

Partielle Korrelation

Einen Zusammenhang nach linearer Bereinigung um eine Drittvariable verstehen

Ratiomera Statistics

Inhaltsverzeichnis

Zusammenfassung der Kursseite	2
Zwei gleichwertige Wege	2
Vergleiche die rohe und die bereinigte Frage	2
Prüfungen und Grenzen	2
Erweiterte Referenz	3
Zweck und Grundlagen	3
Zentrale Ideen	3
Formelleitfaden	4
Die erklärende Abbildung lesen	5
Checkliste zur Interpretation	6
Verbindung zu anderen Themen	6

Zusammenfassung der Kursseite

Die partielle Korrelation fragt, ob zwei quantitative Variablen weiterhin linear zusammenhängen, nachdem beide für dieselbe dritte Variable bereinigt wurden. Sie ist ein bereinigter Zusammenhang, kein experimenteller Effekt und keine Vorhersagegleichung.

Zwei gleichwertige Wege

Seien X und Y die interessierenden Variablen und Z die Bereinigungsvariable.

1. Regressiere X auf Z und behalte die Residuen e_X .
2. Regressiere Y auf Z und behalte die Residuen e_Y .
3. Berechne die Korrelation zwischen e_X und e_Y .

Jedes Residuum ist der Teil seiner interessierenden Variable, den Z im angepassten Modell nicht linear vorhersagt. Ihre Korrelation ist $r_{XY \cdot Z}$.

Mit einer Bereinigungsvariable folgt dasselbe Ergebnis aus den drei paarweisen Korrelationen:

$$r_{XY \cdot Z} = \frac{r_{XY} - r_{XZ}r_{YZ}}{\sqrt{(1 - r_{XZ}^2)(1 - r_{YZ}^2)}}.$$

Vergleiche die rohe und die bereinigte Frage

Tabelle 1: Rohe und partielle Korrelation beantworten verwandte, aber verschiedene Fragen.

Koeffizient	Frage
r_{XY}	Wie sind X und Y vor der Bereinigung linear verbunden?
$r_{XY \cdot Z}$	Wie sind die Residualteile von X und Y verbunden, nachdem ihre separaten Beziehungen mit Z dargestellt wurden?

Ein bereinigter Koeffizient kann kleiner werden, ähnlich bleiben, das Vorzeichen wechseln oder grösser werden. Ein kleinerer Wert deutet darauf hin, dass Z einen Teil des rohen Musters darstellt. Ein grösserer Wert kann Suppression widerspiegeln, bei der die Bereinigung einen zuvor teilweise verdeckten Zusammenhang sichtbar macht. Keine dieser Änderungen beweist eine kausale Rolle von Z .

Prüfungen und Grenzen

- Verwende für jede Korrelation in der Berechnung dieselbe vollständige Fallmenge.
- Betrachte X mit Z , Y mit Z und das Residuen-Residuen-Diagramm. Lineare Bereinigung entfernt keine nicht modellierte gekrümmte Beziehung.
- Nenne genau, für welche Variable bereinigt wurde und weshalb sie zur Frage gehört.
- Bereinigung stellt keine zeitliche Reihenfolge her, entfernt keine unbeobachteten Variablen, korrigiert keine Messfehler und ersetzt keine Randomisierung.
- Die partielle Korrelation ist symmetrisch. Ein Tausch von X und Y verändert ihren Wert nicht.

Thema 7 behält eine benannte Ergebnisvariable und setzt mehrere Prädiktoren in eine Regressionsgleichung. Seine bedingten Koeffizienten tragen dieselbe Bereinigungs-idee, drücken Änderungen aber in Ergebniseinheiten aus und können kategoriale Prädiktoren und Interaktionen enthalten.

Erweiterte Referenz

Zweck und Grundlagen

Die partielle Korrelation beschreibt den linearen Zusammenhang zwischen zwei Variablen, nachdem aus beiden der angepasste lineare Zusammenhang mit einer gemessenen Drittvariable entfernt wurde. Mit den interessierenden Variablen X und Y sowie der Kontrollvariable Z wird die partielle Korrelation erster Ordnung als $r_{XY \cdot Z}$ geschrieben. Der Punkt wird als «unter Kontrolle von» gelesen. Der Koeffizient fragt, ob Fälle, die bei X unter Berücksichtigung von Z höher als erwartet liegen, auch bei Y unter Berücksichtigung desselben Z tendenziell höher als erwartet liegen.

Die Formulierung Z **konstant halten** beschreibt einen modellbasierten Vergleich. Sie bedeutet nicht, dass alle beobachteten Fälle tatsächlich denselben Z -Wert besitzen. Mithilfe der linearen Regression wird X aus Z und Y aus Z vorhergesagt. Das Residuum jedes Falles hält fest, wie weit sein beobachteter Wert über oder unter dieser Vorhersage liegt. Die partielle Korrelation ist die gewöhnliche Pearson-Korrelation zwischen diesen beiden Gruppen von Residuen.

