

<https://doi.org/10.31891/2307-5732-2025-357-27>

УДК 004.4:519.85

ДЕМКІВСЬКИЙ ЄВГЕН

Київський національний університет імені Тараса Шевченка

<https://orcid.org/0000-0002-9382-7466>e-mail: demkivsky@gmail.com**ДЕМКІВСЬКА ТЕТЯНА**

Київський національний університет технологій та дизайну

<https://orcid.org/0000-0002-2176-163X>e-mail: demkivskiy@gmail.com**КОРОГОД ГАННА**

Київський національний університет технологій та дизайну

<http://orcid.org/0000-0003-1670-3125>e-mail: korogod.go@knuutd.com.ua**КИРИЧЕНКО АНТОН**

Київський національний університет технологій та дизайну

<https://orcid.org/0000-0003-0041-3799>e-mail: kirsto@ukr.net

СТВОРЕННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ, ЩО ВИКОРИСТОВУЄ АВТОРЕГРЕСІЙНІ МОДЕЛІ ДЛЯ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

В роботі вирішено завдання розробки програмного забезпечення для автоматизованого вибору адекватної авторегресійної моделі з множини попередньо побудованих моделей з метою прогнозування динаміки процесів, представлених у вигляді часових рядів.

Ключові слова: часовий ряд, авторегресійна модель, критерій адекватності, інтерфейс, статичні характеристики, автокореляційна та часткова автокореляційна функція

DEMKIVSKYI IEVGEN

Taras Shevchenko National University

DEMKIVSKA TETIANA, KOROHOD HANNA,**KYRYCHENKO ANTON**

Kyiv National University of Technologies and Design

CREATION OF SOFTWARE USING AUTOREGRESSION MODELS FOR TIME SERIES ANALYSIS AND FORECASTING

The work performed time series modeling using the method of constructing autoregressive models. At the first stage, the given series was checked for stationarity and the autocorrelation and partial autocorrelation function were constructed. Based on the analysis of the autocorrelation and partial autocorrelation function, a number of autoregressive time series models of different orders were constructed, and the best (adequate) model was determined among the set of constructed ones. The determination of the best model was carried out based on the analysis of their statistical characteristics using the consolidated criterion of adequacy of models.

The modeling was carried out in the econometric package EViews.

The optimal model was presented, a forecast was constructed in 3 steps, and the forecast accuracy was assessed using the MAPE metric. The results obtained are illustrated in figures and graphs. Software was developed to determine the best model among the constructed models for predicting the dynamics of the time series. The software development process was based on the Python programming language, the choice of which was justified by its concise syntax, wide application capabilities and the presence of a developed ecosystem of libraries, in particular pandas (for data processing), numpy (for numerical calculations), statsmodels (for statistical modeling) and scikit-learn (for machine learning tasks). The development and debugging of the program code was carried out in the integrated environment PyCharm.

The result obtained as a result of program execution was verified by comparison with the control example calculated in EViews, and its identity was confirmed.

Keywords: time series, autoregressive model, adequacy criterion, interface, static characteristics, autocorrelation and partial autocorrelation function

Стаття надійшла до редакції / Received 21.07.2025

Прийнята до друку / Accepted 15.08.2025

Постановка проблеми

Прогнозування відіграє критично важливу роль для ефективного управління економікою на всіх рівнях – від державного до рівня окремого підприємства. Компанії використовують прогнози для планування виробництва, управління запасами, розробки маркетингових стратегій, прийняття інвестиційних рішень. Прогнозування допомагає зменшити невизначеність щодо майбутнього стану економіки, що дозволяє суб'єктам приймати більш обережні та зважені рішення. Прогнози можуть виявляти потенційні ризики та загрози (наприклад, рецесію, зростання інфляції, зміни в законодавстві), що дає змогу завчасно розробити стратегії їхнього мінімізації.

Часові ряди є невід'ємною частиною багатьох сфер діяльності, включаючи фінанси (ціни на акції, обмінні курси), економіку (ВВП, інфляція), метеорологію (температура, опади), енергетику

(споживання електроенергії), телекомунікації (трафік мережі), охорону здоров'я (поширення захворювань) та багато інших. Ефективне прогнозування цих рядів є критично важливим для прийняття обґрунтованих рішень, планування та оптимізації процесів.

