

Cuprins

<i>Curier înainte</i>	9
Capitolul 1. Inferența statistică – estimarea parametrilor	11
1.1. Ce este inferența statistică?	13
1.2. Indicatori și parametri statistici	14
1.3. Distribuția de eșantionare	16
1.4. Erorile standard	17
1.5. Curba normală (Gauss)	23
1.6. Teorema limitei centrale	24
1.7. Intervalul de încredere pentru media populației	25
1.8. Intervalul de încredere pentru proporția unui eșantion	29
1.9. Intervalul de încredere pentru diferența dintre proporțiile a două eșantioane independente ...	31
1.10. Calcularea intervalului de încredere prin metoda de resamplare <i>bootstrap</i>	32
1.11. Întrebări recapitulative	35
Capitolul 2. Inferența statistică – testarea ipotezelor	37
2.1. Mecanismul inferenței cu privire la ipoteza cercetării	40
2.1.1. <i>Ipoteza cercetării</i>	41
2.1.2. <i>Ipoteza de nul</i>	41
2.1.3. <i>Scorul standard z pentru media grupului</i>	43
2.1.4. <i>Decizia cu privire la ipoteza de nul</i>	44
2.1.5. <i>Decizia statistică unilaterală/bilaterală</i>	45
2.2. Erori statistice	46
2.3. Teste statistice bivariate	49
2.4. Aspecte critice ale testării ipotezei de nul	52
2.5. Intervalul de încredere pentru testul statistic la nivelul populației	55
2.6. Metoda <i>bootstrap</i> pentru calcularea intervalului de încredere al testului statistic	58
2.7. Întrebări recapitulative	59
Capitolul 3. Puterea testului și mărimea efectului	61
3.1. Puterea testului	63
3.1.1. <i>Distribuția ipotezei cercetării</i>	63
3.1.2. <i>Mijlocul de creștere a puterii</i>	65
3.2. Mărimea efectului	67
3.2.1. <i>Indicatori ai mărimii efectului</i>	70
3.2.2. <i>Intervalul de încredere pentru mărimea efectului</i>	71
3.2.3. <i>Raportarea mărimii efectului</i>	72
3.3. Analiza de putere	75
3.3.1. <i>Tipuri de analiză de putere</i>	75
3.3.2. <i>Analiza de putere cu G*Power 3</i>	78

3.4. Concluzii	83
3.5. Întrebări recapitulative	84
Capitolul 4. Analiza statistică multivariată - noțiuni generale	85
4.1. Ce este statistica multivariată?	87
4.2. Obiectivele statisticilor multivariate	89
4.3. Structura și natura variabilelor	90
4.4. Numărul variabilelor	92
4.5. Strategii ale analizei multivariate	93
4.6. Interpretarea rezultatelor	94
4.7. Întrebări recapitulative	94
Capitolul 5. Analiza preliminară a datelor statistice	97
5.1. Corectitudinea și validitatea datelor	100
5.2. Valorile excesive	101
5.2.1. Sursele valorilor excesive	102
5.2.2. Impactul valorilor excesive	103
5.2.3. Detectarea valorilor excesive	103
5.2.4. Tratarea valorilor excesive	109
5.3. Valorile lipsă	110
5.3.1. Natura valorilor lipsă	111
5.3.2. Impactul valorilor lipsă	111
5.3.3. Analiza valorilor lipsă	112
5.3.4. Tratarea valorilor lipsă	115
5.4. Normalitatea distribuției	116
5.4.1. Modalități de ilustrare și evaluare a normalității	117
5.4.2. Normalizarea distribuției	119
5.4.3. Observații cu privire la normalizare	121
5.5. Linaritatea	123
5.6. Omogenitatea varianței	126
5.7. Concluzii	128
5.8. Întrebări recapitulative	128
Capitolul 6. Regresia liniară multiplă	131
6.1. Noțiuni de bază despre predicție - regresia liniară simplă	133
6.2. Modelul de predicție multivariată	139
6.3. Condiții și limitări	142
6.4. Alegerea modelului de analiză	144
6.5. Volumul eșantionului	145
6.6. Obiectivele analizei de regresie multiplă	148
6.7. Regresia multiplă cu SPSS	148
6.7.1. Operațiile preliminare	148
6.7.2. Procedura de calcul	150
6.7.3. Interpretarea rezultatelor	155
6.8. Validarea modelului de predicție	161
6.9. Realizarea regresiei liniare multiple cu metoda <i>least squares</i>	162
6.10. Raportarea rezultatelor analizei de regresie	164
6.11. Cercedări ilustrative	164
6.12. Întrebări recapitulative	166
Capitolul 7. Corelația parțială și corelația semiparțială	167
7.1. Conceptul de corelație parțială	169
7.2. Interpretarea corelației parțiale	171

