

5. lecke

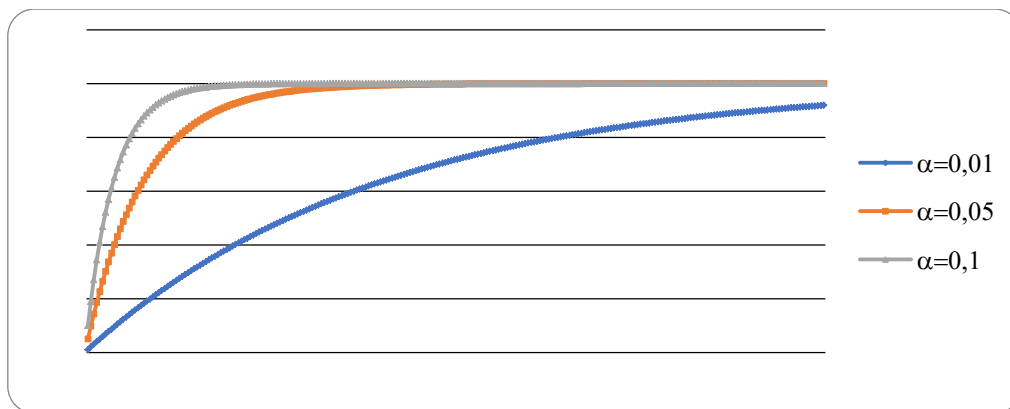
Varianciaanalízis

1. A varianciaanalízis alkalmazási területei

A varianciaanalízis három alkalmazását érdemes megemlíteni:

1. az első alkalmazás több, egyenlő szórású, normális eloszlású csoport/sokaság várható értékének összehasonlítására alkalmas statisztikai módszer, melyet angol elnevezésének kezdőbetűiből adódóan ANOVA-ként (ANalysis Of VAriance) is emlegetnek. Az eljárás a független kétmintás t-próbák általánosításának tekinthető. Kettőnél több minta alapján történő várható értékek összehasonlítására is elvileg működnek a kétmintás próbák, a minták összes lehetséges páronkénti összehasonlításával. Azonban, ez az eljárás nem ajánlott, ugyanis az összehasonlítások számának növekedésével drasztikusan emelkedik az elsőfajú hiba elkövetésének valószínűsége.

Az elsőfajú hiba elkövetésének valószínűsége az összehasonlítások számának függvényében



2. Varianciaanalízis segítségével vizsgálhatjuk egy metrikus függő és kategorikus csoportképző, illetve magyarázóváltozó(k) kapcsolatát. Ezt a kapcsolatot vegyes kapcsolatnak nevezzük.

Már az egyszerűbb jelenségek vizsgálatakor is felmerül, hogy nem csak egy tényező befolyásolja a vizsgált metrikus változónk értékét. Attól függően, hogy hány csoportosító ismérv (hatótényező vagy faktor) hatását vizsgáljuk, beszélünk egy szempontos (egy utas), két szempontos (két utas), illetve több szempontos (több utas) varianciaanalízisről. Például, a jelenlegi fizetések különbözősége magyarázható-e a nemmel (egy utas), illetve a nem és a végzettség (két utas) együttesével. E képzési szinten mi egy utas varianciaanalízis alkalmazásával foglalkozunk.

3. A harmadik alkalmazás segítségével regressziós modellek illeszkedése, illetve a többszörös korrelációs együttható tesztelhető.

Ebben a fejezetben az első és a második alkalmazást taglaljuk, melyek gyakorlatilag csak megfogalmazásban térnek el egymástól, kezelés szempontjából nem. Ugyanis kapcsolatvizsgálat szempontjából a függetlenség (nem szignifikáns kapcsolat) várható értékek összehasonlításának szempontjából úgy is megfogalmazhatóak, hogy a csoportok várható értékei, átlagai között nincs szignifikáns különbség.

Néhány alkalmazási példa:

- Milyen fogyókúra módszer a leghatékonyabb a kilók leadására?

- Befolyásolja-e a keresetek nagyságát a beosztás és/vagy a nem?
- Függ-e egy termék értékesítése a termék elhelyezésétől, csomagolásától?
- Kimutatható-e valamilyen területi és/vagy ágazati egyenlőtlenség a GDP-ben?
- Különböző fajtájú, de azonos növények termésátlagában van-e szignifikáns különbség?
- Egy adott termék átlagára szignifikánsan különbözik-e, az egyes településeken, és/vagy üzletláncoknál.
- Szignifikánsan különbözik-e az amerikai a japán és az európai autók átlagos gyorsulása?

