

Проинтерпретировав данные статистические связи на конкретном хозяйстве, можно сделать выводы о наличии у него резервов повышения эффективности экономической деятельности, о задействовании наиболее перспективных рычагов управления, предложить наиболее подходящие решения по выходу из кризисной ситуации.

Литература.

1. Осипенко А. Н. Метод и средства автоматизации моделирования активных систем: Автореф. дис... канд.техн.наук.: ГГУ, Гомель(1997)

МЕТОД И СРЕДСТВА ОДНОМЕРНОГО АНАЛИЗА ДАННЫХ ПРИ ИССЛЕДОВАНИИ СОЦИАЛЬНЫХ И ПРИРОДНЫХ ОБЪЕКТОВ

П.Н Стрибук

(ГГУ, г. Гомель)

С целью обеспечения адекватности одномерного анализа данных о социальной или природной системе в предлагаемом методе реализованы возможности: учета объема (при «ядерной» аппроксимации функции распределения признаков значений и при проверке гипотез), взвешивания конкретных значений (что, например, позволяет корректно обрабатывать данные, полученные из различных источников), обработки данных с пропусками и сжатых выборок, расщепления смеси, подбора преобразования, выделения репрезентативной подвыборки для последующего регрессионного анализа, визуализации эмпирического и теоретического распределений.

Одномерный анализ проходит по следующим этапам:

1. Нахождение эмпирических статистических характеристик, которые будут использоваться при дальнейшей работе с данными в том случае, если не удастся подобрать согласующееся теоретическое распределение и рассчитать по нему теоретические параметры.

2. «Ядерная» аппроксимация функции распределения. Выбор данного вида аппроксимации вызван тем, что на практике очень часто приходится работать с малыми выборками, по которым «ядерная» аппроксимация дает лучшую оценку параметров, чем стандартные методы (гистограмма и полигон частот) [1]. К тому же на «сверхмалых» выборках они не работают. Особенно это актуально при построении регрессионных моделей, так как детальная информация имеется лишь по небольшому числу объектов.

3. Проверка гипотез о нормальности и логнормальности распределения. Обнаружение нормальности или логнормальности упрощает дальнейшую работу с выборкой (легко находятся теоретические параметры, нет необходимости подбирать преобразования для приведения к нормальному виду, при выделении подвыборки признак, согласующийся с нормальным распределением, не приводит к сильному разрежению выборки и, как следствие, удалению информативных объектов).

4. Проверка на бимодальность и расщепление смеси. Бимодальность, как правило, возникает в случае, когда не была обеспечена однородность выборки и информация бралась для объектов из разных классов (например, измерения биохимического состава крови стада коров, в котором наряду со здоровыми коровами имеются больные).

5. Подбор преобразования выборки. Практически все методы многомерного анализа опираются на предположение о нормальности распределения входящих в модель случайных величин. Когда мы имеем несимметричную эмпирическую функцию распределения, имеет смысл преобразовать статистические данные так, чтобы получить вид функции плотности, наиболее приближенный к симметричному с тем, чтобы после проверки гипотезы о согласии подобрать теоретические параметры распределения и использовать данный признак в дальнейшем анализе.

6. Выделение подвыборки. На практике нередко возникает задача планирования пассивного регрессионного эксперимента путем выделения по данным массового обследования из исходного множества объектов такого его подмножества, которое удовлетворяет требованиям корректности регрессионных построений (обеспечение по каждому признаку, по крайней мере, одномодалности и симметричности распределения).

7. Оценка теоретических параметров распределения и формирование отчета.

Имеется возможность проводить автоматическое нормирование данных на основе полученной после работы программы информации о выборке.

После проведения анализа по каждому признаку формируется отчет (таблица) следующего содержания:

– Номер признака – NOMER;

– Объем выборки по данному признаку без учета (с учетом) пропусков в данных – N ;

– Доли подвыборок в смеси в случае его обнаружения – Q_1 и Q_2 ;

– Наилучшее значение параметра преобразования выборки (содержит единицу или нуль при обнаружении нормальности или логнормальности соответственно) – $LAMBDA$;

– Эмпирические статистические характеристики (среднее арифметическое – $SRARIFM$; среднее квадратическое отклонение – $SRKVOTKL$; медиана – E_MED ; среднее геометрическое – E_SRGEO ; среднегеометрическое отклонение – E_SGOTKL ; коэффициент вариации – E_VARIAC ; асимметрия и эксцесс – E_ASIMM и E_EKSC ; минимум и максимум – MIN и MAX ; 0.01, 0.05, 0.10, 0.15, ..., 0.90, 0.95, 0.99-й эмпирические квантили – E_Q01, \dots, E_Q99);

