

Корреляционно-регрессионный анализ объема платных услуг в Амурской области

Горяйнова В.А., студент 2 курса направления подготовки 38.03.02-«Менеджмент организации», научный руководитель – Заломская Г.А., Амурский государственный университет, г.Благовещенск, Россия

Аннотация. Актуальность изучения темы объясняется тем, что на основе результатов корреляционно-регрессионного анализа можно спрогнозировать «поведение» того или иного социально-экономического процесса, выявить наиболее значимые факторы, влияющие на него и принять продуманное, эффективное решение. Целью корреляционно-регрессионного анализа является нахождение зависимости между двумя параметрами и ее степени с последующим выводением уравнения. В данной статье будет выявляться зависимость между среднемесячной заработной платой одного работника за 2007-2017 гг. (факторный признак) и объемом платных услуг в Амурской области за те же годы (результативный признак).

Ключевые слова: корреляционно-регрессионный анализ, среднемесячная заработная плата, объема платных услуг, статистика Амурской области, факторный и результативный признаки, адекватность модели, направление и теснота связи

Correlation and regression analysis of the volume of paid services in the Amur region

Goryaynova V.A., 2nd year student of the direction of preparation 38.03.02 "Management of the organization», supervisor – Zalomskaya G.A., Amur state University, Blagoveshchensk, Russia

Annotation. The relevance of the study is explained by the fact that on the basis of the results of correlation and regression analysis it is possible to predict the

"behavior" of a socio-economic process, to identify the most significant factors affecting it and make a thoughtful, effective decision. The purpose of correlation and regression analysis is to find the relationship between the two parameters and its degree, followed by the derivation of the equation. This article will identify the relationship between the average monthly wage of one employee for 2007-2017. (factorial sign) and the volume of paid services in the Amur region for the same years (effective sign).

Keywords: correlation and regression analysis, average monthly salary, volume of paid services, statistics of the Amur region, factor and effective features, the adequacy of the model, direction and closeness of communication

Объект исследования - статистические данные по размеру среднемесячной заработной платы одного работника, данные по объему платных услуг в области.

Предмет исследования - определение характера и силы зависимости между исследуемыми признаками.

С помощью корреляционно-регрессионного анализа мы установим зависимость между факторным и результативным признаками, установим направление и тесноту связи, а также составим уравнение регрессии.

Для нахождения параметров уравнения $\bar{y}_x = a_0 + a_1 * x$ произведем вычисления вспомогательных величин, которые запишем в таблице 1.

Год	x	y	y ²	x ²	xy	y _x
1	2	3	4	5	6	7
2007	13534,4	14333,1	205437755,6	183179983,4	193989908,6	13362,8
2008	16665,0	17783	316235089,0	277722225,0	296353695,0	17432,6
2009	19019,0	19900,7	396037860,5	361722361,0	378491413,3	20492,8
2010	21207,5	24749,6	612542700,2	449758056,3	524877142,0	23337,9
2011	24202,1	29038,6	843240290,0	585741644,4	702795101,1	27230,8
2012	26789,0	30812,5	949410156,3	717650521,0	825436062,5	30593,8
2013	30541,7	36852,1	1358077274,4	932795438,9	1125525782,6	35472,3
2014	32396,5	38510,7	1483074014,5	1049533212,3	1247611892,6	37883,6
2015	32901,7	41106,6	1689752563,6	1082521862,9	1352477021,2	38540,3
2016	33836,8	42752,4	1827767705,8	1144929034,2	1446604408,3	39755,9
2017	37367,7	44051,5	1940534652,3	1396345003,3	1646103236,6	44346,1
Итого	288461,4	339890,8	11622110061,9	8181899342,6	9740265663,7	328448,9

Средние значения	26223,8	30899,2	-	-	-	-
------------------	---------	---------	---	---	---	---

Таблица 1. К расчету параметров и оценке линейной однофакторной регрессионной модели

