

Содержание

Введение

1. Модель парной регрессии
2. Методы выбора математической модели в парной регрессии
3. Линейная модель парной регрессии и корреляции
4. Определение тесноты связи и оценка существенности уравнения регрессии
5. Виды нелинейных регрессионных моделей, расчет их параметров
6. Метод максимального правдоподобия

Заключение

Список используемой литературы

Введение

Регрессионный анализ заключается в определении аналитического выражения связи, в котором изменение одной величины обусловлено влиянием одной или несколько независимых величин, а множество всех факторов, также оказывающих влияние на зависимую величину, принимается за постоянные и средние значения. Регрессия может быть однофакторной (парной) и многофакторной (множественной). Уравнение регрессии, или статистическая модель связи социально-экономических явлений выражается в общем случае уравнением:

,

которое адекватно отражает реальное моделируемое явление.

Одной из проблем построения уравнения регрессии является её размерность, т.е. определение числа факторных признаков, включаемых в модель. Их число должно быть оптимальным. Практика выработала критерий, позволяющий установить количество факторных признаков, включаемых в модель. Число факторных признаков (к) должно быть в 5-6 раз меньше объема изучаемой совокупности.

1. Модель парной регрессии

Парная регрессия представляет собой регрессию между двумя переменными - у и х, т. е. модель вида:

,

где у - зависимая переменная (результативный признак); х - независимая, или объясняющая, переменная (признак-фактор). Знак " \wedge " означает, что между переменными х и у нет строгой функциональной зависимости, поэтому практически в каждом отдельном случае величина у складывается из двух слагаемых:

,

где у - фактическое значение результативного признака; - теоретическое значение

результативного признака, найденное исходя из уравнения регрессии; e - случайная величина, характеризующая отклонения реального значения результативного признака от теоретического, найденного по уравнению регрессии.

Случайная величина e называется также возмущением. Она включает влияние не учтенных в модели факторов, случайных ошибок и особенностей измерения. Ее присутствие в модели порождено тремя источниками: спецификацией модели, выборочным характером исходных данных, особенностями измерения переменных. От правильно выбранной спецификации модели зависит величина случайных ошибок: они тем меньше, чем в большей мере теоретические значения результативного признака, подходят к фактическим данным u . К ошибкам спецификации относятся неправильный выбор той или иной математической функции для u и недоучет в уравнении регрессии какого-либо существенного фактора, т. е. использование парной регрессии вместо множественной.

Наряду с ошибками спецификации могут иметь место ошибки выборки, которые имеют место в силу неоднородности данных в исходной статистической совокупности, что, как правило, бывает при изучении экономических процессов. Если совокупность неоднородна, то уравнение регрессии не имеет практического смысла. Для получения хорошего результата обычно исключают из совокупности единицы с аномальными значениями исследуемых признаков. И в этом случае результаты регрессии представляют собой выборочные характеристики.

Использование временной информации также представляет собой выборку из всего множества хронологических дат. Изменив временной интервал, можно получить другие результаты регрессии.

Наибольшую опасность в практическом использовании методов регрессии представляют ошибки измерения. Если ошибки спецификации можно уменьшить, изменяя форму модели (вид математической формулы), а ошибки выборки - увеличивая объем исходных данных, то ошибки измерения практически сводят на нет все усилия по количественной оценке связи между признаками.

Особенно велика роль ошибок измерения при исследовании на макроуровне. Так, в исследованиях спроса и потребления в качестве объясняющей переменной широко используется "доход на душу населения". Вместе с тем, статистическое измерение величины дохода сопряжено с рядом трудностей и не лишено возможных ошибок, например, в результате наличия скрытых доходов.

2. Методы выбора математической модели в парной регрессии

Предполагая, что ошибки измерения сведены к минимуму, основное внимание в эконометрических исследованиях уделяется ошибкам спецификации модели. В парной регрессии выбор вида математической функции может быть осуществлен тремя методами:

- 1) графическим;
- 2) аналитическим, т.е. исходя из теории изучаемой взаимосвязи;
- 3) экспериментальным.

