МИНОБРНАУКИ РОССИИ

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

Кафедра высшей математики

УТВЕРЖДАЮ
Проректор по учебной работе
__________ О.Г. Локтионова
«16» 12 __________ 2019 г.

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методические указания к лабораторным работам
для студентов направления подготовки 02.03.03 Математическое обеспечение и администрирование информационных систем

Курск 2019

УДК 519.2

Составитель: Н.А. Хохлов,

Рецензент
Кандидат технических наук, доцент *Ю.А. Халин*

Теория вероятностей и математическая статистика: методические указания к лабораторным работам для студентов направлений подготовки 02.03.03 Математическое обеспечение и администрирование информационных систем / Юго-Зап. гос. ун-т; сост.; Н.А. Хохлов – Курск, 2019. - 26 с.: - ил. 6 , табл. 4.– Библиогр.: с. 26

Содержат сведения по вопросам применения современных программных средств решения задач теории вероятностей и математической статистики.

Предназначены для студентов направления подготовки 02.03.03 Математическое обеспечение и администрирование информационных систем очной формы обучения.

Методические указания соответствуют рабочей программе дисциплины «Теория вероятностей и математическая статистика».

Текст печатается в авторской редакции

Подписано в печать 16.12. Формат 60*84 1/16.
Усл. печ. л.____. Уч.-изд. л. _____. Тираж 50 экз. Заказ 989. Бесплатно.
Юго-Западный государственный университет.
305040 Курск, ул. 50 лет Октября, 94.

ЛАБОРАТОРНАЯ РАБОТА №1
МЕТОД НАИМЕНЬШИХ КВАДРАТОВ

Цель работы: 1. Изучить основы метода наименьших квадратов.

2. Научиться решать задачу аппроксимации дискретной зависимости $y(x_i)$ непрерывной функцией $y = f(x)$ определенного класса.

3. Освоить методику применения программных продуктов MathCAD и MSExcel для построения линейной и полиномиальной зависимостей по заданным эмпирическим данным.

Задание

Методом наименьших квадратов по заданным эмпирическим данным построить

1. линейную регрессию $y = kx + b$.
2. квадратичную регрессию $y = a_2 x^2 + a_1 x + a_0$.

Студентам инженерных специальностей рекомендуется выполнять задания, используя программный продукты MathCAD, MSExcel.

1. Теоретические сведения

Метод наименьших квадратов (МНК) – один из наиболее часто используемых методов при обработке эмпирических данных, построении и анализе физических, биологических, технических, экономических и социальных моделей[1].

С помощью МНК решают задачу выбора параметров функции (заранее заданного вида) для приближённого описания зависимости величины *y* от величины *x*.

Исходные данные могут носить самый разнообразный характер и относиться к различным отраслям науки или техники, например:

---

[1] Впервые МНК был предложен К. Гауссом и А. Лежандром на рубеже 18-19 веков. Первоначально МНК использовался для обработки результатов астрономических и геодезических наблюдений. Строгое математическое обоснование и установление границ содержательной применимости МНК даны А. А. Марковым и А. Н. Колмогоровым.

- ✓ зависимость продолжительности службы электрических ламп (y) от поданного на них напряжения (x);
- ✓ зависимость пробивного напряжения конденсаторов (y) от температуры окружающей среды (x);
- ✓ зависимость предела прочности стали (y) от содержания углерода (x);
- ✓ зависимость показателей безработицы (y) и инфляции (x);
- ✓ зависимость роста преступности(y),% и роста безработицы (x),%
- ✓ зависимость цен товара (y) от спроса (x) на этот товар;
- ✓ зависимость частного потребления (y) от располагаемого дохода (x);
- ✓ зависимость температура воздуха (y) от высоты над уровнем моря (x) и другие зависимости.

Пусть необходимо установить функциональную зависимость между двумя эмпирическими данными x и y, значения которых занесены в следующую таблицу:

x	x_1	x_2	…	x_i	…	x_n
y	y_1	y_2	…	y_i	…	y_n

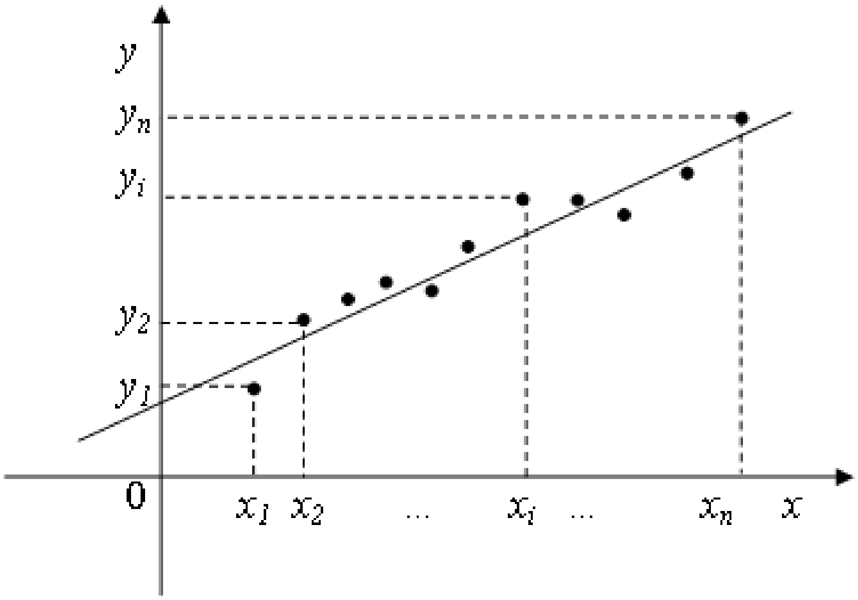
Точки $(x_i; y_i)$ координатной плоскости принято называть *экспериментальными*.

Установим вид функции $y = f(x)$ по характеру расположения на координатной плоскости экспериментальных точек.

Если точки расположены так, как показано на рис.1, то разумно предположить, что между x и y существует линейная зависимость, выражающаяся формулой:

$$y = kx + b. \tag{1}$$

Рассмотрим случай такой зависимости.



