

Причинно-следственный анализ, часть 2

Из этого текста ты узнаешь:

- как провести оценку эффекта, когда воздействие оказывается не выборочно;
- как провести оценку эффекта, когда воздействие растянуто во времени;
- как провести оценку эффекта, когда нет хорошей контрольной группы.

1 Введение

В экспериментальной экономике и других социальных науках одним из ключевых вопросов является оценка причинно-следственных связей. В идеальном мире для этого можно было бы использовать рандомизированные эксперименты, однако в реальных условиях это часто невозможно или неэтично. Поэтому квазиэкспериментальные методы, такие как regression discontinuity design (RDD), synthetic control method (SCM) и difference-in-difference (DiD), играют важную роль в анализе данных и оценке воздействия.

Regression discontinuity design (RDD) используется для оценки эффектов при наличии чёткого порога, разделяющего группы. Этот метод нацелен на исследование наблюдений, которые находятся по обе стороны от порога.

Synthetic control method (SCM) помогает оценить воздействие на уровне группы, сравнивая результаты наблюдаемой группы с «синтетической» контрольной группой, созданной путём взвешивания данных нескольких групп.

Difference-in-difference (DiD) используется для анализа влияния событий или изменений на группы, предполагая, что если бы не произошло воздействия, разница в показателях между группами оставалась бы постоянной с течением времени.

Ниже будут рассмотрены все три перечисленных метода.

2 RDD

Regression discontinuity design (RDD) — это квазиэкспериментальный метод оценки причинно-следственных связей, который используется, когда решение о применении воздействия основано на некотором пороговом значении. Наблюдения, которые находятся по разные стороны от порога, можно рассматривать как «случайно» получившие или не получившие воздействия. Основная идея метода заключается в том, что наблюдения, которые находятся близко к порогу, имеют схожие характеристики, а потому разница между ними может быть объяснена только наличием или отсутствием воздействия.

Метод RDD был впервые предложен Дональдом Кэмпбеллом в 1960 году для анализа эффекта финансовой помощи в образовании. С тех пор метод широко используется в социальных науках, экономике, медицине и маркетинге.

Примером применения метода RDD может служить программа субсидирования студентов, где студенты с баллом выше определённого порога получают субсидии, а те, чей балл ниже, не получают. В этом случае порог — это то самое значение, которое используется для определения того, будет ли к данному объекту применено воздействие или нет. Цель, которая ставится перед данным исследованием — выявить влияние субсидирования на средний балл. Очевидно, что в этом случае нельзя разбить выборки на тестовую и контрольную, поскольку воздействие применяется по определённому признаку. Для обхода этого ограничения выдвигается теория, что студенты, близкие к пороговому значению, должны быть очень схожи между собой. Если же наблюдаются какие-то статистически значимые разрывы вблизи порогового значения, значит, воздействие действительно имеет эффект.

2.1 Математическая формулировка RDD

Пусть T — переменная, определяющая воздействие (treatment). Пусть Z — переменная, описывающая пороговое значение такое, что если $X_i \geq Z$, то наблюдение i получает воздействие, а если $X_i < Z$, то воздействие не оказывается. Мы будем изучать разрыв в зависимости отклика Y_i (результат) от переменной X_i (переменная порога).

Предположим, что истинная зависимость результата Y_i от переменной X_i и наличия воздействия может быть описана следующей моделью:

$$Y_i = \alpha + \tau \cdot D_i + f(X_i) + \epsilon_i, \quad (1)$$

где:

- Y_i — результат для наблюдения i (например, успеваемость студента);
- D_i — индикатор воздействия, который равен 1, если $X_i \geq Z$ (наблюдение получает воздействие), и 0, если $X_i < Z$;
- $f(X_i)$ — гладкая функция, описывающая зависимость результата от X_i ;
- ϵ_i — ошибка.

Цель метода — оценить эффект воздействия τ .

