УДК 004.94

И.Н. Сюльжин[1], С.Г. Бильчинская[2], Ю.А. Чернявский[3], Е.В. Шабинская[4]

РЕГРЕССИОННЫЙ АНАЛИЗ ПОТРЕБИТЕЛЬСКИХ РАСХОДОВ НА ПИТАНИЕ В ДОМОХОЗЯЙСТВАХ РЕГИОНА КАК ФУНКЦИИ РАСПОЛАГАЕМОГО ЛИЧНОГО ДОХОДА И ЦЕНЫ

С помощью метода множественного регрессионного анализа, широко используемого для описания и исследования некоторых процессов в экономике регионов, решается важная проблема разграничения эффектов различных независимых или частично зависимых переменных, зависящих от нескольких, в том числе коррелированных, факторов. Рассмотренная регрессионная модель включает спецификацию составляющих ее соотношений, выбор переменных, входящих в каждое соотношение, и определение соответствующих этим соотношениям математических функций. На основе официальной поквартальной статистической информации (за 2015 и 2016 гг.), представленной в интерполируемом ежемесячном варианте, сформирована регрессионная модель расходов на питание с двумя независимыми переменными как функции дохода и цены. Регрессионный анализ позволил конкретизировать зависимость расходов на питание в домохозяйствах республики от располагаемого личного дохода и относительной цены на продукты питания, что необходимо учитывать при решении задач социально-экономического развития региона.

Введение

В статье рассматривается зависимость общей величины расходов на продукты питания от располагаемого личного дохода с учетом влияния изменений относительной цены этих расходов, которую можно выразить следующим образом:

$$y = \alpha + \beta_1 x + \beta_2 p + u, \qquad (1)$$

где y – общая величина расходов на питание; x – располагаемый личный доход за вычетом личных сбережений (равен потребительским расходам); p – индекс относительной цены продуктов питания; α, β_1 и β_2 – постоянные величины (параметры уравнения); u – случайный член.

В предположении истинности модели (1), в котором допущено упрощение в том плане, что расходы на питание не влияют на доходы и цену, для наблюдаемых (статистических) данных рассмотрим уравнение регрессионной зависимости

$$\hat{y} = a + b_1 x + b_2 p, \qquad (2)$$

где величины a, b_1 и b_2 являются оценками коэффициентов α, β_1 и β_2 исходного уравнения (1) соответственно.

1. Определение и анализ параметров регрессионной модели расходов домохозяйств на питание с двумя независимыми переменными как функции дохода и цены

Для уравнения (2) вычисляются и интерпретируются коэффициенты множественной регрессии a, b_1 и b_2 с использованием статистических данных по экономике региона за 2015–2016 гг. для независимых переменных (x_i и p_i) и зависимой переменной (y_i) на том же временном интервале (номер месяца $i = \overline{1, 24}$). В табл. 1 приведены данные, сформированные на основе официальной информации [1], при этом исходная поквартальная статистическая информация (за 2015 и 2016 гг.) представлена в интерполируемом ежемесячном варианте.

Процедура преобразования для каждого квартала исходных экономических данных в ежемесячные значения заключалась в следующем. Исходные данные n-го квартала отнесены к среднему (второму) месяцу квартала. Для первого и третьего месяца каждого квартала этот результат корректировался в большую или меньшую сторону в зависимости от того, в какой мере исходные данные соседних (n–1)-го и (n+1)-го кварталов соотносятся с данными n-го квартала ($n = \overline{2,7}$).

Таблица 1

Статистические данные для экономики региона за 2015–2016 гг.

