

## 7 Gaia

# Aldagai koalitatiboak

### Aurkibidea

7.1	Sarrera . . . . .	106
7.2	Aldagai koalitatibo bakarra duen eredia . . . . .	106
	Jatorri aldaketa . . . . .	112
	Jatorri eta malda aldaketa . . . . .	114
7.3	Aldagai koalitatibo bi edo gehiago dituen eredia . . . . .	116
7.4	Egitura aldaketa . . . . .	119

## 7.1 Sarrera

Orain arte aldagai koantitatiboak dituzten ereduak besterik ez dugu zehaztu. Hau da, zenbakizko balioak hartzen dituzten aldagaiak erabili ditugu. Hala ere, aldagaiak koalitatiboak ere izan daitezke, hau da, zenbakizko balioak hartzen ez dituztenak, halaber kategoriak, klaseak edo atributuak dituztenak. Besteak beste, banakoen generoa, egoera zibila, arraza, kokapena edo gertakizun historiko baten barnerapena, urtaroak,... aldagai koalitatiboak dira. Adibidez, langileen soldata generoaren funtzioan egon daiteke, banakoen arrazak edo bizilekuak kriminalitate-tasa azaldu dezakete, gerrateek edo beste gertakizun historikoek eragina izan dezakete herrialdeen BPGean, produktu baten salmentak ezberdinak izan daitezke urtaroen arabera, etab.

Nahiz eta aldagai azaldua koantitatiboa bezala mantentzen jarraitu, gai honetan aldagai azaltzaile koalitatiboak edota koantitatiboak dituzten ereduak kontsideratuko ditugu.

Aldagaien kategoriak koantitatiboki neurgarriak ez direnez, zenbakizko aldagaiak sortu behar dira, fikzio-aldagaiak izenekoak. Fikzio-aldagaiek 1 balioa hartzen dute banakoa kategoria konkretu batean barneratua baldin badago eta 0 balioa bestelako kasuetan<sup>1</sup>.

$$D_i = \begin{cases} 1 & \text{i banakoa kategorian baldin badago} \\ 0 & \text{en caso contrario} \end{cases}$$

Gai honetan, estimazioan, koefizienteen interpretazioan eta hipotesien kontrasteetan oinarrituko gara aldagai azaltzaile koalitatiboak dituzten erduentzat. Ramanathaneko (2002) data7-3 eta data7-19 fitxategiak erabiliz, zenbait eredu aztertuko ditugu.

## 7.2 Aldagai koalitatibo bakarra duen ereduak

### Bi kategoriako aldagai koalitatiboa

Aldagai koalitatibo bakarra duen ereduak, kasurik errezena da. Suposa dezagun etxebizitzaren prezioa azaltzeko etxebizitzak igerilekua izatea bakarrik kontuan hartzen dugula<sup>2</sup>. Horretarako, ondoko fikzio-aldagaia definitzen dugu:

$$POOL_i = \begin{cases} 1 & \text{i-etxebizitzak igerilekua badu} \\ 0 & \text{en caso contrario} \end{cases}$$

Ramanathaneko (2002) data7-3 fitxategia irekitzen dugu. Fitxategi honek 14 etxebizitzaren hurrengo datuak ditu: aurreko gaietan erabilitako etxebizitzaren salmenta prezioa (PRICE),

<sup>1</sup>Fikzio-aldagaiek edozein balio har ditzakete, baina koefizienteen interpretazioa errezagoa da 0 eta 1 balioak hartzen baldin badituzte.

<sup>2</sup>Errazteko asmoz, etxebizitzaren prezioari eragiten dioten gainontzeko aldagaiak alde batera uzten ditugu.

oin karratu bizigarriak (SQFT), gela kopurua (BEDRMS) eta komun kopurua (BATHS) eta aurrean definitutako POOL fikzio-aldagaia eransten da.

PRICE eta POOL aldagaiak aukeratu eta beraien balioak behatuz hurrengo balioak ditugu:

7.1 Taula: Pool fikzio-aldagaiaren balioak

Obs	price	pool
1	199,9	1
2	228,0	0
3	235,0	1
4	285,0	0
5	239,0	0
6	293,0	0
7	285,0	0
8	365,0	1
9	295,0	0
10	290,0	0
11	385,0	1
12	505,0	1
13	425,0	0
14	415,0	0

Adibidez, laginaren lehenengo etxebizitzaren prezioa 199900 dolarrekoa da eta igerilekua dauka POOL aldagaiak 1 balioa hartzen duelako. Bigarrenak, berriz, ez dauka igerilekurik (POOL aldagaiak 0 balioa hartzen du) eta bere salmenta prezioa 228000 dolarrekoa da.

Aurreko PRICE aldagaiari dagozkion oinarrizko estatistikoak ateraz, oso erraz lortu daiteke etxebizitzaren batezbesteko salmenta prezioa 317493 dolarrekoa dela:

7.2 Taula: Price aldagaiaren estatistiko nagusiak

Estatistikoen laburpena, 1 - 14 behaketak erabiliz  
'price' aldagaiarentzat (14 behaketa baliagarriak)

Batezbestekoa	317,49
Mediana	291,50
Minimoa	199,90
Maximoa	505,00
Desbiderazio tipikoa	88,498
Aldakuntza Koefizientea (A.K.)	0,27874
Asimetria	0,65346
Kurtosis soberakina	-0,52983

Hala ere, alde batetik, igerilekurik ez duten etxebizitzaren batezbesteko prezioa eta bestaldetik,

igerilekua dutenena kalkulatu daitezke. Horretarako, igerilekua duten etxebizitzaren salmenta prezioa hautatzen dugu, lehendabizi, PRICE aldagaia aukeratzen dugu, gero *Lagina* → *Definitu*, *fikzio-aldagaiaren oinarrituz ...* klikatzen dugu eta POOL aldagaia aukeratu ondoren *Ados* klikatu. Ondoren, oinarritzko estatistikoak lortzen ditugu:

7.3 Taula: Igerilekua duten etxebizitzaren *price* aldagaiaren estatistiko nagusiak

Estatistikoaren laburpena, 1 - 5 behaketak erabiliz  
'price' aldagaiarentzat (5 behaketa baliagarriak)

Batezbestekoa	337,98
Mediana	365,00
Minimoa	199,90
Maximoa	505,00
Desbiderazio tipikoa	122,99
Aldakuntza Koeffizientea (A.K.)	0,36390
Asimetria	0,15896
Kurtosis soberakina	-1,2798

Lagin osoaren tamaina berreskuratzeko *Lagina* → *Berrezarri lagin osoa* klikatu.