Mit jelent a varianciaanalízis alkalmazása? Először megfogalmazzuk hipotéziseinket. **A próba nullhipotézise szerint a csoportok várható értékei megegyeznek, azaz a csoportosító ismerv nem befolyásolja a metrikus változót. Míg az alternatív hipotézis ennek tagadása. Tehát az alternatív hipotézis nem azt jelenti, hogy mindegyik csoport várható értéke különbözik, hanem csak azt, hogy nem tekinthető mindegyik azonosnak.** Például, ha a nullhipotézisünk szerint a régiók munkanélküliségi rátái azonosnak tekinthetők, akkor ezt a nullhipotézist akkor is elvetjük, ha az összes régió munkanélküliségi rátája különböző, de akkor is, ha csak egy tér el szignifikánsan az összes többitől. Ha ennél több információra van szükségünk, azaz kíváncsiak vagyunk arra, hogy a nullhipotézis miért bukott el, akkor úgynevezett **Post Hoc tesztet** kellene végrehajtanunk.

A varianciaanalízis alkalmazásának két feltétele van. Az egyik a sokaság normális eloszlása, a másik pedig a varianciák egyezősége. Az Excel Adatelemzés modulja nem ellenőrzi le a feltételek teljesülését, így a kapott eredményeket óvatosan kell kezelnünk.

A vizsgálat eredménye az **ANOVA táblázatból** olvasható ki, melynek szerkezete az alábbi. A táblázatban n a minta elemszámát, k a csoportok számát, SSK, SSB, SST, pedig a külső a belső és a teljes eltérés-négyzetösszeget jelenti.

<i>Eltérések (Tényezők)</i>	<i>Eltérés-négyzetösszeg</i>	<i>szabadságfok</i>	<i>becsült szórásnégyzet</i>	<i>F</i>	<i>p-érték</i>
Külső (Csoportok közötti eltérés)	SSK	$k-1$	$SSK/(k-1)$	$(SSK/(k-1)) / (SSB/(n-k))$	
Belső (Csoporton belül)	SSB	$n-k$	$SSB/(n-k)$		
Összesen	SST	$n-1$			

Magyarázó modellek esetén az ANOVA táblázatból kiolvasható a modell magyarázó ereje, mely nem más, mint az SSK/SST varianciarányados, mely megmutatja, hogy a metrikus változó értékeinek különbözőségét hány százalékban magyarázhatjuk a csoportképző változókkal. Az SSB/SST varianciarányad pedig a modell által nem magyarázott (tehát más figyelembe nem vett tényezőkkel és a véletlennel magyarázható) rész.

2. Döntés szoftverrel

2.1. Az SPSS alkalmazása

A vegyes kapcsolat esetén vizsgálhatjuk a kapcsolat erősségét, illetve a magyarázóerőt. A kapcsolatmérő mutató a H mutató, amely az SPSS-ben az Eta nevet viseli, a magyarázóerő pedig a H^2 mutató, amely az SPSS-ben az Eta Squared nevet viseli. Ezen mutatókat az **Analyze/Compare Means/Means** menüben tudjuk kiírni, azon belül is az **Options** menüpontban kiválasztva az **Anova table and eta** opciót.

A H mutató 0 és 1 közötti értéket vehet fel, minél közelebb van a nullához, annál gyengébb, és minél közelebb van az egyhez, annál erősebb kapcsolatról beszélhetünk a két változó között.

A H^2 mutató a kimeneten 0 és 1 közötti értéket vesz fel, ezt beszorozva 100-zal viszont megkapjuk a magyarázóerőt, amely azt mutatja meg, hogy a metrikus változó különbözőségeit hány százalékban magyarázhatja a kategorikus változó, vagyis **azt mutatja meg, hogy a csoportképző (kategorialis) ismérv a mennyiségi ismérv szórásnégyzetének mekkora részét (százalékát) magyarázza meg.** Kiszámítása:

$$H^2 = \frac{SST - SSB}{SST} = \frac{SSK}{SST}$$

A varianciaanalízis (éés a nullhipotézis bukása esetén a Post Hoc teszt) az SPSS **Analyze/Compare Means/One-way ANOVA** menüpontjából érhető el. A vizsgálat egy F-próbát hajt végre, melynek képlete az alábbi:

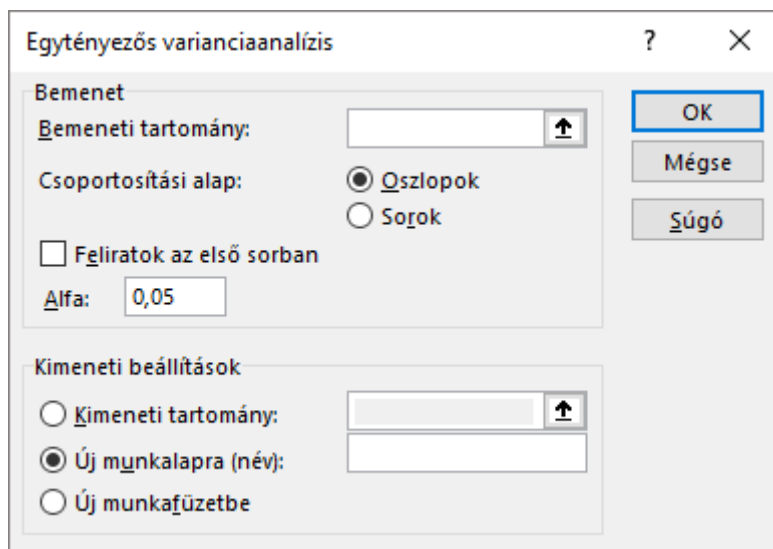
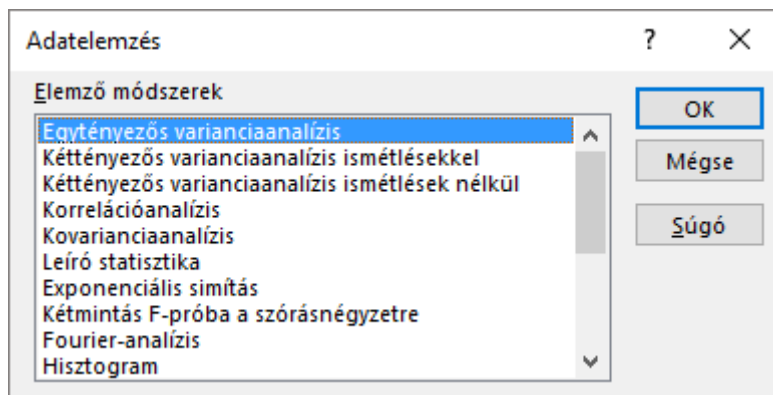
$$F = \frac{SSK / (k - 1)}{SSB / (n - k)} = \frac{s_K^2}{s_B^2}$$

Az F próbafüggvény alkalmazási feltételei:

1. A metrikus változó normális eloszlású minden (k db) csoporton belül, azaz minden egyes csoportban
 - ♦ vagy nagy elemszámú mintánk van,
 - ♦ vagy 30 alatti elemszámú mintánk van és normális eloszlású a változó a csoporton belül (a normális eloszlás vizsgálatát az **Analyze/Descriptive Statistics/Explore** menüben tudjuk megtenni → **Plots: Normality plots with tests**)
 - ♦ vagy 30-100 közötti elemszámú mintánk van és a csoporton belül az aszimmetria mutató értéke +1 alatti (ezt szintén az **Explore** menüben tudjuk vizsgálni)
2. Varianciahomogenitás. (Ennek ellenőrzése itt is Levene tesztel történik) (Szórásazonosság)
 - ♦ Ha ez teljesült: ANOVA
 - ♦ Ha nem teljesül Welch-tesztet alkalmazunk. Ez egy korrigált F-próba (=módosul a próbafüggvény)

2.2. Az Excel alkalmazása

Az egy utas varianciaanalízis bemenete egy egyszerű bemenet, melyben az adatok forrását kell megadnunk.



Az elemzés kimenete két táblázatból áll.

Az *Összesítés* táblázatban láthatjuk az egyes csoportok, változók elemszámát, átlagát, szórásnégyzetét. Sajnos, az Excel nem ellenőrzi a varianciaanalízis alkalmazásának feltételeit.

Egytényezős varianciaanalízis

ÖSSZESÍTÉS

Csoportok	Darabszám	Összeg	Átlag	Variancia
A	30	897	29,9	0,851724138
B	32	952	29,75	0,838709677
C	30	877	29,23333333	1,012643678

D	31	929	29,96774194	0,832258065
---	----	-----	-------------	-------------

VARIANCIAANALÍZIS

<i>Tényezők</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p-érték</i>
Csoportok között	10,0062418	3	3,335413935	3,78	0,0124
Csoporton belül	105,0344086	119	0,882642089		
Összesen	115,0406504	122			

A második táblázat nem más, mint az ANOVA táblázat. Mivel a $p\text{-érték} < 0,05$; ezért a nullhipotézist ötszázalékos szignifikanciaszint mellett elvetjük. Ezek szerint a csoportok várható értékei nem tekinthetők azonosnak. Más megfogalmazásban ez azt jelenti, hogy a csoportosítás befolyásolja a mennyiségi változót.

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

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

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

SZÉCHENYI 2020



MAGYARORSZÁG
KORMÁNYA

Európai Unió
Európai Szociális
Alap



BEFEKTETÉS A JÖVŐBE