

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Локтионова Оксана Геннадьевна

Должность: проректор по учебной работе

Дата подписания: 08.09.2026 03:49:47

Уникальный программный ключ:

0b817ca911e6668abb13a5d426d39e5f1c11eabbf73e943df4a4851fda56d089

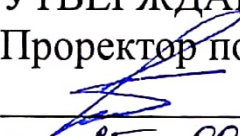
МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра биомедицинской инженерии

УТВЕРЖДАЮ

Проректор по учебной работе


О.Г. Локтионова

« 25 » 09



СТАНДАРТНЫЕ ПРОГРАММНЫЕ СРЕДСТВА В ИМИТАЦИОННОМ МОДЕЛИРОВАНИИ БИОТЕХНИЧЕСКИХ СИСТЕМ

Методические указания по выполнению практических занятий
для студентов направления подготовки
12.03.04 «Биотехнические системы и технологии»

Курск 2023

УДК 004.93:61

Составители: О.В. Шаталова.

Рецензент

Кандидат технических наук, доцент *Т.Н. Конаныхина*

Стандартные программные средства в имитационном моделировании биотехнических систем: методические указания по выполнению практических занятий для студентов направления подготовки 12.03.04 «Биотехнические системы и технологии» / Юго-Зап. гос. ун-т; сост.: О.В. Шаталова. - Курск, 2023. - 88 с.

Предназначено для студентов направления подготовки 12.03.04 «Биотехнические системы и технологии» по дисциплине «Стандартные программные средства в имитационном моделировании биотехнических систем». Может быть использована аспирантами, обучающимися по направлениям подготовки 2.3.1. Системный анализ, управление и обработка информации, статистика и 2.2.12. Приборы, системы и изделия медицинского назначения.

Текст печатается в авторской редакции

Подписано в печать . Формат 60×84 1/16.
Усл. печ. л. 5,12. Уч.-изд. л. 4,63. Тираж 100 экз. Заказ *940*. Бесплатно.
Юго-Западный государственный университет.
305040, г. Курск, ул. 50 лет Октября, 94.

Практическое занятие №1

Статистическая обработка результатов эксперимента

Цель занятия: изучить элементарные статистические процедуры при выполнении экспериментальных исследований и методы оценивания основных характеристик случайных величин.

Теоретические сведения

Весь ход статистического изучения заключается в следующем: производство, наблюдение, группировка материала, вычисление обобщающих показателей и их анализ. Два последних этапа и есть задачи математической статистики.

Оценка точности производимых измерений имеет большое экономическое и прикладное значение. Оценку точности результатов каждого из проводимых измерений делает сам экспериментатор. Поэтому знание практических методов оценки погрешностей результатов измерений необходимо материаловедам, участвующим в многочисленных технологических процессах и научных исследованиях.

Рассмотрим основные практические методы статистической оценки погрешностей результатов измерений.

Различают генеральную и выборочную совокупность измерений. Под генеральной совокупностью подразумевают всё множество возможных значений измерений x_j . Для выбора совокупности число измерений $x_j N$ ограничено. Пусть имеется набор (выборка) экспериментальных данных $x_1, x_2, \dots, x_N$. Обработку этих данных производят обычно в следующем порядке:

1. Определяют точечные оценки математического ожидания $\bar{X}$ дисперсии S_x^2 и среднего квадратичного отклонения S_x :

$$\bar{X} = \frac{1}{N} \sum X_i \quad (1.1)$$

$$S_x^2 = \frac{1}{N-1} \sum (X_i - \bar{X})^2 \quad (1.2)$$

$$S_x = +\sqrt{S_x^2} \quad (1.3)$$

где $\bar{X}$ - среднее значение случайной величины или точечная оценка математического ожидания M ; S_x^2 - точечная оценка дисперсии σ^2 ; S_x - точечная оценка среднего квадратичного отклонения σ .

Точечные оценки стремятся к соответствующим параметрам распределения (M , σ^2 , σ) при числе измерений N стремящемся к бесконечности.

Оценка дисперсии характеризует однородность изменений. Чем выше оценка дисперсии, тем больше разброс значений.

Коэффициент вариации K_B характеризует изменчивость измерений. Чем выше коэффициент вариации, тем больше изменчивость измерений относительно средних значений. K_B оценивает также разброс значений при сравнении нескольких выборок:

$$K_B = \frac{S_x}{\bar{X}} \cdot 100\%. \quad (1.4)$$

2. Анализируют ряд в целях обнаружения грубых ошибок и промахов. Появление этих ошибок вполне вероятно, а наличие их ощутимо влияет на результат измерений. Однако прежде чем исключить то или иное измерение, нужно убедиться, что это действительно грубая ошибка, а не отклонение вследствие статистического разброса. Известно множество методов определения грубых ошибок статистического ряда. Наиболее простой метод - правило трех сигм: разброс случайных величин от среднего значения не должен превышать:

$$X_{\max, \min} = \bar{X} \pm 3\sigma = \bar{X} \pm 3S_x. \quad (1.5)$$

Более достоверными являются методы, базирующиеся на

использовании доверительного интервала. При наличии грубых ошибок критерии их появления вычисляются по формулам

$$\begin{aligned}\beta_1 &= (X_{\max} - \bar{X}) / S_x \sqrt{(N-1) / N}; \\ \beta_2 &= (\bar{X} - X_{\min}) / S_x \sqrt{(N-1) / N}.\end{aligned}\quad (1.6)$$

3. Находят точечные оценки очищенного ряда.

4. Определяют интервальные оценки математического ожидания и дисперсии с помощью доверительной вероятности.

Точечные оценки (особенно для выбора малого объема) могут значительно отличаться от истинного значения оцениваемого параметра. Для того, чтобы дать представление о точности и надежности такой оценки используют доверительные интервалы и доверительные вероятности.

Доверительной вероятностью (достоверностью) измерения называется вероятность того, что истинное значение измеряемой величины попадает в данный доверительный интервал $a \leq X_g \ll b$.

Доверительная вероятность P_g описывается выражением:

$$P_g = P[a \leq X_g \ll b] = \frac{1}{2} \left[\frac{\varphi(b - \bar{X})}{S_x} - \frac{\varphi(a - \bar{X})}{S_x} \right], \quad (1.7)$$

где X_g - действительное (истинное) значение случайной величины; $\varphi(t)$ интегральная функция Лапласа, определяемая выражением:

$$\varphi(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-t^2/2} dt. \quad (1.8)$$

Значение $\alpha = (1 - P_g)$ называют уровнем значимости. Значения интегральной функции Лапласа приводятся в соответствующих таблицах.

Аргументом этой функции является отношение половины доверительного интервала μ к среднеквадратичному отклонению S_x , т.е.:

$$t = \mu \sqrt{N} / S_x; \mu = b - \bar{X}; \mu = -(a - \bar{X}). \quad (1.9)$$

Если на основе определенных данных установлена доверительная вероятность $P_g(0,90; 0,95)$, то устанавливается точность измерений (доверительный интервал 2μ) на основе соотношения $P_d = \Phi(\mu / S_x)$. Половина доверительного интервала равна:

$$\mu = \frac{S_x \cdot t}{\sqrt{N}}, \quad (1.10)$$

где t - аргумент функции Лапласа (при $N > 30$) или функции Стьюдента (при $N \leq 30$). Требуемую точность измерений можно определить для разных уровней доверительной вероятности, приняв значение t по таблице 1.1 при числе степеней свободы $f = N - 1$.

Зная величину доверительного интервала, можно получить интервальную оценку математического ожидания:

$$\bar{X} - \mu < M < \bar{X} + \mu \text{ или } X_d = \bar{X} \pm \mu. \quad (1.11)$$

Подставив значение μ из (10), получим:

$$\bar{X} - \frac{S_x \cdot t}{\sqrt{N}} < M < \bar{X} + \frac{S_x \cdot t}{\sqrt{N}} \text{ или } X_d = \bar{X} \pm \frac{S_x \cdot t}{\sqrt{N}}. \quad (1.12)$$

Следует иметь в виду, что определяемый указанным образом доверительный интервал характеризует абсолютную погрешность оценки исследуемой величины за счет случайных ошибок. Однако в эксперименте может присутствовать и систематическая

погрешность связанная с погрешностями средств измерений. При этом возможны три случая:

Если приборная погрешность Δ по сравнению со случайной ошибкой μ мала, то величиной Δ можно пренебречь.

Если приборная погрешность велика, то пренебрегают случайной ошибкой μ .

Если погрешность серии измерений соизмерима с погрешностью прибора, то границы доверительного интервала:

$$\mu = \sqrt{\frac{S_x^2 t^2}{\sqrt{N}} + \left[\frac{t(\infty)}{3} \right]^2}. \quad (1.13)$$

Формулой (1.13) следует пользоваться при $tS_x / \sqrt{N} \leq 3\Delta$. Если же $tS_x / \sqrt{N} > 3\Delta$ то доверительный интервал вычисляют по формуле (1.10). Для дисперсии σ^2 интервальная оценка находится с помощью χ^2 - распределения по формуле

$$fS_x^2 / \chi_{1f;1-\alpha/2}^2 < \sigma^2 < fS_x^2 / \chi_{2f;\alpha/2}^2, \quad (1.14)$$

где $\chi_{1f;1-\alpha/2}^2$ - значение критерия χ^2 при числе степеней свободы $f=N-1$ и уровне значимости $1-\alpha/2$; $\chi_{2f;\alpha/2}^2$ - значение критерия χ^2 при числе степеней свободы $f=N-1$ и уровне значимости $\alpha/2$. Значение χ^2 приводится в таблице 1.1.

Интервальная оценка S_x находится по формуле:

$$\sqrt{fS_x^2 / x_1^2} \leq \sigma \leq \sqrt{fS_x^2 / x_2^2}. \quad (1.15)$$

5. Оценивают относительную погрешность результатов серии измерений при заданной доверительной вероятности:

$$\varepsilon = \frac{\mu}{\bar{X}} \cdot 100\%. \quad (1.16)$$

Таблица 1.1 – Процентные точки χ^2 - распределения

f / P_d	0,025	0,975
1	0,00098	5,024
2	0,0506	7,378
3	0,2158	9,348
4	0,4844	11,14
5	0,312	12,83
6	1,2373	14,45
7	1,690	16,01
8	2,180	17,53
9	2,700	19,02
10	3,247	20,48
11	3,816	21,92
12	4,404	23,34
13	5,009	24,74
14	5,629	26,12
15	6,262	27,49
16	6,908	28,85
17	7,564	30,19
18	8,231	31,53
19	8,907	32,85
20	9,591	34,17

В случае необходимости определяют минимальное количество измерений при заданной точности. Для проведения опытов с заданной точностью и достоверностью необходимо знать минимальное, но достаточное число измерений для данных условий. Задача сводится к установлению минимального объема выборки при заданных значениях доверительного интервала и доверительной вероятности. Зная точность измерений (или относительную погрешность) при $N_{\min} = N$ из (1.14), подставив выражение для μ , получим:

$$N_{\min} = \frac{S_x^2 \cdot t^2}{\varepsilon^2 (\bar{X})^2} \text{ или } N_{\min} = \frac{K_B^2 \cdot t^2}{\varepsilon^2}. \quad (1.17)$$

Расчеты по описанному выше алгоритму реализуются на ПЭВМ в табличном процессоре Excel пакета Microsoft Office 2007.

Технология выполнения

1. Получить выборку экспериментальных данных. Представить результаты эксперимента в виде таблицы.
2. Провести статистическую обработку данных по формулам (1.1-1.17).
3. Провести обработку результатов эксперимента на ПЭВМ, используя лист Microsoft Excel 2007 (или более новую версию).
4. Проанализировать полученные результаты и сделать выводы. Подробнее последовательность выполнения показана на конкретном примере.

Пример выполнения задания

Для полученной выборки из 18 ($N=18$) измерений провести статистическую обработку результатов эксперимента:

6	6	6	6	6	7	7	7	7	7	7	7	7	7	7	8	8	9
7	7	8	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2

1. Проанализируем ряд в целях обнаружения грубых ошибок и промахов. Точечная оценка математического ожидания вычисляется по формуле (1.1):

$$\bar{X} = 74,28.$$

С помощью формул (1.2, 1.3) вычисляют точечную оценку среднеквадратичного отклонения:

$$S_x = 6,36.$$

Пользуясь правилом трех сигм, по формуле (1.5), вычисляем допустимый разброс случайных величин:

$$X_{\max, \min} = 74,28 \pm (3 \cdot 6,36) = 93,36 \dots 55,2,$$

т.е. грубой ошибки в ряде нет. Однако значение 92 следует проверить с помощью критерия β . По формулам (1.6), (коэффициент β_1 используют для максимального значения в выборке):

$$\beta_1 = \frac{92 - 74,22}{6,2\sqrt{17/18}} = 2,87.$$

При доверительной вероятности $P_d = 0,99$ и $N = 18$ $\beta_{\max} = 2,96$ (таблица 1.2). Так как $2,87 < \beta_{\max}$, измерение 92 не является грубым промахом. В технических расчетах обычно принимают доверительную вероятность $P_g = 0,95$, то проверим соответствующее такой вероятности значение $\beta_{\max}$. $P_d = 0,95$, $N = 18$, то $\beta_{\max} = 2,62$ в этом случае $2,87 > \beta_{\max}$ и значение 92 следует исключить как грубых промах. Запишем очищенный ряд ($N=17$):

67	67	68	68	69	70	71	72	73	74	75	76	77	78	79	80	81
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

2. По формулам (1.1-1.3) найдем точечные оценки очищенного ряда:

$$\bar{X} = 73,24; S_x^2 = 22,19; S_x = 4,71.$$

По формуле (1.4) вычисляем коэффициент вариации:

$$K_v = \frac{4,71}{73,24} \cdot 100\% = 6,43\%.$$

3. Найдем интервальные оценки очищенного ряда. Поскольку $N < 30$, ряд следует отнести к малой выборке, и доверительный интервал вычисляется с применением коэффициента Стьюдента t .

При $P_d = 0,95$ и $f = 17 - 1 = 16$ коэффициент Стьюдента $t = 1,746$ (таблица 1.3).

Таблица 1.2 – Критерий появления грубых ошибок

N	При Р _д		
	0,90	0,95	0,99
1	1,41	1,41	1,41
2	1,64	1,69	1,72
3	1,79	1,87	1,96
4	1,89	2,00	2,13
5	1,97	2,09	2,26
6	2,04	2,017	2,37
7	2,10	2,24	2,46
8	2,15	2,29	2,54
9	2,19	2,34	2,61
10	2,23	2,39	2,66
11	2,26	2,43	2,71
12	2,30	2,46	2,76
13	2,33	2,49	2,80
14	2,35	2,52	2,84
15	2,38	2,55	2,87
16	2,40	2,58	2,90
17	2,43	2,60	2,93
18	2,45	2,62	2,96

Доверительный интервал вычисляется по формуле (1.10) при N=17:

$$\mu = \pm \left(\frac{4,71 \cdot 1,746}{\sqrt{17}} \right) = 1,99.$$

Относительная погрешность результатов измерений вычисляют по формуле (1.16):

$$\varepsilon = \frac{1,99}{73,24} \cdot 100\% = 2,72\%.$$

Запишем истинное значение измеряемой величины:

Таблица 1.3 – Процентные точки одностороннего распределения Стьюдента

f	Доверительная вероятность P_d				
	0,750	0,900	0,950	0,975	0,99
1	1,000	3,078	6,314	12,706	31,821
2	0,816	1,886	2,920	4,303	6,965
3	0,765	1,638	2,353	3,182	4,541
4	0,741	1,533	2,132	2,776	3,747
5	0,727	1,476	2,015	2,571	3,365
6	0,718	1,440	1,943	2,447	3,143
7	0,711	1,415	1,895	2,365	2,998
8	0,706	1,397	1,860	2,306	2,896
9	0,703	1,383	1,833	2,262	2,821
10	0,700	1,372	1,812	2,228	2,764
11	0,697	1,363	1,796	2,201	2,718
12	0,695	1,356	1,782	2,179	2,681
13	0,694	1,350	1,771	2,160	2,650
14	0,692	1,345	1,761	2,145	2,624
15	0,691	1,341	1,753	2,131	2,602
16	0,690	1,337	1,746	2,120	2,583
17	0,689	1,333	1,740	2,110	2,567
18	0,688	1,330	1,734	2,101	2,552
19	0,688	1,328	1,729	2,093	2,539
20	0,687	1,325	1,725	2,086	2,528

$$X_d = \bar{X} \pm \mu = 73,24 \pm 1,99.$$

Найдем интервальную оценку дисперсии σ^2 по формуле (1.14) при уровне значимости $\alpha = 1 - 0,95 = 0,05$:

$$12,5 \leq \sigma^2 \leq 49,87.$$

Найдем интервальную оценку среднеквадратичного отклонения по формуле (1.15):

$$\sqrt{12,5} \leq \sigma < \sqrt{49,87}$$

$$3,53 \leq \sigma < 7,06.$$

4. Пусть необходимая точность измерений составляет 5%. Определим минимальное количество измерений для достижения заданной точности. По формуле (1.17):

$$N_{\min} = \frac{6,43^2 \cdot 1,746^2}{5^2} = 5,045.$$

Округляем полученное значение $N_{\min}$ до большего целого, $N_{\min} = 6$. Для достижения заданной точности достаточно сделать 6 измерений.

*Составление программы для статистических расчетов в
табличном процессоре Excel пакета Microsoft Office*

В таблице 1.4 приводятся номера использованных формул и соответствующие им выражения в строке Excel. Не приводятся логические операции для удобства работы вычислений и приемы работы с массивами, т.к. их легко можно реализовать с помощью встроенной справки в Excel.

