

Kovarianca dhe korrelacioni

Të kuptosh si ndryshojnë së bashku dy ndryshore

Ratiomera Statistics

Përmbajtja

Përmbledhja e faqes së kursit	2
Nga devijimet e çiftuara te korrelacioni	2
Çfarë duhet kontrolluar para interpretimit të r	2
Ura drejt regresionit	3
Referencë e zgjeruar	4
Qëllimi dhe bazat	4
Idetë kryesore	4
Udhëzuesi i formulave	5
Si lexohet figura shpjeguese	7
Lista e kontrollit për interpretim	7
Si lidhet kjo temë me të tjerat	8

Përmbledhja e faqes së kursit

Kovarianca dhe korrelacioni përshkruajnë si ndryshojnë së bashku dy ndryshore sasiore në raste të çiftuara siç duhet. Gjithmonë shqyrto fillimisht diagramin e shpërndarjes. Koeficienti është përmbledhje e atij diagrami, jo zëvendësim i tij.

Nga devijimet e çiftuara te korrelacioni

Për vrojtimitet e çiftuara (x_i, y_i) , kovarianca e kampionit është

$$s_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1}.$$

Prodhimet janë pozitive kur të dyja devijimet kanë të njëjtin drejtim dhe negative kur kanë drejtime të kundërta. Prandaj shenja tregon drejtimin. Madhësia varet nga njësitë e matjes, gjë që e bën të vështirë krahasimin e kovariancave mes shkallëve të ndryshme.

Korrelacioni i Pearson-it e standardizon kovariancën:

$$r = \frac{s_{XY}}{s_X s_Y}.$$

Ai shtrihet nga -1 deri në $+1$. Shenja jep drejtimin e modelit linear, ndërsa $|r|$ përshkruan sa afër një vije të drejtë qëndrojnë pikat. Një vlerë afër zeros do të thotë pak lidhje lineare; një model i fortë i lakuar mund të ketë prapë korrelacion të vogël të Pearson-it.

Çfarë duhet kontrolluar para interpretimit të r

Tabela 1: Lexo modelin e të dhënave dhe dizajnin e studimit para se ta lexosh koeficientin.

Kontrolli	Pse ka rëndësi
Rreshtat e çiftuar	Çdo x_i duhet të qëndrojë me y_i përkatës
Trajta e diagramit të shpërndarjes	r i Pearson-it përmbledh lidhjen lineare, jo çdo formë varësie
Pikat e pazakonta dhe me ndikim	Një pikë mund ta ndryshojë ndjeshëm koeficientin
Intervali i vrojtuar	Kufizimi i intervalit të njëres ndryshore mund ta dobësojë koeficientin e vrojtuar
Nëngrupet	Modeli i përbashkët mund të fshehë modele të ndryshme brenda grupeve
Matja dhe kampionimi	Gabimi, përzgjedhja dhe mbulimi i dobët mund ta shtrembërojnë lidhjen

Korrelacioni i rangjeve të Spearman-it e zbaton idenë e Pearson-it te rangjet. Ai përshkruan model monoton, që do të thotë se njëra ndryshore priret të lëvizë vazhdimisht lart ose poshtë kur ndryshon tjetra, pa kërkuar vijë të drejtë. Vlerave të barabarta u caktohen rangje të përbashkëta të përshtatshme.

Për inferencë rreth korrelacionit të Pearson-it në popullatë nën modelin që përdoret këtu,

$$t = r \sqrt{\frac{n-2}{1-r^2}}, \quad df = n - 2.$$

Testi pyet nëse korrelacioni linear në popullatë mund të jetë zero nën supozimet e deklaruar. Nuk e kthen një lidhje vrojtuese në efekt shkakor.

Ura drejt regresionit

Korrelacioni është simetrik dhe pa njësi. Regresioni i thjeshtë e ruan të njëjtin model të çiftuar, por u jep ndryshoreve role të ndryshme: njëra bëhet parashikuese dhe tjetra rezultat. Tema 5 e shndërron renë e pikave në vijë të përshtatur, i rikthen njësitë fillestare dhe studion rezidualet që tregojnë çfarë nuk arrin të përshkruajë vija.

