Использование информационных технологий в экономических моделях

Литература

1. Орлова И.В., Половников В.А. Экономико-математические методы и модели: компьютерное моделирование: Учеб.пособие. – М.: Вузовский учебник: ИНФРА-М, 2014. 369 с.
2. Орлова И.В. Экономико-математические методы и модели. Выполнение расчетов в среде EXCEL. Практикум: Учеб. пособие для вузов. – М.: Финстатинформ, 2000.
3. Эконометрика: Учебник / Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2002. – 344 с.
4. Кремер Н.Ш., Путко Б.А. Эконометрика: Учебник для вузов. – М.: ЮНИТИ-ДАНА, 2002.
5. Линейная алгебра: учебник и практикум / Н.Ш. Кремер, М.Н. Фридман. М.: Изд-во Юрайт, 2014. – 307 с.
6. Высшая математика для экономистов: учебник / под ред. Н.Ш. Кремера. М.: ЮНИТИ-ДАНА, 2010.
7. Многомерный статистический анализ в экономических задачах: компьютерное моделирование в SPSS: Учебное пособие / Под ред. И.В. Орловой. – М.: Вузовский учебник, 2009-2011.

Эконометрические модели

Переменные, применяемые в эконометрических моделях:

1. Результирующая (зависимая, эндогенная) Y – результативные признаки
2. Объясняющая (независимая, экзогенная) X – факторные признаки

Классы эконометрических моделей

1. **Модели временных рядов**: адаптивные модели, модели кривых роста (трендовые), модели авторегрессии, модели скользящего среднего.

2. **Регрессионные модели с одним уравнением:** линейные и нелинейные, однофакторные (парная регрессия) и многофакторные (множественная регрессия)

$$Y = f\left(X_1, X_2, \ldots, X_k\right)$$

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 \qquad Y = a_0K^{a_1}L^{a_2} \qquad y = \frac{a_0}{1 + a_1e^{-a_2x}}$$

3. Системы эконометрических уравнений:

системы независимых уравнений

$$
\begin{cases}
y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1, \\
y_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \varepsilon_2, \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
y_n = a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m + \varepsilon_n.
\end{cases}
$$

системы рекурсивных уравнений

$$
\begin{cases}
y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1, \\
y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \varepsilon_2, \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
y_n = b_{n1}y_1 + b_{n2}y_2 + \ldots + b_{nn-1}y_{n-1} + a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m + \varepsilon_n.
\end{cases}
$$

системы взаимосвязанных уравнений (система совместных, система одновременных уравнений, структурная форма модели)

$$\begin{cases} y_1 = b_{12}y_2 + b_{13}y_3 + \ldots + b_{1n}y_n + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1, \\ y_2 = b_{21}y_1 + b_{23}y_3 + \ldots + b_{2n}y_n + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \varepsilon_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ y_n = b_{n1}y_1 + b_{n2}y_2 + \ldots + b_{nn-1}y_{n-1} + a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m + \varepsilon_n. \end{cases}$$

Корреляционный анализ

Типы связей между признаками:

- функциональные

- корреляционные

По направлению: прямые и обратные

По форме: линейные и нелинейные

По количеству факторов, действующих на результативный признак: однофакторные (парные) и многофакторные

$$X = (x_1, x_2, \ldots, x_n) \quad Y = (y_1, y_2, \ldots, y_n)$$

Ковариация

$$\mathrm{Cov}(X, Y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n) \qquad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

Коэффициент парной корреляции

$$r_{y,x} = \frac{\mathrm{Cov}(X,Y)}{S_x S_y} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{S_x S_y} =$$

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

$$-1 \le r_{y,x} \le 1$$

Дисперсия $S_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

Среднеквадратическое отклонение (стандартное отклонение, стандартная ошибка)

$$S_x = \sqrt{S_x^2}$$

Качественная оценка коэффициента корреляции (шкала Чеддока)

0,1–0,3 — слабая;
0,3–0,5 — заметная;
0,5–0,7 — умеренная;
0,7–0,9 — высокая;
0,9–1,0 — весьма высокая.

Оценка существенности коэффициента корреляции (t-критерий Стьюдента)

$$t_{набл} = \sqrt{\frac{r_{y,x}^2}{1 - r_{y,x}^2}(n-2)} \qquad t_{набл} > t_{табл}$$

Значения t-критерия Стьюдента (двустороннего) при уровнях значимости 0,10; 0,05; 0,01

Число степеней свободы	α			Число степеней свободы	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

Матрица коэффициентов парно корреляции

$$R = \begin{pmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_m} \\ r_{yx_1} & 1 & r_{x_1x_2} & \dots & r_{x_1x_m} \\ r_{yx_2} & r_{x_1x_2} & 1 & \dots & r_{x_2x_m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{yx_m} & r_{x_1x_m} & r_{x_2x_m} & \dots & 1 \end{pmatrix}.$$

Выборочный коэффициент множественной корреляции

$$R_{j,1,2,\dots,j-1,j+1,\dots,m} = \sqrt{1 - \frac{|R|}{R_{jj}}}$$

Выборочный множественный коэффициент детерминации

$$R^2_{j,1,2,\dots,j-1,j+1,\dots,m}$$

$$R^2_{j,1,2,\dots,j-1,j+1,\dots,m}, \quad R_{j,1,2,\dots,j-1,j+1,\dots,m} \geq 0$$

Проверка значимости коэффициента детерминации (F-критерий Фишера)

$$F_{\text{расч}} = \frac{R^2/(p-1)}{(1-R^2)/(n-p)} \qquad F_{\text{расч}} > F_{\text{табл}}.$$

Выборочный частный коэффициент корреляции

$$r_{jk(1,2,\ldots,m)} = -\frac{R_{jk}}{\sqrt{R_{jj}R_{kk}}} \qquad r_{12(3)} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}$$

