

**Szegedi Tudományegyetem
Gazdaságtudományi Kar
Közgazdaságtudományi Doktori Iskola**

**A multikollinearitás vizsgálata lineáris regressziós modellekben
A PETRES-féle *Red*-mutató vizsgálata**

Doktori értekezés tézisei

Készítette:
Kovács Péter

Témavezető:
Dr. Petres Tibor
egyetemi docens

Szeged
2008

I. A kutatás problémájának definiálása, céljai, hipotézisei

I. 1. A probléma megfogalmazása

A mai globalizálódó világban nagymértékben növekszik a döntéshozók információigénye. Az adatok mennyiségének nagymértékű növekedése nem jár együtt automatikusan a megfelelő mértékű információnövekedéssel. Igazából a döntéshozóknak ma már nem az adatok hiányával, hanem azok bőségével kell szembenéznük, viszont, a nagyon nagy mennyiségű adattömeg gyakran kevés információt hordoz, azaz nagymértékű a **redundancia**. Ez utóbbi alatt a vizsgálat szempontjából újabb információt, érdemleges közlést már nem tartalmazó, „felesleges” adatokat értjük. Éppen ezért, empirikus elemzéseknél **lényeges kérdés a metrikus adatok információtartalma**, Különösen igaz ez a lineáris regressziós modellek alkalmazásakor. **A multikollinearitás lineáris regressziós modellek esetén a redundancia egy fajtájaként értelmezhető**. Ez mátrixalgebrai jelöléssel az $\tilde{\mathbf{y}} = \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\varepsilon}}$ formában is felírható, ahol $\tilde{\mathbf{y}}$ az eredményváltozó n komponensű oszlopvektora; $\tilde{\mathbf{X}}$ a tényezőváltozók n sorból és $(m+1)$ oszlopból álló mátrixa, ahol az első oszlop mindig egy $\tilde{\mathbf{x}}_0$ összegező vektor; $\tilde{\boldsymbol{\beta}}$ a modell számunkra ismeretlen paramétereinek $(m+1)$ komponensű oszlopvektora; m a magyarázóváltozók (tényezőváltozók) száma; $\tilde{\boldsymbol{\varepsilon}}$ a hibatag n komponensű oszlopvektora.

A multikollinearitás fogalma a szakirodalomban látszólag egységes. Az egyes definíciók általában egy-egy szóban térnek el egymástól, de ez jelentős tartalmi változást jelent.

A multikollinearitást, mint kifejezést RAGNAR FRISCH használta először. Olyan esetek leírására használta, amikor egy változó több összefüggésben szerepel. Ezekben a vizsgálataiban nem különböztette meg az eredményváltozót a magyarázóváltozóktól. Feltételezése szerint, minden változó mérése hibás, ez alapján kell megbecsülni a változók tényleges értékei közötti korrelációt.

Nagyon felületes meghatározás az, hogy a multikollinearitás a tényezőváltozók függetlenségének a hiánya. Ezzel a meghatározással az a probléma, hogy nem derül ki egyértelműen, hogy mit értünk a magyarázóváltozók függetlensége alatt. Netán ezek lineáris függetlenségét, esetleg statisztikai értelemben vett függetlenségét?

A standard lineáris regressziós modell egyik alapfeltétele a magyarázóváltozók lineáris függetlensége (KENNEDY). Ezért, egyes forrásokban multikollinearitás alatt a tényezőváltozók lineáris függetlenségének hiányát értik. Ezt a megközelítést a multikollinearitás egy speciális esetének tekinthető, melyet extrém multikollinearitásnak nevezünk. Ez az eset a gyakorlatban nem okoz különösebb problémát, mivel könnyen kezelhető.

Az empirikus elemzések során nagyon gyakran találkozhatunk az extrém multikollinearitáshoz közeli esetekkel, amikor is egyes becsült paraméterek varianciái nagymértékben növekednek a hibátag szórásnégyzetéhez képest. A multikollinearitással foglalkozó szakirodalmak döntő többsége ezzel az esettel foglalkozik. Azonban, megjegyzem, hogy multikollinearitás alatt sokkal általánosabb jelenséget is lehetne érteni, mégpedig **a tényezőváltozók együttmozgását**. Természetesen ennek a meghatározásnak a speciális esetei mindenki számára visszaadnák azt a tartalmat, amit a multikollinearitás alatt ért.

Empirikus vizsgálatoknál gyakran **komoly problémát jelent a multikollinearitás felismerése és okának megtalálása**, hiszen egyrészt **a multikollinearitás negatív következményei nem mindig lépnek fel**, másrészt **a multikollinearitást nem csak egy változó, hanem egy változócsoport is okozhatja**. Így sejtethető, hogy **a multikollinearitás mérőszámai nem minden esetben jellemzik megfelelően ezt a jelenséget**.

A multikollinearitás mérőszámainak értelmezése sokszor meglehetősen szubjektív. Ugyanis, egyrészt a mérőszámok többsége arra ad választ, hogy a vizsgált adatállomány mennyire nem ideális, azaz milyen mértékben térünk el az „ideális esettől”, amikor is minden tényezőváltozó lineárisan független egymástól. Néhány mérőszám esetén nincs egyértelmű határ az „eltérés” káros mértékű jelzésére. Másrészt, ha az alkalmazott modell specifikációja megfelelő, akkor a multikollinearitás csak a megfelelő információ hiányának következménye.

A multikollinearitás negatív hatásainak csökkentésére, illetve kiküszöbölésére gyakrabban használt módszerek sikeressége nagymértékben függhet a multikollinearitás pontos felismerésétől. Ezen módszerek többségének alkalmazása ugyan csökkenti, pontosabban csökkentheti a multikollinearitás negatív következményeinek mértékét, de ez más negatív következményekkel – például jelentős információvesztéssel, az eredmények nem megfelelő értelmezhetőségével – járhat.

A téma aktualitását az adja, hogy ezek a problémák a gazdasági elemzések során szinte kivétel nélkül jelentkeznek. Különösen igaz ez, ha a magyarázóváltozóban erős trend van, vagy ha túlságosan kevés információ áll rendelkezésre ahhoz, hogy a tényezőváltozóknak az eredményváltozóra gyakorolt hatását vizsgáljuk.

Összefoglalva, empirikus elemzéseknél gyakori eset, hogy a vizsgálat szempontjából nem minden adat hordoz hasznos tartalmat, azaz az adatállomány redundáns. Többváltozós lineáris regressziószámításnál a multikollinearitás a redundancia egy fajtájaként értelmezhető. Ezért a regressziószámítás során fontos tudni a $\hat{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$ becslőfüggvény szempontjából hasznos tartalmat hordozó adatok arányát, de probléma ennek a megfelelő mérése. Kérdéses, hogy mit jeleznek a multikollinearitás mérőszámai, illetve az, hogy a multikollinearitás jelenlétének negatív következményei hogyan csökkenthetőek.

I. 2. Az értekezés célja

A $\hat{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{y}$ becslőfüggvény szempontjából hasznos tartalmat hordozó adatok arányának mérésére, egy lehetőség a PETRES-féle *Red*-mutató. A redundancia és így a multikollinearitás egy új, lehetséges mérőszáma a PETRES-féle *Red*-mutató. A *Red*-mutató definiálásakor a tényezőváltozók \mathbf{R} korrelációs mátrixának λ_j ($j=1,2,\dots,m$) sajátértékeit alkalmazzuk. A *Red*-mutató az alábbi gondolatmeneten alapszik. Ha a magyarázóváltozók forrásául szolgáló adatállomány a $\tilde{\beta}$ becslőfüggvény szempontjából redundáns, azaz nagymértékű az adatok együttmozgása, akkor nem mindegyik adat hordoz hasznos tartalmat. Minél kisebb a hasznos tartalmat hordozó adatok aránya, annál nagyobb a redundancia mértéke. Minél nagyobb mértékben szóródnak a sajátértékek, annál nagyobb mértékű az adatállományban szereplő magyarázóváltozók együttmozgása. Két szélsőséges eset létezik: minden sajátérték egyenlő egymással (azaz értékük egy), illetve egy sajátérték kivételével mindegyik sajátérték nullával egyenlő. A diszperzió mértékét számszerűsíthetjük a sajátértékek relatív szórásával vagy (ebben az esetben az ezzel egyenlő) szórásával.

$$v_{\lambda} = \frac{\sigma_{\lambda}}{\bar{\lambda}} = \frac{\sqrt{\frac{\sum_{j=1}^m (\lambda_j - \bar{\lambda})^2}{m}}}{\frac{\sum_{j=1}^m \lambda_j}{m}} = \frac{\sqrt{\frac{\sum_{j=1}^m (\lambda_j - \bar{\lambda})^2}{m}}}{\frac{m}{m}} = \sqrt{\frac{\sum_{j=1}^m (\lambda_j - 1)^2}{m}} = \sigma_{\lambda}$$

Különböző adatállományok redundanciájának összevethetősége végett a fenti mutatót normálni kell. Mivel a sajátértékek nemnegatívak, ezért a relatív szórásra vonatkozó $0 \leq v_{\lambda} \leq \sqrt{m-1}$ összefüggés miatt, a normálás $\sqrt{m-1}$ értékével történik.

Az így kapott mutatót a redundancia mértékének számszerűsítésére használhatjuk, és segítségével a *Red*-mutatót az alábbiak szerint definiáljuk.

$$Red = \frac{v_{\lambda}}{\sqrt{m-1}}$$

A redundancia hiánya esetén a fenti mutató értéke nulla, illetve nulla százalék, míg maximális redundancia esetén egy, illetve száz százalék.

A *Red*-mutató a vizsgált, adott méretű adatállomány redundanciáját méri. Két vagy több különböző méretű adatállomány redundanciájának összevetésekor a *Red*-mutatók alapján csak annyi állítható, hogy az egyes adatállományok mennyire redundánsak, de arra vonatkozó közvetlen kijelentés nem tehető, hogy ezek közül melyiknek van több hasznosítható adata.

