

XALQARO NORDIK UNIVERSITETI

Iqtisodiyot va pedagogika fakulteti, Iqtisodiyot va biznesni boshqarish kafedrası

Fan o'qituvchisi: Sabirov Xasan Nusratovich

Mavzu: Ko'p omilli regressiya modellarini baholash

Reja:

1. F-test ni hisoblash
2. Namuna bo'lmagan ma'lumotlar orqali model tuzish
3. Model xususiyatlari
4. Ma'lumotlar yetishmaslik, koleneartlik va ahamiyatsizlik

6.1. F-test ni hisoblash

Ushbu bobda 5-bobda ko'rib chiqilgan savdo modeli yoki Big Andy's Burger Barn modelidan foydalanamiz. Model uchta mustaqil o'zgaruvchini va natijaviy omilni o'z ichiga oladi:

$$SALES_i = \beta_1 + \beta_2 PRICE_i + \beta_3 ADVERT_i + \beta_4 ADVERT_i^2 + e_i$$

bu yerda $SALES_i$ ma'lum bir shahardagi oylik sotuvlar va u 1000 AQSh dollari miqdorida o'lchanadi, $PRICE_i$ - gamburgerning dollarda o'lchanadigan narxi, $ADVERT_i$ - ming dollarda o'lchanadigan reklama xarajatlari va bularda $i=1,2,\dots,N$ ni tashki etadi.

Nol gipoteza shundan iboratki, reklama o'rtacha savdoga ta'sir qilmaydi. Reklamaning barcha qiymatlari uchun marjinal effekt nolga teng bo'lishi uchun $\beta_3 = 0$ va $\beta_4 = 0$ bo'lishi kerak. Muqobil variant $\beta_3 \neq 0$ yoki $\beta_4 \neq 0$. Modelning nol gipotezasi ostidagi parametrlari nolga teng, muqobil ostidagi parametrlar esa cheklanmagan.

F-testi cheklanmagan modeldan cheklangan modelga kvadrat xatolar yig'indisi solishtiradi. Katta farq cheklovlarning noto'g'ri ekanligiga dalil sifatida qabul qilinadi. Nol gipotezani (cheklovlarni) tekshirish uchun foydalaniladigan statistika.

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N - K)}$$

cheklovlar to'g'ri bo'lganda, J numerator va N-K maxraj erkinlik darajalari bilan F-taqsimotiga ega.

Statistik ma'lumotlar ikkita regressiyani bajarish orqali hisoblanadi. Birinchisi cheklanmagan; ikkinchisida cheklovlar mavjud. Har bir regressiyadan kvadratik xatolar yig'indisini, cheklanmagan regressiyadan erkinlik darajalarini (N-K) va o'rnatilgan mustaqil cheklovlar sonini (J) ifodalang. Keyin, quyidagilarni hisoblang:

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(N - K)} = \frac{(1896,91 - 1532,084)/2}{1532,084/(75 - 4)} = 8,44$$

Ushbu modelni taxmin qilish uchun *andy.dta* ma'lumotlar faylidan foydalaning.

use andy, clear

Stata o'zgaruvchilar oynasida siz ma'lumotlar uchta o'zgaruvchidan iborat ekanligini ko'rasiz: ***sales, price, va advert***. Ular cheklanmagan modelni ***regress*** baholash funksiyasi bilan ishlatiladi.

regress sales price advert c.advert#c.advert

`. regress sales price advert c.advert#c.advert`

Source	SS	df	MS			
Model	1583.39741	3	527.799136	Number of obs =	75	
Residual	1532.08446	71	21.5786543	F(3, 71) =	24.46	
Total	3115.48187	74	42.1011063	Prob > F =	0.0000	
				R-squared =	0.5082	
				Adj R-squared =	0.4875	
				Root MSE =	4.6453	

	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
sales						
price	-7.64	1.045939	-7.30	0.000	-9.725543	-5.554457
advert	12.15124	3.556164	3.42	0.001	5.060446	19.24203
c.advert#						
c.advert	-2.767963	.940624	-2.94	0.004	-4.643514	-.892412
_cons	109.719	6.799045	16.14	0.000	96.16212	123.2759

Kvadrat xatolar yig'indisini ***e(ssr)*** yordamida ***sseu*** deb nomlangan yangi skalarga va dispersiya jadvali tahlilidan qolgan erkinlik darajalarini ***e(df_r)*** yordamida ***df_unrest*** deb nomlangan o'zgaruvchiga ifodalang.

scalar sseu = e(ssr) scalar df_unrest = e(df_r)

Keyinchalik, modelga cheklov qo'ying va uni eng kichik kvadratlar yordamida qayta baholang. Shunga qaramay, kvadrat xatolar yig'indisini va qoldiq erkinlik darajalarini saqlang.

regress sales price

```
. regress sales price
```

Source	SS	df	MS			
Model	1219.09103	1	1219.09103	Number of obs =	75	
Residual	1896.39084	73	25.9779567	F(1, 73) =	46.93	
Total	3115.48187	74	42.1011063	Prob > F =	0.0000	
				R-squared =	0.3913	
				Adj R-squared =	0.3830	
				Root MSE =	5.0969	

sales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
price	-7.829074	1.142865	-6.85	0.000	-10.1068	-5.551348
_cons	121.9002	6.526291	18.68	0.000	108.8933	134.9071

scalar sser = e(ssr)

scalar df_rest = e(df_r)

Cheklangan modeldan saqlangan qoldiq erkinlik darajasidan kiritilgan cheklovlar sonini olish uchun foydalanish mumkin. Chiziqli modeldagi har bir noyob cheklash modeldagi parametrlar sonini bittaga kamaytiradi. Shunday qilib, uchta parametrlilik cheklanmagan modelga (masalan, Big Andy's) bitta cheklov qo'yish cheklangan modeldagi parametrlar sonini ikkitaga qisqartiradi. Cheklangan modeldagi regressorlar soni K_r , cheklanmagan modeldagi esa K_u soni bo'lsin. Cheklanmagan modeldagi erkinlik darajalarini $(N-K_u)$ cheklangan modeldagidan $(N-K_r)$ ayirish siz kiritgan cheklovlar sonini beradi, ya'ni $(N-K_r)-(N-K_u) = (K_u-K_r)=J$. Statada,

scalar J = df_rest - df_unrest

Keyin F-statistikasini hisoblash mumkin

scalar fstat = ((s-ser-sseu)/J)/(sseu/(df_unrest))

$F_{(J,N-K)}$ taqsimotidan kritik qiymat va hisoblangan statistika uchun p-qiymati odatdagi usulda hisoblanishi mumkin. Bu holda **invFtail(J,N-K, α)** erkinlik darajalari J numerator va N-K maxraj bilan F-tarqatishdan α darajadagi kritik qiymat hosil qiladi. **Ftail(J,N-K,fstat)** funksiyasi hisoblangan statistika uchun p-qiymatini qaytarish uchun xuddi shunday ishlaydi, **fstat**.

scalar crit1 = invFtail(J,df_unrest,.05)

scalar pvalue = Ftail(J,df_unrest,fstat)

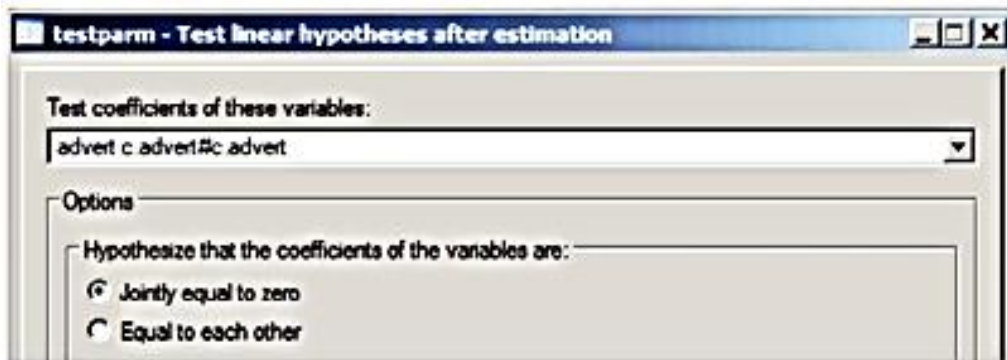
scalar list sseu sser J df_unrest fstat pvalue crit1

Natijada:

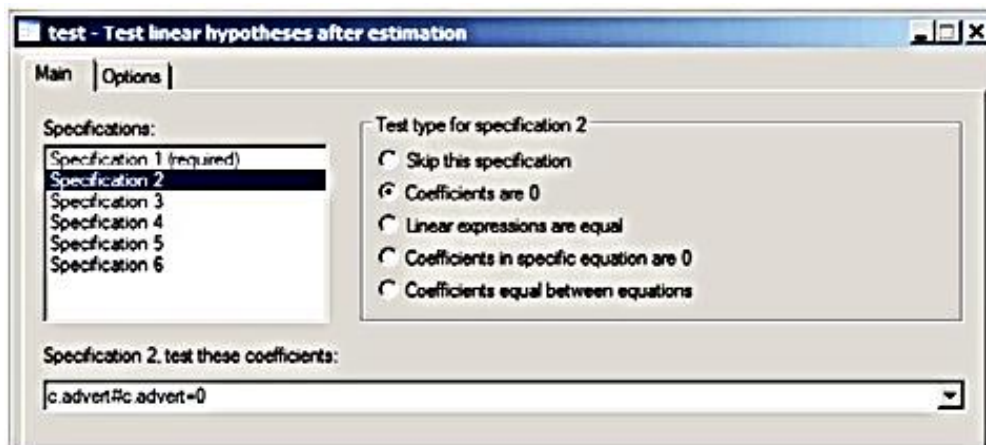
```
. scalar list sseu sser J df_unrest fstat pvalue crit1
      sseu = 1532.0845
      sser = 1896.3908
         J =          2
df_unrest =          71
      fstat =      8.44136
      pvalue = .00051416
      crit1 =  3.1257642
```

Muloqot oynalari model parametrlari bo'yicha cheklovlarni sinab ko'rish uchun ham ishlatilishi mumkin. Birinchi qadam **regress** yordamida modelni baholashdir. Bu xuddi yuqoridagi 5.1-bo'limda bo'lgani kabi davom etadi. Stata ning ochiladigan menyusidan

Select **Statistics>Linear models and related>Linear regression** -ni tanlang. Bu **regress** dialog oynasini ochiladi. **Regress–Linear regression** dialog oynasidagi qaram o‘zgaruvchi sifatida **sales** dan va **price, advert** hamda o‘zaro ta’sir **c.advert#c.advertrt** dan mustaqil o‘zgaruvchilar sifatida foydalanib, **OK** tugmasini bosish orqali regressiyani hisoblang. Regressiya taxmin qilingandan so'ng, gipotezani tekshirish uchun baholashdan keyingi buyruqlar qo'llaniladi. Ochilgan menyudan **Statistics>Postestimation>Tests>Test parameters** -ni tanlang, bu **testparm** dialog oynasini ochadi:



Shuningdek, test dialog oynasidan **Statistics>Postestimation>Tests>Test linear hypotheses** -ni tanlash orqali ham foydalanish mumkin. Test dialogidan foydalanish qiyinroq. Har bir chiziqli gipoteza spetsifikatsiya sifatida kiritilishi kerak. Spetsifikatsiya 1 (majburiy) uchun **advert=0** kiriting va koeffitsientlar nolga teng yoki chiziqli ifodalar teng radio tugma tanlanganligiga ishonch hosil qiling. Keyin Spetsifikatsiya 2 ni belgilang va **c.advert#c.advert=0** kiriting va **Submit** tugmasini bosing. Ushbu bosqich uchun dialog oynasi quyida ko'rsatilgan:



Ikkala holatda ham **Command** oynasidan foydalanish ancha oson. **Testparm** bayonoti parametrlar bo'yicha nol cheklovlarni sinab ko'rish uchun foydalanish uchun eng oddiy hisoblanadi. Sintaksis quyidagicha:

testparm varlist

Bu shuni anglatadiki, null ostida nol koeffitsientga ega bo'lgan o'zgaruvchilarni oddiygina ro'yxatga olish mumkin. Bundan tashqari, teng variantdan foydalangan holda koeffitsientlar bir-biriga teng ekanligini tekshirishga ko'maklashish mumkin.

Test buyrug'i Wald testi yordamida eng so'nggi moslashtirilgan modelning parametrlari haqidagi qo'shma gipotezalarni tekshirish uchun ishlatilishi mumkin. Gipotezalarni aniqlashning bir necha xil usullari mavjud va ulardan bir nechtasi bu qo'llanmada o'rganiladi. Umumiy sintaksis:

test (hypothesis 1) (hypothesis 2)

Qo'shma gipotezalarning har biri qavslar to'plamiga kiritilgan. Chiziqli modelda koeffitsientlarni o'zgaruvchilar nomlari bilan aniqlash mumkin, chunki ularning ma'nosi aniq. Umuman olganda, agar avval belgilangan bo'lsa, parametr nomidan ham, chiziqli modelda ham foydalanish mumkin **_b[variable name]** sintaksisi. Bu yerda qo'shma nullni tekshirishning uchta ekvivalent usuli mavjud.

regress sales price advert c.advert#c.advert

testparm advert c.advert#c.advert

test (advert=0)(c.advert#c.advert=0)

test (_b[advert]=0)(_b[c.advert#c.advert]=0)

```
. testparm advert c.advert#c.advert
( 1)  advert = 0
( 2)  c.advert#c.advert = 0
      F( 2, 71) = 8.44
      Prob > F = 0.0005

. test (advert=0)(c.advert#c.advert=0)
( 1)  advert = 0
( 2)  c.advert#c.advert = 0
      F( 2, 71) = 8.44
      Prob > F = 0.0005

. test (_b[advert]=0)(_b[c.advert#c.advert]=0)
( 1)  advert = 0
( 2)  c.advert#c.advert = 0
      F( 2, 71) = 8.44
      Prob > F = 0.0005
```

6.1.1. Modelning ahamiyatini tekshirish

F-testning ushbu formula natijasida siz modelingiz kerakli statistik ahamiyatlik darajasi muhim yoki yo'qligini aniqlaymiz. K regressorlari bilan umumiy chiziqli modelni ko'rib chiqamiz.

$$y_i = \beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + \dots + x_{iK}\beta_K + e_i$$

Agar erkli o'zgaruvchilar y ning o'rtacha qiymatiga ta'sir qilmasa, unda har bir qiyalik nolga teng bo'lib, nol va muqobil gipotezalarga olib keladi:

$$H_0: \beta_2 = 0, \beta_3 = 0, \dots, \beta_K = 0$$

$$H_1: k = 2, 3, \dots, K \text{ uchun } \beta_k \text{ ning kamida bittasi nolga teng}$$

Bu $J=K-1$ cheklovlariga teng. Shunga qaramay, modelni har biri uchun cheklanmagan va cheklangan erkinlik darajalarini baholang. Keyin, test statistikasini hisoblash uchun yuqoridagi Stata buyrug`idan foydalaning:

$$F = \frac{(SST - SSE)/(K - 1)}{SSE/(N - K)} = \frac{(3115,485 - 1532,08)/3}{1532,084/(75 - 4)} = 24,459$$

Stata buyrug`i:

** Unrestricted Model (all variables)*

regress sales price advert c.advert#c.advert

scalar sseu = e(rss)

scalar df_unrest = e(df_r)

** Restricted Model (no explanatory variables)*

regress sales

scalar sser = e(rss)

scalar df_rest = e(df_r)

scalar J = df_rest - df_unrest

** F-statistic, critical value, pvalue*

scalar fstat = ((sser - sseu)/J)/(sseu/(df_unrest))

scalar crit2 = invFtail(J,df_unrest,.05)

scalar pvalue = Ftail(J,df_unrest,fstat)

scalar list sseu sser J df_unrest fstat pvalue crit2

Buyruq natijasi:

```
. scalar list sseu sser J df_unrest fstat pvalue crit2
      sseu = 1532.0845
      sser = 3115.4819
         J =          3
df_unrest =          71
      fstat = 24.459316
      pvalue = 5.600e-11
      crit2 = 2.7336472
```

Regressiya ahamiyatining ushbu maxsus testi Stata yordamida hisoblangan har bir chiziqli regressiyaning standart natijasida paydo bo'ladigan darajada muhim. Quyidagi natijada ushbu test uchun F-statistika 24,4595 va uning p-qiymati 5% dan ancha past. Shuning uchun biz modelning besh foiz darajasida ahamiyatsizligi haqidagi nol gipotezani rad qilamiz.

. regress sales price advert c.advert#c.advert

Source	SS	df	MS			
Model	1583.39741	3	527.799136	Number of obs =	75	
Residual	1532.08446	71	21.5786543	F(3, 71) =	24.46	
Total	3115.48187	74	42.1011063	Prob > F =	0.0000	
				R-squared =	0.5082	
				Adj R-squared =	0.4875	
				Root MSE =	4.6453	

sales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
price	-7.64	1.045939	-7.30	0.000	-9.725543	-5.554457
advert	12.15124	3.556164	3.42	0.001	5.060446	19.24203
c.advert# c.advert	-2.767963	.940624	-2.94	0.004	-4.643514	-.892412
_cons	109.719	6.799045	16.14	0.000	96.16212	123.2759

6.1.2. t- va F- testlari o'rtasidagi bog'liqlik

Ushbu misolda muhimlik uchun t-testi va F-testning ekvivalentligi ko'rsatilgan. Asosiy model:

$$SALES_i = \beta_1 + \beta_2 PRICE_i + \beta_3 ADVERT_i + \beta_4 ADVERT_i^2 + e_i$$

β_2 uchun t-nisbati -7,30 ga teng (6.1.2-bo'lim oxiridagi natijada ifodalanilgan). F-testi nol emas degan ikki tomonlama muqobilga nisbatan $\beta_2=0$ degan gipotezani tekshirish uchun ishlatilishi mumkin. Cheklangan model esa:

$$SALES_i = \beta_1 + \beta_3 ADVERT_i + \beta_4 ADVERT_i^2 + e_i$$

Cheklanmagan modelni va cheklanmagan modelni baholash hamda Stata-da F-statistikani hisoblash uchun buyruqlar ketma-ketligi:

*** Unrestricted Regression**

regress sales price advert c.advert#c.advert

scalar sseu = e(rss)

scalar df_unrest = e(df_r)

scalar tratio = _b[price]/_se[price]

scalar t_sq = tratio^2

*** Restricted Regression**

regress sales advert c.advert#c.advert

scalar sser = e(rss)

scalar df_rest = e(df_r)

scalar J = df_rest - df_unrest

*** F-statistic, critical value, pvalue**

scalar fstat = ((sser - sseu)/J)/(sseu/(df_unrest))

scalar crit = invFtail(J,df_unrest,.05)

scalar pvalue = Ftail(J,df_unrest,fstat)

scalar list sseu sser J df_unrest fstat pvalue crit tratio t_sq

Buyruqlar natijasida quyidagi hisob-kitoblar ifodalaniladi:

```

. scalar list sseu sser J df_unrest fstat pvalue crit tratio t_sq
      sseu = 1532.0845
      sser = 2683.4109
      J = 1
df_unrest = 71
      fstat = 53.354875
      pvalue = 3.236e-10
      crit = 3.9758102
      tratio = -7.3044421
      t_sq = 53.354875

```

F-statistika 53,35 ni tashkil qiladi. t-nisbatining kvadrati F ga teng ekanligi bejiz emas: $-7,304^2 = 53,35$. Buning sababi t- va F-tarqatishlar o'rtasidagi aniq bog'liqlikdir.

Erkinlik darajasi df bo'lgan t tasodifiy o'zgaruvchining kvadrati F tasodifiy o'zgaruvchi bo'lib, hisoblagichda erkinlik darajasi 1 va maxrajda erkinlik darajasi df bo'ladi.

6.1.3. Qo'shimcha umumiy F-testlari

F-testi regressiya koeffitsientlari bo'yicha nol cheklovlarni o'z ichiga olgan gipotezalarga qaraganda umumiyroq gipotezalarni tekshirish uchun ham ishlatilishi mumkin. Teng belgilarga ega chiziqli gipotezalarni o'z ichiga olgan K gacha farazlarni tekshirish mumkin. Test xuddi shu tarzda kvadrat xatolarning cheklangan yig'indisini uning cheklanmagan qiymatiga solishtirish orqali amalga oshiriladi. Buning uchun foydalanuvchi tomonidan ba'zi algebrai amallar talab qilinadi.

Ko'rib chiqilgan misol avval 5-bobda ko'rib chiqilgan reklamaning optimal darajasiga asoslangan. Agar reklamaga qaytishlar kamaysa, reklamaga sarflangan keyingi dollar sotishning faqat yana bir dollarini hosil qilganda reklamaning optimal darajasi yuzaga keladi. Boshqa (ming) dollarning 1 ga teng savdodagi marginal ta'sirini belgilash:

$$\beta_3 + 2\beta_4 A_0 = 1$$

va A_0 uchun yechish natijasida $\hat{A}_0 = (1 - b_3)/2b_4$ hosil bo'ladi, bunda b_3 va b_4 eng kichik kvadratlar orqali olingan natijalardir. Hisoblangan model natijalarini kiritish reklamaning taxminiy optimal darajasini 2,014 (\$2014) tashkil qiladi. Aytaylik, Andy reklamaning optimal darajasi 1900-dollar degan taxminni sinab ko'rmoqchi 1.9 ni almashtirish (esda tutingki, ma'lumotlardagi reklama 1000-dollar bilan o'lchanadi) nol va muqobil farazlarga olib keladi:

$$H_0: \beta_3 + 3.8\beta_4 = 1 \quad H_1: \beta_3 + 3.8\beta_4 \neq 1$$

Stata bu taxminning qiymatini nol gipoteza va uning standart xatosi ostida hisoblashni quyidagi buyruq orqali amalga oshiradi.

lincom _b[advert]+3.8*_b[c.advert#c.advert]-1

Oldingi boblardan **lincomm** buyrug'i o'zidan oldingi regressiya asosida parametrlarning chiziqli birikmalarini topgandik va shu buyruq asosida natijalarni olamiz. **Lincom** dan olingan natija va t-nisbatini hisoblash:

(1) `advert + 3.8*c.advert#c.advert = 1`

sales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
(1)	.6329759	.6541901	0.97	0.337	-.6714421 1.937394

Regressiya chiziqli bo'lgani uchun oddiyroq sintaksisdan bir xil natijalarni olish uchun ham foydalanish mumkin:

lincom advert+3.8*c.advert#c.advert-1

Ikkala holatda ham taxminiy va standart xato hosil bo'ladi va bu miqdorlar mos ravishda **r(estimate)** va **r(se)**da saqlanadi. Shunday qilib, siz ularni eslab qolishingiz va t-nisbatini qo'lda hisoblash uchun **scalar** buyrug'idan foydalanishingiz mumkin. Buni amalga oshirish uchun quyidagi buyruqlardan foydalanamiz:

scalar t = r(estimate)/r(se)

scalar pvalue2tail = 2*ttail(e(df_r),t)

scalar pvalue1tail = ttail(e(df_r),t)

scalar list t pvalue2tail pvalue1tail

Hisoblangan t-nisbat uchun bir tomonlama p-qiymatini olish uchun **ttail()** buyrug'idan foydalaniladi. Unda **e(df_r)** dan foydalaniladi, bu uni ishlatishdan oldingi savdo regressiyasidan qolgan erkinlik darajasini ifodalaydi.

Natijada:

```
. scalar list t pvalue2tail pvalue1tail
      t = .96757186
pvalue2tail = .33654267
pvalue1tail = .16827134
```

Istalgan cheklovni o'zida mujassam etgan yangi parametr bo'yicha modelni qayta tartiblash imkonini beradigan algebraik formulalarni ishlatish mumkin. Bu **lincom** buyrug'i kabi biror narsani o'z ichiga olmaydigan dasturiy ta'minotdan foydalanilganda foydali bo'ladi.

$$SALES_i - ADVERT_i = \beta_1 + \beta_2 PRICE_i + \theta ADVERT_i + \beta_4 (ADVERT_i^2 - 3.8ADVERT_i) + e_i$$

Reklama koeffitsienti to'liq cheklovni o'z ichiga oladi. Cheklovning to'g'ri yoki yo'qligini tekshirish uchun uning koeffitsienti bo'yicha t-nisbatidan foydalaning. Stata-da ikkita yangi o'zgaruvchi kiritamiz:

$$x_i = ADVERT_i^2 - 3.8ADVERT_i \text{ and } y_i = SALES_i - ADVERT_i$$

gen xstar = c.advert#c.advert-3.8*advert

gen ystar = sales - advert

Kiritilgan yangi o'zgaruvchilardan regressiyada foydalaniladi.

regress ystar price advert xstar

O'zgaruvchan *advert* bo'yicha t-nisbati kerakli statistik qiymatdir. Uning ikki tomonlama p-qiymati natijada berilgan. Agar buni qo'lda hisoblamoqchi bo'lsangiz, quyidagi buyruq orqali amalga oshirish mumkin:

scalar t = (_b[advert])/_se[advert]

scalar pvalue = ttail(e(df_r),t)

scalar list t pvalue

Yuqoridagi buyruqlar uchun quyidagicha natijaga erishiladi:

```
. regress ystar price advert xstar
```

Source	SS	df	MS			
Model	1457.21493	3	485.738311	Number of obs =	75	
Residual	1532.08447	71	21.5786545	F(3, 71) =	22.51	
Total	2989.2994	74	40.3959379	Prob > F =	0.0000	
				R-squared =	0.4875	
				Adj R-squared =	0.4658	
				Root MSE =	4.6453	

ystar	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
price	-7.64	1.045939	-7.30	0.000	-9.725542	-5.554457
advert	.632976	.6541901	0.97	0.337	-.671442	1.937394
xstar	-2.767964	.9406241	-2.94	0.004	-4.643515	-.8924125
_cons	109.719	6.799046	16.14	0.000	96.16212	123.2759


```
. scalar t = (_b[advert])/_se[advert]
. scalar pvalue = ttail(e(df_r),t)
. scalar list t pvalue
      t = .96757201
      pvalue = .1682713
```

Regressiya jadvalidagi *advert* o'zgaruvchining t-nisbati 0,97 ga va ikki tomonlama p-qiymati 0,337 ga teng. Skalyar buyrug'i yordamida hisoblangan t-nisbat bir xil (ko'proq raqamlarga o'tkazilgan bo'lsa ham) va uning bir tomonlama p-qiymati jadvaldagi ikki tomonlama nisbatning yarmiga teng. Natijalar bevosita mos keladi.