По формуле и данным таблицы 1 рассчитаем параметр a_1 :

$$a_1 = \frac{n \sum xy - \sum y * \sum x}{n * \sum x^2 - \sum x * \sum x} ;$$

$$a_1 = \frac{11 * 9740265663,7 - 288461,4 * 339890,8}{11 * 8181899342,6 - 288461,4 * 288461,4} = 1,3.$$

По формуле и данным таблицы 1 рассчитаем параметр a_0 :

$$a_0 = \frac{\sum y * \sum x^2 - \sum yx * \sum x}{n * \sum x^2 - \sum x * \sum x} ,$$

$$a_0 = \frac{339890,8 * 8181899342,6 - 9740265663,7 * 288461,4}{11 * 8181899342,6 - 288461,4 * 288461,4} = -4231,9.$$

Уравнение регрессии принимает вид:

$$y_x = -4231,9 + 1,3 * x.$$

При увеличении среднемесячной заработной платы одного работника на 1 рубль – объем платных услуг Амурской области в среднем увеличивался на 1,3 млн рублей.

Параметр $a_0 = -4231,9$ показывает усредненное влияние на результативный признак неучтенных факторов.

Подставив данные параметров в уравнение регрессии, получим теоретическое значение y_x за 2007 год:

$$y_x = -4231,9 + 1,3 * 13534,4 = 13362,8.$$

Аналогично рассчитаем теоретические значения y_x остальных лет (колонка 7 таблица 1).

Для измерения тесноты связи предварительно рассчитаем дисперсии признаков, а также среднее квадратическое отклонение соответственно:

$$\sigma_x^2 = \frac{617355771}{11} = 56123251,9;$$

$$\sigma_y^2 = \frac{1119768614,2}{11} = 101797146,7;$$

$$\sigma_x = \sqrt{56123251,9} = 7491,5;$$

$$\sigma_y = \sqrt{101797146,7} = 10089,5.$$

Рассчитаем линейный коэффициент корреляции по формуле :

$$r = 1,3 * \frac{7491,5}{10089,5} = 0,97.$$

По шкале Чеддока связь между факторным и результативным признаком является весьма высокой и прямой.

Для расчета теоретического корреляционного отношения необходимо предварительно вычислить дисперсии по формулам:

$$\sigma_{y_x}^2 = \frac{\sum(y_x - \bar{y})^2}{n}, \quad (33)$$

где $\sigma_{y_x}^2$ – факторная дисперсия.

$$\sigma_{\text{ост}}^2 = \frac{\sum(y - y_x)^2}{n}, \quad (34)$$

Где $\sigma_{\text{ост}}^2$ – остаточная дисперсия.

Для факторной дисперсии предварительно рассчитаем вспомогательные величины и запишем их в таблицу 2:

у	у _х	у _х - $\bar{y}$	(у _х - $\bar{y}$) ²	у - у _х	(у - у _х) ²	у - $\bar{y}$	(у - $\bar{y}$) ²
14333,1	13362,8	-970,3	941443,3	970,3	941443,3	-16566,1	274434464,4
17783	17432,6	-350,4	122780,2	350,4	122780,2	-13116,2	172033748,5
19900,7	20492,8	592,1	350582,4	-592,1	350582,4	-10998,5	120966202,4
24749,6	23337,9	-1411,8	1993038,1	1411,8	1993038,1	-6149,6	37817132,9
29038,6	27230,8	-1807,8	3268032,4	1807,8	3268032,4	-1860,6	3461697,0
30812,5	30593,8	-218,7	47829,7	218,7	47829,7	-86,7	7510,6
36852,1	35472,3	-1379,8	1903820,4	1379,8	1903820,4	5952,9	35437451,3
38510,7	37883,6	-627,1	393317,1	627,1	393317,1	7611,5	57935485,8
41106,6	38540,3	-2566,3	6585844,4	2566,3	6585844,4	10207,4	104191757,1
42752,4	39755,9	-2996,5	8978772,5	2996,5	8978772,5	11853,2	140499212,3
44051,5	44346,1	294,6	86795,1	-294,6	86795,1	13152,3	172983951,8
Итого	328448,9	-	24672255,5	-	24672255,5	-	1119768614,2

