
DOI: 10.12731/2227-930X-2016-2-55-66

КОМПЬЮТЕРНЫЕ МЕТОДЫ ГЕНЕТИЧЕСКОГО АНАЛИЗА

Осипов А.Л., Трушина В.П., Чентаева Е.А.

Рассматриваются основные статистические методы, которые используются при проведении генетического анализа признаков человека. Исследованы методы сегрегационного анализа, анализа сцепления и аллельных ассоциаций. Разработано программное обеспечение для реализации этих методов.

Ключевые слова: сцепление; ассоциация; принятие решений; статистические методы; программирование; картирование; выборка; анализ.

COMPUTER METHODS OF GENETIC ANALYSIS

Osipov A.L., Trushina V.P., Chentaeva E.A.

The basic statistical methods used in conducting the genetic analysis of human traits. We studied by segregation analysis, linkage analysis and allelic associations. Developed software for the implementation of these methods support.

Keywords: clutch; association; decision making; statistical methods; programming; mapping; sampling; analysis.

Генетический анализ — это область генетики, занимающаяся выяснением механизмов генетической детерминации различных признаков. В рамках этого раздела решаются вопросы о том, сколько и, каких генов участвует в обеспечении полиморфизма

признака, где они локализованы, как функционируют, взаимодействуют ли друг с другом, модифицируются ли внешними факторами и др. [2, с. 6]. Одним из важнейших этапов генетического анализа является картирование генов, которое позволяет получать информацию о позиции гена и оказывать огромное влияние на понимание механизма зарождения и развития многих болезней.

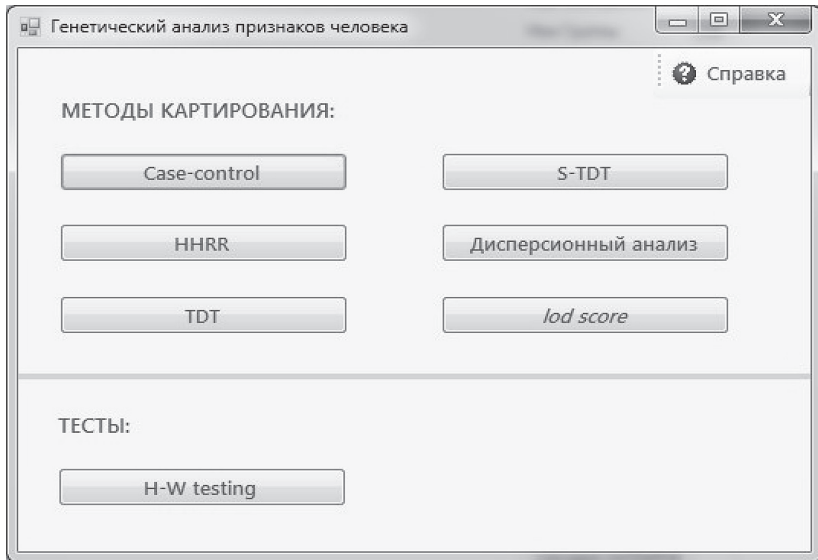


Рис. 1. Главная форма приложения

Целью работы является разработка программного обеспечения для генетического анализа признаков человека с помощью статистических методов анализа сцепления и ассоциаций. Основная идея картирования с помощью анализа ассоциаций заключается в следующем. Если у большинства больных в популяции мутантный аллель имеет общее происхождение, окружающие маркеры находятся с ним в неравновесии по сцеплению. Для локализации гена, контролирующего болезнь, надо найти такой маркер, один из аллелей которого преобладает у больных [3, с. 381]. С помо-

щью анализа ассоциаций можно картировать как простые бинарные, так и количественные признаки. Материалом для анализа могут служить случайные выборки из популяции, независимые группы больных и здоровых людей или выборки родственных особей [1, с. 34]. Многообразии дизайнов эксперимента привело к созданию большого количества статистических методов анализа, соответствующих типу признака, числу и типу маркеров, структуре выборки. В разрабатываемом программном приложении имеется одна главная форма, представленная на рис. 1.

Рассмотрим работу приложения на примерах с использованием реальных данных. Пусть маркерный локус представлен тремя аллелями, распределение которых у больных и здоровых показано в таблице 1.

Таблица 1.

