

Первинне опрацювання статистичних даних та оцінювання параметрів розподілу засобами OpenOffice Calc

Н.В. Шульга*

Черкаський національний університет ім. Б. Хмельницького, Черкаси, Україна

*Corresponding author. E-mail: schulganv@mail.ru

Paper received 18.07.15; Accepted for publication 31.07.15.

Анотація. В статті розкриваються методичні особливості застосування пакету прикладних програм OpenOffice в процесі навчання стохастичності студентів економічних факультетів університетів. Розглянуто методичні особливості застосування вбудованих функцій додатку Calc у первинному опрацюванні статистичних даних (визначення обсягу вибірки, найбільшого та найменшого її значення, частоти варіант; обчислення числових характеристик вибірки), статистичному та інтервальному оцінюванню параметрів розподілу, а також побудові полігону та гістограми розподілу.

Ключові слова: навчання стохастичності, комп'ютерні технології у навчанні, пакет прикладних програм OpenOffice, статистичний аналіз даних, теорія статистичних оцінок, підготовка економістів

Вступ. На початку ХХІ сторіччя суспільно-економічний розвиток людства може бути охарактеризовано як перехід до інформаційного суспільства. Основою та рушійною силою даного процесу є інформація, темпи зростання якої прискорюються з кожним роком. Інформатизація суспільства має наслідки як позитивні (збільшення інформації приводить до утворення нових знань, нових технологій, нових продуктів), так і негативні (через велику кількість другорядної інформації зростає час на її опрацювання, збільшується ймовірність втрати більш важливих знань, отже може з'явитись така парадоксальна ситуація, коли через надлишок інформації виникає інформаційний голод). В результаті, зростає необхідність у фахівців, що мають здатність швидко адаптуватись до нових, або суттєво змінених, умов праці, знаходити та використовувати інноваційні технології, що дають можливість швидко отримати результат.

Огляд публікацій за темою. В сучасних умовах об'єктивним фактором є експансія технічних і технологічних інновацій у всі сфери соціально-економічного буття. Якісна професійна підготовка сьогодні включає в себе необхідність у формуванні компетенцій застосування комп'ютерної техніки та інформаційних технологій в майбутній фаховій діяльності. Дослідженню даної проблеми присвячено роботи Л. Боднар, А. Борисенка, А. Ванюріна, І.Воронікової, О. Зайцевої, Т. Захарової, О. Гавриліної, Н. Канаєвої, І. Керницького, М. Кобякової, Т. Коломієць, І. Костікової, О. Кривоноса, Л. Морської, В. Пономаренка, Л. Пуханової, М. Рагуліної, Л. Мартіросян, І. Роберт, І. Смірної, Є. Третьякової, О. Федорчук, В. Шухардіної, М. Яшанова та ін.

Водночас, комп'ютерні технології мають потужний потенціал застосування в освітньому процесі. Широке розповсюдження та вільний доступ до інформаційних технологій надають можливість створення та використання нових форм навчання, таких, наприклад, як віртуальне навчання, проблеми впровадження якого в навчальний процес стали предметом дослідження Я. Воловця, В. Григор'єва, О. Зачка, Л. Іваненко, М. Козяра, С. Литвиної, О. Малярчук, Г. Мошек, Р. Павлюка, А. Петриці, Т. Рака, А. Столяревської, Г. Сухорукової, В. Хандецького, В. Хомич та ін.

Сучасні інновації змінюють і технічні засоби навчання. Наразі актуальним питанням є використання в

навчальному процесі електронних інтерактивних посібників та навчальних програм-тренажерів (Р. Аршинянц, Я. Ваграменко, С. Григор'єв, Л. Зайнугдінова, О. Зіміна, А. Кіріллово, С. Макаров, Г. Можасва, І. Тубалова та ін.).

Аналіз недетермінованих процесів та явищ, статистична обробка емпіричних даних значно спрощується за рахунок використання пакетів математичних прикладних програм. Саме тому, важливим аспектом у процесі навчання стохастичності майбутніх економістів є формування інформаційних компетенцій. Можливості застосування пакетів комп'ютерних математичних програм у навчанні стохастичності відображено в працях Г. Горелової, М. Жалдака, Е. Железнякової, Н. Кузьміної, І. Лебедевої, Л. Малярця, Г. Михаліна, С. Самсонової, М. Суворової, С. Щербатих та ін.