I. 3. Az értekezés felépítése

Disszertációm **célja a *Red*-mutató tulajdonságainak vizsgálata, valamint más mérőszámokkal történő összehasonlítása, a többváltozós, lineáris regressziós modellen bemutatva.** A disszertáció céljának megfelelően **értekezésem az alábbi felépítést követi.**

Az értekezés I. fejezetében történik a disszertáció problémájának, feladatainak és célkitűzéseinek megfogalmazása. Ehhez, a fejezetben röviden összefoglaltam a regressziószámítás azon alapismereteit, amelyek a dolgozat megértéséhez szükségesek.

A II. fejezetben áttekintettem a multikollinearitással kapcsolatos szakirodalmat. Ebben a fejezetben tárgyalásra kerül a multikollinearitás számos ismert, illetve kevésbé ismert mérőszáma, detektálási módja, lehetséges következményei, valamint ezek negatív hatásának csökkentési lehetőségei:

Taglalt detektálási eljárások és mérőszámok

- A tényezőváltozók korrelációs mátrixának vizsgálata
- KLEIN-féle hüvelykujj szabály
- MASON és PERREAULT javaslata
- M_1 -mutató
- M -mutató
- FARRAR–GLAUBER teszt
- WILKS teszt
- A korrelációs együtthatók és a parciális korrelációs együtthatók különbözőségeinek vizsgálata
- FRISCH sugárkéve-térképek módszere
- VIF -mutató
- BELSLEY-féle gamma-mutató
- FELLMAN L -mutatója
- MAHAYAN és LAWLES M_1 mutatója
- THISTED mci (multicollinearity index) és $pmci$ (predicted multicollinearity index) mutatója,
- $ISRM$ -mutató (Index of Stability of Relative Magnitudes)
- DEF -mutató (Direct Effect Factor),

Taglalt eljárások a multikollinearitás káros hatásainak csökkentésére

- Tényezőváltozók elhagyása a modelltől
- A minta elemszámának növelése

- Külső információk felhasználása
- MOORE–PENROSE inverz használata
- Főkomponensanalízis
- Ridge regresszió
- *Nested estimate* eljárás
- A tényezőváltozók ortogonalitásának vizsgálata

A fejezet zárásaként egy példán szemléltettem az említett eljárásokat, mutatókat.

A szakirodalom áttekintése és az empirikus példák alapján az alábbi megállapításokat tettem.

- 1. A multikollinearitás negatív következményeként leggyakrabban említik a becsült paraméterek varianciáinak növekedést, azonban ezeknek nem az abszolút nagyságát kellene nézni, hanem azt, hogy mekkora ezeknek a „felfújódása” a hibatag varianciájához képest.**
- 2. A multikollinearitás detektálásának és mérésének számos módja ismert, azonban ezek közül kevés a széles körben elfogadott mivel, egyrészt a multikollinearitás detektálása sokszor nagyon nehéz feladat, másrészt a mutatók többségének értelmezése meglehetősen szubjektív. A mérőszámok, eljárások egy része általában csak detektálják a multikollinearitást, de – általában szintetikus jellegük miatt – nem lokalizálják a problémát. Ezzel szemben a mérőszámoknak és eljárásoknak egy csoportja – több-kevesebb sikerrel – megpróbálja lokalizálni a multikollinearitást.**
- 3. A sajátértékek reciprokait használó indikátorok nagy hátránya, hogy értelmezésük szubjektív, azaz nincs egy olyan egyértelmű küszöbszám, ami már erős multikollinearitást jelez. A mutatók értékei egymással nem összehasonlíthatóak. Továbbá ezen mutatók értékei főleg csak a legkisebb sajátértéktől függenek.**
- 4. A bemutatott mérőszámok más és más szempontból jellemzik a multikollinearitást.**

5. Nincs olyan eljárás a multikollinearitás negatív következményeinek csökkentésére, mely általános érvényű lenne, azaz – más szempontból - minden eljárásnak lehetnek káros mellékhatásai.
6. Az ismertett és alkalmazott mutatók, gondolatok, algoritmusok összegzéseként elmondható, hogy az említett mérőszámok és eljárások nem általános érvényűek abban az értelemben, hogy a multikollinearitás jelenségét csak speciális esetekben jellemzik, illetve kezelik megfelelően.

A III. fejezetben ismertetem a kutatásaim során alkalmazott módszereket, illetve ezek eredményeit. Megvizsgálom a *Red*-mutató főbb tulajdonságait. Itt ismertetem más, hasonló vizsgálati módszerek eredményeit, összevetve az általam kapott eredményekkel.

A dolgozat további fejezetei a kutatási tevékenységem és eredményeim értékelését, a felhasznált irodalmak, ábrák, táblázatok jegyzékét, a hosszabb számítógépes elemzések kimeneteit, és publikációim felsorolását tartalmazza.

I. 4. A kutatási hipotézisek

A disszertáció céljának eléréséhez az alábbiakban ismertett problémaköröket, illetve hipotéziseket vizsgálom.

