

XALQARO NORDIK UNIVERSITETI

Iqtisodiyot va pedagogika fakulteti, Iqtisodiyot va biznesni boshqarish kafedrası

Fan o'qituvchisi: Sabirov Xasan Nusratovich

Mavzu: KO`P OMILLI KORRELYATSION-REGRESSION TAHLIL

Reja:

- 1. Ko`p omilli korrelyatsion tahlil**
- 2. Ko`p o`zgaruvchili regressiya tenglamasi**
- 3. Ko`p o`zgaruvchili regressiya modeli uchun omillarni tanlash**
- 4. VIF tahlil**

Ko`p omilli korrelyatsion tahlil

Ko`p o`zgaruvchili korrelyatsio-regression tahlil ekonometrikada juda keng qo`llaniladigan usullardan biridir. Ko`p o`zgaruvchili regressiyaning asosiy maqsadi bir nechta omillarning o`zaro hamda natijaviy o`zgaruvchiga ta`sirini o`rganib chiqib, ularning ana shu natijaviy ko`rsatkichga birgalikdagi ta`sirini ifodalovchi optimal modelni topishdan iboratdir.

Ko`p o`zgaruvchili regressiyaning ko`p mantiqiy xulosalari xuddi oddiy regressiyadagi kabi: o`zgarishi tahlil qilinadigan va tushuntirilishi lozim bo`lgan yagona bog`liq o`zgaruvchi Y va ta`sir etuvchi o`zgaruvchilar X_1, X_2, \dots, X_k aniqlanadi. Bu o`zgaruvchilardan iborat tanlanma bir nechta bosh to`plamdan olinadi. Bashorat uchun qulay bo`lgan va ma`lum ma`noda, Y ni —yaxshi tushuntiruvchi o`zgaruvchilarning kombinatsiyalarini tuzishga harakat qilish lozim bo`ladi. Qisqacha qilib aytganda, Y ni statistik ma`noda yaxshi tushuntiradigan omillarni tanlab, ulardan tanlanma olish zarur. Biznes olamida va umuman iqtisodiyotning turli sohalarida bir nechta miqdoriy o`zgaruvchilarning o`zaro munosabatlarini, ularning o`zaro ta`sirini

tahlil qilish bilan bog'liq ko'plab misollarni keltirish mumkin. Shu sababli ko'p o'zgaruvchili regressiyaning amaliy tatbig'i juda keng qamrovlidir. Misol uchun, kompyuter texnikasi sohasida bir soat davomida tuzilgan shartnomalar soni yaxshi bog'liq o'zgaruvchi bo'lishi mumkin. Unga ta'sir etuvchi omillar esa, masalan, reklamaning tashkil etilishi, mahsulot sifati, dizayni, narxi, rangi va h.k. bo'lishi mumkin. Bu omillar *erkin yoki tushuntiruvchi o'zgaruvchilar* deb ataladi. Ular aniqlanib, identifikatsiya qilinadi hamda har birining Y bilan korrelyatsiyasi hisoblanadi.

Ko'p o'zgaruvchili regressiyada *identifikatsiyalash* deganda o'zgaruvchilarning to'g'ri nomlanishi, ularning o'lchov birliklarining aniqlanishi, har bir o'zgaruvchiga mos tanlanma hajmlarining bir xil bo'lishi va h.k.– bir so'z bilan aytganda tanlanmani statistik tahlil uchun qulay ko'rinishga keltirishni tushunish kerak. k ta o'zgaruvchi uchun (k*k) o'lchovli *korrelyatsiya matritsasi* quyidagi ko'rinishda bo'ladi:

O'zgaruvchilar	1	2	3	.	.	.	k
1	r_{11}	r_{12}	r_{13}	.	.	.	r_{1k}
2	r_{21}	r_{22}	r_{23}	.	.	.	r_{2k}
3	r_{31}	r_{32}	r_{33}	.	.	.	r_{3k}
.
.
.
k	r_{k1}	r_{k2}	r_{k3}	.	.	.	r_{kk}