Уравнение (1) можно представить в виде

$$y-(kx+b)=0.$$

Так как точки $(x_1;y_1)$, $(x_2;y_2)$, …, $(x_n;y_n)$ не обязательно лежат на одной прямой, то, подставляя вместо x и y значения координат этих точек в выражение $y-(kx+b)$, получаем равенства:

$$y_1-(kx_1+b)=\delta_1,\ \ y_2-(kx_2+b)=\delta_2,\ \ldots,\ y_n-(kx_n+b)=\delta_n,$$

где δ_1, δ_2, …, δ_n – некоторые числа, которые называют *погрешностями* (*отклонениями*, *невязками*).

Понятно, что чем меньше эти погрешности по абсолютной величине, тем лучше прямая, задаваемая уравнением $y=kx+b$, описывает зависимость между экспериментально полученными значениями x и y.

Сущность метода наименьших квадратов заключается в подборе коэффициентов k и b таким образом, чтобы сумма квадратов погрешностей была как можно меньшей:

$$S=\delta_1^2+\delta_2^2+\ldots+\delta_n^2=\sum_{i=1}^{n}\delta_i^2=\sum_{i=1}^{n}(y_i-(kx_i+b))^2\to\min \qquad (2)$$

Отметим, что в равенстве (2) находится сумма именно квадратов погрешностей, так как в случае суммирования самих погрешностей δ_i сумма может оказаться малой за счет разных знаков погрешностей.

Так как в равенстве (2) x_i и y_i – заданные числа, а k и b – неизвестные, то сумму S можно рассмотреть как функцию двух переменных k и b: $S=S(k,b)$. Исследуем ее на экстремум:

Необходимое условие существования экстремума функции двух переменных:

$$\begin{cases} \dfrac{\partial S}{\partial k} = 0, \\ \dfrac{\partial S}{\partial b} = 0; \end{cases}$$

$$\frac{\partial S}{\partial k} = 2\sum_{i=1}^{n}(y_i - (kx_i + b))(-x_i) = -2\sum_{i=1}^{n}(y_i - (kx_i + b))\,x_i,$$

$$\frac{\partial S}{\partial b} = 2\sum_{i=1}^{n}(y_i - (kx_i + b))(-1) = -2\sum_{i=1}^{n}(y_i - (kx_i + b)).$$

Приравнивая эти частные производные к нулю, получаем линейную систему двух уравнений с двумя переменными k и b:

$$\begin{cases} -2\sum\limits_{i=1}^{n}(y_i - (kx_i + b))\,x_i = 0, \\ -2\sum\limits_{i=1}^{n}(y_i - (kx_i + b)) = 0. \end{cases}$$

Преобразуя первое уравнение системы, получим

$$-\sum_{i=1}^{n} y_i x_i + k\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = 0.$$

Преобразуя второе уравнение системы, получим

$$-\sum_{i=1}^{n} y_i + k\sum_{i=1}^{n} x_i + bn = 0.$$

Откуда имеем систему:

$$\begin{cases} k\sum\limits_{i=1}^{n} x_i^2 + b\sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} y_i x_i, \\ k\sum\limits_{i=1}^{n} x_i + bn = \sum\limits_{i=1}^{n} y_i. \end{cases} \tag{3}$$

Система (3) называется *нормальной системой*.

Из этой системы находим k и b, которые затем подставляем в уравнение (1) и получаем искомое уравнение прямой.

Тот факт, что функция $S=S(k,b)$ в найденной точке (k,b) имеет именно минимум, устанавливается с помощью частных производных второго порядка.

$$\frac{\partial^2 S}{\partial k^2}=-2\sum_{i=1}^{n}(-x_i)\,x_i=2\sum_{i=1}^{n}(x_i)^2,$$

$$\frac{\partial^2 S}{\partial b^2}=-2\cdot\sum_{i=1}^{n}(-1)=2n,$$

$$\frac{\partial^2 S}{\partial k\,\partial b}=-2\sum_{i=1}^{n}(-x_i)=2\sum_{i=1}^{n}x_i\,.$$

Вычислим $\Delta=\dfrac{\partial^2 S}{\partial k^2}\cdot\dfrac{\partial^2 S}{\partial b^2}-\left(\dfrac{\partial^2 S}{\partial k\,\partial b}\right)^2.$

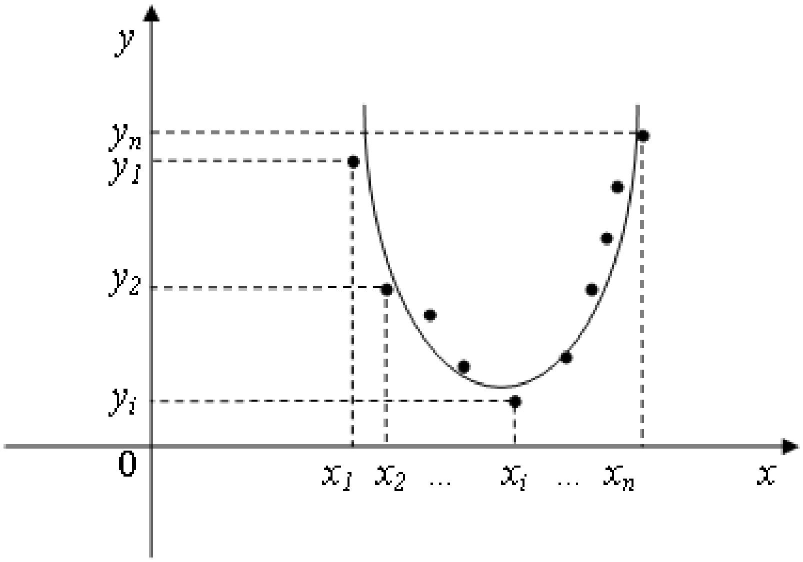
$$\Delta=4n\sum_{i=1}^{n}(x_i)^2-\left(2\sum_{i=1}^{n}x_i\right)^2=2\sum_{i=1}^{n}\sum_{j=1}^{n}(x_i-x_j)^2.$$ [2]

Очевидно, $\Delta>0$, следовательно, в найденной точке (k,b) функция $S=S(k,b)$ имеет экстремум; а так как $\dfrac{\partial^2 S}{\partial k^2}>0$, то, согласно достаточному условию экстремума функции двух переменных, в точке (k,b) функция имеет минимум.