Referencë e zgjeruar

Qëllimi dhe bazat

Kovarianca dhe korrelacioni përshkruajnë si ndryshojnë së bashku dy ndryshore numerike të të njëjtat raste. Çdo rast duhet të japë një çift të përputhur (x_i, y_i) . Diagrami i shpërndarjes është pika e domosdoshme e fillimit: vendose një ndryshore në boshtin horizontal, tjetrën në boshtin vertikal dhe paraqite çdo rast me një pikë. Reja që krijohet tregon drejtimin, formën, forcën, grumbullimet dhe vërtetimet e pazakonta në një mënyrë që një koeficient i vetëm nuk mund ta bëjë.

Një lidhje pozitive do të thotë se rastet me vlera më të mëdha të x priren të kenë vlera më të mëdha të y . Një lidhje negative do të thotë se vlerat më të mëdha të x priren të shoqërohen me vlera më të vogla të y . Një vlerë pranë zeros për një koeficient linear tregon pak lidhje lineare, jo domosdoshmërisht mungesë marrëdhënieje. Modelet e lakuara, nëngrupet e ndara ose një diapazon i kufizuar mund ta bëjnë koeficientin të paplotë ose çorientues.

Kovarianca fillon me devijimet nga dy mesataret. Një rast jep kontribut pozitiv kur të dyja vlerat janë mbi mesataret e tyre ose të dyja janë poshtë. Ai jep kontribut negativ kur njëra vlerë është mbi mesataren e vet dhe tjetra poshtë. Mesatarizimi i këtyre prodhimeve të kryqëzuara jep kovariancën e kampionit. Shenja e saj është informative, por madhësia varet nga njësitë e matjes. Për shembull, matja e orëve në minuta e ndryshon kovariancën edhe pse përputhja bazë e rasteve nuk ka ndryshuar.

Veçoria	Kovarianca	Korrelacioni i Pearson-it
Drejtimi	Shenja tregon bashkëndryshim pozitiv ose negativ	Shenja tregon lidhje lineare pozitive ose negative
Shkalla	Varet nga njësitë e të dyja ndryshoreve	Pa njësi, sepse të dyja ndryshoret standardizohen
Diapazoni numerik	Nuk kufizohet në një interval të pandryshueshëm	Gjithmonë mes -1 dhe 1
Roli kryesor	Bllok ndërtues për lidhjen dhe regresionin	Përmbledhje e krahasueshme e drejtimit dhe forcës lineare

Idetë kryesore

Korrelacioni i Pearson-it r e standardizon kovariancën duke e pjesëtuar me dy devijimet standarde të kampionit. Një vlerë pranë 1 tregon se pikat ndjekin një model të fortë pozitiv në vijë të drejtë. Një vlerë pranë -1 tregon një model të fortë negativ në vijë të drejtë. Një vlerë pranë zeros tregon pak model drejtvizor. Koeficienti përshkruan kampionin. Korrelacioni i popullatës zakonisht shënohet me ρ , ndërsa inferenca nevojitet kur synohet një përfundim për popullatën.

Korrelacioni i rangjeve të Spearman-it i zëvendëson vlerat e vrojtues me rangjet e tyre dhe vlerëson nëse ndryshoret ndjekin një marrëdhënie **monotone**. Monotone do të thotë se prirja ecën vazhdimisht në një drejtim: ndërsa një ndryshore rritet, tjetra përgjithësisht rritet ose përgjithësisht ulët. Modeli mund të lakohet duke e ruajtur atë rend. Prandaj korrelacioni i Spearman-it mund të mbetet i lartë për një lidhje monotone të lakuar, të cilën korrelacioni i Pearson-it e përmbledh më pak plotësisht. Asnjë nga koeficientët nuk e përfaqëson mirë një marrëdhënie në formë U-je, sepse drejtimi kthehet nëpër diapazon.

