

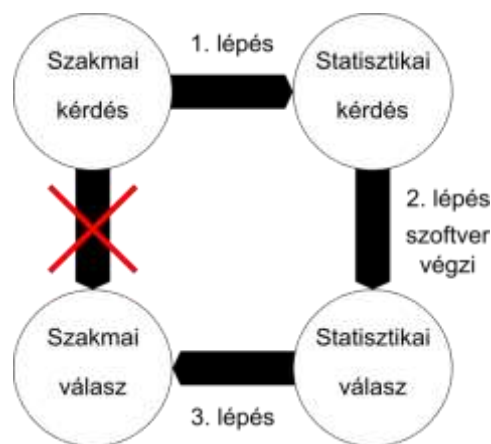
## Segítség ANOVA feladatok megoldásához

Ennek a dokumentumnak a célja az, hogy segítsen a feladat megoldójának beleélni magát abba helyzetbe, amibe bele kell magát képzelnie a megoldás során. Továbbá segítséget kívánok nyújtani a szoftver kezelésének olyan részleteibe is, ami a „Statistica-könyv” jelenlegi kiadásából hiányzik.

### Statisztikai problémák megoldása

A most következő rövid elmélkedés elhangzott az előadáson is, de mivel nem mindenki jár előadásra, így itt is megemlítem, mert rövid, és mert ennek ismerete nélkül az olvasónak nehezebbé eshet a megoldás menetének követése.

A feladat minden esetben egy szakmai kérdéssel kezdődik (vagy kérdésekkel, egyszerűség kedvéért a továbbiakban egyes számot használok). Erre szeretnénk megadni a választ, ami a szakmai válasz. Amint a szakmai válasz megvan, készen vagyunk, nem kell tovább dolgozni. Azonban a legtöbbször abba a problémába ütközünk, hogy a szakmai kérdésre nem tudjuk megadni a választ közvetlenül (a való életben általában, most a feladat megoldása során biztosan nem), emiatt a szakmai kérdésnek megfelelő statisztikai problémát kell megfogalmaznunk. A statisztikai probléma megfogalmazása után ezt kell megoldanunk, ennek a megoldása a statisztikai válasz. A statisztikai válasz megfogalmazása után tudjuk ezt átfogalmazni szakmai válasszá. E három lépés egymástól teljesen eltérő jellegű, és a megfelelő megoldáshoz mindháromat jól kell megcsinálni. Azonban észre kell vennünk azt, hogy az alapozó tárgyak során a harmadik lépést (statisztikai válaszból szakmai válasz) nagyon körüljártuk már, így most erre nem fogunk fókuszálni.



1. ábra: A statisztikai feladatok általános sémája

A második lépéssel (statisztikai kérdésre statisztikai válasz) is sokat foglalkoztunk az alapozó tárgyak során, ráadásul ez az, amit a szoftver végez. Mivel a feladat megoldása során először valószínűleg azzal a problémával szembesülünk, hogy hogyan is kellene a szoftvert helyesen használni, így könnyen az a tévképzetünk lehet, hogy a feladat arról szól, hogy azt kívánjuk megtanítani/ellenőrizni, hogy ezt hogy kell. Azonban az az igazság, hogy bár ez is fontos eleme a megoldásnak, de ez egyrészt a szoftverkörnyezettől is függ (ugye nem a STATISTICA szoftvert, hanem a statisztikát mint tudományt tanítjuk), másrészt pedig az egyéni igényektől függően több-kevesebb gyakorlással, illetve a *Help* menü helyes használatával ez elsajátítható.

Ezek alapján egyértelművé kellett váljon, hogy a feladat lényegi része az első lépés (szakmai kérdésből statisztikai kérdés). Ezt általában az emberek a legbátorteljebb résznek tartják, annak ellenére, hogy messze a legfontosabb. Ez leggyakrabban valamilyen

hipotézisvizsgálatról szól, de előfordulhatnak más jellegű problémák (konfidencia intervallum megadása, pontbecslések stb.)

