

## **О выборе адекватной линейной модели эксперимента и исключении грубых наблюдений: тезисы доклада**

*А. М. Ревякин, И. В. Бардушкина*

*Национальный исследовательский университет «МИЭТ», Москва, Россия*

*arevyakin@mail.ru*

## **On Choosing an Adequate Linear Model for the Experiment and Dropping of Structurally Stable Observations**

*A. M. Revyakin, I. V. Bardushkina*

*National Research University of Electronic Technology, Moscow, Russia*

*arevyakin@mail.ru*

The authors did consider the task of choosing an adequate model for the experiment data. For the most accurate estimates of the model parameters and optimization of calculations they have proposed to make the columns of the experiment matrix orthogonal. For determination of the model dimension the authors did use confidence intervals for the selected model. The hypothesis about the outlier presence is verified using the criteria of L. N. Bolshev. This type of task is offered to students as a course work and is performed using MATLAB, Excel or specialized packages for statistical data analysis (STATISTICA, STATGRAPHIC, STADIA and SPSS).

**Keywords:** Regression Analysis; adequate model; Ordinary Least Squares; condition number; outlier; orthogonal matrix; orthogonalization process; course work.

Курсовая работа по теории вероятностей и математической статистике выполняется студентами младших курсов бакалавриата и является одной из первых работ в учебной программе. Особенности ее выполнения детально обсуждены нами в статье [1]. Тема курсовой «Выбор адекватной модели эксперимента с помощью методов линейного регрессионного анализа» актуальна как для будущих инженеров, так и для менеджеров.

В компьютерных классах Национального исследовательского университета «МИЭТ» нет пакетов прикладных программ, специализирующихся на статистической обработке данных, и их приобретение не планируется. Поэтому в курсовой работе предлагается использовать лицензионные пакеты MATLAB или Excel, что сокращает время выполнения работы, но часто приводит к неверным результатам. Учитывая, что при изучении других предметов

и в дальнейшей практической деятельности большинству студентов статистическая обработка данных будет необходима, мы рекомендуем освоить работу с другими пакетами, например, с STATISTICA или STATGRAPHIC.

Самое трудное в выполнении курсовой работы — постановка задачи [2; 3]. Сформулировать задачу требуется либо для реальных показателей (данные из электронных источников), либо для псевдореальных (данные смоделированы студентом или выбраны из предложенных массивов). Выбранные данные должны удовлетворять требованиям проведения регрессионного анализа.

Часто при проведении различных экспериментов возникает необходимость установить функциональную зависимость результатов наблюдения от условий эксперимента. Результаты наблюдения могут быть сопряжены с ошибками измерений, шумами и другими случайными отклонениями. Статистическое оценивание характера зависимости результатов наблюдения от условий эксперимента называется регрессионным анализом [4; 5]. Цель курсовой работы — изучить на конкретном примере методы статистического анализа функциональной зависимости для заданного класса функций  $F(X)$ .

Будем предполагать, что  $F(X)$  есть либо тригонометрический многочлен (модель Т)

$$F(X) = A_0 + \sum_{k=1}^r A_k \cos \pi k X + B_k \sin \pi k X, \\ r \leq 3, X \in [-1; 1],$$

либо алгебраический многочлен (модель А)

$$F(X) = \sum_{k=0}^r C_k X^k, \\ r \leq 6, X \in [-1; 1].$$

Массив исходных данных, подлежащих статистической обработке, содержит результаты измерений переменной в тридцати двух точках заданного отрезка, причем в каждой точке проведено два измерения.

Предположим, что ошибки измерений независимы и все, за исключением, быть может, одной, имеют нормальное распределение с нулевым математическим ожиданием и неизвестной дисперсией. Распределение одной из случайных ошибок может быть отклонено от нормального и сильно отличаться, например, вследствие какого-либо сбоя в процессе измерений. Соответствующее значение называется выбросом.

