

Regresioni i shumëfishtë

Koeficientët e kushtëzuar, krahasimi i modeleve dhe informacioni i përbashkët parashikues

Ratiomera Statistics

Përmbajtja

Përmbledhja e faqes së kursit	2
Një ekuacion, disa role	2
Testet dhe krahasimi i modeleve	2
Ndërveprimet dhe diagnostikimet	3
Referencë e zgjeruar	4
Qëllimi dhe bazat	4
Idetë kryesore	4
Udhëzuesi i formulave	5
Si lexohet figura shpjeguese	8
Lista e kontrollit për interpretim	8
Si lidhet kjo temë me të tjerat	9

Përmbledhja e faqes së kursit

Regresioni i shumëfishtë modelon një rezultat sasior nga disa terma parashikues njëkohësisht. Çdo koeficient është i kushtëzuar: ai përshkruan një ndryshim të përshtatur që lidhet me një ndryshore parashikuese, ndërsa termat e tjerë në ekuacion mbahen të pandryshuar.

Një ekuacion, disa role

Për dy ndryshore parashikuese sasiore,

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i.$$

Tabela 1: Modeli i bashkon kontributet parashikuese, por secili term ruan një kuptim të caktuar të kushtëzuar.

Pjesa	Interpretimi
Prerja	Mesatarja e përshtatur kur çdo ndryshore parashikuese sasiore është zero dhe çdo ndryshore parashikuese kategorike është në nivelin e saj referues
Koeficienti sasior	Ndryshimi i përshtatur i rezultatit për një njësi shtesë të ndryshores parashikuese, i kushtëzuar nga termat e tjerë
Koeficienti tregues	Diferenca e kushtëzuar nga kategoria referuese e emërtuar te vlerat referuese sasiore
Koeficienti i ndërveprimit	Ndryshimi në një pjerrësi ose diferencë grupi ndërsa ndryshon një ndryshore tjetër
Vlera e përshtatur dhe reziduali	Parashikimi i bashkuar i modelit dhe diferenca mes vlerës së vëzhguar dhe asaj të përshtatur për një rast

Qendërzimi i një ndryshoreje parashikuese sasiore zbret një referencë kuptimplote, si mesatarja e saj në kampion. Kjo e ndryshon prerjen dhe çdo term të rendit më të ulët që përfshihet në një ndërveprim, por nuk i ndryshon vlerat e përshtatura ose përshtatjen e përgjithshme. Qendërzimi e përmirëson interpretimin. Ai nuk e heq mbivendosjen mes ndryshoreve parashikuese.

Testet dhe krahasimi i modeleve

- Testi t i një koeficienti individual pyet nëse një koeficient i popullatës është zero, duke u kushtëzuar nga termat e tjerë të përfshirë.
- Testi global F pyet nëse të gjitha pjerrësitë e testuara janë zero së bashku.
- Testi F i modeleve të ndërfutura pyet nëse një bllok i caktuar termash të shtuar e përmirëson përshtatjen përtej një modeli më të vogël.
- R^2 nuk mund të zvogëlohet kur shtohen ndryshore parashikuese. R^2 i përshtatur përfshin një dënim për ndërlikimin dhe mund të zvogëlohet.
- AIC-ja krahason ekuilibrin relativ mes përshtatjes dhe ndërlikimit të modeleve të kandidatit të shënuara. Nuk identifikon një model të vërtetë ose shkakësor.

Mbivendosja e ndryshoreve parashikuese mund t'i bëjë vlerësimet individuale më pak të sakta. Faktori i fryrjes së variancës përmbledh sa fort shpjegohet në mënyrë lineare një ndryshore parashikuese nga ndryshoret e tjera parashikuese. Është diagnostikim për pasigurinë e koeficientit, jo rregull universal i llojit “kalon/dështon”.

