МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Перечень примерных вопросов к коллоквиуму

Теория вероятностей

1. Размещения без повторений;
2. Перестановки без повторений;
3. Сочетания без повторений;
4. Размещения с повторениями;
5. Перестановки с повторениями;
6. Сочетания с повторениями;
7. Правила комбинаторики (суммы, произведения)
8. Элементарное событие;
9. Пространство элементарных событий;
10. Случайное событие;
11. Достоверное событие;
12. Невозможное событие;
13. Совместные и несовместные события;
14. Равновозможные события;
15. Единственно возможные события;
16. Полная группа событий;
17. Противоположные события;
18. Классическое определение вероятностей;
19. Статистическое определение вероятности;
20. Геометрические вероятности;
21. Задача о выборке. Гипергеометрическая формула;
22. Вероятность суммы несовместных событий;
23. Вероятность событий, образующих полную группу;
24. Вероятность противоположных событий;
25. Независимые и зависимые события;
26. Вероятность совмещения (произведения) независимых событий;
27. Вероятность суммы совместных событий;
28. Условная вероятность;
29. Вероятность совместного наступления зависимых событий;
30. Вероятность появления хотя бы одного события;
31. Формула полной вероятности. Формула Бейеса;
32. Повторение испытаний. Схема Бернулли;
33. Наивероятнейшее число событий;
34. Формула Бернулли;
35. Локальная и интегральная теоремы Лапласа;
36. Формула Пуассона;
37. Дискретная случайная величина (дсв);
38. Закон распределения дсв;
39. Многоугольник распределения;
40. Функция распределения, ее свойства;
41. Биномиальное распределение;
42. Геометрическое распределение;
43. Гипергеометрическое распределение;
44. Распределение Пуассона;
45. Математическое ожидание дсв, его свойства;
46. Дисперсия дсв, ее свойства;
47. Среднее квадратическое отклонение;
48. Непрерывная случайная величина (нсв);
49. Функция распределения нсв, ее свойства;
50. Плотность распределения вероятностей нсв, ее свойства;
51. Математическое ожидание нсв, его свойства;
52. Дисперсия нсв, ее свойства;
53. Среднее квадратическое отклонение;
54. Равномерное распределение;
55. Нормальное распределение;
56. Показательное распределение;

Элементы математической статистики.

57. Генеральная совокупность;
58. Выборка, объем выборки;
59. Ранжирование. Вариационный и статистический ряды;
60. Размах выборки;
61. Дискретное и интервальное распределения ряды;
62. Эмпирическая функция распределения, ее свойства;
63. Полигон. Гистограмма;
64. Точечные оценки параметров распределение;
65. Несмещенные и смещенные точечные оценки;
66. Выборочная средняя.
67. Выборочная дисперсия. Исправленная выборочная дисперсия;
68. Интервальные оценки параметров распределения;
69. Доверительный интервал;
70. Регрессионная зависимость;
71. Статистическая зависимость;
72. Корреляционная зависимость;
73. Функциональная зависимость;
74. Парная корреляция;
75. Корреляционное поле точек;
76. Уравнение линейной регрессии;
77. Коэффициент регрессии;
78. Метод наименьших квадратов;
79. Выборочный коэффициент корреляции, его свойства;
80. Прямая и обратная корреляционные связи;
81. Гипотеза о значимости выборочного коэффициента линейной корреляции.

Общие требования к выполнению расчетно-графической работы

Изучить соответствующий теоретический материал по учебнику или конспекту лекций и подробно рассмотреть приведенные там примеры; разобрать задачи, решенные на практических занятиях.

Разобраться в условии задачи и выполнить схематичный рисунок или чертеж, если это необходимо.

Чертежи, схемы следует выполнять при помощи чертежных принадлежностей, возможно на миллиметровой бумаге, формат листа А4. Все параметры, необходимые для расчета: векторы, оси координат, углы, размеры должны быть изображены на рисунке.

Решение должно сопровождаться краткими, последовательными и грамотными без сокращения слов объяснениями, без многословных пояснений и пересказа учебника. При пользовании формулами или данными, отсутствующими в учебнике, необходимо кратко и точно указывать источник (автор, название, издание, страница, номер формулы).

На каждой странице следует оставлять поля для замечаний рецензента.

Работы выполняются на писчей бумаге формата А4, чернилами (черными или синими), четким почерком, с полями. Также работа может быть набрана на компьютере в текстовом редакторе MS Word.

Номер варианта определяется условием расчетно-графической работы.

Задание, выполненное не по своему варианту, к защите не принимается.

В возвращенной на исправление расчетно-графической работе студент должен в кратчайший срок доработать все отмеченные ошибки и выполнить все данные ему указания на отдельных листах, которые должны быть вложены в соответствующие места рецензированной работы. Сдать работу на повторную проверку. Отдельно от работы исправления не рассматриваются.

Защита расчетно-графических работ производится в соответствии с графиком учебного процесса. При защите задания студент должен дать объяснение по его содержанию, уметь решать типовые задачи и давать ответы по теории соответствующего раздела курса.

Правила оформления

Пояснительная записка к расчетно-графической работе должна включать в указанной последовательности следующие разделы: титульный лист установленного образца; содержание, которое включает наименование всех разделов расчетно-графической работы; введение, которое содержит описание темы, краткий анализ возможных методов решения заданий работы; основную часть, которая содержит описание заданий и используемых методов решения, подробное решение заданий; заключение, которое содержит качественные и количественные оценки результатов расчетно-графической работы, выводы; список использованной литературы, который содержит перечень источников, использованных при выполнении расчетно-графической работы. Следует указывать только те источники, на которые имеются ссылки в тексте пояснительной записки; приложение (при необходимости), которое содержит вспомогательный материал.

