МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Кузбасский государственный технический университет имени Т. Ф. Горбачева»

Кафедра металлорежущих станков и инструментов

Составители
А. Н. Коротков
Д. М. Дубинкин

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ИНЖЕНЕРНЫХ РАСЧЕТАХ

Методические указания к практическим занятиям

Рекомендованы учебно-методической комиссией специальности 15.05.01 Проектирование технологических машин и комплексов в качестве электронного издания для использования в образовательном процессе

Кемерово 2019

Рецензенты

Захарова Л. М. – доктор технических наук, профессор кафедры металлорежущих станков и инструментов.

Дягилева А. В. – кандидат технических наук, доцент кафедры математики.

Коротков Александр Николаевич,
Дубинкин Дмитрий Михайлович

Математические методы в инженерных расчетах: сборник практических занятий [Электронный ресурс] для обучающихся специальности 15.05.01 Проектирование технологических машин и комплексов всех форм обучения / сост.: А. Н. Коротков, Д. М. Дубинкин; КузГТУ. – Электрон. издан. – Кемерово, 2019.

Приведен теоретический и практический материал, необходимый для успешного изучения дисциплины.

Назначение издания – помощь обучающимся в получении практических знаний в области математических методов обработки результатов измерений применяемые в инженерных расчетах.

СОДЕРЖАНИЕ

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ №1 «ПРОВЕРКА ОДНОРОДНОСТИ НАБЛЮДЕНИЙ И ВЫБОРКИ НА НОРМАЛЬНОСТЬ»

Составители: Коротков А. Н., Дубинкин Д. М.

1. ЦЕЛЬ ПРАКТИЧЕСКОГО ЗАНЯТИЯ

Ознакомиться с основными характеристиками выборки, изучить методы проверки выборки на нормальность и однородность наблюдений.

2. КРАТКИЕ ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

Характеристики выборки. Пусть результатами наблюдения или измерения случайной величины X являются n её значений, X_1, X_2, …, X_n. Число n называется объемом выборки. Основными числовыми характеристиками выборки являются выборочное среднее ($\overline{X}$) и выборочная дисперсия (S^2):

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \ , \tag{1}$$

$$S = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}} \tag{2}$$

Размах выборки определяется как:

$$R = X_{\max} - X_{\min} \, . \tag{3}$$

Кроме перечисленных числовых характеристик выборки $\overline{X}$, S^2, R используются также среднеквадратическое отклонение или стандарт S (корень квадратный из дисперсии) и выборочный коэффициент вариации:

$$W = \frac{S}{\overline{\overline{X}}}. \tag{4}$$

При проверке статистических гипотез, относящихся к инженерным задачам, уровень значимости чаще всего принимается равным $\alpha = 0{,}05$. Для всех заданий принято именно это значение.

Выборка называется нормальной, если она получена из нормально распределенной генеральной совокупности. Для проверки выборки на нормальность вычисляется статистика:

$$N = \frac{R}{S}. \tag{5}$$

Если найденное значение статистики принадлежит критическому интервалу (a, b), то выборку можно считать нормальной (табл. 1).

Таблица 1

Значения критического интервала (a, b)

n	10	15	20	25	30	40	50	60	70	80	90	100
a	2,7	3,0	3,2	3,3	3,5	3,7	3,8	4,0	4,1	4,2	4,2	4,3
b	3,7	4,2	4,5	4,7	4,9	5,2	5,4	5,5	5,6	5,7	5,8	5,9

При проверке однородности наблюдений ошибочными элементами в выборке могут быть либо элемент с самым большим значением, либо с самым малым. Данная статистика имеет вид:

$$\tau = \frac{\left| X - \overline{X} \right|}{S}, \tag{6}$$

где X – сомнительный элемент выборки.

Сомнительный элемент отбрасывается как ошибочный, если вычисленное значение статистики превосходит критическое. В табл. 2 приведены критические значения, зависящие от объема выборки n.

Таблица 2

Критические значения

n	3	4	5	6	7	8	9	10
τ	1,41	1,69	1,87	2,00	2,09	2,17	2,24	2,29
n	11	12	13	14	15	16	17	18
τ	2,34	2,39	2,43	2,46	2,49	2,52	2,55	2,58
n	19	20	21	22	23	24	25	26
τ	2,60	2,62	2,64	2,66	2,68	2,70	2,72	2,73

Также сомнительный элемент или ошибочные результаты измерений (грубые погрешности) определяются и исключатся по критерию Граббса, формула (7). Статистический критерий Граббса исключения грубых погрешностей основан на предположении о том, что группа результатов измерений принадлежит нормальному распределению. Для этого вычисляют критерии Граббса G_1 и G_2, предполагая, что наибольший X_{max} или наименьший X_{min} результат измерений вызван грубыми погрешностями:

$$G_1 = \frac{\left|X_{max} - \overline{X}\right|}{S}, G_2 = \frac{\left|\overline{X} - X_{min}\right|}{S} \tag{7}$$

где $\bar{X}$ – среднее арифметическое результатов измерений, рассчитывается по формуле (1); S – среднеквадратическое отклонение результатов измерений, находится по формуле (2).

Сравнивают G_1 и G_2 с теоретическим значением $G_Т$ критерия Граббса при выбранном уровне значимости q. Критические значения критерия Граббса приведены в табл. 3.

Если $G_1 > G_Т$ или $G_2 > G_Т$, то X_{max} или X_{min} исключают как маловероятное значение. Далее вновь вычисляют среднее арифметическое (1) и среднее квадратическое отклонения (9) ряда результатов измерений и процедуру проверки наличия грубых погрешностей повторяют.

Если $G_1 < G_Т$ или $G_2 < G_Т$, то X_{max} или X_{min} не считают промахом и его сохраняют в ряду результатов измерений.

Возможно использование и других критериев исключения ошибочных результатов измерений (грубых погрешностей).

Таблица 3

Критические значения G_T для критерия Граббса

n	Одно наибольшее или одно наименьшее значение при уровне значимости *q*	
	Свыше 1%	Свыше 5%
3	1,155	1,155
4	1,496	1,481
5	1,764	1,715
6	1,973	1,887
7	2,139	2,020
8	2,274	2,126
9	2,387	2,215
10	2,482	2,290
11	2,564	2,355
12	2,636	2,412
13	2,699	2,462
14	2,755	2,507
15	2,806	2,549
16	2,852	2,585
17	2,894	2,620
18	2,932	2,651
19	2,968	2,681
20	3,001	2,709
21	3,031	2,733
22	3,060	2,758
23	3,087	2,781
24	3,112	2,802
25	3,135	2,822
26	3,157	2,841
27	3,178	2,859
28	3,199	2,876
29	3,218	2,893
30	3,236	2,908
31	3,253	2,924
32	3,270	2,938
33	3,286	2,952
34	3,301	2,965
36	3,330	2,991
38	3,356	3,014
40	3,381	3,036

3. ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ

1. Изучить основные теоретические положения.

2. Оформить отчет. Отчет должен содержать: наименование и цель работы; описание основных теоретических положений; ответы на контрольные вопросы.

Отчет по практическому занятию оформляется в тетради для практических занятий и в виде рабочей книги MS Excel и должен включать результаты выполнения индивидуальных заданий.

3. Решить задачи (определить нормальность и однородность наблюдений): получить задание (прил. 1) и произвести расчеты по формулам (1)÷(7), сделать выводы.

4. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Назовите основные характеристики выборки.
2. Как осуществляется проверка выборки на нормальность?
3. Как осуществляется проверка выборки на однородность наблюдений?
4. Критерии Граббса.

5. СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

1. Адлер, Ю. П. Планирование эксперимента при поиске оптимальных условий / Ю. П. Адлер и др. – Москва: Наука, 1976.

2. Математические методы и модели исследования операций [Электронный ресурс]. – Москва: Юнити-Дана, 2015. – 592 с. – Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=114719. – Загл. с экрана.

3. Зайцева, О. Н. Математические методы в приложениях. Дискретная математика [Электронный ресурс]. – Казань: Издательство КНИТУ, 2014. – 173 с. – Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=428299. – Загл. с экрана.

4. Математические методы исследования [Электронный ресурс] / сост.: Э. Н. Огнева – Кемерово: КемГУКИ, 2014. – 98 с. – Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=275375. – Загл. с экрана.

Приложение 1

Содержание свинца в питании флотации, %

Вариант 1		Вариант 2		Вариант 3	
№ проб	%	№ проб	%	№ проб	%
1	3,43	1	2,35	1	4,62
2	3,70	2	0,65	2	4,45
3	3,76	3	2,42	3	4,26
4	4,45	4	2,52	4	4,46
5	4,62	5	2,69	5	3,28
6	4,95	6	2,93	6	3,49
7	4,88	7	2,77	7	3,68
8	4,68	8	2,66	8	3,61
9	5,15	9	2,41	9	2,85
10	4,76	10	2,55	10	2,85
11	4,32	11	2,59	11	2,64
12	4,29	12	2,59	12	3,16
13	4,28	13	4,88	13	3,36
14	3,60	14	2,55	14	3,28
15	3,11	15	5,22	15	2,83
16	3,82	16	5,10	16	2,57
17	3,75	17	3,90	17	2,73
18	4,77	18	3,97	18	2,84
19	4,59	19	3,07	19	2,08
20	3,81	20	2,55	20	1,49
21	3,46	21	2,34	21	1,46
22	3,27	22	1,77	22	1,71
23	3,10	23	2,90	23	2,15
24	3,50	24	3,68	24	1,98
25	3,27	25	3,61	25	2,04
26	3,34	26	3,74	26	2,05
27	3,31	27	1,45	27	2,05
28	3,11	28	3,71	28	2,10
29	2,69	29	3,28	29	2,46
30	2,52	30	3,87	30	2,49
31	2,63	31	3,78	31	2,73
32	2,66	32	4,35	32	1,72
33	2,27	33	4,37	33	1,97

Продолжение приложения 1

Содержание свинца в питании флотации, %

Вариант 4		Вариант 5		Вариант 6	
№ проб	%	№ проб	%	№ проб	%
1	1,73	1	2,72	1	4,27
2	0,16	2	2,62	2	3,82
3	2,44	3	2,77	3	4,15
4	2,99	4	3,41	4	3,76
5	3,41	5	3,76	5	2,25
6	3,11	6	4,46	6	1,90
7	3,06	7	2,88	7	2,41
8	2,85	8	2,88	8	2,54
9	2,26	9	2,84	9	2,45
10	2,46	10	3,06	10	2,99
11	1,90	11	5,00	11	2,77
12	2,34	12	4,20	12	3,06
13	2,52	13	6,24	13	2,54
14	2,94	14	3,52	14	3,58
15	2,72	15	3,63	15	4,05
16	3,28	16	4,13	16	3,24
17	2,90	17	5,15	17	2,88
18	2,84	18	4,71	18	2,53
19	3,86	19	6,05	19	3,23
20	2,90	20	6,50	20	4,02
21	2,66	21	1,34	21	4,87
22	2,97	22	1,93	22	5,71
23	3,34	23	2,66	23	5,50
24	3,24	24	2,97	24	5,50
25	2,60	25	3,09	25	3,20
26	2,85	26	3,20	26	3,44
27	2,17	27	3,96	27	3,55
28	1,79	28	3,35	28	2,96
29	1,67	29	3,06	29	2,60
30	1,73	30	2,47	30	2,92
31	1,66	31	2,27	31	3,60
32	1,66	32	2,47	32	3,36
33	1,97	33	3,47	33	3,44

Продолжение приложения 1

Содержание цинка в питании флотации, %

Вариант 7		Вариант 8		Вариант 9	
№ проб	%	№ проб	%	№ проб	%
1	6,93	1	4,58	1	6,50
2	7,12	2	3,31	2	6,53
3	6,49	3	4,35	3	6,52
4	7,15	4	3,98	4	6,56
5	6,99	5	4,22	5	5,74
6	5,72	6	4,76	6	5,09
7	5,61	7	4,84	7	6,45
8	6,63	8	4,54	8	6,15
9	6,43	9	4,30	9	5,83
10	4,82	10	4,71	10	5,51
11	4,73	11	4,84	11	5,30
12	5,11	12	5,13	12	5,63
13	5,31	13	7,60	13	5,64
14	4,76	14	5,86	14	4,02
15	5,07	15	7,84	15	6,14
16	5,03	16	6,94	16	5,33
17	6,20	17	6,18	17	6,03
18	6,00	18	6,73	18	5,32
19	5,52	19	6,32	19	4,71
20	4,91	20	5,86	20	4,30
21	4,42	21	5,15	21	4,92
22	4,04	22	3,45	22	5,27
23	4,23	23	5,50	23	5,33
24	5,35	24	6,62	24	5,43
25	4,58	25	6,62	25	5,09
26	4,58	26	6,52	26	5,04
27	5,18	27	6,57	27	5,68
28	4,91	28	7,38	28	5,13
29	4,93	29	7,38	29	4,75
30	4,89	30	7,24	30	4,66
31	4,79	31	6,95	31	4,81
32	5,32	32	7,00	32	4,60
33	4,35	33	6,44	33	4,92

Продолжение приложения 1

Содержание цинка в питании флотации, %

Вариант 10		Вариант 11		Вариант 12	
№ проб	%	№ проб	%	№ проб	%
1	4,18	1	5,83	1	6,87
2	5,02	2	6,03	2	6,27
3	5,43	3	6,28	3	6,28
4	5,33	4	6,60	4	5,92
5	5,23	5	6,25	5	6,09
6	5,28	6	6,25	6	5,80
7	5,10	7	6,14	7	6,88
8	5,08	8	5,67	8	7,14
9	4,16	9	6,32	9	6,57
10	4,28	10	6,68	10	8,15
11	4,44	11	5,68	11	7,65
12	4,54	12	6,68	12	7,64
13	5,35	13	8,33	13	6,64
14	4,96	14	7,33	14	7,53
15	5,58	15	8,97	15	8,24
16	5,22	16	9,16	16	7,69
17	5,08	17	9,52	17	7,06
18	4,80	18	9,61	18	6,72
19	5,48	19	11,10	19	7,76
20	5,33	20	11,25	20	8,92
21	5,04	21	4,71	21	9,55
22	5,30	22	4,78	22	9,83
23	5,40	23	5,14	23	9,41
24	5,14	24	7,37	24	9,40
25	4,95	25	6,69	25	7,60
26	4,59	26	6,42	26	7,32
27	4,41	27	7,98	27	7,14
28	4,05	28	7,01	28	6,28
29	4,14	29	7,37	29	6,04
30	5,49	30	6,40	30	6,64
31	4,28	31	6,28	31	6,28
32	9,61	32	6,52	32	7,13
33	4,85	33	7,10	33	6,28

Продолжение приложения 1

Содержание меди в питании флотации, %

Вариант 13		**Вариант 14**		**Вариант 15**	
№ проб	%	№ проб	%	№ проб	%
1	0,79	1	0,57	1	1,21
2	0,85	2	0,14	2	1,13
3	0,87	3	0,59	3	1,03
4	1,01	4	0,60	4	1,09
5	1,06	5	0,68	5	0,87
6	1,11	6	0,72	6	0,82
7	1,02	7	0,75	7	0,86
8	1,06	8	0,73	8	0,73
9	1,04	9	0,67	9	0,57
10	0,88	10	0,64	10	0,57
11	0,82	11	0,80	11	0,56
12	0,80	12	0,75	12	0,71
13	0,83	13	0,88	13	0,78
14	0,63	14	0,75	14	0,78
15	0,65	15	0,96	15	0,73
16	0,66	16	1,00	16	0,67
17	0,74	17	0,69	17	0,72
18	0,85	18	0,76	18	0,70
19	0,85	19	0,55	19	0,52
20	0,75	20	0,44	20	0,34
21	0,73	21	0,43	21	0,34
22	0,66	22	0,32	22	0,39
23	0,56	23	0,50	23	0,45
24	0,82	24	0,92	24	0,48
25	0,62	25	0,89	25	0,48
26	0,66	26	0,98	26	0,46
27	0,67	27	0,86	27	0,44
28	0,65	28	0,91	28	0,46
29	0,61	29	0,92	29	0,51
30	0,64	30	0,88	30	0,52
31	0,63	31	0,96	31	0,57
32	0,68	32	1,03	32	0,42
33	0,52	33	0,92	33	0,50

Продолжение приложения 1

Содержание меди в питании флотации, %

Вариант 16		Вариант 17		Вариант 18	
№ проб	%	№ проб	%	№ проб	%
1	0,43	1	0,56	1	1,23
2	0,53	2	0,50	2	1,03
3	0,62	3	0,63	3	1,13
4	0,77	4	0,68	4	1,01
5	0,76	5	0,78	5	0,58
6	0,84	6	0,96	6	0,50
7	0,81	7	0,68	7	0,68
8	0,73	8	0,63	8	0,76
9	0,55	9	0,60	9	0,69
10	0,55	10	0,63	10	0,82
11	0,48	11	0,94	11	0,86
12	0,59	12	0,83	12	0,93
13	0,62	13	1,43	13	0,74
14	0,57	14	0,81	14	0,96
15	0,60	15	0,89	15	1,08
16	0,67	16	1,00	16	0,78
17	0,55	17	1,10	17	0,63
18	0,57	18	1,07	18	0,55
19	0,80	19	1,53	19	0,78
20	0,59	20	1,64	20	1,03
21	0,59	21	0,38	21	1,30
22	0,68	22	0,55	22	1,30
23	0,73	23	0,66	23	1,37
24	0,64	24	0,83	24	1,37
25	0,56	25	0,89	25	0,66
26	0,54	26	0,78	26	0,70
27	0,44	27	1,06	27	0,70
28	0,37	28	1,04	28	0,55
29	0,37	29	0,91	29	0,47
30	0,43	30	0,65	30	0,56
31	0,29	31	0,65	31	0,69
32	0,40	32	0,75	32	0,68
33	0,43	33	1,12	33	0,63

Продолжение приложения 1

Остаточная концентрация ксантогената
в камере основной флотации, %

Вариант 19		Вариант 20		Вариант 21	
№ проб	%	№ проб	%	№ проб	%
1	41	1	71	1	70
2	44	2	87	2	67
3	50	3	70	3	68
4	52	4	81	4	60
5	63	5	81	5	83
6	66	6	87	6	73
7	55	7	65	7	67
8	59	8	72	8	62
9	60	9	71	9	71
10	62	10	81	10	85
11	57	11	87	11	74
12	55	12	85	12	73
13	56	13	75	13	80
14	55	14	96	14	78
15	54	15	100	15	74
16	73	16	75	16	62
17	66	17	68	17	51
18	70	18	79	18	58
19	76	19	62	19	57
20	56	20	42	20	42
21	59	21	68	21	38
22	66	22	70	22	40
23	56	23	82	23	45
24	61	24	74	24	50
25	67	25	65	25	34
26	67	26	81	26	57
27	73	27	82	27	64
28	86	28	86	28	57
29	72	29	81	29	43
30	62	30	83	30	48
31	73	31	79	31	50
32	63	32	84	32	53
33	75	33	93	33	56

Продолжение приложения 1

Остаточная концентрация ксантогената
в камере основной флотации, %

Вариант 22		Вариант 23		Вариант 24	
№ проб	%	№ проб	%	№ проб	%
1	44	1	80	1	73
2	45	2	83	2	68
3	55	3	63	3	61
4	45	4	78	4	57
5	56	5	82	5	85
6	49	6	81	6	75
7	52	7	67	7	62
8	65	8	80	8	53
9	64	9	73	9	72
10	58	10	88	10	67
11	60	11	82	11	71
12	49	12	83	12	84
13	61	13	80	13	78
14	63	14	85	14	76
15	56	15	78	15	70
16	70	16	74	16	61
17	67	17	73	17	56
18	69	18	60	18	45
19	70	19	44	19	58
20	61	20	60	20	46
21	55	21	72	21	46
22	77	22	67	22	35
23	60	23	77	23	41
24	70	24	77	24	82
25	66	25	76	25	52
26	68	26	83	26	47
27	65	27	78	27	60
28	68	28	78	28	44
29	77	29	86	29	40
30	70	30	82	30	49
31	74	31	86	31	49
32	68	32	88	32	55
33	72	33	68	33	55

Продолжение приложения 1

Остаточная концентрация ксантогената
в камере основной флотации, %

Вариант 25		Вариант 26		Вариант 27	
№ проб	%	№ проб	%	№ проб	%
1	42	1	71	1	67
2	45	2	84	2	64
3	55	3	82	3	58
4	68	4	80	4	54
5	78	5	78	5	70
6	84	6	71	6	65
7	55	7	83	7	63
8	56	8	70	8	56
9	58	9	75	9	64
10	61	10	78	10	63
11	60	11	70	11	73
12	52	12	81	12	81
13	60	13	72	13	73
14	60	14	86	14	80
15	51	15	74	15	68
16	60	16	70	16	54
17	60	17	77	17	50
18	64	18	50	18	57
19	74	19	41	19	57
20	61	20	61	20	36
21	50	21	71	21	37
22	59	22	81	22	35
23	60	23	100	23	42
24	55	24	75	24	72
25	60	25	66	25	50
26	75	26	75	26	50
27	53	27	80	27	55
28	67	28	86	28	56
29	83	29	76	29	42
30	57	30	65	30	50
31	75	31	81	31	47
32	67	32	85	32	53
33	72	33	69	33	54

Продолжение приложения 1

Гранулометрический состав шихты окиси титана, мм

Вариант 28		**Вариант 29**		**Вариант 30**	
№ п/п	размеры	№ п/п	размеры	№ п/п	размеры
1	0,08	1	0,55	1	0,99
2	0,12	2	0,51	2	0,96
3	0,16	3	0,57	3	0,91
4	0,18	4	0,51	4	0,95
5	0,22	5	0,52	5	0,97
6	0,28	6	0,53	6	0,97
7	0,27	7	0,51	7	0,95
8	0,29	8	0,53	8	0,98
9	0,24	9	0,59	9	0,97
10	0,30	10	0,59	10	0,69
11	0,33	11	0,53	11	0,65
12	0,37	12	0,57	12	0,69
13	0,38	13	0,54	13	0,63
14	0,39	14	0,53	14	0,61
15	0,33	15	0,51	15	0,66
16	0,38	16	0,55	16	0,61
17	0,40	17	0,57	17	0,68
18	0,45	18	0,59	18	0,64
19	0,47	19	0,58	19	0,69
20	0,45	20	0,52	20	0,63
21	0,45	21	0,55	21	0,57
22	0,47	22	0,55	22	0,68
23	0,41	23	0,53	23	0,64
24	0,47	24	0,54	24	0,67
25	0,47	25	0,60	25	0,63
26	0,42	26	0,63	26	0,67
27	0,41	27	0,68	27	0,66
28	0,50	28	0,93	28	0,62
29	0,57	29	0,91	29	0,61
30	0,53	30	0,96	30	0,65
31	0,55	31	0,94	31	0,68
32	0,51	32	0,91	32	0,66
33	0,57	33	0,97	33	0,68

Продолжение приложения 1

Гранулометрический состав шихты окиси титана, мм

Вариант 31		Вариант 32		Вариант 33	
№ п/п	размеры	№ п/п	размеры	№ п/п	размеры
1	0,66	1	0,75	1	0,83
2	0,68	2	0,74	2	0,85
3	0,62	3	0,74	3	0,88
4	0,70	4	0,78	4	0,80
5	0,77	5	0,74	5	1,13
6	0,76	6	0,73	6	1,15
7	0,71	7	0,75	7	1,19
8	0,75	8	0,76	8	1,17
9	0,79	9	0,77	9	1,14
10	0,79	10	0,75	10	0,86
11	0,76	11	0,71	11	0,82
12	0,72	12	0,73	12	0,85
13	0,79	13	1,14	13	0,86
14	0,75	14	1,16	14	0,88
15	0,72	15	1,19	15	0,87
16	0,93	16	1,16	16	0,83
17	1,01	17	1,17	17	0,81
18	1,05	18	0,77	18	0,84
19	1,03	19	0,73	19	0,89
20	1,09	20	0,78	20	0,89
21	1,06	21	0,74	21	0,84
22	1,03	22	0,77	22	0,81
23	1,07	23	0,76	23	0,84
24	1,02	24	0,72	24	0,87
25	1,09	25	0,79	25	0,83
26	0,78	26	0,75	26	0,88
27	0,74	27	0,80	27	0,85
28	0,72	28	0,87	28	0,80
29	0,79	29	0,83	29	0,90
30	0,72	30	0,88	30	0,97
31	0,73	31	0,84	31	0,95
32	0,78	32	0,84	32	0,94
33	0,79	33	0,87	33	0,94

Продолжение приложения 1

Гранулометрический состав шихты окиси титана, мм

Вариант 34		Вариант 35		Вариант 36	
№ п/п	размеры	№ п/п	размеры	№ п/п	размеры
1	0,09	1	0,51	1	0,68
2	0,10	2	0,57	2	0,63
3	0,14	3	0,58	3	0,66
4	0,19	4	0,56	4	0,62
5	0,20	5	0,59	5	0,61
6	0,25	6	0,54	6	0,65
7	0,29	7	0,55	7	0,62
8	0,21	8	0,55	8	0,66
9	0,22	9	0,54	9	0,62
10	0,27	10	0,57	10	0,61
11	0,35	11	0,51	11	0,65
12	0,32	12	0,55	12	0,68
13	0,31	13	0,55	13	0,64
14	0,31	14	0,55	14	0,69
15	0,38	15	0,54	15	0,63
16	0,31	16	0,54	16	0,67
17	0,37	17	0,57	17	0,69
18	0,42	18	0,59	18	0,63
19	0,44	19	0,51	19	0,66
20	0,48	20	0,58	20	0,63
21	0,47	21	0,53	21	0,69
22	0,44	22	0,57	22	0,65
23	0,48	23	0,58	23	0,64
24	0,42	24	0,52	24	0,72
25	0,43	25	0,56	25	0,74
26	0,45	26	0,62	26	0,78
27	0,44	27	0,64	27	0,73
28	0,44	28	0,67	28	0,77
29	0,51	29	0,65	29	0,72
30	0,51	30	0,66	30	0,79
31	0,52	31	0,65	31	0,76
32	0,54	32	0,62	32	0,72
33	0,55	33	0,64	33	0,79

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ №2 «ОЦЕНКА РЕЗУЛЬТАТОВ ИЗМЕРЕНИЙ ПО КРИТЕРИЮ ФИШЕРА И СТЬЮДЕНТА»

Составители: Коротков А. Н., Дубинкин Д. М.

