

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ГУМАНИТАРНО-ПЕДАГОГИЧЕСКАЯ АКАДЕМИЯ (ФИЛИАЛ)
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «КРЫМСКИЙ ФЕДЕРАЛЬНЫЙ
УНИВЕРСИТЕТ ИМЕНИ В.И. ВЕРНАДСКОГО» В г. ЯЛТЕ**

УТВЕРЖДАЮ

**Заместитель директора Гуманитарно-
педагогической академии (филиал)
ФГАОУ ВО «КФУ имени В.И. Вернадского»
в г. Ялте**

Н. В. Горбунова

«____» 2019 г.

**Учебно-методические рекомендации по статистике для студентов
экономических специальностей**

Статистические показатели: корреляция и регрессия

Направление подготовки: 38.03.01 «Экономика»

Профиль подготовки: «Финансы и

Ялта – 2019 г.

Содержание

Цель	3
Студент должен уметь.....	3
Студент должен знать.....	3
Информационный материал.....	4
Последовательность расчета коэффициента корреляции.....	5
Расчет средних и среднего произведения.....	6
Расчет ковариации.....	7
Расчет дисперсии.....	8
Расчет сигмы (среднеквадратическое отклонение)	9
Расчет коэффициента корреляции.....	10
Расчет эластичности.....	11
Расчет критериев k и b	12
Расчет теоретического значения результативного признака Y_{meop}	13
Расчет отклонений.....	14
Расчет $Ei-Ei-1$	16
Расчет остаточной дисперсии.....	17
Расчет внутренней дисперсии.....	18
Расчет показателя Дарбина-Уотсона.....	18
График отклонений.....	19
Пример расчета индекса сезонности с использованием метода регрессии и корреляции	20
Задачи на самостоятельное выполнение.....	23

Статистика – это прикладная наука, базирующаяся на математической статистике. Статистика развивает методы применительно к описанию и исследованию явлений и процессов в реальной жизни. Она имеет отношение к сфере социально-экономических процессов и к исследованиям в области естественных наук.

Рассматриваются основные приёмы и методы обработки и анализа статистических данных. Излагаются методы обобщения статистических данных. Приводятся способы выявления взаимосвязи между показателями.

Цель: ознакомление студентов с методами расчета корреляции, регрессионного уравнения, дисперсии и других статистических показателей с помощью электронной таблицы *Exsel*. Благодаря изучению общей теории статистики студенты экономических специальностей начинают овладевать научными методами познания социально-экономических явлений и процессов с количественной стороны. В курсе статистики излагается методология получения и обработки статистической информации, контроля ее достоверности. Курс предполагает получение студентами необходимых знаний и умений в области анализа статистической информации, которые будут необходимы в дальнейшей их исследовательской деятельности. В связи, с чем необходимо подробно изучить такие статистические показатели как: средняя арифметическая, ковариация, дисперсия, сигма, корреляция, эластичность в уравнении регрессии, критерии уравнения регрессии k и b , уравнение регрессии, отклонения, остаточная внутренняя дисперсия, критерий Дарбина-Уотсона и остаточная сигма.

Студент должен уметь:

- выявлять основную закономерность изучаемого признака путем вычисления средних величин;
- рассчитывать средние величины, используя функций Excel;
- обосновывать методику применения коэффициентов и критериев корреляции и регрессии;
- давать характеристику разнообразиям дисперсии и сигмы;

- делать выводы о типичности обобщающей характеристики признака в изучаемой совокупности.

Студент должен знать:

- основные понятия темы (дисперсия, сигма, корреляция, эластичность уравнения регрессии, регрессия, отклонение, остатки);

- методику расчета средних величин и критериев регрессионного уравнения, дисперсий и сигм, корреляции и эластичности;

- как с помощью корреляционного анализа определять характер и тесноту связи между случайными величинами.

Информационный материал

Один из наиболее общих законов объективного мира – закон всеобщей связи и зависимости между явлениями. Естественно, что, исследуя явления в самых различных областях, статистика неизбежно сталкивается с зависимостями как между количественными, так и между качественными показателями, признаками. Ее задача – обнаружить (выявить) такие зависимости и дать им количественную характеристику.

Существует 2 вида связи между отдельными признаками: функциональная и стохастическая (статистическая), частным случаем которой является корреляционная.

Корреляционная связь — связь, проявляющаяся не в каждом отдельном случае, а в массе случаев в средних величинах в форме тенденции.

Статистическое исследование корреляционной связи ставит своей конечной целью получение модели зависимости для ее практического использования.

Изучение корреляционных связей сводится в основном к решению следующих задач:

1) выявление наличия (отсутствия) корреляционной связи между изучаемыми признаками;

2) измерение тесноты связи между двумя (и более) признаками с помощью специальных коэффициентов (эта часть исследования называется корреляционным анализом);

3) определение уравнения регрессии – математической модели, в которой среднее значение результативного признака y рассматривается как функция одной или нескольких переменных – факторных признаков (эта часть исследования именуется регрессионным анализом).

Оснащение занятия мультимедийным проектором и наглядным материалом в виде мультимедийных презентаций.

Последовательность расчета коэффициента корреляции. Рассмотрим расчет коэффициента корреляции и уравнения регрессии на примере нашей задачи поэтапно в такой последовательности.

Существует предположение, что изменение данных величин X и Y зависимо. Проверим это с помощью коэффициента корреляции:

$$Kop. = \frac{[\sum(X_i - \bar{X})(Y_i - \bar{Y})]}{\delta X * \delta Y}$$

Где X_i - значения, принимаемые переменной X ,

Y_i - значения, принимаемые переменной Y ,

\bar{X} - средняя по X ,

\bar{Y} - средняя по Y .

Вначале произведем ввод данных X и Y . Всего 10 значений .

	A1	B1	C1	D1	E1	F1	G1	H1	I1	J1	K1	L1
1	Xi	Ui	XY	X^2	Y^2							
2	38	42	1596	1444	1764							
3	40	45	1800	1600	2025							
4	42	49	2058	1764	2401							
5	45	56	2520	2025	3136							
6	49	54	2646	2401	2916							
7	56	50	2800	3136	2500							
8	54	48	2592	2916	2304							
9	50	45	2250	2500	2025							
10	48	44	2112	2304	1936							
11	45	41	1845	2025	1681							
12												

Рис. 1 Ввод данных

1. Расчет средних и среднего произведения.

Для расчета \bar{X} (средних X и Y) и \bar{Y} будем использовать среднюю арифметическую простую:

$$\bar{X} = \frac{\sum X}{n}$$

Где, $\sum X$ - сумма вариантов признака;

n – число единиц, обладающих данным признаком.

В строке A14 числитель формулы (467). В строке A17 указано количество этих признаков (10). В строке A19 представлен расчет средней арифметической простой:

$$\bar{X} = \text{СРЗНАЧ}(A14/A17) = 46,7$$

Тоже делаем в строке B19 для $\bar{Y} = \text{СРЗНАЧ}(B14/B17) = 47,4$

	A	B	C	D	E	F	G	H	I	J	K	L	M
7	56	50	2800	3136	2500								
8	54	48	2592	2916	2304								
9	50	45	2250	2500	2025								
10	48	44	2112	2304	1936								
11	45	41	1845	2025	1681								
12													
13													
14	467	474	22219	22115	22688								
15													
16	количество												
17	10	10	10	10	10								
18	X ср	У ср											
19	46,7	47,4											
20													
21													

Рис. 2 Расчет средних значений X и Y

Для расчета среднего произведения \bar{XY} используем формулу:

$$\bar{XY} = \frac{\sum XY}{n}$$

Где $\sum XY$ - сумма произведений вариантов признака;

n - число единиц, обладающих данным признаком.