При изучении зависимости между двумя признаками графический метод подбора

вида уравнения регрессии достаточно нагляден. Он основан на поле корреляции. Основные типы кривых, используемые при количественной оценке связей, представлены на рис. 1.1:

Рис 1. Основные типы кривых, используемые при количественной оценке связей между двумя переменными.

Значительный интерес представляет аналитический метод выбора типа уравнения регрессии. Он основан на изучении материальной природы связи исследуемых признаков.

При обработке информации на компьютере выбор вида уравнения регрессии обычно осуществляется экспериментальным методом, т. е. путем сравнения величины остаточной дисперсии, рассчитанной при разных моделях.

Если уравнение регрессии проходит через все точки корреляционного поля, что возможно только при функциональной связи, когда все точки лежат на линии регрессии

,
то фактические значения результативного признака совпадают с теоретическими, т.е. они полностью обусловлены влиянием фактора x . В этом случае остаточная дисперсия.

В практических исследованиях, как правило, имеет место некоторое рассеяние точек относительно линии регрессии. Оно обусловлено влиянием прочих, не учитываемых в уравнении регрессии, факторов. Иными словами, имеют место отклонения фактических данных от теоретических. Величина этих отклонений и лежит в основе расчета остаточной дисперсии:

.
Чем меньше величина остаточной дисперсии, тем меньше влияние не учитываемых в уравнении регрессии факторов и тем лучше уравнение регрессии подходит к исходным данным.

Считается, что число наблюдений должно в 7-8 раз превышать число рассчитываемых параметров при переменной x . Это означает, что искать линейную регрессию, имея менее 7 наблюдений, вообще не имеет смысла. Если вид функции усложняется, то требуется увеличение объема наблюдений, ибо каждый параметр при x должен рассчитываться хотя бы по 7 наблюдениям. Значит, если мы выбираем параболу второй степени

,
то требуется объем информации уже не менее 14 наблюдений.

3. Линейная модель парной регрессии и корреляции

Рассмотрим простейшую модель парной регрессии - линейную регрессию. Линейная регрессия находит широкое применение в эконометрике ввиду четкой экономической интерпретации ее параметров.

Построение уравнения регрессии сводится в первую очередь к расчету его параметров - a и b . Они могут быть определены разными методами. Наиболее распространенным методом, является метод наименьших квадратов (МНК).

Допустим, что заданы n наблюдаемых значений результативного признака (y) и признака-фактора (x).

Следует отметить, что рассчитываются не истинные значения a и b , а только оценки, которые могут быть хорошими или плохими.

Возникает вопрос: существует ли способ достаточно точной оценки a и b алгебраическим путем?

Вначале на поле корреляции построим точки соответствующие наблюдаемым значениям x и y и прямую, выражающую линейную регрессию (рис.2).

регрессионный математический аналитический

Первым шагом является определение остатка для каждого наблюдения. Разность между фактическим и расчетным значением, соответствующим x_i , описывается как остаток в i -м приближении:

Рис.2 Точки рассеивания и прямая, выражающая линейную регрессию

Очевидно, что нужно построить такую линию регрессии, чтобы остатки были минимальными. Необходимо выбрать какой-то критерий подбора, который будет одновременно учитывать величину всех остатков.

Критерий минимизации суммы квадратов отклонений, фактических значений результативного признака (y) от расчетных (теоретических): заложен в основу МНК.

Обозначим через S , тогда

Чтобы найти \min , надо вычислить частные производные по каждому из параметров a и b и приравнять их к нулю.

Преобразуя систему (2.1), получаем следующую систему нормальных уравнений для оценки параметров a и b :

Решая систему (2.2), получим

Параметр b называется коэффициентом регрессии. Его величина показывает, насколько единиц изменится результат с изменением фактора на одну единицу.

Параметр a , вообще говоря, не имеет экономической интерпретации. Например, если $a < 0$, то попытка его экономической интерпретации приводят к абсурду.

