

Лекция 2. Парная регрессия и корреляция

Первый вопрос «Понятие, сущность и содержание парной регрессии и корреляции»

Понятие парной регрессии

Парная регрессия – это уравнение, описывающее корреляционную связь между парой переменных: зависимой переменной (результатом) y и независимой переменной (фактором) x .

Цель парной регрессии

Целью парной регрессии является количественное описание взаимосвязей между экономическими переменными.

Задачи парной регрессии

1. установление факта наличия или отсутствия статистических зависимостей между переменными величинами;
2. выявление причинных связей между переменными величинами;
3. прогноз или восстановление неизвестных значений зависимых переменных по заданным значениям независимых переменных.

**Функция парной регрессии может
быть как линейной, так и
нелинейной.**

Линейная функция парной регрессии

Линейная регрессия – это выраженная в виде прямой зависимость среднего значения какой-либо величины от некоторой другой величины. В модели линейной парной регрессии зависимость между переменными в генеральной совокупности представляется в виде:

$$y = \alpha + \beta \cdot x + \varepsilon,$$

где y – зависимая переменная;

x – независимая переменная;

α и β - параметры уравнения (параметр α – свободный член регрессии; параметр β – коэффициент регрессии, который измеряет, на сколько единиц в среднем изменится результативный признак при изменении фактора на одну единицу;

ε – случайная ошибка.

Виды нелинейной функции парной регрессии

Степенная функция нелинейной парной регрессии
имеет вид: $y = ax^b + \varepsilon$

Функция экспоненты нелинейной парной регрессии
имеет вид: $y = ae^{bx} + \varepsilon$

Показательная функция нелинейной парной регрессии
имеет вид: $y = ab^x + \varepsilon$

Функция равносторонней гиперболы нелинейной парной
регрессии имеет вид: $y = a + b/x + \varepsilon$

Обратная функция нелинейной парной регрессии
имеет вид: $y = 1/(a + bx) + \varepsilon$

Понятие случайной величины ε

Случайная величина ε или возмущение, включает влияние неучтенных в модели факторов, случайных ошибок и особенностей измерения.

Ее присутствие в модели обусловлено тремя источниками:

- спецификацией модели;
- выборочным характером исходных данных;
- особенностями измерения переменных.

Понятие случайной величины ε

Наибольшую опасность в практическом использовании методов регрессии представляют ошибки измерения.

Если ошибки спецификации можно уменьшить, изменяя форму модели, а ошибки выборки – увеличивая объем исходных данных, то ошибки измерения практически аннулируют все усилия по количественной оценке связи между признаками.

Парная корреляция

Коэффициент парной корреляции – это статистический показатель, характеризующий степень тесноты связи между двумя переменными.

Его значение может варьироваться от -1 до 1. При значении коэффициента корреляции от 0,3 до 0,5 связь между переменными слабая, при значении от 0,5 до 0,7 связь умеренная, при значении свыше 0,7 связь тесная. При значении, равном нулю, связь отсутствует.

Положительное значение коэффициента свидетельствует о наличии прямой связи, а отрицательное – обратной связи. Прямая связь характеризуется тем, что с увеличением факторного признака произойдет увеличение результативного признака. И, наоборот, при отрицательной связи рост факторного признака приводит к снижению результативного признака.

Первая формула расчета коэффициента парной корреляции

$$r = b \cdot \frac{\sigma_x}{\sigma_y}$$

Вторая формула расчета коэффициента парной корреляции

$$r = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\sigma_x \cdot \sigma_y}$$

Коэффициент детерминации

Квадрат коэффициента корреляции называется коэффициентом детерминации.

Он показывает долю дисперсии, объясняемую регрессией, в общей дисперсии результативного признака.

Коэффициент эластичности

Коэффициент эластичности определяется по формуле:

$$\varepsilon = b \cdot \frac{x}{y}$$

Коэффициент эластичности

Коэффициент эластичности показывает, на сколько процентов от его среднего значения изменится результативный признак при изменении факторного признака на 1% от его среднего значения.

Второй вопрос «Проверка адекватности модели парной регрессии и корреляции»

Ошибка аппроксимации

Для практического использования моделей регрессии большое значение имеет их адекватность, то есть соответствие расчетных статистическим данным.