Требования к оформлению. Поля при оформлении расчетно-графической работы: слева – 20 мм, справа – 10 мм, сверху – 15 мм, снизу – 20 мм. Шрифт Times New Roman, 14pt, интервал 1,2. Абзацный отступ – 10 мм. Слова разделяются одним пробелом (включить автоматическую расстановку переносов). Нумерация страниц сквозная, первая страница не нумеруется. Все рисунки выровнены по центру. Подписи к рисункам имеют формат «Рис. X. Название рисунка», где X – номер рисунка в документе, и располагаются под рисунком. Таблицы также нумеруются, подпись «Таблица X» располагается в отдельной строке, выровнена по правому краю. Следующая строка – название таблицы, выровненное по центру.

Сроки сдачи. Оформленную в соответствии с требованиями пояснительную записку к расчетно-графической работе необходимо представить на проверку в соответствии с графиком самостоятельных работ текущего семестра.

Рекомендуемая литература и методические указания

С теоретическим материалом по темам расчетно-графических работ и подробными методическими указаниями для их выполнения можно познакомиться в следующих изданиях:

1. Математический практикум. Часть 5.: Учебно-методическое пособие / А.П. Господариков и др. Национальный минерально-сырьевой университет «Горный». – СПб, 2014. – 114 с.

2. Данко П.Е. Высшая математика в упражнениях и задачах. Учебное пособие для студентов ВУЗов / Данко П.Е., Попов А.Г., КожевниковТ..Я. – М.: АСТ, 2014.

3. Гмурман, В**.** Е**.**. Теория вероятнос**тей и математическая статистика**: учебное пособие для бакалавров : учебное пособие для студентов вузов / В. Е. Гмурман. - 12-е изд. - Москва : Юрайт, 2014. - 478, [1] с. : ил., табл.; 22 см. - (Бакалавр. Базовый курс) (Министерство образования и науки РФ рекомендует).; ISBN 978-5-9916-3461-8

ПЕРВОЕ ВЫСШЕЕ ТЕХНИЧЕСКОЕ УЧЕБНОЕ ЗАВЕДЕНИЕ РОССИИ

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение высшего образования
«САНКТ-ПЕТЕРБУРГСКИЙ ГОРНЫЙ УНИВЕРСИТЕТ»

Кафедра высшей математики

РАСЧЕТНО-ГРАФИЧЕСКАЯ РАБОТА

на тему: «____»

Выполнил: студент гр. _______ _____________________ /_______________ /
(шифр) **(подпись)** **(Ф.И.О.)**

Проверил: ________________ _______________ /________________ /
(должность) (подпись) (Ф.И.О.)

Санкт-Петербург
20__

Задача 1

1. По выборкам А и В решить следующие подзадачи:

- составить вариационный ряд (по выборке А – дискретный вариационный ряд, по выборке В – интервальный вариационный ряд);
- построить графики вариационных рядов (полигон и гистограмму);
- построить эмпирическую функцию распределения;
- вычислить числовые характеристики вариационного ряда.

2. Для столбцов выборки С (не сгруппированных данных) вычислить числовые характеристики $\overline{x},\ S^2,\ \overline{S}$.

3. Для столбцов выборки С вычислить несмещенные оценки параметров генеральной совокупности: $\overline{x},\ S^2,\ S$.

4. Вычислить несмещенные оценки параметров генеральной совокупности $\overline{x},\ S^2,\ S$ по выборкам А и В, используя результаты задачи 1.

5. По выборке В при уровне значимости α проверить гипотезу о нормальном законе распределения соответствующей генеральной совокупности.

$$k = \text{остаток}\left(\frac{\text{вариант}}{4}\right), \qquad \alpha = \begin{cases} 0{,}1; & k = 0, \\ 0{,}05; & k = 1, \\ 0{,}02; & k = 2, \\ 0{,}01; & k = 3. \end{cases}$$

Таблица 15

К выборкам *A* и *B*

Вариант	К выборке *A*				К выборке *B*				
	a	***b***	***c***	***d***	***a***	***b***	***c***	***d***	***e***
1	1	2	3	4	62	63	65	67	81
2	1	0	2	1	93	63	77	91	76
3	3	2	4	5	89	101	64	67	71
4	2	3	4	5	95	64	56	78	98
5	6	7	3	4	98	76	54	54	76
6	2	3	6	6	92	65	78	90	67
7	6	7	5	6	90	67	83	54	100
8	0	8	3	0	87	63	82	90	85
9	6	7	5	4	109	65	74	73	71
10	5	6	7	3	64	75	85	95	100
11	6	4	3	3	74	85	96	90	76
12	2	2	3	3	71	71	63	91	95
13	6	4	4	4	90	63	64	75	75
14	0	0	0	5	78	67	98	98	90
15	3	4	3	3	79	97	87	67	100
16	4	6	5	5	22	22	34	45	56
17	5	5	6	6	22	40	40	26	26
18	5	6	6	6	34	41	18	49	27
19	7	7	8	8	45	27	27	23	23
20	6	4	5	6	35	34	22	20	20
21	4	5	6	9	13	13	24	35	27
22	7	7	9	4	49	38	11	11	38
22	8	5	5	5	6	12	66	26	26
23	9	7	7	7	24	34	34	37	37
24	7	6	5	4	14	19	19	31	15
25	7	7	6	6	12	14	19	19	189
26	4	4	5	5	27	53	53	27	29
27	10	10	11	8	49	41	46	12	12
28	8	5	7	5	66	23	34	29	23
29	6	6	5	5	23	10	10	10	32
30	4	5	8	11	22	31	15	32	22

Выборка *A* (*вариант 1–15*)

2 0 2 6 2 3 5 3 8 3 6 4 5 2 6 6 5 5 8 8
3 5 $\boxed{a}$ 2 4 5 2 1 $\boxed{b}$ 9 7 6 7 4 5 6 5 6 8 3
6 5 5 1 7 6 4 1 5 $\boxed{c}$ 4 7 2 8 8 2 8 2 1 6
5 2 3 6 3 3 5 3 3 7 5 $\boxed{d}$ 6 3 4 6 7 4 6 2
7 7 1 2 3 6 6 3 2 6 4 2 4 8