Ndërveprimet dhe diagnostikimet

Kur ka një ndërveprim, interpretoji pjerrësitë e kushtëzuara ose vlerat e përshtatura në kombinime kuptimplota. Mos e interpreto koeficientin e efektit kryesor si një efekt të vetëm universal nëpër të gjitha nivelet e ndryshores që ndërvepron.

Përdori grafikët e rezidualëve kundrejt vlerave të përshtatura dhe grafikët normalë Q-Q për të shqyrtuar formën e modeluar, shpërndarjen dhe formën e përgjithshme të gabimit. Leverage-i përshkruan kombinime të pazakonta të ndryshoreve parashikuese. Distanca e Cook-ut bashkon pozicionin parashikues dhe informacionin e rezidualit për t'u dhënë përparësi kontrolleve të ndjeshmërisë. Rastet e vlefshme nuk fshihen vetëm sepse një diagnostikim është i madh.

Rruga nga Temat 4 deri në 7 është një histori e lidhur: korrelacioni përshkruan lëvizjen në çift, regresioni i thjeshtë i jep asaj lëvizjeje drejtim dhe njësi, korrelacioni i pjesshëm e bën të dukshme përshtatjen dhe regresioni i shumëfishtë vendos disa marrëdhënie të kushtëzuara në një ekuacion të përshtatur. Tema 8 tregon se krahasimet e grupeve kategorike i përkasin së njëjtës familje të përgjithshme të modeleve lineare.

Referencë e zgjeruar

Qëllimi dhe bazat

Regresioni i shumëfishtë linear modelon mesataren e kushtëzuar të një rezultati numerik duke përdorur dy ose më shumë ndryshore parashikuese. Ai e zgjeron idenë e vijës së përshtatur nga regresioni i thjeshtë. Në vend që të lëvizë përgjatë një boshti të vetëm parashikues, rezultati i përshtatur mund të ndryshojë në disa dimensione parashikuese. Secili koeficient përshkruan ndryshimin e përshtatur që lidhet me një ndryshore parashikuese, ndërsa ndryshoret e tjera parashikuese në model mbahen të pandryshuara.

Kjo shprehje e fundit është një rregull krahasimi dhe jo një veprim fizik. Merr një model që parashikon pikëzimin e arsytimit nga orët e praktikës dhe pikëzimi paraprak. Koeficienti i praktikës krahason raste që ndryshojnë me një orë praktike, por kanë të njëjtin pikëzim paraprak të modeluar. Koeficienti i pikëzimit paraprak krahason raste që ndryshojnë me një njësi në atë pikëzim, por kanë të njëjtat orë praktike të modeluara. Nëse këto krahasime mbështeten mirë, varet nga kombinimet e vrojtuar të ndryshoreve parashikuese dhe përshtatshmëria e modelit linear.

Ndryshoret parashikuese mund të kenë informacion të përbashkët. Orët e praktikës dhe përgatitja paraprake mund të lidhen me të njëjtat pjesë të rezultatit dhe me njëra-tjetrën. Pjerrësia e regresionit të thjeshtë përfshin të gjithë ndryshueshmërinë e rezultatit që përputhet me ndryshoren e vetme parashikuese. Koeficienti i regresionit të shumëfishtë izolon përbërësin linear të kushtëzuar të ndryshores parashikuese, duke pasur parasysh të tjerat. Ndryshimi mes koeficientit të thjeshtë dhe atij të kushtëzuar është i pritshëm dhe informativ, por kërkon interpretim përmbajtësor.

Pjesa e modelit	Kuptimi	Pyetja që duhet bërë
Konstanta b_0	Rezultati i përshtatur kur çdo ndryshore parashikuese numerike është zero dhe ndryshoret kategoriale janë në nivelet referuese	A kanë kuptim dhe a përfaqësohen këto vlera referuese?
Koeficienti numerik b_j	Dallimi i përshtatur i rezultatit për një rritje prej një njësie në ndryshoren parashikuese, duke mbajtur të tjerat të pandryshuara	Cilat ndryshore mbahen të pandryshuara dhe në çfarë njësisë?
Koeficienti tregues	Dallimi i përshtatur nga një kategori referuese e deklaruar	Cila kategori është referuese?
Koeficienti i ndërveprimit	Ndryshimi i pjerrësisë së njërës ndryshore nëpër vlerat ose grupet e një ndryshoreje tjetër	Cila pjerrësi e kushtëzuar po ndryshohet?