Задание курсовой работы — выбрать подходящую модель (Т или А) и оценить ее неизвестные параметры, пользуясь статистическими методами.

#### *Этапы курсовой работы.*

1. Для каждой из двух возможных моделей с помощью метода наименьших квадратов (см. [4; 6]) оценить коэффициенты регрессии для модели Т и для модели А. При этом устранить выброс, если он будет обнаружен.

2. Проверить адекватность каждой модели и неадекватную модель исключить из дальнейшего рассмотрения.

3. Для адекватной модели:

- а) определить размерность модели;
- б) найти оценки наименьших квадратов для коэффициентов новой модели, полученной в результате понижения размерности, и построить для них доверительные интервалы;
- в) найти оценку и построить доверительный интервал для дисперсии ошибки;
- г) найти оценку и построить доверительный интервал для значения функции в точке выброса.

Если согласно используемым статистическим критериям обе модели адекватны или, наоборот, неадекватны, то для выбора лучшей из них можно применить некоторые эвристические соображения, связанные, например, с хаотичностью поведения остатков, возможностью понижения размерности, относительной малостью значения  $F$ -статистики.

Рассматриваемые в курсовой работе модели А и Т являются линейными. Поэтому оценка коэффициентов регрессии сводится к решению системы нормальных уравнений [6]. Большая погрешность решения возникает даже в случае неплохой обусловленности матрицы, поэтому мы рекомендуем использовать  $QR$ -разложение [6]. Этот метод более эффективен при обработке результатов эксперимента. (см. пример в [7]).

В постановке задачи курсовой работы предполагается наличие выброса. Поэтому прежде чем оценивать параметры регрессионной модели и дисперсию ошибок, необходимо исключить выброс, иначе оценки будут сильно смещенными. Для проверки гипотезы о наличии выброса и для его поиска рекомендуем использовать критерий Большева [8]. Результаты проверки желательно сравнить с аналогичными результатами, полученными с помощью процедур, встроенных в статистические пакеты.

### Литература

1. **Ревякин А. М., Бардушкина И. В.** Об особенностях выполнения курсовой работы по статистике с применением электронного компонента // Экономические и социально-гуманитарные исследования. 2017. № 1 (13). С. 112—122.
2. **Бардушкина И. В., Ревякин А. М.** К вопросу о повышении качества самостоятельной работы студентов по высшей математике // Экономические и социально-гуманитарные исследования. 2015. № 4 (8). С. 57—62.
3. Задания для выполнения лабораторных и индивидуальных работ по курсу «Теория вероятностей и математическая статистика»

с использованием пакета MATLAB / В. В. Бардушкин, И. В. Бардушкина, В. В. Лесин, А. М. Ревякин // Проектирование инженерных и научных приложений в среде MATLAB: Матлы V Междунар. науч. конф. (г. Харьков, 11—13 мая 2011 г.) / Сост. В. В. Замаруев. Харьков: БЭТ, 2011. С. 471—533.

4. Теория и практика статистических исследований / Под ред. А. М. Ревякина, В. В. Костылева. М.: МГАДА, 2007. 354 с.

5. **Драйвер Н., Смит Г.** Прикладной регрессионный анализ: в 2 кн. М.: Финансы и статистика, 1986.

6. **Ревякин А. М.** Высшая алгебра. М.: МИЭТ, 2007. 504 с.

7. Особенности вычисления тензора по дифрактометрическим измерениям продольной деформации / Д. М. Гулидов, М. О. Варгулевич, К. А. Тузовский, А. М. Ревякин, Б. Малюков // Заводская лаборатория. 1989. № 2. С. 57—59.

8. **Большев Л. Н.** Об исключении грубых наблюдений // Теория вероятностей и ее применение. М.: Наука, 1961. Т. 6. Вып. 4. С. 482—484.