В той же час, економічна криза в світі змушує шукати такі технології, оплата за ліцензію на використання яких була б по кишені навчальним закладам та роботодавцям. Розв'язати дану проблему можна використавши в стохастичних дослідженнях пакети програм з вільним доступом. Однією з таких програм є пакет офісних програм OpenOffice, що представляє собою повнофункціональний пакет офісних додатків, який було розроблено як альтернативу пакету Microsoft Office. Даний пакет підтримується усіма основними операційними системами, в тому числі і Windows.

Мета статті. Широкі можливості застосування пакету прикладних офісних програм OpenOffice наряду з відсутністю оплати за ліцензію вказують на те, що в найближчому майбутньому даний пакет програм вийде на перше місце по застосуванню в навчальному процесі як в школах, так і у вищих навчальних закладах України. Разом з тим, аналіз доступної літератури з даної теми, вказує на те, що практично немає методичних розробок присвячених особливостям застосування пакету OpenOffice та його додатку OpenOffice Calc у навчанні стохастичності. Саме тому, наше дослідження було спрямоване на аналіз можливостей застосування вбудованих статистичних функцій, що входять до додатку OpenOffice Calc в процесі навчання стохастичності студентів економічних спеціальностей університетів.

Матеріали та методи дослідження. В процесі дослідження було використано наступні методи:

– аналіз теоретичних джерел з проблеми застосування вбудованих функцій додатку OpenOffice Calc в

процесі навчання. Принципи роботи з програмами пакету OpenOffice досить широко висвітлено на спеціальних сайтах в мережі Інтернет, зокрема База знань Open Office.org [5], Центр нових інформаційних технологій [10], Prosto Pro Office [3], Основи преподавания офісних технологий на базе OpenOffice.org [1] тощо;

– аналіз методичних підходів до застосування статистичних функцій OpenOffice Calc в процесі навчання взагалі та під час стохастичної підготовки майбутніх економістів. В роботах І. Акімової [4], С. Тімохіної [8], Є. Шереметьєвої [11] представлено методичні розробки щодо застосування офісного додатку OpenOffice в процесі навчання. В. Клячкін [6], Д. Расторгуев [6] розглядають можливості застосування статистичних функцій в економічних дослідженнях.

Результати дослідження. OpenOffice – це повнофункціональний пакет офісних додатків, який було розроблено як альтернативу пакету Microsoft Office як на рівні форматів так і на рівні інтерфейсу. Одним із переваг пакету OpenOffice є те, що він не потребує ніякої плати за ліцензією, в той же час може вільно замінити будь-який з пакетів прикладних офісних програм. В пакеті OpenOffice передбачено зберігання будь-яких даних у відкритому форматі Open Document Format (ODF), який був схвалений міжнародною організацією по стандартизації ISO (ISO/IEC 26300:2006). Документи, що створено за допомогою OpenOffice, можуть бути збережені в інших поширених форматах (в тому числі і тих, що підтримуються в Microsoft Office), в пакеті також передбачено можливість експорту документів в формат PDF [2, 9]. До складу пакету входять текстовий редактор Writer, програма для підготовки презентацій Impress, графічний редактор Draw, табличний процесор Calc, система управління базами даних Base, редактор формул Math.

Розглянемо можливості застосування табличного процесора OpenOffice Calc під час вивчення навчального матеріалу з математичної статистики студентами економічних спеціальностей університетів.

В процесі вивчення теми **Первинне опрацювання статистичних даних** доцільно використати наступні вбудовані функції:

1. Представлення статистичних даних

- COUNT(значення1; значення2; ...), що належать до категорії Статистичні, для визначення обсягу вибірки n ;
- FREQUENCY(дані; класи), що належать до категорії Масив, для визначення частоти n_i появи варіанти та попадання варіанти у визначений інтервал;
- побудувати варіаційний ряд вибірки можна за допомогою сортування, виконавши команди Данные – Сортировать;
- розмах варіації $w = x_{\max} - x_{\min}$ може бути обчислений за допомогою статистичних функцій MAX(число1; число2;...) та MIN(число1; число2;...)
- для обчислення значень емпіричної функції можна скористатись математичною функцією SUM(число1; число2;...);
- передбачено також можливість графічного представлення даних у вигляді полігону розподілу частот та гістограми.

