

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	12,935	1	12,935	7,392	,007
Within Groups	603,756	345	1,750		
Total	616,692	346			

Visszatekintve a 23. táblázatra, szembetűnik, hogy a két tábla tartalmilag teljesen megegyezik.

Gyakorlatilag teljesen ugyanazokat az outputokat kaptuk mindkét elemzés esetén, ennek ellenére javasolt mindkettőt elvégezni, hiszen az egyik a Levene tesztet, a másik az  $\eta^2$  mutatót nyújtja pluszban az elemzőnek, márpedig mindkettő értékes lehet.

*P: Vizsgáljuk meg, van-e összefüggés a havi megélhetési kiadások (k23\_6) és a család életszínvonala (k25) között! (A nyaralás témában nem visz előre a kérdés, viszont remekül bemutatható rajta a homogenitási probléma egyik orvoslási módja.)*

*M: Analyze>Compare Means>One-Way ANOVA>Dependent List: k23\_6, Factor: k25*

*Options: Statistics: Descriptives, Homogeneity-of-variance*

28. táblázat

**Descriptives**

A következő kiadási tételeket tekintve... - ÖSSZESEN

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Az átlag alatt	29	35617,93	15223,05046	2826,850	29827,3923	41408,4697	11000,00	79000,00
Átlagos	188	34983,67	17168,09952	1252,112	32513,5893	37453,7511	5000,00	105000,00
Az átlag felett	128	50230,47	72171,42635	6379,113	37607,3551	62853,5824	4000,00	790000,00
Total	345	40693,77	46431,71288	2499,799	35776,9526	45610,5836	4000,00	790000,00

Ránézve a táblára, sejthetjük, hogy gondot fog okozni a csoportokon belüli szórások különbsége.

29. táblázat

**Test of Homogeneity of Variances**

A következő kiadási tételeket tekintve... - ÖSSZESEN

Levene Statistic	df1	df2	Sig.
4,880	2	342	,008

A Levene teszt igazolja félelmünket, a szórások különbözőek, nem érdemes elemezni az ANOVA táblát.

30. táblázat

#### ANOVA

A következő kiadási tételeket tekintve... - ÖSSZESEN

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1,85E+10	2	9259184443	4,379	,013
Within Groups	7,23E+11	342	2114364309		
Total	7,42E+11	344			

Hiába mutat tehát összefüggést az ANOVA, nem fogadjuk el az eredményét. Mit tehetünk ilyenkor? A megoldás: csökkentjük a mintát, szűrjük ki a nagyon kiugró válaszokat, és futtassuk újra az elemzést!

*M: Data>Select Cases>If k23\_6<=80 000 (A megfelelő érték próbálkozással, vagy a kiugró válaszok szemrevételezésével található meg.)*

*Az ANOVA beállítása marad.*

31. táblázat

#### Descriptives

A következő kiadási tételeket tekintve... - ÖSSZESEN

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Az átlag alatt	29	35617,93	15223,05046	2826,850	29827,3923	41408,4697	11000,00	79000,00
Átlagos	185	33940,16	15187,29242	1116,592	31737,1926	36143,1317	5000,00	76000,00
Az átlag felett	116	37944,66	17216,03762	1598,469	34778,3958	41110,9145	4000,00	80000,00
Total	330	35495,24	15993,72141	880,42539	33763,2690	37227,2158	4000,00	80000,00

Rögtön látjuk, hogy javult a helyzet, a szórások közeledtek egymáshoz, ugyanakkor csak 15 esetről (4,3%) kellett lemondanunk, amelyekből 3 az átlagos, 12 az átlag feletti kategóriába tartozott. A veszteség elfogadhatónak mondható.

32. táblázat

#### Test of Homogeneity of Variances

A következő kiadási tételeket tekintve... - ÖSSZESEN

Levene Statistic	df1	df2	Sig.
1,428	2	327	,241

Most már a Levene teszt is viszonylag homogén varianciákat jelez, az ANOVA tábla eredménye elfogadható.

**ANOVA**

A következő kiadási tételeket tekintve... - ÖSSZESEN

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1,14E+09	2	571886786,7	2,253	,107
Within Groups	8,30E+10	327	253865866,7		
Total	8,42E+10	329			

Az F próba már korántsem erősíti meg olyan biztosan az összefüggés létét, sőt a szokásos szignifikanciaszint (5%) mellett nem is fogadhatjuk el az alternatív hipotézist. Ez az eset például szolgál számunkra, hogy a variancia homogenitást érdemes komolyan venni, hiszen ha lemondtunk volna a Levene tesztről, akkor elfogadtunk volna egy olyan összefüggést, ami nem mutatható ki.

## 5.2 Többszemponatos varianciaanalízis

Az SPSS az **Analyze > General Linear Model** menüpont alatt kínálja a magasabb fokú varianciaelemzési módszereket, a **Univariate** alkalmas többszemponatos varianciaelemzésre, a **Multivariate**-tel több függő változót is kezelhetünk, a **Repeated Measures** pedig ismételt mérések varianciaelemzésre ad lehetőséget. Ezek közül az egyszerű több szemponatos varianciaelemzéssel fogunk megismerkedni.

Az említett **Univariate** menüpontban a metrikus függő változót (**Dependent Variable**) és a független változókat (**Fixed Factors**) kell megadni a többszemponatos varianciaelemzés modelljének meghatározásához. A gombok a szokásos lehetőségeket kínálják, a leíró statisztikák (**Descriptive Statistics**), az  $\eta^2$  mutató (**Estimates of effect size**) és a Levene teszt (**Homogeneity tests**) megjelenítését az **Options** menüben tudjuk beállítani. (Figyeljünk arra, hogy az **Estimated Marginal Means** tábla csak akkor jelenik meg, ha az elemzendő változókat áttettük a **Display Means for** ablakba.)

*F: Vizsgáljuk meg, hogy van-e összefüggés a teljes nyaralásköltség és a nyaralás hossza, valamint az úti cél között. Az elemzést többszemponatos ANOVA-val végezzük el!*

*M: Analyze>General Linear Model>Univariate>Dependent Variable: k13\_s, Fixed Factors: k3,k9\_kat*

*Options:      Display Means for:    OVERALL, k3, k9\_kat*

*Display:            Descriptive Statistics*

*Estimates of effect size*

*Homogeneity tests*

34. táblázat

Between-Subjects Factors			
		Value Label	N
Hol nyaraltál az elmúlt nyáron?	1,00	Magyarországon	178
	2,00	Külföldön	146
Hánynap kategorizálva	1	0-6 nap	88
	2	7 nap	76
	3	8-10 nap	95
	4	11-nél több nap	65

A független változók szerinti gyakoriságokat láthatjuk a fenti táblában.

35. táblázat

Descriptive Statistics				
Dependent Variable: Hozzávetőlegesen mennyit költöttél összesen?				
Hol nyaraltál az	Hánynap kategorizálva	Mean	Std. Deviation	N
Magyarországon	0-6 nap	18065,59	18312,46681	70
	7 nap	19561,11	14350,69245	45
	8-10 nap	19753,41	14937,31630	44
	11-nél több nap	34352,89	54505,92978	19
	Total	20599,42	23697,03339	178
Külföldön	0-6 nap	66494,44	69072,42978	18
	7 nap	87774,19	62605,09014	31
	8-10 nap	94321,57	92720,51836	51
	11-nél több nap	191571,8	190801,25060	46
	Total	120141,1	134414,70058	146
Total	0-6 nap	27971,49	39802,25893	88
	7 nap	47384,87	53172,78440	76
	8-10 nap	59784,74	77924,66222	95
	11-nél több nap	145615,5	177837,06840	65
	Total	65454,62	104302,45851	324

A  $2 \times 4$  csoporton belüli átlagok és szórások szemlevételezésével már sejthetjük, hogy a homogenitás nem fog teljesülni.

36. táblázat

#### Levene's Test of Equality of Error Variances

Dependent Variable: Hozzávetőlegesen mennyit költöttél összesen?

F	df1	df2	Sig.
16,649	7	316	,000

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+K3+K9\_KAT+K3 \* K9\_KAT

A már megismert Levene tesztre (lásd Paraméteres próbák vagy Egyszempontos varianciaanalízis fejezet) pillantva beigazolódik a gyanúnk, a csoporton belüli szórások nem egyeznek meg, így nem elemezhetjük az ANOVA eredményeit. A vizsgálatot csakis a példa kedvéért folytatjuk.

## Tests of Between-Subjects Effects

Dependent Variable: Hozzávetőlegesen mennyit költöttél összesen?

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1,152E+12 <sup>a</sup>	7	1,646E+11	22,014	,000	,328
Intercept	1,174E+12	1	1,174E+12	157,031	,000	,332
K3	5,037E+11	1	5,037E+11	67,385	,000	,176
K9_KAT	1,684E+11	3	5,612E+10	7,508	,000	,067
K3 * K9_KAT	9,686E+10	3	3,229E+10	4,319	,005	,039
Error	2,362E+12	316	7474825130			
Total	4,902E+12	324				
Corrected Total	3,514E+12	323				

a. R Squared = ,328 (Adjusted R Squared = ,313)

A táblázatból kiolvasható a varianciaelemzés Malhotra könyvből megismert outputja (lásd Malhotra (2002) 585/16.5. táblázat). Az egyes oszlopok rendre az eltérésnégyzetösszegeket, a szabadságfokokat, az átlagos négyzetösszegeket, az F értékeket, ennek szignifikanciáit és a parciális  $\eta^2$  mutatókat tartalmazzák a teljes modellre, a magyarázó változókra és együttes hatásukra illetve az interakcióra vonatkozóan. Sajnos az ANOVA leggyakoribb mérőszámát, az  $\omega^2$ -t az SPSS nem kalkulálja, ha mégis szükségünk van rá, akkor a táblázat adatait felhasználva magunknak kell a képlet segítségével kiszámítani (például Excelben).

Ha eltekintünk a Levene teszt számunkra negatív eredményétől, akkor a következőképp elemezhetnénk a táblát: az úti cél és az eltöltött napok száma külön-külön és együtt is befolyásolja a nyaralásköltést, hiszen az F próba szignifikanciaszintje nullához közeli mindegyik esetben, ugyanakkor a két magyarázó változó között erős interakció jelentősen befolyásolja az eredményt. A faktorok együttes hatása mindemellett közepes ( $\eta^2 = 0,328$ ), és az egyes faktorok parciális hatása is gyenge.

Az Estimated Marginal Means táblák a beállításoknak megfelelően a függő változó átlagát illetve részátlagát tartalmazzák a kategorizált magyarázó változók szerinti csoportok szerint.