Выборка *A* (*вариант 16–30*)

6 $\boxed{a}$ 5 6 11 8 7 4 4 8 3 2 3 9 7
6 9 5 8 8 7 10 8 6 9 9 10 3 10 5
7 6 $\boxed{b}$ 4 3 6 12 10 2 2 3 8 6 8 2 3
7 6 8 9 9 3 8 4 11 4 $\boxed{c}$ 9 2 8 $\boxed{d}$
8 8 7 6 9 4 4 7 6 9 6

Выборка *B* (*вариант 1–15*)

57 61 60 63 66 68 64 72 69 59 71 62 69 57 61 58 60 66
$\boxed{a}$ 62 64 53 50 50 55 70 61 77 70 65 66 72 71 $\boxed{b}$ 74
62 49 62 76 66 64 62 60 53 65 49 79 58 $\boxed{c}$ 61 63 64 59
55 70 62 61 68 69 67 64 42 73 $\boxed{d}$ 69 60 64 69 62 67
67 72 57 51 77 58 63 71 $\boxed{e}$ 68 80 54 64 53 64 68 58 73
68 61 54 73 59 69 60 67 57 54 69 55 70 65 61 65 62 71
55 67 57 64 70 55 65 69 65 65 60 66 63 74 60 54 75 62
74 63 64 76 59 71 68 55 68 61 57 73 54 57 56 65 53 64
58 67 48 66 68 55 77 59 58 58 62 58 52 62 65 71 64 66
65 58 66 73 73 72 43 63 59 76 67 63 71 66 59 69 65 66
50 65 57. Длина интервала 4

Выборка *B* (*вариант 16–30*)

22 $\boxed{a}$ 18 44 52 31 18 20 27 35 41 28 29 45 36 40 41 37
18 40 25 $\boxed{b}$ 46 37 50 41 37 37 21 37 27 27 32 34 28 40
31 20 22 25 31 34 56 35 37 47 40 29 28 29 3 $\boxed{c}$ 12 41
$\boxed{d}$ 40 57 49 57 49 37 34 23 38 19 29 27 32 21 21 13 40
24 37 7 24 34 52 38 32 49 43 25 16 33 22 6 41 48 35 55
35 4 31 18 19 17 23 6 36 40 12 66 26 23 30 28 49 30 50
13 33 46 26 37 30 46 41 18 28 14 50 26 25
30 53 46 30 $\boxed{e}$ 40 40 24 16 24 28 29 25 10 19 35 27 22
38 32 41 21 46 27 49 34 53 32 31 15 24 38 25 34 22
35 42 38 33
Длина интервала 7

Выборки *C*

Вари-ант	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
	59	420	57	61	435	46	63	460	50	62	444	90	192	172	91
	71	509	67	63	457	55	65	462	53	62	437	83	171	147	79
	61	435	43	59	422	57	64	456	61	64	458	81	162	142	82
	67	469	60	64	454	49	68	485	50	64	453	83	175	149	103
	62	449	55	63	458	55	70	490	63	62	443	86	178	166	93
	62	450	59	62	449	47	57	409	41	56	410	87	185	157	89
	59	437	54	68	486	67	65	472	53	65	473	90	185	168	93
	61	422	49	65	468	57	69	502	66	66	464	85	189	167	91
	65	463	47	68	478	56	62	436	59	64	458	99	214	191	83
	63	455	62	65	463	45	65	457	64	61	431	91	192	168	104
	65	472	56	62	441	46	66	475	62	62	434	92	189	166	95
	62	448	49	68	491	50	66	474	65	69	486	98	214	193	97
	62	443	55	64	450	55	62	452	45	65	465	87	192	156	100
	65	462	50	60	432	47	60	435	44	62	439	108	189	167	92
	67	484	50	64	453	59	64	465	44	66	466	82	201	212	84
	63	442	54	67	478	58	60	431	40	63	453	79	178	148	92
	58	419	53	68	481	54	60	432	55	57	408	93	179	142	93
	64	456	45	62	438	50	63	446	43	55	396	91	225	182	98
	64	451	49	67	487	59						98	180	168	97
												90	159	168	84

Вари-ант	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
	186	164	95	195	92	14	86	13	66	32	37	151	15	30	45
	158	143	97	207	81	16	86	16	84	40	39	137	18	32	41
	176	146	83	182	92	16	80	15	71	36	30	131	16	31	30
	211	193	97	208	93	15	80	17	86	40	32	128	18	30	43
	188	181	79	161	108	15	88	15	85	36	40	148	18	30	45
	194	176	91	193	78	13	74	16	78	34	34	151	15	30	57
	188	182	85	186	98	17	86	17	75	33	33	130	18	31	35
	186	165	90	187	99	14	80	17	97	35	40	138	17	30	48
	174	150	89	185	90	13	79	16	67	36	38	140	18	29	58
	211	194	91	193	93	15	92	12	83	43	35	147	15	30	34
	199	176	80	169	92	14	77	13	74	29	37	148	17	30	39
	204	178	86	190	76	14	81	13	93	41	35	140	17	29	52
	201	199	90	199	90	16	87	14	109	36	26	145	16	29	42
	187	173	91	192	90	11	71	12	99	37	29	149	18	30	40
	184	173	90	193	87	13	81	14	76	40	37	158	16	30	49
	175	154	92	191	85	15	76	16	77	39	31	130	18	30	42
	205	180	84	174	88	13	77	13	70	28	38	137	16	29	33
	197	183	77	156	103	16	95	19	92	39	33	159	16	28	45
	199	193	88	190	91	15	79	14	79	32	36	150	16	30	52
	218	207	89	183	94	16	99	17	74	37	34	138	17	29	38

КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ

Исследуя природу, общество, экономику, необходимо считаться с взаимосвязью наблюдаемых процессов и явлений. При этом полнота описания так или иначе определяется количественными характеристиками причинно-следственных связей между ними. Оценка наиболее существенных из них, а также воздействия одних факторов на другие является одной из основных задач статистики. Изучение реальных процессов обычно предполагает наблюдение над целым рядом случайных величин. Возникает задача изучения взаимосвязи между случайными величинами. Формы проявления взаимосвязей разнообразны.

