

Разрывной дизайн

Введение

В данной работе используются переменные из базы данных kids. В основном мы будем изучать IQ ребенка (iq), а точнее то, что может оказывать эффект на него. Среди предикторов находятся IQ матери на момент рождения ребенка (miq), включение в программу раннего развития (tr), пол ребенка (sex), возраст матери на момент рождения ребенка ($mage$), образование матери (med) и шкала оценки состояния новорожденного (значение шкалы Апгар при рождении - $apgr$). Из всех вышеперечисленных независимых переменных основным эффектом воздействия для нас будет программа раннего развития, попадание в которую зависит от IQ матери: если он был ниже 85, то ребенок был включен в программу (в противном случае не включен).

Для того, чтобы проверить эффект от включения в программу на IQ ребенка, будет использоваться разрывной дизайн. Выбор данного метода обусловлен тем, что наша основная воздействующая переменная (включение в программу раннего развития) полностью зависит от значения IQ матери. А именно – если IQ матери ниже 85, тогда ребенок включается в программу, если выше – то не включается. Исходя из этой предпосылки метод разрывного дизайна является наиболее подходящим. Кроме того, наши данные позволяют нам использовать строгий дизайн. Таким образом экспериментальная группа будет представлять собой наблюдения, которым соответствует более низкий IQ матери (до 85 (не включительно)). На рисунке 1 представлена небольшая схема того, как именно определялось, какая группа является воздействующей, а какая контрольной.



Рисунок 1. Схематичное объяснение распределения наблюдений в экспериментальную и контрольную группы.

Перед тем, как провести оценку локального эффекта, необходимо проверить основные допущения для проведения анализа: (1) о сбалансированности контрольной и экспериментальной групп, (2) о гладкости распределения. И только после определения всех вышеперечисленных допущений мы сможем корректно оценить эффект и проинтерпретировать полученные результаты. По итогу также необходимо проверить допущение об устойчивости к альтернативным спецификациям.

Разведка данных

Перед тем, как приступить к непосредственному анализу данных необходимо немного поближе рассмотреть наши контрольную и экспериментальную группы, чтобы на первоначальном этапе определить, что действительно могут быть какие-то различия и наш анализ релевантен. В базе данных всего 78 наблюдений, из которых 32 находятся в группе воздействия, а 46 в контрольной группе.

На рисунке 2 мы видим, что средний уровень IQ ребенка выше в группе воздействия (слева от вертикального линии). Отсюда может следовать гипотеза о том, что программа

раннего развития (попадание в которую зависит от IQ матери) действительно оказывает эффект на развитие ребенка. Для того, чтобы подтвердить (или опровергнуть) данное предположение можно также посмотреть на тренды в каждой группе по отдельности.

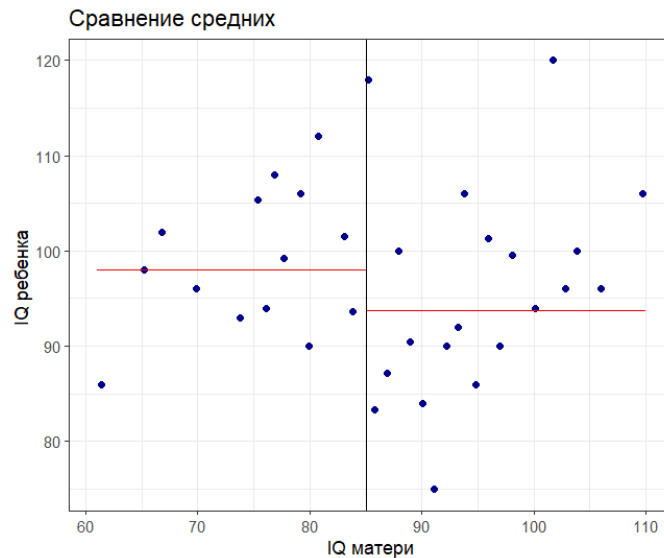


Рисунок 2. Сравнение средних значений в экспериментальной и контрольной группах.