Ручне застосування авторегресійних моделей для прогнозування може бути трудомістким і вимагати глибоких знань статистики та програмування. Існує зростаюча потреба в зручному, автоматизованому та гнучкому програмному забезпеченні, яке дозволить фахівцям з різних галузей ефективно використовувати ці моделі без необхідності глибокого занурення в їхню математичну основу.

Отже, розробка програмного забезпечення для прогнозування процесів на основі авторегресійних моделей є актуальною проблемою через важливість точного прогнозування, потребу в автоматизації та гнучкості, необхідність інтеграції з іншими інструментами та даними, а також можливості, що відкриваються завдяки розвитку технологій.

Аналіз останніх досліджень

Прогнозування динаміки часових рядів на основі авторегресійних моделей є популярним підходом до прогнозування технічних та економічних процесів.

Розглянемо ряд публікацій, в яких висвітлюється проблема прогнозування процесів, описаних часовими рядами на основі моделей авторегресії.

В роботі [1] розглянуто застосування часових рядів для прогнозування фінансових процесів, представлено методичні підходи до аналізу та прогнозування, зокрема основні етапи аналізу, характеристики і розподіл фінансових даних, а також врахування горизонту прогнозування та сезонності. У статтях [2, 3] досліджено ефективність методу авторегресії (AR), інтегрованої моделі авторегресії ковзного середнього (ARIMA), а також описано та порівняно моделі прогнозування часових рядів, а також досліджено їхню застосовність до задач прогнозування з різними характеристиками даних, такими як часові інтервали, наявність сезонності чи трендів. У дослідженні [4] розробляється удосконалена прогностична модель, яка інтегрує переваги базових моделей, щоб забезпечити більш точні та волатильні прогнози. Робота [5] присвячена дослідженню методів прогнозування нелінійних нестационарних процесів, які моделюються часовими рядами і можуть характеризувати динаміку як технічних, так і економічних систем.

Аналіз останніх досліджень проблеми застосування авторегресійних моделей для прогнозування часових рядів показав їхню ефективність у моделюванні лінійної залежності між послідовними спостереженнями, проте також виявив обмеження у випадках складної нелінійної динаміки та необхідність ретельного підбору порядку моделі та перевірки стаціонарності даних для забезпечення надійних прогнозів.

Формулювання цілей статті

Метою роботи є побудова авторегресійних моделей часового ряду, визначення кращої моделі серед множини побудованих моделей кандидатів, розроблення програмного забезпечення для вибору адекватної моделі на основі аналізу множини статистичних характеристик моделей, а також побудова прогнозу на декілька кроків та оцінка якості прогнозу.

Виклад основного матеріалу

Дослідимо експериментальний набір даних зі 100 спостережень, де кожне спостереження є результатом вимірювання показника дохідності депозиту банку.

Динаміка заданого ряду представлена на Рис.1.

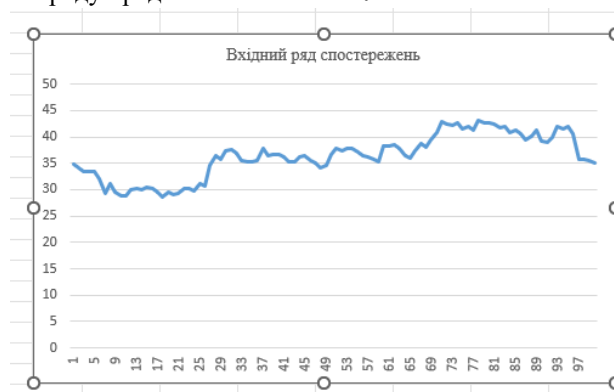


Рис.1. Динаміка заданого ряду спостережень

Першим кроком аналізу даних є перевірка ряду на стаціонарність.

Стаціонарність часового ряду означає, що його статистичні властивості, такі як середнє значення, дисперсія та автокореляція, не змінюються з часом. Іншими словами, ряд є стаціонарним, якщо його поведінка не залежить від того, в який момент часу ми його спостерігаємо.

Існує два основних типи стаціонарності: сильна стаціонарність, коли усі статистичні властивості ряду є незмінними з часом і слабка стаціонарність (стаціонарність другого порядку), коли лише перші два моменти розподілу (середнє значення та автоковаріаційна функція) є незмінними з часом. Для багатьох практичних застосувань достатньо слабкої стаціонарності.