В естествознании и технике мы часто имеем дело с понятием функциональной зависимости, существо которой заключается в том, что какая–либо физическая величина определяется как однозначная функция одной или нескольких величин. Вообще, для функции одной переменной ***функциональная зависимость*** $y = f(x)$, это такая зависимость при которой для каждой независимой переменной X существует вполне определенное значение зависимой переменной Y (рис. 2.1).

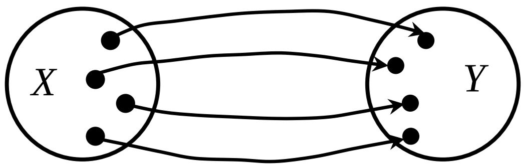

РИСУНОК 2.1

Между случайными величинами может существовать связь другого рода, проявляющаяся в том, что одна из них реагирует на изменение другой изменениями своего закона распределения. Такую связь называют стохастической. ***Стохастическая*** (статистическая) ***зависимость*** – это зависимость, при которой каждому значению одной переменной соответствует определенное (условное) распределение другой (рис. 2.2).

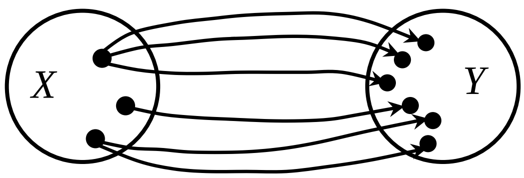

РИСУНОК 2.2

Стохастическая (статистическая) зависимость называется ***корреляционной***, если при изменении значений одной величины изменяется среднее значение другой. Если переменные не равноправны, т.е. четко ясно, какая из них причина, какая – следствие, то такая зависимость, при которой одна из переменных служит причиной изменения другой, называется ***регрессионной.***

В наиболее общем виде задача статистики в области изучения взаимосвязей состоит в количественной оценке их наличия и направления, а также характеристике силы и формы влияния одних факторов на другие. Для ее решения применяют две группы методов, одна из которых включает в себя методы корреляционного анализа, а другая – регрессионный анализ. Ряд исследователей объединяют эти методы в корреляционно–регрессионный анализ, что имеет под собой некоторые основания: наличие целого ряда общих вычислительных процедур, взаимодополнения при интерпретации результатов и др.

Таблица 2.1.

КОРРЕЛЯЦИОННО–РЕГРЕССИОННЫЙ АНАЛИЗ	
Корреляционный анализ	**Регрессионный анализ**
Определение	
Статистический метод, изучающий методы оценки выборочного	Статистический метод, изучающий зависимость между результативным

коэффициента корреляции, проверку значимости коэффициента корреляции, построение для него (в случае значимости) доверительных интервалов.	признаком Y и входной переменной X (оценивание степени и формы связи между величинами).
Основные задачи	
Выявление связи между случайными переменными и оценка ее тесноты.	Установление формы и изучение зависимости между переменными.

1. Корреляционный анализ

Корреляционный анализ можно применять только в том случае, когда данные наблюдений или эксперимента можно считать случайными и выбранными из нормальной совокупности. Типичная корреляционная задача возникает в исследованиях, где каждая испытуемая величина характеризуется двумя или большим числом показателей.

Простейший случай задания экспериментальных данных – несгруппированные данные, т.е. набор пар чисел (x_i, y_i), где $x_1, \ldots, x_n$ – выборка значений переменной X; $y_1, \ldots, y_n$ – выборка значений переменной Y (табл. 2.1.1).

Таблица 2.1.1

X	x_1	…	x_i	…	x_n
Y	y_1	…	y_i	…	y_n

В том случае, когда варианты парной выборки встречаются по несколько раз, причем с одним значением варианты x_i может встретиться несколько вариант y_i, их обычно задают в виде корреляционной таблицы. На пересечении строк и столбцов этой таблицы отмечается частота n_{ij} выбора соответствующей пары (x_i, y_i), а частоты вариант x_i, y_i находятся как суммы значений n_{ij} по соответствующей строке (табл. 2.1.2).

Для наглядности полученного материала каждую пару можно представить в виде точки на координатной плоскости. По оси абсцисс откладывают значения одного вариационного ряда – X, по оси ординат – другого – Y. Такое изображение статистической зависимости называют ***полем корреляции*** или корреляционным полем точек.

Виды корреляционной зависимости между измеренными признаками могут быть различны: так, корреляция бывает линейной и нелинейной, отрицательной и положительной. Она ***линейна*** – если с увеличением или уменьшением одной переменной X, вторая переменная Y в среднем либо также растет, либо

Таблица 2.1.2

	i	1	2	3	…	v	
j	Y \ X	x_1	x_2	x_3	…	x_v	n_j
1	y_1	n_{11}	n_{12}	n_{13}	…	n_{1v}	n_1
2	y_2	n_{21}	n_{22}	n_{23}	…	n_{2v}	n_2
3	y_3	n_{31}	n_{32}	n_{33}	…	n_{3v}	n_3
…	…	…	…	…	…	…	…
q	y_q	n_{q1}	n_{q2}	n_{q3}	…	n_{qv}	n_q
	n_i	n_1	n_2	n_3	…	n_v	$n=\sum_{i=1}^{v} n_i = \sum_{j=1}^{q} n_j$

убывает. Она ***нелинейна***, если при увеличении одной величины характер изменения другой не линеен, а описывается другими законами. Корреляция будет ***положительной***, если с увеличением переменной X переменная Y в среднем также увеличивается, а если с увеличением X переменная Y имеет в среднем тенденцию к уменьшению, то говорят о наличии ***отрицательной*** корреляции. Возможна ситуация, когда между переменными невозможно установить какую-либо зависимость. В этом случае говорят об отсутствии корреляционной зависимости.