Полученная функция $y=kx+b$ называется *линейной регрессией*, а коэффициенты k и b – *коэффициентами регрессии* (величины y на x).

Зависимость между экспериментально полученными величинами может быть близка к квадратичной (рис.2). В этом случае задача состоит в нахождении коэффициентов a_2, a_1, a_0 для составления уравнения вида $y=a_2x^2+a_1x+a_0$.

[2] Последнее равенство читатель может установить самостоятельно, воспользовавшись неравенством Коши-Буняковского.



Можно доказать, что для определения коэффициентов a_2, a_1, a_0 следует решить систему уравнений:

$$\begin{cases} n a_0 + a_1 \sum_{i=1}^{n} x_i + a_2 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i, \\ a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 + a_2 \sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} x_i y_i, \\ a_0 \sum_{i=1}^{n} x_i^2 + a_1 \sum_{i=1}^{n} x_i^3 + a_2 \sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} x_i^2 y_i. \end{cases}$$

В экспериментальной практике в качестве приближающих функций, помимо линейной $y = kx + b$ и квадратичной $y = a_2 x^2 + a_1 x + a_0$, в зависимости от характера точечного графика часто используются следующие приближающие функции:

$$y = ax^m, \quad y = ae^{mx}, \quad y = \frac{1}{ax+b}, \quad y = \frac{a}{x} + b, \quad y = \frac{x}{ax+b}, \quad y = a\ln x + b.$$

Очевидно, что когда вид приближающей функции установлен, задача сводится только к отысканию значений параметров.

Пример

Д.И. Менделеев в труде «Основы химии» приводит данные растворимости y натриевой селитры $NaNO_3$ на 100 г воды в зависимости от температуры t^0:

t_i^0	0	4	10	15	21	29	35	51	68
y_i	66,7	71,0	76,3	80,6	85,7	92,9	99,4	113,6	125,1

Соответствующая зависимость может быть представлена линейной функцией $y = kt + b$.

Требуется найти аппроксимирующую (приближаемую) функцию в предположении, что она является линейной.

Найдем коэффициенты k и b.

Для этого составим и решим нормальную систему уравнений

$$\begin{cases} k\sum_{i=1}^{n} t_i^2 + b\sum_{i=1}^{n} t_i = \sum_{i=1}^{n} y_i t_i, \\ k\sum_{i=1}^{n} t_i + bn = \sum_{i=1}^{n} y_i. \end{cases}$$

n – число эмпирических точек, n = 9.

Выполним предварительные расчеты и для удобства занесем их в таблицу (столбцы t_i, y_i, t_i^2, $t_i y_i$)

№	t_i	y_i	t_i^2	$t_i y_i$	$y_{рас.i} = kt_i + b_i$	δ_i	δ_i^2
1	0	66,7	0	0	67,55	-0,85	0,7225
2	4	71,0	16	284	71,03	-0,03	0,0009
3	10	76,3	100	763	76,25	0,05	0,0025
4	15	80,6	225	1209	80,6	0	0
5	21	85,7	441	1799,7	85,82	-0,12	0,0144
6	29	92,9	841	2694,1	92,78	0,12	0,0144
7	35	99,4	1225	3479	98	1,4	1,96
8	51	113,6	2601	5793,6	111,92	1,68	2,8224
9	68	125,1	4624	8506,8	126,71	-1,61	2,5921
$\sum$	**233**	**811,3**	**10073**	**24529,2**			**8,19**

Таким образом, нормальная система принимает вид

$$\begin{cases} k \cdot 10073 + b \cdot 233 = 24529{,}2 \\ k \cdot 233 + b \cdot 9 = 811{,}3. \end{cases}$$

Решая систему, находим

$$k \approx 0{,}87$$

$$b \approx 67{,}55$$

Следовательно, уравнение искомой прямой

$$y = 0{,}87t + 67{,}55$$

Вычислим теперь для исходных значений t_i расчетные значения $y_{рас.i} = kt_i + b_i$ и занесем полученные результаты в таблицу (столбец $y_{рас.i} = kt_i + b_i$)

Найдем $\delta_i = y_i - (kx_i + b)$ и занесем результаты в таблицу (столбец δ_i).

Вычислим сумму квадратов отклонений

$$S = \sum_{i=1}^{n} \delta_i^2 \approx 8{,}19.$$

Задание

<table>
<tr><th>Вариант</th><th>З а д а н и е</th></tr>
<tr><td>1</td><td>В таблице приведены данные численности занятого населения (x, млн.) и валового выпуска продукции (y, у.е.).

x_i	80	82	83	84	85	86	88	89	90	91
y_i	32	34	35	36	36	37	38	40	39	40

В предположении, что между x и y существует линейная зависимость, определить параметры линейной регрессии $y = kx + b$ методом наименьших квадратов. Спрогнозировать валовой выпуск продукции в случае, если занятое население увеличится на 10% по сравнению с последними данными (90 млн.)</td></tr>
<tr><td>2</td><td>В таблице приведены данные об уровне безработицы (x) и уровне преступности (y) в некотором населенном пункте.

x_i	0,5	1,2	2	3,1	4	5,2	5,9	6,1	6,2	6,3
y_i	4,25	4,32	4,4	4,51	4,6	4,72	4,79	4,9	5,0	5,2

</td></tr>
</table>

В предположении, что между x и y существует линейная зависимость, определить параметры линейной регрессии $y = kx + b$ методом наименьших квадратов. Спрогнозировать уровень преступности в случае, когда безработица отсутствует.

3

В таблице приведены данные о динамике темпов прироста курса акций (y, в %) за определенный период (t – одна неделя).

t_i	1	2	3	4	5	6	7	8	9	10
y_i	10,2	8,3	5,4	4,1	2,2	0	-1,6	-3,9	-5,9	-7,8

В предположении, что между t и y существует линейная зависимость, определить параметры линейной регрессии $y = kt + b$ методом наименьших квадратов. Сделать выводы о возможной динамике темпов прироста на 12 неделе.

