

# В помощь изучающим эконометрику

## Мини-словарь англоязычных эконометрических терминов, часть 2\*

Александр Цыплаков<sup>†</sup>

*Новосибирский государственный университет, Новосибирск, Россия*

В этой части словаря комментируются англоязычные эконометрические термины *kurtosis*, *skewness*, *critical region*, *significance level*, *confidence level* и др. Акцент вновь делается на уточнении значения терминов с целью избежать возможной путаницы и некорректностей при интерпретации.

### **Kurtosis, skewness**

Куртозис (*kurtosis*) является показателем, отражающим остроту вершины и толщину хвостов одномерного распределения. Термин происходит от греческого слова *κυρτός*, что означает «выгнутый», «раздутый». Этот показатель строится на основе четвертого центрального момента распределения ( $\mu_4$ ). Обычно в эконометрике используют следующее определение:

$$\kappa = \frac{\mu_4}{\sigma^4},$$

где  $\sigma^4$  — квадрат дисперсии. У нормального распределения  $\kappa = 3$ , в связи с чем распространение получил модифицированный показатель

$$\kappa - 3 = \frac{\mu_4}{\sigma^4} - 3,$$

называемый *excess kurtosis*. Прилагательное *excess* здесь буквально означает «превышающий норму».

С употреблением и переводом этого слова связан ряд проблем. Во-первых, термин *kurtosis* может использоваться как родовое название для ряда показателей формы распределения, связанных с остротой вершины и толщиной хвостов. В частности, показатель  $\kappa - 3$  часто тоже называют просто *kurtosis*. Во-вторых, в качестве перевода термина *kurtosis* в русскоязычной литературе закрепилось слово «эксцесс». Возможно, это связано с переводом на русский язык в 1948 г. известной книги Гаральда Крамера по статистике (Крамер, 1975). Крамер использовал термин *excess* без добавления *kurtosis* для показателя  $\kappa - 3$ . В англо-русских словарях для *kurtosis* предлагается перевод «эксцесс». Однако ясно, что показатель  $\kappa$  некорректно называть эксцессом, поскольку он не является «излишком» по отношению к 3, как  $\kappa - 3$ . В связи с этим желательнее переводить *kurtosis* на русский язык как «куртозис» или «куртосис», а *excess kurtosis* как «эксцесс». Слово *excess* в значении  $\kappa - 3$  в английском языке практически не употребляется, и во избежание недоразумений лучше использовать *excess kurtosis*.

**Kurtosis:** куртозис, куртосис.

**Excess kurtosis:** эксцесс.

\*Цитировать как: Цыплаков, Александр (2008) «Мини-словарь англоязычных эконометрических терминов, часть 2», Квантиль, №5, стр. 41–48. Citation: Tsyplakov, Alexander (2008) “A mini-dictionary of English econometric terminology II,” *Quantile*, No.5, pp. 41–48.

<sup>†</sup>Адрес: 630090, г. Новосибирск, Весенний проезд, 6–44. Электронная почта: [tsy@academ.org](mailto:tsy@academ.org)

Имеется ряд терминов, которые показывают, насколько велик куртозис распределения, и сравнивают его с куртозисом нормального распределения. Так, *leptokurtic distribution* — это распределение с положительным эксцессом, имеющее острую вершину и толстые хвосты, а *platykurtic distribution* — это распределение с отрицательным эксцессом, имеющее плоскую вершину и тонкие хвосты. Соответственно, *mesokurtic distribution* — это распределение, похожее по куртозису на нормальное распределение.

**Leptokurtic:** лептокуртический, островершинный, имеющий высокий куртозис, имеющий положительный (выше нормального) эксцесс.

**Platykurtic:** платикуртический, плосковершинный, имеющий низкий куртозис, имеющий отрицательный (ниже нормального) эксцесс.

**Mesokurtic:** мезокуртический, имеющий нормальный куртозис, имеющий нулевой эксцесс.

**Kurtotic:** то же, что *leptokurtic*.

**Leptokurtosis:** высокий куртозис, превышающий куртозис нормального распределения, положительный эксцесс.

**Platykurtosis:** низкий куртозис, меньший куртозиса нормального распределения, отрицательный эксцесс.

**Tail:** хвост (распределения).

**Heavy tail, fat tail, thick tail:** тяжелый хвост, утяжеленный хвост, толстый хвост.

**Long tail:** вытянутый хвост, удлиненный хвост.

**Thin tail:** тонкий хвост.

**Long-tailed:** имеющий удлиненный хвост.

**Heavy-tailed:** имеющий утяжеленный хвост.

**Heavy-tailedness:** утяжеленность хвоста.

Несимметричность распределения может характеризоваться терминами *skewness* или *asymmetry*. В качестве формального показателя асимметрии распределения обычно используется

$$\frac{\mu_3}{\sigma^3},$$

где  $\mu_3$  — третий центральный момент, а  $\sigma^3$  — дисперсия в степени  $3/2$ . Этот показатель также называют *skewness*.