7.3. Corelația parțială multiplă	175
7.4. Calcularea coeficientului de corelație parțială cu SPSS	176
7.5. Corelația semiparțială	178
7.6. Cercetări ilustrative	180
7.7. Întrebări recapitulative	182
Capitolul 8. Regresia logistică binomială	183
8.1. Introducere	185
8.1.1. Similitudini cu regresia liniară	186
8.1.2. Caracteristici specifice regresiei logistice	186
8.1.3. Volumul eșantionului și numărul predictorilor	187
8.2. Modelul regresiei logistice	188
8.2.1. Expresia grafică a regresiei	188
8.2.2. Concepte-cheie în regresia logistică	189
8.2.3. Expresia matematică a regresiei logistice	191
8.3. Efectuarea regresiei logistice cu SPSS	192
8.3.1. Setarea procedurii SPSS	192
8.3.2. Analiza rezultatelor	194
8.4. Raportarea regresiei logistice	201
8.5. Cercetări ilustrative pentru analiza de regresie logistică	201
8.6. Întrebări recapitulative	203
Capitolul 9. Analiza de varianță (ANOVA) factorială	205
9.1. ANOVA și modelul de cercetare inter-subiecți	207
9.2. Familia de proceduri ANOVA	208
9.3. Fundamentarea conceptuală pentru ANOVA factorială	209
9.4. Interpretarea rezultatelor pentru ANOVA factorială	213
9.5. Mărimea efectului și puterea în ANOVA factorială	214
9.6. Condiții	216
9.7. Efectuarea procedurii ANOVA factoriale cu SPSS	217
9.8. Analiza rezultatelor	220
9.9. Raportarea rezultatelor	223
9.10. Cercetări ilustrative	224
9.11. Întrebări recapitulative	226
Capitolul 10. Analiza de covarianță factorială (ANCOVA)	227
10.1. Structura modelului ANCOVA	229
10.2. Numărul covarianțelor	230
10.3. Condiții	231
10.4. Utilitatea modelului ANCOVA	233
10.5. Interpretarea rezultatelor	234
10.6. Efectuarea ANCOVA cu SPSS	235
10.7. Cercetări ilustrative	241
10.8. Întrebări recapitulative	243
Capitolul 11. Analiza multivariată a varianței (MANOVA) și a covarianței (MANCOVA)	245
11.1. Structura modelelor MANOVA și MANCOVA	247
11.2. Avantaje	248
11.3. Dezavantaje	249
11.4. Interpretare	250
11.5. Condiții	252

7.3. Corelația parțială multiplă	175
7.4. Calcularea coeficientului de corelație parțială cu SPSS	176
7.5. Corelația semiparțială	178
7.6. Cercetări ilustrative	180
7.7. Întrebări recapitulative	182
Capitolul 8. Regresia logistică binomială	183
8.1. Introducere	185
8.1.1. Similitudini cu regresia liniară	186
8.1.2. Caracteristicile specifice regresiei logistice	186
8.1.3. Volumul eșantionului și numărul predictorilor	187
8.2. Modelul regresiei logistice	188
8.2.1. Expresia grafică a regresiei	188
8.2.2. Concepte-cheie la regresia logistică	189
8.2.3. Expresia matematică a regresiei logistice	191
8.3. Efectuarea regresiei logistice cu SPSS	192
8.3.1. Setarea procedurii SPSS	192
8.3.2. Analiza rezultatelor	194
8.4. Raportarea regresiei logistice	201
8.5. Cercetări ilustrative pentru analiza de regresie logistică	201
8.6. Întrebări recapitulative	203
Capitolul 9. Analiza de varianță (ANOVA) factorială	205
9.1. ANOVA și modelul de cercetare inter-subiecți	207
9.2. Familia de proceduri ANOVA	208
9.3. Fundamentarea conceptuală pentru ANOVA factorială	209
9.4. Interpretarea rezultatelor pentru ANOVA factorială	213
9.5. Mărimea efectului și puterea în ANOVA factorială	214
9.6. Condiții	216
9.7. Efectuarea procedurii ANOVA factoriale cu SPSS	217
9.8. Analiza rezultatelor	220
9.9. Raportarea rezultatelor	223
9.10. Cercetări ilustrative	224
9.11. Întrebări recapitulative	226
Capitolul 10. Analiza de covarianță factorială (ANCOVA)	227
10.1. Structura modelului ANCOVA	229
10.2. Numărul covarianțelor	230
10.3. Condiții	231
10.4. Utilitatea modelului ANCOVA	233
10.5. Interpretarea rezultatelor	234
10.6. Efectuarea ANCOVA cu SPSS	235
10.7. Cercetări ilustrative	241
10.8. Întrebări recapitulative	243
Capitolul 11. Analiza multivariată a varianței (MANOVA) și a covarianței (MANCOVA)	245
11.1. Structura modelelor MANOVA și MANCOVA	247
11.2. Avantaje	248
11.3. Dezavantaje	249
11.4. Interpretare	250
11.5. Condiții	252