На рисунке 3 мы видим немного иные результаты. В группе воздействия в среднем наблюдается почти горизонтальный тренд, в то время как в контрольной группе он скорее положительный. Однако для нас более интересным наблюдением будет рассмотрение около пороговых значений (эффект на которые мы и оцениваем) и трендов. Мы можем отметить, что тренд в группе воздействия отрицательный с приближением к пороговому значению 85, в то время как в контрольной группе он положительный. Следовательно, мы можем предположить, что включение в программу раннего развития имеет больший эффект не для всех детей, а только для тех, IQ матери которых был достаточно близким к значению 85 (компенсируя тем самым генетический аспект).

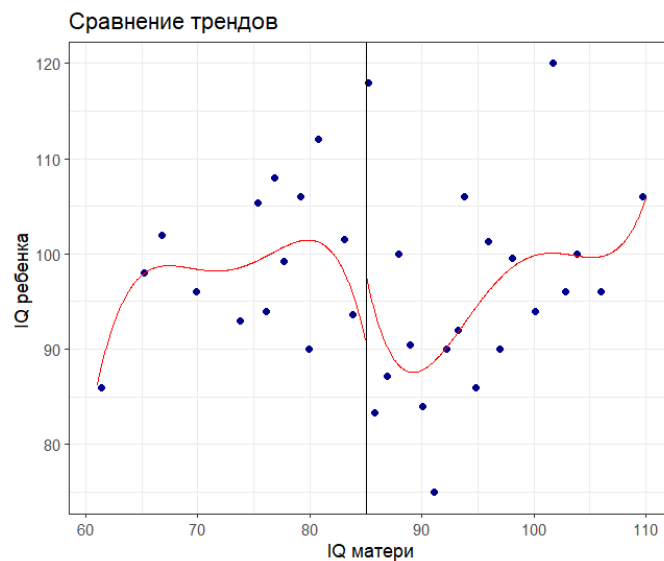


Рисунок 3. Сравнение трендов в экспериментальной и контрольной группах.

Проверка допущения о сбалансированности контрольной и экспериментальной групп

На рисунках ниже предоставлены графические результаты проверки баланса по трем ковариатам (Значения шкалы Апгара; Образование матери; Возраст матери) в экспериментальной и контрольной группах. Стоит отметить, что на рисунке 4 (с двумя ковариатами) наблюдаются относительно большие (хотя бы по сравнению с рисунком 5) дистанции между точками пересечения регрессионных кривых слева и справа от вертикально прямой, что позволяет нам усомниться в наличии баланса по этим двум переменным в экспериментальной и контрольной группах. Однако для того, чтобы в этом убедиться, стоит обратиться к формальным тестам.

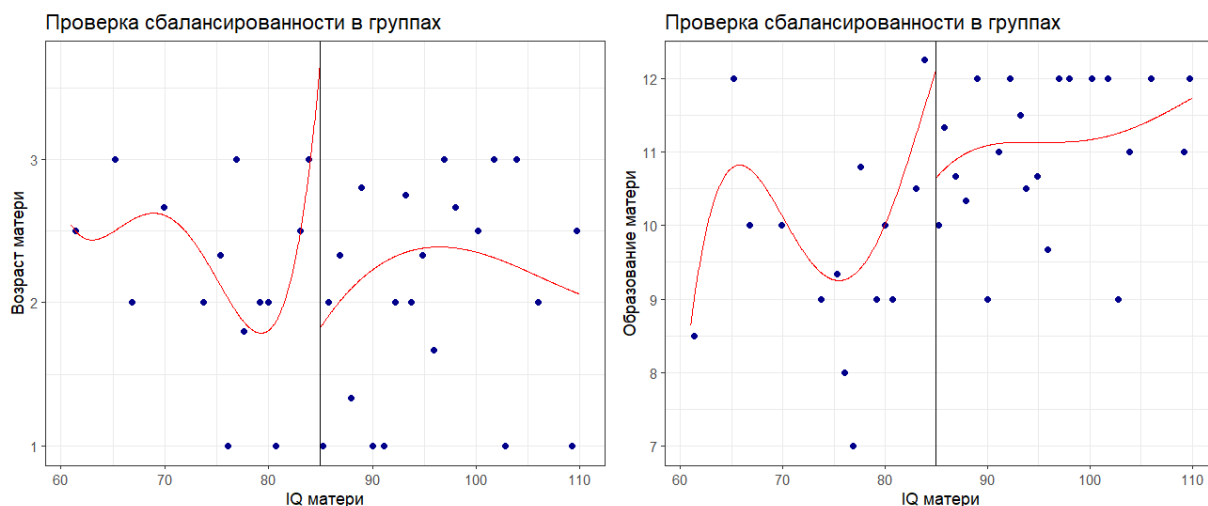


Рисунок 4. Проверка баланса в экспериментальной и контрольной группах по переменным Возраст и Образование матери.