Idetë kryesore

Fillo me një model të zgjedhur nga pyetja kërkimore dhe jo nga një kërkim mekanik nëpër çdo ndryshore të disponueshme. Një ndryshore parashikuese mund të përfaqësojë ekspozimin qendror, një kontroll të planifikuar, një krahasim grupesh ose një term të nevojshëm për ta përfaqësuar formën funksionale. Shpjegoje secilin rol. Shtimi i një ndryshoreje e ndryshon pyetjen që merr përgjigje nga secili koeficient i kushtëzuar, prandaj dy modele me grupe të ndryshme ndryshoresh parashikuese nuk janë përshkrime të këmbyeshme.

Ndryshoret kategoriale hyjnë përmes ndryshoreve treguese. Me tri formate ushtrimesh, një kategori bëhet referuese dhe dy tregues i krahasojnë formatet e tjera me të. Ndryshimi i kategorisë referuese i ndryshon konstantën dhe krahasimet e shtypura, por nuk i ndryshon vlerat e përshtatura. Referenca duhet deklaruar në tabela dhe në tekst.

Ndërveprimi do të thotë se lidhja e kushtëzuar e njëres ndryshore parashikuese ndryshon nëpër vlerat e tjetrës. Në një ndërveprim praktikë me format, nuk ka një pjerrësi të vetme të praktikës për të gjitha formatet. Koeficienti kryesor i praktikës është pjerrësia brenda formatit referues; secili koeficient ndërveprimi tregon si ndryshon pjerrësia e një formati tjetër. Interpretimi së bashku koeficientët përbërëse dhe trego vija të përshtatura ose vlera të parashikuara.

Niveli i vlerësimit	Madhësi ose paraqitje e dobishme	Çfarë kontribuon
Koeficienti individual	Vlerësimi, gabimi standard, intervali, testi t	Drejtimi, madhësia dhe pasiguria e kushtëzuar
Blloku i shtuar i ndryshoreve	Testi F i modeleve të folezuara dhe ndryshimi në R^2	Nëse blloku shton ndryshueshmëri të modeluar të rezultatit
Modeli i plotë	R^2 , R^2 i përshtatur, testi i përgjithshëm F	Përshtatja e kampionit dhe evidenca e përbashkët për grupin e ndryshoreve parashikuese
Përshtatshmëria e modelit	Grafikët e rezidualeve, kuantileve, levës dhe ndikimit	Nëse forma e përshtatur dhe supozimet për pasigurinë janë të besueshme

R^2 nuk mund të ulet kur shtohen ndryshore parashikuese, edhe nëse ato japin pak informacion të dobishëm. R^2 i përshtatur përfshin një ndëshkim për numrin e ndryshoreve parashikuese dhe mund të ulet. Kriteret e informacionit si AIC e baraspeshojnë gjithashtu përshtatjen me ndërlikueshmërinë e modelit, por krahasimet kanë kuptim vetëm mes modeleve të përshtatura me të njëjtin rezultat dhe të njëjtat vrojtime. Asnjë numër i vetëm përshtatjeje nuk i zëvendëson diagnostikën e rezidualeve ose gjykimin përmbajtësor.

Mbivendosja e fortë e ndryshoreve parashikuese do të thotë se ato përmbajnë informacion linear që mbivendoset shumë. Ajo mund t'i zmadhojë gabimet standarde të koeficienteve dhe t'i bëjë vlerësimet individuale të paqëndrueshme, ndërsa vlerat e përshtatura mbeten të dobishme. Në një model tjetër të përshtatshëm, kjo nuk krijon vetvetiu anshmëri. Shqyrto marrëdhëniet mes ndryshoreve parashikuese, pasigurinë e koeficienteve dhe dizajnin. Mos e hiq një ndryshore të nevojshme konceptualisht vetëm për ta bërë një koeficient tjetër statistikisht të rëndësishëm dhe mos zbato një prag numerik universal që materialet e Statistikës 1 të dhëna për këtë temë nuk e mbështesin.