Столбец С представляет собой значения произведения X и Y, например, C2 = A2*B2 и C11 = A11*B11, а строка C14 представляет собой сумму этих всех

произведений [=СУММ(С2:С11)=22219]. В строке С19 отражен расчет среднего произведения \bar{XY} [=СУММ(С14/С17)=2221,9].

The screenshot shows a Microsoft Excel spreadsheet titled "Книга1 - Microsoft Excel (Свойства активации продукта)". The data is organized into several sections:

- Data Section:** Rows 8 through 14 contain numerical values in columns A through E. Row 14 is a summary row with red text.
- Statistical Calculations:** Rows 16 through 19 show statistical measures:
 - Row 16: "количество"
 - Row 17: Values 10, 10, 10, 10, 10
 - Row 18: Labels "X cp", "Уср", "ХУ усредн", "Х^2 cp", "У^2 cp".
 - Row 19: Numerical results 46,7, 47,4, 2221,9, 2211,5, 2268,8.

The formula bar at the top shows the formula =СУММ(С14:С17) in cell C19. The ribbon tabs include Файл, Главная, Вставка, Разметка страницы, Формулы, Данные, Рецензирование, and Вид. The status bar at the bottom right shows "RU" and "166%".

Рис. 3 Расчет среднего произведения X и Y

2. Расчет коэффициента ковариации (то есть числителя формулы).

Простое определение ковариации - это мера взаимодействия двух случайных переменных. Найдем произведение средних \bar{X} и \bar{Y} . Оно представлено в строке A21. [A21=СУММ(A19*B19)=2213,58].

Рис. 4 Расчет произведения средних \bar{X} и \bar{Y}

Затем найдем ковариацию, вычитая из строки С19 строку А21 В строке А23 получим значение ковариации. [A23=СУММ(С19-А21)=8,32].

	A	B	C	D	E	F	G	H	I	J	K
15											
16	количество										
17	10	10	10	10	10						
18	Xср	Уср	ХУ усредн	Х^2ср	У^2ср						
19	46,7	47,4	2221,9	2211,5	2268,8						
20	X*Yср		Xср в кв	Уср в кв							
21	2213,58			2180,89	2246,76						
22	ковар										
23	8,32										
24											
25											
26											
27											

Рис. 5 Расчет произведения показателя ковариации

3. Расчет дисперсии

Дисперсия представляет собой средний квадрат отклонений индивидуальных значений признака от их средней величины.

Найдем дисперсию X (сокращенный способ):

$$D = \bar{X}^2 - (\bar{X})^2 = \text{СУММ}(D19-D21) = 30,61$$

Где \bar{X}^2 - средний квадрат;

$(\bar{X})^2$ - квадрат среднего.

Для расчета среднего квадрата будем использовать среднюю арифметическую простую, но с одной лишь разницей – возведения всех значений X в квадрат и потом уже их суммирования. Для расчета числителя формулы будем все X возводить в квадрат и сделаем их сумму. В колонке D2 представлен расчет признаков X возвещенных в квадрат [=СУММ(A2^2)=1444]. В строке D14 – сумма этих значений, [=СУММ(D2:D11)=22115]. В строке D19 – средний квадрат [D19=СУММ(D14/D17)=2211,5].

Далее для формулы остается рассчитать квадрат среднего [D21=СУММ(A19^2)=2180,89]. Затем рассчитаем разницу между ними: [A27=СУММ(D19-D21)=30,61].

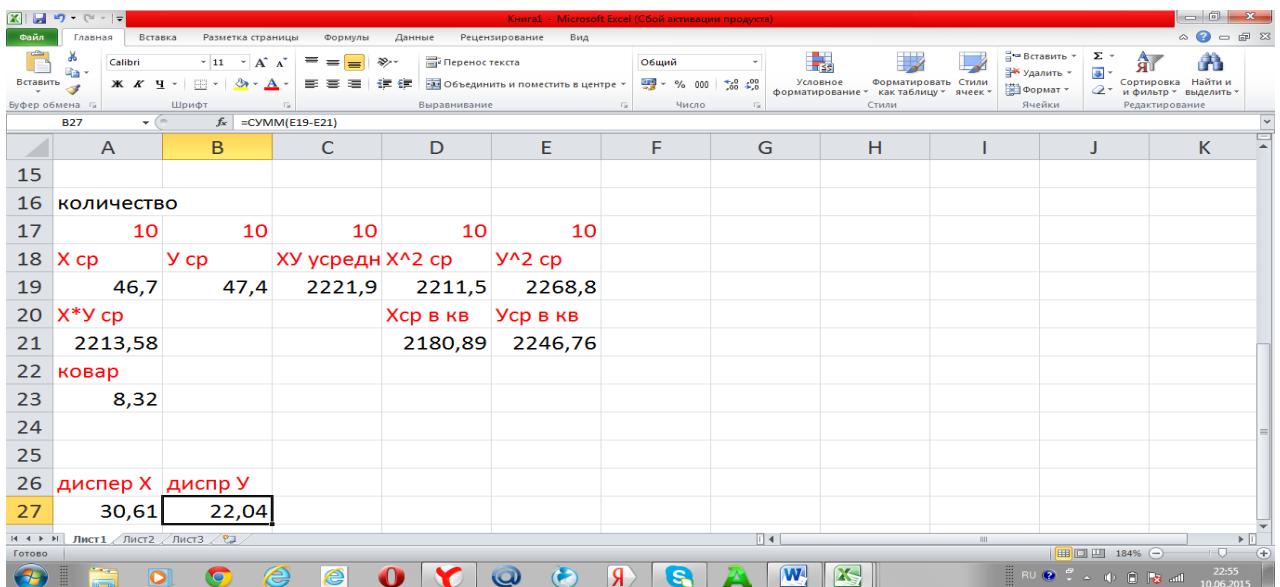


Рис. 6 Расчет дисперсии

4. Расчет сигмы δ (среднеквадратическое отклонение)

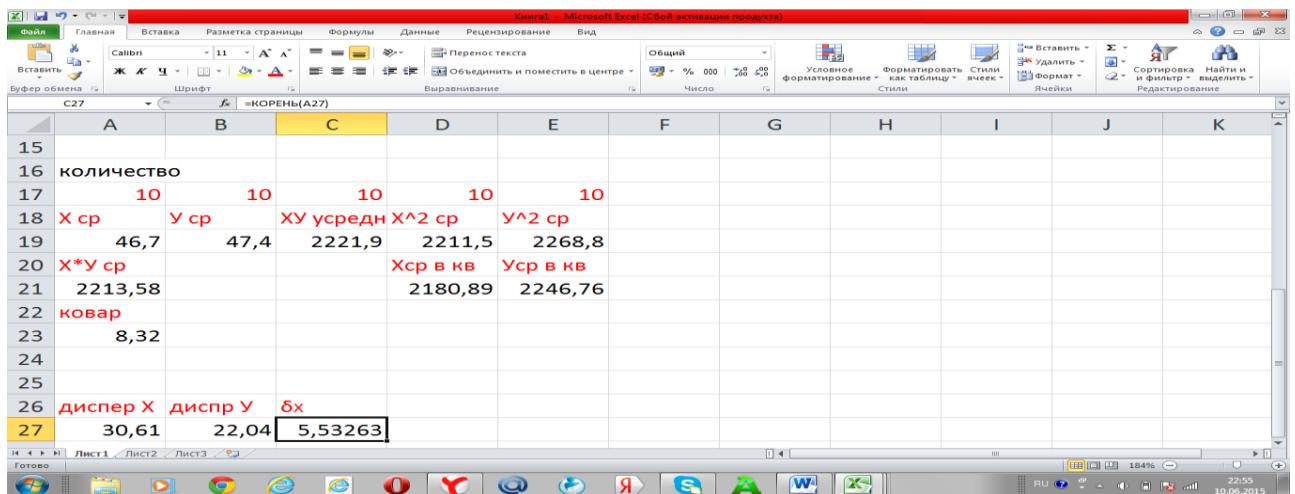
Сигмой (δ) в статистическом анализе обозначают стандартное отклонение.