y_i	x_i	p', %	p_Σ	$p_i = p'/p_\Sigma \times 100$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$p_i - \bar{p}$	$(p_i - \bar{p})^2$
349,98	827,40	93,55	91,13	102,66	–111,79	12497,00	1,53	2,34
352,80	834,00	94,30	91,85	102,66	–105,19	11064,94	1,54	2,37
355,59	840,60	95,04	92,58	102,66	–98,59	9719,99	1,54	2,36
393,99	893,04	105,31	98,35	107,07	–46,15	2129,82	5,94	35,34
389,70	900,00	104,16	99,12	105,09	–39,19	1535,86	3,96	15,68
385,50	909,90	103,04	100,21	102,82	–29,29	857,90	1,70	2,88
383,28	937,56	102,45	103,26	99,21	–1,63	2,66	–1,91	3,66
381,00	931,98	101,84	102,64	99,21	–7,21	51,98	–1,91	3,65
379,20	927,57	101,36	102,16	99,21	–11,62	135,02	–1,91	3,65
371,61	914,01	99,33	100,66	98,67	–25,18	634,03	–2,45	6,02
374,13	907,98	100,00	100,00	100,00	–31,21	974,06	–1,13	1,27
376,59	905,58	100,66	99,74	100,92	–33,61	1129,63	–0,20	0,04
378,21	891,66	101,09	98,20	102,94	–47,53	2259,10	1,82	3,30
382,41	903,99	102,21	99,56	102,66	–35,20	1239,04	1,54	2,37
386,61	916,26	103,34	100,91	102,40	–22,93	525,78	1,28	1,63
403,29	950,76	107,79	104,71	102,94	11,57	133,86	1,82	3,31
416,19	990,99	111,24	109,14	101,92	51,80	2683,24	0,80	0,64
429,09	1031,28	114,69	113,58	100,98	92,09	8480,57	–0,15	0,02
393,00	1019,46	105,04	112,28	93,56	80,27	6443,27	–7,57	57,28
411,69	1014,00	110,04	111,68	98,53	74,81	5596,54	–2,59	6,72
415,41	1008,60	111,03	111,08	99,96	69,41	4817,75	–1,17	1,37
415,98	1012,02	111,19	111,46	99,76	72,83	5304,21	–1,37	1,88
424,89	1027,98	113,57	113,22	100,31	88,79	7883,66	–0,82	0,66
433,80	1043,94	115,95	114,97	100,85	104,75	10972,56	–0,28	0,08
391,00	**939,19**	**104,51**	**103,44**	**101,13**	**0,00**	**4044,69**	**0,00**	**6,60**

Примечание: полужирным шрифтом выделены средние значения данных.

Предположим, что наблюдаемая величина (среднемесячная) в n-м квартале $\bar{z}_n$ (z_i – это x_i или y_i) меньше по сравнению со значением $\bar{z}_{n+1}$ этой величины в (n+1)-м квартале: $\bar{z}_n = \bar{z}_{n+1} - \Delta z$. Тогда наблюдения третьего месяца n-го квартала корректируются так: $z_n^3 = (\bar{z}_n + \Delta z')$, случайная величина $\Delta z' < \Delta z$. Скорректированное наблюдение первого месяца (n+1)-го квартала $z_{n+1}^1 = \bar{z}_n - \Delta z'$. Аналогичным образом осуществляются процедуры коррекции смежных пар наблюдений z_{n-1}^3 и z_n^1.

При такой коррекции необходимо строго соблюдать исходные значения наблюдений $\bar{z}_n$ и $\bar{z}_{n+1}$. В приведенном примере увеличение n-го квартального значения наблюдения на $\Delta z'$ должно быть скомпенсировано уменьшением на соответствующую величину значений z_n^1 либо z_n^2.

Данная коррекция, не нарушая исходные средние значения для каждого квартала, во-первых, вносит определенную случайную (как будет установлено далее, несущественную) составляющую в результаты наблюдений. Во-вторых, в определенной мере она сглаживает существенно различающиеся межквартальные экономические данные, которые реально в месяцах, расположенных на стыке кварталов, не могут значительно отличаться друг от друга.

В табл. 1 приведены также вычисленные выборочные месячные значения $\overline{x}, \overline{p}$ и $\overline{y}$ и выборочные дисперсии $\mathrm{Var}(x)$, $\mathrm{Var}(p)$ и $\mathrm{Var}(y)$ по формулам $\overline{z}=\frac{1}{24}\sum_{i=1}^{24} z_i$ и $\mathrm{Var}(z)=\frac{1}{24}\sum_{i=1}^{24}(z_i-\overline{z})^2$, где $z\in\{x,\ p,\ y\}$.

Диаграмма потребительских расходов домохозяйств в регионе за период 2015–2016 гг. изображена на рис. 1.

