

Мультиколлинеарность в функционировании процессов информационно-управляющей системы предприятия

Рубцов А.В., к.э.н., доцент кафедры базовых дисциплин, Лесосибирского педагогического института – филиала СФУ, Лесосибирск, Россия

Мамаева С.В., к.фил.н., доцент кафедры базовых дисциплин, Лесосибирского педагогического института – филиала СФУ, Лесосибирск, Россия

Мальцева М.В., студентка 3 курса филологического факультета, бакалавриат, направление: Педагогическое образование с двумя профилями: Иностранный язык (английский, немецкий), Лесосибирского педагогического института – филиала СФУ, Лесосибирск, Россия

Храмов И.В., ассистент кафедры базовых дисциплин, Лесосибирского педагогического института – филиала СФУ, Лесосибирск, Россия

Аннотация. В данной работе подчеркиваются особенности процессов информационно-управляющей системы, базирующихся на способах их функционирования. Выделяется явление мультиколлинеарности и приводится её математическое обоснование. Статья направлена на улучшение функционирования информационно-управляющих систем предприятий.

Ключевые слова: мультиколлинеарность, эконометрическое моделирование, информационно-управляющая система предприятия, информационный менеджмент

Multicollinearity in the functioning of the enterprise's information management system processes

Rubtsov A.V., Ph.D., associate Professor of the Department of basic disciplines, Lesosibirsk pedagogical Institute – branch of SibFU, Lesosibirsk, Russia

Mamayeva S.V., Ph.D., associate Professor of the Department of basic disciplines, Lesosibirsk pedagogical Institute – branch of SibFU, Lesosibirsk, Russia

Maltseva M.V., 3rd year student of the faculty of Philology, bachelor's degree, direction: Pedagogical education with two profiles: Foreign language (English, German), Lesosibirsk pedagogical Institute – branch of SibFU, Lesosibirsk, Russia

Khramov I.V., assistant of the Department of basic disciplines, Lesosibirsk pedagogical Institute – branch of SFU, Lesosibirsk, Russia

Annotation. This paper highlights the features of the information management system processes based on the methods of their functioning. The phenomenon of multicollinearity is highlighted and its mathematical justification is given. The article is aimed at improving the functioning of information management systems of enterprises.

Keywords: multicollinearity, econometric modeling, enterprise information management system, information management

Введение

При анализе функционирования всякой системы возникает потребность в прогнозировании выполнения ею тех или иных процессов. Причем, очень часто такое прогнозирование заключается в поиске неизвестных величин, исходя из имеющихся реальных данных. Это, в свою очередь, порождает множество ошибок, неточность полученных данных или их очень большую вариативность (дисперсию). То же касается и такой многозадачной и важной для любой организации составляющей, как информационно-управляющей системы.

Информационно-управляющая система (далее ИУС) – это совокупность компонентов, подготавливающих и предоставляющих управленческую, производственную и финансовую информацию для принятия решений [9; с.32]

Функционирование информационно-управляющей системы отличается рядом многоплановых задач, поскольку действующие в ней информационные потоки охватывают различные сферы любой организации: логистика, бухгалтерия, кадры, производительность, качество и др. По этой причине

моделирование всех процессов должно отличаться разными методами, способными учитывать особенные параметры каждого из процессов.

Проблемам, связанным с функционированием информационно-управляющей системы посвящено множество работ таких отечественных исследователей, как Э.В. Мельника, А.Б. Клименко, А.А. Соколовой, Т.В. Малачиханова, Ю.А. Кропотова, А.А. Белова, А.Ю. Проскурякова и др.

Так, в исследованиях Э.В. Мельника рассматривается вопрос моделирования информационно-управляющих систем в совокупности с внедрением новых методик [7]. В работах Ю.А. Кропотова, А.А. Белова, А.Ю. Проскурякова исследуются вопросы прогнозирования изменений параметров информационно-управляющих систем [5]. Исследования А.А. Соколовой посвящены формулированию требований к современной информации, необходимой для подготовки, принятия, успешной реализации максимально эффективных управленческих решений [9].

В данном исследовании рассматриваются аспекты математического инструментария функционирования информационно-управляющей системы, а именно самое распространенное искомое свойство – свойство выявления зависимости между переменными, через которые строится прогноз (мультиколлинearности). По мнению многих исследователей, в частности, М.А. Муминовой, С.А. Сайдуллаевой и др. мультиколлинearность является проблемой для математического обоснования того или иного процесса [8; с.16].

Объектом настоящего исследования является функционирование информационно-управляющей системы.

Предмет настоящего исследования – мультиколлинearность процессов информационно-управляющей системы предприятия.

Цель настоящей статьи – выявление особенностей мультиколлинearности в процессах информационно-управляющей системы предприятия.

1. Определение и виды мультиколлинearности

Как уже было отмечено выше, многие исследователи считают мультиколлинearность проблемой, препятствующей проведению вычислений

как экспертным, так и автоматизированным путём. Для того, чтобы избежать появления большой вариативности между данными, широко используется метод наименьших квадратов.

Метод наименьших квадратов (МНК) – один из основных способов уменьшить размерность данных, потеряв наименьшее количество информации. Используется в различных задачах для представления набора данных, его визуализации или сохранения максимума информации о данных в минимальном количестве переменных.

Мультиколлинеарность — это явление, состоящее в том, что среди объясняющих переменных (регрессоров) существует линейная зависимость [2; с.3].

Так, линейная зависимость между регрессорами классифицируется на:

- строгую (идеальную) линейную зависимость;
- нестрогую (примерную) линейную зависимость.

Рассмотрим пример строгой мультиколлинеарности:

$$X = \begin{pmatrix} 1 & 4 & 12 & 8 \\ 1 & 3 & 3 & 3 \\ 1 & 1 & 7 & 4 \\ 1 & 2 & 4 & 3 \\ 1 & 3 & 5 & 4 \end{pmatrix} \quad (1)$$

В данном примере можно заметить, что если сложить второй столбец матрицы (второй регрессор) и третий столбец матрицы (третий регрессор), то получим значение четвертого столбца (8, 3, 4, 3, 4) умноженное на два:

$$x_2 + x_3 = 2x_4 \quad (2)$$

В реальных условиях строгая мультиколлинеарность может случиться при ошибке исследователя во введении данных для последующих автоматизированных расчетов, или при ошибке введения дамми-переменных (переменных, принимающих значение 0 или 1) [3; с. 698]. Рассмотрим ошибочно составленную модель с использованием дамми-переменных. В данной модели,

условно, исследователь попытался оценить зависимость заработной платы от количества лет обучения и нахождения сотрудников двух подразделений.

$$wage_i = \beta_1 + \beta_2 first_i + \beta_3 second_i + \beta_4 educ_i + \varepsilon_i \quad (3)$$

где:

$first = 1$, переменная, обозначающая сотрудников первого подразделения;

$second = 2$, переменная, обозначающая сотрудников второго подразделения;

β_1 – свободный член;

$educ$ – переменная, показывающая количество лет обучения.

Для выявления строгой линейной зависимости в данном примере составим матрицу всех входящих в модель регрессоров. Получим:

$$X = \begin{pmatrix} 1 & 1 & 0 & 16 \\ 1 & 1 & 0 & 11 \\ 1 & 0 & 1 & 18 \end{pmatrix} \quad (4)$$

Таким образом сумма второго и третьего регрессоров (β_2, β_3) равны значению первого регрессора β_1 : $x_2 + x_3 = x_1$. Следовательно, мы обнаружили строгую линейную зависимость, которая свидетельствует о наличии подмножества линейного пространства.

Рассмотрим последствия строгой линейной зависимости в теоретическом аспекте. Главная проблема состоит в том, что при использовании вышеупомянутого метода наименьших квадратов в решении задач на избежание мультиколлинеарности, оценки, полученные таким методом, не являются единственными [1; с.6]. Рассмотрим эту проблему в трёх оценённых уравнениях:

$$\widehat{wage}_i = 15 + 3first_i - 2second_i + 3educ_i \quad (5)$$

$$\widehat{wage}_i = 28 - 10first_i - 15second_i + 3educ_i \quad (6)$$

$$\widehat{wage}_i = 18 + 0first_i - 5second_i + 3educ_i \quad (7)$$

Если мы рассмотрим уравнение (2), где $first = 1$, $second = 0$, то искомая величина прогнозируемой заработной платы в объясняющем её процессе ИУС будет равна $18 + 3educ_i$. Рассмотрим модель с отличными от (6) коэффициентами – уравнение (7), где $first = 1$, $second = 0$. Получим $18 + 3educ_i$. Таким образом, модели с абсолютно разными коэффициентами дают одинаковые прогнозы, что является следствием строгой линейной зависимости между объясняющими переменными. В результате, ни эксперту, ни единице определённого программного обеспечения не удастся выбрать модель выполнения той или иной задачи подразделов ИУС [6]. Основываясь на международном опыте, в настоящей статье мы рекомендуем наиболее общий способ устранения мультиколлинеарности – правильное введение дамми-переменных [10; с.3]. А именно, – выявление той переменной, значение которой будет равно 1 и 0 соответственно.

Перейдём к рассмотрению нестрогой мультиколлинеарности. Нестрогая мультиколлинеарность характеризуется тем, что между объясняющими переменными существует примерная зависимость [8; с.14]. В отличие от первого случая, данная проблема является более распространённой в функционировании как информационно-управляющей системы, так и всего информационного обеспечения предприятия. Нестрогая мультиколлинеарность в основном возникает в двух случаях. Во-первых, когда регрессоры моделируемой задачи или процесса измеряют один и тот же показатель с математической точки зрения (например, производительность серверов на начало и конец рабочего дня). Во-вторых, когда между регрессорами существуют естественные соотношения: возраст, стаж, количество лет обучения.

Оценим последствия нестрогой мультиколлинеарности:

– нестрогая мультиколлинеарность не нарушает стандартных предпосылок, например, предпосылок теоремы Гаусса-Маркова. Таким образом отмечается существование оценок, полученных МНК, которые характеризуются

несмещённостью, асимптотической нормальностью (обладают наименьшей дисперсией), возможностью проверки гипотез и построений доверительных интервалов.

– в результате нестрогой мультиколлинеарности один из регрессоров возможно объяснить через другие. Рассмотрим модель выявления высоких стандартных ошибок, характерных для процессов прогнозирования ИУС:

$$se^2(\hat{\beta}_j) = \frac{\hat{\sigma}^2}{RSS_j} = \frac{\hat{\sigma}^2}{TSS_j*(1-R_j^2)} = \frac{1}{1-R_j^2} \frac{\hat{\sigma}^2}{TSS_j} \quad (8)$$

Предположим, что существует некая линейная зависимость между объясняющими переменными уравнения (8), тогда j -объясняющая переменная входит в формулу зависимости между регрессорами. Соответственно, один из регрессоров будет объясняться другими: производительность отдела будет объясняться числом сотрудников и количеством лет обучения, а сумма квадратов остатков RSS_j будет маленькой. Это объясняет, что $\hat{\beta}_j$ будет достаточно большой. В данном случае, последствием нестрогой мультиколлинеарности будут являться высокие стандартные ошибки.

Однако, высокие стандартные ошибки зачастую приводят к тому, что доверительный интервалы для настоящих неизвестных коэффициентов становятся шире, сами коэффициенты получаются незначимыми (т.е. гипотеза о том, что один из коэффициентов равен нулю не отвергается). Таким образом, модель, отражающая функционирование того или иного процесса ИУС становится неустойчивой, т.е. при добавлении или изъятии одного из наблюдений оценки коэффициентов резко меняются.

С точки зрения исследователя функционирования ИУС, обычным проявлением нестрогой мультиколлинеарности является следующее. Наличие нескольких коэффициентов, незначимых по отдельности. Следовательно, гипотеза об их одновременном равенстве нулю отвергается, а модель является неэффективной.

Для определения наличия мультиколлинеарности используются следующие количественные признаки:

– показатель вздутия дисперсии (*Variance Inflation Factor*), считаемый отдельно для каждой объясняющей переменной. Мультиколлинеарность существует тогда, когда $VIF_j > 10$.

– выборочные корреляции между регрессорами. Мультиколлинеарность существует тогда, когда $sCorr(x, y) > 0.9$.

2. Способы решения проблемы нестрогой мультиколлинеарности в функционировании информационно-управляющей системы

Прежде чем приступить к избавлению от мультиколлинеарности в моделях информационно-управляющей системы, необходимо определить, является ли это свойство проблемой на данном этапе. Поскольку информационно-управляющую систему отличает многозадачность, нужно учитывать характеристики каждого процесса отдельно.