Elsőként a hipotézisvizsgálattal foglalkozunk. A hipotézisvizsgálat első lépése a nullhipotézis megfogalmazása. Itt több fontos szabályt figyelembe kell venni. Az egyik az, hogy az egyenlőségjelnek a nullhipotézisben kell lennie. Ez azért kell, mert a próbastatisztikát úgy kell megalkotni, hogy feltételezzük, hogy a nullhipotézis egyenlőség része igaz (ez ugye kényelmetlen lenne, ha nem lenne az egyenlőség a nullhipotézisben). A másik dolog az, hogy az ellenhipotézisnek a nullhipotézis ellenkezőjét kell megfogalmaznia. Bár mivel a nullhipotézisből az ellenhipotézis egyértelműen következik, így ezt nem is feltétlenül kell megfogalmazni. A harmadik dolog az, hogy a nullhipotézisnek igazodnia kell a szakmai kérdéshez. Ez nagyon magától értetődőnek tűnhet, mégis sokszor hibát vétenek ebben a hallgatók. A legfontosabb dolog, hogy a szakmai kérdés sosem a kapott adatokra (mintára) vonatkozik. (Például a szakmai kérdés sosem hangzik úgy, hogy az adatok átlaga valamilyen  $x$  érték-e.) Ezt mindenféle hipotézisvizsgálat nélkül el tudnánk dönteni, viszont nem lenne túl informatív. A szakmai kérdés mindig az adatok mögött megbújó magasabb szintű összefüggésre kérdez rá, így a nullhipotézisnek is mindig az adatok mögött megbújó sokaság(ok) paraméterére vagy valamilyen – a vizsgált jelenséget leíró – modell paraméterére kell vonatkoznia!

Ezen ismereteket jó mindig a gondolatainkban tartani, amikor a feladatot (bármilyen statisztikai jellegű problémát) megpróbáljuk megoldani.

Hogy (és hogy ne) használjuk a STATISTICA-könyvet?

A könyv kifejezetten arról szól, hogy bemutatja a szoftver egyes funkcióit különböző példákon keresztül, tehát inkább törekszik arra, hogy az összes elérhető funkciót bemutassa, mint hogy egy logikai íven keresztül bemutassa egy feladat megoldásának menetét. Ráadásul a szoftver használatához nem elengedhetetlen elemekre kis mértékben vagy egyáltalán nem tér ki.

Tehát a megoldás során

- NE próbáljunk szövegrészeket, táblázatokat egy az egyben a könyvből kimásolni (akkor sem, ha az adatokat átírjuk a sajátunkra),
- NE illesszünk be táblázatot/elemzést csak azért, mert a könyvben van olyan,
- NE változtassunk meg beállításokat a programban csak azért, hogy a könyvben szereplő elemzéseket el tudjuk végezni!

A könyvben szereplő példákat érdemes lehet végigkövetni, akár a saját adatainkkal is végigvinni az elemzéseket, de nem tanácsos megkeresni azt, hogy a mi feladatunk melyik könyvbeli példához hasonlít a legjobban, és gépiesen lemásolni a lépéseket a saját feladatunkra. És sose feledjük: Elképzelhető, hogy a kapott feladatunk nem illeszkedik a könyvben található egyetlen sémához sem! (Például a könyvben mindig egy vagy két faktor szerinti ANOVA-t végzünk, de képes a szoftver többet is kezelni.)

## Általános "ezt ne csináld" szabályok

Alapvető oktatási alapelv, hogy nem írunk le olyat, ami rossz, de most itt olyan dolgok következnek, amiket erősen ajánlott nem elkövetni. (Csak hogy egyértelmű legyen, hogy ezek hibásak, minden mondat NE-vel kezdődik.)

- NE illesszük be a teljes adatfájlt a megoldásba! (Főleg, ha több oldalnyi nagyságú.)
- NE hivatkozzunk olyan betűkre (modellparaméter, adat, sokaság paraméter), amiket nem vezettünk be!
- NE rakjunk a megoldásba olyan táblázatot vagy ábrát, amiről nem vonunk le következtetést! (Az nem következtetés, hogy kimásoljuk a könyvből azt, hogy mit látunk az ábrán.)
- NE másoljuk a megoldásba ennek az írásnak semmilyen részét!

