

Scientific journal
PHYSICAL AND MATHEMATICAL EDUCATION
Has been issued since 2013.

Науковий журнал
ФІЗИКО-МАТЕМАТИЧНА ОСВІТА
Видається з 2013.

ISSN 2413-158X (online)
ISSN 2413-1571 (print)



<http://fmo-journal.fizmatsspu.sumy.ua/>

Кобильник Т.П. Методичні аспекти навчання множинного лінійного регресійного аналізу з використанням статистичного середовища R. Фізико-математична освіта. 2018. Випуск 1(15). С. 57-62.

Kobylnyk T. Methodological Aspects Of Learning Multiple Linear Regression Analysis With R Package. Physical and Mathematical Education. 2018. Issue 1(15). P. 57-62.

УДК 378:004.6

Т.П. Кобильник

Дрогобицький державний педагогічний університет імені Івана Франка, Україна

kobylnyktaras@gmail.com

DOI 10.31110/2413-1571-2018-015-1-008

МЕТОДИЧНІ АСПЕКТИ НАВЧАННЯ МНОЖИННОГО ЛІНІЙНОГО РЕГРЕСІЙНОГО АНАЛІЗУ З ВИКОРИСТАННЯМ СТАТИСТИЧНОГО СЕРЕДОВИЩА R

Анотація. Однією з важливих умов підготовки вчителів інформатики є вміння та навички використовувати різноманітні статистичні методи аналізу експериментальних даних. Упровадження статистичних методів аналізу у навчальний процес дає можливість підвищити рівень підготовки майбутнього педагога.

Методи регресійного аналізу використовуються в різних галузях науки для визначення вигляду залежностей між досліджуваними ознаками. Регресійний аналіз є основним статистичним методом побудови математичних моделей об'єктів або явищ на основі експериментальних даних. Основні результати в даний час отримані стосовно до лінійних регресійних моделей, які можуть служити як початковий етап дослідження.

Опрацювання даних неможливе без використання комп'ютера з відповідним програмним забезпеченням. Перед користувачем виникає проблема вибору програмного забезпечення для дослідження. Пропонується використовувати статистичне середовище R для наукових досліджень та супроводу навчального процесу у вищих навчальних закладах. Для проведення регресійного аналізу у статті використано вільно поширюваний пакет R – однією з кращих програм для проведення статистичного аналізу. Нелінійні зв'язки за певними перетвореннями (заміною змінних чи логарифмуванням) можна звести до лінійного вигляду, тобто апроксимувати відповідні залежності лінійними функціями. Модель лінійної регресії є найпростішим і найчастіше використовуваним видом залежності між змінними. Тому під час вивчення елементів регресійного аналізу значну увагу слід приділити лінійній моделі. У статті на конкретному прикладі показано побудову та дослідження множинної лінійної регресійної моделі з використанням статистичного середовища R. Перевірку на відповідність нормальному розподілу проведено за допомогою побудови q-q діаграми.

Перспективи подальших досліджень будуть спрямовані на вивчення можливостей використання пакету R для статистичного аналізу даних та методиці навчання основ імітаційного моделювання студентів інформатичних спеціальностей в педагогічних університетах.

Ключові слова: множинний лінійний регресійний аналіз, статистичний пакет R.

Постановка проблеми. За допомогою методів дисперсійного та кореляційного аналізів виявляють наявність зв'язку між випадковими величинами та оцінюють силу цього зв'язку. Наступним кроком є встановлення певної залежності між випадковими величинами. Це здійснюється методами регресійного аналізу. Розрізняють парний та множинний регресійні аналізи. Студенти починаються знайомитися з парним регресійним аналізом, як правило, лінійним. Це пояснюється тим, що значну кількість процесів можна описати лінійною функцією або звести функцію до лінійного вигляду (наприклад, логарифмування чи заміною змінних). Проте у дослідженнях багатьох процесів виявляється, що деяка результативна ознака змінюється під впливом не одного, а кількох факторів. Таким чином, встановлення залежності між випадковими величинами здійснюється за допомогою множинного регресійного аналізу.