Supozimet për rezidualet i zgjerojnë ato të regresionit të thjeshtë: një mesatare lineare e kushtëzuar e përshtatshme, gabime të pavarura, variancë e përshtatshme nëpër vlerat e përshtatura dhe një shpërndarje reziduale e mjaftueshme për inferencën e synuar. Rastet me ndikim mund të ndryshojnë disa koeficiente. Ekstrapolimi mund të ndodhë edhe në kombinime ndryshoresh parashikuese, edhe kur secila vlerë veçmas është brenda diapazonit të vërtetuar.

Udhëzuesi i formulave

Për p ndryshore parashikuese, modeli i popullatës është:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + \varepsilon_i$$

Vlera e përshtatur e kampionit përdor koeficientët e vlerësuar, ndërsa reziduali mbetet vlera e vërtetuar minus vlerën e përshtatur:

$$\hat{y}_i = b_0 + \sum_{j=1}^p b_j x_{ji}, \quad e_i = y_i - \hat{y}_i$$

Për dy ndryshore sasiore parashikuese, pjerrësitë e kushtëzuara mund të shprehen përmes tri korrelacioneve dyshe dhe devijimeve standarde të ndryshoreve:

$$b_1 = \frac{r_{Y1} - r_{Y2}r_{12}}{1 - r_{12}^2} \frac{s_Y}{s_{X_1}}, \quad b_2 = \frac{r_{Y2} - r_{Y1}r_{12}}{1 - r_{12}^2} \frac{s_Y}{s_{X_2}}.$$

Zbritja përfaqëson informacionin e korrelacionit që ndahet me ndryshoren tjetër parashikuese, ndërsa raporti i devijimeve standarde e kthen rezultatin në njësi rezultati për njësi parashikuese. Nëse $|r_{12}| = 1$, emëruesi është zero dhe dy pjerrësitë e veçanta nuk mund të vlerësohen nga ai model.

Gabimi standard i rezidualeve raporton shpërhapjen tipike të pashpjeguar në njësitë e rezultatit:

$$s_e = \sqrt{\frac{\sum_{i=1}^n e_i^2}{n - p - 1}}.$$

Këtu p numëron parametrat parashikues pa konstantën. Një ndryshore kategoriale mund të kërkojë më shumë se një parametër.

Për një ndryshore sasiore parashikuese, koeficienti i standardizuar është

$$\hat{\beta}_j = b_j \frac{s_{X_j}}{s_Y}.$$

Ai përshkruan ndryshimin e kushtëzuar të përshtatur në devijime standarde të rezultatit për një dallim prej një devijimi standard në ndryshoren parashikuese. Ndryshe nga korrelacioni dysh, ai kushtëzohet nga termat e tjerë të modelit dhe nuk kufizohet në intervalin nga -1 deri në $+1$.

Koeficienti i përcaktimit krahason shumat reziduale dhe të përgjithshme të katrorëve:

$$R^2 = 1 - \frac{SS_{\text{residual}}}{SS_{\text{total}}}$$

R^2 i përshtatur merr parasysh madhësinë e kampionit n dhe numrin e ndryshoreve parashikuese p :

$$R_{\text{adjusted}}^2 = 1 - (1 - R^2) \frac{n - 1}{n - p - 1}$$

Statistika e përgjithshme F krahason mesataren e katrorëve të modelit me mesataren e katrorëve rezidualë:

$$F = \frac{SS_{\text{model}}/p}{SS_{\text{residual}}/(n - p - 1)}$$

Hipoteza e përgjithshme zero është $H_0 : \beta_1 = \dots = \beta_p = 0$. Një rezultat statistikisht i rëndësishëm thotë se, sipas modelit, të paktën një koeficient i popullatës pa konstantën ndryshon nga zeroja, por nuk tregon cili koeficient. Për një koeficient të vetëm përdoret

$$t = \frac{b_j}{SE(b_j)}, \quad df = n - p - 1.$$

Ky test lidhet me koeficientin j të kushtëzuar nga termat e saktë të tjerë në model. Gabimi i tij standard nuk është gabimi standard i rezidualeve.