7.2. Interpretarea corelației parțiale

În contextul corelației parțiale, variabila Z joacă un rol explicativ pentru relația dintre X și Y , prin controlul pe care îl exercită asupra acestor două variabile. Modul efectiv în care se manifestă acest control poate diferi de la o situație la alta, iar pentru a-l înțelege și interpreta, trebuie să raportăm corelația parțială la corelația obișnuită. Situațiile care pot apărea din acest punct de vedere sunt următoarele :

- a) *Corelația parțială și corelația totală sunt aproximativ egale.* În această situație, variabila Z nu are nici o contribuție explicativă cu privire la relația dintre variabilele X și Y (figura 7.3). Această constatare nu exclude însă posibilitatea să existe altă variabilă Z care să joace un rol în relația dintre X și Y .

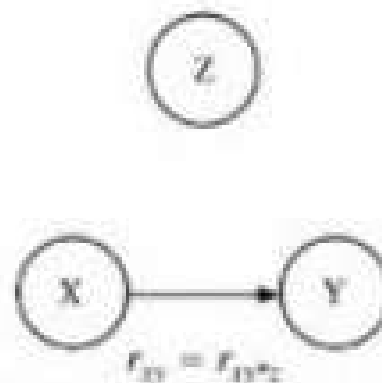


Figura 7.3. Variabila Z nu explică relația dintre X și Y

- b) *Corelația dintre X și Y este mai mică după excluderea variabilei Z .* Această situație este, de regulă, aceea pe care dorim să o punem în evidență și înseamnă că variabila Z oferă o anumită „explicație” în raport cu asocierea dintre variabilele X și Y . Ne aflăm într-o situație de *explicație totală* atunci când variabila Z este simultan o cauză atât pentru X , cât și pentru Y , iar între X și Y nu există o legătură causală directă. Evident, aprecierea antecedenței cauzale a unei variabile în raport cu altă variabilă nu poate decurge din calculele statistice, ci din cunoașterea succesiunii temporale reale dintre acestea (Garson, 2008c). Atunci când corelația parțială se apropie de zero și este nesemnificativă statistic înseamnă că variabila Z se mai murește și variabilă supresoare, iar corelația dintre X și Y este o *corelație falsă* (figura 7.4).

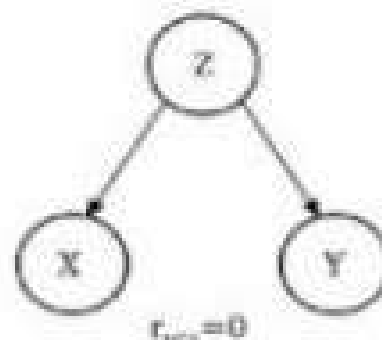
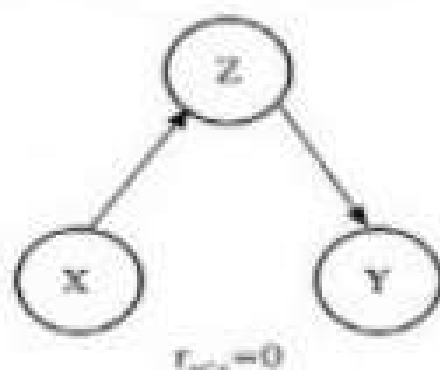
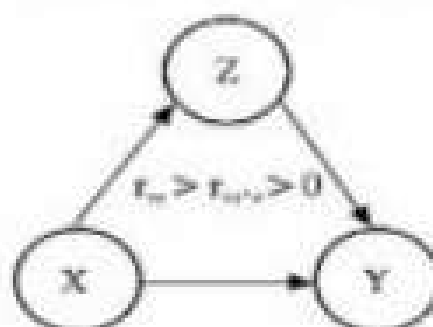
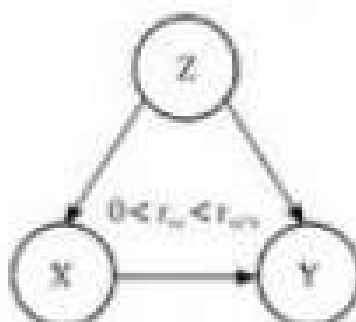