**Ревякин Александр Михайлович** — кандидат физико-математических наук, доцент кафедры высшей математики № 2 Национального исследовательского университета «МИЭТ» (Россия, 124498, Москва, г. Зеленоград, пл. Шокина, д. 1), [arevyakin@mail.ru](mailto:arevyakin@mail.ru)

**Бардушкина Ирина Вячеславовна** — кандидат физико-математических наук, доцент кафедры высшей математики № 2 Национального исследовательского университета «МИЭТ» (Россия, 124498, Москва, г. Зеленоград, пл. Шокина, д. 1), [i\\_v\\_bars@mail.ru](mailto:i_v_bars@mail.ru)

### References

1. Revyakin A. M., Bardushkina I. V. Ob osobennostyakh vypolneniya kursovoi raboty po statistike s primeneniem elektronnoho komponenta (On the Specifics of Statistics Coursework Realization with the Use of Electronic Components for the Students' Independent Work), *Ekonomicheskie i sotsial'no-gumanitarnye issledovaniya*, 2017, No. 1 (13), pp. 112—122.
2. Bardushkina I. V., Revyakin A. M. K voprosu o povyshenii kachestva samostoyatel'noi raboty studentov po vysshei matematike (Towards Improving Quality of Students' Independent Work on the Higher Mathematics), *Ekonomicheskie i sotsial'no-gumanitarnye issledovaniya*, 2015, No. 4 (8), pp. 57—62.
3. Zadaniya dlya vypolneniya laboratornykh i individual'nykh rabot po kursu "Teoriya veroyatnostei i matematicheskaya statistika" s ispol'zovaniem

paketa MATLAB (Tasks for Laboratory and Independent Work in “Theory of Probability and Mathematical Statistics” Course Using MATLAB Software), by V. V. Bardushkin, I. V. Bardushkina, V. V. Lesin, A. M. Revyakin, *Proektirovanie inzhernykh i nauchnykh prilozhenii v srede MATLAB: Mat-ly V Mezhdunar. nauch. konf. (g. Khar’kov, 11–13 maya 2011 g.)*, Sost. V. V. Zamaruev, Khar’kov, BET, 2011, pp. 471–533.

4. Teoriya i praktika statisticheskikh issledovaniy (Theory and Practice of Statistical Investigations), Pod red. A. M. Revyakina, V. V. Kostyleva, M., MGADA, 2007, 354 p.

5. Draiper N., Smit G. Prikladnoi regressionnyi analiz, v 2 kn. (Applied Regression Analysis, in 2 Books), M., Finansy i statistika, 1986.

6. Revyakin A. M. Vysshaya algebra (Higher Algebra), M., MIET, 2007, 504 p.

7. Osobennosti vychisleniya tenzora po diffraktoметрическим измерениям продольной деформации (Peculiarities of Tensor Calculation by Diffractometric Measurement of Longitudinal Strains),

D. M. Gulidov, M. O. Vargulevich, K. A. Tuzovskii, A. M. Revyakin, B. Malyukov, *Zavodskaya laboratoriya*, 1989, No. 2, pp. 57–59.

8. Bol’shev L. N. Ob isklyuchenii grubyykh nablyudenii (On Dropping of Structurally Stable Observations), *Teoriya veroyatnostei i ee primeneniye*, M., Nauka, 1961, T. 6, Vyp. 4, pp. 482–484.

**Revyakin Alexander M.**, Ph.D. of physical and mathematical sciences, associate professor, associate professor of Higher Mathematics Department No. 2, National Research University of Electronic Technology (Shokin Square, 1, 124498, Moscow, Zelenograd, Russia), *arevyakin@mail.ru*

**Bardushkina Irina V.**, candidate of physical and mathematical sciences, associate professor, associate professor of Higher Mathematics Department No. 2, National Research University of Electronic Technology (Shokin Square, 1, 124498, Moscow, Zelenograd, Russia), *i\_v\_bars@mail.ru*