

Матрица-мать всех регрессий

Борис Демешев*

31 декабря 2016 г.

Аннотация

Корреляция — это способ описать силу линейной зависимости между двумя случайными величинами одним числом. Каков геометрический смысл корреляции? Что такое частная корреляция? Как связаны частная и условная корреляция? Как связаны частные корреляции, коэффициенты регрессии и стандартизованные коэффициенты регрессии? И, главное, оказывается, существует матрица-мать, порождающая все эти меры зависимости.

Ключевые слова: корреляция, частная корреляция, условная корреляция, косинус, проекция, стандартизованный коэффициент, теорема Фриша-Во-Ловелла, матрица-мать всех регрессий.

1 Сколько вешать в граммах?

Everything Should Be Made as Simple
as Possible, But Not Simpler.

*Roger Sessions,
paraphrase of Albert Einstein*

There is always a well-known solution
to every human problem — neat,
plausible, and wrong.

Henry Louis Mencken

Почему при болезни мы измеряем температуру тела? По всей видимости важны два факта:

1. Измерить температуру очень удобно
2. Это измерение несёт в себе информацию о здоровье

*НИУ ВШЭ, Москва.

Сложное описание состояния здоровья сводится измерением температуры к одной цифре. Естественно, куча информации теряется в этой цифре и бессмысленно лечить человека, руководствуясь только температурой его тела. Температура 40° говорит, что что-то не так, но что — непонятно. А температура 36.6° ещё не говорит о том, что у человека идеальное здоровье. Однако процедура очень проста и в некоторых ситуациях, например, при обыкновенной простуде, её достаточно для принятия решения о приёме жаропонижающего.

Если бы для измерения температуры нужно было специальное устройство размером с половину комнаты, никто бы дома её не мерял. Люди любят упрощать, а здесь простота измерения очень важна!

Подобном образом дела обстоят и с описанием зависимости между случайными величинами. Зависимость между случайными величинами полностью описывается их совместной функцией распределения, $F(x, y) = \mathbb{P}(X \leq x, Y \leq y)$. Это слишком сложно! Вместо сложной функции распределения мы хотим получить одно число. Некую «силу зависимости». Назовём это мифическое число $\text{Dep}(X, Y)$.

Чего бы нам хотелось от этого числа?

1. Посчитать это число очень удобно
2. Это число несёт в себе информацию о зависимости

2 Ковариация

В каком смысле «удобно» считать?

Очень часто возникают суммы случайных величин, поэтому было бы здорово, чтобы у суммы легко считалась наша характеристика $\text{Dep}(X + Z, Y)$. Проще всего было бы, если бы:

$$\text{Dep}(X + Z, Y) = \text{Dep}(X, Y) + \text{Dep}(Z, Y)$$

Кроме того, хотелось бы, чтобы сила зависимости X и Y была бы такой же, как и сила зависимости Y и X :

$$\text{Dep}(X, Y) = \text{Dep}(Y, X)$$

И, конечно, мы ждём, что у независимых случайных величин нулевая сила зависимости, $\text{Dep}(X, Y) = 0$, а ненулевая сила зависимости, $\text{Dep}(X, Y) \neq 0$, была бы возможна только у зависимых случайных величин.

Этим трём свойствам, аддитивности, симметричности и занулению при независимости, удовлетворяет ковариация.

Определение 1. Ковариация величин X и Y равна

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

здесь рассказать про прямоугольники и площадь с плюсом/минусом? почему она такая?

Можно доказать, что любая другая непрерывная функция удовлетворяющая трём заявленным свойствам обязана быть пропорциональна ковариации (?).

3 Коэффициенты в парной регрессии

Наиболее легко интерпретировать коэффициенты в парной регрессии. Пусть у нас есть две случайных величины X_1 и X_2 . Мы можем получить два коэффициента:

- коэффициент β_{12} — на сколько в среднем растёт X_1 при росте X_2 на единицу
- коэффициент β_{21} — на сколько в среднем растёт X_2 при росте X_1 на единицу

А теперь из этого словесного определения мы получим формулу для β_{12} . Разложим величину X_1 на два слагаемых. Первое слагаемое вбирает в себя всю ту часть X_1 , которая линейно зависит от X_2 , а второе — всё оставшееся:

$$X_1 = \beta_{12} \cdot X_2 + u$$

Здесь β_{12} показывает на сколько единиц в среднем растёт X_1 при росте X_2 на одну единицу.

Естественно, мы хотим, чтобы с ростом X_2 величина u в среднем не менялась, то есть хотим нулевую «силу зависимости» между ними, $\text{Cov}(X_2, u) = 0$.

$$\text{Cov}(X_2, X_1 - \beta_{12}X_2) = 0$$

По свойствам ковариации получаем

$$\text{Cov}(X_2, X_1) - \beta_{12} \text{Cov}(X_2, X_2) = 0$$

И, тадам, выражаем коэффициент β_{12} :

$$\beta_{12} = \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_2)}$$

По аналогии окажется, что

$$\beta_{21} = \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_1)}$$

Заметим, что хотя знак β_{12} и знак β_{21} совпадают, сами они будут совпадать только если $\text{Var}(X_1) = \text{Var}(X_2)$.

Коэффициент β_{12} , естественно, будет реагировать на изменение единиц измерения X_1 или X_2 . Если увеличить X_1 в 10 раз, то и β_{12} увеличится в 10 раз. Иногда чувствительность к единицам измерения является нежелательной.

Коэффициент регрессии можно избавить от чувствительности к единицам измерения, если стандартизировать исходные переменные, то есть поделить их на стандартные отклонения. При этом мы получим стандартизированный коэффициент регрессии:

$$\beta_{12}^{st} = \frac{\text{Cov}(X_1/\sigma_1, X_2/\sigma_2)}{\text{Var}(X_2/\sigma_2)} = \beta_{12} \frac{\sigma_2}{\sigma_1} = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1) \text{Var}(X_2)}}$$

По своему смыслу β_{12}^{st} показывает, на сколько своих стандартных отклонений в среднем изменяется X_1 при изменении X_2 на одно своё стандартное отклонение.