Зато можно интерпретировать знак при параметре a . Если, $a > 0$, то относительное изменение результата происходит медленнее, чем изменение фактора.

4. Определение тесноты связи и оценка существенности уравнения регрессии

Уравнение регрессии всегда дополняется показателем тесноты связи. При использовании линейной регрессии в качестве такового показателя выступает линейный коэффициент корреляции r . Одна из формул линейного коэффициента корреляции имеет вид:

Коэффициент корреляции находится в пределах: $-1 < r < 1$. Если $b > 0$, то $0 < r < 1$, и, наоборот, при $b < 0$, $-1 < r < 0$.

Линейный коэффициент корреляции оценивает тесноту связи рассматриваемых признаков в ее линейной форме. Поэтому близость абсолютного значения линейного коэффициента корреляции к нулю еще не означает отсутствие связи между признаками. При нелинейном виде модели связь может оказаться достаточно тесной.

Квадрат линейного коэффициента корреляции называется коэффициентом детерминации. Он характеризует долю дисперсии резульативного показателя y , объясняемую регрессией.

Соответственно величина $1 - r^2$ характеризует долю дисперсии y , вызванную влиянием остальных, неучтенных в модели, факторов.

После того как построено уравнение линейной регрессии, проводится оценка значимости как уравнения в целом, так и отдельных ее параметров.

Оценка значимости уравнения регрессии в целом производится с помощью F-критерия Фишера.

С F-критерием тесно связана характеристика, называемая числом степеней свободы, которая применительно к исследуемой проблеме показывает, сколько независимых отклонений из n -возможных требуется для образования данной суммы квадратов.

Существует равенство между числом степеней свободы общей, факторной и остаточной суммы квадратов.

Число степеней свободы для факторной суммы квадратов равно 1, для общей суммы квадратов равно $(n - 1)$, для остаточной суммы квадратов составляет $(n - 2)$.

Разделив каждую сумму квадратов на соответствующее ей число степеней свободы, получаем дисперсию на одну степень свободы:

Сопоставляя факторную и остаточную дисперсию на одну степень свободы, получим величину F- отношения (F - критерий):

Величина F- критерия связана с коэффициентом детерминации r^2 :

F - критерий для проверки нулевой гипотезы $H_0 : D_{\text{факт}} = D_{\text{ост}}$.

Т.е. если нулевая гипотеза справедлива, то факторная и остаточная дисперсии не отличаются друг от друга. Это дает основание считать, что влияние объясняющей переменной x модели несущественно, а, следовательно, общее качество модели невысоко.

Английским статистиком Снедекором разработаны таблицы критических значений F - отношений при разных уровнях существенности нулевой гипотезы и различном числе степеней свободы. Табличное значение F - критерия - это максимальная величина отношения дисперсии, которая может иметь место при случайном их расхождении для данного уровня вероятности наличия нулевой гипотезы.

Если $F_{\text{факт}} > F_{\text{табл}}$, то нулевая гипотеза H_0 об отсутствии связи признаков отклоняется и делается вывод о существенности этой связи.

Если $F_{\text{факт}} < F_{\text{табл}}$, то H_0 не отклоняется и уравнение регрессии считается статистически незначимым.

В линейной регрессии обычно оценивается значимость не только уравнения в целом, но и отдельных его параметров. Для этого по каждому из параметров определяется его стандартная ошибка: m_b и m_a :

Для оценки существенности коэффициента регрессии его величина сравнивается с его стандартной ошибкой, т. е. определяется фактическое значение t - критерия Стьюдента, которое затем сравнивается с табличным значением при заданном уровне значимости α и числе степеней свободы $(n - 2)$

Имеет место равенство:

Для оценивания существенности параметра a определяется:
и его величина сравнивается с табличным значением.

Если табличное значение t - критерия превышает фактическое, то делается вывод о несущественности данного коэффициента, а если наоборот, табличное значение меньше фактического - вывод о существенности данного коэффициента.