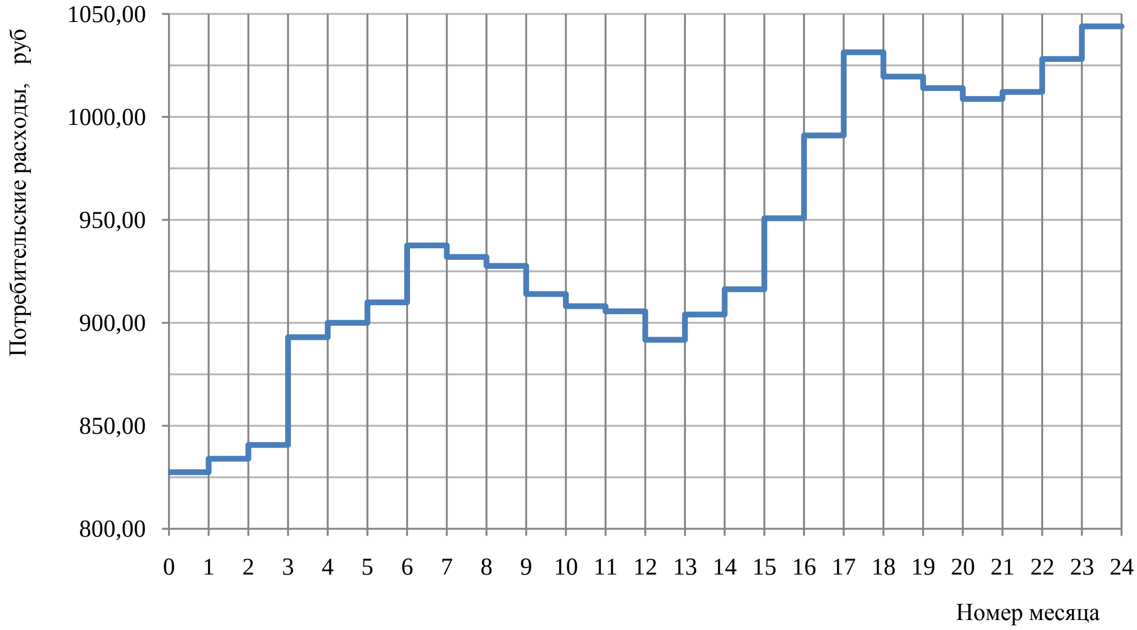

Рис. 1. Потребительские расходы домохозяйств в регионе за 2015–2016 гг.

Индекс относительной цены p_i вычисляется путем деления неявного дефлятора цен продуктов питания $p_i'=\frac{y_i}{374{,}13}\times 100$ на неявный дефлятор общих расходов $p_\Sigma^i=\frac{x_i}{907{,}98}\times 100$ и умножения полученного результата на 100: $p=\frac{p_i'}{p_\Sigma^i}\times 100$. Поскольку в ноябре 2015 г. переменные исходного уравнения $y_{i=11}=374{,}13$ и $x_{i=11}=907{,}98$, для ноября 2015 г. дефлятор цен на продукты питания $p_{i=11}'=100$ и дефлятор общих расходов $p_\Sigma^i=100$ [2].

В табл. 2 приведены данные, необходимые для определения автокорреляционных зависимостей между всеми переменными в уравнении (2):

$$\mathrm{Cov}(x,y)=\overline{x_i\,y_i}-\overline{x}\,\overline{y}=1357{,}08;\ \mathrm{Cov}(p,y)=\overline{p_i\,y_i}-\overline{p}\,\overline{y}=-9{,}29;\ \mathrm{Cov}(x,p)=\overline{x_i\,p_i}-\overline{x}\,\overline{p}=-84{,}394.$$

Таблица 2

Месячные данные для определения автокорреляционных зависимостей
между всеми переменными в уравнении (2)