Найпростіший спосіб перевірки ряду на стаціонарність є аналіз графіку часового ряду. На графіку не повинно бути чітко вираженого тренду (зростання або спадання) або сезонності (періодичних повторюваних патернів). Значення повинні коливатися навколо постійного середнього рівня. Дисперсія (розсіювання точок навколо середнього) також повинна бути приблизно постійною.

Іншим способом перевірки на стаціонарність є аналіз графіку автокореляційної функції (ACF). ACF для стаціонарного ряду зазвичай швидко спадає до нуля або залишається в межах значущості для невеликої кількості лагів, що спостерігається з Рис. 2.

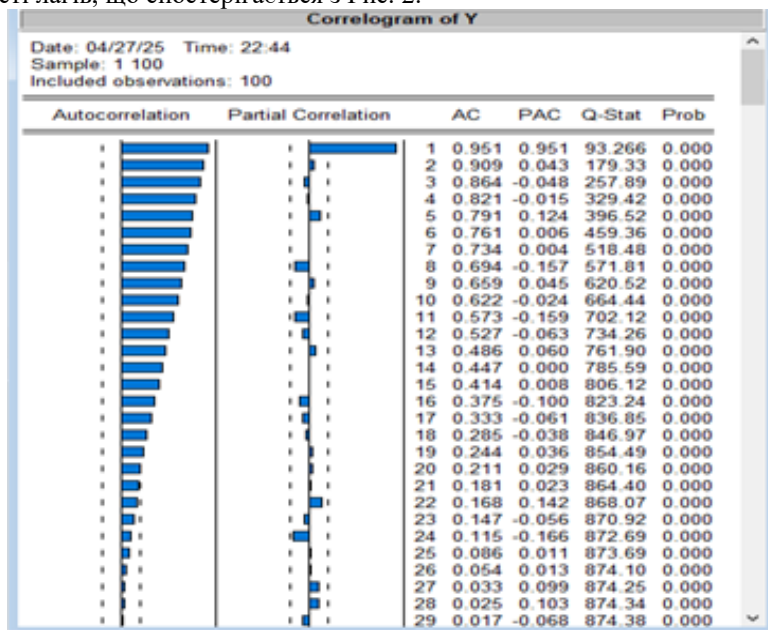


Рис.2 ACF та PACF заданого ряду спостережень

З аналізу автокореляційної та часткової автокореляційної функції ряду спостережень видно, що ACF плавно спадає, а PACF має значимі лаги 1, 5, 8 побудуємо моделі, використовуючи методику побудови авторегресійних моделей, включивши ці лаги в моделі. Для моделювання використовується економетричний пакет EViews.

Це будуть моделі:

- авторегресійна модель першого порядку

$$y_t = c(1) + c(2) * y_{t-1}$$

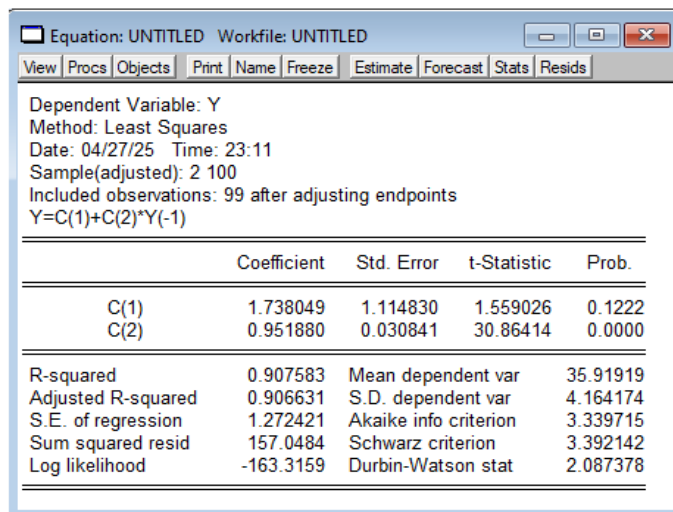


Рис.3 Статистичні характеристики і параметри авторегресійної моделі першого порядку AR(1)