$$-1 \le r_{jk(1,2,\ldots,m)} \le 1$$

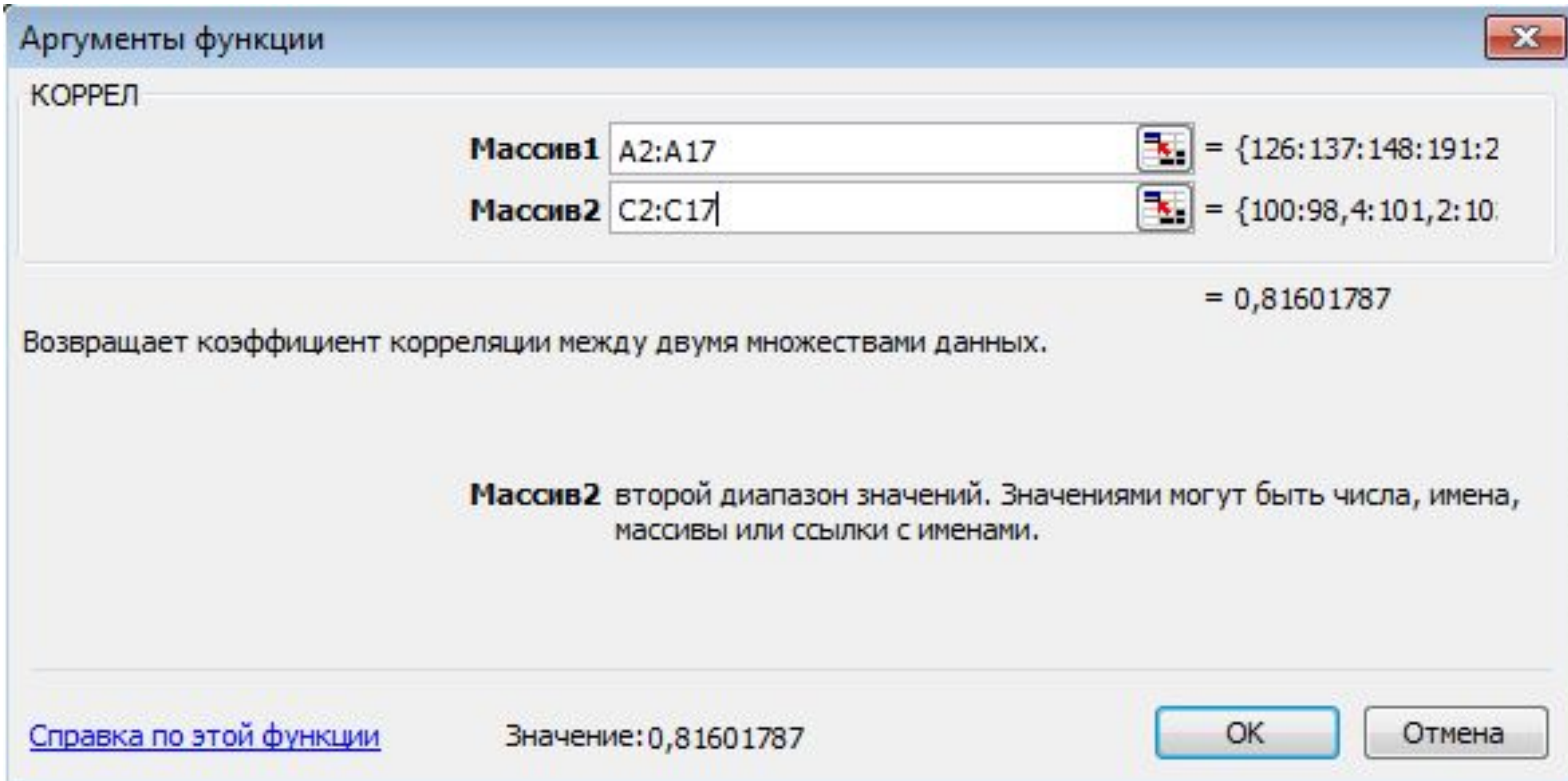
Пример. Имеются данные об объеме продаж и затратах на рекламу одной фирмы, а также индекс потребительских расходов за ряд текущих лет

	A	B	C
1	Объем продаж, Y, тыс. руб.	Затраты на рекламу X1, тыс.руб.	Индекс потребительских расходов, X2, %
2	126	4	100
3	137	4,8	98,4
4	148	3,8	101,2
5	191	8,7	103,5
6	274	8,2	104,1
7	370	9,7	107
8	432	14,7	107,4
9	445	18,7	108,5
10	367	19,8	108,3
11	367	10,6	109,2
12	321	8,6	110,1
13	307	6,5	110,7
14	331	12,6	110,3
15	345	6,5	111,8
16	364	5,8	112,3
17	384	5,7	112,9

1. Построить диаграмму рассеяния (корреляционное поле) для переменных «Объем продаж» и «Индекс потребительских расходов»

2. Определить степень влияния индекса потребительских расходов на объем продаж (вычислить коэффициент парной корреляции)

Функция КОРРЕЛ из категории Статистические



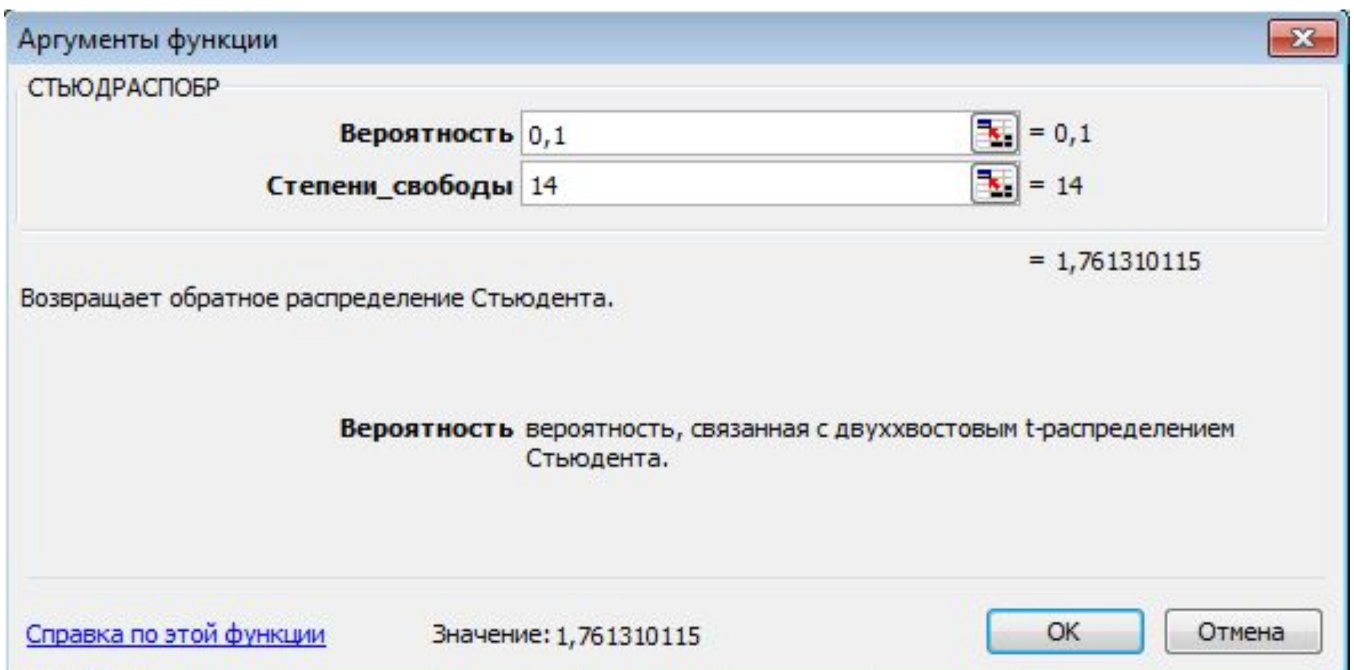
3. Оценить значимость вычисленного коэффициента парной корреляции

Фактическое значение t-критерия

$$t_{факт} = \sqrt{\frac{r^2_{y,x_2}}{1 - r^2_{y,x_2}}(n-2)} = \sqrt{\frac{0{,}816^2}{1 - 0{,}816^2}(16-2)} = 5{,}282$$