**Asymmetry:** асимметрия.

**Skewness:** скошенность, асимметрия.

**Skew (сущ.):** то же, что *skewness*.

**Asymmetric/asymmetrical distribution:** асимметричное распределение.

**Skewed distribution, skew distribution:** скошенное распределение.

При положительной скошенности распределения правый хвост, как правило, толще левого, а вершина смещена влево. Такое распределение часто называют скошенным вправо (исходя из соотношения толщины хвостов, а не из положения вершины). При отрицательной скошенности, соответственно, левый хвост, как правило, толще правого, а вершина смещена вправо. Такое распределение называют скошенным влево.

Right-skewed distribution: скошенное вправо распределение.

Left-skewed distribution: скошенное влево распределение.

### Critical region, significance level

В этой словарной статье мы обсудим некоторые термины, связанные с проверкой гипотез. Стандартная процедура проверки гипотезы о параметрах статистической модели состоит в том, что выбираются нулевая гипотеза и конкурирующая с ней альтернативная гипотеза, а также некоторая функция данных — тестовая статистика (*test statistic*). Для тестовой статистики определяются две области: область принятия (область значений статистики, при попадании в которую принимается нулевая гипотеза) и критическая область (область, при попадании в которую отклоняется нулевая гипотеза и принимается альтернативная).

В английском языке слово «гипотеза» часто опускается, и говорится просто *the null, the alternative*. Следует также помнить, что «нулевая гипотеза» — это именно *null hypothesis*, а не *zero hypothesis* (хотя такое словоупотребление изредка встречается)<sup>1</sup>.

Чаще всего для одномерных статистик критическая область задается при помощи критического значения (*critical value*). Это граничная точка между критической областью и областью принятия нулевой гипотезы. Обычно если значение тестовой статистики превышает критическое значение, то нулевая гипотеза отклоняется и принимается альтернативная гипотеза. Если значение статистики ниже критического значения, то наоборот.

To test a hypothesis: проверить гипотезу.

Statistical test: статистический тест, статистический критерий.

Test statistic: тестовая статистика, статистика критерия.

Null hypothesis, null,  $H_0$ : нулевая гипотеза.

Alternative hypothesis, alternative,  $H_1$ : альтернативная гипотеза.

To accept a hypothesis: принять гипотезу.

To reject a hypothesis: отклонить, отвергнуть гипотезу.

Critical region: критическая область.

Rejection region: область отклонения (неприятя) нулевой гипотезы, критическая область.

Acceptance region: область принятия нулевой гипотезы.

Critical value: критическое значение.

<sup>1</sup>Zero hypothesis можно понять и в смысле ограничения, состоящего в том, что некий коэффициент равен нулю. Лучше, конечно, такие неоднозначные термины не использовать.

С процедурой проверки гипотезы связаны различные события и их вероятности. Результат проверки может быть верным (принята верная нулевая гипотеза или отклонена неверная нулевая гипотеза) или ошибочным. Ошибка первого рода состоит в отклонении верной нулевой гипотезы, а ошибка второго рода — в принятии неверной нулевой гипотезы.

Вероятность ошибки первого рода обычно обозначается буквой  $\alpha$ . (Отсюда происходит еще один термин для ошибки первого рода —  $\alpha$  error.) Эту вероятность принято называть уровнем значимости или размером теста. Термины *significance level* и *size* обычно употребляются как синонимы, хотя есть некоторые отличия. Например, обычно говорят, что «гипотеза проверяется на таком-то уровне значимости»; термин *size* здесь не годится.

Если проверяется сложная нулевая гипотеза (т. е. гипотеза, которая неоднозначно задает распределение), то вероятность ошибки первого рода меняется в зависимости от распределения, и в качестве уровня значимости (размера) берется верхняя граница по рассматриваемому семейству распределений, удовлетворяющих нулевой гипотезе. Поэтому *type I error probability* и *significance level/size* — это не полные синонимы.

Близкий к уровню значимости термин — это «р-значение» *p-value*. *P* в данном случае — это сокращение слова *probability* («вероятность»). *P*-значение — это уровень значимости, который наблюдался бы в том случае, если бы критическое значение было выбрано равным текущему значению статистики, или другими словами, это граничный уровень значимости — если бы использовался такой уровень значимости, то при текущем значении статистики мы были бы на грани отклонения нулевой гипотезы. Отсюда синоним *p*-значения — *marginal significance level*.

Вероятность не совершить ошибку второго рода называется мощностью теста (*power*). Это единица минус вероятность ошибки второго рода. Мощность обычно обозначается буквой  $\beta$ . (Отсюда происходит еще один термин для ошибки второго рода —  $\beta$  error.)