Ushbu bo'lim Big Andyning ikkita taxminini birgalikda sinovdan o'tkazish bilan yakunlanadi. Oylik reklama xarajatlarning optimal darajasi 1900 AQSh dollarini tashkil etishini taklif qilishdan tashqari, Big Andy PRICE \$6 narxi va ADVERT 1.9 reklama xarajatlari o'rtacha hisobda sotishdan tushgan daromad keltiradi degan faraz asosida xodimlarni ta'minlash va 80 000 AQSh dollar sotuvda olish rejalashtirmoqda. Qo'shma nol gipoteza:

$$H_0: \beta_3 + 3.8\beta_4 = 1 \quad va \quad \beta_1 + 6\beta_2 + 1.9\beta_3 + 3.61\beta_4 = 80$$

Bunga qarshi alternativdan farazlarning bittasi to'g'ri emas deb xulosa qilinadi. Qo'shmcha test uchun Stata quyidagi buyruqdan foydalanamiz:

regress sales price advert c.advert#c.advert

*test (_b[advert]+3.8*_b[c.advert#c.advert]=1) (_b[_cons] + ///*

*6*_b[price] + 1.9*_b[advert]+3.61*_b[c.advert#c.advert]= 80)*

Bu misolda *test* buyrug'i qo'llaniladi, undan so'ng har ikkala cheklov qo'yiladi, ularning har biri alohida qavslar to'plamida joylashgan. E'tibor bering, *test* oldingi

regressiyadan saqlangan **_b[varname]** koeffitsientlari ishlatiladi. Yana bir bor, buni faqat o'zgaruvchilar nomlaridan foydalangan holda chiziqli regressiyada soddalashtirish mumkin.
test (advert+3.8*c.advert#c.advert=1) (_cons + 6*price + ///
1.9*advert+3.61*c.advert#c.advert= 80)

Natijalar:

```
. test (_b[advert]+3.8*_b[c.advert#c.advert]=1) ///
>      (_b[_cons]+6*_b[price]+1.9*_b[advert]+3.61*_b[c.advert#c.advert]= 80)

( 1)  advert + 3.8*c.advert#c.advert = 1
( 2)  6*price + 1.9*advert + 3.61*c.advert#c.advert + _cons = 80

      F( 2,      71) =    5.74
      Prob > F =    0.0049
```

p-qiymati 0,0049 va 5% dan kam bo'lganligi sababli, ushbu muhimlik darajasida nol (qo'shma) gipoteza rad etiladi.

6.2. Namuna bo'lmagan ma'lumotlar orqali model tuzish

Ba'zan sizda modelni baholashda foydalanmoqchi bo'lgan aniq namuna bo'lmagan ma'lumotlar mavjud. Namuna bo'lmagan ma'lumotlardan foydalanish qolgan parametrlarni baholashning aniqligini oshiradi. *Principle of Econometrics-4th* nashirdagi adabiyotdan olingan ushbu misolda mualliflar pivo sotish modelini pivo narxlari, spirtli ichimliklar narxi, boshqa tovarlar narxi va daromadning funktsiyasi sifatida ko'rib chiqishgan. Shu asnoda ushbu ma'lumotlar asosida modelni tuzib olamiz. O'zgaruvchilar tabiiy logarifmlarida ifodalaymiz:

$$\ln(Q_t) = \beta_1 + \beta_2 \ln(PB_t) + \beta_3 \ln(PL_t) + \beta_4 \ln(PR_t) + \beta_5 \ln(I_t) + e$$

Modelga asosan iqtisodiy nazariya shuni ifodalaydi:

$$\beta_2 + \beta_3 + \beta_4 + \beta_5 = 0$$

Modelni baholash uchun *beer.dta* ma'lumotlar faylidan foydalanamiz. Ma'lumotlar faylini quyidagi buyruq orqali ochamiz:

use beer, clear

Keyin ma'lumotlar to'plami uchun har bir o'zgaruvchining natural logarifmlarini tuzing. Stata ning buyrug'i **log(variable)** o'zgaruvchining natural logarifmini olish uchun ishlatiladi. Shunday qilib, har bir o'zgaruvchining natural logarifimli o'zgaruvchilarini yaratish uchun foydalaning:

use beer, clear

gen lq = ln(q)

gen lpb = ln(pb)

gen lpl = ln(pl)

gen lpr = ln(pr)

gen li = ln(i)

Chiziqli cheklovlarni o'rnatish uchun siz Stata **constrained regression** deb ataladigan buyruqdan foydalanasiz. Stata cheklashni **constraint** deb ataladi va chiziqli regressiya

modeliga ushbu cheklovlarni qo'yish uchun foydalanadigan buyruq **cnsreg** hisoblanadi. Sintaksi quyidagicha ifodalanadi:

constraint 1

constraint 2

cnsreg depvar indepvars [if] [in] [weight], constraints(1 2)

Cheklovlarning (constraints) har biri birinchi navbatda sanab o'tiladi va ularga noyob raqam beriladi. Natijalar olinganidan keyin **cnsreg** buyrug'i **regress** kabi ishlatiladi; regressiya modeliga vergul bilan amal qiling va cheklash raqamlari ro'yxati **constraints (1 2 ...)** va Stata sanab o'tilgan cheklovlarni qo'yadi hamda qolgan parametrlarni baholash uchun eng kichik kvadratlardan foydalanadi. **Constraints** buyrug'i quyida ko'rsatilgandek **c(1 2)** qisqartirilishi mumkin. Pivo misoli asosida quyidagi sintaksi orqali amalga oshiriladi:

constraint 1 lpb+lpl+lpr+li=0

cnsreg lq lpb lpl lpr li, c(1)

Natijada:

```
. constraint 1 lpb+lpl+lpr+li=0
. cnsreg lq lpb lpl lpr li, c(1)
```

Constrained linear regression

Number of obs	=	30
F(3, 26)	=	36.46
Prob > F	=	0.0000
ROOT MSE	=	0.0617

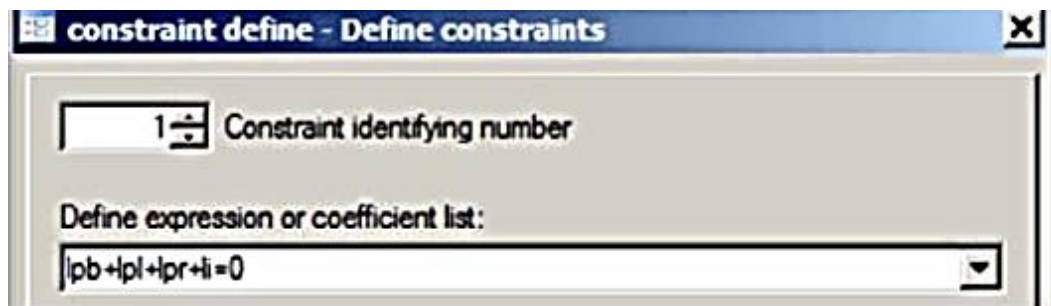
(1) lpb + lpl + lpr + li = 0

	lq	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
	lpb	-1.299387	.1657377	-7.84	0.000	-1.640065 - .958708
	lpl	.1868161	.2843833	0.66	0.517	-.3977422 .7713744
	lpr	.1667424	.0770752	2.16	0.040	.0083121 .3251727
	li	.9458282	.4270468	2.21	0.036	.0680209 1.823635
	_cons	-4.797793	3.713905	-1.29	0.208	-12.43183 2.836247

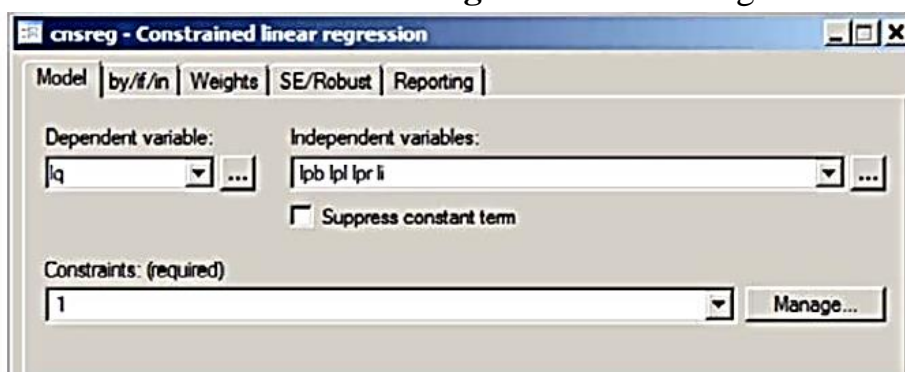
Ko'proq harakat talab qilsa ham, ushbu natijalarni olish uchun ochiladigan menyulardan ham foydalanish mumkin. Birinchidan, cheklovni aniqlash kerak. Ketma-ketlik: **Statistics > Other > Manage Constraints**



Raqamlash va cheklovlarni aniqlash uchun ishlatiladigan dialog oynasini ochish uchun **Create** tugmasini bosing.



Cheklov raqamini tanlang va **Define expression** maydoniga kerakli cheklovni kiriting. Cheklovni qabul qilish va qutini yopish uchun **OK** tugmasini bosing. Cheklovlarni qo'shish uchun **constraint—Manage constraints** maydonidagi **Create** tugmasini yana bir marta bosing. Ish tugagach, oynani yopish uchun **Close** tugmasini bosing. Cheklangan modelni baholash uchun ko'rsatilganidek, ochiladigan menyudan **Statistics > Linear models and related > Constrained linear regression**-ni tanlang:



Cheklangan modelni baholash uchun **OK** yoki **Submit** tugmasini bosing.

