Министерство сельского хозяйства Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Саратовский государственный аграрный университет
имени Н. И. Вавилова»

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

**Методические указания и задания
для выполнения типового расчета
по курсу «Математика»**

Направление подготовки
**35.03.07 Технология производства и
переработки сельскохозяйственной продукции**

Профиль подготовки
Технологии пищевых производств в АПК

Саратов 2018

Математическая статистика: методические указания и задания для выполнения типового расчета по дисциплине «Математика» направления подготовки 35.03.07 Технология производства и переработки сельскохозяйственной продукции, / сост. Т.В.Кириллова //ФГБОУ ВО «Саратовский ГАУ».- Саратов, 2018.-

Методические указания и задания для выполнения типового расчета по дисциплине «Математика» составлены в соответствии с программой и предназначены для студентов направления подготовки 35.03.07 Технология производства и переработки сельскохозяйственной продукции. Они содержат рекомендации, примеры и задания к выполнению типового расчета. Позволяют студентам освоить основные методы математической статистики, необходимые для анализа процессов и явлений в ходе поиска оптимальных решений практических задач.

В данном пособии представлена также необходимая для выполнения типового расчета теоретическая информация и примеры ее применения.

В конце приводится список литературы, которую можно порекомендовать студентам для изучения данного раздела математики.

Содержние

1. ОБЩИЕ МЕТОДИЧЕСКИЕ УКАЗАНИЯ

Типовой расчет «Математическая статистика» содержит 2 задачи. Первые задача содержит девять заданий, они связано с исследованием выборки, вторая задача содержит три задания, связанные с корреляционно-регрессионным анализом выборки. Задания типового расчета охватывают следующие разделы математической статистики:

- основы выборочного метода и элементы статистической теории оценивания;
- методы статистической проверки гипотез;
- статистическое исследование зависимостей (корреляционный и регрессионный анализ).

Перед выполнением типового расчета необходимо изучить соответствующие разделы литературы и закрепить с помощью упражнений для самостоятельной работы основные понятия, определения и методы математической статистики.

Так же перед решением заданий рекомендуется ознакомиться со всеми примерами, рассмотренными ниже. По каждой задаче типового расчета в методических указаниях приводится основной теоретический материал и разбирается типовой пример.

Типовой расчет сдается студентом на проверку после изучения всех разделов математической статистики.

Защита осуществляется в письменной форме во время занятий по расписанию Повторная защита – проводится вне сетки расписания в письменной форме или в форме собеседования. Работа выполняется на листах формата А4 (210х297), которые затем скрепляются. Решение заданий следует сопровождать краткими пояснениями. Исходные данные для заданий типового расчета выбирается с номерами вариант, которые соответствуют номеру в списке группового журнала.

2. КРАТКИЕ ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ И ПРИМЕРЫ ТИПОВЫХ ЗАДАНИЙ

Тема 1

Основы выборочного метода и элементы статистической теории оценивания. Методы статистической проверки гипотез

Определим основные понятия математической статистики.

Генеральная совокупность – все множество имеющихся объектов.

Выборка – набор объектов, случайно отобранных из генеральной совокупности.

Объем генеральной совокупности N и объем выборки n – число объектов в рассматриваемой совокупности.

Виды выборки:

Повторная – каждый отобранный объект перед выбором следующего возвращается в генеральную совокупность;

Бесповторная – отобранный объект в генеральную совокупность не возвращается.

Замечание. Для того, чтобы по исследованию выборки можно было сделать выводы о поведении интересующего нас признака генеральной совокупности, нужно, чтобы выборка правильно представляла пропорции генеральной совокупности, то есть была **репрезентативной** (представительной). Учитывая закон больших чисел, можно утверждать, что это условие выполняется, если каждый объект выбран случайно, причем для любого объекта вероятность попасть в выборку одинакова.

Первичная обработка результатов.

Пусть интересующая нас случайная величина X принимает в выборке значение x_1 n_1 раз, x_2 – n_2 раз, …, $x_к$ – $n_к$ раз, причем $\sum_{i=1}^{k} n_k = n$, где n – объем выборки. Тогда наблюдаемые значения случайной величины x_1, x_2,…, $x_к$ называют **вариантами**, а n_1, n_2,…, $n_к$ – **частотами**. Если разделить каждую частоту на объем выборки, то получим **относительные частоты** $w_i = \frac{n_i}{n}$. Последовательность вариант, записанных в порядке возрастания, называют **дискретным вариационным рядом,** а перечень вариант и соответствующих им частот или относительных частот – **статистическим рядом** или **статистическим распределением выборки:**

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k
w_i	w_1	w_2	…	w_k

Если исследуется некоторый непрерывный признак, то вариационный ряд может состоять из очень большого количества чисел. В этом случае удобнее использовать **группированную выборку**. Для ее получения интервал, в котором заключены все наблюдаемые значения признака, разбивают на несколько равных частичных интервалов длиной h, а затем находят для каждого частичного интервала n_i – сумму частот вариант, попавших в i-й интервал. Составленная по этим результатам таблица называется **группированным статистическим рядом** или **интервальным вариационным рядом:**

Номера интервалов	1	2	…	k
Границы интервалов	$(a, a + h)$	$(a + h, a + 2h)$	…	$(b – h, b)$
Сумма частот вариант, попавших в интервал	n_1	n_2	…	n_k

Полигоны частот. Выборочная функция распределения и гистограммы.

