



Регрессии: почему экономисты уделяют им так много внимания?

Родни Рамчаран

ЧТЕНИЕ — важный навык, и учителя начальных школ заметили, что умение читать у их учеников обычно улучшается вместе с увеличением размера обуви. Должны ли разработчики политики предлагать ученым награды за разработку методов увеличения размера обуви у учеников начальных школ, чтобы способствовать улучшению навыков чтения? Очевидно, что тенденция к одновременному увеличению размера обуви и улучшению умения читать не означает, что большой размер ноги является причиной улучшения навыков чтения. У детей более старшего возраста размер ноги больше, но они также имеют более развитый мозг. Такое естественное развитие детей объясняет то простое наблюдение, что размер ноги и умение читать имеют тенденцию к одновременному возрастанию, то есть между ними имеется положительная корреляция. Однако ясно, что здесь нет зависимости: больший размер ноги не является причиной улучшения умения читать.

В экономической науке корреляции широко распространены. Но нечасто бывает так уж легко определить, представляет ли корреляция между двумя или более переменными причинно-следственную связь. Страны, имеющие больший объем торговли с остальным миром, характеризуются также более высокими уровнями доходов, но означает ли это, что торговля приводит к повышению уровня доходов? Более образованные люди, как правило, зарабатывают больше, но следует ли из этого, что образование приводит к более высоким доходам? Знать точные

ответы на эти вопросы важно. Если бы дополнительные годы обучения приводили к повышению доходов, разработчики политики могли бы сократить бедность, выделив больше средств на образование. Если бы один лишний год образования вызывал увеличение доходов на 20 000 долларов в год, тогда выгоды от расходов на образование были бы гораздо больше, чем если бы один лишний год образования вызывал увеличение всего на 2 доллара в год.

Чтобы ответить на такие вопросы, экономисты применяют статистический инструмент, известный как регрессионный анализ. Регрессии используются для количественной оценки связи между одной переменной и другими переменными, которые, как предполагается, объясняют ее изменения; регрессии также позволяют установить, насколько тесной и четко определенной является такая связь. В настоящее время проведение расчетов по нескольким тысячам регрессий стало делом обычным и простым — хотя так было не всегда (см. вставку), — более того, трудно отыскать эмпирическое экономическое исследование, в котором не было бы регрессии. Регрессии также активно применяются в других областях, включая социологию, статистику и психологию.

Как построить регрессию

Чтобы проиллюстрировать механизм регрессии, рассмотрим более внимательно задачу, которая заключается в том, чтобы попытаться определить отдачу от образования. Государственные органы собирают данные об уровне образования населения и их доходах после его получения. Однако люди получают образование по разным причинам: одним легче учиться, чем другим, или они просто хотят оставаться в учебном заведении подольше. Другие могут успешно продвигаться по служебной лестнице в неакадемических областях и при этом достигать более высоких уровней доходов. Эти различные причины для посещения учебного заведения могут оказывать влияние на доходы, что затрудняет выяснение вопроса о том, представляет ли корреляция между обучением и доходами причинно-следственную связь или определяется каким-либо другим фактором. Людям, которым учение дается легче, возможно, также легче обучаться на рабочем месте, что приводит к более высоким заработкам. Таким образом, положительная связь между более высокими доходами и уровнем образования может отражать врожденные способности, а не результат образования.

Прежде чем строить регрессию, полезно иметь теоретическую модель, позволяющую объяснить, каким образом и

Волшебная сила компьютеров

Первый этап формирования концепции регрессии относится к XIX веку, но в действительности стремительный подъем регрессионного анализа начался в XX веке благодаря технологической революции, которая сделала настольные компьютеры его главной опорой. В 1950-е и 1960-е годы экономистам приходилось рассчитывать регрессии при помощи электромеханических настольных калькуляторов. Еще совсем недавно, в 1970-е годы, до 24 часов уходило только на то, чтобы получить результаты одной регрессии из центральной компьютерной лаборатории — и это после нескольких часов или дней, затраченных на пробивку компьютерных перфокарт. Одна неверная пробивка (ошибка в написании управляющей команды или неправильное значение показателя) могла свести на нет всю работу.

почему одна «зависимая» переменная определяется одной или несколькими «независимыми» или «объясняющими» переменными. Утверждение о том, что доходы человека зависят от его уровня образования, является одним из примеров простой модели с одной объясняющей переменной. Соответствующее уравнение регрессии, которое, как предполагается, является линейным, будет выглядеть следующим образом:

$$Y = a + bX$$

В левой стороне уравнения находится Y , зависимая переменная — доходы. В правой стороне находятся: a — константа (или отрезок) и b — коэффициент (или наклон), умноженный на X , независимую (или объясняющую) переменную — уровень образования. Эта регрессия показывает (в алгебраических символах), что «доходы зависят только от уровня образования, и эта зависимость линейна»; другие объясняющие факторы, если таковые есть, опускаются.