6.3. Model xususiyatlari

Modelni tanlashning uchta asosiy xususiyati quyidagilardan iborat: (1) funktsional shaklni tanlash, (2) modelga kiritiladigan erkli o'zgaruvchilarni (regressorlar) tanlash va (3) 5-bobda sanab o'tilgan ko'plik regressiya modelining MR1-MR6 farazlari tanlash. Ushbu bo'lim ularning dastlabki ikkitasini o'rganadi.

6.3.1. Yo'qolgan o'zgaruvchilar

Agar siz modelingizdagi tegishli o'zgaruvchilarni o'tkazib yuborsangiz, eng kichik kvadratlar usuli noto'g'ri bo'ladi. **Omitted variable problem** bilan bog'liq ko'rsatish uchun biz ishlaydigan er va xotin juftliklar namunasini ko'rib chiqamiz. Ma'lumotlar *edu_inc.dta* faylida saqlanadi.

Ma'lumotlar faylini oching va Stata xotirasidan avval saqlangan ma'lumotlarni o'chiring.

use edu_inc, clear

Birinchi regressiya oila daromadini bog'liq o'zgaruvchi sifatida (**famine**) va tushuntirish o'zgaruvchisi sifatida erning (**he**) va xotinning ma'lumotini (**we**) o'z ichiga oladi. Buyruqlar qatoridan:

regress faminc he we

Natijada:

. regress faminc he we

Source	SS	df	MS			
Model	1.3405e+11	2	6.7027e+10	Number of obs =	428	
Residual	6.9703e+11	425	1.6401e+09	F(2, 425) =	40.87	
Total	8.3109e+11	427	1.9463e+09	Prob > F =	0.0000	
				R-squared =	0.1613	
				Adj R-squared =	0.1574	
				Root MSE =	40498	

faminc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
he	3131.509	802.908	3.90	0.000	1553.344	4709.674
we	4522.641	1066.327	4.24	0.000	2426.711	6618.572
_cons	-5533.629	11229.53	-0.49	0.622	-27605.97	16538.71

Xotinning ma'lumotsizligi (**we**) quyidagilarni beradi:

. regress faminc he

Source	SS	df	MS			
Model	1.0455e+11	1	1.0455e+11	Number of obs =	428	
Residual	7.2654e+11	426	1.7055e+09	F(1, 426) =	61.30	
Total	8.3109e+11	427	1.9463e+09	Prob > F =	0.0000	
				R-squared =	0.1258	
				Adj R-squared =	0.1237	
				Root MSE =	41297	

faminc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
he	5155.483	658.4574	7.83	0.000	3861.254	6449.713
_cons	26191.27	8541.108	3.07	0.002	9403.309	42979.23

Oddiy korrelyatsiya tahlili shuni ko'rsatadiki, er va xotinning ta'lim darajasi ijobiy bog'liqdir. Matnda taklif qilinganidek, bu **we** modeldan chiqarib tashlash **he** koeffitsientida ijobiy natija olib kelishi mumkinligini anglatadi. Bu baholash modellar tomonidan qo'llab-quvvatlanadi.

. correlate
(obs=428)

	faminc	he	we	kl6	xtra_x5	xtra_x6
faminc	1.0000					
he	0.3547	1.0000				
we	0.3623	0.5943	1.0000			
kl6	-0.0720	0.1049	0.1293	1.0000		
xtra_x5	0.2898	0.8362	0.5178	0.1487	1.0000	
xtra_x6	0.3514	0.8206	0.7993	0.1595	0.9002	1.0000

Xotinning ma'lumoti va maktabgacha yoshdagi bolalar soni (kl6) hisobga olinsa, quyidagi natijalar hosil bo'ladi:

. regress faminc he we kl6

Source	SS	df	MS			
Model	1.4725e+11	3	4.9082e+10	Number of obs =	428	
Residual	6.8384e+11	424	1.6128e+09	F(3, 424) =	30.43	
Total	8.3109e+11	427	1.9463e+09	Prob > F =	0.0000	
				R-squared =	0.1772	
				Adj R-squared =	0.1714	
				Root MSE =	40160	

faminc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
he	3211.526	796.7026	4.03	0.000	1645.547	4777.504
we	4776.907	1061.164	4.50	0.000	2691.111	6862.704
kl6	-14310.92	5003.928	-2.86	0.004	-24146.52	-4475.326
_cons	-7755.33	11162.93	-0.69	0.488	-29696.91	14186.25

E'tibor bering, *we* va *he* uchun koeffitsientlarni baholash avvalgi regressiyaga nisbatan unchalik o'zgaradi. Buning sababi shundaki, *kl6* hech qanday shakllanish o'zgaruvchilari bilan kuchli bog'liq emas. Bu shuni anglatadiki, tegishli o'zgaruvchi ko'rsatilmagan taqdirda ham foydali natijalarga erishish mumkin. O'tkazib yuborilgan o'zgaruvchining ushbu misolda ta'lim o'zgaruvchilari bo'lgan qiziqish o'zgaruvchilari bilan bog'liq bo'lmasligi talab qilinadi. Bunday holda, tegishli o'zgaruvchini o'tkazib yuborish testlarning haqiqiyligiga va *we* yoki *he* ishtirok etadigan ishonch oralig'iga ta'sir qilmaydi.

6.3.2. Ahamiyatsiz o'zgaruvchilar

Modelga ahamiyatsiz o'zgaruvchilarni kiritish eng kichik kvadratlarni baholashning aniqligini pasaytiradi. Eng kichik kvadratlar usuli xolis, ammo koeffitsientlarning standart xatolari zarur bo'lganidan kattaroq bo'ladi. Ushbu misol modelga ikkita ahamiyatsiz o'zgaruvchini qo'shadi (*xtra_x5* va *xtra_x6*). Bu o'zgaruvchilar *he* va *we* bilan bog'liq, lekin oila daromadining o'rtacha qiymati bilan bog'liq emas. Quyidagilarni olish uchun chiziqli regressiya yordamida modelni baholang:

. regress faminc he we kl6 xtra_x5 xtra_x6

Source	SS	df	MS			
Model	1.4776e+11	5	2.9553e+10	Number of obs =	428	
Residual	6.8332e+11	422	1.6192e+09	F(5, 422) =	18.25	
Total	8.3109e+11	427	1.9463e+09	Prob > F =	0.0000	
				R-squared =	0.1778	
				Adj R-squared =	0.1681	
				Root MSE =	40240	

faminc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
he	3339.792	1250.039	2.67	0.008	882.7131	5796.871
we	5868.677	2278.067	2.58	0.010	1390.905	10346.45
kl6	-14200.18	5043.72	-2.82	0.005	-24114.13	-4286.242
xtra_x5	888.8426	2242.491	0.40	0.692	-3519.001	5296.686
xtra_x6	-1067.186	1981.685	-0.54	0.590	-4962.389	2828.018
_cons	-7558.613	11195.41	-0.68	0.500	-29564.33	14447.1

Hisoblangan standart xatolar oldingi regressiya xatolariga nisbatan qanchalik ko'payganiga e'tibor bering. Agar ular *he* va *we* bilan bog'lanmagan bo'lsa, biz ularning standart xatolariga juda kam ta'sir ko'rsatishini kutgan bo'lardik.

6.3.3. Model tanlash

To'g'ri o'zgaruvchilar to'plamini va to'g'ri funktsional shaklni tanlash nafaqat fan, balki san'at hamdir. Ideal holda, siz qiziqtirgan parametrlarni izchil va samarali baholash imkonini beruvchi ma'lumotlar va o'zgaruvchilarning tegishli xususiyatlarini qamrab oladigan funktsional shaklni tanlashni xohlaysiz.

Ushbu bo'limda ko'pincha o'zgaruvchilarni vaqtincha tanlash va funktsional shaklning muvofiqligini tekshirish uchun foydalaniladigan statistik ma'lumotlar muhokama qilinadi.

Modelni tanlash mezonlari

Modelni tanlash uchun uchta mezon ko'rib chiqiladi: hisoblangan R^2 , AIC va SC (BIC). Ushbu statistika muqobil modellarni tanlashda foydali bo'lishi mumkin, garchi ulardan foydalanish munozaralidir. Qanday bo'lmasin, ular faqat model spetsifikatsiyasining barcha boshqa manbalari tugaganidan keyin ishlatilishi kerak. Ya'ni, iloji boricha nazariya va aql-idrokdan foydalaning va model tanlash qoidalariga faqat muqobil modellarning nisbiy afzalliklari haqida ko'proq ma'lumot olish uchun murojaat qiling. Ushbu ogohlantirishga e'tibor bermay, davom etaylik.