## 6 Korrelációelemzés

A korrelációs számítás két módját fogjuk megismerni az SPSS segítségével, a kétváltozós és a parciális korrelációs számítást. Mindkét elemzés esetén fontos megjegyezni, hogy csak a változók közti kapcsolat szorosságát és előjelét mérjük, az irányát nem (nincs jelentősége, hogy melyik a függő, és melyik a független változó). Ráadásul a nulla korrelációs együttható csak a lineáris kapcsolatot cáfolja, ugyanakkor nem zárja ki más típusú kapcsolat létezését.

### 6.1 Pearson-féle (szorzat-momentum) korreláció

A Pearson-féle korrelációs együttható ( $r$ ), amely két metrikus változó közötti kapcsolat erősségét méri, az SPSS **Analyze > Correlate > Bivariate** menüpontjában érhető el. A páronként vizsgálni kívánt változók beadása után (**Variables**) beállíthatjuk, hogy a Pearson-féle  $\rho=0$  nullhipotézist egy vagy két oldalról tesztelje a program (**Test of Significance**), illetve lehetőség van az alapstatisztikák megjelenítésére az **Options** gomb alatt. Nem metrikus (rang) korrelációt is számíthatunk (sorrendi skálán mért vagy numerikus változók esetén) a **Correlation Coefficients** opció **Kendall's tau-b** illetve **Spearman** kockáinak kipipálásával.

*F: Vizsgáljuk meg milyen szoros a kapcsolat a nyaralással töltött napok száma ( $k_9$ ) és az összes nyaralási kiadás ( $k13\_s$ ) között!*

*M: Analyze>Correlate>Bivariate>Variables:  $k_9, k13\_s$ ; Correlation Coefficients: Pearson*

*Test of Significance: Two-tailed*

*Options: Statistics: mindkettő*

38. táblázat

**Descriptive Statistics**

	Mean	Std. Deviation	N
Hány napig tartott ez a nyaralás?	9,7522	8,37915	347
Hozzávetőlegesen mennyit költöttél összesen?	64894,52	103680,60750	329

A táblázat az alapstatisztikákat tartalmazza, ez csak akkor jelenik meg, ha beállítottuk az **Options** menüben.



## Correlations

		Hány napig tartott ez a nyaralás?	Hozzávetőlegesen mennyit költöttél összesen?
Hány napig tartott ez a nyaralás?	Pearson Correlation	1	,558**
	Sig. (2-tailed)	,	,000
	Sum of Squares and Cross-products	24292,686	161059315,0
	Covariance	70,210	494046,978
	N	347	327
Hozzávetőlegesen mennyit költöttél összesen?	Pearson Correlation	,558**	1
	Sig. (2-tailed)	,000	,
	Sum of Squares and Cross-products	161059315	3,5259E+12
	Covariance	494046,978	10749668371
	N	327	329

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

A megfelelő változók sor-oszlop találkozásánál kiolvasható a keresett  $r$  érték, jelen esetben ez 0,558, ami közepesen erős pozitív kapcsolatot jelez a nyaralás időtartama és a költségek között. (Evidens: minél hosszabb a nyaralás, annál többet költ az ember.)

A hipotézisvizsgálat is megerősíti a kapcsolat létét, hiszen a szignifikancia 0,000. A táblázatban található egyéb adatok a korreláció számítás részeredményei (pl.: kovariancia), ezek az Options menü megfelelő beállításai mellett tűnnek fel.

Hasonlóan elemezhető adatokat kapunk rangkorreláció számításakor is, csak akkor a Nonparametric Correlations táblát kell elemezni.

## 6.2 Parciális korreláció

A parciális korrelációszámítás értelemszerűen az Analyze > Correlate > Partial útvonalon érhető el. Két változó közötti kapcsolat szorosságának mérésére ad lehetőséget, egy vagy több más változó hatásának kontrollálása mellett. A Bivariate parancshoz képest csak annyi a változás, hogy a Controlling for mezőben meg kell adnunk a kontrollált változót is.

*F: Vizsgáljuk az előző feladatban megismert kapcsolat szorosságát a nyaraláskori pénzköltséssel szembeni attitűd (k18\_5) kontrollálása mellett!*

*M: Analyze>Correlate>Partial>Variables: k13\_s, k9 ; Controlling for: k18\_5*

*Többi marad.*

40. táblázat

Variable	Mean	Standard Dev	Cases
K13_S	65345,3642	104364,4940	324
K9	9,7438	8,5601	324
K18_5	3,5741	1,6310	324

Az eddig megszokott output táblákhoz képest ezek kicsit más, egyszerűbb felépítésűek, de szerencsére ugyanolyan jól elemezhetők. Kényelmetlenséget csak a változók azonosítása okozhat.

41. táblázat

- - - P A R T I A L C O R R E L A T I O N C O E F F I C I E N T S - -  
-

Zero Order Partial

	K13_S	K9	K18_5
K13_S	1,0000 ( 0) P= ,	,5582 ( 322) P= ,000	,0665 ( 322) P= ,233
K9	,5582 ( 322) P= ,000	1,0000 ( 0) P= ,	-,0619 ( 322) P= ,266
K18_5	,0665 ( 322) P= ,233	-,0619 ( 322) P= ,266	1,0000 ( 0) P= ,

(Coefficient / (D.F.) / 2-tailed Significance)

" , " is printed if a coefficient cannot be computed

A fenti táblázat az egyszerű korrelációs együtthatókat (r értékeket) tartalmazza; megfigyelhetjük, hogy a megismert kapcsolat szorossági mutatója ezúttal is 0,558 (ez csak akkor van így, ha a vizsgált minta megegyezik).

42. táblázat

- - - P A R T I A L C O R R E L A T I O N C O E F F I C I E N T S - - -

-

Controlling for.. K18\_5

	K13_S	K9
K13_S	1,0000 ( 0 ) P= ,	,5647 ( 321 ) P= ,000
K9	,5647 ( 321 ) P= ,000	1,0000 ( 0 ) P= ,

(Coefficient / (D.F.) / 2-tailed Significance)

" , " is printed if a coefficient cannot be computed

Ebből a táblából olvasható ki a számunkra fontos parciális korrelációs együttható, amelynek értéke (0,5647) azt jelzi, hogy az attitűd kontrollálásának viszonylag gyenge a hatása a kapcsolatra, hiszen az eredeti r értékhez képest csak kicsi növekedést észlelhetünk.

## 7 Regresszióelemzés

A regresszióelemzés során egy metrikus függő és több független változó kapcsolatát vizsgáljuk valamilyen becselőfüggvény illesztésével. Az SPSS-ben lehetőség van különféle modellek felépítésére, ezek közül mi a lineáris illesztéssel fogunk megismerkedni. A két- és a többváltozós lineáris regresszió ugyanazon menüpontban érhető el, így a Malhotra könyvvel ellentétben most együtt tárgyaljuk ezeket. Azonban először – követve a regresszióelemzés folyamatát – az általános modellt kell meghatároznunk. Ehhez nyújt segítséget az **Analyze > Regression > Curve Estimation** menüpont, ahol a kapcsolat (jelen esetben a lineáris kapcsolat) meglétét illeszkedésvizsgálattal ellenőrizhetjük. Az illeszkedésvizsgálat futtatásához meg kell adnunk a függő (**Dependent**) és független változókat (**Independent**). A lineáris regresszió változóinak metrikusnak kell lenniük, hogy alkalmasak legyenek a modellbe illesztéshez, ugyanakkor nem metrikus adatokat is lehet kezelni mesterséges (ún. Dummy) változók létrehozásával. Mivel most a lineáris regresszióval foglalkozunk, így ki kell jelölni a lineáris modellt (**Models > Linear**), illetve célszerű még kipipálni az **Include constant in equation** (konstans elem beépítése a modellbe), a **Plot models** (a diagram megjelenítése) és a **Display ANOVA table** lehetőségeket is.

*F: Vizsgáljuk tovább a fenti kapcsolatot! Milyen lineáris modellel becsülhető a nyaralási költség (k13\_s) az időtartam (k9) tekintetében?*

*M: Analyze > Regression > Curve Estimation > Dependent: k13\_s, Independent/Variable: k9*

*Include constant in equation*

*Plot models*

*Models:           Linear*

*Display ANOVA table*

43. táblázat

MODEL: MOD\_2.

—

Dependent variable.. K13\_S

Method.. LINEAR

Listwise Deletion of Missing Data

Multiple R ,55800

R Square ,31137

Adjusted R Square ,30925

Standard Error 86360,79065

Analysis of Variance:

	DF	Sum of Squares	Mean Square
Regression	1	1095979398894	1095979398894
Residuals	325	2423910502324	7458186161,0

F = 146,94986 Signif F = ,0000

----- Variables in the Equation -----

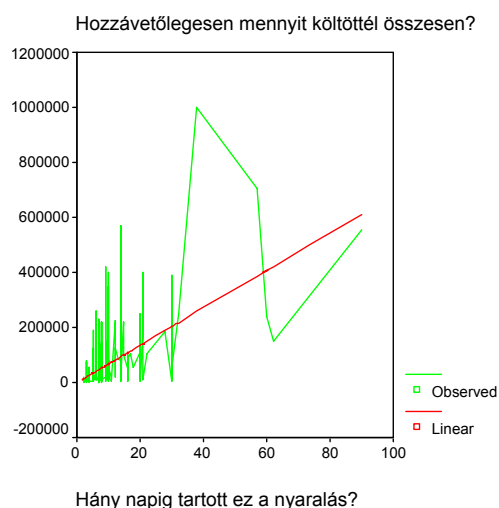
Variable	B	SE B	Beta	T	Sig T
K9	6804,818456	561,347713	,558003	12,122	,0000
(Constant)	-1079,062995	7260,916436		-,149	,8820

A fenti táblázatban megtalálhatjuk az  $R^2$  értékét (0,31137), ami a közepesnél kicsit gyengébb kapcsolatot jelez, egy F próbát, amelynek szignifikanciája a kapcsolat létét engedi sejtetni, és t próbát a  $\beta$  értékekre, amelyek közül a meredekség szignifikanciája kisebb, mint 5 %, így azt mondhatjuk, hogy az eltöltött idő befolyásolja a nyaralásköltést.

A lineáris modellt is megtalálhatjuk a táblában, amely a következőképpen írható fel:

Nyaralásköltés = - 1079,063 + 6804,818 × (nyaralással töltött napok száma)

V. ábra



Az ábra a függő változó (nyaralásköltség) eredeti és becsült értékeit tartalmazza a független változó (nyaralással töltött napok száma) függvényében. Világosan látszik, hogy az illeszkedés nem tökéletes, ugyanakkor valamiféle lineáris trend fellelhető.

