XALQARO NORDIK UNIVERSITETI

Iqtisodiyot va pedagogika fakulteti,
Iqtisodiyot va biznesni boshqarish kafedrasi

Fan o‘qituvchisi: Sabirov Xasan Nusratovich

Mavzu: O‘zgaruvchilar indikatoridan foydalanish

Reja:

1. **O’zgaruvchilar indikatori**
2. **O’zgaruvchilar indikatorini qo’llash**
3. **Chiziqli ehtimollik model**

7.1. O’zgaruvchilar indikatori

Indikator, yoki dummy model, o’zgaruvchilar ikkillik sanoq sistemasi 0/1 o’zgaruvchilari bo’lib, a ning mavjud yoki yoqligini harakterlaydi. Ushbu bo’limda biz ko’chmas mulk misolida o’zgaruvchi indikatorilaridan foydalanishni o’rganamiz. Yangi log faylini oching va ***utown.dta*** data faylini oching.

log using chap07_utown, replace text
use utown, clear
describe
summarize

Ma'lumotlarni va birinchi oltita kuzatuvni umumlashtiring.

list in 1/6
list in 501/506

```
. list in 1/6
```

	price	sqft	age	utown	pool	fplace
1.	205.452	23.46	6	0	0	1
2.	185.328	20.03	5	0	0	1
3.	248.422	27.77	6	0	0	0
4.	154.69	20.17	1	0	0	0
5.	221.801	26.45	0	0	0	1
6.	199.119	21.56	6	0	0	1

```
. list in 501/506
```

	price	sqft	age	utown	pool	fplace
501.	314.65	29.28	24	1	1	0
502.	288.556	24.48	4	1	0	1
503.	302.834	27.02	1	1	0	1
504.	247.82	21.26	2	1	0	1
505.	269.971	22.76	4	1	0	0
506.	292.926	26	17	1	0	1

7.1.1. O’zgaruvchilar indikatorlarini yaratish

O’zgaruvchilar indikatori uchun POE (Principles of Econometrics) kitobida ko’plab misollar avvaldan o’ylab topilgan va qo’llanib kelinmoqda. Stata dasturida yangi indikatorli o’zgaruvchilarni yaratish uchun ko’plab xususiyatlar mavjud bo’lib, boshqa indikatorli o’zgaruvchilarning qiymatlariga asoslangan holda yangi indikatorli o’zgaruvchilarni hosil qilish uchun “**generate**” (yoki “**gen**”) buyrug’idan foydalaniladi.

Yakuniy satatistik xulosaning har bir detalalarini hisoblash uchun **price** va **sqft** buyruqlaridan foydalaniladi.

summarize price sqft, detail

Katta uylar sifatidagi o’zgaruvching o’zgaruvchilar indikatorini yarating, ya’ni o’lchami 2500 kvadrat futdan katta bo’lgan uylar uchun. Buni amalga oshirishda “**generate**” buyrug’i bilan birgalikda quyidagicha shart qo’yamiz (SQFT>25)

gen large = (sqft > 25)

Agar qo’yilgan shart bo’yicha uylarning maydoni (SQFT>25) 2500 dan katta bo’lsa natija “to’g’ri” va **“generate”** funksiyasi natijaga1 qiymat beradi. Agar aksincha bo’lsa natija “notog’ri” yoki qiymati 0 ga teng bo’ladi.

Eslatma: Biz taqdim etgan darslik ma'lumotlarida yetishmayotgan qiymatlar yo'q. Agar ma'lumotlar yetishmayotgan qiymatlarga ega bo'lsa, yuqorida tavsiflangan "mantiqiy operatorlar" dan foydalanish xavflidir. Misol uchun, agar sqft qiymatlari yetishmayotgan bo'lsa, ular "katta" deb tasniflanadi, bu siz

O’zgaruvchi indikatorlarini yaratilayotganda “o’rta narxli” uylar uchun 1 qiymatga teng o’zgaruvchi foydalaniladi.

gen midprice = (215 < price) & (price < 275)

“**&**” – bu mantiqiy amal bo’lib, ikki shart kiritilganda ularning har ikkisi shartga muvofiqligi yoki aksinchaligi ifodalash uchun ishlatiladi. Ya’ni kiritilgan ikki shart ***PRICE>215*** va ***PRICE<275*** lar har ikkisi ham to’g’ri bo’lsa, ***MIDPRICE*** o’zgaruvchisining qiymati 1 ga teng bo’ladi. Quyida keltirilgan bir necha kuzatuvlarning natijalarini ko’rishingiz mumkin.

list sqft price large midprice in 1/5

```
. list sqft price large midprice in 1/5
```

	sqft	price	large	midprice
1.	23.46	205.452	0	0
2.	20.03	185.328	0	0
3.	27.77	248.422	1	1
4.	20.17	154.69	0	0
5.	26.45	221.801	1	1

7.1.2. O’zgaruvchilar indikatorining regressiyasini baholash

Aslida o’zgaruvchili indikatorli modellarni aniqlash (baholash) boshqa reggresiyalardan farq qilmaydi. Quyidagi modelni ko’rib chiqamiz:

$$PRICE = \beta_1 - \delta_1 UTOWN + \delta_2 SQFT + \gamma(SQFT * UTOWN) + \beta_3 AGE + \delta_2 POOL + \delta_3 FPLACE + e$$