Самой простой оценкой корреляционной зависимости является оценка по расположению точек на корреляционном поле (рис. 2.1.1). Такую же приближенную картину дает и корреляционная таблица. Если заполнены клетки вблизи той или другой диагонали, то речь идет о линейной корреляции. Если заполнены большинство клеток или заполненные клетки образуют какую-то кривую, то можно говорить о нелинейной корреляции.

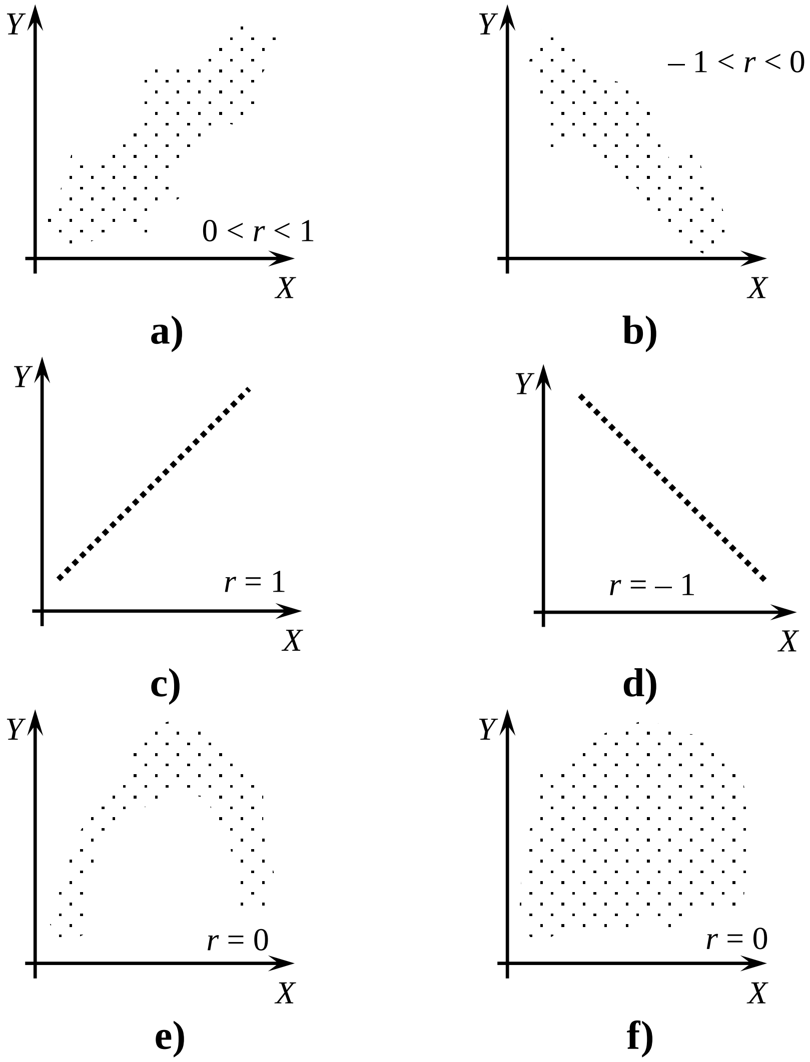

Рисунок 2.1.1. a) – положительная линейная корреляция;
b) – отрицательная линейная корреляция;
c), d) – линейная функциональная зависимость;
e), f) – отсутствие линейной корреляционной зависимости.

На практике часто интересует не сама зависимость одной величины от другой, а именно характеристика тесноты связи между ними, которую можно было бы выразить одним числом. Эта характеристика называется ***выборочным коэффициентом линейной корреляции*** *r*. Рассмотрим некоторые его свойства.

1. $|r| \leq 1$.

2. $|r| = 1$ тогда и только тогда, когда точки $(x; y)$ лежат на одной прямой.

3. Обычно степень тесноты связи определяют по шкале Шеддока: если $r < 0{,}2$ – связи нет; если $0{,}2 \leq r < 0{,}5$ – связь слабая; если $0{,}5 \leq r < 0{,}75$ – связь средняя; если $0{,}75 \leq r < 0{,}95$ – связь тесная; если $0{,}95 \leq r < 1$ – связь очень тесная.

4. Коэффициенты корреляции X на Y и Y на X совпадают.

5. Если $r = 0$, то случайные величины X и Y некоррелированы, что не означает независимости вообще, зависимость между ними можно описать другими законами, например, параболой.

6. Если $r > 0$, то корреляционная связь между переменными прямая, при $r < 0$ – связь обратная.

Аналитически выборочный коэффициент линейной корреляции находится следующим образом:

Таблица 2.1.3

Для несгруппированных данных	$$r=\frac{\sum\limits_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sqrt{\sum\limits_{i=1}^{n}\left(x_i-\bar{x}\right)^2\sum\limits_{i=1}^{n}\left(y_i-\bar{y}\right)^2}}$$ (2.1.1)
Для сгруппированных данных	$$r=\frac{n\cdot\sum\limits_{i=1}^{v}\sum\limits_{j=1}^{q}x_i y_j n_{ij}-\sum\limits_{i=1}^{v}x_i n_i\sum\limits_{j=1}^{q}y_j n_j}{\sqrt{n\sum\limits_{i=1}^{v}x_i^2 n_i-\left(\sum\limits_{i=1}^{q}x_i n_i\right)^2}\cdot\sqrt{n\sum\limits_{j=1}^{v}y_j^2 n_j-\left(\sum\limits_{j=1}^{q}y_j n_j\right)^2}}$$ (2.1.2)

2. Регрессионный анализ

Регрессионный анализ – один из основных методов современной математической статистики, позволяющий аналитически представить связь между переменными объекта. Чаще всего регрессионный анализ используется для прогноза, т.е. предсказания значений ряда зависимых переменных по известным значениям других. Регрессионная модель представляет собой математическое выражение, связывающее входные переменные X с одним выходом Y. В этом случае задача будет состоять в нахождении зависимости вида $Y = F(X)$ или, напротив, в нахождении зависимости вида $X = F(Y)$. При этом изменение функции в зависимости от изменений одного или нескольких аргументов называется регрессией. ***Регрессия*** – это функция, позволяющая по величине одного признака X, находить среднее ожидаемое значение другого признака Y, корреляционно связанного с X. В регрессионном анализе рассматривают:

1. Линейную относительно параметров регрессию: парную линейную, парную криволинейную, множественную линейную, множественную нелинейную.

