

Научная статья
УДК 519.6; 004.67

[https://doi.org/10.14258/izvasu\(2026\)1-11](https://doi.org/10.14258/izvasu(2026)1-11)

Сравнительный анализ методов оценки среднего причинного эффекта в регрессионном моделировании на основе синтетических выборок

*Анастасия Павловна Боронкина¹, Елена Владимировна Понькина²,
Салтанат Сайларбековна Адиканова³*

¹Алтайский государственный университет, Барнаул, Россия, fomenkoap@vc.asu.ru

²Алтайский государственный университет, Барнаул, Россия, ponkinaelena77@mail.ru

³Восточно-Казахстанский университет им. Сарсена Аманжолова,
Усть-Каменогорск, Казахстан, Ersal_7882@mail.ru

Original article

Comparative Analysis of Methods for Evaluation of the Average Causal Effect in Regression Modeling Based on Synthetic Samples

Anastasia P. Boronkina¹, Elena V. Ponkina², Saltanat S. Adikanova³

¹Altai State University, Barnaul, Russia, fomenkoap@vc.asu.ru

²Altai State University, Barnaul, Russia, ponkina@math.asu.ru

³Sarsen Amanzholov East Kazakhstan University, Ust-Kamenogorsk, Kazakhstan,
Ersal_7882@mail.ru

Аннотация. В работе рассматриваются результаты сравнительного анализа методов оценки причинного эффекта (ATE) на основе t-теста, линейной регрессии и техники матчинга данных (ближайший сосед, полное соответствие и подклассификация). Выполнена генерация синтетических выборок варьируемого объема от 250 до 1000, различных пропорций в распределении данных на экспериментальную и контрольную группы (50/50, 70/20, 90/10). Генерация данных осуществлена при предположении о нормальном, экспоненциальном и логарифмическом распределении предикторов. Для оценки точности метода используется метрика погрешности MAPE. Результаты численного эксперимента показали, что методы сравнения средних и линейной регрессии приводят к недооценке или переоценке ATE, а метод полного матчинга (Full Matching) обеспечивает более точные результаты независимо от формы распределения данных. Выявлено, что сбалансированность и больший объем выборок позволяют получить более точные оценки ATE. Полученные результаты могут использоваться для подбора и обоснования метода оценки причинных эффектов в случае

Abstract. The study focuses on comparative analysis of accuracy for various methods for evaluation of the average causal effect (ATE). A range of methods includes the t-test, linear regression, and several data matching techniques such as nearest neighbor, full matching, and subclassification. The methods are tested using the generated synthetic samples of varying dimensions from 250 to 1000, with different proportions of data divided into experimental and control groups (50/50, 70/20, and 90/10), under the assumption of normal, exponential, and logarithmic distribution of predictors. The Mean Absolute Percentage Error (MAPE) error metric is used to evaluate the accuracy of methods. The numerical experiments reveal that using t-test and linear regression techniques leads to a significant under- or overestimation of the ATE values, while the full matching methods produce more accurate results regardless of covariate probability distributions. Also, the balanced and larger samples provide the ATE evaluations with higher accuracy. The obtained results can be used to justify the proper method for evaluation of casual effects under the direct influence of the binary nature and nonlinearity of one of the affecting factors.

наличия прямого вмешательства бинарного характера и с учетом нелинейности действия одного из влияющих факторов.

Ключевые слова: вычисленный эксперимент, t-тест, линейная регрессия, матчинг, полное соответствие, подклассификация

Для цитирования: Боронкина А.П., Понькина Е.В., Адиканова С.С. Сравнительный анализ методов оценки среднего причинного эффекта в регрессионном моделировании на основе синтетических выборок // Известия Алтайского государственного университета. 2026. № 1 (147). С. 84–90. [https://doi.org/10.14258/izvasu\(2026\)1-11](https://doi.org/10.14258/izvasu(2026)1-11).

Введение

Фундаментальной основой исследования причинных эффектов в прикладных научных исследованиях является теория о потенциальных исходах (*Potential Outcome Theory*), предложенная в исследованиях Дж. Неймана и Д. Рубина [1]. Центральной проблемой теории потенциальных исходов является количественная оценка причинного эффекта некоторого воздействия или вмешательства на состояние объекта. Рассмотрим задачу, в которой некоторое вмешательство (*treatment* — Z) влияет на исход (*outcome* — Y), а результаты измерений представлены в виде набора данных, содержащих значения Y , Z , и прочих влияющих на результат факторов X или вектора ковариат. Рассматривается случай, когда вмешательство представляется в виде бинарной переменной — $Z=\{1, 0\}$, а Y , X определены на R .