1. ЦЕЛЬ ПРАКТИЧЕСКОГО ЗАНЯТИЯ

Научиться применять критерии Фишера и Стьюдента на практике при обработке результатов измерений.

2. КРАТКИЕ ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

Критерий Фишера. Имеются две нормальные выборки объемом n_1 и n_2 с дисперсиями S_1^2 и S_2^2. Требуется проверить нуль-гипотезу: обе выборочные дисперсии являются оценками одной и той же генеральной дисперсии. F-статистика со степенями свободы числителя $f_1=(n_1-1)$ и знаменателя $f_2=(n_2-1)$ имеет вид:

$$F=\frac{S_1^2}{S_2^2}, \qquad S_1^2>S_2^2. \tag{1}$$

Нуль-гипотеза отклоняется, если найденное значение F-статистики превосходит критическое (табл. 1).

Таблица 1

Критические значения для F-статистики

f_2	f_1								
	4	6	8	10	20	30	40	60	∞
4	6,39	6,16	6,04	5,96	5,80	5,75	5,72	5,69	5,63
6	4,53	4,28	4,15	4,06	3,87	3,81	3,77	3,74	3,67
8	3,84	3,58	3,44	3,35	3,15	3,08	3,04	3,01	2,93
10	3,48	3,22	3,07	2,98	2,77	2,70	2,66	2,62	2,54
20	2,87	2,60	2,45	2,65	2,12	2,04	1,99	1,95	1,84
30	2,69	2,42	2,27	2,16	1,93	1,84	1,79	1,74	1,62
40	2,61	2,34	2,18	2,08	1,84	1,74	1,69	1,64	1,51
60	2,53	2,25	2,10	1,99	1,75	1,65	1,59	1,53	1,39
∞	2,37	2,10	1,94	1,83	1,57	1,46	1,39	1,32	1,00

Нулевая гипотеза – принимаемое по умолчанию предположение о том, что не существует связи между двумя наблюдаемыми событиями, феноменами. Так, нулевая гипотеза считается верной до того момента, пока нельзя доказать обратное. Опровержение нулевой гипотезы, то есть приход к заключению о том, что связь между двумя событиями, феноменами существует, – главная задача современной науки. Статистика как наука даёт чёткие условия, при наступлении которых нулевая гипотеза может быть отвергнута.

Часто в качестве нулевой гипотезы выступают предположения об отсутствии взаимосвязи или корреляции между исследуемыми переменными, об отсутствии различий (однородности) в распределениях (параметрах распределений) в двух и/или более выборках. Для обозначения нулевой гипотезы часто используют символ H_0.

Критерий Стьюдента. Имеются две нормальные выборки объемом n_1 и n_2 с дисперсиями S_1^2, S_2^2 и средними $\overline{X_1}$, $\overline{X_2}$. Требуется проверить нуль-гипотезу: $\overline{X_1}$ и $\overline{X_2}$ являются оценками одного и того же генерального среднего. Если выборочные дисперсии отличаются незначимо, то находится средневзвешенная дисперсия:

$$S^2 = \frac{S_1^2 \cdot (n_1 - 1) + S_2^2 \cdot (n_2 - 1)}{n_1 + n_2 - 2}, \tag{2}$$

и вычисляется статистика:

$$T = t \cdot S \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \tag{3}$$

где t – критерий имеет $f = n_1 + n_2 - 2$ степеней свободы (табл. 2).

При этом нуль гипотеза отклоняется, если $\left|\overline{X_1} - \overline{X_2}\right| > T$.

Таблица 2

Значения t – критерия

f	5	10	20	30	40	60	∞
t	2,57	2,23	2,08	2,04	2,02	2,00	1,96

3. ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ

1. Изучить основные теоретические положения.

2. Оформить отчет. Отчет должен содержать: наименование и цель работы; описание основных теоретических положений; ответы на контрольные вопросы.

Отчет по практическому занятию оформляется в тетради для практических занятий и в виде рабочей книги MS Excel и должен включать результаты выполнения индивидуальных заданий.

3. Решить задачи (определить критерий Фишера и Стьюдента): получить задание (см. практическое занятие 1 «Проверка однородности наблюдений и выборки на нормальность», прил. 1); определить для двух выборок $\overline{X}$, S^2 и R; проверить выборки на нормальность; сравнить дисперсии по критерию Фишера и по критерию Стьюдента; сделать выводы.

4. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Как осуществляется проверка критерию Фишера?
2. Как осуществляется проверка критерию Стьюдента?
3. Дайте определение нулевой гипотезы.
4. Для чего применяется понятие нулевой гипотезы?

5. СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

1. Адлер, Ю. П. Планирование эксперимента при поиске оптимальных условий / Ю. П. Адлер и др. – Москва: Наука, 1976.

2. Математические методы и модели исследования операций [Электронный ресурс]. – Москва: Юнити-Дана, 2015. – 592 с. – Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=114719. – Загл. с экрана.

3. Математические методы исследования [Электронный ресурс] / сост.: Э. Н. Огнева – Кемерово: КемГУКИ, 2014. – 98 с. – Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=275375. – Загл. с экрана.

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ №3 «ПОСТРОЕНИЕ ГИСТОГРАММЫ»

Составитель: Дубинкин Д. М.

1. ЦЕЛЬ ПРАКТИЧЕСКОГО ЗАНЯТИЯ

Получить практические навыки построения гистограммы. Ознакомиться с методикой построения гистограммы, преимуществами и недостатками данного статистического метода контроля качества.

2. КРАТКИЕ ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

Гистограмма – инструмент, который позволяет наглядно изобразить и легко выявить структуру и характер изменения полученных данных (оценить распределение), которые трудно заметить при их табличном представлении. Проведя анализ формы полученной гистограммы и ее местоположения относительно интервала допуска можно сделать заключение о качестве рассматриваемой продукции или состоянии изучаемого процесса. На основе заключения вырабатываются меры по устранению отклонений качества продукции или состояния процесса от нормы.

В зависимости от способа представления (сбора) исходных данных, существуют 2 методики построения гистограммы

Для сбора статистических данных разрабатываются контрольные листки показателей продукции или процесса.

Статистические данные уже собраны (например, проставлены в журналах регистрации) или их предполагается собрать в виде точно измеренных значений.

Методика №1 построения гистограммы

1. Определить количество и ширину интервалов для контрольного листка. Точное количество и ширину интервалов стоит выбирать исходя из удобства использования. Если для измеряемого показателя существуют допуски, то стоит ориентироваться на 6÷12 интервалов внутри допуска и 2÷3 интервала за пределами допуска. Если допусков нет, то оцениваем возможный разброс

значений показателя и тоже делим на 6÷12 интервалов. При этом ширина интервалов обязательно должна быть одинаковой.

2. Разработать контрольные листки и с их помощью произвести сбор необходимых данных.

3. С помощью заполненных контрольных листков подсчитать частоту попадания полученных значений показателя в каждый интервал. Обычно для этого выделяют отдельный столбец, расположенный в конце таблицы регистрации данных. Если значение показателя точно соответствует границе интервала, то добавьте по половинке обоим интервалам на границу которых попало значение показателя.

Для построения гистограммы необходимо использовать только те интервалы, в которые попало хотя бы одно значение показателя. Если между интервалами, в которые попали значения показателя, имеются пустые интервалы, то их тоже нужно построить на гистограмме.

4. Вычислить среднее значение результатов наблюдения по формуле (1):

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \,, \tag{1}$$

где X_i – результат наблюдения или измерения случайной величины; n – общее количество полученных данных в выборке.

5. Постройте горизонтальную и вертикальную оси.

6. На горизонтальную ось нанести границы выбранных интервалов. Для удобства восприятия рекомендуется перед первым и после последнего интервалов оставить место размером не менее одного интервала. Также необходимо предусмотреть место для нанесения границ допуска, если он есть.

Если в дальнейшем планируется сравнивать гистограммы, описывающие похожие факторы или характеристики, то стоит при нанесении шкалы на ось абсцисс руководствоваться не интервалами, а единицами измерения данных.

7. На вертикальную ось нанести шкалу значений в соответствии с выбранным масштабом и диапазоном.

8. Для каждого выбранного интервала построить столбик (прямоугольник), ширина которого равна интервалу, а высота

равна частоте попадания результатов наблюдений в соответствующий интервал (частота уже подсчитана ранее).

9. Нанести на график линию, соответствующую среднему арифметическому значению исследуемого показателя. При наличии поля допуска построить линии, соответствующие границам и центру интервала допуска.

10. Провести анализ построенной гистограммы.

Методика №2 построения гистограммы

Полученные данные для анализа необходимо свести в один документ в удобном для дальнейшей обработки виде (например, в виде таблицы).

Вычислить диапазон значений показателя (выборочный размах) по формуле (2):

$$R = X_{\max} - X_{\min}, \tag{2}$$

где $X_{\max}$ и $X_{\min}$ – соответственно наибольшее полученное значение и наименьшее полученное значение.

3. Определить количество интервалов гистограммы. Для этого можно воспользоваться значениями, представленными в табл. 1.

Таблица 1

Определение количества интервалов гистограммы

Объем выборки (n)	Число интервалов (k)
23÷45	6
46÷90	7
91÷180	8
181÷361	9
362÷723	10
724÷1447	11
1448÷2895	12

4. Определить ширину (размер) интервалов по формуле (3):

$$H = R / k \tag{3}$$

Округлить полученный результат в большую сторону до удобного значения. Также необходимо обратить внимание на то, что вся выборка должна быть разделена на интервалы одинакового размера.

5. Определить границы интервалов. Сначала определить нижнюю границу первого интервала таким образом, чтобы она была меньше $X_{\min}$. К ней прибавить ширину интервала, чтобы получить границу между первым и вторым интервалами. Далее необходимо прибавлять ширину интервала (H) к предыдущему значению для получения второй границы, затем третьей и т. д.

После произведенных действий следует удостовериться, что верхняя граница последнего интервала больше $X_{\max}$.

6. Для выбранных интервалов подсчитать частоты попадания значений исследуемого показателя в каждый интервал. Если значение показателя точно соответствует границе интервала, то необходимо добавить по половинке обоим интервалам, на границу которых попало значение показателя.

7. Вычислить среднее значение исследуемого показателя по формуле (1).

8. Постройте горизонтальную и вертикальную оси.

9. На горизонтальную ось нанести границы выбранных интервалов. Для удобства восприятия рекомендуется перед первым и после последнего интервалов оставить место размером не менее одного интервала. Также необходимо предусмотреть место для нанесения границ допуска, если он есть.

Если в дальнейшем планируется сравнивать гистограммы, описывающие похожие факторы или характеристики, то стоит при нанесении шкалы на ось абсцисс руководствоваться не интервалами, а единицами измерения данных.

10. На вертикальную ось нанести шкалу значений в соответствии с выбранным масштабом и диапазоном.

11. Для каждого выбранного интервала построить столбик (прямоугольник), ширина которого равна интервалу, а высота равна частоте попадания результатов наблюдений в соответствующий интервал (частота уже подсчитана ранее).

12. Нанести на график линию, соответствующую среднему арифметическому значению исследуемого показателя. При нали-

чии поля допуска построить линии, соответствующие границам и центру интервала допуска.

13. Провести анализ построенной гистограммы.

Анализ гистограммы также разбивается на 2 варианта, в зависимости от наличия технологического допуска

1. Допуски для показателя не заданы. В этом случае производим анализ формы гистограммы (табл. 2).

2. Для исследуемого показателя существует технологический допуск. В этом случае производится анализ, как формы гистограммы, так и ее расположение по отношению к полю допуска (табл. 3).

Достоинства гистограммы:

– наглядность, простота освоения и применения;

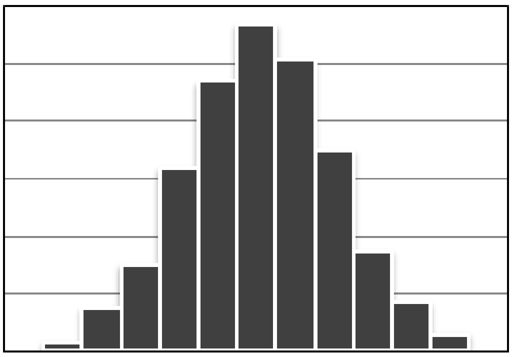
– управление с помощью фактов, а не мнений;

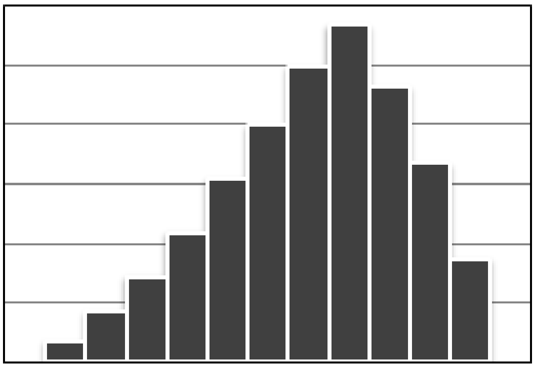
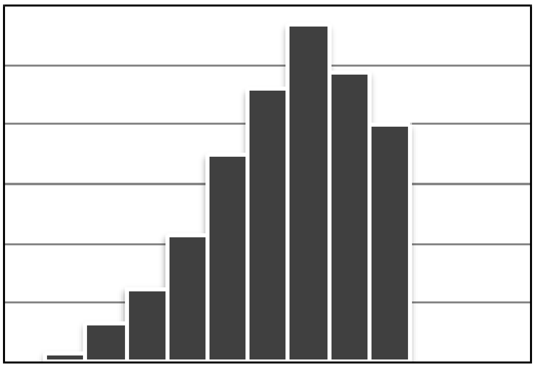
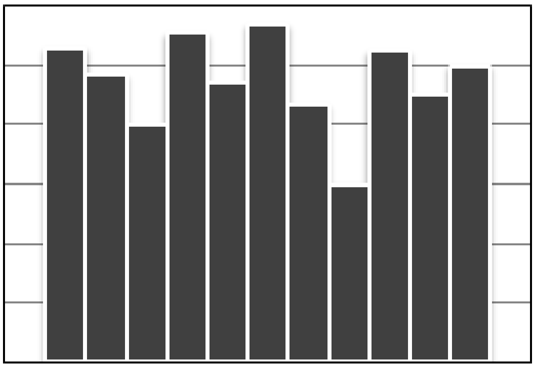
– позволяет лучше понять вариабельность, присущую процессу, глубже взглянуть на проблему и облегчить нахождение путей ее решения.

Недостатки гистограммы: интерпретация гистограммы, построенной по малым выборкам, не позволяет сделать правильные выводы.

Таблица 2

Анализ гистограммы в случае, когда технологические допуски для показателя не заданы

№ п/п	Анализ гистограммы	Гистограмма
1	Обычная (симметричная, колоколообразная) форма. Среднее значение гистограммы соответствует середине размаха данных. Максимальная частота также приходится на середину и постепенно уменьшается к обоим концам. Форма симметричная. Такая форма гистограммы встречается наиболее часто. Она свидетельствует о стабильности процесса	

№ п/п	Анализ гистограммы	Гистограмма
2	Отрицательно скошенное распределение (положительно скошенное распределение). Среднее значение гистограммы располагается правее (левее) середины размаха данных. Частоты резко уменьшаются при движении от центра гистограммы вправо (влево) и медленно влево (вправо). Форма ассиметричная. Такая форма образуется либо, если верхняя (нижняя) граница регулируется теоретически или по значению допуска либо, если правое (левое) значение невозможно достигнуть	
3	Распределение с обрывом справа (распределение с обрывом слева). Среднее значение гистограммы располагается далеко правее (левее) середины размаха данных. Частоты очень резко уменьшаются при движении от центра гистограммы вправо (влево) и медленно влево (вправо). Форма ассиметричная. Такая форма часто встречается в ситуации 100 %-го контроля изделий по причине плохой воспроизводимости процесса	
4	Гребенка (мультимодальный тип). Интервалы через один или два обладают более низкими (высокими) частотами	

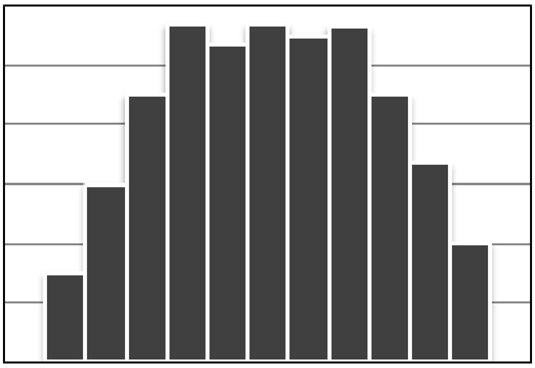
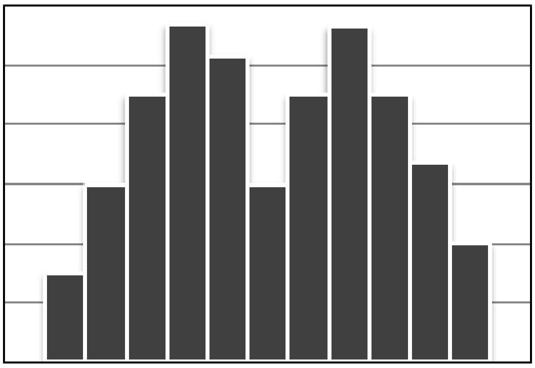
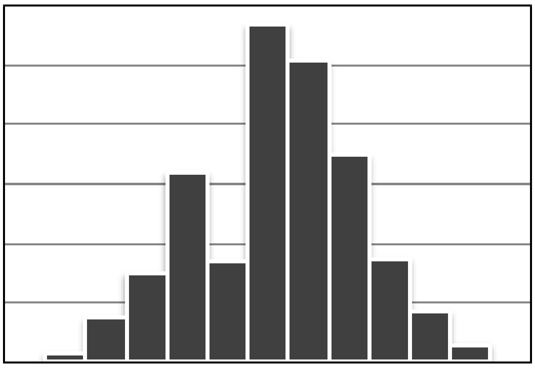
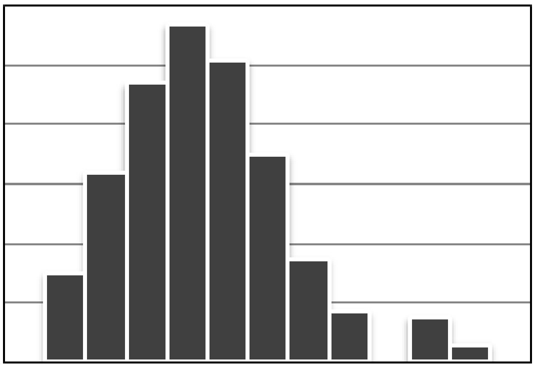
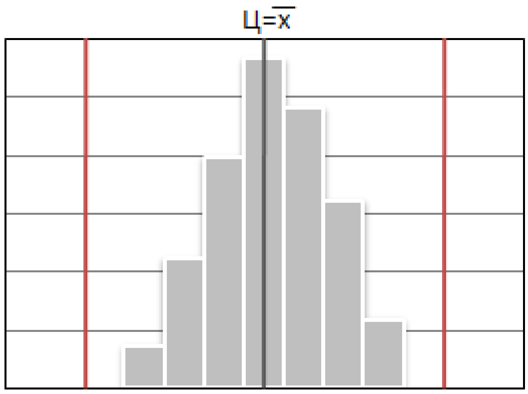
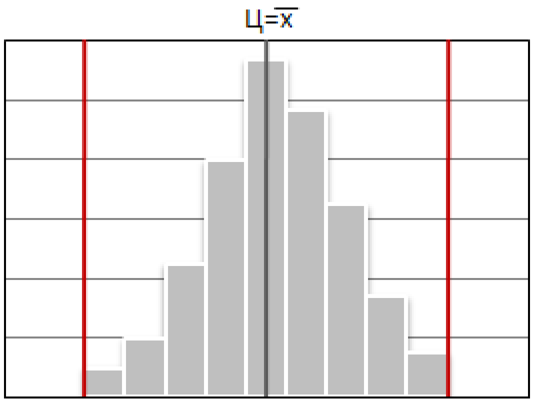
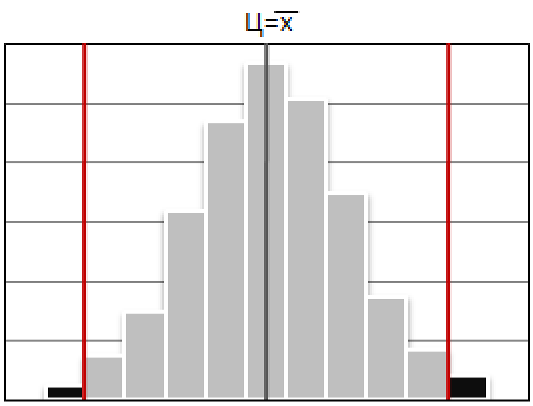
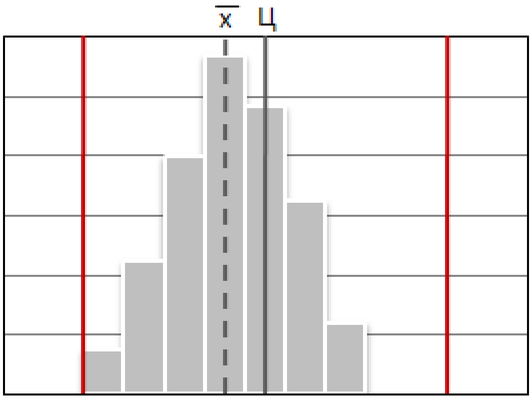
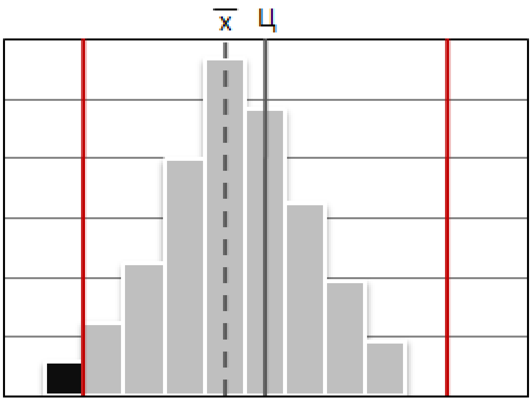
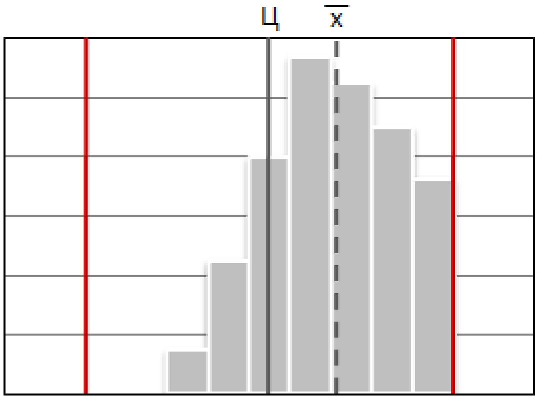
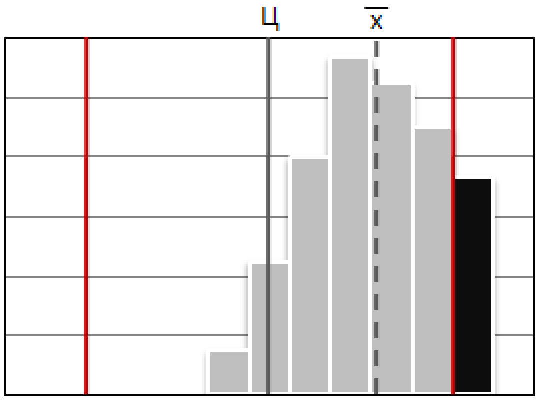
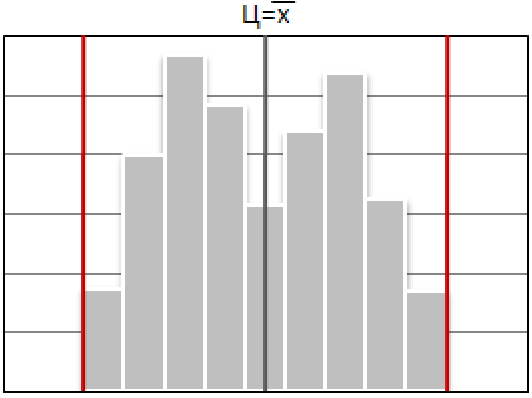
№ п/п	Анализ гистограммы	Гистограмма
5	Гистограмма, не имеющая высокой центральной части (плато). Частоты в середине гистограммы примерно одинаковые (для плато все частоты примерно равны). Такая форма встречается, если объединяется несколько распределений со средними значениями близко расположенными друг к другу. Для дальнейшего анализа рекомендуется применить метод стратификации	
6	Двухпиковый тип (бимодальный тип). В окрестностях середины гистограммы частота низкая, но с каждой стороны есть по пику частот. Данная форма встречается, если объединяется два распределения со средними значениями, далеко отстоящими друг от друга. Для дальнейшего анализа рекомендуется применить метод стратификации.	
7	Гистограмма с провалом (с вырванным зубом). Форма гистограммы близка к распределению обычного типа, но есть интервал с частотой ниже, чем в обоих соседних интервалах. Данная форма встречается, если ширина интервала не кратна единице измерения, если неправильно считаны показания шкалы и др.	
8	Распределение с изолированным пиком. Совместно с обычной формой гистограммы появляется небольшой изолированный пик. Такая форма образуется при включении небольшого количества данных из другого распределения, например, если нарушена управляемость процесса, произошли ошибки при измерении или произошло включение данных из другого процесса	