2. Нелинейную относительно параметров.

Графическое выражение регрессионного уравнения называется ***линией регрессии***.

Рассмотрим ***линейную парную регрессию***. В линейной математической модели для несгруппированных данных уравнение линейной регрессии имеет вид: $y = a + bx$, где a, b – параметры уравнения линейной регрессии; b – это коэффициент регрессии, показывающий насколько в среднем величина одного признака Y изменяется при изменении на единицу меры другого признака X, корреляционно связанного с Y. Чем больше b, тем круче прямая; a – свободный член в уравнении, определяет y при $x = 0$; y – это предсказанное значение для данного x при определенных значениях регрессионных параметров. Линию регрессии можно задать также при помощи линейного уравнения $x = c + dy$.

Для определения неизвестных параметров регрессии используется ***метод наименьших квадратов***, рассматриваемый ниже, который для линейной регрессии в результате преобразований сводится к решению систем линейных уравнений:

Таблица 2.2.1

Y на X	X на Y
$$\begin{cases} a\cdot n+b\cdot\sum x_i=\sum y_i, \\ a\cdot\sum x_i+b\cdot\sum x_i^2=\sum x_i y_i \end{cases}$$ (2.2.1)	$$\begin{cases} c\cdot n+d\cdot\sum y_i=\sum x_i, \\ c\cdot\sum y_i+d\cdot\sum y_i^2=\sum x_i y_i \end{cases}$$ (2.2.2)

Для сгруппированных данных уравнение линейной регрессии удобнее записать в виде:

Y на X: $y-\bar{y}=r\frac{\bar{S}_y}{\bar{S}_x}(x-\bar{x})$, X на Y: $x-\bar{x}=r\frac{\bar{S}_x}{\bar{S}_y}(y-\bar{y})$,

где r – коэффициент корреляции;

$\bar{x}$, $\bar{y}$ – средние значения признаков x и y;

$\bar{S}_x=\sqrt{n\sum x_i^2\cdot n_i-\sum(x_i n_i)^2}$, $\bar{S}_y=\sqrt{n\sum y_j^2\cdot n_j-\sum(y_j n_j)^2}$ – средние значения квадратичных отклонений.

Линии регрессии, заданные этими уравнениями пересекаются в точке $M(\bar{x},\bar{y})$, с координатами, соответствующими средним арифметическим значениям корреляционно связанных между собой переменных X и Y.

3. Оценка значимости параметров взаимосвязи

Получив оценки корреляции и регрессии, необходимо проверить их на соответствие истинным параметрам взаимосвязи.

Проверим гипотезу о значимости выборочного коэффициента линейной корреляции. Это ответ на вопрос, существует ли вообще эта связь. Эмпирический коэффициент корреляции, как и любой другой выборочный показатель, служит оценкой своего генерального параметра. Он является величиной случайной, так как определяется по значениям переменных, случайно попавших в выборку из генеральной совокупности, а значит, как и любая случайная величина имеет ошибку m_r. Для оценки значимости коэффициента парной корреляции эту ошибку рассчитывают следующим образом:

Таблица 3.1

Для выборки $n > 100$	Для выборки $n \le 100$
$m_r=\frac{1-r^2}{\sqrt{n}}$	$m_r=\sqrt{\frac{1-r^2}{n-2}}$

Значимость коэффициента корреляции проверяется его сопоставлением с m_r, при этом получают $t_{набл}=\frac{r}{m_r}$, где $t_{набл}$ – расчетное значение t-критерия с числом степеней свободы $n-2$. Гипотезу проверяют по таблицам распределения Стьюдента в соответствии с выбранным уровнем значимости. Если $t_{набл}$ больше теоретического (табличного) значения критерия Стьюдента ($t_{табл}$) для заданного уровня значимости, то можно утверждать, что r значимо. Если же $t_{набл} < t_{табл}$, то r статистически незначим, эта связь случайна.

Таблица 3.2

Алгоритм корреляционно–регрессионного анализа
1. Исходя из целей и задач исследования зависимости, устанавливается результативный (y) и факторные (x) признаки.
2. По совокупности объектов определяется значение результативного и факторных признаков.
3. Обосновывается, для случая парной зависимости – обычно графическим методом, модель уравнения регрессии.
4. Методом наименьших квадратов определяются параметры уравнения регрессии.
5. Определяется теснота связи между изучаемыми признаками.
6. Оценивается значимость уравнения связи, его параметров и показателей тесноты связи.

Пример 3.1. По заданной выборке *X*, *Y* несгруппированных данных построить корреляционное поле для двумерной выборки *XY*. Найти соответствующие уравнения регрессии X_y и Y_x, построить их графики. Найти выборочный коэффициент регрессии, проверить его значимость.

x	– 0,1	0,2	0,5	0,9	1,2
y	– 7,1	– 6,2	– 4,3	– 2,7	– 0,9

Р е ш е н и е. Построим корреляционное поле точек (рис. 3.1).

