

Торайғыров университетінің
ҒЫЛЫМИ ЖУРНАЛ

НАУЧНЫЙ ЖУРНАЛ
Торайғыров университета

ТОРАЙҒЫРОВ УНИВЕРСИТЕТІНІҢ ХАБАРШЫСЫ

Физика, математика және компьютерлік
ғылымдар сериясы
1997 жылдан бастап шығады



ВЕСТНИК ТОРАЙҒЫРОВ УНИВЕРСИТЕТА

Серия: Физика, математика
и компьютерные науки
Издается с 1997 года

ISSN 2959-068X

№ 2 (2023)
Павлодар

**НАУЧНЫЙ ЖУРНАЛ
ТОРАЙГЫРОВ УНИВЕРСИТЕТА**

Серия: Физика, математика и компьютерные науки
выходит 4 раза в год

СВИДЕТЕЛЬСТВО

о постановке на переучет периодического печатного издания,
информационного агентства и сетевого издания
№ KZ91VPY00046988

выдано

Министерством информации и общественного развития
Республики Казахстан

Тематическая направленность

публикация материалов в области физики, математики,
механики и информатики

Подписной индекс – 76208

<https://doi.org/10.48081/GGBG9646>

Бас редакторы – главный редактор

Тлеукинов С. К., *д.ф.-м.н., профессор*

Заместитель главного редактора Искулов Н. А., *к.ф.-м.н., профессор*

Ответственный секретарь Жумабеков А. Ж., *PhD доктор*

Редакция алкасы – Редакционная коллегия

Esref Adali,	<i>PhD доктор, профессор (Турция);</i>
Abdul Qadir Rahimoon,	<i>PhD доктор, профессор (Пакистан);</i>
Донбаев К. М.,	<i>д.ф.-м.н., профессор;</i>
Демкин В. П.,	<i>д.ф.-м.н., профессор (Российская Федерация);</i>
Жумадиллаева А. К.,	<i>к.т.н., профессор;</i>
Ибраев Н. Х.,	<i>д.ф.-м.н., профессор;</i>
Косов В. Н.,	<i>д.ф.-м.н., профессор;</i>
Сенцова С. М.,	<i>д.пед.н., профессор;</i>
Шоканов А. К.,	<i>д.ф.-м.н., профессор</i>
Омарова А. Р.,	<i>технический редактор</i>

За достоверность материалов и рекламы ответственность несут авторы и рекламодатели

Редакция оставляет за собой право на отклонение материалов

При использовании материалов журнала ссылка на «Вестник Торайгыров
университета» обязательна

© Торайгыров университет

СЕКЦИЯ «КОМПЬЮТЕРНЫЕ НАУКИ»

FTAMP 20.53.19

<https://doi.org/10.48081/SKNV4152>

***Ж. С. Алимова¹, Н. К. Токжигитова²,
М. А. Токкожина³, Ж. Айсаұлы⁴**

^{1,2,4}Торайғыров университеті, Қазақстан Республикасы, Павлодар қ.

³Алматы мемлекеттік политехникалық колледжі,
Қазақстан Республикасы, Алматы қ.

*e-mail: jarasovajanar@mail.ru

ДЕРЕКТЕРДІ ТАЛДАУДЫҢ СИПАТТАМАЛЫҚ СТАТИСТИКА ӘДІСІ МЕН МОДЕЛІ

Бұл мақалада, деректерге сипаттамалық талдау әдістерін қолдана отырып, зерттеу нәтижелерін ұсынуда орын алатын қателіктерге көңіл аудару арқылы, бұл қателіктерді болдырмау үшін қандай ережелерге сүйенген жөн? – деген сұрақ төңірегінде зерттеулер жүргізілді. Нәтижесінде, деректерді талдаудың сипаттамалық әдісімен талдау жасаудың концептуалды моделі ұсынылды.

Жалпы алғанда, деректерге сипаттамалық талдау жасауда қарапайым математикалық формулалар қолданылғанымен, әртүрлі деректер үшін, олардан алынатын мақсатты нәтижелерге байланысты бұл формулалар бәріне бірдей қолданылмайды. Сондықтан, мақсатты нәтижелер анықталғаннан кейін, сол мақсатқа жету үшін қажетті ең тиімді жолдарды таңдау қажеттігі туындайды. Осыған орай, бұл мақалада, орта мәннің орнына – медиана және мода, стандартты ауытқуды қолдану, ал өзгерістердің динамикасын Стюденттің t-критеріі арқылы көрсету ұсынылып отыр.