- авторегресійна модель другого порядку

$$y_t = c(1) + c(2) * y_{t-1} + c(3) * y_{t-5}$$

Equation: UNTITLED Workfile: UNTITLED				
View Procs Objects Print Name Freeze Estimate Forecast Stats Resids				
Dependent Variable: Y				
Method: Least Squares				
Date: 04/27/25 Time: 22:59				
Sample(adjusted): 6 100				
Included observations: 95 after adjusting endpoints				
Y=C(1)+C(2)*Y(-1)+C(3)*Y(-5)				
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	1.650571	1.197535	1.378307	0.1715
C(2)	0.924770	0.057499	16.08319	0.0000
C(3)	0.030076	0.057397	0.523997	0.6015
R-squared	0.906954	Mean dependent var	36.03158	
Adjusted R-squared	0.904931	S.D. dependent var	4.213587	
S.E. of regression	1.299184	Akaike info criterion	3.392419	
Sum squared resid	155.2848	Schwarz criterion	3.473068	
Log likelihood	-158.1399	Durbin-Watson stat	2.030459	

Рис.4. Статистичні характеристики і параметри авторегресійної моделі другого порядку AR(2)

- авторегресійна модель третього порядку

$$y_t = c(1) + c(2) * y_{t-1} + c(3) * y_{t-5} + c(4) * y_{t-8}$$

Equation: UNTITLED Workfile: UNTITLED				
View Procs Objects Print Name Freeze Estimate Forecast Stats Resids				
Dependent Variable: Y				
Method: Least Squares				
Date: 04/27/25 Time: 23:16				
Sample(adjusted): 9 100				
Included observations: 92 after adjusting endpoints				
Y=C(1)+C(2)*Y(-1)+C(3)*Y(-5)+C(4)*Y(-8)				
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	2.015481	1.205771	1.671529	0.0982
C(2)	0.926171	0.058148	15.92789	0.0000
C(3)	0.080365	0.080199	1.002078	0.3191
C(4)	-0.061390	0.073233	-0.838279	0.4041
R-squared	0.911370	Mean dependent var	36.20652	
Adjusted R-squared	0.908348	S.D. dependent var	4.160352	
S.E. of regression	1.259507	Akaike info criterion	3.341822	
Sum squared resid	139.5994	Schwarz criterion	3.451465	
Log likelihood	-149.7238	Durbin-Watson stat	1.916066	

Рис.5. Статистичні характеристики і параметри авторегресійної моделі третього порядку AR(3)

Для визначення кращої моделі серед множини побудованих моделей використовуємо консолідований критерій адекватності моделей, включно з перевіркою якості моделі для прогнозування, що забезпечується коефіцієнтом Тейла [6].

$$KK = e^{1-R^2} + \frac{RSS}{N} + \left\{ \begin{matrix} \ln(AIC + BSC), & AIC + BSC > 0 \\ e^{AIC+BSC}, & AIC + BSC \leq 0 \end{matrix} \right\} + e^{2-DW} + e^U$$

Мінімальне значення критерію відповідає кращій моделі. Результати моделювання представлені на рисунку 6.

Модель	R ²	RSS	AIC	BSC	DW	Theil	KK
$y = 1.738 + 0.95 * y(-1)$	0,906954	155,2824	3,392419	3,473068	2,030459	0,04875	6,596801226
$y = 1.65 + 0.92 * y(-1) + 0.03 * y(-5)$	0,907583	157,0484	3,339715	3,392142	2,087378	0,052706	6,544607481
$y = 2.01 + 0.92 * y(-1) + 0.08 * y(-5) - 0.06 * y(-8)$	0,91137	139,5994	3,341822	3,451465	1,916066	0,042486	6,535563771

Рис.6. Результати моделювання

В результаті проведеного моделювання на основі аналізу статистичних характеристик отримали кращу модель. Це модель AR(3).

$$y_t = 2,0155 + 0,9262 * y_{t-1} + 0,0836 * y_{t-5} - 0,0639 * y_{t-8} + e_t$$

Заданий ряд містив 100 спостережень. Із застосуванням адекватної моделі побудовано прогноз на 3 кроки: 101, 102, 103.