Ezek után továbbléphetünk a regresszióelemzés folyamatának következő lépéseire, vagyis ellenőrizhetjük az előrejelzés pontosságát, és megvizsgálhatjuk a reziduumokat, illetve a regressziószámítás egyéb előfeltételeinek teljesülését (pl.: változók normális eloszlása, heteroszkedaszticitás, autokorreláció, multikollinearitás). Mindehhez az SPSS **Analyze > Regression > Linear** menüpontja nyújt segítséget.

Az **Curve Estimates**-hez hasonlóan itt is meg kell adnunk a függő változót (**Dependent**), illetve a független változókat (**Independents**). A lineáris regresszió változóinak metrikusnak kell lenniük, hogy alkalmasak legyenek a modellbe illesztéshez, ugyanakkor nem metrikus adatokat is lehet kezelni mesterséges (ún. Dummy) változók létrehozásával.

A regresszió felépítésének módszerét a **Method** sorban állíthatjuk be, választhatunk a bevonásos, a lépésenkénti, a forward és backward eljárások közül. A **Selection Variable** mező lehetőséget biztosít a válaszadók közti szelektálásra, itt megadhatjuk, hogy melyik változó milyen értékét (megadás: **Rule**) teljesítő esetek kerüljenek a mintába. Hasznos lehet, ha csak egy válaszadói szegmensre akarjuk a modellt felépíteni, illetve kényelmesen végezhetünk eme opció segítségével keresztvényesség vizsgálatot.

A regresszióvizsgálat statisztikái a **Statistics** gombra klikkeléssel érhetőek el. Ha semmit sem jelölünk be, akkor is megkapjuk a legfontosabb eredményeket:  $R^2$  mutató, modell paraméterei és érvényességvizsgálat, ANOVA tábla. Az egyes lehetőségek kipipálásával hozzájuthatunk még egy sor hasznos adathoz: Durbin-Watson mutató az autokorreláció méréséhez,  $R^2$

változásának mutatója a modellbe vont változók kiválasztásához, egyszerű, rész- és parciális korrelációs együtthatók a kapcsolatok szorosságának jellemzéséhez, kollinearitás vizsgálat a multikollinearitás kizárásához, leíró statisztikák, konfidencia intervallumok, kovariancia értékek stb. (Érdemes ezeket kipróbálgatni, tényleg rengeteg érdekes adathoz juthatunk.)

A **Plot** menü segítségével vizsgálhatjuk meg a reziduumokat; a **Scatter** lehetővé teszi pontdiagrammok készítését például a reziduumok (\*ZRESID) és a becsült változó (\*ZPRED) közötti összefüggés vizsgálatára, míg a **Standardized Residual Plots** mezőben a **Histogram** és a **Normal probability plot** négyzeteit kiikszelve lehetőség nyílik a hibátényező normális eloszlására vonatkozó feltételezés tesztelésére. (A **Plot** minden lehetőségét vigyázva alkalmazzuk, mert gyengébb gépeken fagyást okozhat!)

A **Save** és az **Options** gombok a szokásos lehetőségeket kínálják, mindamellett érdekes lehet, hogy a változók bevonásához alkalmazott F próbát milyen paraméterekkel teszteljük (ezt az **Options > Stepping Method Criteria** mezőben definiálhatjuk).

*F: Folytassuk a fenti modell vizsgálatát! Ellenőrizzük a modell pontosságát és a vizsgáljuk a regresszió előfeltevéseinek teljesülését a nyaralási költség (k13\_s) az időtartam (k9) közötti lineáris összefüggésre?*

*M: Analyze>Regression>Linear>Dependent: k13\_s, Independent: k9, Method: Enter*

*Statistics: Estimates  
Model fit  
Descriptives  
Durbin-Watson  
Plots: Scatter Y: \*ZRESID  
X: \*ZPRED  
Standardized Residual Plots (mindkettő)  
Options: marad*

44. táblázat

**Descriptive Statistics**

	Mean	Std. Deviation	N
Hozzávetőlegesen mennyit költöttél összesen?	65221,09	103909,61595	327
Hány napig tartott ez a nyaralás?	9,7431	8,52071	327

A leíró statisztikák mindig hasznosak, legfőképp az átlag és a szórás miatt.

45. táblázat

**Correlations**

		Hozzávetőlegesen mennyit költöttél összesen?	Hány napig tartott ez a nyaralás?
Pearson Correlation	Hozzávetőlegesen mennyit költöttél összesen?	1,000	,558
	Hány napig tartott ez a nyaralás?	,558	1,000
Sig. (1-tailed)	Hozzávetőlegesen mennyit költöttél összesen?	,	,000
	Hány napig tartott ez a nyaralás?	,000	,
N	Hozzávetőlegesen mennyit költöttél összesen?	327	327
	Hány napig tartott ez a nyaralás?	327	327

A korrelációvizsgálat alkalmával már megismert eredményeket itt is elérhetjük, egyébként a korrelációs mátrixnak a többváltozós regresszió esetén van értelme a multikollinearitás megállapítása miatt.

46. táblázat

**Variables Entered/Removed<sup>a</sup>**

Model	Variables Entered	Variables Removed	Method
1	Hány napig tartott ez a nyaralás?	,	Enter

a. All requested variables entered.

b. Dependent Variable: Hozzávetőlegesen mennyit költöttél összesen?

A modell felépítésekor a bevonásos módszerrel dolgoztunk, de ennek is csak több változó esetén van jelentősége.



47. táblázat

**Model Summary<sup>b</sup>**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,558 <sup>a</sup>	,311	,309	86360,79065	1,948

a. Predictors: (Constant), Hány napig tartott ez a nyaralás?

b. Dependent Variable: Hozzávetőlegesen mennyit költöttél összesen?

Az előző elemzésben megismert  $R^2$ -t itt is azt jelzi, hogy a modellünk magyarázó ereje a közepesnél gyengébb. A Durbin-Watson mutató értéke közel 2, így nem veszélyezteti a modell érvényességét az autokorreláció.

48. táblázat

**ANOVA<sup>b</sup>**

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1,10E+12	1	1,096E+12	146,950	,000 <sup>a</sup>
	Residual	2,42E+12	325	7458186161		
	Total	3,52E+12	326			

a. Predictors: (Constant), Hány napig tartott ez a nyaralás?

b. Dependent Variable: Hozzávetőlegesen mennyit költöttél összesen?

Az ANOVA tábla is előkerült már a Curve Estimation-nél is.

49. táblázat

**Coefficients<sup>a</sup>**

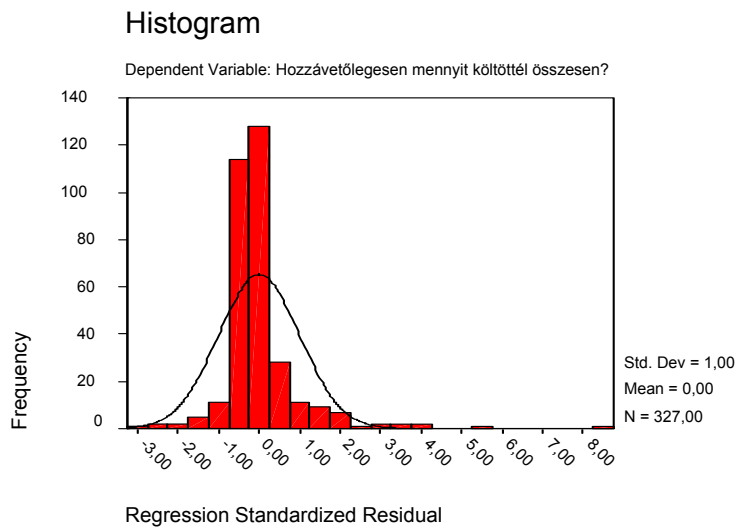
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-1079,063	7260,916		-,149	,882
	Hány napig tartott ez a nyaralás?	6804,818	561,348	,558	12,122	,000

a. Dependent Variable: Hozzávetőlegesen mennyit költöttél összesen?

A lineáris modellt itt is kiolvashatjuk a táblából, vagyis

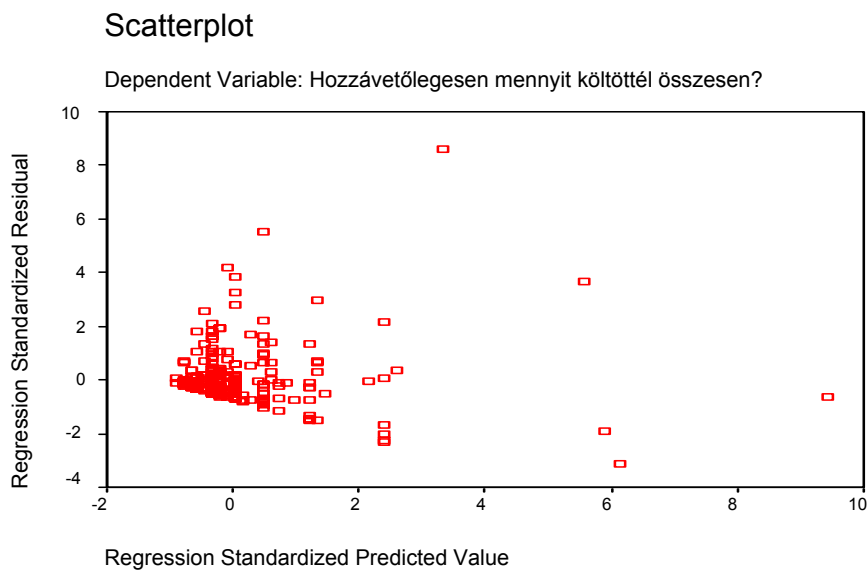
Nyaralásköltség = - 1079,063 + 6804,818 × (nyaralással töltött napok száma). A konstans továbbra sem szignifikáns, innentől kezdve kutatói döntés, hogy bennhagyjuk-e a modellben.

VI. ábra



Jelen esetben elégedettek lehetünk az ábrával, hiszen szép, normális eloszlást mutatnak a standardizált reziduumok.

VII. ábra



A **Scatterplot**-tal már sajnos kevésbé lehetünk elégedettek, ugyanis az ábra jelzi, hogy a reziduumok varianciája nem konstans (lásd bővebben: Malhotra (2002) 627/17.6. ábra), márpedig ez a regressziószámítás egyik alapfeltevése (heteroszkedaszticitás). Ilyen körülmények között a modellt el kell vetnünk; a tanulság az, hogy érdemes mindig először a feltevések teljesülését megvizsgálni, hogy ne elemezzünk feleslegesen.