Заметим, что произошло чудо! В силу симметричности выражения для β_{12}^{st} оказывается, что $\beta_{12}^{st} = \beta_{21}^{st} = \beta^{st}$. Получается, с одной стороны, что:

$$\frac{X_1}{\sigma_1} = \beta^{st} \frac{X_2}{\sigma_2} + u_1$$

А с другой стороны, в противоположном разложении будет тот же коэффициент:

$$\frac{X_2}{\sigma_2} = \beta^{st} \frac{X_1}{\sigma_1} + u_2$$

Это чудо слегка контр-интуитивно! Если бы зависимость между X_1 и X_2 была бы жёстко детерминистически линейной, и с ростом X_2 на единицу величина X_1 росла бы на Δ , то с ростом X_1 на единицу величина X_2 росла бы на $1/\Delta$. Для случайных величин обращения не происходит! Если с ростом X_1 на одно своё стандартное отклонение величина X_2 в среднем растёт на Δ своих стандартных отклонений, то и с ростом X_2 на одно своё стандартное отклонение величина X_1 в среднем растёт на Δ своих стандартных отклонений.

Здесь задачи от Фрэнсиса Галтона?

4 Корреляция по-русски

Обычно в учебниках даётся такое определение корреляции

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}. \quad (4.1)$$

Естественно, возникает вопрос: «С какого перепугу? Почему это мы делим ковариацию на что-то там?»

Мы дадим естественное определение корреляции:

Определение 2. *Корреляция величин X_1 и X_2 — это среднее геометрическое среднего изменения X_1 при росте X_2 на единицу и среднего изменения X_2 при росте X_1 на единицу, совпадающее знаком с этими изменениями:*

$$\text{Corr}(X, Y) = \sqrt{\beta_{XY} \beta_{YX}} \cdot \text{sgn}(\beta_{XY})$$

Тут может возникнуть естественный вопрос: а почему именно среднее геометрическое, а не более привычное среднее арифметическое? Проблема со средним арифметическим становится понятна, если посмотреть на единицы измерения. Если β_{12} измеряется в килограммах на сантиметр, то β_{21} будет измеряться в сантиметрах на килограмм. И тогда совершенная ерунда получается, если считать среднее арифметическое. А среднее геометрическое автоматически оказывается безразмерной величиной.

Два определения корреляции полностью эквивалентны, действительно:

$$\text{Corr}(X, Y)^2 = \frac{\text{Cov}(X, Y)^2}{\text{Var}(X) \text{Var}(Y)} = \beta_{12} \cdot \beta_{21}$$

Посмотрим, как коэффициент корреляции связан со стандартизованными коэффициентам регрессии:

$$\beta_{12} \cdot \beta_{21} = \beta_{12}^{st} \cdot \beta_{21}^{st} = (\beta_{12}^{st})^2 = (\beta_{21}^{st})^2$$

То есть коэффициент корреляции просто равен стандартизованному коэффициенту регрессии! Поэтому корреляция между величинами X и Y также показывает, на сколько своих стандартных отклонений в среднем изменяется X при изменении Y на одно своё стандартное отклонение.

5 Геометрический смысл корреляции

Давайте рисовать случайные величины векторами-стрелочками! Не в том смысле, что у стрелочки случайное направление или длина, а в том смысле, что направление и длина стрелочки описывают характеристики этой случайной величины.

Любую геометрию можно задать, задав скалярное произведение. Действительно, если мы умеем считать скалярное произведение двух любых векторов, $\langle \vec{a}, \vec{b} \rangle$, то длина вектора считается ровно как в 9-м классе:

$$|\vec{a}| = \sqrt{\langle \vec{a}, \vec{a} \rangle}$$

И также любой девятиклассник помнит, что косинус угла между векторами считается как

$$\cos(\vec{a}, \vec{b}) = \frac{\langle \vec{a}, \vec{b} \rangle}{|\vec{a}| |\vec{b}|}$$

Мы определим скалярное произведение двух случайных величин как их ковариацию:

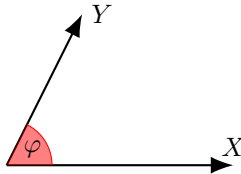
$$\langle X, Y \rangle = \text{Cov}(X, Y)$$

При таком подходе длиной случайной величины окажется стандартное отклонение:

$$\sqrt{\text{Cov}(X, X)} = \sqrt{\text{Var}(X)} = \sigma_X$$

А корреляция окажется косинусом угла между случайными величинами:

$$\cos \varphi = \cos(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \text{Corr}(X, Y)$$

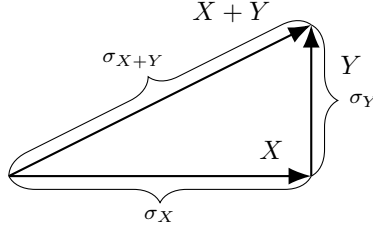


Значит в нашей геометрии длина стрелочки — стандартное отклонение случайной величины, а косинус угла между двумя стрелочками — это корреляция двух случайных величин. Дисперсия, следовательно, это квадрат длины случайной величины. Перпендикулярными случайными величинами будут те, косинус угла между которыми равен нулю, то есть некоррелированные. Автоматически оказывается, что корреляция лежит в диапазоне от (-1) до 1 .

Например, сформулируем в данной геометрии теорему Пифагора. Если случайные величины X и Y перпендикулярны (корреляция или ковариация равны нулю), то

дисперсия их суммы (квадрат длины гипотенузы) равен сумме их дисперсий (сумму квадратов длин катетов):

$$\text{Var}(X + Y) = \sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 = \text{Var}(X) + \text{Var}(Y)$$



Введение геометрии позволяет говорить о проекции. Например, можно спроецировать случайную величину Y на множество случайных величин пропорциональных величине X , $\{cX | c \in \mathbb{R}\}$. Если на обычной плоскости спроецировать вектор \vec{a} на прямую, порожденную вектором \vec{b} , то получится $\cos(\vec{a}, \vec{b}) \cdot \vec{b}$. По аналогии, если спроецировать случайную величину Y на множество $\{cX | c \in \mathbb{R}\}$, то получится $\text{Corr}(X, Y) \cdot X$. Другими словами, среди случайных величин пропорциональных X величина $\hat{Y} = \text{Corr}(X, Y) \cdot X$ — самая похожая на величину Y .