Пример данных для метода Case-control

Аллели	Больные	Здоровые
1	104	75
2	86	33
3	56	104

На рис. 2 показана форма Case-control и расчет показателей:

Данные:

	Алель	Больные	Здоровые	Всего
▶	1	104	75	179
	2	86	33	119
	3	56	104	160
*	Всего	246	212	458

Критерий: Хи-квадрат

χ² = 40.4

Рис. 2. Форма Case-control

Результат теста, полученный с помощью метода case-control при расчете критерия хи-квадрат, показан на рис. 3:

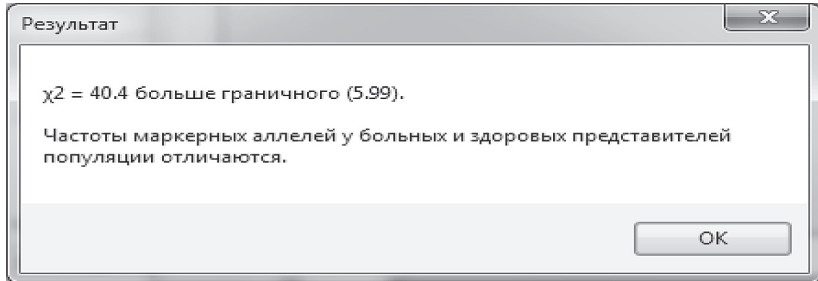


Рис. 3. Результат теста с помощью метода case-control

Рассмотрим еще один пример. Пусть в группу больных попали 28 человек с генотипом ММ, 48 – с генотипом Мм и 19 человек с генотипом мм. В группе здоровых 29 человек с генотипом ММ, 17 – с генотипом Мм и 8 человек с генотипом мм. Из этих данных легко получить распределение аллелей маркерного локуса среди больных и здоровых людей, которое и заносится в таблицу на форме Case-control. Произведем расчет критерия Odds Ratio, выбрав его из списка справа, от таблицы. На рис. 4 показана форма Case-control после ввода данных контрольного примера и расчета показателей.

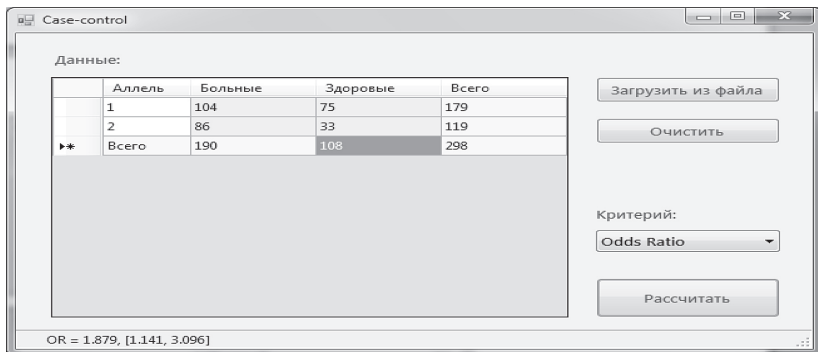


Рис. 4. Форма Case-control после ввода данных

Результат теста, полученный с помощью метода case-control при расчете критерия Odds Ratio, показан на рис. 5:

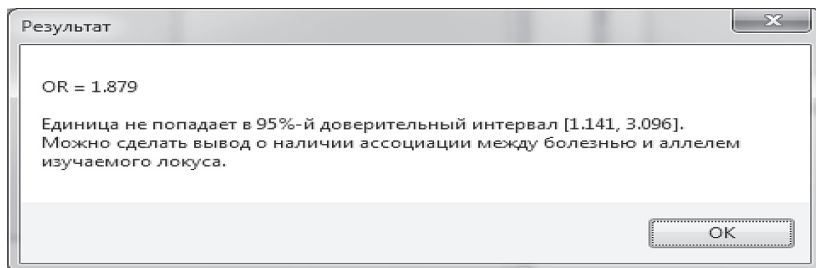


Рис. 5. Результат теста с помощью метода case-control при расчете OR

В качестве примера для метода NHRR рассмотрим выборку, состоящую из 100 больных детей и их родителей, и сформируем таблицу, характеризующую трансмиссию аллелей, данные которой приведены в таблице 2:

Таблица 2.