Для наглядного представления о поведении исследуемой случайной величины в выборке можно строить различные графики. Один из них – **полигон частот**: ломаная, отрезки которой соединяют точки с координатами (x_1, n_1), (x_2, n_2),…, (x_k, n_k), где x_i откладываются на оси абсцисс, а n_i – на оси ординат. Если на оси ординат откладывать не абсолютные (n_i), а относительные (w_i) частоты, то получим **полигон относительных частот** (рис.1)**.**

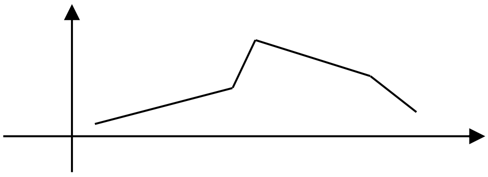


Рис. 1.

По аналогии с функцией распределения случайной величины можно задать некоторую функцию, относительную частоту события $X < x$.

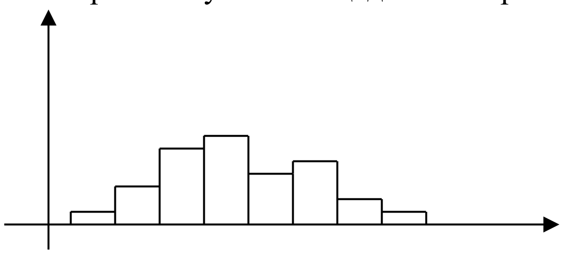
Определение. **Выборочной (эмпирической) функцией распределения** называют функцию $F^*(x)$, определяющую для каждого значения x относительную частоту события $X < x$. Таким образом,

$$F^*(x) = \frac{n_x}{n},$$

где n_x – число вариант, меньших x, n – объем выборки.

Замечание. В отличие от эмпирической функции распределения, найденной опытным путем, функцию распределения $F(x)$ генеральной совокупности называют *теоретической функцией распределения*. $F(x)$ определяет вероятность события $X < x$, а $F^*(x)$ – его относительную частоту. При достаточно больших n, как следует из теоремы Бернулли, $F^*(x)$ стремится по вероятности к $F(x)$.

Для непрерывного признака графической иллюстрацией служат **гистограммы**, то есть ступенчатые фигуры, состоящие из прямоугольников, основаниями которых служат частичные интервалы длиной h, а высотами – отрезки длиной n_i /h (гистограмма частот) или w_i /h (гистограмма относительных частот). В первом случае площадь гистограммы равна объему



выборки, во втором – единице (рис.2).

Одна из задач математической статистики: по имеющейся выборке оценить значения числовых характеристик исследуемой случайной величины или признака.

Определение. **Выборочным средним** называется среднее арифметическое значений случайной величины, принимаемых в выборке:

$$\bar{x}_B = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{x_1 n_1 + x_2 n_2 + \ldots + x_k n_k}{n} = \frac{\sum_{i=1}^{k} x_i n_i}{n},$$

где x_i – варианты, n_i - частоты.

Замечание. Выборочное среднее служит для оценки математического ожидания исследуемой случайной величины. В дальнейшем будет рассмотрен вопрос, насколько точной является такая оценка.

Определение. **Выборочной дисперсией** называется -

$$D_B = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_B)^2}{n} = \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x}_B)^2}{n},$$

а **выборочным средним квадратическим отклонением** –

$$\sigma_B = \sqrt{D_B}.$$

Так же, как в теории случайных величин, можно доказать, что справедлива следующая формула для вычисления выборочной дисперсии:

$$D = \overline{x^2} - (\bar{x})^2.$$

Другими характеристиками вариационного ряда являются:

- **мода** M_0 – варианта, имеющая наибольшую частоту ;

- **медиана** m_e - варианта, которая делит вариационный ряд на две части, равные по числу вариант. Если число вариант нечетно ($n = 2k + 1$), то $m_e = x_{k+1}$, а при четном $n = 2k$ $m_e = \frac{x_k + x_{k+1}}{2}$

.

Одна из задач математической статистики: по имеющейся выборке оценить значения числовых характеристик исследуемой случайной величины или признака.

Получив статистические оценки параметров распределения (выборочное среднее, выборочную дисперсию и т.д.), нужно убедиться, что они в достаточной степени служат приближением соответствующих характеристик генеральной совокупности. Требования к оценкам, которые должны при этом выполняться, служат требования несмещенности, эффективности и состоятельности.

При выборке малого объема точечная оценка может значительно отличаться от оцениваемого параметра, что приводит к грубым ошибкам. Поэтому в таком случае лучше пользоваться *интервальными оценками*, то есть указывать интервал, в который с заданной вероятностью попадает истинное значение оцениваемого параметра. Разумеется, чем меньше длина этого интервала, тем точнее оценка параметра.

Определение **Статистической гипотезой** называют гипотезу о виде неизвестного распределения генеральной совокупности или о параметрах известных распределений.

Определение. **Нулевой (основной)** называют выдвинутую гипотезу H_0. **Конкурирующей (альтернативной)** называют гипотезу H_1, которая противоречит нулевой.

Определение. **Простой** называют гипотезу, содержащую только одно предположение, **сложной** – гипотезу, состоящую из конечного или бесконечного числа простых гипотез.

В результате проверки правильности выдвинутой нулевой гипотезы (такая проверка называется **статистической**, так как производится с применением методов математической статистики) возможны ошибки двух видов: **ошибка первого рода**, состоящая в том, что будет отвергнута правильная нулевая гипотеза, и **ошибка второго рода**, заключающаяся в том, что будет принята неверная гипотеза.