*F (többváltozós regresszió): Nézzük meg, hogy felépíthető-e lineáris regressziós modell a nyaralással való általános elégedettség, mint függő, és a napok száma illetve a nyaralás fontossága, mint független változók között!*

*M: Először tesztelnünk kell, hogy az egyes független változók lineáris kapcsolatban vannak-e a függő változóval, illetve milyen szoros a kapcsolat, megfelelnek-e a változók az előfeltevéseknek, stb. Ezt az előző feladatban részletesen bemutatott Curve Estimate, illetve Linear menüpontokban végezhetjük el, most eltekintünk ennek a leírásától.*

*Ezután építhetjük fel a többváltozós modellt, szintén a Linear menüpont alatt.*

*Analyze>Regression>Linear>Dependent: k12, Independents: k9, k18\_15; Method: Enter*

*Statistics:       Estimates  
                     Covariance matrix  
                     Model fit  
                     Descriptives  
                     Durbin-Watson*

Csak azokat a táblákat nézzük meg, ahol számottevő újdonság fedezhető fel.

50. táblázat

Correlations				
		Összességében mennyire voltál elégedett ezzel a nyaralással?	Hány napig tartott ez a nyaralás?	Mennyire értesz egyet vele... - A nyaralás az életem egyik központi része.
Pearson Correlation	Összességében mennyire voltál elégedett ezzel a nyaralással? Hány napig tartott ez a nyaralás? Mennyire értesz egyet vele... - A nyaralás az életem egyik központi része.	1,000 ,160 ,192	,160 1,000 ,195	,192 ,195 1,000
Sig. (1-tailed)	Összességében mennyire voltál elégedett ezzel a nyaralással? Hány napig tartott ez a nyaralás? Mennyire értesz egyet vele... - A nyaralás az életem egyik központi része.	, ,001 ,000	,001 , ,000	,000 ,000 ,
N	Összességében mennyire voltál elégedett ezzel a nyaralással? Hány napig tartott ez a nyaralás? Mennyire értesz egyet vele... - A nyaralás az életem egyik központi része.	343 343 343	343 343 343	343 343 343

A korrelációs mátrix lehetőséget nyújt a multikollinearitás vizsgálatára, jelen esetben láthatjuk, hogy a két független változó közötti korreláció kicsi (0,195), tehát nem áll fenn multikollinearitás. Ugyanakkor azt is láthatjuk, hogy a független és függő változók közötti

Pearson korreláció mindkét magyarázó változó esetében alacsony, tehát sejthetjük, hogy a modell magyarázó ereje is alacsony lesz.

51. táblázat

**Model Summary<sup>a</sup>**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,229 <sup>a</sup>	,052	,047	,92319	1,843

a. Predictors: (Constant), Mennyire értesz egyet vele... - A nyaralás az életem egyik központi része., Hány napig tartott ez a nyaralás?

b. Dependent Variable: Összességében mennyire voltál elégedett ezzel a nyaralással?

A sejtést máris igazolja az  $R^2$ -t tartalmazó tábla: a modell magyarázó ereje nagyon gyenge ( $R^2=0,052$ ), ezt a modellt a gyakorlatban nem elemeznénk tovább, most csak azért folytatjuk, hogy bemutathassuk a többi tábla tartalmát. A Durbin-Watson együttható értéke elfogadható, nagyon gyenge pozitív autokorrelációt jelez.

52. táblázat

**Coefficients<sup>a</sup>**

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5,389	,133		40,461	,000
	Hány napig tartott ez a nyaralás?	1,427E-02	,006	,127	2,361	,019
	Mennyire értesz egyet vele... - A nyaralás az életem egyik központi része.	9,412E-02	,030	,167	3,109	,002

a. Dependent Variable: Összességében mennyire voltál elégedett ezzel a nyaralással?

A koefficiens mátrixból kiolvasható (az egyébként gyenge magyarázóerejű modellünk), vagyis:

Nyaralási elégedettség =  $5,389 + 0,01427 \times (\text{nyaralással töltött napok száma}) + 0,09412 \times (\text{nyaralás fontossága})$

Bár a magyarázóerő gyenge, de a t próba alapján a konstans és az együtthatók szignifikánsan különböznek nullától. (Vagyis, ha az  $R^2$ -t nagyobb lenne, akkor ez egy nagyon jó modell lenne, de sajnos a gyakorlatban nem könnyű olyan szép modelleket építeni, mint a Malhotra könyv (2002) mintaadatbázisából.)

Coefficient Correlations<sup>a</sup>

Model			Mennyire értesz egyet vele... - A nyaralás az életem egyik központi része.	Hány napig tartott ez a nyaralás?
1	Correlations	Mennyire értesz egyet vele... - A nyaralás az életem egyik központi része.	1,000	-,195
		Hány napig tartott ez a nyaralás?	-,195	1,000
	Covariances	Mennyire értesz egyet vele... - A nyaralás az életem egyik központi része.	9,164E-04	-3,575E-05
		Hány napig tartott ez a nyaralás?	-3,575E-05	3,651E-05

a. Dependent Variable: Összességében mennyire voltál elégedett ezzel a nyaralással?

A koefficiens mátrix a már megismert korreláción kívül a kovariancia értékeket is tartalmazza.

## 8 Faktorelemzés

A faktorelemzés célja az, hogy a sok, nehezen kezelhető, ugyanakkor egymással korreláló változóból integrált mesterséges változókat állítson elő, így jellemzően a különféle értékelő skálákkal (pl.: Likert skála) mért attitűdkérdések a legmegfelelőbb inputok számára. Statisztikai programunkban a faktoranalízis az **Analyze > Data Reduction > Factor** menüpontban található. A felnyíló **Factor Analysis** ablakban kell beadnunk a kiinduló változókat (**Variables**), és természetesen itt is lehetőség nyílik a válaszadók kiválogatására (**Selection Variable > Value**).

Faktorelemzéskor különösen érdemes odafigyelni az opciók megadására, mert az output táblák óriási mennyiségű adatot tartalmaznak már a legegyszerűbb esetben is, tehát fontos, hogy tényleg csak a használni kívánt elemzéseket állítsuk be, különben könnyen agyonnyomhat minket az adathalmaz. Éppen ezért, most a fontosakra fogunk koncentrálni.

A **Descriptives** gomb alatti lehetőségekből a **Statistics** közül kihagyhatatlan az **Initial solution**, a **Correlation Matrix**-nél pedig a **KMO** és a **Bartlett-próba**. Hasznos lehet még a korrelációs együtthatók (**Coefficients**) és szignifikanciájuk (**Significance Levels**), továbbá a leíró statisztikák (**Univariate Descriptives**) megjelenítése is.

Az **Extraction** menüpontban állíthatjuk be a faktorelemzés módját (**Method**), leggyakrabban a főkomponens módszer szoktuk alkalmazni (**Principal components**). A közös faktorok kiválasztásának feltételét is itt tudjuk megváltoztatni (**Extract**); az alapállapot a Kaiser-kritérium, ekkor az egynél nagyobb sajátértékű faktorokat választjuk ki (**Eigenvalues over 1**), de megadható más küszöbérték, vagy konkrét faktorszám is (**Number of factors**) kutatói döntéstől függően (a faktorok számának meghatározásával később bővebben foglalkozunk). Ugyanitt, a **Display** mezőben adhatjuk meg, hogy megjelenjen-e a rotálás nélküli komponens mátrix (**Unrotated factor solution**). A mező másik sora (**Scree plot**) a sajátértékábra megjelenítésének lehetőségét biztosítja, ez szintén segítségünkre lehet a faktorszám döntésnél. A **Rotation** a faktorok forgatásának beállítását teszi lehetővé, általában a **Varimax**-ot szoktuk használni, de érdemes kipróbálni más lehetőségeket is, bár általában csak nehezen észlelhető változást okoz. A **Display** mezőben hasonló dolgokat állíthatunk be, mint az előbb, a **Rotated solution**-t mindenképpen pipáljuk ki, a **Loading plots** kihagyható (forgatott komponenseket helyezi el egy háromdimenziós, így szükségszerűen nehezen átlátható ábrában).

A **Score** gomb a kapott eredmény elmentésére szolgál; ha megfelelnek a faktorok, válasszuk a **Save as variables**-t, így további elemzések alá vonhatjuk ezeket (Method: Regression marad).

Az **Options** gomb által felkínált lehetőségek közül számunkra a **Coefficient Display Format** a fontosabb, itt érdemes a méret szerint csökkenőt beállítani (**Sorted by size**), a faktorok könnyebb értelmezése és áttekintése érdekében. (Ha nagyon kényelmesek vagyunk, úgy a másik opció is segítséget jelenthet, ez eltünteti a valószínűleg nem releváns értékeket.)

*F: A 18. kérdés remek lehetőséget biztosít számunkra egy faktorelemzés elvégzéséhez.*

*M: Analyze>Data Reduction>Factor>Variables: k18\_1 – k18\_17*

<i>Descriptives:</i>	<i>Statistics:</i>	<i>Initial solution</i>
	<i>Correlation Matrix:</i>	<i>Coefficients</i>
		<i>Significance levels</i>
		<i>KMO and Bartlett's test</i>
<i>Extraction:</i>	<i>Method:</i>	<i>Principal components</i>
	<i>Analyze:</i>	<i>Correlation matrix</i>
	<i>Display:</i>	<i>Unrotated factor solution</i>
		<i>Scree plot</i>
	<i>Extract:</i>	<i>Eigenvalues over 1</i>
<i>Scores:</i>	<i>egyelőre semmi</i>	
<i>Rotation:</i>	<i>Varimax</i>	
<i>Options:</i>	<i>Coefficient Display Format: Sorted by size</i>	

Az első táblázat már ismerősnek tűnhet, ez egy óriási korrelációs mátrix, amely megmutatja a változók közötti páronkénti kapcsolat szorosságát. Akkor alkalmasak a kiinduló adatok faktoranalízisre, ha vannak abszolút értelemben vett magas  $r$  értékek, illetve nullához közeli szignifikanciák a táblázatban.

54. táblázat

**KMO and Bartlett's Test**

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,725
Bartlett's Test of Sphericity	Approx. Chi-Square	1012,382
	df	136
	Sig.	,000

A KMO már egy könnyebben kezelhető mutató a változók közti korrelációra; jelen esetben az értéke 0,725, ez nagyobb, mint 0,6, vagyis a változóink alkalmasak a faktoranalízisre. A Bartlett próba nullhipotézisét is elvethetjük, miszerint a kiinduló változók közötti korreláció nulla.