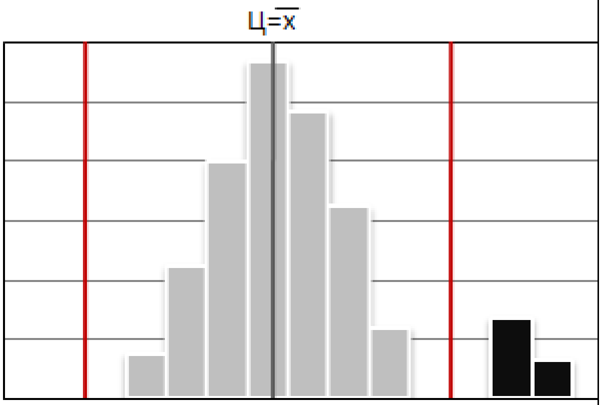
Таблица 2

Анализ гистограммы в случае, когда технологические допуски для показателя заданы

№ п/п	Анализ гистограммы	Гистограмма
1	Гистограмма имеет вид обычного распределения. Среднее значение гистограммы совпадает с центром поля допуска. Ширина гистограммы меньше ширины поля допуска с запасом. В данной ситуации процесс не нуждается в корректировке	
2	Гистограмма имеет вид обычного распределения. Среднее значение гистограммы совпадает с центром поля допуска. Ширина гистограммы равна ширине интервала допуска, в связи с чем возникают опасения появления не кондиционных деталей как со стороны верхнего, так и со стороны нижнего полей допуска. В этом случае необходимо либо рассмотреть возможность изменения технологического процесса с целью уменьшения ширины гистограммы либо расширить поле допуска, т. к. требования к качеству деталей в данном случае трудновыполнимы	
3	Гистограмма имеет вид обычного распределения. Среднее значение гистограммы совпадает с центром поля допуска. Ширина гистограммы больше ширины интервала допуска, в связи с чем обнаруживаются некондиционные детали как со стороны верхнего, так и со стороны нижнего полей допуска. В этом случае необходимо реализовать меры, описанные в пункте 2	

№ п/п	Анализ гистограммы	Гистограмма
4	Гистограмма имеет вид обычного распределения. Ширина гистограммы меньше ширины поля допуска с запасом. Среднее значение гистограммы сдвинуто влево (вправо) относительно центра интервала допуска, в связи с чем имеются опасения, что могут находится некондиционные детали со стороны нижней (верхней) границы поля допуска. В данной ситуации необходимо проверить, не вносят ли систематическую ошибку применяемые средства измерения. Если средства измерения исправны, следует отрегулировать процесс таким образом, чтобы центр гистограммы совпал с центром поля допуска	
5	Гистограмма имеет вид обычного распределения. Ширина гистограммы примерно равна ширине поля допуска. Среднее значение гистограммы сдвинуто влево (вправо) относительно центра интервала допуска, причем один или несколько интервалов выходят за границу поля допуска, что свидетельствует о наличии дефектных деталей. В этом случае первоначально необходимо отрегулировать технологические операции таким образом, чтобы центр гистограммы совпадал с центром поля допуска. После этого нужно принять меры для уменьшения размаха гистограммы или увеличения размера интервала допуска	

№ п/п	Анализ гистограммы	Гистограмма
6	Центр гистограммы смещен к верхнему (нижнему) пределу допуска, причем правая (левая) сторона гистограммы рядом с верхней (нижней) границей допуска имеет резкий обрыв. В этом случае можно сделать вывод, что изделия со значением показателя, выходящим за пределы поля допуска, исключили из партии или умышленно распределили как годные, для включения в пределы допуска. Следовательно, необходимо выявить причину, которая привела к появлению данного явления	
7	Центр гистограммы смещен к верхнему (нижнему) пределу допуска, причем правая (левая) сторона гистограммы рядом с верхней (нижней) границей допуска имеет резкий обрыв. Кроме того один или несколько интервалов выходят за границы поля допуска. Случай аналогичен 6, но интервалы гистограммы, выходящие за границы поля допуска, указывают на то, что измерительное средство было неисправно. В связи с эти необходимо провести поверку средств измерения, а также провести повторный инструктаж работникам по правилам выполнения измерений	
8	Гистограмма имеет два пика, хотя измерение значений показателя проводилось у изделий из одной партии	

№ п/п	Анализ гистограммы	Гистограмма
9	Основные характеристики гистограммы в порядке (соответствуют случаю 1), при этом имеются дефектные изделия со значениями показателя, выходящими за пределы поля допуска, которые образуют обособленный «островок» (изолированный пик). Данная ситуация могла возникнуть в результате небрежности, при которой дефектные детали были перемешаны с доброкачественными. В этом случае необходимо выявить причины и обстоятельства, приводящие к возникновению данной ситуации, а также принять меры к их устранению	

4. ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ

1. Изучить основные теоретические положения.

2. Оформить отчет. Отчет должен содержать: наименование и цель работы; описание основных теоретических положений; ответы на контрольные вопросы.

Отчет по практическому занятию оформляется в тетради для практических занятий и в виде рабочей книги MS Excel и должен включать результаты выполнения индивидуальных заданий.

3. По заданию преподавателя (по исходным данным практического занятия 1 «Проверка однородности наблюдений и выборки на нормальность», прил. 1) построить и провести анализ гистограммы, сделать выводы.

5. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Понятие гистограммы.

2. Какие существуют методики построения гистограммы?

3. Описать методику построения гистограммы в случае, когда технологический допуск для показателя не задан.

4. Описать методику построения гистограммы в случае, когда технологический допуск для показателя задан.

5. От чего зависит анализ гистограммы?

6. Преимущества применения гистограммы.

7. Недостатки применения гистограммы.

6. СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

1. Мазур, И. И. Управление качеством: учеб. пособие для студентов вузов обучающихся по специальности «Управление качеством» / И. И. Мазур, В. Д. Шапиро; под общ. ред. И. И. Мазура. – Москва: Омега-Л , 2006. – 400 с.

2. Кане, М. М. Системы, методы и инструменты менеджмента качества / М. М. Кане, Б. В. Иванов, В. Н. Корешков. – Санкт-Петербург: Питер , 2008. – 560 с.

3. Голоктеев, К. Н. Управление производством: инструменты, которые работают / К. Н. Голоктеев, И. А. Матвеев. – Санкт-Петербург: Питер , 2008. – 251 с.

4. Методы менеджмента качества: ежемесячный научно-технический журнал / Госстандарт России [и др.]. – Москва: РИА «Стандарты и качество».

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ №4 «ПОСТРОЕНИЕ ДИАГРАММЫ И ГРАФИКА»

Составитель: Дубинкин Д. М.

1. ЦЕЛЬ ПРАКТИЧЕСКОГО ЗАНЯТИЯ

Получить практические навыки построения диаграмм и графиков. Ознакомиться с методикой построения диаграммы и графика, преимуществами и недостатками данного статистического анализа.

2. КРАТКИЕ ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

MS Office Excel (рис. 1) – программа для работы с электронными таблицами. Она необходима для проведения расчетов, составления таблиц и диаграмм, вычисления простых и сложных функций, анализа данных и представления отчетности.

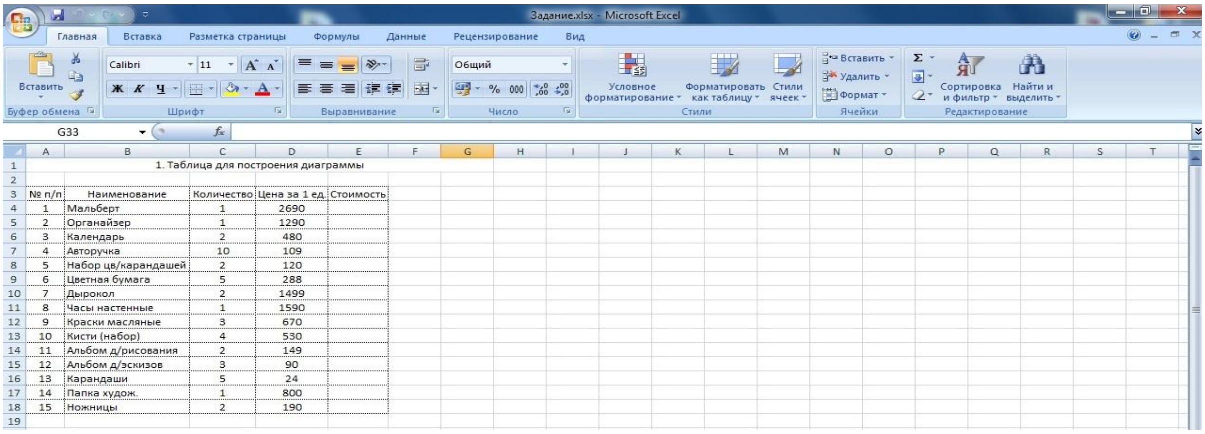


Рис. 1. Общий вид MS Office Excel

Каждый документ MS Office Excel – это отдельная книга, состоящая из листов, строк, столбцов и ячеек. Табличные данные из нескольких различных документов можно синхронизировать в один файл для дальнейшей общей обработки.

Диаграмма (график) (рис. 2) – это графическое представление данных, позволяющее оценить соотношение нескольких величин. Они используются для сравнения и анализа данных, представления их в наглядном виде.

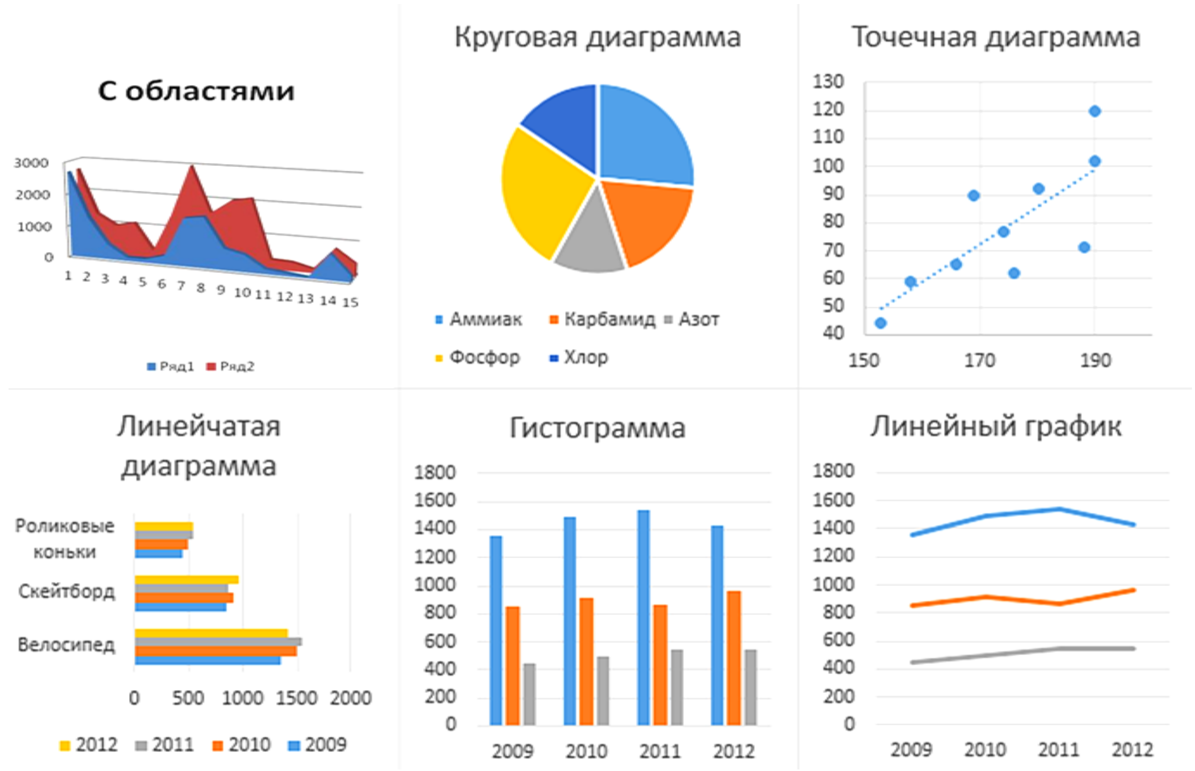


Рис. 2. Основные типы диаграмм и графиков

Диаграмма состоит из:

– геометрических объектов: точек, линий, фигур различной формы и цвета.

– вспомогательных элементов: осей координат, условных обозначений, заголовков и т. п.

Легенда – это условные обозначения: названия, маркеры и значение различных рядов данных на диаграмме.

Мастер диаграмм в MS Excel позволяет создавать 14 стандартных типов плоскостного и объемного представления (гистограмма, график, кольцевая, линейчатая, лепестковая, точечная, пузырьковая, поверхностная, круговая, биржевая, с областями и др.) и 22 нестандартных типа.

Мастер диаграмм осуществляет построение новой диаграммы в интерактивном режиме за 4 шага:

– выбор типа и формата диаграммы;

– выбор и указание диапазона данных для построения диаграммы;

– задание параметров диаграммы;

– размещение диаграммы

При изменении данных в таблице, диаграмма меняется автоматически.

Круговая диаграмма (рис. 3) показывает отношение размеров элементов, образующих ряд данных, к сумме элементов. Всегда отображается только один ряд данных. Такой тип диаграмм целесообразно использовать, когда необходимо подчеркнуть важный элемент. Для облегчения работы с маленькими секторами в основной диаграмме их можно объединить в один элемент, а затем разбить в отдельную диаграмму рядом с основной.

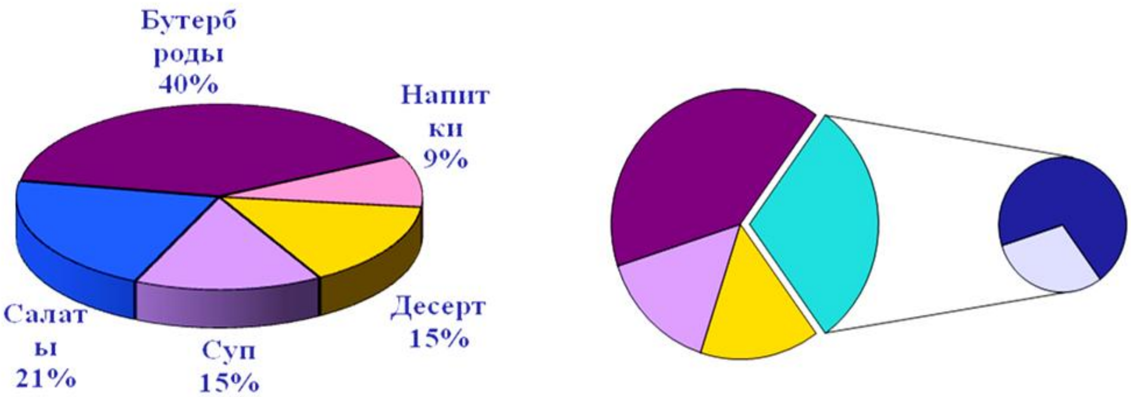


Рис. 3. Круговая диаграмма

Кольцевая диаграмма (рис. 4), как и круговая диаграмма, кольцевая показывает отношение частей к целому, но этот тип может включать несколько рядов данных. Каждое кольцо соответствует одному ряду данных.

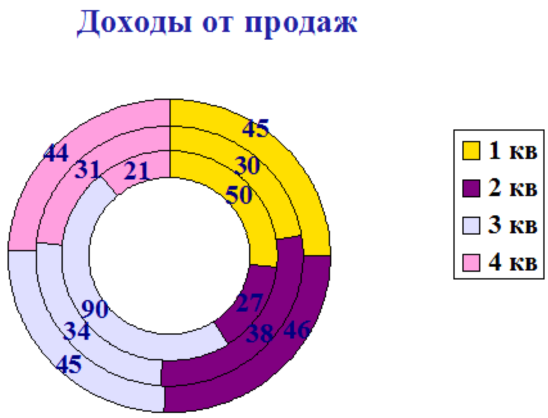
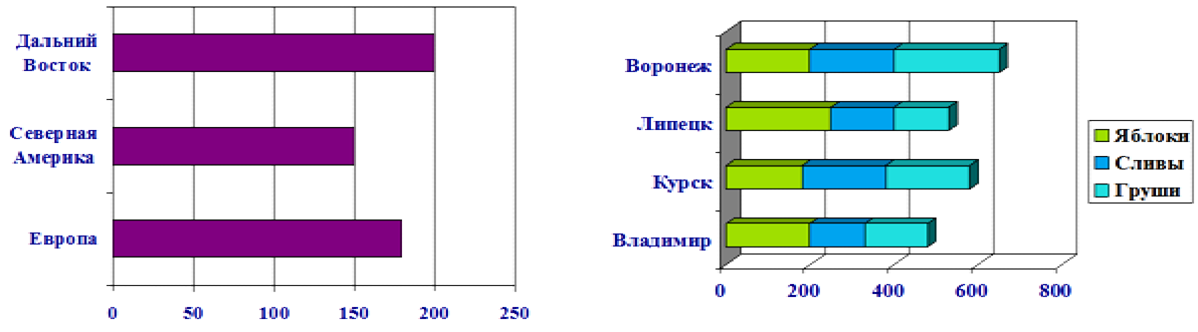


Рис. 4. Кольцевая диаграмма

Линейчатая диаграмма (рис. 5, *а*) позволяет сравнивать отдельные значения. Ось категорий расположена по вертикали, ось значений – по горизонтали. Это позволяет обратить большее внимание на сравниваемые значения, чем на время.

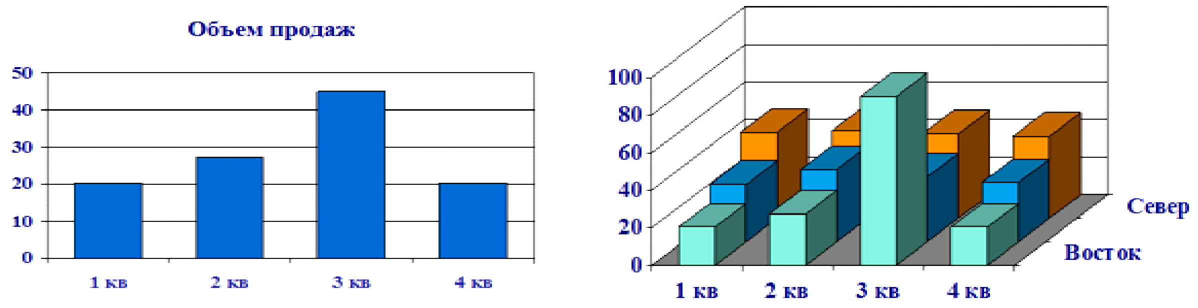


а) *б*)

Рис. 5. Линейчатая диаграмма (*а*) и линейчатая диаграмма с накоплением (*б*)

Линейчатая диаграмма с накоплением (рис. 5, *б*) показывает вклад отдельных элементов в общую сумму.

Гистограмма (рис. 6, *а*) позволяет представить изменение данных на протяжении отрезка времени. Диаграммы этого типа удобны также для наглядного сравнения отдельных величин. Ось категорий в гистограмме располагается по горизонтали, ось значений – по вертикали. Такое расположение осей подчеркивает характер изменения значений с течением времени.