Процесс проверки гипотезы состоит из следующих этапов:

1) выбирается статистический критерий K;

2) вычисляется его наблюдаемое значение $K_{набл}$ по имеющейся выборке;

3) поскольку закон распределения K известен, определяется (по известному уровню значимости α) **критическое значение** $k_{кр}$, разделяющее критическую область и область принятия гипотезы (например, если $p(K > k_{кр}) = \alpha$, то справа от $k_{кр}$ располагается критическая область, а слева – область принятия гипотезы);

4) если вычисленное значение $K_{набл}$ попадает в область принятия гипотезы, то нулевая гипотеза принимается, если в критическую область – нулевая гипотеза отвергается.

Типовая задача 1

Приведены результаты тестирования студентов по математике (ответы на 50 вопросов программы). Требуется:

1. Построить интервальные статистические ряды распределения частот и относительных частот (частостей) наблюдаемых значений;

2. Найти размах вариации и разбить его на 9 интервалов;

3. Построить гистограмму и полигон относительных частот, кумуляту. Указать, графикам каких функции в теории вероятностей они соответствуют;

4. Найти эмпирическую функцию распределения и построить ее график;

5. Вычислить числовые характеристики ряда распределения: выборочную среднюю, выборочные моду M_0^* и медиану M_e^*, выборочную дисперсию s^2 , выборочное среднее квадратичное отклонение s и выборочный коэффициент вариации V_s^*. Вычислить выборочные начальные и центральные моменты до четвертого порядка включительно, а также выборочные коэффициент асимметрии A_c^* и эксцесса $E_к^*$;

6. Рассчитать теоретическую нормальную кривую распределения и построить ее на эмпирическом графике;

7. Приняв в качестве нулевой гипотезы H_o (генеральная совокупность, из которой извлечена выборка, имеет нормальное распределение), проверить гипотезу, пользуясь критерием согласия Пирсона (χ^2) при уровне значимости $\alpha = 0{,}025$;

8. Найти доверительные интервалы для математического ожидания и среднего квадратичного отклонения.

Исходные выборочные данные

43	**49**	**25**	**22**
28	**36**	**36**	**28**
45	**21**	**48**	**49**
29	**25**	**31**	**23**
31	**40**	**35**	**32**
18	**26**	**43**	**33**
36	**25**	**38**	**27**
39	**33**	**26**	**43**
32	**34**	**35**	**35**
44	**21**	**31**	**37**

Решение:

1) Минимальное значение признака x_{min} = 18 вопросов., максимальное - x_{max} = 49 вопросов. Для определения границ интервалов находим шаг интервала: h = $\frac{x_{max}-x_{min}}{9}$. Шаг интервала округляем

h = $\frac{49-18}{9} = \frac{31}{9} \approx 4$.

Принимаем, что интервалы включают правую границу.

2) Для составления интервального распределения составим таблицу. В первой строке расположим в порядке возрастания интервалы, длина каждого из которых h=4. Во второй строке запишем количество значений признака в выборке, попавших в этот интервал (т.е. сумму частот вариант, попавших в соответствующий интервал). Интервальный статистический ряд таков:

(x_i, x_{i+1})	16–20	20–24	24–28	28–32	32–36	36-40	40-44	44-48	48-52
n_i	1	4	6	6	8	6	4	2	3

Объем выборки n=1 +4+6+6+8+6+ 4+2 +3=40.

Распределение относительных частот .

(x_i, x_{i+1})	16–20	20–24	24–28	28–32	32–36	36-40	40-44	44-48	48-52
n_i/n	0,025	0,1	0,15	0,15	0,2	0,15	0,1	0,05	0,075

Для построения гистограммы частот на оси абсцисс откладываем частичные интервалы; на каждом из них строим прямоугольники высотой $\frac{n_i}{n}$

Дискретный ряд распределения

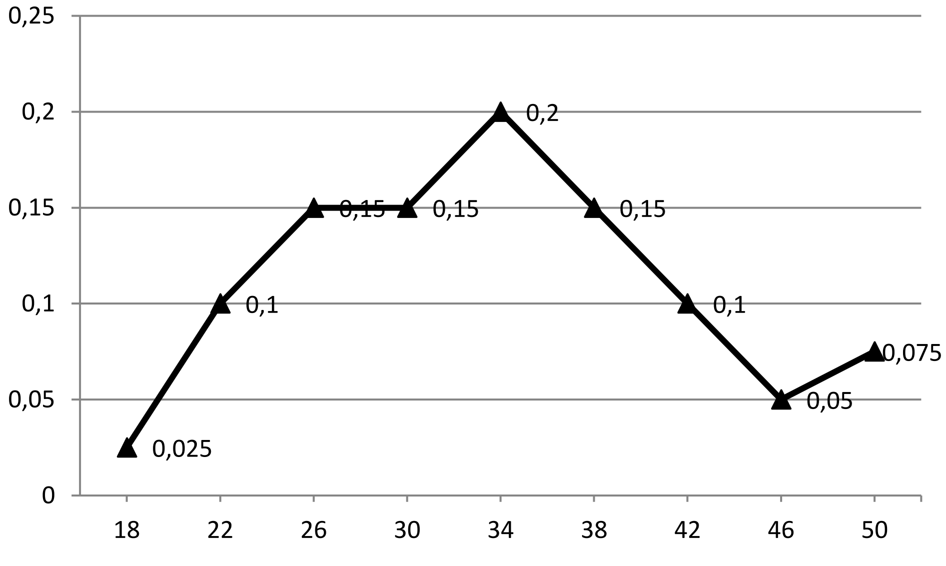
x_i	18	22	26	30	34	38	42	46	50
n_i/n	0,025	0,1	0,15	0,15	0,2	0,15	0,1	0,05	0,075

