

Notkun flokkabreyta í aðfallsgreiningu

Hægt er að nota flokkabreytur í aðfallsgreiningu ef þær eru kóðaðar á viðeigandi hátt áður en þær eru settar í aðfallsgreiningarforritið. Ýmsar aðferðir eru færar við þessa kóðun en öllum er þeim sameiginlegt að hver flokkabreytu er táknuð með einni eða fleirum tvíkostabreytum. Fjöldi breyta fer eftir fjölda þeirra flokka sem flokkabreytan tekur; þær eru ætíð einum færri en flokkarnir. Hér verður lýst algengustu aðferð við að kóða flokkabreytur í aðfallsgreiningu, því að nota „dummy“-breytur.

Enska orðið „dummy“ skírskotar hér ekki til heimsku eða kjánaskapar heldur visar það að öllum líkindum þess að breyturarnar koma í stað upprunalegu flokkabreytanna. Þær eru því staðgenglar í sama skilningi og tálbeitur (*dummies*) gæsaskyttunnar sem hann stillir upp til að laða að bráðina. Við kóðum flokkabreytur því á forni einnar eða fleiri staðgengilsbreyta.

Við ætlum að nota gagnasafnið yfir samræmdar einkunnir grunnskóla í Reykjavík. Við höfum skipað skólunum í fjögur hverfi og miðum við að hvert hverfi hafi byggst upp á svipuðum tíma. Okkar viðfangsefni er þá að kóða þessa fjórskiptu flokkabreytu á form nokkurra staðgengilsbreyta.

Hverfin eru fjögur. Almenn séð þurfum við $k-1$ staðgengilsbreytur þar sem k er fjöldi flokka. Í okkar tilfalli eru fjórir flokkar og því notum við þrjár staðgengilsbreytur.

Við þurfum að velja okkur viðmiðunarhóp og kjósum að nota vesturborgina, Hagaskóla og Austurbæjarskóla, sem viðmiðun fyrir hin hverfin. Síðan kóðum við breyturarnar þrjár þannig að hin hverfin þrjú séu borin saman við þetta viðmiðunarhverfi.

Við notum þrjár staðgengilsbreytur, X_1 , X_2 og X_3 . Breytuna X_1 notum við til að tákna muninn á hverfi 2, svæðinu austan Snorrabrautar og vestan Elliðaár, og vesturborginni. Við látum alla skóla fá núll á þessari breytu nema þá átta grunnskóla sem eru á þessu svæði. Breytan X_2 stendur fyrir Breiðholt og Árbæ og því fá þeir skólar 1 á henni en aðrir núll. Að síðustu stendur breytan X_3 fyrir Grafarvog og nýju hverfin þar í kring. Það eru fjórir skólar sem fá einn á þeirri breytu en aðrir fá núll.

Það er mikilvægt að átta sig á því að við höfum allar upplýsingar flokkabreytunnar með þessum þremur staðgengilsbreytum. Ef við skoðum hvert hverfi fyrir sig með tilliti til vektors þessara þriggja breyta, sjáum við að við þekkjum skólanna tvö í vesturborginni af mynstrinu [0; 0; 0], skólanna átta þar austur af þekkjum við af mynstrinu [1; 0; 0], Breiðholt og Árbæ þekkjum við af [0; 1; 0] og nýju hverfin við Grafarvog og þar norður af einkennast af mynstrinu [0; 0; 1]. Þannig hafa allar þær upplýsingar sem fölgna voru í fjórskiptu flokkabreytunni varðveist í þessum fjórum einkennismynstrum.

Til að prófa áhrif flokkabreytuna setjum við allar þrjár staðgengilsbreytur inn í líkanið með öðrum breytum líkansins (MenntStig og Fjarvist); þetta samsvarar líkani 1 í rammanum hér til hliðar. Við fáum hallastuðla fyrir bæði menntunarstig og fjarvistir en auk þess þrjá mismunandi hallastuðla fyrir flokkabreytuna Hverfi. Þessa hallastuðla má túlka sem meðaltalsmun viðkomandi hverfis og viðmiðunarhverfisins eftir að leiðrétt hefur verið fyrir hinar breyturarnar tvær. Þannig er hverfi 2 um 0,1 lægra í samræmdum einkunnum en vesturborgin eftir að leiðrétt hefur verið fyrir mun á menntunarstigi og fjarvistum. Það munar hins vegar 0,2 á Breiðholt og Árbæ annars vegar og vesturborginni hins vegar og svipuðu munar á Grafarvogi og vesturborg. Sá munur sem virðist vera á samræmdum meðaleinkunnum skóla eftir skóla-

Skóli	Hverfi	X ₁	X ₂	X ₃
Hagaskóli	1	0	0	0
Austurbæjarskóli	1	0	0	0
Hlíðaskóli	2	1	0	0
Hvassaleitisskóli	2	1	0	0
Álftamýraskóli	2	1	0	0
Háteigsskóli	2	1	0	0
Laugalækjarskóli	2	1	0	0
Réttarholtsskóli	2	1	0	0
Vogaskóli	2	1	0	0
Langholtsskóli	2	1	0	0
Ölduselsskóli	3	0	1	0
Árbæjarskóli	3	0	1	0
Seljaskóli	3	0	1	0
Hólabrekkuskóli	3	0	1	0
Breiðholtsskóli	3	0	1	0
Fellaskóli	3	0	1	0
Foldaskóli	4	0	0	1
Hamraskóli	4	0	0	1
Rimaskóli	4	0	0	1
Húsaskóli	4	0	0	1