*Type I error, type 1 error, error of the first kind*: ошибка первого рода.

$\alpha$  error:  $\alpha$ -ошибка, ошибка первого рода.

*Type II error, type 2 error, error of the second kind*: ошибка второго рода.

$\beta$  error:  $\beta$ -ошибка, ошибка второго рода.

*Rejection probability/rate*: вероятность отклонения нулевой гипотезы.

*Test significance level, test significance*: уровень значимости теста (критерия), значимость теста.

*Conventional significance level*: обычный (традиционный) уровень значимости.

*To test a hypothesis at the 5% significance level*: проверять гипотезу на уровне значимости 5%.

*Test size, size of a test*: размер теста (критерия).

*P-value, probability value*: *p*-значение.

*Marginal significance level*: граничный уровень значимости, *p*-значение.

*Power of a test*: мощность теста (критерия).

При использовании асимптотических аппроксимаций бывает трудно контролировать размер теста, так что реальная вероятность отклонения нулевой гипотезы (*actual rejection rate*) в конечных выборках может отличаться от номинальной вероятности (*nominal rejection rate*). При этом исследователь будет отклонять нулевую гипотезу слишком часто или слишком редко (*over-reject/under-reject*).

Nominal rejection rate: номинальная вероятность отклонения.

Nominal test size/significance level: номинальный размер/уровень значимости теста.

Actual (true) rejection rate: реальная (фактическая) вероятность отклонения.

Over-reject, overreject: отвергать нулевую гипотезу слишком часто (по сравнению с номинальным размером).

Under-reject, underreject: отвергать нулевую гипотезу слишком редко (по сравнению с номинальным размером).

### Confidence level

Близкая к проверке гипотез тема — интервальное оценивание (в общем случае — построение доверительных областей). Вероятность накрытия для доверительного интервала — это вероятность того, что интервал содержит (накрывает) оцениваемую величину при данном значении неизвестных параметров. Доверительный интервал строится таким образом, чтобы вероятность накрытия была равна некоторому уровню, называемому доверительным уровнем (коэффициентом доверия). Если вероятность накрытия зависит от неизвестных параметров, то доверительный уровень интервала приравнивается к наименьшей вероятности накрытия по всем возможным значениям параметров, так что вероятность накрытия при любом значении параметров заведомо не меньше доверительного уровня. Таким образом, coverage probability и confidence level близки по значению, но это не полные синонимы. Доверительная граница (доверительный предел) — это граница доверительного интервала.

Confidence interval: доверительный интервал.

Confidence region/set: доверительная область/доверительное множество.

Coverage probability: вероятность накрытия.

Confidence level: доверительный уровень, уровень доверительной вероятности.

Confidence coefficient: коэффициент доверия.

Confidence bound/limit: доверительная граница/доверительный предел.

Существуют также статистические термины, которые похожи на термин confidence interval, но имеют другое значение, в связи с чем может возникнуть путаница. В частности, это fiducial interval — один из давно известных видов интервального оценивания, но не очень распространенный из-за того, что применим только в очень специфических случаях.

Другой термин, который может вызвать путаницу, — это credible interval. Он относится к байесовскому интервальному оцениванию. Неискушенные в статистике люди часто неверно интерпретируют обычный классический доверительный интервал как такой фиксированный интервал, в который оцениваемая величина попадает с заданной вероятностью. Такая интерпретация скорее подходит для байесовского апостериорного интервала.

Fiducial interval: фидуциальный интервал.

Fiducial probability: фидуциальная вероятность.

Credible interval: правдоподобный интервал, байесовский апостериорный интервал.

Credible set: правдоподобное множество.

## Ad hoc

*Ad hoc* с латинского буквально переводится «для этого». В научных исследованиях иногда приходится делать предположения или использовать приемы, которые берутся не из какой-то общей теории или общего метода, а придумываются специально для данного случая. Для определения подобных приемов и используется термин *ad hoc*. Он может иметь как нейтральный, так и неодобрительный характер. Образованное от него существительное *ad hockery* не случайно похоже на *trickery* — «жульничество».

*Ad hoc*: специальный, устроенный/придуманный для данной цели, для конкретного случая, специфический, произвольный, необоснованный.

*Ad hoc assumption*: специфическое предположение, произвольное, необоснованное допущение.

*Ad hoc approach*: специальный подход, подход, использующий специфические или произвольные приемы.

*Ad hockery, adhockery, ad hocery*: приемы, правила, решения и т. д. предназначенные специально для данного случая.

## Intercept

*Intercept* буквально означает «пересечение». В эконометрике *intercept* означает точку, в которой линия регрессии пересекает ось  $Y$ . В более широком смысле — это параметр регрессии, равный координате, в которой гиперплоскость регрессии пересекает ось, соответствующую зависимой переменной. Синонимом является *constant* или *constant term*.