Сондай-ақ, ұсынылып отырған модель бойынша, сипаттамалық статистика әдісімен талдау жасау, өз ішінде деректер қорын қамтитын, үш блоктан тұрады: қандай да бір көрсеткішінің статистикалық мәліметтерін жинақтайтын – бастапқы деректер қоры; деректер массиві мен бағдарламалық құралдар арқылы математикалық өңдеулердің нәтижелері орналасқан –

аралық деректер қоры; офис пакеттері – қорытынды деректер қоры.

Мақалада, 6B06101-Информатика мамандығының түлегі Ж.Айсаұлының бітіру жұмысында алынған нәтижелер қолданылды.

Кілтті сөздер: деректерді талдау, сипаттамалық статистика, талдау әдісі, сипаттамалық талдау, деректер қоры.

Кіріспе

Қазіргі таңда, деректерді статистикалық және математикалық талдау және есептеулер компьютерде арнайы бағдарламалардың көмегімен жүзеге асырылатыны белгілі. Сондай-ақ, зерттеулердің нәтижелеріне сүйенсек, деректерге талдау жасауға арналған бағдарламалық құралдар, негізінен, деректердің көптеген көздерінен әртүрлі форматтағы ақпараттарды жүктеуге, және осы деректер массивіне математикалық өңдеу әдістерінің көптеген түрлерін қолдана отырып, талдаудың – ұйымдастыру, талдау, визуализация, презентация әдістерімен нақты зерттеу деректерін талдауға мүмкіндік беретіні анықталды. Яғни, математикалық және статистикалық талдау кезінде сипаттамалық нәтижелерді алуға толық мүмкіндік береді [1].

Дегенмен, деректерге сипаттамалық статистика әдістерін қолдана отырып, зерттеу нәтижелерін ұсынуда біршама қателіктер орын алады: зерттеу нәтижелерін орташа мәндер бойынша ұсыну; деректерді % түрінде көрсету; өзгерістердің динамикасы да % немесе орташа көрсетіледі [2]. Осындай қателіктермен берілген талдау нәтижелері қате тұжырымдарға әкеледі. Бұл қателіктерді болдырмау үшін қандай ережелерге сүйенген жөн? Осы сұрақты қарастыру арқылы, бұл мақалада, деректерді талдаудың сипаттамалық әдісі мен деректерге қарапайым сипаттамалық статистика әдісімен талдау жасаудың концептуалды моделі ұсынылады.

Деректерді зерттеу нәтижелері ретінде ұсынылатын ақпараттарды талдауда жиі қолданылатын математикалық әдістер келесілер екені бәріне мәлім: сипаттамалық статистика (негізгі статистикалық әдістер, мәліметтердің қалыпты таралуын тексеру); Корреляциялық талдау; Регрессиялық талдау; Факторлық талдау. Екінші жағынан, үлкен массивті деректерді талдаудың әдістерін қарастыратын болсақ, оған мыналар жатады: Сипаттамалық статистика; Деректерді терең талдау, сыныптау; Краудсорсинг; Сплит-тестілеу; Болжау; Машиналық оқыту; Желілік белсенділікті талдау. Big data әдісі [3].

Материалдар мен әдістері

Сипаттамалық талдау қарапайым екі статистикалық көрсеткіштерге – орта мәні және стандартты ауытқуына негізделген. Ал, орта мәндерді

салыстыру әдістерін айырмашылықтардың статистикалық талдау әдісі десек те болады.

Орташа шамалар мынадай екі үлкен класқа бөлінеді:

- а) дәрежелік орташа шамалар;
- б) құрылымдық орташа шамалар.

Құрылымдық орташа шамаларға мода мен медиана жатады. Оларды вариациялық қатарларды сипаттау үшін қолданады. Дәрежелік орташа шамалардың мынадай түрлері болады:

- а) арифметикалық орташа шама;
- б) гармониялық (үйлесімді) орташа шама;
- в) квадраттық орташа шама;
- г) геометриялық орташа шама.

Сонымен, сипаттамалық статистиканы есептеу – бұл объектілерді диагностикалаудың дайын нәтижелеріне қолданылуы керек статистикалық өңдеудің алғашқы әдісі [4].