Igerilekurik ez duten etxebizitzaren prezioa aukeratzeko, berriz, *Lagina* → *Murriztu honako irizpidean oinarrituz...* klikatu, *pool=0* baldintza sartu eta adostu. Horrela, NOPOOL aldagaia sortzen dugu, zeinak 1 balioa hartzen duen igerilekurik gabeko etxebizitza bada eta 0 balioa bestelako kasuan. PRICE aldagaiaren oinarritzko estatistikoak ondokoak dira:

7.4 Taula: Igerilekua ez duten etxebizitzaren *price* aldagaiaren estatistiko nagusiak

Estatistikoaren laburpena, 1 - 9 behaketak erabiliz  
'price' aldagaiarentzat (9 behaketa baliagarriak)

Batezbestekoa	306,11
Mediana	290,00
Minimoa	228,00
Maximoa	425,00
Desbiderazio tipikoa	68,959
Aldakuntza Koeffizientea (A.K.)	0,22527
Asimetria	0,87575
Kurtosis soberakina	-0,52255

Hortaz, igerilekua duten etxebizitzaren batezbesteko prezioa 337980 dolarrekoa da eta igerilekurik ez dutenena, aldiz, 306110ekoa. Hala ere, emaitza hauetatik ezin da ondorioztatu igerilekua duen etxebizitzaren batezbesteko prezioa 31870 dolar garestiagoa denik zeren eta etxebizitzaren prezioan eragina izan dezaketen bestelako faktoreak (oin karratu bizigarriak, gela kopurua,...) ez baitira kontuan hartu.

POOL aldagai azaltzailea erabiliz, eredu ekonometriko bat zehaztu, estimatu, kontrasteak egin eta etxebizitzaren prezioetan eragina izan dezaketen bestelako ezaugarriak barneratu ditzaitegu. Hasteko, ondoko ereduak kontsideratzen dugu:

$$PRICE_i = \alpha + \beta POOL_i + u_i \quad i = 1, \dots, 14 \quad (7.1)$$

### Koefizienteen interpretazioa eta estimazioa

Gure adibidean, populazioko erregresio funtzioa aldatzen da etxebizitzak igerilekua izatearen arabera:

- $E(PRICE_i | i \text{ etxebizitzak igerilekua du}) = \alpha + \beta$  da, POOL aldagaiak 1 balioa hartzen duelako eta  $E(u_i) = 0$  delako.
- $E(PRICE_i | i \text{ etxebizitzak ez du igerilekurik}) = \alpha$  da, POOL aldagaiak 0 balioa hartzen duelako eta  $E(u_i) = 0$  delako.

Beraz, koefizienteak horrela interpretatzen dira:

- $\alpha$ : igerilekurik gabeko etxebizitzaren batezbesteko prezioa.
- $\alpha + \beta$ : igerilekua duen etxebizitzaren batezbesteko prezioa.

Hortaz:

- $\beta$ : igerilekua duen etxebizitzaren batezbesteko prezioaren diferentzia igerilekurik ez duenarekiko.

Eredua KTAen bidez estimatuz, ondoko emaitza lortzen dugu Gretlen bitartez:

Eredua 1: KTA estimazioak 14 behaketak erabiliz 1–14  
Aldagai azaldua: price

Aldagaia	Koefizientea	Desb. Tipikoa	$t$ -estatistikoa	p-balioa
const	306,111	30,2077	10,1335	0,0000
pool	31,8689	50,5471	0,6305	0,5402
Aldagai azalduaren batezbestekoa			317,493	
Aldagai azalduaren Desb. Tip.			88,4982	
Hondar Karratuen Batura			98550,5	
Hondarren desbiderazio tipikoa ( $\hat{\sigma}$ )			90,6231	
$R^2$			0,0320632	
Zuzendutako $\bar{R}^2$			-0,0485982	
Askatasun graduak			12	
Log-egiantza			-81,880	
Akaike Informazio Irizpidea			167,760	
Schwarz Bayesian Irizpidea			169,038	
Hannan–Quinn Irizpidea			167,642	

Erregresio bakuneko eredua estimatzeko, 2. gaian lortu genituen ekuazio normalak erabiliz eta POOL fikzio-aldagaiaren karratua aldagai bera dela kontuan hartuz (0 eta 1 balioak hartzen dituelako), (7.1) ereduko koefizienteen estimatzaileak lagin batezbestekoen bitartez lortzen dira<sup>3</sup>:

- $\hat{\alpha} = \overline{PRICE}_{nopool} = 306,11 \Rightarrow$  igerilekurik gabeko etxebizitzaren batezbesteko prezio estimatua.
- $\hat{\beta} = \overline{PRICE}_{pool} - \overline{PRICE}_{nopool} = 337,98 - 306,11 = 31,87 \Rightarrow$  igerilekua duen etxebizitzaren batezbesteko prezioaren diferentzia estimatua igerilekurik ez duenarekiko.

Etxebizitzaren prezio aldakuntzak azaltzeko etxebizitzak igerilekua izatearen funtzioan jar-tzea ez da zehazpen zuzen bakarra. NOPOOL fikzio-aldagaia erabiliz ondorengo bi eredu baliokideak zehaz daitezke:

$$PRICE_i = \gamma_1 + \gamma_2 NOPOOL_i + u_i \quad i = 1, \dots, 14 \quad (7.2)$$

$$PRICE_i = \beta_1 POOL_i + \beta_2 NOPOOL_i + u_i \quad i = 1, \dots, 14 \quad (7.3)$$

Koefizienteak interpretatzerakoan, (7.1), (7.2) eta (7.3) eredu-tako baliokidetasuna ondorengo da:

- $E(PRICE_i | i \text{ etxebizitzak igerilekua du}) = \alpha + \beta = \gamma_1 = \beta_1$
- $E(PRICE_i | i \text{ etxebizitzak ez du igerilekurik}) = \alpha = \gamma_1 + \gamma_2 = \beta_2$

Oker legokeen zehazpen bat ondorengo litzateke:

$$PRICE_i = \alpha + \beta_1 POOL_i + \beta_2 NOPOOL_i + u_i \quad i = 1, \dots, 14$$

zeren eredu honetan  $X$  datu matrizea aztertzerakoan, bigarren eta hirugarren zutabeen batuketara lehenengoa baita eta aipatutako oinarrizko hipotesi bat,  $h(X) < k$ , ez da betetzen. Kolinealitate anizkoitz zehatza dago, hau da, aldagaien arteko konbinazio lineal bat dago. Ondorioz,  $X'X$  singularra denez ezin dugu bere alderantzizkoa atera eta ezin izango genuke eredua KTAen bitartez era bakar batean estimatu.

