

ПЕРВИННЕ ОПРАЦЮВАННЯ СТАТИСТИЧНИХ ДАНИХ З ВИКОРИСТАННЯМ СИСТЕМИ MAPLE

Тарас КОБИЛЬНИК

У статті проаналізовані можливості використання системи комп'ютерної математики Maple для опрацювання експериментальних даних. Проаналізовано деякі функції пакету stats для первинного опрацювання статистичних даних. Теоретичні відомості ілюструються конкретними прикладами.

The paper analyzed the possibility of using the system Maple computer mathematics to process experimental data. Analyzed some features stats package for data processing statistics. Theoretical data illustrated by concrete examples.

Постановка проблеми та аналіз останніх досліджень. Важливим аспектом застосування статистичних методів є їх комп'ютерна реалізація. Опрацювання експериментальних даних неможлива без використання відповідного програмного забезпечення. Найбільш розповсюдженим засобом для опрацювання експериментальних даних є MS Excel, що містить статистичні функції та надбудову «Аналіз даних» [8]. Система Statistica – це універсальне інтегроване середовище, призначене для проведення статистичного аналізу та візуалізації даних, керування базами даних і розробки власних додатків. У посібниках [1, 7] розглядаються найчастіше вживані методи прикладного статистичного аналізу даних. Кожен розділ посібника відповідає певній групі методів і містить опис прикладної задачі, статистичних засобів її дослідження, необхідне теоретичне обґрунтування та приклади застосування до реальних даних за допомогою пакету певного програмного продукту (MS Excel, Statistica). У посібнику [6] показано у порівнянні інструментарій пакетів MS Excel, SPSS, Statistica для проведення статистичного аналізу. Окремо слід виділити підручник [4], у якому для побудови графічних зображень, аналізу статистичних даних, визначення числових характеристик розподілів ймовірностей, в тому числі статистичних, передбачається використання відповідних комп'ютерних програм, зокрема Gran1 [3]. Крім програмного забезпечення загального та спеціального призначення, для опрацювання статистичних даних можна використовувати й універсальні математичні пакети (системи комп'ютерної математики). Власне й стаття присвячена дослідженню можливостей використання системи Maple [5] для опрацювання статистичних даних.

Мета статті: характеристика пакету stats системи Maple для первинного опрацювання статистичних даних.

Виклад основного матеріалу. До первинного опрацювання статистичних даних відносять визначення числових характеристик вибірки, побудова дискретного та інтервального варіаційних рядів та їх графічне подання, перевірка статистичних гіпотез (наприклад, узгодженість емпіричного розподілу з деяким теоретичним).

Пакет stats містить команди для аналізу та графічного подання статистичних даних, а також розподілів статистичних ймовірностей. Пакет складається з семи підпакетів та однієї функції importdata, за якою імпортуються дані з файлу (див.табл.1).

Табл.1

Пакет	Призначення
anova	Дисперсійний аналіз
describe	Обчислення числових характеристик масивів даних
fit	Регресійний аналіз
random	Генерація випадкових величин за заданим законом розподілу
statevalf	Обчислення статистичних функцій та оцінок масивів даних
statplots	Графічне подання статистичних даних
transform	Функції перетворення даних

Розглянемо команди підпакету statevalf. Формат виклику команд:

stats[statevalf, function, distribution](args);

або

statevalf[function, distribution](args);

де function може набувати одного з наступних значень:

1) для функцій неперервних розподілів ймовірностей:

а) cdf – функція неперервного розподілу ймовірностей (інтегральна функція);

б) pdf – щільність розподілу ймовірностей (диференціальна функція);

2) для функцій дискретних розподілів ймовірностей:

а) dcdf – інтегральна функція розподілу ймовірностей;

б) pf – розподіл ймовірностей;

a distribution – певний розподілу ймовірностей.

Необхідність статистичного опрацювання даних виникає тоді, коли задано деякий масив даних – вибірка. Для задання вибірки, яка складається з набору випадкових спостережених значень, використовуються команди підпакету random, формат виклику яких наступний:

random[distribution](quantity,uniform,method)

де distribution – опис розподілу ймовірностей, на множині якої отримуються числа, quantity – опція, яerez яку задається додатне ціле число (за замовчуванням дорівнює 1), за яким визначається, скільки випадкових чисел потрібно отримати, uniform – опція, за якою генеруються числа з рівномірним розподілом ймовірностей (за замовчуванням 'default'), method – одне з наступних ключових слів 'auto', 'inverse' або 'builtin' (за замовчуванням 'auto').