Figura 7.4. Corelație r_{xy} falsă (Z = variabilă supresoare explică total asocierea dintre X și Y)

De exemplu, în cazul corelației dintre consumul de înghețată și numărul actelor violente, temperatura ambiantă este o variabilă supresoare, deoarece ea reprezintă atât cauza variabilei X , cât și cauza variabilei Y . Avem, de asemenea, o explicație totală atunci când variabila X determină variabila Z , iar aceasta, la rândul ei, determină variabila Y (figura 7.5a). Această situație descrie ceea ce se numește o mediere totală între X și Y . Dacă Z este un mediator eficient, el corelează mai puțin cu X . Cu cât această corelație este însă mai mare, cu atât rămâne o cantitate mai mică de variabilitate pe care Z o poate explica în raport cu Y . Acesta este fenomenul de coliniaritate, care, în acest caz, nu poate fi evitat. Singura soluție cu Z să-și păstreze puterea explicativă în raport cu Y este existența unui eșantion suficient de mare – mai exact, $N \cdot (1 - r_{xz}^2)$, – cu cât r_{xz} este mai mare, cu atât mai mare trebuie să fie eșantionul corelației dintre Z și Y (Kenney, 2009). Atunci când corelația parțială este mai mică decât corelația totală, dar rămâne semnificativă statistic, Z oferă doar o *explicație parțială* și vorbim de o mediere incompletă între X și Y (figura 7.5b). În exemplul de mai sus, corelația dintre vârstă și nivelul salariului este mai mică după ce am exclus efectul nivelului de educație, ceea ce indică o mediere incompletă a relației dintre vârstă și salariu de către nivelul de instruire.

a) Mediere totală între X și Y b) Mediere parțială între X și Y **Figura 7.5.** Variabila Z explică total asocierea dintre X și Y

- c) Corelația dintre X și Y este mai mare după excluderea variabilei Z . În acest caz, conchidem că relația dintre X și Y este mascată de prezența variabilei Z (figura 7.6). De exemplu, să presupunem că am măsurat variabilele ore de studiu pentru un examen (X), nota la examen (Y) și nivelul anxietății de examen (Z), iar corelațiile dintre ele sunt următoarele: $r_{xy} = +0.20$ ($r_{xy}^2 = 0.04$); $r_{xz} = +0.80$ ($r_{xz}^2 = 0.64$); $r_{yz} = -0.40$ ($r_{yz}^2 = 0.16$). În acest caz, corelația parțială dintre orele de studiu și nota la examen, sub controlul nivelului anxietății de examen, este $r_{xy.z} = +0.96$, substanțial mai mare decât corelația totală între X și Y . Cu alte cuvinte, anxietatea maschează corelația dintre numărul orelor de studiu și nota obținută.

**Figura 7.6.** Variabila Z maschează efectul lui X asupra lui Y

Se poate întâmpla uneori ca semnul corelației parțiale dintre X și Y să fie diferit de semnul corelației obișnuite dintre aceleași variabile (figura 7.7). Bobko și Donnelly (*apud* Bobko, 2001) au testat ipoteza că valoarea percepută a postului de muncă (Y) crește pe măsură ce aptitudinea minimă necesară este mai ridicată (X). Rezultatele cercetării i-au pus însă în fața unei corelații negative surprinzătoare între aceste două variabile. Analiza ulterioară a condus la constatarea că subiecții care se aflau pe posturi ce implicau participarea la misiuni de luptă aveau o tendință de valorizare foarte scăzută a postului. Utilizând un indice de probabilitate a participării la activități de luptă drept variabilă de control, corelația parțială dintre variabilele X și Y a devenit pozitivă și semnificativă statistic.