Për dy modele të folezuara, kontributi i shtuar i q ndryshoreve të reja parashikuese mund të testohet duke krahasuar uljen e tyre në shumën reziduale të katrorëve me mesataren reziduale të katrorëve të modelit më të madh:

$$F = \frac{(SS_{\text{residual, reduced}} - SS_{\text{residual, full}})/q}{SS_{\text{residual, full}}/(n - p - 1)}$$

Modeli i reduktuar duhet të përftohet duke i vendosur në zero koeficientët e shtuara të modelit të plotë dhe të dy modelet duhet të përdorin të njëjtin rezultat dhe të njëjtat raste të analizuara. Në këtë formulë, p është numri i parametrave parashikues pa konstantën në modelin e plotë, prandaj emëruesi përdor shkallët e lirisë të rezidualeve të modelit të plotë. Për një ndryshore të shtuar parashikuese, korrelacioni gjysmëpartial jep të njëjtën rritje të përshtatjes:

$$sr_j^2 = R_{\text{larger}}^2 - R_{\text{smaller}}^2 = \Delta R^2.$$

Për një korrelacion gjysmëpartial, rezidualizohet vetëm ndryshorja parashikuese kandidate. Korrelacioni i pjesshëm në Temën 6 i rezidualizon të dyja ndryshoret qendrore.

Ndryshoret kategoriale kërkojnë tregues. Me një konstantë dhe k kategori, përdor $k - 1$ tregues. Për një ndryshore sasiore parashikuese X dhe një tregues dyvlerësh D , modeli mbledhës është

$$\hat{Y} = b_0 + b_1X + b_2D.$$

Kur $D = 0$, vija e përshtatur është $b_0 + b_1X$. Kur $D = 1$, ajo është $(b_0 + b_2) + b_1X$. Vijat janë paralele dhe b_2 është dallimi i përshtatur mes grupeve të njëjta vlerë të X .

Ndërveprimi lejon që pjerrësitë të ndryshojnë:

$$\hat{Y} = b_0 + b_1X + b_2D + b_3XD.$$

Pjerrësia e grupit referues është b_1 , pjerrësia e grupit krahasues është $b_1 + b_3$ dhe b_3 është dallimi mes pjerrësive. Koeficienti b_2 është dallimi mes grupeve kur $X = 0$, prandaj qendërzimi i X mund t'i japë atij krahasimi një pikë referimi më të dobishme.

Kriteri informues i Akaike-s, i përdorur për krahasimin e modeleve kandidate, është

$$AIC = -2\log(L) + 2k,$$

ku L është gjasësia e përshtatur dhe k është numri i parametrave të vlerësuar në gjasësi. Një AIC më i vogël tregon një ekuilibër relativ më të mirë mes përshtatjes dhe ndërlikimit vetëm mes modeleve të përshtatura me të njëjtin rezultat dhe të njëjtat raste. Ai nuk vërteton se modeli i zgjedhur është i vërtetë, shkakor ose i saktë për të dhëna të reja.