Faktor o'zgaruvchisi yozuvidan foydalanish model

reg price i.utown sqft i.utown#c.sqft age i.pool i.fplace

Regressiya buyrug’i ya’ni ***regress*** qisqartirib ***reg*** sharklida qo’llaniladi. Bu Stata tez-tez ishlatiladigan buyruqlarni qisqartma shaklini qabul qilishiga yaqqol misol (masalan, ***generate*** buyrug’ini ***gen*** shaklda qo’llaniladi). Uzliksiz uchun omil o’zgaruvchisi ***c.*** faqatgina o’zgaruvchilarning o’zaro ta’sir etivchi holatlarda qo’llaniladi.

i.utown#c.sqft kiritmasi **UTOWN** va **SQFT** orasidagi o’zaro tasirni ifodalaydi. Biz Stata ***A,B*** va ***A#B*** larni qabul qilishi uchun ***“A##B”*** dan foydalanishimizning o’zi kifoya.

reg price i.utown##c.sqft age i.pool i.fplace

Hosil bo’ladigan natijada o’zgaruvchi indikatori UTOWN ***1.utown*** kabi va o’zaro ta’sir koffisienti iborasi **SQFT *UTOWN *utown#c.sqft.*** kabi nomoyon bo’ladi.

```
. reg price i.utown##c.sqft age i.pool i.fplace
```

Source	SS	df	MS
Model	1548261.71	6	258043.619
Residual	230184.426	993	231.807076
Total	1778446.14	999	1780.22637

```
Number of obs =    1000
F(  6,   993) = 1113.18
Prob > F      =  0.0000
R-squared     =  0.8706
Adj R-squared =  0.8698
Root MSE      =  15.225
```

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.utown	27.45295	8.422582	3.26	0.001	10.92485	43.98106
sqft	7.612177	.2451765	31.05	0.000	7.131053	8.0933
utown#c.sqft 1	1.299405	.3320478	3.91	0.000	.6478091	1.951001
age	-.1900864	.0512046	-3.71	0.000	-.2905681	-.0896048
1.pool	4.377163	1.196692	3.66	0.000	2.028828	6.725498
1.fplace	1.649176	.9719568	1.70	0.090	-.2581495	3.556501
_cons	24.49998	6.191721	3.96	0.000	12.34962	36.65035

7.1.3. O’zgaruvchilar indikatorining ahamiyatlilik testi

O’zgaruvchili indikator koffisienti haqidagi gipotezalarini tekshirish boshqa koffesientlar haqidagi gipotezalarni tekshirishdan farq qilmaydi. Unversitet shaxarchasining joylashuvining ahamiyatlilik testini o’tkazish uchun biz boshqa nolli bo’lmagan gipotezalarga qarshi $H_0: \delta_1 = 0, \gamma = 0$ nolli gipotezasini qo’yamiz.

Bu gipotezani F-testiga tekshirish uchun post-estimation buyrig’i yordamida amalga oshiriladi. Stata menyusining yuqorisidagi ***Statistics>Postestimation >Tests>Test linear hypotheses*** ketma-ketlikni amalga oshiring yoki regressiyadan so’ng test dialog oynasini ochish uchun ***db test*** ni kiriting. Stata buyrug’i esa quydagicha:

test 1.utown = 0 1.untown#c.sqft

Natija:

```
. test 1.utown 1.utown#c.sqft

 ( 1)  1.utown = 0
 ( 2)  1.utown#c.sqft = 0

       F(  2,   993) = 1954.83
            Prob > F =    0.0000
```

Sinov natijasi asosida, p-qiymati 0.0000 bilan, ahamiyatlilik darajasi α=0.05 yoki hatto α=0.001bo’lmagan sababli biz nolli gipotezamizni inkor etamiz.

7.1.4. Qo’shimcha hisob-kitoblar

Uylarning unversitet oldida joylashuvining funksiyasining reggressiyasi quyidagicha hisoblanadi:

$\widehat{PRICE}$ = (24.5+27.453) + (7.6122+1.2994)SQFT - .1901AGE + 4.3772POOL + 1.6492FPLACE = 51.953+89116SQFT - .1901AGE + 4.3772POOL + 1.6492FPLACE

Universitet yaqinida joylashgan uylar regressiya og’ishi va kesishishini aniqlash uchun ***lincom*** buyrug’idan foydalanamiz.

lincom _cons +1 .utown

lincom c.sqft + 1 .utown#c.sqft

natijalar nafaqat ehtimollikni ko’rsatadi balki ularning 95% ehtimollik orlig’ida ekanini ham ko’rsatadi.

```
. lincom _cons + 1.utown

 ( 1)  1.utown + _cons = 0
```

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	51.95294	5.767235	9.01	0.000	40.63557	63.2703

```
. lincom c.sqft + 1.utown#c.sqft

 ( 1)  sqft + 1.utown#c.sqft = 0
```

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	8.911581	.2247944	39.64	0.000	8.470455	9.352708

7.1.5. O’rtacha marjinal ta`sirni hisoblash

Faktorli o’zgaruvchilarni yozma ko’rinishda foydalanishning yana bir qulaylik tomoni shundaki, marjinal effektlarni hisoblayotganda ***margins*** buyrug’idan to’g’ridan to’g’ri foydalaniladi:

marjins, dydx(*)