Таблица 1.4 – Расчет формул для статистической обработки в табличном процессоре Excel пакета Microsoft Office

№	Выражение
1	=ОКРУГЛ((СРЗНАЧ("Диапазон выборки"));2)
2	=ОКРУГЛ((ДИСП("Диапазон выборки"));2)
3	=ОКРУГЛ((СТАНДОТКЛОН("Диапазон выборки"));2)
4	=СТАНДОТКЛОН("Диапазон выборки")/СРЗНАЧ("Диапазон выборки")
5	Максимум: =ОКРУГЛ((СРЗНАЧ("Диапазон выборки")+(3*СТАНДОТКЛОН("Диапазон выборки")));2); Минимум: =ОКРУГЛ((СРЗНАЧ("Диапазон выборки")-3*СТАНДОТКЛОН("Диапазон выборки"));2)

Продолжение таблицы 1.4

6	Бэта 1: $= \text{ОКРУГЛ}(((\text{МАКС}(\text{"Диапазон выборки"}) - \text{СРЗНАЧ}(\text{"Диапазон выборки"})) / (\text{СТАНДОТКЛОН}(\text{"Диапазон выборки"}) * \text{КОРЕНЬ}(((\text{СЧЁТ}(\text{"Диапазон выборки"}) - 1) / \text{СЧЁТ}(\text{"Диапазон выборки"})))))); 2)$ Бэта 2: $= \text{ОКРУГЛ}(((\text{СРЗНАЧ}(\text{"Диапазон выборки"}) - \text{МИН}(\text{"Диапазон выборки"})) / (\text{СТАНДОТКЛОН}(\text{"Диапазон выборки"}) * \text{КОРЕНЬ}(((\text{СЧЁТ}(\text{"Диапазон выборки"}) - 1) / \text{СЧЁТ}(\text{"Диапазон выборки"})))))); 2)$
10	$= \text{ОКРУГЛ}(((\text{СТАНДОТКЛОН}_{\text{ф5:Ш}} * (\text{СТЮДРАСПОБР}((2 * (1 - \text{"Доверительная вероятность"})) / \text{СЧЁТ}(\text{"Диапазон выборки"}) - 1))) / \text{КОРЕНЬ}(\text{СЧЁТ}(\text{"Диапазон выборки"}))))); 2)$
14	Меньшее: $= \text{ОКРУГЛ}((\text{СЧЁТ}(\text{"Диапазон выборки"}) * \text{ДИСП}(\text{"Диапазон выборки"})) / (\text{ИНДЕКС}(\text{"Таблица 3'!С3:D22; СЧЁТ}(\text{"Диапазон выборки"}); 2))); 2)$ Большее: $= \text{ОКРУГЛ}((\text{СЧЁТ}(\text{"Диапазон выборки"}) * \text{ДИСП}(\text{"Диапазон выборки"})) / (\text{ИНДЕКС}(\text{"Таблица 3'!С3:D22; СЧЁТ}(\text{"Диапазон выборки"}); 1))); 2)$
15	Меньшее: $= \text{ОКРУГЛ}((\text{КОРЕНЬ}((\text{СЧЁТ}(\text{"Диапазон выборки"}) * \text{ДИСП}(\text{"Диапазон выборки"})) / (\text{ИНДЕКС}(\text{"Таблица 3'!С3:D22; СЧЁТ}(\text{"Диапазон выборки"}); 2))))); 2)$ Большее: $= \text{ОКРУГЛ}((\text{КОРЕНЬ}((\text{СЧЁТ}(\text{"Диапазон выборки"}) * \text{ДИСП}(\text{"Диапазон выборки"})) / (\text{ИНДЕКС}(\text{"Таблица 3'!С3:D22; СЧЁТ}(\text{"Диапазон выборки"}); 1))))); 2)$
16	$= (\text{СТАНДОТКЛОН}(\text{"Диапазон выборки"}) * (\text{СТЮДРАСПОБР}((2 * (1 - \text{"Доверительная вероятность"})) / \text{СЧЁТ}(\text{"Диапазон выборки"}) - 1))) / \text{КОРЕНЬ}(\text{СЧЁТ}(\text{"Диапазон выборки"})) / \text{СРЗНАЧ}(\text{"Диапазон выборки"})$

Продолжение таблицы 1.4

17	=ОКРУГЛВВЕРХ(((СТЕПЕНЬ((СТАНДОТКЛОН("Диапазон выборки")* 100/СРЗНАЧ("Диапазон выборки")));2)*СТЕПЕНЬ((СТЮДРАСПОБР((2*(1- "Доверительная вероятность")));СЧЁТ("Диапазон выборки")-1));2))/СТЕПЕНЬ("Заданная точность";2));0)
Примечание: Формат ячеек - Процентный устанавливается в формуле 4 и 16, все остальные ячейки - Общий.	

Задание

Получить статистический ряд и провести экспериментальные статистические процедуры. Статистический ряд можно получить с помощью пакета MathCad (приложение 1).

Содержание отчета

1. Цель занятия.
2. Краткие теоретические положения.
3. Технология выполнения работы и полученные результаты: задание; результаты измерений; формулы для получения оценок выборки и результаты расчетов; результаты вычислений на ПЭВМ.
4. Выводы.

Контрольные вопросы

1. В чем разница между понятиями «дисперсия» и «оценка дисперсии»?
2. Какие оценки параметров распределения вы знаете?
3. Что называют коэффициентов вариации?
4. Что такое точность измерений, достоверность измерений?
5. В каких случаях при определении доверительного интервала следует пользоваться интегральной функцией Лапласа?

Практическое занятие №2

Статистические характеристики случайных величин

Цель занятия: научиться вычислять оценки основных числовых характеристик по экспериментальным данным.

Теоретические сведения

Математическим ожиданием дискретной случайной величины X , принимающей конечное число значений x_i с вероятностями p_i , называется сумма:

$$M(X) = x_1 \cdot p_1 + x_2 \cdot p_2 + x_3 \cdot p_3 + \dots + x_n p_n.$$

Свойства математического ожидания:

1. $M(c \cdot X) = c \cdot M(X)$, $c \in R$,
2. $M(X + Y) = M(X) + M(Y)$, $X, Y \in E$,
3. $M(X \cdot Y) = M(X) \cdot M(Y)$, для независимых случайных величин X и Y .

Дисперсией случайной величины X называется число:

$$D(X) = M\{[X - M(X)]^2\} = M(X^2) - [M(X)]^2.$$

Свойства дисперсии:

1. $D(c \cdot X) = c^2 \cdot D(X)$, $c \in R$,
2. $D(X + Y) = D(X) + D(Y)$, для независимых случайных величин X и Y .

Среднеквадратичное отклонение определяется по формуле:

$$\sigma = \sqrt{D(X)}.$$

Статистическая функция СЧЕТ используется для определения числа значений (рисунок 2.1).

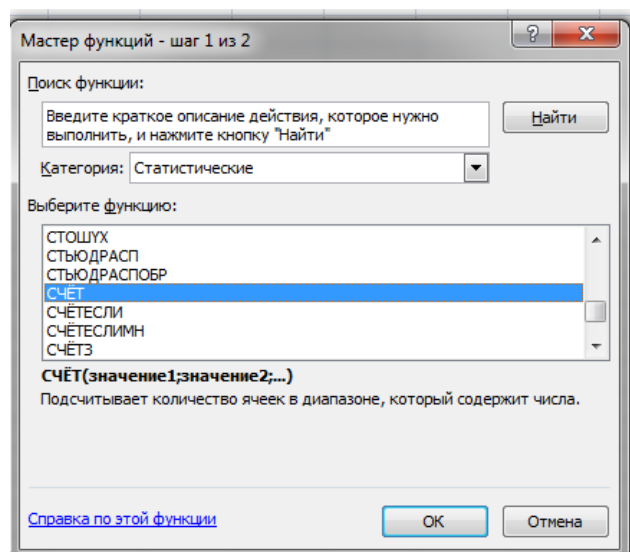


Рисунок 2.1 – Статистическая функция СЧЁТ

Диапазон ячеек указывается адресами первой и последней ячейки данных, записанными через двоеточие, например B6:B15 (рисунок 2.2).

Функция СРЗНАЧ рассчитывает среднее значение выборки.

Функция СТАНДОТКЛОН - стандартное отклонение выборки.

Аргументами этих функций служит все тот же диапазон ячеек.

Статистическая функция СТЬЮДРАСПОБР используется для нахождения коэффициента Стьюдента. Этот коэффициент зависит от вероятности ошибки (при обычно задаваемой надежности 95% вероятность ошибки составляет 5%) и от числа степеней свободы $n-1$. Также это значение можно найти по таблице критических значений t-критерия (приложение 2).

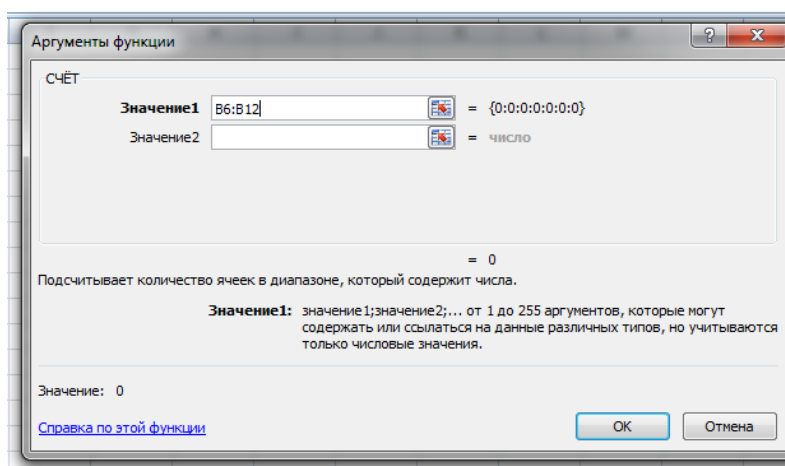


Рисунок 2.2 – Диапазон ячеек

Для нахождения доверительного интервала используется обычная формула умножения «=СТЮДРАСПОБР* $X_{\text{ср}}$ ».

Технология выполнения

В результате эксперимента были получены значения переменной X:

i	1	2	3	4	5	6	7	8	9	10
X	14,84	14,80	14,84	14,81	14,63	14,81	14,80	14,85	14,84	14,80

Данные введем в столбец В (рисунок 2.3). В столбцах D и E - подсказки характеристик, которые мы будем рассчитывать. Используя приведенные выше функции, в столбец F поместим результаты. По найденному значению ячейки F11 окончательный результат доверительного интервала можно записать так: с 95%-ной надежностью $X = 14,81 \pm 0,046$. В заключение вычислим относительную ошибку определения доверительного интервала: $S = \text{ДИ} / X_{\text{ср}}$ (формула: «=F11/F7»). Значение относительной ошибки обычно выражают в процентах, в нашем случае 0,3 %.

	A	B	C	D	E	F
1						
2						
3	Данные эксперимента				Обработка	
4						
5	№	X				
6	1	14,85		Число значений n	СЧЁТ	10
7	2	14,8		Среднее значение X _{ср}	СРЗНАЧ	14,803
8	3	14,84		Станд. Отклонение S	СТАНДОТКЛОН	0,0643
9	4	14,81		Ст. откл. Среднего S _{ср}	\$/КОРЕНЬ(n)	0,0203
10	5	14,63		К. Стюд (5%, n-1) t	СТЮДРАСПОБР	2,2622
11	6	14,81		Доверит. Интервал ДИ	t*S _{ср}	0,046
12	7	14,8		Относит. ошибка s	ДИ/X _{ср}	0,0031
13	8	14,85				
14	9	14,84				
15	10	14,8				

Рисунок 2.3 – Расчет данных

Задание

Получить статистический ряд и определить основные характеристики случайных величин:

- число значений;

- среднее значение выборки;
- стандартное отклонение выборки;
- коэффициент Стьюдента;
- доверительный интервал;
- относительную ошибку.

Статистический ряд можно получить с помощью пакета MathCad (приложение 1).

Содержание отчета

1. Цель занятия.
2. Краткие теоретические положения.
3. Технология выполнения работы и полученные результаты: задание; результаты измерений; формулы для получения оценок выборки и результаты расчетов; результаты вычислений на ПЭВМ.
4. Выводы.

Контрольные вопросы

1. Что такое математическое ожидание?
2. Перечислите свойства математического ожидания.
3. Что такое дисперсия?
4. Перечислите свойства дисперсии?
5. Что такое среднеквадратичное отклонение?

Практическое занятие №2.1

Оценка параметров и определение закона распределения

Цель занятия: приобрести навыки первичной статистической обработки данных, проверки гипотезы о виде распределения случайной величины с помощью критерия согласия Пирсона.

Теоретические сведения

Оценку параметров и определение закона распределения рассмотрим на примере.

Исследуется случайная величина X - число правонарушений в течение одних суток в некотором городе N . Получены данные за первые 150 суток года.

3	5	4	4	5	8	2	3	1	6	6	1	2	5	5
4	4	4	3	4	5	5	2	2	3	4	3	2	4	4
8	10	1	4	3	3	2	5	7	5	3	6	7	5	6
1	4	6	4	5	4	5	7	6	5	3	5	5	8	7
7	5	5	4	5	3	3	6	3	5	2	2	2	6	2
5	6	8	4	4	8	3	6	4	4	5	5	7	5	5
3	5	4	5	5	4	7	6	9	3	3	5	6	6	3
4	5	2	6	7	5	5	4	2	5	4	2	6	2	7
5	5	8	5	3	5	2	5	3	7	4	6	3	6	0
4	4	4	5	2	7	7	3	1	1	3	6	5	7	6

Данная задача решается с помощью статистических процедур Анализа данных и статистических функций библиотеки встроенных функций MS Excel. Приведем алгоритм решения задачи.

1. Ввод данных. В диапазон ячеек A1:AN ввести выборочные значения $X_1, X_2, \dots, X_N$.

2. Построение вариационного ряда. Скопировать содержимое ячеек A1:AN в ячейки B1:BN. Упорядочить выборочные значения, используя кнопку сортировки по возрастанию.

3. Построение статистического ряда выборки. В ячейки C1:CK ввести k различных выборочных значений. В меню Данные выделить строку Анализ данных, выделить процедуру

Гистограмма. В поле Входной интервал диалогового окна Гистограмма ввести ссылку на диапазон A1:AN. В поле Интервал карманов ввести ссылку на диапазон C1:CN. Активизировать поле Выходной интервал и ввести в это поле ссылку - левая верхняя ячейка, в которую будет введена таблица результатов решений. Установить флажок Вывод графика. Составить таблицу 2.1.1 статистического ряда по следующему образцу:

Таблица 2.1.1 – Таблица результатов решений

x_i различные выборочные значения	n_i частота выборочного значения x_i	$\frac{n_i}{N}$ относительная частота выборочного значения x_i	$\frac{n_i^*}{N}$ накопленная относительная частота
--	---	--	--

Первые столбцы заполнить копированием. Относительные и накопленные частоты вычислить с использованием формул.

4. Построение полигонов относительных и накопленных относительных частот.

Скопировать первый и третий столбцы таблицы 2.1.1. Выделить их, используя меню Вставка, применить к выделенным числам средства диаграммы Точечная. Полученный график есть полигон относительных частот. Если же действия проделать с первым и четвертым столбцами таблицы 2.1.1, то получим полигон накопленных частот – сглаженный график эмпирической функции распределения.

5. Определение выборочных характеристик. В меню Данные выделить подменю Анализ данных, выделить процедуру Описательная статистика, в поле ввода Входной интервал ввести ссылку на диапазон ячеек, содержащий статистические данные A1:AN Установить флажок Итоговая статистика. Активизировать поле Выходной интервал, ввести в это поле ссылку - левая верхняя ячейка, в которую будет введена таблица результатов решений.

6. Проверка гипотезы о виде распределения случайной величины с помощью критерия согласия Пирсона. Заполнить таблицу 2.1.2.

Таблица 2.1.2 – Таблица результатов решений

x_i различные выборочные значения	n_i частота выборочного значения x_i	p_i Теоретическая вероятность выборочного значения x_i	$n_i = N \cdot p_i$ Теоретическая частота выборочного значения x_i	$\frac{(n'_i - n_i)^2}{n'_i}$
--	--	--	--	-------------------------------

Первые столбцы заполнить копированием, а оставшиеся - вычисленными по формулам значениями.

Если проверяется гипотеза о распределении Пуассона, то теоретические вероятности p_i вычислить с помощью функции ПУАССОН ($x_i, \bar{x}, 0$). Здесь $\bar{x}$ — выборочное среднее, оно определяется в пункте 5, 0 - параметр, показывающий, что вычисляется вероятность того, что случайная величина, распределенная по закону Пуассона, принимает значение x_i .

Если проверяется гипотеза о биномиальном распределении случайной величины, то теоретические вероятности p_i вычислить с помощью функции БИНОМРАСП ($x_i, N, P, 0$), при этом вероятность успеха P в одном испытании определить по формуле $P = \frac{\bar{x}}{N}$, где $\bar{x}$ - выборочное среднее.

В случае других распределений воспользоваться справкой о статистических функциях библиотеки встроенных функций MS Excel.

Значение

$$\chi^2 = \sum_{i=1}^k \frac{(n'_i - n_i)^2}{n'_i}$$

является наблюдаемым значением случайной величины χ^2 . Число степеней свободы этой случайной величины равно $r = k - 2$, при проверке гипотезы о распределении Пуассона и $r = k - 3$, если проверяется гипотеза о биномиальном распределении. Критическое значение случайной величины χ^2 определить с помощью функции

$\chi^2_{кр} = \text{ХИ2ОБР}(\alpha, r)$, где α - уровень значимости. Полученное наблюдаемое значение χ^2 сравнить с $\chi^2_{кр}$. Если $\chi^2 < \chi^2_{кр}$, то гипотеза о виде распределения принимается при уровне значимости α . Если $\chi^2 > \chi^2_{кр}$, то гипотеза отвергается с уровнем значимости α .

Технология выполнения

По предложенному в теоретических сведениях алгоритму проведем первичную статистическую обработку данных. Согласно пункту 3 алгоритма находим $k=11$ различных выборочных значений

0 1 2 3 4 5 6 7 8 9 10

С помощью пакета Анализ данных получаем статистический ряд выборки и его графическое представление (рисунок 2.1.1).



Рисунок 2.1.1 – Статистический ряд выборки

Построенная гистограмма позволяет сделать предложение о виде распределения случайной величины X . В результате заполнения таблицы 2.1.1 получим таблицу 2.1.3, в третьем столбце которой, представлены относительные, а в четвертом — накопленные относительные частоты выборочных значений.

Таблица 2.1.3 - Таблица результатов решений

x_i	n_i	$\frac{n_i}{N}$	$\frac{n_i^*}{N}$
0	1	0,006667	0,006667
1	6	0,04	0,046667
2	16	0,106667	0,153333
3	22	0,146667	0,3
4	27	0,18	0,48
5	38	0,253333	0,733333
6	19	0,126667	0,86
7	13	0,086667	0,946667
8	6	0,04	0,986667
9	1	0,006667	0,993333
10	1	0,006667	1

Согласно пункту 4 алгоритма получим полигон относительных и накопленных частот (рисунки 2.1.2, 2.1.3).

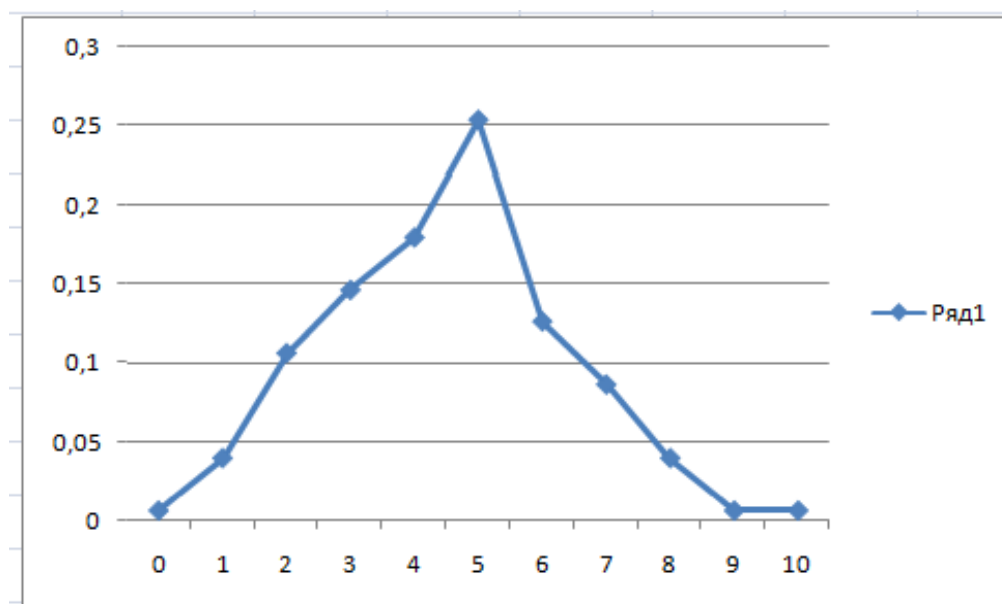


Рисунок 2.1.2 – Полигон относительных частот

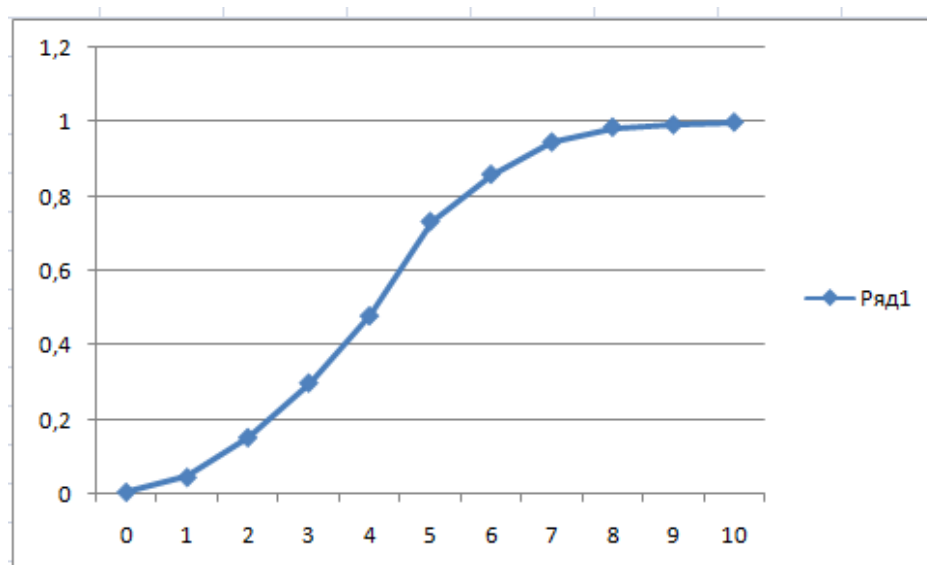


Рисунок 2.1.3 – Полигон накопленных частот

Согласно пункту 5 алгоритма получаем выборочные характеристики (таблица 2.1.4).

Таблица 2.1.4 - Выборочные характеристики

Среднее	4,493333
Стандартная ошибка	0,150687
Медиана	5
Мода	5
Стандартное отклонение	1,845534
Дисперсия выборки	3,405996
Эксцесс	-0,12004
Асимметричность	0,130829
Интервал	10
Минимум	0
Максимум	10
Сумма	674
Счет	150

Проверим гипотезу о распределении случайной величины по закону Пуассона. В качестве точечной оценки параметра распределения выбираем выборочное среднее $\lambda = \bar{x} = 4,49$. В результате заполнения таблицы 2.1.2 получим таблицу 2.1.5. наблюдаемое значение случайной величины $\chi^2 = 9,261528$. Оно

получено суммированием чисел последнего столбца таблицы 2.1.4. Критическое значение $\chi^2_{\text{кр}} = \text{ХИ25ОБР}(\alpha, r) = 16,91898$, где $\alpha = 0,05, r = k - 2 = 11 - 2 = 9$. Так как $\chi^2 < \chi^2_{\text{кр}}$, то гипотеза о распределении по закону Пуассона при уровне значимости $\alpha = 0,05$ не противоречит опытным данным.

Таблица 2.1.5 - Таблица результатов решений

x_i	n_i	p_i	p'_i	χ^2
0	1	0,011183	1,677501	0,273626
1	6	0,05025	7,537566	0,313644
2	16	0,112896	16,93439	0,051557
3	22	0,169093	25,36393	0,446146
4	27	0,189948	28,49213	0,078142
5	38	0,170699	25,6049	6,00035
6	19	0,127835	19,17521	0,001601
7	13	0,082058	12,30865	0,038831
8	6	0,046089	6,913355	0,120667
9	1	0,02301	3,451554	1,741278
10	1	0,010339	1,550897	0,195685
				9,261528

Задание

Получить статистический ряд и провести первичную статистическую обработку данных, проверить гипотезу о виде распределения случайной величины с помощью критерия согласия Пирсона.

Статистический ряд можно получить с помощью пакета MathCad (приложение 1).

Содержание отчета

1. Цель занятия.
2. Краткие теоретические положения.

3. Технология выполнения работы и полученные результаты: задание; результаты измерений; формулы для получения оценок выборки и результаты расчетов; результаты вычислений на ПЭВМ.

4. Выводы.

Контрольные вопросы

1. Приведите алгоритм построения вариационного ряда.
2. Приведите алгоритм построения статистического ряда выборки.
3. Приведите алгоритм построения полигонов относительных и накопительных относительных частот.
4. Что такое распределение Пуассона?
5. Что такое биномиальное распределение?
6. Перечислите алгоритм проверки гипотезы о виде распределения величины с помощью критерия согласия Пирсона.
7. Как находится теоретическая частота выборочного значения x_i ?

Практическая работа №3

Вычисление коэффициентов уравнения линейной регрессии методом наименьших квадратов

Цель занятия: научиться вычислять коэффициенты уравнения линейной регрессии по пространственной выборке.

Теоретические сведения

Выбрав вид функции регрессии, т.е. вид рассматриваемой модели зависимости Y от X , например, линейную модель $y_x = a + bx$, необходимо определить конкретные значения коэффициентов модели.

При различных значениях a и b можно построить бесконечное число зависимостей вида $y_x = a + bx$, т.е. на координатной плоскости имеется бесконечное количество прямых, нам же необходима такая зависимость, которая соответствует наблюдаемым значениям наилучшим образом. Таким образом, задача сводится к подбору наилучших коэффициентов.

Линейную функцию $a+bx$ ищем, исходя лишь из некоторого количества имеющихся наблюдений. Для нахождения функции с наилучшим соответствием наблюдаемым значениям используем метод наименьших квадратов.

Обозначим: Y_i - значение, вычисленное по уравнению $Y_i = a + bx_i$; y_i - измеренное значение, $\varepsilon_i = y_i - Y_i$ - разность между измеренными и вычисленными по уравнению значениями, $\varepsilon_i = y_i - a - bx_i$.

В методе наименьших квадратов требуется, чтобы ε_i , разность между измеренными y_i и вычисленными по уравнению значениями Y_i , была минимальной. Следовательно, находим коэффициенты a и b так, чтобы сумма квадратов отклонений наблюдаемых значений от значений на прямой линии регрессии оказалась наименьшей:

$$S = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - Y_i)^2 = \sum_{i=1}^n (y_i - a - bx_i)^2 \rightarrow \min. \quad (3.1)$$

Исследуя на экстремум эту функцию аргументов a и b с помощью производных, можно доказать, что функция принимает минимальное значение, если коэффициенты a и b являются решениями системы:

$$\begin{cases} a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \\ na + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \end{cases} \quad (3.2)$$

Если разделить обе части нормальных уравнений на n , то получим:

$$\begin{cases} a \frac{\sum_{i=1}^n x_i}{n} + b \frac{\sum_{i=1}^n x_i^2}{n} = \frac{\sum_{i=1}^n x_i y_i}{n} \\ a + b \frac{\sum_{i=1}^n x_i}{n} = \frac{\sum_{i=1}^n y_i}{n} \end{cases}$$

Учитывая, что

$$\frac{\sum_{i=1}^n x_i}{n} = \bar{x}, \quad \frac{\sum_{i=1}^n y_i}{n} = \bar{y}, \quad \frac{\sum_{i=1}^n x_i^2}{n} = \bar{x}^2, \quad \frac{\sum_{i=1}^n x_i y_i}{n} = \overline{xy}. \quad (3.3)$$

Получим

$$\begin{cases} a\bar{x} + b\bar{x}^2 = \overline{xy} \\ a + b\bar{x} = \bar{y} \end{cases},$$

отсюда $a = \bar{y} - b\bar{x}$, подставляя значение a в первое уравнение, получим:

$$\begin{aligned}\bar{x}(\bar{y} - b\bar{x}) + b\bar{x}^2 &= \overline{xy} \Rightarrow \bar{x}\bar{y} - b\bar{x}^2 + b\bar{x}^2 = \overline{xy} \Rightarrow b(\bar{x}^2 - \bar{x}^2) = \overline{xy} - \bar{x}\bar{y}, \\ b &= \frac{\overline{xy} - \bar{x}\bar{y}}{\bar{x}^2 - \bar{x}^2} = \frac{\text{cov}(x, y)}{S_x^2},\end{aligned}\quad (3.4)$$

где $\text{cov}(x, y) = \overline{xy} - \bar{x}\bar{y}$ - выборочная ковариация; S_x^2 - выборочное значение дисперсии величины x , определяемой по формуле

$$S_x^2 = \bar{x}^2 - \bar{x}^2. \quad (3.5)$$

$$a = \frac{\bar{x}^2\bar{y} - \bar{x}\overline{xy}}{\bar{x}^2 - \bar{x}^2}. \quad (3.6)$$

При этом b называют коэффициентом регрессии; a называют свободным членом уравнения регрессии и вычисляют по формуле:

$$a = \bar{y} - b\bar{x}. \quad (3.7)$$

Полученная прямая является оценкой для теоретической линии регрессии. Имеем:

$$Y = a + bx = \bar{y} - b\bar{x} + bx = \bar{y} + b(x - \bar{x}).$$

Итак, $Y = \bar{y} + b(x - \bar{x})$ является уравнением линейной регрессии.

Регрессия может быть прямой ($b > 0$) и обратной ($b < 0$). Прямая регрессия означает, что при росте одного параметра, значения другого параметра тоже увеличиваются. А обратная, что при росте одного параметра, значения другого параметра уменьшаются.

Технология выполнения

Для определения зависимости между сменной добычей угля на одного рабочего (переменная Y , измеряемая в тоннах) и мощностью угольного пласта (переменная X , измеряемая в метрах) на 10 шахтах были проведены исследования, результаты которых представлены в таблице 3.1.

Таблица 3.1 – Результаты исследования

i	1	2	3	4	5	6	7	8	9	10
X	8	11	12	9	8	8	9	9	8	12
Y	5	10	10	7	5	6	6	5	6	8

Вычислим эти коэффициенты, используя табличный процессор Excel. На рисунке 3.1 показан фрагмент документа Excel, в котором: а) размещены выборочные данные таблицы 1; б) запрограммировано вычисление коэффициентов системы

$$\begin{cases} b_0 + b_1 \cdot \bar{x} = \bar{y} \\ b \cdot \bar{x} + b_1 \cdot \bar{x}^2 = \overline{xy} \end{cases};$$

в) запрограммировано вычисление b_0 , b_1 по формулам

$$b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\bar{x}^2 - \bar{x}^2} = \frac{m_{XY}}{S_X^2};$$

$$b_0 = \bar{y} - b_1 \cdot \bar{x}$$

соответственно.

	A	B	C	D	E	F	G	H
1		Исходные данные			=B3^2		=B3*C3	
2		x_i	y_i	x_i^2	$x_i \cdot y_i$			
3		8	5	64	40	=(E13-B13*C13)/(D13-B13^2)		
4		11	10	121	110			
5		12	10	144	120	b_1	1,016	
6		9	7	81	63	b_0	-2,75	
7		8	5	64	40			
8		8	6	64	48		=C13-G5*B13	
9		9	6	81	54			
10		9	5	81	45			
11		8	6	64	48			
12		12	8	144	96			
13	Средние значения	9,4	6,8	90,8	66,4			
14								
15	=СРЗНАЧ(B3:B12)				=СРЗНАЧ(E3:E12)			

Рисунок 3.1 - Вычисление коэффициентов линейной регрессии

Заметим, что для вычисления средних значений используется функция Excel СРЗНАЧ(диапазон ячеек).

Вычислим значения, используя данные таблицы 3.2.

$$\bar{x} = 9,4; \bar{y} = 6,8; \overline{x^2} = 90,8; \overline{xy} = 66,4.$$

Таблица 3.2 – Расчет результатов исследования

№ п/п	x	y	xy	x^2	y^2	$(y - \bar{y})$	$(y - \bar{y})^2$	$\hat{y}_i$	$(y - \hat{y})$	$(y - \hat{y})^2$
1	8	5	40	64	25	-1,8	3,24	5,38	0,38	0,1429
2	8	5	40	64	25	-1,8	3,24	5,38	0,38	0,1429
3	8	6	48	64	36	-0,8	0,64	5,38	-0,62	0,3869
4	8	6	48	64	36	-0,8	0,64	5,38	-0,62	0,3869
5	9	7	63	81	49	0,2	0,04	6,39	-0,61	0,3672
6	9	6	54	81	36	-0,8	0,64	6,39	0,39	0,1552
7	9	5	45	81	25	-1,8	3,24	6,39	1,39	1,9432
8	11	10	110	121	100	3,2	10,24	8,43	-1,57	2,4775
9	12	10	120	144	100	3,2	10,24	9,44	-0,56	0,3114
10	12	8	96	144	64	1,2	1,44	9,44	1,44	2,0794
Сумма	94	68	664	908	496	0	33,6	68	0	8,3934
Среднее	9,4	6,8	66,4	90,8	49,6					

Коэффициент a , b вычислим по формулам (3.7) и (3.4) соответственно:

$$a = -2,75; b = 1,02.$$

а само уравнение регрессии $\hat{y} = f(x) = a + bx$ примет вид:

$$\hat{y} = f(x) = -2,75 + 1,02x. \quad (3.8)$$

Задание

Получить статистический ряд из 15 значений и вычисление коэффициентов линейной регрессии.

Статистический ряд можно получить с помощью пакета MathCad (приложение 1).

Содержание отчета

1. Цель занятия.
2. Краткие теоретические положения.
3. Технология выполнения работы и полученные результаты: задание; результаты измерений; формулы для получения оценок выборки и результаты расчетов.
4. Выводы.

Контрольные вопросы

1. Приведите формулу линейной регрессии.
2. Что такое регрессия?
3. На каком методе основан классический подход к оцениванию параметров линейной регрессии?
4. Что такое прямая регрессия?
5. Что такое обратная регрессия?

Практическое занятие №3.1

Вычисление выборочного коэффициента корреляции

Цель занятия: приобрести навыки вычисления выборочного коэффициента корреляции.

Теоретические сведения

Коэффициент корреляции вводится как величина, которая служит мерой тесноты (силы) линейной корреляционной зависимости признаков.

Рассмотрим сначала эксперименты, в которых одновременно измеряются только две величины (обозначим их Y и X). Пусть проведено n независимых совместных измерений признаков Y и X , в результате которых получено n пар чисел:

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_i, y_i), (x_n, y_n). \quad (3.1.1)$$

Такую совокупность пар чисел можно рассматривать как случайную выборку из генеральной совокупности всех возможных значений пары (Y, X) . Поэтому получаемое по этим данным значение коэффициента корреляции называют выборочным.