Стандартное отклонение — это та погрешность, то « \pm сколько-то», которым обязательно сопровождают измерение величины.

Найдем сигму (δ) от X. Средним квадратическим отклонением случайной величины X (дискретной или непрерывной) называют квадратный корень из дисперсии:

$$\delta = \sqrt{D} = \text{КОРЕНЬ}(A27) = 5.53263$$

Где \sqrt{D} - квадратный корень из дисперсии.

Рис.6 Расчет δ_x среднеквадратического отклонения X

По аналогии найдем δ_y сигму от Y:

$$\delta_y = \text{КОРЕНЬ}(B27) = 4,69468$$

Книга1 - Microsoft Excel (Сбой активации продукта)											
Файл Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид											
	Буфер обмена	Стиль	Шрифт	Выравнивание	Формат	Условное форматирование	Стили	Форматировать как таблицу	Стили ячеек	Вставить	Удалить
D27			f _x =КОРЕНЬ(B27)								
15	A	B	C	D	E	F	G	H	I	J	K
16	количество										
17	10	10	10	10	10						
18	Xср	Yср	XУ усредн	X^2ср	Y^2ср						
19	46,7	47,4	2221,9	2211,5	2268,8						
20	X*Yср			Xср в кв	Yср в кв						
21	2213,58			2180,89	2246,76						
22	ковар										
23	8,32										
24											
25											
26	диспер X	диспр Y	δx	δy							
27	30,61	22,04	5,53263	4,69468							

Рис.7 Расчет δ_y среднеквадратического отклонения от Y

5. Расчет коэффициента корреляции. Корреляция - это статистическая взаимосвязь двух или нескольких случайных величин (либо величин, которые можно с некоторой допустимой степенью точности считать таковыми). При этом изменения значений одной или нескольких из этих величин сопутствуют систематическому изменению значений другой или других величин.

$$\text{Коррел.} = \frac{\text{Ковариация}(A23)}{[\delta X(C27)*\delta Y(D27)]} = \text{СУММ}(A23/C27*D27) = 7,05988$$

Книга1 - Microsoft Excel (Сбой активации продукта)											
Файл Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид											
	Буфер обмена	Стиль	Шрифт	Выравнивание	Формат	Условное форматирование	Стили	Форматировать как таблицу	Стили ячеек	Вставить	Удалить
E27			f _x =СУММ(A23/C27*D27)								
15	A	B	C	D	E	F	G	H	I	J	K
16	количество										
17	10	10	10	10	10						
18	Xср	Yср	XУ усредн	X^2ср	Y^2ср						
19	46,7	47,4	2221,9	2211,5	2268,8						
20	X*Yср			Xср в кв	Yср в кв						
21	2213,58			2180,89	2246,76						
22	ковар										
23	8,32										
24											
25											
26	диспер X	диспр Y	δx	δy	корел						
27	30,61	22,04	5,53263	4,69468	7,05988						

Рис. 8 Расчет коэффициента корреляции

Значение коэффициента корреляции изменяется от -1 до +1. Его отрицательное значение говорит о том, что между переменными наблюдается

обратная взаимосвязь. Например, когда доходность ценной бумаги будет расти, то доходность портфеля будет падать, и наоборот.

Если абсолютное значение коэффициента корреляции находится ближе к 1, то это свидетельствует о сильной взаимосвязи между переменными, а если ближе к 0 - то это говорит о слабой связи или ее отсутствии. Если его значение равно -1 или +1, то можно говорить о существовании функциональной взаимосвязи между переменными, то есть одну из них можно выразить через другую посредством математической функции.

6. Расчет эластичности

В статистике часто используют рассчитываемые на основе уравнений регрессии коэффициенты эластичности результирующего признака относительно факторного. Коэффициент эластичности (\mathcal{E}) показывает, на сколько изменяется в среднем результирующий признак (y) при изменении факторного признака (x) на одну единицу значения. Коэффициент эластичности для большинства форм связи величина переменная, т.е. изменяется с изменением фактора (x).

В строке A29 найдем эластичность, деля строку C27 на строку D27:

$$\mathcal{E} = \delta X(C27) / \delta Y(D27) = 1,17849$$

	A	B	C	D	E	F	G	H	I	J	K
17	10	10	10	10	10						
18	Xср	Уср	XУ усредн	X^2ср	У^2ср						
19	46,7	47,4	2221,9	2211,5	2268,8						
20	X*Yср			Xср в кв	Уср в кв						
21	2213,58			2180,89	2246,76						
22	ковар										
23	8,32										
24											
25											
26	диспер X	диспер У	δх	δу	корел						
27	30,61	22,04	5,53263	4,694678	7,059882						
28	эластич										
29	1,17849										
30											

Рис. 9 Расчет для уравнения регрессии коэффициента эластичности

Стоит отметить, что использование результатов эластичности следует вместе с значением коэффициента корреляции перемножая их. В дальнейшем это действие именуется критерием «k».

7. Расчет элементов уравнения регрессии

Для расчета уравнения регрессии ($Y=kX+b$) необходимо обратить внимание на критерии k и b . Первым найдем « k » критерий (B29). Для этого необходимо умножить корреляцию (E27) на эластичность (A29) или $k = \text{коррел.}^* (\sigma_{\text{X}} (\text{C27}) / \sigma_{\text{Y}} (\text{D27}))$.

$$k \text{ крит.} = E27 * A29 = 8,32$$

Книга1 - Microsoft Excel (Сбой активации продукта)										
Файл	Главная	Вставка	Разметка страницы	Формулы	Данные	Рецензирование	Вид			
		Сабри	11	A						
		Вставить								
		Шрифт								
		Выравнивание								
		Число								
		Форматирование								
		Стили								
		Стили ячеек								
		Формат ячеек								
		Редактирование								
B29		f _x	=E27*A29							
18	X ср	У ср	XУ усредн	X ² ср	У ² ср					
19	46,7	47,4	2221,9	2211,5	2268,8					
20	X*У ср			Xср в кв	Уср в кв					
21	2213,58			2180,89	2246,76					
22	ковар									
23	8,32									
24										
25										
26	диспер X	диспр У	δх	δу	корел					
27	30,61	22,04	5,53263	4,694678	7,059882					
28	эластич	к крит.								
29	1,17849	8,32								
30										
31										

Рис. 10 Расчет для уравнения регрессии « k » критерия

Далее рассчитаем « b » критерий (C29). Чтобы найти b критерий следует от \bar{Y} отнять произведение \bar{X} на « K » критерий.

$$b = \bar{Y} (B19) - K(A19) * \bar{X} (B29) = -341,144$$

Книга1 - Microsoft Excel (Сбой активации продукта)										
Файл	Главная	Вставка	Разметка страницы	Формулы	Данные	Рецензирование	Вид			
		Сабри	11	A						
		Вставить								
		Шрифт								
		Выравнивание								
		Число								
		Форматирование								
		Стили								
		Стили ячеек								
		Формат ячеек								
		Редактирование								
C29		f _x	=B19-(A19*B29)							
18	X ср	У ср	XУ усредн	X ² ср	У ² ср					
19	46,7	47,4	2221,9	2211,5	2268,8					
20	X*У ср			Xср в кв	Уср в кв					
21	2213,58			2180,89	2246,76					
22	ковар									
23	8,32									
24										
25										
26	диспер X	диспр У	δх	δу	корел					
27	30,61	22,04	5,53263	4,694678	7,059882					
28	эластич	к крит.								
29	1,17849	8,32			-341,144					
30										
31										

Рис. 11 Расчет для уравнения регрессии « b » критерия

Получив значение этих критериев можно найти Y теоретическое.