Понятие проекции позволяет интерпретировать квадрат корреляции. Квадрат косинуса равен отношению квадрата длины прилежащего катета $\text{Var}(\hat{Y})$ к квадрату гипотенузы $\text{Var}(Y)$.

(картинка)

Следовательно, $\text{Corr}(X, Y)^2 = \frac{\text{Var}(\hat{Y})}{\text{Var}(Y)}$, то есть квадрат корреляции показывает долю дисперсии Y , которую можно объяснить с помощью величин пропорциональных X .

6 Великое равенство

Если собрать воедино все кусочки, то мы доказали, что для двух случайных величин X_1 и X_2 :

$$\text{Corr}^2(X_1, X_2) = \beta_{12} \cdot \beta_{21} = \beta_{12}^{st} \cdot \beta_{21}^{st} = (\beta_{12}^{st})^2 = (\beta_{21}^{st})^2 = \frac{\text{Var}(\hat{X}_1)}{\text{Var}(X_1)} = \frac{\text{Var}(\hat{X}_2)}{\text{Var}(X_2)}$$

7 Корреляция и независимость

Теорема 1. Случайные величины X и Y независимы тогда и только тогда, когда некоррелированы любые функции $f(X)$ и $g(Y)$.

Другими словами для независимости X и Y необходима некоррелированность пар X и Y , X^2 и $\cos(Y)$, $\exp(X)$ и $1/Y$, и так далее. Из этого следует, что некоррелированность X и Y является необходимым, но недостаточным условием для независимости.

Можно выделить три «степени» независимости случайных величин X и Y :

Некоррелированность Y и X	$\text{Corr}(X, Y) = 0$
$\mathbb{E}(Y X) = \mathbb{E}(Y)$	$\text{Corr}(f(X), Y) = 0$ для всех $f()$
Независимость Y и X	$\text{Corr}(f(X), g(Y)) = 0$ для всех $f()$ и $g()$

Многие ошибочно считают, что если величина X имеет нормальное распределение $N(\mu_X, \sigma_X^2)$ и величина Y имеет нормальное распределение $N(\mu_Y, \sigma_Y^2)$, и X и Y некоррелированы, то они независимы. Это неверно.

Контрпример. Случайная величина X имеет стандартное нормальное распределение $N(0; 1)$, случайная величина Z независима от X и равновероятно принимает значения -1 и $+1$. Определим величину Y как их произведение, $Y = XZ$.

В этом примере величины X и Y зависимы, так как $|X| = |Y|$. Однако Y распределена нормально стандартно и $\text{Corr}(X, Y) = 0$.

Правильная теорема звучит так:

Теорема 2. *Если некоррелированные случайные величины X и Y имеют совместное нормальное распределение, то X и Y независимы.*

Попутно упомянем ещё одно неожиданное свойство предъявленного контрпримера. Если случайные величины нормальны по отдельности, то вполне возможно, что их сумма ненормальна. Для пары величин, имеющих совместное нормальное распределение, это невозможно.

8 Более двух случайных величин

Если случайных оказывается больше двух, то для описания зависимости с помощью одного числа оказывается очень много вариантов. Среди популярных можно назвать:

1. коэффициент бета в множественной регрессии
2. стандартизированный коэффициент множественной регрессии
3. частная корреляция
4. каноническая корреляция

9 Стандартизованный коэффициент регрессии

Для нахождения стандартизованного коэффициента множественной регрессии используется разложение

$$\frac{Y}{\sigma_Y} = \rho_{XY|Z} \cdot \frac{X}{\sigma_X} + \rho_{YZ|X} \frac{Z}{\sigma_Z} + \varepsilon$$

Два подхода к определению стандартизованного коэффициента регрессии эквивалентны в силу теоремы Фриша-Во-Ловелла (Frisch–Waugh–Lovell). Обычно эта теорема формулируется применительно к регрессии, а здесь мы приведём её вариант для случайных величин.

Теорема 3. *Если имеют место разложения:*

$$Y = a_1 Z_1 + a_2 Z_2 + \dots + a_n Z_n + \tilde{Y}, \text{ где } \tilde{Y} \perp Z_1, Z_2, \dots, Z_n$$

и

$$X = b_1 Z_1 + b_2 Z_2 + \dots + b_n Z_n + \tilde{X}, \text{ где } \tilde{X} \perp Z_1, Z_2, \dots, Z_n$$

То в разложениях

$$\tilde{Y} = d\tilde{X} + \varepsilon, \text{ где } \varepsilon \perp \tilde{X}$$

и

$$Y = c_1 Z_1 + c_2 Z_2 + \dots + c_n Z_n + dX + u, \text{ где } u \perp Z_1, Z_2, \dots, Z_n, X$$

коэффициенты при \tilde{X} и X совпадают.

10 Частная корреляция

Есть два словесных определения частной корреляции. Первое — через коэффициенты в регрессиях:

Определение 3. Частная корреляция между величинами X_1 и X_2 при фиксированной величине X_3 — это среднее геометрическое между коэффициентом при X_2 в регрессии X_1 на остальные переменные и коэффициентом при X_1 в регрессии X_2 на остальные переменные, совпадающее знаком с этими коэффициентами.

Попутно первое определение использует нетривиальный факт, что знак при X_1 в регрессии X_2 на остальные переменные и знак при X_2 в регрессии X_1 на остальные переменные совпадают.

Второе определение — через остатки и обычную корреляцию:

Определение 4. Частная корреляция между величинами X_1 и X_2 при фиксированной величине X_3 — это есть обычная корреляция между остатками регрессии X_1 на X_3 и остатками регрессии X_2 на X_3 .

Согласно второму определению корреляция показывает на сколько своих стандартных отклонений в среднем вырастет величина Y , очищенная от связи с Z , при росте величины X , очищенной от связи с Z , на одно своё стандартное отклонение.

Второе определение даёт следующий подход к подсчёту частной корреляции:

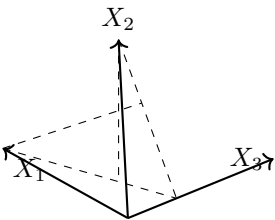
1. Спроецируем X на множество величин, некоррелированных с Z . Получим \tilde{X} .
2. Спроецируем Y на множество величин, некоррелированных с Z . Получим \tilde{Y} .
3. Частная корреляция между X и Y при фиксированной Z — это обычная корреляция между \tilde{X} и \tilde{Y} .

