КЫРГЫЗСКО-РОССИЙСКИЙ
СЛАВЯНСКИЙ УНИВЕРСИТЕТ

КАФЕДРА МАТЕМАТИКИ

Т.А. Давидюк,
Ш.А. Ильясов,
Т.К. Кадыров

РУКОВОДСТВО К РЕШЕНИЮ ЗАДАНИЙ ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Методическое пособие
для студентов дневного и заочного
обучения

Издательство Кыргызско-Российского
Славянского университета

Бишкек 2001

Д 13

Давидюк Т.А., Ильясов Ш.А., Кадыров Т.К. РУКОВОДСТВО К РЕШЕНИЮ ЗАДАНИЙ ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ: Методическое пособие для студентов дневного и заочного обучения /Кыргызско-Российский Славянский университет. – Бишкек, 2001. – 56 с.

Рекомендовано к изданию
кафедрой математики КРСУ
и РИСО КРСУ

Рецензент Е.С. Федорова

Часть I. АНАЛИЗ РАСПРЕДЕЛЕНИЙ

Изучаемый объект или процесс весьма часто находится под действием множества факторов, форма влияния которых или неизвестна, или слишком сложна для решения в рамках конкретной задачи. Кроме того, не любой из этих влияющих факторов доступен для контроля. В этом случае для изучения применяют статистические методы. Суть этих методов заключается в том, что изучается не единичное явление, а массовая совокупность однородных явлений. В такой массовой совокупности влияние каждого второстепенного фактора носит случайный характер и в общей массе взаимно погашается. В результате проявляются общие для всей совокупности статистические закономерности. Если нас интересует некоторый количественный или качественный признак, то подвергают наблюдению множество объектов, характеризующихся этим признаком. Если наблюдению подвергаются абсолютно все объекты – носители этого признака, то наблюдение называют сплошным (перепись населения). В силу множества причин (временные, материальные и т. д.) наблюдению чаше всего подвергают определенную часть объектов, отобранных по специальным правилам. В этом случае наблюдение называют выборочным. Целью статистического наблюдения является изучение изменения вариации признака в данной совокупности.

Занятие 1. ПОСТРОЕНИЕ СТАТИСТИЧЕСКОГО РАСПРЕДЕЛЕНИЯ ВЫБОРКИ. ГРАФИК СТАТИСТИЧЕСКОГО РАСПРЕДЕЛЕНИЯ

Произвольное множество однородных объектов называют генеральной совокупностью. Множество случайно отобранных из генеральной совокупности объектов называют выборочной совокупностью или выборкой.

Число объектов совокупности называют объемом совокупности. Объем генеральной совокупности обозначают N, а выборочной – n.

Пусть для изучения количественного признака X из генеральной совокупности извлечена выборка объема n. Первичным результатом статистического наблюдения является перечень членов выборочной совокупности и соответствующих им значений признака. Значения признака иначе называют вариантами. Такая сводка называется рядом вариант или простым статистическим рядом. Простой статистический ряд можно подвергнуть первичной обработке, заключающейся в группировке данных членов совокупности. Если количественный признак дискретный (принимает множество определенных значений), то при группировке выбирают члены совокупности, принимающие одни и те же значения. В результате группировки приходят к таблице.

Таблица 1

Значения признака x_i	x_1	x_2	x_3	…	x_k	Итого
Число объектов	n_1	n_2	n_3	…	n_k	n

Число n_i значений признака x_i называют частотой соответствующего значения x_i. Очевидно, что $\sum_{i=1}^{k} n_i = n$, где n – объем выборки.

Перечень вариант x_i и соответствующих им частот n_i носит название статистического распределения выборки, заданного в табличном виде.

Если количественный признак принимает любые значения из некоторого интервала, то его считают непрерывным. При группировке данных в этом случае все значения разбивают на интервалы и подсчитывают, сколько значений признака попадает в каждый интервал. Статистическое распределение выборки принимает вид:

Таблица 2

Интервалы	(x_0, x_1)	(x_1, x_2)	…	(x_{k-1}, x_k)	Итого
Число значений признака	n_1	n_2	…	N_k	n

Отношение n_i/n называют относительной частотой и обозначают w_i. Очевидно, $\sum_{i=1}^{k} w_i = 1$.

Замечание 1. Кроме статистических распределений, заданных в виде таблиц 1 и 2, можно получить статистические распределения относительных частот в виде следующих таблиц:

Таблица 3

Значение признака	x_1	x_2	…	x_k	Итого
Относительная частота	w_1	w_2	…	w_k	n

Таблица 4

Интервалы	(x_0, x_1)	(x_1, x_2)	…	(x_{k-1}, x_k)	Итого
Относительная частота	w_1	w_2	…	w_k	1

Замечание 2. При составлении статистического распределения в виде таблиц 2 и 4 частоту вариант, точно попавших на границы интервалов,

делят поровну между ними или относят к левому (правому) интервалу. Выбор типа отнесения произволен, но он должен быть единообразным для всей выборки.

Для наглядного представления статистических распределений признака строят полигон для дискретного признака или гистограмму для непрерывного распределения признака.

Полигоном частот (относительных частот) называют ломаную с вершинами в точках (x_i, n_i) или (x_i, w_i).

Гистограммой частот (относительных частот) называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиной (x_{i+1} - x_i), а высотами величины $n_i/(x_{i+1} - x_i)$ или $w_i/(x_{i+1} - x_i)$.

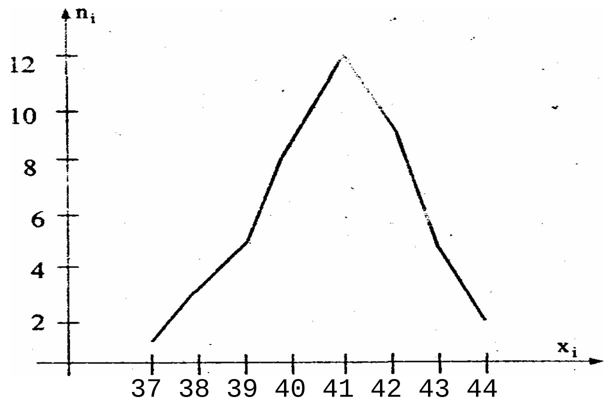
Пример 1.1. Построить статистическое распределение выборки и изобразить его графически для следующего распределения размеров 45 пар мужской обуви, проданных магазином за день.
39 41 40 42 41 40 42 44 40 43 42 41 43 39 42 41 42 39 41 37 43 41 38 43 42 41 40 41 38 44 40 39 41 40 42 40 41 42 40 43 38 39 41 41 42

Решение

Сгруппируем данные, расположив их в порядке возрастания, и подсчитав, сколько раз встречалось каждое значение признака, получим следующее статистическое распределение в табличном виде.

Размер обуви	37	38	39	40	41	42	43	44	Итого
Число пар	1	3	5	8	12	9	5	2	45

Изобразим полученное распределение графически. Графиком распределения будет полигон.



Пример 1.2. Выработка валовой продукции на одного работающего за год по деревообрабатывающим предприятиям составила (тыс. сом).
5.0 4.0 3.9 7.4 4.5 5.3 5.6 7.3 3.8 5.2 5.1 6.3 5.5 4.8 5.1 7.2 4.1 6.5 5.4 5.8 6.0 6.4 5.8 5.3 4.9 6.2 7.2 5.9 5.7 5.5

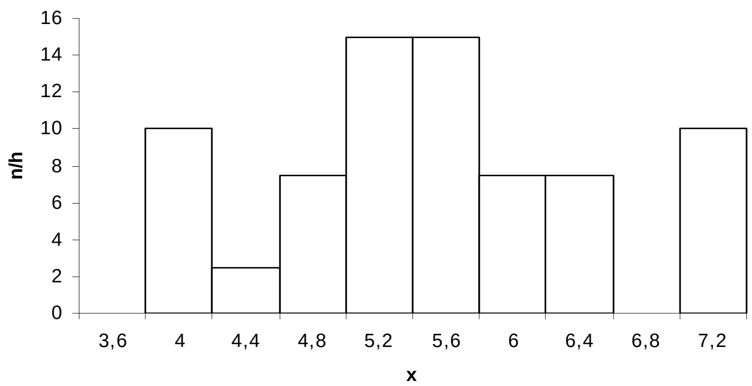
Составить статистическое распределение выборки и изобразить его графически.

Решение

Признак принимает любые значения от 3.8 до 7.4. Разобьем весь интервал на 9 частичных интервалов с одинаковой длиной h = (7.4-3.8)/9 = = 0.4 и подсчитаем, сколько значений признака попадает в каждый частичный интервал (значения, совпадающие с граничными, будем относить к левому интервалу). Статистическое распределение частот оказывается следующим в табличном виде.

Интервалы	3.8-4.2	4.2-4.6	4.6-5.0	5.0-5.4	5.4-5.8	5.8-6.2	6.2-6.6	6.6-7.0	7.0-7.4	Итого
Количество значений	4	1	3	6	6	3	3	0	4	30

Графиком полученного распределения будет гистограмма.



Задания для работы в аудитории

1.1. Число рабочих мест в магазинах составляло:
2 5 4 3 2 1 2 3 4 2 3 2 2
2 2 1 3 4 3 2 3 1 2 3 3 3
2 2 4 1
Составить статистическое распределение частот и изобразить его графически.

1.2. Данные о возрасте 56 студентов первого курса оказались следующими (приводится число полных лет):

17	17	18	17	18	17	19	17	20	19	17	18	20
18	17	20	21	20	17	18	22	18	17	18	17	18
19	17	22	19	21	19	21	18	17	21	21	17	19
18	20	17	18	19	19	17	18	19	17	20	17	18
18	18	20	17									

Построить статистическое распределение и изобразить его графически.

1.3. Результаты хронометража при выполнении одной и той же операции 20 рабочими оказались следующими (длительность операции задается в минутах):

42	56	45	51,5	43	47	49,5	47,5	51	49	45,5	52,3	53
48	46,5	44	50	47,5	55,5	45						

Построить интервальный ряд с пятью равными интервалами и изобразить его графически.

Задания для самостоятельной работы

1.4. По имеющимся данным о возрасте рабочих одной строительной организации (приводится число полных лет):

25	17	53	18	19	46	18	35	25	23	22	25	17
23	35	21	26	19	17	22	18	25	27	44	35	25
24	25	17	24	18	23	32	23	33	26	18	27	38
28	18	18	17	21	18	19	22	42	22			

Составить статистическое распределение частот и изобразить его графически.

1.5. По данным измерения диаметра валиков (в см) построить статистическое распределение относительных частот и изобразить его графически:

4,8	4,7	5.3	5.2	5.3	4.7	5.0	5.1	4.7	5.0	5.0	4.8	5,1
5,0	4.8	5.2	5.2	5.3	5.0	4.9	5.1	4.9	4.9	5.1	4,8	5,0
4.9	4.9	5.1	4.8	5.2	4.7	5.0	4.8	5.0	4.8	5,0	5,0	5.3
5.0	4.9	5.1	5.1	5.0	5.0	5.1	5.1	5.2	4,9	5,1		

З а н я т и е 2. СТАТИСТИЧЕСКИЕ ОЦЕНКИ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

Пусть для изучения генеральной совокупности относительно количественного признака X извлечена выборка объема n. Функцию от наблюдаемых случайных значений называют статистической оценкой неизвестного параметра теоретического распределения генеральной совокупности.

Оценкой математического ожидания (выборочной средней) $\overline{X}$ называют среднее арифметическое значение признака выборочной совокупности и вычисляют по формулам:

$$\bar{x} = \sum_{i=1}^{n} x_i/n,$$

$$\bar{x} = \sum_{i=1}^{k} x_i n_i/n.$$

Первую формулу применяют, когда все значения признака X различны или не сгруппированы, а вторую – в случае предварительно сгруппированной выборки, т.е. когда значения X имеют частоты n_1, n_2, n_3, …, n_k.

Несмещенной оценкой дисперсии называют среднее арифметическое квадратов отклонения наблюдаемых значений признака от их среднего

$$S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/(n - 1) \text{ или } S^2 = \sum_{i=1}^{k} (x_i - \bar{x})^2 n_i/(n - 1).$$

Иногда оценку дисперсии удобнее вычислять по формуле

$$S^2 = (\sum_{i=1}^{n} x_i^{\ 2} - n(\sum_{i=1}^{n} x_i)^2)/(n - 1).$$

Оценкой среднего квадратического отклонения называют квадратный корень из оценки дисперсии: $s = \sqrt{S^2}$.

Модой Mo называют варианту, имеющую наибольшую частоту. Медианой Me называют варианту, которая делит вариационный ряд на две части, равные по числу вариант. Если число вариант нечетно n = 2k + 1, то $Me = x_{k+1}$; если n = 2k – четное, то $Me = (x_k + x_{k+1})/2$.

Размахом вариации R называют разность между наибольшей и наименьшей вариантами $R = x_{max} - x_{min}$.

Коэффициентом вариации V называют отношение выборочного среднего квадратического отклонения к выборочной средней, взятое в процентах:

$V = (s / \overline{X}_в) \times 100\%$.

Эмпирической функцией распределения называют функцию F*(x), которая определяет для каждого значения x относительную частоту события X < x.

$F^*(x) = n_x/n$, где n_x - число вариант, меньших x.

Пример 2.1. Для статистического распределения частот, полученного в примере 1.1, вычислить $\overline{X}$, S^2, s, Mo, Me, R, V. Найти функцию F*(x) и изобразить ее графически.

Решение

Распределение было получено в виде:

x_i	37	38	39	40	41	42	43	44	Итого
n_i	1	3	5	8	12	9	5	2	45

Значения признака имеют частоты, т.е. сгруппированы, поэтому вычисления будем производить по соответствующим формулам.

$\bar{x} = (37\times1 + 38\times3 + 39\times5 + 40\times8 + 41\times12 + 42\times9 + 43\times5 + 44\times2)/45 = 40.87.$

Для вычисления несмещенной оценки дисперсии используем формулу

$$S^2 = \sum_{i=1}^{k}(x_i - \overline{X})^2 n_i/(n - 1) \approx 2.66,\ s \approx 1.63,$$

$Mo = 41$, $Me = (40 + 41)/2 = 40.5$,

$R = x_{max} - x_{min} = 44 - 37 = 7$, $V = \sigma_в / \overline{X}_в \times 100\% = 1.63/40.87*100\% = 3.99\%$.

Построим эмпирическую функцию F*(x).

Для X < 37, F*(37) = 0, так как значений, меньших 37, не наблюдалось.

Значения X < 38, а именно равное 37, наблюдалось один раз, следовательно, F* (38) = 1/45 = 0.02.

Значения X < 39, а именно равное 37 и 38, наблюдались 1+3 = 4 раза, следовательно, F* (39) = 3/45 = 0.09.

Значения X < 40, а именно равное 37, 38 и 39, наблюдались 1+3+5 = 9 раз, следовательно, F*(40) = 9/45 = 0.2.

Значения X < 41, а именно равное 37, 38, 39 и 40, наблюдались 1+3+5+8 = 17 раз, следовательно, F*(41) = 17/45 = 0.4.

Аналогично F*(42) = 29/45 = 0.6; F*(43) = 38/45 = 0.8; F*(44) = 43/45 = 0.96. Для всех значений X > 44 очевидно, что F*(x) = 1.

Искомая эмпирическая функция:

F*(x) = 0, для $x \le 37$;

F*(x) = 0.02, для $37 < x \le 38$;

F*(x) = 0.09, для $38 < x \le 39$;

F*(x) = 0.2, для $39 < x \le 40$;

F*(x) = 0.4, для $40 < x \le 41$;

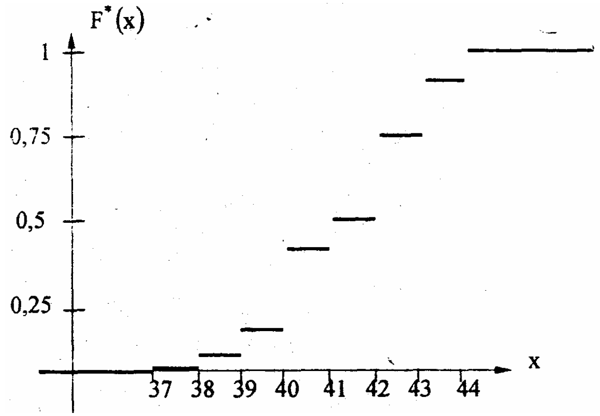
F*(x) = 0.6, для $41 < x \le 42$;

F*(x) = 0.8, для $42 < x \le 43$;

F*(x) = 0.96, для $43 < x \le 44$;

F*(x) = 1.0, для $44 < x$.

Построим график функции F*(x)



Задания для работы в аудитории

2.1. Данные о распределении шести рабочих столярного цеха по уровню квалификации (тарифным разрядам) оказались:

Порядковые номера рабочих	1	2	3	4	5	6
Тарифные разряды	4	6	2	2	3	5

Вычислить средний тарифный разряд рабочих.

2.2. Для статистического распределения, полученного в задаче 1.2, вычислить $\overline{X}$, S^2, s, Mo, Me, R, V. Найти функцию F*(x) и изобразить ее графически.

Задания для самостоятельной работы

2.3. Для статистического распределения, полученного в задаче 1.3, вычислить $\overline{X}$, S^2, s, Mo, Me, R, V. Найти функцию F*(x) и изобразить ее графически.

Указание. В качестве конкретных значений x берут середины интервалов.

З а н я т и е 3. МЕТОД ПРОИЗВЕДЕНИЙ (МЕТОД УСЛОВНОГО НУЛЯ). ПРАВИЛО СЛОЖЕНИЯ ДИСПЕРСИЙ

Если разность между любыми двумя соседними вариантами есть величина постоянная, то говорят о равноотстоящих вариантах, т.е. $x_{k+1} - x_k = h$ – const, h – шаг. В этом случае удобно выборочные среднюю и дисперсию находить методом произведений по формулам

$\bar{x} = U'h + C$, $S^2 = (U'' - (U')^2)h^2 n/(n-1)$,

где $U' = \sum_{i=1}^{k} U_i n_i / n$ и $U'' = \sum_{i=1}^{k} U^2_i n_i / n$, $U_i = (x_i - C)/h$ – условные варианты, $i = 1,2,...,k$.

C – ложный нуль или новое начало отсчета. В качестве C рекомендуется выбирать варианту, стоящую в середине. В случае четного числа вариант из двух вариант, стоящих в середине, выбирают варианту с наибольшей частотой.

Пример 3.1. Распределение месячной заработной платы 100 строителей задано таблицей:

Заработная плата (сом)	500-600	600-700	700-800	800-900	900-1000	1000-1100	1100-1200
Число рабочих	3	11	20	30	19	12	5

Вычислить $\overline{X}$, S^2, s методом произведений.

Решение

В качестве конкретных значений признака возьмем середины интервалов: $x_1 = 550$, $x_2 = 650$, $x_3 = 750$, $x_4 = 850$, $x_5 = 950$, $x_6 = 1050$, $x_7 = 1150$. Имеем случай равноотстоящих вариант с h = 100. Выберем C = 850.

Тогда

$U_1 = (550 - 850)/100 = -3$, $U_2 = (650 - 850)/100 = -2$,
$U_3 = (750 - 850)/100 = -1$, $U_4 = (850 - 850)/100 = 0$,
$U_5 = (950 - 850)/100 = 1$, $U_6 = (1050 - 850)/100 = 2$,
$U_7 = (1150 - 850)/100 = 3$.

Ряд из условных вариант будет иметь вид:

U_i	-3	-2	-1	0	1	2	3	Итого
n_i	3	11	20	30	19	12	5	100

Вычислим

$U' = (-3\times3 - 2\times11 - 1\times20 + 0\times30 + 1\times19 + 2\times12 + 3\times5)/100 = 0.07$

$U'' = ((-3)^2\times3 + (-2)^2\times11 + (-1)^2\times20 + 1^2\times19 + 2^2\times12 + 3^2\times5)/100 = 2.03$

Тогда $\overline{X} = U'h + C = 0.07\times100 + 850 = 857$

$S^2 = (U'' - (U')^2)h^2 n/(n-1) = (2.03 - (0.07)^2)100^3/99 = 20456$, s = 143.

Пусть вся совокупность разбита на группы. Рассматривая каждую группу как самостоятельную совокупность, можно найти ее среднюю и дисперсию.

Групповой средней называют среднее арифметическое значений признака, принадлежащих группе.

Групповой дисперсией называют дисперсию значений признака, принадлежащих группе, относительно групповой средней.

Если j – номер группы, j = 1,2,…,k, где k – количество групп,

x_i^j – значения признака в j-й группе,

n_i^j – частоты признака в j-й группе,

n_j - объем j-й группы,

то $\bar{x}_j = \sum_{i=1}^{k_j} x_i^j n_i^j / n_j$ групповые средние, а $S_j^2 = \sum_{i=1}^{k_j} (x_{ij} - \bar{x}_j)^2 n_i^j / (n_j - 1)$ – групповые дисперсии.

Внутригрупповой дисперсией называют среднюю арифметическую групповых дисперсий, взвешенных по объемам групп.

$$S^2_{\text{внутр}} = \sum_{j=1}^{k} S^2{}_j n_j \Big/ \sum_{j=1}^{k} n_j .$$

Межгрупповой дисперсией называют дисперсию групповых средних относительно общей средней.

$$S^2_{\text{межгр}} = \sum_{j=1}^{k} (\bar{x}_j - \bar{x}_{\text{общ}})^2 n_j \Big/ \sum_{j=1}^{k} n_j .$$

Среднюю и дисперсию, вычисленные для всей совокупности, называют общими и обозначают $\bar{x}_{\text{общ}}$ и $\bar{S}^2{}_{\text{общ}}$.

Справедлива формула $\bar{S}^2{}_{\text{общ}} = \bar{S}^2{}_{\text{внутр.}} + \bar{S}^2{}_{\text{межгр.}}$, известная как правило сложения дисперсий.

Пример 3.2. Распределение рабочих строительного треста по стажу работы оказалось следующим:

U_I	Стаж работы, лет	Число рабочих			
		СМУ-1	СМУ-2	СМУ-3	Всего
-2	0 – 5	1	13	21	35
-1	5 – 10	5	5	15	25
0	10 – 15	6	4	7	17

U_I	Стаж работы, лет	Число рабочих			
		СМУ-1	СМУ-2	СМУ-3	Всего
1	15 – 20	4	3	4	11
2	20 – 25	3	3	1	7
3	25 – 30	1	2	2	5
Итого		20	30	50	100

Проверить правило сложения дисперсий.

Решение

Конкретные значения признака: $x_1 = 2{,}5$; $x_2 = 7{,}5$; $x_3 = 12{,}5$; $x_4 = 17{,}5$; $x_5 = 22{,}5$; $x_6 = 27{,}5$. Варианты равноотстоящие, с шагом $h = 5$.

Перейдем к условным вариантам $U_i = (x_i - C)/h$, выбрав $C = 12{,}5$. Тогда $U_1 = -2$, $U_2 = -1$, $U_3 = 0$, $U_4 = 1$, $U_5 = 2$, $U_6 = 3$. Запишем их в столбец, левее столбца x. Вся совокупность – строительный трест – разбита на три группы. Вычислим групповые и общие средние и дисперсии методом произведений.

$\overline{U}_1 = (-2\times1 -1\times5 + 1\times4 + 2\times3 + 3\times1)/20 = 0.3$ – средняя из условных вариант для 1-й группы,

$\overline{U}_2 = (-2\times13 -1\times5 + 1\times3 + 2\times3 + 3\times2)/30 = -0.53$ – средняя из условных вариант для 2-й группы,

$\overline{U}_3 = (-2\times21 -1\times15 + 1\times4 + 2\times1 + 3\times2)/50 = -0.9$ – средняя из условных вариант для 3-й группы,

$\overline{U}_{общ} = (-2\times35 -1\times25 + 1\times11 + 2\times7 + 3\times5)/100 = -0.55$ – средняя из условных вариант для всей совокупности.

$\overline{U}_1'' = ((-2)^2\times1 + (-1)^2\times5 + 1^2\times4 + 2^2\times3 + 3^2\times1)/20 = 1.7$,

$\overline{U}_2'' = ((-2)^2\times13 + (-1)^2\times5 + 1^2\times3 + 2^2\times3 + 3^2\times2)/30 = 3$,

$\overline{U}_3'' = ((-2)^2\times21 + (-1)^2\times15 + 1^2\times4 + 2^2\times1 + 3^2\times2)/50 = 2.5$,

$\overline{U}_{общ}'' = ((-2)^2\times35 + (-1)^2\times25 + 1^2\times11 + 2^2\times7 + 3^2\times5)/100 = 2.49$.

Вычислим: $\overline{x}_1 = \overline{U}_1\times h + C = 0.3\times5 + 12.5 = 14$;

$\overline{x}_2 = \overline{U}_2\times h + C = -0.53\times5 + 12.5 = 9.85$;

$\overline{x}_3 = \overline{U}_3\times h + C = -0.9\times5 + 12.5 = 8$;

$\overline{x}_{общ} = \overline{U}_{общ}\times h + C = -0.55\times5 + 12.5 = 9.75$.

Вычислим дисперсии по формуле

$S^2 = (U'' - (U')^2)h^2\, n/(n - 1)$

$S^2_1 = (1.7 - (0.3)^2)\times5^2\times20/19 = 42.37$,

$S^2_2 = (3 - (-0.53)^2)\times5^2\times20/29 = 70.23$,

$S^2_3 = (2.5 - (-0.9)^2)\times5^2\times50/49 = 43.11$.

Теперь вычислим

$S^2_{внутр} = (42.37\times20 + 70.23\times30 + 43.11\times50)/100 = 50.67$,

$S^2_{межгр} = ((14 - 9.75)^2\times20 + (9.85 - 9.75)^2\times30 + (8 - 9.75)^2\times50)/100 = 5.15$,

$\overline{S}^2_{общ} = 50.67 + 5.15 = 55.82$.

Задания для работы в аудитории

3.1. Для определения крепости нити проведено 1000 испытаний

Крепость нити, г	180-190	190-200	200-210	210-220	220-230	230-240	240-250	Всего
Число проб	60	95	135	180	280	150	100	1000

Вычислить $\overline{X}$, S^2, s методом произведений.

3.2. Распределение сотрудников одной организации по размеру заработной платы оказалось следующим:

Заработная плата, сом	Число сотрудников			
	Отд. 1	Отд. 2	Отд. 3	Всего
400 -500	1	3	1	5
500 - 600	2	4	4	10
600 - 700	5	1	14	20
700 - 800	5	8	12	25
800 - 900	5	7	8	20
900 - 1000	2	7	6	15
1000 - 1100	-	-	5	5
Всего	20	30	50	100

Проверить правило сложения дисперсий.

У к а з а н и е : От заданных вариант x, перейти к условным U.

Задания для самостоятельной работы

3.3. Результаты хронометража времени, затрачиваемого одним рабочим на изготовление одной детали, приведены в таблице:

Время, мин	5	7	9	11	Всего
Число рабочих	2	5	12	6	25

Определить среднее время, затрачиваемое одним рабочим на изготовление одной детали.

3.4. Распределение рабочих мебельной фабрики по уровню квалификации было следующим:

Тарифные разряды	Количество рабочих			
	Цех 1	Цех 2	Цех 3	Всего
1	1	6	8	15
2	2	7	21	30
3	3	6	11	20
4	5	5	5	15
5	8	4	3	15
6	1	2	2	5
Всего	20	30	50	100

Проверить правило сложения дисперсий.

З а н я т и е 4. ПОСТРОЕНИЕ ДОВЕРИТЕЛЬНЫХ ИНТЕРВАЛОВ ДЛЯ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ И СРЕДНЕГО КВАДРАТИЧЕСКОГО ОТКЛОНЕНИЯ СЛУЧАЙНОЙ ВЕЛИЧИНЫ, РАСПРЕДЕЛЕННОЙ ПО НОРМАЛЬНОМУ ЗАКОНУ

Статистическую оценку, определяемую одним числом, называют точечной.

Статистическую оценку, определяемую двумя числами - концами интервала, покрывающего оцениваемый параметр, называют интервальной.

Пусть θ^* – статистическая оценка для оцениваемого параметра θ. Оценка θ^* тем точнее оценивает параметр θ, чем меньше разность $|\theta - \theta^*|$. Если $|\theta - \theta^*| < \delta$, $\delta > 0$, то чем меньше δ, тем точнее оценка θ^*. Положительное число δ называют точностью оценки. Найти точность оценки можно только с заданной вероятностью.

Надежностью (доверительной вероятностью) оценки θ по θ^* называют вероятность γ, с которой осуществляется неравенство

$|\theta - \theta^*| < \delta$, $\delta > 0$, т. е. $P(|\theta - \theta^*| < \delta) = \gamma$.

Надежность оценки задается наперед, причем в качестве γ берут число, близкое к единице. Чаще всего задают надежность, равную 0.9; 0.95; 0.99.

Интервал $(\theta^* - \delta;\ \theta^* + \delta)$, который покрывает неизвестный параметр θ с заданной надежностью γ, называют доверительным.

Доверительный интервал и есть интервальная оценка.

Математическое ожидание M(x) = m нормально распределенной величины X оценивается по выборочной средней $\bar{X}$. Доверительный интервал при известном среднем квадратическом отклонении σ имеет вид

$\bar{x} - z\times\sigma/\sqrt{n} < m < \bar{x} + z\times\sigma/\sqrt{n}$,

где n – объем выборки, z – параметр.

Параметр z определяется из условия $F(z) = \gamma/2$, где F(z) – функция Лапласа. Значения этой функции берут из таблиц (см. Приложение 1). Величина $z\times\sigma/\sqrt{n} = \delta$ – точность оценки. При неизвестном среднем квадратическом отклонении его значение заменяется оценкой по выборке и доверительный интервал для математического ожидания принимает вид:

$\bar{x} - t_\gamma\times S/\sqrt{n} < m < \bar{x} + t_\gamma\times S/\sqrt{n}$,

s – несмещенное выборочное среднее квадратическое отклонение, вычисляемое по формуле

$$s = \sqrt{\sum_{i=1}^{k}(x_i - \bar{x}_в)^2 n_i/(n-1)},$$

$t_\gamma(\nu)$ – параметр, значения которого берут из специальных таблиц (см. Приложение 2) в зависимости от объема выборки n и надежности γ. Число степеней свободы ν определяется как объем выборки минус количество параметров, определенных по выборке, $\nu = n - k$, где k – количество параметров, определенных по выборке.

Среднее квадратическое отклонение $\sigma(x)$ нормально распределенной случайной величины оценивают по несмещенному выборочному среднему квадратическому отклонению. Доверительный интервал для σ имеет вид

$s\times(1 - q) < \sigma < s\times(1 + q)$,

q – параметр, значения которого зависят от объема выборки n и надежности γ и берут из таблиц (см. Приложение 3).

Пример 4.1. Найти доверительный интервал для неизвестного математического ожидания нормально распределенного признака с надежностью $\gamma = 0{,}99$, если известно генеральное среднее квадратическое отклонение $\sigma = 3$ и по данным выборки n = 25 вычислена $\bar{x} = 20.12$.

Решение

Доверительный интервал ищем в виде $\bar{x} - z_\gamma\times\sigma/\sqrt{n} < m < \bar{x} + z_\gamma\times\sigma/\sqrt{n}$. Найдем параметр z из условия F(z) = 0.99, а с учетом того, что критическая область двухсторонняя – 0.995.

По таблице (Приложение 1) находим $z_\gamma = 2.58$.

Доверительный интервал

$20.12 - 2.58\times3/\sqrt{25} < m < 20.12 + 2.58\times3/\sqrt{25}$

$8.57 < m < 21.67$.

Пример 4.2. Найти доверительный интервал с надежностью $\gamma = 0.95$ для $M(x) = m$ нормально распределенного признака, если по данным выборки объема $n = 12$ вычислены выборочная средняя $\bar{X}=16.8$ и несмещенное выборочное среднее квадратическое отклонение $s = 1{,}5$.

Решение

Доверительный интервал ищем в виде

$\bar{x} - t_\gamma\times s/\sqrt{n} < m < \bar{x} + t_\gamma\times s/\sqrt{n}$.

Параметр $t_\gamma = t_\gamma(12 - 1, 0.975)$ найдем по таблице (Приложение 2). $t_\gamma = 2.20$.

Тогда

$6.8 - 2.2\times1.5/\sqrt{12} < m < 16.8 + 2.2\times1.5/\sqrt{12}$,

$5.85 < m < 17.75$.

Пример 4.3. Найти с надежностью 0.99 доверительный интервал для σ нормально распределенного признака, если по данным выборки объема $n = 16$ вычислено $s = 0.400$.

Решение

Доверительный интервал ищем в виде

$s\times(1 - q) < \sigma < s\times(1 + q)$

Из таблицы (Приложение 3) находим $q = q(16; 0,.99)$, $q = 1.07$, тогда $0 < q < 0.400(1 +1.07)$, $0 < q < 0.828$.

Задания для работы в аудитории

4.1. Рост призывников нормальная случайная величина со средним квадратическим отклонением $\sigma = 5.92$. Определить, в каких доверительных границах с вероятностью 0.95 находится рост призывников в генеральной совокупности, если по данным выборки объема $n = 1000$ вычислено $\bar{X} = 168$.

4.2. По данным 16 независимых равноточных измерений некоторой физической величины найдены выборочная средняя $\bar{X} = 42.8$ и несмещенное выборочное среднее квадратичное отклонение $s = 8$. Оценить с надежностью 0.99 истинное значение измеряемой величины.

4.5. Выборочное обследование величины вклада в одном из банков по 100 лицевым счетам дало следующие результаты:

Интервалы вклада, $	0-50	50-100	100-500	500-1000	1000-1500	1500-2000	2000-5000
Число счетов	5	10	15	30	25	10	5

Определить с доверительной вероятностью 0.99 возможные пределы для средней величины вклада в данном банке.

4.4. Найти минимальный объем выборки, при котором с надежностью 0.925 точность оценки математического ожидания нормально распределенного признака по выборочной средней будет равна 0.2, если известно $\sigma = 1.5$.

4.5. Произведено 10 измерений одним прибором некоторой физической величины, причем исправленное среднее квадратическое отклонение случайных ошибок измерений оказалось равным 0.8. Найти точность прибора с надежностью 0.95. Точность прибора характеризуется средним квадратическим отклонением случайных ошибок измерений.

Задания для самостоятельной работы

4.6. Испытание крепости отобранных 150 нитей дало следующие результаты:

Крепость нити, г	210-250	250-290	290-330	330-370
Число проб	30	70	40	10

Определить с надежностью 0.99 среднюю крепость нитей всей партии, считая ее нормальной случайной величиной.

4.7. Для определения точности измерительного прибора было проведено 10 независимых измерений, на основании которых вычислена несмещенная выборочная дисперсия $S^2 = 4$ мм^2. Найти с точностью 0.95 доверительный интервал точности этого измерительного прибора.

З а н я т и е 5. КРИТЕРИЙ СОГЛАСИЯ χ^2

Пусть относительно интересующей случайной величины X проведено выборочное наблюдение, давшее результаты:

x_1	x_2	…	x_k
n_1	n_2	…	n_k

На предыдущих занятиях были рассмотрены вопросы оценки неизвестных параметров этой величины. Если неизвестен и закон распределения этой величины, но имеются основания предположить, что он имеет определенный вид (назовем А), то выдвигают гипотезу: генеральная совокупность распределена по закону А.

Статистической называют гипотезу о виде неизвестного распределения или о параметрах известных распределений. Выдвинутая гипотеза может быть правильной или неправильной, поэтому возникает необходимость ее проверки. Для статистической проверки гипотезы о том, что выборочная совокупность (выборка) имеет предполагаемый закон распределения, применяются различные критерии согласия. Критерием согласия (это специально подобранная случайная величина) называют критерий проверки гипотезы о предполагаемом законе неизвестного распределения.

Рассмотрим критерий согласия χ^2 (читается хи-квадрат) или критерий Пирсона, названный так по имени автора, впервые предложившего его. Сравнивают эмпирические (наблюденные) и теоретические (вычисленные в предположении о данном законе распределения) частоты.

$$\chi^2 = \sum_{i=1}^{k} (f_{oi} - f_{ti})^2 / f_{ti} ,$$

где f_{oi} – наблюдаемая частота для каждой группы i;
f_{ti} – теоретическая частота для каждой группы.

Если вычисленное значение критерия равно нулю, то наблюдаемые и теоретические предполагаемые значения частот точно совпадают и распределение выборочной совокупности считается точно совпадающим с предполагаемым нами распределением. Если значение критерия не равно нулю, расхождение существует и проверка значимости расхождения (со статистической точки зрения) проводится по таблицам для выбранного (заданного) уровня значимости α (в некоторой литературе приведена обратная величина – уровень доверительной вероятности $(1 - \alpha)$ и имеющегося числа степеней свободы. Таблицы критерия χ^2 приведены в Приложении 4.

Основные правила применения критерия Пирсона:

1. Объем выборочной совокупности должен быть не менее 100, в противном случае необходимо применять другие критерии, например, Колмогорова-Смирнова или Крамера-фон Мизеса;

2. Группы необходимо составлять таким образом, чтобы в каждой из них частота (как наблюдаемая, так и теоретическая) была не менее 5. Если в группе какая-либо из частот менее 5, то необходимо ее объединять с предшествующей группой. Правила составления групп в общем случае произвольные, но некоторые приемы можно конкретизировать.

Для дискретных распределений понятие группы совпадает с фиксированным значением аргумента.

Для непрерывных распределений проще всего разбить теоретическую функцию плотности предполагаемого распределения на одинаковые по площади участки, границы этих участков и будут границами групп. Теоретическая частота в этом случае будет постоянной для всех групп и равна площади одного участка под теоретической функцией распределения, умноженной на объем выборки. Наблюдаемая частота для каждой группы определяется как количество элементов выборки, попавших в границы конкретного участка. Если непрерывное распределение задано выборкой небольшого числа сгруппированных значений, то можно принять границы участков в середине между каждыми значениями. Левой границей первого участка в этом случае будет $-\infty$, а правой границей последнего участка $+\infty$;

3. Число степеней свободы определяется по формуле $m = k - p - 1$, где k – количество групп после проведения операций объединения, если они оказались необходимы, p – количество параметров, определенных по выборке для построения теоретических частот. В каждом конкретном случае это число различно и определяется видом предполагаемого распределения, с которым мы хотели бы провести сравнение. Например, для распределения Пуассона p = 1, а для нормального распределения p = 2.

Пример 5.1. По результатам наблюдения работы АТС были получены следующие данные о частоте звонков за одноминутный интервал

Число запросов, x	0	1	2	3	4	5
Число одноминутных интервалов с x запросами	315	142	40	9	2	1

Требуется проверить, имеет ли эта выборка распределение Пуассона. Уровень значимости принять равным 0.05.

Решение

Общее число одноминутных интервалов равно 509. Распределение Пуассона имеет только один параметр λ, который нетрудно определить, учитывая, что он также является средним значением.
λ=0×315/509+1×142/509+2×40/509+3×9/509+4×2/509+ 5×1/509=0.5577.

Взяв в формуле Пуассона $P(x) = \lambda^x e^{-\lambda}/x!$ последовательно значения x от 0 до 5, получим теоретические верятности каждой группы. Умножая эти вероятности на общее число одноминутных интервалов, получим теоретические частоты. Все действия внесем в следующую таблицу:

<table>
<tr><th>X</th><th>P(x)</th><th colspan="2">f_t</th><th colspan="2">f_o</th><th>$\chi^2 = \sum_{i=1}^{k}(f_{oi} - f_{ti})^2 / f_{ti}$</th></tr>
<tr><td>0</td><td>0.571</td><td colspan="2">291</td><td colspan="2">315</td><td>1.98</td></tr>
<tr><td>1</td><td>0.319</td><td colspan="2">162</td><td colspan="2">142</td><td>2.47</td></tr>
<tr><td>2</td><td>0.089</td><td colspan="2">45</td><td colspan="2">40</td><td>0.56</td></tr>
<tr><td>3</td><td>0.017</td><td>9</td><td rowspan="3">11</td><td>9</td><td rowspan="3">12</td><td rowspan="3">0.09</td></tr>
<tr><td>4</td><td>0.003</td><td>1</td><td>2</td></tr>
<tr><td>5</td><td>0.001</td><td>1</td><td>1</td></tr>
<tr><td>Всего</td><td>1.000</td><td colspan="2">509</td><td colspan="2">509</td><td>5.10</td></tr>
</table>

Последние три группы объединены в одну, так как частоты в них меньше 5.

Расчетная величина критерия χ^2 = 5.10. По таблице Приложения 4 находим, что для уровня значимости 0.05 и числа степеней свободы m = = 4 – 1 – 1 = 2 критическое значение величины $\chi^2_{кр}$ = 5.99. Следовательно, поскольку расчетная величина критерия меньше табличного критического значения, нельзя отбросить гипотезу, что данная выборка имеет распределение Пуассона.