8. Расчет уравнения регрессии Y теоретического.

Теперь рассчитаем Y_{teor} . (уравнение регрессии).

$$Y_{teor} = b + (k * X_i)$$

Уравнение регрессии - это уравнение, отражающее изменение средней величины одного признака (y) в зависимости от второй (x).

$$Y_{teor} = C29 + (B29 * A2) = -24,98$$

The screenshot shows an Excel spreadsheet titled "Книга1 - Microsoft Excel (Сбой активации продукта)". The data is organized into columns A through F:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	X_i	Y_i	XY	X^2	Y^2	Y_{teor}																
2	38	42	1596	1444	1764	-24,98																
3	40	45	1800	1600	2025																	
4	42	49	2058	1764	2401																	
5	45	56	2520	2025	3136																	
6	49	54	2646	2401	2916																	
7	56	50	2800	3136	2500																	
8	54	48	2592	2916	2304																	
9	50	45	2250	2500	2025																	
10	48	44	2112	2304	1936																	
11	45	41	1845	2025	1681																	
12																						
13																						
14	467	474	22219	22115	22688																	
15																						
16	количество																					
17	10	10	10	10	10																	
18	$X_{ср}$	$Y_{ср}$	XY	X^2	Y^2																	
19	46,7	47,4	2221,9	2211,5	2268,8																	
20	$X^*Y_{ср}$	$X^*Y_{ср}$	X^2 средн X ² средн	Y^2 средн Y^2 средн																		
21	2213,58																					
22	новар																					
23	8,32																					
24																						

Рис. 12 Расчет уравнения регрессии Y теоретического

Аналогичную процедуру проделываем с каждым последующим значением.

The screenshot shows an Excel spreadsheet titled "Книга1 - Microsoft Excel (Сбой активации продукта)". The data is organized into columns A through F:

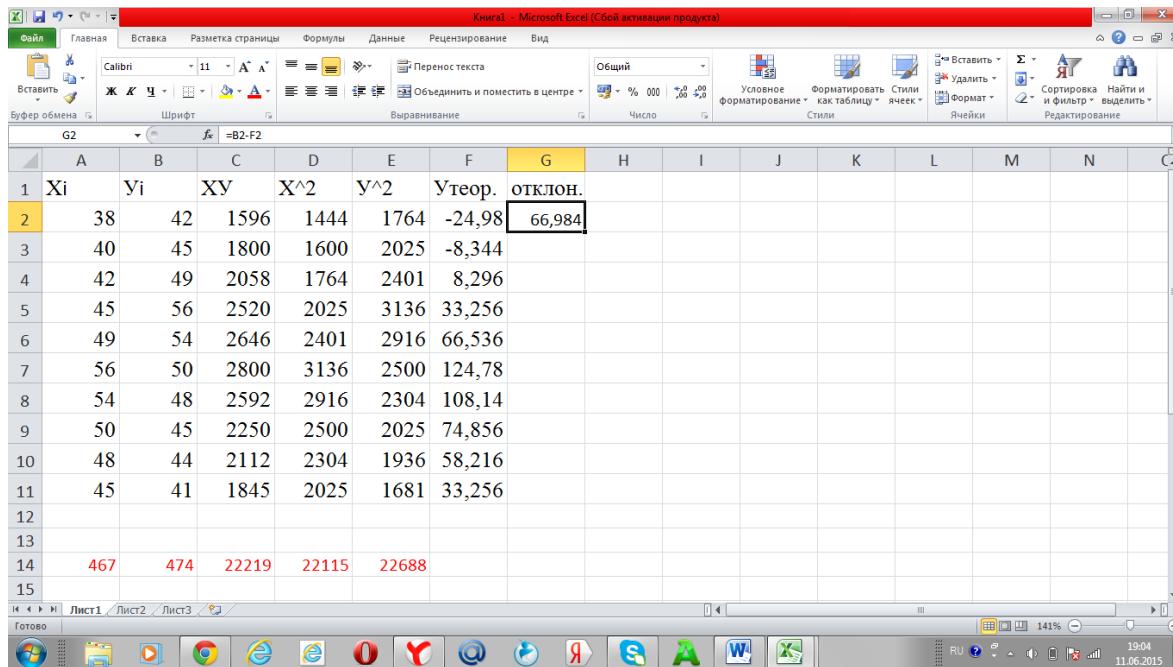
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	X_i	Y_i	XY	X^2	Y^2	Y_{teor}																
2	38	42	1596	1444	1764	-24,98																
3	40	45	1800	1600	2025	-8,344																
4	42	49	2058	1764	2401	8,296																
5	45	56	2520	2025	3136	33,256																
6	49	54	2646	2401	2916	66,536																
7	56	50	2800	3136	2500	124,78																
8	54	48	2592	2916	2304	108,14																
9	50	45	2250	2500	2025	74,856																
10	48	44	2112	2304	1936	58,216																
11	45	41	1845	2025	1681	33,256																
12																						
13																						
14	467	474	22219	22115	22688																	
15																						

Рис. 13 Расчет каждого значения уравнения регрессии Y теоретического

9. Расчет отклонения

Рассчитаем отклонение. Отклонение Y_i от $Y_{теор.}$ - это статистический термин, который является индикатором изменчивости. Он измеряет, на сколько широко значения рассеяны от фактического значения. Для расчета отклонения необходимо от Y_i отнять $Y_{теор.}$:

$$\text{Отклон. } Y_i \text{ от } Y_{теор.} = Y_i - Y_{теор.} = B2 - F2 = 66,984$$

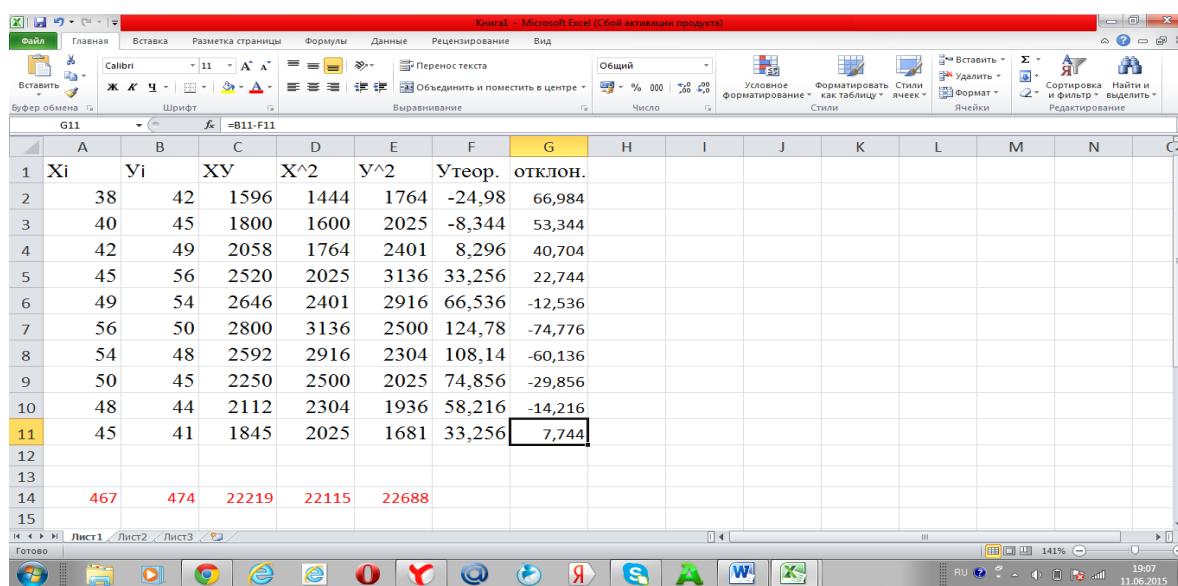


The screenshot shows a Microsoft Excel spreadsheet titled "Книга1 - Microsoft Excel (Сбой активации продукта)". The data is organized into columns A through G:

	A	B	C	D	E	F	G
1	X_i	Y_i	X^2	Y^2	Утеор.	отклон.	
2	38	42	1596	1444	1764	-24,98	66,984
3	40	45	1800	1600	2025	-8,344	
4	42	49	2058	1764	2401	8,296	
5	45	56	2520	2025	3136	33,256	
6	49	54	2646	2401	2916	66,536	
7	56	50	2800	3136	2500	124,78	
8	54	48	2592	2916	2304	108,14	
9	50	45	2250	2500	2025	74,856	
10	48	44	2112	2304	1936	58,216	
11	45	41	1845	2025	1681	33,256	
12							
13							
14	467	474	22219	22115	22688		
15							

Рис. 14 Расчет отклонение Y_i от $Y_{теор.}$

Остальные показатели рассчитаем аналогично первому



The screenshot shows a Microsoft Excel spreadsheet titled "Книга1 - Microsoft Excel (Сбой активации продукта)". The data is organized into columns A through G:

	A	B	C	D	E	F	G
1	X_i	Y_i	X^2	Y^2	Утеор.	отклон.	
2	38	42	1596	1444	1764	-24,98	66,984
3	40	45	1800	1600	2025	-8,344	53,344
4	42	49	2058	1764	2401	8,296	40,704
5	45	56	2520	2025	3136	33,256	22,744
6	49	54	2646	2401	2916	66,536	-12,536
7	56	50	2800	3136	2500	124,78	-74,776
8	54	48	2592	2916	2304	108,14	-60,136
9	50	45	2250	2500	2025	74,856	-29,856
10	48	44	2112	2304	1936	58,216	-14,216
11	45	41	1845	2025	1681	33,256	7,744
12							
13							
14	467	474	22219	22115	22688		
15							

Рис. 15 Расчет каждого значения отклонение Y_i от $Y_{теор.}$

После этого каждый из показателей отклонения возведем в квадрат.

$$\text{Откл.}^2 = G2^2 = 4486.86$$

The screenshot shows a Microsoft Excel spreadsheet titled "Книга1 - Microsoft Excel (Сбой активации продукта)". The data is organized into columns A through H. Columns A, B, C, D, E, and F contain numerical values. Column G contains the formula "Утеор." (Theoretical) followed by the value "откл.". Column H contains the formula "откл.^2" followed by the value "4486,86". The formula bar at the top shows "=G2^2". The status bar at the bottom right indicates the date as 11.06.2015 and the time as 19:11.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	C
1	Xi	Ui	XU	X^2	U^2	Утеор.	откл.	откл.^2							
2	38	42	1596	1444	1764	-24,98	66,984	4486,86							
3	40	45	1800	1600	2025	-8,344	53,344	2845,58							
4	42	49	2058	1764	2401	8,296	40,704	1656,82							
5	45	56	2520	2025	3136	33,256	22,744	517,29							
6	49	54	2646	2401	2916	66,536	-12,536	157,151							
7	56	50	2800	3136	2500	124,78	-74,776	5591,45							
8	54	48	2592	2916	2304	108,14	-60,136	3616,34							
9	50	45	2250	2500	2025	74,856	-29,856	891,381							
10	48	44	2112	2304	1936	58,216	-14,216	202,095							
11	45	41	1845	2025	1681	33,256	7,744	59,9695							
12															
13															
14	467	474	22219	22115	22688										
15															

Рис. 16 Расчет каждого значения отклонение в квадрате

Затем найдем сумму всех отклонений: =СУММ(Н2:Н11)

The screenshot shows a Microsoft Excel spreadsheet titled "Книга1 - Microsoft Excel (Сбой активации продукта)". The data is organized into columns A through H. Columns A, B, C, D, E, and F contain numerical values. Column G contains the formula "Утеор." (Theoretical) followed by the value "откл.". Column H contains the formula "=СУММ(Н2:Н11)" followed by the value "20024,9". The formula bar at the top shows "=СУММ(Н2:Н11)". The status bar at the bottom right indicates the date as 11.06.2015 and the time as 19:13.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	C
1	Xi	Ui	XU	X^2	U^2	Утеор.	откл.	откл.^2							
2	38	42	1596	1444	1764	-24,98	66,984	4486,86							
3	40	45	1800	1600	2025	-8,344	53,344	2845,58							
4	42	49	2058	1764	2401	8,296	40,704	1656,82							
5	45	56	2520	2025	3136	33,256	22,744	517,29							
6	49	54	2646	2401	2916	66,536	-12,536	157,151							
7	56	50	2800	3136	2500	124,78	-74,776	5591,45							
8	54	48	2592	2916	2304	108,14	-60,136	3616,34							
9	50	45	2250	2500	2025	74,856	-29,856	891,381							
10	48	44	2112	2304	1936	58,216	-14,216	202,095							
11	45	41	1845	2025	1681	33,256	7,744	59,9695							
12															
13															
14	467	474	22219	22115	22688			20024,9							
15															

Рис. 17 Расчет значения суммы всех отклонений

10. Расчет Еi-Еi-1

Рассчитаем $Ei-Ei-1$ -остатки, т.е. разницы значений результативного показателя от его предыдущего значения. Для этого необходимо от ячейки Н3 отнять ячейку Н2. Каждая последующая ячейка рассчитывается аналогично.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	X_i	Y_i	XY	X^2	Y^2	Утеор.	откл.	откл. $\cdot 2$	$Ei-Ei-1$	$(Ei-Ei-1)^2$				
2	38	42	1596	1444	1764	-24,98	66,984	4486,9	-	-				
3	40	45	1800	1600	2025	-8,344	53,344	2845,6	-13,64	-186,05				
4	42	49	2058	1764	2401	8,296	40,704	1656,8	-12,64	-159,77				
5	45	56	2520	2025	3136	33,256	22,744	517,29	-17,96	-322,562				
6	49	54	2646	2401	2916	66,536	-12,54	157,15	-35,28	-1244,68				
7	56	50	2800	3136	2500	124,78	-74,78	5591,5	-62,24	-3873,82				
8	54	48	2592	2916	2304	108,14	-60,14	3616,3	14,64	214,33				
9	50	45	2250	2500	2025	74,856	-29,86	891,38	30,28	916,878				
10	48	44	2112	2304	1936	58,216	-14,22	202,09	15,64	244,61				
11	45	41	1845	2025	1681	33,256	7,744	59,97	21,96	482,242				
12														
13														
14	467	474	22219	22115	22688					20024,9				
15														

Рис. 18 Расчет значений $Ei-Ei-1$

После этого возведем показатель $Ei-Ei-1$ в квадрат.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	X_i	Y_i	XY	X^2	Y^2	Утеор.	откл.	откл. $\cdot 2$	$Ei-Ei-1$	$(Ei-Ei-1)^2$				
2	38	42	1596	1444	1764	-24,98	66,984	4486,9	-	-				
3	40	45	1800	1600	2025	-8,344	53,344	2845,6	-13,64	186,05				
4	42	49	2058	1764	2401	8,296	40,704	1656,8	-12,64	159,77				
5	45	56	2520	2025	3136	33,256	22,744	517,29	-17,96	322,562				
6	49	54	2646	2401	2916	66,536	-12,54	157,15	-35,28	1244,68				
7	56	50	2800	3136	2500	124,78	-74,78	5591,5	-62,24	3873,82				
8	54	48	2592	2916	2304	108,14	-60,14	3616,3	14,64	214,33				
9	50	45	2250	2500	2025	74,856	-29,86	891,38	30,28	916,878				
10	48	44	2112	2304	1936	58,216	-14,22	202,09	15,64	244,61				
11	45	41	1845	2025	1681	33,256	7,744	59,97	21,96	482,242				
12														
13														
14	467	474	22219	22115	22688					20024,9				
15														