---

<sup>3</sup> $\overline{PRICE}_{pool}$  igerilekua duen etxebizitzaren lagin batezbesteko prezioa da eta  $\overline{PRICE}_{nopool}$  igerilekurik gabeko etxebizitzaren lagin batezbesteko prezioa da.

$$X = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

### Hipotesien kontrastea

Hipotesien kontrasteak aurreko gaietan ikusitako metodologia erabiliz egiten dira. Adibidez, etxebizitzaren batezbesteko prezioaren diferentzia igerilekua izateagatik, esanguratsua dela kontrastatzeko hipotesi hutsa (7.1) erudian  $\beta = 0$  da<sup>4</sup>. Kontraste hau egiteko  $t$ -estatistikoa erabili daiteke. Kasu honetan bere  $p$ -balioa 0,5402 denez, ez dugu hipotesi hutsa baztertzen %5eko esangura-mailarekin, hau da, etxebizitzaren batezbesteko prezioaren diferentzia igerilekua izateagatik ez da esanguratsua. Kontraste bera hondar karratuen baturetan oinarritutako  $F$ -estatistikoa erabiliz ere egin daiteke. Honela, (7.1) murriztu gabeko eredua eta  $PRICE_i = \alpha + u_i \quad i = 1, \dots, 14$  murriztutako eredua izanik.

### Kategoria bi baino gehiagoko aldagai koalitativoa

Aurrean ikusitako kasua orokortu daiteke kategoria bi baino gehiagoko aldagai koalitativoa erudian barneratzean.

Demagun igerilekua duten etxebizitzaren prezioa azaldu nahi dugula eta, errazteko asmoz, igerilekuaren tamaina dela aldagai azaltzaile bakarra. Horretarako, ondoko fikzio-aldagaiak definitzen ditugu:

$$D_{1i} = \begin{cases} 1 & \text{i etxebizitzaren igerilekua txikia baldin bada (10 m}^2 \text{ baino gutxiago)} \\ 0 & \text{en caso contrario} \end{cases}$$

$$D_{2i} = \begin{cases} 1 & \text{i etxebizitzaren igerilekua ertaina baldin bada (10 eta 20 m}^2 \text{ artean)} \\ 0 & \text{en caso contrario} \end{cases}$$

$$D_{3i} = \begin{cases} 1 & \text{i etxebizitzaren igerilekua handia baldin bada (20 m}^2 \text{ baino gehiago)} \\ 0 & \text{en caso contrario} \end{cases}$$

<sup>4</sup>Era berean,  $H_0 : \gamma_2 = 0$  edo  $H_0 : \beta_1 = \beta_2$  (7.2) eta (7.3) eruedetan, hurrenez hurren.

Eredua zehazterakoan, termino konstantea eta hiru fikzio-aldagaiak barneratzen baditugu kolinealitate anizkoitzaren arazoa daukagu. Arazo hau ekiditzeko termino konstantea edo fikzio-aldagaietatik bat kendu behar da. Adibidez, zehazpen baliokide posible batzuk ondo-koak dira:

$$PRICE_i = \alpha_1 D_{1i} + \alpha_2 D_{2i} + \alpha_3 D_{3i} + u_i \quad i = 1, \dots, N \quad (7.4)$$

$$PRICE_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + u_i \quad i = 1, \dots, N \quad (7.5)$$

non  $\alpha_1 = \beta_1$  igerileku txikia duen etxebizitzaren batezbesteko prezioa den,  $\alpha_2 = \beta_1 + \beta_2$  igerileku ertaina duenarena eta  $\alpha_3 = \beta_1 + \beta_3$  igerileku handia duenarena. Hortaz,  $\beta_2$  igerileku ertaina duen etxebizitzaren batezbesteko salmenta prezioaren diferentzia igerileku txikia duenarekiko da eta  $\beta_3$  igerileku handia duen etxebizitzaren batezbesteko salmenta prezioaren diferentzia igerileku txikia duenarekiko.

Lehen aipatu dugun bezala, hipotesien kontrasteak aurreko gaietan ikusitako metodologia erabiliz egiten dira. Adibidez, (7.4) ereduan igerilekuaren tamainak ez daukala eraginik etxebizitzaren batezbesteko prezioan kontrastatzeko hipotesi hutsa  $H_0 : \alpha_1 = \alpha_2 = \alpha_3$  da, (7.4) murriztu gabeko eredua izanik eta  $PRICE_i = \alpha + u_i \quad i = 1, \dots, N$  murriztutako eredua. Era berean, (7.5) ereduan hipotesi hutsa  $H_0 : \beta_2 = \beta_3 = 0$  da, (7.5) murriztu gabeko eredua izanik eta  $PRICE_i = \beta_1 + u_i \quad i = 1, \dots, N$  murriztutako eredua.

### Aldagai koantitatiboen barnerapena

Orain arte etxebizitzaren prezioa azaltzeko aldagai azaltzaile bakarra koalatitiboa izan dugu. Hala ere, askotan aldagai koalatitibo batez gain, aldagai koantitatiboak barneratzen dira aldagai endogenoa azaltzeko. Gure adibidean, etxebizitzaren prezioa etxebizitzak igerilekua izatearen funtzioan zehaztea ez da oso errealista, beraz, SQFT (etxebizitzaren oin karratu bizigarriak) aldagai koantitatiboa aldagai azaltzaile moduan erantsiko dugu ereduan. Horretarako, bi aukera ikusiko ditugu. Lehenengoan, SQFT aldagaiak  $E(PRICE_i)$  zuzenean jatorri aldaketa eragiten du. Bigarrean, berriz, bai jatorrian baita zuzenaren maldan ere aldaketa eragiten du.