4

Торговое предприятие имеет сеть, состоящую из 10 магазинов, информация о деятельности которых: годовой товарооборот (y, млн. руб.) и торговая площадь (x, тыс. $м^2$) представлена в таблице.

x_i	0,24	0,41	0,55	0,58	0,78	0,94	0,98	1,21	1,28	1,32
y_i	19,8	38,1	41,0	43,1	56,3	68,5	75,0	89,1	91,1	91,3

В предположении, что между x и y существует линейная зависимость, определить параметры линейной регрессии $y = kx + b$ методом наименьших квадратов. Спрогнозировать годовой товарооборот в случае, если торговая площадь составит ровно 1 тыс. $м^2$.

5

Показатели по объему производства (x, у.е.) и затратам (y, тыс. руб.), взятые из отчетной ведомости предприятия за 10 месяцев, приведены в таблице.

x_i	2,32	2,33	2,38	2,41	2,44	2,48	2,51	2,55	2,58	2,60
y_i	427	430	440	444	448	455	460	462	465	466

Полагая, что зависимость между x и y задается формулой $y = kx + b$, где b – постоянные затраты в тыс. руб., k – переменные затраты на 1 условную единицу продукции, определить параметры k и b методом наименьших квадратов. Рассчитать возможные затраты на производство в случае, если

<table>
<tr><td></td><td>объем производства достигнет 3 у.е.</td></tr>
<tr><td>6</td><td>В таблице приведена динамика валового выпуска (y, у.е.) за последние 10 лет (x – год)
<table>
<tr><td>x_i</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td></tr>
<tr><td>y_i</td><td>178</td><td>182</td><td>190</td><td>199</td><td>200</td><td>213</td><td>220</td><td>231</td><td>235</td><td>242</td></tr>
</table>
Предполагая линейную зависимость валового выпуска от времени, определить параметры линейной регрессии $y = kx + b$, используя метод наименьших квадратов. Получить прогноз валового выпуска на следующий год.</td></tr>
<tr><td>7</td><td>Показатели стоимости основных производственных фондов (x, млн. руб.) и среднесуточной производительности (y, тонны) приведены в таблице.
<table>
<tr><td>x_i</td><td>2,1</td><td>2,3</td><td>2,4</td><td>2,9</td><td>4,1</td><td>4,7</td><td>5,5</td><td>7,2</td><td>10,2</td><td>14,3</td></tr>
<tr><td>y_i</td><td>27</td><td>29</td><td>30</td><td>35</td><td>36</td><td>44</td><td>47</td><td>55</td><td>63</td><td>73</td></tr>
</table>
Предполагая линейную зависимость y от x, определить параметры линейной регрессии $y = kx + b$, используя метод наименьших квадратов. Получить прогноз среднесуточной производительности при стоимости основных производственных фондов 16 млн. руб.</td></tr>
<tr><td>8</td><td>В таблице приведены данные о количестве пропусков занятий (x) студентом в течение учебного семестра и результатах (y, %) написания экзаменационного теста.
<table>
<tr><td>x_i</td><td>1</td><td>3</td><td>5</td><td>6</td><td>8</td><td>10</td><td>12</td><td>14</td><td>15</td><td>16</td></tr>
<tr><td>y_i</td><td>85</td><td>75</td><td>70</td><td>60</td><td>50</td><td>40</td><td>20</td><td>10</td><td>10</td><td>5</td></tr>
</table>
Предполагая наличие линейной зависимости между x и y определить параметры линейной регрессии $y = kx + b$, используя метод наименьших квадратов. Получить прогноз результатов теста при отсутствии пропусков.</td></tr>
<tr><td>9</td><td>В таблице приведены данные об объемах производства (x, у.е.) некоторой компании в течение 10 месяцев и соответствующей операционной прибылью (y,тыс. руб.).</td></tr>
</table>

x_i	500	520	523	530	550	555	560	562	565	570
y_i	61	66,8	67	69	74	76,7	78	79	79,3	81

В предположении, что между x и y существует линейная зависимость, определить параметры линейной регрессии $y = kx + b$ методом наименьших квадратов. Сделать выводы о возможной месячной прибыли, если объем производства достигнет 600 у.е.

10

В таблице приведены данные об уровне безработицы (x) и уровне преступности (y) в некотором населенном пункте.

x_i	0,6	1,3	2,2	3,3	4,2	5,3	6,0	6,3	6,4	6,5
y_i	4,2	4,27	4,32	4,47	4,53	4,68	4,85	5,01	5,15	5,22

В предположении, что между x и y существует линейная зависимость, определить параметры линейной регрессии $y = kx + b$ методом наименьших квадратов. Спрогнозировать уровень преступности в случае, когда безработица отсутствует.

11

В таблице приведены данные численности занятого населения (x, млн.) и валового выпуска продукции (y, у.е.).

x_i	70	73	74	75	76	77	79	80	81	83
y_i	219	241	250	264	265	272	281	291	309	320

В предположении, что между x и y существует линейная зависимость, определить параметры линейной регрессии $y = kx + b$ методом наименьших квадратов. Спрогнозировать валовой выпуск продукции в случае, если занятое население увеличится на 10% по сравнению с начальными данными (80 млн.)

12

Показатели по объему производства (x, у.е.) и затратам (y, тыс. руб.), взятые из отчетной ведомости предприятия за 10 месяцев, приведены в таблице.

x_i	4,25	4,3	4,4	4,42	4,45	4,5	4,53	4,55	4,6	4,62
y_i	530	540	553	554	557	560	565	568	571	572

Полагая, что зависимость между x и y задается формулой $y = kx + b$, где b – постоянные затраты в тыс. руб., k – переменные затраты на 1 условную единицу продукции, определить параметры k и b методом наименьших квадратов. Рассчитать возможные затраты на производство в случае, если объем производства достигнет 3 у.е.

13

В таблице приведена сведения об объеме спроса (y, у.е.) на некоторую продукцию и цены на эту продукцию (x, тыс. руб.).

x_i	10	10,6	11	12	12,5	12,8	13	13,2	13,3	13,7
y_i	68	64	59	52	45	42	38	37	35	34

Предполагая линейную зависимость объема спроса от цены на продукцию, определить параметры линейной регрессии $y = kx + b$, используя метод наименьших квадратов. Получить прогноз объема спроса в случае, если цена на продукцию достигнет 14 тыс. руб.