Рис. 19 Расчет значений $Ei-Ei-1$ в квадрате

Затем найдем сумму этих показателей в строке J14 $[(Ei-Ei-1)^2]$:

$$J14=\text{СУММ}(J3:J11)=7644,94$$

The screenshot shows a Microsoft Excel spreadsheet titled "Книга1 - Microsoft Excel (Сбой активации продукта)". The data is organized in columns A through C and rows 1 through 14. Column A contains values for X_i , column B for Y_i , and column C contains formulas and results. Row 14 is highlighted in yellow and contains the formula =SUMM(J3:J11) in cell J14, which calculates the sum of the squared differences in row 11. The value 7644,94 is displayed in cell H14.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	X_i	Y_i	X^2	Y^2	Утеор.	откл.	откл.^2	$Ei-Ei-1$	$(Ei-Ei-1)^2$					
2	38	42	1596	1444	1764	-24,98	66,984	4486,9	-	-				
3	40	45	1800	1600	2025	-8,344	53,344	2845,6	-13,64	186,05				
4	42	49	2058	1764	2401	8,296	40,704	1656,8	-12,64	159,77				
5	45	56	2520	2025	3136	33,256	22,744	517,29	-17,96	322,562				
6	49	54	2646	2401	2916	66,536	-12,54	157,15	-35,28	1244,68				
7	56	50	2800	3136	2500	124,78	-74,78	5591,5	-62,24	3873,82				
8	54	48	2592	2916	2304	108,14	-60,14	3616,3	14,64	214,33				
9	50	45	2250	2500	2025	74,856	-29,86	891,38	30,28	916,878				
10	48	44	2112	2304	1936	58,216	-14,22	202,09	15,64	244,61				
11	45	41	1845	2025	1681	33,256	7,744	59,97	21,96	482,242				
12														
13														
14	467	474	22219	22115	22688			20024,9	7644,94					
15														

Рис. 20 Расчет значения суммы всех $Ei-Ei-1$ в квадрате

11. Расчет остаточной дисперсии

Найдем остаточную дисперсию которая показывает воздействие других признаков на изучаемый признак ($Дост.=\sum Ei-Ei-1/(n-2)$). Для этого нам необходимо сумму отклонений возвещенных в квадрат разделить на разницу n и двух:

$$\text{Дост.} = H14/(A27-2) = 699,928$$

The screenshot shows a Microsoft Excel spreadsheet titled "Книга1 - Microsoft Excel (Сбой активации продукта)". The data is organized in columns A through F and rows 13 through 27. Row 27 is highlighted in yellow and contains the formula =H14/(A27-2) in cell F27, which calculates the residual variance. The value 699,928 is displayed in cell H27.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
13														
14	467	474	22219	22115	22688			20024,9	7644,94					
15														
16	количество													
17	10	10	10	10	10	10								
18	$X_{ср}$	$Y_{ср}$	XU	усредн.	X^2	$ср$	Y^2	$ср$						
19	46,7	47,4	2221,9		2211,5		2268,8							
20	X^*Y	$ср$			$X_{ср}$	в кв	$Y_{ср}$	в кв						
21	2213,58				2180,89		2246,76							
22	ковар													
23	8,32													
24														
25														
26	диспер X	диспр Y	бх	бу	корел	Дост.	Двнутр.							
27	30,61	22,04	5,53263	4,69468	7,05988	699,928								
28	эластич	К крит.	В крит.											
29	1,17849	8,32	-341,144											
30														

Рис. 21 Расчет значения остаточной дисперсии

12. Расчет внутренней дисперсии

Произведем расчет внутренней дисперсии. Для этого необходимо сумму отклонений, возведенных в квадрат разделить на дисперсию Х:

$$D_{\text{внутр.}} = H14/A27 = 654,196$$

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
13														
14	467	474	22219	22115	22688			20024,9		7644,94				
15														
16	количество													
17	10	10	10	10	10									
18	Xср	Уср	XУ усредн.	X^2ср	У^2ср									
19	46,7	47,4	2221,9	2211,5	2268,8									
20	X*Уср		Xср в кв	Уср в кв										
21	2213,58		2180,89	2246,76										
22	ковар													
23	8,32													
24														
25														
26	диспер X	диспер У	δх	δу	корел	Дост.	Dвнутр.							
27	30,61	22,04	5,53263	4,69468	7,05988	699,928	654,196							
28	эластич	К крит.	В крит.											
29	1,17849	8,32	-341,144											
30														

Рис. 22 Расчет значения внутренней регрессии

13. Расчет критерия Дарбина Уотсона (DW)

Найдем показатель Дарбина-Уотсона (DW)- это условный показатель, который применяется для выявления автокорреляции во временных рядах.

$$DW = \frac{(\sum E_i - E_i - 1)^2}{\sum \text{откл.}^2}$$

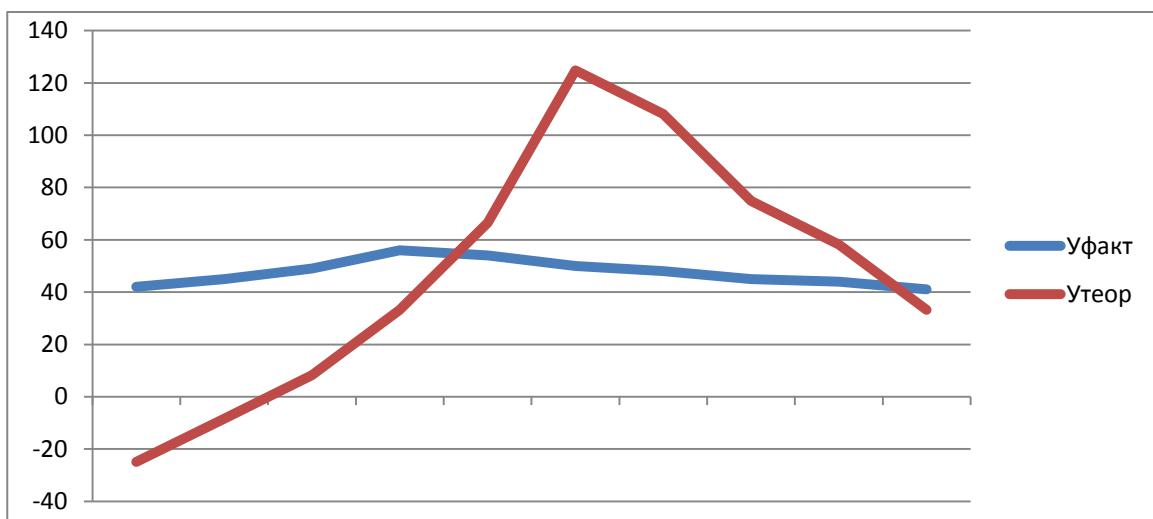
Значение статистики Дарбина-Уотсона изменяется в диапазоне от 0 до 4. При этом $d = 2$ указывает на отсутствие автокорреляции элементов временного ряда. Если d меньше двух, то имеет место положительная автокорреляция, а больше двух - отрицательная.

$$DW = J14/H14 = 0,38177$$

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	C
13															
14	467	474	22219	22115	22688			20024,9		7644,94					
15															
16	количество														
17	10	10	10	10	10										
18	Xср	Уср	XУусредн.	X^2ср	У^2ср										
19	46,7	47,4	2221,9	2211,5	2268,8										
20	X*Уср		Xср в кв	Уср в кв											
21	2213,58			2180,89	2246,76										
22	ковар														
23	8,32														
24															
25															
26	диспер	Х диспер	Y дх	Yу	корел	Дост.	Двнутр.	DW	о ост						
27	Х	22,04	5,53263	4,69468	7,05988	699,928	654,196	0,38177							
28	эластич	К крит.	В крит.												
29	1,17849	8,32	-341,144												
30															

Рис. 23 Расчет критерия Дарбина Уотсона (DW)

Построение графика отклонения $Y_{факт}$ и $Y_{теор}$.