Сипаттамалық статистика эмпирикалық зерттеу қолданылатын педагогика мен психологиядағы, экономикадағы барлық зерттеу жұмыстарында есептеледі [5]. Негізгі сипаттамалық статистика екі топқа бөлінеді – орта тенденция өлшемдері және өзгергіштік өлшемдері.

Әдепкі жағдайларда орта тенденцияның өлшемі ретінде – орта мән, ал өзгергіштіктің негізгі өлшемі ретінде – стандартты ауытқу болып табылады. Сәйкесінше, әлеуметтік салаларда сипаттамалық талдаулар жасалғанда, осы екі қарапайым параметрлерді статистикалық көрсеткіш ретінде алады – орта мән және стандартты ауытқу.

Орта мән дегеніміз – бұл бізге мектеп курсынан белгілі қарапайым арифметикалық орташа (1) [2]. Орта мән, стандартты жағдайларда орта тенденцияның өлшемі бола алады. Ал, стандартты емес жағдайларда, орта тенденцияның өлшемі ретінде басқа мәндер – медиана және мода алынады.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

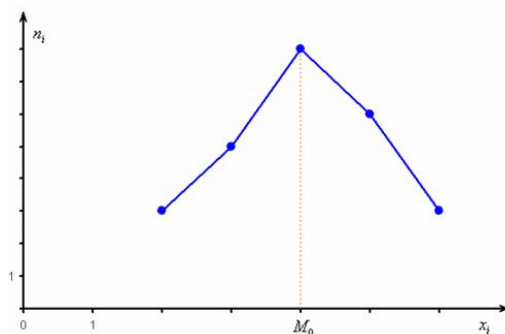
Формула 1 – Орташа арифметикалық мән

Стандартты ауытқуды (2) әртүрлі атайды: орта квадратты ауытқу, квадратты ауытқу және т.б, бірақ бұлар бәрі бір мағынаны білдіреді [2].

$$\delta = \frac{1}{n} \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2)$$

Формула 2 – Стандартты ауытқу (мұндағы x_i және i – выборкадағы элементтер, n – выборканың мөлшері, $\bar{x}$ – орташа арифметикалық мәні)

Дискретті вариациялық қатардың модасы M_0 – бұл максималды жиіліктегі варианта (1). Графикте түсініктірек болады – ең биік нүтесінің абсциссасы (1-сурет):



Сурет 1 – Дискретті вариациялық қатардың модасы M_0 графигі

Вариациялық қатардың медианасы – бұл деректердің жалпы санын (варианттардың саны бойынша) тең екіге бөлетін мән (2-сурет). Сандар қатары берілсе, оларды өсуі немесе кемуі бойынша реттейміз де, осы реттелген қатардың ортасын табамыз. Яғни, осы медиананың оң жағында және сол жағында элементтердің бірдей саны қалады (мысалы, медиананың оң жағында 4 сан болса, сол жағында да 4 сан бар).



Сурет 2 – Мода, медиана, орташа мән графигі

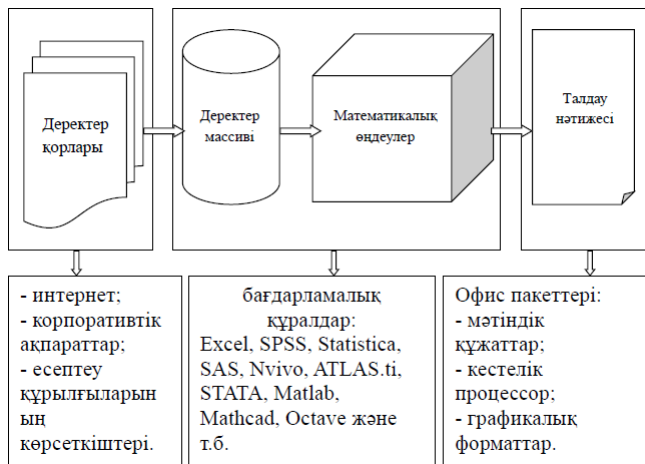
Осылайша, зерттеу нәтижелерін ұсынуға кететін қателіктерді болдырмау үшін, мына ережелерді ұстанған жөн [1]:

- а) орта мәннің орнына Медиананы (μ) және стандартты ауытқуды (δ) қолдану;
- ә) өзгерістердің динамикасын % немесе орташа мәнмен көрсетпей, оның орнына Стьюденттің t-критерийін қолдану;
- б) нәтижелерді жиілік талдау;
- в) зерттеліп отырған көрсеткіштердің арасындағы өзара байланысты орнату.