2. Числові характеристики вибірки:

- статистичну функцію AVERAGE(число1; число2; ...) використовують для обчислення вибіркової середньої $\bar{x}_B = \frac{\sum_{i=1}^n x_i}{n}$, де x_i - значення варіант, n - обсяг вибірки.
- функція GEOMEAN(число1;число2;...) обчислює середнє геометричне вибірки $\bar{x}_{\text{геом}} = \sqrt[n]{\prod_{i=1}^n x_i}$, а функція HARMEAN(число1;число2;...) – середнє гармонійне $\bar{x}_{\text{гар}} = n \cdot \left(\sum_{i=1}^n \frac{1}{x_i}\right)^{-1}$;
- вибіркoву дисперсію $D_B = \frac{\sum_{i=1}^n (x_i - \bar{x}_B)^2}{n}$ обчислюють за допомогою функції VARP(число1; число2;...);
- середнє квадратичне значення вибірки $\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x}_B)^2}{n}}$ обчислюють за допомогою функції STDEVP(число1; число2;...);
- мода вибірки може бути визначена із використанням функції MODE(число1; число2; ...), а медіана – MEDIAN(число1; число2; ...).

В табличному процесорі Calc передбачено також можливість графічного представлення даних у вигляді полігону розподілу частот та гістограми.

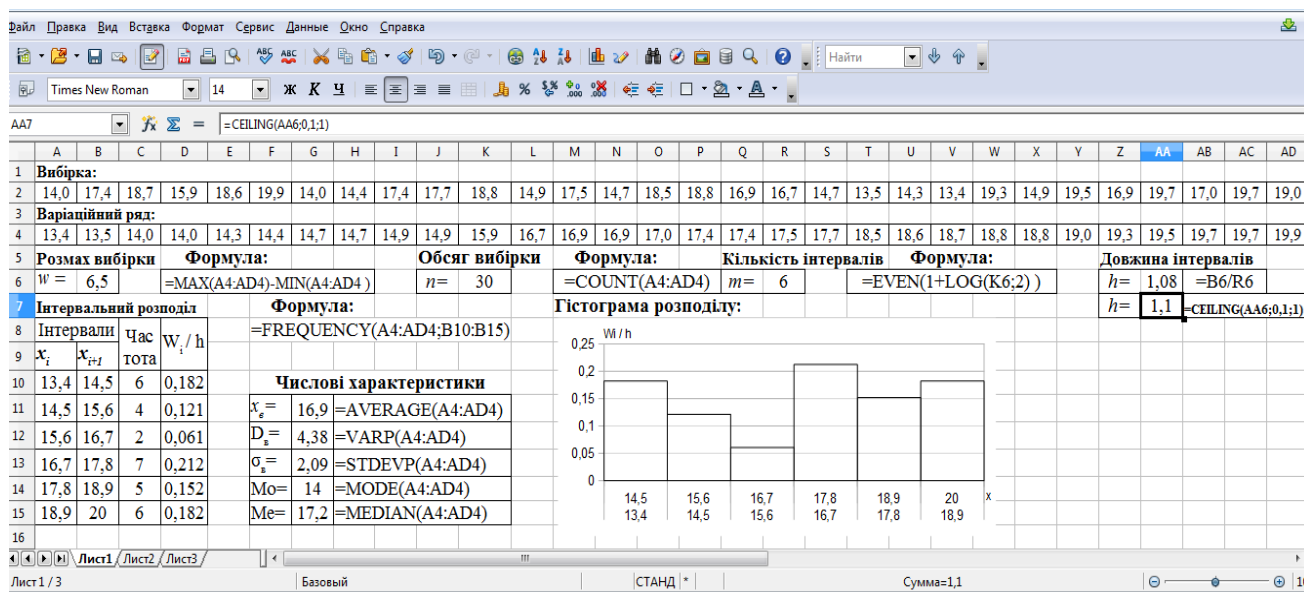


Рис. 1.

Розглянемо наступний *приклад*: Проводяться вимірювання ваги виробу. Результати випробувань: 14; 17,4; 18,7; 15,9; 18,6; 19,9; 14,0; 14,4; 17,4; 17,7; 18,8; 14,9; 17,5; 14,7; 18,5; 18,8; 16,9; 16,7; 14,7; 13,5; 14,3; 13,4; 19,3; 14,9; 19,5; 16,9; 19,7; 17,0; 19,7; 19,0. Провести статистичний та графічний аналіз вибірки.

