

Лабораторная работа №5

Тема: Метод оптимального выбора предикторов.

Цель работы: Получить навыки отбора объяснительных признаков для построения эконометрической модели.

Методические указания к выполнению работы

Подбор объяснительных признаков выполняется по методу оптимального выбора предикторов. Его сущность заключается в следующем: сначала для исходной совокупности признаков строятся все одноэлементные подмножества, дальше все двухэлементные, трёхэлементные и т.д. Последнее подмножество будет состоять из всех исходных признаков. Дальше для каждого подмножества строится показатель ёмкости информации H_m :

$$H_m = \sum_{i=1}^m h_i,$$
$$h_i = \frac{r_i^2}{1 + \sum_{j \neq i} |r_{ij}|},$$

где r_i – коэффициент корреляции между признаком X_i и объяснительным признаком Y ; r_{ij} – коэффициент корреляции между признаком X_i и другими признаками, входящими в состав анализируемого подмножества признаков; m – количество признаков в рассматриваемом подмножестве.

Наилучшей будет та комбинация признаков, для которой значение H_m будет наибольшим.

Для упрощения расчетов рекомендуется вычислить матрицы взаимных корреляций исходных признаков и вектор корреляций как исходных признаков, так и объяснительного признака.

Задание для индивидуальной отработки

1) Для данных лабораторной работы №4 провести отбор объяснительных признаков. При этом допустить, что первые шесть признаков $X_1 - X_6$ являются объяснительными, а последняя X_7 – результативной. Для этого избирательно рассмотреть:

- а) все одноэлементные подмножества (шесть подмножеств);
 - б) произвольные две двух-, трёх-, четырех- и пятиэлементные подмножества, учитывая разбиение исходного множества признаков на однородные подмножества (результаты лабораторной работы №4);
 - в) шестиэлементное подмножество.
- 2) Для каждого из подмножеств вычислить значение H_m .

3) Среди отобранных подмножеств избрать две наилучших. Построить регрессионную зависимость для каждого отобранного подмножества, и сравнить по сумме квадратов остаточных разниц, которая из них является наилучшей.

4. Сделать выводы.

5. Для отчета представить:

а) подмножества рассматриваемых признаков;

б) значение величин H_m для каждого подмножества;

в) отобранные подмножества;

г) регрессионное уравнение;

д) суммы квадратов остатков.

Рекомендуемая литература

1. Вознесенский В.А. Статистические методы планирования эксперимента в технико-экономических исследованиях / В.А. Вознесенский. – М. : Статистика, 1974. – 192 с.

2. Гмурман В.Е. Теория вероятностей и математическая статистика. Уч. пособие для втузов. – М.: Высш. школа, 2002. – 479 с.

3. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. Уч. пособие для втузов. – М.: Высш. Школа, 2002. – 400 с.

4. Жалдак М.И. Теория вероятностей с элементами информатики. Практикум: Уч. Пособие / М.И. Жалдак, А.Н. Квитко / Под общ. ред. М.И. Ядренко. – К.: Вища шк., 1989. – 263 с.

5. Ивченко Г.И. Математическая статистика: учеб. пособие для втузов / Г.И. Ивченко, Ю.И. Медведев. – М.: Высш. шк., 1984. – 248 с.

6. Методичні рекомендації до виконання лабораторних робіт з навчальної дисципліни "Моделювання систем" для студентів напряму підготовки 0804 "Комп'ютерні науки" всіх форм навчання / укл. В. М. Задачин, І. Г. Конюшенко. – Харків : Вид. ХНЕУ, 2007. – 96 с.

7. Советов Б.Я. Моделирование систем : учеб. для вузов / Б.Я. Советов, С.А. Яковлев. – М. : Высш. шк., 2001. – 343 с.

8. Советов Б.Я. Моделирование систем: практикум : учеб. пособие для студентов вузов. Изд. 2-е, перераб. и доп. / Б. Я. Советов, С. А. Яковлев. – М.: Высшая школа, 2003. – 295 с.