Пример 5.2. Генератор случайных чисел равномерного распределения выдал 500 цифр в диапазоне от 0 до 9. Частоты, с которыми появлялась каждая цифра, приведены в таблице. Используя уровень значимости 0.01, проверить, насколько полученные результаты соответствуют равномерному закону распределения.

Решение

Если бы цифры генерировались действительно по равномерному закону распределения, то можно было бы ожидать, что каждая цифра появится 50 раз. Ход расчетов иллюстрируется таблицей.

Цифра	0	1	2	3	4	5	6	7	8	9	Всего
f_o	62	58	36	28	40	70	60	40	72	34	500
f_t	50	50	50	50	50	50	50	50	50	50	500
χ^2	2.88	1.28	3.92	9.68	2.00	8.00	2.00	2.00	9.68	5.12	46.56

Табличная величина критерия для m = 10 – 1 числа степеней свободы и уровня значимости 0.01 χ^2 = 21.7. Поскольку $\chi^2_{расч} > \chi^2_{табл}$, то гипотеза о равномерном распределении цифр, полученных с генератора, отвергается.

Пример 5.3. Имеется выборка сгруппированных значений.

X	38.5	41.5	44.5	47.5	50.5	53.5	56.5	59.5	Всего
n	8	12	21	27	15	9	4	4	100

Используя уровень значимости 0.05, проверить, насколько полученные результаты соответствуют нормальному закону распределения.

Определим выборочное среднее

$$\bar{x} = \sum_{i=1}^{k} x_i n_i / n = 47.26.$$

Определим несмещенную оценку дисперсии

$$S^2 = \sum_{i=1}^{k} (x_i - \bar{x})^2 n_i \ /(n - 1) = 26.48.$$

Отсюда оценка среднего квадратического отклонения S = 5.15.

Все последующие действия внесем в таблицу

<table>
<tr><th>x</th><th colspan="2">f_o</th><th>Границы групп по x</th><th>Границы групп по z</th><th>F(z) по границам</th><th>Вероятность попадания в группу</th><th>f_t</th><th>χ^2</th></tr>
<tr><td>38.5</td><td colspan="2">8</td><td>-∞, 40</td><td>-∞, -1.41</td><td>0, 0.0793</td><td>0.0793</td><td>7.9</td><td>0.0013</td></tr>
<tr><td>41.5</td><td colspan="2">12</td><td>40, 43</td><td>-1.41, -0.83</td><td>0.0793, 0.2033</td><td>0.124</td><td>12.4</td><td>0.0129</td></tr>
<tr><td>44.5</td><td colspan="2">21</td><td>43, 46</td><td>-0.83, -0.24</td><td>0.2033, 0.409</td><td>0.2057</td><td>20.6</td><td>0.0078</td></tr>
<tr><td>47.5</td><td colspan="2">27</td><td>46, 49</td><td>-0.24, 0.34</td><td>0.409, 0.6331</td><td>0.2241</td><td>22.4</td><td>0.9446</td></tr>
<tr><td>50.5</td><td colspan="2">15</td><td>49, 52</td><td>0.34, 0.92</td><td>0.6331, 0.8212</td><td>0.1881</td><td>18.8</td><td>0.7681</td></tr>
<tr><td>53.5</td><td colspan="2">9</td><td>52, 55</td><td>0.92, 1.50</td><td>0.8212, 0.9332</td><td>0.112</td><td>11.2</td><td>0.4321</td></tr>
<tr><td>56.5</td><td>4</td><td rowspan="2">8</td><td rowspan="2">55, +∞</td><td rowspan="2">1.50, +∞</td><td rowspan="2">0.9332, 1.0</td><td rowspan="2">0.0668</td><td rowspan="2">6.7</td><td rowspan="2">0.2522</td></tr>
<tr><td>59.5</td><td>4</td></tr>
<tr><td colspan="6">Итого</td><td>1.0</td><td>100</td><td>2.4190</td></tr>
</table>

Границы групп по x (столбец 3), в соответствии с приведенными выше рекомендациями, выбраны по середине между значениями x. Последние две группы объединены, так как частоты в них менее 5.

Границы групп по z (столбец 4) определены нормированием и центрированием значений границ по x с использованием формулы

$$z_i = (x_i - \bar{X})/s.$$

Значения функции распределения вероятности F(z) (столбец 5) для каждой границы определены по таблице для стандартного нормального закона распределения, приведенной в Приложении 1.

Вероятность попадания в группу P_i (столбец 6) определена как разность значений функции распределения между границами каждой группы.

Величина теоретической частоты для каждой группы (столбец 7) рассчитана из выражения по формуле $f_t = P_i \times n$.

Замечание. Для проверки правильности расчетов рекомендуется определять сумму строк по столбцам 6 и 7. Эти суммы должны быть равны соответственно 1.0 и n (в данном примере n = 100).

Расчетная величина критерия $\chi^2 = 2.419$. По таблице Приложения 4 находим, что для уровня значимости 0.05 и числа степеней свободы $m = 7 - 2 - 1 = 4$ критическое значение величины $\chi^2 = 9.49$. Следовательно, поскольку расчетная величина критерия меньше табличного критического значения, нельзя отбросить гипотезу, что данная выборка имеет нормальное распределение.

Задания для работы в аудитории

5.1. По данным, приведенным в задаче 3.1, построить теоретическую нормальную кривую. Проверить гипотезу о нормальном распределении при уровне значимости $\alpha = 0.01$, используя критерий χ^2.

5.2. Установить, пользуясь критерием χ^2 при уровне значимости 0.05 случайно или значимо расхождение между эмпирическими и теоретическими частотами?

f_o	4	18	32	70	20	36	10
f_t	10	24	34	80	18	22	12

Задания для самостоятельной работы

5.3. Считая, что генеральная совокупность, из которой была произведена выборка в задаче 3.3, распределена по нормальномузакону, построить теоретическую нормальную кривую. Проверить согласованность гипотезы о нормальном распределении при уровне значимости 0.025, используя критерий χ^2.

5.4. Установить, пользуясь критерием χ^2, при уровне значимости 0.05, случайно или значимо расхождение между эмпирическими и теоретическими частотами, которые вычислены исходя из предположения, что генеральная совокупность распределена нормально.

f_o	5	7	15	14	21	16	9	7	6
f_t	6	6	14	15	22	15	8	8	6

Часть II. КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ

При рассмотрении двух случайных величин X и Y может оказаться, что либо они связаны некоторой зависимостью, либо независимы. Из математического анализа, согласно определению взаимно однозначной функциональной зависимости, каждому значению одной переменной соответствует только одно определенное значение другой переменной. Такого вида зависимость на практике встречается сравнительно редко. При изучении взаимосвязей между различными величинами весьма часто приходится встречаться с таким положением, когда каждому значению одной переменной соответствует несколько значений другой, т.к. обе величины или одна из них подвержены еще действию случайных факторов. Иными словами, каждому значению одной переменной соответствует закон распределения другой переменной. Подобного вида зависимость называется статистической. Например, рассматривается зависимость между себестоимостью продукции (признак Y) и объемом продукции (признак X) некоторого количества однотипных предприятий. Ясно, что при одной и той же величине объема продукции X на различных предприятиях наблюдается разное значение величины себестоимости Y, т.е. каждому значению X соответствует некоторое распределение Y. В приведенном примере признаки X, Y связаны статистической зависимостью. Пусть для значения X = x признак Y принимает значения y_1, y_2, …y_n. Можно вычислить среднее значение признака Y, соответствующее X= x, называемое условным средним и обозначаемое

$$\bar{y}_x = \frac{y_1 + y_2 + ... + y_n}{n}.$$

Математически зависимость Y от X выражается в виде функциональной зависимости условной средней от аргумента x

$$\bar{y}_x = f(x). \qquad (1)$$

Уравнение (1) называют уравнением регрессии Y на X, а функцию f(x) – регрессией Y на X, а ее график – линией регрессии Y на X.

Аналогично можно рассматривать зависимость X от Y, тогда

$$\bar{x}_y = \varphi(y) \qquad (2)$$

будет уравнение регрессии X на Y, φ(y) – регрессия X на Y, а ее график – линия регрессии X на Y. Регрессии, заданные уравнениями (1), (2), называются сопряженными. Признак, играющий роль аргумента в уравнении регрессии, называется факторным, а признак, играющий роль функции, – результативным. Во многих случаях оба признака X, У целесообразно считать равноправными. При изучении зависимости таких признаков необходимо рассматривать обе сопряженные регрессии

$\hat{y} = f(x)$ и € = φ(y).

Количественное изучение зависимости заключается в решении двух основных вопросов:

1. Определение формы зависимости (вида функций) f(x) и φ (y);
2. Установление силы (тесноты) связи между X и Y.

З а н я т и е 6. ОТЫСКАНИЕ ПАРАМЕТРОВ ЛИНЕЙНОЙ ЗАВИСИМОСТИ И ОЦЕНКА ТЕСНОТЫ СВЯЗИ (СЛУЧАЙ НЕГРУППИРОВАННЫХ ДАННЫХ)

Корреляционная зависимость считается линейной, если графиком регрессии является прямая линия. Регрессию в этом случае ищут в виде

$$\hat{y} = kx + b \qquad (3)$$

или

$$\text{€} = ky + b. \qquad (4)$$

Неизвестные параметры k и b находят методом наименьших квадратов. Пусть в результате n независимых опытов получено n пар чисел $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$. Эти данные можно записать в виде следующей таблицы:

X	x_1	x_2	…	x_n
Y	y_1	y_2	…	y_n

Согласно методу наименьших квадратов

$$J(k,b) = \sum_{i=1}^{n} (\text{€}_i - y_i)^2 = \sum_{i=1}^{n} (kx_i + b - y_i)^2 \to \min .$$

Необходимое условие минимума функции J(k,b*)* приводит к системе

$$\begin{cases} k\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i \\ k\sum_{i=1}^{n} x_i + nb = \sum_{i=1}^{n} y_i \end{cases} \qquad (5)$$

Решив систему (5) относительно неизвестных коэффициентов k, b и подставив в уравнение (3), получим регрессию y на x.

Замечание: при отыскании уравнения регрессии (4) x и y необходимо поменять ролями.

В случае линейной зависимости тесноту связи оценивают с помощью выборочного коэффициента корреляции r, вычисляемого по формуле

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{(n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2)(n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2)}} . \qquad (6)$$

Основные свойства коэффициента корреляции:

1. $-1 \le r \le 1$.
2. Если |r| =1, то признаки X и Y связаны функциональной зависимостью.
3. Если r = 0, то признак Y не коррелирован с признаком X.
4. Если от данных вариант X, Y перейти к условным U и V, то формула вычисления r не меняется (X заменяется на U, Y заменяется на V).
5. Чем ближе |r| к единице, тем связь сильнее; чем ближе |r| к нулю, тем связь слабее. Значимость коэффициента корреляции можно проверить по таблице (см. Приложение 5) в зависимости от объема используемой выборки и принятого (заданного) уровня значимости.

Пример 6.1. За 12 месяцев зарегистрированы следующие данные о величине средней выработки на одного рабочего (Y, тыс. сом) и удельном весе стоимости покупных полуфабрикатов в общей стоимости продукции (X) по каждому месяцу:

X	37	39	43	40	36	45	36,5	39	40	35	41	42
У	6,3	7	11	8	5	12	7,5	8	10	6,1	9	9,5

Найти уравнение прямой регрессии y на x и оценить тесноту связи.

Решение. Уравнение регрессии Y на X ищем в виде $\overline{y}$ = kx + b. Параметры k и b найдем из системы (5).

Составим вспомогательную таблицу для определения коэффициентов системы:

I	x_i	y_i	x_i^2	x_iy_I	y_i^2
1	37	6,3	1369	233,1	39,69
2	39	7	1521	273	49
3	43	11	1849	473	121
4	40	8	1600	320	64
5	36	5	1296	180	25
6	45	12	2025	540	144
7	36,5	7,5	1332,25	273,75	56,25
8	39	8	1521	312	64
9	40	10	1600	400	100
10	35	6,1	1225	213,5	37,21
11	41	9	1681	369	81
12	42	9,5	1764	399	90,25
$\sum$	473,5	99,4	18783,25	3986,35	871,4

Система принимает вид

$$\begin{cases} 18783{,}25k + 473{,}5b = 3986{,}35 \\ 473{,}5k + 12b = 99{,}4 \end{cases}$$

Решив систему, получаем: k ≈ 0.64, b ≈ - 17.11.

Уравнение линейной регрессии Y на X.

y = 0.64x - 17.11.

Оценим тесноту связи, вычислив r

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{473{,}5}{12} \approx 39{,}46; \qquad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{99{,}4}{12} \approx 8{,}28;$$

$$\overline{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n} = \frac{3986{,}35}{12} \approx 332{,}20; \qquad \overline{x^2} = \frac{\sum_{i=1}^{n} x_i^2}{n} = \frac{18783{,}25}{12} \approx 1565{,}27;$$

$$\overline{y^2} = \frac{\sum_{i=1}^{n} y_i^2}{n} = \frac{871{,}4}{12} \approx 72{,}62.$$

$$s_x = \sqrt{S_x^2} = \sqrt{\overline{x^2} - (\bar{x})^2} = \sqrt{1565{,}27 - (39{,}46)^2} \approx 2{,}86.$$

$$s_y = \sqrt{S_y^2} = \sqrt{\overline{y^2} - (\bar{y})^2} = \sqrt{72{,}62 - (8{,}28)^2} \approx 2{,}02.$$

$$r = \frac{332{,}20 - 39{,}46 \cdot 8{,}28}{2{,}86 \cdot 2{,}02} \approx \frac{5{,}4712}{5{,}7772} \approx 0{,}95.$$

По таблице 5 Приложения для уровня значимости α = 0.05 и числа степеней свободы 12 -2 = 10 найдем критическое значение коэффициента корреляции r(α, ν) = 0.576. Так как расчетное значение больше критического, то гипотеза о независимости X и У отвергается.