1. A *Red*-mutató más módon történő kiszámítása.

A *Red*-mutató definíciója szerint a korrelációs mátrix sajátértékei alapján számítható ki. Felmerülhet a kérdés, hogy a sajátértékek ismerete nélkül kiszámítható-e a mutató értéke, pusztán a tényezőváltozók korrelációs mátrixának elemei alapján. A III.1. fejezetben megvizsgáltam az alábbi hipotézist.

1. Hipotézis: A Red-mutató kifejezhető a tényezőváltozók korrelációs mátrixa sajátértékeinek ismerete nélkül, pusztán a páronkénti korrelációs együtthatók alapján.

2. A multikollinearitás vizsgálati módszerének általánosítása

Úgy gondolom, hogy a multikollinearitás vizsgálatakor nem csak változópárok együttmozgása, hanem változócsoportok együttmozgása is problémát jelenthet. Ennek

azonban még nincs részletesen kidolgozott módszertana. Úgy látom, hogy a problémára megoldást jelenthet a kanonikus korrelációelemzés használata, melynek egy speciális helyzete vizsgálható a *Red*-mutató segítségével. A III.1. fejezetben megvizsgáltam az alábbi hipotézist.

2. Hipotézis: Tényezőváltozók két csoportja együttmozgásának vizsgálata speciális esetekben a *Red*-mutató segítségével lehetséges.

3. A multikollinearitás új modellezési lehetőségének vizsgálata.

A multikollinearitás modellezésének egy módja a tényezőváltozók ortogonalitásának, azaz a tényezőváltozók tere „kifeszítettségének” vizsgálata. Jogos kérdés, hogy lehet-e másképpen modellezni a multikollinearitást. A III.2. fejezetben megvizsgáltam az alábbi hipotézist.

3. Hipotézis: Új megközelítésként megalkotható a multikollinearitás elliptikus modellje a *Red*-mutató alapján.

4. Valamilyen kapcsolat keresése a becsült regressziós paraméterek varianciái és a *Red*-mutató között.

Mivel a multikollinearitás egyik leggyakrabban említett negatív következménye a becsült regressziós paraméterek varianciáinak, illetve ezek felfújódásának növekedése, ezért célszerű megvizsgálni a *Red*-mutató és a becsült regressziós paraméterek varianciáinak kapcsolatát. A III.3. fejezetben megvizsgáltam az alábbi hipotézist.

4. Hipotézis: Megadható a *Red*-mutató egy olyan kritikus értéke, amely szükséges feltétele annak, hogy a becsült paraméterek varianciái ne legyenek végtelenek.

5. A *Red*-mutató eloszlásának vizsgálata.

A III.4. fejezetben megpróbáltam a *Red*-mutató empirikus eloszlásfüggvényét elkészíteni, illetve az elméleti eloszlását meghatározni.

6. A *Red*-mutató alkalmazási lehetőségeinek vizsgálata.

Érdekes kérdés, hogy a *Red*-mutató milyen területeken alkalmazható. A III.5. fejezetben megvizsgáltam az alábbi hipotézist.

5. Hipotézis: A *Red*-mutató alapján kifejezhető a faktoranalízis során használt *KMO*-mutató.

7. A *Red*-mutatóhoz hasonló mérőszám megadása.

Mivel a *Red*-mutató a tényezőváltozók korrelációs mátrixának sajátértékei alapján számított normált relatív szórás, ezért úgy gondolom, hogy a multikollinearitás mérhető a sajátértékek más szóródási mérőszámával is, melynek alap gondolata megegyezik a *Red*-mutató alapötletével. A III.6. fejezetben bebizonyítottam az alábbi hipotézist.

6. Hipotézis: A *Red*-mutató definiálásának gondolatmenetén alapuló hasonló multikollinearitás mérőszám a tényezőváltozók korrelációs mátrixa sajátértékeinek GINI-együtthatója.

II. Kutatásaim eredményei, megállapításaim

Az értekezés III. fejezete tartalmazza a dolgozat új eredményeit. A vizsgálatok egy része elméleti megfontolásokon alapszik, míg egy másik részéhez különböző mintákat kellett létrehozni és ezek eredményeit elemezni. Az elemzésekhez az SPSS 13.0; illetve Microsoft Excel programokat használtam. A geometriai ábrázolás elkészítése a Derive 6.0 segítségével történt. Összefoglalva értekezésem az alábbi téziseket tartalmazza.

1. Tézis: A *Red*-mutató kifejezhető a tényezőváltozók korrelációs mátrixa sajátértékeinek ismerete nélkül, pusztán a páronkénti korrelációs együtthatók négyzetes átlagaként.

A *Red*-mutatót sikerült kifejeznem – a sajátértékek ismerete nélkül – a tényezőváltozók korrelációs mátrixának főátlón kívüli elemeinek négyzetes átlagaként. Ez azt jelenti, hogy a mutató nem csak a $\tilde{\beta}$ becslőfüggvény szempontjából hasznos tartalmat hordozó adatok arányát mutatja, hanem a tényezőváltozók együttmozgásának átlagos mértékét

is. Ezt az eredményt több nemzetközi konferencián elismerték, illetve rangos nemzetközi szaklapban is hivatkoztak erre.

