



Практическое занятие 6

Проведение регрессионного анализа и оценка результатов

Цель работы – приобрести практические навыки проведения регрессионного анализа и оценки его

результатов

Задание

1. Определить коэффициенты линейной множественной регрессии.
2. Определить дисперсии (общая, факторная, остаточная) и корреляционную связь признака и факторов.
3. Определить коэффициент детерминации и множественный R .
4. Определить критерий Фишера и ошибки аппроксимации.
5. Определить t критерий и границы доверительной вероятности.
6. Произвести оценку значений определенных параметров.
7. Провести регрессионный анализ с помощью MS Excel.
8. Сделать выводы по работе.

Исходные данные – результаты работы №5

№ опыта	X1	X2	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12	Y13	Y14	Y15	Y16	Y17	Y18	Y19	Y20	Ycp
1	1	50	2,79	2,82	2,94	2,85	2,89	2,98	2,75	2,99	2,74	2,71	2,85	2,58	2,78	2,80	2,79	2,83	2,89	2,98	2,97	3,08	2,83
2	1	300	13,38	13,91	14,23	12,79	14,12	14,05	12,92	13,51	13,53	13,37	13,94	13,04	13,88	14,08	13,79	13,38	13,24	15,05	13,83	14,82	13,72
3	10	50	8,90	8,28	9,08	8,27	8,70	9,83	9,24	10,98	8,00	7,88	8,58	7,02	7,99	7,70	8,11	9,39	7,90	8,02	10,18	9,98	8,89
4	10	300	19,48	19,34	20,35	18,41	19,93	20,72	19,41	21,50	18,79	18,54	19,87	17,48	18,90	18,98	19,11	19,92	18,45	20,10	21,02	21,52	19,58

1. Определение коэффициентов линейной множественной регрессии

Линейный полином

$$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2$$

Для получения вида математической модели необходимо определить коэффициенты уравнения регрессии b_0 , b_1 и b_2 . решим систему трехлинейных уравнений с тремя неизвестными b_0, b_1, b_2 :

$$\begin{cases} \sum y_i = nb_0 + b_1 \sum x_{i1} + b_2 \sum x_{i2}; \\ \sum x_{i1}y_i = b_0 \sum x_{i1} + b_1 \sum x_{i1}^2 + b_2 \sum x_{i1}x_{i2}; \\ \sum x_{i2}y_i = b_0 \sum x_{i2} + b_1 \sum x_{i1}x_{i2} + b_2 \sum x_{i2}^2. \end{cases}$$

Для решения системы можете воспользоваться [решение системы методом Крамера](#)

Определим значения используя формулы

$$b_0 = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2;$$

$$b_1 = \frac{\sum (x_{i1} - \bar{x}_1)(y_i - \bar{y}) \cdot \sum (x_{i2} - \bar{x}_2)^2 - \sum (x_{i2} - \bar{x}_2)(y_i - \bar{y}) \cdot \sum (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{\sum (x_{i1} - \bar{x}_1)^2 \sum (x_{i2} - \bar{x}_2)^2 - (\sum (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2))^2};$$

$$b_2 = \frac{\sum (x_{i2} - \bar{x}_2)(y_i - \bar{y}) \cdot \sum (x_{i1} - \bar{x}_1)^2 - \sum (x_{i1} - \bar{x}_1)(y_i - \bar{y}) \cdot \sum (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)}{\sum (x_{i1} - \bar{x}_1)^2 \sum (x_{i2} - \bar{x}_2)^2 - (\sum (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2))^2}.$$

Для решения построим таблицу

Y	x ₁	x ₂	(y-y _{ср}) ²	(x ₁ -x _{1ср}) ²	(x ₂ -x _{2ср}) ²	(y-y _{ср})(x ₁ -x _{1ср})	(y-y _{ср})(x ₂ -x _{2ср})	(x ₁ -x _{1ср})(x ₂ -x _{2ср})

На основании полученных значений
строим функцию

2. Определение дисперсий и корреляционной связи признака и факторов

Общая дисперсия результативного признака Y , отображающую влияние как основных, так и остаточных факторов:

$$\sigma_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n},$$

где \bar{y} - среднее значение результативного признака Y .

Факторная дисперсия результативного признака Y ,
отображающую влияние только основных факторов

$$\sigma_F^2 = \frac{\sum_{i=1}^n (\bar{y}_i - \bar{y})^2}{n};$$

Остаточная дисперсия результативного признака Y ,
отображающую влияние только остаточных факторов

$$\sigma_O^2 = \frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{n - (m + 1)}.$$

При корреляционной связи результативного признака и факторов
выполняется соотношение

$$\sigma_F^2 < \sigma_Y^2,$$

при этом

$$\sigma_Y^2 = \sigma_F^2 + \sigma_O^2.$$

3. Определение коэффициента детерминации множественного R

Для анализа общего качества уравнения линейной многофакторной регрессии используют множественный коэффициент детерминации

$$R^2 = \frac{\sigma_F^2}{\sigma_y}$$

Множественный R равен квадратному корню из коэффициента детерминации, эта величина принимает значения в интервале от нуля до единицы.