Оскільки в результаті експерименту отримали значну кількість різних варіант, то доцільно побудувати інтервальний розподіл вибірки. Довжину інтервалу визначимо за формулою $h = \frac{w}{m}$, де $w = x_{\max} - x_{\min}$ - розмах вибірки, $m = 1 + \log_2 n$ - кількість інтервалів. Для того, щоб округлити число m , що отримали в результаті розрахунків, до найбільшого цілого, можна використати математичну функцію EVEN(число). Округлити довжину інтервалу до числа із заданою точністю можна за рахунок використання функції CEILING(число, точність, спосіб). Для побудови гістограми за відносними частотами, обчислимо відношення $\frac{W_i}{n}$, де $W_i = \frac{n_i}{n}$. Модель розв'язання задачі наведено на рис. 1.

В процесі вивчення теми Статистичне та інтервальне оцінювання параметрів розподілу

1. Для визначення точкових оцінок параметрів генеральної сукупності доцільно використати наступні вбудовані функції:

- VAR(число1; число2;...), яка обчислює виправлену дисперсію $S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x}_B)^2}{n-1}$, яка є незсунутою та обґрунтованою оцінкою генеральної дисперсії;

- STDEV(число1; число2;...), що визначає величину стандарту $S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x}_B)^2}{n-1}}$ (виправленого середньоквадратичного відхилення) генеральної сукупності;

- SKEW(число1; число2;...), що обчислює величину асиметрії розподілу $As = \frac{\mu_3}{\sigma^3}$ і дає змогу оцінити симетричність розподілу генеральної сукупності відносно її математичного сподівання;

- CONFIDENCE (рівень значущості α ; величина стандарту S ; обсяг вибірки n) - визначає величину відхилення $\delta = \pm \frac{t\sigma}{\sqrt{n}}$ довірчого інтервалу $[\bar{x}_B - \frac{t\sigma}{\sqrt{n}}; \bar{x}_B + \frac{t\sigma}{\sqrt{n}}]$ для оцінки математичного сподівання, якщо відома дисперсія σ^2 , та задано надійність $\gamma = 1 - \alpha$, де $\Phi_0(t) = \frac{\gamma}{2}$ - функція Лапласа;

- TINV(рівень значущості α ; кількість ступенів свободи k) - визначає параметр t_γ розподілу Стьюдента, необхідний для обчислення із надійністю $\gamma = 1 - \alpha$ границь довірчого інтервалу $[\bar{x}_B - \frac{t_\gamma \cdot S}{\sqrt{n}}; \bar{x}_B + \frac{t_\gamma \cdot S}{\sqrt{n}}]$ для математичного сподівання, якщо невідома дисперсія;

- KURT(число1; число2;...), що обчислює величину ексцесу розподілу $Es = \frac{\mu_4}{\sigma^4} - 3$ і дає змогу оцінити гостровершинність (чи плосковершинність) розподілу генеральної сукупності;

2. Для визначення інтервальних оцінок параметрів генеральної сукупності доцільно використати наступні вбудовані функції:

- CHINV(рівень значущості α ; кількість ступенів свободи k) - визначає параметри $\chi_1(\frac{1+\gamma}{2}, k)$ та $\chi_2(\frac{1-\gamma}{2}, k)$ розподілу $\chi^2(\alpha, k)$, необхідний для обчислення із надійністю γ границь довірчого інтервалу

$[\frac{\sqrt{n} \cdot S}{\chi_2}; \frac{\sqrt{n} \cdot S}{\chi_1}]$ для середнього квадратичного відхилення, де $k = n$, якщо математичне сподівання - відоме, $k = n - 1$, якщо математичне сподівання - невідоме.

Застосування функцій розглянемо на наступних прикладах.