### Jatorri aldaketa

SQFT aldagaia barneratzen dugu gure ereduan aldagai azaltzaile moduan:

$$PRICE_i = \alpha_1 + \alpha_2 POOL_i + \beta_1 SQFT_i + u_i \quad i = 1, \dots, 14 \quad (7.6)$$

### Koefizienteen interpretazioa eta estimazioa

Populazioko erregresio funtzioa ondokoa da:

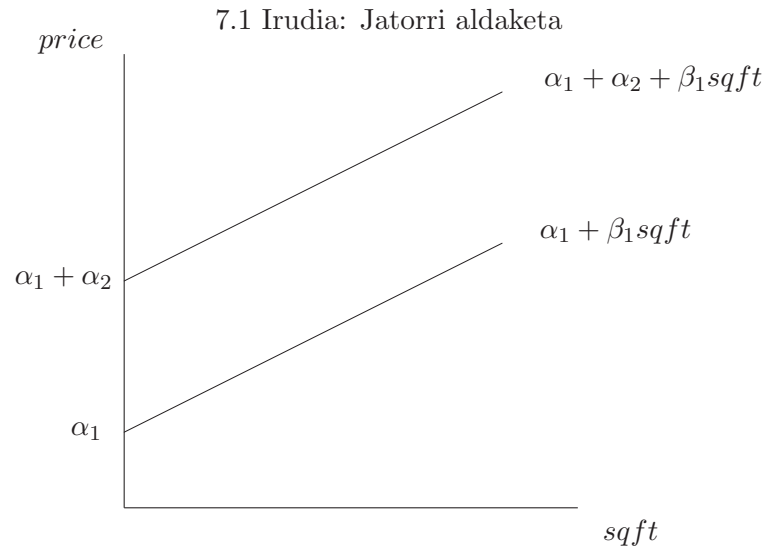
- $E(PRICE_i | i \text{ etxebizitzak igerilekua dauka}) = \alpha_1 + \alpha_2 + \beta_1 SQFT_i$



- $E(PRICE_i | i \text{ etxebizitzak ez dauka igerilekurik}) = \alpha_1 + \beta_1 SQFT_i$

Beraz,  $\alpha_1$  igerileku gabeko eta oin karratu bizigarririk gabeko etxebizitzaren batezbesteko prezioa da,  $\alpha_2$  igerilekua duen etxebizitzaren batezbesteko prezioaren diferentzia igerilekurik ez duenarekiko, oin karratu bizigarrien kopurua berdina izanik eta  $\beta_1$ -k etxebizitza oin karratu bizigarri batean handitzen denean batezbesteko prezioaren gehikuntza adierazten du.

Grafikoki, malda berdineko ( $\beta_1$ ) eta jatorri ezberdineko zuzen bi dauzkagu:



Ondoren (7.6) ereduko KTA bidezko estimazioaren emaitzak ematen dira:

Eredua 2: KTA estimazioak 14 behaketak erabiliz 1–14  
Aldagai azaldua: price

Aldagaia	Koefizientea	Desb. Tipikoa	$t$ -estatistikoa	p-balioa
const	22,6728	29,5058	0,7684	0,4584
pool	52,7898	16,4817	3,2029	0,0084
sqft	0,144415	0,0141849	10,1809	0,0000
Aldagai azalduaren batezbestekoa			317,493	
Aldagai azalduaren Desb. Tip.			88,4982	
Hondar Karratuen Batura			9455,36	
Hondarren desbiderazio tipikoa ( $\hat{\sigma}$ )			29,3186	
$R^2$			0,907132	
Zuzendutako $\bar{R}^2$			0,890247	
$F(2, 11)$			53,7238	
Log-egiantza			-65,472	
Akaike Informazio Irizpidea			136,944	
Schwarz Bayesian Irizpidea			138,861	
Hannan–Quinn Irizpidea			136,767	

Emaitza hauetan oinarrituriko lagineko erregresio funtzioa hurrengoa da:

$$\widehat{PRICE}_i = 22,673 + 52,790 POOL_i + 0,144 SQFT_i$$

$$\begin{matrix} (t_{est}) & (0,768) & (3,203) & (10,181) \end{matrix}$$

Aldagaiak aztertzerakoan, aldagai azaltzaile biak esanguratsuak direla kontrastatu daiteke etxebizitzaren batezbesteko prezioa azaltzeko. Izatez, dagozkien koefizienteen t-estatistikoak Student- $t$  tauletako  $t_{(N-K)0,05/2} = t_{(14-3)0,05/2} = 2,201$  balioa baino handiagoak dira. Beraz, alde batetik, etxebizitzaren batezbesteko prezioen diferentzia etxebizitzak igerilekua izateagatik esanguratsua da. Bestaldetik, etxebizitzaren oin karratu bizigarrien kopurua garrantzitsua da prezioa azaltzeko.

Koefiziente estimatuen interpretazioa ondorengoa da:

- $\hat{\alpha}_1 = 22,673 \Rightarrow$  igerileku gabeko eta oin karratu bizigarriarik gabeko etxebizitzaren batezbesteko prezio estimatua 22673 dolarrekoa da.
- $\hat{\alpha}_2 = 52,790 \Rightarrow$  igerilekua duen etxebizitzaren batezbesteko prezio estimatua 52790 dolar garestiagoa da igerilekua ez duen etxebizitzaren batezbesteko prezio estimatuarekiko, beste guztia konstante mantenduz, hau da, oin karratu bizigarrien kopuru berdina izanik.
- $\hat{\beta}_1 = 0,144 \Rightarrow$  etxebizitza oin karratu bizigarri batean handitzen denean batezbesteko prezio estimatua 144 dolarretan gehitzen da.