14

Показатели стоимости основных производственных фондов (x, млн. руб.) и среднесуточной производительности (y, тонны) приведены в таблице.

x_i	2,6	2,8	2,9	3,4	4,6	5,2	6,1	7,7	10,6	14,0
y_i	19	18	20	23	26	31	37	45	53	68

Предполагая линейную зависимость y от x, определить параметры линейной регрессии $y = kx + b$, используя метод наименьших квадратов. Получить прогноз среднесуточной производительности при стоимости основных производственных фондов 2 млн. руб.

15

Торговое предприятие имеет сеть, состоящую из 10 магазинов, информация о деятельности которых: годовой товарооборот (y, млн. руб.) и торговая площадь (x, тыс. м^2) представлена в таблице.

x_i	0,25	0,42	0,57	0,59	0,79	0,95	0,99	1,23	1,29	1,33
y_i	21,9	40,1	43,2	44,3	58,3	70,6	77,2	91,2	93,2	93,4

	В предположении, что между x и y существует линейная зависимость, определить параметры линейной регрессии $y = kx + b$ методом наименьших квадратов. Спрогнозировать годовой товарооборот в случае, если торговая площадь составит ровно 1 тыс. м^2.

ЛАБОРАТОРНАЯ РАБОТА №2
ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ СЛУЧАЙНЫХ ВЕЛИЧИН

Задание 1

По заданному в нижеследующих задачах статистическому ряду выборки найти числовые характеристики:

а) выборочную дисперсию;

б) выборочное среднее квадратическое отклонение;

в) размах выборки;

г) асимметрию;

д) эксцесс.

Замечание. Числовые характеристики пунктов г) и д) можно найти с помощью программных продуктов MATHCAD или Excel.

1. Имеются следующие данные об уровне энерговооруженности труда (кВт): 50; 52; 50; 52; 52; 50 + N; 60 – N; 60; 63; 60; 50 + N; 55; 55; 54; 54; 54; 60 – N; 63; 63; 55; 60 – N; 60 – N; 50; 50 + N; 55; 55; 50; 54; 52; 52. Найти среднюю энерговооруженность труда.
2. Имеются следующие данные о себестоимости одной единицы продукции (тыс. руб.): 13; 13; 12; 11; 12; 12; 10; 9; 9; 8 + N; 10; 10; 8; 12; 9 + N; 8 + N; 8 + N; 9 + N; 8; 8; 8; 9 + N; 9; 9; 8; 11; 11; 11; 13; 13. Найти среднюю себестоимость одной единицы продукции.
3. Имеются данные по заводам за отчетный период о среднегодовой стоимости основных промышленно – производственных фондов (млн. руб.): 100; 130; 150; 140; 100 + 10N; 100; 100 + 10N; 100; 120; 110; 120; 100 +10N; 160; 160; 100; 100; 130; 130; 130; 150; 150; 140; 140; 150; 140; 160; 110; 120; 110; 120. . Найти среднегодовую стоимость основных промышленно – производственных фондов по всем заводам.
4. Имеются следующие данные по заводам за отчетный период о фактическом выпуске продукции (млн. руб.): 130; 140; 130; 140; 130; 140; 150; 200 – 10N; 180; 180; 200 – 10N; 180; 170; 170; 130; 170; 170; 120; 150; 120; 150; 180; 120; 140; 110; 180; 120; 110; 100; 200 – 10N. Найти средний фактический выпуск продукции заводами.

5. Имеются данные по группе предприятий об основных производственных фондах (млн. руб.): 3; 4; 5; 8; N + 5; 10; 7; 6; 5; 4; N + 5; 10; N + 5; 11; N + 5; 3; 3; 7; 7; 10; 11; 11; 11; 4; 5; 5; 4; 3; 8; 8. Найти среднее значение основных производственных фондов по всей группе.
6. Имеются данные по группе предприятий о валовой продукции (млн. руб.): 3; 5; 10; N + 6; 6; 4; 7; N + 7; 8; 8; 3; 5; 10; 6; 6; N+6; N+7; 10; 3; 5; 5; 4; 4; 6; 7; 10; 3; 3; N+6; N+7. Найти среднее значение выпускаемой валовой продукции.
7. Имеются данные о росте производительности труда предприятия (прирост в процентах):7; N; 8; 4; 8; 4; 8; 4; 5; 7; 5; 8;6; 6; 6; 3; 5; 5; 5; N; 6; 9; 3; 5; 3; 4; 3; 3; 3; 7. Найти средний рост производительности труда на предприятиях.
8. Имеются данные о росте фондовооруженности предприятия (прирост в процентах): 5; 7; 9; 10; 8; 6; 4; N + 2; N + 1; 7; 9; 5; 5; 7; 6; N+1; N+2; N+2; 6; 6; 6; 7; 5; 5; 4; 4; 9; 9; 9; 4. Найти среднее значение роста фондовооруженности предприятия.
9. Имеются следующие данные по предприятиям о выпуске готовой продукции на одного рабочего (тыс. руб.): 3; 6; 4; 6; 4; 8; 6; N – 1; N – 1; 5; 5; 7; 8; 10; 8; 6; 6; 6; 4; 3; 3; 8; 8; 10; 3; 3; 8; 10; 10; 10. Найти средний выпуск готовой продукции.
10. Имеются данные по предприятиям об электровооруженности труда на одного работающего (кВт – ч): 2; N + 4; 3; 7; 2; 6; 4; 10 – N; 8; 4; 6; 7; 7; 8; 8; 2; N+4; 10 – N; 10 – N; 7; 7; 7; 6; 3; 3; 3; 3; N+4; N+4; 2. Найти среднее значение электровооруженности труда на одного работающего.
11. Имеются данные о продаже товаров по ряду товарных групп за год (млн. руб.): 3,8; 2,4; 2,7; 2,6; 2,6; 2,5 + 0,1N; 2,5 + 0,1N; 2,3; 2,2; 2,3; 2,5; 2,6; 2,2; 2,0; 2,1; 3,8; 2,1; 2,1; 2,0; 2,0; 2,0; 3,8; 2,2; 2,4; 2,2; 2,1; 3,8; 2,4; 2,3; 2,3. Найти среднее значение проданных товаров.
12. Имеются данные о тарифных разрядах рабочих на предприятии: 3; N; 3; N; 3; 10 – N; 3; 3; 3; 3; 4; 5; 4; 5; 4; 5; 6; 10 – N; 6; 10 – N; 6; N; 5; 4; 5; 6; 10 – N; N; N; 3. Найти средний тарифный разряд на данном предприятии.
13. Имеются данные об основных производственных фондах ряда заводов (млрд. руб.): N + 1,4; N + 1,4; 4,8; 9,0; 7,8; 5,0; 5,5; 4,0; 6,4;