2. Tézis: Tényezőváltozók két csoportja együttmozgásának vizsgálata egy-egy elemű csoportok esetén a *Red*-mutatóval, míg egy- $(m-1)$ elemű csoportok esetén a VIF_j mutatók harmonikus átlagának segítségével lehetséges.

Megállapítottam, hogy a multikollinearitást nem csak változók, hanem változócsoportok is okozhatják. Mivel ennek nincs bőséges szakirodalma, ezért a későbbiekben vizsgálni fogom a változócsoportok együttmozgásának hatását. **Megállapítottam, hogy ennek egyik speciális esete a *Red*-mutató segítségével, míg egy másik speciális esete a VIF_j -mutatók harmonikus átlagának segítségével mérhető.**

3. Tézis: Új megközelítésként megalkotható a multikollinearitás elliptikus modellje a *Red*-mutató alapján.

Új megközelítésként, megalkottam a multikollinearitás elliptikus modelljét. A változók átlagos együttmozgása mértékének növekedésével párhuzamosan, a „lehetséges sajátértékek”, egy nagyobb sugarú m -dimenziós gömbön helyezkednek el. **A „lehetséges sajátértékek” az m -dimenziós gömbnek egy metszetén helyezkednek el úgy, hogy rögzített *Red* érték mellett ezek egy $(m-1)$ -dimenziós ellipszoidon helyezkednek el.**

Sajnos, minél nagyobb a modell dimenziószáma, annál több feltételt kell megadni a „lehetséges sajátértékek” tartományának körülhatárolásához, vizsgálatához. Ezért ennek a tartománynak, illetve az elliptikus görbéknek a részletes vizsgálata csak három magyarázóváltozó esetén történt meg. Egy sajátérték függvényében behatároltam a *Red*-mutató lehetséges értékeit, illetve a *Red*-mutató értékétől függően meg tudtam adni az egyes sajátértékek lehetséges értékeit. Összehasonlítottam, hogy a „lehetséges sajátértékek” tartományát hogyan „járják be” az ellipszisek, illetve a sajátértékek legnagyobb és legkisebb értékének azonos értékű hányadosait tartalmazó egyenesek.

A későbbiekben megpróbálom a modellt továbbfejleszteni, illetve a vizsgálatot magasabb dimenziókra kiterjeszteni.

4. Tézis: Megadható a *Red*-mutató egy olyan kritikus értéke, amely szükséges feltétele annak, hogy a becült paraméterek varianciái ne legyenek végtelenek.

Mivel a *Red*-mutató egy szintetikus mutató, ezért a becült regressziós paraméterek varianciáihoz külön-külön nem lehet kötni. Megállapítottam, hogy a becült regressziós paraméterek varianciáinak nem az abszolút nagyságát kell vizsgálni, hanem ezeknek a hibatag szórásnégyzetéhez viszonyított „felfújódását”. Ezek összege, átlaga pedig a sajátértékek reciprokösszegétől függ. Bebizonyítottam, hogy a sajátértékek harmonikus átlagának és a becült regressziós paraméterek számtani átlagának szorzata megegyezik a hibatag szórásnégyzetével, illetve a sajátértékek harmonikus átlagának és a VIF_j-mutatók számtani átlagának szorzata egy.

Egy korábbi állításom cáfolása után, megadtam a *Red*-mutatók egy olyan kritikus értékét, amely szükséges feltétele annak, hogy a becült paraméterek varianciái ne legyenek végtelenek, illetve olyan kritikus értékeket, amelyek szükséges feltételei annak, hogy a zéró sajátértékek száma k darabnál kevesebb legyen. Mivel ennek így önmagában kevés gyakorlati jelentősége van, ezért ez további részletes vizsgálatokat igényel. Azonban, ezeket a vizsgálatokat három tényezőváltozó esetén az elliptikus modell használatával elvégeztem. Megfigyeltem, hogy a „lehetséges sajátértékek” tartományának alsó határától távolodva a sajátértékek reciprokösszege nő. Ez alapján a *Red*-mutató függvényében megadtam a becült paraméterek varianciái összegének a hibatag szórásnégyzetéhez viszonyított „felfújódásának” legkisebb és – ha lehetett – legnagyobb mértékét. Ez alapján a *Red*-mutató függvényében megadható egy olyan kritikus érték, amely szükséges feltétele annak, hogy a becült paraméterek varianciáinak összege egy előre adott mértéknél jobban ne „fújódjanak fel” a hibatag szórásnégyzetéhez viszonyítva.

A *Red*-mutató eloszlásának vizsgálata során elkészítettem néhány dimenzióban az empirikus eloszlásfüggvényt. Az elemzés során csak létező korrelációs struktúrákat vizsgáltam. A „lehetséges sajátértékek” generálásához és a *Red*-mutató eloszlásának elkészítéséhez egy saját magam által írt algoritmust használtam. A készített algoritmus lényege az, hogy adott pontosság mellett előállítjuk az összes lehetséges sajátérték kombinációt. Az elemzést nehezítette, hogy a legenerált sajátértékek száma már

durvább közelítés esetén is több százezer, több százmillió lehet. A **Red-mutató eloszlásának azonosítása sikertelen volt.** További vizsgálatok elvégzéséhez nagy teljesítményű számítógépekre lenne szükség.

5. Tézis: A Red-mutató alapján kifejezhető a faktoranalízis során használt KMO-mutató.

Javasoltam a *Red*-mutató egy alkalmazási lehetőségét. A *Red*-mutató alapján kifejezhető a faktoranalízis során használt *KMO*-mutató. Ezek alapján megállapítottam, hogy a parciális korrelációs együtthatók átlagos együttmozgása nem lehet kisebb a korrelációs együtthatók átlagos együttmozgásánál.

6. Tézis: A Red-mutató definiálásának gondolatmenetén alapuló hasonló multikollinearitás mérőszám a tényezőváltozók korrelációs mátrixa sajátértékeinek GINI-együtthatója.

Megadtam a multikollinearitás egy másik lehetséges mérőszámát, ami a *Red*-mutatóval megegyező gondolatmeneten alapul. Ez a mutató a sajátértékek GINI-együtthatója. Megadtam a mérőszám egy könnyen kezelhető kiszámítási módját. Megvizsgáltam a mutató viselkedését – három tényezőváltozó esetén – a „lehetséges sajátértékek” tartományán. A mutató viselkedése további részletes vizsgálatokat igényel.

II. 1. Kutatási irányok a jövőre vonatkozóan

Dolgozatom zárásaként a jövőben tervezett kutatási irányokat foglalom össze a dolgozat felépítésében szereplő sorrendben.

1. Nagyon fontos gyakorlati probléma a multikollinearitás negatív következményeinek csökkenthetősége. Ezért
 - a. egyrészt meg szeretném vizsgálni azt, hogy a ridge-regresszióban alkalmazott torzító paraméterre lehetséges-e valamilyen optimális becslés a *Red*-mutató értéke alapján.
 - b. Másrészt egy változószelekciós eljárást szeretnék készíteni a *Red*-mutató értéke alapján, úgy, hogy a mutatót tényezőváltozónként, parciálisan is

definiálom, mint egy adott tényezőváltozóknak az összes többi tényezőváltozóval vett átlagos együttmozgása.

2. Szeretném folytatni a multikollinearitás kiterjesztésének vizsgálatát, azaz két vagy több tetszőleges számú tényezőváltozóból álló csoport együttmozgása hogyan mérhető, illetve a jelenségnek milyen negatív következményei vannak.
3. A későbbiekben az elliptikus modell további tulajdonságaira szeretnék fényt deríteni mind három tényezőváltozó esetén, mind pedig a magasabb dimenziókban.
4. Szeretném mélyebben megvizsgálni a *Red*-mutató, illetve a parciálisan definiálható *Red*-mutató kapcsolatát a becült regressziós paraméterek „felfújódásával”.
5. A *Red*-mutató elméleti eloszlásának, illetve az empirikus eloszlásának meghatározása egy komoly jövőbeni feladatot jelenthet.
6. A *Red*-mutató hipotetikus értékére vonatkozóan valamilyen statisztikai tesztet szeretnék készíteni.
7. A *Red*-mutató alkalmazási körét mind elméleti módszerek, mind közgazdasági vizsgálatok során szeretném gyarapítani.

III. Publikációim, konferencia-előadásaim

Lektorált, tudományos publikációk

- [1] SZONDI I. – KOVÁCS P. – IDOVKA B. [2002]: A családok helyzete Szeged város lakótelepein, *ACTA JURIDICA ET POLITICA*, Tomus LXII. Fasc. 18., Szeged, 30 oldal.
- [2] FÜLÖP V. – SZONDI I. – KOVÁCS P. [2003]: Lakáscélú állami támogatások és egyéb, a lakáshoz jutást segítő ellátási formák, *PUBLICATIONES DOKTORANDUM JURIDICORUM*, Tomus II. Fasc. 6. Szeged, 30 oldal.
- [3] KOVÁCS P. – SZONDI I. [2003]: Úton az információs társadalom felé, *ACTA JURIDICA ET POLITICA*, Tomus LXIII. Fasc. 13., Szeged, 20 oldal.
- [4] KOVÁCS P. – PETRES T. – TÓTH L. [2004]: Adatállományok redundanciájának mérése, *Statisztikai Szemle*, Budapest, 82. évfolyam 6.-7. szám, 595-604. oldal.
- [5] GYÉMÁNT R. – PETRES T. – KOVÁCS P. [2005]: A Szandzsák, az egyedülálló vallási régió, *Területi statisztika*, 8. (45.) évfolyam 3. szám 278-287. oldal.