Опрацювання експериментальних даних неможливе без використання комп'ютерів з відповідним програмним забезпеченням. Існує широкий вибір програмного забезпечення як загального, так спеціального призначення для опрацювання експериментальних даних. Стандартні статистичні методи опрацювання експериментальних даних реалізовані в електронних таблицях та табличних процесорах (Lotus, QuatroPro, MS Excel, OpenOffice.org Calc та ін.), системах комп'ютерної математики (Gran1, Maple, MathCAD, Mathematica, Matlab, Maxima та ін.), спеціалізованих статистичних пакетах (R, IBM SPSS, Statistica, SAS та ін.).

Аналіз актуальних досліджень. У монографії [17] сформульовано проблеми і завдання створення методології регресійного аналізу стосовно отримання багатфакторних статистичних моделей прикладних складних систем.

Охарактеризовано лінійний регресійний аналіз та наведено рекомендації щодо його використання у випадках, якщо передумови аналізу певною мірою не виконуються. Книга [8] – повне класичне введення в фундаментальні основи множинного регресійного аналізу. У ній описуються методи вибору і дослідження лінійних і нелінійних регресійних моделей, а також розглядаються практичні аспекти їх застосування, в тому числі з використанням спеціалізованого програмного забезпечення, зокрема Minitab. У статті [6] сформульована та обґрунтована конструктивна схема побудови та дослідження економетричної моделі множинної регресії та продемонстровано основні етапи її побудови та дослідження.

Посібник [4] має чітке практичне спрямування: розглядається як практикум із застосування табличного процесора MS Excel для дослідження задач економетричного моделювання та математичного програмування. Усі приклади, що розглядаються у ньому, реалізовано у середовищі табличного процесора MS Excel 2007. Також у посібнику наведено основні відмінності версій MS Excel 2010 та MS Excel 2013 від MS Excel 2007 і особливості використання наведеного у посібнику інструментарію для MS Excel 2010/2013 у порівнянні з MS Excel 2007. Автори вважають MS Excel «...оптимальним програмним засобом для широкого кола користувачів» [4, с.7]. Навчальний посібник [9] містить лабораторний практикум з первинного опрацювання даних, побудови множинних лінійних і нелінійних моделей, оцінювання ступеня мультиколінеарності та методів її усунення, побудови економетричних моделей динаміки, моделей з лаговими змінними, систем одночасних рівнянь за допомогою системи Statistica. Навчальний посібник [5] містить теоретичні відомості про основні математично-статистичні методи, що використовуються в прикладних дослідженнях, та шляхи їх реалізації засобами Microsoft Excel та SPSS. У навчальному посібнику [2] наведені прийоми роботи з універсальними статистичними пакетами Statgraphics, SPSS, Statistica. У навчальному посібнику [3] окремий розділ відведено для опису використання системи SAS для розв'язування задач статистичного опрацювання, дослідженню розподілів, змінних аналізу та використанню ANOVA аналізу.

У монографії [7, с.303-313] наведено приклади використання програми Gran1 [11] для визначення та дослідження математичних моделей результатів певних експериментів, застосовуючи метод найменших квадратів для відшукування аналітичного виразу функції, значення якої якомога менше відхиляється від одержаних експериментально значень у заданих точках.

У книзі [1] аналізуються можливості використання електронних таблиць OpenOffice.org Calc на прикладах задач з курсів економічного циклу. У посібнику [13] показано у порівнянні пакетів SPSS, Statistica та табличного процесора MS Excel для обчислення параметрів розподілів ймовірностей випадкової величини, виконання параметричного та непараметричного порівняння двох та більше зв'язних та незв'язних вибірок, у тому числі одно- та двофакторного дисперсійного аналізу, критерії порівняння емпіричних розподілів та перевірки нормальності емпіричних розподілів випадкової величини, виконання дискримінантного та кластерного аналізу; інструментарій для графічного подання результатів та засоби формування звітів. У навчальному посібнику [10] показано можливості практичного застосування комп'ютерних програм, зокрема, табличного процесора MS Excel і прикладних програм Statistica і EViews для моделювання діяльності підприємств.