Dieses Verfahren beantwortet eine Frage nach einem bedingten Zusammenhang. Es kann zeigen, dass eine rohe Korrelation nach der Bereinigung kleiner oder grösser wird oder ihr Vorzeichen wechselt. Es kann jedoch nicht von selbst entscheiden, ob Z eine Konfundierungsvariable, ein Mediator, ein Collider oder eine geeignete Kontrollvariable ist. Diese Rollen betreffen den Vorgang, der die Daten erzeugt, und erfordern inhaltliches Wissen, zeitliche Reihenfolge und das Forschungsdesign.

Grösse	Was korreliert wird	Beantwortete Frage
Rohe Korrelation r_{XY}	Beobachtetes X mit beobachtetem Y	Wie hängen die beiden gemessenen Variablen linear zusammen?
Partielle Korrelation $r_{XY \cdot Z}$	Residualisiertes X mit residualisiertem Y	Wie hängen ihre verbleibenden linearen Bestandteile nach der Bereinigung um Z zusammen?

Zentrale Ideen

Die Residualisierung erfolgt in drei nachvollziehbaren Schritten. Regrediere zuerst X auf Z und speichere jedes Residuum e_{Xi} . Regrediere danach Y auf Z und speichere jedes Residuum e_{Yi} . Berechne schliesslich die Pearson-Korrelation zwischen e_X und e_Y . Werte über null bei einer residualisierten Variable bedeuten «höher als die lineare Vorhersage auf Grundlage von Z ». Werte unter null bedeuten «niedriger als vorhergesagt».

Eine Standardisierung der ursprünglichen Variablen führt diese Bereinigung nicht durch. Bei der Standardisierung wird ein Mittelwert abgezogen und durch eine Standardabweichung geteilt. Dadurch werden Einheiten und Bezugspunkte angeglichen, während die Pearson-Korrelation unverändert bleibt. Die Residualisierung entfernt dagegen den angepassten linearen Bestandteil, der mit der Kontrollvariable verbunden ist. Diese Unterscheidung ist wichtig, weil zwei Grafiken gleichartige standardisierte Achsen besitzen können und dennoch nach der Residualisierung verschiedene Korrelationen zeigen.

Beobachtete Veränderung des Koeffizienten	Mögliche Lesart	Was weiterhin geprüft werden muss
Partieller Koeffizient ist kleiner	Ein Teil der rohen Überlappung wurde mit Z geteilt	Ob Z eine vertretbare Kontrollvariable ist und die Modelle angemessen sind

Beobachtete Veränderung des Koeffizienten	Mögliche Lesart	Was weiterhin geprüft werden muss
Partieller Koeffizient ist ähnlich	Die lineare Bereinigung um Z veränderte wenig	Nicht lineare Effekte, Messung, Wertebereich und Stichprobenunsicherheit
Partieller Koeffizient ist grösser	Die Bereinigung machte einen zuvor verdeckten Zusammenhang sichtbar	Ob Suppression inhaltlich plausibel und nicht nur ein zufälliges Muster ist

Die Bereinigung ist linear. Wenn Z mit X oder Y gekrümmt zusammenhängt, kann eine geradlinige Residualisierung systematische Struktur zurücklassen. Das Verfahren übernimmt ausserdem die Empfindlichkeit der Pearson-Korrelation und Regression gegenüber einflussreichen Beobachtungen und eingeschränkten Wertebereichen. Untersuche die rohen Beziehungen von X mit Y , X mit Z und Y mit Z und danach die residualisierte Beziehung.

Die partielle Korrelation ist konzeptuell eng mit der multiplen Regression verbunden. Beide stellen bedingte lineare Fragen, nachdem weitere gemessene Informationen berücksichtigt wurden. Ihre numerischen Koeffizienten besitzen jedoch verschiedene Skalen: Die partielle Korrelation ist auf das Intervall von -1 bis 1 standardisiert, während eine spätere Regressionssteigung eine angepasste Differenz der Zielvariable pro Prädiktoreinheit ausdrückt. Thema 7 entwickelt diesen umfassenderen Rahmen mit mehreren Prädiktoren.