Существует два основных типа RDD:

- **sharp RDD** — в этом случае порог чётко определяет наличие или отсутствие воздействия. Если $X_i \geq Z$, то воздействие точно оказывается;
- **fuzzy RDD** — в этом случае наличие воздействия зависит от вероятности, и порог лишь увеличивает вероятность получения воздействия, но не определяет его точно.

2.2 Графический анализ в RDD

Одним из ключевых этапов в применении метода RDD является графический анализ. Мы строим график зависимости Y от X , чтобы увидеть разрыв в значении отклика на пороге Z .

Для этого часто используется регрессия, обученная отдельно на различных сторонах от порога, для оценки функции $f(X)$ с обеих сторон от порога. Разрыв в значении отклика указывает на эффект воздействия.

Для оценки эффекта в точке порога Z обычно применяются полиномиальные регрессии, где исследуется изменение наклона функции отклика $Y(X)$ на пороге.

Основная задача — оценить разницу в ожидаемых значениях результата Y на границе порога:

$$\tau = \lim_{X \rightarrow Z^+} E[Y_i | X_i = Z] - \lim_{X \rightarrow Z^-} E[Y_i | X_i = Z].$$

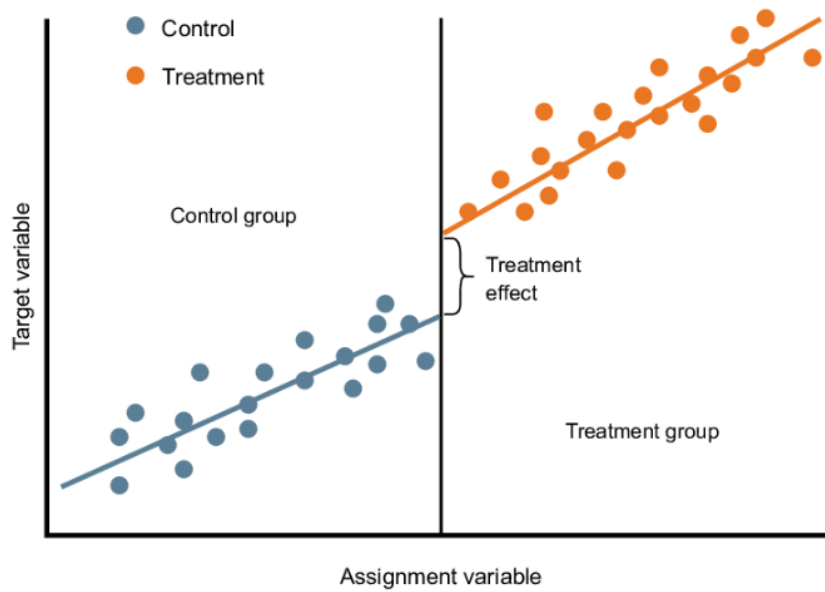


Рис. 1: Пример графического представления RDD

Для проверки значимости эффекта также часто используется тест с применением бутстрапа или других методов для вычисления доверительных интервалов для разрыва в точке порога.

Хотя метод RDD является мощным инструментом, он имеет свои ограничения.

- **Требования к данным:** данные должны быть достаточно плотными около порога, чтобы можно было точно оценить эффект.
- **Нелинейность зависимостей:** если функция $f(X)$ существенно нелинейна, результаты могут быть ошибочными.
- **Манипуляции с порогом:** если участники могут манипулировать своей переменной X для того, чтобы попасть в группу с воздействием, результаты могут быть искажены.

3 Synthetic control method (SCM)

Synthetic control method (SCM) — это метод оценки, который используется для анализа влияния политик или вмешательств на уровне групп, особенно в случае когда нет явной контрольной группы. SCM позволяет создать искусственную контрольную группу (synthetic control) путём взвешивания нескольких наблюдаемых единиц, которые не получили воздействия, но могут служить адекватным сравнением для изучаемой группы.