а) *б*)

Рис. 6. Гистограмма (*а*) и гистограмма с накоплением (*б*)

Гистограмма с накоплением (рис. 6, *б*) позволяет представить отношение отдельных составляющих к их совокупному значению. На объемной гистограмме с перспективой сравнения значения располагаются в плоскости (вдоль двух осей).

Назначение основных типов диаграмм MS Excel:

Таблица 1

Название	Описание
Стандартные типы диаграмм	
Гистограммы	Используются для сравнения отдельных величин или их изменений в течение некоторого периода времени Удобны для отображения дискретных данных
Линейчатые диаграммы	Похожи на гистограммы (отличие – повернуты на 90° по часовой стрелке). Используются для сопоставления отдельных значений в определенный момент времени, не дают представления об изменении объектов во времени. Горизонтальное расположение полос позволяет подчеркнуть положительные или отрицательные отклонения от некоторой величины Линейчатые диаграммы можно использовать для отображения отклонений по разным статьям бюджета в определенный момент времени. Можно перетаскивать точки в любое положение
Графики	Отображают зависимость данных (ось *Y*) от величины, которая меняется с постоянным шагом (ось *X*) Метки оси категорий должны располагаться по возрастанию или убыванию. Графики чаще используют для коммерческих или финансовых данных, равномерно распределенных во времени (отображение непрерывных данных), или таких категорий, как продажи, цены и т. п.
Круговые диаграммы	Отображают соотношение частей и целого и строятся только по одному ряду данных, первому в выделенном диапазоне. Эти диаграммы можно использовать, когда компоненты в сумме составляют 100%
Точечные диаграммы	Хорошо демонстрируют тенденции изменения данных при неравных интервалах времени или других интервалах измерения, отложенных по оси категорий. Можно использовать для представления дискретных измерений по осям X и Y. В точечной диаграмме деления на оси категорий наносятся равномерно между самым низким и самым высоким значением X

Название	Описание
Диаграммы с областями	Позволяют отслеживать непрерывное изменение суммы значений всех рядов данных и вклад каждого ряда в эту сумму. Этот тип применяется для отображения процесса производства или продажи изделий (с равно отстоящими интервалами)
Поверхность	Показывает низкие и высокие точки поверхности. Эти диаграммы используются для набора данных, который зависит от двух переменных. Диаграмму можно поворачивать и рассматривать с разных точек зрения
Пузырьковые диаграммы	Позволяют отображать на плоскости наборы из трех значений. Первые два значения откладываются по осям X и Y. Третье значение представляется размером пузырька
Биржевая	Используется для отображения изменения информации о ценах на бирже. Отображает наборы данных из трех значений
Цилиндрические и др.	Являются объемными вариантами гистограмм и линейчатых диаграмм
Нестандартные типы диаграмм	
–	Нестандартные типы диаграмм основаны на стандартных типах, но имеют некоторые улучшения в форматировании и отображении
Пользовательские форматы диаграмм	
–	Добавляются в список дополнительных типов диаграмм. Пользовательские форматы создаются на основе базовых с применением различных средств форматирования

Для создания диаграмм в MS Excel, прежде всего, следует подготовить данные для построения диаграммы и определить ее тип.

При этом необходимо учитывать следующее:

1. MS Excel предполагает, что количество рядов данных (Y) должно быть меньше, чем категорий (X). Исходя из этого, определяется расположение рядов (в строках или столбцах), а также – снабжены ли ряды и категории именами.

а) если диаграмма строится для диапазона ячеек, имеющего больше столбцов, чем строк, или равное их число, то рядами данных считаются строки.

б) если диапазон ячеек имеет больше строк, то рядами данных считаются столбцы.

2. MS Excel предполагает, что названия, связанные с рядами данных, считаются их именами и составляют *легенду* диаграммы. Данные, интерпретируемые как категории, считаются названиями категорий и выводятся вдоль оси *X*.

3. Если в ячейках, которые MS Excel будет использовать как названия категорий, содержатся числа (не текст и не даты), то MS Excel предполагает, что в этих ячейках содержится ряд данных, и строит диаграмму без меток на оси категорий (*X*), вместо этого нумеруя категории.

4. Если в ячейках, которые MS Excel намерен использовать как названия рядов, содержатся числа (не текст и не даты), то MS Excel предполагает, что в этих ячейках содержатся первые точки рядов данных, а в каждом ряду данных присваивается имя Ряд 1, Ряд 2 и т. д.

С диаграммами можно производить следующие операции:

1. Добавлять и удалять ряды данных с помощью мастера диаграмм либо из контекстного меню диаграммы командой Исходные данные. Возможно также использование клавиши <Delete>, перетаскивание мышью данных на построенную диаграмму и др.

2. Изменять (редактировать) данные в диаграмме и на рабочем листе с помощью средства Подбор параметра (если данные, на которых построена диаграмма, выражены через формулу).

3. Переставлять ряды данных на диаграмме это касается, в основном, диаграмм гистограммного типа.

4. Вставлять текст в любом месте диаграммы выделить диаграмму (т. е. щелкнуть на ней мышью), а затем в строке формул ввести необходимый текст, который можно буксировать по всей диаграмме и форматировать как надпись.

5. Редактировать, форматировать и добавлять различные элементы диаграмм с помощью контекстного меню для необходимого объекта диаграммы.

6. Изменять пространственную ориентацию трехмерных диаграмм выделить диаграмму и воспользоваться командой меню Диаграмма | Объемный вид, можно также щелкнуть мышью на конце любой оси координат появятся черные крестики, а затем, удерживая мышь на любом из них, изменять расположение трехмерной диаграммы в пространстве.

7. Добавлять различные графические объекты (например, стрелки, выноски и т. д.) с помощью кнопок панели инструментов Рисование либо посредством команд меню Вставка | Рисунок.

8. Настраивать оси и выбирать шкалу с помощью контекстного меню для данной оси.

9. Строить составные диаграммы (различные типы графиков в одной системе координат) используя нестандартные типы диаграмм.

10. Изменять типы диаграмм, выбрав команду Тип диаграммы из ее контекстного меню.

11. Создавать рисованные диаграммы (вместо цветовой заливки рисунки). В данном случае необходимо выбрать некоторый ряд данных и использовать для него команду контекстного меню Формат рядов данных.

12. Связывать текст на диаграмме с ячейками рабочего листа.

13. Создавать диаграммы на основе структурированных данных.

14. Применять диаграммы для анализа данных, т. е. строить различные линии тренда и делать прогнозы.

4. ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ

1. Изучить основные теоретические положения.

2. Оформить отчет. Отчет должен содержать: наименование и цель работы; описание основных теоретических положений; ответы на контрольные вопросы.

3. По заданию преподавателя (см. приложение 1) построить 2 диаграммы и 2 графика, сделать выводы.

Отчет по практическому занятию оформляется в виде рабочей книги MS Excel (по каждой задаче отдельная страница рабочей книги) и должен включать результаты выполнения индивидуальных заданий.

Результаты выполнения индивидуального задания сохраняются в электроном обучении, и представляется студентом преподавателю для проверки и последующей защиты. Защита результатов практического занятия производится студентом только индивидуально. В ходе защите работы студент отвечает на вопросы преподавателя (поясняет методику выполнения заданий, отвечает на контрольные вопросы и т. д.).

Вопросы, на которые необходимо отвечать, когда будете выполнять практическое задание:

1. Вы собрали данные, упорядочили их, обобщили и отобразили в нескольких диаграммах. Однако прежде чем представить свое творение внешнему миру, попытайтесь ответить на ряд вопросов.
2. Иллюстрирует ли диаграмма закономерность, которую она должна иллюстрировать?
3. Уместен ли стиль диаграммы (формальная или неформальная) для предполагаемой аудитории?
4. Нет ли ошибок в данных? Вы проверили все ваши формулы дважды?
5. Правильно ли выбран тип диаграммы? Вы должны были рассмотреть несколько типов.
6. Если это круговая диаграмма, то нельзя ли изменить ее тип на более подходящий?
7. Если это объемная диаграмма, то может ли пользователь, глядя на нее, определить значение каждой точки данных?
8. Может ли диаграмма ввести пользователя в заблуждение?
9. Правильно ли выбраны типы осей? Если по оси категорий выведены числовые значения, равномерно ли размещены интервалы?
10. Все ли элементы диаграммы останутся разборчивыми, если распечатать ее на черно-белом принтере?
11. Идентифицируется ли числовая шкала оси значений (например, может ли пользователь увидеть, что выводится – тысячи, или миллионы)? Определены ли единицы измерения?
12. Если в диаграмме используются две оси значений, может ли пользователь понять, какая из них ассоциирована с каждым рядом данных?
13. Если создано несколько диаграмм, одинаково ли они выглядят?
14. Используется ли в них одна и та же цветовая гамма?
15. Используются ли в них одни и те же гарнитуры и размеры шрифтов?

16. Все ли текстовые элементы можно прочитать? Все ли тексты необходимы? Можно ли сократить их?

17. Используется ли в текстах более одной гарнитуры шрифтов? Если да, то постарайтесь применить единственную гарнитуру, например Arial. Не используются ли в текстах буквы только верхнего (или только нижнего) регистра?

18. Нет ли грамматических ошибок?

5. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что такое диаграмма и график?
2. Для чего используется диаграмма и график?
3. Типы диаграмм.
4. Из чего состоит диаграмма (график)?
5. Что такое легенда?
6. Как осуществляется построение диаграммы (графика)?
7. Что такое круговая и кольцевая диаграмма?
8. Что такое линейчатая диаграмма?

6. СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

1. Адлер, Ю. П. Планирование эксперимента при поиске оптимальных условий / Ю. П. Адлер и др. – Москва: Наука, 1976.

2. Математические методы и модели исследования операций [Электронный ресурс]. – Москва: Юнити-Дана, 2015. – 592 с. – Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=114719. – Загл. с экрана.

3. Математические методы исследования [Электронный ресурс] / сост.: Э. Н. Огнева – Кемерово: КемГУКИ, 2014. – 98 с. – Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=275375. – Загл. с экрана.

Приложение 1

Таблица 1

Численность населения мира, млн. чел.

№ варианта		*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
	Страна\год	1900	1913	1929	1938	1950	1960	1970	1980	1990	2000
1	США	76,4	97,6	122,2	130,5	153	176	200,5	227	247	277
2	Германия	45,7	54,7	58,7	62,3	67	72	77	78,5	79	82
3	Франция	40,8	41,8	42	42	42	46	50,5	54	56,5	59
4	Япония	44	51,6	63,2	71,8	83	93	104	116,8	123,5	127
5	СССР	123	158	171,5	186,5	205,5	226,5	247	258,5	290	290

Таблица 2

Численность занятых в мировой экономике, млн. чел.

№ варианта		*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*
	Страна\год	1900	1913	1929	1938	1950	1960	1970	1980	1990	2000
6	Германия	18,5	23,5	25	26,5	29	31	34	35	37	38,5
7	Франция	20	20	20	19,5	19	21	23	25	26,5	27,5
8	Великобритания	16,5	18,5	20	20,5	22,5	24	25	25,5	26	26,5
9	Италия	15	16,5	17	18	18,5	20	22	24	24,5	25

Таблица 3

Промышленное производство: добавленная стоимость, в ценах и по ППС нац. валют 2000 г., млрд. долл.

№ варианта		***21***	***22***	***23***	***24***	***25***	***26***	***27***	***28***	***29***	***30***
	Страна\год	1900	1913	1929	1938	1950	1960	1970	1980	1990	2000
10	Германия	29	51	59	478	93	244	420	510	575	625
11	Франция	28	46	57	52	63	93	190	275	310	355
12	Великобритания	53	73	84	105	130	180	245	265	300	335
13	Советский Союз	40	70	80	105	205	480	725	935	1000	545

Продолжение приложения 1

Таблица 4

Мировое сельскохозяйственное производство: добавленная стоимость в ценах и по ППС 2000 г., млрд. долл.

№ варианта	Страна\год	1900	1913	1929	1938	1950	1960	1970	1980	1990	2000
14	США	43	56	69	76,5	93,5	105	128,5	146	157,5	175
15	Германия	16	19	20	21,5	23	29	37	40,5	46,5	52,5
16	Франция	21,5	22	22,5	23	23,5	29,5	47	53	65	76,5
17	Италия	13,5	14,5	16	17	18,5	30,5	42	44,5	49	56
18	Советский Союз	37	50,5	58,8	63	75	81,5	87,5	98	120	100

Таблица 5

Мировой товарный экспорт, в ценах и по ППС 2000 г., млрд. долл.

№ варианта	Страна\год	1900	1913	1929	1938	1950	1960	1970	1980	1990	2000
19	Германия	21,5	54	58	64,1	36,5	87,5	185	385	600	710
20	Франция	22	28,5	40,5	40	31,5	62,5	140	235	330	420
21	Великобритания	38,5	54,5	73	76	66	105	160	235	320	400
22	Бельгия	12,2	15,5	18,4	16,8	12,3	27,5	63	112	176	214

Таблица 6

Динамика добычи нефти крупнейшими российскими компаниями, млн. тонн.

№ варианта	Компания\год	1996	1997	1998	1999	2000	2001	2002	2003
23	Сургутнефтегаз	0	33,9	35,2	37,6	41	44	48	52
24	ЛУКОЙЛ	53	57,1	64,2	73,5	75	78,3	81,4	84,7
25	Татнефть	25	25,5	25,8	26,3	23,1	24,6	24,7	26
26	Сибнефть	19	18,2	17,3	16,3	17,2	20,7	26,8	31,5

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ №5 «АППРОКСИМАЦИЯ РЕЗУЛЬТАТОВ ИЗМЕРЕНИЙ»

Составитель: Дубинкин Д. М.

1. ЦЕЛЬ ПРАКТИЧЕСКОГО ЗАНЯТИЯ

Получить практические навыками и правилами работы с Мастером диаграмм. Ознакомиться с методикой аппроксимации результатов измерений.

2. КРАТКИЕ ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

Сущность аппроксимации (подбора зависимостей) состоит в том, чтобы на основе имеющихся двух числовых рядов, которые можно представить как значения по оси X (независимая переменная) и по оси Y (зависимая переменная), получить соответствующее выражение для подбираемой функциональной зависимости Y от X. На практике такая задача может возникнуть при проведении каких-либо опытов, экспериментов, испытаний устройств и других подобных действий.

Коэффициент корреляции. Имеются результаты одновременных наблюдений $X_1, X_2, \ldots, X_n$ и $Y_1, Y_2, \ldots, Y_n$. Характеристики выборок $\overline{X}$, $\overline{Y}$, S_X^2 и S_Y^2.

Для нахождения коэффициента корреляции, который характеризует степень близости взаимосвязи между X и Y к линейной, вычисляется статистика:

$$C = \frac{(\overline{X \cdot Y} - \overline{X} \cdot \overline{Y})}{S_X \cdot S_Y}. \qquad (1)$$

Если вычисленное значение статистики C по абсолютной величине меньше критического (табл. 1), то корреляция отсутствует.

Таблица 1

Критические значения статистики *C*

n	5	6	7	8	9	10	11	12	13	14
C	0,669	0,621	0,582	0,549	0,521	0,497	0,476	0457	0,441	0,426
n	15	16	17	18	19	20	25	30	35	40
C	0,412	0,400	0,389	0,378	0,369	0,360	0,323	0,296	0,275	0,257
n	45	50	60	70	80	90	100	110	120	130
C	0,243	0,231	0,211	0,195	0,183	0,173	0,164	0,155	0,146	0,137

3. ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ

1. Изучить основные теоретические положения.

2. Оформить отчет. Отчет должен содержать: наименование и цель работы; описание основных теоретических положений; ответы на контрольные вопросы.

3. По заданию преподавателя (см. приложение 1) выполнить аппроксимацию результатов измерений по трем функциям, сравнить коэффициенты корреляции, сделать выводы.

Отчет по практическому занятию оформляется в виде рабочей книги MS Excel (по каждой задаче отдельная страница рабочей книги) и должен включать результаты выполнения индивидуальных заданий.

Результаты выполнения индивидуального задания сохраняются в электроном обучении, и представляется студентом преподавателю для проверки и последующей защиты. Защита результатов практического занятия производится студентом только индивидуально. В ходе защите работы студент отвечает на вопросы преподавателя (поясняет методику выполнения заданий, отвечает на контрольные вопросы и т. д.).

Технология выполнения работы.

1. Создать новый лист, переименовать. В ячейку C2 ввести текст Аппроксимация функции одной переменной.

2. В ячейку B4 ввести № п/п, в ячейку C4 ввести X, в ячейку D4 ввести Y.

3. В ячейки B5 и B6 ввести числа 1, 2, затем, используя маркер заполнения, протянуть до 15.

4. В столбец C, начиная с ячейки C5, ввести значения независимой переменной X согласно своему варианту. В столбец D, начиная с ячейки D5, ввести значения зависимой переменной Y согласно своему варианту.

5. Оформить рамку (границы) таблицы и расположить содержимое столбцов по центру ячеек.

6. Создать диаграмму-график, выделив диапазон ячеек C5:D15 со значениями X и Y, выбрать на вкладке Вставка График - График.

7. Применить подходящий макет к графику. Вкладка Работа с диаграммами – Конструктор – Макеты диаграмм. Дать название, удалить легенду.

8. Отформатировать остальные элементы диаграммы: увеличить толщину осевых линий, линий графиков и изменить цвет графика. Вкладки Работа с диаграммами – Макет и Работа с диаграммами – Формат.

9. Вывести в Область построения диаграммы искомое выражение для функции. Для этого необходимо выбрать на вкладке Работа с диаграммами – Макет – Линяя тренда. В диалоговом окне Линия тренда выбрать наиболее приемлемый вариант из числа предлагаемых и щёлкнуть по нему мышью. Для полиномиальной линии надо ещё установить предполагаемый показатель степени. Затем щелчком мыши установить флажки в окошечках: показывать уравнение на диаграмме и поместить на диаграмму величину достоверности аппроксимации (R^2). После этого щёлкнуть по кнопке ОК.

10. Переместить появившееся уравнение, правая часть которого и будет выражением для искомой функции, на свободное место Области диаграммы внизу графика. Для этого установить указатель мыши на выражение и щёлкнуть левой кнопкой мыши, а затем, зацепив мышью за появившуюся рамку, перетащить на свободное место.

5. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что такое точечный график?
2. Для чего используется точечный график?
3. Как осуществляется построение точечного графика?

4. Что такое круговая и кольцевая диаграмма?
5. Что такое линейчатая диаграмма?
6. Что такое коэффициент корреляции?

6. СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

1. Адлер, Ю. П. Планирование эксперимента при поиске оптимальных условий / Ю. П. Адлер и др. – Москва: Наука, 1976.

2. Математические методы и модели исследования операций [Электронный ресурс]. – Москва: Юнити-Дана, 2015. – 592 с. – Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=114719. – Загл. с экрана.

3. Математические методы исследования [Электронный ресурс]. / сост.: Э. Н. Огнева – Кемерово: КемГУКИ, 2014. – 98 с. – Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=275375. – Загл. с экрана.

Приложение 1

Таблица 1

№ п/п	ВАРИАНТЫ							
	1		2		3		4	
	x	***y***	***x***	***y***	***x***	***y***	***x***	***y***
1	–2,1	–0,86	–0,7	–0,8	–2,1	–0,5	–0,7	2,35
2	–1,8	–0,97	–0,6	–0,6	–1,8	–0,2	–0,6	2,2
3	–1,5	–1	–0,5	–0,5	–1,5	0,1	–0,5	2,1
4	–1,2	–0,93	–0,4	–0,4	–1,2	0,4	–0,4	1,98
5	–0,9	–0,78	–0,3	–0,3	–0,9	0,6	–0,3	1,88
6	–0,6	–0,56	–0,2	–0,2	–0,6	0,8	–0,2	1,77
7	–0,3	–0,29	–0,1	–0,1	–0,3	0,9	–0,1	1,67
8	0	0	0	0	0	0,98	0	1,57
9	0,3	0,29	0,1	0,1	0,3	0,9	0,1	1,47
10	0,6	0,56	0,2	0,2	0,6	0,8	0,2	1,37
11	0,9	0,78	0,3	0,3	0,9	0,6	0,3	1,27
12	1,2	0,93	0,4	0,4	1,2	0,4	0,4	1,16
13	1,5	1	0,5	0,5	1,5	0,1	0,5	1,04
14	1,8	0,97	0,6	0,6	1,8	-0,2	0,6	0,93
15	2,1	0,86	0,7	0,8	2,1	-0,5	0,7	0,8

Таблица 2

№ п/п	ВАРИАНТЫ							
	5		6		7		8	
	x	***y***	***x***	***y***	***x***	***y***	***x***	***y***
1	–0,35	1,93	–0,7	2	–0,7	0,5	0,3	0,55
2	–0,3	1,88	–0,6	1,8	–0,6	0,55	0,4	0,63
3	–0,25	1,82	–0,5	1,6	–0,5	0,61	0,5	0,7
4	–0,2	1,77	–0,4	1,5	–0,4	0,67	0,6	0,77
5	–0,15	1,72	–0,3	1,3	–0,3	0,74	0,7	0,84
6	–0,1	1,67	–0,2	1,2	–0,2	0,82	0,8	0,89
7	–0,05	1,62	–0,1	1,1	–0,1	0,9	0,9	0,95
8	0	1,57	0	1	0	1	1	1
9	0,05	1,52	0,1	0,9	0,1	1,1	1,1	1,05
10	0,1	1,47	0,2	0,8	0,2	1,2	1,2	1,09
11	0,15	1,42	0,3	0,7	0,3	1,35	1,3	1,14
12	0,2	1,37	0,4	0,67	0,4	1,5	1,4	1,18
13	0,25	1,32	0,5	0,6	0,5	1,65	1,5	1,22
14	0,3	1,27	0,6	0,56	0,6	1,82	1,6	1,26
15	0,35	1,21	0,7	0,5	0,7	2	1,7	1,3

Продолжение приложения 1

Таблица 3

№ п/п	ВАРИАНТЫ							
	9		10		11		12	
	x	*y*	*x*	*y*	*x*	*y*	*x*	*y*
1	2	0,3	2	0,69	–3,5	12,25	–1,4	–0,83
2	2,5	0,4	2,5	0,92	–3	9	–1,2	–0,64
3	3	0,48	3	1,1	–2,5	6,25	–1	–0,46
4	3,5	0,54	3,5	1,25	–2	4	–0,8	–0,3
5	4	0,6	4	1,39	–1,5	2,25	–0,6	–0,17
6	4,5	0,65	4,5	1,5	–1	1	–0,4	–0,08
7	5	0,7	5	1,61	–0,5	0,25	–0,2	–0,02
8	5,5	0,74	5,5	1,7	0	0	0	0
9	6	0,78	6	1,79	0,5	0,25	0,2	–0,02
10	6,5	0,81	6,5	1,87	1	1	0,4	–0,08
11	7	0,85	7	1,95	1,5	2,25	0,6	–0,17
12	7,5	0,88	7,5	2,01	2	4	0,8	–0,3
13	8	0,9	8	2,08	2,5	6,25	1	–0,46
14	8,5	0,93	8,5	2,14	3	9	1,2	–0,64
15	9	0,95	9	2,2	3,5	12,25	1,4	–0,83

