

В помощь изучающим эконометрику

Мини-словарь англоязычных эконометрических терминов, часть 1*

Александр Цыплаков[†]

Новосибирский государственный университет, Новосибирск, Россия

В словаре комментируются отдельные англоязычные эконометрические термины: *dummy*, *sample*, *population*, *score*, *inference*. Акцент делается на уточнении значения терминов с целью избежать возможной путаницы и некорректностей при интерпретации.

Dummy

Буквально слово *dummy* означает нечто ненастоящее, подставное. Для эконометрического термина *dummy* имеется достаточно общепринятый перевод на русский язык — «фиктивная переменная». Это переменная, принимающая значение 1 или 0 в зависимости от того, выполняется или нет определенное условие. Например, пол человека может кодироваться следующим образом: если женский, то 1, если мужской, то 0. О фиктивных переменных см. Wooldridge (2005), Доугерти (2007).

Dummy, dummy variable: фиктивная переменная, переменная-манекен, дамми, псевдопеременная.

Dummy — это самый распространенный термин для названия переменной, принимающей два значения, но есть еще много синонимов для обозначения того же объекта.

Binary variable: бинарная переменная.

Indicator variable: индикаторная переменная.

Dichotomous variable: дихотомическая переменная.

Если бинарная переменная используется в качестве зависимой в модели типа регрессии (такой как пробит или логит), то для ее обозначения могут использоваться другие термины. Основным термин — **binomial variable** — связан с тем, что отдельное наблюдение имеет биномиальное распределение с $n = 1$ (т. е. соответствует одному испытанию Бернулли). О моделях с биномиальной зависимой переменной см. Wooldridge (2005), Доугерти (2007).

Binomial variable: биномиальная переменная.

Qualitative dependent variable: качественная зависимая переменная.

Часто фиктивные переменные используются для моделирования сезонности.

Seasonal dummy: сезонная фиктивная переменная.

Сезонная фиктивная переменная может быть названа **seasonal**, однако термин **seasonal** более широкий.

Seasonal (сущ.): сезонная переменная, сезонная составляющая, сезонное колебание.

*Цитировать как: Цыплаков, Александр (2007) «Мини-словарь англоязычных эконометрических терминов, часть 1», Квантиль, №3, стр. 67–72. Citation: Tsyplakov, Alexander (2007) “A mini-dictionary of English econometric terminology I,” *Quantile*, No.3, pp. 67–72.

[†]Адрес: 630090, г. Новосибирск, Весенний проезд, 6–44. Электронная почта: tsy@academ.org

Sample

В исходном смысле **sample** (по-русски «выборка») — это набор образцов, взятых из некоторой более широкой совокупности объектов (**population**, см. ниже). Задача статистики — по характеристикам данной выборки оценить характеристики той совокупности, из которой она была взята. Однако этот термин практически потерял свой исходный смысл, и под выборкой в статистической теории обычно понимают набор случайных величин, которые независимы в совокупности и имеют одно и то же распределение. В эконометрике термин **sample** во многом потерял и это последнее значение: очень часто выборкой называют набор наблюдений, которые не обязательно являются независимыми и одинаково распределенными.

Sample (сущ.): выборка, выбор, набор выбранных объектов, выборочная совокупность.

Random sample: случайная выборка.

Sample with replacement: выборка с возвращением.

Sample without replacement: выборка без возвращения.

Selective sample: избирательный выбор, избирательная выборка.

Sample selection, sample selectivity: избирательность, селективность при составлении выборки

Truncated sample: усеченная выборка.

Свойства статистик, построенных на основе выборки, зависят от ее размера.

Sample size: размер выборки, объем выборки.

Finite sample: конечная выборка.

Infinite sample: бесконечная выборка.

Small sample: малая выборка, выборка малого размера (объема).

Moderate sample: выборка среднего размера.

Large sample: большая выборка, выборка большого размера (объема).

Infinite sample и **large sample** обычно подразумевают предельный случай, когда размер выборки стремиться к бесконечности. Соответственно, синонимами прилагательного **large-sample** являются термины **asymptotic** и **limiting**.

In large samples: в больших выборках, при больших размерах выборок, асимптотически, в пределе.

Large-sample: относящийся к большой выборке, асимптотический.

Large-sample distribution: распределение в больших выборках.

Asymptotic, asymptotical: асимптотический.

Limiting: предельный.

Терминология, позиционирующая наблюдение (наблюдения) по отношению к рассматриваемой выборке:

Sample period: выборочный период

In-sample, within-sample: внутривыборочный, относящийся к наблюдениям в пределах выборки.

Out-of-sample: вневыборочный, относящийся к наблюдениям за пределами выборки.

Pre-sample: довыборочный, относящийся к периоду, предшествующему выборке.

Subsample: подвыборка, часть выборки.

To sample означает «отбирать объекты, собирать данные для получения выборки». В более широком смысле речь может идти о получении информации на основе выборки.

To sample: производить выборку, получать выборку, составлять выборку.

To draw sample: производить выборку.

Sampling — это процесс, соответствующий глаголу to sample, т. е. процесс получения выборки.

Sampling (сущ.): выбор, получение выборки, составление выборки, взятие образцов, выборочный контроль.

Методы, основанные на изучении свойств выборок, полученных на основе выборки, в статистике называют *resampling*.

To resample: делать повторную выборку, ресэмплировать.

Resampling: повторная выборка, ресэмплирование, ресэмплинг.

Один из численных алгоритмов получения случайных выборок с заданным распределением, особенно популярный в байесовской статистике, называется *Gibbs sampler*. Первоначальные сведения об этом методе можно получить из Casella & George (1992).

Gibbs sampler, Gibbs sampling: гиббсовский сэмплер, гиббсовская схема.

Термины *sample* и *sampling* используются также как прилагательные в значении «выборочный».

Sample (прил.): относящийся к выборке, выборочный.

Sampling (прил.): относящийся к процессу получения выборки, выборочный.

Sample characteristic: выборочная характеристика.

Sample information: выборочная информация; выборочные данные.

Sample moment: выборочный момент.

Sample data: выборочные данные.

Sample value: выборочное значение.

Sampling frequency, sampling rate: частота выборки, шаг выборки.

У термина *sample* в значении «выборочный» имеется синоним — *empirical*.

Empirical, empiric: эмпирический, полученный из опыта, выборочный.

Empirical distribution function: эмпирическая функция распределения.

Population

В исходном значении **population** — это та совокупность, из которой берется выборка. Этот термин, как и термин **sample**, используют в настоящее время в основном в фигуральном смысле, уподобляя теоретическое распределение воображаемой бесконечной совокупности изучаемых объектов. Например, **normal population** означает бесконечную нормально распределенную совокупность, и является синонимом нормального закона распределения. В русскоязычной статистической литературе принято **population** переводить как «совокупность» и добавлять для уточнения слово «генеральная».

Как прилагательное термин **population** фактически является синонимом термина **theoretical** (и даже синонимом термина **true**).

Population (сущ.): (генеральная) совокупность, популяция

Population (прил.): относящийся к (генеральной) совокупности, генеральный, теоретический

Population characteristic: теоретическая характеристика, характеристика (генеральной) совокупности.

Population mean: теоретическое среднее, математическое ожидание.

Population moment: теоретический момент.

Theoretical distribution: теоретическое распределение.

True parameter: истинный параметр, истинное значение параметра.

Есть также не очень распространенный в эконометрике термин **parent**, который указывает на источник выборочных данных, на что-то, что породило данные. Термин может относиться к совокупности, из которой набраны данные, или же вероятностному закону, которому подчиняются данные. Также иногда используется близкое по значению слово **underlying**.

Parent (прил.): порождающий, относящийся к генеральной совокупности, теоретический.

Parent mean: теоретическое среднее, истинное среднее.

Underlying distribution: лежащее в основе [данных] распределение.

Эконометристы привыкли иметь дело с временными рядами и неоднородными наблюдениями. Обычная для теоретической статистики терминология не совсем подходит для описания того, как порождаются эконометрические данные. В связи с этим получил распространение термин **data generating process**, сокращенно **DGP**. В отличие от *модели* **DGP** должен быть абсолютно точно специфицирован. Например, авторегрессия первого порядка — это модель, а описание порождающего данные процесса может быть таким:

Ряд x_t ($t = 1, \dots, 100$) получен по формуле $x_t = 0.7x_{t-1} + \varepsilon_t$, где ε_t — это гауссовский белый шум с нулевым математическим ожиданием и дисперсией 5, причем $x_1 = 0$.

Источник: Davidson & MacKinnon (2003).