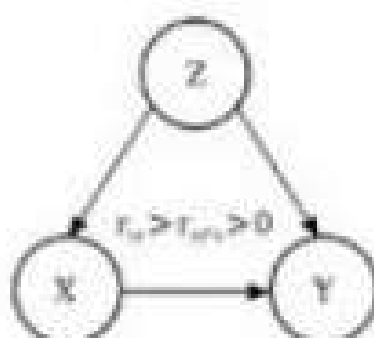


Figura 7.7. Relație de supresione a variabilei Z asupra corelației dintre X și Y , cu r_{xy} negativ

În modelele ilustrate mai sus am evocat raportul de mărime dintre coeficienții de corelație. Va trebui să ținem însă cont de faptul că nu diferența de mărime în sine trebuie luată în considerare, ci diferențele semnificative statistic. Rezolvarea acestei probleme constă în testarea semnificației diferenței dintre coeficienții de corelație, în contextul modelului corelațiilor multiple. Dat fiind faptul că în acest caz coeficienții de corelație sunt calculați pe același eșantion de subiecți, formula de calcul pentru testul de semnificație este diferită de aceea pentru diferența dintre doi coeficienți de corelație independenți și se prezintă astfel (Chen și Popovich, *apud* Field, 2009) :

$$t_{xy} = (r_{xy} - r_{xy.z}) \sqrt{\frac{(n-3) * (1 + r_{xy.z})}{2 * (1 - r_{xy}^2 - r_{xy.z}^2 - r_{xz}^2 + 2 * r_{xy} * r_{xy.z} * r_{xz})}} \quad (\text{formula 7.2})$$

unde :

r_{xy} = coeficientul de corelație dintre X și Y

$r_{xy.z}$ = coeficientul de corelație dintre Z și Y

r_{xz} = coeficientul de corelație dintre X și Z

n = volumul eșantionului

Rezultatul reprezintă o valoare t , a cărei semnificație statistică se citește din tabelul distribuției t , cu $N-3$ grade de libertate, unde N este volumul eșantionului.

Practic, cu această formulă putem testa diferența oricărei perechi de coeficienți de corelație, pe care li plasăm în prima paranteză a formulei. Deși complicată, formula nu conține totuși decât coeficienți de corelație (pe care li obținem cu SPSS) și volumul eșantionului.

Oricât de tentantă ar fi interpretarea causală, în contextul modelului corelației parțiale, Anderson (2001) atrage totuși atenția asupra riscului de utilizare abuzivă a acesteia. Simplul fapt că am eliminat statistic efectul unei variabile (Z) nu înseamnă că am adus la lumină o relație de tip cauză-efect între X și Y . Nimic nu ne garantează că nu există una sau mai multe alte variabile Z , necunoscute, care influențează relația dintre X și Y . Ronald Fischer însuși a susținut această idee în lucrarea sa fundamentală *Statistical methods for research workers*: „Dacă știu că fenomenul A nu influențează alte fenomene B , C , D ..., ci, dimpotrivă, este probabil direct influențat de acestea, atunci calcularea corelațiilor parțiale ale lui A cu B , C , D ..., care elimină de fiecare dată valorile rămase, va susține cea mai valoroasă analiză a cauzalității. Dacă însă alegem un fenomen social referitor la care nu avem cunoștințe anterioare privind ceea ce îl cauzează sau nu îl cauzează, atunci calcularea coeficientului de corelație, total sau parțial, nu ne ajută să avansăm în evaluarea cauzei care îl determină” (Fisher, 1925). Pentru a putea trage o concluzie de ordin causal pe baza corelației, trebuie să fie întrunite următoarelor trei condiții fundamentale (Kenny, 1979):

- existența unei corelații între variabilele X și Y ;
- valorile variabilei X reprezintă o caracteristică cu manifestare anterioară celei care produce valorile variabilei Y - aspect care face imposibil efectul lui Y asupra lui X ;
- asocierea dintre X și Y persistă și după eliminarea efectului uneia sau mai multor variabile covariante.