Таблица 4

№ п/п	ВАРИАНТЫ							
	13		14		15		16	
	x	*y*	*x*	*y*	*x*	*y*	*x*	*y*
1	2	0,3	2	0,69	–3,5	12,25	–1,4	–0,83
2	2,5	0,4	2,5	0,92	–3	9	–1,2	–0,64
3	3	0,48	3	1,1	–2,5	6,25	–1	–0,46
4	3,5	0,54	3,5	1,25	–2	4	–0,8	–0,3
5	4	0,6	4	1,39	–1,5	2,25	–0,6	–0,17
6	4,5	0,65	4,5	1,5	–1	1	–0,4	–0,08
7	5	0,7	5	1,61	–0,5	0,25	–0,2	–0,02
8	5,5	0,74	5,5	1,7	0	0	0	0
9	6	0,78	6	1,79	0,5	0,25	0,2	–0,02
10	6,5	0,81	6,5	1,87	1	1	0,4	–0,08
11	7	0,85	7	1,95	1,5	2,25	0,6	–0,17
12	7,5	0,88	7,5	2,01	2	4	0,8	–0,3
13	8	0,9	8	2,08	2,5	6,25	1	–0,46
14	8,5	0,93	8,5	2,14	3	9	1,2	–0,64
15	9	0,95	9	2,2	3,5	12,25	1,4	–0,83

Продолжение приложения 1

Таблица 5

№ п/п	ВАРИАНТЫ							
	17		18		19		20	
	x	*y*	*x*	*y*	*x*	*y*	*x*	*y*
1	–2,1	–0,86	–0,7	–0,8	–2,1	–0,5	–0,7	2,35
2	–1,8	–0,97	–0,6	–0,6	–1,8	–0,2	–0,6	2,2
3	–1,5	–1	–0,5	–0,5	–1,5	0,1	–0,5	2,1
4	–1,2	–0,93	–0,4	–0,4	–1,2	0,4	–0,4	1,98
5	–0,9	–0,78	–0,3	–0,3	–0,9	0,6	–0,3	1,88
6	–0,6	–0,56	–0,2	–0,2	–0,6	0,8	–0,2	1,77
7	–0,3	–0,29	–0,1	–0,1	–0,3	0,9	–0,1	1,67
8	0	0	0	0	0	0,98	0	1,57
9	0,3	0,29	0,1	0,1	0,3	0,9	0,1	1,47
10	0,6	0,56	0,2	0,2	0,6	0,8	0,2	1,37
11	0,9	0,78	0,3	0,3	0,9	0,6	0,3	1,27
12	1,2	0,93	0,4	0,4	1,2	0,4	0,4	1,16
13	1,5	1	0,5	0,5	1,5	0,1	0,5	1,04
14	1,8	0,97	0,6	0,6	1,8	–0,2	0,6	0,93
15	2,1	0,86	0,7	0,8	2,1	–0,5	0,7	0,8

Таблица 6

№ п/п	ВАРИАНТЫ							
	21		22		23		24	
	x	*y*	*x*	*y*	*x*	*y*	*x*	*y*
1	–0,35	1,93	–0,7	2	–0,7	0,5	0,3	0,55
2	–0,3	1,88	–0,6	1,8	–0,6	0,55	0,4	0,63
3	–0,25	1,82	–0,5	1,6	–0,5	0,61	0,5	0,7
4	–0,2	1,77	–0,4	1,5	–0,4	0,67	0,6	0,77
5	–0,15	1,72	–0,3	1,3	–0,3	0,74	0,7	0,84
6	–0,1	1,67	–0,2	1,2	–0,2	0,82	0,8	0,89
7	–0,05	1,62	–0,1	1,1	–0,1	0,9	0,9	0,95
8	0	1,57	0	1	0	1	1	1
9	0,05	1,52	0,1	0,9	0,1	1,1	1,1	1,05
10	0,1	1,47	0,2	0,8	0,2	1,2	1,2	1,09
11	0,15	1,42	0,3	0,7	0,3	1,35	1,3	1,14
12	0,2	1,37	0,4	0,67	0,4	1,5	1,4	1,18
13	0,25	1,32	0,5	0,6	0,5	1,65	1,5	1,22
14	0,3	1,27	0,6	0,56	0,6	1,82	1,6	1,26
15	0,35	1,21	0,7	0,5	0,7	2	1,7	1,3

Продолжение приложения 1

Таблица 7

№ п/п	ВАРИАНТЫ							
	25		26		27		28	
	x	***y***	***x***	***y***	***x***	***y***	***x***	***y***
1	2	0,3	2	0,69	–3,5	12,25	–1,4	–0,83
2	2,5	0,4	2,5	0,92	–3	9	–1,2	–0,64
3	3	0,48	3	1,1	–2,5	6,25	–1	–0,46
4	3,5	0,54	3,5	1,25	–2	4	–0,8	–0,3
5	4	0,6	4	1,39	–1,5	2,25	–0,6	–0,17
6	4,5	0,65	4,5	1,5	–1	1	–0,4	–0,08
7	5	0,7	5	1,61	–0,5	0,25	–0,2	–0,02
8	5,5	0,74	5,5	1,7	0	0	0	0
9	6	0,78	6	1,79	0,5	0,25	0,2	–0,02
10	6,5	0,81	6,5	1,87	1	1	0,4	–0,08
11	7	0,85	7	1,95	1,5	2,25	0,6	–0,17
12	7,5	0,88	7,5	2,01	2	4	0,8	–0,3
13	8	0,9	8	2,08	2,5	6,25	1	–0,46
14	8,5	0,93	8,5	2,14	3	9	1,2	–0,64
15	9	0,95	9	2,2	3,5	12,25	1,4	–0,83

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ №6 «ПАРНАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ»

Составитель: Дубинкин Д. М.

1. ЦЕЛЬ ПРАКТИЧЕСКОГО ЗАНЯТИЯ

Получить практические навыки по парной регрессии и корреляции. Ознакомиться с методикой анализа парной регрессии и корреляции.

2. КРАТКИЕ ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

Парная (простая) линейная регрессия представляет собой модель, где среднее значение зависимой (объясняемой) переменной рассматривается как функция одной независимой (объясняющей) переменной *x*, т. е. это модель вида:

$$\hat{y}_x = f(x) \qquad (2)$$

Так же *y* называют результативным признаком, а *x* признаком-фактором. Знак «^» означает, что между переменными *x* и *y* нет строгой функциональной зависимости.

Практически в каждом отдельном случае величина *y* складывается из двух слагаемых:

$$y = \hat{y}_x + \varepsilon \qquad (2)$$

где *y* – фактическое значение результативного признака; $\hat{y}_x$ – теоретическое значение результативного признака, найденное исходя из уравнения регрессии; *ε* – случайная величина, характеризующая отклонения реального значения результативного признака от теоретического, найденного по уравнению регрессии.

Случайная величина *ε* называется также возмущением. Она включает влияние не учтенных в модели факторов, случайных ошибок и особенностей измерения. Ее присутствие в модели порождено тремя источниками: спецификацией модели, выборочным характером исходных данных, особенностями измерения переменных.

Различают *линейные* и *нелинейные* регрессии:

Линейная регрессия:

$$y = a + b \cdot x + \varepsilon$$

Нелинейные регрессии делятся на два класса: регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам, и регрессии, нелинейные по оцениваемым параметрам. Например:

регрессии, нелинейные по объясняющим переменным:

– полиномы разных степеней:

$$y = a + b_1 \cdot x + b_2 \cdot x^2 + \ldots + b_n \cdot x^n + \varepsilon$$

– равносторонняя гипербола:

$$y = a + \frac{b}{x} + \varepsilon$$

регрессии, нелинейные по оцениваемым:

– степенная:

$$y = a \cdot x^b \cdot \varepsilon$$

– показательная:

$$y = a \cdot b^x \cdot \varepsilon$$

– экспоненциальная:

$$y = e^{a + bx + \varepsilon}$$

Построение уравнения регрессии сводится к оценке ее параметров. Для оценки параметров регрессий, линейных по параметрам, используют метод наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических $\hat{y}_x$ минимальна, т. е.:

$$\sum \left(y - \hat{y}_x \right)^2 \to \min \qquad (3)$$

Для линейных и нелинейных уравнений, приводимых к линейным, решается следующая система относительно a и b:

$$\begin{cases} na + b\sum x = \sum y, \\ a\sum x + b\sum x^2 = \sum xy \end{cases} \qquad (4)$$

Можно воспользоваться готовыми формулами, которые вытекают непосредственно из решения этой системы:

$$a = \bar{y} - b \cdot \bar{x}, \qquad b = \frac{\operatorname{cov}(x, y)}{\sigma_x^2}, \qquad (5)$$

где $\operatorname{cov}(x,y)=\overline{y\cdot x}-\overline{y}\cdot\overline{x}$ ковариация признаков x и y $\sigma_x^2=\overline{x^2}-\overline{x}^2$ дисперсия признака x и

$$\overline{x}=\frac{1}{n}\sum x,\ \overline{y}=\frac{1}{n}\sum y,\ \overline{y\cdot x}=\frac{1}{n}\sum y\cdot x,\ \overline{x^2}=\frac{1}{n}\sum x^2.$$

Ковариация – числовая характеристика совместного распределения двух случайных величин, равная математическому ожиданию произведения отклонений этих случайных величин от их математических ожиданий.

Дисперсия – характеристика случайной величины, определяемая как математическое ожидание квадрата отклонения случайной величины от ее математического ожидания.

Математическое ожидание – сумма произведений значений случайной величины на соответствующие вероятности.

Тесноту связи изучаемых явлений оценивает коэффициент парной корреляции r_{xy} для линейной регрессии ($-1 \le r_{xy} \le 1$):

$$r_{xy}=b\cdot\frac{\sigma_x}{\sigma_y}=\frac{\operatorname{cov}(x,y)}{\sigma_x\cdot\sigma_y} \tag{6}$$

и индекс корреляции ρ_{xy} – для нелинейной регрессии ($0 \le \rho_{xy} \le 1$):

$$\rho_{xy}=\sqrt{1-\frac{\sigma_{\text{ост}}^2}{\sigma_y^2}}=\sqrt{1-\frac{\sum(y-\hat{y}_x)^2}{\sum(y-\overline{y})^2}}, \tag{7}$$

где $\sigma_y^2=\sum(y-\overline{y})^2$ – общая дисперсия результативного признака y; $\sigma_{\text{ост}}^2=\sum(y-\hat{y}_x)^2$ – остаточная дисперсия, определяемая исходя из уравнения регрессии $\hat{y}_x=f(x)$.

Оценку качества построенной модели даст коэффициент (индекс) детерминации r^2_{xy} (для линейной регрессии) либо ρ^2_{xy} (для нелинейной регрессии), а также средняя ошибка аппроксимации.

Средняя ошибка аппроксимации – среднее отклонение расчетных значений от фактических:

$$\overline{A}=\frac{1}{n}\sum\left|\frac{y-\hat{y}}{y}\right|\cdot 100\% \tag{8}$$

Допустимый предел значений $\overline{A}$ не более 10%.

Средний коэффициент эластичности $\overline{Э}$ показывает, на сколько процентов в среднем по совокупности изменится результат от своей средней величины при изменении фактора *x* на 1% от своего среднего значения:

$$\overline{Э} = f'(x)\frac{\overline{x}}{\overline{y}} \tag{9}$$

После того как найдено уравнение линейной регрессии, проводится оценка значимости как уравнения в целом, так и отдельных его параметров.

Проверить значимость уравнения регрессии – значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включенных в уравнение объясняющих переменных (одной или нескольких) для описания зависимой переменной.

Оценка значимости уравнения регрессии в целом производится на основе *F*-критерия Фишера, которому предшествует дисперсионный анализ. Согласно основной идее дисперсионного анализа, общая сумма квадратов отклонений переменной *y* от среднего значения $\overline{y}$ раскладывается на две части – «объясненную» и «необъясненную»:

$$\sum\left(y-\overline{y}\right)^2 = \sum\left(\hat{y}_x-\overline{y}\right)^2 + \sum\left(y-\hat{y}_x\right)^2, \tag{10}$$

где $\sum\left(y-\overline{y}\right)^2$ – общая сумма квадратов отклонений; $\sum\left(\hat{y}_x-\overline{y}\right)^2$ – сумма квадратов отклонений, объясненная регрессией (или факторная сумма квадратов отклонений); $\sum\left(y-\hat{y}_x\right)^2$ – остаточная сумма квадратов отклонений, характеризующая влияние неучтенных в модели факторов.

Схема дисперсионного анализа имеет вид, представленный в таблице 1 (*n* – число наблюдений, *m* – число параметров при переменной *x*).

Таблица 1

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Дисперсия на одну степень свободы
Общая	$\sum(y-\bar{y})^2$	$n-1$	$S^2_{\text{общ}} = \dfrac{\sum(y-\bar{y})^2}{n-1}$
Факторная	$\sum(\hat{y}_x-\bar{y})^2$	m	$S^2_{\text{факт}} = \dfrac{\sum(\hat{y}_x-\bar{y})^2}{m}$
Остаточная	$\sum(y-\hat{y}_x)^2$	$n-m-1$	$S^2_{\text{ост}} = \dfrac{\sum(y-\hat{y}_x)^2}{n-m-1}$

Определение дисперсии на одну степень свободы приводит дисперсии к сравнимому виду (напомним, что степени свободы – это числа, показывающие количество элементов варьирования, которые могут принимать произвольные значения, не изменяющие заданных характеристик). Сопоставляя факторную и остаточную дисперсии в расчете на одну степень свободы, получим величину F-критерия Фишера:

$$F = \frac{S^2_{\text{факт}}}{S^2_{\text{ост}}}$$

Фактическое значение F-критерия Фишера сравнивается с табличным значением $F_{\text{табл}}(\alpha;\ k_1;\ k_2)$ при уровне значимости α и степенях свободы $k_1 = m$ и $k_2 = n - m - 1$. При этом если фактическое значение F-критерия больше табличного, то признается статистическая значимость уравнения в целом.

Для парной линейной регрессии $m = 1$, поэтому

$$F = \frac{S^2_{\text{факт}}}{S^2_{\text{ост}}} = \frac{\sum(\hat{y}_x-\bar{y})^2}{\sum(y-\hat{y}_x)^2}\cdot(n-2)$$

Величина F-критерия связана с коэффициентом детерминации r^2_{xy}, и ее можно рассчитать по следующей формуле:

$$F = \frac{r^2_{xy}}{1-r^2_{xy}}\cdot(n-2) \tag{11}$$

Для оценки статистической значимости параметров регрессии и корреляции рассчитываются t-критерий Стъюдента и доверительные интервалы каждого из показателей. Оценка значимости коэффициентов регрессии и корреляции с помощью t-критерия Стьюдента проводится путем сопоставления их значений с величиной случайной ошибки:

$$t_b = \frac{b}{m_b}; \quad t_a = \frac{a}{m_a}; \quad t_r = \frac{r_{xy}}{m_r} \qquad (12)$$

Стандартные ошибки параметров линейной регрессии и коэффициента корреляции определяются по формулам:

$$m_b = \sqrt{\frac{\sum (y-\hat{y}_x)^2 / (n-2)}{\sum (x-\bar{x})^2}} = \sqrt{\frac{S^2_{\text{ост}}}{n \cdot \sigma_x^2}};$$

$$m_a = \sqrt{\frac{\sum (y-\hat{y}_x)^2}{(n-2)} \cdot \frac{\sum x^2}{n \sum (x-\bar{x})^2}} = \sqrt{S^2_{\text{ост}} \frac{\sum x^2}{n^2 \sigma_x^2}}; \qquad (13)$$

$$m_{r_{xy}} = \sqrt{\frac{1-r_{xy}^2}{n-2}}.$$

Сравнивая фактическое и критическое (табличное) значения t-статистики – $t_{\text{табл}}$ и $t_{\text{факт}}$ – делаем вывод о значимости параметров регрессии и корреляции. Если $t_{\text{табл}} < t_{\text{факт}}$ то параметры a, b и r_{xy} не случайно отличаются от нуля и сформировались под влиянием систематически действующего фактора x. Если $t_{\text{табл}} > t_{\text{факт}}$. то признается случайная природа формирования a, b или r_{xy}.

Для расчета доверительного интервала определяем предельную ошибку Δ для каждого показателя:

$$\Delta_a = t_{\text{табл}} m_a, \qquad \Delta_b = t_{\text{табл}} m_b.$$

Формулы для расчета доверительных интервалов имеют следующий вид:

$$\gamma_a = a \pm \Delta_a; \qquad \gamma_{a_{\min}} = a - \Delta_a; \qquad \gamma_{a_{\max}} = a + \Delta_a;$$

$$\gamma_b = b \pm \Delta_b; \qquad \gamma_{b_{\min}} = b - \Delta_b; \qquad \gamma_{b_{\max}} = b + \Delta_b;$$

Если в границы доверительного интервала попадает ноль, т. е. нижняя граница отрицательна, а верхняя положительна, то оцениваемый параметр принимается нулевым, так как он не может одновременно принимать и положительное, и отрицательное значения.

Связь между F-критерием Фишера и t-статистикой Стьюдента выражается равенством

$$\left|t_r\right| = \left|t_b\right| = \sqrt{F} \tag{14}$$

В прогнозных расчетах по уравнению регрессии определяется предсказываемое индивидуальное значение y_0 как точечный прогноз при $x = x_0$, т. е. путем подстановки в линейное уравнение $\hat{y}_x = a + b \cdot x$ соответствующего значения x. Однако точечный прогноз явно нереален, поэтому он дополняется расчетом стандартной ошибки:

$$m_{\hat{y}_0} = \sqrt{S^2_{\text{ост}}\left(1 + \frac{1}{n} + \frac{\left(x_p - \overline{x}\right)^2}{\sum\left(x - \overline{x}\right)^2}\right)} = \sqrt{S^2_{\text{ост}}\left(1 + \frac{1}{n} + \frac{\left(x_p - \overline{x}\right)^2}{n \cdot \sigma_x^2}\right)} \tag{15}$$

где $S^2_{\text{ост}} = \dfrac{\sum\left(y - \hat{y}_x\right)^2}{n-2}$, и построением доверительного интервала прогнозного значения y^*_0:

$$\hat{y}_0 - m_{\hat{y}_0} \cdot t_{\text{табл}} \le y^*_0 \le \hat{y}_x + m_{\hat{y}_x} \cdot t_{\text{табл}}$$

3. РЕШЕНИЕ ТИПОВОЙ ЗАДАЧИ

По территориям региона приводятся данные в табл. 2 за 199Х г.

Требуется:

1. Построить линейное уравнение парной регрессии y по x.
2. Рассчитать линейный коэффициент парной корреляции, коэффициент детерминации и среднюю ошибку аппроксимации.
3. Оценить статистическую значимость уравнения регрессии в целом и отдельных параметров регрессии и корреляции с помощью F-критерия Фишера и t-критерия Стьюдента.
4. Выполнить прогноз заработной платы y при прогнозном значении среднедушевого прожиточного минимума x, составляющем 107% от среднего уровня.
5. Оценить точность прогноза, рассчитав ошибку прогноза и его доверительный интервал.
6. На одном графике отложить исходные данные и теоретическую прямую.

Таблица 2

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	78	133
2	82	148
3	87	134
4	79	154
5	89	162
6	106	195
7	67	139
8	88	158
9	73	152
10	87	162
11	76	159
12	115	173

Решение:

1. Для расчета параметров уравнения линейной регрессии строим расчетную таблицу 3.

Таблица 3

№	***x***	***y***	***x·y***	x^2	y^2	$\hat{y}_x$	$y-\hat{y}_x$	$\left(y-\hat{y}_x\right)^2$	A_i
1	***2***	***3***	***4***	***6***	***6***	***7***	***8***	***9***	***10***
1	78	133	10374	6084	17689	148,78	–15,78	249,01	11,86
2	82	148	12136	6724	21904	152,46	–4,46	19,89	3,01
3	87	134	11658	7569	17956	157,06	–23,06	531,76	17,21
4	79	154	12166	6241	23716	149,70	4,30	18,49	2,79
5	89	162	14418	7921	26244	158,90	3,10	9,61	1,91
6	106	195	20670	11236	38025	174,54	20,46	418,61	10,49
7	67	139	9313	4489	19321	138,66	0,34	0,12	0,24
8	88	158	13904	7744	24964	157,98	0,02	0,00	0,01
9	73	152	11096	5329	23104	144,18	7,82	61,15	5,14
10	87	162	14094	7569	26244	157,06	4,94	24,40	3,05
11	76	159	12084	5776	25281	146,94	12,06	145,44	7,58
12	115	173	19895	13225	29929	182,82	–9,82	96,43	5,68
Итого	1027	1869	161808	89907	294377	1869,08	–0,08	1574,91	68,97
Среднее значение	85,58	155,75	13484,0	7492,25	24531,4	155,76	–	131,24	5,75

№	x	y	x·y	x^2	y^2	$\hat{y}_x$	$y-\hat{y}_x$	$(y-\hat{y}_x)^2$	A_i
1	**2**	**3**	**4**	**6**	**6**	**7**	**8**	**9**	**10**
σ	12,97	16,53	–	–	–	–	–	–	–
σ^2	168,31	273,34	–	–	–	–	–	–	–

По формулам (5) находим параметры регрессии

$$b=\frac{\overline{y\cdot x}-\overline{y}\cdot\overline{x}}{\overline{x^2}-\overline{x}^2}=\frac{13484-155,75\cdot 85,58}{7492,25-85,58^2}=\frac{154,915}{168,31}=0,92;$$

$$a=\overline{y}-b\cdot\overline{x}=155,75-0,92\cdot 85,58=77,02.$$

Получено уравнение регрессии:

$$y = 77,02 + 0,92\cdot x.$$

Параметр регрессии позволяет сделать вывод, что с увеличением среднедушевого прожиточного минимума на 1 руб. среднедневная заработная плата возрастает в среднем на 0,92 руб. (или 92 коп.).

После нахождения уравнения регрессии заполняем столбцы 7–10 таблицы 3.

2. Тесноту линейной связи оценит коэффициент корреляции (6):

$$r_{xy}=b\cdot\frac{\sigma_x}{\sigma_y}=0,92\cdot\frac{12,97}{16,53}=0,722;$$

Так как значение коэффициента корреляции больше 0,7, то это говорит о наличии весьма тесной линейной связи между признаками.

Коэффициент детерминации:

$$r_{xy}^2=0,521.$$

Это означает, что 52% вариации заработной платы (y) объясняется вариацией фактора x – среднедушевого прожиточного минимума.

Качество модели определяет средняя ошибка аппроксимации (7):

$$\overline{A}=\frac{1}{n}\sum A_i=\frac{68,97}{12}=5,75\%.$$

Качество построенной модели оценивается как хорошее, так как $\overline{A}$ не превышает 10%.

3. Оценку статистической значимости уравнения регрессии в целом проведем с помощью F-критерия Фишера. Фактическое значение F-критерия по формуле (11) составит

$$F_{\text{факт}} = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot (n-2) = \frac{0{,}521}{1 - 0{,}521} \cdot 10 = 10{,}88 .$$

Табличное значение критерия при пятипроцентном уровне значимости и степенях свободы $k_1 = 1$ и k_2=12–2 = 10 составляет $F_{\text{табл}}$ = 4,96. Так как $F_{\text{факт}}$ = 10,41 > $F_{\text{табл}}$ = 4,96, то уравнение регрессии признается статистически значимым.

Оценку статистической значимости параметров регрессии и корреляции проведем с помощью t-статистики Стьюдента и путем расчета доверительного интервала каждого из параметров.

Табличное значение t-критерия для числа степеней свободы $df = n - 2 = 12 - 2 = 10$ и уровня значимости 0,05 составит $t_{\text{табл}}$ = 2,23.