Пример данных для метода NHRR

Переданные аллели	Непереданные аллели	
	аллель 1	аллель 2
Аллель M	93	31
Аллель m	63	13

Значения, приведенные в этой таблице, используем для вычисления критерия NHRR. На рис. 6 показана форма NHRR после ввода данных:

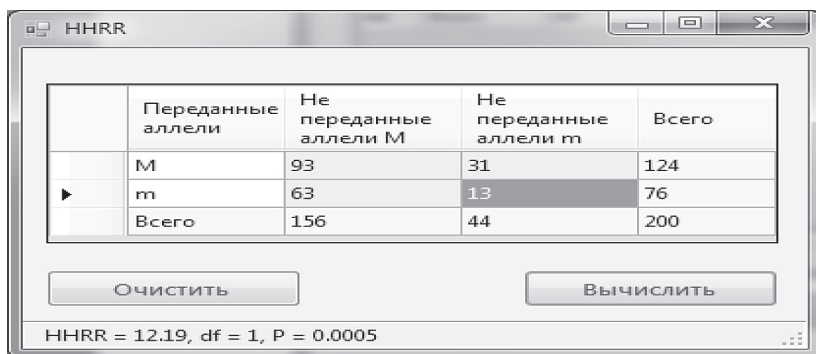


Рис. 6. Форма NHRR после ввода данных

Результат теста с помощью метода HHRR показан на рис. 7:

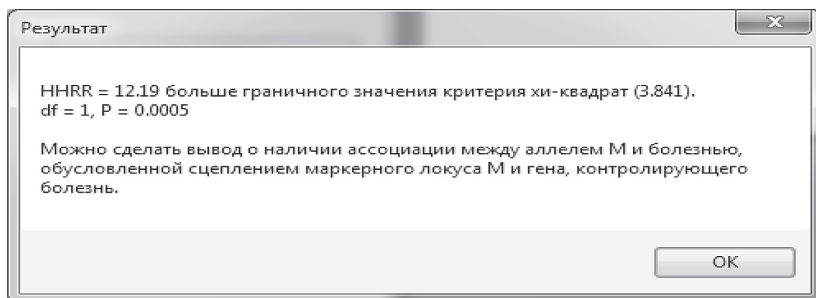


Рис. 7. Результат теста с помощью метода HHRR

В качестве данных для примера тестирования с помощью метода TDT рассмотрим значения, приведенные выше в таблице 2. На рис. 8 показана форма TDT после ввода данных.

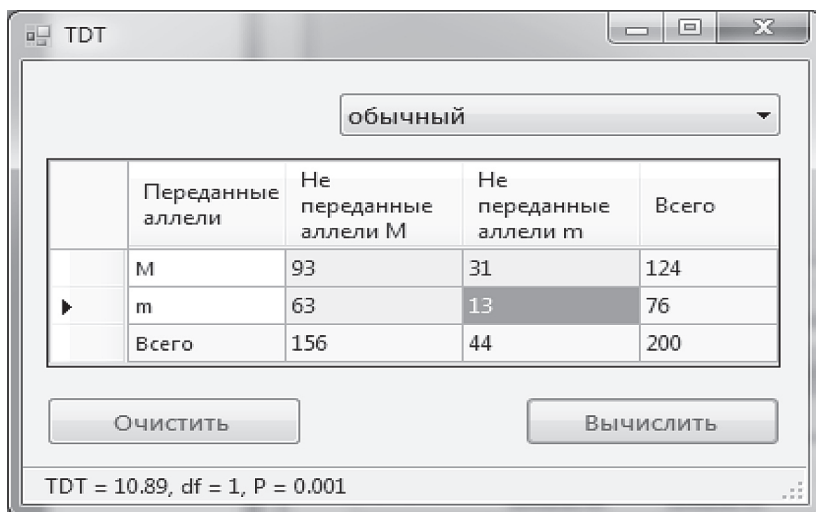


Рис. 8. Форма TDT после ввода данных

Результат теста, полученный с помощью метода TDT при расчете обычного критерия, показан на рис. 9:

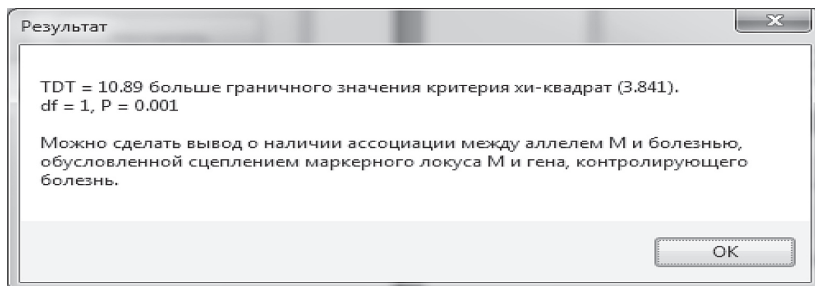


Рис. 9. Результат теста с помощью метода TDT

На рис. 10 показана форма TDT после ввода данных и расчета критерия TDT, основанного на отношении правдоподобия.