Madhësia	Pyetja	Kufizimi thelbësor
R^2	Sa ndryshueshmëri të rezultatit në kampion përfaqëson ky model i përshtatur?	Nuk mund të ulet kur i shtohen terma të njëjtit model OLS
R^2 i përshtatur	A e tejkalon përfitimi i shtuar në përshtatjen e kampionit ndëshkimin për parametrat brenda kampionit?	Nuk është vlerësim në raste të reja
F i modeleve të folezuara	A e përmirësojnë bashkërisht përshtatjen koeficientët e shtuar?	Kërkon modele vërtet të folezuara dhe të njëjtat raste
AIC	Cili kandidat i deklaruar ka ekuilibrin relativ më të mirë mes përshtatjes dhe ndërlikimit?	Nuk ka prag universal kalimi

Këto formula e përshkruajnë numerikisht përshtatjen dhe pasigurinë e kampionit. Ato nuk përcaktojnë cilat ndryshore parashikuese kanë kuptim shkencor ose nëse një koeficient i kushtëzuar ka interpretim shkallor.

Si lexohet figura shpjeguese

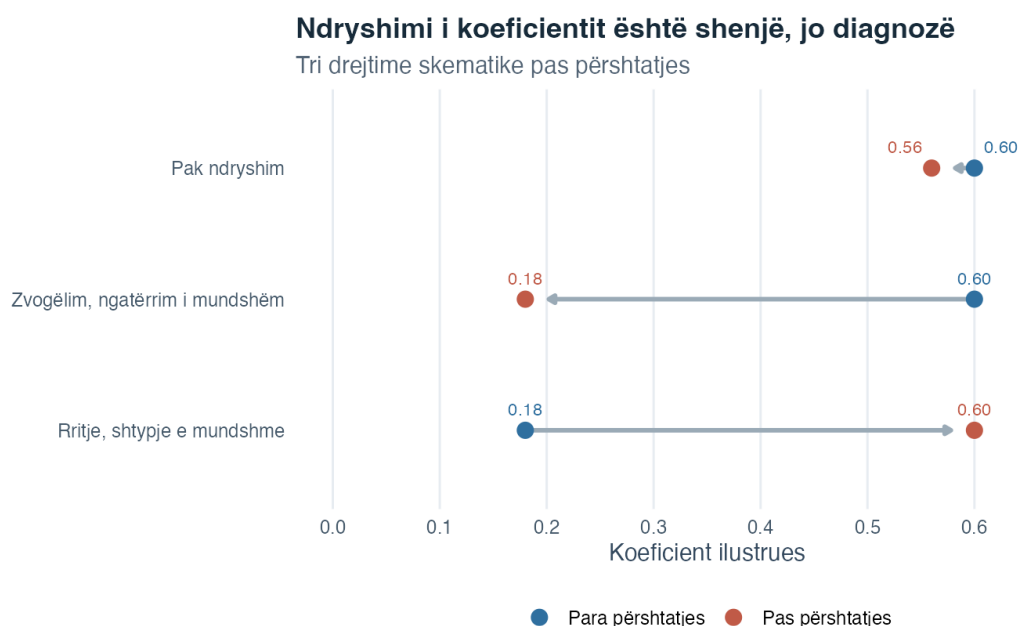


Figura 1: Tri krahasime horizontale të koeficienteve tregojnë ndryshim të vogël, zvogëlim dhe rritje mes vlerave para dhe pas përshtatjes në një model regresioni të shumëfishtë.

Secili rresht krahason një koeficient blu para përshtatjes me një koeficient portokalli pasi në model hyjnë ndryshore të tjera parashikuese. Në rreshtin e sipërm, 0.60 ndryshon në 0.56. Rezultati i kushtëzuar është i ngjashëm me rezultatin e papërshtatur, prandaj ndryshoret e shtuara e ndryshuan pak këtë koeficient. Kjo nuk provon se ndryshoret janë të parëndësishme; ato mund ta përmirësojnë parashikimin ose të kenë rëndësi për koeficiente të tjera.

Në rreshtin e mesëm, 0.60 zvogëlohet në 0.18. Ndryshoret parashikuese qendrore kishte shumë informacion të lidhur me rezultatin të përbashkët me ndryshoret e shtuara. Ngatërrimi është një shpjegim i mundshëm përmbajtësor, por grafiku e etiketon si të mundshëm, sepse vetëm lëvizja e koeficientit nuk mund ta identifikojë një rol shkallor. Mund të kenë rëndësi edhe mbivendosja e matjes, përzgjedhja, forma funksionale ose ndryshueshmëria e kampionimit.