Рис. 24 График отклонения $Y_{факт}$ и $Y_{теор}$.

В процессе изучения сущности корреляции и регрессии были рассмотрены такие статистические показатели как: средняя арифметическая простая, ковариация, дисперсия, сигма (среднеквадратическое отклонение), корреляция, критерий уравнения регрессии k и b , уравнение регрессии, отклонения, остаточная внутренняя дисперсия, критерий *Дарбина-Уотсона* и остаточная сигма. В методичке было рассмотрено значение каждого из вышеперечисленных показателей, были приведены формулы их расчета. Методическое пособие содержит теоретическую и практическую основу.

14. Пример расчета индекса сезонности с использованием метода регрессии и корреляции

По условию примера будем рассматривать вымышленное предприятие и цифры, которые характеризуют загрузку санатория в условиях сезонности. В табл.1 представлены коэффициенты сезонности по месяцам за год. Они рассчитывались простым делением Y_i / Y (теор). В последней колонке в нижнем углу получен общий коэффициент сезонности, который равен 0,84, что можно выразить как непостоянное использование 84% мощности санатория в течение года.

Для количественного описания взаимосвязей между представленными переменными X_i (месяцы) и Y_i (загрузка) использовался метод регрессии и корреляции.

Таблица 1

Сезонная загрузка санатория

Модель вида, где Y_i - зависимая переменная (результативный признак); X_i - независимая, или объясняющая, переменная (признак-фактор).

Линия регрессии - график функции $y = f(x)$.

Два типа взаимосвязей между x и y : 1) может быть неизвестно, какая из двух переменных является независимой, а какая - зависимой, переменные равноправны, это взаимосвязь корреляционного типа; 2) если x и y неравноправны и одна из них рассматривается как объясняющая (независимая) переменная, а другая - как зависимая, то это взаимосвязь регрессионного типа.

В нашем случае значение Y зависит от значения X .

Коэффициент корреляции величин x и y (r_{xy}) свидетельствует о наличии или отсутствии линейной связи между переменными:

$$r_{xy} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} = \frac{xy * x * y}{\sigma_x * \sigma_y}; \quad r_{xy} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

где $r_{xy} \in (-1; 1)$. Если: $r_{xy} = -1$, то наблюдается строгая отрицательная связь;

$r_{xy} = 1$, то наблюдается строгая положительная связь; $r_{xy} = 0$, то линейная связь отсутствует.

Вычисление промежуточных расчетов представлены в табл. 2.

Таблица 2

Расчёт основных показателей для анализа сезонности загрузки

x сред	y сред	xy усредн	x^2 сред	y^2 сред			
6,50	83,83	579,25	54,17	11358,17			
x *y сред			x ср в кв	y ср в кв			
544,92			42,25	7028,03			
ковар							
34,33							
дисп у	σ_u	корреляц	крит к	крит в	Дост	r	σ ост
4330,14	65,80	0,15	2,88	65,11	5077,46	0,15	71,26
дисп x	σ_x	элласт			Двнутр	DW	
11,92	3,45	19,06			4231,22	0,41	

Экономический смысл параметров уравнения линейной парной регрессии.

Параметр b показывает среднее изменение результата y с изменением фактора x на единицу.

Сезонный характер загрузки мощностей – существенный фактор, сдерживающий движение к финансовому благополучию финансовых санатория. Проблема ярко выраженной сезонности в работе санаторно-курортного комплекса стоит наиболее остро. Период курортной активности регионов санатория как и в общем по региону составляет три летних месяца, определяющих «высокий сезон».

Это период наполнения бюджета города, курорта и солидных отчислений в вышестоящие бюджеты.

Динамика загрузки санатория, которая выражает сезонность (расчет в табл. 1 и 2), представлена на рис. 25.

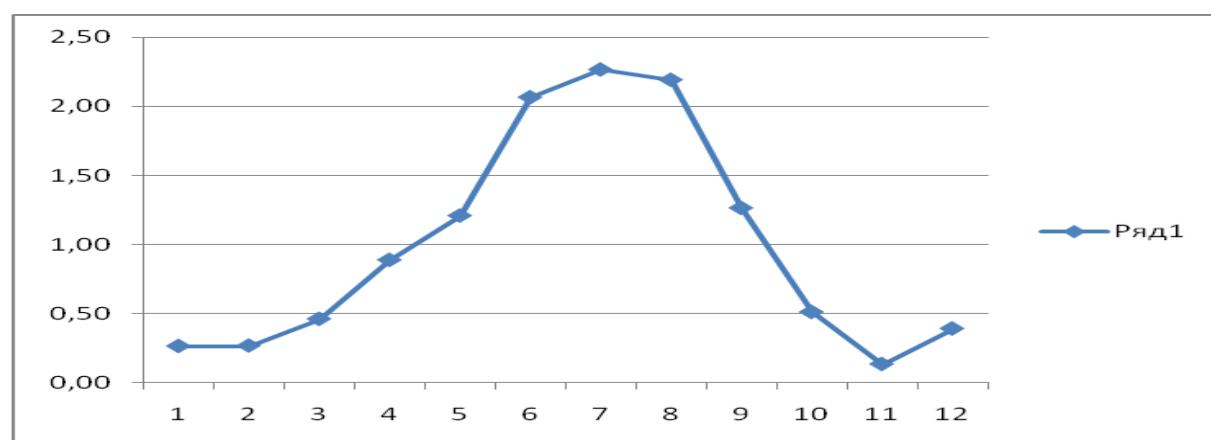


Рис. 25 Сезонность по месяцам санатория

Сезонность стимулирует высокую динамику цен на путевки в сезон повышенного спроса (июнь – сентябрь). Уровень цен на путевки в период повышенного спроса слишком высок и даже недоступен для подавляющей части работающих граждан. В этой связи: резерв роста финансовых ресурсов санатория – снижение цен на путевки в период межсезонья. Снижение стоимости путевок способствовало бы росту спроса на отдых и лечение, положительно сказавшись на показателях загрузки.

Задачи для самостоятельного выполнения

Выполните задания в соответствии с вариантами ниже.

Задача 1

Рассчитать дисперсию и ковариацию X и Y, если каждый последующий X после 6 элемента будет меньше на 5 предыдущего, а и Y на 10 больше предыдущего. Дать свое пояснение как поменяется значение коэффициента корреляции после этих изменений.