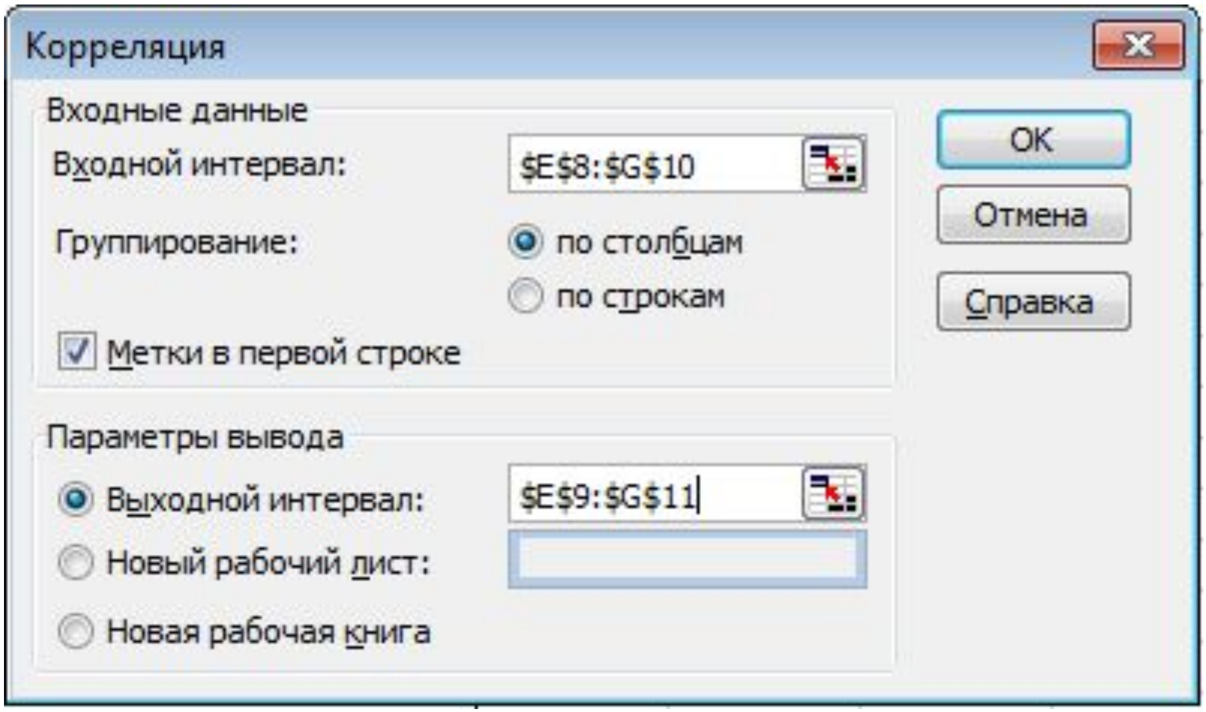
Критическое значение t-критерия

Функция СТЬЮДРАСПОБР категория Статистические



4. Построить матрицу коэффициентов парной корреляции по трем переменным

Сервис -> Анализ данных -> Корреляция



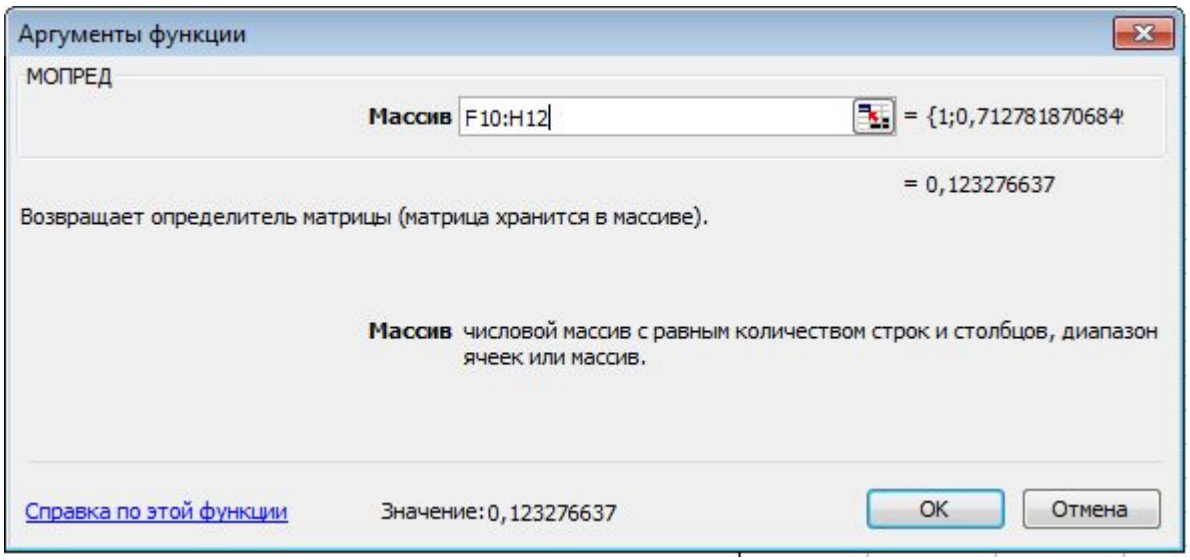
Матрица коэффициентов парной корреляции

	Y	*X1*	*X2*
Y	1		
X1	0,645918	1	
X2	0,816018	0,273373	1

5. Найти оценку множественного коэффициента корреляции

$$R_{1,2,3} = \sqrt{1 - \frac{|R|}{R_{11}}} = \sqrt{1 - \frac{0{,}1304}{0{,}9253}} = 0{,}9269$$

Функция МОПРЕД категория Математические



6. Найти оценки коэффициентов частной корреляции

$$r_{12(3)} = -\frac{R_{12}}{\sqrt{R_{11}R_{22}}} = -\frac{-0,423}{\sqrt{0,925 \cdot 0,334}} = 0,706$$

$$r_{13(2)} = -\frac{R_{13}}{\sqrt{R_{11}R_{22}}} = -\frac{-0,639}{\sqrt{0,925 \cdot 0,334}} = 0,871$$

или

$$r_{12(3)} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{\left(1 - r_{13}^2\right)\left(1 - r_{23}^2\right)}} = \frac{0,646 - 0,816 \cdot 0,272}{\sqrt{\left(1 - 0,816^2\right)\left(1 - 0,273^2\right)}} = 0,706$$

$$r_{13(2)} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{\left(1 - r_{12}^2\right)\left(1 - r_{23}^2\right)}} = \frac{0,816 - 0,646 \cdot 0,272}{\sqrt{\left(1 - 0,646^2\right)\left(1 - 0,273^2\right)}} = 0,871$$

Регрессионный анализ

Линейная модель парной регрессии

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

Предпосылки метода наименьших квадратов

1. Математическое ожидание случайной составляющей в любом наблюдении должно быть равно нулю

$$M(\varepsilon_i) = 0$$

2. Возмущение ε_i (или зависимая переменная y_i) есть величина случайная, а объясняющая переменная x_i – неслучайная.