Выборочный коэффициент корреляции определяется равенством

$$r_B = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n \sigma_x \sigma_y}, \quad (3.1.2)$$

где y_i – экспериментальное значение величины Y , соответствующее x_i ; n – объем выборки; $\bar{x}, \bar{y}$ – выборочные средние признаков X и Y ; σ_x, σ_y – выборочные средние квадратичные отклонения признаков X и Y .

Если вместо выборочных использовать исправленные средние квадратичные отклонения S_x, S_y , то вычисление выборочного коэффициента корреляции производится по формуле

$$r_B = \frac{\sum_i^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{(n-1)S_x S_y}. \quad (3.1.3)$$

Почему введенное таким образом выражение для r_B способно служить мерой тесноты корреляционной связи? Дело в том, что выражение для r_B является суммой произведений так называемых нормированных отклонений. Нормированное отклонение $t(x_i)$ варианты x_i есть разность между значением x_i варианты и ее средним значением, отнесенная к среднему квадратичному отклонению этой варианты:

$$t(x_i) = \frac{x_i - \bar{x}}{\sigma_x}.$$

Аналогично

$$t(y_i) = \frac{y_i - \bar{y}}{\sigma_y}$$

есть нормированное отклонение варианты y_i . Используя нормированные отклонения, выражение для r_B можно представить

в виде $r_B = \frac{1}{n} \sum_i^n t(x_i) \cdot t(y_i)$. При коррелированности (т.е. наличии

сопряженности изменения) признаков X и Y величины $t(x_i)$ и $t(y_i)$ принимают положительные и отрицательные значения некоторым регулярным образом, поэтому произведения $t(x_i) \cdot t(y_i)$ входят в $\sum t(x_i) \cdot t(y_i)$ большей частью с одинаковым знаком, что обеспечивает отличие r_B от нуля. Кроме того, при коррелированности признаков X и Y и модули величин $t(x_i)$ и $t(y_i)$ меняются согласованно: например, в случае линейной корреляции большим по модулю значениям $t(x_i)$ соответствуют,

как правило, большие по модулю значения $t(y_i)$, что также приводит к отличию r_B от нуля. Наоборот, при отсутствии корреляции знаки величин $t(x_i)$ и $t(y_i)$ будут меняться чисто случайным образом, из-за чего число положительных слагаемых в сумме $\sum t(x_i) \cdot t(y_i)$ будет примерно равно числу отрицательных, что приведет к их взаимному сокращению и равенству $r_B = 0$.

Существует много различных рабочих формул для вычисления r_B прямым способом, т.е. при непосредственном использовании полученных в результате измерений значений x_i и y_i . Здесь приведены лишь наиболее употребляемые из них. Выражение

$$r_B = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i^n (x_i - \bar{x})^2 \cdot \sum_i^n (y_i - \bar{y})^2}}$$

получается из (3.1.2) заменой σ_x и σ_y выражениями $\sqrt{\frac{\sum_i^n (x_i - \bar{x})^2}{n}}$

и $\sqrt{\frac{\sum_i^n (y_i - \bar{y})^2}{n}}$ в соответствии с их определением. В этом выражении используются только отклонения вариант от средних.

Еще одна формула для вычисления r_B получается в результате преобразования числителя в выражении (3.1.3):

$$\begin{aligned}
\sum_i^n (x_i - \bar{x}) \cdot (y_i - \bar{y}) &= \sum_i^n x_i \cdot y_i - \bar{x} \cdot \sum_i^n y_i - \bar{y} \cdot \sum_i^n x_i + n \cdot \bar{x} \cdot \bar{y} = \\
&= \sum_i^n x_i y_i - n \bar{x} \frac{\sum_i^n y_i}{n} - n \bar{y} \frac{\sum_i^n x_i}{n} + n \cdot \bar{x} \cdot \bar{y} = \\
&= \sum_i^n x_i y_i - n \cdot \bar{x} \cdot \bar{y} .
\end{aligned}$$

Отсюда получаем

$$r_B = \frac{\sum_i^n x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{n \sigma_x \sigma_y} \quad (3.1.4)$$

или

$$r_B = \frac{\sum_i^n x_i y_i - n \cdot \bar{x} \cdot \bar{y}}{(n-1) s_x s_y} . \quad (3.1.5)$$

Формулы (3.1.2), (3.1.3) и (3.1.4) для вычисления выборочного коэффициента корреляции применимы и в случаях, когда данные измерений не могут быть сгруппированы, из-за того, что различные значения x_i и y_i величин X и Y наблюдаются по одному разу, и в тех случаях, когда данные могли бы быть сгруппированы, но решено группировку не делать.

Если данные n экспериментов по совместному измерению значений величин Y и X сгруппированы и представлены в виде корреляционной таблицы, то выборочный коэффициент корреляции удобно вычислять по формуле

$$r_B = \frac{\sum n_{xy} xy - n \bar{x} \cdot \bar{y}}{n \sigma_x \cdot \sigma_y} , \quad (3.1.6)$$

где x, y – варианты (наблюдавшиеся значения) признаков X и Y ; n_{xy} – частота (число появлений) наблюдавшейся пары вариантов (x, y) ; n – объем выборки ($n = \sum n_{xy}$); $\bar{x}, \bar{y}$ – выборочные средние признаков X и Y ; σ_x, σ_y – выборочные средние квадратичные отклонения признаков X и Y .

Если использовать исправленные средние квадратичные отклонения S_x, S_y то

$$r_b = \frac{\sum n_{xy} xy - n \bar{x} \cdot \bar{y}}{(n-1) S_x \cdot S_y}. \quad (3.1.7)$$

Важно отметить, что выборочный коэффициент корреляции r_b и угловой коэффициент k прямой линии регрессии связаны между собой соотношением

$$r_b = \frac{\sigma_x}{\sigma_y} \cdot k, \quad (3.1.8)$$

или

$$r_b = \frac{S_x}{S_y} \cdot k. \quad (3.1.9)$$

Это выражение не только дает еще один способ вычисления коэффициента корреляции, но и явно указывает на то, что r_b является мерой именно линейной корреляционной связи.

Выборочный коэффициент корреляции определяется соотношением

$$r_{XY} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{S_X \cdot S_Y}, \quad (3.1.10)$$

где

$$S_X = \sqrt{x^2 - \bar{x}^2}; S_Y = \sqrt{y^2 - \bar{y}^2}; \bar{y}^2 = \frac{1}{n} \sum_{i=1}^n y_i^2. \quad (3.1.11)$$

Технология выполнения

Используя вычисления практической работы №3, данные таблицы 3.2 и формулы (3.1.10) и (3.1.11), получим (рисунок 3.1.1):

$$S_X = 1,56; \bar{y}^2 = 49,6; S_Y = 1,83; r_{XY} = \frac{66,4 - 9,4 \cdot 6,8}{1,56 \cdot 1,83} = \frac{2,48}{2,86} = 0,87.$$

	A	B	C	D	E	F	G
1	Исходные данные						
2	x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$		
3	8	5	64	25	40	=КОРЕНЬ(C13-A13^2)	
4	11	10	121	100	110		
5	12	10	144	100	120	S_X	1,562
6	9	7	81	49	63	S_Y	1,833
7	8	5	64	25	40		
8	8	6	64	36	48	r_{XY}	0,866
9	9	6	81	36	54		
10	9	5	81	25	45	=(E13-A13*B13)/(G5*G6)	
11	8	6	64	36	48		
12	12	8	144	64	96		
13	9,4	6,8	90,8	49,6	66,4	Средние значения	

Рисунок 3.1.1 – Вычисление выборочного коэффициента корреляции

Величина $r_{XY} = 0,87$ говорит о сильной положительной линейной связи.

Задание

Вычислить выборочный коэффициент корреляции по данным выборки из практического занятия №3.

Содержание отчета

1. Цель занятия.
2. Краткие теоретические положения.
3. Технология выполнения работы и полученные результаты: задание; результаты измерений; формулы для получения оценок выборки и результаты расчетов.
4. Выводы.

Контрольные вопросы

1. Что такое коэффициент корреляции?
2. Что такое выборочный коэффициент корреляции?
3. Каким равенством определяется коэффициент корреляции?
4. Что такое теснота корреляционной связи?
5. Что такое нормированное отклонение?

Практическое занятие №3.2

Проверка значимости уравнения линейной регрессии по критерию Фишера

Цель занятия: освоить навыки проверки значимости уравнения линейной регрессии по критерию Фишера.

Теоретические сведения

Уравнение парной регрессии значимо при уровне значимости α , если выполняется следующее неравенство:

$$F = \frac{Q_r (n - 2)}{Q_e} > F_{1-\alpha; 1; n-2}. \quad (3.2.1)$$

Величины Q_r , Q_e являются факторной и остаточной суммами квадратов соответственно:

$$Q_r = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2, \quad Q_e = \sum_{i=1}^n (\hat{y}_i - y_i)^2. \quad (3.2.2)$$

Величина $F_{1-\alpha; 1; n-2}$ - табличное значение F-распределения с числами степеней свободы $k_1 = 1$ и $k_2 = n - 2$ квантиля уровня $\gamma = 1 - \alpha$ (приложение 3). Эту вероятность можно также определить с помощью функции FРАСПОБР:

$$F_{1-\alpha; 1; n-2} = F \text{ РАСПОБР}(\alpha; 1; n - 2). \quad (3.2.3)$$

Значение F для линейной парной регрессии можно вычислить через коэффициент корреляции:

$$F = \frac{r_{XY}^2}{1 - r_{XY}^2} (n - 2). \quad (3.2.4)$$

Величина $r_{XY}^2 = R^2$ называется коэффициентом детерминации и характеризует долю дисперсии результативного признака y , объяснимую регрессией, в общей дисперсии результативного признака:

$$R^2 = 1 - \frac{Q_e}{Q}, \text{ где } Q = \sum_{i=1}^n (y_i - \bar{y})^2. \quad (3.2.5)$$

Соответственно величина $1 - r_{XY}^2$ характеризует долю дисперсии y , вызванную влиянием остальных, не учтенных в модели факторов.

Технология выполнения

Вычислим значения $Q_e, Q_r = Q - Q_e$ и критерий F по данным таблицы 3.2 (рисунок 3.2.1). в столбце $\hat{y}_i$ значения вычисляются по формуле (3.8). Итак, получены следующие значения:

$$Q_e = 8,39; Q_r = Q - Q_e = 33,6 - 8,39 = 25,21;$$

$$F = \frac{Q_r(n-2)}{Q_e} = \frac{25,21 \cdot 8}{8,39} = 24,03.$$

По формуле (3.2.3) при $k_1 = 1, k_2 = 8$ или по таблице (приложение 3) вычисляем квантиль $F_{0,95;1;8} = 5,32$. Неравенство (3.2.1) выполняется, т.е. $24,03 > 5,32$, и поэтому уравнение регрессии (3.8) значимо.

	A	B	C	D	E	F
1	b_0	-2,75				
2	b_1	1,016	$=(B5-\$B\$15)^2$		$=(D5-B5)^2$	
3	Исходные данные					
4	x_i	y_i	$(y_i - \bar{y})^2$	$\hat{y}_i$	$(\hat{y}_i - y_i)^2$	
5	8	5	3,240	5,378	0,143	
6	11	10	10,240	8,426	2,477	
7	12	10	10,240	9,442	0,311	
8	9	7	0,040	6,394	0,367	
9	8	5	3,240	5,378	0,143	
10	8	6	0,640	5,378	0,387	
11	9	6	0,640	6,394	0,155	
12	9	5	3,240	6,394	1,943	
13	8	6	0,640	5,378	0,387	
14	12	8	1,440	9,442	2,079	
15		6,800	33,600		8,393	
16		$\bar{y}$	Q		Q_e	
17						
18				$=\text{СУММ}(E5:E14)$		
19	$Q_r = Q - Q_e$		25,207			
20	F		24,025			
21				$=C19*(10-2)/E15$		

Рисунок 3.2.1 – Вычисление величины F-критерия

Задание

По данным из практической работы №3 оценить при уровне $\alpha = 0,05$ значимость уравнения регрессии (3.8), построенного в практической работе №3.

Содержание отчета

1. Цель занятия.
2. Краткие теоретические положения.
3. Технология выполнения работы и полученные результаты: задание; результаты измерений; формулы для получения оценок выборки и результаты расчетов.
4. Выводы.

Контрольные вопросы

1. Что такое критерий Фишера?

2. При каком условии значимо уравнение парной регрессии?
3. Как определяется факторная сумма квадратов?
4. Как определяется остаточная сумма квадратов?
5. Что такое линейная парная регрессия?
6. Что такое коэффициент детерминации?

Практическое занятие №4

Вычисление коэффициентов уравнения линейной регрессии с помощью статистических функций

Цель занятия: освоение навыков вычисления коэффициентов уравнения линейной регрессии.

Теоретические сведения

В вычислительной среде табличного процессора MS Excel эта задача решается при помощи статистических функций НАКЛОН (наклон прямой относительно оси X, коэффициент b) и ОТРЕЗОК (отрезок, отсекаемый прямой на оси Y, коэффициент a).

Для знакомства с этими возможностями введем необходимые исходные данные (таблица 4.1). В столбцах B и C вводятся данные таблицы 4.2, записи в столбце E играют роль подсказок, столбец F заполняется по мере обработки. В ячейку F3 вводится функция НАКЛОН, в ячейку F4 - ОТРЕЗОК. Обе эти функции имеют два аргумента: диапазон ячеек со значениями Y (C3:C12) и диапазон ячеек со значениями X (B3:B12).

Статистическая функция КВПИРСОН вычисляет значение коэффициента детерминации.

Таблица 4.1 – Исходные данные

Линейная зависимость			Обработка	
№	X	Y		
1	8	5	НАКЛОН	1,0164
2	11	10	ОТРЕЗОК	-2,754
3	12	10	КВПИРСОН	0,7502
4	9	7		
5	8	5		
6	8	6		
7	9	6		
8	9	5		
9	8	6		
10	12	8		

Таблица 4.2 - Результаты исследования

i	1	2	3	4	5	6	7	8	9	10
X	8	11	12	9	8	8	9	9	8	12
Y	5	10	10	7	5	6	6	5	6	8

Функция ЛИНЕЙН(изв_знач_y; изв_знач_x; константа; стат) вычисляет коэффициенты линейной регрессии, коэффициент детерминации R, F-статистику.

В поле «изв_знач_y» вводится диапазон значений Y (C3:C12); «изв_знач_x» - диапазон значений X (B3:B12); константа устанавливается на 0, если заранее известно, что свободный член равен 0, и на 1 в противном случае; стат устанавливается на 0, если не нужен вывод дополнительных сведений регрессионного анализа, и на 1 в противном случае.

Порядок использования функции ЛИНЕЙН:

1. Выделить область пустых ячеек 5x2 (5 строк, 2 столбца) для вывода результатов регрессионной статистики и 1x2 для вывода только коэффициентов a, b.

2. Ввести функцию ЛИНЕЙН вручную или через Мастер функций.

3. После корректного ввода функции в левой верхней ячейке выделенной таблицы появится первый итоговый элемент таблицы. Чтобы раскрыть всю таблицу, следует сначала нажать клавишу F2, а затем одновременно нажать клавиши [Ctrl], [Shift], [Enter]. Далее появляется следующая регрессионная статистика, представленная в таблице 4.3.

Таблица 4.3 – Регрессионная статистика

Значение коэффициента b	Значение коэффициента a
Среднеквадратическое отклонение b	Среднеквадратическое отклонение a
Коэффициент детерминации R^2	Среднеквадратическое отклонение y
F-статистика	Число степеней свободы
Регрессионная сумма квадратов	Остаточная сумма квадратов

Технология выполнения

В результате выполнения алгоритма, приведенного в теоретических сведениях, получим таблицу 4.4. Значения в таблице 4.4 совпадают со значениями, полученными в практических занятиях №3, 3.2.

Таблица 4.4 – Регрессионная статистика

Значение коэффициента b	1,0164	-2,754	Значение коэффициента a
Среднеквадратическое отклонение b	0,2074	1,9759	Среднеквадратическое отклонение a
Коэффициент детерминации R^2	0,7502	1,0243	Среднеквадратическое отклонение y
F-статистика	24,025	8	Число степеней свободы
Регрессионная сумма квадратов	25,207	8,3934	Остаточная сумма квадратов

На рисунках 4.1, 4.2 представлено вычисление коэффициентов уравнения линейной регрессии в Excel.

Задание

По данным из практического занятия №3 вычислить коэффициент уравнения линейной регрессии.

	A	B	C	D	E	F
1	b_0	-2,75				
2	b_1	1,016		=B\$1+B\$2*A5		
3	Исходные данные					
4	x_i	y_i	$\hat{y}_i$	$e_i = \hat{y}_i - y_i$	e_i^2	$(x_i - \bar{x})^2$
5	8	5	5,378	0,378	0,143	1,96
6	11	10	8,426	-1,574	2,477	2,56
7	12	10	9,442	-0,558	0,311	6,76
8	9	7	6,394	-0,606	0,367	0,16
9	8	5	5,378	0,378	0,143	1,96
10	8	6	5,378	-0,622	0,387	1,96
11	9	6	6,394	0,394	0,155	0,16
12	9	5	6,394	1,394	1,943	0,16
13	8	6	5,378	-0,622	0,387	1,96
14	12	8	9,442	1,442	2,079	6,76
15	9,4	6,8			8,393	24,40
16				=СУММ(E5:E14)		
17					=СУММ(F5:F14)	
18	$\bar{x}$	9,400			=B20*B19/F15	
19	x^2	90,800		$S_{b_0}^2$	3,904	
20	S^2	1,049		$S_{b_1}^2$	0,043	
21						
22	=E15/(10-2)				=B20/F15	

Рисунок 4.1 - Вычисление дисперсий оценок b_0, b_1

	A	B	C	D	E	F
1	Исходные данные		=ПРЕДСКАЗ(A3;\$B\$3:\$B\$12;\$A\$3:\$A\$12)			
2	x_i	y_i	$\hat{y}_i$			
3	8	5	5,377			
4	11	10	8,426			
5	12	10	9,443			
6	9	7	6,393			
7	8	5	5,377			
8	8	6	5,377			
9	9	6	6,393			
10	9	5	6,393			
11	8	6	5,377			
12	12	8	9,443			
13						
14						
15		b_0	-2,754	=ОТРЕЗОК(B3:B12;A3:A12)		
16		b_1	1,016	=НАКЛОН(B3:B12;A3:A12)		
17		S	1,024	=СТОПУХ(B3:B12;A3:A12)		

Рисунок 4.2 – Использование функция Excel

Содержание отчета

1. Цель занятия.
2. Краткие теоретические положения.