– Теоретические параметры распределения, в случае подтверждения гипотез о нормальности или логнормальности, успешности расщепления смеси или подбора преобразования (математическое ожидание – T_MO ; среднее квадратическое отклонение – T_SKO ; медиана – T_MED ; среднее геометрическое для логнормального распределения – T_SRGEO ; среднегеометрическое отклонение для логнормального распределения – T_SGOTKL ; коэффициент вариации – T_VARIAC ; асимметрия и эксцесс – T_ASIMM и T_EKSC ; 0.01, 0.05, 0.10, 0.15, ..., 0.90, 0.95, 0.99-й теоретические квантили – T_Q01, \dots, T_Q99 ; математические ожидания и среднее квадратическое отклонения подвыборок в смеси – T_MO1, T_SKO1 и T_MO2, T_SKO2).

Также по каждому признаку создаются файлы необходимых графиков.

Разработанное математическое обеспечение одномерного анализа данных, было апробировано в РНИУП «Институт радиологии» при анализе дозиметрических показателей: дозы внутреннего облучения жителей (по СИЧ-измерениям), содержания радионуклидов в молоке, мясе и т. п. Данное математическое обеспечение основано на ряде авторских разработок, а также усовершенствованных автором алгоритмов, приведенных в [2]. Основные результаты одномерного анализа на примере СИЧ-измерений жителей населенных пунктов Брагинского района приведены в табл. 1.

Фрагмент отчета по результатам работы программы одномерного анализа доз внутреннего облучения жителей населенных пунктов.

Наименование населенного пункта	N	Q1	Q2	...	T_Q01	...	T_Q50	...	T_Q99	T_M01	T_SK01	T_M02	T_SK02
Соболи	143	0,7	0,3	...	0,00	...	0,20	...	0,44	0,16	0,07	0,31	0,08
1 день	119	0,7	0,3	...	0,06	...	0,41	...	0,86	0,35	0,13	0,62	0,16
Кирово	214	0,8	0,2	...	0,09	...	0,32	...	0,73	0,29	0,09	0,50	0,16
Асаревичи	386	0,8	0,2	...	0,00	...	0,21	...	0,57	0,19	0,08	0,43	0,10
Галки	111	0,5	0,5	...	0,05	...	0,26	...	0,54	0,21	0,07	0,34	0,10
Верхние Жары	139	0,8	0,2	...	0,08	...	0,25	...	0,50	0,23	0,07	0,43	0,08
Нижние Жары	70	0,1	0,9	...	0,06	...	0,27	...	0,52	0,35	0,13	0,27	0,08
Алексеевка	202	0,9	0,1	...	0,01	...	0,20	...	0,57	0,19	0,08	0,39	0,14
Переносы	93	0,7	0,3	...	0,02	...	0,18	...	0,47	0,16	0,06	0,30	0,12
Городище	101	0,6	0,4	...	0,07	...	0,23	...	0,61	0,19	0,06	0,37	0,12
Заречье	319	0,9	0,1	...	0,02	...	0,20	...	0,56	0,19	0,08	0,46	0,15
Котловица	192	0,7	0,3	...	0,03	...	0,21	...	0,44	0,19	0,07	0,28	0,09
Малейки	652	0,9	0,1	...	0,00	...	0,23	...	0,71	0,21	0,09	0,58	0,10
Селец	435	0,8	0,2	...	0,01	...	0,22	...	0,63	0,20	0,08	0,41	0,14
Старый Мокрец	99	0,3	0,7	...	0,03	...	0,20	...	0,32	0,13	0,05	0,23	0,05
Красная Поляна	14	0,3	0,7	...	0,02	...	0,16	...	0,27	0,10	0,03	0,18	0,04

Литература.

1. Гаскаров Д. В., Шаповалов В. И. Малая выборка. – М.: Статистика, 1978. – 185 с.
2. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: Основы моделирования и первичная обработка данных. – М.: Финансы и статистика, 1983. – 471 с.

ОБРАЗОВАТЕЛЬНЫЙ СЕРВЕР ФАКУЛЬТЕТА НА ПЛАТФОРМЕ СЕРВЕРА ПРИЛОЖЕНИЙ ZORE

А.И. Ступяк, В.С. Скращук, П.В. Картацев, В.С. Никитенко, А.Н. Ваглай
(ГрГУ, г. Гродно)

Характерными особенностями сервера, построенного на платформе сервера приложений Zore являются: