

## Применение однородной вложенной кусочно-линейной регрессии с группировкой переменных для моделирования штатной численности подразделений по защите информации

*С.И. Носков, А.П. Медведев*

*Иркутский государственный университет путей сообщения*

**Аннотация:** Математическое моделирование сложных систем часто требует применения методов группировки переменных для построения эффективных моделей. В данной статье рассматривается задача построения однородной вложенной кусочно-линейной регрессии с группировкой переменных для моделирования штатной численности подразделений по защите информации. По пространственным данным за 2022 год построена соответствующая модель для Социального фонда России. В качестве независимых переменных используются данные о количестве сотрудников организации, электронных подписей, защищенных узлов, защищаемых ресурсов, общего числа структурных подразделений, отдельных зданий и специалистов службы ИТ.

**Ключевые слова:** защита информации, регрессионная модель, однородная вложенная кусочно-линейная регрессия, оценивание параметров, метод наименьших модулей, задача линейно-булева программирования, индексное множество, мощность множества, социальный фонд, критерии адекватности, прогнозирование, штатная численность подразделений по защите информации.

### Введение

Проблема обеспечения кадрами играет важнейшую роль для любой организации. Особую значимость этот вопрос приобретает при определении численности подразделений, занимающихся защитой информации. Несмотря на наличие некоторых нормативно-правовых документов, предлагающих рекомендации по формированию таких подразделений, большинство из них не содержат четких указаний по расчету трудовых затрат в зависимости от функциональных обязанностей сотрудников.

Применение математических моделей, в том числе предполагающих использование методов группирования задействованных факторов, для планирования человеческих ресурсов уже неоднократно обсуждалось в научной литературе. Так, в работе [1] рассматривается метод "fused lasso", который позволяет группировать переменные на основе их сходства, что

может быть полезно для задач регрессии и классификации. Авторы статьи [2] предлагают метод "group lasso", который не только группирует переменные, но и выполняет отбор признаков, что особенно полезно, когда переменные естественным образом объединены в группы. В статье [3] представлен подход к группировке переменных в регрессионных моделях, что позволяет учитывать структуру данных и повышать интерпретируемость моделей. В работе [4] авторы развивают идею "group lasso", предлагая метод "sparse-group lasso", который учитывает, как групповую, так и индивидуальную разреженность переменных. В [5] предоставлен обзор современных методов работы с высокоразмерными данными, включая методы группировки переменных и регуляризации. В [6] показано применение метода "group lasso" для логистической регрессии, что позволяет улучшить качество модели за счет группировки переменных. Авторы статьи [7] предлагают адаптивный метод "elastic-net", который учитывает взаимосвязи между переменными и позволяет группировать их в условиях высокоразмерных данных. В [8] обсуждается метод "group lasso" в контексте многократного ядерного обучения, что способствует улучшению предсказательной способности моделей за счет группировки переменных.

Существует значительное количество исследований, посвященных использованию математических моделей для планирования и прогнозирования человеческих ресурсов, а также объемов финансирования. Так, в работе [9] показано применение методов математического моделирования при оценке штатной численности подразделений по защите информации. В [10] рассматриваются математические подходы к управлению персоналом, включая прогнозирование потребностей. Работа [11] рассматривает классические экономико-математические модели прогнозирования человеческих ресурсов. Большой интерес также представляет работа [12], посвященная методам оптимизации для

---

планирования персонала в производственных системах и [13], касающаяся применения методов нечеткой логики для планирования персонала в здравоохранении. В [14] на основе статистических данных приведена модель объема финансирования фонда социального и пенсионного обеспечения.

Отдельные работы посвящены кластерному анализу и группированию факторов при исследовании численности подразделений. Так, в статье [15] приведен обзор методов машинного обучения, включая кластеризацию для HR-аналитики, а работа [16] посвящена исследованию многокритериальной системы кластеризации компетенций сотрудников.

Вместе с тем, основная сложность в решении подобных задач заключается в корректном определении характеристик объекта исследования и их последующего встраивания в соответствующие математические модели различных классов. Кроме того, традиционные методы расчета численности часто оказываются недостаточно точными из-за сложного характера взаимосвязей между организационными параметрами и требуемым количеством специалистов.