Корреляционный и регрессионный анализ зачастую проводится для небольшой по объему совокупности.

Поэтому показатели парной регрессии и корреляции могут быть искажены действием случайных факторов, что требует проверки адекватности модели.

Ошибка аппроксимации

Проверка адекватности модели позволяет определить, не являются ли полученные значения параметров результатом действия случайных факторов.

Ошибка аппроксимации

Фактические значения результативного признака отличаются от теоретических, рассчитанных по уравнению регрессии. Величина отклонений фактических и расчетных значений результативного признака по каждому наблюдению представляет собой ошибку аппроксимации.

Ошибка аппроксимации

Ввиду того, что разница между фактическим и теоретическим значением может быть как положительной, так и отрицательной величиной, ошибку аппроксимации для каждого наблюдения принято определять в процентах к модулю.

Формула расчета ошибки аппроксимации

$$\bar{A} = \frac{1}{n} \sum \left| \frac{(y - \hat{y})}{y} \right| 100\%$$

Ошибка аппроксимации

Если значение средней ошибки аппроксимации получилось в пределах 5 – 7%, то это свидетельствует о хорошем подборе вида модели к исходным данным.

Если значение в пределах 8 – 10%, то это говорит о повышенной, но допустимой ошибке аппроксимации.

t-статистика Стьюдента и F-критерий Фишера

Значимость параметров парной регрессии и корреляции определяют с помощью t-критерия Стьюдента или F-критерия Фишера. При этом расчетное значение t-критерия или F-критерия сравнивают с табличным. Превышение расчетного значения над табличным свидетельствует о значимости модели.

t- статистика Стьюдента

Оценка статистической значимости параметров регрессии и коэффициента корреляции проводится с помощью t- статистики Стьюдента.

Для этого выдвигается гипотеза, в которой предполагается, что параметры a и b , а также коэффициент корреляции r_{xy} равны нулю.

Для начала определяются случайные ошибки параметров и коэффициента корреляции. Затем по их значениям рассчитываются фактические значения t- статистики. В заключение фактические значения t- статистики сравниваются с табличными и делаются выводы: нулевая гипотеза либо подтверждается, либо отвергается.

F-критерий Фишера

При проверке статистической значимости коэффициента детерминации используется расчетное и табличное значение F-критерия Фишера. Расчетное значение определяется по формуле:

$$F = (R^2 / (1 - R^2)) \cdot ((n - m - 1) / m),$$

где R^2 – коэффициент детерминации;

n – количество выборок;

m – количество факторов.

Табличное значение F-критерия Фишера распределяется в зависимости степеней свободы.

Третий вопрос «Точечный и интервальный прогноз на основе линейного уравнения регрессии»

Точечный и интервальный прогноз

Под прогнозированием в эконометрике понимают построение оценки зависимой переменной для таких значений независимых переменных, которых нет в исходных наблюдениях. Различают точечное и прогнозирование и интервальное.

Точечный прогноз

При точечных прогнозных расчетах по уравнению регрессии определяется прогнозное значение результативного признака путем подстановки в линейное уравнение регрессии соответствующего (точечного) значения фактора. Такой прогноз обязательно дополняется расчетом стандартной ошибки.

Интервальный прогноз

Интервальные прогнозы заключаются в построении доверительного интервала прогноза и строятся на основе точечных прогнозов.

Доверительным интервалом называется интервал, относительно которого можно с заранее выбранной вероятностью утверждать, что он содержит значение прогнозируемого показателя.

Доверительный интервал имеет верхнюю и нижнюю границу. Ширина интервала зависит от качества модели, числа наблюдений, горизонта прогнозирования, выбранного уровня вероятности и других факторов.

Интервальный прогноз

При построении доверительного интервала прогноза используется стандартная ошибка прогноза и табличное значение t -критерия Стьюдента.

Если построенная модель адекватна, то с выбранной вероятностью можно утверждать, что при сохранении сложившихся закономерностей развития прогнозируемая величина попадает в интервал, образованный верхней и нижней границей.

Благодарю за внимание