3. Технология выполнения работы и полученные результаты: задание; результаты измерений; формулы для получения оценок выборки и результаты расчетов.

4. Выводы.

Контрольные вопросы

1. При помощи какой статистической функции в табличном процессоре MS Excel решается задача вычисления коэффициента уравнения линейной регрессии?

2. Какая функция в табличном процессоре MS Excel вычисляет значение коэффициента детерминации?

3. Какая функция в табличном процессоре MS Excel вычисляет F-статистику?

4. Перечислите порядок использования ЛИНЕЙН.

5. Какая функция в табличном процессоре MS Excel вычисляет среднеквадратическое отклонение?

Практическая работа №4.1

Автокорреляция остатков. Статистика Дарбина-Уотсона

Цель занятия: научиться пользоваться статистикой Дарбина-Уотсона.

Теоретические сведения

Одна из предпосылок МНК - это независимость значений случайных отклонений от значений отклонений во всех других наблюдениях. Автокорреляция (последовательная корреляция) - это корреляция между наблюдаемыми показателями, упорядоченными во времени (временные ряды) или в пространстве (перекрестные ряды). Нам неизвестны истинные значения отклонений $e_i = y_i - f(x_i)$, $i = 1, \dots, n$. Поэтому выводы об их независимости делаются на основе оценок $e_i = y_i - f(x_i)$, $i = 1, \dots, n$. При этом обычно проверяется некоррелированность только соседних величин. При наличии автокорреляции остатков полученное уравнение регрессии считается неудовлетворительным.

Критерий Дарбина-Уотсона - наиболее известный способ обнаружения автокорреляции первого порядка. Пусть n - число наблюдений, k - число факторов модели, уровень значимости α . Для n , k , α по таблицам распределения Дарбина-Уотсона (приложение 4) находим числа d_i и d_u . Вычисляем статистику Дарбина-Уотсона

$$DW = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=2}^n e_i^2}.$$

Здесь $e_i = y_i - (ax_i + b)$ - остатки, вычисленные по уравнению линейной регрессии $y = a + bx$ для статистических данных (x_i, y) .

Если $DW < d_i$, то это свидетельствует о положительной автокорреляции остатков. Если $DW > 4 - d_i$, то это свидетельствует об отрицательной автокорреляции остатков. При

$d_i < DW < 4 - d_i$ гипотеза об отсутствии автокорреляции остатков принимается. Если

$$d_i < DW < d_u \text{ или } 4 - d_u < DW < 4 - d_i,$$

то гипотеза об отсутствии автокорреляции остатков не может быть ни принята, ни отвергнута.

Технология выполнения

По данным из практического занятия №3 определим наличие автокорреляции с помощью критерия Дарбина-Уотсона. В столбце $(y - \hat{y}_i)^2$ таблицы 3.2 получили

$$\sum_{i=1}^n e_i^2 = 8,39.$$

Во второй столбец e_i таблицы 4.1.1 введем данные столбца $(y - \hat{y}_i)$ таблицы 3.2. из каждого числа 2-го столбца вычитаем предыдущее число этого же столбца и результат введем в третий столбец $e_i - e_{i-1}$. Статистика Дарбина-Уотсона равна

$$DW = \frac{\sum_{i=1}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} = \frac{16,78}{8,39} = 2.$$

Таблица 4.1.1 – Результаты автокорреляции

№ п/п	e_i	$e_i - e_{i-1}$	$(e_i - e_{i-1})^2$
1	0,38	-	-
2	0,38	0	0
3	-0,62	-1,00	1
4	-0,62	0	0
5	-0,61	0,01	0

Продолжение таблицы 4.1.1

6	0,39	1	1
7	1,39	1	1
8	-1,57	-2,96	8,76
9	-0,56	1,01	1,02
10	1,44	2	4
Сумма			16,78

По таблице распределения Дарбина-Уотсона (приложение 4) находим числа $d_i = 0,88$ и $d_u = 1,32$ при $n=10$, $k=1$. Так как $d_u < DW < 4 - d_u$ ($1,32 < 2 < 2,68$), то гипотеза об отсутствии автокорреляции остатков не отклоняется на уровне значимости 0,05. Это является одним из подтверждений высоко качества модели.

Задание

По данным из практического занятия №3 вычислить автокорреляцию остатков и статистику Дарбина-Уотсона.

Содержание отчета

1. Цель работы.
2. Краткие теоретические положения.
3. Технология выполнения работы и полученные результаты: задание; результаты измерений; формулы для получения оценок выборки и результаты расчетов.
4. Выводы.

Контрольные вопросы

1. Что такое автокорреляция?
2. В каком случае полученное уравнение регрессии считается неудовлетворительным?
3. Что такое критерий Дарбина-Уотсона?
4. В каком случае можно получить положительную автокорреляцию?

5. При каком условии гипотеза об отсутствии автокорреляции остатков принимается?

Практическое занятие №5

Построение нелинейной регрессии

Цель занятия: научиться построению уравнения нелинейной регрессии и вычислению индекса детерминации R^2 .

Теоретические сведения

В случае нелинейных зависимостей с помощью некоторых преобразований можно перейти к линейным. Приведем некоторые из нелинейных функций и их модификацию к линейной функции – таблица 5.1.

Таблица 5.1 – Модификация нелинейных функций

№ п/п	Исходная спецификация	Преобразо вание $x \rightarrow x^*$	Преобразова ние $y \rightarrow y^*$	Вычислен ие b по b^*	Вычис ление a по a^*
1	$y = a + \frac{b}{x} + \varepsilon$	$x^* = \frac{1}{x}$	$y^* = y$	$b = b^*$	$a = a^*$
2	$y = \frac{1}{a + bx + \varepsilon}$	$x^* = x$	$y^* = \frac{1}{y}$	$b = b^*$	$a = a^*$
3	$y = \frac{x}{a + bx + x\varepsilon}$	$x^* = \frac{1}{x}$	$y^* = \frac{1}{y}$	$b = a^*$	$a = b^*$
4	$y = ae^{bx+\varepsilon}$	$x^* = x$	$y^* = \ln y$	$b = b^*$	$a = e^{a^*}$
5	$y = ae^{\frac{b}{x}+\varepsilon}$	$x^* = \frac{1}{x}$	$y^* = \ln y$	$b = b^*$	$a = e^{a^*}$
6	$y = \frac{1}{a + be^{-x} + \varepsilon}$	$x^* = e^{-x}$	$y^* = \frac{1}{y}$	$b = b^*$	$a = a^*$
7	$y = ax^b e^{\varepsilon}$	$x^* = \ln x$	$y^* = \ln y$	$b = b^*$	$a = e^{a^*}$

Технология выполнения

В таблице 3.2 приведены значения независимой переменной X (товарооборот, тыс. р.) и значения зависимой переменной Y (уровень издержек обращения по отношению к товарообороту, %).

Таблица 3.2 - Значения независимой переменной X и значения зависимой переменной Y

X	7	10	15	20	30	45	60	120
Y	10	9	7,2	6	6,3	5,8	5,4	5

На рисунке 5.1 изображены данные в виде диаграммы. Из этого рисунка видно, что зависимость между показателями нелинейная. Будем искать регрессионную зависимость в виде степенной: $y = ax^b$; переходя к логарифмам (модификация таблица 3.1), получим линейную зависимость $y^* = a^* + bx^*$, где $y^* = \ln y$, $x^* = \ln x$.

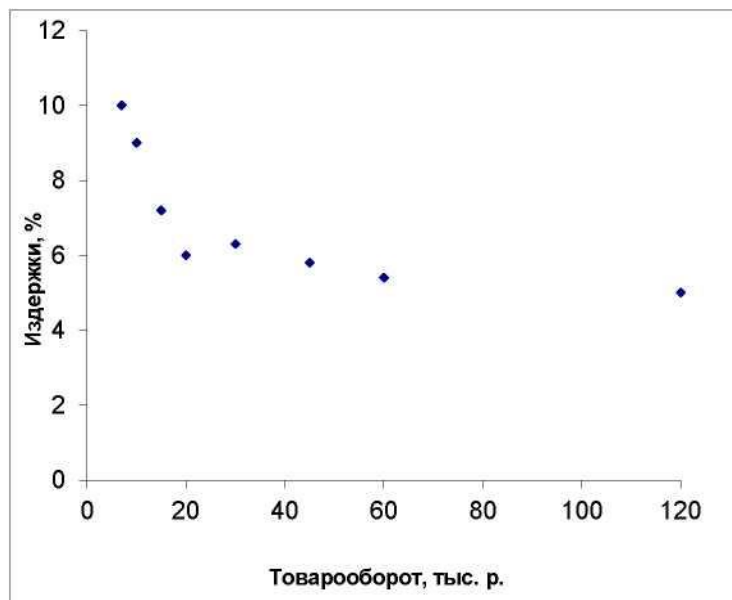


Рисунок 5.1 – Диаграмма товарооборота и уровня издержек обращения по отношению к товарообороту

Определяем коэффициенты a^* , b для модификации данных. Исходные данные и основные вычисления приведены в таблице 5.3.

Таблица 5.3 – Исходные данные

№ п/п	x	$x^* = \ln x$	$y^* = \ln y$	x^2	y^2	$(y - \bar{y}_x)^2$	$\hat{y}_x$	$(y - \hat{y}_x)^2$
1	7	1,95	2,30	3,79	5,30	10,00	9,1 2	0,77
2	10	2,30	2,20	5,30	4,83	4,68	8,3 7	0,40
3	15	2,71	1,97	7,33	3,90	0,13	7,5 9	0,15
4	20	3,00	1,79	8,97	3,21	0,70	7,0 8	1,17
5	30	3,40	1,84	11,5 7	3,39	0,29	6,4 2	0,02
6	45	3,81	1,76	14,4 9	3,09	1,08	5,8 2	0,00
7	60	4,09	1,69	16,7 6	2,84	2,07	5,4 3	0,00
8	120	4,79	1,61	22,9 2	2,59	3,38	4,6 0	0,16
Итого		26,04	15,16	91,1 4	29,1 5	22,32	-	2,67

Аналогично практическому занятию №3 определим коэффициенты a^* и b из соотношений (3.7), (3.4): $a^* \approx 2,68$; $b \approx -0,24$. По величине $a^* \approx 2,68$ потенцированием находим исходный коэффициент a : $a = e^{2,68} \approx 14,58$.

Таким образом, уравнение регрессии имеет вид: $y = 14,58x^{-0,24}$. Изобразим эту кривую на диаграмме с исходными данными (рисунок 5.2).

Аналогично практической работе 3.2 вычислим

$$Q_e = 2,67; Q_r = Q - Q_e = 22,32 - 2,67 = 19,65.$$

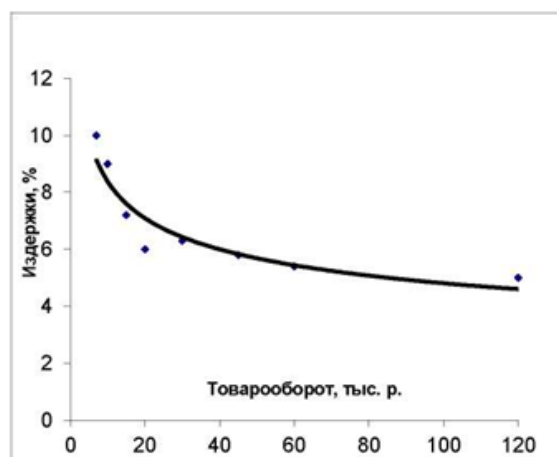


Рисунок 5.2 – Уравнение регрессии на диаграмме с исходными данными

Согласно формуле (3.2.5) индекс детерминации равен

$$R^2 = 1 - \frac{Q_e}{Q} = 1 - \frac{2,67}{22,32} = 1 - 0,12 = 0,88.$$

Вычисленное значение говорит о том, что уравнением регрессии объясняется 88% дисперсии результативного признака, а на долю прочих факторов приходится лишь 12%.

Задание

Построить уравнение нелинейной регрессии и вычислить индекс детерминации R^2 по исходным данным:

X	14*N	20*N	30*N	40*N	60*N	90*N	120*N	240*N
Y	20*N	18*N	14,2*N	12*N	12,6*N	11,6*N	10,8*N	10*N

где N – номер варианта студента по списку.

Содержание отчета

1. Цель занятия.
2. Краткие теоретические положения.

3. Технология выполнения работы и полученные результаты: задание; результаты измерений; формулы для получения оценок выборки и результаты расчетов.

4. Выводы.

Контрольные вопросы

1. Что такое параболическая зависимость величины X от величины Y ?
2. Что такое степенная зависимость величины X от величины Y ?
3. Что такое логарифмическая зависимость величины X от величины Y ?
4. Перечислите алгоритм построения диаграммы в среде Excel.
5. Что такое индекс детерминации?

Практическое занятие №6

Построение нелинейной регрессии с использованием команды «Добавить линию тренда». Взвешенный метод наименьших квадратов

Цель занятия: освоить навыки построения уравнения нелинейной регрессии с использованием команды «Добавить линию тренда»; научиться применять взвешенный метод наименьших квадратов для коррекции гетероскедичности остатков.

Теоретические сведения

6.1 Построение нелинейной регрессии с использованием команды «Добавить линию тренда»

Команда «Добавить линию тренда» используется для выделения тренда (медленных изменений) при анализе временных рядов. Однако эту команду можно использовать и для построения уравнения нелинейной регрессии, рассматривая в качестве времени t независимую переменную x .

Эта команда позволяет построить следующие уравнения регрессии:

- линейная $\hat{y} = a + bx$;
- полиномиальная

$$\hat{y} = a + b_1x + b_2x^2 + b_3x^3 + b_4x^4 + b_5x^5 + b_6x^6$$
;
- Логарифмическая $\hat{y} = a + b$;
- степенная $\hat{y} = ax^b$;
- экспоненциальная $\hat{y} = ae^{bx}$.

Для построения одной из перечисленных регрессий необходимо выполнить следующие шаги:

Шаг 1. Ввести по столбцам исходные данные $\{x_i, y_i\}, i = 1, 2, \dots, n$ (рисунок 6.1).

Шаг 2. По этим данным построить график в декартовой системе координат (рисунок 6.1).

Шаг 3. Установить курсор на любую точку построенного графика, сделать щелчок правой кнопкой и в появившемся

контекстном меню выполнить команду **Добавить линию тренда** (рисунок 6.1).

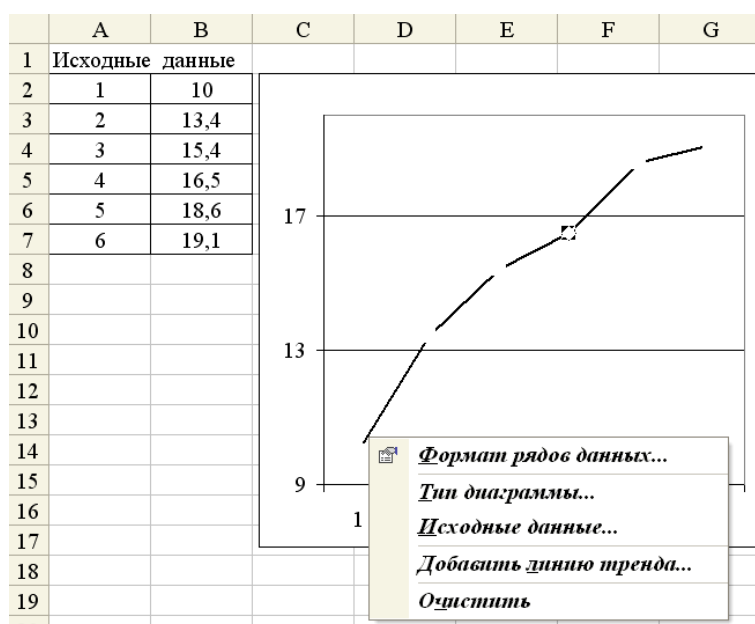


Рисунок 6.1 – Построение регрессии

Шаг 4. В появившемся диалоговом окне (рисунок 6.2) активизировать закладку «Тип» и выбрать нужное уравнение регрессии.