Në rreshtin e poshtëm, 0.18 rritet në 0.60. Përshtatja ka zbuluar një lidhje të kushtëzuar më të fortë, një model që shpesh përshkruhet si shtypje e mundshme. Edhe këtu etiketa është shenjë dhe jo përfundim. Shqyrto marrëdhëniet mes ndryshoreve parashikuese, intervalet e koeficienteve, dizajnin dhe diagnostikën e modelit. Largësia horizontale paraqet lëvizjen numerike; ajo nuk tregon pasigurinë, prandaj një analizë e plotë kërkon edhe intervale besimi.

Lista e kontrollit për interpretim

Trego ndryshoren e rezultatit, çdo ndryshore parashikuese, njësitë, kodimin dhe kategoritë referuese. Shpjego pse përfshihet secila ndryshore parashikuese dhe nëse ndërveprimet ishin planifikuar. Shqyrto shpërndarjet, marrëdhëniet mes ndryshoreve parashikuese, të dhënat që mungojnë dhe kombinimet e mbështetura. Ktheje secilin koeficient në një dallim të përshtatur të kushtëzuar dhe emërto çfarë mbahet e pandryshuar. Për ndërveprimet, raportoj pjerresë të kushtëzuara ose vlera të parashikuara në vend që të interpretosh një term të vetëm.

Krahaso modelet e folezuara vetëm kur përdorin të njëjtat vrojtme dhe të njëjtin rezultat. Raporto vlerësimet dhe intervalet e koeficienteve, R^2 , R^2 të përshtatur, krahasimet përkatëse të modeleve dhe diagnostikën. Kontrolllo formën e rezidualëve, ndryshimin e variancës, levën, ndikimin dhe mbivendosjen e ndryshoreve parashikuese. Mbajti të ndara parashikimin, lidhjen dhe shkakësinë. Nëse koeficientët ndryshojnë mes modeleve, përshkruaje ndryshimin dhe heto burimin e tij në vend që t'i vendosësh automatikisht një etiketë shkakore.

Si lidhet kjo temë me të tjerat

Regresioni i shumëfishtë i mbledh mjetet e mëparshme të lidhjes në një kornizë. Kovarianca dhe korrelacioni paraqitën ndryshueshmërinë e përbashkët lineare. Regresioni i thjeshtë e ktheu atë model në një ekuacion të drejtuar të përshtatur me reziduale. Korrelacioni i pjesshëm tregoi se «mbajtja e një ndryshoreje të pandryshuar» mund të kuptohet duke rezidualizuar ndryshoret qendrore. Koeficienti i regresionit të shumëfishtë zbaton të njëjtën logjikë të kushtëzuar, duke ruajtur njësinë e rezultatit dhe duke lejuar që disa ndryshore parashikuese të vlerësohen së bashku.

Analiza e variancës është shprehja vijuese e kësaj kornize. Përkatësia në grup mund të përfaqësohet me ndryshore treguese, prandaj krahasimi i mesatareve të grupeve bëhet model regresioni me informacion kategorial. Testi F i ANOVA-s pyet nëse termat e grupit shpjegojnë së bashku ndryshueshmëri të rezultatit përtej ndryshueshmërisë reziduale. Ajo që në fillim duket si metodë e veçantë është kështu një pamje tjetër e të njëjtit arsyetim të modelit të përgjithshëm linear.

ID-ja e dokumentit: topic-07-multiple-regression-summary-sq

Shënime 1 nga 3

Regresioni i shumëfishtë | Tema 7

Shënime 2 nga 3

Regresioni i shumëfishtë | Tema 7

Shënime 3 nga 3

Regresioni i shumëfishtë | Tema 7