Formelleitfaden

Die Residualisierung von X bezüglich Z beginnt mit einer angepassten Geraden und behält jenen Teil, den die Gerade nicht erklärt:

$$e_{Xi} = x_i - (a_X + b_X z_i)$$

Führe die entsprechende Berechnung für Y durch:

$$e_{Yi} = y_i - (a_Y + b_Y z_i)$$

Die partielle Korrelation ist danach die Pearson-Korrelation der beiden Residualvariablen:

$$r_{XY \cdot Z} = r(e_X, e_Y)$$

Wenn genau eine Variable Z kontrolliert wird, kann der Koeffizient auch aus den drei paarweisen Korrelationen berechnet werden:

$$r_{XY \cdot Z} = \frac{r_{XY} - r_{XZ}r_{YZ}}{\sqrt{(1 - r_{XZ}^2)(1 - r_{YZ}^2)}}$$

Diese Formel ist kompakt. Die Residualmethode ist jedoch oft leichter zu verstehen und zu diagnostizieren, weil sie die beiden Bereinigungsmodelle direkt sichtbar macht. Beide Wege stimmen überein, wenn sie dieselben vollständigen Fälle, gewöhnliche lineare Regressionen mit Achsenabschnitten und dieselben drei gemessenen Variablen verwenden. Diese Übereinstimmung prüft die Berechnung. Sie belegt nicht, dass die Kontrollvariable kausal angemessen war.

Der Nenner zeigt zugleich, wann die direkte Formel definiert ist. Falls $|r_{XZ}| = 1$ oder $|r_{YZ}| = 1$ gilt, besitzt eine der residualisierten interessierenden Variablen keine verbleibende Variation, und der Nenner ist null. Für eine Variable ohne Streuung kann keine Korrelation berechnet werden. Nahezu perfekte Beziehungen mit Z können das bereinigte Ergebnis ebenfalls sehr empfindlich gegenüber kleinen Veränderungen oder Rundungen machen.

Art der Kontrolle	Was sich verändert	Welche Schlussfolgerung sie stützen kann
Experimentelle Kontrolle	Das Studiendesign weist Bedingungen zu oder hält sie fest, bevor Ergebnisse beobachtet werden	Kann einen kausalen Vergleich stärken, wenn Design und Annahmen ihn rechtfertigen
Statistische Kontrolle	Die Analyse stellt angepasste lineare Beziehungen mit dem gemessenen Z dar	Liefert einen bereinigten Zusammenhang und keine zufällige Zuweisung
Nicht gemessene Drittvariable	Nichts in der Berechnung stellt sie dar	Ihr möglicher Beitrag bleibt ungeklärt

Die Residualisierung entfernt nur den angepassten linearen Bestandteil, der mit der gemessenen Version von Z verbunden ist. Sie erzeugt keine fehlerfreien Variablen, entfernt nicht automatisch nicht lineare Beziehungen und garantiert nicht, dass Z eine geeignete Kontrollvariable war.

Die erklärende Abbildung lesen

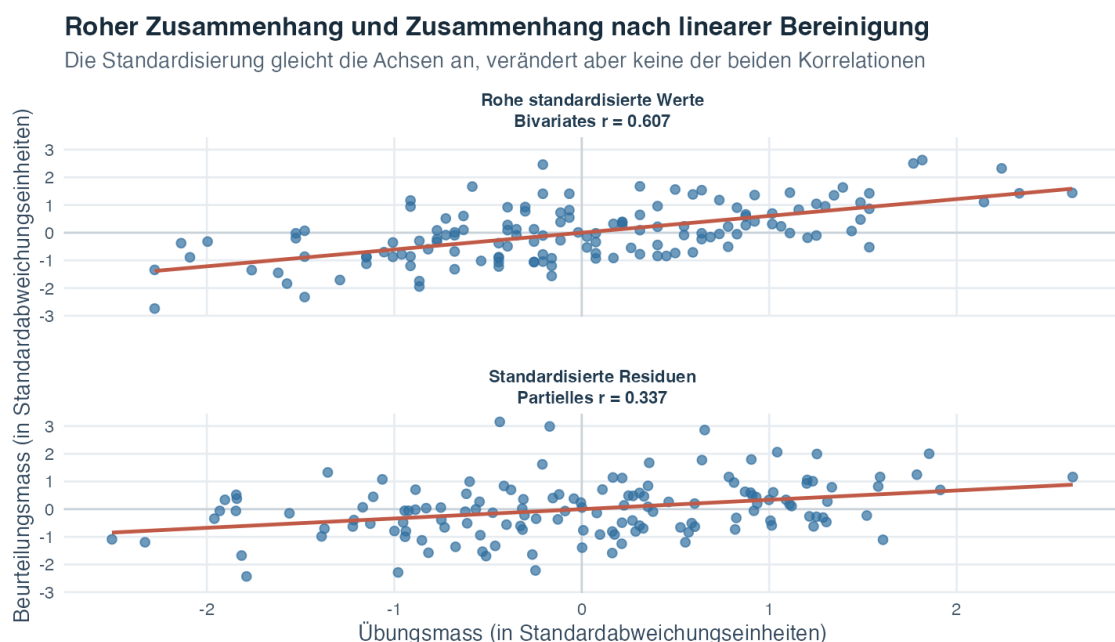


Abbildung 1: Zwei Streudiagramme vergleichen den standardisierten rohen Zusammenhang mit dem Zusammenhang zwischen standardisierten Residuen nach linearer Bereinigung um ein drittes Merkmal.