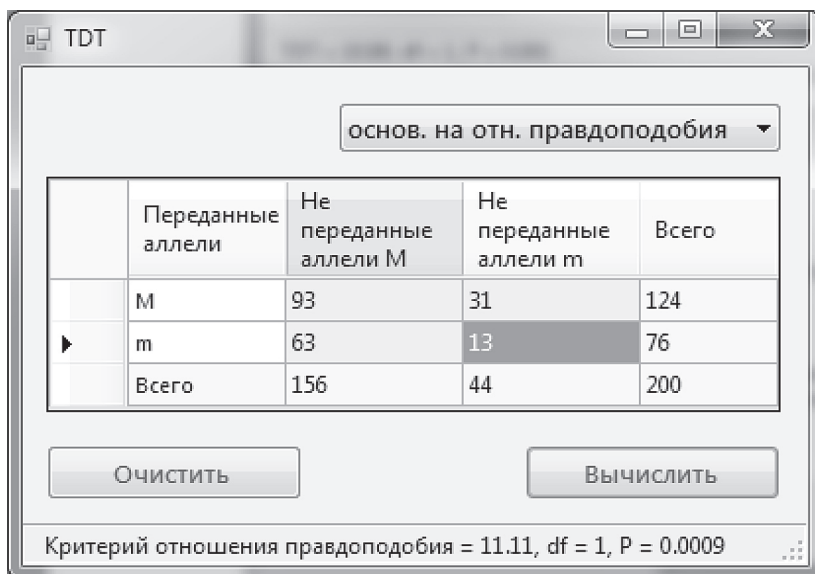


Рис. 10. Форма TDT после ввода данных

Результат теста, полученный с помощью метода TDT при расчете критерия, основанного на отношении правдоподобия, показан на рис. 11:

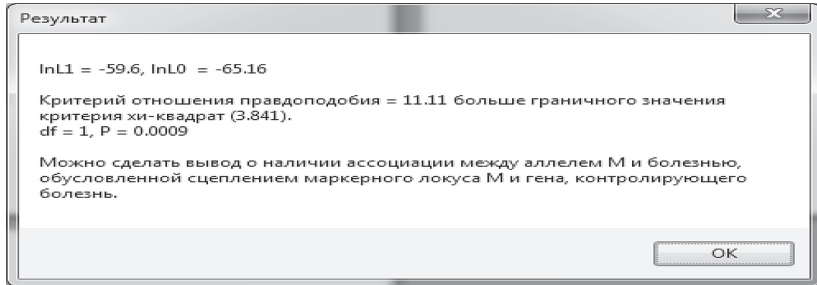


Рис. 11. Результат теста с помощью метода TDT при расчете критерия, основанного на отношении правдоподобия

В качестве примера возьмем выборку, состоящую из трех ядерных родословных, в которых не известны генотипы обоих родителей, но генотипировано по несколько потомков, среди которых встречаются как больные, так и здоровые. Эти данные были внесены в таблицу на форме S-TDT и выделены в ней розовым цветом. На рис. 12 показана форма S-TDT после ввода данных и расчета статистики.

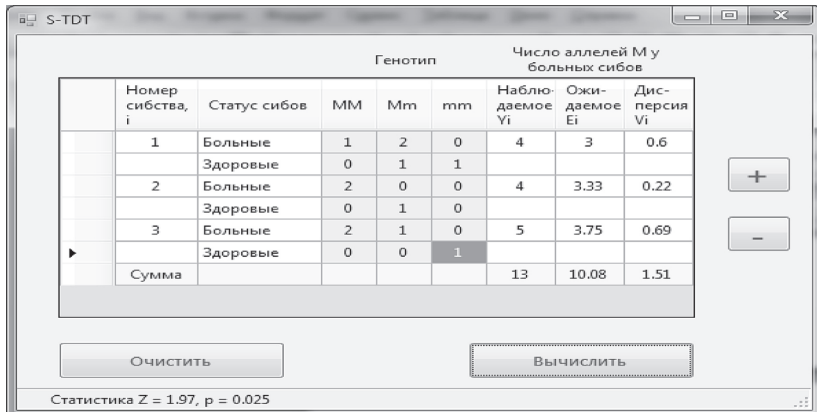


Рис. 12. Форма S-TDT после ввода данных и расчета статистики

Результат, полученный с помощью метода S-TDT, показан на рис. 13:

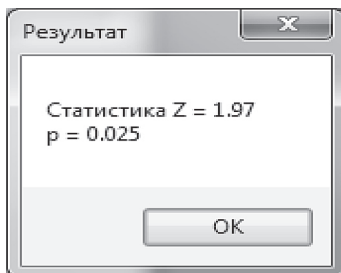


Рис. 13. Результат теста с помощью метода S-TDT

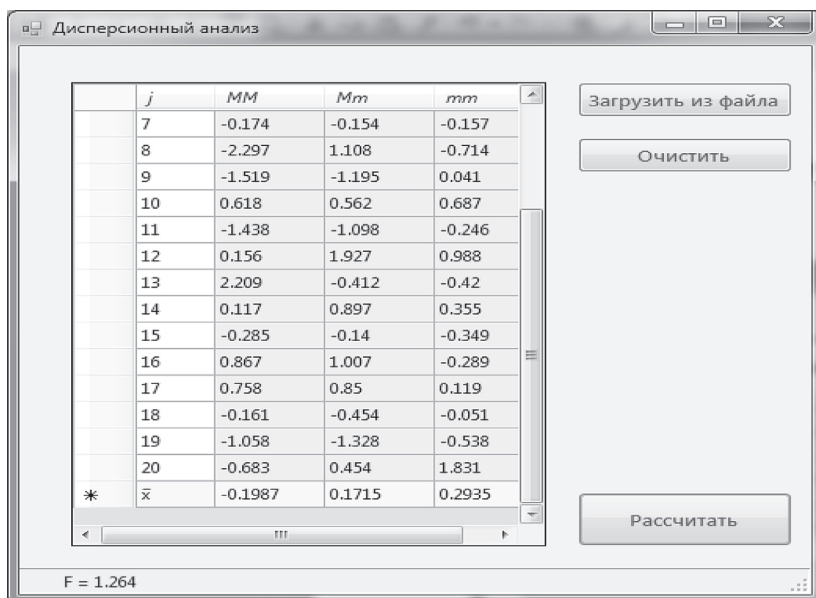


Рис. 14. Форма «Дисперсионный анализ» после ввода данных

Заполним таблицу на форме «Дисперсионный анализ» значениями количественного признака у особей с различными генотипами маркерного локуса.

На рис. 14 показана форма для ввода данных и анализа ассоциаций с помощью дисперсионного анализа после загрузки данных из файла и расчетов:

Результат дисперсионного анализа показан на рис. 15:

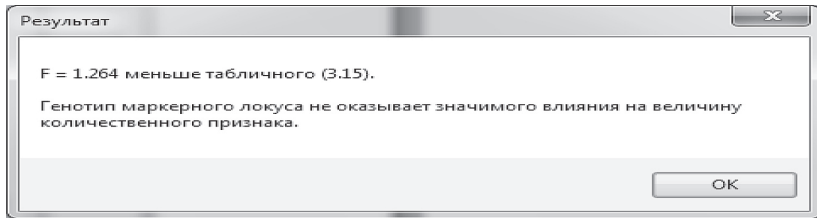


Рис. 15. Результат дисперсионного анализа

Для расчета равновесия Харди-Вайнберга возьмем выборку, состоящую из двухсот неродственных особей, которые были генотипизированы по десяти SNP маркерам. Результаты генотипирования представлены в таблице 3.

Таблица 3.

Пример данных для расчета равновесия Харди-Вайнберга

Номер маркера	Наблюдаемые численности генотипов			
	N_{MM}^0	N_{Mm}^0	N_{mm}^0	N_{NA}
1	114	80	6	0
2	56	111	32	1
3	168	29	3	0
4	167	26	7	0
5	60	96	44	0
6	95	90	14	1
7	51	110	39	0
8	118	48	22	12
9	121	60	19	0
10	79	95	25	1

На рис. 16 показана форма для расчета равновесия Харди-Вайнберга после загрузки данных из файла:

На рис. 17 показана эта же форма после расчета статистики. Полужирным шрифтом и розовым цветом в последнем столбце выделены отличия на уровне значимости $p < 0.5$. Это говорит о том, что численности генотипов трех из десяти маркеров не соответствуют ожидаемым при равновесии Харди-Вайнберга, и указывает на присутствие ошибок. Для устранения этих ошибок проводят коррекцию данных с последующим тестированием.