У навчальному посібнику [14] описано статистичний аналіз даних за допомогою програмного середовища R. Містить загальні рекомендації з користування пакетом. Особливу увагу приділено застосуванню R до прикладних задач математичної статистики, а також розбору прикладів застосування до реальних даних.

Мета статті: методичні аспекти використання статистичного середовища R у процесі навчання лінійного множинного регресійного аналізу.

Виклад основного матеріалу. Як видно з аналізу програмного забезпечення, яке використовується для опрацювання експериментальних даних, часто обирають електронні таблиці (зокрема, MS Excel). Використання табличних процесорів доцільне у випадках, коли необхідно виконувати прості статистичні операції, зокрема побудову варіаційного ряду, обчислення описових статистик, побудови деяких типів графіків, проведення деяких видів аналізу. Тобто повноцінне статистичне опрацювання експериментальних даних засобами табличних процесорів неможливе: це офісні програми, а не спеціалізовані, наукові. Для реалізації складних статистичних процедур в електронних таблицях користувачу необхідні глибокі знання всіх обчислювальних алгоритмів, значне збільшення часу і виконання трудомістких обчислень. Тому використання MS Excel навіть при вивченні статистичних методів аналізу, зокрема факторного, дискримінантного, кластерного, видається, на наш погляд, мало ефективним.

Тому для навчання студентів методів статистичного аналізу рекомендується використовувати статистичне середовище R - вільне програмне середовище з відкритим кодом, що поширюється в рамках проекту GNU [16]. Навчання статистичного аналізу з використанням R і всесвітня підтримка науковим співтовариством обумовили те, що наведення скриптів R поступово стає загальноновизнаним «стандартом» як в журнальних публікаціях, так і в неформальному спілкуванні науковців усього світу [15]. Додаткової популярності R дало створення центральної системи зберігання і розповсюдження пакетів – CRAN (Comprehensive R Archive Network — <http://cran.r-project.org>). У статті [12] як приклад, наведено використання у процесі навчання вищої математики статистичного середовища R, проаналізовано його переваги та недоліки.

Множинна лінійна регресійна модель має вигляд $Y = X\beta + \varepsilon$, де Y – вектор значень залежної змінної, X – матриця значень факторів, β – параметри рівняння (вектор коефіцієнтів рівняння), ε – випадкова змінна (похибка). Вважається, що зв'язок між змінними X та Y , лінійний, тобто існує деяка залежність $Y = Xb$. Задача регресійного аналізу полягає у знаходженні оцінок b_0, b_1, \dots, b_n вектора коефіцієнтів β регресійної моделі. Величина ε показує відхилення значень емпіричних даних від прямої регресії і може бути пов'язаною з похибками вимірювань, неправильно вибраною формою залежності між змінними тощо.

Лінійний регресійний аналіз у R реалізується за допомогою функції `lm`, синтаксис якої є таким:

`lm(formula, data,...)`,

де `formula` – оцінкова модель залежності; `data` – дані, ... - інші додаткові параметри.

Приклад 1. Для деякого підприємства отримано такі результати вибірових статистичних спостережень за останні 24 місяці (2 роки), що містять дані про продуктивність праці та фактори, що на неї впливають (див. таблиця 1) [4, с.56-57].

Таблиця 1.

Дані про продуктивність праці та фактори, що на неї впливають

	Y продуктивність праці, гр.од/люд.-год	X1 Фондомісткість, гр.од/гр.од	X2 Коефіцієнт плинності робочої сили, %	X3 Рівень втрат робочого часу, %
1	60	30	13	15
2	61	35	12,5	14,3
3	58	33	12	12
4	59	34	11	12,8
5	62	36	10	13
6	63	38	9	12,5
7	65	40	8,5	11
8	60	41	8,2	11,5
9	68	45	8	10
10	69	45	5,5	9
11	70	46	5	8
12	72	48	4,7	7,5
13	73	47	4,6	6,5
14	78	50	4	6
15	75	49	4,1	6,2
16	80	51	4,2	5,8
17	81	50	4,5	5,5
18	83	53	4	5
19	81	55	4	4,5
20	85	56	3	4,7
21	87	58	4	5
22	88	58	5	5,1
23	90	59	5	4,8
24	92	60	6	5,2

У припущенні щодо лінійної залежності між наведеними показниками побудувати та дослідити лінійну регресійну модель залежності між продуктивністю праці та іншими факторами.