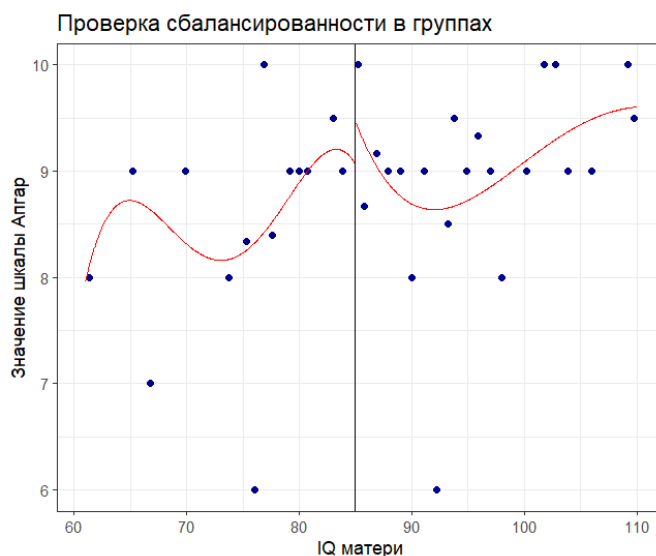


Рисунок 5. Проверка баланса в экспериментальной и контрольной группах по переменной Значение шкалы Апгар.

Для проведения формального теста на проверку сбалансированности в группах использовалась функция `rdrobust`. Результаты по переменной `argr` (значение шкалы Апгар) показывают незначимые различия ($p\text{-value} = 0.702(0.539)$). Это позволяет нам сделать вывод, что распределение наблюдений по данной переменной мы можем считать квазислучайным.

Однако результаты по таким переменным как `med` (образование матери) и `mage` (возраст матери) показывают значимые результаты теста ($p\text{-value} = 0.022$ (0.022), $p\text{-value} =$

0.001 (0.001), соответственно), что говорит об отсутствии сбалансированности по данным переменным в экспериментальной и контрольной группах.

Проверка допущения о гладкости распределения

Суть данного допущения состоит в том, чтобы проверить, что количество наблюдений в непосредственной близости слева от порогового значения должно быть примерно таким же, как и количество наблюдений в непосредственной близости справа от него.

На рисунке 6 мы видим график распределения наблюдений. Стоит отметить, что есть небольшой разрыв в пограничных значениях, что заставляет нас задуматься о соблюдении данного допущения.

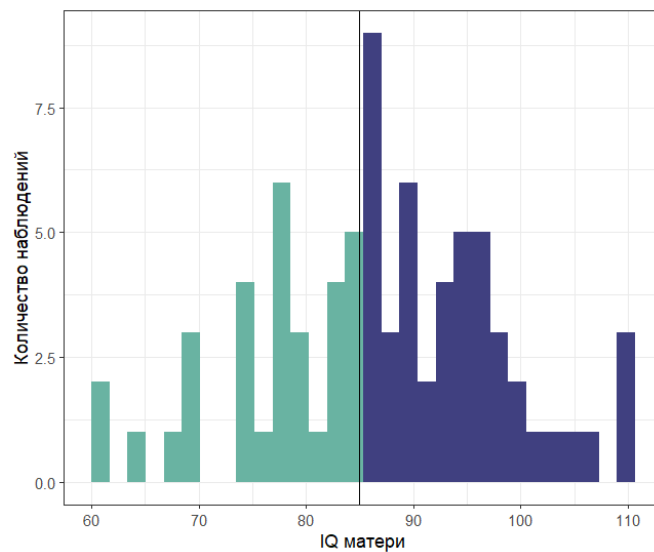


Рисунок 6. Распределение значений IQ матери.

Однако стоит немного более подробно посмотреть на данное распределение с использованием плотности. Рисунок 7 кажется более подробным и показывает то, что нет большого скачка в функции плотности распределения переменной IQ матери в районе её пересечения с вертикальной прямой. Это говорит нам о том, что допущение скорее является соблюденным.