55. táblázat

Communalities		
	Initial	Extraction
Mennyire értesz egyet vele... - Nyaraláskor elsősorban lustálkodni akarok.	1,000	,594
Mennyire értesz egyet vele... - Szeretek nyaralási terveket szőni.	1,000	,423
Mennyire értesz egyet vele... - A nyaralás alatt nagy hangsúlyt helyezek a társaságra.	1,000	,567
Mennyire értesz egyet vele... - A választásnál a tájnak, a környezetnek nagy szerepe van.	1,000	,594
Mennyire értesz egyet vele... - Nyaraláskor nem figyelek a pénzre.	1,000	,525
Mennyire értesz egyet vele... - Egy elrontott nyaralás nagyon tud dühíteni.	1,000	,462
Mennyire értesz egyet vele... - Nyaraláskor sok friss levegőn való mozgásra van szükségem.	1,000	,691
Mennyire értesz egyet vele... - Nyaraláskor fontosnak tartom a kényelmes szállást és a jó vendéglátást.	1,000	,559
Mennyire értesz egyet vele... - A nyaralás mindennek előtt élményteli legyen.	1,000	,353
Mennyire értesz egyet vele... - A nyaralás során szívesen aktívizálom magam csoportos tevékenységekben.	1,000	,577
Mennyire értesz egyet vele... - Az általam kiválasztott üdülőhely kultúrája és történelme nagyon fontos a számomra.	1,000	,636
Mennyire értesz egyet vele... - A nyaralásokról, utazásokról óráig tudnék beszélni.	1,000	,490
Mennyire értesz egyet vele... - Nyaraláskor a nyugalmat és az egyedüllétet keresem.	1,000	,592
Mennyire értesz egyet vele... - Szívesen ismerkedem meg új emberekkel az üdülés során.	1,000	,568
Mennyire értesz egyet vele... - A nyaralás az életem egyik központi része.	1,000	,536
Mennyire értesz egyet vele... - Ha nyaralok, a természet és az eredetiség élményét keresem.	1,000	,595
Mennyire értesz egyet vele... - Nem érdekel az olyan úticél, amely nem szolgál kulturális érdekességekkel.	1,000	,514

Extraction Method: Principal Component Analysis.



A hosszú kommunalitási tábla a bemenő és kijövő kommunalitás értékeket mutatja a faktorokra (kezdetben ez persze egy). A táblázat alján a módszert láthatjuk (jelen esetben főkomponens módszer).

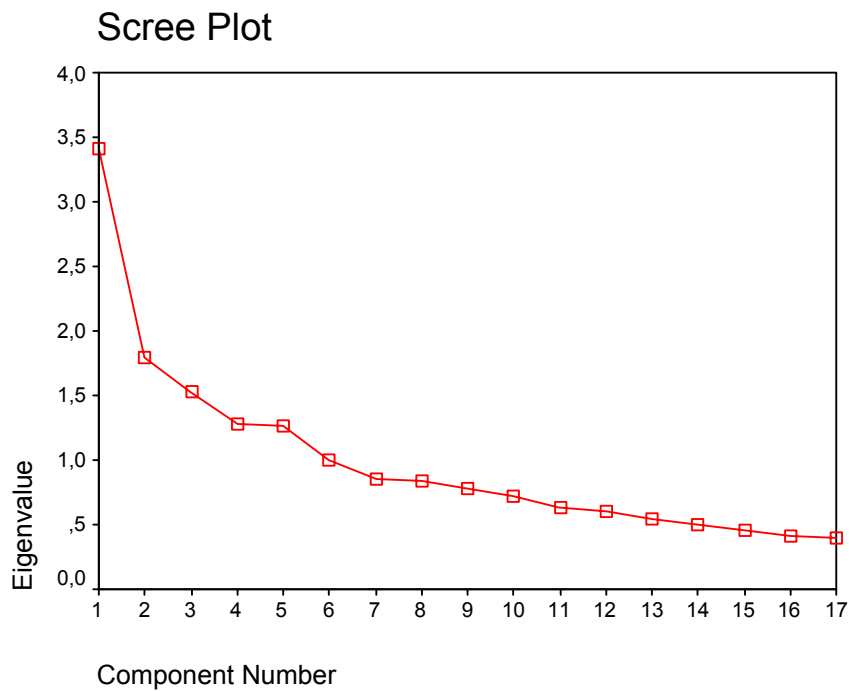
56. táblázat

Total Variance Explained									
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,407	20,041	20,041	3,407	20,041	20,041	2,412	14,186	14,186
2	1,801	10,594	30,634	1,801	10,594	30,634	1,917	11,274	25,460
3	1,527	8,983	39,617	1,527	8,983	39,617	1,754	10,320	35,780
4	1,280	7,530	47,147	1,280	7,530	47,147	1,682	9,894	45,674
5	1,262	7,426	54,573	1,262	7,426	54,573	1,513	8,899	54,573
6	,997	5,866	60,440						
7	,849	4,996	65,435						
8	,836	4,920	70,356						
9	,777	4,569	74,925						
10	,727	4,276	79,201						
11	,627	3,688	82,889						
12	,602	3,540	86,429						
13	,543	3,197	89,626						
14	,497	2,924	92,550						
15	,449	2,641	95,191						
16	,415	2,442	97,633						
17	,402	2,367	100,000						

Extraction Method: Principal Component Analysis.

A főkomponens módszer lényege, hogy azokat a faktorokat választjuk ki, amelyek a legtöbb varianciát magyarázzák, és ez pedig a fenti variancia táblázatban figyelhető meg. Az első oszlop tartalmazza a kiinduló változók számát, a második főoszlop a sajátértékeket és varianciákat a főkomponens módszer alkalmazása után, a harmadik ugyanezeket a kiválasztott faktorokra, a negyedik pedig a rotálás utáni értékeket. A főoszlopokon belül az egyes oszlopokban a sajátértéket, a magyarázott varianciát és ennek kumulált értékét láthatjuk. Jelen esetben a faktorszám meghatározásnál a Kaiser kritériumot használtuk (emlékezzünk, hogy az **Eigenvalues over 1** lehetőséget állítottuk be), így csak az 5 darab egynél nagyobb sajátértékű komponens került kiválasztásra (lásd 2. főoszlop, 1. oszlop). Ez az öt faktor együtt a teljes variancia 54,573%-át magyarázza (lásd 2. főoszlop, 3. oszlop alsó sora). Ez sajnos nem túl jó érték, mert nem éri el a hüvelykujj-szabályként elfogadott 60%-os küszöböt. Felmerül a kérdés, hogy megfelelő-e nekünk az ötös faktorszám, vagy dolgozzunk 6 faktorról, hogy elérjük a 60%-os magyarázott varianciát (60,440%), ráadásul a 6. faktor sajátértéke is majdnem eléri az 1-et (0,997). A döntésben a következő ábra is segítséget nyújthat.

VIII. ábra



A fenti sajátértékábrán könyökpontokat kell keresni (ún. elbow-kritérium), hiszen azokon a helyeken romlik el hirtelen a magyarázott variancia növekedése, ahol törés van a görbén. Ezen az ábrán a 4, illetve 7 faktorszámnál találhatunk könyökpontokat, ezek közül nyilván csak a hetes a releváns, a magyarázott varianciarány nagysága miatt. A 7 faktorszámnál lévő törés megerősít minket abban, hogy a 6 faktoros megoldást érdemes választanunk.

Component Matrix<sup>a</sup>

	Component				
	1	2	3	4	5
Mennyire értesz egyet vele... - Ha nyaralok, a természet és az eredetiség élményét keresem.	,640	,352	-,127	,170	,128
Mennyire értesz egyet vele... - Az általam kiválasztott üdülőhely kultúrája és történelme nagyon fontos a számomra.	,619	,379	-,112	-,277	,139
Mennyire értesz egyet vele... - A nyaralásokról, utazásokról órákig tudnék beszélni.	,593	-,122	,214	-,279	-8,99E-03
Mennyire értesz egyet vele... - A nyaralás az életem egyik központi része.	,586	-,147	,143	-,208	,328
Mennyire értesz egyet vele... - A nyaralás során szívesen aktivizálok magam csoportos tevékenységekben.	,565	-,405	-5,27E-02	,287	9,574E-02
Mennyire értesz egyet vele... - Szeretek nyaralási terveket szőni.	,535	-7,37E-02	4,351E-02	-,292	-,210
Mennyire értesz egyet vele... - A nyaralás mindenképp élménytel legyen.	,513	-,116	-3,20E-02	1,573E-03	-,276
Mennyire értesz egyet vele... - Nem érdekel az olyan úticél, amely nem szolgál kulturális érdekességekkel.	,462	,281	-4,96E-02	-,453	,121
Mennyire értesz egyet vele... - A választásnál a tájnak, a környezetnek nagy szerepe van.	,446	,370	-1,67E-02	,265	-,433
Mennyire értesz egyet vele... - Szívesen ismerkedem meg új emberekkel az üdülés során.	,414	-,380	,144	,372	,305
Mennyire értesz egyet vele... - A nyaralás alatt nagy hangsúlyt helyezek a társaságra.	,179	-,708	,112	5,115E-02	-,139
Mennyire értesz egyet vele... - Nyaraláskor a nyugalmat és az egyedüllétet keresem.	-,115	,563	,261	,203	,391
Mennyire értesz egyet vele... - Nyaraláskor fontosnak tartom a kényelmes szállást és a jó vendéglátást.	-2,57E-02	,222	,705	5,957E-02	-9,58E-02
Mennyire értesz egyet vele... - Nyaraláskor elsősorban lustálkodni akarok.	-,351	3,630E-02	,647	-,103	-,201
Mennyire értesz egyet vele... - Egy elrontott nyaralás nagyon tud dühíteni.	,303	-1,60E-02	,447	-9,18E-02	-,402
Mennyire értesz egyet vele... - Nyaraláskor sok friss levegőn való mozgásra van szükségem.	,450	,250	5,760E-02	,637	-,129
Mennyire értesz egyet vele... - Nyaraláskor nem figyelek a pénzre.	5,998E-02	-,150	,453	1,593E-02	,542

Extraction Method: Principal Component Analysis.

a. 5 components extracted.

A komponens mátrixból olvashatjuk ki, hogy melyik változó melyik faktort jellemzi leginkább. Például az első sorban lévő változó (eredetiség, természet szépségének keresése) leginkább az egyes faktort jellemzi, mert itt van a legnagyobb, abszolút értékben 0,5 feletti faktorsúly (0,640). Megfigyelhetjük, hogy a főkomponens módszer a legtöbb változót az egyes faktorhoz sorolja, ami az értelmezést nehézkessé teszi, ezért van szükség a változók forgatására. A faktorok kialakításánál tehát mindig a rotált komponens mátrixot kell értelmezni.