N	y_i	$\hat{y}_i$
99	35,5	33,54054188
100	35	33,54565718
101	$y(101)$	33,74691399
102	$y(102)$	33,93341892
103	$y(103)$	34,10626068
104		

Рис.7. Прогнозування на 3 кроки

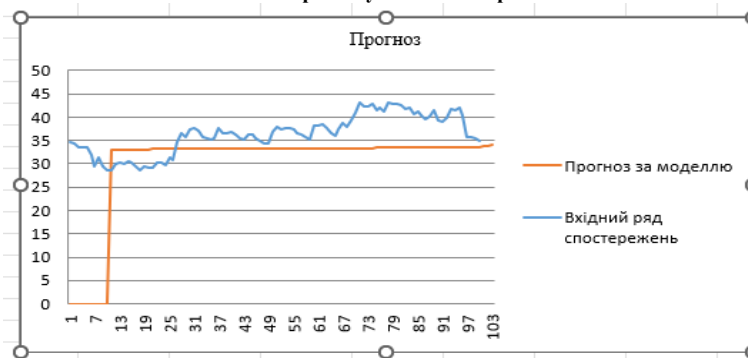


Рис.8. Графічне представлення побудованого прогнозу

Процес розробки програмного забезпечення базувався на використанні мови програмування Python, вибір якої зумовлений її простотою синтаксису, гнучкістю застосування та наявністю широкого спектру бібліотек для опрацювання даних і машинного навчання. В рамках проєкту було використано наступні основні бібліотеки: pandas (для обробки та аналізу даних), numpy (для виконання числових обчислень), statsmodels (для статистичного моделювання та аналізу), scikit-learn (для задач машинного навчання). Інтегроване середовище розробки PyCharm слугувало для написання та налагодження програмного коду.

На Рис. 9 представлено результат роботи програми, визначено кращу модель, виходячи з аналізу її статистичних характеристик.

Модель	R ²	RSS	AIC	BSC	DW	Theil	KK
AR(1)	0,9069	155,2824	3,3924	3,4730	2,0304	0,0487	6,5968
AR(2)	0,9076	157,0484	3,3397	3,3921	2,0873	0,0527	6,5446
AR(3)	0,9113	139,5994	3,3418	3,4514	1,9160	0,0424	6,5356

Рис.9. Знаходження кращої моделі

Отриманий внаслідок виконання програми результат було верифіковано шляхом порівняння з контрольним прикладом, розрахованим в EViews, та підтверджено його ідентичність.

Оцінено точність отриманої моделі для прогнозування за допомогою метрики MAPE (Mean Absolute Percentage Error), яка показує середню абсолютну відсоткову похибку між прогнозованими та фактичними значеннями. Це дає чітке уявлення про те, наскільки відхиляються прогнози від реальних даних у відсотковому вираженні.

Для MAPE менші значення також вказують на кращу точність прогнозу. Зазвичай, значення MAPE менше 10% вважається дуже хорошим прогнозом, 10-20% - хорошим, 20-50% - прийнятним, а понад 50% - поганим. Однак ці межі можуть варіюватися залежно від галузі та конкретної задачі.

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| * 100\% = 8.4\%$$

MAPE на рівні 8.4% вказує на апридатність побудованої моделі для прогнозування.

Висновки

Розроблено програмне забезпечення для визначення кращої моделі серед побудованих моделей для прогнозування динаміки часового ряду. Виконано моделювання в EViews, яке базується на використанні методики побудови авторегресійних моделей та побудовано прогноз на 3 кроки. Для визначення кращої моделі серед побудованих використано консолідований критерій адекватності моделі. Виконано оцінку точності прогнозу за допомогою метрики MAPE.

Побудова прогнозу в банківській сфері дозволяє банку оцінити, наскільки привабливими будуть його депозитні продукти для клієнтів у майбутньому. Розуміння майбутніх тенденцій доходності може допомогти банку розробляти нові, більш конкурентоспроможні та привабливі депозитні

пропозиції. Банк може своєчасно коригувати процентні ставки, терміни та інші умови існуючих депозитів, щоб залишатися конкурентоздатним та залучати кошти а також може використовувати прогнози для формування оптимальної структури своїх пасивів, балансуючи між вартістю залучених коштів та їхньою стабільністю.