$x_i\, p_i$	$x_i\, y_i$	$p_i\, y_i$	y_i	$e_i = y_i - y_i$	$(e_i - \overline{e})^2$	$(y_i - \overline{y})^2$
84 937,01	289 573,45	35 927,31	350,34	0,36	0,13	1682,44
85 621,40	294 235,20	36 219,70	353,12	0,32	0,11	1459,05
86 298,51	298 908,95	36 505,93	355,88	0,29	0,08	1253,69
95 617,84	351 848,83	42 184,53	395,08	1,09	1,20	8,96
94 576,70	350 730,00	40 951,71	390,19	0,49	0,24	1,68
93 557,40	350 766,45	39 637,74	385,43	–0,07	0,00	30,22
93 018,62	359 348,00	38 026,56	382,81	–0,47	0,22	59,56
92 465,29	355 084,38	37 800,46	380,48	–0,52	0,27	99,95
92 028,44	351 734,54	37 622,16	378,64	–0,56	0,31	139,18
90 186,42	339 655,26	36 667,19	370,84	–0,77	0,59	375,88
90 798,00	339 702,56	37 413,00	373,55	–0,58	0,34	284,51
91 395,02	341 032,37	38 007,08	376,18	–0,41	0,17	207,58
91 788,18	337 234,73	38 933,23	378,29	0,08	0,01	163,52
92 807,48	345 694,82	39 259,85	382,35	–0,06	0,00	73,75
93 826,78	354 235,28	39 589,61	386,44	–0,17	0,03	19,25
97 874,87	383 432,00	41 516,22	402,97	–0,32	0,10	151,11
101 005,59	412 440,13	42 419,72	415,76	–0,43	0,18	634,66
104 136,30	442 511,94	43 328,53	428,86	–0,23	0,05	1451,04
95 377,58	400 647,78	36 767,89	394,78	1,78	3,16	4,01
99 913,48	417 453,66	40 565,46	412,05	0,36	0,13	428,18
100 816,29	418 982,53	41 523,00	415,38	–0,03	0,00	595,97
100 954,62	420 980,08	41 496,32	416,02	0,04	0,00	624,13
103 117,00	436 778,42	42 620,85	424,86	–0,03	0,00	1148,70
105 279,37	452 861,17	43 747,91	433,64	–0,16	0,03	1832,05
94 891,59	**368 578,02**	**39 530,50**	**391,00**	**0,00**	**0,31**	**530,38**

Примечание: полужирным шрифтом выделены средние значения данных.

С использованием данных табл. 1 и 2 получено следующее уравнение регрессии:

$$\hat{y} = -398{,}365 + 0{,}4175\, x + 3{,}9284\, p, \qquad (3)$$

$$(\text{с.о.}) \quad (7{,}64) \;\; (0{,}0022) \;\; (0{,}0549)$$

где с.о. – стандартные ошибки [3].

Величины a, b_1 и b_2 уравнения вычислялись по формулам [4]

$$a = \overline{y} - b_1\, \overline{x} - b_2\, \overline{p} = -398{,}365; \qquad (4)$$

$$b_1 = \frac{\mathrm{Cov}(x,y)\,\mathrm{Var}(p) - \mathrm{Cov}(p,y)\;\mathrm{Cov}(x,p)}{\mathrm{Var}(x)\,\mathrm{Var}(p) - (\mathrm{Cov}(x,p))^2} = 0{,}4175; \qquad (5)$$

$$b_2 = \frac{\mathrm{Cov}(p,y)\,\mathrm{Var}(x) - \mathrm{Cov}(x,y)\;\mathrm{Cov}(x,p)}{\mathrm{Var}(p)\,\mathrm{Var}(x) - (\mathrm{Cov}(x,p))^2} = 3{,}9284. \qquad (6)$$

На основании установленной регрессионной зависимости (3) можно прогнозировать специфику изменений затрат на питание следующим образом. В связи с тем что оба коэффициента b_1 и b_2 положительные, общая величина расходов на питание y возрастает при увеличении как располагаемого личного дохода x, так и индекса относительной цены на продукты питания p. Согласно рассматриваемой модели за два года (2015–2016 гг.) увеличение располагаемого личного дохода составило 216,54 (1043,94–827,40), а индекс цен p упал с 102,66 до 100,85, т. е. изменился на 1,8 пункта. Таким образом, совместное воздействие отмеченных факторов привело к увеличению затрат на питание за указанный период в размере $b_1 \times 216{,}54 - b_2 \times 1{,}8 = 90{,}4 - 7{,}07 = 83{,}33$. Полученный аналитически результат достаточно хорошо соответствует фактической разности расходов на питание y в конце рассматриваемого периода (декабрь 2016 г.) относительно начала (январь 2015 г.), т. е. 433,80–349,98 = 83,82.

Предполагалось, что расходы на питание не влияют на доход и цену, что является определенным упрощением действительности. В реальности же чистый эффект воздействия на величину y будет зависеть не только от коэффициентов β_1 и β_2, но также от размеров изменений x и p [5].

Диаграмма расходов домохозяйств на продукты питания в регионе за 2015–2016 гг. показана на рис. 2.