Natijada:

```
. margins, dydx(*)

Average marginal effects                        Number of obs     =       1000
Model VCE    : OLS

Expression   : Linear prediction, predict()
dy/dx w.r.t. : 1.utown sqft age 1.pool 1.fplace
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.utown	60.21049	.9646176	62.42	0.000	58.31988	62.10111
sqft	8.286568	.1661803	49.86	0.000	7.96086	8.612275
age	-.1900864	.0512046	-3.71	0.000	-.2904456	-.0897272
1.pool	4.377163	1.196692	3.66	0.000	2.031691	6.722636
1.fplace	1.649176	.9719568	1.70	0.090	-.2558247	3.554176

```
Note: dy/dx for factor levels is the discrete change from the base level.
```

Yuqoridagi natija bu faqat Stata tomonidan chiqarilayotgan o’rtacha marjinal effektlarning qiymatdir. AGE, POOL va FPLACE uchun buni hisoblash mushkil emas. AGE uchun bu AGE ning PRICE ga marjinal ta’siridir, taxminiy koeffisientidir. POOL va FPLACE indikatorli o’zgaruvchilar bo’lganligi uchun ularning marjinal effeti hosil bo`lmaydi, balki birinchi navbatda hovuzga ega bo'lish yoki yo'qligi va ikkinchidan kaminga ega bo'lish yoki yo'qligi o'rtasidagi PRICEning diskret o'zgarishidir. Shunday bo’lishiga qaramasdan ***1.utown*** va ***sqft*** uchun unchalik oddiy emas.

Ikki qo’shnilar o’rtasidagi kutilayotgan taklif etilayotgan PRICE farqlari quyidagicha:

$$(E(PRICE)|UTOWN=1) - (E(PRICE)|UTOWN=0)=\delta_1 +\gamma SQFT$$

Stata da quyidagicha hisoblaydi:

$$\text{AME(UTOWN)} = \frac{1}{N}\sum_{i=1}^{N}(\delta 1 + \gamma SQFT_1) = \delta_1 + \overline{\gamma SQFT}$$

Hisoblangan marjinal effektning dispersiyasi quyidagicha:

$\text{var}(\widehat{AME}(\text{UTOWN}))=\text{var}(\hat{\delta}_1+ \hat{\gamma}\ \overline{SQFT}\ )=\text{var}\ (\hat{\delta}_1)+ \overline{\gamma SQFT}^2\ \text{var}\ (\hat{\gamma}) + 2\text{SQFTcov}\ (\hat{\delta}_1, \hat{\gamma})$

Bu tog'riligini ko'rish uchun quyidagicha ishlarni amlaga oshiramiz:

quietly summarize sqft

scalar asqft = r(mean)

lincom 1.utown+c.sqft#1.utown*asqft

Natija esa:

```
. quietly summarize sqft

. scalar asqft = r(mean)

. lincom 1.utown+c.sqft#1.utown*asqft

 ( 1)  1.utown + 25.20965*1.utown#c.sqft = 0
```

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	60.21049	.9646176	62.42	0.000	58.31757	62.10342

Huddi shunday SQFT ning marjinal effektning ta'siri:

$$\frac{\partial E(PRICE)}{\partial SQFT} = \beta_2 + \gamma UTOWN$$

Stata quyidagicha hisoblaydi:

$$AME(\text{SQFT}) = \frac{1}{N}\sum_{i=1}^{N}\beta_2 + \gamma UTOWN_i = \beta_2 + \gamma\overline{UTOWN}$$

Hisoblash uchun quyidagicha amallarni bajaring:

quietly summarize utown

scalar autown = r(mean)

lincom c.sqft+c.sqft#1.utown*autown

```
. quietly summarize utown

. scalar autown = r(mean)

. lincom c.sqft+c.sqft#1.utown*autown

 ( 1)  sqft + .519*1.utown#c.sqft = 0
```

price	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	8.286568	.1661803	49.86	0.000	7.960463	8.612673

log close

7.2. O'zgaruvchilar indikatorini qo'llash

Ushbu bo'limda biz indikator o'zgaruvchilarning turli xil ilovalarini ko'rib chiqamiz. ***cps4_small.dta*** data faylini oching. Yangi jurnalni ishga tushiring va ma'lumotlarni tekshiring.

log using chap07_cps4, replace text

use cps4_small, clear

describe

```
Contains data from cps4_small.dta
  obs:         1,000
 vars:            12
 size:        23,000 (99.9% of memory free)
```

variable name	storage type	display format	value label	variable label
wage	double	%10.0g		earnings per hour
educ	byte	%8.0g		years of education
exper	byte	%8.0g		post education years experience
hrswk	byte	%8.0g		usual hours worked per week
married	byte	%8.0g		= 1 if married
female	byte	%8.0g		= 1 if female
metro	byte	%8.0g		= 1 if lives in metropolitan area
midwest	byte	%8.0g		= 1 if lives in midwest
south	byte	%8.0g		= 1 if lives in south
west	byte	%8.0g		= 1 if lives in west
black	byte	%8.0g		= 1 if black
asian	byte	%8.0g		= 1 if asian

summarize

Natijada:

```
. summarize
```

Variable	Obs	Mean	Std. Dev.	Min	Max
wage	1000	20.61566	12.83472	1.97	76.39
educ	1000	13.799	2.711079	0	21
exper	1000	26.508	12.85446	2	65
hrswk	1000	39.952	10.3353	0	90
married	1000	.581	.4936423	0	1
female	1000	.514	.5000541	0	1
metro	1000	.78	.4144536	0	1
midwest	1000	.24	.4272968	0	1
south	1000	.296	.4567194	0	1
west	1000	.24	.4272968	0	1
black	1000	.112	.3155243	0	1
asian	1000	.043	.2029586	0	1

7.2.1. Sifat omillarining o’zaro ta’siri

Birinchidan, biz modeldagi ikkita ko'rsatkich o'zgaruvchisi, qora va ayol o'rtasidagi o'zaro ta'sirni ko'rib chiqamiz.

$$WAGE = \beta_1 + \beta_2 EDUC + \delta_1 EDUC + \delta_2 BLACK + \delta_3 FEMALE + \gamma(BLACK \times FEMALE) + e$$

Faktorli o'zgaruvchilar operatoriga "##" yordamida regressiya hisoblanadi.

reg wage educ i.black##i.female

```
. reg wage educ i.black##i.female
```

Source	SS	df	MS	Number of obs =	1000
Model	34370.7606	4	8592.69016	F(4, 995) =	65.67
Residual	130194.667	995	130.848912	Prob > F =	0.0000
Total	164565.428	999	164.730158	R-squared =	0.2089
				Adj R-squared =	0.2057
				Root MSE =	11.439

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	2.070391	.1348781	15.35	0.000	1.805712	2.335069
1.black	-4.169077	1.774714	-2.35	0.019	-7.651689	-.6864656
1.female	-4.784607	.7734139	-6.19	0.000	-6.302317	-3.266898
black#female						
1 1	3.844294	2.327653	1.65	0.099	-.7233779	8.411966
_cons	-5.281159	1.900468	-2.78	0.006	-9.010544	-1.551774

Biz oq tanli erkaklar va qora tanli ayollar o'rtasidagi ish haqi farqini hisoblaymiz.

lincom 1.black + 1.female + 1.black#1.female

Natijada:

```
. lincom 1.black + 1.female + 1.black#1.female

 ( 1)  1.black + 1.female + 1.black#1.female = 0
```

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	-5.10939	1.510567	-3.38	0.001	-8.073652	-2.145128

FEMALE va BLACK orasidagi o’zaro tasirni tekshirish uchun F-test o’kazing.

test 1.female 1.black 1.black#1.female

Natija quyidagicha

```
. test 1.female 1.black 1.black#1.female

 ( 1)  1.female = 0
 ( 2)  1.black = 0
 ( 3)  1.black#1.female = 0

       F(  3,   995) =   14.21
            Prob > F =    0.0000
```

Marjinal effektlarni ***margins*** buyrig’i yordamida hisoblanadi.

margins, dydx(*)

Natija:

```
. margins, dydx(*)

Average marginal effects                        Number of obs   =       1000
Model VCE    : OLS

Expression   : Linear prediction, predict()
dy/dx w.r.t. : educ 1.black 1.female
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
educ	2.070391	.1348781	15.35	0.000	1.806034	2.334747
1.black	-2.19311	1.160919	-1.89	0.059	-4.468469	.0822488
1.female	-4.354046	.7313539	-5.95	0.000	-5.787474	-2.920619

Note: dy/dx for factor levels is the discrete change from the base level.

Eslatib o'tamiz, model quyidagicha:

WAGE = $\beta_1 + \beta_2$EDUC + β_1BLACK + β_1FEMALE + γ(BLACK ×FEMALE) + e

Ya’ni

E (WAGE|FEMALE=1) = $\beta_1 + \beta_2$EDUC + β_1BLACK + $\beta_2 + \gamma$BLACK = ($\beta_1 + \beta_2$) + β_2EDUC + ($\beta_1 + \gamma$) BLACK

va E(WAGE|FEMALE=1) = $\beta_1 + \beta_2$EDUC + β_1BLACK. Farqi esa quyidagicha:

E(WAGE|FEMALE=1) – E(WAGE|FEMALE=0)= $\beta_2 + \gamma$BLACK

Stata da quyidagicha natijaga topiladi:

$$\text{AME (FEMALE)} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\delta}_2 + \hat{\gamma}BLACK_i) = \hat{\delta}_2 + \hat{\gamma}\overline{BLACK}$$

Ayolning o'rtacha marjinal ta'sirini hisoblash quyidagi yordamida tekshiriladi.

quietly summarize black

scalar ablack = r(mean)

lincom 1.female + 1.black#1.female*ablack

```
. lincom 1.female + 1.black#1.female*ablack

 ( 1)  1.female + .112*1.black#1.female = 0
```

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	-4.354046	.7313539	-5.95	0.000	-5.789219	-2.918873

7.2.2. Mintaqaviy o’zgaruvchilarni qo’shish

Keyin bir nechta toifali ko'rsatkich o'zgaruvchilari, mintaqaviy o’zgaruvchilar qo'shing. Modeli quyidagicha:

WAGE= $\beta_1 + \beta_2$EDUC + β_3BLACK + β_4FEMALE + β_5(BLACK×FEMALE) + δ_1SOUTH + δ_2MIDWEST+ δ_3WEST + e

Mintaqaviy indikator o'zgaruvchilari ma'lumotlar faylida allaqachon aniqlanganligi sababli, biz ularni regressiya modeliga shunchaki qo'shamiz.

reg wage educ i.black##i.female i.south i.midwest i.west

Natijaning bir qismi:

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	2.071231	.1344687	15.40	0.000	1.807355	2.335106
1.black	-3.905465	1.786258	-2.19	0.029	-7.410743	-.4001873
1.female	-4.744129	.7698381	-6.16	0.000	-6.254827	-3.233431
black#female 1 1	3.625021	2.318375	1.56	0.118	-.9244618	8.174504
1.south	-.4499056	1.025024	-0.44	0.661	-2.46137	1.561558
1.midwest	-2.608406	1.059644	-2.46	0.014	-4.687807	-.5290049
1.west	.9866332	1.059815	0.93	0.352	-1.093104	3.06637
_cons	-4.80621	2.028691	-2.37	0.018	-8.787229	-.8251912

Ushbu mintaqaviy o’zgaruvchining umumiy ahamiyatini tekshirish uchun ***test*** buyrug`idan foydalanimiz:

test 1.south 1.midwest 1.west

```
. test 1.south 1.midwest 1.west

 ( 1)  1.south = 0
 ( 2)  1.midwest = 0
 ( 3)  1.west = 0

       F(  3,   992) =    4.25
            Prob > F =    0.0054
```

3 ta numerator va 992 bilan F-statistikadan kritik qiymatlarni hisoblang va maxraj erkinlik darajalari ko'rsating.

di "F(3,992,.95) = " invFtail(3,992,.05)

di "F(3,992,.90) = " invFtail(3,992,.10)

Natija:

```
. di "F(3,992,.95) = " invFtail(3,992,.05)
F(3,992,.95) = 2.6138755

. di "F(3,992,.90) = " invFtail(3,992,.10)
F(3,992,.90) = 2.0893205
```

7.2.3. Ikki regressiya uchun tenglik testini o’tkazish

Janubiy mintaqa uchun ish haqi tenglamalarining qolgan qismiga nisbatan ekvivalentligini tekshirish mamlakat (ya'ni janubiy emas) biz regressiya modelidagi har bir o'zgaruvchi uchun janubiy ko'rsatkich o'zgaruvchisi bilan o'zaro ta'sir o'zgaruvchisini yaratamiz. Biz sinab ko'rmoqchi bo'lgan tenglama quyidagicha:

$$WAGE= \beta_1+ \beta_2EDUC + \beta_1BLACK + \beta_2FEMALE + \gamma(BLACK\times FEMALE) + e$$

Indikator o'zgaruvchan o'zaro ta'sirga ega model quyidagicha ko’rinishda

$$WAGE= \beta_1+ \beta_2EDUC + \beta_1BLACK + \beta_2FEMALE + \gamma\,(BLACK\times FEMALE) + \theta_1SOUTH + \theta_2\,(EDUC\ SOUTH) + \theta_3\,(BLACK \times SOUTH) + \theta_4\,(FEMALE \times SOUTH) + \theta_5\,(BLACK \times FEMALE \times SOUTH) + e$$

Biz SOUTH bilan regressiya modelidagi har bir o‘zgaruvchi bilan, kesishuvchi o‘zaro ta’sir bo’yicha ish olib bordik. ## Stata operatoridan foydalanib barcha o’zaro tasir etuvchilarni yarating. Avvalo, BLACK, FEMALE va ularning o’zaro ta’sir etuvchilarini yaratishda biz ***i.black##i.female***. dan foydalanamiz. So’ng to’liq o’zaro ta’sir modelini yaratamiz

reg wage i.south##(c.educ i.black##i.female)

Natija esa:

```
. reg wage i.south##(c.educ i.black##i.female)
```

Source	SS	df	MS	Number of obs =	1000
Model	34581.0189	9	3842.33543	F(9, 990) =	29.26
Residual	129984.409	990	131.297383	Prob > F =	0.0000
				R-squared =	0.2101
				Adj R-squared =	0.2030
Total	164565.428	999	164.730158	Root MSE =	11.459

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.south	3.94391	4.048453	0.97	0.330	-4.000625	11.88845
educ	2.172554	.1664639	13.05	0.000	1.845891	2.499216
1.black	-5.08936	2.64306	-1.93	0.054	-10.276	.0972837
1.female	-5.005078	.8990074	-5.57	0.000	-6.769257	-3.240899
black#female						
1 1	5.305574	3.497267	1.52	0.130	-1.557333	12.16848
south#c.educ						
1	-.308541	.2857343	-1.08	0.280	-.8692554	.2521734
south#black						
1 1	1.704396	3.633327	0.47	0.639	-5.42551	8.834302
south#female						
1 1	.9011198	1.772665	0.51	0.611	-2.577492	4.379732
south#black#female						
1 1 1	-2.935834	4.787647	-0.61	0.540	-12.33094	6.459268
_cons	-6.605572	2.336628	-2.83	0.005	-11.19088	-2.02026

Janubiy va qolgan qismi uchun model o'rtasida "hech qanday farq" yo'qligi haqidagi gipotezani sinab ko'rish uchun mamlakat $H_0 : \theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta_5 = 0$ qo'shma gipotezasini ***test*** bayonoti yordamida tekshiramiz.