Розв'язування. Задамо дані таблиці 1 через вектори Y, X1, X2, X3, використовуючи функцію c (рис. 1). Для функції lm як параметр вибрана лінійна модель з вільним членом (Y~X1+X2+X3).

```
Y<-c(60,61,58,59,62,63,65,60,68,69,70,72,73,78,75,80,81,83,81,85,87,88,90,92)
X1<-c(30,35,33,34,36,38,40,41,45,45,46,48,47,50,49,51,50,53,55,56,58,58,59,60)
X2<-c(13,12.5,12,11,10,9,8.5,8.2,8,5.5,5,4.7,4.6,4,4.1,4.2,4.5,4,4,3,4,5,5,6)
X3<-c(15,14.3,12,12.8,13,12.5,11,11.5,10,9,8,7.5,6.5,6,6.2,5.8,5.5,5,4.5,4.7,5,5.1,4.8,5.2)
fit=lm(Y~X1+X2+X3);
summary(fit)
```

Рис. 1

Результати регресійного аналізу наведено на рис. 2. Пояснимо їх. Застосовуючи функцію summary до об'єктів функції lm, отримуємо формулу моделі, залишки (Residuals), коефіцієнти оцінкового рівняння, середнє квадратичне відхилення оцінки регресії, коефіцієнт детермінації R² та скоригований коефіцієнт детермінації, F-статистику.

Рівняння регресії має вигляд $Y = 27.7103 + 1.0456X_1 + 1.5142X_2 - 1.5845X_3$. Оцінка стандартного відхилення залишків (Residual standard error) дорівнює 2.332. Вважається, що залишки нормально розподілені із середнім значенням 0 і стандартним відхиленням σ . Власне у цьому рядку і виводиться оцінка значення σ . За значенням коефіцієнта детермінації (Multiple R-squared) та скоригованого коефіцієнта детермінації (Adjusted R-squared), близькими до одиниці (0.9606 та 0.9547) можна зробити висновок про те, що приблизно 96% варіації залежної змінної пояснюється варіаціями незалежних змінних, тобто зміни в продуктивності праці на 96% пояснюються змінами у фондомісткості продукції, коефіцієнтом плинності робочої сили та рівнем втрат робочого часу. Стандартні похибки (Std.Error) параметрів $b_0 = 27.7103$, $b_1 = 1.0456$, $b_2 = 1.5142$, $b_3 = -1,5845$ дорівнюють 12.3718, 0.1833, 0.4756 та 0.6716 відповідно. Ці

значення разом з t-розподілом Стюдента використовують для перевірки істотності оцінок коефіцієнтів регресії та для визначення їхніх надійних інтервалів. Зі стовпця (Pr(>|t|)) видно, що коефіцієнти регресії $b_0 = 27.7103$, $b_1 = 1.0456$, $b_2 = 1.5142$, $b_3 = -1,5845$ за t-критерієм Стюдента статистично відрізняються від нуля (ймовірність відхилення гіпотез про рівність отриманих оцінок істинному значенню: $P(\beta_0 \neq b_0) = 0.03662 < 0.05$, $P(\beta_1 \neq b_1) = 1.34 \cdot 10^{-5} < 0.05$, $P(\beta_2 \neq b_2) = 0.00466 < 0.05$, $P(\beta_3 \neq b_3) = 0.02859 < 0.05$). Адекватність побудованої моделі оцінюється за допомогою F-критерію Фішера. Гіпотезу про відсутність лінійної залежності можна прийняти з ймовірністю $p = 3.302 \cdot 10^{-14}$, тобто з надійністю можна стверджувати, що модель є адекватною щодо статистичних даних, тобто рівняння регресії є статистично значимим.

```
Call:
lm(formula = Y ~ X1 + X2 + X3)

Residuals:
    Min       1Q   Median       3Q      Max
-4.8985 -0.7640  0.1765  1.4982  4.9140

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  27.7103    12.3718   2.240  0.03662 *
X1           1.0486     0.1833   5.721 1.34e-05 ***
X2           1.5142     0.4756   3.184  0.00466 **
X3          -1.5845     0.6716  -2.359  0.02859 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.332 on 20 degrees of freedom
Multiple R-squared:  0.9606, Adjusted R-squared:  0.9547
F-statistic: 162.4 on 3 and 20 DF, p-value: 3.302e-14
```