Таблица 2. К расчету факторной и остаточной дисперсий

$$\sigma_{y_x}^2 = 101797146,7 - 223841,4 = 99554214,4;$$

$$\sigma_{\text{ост}}^2 = \frac{24672255,5}{11} = 2242932,3.$$

Исходя из полученных результатов, общая дисперсия показывает, что вариация результативного признака под влиянием всех факторов, вызывающих

эту вариацию составляет 289,07, факторная дисперсия показывает, что вариация численности занятого населения под влиянием признака-фактора составляет 144.

Зная все необходимые данные, рассчитаем теоретическое корреляционное отношение по формуле :

$$\eta = \sqrt{\frac{\sum(y_x - \bar{y})^2}{\sum(y - \bar{y})^2}};$$

$$\eta = \sqrt{\frac{99554214,4}{101797146,7}} = 0,97.$$

Теоретическое корреляционное отношение близится к 1, значит, говорит о тесной связи признаков.

Адекватность регрессионной модели оценим с помощью критерия Фишера:

$$F_3 = \frac{99554214,4}{2242932,3} * \frac{11-2}{2-1} = 399,5.$$

Табличное значение F-критерия при доверительной вероятности 0,95, т.е. 1-0,05 при $v_1 = m - 1 = 2 - 1 = 1$; $v_2 = n - m = 11 - 2 = 9$ составляет 5,12.

Поскольку $F_{\text{расч}} > F_{\text{табл}}$, то уравнение регрессии следует признать адекватным.

Для вычислим средней ошибки аппроксимации предварительно рассчитаем $|y - y_x|$, $\frac{|y - y_x|}{y}$, $\sum \frac{|y - y_x|}{y}$ и представим в таблице 3.

y	y _x	y - y _x	$\frac{ y - y_x }{y}$
14333,1	13362,8	970,3	0,07
17783	17432,6	350,4	0,02
19900,7	20492,8	592,1	0,03
24749,6	23337,9	1411,8	0,06
29038,6	27230,8	1807,8	0,06
30812,5	30593,8	218,7	0,01
36852,1	35472,3	1379,8	0,04
38510,7	37883,6	627,1	0,02
41106,6	38540,3	2566,3	0,06
42752,4	39755,9	2996,5	0,07
44051,5	44346,1	294,6	0,01
Итого	328448,9	-	0,44

Таблица 3. Расчеты для вычисления средней ошибки аппроксимации

$$\bar{\varepsilon} = \frac{1}{11} * 0,44 * 100 = 4\%.$$

Ошибка аппроксимации показывает, что подбор фактора (среднемесячная заработная плата одного работника за 2007-2017 гг.) произведен верно, а расчеты проведены точно, так как ошибка аппроксимации не превышает 12-15 %.

Таким образом, все показатели тесноты корреляционной связи показывают весьма высокую прямую связь между среднемесячной заработной платой одного работника и объемом платных услуг Амурской области, и действительно, с увеличением наших доходов - объем наших потребностей растет. В связи с тем, что линейный коэффициент корреляции и теоретическое корреляционное отношение равны, то можно заключить, что гипотеза о линейной форме связи подтверждена.

Список используемых источников:

1. Федеральная служба государственной статистики [электронный ресурс]
// Режим доступа: URL: www.gks.ru
2. Амурский статистический ежегодник: сборник/ Амуроблкомстат: Б., 2018
3. Елисеева, И. И. Статистика: [углубленный курс]: учебник для бакалавров / И. И. Елисеева и др.]. – Москва: Юрайт: ИД Юрайт, 2016
4. Айзинова, И. М. Сфера платных услуг населению в контексте социально-экономического развития // Проблемы прогнозирования. – 2010 - №1