Pyetja diagnostikuese	Çfarë të shqyrtosh	Pse e ndryshon interpretimin
A është forma afërsisht lineare?	Diagramin e shpërndarjes dhe një model të mundshëm të lëmuar	r i Pearson-it përmbledh një prirje drejtvizore

Pyetja diagnostikuese	Çfarë të shqyrtosh	Pse e ndryshon interpretimin
A kanë ndikim pikat e pazakonta?	Diagramin me pika të etiketuara dhe krahasimin e ndjeshmërisë	Një pikë e largët mund ta ndryshojë drejtimin ose madhësinë
A janë përzier grupet?	Ngjyra ose panele për grupe me kuptim	Lidhja e bashkuar mund të ndryshojë nga lidhjet brenda grupeve
A është i kufizuar diapazoni i vërtetuar?	Diapazonet e ndryshoreve dhe procesin e kampionimit	Ndryshueshmëria e kufizuar mund ta dobësojë koeficientin e vërtetuar
A është e vlefshme përputhja?	Identifikuesit e rasteve dhe kohën e matjes	Korrelacioni kërkon që të dyja vlerat t'i përkasin të njëjtit rast

Korrelacioni nuk vërteton shkakësinë. Një lidhje e vërtetuar mund të pasqyrojë një ndikim të drejtpërdrejtë, drejtim të kundërt të ndikimit, një ndryshore të tretë që lidhet me të dyja, përzgjedhjen në kampion, artefakte të matjes ose rastësinë. Rendi kohor dhe një dizajn kërkimor i besueshëm japin informacion që një koeficient i vetëm nuk mund ta japë. Edhe kur nuk synohet arsyetim shkakor, konteksti përmbajtësor përcakton nëse ndryshoret e çiftuara dhe interpretimi i tyre kanë kuptim.

Kur synohet një përfundim për popullatën, koeficienti i kampionit mund të testohet kundrejt $H_0 : \rho = 0$. Mbye atë test të ndarë nga madhësia dhe kuptimi praktik. Një vlerë e vogël p lidhet me pajtueshmërinë me një korrelacion zero në popullatë sipas modelit; ajo nuk e bën lidhjen të madhe, të rëndësishme ose shkakore.

Udhëzuesi i formulave

Kovarianca e kampionit mesatarizon prodhimet e çiftuara të devijimeve dhe përdor $n - 1$ në emërues:

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Shenja e secilit prodhim tregon nëse dy devijimet drejtohen në të njëjtën anë apo në anë të kundërta. Devijimet e mëdha marrin më shumë peshë, sepse prodhimi i tyre ka madhësi më të madhe.

Korrelacioni i kampionit i Pearson-it e pjesëton kovariancën me prodhimin e dy devijimeve standarde të kampionit:

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

Kur janë llogaritur shumatat e korigjuara, i njëjti koeficient mund të llogaritet drejtpërdrejt si

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}.$$

Të dyja format kërkojnë të njëjtat raste të çiftuara në të dyja ndryshoret. Ndryshimi i rendit të njëjës kolonë pa partneren e saj i shkatërron çiftet dhe e ndryshon pyetjen.

E njëjta llogaritje mund të shkruhet si shumë e prodhimeve të pikëzimeve të standardizuara. Kjo e bën të dukshme se koeficienti nuk ka njësi:

$$r_{xy} = \frac{1}{n-1} \sum_{i=1}^n z_{xi} z_{yi}$$

Për një kampion pa rangje të barabarta, korrelacioni i rangjeve të Spearman-it mund të llogaritet nga dallimi d_i mes dy rangjeve të secilit rast:

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

Kur ka rangje të barabarta, llogarit korrelacionin e Pearson-it mbi rangjet e caktuara. Në çdo formë, koeficienti duhet lexuar bashkë me diagramin e shpërndarjes ose modelin e rangjeve që e prodhoi.

Për ta testuar korrelacionin e popullatës të Pearson-it kundrejt $H_0 : \rho = 0$, përdor

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}, \quad df = n - 2.$$

Hipoteza alternative përcakton nëse sipërfaqja referuese është e njëanshme apo e dyanshme. Llogaritja mbështetet në raste të çiftuara të pavarura, në një marrëdhënie për të cilën mund të mbrohet një përmbledhje lineare e Pearson-it dhe në mungesën e problemeve të dizajnit ose të pikave me ndikim që do ta bënin interpretimin të pavlefshëm.