Rotated Component Matrix <sup>a</sup>

	Component				
	1	2	3	4	5
Mennyre értesz egyet vele... - Az általam kiválasztott üdülőhely kultúrája és történelme nagyon fontos a számomra.	,736	,237	-1,09E-02	-,146	-,132
Mennyre értesz egyet vele... - Nem érdekel az olyan úticél, amely nem szolgál kulturális érdekességekkel.	,706	7,125E-03	-8,26E-02	-8,14E-02	-5,12E-02
Mennyre értesz egyet vele... - A nyaralásokról, utazásokról órákig tudnék beszélni.	,573	6,554E-02	,246	,262	,171
Mennyre értesz egyet vele... - A nyaralás az életem egyik központi része.	,566	-3,43E-02	,458	6,899E-02	-8,77E-03
Mennyre értesz egyet vele... - Szeretek nyaralási terveket szőni.	,518	,140	2,249E-02	,357	8,440E-02
Mennyre értesz egyet vele... - Nyaraláskor sok friss levegőn való mozgásra van szükségem.	-3,01E-02	,787	,244	-,103	2,852E-02
Mennyre értesz egyet vele... - A választásnál a tájnak, a környezetnek nagy szerepe van.	,184	,720	-,165	6,456E-02	9,939E-02
Mennyre értesz egyet vele... - Ha nyaralok, a természet és az eredetiség élményét keresem.	,460	,531	,175	-,187	-,188
Mennyre értesz egyet vele... - Szívesen ismerkedem meg új emberekkel az üdülés során.	1,505E-02	,171	,722	,113	-6,36E-02
Mennyre értesz egyet vele... - A nyaralás során szívesen aktivizálom magam csoportos tevékenységekben.	,135	,284	,575	,334	-,188
Mennyre értesz egyet vele... - Nyaraláskor nem figyelek a pénzre.	8,613E-02	-,271	,565	-,264	,234
Mennyre értesz egyet vele... - Nyaraláskor a nyugalmat és az egyedüllétet keresem.	6,241E-04	,109	6,423E-02	-,740	,169
Mennyre értesz egyet vele... - A nyaralás alatt nagy hangsúlyt helyezek a társaságra.	-,105	-9,88E-02	,398	,620	5,964E-02
Mennyre értesz egyet vele... - A nyaralás mindenekelőtt élményteli legyen.	,295	,335	8,773E-02	,383	5,670E-03
Mennyre értesz egyet vele... - Nyaraláskor fontosnak tartom a kényelmes szállást és a jó vendéglátást.	2,320E-03	9,746E-02	6,982E-02	-,205	,709
Mennyre értesz egyet vele... - Nyaraláskor elsősorban lustálkodni akarok.	-,194	-,199	-,108	-5,25E-02	,708
Mennyre értesz egyet vele... - Egy elrontott nyaralás nagyon tud dühíteni.	,220	,221	-2,77E-03	,293	,528

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

<sup>a</sup>. Rotation converged in 7 iterations.

Az 5 faktorra Varimax rotálást alkalmazva a faktorok sokkal könnyebben elemezhetőbbé váltak. A faktorsúly nagyság szerinti rendezés hatására a táblázat oszlopaiban felülről lefelé haladva könnyedén megtalálhatók az abszolút értékben 0,5-nél nagyobb számok. Ezek alapján az egyes faktorokhoz a következő változók tartoznak:

59. táblázat

Faktor	Változó állítása	Változó kódja	Faktorsúly érték
<b>F1</b>	Az általam kiválasztott üdülőhely kultúrája és történelme nagyon fontos a számomra.	k18_11	0,736
	Nem érdekel az olyan úti cél, amely nem szolgál kulturális érdekességekkel.	k18_17	0,706
	A nyaralásokról, utazásokról órákig tudnék beszélni.	k18_12	0,573
	A nyaralás az életem egyik központi része.	k18_15	0,566
	Szeretek nyaralási terveket szőni.	k18_2	0,518
<b>F2</b>	Nyaraláskor sok friss levegőn valómozgásra van szükségem.	k18_7	0,787
	A választásnál a tájnak, a környezetnek nagy szerepe van.	k18_4	0,720
	Ha nyaralok, a természet és az eredetiség élményét keresem.	k18_16	0,531
<b>F3</b>	Szívesen ismerkedem meg új emberekkel az üdülés során.	k18_14	0,722
	A nyaralás során szívesen aktivizálok magam csoportos tevékenységekben.	k18_10	0,575
	Nyaraláskor nem figyelek a pénzre.	k18_5	0,565
<b>F4</b>	Nyaraláskor a nyugalmat és az egyedüllétet keresem. (-)	k18_13	-0,740
	A nyaralás alatt nagy hangsúlyt helyezek a társaságra.	k18_3	0,620
	A nyaralás mindenekelőtt élményteli legyen. (0)	k18_9	0,383
<b>F5</b>	Nyaraláskor fontosnak tartom a kényelmes szállást és a jó vendéglátást.	k18_8	0,709
	Nyaraláskor elsősorban lustálkodni akarok.	k18_1	0,708
	Egy elrontott nyaralás nagyon tud dühíteni.	k18_6	0,528

A faktorokat a fenti táblázat alapján lehet értelmezni, elnevezni, de mivel már fent eldöntöttük, hogy ideálisabb lenne a 6 faktoros megoldással dolgozni, így most ettől eltekintünk, ehelyett megvizsgáljuk a 6 faktoros esetet.

*F: Készítsük el a 6 faktoros megoldást, és mentsük is le az új változókat!*

*M: Analyze>Data Reduction>Factor>Variables: k18\_1 – k18\_17*

*Descriptives: Statistics:*

*Initial solution*

*Correlation Matrix: Coefficients*

*Significance levels*

*KMO and Bartlett's test*

*Extraction: Method:*

*Principal components*

*Analyze: Correlation matrix*  
*Display: Unrotated factor solution*  
*Extract: Number of factors: 6*  
*Rotation: Varimax*  
*Scores: Save as variables (Method: Regression)*  
*Options: Coefficient Display Format: Sorted by size*

*Csak azokat a táblákat elemzzük, ahol érdemi változás történt.*

60. táblázat

**Total Variance Explained**

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,407	20,041	20,041	3,407	20,041	20,041	1,944	11,437	11,437
2	1,801	10,594	30,634	1,801	10,594	30,634	1,846	10,858	22,295
3	1,527	8,983	39,617	1,527	8,983	39,617	1,786	10,507	32,802
4	1,280	7,530	47,147	1,280	7,530	47,147	1,737	10,219	43,021
5	1,262	7,426	54,573	1,262	7,426	54,573	1,482	8,720	51,740
6	,997	5,866	60,440	,997	5,866	60,440	1,479	8,699	60,440
7	,849	4,996	65,435						
8	,836	4,920	70,356						
9	,777	4,569	74,925						
10	,727	4,276	79,201						
11	,627	3,688	82,889						
12	,602	3,540	86,429						
13	,543	3,197	89,626						
14	,497	2,924	92,550						
15	,449	2,641	95,191						
16	,415	2,442	97,633						
17	,402	2,367	100,000						

Extraction Method: Principal Component Analysis.

A varianciatáblában láthatjuk, hogy ezúttal 6 faktort választottunk ki, és ezek a teljes variancia 60,440%-át magyarázzák.

Rotated Component Matrix <sup>a</sup>

	Component					
	1	2	3	4	5	6
Mennyire értesz egyet vele... - A nyaralásokról, utazásokról órákig tudnék beszélni.	,699	6,809E-02	,168	,156	8,532E-02	6,137E-02
Mennyire értesz egyet vele... - Szeretek nyaralási terveket szőni.	,695	,160	9,262E-02	-6,65E-02	-2,06E-02	,154
Mennyire értesz egyet vele... - A nyaralás az életem egyik központi része.	,631	-3,33E-02	,216	,361	-8,46E-02	-,128
Mennyire értesz egyet vele... - Nyaraláskor sok friss levegőn való mozgásra van szükségem.	-,101	,762	7,480E-02	,312	7,762E-02	-2,35E-02
Mennyire értesz egyet vele... - A választásnál a tájnak, a környezetnek nagy szerepe van.	,265	,749	1,995E-02	-,183	4,434E-02	8,980E-03
Mennyire értesz egyet vele... - Ha nyaralok, a természet és az eredetiség élményét keresem.	,275	,534	,378	,171	-,196	-,193
Mennyire értesz egyet vele... - Nem érdekel az olyan úticél, amely nem szolgál kulturális érdekességekkel.	,138	-4,75E-02	,815	-8,17E-04	5,671E-02	,106
Mennyire értesz egyet vele... - Az általam kiválasztott üdülőhely kultúrája és történelme nagyon fontos a számomra.	,279	,210	,732	2,655E-02	-7,94E-02	-4,42E-02
Mennyire értesz egyet vele... - Szívesen ismerkedem meg új emberekkel az üdülés során.	3,030E-02	,107	3,611E-02	,767	-7,04E-03	,110
Mennyire értesz egyet vele... - A nyaralás során szívesen aktívizálok magam csoportos tevékenységekben.	,256	,240	-2,25E-03	,595	-,182	,259
Mennyire értesz egyet vele... - Nyaraláskor nem figyelek a pénzre.	,113	-,285	2,504E-02	,504	,224	-,356
Mennyire értesz egyet vele... - Nyaraláskor fontosnak tartom a kényelmes szállást és a jó vendéglátást.	-,158	5,786E-02	,153	,112	,787	-9,30E-02
Mennyire értesz egyet vele... - Nyaraláskor elsősorban lustálkodni akarok.	3,699E-02	-,175	-,295	-,181	,656	-,148
Mennyire értesz egyet vele... - Egy elrontott nyaralás nagyon tud dühíteni.	,329	,204	2,751E-02	-1,70E-02	,509	,230
Mennyire értesz egyet vele... - Nyaraláskor a nyugalmat és az egyedüllétet keresem.	-3,65E-02	,172	3,835E-03	-3,42E-02	,106	-,806
Mennyire értesz egyet vele... - A nyaralás mindenekelőtt élményteli legyen.	,113	,266	,326	,192	9,300E-02	,497
Mennyire értesz egyet vele... - A nyaralás alatt nagy hangsúlyt helyezek a társaságra.	,303	-,126	-,365	,371	1,255E-02	,445

Extraction Method: Principal Component Analysis.  
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 10 iterations.



Természetesen csak a rotált faktor mátrix a releváns számunkra, ez alapján elkészíthetjük a faktorainkat jellemző táblázatunkat.