Задача 2

1. Найти средние значения X и Y
2. Вычислить дисперсии и коэффициенты вариации X и Y
3. Найти коэффициент корреляции X и Y
4. Найти коэффициенты уравнения регрессии
5. Найти остаточную дисперсию регрессии
6. Найти остаточное среднее квадратичное отклонение регрессии
7. Определить автокорреляцию
8. Представить графически теоретическое и фактическое значения Y
9. Спрогнозировать Y на 3 года вперед
10. Данные расчеты произвести в электронной таблице

Задача 3

1. Найти коэффициент корреляции Y с ее предыдущим значением $X = X - 1$
2. Найти коэффициент корреляции Y с ее предыдущим значением $X = X - 2$
3. Найти коэффициенты уравнения регрессии Y с ее предыдущим значением $X = X - 1$
4. Найти коэффициенты уравнения регрессии Y с ее предыдущим значением $X = X - 2$
5. Найти остаточную дисперсию регрессии Y с предыдущим значением $X = X - 1$
6. Найти остаточную дисперсию регрессии Y с предыдущим значением $X = X - 2$
7. Найти остаточное среднее квадратичное отклонение регрессии Y с ее предыдущим значением $X = X - 1$
8. Найти остаточное среднее квадратичное отклонение регрессии Y с ее предыдущим значением $X = X - 2$
9. Определить автокорреляцию
10. Данные расчеты произвести в электронной таблице

Задача 4

Согласно варианту заменим все значения X на последовательность цифр от 1 до 14, а значение Y оставим согласно существующему условию варианта.

- a) сделать прогноз значения Y от 15 до 18
- б) представив первые 12 значений за месяцы рассчитать индекс сезонности.

Варианты заданий

1	X	18	20	30	40	50	60	70	80	90	100	110	120	130	180
	Y	18	19	34	48	67	56	48	35	31	28	27	29	35	48
2	X	1	2	3	4	5	6	7	8	9	10	11	12	9	10
	Y	18	19	34	48	67	56	48	35	31	28	27	29	31	28
3	X	50	60	70	80	90	40	50	60	70	80	90	100	110	120
	Y	67	56	48	35	31	48	67	56	48	35	31	28	56	48
4	X	5	6	30	40	50	60	70	80	90	100	110	30	60	70
	Y	67	56	34	48	67	56	48	35	31	28	27	34	56	48
5	X	31	28	27	34	56	48	31	80	90	100	110	30	60	70
	Y	18	19	34	48	67	56	48	35	31	28	27	34	56	48
6	X	31	28	31	28	27	34	56	48	31	28	27	30	48	35
	Y	18	19	34	48	67	56	48	35	31	28	27	34	50	60
7	X	5	6	7	67	56	48	35	34	48	67	5	6	67	56
	Y	67	56	48	90	40	50	60	3	4	5	67	56	50	60
8	X	90	40	50	31	48	67	56	34	48	67	90	40	67	56
	Y	31	48	67	9	40	50	60	70	80	90	31	48	56	6
9	X	9	40	50	31	48	67	56	48	35	31	9	40	6	7
	Y	31	48	67	48	67	56	6	7	8	9	31	48	50	60
10	X	50	60	70	4	5	6	7	8	9	10	50	60	56	48
	Y	18	19	34	30	40	50	60	70	80	90	100	110	60	70
11	X	5	6	7	34	48	67	56	48	35	31	28	27	56	48
	Y	67	56	48	30	40	50	60	70	80	90	100	110	60	70
12	X	90	40	50	34	48	67	56	48	35	31	28	27	60	70
	Y	31	48	67	30	40	50	60	70	80	90	100	110	56	48
13	X	9	40	50	34	48	67	56	48	35	31	28	27	60	70
	Y	18	19	6	7	8	9	6	7	8	31	48	67	56	48
14	X	1	2	3	4	5	6	7	8	9	50	60	70	80	90
	Y	18	19	34	48	67	56	48	35	31	28	27	29	31	28
15	X	5	6	7	8	9	10	5	6	7	8	9	10	5	6
	Y	67	56	48	35	31	28	67	56	48	35	31	28	67	56
16	X	90	40	50	60	70	80	90	40	50	60	70	80	90	40
	Y	31	48	67	56	48	35	31	48	67	56	48	35	31	48
17	X	9	40	50	60	70	80	9	40	50	60	70	80	9	40
	Y	31	48	67	56	48	35	31	48	67	56	48	35	31	48
18	X	1	2	3	4	5	6	7	8	9	10	11	12	9	10
	Y	18	19	34	48	67	56	48	35	31	28	27	29	31	28
19	X	5	6	7	8	9	10	5	6	7	8	9	10	5	6
	Y	67	56	48	35	31	28	67	56	48	35	31	28	67	56
20	X	90	40	50	60	70	80	90	40	50	60	70	80	90	40
	Y	31	48	67	56	48	35	31	48	67	56	48	35	31	48
21	X	18	20	30	40	50	60	70	80	90	100	110	120	130	180
	Y	18	19	34	48	67	56	48	35	31	28	27	29	35	48
22	X	1	2	3	4	5	6	7	8	9	10	11	12	9	10
	Y	18	19	34	48	67	56	48	35	31	28	34	56	48	31
23	X	50	60	70	80	90	40	50	60	70	80	56	48	35	31
	Y	67	56	48	35	31	48	67	56	48	35	48	35	34	48
24	X	5	6	30	40	50	60	70	80	90	100	50	60	3	4
	Y	67	56	34	48	67	56	48	35	31	28	67	56	34	48

25	34	56	48	31	34	56	48	67	48	90	100	110	30	60	70
	56	48	35	31	48	67	8	9	8	31	28	27	34	56	48
26	48	35	34	48	28	27	35	31	35	31	28	27	30	48	35
	50	60	3	4	48	67	60	70	60	56	48	31	34	50	60
27	67	56	34	48	67	56	56	48	56	48	35	31	6	67	56
	Y	67	56	48	90	40	60	70	60	35	34	48	56	50	60
28	X	90	40	50	31	48	56	48	56	60	3	4	40	67	56
	Y	31	48	67	9	40	4	5	4	56	34	48	48	56	6
29	X	9	40	50	31	48	48	67	48	60	70	80	40	6	7
	Y	31	48	67	48	67	8	9	8	56	48	35	48	50	60
30	X	50	60	70	4	5	35	31	35	6	7	8	60	56	48
	Y	18	19	34	30	40	50	60	6	7	8	9	110	60	70

Литература

1. Балдин, К.В. Общая теория статистики: Учебное пособие / К.В. Балдин, А.В. Рукосуев. - М.: Дашков и К, 2012. - 312 с.
2. Батракова, Л.Г. Теория статистики: Учебное пособие / Л.Г. Батракова. - М.: КноРус, 2013. - 528 с.
3. Годин А.М. Статистика: Учебник. – М.: Дашков и К', 2008.
4. Гусаров, В.М. Общая теория статистики: Учебное пособие для студентов вузов / В.М. Гусаров, С.М. Проява. - М.: ЮНИТИ-ДАНА, 2008. - 207 с.
5. Елисеева И.И. Эконометрика. Учебник / Под ред. Елисеевой И.И.. — 2-е изд. — М.: Финансы и статистика, 2006. — 576 с.
6. Елисеева И.И., Юзбашев М.М. Общая теория статистики: Учебник/ Под ред. чл.-корр. РАН И.И. Елисеевой. - 4-е изд., перераб. и доп, — М.: Финансы и статистика, 2001. - 480 с.:
7. Минашкин В.Г., Шмойлова Р.А. Теория статистики. Учебно-методический комплекс. М.: ЕАОИ, 2008, — 296 с.
8. Мхитарян В.С., Агапова Т.Н., Статистика : учебник для бакалавров / под ред. В. С. Мхитаряна. – М. : Издательство Юрайт, 2013. – 590 с. – Серия : Бакалавр. Базовый курс.
9. Сизова Т.М. Статистика: Учебное пособие. – СПб.: СПб ГУИТМО, 2005. – 80 с.
10. Чалиев А.А., Овчаров А.О. Статистика. Учебно-методическое пособие. Часть 1. – Нижний Новгород: Издательство Нижегородского госуниверситета, 2007.– 87 с.