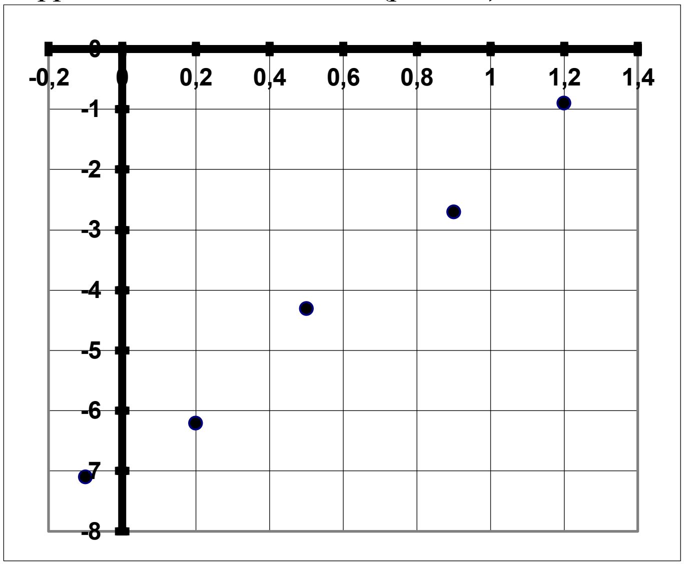

Рисунок 3.1

Для несгруппированных данных уравнения регрессии будут иметь вид $y = a + bx$ и $x = c + dy$. Для определения параметров *a*, *b*, *c*, *d* составим системы уравнений по таблице 2.2.1. Для уравнения регрессии Y_x составим расчетную таблицу.

Таблица 3.3

i	x_i	y_i	x_i^2	y_i^2	x_iy_i
1	– 0,1	– 7,1	0,01	50,41	0,71
2	0,2	– 6,2	0,04	38,44	– 1,24
3	0,5	– 4,3	0,25	18,49	– 2,15
4	0,9	– 2,7	0,81	7,29	– 2,43
5	1,2	– 0,9	1,44	0,81	– 1,08
Σ	2,7	– 21,2	2,55	115,44	– 6,19

По таблице 3.3 составим систему уравнений (2.2.1):

$$\begin{cases} a \cdot 5 + b \cdot 2{,}7 = -21{,}2, \\ a \cdot 2{,}7 + b \cdot 2{,}55 = -6{,}19. \end{cases}$$

Решая эту систему, получаем коэффициенты *a* и *b*: $a = -6{,}84$; $b = 4{,}815$. Таким образом, получили уравнение линейной регрессии *Y* на *X*: $y = -6{,}84 + 4{,}815x$.

Аналогично, используя ту же таблицу 3.3, а также формулу (2.2.2), построим уравнение линейной регрессии *X* на *Y*:

$$\begin{cases} c \cdot 5 + d \cdot (-21{,}2) = 2{,}7, \\ c \cdot (-21{,}2) + d \cdot 115{,}44 = -6{,}19. \end{cases}$$

Из системы находим коэффициенты *c* и *d*: $c = 1{,}412$; $d = 0{,}206$. Тогда уравнение линейной регрессии *X* на *Y* будет иметь вид: $x = 1{,}412 + 0{,}206y$.

Построим графики реальных данных и полученных зависимостей (рис. 3.2).

x	– 0,1	1,3
y_x	– 7,32	– 0,58

y	– 7,3	– 0,5
x_y	– 0,09	1,31

Определим коэффициент r корреляции по таблице 2.1.3 по формуле (2.1.1).

Определим средние значения:

$$\bar{x} = \frac{-0{,}1 + 0{,}2 + 0{,}5 + 0{,}9 + 1{,}2}{5} = 0{,}54,$$

$$\bar{y} = \frac{-7{,}1 - 6{,}2 - 4{,}3 - 2{,}7 - 0{,}9}{5} = -4{,}24$$

Составим расчетную таблицу 3.4.

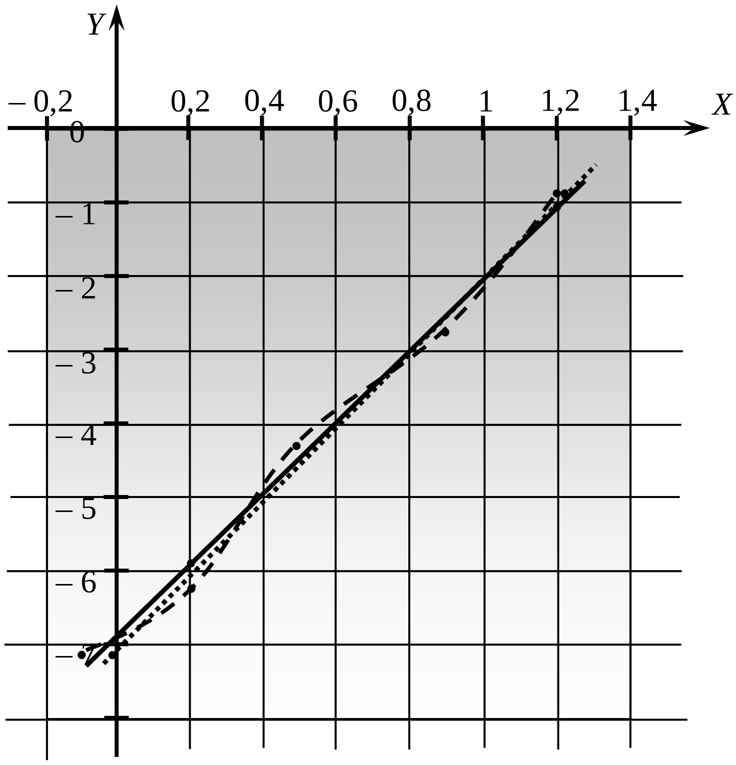

Рисунок 3.2

Таблица 3.4

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
– 0,1	– 7,1	– 0,64	– 2,86	0,41	8,18	1,83
0,2	– 6,2	– 0,34	– 1,96	0,12	3,84	0,67
0,5	– 4,3	– 0,04	– 0,06	0,002	0,004	0,002
0,9	– 2,7	0,36	1,54	0,13	2,37	0,55
1,2	– 0,9	0,66	3,34	0,44	11,16	2,20
Σ				1,102	25,554	5,252

Тогда коэффициент корреляции:

$$r = \frac{5{,}252}{\sqrt{1{,}102 \cdot 25{,}554}} = \frac{5{,}252}{5{,}307} = 0{,}99.$$

Так как $r \in [0{,}95;\ 1)$, то связь между величинами X и Y очень тесная. Это видно и по графикам, линии X_y и Y_x практически совпадают.