3. В любых двух наблюдениях отсутствует систематическая связь между значениями случайной составляющей

$$M(\varepsilon_i, \varepsilon_j) = 0 \quad (i \neq j)$$

4. Дисперсия случайной составляющей должна быть постоянной для всех наблюдений

$$D(\varepsilon_i) = \sigma_\varepsilon^2$$

Оценка параметров регрессионного уравнения

$$\hat{\beta} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}.$$

Условие идентифицируемости

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 \neq 0$$

Оценка качества уравнения регрессии

Коэффициент детерминации

$$R^2 = \frac{\text{Объясненная сумма квадратов}}{\text{Общая сумма квадратов}} =$$

$$= \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Коэффициент множественной корреляции

$$R = \sqrt{1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

$$R = \left| r_{y,x} \right|$$

Средняя относительная ошибка аппроксимации

$$E_{\text{отн}} = \frac{1}{n} \sum_{i=1}^{n} \frac{\left| y_i - \hat{y}_i \right|}{y_i} \cdot 100\% = \frac{1}{n} \sum_{i=1}^{n} \frac{\left| e_i \right|}{y_i} \cdot 100\%.$$

Проверка значимости уравнения регрессии и отдельных параметров

F-критерий

$$F = \frac{R^2}{1 - R^2}(n-2) = \frac{r_{y,x}^2}{1 - r_{y,x}^2}(n-2)$$

$$F > F_{табл}$$

Стандартная ошибка

$$\sigma_e = \sqrt{\frac{1}{n-2} \sum_{i=1}^{n} e_i^2}$$

Стандартные ошибки коэффициентов

$$\sigma_{\hat{\alpha}} = \frac{\sigma_e \sqrt{\sum_{i=1}^{n} x_i^2}}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}} = \sqrt{\frac{\sigma_e^2 \sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

$$\sigma_{\hat{\beta}} = \frac{\sigma_e \sqrt{n}}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}} = \sqrt{\frac{\sigma_e^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

t-критерий

$$t_{\alpha\ расч} = \frac{|\hat{\alpha}|}{\sigma_{\hat{\alpha}}} \qquad t_{\beta\ расч} = \frac{|\hat{\beta}|}{\sigma_{\hat{\beta}}}$$

$$t_{расч} > t_{табл}$$

Интервальная оценка параметров модели

$$\hat{\alpha} \pm t_{кр}\sigma_{\hat{\alpha}} \qquad \hat{\beta} \pm t_{кр}\sigma_{\hat{\beta}}$$

Прогнозирование с применением уравнения регрессии

Точечный прогноз

$$\hat{y}_{прогн} = \hat{\alpha} + \hat{\beta}x_{прогн}$$

Интервальный прогноз

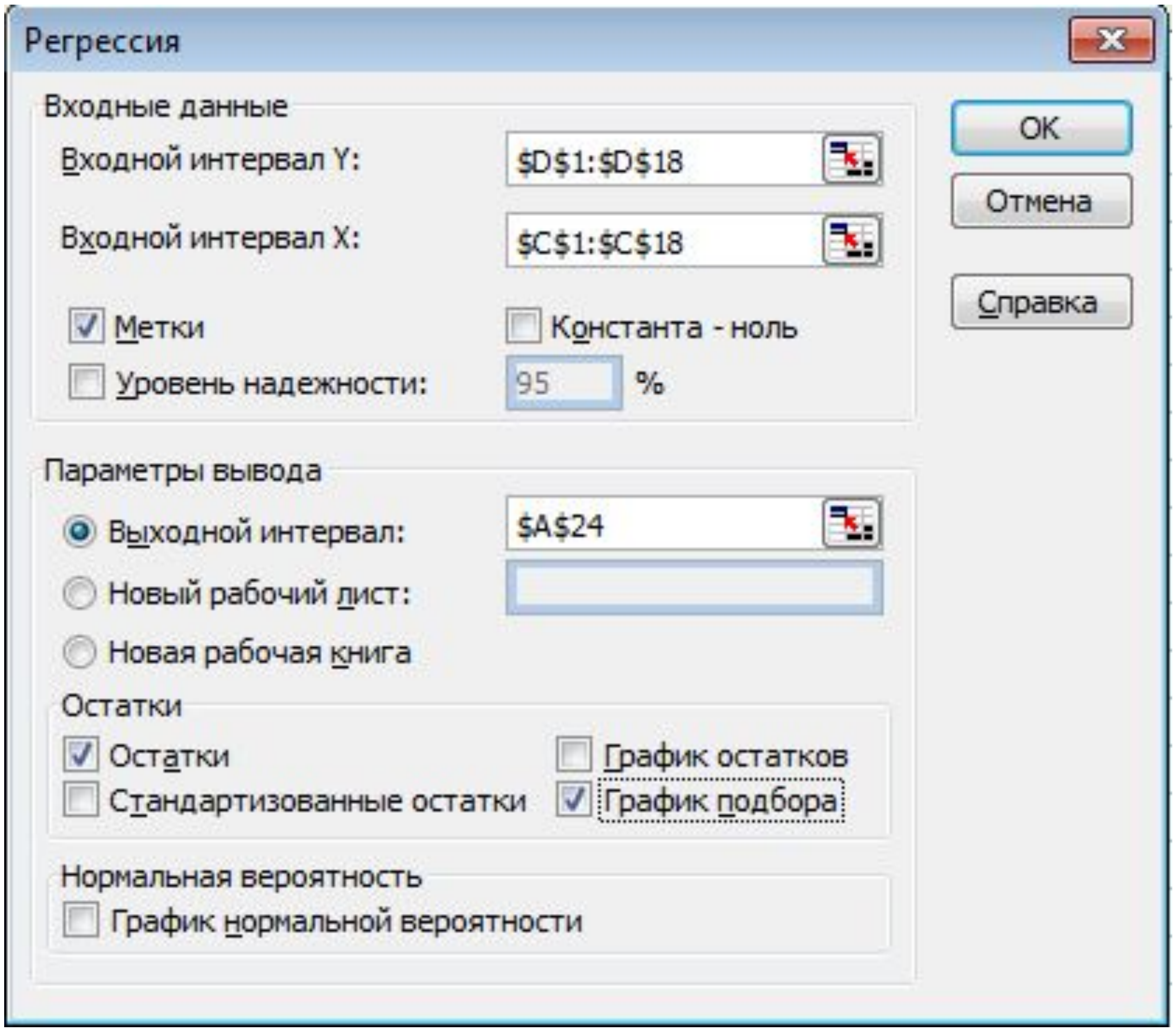
$$y_{прогн} \in \left[\hat{y}_{прогн} \pm \sigma_e t_\alpha \sqrt{1 + \frac{1}{n} + \frac{\left(x_{прогн} - \bar{x}\right)^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}}\right]$$

Пример. Имеются данные о среднедушевых месячных доходах и расходах по Центральному федеральному округу

	A	B	C	D
1	Область	№ п.п.	Доходы, руб.	Расходы, руб.
2	Белгородская	1	27845	24783
3	Брянская	2	22559	20347
4	Владимирская	3	20629	20198
5	Воронежская	4	25538	25015
6	Ивановская	5	15959	16684
7	Калужская	6	22547	21886
8	Костромская	7	23719	22173
9	Курская	8	25185	22021
10	Липецкая	9	27420	23921
11	Московаская (без г. Москва)	10	34168	33546
12	Орловская	11	25408	23475
13	Рязанская	12	25105	23096
14	Смоленская	13	28430	26712
15	Тамбовская	14	26486	22012
16	Твеская	15	22043	19328
17	Тульская	16	25611	21608
18	Ярославская	17	33118	29214

1. Построить однофакторную модель регрессии зависимости расходов от доходов

Сервис –> Анализ данных -> Регрессия



Регрессионная статистика	
Множественный R	0,94057199
R-квадрат	0,884675669
Нормированный R-квадрат	0,87698738
Стандартная ошибка	1371,09149
Наблюдения	17

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*ачимость F*
Регрессия	1	216315315	216315315	115,068	1,97E-08
Остаток	15	28198378,1	1879891,873		
Итого	16	244513693,1			

	Коэффициент	*Стандартная ошибка*	*t-статистика*	*P-Значени*	*ижние 95%*	*ерхние 95%*	*ижние 95,0*	*рхние 95,0%*
Y-пересечение	1701,814423	2040,284233	0,83410654	0,417307	-2646,95	6050,577	-2646,95	6050,577
Доходы, руб.	0,850193749	0,079257562	10,72697328	1,97E-08	0,68126	1,019127	0,68126	1,019127

ВЫВОД ОСТАТКА

Наблюдение	занное Расхо	Остатки
1	25375,45935	-592,4593512
2	20881,3352	-534,3351964
3	19240,46126	957,5387382
4	23414,06237	1600,937627
5	15270,05646	1413,943544
6	20871,13287	1014,867129
7	21867,55994	305,4400553
8	23113,94398	-1092,94398
9	25014,12701	-1093,127008
10	30751,23442	2794,765577
11	23303,53719	171,462814
12	23045,92848	50,07151979
13	25872,82269	839,1773059
14	24220,04605	-2208,046047
15	20442,63522	-1114,635222
16	23476,12652	-1868,126517

Уравнение регрессии

$$\hat{y}_i = \underset{(2040,28)}{1701,81} + \underset{(0,079)}{0,85}\, x_i$$

2. Проверить качество уравнения регрессии и оценить его значимость. Оценить точность модели

Коэффициент детерминации

Регрессионная статистика	
Множественный R	0,94057199
R-квадрат	0,884675669
Нормированный R-квадрат	0,87698738
Стандартная ошибка	1371,09149
Наблюдения	17

F-критерий

Дисперсионный анализ					
	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	216315315	216315315	115,068	1,9697E-08
Остаток	15	28198378,1	1879891,873		
Итого	16	244513693,1			

Табличное значение F-критерия FРАСПРОБР категория Статистические

$$F_{табл} = 4,54$$

Средняя ошибка аппроксимации

45	ВЫВОД ОСТАТКА			
46				
47	*Наблюдение*	*Предсказанное Расходы, руб.*	*Остатки*	*ei/yi*
48	1	25375,45935	-592,4593512	0,023905877
49	2	20881,3352	-534,3351964	0,026261129
50	3	19240,46126	957,5387382	0,047407602
51	4	23414,06237	1600,937627	0,063999106
52	5	15270,05646	1413,943544	0,084748474
53	6	20871,13287	1014,867129	0,046370608
54	7	21867,55994	305,4400553	0,013775315
55	8	23113,94398	-1092,94398	0,049631896
56	9	25014,12701	-1093,127008	0,045697379
57	10	30751,23442	2794,765577	0,08331144
58	11	23303,53719	171,462814	0,00730406
59	12	23045,92848	50,07151979	0,002167974
60	13	25872,82269	839,1773059	0,031415742
61	14	24220,04605	-2208,046047	0,100311014
62	15	20442,63522	-1114,635222	0,057669455
63	16	23476,12652	-1868,126517	0,086455318
64	17	29858,53099	-644,5309872	0,022062401
65				4,661734063

3. Проверить значимость коэффициента модели $\hat{\beta}$,вычислить доверительные интервалы с вероятностью 95% коэффициента модели регрессии

	Коэффициенты	Стандартная ошибка	t-статистика	Р-Значение	Нижние 95%	ерхние 95%	жние 95,0	рхние 95
Y-пересечение	1701,814423	2040,284233	0,83410654	0,417307	-2646,948459	6050,577	-2646,95	6050,57
Доходы, руб.	0,850193749	0,079257562	10,72697328	1,97E-08	0,681260255	1,019127	0,68126	1,01912

Р-значение 0,00<0,01<0,05

Верхняя и нижняя граница имеют одинаковые знаки

4. Построить доверительный интервал для полученной модели регрессии $(\alpha = 0{,}1)$

$$y_i \in \left[\hat{y}_i \pm U_i\right] = \left[\hat{y}_i \pm \sigma_e t \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}\right]$$

	yi расчетное	(xi-xcp)^2	Ui	Нижняя граница	Верхняя граница
1	25375,45935	5986657,526	2496,531088	22878,92826	27871,99044
2	20881,3352	8061257,055	2504,539409	18376,79579	23385,87461
3	19240,46126	22745605,29	2560,507614	16679,95365	21800,96888
4	23414,06237	19534,17301	2473,352392	20940,70998	25887,41477
5	15270,05646	89099162,94	2799,492058	12470,5644	18069,54851
6	20871,13287	8129542,702	2504,802568	18366,3303	23375,93544
7	21867,55994	2819831,173	2484,256797	19383,30315	24351,81674
8	23113,94398	45469,29066	2473,453604	20640,49038	25587,39758
9	25014,12701	4087532,526	2489,177541	22524,94947	27503,30455
10	30751,23442	76908773	2757,139841	27994,09458	33508,37426
11	23303,53719	95,34948097	2473,276529	20830,26066	25776,81371
12	23045,92848	85986,93772	2473,611717	20572,31676	25519,5402
13	25872,82269	9191597,232	2508,891964	23363,93073	28381,71466
14	24220,04605	1183232,055	2477,889679	21742,15637	26697,93573
15	20442,63522	11257603,88	2516,828	17925,80722	22959,46322
16	23476,12652	45268,82007	2473,452822	21002,67369	25949,57934
17	29858,53099	59594767,11	2695,843639	27162,68735	32554,37463
		299261917,1			

5. Оценить расходы, если доход составит 36000 руб. $(\alpha = 0{,}1)$

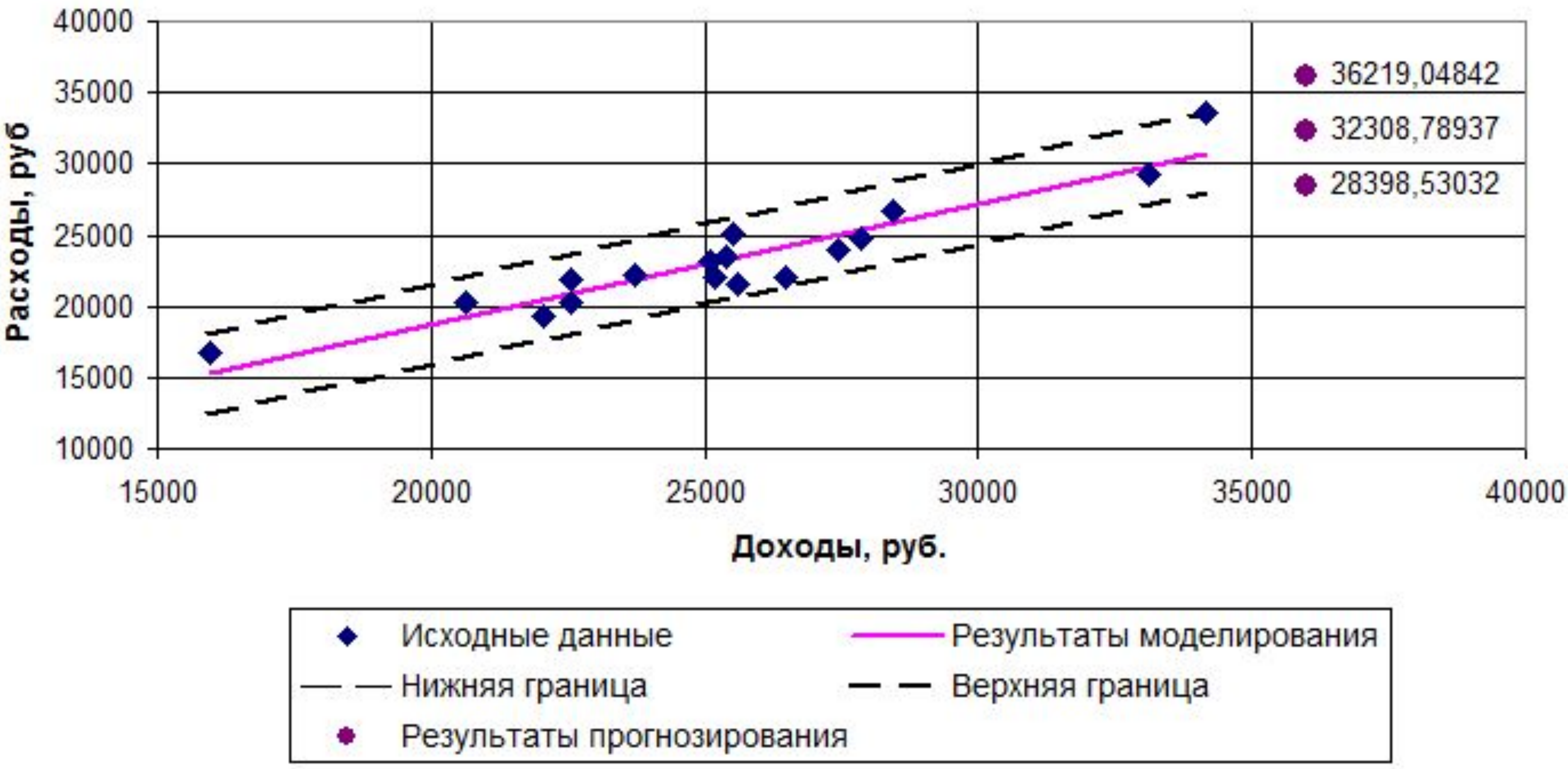
$$\hat{y}_{прогн} = 1701{,}81 + 0{,}85 \cdot 3600 = 32308{,}79$$

$$U_{x=36000,\ n=17,\ \alpha=0{,}1} = 1371{,}091 \cdot 1{,}75\sqrt{1 + \frac{1}{17} + \frac{(36000 - 25398{,}24)^2}{299261917{,}1}} = 3910{,}26$$

Нижняя граница $32308{,}79 - 39308{,}79 = 28398{,}53$

Верхняя граница $32308{,}79 + 3910{,}26 = 36219{,}05$

6. Отобразить на графике исходные данные, результаты моделирования и прогнозирования



Задачи оптимизации

Общая задача математического программирования

Найти максимум или минимум функции

$$f(\overline{X}) = f(x_1, x_2, \ldots, x_n)$$

при ограничениях

$$\varphi_1(x_1, x_2, \ldots, x_n)\{\le, =, \ge\}\, b_1,$$

$$\varphi_2(x_1, x_2, \ldots, x_n)\{\le, =, \ge\}\, b_2,$$

$$\ldots\ldots\ldots$$

$$\varphi_m(x_1, x_2, \ldots, x_n)\{\le, =, \ge\}\, b_m,$$

$$x_j \ge 0,\ j = \overline{1, n}.$$

Общая задача линейного программирования

$$\max(\min)\ f(\overline{X}) = c_1x_1 + c_2x_2 + \ldots + c_nx_n$$

при ограничениях (условиях):

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \{\le, =, \ge\}\, b_1,\\
a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \{\le, =, \ge\}\, b_2,\\
\cdots\cdots\cdots\\
a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \{\le, =, \ge\}\, b_m,
\end{aligned}$$

$$x_j \ge 0;\ j = \overline{1, n},$$

Каноническая задача линейного программирования

Найти $\max\ f(\overline{X}) = \sum_{j=1}^{n} c_j x_j$

при ограничениях $\sum_{j=1}^{n} a_{ij}x_j = b_i,\ i = \overline{1, m},$

$$x_j \ge 0,\ b_i \ge 0,\ i = \overline{1, m},;\ \ j = \overline{1, n}.$$

Пример (задача о смесях). Стандартом предусмотрено, что октановое число автомобильного бензина А-76 должно быть не ниже 76, а содержание серы в нем — не более 0,3%. Для изготовления такого бензина на заводе используется смесь из четырех компонентов. Данные о ресурсах смешиваемых компонентов, их себестоимости и их октановом числе, а также о содержании серы приведены в таблице

Характеристика	*Компонент автомобильного бензина*			
	№ 1	№ 2	№ 3	№ 4
Октановое число	68	72	80	90
Содержание серы, %	0,35	0,35	0,3	0,2
Ресурсы, т	700	600	500	300
Себестоимость, ден.ед./т	40	45	60	90

Требуется определить, сколько тонн каждого компонента следует использовать для получения 1000 т автомобильного бензина А-76, чтобы его себестоимость была минимальной.

Экономико-математическая модель задачи

$$\min f(\overline{X}) = 40x_1 + 45x_2 + 60x_3 + 90x_4,$$

$$x_1 + x_2 + x_3 + x_4 = 1000,$$

$$68x_1 + 72x_2 + 80x_3 + 90x_4 \geq 76 \cdot 1000,$$

$$0{,}35x_1 + 0{,}35x_2 + 0{,}3x_3 + 0{,}2x_4 \leq 0{,}3 \cdot 1000,$$

$$x_1 \leq 700,$$

$$x_2 \leq 600,$$

$$x_3 \leq 500,$$

$$x_4 \leq 300,$$

$$x_j \geq 0,\ j = 1,2,3,4.$$

Теорема 1 $\Delta_j = z_j - c_j < 0$, где $z_j = \sum_{i=1}^{m} c_i a_{ij}$, $j = \overline{1,n}$,

Теорема 2 $\Delta_j = z_j - c_j \geq 0$

A_k $z_k - c_k = \min(z_j - c_j)$.

A_r $Q = \min_i \frac{b_i}{a_{ik}} = \frac{b_r}{a_{rk}}; \quad a_{ik} > 0,\ i = \overline{1,m}.$

$$a'_{rj} = \frac{a_{rj}}{a_{rk}}, \quad j = \overline{1,n}, \qquad a'_{ij} = \frac{a_{ij} \cdot a_{rk} - a_{rj} \cdot a_{ik}}{a_{rk}},\ i = \overline{1,m};\ j = \overline{1,n}.$$

$$b'_r = \frac{b_r}{a_{rk}} \text{ для } i = r;\ b'_i = \frac{b_i \cdot a_{rk} - b_r \cdot a_{ik}}{a_{rk}},\ i = \overline{1,m} \text{ для } i \neq r.$$

Кластерный анализ

Результаты классификации респондентов по предпочтениям

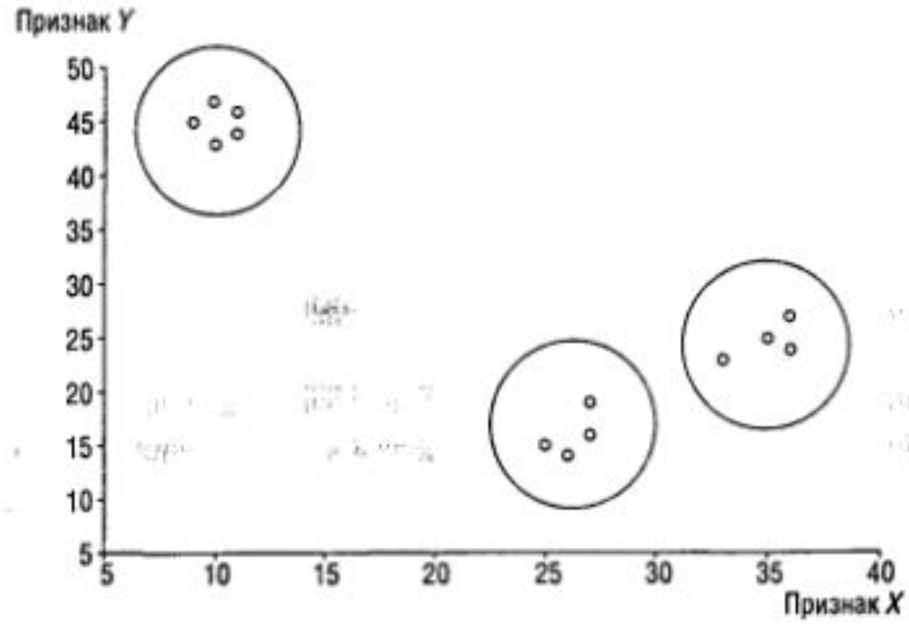
Свойство продукта	Ранги свойств по сегментам			
	1 (18%)	2 (7%)	3 (60%)	4 (15%)
Моющая способность	**3**	8	**2**	7
Отдушка	5	5	7	**1**
Цена	8	7	**1**	**2**
Безвредность	**1**	4	8	**3**
Эффект отбеливания	**2**	6	**3**	6
Подсинивание	4	**3**	6	8
Быстрое растворение	7	**1**	4	5
Отсутствие пыления	6	**2**	5	4

Представление исходных данных

$$X = \begin{pmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1m} \\ \dots & \dots & \dots & \dots & \dots \\ x_{i1} & \dots & x_{ij} & \dots & x_{im} \\ \dots & \dots & \dots & \dots & \dots \\ x_{n1} & \dots & x_{nj} & \dots & x_{nm} \end{pmatrix}$$

Совокупность объектов с двумя признаками

Испытуемый	Признак X	Признак Y
A	27	19
B	11	46
C	25	15
D	36	27
E	35	25
F	10	43
G	11	44
H	36	24
I	26	14
J	9	45
K	33	23
L	27	16
M	10	47



Расстояние между объектами

1) $d_{ij} \geq 0$ (неотрицательность);
2) $d_{ij} = d_{ji}$ (симметрия);
3) $d_{ij} + d_{jq} \geq d_{iq}$ (неравенство треугольника, здесь q — номер объекта);
4) если $d_{ij} \neq 0$, то $i \neq j$ (различимость нетождественных объектов);
5) если $d_{ij} = 0$, то $i = j$ (неразличимость тождественных объектов).

Меры расстояния между объектами

1) евклидово расстояние

$$d_{ij} = \sqrt{\sum_{k=1}^{m}(x_{ik} - x_{jk})^2}$$

2) взвешенное евклидово расстояние

$$d_{ij} = \sqrt{\sum_{k=1}^{m} w_k (x_{ik} - x_{jk})^2}$$

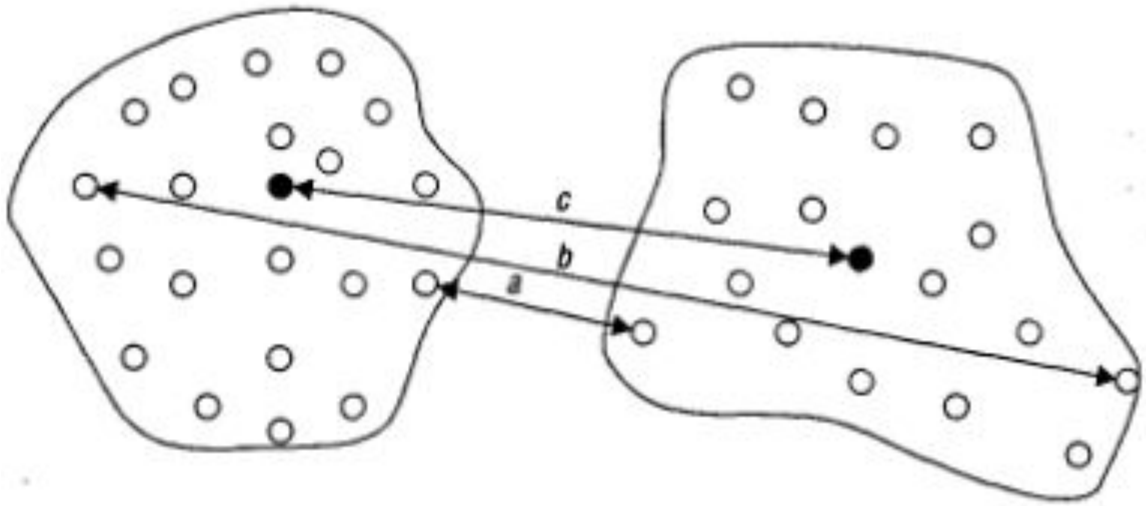
3) расстояние Минковского

$$d_{ij} = \left(\sum_{k=1}^{m}\left|x_{ik} - x_{jk}\right|^p\right)^{1/p};$$

4) расстояние *city-block* (расстояние городских кварталов)

$$d_{ij} = \sum_{k=1}^{m}\left|x_{ik} - x_{jk}\right|,$$

Меры сходства для объединения кластеров



Анализ качества классификации

$Q(S)$ функционал качества разбиения

d метрика в пространстве X

$X_1, X_2, \ldots, X_n$ наблюдения (объекты)

$S = \left(S_1, S_2, \ldots, S_p\right)$ Разбиение объектов на классы

Характеристики функционала качества

- ***сумма внутриклассовых дисперсий расстояний***

$$Q_1(S) = \sum_{l=1}^{p} \sum_{x_i \in S_l} d^2(x_i, \bar{x}_l);$$

- ***сумма попарных внутриклассовых расстояний*** между внутрикластерными элементами

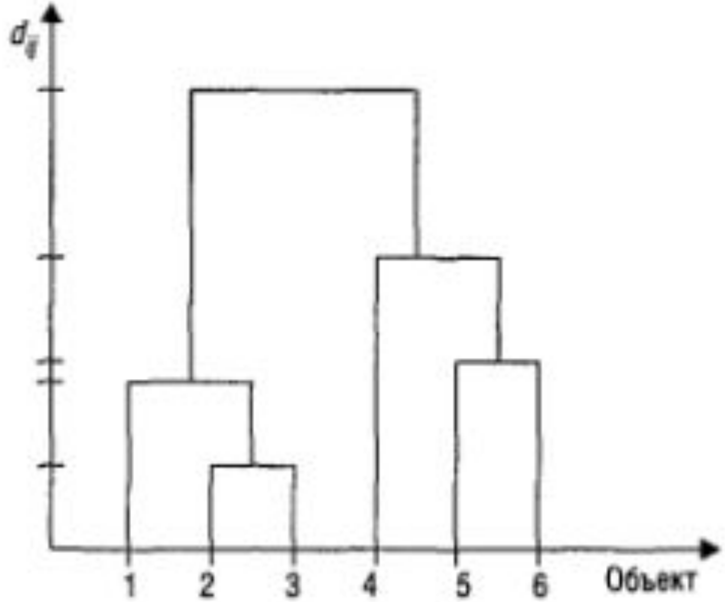
$$Q_2(S) = \sum_{l=1}^{p} \sum_{x_i, x_j \in S_l} d^2(x_i, x_j),$$

Кластерные профили

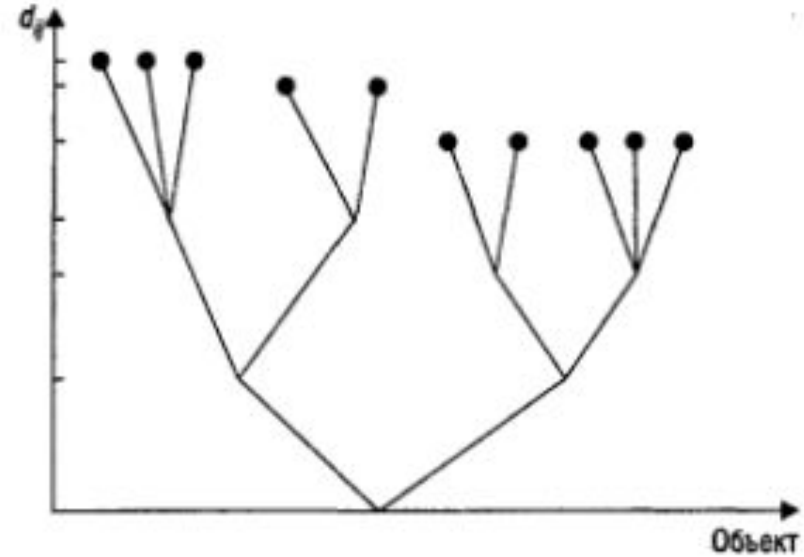
(%)

Кластер	Муж.	30–50 лет	Более 50 лет	Руководящие работники	Мед. обслуживание	Льготы	З/п	Стаж	Образование
1	80	90	5	70	10	12	95	30	30
2	50	70	45	13	60	70	60	40	20
3	90	35	10	5	30	20	70	20	50

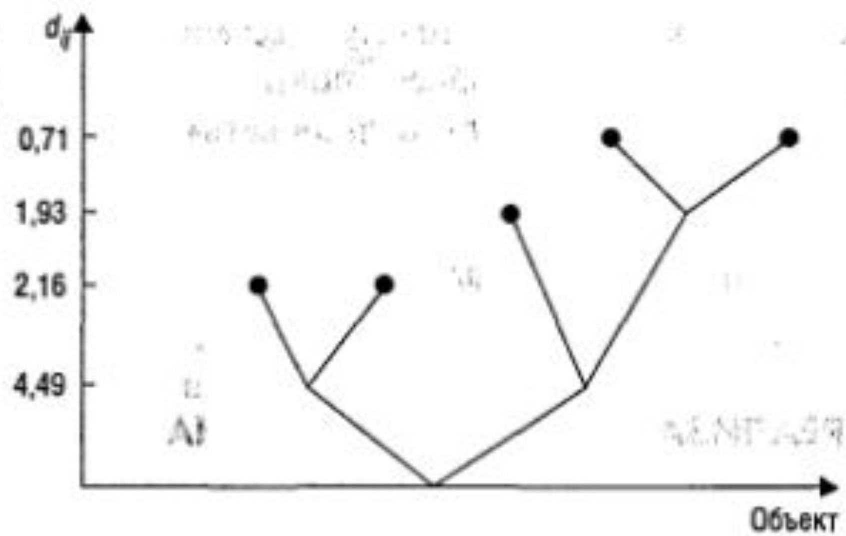
Методы кластерного анализа



Пример дендрограммы иерархического агломеративного кластерного анализа



Дендрограмма иерархического дивизимного алгоритма



Дендрограмма кластеризации объектов

Факторный анализ

Методика факторного анализа в случае одного объясняющего фактора

$$r(x_i, f_j) = a_{ij},\ i = \overline{1,k};\ j = \overline{1,m}.$$

Матрица факторных нагрузок $A = (a_{ij}),\quad i = \overline{1,k};\ j = \overline{1,m}.$

Матрица наблюдений

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{pmatrix}$$

Корреляционная матрица

$$R = (r_{ij}) = \begin{pmatrix} 1 & r_{12} & \cdots & r_{1k} \\ r_{21} & 1 & \cdots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \cdots & 1 \end{pmatrix}.$$

Редуцированная (преобразованная) корреляционная матрица

$$R^{+} = aa^{\mathrm{T}}$$

$$\begin{pmatrix} h_1^2 & r_{12} & \cdots & r_{1k} \\ r_{21} & h_2^2 & \cdots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \cdots & h_k^2 \end{pmatrix} = (a_1\, a_2 \ldots a_k)\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}.$$

Характерность $l_i^2 = 1 - h_i^2$ **Общность** h_i^2

$$r_{ii} = r_{ii}^{+} + l_i^2.$$

Методика факторного анализа в случае нескольких объясняющего фактора

	f_1	f_2	f_3	u_1	u_2	u_3	u_4	u_5	u_6
x_1	×	×		×					
x_2	×	×			×				
x_3	×	×	×			×			
x_4	×		×				×		
x_5	×		×					×	
x_6	×		×						×

Генеральный фактор

Общие факторы

Характерные факторы

$$x_i = \sum_{j=1}^{m} a_{ij} f_j + e_i, \quad i = \overline{1,k}; \; j = \overline{1,m},$$

$$M(ee^{\mathrm{T}}) = E, \quad M(e) = 0, \quad M(ef) = 0, \quad M(f) = 0, \quad M(ff^{\mathrm{T}}) = E$$

$$\sigma_{xi}^2 = \sigma_{ii}^2 = \sum_{j=1}^{m} a_{ij}^2 + l_{ii}^2,$$

$$X = AF + e,$$

Корреляция между переменными

$$R = M(XX^{\mathrm{T}}) = M(AFF^{\mathrm{T}}A^{\mathrm{T}}) + M(LL^{\mathrm{T}}) = \\ = AA^{\mathrm{T}} + L^2 = R^+ + L^2,$$

Характеристическое уравнение

$$|R^+ - \lambda E| = 0.$$

Матрица факторного отображения

$$A = V\Lambda^{-1/2}.$$

Оценки главных компонент и общих факторов

$$f_j = \frac{\sum_{i=1}^{k} a_{ij} x_i}{\sqrt{\lambda_j}},$$

Оценки факторных нагрузок

$$l_{ij} = a_{ij}\sqrt{\lambda_i}, \quad i = \overline{1,k}; \; j = \overline{1,m}.$$

Оценки общностей

$$h_i^2 = \sum_{i=1}^{k} l_{ij}^2.$$

Процедура вращения

$$F = B^{\mathrm{T}} A \Lambda_m^{-1/2} A^{\mathrm{T}} X,$$