### **Формирование групп переменных и оценивание параметров однородной вложенной кусочно-линейной регрессии.**

В работе [17] введена, в том числе, довольно сложная регрессионная конструкция - однородная вложенная кусочно-линейная регрессия первого типа:

$$y_k = \min \{ \min_{i \in J^1} \{ \alpha_i^1 x_{ki} \}, \dots, \min_{i \in J^H} \{ \alpha_i^H x_{ki} \} \} + \varepsilon_k, k = \overline{1, n}. \quad (1)$$

Здесь  $k$  – номер наблюдения,  $n$  – их количество,  $y$  – зависимая переменная,

$x_i, i = \overline{1, m}$  – независимые переменные,  $\alpha_i^j$  – подлежащие оцениванию

параметры,  $\varepsilon_k, k = \overline{1, n}$ , – ошибки аппроксимации, Индексные множества

$J^j, j = \overline{1, H}$  являются подмножествами множества независимых переменных

---

$\{1, 2, \dots, m\}$  и в общем случае могут иметь как пустые, так и непустые пересечения.

В [18] при заранее заданных составах множеств  $J^j, j = \overline{1, H}$  и соответствующей методу наименьших модулей функции потерь

$$I(\alpha^1, \dots, \alpha^H) = \sum_{k=1}^n |\varepsilon_k| \quad (2)$$

задача идентификации параметров регрессии (1) сведена к задаче линейно-булева программирования (ЛБП). Здесь  $\alpha^j = (\alpha_i^j, \dots, \alpha_{|J^j|}^j), j = \overline{1, H}$ , а  $|J^j|$  - число элементов в множестве  $J^j$ , его мощность.

Эту задачу можно существенным образом усложнить, сделав ее более приближенной к реальным ситуациям. Действительно, исследователю весьма затруднительно заранее определять состав множеств  $J^j, j = \overline{1, H}$ , поскольку заранее непонятен сам содержательный подход, который мог бы быть положен в основу такого определения, или, иными словами, кластеризации независимых переменных. Поэтому функция потерь (2) может быть задана в более общей форме:

$$\tilde{I}(\alpha^1, \dots, \alpha^H, J^1, \dots, J^H) = \sum_{k=1}^n |\varepsilon_k|. \quad (3)$$

То есть, необходимо вычислить как состав множеств  $J^j, j = \overline{1, H}$ , так и оценки параметров  $\alpha_i^j, i \in J^j, j = \overline{1, H}$ .

Будем, вместе с тем, считать известными количество элементов в индексных множествах  $J^j, j = \overline{1, H}$ :

$$|J^j| = p_j, j = \overline{1, H}.$$

Тогда, воспользовавшись вычислительными приемами из работ [18,19], задачу минимизации функции потерь (3) также можно свести к задаче ЛБП.

Иногда характер решаемой с помощью модели (1) прикладной задачи может потребовать, чтобы множества  $J^j, j = \overline{1, H}$  не пересекались:

$$J^j \cap J^i = \emptyset, i \neq j. \quad (4)$$

Реализация условия (4) достигается включением в состав условий упомянутой задачи ЛБП соответствующих дополнительных ограничений

### **Вложенная кусочно-линейная модель штатной численности подразделений по защите информации.**

Применим описанный выше подход для построения вложенной кусочно-линейной модели штатной численности подразделений по защите информации органов Социального фонда России (СФР) и проведем сравнительный анализ результатов с моделями, описанными в [9].

Зададим число внутренних минимумов в модели:  $H = 2$ .

Назначим мощность индексных множеств  $J^1$  и  $J^2$ :

$$|J^1| = 4, |J^2| = 3.$$

В результате решения задачи ЛБП на языке программирования C# с помощью библиотеки Google Or-Tools получим модель:

$$y_k = \min \{ \min \{ 50x_{k2}, 0.98x_{k4}, 1.43x_{k6}, 0.4x_{k7} \}, \min \{ 66.37x_{k3}, 0.98x_{k4}, 0.4x_{k7} \} \} + \varepsilon_k, \\ k = \overline{1, 10}. \quad (5)$$

$$E = 6.02\%, J^1 = \{2, 4, 6, 7\}, J^2 = \{3, 4, 7\}.$$

Здесь  $E$  – средняя процентная ошибка.