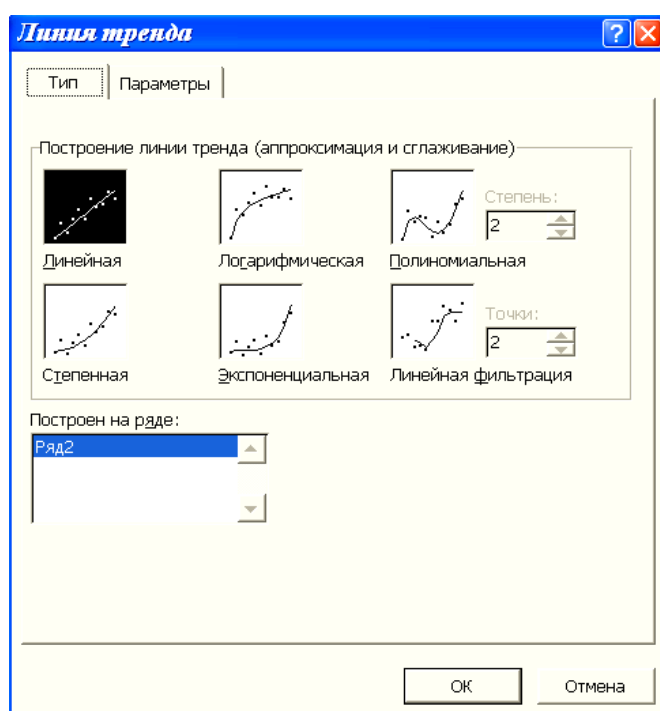


Рисунок 6.2 – Выбор вида уравнения регрессии

Шаг 5. Активизировать закладку «Параметры» (рисунок 6.3) и «включить» необходимые опции:

- «Показать уравнение на диаграмме» - на диаграмме будет показано выбранное уравнение регрессии с вычисленными коэффициентами;
- «Поместить на диаграмму величину достоверности аппроксимации (R^2)» - на диаграмме будет показано значение коэффициента детерминации R^2 (для нелинейной регрессии - индекс детерминации).

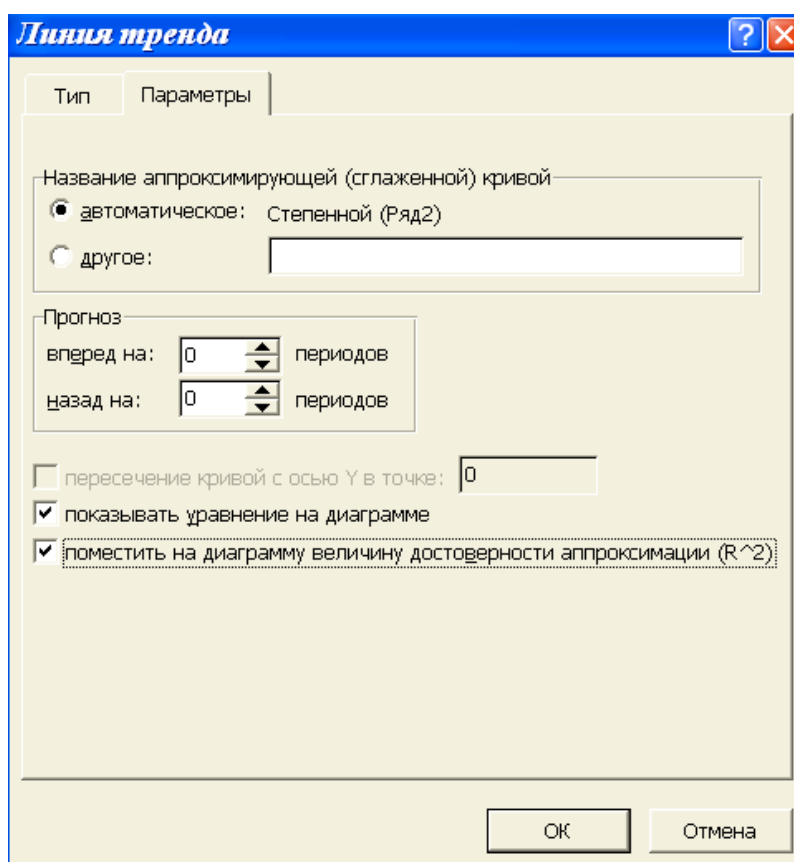


Рисунок 6.3 - Задание опций вывода информации

Если по построенному уравнению регрессии необходимо выполнить прогноз, то нужно указать число периодов прогноза.

Шаг 6. После задания всех перечисленных опций на диаграмме появится формула построенного уравнения регрессии и значение индекса детерминации R^2 .

6.2 Взвешенный метод наименьших квадратов

Существенным условием классической регрессионной модели являются следующие требования:

1. $M(\varepsilon_i \cdot \varepsilon_j) = 0$, если $i \neq j$.
2. $M(\varepsilon_i \cdot \varepsilon_j) = M(\varepsilon_i^2) = \sigma^2$, $i = j$.

Первое равенство означает некоррелированность возмущений между собой и это требование для пространственной выборки, как правило, выполняется. Второе требование – равенство дисперсий возмущений ε_i , называемое условием гомоскедастичности. На практике в ряде случаев это условие не выполняется и имеет место гетероскедастичность модели. Гетероскедастичность «портит хорошие» свойства оценок классической модели и, как правило, ее необходимо «устранить». Поэтому необходимо определить какая ситуация – гомоскедастичность или гетероскедастичность имеет место. Для этого используют специальные статистические тесты. Приведем один из таких тестов.

Тест ранговой корреляции Спирмена. В качестве нулевой гипотезы H_0 примем гипотезу о гомоскедастичности модели. Идея теста заключается в том, что абсолютные величины невязок e_i являются оценками для σ_i и поэтому в случае гетероскедастичности величины $|e_i|$ и x_i будут коррелированы. Степень коррелированности определяется по величине коэффициента ранговой корреляции Спирмена ρ_{xe} .

Для вычисления этого коэффициента необходимо выполнить следующие шаги:

1. Предполагая гомоскедастичность модели, вычислить коэффициенты линейного уравнения регрессии и определить значения невязок e_i .

2. Определить ранг ρ_{e_i} невязки e_i и ранг ρ_{x_i} значения x_i .

Для вычисления рангов невязок необходимо первоначально упорядочить e_i , например, по возрастанию. Порядковый номер e_i в

таком упорядоченном ряду и будет рангом ρ_{e_i} . Аналогичным образом вычисляются ранги ρ_{x_i} .

3. Вычислить коэффициента ранговой корреляции по следующей формуле:

$$\rho_{xe} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n},$$

где $d_i = \rho_{e_i} - \rho_{x_i}$ - разность между рангами значений e_i и x_i . Нетрудно заметить, что если ранги $\rho_{e_i} = \rho_{x_i}$, то $|\rho_{xe}| = 1$.

После определения ρ_{xe} вычисляется критерий

$$T_\rho = \frac{\rho_{xe} \sqrt{n-2}}{\sqrt{1-\rho_{xe}^2}}.$$

Если выполняется неравенство

$$|T_\rho| > t(1-\alpha, n-2),$$

то нулевая гипотеза H_0 отвергается с уровнем значимости α и принимается гипотеза о гетероскедастичности модели. Значение $t(1-\alpha, n-2)$ вычисляется выражением

$$t(1-\alpha, n-2) = \text{СТЬДРАСПОБР}(\alpha, n-2).$$

Заметим, что в приведенном тесте не делается никаких предположений о законе распределений возмущений ε_i и тест следует использовать при $n \geq 10$.

Для вычисления рангов ρ_{e_i} , ρ_{x_i} целесообразно использовать функцию РАНГ из категории статистических функций Excel. Обращение к этой функции имеет вид:

РАНГ(число; значения; порядок),

где число – значение, для которого определяется ранг;

значения – массив исходных числовых данных (как правило, диапазон ячеек), относительно которого вычисляется ранг заданного числа;

порядок – величина, определяющая способ упорядочивания при вычислении ранга: если порядок = 0 или этот параметр опущен в обращении, упорядочивание в порядке убывания; если порядок > 0, то упорядочивание в порядке возрастания.

Замечание: Вычисление ранга значений x_i обуславливает наличие в модели только одной независимой переменной. Поэтому изложенный тест применим только для парной регрессии.

Метод взвешенных наименьших квадратов. Предположим, что неравенство $|T_p| > t(1 - \alpha, n - 2)$, выполнилось, и была принята гипотеза о гетероскедастичности модели, т.е. каждое возмущение ε_i имеет свою дисперсию σ_i^2 . В этом случае ковариационная матрица вектора возмущения ε остается диагональной, но на главной диагонали стоят не равные элементы и такую матрицу можно записать в виде

$$V_\varepsilon = \text{diag}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2\}$$

Гомоскедастичной модели ковариационная матрица вектора возмущения ε определяется выражением

$$V_\varepsilon = \text{diag}\{\sigma^2, \sigma^2, \dots, \sigma^2\}$$

Возникает вопрос: «Каким свойством будет характеризоваться оценка обыкновенного (классического) МНК вида:

$$b = (X^T X)^{-1} \cdot X^T y,$$

Дадим ответ на этот вопрос.

1. Оценка $b = (X^T X)^{-1} \cdot X^T y$, и в этом случае будет несмещенной и состоятельной. Это означает, что определение значений зависимой переменной можно осуществлять по уравнению регрессии с коэффициентами $b = (X^T X)^{-1} \cdot X^T y$.

2. Оценка $b = (X^T X)^{-1} \cdot X^T y$, не будет эффективной, т.е. существуют другие оценки (и соответствующие алгоритмы вычисления коэффициентов уравнения регрессии), которые имеют меньшую дисперсию.

3. Изложенный ранее статистический анализ построенной регрессии (точность модели, оценка значимости, построение доверительных интервалов и т.д.) оказывается неверным для гетероскедастичной модели.

Возникает традиционный вопрос: «Что делать?». Для ответа на этот вопрос обратимся к функционалу

$$F(b_0, b_1) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - b_0 + b_1 x_i)^2 \quad \text{классического МНК, в}$$

котором все квадраты невязок входят с одинаковыми весами, т.к. дисперсия возмущений одинакова. В случае неодинаковых дисперсий квадрат невязки должен «входить» в функционал МНК с весом, обратным величине соответствующей дисперсии.

Таким образом, приходим к функционалу

$$F(b) = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{\sigma_i^2} = \sum_{i=1}^n \frac{(e_i)^2}{\sigma_i^2} = (y - Xb)^T \cdot V_\varepsilon^{-1} \cdot (y - Xb),$$

где матрица V_ε определяется выражением $V_\varepsilon = \text{diag}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2\}$

Можно показать, что вектор b , доставляющий минимум

$$\text{функционалу } F(b) = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{\sigma_i^2} = \sum_{i=1}^n \frac{(e_i)^2}{\sigma_i^2} = (y - Xb)^T \cdot V_\varepsilon^{-1} \cdot (y - Xb)$$

является решением следующей матричной системы уравнений:

$$(X^T V_\varepsilon^{-1} X) b = X^T V_\varepsilon^{-1} y$$

и определяется выражением:

$$b = (X^T V_{\varepsilon}^{-1} X)^{-1} X^T V_{\varepsilon}^{-1} y.$$

Метод вычисления оценок на основе минимизации функционала $F(b) = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{\sigma_i^2} = \sum_{i=1}^n \frac{(e_i)^2}{\sigma_i^2} = (y - Xb)^T \cdot V_{\varepsilon}^{-1} \cdot (y - Xb)$ получил название метода взвешенных наименьших квадратов, а оценка $b = (X^T V_{\varepsilon}^{-1} X)^{-1} X^T V_{\varepsilon}^{-1} y$ называется оценкой метода взвешенных наименьших квадратов. Эта оценка является несмещенной, состоятельной и эффективной. Ковариационная матрица вектора b определяется выражением:

$$V_b = (X^T V_{\varepsilon}^{-1} X)^{-1}.$$

При реализации метода взвешенных наименьших квадратов в Excel будут полезны следующие соотношения:

$$\begin{aligned} [X^T V_{\varepsilon}^{-1} X]_{l,k} &= \sum_{i=1}^n \frac{[X]_{i,l} \cdot [X]_{i,k}}{\sigma_i^2}, \\ [X^T V_{\varepsilon}^{-1} y]_l &= \sum_{i=1}^n \frac{[X]_{i,l} \cdot [y]_i}{\sigma_i^2}, \end{aligned}$$

где $[X]_{i,l}$ - означает i , l -ый элемент матрицы X . Очевидно, что $X^T V_{\varepsilon}^{-1} X$ есть матрица размером $m \times m$, $X^T V_{\varepsilon}^{-1} y$ - вектор из m проекций, m – число коэффициентов регрессионной модели.

Технология выполнения

6.1. По данным таблицы 6.1 построим степенное уравнение регрессии $\hat{y} = ax^b$.

Таблица 6.1 - Значения независимой переменной X и значения зависимой переменной Y

X	7	10	15	20	30	45	60	120
Y	10	9	7,2	6	6,3	5,8	5,4	5

После выполнения алгоритма приведенного в теоретических сведениях получим уравнение $\hat{y} = 14,58x^{-0,24}$, для которого индекс детерминации равен $R^2 = 0,877$ (рисунки 6.4). Такая величина говорит о хорошем соответствии построенного уравнения исходным данным.

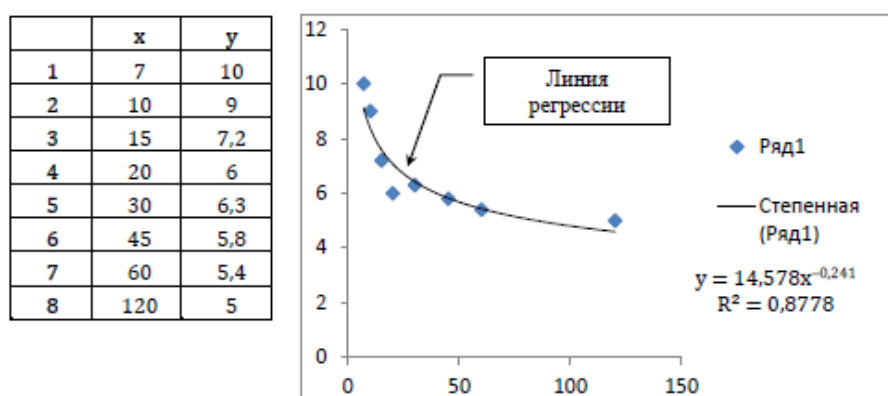


Рисунок 6.4 – Индекс детерминации

6.2. Построить точечную диаграмму исходных данных, поместить на нее линию тренда, его уравнение и индекс детерминации.

1. Сформировать массив остатков.
2. Модифицировать массив зависимой переменной следующим образом: $x_i^* = 1/x_i$.
3. Модифицировать массив зависимой переменной следующим образом: $y_i^* = y_i / x_i$.
4. По полученным данным модифицировать регрессии x_i^*, y_i^* построить диаграмму облака рассеяния, поместить на нее линию тренда с уравнением линии регрессии и индексом детерминации. Сравнить это уравнение с исходной регрессией.

Задание

6.1 Построить уравнение нелинейной регрессии с использованием команды «Добавить линию тренда» по исходным данным.

6.2 Произвести обработку исходных данных взвешенным методом наименьших квадратов.

Исходные данные:

X	7*N	10*N	15*N	20*N	60*N	45*N	60*N	120*N
Y	10*N	9*N	7,2*N	6*N	6,3*N	5,8*N	5,4*N	5*N

где N – номер варианта студента по списку.

Содержание отчета

1. Цель занятия.
2. Краткие теоретические положения.
3. Технология выполнения работы и полученные результаты: задание; результаты измерений; формулы для получения оценок выборки и результаты расчетов.
4. Выводы.

Контрольные вопросы

1. Какие уравнения регрессии позволяет построить команда «Добавить линию тренда»?
2. Что такое метод взвешенных наименьших квадратов?
3. Что такое коэффициент метода взвешенных наименьших квадратов?
4. Что такое гомоскедастичность остатков?
5. Что такое гетероскедастичность остатков?

Практическое занятие №7

Вычисление коэффициентов линейной множественной регрессии

Цель занятия: научиться вычислять коэффициент линейной множественной регрессии с помощью MS Excel.

Теоретические сведения

Вектор коэффициентов, найденный методом наименьших квадратов, является решением следующей системы уравнений:

$$X^T X b = X^T Y,$$

где X – матрица размерности 10×3 , первый столбец которой составлен из 1, а другие два столбца составлены из значений x_{i1}, x_{i2} , т.е. матрица X имеет структуру

$$X = \begin{pmatrix} 1 & 8 & 5 \\ 1 & 11 & 8 \\ \dots & \dots & \dots \\ 1 & 12 & 7 \end{pmatrix},$$

а Y – вектор, составленный из 10 значений y_i , т.е.

$$Y = \begin{pmatrix} 5 \\ 10 \\ \dots \\ 8 \end{pmatrix}.$$

Матрица $X^T X$ имеет обратную матрицу $(X^T X)^{-1}$, и тогда вектор коэффициентов равен

$$b = (X^T X)^{-1} (X^T Y). \quad (7.1)$$

Для реализации этой матричной формулы необходимо выполнить следующие операции: транспонирование; умножение матриц (частный случай - умножение матрицы на вектор); вычисление обратной матрицы. Все эти операции можно реализовать с помощью матричных функций MS Excel категории функций Ссылки и массивы. Для работы с этими функциями можно либо обратиться к Мастеру функций и выбрать нужную категорию функций, затем указать имя функции и задать соответствующие диапазоны ячеек, либо непосредственно ввести с клавиатуры имя функции и задать соответствующие диапазоны ячеек.

Функция ТРАНСП(диапазон ячеек) осуществляет транспонирование матрицы, где параметр диапазон ячеек задает все элементы транспонируемой матрицы (или вектора).

Функция МУМНОЖ(диапазон_1; диапазон_2) осуществляет умножение матриц в категории функций Математические. Параметр диапазон_1 задает элементы первой из перемножаемых матриц, а параметр диапазон_2 - элементы второй матрицы. При этом перемножаемые матрицы должны иметь соответствующие размерности (если первая матрица размерности $n \times k$, вторая - $k \times m$, то результатом будет матрица размерности $n \times m$).

Функция МОБР(диапазон ячеек) осуществляет обращение матрицы (вычисление обратной матрицы) в категории функций Математические. Параметр диапазон ячеек задает все элементы обращаемой матрицы, которая должна быть квадратной и невырожденной.