Bu yerda r_{ij} i- va j- o'zgaruvchi (i, j =1, 2,...,k) orasidagi korrelyatsiya koeffitsiyentining qiymatini bildiradi hamda u i-satr va j-ustunning kesishishida turadi. Yana shuni ta'kidlab o'tish lozimki, $r_{ij}=r_{ji}$ va matritsaning diagonalidagi elementlar ($r_{11}, r_{22}, \dots, r_{kk}$) har doim 1 ga teng, chunki o'zgaruvchi o'zi bilan o'zi qat'iy musbat korrelyatsiyalangan.

Agar modelda qatnashayotgan o'zgaruvchilarning bir nechtasi bir paytda o'zaro chiziqli bog'liq bo'lib chiqsa, bunday holat ko'p o'zgaruvchili regressiya masalasidagi omillar orasida *multikollinearlik* deb ataladi. Multikollinearlik hodisasida omillarning bir paytdagi bir-biriga nisbatan

kuchli ta'siri kuzatiladi, ular xuddi orkestr unisonlikda, ya'ni bir-birini to'ldirib o'ynaganidek effekt beradi. Faqat bunday hodisa orkestr uchungina yaxshi, regressiya masalasida esa har bir o'zgaruvchi boshqalariga bog'liq bo'lmagan holda o'z vazifasini bajarishi - bu ideal holat bo'ladi.

Multikollinearlikning salbiyligi shundan iboratki,

- birinchidan, u o'zgaruvchilar ham o'zaro chiziqli bog'liqligini ko'rsatish bilan bir paytda kuzatilgan xususiyatlar bu o'zgaruvchilarning qaysi biriga ko'proq tegishli degan savolni keltirib chiqaradi.

- ikkinchi tomondan, agar bir nechta o'zgaruvchi kuchli korrelyatsiyalangan bo'lsa, ular Y ni tushuntirish uchun deyarli bir xil ma'lumot yetkazib beradi.

- omillar korrelyatsiyalanganligi sababli ularning tushuntiriluvchi o'zgaruvchiga alohida – alohida tasirini o'rganish va ko'p o'zgaruvchili regressiya parametrlarini izohlash qiyinlashadi.

- parametrlarning bahosi ishonchli bo'lmaydi. Ularning standart xatoliklari katta bo'lib, tanlanma hajmi o'zgarishi bilan ular ham (ba'zan ishoralari ham) o'zgaradi. Bu modelning amaliy ahamiyatini yo'qotadi.

Multikollinearlikdan qutulishning bir nechta yo'llari mavjud:

- modeldan o'zaro kuchli korrelyatsiyalangan bir yoki bir nechta o'zgaruvchi chiqarib tashlanadi.

- o'zgaruvchilar ustida shunday shakl almashtirish bajariladiki, ular orasidagi korrelyatsiya kamayadi. Masalan, vaqt qatorlari qaralayotgan bo'lsa, X_t – X_{t-1} , birinchi tartibli ayirmalar olinadi, natijada berilganlardagi trendning yo'qolishi hisobiga korrelyatsiya ham pasayadi.

- kuchli korrelyatsiyalangan o'zgaruvchilarning chiziqli kombinatsiyalaridan yangi o'zgaruvchi sifatida foydalaniladi. Bu *asosiy komponentalar usuli* deyiladi.