Для аналізу статистичних даних досить важливими є числові характеристики вибірки: середнє арифметичне спостережених значень, середнє гармонійне, середнє геометричне, середнє квадратичне, мода, медіана, а також показники розсіювання спостережених значень випадкової величини – розмах вибірки, середнє абсолютне відхилення, коефіцієнт варіації тощо. За допомогою команд підпакету describe можна обчислювати їх. Формат виклику функцій є таким:

stats[describe, function](args);

або

describe[function](args);

де args – масив даних, function – одна з команд підпакету describe.

Можна використовувати безпосередньо команду для дослідження статистичних даних, попередньо звернувшись до потрібного підпакету пакету stats за допомогою команди

```
with(stats, subpackage);
```

де subpackage – підпакет, у якому містяться команди, необхідні для проведення статистичного дослідження.

Наведемо приклад задання статистичних даних та застосування деяких команд для визначення інтервалу зміни даних, кількості елементів у списку даних та середнього значення:

```
> s1:=[1,2,,1,2.2,5,1,4,2.5,5.4,4.1];#задання статистичних даних
      s1 := [ 1, 2, 0.1, 2.2, 5, 1, 4, 2.5, 5.4, 4.1 ]
> describe[range](s1);#діапазон зміни даних (розмах)
      0.1 .. 5.4
> describe[count](s1);#кількість елементів у списку даних
      10
> describe[mean](s1);#середнє арифметичне
      2.730
> describe[variance](s1);#дисперсія
      2.953
> describe[median](s1);#медіана
      2.350
> transform[statsort](s1);#сортування у порядку зростання
      [0.1, 1, 1, 2, 2.2, 2.5, 4, 4.1, 5, 5.4]
```

Розглянемо кілька прикладів.

Приклад 1. Під час дослідження кількісної ознаки X з генеральної сукупності було отримано вибірку

4,3,6,4,7,2,5,1,2,5,4,4,3,5,6,3,4,1,3,4.

Знайти обсяг вибірки, побудувати варіаційний ряд вибірки та ряд дискретного розподілу статистичних ймовірностей.

Розв’язування. Задамо вибірку

```
> X:=[4,3,6,4,7,2,5,1,2,5,4,4,3,5,6,3,4,1,3,4];
      X := [ 4, 3, 6, 4, 7, 2, 5, 1, 2, 5, 4, 4, 3, 5, 6, 3, 4, 1, 3, 4 ]
```

Обсяг вибірки обчислимо за функцією count:

```
> n:=count(X);
      n := 20
```

Варіаційний ряд (впорядкований за зростанням набір варіант) будуюмо за функцією sort:

```
> Y:=sort(X);
      Y := [ 1, 1, 2, 2, 3, 3, 3, 3, 4, 4, 4, 4, 4, 4, 5, 5, 5, 6, 6, 7 ]
```

Для побудови ряду дискретного розподілу статистичних ймовірностей необхідно встановити відповідність між можливими значеннями випадкової величини X та їхніми статистичними ймовірностями. Для цього за функцією tally побудуємо множину, елементами якої є вирази Weight(val,k), де val – значення варіанти, k – її частота.

Xm:=tally(X);

$X_m := [\text{Weight}(1, 2), \text{Weight}(2, 2), \text{Weight}(3, 4), \text{Weight}(4, 6), \text{Weight}(5, 3), \text{Weight}(6, 2), 7]$

За функцією statvalue визначимо множину різних елементів вибірки X :

> **X1:=statvalue(Xm);**

$X_1 := [1, 2, 3, 4, 5, 6, 7]$

За функцією frequency визначаємо частоти для елементів вибірки X :

> **m:=frequency(Xm);**

$m := [2, 2, 4, 6, 3, 2, 1]$

Далі визначаємо статистичні ймовірності:

> **p:=m/n;**

$p := \left[\frac{1}{10}, \frac{1}{10}, \frac{1}{5}, \frac{3}{10}, \frac{3}{20}, \frac{1}{10}, \frac{1}{20} \right]$

Ряд дискретного розподілу статистичних ймовірностей для вибірки X буде мати вигляд:

> **XP:={};**

for i to k do

XP:=XP union {{X1[i],p[i]}};

end do:XP;

$\{ \{4, \frac{3}{10}\}, \{2, \frac{1}{10}\}, \{1, \frac{1}{10}\}, \{3, \frac{1}{5}\}, \{5, \frac{3}{20}\}, \{6, \frac{1}{10}\}, \{7, \frac{1}{20}\} \}$

Приклад 2. При визначенні похибки вимірювального приладу зафіксовано такі похибки (табл.2).