доказательство эквивалентности подходов?

Второе определение можно дополнить неожиданной теоремой:

Теорема 4. Частная корреляция между величинами X_1 и X_2 при фиксированной величине X_3 — это есть обычная корреляция между остатками регрессии X_1 на X_2, X_3 и остатками регрессии X_2 на X_1, X_3 , взятая с противоположным знаком.

Геометрическое доказательство:



11 Частная дисперсия и ковариация

Определения частной корреляции можно сформулировать с помощью частной ковариации и частной дисперсии.

$$\text{pCorr}(X, Y; Z) = \frac{\text{pCov}(X, Y; Z)}{\sqrt{\text{pVar}(X; Z) \text{pVar}(Y; Z)}}$$

Частная дисперсия величины X очищенной от Z определяется как

$$\text{pVar}(X; Z) = \text{Var}(X - aZ), \text{ где константа } a \text{ такова, что } \text{Cov}(X - aZ, Z) = 0$$

Частная ковариация

$$\text{pCov}(X, Y; Z) = \text{Cov}(X - aZ, Y - bZ), \text{ где константы } a \text{ и } b \text{ такова, что } \text{Cov}(X - aZ, Z) = 0$$

12 Условная корреляция

Определение 5. Условная корреляция между величинами X и Y при известном значении величины Z показывает на сколько своих стандартных отклонений σ_Y в среднем вырастет Y при росте величины X на одно своё стандартное отклонение σ_X при заданном значении величины Z .

Следует подчеркнуть одно существенное отличие условной корреляции от обычной и частной. Обычная и частная корреляция являются константами. Условная корреляция $\text{Corr}(X, Y|Z)$ является функцией от Z . Величина Z является случайной, поэтому и условная корреляция $\text{Corr}(X, Y|Z)$ является случайной величиной.

Здесь регрессионное определение???

Чуть более формальное определение:

Определение 6.

$$\text{Corr}(X, Y|Z) = \frac{\text{Cov}(X, Y|Z)}{\sqrt{\text{Var}(X|Z) \text{Var}(Y|Z)}},$$

где $\text{Cov}(X, Y|Z) = \mathbb{E}(XY|Z) - \mathbb{E}(X|Z)\mathbb{E}(Y|Z)$ и $\text{Var}(X|Z) = \mathbb{E}(X^2|Z) - (\mathbb{E}(X|Z))^2$

Пример подсчета частной и условной корреляций.

Пример 1.

Закон распределения случайных величин X_1, X_2, X_3 задан двумя таблицами:

$X_3 = 0$			$X_3 = 1$		
	$X_2 = 0$	$X_2 = 1$		$X_2 = 0$	$X_2 = 1$
$X_1 = 0$	0.06	0.08	$X_1 = 0$	0.1	0
$X_1 = 1$	0.24	0.32	$X_1 = 1$	0	0.2

Найдите условную корреляцию $\text{Corr}(X_1, X_2|X_3)$ и частную корреляцию $\text{pCorr}(X_1, X_2; X_3)$

Решение.

Эти две таблички на самом деле реализуют простую мысль: при $X_3 = 1$ величины X_1 и X_2 связаны детерминистически линейно, а при $X_3 = 0$ величины X_1 и X_2 независимы.

Считаем две вспомогательные условные корреляции, $\text{Corr}(X_1, X_2|X_3 = 0) = 0$, $\text{Corr}(X_1, X_2|X_3 = 1) = 1$.

Отсюда получаем, что $\text{Corr}(X_1, X_2|X_3) = X_3$. Для дискретных случайных величин запись условного ожидания не однозначна, и, например, ответ $\text{Corr}(X_1, X_2|X_3) = X_3^2$ также будет верным.

Пример 2.

Величины X_1, X_2, X_3 имеют совместное нормальное распределение с математическим ожиданием $\mathbb{E}(X) = (1, 2, -3)'$ и ковариационной матрицей

$$\text{Var}(X) = \begin{pmatrix} 9 & 2 & -1 \\ 2 & 16 & 1 \\ -1 & 1 & 4 \end{pmatrix}$$

Найдите условную корреляцию $\text{Corr}(X_1, X_2|X_3)$ и частную корреляцию $\text{pCorr}(X_1, X_2; X_3)$.

Решение.

...

В данном примере частная и условная корреляция совпали. Это одно из приятных свойств многомерного нормального распределения:

Теорема 5. Если величины X, Y и Z имеют совместное нормальное распределение, то:

1. частная и условная дисперсии совпадают
2. частная и условная ковариации совпадают
3. частная и условная корреляции совпадают

Прикольный пример. $X \sim U[0; 1]$. При фиксированном X величины Y и Z условно независимы и одинаково распределены $Y|X \sim U[0; X^2]$ и $Z|X \sim U[0; X^2]$. В данном случае условная корреляция $\text{Corr}(Y, Z|X) = 0$, однако $\text{pCorr}(Y, Z; X) \neq 0$.

This example was suggested by P. Groeneboom, and published in Kurowicka (2001).

$$\text{pCorr}(Y, Z; X) = \frac{1}{13}$$

13 Случай равных дисперсий

Если дисперсии двух рассматриваемых величин X_1 и X_2 равны, то стандартизированный коэффициент регрессии равен обычному коэффициенту регрессии, $\beta_{12} = \beta_{12}^{st}$.

Если же выполнено более сильное условие стационарности случайного процесса X_1, X_2, \dots , то

$$\text{pCorr}(X_1, X_k; X_{2,3,\dots,(k-1)}) = \beta_{1k} = \beta_{k1} = \beta_{1k}^{st} = \beta_{k1}^{st}$$

К данному случаю относится частная автокорреляционная функция для стационарных процессов:

Пример 3. AR(1) процесс

$$y_t = 0.7y_{t-1} + u_t$$

Найдите $\text{pCorr}(y_t, y_{t-3}; y_{t-1}, y_{t-2})$.