Рис. 2

Застосовуючи функцію confint до об'єктів функції lm, отримуємо межі надійних інтервалів для параметрів $\beta_0, \beta_1, \beta_2, \beta_3$ (рис. 3).

```
> confint(fit)
                2.5 %      97.5 %
(Intercept)  1.9033107 53.5173621
X1           0.6663091 1.4309326
X2           0.5221913 2.5061935
X3          -2.9854687 -0.1835257
```

Рис. 3

Межі надійних інтервалів є такими:

$$1.9033107 < \beta_0 < 53.5173621,$$

$$0.6663091 < \beta_1 < 1.4309326,$$

$$0.5221913 < \beta_2 < 2.5061635,$$

$$-2.9854687 < \beta_3 < -0.1835257.$$

Аналіз меж надійних інтервалів для параметрів $\beta_0, \beta_1, \beta_2, \beta_3$ дозволяє зробити висновок, що з ймовірністю $p = 1 - \alpha = 0.95$, знаходячись у вказаних межах, вони не набувають нульового значення, тобто вони є статистично значимим та суттєво відрізняються від нуля (це підтверджується і перевіркою за t-критерієм Стюдента).

Після побудови моделі та дослідження її на адекватність, доцільно провести аналіз залишків для порівняння з нормальним розподілом ймовірностей, тобто перевірити гіпотезу про нормальний розподіл ймовірностей на множині значень похибок моделі. Для цього побудуємо q-q діаграму (рис. 4).

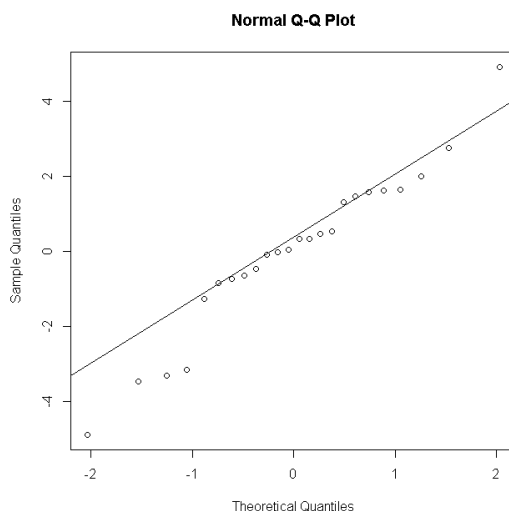


Рис. 4

Як видно з рис. 4, значення залишків не значно відхиляються від нормального розподілу. Студентам можна запропонувати побудувати та дослідити лінійну модель без вільного члена ($Y \sim X_1 + X_2 + X_3 + 0$).

Висновки. У статті проаналізовано можливості використання статистичного пакету R для вивчення множинного регресійного аналізу, зокрема для побудови та дослідження множинної лінійної регресійної моделі. При тому, слід звернути увагу студентів на два важливі аспекти:

- 1) математичних методах побудови моделі та на її застосуванні для дослідження деякого явища;
- 2) використанні статистичного пакету R, що сприяє більш ґрунтовному вивченню методів статистичного аналізу.

Подальші дослідження будуть зосереджені на методичних аспектах навчання майбутніх вчителів інформатики методів статистичного моделювання з використанням пакету R.