Табл.2

i	1	2	3	4	5	6	7	8	9	10
x_i	-2,5	3	4	2	0,5	-1	2	4	-4	0
i	11	12	13	14	15	16	17	18	19	20
x_i	-0,5	-0,5	1	0,5	2,5	-0,5	2	1	-4	-2
i	21	22	23	24	25	26	27	28	29	30
x_i	-1	1,5	0,5	4	-1,5	-1	0	1	0	1
i	31	32	33	34	35	36	37	38	39	40
x_i	-1,5	1,5	0,5	0,5	-0,5	-1,5	-0,5	-1	2	0,5

За цими даними випробувань побудувати інтервальний розподіл статистичних ймовірностей та гістограму, якщо $k = 8$ [4, с.421].

Розв'язування. Задамо вибірку X та обчислимо її обсяг n :

> **X:=[-2.5,3,4,2,0.5,-1,2,4,-4,0,-0.5,-0.5,1,0.5,2.5,-0.5,2,1,-4,**

-2, -1,1.5,0.5,4,-1.5,-1,0,1,0,1,-1.5,1.5,0.5,0.5,-0.5,

-1.5,-0.5,-1,2,0.5];#вибірка

n:=count(X);#обсяг вибірки

$n := 40$

Варіаційний ряд Y (впорядкований за зростанням набір варіант) будуємо за функцією statsort з підпакету transform:

```
> Y:=statsort(X);
Y := [-4, -4, -2.5, -2, -1.5, -1.5, -1.5, -1, -1, -1, -1, -0.5, -0.5, -0.5, -0.5, -0.5, 0, 0, 0, 0.5,
0.5, 0.5, 0.5, 0.5, 0.5, 1, 1, 1, 1, 1.5, 1.5, 2, 2, 2, 2, 2.5, 3, 4, 4, 4]
```

Задамо кількість проміжків k для групування та визначимо розмах вибірки R :

```
> k:=8:#кількість проміжків для групування
R:=Y[n]-Y[1];#Розмах вибірки
R := 8
```

Визначимо довжину h часткового проміжку $[x_{i-1}, x_i)$, $i \in \overline{1, k}$:

```
> #довжина проміжку
h:=R/k;
h := 1
```

Обчислюємо межі часткових проміжків $[x_{i-1}, x_i)$, $i \in \overline{1, k}$:

```
> #межі часткових проміжків
xr:=[Y[1]+i*h $i=0..k]:
xr[k+1]:=xr[k+1]+0.001:
xr;
[-4, -3, -2, -1, 0, 1, 2, 3, 4.001]
```

та визначаємо часткові проміжки:

```
> xrr:=(xr[i]..xr[i+1]) $i=1..k;#проміжки
xrr := [-4 .. -3, -3 .. -2, -2 .. -1, -1 .. 0, 0 .. 1, 1 .. 2, 2 .. 3, 3 .. 4.001]
```

Будуємо інтервальний розподіл статистичних ймовірностей:

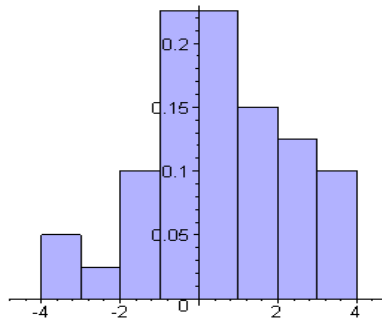
```
> xp:=scaleweight[1/n](statsort(tallyinto(Y,xrr)));
xp := [ Weight(-4 .. -3, 1/20), Weight(-3 .. -2, 1/40), Weight(-2 .. -1, 1/10),
Weight(-1 .. 0, 9/40), Weight(0 .. 1, 9/40), Weight(1 .. 2, 3/20), Weight(2 .. 3, 1/8),
Weight(3 .. 4.001, 1/10) ]
```

Останній результат слід розуміти так:

$[x_{i-1}, x_i)$	$[-4, -3)$	$[-3, -2)$	$[-2, -1)$	$[-1, 0)$	$[0, 1)$	$[1, 2)$	$[2, 3)$	$[3, 4, 00.1)$
$P_n^*([x_{i-1}, x_i))$	$\frac{1}{20}$	$\frac{1}{40}$	$\frac{1}{10}$	$\frac{9}{40}$	$\frac{9}{40}$	$\frac{3}{20}$	$\frac{1}{8}$	$\frac{1}{10}$

Для побудови гістограми розподілу статистичних ймовірностей використовуємо функцію histogram з підпакету statplots:

```
> #побудова гістограми
statplots[histogram](xp);
```



Приклад 3. 20 навмання вибраних студентів виконують стрибки у висоту. При цьому зафіксовано такі результати (в сантиметрах) (табл. 3) [4, с.423]

Табл.3

i	1	2	3	4	5	6	7	8	9	10
x_i	137	140	143	135	142	139	141	137	142	131
i	11	12	13	14	15	16	17	18	19	20
x_i	145	138	141	143	130	138	140	135	137	138

Розв’язування. Ряд розподілу статистичних ймовірностей буде мати вигляд:

> $X := [137, 140, 143, 135, 142, 139, 141, 137, 142, 131, 145, 138, 141, 143, 130, 138, 140, 135, 137, 138]$;

$n := \text{count}(X)$;

$Xm := \text{tally}(X)$;

$\text{transform}[\text{scaleweight}[1/n]](Xm)$;

$\left[\text{Weight}\left(130, \frac{1}{20}\right), \text{Weight}\left(131, \frac{1}{20}\right), \text{Weight}\left(135, \frac{1}{10}\right), \text{Weight}\left(137, \frac{3}{20}\right), \right.$
 $\text{Weight}\left(138, \frac{3}{20}\right), \text{Weight}\left(139, \frac{1}{20}\right), \text{Weight}\left(140, \frac{1}{10}\right), \text{Weight}\left(141, \frac{1}{10}\right),$
 $\left. \text{Weight}\left(142, \frac{1}{10}\right), \text{Weight}\left(143, \frac{1}{10}\right), \text{Weight}\left(145, \frac{1}{20}\right) \right]$

що слід розуміти так:

x_i	130	131	135	137	138	139	140	141	142	143	145
$P_n^*(x_i)$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{10}$	$\frac{3}{20}$	$\frac{3}{20}$	$\frac{1}{20}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{20}$

Для побудови багатокутника розподілу статистичних ймовірностей (полігона частот) за функціями statvalue та frequency відповідно визначаємо різні елементи вибірки X та їх статистичні ймовірності p :

> $X1 := \text{statvalue}(xx)$;

$X1 := [130, 131, 135, 137, 138, 139, 140, 141, 142, 143, 145]$

> $m := \text{frequency}(Xm)$;

> $p := m/n$;

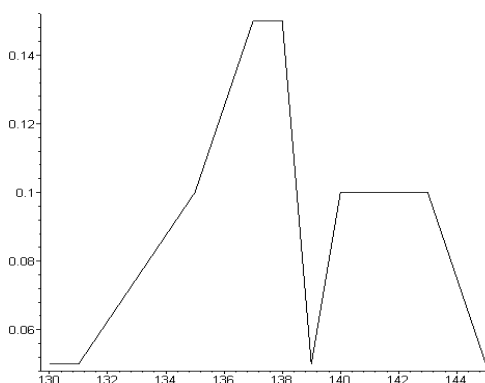
$p := \left[\frac{1}{20}, \frac{1}{20}, \frac{1}{10}, \frac{3}{20}, \frac{3}{20}, \frac{1}{20}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{10}, \frac{1}{20} \right]$

За функцією seq сформуємо список pnt, що містить пари елементів (x_i, p_i) , де x_i – значення варіанти, а p_i – відповідна їй статистична ймовірність.

```
>pnt:= [seq([X1[i],p[i]],i=1..nops(p))];
pic1 := [[ [130, 1/20], [131, 1/20], [135, 1/10], [137, 3/20], [138, 3/20], [139, 1/20], [140, 1/10],
           [141, 1/10], [142, 1/10], [143, 1/10], [145, 1/20]]]
```