Das linke Feld zeigt standardisierte Übungs- und Beurteilungswerte. Standardabweichungseinheiten legen null auf den Mittelwert jeder Variable und machen eine Einheit gleich einer Stichprobenstandardabweichung. Die sichtbare ansteigende Gerade und die angegebene bivariate Korrelation von 0.607 beschreiben in diesen simulierten Werten einen mässig positiven rohen linearen Zusammenhang.

Das rechte Feld zeigt standardisierte Residuen, nachdem beide Variablen aus dem Kontrollmerkmal vorhergesagt wurden. Jede horizontale Koordinate bedeutet nun, wie weit der Übungswert eines Falles über oder unter dem aus der Kontrollvariable vorhergesagten Wert liegt. Jede vertikale Koordinate bezeichnet die entsprechende Abweichung bei der Beurteilung. Die Gerade steigt weiterhin an, in standardisierten Einheiten jedoch weniger stark. Die partielle Korrelation beträgt

0.337. Die Bereinigung hat den gemessenen Zusammenhang somit abgeschwächt und zugleich einen positiven Residualzusammenhang bestehen lassen.

Der Untertitel der Abbildung betont, dass die Standardisierung allein keine der beiden Korrelationen verändert. Der Koeffizient ändert sich, weil das zweite Feld Residuen verwendet, und nicht weil seine Achsen standardisiert sind. Der Unterschied zwischen 0.607 und 0.337 ist ein deskriptiver Hinweis darauf, dass ein Teil der rohen gemeinsamen Variation auf die Kontrollvariable ausgerichtet war. Er ist keine Diagnose von Kausalität. Er sollte zusammen mit den drei rohen paarweisen Grafiken und der inhaltlichen Rolle der Kontrollvariable interpretiert werden.

Checkliste zur Interpretation

Benenne X , Y und jede Kontrollvariable und erkläre, weshalb jede Kontrolle in die Analyse gehört. Beschreibe vor der Bereinigung die angenommene zeitliche oder kausale Reihenfolge in Worten. Untersuche alle rohen paarweisen Grafiken, Muster fehlender Daten, Wertebereiche und einflussreiche Beobachtungen. Prüfe, ob die Bereinigungsbeziehungen annähernd linear sind. Berichte die rohe und die partielle Korrelation gemeinsam, damit die Lesenden erkennen können, was sich verändert hat.

Beschreibe den Koeffizienten als Zusammenhang nach linearer Bereinigung um die namentlich genannte Drittvariable. Berichte die Stichprobengröße, die drei bivariaten Korrelationen und die daraus entstehende partielle Korrelation, damit die Berechnung nachvollziehbar ist. Interpretiere eine Verkleinerung des Koeffizienten nicht als automatischen Beweis für Konfundierung, eine Vergrößerung nicht als automatischen Beweis für Suppression und einen Wert nahe null nicht als Beweis für das Fehlen jeder Beziehung. Berücksichtige durchgehend Messqualität und Stichprobenunsicherheit.

Verbindung zu anderen Themen

Kovarianz und Pearson-Korrelation führten die gepaarte lineare gemeinsame Variation ein. Die einfache Regression zerlegte danach jeden Wert der Zielvariable in einen angepassten Wert und ein Residuum. Die partielle Korrelation verbindet diese Ideen: Sie verwendet zwei Regressionen, um die linear mit Z verbundenen Bestandteile zu entfernen, und korreliert danach die verbleibenden Bestandteile.

Die multiple Regression ist der natürliche nächste Schritt. Sie schätzt in einem einzigen Modell der Zielvariable den bedingten Beitrag mehrerer Prädiktoren und gibt Steigungen in deren ursprünglichen Einheiten an. Die Sichtweise der partiellen Korrelation bleibt hilfreich, weil sie die Bedeutung der Formulierung «andere Prädiktoren konstant halten» erklärt: Verglichen werden Fälle anhand jener Teile eines Prädiktors und der Zielvariable, die nach der linearen Bereinigung verbleiben. Diese Verbindung macht aus einer technischen Formulierung einen konkreten Vergleich von Residuen.

Dokument-ID: topic-06-partial-correlation-summary-de

Notizen 1 von 3

Partielle Korrelation | Thema 6

Notizen 2 von 3

Partielle Korrelation | Thema 6

Notizen 3 von 3

Partielle Korrelation | Thema 6