5-bobda hisoblangan R^2 eng kichik kvadratlarga moslashishning muqobil o'lchovi sifatida taqdim etilgan bo'lib, u oddiy R^2 bilan mashhur muammoni yengib chiqadi, ya'ni modelga regressorlar qo'shilganda u hech qachon kichraymaydi. Hisoblangan R^2 regressor qo'shilganligi sababli tegishli salbiy natijani belgilaydi. Agar moslashuvning yaxshilanishi salbiy natija bilan solishtirganda juda kichik bo'lsa, modelga ahamiyatsiz regressor qo'shilganda hisoblangan R^2 kichikroq bo'lishi mumkin. Hisoblangan R^2 quyidagicha:

$$\bar{R}^2 = 1 - \frac{SSE/(N - K)}{SST/(N - 1)}$$

Ushbu statistika ma'lumot bo'yicha **Regress** Stata buyrug'i yordamida qayta hisoblaydi. Ko'rib chiqilgan boshqa model tanlash qoidalari **Akaike information criterion** (AIC) tomonidan berilgan.

$$AIC = \ln\left(\frac{SSE}{N}\right) + 2\frac{K}{N}$$

va **Bayesian information criterion** (SC) quyidagi formula bilan berilgan.

$$SC = \ln\left(\frac{SSE}{N}\right) + 2\frac{K \ln(N)}{N}$$

Bu ikki statistika juda o'xshash va ikkita atamadan iborat. Birinchisi, muvofiqlik o'lchovidir; moslik qanchalik yaxshi bo'lsa, SSE shunchalik past bo'ladi va uning natural logarifmi shunchalik past bo'ladi. Regressor qo'shish bu formulaning hajmini oshirib bo'lmaydi. Ikkinchisi, regressorni qo'shish mezoniga nisbatan qo'llaniladigan salbiy natija. K qanchalik katta bo'lsa, salbiy natija shunchalik katta bo'ladi. Modellar orasida AIC yoki SC ni minimallashtiradigan modelni tanlash g'oyasidir. Ular faqat salbiy natija hajmida farqlanadi, SC esa biroz kattaroqdir.

Ushbu mezonlar Stata-da mavjud, ammo ular boshqacha hisoblanadi. Stata versiyalari bu yerda muhokama qilinganidan ko'ra kengroq ma'lumotlarni ishlab chiqarish

jarayonlarida foydalanish uchun mo'ljallangan, shuning uchun agar kerak bo'lsa, ulardan foydalanishni unutmang¹.

Ushbu mezonlar "Principles of Econometrics", 4-nashr adabiyotida qayta-qayta qo'llaniladi va ushbu qo'llanmaning maqsadlaridan biri ularning natijalarini takrorlashdir. Shuning uchun, uchta model tanlash qoidalarini hisoblash va ko'rsatish uchun **dastur** yozish tavsiya etiladi; Yozilgandan so'ng, dastur modelning turli xususiyatlarini solishtirish uchun bir necha marta ishga tushirilishi mumkin. 9-bobda model tanlash dasturi qayta ko'rib chiqilgan va dasturlash sikllarida qo'llanilgan.

Stata-da dastur bu shunchaki dastur nomini kiritish orqali kod bloklarini bajarish imkonini beruvchi tuzilmadir. Quyidagi misolda **modelsel** deb nomlangan dastur yaratilgan. Buyruqlar oynasida **modelsel** har gal kiritilganda dastur ichidagi kod satrlari bajariladi. Bunday holda, dastur AIC, SC ni hisoblab chiqadi va sozlangan R^2 qiymatini chop etadi, barchasi avval amalga oshirilgan regressiyaga asoslanadi.

Stata-da dasturlash quyidagicha ishlaydi. Dasturda dastur buyrug'ini berish va unga nom berishdan boshlanadi, masalan, **progrname**. Keyin dastur har safar ishga tushirilganda bajariladigan Stata buyruqlar bloki yoziladi. Dastur oxirida yopiladi. Quyidagi asosiy tuzilma hosil bo'ladi:

```
program progrname
```

```
Stata commands
```

```
end
```

Dasturni yozgandan so'ng, uni kompilyatsiya qilishingiz kerak. Agar dastur alohida do-faylida bo'lsa, oddiy usulda do-faylini ishga tushiring. Agar dastur do-faylidagi boshqa kod bilan birga bo'lsa, kodni ajratib ko'rsating va ma'lumotni odatdagi tarzda ishga tushiring. Dasturni faqat bir marta kompilyatsiya qilish kerak. Dastur Stata so'roviga nuqta bilan dastur nomini, **dastur** nomini kiritish orqali boshlanadi.

Modelselsel dasturi:

```
program modelsel
```

```
scalar aic = ln(e(rss)/e(N))+2*e(rank)/e(N)
```

```
scalar bic = ln(e(rss)/e(N))+e(rank)*ln(e(N))/e(N)
```

```
di "r-square = "e(r2) " and adjusted r_square " e(r2_a)
```

```
scalar list aic bic
```

```
end
```

Stata jarayonni tugatmaguningizcha yoki Stata-ga dasturni xotiradan olib tashlashni aytmaguningizcha dastur xotirada qoladi. Bu ikki yo'l bilan amalga oshiriladi. Dastur nomini tushirish dasturi avval berilgan dasturni (**ya'ni dastur nomini**) xotiradan olib tashlaydi. Yana bir usul - **drop_all programm** yordamida barcha dasturlarni xotiradan olib

¹ Darhaqiqat, Stata ning baholashdan keyingi buyrug'i $estat ic AIC = -2 \ln(L) + 2k$ va $BIC = -2 \ln(L) + k \ln(N)$ dan foydalanadi, bu yerda L - model xatolari normal taqsimlanganda maksimal ehtimollik funksiyasining qiymati.

tashlashdir. Ushbu usuldan faqat Stata xotirasidan barcha foydalanuvchi dasturlarini olib tashlamoqchi bo'lsangiz foydalaning.

Ushbu maxsus dastur Stata regressiyadan so'ng yaratadigan va saqlaydigan natijalardan foydalanadi. Ulardan ba'zilari sizga allaqachon tanish bo'ladi. **e(rss)** kvadrat xatolar yig'indisini o'z ichiga oladi va **e(N)** - namuna hajmi. Ishlatilgan yangi natija **e(rank)** bo'lib, u asosan modelda qancha erkli o'zgaruvchilar borligini o'lchaydi, boshqalar bilan to'liq mos keladiganlar bundan mustasno. Aniqlangan regressiya modelida bu odatda modeldagi koeffitsientlar sonini o'lchaydi, K.

Dasturning jarayonida **aic** va **bic** skayarlari (ba'zan SC - Shvarts mezoni deb ataladi) hisoblab chiqiladi va modeldagi hisoblangan R^2 qiymatini chop etish uchun display buyrug'i beriladi. Nihoyat, hisoblangan **scalar listni** chop etish uchun skalyar ro'yxat buyrug'i beriladi.

Modelni baholash va undan olingan modelni tanlash qoidalarini hisoblash uchun, agar siz hali buni qilmagan bo'lsangiz, **modelsel** dasturini ishga tushiring. Keyin regressiyani baholang va **sel modelini** kiriting. Masalan:

quietly regress faminc he

di "Model 1 (he) "

modelsel

estimates store Model1

Natijada:

```
. di "Model 1 (he) "  
Model 1 (he)  
  
. modelsel  
r-square = .12580103 and adjusted r_square .12374892  
    aic = 21.261776  
    bic = 21.280744
```

Model tanlash qoidalaridan foydalanish uchun har bir modeldan keyin **modelselni** ishga tushiring va eng katta hisoblangan R^2 (odatda yomon fikr) yoki eng kichik AIC yoki BIC (yaxshiroq, lekin ajoyib g'oya emas) bo'lganini tanlang.

quietly regress faminc he we

di "Model 2 (he, we) "

modelsel

estimates store Model2

quietly regress faminc he we kl6

di "Model 3 (he, we, kl6) "

modelsel

estimates store Model3

quietly regress faminc he we kl6 xtra_x5 xtra_x6

di "Model 4 (he, we, kl6. x5, x6) "

modelsel

estimates store Model4

Natijada:

```
. di "Model 2 (he, we) "  
Model 2 (he, we)  
  
. modelsel  
r-square = .16130044 and adjusted r_square .15735362  
    aic = 21.224993  
    bic = 21.253445  
  
. di "Model 3 (he, we, k16) "  
Model 3 (he, we, k16)  
  
. modelsel  
r-square = .17717332 and adjusted r_square .17135143  
    aic = 21.210559  
    bic = 21.248495  
  
. di "Model 4 (he, we, k16. x5, x6) "  
Model 4 (he, we, k16. x5, x6)  
  
. modelsel  
r-square = .17779646 and adjusted r_square .16805472  
    aic = 21.219148  
    bic = 21.276051
```

Bu misolda har bir modeldan keyin Stata ni **estimates store** buyrug'i chiqariladi va natijalar **estimates table** buyrug'i yordamida to'planadi.