## Gyakorlati, technikai jó tanácsok

Ne adjunk a fájlnek jó hosszú, komplikált nevet! A beküldött fájlok hallgatónként, és feladatonként külön-külön mappákban lesznek tárolva, tehát a fájlnev legfontosabb feladata az, hogy a megoldások különböző verzióit megkülönböztethesse a javító egymástól (ez jellemzően a 30-40 karakteres fájlnevek utolsó karaktereiből szokott látszani). Ez persze nem azt jelenti, hogy ne írjuk bele a nevünket a fájl nevébe. Amikor a javaslatok alapján javított változatot készítünk, akkor célszerű megőrizni az eredeti változatot is, és a javított változatot más néven elmenteni, és elküldeni. A javítások készítésére több, különféle technikát követhetünk. Amennyiben rájövünk, hogy a feladat egy részét teljesen másképpen kellene megoldani, akkor természetesen nincs sok választásunk, akkor ezt teljesen át kell írunk. Nem célszerű olyan elemeket meghagyni a dolgozatban, amikről tudjuk, hogy helytelenek. (Nem összekeverendő ez azzal, hogy meghagyjuk a teljes korábbi változatot egy külön fájlban.)

Azonban ha csak kis változtatásokat, kiegészítéseket végzünk, akkor több lehetőségünk is van. Az egyik, hogy a teljes dolgozatot újra elküldjük, és a megfelelő helyen kiegészítéseket teszünk. Ilyenkor tanácsos ezeket más színnel vagy háttérkitöltéssel jelölni, főleg, ha mondjuk egy 10-15 oldalas megoldásban mondjuk összesen 2-3 mondatot változtattunk.

A másik lehetőség, hogy csak a változtatásokat, kiegészítéseket írjuk le a javításban, ilyenkor lehetőleg vagy kérdés-válasz formában vagy célszerűen megjelölve, hogy az eredeti dokumentum mely részére vonatkozik a javítás. (Például: A 3. oldalon a nullhipotézis helyesen: Az A faktorhoz tartozó  $\alpha_i$  értékek mind 0-k.)

Gyakran fordul elő olyan, hogy valaki az első lehetőséget választja (a teljes javított dokumentumot küldi el újra), viszont nem jelöli a javítások helyét, ez a javító számára kényelmetlenséget okozhat. Illetve az is előfordul, hogy a második lehetőséget választja, de a javításokat beszúrja az eredeti dokumentum végére. Ez utóbbi nagyon helyes oly tekintetben, hogy megőrzi az eredeti változatot, de mi minden esetben megőrizzük az eredeti változatot külön fájlban, így ez szükségtelen. Másrészt így megmarad a helytelen információ a megoldásban, és ha mondjuk a javítónak nem mondjuk meg, hogy a „javítások a dokumentum végén vannak összegyűjtve”, ez újabb kellemetlenségeket okoz a javítónak.

Utolsóként javasolom mindenkinek, hogy a megoldását .pdf kiterjesztésű dokumentumban küldje el. Ugyanis a Word programok sokféle változata miatt előfordul, hogy a javító számára teljesen más formában jelenik meg a megoldás. Ez jobb esetben csak esztétikai problémát okoz, rosszabb esetben a megoldásból fontos információk tűnhetnek el. (Igen, ez utóbbi is előfordult már.) A .pdf-ek viszont minden környezetben ugyanúgy jelennek meg (továbbá jellemzően kisebb méretűek is, de ez manapság már kevésbé probléma).

## Az ANOVA feladatok megoldásának logikai sorrendje

A korábbiak alapján az első lépés a statisztikai probléma megfogalmazása. Az ANOVA feladatok esetén kevés kivételtől eltekintve a nullhipotézis valamilyen modellparaméterre vonatkozik. A szakmai kérdés az szokott lenni, hogy az egyes faktoroknak van-e hatása, vagy ezzel analóg kérdés. Ebből következik, hogy a nullhipotézist NEM fogalmazhatjuk meg úgy, hogy a faktornak nincs hatása, mert ez a szakmai kérdés. Tehát a nullhipotézisek értelmes felírásához először a modellt kell megalkotnunk.