Рассчитаем вектор срабатывания независимых переменных  $\lambda$ , указывающий на то, какая из них реализовалась на том или ином наблюдении выборки:

$$\lambda = (7, 7, 7, 7, 6, 7, 7, 7, 7, 4, 4, 7).$$

Таким образом, независимые переменные  $x_4$  и  $x_7$  вошли в оба индексных множества, а, значит, в оба внутренних минимума модели (5).

Теперь наложим на эти множества условие (4) и вновь решим задачу ЛБП. В результате получим модель:

$$y_k = \min \{ \min \{ 0.98x_{k4}, 1.31x_{k5}, 1.43x_{k6}, 0.4x_{k7} \}, \min \{ 12.74x_{k1}, 50x_{k2}, 66.37x_{k3} \} \} + \varepsilon_k, \\ k = \overline{1,10}. \quad (6)$$

$$E = 6.02\%, J^1 = \{4,5,6,7\}, J^2 = \{1,2,3\}.$$

Вектор срабатывания переменных:

$$\lambda = (7,7,7,7,6,7,7,7,5,4,7).$$

Следует отметить, что расширение системы ограничений задачи ЛБП за счет этого условия не повлияло на величину средней процентной ошибки. При анализе моделей наблюдается следующее: в модель (5) не входят независимые переменные  $x_1$  и  $x_5$ , в то время как в модели (6) они присутствуют – одна в множестве  $J^1$ , а вторая – в  $J^2$ .

Экспериментальная проверка проводилась на данных 12 региональных подразделений за 2022 год [9]. Основные результаты:

### **1. Точность аппроксимации.**

Средняя относительная ошибка в построенных моделях составила 6,02%, что сопоставимо с линейными моделями в [9] (показатель ошибки до 10% считается приемлемым и в целом указывает на высокое качество модели), однако последняя аппроксимирует данные одной гиперплоскостью и, как следствие, теряет в точности с увеличением длины выборки, так как реальные данные редко имеют линейный характер. Кусочно-линейная модель, напротив, предполагает разделение данные на сегменты, на каждом из которых проявляется линейный характер межфакторного взаимодействия.

### **2. Значимые факторы.**

В отличие от линейной модели, характеризующейся одним вектором параметров, значимость которых в основном определяется t-критерием Стьюдента, кусочно-линейная модель позволяет учитывать своего рода "веса", проявляющиеся во вложенных минимумах и формировании

различных групп переменных, что, в частности, позволяет автоматически определять наиболее значимый ограничивающий фактор для каждого наблюдения. В данном случае им в основном является количество ИТ-специалистов ( $x_7$ ), на что указывает вектор срабатывания переменных.

Показано также, что помимо численности ИТ-персонала ( $x_7$ ), вторым ключевым фактором является количество защищаемых ресурсов ( $x_4$ ).

### Вывод

В данной работе рассмотрена задача применения однородной вложенной кусочно-линейной регрессии для моделирования штатной численности подразделений по защите информации СФР. На основе статистических данных за 2022 год построены две модели, учитывающие различные способы группировки независимых переменных, таких как количество сотрудников, объем выданных электронных подписей, количество защищенных узлов и другие показатели.

Основные результаты исследования состоят в следующем.

1. Представленный алгоритмический способ позволил с высокой точностью (средняя процентная ошибка 6,02%) оценить зависимость штатной численности подразделений от рассматриваемых факторов.
2. Установлено, что переменные  $x_4$  (количество защищаемых ресурсов) и  $x_7$  (количество специалистов ИТ) оказывают наиболее значимое влияние на выходной показатель и присутствуют в обеих построенных моделях.
3. Показано, что введение дополнительного условия непересечения групп переменных не ухудшает качество модели, но изменяет состав индексных множеств.

4. Сравнительный анализ с линейной моделью показывает на неоспоримые преимущества кусочно-линейных моделей по сравнению с линейной.

Предложенный подход может быть использован для:

- оптимизации штатной численности подразделений;
- прогнозирования потребности в кадрах;
- анализа влияния различных факторов на организационную

структуру предприятий.

Дальнейшие исследования могут быть направлены на расширение набора учитываемых переменных и адаптацию модели для подразделений по защите информации в некоторых других сферах.