Список використаних джерел

1. OpenOffice.org: Теория и практика / И. Хахаев и др. М. : ALT Linux ; БИНОМ. Лаборатория знаний, 2008. 319 с.
2. Берестнева О.Г., Марухина О.В., Шевелёв Г.Е. Прикладная математическая статистика : учебное пособие. Томск : Изд-во Томского политехнического университета, 2012. 188 с.
3. Бідюк П. І., Терентьев О. М., Просьянкіна-Жарова Т. І. Прикладна статистика. Вінниця : ПП "ТД"Едельвейс і К", 2013. 304с.
4. Бредюк В.І., Джоші О. І. Економіко-математичне моделювання в середовищі табличного процесора MS Excel: Навч. посібник. Рівне: НУВГП, 2015. 240 с.
5. Василенко О. А., Сенча І.А. Математично-статистичні методи аналізу у прикладних дослідженнях : навч. посіб. Одеса : ОНАЗ ім. О. С. Попова, 2011. 166 с.
6. Глушак О.М., Семеняка С.О. Економіко-математичне моделювання – перспективний напрямок прикладної математики. Фізико-математична освіта: науковий журнал. 2017. Випуск 1(11). С.28-31.
7. Горошко Ю.В. Інформаційне моделювання у підготовці учителів математики та інформатики : монографія. Чернігів : Лозовий В. М., 2012. 367 с.
8. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. 3-е изд. М. : Диалектика, 2007. 912 с.
9. Економетрика : навчальний посібник для студентів напряму підготовки "Економічна кібернетика" всіх форм навчання / Гур'янова Л.С. та ін.. Харків : ХНЕУ ім. С. Кузнеця, 2015. 384 с.
10. Єлейко В.І., Боднар Р.Д., Демчишин М.Я. Економетричний аналіз діяльності підприємств: Навчальний посібник. Тернопіль : Навчальна книга-Богдан, 2011. 362 с.
11. Жалдак М.І., Горошко Ю.В., Вінниченко Є.Ф. Математика з комп'ютером : посіб. для вчителів. 2-ге вид. К. : НПУ ім. Драгоманова, 2009. 282 с.
12. Книгніцька Т.В. «Від практики до теорії» або як зацікавити студентів математикою. *Фізико-математична освіта : науковий журнал*. 2017. Вип. 4(14). С. 199-204.
13. Лупан І.В., Авраменко О.В. Комп'ютерні статистичні пакети : навчально-методичний посібник. Кіровоград : КОД, 2010. 218 с.
14. Майборода Р.Є., Сугакова О.В. Аналіз даних за допомогою пакета R : Навчальний посібник. 2015. 65 с.
15. Мاستицкий С.Э., Шитиков В.К. Статистический анализ и визуализация данных с помощью R: Электронная книга. URL: <http://docplayer.ru/49774338-Mastickiy-s-e-shitikov-v-k-statisticheskij-analiz-i-vizualizaciya-dannyh-s-pomoshchyu-r.html> (дата звернення 07.03.2018).
16. Наглядная статистика. Используем R! / А.Б. Шипунов и др. М. : ДМК Пресс, 2012. 298 с.
17. Радченко С.Г. Методология регрессионного анализа : монография. К. : Корнійчук, 2011. 376 с.
18. Слепко Ю.Н., Ледовская Т.В. Обработка и интерпретация результатов психологического исследования: учебное пособие. Ярославль : Изд-во ЯГПУ, 2013. 136 с.

