

ОЦЕНКА СТЕПЕНИ ЗАВИСИМОСТИ УРОВНЯ ЗНАНИЙ, ПОЛУЧЕННЫХ СТУДЕНТАМИ ПО ДИСЦИПЛИНЕ «ИНФОРМАТИКА», ОТ ВЫБРАННЫХ ФАКТОРОВ С ПОМОЩЬЮ РЕГРЕССИОННОГО АНАЛИЗА

Введение. Одной из типовых задач обработки многомерных экспериментальных данных является определение количественной зависимости показателей качества объекта от значений его параметров и характеристик внешней среды. Регрессионный анализ является методом, с помощью которого можно определить степень влияния одних параметров на другие, а в результате – получить предсказание (прогноз) одной переменной на основании другой.

Постановка задачи регрессионного анализа: имеется совокупность результатов наблюдений, сведенная в таблицу. В этой таблице один столбец соответствует показателю, для которого необходимо установить функциональную зависимость с параметрами объекта (факторами), представленными остальными столбцами.

Решение задачи регрессионного анализа состоит на несколько этапов:

- предварительная обработка ЭД;
- выбор вида уравнений регрессии;
- вычисление коэффициентов уравнения регрессии;
- проверка адекватности построенной функции результатам наблюдений [1].

Формулировка цели. Целью работы является определение степени влияния исследуемых параметров на уровень знаний, приобретенных студентами на момент окончания курса обучения, и возможности прогнозирования зачетного балла студентов по значениям исследуемых факторов с помощью парного регрессионного анализа по методу наименьших квадратов, описанному в [2].

Регрессионный анализ. При построении уравнений регрессии считаем, что между зачетным баллом, получаемым студентами с помощью тестирования, и остальными данными, принятыми в качестве исходных параметров для построения модели, существует линейная зависимость, вида $y = b_0 + b_1x$, где y – значение зачетного балла студента, рассчитанное по уравнению регрессии; b_0 – свободный член линейной зависимости, равный значению y при $x = 0$; b_1 – коэффициент, определяющий наклон линии зависимости; x – параметр, степень влияния которого оценивается.

Расчет коэффициентов b_0 и b_1 произведен по формулам (1) и (2) соответственно:

$$b_0 = \frac{\sum_{j=1}^n x_{i,j}^2 \sum_{j=1}^n y_j - \sum_{j=1}^n x_{i,j} \sum_{j=1}^n x_{i,j} y_j}{n \sum_{j=1}^n x_{i,j}^2 - \left(\sum_{j=1}^n x_{i,j} \right)^2}, \quad (1)$$

$$b_1 = \frac{n \sum_{j=1}^n x_{i,j} y_j - \sum_{j=1}^n x_{i,j} \sum_{j=1}^n y_j}{n \sum_{j=1}^n x_{i,j}^2 - \left(\sum_{j=1}^n x_{i,j} \right)^2}, \quad (2)$$

где $x_{i,j}$ – значение i -того параметра в j -том измерении; y_j – значение зачетного балла студента, полученное с помощью тестирования в j -том измерении ($y_j = x_{1,j}$); i – порядковый номер влияющего параметра; m – количество исследуемых параметров ($m = 6$); j – порядковый номер измерения; n – количество измерений ($n = 136$).

После расчета коэффициентов для каждого параметра получили уравнения линейных регрессий.

Уравнение линейной регрессии зачетного балла с уровнем начальных знаний студента:

$$y_1 = 29,4 + 0,5x_2. \quad (3)$$

Уравнение линейной регрессии зачетного балла с уровнем знаний, полученных студентом на занятиях:

$$y_2 \approx -1,96 + 0,96x_3. \quad (4)$$

Уравнение линейной регрессии зачетного балла с признаком пропуска студентом занятий:

$$y_3 \approx 68,74 - 10,4x_4. \quad (5)$$

Уравнение линейной регрессии зачетного балла с уровнем знаний, оставшихся у студента с учетом забывания материала:

$$y_4 \approx 11,96 + 1,74x_5. \quad (6)$$

Уравнение линейной регрессии зачетного балла с уровнем знаний, приобретенных студентом после самостоятельного повторения материала:

$$y_5 \approx 11,96 + 1,29x_6. \quad (7)$$

Уравнение линейной регрессии зачетного балла студента, полученного с помощью тестирования, с зачетным баллом студента, полученным с помощью моделирования:

$$y_6 \approx 3,05 + 0,92x_7. \quad (8)$$

Для проверки адекватности полученных моделей и подтверждения значимости уравнений регрессии применили критерий Фишера. Расчетное значение критерия Фишера вычислено по формуле (9):

До обговорення

$$F_{расч} = \frac{S_{ост}^2}{S_{общ}^2}, \quad (9)$$

где $F_{расч}$ – критерий Фишера; $S_{ост}^2 = \frac{\sum_{i=1}^n \Delta y_i^2}{f}$ – остаточная дисперсия;

$S_{общ}^2 = \frac{\sum_{i=1}^n (y_i - y_{cp})^2}{f}$ – общая дисперсия; $\Delta y_i = y_i - y_i$ – ошибка прогнозирования y_i – измеренное значение зачетного балла студента; y_i – прогнозируемое значение зачетного балла студента; y_{cp} – среднее значение измеренного зачетного балла студента; $n = 136$ – количество проведенных экспериментов; $f = n - 2$ – число степеней свободы для парной регрессии.