Формально для оценки причинного эффекта для каждого объекта должны быть определены два потенциальных исхода — (1) исход $Y_1=Y(Z=1 | X)$, который наблюдается в случае, когда объект получил воздействие; (2) исход $Y_0=Y(Z=0 | X)$ в случае, если

$$ATE = E(Y_1) - E(Y_0) = E_{N_1}[Y(Z=1 | X)] - E_{N_2}[Y(Z=0 | X)],$$

где $E(\cdot)$ — математическое ожидание, N_1 , N_2 — объемы экспериментальной и контрольной групп, X — вектор признаков.

В 1986 г. П. Холланд [1, 4] показал, что средний эффект воздействия (ATE) может оцениваться со смещением ($Bias$) и не равен среднему истинному эффекту воздействия (*Average True Treatment Effect* — $ATTE$), измеряемому согласно теории потенциальных исходов:

$$ATTE = E_N(Y_1 - Y_0) = ATE + Bias.$$

Как показано в [4], причинами возникновения $Bias$ являются неоднородность выборки объектов, нарушение принципов рандомизации «назначения» воздействия в случае обзорных исследований, не-

Keywords: calculated experiment, t-test, linear regression, matching, full matching, subclassification

For citation: Boronkina A.P., Ponkina E.V., Adikanova S.S. Comparative Analysis of Methods for Evaluation of the Average Causal Effect in Regression Modeling Based on Synthetic Samples. *Izvestiya of Altai State University*. 2026. No 1 (147). P. 84–90. (In Russ.). [https://doi.org/10.14258/izvasu\(2026\)1-11](https://doi.org/10.14258/izvasu(2026)1-11).

объект не получил воздействие. Истинный причинный эффект воздействия для некоторого объекта i (*True Treatment Effect* — TTE_i) оценивается величиной разности между двумя потенциальными исходами [2, 3]:

$$TTE_i = Y_{1i} - Y_{0i} = Y(Z_i = 1 | X_i) - Y(Z_i = 0 | X_i).$$

TTE доступен к измерению в натуральных экспериментах, когда действительно для одного и того же по свойствам объекта можно получить два сопоставимых измерения Y_1 и Y_0 . По значениям индивидуальных TTE_i оценивается средний причинный эффект по всей группе (*Average Treatment Effect* — ATE).

В экспериментальных исследованиях, когда измерение обоих исходов для одного и того же объекта невозможно, объекты выборки делятся на экспериментальную (испытывающих вмешательство) и контрольную (не испытывающих вмешательство) группы случайным образом, оценка ATE осуществляется как:

сбалансированность данных как по признакам X , так и в распределении объектов между экспериментальной и контрольной группами — N_1 / N_2 .

Для решения задачи минимизации TTE_i используются различные подходы, в частности, регрессионный анализ и взвешивание (*probability weighting*), методы сопоставления (*matching*) [5], включая *Propensity Score Matching* [6], демонстрирующие различную точность приближения $ATTE$. Поэтому в современных исследованиях проблема выбора метода оценки ATE , обеспечивающего его минимальное смещение, является актуальной.

Проведение численного эксперимента для выборки с заданными статистическими свойствами и известной величиной ATE позволит получить оценки погрешностей методов и исследовать влияние как различий в про-

порциях между объемами экспериментальной и контрольной групп, так и форме распределения признаков X на численный результат. Одним из ключевых преимуществ подобных экспериментов является то, что ATE заранее известен, как и известна сама причинная модель, обеспечивающая генерацию исходов Y .

Целью работы является численная оценка погрешности методов расчета ATE по данным синтетических выборок при условии вмешательства бинарного типа, сгенерированных для различных пропорций экспериментальной и контрольной групп и различных вероятностных распределений ковариат X .

Материалы и методы

1. Генерация синтетической выборки

Для генерации рандомизированной синтетической выборки рассмотрим структуру причинной модели, которая в целом соответствует обсервационным наблюдениям, когда назначение вмешательства Z зависит от факторов X_1 и X_2 согласно логистической регрессионной зависимости. В совокупности эти факторы оказывают влияние на результат Y (рис. 1). Подобные конструкции широко встречаются в медицинских исследованиях, где в качестве X_1 и X_2 могут выступать возраст и вес пациента, например, а в качестве Z — прием препарата.