A modellbe a feladatban szereplő összes faktort be kell venni, előfordulhat, hogy 2-nél több faktor van, nincs ebben semmi különleges. Ha meghatároztuk a faktorokat, akkor meg kell határozni ezek egymáshoz való viszonyát. Leegyszerűsítve két faktor vagy „keresztben” van egymással, vagy egymásba vannak ágyazva. Ezt minden faktor-párra vizsgálva kiegészítjük a modellt (a beágyazott faktorokat megfelelően jelöljük, a keresztben levők kölcsönhatását bevesszük a modellbe).

Így megalkottuk a legbővebb, „teljes” modellt, ami leírja a kísérleteinket. A modell a kísérleti tervet írja le, azaz a megoldás során sosem kell újabb modellt alkotni addig, amíg ugyanezeket az adatokat elemezzük (vigyázat, ez nem keverendő össze a modell redukciójával, ami kicsit később jön!).

További fontos elem, hogy minden főhatásról el kell döntenünk, hogy **rögzített** (fix) vagy **véletlen** (random). Ennek eldöntéséhez nem matematikai módszereket kell használni. Egy faktor akkor véletlen, ha „a kísérletek teljes újbóli elvégzése esetén a faktor szintjeit nem tudnánk, vagy nem akarnánk ugyanúgy beállítani”. Ennek megfelelően akkor rögzített, ha **tudjuk és akarjuk** is ugyanúgy beállítani. A hangsúly itt az „akarjuk”-on van. Azt általában könnyen meg tudjuk állapítani, ha egy faktor szintjeit nem tudjuk újra ugyanarra beállítani, ráadásul az órai példák is jellemzően ilyenek voltak, így sokszor megfélekedünk erről a másik kitéletről. Ne tegyük!

Mindenképp érdemes megemlíteni itt, hogy az, hogy akarjuk-e a faktor szintjeit ugyanúgy beállítani, csak a feladat szövegéből derül ki, és nem minden esetben egyértelműen. Tehát előfordulhat, hogy a feladatot többféleképpen lehet értelmezni. Ha a megoldás során úgy érezzük, hogy ilyen helyzettel talákoztunk, akkor mindenképpen írjuk bele a megoldásba, hogy döntöttünk, a faktort rögzítettnek vagy véletlennek vettük, és hogy miért!

A kölcsönhatások rögzített/véletlenségének eldöntéséhez már nem kell sokat elmélkedni. Ugyanis egy kölcsönhatás csak akkor rögzített, ha minden tagja rögzített, minden más esetben véletlen, vagy vegyes (mixed), de ez utóbbi igazából minden tekintetben úgy viselkedik, mint a véletlen, így ennek a megkülönböztetésnek nincs gyakorlati jelentősége.

Az, hogy a faktor rögzített vagy véletlen, a nullhipotézisben jelenik meg. Ha a faktor rögzített, akkor a nullhipotézis a következő formát veszi fel:

$$H_0 : \alpha_i = 0$$

Ha a faktor véletlen, akkor előbb egy feltételezést kell tennünk, hogy a modellben szereplő koefficiens egy normális eloszlású valószínűségi változó  $\sigma_A^2$  varianciával:

$$\alpha_i \sim N(0, \sigma_A^2)$$

És ez után a nullhipotézis már erre a  $\sigma_A^2$ -re vonatkozik (mindegy hogy a szigma négyzetre vagy a szigmára nézzük):

$$H_0 : \sigma_A^2 = 0$$

A szoftver úgy van megalkotva, hogy a modellben szereplő mindegyik faktorra és a becsült függvény várható értékére a vonatkoztatási pontban\* (ezt tengelymetszetnek /intercept/ nevezik a szoftverben) statisztikai próbát végez.

Az intercept-re azt a nullhipotézist vizsgálja, hogy ez 0-e. Ez sokszor az adatok természetéből fakadóan érdektelen és értelmetlen kérdés. (Például ha emberek vérnyomását vizsgáljuk, akkor itt az a nullhipotézis, hogy az összes embert figyelembe véve a várható vérnyomás 0.) Ezt leszámítva viszont érdemes az összes többi nullhipotézist leírni a megoldásban, hogy a próbák értelmezhetőek legyenek. Amennyiben nem tesszük meg, akkor mindenképpen le kell írunk, hogy az általunk felírt nullhipotéziseket az ANOVA tábla mely sorában szereplő statisztikai próba vizsgálja.