O metodă de inferență causală pe baza coeficienților de corelație este „compararea încrucișată longitudinală a corelațiilor” (*crossed-lagged panel design*). Aceasta nu reprezintă o procedură statistică în sine, și nici nu are legătură cu corelația parțială. Dacă o amintim aici, o facem doar pentru a completa discuția cu privire la cauzalitate și corelație. Compararea încrucișată a corelațiilor este un model de cercetare multivariată, în care sunt utilizați patru coeficienți de corelație, calculați în două momente temporale diferite. Să presupunem că dorim să aflăm dacă atitudinea cooperantă a șefilor este cauza unei conștiinciozități mai mari a angajaților sau invers. Pentru a rezolva această problemă se poate realiza un studiu longitudinal, în care aceste variabile sunt măsurate pe același grup de subiecți în două momente diferite, de exemplu, în prima lună de la angajare și după un an. Variabilele măsurate în cele două momente sunt apoi corelate două câte două (figura 7.8).

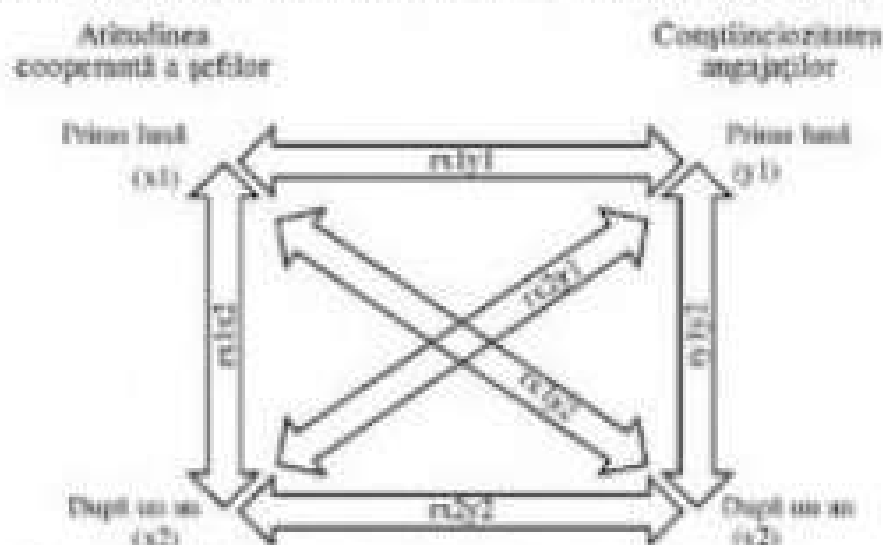


Figura 7.8. Compararea încrucișată longitudinală a corelațiilor

Pentru a interpreta rezultatele, se compară mai întâi corelațiile dintre cele două momente (după prima lună și după un an). Dacă între atitudinea cooperantă a șefilor și conștiinciozitatea angajaților există o corelație, atunci r_{12t2} ar trebui să fie mai mare decât r_{12t1} , acest lucru însemnând, pe de o parte, stabilitate și, pe de altă parte, potențarea asocierii în timp. Mai departe, se compară corelațiile r_{11t2} și r_{22t2} , care măsoară gradul de asociere între fiecare variabilă și ea însăși, în cele două momente diferite. Variabila pe care o bănuim a avea un efect asupra celeilalte ar trebui să aibă o corelație mai mare între cele două momente. De exemplu, dacă atitudinea cooperantă a șefilor este cauza conștiinciozității subordonaților, atunci corelația acestei variabile între prima lună și după un an (r_{11t2}) ar trebui să fie mai mare decât corelația conștiinciozității angajaților între aceleași momente (r_{22t2}). Altfel spus, dacă atitudinea cooperantă este cauza, atunci șefii care au fost cooperanți în prima lună trebuie să fie la fel de cooperanți și după un an. În fine, comparațiile cele mai importante sunt cele încrucișate. Variabila „cauză” trebuie să fie prima în ordine temporală și să aibă o corelație mai mare decât variabila „efect”. Ca urmare, dacă atitudinea cooperantă a șefilor este aceea care determină conștiinciozitatea subordonaților, atunci r_{12t2} ar trebui să fie mai mare decât r_{21t2} . Cu alte cuvinte, dacă atitudinea cooperantă a șefilor este cauza, atunci corelația acestei variabile din prima lună (r_{12t1}) cu conștiinciozitatea subordonaților după un an (r_{21t2}) trebuie să fie mai mare decât corelația dintre atitudinea șefilor după un an și conștiinciozitatea subordonaților din prima lună (r_{21t1}).