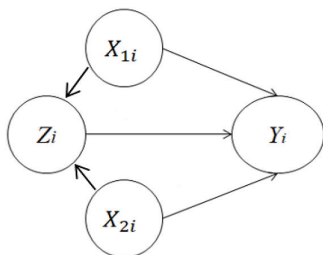


Рис. 1. Структура рассматриваемой причинной модели

Уравнение для генерации Y согласно заданной причинной модели имеет вид:

$$Y_i = a_1 \cdot X_{1i} + a_2 \cdot X_{2i} + 5 \cdot Z_i, \quad i = 1, \dots, N,$$

$$P(Z_i = 1 | X_i) = \frac{1}{1 - (b_1 \cdot X_{1i} + b_2 \cdot X_{2i})},$$

где Y_i — результирующая переменная (outcome); Z_i — бинарная переменная вмешательства (1 — объект испытывает вмешательство, 0 — вмешательство отсутствует); X_{1i}, X_{2i} — факторы или независимые случайные величины с заданными статистическими характеристиками, оказывающие аддитивный эффект на результирующую переменную.

Для генерации Z_i принималась во внимание необходимость оценки ATE для различных соотношений размерностей экспериментальной и контрольной групп (50/50; 70/30; 90/10). При сбалансированной выборке при распределении 50/50 полагалось $Z_i=1$, если $P(Z_i = 1 | X_i) > 0,5$, и 0 в противном случае. При соот-

ношении 70/30 генерация $Z_i=1$, если $P(Z_i = 1 | X_i) > 0,7$, и 0 в ином случае. При пропорции 90/10 $Z_i=1$, если $P(Z_i = 1 | X_i) > 0,9$, и 0 в противном случае. Параметры $b_1=0,05, b_2=-0,05$.

Следует подчеркнуть, что в рассматриваемой модели признак X_1 имеет нелинейный характер влияния на Y , таким образом, нелинейность вносит значительные искажения в оценки Y . Воздействие факторов X_{1i}, X_{2i} и переменной Z_i на результирующую переменную Y_i является равноценным по значимости, что достигается за счет балансировки параметров $a_1=a_2=0,05$. Таким образом, истинный средний эффект вмешательства (ATE) в синтетических данных равен 5.

Для проведения численного эксперимента в работе были сгенерированы синтетические выборки различных объемов 250, 500 и 1000 наблюдений. В каждой симуляции пропорция объектов, испытывающих вмешательство и невмешательство (N_1/N_2), варьировалась 50/50, 70/30, 90/10. Для генерации случайных величин X_1, X_2 использованы нормальное $N(5, 2.5)$, логистическое $log(5, 2.5)$ и экспоненциальное $exp(5)$ распределения. Чтобы исключить влияние конкретных значений выборки на систематическую ошибку и дисперсию, модель принимает во внимание воздействие внешних, не включенных в модель, факторов. Таким образом, используется генерация «шумов» для переменных X_1, X_2 в размере $s=0,1$ относительно среднего. Величина шума по Y относительно его среднего значения $p=0,1$.

Для каждого варианта расчета выполнялась сто- кратная генерация элементов выборки. Таким образом, в исследовании использовано 2700 синтетических выборок.

2. Методы оценки ATE

В ходе сравнительного анализа точности оценки ATE по набору синтетических выборок использованы следующие пять методов:

1. *t-тест Стьюдента* (t-test).
2. *Линейная регрессия* (LR).
3. *Матчинг объектов согласно методу ближайшего соседа* (Nearest Neighbor — NN) — основан на поиске соответствий по принципу 1:1 на базе некоторой заданной метрики [7]. Метод находит соответствия каждому объекту в экспериментальной группе, после чего отбрасывает объекты контрольной группы без пары, при этом каждый объект контрольной группы используется лишь раз.
4. *Метод полного соответствия* (Full Matching — FM) также находит пару каждому объекту, основываясь на метрике дистанции, но при его использовании рассчитывается эффект воздействия в каждой группе отдельно. Веса в группе взвешиваются по числу объектов в них [8]. Также при использовании данного метода можно формировать группы соответствий не только 1:1, а также и 1:k.
5. *Метод подклассификации* (Subclassification — Subclasses) выполняет классификацию оценок склон-

ности объектов на заданное количество подклассов. Метод формирует подклассы так, что подкласс будет иметь примерно одинаковое количество объектов экспериментальной и контрольной групп. Веса рассчитываются по числу объектов в каждом подклассе [9].