3,4; 4,0; 9,4; 3,2; 5,6; 9,8; 9,0; 7,8; 10,6; 3,4; 4,0; 4,8; 5,0; 5,5; 6,4; 9,4; 3,2; 5,6; 9,8; 3,4; 10,6. Найти среднее значение основных производственных фондов.

14. Имеются данные по группе предприятий о фактическом выпуске продукции (млрд. руб.): 7,4; 5,8; 5,6; 3,6; 5,0; 9,0; 4,6; 6,4; 3,0; 6,4; 8,6; N+2,6; 6,8; 5,0; 7,2; 7,8; 7,; 9,0; 3,8; 4,4; N + 2,6; 5,0; 3,0; 5,8; 4,6; 6,4; 9,0; 8,6; 7,4; 3,6. Найти среднее значение фактически выпущенной продукции.

15.По ряду партий деталей, обработанных рабочими производственного участка, имеются следующие данные о количестве операций, выполняемых при обработке детали: N + 1; 3; 3; 4; 5; 5; 6; 8; 11; 12; 14; 20; N + 1; 8; 8; 11; 11; 14; 14; 14; 12; 14; 20; 20; 3; 4; 4; 5; 12; N+1.Найти среднее значение количества операций.

ЛАБОРАТОРНАЯ РАБОТА № 3
ПРОВЕРКА СТАСТИСТИЧЕСКИХ ГИПОТЕЗ

Цель работы: 1. Научиться строить доверительные интервалы для математического ожидания, дисперсии и среднего квадратического отклонения.

2. Научиться проверять гипотезы о нормальном законе распределения, о равенстве средних и дисперсий, применяя пакет прикладных программ EXCEL.

Задание

1. Постройте доверительные интервалы для математического ожидания, дисперсии и среднего квадратического отклонения.
2. Проверьте гипотезу о нормальном законе распределения для изучаемой выборочной совокупности.
3. Разбейте исходные данные на две равные части и проверьте гипотезы о равенстве средних и дисперсий.

Используем результаты выполнения лабораторной работы №1 в качестве исходных данных берем значения столбца «Затраты на производство продукции».

1. Доверительный интервал для математического ожидания находим по формуле:

$$\bar{x} - t(P, n-1) \cdot \frac{S^*}{\sqrt{n}} < M[x] < \bar{x} + t(P, n-1) \cdot \frac{S^*}{\sqrt{n}}.$$

Для изучаемой выборочной совокупности: объем выборки $n = 30$, среднее выборочное $\bar{x} = 36{,}65037$, среднее квадратическое отклонение $S^* = 11{,}33477$.Для доверительной вероятности $P = 0{,}99$ квантиль распределения Стьюдента $t(0{,}99;29) = 2{,}76$ (см. табл. приложения или в свободной ячейке введем = СТЬЮДРАСПОБР (0,01;29)).

В ячейках A1−B4 введены эти данные, а в ячейках F4−F5 вычислены границы интервала для математического ожидания (см. рис. 1 и 2). Итак, $30{,}9387 < M[x] < 42{,}3620$.

Доверительные интервалы для дисперсии и среднего квадратического отклонения находим по формулам:

$$\gamma_1^2 \cdot S^{*2} < \sigma^2 < \gamma_2^2 \cdot S^{*2},$$

$$\gamma_1 \cdot S^* < \sigma < \gamma_2 \cdot S^*.$$

Для доверительной вероятности $P = 0{,}99$ и числа степеней свободы $f = 29$ находим табличные значения $\gamma_1^2 = 0{,}554$ и $\gamma_2^2 = 2{,}21$ (см.табл. приложения).

В ячейках A5−B9 введены необходимые данные, а в ячейках G4−H5 определены границы интервалов для дисперсии и среднего квадратического отклонения (рис.1 и 2). Таким образом, $71{,}1763 < \sigma^2 < 282{,}6494$ и $8{,}4366 < \sigma < 16{,}8122$.

	A	B	C	D	E	F	G	H
1	n=	30						
2	x_c=	36,65037				Доверительные интервалы для		
3	S*=	11,33477				мат.ожид.	дисперсии	сред.кв.откл.
4	t=	2,76		начало интервала		=B2-B4*B3/КОРЕНЬ(B1)	=B5*B6	=B8*B3
5	S^{*2}=	128,477		конец интервала		=B2+B4*B3/КОРЕНЬ(B1)	=B5*B7	=B9*B3
6	Y_1^2=	0,554						
7	Y_2^2=	2,2						
8	Y_1=	=КОРЕНЬ(B6)						
9	Y_2=	=КОРЕНЬ(B7)						
10								

Рисунок 1 – Формульный шаблон расчета доверительных интервалов

	A	B	C	D	E	F	G	H	I	J
1	n=	30								
2	x_c=	36,65037				Доверительные интервалы для				
3	S*=	11,33477				мат.ожид.	дисперсии	сред.кв.откл.		
4	t=	2,76		начало интервала		30,9387	71,1763	8,4366		
5	S^{*2}=	128,477		конец интервала		42,3620	282,6494	16,8122		
6	Y_1^2=	0,554								
7	Y_2^2=	2,2								
8	Y_1=	0,7443118								
9	Y_2=	1,4832397								
10										

Рисунок 2 – Расчет доверительных интервалов

2. Проверим гипотезу о нормальном законе распределения для изучаемой выборочной совокупности. Сравнение эмпирического (интервального ряда) с теоретическим (нормальным) распределением производим согласно критерию Пирсона. Для этого рассчитаем вероятности попадания нормальной случайной величины в каждый из полученных интервалов (x_{i-1}, x_i) по формуле:

$$P_i = \Phi\left(\frac{x_i - \overline{x}}{S^*}\right) - \Phi\left(\frac{x_{i-1} - \overline{x}}{S^*}\right).$$

Зная P_i, рассчитываем теоретические частоты $m_i' = P_i \cdot n$ и значение критерия:

$$\chi^2 = \sum \frac{(m_i - m_i')^2}{m_i'}.$$

Вычисления производим в таблице (рис.3 и 4). В первом столбце указаны концы интервалов, при расчете значений функции Лапласа используем встроенную статистическую функцию НОРМРАСП(x) с параметрами: среднее = 0, стандартное отклонение = 1, интегральный =1.

	A	B	C	D	E	F	G	H
13	x_i	$(x_i-x_c)/S^*$	Ф	P_i	m_i'	m_i	m_i-m_i'	$(m_i-m_i')^2/m_i'$
14	13,628	=(A14-36,65)/11,335	=НОРМРАСП(B14;0;1;1)-0,5	-	-	-	-	-
15	=A14+7,678	=(A15-36,65)/11,335	=НОРМРАСП(B15;0;1;1)-0,5	=C15-C14	=D15*30	3	=F15-E15	=G15*G15/E15
16	=A15+7,678	=(A16-36,65)/11,335	=НОРМРАСП(B16;0;1;1)-0,5	=C16-C15	=D16*30	4	=F16-E16	=G16*G16/E16
17	=A16+7,678	=(A17-36,65)/11,335	=НОРМРАСП(B17;0;1;1)-0,5	=C17-C16	=D17*30	11	=F17-E17	=G17*G17/E17
18	=A17+7,678	=(A18-36,65)/11,335	=НОРМРАСП(B18;0;1;1)-0,5	=C18-C17	=D18*30	4	=F18-E18	=G18*G18/E18
19	=A18+7,678	=(A19-36,65)/11,335	=НОРМРАСП(B19;0;1;1)-0,5	=C19-C18	=D19*30	4	=F19-E19	=G19*G19/E19
20	=A19+7,678	=(A20-36,65)/11,335	=НОРМРАСП(B20;0;1;1)-0,5	=C20-C19	=D20*30	4	=F20-E20	=G20*G20/E20
21	сумма	-	-	=СУММ(D15:	=СУММ(E1	=СУММ(F15	=СУММ(G15:	=СУММ(H15:H20
22								

Рисунок 3 – Формульный шаблон расчета значения $\chi^2_{расч}$

	A	B	C	D	E	F	G	H	I
13	x_i	$(x_i-x_c)/S^*$	Ф	P_i	m_i'	m_i	m_i-m_i'	$(m_i-m_i')^2/m_i'$	
14	13,628	-2,0311	-0,4789	-	-	-	-	-	
15	21,306	-1,3537	-0,4121	0,0668	2,0038	3,0000	0,9962	0,4952	
16	28,984	-0,6763	-0,2506	0,1615	4,8451	4,0000	-0,8451	0,1474	
17	36,662	0,0011	0,0004	0,2510	7,5300	11,0000	3,4700	1,5990	
18	44,340	0,6784	0,2513	0,2508	7,5248	4,0000	-3,5248	1,6511	
19	52,018	1,3558	0,4124	0,1612	4,8351	4,0000	-0,8351	0,1442	
20	59,696	2,0332	0,4790	0,0666	1,9969	4,0000	2,0031	2,0093	
21	сумма	-	-	0,9579	28,7357	30,0000	1,2643	6,0463	
22									

Рисунок 4 – Расчет значения $\chi^2_{расч}$

Таким образом, $\chi^2_{расч} = 6{,}0463$.

Определяем число степеней свободы по формуле $f = \ell - r - 1$, здесь $\ell = 6$ – число интервалов, $r = 2$ – число параметров нормального распределения (математическое ожидание и среднее квадратическое отклонение). Тогда для $f = 6 - 2 - 1 = 3$ и доверительной вероятности $P = 0{,}99$ находим табличное значение $\chi^2_{табл}(0{,}99;3) = 11{,}3$ (см.табл. [3] приложения 5 или в свободной ячейке введем = ХИ2ОБР(0,01;3)).

Так как $\chi^2_{расч} < \chi^2_{табл}$, расхождение эмпирических и теоретических частот незначимое, т.е. данные по затратам на производство продукции подчиняются нормальному распределению.

3. Для технических специальностей рекомендуется в качестве второго ряда взять дополнительные значения из таблиц приложений согласно варианту. Для экономических специальностей рекомендуется разбить исходные данные.

Разобьем исходные данные (затраты на производство) на две равные части, получим два дискретных ряда, для каждого из рядов рассчитаем числовые характеристики. Для этого воспользуемся пакетом «Анализ данных», расположенном в меню «Сервис», и его надстройкой «Описательная статистика». Вывод числовых характеристик можно осуществить на этом же листе, для этого в окошке «Входной интервал» указываем диапазон ячеек первого дискретного ряда, в подзаголовке «Параметры вывода» отмечаем метками «Выходной интервал», «Итоговая статистика», «Уровень надежности». В окошке «Выходной интервал» указываем диапазон ячеек A44−B54, куда будут выведены числовые характеристики для первого дискретного ряда (рис.5). Аналогично, в ячейках D44−E54 будут выведены числовые характеристики для второго дискретного ряда.

Итак, $n_1 = 15$, $n_2 = 15$, $\overline{x_1} = 35{,}8842$, $\overline{x_2} = 37{,}4165$, $S_1^{*2} = 148{,}8942$, $S_2^{*2} = 115{,}9788$.