За функцією listplot будуємо полігон відносних частот (многокутник розподілу статистичних ймовірностей):

```
> with(plots):# звернення до команд пакету
listplot(pnt);# побудова полігона
```



Приклад 4. Використовуючи критерій Пірсона, перевірити гіпотезу про нормальний розподіл ймовірностей при рівні значущості $\alpha = 0.05$, якщо відомі емпіричні частоти (табл.4) [2, с.333-334]:

Табл.4

x_i	5	7	9	11	13	15	17	19	21
n_i	15	26	25	30	26	21	24	20	13

Розв’язування. Значення варіанти та її частоти зручно задавати за командою Weight(значення_варіанти, частота). Введемо заданий статистичний розподіл:

```
>W:= [Weight(5,15),Weight(7,26),Weight(9,25),Weight(11,30),Weight(13,26),
Weight(15,21),Weight(17,24),Weight(19,20),Weight(21,13)];
W := [ Weight ( 5, 15), Weight ( 7, 26), Weight ( 9, 25), Weight ( 11, 30), Weight ( 13, 26),
       Weight ( 15, 21), Weight ( 17, 24), Weight ( 19, 20), Weight ( 21, 13) ]
```

Визначимо кількість елементів у вибірці W:

```
> N:=describe[count](W);
N := 200
```

вибіркове середнє:

```
> xB:=describe[mean](W):evalf(%)
12.63
```

та вибіркове середньоквадратичне відхилення:

```
> q:=describe[standarddeviation](W):evalf(%)
4.695
```

Виокремимо списки варіант та частот:

```
> x:=transform[statvalue](W);
      x := [ 5, 7, 9, 11, 13, 15, 17, 19, 21 ]
> n:=transform[frequency](W);
      n := [ 15, 26, 25, 30, 26, 21, 24, 20, 13 ]
```

Обчислюємо гіпотетичні частоти за формулою $m_i = \frac{N \cdot h}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(x_i - \bar{x}_B)^2}{2\sigma^2}\right)$, де

N – кількість елементів у вибірці (сума всіх частот), h – крок (різниця між сусідніми варіантами), σ – середнє квадратичне відхилення, \bar{x}_B – вибіркове середнє:

```
> h:=2:m:=[seq(N*h*exp(-(x[i]-xB)^2/(2*q^2))/(q*sqrt(2*Pi)),i=1..9)]:
evalf(m);
      [9.068, 16.55, 25.20, 31.98, 33.87, 29.90, 22.02, 13.53, 6.934]
```

Обчислюємо спостережене значення величини χ_{ci}^2 :

```
> sum((n[i]-m[i])^2/m[i],i=1..9):evalf(%)
      22.46
```

За таблицею критичних точок розподілу χ^2 за рівнем значимості $\alpha = 0.05$ і кількістю ступенів вільності $k = s - 3 = 9 - 3 = 6$ відшукуємо $\chi_{\alpha\delta}^2(0.05, 6) = 12.6$.

Оскільки $\chi_{ci}^2 > \chi_{\alpha\delta}^2$ – гіпотеза про нормальний розподіл ймовірностей відхиляється як така, що не узгоджується з емпіричними даними.

Висновок. У статті охарактеризовано можливості використання системи тільки Maple для первинного опрацювання статистичних даних. Проаналізовано тільки незначна частина функцій пакету stats для опрацювання експериментальних даних – можливості використання СКМ Maple значно більші. З використанням системи Maple можна також здійснювати дисперсійний, кореляційно-регресійний аналізи. Крім того, у системі Maple міститься й інший пакет для здійснення статистичного аналізу – Statistics. Власне на це й будуть спрямовані подальші дослідження.