3.1 Математическая формулировка SCM

SCM создаёт синтетическую контрольную группу как взвешенное среднее из нескольких групп, которые не получили воздействия. Весовые коэффициенты выбираются так, чтобы минимизировать разницу в предшествующих наблюдениях (до воздействия) между синтетической контрольной группой и наблюдаемой группой. После того как синтетическая контрольная группа сформирована, её результаты сравниваются с результатами группы, подвергшейся воздействию.

Рассмотрим описание алгоритма синтетического контроля с использованием матриц ковариат и оценкой среднего воздействия на уровне группы.

Пусть:

- X_1 — это матрица ковариат для тестовой (или целевой) выборки **до начала эксперимента**. Она включает информацию о факторах, которые могут влиять на целевую метрику;
- X_0 — это матрица ковариат для контрольной выборки. Контрольная группа состоит из наблюдений, которые не подверглись воздействию, и используется для построения синтетического контроля;
- $W = (w_1, w_2, \dots, w_J)$ — это вектор весов, который назначается контрольным наблюдением. Эти веса будут использованы для создания синтетической контрольной группы;
- $Y(T, t)$ — значение метрики (например, бизнес-метрики) для тестовой группы в момент времени t . Это значение наблюдается на протяжении эксперимента;
- $Y(C, t)^*$ — соответствующее значение для синтетической контрольной группы, вычисленное как взвешенная сумма значений метрики для контрольных наблюдений.

3.2 Оптимизация весов

Цель метода SCM — найти такие веса W , которые минимизируют расхождение между тестовой выборкой и взвешенной контрольной группой до начала эксперимента. Это можно записать как задачу минимизации следующего выражения:

$$\min_W \|X_1 - X_0 W\|,$$

где $X_0 W$ — это линейная комбинация ковариат контрольной группы с весами W . Вес w_j показывает вклад j -го контрольного наблюдения в синтетическую контрольную группу.

Для решения этой задачи накладываются следующие ограничения на веса:

$$w_j \geq 0 \quad \text{и} \quad \sum_{j=1}^J w_j = 1.$$

Эти ограничения означают, что все веса должны быть неотрицательными (чтобы контрольная группа имела реальный физический смысл) и сумма весов должна быть равна 1 (чтобы синтетическая группа представляла взвешенную сумму наблюдений).

После нахождения оптимальных весов W для контрольных наблюдений синтетическая контрольная группа вычисляется как взвешенная сумма наблюдений контрольной группы:

$$Y(C, t)^* = \sum_{j=1}^J w_j Y_j(t),$$

где $Y_j(t)$ — это значение метрики для j -го контрольного наблюдения в момент времени t , а w_j — это оптимальные веса, найденные в предыдущем шаге.

3.3 Оценка среднего воздействия (average treatment effect, ATE)

Для оценки воздействия на тестовую группу в момент времени t используется разница между значением метрики для тестовой группы и соответствующим значением для синтетической контрольной группы:

$$\Delta_t = Y(T, t) - Y(C, t)^*.$$

Здесь Δ_t — это разница между фактическим результатом для тестовой группы и контрфактическим результатом, который мы бы ожидали, если бы воздействие не было применено.

Если разница Δ_t значительна и стабильна на протяжении всего эксперимента, то можно сделать вывод о наличии эффекта воздействия на тестовую группу. Для оценки статистической значимости эффекта часто применяются различные статистические тесты, например бутстрап, для анализа распределения Δ_t по времени и определения, является ли это различие случайным или действительно связано с воздействием.