### Jatorri eta malda aldaketa

Etxebizitzaren azalera oin karratu batean handitzerakoan, dakartzan prezioaren batezbesteko aldakuntza, igerilekua izatearen menpekoa dela pentsa daiteke. Hala bada, POOL fikzio-aldagaiak eragina dauka bai zuzenaren jatorrian eta baita maldan ere. Ondoko ereduak aipatutako eragina jasotzen du:

$$PRICE_i = \alpha_1 + \alpha_2 POOL_i + \beta_1 SQFT_i + \beta_2 POOL \times SQFT_i + u_i \quad i = 1, \dots, 14 \quad (7.7)$$

$POOL \times SQFT$  terminoak igerilekua duten etxebizitzaren oin karratu bizigarriak biltzen ditu eta igerilekurik ez dutenentzat berriz, 0 balioa hartzen du.

### Koefizienteen estimazioa eta interpretazioa

$POOL \times SQFT$  aldagai berria definitu ondoren, (7.7) eredu estimatzen dugu:

Eredua 3: KTA estimazioak 14 behaketak erabiliz 1–14  
Aldagai azaldua: price

Aldagaia	Koefizientea	Desb. Tipikoa	$t$ -estatistikoa	p-balioa
const	77,1332	25,6379	3,0086	0,0131
pool	-82,648	39,7759	-2,0779	0,0644
sqft	0,116667	0,0125934	9,2641	0,0000
poolxsqft	0,0722955	0,0203274	3,5566	0,0052

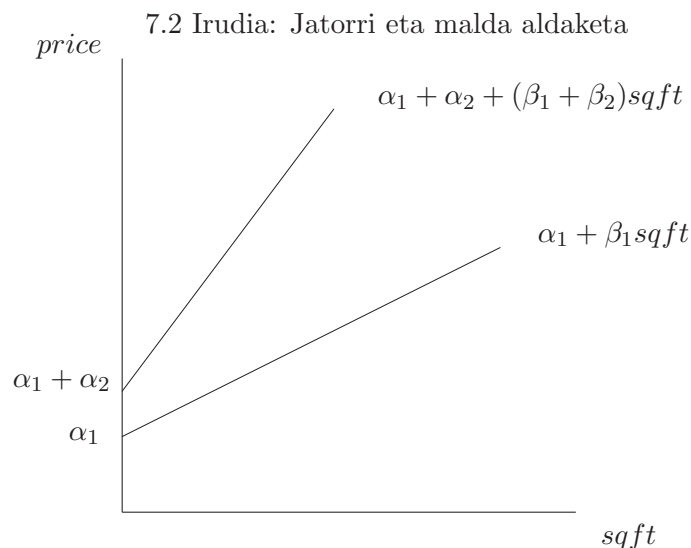
Aldagai azalduaren batezbestekoa	317,493
Aldagai azalduaren Desb. Tip.	88,4982
Hondar Karratuen Batura	4174,72
Hondarren desbiderazio tipikoa ( $\hat{\sigma}$ )	20,4321
$R^2$	0,958997
Zuzendutako $\bar{R}^2$	0,946696
$F(3, 10)$	77,9615
Log-egiantza	-59,749
Akaike Informazio Irizpidea	127,499
Schwarz Bayesian Irizpidea	130,055
Hannan–Quinn Irizpidea	127,262

Populazioko erregresio funtzioak:

- $E(PRICE_i|i \text{ etxebizitzak igerilekua du}) = \alpha_1 + \alpha_2 + (\beta_1 + \beta_2)SQFT_i$
- $E(PRICE_i|i \text{ etxebizitzak ez du igerilekurik}) = \alpha_1 + \beta_1 SQFT_i$

$\alpha_1$  parametroa aurreko kasuan ikusi dugun moduan interpretatzen da.  $\alpha_2$  azalera zero eta igerilekua duen etxebizitzaren batezbesteko salmenta prezioaren diferentzia igerilekurik ez duenarekiko da,  $\beta_1$ -ek etxebizitzak oin karratu bat gehiago izaterakoan, igerilekurik ez duen etxebizitzaren batezbesteko prezioaren gehikuntza neurtzen du,  $(\beta_1 + \beta_2)$ -k etxebizitzak oin karratu bat gehiago izaterakoan, igerilekua duen etxebizitzaren batezbesteko prezioaren gehikuntza neurtzen du eta hortaz,  $\beta_2$ -k etxebizitzak oin karratu bat gehiago izaterakoan, igerilekua duen etxebizitzaren batezbesteko prezioaren diferentziaren gehikuntza neurtzen du.

Grafikoki, malda eta jatorri ezberdineko bi zuzen dauzkagu:



Koefiziente estimatuen interpretazioa:

- $\hat{\alpha}_1 = 77,133 \Rightarrow$  etxebizitzaren batezbesteko salmenta prezio estimatua, azalera zero denean eta igerilekurik ez duenean, 77133 dolarrekoa da.

- $\hat{\alpha}_2 = -82,648 \Rightarrow$  azalera zero eta igerilekua duen etxebizitzaren batezbesteko salmenta prezioaren diferentzia estimatua igerilekurik ez duenarekiko -82648 dolarrekoa da.
- $\hat{\beta}_1 = 0,117 \Rightarrow$  etxebizitzak oin karratu bat gehiago izaterakoan, igerilekurik ez duen etxebizitzaren prezioaren batezbesteko gehikuntza estimatua 117 dolarrekoa da.
- $\hat{\beta}_2 = 0,072 \Rightarrow$  etxebizitzak oin karratu bat gehiago izaterakoan, igerilekua duen etxebizitzaren batezbesteko prezio estimatuaren gehikuntza, igerilekua ez duen etxebizitzaren batezbesteko prezio estimatuaren gehikuntzarekiko diferentzia 72 dolarrekoa da.

### Hipotesien kontrastea

Etxebizitzaren batezbesteko prezioa desberdina dela igerilekua izateagatik kontrastatzeko hipotesi hutsa  $H_0 : \alpha_2 = \beta_2 = 0$  eta aurkakoa  $H_a : \alpha_2 \neq 0$  edota  $\beta_2 \neq 0$  dira.

Kontrastearen emaitza ondokoa da:

Hipotesi hutsa: aldagaien erregresio koefizienteak zero dira

```
pool
poolxsqft
```

Estatistikoa:  $F(2, 10) = 16,886$ , p-balioarekin = 0,000622329

hortaz, %5eko esangura-mailarekin hipotesi hutsa baztertzen dugu eta igerilekua izateak eragina dauka etxebizitzaren batezbesteko prezioetan.