4. Определение критерия Фишера и ошибки аппроксимации.

F-критерий Фишера

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m},$$

где n – число наблюдений;

m – число факторов в уравнении регрессии.

Показатель средней ошибки аппроксимации

$$\bar{\varepsilon} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}|}{y_i} \cdot 100\%.$$

5. Определение t критерия и границы доверительной вероятности.

Показатель средней ошибки аппроксимации

$$t = \frac{a_i}{\sigma_{a_i}},$$

где σ_{a_i} - стандартное значение ошибки для коэффициента регрессии

Границы доверительных интервалов для коэффициентов регрессии

$$a_i^{\min} = a_i - t_{kp} \sigma_{a_i};$$

$$a_i^{\max} = a_i + t_{kp} \sigma_{a_i}.$$

6. Оценка значений определенных параметров.

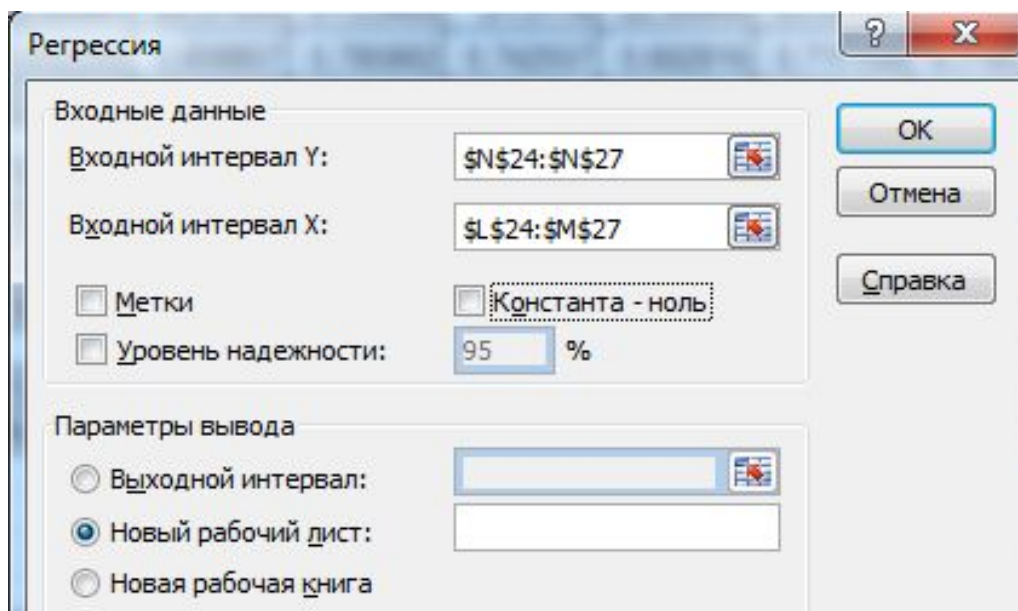
- 1) Проверяем R^2 – должен стремиться к 1.
- 2) Проверка значения критерия Фишера - должен быть в интервале $(F_{кр}; +\infty)$, определяется в Excel =FРАСПОБР(0,05;кп;ко)
- 3) Оценка значимости – значение должно быть меньше 0,05.
- 4) Проверка t-критерия - должен попадать в критическую область: $(-\infty; t_{кр}) \cup (t_{кр}; +\infty)$. Определяется: =СТЮДРАСПРОБР(0,05; n-k-1).
- 5) Проверка нижних и верхних границ доверительных интервалов - не должен проходить через 0.

7. Регрессионный анализ с помощью MS Excel.

Исходные данные для регрессионного анализа

№ опыта	X1	X2	Y _{ср}
1	1	50	123,1551
2	1	300	568,5452
3	10	50	429,8486
4	10	300	875,2387

Ввод данных в MS Excel.



Результаты регрессионного анализа в MS Excel.

	A	B	C	D	E	F	G
1	ВЫВОД ИТОГОВ						
2							
3	<i>Регрессионная статистика</i>						
4	Множественный R	1					
5	R-квадрат	1					
6	Нормированный R-кв	1					
7	Стандартная ошибка	5,96856E-14					
8	Наблюдения	4					
9							
10	<i>Дисперсионный анализ</i>						
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>	
12	Регрессия	2	292433,2458	146216,6229	4,10448E+31	1,10371E-16	
13	Остаток	1	3,56237E-27	3,56237E-27			
14	Итого	3	292433,2458				
15							
16		<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
17	Y-пересечение	5,68434E-14	6,29805E-14	0,902555707	0,532577656	-7,434E-13	8,57087E-13
18	Переменная X 1	34,07705953	6,63173E-15	5,13849E+15	1,23892E-16	34,07705953	34,07705953
19	Переменная X 2	1,781560309	2,38742E-16	7,46227E+15	8,53118E-17	1,781560309	1,781560309
20							

8. Формирование выводов

Необходимо представить анализ оценки регрессионного анализа произведенного вручную и с помощью MS Excel. Далее сравниваем полученные результаты. Даем рекомендации по использованию моделей для прогнозирования.