Приклад 1. Розмір деталі визначається за допомогою приладу, що має стандартне відхилення 5 мм. Проведено 20 незалежних вимірювань випадкової величини $X = \{ \text{Розмір деталі} \}$. Результати вимірювань наступні (см): 58,08; 58,74; 55,66; 57,86; 55,88; 59,40; 56,98; 55,88; 59,18; 58,30; 57,42; 56,10; 56,32; 59,18; 57,42; 57,20; 56,10; 56,98; 56,32; 59,40. Припустивши, що випадкова величина X розподілена за нормальним законом, побудувати 95% довірчий інтервал для оцінки її математичного сподівання. (Модель розв'язування представлено на рис. 2)

	A	B	C	D	E
1	Вибірка:				
2	58,08	58,74	55,66	57,86	55,88
3	59,4	56,98	55,88	59,18	58,3
4	57,42	56,1	56,32	59,18	57,42
5	57,2	56,1	56,98	56,32	59,4
6					
7	$\sigma =$	0,05			
8	$\gamma =$	0,95			
9					
10	Вибіркова середня		Формула		
11	$\bar{x}_B =$	57,42	=AVERAGE(A2:E5)		
12					
13	Відхилення:		Формула		
14	$\delta =$	0,0219131	=CONFIDENCE(1-B8;B7;20)		
15					
16	Довірчий інтервал				
17		57,398087	$\leq a \leq$	57,441913	
18		=B11-B14		=B11+B14	
19					

Рис. 2

Приклад 2. Проведено 10 незалежних вимірювань випадкової величини $X = \{ \text{Вага ящика продукції} \}$. Результати вимірювань наступні (кг): 28,08; 28,74; 25,66; 27,86; 25,88; 29,40; 26,98; 25,88; 29,18; 28,30. Припустивши, що випадкова величина X розподілена за нормальним законом, побудувати 95% довірчий інтервал для оцінки її математичного сподівання. (Модель розв'язування представлено на рис. 3)

	H	I	J	K	L
1	Вибірка:				
2	28,08	28,74	25,66	27,86	25,88
3	29,4	26,98	25,88	29,18	28,3
4					
5					
7	$\gamma =$	0,95			
8					
9	Вибіркова середня		Формула		
10	$\bar{x}_B =$	27,60	=AVERAGE(H2:L3)		
11	Виправлене СКВ		Формула		
12	$S =$	1,41	=STDEV(H2:L3)		
13	Параметр t_γ		Формула		
14	$t_\gamma =$	2,26	=TINV(1-I7;9)		
15					
16	Відхилення:		Формула		
17	$\delta =$	1,01	=(I14*I12)/SQRT(10)		
18					
19	Довірчий інтервал				
20		26,59	$\leq a \leq$	28,60	
21		Формула			
22		=I10-I17		=I10+I17	
23					

Рис. 3

	Н	І	Ј	К	Л
1	Вибірка:				
2	98,08	98,74	95,66	97,86	95,88
3	99,4	96,98	95,88	99,18	98,3
4					
5					
6					
7	$\gamma =$	0,95			
8					
9	Вибіркова середня		Формула		
10	$\bar{x}_n =$	97,60	=AVERAGE(H2:L3)		
11	Виправлене СКВ		Формула		
12	S=	1,41	=STDEV(H2:L3)		
13	Параметр χ^2_1		Формула		
14	$\chi^2_1 =$	2,70	=CHINV((1+17)/2;9)		
15	Параметр χ^2_2		Формула		
16	$\chi^2_2 =$	19,02	=CHINV((1-17)/2;9)		
17					
18					
19		Довірчий інтервал			
20		0,97	$\leq a \leq$	2,57	
21			Формула		
22	=(12*SQRT(10-1))/SQRT(116)			=(12*SQRT(10-1))/SQRT(114)	

Рис. 4

Приклад 3. Проведено 10 незалежних вимірювань випадкової величини $X = \{\text{Вага одиниці продукції}\}$. Результати вимірювань наступні (γ): 98,08; 98,74; 95,66; 97,86; 95,88; 99,40; 96,98; 95,88; 99,18; 98,30. Припустивши, що випадкова величина X розподілена за нормальним законом, побудувати 95% довірчий інтервал для оцінки її математичного сподівання. (Модель розв'язування представлено на рис. 4)

Висновки. Застосування табличного процесору OpenOffice Calc у навчанні стохастички майбутніх економістів надає можливість застосування розширеного методичного інструментарію, крім того спрощує процеси обчислень, пошуку та визначення табличних даних, представлення графічних даних та побудови розрахункових таблиць.