Для методов поиска ближайшего соседа и полного соответствия использовались два способа установления меры сходства между наблюдениями: расстояние Евклида [10] и Махаланобиса [11].

3. Оценка точности

Для определения точности методов оценки ATE вычислена средняя абсолютная ошибка в процентах (*Mean Absolute Percentage Error — MAPE*):

$$MAPE = \sum_{i=1}^N \left| \frac{ATTE - ATE}{ATTE} \right| \cdot \frac{1}{N}.$$

Показателя $MAPE$ достаточно для сравнительной оценки точности методов, поскольку генерированные синтетические выборки имеют идентичные статистические свойства. Все вычисления выполнены в среде R с использованием пакета *Matchit* [12].

Результаты

Результаты оценки точности ATE ($MAPE$, %) для различных распределений X_1, X_2 тестируемыми методами приведены в таблице.

Минимальные средние абсолютные ошибки оценки ATE по тестируемым методам

Распределение		N_1/N_2	Объем выборки		
			250	500	1000
$N(5, 2.5)$	Метод MAPE, %	50/50	$\frac{FM_M}{0,02}$	$\frac{FM_M}{0,02}$	$\frac{FM_M}{0,01}$
	Метод MAPE, %	70/30	$\frac{FM_M}{0,03}$	$\frac{NN_E}{0,03}$	$\frac{NN_E}{0,02}$
	Метод MAPE, %	90/10	$\frac{FM_M}{0,07}$	$\frac{FM_M}{0,06}$	$\frac{FM_M}{0,03}$
$\log(5, 2.5)$	Метод MAPE, %	50/50	$\frac{Subclass}{0,05}$	$\frac{Subclass}{0,02}$	$\frac{Subclass}{0,01}$
	Метод MAPE, %	70/30	$\frac{Subclass}{0,07}$	$\frac{Subclass}{0,04}$	$\frac{Subclass}{0,02}$
	Метод MAPE, %	90/10	$\frac{FM_E}{0,09}$	$\frac{FM_E}{0,06}$	$\frac{FM_E}{0,03}$
$\exp(5)$	Метод MAPE, %	50/50	$\frac{Subclass}{0,13}$	$\frac{Subclass}{0,11}$	$\frac{Subclass}{0,08}$
	Метод MAPE, %	70/30	$\frac{Subclass}{0,17}$	$\frac{Subclass}{0,14}$	$\frac{Subclass}{0,11}$
	Метод MAPE, %	90/10	$\frac{FM_E}{0,22}$	$\frac{FM_E}{0,22}$	$\frac{FM_E}{0,17}$

Источник: Расчеты авторов.

Численный эксперимент показал, что для тестируемого набора данных с нормальным распределением метод полного соответствия (*Full Matching — FM*) в сочетании с метрикой Махаланобиса обеспечивает наиболее точные оценки. В случае если распределение данных соответствует логарифмическому закону, наиболее точные оценки дает метод полного соответствия (*FM*), использующий меру расстояния Махаланобиса, и метод подклассификации (*Subclass*). Набор данных, подчиняющихся экспоненциальному закону распределения, также обеспечивает наиболее точные оценки методом подклассификации и методом полного соответствия (*FM*), но с метрикой Евклида. Стоит подчеркнуть, что результаты оценки, достигнутые методом матчинга с использованием метрики Махаланобиса, очень близки с результатами, полученными с применением евклидовой

метрики, хотя полное совпадение отсутствует. В отличие от евклидовой метрики, определяющей дистанцию объектов без учета взаимосвязей между признаками, метрика Махаланобиса принимает во внимание эти взаимосвязи, корректируя расстояния с учетом ковариационной структуры данных. Методы линейной регрессии и статистические оценки сравнения средних на основе t -теста дают сильно смещенные оценки, величины $MAPE$ значительны. Важно подчеркнуть, что включение квадратичного члена X_1 (т.е. X_1^2) в модель линейной регрессии могло бы повысить точность оценки метода.

Вычислительный эксперимент показал, что с увеличением размерности выборки повышается точность оценивания ATE (рис. 2). На графике представлены различия в погрешности оценки ATE для выборок наибольшего и наименьшего объема.