Задания для работы в аудитории

Задача 6.1. Данные о среднесписочной численности работающих (X, чел.) и валовой продукции (Y, млн. сом) группы однородных предприятий заданы таблицей

X	273	200	300	366	121	99	88	96	191	85
Y	19.1	1.0	16.1	1.2	6.2	3.0	2.7	4.8	10.3	2.4

Найти уравнение линейной регрессии Y на X и оценить тесноту связи.

Задача 6.2. Имеются следующие данные о посевной площади кормовых культур (X, тыс. га) и среднегодовом поголовье скота (Y) группы однородных фермерских хозяйств

X	0,5	1,8	1,4	1,1	1,7	1,2	0,6	1,4	1,3	1,1	1,2	1,5
Y	180	418	400	220	400	230	150	300	200	248	220	300

Считая, что X и Y связаны линейной зависимостью, оценить тесноту связи.

Задания для самостоятельной работы

Задача 6.3. В таблице приведены данные, характеризующие зависимость среднесуточного съема стали с 1 м^2 площади пола мартеновских печей Y (т) от простоя X (%)

X	35,4	22,8	29,6	16,6	14,5
Y	2,01	4,43	3,67	5,63	6,55

Найти уравнение линейной регрессии Y на X и оценить тесноту связи.

Задача 6.4. Результаты исследования изменения веса поросят Y (кг) с возрастом X (недель) заданы таблицей

X	0	1	2	3	4	5	6	7	8
Y	1,3	2,5	3,9	5,2	6,3	7,5	9,0	10,8	13,1

Оценить тесноту связи между X и Y для случая линейной зависимости.

ЗАНЯТИЕ 7. ОТЫСКАНИЕ ПАРАМЕТРОВ ЛИНЕЙНОЙ ЗАВИСИМОСТИ И ОЦЕНКА ТЕСНОТЫ СВЯЗИ (СЛУЧАЙ СГРУППИРОВАННЫХ ДАННЫХ)

Пусть в результате n независимых опытов признак X принимал каждое значение x_i - n_{xi} раз, признак Y – каждое значение y_i - n_{yj} раз, а каждая пара (x_i, y_j) наблюдалась n_{xiyj} раз. В этом случае данные группируют и записывают в виде таблицы, которую называют корреляционной.

Y	X				
	x_1	x_2	…	x_k	n_y
y_1	$n_{x_1y_1}$	$n_{x_2y_1}$	…	$n_{x_ky_1}$	n_{y_1}
y_2	$n_{x_1y_2}$	$n_{x_2y_2}$	…	$n_{x_ky_2}$	n_{y_2}
…	…	…	…	…	…
y_e	$n_{x_1y_e}$	$n_{x_2y_e}$	…	$n_{x_ky_e}$	n_{y_e}
n_x	n_{x_1}	n_{x_2}	…	n_{x_k}	n

$$\sum n_x = \sum n_y = n$$

Замечание: если какие-то пары (x, y) не наблюдались, то в соответствующей клетке ставят прочерк.

Уравнения регрессий Y на X и X на Y в этом случае ищут в виде

$$\frac{\overline{y_x} - \overline{y}}{\sigma_y} = r\frac{x - \overline{x}}{\sigma_x} \qquad (7)$$

$$\frac{\overline{x_y} - \overline{x}}{\sigma_x} = r\frac{y - \overline{y}}{\sigma_y} \qquad (8)$$

r – выборочный коэффициент корреляции.

Пример 7.1. Результаты обследования роста X и веса Y 50 учеников приведены в таблице

X	Y					
	22,5-25,5	25,5-28,5	28,5-31,5	31,5-34,5	34,5-37,5	n_x
117,5-122,5	1	3	-	-	-	4
122,5-127,5	-	2	6	1	-	9
127,5-132,5	-	1	5	5	-	11
132,5-137,5	-	1	6	7	2	16
137,5-142,5	-	-	1	4	2	7
142,5-147,5	-		-	1	1	2
147,5-152,5	-		-	-	1	1
n_y	1	7	18	18	6	50

Найти уравнения линейных регрессий У на X и X на У и построить их графики. Оценить тесноту связи между X и Y.

Решение. Уравнения регрессий (мы имеем случай сгруппированных данных) ищем в виде (7) и (8). Вычислим $\bar{x}$, $\bar{y}$, s_x, s_y и r. В качестве конкретных значений признаков X и Y следует взять середины интервалов, т.е. x_1=120, x_2=125, x_3=130, x_4=135, x_5=140, x_6=145, x_7=150, y_1=24, y_2=27, y_3=30, y_4=33, y_5=36.

Значения признаков представляют равноотстоящие варианты. Можно перейти к условным вариантам

$$U_i = \frac{x_i - C_x}{h_x}, \quad \vartheta_j = \frac{y_j - C_y}{h_y}, \; i = \overline{1,7}, \quad j = \overline{1,5}.$$

C_x=135 и – C_y = 30 “условные нули” признаков X и Y.
h_x=5 – шаг для X, h_y= 3 – шаг для Y.
Тогда корреляционная таблица в условных вариантах:

U	V					
	-2	-1	0	1	2	n_u
-3	1	3	-	-	-	4
-2	-	2	6	1	-	9
-1	-	1	5	5	-	11
0	-	1	6	7	2	16
1	-	-	1	4	2	7
2	-	-	-	1	1	2
3	-	-	-	-	1	1
n_v	1	7	18	18	6	50

Вычислим:

$$\overline{U} = \frac{\sum_i U_i n_i}{n} = \frac{-3\cdot 4 + (-2)\cdot 9 + (-1)\cdot 11 + 1\cdot 7 + 2\cdot 2 + 3\cdot 1}{50} = -0{,}54.$$

$$\overline{U^2} = \frac{\sum_i U_i^2 n_i}{n} = \frac{(-3)^2\cdot 4 + (-2)^2\cdot 9 + (-1)^2\cdot 11 + 1^2\cdot 7 + 2^2\cdot 2 + 3^2\cdot 1}{50} = 2{,}14.$$

$$\overline{\vartheta} = \frac{\sum_j \vartheta_j n_j}{n} = \frac{-2\cdot 1 + (-1)\cdot 7 + 1\cdot 18 + 2\cdot 6}{50} = 0{,}42.$$

$$\overline{\vartheta^2}=\frac{\sum\limits_j \vartheta_j^2\cdot n_j}{n}=\frac{4\cdot 1+7+18+4\cdot 6}{50}=1{,}06.$$

Тогда

$\bar{x}=\bar{U}\cdot h_x+C_x=-0{,}54\cdot 5+135=132{,}3$.

$\bar{y}=\bar{\vartheta}\cdot h_y+C_y=0{,}42\cdot 3+30=31{,}26$.

$$s_x=\sqrt{S_x^2}=\sqrt{(\overline{U^2}-(\bar{U})^2)\cdot h_x^2}=h_x\sqrt{\overline{U^2}-(\bar{U})^2}=$$

$$=5\sqrt{2{,}14-(-0{,}54)^2}\approx 5\cdot 1{,}36=6{,}68.$$

$$s_y=\sqrt{S_y^2}=\sqrt{(\overline{\vartheta^2}-(\bar{\vartheta})^2)\cdot h_y^2}=h_y\sqrt{\overline{\vartheta^2}-(\bar{\vartheta})^2}=$$

$$=3\sqrt{1{,}06-(0{,}42)^2}\approx 3\cdot 0{,}94=2{,}82.$$

Для вычисления r используем свойство: если от данных вариант X, Y перейти к условным U и ϑ, то выборочный коэффициент корреляции

$$r=\frac{\sum n_{u\vartheta}\cdot U\cdot\vartheta-n\bar{U}\bar{\vartheta}}{n\cdot\sigma_u\cdot\sigma_\vartheta},\text{ где}$$

$$s_u=\sqrt{\overline{U^2}-(\bar{U})^2}=\sqrt{2{,}14-(-0{,}54)^2}\approx 1{,}36.$$

$$s_\vartheta=\sqrt{\overline{\vartheta^2}-(\bar{\vartheta})^2}=\sqrt{1{,}06-(0{,}42)^2}=0{,}94.$$

Вычислим сумму

$$\sum n_{U\vartheta}\cdot U\cdot\vartheta=(-3)\cdot(-2)\cdot 1+(-3)\cdot(-1)\cdot 3+(-2)\cdot(-1)\cdot 2+(-2)\cdot 6\cdot 0+$$
$$+(-2)\cdot 1\cdot 1+(-1)\cdot(-1)\cdot 1+(-1)\cdot 5\cdot 0+(-1)\cdot 5\cdot 1+1\cdot 4\cdot 1+1\cdot 2\cdot 2+2\cdot 1\cdot 1+$$
$$+2\cdot 1\cdot 2+3\cdot 1\cdot 2=33.$$

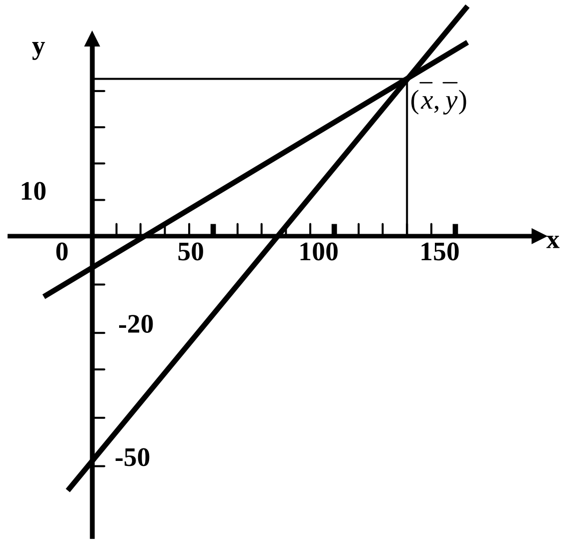
Выборочный коэффициент корреляции

$$r=\frac{33-50\cdot(-0{,}54)\cdot 0{,}42}{50\cdot 1{,}36\cdot 0{,}94}=\frac{44{,}34}{63{,}92}\approx 0{,}69.$$

Уравнения регрессий принимают вид

$\frac{\overline{y_x}-31{,}26}{2{,}82}=0{,}69\frac{x-132{,}3}{6{,}8}$ и $\frac{\overline{x_y}-132{,}3}{6{,}8}=0{,}69\frac{y-31{,}26}{2{,}82}$ или y = 0.29x - 6.6

– уравнение линейной регрессии Y на X, а x = 1.66y + 80.29 – уравнение линейной регрессии X на Y. Построим графики этих прямых



Прямые пересекаются в точке $(\bar{x},\bar{y})$.

Задания для работы в аудитории

Задача 7.1. В таблице представлены данные распределения возрастов мужей (Y) и жен (X) (X, Y представляют середины возрастных интервалов, а частоты, указанные в таблице, выражены в десятках тыс. супружеских пар).

Y	X							
	20	30	40	50	60	70	80	n_y
20	19	5	-	-	-	-	-	24
30	23	116	11	-	-	-	-	150
40	1	41	98	9	-	-	-	149
50	-	4	32	65	7	-	-	108
60	-	1	4	21	36	3	-	65
70	-	-	1	2	11	13	1	28
80	-	-	-	-	1	3	2	6
n_x	43	167	146	97	55	19	3	530

Составить уравнения регрессии Y на X и X на Y и оценить тесноту связи.

Задача 7.2. В таблице представлены данные анализа 100 проб руды, добытой на руднике, относительно содержания окиси железа в процентах (X) и закиси железа в процентах (Y). Считая, что признаки связаны линейной зависимостью, оценить тесноту связи.

X	Y						
	0-6	6-12	12-18	18-24	24-30	30-36	n_x
30 – 40	-	-	-	1	-	1	2
40 – 50	-	-	1	5	4	5	15
50 – 60	-	-	2	18	10	2	32
60 – 70	-	6	14	2	2	-	24
70 – 80	-	6	3	-	-	-	9
80 - 90	4	8	-	-	-	-	12
90 - 100	6	-	-	-	-	-	6
n_y	10	20	20	26	16	8	100

Задания для самостоятельной работы

Задача 7.3. В результате группировки 50 промышленных предприятий по стоимости основных фондов в млн. сом (X) и объему валовой продукции в млн. сом (Y) получена корреляционная таблица

X	Y							
	2	5	8	11	14	16	20	n_y
2	2	18	7	3	-	-	-	30
5	-	4	5	3	2	1	-	15
10	-	-	-	-	-	2	3	5
n_y	2	22	12	6	2	3	3	50

Построить прямые регрессий Y на X и X на Y. Оценить тесноту линейной корреляционной зависимости.

Задача 7.4. Найти уравнение линейной регрессии У на X по данным исследования 100 опытных участков по урожайности Y (ц/га) и срокам уборки после достижения полной спелости X (в днях).

X	Y							
	18	20	22	24	26	28	30	n_x
5	-	-	-	-	2	8	10	20
10	-	-	4	6	10	7	3	30
15	-	2	8	10	7	3	-	30
20	2	3	5	6	4	-	-	20
n_y	2	5	17	22	23	18	13	100

З а н я т и е 8. КРИВОЛИНЕЙНАЯ КОРРЕЛЯЦИЯ.
ОЦЕНКА ТЕСНОТЫ СВЯЗИ НЕЛИНЕЙНОЙ ЗАВИСИМОСТИ

Корреляцию называют нелинейной, если график регрессии y = f(x) или x = φ(y) изображается кривой линией. На практике чаще всего (исходя из удобства вычислений) рассматривают следующие функции регрессии:

$y = ax^2 + bx + c$ – параболическая зависимость 2-го порядка,

$y = ax^3 + bx^2 + cx + d$ – параболическая зависимость 3-го порядка,

$y = a/x + b$ – гиперболическая зависимость.