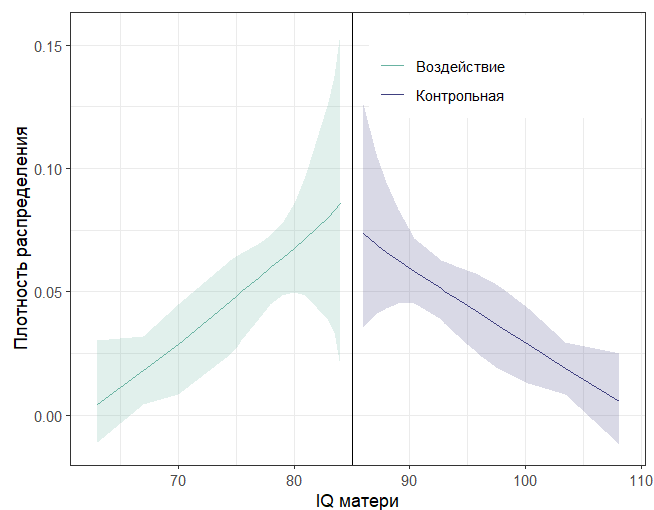


Рисунок 7. Плотность распределения значений IQ матери.

Для проведения формального теста использовалась функция `rddensity`, которая подтверждает этот вывод. Мы видим, что результаты теста статистически не значимы ($p\text{-value} = 0.5361$), следовательно, можем сделать вывод, что распределения наблюдений вокруг порогового значения является квазислучайным.

Оценка эффекта

Для более наглядной иллюстрации результатов и настроек, которые при этом использовались я буду стараться показывать каждую модель графически, сопровождая изображения результатами формальных тестов.

Самым первым вариантом, базовым, является оценка связи между нашей зависимой переменной (IQ ребенка) и переменной IQ матери. Как было уже отмечено ранее, IQ матери может на прямую не иметь эффекта на IQ ребенка, но именно благодаря ему определяется то, попадает ли ребенок в программу раннего развития, которая в свою очередь может иметь эффект на интеллектуальные способности ребенка. Таким образом IQ матери полностью определяет другую переменную – попадание в программу раннего развития. А значит мы можем использовать ее как основную переменную, которая может определять IQ ребенка.

На рисунке 8 мы видим базовую модель взаимосвязи IQ ребенка и IQ матери без включения ковариат (`kernel = 'triangular'`, $p = 1$, `bwselect = 'mserd'`). Результаты формального теста показывают отсутствие статистической значимости ($p\text{-value} = 0.959$ (0.888)), что говорит нам также об отсутствии локального эффекта воздействия. Добавление ковариат в данную модель также не приводит к заметным улучшениям ($p\text{-value} = 0.743$ (0.975)).

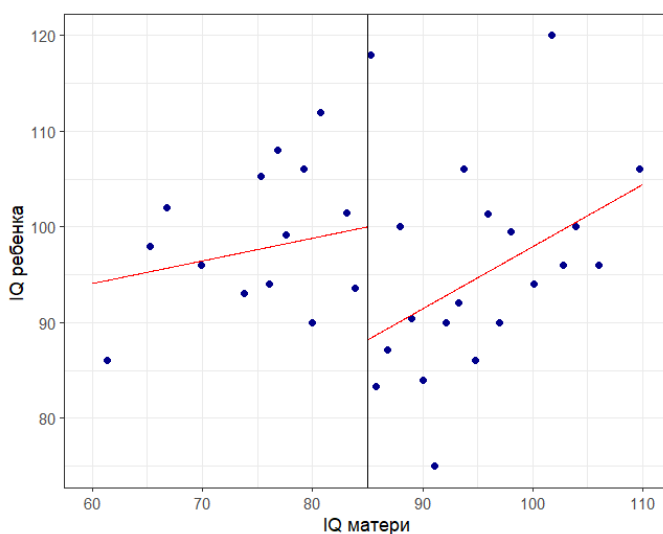


Рисунок 8. Базовая модель оценки эффекта (`kernel = 'triangular'`, $p = 1$, `bwselect = 'mserd'`).