Metoda comparației încrucișate longitudinale a corelațiilor poate fi utilă pentru aprecierea direcției cauzale. Intensitatea relației cauzale este dată de mărimea coeficientului de corelație, evaluată ca mărime a efectului (coeficientul de determinare r^2). În ciuda posibilității de a detecta sensul cauzalității, această metodă nu este lipsită de vulnerabilități: în primul rând, lipsește raportarea la un nivel de bază al variabilei „cauză”; în al doilea rând, nu exclude variabilele covariante care să se interpună între „cauză” și „efect”. De aceea, aplicarea ei ar fi recomandabilă numai după ce au fost eliminate eventuale variabile covariante prin metoda corelației parțiale. În plus, aplicarea acestei metode implică o durată mai mare a cercetării, pentru a acoperi cele două momente. În pofida limitelor, Locascio (1982) apreciază că analiza încrucișată a corelațiilor nu trebuie respinsă și își merită locul printre tehnicile exploratorii ale relațiilor dintre variabile.

7.3. Corelația parțială multiplă

Coeficientul de corelație parțială oferă posibilitatea de a calcula nivelul asocierii dintre două variabile menținând constantă (controlând) influența externă din partea altei variabile. Chiar dacă în exemplele prezentate am sugerat doar influența unei a treia variabile, modelul de analiză și procedura de calcul pot fi extinse la oricât de multe variabile externe (corelație parțială multiplă) (figura 7.9).

Controlul simultan al mai multor variabile poate fi tentant, dar trebuie utilizat cu grijă, deoarece ridică probleme în elaborarea unui model de interpretare a rezultatelor. O soluție recomandabilă ar fi în acest caz efectuarea unor analize repetate, cu eliminarea succesivă a efectului câte unei singure variabile. În general, se consideră că utilizarea

corelației parțiale trebuie să se limiteze la un număr de 3-4, maxim 5 variabile. Pentru un număr mai mare de variabile, fie se va apela la divizarea modelului de analiză în submodele mai simple, fie se vor prefera alte proceduri statistice mai evoluate, cum ar fi analiza de cale, modelarea ecuației de structură sau analiza factorială exploratorie, capabile să surprindă modele complexe într-un număr mare de variabile (Garson, 2008c).

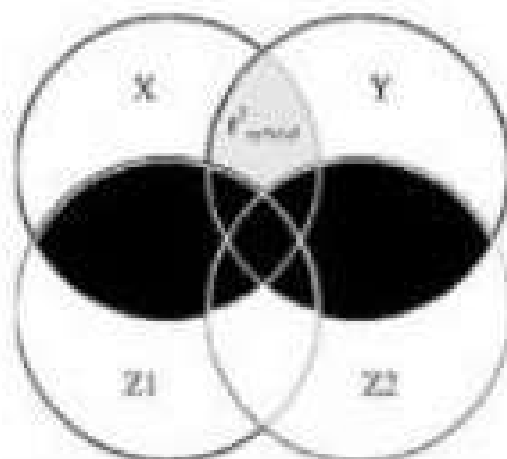


Figura 7.9. Corelația parțială multiplă dintre X și Y după excluderea efectului variabilelor $Z1$ și $Z2$

În funcție de numărul variabilelor controlate, modelul corelației parțiale poate fi de *ordinul întâi* (atunci când se exclude o singură variabilă) sau de *ordinul doi* (atunci când se exclude două variabile). Drept urmare, corelația obișnuită, care nu exclude nici o variabilă, este de *ordin zero*.

7.4. Calcularea coeficientului de corelație parțială cu SPSS

Pentru exemplificarea modului de calcul cu SPSS, vom utiliza trei variabile ale chestionarului de personalitate ALAPS (Popa, 2002): agresivitate, impulsivitate și labilitate afectivă. Ipoteza pe care dorim să o testăm este aceea că relația dintre agresivitate și impulsivitate este mediată de echilibrul emoțional. Ca urmare, ne așteptăm la o corelație diferită între primele două variabile, dacă o menținem constantă pe a treia.

Extrus din matricea de date SPSS

	id	mpu	agre	afec
1	3503	8	11	7
2	3504	4	4	5
3	3505	8	7	10
4	3506	8	8	9
5	3507	10	9	6
6	3508	9	11	14

Procedura de calcul se lansează din meniul *Analyze/Correlate-Partial*, cu deschiderea casetei *Partial Correlation*, în care variabilele cercetate sunt trecute în lista *Variables*, iar variabila pe care dorim să o izolăm, în lista *Controlling for*: