

Задание 1.

Имеются следующие данные о выработке литья на одного работающего X_1 (т), браке литья X_2 (%) и себестоимости 1 т литья Y (руб.) по 20 литейным цехам различных заводов:

i	1	2	3	4	5	6	7	8	9	10
x_{1i}	44,1	16,4	44,5	83,9	76,8	42,3	80,3	32,5	63,2	67,5
x_{2i}	5,9	7,5	8	1,3	8,6	6,6	3,5	6,3	3,4	7,5
y_i	228,6	270,7	231,5	111,8	198,6	262,7	147,6	239,2	157,9	226,6

i	11	12	13	14	15	16	17	18	19	20
x_{1i}	53,8	56,6	42,7	58,8	38,3	20,5	42,2	48,4	38,6	85
x_{2i}	8,7	6	3,1	3,9	3,4	2,8	7,7	2,9	5,6	3,7
y_i	213,8	222,6	143	177,2	178,5	230,5	223,9	187,4	213,3	119,7

Необходимо установить связь между себестоимостью литья и выработкой литья на одного работающего

- 1) без учёта производственного брака (найти уравнение парной регрессии Y по X_1);
- 2) и с учётом производственного брака (найти уравнение множественной регрессии Y по X_1 и X_2);
- 3) оценить значимость полученных уравнений на уровне $\alpha = 0,05$;
- 4) установить значимость коэффициента регрессии при X_2 на уровне $\alpha = 0,05$;
- 5) получить точечную оценку среднего значения себестоимости 1т литья в цехах, в которых выработка литья на одного работающего составляет 40 т, а брак литья составляет 5%.

Решение:

1) Общий вид уравнения регрессии:

$$\hat{y}_{x_1} = a + bx_1$$

Система нормальных уравнений в общем виде:

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

Приведём расчётную таблицу:

i	y	x_1	x_2	y^2	x_1^2	x_2^2	yx_1	yx_2	x_1x_2
1	228,6	44,1	5,9	52257,96	1944,81	34,81	10081,26	1348,74	260,19
2	270,7	16,4	7,5	73278,49	268,96	56,25	4439,48	2030,25	123
3	231,5	44,5	8	53592,25	1980,25	64	10301,75	1852	356
4	111,8	83,9	1,3	12499,24	7039,21	1,69	9380,02	145,34	109,07
5	198,6	76,8	8,6	39441,96	5898,24	73,96	15252,48	1707,96	660,48
6	262,7	42,3	6,6	69011,29	1789,29	43,56	11112,21	1733,82	279,18
7	147,6	80,3	3,5	21785,76	6448,09	12,25	11852,28	516,6	281,05
8	239,2	32,5	6,3	57216,64	1056,25	39,69	7774	1506,96	204,75
9	157,9	63,2	3,4	24932,41	3994,24	11,56	9979,28	536,86	214,88

10	226,6	67,5	7,5	51347,56	4556,25	56,25	15295,5	1699,5	506,25
11	213,8	53,8	8,7	45710,44	2894,44	75,69	11502,44	1860,06	468,06
12	222,6	56,6	6	49550,76	3203,56	36	12599,16	1335,6	339,6
13	143	42,7	3,1	20449	1823,29	9,61	6106,1	443,3	132,37
14	177,2	58,8	3,9	31399,84	3457,44	15,21	10419,36	691,08	229,32
15	178,5	38,3	3,4	31862,25	1466,89	11,56	6836,55	606,9	130,22
16	230,5	20,5	2,8	53130,25	420,25	7,84	4725,25	645,4	57,4
17	223,9	42,2	7,7	50131,21	1780,84	59,29	9448,58	1724,03	324,94
18	187,4	48,4	2,9	35118,76	2342,56	8,41	9070,16	543,46	140,36
19	213,3	38,6	5,6	45496,89	1489,96	31,36	8233,38	1194,48	216,16
20	119,7	85	3,7	14328,09	7225	13,69	10174,5	442,89	314,5
Сумма	3985,1	1036,4	106,4	832541,05	61079,82	662,68	194583,74	22565,23	5347,78
Среднее	199,26	51,82	5,32	41627,05	3053,99	33,13	9729,19	1128,26	267,39

Система нормальных уравнений с вычисленными коэффициентами:

$$\begin{cases} 20 \cdot a + 1036,4 \cdot b = 3985,1; \\ 1036,4 \cdot a + 61079,82 \cdot b = 194583,74 \end{cases}$$

Решение системы:

$$\begin{cases} a = 283,06; \\ b = -1,62 \end{cases}$$

Построенное уравнение регрессии:

$$\hat{y}_{x_1} = 283,06 - 1,62x_1$$

Соответственно, без учёта производственного брака при увеличении выработки литья на одного работающего на 1 т себестоимость 1 т. литья снижается в среднем на 1,62 руб.

2) Общий вид уравнения регрессии:

$$\hat{y}_{x_1x_2} = a + b_1x_1 + b_2x_2$$

Для нахождения коэффициентов необходимо решить систему:

$$\begin{cases} a \cdot n + b_1 \cdot \sum_{i=1}^n x_1 + b_2 \cdot \sum_{i=1}^n x_2 = \sum_{i=1}^n y; \\ a \cdot \sum_{i=1}^n x_1 + b_1 \cdot \sum_{i=1}^n (x_1)^2 + b_2 \cdot \sum_{i=1}^n x_1x_2 = \sum_{i=1}^n yx_1 \\ a \cdot \sum_{i=1}^n x_2 + b_1 \cdot \sum_{i=1}^n x_1x_2 + b_2 \cdot \sum_{i=1}^n (x_2)^2 = \sum_{i=1}^n yx_2 \end{cases}$$

Система уравнений с вычисленными коэффициентами:

$$\begin{cases} 20a + 1036,4b_1 + 106,4b_2 = 3985,1; \\ 1036,4a + 61079,82b_1 + 5347,78b_2 = 194583,74 \\ 106,4a + 5347,78b_1 + 662,68b_2 = 22565,23 \end{cases}$$

Решение системы:

$$\begin{cases} a = 206,52 \\ b_1 = -1,35 \\ b_2 = 11,8 \end{cases}$$

Построенное уравнение регрессии:

$$\hat{y}_{x_1x_2} = 206,52 - 1,35x_1 + 11,8x_2$$

При увеличении выработки литья на одного работающего на 1 т себестоимость и неизменном проценте брака себестоимость 1 т. литья снижается в среднем на 1,35 руб. При увеличении брака литья на 1% и неизменной выработке литья себестоимость 1 т. литья увеличивается в среднем на 11,8 руб.

3) Проведём в таблице расчёты:

i	y	x ₁	x ₂	\hat{y}_{x_1}	$\hat{y}_{x_1x_2}$	$(y - \hat{y}_{x_1})^2$	$(y - \hat{y}_{x_1x_2})^2$	$(y - \bar{y})^2$	$(x_2 - \bar{x}_2)^2$
1	228,6	44,1	5,9	211,74	216,53	284,28	145,58	861,13	0,34
2	270,7	16,4	7,5	256,53	272,86	200,67	4,65	5104,39	4,75
3	231,5	44,5	8	211,09	240,77	416,47	86,02	1039,74	7,18
4	111,8	83,9	1,3	147,38	108,46	1265,72	11,19	7648,38	16,16
5	198,6	76,8	8,6	158,86	204,19	1579,37	31,30	0,43	10,76
6	262,7	42,3	6,6	214,65	227,23	2308,78	1258,29	4025,27	1,64
7	147,6	80,3	3,5	153,20	139,28	31,35	69,19	2668,24	3,31
8	239,2	32,5	6,3	230,50	236,93	75,72	5,13	1595,60	0,96
9	157,9	63,2	3,4	180,85	161,22	526,79	11,00	1710,24	3,69
10	226,6	67,5	7,5	173,90	203,79	2777,48	520,51	747,75	4,75
11	213,8	53,8	8,7	196,05	236,46	314,95	513,65	211,56	11,42
12	222,6	56,6	6	191,53	200,82	965,65	474,45	544,99	0,46
13	143	42,7	3,1	214,00	185,39	5041,48	1796,54	3164,63	4,93
14	177,2	58,8	3,9	187,97	173,06	115,94	17,11	486,42	2,02
15	178,5	38,3	3,4	221,12	194,87	1816,36	268,08	430,77	3,69
16	230,5	20,5	2,8	249,90	211,85	376,52	347,70	976,25	6,35
17	223,9	42,2	7,7	214,81	240,34	82,59	270,38	607,38	5,66
18	187,4	48,4	2,9	204,79	175,32	302,26	145,91	140,54	5,86
19	213,3	38,6	5,6	220,63	220,43	53,78	50,82	197,26	0,08
20	119,7	85	3,7	145,60	135,29	670,71	243,02	6329,00	2,62
Сумма	3985,1	1036,4	106,4	3985,1	3985,1	19206,87	6270,52	38489,95	96,63

Коэффициент детерминации определяется по формуле:

$$R^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2}$$

Тогда для первого уравнения:

$$R^2 = 1 - \frac{19206,87}{38489,95} = 0,5$$

Для второго уравнения:

$$R^2 = 1 - \frac{6270,52}{38489,95} = 0,837$$

Для определения значимости полученного уравнения регрессии необходимо рассчитать эмпирическую величину F-критерия Фишера:

$$F_p = \frac{R^2}{1 - R^2} * \frac{n - p - 1}{p}$$

Для первого уравнения:

$$F_p = \frac{0,5}{1-0,5} * \frac{20-1-1}{1} = 18,07$$

Для второго уравнения:

$$F_p = \frac{0,837}{1-0,837} * \frac{20-2-1}{2} = 43,68$$

По таблице F-распределения Снедекора-Фишера при $\alpha = 0,05$ и $K_1 = 2$, $K_2 = 20 - 3 = 17$ величина $F_T = 3,59$. При $K_1 = 1$, $K_2 = 20 - 2 = 18$ величина $F_T = 4,41$. Это означает, что оба уравнения статистически значимы.

4) На основании данных расчётной таблицы рассчитаем остаточное среднее квадратическое отклонение:

$$S_{\text{ост}} = \sqrt{\frac{\sum (y - \hat{y}_x)^2}{n - p - 1}} = \sqrt{\frac{6270,52}{20 - 2 - 1}} = 19,21$$

Определим стандартную ошибку коэффициента:

$$m_{b_2} = S_{\text{ост}} * \sqrt{\frac{1}{\sum (x_2 - \bar{x}_2)^2}} = 19,21 * \sqrt{\frac{1}{96,63}} = 1,95$$

Рассчитаем t-статистику:

$$t_{b_2} = \frac{b_2}{m_{b_2}} = \frac{11,8}{1,95} = 6,04$$

Табличное значение t-критерия Стьюдента при $\alpha = 0,05$ и $k = 20 - 2 - 1 = 17$ будет равно $t_{\text{табл}} = 2,11$. Таким образом, коэффициент регрессии при X_2 является статистически значимым на уровне $\alpha = 0,05$.

5) Выполним прогноз:

$$\hat{y}_{x_1x_2} = 206,52 - 1,35 * 40 + 11,8 * 5 = 211,46 \text{ руб./т.}$$

Задание 2.

Имеются следующие данные о численности населения США в 1950-1985 гг. (млн. чел.)

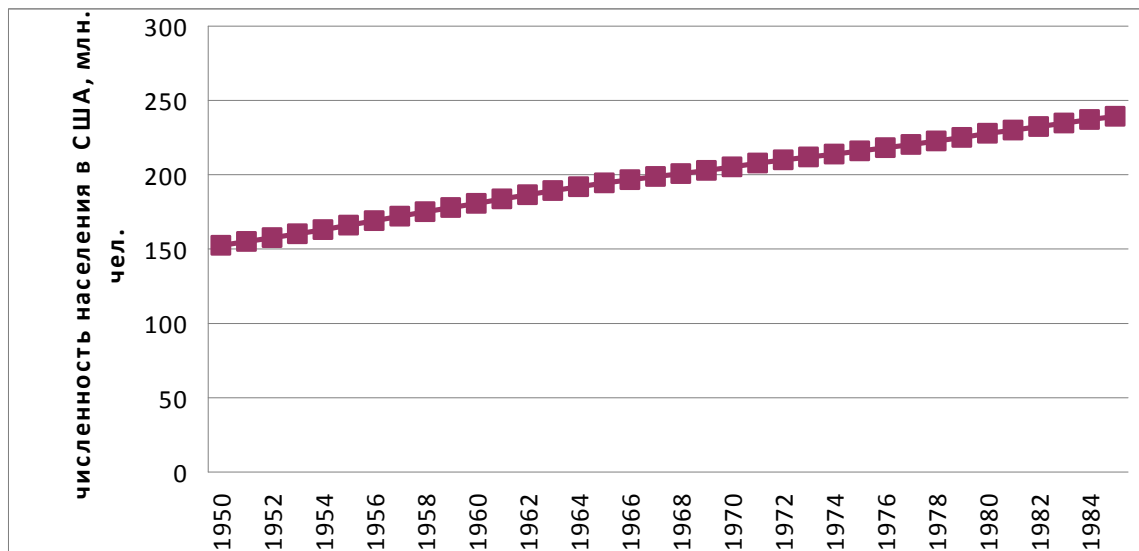
Год	1950	1951	1952	1953	1954	1955	1956	1957	1958
y_t	152,3	154,9	157,6	160,2	163,0	165,9	168,9	172,0	174,9
Год	1959	1960	1961	1962	1963	1964	1965	1966	1967
y_t	177,8	180,7	183,7	186,5	189,2	191,9	194,3	196,6	198,7
Год	1968	1969	1970	1971	1972	1973	1974	1975	1976
y_t	200,7	202,7	205,1	207,7	209,9	211,9	213,8	216,0	218,0
Год	1977	1978	1979	1980	1981	1982	1983	1984	1985
y_t	220,2	222,6	225,1	227,7	230,0	232,3	234,8	237,0	239,3

Требуется обработать эти данные, выполнив следующие действия:

- 1) представить ряд графически;
- 2) подобрать подходящее уравнение тренда по методу наименьших квадратов или подходящую скользящую среднюю, если характер тренда неясен;
- 3) удалить трендовую составляющую из временного ряда и построить график остатков;
- 4) проанализировать поведение ряда остатков.

Решение:

1) Построим график:



2) Видим, что тренд близок к линейному, поэтому построим линейное уравнение тренда.

Общий вид уравнения тренда:

$$\hat{y}_i = a + bt$$

Система нормальных уравнений в общем виде:

$$\begin{cases} a \cdot n + b \cdot \sum_{i=1}^n t_i = \sum_{i=1}^n y_i; \\ a \cdot \sum_{i=1}^n t_i + b \cdot \sum_{i=1}^n t_i^2 = \sum_{i=1}^n t_i \cdot y_i \end{cases}$$

Приведём расчётную таблицу:

Год	t	y	t ²	ty
1950	1	152,3	1	152,3
1951	2	154,9	4	309,8
1952	3	157,6	9	472,8
1953	4	160,2	16	640,8
1954	5	163	25	815
1955	6	165,9	36	995,4
1956	7	168,9	49	1182,3
1957	8	172	64	1376
1958	9	174,9	81	1574,1
1959	10	177,8	100	1778
1960	11	180,7	121	1987,7
1961	12	183,7	144	2204,4
1962	13	186,5	169	2424,5
1963	14	189,2	196	2648,8
1964	15	191,9	225	2878,5
1965	16	194,3	256	3108,8

1966	17	196,6	289	3342,2
1967	18	198,7	324	3576,6
1968	19	200,7	361	3813,3
1969	20	202,7	400	4054
1970	21	205,1	441	4307,1
1971	22	207,7	484	4569,4
1972	23	209,9	529	4827,7
1973	24	211,9	576	5085,6
1974	25	213,8	625	5345
1975	26	216	676	5616
1976	27	218	729	5886
1977	28	220,2	784	6165,6
1978	29	222,6	841	6455,4
1979	30	225,1	900	6753
1980	31	227,7	961	7058,7
1981	32	230	1024	7360
1982	33	232,3	1089	7665,9
1983	34	234,8	1156	7983,2
1984	35	237	1225	8295
1985	36	239,3	1296	8614,8
Сумма:	666	7123,9	16206	141323,7

Система нормальных уравнений с вычисленными коэффициентами:

$$\begin{cases} 36 \cdot a + 666 \cdot b = 7123,9; \\ 666 \cdot a + 16206 \cdot b = 141323,7 \end{cases}$$

Решение системы:

$$\begin{cases} a = 152,5; \\ b = 2,45 \end{cases}$$

Построенное уравнение тренда:

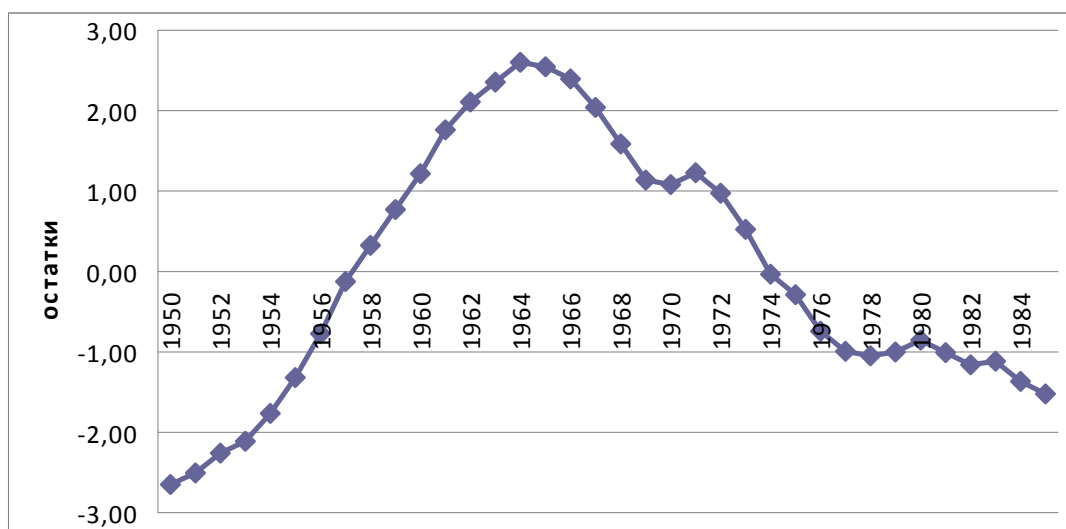
$$\hat{y}_t = 152,5 + 2,45t$$

3) Удалим трендовую составляющую из временного ряда:

Год	t	y	$\hat{y}_t = 152,5 + 2,45t$	$e = y - \hat{y}_t$
1950	1	152,3	154,95	-2,65
1951	2	154,9	157,40	-2,50
1952	3	157,6	159,86	-2,26
1953	4	160,2	162,31	-2,11
1954	5	163	164,76	-1,76
1955	6	165,9	167,22	-1,32
1956	7	168,9	169,67	-0,77
1957	8	172	172,13	-0,13
1958	9	174,9	174,58	0,32
1959	10	177,8	177,03	0,77
1960	11	180,7	179,49	1,21

1961	12	183,7	181,94	1,76
1962	13	186,5	184,39	2,11
1963	14	189,2	186,85	2,35
1964	15	191,9	189,30	2,60
1965	16	194,3	191,75	2,55
1966	17	196,6	194,21	2,39
1967	18	198,7	196,66	2,04
1968	19	200,7	199,11	1,59
1969	20	202,7	201,57	1,13
1970	21	205,1	204,02	1,08
1971	22	207,7	206,47	1,23
1972	23	209,9	208,93	0,97
1973	24	211,9	211,38	0,52
1974	25	213,8	213,83	-0,03
1975	26	216	216,29	-0,29
1976	27	218	218,74	-0,74
1977	28	220,2	221,19	-0,99
1978	29	222,6	223,65	-1,05
1979	30	225,1	226,10	-1,00
1980	31	227,7	228,55	-0,85
1981	32	230	231,01	-1,01
1982	33	232,3	233,46	-1,16
1983	34	234,8	235,91	-1,11
1984	35	237	238,37	-1,37
1985	36	239,3	240,82	-1,52
Сумма:	666	7123,9	7123,9	0

Построим график остатков:



4) Видим, что значительные положительные остатки наблюдаются в период 1967-1974 гг, когда численность населения сильно отклонялась от тренда в положительную сторону.