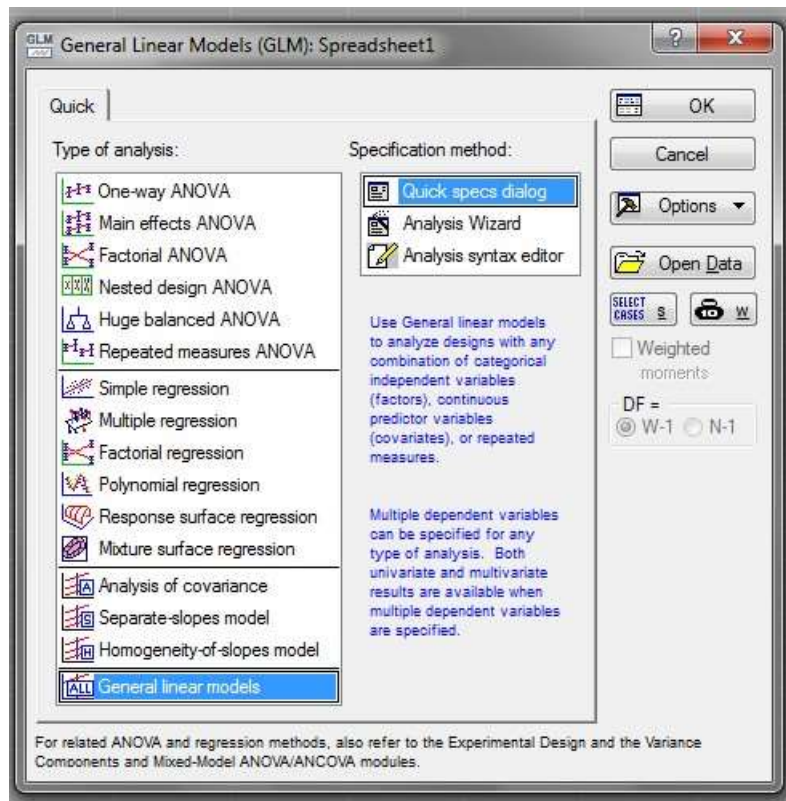
Ezzel befejeztük a feladat első lépését, egyszerűbb modellek esetén az ez utáni lépések már igen egyszerűek és mechanisztikusak. Azonban előfordulhat, hogy egy bonyolultabb modellbe ütközünk.

Hogy kezeljük a bonyolultabb modelleket a STATISTICA-ban?

A probléma úgy merül fel, hogy a szoftverben ANOVA vizsgálatot választva vagy a *factorial ANOVA* menüpontot választjuk (azaz a keresztosztályozást) vagy a *nested design*-t (azaz a hierarchikusat). De mi van akkor, ha mi olyan modellt írtunk fel, amiben van hierarchikus és kereszt-osztályozás szerinti szerkezet is? Emlékeztetőül: a modell a kísérleti terv szerkezetétől függ, és nem választás kérdése! Ekkor a legáltalánosabb GLM modult kell használnunk. (Ebben a modulban lehetőségünk van folytonos faktorok megadására is, amit most nem fogunk használni.)

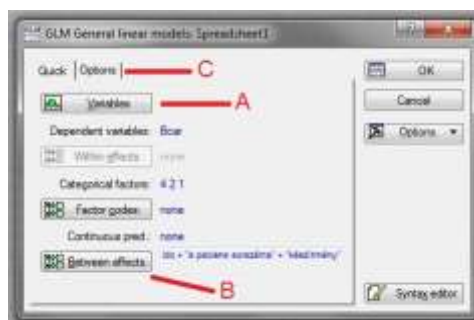
A modul előhívásához a *Statistics* menü *Advanced Linear/Nonlinear Models*, ezen belül *General Linear Models* pontját válasszuk! Az itt megjelenő ablakban válasszuk a legalsó lehetőséget (ennek szintén *General linear models* a neve)! (2. ábra)

\*Ez lehet a globális átlag (sum to zero vonatkoztatás esetén) vagy az utolsó szintbeli függvényérték (set to zero – overparameterized vonatkoztatás esetén).