Однако необходимо продемонстрировать устойчивость результатов к выбору метода оценки – для этого попробуем поменять пару настроек, а именно: степень полинома (p), ширину границ области учитываемых значений (h), веса наблюдений (`kernel`). На рисунке 9 представлены для сравнения результаты смены значения степени полинома – значения теста немного улучшились, однако по-прежнему остаются статистически не значимыми ($p\text{-value} = 0.561$ (0.465)).

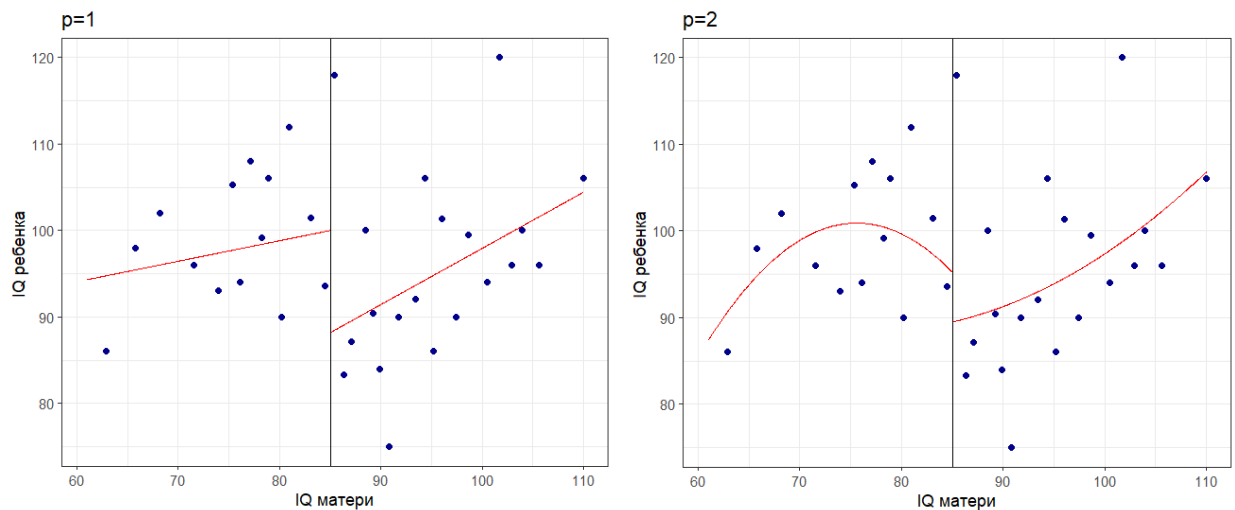


Рисунок 9. Оценка эффекта воздействия при смене значения полинома (p).

Посмотрим на изменения при смене ширины области учитываемых значений (Рисунок 10). Если мы уменьшим ширину, то увидим, что эффект в целом не меняется. Результаты теста продолжают показывать не значимый эффект как для $h=25$ ($p\text{-value} = 0.124(0.640)$), так и для $h=10$ ($p\text{-value} = 0.661(0.671)$).