Результаты расчетов F -критерия Фишера для проверки адекватности полученных уравнений линейной регрессии приведены в табл. 1:

Таблица 1

Рассчитанные значения F -критерия Фишера

Начальные знания $F_{расч1}$	Средний балл на занятиях $F_{расч2}$	Признак пропуска занятия $F_{расч3}$	Ср. балл после за- бывания $F_{расч4}$	Ср. балл после по- вторения $F_{расч5}$	Результаты моделиро- вания $F_{расч6}$
0,93	0,26	0,06	0,03	0,03	0,04

По таблице критических значений F -критерия Фишера для уровня значимости $\alpha = 0,05$ и количестве проведенных экспериментов $n = 136$ определяем, что критическое значение критерия Фишера $F_{кр} = 1,18$. В результате сравнения рассчитанных значений F -критерия Фишера из табл. 1 с критическим значением критерия $F_{кр} = 1,18$ имеем, что все рассчитанные критерии меньше табличного критического значения. Следовательно, все оцениваемые уравнения адекватны и способны предсказывать зачетный балл студентов с достоверностью 95%.

Для подтверждения значимости проведенных вычислений рассчитаем коэффициенты детерминации по формуле (10):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y_i)^2}{\sum_{i=1}^n (y_i - y_{cp})^2}, \quad (10)$$

До обговорення

где R^2 – коэффициент детерминации; y_i – измеренное значение зачетного балла студента; y_i – прогнозируемое значение зачетного балла студента; y_{cp} – среднее значение измеренного зачетного балла студента; $n = 136$ – количество проведенных экспериментов.

Результаты расчета коэффициентов детерминации сведены в табл. 2.

Таблица 2

Коэффициенты детерминации

Начальные знания R^2_1	Средний балл на занятиях R^2_2	Признак пропуска занятия R^2_3	Ср. балл после за- бывания R^2_4	Ср. балл после по- вторения R^2_5	Результаты моделиро- вания R^2_6
0,0651	0,7386	0,3988	0,7031	0,7031	0,6047

Выводы.

1. По данным табл. 2 можно сделать вывод, что зачетный балл студентов зависит более чем на 70% от знаний, приобретенных студентами на занятиях ($R^2 = 0,7386$), от объема забытого материала изученных ранее тем ($R^2 = 0,7031$) и от эффективности самостоятельной работы студентов при повторении учебного материала ($R^2 = 0,7031$).

2. Начальные знания студентов оказывают незначительное влияние на уровень знаний студентов в момент окончания обучения (зачетный балл), так как только около 6% зачетных знаний студентов зависят от их уровня начальных знаний ($R^2_1 = 0,0651$).

3. Исследуемая модель в целом прогнозирует зачетный балл студентов примерно на 60% ($R^2_6 = 0,6047$).

Литература: 1. Ходасевич Г.Б. Обработка экспериментальных данных на ЭВМ. Часть 2. Обработка многомерных данных / Г.Б. Ходасевич – СПб., СПбГУТ, 2002. – 54 с. 2. Бараз В. Р. Корреляционно-регрессионный анализ связи показателей коммерческой деятельности с использованием программы Excel: учебное пособие / В. Р. Бараз – Екатеринбург, ГОУ ВПО УГТУ-УПИ, 2005. – 102 с.

Bibliography (transliterated): 1. Khodasevich G.B. Obrabotrka eksperimental'nyh dannyh na EVM. Chast' 2. Obrabotrka mnogomernyh / G.B. Khodasevich – SPb., SPbGUT, 2002. – 54 s. 2. Baraz V. R. Korreljacionno-regressionnyj analiz svyazi pokazatelej kommercheskoj dejatel'nosti s ispol'zovanijem programmy Excel: uchebnoje posobije / V. R. Baraz – Jekaterinburg, GOU VPO UGTU-UPI, 2005. – 102 s.

Шевченко В. О.

**ОЦІНКА СТУПЕНЯ ЗАЛЕЖНОСТІ РІВНЯ ЗНАТЬ,
ОДЕРЖАНИХ СТУДЕНТАМИ З ДИСЦИПЛІНИ «ІНФОРМАТИКА»,
ВІД ВИБРАНИХ ЧИННИКІВ ЗА ДОПОМОГОЮ РЕГРЕСІЙНОГО АНАЛІЗУ**

Описане використання регресійного аналізу для визначення ступеня залежності залікового балу студента від вибраних раніше чинників і підтвердження адекватності побудованої моделі придбання знань студентами з дисципліни «Інформатика».

Шевченко В. А.

ОЦЕНКА СТЕПЕНИ ЗАВИСИМОСТИ УРОВНЯ ЗНАНИЙ, ПОЛУЧЕННЫХ
СТУДЕНТАМИ ПО ДИСЦИПЛИНЕ «ИНФОРМАТИКА», ОТ ВЫБРАННЫХ
ФАКТОРОВ С ПОМОЩЬЮ РЕГРЕССИОННОГО АНАЛИЗА

Описано применение регрессионного анализа для определения степени зависимости зачетного бала студента от выбранных ранее факторов и подтверждения адекватности построенной модели приобретения знаний студентами по дисциплине «Информатика».

Shevchenko V.

ASSESSMENT OF KNOWLEDGE LEVEL DEPENDENCE DEGREE OBTAINED BY
STUDENTS ON ACADEMIC DISCIPLINE "COMPUTER SCIENCE" FROM SELECTED
FACTORS USING REGRESSION ANALYSIS

The application of regression analysis for determining the degree of dependence of students credits from previously selected factors and confirmation of the adequacy of created model of learning by students on academic discipline "Computer Science" is described.
