

Эконометрика: типичные ошибки студентов и аспирантов

Борис Демешев, Кирилл Фурманов*

29 января 2016 г.

Аннотация

В этой короткой статье перечислены ошибки, наиболее часто допускаемые студентами и аспирантами при интерпретации эконометрических моделей, написании работ и презентации своих результатов окружающим.

Ключевые слова: эконометрическая культура, статистическая мудрость, довольная комиссия, распространённые ошибки.

1 Советы Кирилла Фурманова

1. Коэффициенты в регрессии показывают наличие лишь статистической взаимосвязи. Причинно-следственная интерпретация часто ошибочна. Она возможна в некоторых случаях — например, когда данные получены в результате эксперимента.
2. Если нулевая гипотеза не отвергается, это не означает, что она верна. Корректно говорить, что недостаточно данных, чтобы отвергнуть \mathcal{H}_0 , или что данные не противоречат \mathcal{H}_0 . При этом они ещё много чему могут не противоречить.
3. Нужно понимать соответствие между содержательной гипотезой, формулируемой без статистических терминов, и формулировкой в терминах нулевой-альтернативной гипотезы.
4. «Мы говорим — Ленин, подразумеваем — партия, мы говорим — партия, подразумеваем — Ленин». (Маяковский, 1957.) Мы говорим, что проверяем гипотезу о значимости регрессии, хотя на самом деле проверяем гипотезу о незначимости регрессии, то есть $\mathcal{H}_0: \beta_2 = \beta_3 = \dots = \beta_k = 0$.

*НИУ ВШЭ, Москва.

5. Если правильно интерпретировать коэффициенты, то разумными и верными могут оказаться одновременно несколько разных моделей. При этом все оценки коэффициентов будут несмещёнными! Например, и регрессия игрека на икс, и регрессия игрека на икс и зет могут быть осмысленны и интересны. Коэффициент при иксе в первой модели показывает, на сколько в среднем изменяется игрек, когда икс меняется на единицу, а во второй — насколько в среднем изменяется игрек, когда икс меняется на единицу, а зет не изменяется.
6. Значимость — это не то же самое, что существенность. Коэффициент может быть значимым, но совершенно бесполезным. Если месячная зарплата мужчин и женщин значимо отличается, но это отличие составляет два рубля, то можно считать, что его нет. Возможно домножать коэффициент на стандартную ошибку регрессора или на квантили.

2 Советы Бориса Демешева

1. Больше графиков! Работа по эконометрике без картинок скучна и бессмысленна. Если сомневаешься, нужно ли построить ещё один график, значит, нужно. Графиков много разных, не бойся экспериментировать! Например, в R пакет **mvtsplot** позволяет на одном графике осмысленно изобразить 60 временных рядов (см. Peng, 2012).
2. Рассказывай презентацию для идиотов. Методов и моделей слишком много. Во время презентации исходи из предположения, что комиссия не знает эту тему. Выбирай простые примеры, а не рассказывай про общий случай. Меньше буковок на слайдах!
3. «Ларису Ивановну хочу!» — «Хочу построить модель, которая описывает...» — это бяка. Так бессмысленно ставить вопрос! Хочу спасти город Энск от бедности! Это — хорошо! Хочу заработать миллион баксов — чуть менее прикольно, но тоже неплохо. Модель не строят просто так, модель строят для того, чтобы ответить на содержательный вопрос или чтобы использовать для некоторого действия.
4. Зачастую не нужны сложные модели. Всегда проверяй сложную модель против самой простейшей, которая приходит в голову. Например, качество прогнозов во временных рядах стоит проверить против модели «завтра будет так же, как сегодня».
5. После написания курсовой, ВКР, диссертации напиши для себя мораль. Там, конечно, аршинными буквами будет «В СЛЕДУЮЩЕМ ГОДУ Я НАЧНУ ПИСАТЬ ДО НОВОГО ГОДА». Чему тебя научила ВКР? Ещё можно написать протокол воспроизведения всех регрессий и обработки данных. Полезно выложить в публичный доступ.
6. Осваивай открытый софт: R, gretl, L^AT_EX, Markdown, Python, SQL и ещё куча страшных слов! Это бесплатно, и сообщество просто огромное.

Модно, стильно, молодёжно!

7. А ты знаешь, что такое p -value? Учи матстат!
8. Мелочи по представлению результатов, способные вызвать праведный гнев членов комиссии:
 - а) *Техническая подготовка*. Проверь флешку, выложи файл в интернет заранее, чтобы, если флешка не работает, его можно было быстро скачать. Будь готов к мелочам вроде проектора, не отличающего красного от розового, яркого солнца, и полям слайда, вылезающим за доску.
 - б) *Отсутствие номера на слайде*. А покажи-ка мне слайд, ну, этот... на котором... Делай номер внизу слайда.
 - в) *Неподписанные оси на графике*. Принцип идеального графика: идеальный график можно понять, не читая оставшуюся часть работы. Подписывай оси, приводи единицы измерения, расшифровывай названия переменных. Чем больше нужно устных комментариев к графику, чтобы понять его, тем хуже график.
 - г) *Семь знаков после запятой*. По некоторым данным (Frank [и др.], 2008), люди народности пираха (Бразилия) считают так: один, два и много. Пираха знают толк в знаках после запятой!

Список литературы

- Frank M. C. [и др.] Number as a cognitive technology: Evidence from Pirahã language and cognition // Cognition. — 2008. — Т. 108, № 3. — С. 819—824.
- Peng R. D. mvtsplot: Multivariate Time Series Plot. — 2012. — R package version 1.0-1.
- Маяковский В. В. Полное собрание сочинений в тринадцати томах : в 13 т. Т. 6. — М. : Государственное издательство художественной литературы, 1957.