Алдағы уақытта бұл ережелерді, деректерге сипаттамалық статистика әдісімен талдау жасаудың концептуалды моделіндегі «Математикалық талдаулар» блогында ескеру қажет.

Сонымен, зерттеулерге сүйенсек [1], деректерге қарапайым сипаттамалық статистика әдісімен талдау, зерттелетін объектінің қандай да бір көрсеткішінің статистикалық мәліметтерін жинақтаудан басталады. Одан кейін, объектіге талдау жасайтын бағдарламалық құралдар таңдалады. Осы бағдарламалық құралдар арқылы математикалық өңдеулер жүргізіледі. Талдау нәтижесі офис пакеттеріне: мәтіндік құжаттар; кестелік процессор; графикалық форматтарға жүктеп алынады [6].

Осыған сәйкес, деректерге қарапайым сипаттамалық статистика әдісімен талдау жасаудың концептуалды моделі 3-суретте ұсынылады.



Сурет 3 – Деректерге қарапайым сипаттамалық статистика әдісімен талдау жасаудың концептуалды моделі

Ұсынылып отырған модельге сүйенсек, деректерді сипаттамалық талдау жасаудың моделі 3 блокты қамтиды:

1-ші блокта интернет, корпоративтік ақпараттар, есептеу құрылғыларының көрсеткіштерінен алынған деректер қорларынан – бастапқы деректер қоры қалыптастырылады [7–10].

2-ші блокта бағдарламалық құралдың көмегімен математикалық өңдеулер жүргізіледі. Бұл блок өз ішінде бағдарламалық құралдың ерекшеліктерін ескере отырып ұйымдастырылған деректер массиві мен Excel, SPSS, Statistica, SAS, Nvivo, ATLAS.ti, STATA, Matlab, Mathcad, Octave және т.б. сияқты бағдарламалық құралдар арқылы математикалық өңдеулердің нәтижелері орналасқан – аралық деректер қорын қамтиды [7–10].

3-ші блокта офис пакеттері: мәтіндік құжаттар; кестелік процессор; графикалық форматтарға жүктеп алынған талдау нәтижесі – қорытынды мәліметтерді құрайды.

Нәтижелер және талқылау

Жалпы алғанда, деректерге сипаттамалық талдау жасауда қарапайым математикалық формулалар қолданылғанымен, әртүрлі деректер үшін, олардан алынатын мақсатты нәтижелерге байланысты, бұл математикалық формулалар бәріне бірдей қолданылмайды. Сондықтан, мақсатты нәтижелер анықталғаннан кейін, сол мақсатқа жету үшін қажетті ең тиімді жолдарды таңдау қажеттігі туындайды. Осыған орай, бұл мақалада, орта мәннің орнына

– медиана және мода, стандартты ауытқуды қолдану, ал өзгерістердің динамикасын Стьюденттің t-критеріі арқылы көрсету ұсынылып отыр.

Сондай-ақ, ұсынылып отырған модель бойынша, сипаттамалық статистика әдісімен талдау жасау, өз ішінде деректер қорын қамтитын, үш блоктан тұрады:

1) Деректерге қарапайым сипаттамалық статистика әдісімен талдау, зерттелетін объектінің қандай да бір көрсеткішінің статистикалық мәліметтерін жинақтаудан басталады – бастапқы деректер қоры.

2) Бұл модельдің келесі қадамы, объектіге талдау жасайтын бағдарламалық құралдармен жұмыс істеуге негізделген. Бұл блок өз ішінде бағдарламалық құралдың ерекшеліктерін ескере отырып ұйымдастырылған деректер массиві мен бағдарламалық құралдар арқылы математикалық өңдеулердің нәтижелері орналасқан – аралық деректер қорын қамтиды. Деректермен жұмыс істеу үшін, алдымен оларды талдауға ыңғайлы түрге келтіріп реттеу керек.

3) Келесі блокта офис пакеттері: мәтіндік құжаттар; кестелік процессор; графикалық форматтарға жүктеп алынған талдау нәтижесі – қорытынды мәліметтерді құрайды.