62. táblázat

Faktor	Változó állítása	Változó kódja	Faktorsúly érték
F1	A nyaralásokról, utazásokról órákig tudnék beszélni	k18_12	0,699
	Szeretek nyaralási terveket szőni.	k18_2	0,695
	A nyaralás az életem egyik központi része.	k18_15	0,631
F2	Nyaraláskor sok friss levegőn való mozgásra van szükségem.	k18_7	0,762
	A választásnál a tájnak, a környezetnek nagy szerepe van.	k18_4	0,749
	Ha nyaralok, a természet és az eredetiség élményét keresem.	k18_16	0,534
F3	Nem érdekel az olyan úti cél, amely nem szolgál kulturális érdekességekkel.	k18_17	0,815
	Az általam kiválasztott üdülőhely kultúrája és történelme nagyon fontos a számomra.	k18_11	0,732
F4	Szívesen ismerkedem meg új emberekkel az üdülés során.	k18_14	0,767
	A nyaralás során szívesen aktivizálom magam csoportos tevékenységekben.	k18_10	0,595
	Nyaraláskor nem figyelek a pénzre.	k18_5	0,504
F5	Nyaraláskor fontosnak tartom a kényelmes szállást és a jó vendéglátást.	k18_8	0,787
	Nyaraláskor elsősorban lustálkodni akarok.	k18_1	0,656
	Egy elrontott nyaralás nagyon tud dühíteni.	k18_6	0,509
F6	Nyaraláskor a nyugalmat és az egyedüllétet keresem. (-)	k18_13	-0,806
	A nyaralás mindenekelőtt élményteli legyen.	k18_9	0,497
	A nyaralás alatt nagy hangsúlyt helyezek a társaságra.	k18_3	0,445

A táblázat alapján el tudjuk nevezni a 6 faktort (ne feledjük, hogy ezek attitűdöket jelölnek), és mivel ezeket elmentettük a változók közé, így érdemes ellátni az új változókat (fac1\_1, ..., fac1\_6) címkével is.

F1 (fac1\_1): fontos a nyaralás

F2 (fac2\_1): természet- és mozgásorientáltság

F3 (fac3\_1): kultúraorientáltság

F4 (fac4\_1): társaságkeresés és kiadási hajlandóság

F5 (fac5\_1): komfort és kényelem fontossága

F6 (fac6\_1): élményorientáltság

*F: Vizsgáljuk meg, hogy az elmentett faktorainknak mekkora az átlaga és szórása!*

*M: Analyze>Descriptive Statistics>Descriptives>Variables: fac1\_1, ..., fac6\_1*

## Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Faktor: fontos a nyaralás	337	-2,79639	2,56464	,0000000	1,00000000
Faktor: természet és mozgás	337	-3,93629	2,45399	,0000000	1,00000000
Faktor: kultúra	337	-2,81293	2,68820	,0000000	1,00000000
Faktor: társaság és kiadás	337	-3,26322	3,35619	,0000000	1,00000000
Faktor: komfort és kényelem	337	-2,84320	2,48126	,0000000	1,00000000
Faktor: élmények	337	-2,81555	1,96126	,0000000	1,00000000
Valid N (listwise)	337				

Minthogy a faktorok standardizált változók, az átlaguk nulla, a szórásuk 1, ezt igazolja a táblázat is.

*F: Vizsgáljuk meg, hogy az első 20 válaszadóra milyen értékekkel rendelkeznek a faktorok, és jellemezzük a 15. elemet!*

*M: Analyze>Report>Case Summaries>Variables: fac1\_1, ..., fac6\_1*

*Display cases*

*Limit cases to first 20*

*Show only valid cases*

*Show case numbers*

Case Summaries<sup>a</sup>

	Case Number	Faktor: fontos a nyaralás	Faktor: természet és mozgás	Faktor: kultúra	Faktor: társaság és kiadás	Faktor: komfort és kényelem	Faktor: élmények
1	1	-,32220	-,44763	1,49027	-,44772	,59914	-1,74464
2	2	,10439	,17210	-,51451	,46469	,00666	,04874
3	3	-2,09989	,66861	2,09574	1,55166	,33588	-,28199
4	4	1,45928	-1,12860	,26350	1,08668	-1,28934	-1,44797
5	5	-,27775	-,74854	-,06184	,19132	,40193	,10111
6	6	1,73643	-,08853	,69687	1,07864	,35258	,45704
7	7	-,03400	-2,65006	1,33604	-,28541	-,83721	1,08983
8	8	-,12411	-1,43300	1,71789	-1,19862	-,80693	,42969
9	9	,81350	-,22644	,25553	,36305	1,01478	-,56233
10	10	,92965	-,24212	1,56380	-,64397	,92437	-,36369
11	11	,36306	-,32608	1,25062	1,31767	,05815	1,51967
12	12	,51969	-1,07285	-,49521	,88164	-,28668	1,11780
13	13	,04386	-2,36989	,40764	-,70871	,29227	-,33343
14	14	-,14211	,70954	1,22283	-1,46278	-,10219	-,83781
15	15	-,19680	-,63622	-,86387	-,34892	1,67514	-2,34621
16	16	,43316	1,57528	-,10973	,89987	-1,14150	,21083
17	17	,81967	-2,63720	-2,29962	,05655	-2,84320	-1,36368
18	18	-,38129	,46533	-,39262	-,14029	,98764	-,47571
19	19	,04502	-,56651	,70352	1,25699	-,15892	,63802
20	20	-,26533	-,80001	-,46967	-,46833	,69613	,64593
Total	N	20	20	20	20	20	20

a. Limited to first 20 cases.

Egy egyszerű Report-tal megvizsgálhatjuk az első 20 eset értékeit. A 15. válaszadó tipikus passzív pihenő: nem fontos számára a természet és a mozgás (-0,63622), a kultúra (-0,86387), az élmények (-2,34621), viszont fontos a kényelem és a komfort (1,67514).

## 9 Klaszterelemzés

A klaszterelemzés célja, hogy a megfigyelt eseteket homogén csoportokba rendezze a kiválasztott változók szerint. Féléves tanulmányaink során a metrikus változókra alkalmazható hierarchikus klasztermódszerrel ismerkedünk meg, ezt az SPSS az **Analyze > Classify > Hierarchical Cluster** menüpontban teszi elérhetővé. A klaszterképzéshez mindenekelőtt meg kell adnunk a csoportosító változókat (**Variables**), amelyeknek metrikusnak kell lenniük. A **Cluster** mezőben a **Cases** opciót kell kijelölnünk, ugyanis ez jelenti a válaszadók csoportosítását (a **Variables** a változók csoportosítására szolgál, ez a faktoranalízishez hasonló eljárás, de ezzel nem foglalkozunk). A **Display** mező lehetőségeit kipipálva lesz aktív a két gomb (**Statistics**, **Plots**). Két dologra kell ügyelni: 1) ha nem jelölünk ki semmit, akkor nem lesz output táblája az elemzésnek (erre figyelmeztet is az SPSS); 2) a **Plots** opció nagyobb ábrái könnyen lefagyást okozhatnak gyengébb gépeken. Összegezve: a gyakorlatban általában csak a **Statistics** opciót használjuk. Nézzük meg, hogy mit lehet beállítani ebben az opcióban!

Az **Agglomeration schedule** kipipálásával megjeleníthetjük az összevonási séma táblázatát (mint Malhotra (2002) 708/20.2. táblázat), a **Proximity matrix** kijelölésével pedig az elemtávolság mátrixot. A **Cluster Membership** alatt beállíthatjuk, hogy milyen klaszterszámú megoldásokat kívánunk megjeleníteni.

A **Plots** menüpont alatt érhető el a dendrogram (Malhotra (2002) 711/20.8. ábra) és a jégcsap diagram (**Icicle**), különböző tájolással (**Horizontal/Vertical**); különösen ez utóbbi kiírása okozhat fennakadást a program futásában (ráadásul igencsak hely és memóriaigényes).

A **Method** gomb lenyomásakor felnyíló ablakban állíthatjuk be a klaszterelemzés módszerét. A Ward-féle eljárás (**Ward's method**) a **Cluster Method**, az euklideszi távolságmérték (**Squared Euclidean distance**) pedig a **Measure > Interval** legördülő menüjében választható ki, míg a további mezőkben az értékek (**Transform Values**) és mértékek (**Transform Measures**) átalakítása oldható meg.

A **Save** opcióban a számunkra megfelelő számú klasztermegoldás elmentéséről gondoskodhatunk; az új változó (pl.: **clu6\_1**) megmutatja, hogy az adott esetet melyik klaszterbe sorolta az elemzés.

*F: A faktoranalízis során előállított változók segítségével soroljuk válaszadóinkat homogén csoportokba!*

*M: Analyze>Classify>Hierarchical Cluster>Variables: fac1\_1, ..., fac6\_1; Cluster: Cases, Display: Statistics*

Statistics: Agglomeration schedule  
Cluster Membership: None  
Method: Ward's method  
Measure>Interval: Squared Euclidean distance  
Save: egyelőre None

65. táblázat

Case Processing Summary<sup>a,b</sup>

Cases					
Valid		Missing		Total	
N	Percent	N	Percent	N	Percent
337	91,6	31	8,4	368	100,0

a. Squared Euclidean Distance used

b. Ward Linkage

Az elemzésbe 337 esetet tudtunk bevonni. A megjegyzésekben láthatjuk, hogy milyen távolságmértéket és eljárást alkalmaztunk.

66. táblázat

Agglomeration Schedule

Stage	Cluster Combined		Coefficient s	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	59	60	,000	0	0	136
2	21	49	,000	0	0	322
3	101	159	,069	0	0	77
4	229	315	,162	0	0	39
5	58	72	,282	0	0	54
6	96	149	,406	0	0	40
7	56	312	,565	0	0	80
8	225	329	,730	0	0	112
9	18	80	,903	0	0	109
10	26	222	1,089	0	0	20
...	...	...	...	...	...	...
321	2	3	935,867	315	296	325
322	21	31	965,350	2	318	332
323	7	38	1000,297	307	297	333
324	1	28	1036,301	310	319	329
325	2	18	1072,359	321	283	328
326	13	17	1109,119	316	306	331
327	15	32	1149,813	308	314	331
328	2	6	1199,888	325	317	332
329	1	4	1254,895	324	303	334
330	16	24	1324,243	320	313	335
331	13	15	1397,843	326	327	334
332	2	21	1475,658	328	322	333
333	2	7	1581,332	332	323	336
334	1	13	1704,684	329	331	335
335	1	16	1849,583	334	330	336
336	1	2	2016,000	335	333	0

Az összevonási séma táblázata a legfontosabb számunkra a klaszterelemzés output táblái közül, hiszen ez alapján hozhatunk döntést a klaszterek számát illetően. A koefficiens oszlop adatai alapján alkalmazhatjuk a könyök-kritériumot, vagy az 50%-os szabályt, de bizonyos esetekben egyedi kutatói döntésre is hagyatkozhatunk. Jelen helyzetben az 50%-os szabály irreálisan sok klasztert eredményezne (akár 14 darab is lehetne), a könyök-kritérium használhatatlan, mert nincs nagyléptékű homogenitás növekedés, így egyedi döntésre kell hagyatkoznunk.

*F: Ennek meghozatalához mentsük el az 5, 6, 7 klaszteres megoldásokat, és elemezzük ezeket tovább!*

*M: Analyze>Classify>Hierarchical Cluster> minden beállítás marad*

*Save: Range of solutions: From 5 through 7 clusters*

A különböző megoldások mind ANOVA-val, mind Means-zel jól elemezhetők, most a Means lehetőséget mutatjuk be, mert talán egy hajszálnyival könnyebben értelmezhető (viszont le kell mondanunk a Levene tesztről).