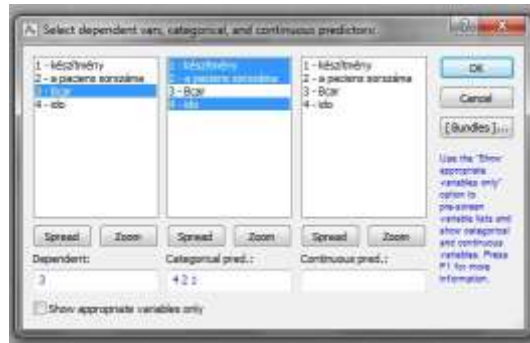


2. ábra: A *General Linear Models* ablakja.

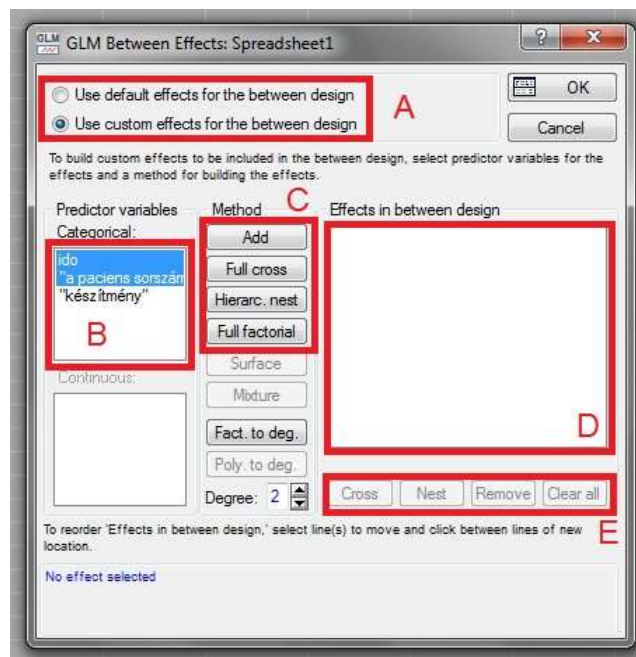
A megjelenő ablakban a szokásos módon a *Variables* gombra (3. ábra, A) kattintva tudjuk a megfelelő változókat kiválasztani (4. ábra). Három mezőben tudjuk a változókat kiválasztani. A bal oldaliban kell a függő változót, a középsőben pedig a faktorokat, a jobb oldaliban választjuk ki a folytonos független változót, erre akkor van szükségünk, ha kovarianciaanalízist szeretnénk végezni.



3. ábra: A GLM General linear models főablaka



4. ábra: Ebben az ablakban jelöljük ki a függő változót, illetve a faktorokat. Miután kiválasztottuk a változókat, a főablakban a *Between effects* gombra (3. ábra, B) kattintva tudjuk megadni a modell alakját.



5. ábra: Ebben az ablakban tudjuk a modell pontos alakját megadni.

A megjelenő ablakban először a legtetején (5. ábra, A) kell a „*Use default effects for the between design*”-ről (ez azt jelenti, hogy csak a főhatásokat vizsgáljuk) a „*Use custom effects for the between design*”-ra (ez azt jelenti, hogy mi adjuk meg a modellt) váltani az ablakot. Ezzel aktivizálódik az ablak további része, ahol a modellt tudjuk megadni.

Az ablak bal oldali mezőjében (5. ábra, B) szerepelnek azok a változók, amiket korábban kategorikus faktoroknak adtunk meg. Ha ezek közül (legalább) kettőt kijelölünk, akkor aktivizálódnak a középen lévő gombok (5. ábra, C). Ezek közül csak a most használatosak funkcióját írrom le.

Az *Add* gombbal egyszerűen hozzáadjuk a megadott főhatásokat a modellhez (ez az egyetlen gomb, ami úgy is aktivizálódik, ha csak egy faktort jelölünk ki).

A *Full cross* gomb a megadott faktorok kölcsönhatását adja hozzá a modellhez (ha két faktort jelöltünk ki, akkor a kettős kölcsönhatást, ha hármat, akkor csak a hármas kölcsönhatást stb.).