Бұл нәтижелер Қазақстан Республикасы Ғылым және жоғары білім министрлігі Ғылым комитетінің ИРН АР14972847 «Деректердегі айқын емес байланыстарды анықтау мен терең талдауды жүзеге асыратын алгоритмі мен компьютерлік бағдарламасын әзірлеу» мемлекеттік бюджеттік жобасын орындау барысында, студенттерді осы тақырыптың төңірегінде дипломдық жұмыс жазуға тарту мақсатында, В06101-Информатика мамандығының түлегі Ж. Айсаұлының бітіру жұмысында шолу ретінде қарастырылды.

Қорытынды

Қорыта айтқанда, мақалада, деректерге сипаттамалық статистика әдістерін қолдана отырып, зерттеу нәтижелерін ұсынуда орын алатын қателіктерге көңіл аудару арқылы, бұл қателіктерді болдырмау үшін қандай ережелерге сүйенген жөн? – деген сұрақ төңірегінде зерттеулер жүргізілді. Нәтижесінде, деректерді талдаудың сипаттамалық әдісі мен осы таңдап алынған сипаттамалық талдаудың қарапайым моделін негізге ала отырып, деректерге сипаттамалық статистика әдісімен талдау жасаудың концептуалды моделі ұсынылды.

ПАЙДАЛАНЫЛҒАН ДЕРЕКТЕР ТІЗІМІ

1 **Алимова, Ж. С.** Деректерді талдаудың бағдарламалық құралдарына шолу [Текст] // ТоУ Хабаршысы. – 2022. – № 4.

2 **Абилдаева, Г. Б.** Дискреттік математика : Электрондық оқулық. – Қарағанды : ҚарMTU, 2017.

3 **Smolan, R., Erwit, J.** The Human Face of Big Data [Text]. – Publisher: Against All Odds Productions, 2012. – P. 224.

4 **Pang Ning Tan, V., Steinbach, M.** Introduction to Data Mining. – 2nd Edition. [Text]. – Publisher : AK.R, 2015. – P. 152.

5 **Кричевец, А. Н., Корнеев, А. А., Рассказова, Е. И.** Основы статистики для психологов [Текст]. – М. : Акрополь, 2019. – 286 с.

6 **Гунарто, Х.** Әлеуметтік зерттеулерге арналған параметрлік және параметрлік емес деректерді талдау : IBM SPSS [Текст]. – LAP академиялық баспа. – ISBN 978-6200118721. – 2019.

7 <https://data.egov.kz/> [Электронды ресурс].

8 <https://www.nationalbank.kz/ru>. [Электронды ресурс].

9 <https://data.worldbank.org> [Электронды ресурс].

10 <https://stats.oecd.org/> [Электронды ресурс].

REFERENCES

1 **Alimova, Zh.** Dereakterdi taldaudy`n baғdarlamaly`q qўraldary`na sholu [An overview of data analysis software] [Text] // Bulletin of ToU. – 2022. – № 4.

2 **Abildaeva, G. B.** Diskreterik matematika [Discrete mathematics] : Electronic textbook [Text] – Karagandy : KarMTU, 2017.

3 **Smolan, R., Erwit, J.** «The Human Face of Big Data» [Text]. – Publisher : Against All Odds Productions, 2012. – p. 224.

4 **Pang Ning Tan, V., Steinbach, M.** «Introduction to Data Mining»-2nd-Edition [Text]. – Publisher : AK.R, 2015.– P. 152.

5 **Krichevets, A. N., Korneev, A. A., Rasskazova, E. I.** Osnovy` statistiki dlya psixologov. [Fundamentals of statistics for psychologists.] [Text]. – Moscow : Akropol`, 2019. – 286 p.

6 **Gunarto, X.** Әлеуметтік зерттеулерге арналған параметрлік және параметрлік емес деректерді талдау : IBM SPSS [Parametric and Nonparametric Data Analysis for Social Research : IBM SPSS] [Text]. – LAP akademiya`q baspa. ISBN 978-6200118721. – 2019.

7 <https://data.egov.kz/> [Electronic resource].

8 <https://www.nationalbank.kz/ru> [Electronic resource].

9 <https://data.worldbank.org> [Electronic resource].

10 <https://stats.oecd.org/> [Electronic resource].

Материал 16.05.23 баспаға түсті.

***Ж. С. Алимова¹, Н. К. Токжигитова², М. А. Токкожина³, Ж. Айсаулы⁴**

^{1,2,4}Торайгыров университет, Республика Казахстан, г. Павлодар

³Алматинский государственный политехнический колледж,
Республика Казахстан, г. Алматы.