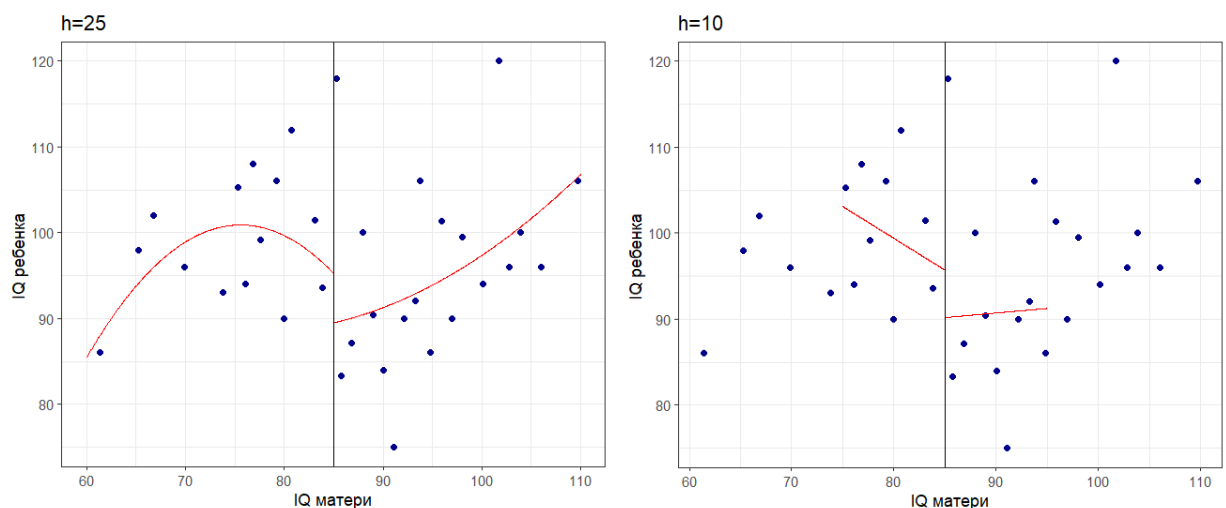


Рисунок 10. Оценка эффекта воздействия при смене ширины области учитываемых значений (h).

Теперь попробуем изменить веса наблюдений. Результаты теста показывают, что смена расчета настройки kernel со значения «triangular» на «uniform» также не приводит к заметным улучшениям ($p\text{-value} = 0.945(0.932)$).

Однако если попробовать изменять несколько настроек сразу, то рано или поздно можно достичь значимого результата $p\text{-value}$. В нашем случае такого можно достичь при $\text{kernel} = \text{'uniform'}$, $p = 1$, $h=25$. Значение $p\text{-value}$ становится значимым и равным 0.013, однако конвенциональная оценка (проверка на робастность) показывает не значимый результат (0.431), что также подтверждает отсутствие эффекта.

Проверка допущения об устойчивости к альтернативным спецификациям

Прежде чем перейти к очевидному выводу, все-таки нужно проверить еще одно важное допущение. Оно заключается в том, что мы должны убедиться, что функция зависимости IQ ребенка от IQ матери является непрерывной в других областях (IQ матери

< 85 или IQ матери > 85). Для этого мы искусственно изменим значение порога с 85 на 91 (ближайшее доступное значение) и проверим, наблюдаются ли скачки функции.

На рисунке 10 мы видим две функции – со значениями до нового порога (IQ матери = 91) и после. Данная иллюстрация показывает, что есть кое-какие скачки до порогового значения, однако после они приобретают более равномерные очертания. У самого порога скачков в целом не наблюдается – уход тренда в экспериментальной группе скорее связан с техническими расчетами (но это не точно). Формальный тест подтверждает наши предположения и свидетельствует об отсутствии сильных скачков в пределах нового порога ($p\text{-value} = 0.620(0.687)$).

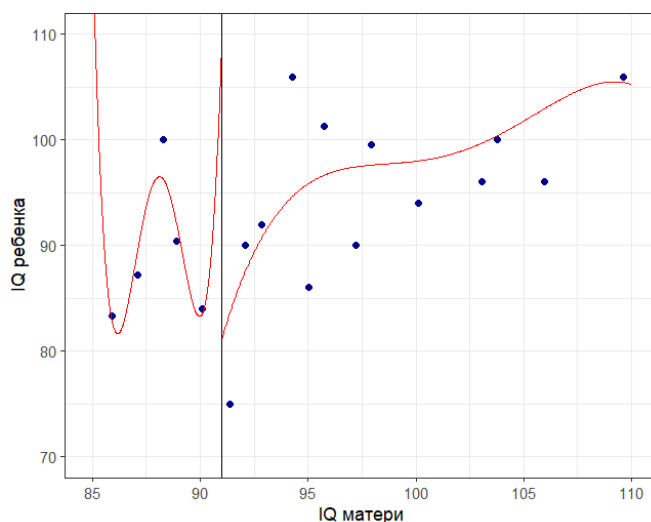


Рисунок 4. Оценка эффекта воздействия при смене порогового значения на 91.

Выводы

Подводя итоги, необходимо отметить, что при проверке сбалансированности в группе воздействия и контрольной группе по некоторым переменным наблюдается дисбаланс, что позволяет засомневаться в распределении наблюдений вокруг порогового значения как квазислучайного.

В дальнейшем при оценке эффекта IQ матери на IQ ребенка было установлено, что он статистически не значим. Следовательно, мы не можем с определенной долей вероятности говорить о наличии локального эффекта.