	A	B	C	D	E	F	G
24	x_{1i}	x_{2i}					
25	31,355	32,126					
26	21,224	43,814					
27	39,263	34,72					
28	48,304	45,087					
29	34,646	16,752					
30	23,931	27,494					
31	58,98	33,639					
32	44,876	46,802					
33	34,248	24,99					
34	26,476	36,642					
35	35,459	55,554					
36	52,114	35,402					
37	42,906	55,189					
38	30,853	31,259					
39	13,628	41,778					
40							
41							
42	*Столбец1*			*Столбец1*			
43							
44	Среднее	35,8842		Среднее	37,4165		
45	Стандартная ошибка	3,1506		Стандартная ошибка	2,7806		
46	Медиана	34,6460		Медиана	35,4020		
47	Мода	#Н/Д		Мода	#Н/Д		
48	Стандартное отклонение	12,2022		Стандартное отклонение	10,7693		
49	Дисперсия выборки	148,8942		Дисперсия выборки	115,9788		
50	Эксцесс	-0,2614		Эксцесс	-0,1776		
51	Асимметричность	0,1216		Асимметричность	0,0673		
52	Интервал	45,3520		Интервал	38,8020		
53	Минимум	13,6280		Минимум	16,7520		
54	Максимум	58,9800		Максимум	55,5540		

Лист1 / Лист4 / Лист2 / Лист3

Рисунок 5 –Дискретные ряды и вывод их числовых характеристик

Проверим гипотезу о равенстве средних. Для этого найдем

$$t_{расч} = \frac{\left|\overline{x}_1 - \overline{x}_2\right|}{\hat{S}\cdot\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}},$$

где
$$\hat{S} = \sqrt{\frac{(n_1-1)\cdot S_1^{*2} + (n_2-1)\cdot S_2^{*2}}{n_1+n_2-2}}.$$

В ячейках A60−B63, A65−B66 введены необходимые данные, в ячейках B64, B67 рассчитаны значения $\hat{S}$ и $t_{расч}$ (рис. 6 и 7).

Для доверительной вероятности $P = 0{,}99$ и числа степеней свободы $f = n_1 + n_2 - 2 = 15 + 15 - 2 = 28$ находим $t_{табл}(0{,}99;28) = 2{,}77$.

Так как $t_{расч} < t_{табл}$, то расхождение средних можно считать незначительным.

	A	B	C	D
60	n_1=	15		
61	n_2=	15		
62	$S*_1^2$=	=B49		
63	$S*_2^2$=	=E49		
64	S^=	=КОРЕНЬ(((B60-1)*B62+(B61-1)*B63)/(B60+B61-2))		
65	x_1=	=B44		
66	x_2=	=E44		
67	$t_{расч}$=	=(B66-B65)/B64/КОРЕНЬ(1/B60+1/B61)		
68	$F_{расч}$=	=B62/B63		
69				
70				

Рисунок 6 – Формульный шаблон расчета значений $t_{расч}$ и $F_{расч}$

	A	B	C	D	E	F	G	H
58								
59								
60	n_1=	15						
61	n_2=	15						
62	$S*_1^2$=	148,894						
63	$S*_2^2$=	115,979						
64	S^=	11,508						
65	x_1=	35,884						
66	x_2=	37,417						
67	$t_{расч}$=	0,365						
68	$F_{расч}$=	1,284						
69								
70								

Рисунок 7 – Расчет значений $t_{расч}$ и $F_{расч}$

Табличное значение можно найти по таблице приложения 2 или в свободной ячейке ввести = СТЬЮДРАСПОБР (0,01;28).

Проверим гипотезу о равенстве дисперсий. Для этого находим $F_{расч} = \dfrac{S_1^2}{S_2^2}$, где $S_1^2 = \max\{S_1^{*2}, S_2^{*2}\}$, $S_2^2 = \min\{S_1^{*2}, S_2^{*2}\}$.

В нашем случае $S_1^{*2} > S_2^{*2}$, следовательно, $S_1^2 = S_1^{*2}$, $S_2^2 = S_2^{*2}$. В ячейке B68 рассчитана величина $F_{расч} = 1{,}284$. Для доверительной вероятности 0,99 табличное значение $F_{табл}(0{,}99;14;14) = 3{,}63$.

Так как $F_{расч} < F_{табл}$, то расхождение дисперсий можно считать незначимым.

Табличное значение можно найти по таблице F−распределения приложения 6 или в свободной ячейке ввести = FРАСПОБР (0,01;14;14).

ЛАБОРАТОРНАЯ РАБОТА № 4
КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Для двух случайных величин X и Y проведена серия испытаний. Результаты испытаний записаны в следующую корреляционную таблицу. Четные варианты индивидуальные задания берут из таблицы 2.4, а нечетные – из таблицы 2.5.

Таблица 2.4

Индивидуальные данные к заданию 6

Y \ X	0	1	2	3	4	5
1	D	C	–	–	–	–
2	–	C	B	–	–	–
3	–	–	A	B	A	1
4	–	–	–	–	D	C

Таблица 2.5

Индивидуальные данные к заданию 6

Y \ X	0	1	2	3	4	5
1						A
2				C	B	A
3		D	3	B		
4	C	2				

Для этих случайных величин:

1. Вычислить числовые характеристики выборочные средние, выборочные дисперсии, ковариацию и выборочный коэффициент корреляции ρ_{XY}.
2. Проверить для доверительной вероятности P = 0,95 значимость коэффициента корреляции ρ_{XY}. Сделать вывод о тесноте взаимосвязи.
3. Написать уравнения прямых регрессий Y на X и X на Y.
4. В подходящем масштабе изобразить на графике точки (x, y) из корреляционной таблицы и прямые регрессии.

2.7 Задание 7

Над случайными величинами X, Y, Z проведена серия из 8 наблюдений. Результаты записаны в таблицу

Таблица 2.6

Индивидуальные данные к заданию 7

	X	Y	Z
1	1	A	0
2	0	1	A
3	2	B	3
4	C	2	3
5	3	1	1
6	2	0	-1
7	A	3	B
8	1	C	D

Составить матрицы моментов и корреляционную. Вычислить коэффициент множественной корреляции между переменной Z (как функции от X, Y) и переменными X, Y.

Список рекомендуемой литературы

1. Вентцель Е. С., Овчаров Л.А. Теория вероятностей и её инженерные приложения. М.:1986.
2. Гмурман В.Е. Теория вероятностей и математическая статистика. М.: Высш. шк., 2007.
3. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. М.: Высш. шк., 2007.
4. Пискунов Н.С. Дифференциальное и интегральное исчисления. Т.2. −М.: Интеграл-Пресс, 2003.