Література

1. Hamdy A. Taha. (2017). Operations Research: An Introduction Pearson Education Limited. P. 833. URL: <http://zalamasyah.staff.unja.ac.id/wp-content/uploads/sites/286/2019/11/9-Operations-Research-An-Introduction-10th-Ed.-Hamdy-A-Taha.pdf>
2. Ronald L. Rardin (2017) Optimizations in operations research, Pearson Higer educations, P. 1127:URL: <https://industri.fatek.unpatti.ac.id/wp-content/uploads/2019/03/173-Optimization-in-Operations-Research-Ronald-L.-Rardin-Edisi-2-2015.pdf>
3. F. Werner, L. Burtneva, Y. Sotokov. (2018), Algotitms for Scheduling Problems . – MDPI, P. 192 URL: https://books.google.com.ua/books?hl=uk&lr=&id=Da1qDwAAQBAJ&oi=fnd&pg=PR7&dq=related:xrREKSP4fp0J:scholar.google.com/&ots=9SEhWKyEPz&sig=NcoH5QHB3CLL7uufHdzhuMjzsbo&redir_esc=y#v=onepage&q&f=false
4. Pinedo, M.L. (2012) Scheduling: Theory, Algorithms, and Systems. Springer, NY, USA. P. 672. URL: <https://www.google.com.ua/books/edition/Scheduling/QRiDnuXSnVwC?hl=ru&gbpv=1&dq=inauthor:%22Michael+L.+Pinedo%22&printsec=frontcover>
5. Low, C., Li, R. K., Wu, G. H., & Huang, C. L. (2015). Minimizing the sum of absolute deviations under a common due date for a single-machine scheduling problem with availability constraints. Journal of Industrial and Production Engineering, 32(3), 204-217.2017. URL: <https://doi.org/10.1080/21681015.2015.1031196>
6. Демківський О.Б. Статистична обробка результатів вимірювань та експериментальних даних в текстильній промисловості / О.Б. Демківський, С.М. Кравснитський, Ю.М. Пимлипенко, А.М. Слізков // Навчальний посібник. – Київ. - КНУТД,. - 2012. – 105 с. URL: https://er.knutd.edu.ua/bitstream/123456789/6123/1/20120424_Demkivskiy.pdf

References

1. Hamdy A. Taha. (2017). Operations Research: An Introduction Pearson Education Limited. P. 833. URL: <http://zalamasyah.staff.unja.ac.id/wp-content/uploads/sites/286/2019/11/9-Operations-Research-An-Introduction-10th-Ed.-Hamdy-A-Taha.pdf>
2. Ronald L. Rardin (2017) Optimizations in operations research, Pearson Higer educations, P. 1127 URL: <https://industri.fatek.unpatti.ac.id/wp-content/uploads/2019/03/173-Optimization-in-Operations-Research-Ronald-L.-Rardin-Edisi-2-2015.pdf>
3. F. Werner, L. Burtneva, Y. Sotokov. (2018), Algotitms for Scheduling Problems . – MDPI, P. 192 URL: https://books.google.com.ua/books?hl=uk&lr=&id=Da1qDwAAQBAJ&oi=fnd&pg=PR7&dq=related:xrREKSP4fp0J:scholar.google.com/&ots=9SEhWKyEPz&sig=NcoH5QHB3CLL7uufHdzhuMjzsbo&redir_esc=y#v=onepage&q&f=false
4. Pinedo, M.L. (2012) Scheduling: Theory, Algorithms, and Systems. Springer, NY, USA. P. 672. URL: <https://www.google.com.ua/books/edition/Scheduling/QRiDnuXSnVwC?hl=ru&gbpv=1&dq=inauthor:%22Michael+L.+Pinedo%22&printsec=frontcover>
5. Low, C., Li, R. K., Wu, G. H., & Huang, C. L. (2015). Minimizing the sum of absolute deviations under a common due date for a single-machine scheduling problem with availability constraints. Journal of Industrial and Production Engineering, 32(3), 204-217.2017. URL: <https://doi.org/10.1080/21681015.2015.1031196>
6. Demkivskiy, O.B., Krasnytskyi, S.M., Pylypenko, Yu.M., Slizkov A.M., (2012)– *Statystychna obrobka rezul'tativ vymiriuvan ta eksperymentalnykh danykh v tekstylnii promyslovosti* [Statistical processing of measurement results and experimental data in the textile industry] Kyiv.:KNTUD [in Ukrainian]. URL: https://er.knutd.edu.ua/bitstream/123456789/6123/1/20120424_Demkivskiy.pdf