test 1.south 1.south#c.educ 1.south#1.black 1.south#1.female /// 1.south#1.black#1.female

Natija:

```
. test 1.south 1.south#c.educ 1.south#1.black 1.south#1.female ///
>      1.south#1.black#1.female

 ( 1)  1.south = 0
 ( 2)  1.south#c.educ = 0
 ( 3)  1.south#1.black = 0
 ( 4)  1.south#1.female = 0
 ( 5)  1.south#1.black#1.female = 0

       F(  5,   990) =    0.32
            Prob > F =    0.9009
```

To'liq o'zaro ta'sirlar modelidan biz QORA va JANUBIY o`zgaruvchilar yordamida birgalikdagi effektni olishimiz mumkin.

lincom 1.black + 1.black#1.south

Natija quyidagicha:

```
. lincom 1.black + 1.black#1.south
 ( 1)  1.black + 1.south#1.black = 0
```

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	-3.384964	2.49305	-1.36	0.175	-8.277233	1.507305

Xuddi shunday, AYOL va SOUTHning birgalikdagi ta'siri yordamida:

lincom 1.female + 1.female#1.south

Natija:

```
. lincom 1.female + 1.female#1.south
 ( 1)  1.female + 1.south#1.female = 0
```

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	-4.103958	1.527785	-2.69	0.007	-7.102027	-1.105889

7.2.4. Alohida regressiyalarni baholash

Oldingi bo'limda to'liq o'zaro ta'sirlangan model yondashuvidan foydalanish o'rniga, F-statistik bo'lishi mumkin kvadrat qoldiqlarning cheklangan va cheklanmagan yig'indisi yordamida hisoblangan. Kvadrat yig'indisi to'liq model uchun qoldiqlar ikkita alohida regressiyadan olingan SSE yig'indisidir.

$SSE_{full} = SSE_{non\text{-}south} + SSE_{south} = 89088.5 + 40895.9 = 129984.4$

Ikki mintaqa bo'yicha hisob-kitoblar standart regress yordamida ***bysort*** buyrug'i bilan samarali amalga oshirilishi mumkin, bu Stata buyruqlarini ma'lumotlarning kichik to'plamlarida takrorlash imkonini beradi. ***help bysort*** ni kiriting. By prefiksdan foydalanish uchun ma'lumotlarni guruhlash o'zgaruvchisi yoki o'zgaruvchilari bo'yicha tartiblash kerak. Agar bu hali bajirmagan bo'lsa, ***by*** va ***sort*** buyriqlarini ***bysort*** shaklida kiriting. Sintaksis buyruqlari quyidagicha:

by varlist: stata_cmd

bysort varlist: stata_cmd

Ikki mintaqa uchun regressiyalarni amalga oshirish uchun bizda mavjud:

bysort south: reg wage educ i.black##i.female

Ya'ni, biz avval SOUTH qiymatlari bo'yicha tartiblaymiz va keyin har bir kuzatish guruhi uchun regressiyani amalga oshiramiz.

```
. bysort south: reg wage educ i.black##i.female

-> south = 0

      Source |       SS       df       MS              Number of obs =     704
-------------+------------------------------           F(  4,   699) =   49.72
       Model |  25346.0083     4  6336.50209            Prob > F      =  0.0000
    Residual |  89088.4615   699  127.451304           R-squared     =  0.2215
-------------+------------------------------           Adj R-squared =  0.2170
       Total |  114434.47    703  162.780185           Root MSE      =  11.289

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |    2.172554   .1640077    13.25   0.000     1.850547     2.49456
     1.black |   -5.08936   2.604061    -1.95   0.051    -10.20208    .0233585
    1.female |   -5.005078   .8857423    -5.65   0.000    -6.744112   -3.266044
             |
black#female |
         1 1 |    5.305574   3.445664     1.54   0.124    -1.459516    12.07066
             |
       _cons |   -6.605572    2.30215    -2.87   0.004    -11.12553   -2.085615
------------------------------------------------------------------------------

-> south = 1

      Source |       SS       df       MS              Number of obs =     296
-------------+------------------------------           F(  4,   291) =   16.43
       Model |  9234.26014     4  2308.56503            Prob > F      =  0.0000
    Residual |  40895.9474   291  140.535902           R-squared     =  0.1842
-------------+------------------------------           Adj R-squared =  0.1730
       Total |  50130.2075   295  169.932907           Root MSE      =  11.855

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |    1.864013   .2402682     7.76   0.000     1.391129    2.336896
     1.black |   -3.384964   2.579268    -1.31   0.190     -8.46135    1.691422
    1.female |   -4.103958   1.580621    -2.60   0.010    -7.214857    -.993059
             |
black#female |
         1 1 |     2.36974   3.382739     0.70   0.484    -4.287995    9.027476
             |
       _cons |   -2.661662   3.420413    -0.78   0.437    -9.393547    4.070223
------------------------------------------------------------------------------
```

Ikki SSE alohida regressiyalardan dispersiya jadvalining tahlilidan kelib chiqadi.