Частная корреляция в данном случае обычному коэффициенту регрессии, то есть коэффициенту перед y_{t-3} в регрессии y_t . Замечаем, что

$$y_t = 0.7y_{t-1} + 0y_{t-2} + 0y_{t-3} + u_t$$

Значит, $\text{pCorr}(y_t, y_{t-3}; y_{t-1}, y_{t-2}) = 0$.

14 Каноническая корреляция

Если разбить все случайные величины на две группы, скажем на иксы и игреки, то каноническая корреляция — это максимальная корреляция между линейной комбинацией иксов и линейной комбинацией игреков.

$$\text{canCorr}(Y_1, Y_2, Y_3; X_1, X_2) = \max_{\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3} \text{Corr}(\alpha_1 X_1 + \alpha_2 X_2, \beta_1 Y_1 + \beta_2 Y_2 + \beta_3 Y_3)$$

Если группа игреков состоит из одной случайной величины, то мы получаем коэффициент множественной корреляции, квадрат которого можно назвать «теоретическим R^2 ».

$$\text{canCorr}(Y; X_1, X_2, \dots, X_k)$$

Каноническая корреляция — это замечательное оправдание для скорректированного коэффициента R_{adj}^2 . А именно:

Если $\text{canCorr}(Y; X_1, X_2, \dots, X_k) = 0$, то $\mathbb{E}(R_{adj}^2) = 0$.

И Шахерезаду застигло утро, и она прекратила дозволенные речи, заметив, что вышла замечательная статья Димы Борзых о методе канонических корреляций.

15 О двойственности

В математике часто встречаются двойственные преобразования. Если двойственное волшебство f применить два раза, то получится исходный объект, то есть $f(f(A)) = A$.

- Положительные числа и $f(x) = 1/x$. Действительно, $f(f(x)) = x$.
 - Матрицы и транспонирование. Конечно же, $(A^T)^T = A$.
 - Случайные величины с положительной дисперсией и $f(X) = X/\text{Var}(X)$. Убедитесь самостоятельно, что $f(f(X)) = X$.
 - Папа как двойственное преобразование орущего младенца.
- Объекты A и $f(A)$ также называют двойственными. Приведите свои примеры двойственных преобразований.

16 Матрица—Мать всех регрессий и корреляций

Оказывается есть матрица M , глянув на которую, можно с помощью пары арифметических действий посчитать любой коэффициент в любой регрессии.

Пусть X — это вектор случайных величин $X = (X_1, X_2, \dots, X_k)'$ с ковариационной матрицей $C = \text{Var}(X)$.

От величин X_1, X_2, \dots, X_k мы перейдём к двойственным величинам $X_1^*, X_2^*, \dots, X_k^*$. Чтобы перейти от $\{X_i\}$ к $\{X_i^*\}$ надо применить ровно то же преобразование, что и для перехода от $\{X_i^*\}$ к $\{X_i\}$. Сейчас мы это преобразование опишем.

По своей сути X_i^* будет пропорциональна той части X_i , которая не коррелирована ни с одним оставшимся X_j . Строить величину X_i^* мы будем как линейную комбинацию X_1, X_2, \dots, X_k с некими весами μ_{ij} ,

$$X_i^* = \sum_{t=1}^k \mu_{it} X_t$$

Коэффициенты μ_{ik} должны быть такими, чтобы выполнялось условие:

$$\text{Cov}(X_i^*, X_j) = \begin{cases} 1, & \text{при } i = j \\ 0, & \text{при } i \neq j \end{cases}$$

Из этой системы видна двойственность $\{X_i^*\}$ и $\{X_i\}$. Если поменять «звезданутые» X_i и исходные X_i , система уравнений ни капли не изменится.

Первый смысл коэффициента μ_{it} — вес, с которым X_t входит в величину X_i^* . Заметим, что есть и другая интерпретация, а именно, эти коэффициенты показывают ковариацию между новыми X_i^* :

$$\text{Cov}(X_i^*, X_j^*) = \text{Cov}(X_i^*, \sum_t \mu_{jt} X_t) = \mu_{ji}$$

Из этого, в частности, следует, что $\mu_{ij} = \mu_{ji} = \text{Cov}(X_i^*, X_j^*)$ и $\mu_{ii} = \text{Var}(X_i^*)$.

Из условия:

$$X_i^* = \sum_{t=1}^k \mu_{ik} X_t$$

Следует также представление:

$$X_i = \sum_{t \neq i} -\frac{\mu_{ik}}{\mu_{ii}} X_t + \frac{1}{\mu_{ii}} X_i^*$$

То есть $\frac{1}{\mu_{ii}} X_i^*$ — это остаток от регрессии X_i на остальные X_j . Заметим попутно, что дисперсия этого остатка равна

$$\text{Var}\left(\frac{1}{\mu_{ii}} X_i^*\right) = \frac{\mu_{ii}}{\mu_{ii}^2} = \frac{1}{\mu_{ii}}$$

Выпишем, как связаны коэффициенты μ_{ij} , ковариации между новыми X_i^* , со ковариациями между X_i :

$$\text{Cov}(X_i^*, X_j) = \sum_{k=1}^n \mu_{ik} \text{Cov}(X_k, X_j) = \begin{cases} 1, & \text{при } i = j \\ 0, & \text{при } i \neq j \end{cases} \quad (16.1)$$

В матричном виде это же соотношение записывается потрясающе просто:

$$M \cdot C = I$$

То есть $M = C^{-1}$. Таким образом получается, что обратная ковариационная матрица позволяет легко посчитать коэффициенты любой регрессии!

Например, мы хотим посчитать коэффициенты в регрессии X_i на остальные иксы:

$$X_i = \beta_{i1}X_1 + \beta_{i2}X_2 + \dots + \beta_{in}X_k + u_i$$

Оказывается, что

$$\beta_{it} = -\frac{\mu_{it}}{\mu_{ii}} = -\frac{C_{it}^{-1}}{C_{ii}^{-1}}$$

Кроме того, дисперсия остатка равна

$$\text{Var}(u_i) = \frac{1}{\mu_{ii}} = \frac{1}{C_{ii}^{-1}}$$

Осталось заметить, что

$$\text{pCorr}(X_i, X_j; \dots) = -\frac{\text{Cov}(X_i^*, X_j^*)}{\sqrt{\text{Var}(X_i^*) \text{Var}(X_j^*)}} = -\frac{\mu_{ij}}{\sqrt{\mu_{ii}\mu_{jj}}} = -\frac{C_{ij}^{-1}}{\sqrt{C_{ii}^{-1}C_{jj}^{-1}}}$$

Почему эта формула не срабатывает при $i = j$? При $i = j$ не верной оказывается неожиданная теорема о противоположном знаке частных корреляций.