- [6] KOVÁCS P. – PETRES T. – TÓTH L. [2005]: A new measure of multicollinearity in linear regression models, *International Statistical Review (ISR)*, Volume 73 Number 3, Voorburg, The Netherlands, 405-412. oldal.
- [7] KOVÁCS P. – PETRES T. – TÓTH L. [2006]: Válogatott fejezetek Statisztikából, Többváltozós statisztikai módszerek, JATEPress, Szeged, 167 oldal.
- [8] KOVÁCS P. [2008]: A multikollinearitás vizsgálata lineáris regressziós modellekben, *Statisztikai Szemle*, Budapest, 86. évfolyam 1 szám, 38-67. oldal.
- [9] LUKOVICS M.– KOVÁCS P. [2008]: Eljárás a területi versenyképesség mérésére, *Területi statisztika*, KSH, 11. (48.) évfolyam, 20 oldal, (megjelenés alatt).
- [10] VILMÁNYI M. – KOVÁCS P. [2008]: Egyetemi-ipar együttműködések teljesítménye és lehetséges vizsgálati módszere, *Kérdőjelek a régiók gazdasági fejlődésében* (szerk. LENGYEL I. – LUKOVICS M.), JATEPress, Szeged, 25 oldal, (megjelenés alatt).
- [11] KOVÁCS P. [2008]: Az információs társadalom szerinti területi egyenlőtlenségek mérése, *Kérdőjelek a régiók gazdasági fejlődésében* (szerk. LENGYEL I. – LUKOVICS M.), JATEPress, Szeged, 11 oldal, (megjelenés alatt).

Oktatási segédanyagok, jegyzetek

- [1] KOVÁCS P. – PETRES T. [2004]: Statisztika Feladatgyűjtemény (közgazdász hallgatók számára), SZTE GTK, 120 oldal.
- [2] KOVÁCS P. – PETRES T. [2004]: Statisztika Feladatgyűjtemény, Dunaújvárosi Főiskola, Dunaújváros, 284 oldal.
- [3] KOVÁCS P. – PETRES T. [2004]: Statisztika Képletgyűjtemény, Dunaújvárosi Főiskola, Dunaújváros, 50 oldal.
- [4] KATONA T. – KOVÁCS P. – PETRES T. [2006]: Általános statisztika, tankönyv, JATEPress, Szeged, 225 oldal.
- [5] KOVÁCS P. – PETRES T. [2006]: Általános Statisztika Feladatgyűjtemény (joghallgatók részére), JATEPress, Szeged, 2005, 132 oldal.
- [6] KOVÁCS P. [2006]: Általános statisztikai alapismeretek, EU távoktatás elektronikus jegyzet, 80 oldal.

- [7] KOVÁCS P. — PETRES T. [2007]: Tanulási útmutató a főiskolák és egyetemek Általános statisztika című tantárgyához, Dunaújvárosi Főiskola, Dunaújváros, 190 oldal.
- [8] KOVÁCS P. — PETRES T. [2008]: Szoftverek alkalmazása az üzleti életben: statisztikai programok, Dunaújvárosi Főiskola, Dunaújváros, 83 oldal.
- [9] KOVÁCS P. — PETRES T. [2008]: Tanulási útmutató a Szoftverek alkalmazása az üzleti életben: statisztikai programok tantárgyához, Dunaújvárosi Főiskola, Dunaújváros, 78 oldal.

Idegen-nyelvű konferencia-kiadványok

- [1] KOVÁCS P. – SZONDI I. [2006]: E-europe- E-Hungary, Ungarn auf der Schwelle in die EU, A Pólay Elemér Alapítvány Könyvtára, sorozatszerkesztő: Balogh Elemér, Szeged, 29.-48. oldal.
- [2] KOVÁCS P. – PETRES T. [2006]: A New Measure of Multicollinearity in Linear Regression Models, International Conference Applied Statistics (2006, Ribno, Slovenia), Program and Abstract, Statistical Society of Slovenia, Ljubljana.
- [3] KOVÁCS P. – LUKOVICS M. [2006]: Classifying Hungarian sub-regions by their competitiveness, Globalization Impact on Regional and Urban Statistics, 25th SCORUS Conference on Regional and Urban Statistics Research, Wroclaw, Poland, <http://www.scorus2006.ae.wroc.pl>, 12 oldal.
- [4] KOVÁCS P. – PETRES T. [2007]: Measure of Multicollinearity with a New, Original Indicator (PETRES' Red) in Linear Regression Models, International Conference on Mathematics & Statistics, Athens Institute for Education Research, Athens, (KIADÁS ALATT)

Magyar nyelvű konferencia-kiadványok

- [1] KOVÁCS P. – LAMPERTNÉ A. I. – PETRES T. [2005]: A multikollinearitás mérése lineáris regressziós modellekben, A Dunaújvárosi Főiskola Közleményei XXVI/II., Dunaújváros, 355-365. oldal.
- [2] KOVÁCS P. [2005]: Statisztikai mintákat generáló algoritmusok, A Dunaújvárosi Főiskola Közleményei XXVI/II., Dunaújváros, 347-354. oldal.