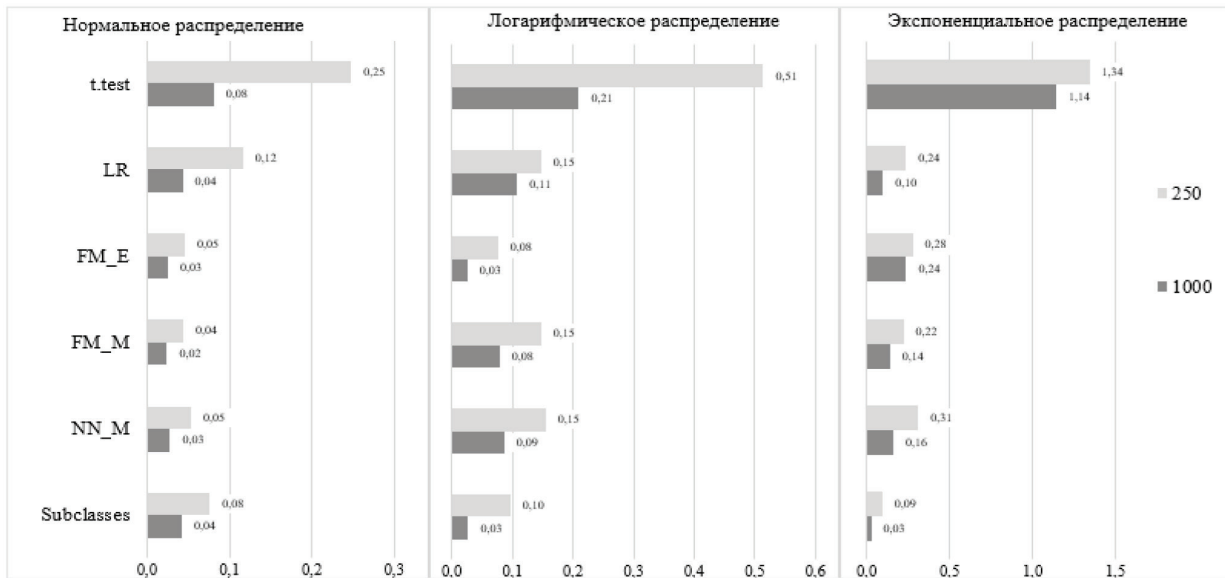


Рис. 2. Оценки MAPE (%) величины ATE для выборок разного объема в среднем для рассматриваемых пропорций вмешательства (50/50, 70/30, 90/10). Источник: Расчеты авторов