При использовании этих функций необходимо соблюдать следующий порядок действий:

1. Выделить фрагмент ячеек, в которые будет занесен результат выполнения матричных функций (при этом надо учитывать размеры исходных матриц);
2. Ввести арифметическое выражение, содержащее обращение к матричным функциям MS Excel;
3. Одновременно нажать клавиши [Ctrl], [Shift], [Enter]. Если этого не сделать, то вычисляется только один элемент результирующей матрицы или вектора.

Технология выполнения

Данные о сменной добыче угля на одного рабочего (переменная Y , измеряемая в тоннах), мощности пласта (переменная X_1 , измеряемая в метрах) и об уровне механизации работ в шахте (переменная X_2 , измеряемая в процентах), характеризующие процесс добычи угля в 10 шахтах, приведены в таблице 7.1.

Таблица 7.1 – Исходные данные

Номер шахты i	X_1	X_2	Y
1	8	5	5
2	11	8	10
3	12	8	10
4	9	5	7
5	8	7	5
6	8	8	6
7	9	6	6
8	9	4	5
9	8	5	6
10	12	7	8

Предполагая, что между переменными Y , X_1 , X_2 существует линейная зависимость, необходимо найти аналитическое выражение для этой зависимости, т.е. построить уравнение

линейной регрессии, вычислить коэффициенты $b = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}$ уравнения

регрессии.

Сформируем матрицу X и вектор Y . Затем выполним формирование матрицы $X^T X$, вектора $X^T Y$ и вычисление вектора $b = (b_0, b_1, b_2)^T$ по формуле (7.1). В результате получим вектор

коэффициентов $b = \begin{pmatrix} -3,5393 \\ 0,8539 \\ 0,3670 \end{pmatrix}$ и тогда уравнение регрессии

$\hat{y}(x_1, x_2) = b_0 + b_1 x_1 + b_2 x_2$ имеет вид

$$\hat{y}(x_1, x_2) = -3,5393 + 0,8539x_1 + 0,367x_2.$$

На рисунках 7.1, 7.2 представлен результат вычисления линейной множественной регрессии.

	A	B	C	D	E	F	G
1			Исходные матрица A и вектор x				
2		2	2	6			
3		5	4	8			5
4	матрица A =	7	3	5	вектор x =		6
5		5	1	3			7
6		6	5	2			
7	Транспонирование и умножение матриц						
8				=ТРАНСП(B2:D6)			
9							
10		2	5	7	5	6	
11	матрица $A^T =$	2	4	3	1	5	
12		6	8	5	3	2	
13							
14		44	66	50			
15	матрица $A^T A =$	66	105	87			
16		50	87	83			
17							
18		64		=МУМНОЖ(B2:D6;B10:F12)			
19	вектор $A \cdot x =$	105					
20		88		=МУМНОЖ(B2:D6;G3:G5)			

Рисунок 7.1 – Матричные функции MS Excel

	A	B	C	D	E	F	
1		1	8	5		5	
2		1	11	8		10	
3		1	12	8		10	
4		1	9	5		7	
5	$X =$	1	8	7	$y =$	5	
6		1	8	8		6	
7		1	9	6		6	
8		1	9	4		5	
9		1	8	5		6	
10		1	12	7		8	
11							
12		10	94	63		68	
13	$X^T \cdot X =$	94	908	603	$X^T y =$	664	
14		63	603	417		445	
15	=МУМНОЖ(ТРАНСП(B1:D10);B1:D10)						
16	=МУМНОЖ(ТРАНСП(B1:D10);F1:F10)						
17		4,0201	-0,323	-0,14		-3,5393	
18	$(X^T \cdot X)^{-1} =$	-0,323	0,054	-0,029	$b =$	0,8539	
19		-0,14	-0,029	0,0653		0,3670	
20							
21	=МОБР(B12:D14)			=МУМНОЖ(B17:D19;F12:F14)			

Рисунок 7.2 – Вычисление коэффициентов регрессии

Задание

По исходным данным построить уравнение линейной регрессии и вычислить вектор коэффициентов $b = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}$ уравнения регрессии.

Исходные данные

X_1	X_2	Y
8*N	5*N	5*N
11*N	8*N	10*N
12*N	8*N	10*N
9*N	5*N	7*N
8*N	7*N	5*N
8*N	8*N	6*N
9*N	6*N	6*N
9*N	4*N	5*N
8*N	5*N	6*N
12*N	7*N	8*N

где N – номер варианта студента по списку.

Содержание отчета

1. Цель занятия.
2. Краткие теоретические положения.
3. Технология выполнения работы и полученные результаты: задание; результаты измерений; формулы для получения оценок выборки и результаты расчетов.
4. Выводы.

Контрольные вопросы

1. Что осуществляет функция ТРАНСП (диапазон ячеек) в Excel?
2. Что осуществляет функция МУМНОЖ (диапазон_1; диапазон_2) в Excel?
3. Что осуществляет функция МОБР (диапазон ячеек) в Excel?
4. Что такое коэффициент линейной множественной регрессии?
5. Приведите уравнение линейной множественной регрессии?

Практическое занятие №8

Вычисление коэффициентов линейной множественной регрессии и проверка значимости в режиме «Регрессия»

Цель занятия: научиться вычислять коэффициент линейной множественной регрессии и проверять значимость экспериментальных данных с помощью MS Excel.

Теоретические сведения

Табличный процессор MS Excel содержит модуль Анализ данных. Этот модуль позволяет выполнить статистический анализ выборочных данных (построение гистограмм, вычисление числовых характеристик и т. д.). Режим работы Регрессия этого модуля осуществляет вычисление коэффициентов линейной множественной регрессии с k переменными, построение доверительных интервалов и проверку значимости уравнения регрессии.

После вызова режима Регрессия на экране появляется диалоговое окно (рисунок 8.1), в котором задаются следующие параметры:

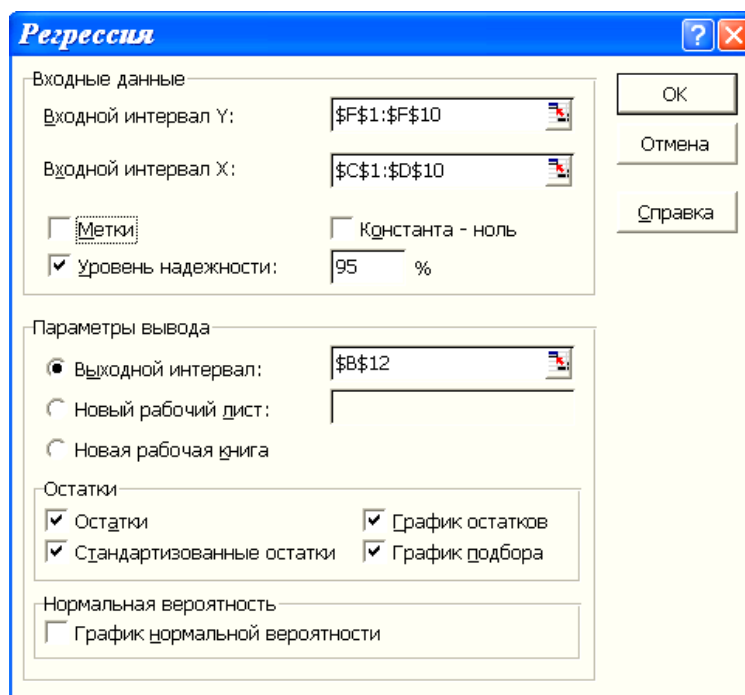


Рисунок 8.1 – Диалоговое окно режима Регрессия

1. Входной интервал Y - вводится диапазон адресов ячеек, содержащих значения y_i - (ячейки должны составлять один столбец).

2. Входной интервал X - вводится диапазон адресов ячеек, содержащих значения независимых переменных. Значения каждой переменной представляются одним столбцом. Количество переменных - не более 16.

3. Метки - включается, если первая строка во входном диапазоне содержит заголовок. В этом случае автоматически будут созданы стандартные названия.

4. Уровень надежности - при включении этого параметра задается надежность Y при построении доверительных интервалов.

5. Константа-ноль - при включении этого параметра коэффициент $b_0 = 0$.

6. Выходной интервал - при включении активизируется поле, в которое необходимо ввести адрес левой верхней ячейки выходного диапазона, который содержит ячейки с результатами вычислений режима Регрессия.

7. Новый рабочий лист - при включении этого параметра открывается новый лист, в который, начиная с ячейки A1, вставляются результаты работы режима Регрессия.

8. Новая рабочая книга - при включении этого параметра открывается новая книга, на первом листе которой, начиная с ячейки A1, вставляются результаты работы режима Регрессия.

9. Остатки - при включении вычисляется столбец, содержащий невязки $y_i - \hat{y}_i$, $i = 1, \dots, n$.

10. Стандартизованные остатки - при включении вычисляется столбец, содержащий стандартизованные остатки.

11. График остатков - при включении выводятся точечные графики невязки в зависимости от значений переменных x_j , $j = 1, \dots, k$. Количество графиков равно числу k переменных x_j .

12. График подбора - при включении выводятся точечные графики предсказанных по построенной регрессии значений y_i от значений переменных x_j , $j = 1, \dots, k$. Количество графиков равно числу k переменных x_j .

Технология выполнения

Данные о сменной добыче угля на одного рабочего (переменная Y , измеряемая в тоннах), мощности пласта (переменная X_1 , измеряемая в метрах) и об уровне механизации работ в шахте (переменная X_2 , измеряемая в процентах), характеризующие процесс добычи угля в 10 шахтах, приведены в таблице 8.1.

Таблица 8.1 – Исходные данные

Номер шахты i	X_1	X_2	Y
1	8	5	5
2	11	8	10
3	12	8	10
4	9	5	7
5	8	7	5
6	8	8	6
7	9	6	6
8	9	4	5
9	8	5	6
10	12	7	8

Первоначально введем в столбец C десять значений первой переменной X_1 в столбец D - десять значений второй переменной X_2 , а в столбец F - десять значений зависимой переменной Y .

После этого вызовем режим Регрессия и в диалоговом окне зададим необходимые параметры (рисунок 8.1). Результаты работы приведены на рисунке 8.2 - 8.3. Заметим, что из-за большой «ширины» таблиц, в которых выводятся результаты работы режима Регрессия, часть результатов помещена в другие ячейки.

Дадим краткую интерпретацию показателям, значения которых вычисляются в режиме Регрессия. Первоначально рассмотрим показатели, объединенные названием Регрессионная статистика (рисунок 8.2).

ВЫВОДИТОГОВ			
<i>Регрессионная статистика</i>			
Множественный R	0,9009		
R-квадрат	0,8116		
Нормированный R-квадрат	0,7578		
Стандартная ошибка	0,9509		
Наблюдения	10		
<i>Дисперсионный анализ</i>			
	<i>df</i>	<i>SS</i>	<i>MS</i>
Регрессия	2	27,2704	13,635
Остаток	7	6,3296	0,904
Итого	9	33,6000	
		<i>F</i>	<i>Значимость F</i>
		15,0794	0,0029

Рисунок 8.2 – Результат работы режима Регрессия

Множественный R - корень квадратный из коэффициента детерминации. R-квадрат - коэффициент детерминации R^2 .

Нормированный R-квадрат - приведенный коэффициент детерминации

$$\hat{R}^2 = 1 - \frac{(n-1)Q_e}{(n-k)Q} = 1 - \frac{n-1}{n-k}(1 - R^2).$$

Стандартная ошибка – оценка S для среднеквадратического отклонения σ .

Наблюдения - число наблюдений n.

Перейдем к показателям, объединенным названием Дисперсионный анализ (рисунок 8.2).

Столбец df - число степеней свободы. Для строки Регрессия показатель равен числу независимых переменных $k_1 = k - 1$; для строки Остаток он равен $k_2 = n - (k + 1) = n - k$; для строки Итого $k_1 + k_2 = n - 1$.

Столбец SS - сумма квадратов отклонений. Для строки Регрессия показатель равен величине Q_r (формулы (3.2.2)); для строки Остаток - равен величине Q_e (см. формулы (3.2.2)); для

строки Итого $Q = Q_r + Q_e$.

Столбец MS – дисперсии, вычисленные по формуле $MS = \frac{SS}{df}$,

т.е. дисперсия на одну степень свободы.

Столбец F – значение F_c , равное F-критерию Фишера, вычисленному по формуле

$$F_c = \frac{Q_r / k_1}{Q_e / k_2}.$$

Столбец Значимость F – значение уровня значимости, соответствующее вычисленной величине f-критерия и равное вероятности $P(F(k_1, k_2) \geq F_c)$, где $F(k_1, k_2)$ – случайная величина, подчиняющаяся распределению Фишера с k_1, k_2 степенями свободы. Если вероятность меньше уровня значимости α (обычно $\alpha = 0,05$), то построенная регрессия является значимой.

Перейдем к следующей группе показателей, объединенных в таблице 8.2.

Таблица 8.2 – Продолжение результатов работы режима Регрессия

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>
Y-пересечение	-3,539	1,907	-1,8564
Переменная X 1	0,854	0,221	3,8726
Переменная X 2	0,367	0,243	1,5108
	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
	0,1058	-8,0477	0,9690
	0,0061	0,3325	1,3753
	0,1746	-0,2074	0,9415

Столбец Коэффициенты – вычисленные значения коэффициентов $b_0, b_1, \dots, b_k$, расположенных сверху вниз.

Столбец Стандартная ошибка – значения S_{bj} , $j = 0, \dots, k$, вычисленные по формуле

$$S_{bj} = \sqrt{S^2 \left\{ (X^T X)^{-1} \right\}_{jj}}.$$

Столбец t-статистика - значения статистик T_{bj} .

Столбец Р-значение - содержит вероятности случайных событий $P(t(n-k) \geq T_{bj})$, где $t(n-k)$ - случайная величина, подчиняющаяся распределению Стьюдента с $n-k$ степенями свободы.

Если эта вероятность меньше уровня значимости α , то принимается гипотеза о значимости соответствующего коэффициента регрессии.

Из таблицы 8.2 видно, что значимым коэффициентом является только коэффициент при x_1 : $b_1 = 0,854$, так как Р-значение при x_1 равно 0,0061, что меньше α .

Столбцы Нижние 95 % и Верхние 95 % - соответственно нижние и верхние интервалы для оцениваемых коэффициентов β_j .

Перейдем к следующей группе показателей, объединенных в таблице 8.3.

Столбец Наблюдение - содержит номера наблюдений.

Столбец Предсказанное Y - значения $\hat{y}_i$, вычисленные по построенному уравнению регрессии.

Столбец Остатки - значения невязок $y_i - \hat{y}_i$.

Таблица 8.3 - Продолжение результатов работы режима Регрессия

ВЫВОД ОСТАТКА			
<i>Наблюдение</i>	<i>Предсказанное Y</i>	<i>Остатки</i>	<i>Стандартные остатки</i>
1	5,127	-0,127	-0,152
2	8,790	1,210	1,443
3	9,644	0,356	0,424
4	5,981	1,019	1,215
5	5,861	-0,861	-1,027
6	6,228	-0,228	-0,272
7	6,348	-0,348	-0,415
8	5,614	-0,614	-0,732
9	5,127	0,873	1,041
10	9,277	-1,277	-1,523

В заключение рассмотрения результатов работы режима Регрессия приведем график невязок - остатков $y_i - \hat{y}_i$ при заданных значениях только второй переменной (рисунок 8.3). Наличие чередующихся положительных и отрицательных значений невязок является косвенным признаком отсутствия систематической ошибки (неучтенной независимой переменной) в построенном уравнении регрессии.

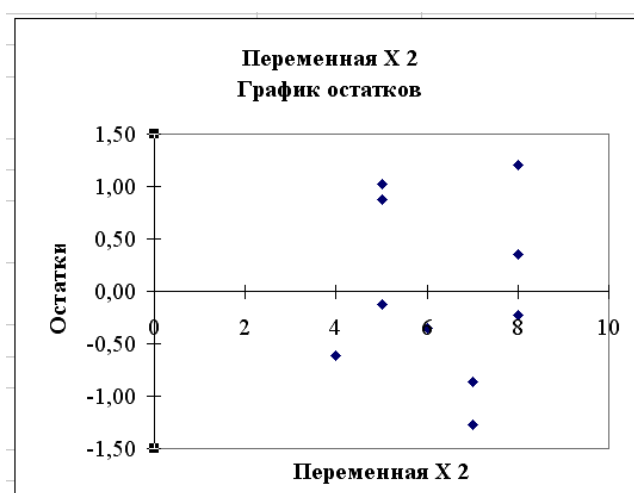


Рисунок 8.3 – График невязок как функция переменной X_2

Задание

По исходным данным, используя режим Регрессия, вычислить вектор коэффициентов уравнения регрессии.

Исходные данные

X_1	X_2	Y
8*N	5*N	5*N
11*N	8*N	10*N
12*N	8*N	10*N
9*N	5*N	7*N
8*N	7*N	5*N
8*N	8*N	6*N
9*N	6*N	6*N
9*N	4*N	5*N
8*N	5*N	6*N
12*N	7*N	8*N

где N – номер варианта студента по списку.

Содержание отчета

1. Цель занятия.
2. Краткие теоретические положения.
3. Технология выполнения работы и полученные результаты: задание; результаты измерений; формулы для получения оценок выборки и результаты расчетов.
4. Выводы.

Контрольные вопросы

1. Перечислите возможности режима Регрессия.
2. Какие параметры задаются после появления диалогового окна Регрессия?
3. Что такое Метки в диалоговом окне Регрессия?
4. Что такое Остатки в диалоговом окне Регрессия?
5. Перечислите алгоритм обработки данных, используя режим Регрессия.