### Литература

1. Tibshirani R., Saunders M., Rosset S., Zhu J., Knight K. Sparsity and smoothness via the fused lasso // Journal of the Royal Statistical Society Series B: Statistical Methodology, 2005, Vol. 67, Issue 1, pp. 91-108. URL: doi.org/10.1111/j.1467-9868.2005.00490.x.

2. Friedman J., Hastie T., Tibshirani R. A note on the group lasso and a sparse group lasso // Cornell University, 2010. URL: doi.org/10.48550/arXiv.1001.0736.

3. Yuan M., Lin Y. Model selection and estimation in regression with grouped variables // Journal of the Royal Statistical Society Series B: Statistical Methodology, 2006, Vol. 68, Issue 1, pp. 49-67. URL: doi.org/10.1111/j.1467-9868.2005.00532.x.

4. Simon N., Friedman J., Hastie T., Tibshirani R. A sparse-group lasso // Journal of Computational and Graphical Statistics, 2013, Vol. 22, Issue 2, pp. 231-245. URL: doi.org/10.1080/10618600.2012.681250.

5. Bühlmann P., Sara van de Geer S. Statistics for High-Dimensional Data: Methods, Theory and Applications, 2011, 555p.
  6. Meier L., Sara van de Geer, S., Bühlmann P. The group lasso for logistic regression // Journal of the Royal Statistical Society Series B: Statistical Methodology, 2008, Vol.70 Issue 1, pp. 53-71. URL: doi.org/10.1111/j.1467-9868.2007.00627.x.
  7. Zou H., Zhang H.H. On the adaptive elastic-net with a diverging number of parameters // The Annals of Statistics, 2009, Vol. 37, Issue 4, pp. 1733-1751. URL: doi.org/10.1214/08-AOS625.
  8. Bach F.R. Consistency of the group lasso and multiple kernel learning // Journal of Machine Learning Research, 2008, Vol. 9, pp. 1179-1225.
  9. Носков С.И., Медведев А.П., Глухов Н.И. Регрессионное моделирование штатной численности подразделений по защите информации // Инженерный вестник Дона, 2024, № 6. URL: ivdon.ru/ru/magazine/archive/n6y2024/9283 (дата обращения: 01.03.2025).
  10. Ahmad S., Schroeder R. G. The impact of human resource management practices on operational performance: recognizing country and industry differences // Journal of Operations Management, 2003, Vol. 21, Issue 1, pp. 19-43.
  11. Cascio W.F. Costing Human Resources: The Financial Impact of Behavior in Organizations // Academy of Management Review, 1984, Vol. 9, No 2, pp. 370-371. URL: doi.org/10.2307/258455.
  12. Heuser P., Letmathe P., Schinner M. Workforce planning in production with flexible or budgeted employee training and volatile demand // Journal of Business Economics, 2022, No. 92, pp. 1093-1124. URL: doi.org/10.1007/s11573-022-01090-z.
  13. Topaloglu S., Selim H. Nurse scheduling using fuzzy multiple objective programming // Conference: New Trends in Applied Artificial Intelligence, 20th International Conference on Industrial, Engineering and Other Applications of
-

Applied Intelligent Systems, IEA/AIE, 2007, pp. 54-63. URL: [dx.doi.org/10.1007/978-3-540-73325-6\\_6](https://dx.doi.org/10.1007/978-3-540-73325-6_6).

14. Носков С.И., Медведев А.П. Реализация конкурса регрессионных моделей при оценке объема финансирования социального и пенсионного обеспечения // Инженерный вестник Дона, 2024, №4. URL: [ivdon.ru/ru/magazine/archive/n4y2024/9155](https://ivdon.ru/ru/magazine/archive/n4y2024/9155) (дата обращения: 01.03.2025).

15. Ersoz T., Ersoz F., Bedir E. Predictive analytics in human resources using machine learning and data mining // International Journal of 3D Printing Technologies and Digital Industry, 2023, No. 7 (3). URL: [dx.doi.org/10.46519/ij3dptdi.1379628](https://dx.doi.org/10.46519/ij3dptdi.1379628).

16. Van der Velden R.D. Multicriteria competency clustering framework. Exploiting employees' competencies for strategic workforce planning // Tilburg School of Economics and Management. Tilburg University, 2022, 104p.

17. Носков С.И. Некоторые формы вложенной кусочно-линейной регрессии // Известия Тульского государственного университета. Технические науки, 2023, № 3, С. 467-469.