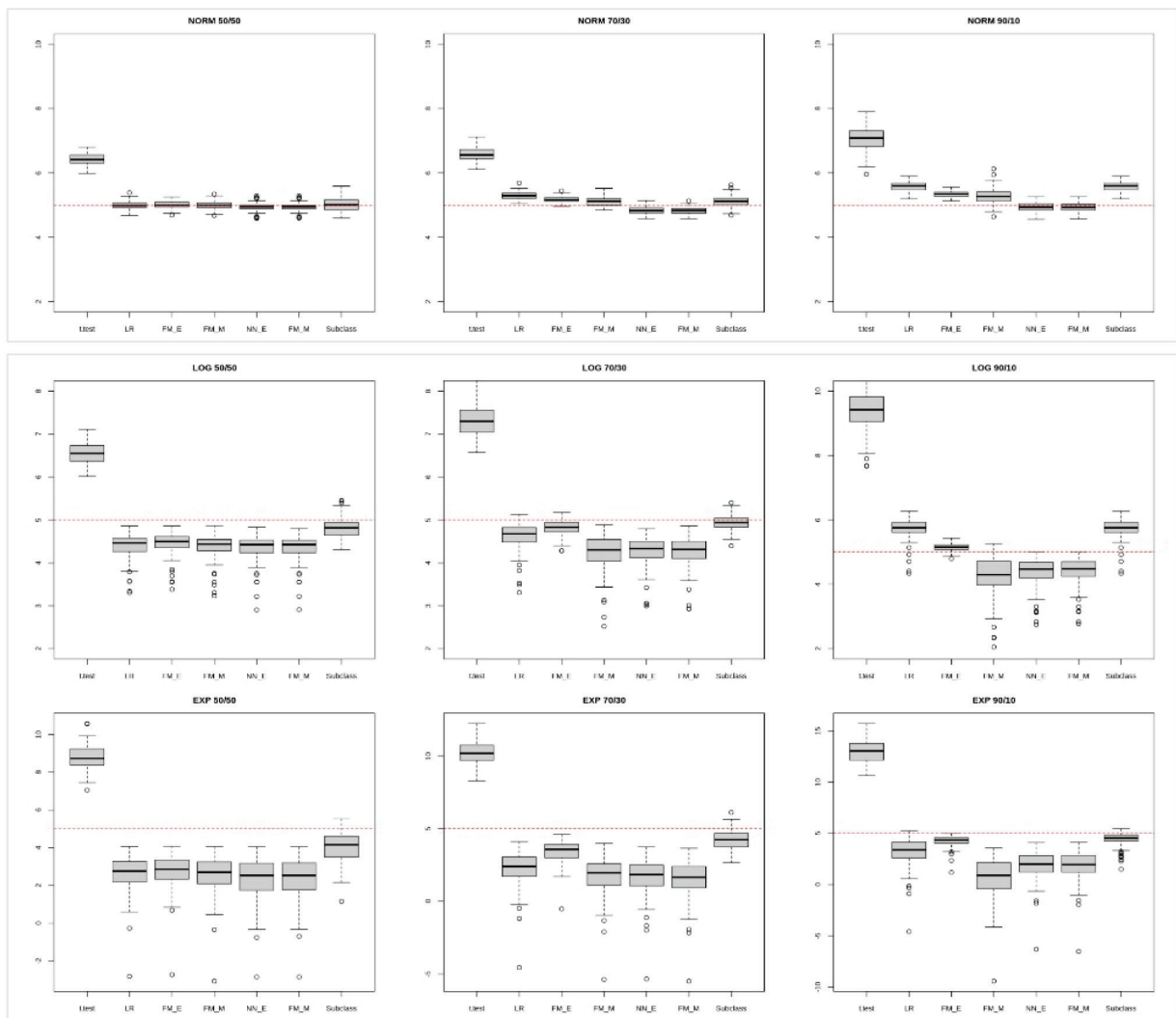


Рис. 3. Оценки ATE, полученные в выборке объемом 250 объектов. Пунктирной линией обозначено значение ATE=5. Источник: Расчеты авторов

Оценка влияния пропорции распределения данных на точность оценки различными методами показала, что нарушение сбалансированности данных смещением к предельным величинам (1 или 0) приводит к увеличению ошибки при определении среднего причинного эффекта (рис. 3). Видно, что на выборках малого объема оценки АТЕ могут значительно отклоняться от истинного значения.

Заключение

Использование метода математической статистики определяется тем, что каждый из методов имеет определенные границы применимости. Использование же методов вне границ их возможностей приводит часто к искаженным или даже ошибочным выводам. Оценка эффективности различных подходов к определению АТЕ, проведенная на искус-

ственно созданных данных с заданным распределением и нелинейной причинной зависимостью, выявила, что оценки, полученные с использованием методов полного соответствия и подклассификации, демонстрируют близкие значения к истинным и являются наиболее точными.

Перспективы развития метода полного соответствия связаны с его адаптацией к неполным данным и данным с малой размерностью. Разработка метода, позволяющего эффективно обрабатывать такие данные, позволит расширить область применения метода и решать более сложные задачи анализа данных. Кроме того, представляется интересным исследовать возможности применения метода в сочетании с другими методами машинного обучения, такими как нейронные сети и генетические алгоритмы.