7.2.5. Log-chiziqli modellariga o’zgaruvchilar indikatorining qo`llanilishi

Log-chiziqli modeldagi indikator o'zgaruvchisining aniq ta'sirini hisoblash murakkab ko'rinadi, ammo Stataning ***nlcom*** buyrug'i buni ancha osonlashtiradi.

ln(wage) tuzing va quyidagi tenglamani hisoblang:

$$\ln(WAGE) = \beta_1 + \beta_2 EDUC + \delta FEMALE$$

gen lwage = ln(wage)

reg lwage educ i.female

Natija quyidagicha:

```
. reg lwage educ i.female
```

Source	SS	df	MS	Number of obs =	1000
Model	74.5420655	2	37.2710328	F(2, 997) =	141.70
Residual	262.238666	997	.263027749	Prob > F =	0.0000
Total	336.780731	999	.337117849	R-squared =	0.2213
				Adj R-squared =	0.2198
				Root MSE =	.51286

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0962484	.0060365	15.94	0.000	.0844026	.1080942
1.female	-.243214	.0327275	-7.43	0.000	-.3074367	-.1789913
_cons	1.653868	.0843786	19.60	0.000	1.488288	1.819448

Ko'rsatkichning aniq ta'siri va indikatorli o’zgaruvchisi:

$$100(e^{\delta} - 1)\%$$

Bu ***nlcom.***dan foydalanishni talab qiluvchi parametrlarning chiziqli bo'lmagan funktsiyasi:

nlcom.100*(exp(_b[1.female]) - 1) 100*(exp(_b[1.female]) - 1)

Buning natijasi:

```
. nlcom 100*(exp(_b[1.female]) - 1)

       _nl_1:  100*(exp(_b[1.female]) - 1)
```

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
_nl_1	-21.58963	2.566176	-8.41	0.000	-26.62535	-16.5539

Ya'ni, qolgan hamma narsani o'zgarmagan holda, biz ayol ishchilar erkaklarnikiga qaraganda 21,6% kamroq maosh olishlarini taxmin qilamiz.

Xuddi shunday, biz boshqa chiziqli bo'lmagan marjinal effektlarni hisoblashimiz mumkin. Modelni ko'rib chiqamiz.

$$ln(WAGE) = \beta_1 + \beta_2 EDUC + \beta_3 EXPER + \gamma(EDUC * EXPER$$

Taxminiy marjinal ta'sir

100(β3 + γEDUC)%

Ta'lim va tajribaning o'zaro ta'sirini yarating va uni regressiya modeliga qo'shing

reg lwage c.educ##c.exper

Baholash natijalari:

```
. reg lwage c.educ##c.exper
```

Source	SS	df	MS	Number of obs =	1000
Model	65.5449479	3	21.848316	F(3, 996) =	80.23
Residual	271.235783	996	.272325084	Prob > F =	0.0000
				R-squared =	0.1946
				Adj R-squared =	0.1922
Total	336.780731	999	.337117849	Root MSE =	.52185

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0949385	.0146246	6.49	0.000	.06624	.123637
exper	.0063295	.0066985	0.94	0.345	-.0068153	.0194743
c.educ# c.exper	-.0000364	.0004838	-0.08	0.940	-.0009858	.0009129
_cons	1.392318	.2066447	6.74	0.000	.986809	1.797827

Lincom yoki ***nlcom***-dan foydalanish uchun ***_b [o'zgaruvchi]*** ko'rinishidagi koeffitsient nomlari ba'zan faktor belgilaridan foydalanganda darhol aniqlanmaydi. Buni koeffitsientlar jadvalini emas, balki ***coeflegend*** afsonasini ko'rsatadigan variant koeflegendasi bilan ***regress*** buyrug'ini ko'rsatish orqali aniqlash mumkin.

reg, coeflegend

```
. reg, coeflegend
```

Source	SS	df	MS	Number of obs =	1000
Model	65.5449479	3	21.848316	F(3, 996) =	80.23
Residual	271.235783	996	.272325084	Prob > F =	0.0000
				R-squared =	0.1946
				Adj R-squared =	0.1922
Total	336.780731	999	.337117849	Root MSE =	.52185

lwage	Coef.	Legend
educ	.0949385	_b[educ]
exper	.0063295	_b[exper]
c.educ# c.exper	-.0000364	_b[c.educ#c.exper]
_cons	1.392318	_b[_cons]

Taxminiy va aniq effektlar yordamida hisoblab chiqiladi

lincom 100*(exper+ c.educ#c.exper*16)

```
. lincom 100*(exper+ c.educ#c.exper*16)
 ( 1)  100*exper + 1600*c.educ#c.exper = 0
```

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	.574639	.174402	3.29	0.001	.2324014	.9168765

nlcom 100*(exp(_b[exper]+_b[c.educ#c.exper]*16) - 1)

```
. nlcom 100*(exp( _b[exper]+_b[c.educ#c.exper]*16) - 1)
       _nl_1:  100*(exp( _b[exper]+_b[c.educ#c.exper]*16) - 1)
```

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
_nl_1	.5762932	.1754071	3.29	0.001	.2320833	.920503

log close

7.3. Chiziqli empirik model

Ikki muqobil o'rtasida tanlovni modellashtirishda indikator o'zgaruvchisi bog'liq bo'ladi, regressiya modelida mustaqil o'zgaruvchi emas, balki erksiz o'zgaruvchi. Faraz qilaylik:

$$Y = \begin{cases} 1 & \text{birinchi muqobil tanlansa} \\ 0 & \text{agar ikkinchi variant tanlansa} \end{cases}$$

Agar p birinchi muqobil tanlanish ehtimoli bo‘lsa, u holda P[y=1]=p bo‘lsa, y ning kutilayotgan qiymati E (y) = p, dispersiyasi esa var(y)=p(p-1) bo‘ladi.