Halaber, banakako esanguratasun kontrastearen bidez, kontrastatu daiteke ea etxebizitzak igerilekua izateak batezbesteko prezioan eragiten duen jatorrian bakarrik ( $H_0 : \alpha_2 = 0$ ) edo maldan bakarrik ( $H_0 : \beta_2 = 0$ ).

## 7.3 Aldagai koalatibo bi edo gehiago dituen eredua

Zehaztutako (7.6) ereduan, etxebizitzaren prezioa azaldu lezaketeen aldagai azaltzaile gehiago barneratu daitezke. Adibidez, etxebizitzak egongela izatea, tximinia izatea, logela kopurua eta komun kopurua. Lehenengo bi aldagaiak (egongela eta tximinia izatea) koalatiboak dira eta azken biak (logela eta komun kopurua), berriz, koantitatiboak dira. Demagun ondoko eredua zehazten dugula:

$$PRICE_i = \alpha_1 + \alpha_2 POOL_i + \alpha_3 FAMROOM_i + \alpha_4 FIREPL_i + \beta_1 SQFT_i + \beta_2 BEDRMS_i + \beta_3 BATHS_i + u_i \quad i = 1, \dots, 14 \quad (7.8)$$

non FAMROOM eta FIREPL hurrengo fikzio-aldagaiak diren:

$$FAMROOM_i = \begin{cases} 1 & \text{i etxebizitzak egongela baldin badu} \\ 0 & \text{en caso contrario} \end{cases}$$

$$FIREPL_i = \begin{cases} 1 & \text{i etxebizitzak tximinia baldin badu} \\ 0 & \text{en caso contrario} \end{cases}$$

BEDRMS<sub>*i*</sub>-k, *i* etxebizitzaren logela kopurua adierazten du eta BATHS<sub>*i*</sub>-k, *i* etxebizitzaren komun kopurua adierazten du.

Aukeratutako zehazpenean, (7.8) ereduan, hiru aldagai koalitatibo (igerilekua izatea, tximinia izatea eta egongela izatea) daude eta jatorrian bakarrik eragiten dute. Zehazpenean, termino konstantea mantendu da eta aldagai koalitatibo bakoitzatik kategoria bat kendu da. Koefizienteak aurreko ataletan ikusi dugun moduan interpretatzen dira. Adibidez,  $\alpha_3$  koefizienteak egongela duen etxebizitzaren batezbesteko prezioaren diferentzia egongelarik ez duenarekiko adierazten du beste aldagai azaltzaile guztiak konstante mantenduz. Estimazioaren emaitzak ondokoak dira:

Eredua 4: KTA estimazioak 14 behaketak erabiliz 1–14  
Aldagai azaldua: price

Aldagaia	Koefizientea	Desb. Tipikoa	<i>t</i> -estatistikoa	p-balioa
const	39,0571	89,5397	0,4362	0,6758
pool	53,1958	22,0635	2,4110	0,0467
famroom	-21,344	42,8734	-0,4979	0,6338
firepl	26,1880	53,8454	0,4864	0,6416
sqft	0,146551	0,0301014	4,8686	0,0018
bedrms	-7,0455	28,7363	-0,2452	0,8134
baths	-0,263691	41,4547	-0,0064	0,9951

Aldagai azalduaren batezbestekoa	317,493
Aldagai azalduaren Desb. Tip.	88,4982
Hondar Karratuen Batura	9010,24
Hondarren desbiderazio tipikoa ( $\hat{\sigma}$ )	35,8773
$R^2$	0,911504
Zuzendutako $\bar{R}^2$	0,835650
$F(6, 7)$	12,0166
p-balioa $F()$	0,00221290
Log-egiantza	-65,134
Akaike Informazio Irizpidea	144,269
Schwarz Bayesian Irizpidea	148,743
Hannan–Quinn Irizpidea	143,855

Koefiziente estimatuen interpretazioa:

- $\hat{\alpha}_1 = 39,057$ : etxebizitzaren batezbesteko salmenta prezio estimatua, azalera zero duen eta igerilekurik, komunik, logelik eta tximinirik ez dituenen, 39057 dolarrekoa da.
- $\hat{\alpha}_2 = 53,1958$ : igerilekua duen etxebizitzaren batezbesteko salmenta prezioaren diferentzia estimatua igerilekurik ez duenarekiko 53196 dolarrekoa da, gainontzeko ezaugarriak (azalera, logelak, komunak, egongela eta tximinia) berdinak izanik.
- $\hat{\alpha}_3 = -21,34$ : egongela duen etxebizitzaren batezbesteko prezio estimatua egongelarik ez duenarena baino 21340 dolar merkeagoa da, gainontzeko ezaugarriak berdinak izanik. Zeinu hau izatearen arrazoia, etxebizitzaren azalera, logela eta komun kopurua konstante mantenduz, egongela izateak gelak edota komunak txikiagoak izan behar direla inplikatzeko duela, da.
- $\hat{\alpha}_4 = 26,188$ : tximinia duen etxebizitzaren batezbesteko prezio estimatua tximinirik ez duenarena baino 26188 dolar garestiagoa da, gainontzeko ezaugarriak berdinak izanik.
- $\hat{\beta}_1 = 0,147$ : etxebizitzak oin karratu bat gehiago izaterakoan, gainontzeko aldagaiak konstante mantenduz, espero den salmenta prezioaren batezbesteko gehikuntza 147000 dolarrekoa da.
- $\hat{\beta}_2 = -7,046$ : etxebizitzak logela bat gehiago izaterakoan, gainontzeko aldagaiak konstante mantenduz, espero den salmenta prezioaren batezbesteko jaitsiera 7046 dolarrekoa da. Honen arrazoia azalera eta komun kopurua mantentzen direnez, logelak txikiagoak izan beharko dutela, da.
- $\hat{\beta}_3 = -0,264$ : etxebizitzak komun bat gehiago izaterakoan, gainontzeko aldagaiak konstante mantenduz, espero den salmenta prezioaren batezbesteko jaitsiera 264 dolarrekoa da. Lehen bezala, logelak txikiagoak izango baitira.