Библиографический список

1. Rubin D.B. Teaching Statistical Inference for Causal Effects in Experiments and Observational Studies // *Journal of Educational and Behavioral Statistics*. 2004. № 29 (3). P. 343–367.
2. King G., Lucas C., Nielsen R.A. The Balance-Sample Size Frontier in Matching Methods for Causal Inference // *American journal of political science*. 2016. № 61 (2). P. 1–17.
3. Rubin D.B. Causal Inference Using Potential Outcomes // *Journal of the American Statistical Association*. 2011. Vol. 469. No 100. P. 322–331.
4. Holland P.W. Statistics and Causal Inference // *Journal of the American Statistical Association*. 1986. T. 81. № 396. P. 945–960.
5. Френкель А.А., Сурков А.А. Определение весовых коэффициентов при объединении прогнозов // *Вопросы статистики*. 2017. № 12. С. 3–15.
6. Rosenbaum P.R. The Central Role of the Propensity Score in Observational Studies for Causal Effects // *Biometrika*. 1983. Vol. 70. № 1. P. 41–55.
7. Rubin D.B. Matching to Remove Bias in Observational Studies // *Biometrics*. 1967. № 29. P. 159–184.
8. Ming K., Rosenbaum P.R. Substantial Gains in Bias Reduction from Matching with a Variable Number of Controls // *Biometrics*. 2000. № 56. P. 118–124.
9. Rosenbaum P.R., Rubin D.B. Reducing Bias in Observational Studies Using Subclassification on the Propensity Score // *Journal of the American Statistical Association*. 1984. № 79. P. 516–524.
10. Abdi M.A. Novel Approach for Off-Line Arabic Writer Identification Based on Stroke Feature Combination // *24th International Symposium on Computer Information Science, IEEE*. 2009. P. 597–600.
11. Diamond A. Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies // *Review of Economics and Statistics*. 2013. Vol. 95. № 3. P. 932–945.
12. Ho D.E., Imai K., King G., Stuart E.A. Matchit: Non-parametric Preprocessing for Parametric Causal Inference. Software for Using Matching Methods in R // *Journal of Statistical Software*. 2011. № 42 (8) P. 1–28.

References

1. Rubin D.B. Teaching Statistical Inference for Causal Effects in Experiments and Observational Studies. *Journal of Educational and Behavioral Statistics*. 2004. No 29 (3). P. 343–367.
2. King G., Lucas C., Nielsen R.A. The Balance-Sample Size Frontier in Matching Methods for Causal Inference. *American Journal of Political Science*. 2016. № 61(2). P. 1–17.
3. Rubin D.B. Causal Inference Using Potential Outcomes. *Journal of the American Statistical Association*. 2011. Vol. 469. No 100. 322–331.
4. Holland P.W. Statistics and Causal Inference. *Journal of the American Statistical Association*. 1986. T. 81. № 396. P. 945–960.
5. Frenkel A.A., Surkov A.A. Determination of Weighting Coefficients When Combining Forecasts. *Voprosy Statistiki*. 2017. No 12. P. 3–15. (In Russ.).
6. Rosenbaum P.R. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*. 1983. Vol. 70. № 1. P. 41–55.
7. Rubin D.B. Matching to Remove Bias in Observational Studies. *Biometrics*. 1967. № 29. P. 159–184.
8. Ming K., Rosenbaum P.R. Substantial Gains in Bias Reduction from Matching with a Variable Number of Controls. *Biometrics*. 2000. № 56. P. 118–124.

9. Rosenbaum P.R., Rubin D.B. Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *Journal of the American Statistical Association*. 1984. № 79. P. 516–524.

10. Abdi M.A. Novel Approach for Off-Line Arabic Writer Identification Based on Stroke Feature Combination. *24th International Symposium on Computer Information Science, IEEE*. 2009. P. 597–600.

11. Diamond A. Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. *Review of Economics and Statistics*. 2013. Vol. 95. № 3. P. 932–945.

12. Ho D.E., Imai K., King G., Stuart E.A. Matchit: Non-parametric Preprocessing for Parametric Causal Inference. Software for Using Matching Methods in R. *Journal of Statistical Software*. 2011. № 42 (8) P. 1–28.

Информация об авторах

А.П. Боронкина, аспирантка Института математики и информационных технологий, Алтайский государственный университет, Барнаул, Россия;

Е.В. Понькина, кандидат технических наук, доцент, заведующая кафедрой теоретической кибернетики и прикладной математики, Алтайский государственный университет, Барнаул, Россия;

С.С. Адиканова, PhD, декан высшей школы ИТ и естественных наук, Восточно-Казахстанский университет им. Сарсена Аманжолова, Усть-Каменогорск, Казахстан.

Information about the authors

A.P. Boronkina, Postgraduate Student of the Institute of Mathematics and Information Technology, Altai State University, Barnaul, Russia;

E.V. Ponkina, Candidate of Sciences in Technology, Associate Professor, Head of the Department of Theoretical Cybernetics and Applied Mathematics, Altai State University, Barnaul, Russia;

S.S. Adikanova, PhD, Dean of the Higher School of IT and Natural Sciences, Sarsen Amanzholov East Kazakhstan University, Ust-Kamenogorsk, Kazakhstan.