Определим стандартные ошибки m_a, m_b, $m_{r_{xy}}$ (остаточная дисперсия на одну степень свободы $S^2_{\text{ост}} = \frac{\sum (y - \hat{y}_x)^2}{n-2} = \frac{1574{,}91}{10} = 157{,}49$):

$$m_a = \sqrt{S^2_{\text{ост}} \frac{\sum x^2}{n^2 \sigma_x^2}} = \sqrt{157{,}49 \cdot \frac{89907}{12^2 \cdot 164{,}94}} = 24{,}42;$$

$$m_b = \sqrt{\frac{S^2_{\text{ост}}}{n \cdot \sigma_x^2}} = \sqrt{\frac{157{,}49}{12 \cdot 164{,}94}} = 0{,}282;$$

$$m_{r_{xy}} = \sqrt{\frac{1 - r_{xy}^2}{n-2}} = \sqrt{\frac{1 - 0{,}521}{12 - 2}} = 0{,}219 .$$

Тогда:

$$t_a = \frac{a}{m_a} = \frac{77{,}02}{24{,}42} = 3{,}15;$$

$$t_b = \frac{b}{m_b} = \frac{0{,}92}{0{,}282} = 3{,}26;$$

$$t_{r_{xy}} = \frac{r_{xy}}{m_{r_{xy}}} = \frac{0{,}722}{0{,}219} = 3{,}30 .$$

Фактические значения t – статистики превосходят табличные значения:

$$t_a = 3{,}26 > t_{\text{табл}} = 2{,}3;\; t_b = 3{,}16 > t_{\text{табл}} = 2{,}3;\; t_{r_{xy}} = 3{,}25 > t_{\text{табл}} = 2{,}3,$$

поэтому параметры a, b и r_{xy} не случайно отличаются от нуля, а статистически значимы.

Рассчитаем доверительные интервалы для параметров регрессии a и b. Для этого определим предельную ошибку для каждого показателя:

$$\Delta_a = t_{\text{табл}} \cdot m_a = 2{,}23 \cdot 24{,}42 = 54{,}46;$$

$$\Delta_b = t_{\text{табл}} \cdot m_b = 2{,}23 \cdot 0{,}282 = 0{,}63.$$

Доверительные интервалы

$$\gamma_a = a \pm \Delta_a = 77{,}02 \pm 54{,}46 \text{ и } 22{,}56 \le a^* \le 131{,}48;$$

$$\gamma_b = b \pm \Delta_b = 0{,}92 \pm 0{,}63 \text{ и } 0{,}29 \le b^* \le 1{,}55$$

Анализ верхней и нижней границ доверительных интервалов приводит к выводу о том, что с вероятностью $\rho = 1 - \alpha = 0{,}95$ параметры a и b, находясь в указанных границах, не принимают нулевых значений, т. е. являются статистически значимыми и существенно отличны от нуля.

4. Полученные оценки уравнения регрессии позволяют использовать его для прогноза. Если прогнозное значение прожиточного минимума составит: $x_0 = \bar{x} \cdot 1{,}07 = 85{,}6 \cdot 1{,}07 = 91{,}6$ руб., тогда индивидуальное прогнозное значение заработной платы составит:

$$\hat{y}_0 = 77{,}02 + 0{,}92 \cdot 91{,}6 = 161{,}29 \text{ руб.}$$

5. Ошибка прогноза составит:

$$m_{\hat{y}_0} = \sqrt{S^2_{\text{ост}}\left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{n \cdot \sigma_x^2}\right)} = \sqrt{157{,}49 \cdot \left(1 + \frac{1}{12} + \frac{(91{,}6 - 85{,}6)^2}{12 \cdot 164{,}94}\right)} = 13{,}17.$$

Предельная ошибка прогноза, которая в 95% случаев не будет превышена, составит:

$$\Delta_{\hat{y}_0} = t_{\text{табл}} \cdot m_{\hat{y}_0} = 2{,}23 \cdot 13{,}17 = 29{,}37.$$

Доверительный интервал прогноза:

$$\gamma_{\hat{y}_0} = \hat{y}_0 \pm \Delta_{\hat{y}_0} = 161{,}29 \pm 29{,}37 \text{ и } 131{,}92 \le y_0^* \le 190{,}66.$$

Выполненный прогноз среднемесячной заработной платы является надежным ($\rho = 1 - \alpha = 1 - 0{,}05 = 0{,}95$) и находится в пределах от 131,92 руб. до 190,66 руб.

6. В заключение решения задачи построим на одном графике исходные данные и теоретическую прямую (рис. 1).

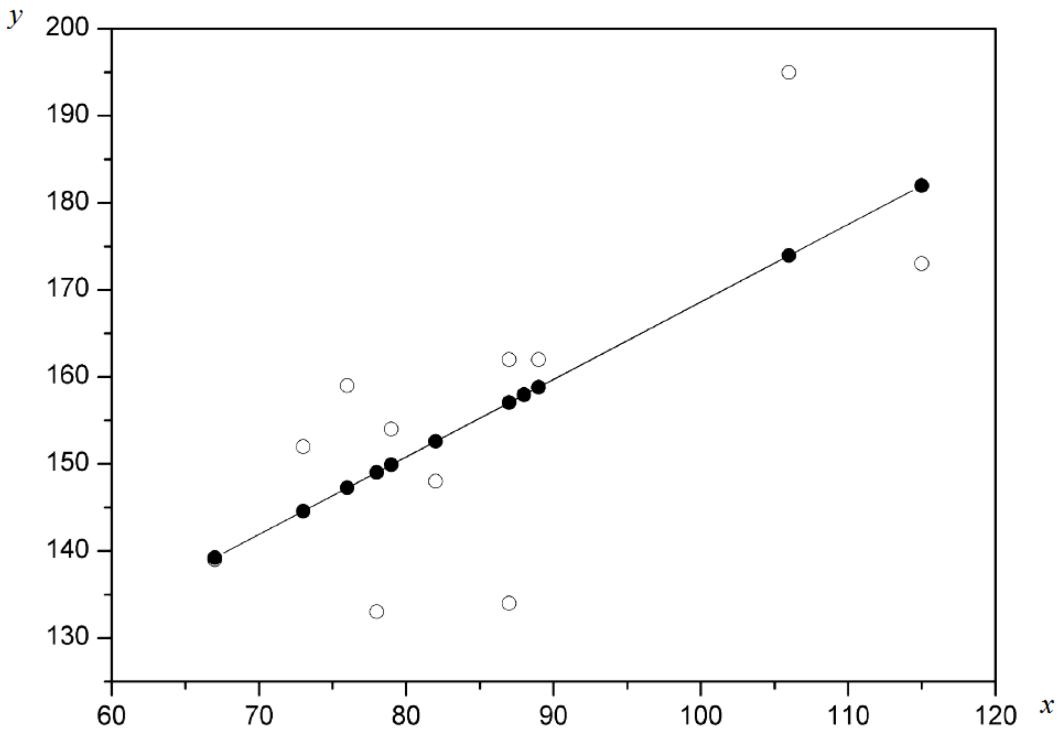


Рис. 1. Исходные данные и теоретическая прямая

4. РЕШЕНИЕ ТИПОВОЙ ЗАДАЧИ В MS EXCEL

С помощью инструмента анализа данных Регрессия можно получить результаты регрессионной статистики, дисперсионного анализа, доверительных интервалов, остатки и графики подбора линии регрессии.

Если в меню сервис еще нет команды Анализ данных, то необходимо сделать следующее. В главном меню последовательно выбираем Сервис → Надстройки и устанавливаем «флажок» в строке Пакет анализа (рис. 2).

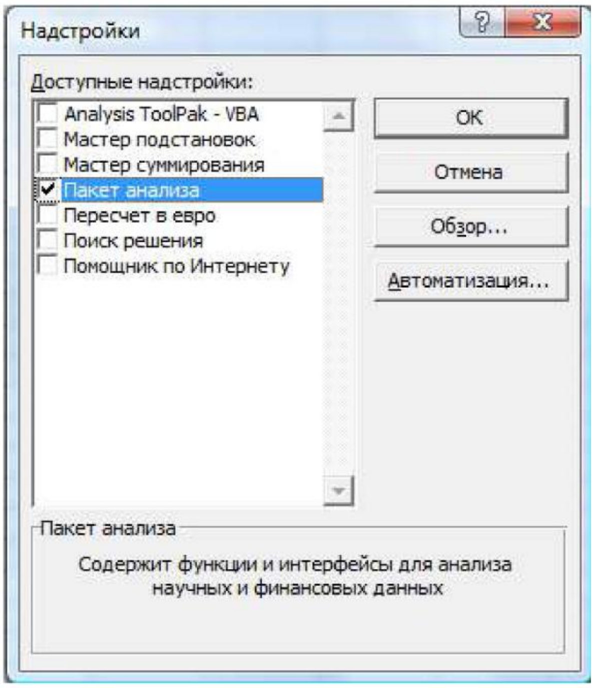


Рис. 2. Активация пакета «Анализ данных»

Далее следуем по следующему плану.

1. Если исходные данные уже внесены, то выбираем Сервис → Анализ данных → Регрессия.

2. Заполняем диалоговое окно ввода данных и параметров вывода (рис. 3).

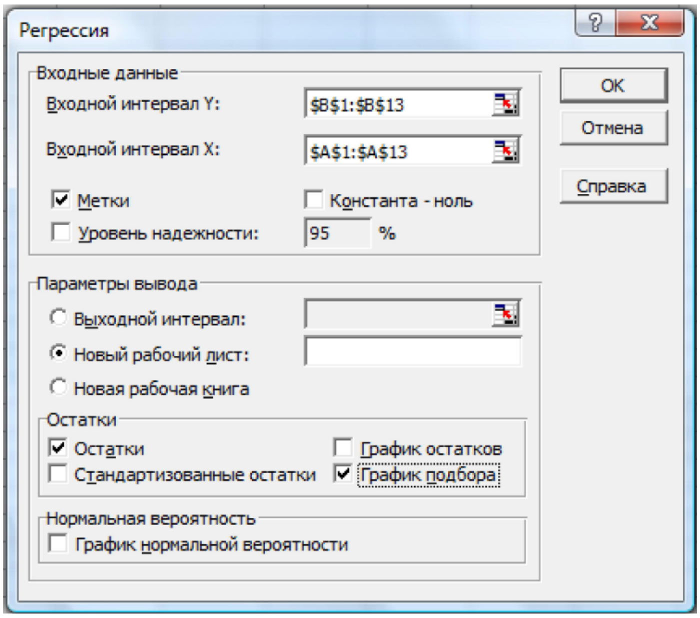


Рис. 3. Диалоговое окно ввода данных и параметров вывода

Здесь:

Входной интервал Y – диапазон, содержащий данные результативного признака;

Входной интервал X – диапазон, содержащий данные признака-фактора;

Метки – «флажок», который указывает, содержи ли первая строка названия столбцов;

Константа - ноль – «флажок», указывающий на наличие или отсутствие свободного члена в уравнении;

Выходной интервал – достаточно указать левую верхнюю ячейку будущего диапазона;

Новый рабочий лист – можно указать произвольное имя нового листа (или не указывать, тогда результаты выводятся на вновь созданный лист).

Получаем следующие результаты для рассмотренного выше примера:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ВЫВОД ИТОГОВ												
2													
3	*Регрессионная статистика*												
4	Множественный R	0,721025214											
5	R-квадрат	0,519877359											
6	Нормированный R-кв	0,471865095											
7	Стандартная ошибка	12,5495908											
8	Наблюдения	12											
9													
10	Дисперсионный анализ												
11		*df*	*SS*	*MS*	*F*	*Значимость F*							
12	Регрессия	1	1705,327706	1705,328	10,82801	0,008142							
13	Остаток	10	1574,922294	157,4922									
14	Итого	11	3280,25										
15													
16		*Коэффициенты*	*Стандартная*	*t-статисти*	*P-Значени*	*Нижние 95*	*Верхние 9*	*Нижние 95*	*Верхние 95,0%*				
17	Y-пересечение	76,9764852	24,21156138	3,179328	0,009831	23,02976	130,9232	23,02976	130,9232				
18	x	0,920430553	0,279715587	3,290594	0,008142	0,297185	1,543676	0,297185	1,543676				
19													
20													
21													
22	ВЫВОД ОСТАТКА												
23													
24	*Наблюдение*	*Предсказанное y*	*Остатки*										
25	1	148,7700683	-15,77006831										
26	2	152,4517905	-4,45179052										
27	3	157,0539433	-23,05394328										
28	4	149,6904989	4,309501138										
29	5	158,8948044	3,105195612										
30	6	174,5421238	20,45787622										
31	7	138,6453322	0,354667771										
32	8	157,9743738	0,025626164										
33	9	144,1679155	7,832084455										
34	10	157,0539433	4,946056717										
35	11	146,9292072	12,0707928										
36	12	182,8259988	-9,825998758										

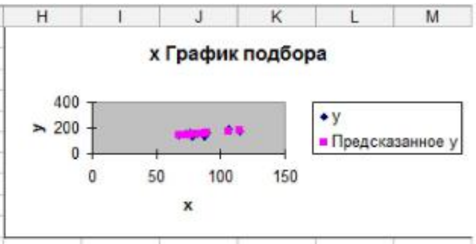


Рис. 4. Результаты анализа данных

Откуда выписываем, округляя до 4 знаков после запятой и переходя к нашим обозначениям:

Уравнение регрессии:

$$\hat{y}_x = 76,9765 + 0,9204x .$$

Коэффициент корреляции:

$$r_{xy} = 0,7210 .$$

Коэффициент детерминации:

$$r_{xy}^2 = 0,5199 .$$

Фактическое значение F-критерия Фишера:

$$F = 10,8280$$

Остаточная дисперсия на одну степень свободы:

$$S_{\text{ост}}^2 = 157,4922$$

Корень квадратный из остаточной дисперсии (стандартная ошибка):

$$S_{\text{ост}} = 12,5496 .$$

Стандартные ошибки для параметров регрессии:

$$m_a = 24,2116, \qquad m_b = 0,2797 .$$

Фактические значения t-критерия Стьюдента:

$$t_a = 3,1793, \qquad t_b = 3,2906 .$$

Доверительные интервалы:

$$23,0298 \le a^* \le 130,9232,$$

$$0,2972 \le b^* \le 1,5437 .$$

Как видим, найдены все рассмотренные выше параметры и характеристики уравнения регрессии, за исключением средней ошибки аппроксимации (значение t-критерия Стьюдента для коэффициента корреляции совпадает с t_b). Результаты «ручного счета» от машинного отличаются незначительно (отличия связаны с ошибками округления).

5. ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ

1. Изучить основные теоретические положения.
2. Оформить отчет. Отчет должен содержать: наименование и цель работы; описание основных теоретических положений; ответы на контрольные вопросы.
3. По заданию преподавателя (см. приложение 1) выполнить расчеты по парной регрессии и корреляции, сделать выводы.

Отчет по практическому занятию оформляется в тетради для практических занятий и в виде рабочей книги MS Excel и должен включать результаты выполнения индивидуальных заданий.

Результаты выполнения индивидуального задания сохраняются в электронном обучении, и представляется студентом преподавателю для проверки и последующей защиты. Защита результатов практического занятия производится студентом только индивидуально. В ходе защите работы студент отвечает на вопросы преподавателя (поясняет методику выполнения заданий, отвечает на контрольные вопросы и т. д.).

6. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что такое парная регрессия?
2. Какие бывают регрессии?
3. Как осуществляется построение уравнения регрессии?
4. Что такое дисперсия?
5. Что такое математическое ожидание?
6. Опишите алгоритм решения типовой задачи.
7. Опишите алгоритм решения типовой задачи в MS EXCEL.

7. СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

1. Адлер, Ю. П. Планирование эксперимента при поиске оптимальных условий / Ю. П. Адлер и др. – Москва: Наука, 1976.

2. Математические методы и модели исследования операций [Электронный ресурс]. – Москва : Юнити-Дана, 2015. – 592 с. – Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=114719. – Загл. с экрана.

3. Математические методы исследования [Электронный ресурс] / сост.: Э. Н. Огнева – Кемерово: КемГУКИ, 2014. – 98 с. – Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=275375. – Загл. с экрана.

Приложение 1

Задача. По территориям региона приводятся данные за 199Х г. (p_1 – число букв в полном имени, p_2 – число букв в фамилии):

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., ***x***	Среднедневная заработная плата, руб., ***y***
1	$78 + p_1$	$133 + p_2$
2	$80 + p_2$	148
3	87	$135 + p_1$
4	79	154
5	106	$157 + p_1$
6	$106 + p_1$	195
7	61	139
8	98	$158 + p_2$
9	$73 + p_2$	152
10	87	162
11	86	$146 + p_2$
12	$110 + p_2$	173

Требуется:

1. Построить линейное уравнение парной регрессии по x.
2. Рассчитать линейный коэффициент парной корреляции, коэффициент детерминации и среднюю ошибку аппроксимации.
3. Оценить статистическую значимость уравнения регрессии в целом и отдельных параметров регрессии и корреляции с помощью F-критерия Фишера и t-критерия Стьюдента.
4. Выполнить прогноз заработной платы y при прогнозном значении среднедушевого прожиточного минимума x, составляющем 107% от среднего уровня.
5. Оценить точность прогноза, рассчитав ошибку прогноза и его доверительный интервал.
6. На одном графике отложить исходные данные и теоретическую прямую.
7. Проверить вычисления в MS Excel.

ПРАКТИЧЕСКОЕ ЗАНЯТИЕ №7 «МНОЖЕСТВЕННАЯ РЕГРЕССИЯ И КОРРЕЛЯЦИЯ»

Составитель: Дубинкин Д. М.

1. ЦЕЛЬ ПРАКТИЧЕСКОГО ЗАНЯТИЯ

Получить практические навыки множественной регрессии и корреляции. Ознакомиться с методикой анализа множественной регрессии и корреляции.

2. КРАТКИЕ ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

Множественная регрессия – это уравнение связи с несколькими независимыми переменными:

$$y = f\left(x_1, x_2, ..., x_m\right) + \varepsilon, \tag{1}$$

где y – зависимая переменная (результативный признак); x_1, x_2, … x_m – независимые переменные (признаки-факторы); ε – случайная величина, характеризующая отклонения реального значения результативного признака от теоретического, найденного по уравнению регрессии..

Для построения уравнения множественной регрессии чаще используются следующие функции:

– линейная:

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + ... + b_m \cdot x_m + \varepsilon \tag{2}$$

– степенная:

$$y = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot ... \cdot x_m^{b_m} \cdot \varepsilon \tag{3}$$

– экспонента:

$$y = \mathrm{e}^{a + b_1 \cdot x_1 + b_2 \cdot x_2 + ... + b_m \cdot x_m + \varepsilon} \tag{4}$$

– гипербола:

$$y = \frac{1}{a + b_1 \cdot x_1 + b_2 \cdot x_2 + ... + b_m \cdot x_m + \varepsilon} \tag{5}$$

Можно использовать и другие функции, приводимые к линейному виду.

Для оценки параметров уравнения множественной регрессий применяют метод наименьших квадратов (МНК). Для линейных уравнений:

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \ldots + b_m \cdot x_m + \varepsilon \tag{6}$$

строится следующая система нормальных уравнений, решение которой позволяет получить оценки параметров регрессии:

$$\begin{cases} \sum y = na + b_1 \sum x_1 + b_2 \sum x_2 + \ldots + b_m \sum x_m, \\ \sum yx_1 = a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 + \ldots + b_m \sum x_m x_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \sum yx_m = a \sum x_m + b_1 \sum x_1 x_m + b_2 \sum x_2 x_m + b_m \sum x_m^2. \end{cases} \tag{7}$$

Для двухфакторной модели данная система будет иметь вид:

$$\begin{cases} na + b_1 \sum x_1 + b_2 \sum x_2 = \sum y, \\ a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 = \sum yx_1, \\ a \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2 = \sum yx_2. \end{cases} \tag{8}$$

Так же можно воспользоваться готовыми формулами, которые являются следствием из этой системы:

$$\begin{gathered} b_1 = \frac{\sigma_y}{\sigma_{x_1}} \cdot \frac{r_{yx_1} - r_{yx_2} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}; \\ b_2 = \frac{\sigma_y}{\sigma_{x_2}} \cdot \frac{r_{yx_2} - r_{yx_1} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}; \\ a = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2. \end{gathered} \tag{9}$$

В линейной множественной регрессии параметры при x называются коэффициентами «чистой» регрессии. Они характеризуют среднее изменение результата с изменением соответствующего фактора на единицу при неизмененном значении других факторов, закрепленных на среднем уровне.

Метод наименьших квадратов применим и к уравнению множественной регрессии в стандартизированном:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \ldots + \beta_m t_{x_m} + \varepsilon \tag{10}$$

где t_y, t_{x_1}, ..., t_{x_m} – стандартизированные переменные; β_i – стандартизированные коэффициенты регрессии.

$$t_y = \frac{y - \bar{y}}{\sigma_y}, \; t_{x_i} = \frac{x_i - \bar{x}_i}{\sigma_{x_i}} \tag{11}$$

Для t_y, t_{x_i} среднее значение равно нулю:

$$\overline{t_y} = \overline{t_{x_i}} = 0 \qquad (12)$$

Для t_y, t_{xi} среднее квадратическое отклонение равно единице:

$$\sigma_{t_y} = \sigma_{t_{x_i}} = 1 \qquad (13)$$

В силу того, что все переменные заданы как центрированные и нормированные, стандартизованные коэффициенты регрессии β_i можно сравнивать между собой. Сравнивая их друг с другом, можно ранжировать факторы по силе их воздействия на результат. В этом основное достоинство стандартизованных коэффициентов регрессии в отличие от коэффициентов «чистой» регрессии, которые несравнимы между собой.

Применяя МНК к уравнению множественной регрессии в стандартизированном масштабе, получим систему нормальных уравнений вида:

$$\begin{cases} r_{yx_1} = \beta_1 + \beta_2 r_{x_1x_2} + \beta_3 r_{x_1x_3} + \ldots + \beta_m r_{x_1x_m}, \\ r_{yx_2} = \beta_1 r_{x_1x_2} + \beta_2 + \beta_3 r_{x_1x_3} + \ldots + \beta_m r_{x_1x_m}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ r_{yx_m} = \beta_1 r_{x_1x_m} + \beta_2 r_{x_2x_m} + \beta_3 r_{x_3x_m} + \ldots + \beta_m, \end{cases} \qquad (14)$$

где r_{yxi} и $r_{xi\ xj}$ – коэффициенты парной и межфакторной корреляции.

Коэффициенты «чистой» регрессии b_i связаны со стандартизованными коэффициентами регрессии β_i следующим образом:

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}} \quad \left(\beta_i = b_i \frac{\sigma_{x_i}}{\sigma_y} \right) \qquad (15)$$

Поэтому можно переходить от уравнения регрессии в стандартизованном масштабе (10) к уравнению регрессии в натуральном масштабе переменных (6), при этом параметр a определяется как:

$$a = \overline{y} - b_1\overline{x}_1 - b_2\overline{x}_2 - \ldots - b_m\overline{x}_m \qquad (16)$$

Рассмотренный смысл стандартизованных коэффициентов регрессии позволяет их использовать при отсеве факторов – из модели исключаются факторы с наименьшим значением β_i.

Средние коэффициенты эластичности для линейной регрессии рассчитываются по формуле:

$$\overline{Э}_{yx_j} = b_j \frac{\overline{x}_j}{\overline{y}} \qquad (17)$$

Средние коэффициенты эластичности для линейной регрессии показывают, на сколько процентов в среднем изменится результат, при изменении соответствующего фактора на 1%. Средние показатели эластичности можно сравнивать друг с другом и соответственно ранжировать факторы по силе их воздействия на результат.