Предположим, однако, что мир устроен гораздо сложнее, и что влияние уровня образования на доходы может зависеть от целого ряда других факторов. В этом случае мы построили бы множественную регрессию, которая выглядела бы следующим образом:

$$Y = a + b_1X_1 + b_2X_2 + \dots$$

Теперь у нас имеется несколько переменных X для объяснения величины доходов Y — таких как способности, интеллект, возраст, уровень образования, семейное положение и уровень образования родителей. Коэффициент b попросту измеряет влияние каждой из этих переменных на доходы, при предположении о том, что другие переменные остаются постоянными.

Чем умнее, тем богаче?

Попытаемся построить регрессию на основе теории, согласно которой почасовая заработная плата (зависимая переменная) зависит от уровня образования (объясняющая переменная). Предположим, что другая возможная объясняющая переменная — способности, измеряемые при помощи коэффициента интеллектуальности (KI), — не оказывает влияния на заработную плату, помимо любого влияния, которое она может иметь благодаря образованию. Мы вводим в уравнение все данные о доходах и уровнях образования. Строим регрессию и получаем:

$$Y = 5,40 + 1,06 ОБР$$

Коэффициент b говорит о том, что один дополнительный год образования приводит к увеличению почасовой заработной платы в размере 1,06 доллара. А для тех, кто не имеет образования ($ОБР=0$), константа показывает, что средняя заработная плата составляет 5,40 доллара в час.

Но что произойдет, если включить в уравнение KI , то есть предположить, что доходы зависят и от уровня образования, и от KI ? Вводим в уравнение данные о результатах теста KI и получаем:

$$Y = 5,40 + 0,83EDU + 0,001KI$$

Мы узнаем, что лица, которые имели более высокие результаты теста KI , также получали более высокую почасовую

заработную плату. Кроме того, хотя влияние образования на заработную плату остается положительным, оно примерно на 27 процентов меньше, чем в случае, если бы мы не включили результаты KI (27 процентов получены на основе разности коэффициентов: $100 \cdot (1,06 - 0,83) / 0,83$). Это означает, что ранее мы переоценили воздействие образования на заработную плату, так как не учли влияния KI , который коррелирован с уровнем образования.

Потенциальные ошибки

Несмотря на свои достоинства, регрессии подвержены ошибкам и часто используются неправильно. Рассмотрим три основные трудности.

Пропущенные переменные. Необходимо иметь хорошую теоретическую модель для выбора переменных, объясняющих зависимую переменную. В случае простой регрессии с двумя переменными следует рассмотреть другие факторы, которые могут объяснять зависимую переменную. В нашем примере даже при включении KI корреляция между уровнем образования и доходами может отражать также некоторые другие факторы, не включенные в уравнение. То есть лица в выборке могут все же быть разными в том или ином «ненаблюдаемом» отношении, которое объясняет их последующие доходы, возможно, из-за выбора ими того или иного образования. Лица из состоятельных семей обычно имеют лучший доступ к образованию, но благосостояние семьи может также обеспечивать больше связей на рынке труда, приводя к более высоким доходам. Таким образом, благосостояние родителей, возможно, является еще одной переменной, которую следует включить в уравнение.

Обратная причинно-следственная связь. Многие теоретические модели прогнозируют двустороннюю причинно-следственную связь, то есть когда зависимая переменная может вызывать изменения в одной или нескольких объясняющих переменных. Например, более высокие доходы позволяют людям инвестировать больше средств в свое собственное образование, что в свою очередь, приводит к повышению их доходов. Это усложняет способ, при помощи которого должны оцениваться регрессии, и требует применения специальных методов.

Ошибки измерения. Факторы могут быть измерены неправильно. Например, способности с трудом поддаются измерению, а тесты KI страдают хорошо известными недостатками. Как результат, регрессия, в которой используется KI , может недостаточно хорошо фиксировать показатель способностей, что ведет к неточным или смещенным корреляциям между уровнем образования и доходами.

Излишне узкая направленность. Коэффициенты регрессии дают информацию только о том, как малые — а не крупные, — изменения одной переменной связаны с изменениями другой. Регрессия показывает, как небольшое изменение уровня образования скорее всего повлияет на доходы, но не позволит исследователю сделать общий вывод о воздействии крупных изменений. Если все одновременно получают высшее образование, новый выпускник вуза вряд ли сможет заработать значительно больше, так как общее количество выпускников вузов резко увеличится. ■

Родни Рамчаран — экономист в Исследовательском департаменте МВФ.