- [3] KOVÁCS P. [2005]: Az informatika alkalmazása a közgazdasági képzésben, Informatika a felsőoktatásban konferencia 2005 CD-melléklete, Debreceni Egyetem Informatikai Kar, Debrecen, 6 oldal.
- [4] KOVÁCS P. [2005]: Az informatika oktatása és lehetőségei a jogászképzésben, Informatika a felsőoktatásban konferencia 2005 CD-melléklete, Debreceni Egyetem Informatikai Kar, Debrecen, 6 oldal.
- [5] KOVÁCS P. – PETRES T. [2006]: A PETRES-féle *Red*-mutató eloszlásának vizsgálata, A Dunaújvárosi Főiskola Közleményei XXVII/II., Dunaújváros, 2006, 521-530. oldal.
- [6] KOVÁCS P. – PETRES T. – LUKOVICS M. [2006]: A PETRES-féle *Red*-mutató alkalmazásának lehetőségei, A Dunaújvárosi Főiskola Közleményei XXVIII., Dunaújváros, 304-316. oldal.

Egyéb tanulmányok

- [1] KOVÁCS P. [2006]: A statisztika oktatásának és oktatásmódszertanának reformálása a saját gyakorlatomban, „A felsőoktatás szerkezeti és tartalmi fejlesztése” tárgyú Humán erőforrás-fejlesztési Operatív Program (HEFOP 3.3.) „Partnerközpontú önértékelési modell megalkotása és továbbképzések a felsőoktatási intézmények humán erőforrásainak fejlesztéséért” tanulmányainak CD gyűjteménye, Dunaújváros, 52 oldal.

Idegennyelvű konferencia-előadások

- [1] KOVÁCS P. – SZONDI I.: eEurope, eHungary, Társadalmi és gazdasági kihívások az Eu-csatlakozás küszöbén, angol nyelvű, nemzetközi konferencia előadás, Szeged, 2004. június 12.
- [2] KOVÁCS P. – LUKOVICS M.: Classifying Hungarian sub-regions by their competitiveness, Globalization Impact on Regional and Urban Statistics, 25th SCORUS Conference on Regional and Urban Statistics Research, Wroclaw, Poland, 2006. augusztus 30.-szeptember 1.

- [3] KOVÁCS P. – PETRES T.: A new measure of multicollinearity in linear regression models, Applied Statistics 2006 International Conference, Ribno (Bled), Slovenia, 2006. szeptember 17.-20.
- [4] KOVÁCS P. – PETRES T.: Measure of Multicollinearity with a New, Original Indicator (PETRES' *Red*) in Linear Regression Models, International Conference on Mathematics & Statistics, ATINER, 2007. június 11.

Magyarnyelvű konferencia-előadások

- [1] KOVÁCS P. – LAMPERTNÉ A. I. – PETRES T.: A multikollinearitás mérése lineáris regressziós modellekben, DUF Közgazdasági szimpózium, Dunaújváros, 2004. november.
- [2] KOVÁCS P.: Statisztikai mintákat generáló algoritmusok, DUF Informatikai szimpózium, Dunaújváros, 2004. november.
- [3] KOVÁCS P.: Az informatika alkalmazása a közgazdasági képzésben, Informatika a felsőoktatásban konferencia 2005, B szekció, Debreceni Egyetem Informatikai Kar, Debrecen, 2005. augusztus 24.-26.
- [4] KOVÁCS P.: Az informatika oktatása és lehetőségei a jogászképzésben, Informatika a felsőoktatásban konferencia 2005, F szekció, Debreceni Egyetem Informatikai Kar, Debrecen, 2005. augusztus 24.-26.
- [5] KOVÁCS P. – PETRES T.: A PETRES-féle *Red*-mutató eloszlásának vizsgálata, Magyar Tudomány Hete a Dunaújvárosi Főiskolán Közgazdasági és menedzsment Konferencia, Dunaújváros, 2005. november 22.
- [6] KOVÁCS P. – PETRES T. – LUKOVICS M.: A PETRES-féle *Red*-mutató alkalmazásának lehetőségei, Magyar Tudomány Hete a Dunaújvárosi Főiskolán Közgazdasági és menedzsment Konferencia, Dunaújváros, 2006. november 16.
- [7] KOVÁCS P. – PETRES T.: A PETRES-féle *Red*-mutató ismertetése, VI. Természet-, Műszaki- és Gazdaságtudományok Alkalmazása Nemzetközi Konferencia, Szombathely, 2007. május 18.
- [8] VILMÁNYI M. – KOVÁCS P.: Egyetemi-ipar együttműködések teljesítménye és lehetséges vizsgálati módszere, "*Kérdőjelek a régiók gazdasági fejlődésében*" Konferencia, Szeged, 2007. november 12.

- [9] KOVÁCS P. : Az információs társadalom szerinti területi egyenlőtlenségek mérése, *"Kérdőjelek a régiók gazdasági fejlődésében"* Konferencia, Szeged, 2007. november 13.