Для построения полигона частот по оси абсцисс откладываем середины интервалов, по оси ординат относительные частоты

Накопленные частоты

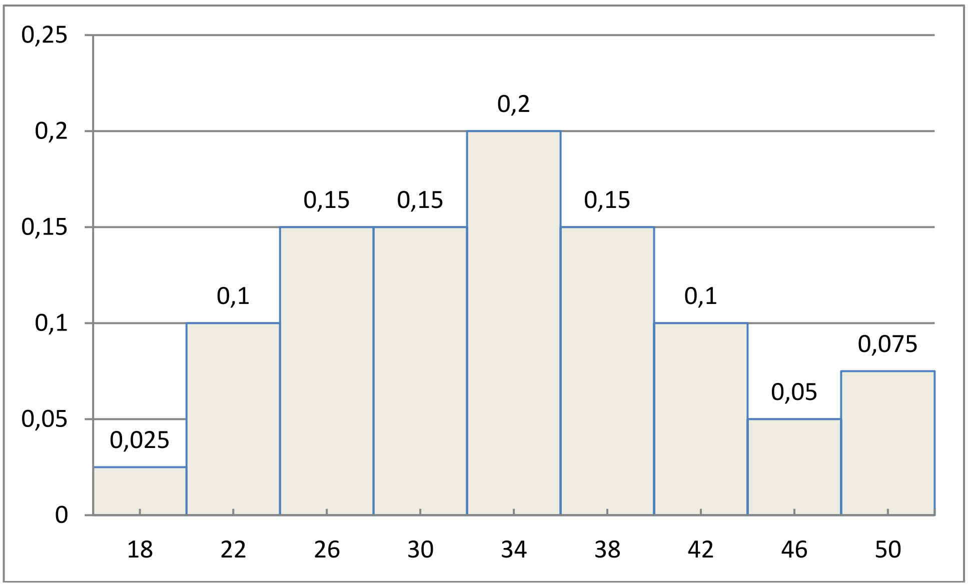
x_i	18	22	26	30	34	38	42	46	50
n_i	1	5	11	17	25	31	35	37	40
n_i/n	0,025	0,125	0,275	0,425	0,625	0,775	0,875	0,925	1

3) **Полигон** относительных частот соответствует графику плотности распределения, кумулята соответствует функции распределения

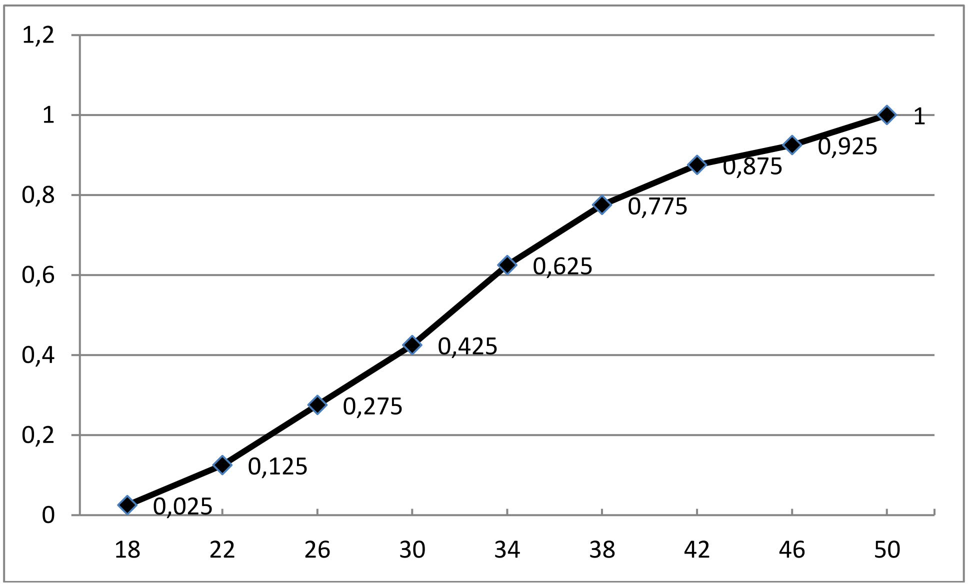


Полигон относительных частот

4) **Гистограмма и кумулята** относительных частот соответствует графику плотности распределения, кумулята соответствует функции распределения



Гистограмма относительных частот



Кумулята

5) **Эмпирической функцией** распределения (функцией распределения выборки) называют функцию $F^*(\text{x})$, определяющую для каждого значения x относительную частоту события X<x

$F^*(\text{x}) = \frac{\text{n}_\text{x}}{\text{n}}$, число вариант ,меньших x, n - объем выборки

Если x$\leq$ 18 то $\text{F}(x)$ =0;
Если 18 <x$\leq$ 22 то $\text{F}(x)$ =0, 0,025;
Если 22 <x$\leq$ 26 то $\text{F}(x)$ =0, 125;
Если 26 <x$\leq$ 30 то $\text{F}(x)$ =0, 275;
Если 30 <x$\leq$ 34 то $\text{F}(x)$ =0,425;
Если 34 <x$\leq$ 38 то $\text{F}(x)$ =0,625;