БІБЛІОГРАФІЯ:

1. Бабенко В.В. Основи теорії ймовірностей і статистичні методи аналізу даних у психологічних і педагогічних експериментах: навч. посібник / В.В. Бабенко. – Львів: Видавничий центр ЛНУ імені Івана Франка, 2009. – 184 с.
2. Гмурман В.Е. Теория вероятностей и математическая статистика / В.Е. Гмурман. – М., Высш.шк., 2003. – 479 с.
3. Жалдак М.І. Математика з комп'ютером: посібник для вчителів. — 2-ге вид. / Жалдак М.І., Горошко Ю.В., Вінниченко Є. Ф. — К. : НПУ імені Драгоманова, 2009. — 282 с.
4. Жалдак М.І. Теорія ймовірностей і математична статистика: Підручник для студентів фізико-математичних спеціальностей педагогічних університетів. – Вид.2, перероб. і доп. / Жалдак М.І., Кузьміна Н.М., Михалін Г.О. – Полтава: Довкілля-К, 2009. – 500 с.
5. Лазурчак І.І. Система комп'ютерної математики: навч. посібник / І.І. Лазурчак, Т.П. Кобильник. – Дрогобич: Коло, 2013. – 256 с.
6. Лупан І.В. Комп'ютерні статистичні пакети: навчально-методичний посібник / І.В. Лупан, О.В. Авраменко. – Кіровоград, 2010. – 218 с.
7. Мамчич Т.І. Статистичний аналіз даних з пакетом STATISTICA. Навчально-методичний посібник / Мамчич Т.І., Оленко А.Я., Осипчук М.М., Шпортюк В.Г. – Дрогобич: Видавнича фірма «Відродження», 2006. – 208 с.

8. Руденко В. М. Математична статистика. Навч. посіб. / В.М. Руденко – К.: Центр учбової літератури, 2012. – 304 с.

ВІДОМОСТІ ПРО АВТОРІВ

Кобильник Тарас Петрович – доцент, кандидат педагогічних наук, доцент кафедри інформатики та обчислюваної математики Дрогобицького державного педагогічного університету імені Івана Франка

Коло наукових інтересів: математична інформатика, системи комп'ютерної математики, web-орієнтоване програмне забезпечення, методика навчання інформатики.

ОСОБЛИВОСТІ РОЗРОБКИ КРИТЕРІЇВ ОЦІНЮВАННЯ ЕЛЕКТРОННИХ ОСВІТНІХ РЕСУРСІВ

Світлана ЛИТВИНОВА

У статті проаналізовано стан та перспективи використання електронних освітніх ресурсів (ЕОР) у загальноосвітніх навчальних закладах, розкрито особливості аналізу, оцінювання та систематизації вимог до використання електронних освітніх ресурсів, визначено особливості розробки критеріїв та структурні елементи ЕОР (змістовна, програмна, методична), вимоги до структурних елементів їх вмісту; до особливостей ЕОР віднесено режим конструктора уроку.

The paper analyzes the status and prospects of EER in secondary schools, disclosed terms of analysis, evaluation and systematization requirements for electronic educational resources peculiarities of criteria of design and structural elements of EER (content, software and methodology), the requirements for the structural elements of their contents, to EER include constructor features designer lesson.

Постановка проблеми. Відмінною особливістю сучасної загальної середньої освіти є активна реалізація можливостей інформаційних і комунікаційних технологій (ІКТ) у підвищенні якості методичного та дидактичного забезпечення навчально-виховного процесу, вивленні обдарованості та розвитку особистості учня. У сучасній школі більшість видів навчальної діяльності вимагають від вчителів готовності до застосування засобів ІКТ у своїй професійній діяльності. У цих умовах одним з пріоритетних напрямів модернізації загальної середньої освіти є широке впровадження засобів ІКТ у процес навчання і виховання, зокрема, використання електронних освітніх ресурсів (ЕОР) як під час проведення уроків, так і у позаурочний час. Барретт Крейг, президент та виконавчий директор корпорації Intel, підкреслює, що всі освітні технології нічого не варті, якщо вчителі не знають, як ними ефективно користуватися. Дива в освіті творять не комп'ютери, а вчителі [5, с.18].

Педагоги все частіше піднімають питання про відповідність ЕОР психолого-педагогічним вимогам. Однією із основних причин такої стурбованості є недостатня розробленість теоретичних засад оцінювання якості ЕОР. Відкритий доступ до ЕОР створив умови для аналізу, оцінювання та систематизації вимог, які вчителі формують під час активного використання ресурсів. Виникає необхідність обґрунтування критеріїв оцінювання, дослідження методів комплексної оцінки якості, визначення та апробація дієвих методик встановлення відповідності електронних засобів і технологій навчального призначення певним об'єктивним психолого-педагогічним вимогам до їх якості.