### Hipotesien kontrastea

Adibide moduan, etxebizitzaren batezbesteko prezioaren diferentzia etxebizitzak tximinia izateagatik esanguratsua dela kontrastatzeko, FIREPL aldagaiaren banakako esanguratasun kontrastea egin behar da. Kasu honetan, dagokion p-balioa 0,6416 denez, %5eko esanguramailearekin diferentzia ez dela esanguratsua ondorioztatu daiteke.

Aurreko (7.6) eta (7.8) ereduak konparatuz, azken ereduaren erantsitako aldagaiak ez dira banaka adierazgarriak. Halaber,  $\bar{R}^2$  txikiagoa da. Erantsitako aldagaien baterako esanguratasuna kontrastatzeko hipotesiak  $H_0 : \alpha_3 = \alpha_4 = \beta_2 = \beta_3 = 0$  eta  $H_a : \alpha_3 \neq 0$  edota  $\alpha_4 \neq 0$  edota  $\beta_2 \neq 0$  edota  $\beta_3 \neq 0$  dira. Kontrastea hondar karratuen baturetan oinarritutako  $F$ -estatistikoa erabiliz egin daiteke, (7.6) murriztutako eta (7.8) murriztu gabeko ereduak izanik.

Kontrastearen emaitza ikusirik:

Hipotesi hutsa: aldagaien erregresio koefizienteak zero dira

```
famroom
firepl
bedrms
baths
```

Estatistikoa:  $F(4, 7) = 0,0864517$ , p-balioarekin = 0,983881

Ez dugu hipotesi hutsa baztertzen %5eko esangura-mailarekin, hau da, (7.6) erduan erantsitako aldagaiak ez dira batera esanguratsuak. Aldagai horiek omititzen baditugu, lortutako erdua hobeagoa da koefizienteen esanguratasun eta  $\bar{R}^2$  koefizientearen aldetik. Beraz, POOL eta SQFT aldagaiak erduan mantenduz, gainontzekoak (FIREPL, FAMROOM, BATHS, BEDRMS) ez dira etxebizitzaren prezioaren azalpenean lagungarriak.

## 7.4 Egitura aldaketa

Batzuetan, aldagai azaldua eta azaltzaileen arteko erlazioa laginean zehar aldatu daiteke, hau da, egitura aldaketa bat gerta daiteke. Adibidez, tabakoaren kontsumoa analizatu nahi badugu eta laginean zehar, tabakoaren kalte informatzeko asmoz, kanpaina bat abian jartzen bada, kanpainaren ondorioz tabakoaren eskaera murriztu daitekela pentsa dezakegu. Hala bada, ezin da eskaera funtzio bakarra zehaztu, bi baizik. Bata kanpaina egiten den momenturaino eta bestea, kanpainatik aurrera. Beraz, egitura aldaketaren bat dagoela susmatzen badugu funtzioaren koefizienteen egonkortasuna kontrastatu beharko dugu.

Egitura aldaketaren kontrastea, Chow-en kontrastea deritzona, era desberdinetara burutu daiteke. Alde batetik, hondar karratuen baturetan oinarritutako  $F$  estatistikoa erabiliz eta bestaldetik, fikzio-aldagaiak erabiliz.

Ramanathaneko (2002) data7-19 fitxategian 1960 eta 1988 tartean Turkian tabakoaren eskaerari buruzko ikasketa bat egiteko datuak daude:

Q: Tabakoaren kontsumoa kilotan (Ibiltartea 1,86 - 2,723)  
 Y: NPG per capita 1968ko Turkiako liratean (Ibiltartea 2560 - 5723)  
 P: Tabakoaren salmenta prezioa Turkiako liratean (Ibiltartea 1,361 - 3,968)  
 D82: 1982 urtetik 1 balioa hartzen duen fikzio-aldagaia

Tabakoaren eskaera azaltzeko, ondoko erdua zehazten dugu:

$$\ln Q_t = \alpha + \beta \ln Y_t + \gamma \ln P_t + u_t \quad t = 1960, \dots, 1988 \quad (7.9)$$

Hala ere, 1981 urtean Turkiako gobernuak tabakoaren aurkako kanpaina bati hasiera eman zion eta horren ondorioz, tabakoaren kontsumoa aldatu zen. Gure helburua aldaketa hori nabaria izan zen ala ez aztertzea da. Nabaria izan bazen, ondoko erduak zehaztu beharko genituzke:

$$\ln Q_t = \alpha_1 + \beta_1 \ln Y_t + \gamma_1 \ln P_t + u_{1t} \quad t = 1960, \dots, 1981 \quad (7.10)$$

$$\text{Ln}Q_t = \alpha_2 + \beta_2 \text{Ln}Y_t + \gamma_2 \text{Ln}P_t + u_{2t} \quad t = 1982, \dots, 1988 \quad (7.11)$$

Egitura aldaketarik ez dagoelaren hipotesi hutsa  $H_0 : \alpha_1 = \alpha_2, \beta_1 = \beta_2$  eta  $\gamma_1 = \gamma_2$  da eta hondar karratuen baturetan oinarritutako  $F$ -estatistikoa erabil daiteke kontrastea burutzeko. Kasu honetan, (7.10) eta (7.11) ekuazioek murriztu gabeko eredia osatzen dute eta (7.9) murriztutako eredia da.