Неизвестные параметры, входящие в каждое из уравнений, находят методом наименьших квадратов, который приводит к системе уравнений линейных относительно неизвестных коэффициентов.

Для негруппированных данных при параболической зависимости 2-го порядка имеем систему уравнений

$$\begin{cases} a\sum_{i=1}^{n} x_i^4 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i^2 y_i \\ a\sum_{i=1}^{n} x_i^3 + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i \\ a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i + nc = \sum_{i=1}^{n} y_i \end{cases} \qquad (9)$$

Если данные сгруппированы, то система уравнений для этой же зависимости имеет вид

$$\begin{cases} a\sum_{i=1}^{k} n_{x_i} x_i^4 + b\sum_{i=1}^{k} n_{x_i} x_i^3 + c\sum_{i=1}^{k} n_{x_i} x_i^2 = \sum_{i=1}^{k} n_{x_i} x_i^2 y_i \\ a\sum_{i=1}^{k} n_{x_i} x_i^3 + b\sum_{i=1}^{k} n_{x_i} x_i^2 + c\sum_{i=1}^{k} n_{x_i} x_i = \sum_{i=1}^{k} n_{x_i} x_i y_i \\ a\sum_{i=1}^{k} n_{x_i} x_i^2 + b\sum_{i=1}^{k} n_{x_i} x_i + nc = \sum_{i=1}^{k} n_{x_i} y_i \end{cases} \qquad (10)$$

Для гиперболической зависимости неизвестные коэффициенты можно определить из следующей системы уравнений (случай негруппированных данных):

$$\begin{cases} a\sum_{i} \frac{1}{x_i^2} + b\sum_{i} \frac{1}{x_i} = \sum_{i}^{n} \frac{y_i}{x_i} \\ a\sum \frac{1}{x_i} + nb = \sum_{i} y_i \end{cases} \qquad (11)$$

Для оценки тесноты связи криволинейной корреляционной зависимости пользуются характеристикой, которая называется корреляционным отношением и вычисляется по формуле

$$\eta_{y/x} = \frac{s_{межгр}}{s_{общ}},$$

где $s_{межгр}$ – межгрупповое среднее квадратическое отклонение;

$s_{общ}$ – общее среднее квадратическое отклонение.

$$s_{межгр} = \sqrt{\frac{\sum(\overline{y}_{x_i} - \overline{y})^2 n_{x_i}}{n}}, \qquad s_{общ} = \sqrt{\frac{\sum(y_i - \overline{y})^2 n_{y_i}}{n}}.$$

Тогда $\eta_{y/x} = \sqrt{\dfrac{\sum\limits_i(\overline{y}_{x_i} - \overline{y})^2 n_{x_i}}{\sum\limits_i(y_i - \overline{y})^2 n_{y_i}}}$.

Корреляционное отношение обладает свойствами:

1. $0 \le \eta \le 1$;
2. При $\eta = 0$ признак Y с признаком X корреляционной зависимостью не связан;
3. При $\eta = 1$признак Y связан с признаком X функциональной зависимостью;
4. С ростом корреляционного отношения η от 0 до 1 связь между Yи X становится более тесной, приближаясь к функциональной;
5. $\eta \ge |r|$;
6. При $\eta = |r|$ имеет место точная линейная корреляционная зависимость.

Замечания:

1. Корреляционное отношение служит мерой тесноты связи любой, в том числе и линейной зависимости (выборочный коэффициент корреляции r оценивает тесноту лишь линейной зависимости).
2. При нахождении регрессии X на Y теснота связи определяется по формуле

$$\eta_{x/y} = \sqrt{\frac{\sum(\overline{x}_y - \overline{x})^2 n_y}{\sum(x - \overline{x})^2 n_x}}.$$

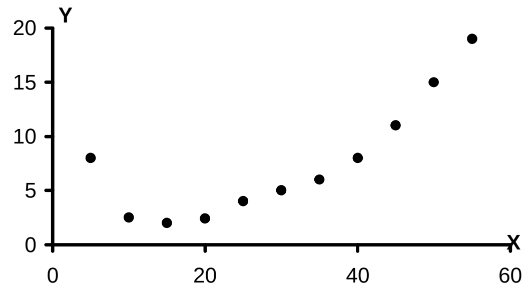
Пример 8.1. В таблице приведены значения повозрастного коэффициента смертности (Y) для разных возрастов (X), вычисленные для населения определенного города. (Повозрастной коэффициент смертности равен

отношению числа умерших данного возраста к числу живущих того же возраста в течение определенного года в процентах).

X	5	10	15	20	25	30	35	40	45	50	55
Y	8	2,5	2,0	2,4	4	5	6	8	11	15	19

Исследовать форму зависимости между Y и X. Найти уравнение регрессии Y на X и оценить тесноту связи.

Решение. Для определения формы зависимости между Y и X поступим так: каждую пару значений (x_i, y_i) из таблицы изобразим в виде точки (x_i, y_i) в системе координат xOy.



Построенное таким образом изображение корреляционной таблицы называется полем корреляции. Расположение точек говорит о том, что уравнение регрессии целесообразно искать в виде

$y = ax^2 + bx + c$.

Параметры a, b, c найдем из системы (9). Составим вспомогательную таблицу

i	x_i	x_i^2	x_i^3	x_i^4	y_i	x_iy_i	$x_i^2y_i$
1	5	25	125	625	8	40	200
2	10	100	1000	10000	2,5	25	250
3	15	225	3375	50625	2	30	450
4	20	400	8000	160000	2,4	48	960
5	25	625	15625	390625	4	100	2500
6	30	900	27000	810000	5	150	4500
7	35	1225	42875	1500625	6	210	7350
8	40	1600	64000	2560000	8	320	12800
9	45	2025	91125	4100625	11	495	22275
10	50	2500	125000	6250000	15	750	37500
11	55	3025	166375	9150625	19	1045	57425
Σ	330	12650	544500	24983750	82,9	3213	146210

Система для отыскания a, b, c

$$\begin{cases} 24983750a + 544500b + 12650C = 146210 \\ 544500a + 12650b + 330C = 3213 \\ 12650a + 330b + 11C = 82{,}9 \end{cases}$$

Решив систему любым известным методом, получим
a = 0,0136, b = – 0,55, c = 8,46
Уравнение регрессии Y на X принимает вид

$$\overline{y_x} = 0{,}0136x^2 - 0{,}55x + 8{,}46$$

Тесноту связи оценим по $\eta_{y/x}$, для чего вычислим:

$$\overline{y_{x_1}} = 0{,}0136 \cdot 5^2 - 0{,}55 \cdot 5 + 8{,}46 = 6{,}04$$

$$\overline{y_{x_2}} = 0{,}0136 \cdot 10^2 - 0{,}55 \cdot 10 + 8{,}46 = 4{,}32$$

$$\overline{y_{x_3}} = 0{,}0136 \cdot 15^2 - 0{,}55 \cdot 15 + 8{,}46 = 3{,}27$$

$$\overline{y_{x_4}} = 0{,}0136 \cdot 20^2 - 0{,}55 \cdot 20 + 8{,}46 = 2{,}9$$

$$\overline{y_{x_5}} = 0{,}0136 \cdot 25^2 - 0{,}55 \cdot 25 + 8{,}46 = 2{,}71$$

$$\overline{y_{x_6}} = 0{,}0136 \cdot 30^2 - 0{,}55 \cdot 30 + 8{,}46 = 4{,}20$$

$$\overline{y_{x_7}} = 0{,}0136 \cdot 35^2 - 0{,}55 \cdot 35 + 8{,}46 = 5{,}87$$

$$\overline{y_{x_8}} = 0{,}0136 \cdot 40^2 - 0{,}55 \cdot 40 + 8{,}46 = 8{,}22$$

$$\overline{y_{x_9}} = 0{,}0136 \cdot 45^2 - 0{,}55 \cdot 45 + 8{,}46 = 11{,}25$$

$$\overline{y_{x_{10}}} = 0{,}0136 \cdot 50^2 - 0{,}55 \cdot 50 + 8{,}46 = 14{,}96$$

$$\overline{y_{x_{11}}} = 0{,}0136 \cdot 55^2 - 0{,}55 \cdot 55 + 8{,}46 = 19{,}35$$

Теперь вычислим $\bar{y}$ по формуле средней арифметической

$$\bar{y} = \frac{\sum y}{n}; \quad \bar{y} = \frac{82{,}9}{11} \approx 7{,}54;$$

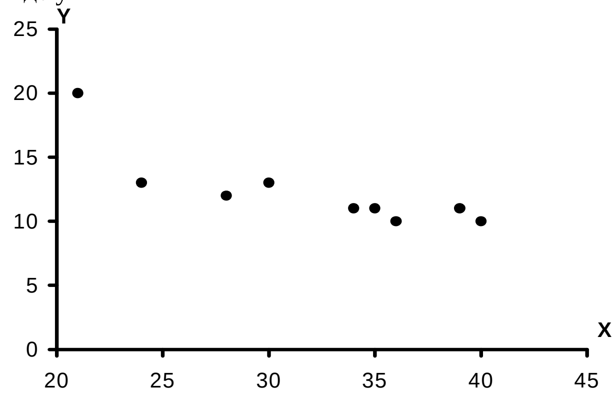
$$\eta_{y/x} = \sqrt{\frac{\sum(\overline{y_x} - \bar{y})^2}{\sum(y - \bar{y})^2}}; \quad \eta_{y/x} = \sqrt{\frac{298{,}4133}{303{,}2546}} \approx 0{,}99.$$

Пример 8.2. Результаты анализа зависимости себестоимости тонны угля (Y) от среднемесячной производительности труда рабочего (X) приведены в таблице

X, т	21	24	28	30	34	35	36	39	40
Y, руб.	20	13	12	13	11	11	10	11	10

Исследовать форму зависимости между Y и X и найти уравнение регрессии Y на X.

Решение. Построим в системе координат xOy точки (x_i, y_i), i = 1,9.
По расположению точек видно, что уравнение регрессии целесообразно искать в виде y = a/x + b



Неизвестные параметры a и b найдем из системы

$$\begin{cases} a\sum_{i=1}^{9}\frac{1}{x_i^2} + b\sum_{i=1}^{9}\frac{1}{x_i} = \sum_{i=1}^{9}\frac{y_i}{x_i} \\ a\sum_{i=1}^{9}\frac{1}{x_i} + nb = \sum_{i=1}^{9} y_i \end{cases}$$

Составим вспомогательную расчетную таблицу

x_i	$1/x_i$	$1/x_i^2$	y	y_i/x_i
21	0,0476	0,0023	20	0,9524
24	0,0417	0,0017	13	0,5417
28	0,0357	0,0013	12	0,4286
30	0,0333	0.0011	13	0,4333
34	0,0294	0,0009	11	0,3235
35	0,0286	0,0008	11	0,3143
36	0,0278	0,0008	10	0,2778
39	0,0256	0,0007	11	0,2821
40	0,025	0,0006	10	0,25
Σ	0,2947	0,0102	111	3,8037

Система для отыскания параметров

$$\begin{cases} 0{,}0102a + 0{,}2947b = 3{,}8037 \\ 0{,}2947a + 9b = 111 \end{cases}$$

Решив систему, получаем $a \approx 303.06$; $b \approx 2.3$. Уравнение регрессии Y на X имеет вид $y = 303.6/x + 2.3$.

Задания для работы в аудитории

Задача 8.1. В результате лабораторных испытаний прочности стальных проволок различных диаметров были получены следующие данные

Диаметр проволоки X, мм	0,6	2	2,2	2,45	2,6
Разрывное усилие Y, кг	50	560	690	760	900

Полагая, что между X и Y существует параболическая зависимость, найти уравнение регрессии Y на X в виде $y = ax^2 + bx + c$ и оценить тесноту связи.

Задача 8.2. В результате исследования корреляционной зависимости себестоимости единицы продукции Y от объема выпускаемой продукции за год X (тыс. т) получены следующие данные:

X	85	237	414	551	820	1150
У	18,4	12,1	11,5	10,3	9,6	7,7

Найти уравнение регрессии вида $y = a/x + b$. Оценить тесноту связи Y с X.

Задача 8.3. Найти уравнение регрессии $y = ax^2 + bx + c$ и вычислить корреляционное отношение по данным корреляционной таблицы

Y\X	0	4	6	n_y
1	50	5	1	56
35	-	44	-	44
50	-	5	45	50
n_x	50	54	46	150

Задания для самостоятельной работы

Задача 8.4. Группировка цементных заводов по размерам производства за год представлена в таблице

Производство продукции X, тыс. т	0-100	100-200	200-300	300-450	450-600
Средняя выработка на одного рабочего Y, т	210	360	390	550	600

Найти уравнение регрессии вида $y = ax^2 + bx + c$ и оценить тесноту связи.

Задача 8.5. В таблице приведены данные о количестве выпускаемых деталей X, тыс. шт. и о полных затратах на их изготовление Y (сотни сом).

X	2	4	4	2	2	18	4	2	9	9	4	9	18	9	4
Y	19	-3	4	20	24	5	8	26	3	7	13	10	8	12	14

Исследовав форму зависимости, найти уравнение регрессии Y на X и оценить тесноту связи.

Задача 8.6. В результате группировки 100 га посевов по количеству внесенных удобрений X и урожайности Y (ц/га) получена корреляционная таблица

Y\X	12	14	16	18	20	22	24	n_X
10	7	2	1	-	-	-	-	10
30	2	10	4	3	1	-	-	20
50	-	3	17	18	2	-	-	40
70	-	-	1	7	11	7	4	30
N_y	9	15	23	28	14	7	4	100

Вычислить корреляционное отношение $\eta_{y/x}$.