При нестрогой мультиколлинеарности, встречающейся намного чаще строгой, соблюдается теорема Гаусса-Маркова, что свидетельствует о том, что получаемые оценки $\hat{\beta}_j$ обладают наименьшей дисперсией среди несмещённых оценок. Следовательно, на доверительных интервалах мультиколлинеарность не сказывается. Поэтому, если процесс ИУС рассчитан только лишь на вычисление прогноза без интерпретации коэффициентов, то мультиколлинеарность не является проблемной задачей.

Однако, если необходимо выявить зависимость в моделируемом процессе, можно уменьшить дисперсию оценок, пожертвовав их несмещённостью [4]. В данном случае придётся избежать нестрогой мультиколлинеарности.

Рассмотрим задачу, где выявим существование мультиколлинеарности и определим способ её устранения. Предположим, что исследователем была оценена следующая модель:

$$y_i = \beta_1 + \beta_2 * x_i + \beta_3 * z_i + \beta_4 * w_i + \varepsilon_i \quad (9)$$

где $\beta_1, \beta_2, \beta_3, \beta_4$ – коэффициенты, x_i, z_i, w_i – объясняющие переменные, ε_i – случайная составляющая. Пусть между x_i, z_i, w_i существует некая линейная зависимость. Известно, что:

квадрат вспомогательной регрессии x_i $R_2^2 = 0.5$, в регрессии ($x_i = \gamma_1 + \gamma_2 * z_i + \gamma_3 * w_i + u_i$) (10);

квадрат вспомогательной регрессии z_i $R_3^2 = 0.95$ в регрессии ($z_i = \gamma_1 + \gamma_2 * x_i + \gamma_3 * w_i + u_i$) (11);

квадрат вспомогательной регрессии w_i $R_4^2 = 0.98$ в регрессии ($w_i = \gamma_1 + \gamma_2 * x_i + \gamma_3 * z_i + u_i$) (12);

Необходимо выявить наличие мультиколлинеарности в модели (5) и между какими переменными модели (9) наблюдается линейная связь. Для определения ответов на поставленные вопросы, воспользуемся показателем вздутия дисперсии (*Variance Inflation Factor*), рассчитываемым как:

$$VIF_j = \frac{1}{1-R_j^2} \quad (10)$$

Таким образом, рассчитывая показатель вздутия дисперсии для каждого регрессора, получаем: $VIF_x = 2$; $VIF_z = 20$; $VIF_w = 50$.

Получив оценки данного показателя, видим, что переменная x_i плохо объясняется через другие ($VIF_x < 10$), следовательно можно говорить об её независимости от остальных регрессоров. Переменные z_i, w_i достаточно сильно зависят от остальных регрессоров, т.к. $VIF_{z,w} > 10$. В данном случае мы видим, что в модели имеется мультиколлинеарность, вызванная тем, что z_i, w_i объясняются другими регрессорами ввиду оценивания показателя вздутия их дисперсии. Поэтому, зависимыми друг от друга переменными будут считаться z_i, w_i .

Таким образом, без ввода дополнительных переменных, мы выявили наличие мультиколлинеарности в задаче при заданных условиях, чем можно обосновать, что при последующем моделировании определённого процесса ИУС можно будет пожертвовать одной из переменных, не входящих в зависимость. Тем самым снизив чувствительность и повысив устойчивость модели.

Заключение

В данной статье мы обратили внимание на процессы информационно-управляющей системы, функционированию которых препятствуют некоторые математические проблемы. В результате проведенного исследования, нами было выявлено определение и природа такого явления как мультиколлинеарность, были охарактеризованы основные виды мультиколлинеарности, второй из которых оказался, по нашим доводам, более часто выявляемым на практике. Сравнивая настоящую работу с предыдущими исследованиями, где подчеркивались особенности функционирования процессов ИУС [5], их многозадачность и многопрофильность [7], нами было продемонстрировано явление мультиколлинеарности в задачах одного из таких процессов, описаны особенности и способы устранения данного свойства, а также на примере решения задачи было показано выявление зависимости между переменными.