Значимость линейного коэффициента корреляции проверяется на основе величины ошибки коэффициента корреляции:

Т.о. проверка гипотез о значимости коэффициентов регрессии и корреляции равносильна проверке гипотезы о существенности линейного уравнения регрессии. В прогнозных расчетах по уравнению регрессии определяется предсказываемое (y_p) значение как точечный прогноз u_x при $x_p = x_k$ т.е. путем подстановки в уравнение регрессии $u_x = a + bx$ соответствующего значения x . Точечный прогноз явно не реален, поэтому он всегда дополняется расчетом стандартной ошибки u_x , т.е. m_{u_x} , и соответственно интервальной оценкой прогнозного значения:

Стандартная ошибка предсказываемого среднего значения y , при заданном значении x , определяется по формуле:

При прогнозировании на основе уравнения регрессии следует помнить, что величина прогноза зависит не только от стандартной ошибки индивидуального значения y , но и от точности прогноза значения фактора x .

Доверительные интервалы прогноза индивидуальных значений y при фиксированных значениях x с различными вероятностями имеют вид:

где $t_a = 1$ при 68% вероятности

$t_a = 2,0$ при 95% вероятности

$t_a = 2,58$ при 99% вероятности

Для экономических расчетов степень вероятности обычно принимается равной 95%.

5. Виды нелинейных регрессионных моделей, расчет их параметров

Хотя во многих практических случаях моделирование экономических зависимостей линейными уравнениями дает вполне удовлетворительный результат, однако ограничиться рассмотрением лишь линейных регрессионных моделей невозможно. Так близость линейного коэффициента корреляции к нулю еще не значит, что связь между соответствующими экономическими переменными отсутствует. При слабой линейной связи может быть очень тесной, например, не линейная связь. Поэтому необходимо рассмотреть и нелинейные регрессии, построение и анализ которых имеют свою специфику.

В случае, когда между экономическими явлениями существует нелинейные соотношения, то они выражаются с помощью соответствующих нелинейных эконометрических моделей.

Различает две группы нелинейных регрессионных моделей:

- модели, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам;
- модели нелинейные по оцениваемым параметрам.

К первой группе относятся, например, следующие виды функций:

- полином 2-й степени
- полином 3-й степени;
- гипербола.

Ко второй группе относятся:

- степенная;
- показательная;
- экспоненциальная и др. виды функций.

Первая группа нелинейных функций легко может быть линеаризована (приведены к линейному виду). Например, для полинома k -го порядка производя замену:

$$x = x_1, x^2 = x_2, \dots, x^k = x_k$$

Аналогично могут быть линеаризованы и другие виды нелинейных функций 1-й группы, производя соответствующие замены.

Для оценки параметров нелинейных функций первой группы можно использовать, обычный МНК, аналогично, как и в случае линейных функций.

Иначе обстоит дело с группой регрессионных, нелинейных функций по оцениваемым параметрам. Данную группу функций можно разбить на две подгруппы: нелинейные модели внутренне линейные;

нелинейные модели внутренне нелинейные.

Рассмотрим степенную функцию

Она нелинейна относительно параметров a и b . Однако ее можно считать внутренне линейной, так как, прологарифмировав ее можно привести к линейному виду:

Следовательно, ее параметры могут быть найдены обычным МНК.

В исследованиях, часто к нелинейным относят модели, только внутренне нелинейные по оцениваемым параметрам, а все другие модели, которые легко преобразуются в линейный вид, относятся к группе линейных моделей.

Если, модель внутренне нелинейна по параметрам, то для оценки параметров используются итеративные методы, успешность которых зависит от вида функции и особенностей применяемого итеративного подхода.

6. Метод максимального правдоподобия

Метод максимального правдоподобия или метод наибольшего правдоподобия в математической статистике - это метод оценивания неизвестного параметра путём максимизации функции правдоподобия. Основан на предположении о том, что вся информация о статистической выборке содержится в функции правдоподобия.