ЗАДАНИЕ ДЛЯ ТИПОВОГО РАСЧЕТА 1

По статистическим данным, полученным в результате опыта, требуется:

1. Произвести группировку; построить статистическое распределение выборки.
2. Построить график полученного статистического распределения.
3. Найти эмпирическую функцию распределения F*(x).
4. Построить график функции F*(x).
5. Вычислить выборочную среднюю $\overline{x}_в$; выборочную дисперсию S^2; выборочное среднее квадратическое отклонение S; моду Mo; медиану Me.
6. С надежностью 0.99 найти доверительный интервал для истинного значения рассматриваемой величины.
7. Построить теоретическую нормальную кривую.
8. Пользуясь критерием χ^2 при уровне значимости 0.01 и предполагая, что генеральная совокупность имеет нормальное распределение, установить случайно или значимо расхождение между формой распределения выборки и генеральной совокупности.

№ варианта и исходные данные, n = 100										
1.	40.26	40.35	40.44	40.35	40.39	40.40	40.42	40.32	40.36	40.33
	40.37	40.35	40.44	40.35	40.30	40.34	40.31	40.32	40.31	40.33
	40.33	40.41	40.35	40.30	40.33	40.38	40.33	40.33	40.36	40.34
	40.28	40.30	40.40	40.36	40.32	40.32	40.42	40.35	40.38	40.35
	40.29	40.33	40.31	40.33	40.36	40.34	40.30	40.30	40.34	40.31
	40.41	40.40	40.33	40.37	40.34	40.30	40.43	40.34	40.38	40.35
	40.35	40.34	40.34	40.31	40.43	40.36	40.34	40.34	40.32	40.35
	40.28	40.44	40.32	40.34	40.31	40.31	40.36	40.34	40.35	40.35
	40.29	40.39	40.39	40.37	40.37	40.38	40.36	40.41	40.38	40.42
	40.27	40.38	40.37	40.37	40.36	40.35	40.32	40.36	40.38	40.37
2.	40.33	40.27	40.36	40.45	40.36	40.40	40.41	40.43	40.25	40.38
	40.33	40.38	40.36	40.45	40.31	40.31	40.35	40.32	40.36	40.29
	40.34	40.34	40.42	40.36	40.36	40.34	40.39	40.34	40.32	40.32
	40.35	40.29	40.31	40.41	40.38	40.33	40.33	40.43	40.27	40.31
	40.31	40.30	40.34	40.32	40.34	40.37	40.35	40.31	40.28	40.35
	40.35	40.42	40.41	40.34	40.38	40.35	40.31	40.44	40.40	40.33
	40.35	40.36	40.35	40.35	40.32	40.44	40.37	40.34	40.34	40.42
	40.35	40.29	40.45	40.33	40.35	40.32	40.32	40.37	40.27	40.30
	40.42	40.30	40.40	40.40	40.38	40.38	40.39	40.37	40.28	40.36
	40.37	40.28	40.39	40.38	40.38	40.37	40.36	40.33	40.26	40.35

№ варианта и исходные данные, n = 100										
3.	40.34	40.25	40.43	40.34	40.38	40.39	40.41	40.31	40.41	40.42
	40.34	40.36	40.43	40.34	40.29	40.33	40.30	40.31	40.32	40.36
	40.40	40.32	40.34	40.29	40.32	40.37	40.32	40.32	40.35	40.40
	40.29	40.27	40.39	40.35	40.31	40.31	40.41	40.34	40.34	40.34
	40.32	40.28	40.30	40.32	40.35	40.33	40.29	40.29	40.38	40.36
	40.39	40.40	40.32	40.36	40.33	40.29	40.42	40.33	40.36	40.32
	40.33	40.34	40.33	40.30	40.42	40.29	40.33	40.33	40.45	40.38
	40.43	40.27	40.31	40.33	40.30	40.30	40.35	40.33	40.33	40.33
	40.38	40.28	40.38	40.36	40.36	40.37	40.35	40.40	40.39	40.40
	40.37	40.26	40.36	40.36	40.35	40.34	40.31	40.35	40.38	40.37
4.	40.37	40.41	40.42	40.44	40.34	40.28	40.37	40.37	40.41	40.31
	40.37	40.32	40.36	40.33	40.35	40.39	40.37	40.46	40.30	40.31
	40.32	40.35	40.40	40.35	40.34	40.35	40.43	40.46	40.32	40.32
	40.38	40.34	40.34	40.44	40.37	40.30	40.32	40.42	40.41	40.34
	40.35	40.38	40.36	40.32	40.32	40.31	40.35	40.33	40.29	40.29
	40.39	40.36	40.32	40.34	40.36	40.43	40.42	40.36	40.42	40.33
	40.33	40.45	40.38	40.36	40.36	40.37	40.36	40.34	40.33	40.33
	40.36	40.33	40.33	40.38	40.36	40.30	40.46	40.41	40.35	40.33
	40.39	40.39	40.40	40.38	40.43	40.31	40.41	40.39	40.35	40.40
	40.39	40.38	40.37	40.34	40.38	40.29	40.40	40.23	40.31	40.35
5.	40.30	40.24	40.33	40.42	40.28	40.37	40.38	40.40	40.41	40.31
	40.30	40.35	40.39	40.42	40.33	40.28	40.32	40.29	40.30	40.31
	40.31	40.31	40.33	40.33	40.33	40.31	40.36	40.31	40.32	40.32
	40.33	40.26	40.28	40.38	40.34	40.34	40.30	40.40	40.41	40.34
	40.28	40.27	40.31	40.29	40.31	40.32	40.32	40.28	40.29	40.29
	40.32	40.39	40.38	40.31	40.35	40.41	40.28	40.41	40.42	40.33
	40.32	40.33	40.32	40.32	40.29	40.29	40.34	40.32	40.33	40.33
	40.32	40.26	40.42	40.30	40.32	40.35	40.29	40.34	40.35	40.33
	40.39	40.27	40.37	40.37	40.35	40.34	40.36	40.34	40.35	40.40
	40.34	40.25	40.36	40.35	40.35	40.30	40.33	40.30	40.31	40.35

№ варианта и исходные данные, n = 102												
6.	440	443	475	418	462	433	438	469	465	500	524	491
	479	442	463	423	447	462	498	497	473	485	489	440
	442	451	423	475	452	526	490	511	492	509	472	428
	456	417	452	468	475	502	507	477	443	432	487	
	499	434	505	449	533	488	444	504	493	512	452	
	501	424	511	458	497	445	497	478	528	456	509	
	485	459	489	430	515	469	491	456	493	519	522	
	475	471	454	487	479	429	479	452	449	478	485	
	502	484	477	459	466	451	505	431	429	462	444	

№ варианта и исходные данные, n = 102												
7.	448	445	423	480	467	438	443	474	469	505	529	496
	484	447	468	428	452	457	503	502	478	490	494	445
	447	456	428	480	457	531	495	516	497	514	477	461
	433	422	457	473	480	507	512	482	450	509	442	
	502	498	442	509	515	463	479	449	451	522	451	
	488	490	456	522	479	423	442	458	417	485	475	
	445	456	473	485	466	452	456	430	468	444	533	
	499	519	492	442	519	452	462	487	428	521	526	
	501	477	443	451	478	431	526	459	457	509	502	
8.	413	435	428	470	457	428	433	464	460	495	519	486
	474	437	458	418	442	457	493	492	468	480	483	435
	437	446	418	470	447	521	485	506	487	504	467	423
	451	402	447	463	470	497	502	472	440	423	445	
	445	512	424	445	511	494	488	473	505	468	447	
	447	456	459	469	489	477	445	509	511	428	456	
	456	519	471	429	480	442	469	522	489	457	422	
	422	478	484	456	515	479	429	521	454	442	498	
	498	462	430	519	479	466	451	509	487	512	469	
9.	450	443	485	428	483	443	448	479	475	510	534	501
	489	452	473	433	457	462	508	507	483	495	498	450
	452	461	433	485	462	536	500	521	502	519	482	438
	466	427	462	478	485	512	517	487	455	493	468	
	442	495	519	462	447	475	423	432	417	528	449	
	456	480	483	447	456	468	452	512	434	493	458	
	473	504	467	452	422	449	505	456	424	449	430	
	492	423	445	475	498	458	511	519	459	429	487	
	443	468	447	533	490	430	489	478	471	428	446	
10	423	430	465	408	452	423	428	459	455	490	514	481
	469	432	453	413	437	442	488	487	463	475	478	430
	432	441	413	465	442	516	480	501	482	499	462	418
	446	407	442	458	465	492	497	467	435	433	485	
	474	402	424	418	497	506	468	495	498	462	447	
	437	512	459	470	494	472	487	480	482	519	456	
	451	456	471	463	477	473	440	504	468	483	422	
	445	519	484	445	442	509	505	423	449	467	498	
	447	478	430	469	479	522	515	468	458	495	465	

№ варианта и исходные данные для сгруппированных данных									
11.	x	308	401	404	405	505	508	605	695
	n	5	9	18	24	18	12	7	7
12.	x	12	14	16	17	18	20	22	25
	n	8	20	22	8	15	13	12	2

№ варианта и исходные данные для сгруппированных данных									
13.	x	10	12	14	16	18	20	22	24
	n	10	12	14	14	14	14	12	20
14.	x	1991	1992	1993	1994	1995	1996	1997	1998
	n	2	3	20	25	35	5	15	10
15.	x	1.5	1.7	1.8	1.9	2.0	2.2	2.3	2.4
	n	12	13	14	15	10	10	16	20
16.	x	1	2	3	4	5	6	7	9
	n	15	15	15	15	15	15	15	15
17.	x	12.1	12.2	12.3	12.5	12.6	12.7	12.8	12.9
	n	20	18	19	40	50	30	20	3
18.	x	3	5	7	9	11	13	15	17
	n	4	5	15	25	46	34	9	1
19.	x	-6	-4	-3	-2	-1	0	1	2
	n	3	10	40	70	5	6	2	1
20.	x	-10	0	10	20	30	40	50	55
	n	4	5	7	10	30	50	40	10

ЗАДАНИЯ ДЛЯ ТИПОВОГО РАСЧЕТА 2

По данным, полученным в результате выборочного наблюдения, требуется:

1. Составить корреляционную таблицу.
2. Найти уравнения линейных регрессий Y на X и X на Y.
3. Оценить тесноту связи.
4. Построить графики регрессий.

В вариантах 1–4 приводятся данные исследования заработной платы (условные единицы) плотника Y и его тарифного разряда X.

1.

Порядковый номер рабочего	1	2	3	4	5	6	7	8	9	10	11	12
Тарифный разряд	1	1	1	2	2	2	3	3	3	4	4	6
Заработная плата	3	3	3	4	4	5	5	5	6	7	6	10

Порядковый номер рабочего	13	14	15	16	17	18	19	20	21	22	23	24	25
Тарифный разряд	5	4	6	6	5	4	4	5	6	3	4	5	6
Заработная плата	8	7	7	6	8	8	7	9	8	6	5	7	8

2.

Порядковый номер рабочего	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Тарифный разряд	3	1	4	6	6	2	3	4	5	4	5	5	6	2	3
Заработная плата	3	2	7	9	8	3	4	6	7	6	8	7	8	4	5

3.

Порядковый номер рабочего	1	2	3	4	5	6	7	8	9	10
Тарифный разряд	1	1	1	2	2	2	3	3	4	4
Заработная плата	3	3	3	4	4	5	5	5	6	7

4.

Порядковый номер рабочего	1	2	3	4	5	6	7	8	9	10	11	12
Тарифный разряд	6	4	4	5	1	5	2	3	4	6	4	3
Заработная плата	9	6	5	7	3	9	3	3	7	10	7	4

Порядковый номер рабочего	13	14	15	16	17	18	19	20
Тарифный разряд	6	3	4	3	5	3	2	5
Заработная плата	9	4	8	4	8	5	6	4

В вариантах 5–8 приводятся результаты измерений предела прочности в кг/см^2 Y и предела текучести в кг/мм^2 X разных марок стали.

5.

Y	77	96	86	92	98	53	63	80	64	66	81
X	81	77	76	86	53	47	36	40	49	60	54

57	86	80	87	63	153	133	159	134	129
40	61	68	88	145	136	129	126	96	100

145	142	120	95
95	106	118	109

6.

Y	107	103	140	149	147	104	108	93	124	112
X	107	120	114	113	123	94	83	73	107	94

113	95	112	116	93	96	112	136	104	103
107	99	106	104	88	84	94	152	98	77

115	123	111	127	129
88	94	76	84	73

7.

Y	96	121	143	146	130	135	160	134	154	104
X	110	119	107	96	101	97	127	130	137	146

88	81	87	68	67	65	81	64	54	99
89	89	69	62	41	55	61	50	48	54

93	87	97	78
87	77	78	82

8.

Y	94	97	113	137	105	104	116	124	112	128
X	89	85	95	153	99	78	89	95	77	84

130	108	134	141	150	148	105	109	94	125
74	108	121	115	114	123	95	84	74	108

113	114	96	113	116
95	108	100	101	105

В вариантах 9–12 приведены результаты лабораторного анализа образцов сланцевых пород на содержание кремния и алюминия.

9.

Кремний SiO_2	57,8	54,6	54,8	57,7	61,1	62,3	52,2
Алюминий Al_2O_3	17,2	17,9	18,8	19,9	16,0	17,8	18,8

49,2	53,9	60,0	56,2	55,2	53,3	57,9	54,0	52,6
19,3	16,1	14,8	17,0	17,8	19,9	17,1	15,5	17,6

53,8	53,6	51,5	54,0	50,4	53,0	53,3	51,6	50,9
16,3	17,2	15,8	15,0	14,4	15,3	16,6	14,9	14,7

10.

Кремний SiO_2	48,8	53,5	52,8	52,9	52,1	47,3	49,8
Алюминий Al_2O_3	16,4	15,9	15,9	14,8.	19,8	18,7	20,2

49,3	50,1	54,4	49,0	48,9	51,3	51,6	46,2	50,4
17,6	19,2	18,2	16,8	18,2	19,7	19,6	19,1	20,2

50,7	53,1	52,9	51,3
21,5	21,3	20,3	20,1

11.

