ЮЖНО – КАЗАХСТАНСКАЯ ГОСУДАРСТВЕННАЯ ФАРМАЦЕВТИЧЕСКАЯ АКАДЕМИЯ
КАФЕДРА МЕДИЦИНСКОЙ БИОФИЗИКИ, ИНФОРМАТИКИ И МАТЕМАТИКИ

ПРЕЗЕНТАЦИЯ

Множественная регрессия

Содержание

Введение

Основная часть

1. Регрессионный анализ.
2. Виды регрессии.
3. Виды уравнений множественной регрессии.
4. Линейное уравнение множественной регресии: определение коэффициентов.
5. Практический пример построения линейного уравнения множественной регрессии.

Заключение

Список использованной литературы

Введение

Регрессия – величина, выражающая зависимость среднего значения случайной величины y от значений случайной величины *x*.

Впервые термин «регрессия» был введен основателем биометрии Ф. Гальтоном (XIX в.), идеи которого были развиты его последователем К. Пирсоном.

Регрессионный анализ

- метод статистической обработки данных, позволяющий измерить связь между одной или несколькими причинами (факторными признаками) и следствием (результативным признаком).

Признак - это основная отличительная черта, особенность изучаемого явления или процесса.

Целью регрессионного анализа является оценка функциональной зависимости среднего значения результативного признака (y) от факторных (x_1, x_2, …, x_N), выражаемой в виде уравнения регрессии

$$y = f(x_1, x_2, \ldots, x_n).$$

Различают два вида регрессии:

парную

множественную

Парная (простая) регрессия

Парная (простая) регрессия - уравнение вида:

$$y = f(x).$$

Результативный признак при парной регрессии рассматривается как функция от одного аргумента, т.е. одного факторного признака.

Множественная регрессия

Множественная регрессия - уравнение вида:

$$y = f(x_1, x_2, \ldots, x_n).$$

Результативный признак рассматривается как функция от нескольких аргументов, т.е. много факторных признаков.

$$y = f(x_1, x_2, \ldots, x_n)$$

где y – зависимая переменная (результативный признак),
$x_1, x_2, \ldots, x_n$ – независимые, объясняющие переменные (признаки-факторы),
E – возмущение или стохастическая переменная, включающая влияние неучтенных факторов в модели.

По направлению связи регрессия делится на:

прямую регрессию, возникающую при условии, что с увеличением или уменьшением независимой величины «*x*» значения зависимой величины «*y*» также соответственно увеличиваются или уменьшаются;

обратную регрессию, возникающую при условии, что с увеличением или уменьшением независимой величины «*x*» зависимая величина «*y*» соответственно уменьшается или увеличивается

Для характеристики связей используют следующие виды уравнений парной регрессии:

- $y=a+bx$ – *линейное;*
- $y=e^{ax+b}$ – *экспоненциальное;*
- $y=a+b/x$ – *гиперболическое;*
- $y=a+b_1x+b_2x^2$ – *параболическое;*
- $y=ab^x$ – *показательное* и др.

где a, b_1, b_2 - коэффициенты (параметры) уравнения; y - результативный признак; x - факторный признак.

Построение уравнения регрессии сводится к оценке его коэффициентов (параметров), для этого используют метод наименьших квадратов (МНК).

Метод наименьших квадратов позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака «y» от теоретических «y_x» минимальна, то есть

$$\sum (y - y_x)^2 \to \min$$

Параметры уравнения регрессии $y=a+bx$ по методу наименьших квадратов оцениваются с помощью формул:

$$a = \bar{y} - b\bar{x} \qquad b = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2}$$

где a – свободный коэффициент, b - коэффициент регрессии, показывает на сколько изменится результативный признак «y» при изменении факторного признака «x» на единицу измерения.

Для оценки статистической значимости коэффициентов регрессии используется -критерий Стьюдента.

Схема проверки значимости коэффициентов регрессии:

1. H_0: $a=0$, $b=0$ - коэффициенты регрессии незначимо отличаются от нуля

H_1: $a\neq 0$, $b\neq 0$ - коэффициенты регрессии значимо отличаются от нуля

2. $p=0{,}05$ – уровень значимости.

3. $t_{b\ расч} = \dfrac{b}{m_b}$ $\quad t_{a\ расч} = \dfrac{a}{m_a}$

где m_b, m_a - случайные ошибки:

4. $t_{табл}(p; f)$,

где $f=n-k-1$ - число степеней свободы (табличное значение), n - число наблюдений, k - число параметров в уравнении при переменных «x».

5) Если $t_{расч} > t_{табл}$, то H_0 отклоняется, т.е. коэффициент значимый.

Если $t_{расч} < t_{табл}$, то H_0 принимается, т.е. коэффициент незначимый.

Для проверки правильности построенного уравнения регрессии применяется критерий Фишера.

Схема проверки значимости уравнения регрессии:

1) H_0: уравнение регрессии незначимо.
   H_1: уравнение регрессии значимо.
2) p=0,05 – уровень значимости.
3) $$F_{расч} = \frac{\frac{\sum (y_x - \bar{y})^2}{k}}{\frac{\sum (y - y_x)^2}{n-k-1}} = (n-2)\frac{r_{xy}^2}{1-r_{xy}^2}$$

- число наблюдений; k - число параметров в уравнении при переменных «x»; y - фактическое значение результативного признака; y_x - теоретическое значение результативного признака;

4) $F_{табл}(p; f_1; f_2)$,
где $f_1=k$, $f_2=n-k-1$- число степеней свободы (табличные значения).

5) Если $F_{расч} > F_{табл}$, то уравнение регрессии подобрано верно и может применяться на практике.

Если $F_{расч} < F_{табл}$, то уравнение регрессии подобрано неверно.

Основным показателем, отражающим меру качества регрессионного анализа, является коэффициент детерминации (R^2).

$$R^2 = r_{xy}^2$$

Коэффициент детерминации показывает, какая доля зависимой переменной «*y*» учтена в анализе и вызвана влиянием на нее факторов, включенных в анализ.

Коэффициент детерминации (R^2) принимает значения в промежутке [0, 1].

Уравнение регрессии является качественным, если $R^2 \geq 0{,}8$.

Коэффициент детерминации равен квадрату коэффициента корреляции

ЛИНЕЙНОЕ УРАВНЕНИЯ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

$$\widehat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_m x_m.$$

Заключение

Таким образом, регрессионный анализ включает в себя следующие этапы:

- определение типа функции;
- определение коэффициентов регрессии;
- расчет теоретических значений результативного признака;
- проверку статистической значимости коэффициентов регрессии;
- проверку статистической значимости уравнения регрессии.

Список использованной литературы

1. Айвазян С.А. Прикладная статистика. Основы эконометрики: Учебник для ВУЗов в 2-х т. - Т.2. - М.: ЮНИТИ - ДАНА, 2013. - 432 с.
2. Рудакова Р.П., Букин Л.Л., Гаврилов В.И. Статистика. 2-е изд. – СПб.: Питер, 2007 – 288 с.: ил.
3. Статистика: Учеб. пособие / Багат А.В., Конкина М.М., Симчера В.М. и др.; Под ред. В.М. Симчеры. – М.: Финансы и статистика, 2009. – 368 с.: ил.
4. Эконометрика. Учебник для вузов.; Под ред. чл. - кор. РАН И.И. Елисеевой. - М.: Финансы и статистика, 2008. – 344с.

Спасибо за внимание!!!