Rezultati që duhet raportuar	Pyetja së cilës i përgjigjet	Çfarë nuk mund të vërtetojë i vetëm
Forma e diagramit të shpërndarjes	Cili model, cilat grumbullime, cili diapazon dhe cilat pika të pazakonta duken?	Një përfundim për popullatën
r ose r_s	Cili drejtim dhe cila forcë lineare ose monotone shfaqet në kampion?	Shkakësinë
t , df dhe vlera p	Sa pajtohet koeficienti i kampionit i Pearson-it me $\rho = 0$ sipas modelit?	Rëndësinë praktike ose një lidhje të madhe

Si lexohet figura shpjeguese

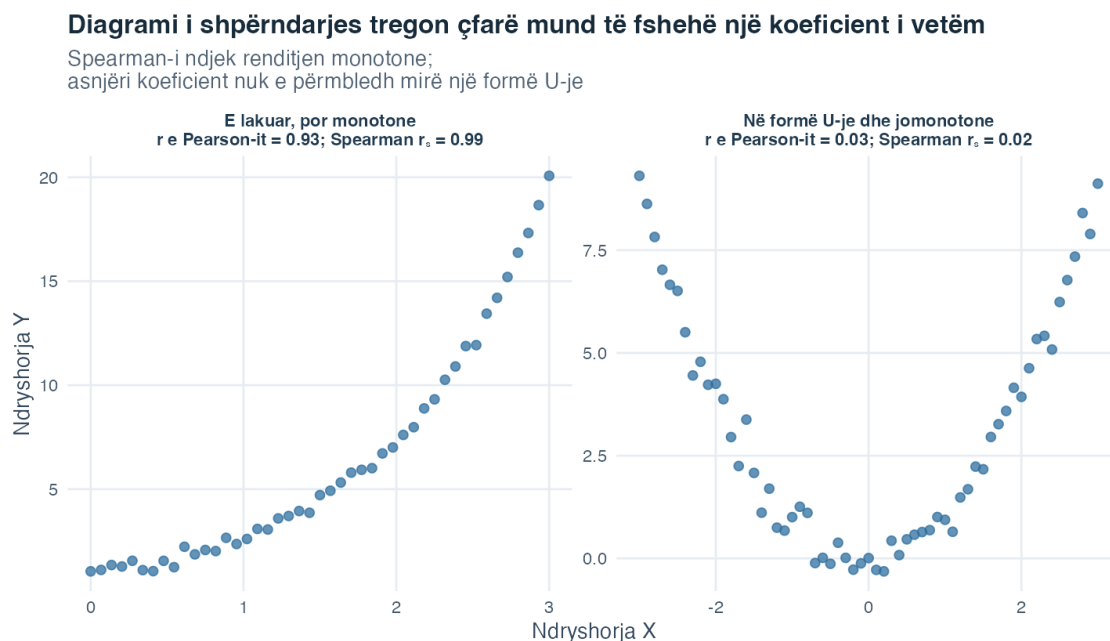


Figura 1: Dy diagrame shpërndarjeje krahasojnë një model monotone të lakuar me një model në formë U-je dhe raportojnë mbi secilin panel korrelacionet e Pearson-it dhe Spearman-it.

Paneli i majtë ngrihet gjatë gjithë diapazonit të vërtetuar. Rritja është e lakuar dhe jo e drejtë, por renditja është shumë e qëndrueshme: vlerat më të mëdha të x pothuajse gjithmonë shoqërohen me vlera më të mëdha të y . Prandaj korrelacioni i Spearman-it është pranë njëshit, sepse rangjet e ruajnë këtë rend në rritje. Edhe korrelacioni i Pearson-it është fort pozitiv, por ai vazhdon të përqendrohet te përbërësi i përgjithshëm drejtvizor i modelit. Lakorja e dukshme të tregon se një përmbledhje drejtvizore nuk e kap çdo veçori.