Кремний SiO_2	51,5	57,6	54,4	59,9	62,1	52,0	49,0
Алюминий Al_2O_3	20,0	17,3	18,0	18,9	20,0	16,1	17,9

53,7	59,8	56,0	55,0	53,1	57,7	53,8	524	53,6
18,9	19,4	16,2	14,9	17,1	17,9	20,0	17,2	15,6

53,4	51,3	53,8	50,2
17,7	15,9	15,1	14,5

12.

Кремний SiO_2	53,0	53,7	49,0	53,1	52,3	475	50,0
Алюминий Al_2O_3	16,0	16.0	16,5	14.9	19.9	18,8	20.3

49,5	50,3	54,6	49,2	49,1	51,5	51,8	46,4	50,6
17,7	19,3	18,3	16,9	18,3	19,8	19,7	19,2	20,3

50,9	53,3	53,1	51,5
21,6	21,4	20,4	20,2

В вариантах 13–15 приведены результаты исследования стоимости основных производственных фондов в млн. сом X и объемов строительно-монтажных работ, выполненных в течение года Y.

13.

X	8,35	8,74	9,25	9,5	9,75	10,24	13,65	15,25	14,51
Y	3,5	1,49	6,40	4,50	5,0	7,0	9,5	12,50	9,50

10,50	10,75	10,76	11,0	11,0	11,25	14,50	14,23	16,25
6,0	2,5	5,74	8,5	5,26	8,0	10,0	8,4	12,0

11,35	11,50
9,5	6,0

14.

X	11,75	12,0	13,75	16,0	12,15	12,25	12,35
Y	10,0	9,6	8,51	11,5	6,0	8,05	5,01

12,50	12,76	12,85	14,75	14,26	12,85	13,15	13,25	13,26
7,03	7,53	6,01	12,0	10,0	9,5	9,02	6,49	10,5

13,4	13,5	14,0	16,0
7,51	10,0	11,0	13,0

15.

X	11,62	11,5	10,25	13,66	15,26	14,52	12,16
Y	9,2	6,4	10,4	7,5	10,0	9,02	6,4

12,26	12,36	12,51	12,77	12,86	13,16	13,26	13,27	13,41
9,2	6,4	10,4	7,5	10,0	9,02	6,4	10,4	7,5

13,51	14,0	16,0	14,52
9,8	11,0	12,0	9,4

Литература

1. Общий курс высшей математики для экономистов /Под ред. проф. В.И. Ермолаева. – М.: Инфра-М, 1999.
2. *Колемаев В.А., Калинина В.Н.* Теория вероятностей и математическая статистика. – М.: Инфра-М, 1999.
3. *Феллер В.* Введение в терию вероятностей и ее приложения. – М.: Мир, 1984.
4. Справочник по теории вероятностей и математической статистике. – М.: Наука, 1985.
5. *Мюллер П.* и др. Таблицы по математической статистике. – М.: Финансы и статистика, 1982.
6. *Поллард Дж.* Справочник по вычислительным методам статистики. – М.: Финансы и статистика, 1982.
7. *Гмурман В.Е.* Руководство к решению задач по теории вероятностей и математической статистике. – М.: Высшая школа, 1999.

ПРИЛОЖЕНИЯ

Приложение 1

Функция стандартного нормального распределения

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp(-x^2/2)dx$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5971	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6806	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.857	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Распределение Стьюдента

ν	$t_{0.995}$	$t_{0.99}$	$t_{0.975}$	$t_{0.95}$	$t_{0.9}$
1	63.66	31.82	12.71	6.31	3.08
2	9.92	6.96	4.30	2.92	1.89
3	5.84	4.54	3.18	2.35	1.64
4	4.60	3.75	2.78	2.13	1.53
5	4.03	3.36	2.57	2.02	1.48
6	3.71	3.14	2.45	1.94	1.44
7	3.50	3.00	2.36	1.90	1.42
8	3.36	2.90	2.31	1.86	1.40
9	3.25	2.82	2.26	1.83	1.38
10	3.17	2.76	2.23	1.81	1.37
11	3.11	2.72	2.20	1.80	1.36
12	3.06	2.68	2.18	1.78	1.36
13	3.01	2.65	2.16	1.77	1.35
14	2.98	2.62	2.14	1.76	1.34
15	2.95	2.60	2.13	1.75	1.34
16	2.92	2.58	2.12	1.75	1.34
17	2.90	2.57	2.11	1.74	1.33
18	2.88	2.55	2.10	1.73	1.33
19	2.86	2.54	2.09	1.73	1.33
20	2.84	2.53	2.09	1.72	1.32
21	2.83	2.52	2.08	1.72	1.32
22	2.82	2.51	2.07	1.72	1.32
23	2.81	2.50	2.07	1.71	1.32
24	2.80	2.49	2.06	1.71	1.32
25	2.79	2.48	2.06	1.71	1.32
26	2.78	2.48	2.06	1.71	1.32
27	2.77	2.47	2.05	1.70	1.31
28	2.76	2.47	2.05	1.70	1.31
29	2.76	2.46	2.04	1.70	1.31
30	2.75	2.46	2.04	1.70	1.31
40	2.70	2.42	2.02	1.68	1.30
60	2.66	2.39	2.00	1.67	1.30
120	2.62	2.36	1.98	1.66	1.29
∞	2.58	2.33	1.96	1.645	1.28

Таблица значений $q = q(\gamma, n)$

n	γ			n	γ		
	0.95	0.99	0.999		0.95	0.99	0.999
5	1.37	2.67	5.64	20	0.37	0.58	0.88
6	1.09	2.01	3.88	25	0.32	0.49	0.73
7	0.92	1.62	2.98	30	0.28	0.43	0.63
8	0.80	1.38	2.42	35	0.26	0.38	0.56
9	0.71	1.20	2.06	40	0.24	0.35	0.50
10	0.65	1.08	1.80	45	0.22	0.32	0.46
11	0.59	0.98	1.60	50	0.21	0.30	0.43
12	0.55	0.90	1.45	60	0.188	0.269	0.38
13	0.52	0.83	1.33	70	0.174	0.245	0.34
14	0.48	0.78	1.23	80	0.161	0.226	0.31
15	0.46	0.73	1.15	90	0.151	0.211	0.29
16	0.44	0.70	1.07	100	0.143	0.198	0.27
17	0.42	0.66	1.01	150	0.115	0.160	0.221
18	0.40	0.63	0.96	200	0.099	0.136	0.185
19	0.39	0.60	0.92	250	0.089	0.120	0.162

Приложение 4

Распределение $\chi 2$

ν	$\chi^2_{0.995}$	$\chi^2_{0.99}$	$\chi^2_{0.975}$	$\chi^2_{0.95}$	$\chi^2_{0.90}$
1	7,88	6,63	5,02	3,84	2,71
2	10,60	9,21	7,38	5,99	4,61
3	12,84	11,34	9,35	7,81	6,25
4	14,96	13,28	11,14	9,49	7,78
5	16,7	15,1	12,8	11,1	9,2
6	18,5	16,8	14,4	12,6	10,6
7	20,3	18,5	16,0	14,1	12,0
8	22,0	20,1	17,5	15,5	13,4
9	23,6	21,7	19,0	16,9	14,7
10	25,2	23,2	20,5	18,3	16,0
11	26,8	24,7	21,9	19,7	17,3
12	28,3	26,2	23,3	21,0	18,5
13	29,8	27,7	24,7	22,4	19,8
14	31,3	29,1	26,1	23,7	21,1
15	32,8	30,6	27,5	25,0	22,3
16	34,3	32,0	28,8	26,3	23,5
17	35,7	33,4	30,2	27,6	24,8
18	37,2	34,8	31,5	28,9	26,0
19	38,6	36,2	32,9	30,1	27,2
20	40,0	37,6	34,2	31,4	28,4
21	41,4	38,9	35,5	32,7	29,6
22	42,8	40,3	36,8	33,9	30,8
23	44,2	41,6	38,1	35,2	32,0
24	45,6	43,0	39,4	36,4	33,2
25	49,6	44,3	40,6	37,7	34,4
26	48,3	45,6	41,9	38,9	35,6
27	49,6	47,0	43,2	40,1	36,7
28	51,0	48,3	44,5	41,3	37,9
29	52,3	49,6	45,7	42,6	39,1
30	53,7	50,9	47,0	43,8	4,03
40	66,8	63,7	59,3	55,8	51,8
50	79,5	76,2	71,4	67,5	63,2
60	92,0	88,4	83,3	79,1	74,4
70	104,2	100,4	95,0	90,5	85,5
80	116,3	112,3	106,6	101,9	96,6
90	128,3	124,1	118,1	113,1	107,6
100	140,2	135,8	129,6	124,3	118,5

Приложение 5

Критические значения для коэффициента корреляции

m	q			m	q		
	0.95	0.975	0.995		0.95	0.975	0.995
1	0.9877	0.9969	0.9999	55	0.2201	0.2609	0.3385
2	0.9000	0.9500	0.9900	60	0.2108	0.2500	0.3248
3	0.8054	0.8783	0.9587	65	0.2027	0.2405	0.3127
4	0.7293	0.8114	0.9172	70	0.1954	0.2319	0.3017
5	0.6694	0.7545	0.8745	75	0.1889	0.2242	0.2919
6	0.6215	0.7067	0.8343	80	0.1829	0.2172	0.2830
7	0.5822	0.6664	0.7977	85	0.1775	0.2108	0.2748
8	0.5494	0.6319	0.7646	90	0.1726	0.2050	0.2673
9	0.5214	0.6021	0.7348	95	0.1680	0.1996	0.2604
10	0.4973	0.5760	0.7079	100	0.1638	0.1946	0.2540
12	0.4575	0.5324	0.6614	110	0.1562	0.1857	0.2425
14	0.4259	0.4973	0.6226	120	0.1496	0.1779	0.2324
16	0.4000	0.4683	0.5897	130	0.1438	0.1710	0.2235
18	0.3783	0.4438	0.5614	140	0.1386	0.1648	0.2155
20	0.3598	0.4227	0.5368	150	0.1339	0.1593	0.2084
22	0.3438	0.4044	0.5131	160	0.1297	0.1543	0.2019
24	0.3297	0.3882	0.4958	170	0.1258	0.1497	0.1959
26	0.3172	0.3739	0.4785	180	0.1223	0.1455	0.1905
28	0.3061	0.3610	0.4629	190	0.1191	0.1417	0.1855
30	0.2960	0.3494	0.4487	200	0.1161	0.1381	0.1809
32	0.2869	0.3388	0.4357	250	0.1039	0.1236	0.1620
34	0.2785	0.3291	0.4238	300	0.0948	0.1129	0.1480
36	0.2709	0.3202	0.4128	350	0.0878	0.1046	0.1371
38	0.2638	0.3120	0.4026	400	0.0822	0.0978	0.1283
40	0.2573	0.3044	0.3932	450	0.0775	0.0922	0.1210
42	0.2512	0.2973	0.3843	500	0.0735	0.0875	0.1149
44	0.2455	0.2907	0.3761	600	0.0671	0.0799	0.1049
46	0.2403	0.2845	0.3683	700	0.0621	0.0740	0.0972
48	0.2353	0.2787	0.3610	800	0.0581	0.0692	0.0909
50	0.2306	0.2732	0.3541	1000	0.0520	0.0619	0.0813

Содержание

Ч А С Т Ь I. АНАЛИЗ РАСПРЕДЕЛЕНИЙ 3
Занятие 1. Построение статистического распределения выборки. График статистического распределения 3
Занятие 2. Статистические оценки параметров распределения................7
Занятие 3. Метод произведений (метод условного нуля). Правило сложения дисперсий 11
Занятие 4. Построение доверительных интервалов для математического ожидания и среднего квадратического отклонения случайной величины, распределенной по нормальному закону.................................15
Занятие 5. Критерий согласия χ^2 18
Ч А С Т Ь II. КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ 24
Занятие 6. Отыскание параметров линейной зависимости и оценка тесноты связи (случай негруппированных данных).................................25
Занятие 7. Отыскание параметров линейной зависимости и оценка тесноты связи (случай группированных данных) 29
Занятие 8. Криволинейная корреляция. Оценка тесноты связи нелинейной зависимости...34
Задание для типового расчета 1...41
Задание для типового расчета 2...44
Литература...48
Приложения
Приложение 1. Функция стандартного нормального распределения 50
Приложение 2. Распределение Стьюдента 51
Приложение 3. Таблица значений q = q(γ, n) 52
Приложение 4. Распределение χ^2 53
Приложение 5. Критические значения для коэффициента корреляции..54

РУКОВОДСТВО К РЕШЕНИЮ ЗАДАНИЙ
ПО МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Методическое пособие
для студентов дневного и заочного обучения

Редактор Т.К. Песковая
Технический рдактор Т.К. Песковая
Корректор О.А. Матвеева
Компьютерная верстка Д.Р. Зайнулиной

Подписано к печати 28.05.01. Формат 60×84 $^1/_{16}$.
Печать офсетная. Объем 3,4 п.л.
Тираж 200 экз. Заказ 120.

Издательство Кыргызско-Российского Славянского университета
720000, Бишкек, Киевская, 44

Отпечатано в типографии КРСУ
720000, Бишкек, Шопокова, 68

КЫРГЫЗСКО-РОССИЙСКИЙ
СЛАВЯНСКИЙ УНИВЕРСИТЕТ

КАФЕДРА МАТЕМАТИКИ

Т.А. Давидюк,
Ш.А. Ильясов,
Т.К. Кадыров

РУКОВОДСТВО
К РЕШЕНИЮ ЗАДАНИЙ
ПО МАТЕМАТИЧЕСКОЙ
СТАТИСТИКЕ

Методическое пособие

Бишкек 2001