*M: Analyze>Compare Means>Means>Dependent List: fac1\_1, ..., fac6\_1; Independent List: clu7\_1, clu6\_1, clu5\_1*

*Options: Cell Statistics: Mean, Number of Cases, Standard Deviation*

*Statistics for First Layer: Anova table and eta*

67. táblázat

Report							
Ward Method		Faktor: fontos a nyaralás	Faktor: természet és mozgás	Faktor: kultúra	Faktor: társaság és kiadás	Faktor: komfort és kényelem	Faktor: élmények
1	Mean	,8772143	,2979500	,5039085	-,5306434	,4932752	-,5239461
	N	57	57	57	57	57	57
	Std. Deviation	,77086327	,64613787	,88349043	1,0009469	,76314787	1,0543040
2	Mean	-,0368256	,0790914	,0522119	,6531306	,6147837	,2683635
	N	102	102	102	102	102	102
	Std. Deviation	,77748603	,80569343	,86721864	,59203796	,62642374	,70041637
3	Mean	-,2304132	-1,1729451	1,4142204	,0000739	-,4912386	,7867606
	N	24	24	24	24	24	24
	Std. Deviation	1,09213862	,98531752	,52582942	,79416440	,93711646	,61906533
4	Mean	,3998040	-1,1561123	-,8884123	-,6747493	-,5858157	-,1759571
	N	27	27	27	27	27	27
	Std. Deviation	,70569468	,89069538	1,03110405	1,0442147	,89099457	,86388954
5	Mean	-,9405108	,0444844	-,7849394	-,9106595	,5212924	-,7765359
	N	34	34	34	34	34	34
	Std. Deviation	,84917751	1,15169791	,79314670	,71873215	,72709986	1,0603407
6	Mean	,0359515	,5770567	-,1370760	-,1234147	-,9618310	,3984597
	N	76	76	76	76	76	76
	Std. Deviation	,89137858	,59510630	,76758988	,73386426	,72918856	,89362038
7	Mean	-1,3096903	-,6502082	-,4056966	1,3050446	-,4613329	-,9129574
	N	17	17	17	17	17	17
	Std. Deviation	,74904556	1,28453232	,71333969	,93663156	,96643476	,92085282
Total	Mean	,0000000	,0000000	,0000000	,0000000	,0000000	,0000000
	N	337	337	337	337	337	337
	Std. Deviation	1,00000000	1,00000000	1,00000000	1,00000000	1,00000000	1,00000000

A táblázatot elemezve megfigyelhetjük a klasztertagságot: a klaszterek rendre 57, 102, 24, 27, 34, 76 illetve 17 elemet tartalmaznak a 7 klaszteres megoldás esetén. Általában nem érdemes olyan megoldást elemezni, ahol vannak nagyon alacsony elemszámú klaszterek. Fontos még az is, hogy az egyes csoportokon belül ne legyen egynél nagyobb szórás, hiszen standardizált változókat vontunk be az elemzésbe (amelyek átlagszórása 1). Ebben az esetben elég sok helyen fedezhetünk fel egynél nagyobb szórást (1,09; 1,15; 1,28; stb), ezért érdemes megvizsgálni az 5 és 6 klaszteres esetet is.

68. táblázat

Report							
Ward Method		Faktor: fontos a nyaralás	Faktor: természet és mozgás	Faktor: kultúra	Faktor: társaság és kiadás	Faktor: komfort és kényelem	Faktor: élmények
1	Mean	,8772143	,2979500	,5039085	-,5306434	,4932752	-,5239461
	N	57	57	57	57	57	57
	Std. Deviation	,77086327	,64613787	,88349043	1,0009469	,76314787	1,0543040
2	Mean	-,0368256	,0790914	,0522119	,6531306	,6147837	,2683635
	N	102	102	102	102	102	102
	Std. Deviation	,77748603	,80569343	,86721864	,59203796	,62642374	,70041637
3	Mean	-,2304132	-1,1729451	1,4142204	,0000739	-,4912386	,7867606
	N	24	24	24	24	24	24
	Std. Deviation	1,09213862	,98531752	,52582942	,79416440	,93711646	,61906533
4	Mean	-,3472567	-,4869273	-,8307389	-,8062403	,0312610	-,5107060
	N	61	61	61	61	61	61
	Std. Deviation	1,03101502	1,19784974	,89966092	,87782438	,97067424	1,0159949
5	Mean	,0359515	,5770567	-,1370760	-,1234147	-,9618310	,3984597
	N	76	76	76	76	76	76
	Std. Deviation	,89137858	,59510630	,76758988	,73386426	,72918856	,89362038
6	Mean	-1,3096903	-,6502082	-,4056966	1,3050446	-,4613329	-,9129574
	N	17	17	17	17	17	17
	Std. Deviation	,74904556	1,28453232	,71333969	,93663156	,96643476	,92085282
Total	Mean	,0000000	,0000000	,0000000	,0000000	,0000000	,0000000
	N	337	337	337	337	337	337
	Std. Deviation	1,00000000	1,00000000	1,00000000	1,00000000	1,00000000	1,00000000

A 6 klaszteres esetben már jóval kevesebb helyen fedezhető fel 1-nél nagyobb szórás, ugyanakkor a tagszámok némiképp kiegyenlítődték, tehát javult a helyzet. Láthatjuk, hogy a 4-es és 5-ös klaszter összevonásával csökkent a klaszterszám, ezekből jött létre az új 4-es klaszter (elemszám:  $61 = 34 + 27$ ). Ugyanígy megvizsgálhatjuk az 5 klaszteres esetet is, látni fogjuk, hogy a helyzet már nem sokat javul, ráadásul a legkisebb klasztert a legnagyobbval vonja össze, tehát a tagszámok nem egyenlítődnének ki. További vizsgálat alá vonhatjuk az ANOVA és  $\eta$  táblákat, ezek nem meglepő módon azt jelzik, hogy minél több klaszter van, annál szorosabb a kapcsolat. Mindegyik esetben lehet jellemezni a klasztereket, majd mindezek alapján eldönteni, hogy melyik számunkra az ideális megoldás (nincs általános szabály, legtöbbször szuverén kutatói döntés). Mi a 6 klaszteres megoldás mellett döntöttünk, ezt fogjuk a továbbiakban vizsgálni.

## **9.1 A klaszterek jellemzése**

A klasztereket legegyszerűbben a klasztercentroidok elemzésével jellemezhetjük, vagyis a csoportokon belül meg kell határozni a változó átlagát, és ez alapján következtetést levonni a csoportra vonatkozóan. Az átlagok egy már megismert táblában találhatók, jelen esetben ez az 59. táblázat. Ez alapján könnyen készíthetünk egy bővített táblát (mint: Malhotra (2002) 712/20.3. táblázat), amelyben még egyéb változókkal való kapcsolatokat is vizsgálhatunk a klaszterek jobb jellemzésének céljából. A klaszterváltozót (jelen esetben ez a clu6\_1) egyszerű nominális változóként kezelve bevonhatjuk keresztábra elemzésbe, részátlag számításba, illetve egyéb nominális változókra alkalmazható elemzésekbe, és így bővíthetjük a táblát a klasztert leíró egyéb jellemzőkkel (mint például tanulmányi eredmény, nem, átlagos nyaralásköltés, stb.).



69. táblázat

Csoportátlagok	1. klaszter	2. klaszter	3. klaszter	4. klaszter	5. klaszter	6. klaszter
<b>Tagság (db/%)</b>	<b>57</b> <b>16,9%</b>	<b>102</b> <b>30,3%</b>	<b>24</b> <b>7,1%</b>	<b>61</b> <b>18,1%</b>	<b>76</b> <b>22,6%</b>	<b>17</b> <b>5,0%</b>
<b>F1</b> <b>Fontos a nyaralás</b>	0,88	0	-0,23	-0,35	0	-1,31
<b>F2</b> <b>Természet és mozgás</b>	0,30	0	-1,17	-0,49	0,58	-0,65
<b>F3</b> <b>Kultúra</b>	0,50	0	1,41	-0,83	-0,14	-0,41
<b>F4</b> <b>Társaság és kiadás</b>	-0,53	0,65	0	-0,81	-0,12	1,3
<b>F5</b> <b>Komfort</b>	0,49	0,61	-0,49	0	-0,96	-0,46
<b>F6</b> <b>Élmények</b>	-0,52	0,27	0,79	-0,51	0,40	-0,91
<b>Egyéb változók</b>						
<b>Mennyit költött átlagosan a nyaralásra</b>	89 253	56 650	89 023	55 251	55 929	44 343
<b>Hány napig tartott a nyaralás?</b>	10,27	8,91	12,77	8,31	10,29	8,67
<b>Nem</b>	Sok lány	Is-is	Több lány	Több fiú	Több lány	Sok fiú
<b>Tanulmányi átlag</b>	Átlagos	Átlagos	Kissé az átlag alatt	Erősen átlagos	Jóval az átlag felett	Jóval az átlag alatt
<b>Klaszterek elnevezése</b>	Igényes lányok	Társaságkedvelő nagyközép	Kultúrtündérek	Apatikusak	Jó tanuló erdőjárók	Bulizó fiúk

A fenti táblázat alapján már lehet jellemezni a klasztereket, nézzünk is pár példát!

Igényes lányok: imádnak nyaralni menni, ilyenkor nem sajnálják a pénzt, a kultúra és a komfort egyaránt fontos számukra, de nem vágnak különösen társaságra és élményekre, inkább a nyugalmat keresik.

Bulizó fiúk: semmi sem számít számukra a nyaralás folyamán, csak a jó társaság és a buli, inkább rövidebb időre mennek nyaralni, és nem is költenek sokat. Ne a jó tanulók között keressük őket!

## Függelék

70. táblázat

Elemzési módszerek	Input változók		Output változók
Gyakorisági eloszlások	Nem metrikus		
Kereszt tábla	Nem metrikus		
Hipotézisvizsgálat – paraméteres próbák	Metrikus		
Hipotézisvizsgálat – nem paraméteres próbák	Nem metrikus		
Varianciaelemzés	Fgtln: kategorizált	Függő: metrikus	
Kovarianciaelemzés	Fgtln: kategorizált és metrikus	Függő: metrikus	
Korreláció – Pearson-féle	Mindkettő metrikus		
Korreláció – parciális	Mindegyik metrikus		
Korreláció – nem metrikus (Spearman, Kendall)	Ordinális		
Regresszió	Fgtln: metrikus	Függő: metrikus	
Faktorelemzés	Metrikus		standardizált
Klaszterelemzés	Nem metrikus és metrikus		kategorizált