Paneli i djathtë ka formë U-je. Duke lëvizur nga skaji i majtë drejt qendrës, y ulet ndërsa x rritet. Duke lëvizur nga qendra drejt skajit të djathtë, y rritet ndërsa x rritet. Këto drejtime të kundërta e anulojnë njëra-tjetrën në llogaritjet e Pearson-it dhe Spearman-it, duke prodhuar vlera pranë zeros. Megjithatë, ndryshoret kanë një marrëdhënie të theksuar. Përfundimi i saktë nuk është «nuk ka lidhje». Përfundimi është se as koeficienti linear, as ai monotone nuk e përmbledh mirë këtë formë.

Vijat lidhëse në figurë ndihmojnë për të zbuluar rendin e pikave; ato nuk janë vija regresioni të përshtatura. Koeficientet e shtypur mbi panelet përshkruajnë vlerat e simuluar që paraqiten. Ato janë rezultate mesimore, jo vlerësime nga pjesëmarrës realë. Ky shembull të kujton ta lejosh grafikun të identifikojë formën dhe pastaj ta lejosh koeficientin të përmbledhë veçorinë për të cilën është ndërtuar.

Lista e kontrollit për interpretim

Sigurohu se ndryshoret janë numerike ose se një analizë e bazuar në rangje është e përshtatshme. Verifiko se vlerat janë çiftuar sipas rasteve. Shqyrto një diagram shpërndarjeje para se të llogaritësh koeficientin. Përshkruaj drejtimin, formën, forcën, grumbullimet, diapazonin dhe pikat e pazakonta. Zgjidh korrelacionin e Pearson-it për një përmbledhje lineare me kuptim përmbajtësor dhe korrelacionin e Spearman-it për një përmbledhje monotone të bazuar në rangje. Raporto madhësinë e kampionit dhe koeficientin, dhe shto një interval ose test kur kërkohet përfundim për popullatën.

Mos i emërto kufijtë universalë të pandryshueshëm si të dobët, mesatarë ose të fortë pa kontekst. Kuptimi praktik i një koeficienti varet nga besueshmëria e matjes, fusha, dizajni dhe pasojat. Kontrolllo si ndryshon rezultati kur shqyrtohet një pikë me ndikim ose një nëngrup me kuptim, por

mos i hiq rastet vetëm për ta përmirësuar koeficientin. Shmang foljet shkakore nëse dizajni nuk i mbështet.

Si lidhet kjo temë me të tjerat

Kovarianca është ura nga ndryshueshmëria përshkruese te regresioni. Në regresionin e thjeshtë linear, pjerrësia mund të shkruhet si kovarianca mes ndryshores parashikuese dhe ndryshores së rezultatit, e pjesëtuar me variancën e ndryshores parashikuese. Korrelacioni e standardizon të njëjtën prirje të çiftuar, ndërsa regresioni e ruan njësinë e rezultatit dhe jep ndryshimin e përshtatur në atë rezultat për një ndryshim prej një njësie në ndryshoren parashikuese.

Korrelacioni i pjesshëm pyet më vonë si mbeten të lidhura dy ndryshore pas përshtatjes lineare për një të tretë. Regresioni i shumëfishtë e zgjeron të njëjtën logjikë duke vlerësuar lidhjen e kushtëzuar të secilës ndryshore parashikuese ndërsa të tjerat mbahen të pandryshuara. Edhe analiza e variancës i përket kësaj familjeje: ajo shpjegon ndryshueshmërinë e një rezultati duke përdorur përkatësinë në grup në vend që të fillojë me një ndryshore parashikuese numerike. Pyetja e përbashkët është si përputhet ndryshueshmëria e rezultatit me informacionin që jep një ose më shumë ndryshore parashikuese.

ID-ja e dokumentit: topic-04-covariance-correlation-summary-sq

Shënime 1 nga 3

Kovarianca dhe korrelacioni | Tema 4

Shënime 2 nga 3

Kovarianca dhe korrelacioni | Tema 4

Shënime 3 nga 3

Kovarianca dhe korrelacioni | Tema 4