Если 38 <x≤ 42 то F(x) =0,775;
Если 42 <x≤ 46 то F(x) =0,875;
Если 46 <x≤ 50 то F(x) =0,925;
Если x> 50 то F(x) =1;

Строим график функции

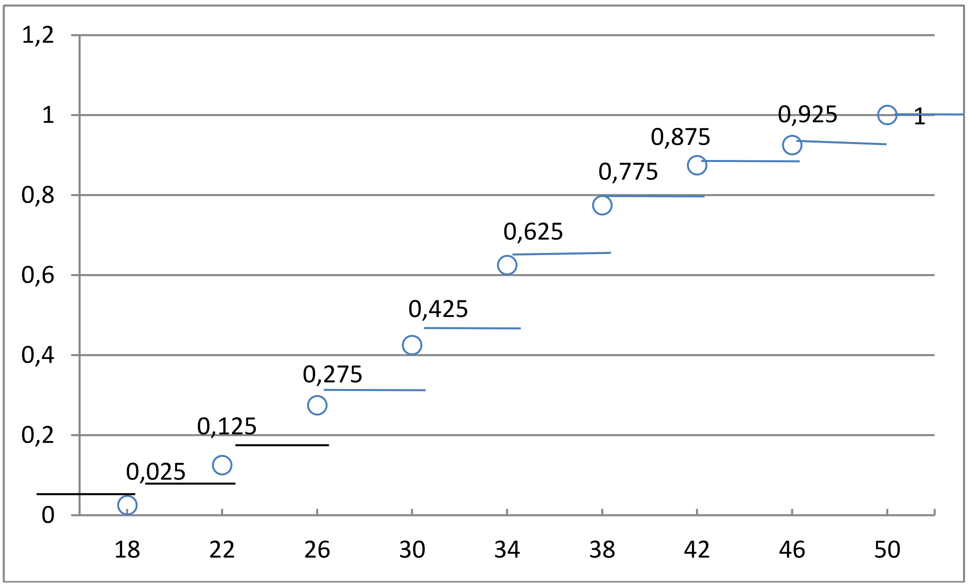


График эмпирической функции распределения

5) Мода $x_{мо}$ - значение признака с наибольшей частотой;
Медиана $x_{ме}$ значение признака, расположенного в середине ряда распределения. Мода и медиана являются структурными (распределительными) средними.
Для определения моды сначала находят интервал с наибольшей частотой $n_{мо}$= 8. В этом интервале число правильных ответов 32-36. Точное значение моды $x_{мо}$ находят путем интерполяции по формуле

$$x_{мо} = x_o + h\frac{n_{мо} - n_{мо-1}}{2n_{мо} - n_{мо-1} - n_{мо+1}}$$

Где h - шаг интервала, $n_{мо-1}$ - частота предмодального интервала, $n_{мо+1}$ - частота постмодального интервала.

$$x_{мо} = x_o + h\frac{n_{мо} - n_{мо-1}}{2n_{мо} - n_{мо-1} - n_{мо+1}} = 32 + \ 4\frac{8 - \ 6}{16 - 6 - \ 6}$$

$= 32 + \frac{2}{4} \cdot 4 = 34$

Значение медианы $x_{ме}$также определяем   путем интерполяции по формуле

$x_{ме} = x_o + h\frac{\frac{n}{2} - S_{ме-1}}{n_{ме}}$.

$S_{ме-1}$ - накопленные частоты интервалов, предшествующих меданному.
$n_{ме}$ - локальная частота интервала, в котором находятся   единицы совокупности, делящие ряд пополам, $x_o -$ начало медианного интервала.
$\frac{n}{2} = \frac{40}{2} = 20$, следовательно   медианным является интервал с накопленной частотой 20, его частота составляет $n_{ме}$ =10, $S_{ме-1}$=14.

$x_{\text{ме}} = x_o + h\frac{\frac{n}{2} - S_{\text{ме}-1}}{n_{\text{ме}}} = 32 + 4 \cdot \frac{20-17}{8} = 32 + 0{,}375 \cdot 4 = 33{,}5$

x_i	18	22	26	30	34	38	42	46	50
n_i	1	5	11	17	25	31	35	37	40

Найдем методом произведений выборочные: среднюю, дисперсию, среднее квадратическое отклонение, начальные и центральные моменты до четвертого порядка включительно.

Составляем таблицу:

Таблица 1.

x_i	n_i	u_i	n_iu_i	$n_iu_i^2$	$n_iu_i^3$	$n_iu_i^4$	$n_i(u_i+1)^4$
18	1	-4	-4	16	-64	256	81
22	4	-3	-12	36	-108	324	64
26	6	-2	-12	24	-48	96	6
30	6	-1	-6	6	-6	6	0
34	8	0	0	0	0	0	8
38	6	1	6	6	6	6	96
42	4	2	8	16	32	64	324
46	2	3	6	18	54	162	512
50	3	4	12	48	192	768	1875
$\sum$	n=40		$\sum n_iu_i$ = -2	$\sum n_iu_i^2$=170	$\sum n_iu_i^3 = 58$	$\sum n_iu_i^4 = 1682$	$n_i(u_i+1)^4 = 2966$

В качестве ложного нуля принимаем С= 34– варианта с наибольшей частотой 10 и находящаяся в середине вариационного ряда. Шаг выборки h=4. Тогда условные варианты определяются по формуле

$u_i = \frac{x_i - C}{h} = \frac{x_i - 34}{4}$.