Omillarning ichki korrelyatsiyasini hisobga olish yo'llaridan biri – bu o'rindosh tenglamalarga o'tishdir, ya'ni faqat omillarninggina emas, balki ularning o'zaro ta'sirini ham aks ettiruvchi tenglamalarga. Masalan, agar $y=f(x_1, x_2, x_3)$ bo'lsa, u holda quyidagi o'rindosh tenglama qurish mumkin:

$$Y = a + b_1X_1 + b_2X_2 + b_3 \cdot X_3 + b_{12}X_1X_2 + \\ + b_{13}X_1X_3 + b_{23}X_2X_3 + \varepsilon.$$

Bu tenglamaga ikki omilning o'zaro ta'siri kiradi. Undan yuqori tartibli o'zaro ta'sirni ham kiritish mumkin. Agar Fisherning mezoni bo'yicha statistik ahamiyatliligi isbotlansa, ikkinchi tartibli o'zaro ta'sirning ko'rinishi, masalan, $b_{123}X_1X_2X_3$ va h.k. ko'rinishida bo'ladi. Odatda, undan yuqori tartibli o'zaro ta'sirlar statistik ahamiyatsiz bo'lib chiqadi va birinchi, ikkinchi tartiblilari yetarli bo'ladi. Masalan, agar tahlil natijasi X_1 va X_3 ning o'zaro ta'siri statistik ahamiyatliligini ko'rsatgan bo'lsa, u holda tenglamaning ko'rinishi quyidagicha bo'ladi:

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + b_{13}X_1X_3 + \varepsilon$$

Ko'p o'zgaruvchili regressiyada ideal holatda o'zaro kuchli korrelyatsiyalanmagan o'zgaruvchilar mavjud bo'ladi hamda ularning har biri bog'liq o'zgaruvchi o'zgaruvchanligining alohida bir qismini tushuntiradi. Agar ikkita o'zgaruvchi o'zaro kuchli korrelyatsiyalangan bo'lsa, multikollinearlik holati vujudga keladi. Ko'p o'zgaruvchili regressiyada o'zgaruvchilarni tanlashda quyidagi asosiy qoidalarga amal qilish lozim:

1. O'zgaruvchi bog'liq o'zgaruvchi bilan kuchli korrelyatsiyalangan bo'lishi;
2. O'zgaruvchi boshqa o'zgaruvchilar bilan kuchli korrelyatsiyalanmagan bo'lishi kerak.

1-misol. *Quyidagi korrelyatsiya matritsasini tahlil qilib, multikollinearlik muammosi bor yoki yo'qligini aniqlang.*

	Y	X_1	X_2
Y	1		
X_1	0,9872	1	
X_2	-0,9095	0,9194	1

Multikollinearlikni aniqlash uchun korrelyatsiya matritsasining tahlilida quyidagi qoidaga amal qilinadi: ikki o'zgaruvchi orasidagi korrelyatsiya ular va bog'liq o'zgaruvchi orasidagi korrelyatsiyadan absolyut qiymati bo'yicha kichik bo'lishi kerak, ya'ni bunda ishora e'tiborga olinmaydi. Ko'rilayotgan misolda ikki o'zgaruvchi orasidagi korrelyatsiya ($r_{12}=0,92$) absolyut qiymati bo'yicha Y va o'zgaruvchilar orasidagi korrelyatsiyaning kichigidan katta ($r_{Y1}=0,99$ va $r_{Y2}=0,91$): $r_{12} \geq \min\{r_{Y1}; r_{Y2}\}$. Shubhasiz, multikollinearlik mavjud.

Ko'p o'zgaruvchili regressiya tenglamasi

Ko'p o'zgaruvchili regressiya tenglamasi quyidagicha ko'rinishda:

$$Y = \beta_0 + \beta_1 \xi_1 + \beta_2 \xi_2 + \dots + \beta_{k-1} \xi_{k-1} + \varepsilon,$$

bu yerda

Y – bog'liq o'zgaruvchi;

$\xi_1, \xi_2, \dots, \xi_{k-1}$ – tushuntiruvchi o'zgaruvchilar yoki o'zgaruvchilar;

$\beta_0, \beta_1, \beta_2, \dots, \beta_{k-1}$ – modelning bosh to'plamdagi parametrlari;

ε – tasodifiy xatolik.

Ko'p o'zgaruvchili regressiya tenglamasi Y ning haqiqiy qiymatlari va ularning regressiya tekisligida joylashgan bahosi orasidagi masofalar kvadratini, ya'ni

$$\sum (Y - \hat{Y})^2$$

ni minimallashtiruvchi kichik kvadratlar usulidan oddiy regressiyada keng foydalangan edik. Bu yerda ham kichik kvadratlar tamoyili xuddi oddiy regressiyadagi kabi. Faqat ko'p o'zgaruvchili regressiyada o'zgaruvchi soni bittadan ortiq, demak baholanadigan parametrlar soni ikkitadan ortiq bo'ladi. Odatda, regressiya tenglamasi parametrlari $\beta_0, \beta_1, \beta_2, \dots, \beta_{k-1}$ va ε noma'lum bo'ladi, lekin ularning tanlanma qiymatlar yordamida topilgan baholari mavjud. Tanlanma qiymatlar yordamida parametrlarning baholari

topiladi va berilgan nuqtalarga eng yaxshi yaqinlashuvchi *tanlanma regressiya gipertekisligi* deb ataluvchi tekislik aniqlanadi. *Tanlanma regressiya tenglamasi* quyidagi ko'rinishga ega:

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + \dots + b_{k-1}X_{k-1}$$

bu yerda

\hat{Y} – Y ning tanlanma regressiya tenglamasi bo'yicha bahosi;

X_1, X_2, \dots, X_{k-1} , - o'zgaruvchilar;

$b_0, b_1, b_2, \dots, b_{k-1}$ – bosh to'plamdagi $\beta_0, \beta_1, \beta_2, \dots, \beta_{k-1}$ koeffitsiyentlarning bahosi.

Xuddi oddiy regressiya modelidagi kabi parametrlarning $b_0, b_1, b_2, \dots, b_{k-1}$ baholari *kichik kvadratlar usulini* qo'llashdan hosil bo'lgan k ta chiziqli tenglamalar sistemasining yechimidan iborat. Bunday sistemani kompyuter dasturidan foydalanmay yechish katta k larda juda ko'p hisob-kitobni talab qilishi mumkin.

Oddiy regressiyadagi kabi b_0 - Y -kesimning qiymati, ya'ni bir paytda $X_1=X_2=\dots=X_{k-1}=0$ bo'lsa, u holda $\hat{Y}=b_0$ bo'ladi. Bu talqin faqat X_2, X_3, \dots, X_k lar bir paytda 0 ga teng bo'lishi mumkin bo'lgan holdagina qiziqarli bo'lishi mumkin. Biroq ko'p hollarda o'zgaruvchilarning qiymati noldan ancha farqli bo'ladi. Demak, oddiy regressiyadagi kabi ko'p o'zgaruvchili regressiyada ham Y -kesim har doim ham tadqiqotchi uchun foydali ma'lumot bera olmaydi. $b_0, b_1, b_2, \dots, b_{k-1}$ larning qiymati bosh to'plamdagi regressiya koeffitsiyentlari $\beta_0, \beta_1, \beta_2, \dots, \beta_{k-1}$ uchun baho bo'lib xizmat qiladi.

Har bir koeffitsiyent mos o'zgaruvchi bir birlikka o'zgarib, qolganlari o'zgarmay turganida, Y o'zgaruvchining necha birlikka o'zgarishini bildiradi. Demak, regressiya tenglamasi yordamida barcha o'zgaruvchilarning Y ga bir paytdagi ta'siri o'lchanayotganidan qat'iy nazar, boshqa o'zgaruvchilarning ta'siridan mustasno holda, X_i ($i=1, 2, \dots, k-1$) ning Y ga alohida ta'sirini ham hisoblash mumkin ekan. Ya'ni bi, boshqa barcha o'zgaruvchilar o'zgarmas bo'lib turganida, X_i bir birlikka o'zgarganida \hat{Y} ning qanchaga o'zgarishini ko'rsatadi.

Shunday qilib, regressiya koeffitsiyentining har bir bahosi mos o'zgaruvchi bir birlikka o'zgarib, qolganlari o'zgarmay turganida \hat{Y} ning qanday kattalikka o'zgarishini ko'rsatadi.

Yuqorida ko'p o'zgaruvchili regressiya tenglamasi parametrlarining (ozod haddan tashqari) aniq iqtisodiy interpretatsiyasi bor deb aytdik. Biroq, aksariyat hollarda, parametrlarning qaysi birining o'zgarishi tushuntiriluvchi o'zgaruvchiga ko'proq ta'sir ko'rsatishi —sof modeldan birdaniga ko'zga tashlanmaydi, chunki tenglamada qatnashayotgan o'zgaruvchilar turli o'lchov birligiga ega. Standartlashtirilgan o'zgaruvchilardan tuzilgan model maqsadga muvofiq bo'ladi. Korrelyatsiya matritsasining tahlilidan so'ng standartlashtirilgan o'zgaruvchilar uchun quyidagi regressiya tenglamasi olinadi:

$$\hat{Z}_Y = b_1 \cdot Z_{X_1} + b_2 \cdot Z_{X_2} + \dots + b_k \cdot Z_{X_k}$$

Bu yerda b_i - standartlashtirilgan regressiya koeffitsiyentlari;

$$\hat{Z}_Y = \frac{Y - \bar{Y}}{\sigma_Y};$$

$$Z_{X_i} = \frac{X_i - \bar{X}_i}{\sigma_{X_i}}.$$

standartlashtirilgan o'zgaruvchilar bo'lib, ularning o'rta qiymati nolga, o'rtacha kvadratik og'ishi birga teng:

Ko'p o'zgaruvchili regressiya modeli uchun omillarni tanlash

Ko'p o'zgaruvchili regressiya tenglamasiga u yoki bu omillarni kiritish, avvalambor, model qurish bilan shug'ullanayotgan mutaxassisning modellashtirilayotgan ko'rsatkichning boshqa iqtisodiy omillar bilan o'zaro ta'siri haqidagi tasavvuriga bog'liq. Modelda qatnashadigan omillar haqida ma'lumot yig'ishni boshlagan mutaxassis quyidagilarga e'tibor berishi lozim:

- Bu *omillar miqdoriy xarakterga ega* bo'lishi lozim, aks holda, ya'ni ular nomiqdoriy o'zgaruvchilar bo'lsa, ularga miqdoriy tus berish zarur

bo'ladi. Masalan, agar modelga ishchilarning yoki sportchilarning jinsini kiritish lozim bo'lsa, ayolga "1", erkakka — "0" raqamlarini mos qo'yish, ko'chmas mulkning narxi haqidagi masalalarda mulkning joylashgan yeri ranglarga ajratilishi mumkin: metroga yaqin – "0", shahar markazida – "1" va h.k.

• Omillar o'zaro kuchli korrelyatsiyalanmagan va funksional bog'lanmagan bo'lishi lozim. Ko'p o'zgaruvchili regressiya masalasida agar $r_{ij} > 0,7$ bo'lsa, X_i va X_j kuchli korrelyatsiyalangan deyiladi. Bu yerda $r_{ij} - X_i$ va X_j o'zgaruvchilar orasidagi korrelyatsiya koeffitsiyenti. Modelga o'zaro kuchli korrelyatsiyalangan $r_{ij} > 0,7$ va $r_{yxi} < r_{ij}$ bo'lgan omillarning kiritilishi qurilgan modelda statistik ahamiyatsiz parametrlarning qatnashishiga olib kelishi mumkin.