*estimates table Model1 Model2 Model3 Model4, b(%9.3f) stfmt(%9.3f) ///
se stats(N r2 r2_a aic bic)*

variable	Model1	Model2	Model3	Model4
he	5155.483	3131.509	3211.526	3339.792
	658.457	802.908	796.703	1250.039
we		4522.641	4776.907	5868.677
		1066.327	1061.164	2278.067
k16			-1.43e+04	-1.42e+04
			5003.928	5043.720
xtra_x5				888.843
				2242.491
xtra_x6				-1067.186
				1981.685
_cons	26191.270	-5533.629	-7755.330	-7558.613
	8541.108	11229.533	11162.935	11195.411
N	428	428	428	428
r2	0.126	0.161	0.177	0.178
r2_a	0.124	0.157	0.171	0.168
aic	10314.652	10298.909	10292.731	10296.407
bic	10322.770	10311.086	10308.967	10320.761

Legend: b/se

Stata tomonidan yaratilgan ushbu jadval har bir regressiya uchun hisoblangan **aic** va **bic** statistikasing Stata versiyalaridan foydalanadi. Shubhasiz, Stata boshqa hisoblashdan foydalanadi! Xavotir olmang, ikkalasi ham hisoblash to'plamlari haqiqiy va bir xil xulosaga olib keladi. Eng katta R^2 , eng kichik **aic** va **bic** statistikasi kabi 3-modeldan olingan. Ushbu misolda 3-model afzal ko'rilgan spetsifikatsiya ekanligi tushuniladi.

Funksional shakl

Funksional shaklni tanlashda nazariy mulohazalar asosiy qo'llanma bo'lishi kerak bo'lsa-da, ko'p hollarda iqtisodiy nazariya yoki sog'lom fikr yetarli emas. Bu yerda RESET testi yordam beradi. RESET funksional shaklni belgilashda aniq xatoga yo'l qo'yganingizni

aniqlash uchun qo'pol tekshirish sifatida ishlatilishi mumkin. Bu aslida yetishmayotgan o'zgaruvchilar uchun test EMAS; Buning o'rniga, bu sizning funktsional shaklingizning mosligini tekshiramiz.

Sinov oddiy. Nol gipoteza shundaki, sizning funktsional shaklingiz adekvat; muqobil - bunday emas. Funktsional shaklni to'g'ri deb hisoblab, regressiyani baholang va taxmin qilingan qiymatlarni oling. Ularni kvadratga aylantiring va kubga aylantiring, ularni yana modelga qo'shing, regressiyani qayta baholang va \hat{y}^2 va \hat{y}^3 ning ahamiyatini birgalikda sinab ko'ring.

Aslida, ushbu testning bir nechta variantlari mavjud. Birinchisi modelga faqat \hat{y}^2 qo'shadi va uning ahamiyatini F-testi yoki unga tenglashtirilgan t-testi yordamida tekshiradi. Ikkinchisi \hat{y}^2 va \hat{y}^3 ni qo'shadi va keyin ularning ahamiyatini birgalikda tekshirishni amalga oshiradi. Biz ularni mos ravishda RESET (1) va RESET (2) deb ataymiz.

Misol yana oila daromadining regressiyasiga asoslangan. Modelni eng kichik kvadratlar yordamida baholang va chiziqli hisoblangan qiymatlarni regressiyada saqlash uchun **predict** buyrug'idan foydalaning.

```
regress faminc he we kl6
```

```
predict yhat
```

Eslatib o'tamiz, regressiya namunasida **predict** qiymatlarni olish sintaksisi, \hat{y}_i , **yhat**, **xb** ni taxmin qiladi. Ushbu buyruqda **yhat** siz tayinlayotgan nomdir. Biz **xb** parametrini xavfsiz o'tkazib yuborishimiz mumkin, chunki bu standart Stata sozlamasidir. Endi \hat{y}_i yordamida kvadrat va kublarni hosil qiling.

```
gen yhat2 = yhat^2
```

```
gen yhat3 = yhat^3
```

Modelga qo'shilgan **yhat2** bilan dastlabki regressiyani taxmin qiling. **Yhat2** ning ahamiyatini t-test yoki F-test yordamida tekshiring. Ikkinchisi uchun ko'rsatilgandek Stata **test** buyrug'idan foydalaning.

```
regress faminc he we kl6 yhat2
```

```
test yhat2
```

Test natijasi:

```
. test yhat2
( 1) yhat2 = 0
    Constraint 1 dropped

F( 0, 423) = .
Prob > F = .
```

Shubhasiz, bu iborada muammo bor. Stata cheklov olib tashlanganligini va tekshirish uchun hech narsa qoldirmagani haqida bizga xabar beradi! Muammo shundaki, ma'lumotlar **noto'g'ri**. Kompyuter arifmetik amallarni bajarishi uchun ma'lumotlar to'plamidagi o'zgaruvchilar bir xil qiymatga ega bo'lishi kerak. Modeldagi o'zgaruvchilar uchun umumiy statistik ma'lumotlarni ko'rib chiqing.

. summarize faminc he we kl6

Variable	Obs	Mean	Std. Dev.	Min	Max
faminc	428	91213	44117.35	9072	344146.3
he	428	12.61215	3.035163	4	17
we	428	12.65888	2.285376	5	17
kl6	428	.1401869	.3919231	0	2

faminc darajasi boshqa o'zgaruvchilarga qaraganda 1000 baravar ko'p. Chiziqli regressiya bashoratlari ham xuddi shunday masshtabga ega bo'ladi. Eslatib o'tamiz, chiziqli regressiyada bog'liq va mustaqil o'zgaruvchilar shkalasining o'zgarishi faqat koeffitsientlarning qiymatlariga ta'sir qiladi, ammo sezilarli regressiya natijalari emas. Shunday qilib, ma'lumotlardan yomon shartli bashoratlarni olib tashlang va **faminc** ko'lamini 10 000 ga bo'lish orqali o'zgartiring.

drop yhat yhat2 yhat3

gen faminc_sc = faminc/10000

Endi modelni baholaymiz, hisoblangan qiymatlarni saqlang va kvadrat hamda kublarni yarating.

regress faminc_sc he we kl6

. regress faminc_sc he we kl6

Source	SS	df	MS			
Model	1472.46499	3	490.821663	Number of obs =	428	
Residual	6838.40844	424	16.1283218	F(3, 424) =	30.43	
Total	8310.87343	427	19.4634038	Prob > F =	0.0000	
				R-squared =	0.1772	
				Adj R-squared =	0.1714	
				Root MSE =	4.016	

faminc_sc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
he	.3211526	.0796703	4.03	0.000	.1645547	.4777504
we	.4776908	.1061164	4.50	0.000	.2691111	.6862704
kl6	-1.431092	.5003928	-2.86	0.004	-2.414652	-.4475326
_cons	-.775533	1.116293	-0.69	0.488	-2.969691	1.418625

predict yhat

gen yhat2 = yhat^2

gen yhat3 = yhat^3

RESET(1) uchun modelga **yhat2** qo'shing va uning t-testi yoki F-testi yordamida uning ahamiyatini tekshiring.

```
. regress faminc_sc he we k16 yhat2
```

Source	SS	df	MS			
Model	1567.85524	4	391.963811	Number of obs =	428	
Residual	6743.01819	423	15.9409413	F(4, 423) =	24.59	
Total	8310.87343	427	19.4634038	Prob > F =	0.0000	
				R-squared =	0.1887	
				Adj R-squared =	0.1810	
				Root MSE =	3.9926	

faminc_sc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
he	-.2381465	.2419692	-0.98	0.326	-.7137582	.2374653
we	-.4235108	.383214	-1.11	0.270	-1.176752	.32973
k16	1.088733	1.143928	0.95	0.342	-1.159757	3.337224
yhat2	.099368	.0406211	2.45	0.015	.0195237	.1792123
_cons	8.724297	4.03894	2.16	0.031	.785406	16.66319

```
. test yhat2
( 1) yhat2 = 0
F( 1, 423) = 5.98
Prob > F = 0.0148
```

Yana bir bor, kvadrat t-nisbati F-statistikaga teng va ular bir xil p-qiymatiga ega. RESET(2) uchun **yhat3** ni qo'shing va kvadrat hamda kubik bashoratlarning umumiy ahamiyatini tekshiring:

```
. regress faminc_sc he we k16 yhat2 yhat3
```

Source	SS	df	MS			
Model	1572.19036	5	314.438072	Number of obs =	428	
Residual	6738.68307	422	15.9684433	F(5, 422) =	19.69	
Total	8310.87343	427	19.4634038	Prob > F =	0.0000	
				R-squared =	0.1892	
				Adj R-squared =	0.1796	
				Root MSE =	3.9961	

faminc_sc	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
he	-.8451478	1.189891	-0.71	0.478	-3.184	1.493704
we	-1.301625	1.72841	-0.75	0.452	-4.698991	2.095741
k16	3.741007	5.217535	0.72	0.474	-6.514588	13.9966
yhat2	.3234728	.4320297	0.75	0.454	-.5257254	1.172671
yhat3	-.0085692	.0164465	-0.52	0.603	-.0408965	.023758
_cons	15.01857	12.73868	1.18	0.239	-10.0206	40.05774

```
. test yhat2 yhat3
( 1) yhat2 = 0
( 2) yhat3 = 0
F( 2, 422) = 3.12
Prob > F = 0.0451
```

RESET(1) va RESET(2) ikkalasi ham 5% darajasida muhim ahamiyatga ega va asl chiziqli funktsional shakl ushbu munosabatlarni modellashtirish uchun mos emas degan xulosaga kelish mumkin.