Подсчитываем все варианты u_i и заполняем все столбцы.

Последний столбец служит для контроля вычислений по тождеству:

$\sum n_i(u_i+1)^4 = \sum n_iu_i^4 + 4\sum n_iu_i^3 + 6\sum n_iu_i^2 + 4\sum n_iu_i$ +n= 1682 +4· 58+6·170 +4·(-2) +40=2966.

Вычисления произведены верно. Найдем условные начальные моменты.

$M_1^* = \frac{\sum n_iu_i}{n} = \frac{-2}{40} = -0{,}05.$ $M_2^* = \frac{\sum n_iu_i^2}{n} = \frac{170}{40}$ = 4, 25.

$M_{\text{к}}^* = \frac{\sum n_iu_i^{\text{к}}}{n}$ - условные начальные моменты к- го порядка

Вычисляем выборочную среднюю:

$\bar{x}_{\text{в}} = M_1^* \cdot h + C = \frac{-2}{40} \cdot 4 + 34 = 33{,}8.$

Находим выборочную дисперсию:

$D_{\text{в}} = [M_2^* - (M_1^*)^2] \cdot h^2 = [4{,}25 - (-0{,}05)^2] \cdot 4^2$ = 4,2475· 16= 67,96.

Определяем выборочное среднее квадратическое отклонение:

$\sigma_B = \sqrt{D_B} = \sqrt{67{,}96}$ =8,2437.

$m_3 = \frac{\sum (x_i - \bar{x})^3 n_i}{\sum n_i}$. $m_4 = \frac{\sum (x_i - \bar{x})^4 n_i}{\sum n_i}$ центральные эмпирические моменты третьего и четвертого порядков.

Эти моменты в случае равноотстоящих вариант с шагом h вычисляются по формулам:

$$m_3 = [M_3^* - 3M_1^*M_2^* + 2(M_1^*)^3]h^3$$
$$m_4 = [M_4^* - 4M_1^*M_3^* + 6(M_1^*)^2M_2^* - 3(M_1^*)^4]h^4$$

Ассиметрия и эксцесс определяются равенствами: $a_s = m_3/\sigma_B^3$, $e_k = \frac{m_4}{\sigma_B^4} - 3$

$M_3^* = \frac{\sum n_i u_i^3}{40} = \frac{58}{40} = 1{,}45$,

$M_4^* = \frac{\sum n_i u_i^4}{40} = \frac{1682}{40} = 42{,}05$,

$M_1^* = \frac{\sum n_i u_i}{n} = \frac{-2}{40} = -0{,}05$. $M_2^* = \frac{\sum n_i u_i^2}{n} = \frac{170}{40} = 4{,}25$.

$m_3 = [M_3^* - 3M_1^*M_2^* + 2(M_1^*)^3]h^3 = [1{,}45 - 3\cdot(-0{,}05)\cdot 4{,}25 + 2(-0{,}05)^3]4^3 = 2{,}08725\cdot 64 = 133{,}584$

$m_4 = [M_4^* - 4M_1^*M_3^* + 6(M_1^*)^2M_2^* - 3(M_1^*)^4]h^4 = [42{,}05 - 4\cdot(-0{,}05)\cdot 1{,}45 + 6\cdot(-0{,}05)^2\cdot 4{,}25 - 3(-0{,}05)^4]4^4 = 42{,}4\cdot 256 = 10855{,}3552$

Коэффициент вариации находим по формуле:

$\upsilon = \frac{\sigma}{\bar{x}}\cdot 100\%$. $\upsilon = \frac{8{,}2437}{33{,}8}\cdot 100\% = 24{,}39\ \%$.

$$a_s = \frac{m_3}{\sigma_B^3} = \frac{133{,}584}{8{,}2437^3} = 0{,}2384$$

$$e_k = \frac{m_4}{\sigma_B^4} - 3 = \frac{10855{,}3552}{8{,}2437^4} - 3 = 2{,}35 - 3 = -0{,}65$$

6. **Строим нормальную кривую.**

Для облегчения вычислений все расчеты сводим в таблицу 2

Таблица 2.

x_i	n_i	$x_i - \bar{x}_в = x_i - 33{,}8$	$u_i = \frac{x_i - \bar{x}_в}{\sigma_B} = \frac{x_i - 33{,}8}{8{,}2437}$	$\varphi(u_i)$	$n_i' = 19{,}4\cdot\varphi(u_i)$
18	1	-15,8	-1,91654	0,0644	1,24936≅ 1
22	4	-11,8	-1,43134	0,1435	2,7839≅ 3
26	6	-7,8	-0,94614	0,2565	4,9761≅ 5
30	6	-3,8	-0,46094	0,3589	6,96266≅ 7
34	8	0,2	0,02426	0,3989	7,73866≅ 8
38	6	4,2	0,50946	0,3521	6,83074≅ 7
42	4	8,2	0,99466	0,2444	4,74136≅5
46	2	12,2	1,47986	0,1334	2,58796≅ 3
50	3	16,2	1,96506	0,058	1,1252≅ 1
	n=40				n=40

Заполняем первые три столбца.

В четвертом столбце записываем условные варианты по формуле, указанной в «шапке» таблицы. В пятом столбце находим значения функции $\varphi(u_i) = \frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{u_i^2}{2}}$.

Функция $\varphi(u_i)$ четная ,т.е. $\varphi(u_i) = \varphi(-u_i)$.Значения функции $\varphi(u_i)$ в зависимости от аргумента u_i(берутся положительные u_i, т.к. $\varphi(u_i)$ четная) находим из таблицы.

