

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 регрессии: 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.

ЦЕЛЮ РЕГРЕССИОННОГО АНАЛИЗА ЯВЛЯЕТСЯ ОЦЕНКА
ФУНКЦИОНАЛЬНОЙ ЗАВИСИМОСТИ СРЕДНЕГО ЗНАЧЕНИЯ
РЕЗУЛЬТАТИВНОГО ПРИЗНАКА (y) ОТ ФАКТОРНЫХ (x_1 ,
 x_2, \dots, x_n), ВЫРАЖАЕМОЙ В ВИДЕ УРАВНЕНИЯ
РЕГРЕССИИ

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



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:

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

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



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

где y – зависимая переменная (результативный признак),
 x_1, x_2, \dots, 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

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

ПАРАМЕТРЫ УРАВНЕНИЯ РЕГРЕССИИ $y=A+Bx$ ПО
МЕТОДУ НАИМЕНЬШИХ КВАДРАТОВ ОЦЕНИВАЮТСЯ С
ПОМОЩЬЮ ФОРМУЛ:

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

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.
$$t_b \text{ расч} = \frac{b}{m_b} \quad t_a \text{ расч} = \frac{a}{m_a}$$

где m_b, m_a - случайные ошибки:

4. $t_{\text{табл}}(p; f),$

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

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

Если $t_{\text{расч}} < t_{\text{табл}}$, 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.

$$3) F_{расч} = \frac{\sum (y_x - \bar{y})^2}{k} = (n - 2) \frac{r_{xy}^2}{1 - r_{xy}^2} \frac{1}{n - k - 1}$$



- 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 «у» 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 УРАВНЕНИЕ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m.$$

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. Елисейевой. - М.: Финансы и статистика, 2008. – 344с.

Спасибо за внимание!!!