Тест на равновесие Харди-Вайнберга

Номер маркера	q	Наблюдаемые численности генотипов				Ожидаемые численности генотипов				N	χ^2	P при df = 1
		NO(MM)	NO(Mm)	NO(mm)	N(NA)	NE(MM)	NE(Mm)	NE(mm)	N			
1		114	80	6	0							
2		56	111	32	1							
3		168	29	3	0							
4		167	26	7	0							
5		60	96	44	0							
6		95	90	14	1							
7		51	110	39	0							
8		118	48	22	12							

Загрузить из файла Очистить Вычислить

Рис. 16. Форма «Тест на равновесие Харди-Вайнберга» после загрузки данных из файла

Тест на равновесие Харди-Вайнберга

Номер маркера	q	Наблюдаемые численности генотипов				Ожидаемые численности генотипов				N	χ^2	P при df = 1
		NO(MM)	NO(Mm)	NO(mm)	N(NA)	NE(MM)	NE(Mm)	NE(mm)	N			
2	0.44	56	111	32	1	62.41	98.07	38.53	199	3.47	0.0625	
3	0.088	168	29	3	0	166.35	32.1	1.55	200	1.67	0.1963	
4	0.1	167	26	7	0	162	36	2	200	15.43	0.0001	
5	0.46	60	96	44	0	58.32	99.36	42.32	200	0.23	0.6315	
6	0.296	95	90	14	1	98.63	82.94	17.44	199	1.41	0.2351	
7	0.47	51	110	39	0	56.18	99.64	44.18	200	2.16	0.1416	
8	0.245	118	48	22	12	107.16	69.55	11.28	188	17.96	0	
9	0.245	121	60	19	0	114	73.99	12	200	7.16	0.0075	

Загрузить из файла Очистить Вычислить

Рис. 17. Форма «Тест на равновесие Харди-Вайнберга» после расчета статистики

Перечисленные выше методы, а также тест Хайди-Вайнберга, применяемый для контроля генотипирования случайной выборки особей из популяции, были реализованы в виде программного приложения для генетического анализа признаков человека на языке C# в среде Visual Studio 2012. Созданная программа будет полезна специалистам в области генетики, и поможет гораздо быстрее проводить генетические тесты.

Список литературы

1. Аксенович Т.И., Белоногова Н.М. Картирование генов с помощью неравновесия по сцеплению или аллельных ассоциаций: учебное

- пособие / Новосибирский государственный ун-т. Новосибирск, 2008. 98 с.
2. Аксенович Т.И. Статистические методы генетического анализа признаков человека: учеб. пособие / Новосиб. гос. ун-т Новосибирск, 2003. 160 с.
 3. Трушина В.П., Пятницев Д.В. Мобильное приложение для принятия решений с помощью методов анализа ассоциаций // В мире научных открытий. 2015. № 8.1 (68). С. 377–384.

References

1. Akseovich T.I., Belonogova N.M. *Kartirovanie genov s pomoshch'yu neravnovesiya po stsepleniyu ili allel'nykh assotsiatsiy* [Mapping genes using linkage disequilibrium or allelic association]. Novosibirsk, 2008. 98 p.
2. Akseovich T.I. *Statisticheskie metody geneticheskogo analiza priznakov cheloveka* [Statistical methods for the genetic analysis of human traits]. Novosibirsk, 2003. 160 p.
3. Trushina V.P., Pyatnitsev D.V. *V mire nauchnykh otkrytiy*. 2015. № 8.1 (68), pp. 377–384.

ДАнные ОБ АВТОРАХ

Осипов А.Л., Трушина В.П., Чентаева Е.А.

Новосибирский государственный университет экономики и управления

ул. Каменская 52/1, г. Новосибирск, 630099, Российская Федерация

alosip@mail.ru

DATA ABOUT THE AUTHORS

Osipov A.L., Trushina V.P., Chentaeva E.A.

*Novosibirsk State University of Economics and Management
52/1, Kamenskaya Str., Novosibirsk, 630099, Russian Federation
alosip@mail.ru*