Data generating process, data generation process: порождающий данные процесс.

DGP: сокращение от **data generating process**.

Score

Слово *score* довольно многозначно и трудно для перевода. Употребление этого термина в статистике в основном ограничивается значением «некий количественный показатель».

По историческим причинам в рамках метода максимального правдоподобия за термином *score* закрепилось специфическое значение «градиент логарифмической функции правдоподобия». В этом значении предлагается переводить его как «скор» с добавлением другого слова («функция», «вектор»), чтобы термин склонялся более благозвучно. Более полный вариант термина — *efficient score*, но слово *efficient* добавляется достаточно редко.

В имеющейся на русском языке литературе по методу максимального правдоподобия этот объект обычно не называется отдельным термином. В книге Кокс & Хинкли (1978) термин *efficient score* переведен как «эффективный вклад». Такой перевод неудачен, поскольку в методе максимального правдоподобия слово «вклад» (*contribution*) используется в прямом смысле (например, может идти речь о вкладе отдельного наблюдения в скор-вектор).

От термина *score* произошли термины *score test* и *scoring*. *Score test* — это статистический критерий, основанный на скор-функции; его также называют *efficient score test*, *Rao score test* и *Lagrange multiplier test*. *Scoring* — это численный метод максимизации функции правдоподобия, относящийся к классу градиентных методов оптимизации. Подробнее ознакомится с этими понятиями и методом максимального правдоподобия в целом можно, например, в Davidson & MacKinnon (2003).

Score: скор-функция, скор-вектор, количественный показатель, числовая характеристика.

Score vector: скор-вектор.

Efficient score: эффективная скор-функция, то же, что и *score*.

Score test: скор-тест, скор-критерий.

Lagrange multiplier test: тест множителя Лагранжа, критерий множителя Лагранжа.

Scoring, method of scoring: скоринг, метод скоринга.

Термин *score* также употребляется в статистических методах, основанных на рангах (см., например, Хеттманспергер, 1987). При этом рангу наблюдения (т. е. его порядковому номеру при расположении наблюдений в порядке возрастания) сопоставляется некоторое число — *rank score* (или просто *score*). В данном контексте *score* принято переводить «метка».

Rank score: метка.

В прямом значении количественного показателя термин *score* используется в названии одного из методов оценивания моделей с биномиальной зависимой переменной — *MSCORE* (см., например, Greene, 2007).

Maximum score estimator: оценка, основанная на максимуме очков.

MSCORE: сокращение от *maximum score estimator*.

Inference

Statistical inference — это получение информации о теоретическом распределении случайных величин (о генеральной совокупности) по выборочным данным. В самом общем смысле это понятие включает в себя оценивание (включая построение доверительных областей), проверку гипотез и прогнозирование. В узком смысле это построение доверительных областей и проверка гипотез.

Inference: выведение, вывод (как результат и как процесс), (умо-)заключение, инференция.

Statistical inference: статистические выводы, статистический вывод, статистическая инференция.

Inferential: относящийся к выводу, заключению, выведенный/выводимый путем заключения, инференциальный, инференционный.

Estimation: оценивание.

Confidence region: доверительная область.

Hypothesis testing: проверка гипотезы, тестирование гипотезы.

Список литературы

Доугерти, К. (2007). *Введение в эконометрику* (2-е издание). Москва: ИНФРА-М.

Кокс, Д. & Д. Хинкли (1978). *Теоретическая статистика*. Москва: Мир.

Хеттманспергер, Т. (1987). *Статистические выводы, основанные на рангах*. Москва: Финансы и статистика.

Casella, G. & E. I. George (1992). Explaining the Gibbs Sampler. *American Statistician* 46, 167–174.

Davidson, R. & J. G. MacKinnon (2003). *Econometric Theory and Methods*. Oxford University Press.

Greene, W. H. (2007). *Econometric Analysis* (6th edition). Prentice Hall.

Wooldridge, J. (2005). *Introductory Econometrics: A Modern Approach* (3rd edition). South-Western College Publishing.

A mini-dictionary of English econometric terminology I

Alexander Tsyplakov

Novosibirsk State University, Novosibirsk, Russia

This dictionary comments on some English econometric terms: dummy, sample, population, score, inference. Emphasis is placed on accurate definitions of their meaning to avoid possible confusion and incorrect interpretation.