ЛІТЕРАТУРА

- [1] MDV101: Основы преподавания офисных технологий на базе OpenOffice.org / URL: http://www.raspo.ru/files/files/prepodavanie_ooo.pdf
- [2] OpenOffice 4.1.0 Beta / URL: <http://new-build.info/office/6570-openoffice-410-beta.html>
- [3] Prosto Pro Office / Просто Про Office / URL: <http://prostoprooffice.com/openoffice/>
- [4] Акимова И.В. Методика работы в СУБД OPEN OFFICE BASE / И.В. Акимова // Известия Пензенского государственного педагогического университета им. В.Г. Белинского. - 2011. - №26. - С. 347-357.
- [5] База знаний OpenOffice.org / URL: https://wiki.openoffice.org/wiki/RU/knowledge_base
- [6] Клячкин В.Н. Практикум по статистике, контролю качества и расчетам надежности в OpenOffice.org Calc: учебное пособие / В.Н. Клячкин – Ульяновск : УлГТУ, 2009. – 133 с.
- [7] Расторгуев Д.Н. Методические рекомендации к практическим занятиям по компьютерному моделированию социально-экономических процессов в OPENOFFICE.ORG CALC / Д.Н. Расторгуев. - Ульяновск : УлГТУ, 2011. – 40 с.
- [8] Тимохина С.Д. Сравнительный анализ работы программного обеспечения офисных приложений Microsoft Office и OpenOffice.org / С.Д. Тимохина / URL: http://mioo.edu.ru/attachments/article/247/Тимохина%20СД%20Сравнительный_анализ_работы_ПО_офисных_приложений.pdf
- [9] Формулы и функции в OpenOffice.org Calc / URL: https://sites.google.com/site/calcfandf/lekci/lekcia_1
- [10] Центр новых информационных технологий / URL: <http://edu.tltsu.ru/sites/site.php?s=117&m=29487>
- [11] Шереметьева Е.Г. Основы работы с офисным пакетом OpenOffice.org 3.2 : Работа с электронной таблицей Open Office.org Calc / Е.Г. Шереметьева / URL: <http://window.edu.ru/resource/931/72931/files/stup590.pdf>

REFERENCES

- [1] MDV101: Principles of learning the office applications technologies on the OpenOffice.org base / URL: http://www.raspo.ru/files/files/prepodavanie_ooo.pdf
- [2] OpenOffice 4.1.0 Beta / URL: <http://new-build.info/office/6570-openoffice-410-beta.html>
- [3] Prosto Pro Office / URL: <http://prostoprooffice.com/openoffice/>
- [4] Akimova I.V. Methods of work in database OPEN OFFICE BASE / I.V. Akimova // News of Belinsky Penza State Pedagogical University. - 2011. - №26. - P. 347-357.
- [5] Knowledge Base OpenOffice.org / URL: https://wiki.openoffice.org/wiki/RU/knowledge_base
- [6] Klyachkin, V.N. Practical work on statistics, quality control and estimates of reliability in OpenOffice.org Calc: Tutorial / V.N. Klyachkin - Ulyanovsk: UIGTU, 2009. - 133 p.
- [7] Rastorguev, D.N. Methodical recommendations to practical training in computer modeling of socio-economic processes in OPENOFFICE.ORG CALC / D.N. Rastorguev. - Ulyanovsk: UIGTU, 2011. - 40 p.
- [8] Timokhina, S.D. Comparative analysis of software office applications Microsoft Office and OpenOffice.org / S.D. Timokhina / URL: http://mioo.edu.ru/attachments/article/247/Тимохина%20СД%20Сравнительный_анализ_работы_ПО_офисных_приложений.pdf
- [9] Formulas and functions in OpenOffice.org Calc / URL: https://sites.google.com/site/calcfandf/lekci/lekcia_1
- [10] Center of New Information Technologies / URL: <http://edu.tltsu.ru/sites/site.php?s=117&m=29487>
- [11] Sheremetyeva, E.G. Basics office application OpenOffice.org 3.2: working with electronic worksheets OpenOffice.org Calc / E.G. Sheremetyeva / URL: <http://window.edu.ru/resource/931/72931/files/stup590.pdf>

Initial processing of statistical data and estimation of distribution parameters tools OpenOffice Calc

N. V. Shulga

Abstract. Methodological features of the application package OpenOffice applications in learning stochastics students of economic faculties of universities are revealed in the article. The ability to use embedded functions of Calc application examined: primary processing of statistics (determination of sample size, its largest and smallest values, frequency variant, calculation of numerical characteristics of the sample), statistical and interval estimation of distribution parameters, and construction of the polygon and the histogram distribution.

Keywords: Stochastics teaching, computer technology in education, application program package OpenOffice, statistical data analysis, statistical theory estimates, training economists