Библиографический список

1. Ахлюстин С.Б., Мельников А.В., Черников Д.Н. Моделирование и прогноз обобщенных нелинейных показателей качества элементов интегрированных систем безопасности // Вестник Воронежского института МВД России. – 2019. – № 1. – С. 1-12.
2. Давлятова Б. Проблемы мультиколлинеарности и гетероскедастичности в построении модели конечного потребления на примере Кыргызской республики // Вестник науки и образования. – 2019. – № 2. – С. 1-5.
3. Демешев Б.Б., Малаховская О.А. Макроэкономическое прогнозирование с помощью BVAR Литтермана // Экономический журнал Высшей школы экономики. – 2016. – № 2. – С. 691-710.
4. Заернюк В. М., Филимонова Н.Н. Эконометрическое прогнозирование влияния производственных факторов на снижение себестоимости продукции // Сервис plus. – 2018. – № 1 (12). – С. 56-66.
5. Кропотов Ю. А., Белов А.А., Проскуряков А.Ю. Прогнозирование изменений параметров временных рядов в цифровых информационно-

управляющих системах // Системы управления, связи и безопасности. – 2017. – № 2. – С. 1-17.

6. Малачиханов Т.В. Информационное обеспечение менеджмента // Проблемы науки. – 2017. – № 1. – С. 1-2.

7. Мельник Э.В., Клименко А.Б. Применение концепции «Туманных» вычислений при проектировании высоконадежных информационно-управляющих систем // Известия Тульского государственного университета. Технические науки. – 2020. – № 2. – С. 273-283.

8. Нуруллаева Ш.Т., Рузуметова Н.Ш., Муминова М.А., Сайдуллаева С.А. Проблема мультиколлинеарности при построении эконометрических моделей // International scientific and practical conference world science. – 2018. – № 4 (32). – С. 14-18.

9. Соколова А.А. Информационно-управляющие системы в условиях чрезвычайных ситуаций: проблемы и перспективы // Вестник Полесского государственного университета. Серия общественных и гуманитарных наук. – 2019. – № 2. – С. 31-35.

10. Черемухин А.Д. Оценка вероятности прекращения деятельности и предоставления недостоверной информации сельскохозяйственными организациями // Вестник НГИЭИ. – 2019. – № 7 (98). – С. 1-10.

References

1. Akhlyustin S. B., Melnikov A.V., Chernikov D. N. Modeling and forecast of generalized nonlinear quality indicators of integrated security systems elements // Bulletin of the Voronezh Institute of the Ministry of internal Affairs of Russia, – 2019. – № 1. – Pp. 1-12.

2. Davlyatova B. Problems of multicollinearity and heteroscedasticity in building a model of final consumption on the example of the Kyrgyz Republic // Bulletin of science and education. – 2019. – № 2. – P. 1-5.

3. Demeshev B. B., Malakhovskaya O. A. Macroeconomic forecasting using Letterman's BVAR // Economic journal of the Higher school of Economics, – 2016. – № 2. – Pp. 691-710.
4. Zaernyuk V. M., Filimonova N. N. Econometric forecasting of the impact of production factors on reducing the cost of production // Service plus. – 2018. – № 1 (12). – Pp. 56-66.
5. Kropotov Yu. a., Belov A. A., Proskuryakov A. Yu. Predicting changes in time series parameters in digital information and control systems // management Systems, communications and security. – 2017. – № 2. – Pp. 1-17.
6. Malachikhanov T. V. Information support of management // Problems of science, – 2017. – № 1. – Pp. 1-2.
7. Melnik E.V., Klimenko A.B. Application of the concept of «Foggy» computing in the design of highly reliable information and control systems // Izvestiya of the Tula state University. Technical Sciences, – 2020. – № 2. – Pp. 273-283.
8. Nurullayeva Sh.T., Ruzumetova N.Sh., Muminova M.A., Saidullayeva S.A. the Problem of multicollinearity in the construction of econometric models // INTERNATIONAL SCIENTIFIC AND PRACTICAL CONFERENCE WORLD SCIENCE. – 2018. – № 4 (32). – Pp. 14-18.
9. Sokolova A.A. Information and control systems in emergency situations: problems and prospects // Bulletin of the Polesie state University. Series of social and humanitarian Sciences, – 2019. – № 2. – Pp. 31-35.
10. Cheremukhin A.D. Assessment of the probability of termination of activities and provision of false information by agricultural organizations // Bulletin of the NGIEI. – 2019. – № 7 (98). – Pp. 1-10.