Breyta	Hallatala	SE	p
Líkan 1			
Fasti	2,77	1,34	0,06
MenntStig	1,22	0,44	0,02
Fjarvist	-0,19	0,12	0,12
X ₁	-0,09	0,27	0,75
X ₂	-0,19	0,34	0,58
X ₃	-0,22	0,35	0,55
R ² = 0,77 SE _{Est} = 0,31			
F(5, 14)= 9,4, p < 0,001			
Líkan 2			
Fasti	2,06	0,76	0,01
MenntStig	1,44	0,27	0,01
Fjarvist	-0,17	0,10	0,10
R ² = 0,76 SE _{Est} = 0,28			
F(2, 17)= 27,2, p < 0,001			
Líkan 3			
Fasti	5,95	0,32	0,00
X ₁	-0,32	0,36	0,40
X ₂	-0,94	0,37	0,02
X ₃	-0,86	0,40	0,05
R ² = 0,64 SE _{Est} = 0,46			
F(3, 16)= 3,7, p = 0,03			

hverfum virðist því að verulegu leyti vera tilkominn vegna ólíks menntunarstigs og mismiklum fjarvistum eftir skólahverfum.

Ef við viljum athuga hvort flokkabreytan hafi áhrif í þýði þurfum við að framkvæma marktektarpróf. Við setjum fram núlltilgátu: $H_0: b_{x1} = b_{x2} = b_{x3} = 0$. Aðfallsgreiningarforritið gefur okkur t -próf fyrir hvern og einn þessara þriggja hallastuðla; við prófum hins vegar aðeins eina tilgátu, þ.e. hvort flokkabreytan Hverfi skýri meðaleinkunnir eftir að tekið hefur verið tillit til menntunarstigs og fjarvista. Þetta samsvarar því að prófa í einu tölfræðiprófi hvort þessir þrjár hallastuðlar séu hærri en 0,0 í þýði, einn eða fleiri, samanber núlltilgátuna. Þetta má gera með því að athuga þá viðbótarskýringu sem staðgengilsbreyturnar þrjár veita, þ.e. bera saman líkön með og án staðgengilsbreytanna. Í okkar tilfalli samsvarar þetta því að bera saman líkan 1 og líkan 2 og setja niðurstöðurnar inn í formúluna fyrir F -próf á viðbótarskýringu sem sést hér til hliðar.

$$F_{inc} = \frac{n-k-1}{q} \cdot \frac{R_1^2 - R_0^2}{1 - R_1^2}$$

F -prófið dreifist með q og $n-k-1$ frígráðu, þar sem q er fjöldi viðbótarbreyta og k er heildarfjöldi frumbreyta. Í okkar tilfalli eru þetta 3 og 14 frígráður. Við erum með R^2 fyrir bæði líkan 1 og líkan 2 og getum því sett beint inn í formúluna. Niðurstaðan er $F(3, 14) = 0,162$ sem má fletta upp í viðeigandi töflu. Niðurstaðan er langt frá því að vera marktæk og því hæpið að álykta sem svo að munur sé á milli hverfa eftir að tekið hefur verið tillit til menntunarstigs og fjarvista.

$$F_{inc} = \frac{20-5-1}{3} \cdot \frac{0,770-0,762}{1-0,770}$$

$$= \frac{14 \cdot 0,008}{3 \cdot 0,23} = 0,112 / 0,69 = 0,162$$

Við gætum viljað vita hvort Hverfi skýri samræmdar meðaleinkunnir skóla þegar ekki er tekið tillit til menntunarstigs skólahverfis og fjölda óheimilla fjarvista. Þetta samsvarar því að setja staðgengilsbreyturnar þrjár einar inn í líkanið, sbr. líkan 3. Hér þarf ekki að skoða viðbótarskýringu þar sem þetta eru einu breyturnar í líkaninu. Marktektarpróf er því gefið beint með F -prófinu fyrir allt líkanið sem birt er neðst í töflunni.

Hverfi	Útreikningar	Spágildi
1	5,95-0,32-0-0,94-0-0,86-0	5,95
2	5,95-0,32-1-0,94-0-0,86-0	5,64
3	5,95-0,32-0-0,94-1-0,86-0	5,01
4	5,95-0,32-0-0,94-0-0,86-1	5,09

Við getum notað hallastuðlana í líkani 3 til að reikna spágildin fyrir öll fjögur gildi Hverfis. Útreikningar og spágildi sjást í töflunni hér til hliðar. Ef við berum þetta saman við meðaltöl hverfanna sjáum við að spágildin samsvara nákvæmlega meðaltölunum. Þetta er auðvitað eins og vera ber ef „dummy“-kóðunin varðveitir allar þær upplýsinga sem fylgnar eru í flokkabreytunni.

Hverfi	Meðaltal	Staðal-	
		frávik	Fjöldi
1	5,95	0,424	2
2	5,64	0,375	8
3	5,01	0,607	6
4	5,09	0,344	4