Попытожим. Мы доказали, что обратная ковариационная матрица имеет следующую структуру:

$$C^{-1} = \begin{pmatrix} \frac{1}{\text{pVar}(X_1; X_{-1})} & -\frac{\beta_{12}}{\text{pVar}(X_1; X_{-1})} \\ -\frac{\beta_{21}}{\text{pVar}(X_1; X_{-1})} & \frac{1}{\text{pVar}(X_2; X_{-2})} \\ & & \frac{1}{\text{pVar}(X_3; X_{-3})} \end{pmatrix},$$

Где β_{it} — коэффициент перед X_t в регрессии X_i на остальные иксы, а $\text{pVar}(X_1; X_{-1})$ — дисперсия остатка в той же регрессии.

Вспомним, что наши исходные переменные $\{X_i\}$ и «звезданутые» переменные $\{X_i^*\}$ являются двойственными. По смыслу X_i^* является делённым на свою дисперсию остатком в регрессии X_i на остальные переменные. Значит и X_i является делённым на свою дисперсию остатком в регрессии X_i^* на остальные «звезданутые» переменные.

Похожий результат верен и для обратной корреляционной матрицы. Если A — корреляционная матрица, то:

$$A^{-1} = \begin{pmatrix} \frac{\text{Var}(X_1)}{\text{pVar}(X_1; X_{-1})} & & \\ & \frac{\text{Var}(X_2)}{\text{pVar}(X_2; X_{-2})} & \\ & & \frac{\text{Var}(X_3)}{\text{pVar}(X_3; X_{-3})} \end{pmatrix}$$

Частная корреляция между X_i и X_j при фиксированных остальных иксах равна:

$$\text{pCorr}(X_i, X_j; \dots) = -\frac{A_{ij}}{\sqrt{A_{ii}A_{jj}}}$$

Стандартизованный коэффициент при X_j в регрессии X_i на остальные иксы равен:

$$\beta_{ij}^{st} = -\frac{A_{ij}}{A_{ii}}$$

Квадрат коэффициента множественной корреляции X_i с остальными иксами равен:

$$\text{canCorr}(X_i; \dots) = 1 - \frac{1}{A_{ii}}$$

17 Численный пример

Дана конкретная ковариационная матрица...

18 Разложение R^2 по факторам

Возьмём 1-ую строку из матрицы M и помножим на 1-ый столбец из матрицы C . Получим равенство:

$$\frac{\text{Var}(X_1)}{\text{pVar}(X_1; X_{-1})} - \beta_{12} \frac{\text{Cov}(X_1, X_2)}{\text{pVar}(X_1; X_{-1})} - \beta_{13} \frac{\text{Cov}(X_1, X_3)}{\text{pVar}(X_1; X_{-1})} - \dots = 1$$

Домножив на общий знаменатель, поделив на $\text{Var}(X_1)$, и, перенеся слагаемые, получаем:

$$\beta_{12} \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_1)} + \beta_{13} \frac{\text{Cov}(X_1, X_3)}{\text{Var}(X_1)} + \dots = 1 - \frac{\text{pVar}(X_1; X_{-1})}{\text{Var}(X_1)}$$

У данной теоретической формулы есть выборочный собрат:

$$\hat{\beta}_{12} \frac{s\text{Cov}(X_1, X_2)}{s\text{Var}(X_1)} + \hat{\beta}_{13} \frac{s\text{Cov}(X_1, X_3)}{s\text{Var}(X_1)} + \dots = R^2$$

Здесь под R^2 мы подразумеваем коэффициент детерминации в регрессии переменной X_1 на остальные.

19 Частные корреляции у стационарного процесса

Произвольная матрица не может быть корреляционной матрицей вектора случайных величин. Она должна быть симметричной и неотрицательно определённой, а на диагонали должны стоять единицы.

Под матрицей частных корреляций можно подразумевать разные вещи.

Вариант А: элемент (i, j) — это частная корреляция между X_i и X_j при фиксированной посторонней величине Z .

Здесь неотрицательно определена?

Вариант Б: элемент (i, j) — это частная корреляция между X_i и X_j при фиксированных всех остальных X_k .

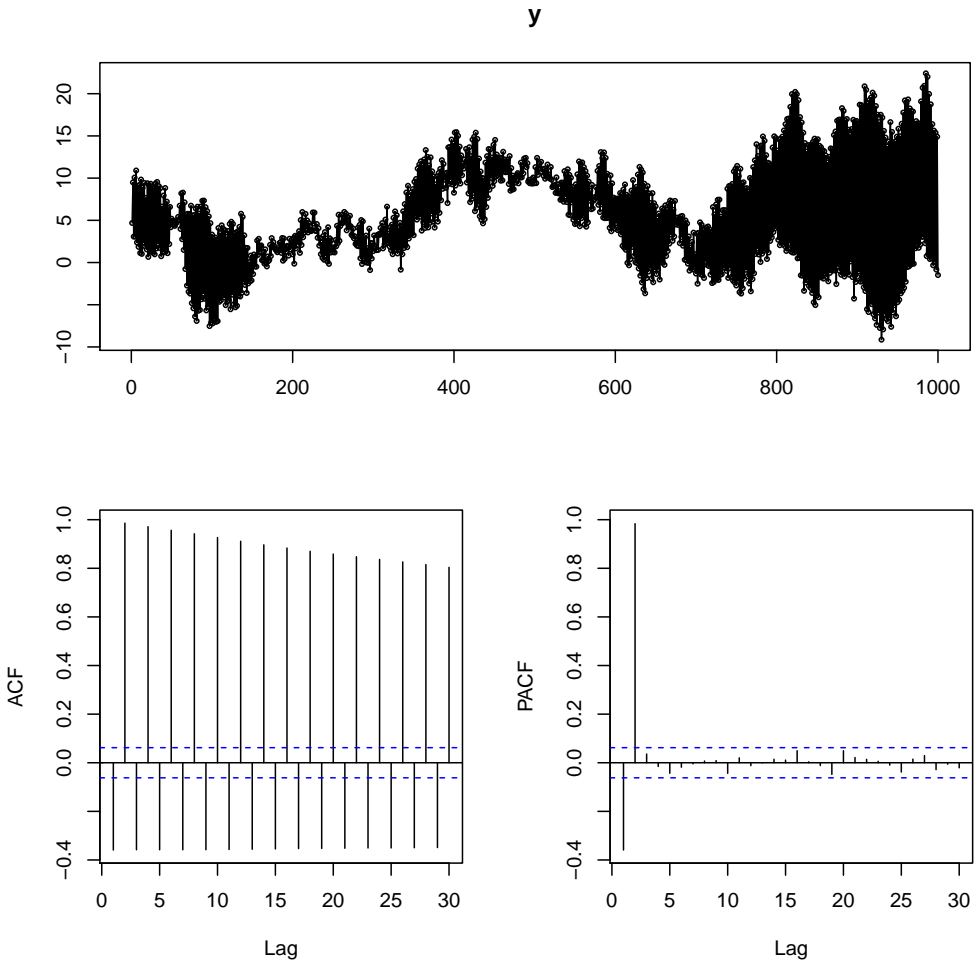
Здесь $B = 2I - C$, где C — корреляционная матрица?