Теоретические частоты теоретической кривой находим по формуле

n_i'=n$\cdot p_i$,

где p_i - вероятность попадания X в i-частичный интервал с концами $x_i - \frac{h}{2}$ и $x_i + \frac{h}{2}$.

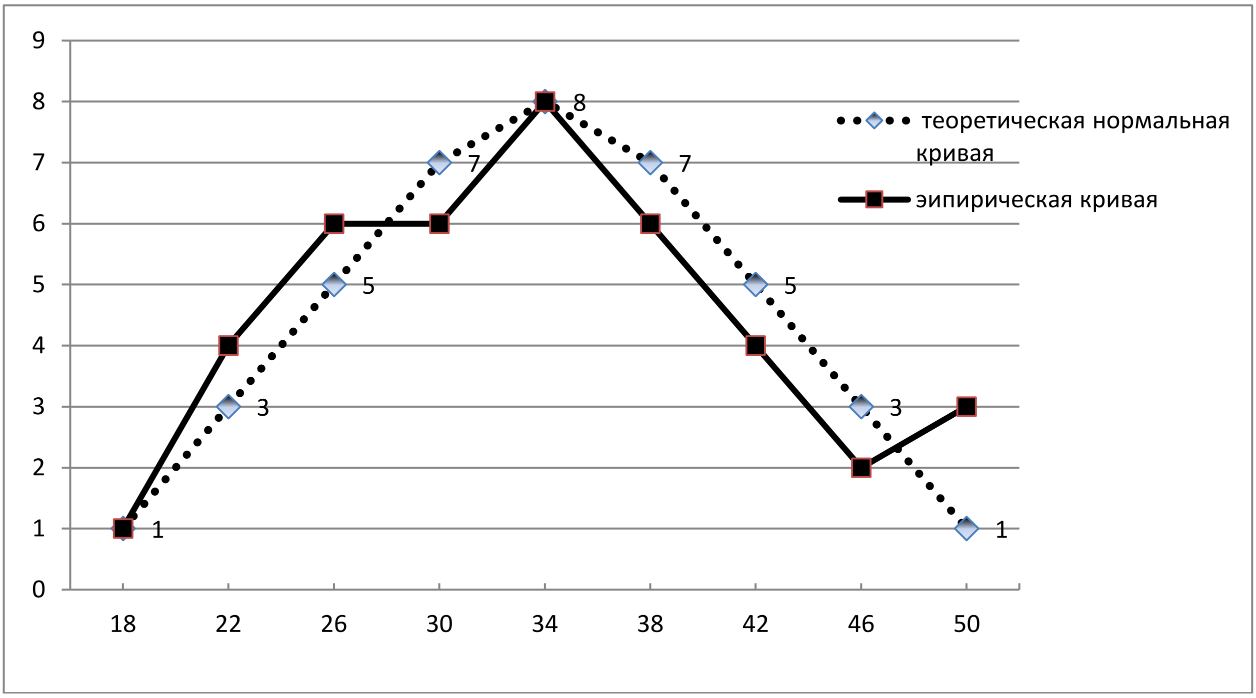
Приближенно вероятности могут быть найдены по формуле $p_i = \frac{h}{\sigma_B} \varphi(u_i)$.

Тогда теоретические частоты равны равны

n_i'=n$\cdot \frac{h}{\sigma_B} \varphi(u_i)$=$\frac{40\cdot 4}{8{,}2437} \varphi(u_i)$ = 19, 4$\cdot \varphi(u_i)$.

Заполняем последний столбец. В последнем столбце частоты n_i' округляются до целого числа и $\sum n_i'$= $\sum n_i$ =40.

В системе координат $(x_i; y_i = n_i')$ строим нормальную (теоретическую кривую)кривую по выравнивающим частотам n_i' и полигон наблюдаемых частот n_i. Полигон наблюдаемых частот построен в системе координат $(x_i; y_i = n_i)$.



7. **Проверяем гипотезу** о нормальности X при уровне значимости α=0,05.
В качестве статистики θ выбирают СВ χ^2:

$$\chi^2=\sum_{i=1}^{s} \frac{(n_i - n_i')^2}{n_i'}.$$

Она подчиняется распределению χ^2 с числом степеней свободы $\nu = s - r - 1$, где s - число различных значений x_i ; r - число параметров, откоторых зависит распределение. Для нормального закона таких параметров два: a=$\bar{x}_в = M(x)$ и $\sigma = s = \sqrt{D_B \cdot \frac{n}{n-1}}$, т.е. $r = 2$, и $\nu = s - 3$. По данному уровню значимости α и числу степеней свободы ν в таблице распределения χ^2 находят критическое значение $\chi^2_{крит.}$ и находят критическую область: $\chi^2 < \chi^2_{крит.}$, $\omega_0 = \{\chi^2 : \chi^2 \geq \chi^2_{крит.}\}$. Затем вычисляем наблюдаемое значение χ^2, т.е. $\chi^2_{набл.}$ по формуле

$$\chi^2_{набл.} = \sum_{i=1}^{s} \frac{(n_i - n_i')^2}{n_i'}.$$

Если окажется, что $\chi^2_{набл.} < \chi^2_{крит.}$, то нулевую гипотезу H_0 о том ,что X имеет нормальное распределение, принимают. В этом случае опытные данные хорошо согласуются с гипотезой о нормальном распределении генеральной совокупности.