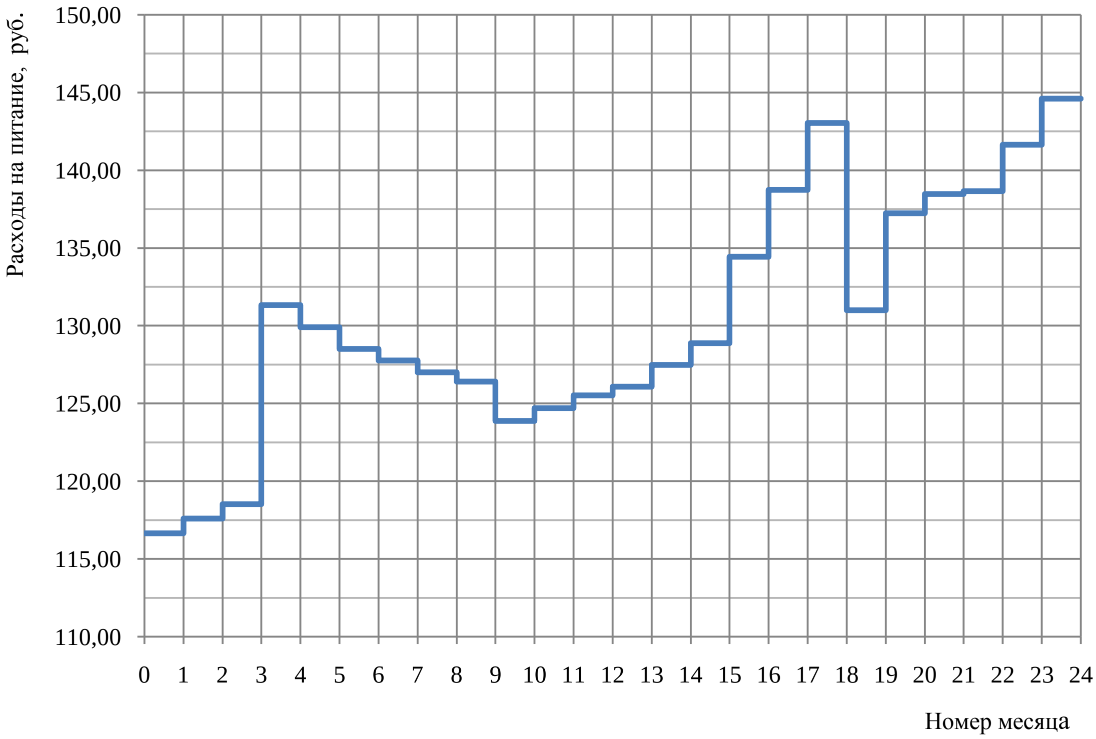

Рис. 2. Расходы домохозяйств на питание в регионе за 2015–2016 гг.

2. Оценка стандартных ошибок коэффициентов регрессии. Коэффициент детерминации R^2, t-тесты и доверительные интервалы

Стандартные ошибки коэффициентов регрессии b_1 и b_2 оценивались по формулам [4]

$$\text{с.о.}(b_1) = \sqrt{\frac{\operatorname{Var}(e)}{(n-3)\operatorname{Var}(x)} \times \frac{1}{1 - r_{xp}^2}} = 0{,}0022; \tag{7}$$

$$\text{c.o.}\,(b_2) = \sqrt{\frac{\text{Var}(e)}{(n-3)\text{Var}(p)} \times \frac{1}{1-r_{xp}^2}} = 0{,}0549; \tag{8}$$

$$\text{c.o.}\,(a) = \sqrt{\frac{\text{Var}(e)}{(n-3)} \times \left(1 + \frac{(\bar{x})^2}{\text{Var}(x)}\right) \times \left(\frac{1}{1-r_{xp}^2}\right)} + \sqrt{\frac{\text{Var}(e)}{(n-3)} \times \left(1 + \frac{(\bar{p})^2}{\text{Var}\,p}\right) \times \left(\frac{1}{1-r_{xp}^2}\right)} = 7{,}64, \tag{9}$$

где разница между вычисленным регрессионным и наблюдаемым значениями зависимой переменной $e_i = \hat{y}_i - y_i$; выборочная дисперсия ошибки

$$\text{Var}(e) = \frac{1}{24}\sum_{i=1}^{24}(e_i - \bar{e})^2 = 0{,}3067; \tag{10}$$

коэффициент корреляции

$$r_{x,p} = \frac{\text{Cov}(x,p)}{\sqrt{\text{Var}(x)\,\text{Var}(p)}} = 0{,}5164. \tag{11}$$

Коэффициент детерминации R^2, который оценивает долю дисперсии зависимой переменной y, обусловленную регрессией, определяется следующим образом:

$$R^2 = 1 - \frac{\text{Var}(e)}{\text{Var}(y)} = 0{,}9994. \tag{12}$$

В связи с тем что коэффициент R^2 близок к единице, можно сделать заключение о достаточно точном соответствии линии регрессии статистическим данным. С учетом этого, а также не слишком тесной корреляции между переменными x и p ($r_{x,p}$ = 0,5164) результаты оценивания регрессии можно считать надежными.