Метод максимального правдоподобия был проанализирован, рекомендован и значительно популяризирован Р. Фишером между 1912 и 1922 годами (хотя ранее он был использован Гауссом, Лапласом и другими).

Оценка максимального правдоподобия является популярным статистическим методом, который используется для создания статистической модели на основе данных, и обеспечения оценки параметров модели.

Метод максимального правдоподобия соответствует многим известным методам оценки в области статистики. Например, вы заинтересуетесь таким

антропометрическим параметром, как рост жителей России. Предположим, у вас имеются данные о росте некоторого количества людей, а не всего населения. Кроме того предполагается, что рост является нормально распределённой величиной с неизвестной дисперсией и средним значением. Среднее значение и дисперсия роста в выборке являются максимально правдоподобными к среднему значению и дисперсии всего населения.

Для фиксированного набора данных и базовой вероятностной модели, используя метод максимального правдоподобия, мы получим значения параметров модели, которые делают данные "более близкими" к реальным. Оценка максимального правдоподобия даёт уникальный и простой способ определить решения в случае нормального распределения.

Метод оценки максимального правдоподобия применяется для широкого круга статистических моделей, в том числе:

линейные модели и обобщённые линейные модели;

факторный анализ;

моделирование структурных уравнений;

многие ситуации, в рамках проверки гипотезы и доверительного интервала

формирования;

дискретные модели выбора.

Пусть X - независимая выборка из непрерывного равномерного распределения на отрезке $[a, b]$, где θ - неизвестный параметр. Тогда функция правдоподобия имеет вид

Последнее равенство может быть переписано в виде:

где

,

откуда видно, что своего максимума функция правдоподобия достигает в точке

.

Таким образом

.

Такая оценка будет смещенной:

,

откуда

Пусть

-

независимая выборка из нормального распределения с неизвестными средним и дисперсией. Построим оценку максимального правдоподобия для неизвестного вектора параметров θ . Логарифмическая функция правдоподобия принимает вид

.

Чтобы найти её максимум, приравняем к нулю частные производные:

откуда

- выборочное среднее, а

- выборочная дисперсия.

Заключение

Таким образом, по данной теме можно сказать, что парная регрессия характеризует связь между двумя признаками.

Аналитически связь между ними описывается следующими уравнениями:

- прямой $Y(X)=A_0 + A_1 *X$

- параболы $Y(X)=A_0 +A_1 *X+A_2 *X^2$

-гиперболы $Y(X)=A_0 +A_1 + 1/X$

Определить тип уравнения можно в первую очередь графическим способом. Помимо этого существует более общее указание: если результативный и факторный признаки возрастают одинаково, то это свидетельствует о наличии линейной связи между ними. Если результативный признак увеличивается в арифметической прогрессии, а факторный значительно быстрее, то используется параболическая функция.

Также мы узнали, что наряду с методом наименьших квадратов (МНК) возможен и другой подход к оцениванию параметров линейного регрессионного уравнения по данным наблюдений - метод максимального правдоподобия.

Оценка максимального правдоподобия является популярным статистическим методом, который используется для создания статистической модели на основе данных, и обеспечения оценки параметров модели.

Метод оценки максимального правдоподобия применяется для широкого круга статистических моделей, в том числе:

- линейные модели и обобщённые линейные модели;
- факторный анализ;
- моделирование структурных уравнений;
- многие ситуации, в рамках проверки гипотезы и доверительного интервала формирования;
- дискретные модели выбора.

Список используемой литературы

1. Магнус А.Р., Катыхев П.К., Персецкий А.А. Эконометрика. Начальный курс. - М.: "Дело", 2001.
2. Эконометрика. Под ред. И.И. Елисеевой. М.: "Финансы и статистика", 2002.
3. Мардас А.Н. Эконометрика. - С-Пб.: "Питер", 2001
4. <http://pandia.ru/text/78/213/97434.php> - Эконометрика, вход свободный.