Adibide honetarako Chow-en kontrastea burutuko dugu Gretl erabiliz. Data7-19 fitxategia ireki eta aldagaiak eraldatu ondoren, (7.9) eredia estimatzen dugu KTA bitartez. Estimazioaren emaitzen leihoan *Kontrasteak* klikatu, gero *Chow kontrastea* aukeratu eta *Lagina zatitzeko behaketa* leihoan 1982 idatzi. Estimazio eta kontrastearen emaitza hurrengoa da:

Eredua 1: KTA estimazioak 29 behaketak erabiliz 1960–1988

Aldagai azaldua: LnQ

Aldagaia	Koefizientea	Desb. Tipikoa	$t$ -estatistikoa	p-balioa
const	-4,5898	0,724913	-6,3316	0,0000
LnY	0,688498	0,0947276	7,2682	0,0000
LnP	-0,485683	0,101394	-4,7900	0,0001
Aldagai azalduaren batezbestekoa			0,784827	
Aldagai azalduaren Desb. Tip.			0,108499	
Hondar Karratuen Batura			0,0949108	
Hondarren desbiderazio tipikoa ( $\hat{\sigma}$ )			0,0604187	
$R^2$			0,712058	
Zuzendutako $\bar{R}^2$			0,689908	
$F(2, 26)$			32,1480	
Durbin–Watson estatistikoa			1,00057	
Lehen ordenako autokoerlazio koefizientea			0,489867	
Log-egiantza			41,8214	
Akaike Informazio Irizpidea			-77,642	
Schwarz Bayesian Irizpidea			-73,541	
Hannan–Quinn Irizpidea			-76,358	

Chow egitura aldaketaren estatistikoa (1982 behaketan) –

Hipotesi hutsa: egitura aldaketarik ez

Estatistikoa:  $F(3, 23) = 20,1355$

p-balioarekin =  $P(F(3, 23) > 20,1355) = 1,25619\text{e-}006$

Emaitzetan ikusten dugunez, kalkulaturako  $F$ -estatistikoa  $20,1355 > F_{(3,23)0,05}$  da, beraz hipotesi hutsa baztertzen dugu %5eko esangura-mailarekin. Honen ondorioz, Turkiako gobernuak hasitako tabakoaren aurkako kanpainak eragina izan zuela kontsumoan esan dezakegu eta (7.10) eta (7.11) ekuazioek azaltzen dutela tabakoaren eskaera. Estimazioaren emaitzak ondokoak dira:

$$\widehat{\text{Ln}Q}_t = -5,024 + 0,735 \text{Ln}Y_t - 0,381 \text{Ln}P_t \quad t = 1960, \dots, 1981 \quad R^2 = 0,9456$$

$(t_{est}) \quad (-10,614) \quad (11,587) \quad (-4,227)$



$$\widehat{LnQ}_t = 8,837 - 0,953 LnY_t + 0,108 LnP_t \quad t = 1982, \dots, 1988 \quad R^2 = 0,6203$$

$$\begin{matrix} (t_{est}) & (2,170) & (-1,941) & (0,654) \end{matrix}$$

### Egitura aldaketaren kontrastea fikzio-aldagaiak erabiliz

Aurreko kontrastea D82 fikzio-aldagaia erabiliz burutu daiteke. Horretarako, hurrengo eredua zehazten dugu:

$$LnQ_t = \beta_1 + \beta_2 LnY_t + \beta_3 LnP_t + \beta_4 D82_t + \beta_5 D82_t \cdot LnY_t + \beta_6 D82_t \cdot LnP_t + u_t \quad (7.12)$$

$$t = 1960, \dots, 1988$$

non egitura aldaketarik ez dagoelaren hipotesi hutsa  $H_0 : \beta_4 = \beta_5 = \beta_6 = 0$  den. Hondar karratuen baturetan oinarritutako  $F$ -estatistikoa erabil daiteke kontrastea burutzeko, (7.12) murriztu gabeko eredua izanik eta

$$LnQ_t = \beta_1 + \beta_2 LnY_t + \beta_3 LnP_t + u_t \quad t = 1960, \dots, 1988 \quad (7.13)$$

murriztutako eredua.

Ondoren, Gretlen jarraitu behar diren pausuak egitura aldaketaren kontrastea burutzeko aztertuko dugu. Aldagaiak eraldatu ondoren<sup>5</sup>, (7.12) eredua estimatzen dugu KTAren bitartez. Estimazioaren emaitzen leihuan *Kontrasteak* klikatu, gero *Omititu aldagaiak* eta D82, D82 × LnY eta D82 × LnP aldagaiak aukeratzen ditugu. Kontrastearen emaitza ondokoa da:

Eredua 2: KTA estimazioak 29 behaketak erabiliz 1960–1988

Aldagai azaldua: LnQ

Aldagaia	Koefizientea	Desb. Tipikoa	t-estatistikoa	p-balioa
const	-5,0248	0,541262	-9,2837	0,0000
LnY	0,735837	0,0726077	10,1344	0,0000
LnP	-0,381857	0,103289	-3,6970	0,0012
D82	13,8623	2,85197	4,8606	0,0001
D82LnY	-1,6892	0,345466	-4,8898	0,0001
D82LnP	0,490526	0,153989	3,1855	0,0041

Aldagai azalduaren batezbestekoa 0,784827

Aldagai azalduaren Desb. Tip. 0,108499

Hondar Karratuen Batura 0,0261724

Hondarren desbiderazio tipikoa ( $\hat{\sigma}$ ) 0,0337332

$R^2$  0,920598

Zuzendutako  $\bar{R}^2$  0,903336

$F(5, 23)$  53,3328

Durbin–Watson estatistikoa 2,02153

Lehen ordenako autokoerlazio koefizientea -0,0136939

Log-egiantza 60,5008

Akaike Informazio Irizpidea -109,00

Schwarz Bayesian Irizpidea -100,79

Hannan–Quinn Irizpidea -106,43

<sup>5</sup>Adibidez, LnQ sortzeko Gretlen, *Aldagaia* → *Definitu aldagai berria...* klikatu eta agertzen den kaxan  $LnQ = \ln(Q)$  idatzi.

Omititutako aldagaien kontrastea –

Hipotesi hutsa: aldagai guztientzat parametroak zero dira

D82

D82LnY

D82LnP

Estatistikoa:  $F(3, 23) = 20,1355$

p-balioarekin =  $P(F(3, 23) > 20,1355) = 1,25619e-006$

Emaitzetan ikus daitekenez, p-balioa 1,25619e-006 da eta hipotesi hutsa baztertzeko %5eko esangura-mailarekin. Beraz, Turkiako gobernuak hasitako tabakoaren aurkako kanpainak eragina izan zuen kontsumoan. Honen ondorioz, (7.10) eta (7.11) ekuazioek azaltzen dute tabakoaren kontsumoa.

## Bibliografia

**Ramanathan, R.** (2002), *Introductory Econometrics with applications*, 5. ed., South-Western, Ohio.