Библиографический список

1. Брандт, З. Анализ данных. Статистические и вычислительные методы для научных работников и инженеров: учеб. Пособие / З. Брандт; пер. с англ. О.И. Волкова. – М.: Мир, 2003. – 686 с.
2. Вадзинский, Р. Статистические вычисления в среде Excel / Р. Вадзинский. – СПб.: Питер, 2008. – 608 с.
3. Горелова, Г.В. Теория вероятностей и математическая статистика в примерах и задачах с применением Excel: учеб. Пособие / Г.В. Горелова, И.А. Кацко. – 3-е изд., доп. и перераб. – Ростов н/Д: Феникс, 2005. – 480 с.
4. Джонсон, Н. Дион Ф. Статистика и планирование эксперимента в науке и технике. Методы планирования эксперимента: [пер. с англ.] /Н. Дион Ф. Джонсон. – М.: Мир, 1991. – 520 с.
5. Заворыкин, В.М. Численные методы / В.М. Заворыкин, В.Г. Житомирский, М.П. Лапчик. – М.: Просвещение, 1990. – 176 с.
6. Колемаев, В.А. Теория вероятностей и математическая статистика : учебник / В. А. Колемаев, В. Н. Калинина. - М.: ИНФРА-М, 2000. - 336 с.
7. Минько, А.А. Статистический анализ в MS Excel / А.А. Минько. - М. : Издательский дом «Вильямс», 2004. - 448 с.
8. Розанов, Ю.А. Теория вероятностей, случайные процессы и математическая статистика : учебник / Ю.А. Розанов. - 2-е изд., доп. - М.: Наука, 1989. - 312 с.
9. Себер, Дж. Линейный регрессионный анализ / Дж. Себер; пер. с англ. В. П. Носко. - М.: Мир, 1980. - 456 с.
10. Соболев, Б.В. Практикум по статистике в Excel / Б.В. Соболев, Л.В. Борисова, Т.А. Иваночкина, И.М. Пешхоев. – М.: Феникс, 2010. – 384 с.
11. Тюрин, Ю.Н. Анализ данных на компьютере / Ю.Н. Тюрин, А.А. Макаров; под ред. В. Э. Фигурнова. - 3-е изд., перераб. и доп. - М. : ИНФРА-М, 2003. - 544 с.

Приложение 1

Получение контрольных и обучающих выборок

Используя генератор случайных чисел пакета MathCad можно создать как обучающие, так и контрольные выборки с любым объемом и любой размерности пространства информативных признаков. Пример листа MathCad, в котором создается выборка из 28 объектов с нормально распределенным законом распределения признаков (признаковое пространство четырехмерное) показан на рисунок 1.1.

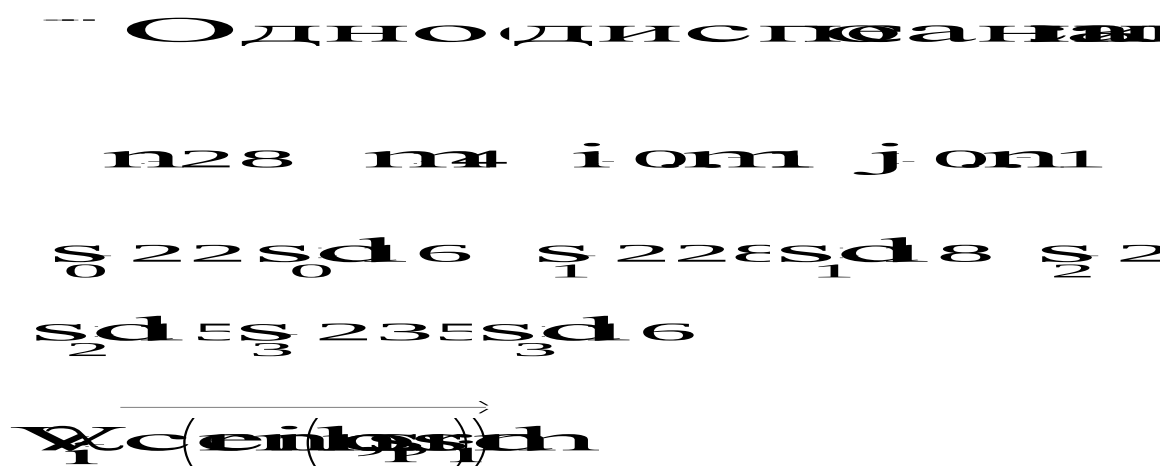


Рисунок 1.1 - Лист MathCad с генератором случайных чисел

Каждый информативный признак в выборке представлен вектором-столбцом V_{λ} из 28 элементов. Случайные числа, распределенные по нормальному закону распределения, генерируются функцией

$$rnorm(n, s, sd),$$

где n -число элементов в выборке, s – математическое ожидание информативного признака, sd – дисперсия информативного признака.

Варьируя эти параметры (s и sd) вы можете менять структуру распределения классов в признаковом пространстве и исследовать

эффективность классификации при различных классовых структурах.

Функция `ceil` округляет случайное число до ближайшего целого.

Значок $\longrightarrow$ обозначает операцию векторизации, то есть одновременное выполнение скалярной операции над всеми элементами вектора.

Пример полученной выборки в четырехмерном признаковом пространстве показан на рисунке 1.2.

$\mathbf{V}_0 =$	<table> <tr><td>0</td><td>218</td></tr> <tr><td>1</td><td>215</td></tr> <tr><td>2</td><td>218</td></tr> <tr><td>3</td><td>210</td></tr> <tr><td>4</td><td>199</td></tr> <tr><td>5</td><td>226</td></tr> <tr><td>6</td><td>224</td></tr> <tr><td>7</td><td>234</td></tr> <tr><td>8</td><td>261</td></tr> <tr><td>9</td><td>238</td></tr> <tr><td>10</td><td>241</td></tr> <tr><td>11</td><td>239</td></tr> <tr><td>12</td><td>240</td></tr> <tr><td>13</td><td>236</td></tr> <tr><td>14</td><td>209</td></tr> <tr><td>15</td><td>227</td></tr> <tr><td>16</td><td>213</td></tr> <tr><td>17</td><td>237</td></tr> <tr><td>18</td><td>223</td></tr> <tr><td>19</td><td>215</td></tr> <tr><td>20</td><td>214</td></tr> <tr><td>21</td><td>217</td></tr> <tr><td>22</td><td>234</td></tr> <tr><td>23</td><td>222</td></tr> <tr><td>24</td><td>227</td></tr> <tr><td>25</td><td>246</td></tr> <tr><td>26</td><td>214</td></tr> <tr><td>27</td><td>226</td></tr> </table>	0	218	1	215	2	218	3	210	4	199	5	226	6	224	7	234	8	261	9	238	10	241	11	239	12	240	13	236	14	209	15	227	16	213	17	237	18	223	19	215	20	214	21	217	22	234	23	222	24	227	25	246	26	214	27	226
0	218																																																								
1	215																																																								
2	218																																																								
3	210																																																								
4	199																																																								
5	226																																																								
6	224																																																								
7	234																																																								
8	261																																																								
9	238																																																								
10	241																																																								
11	239																																																								
12	240																																																								
13	236																																																								
14	209																																																								
15	227																																																								
16	213																																																								
17	237																																																								
18	223																																																								
19	215																																																								
20	214																																																								
21	217																																																								
22	234																																																								
23	222																																																								
24	227																																																								
25	246																																																								
26	214																																																								
27	226																																																								
$\mathbf{V}_1 =$	<table> <tr><td>0</td><td>248</td></tr> <tr><td>1</td><td>245</td></tr> <tr><td>2</td><td>176</td></tr> <tr><td>3</td><td>190</td></tr> <tr><td>4</td><td>232</td></tr> <tr><td>5</td><td>217</td></tr> <tr><td>6</td><td>207</td></tr> <tr><td>7</td><td>230</td></tr> <tr><td>8</td><td>242</td></tr> <tr><td>9</td><td>234</td></tr> <tr><td>10</td><td>229</td></tr> <tr><td>11</td><td>215</td></tr> <tr><td>12</td><td>221</td></tr> <tr><td>13</td><td>216</td></tr> <tr><td>14</td><td>226</td></tr> <tr><td>15</td><td>247</td></tr> <tr><td>16</td><td>232</td></tr> <tr><td>17</td><td>238</td></tr> <tr><td>18</td><td>242</td></tr> <tr><td>19</td><td>228</td></tr> <tr><td>20</td><td>234</td></tr> <tr><td>21</td><td>231</td></tr> <tr><td>22</td><td>236</td></tr> <tr><td>23</td><td>220</td></tr> <tr><td>24</td><td>216</td></tr> <tr><td>25</td><td>283</td></tr> <tr><td>26</td><td>223</td></tr> <tr><td>27</td><td>205</td></tr> </table>	0	248	1	245	2	176	3	190	4	232	5	217	6	207	7	230	8	242	9	234	10	229	11	215	12	221	13	216	14	226	15	247	16	232	17	238	18	242	19	228	20	234	21	231	22	236	23	220	24	216	25	283	26	223	27	205
0	248																																																								
1	245																																																								
2	176																																																								
3	190																																																								
4	232																																																								
5	217																																																								
6	207																																																								
7	230																																																								
8	242																																																								
9	234																																																								
10	229																																																								
11	215																																																								
12	221																																																								
13	216																																																								
14	226																																																								
15	247																																																								
16	232																																																								
17	238																																																								
18	242																																																								
19	228																																																								
20	234																																																								
21	231																																																								
22	236																																																								
23	220																																																								
24	216																																																								
25	283																																																								
26	223																																																								
27	205																																																								
$\mathbf{V}_2 =$	<table> <tr><td>0</td><td>243</td></tr> <tr><td>1</td><td>226</td></tr> <tr><td>2</td><td>213</td></tr> <tr><td>3</td><td>216</td></tr> <tr><td>4</td><td>231</td></tr> <tr><td>5</td><td>232</td></tr> <tr><td>6</td><td>235</td></tr> <tr><td>7</td><td>239</td></tr> <tr><td>8</td><td>235</td></tr> <tr><td>9</td><td>221</td></tr> <tr><td>10</td><td>207</td></tr> <tr><td>11</td><td>253</td></tr> <tr><td>12</td><td>246</td></tr> <tr><td>13</td><td>214</td></tr> <tr><td>14</td><td>234</td></tr> <tr><td>15</td><td>257</td></tr> <tr><td>16</td><td>227</td></tr> <tr><td>17</td><td>201</td></tr> <tr><td>18</td><td>210</td></tr> <tr><td>19</td><td>212</td></tr> <tr><td>20</td><td>224</td></tr> <tr><td>21</td><td>232</td></tr> <tr><td>22</td><td>223</td></tr> <tr><td>23</td><td>227</td></tr> <tr><td>24</td><td>240</td></tr> <tr><td>25</td><td>235</td></tr> <tr><td>26</td><td>238</td></tr> <tr><td>27</td><td>214</td></tr> </table>	0	243	1	226	2	213	3	216	4	231	5	232	6	235	7	239	8	235	9	221	10	207	11	253	12	246	13	214	14	234	15	257	16	227	17	201	18	210	19	212	20	224	21	232	22	223	23	227	24	240	25	235	26	238	27	214
0	243																																																								
1	226																																																								
2	213																																																								
3	216																																																								
4	231																																																								
5	232																																																								
6	235																																																								
7	239																																																								
8	235																																																								
9	221																																																								
10	207																																																								
11	253																																																								
12	246																																																								
13	214																																																								
14	234																																																								
15	257																																																								
16	227																																																								
17	201																																																								
18	210																																																								
19	212																																																								
20	224																																																								
21	232																																																								
22	223																																																								
23	227																																																								
24	240																																																								
25	235																																																								
26	238																																																								
27	214																																																								
$\mathbf{V}_3 =$	<table> <tr><td>0</td><td>218</td></tr> <tr><td>1</td><td>218</td></tr> <tr><td>2</td><td>229</td></tr> <tr><td>3</td><td>214</td></tr> <tr><td>4</td><td>190</td></tr> <tr><td>5</td><td>256</td></tr> <tr><td>6</td><td>248</td></tr> <tr><td>7</td><td>238</td></tr> <tr><td>8</td><td>219</td></tr> <tr><td>9</td><td>221</td></tr> <tr><td>10</td><td>221</td></tr> <tr><td>11</td><td>240</td></tr> <tr><td>12</td><td>229</td></tr> <tr><td>13</td><td>256</td></tr> <tr><td>14</td><td>213</td></tr> <tr><td>15</td><td>242</td></tr> <tr><td>16</td><td>243</td></tr> <tr><td>17</td><td>264</td></tr> <tr><td>18</td><td>225</td></tr> <tr><td>19</td><td>244</td></tr> <tr><td>20</td><td>260</td></tr> <tr><td>21</td><td>206</td></tr> <tr><td>22</td><td>256</td></tr> <tr><td>23</td><td>219</td></tr> <tr><td>24</td><td>243</td></tr> <tr><td>25</td><td>219</td></tr> <tr><td>26</td><td>216</td></tr> <tr><td>27</td><td>242</td></tr> </table>	0	218	1	218	2	229	3	214	4	190	5	256	6	248	7	238	8	219	9	221	10	221	11	240	12	229	13	256	14	213	15	242	16	243	17	264	18	225	19	244	20	260	21	206	22	256	23	219	24	243	25	219	26	216	27	242
0	218																																																								
1	218																																																								
2	229																																																								
3	214																																																								
4	190																																																								
5	256																																																								
6	248																																																								
7	238																																																								
8	219																																																								
9	221																																																								
10	221																																																								
11	240																																																								
12	229																																																								
13	256																																																								
14	213																																																								
15	242																																																								
16	243																																																								
17	264																																																								
18	225																																																								
19	244																																																								
20	260																																																								
21	206																																																								
22	256																																																								
23	219																																																								
24	243																																																								
25	219																																																								
26	216																																																								
27	242																																																								

Рисунок 1.2 – Пример выборки из 28 элементов в четырехмерном признаковом пространстве

Приложение 2

Критические значения t-критерия Стьюдента при уровне
значимости 0,10; 0,05; 0,01 (двухсторонний)

Число степеней свободы d.f.	α			Число степеней свободы d.f.	α		
	0,10	0,05	0,01		0,10	0,05	0,01
1	6,3138	12,706	63,657	18	1,7341	2,1009	2,8784
2	2,9200	4,3027	9,9248	19	1,7291	2,0930	2,8609
3	2,3534	3,1825	5,8409	20	1,7247	2,0860	2,8453
4	2,1318	2,7764	4,5041	21	1,7207	2,0796	2,8314
5	2,0150	2,5706	4,0321	22	1,7171	2,0739	2,8188
6	1,9432	2,4469	4,7074	23	1,7139	2,0687	2,8073
7	1,8946	2,3646	3,4995	24	1,7109	2,0639	2,7969
8	1,8595	2,3060	3,3554	25	1,7081	2,0595	2,7874
9	1,8331	2,2622	3,2498	26	1,7056	2,0555	2,7787
10	1,8125	2,2281	3,1693	27	1,7033	2,0518	2,7707
11	1,7959	2,2010	3,1058	28	1,7011	2,0484	2,7633
12	1,7823	2,1788	3,0545	29	1,6991	2,0452	2,7564
13	1,7709	2,1604	3,0123	30	1,6973	2,0423	2,7500
14	1,7613	2,1448	3,9768	40	1,6839	2,0211	2,7045
15	1,7530	2,1315	2,9467	60	1,6707	2,0003	2,6603
16	1,7459	2,1199	2,9208	120	1,6577	1,9799	2,6174
17	1,7396	2,1098	2,8982	∞	1,6449	1,9600	2,5758

Приложение 3

Таблица значений F-критерия Фишера при уровне значимости $\alpha = 0,05$

k_2	k_1									
	1	2	3	4	5	6	8	12	24	∞
1	161,5	199,5	215,7	224,6	230,2	233,9	238,9	243,9	249,0	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,41	19,45	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,74	8,64	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,91	5,77	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,68	4,53	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95	2,79	2,61	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,69	2,50	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38	2,20	2,00	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36	2,18	1,98	1,73
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,16	1,96	1,71
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29	2,12	1,91	1,65
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,09	1,89	1,62
35	4,12	3,26	2,87	2,64	2,48	2,37	2,22	2,04	1,83	1,57
40	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,00	1,79	1,51
45	4,06	3,21	2,81	2,58	2,42	2,31	2,15	1,97	1,76	1,48
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	1,95	1,74	1,44
60	4,00	3,15	2,76	2,52	2,37	2,25	2,10	1,92	1,70	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,07	1,89	1,67	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,06	1,88	1,65	1,31
90	3,95	3,10	2,71	2,47	2,32	2,20	2,04	1,86	1,64	1,28
100	3,94	3,09	2,70	2,46	2,30	2,19	2,03	1,85	1,63	1,26
125	3,92	3,07	2,68	2,44	2,29	2,17	2,01	1,83	1,60	1,21
150	3,90	3,06	2,66	2,43	2,27	2,16	2,00	1,82	1,59	1,18
200	3,89	3,04	2,65	2,42	2,26	2,14	1,98	1,80	1,57	1,14

Приложение 4

Значение статистик Дарбина-Уотсона d_l, d_u при 5%-ном уровне значимости

n	$k = 1$		$k = 2$		$k = 3$		$k = 4$		$k = 5$	
	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u
6	0,61	1,40								
7	0,70	1,36	0,47	1,90						
8	0,76	1,33	0,56	1,78	0,37	2,29				
9	0,82	1,32	0,63	1,70	0,46	2,13				
10	0,88	1,32	0,70	1,64	0,53	2,02				
11	0,93	1,32	0,66	1,60	0,60	1,93				
12	0,97	1,33	0,81	1,58	0,66	1,86				
13	1,01	1,34	0,86	1,56	0,72	1,82				
14	1,05	1,35	0,91	1,55	0,77	1,78				
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,85	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,99
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83