Stata regressiyadan so'ng RESET(3) testini bajaradigan keyingi baholash buyrug'ini o'z ichiga oladi. Sintaksi quyidagicha:

```
regress faminc he we k16
estat ovtest
```

```
. estat ovtest
```

```
Ramsey RESET test using powers of the fitted values of faminc
Ho: model has no omitted variables
F(3, 421) = 2.15
Prob > F = 0.0931
```

RESET ning ushbu versiyasi modelga \hat{y}^2 , \hat{y}^3 va \hat{y}^4 qo'shadi va ularning umumiy ahamiyatini tekshiradi. Texnik jihatdan, buning hech qanday yomon joyi yo'q. Biroq, juda ko'p \hat{y} kuchlarini kiritish ko'pincha tavsiya etilmaydi, chunki RESET \hat{y} kuchlari qo'shilganda statistik quvvatni tezda yo'qotadi.

6.4. Ma'lumotlar yetishmaslik, kolinearlik va ahamiyatsizlik

Oldingi bo'limda biz Stata hisob-kitoblaridan biri ma'lumotlarning yomon ishlashi tufayli muvaffaqiyatsizlikka uchraganini aytib o'tgan edik. Bu kollinearlik regressiya bilan nima qilishiga o'xshaydi. Kollinearlik parametrlarni baholash va boshqa turli statistik ma'lumotlarni katta aniqlik bilan hisoblashni qiyinlashtiradi yoki imkonsiz qiladi. Statistik modelda kollinearlik noto'g'ri eksperimental dizayndan kelib chiqadi yoki bizning holatlarimizda, parametrlarni aniq o'lchash imkonini beradigan ma'lumotlar yetarli darajada farq qilmaydi. Afsuski, buning oddiy davosi yo'q; ma'lumotlarning o'lchamini o'zgartirish uning tarkibidagi chiziqli munosabatlarga ta'sir qilmaydi.

Bu misol *cars.dta* dan foydalanadi. Oldingi barcha ma'lumotlarni xotiradan o'chirish orqali avtomobil ma'lumotlarini yuklab oling.

use cars, clear

Xulosa statistik ma'lumotlarga nazar tashlash (**summarize**) ma'lumotlardagi asosli farqlarni ko'rsatadi.

`. summarize`

Variable	Obs	Mean	Std. Dev.	Min	Max
mpg	392	23.44592	7.805007	9	46.6
cyl	392	5.471939	1.705783	3	8
eng	392	194.412	104.644	68	455
wgt	392	2977.584	849.4026	1613	5140

O'zgaruvchilarning har biri ularning diapazoni va standart og'ishlari bilan o'lchanadigan o'zgarishlarni o'z ichiga oladi. Oddiy korrelyatsiyalar (**corr**) potentsial muammoni ochib beradi.

`. corr`
(obs=392)

	mpg	cyl	eng	wgt
mpg	1.0000			
cyl	-0.7776	1.0000		
eng	-0.8051	0.9508	1.0000	
wgt	-0.8322	0.8975	0.9330	1.0000

E'tibor bering, potentsial tushuntirish o'zgaruvchilari (**cyl**, **eng**, **wgt**) orasida korrelyatsiyalar juda yuqori; eng kichik qiymat **cyl** va **wgt** o'rtasida sodir bo'ladi va deyarli 0,9 ni tashkil qiladi. Ushbu o'zgaruvchilarning har birining gallon uchun milga mustaqil ta'sirini baholash qiyin bo'ladi.

Birinchiidan, dvigateldagi tsilindrlar (**mpg**) soniga bog'liq bo'lgan gallon boshiga milning oddiy modelini taxmin qiling.

regress mpg cyl

. regress mpg cyl

Source	SS	df	MS			
Model	14403.0831	1	14403.0831	Number of obs =	392	
Residual	9415.91039	390	24.14336	F(1, 390) =	596.56	
Total	23818.9935	391	60.9181419	Prob > F =	0.0000	
				R-squared =	0.6047	
				Adj R-squared =	0.6037	
				Root MSE =	4.9136	

mpg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
cyl	-3.558078	.1456755	-24.42	0.000	-3.844486	-3.271671
_cons	42.91551	.8348668	51.40	0.000	41.2741	44.55691

Avtomobil dvigatelining hajmini kub dyuymdagi modelning og'irligiga (**wgt**) qo'shing.

regress mpg cyl eng wgt

. regress mpg cyl eng wgt

Source	SS	df	MS			
Model	16656.4443	3	5552.1481	Number of obs =	392	
Residual	7162.54916	388	18.4601782	F(3, 388) =	300.76	
Total	23818.9935	391	60.9181419	Prob > F =	0.0000	
				R-squared =	0.6993	
				Adj R-squared =	0.6970	
				Root MSE =	4.2965	

mpg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
cyl	-.2677968	.4130673	-0.65	0.517	-1.079927	.5443336
eng	-.012674	.0082501	-1.54	0.125	-.0288944	.0035465
wgt	-.0057079	.0007139	-8.00	0.000	-.0071115	-.0043043
_cons	44.37096	1.480685	29.97	0.000	41.45979	47.28213

Endi bir qator farazlarni sinab ko'ring. Birinchisi **cyl** qiymati uchun, ikkinchisi **eng** qiymati uchun, uchinchisi esa ularning birlashgan qiymati uchun.

test cyl

test eng

test eng cyl

Natijalar:

```
. test cyl
( 1)  cyl = 0
      F( 1, 388) = 0.42
      Prob > F = 0.5172

. test eng
( 1)  eng = 0
      F( 1, 388) = 2.36
      Prob > F = 0.1253

. test eng cyl
( 1)  eng = 0
( 2)  cyl = 0
      F( 2, 388) = 4.30
      Prob > F = 0.0142
```

Aslida, o'zgaruvchilarning hech biri alohida ahamiyatga ega emas, lekin ular birgalikda 5% darajasida muhim ahamiyatga ega. Bu ularning individual ta'sirini aniq o'lchay olmaganligingiz bilan bog'liq bo'lishi mumkin. Oddiy korrelyatsiyalar **cyl**, **eng** va

wgt mustaqil o'zgaruvchilari bir-biri bilan yuqori korrelyatsiya qilishini ko'rsatdi. Bu mustaqil o'zgaruvchilarning har biri boshqalar uchun regressga uchragan bir nechta yordamchi regressiyalarni baholash orqali tasdiqlanishi mumkin.

regress cyl eng wgt

$$\text{scalar } r1 = e(r2)$$

regress eng wgt cyl

$$\text{scalar } r2 = e(r2)$$

regress wgt eng cyl

$$\text{scalar } r3 = e(r2)$$

0,8 dan yuqori bo'lgan R^2 qiymati kuchli kollinearlikni ko'rsatadi, bu barcha o'zgaruvchilarni o'z ichiga olgan model parametrlarini baholashning aniqligiga salbiy ta'sir ko'rsatishi mumkin. Ushbu misolda R^2 0,93, 0,90 va 0,87 bo'lib, 0,8 chegaradan ancha yuqori. Bu avtomobilning gaz yurishiga siljishning individual hissasi va silindrlar sonini farqlash qiyin bo'lishining yana bir tasdig'idir.

```
. scalar list r1 r2 r3
      r1 = .90490236
      r2 = .93665456
      r3 = .87160914
```

Kollinearlikni aniqlash uchun oddiy korrelyatsiyalar o'rniga yordamchi regressiyalardan foydalanishning afzalligi ushbu misolda unchalik aniq emas. Regressiyada ko'plab o'zgaruvchilar mavjud bo'lganda, korrelyatsiyalar bilan kollinearlikni aniqlash qiyin bo'lishi mumkin. Ikkita o'zgaruvchining yuqori korrelyatsiyasi bo'lmasa-da, ba'zi o'zgaruvchilar aniq bo'lmagan chiziqli munosabatlarga ega bo'lishi mumkin. Bunday vaziyatlarda R^2 ni yordamchi ko'p regressiyadan ko'rib chiqish foydaliroqdir.

VI bob mavzularini mustahkamlash uchun savollar

1. F-testi nimaga xizmat qiladi?
2. Nol gipotezani (cheklovlarni) tekshirish uchun qanday formuladan foydalaniladi?
3. F-testning ahamiyatini tekshirish qanday ifodalanadi?
4. t- va F- testlari o'rtasidagi bog'liqlikni tushuntirib bering.
5. Namuna bo'lmagan ma'lumotlar orqali model tuzish ketma ketligini tushuntiring.
6. O'zgaruvchilar tabiiy logarifmlarida ifodalash formulasini ifodalang.
7. Cheklangan modelni baholash uchun qanday buyruqlar ketma ketligi bajariladi?
8. Yo'qolgan o'zgaruvchilar ma'lumotlari qaysi faylda saqlanadi?
9. Akaike information criterion (AIC) qanday amal bajarishga xizmat qiladi?
10. Bayesian information criterion (SC) qanday formula orqali ifodalanadi?
11. RESET testi qanday vazifa bajaradi?