Biz chiziqli regressiya funksiyasidan yoki shu nuqtai nazardan chiziqli ehtimollik modelidan foydalangan holda p ehtimolligiga ta'sir etuvchi omillarni aniqlashdan manfaatdormiz.

$E(y) = p = \beta_1 + \beta_2 x_2 + \ldots + \beta_K x_K$

Chiziqli ehtimollik regressiya modeli:

$y = \beta 1 + \beta 2\ x2 + \ldots + \beta_K x_K + e$

Xato atamasining dispersiyasi e ga teng.

$$var(e) = (\beta_1 + \beta_2 x_2 + \ldots + \beta_K x_K)(1 - \beta_1 - \beta_2 x_2 - \ldots - \beta_K x_K)$$

Bu xato homosedastik emas va buni 8-bobda to`liq ko'rib chiqamiz.

Misol sifatida Cola va Pepsi o'rtasidagi tanlovni ko'rib chiqing. ***Coke.dta*** ni oching va uning mazmunini tekshiring.

log using chap07_coke, replace text
use coke, clear
describe
summarize

variable name	storage type	display format	value label	variable label
coke	byte	%8.0g		=1 if coke chosen, =0 if pepsi chosen
pr_pepsi	double	%10.0g		price of 2 liter bottle of pepsi
pr_coke	double	%10.0g		price of 2 liter bottle of coke
disp_pepsi	byte	%8.0g		= 1 if pepsi is displayed at time of purchase, otherwise = 0
disp_coke	byte	%8.0g		= 1 if coke is displayed at time of purchase, otherwise = 0
pratio	double	%10.0g		price of coke relative to price of pepsi

Eng kichik kvadratlar regressiyasidan foydalanib, kola tanlash uchun chiziqli ehtimollik modelini taxmin qilamiz.

reg coke pratio disp_coke disp_pepsi

```
. reg coke pratio disp_coke disp_pepsi
```

Source	SS	df	MS	Number of obs =	1140
Model	33.8378078	3	11.2792693	F(3, 1136) =	51.67
Residual	248.004297	1136	.218313642	Prob > F =	0.0000
				R-squared =	0.1201
Total	281.842105	1139	.247446976	Adj R-squared =	0.1177
				Root MSE =	.46724

coke	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
pratio	-.4008614	.0613494	-6.53	0.000	-.5212324	-.2804904
disp_coke	.0771745	.0343919	2.24	0.025	.0096956	.1446533
disp_pepsi	-.1656637	.0355997	-4.65	0.000	-.2355122	-.0958152
_cons	.8902151	.0654849	13.59	0.000	.7617301	1.0187

Chiziqli regressiya yondashuvi bilan bog'liq tashvish shundaki, bashorat qilingan ehtimolliklar [0, 1] oraliqdan tashqarida bo'lishi mumkin. Bu holda ehtimollik bo'lgan bashorat qilingan qiymatlarni oling va umumlashtiring.

predict phat

summarize phat

```
. summarize phat
```

Variable	Obs	Mean	Std. Dev.	Min	Max
phat	1140	.4473684	.1723611	-.2073211	.7680784

Biz minimal qiymat noldan kichik ekanligini ko'ramiz, lekin bashorat qilingan ehtimollarning hech biri birdan katta emas. Qancha bashorat qilingan ehtimollar salbiy ekanligini ko'rish uchun ulardan foydalanib, ularni umumlashtiring

summarize phat if phat<=0

```
. summarize phat if phat<=0
```

Variable	Obs	Mean	Std. Dev.	Min	Max
phat	16	-.0183585	.0523201	-.2073211	-.0002385

1140 ta kuzatuvdan 16 tasi salbiy bashorat qilingan ehtimollarga ega ekanligini ko'ramiz.

log close

VII bob mavzularini mustahkamlash uchun savollar

1. O’zgaruvchilar indikatori deganda nimani tushumasiz?
2. O’zgaruvchilarning qiymatlariga asoslangan holda yangi indikatorli o’zgaruvchilarni hosil qilish uchun qanday buyruqdan foydalaniladi?
3. Yakuniy satatistik xulosaning har bir detalalarini hisoblash uchun qaysi buyruqlardan foydalanish lozim?
4. O’zgaruvchili indikator koffisienti gipotezalarini tekshirish uchun qanday buyruqlarni bajarish talab etiladi?
5. O’rtacha marjinal ta`sirni hisoblash buyrug’ini ifodalang.
6. indikator o'zgaruvchilarini ifodalashda qanday fayl orqali amalga oshiriladi?
7. Mintaqaviy indikator o'zgaruvchilari ma'lumotlarini regressiya modeliga kiritish uchun qanday amallar bajarilishi lozim?
8. Ikki mintaqa bo'yicha hisob-kitoblar qaysi buyruq yordamida amalga oshiriladi?
9. Chiziqli empirik model tuzish qanday amallarni bajarishni talab etadi?
10. Chiziqli ehtimollik regressiya modelini ifodalang.