References

1. OpenOffice.org: Theory and Practice / I. Hahaev i dr. M. : ALT Linux ; BINOM. Laboratorija znanij, 2008. 319 s. (in Russian)
2. Berestneva O.G., Maruhina O.V., Sheveljov G.E. Applied Mathematical Statistics : a tutorial. Tomsk : Izd-vo. Tomskogo politehnicheskogo universiteta, 2012. 188 s. (in Russian)
3. Bidiuk P.I., Terentiev O. M., Prosiankina-Zharova T. I. Applied Statistics. Vinnytsia : PP "TD"Edelweis i K". 304 s. (in Ukrainian)
4. Brediuk V.I., Dzhoshi O.I. Economic and Mathematical Modeling by the Environment of the MS Excel Spreadsheet : teaching aid. Rivne : NUVHP, 2015. 240 s. (in Ukrainian)
5. Vasylenko O.A., Sencha I.A. Mathematical and statistical methods of analysis in applied research. Odesa : ONAZ im. O. S. Popova, 2011. 166 s. (in Ukrainian)
6. Hlushak O.M., Semeniaka O.M. Economic and Mathematical Modeling Is A Promising Area Of Applied Mathematics. *Physical and Mathematical Education : Scientific Journal*. 2017. Issue 1(15). S.28-31. (in Ukrainian)
7. Horoshko Yu.V. Information modeling in the teacher training of mathematics and computer science: monograph. Chernihiv : Lozovyi V. M., 2012. 367 s. (in Ukrainian)
8. Draper N., Smith H. Applied Regression Analysis. 3rd ed. M.: Dialektika, 2007. 912 s. (in Russian)
9. Econometrics: a manual for students in the field of training "Economic Cybernetics" of all forms of training / Hurianova L.S., Klebanova T.S., Serhiienko O.A., Prokopovych S.V. Kharkiv. : KhNEU im. S. Kuznetsia, 2015. 384 s. (in Ukrainian)
10. Yeleiko V.I., Bodnar R.D., Demchyshyn M.Ya. Econometric Analysis of Enterprises : teaching aid. Ternopil : Navchalna knyha-Bohdan, 2011. 362 c. (in Ukrainian)
11. Zhaldak M.I., Horoshko Yu.V., Vinnychenko Ye.F. Mathematics with a Computer : Teacher's Manual. 2-nd ed. K. : NPU im. Drahomanova, 2009. 282 s. (in Ukrainian)
12. Knihnitska T.V. «From The Practice To Theory» Or How To Interest The Students By Mathematics. *Physical and Mathematical Education : Scientific Journal*. 2017. Issue 4(14). S. 199-204. (in Ukrainian)
13. Lupan I.V., Avramenko O.V. Computer Statistical Packages : teaching aid. Kirovohrad: KOD, 2010. 218 s. (in Ukrainian)
14. Maiboroda R.Ye., Suhakova O.V. Data Analysis by the R Package Tools : teaching aid. 2015. 65 s. (in Ukrainian)

15. Mastickij S.Je., Shitikov V.K. Statistical analysis and visualization of data by R tools : EBook. URL: <http://docplayer.ru/49774338-Mastickiy-s-e-shitikov-v-k-statisticheskij-analiz-i-vizualizaciya-dannyh-s-pomoshchyu-r.html> (application date 07.03.2018).
16. Visual Statistics. Use R! / A.B. Shipunov A.B. and other. M. : DMK Press, 2012. 298 s.
17. Radchenko S.G. Regression Analysis Methodology : monograph. K. : Kornijchuk, 2011. 376 s.
18. Slepko Ju.N. Ledovskaja T.V. Processing and Interpretation of the Results of Psychological Research: teaching aid. Jaroslavl' : IZD-VO JA GPU, 2013. 136 s.

METHODOLOGICAL ASPECTS OF LEARNING MULTIPLE LINEAR REGRESSION ANALYSIS WITH R PACKAGE

Taras Kobyl'nyk

Drohobych Ivan Franko State Pedagogical University, Ukraine

Abstract. *One of the important conditions of preparation of Informatics teachers is the ability and skill to use various statistical methods of analysis of experimental data. Introduction of statistical analysis methods in educational process gives the opportunity to increase the level of training of a future teacher.*

Regression analysis methods are used in various fields of science to determine the types of dependencies between the examined features. Regression analysis is the main statistical method for constructing mathematical models of objects or phenomena on the basis of experimental data. The main results currently obtained with reference to linear regression models that can serve as the initial phase of the study.

Data processing is impossible without the use of a computer with appropriate software. To the user there is a problem of choice of software for the study. It is proposed to use the statistical environment R for research and support of educational process in higher educational institutions. For regression analysis we use the freely available R package - one of the best programs for statistical analysis. Nonlinear with respect to certain transformations (change of variables or logarithms) can be reduced to linear form, that is, to approximate these dependencies by linear functions. A linear regression model is the simplest and most commonly used kind of dependence between the variables. Therefore, the study of the elements of the regression analysis, considerable attention should be paid to linear models. In the article on the specific example shows the construction and study of multiple linear regression models using the statistical environment R. the Test of conformity to the normal distribution carried out by constructing a q-q chart.

Prospects of further researches will be aimed at exploring the use of R package for statistical data analysis and methods of teaching the fundamentals of simulation modeling of students of technical specialties at pedagogical universities.

Key words: *multiple linear regression analysis, R package.*