Тесноту совместного влияния факторов на результат оценивает индекс множественной корреляции:

$$R_{yx_1x_2\dots x_m} = \sqrt{1 - \frac{\sigma^2_{y_{\text{ост}}}}{\sigma^2_y}} \quad (18)$$

Значение индекса множественной корреляции лежит в пределах от 0 до 1 и должно быть больше или равно максимальному парному индексу корреляции:

$$R_{yx_1x_2\dots x_m} \geq r_{yx_i} \quad \left(i = \overline{1, m}\right) \quad (19)$$

При линейной зависимости коэффициент множественной корреляции можно определить через матрицы парных коэффициентов корреляции:

$$R_{yx_1x_2\dots x_m} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}} \quad (20)$$

где Δr – определитель матрицы парных коэффициентов корреляции, определяется по формуле (21); Δr_{11} – определитель матрицы межфакторной корреляции.

$$\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_m} \\ r_{yx_1} & 1 & r_{x_1x_2} & \dots & r_{x_1x_m} \\ r_{yx_2} & r_{x_2x_1} & 1 & \dots & r_{x_2x_m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{yx_m} & r_{x_mx_1} & r_{x_mx_2} & \dots & 1 \end{vmatrix} \quad (21)$$

$$\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} & \dots & r_{x_1x_m} \\ r_{x_2x_1} & 1 & \dots & r_{x_2x_m} \\ \dots & \dots & \dots & \dots \\ r_{x_mx_1} & r_{x_mx_2} & \dots & 1 \end{vmatrix} \quad (21)$$

Так же при линейной зависимости признаков формула коэффициента множественной корреляции может быть также представлена следующим выражением:

$$R_{yx_1x_2\ldots x_m} = \sqrt{\sum \beta_i \cdot r_{yx_i}} \qquad (22)$$

где β_i – стандартизованные коэффициенты регрессии; r_{yxi} – парные коэффициенты корреляции результата с каждым фактором.

Качество построенной модели в целом оценивает коэффициент (индекс) детерминации. Коэффициент множественной детерминации рассчитывается как квадрат индекса множественной корреляции $R^2_{yx_1x_2\ldots x_m}$:

Для того чтобы не допустить преувеличения тесноты связи, применяется скорректированный индекс множественной детерминации, который содержит поправку на число степеней свободы и рассчитывается по формуле:

$$\hat{R}^2 = 1 - \left(1 - R^2\right)\frac{(n-1)}{(n-m-1)} \qquad (23)$$

где n – число наблюдений, m – число факторов. При небольшом числе наблюдений нескорректированная величина коэффициента множественной детерминации R^2 имеет тенденцию переоценивать долю вариации результативного признака, связанную с влиянием факторов, включенных в регрессионную модель.

Частные коэффициенты (или индексы) корреляции, измеряющие влияние на y фактора x_i, при элиминировании (исключении влияния) других факторов, можно определить по формуле:

$$r_{yx_i \cdot x_1x_2\ldots x_{i-1}x_{i+1}\ldots x_m} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2\ldots x_i\ldots x_m}}{1 - R^2_{yx_1x_2\ldots x_{i-1}x_{i+1}\ldots x_m}}} \qquad (24)$$

или по рекуррентной формуле:

$$r_{yx_i \cdot x_1x_2\ldots x_{i-1}x_{i+1}\ldots x_m} = \frac{r_{yx_i \cdot x_1x_2\ldots x_{i-1}x_{i+1}\ldots x_{m-1}} - r_{yx_m \cdot x_1x_2\ldots x_{m-1}} \cdot r_{x_ix_m \cdot x_1x_2\ldots x_{i-1}x_{i+1}\ldots x_{m-1}}}{\sqrt{\left(1 - r^2_{yx_m \cdot x_1x_m\ldots x_{m-1}}\right)\left(1 - r^2_{x_ix_m \cdot x_1x_2\ldots x_{i-1}x_{i+1}\ldots x_{m-1}}\right)}} \qquad (25)$$

Рассчитанные по рекуррентной формуле частные коэффициенты корреляции изменяются в пределах от –1 до +1, а по формулам через множественные коэффициенты детерминации – от 0 до 1. Сравнение их друг с другом позволяет ранжировать факторы по тесноте их связи с результатом. Частные коэффициенты корреляции дают меру тесноты связи каждого фактора с результатом в чистом виде.

При двух факторах формулы (23) и (24) примут вид:

$$r_{yx_1 \cdot x_2} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - r^2_{yx_2}}}; \; r_{yx_2 \cdot x_1} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - r^2_{yx_1}}} \tag{26}$$

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{\left(1 - r^2_{yx_2}\right) \cdot \left(1 - r^2_{x_1x_2}\right)}}; \; r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{\left(1 - r^2_{yx_1}\right) \cdot \left(1 - r^2_{x_1x_2}\right)}} \tag{27}$$

Значимость уравнения множественной регрессии в целом оценивается с помощью F-критерия Фишера:

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m} \tag{28}$$

Частный F-критерий оценивает статистическую значимость присутствия каждого из факторов в уравнении. В общем виде для фактора x частный F-критерий определится как:

$$F_{x_i} = \frac{R^2_{yx_1 \ldots x_i \ldots x_m} - R^2_{yx_1 \ldots x_{i-1} x_{i+1} \ldots x_m}}{1 - R^2_{yx_1 \ldots x_i \ldots x_m}} \cdot \frac{n - m - 1}{1} \tag{29}$$

Фактическое значение частного F-критерия сравнивается с табличным при уровне значимости α и числе степеней свободы: $k_1 = 1$ и $k_2 = n - m - 1$. Если фактическое значение F_{x_i} превышает $F_{табл}$ (α, k_1, k_2)то дополнительное включение фактора х. в модель статистически оправданно и коэффициент чистой регрессии b_i при факторе x_i статистически значим.

Если же фактическое значение F_{x_i} меньше табличного, то дополнительное включение в модель фактора x_i не увеличивает существенно долю объясненной вариации признака y, следовательно, нецелесообразно его включение в модель; коэффициент регрессии при данном факторе в этом случае статистически незначим.

Оценка значимости коэффициентов чистой регрессии проводится по t-критерию Стьюдента. В этом случае, как и в парной регрессии, для каждого фактора используется формула

$$t_{b_i} = \frac{b_i}{m_{b_i}} \tag{29}$$

Для уравнения множественной регрессии (6) средняя квадратическая ошибка коэффициента регрессии может быть определена по формуле:

$$m_{b_i} = \frac{\sigma_y \cdot \sqrt{1 - R^2_{yx_1 \ldots x_m}}}{\sigma_{x_i} \cdot \sqrt{1 - R^2_{x_i x_1 \ldots x_m}}} \cdot \frac{1}{\sqrt{n - m - 1}} \tag{30}$$

где $R^2_{x_i\, x_1 \ldots\, x_m}$ – коэффициент детерминации для зависимости фактора x_i со всеми другими факторами уравнения множественной регрессии. Для двухфакторной модели (m = 2) имеем:

$$m_{b_1} = \frac{\sigma_y \cdot \sqrt{1 - R^2_{yx_1 x_2}}}{\sigma_{x_1} \cdot \sqrt{1 - r^2_{x_1 x_2}}} \cdot \frac{1}{\sqrt{n - 3}} \tag{31}$$

$$m_{b_2} = \frac{\sigma_y \cdot \sqrt{1 - R^2_{yx_1 x_2}}}{\sigma_{x_2} \cdot \sqrt{1 - r^2_{x_1 x_2}}} \cdot \frac{1}{\sqrt{n - 3}} \tag{32}$$

Существует связь между t-критерием Стьюдента и частным F-критерием Фишера:

$$\left| t_{b_i} \right| = \sqrt{F_{x_i}} \tag{33}$$

Уравнения множественной регрессии могут включать в качестве независимых переменных качественные признаки (например, профессия, пол, образование, климатические условия, отдельные регионы и т. д.). Чтобы ввести такие переменные в регрессионную модель, их необходимо упорядочить и присвоить им те или иные значения, т. е. качественные переменные преобразовать в количественные.

Такого вида сконструированные переменные принято называть фиктивными переменными. Например, включать в модель фактор «пол» в виде фиктивной переменной можно в следующем виде:

$$z = \begin{cases} 1 \ - \ \text{мужской пол,} \\ 0 \ - \ \text{женский пол.} \end{cases} \tag{34}$$

Коэффициент регрессии при фиктивной переменной интерпретируется как среднее изменение зависимой переменной при переходе от одной категории (женский пол) к другой (мужской пол) при неизменных значениях остальных параметров.

3. РЕШЕНИЕ ТИПОВОЙ ЗАДАЧИ

По 20 предприятиям региона (табл. 1)изучается зависимость выработки продукции на одного работника y (тыс. руб.) от ввода в действие новых основных фондов x_1, (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x_2 (%).

Требуется:

1. Построить линейную модель множественной регрессии. Записать стандартизованное уравнение множественной регрессии. На основе стандартизованных коэффициентов регрессии и средних коэффициентов эластичности ранжировать факторы по степени их влияния на результат.

Таблица 1

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7,0	3,9	10,0	11	9,0	6,0	21,0
2	7,0	3,9	14,0	12	11,0	6,4	22,0
3	7,0	3,7	15,0	13	9,0	6,8	22,0
4	7,0	4,0	16,0	14	11,0	7,2	25,0
5	7,0	3,8	17,0	15	12,0	8,0	28,0
6	7,0	4,8	19,0	16	12,0	8,2	29,0
7	8,0	5,4	19,0	17	12,0	8,1	30,0
8	8,0	4,4	20,0	18	12,0	8,5	31,0
9	8,0	5,3	20,0	19	14,0	9,6	32,0
10	10,0	6,8	20,0	20	14,0	9,0	36,0

2. Найти коэффициенты парной, частной и множественной корреляции. Проанализировать их.

3. Найти скорректированный коэффициент множественной детерминации. Сравнить его с нескорректированным (общим) коэффициентом детерминации.

4. С помощью F-критерия Фишера оценить статистическую надежность уравнения регрессии и коэффициента детерминации $R^2_{yx_1x_2}$.

5. С помощью t-критерия оценить статистическую значимость коэффициентов чистой регрессии.

6. С помощью частных F-критериев Фишера оценить целесообразность включения в уравнение множественной регрессии фактора x_1 после x_2 и фактора x_2 после x_1.

7. Составить уравнение линейной парной регрессии, оставив лишь один значащий фактор.

Решение:

Для удобства проведения расчетов поместим результаты промежуточных расчетов в таблицу 2:

Таблица 2

№	y	x_1	x_2	$y\,x_1$	$y\,x_2$	$x_1\,x_2$	x_1^2	x_2^2	y^2
1	7,0	3,9	10,0	27,3	70,0	39,0	15,21	100,0	49,0
2	7,0	3,9	14,0	27,:3	98,0	54,6	15,21	196,0	49,0
3	7,0	3,7	15,0	25,9	105,0	55,5	13,69	225,0	49,0
4	7,0	4,0	16,0	28,0	112,0	64,0	16,0	256,0	49,0
5	7,0	3,8	17,0	26,6	119,0	64,6	14,44	289,0	49,0
6	7,0	4,8	19,0	33,6	133,0	91,2	23,04	361,0	49,0
7	8,0	5,4	19,0	43,2	152,0	102,6	29,16	361,0	64,0
8	8,0	4,4	20,0	35,2	160,0	88,0	19,36	400,0	64,0
9	8,0	5,3	20,0	42,4	160,0	106,0	28,09	400,0	64,0
10	10,0	6,8	20,0	68,0	200,0	136,0	46,24	400,0	100,0
11	9,0	6,0	21,0	54,0	189,0	126,0	36,0	441,0	81,0
12	11,0	6,4	22,0	70,4	242,0	140,8	40,96	484,0	121,0
13	9,0	6,8	22,0	61,2	198,0	149,6	46,24	484,0	81,0
14	11,0	7,2	25,0	79,2	275,0	180,0	51,84	625,0	121,0
15	12,0	8,0	28,0	96,0	336,0	224,0	64,0	784,0	144,0
16	12,0	8,2	29,0	98,4	348,0	237,8	67,24	841,0	144,0
17	12,0	8,1	30,0	97,2	360,0	243,0	65,61	900,0	144,0
18	12,0	8,5	31,0	102,0	372,0	263,5	72,25	961,0	144,0
19	14,0	9,6	32,0	134,4	448,0	307,2	92,16	1024,0	196,0
20	14,0	9,0	36,0	126,0	504,0	324,0	81,0	1296,0	196,0
Сумма	192	123,8	446	1276,3	4581	2997,4	837,74	10828,0	1958,0
Ср. знач.	9,6	6,19	22,3	63,815	229,05	149,87	41,887	541,4	97,9

Найдем средние квадратические отклонения признаков:

$$\sigma_y = \sqrt{\overline{y^2} - \bar{y}^2} = \sqrt{97{,}9 - 9{,}6^2} = \sqrt{5{,}74} = 2{,}396\,;$$

$$\sigma_{x_1} = \sqrt{\overline{x_1^2} - \bar{x}_1^2} = \sqrt{41{,}887 - 6{,}19^2} = \sqrt{3{,}571} = 1{,}890\,;$$

$$\sigma_{x_2} = \sqrt{\overline{x_2^2} - \bar{x}_2^2} = \sqrt{541{,}4 - 22{,}3^2} = \sqrt{44{,}11} = 6{,}642\,.$$

1. Для нахождения параметров линейного уравнения множественной регрессии:

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

необходимо решить систему линейных уравнений относительно неизвестных параметров a, b_1, b_2 (8) либо воспользоваться готовыми формулами (9).

Рассчитаем сначала парные коэффициенты корреляции:

$$r_{yx_1} = \frac{\operatorname{cov}(y, x_1)}{\sigma_y \cdot \sigma_{x_1}} = \frac{63,815 - 6,19 \cdot 9,6}{1,890 \cdot 2,396} = 0,970;$$

$$r_{yx_2} = \frac{\operatorname{cov}(y, x_2)}{\sigma_y \cdot \sigma_{x_2}} = \frac{229,05 - 22,3 \cdot 9,6}{6,642 \cdot 2,396} = 0,941;$$

$$r_{x_1 x_2} = \frac{\operatorname{cov}(x_1, x_2)}{\sigma_{x_1} \cdot \sigma_{x_2}} = \frac{149,87 - 6,19 \cdot 22,3}{1,890 \cdot 6,642} = 0,943.$$

Находим по формулам (9) коэффициенты чистой регрессии и параметр a:

$$b_1 = \frac{\sigma_y}{\sigma_{x_1}} \cdot \frac{r_{yx_1} - r_{yx_2} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2} = \frac{2,396}{1,890} \cdot \frac{0,970 - 0,941 \cdot 0,943}{1 - 0,943^2} = 0,946;$$

$$b_2 = \frac{\sigma_y}{\sigma_{x_2}} \cdot \frac{r_{yx_2} - r_{yx_1} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2} = \frac{2,396}{6,642} \cdot \frac{0,941 - 0,970 \cdot 0,943}{1 - 0,943^2} = 0,0856;$$

$$a = \overline{y} - b_1 \overline{x}_1 - b_2 \overline{x}_2 = 9,6 - 0,946 \cdot 6,19 - 0,0856 \cdot 22,3 = 1,835.$$

Таким образом, получили следующее уравнение множественной регрессии:

$$\hat{y} = 1,835 + 0,946 \cdot x_1 + 0,0856 \cdot x_2.$$

Уравнение регрессии показывает, что при увеличении ввода в действие основных фондов на 1% (при неизменном уровне удельного веса рабочих высокой квалификации) выработка продукции на одного рабочего увеличивается в среднем на 0,946 тыс. руб., а при увеличении удельного веса рабочих высокой квалификации в общей численности рабочих на 1% (при неизменном уровне ввода в действие новых основных фондов) выработка продукции на одного рабочего увеличивается в среднем на 0,086 тыс. руб.

После нахождения уравнения регрессии составим таблицу 3 для определения теоретических значений результативного признака, остаточной дисперсии и средней ошибки аппроксимации.

Таблица 3

№	y	x_1	x_2	$\hat{y}$	$y-\hat{y}$	$(y-\hat{y})^2$	A_i, %
1	7,0	3,9	10,0	6,380	0,620	0,384	8,851
2	7,0	3,9	14,0	6,723	0,277	0,077	3,960
3	7,0	3,7	15,0	6,619	0,381	0,145	5,440
4	7,0	4,0	16,0	6,989	0,011	0,000	0,163
5	7,0	3,8	17,0	6,885	0,115	0,013	1,643
6	7,0	4,8	19,0	8,002	–1,002	1,004	14,317
7	8,0	5,4	19,0	8,570	–0,570	0,325	7,123
8	8,0	4,4	20,0	7,709	0,291	0,084	3,633
9	8,0	5,3	20,0	8,561	–0,561	0,315	7,010
10	10,0	6,8	20,0	9,980	0,020	0,000	0,202
11	9,0	6,0	21,0	9,309	–0,309	0,095	3,429
12	11,0	6,4	22,0	9,773	1,227	1,507	11,158
13	9,0	6,8	22,0	10,151	–1,151	1,325	12,789
14	11,0	7,2	25,0	10,786	0,214	0,046	1,944
15	12,0	8,0	28,0	11,800	0,200	0,040	1,668
16	12,0	8,2	29,0	12,075	–0,075	0,006	0,622
17	12,0	8,1	30,0	12,066	–0,066	0,004	0,547
18	12,0	8,5	31,0	12,530	–0,530	0,280	4,413
19	14,0	9,6	32,0	13,656	0,344	0,118	2,459
20	14,0	9,0	36,0	13,431	0,569	0,324	4,067
Сумма	192	123,8	446	191,992	0,008	6,093	95,437
Ср. знач.	9,6	6,19	22,3	9,6	–	0,305	4,77

Остаточная дисперсия:

$$\sigma^2_{\text{ост}} = \frac{\sum\left(y-\hat{y}\right)^2}{n} = \frac{6,093}{20} = 0,305.$$

Средняя ошибка аппроксимации:

$$\bar{A} = \frac{1}{n}\sum\left|\frac{y-\hat{y}}{y}\right| \cdot 100\% = \frac{95,437\%}{20} = 4,77\%.$$

Качество модели, исходя из относительных отклонений по каждому наблюдению, признается хорошим, т. к. средняя ошибка аппроксимации не превышает 10%.

Коэффициенты β_1 и β_2 стандартизованного уравнения регрессии находятся по формуле (15):

$$\hat{t}_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \varepsilon$$

$$\beta_1 = b_1 \frac{\sigma_{x_1}}{\sigma_y} = 0,946 \cdot \frac{1,890}{2,396} = 0,746;$$

$$\beta_2 = b_2 \frac{\sigma_{x_2}}{\sigma_y} = 0,0856 \cdot \frac{6,642}{2,396} = 0,237.$$

То есть уравнение будет выглядеть следующим образом:

$$\hat{t}_y = 0,746 \cdot t_{x_1} + 0,237 \cdot t_{x_2}.$$

Так как стандартизованные коэффициенты регрессии можно сравнивать между собой, то можно сказать, что ввод в действие новых основных фондов оказывает большее влияние на выработку продукции, чем удельный вес рабочих высокой квалификации.

Сравнивать влияние факторов на результат можно также при помощи средних коэффициентов эластичности (17):

$$\bar{Э}_i = b_i \cdot \frac{\bar{x}_i}{\bar{y}_{x_i}}.$$

$$\bar{Э}_1 = 0,946 \cdot \frac{6,19}{9,6} = 0,61; \ \bar{Э}_2 = 0,0856 \cdot \frac{22,3}{9,6} = 0,20.$$

То есть увеличение только основных фондов (от своего среднего значения) или только удельного веса рабочих высокой квалификации на 1% увеличивает в среднем выработку продукции на 0,61% или 0,20% соответственно. Таким образом, подтверждается большее влияние на результат y фактора x_1, чем фактора x_2.

2. Коэффициенты парной корреляции мы уже нашли:

$$r_{yx_1} = 0,970; \ r_{yx_2} = 0,941; \ r_{x_1x_2} = 0,943.$$

Они указывают на весьма сильную связь каждого фактора с результатом, а также высокую межфакторную зависимость (факторы x_1 и x_2 явно коллинеарны, т. к. $r_{x1\,x2} = 0,943 > 0,7$). При такой сильной межфакторной зависимости рекомендуется один из факторов исключить из рассмотрения.

Частные коэффициенты корреляции характеризуют тесноту связи между результатом и соответствующим фактором при элиминировании (устранении влияния) других факторов, включенных в уравнение регрессии.

При двух факторах частные коэффициенты корреляции рассчитываются следующим образом:

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{\left(1 - r_{yx_2}^2\right) \cdot \left(1 - r_{x_1x_2}^2\right)}} = \frac{0,970 - 0,941 \cdot 0,943}{\sqrt{\left(1 - 0,941^2\right) \cdot \left(1 - 0,943^2\right)}} = 0,734;$$

$$r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{\left(1 - r_{yx_1}^2\right) \cdot \left(1 - r_{x_1x_2}^2\right)}} = \frac{0,941 - 0,970 \cdot 0,943}{\sqrt{\left(1 - 0,970^2\right) \cdot \left(1 - 0,943^2\right)}} = 0,325.$$

Если сравнить коэффициенты парной и частной корреляции, то можно увидеть, что из-за высокой межфакторной зависимости коэффициенты парной корреляции дают завышенные оценки тесноты связи. Именно по этой причине рекомендуется при наличии сильной коллинеарности (взаимосвязи) факторов исключать из исследования тот фактор, у которого теснота парной зависимости меньше, чем теснота межфакторной связи.

Коэффициент множественной корреляции определить через матрицы парных коэффициентов корреляции (18):

$$R_{yx_1x_2} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}},$$

где Δr – определитель матрицы парных коэффициентов корреляции:

$$\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} \\ r_{yx_1} & 1 & r_{x_1x_2} \\ r_{yx_2} & r_{x_2x_1} & 1 \end{vmatrix}$$

Δr_{11} – определитель матрицы межфакторной корреляции:

$$\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} \\ r_{x_2x_1} & 1 \end{vmatrix}$$

Находим:

$$\Delta r = \begin{vmatrix} 1 & 0,970 & 0,941 \\ 0,970 & 1 & 0,943 \\ 0,941 & 0,943 & 1 \end{vmatrix} = 1 + 0,8607 + 0,8607 -$$

$$-0,8855 - 0,8892 - 0,9409 = 0,0058;$$

$$\Delta r_{11} = \begin{vmatrix} 1 & 0,943 \\ 0,943 & 1 \end{vmatrix} = 1 - 0,8892 = 0,1108.$$

Коэффициент множественной корреляции:

$$R_{yx_1x_2} = \sqrt{1 - \frac{0,0058}{0,1108}} = 0,973.$$

Аналогичный результат получим при использовании формул (17) и (20):

$$R_{yx_1x_2} = \sqrt{1 - \frac{\sigma^2_{\text{ост}}}{\sigma^2_y}} = \sqrt{1 - \frac{0,305}{5,74}} = 0,973;$$

$$R_{yx_1x_2} = \sqrt{\sum \beta_i \cdot r_{yx_i}} = \sqrt{0,746 \cdot 0,970 + 0,237 \cdot 0,941} = 0,973$$

Коэффициент множественной корреляции указывает на весьма сильную связь всего набора факторов с результатом.

3. Нескорректированный коэффициент множественной детерминации $R^2_{y\,x_1\,x_2} = 0,947$ оценивает долю дисперсии результата за счет представленных в уравнении факторов в общей вариации результата. Здесь эта доля составляет 94,7% и указывает на весьма высокую степень обусловленности вариации результата вариацией факторов, иными словами – на весьма тесную связь факторов с результатом.

Скорректированный коэффициент множественной детерминации:

$$\hat{R}^2 = 1 - \left(1 - R^2\right)\frac{(n-1)}{(n-m-1)} = 1 - (1 - 0,947)\frac{20-1}{20-2-1} = 0,941$$

определяет тесноту связи с учетом степеней свободы общей и остаточной дисперсий. Он дает такую оценку тесноты связи, которая не зависит от числа факторов и поэтому может сравниваться по разным моделям с разным числом факторов. Оба коэффициента указывают на весьма высокую (более 94%) детерминированность результата y в модели факторами x_1 и x_2.

4. Оценку надежности уравнения регрессии в целом и показателя тесноты связи $R_{y\,x_1\,x_2}$ дает F-критерий Фишера:

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

В нашем случае фактическое значение F -критерия Фишера:

$$F_{\text{факт}} = \frac{0{,}973^2}{1-0{,}973^2} \cdot \frac{20-2-1}{2} = 151{,}88.$$

Получили, что $F_{\text{факт}}$ = 151,88 > $F_{\text{табл}}$ = 3,59 (при n =20), т. е. вероятность случайно получить такое значение F-критерия не превышает допустимый уровень значимости 5%. Следовательно, полученное значение не случайно, оно сформировалось под влиянием существенных факторов, т. е. подтверждается статистическая значимость всего уравнения и показателя тесноты связи $R_{y\,x_1\,x_2}$.

5. Оценим статистическую значимость параметров чистой регрессии помощью t-критерия Стьюдента. Рассчитаем стандартные ошибки оэффициентов регрессии по формулам (31) и (32):

$$m_{b_1} = \frac{\sigma_y \cdot \sqrt{1-R^2_{yx_1x_2}}}{\sigma_{x_1} \cdot \sqrt{1-r^2_{x_1x_2}}} \cdot \frac{1}{\sqrt{n-3}} = \frac{2{,}396 \cdot \sqrt{1-0{,}973^2}}{1{,}890 \cdot \sqrt{1-0{,}943^2}} \cdot \frac{1}{\sqrt{20-3}} = 0{,}2132;$$

$$m_{b_2} = \frac{\sigma_y \cdot \sqrt{1-R^2_{yx_1x_2}}}{\sigma_{x_2} \cdot \sqrt{1-r^2_{x_1x_2}}} \cdot \frac{1}{\sqrt{n-3}} = \frac{2{,}396 \cdot \sqrt{1-0{,}973^2}}{6{,}642 \cdot \sqrt{1-0{,}943^2}} \cdot \frac{1}{\sqrt{20-3}} = 0{,}0607.$$

Фактические значения t-критерия Стьюдента:

$$t_{b_1} = \frac{b_1}{m_{b_1}} = \frac{0{,}946}{0{,}2132} = 4{,}44, \quad t_{b_2} = \frac{b_2}{m_{b_2}} = \frac{0{,}0856}{0{,}0607} = 1{,}41.$$

Табличное значение критерия при уровне значимости $\alpha = 0{,}05$ и числе степеней свободы $k = 17$ составит $t_{\text{табл}} = (\alpha = 0{,}05;\ k = 17) = 2{,}11$. Таким образом, признается статистическая значимость параметра b_1, т. к. $t_{b_1} > t_{\text{табл}}$ и случайная природа формирования параметра b_2, т. к. $t_{b_2} < t_{\text{табл}}$.

Доверительные интервалы для параметров чистой регрессии:

$$b_1 - m_{b_1} \cdot t_{\text{табл}} \le b_1^* \le b_1 + m_{b_1} \cdot t_{\text{табл}}, \; 0{,}496 \le b_1^* \le 1{,}396$$

$$b_2 - m_{b_2} \cdot t_{\text{табл}} \le b_2^* \le b_2 + m_{b_2} \cdot t_{\text{табл}}, \; -0{,}0425 \le b_1^* \le 0{,}2137 .$$

6. С помощью частных F-критериев Фишера оценим целесообразность включения в уравнение множественной регрессии фактора x_1 после x_2 и фактора x_2 после x_1 при помощи формул (29):

$$F_{x_1} = \frac{R^2_{yx_1x_2} - R^2_{yx_2}}{1 - R^2_{yx_1x_2}} \cdot \frac{n-m-1}{1}; \; F_{x_2} = \frac{R^2_{yx_1x_2} - R^2_{yx_1}}{1 - R^2_{yx_2x_2}} \cdot \frac{n-m-1}{1}.$$

Найдем $R^2{}_{yx1}$ и $R^2{}_{yx2}$:

$$R^2_{yx_1} = r^2_{yx_1} = 0{,}970^2 = 0{,}941;$$

$$R^2_{yx_2} = r^2_{yx_2} = 0{,}941^2 = 0{,}885.$$

Имеем:

$$F_{x_1} = \frac{0{,}947 - 0{,}885}{1 - 0{,}947} \cdot \frac{20-2-1}{1} = 19{,}89;$$

$$F_{x_2} = \frac{0{,}947 - 0{,}941}{1 - 0{,}947} \cdot \frac{20-2-1}{1} = 1{,}924.$$

Получили, что F_{x2} = 0,89 < $F_{\text{табл}}$ (α = 0,05; k_1 = 17; k_2 = 17) = 4,45. Следовательно, включение в модель фактора x_2 после того, как в модель включен фактор x_1 статистически нецелесообразно: прирост факторной дисперсии за счет дополнительного признака x_2 оказывается незначительным, несущественным; фактор x_2 включать в уравнение после фактора x_1 не следует.

Если поменять первоначальный порядок включения факторов в модель и рассмотреть вариант включения x_1 после x_2, то результат расчета частного F-критерия для x_1 будет иным F_{x1} = 17,86 > $F_{\text{табл}}$ = 4,45, т. е. вероятность его случайного формирования меньше принятого стандарта a = 0,05 (5%). Следовательно, значение частного F-критерия для дополнительно включенного фактора x_1 не случайно, является статистически значимым, надежным, достоверным: прирост факторной дисперсии за счет дополнительного фактора x_1 является существенным. Фактор x_1 должен присутствовать в уравнении, в том числе в варианте, когда он дополнительно включается после фактора x_2.

7. Общий вывод состоит в том, что множественная модель с факторами x_1 и x_2 с $R^2_{y\,x1\;x2}$ = 0,947 содержит неинформативный фактор x_2.

Если исключить фактор x_2, то можно ограничиться уравнением парной регрессии:

$$\hat{y}_{x_1} = \alpha + \beta x_1 .$$

Найдем его параметры:

$$\beta = \frac{\operatorname{cov}(y,\quad x_1)}{\sigma_{x_2}} = \frac{63{,}815 - 6{,}19 \cdot 9{,}6}{3{,}571} = 1{,}23;$$

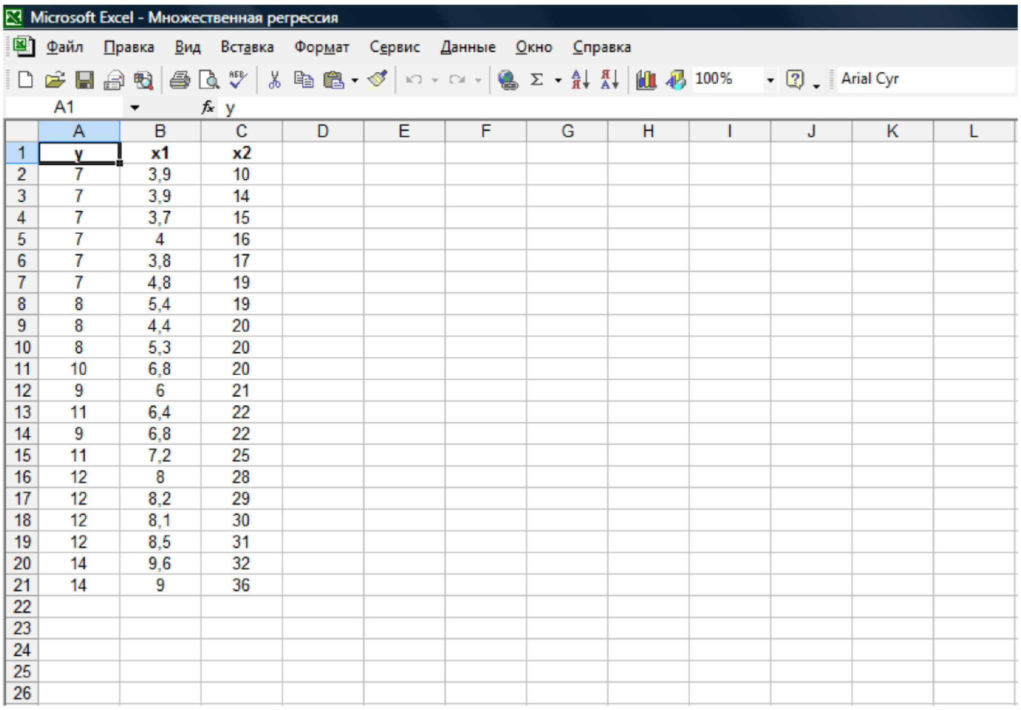
$$\alpha = \overline{y} - \beta \cdot \overline{x} = 9{,}6 - 1{,}23 \cdot 6{,}19 = 1{,}99.$$

Таким образом,

$$\hat{y}_{x_1} = 1{,}99 + 1{,}23 \cdot x_1, \qquad r^2_{yx_1} = 0{,}941.$$

4. РЕШЕНИЕ ТИПОВОЙ ЗАДАЧИ В MS EXCEL

Вносим исходные данные в таблицу MS Excel (рис. 1).



	A	B	C
1	y	x1	x2
2	7	3,9	10
3	7	3,9	14
4	7	3,7	15
5	7	4	16
6	7	3,8	17
7	7	4,8	19
8	8	5,4	19
9	8	4,4	20
10	8	5,3	20
11	10	6,8	20
12	9	6	21
13	11	6,4	22
14	9	6,8	22
15	11	7,2	25
16	12	8	28
17	12	8,2	29
18	12	8,1	30
19	12	8,5	31
20	14	9,6	32
21	14	9	36

Рис. 1. Подготовка данных к анализу

С помощью инструмента анализа данных Регрессия можно получить результаты регрессионной статистики, дисперсионного анализа, доверительных интервалов, остатки и графики подбора линии регрессии и т. п.

Если в меню сервис еще нет команды Анализ данных, то необходимо сделать следующее. В главном меню последовательно выбираем Сервис → Надстройки и устанавливаем «флажок» в строке Пакет анализа (рис. 2):

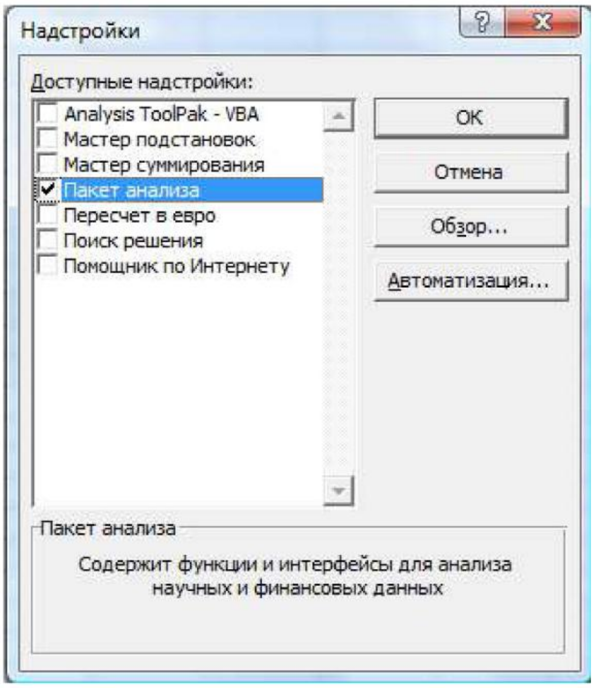


Рис. 2. Активация пакета «Анализ данных»

Далее найдем матрицу парных коэффициентов (Сервис → Анализ данных → Корреляция) (рис. 3):

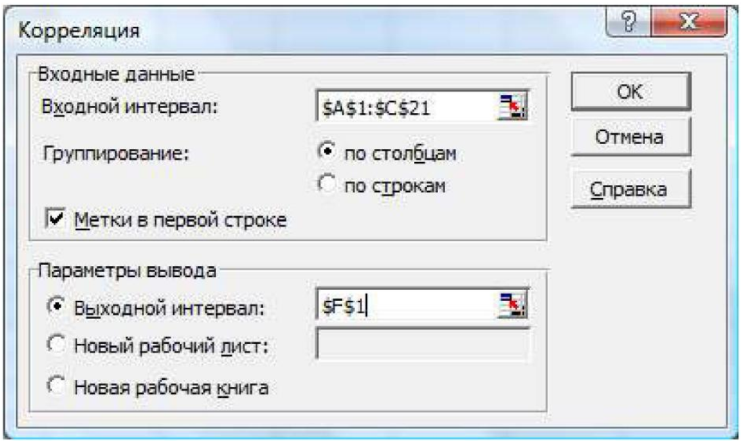


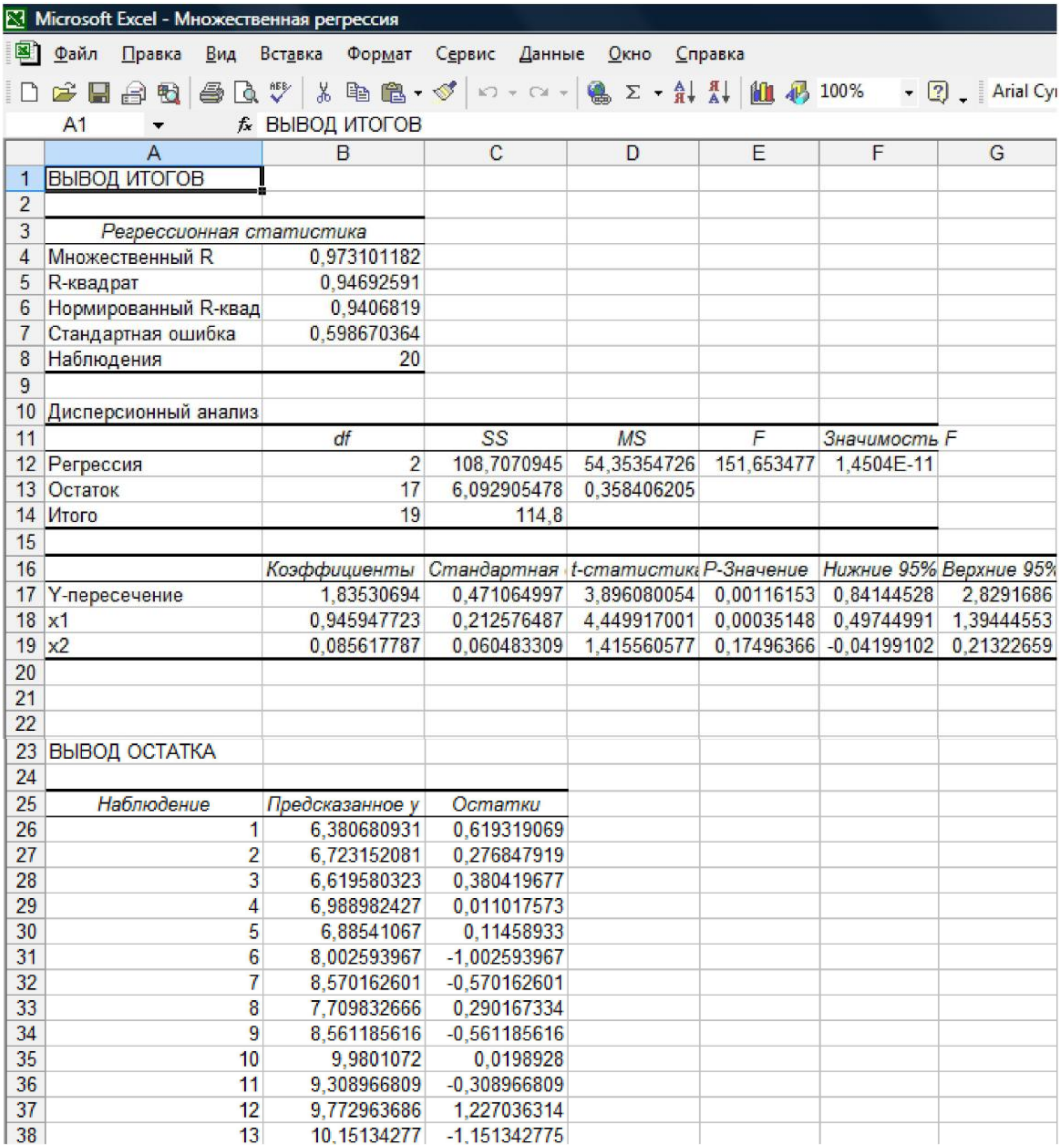
Рис. 3. Корреляция

Получаем следующий результат:

$$\begin{pmatrix} 1 & & \\ 0{,}9699 & 1 & \\ 0{,}9408 & 0{,}9428 & 1 \end{pmatrix}$$

т. е. $r_{yx1} = 0{,}9699$; $r_{y\,x2} = 0{,}9408$; $r_{x1\,x2} = 0{,}9428$.

С помощью инструмента Регрессия (Сервис → Анализ данных → Регрессия) получаем следующие результаты (рис. 4):



Microsoft Excel - Множественная регрессия

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

A1 ВЫВОД ИТОГОВ

	A	B	C	D	E	F	G
1	ВЫВОД ИТОГОВ						
2							
3	*Регрессионная статистика*						
4	Множественный R	0,973101182					
5	R-квадрат	0,94692591					
6	Нормированный R-квад	0,9406819					
7	Стандартная ошибка	0,598670364					
8	Наблюдения	20					
9							
10	Дисперсионный анализ						
11		*df*	*SS*	*MS*	*F*	*Значимость F*	
12	Регрессия	2	108,7070945	54,35354726	151,653477	1,4504E-11	
13	Остаток	17	6,092905478	0,358406205			
14	Итого	19	114,8				
15							
16		*Коэффициенты*	*Стандартная*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
17	Y-пересечение	1,83530694	0,471064997	3,896080054	0,00116153	0,84144528	2,8291686
18	x1	0,945947723	0,212576487	4,449917001	0,00035148	0,49744991	1,39444553
19	x2	0,085617787	0,060483309	1,415560577	0,17496366	-0,04199102	0,21322659
20							
21							
22							
23	ВЫВОД ОСТАТКА						
24							
25	*Наблюдение*	*Предсказанное у*	*Остатки*				
26	1	6,380680931	0,619319069				
27	2	6,723152081	0,276847919				
28	3	6,619580323	0,380419677				
29	4	6,988982427	0,011017573				
30	5	6,88541067	0,11458933				
31	6	8,002593967	-1,002593967				
32	7	8,570162601	-0,570162601				
33	8	7,709832666	0,290167334				
34	9	8,561185616	-0,561185616				
35	10	9,9801072	0,0198928				
36	11	9,308966809	-0,308966809				
37	12	9,772963686	1,227036314				
38	13	10,15134277	-1,151342775				

Рис. 4. Лист с результатами регрессионного анализа

Уравнение регрессии:

$$\hat{y} = 1{,}8353 + 0{,}9459x_1 + 0{,}0856x_2 .$$

Множественный коэффициент корреляции:

$$R=0{,}9731.$$

Коэффициент детерминации:

$$R^2 = 0{,}9469.$$

Скорректированный коэффициент детерминации:

$$\hat{R}^2 = 0{,}9407 .$$

Фактическое значение F-критерия Фишера:

$$F = 151{,}653.$$

Фактические значения t-критерия Стьюдента:

$$t_{b_1} = 4{,}450, \quad t_{b_2} = 1{,}416 .$$

Доверительные интервалы для параметров регрессии:

$$0{,}4974 \le b_1^* \le 1{,}3944 ,$$

$$-0{,}0420 \le b_2^* \le 0{,}2132 .$$

Значения частного F-критерия Фишера можно найти как квадрат соответствующего значении t-критерия Стьюдента:

$$F_{x_1} = 4{,}450^2 = 19{,}803 , \; F_{x_2} = 1{,}416^2 = 2{,}005 .$$

Оставшиеся характеристики можно найти, используя известные формулы и полученные здесь результаты.

5. ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ

1. Изучить основные теоретические положения.
2. Оформить отчет. Отчет должен содержать: наименование и цель работы; описание основных теоретических положений; ответы на контрольные вопросы.
3. По заданию преподавателя (см. приложение 1) выполнить расчеты по множественной регрессии и корреляции, сделать выводы.

Отчет по практическому занятию оформляется в тетради для практических занятий и в виде рабочей книги MS Excel и должен включать результаты выполнения индивидуальных заданий.

Результаты выполнения индивидуального задания сохраняются в электроном обучении, и представляется студентом преподавателю для проверки и последующей защиты. Защита результатов практического занятия производится студентом только индивидуально. В ходе защите работы студент отвечает на вопросы преподавателя (поясняет методику выполнения заданий, отвечает на контрольные вопросы и т. д.).

6. КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что такое парная множественная регрессия?
2. Какие бывают множественные регрессии?
3. Как осуществляется построение уравнения множественной регрессии?
4. Что такое дисперсия?
5. Что такое математическое ожидание?
6. Опишите алгоритм решения типовой задачи.
7. Опишите алгоритм решения типовой задачи в MS EXCEL.

7. СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

1. Адлер, Ю. П. Планирование эксперимента при поиске оптимальных условий / Ю. П. Адлер и др. – Москва: Наука, 1976.
2. Математические методы и модели исследования операций [Электронный ресурс]. – Москва: Юнити-Дана, 2015. – 592 с. – Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=114719. – Загл. с экрана.
3. Математические методы исследования [Электронный ресурс] / сост.: Э. Н. Огнева – Кемерово: КемГУКИ, 2014. – 98 с. – Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=275375. – Загл. с экрана.

Приложение 1

Задача.

По 20 предприятиям региона (табл. 4) изучается зависимость выработки продукции на одного работника y (тыс. руб.) от ввода в действие новых основных фондов x_1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x_2 (%) (p_1 – число букв в полном имени, p_2 – число букв в фамилии).

Таблица 4

Номер предприятия	y	x_1	x_2	Номер предприятия	y	x_1	x_2
1	7,0	$3{,}6+0{,}1\cdot p_1$	11,0	11	9,0	$6{,}0+0{,}1\cdot p_2$	21,0
2	7,0	3,7	13,0	12	11,0	6,4	22,0
3	7,0	3,9	15,0	13	9,0	6,9	22,0
4	7,0	4,0	17,0	14	11,0	7,2	25,0
5	7,0	$3{,}8+0{,}1\cdot p_1$	18,0	15	12,0	$8{,}0-0{,}1\cdot p_2$	28,0
6	7,0	4,8	19,0	16	12,0	8,2	29,0
7	8,0	5,3	19,0	17	12,0	8,1	30,0
8	8,0	5,4	20,0	18	12,0	8,6	31,0
9	8,0	$5{,}6-0{,}1\cdot p_1$	20,0	19	14,0	9,6	32,0
10	10,0	6,8	21,0	20	14,0	$9{,}0+0{,}1\cdot p_2$	36,0

Требуется:

1. Построить линейную модель множественной регрессии. Записать стандартизованное уравнение множественной регрессии. На основе стандартизованных коэффициентов регрессии и средних коэффициентов эластичности ранжировать факторы по степени их влияния на результат.

2. Найти коэффициенты парной, частной и множественной корреляции. Проанализировать их.

3. Найти скорректированный коэффициент множественной детерминации. Сравнить его с нескорректированным (общим) коэффициентом детерминации.

4. С помощью F-критерия Фишера оценить статистическую надежность уравнения регрессии и коэффициента детерминации $R^2_{yx_1x_2}$.

5. С помощью t-критерия Стьюдента оценить статистическую значимость параметров чистой регрессии.

6. С помощью частных F-критериев Фишера оценить целесообразность включения в уравнение множественной регрессии фактора x_1 после x_2 и фактора x_2 после x_1.

7. Составить уравнение линейной парной регрессии, оставив лишь один значащий фактор.

8. Проверить вычисления в MS Excel.