У стационарного процесса может быть совершенная любая частная автокорреляционная функция со значениями в диапазоне $(-1; 1)$. Например, у стационарного процесса может быть частная автокорреляционная функция

$$\varphi_{kk} = \begin{cases} 0, & \text{если } k = 1 \\ 0.99, & \text{если } k = 2 \\ 0, & \text{если } k \geq 3 \end{cases}$$

А вот и он, $y_t = 0.99y_{t-2} + \varepsilon_t$:

```
library("forecast")
y <- arima.sim(n = 1000, model = list(ar = c(0, 0.99)))
tsdisplay(y)
```



20 Почиташки

https://en.wikipedia.org/wiki/Schur_complement#Applications_to_probability_theory_and_statistics
<https://stats.stackexchange.com/questions/140080/why-does-inversion-of-a-c>
<https://stats.stackexchange.com/questions/10795/how-to-interpret-an-invers>
<http://epublications.bond.edu.au/cgi/viewcontent.cgi?article=1147&context=ejsie>
<https://cran.r-project.org/web/packages/corpcor/corpcor.pdf>
<http://projecteuclid.org/euclid.aos/1176342881>
http://www.gragusa.org/teaching/et/files/matrix_notes.pdf
<http://www.stata-journal.com/sjpdf.html?articlenum=st0285>
https://www.hse.ru/data/2016/02/15/1140408320/WP2_2016_01_F.pdf
http://www.ewi.tudelft.nl/fileadmin/Faculteit/EWI/Over_de_faculteit/Afdeling_Applied_Mathematics/Risico_en_Beslissingen_Analyse/Theses/MWawrzyzniak_thesis.pdf
http://www.ewi.tudelft.nl/fileadmin/Faculteit/EWI/Over_de_faculteit/Afdeling_Applied_Mathematics/Risico_en_Beslissingen_Analyse/Theses/MWawrzyzniak_thesis.pdf
<http://iroha.scitech.lib.keio.ac.jp:8080/sigma/bitstream/handle/10721/1899/document.pdf?sequence=4>

21 Выборочные характеристики

это в статью два?

В теории обычную корреляцию и частную корреляцию можно посчитать, если известен закон распределения случайных величин. На практике закон распределения не известен, однако доступны наблюдения. Как по имеющимся наблюдениям оценить неизвестные корреляции?

Несколько способов оценки корреляции

Здесь пара картинок: википедийная с корреляциями и два ряда случайного блуждания/тренда

Несколько способов оценки частной корреляции

Оценкой для частной дисперсии X_1 при фиксированных X_2, \dots, X_n является оценка дисперсии случайной составляющей в регрессии X_1 на X_2, \dots, X_n

А ещё частную корреляцию можно посчитать в два шага. Построить регрессию X_1 на остальные переменные. Взять t -статистику перед X_2 и воспользоваться формулой:

$$\text{pCorr}(X_1, X_2; X_{-1, -2}) = \frac{t^2}{t^2 + (n - k)}$$

Здесь про зависимости возникающие при оценке автокорреляционной и частной автокорреляционной функции для случайных процессов.

Несколько способов оценки множественной корреляции:

...

Обратная к выборочной ковариационной матрице

$$\hat{C}^{-1} = \begin{pmatrix} \frac{1}{\hat{\sigma}_1^2} & & \\ & \frac{1}{\hat{\sigma}_2^2} & \\ & & \frac{1}{\hat{\sigma}_3^2} \end{pmatrix}$$

Если мы предположим, что разные наблюдения в выборке независимы и одинаково распределены, то получим такие выборочные оценки

Теоретический	Выборочный
$\text{pVar}(y_i; z_i)$	$\hat{\sigma}_u^2$ в регрессии y
$\text{pCov}(y_i, x_i; z_i)$	выборочная кова
$\text{capCorr}(y_i; x_{i1}, x_{i2}, \dots, x_{ik})$	R^2, R_{adj}^2
коэффициент в множественной регрессии	$\hat{\beta} = (X'X)^{-1}X'y$
стандартизированный коэффициент в множественной регрессии	$\hat{\beta}_j^{st} = \hat{\beta}_j \frac{\hat{\sigma}_x}{\hat{\sigma}_y}, \hat{\beta}^{st} = \Sigma$
обычная корреляция	$sCorr(y, x) = \Sigma$
частная корреляция	корреляция оста

22 Дисперсия МНК-оценок

Обратные матрицы окружают нас, они повсюду! Как известно, в модели множе-ственной регрессии

$$\text{Var}(\hat{\beta}|X) = \sigma^2(X'X)^{-1}$$

На диагонали этой матрицы стоят дисперсии отдельных $\hat{\beta}_j$. И их можно получить в явном виде! Снова нам на помощь придёт матрица-мать всех регрессий!