18. Носков С.И., Белинская С.И. Вычисление оценок параметров однородной вложенной кусочно-линейной регрессии // Вестник Дагестанского государственного технического университета. Технические науки, 2023, № 50 (4), С. 115-120.

19. Носков С.И., Чекалова А.Р. Использование комбинированных функций потерь при построении вложенных кусочно-линейных регрессионных моделей // Вестник Череповецкого государственного университета, 2025, № 1 (124), С. 59-69.

## References

1. Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., Knight, K. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2005, Vol. 67, Issue 1, pp.91-108. URL: doi.org/10.1111/j.1467-9868.2005.00490.x.
  2. Friedman J., Hastie T., Tibshirani R. Cornell University, 2010. URL: doi.org/10.48550/arXiv.1001.0736.
  3. Yuan M., Lin Y. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2006, Vol. 68, Issue 1, pp. 49-67. URL: doi.org 10.1111/j.1467-9868.2005.00532.x.
  4. Simon N., Friedman J., Hastie T., Tibshirani R. Journal of Computational and Graphical Statistics, 2013, Vol. 22, Issue 2, pp. 231-245. URL: doi.org/10.1080/10618600.2012.681250.
  5. Bühlmann P., Sara van de Geer, S. Statistics for High-Dimensional Data: Methods, Theory and Applications, 2011, 555p.
  6. Meier L., van de Geer, S., Bühlmann P. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2008, Vol.70 Issue 1, pp. 53-71. URL: doi.org/10.1111/j.1467-9868.2007.00627.x.
  7. Zou H., Zhang H.H. The Annals of Statistics, 2009, Vol. 37 Issue 4, pp. 1733-1751. URL: doi.org//10.1214/08-AOS625.
  8. Bach F.R. Journal of Machine Learning Research, 2008, Vol. 9, pp. 1179-1225.
  9. Noskov S.I., Medvedev A.P., Gluhov N.I. Inzhenernyj vestnik Dona, 2024, № 6. URL: ivdon.ru/ru/magazine/archive/n6y2024/9283.
  10. Ahmad S., Schroeder R. G. Journal of Operations Management, 2003, Vol. 21, Issue 1, pp. 19-43.
  11. Cascio W.F. Academy of Management Review, 1984, Vol. 9, No 2, pp. 370-371. URL: doi.org/10.2307/258455.
-



12. Heuser P., Letmathe P., Schinner M. Journal of Business Economics, 2022, No. 92, pp. 1093-1124. URL: [doi.org/10.1007/s11573-022-01090-z](https://doi.org/10.1007/s11573-022-01090-z).

13. Topaloglu S., Selim H. Conference: new trends in applied artificial intelligence, 20th international conference on industrial, engineering and other applications of applied intelligent systems, IEA/AIE, 2007, pp. 54-63. URL: [dx.doi.org/10.1007/978-3-540-73325-6\\_6](https://dx.doi.org/10.1007/978-3-540-73325-6_6).

14. Noskov S.I., Medvedev A.P. Inzhenernyj vestnik Dona, 2024, No.4. URL: [ivdon.ru/en/magazine/archive/n4y2024/9155](https://ivdon.ru/en/magazine/archive/n4y2024/9155).

15. Ersoz T., Ersoz F., Bedir E. International Journal of 3D Printing Technologies and Digital Industry, 2023, No. 7 (3). URL: [dx.doi.org/10.46519/ij3dptdi.1379628](https://dx.doi.org/10.46519/ij3dptdi.1379628).

16. Van der Velden R.D. Tilburg School of Economics and Management. Tilburg University, 2022, 104p.

17. Noskov S.I. Izvestija Tul'skogo gosudarstvennogo universiteta. Tehnicheskie nauki, 2023, № 3, pp. 467-469.

18. Noskov S.I., Belinskaja S.I. Vestnik Dagestanskogo gosudarstvennogo tehničeskogo universiteta. Tehnicheskie nauki, 2023, № 50 (4), pp. 115-120.

19. Noskov S.I., Chekalova A.R. Vestnik Cherepoveckogo gosudarstvennogo universiteta, 2025. No. 1 (124), pp. 59-69.

**Дата поступления: 4.04.2025**

**Дата публикации: 25.06.2025**