A *Hierarc. nest* gombbal tudunk hierarchikus elrendezést megadni. Ha ezt megnyomjuk, akkor egy újabb ablak jelenik meg (6. ábra), ahol megadhatjuk, hogy melyik faktor melyikbe van ágyazva. Ha egy faktort egy másikba ágyazunk, a program magától nem gondoskodik az amúgy logikus további beágyazásokról. Például az injekció beadását a mintába ágyazzuk, a mintát a laborba, akkor külön meg kell adni, hogy az injekció beadását a laborba is beágyazzuk.



A *Full factorial* gombbal a kiválasztott faktorok összes lehetséges fő- és kölcsönhatását bevesszük a modellbe. (Két faktor esetén két főhatást, és a kölcsönhatást; három faktor esetén három főhatást, három kettes kölcsönhatást, és egy hármas kölcsönhatást; stb.)



6. ábra: Ebben az ablakban adhatjuk meg, hogy a faktorok közül melyik van melyikbe beágyazva

A modellbe bevett hatások a jobb oldali mezőben jelennek meg (5. ábra, D). Ez alatt négy gomb látható (5. ábra, E). Ezek közül a 4. „Clear all” gombbal mindent kitörölhetünk a mezőből, a 3. „Remove” gombbal pedig a kijelölt hatásokat törölhetjük.

Amennyiben több mint 1 hatást jelölünk ki a jobb oldali mezőben, akkor az első kettő gomb is aktivizálódik. A „Cross” feliratúval a kijelölt hatások kölcsönhatását hozhatjuk létre, a „Nest” feliratúval pedig egymásba ágyazhatjuk a hatásokat. Ezekkel el lehet végezni azokat a műveleteket, amiket a korábban említett gombokkal, de igazán akkor hasznosak, ha mondjuk egy kölcsönhatást kell beágyazni egy faktorba, vagy egy beágyazott faktor kölcsönhatását akarjuk venni egy másik faktoral, ugyanis ezt csak így lehet megadni.

Amennyiben az összes hatást felvettük a modellbe, OK nyomással visszajutunk az előző ablakhoz. Itt az *Options* fülön (3. ábra, C) tudjuk megadni a véletlen faktorokat. Ezután újabb OK nyomással eljuthatunk a terv kiértékeléséhez, amit már a hagyományos módon tudunk végrehajtani.

Tanácsos a modell kiértékelése során figyelni az ANOVA táblában a nevezőket és szabadsági fokokat, ugyanis ebből lehet a legkönnyebben észrevenni, ha mégsem azt a modellt sikerült beadni a szoftvernek, amit szerettünk volna.

Ha egy hatás szignifikáns, a becsült nagysága is érdekes, ez a szoftverben a *Summary* gombra kattintva kapható meg.

A feladat továbbgondolása, tegyünk fel kérdéseket magunknak!

Korábban volt róla szó, hogy amennyiben a megfogalmazott szakmai kérdést megválaszoltuk, akkor készen vagyunk. Valójában viszont a legtöbbször a kérdésre adott válaszok újabb kérdéseket generálnak, amiknek a megválaszolásához a legtöbbször újabb kísérletek elvégzése szükséges. Mivel ez a feladat kereti között nem lehetséges, ezért most kizárólag arra kell szorítkoznunk, hogy milyen további információkat tudunk kinyerni az adatokból.

Az első lehetőség a használt próbák erejének növelése. A próba ereje a szabadsági fokoktól függ, jelen esetben F-próbákról van szó, itt a nevező szabadsági foka az érdekes (ezt befolyásolhatjuk). A nevező szabadsági fokát úgy tudjuk növelni, hogy a nevezőben lévő tagba plusz szabadsági fokokat vonunk be azzal, hogy egy nem szignifikánsnak talált tagot elhagyunk a modellből. (A leggyakoribb esetben így a maradék szórásnégyzet szabadsági fokát lehet növelni, de ugyeszen meg lehet ezt csinálni másik tagokkal is. Mindig figyeljük a szabadsági fokokat az ANOVA táblában!) Ezt a műveletet hívják a modell redukálásának. A redukált modellre kapott eredményeket mindig értelmezni kell, illetve célszerű azt is megindokolni, hogy miért gondoltuk azt, hogy célszerű a modellt redukálni.

Rögzített faktorok esetén a nullhipotézis az egyes szintek „hozzáadott értékéről” szól. Amennyiben elutasítottuk a nullhipotézist, akkor azt találtuk, hogy nem minden szinten azonos ez az érték. Azonban ha több, mint 2 szintje van a faktornak, akkor értelmes feltenni



azt a kérdést, hogy mindegyik különböző-e, vagy vannak egyformák. Ezt a szoftverben a *Post-hoc* fülön lehet megoldani, és a könyvben kellő részletességgel van leírva a módja. Ugyancsak érdekes lehet a szignifikáns hatások nagyságának becslése. Itt vigyázni kell arra, hogy a becslt paraméterek értelmezése függ a megkötéstől (sum to zero = sigma restricted vagy set to zero = overparameterized)

Véletlen faktorok esetén a nullhipotézis a faktor szintjének véletlenszerű beállításából származó többletingadozásról szól, amit egy varianciával jellemzünk. A nullhipotézisben arra kérdezzük rá, hogy ez a variancia 0-e. Amennyiben azt találjuk, hogy nem, akkor érdekes lehet a nagyságának becslése. (A szoftverben a *Summary* gombra kattintva választható *Variance Components* rész végzi el a számítást. A könyvben ez is megfelelő részletességgel van leírva.) Az errorhoz tartozó varianciához nem tartozik nullhipotézis, erről minden esetben értelmes dolog megkérdezni, hogy mennyi. Első körben a variancia pontbecslését kaphatjuk meg, de adott esetben intervallumbecslést is adhatunk (ez a szoftverben csak a VEPAC modulban érhető el, amit sajnos a BME-s hallgatók rendelkezésére bocsátott verzió nem tartalmaz). Ne feledjük, hogy a szórásnégyzet khí-négyzet (és nem normális!) eloszlású!

## Kovarianciaanalízis

Kovarianciaanalízisnek hívjuk azokat az elemzéseket, amelyekben a kategorikus faktorok mellett folytonos független változók (ún. kovariánsok) is szerepelnek. A kovariánsok szerepeltetése jelentősen megnöveli az elképzelhető modellek számát. Általános esetben elsőként egy bonyolult modellből indulunk ki, majd az elemzés eredménye adatok alapján egyszerűsítjük le a modellt azzal, hogy elhagyjuk belőle a nem szignifikáns hatásokat.

Mivel a kovarianciaanalízis témáját időhiány miatt csak érintőlegesen vettük, így ezt nem is várjuk el a megoldásban. Ha a feladat megoldása során kovarianciaanalízist kell végezni, akkor a STATISTICAban a *Statistics* menü *Advanced Linear/Nonlinear Models*, ezen belül *General Linear Models* menüpontját válasszuk, az itt megjelenő ablakban (2. ábra) pedig az *Analysis of Covariance* menüpontot! Itt a *Help* menüpontban leírtakkal ellentétben alapértelmezésként a faktoroknak és a kovariánsoknak a főhatását teszi a modellbe a szoftver. Ez most számunkra megfelelő, így ennél a beállításnál maradhatunk.

Kovarianciaanalízis során is kapunk ANOVA táblázatot. Ebben a kovariánsokhoz is tartozik egy sor. Itt eltérés-négyzetösszeget úgy kell számítani, hogy minden pont esetében az egyenesillesztéssel becslt értékből vonjuk ki az átlagos  $x$  értékre becslt értéket, ezt emeljük négyzetre, és adjuk őket össze. Ha a kovariánsokhoz tartozó egyenes meredekségére vagyunk kíváncsiak, akkor azt a *Coefficients* gombra kattintva kapjuk meg. Arra érdemes odafigyelni viszont, hogy az így kapott táblázatban szereplő statisztikai próbák egymástól nem függetlenek, így nem alkalmasak az egyes hatások szignifikanciájának kiértékelésére!

Ezek alapján kitalálható, hogy ANOVA táblázatot úgy is konstruálhatunk, ha nincs is kategorikus faktor a modellünkben (regressziós feladatok). Természetesen ez ilyenkor is alkalmas az egyes faktorok szignifikanciájának kiértékelésére.

Kunovszki Péter