Чтобы понять, что у неё на диагонали применим аналогичный трюк. Пусть x_i — это столбцы матрицы X . От векторов x_i перейдём к векторам x_i^* . По своей сути x_i^* будет пропорционален остатку в регрессии x_i на остальные x_j . Строить величину x_i^* мы будем как линейную комбинацию x_1, x_2, \dots, x_k с некими весами μ_{ij} ,

$$x_i^* = \sum_{t=1}^k \mu_{it} x_t$$

Коэффициенты μ_{it} должны быть такими, чтобы выполнялось условие:

$$\langle x_i^*, x_j \rangle = \begin{cases} 1, & \text{при } i = j \\ 0, & \text{при } i \neq j \end{cases}$$

Первый смысл коэффициента μ_{it} — вес, с которым вектор x_t входит в вектор x_i^* . Заметим, что есть и другая интерпретация, а именно, эти коэффициенты показывают скалярное произведение между новыми x_i^* :

$$\langle x_i^*, x_j^* \rangle = \langle x_i^*, \sum_k \mu_{jk} x_k \rangle = \mu_{ji}$$

Из этого, в частности, следует, что $\mu_{ij} = \mu_{ji} = \langle x_i^*, x_j^* \rangle$ и $\mu_{ii} = |x_i^*|^2$.

Из условия:

$$x_i^* = \sum_{t=1}^k \mu_{ik} x_t$$

Следует также представление:

$$x_i = \sum_{t \neq i} -\frac{\mu_{it}}{\mu_{ii}} x_t + \frac{1}{\mu_{ii}} x_i^*$$

То есть $\frac{1}{\mu_{ii}} x_i^*$ — это остаток от регрессии x_i на остальные x_j . Заметим попутно, что квадрат длины этого остатка, или RSS_i , равен:

$$\left| \frac{1}{\mu_{ii}} x_i^* \right|^2 = \frac{\mu_{ii}}{\mu_{ii}^2} = \frac{1}{\mu_{ii}}$$

Выпишем, как связаны коэффициенты μ_{ij} со скалярными произведениями между x_i :

$$\langle x_i^*, x_j \rangle = \sum_{t=1}^n \mu_{it} \langle x_t, x_j \rangle = \begin{cases} 1, & \text{при } i = j \\ 0, & \text{при } i \neq j \end{cases} \quad (22.1)$$

В матричном виде это же соотношение записывается потрясающе просто:

$$M \cdot C = I$$

То есть $M = C^{-1}$. Таким образом получается, что обратная ковариационная матрица позволяет легко посчитать коэффициенты любой регрессии!

В частности, на диагонали у неё находятся числа $\mu_{ii} = 1/RSS_i$.

23 Геометрия обратной матрицы

Если:

1. r_1, r_2, \dots, r_n — это строки матрицы A ,
2. $c_1^{-1}, c_2^{-1}, \dots, c_n^{-1}$ — это столбцы матрицы A^{-1} ,
3. V_i — это гипер-плоскость, линейная оболочка всех столбцов обратной матрицы кроме столбца c_i^{-1} , то есть $V_i = \text{Lin}(c_1^{-1}, \dots, c_{i-1}^{-1}, c_{i+1}^{-1}, \dots, c_n^{-1})$

то:

1. Строка r_i перпендикулярна столбцу c_j^{-1} при $j \neq i$, $\langle r_i, c_j^{-1} \rangle = 0$
2. Скалярное произведение r_i и c_i^{-1} равно единице, $\langle r_i, c_i^{-1} \rangle = 1$
3. $\text{length}(r_i) \cdot \text{distance}(c_i^{-1}, V_i) = 1$.

Действительно, спроецируем столбец c_i^{-1} на V_i , получим вектор \hat{c}_i^{-1} . Получаем цепочку,

$$1 = \langle r_i, c_i^{-1} \rangle = \langle r_i, \hat{c}_i^{-1} + (c_i^{-1} - \hat{c}_i^{-1}) \rangle$$

Заметим, что и вектор r_i , и вектор $(c_i^{-1} - \hat{c}_i^{-1})$, перпендикулярны гипер-плоскости V_i . Поэтому угол между ними равен нулю. И, следовательно,

$$\langle r_i, \hat{c}_i^{-1} + (c_i^{-1} - \hat{c}_i^{-1}) \rangle = \langle r_i, (c_i^{-1} - \hat{c}_i^{-1}) \rangle = |r_i| \cdot |c_i^{-1} - \hat{c}_i^{-1}| \cdot \cos(0) = |r_i| \cdot |c_i^{-1} - \hat{c}_i^{-1}|$$

Для симметричной матрицы аналогичное утверждение верно про строки исходной матрицы и строки обратной.

<http://mathoverflow.net/questions/120884/>

Ещё один геометрический факт:

Если каждую строку r_i интерпретировать как точку в \mathbb{R}^n , то через эти n точек можно провести гипер-плоскость α (она вовсе не обязательно содержит начало координат, не стоит путать её с $\text{Lin}(r_1, r_2, \dots, r_n)$).

Обозначим также $\mathbb{1}$ — вектор столбец из единиц. Тогда вектор построчных сумм $A^{-1} \cdot \mathbb{1}$ перпендикулярен гипер-плоскости α .

Как это использовать в метрике?

24 До кучи

Есть малоизвестная теорема Кроба:

Теорема 6. *Рациональное равенство является тождеством для произвольного кольца, если и только если его можно доказать дополнительно используя бессмысленную в кольце формулу суммы бесконечно убывающей геометрической прогрессии $(1 - a)^2 = 1 + a + a^2 + a^3 + \dots$*

<http://mathoverflow.net/questions/31595>

Эта теорема позволяет легко и строго выводить многие матричные тождества:

Например. Докажите, что

$$(I - BA)^{-1} = I + B(I - AB)^{-1}A$$

Решение. Применим разрешённую Кробом формулу бесконечно убывающей геометрической прогрессии два раза:

$$\begin{aligned} (I - BA)^{-1} &= I + BA + BABA + BABABA + \dots = \\ &= I + B(I + AB + ABAB + \dots)A = I + B(I - AB)^{-1}A \end{aligned}$$

25 Упражнения

1. Докажите, что

$$(I + A)(I - BA)^{-1}(I + B) = (I + B)(I - AA)^{-1}(I + A)$$