Проверим значимость коэффициента корреляции $r = 0{,}99$. Ошибка m_r для выборки $n = 5$, $n < 100$ имеет вид:

$$m_r = \sqrt{\frac{1 - r^2}{n - 2}} = \sqrt{\frac{1 - 0{,}99^2}{5 - 2}} = \frac{0{,}141}{1{,}732} = 0{,}081;$$

$$t_{\text{наб}} = \frac{0{,}99}{0{,}081} = 12{,}222.$$

По таблицам распределения Стьюдента $t_{табл}$ при $\alpha = 0{,}05$ и числе степеней свободы $n - 2 = 5 - 2 = 3$ имеет вид $t_{табл}$ (0,05; 3) = 3,18. Так как $t_{наб} > t_{табл}$, коэффициент корреляции значим, связь между признаками неслучайна.

Задачи для самостоятельного решения

Задача 3.1. По заданной выборке *X*, *Y* несгруппированных данных построить корреляционное поле для двумерной выборки *XY*. Найти соответствующие уравнения регрессии X_y и Y_x, построить их графики. Найти выборочный коэффициент регрессии, проверить его значимость.

№	Распределение						№	Распределение					
1	*x*	– 1	– 0,8	– 0,5	– 0,1	0,2	**16**	*x*	– 3,8	5,5	7,0	8,4	14,2
	y	– 7	– 6,3	– 5,4	– 4,7	– 4,0		*y*	–16,0	–12,8	– 2,6	3,4	4,1
2	*x*	0,6	1,4	2,5	3,4	4,0	**17**	*x*	13,1	14,2	15,4	18,3	20,1
	y	– 1,0	– 0,1	1,0	1,2	3,4		*y*	20,8	13,5	12,6	5,0	– 1,5
3	*x*	0,5	1,3	2,3	3,3	3,9	**18**	*x*	5,1	15,8	22,4	28,3	30,5
	y	5,5	4,7	3,7	3,0	2,2		*y*	0,5	0,4	0,1	– 0,2	– 3,1
4	*x*	10,0	12,2	14,2	16,7	19,5	**19**	*x*	– 4,3	– 1,8	2,5	8,4	13,1
	y	6,0	4,0	3,6	2,5	2,4		*y*	6,3	12,8	16,4	23,6	28,8
5	*x*	–11,0	– 9,1	– 7,4	– 5,0	– 3,3	**20**	*x*	– 0,3	0,4	0,4	0,8	1,9
	y	– 0,5	– 1,1	– 1,5	– 2,2	– 1,7		*y*	0,5	0,17	0,14	0,13	–0,19
6	*x*	–10,8	– 8,9	– 7,2	– 4,8	– 3,1	**21**	*x*	– 1,5	5,0	12,6	13,5	20,8
	y	0,5	1,1	1,5	2,2	1,7		*y*	– 1,9	1,3	1,4	1,7	5,0
7	*x*	0,3	1,0	1,8	2,4	3,1	**22**	*x*	0,2	0,8	1,2	2,1	2,4
	y	– 2,1	– 2,9	– 3,5	– 4,4	– 5,2		*y*	–12,4	–10,5	– 3,9	– 4,2	– 1,5
8	*x*	– 3,0	– 1,1	0,9	3,0	4,8	**23**	*x*	–25,4	–21,6	–18,7	–16,4	–11,1
	y	6,0	8,5	10,0	11,8	15,5		*y*	– 3,1	–6,5	–10,0	–11,2	–15,4
9	*x*	– 2,5	– 1,9	– 1,1	– 0,6	0,2	**24**	*x*	– 9,3	– 7,4	– 3,1	0,5	2,9
	y	2,7	2,6	0,3	0,1	– 0,3		*y*	2,6	3,7	5,1	8,3	7,1
10	*x*	– 0,1	0,2	0,6	0,8	1,5	**25**	*x*	– 7,2	– 8,4	– 5,3	– 3,8	– 2,7
	y	–13,3	–10,2	– 7,3	– 5,3	– 4,0		*y*	–12,5	–10,6	– 3,8	– 4,3	– 1,4
11	*x*	–13,4	–10,3	– 7,4	– 5,4	– 4,1	**26**	*x*	–16,2	–13,0	– 2,8	3,6	4,3
	y	2,2	1,1	0,2	– 1,3	0,1		*y*	– 3,9	5,6	7,1	8,5	14,3
12	*x*	16,2	20,4	22,3	30,5	36,1	**27**	*x*	– 4,0	– 1,0	2,0	8,0	13,0
	y	– 8,0	– 2,4	5,0	6,4	10,8		*y*	5,2	15,8	22,5	28,4	30,6
13	*x*	0,01	0,04	0,08	0,14	0,22	**28**	*x*	– 3,2	– 0,3	0,2	0,5	0,6
	y	– 5,3	1,2	4,8	8,1	11,3		*y*	4,3	3,6	– 2,8	–12,6	–15,3
14	*x*	– 4,1	– 1,3	2,2,	8,4	13,8	**29**	*x*	1,8	2,1	2,2	3,0	4,1
	y	12,4	8,8	3,1	– 3,5	– 7,2		*y*	0,3	1,4	1,7	2,6	5,3
15	*x*	2,1	6,6	8,5	9,4	11,7	**30**	*x*	8,0	11,0	14,0	15,0	20,0
	y	20,4	17,1	11,2	6,3	1,8		*y*	6,0	4,3	4,0	2,0	– 3,0