Предположим, что вы хотите оценить влияние рекламной кампании на продажи в определённом регионе. В этом случае:

- X_1 — это матрица ковариат, включающая данные по продажам, экономическим показателям и другим факторам для региона, где была проведена рекламная кампания, до её начала;
- X_0 — это аналогичные данные для нескольких других регионов, где реклама не проводилась;
- алгоритм SCM используется для создания синтетической контрольной группы, которая представляет собой взвешенное среднее данных по другим регионам;
- значение $Y(T, t)$ — это фактические продажи в тестовом регионе после начала кампании, а $Y(C, t)^*$ — это предсказанные продажи для синтетической контрольной группы;
- разница Δ_t между фактическими и предсказанными продажами используется для оценки эффекта рекламы.

3.4 Ограничения SCM

Основные ограничения SCM заключаются в том, что метод требует наличия хороших доноров (групп для создания синтетического контроля) и большого объёма данных до воздействия для точной калибровки модели. Метод также не всегда применим, если изучаемый объект слишком уникален.

Метод синтетического контроля является мощным инструментом для оценки воздействия в ситуациях, когда невозможно провести классический рандомизированный эксперимент. SCM позволяет создать синтетическую контрольную группу путём взвешивания нескольких контрольных наблюдений, что даёт возможность оценить контрфактический результат. Если разница между фактическими и синтетическими результатами значительна и стабильна, это указывает на наличие воздействия. Однако SCM требует тщательной настройки и правильного выбора ковариат, чтобы результаты были надёжными и интерпретируемыми.

4 Difference-in-difference (DiD)

4.1 Описание метода

Difference-in-difference (DiD) — это метод, который используется для оценки воздействия политик или событий, когда доступна панельная или кросс-секционная информация. DiD сравнивает изменения в результатах между двумя группами до и после воздействия. Предполагается, что без воздействия разница в результатах между группами оставалась бы постоянной.

4.2 Теоретические основы DiD

Пусть у нас есть две группы: *тестовая группа* (которая подвергается воздействию) и *контрольная группа* (которая не подвергается воздействию). Пусть $Y_{T,t}$ и $Y_{C,t}$ обозначают результаты для тестовой и контрольной групп в момент времени t . В модели DiD мы предполагаем, что разница в результатах между группами до воздействия была постоянной, а после воздействия тестовая группа испытывает изменение в результате воздействия.

Эффект воздействия можно оценить как разницу разниц:

$$\Delta Y = (Y_{T,\text{post}} - Y_{T,\text{pre}}) - (Y_{C,\text{post}} - Y_{C,\text{pre}}),$$

где pre и post обозначают периоды до и после воздействия соответственно.

4.3 Пример применения DiD

Примером применения DiD является исследование влияния увеличения минимальной заработной платы на уровень занятости. Сравниваются изменения в уровне занятости в регионе, где произошло увеличение минимальной зарплаты, с регионом, где таких изменений не было.

4.4 Ограничения метода DiD

Метод DiD предполагает, что тренды в результатах для тестовой и контрольной группы до воздействия должны быть одинаковыми. Если это условие нарушается (например, тестовая группа развивается быстрее или медленнее), результаты могут быть искажены.

5 Заключение

Квазиэкспериментальные методы, такие как regression discontinuity design (RDD), synthetic control method (SCM) и difference-in-difference (DiD), играют важную роль в оценке воздействия политик, программ и событий, когда рандомизация невозможна.

RDD полезен, когда есть чёткий порог для разделения групп. Он позволяет анализировать наблюдения, которые находятся по разные стороны от порога. Этот метод эффективен, когда порог объективен и не поддаётся манипуляциям.

SCM применяется, когда нужно оценить воздействие на уровень группы и нет явной контрольной группы. Этот метод использует комбинацию нескольких групп для создания синтетической контрольной группы, что позволяет точно оценить эффект.

DiD особенно полезен для анализа панельных данных и сравнений между группами до и после воздействия. Этот метод предполагает, что в отсутствие воздействия разница в результатах между группами оставалась бы постоянной.

6 Результаты

Из этого текста ты узнал:

- что делать, когда воздействие оказывается не выборочно;
- что делать, когда воздействие растянуто во времени;
- что делать, когда нет контрольной группы.