Agar o'zgaruvchilar kuchli korrelyatsiyalangan bo'lsa, ularning natijaviy ko'rsatkichga individual ta'sirini aniqlab bo'lmaydi, demak, regressiya tenglamasining parametrlarini izohlab bo'lmaydi. Masalan,

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2$$

modelda statistik jihatdan $r_{12} = 0$ bo'lsa, X_1 va X_2 omillar o'zaro chiziqli bog'liq bo'lmaydi va X_2 o'zgarmay turganida, X_1 ning bir birlikka o'zgarishi Y ning o'zgarishiga qanday ta'sir etishi haqida gapirish mumkin va, aksincha. Agar $r_{12} = 1$ bo'lsa, u holda X_1 ning bir birlikka o'zgarishi bilan X_2 o'zgarmay qolishi mumkin emas. Shu sababli o'zgaruvchilarning Y ga alohida ta'sirini o'rganish va ularning har biri ko'paytirilayotgan parametr, ya'ni koeffitsiyentlarga to'g'ri izoh berib bo'lmaydi.

VIF tahlil

Kollinearlikni o'lchaydigan yana bir omil –bu **VIF** (*variance inflationary factor*). U quyidagi formula yordamida hisoblanadi:

$$VIF_j = \frac{1}{(1 - R^2)}$$

bu yerda R^2_j – X_j tushuntiruvchi o'zgaruvchining qolgan o'zgaruvchilar bilan ko'p o'zgaruvchili determinatsiya koeffitsiyenti.

Agar o'zgaruvchilar to'plami korrelyatsiyalanmagan bo'lsa, u holda $VIF_j=1$ bo'ladi. Ba'zi konservativ mutaxassislar agar $VIF_j \geq 10$ bo'lsa, u holda X_j va boshqa o'zgaruvchilar orasida juda katta korrelyatsiya mavjud va uni modeldan chiqarib tashlash zarur deb hisoblaydilar. Biroq boshqa ekonometristlar, ko'p hollarda, agar $\max VIF_j \geq 5$ bo'lsa, X_j ni tahlildan chiqarib tashlash, aks holda, t-test o'tkazilgunicha qoldirish mumkin deb hisoblaydilar.

VIF_j tahlil asosida multikollinearlik muammosini *qisman hal etish mumkin*. Ko'p o'zgaruvchili regressiya tenglamasi qurishning eng keng tarqalgan usullari quyidagilardir:

- o'zgaruvchilarni chiqarib tashlash usuli;
- o'zgaruvchilarni kiritish usuli;
- qadamma – qadam regressiya usuli.

Bu usullarning har biri regressiya masalasini o'ziga xos ravishda hal etadi. Birinchi usulda barcha tanlangan o'zgaruvchilar tahlil qilinib, statistik ahamiyatli deb topilganlari yordamida regressiya tenglamasi quriladi. Ikkinchisida avval kiritilmagan qo'shimcha omillar kiritiladi. Uchinchi usul ikkinchisiga o'xshab ketadi, biroq qo'shimcha omillar ketma-ket bitta-bittalab kiritilib boriladi va har bir qadamda korrelyatsiya, dispersiya tahlili amalga oshirilib, determinatsiya koeffitsiyentining o'zgarishi kuzatib boriladi. Umuman olganda esa, ko'p o'zgaruvchili regressiya masalasida bu uchchala usul kompleks holatda ishlatiladi.

Nazorat uchun savollar

1. Ko'p o'zgaruvchili regressiya nima?
2. Oddiy va ko'p o'zgaruvchili regressiya orasidagi farq nimada?
3. Korrelyatsiya matritsasi nimani aks ettiradi?
4. Korrelyatsiya matritsasining xossalari?
5. Multikollinearlik nimani bildiradi?
6. Multikollinearlikning qanday salbiy oqibatlarini sanab o'ting?

7. Multikollinearlikni qanday qilib yo'qotish mumkin?

8. Ko'p o'zgaruvchili regressiyada o'zgaruvchilarni tanlashning asosiy tamoyillarini sanab o'ting.

9. VIF –tahlilni qachon o'tkazish kerak?