Критический уровень t-тестов для коэффициентов множественной регрессии при любом уровне значимости зависит от числа степеней свободы, которое равно (n–k–1): число наблюдений n=25 минус число оцениваемых параметров k = 2 и минус постоянный член 1. В рассматриваемом случае при числе степеней свободы 22 и уровне значимости 5 % критическое значение $t_{\text{крит}}$=2,074 [6]. Таким образом, гипотетические значения β_1 и β_2 будут совместимыми с результатом оценивания регрессии, когда удовлетворяются двойные неравенства

$$\begin{gathered} b_1 - \text{c.o.}\,(b_1) \times t_{\text{крит}} < \beta_1 < b_1 + \text{c.o.}\,(b_1)\,) \times t_{\text{крит}}, \\ b_2 - \text{c.o.}\,(b_2) \times t_{\text{крит}} < \beta_2 < b_2 + \text{c.o.}\,(b_2)\,) \times t_{\text{крит}}. \end{gathered} \tag{13}$$

Подставляя полученные в результате регрессионного анализа значения в неравенства (13), получаем

$$\begin{gathered} 0{,}413 < \beta_1 < 0{,}422, \\ 3{,}815 < \beta_2 < 4{,}042. \end{gathered} \tag{14}$$

Любые гипотетические значения β_1 и β_2, которые удовлетворяют соотношениям (14), будут автоматически совместимы с оценками b_1 и b_2; иными словами, не будут ими опровергаться [6]. При условии выполнения четвертого условия Гаусса – Маркова, т. е. когда объясняющие (независимые) переменные не являются стохастическими (не имеют случайной составляющей), математическое ожидание величины b_1 находится из выражения [4]

$$\bar{b} = \beta_1 + \beta_2 \frac{\text{Cov}(x, p)}{\text{Var}(x)}. \qquad (15)$$

Таким образом, оценка b_1 будет смещена относительно значения β_1 на величину $\beta_2 \frac{\text{Cov}(x, p)}{\text{Var}(x)} = 3{,}9284 \times \frac{-84{,}394}{4044{,}69} = -0{,}08$, т. е. выражение (15) будет в среднем давать несколько заниженную оценку β_1. Оценка b_1 будет несмещенной только в исключительном случае, когда $\text{Cov}(x, p) = 0$.

В неравенствах (14) обозначены границы доверительных интервалов для величин β_1 и β_2. Относительно небольшой доверительный интервал для коэффициента β_1 обусловлен значительной дисперсией потребительских расходов домохозяйств $\text{Var}(x) = 4044{,}687$ при среднем значении $\bar{x} = 939{,}19$.

Дополнительно для проверки качества оценивания регрессии, т. е. проверки того, действительно ли полученное при оценке регрессии значение коэффициента $R^2 = 0{,}9994$ отражает истинную зависимость или оно появилось случайно, определим *F*-статистику [4, 6]. С этой целью вначале разложим выборочную дисперсию $\text{Var}(y)$ зависимой переменной на «объясненную» $\text{Var}(\hat{y})$ и «необъясненную» $\text{Var}(e)$ составляющие:

$$\text{Var}(y) = \text{Var}(\hat{y}) + \text{Var}(e). \qquad (16)$$

Сравнивая значение выборочной дисперсии зависимой переменной $\text{Var}(y) = 530{,}3772$ со значениями «объясненной» $\text{Var}(\hat{y}) = 530$ и «необъясненной» $\text{Var}(e) = 0{,}3067$ составляющих ошибок, можно сделать заключение о хорошем соответствии полученных результатов уравнению (16). В связи с незначительностью «необъясненной» $\text{Var}(e) = 0{,}3067$ составляющей ошибок отсутствует необходимость анализировать их особенности.

Используя определение выборочной дисперсии и умножая обе части уравнения (16) на число *n*, получим уравнение, связывающее общую сумму квадратов отклонений *TSS* зависимой переменной *y* от ее выборочного среднего значения, «объясненную» сумму квадратов отклонений *ESS* и «необъясненную» сумму квадратов отклонений *RSS*:

$$TSS = ESS + RSS. \qquad (17)$$

Для определения качества оценивания регрессии используем *F*-статистику, которую запишем как отношение «объясненной» суммы квадратов отклонений *ESS* в расчете на одну независимую переменную к «необъясненной» (остаточной) сумме квадратов *RSS* в расчете на одну степень свободы:

$$F = \frac{ESS / k}{RSS / (n - k - 1)}. \qquad (18)$$

Учитывая равенство коэффициента $R^2 = ESS / TSS$, *F*-статистику определим следующим образом:

$$F = \frac{R^2 / k}{(1 - R^2) / (n - k - 1)} = 4997. \qquad (19)$$

Критический уровень *F* с 22 степенями свободы при уровне значимости в 5 % составляет $F_{\text{крит}} = 3{,}44$ [4]. Таким образом, соответствующая выполненному регрессионному анализу *F*-статистика, равная 4997, указывает на значимый уровень обоснования полученных результатов.

Заключение

В результате выполненного в работе регрессионного анализа зависимости затрат на продукты питания от двух факторов: располагаемого личного дохода домохозяйств региона и относительного индекса цен – определены оптимальные оценки коэффициентов регрессии, которые обеспечивают хорошее соответствие между неизвестными (на момент наблюдения) фактическими значениями затрат на продукты питания и значениями, прогнозируемыми по уравнению регрессии. Определены стандартные ошибки коэффициентов регрессии, коэффициент детерминации R^2, F-статистика, t-тесты и обозначены границы доверительных интервалов для коэффициентов множественной регрессии. Установлено пренебрежимо малое значение «необъясненной» составляющей ошибок по сравнению с «объясненной» составляющей. Математические выкладки и полученные результаты могут использоваться для оценки эффективности функционирования системы управления развитием экономики региона и его прогнозирования.

Список литературы

1. Национальный статистический комитет Республики Беларусь [Электронный ресурс]. – Режим доступа : http://www.belstat.gov.by/ofitsialnaya-statistika/solialnaya-sfera/uroven-zhizni-naseleniya/operativnaya-informatsiya_7/potrebitelskie-rashody-domashnih-hozyaistv_2/. – Дата доступа : 01.03.2017.

2. Сюльжин, И.Н. Анализ взаимосвязей между компонентами СППР / И.Н. Сюльжин // Материалы Третьей Междунар. науч.-практ. конф. «Прикладные проблемы оптики, информатики, радиофизики и физики конденсированного состояния», Минск, Беларусь, 28–29 апреля 2015 / НИУ «Институт прикладных физических проблем им. А.Н. Севченко» БГУ. − Минск, 2015. − С. 197−199.

3. Регрессионный двухкомпонентный анализ инфляционной составляющей в системных показателях экономической деятельности регионов / С.Г. Бильчинская [и др.] // Вестник КамчатГТУ. – 2016. – № 38. – С. 90–99.

4. Доугерти, К. Введение в эконометрику / К. Доугерти ; пер. с англ. – М. : ИНФРА-М, 1999. – XIV. – 402 с.

5. Макконнелл, К.Р. Экономикс: принципы, проблемы и политика / К.Р. Макконнелл, С.Л. Брю ; пер. с 13-го англ. изд. – М. : ИНФРА-М, 1999. – XXXIV. – 974 с.

6. Harvey, A. The Econometric Analysis of Time Series / A. Harvey. – Oxford : Philip. Allan, 1981. – XVII. – 410 с.

Поступила 09.06.2017

[1] *Белорусский государственный университет,*
Минск, ул. Курчатова, 5;

[2] *Академия управления при Президенте Республики Беларусь,*
Минск, ул. Московская, 13;

[3] *Институт информационных технологий Белорусского*
государственного университета информатики и радиоэлектроники,
Минск, ул. Козлова, 28;

[4] *НИУ «Институт прикладных*
физических проблем им. А.Н. Севченко» БГУ,
Минск, ул. Курчатова, 7
e-mail: ivan.syulzhin@yandex.ru,
shabinskaya@rambler.ru