Материал поступил в редакцию 16.05.23.

МЕТОД И МОДЕЛЬ ОПИСАТЕЛЬНОЙ СТАТИСТИКИ АНАЛИЗА ДАННЫХ

В этой статье, сосредоточив внимание на ошибках, возникающих при представлении результатов исследования с использованием методов описательного анализа данных, были проведены исследования, посвященные вопросу о том, на каких правилах следует полагаться, чтобы избежать этих ошибок. В результате предложена концептуальная модель анализа данных описательным методом анализа.

В целом, хотя при описательном анализе данных используются простые математические формулы, для различных данных эти формулы применимы не ко всем из-за целевых результатов, полученных из них. Поэтому после определения целевых результатов возникает необходимость выбора наиболее эффективных путей, необходимых для достижения этой цели. В этой статье предлагается использовать вместо среднего значения – медиана и мода и стандартное отклонение, а динамику изменений показать с помощью Т-критерия Стьюдента.

Также по предлагаемой модели анализ методом описательной статистики состоит из трех блоков, включающих в себя базу данных: исходную базу данных, которая суммирует статистические данные того или иного показателя; промежуточную базу данных, в которой расположены результаты математической обработки с помощью массива данных и программных средств; офисные пакеты – итоговую базу данных.

В статье использованы результаты, полученные в выпускной работе выпускника специальности 6В06101-Информатика Ж. Айсаулы.

Ключевые слова: анализ данных, описательная статистика, метод анализа, описательный анализ, база данных.

***Zh. S. Alimova¹, N. K. Tokzhigitova², M. A. Tokkozhina³, Zh. Aisaully⁴**

^{1,2,4}Toraighyrov University, Republic of Kazakhstan, Pavlodar;

³Almaty State Polytechnic College, Republic of Kazakhstan, Almaty.

Material received on 16.05.23.

DESCRIPTIVE STATISTICS METHOD AND MODEL OF DATA ANALYSIS

In this article, by focusing on the errors that occur in the presentation of research results using descriptive data analysis methods, research was carried out around the question of what rules should be followed to avoid these errors. As a result, a conceptual model of analysis by the descriptive method of data analysis was proposed.

In general, although simple mathematical formulas are used in descriptive data analysis, for different data, depending on the target results obtained from them, these formulas are not applicable to everyone. Therefore, once the target results are identified, it becomes necessary to choose the most effective ways necessary to achieve that goal. In this regard, in this article, instead of the average value – the median and Moda, it is proposed to use the standard deviation, and to show the dynamics of changes using the T-criterion of the Steudent.

Also, according to the proposed model, analysis by the method of descriptive statistics consists of three blocks: primary database, which summarizes statistical data of some indicator; intermediate database, where the data array and the results of mathematical processing by software tools are located; Office packages – final database.

In the article, the results obtained in the final work of a graduate of the specialty 6B06101-Computer Science Zh. Aisaully were used.

Keywords: data analysis, descriptive statistics, analysis method, descriptive analysis, database.

Теруге 16.05.2023 ж. жіберілді. Басуға 31.05.2023 ж. қол қойылды.

Электрондық баспа

7,50 Mb RAM

Шартты баспа табағы 10,01. Таралымы 300 дана. Бағасы келісім бойынша.

Компьютерде беттеген: Е. Е. Калихан

Корректор: А. Р. Омарова, Д. А. Кожас

Тапсырыс № 4081

Сдано в набор 16.05.2023 г. Подписано в печать 31.05.2023 г.

Электронное издание

7,50 Mb RAM

Усл.печ.л. 10,01. Тираж 300 экз. Цена договорная.

Компьютерная верстка Е. Е. Калихан

Корректор: А. Р. Омарова, Д. А. Кожас

Заказ № 4081

«Toraighyrov University» баспасынан басылып шығарылған

«Торайғыров университеті» КЕ АҚ

140008, Павлодар қ., Ломов к., 64, 137 каб.

«Toraighyrov University» баспасы

«Торайғыров университеті» КЕ АҚ

140008, Павлодар қ., Ломов к., 64, 137 каб.

+7(718)267-36-69

e-mail: kereku@tou.edu.kz

www.vestnik.tou.edu.kz

<https://vestnik-pm.tou.edu.kz/>