*Intercept*: параметр пересечения, константа.

*Constant term*: константа, постоянный член.

## Bayesian, frequentist

Статистические методы делятся на две группы: классические и байесовские. Идея байесовского метода состоит в том, что параметры модели являются случайными величинами, и у исследователя есть некоторые исходные (априорные) представления об их распределении. Эти априорные представления затем корректируются с учетом наблюдаемых данных, результатом чего являются апостериорные представления о распределении параметров. Эти методы называют байесовскими (*Bayesian*) из-за того, что в них часто используется теорема Байеса для условных вероятностей. Классические статистические методы в отличие от байесовских направлены на то, чтобы получать результаты, которые бы не зависели от априорного распределения. Классический подход называют также частотным (*frequentist*). Это связано с тем, что принято различать два взгляда на вероятность — частотный и субъективный (байесовский). (Следует однако понимать, что различие байесовской и классической статистики основано на несколько других принципах, чем различие байесовского и частотного взгляда на вероятность. Байесовская статистика не обязательно основывается на субъективных вероятностях, а классическая статистика не обязательно основывается на частотном определении вероятности.)

*Bayes' theorem*: теорема Байеса.

*Bayes' formula/law/rule/equation*: формула/закон/правило/уравнение Байеса.

Bayesian (прил.): байесовский, бейесовский.

Bayesian (сущ.): байесовец, сторонник байесовского подхода.

Bayesian approach: байесовский подход.

Classical approach: классический подход.

Frequentist: частотник, сторонник частотного подхода.

Frequentist approach: частотный подход.

## Hazard

Одно из ключевых понятий в анализе длительностей (дюраций) — это **hazard rate**. Это показатель вероятности наступления некоторого события при том, что это событие до текущего момента еще не наступило. Для непрерывного распределения длительностей

$$\lambda(d) = \frac{f(d)}{1 - F(d)} = \frac{f(d)}{S(d)},$$

где  $F(d)$  — функция распределения,  $f(d) = F'(d)$  — плотность распределения,  $S(d) = 1 - F(d)$  — так называемая **survival function**. Показатель  $\lambda(d)$  как функцию длительности  $d$  называют **hazard function**. Вместо **hazard function** часто используют просто **hazard**.

Данную терминологию трудно перевести на русский язык. Само слово **hazard** буквально означает «риск». Использование слова **hazard** подразумевает, что изучаемые события имеют неблагоприятный характер (например, при изучении продолжительности жизни это смерть). В теории надежности в качестве синонима используют слово **failure**, которое на русский переводят как «отказ». Имеющиеся в русскоязычной литературе переводы связаны с в основном с теорией надежности или статистикой продолжительности жизни. Например, **hazard rate** переводится как «интенсивность отказов».

Подробнее об этих и связанных с ними понятиях можно прочитать в методологической секции настоящего выпуска журнала «Квантиль». Также с данной темой можно познакомиться в Кокс & Оукс (1978), Wooldridge (2001) и Greene (2007).

Duration: длительность, продолжительность, дюрация.

Survival function, survivor function: функция выживаемости, функция надежности, функция долговечности.

Hazard rate, hazard ratio: интенсивность отказов.

Hazard function: функция риска, функция интенсивности отказов.

Proportional hazard model: модель пропорционального риска, модель пропорциональной интенсивности отказов.

Baseline hazard function: базовая функция риска.

## Proxy

**Proxy variable** или просто **proxy** — это такая переменная, которая сама по себе не представляет интереса, но которая заменяет другую переменную. В экономике часто вместо тех переменных, которые действительно интересны, используют их не совсем совершенные заменители.

Например, природные способности человека невозможно измерить непосредственно, но существуют различные показатели, такие как известный коэффициент IQ. ВВП на душу населения используют как показатель степени экономического развития страны или как показатель качества жизни. Об использовании переменных-заменителей в регрессии см. Wooldridge (2001) или Wooldridge (2008).

Proxy variable, проку: переменная-заменитель, суррогатная переменная, эрзац-переменная, прокси.

## Список литературы

Кокс, Д.Р. & Д. Оукс (1978). *Анализ данных типа времени жизни*. Москва: Финансы и статистика.

Крамер, Г. (1975). *Математические методы статистики*. Москва: Мир.

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

Wooldridge, J. (2001). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Wooldridge, J. (2008). *Introductory Econometrics: A Modern Approach* (4<sup>th</sup> edition). South-Western College Publishing.

# A mini-dictionary of English econometric terminology II

Alexander Tsyplakov

*Novosibirsk State University, Novosibirsk, Russia*

This part of the dictionary comments on English econometric terms kurtosis, skewness, critical region, significance level, confidence level, and some others. Emphasis is again placed on accurate definitions of their meaning to avoid possible confusion and incorrect interpretation.