Вычислим $\chi^2_{\text{набл.}}$, для чего составим расчетную таблицу 3.

Таблица 3

1	2	3	4	5	6	7
n_i	n_i'	$n_i - n_i'$	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$	n_i^2	$\frac{n_i^2}{n_i'}$
1	1	0	0	0	1	1
4	3	1	1	0,333333	16	5,333333
6	5	1	1	0,2	36	7,2
6	7	-1	1	0,142857	36	5,142857
8	8	0	0	0	64	8
6	7	-1	1	0,142857	36	5,142857
4	5	-1	1	0,2	16	3,2
2	3	-1	1	0,333333	4	1,333333
3	1	2	4	4	9	9
	n=40			$\chi^2_{\text{набл.}}$= 5,352381		45,35238

Суммируя числа пятого столбца, получаем $\chi^2_{\text{набл.}}$= 5,352381

Суммируя числа последнего столбца, получаем 45,35238

Контроль: $\chi^2_{\text{набл.}}$=5,352381

$\sum \frac{n_i^2}{n_i'}$ - $\sum n_i$= 45,35238-40=5,352381

Совпадение результатов подтверждает правильность вычислений.

Найдем число степеней свободы, учитывая, что число групп выборки (число различных вариантов) 7, $\nu = s - 3$=9-3=6.

По таблице критических точек распределения χ^2, по уровню значимости $\alpha = 0.025$ и числу степеней свободы ν =6 находим $\chi^2_{\text{крит.}} = 14{,}4$.

Так как $\chi^2_{\text{набл.}} < \chi^2_{\text{крит.}}$,то нет оснований отвергнуть нулевую гипотезу. Расхождение эмпирических и теоретических частот незначимое. Следовательно, данные наблюдений согласуются с гипотезой о нормальном распределении генеральной совокупности.

8.**Найдем доверительный интервал** для оценки неизвестного математического ожидания, полагая, что X имеет нормальное распределение, среднее квадратическое отклонение $\sigma = \sigma_{\text{X}} = \sigma_B$ = 8,2437и доверительную вероятность $\gamma = 0{,}95$.

Известен объем выборки: n=40, выборочная средняя $\bar{x}_{\text{в}} = 33{,}8$.

Из соотношения $2\Phi(t) = \gamma$ получим $\Phi(t)$ =0,475. По таблице находим параметр t=1,96.

Найдем точность оценки

$$\delta = \frac{t\sigma}{\sqrt{n}} = \frac{1{,}96 \cdot 8{,}2437}{\sqrt{40}} = 2{,}55$$

Доверительный интервал таков:

$\bar{x}_{\text{в}} - \delta < M(X) < \bar{x}_{\text{в}} + \delta$ или $33{,}8 - 2{,}55 < M(X) < 33{,}8 + 2{,}55$ $\Leftrightarrow$ или $31{,}25 < M(X) < 36{,}35$.

Надежность $\gamma = 0{,}95$ указывает, что если произведено достаточно большое число выборок, то 95 % из них определяет такие доверительные интервалы, в которых параметр действительно заключен.

Интервальная оценка для среднего квадратического отклонения:

$s(1-q) < \sigma < s(1+q)$. q находим по таблице по заданным n и $\gamma = 0{,}95$.

$q(40;0{,}95) = 0{,}24$

$$8{,}2437 \cdot 0{,}76 < \sigma < 8{,}2437 \cdot 1{,}24;$$
$$6{,}26 < \sigma < 10{,}222$$

Тема 2

Статистическое исследование зависимостей (корреляционно- регрессионный анализ)

Рассмотрим выборку двумерной случайной величины (X, Y). Примем в качестве оценок условных математических ожиданий компонент их условные средние значения, а именно: **условным средним** $\bar{y}_x$ назовем среднее арифметическое наблюдавшихся значений Y, соответствующих $X = x$. Аналогично **условное среднее** $\bar{x}_y$ **-** среднее арифметическое наблюдавшихся значений X, соответствующих $Y = y$. Уравнения регрессии Y на X и X на Y имеют вид :

$$\bar{y}_x = f^*(x) \text{ -}$$

- **выборочное уравнение регрессии Y на X,**

$$\bar{x}_y = \varphi^*(y) \text{ -}$$

- **выборочное уравнение регрессии X на Y.**

Соответственно функции $f^*(x)$ и $\varphi^*(y)$ называются **выборочной регрессией Y на X и X на Y** , а их графики – **выборочными линиями регрессии**. Выясним, как определять параметры выборочных уравнений регрессии, если сам вид этих уравнений известен.

Пусть изучается двумерная случайная величина (X, Y), и получена выборка из n пар чисел $(x_1, y_1), (x_2, y_2),\ldots, (x_n, y_n)$. Будем искать параметры прямой линии регрессии Y на X вида

$$Y = \rho_{yx}x + b,$$

подбирая параметры ρ_{yx} и b так, чтобы точки на плоскости с координатами $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ лежали как можно ближе к прямой. Используем для этого метод наименьших квадратов и найдем минимум функции

$$F(\rho, b) = \sum_{i=1}^{n}(Y_i - y_i)^2 = \sum_{i=1}^{n}(\rho x_i + b - y_i)^2.$$

Приравняем нулю соответствующие частные производные:

$$\frac{\partial F}{\partial \rho} = 2\sum_{i=1}^{n}(\rho x_i + b - y_i)x_i = 0$$
$$\frac{\partial F}{\partial b} = 2\sum_{i=1}^{n}(\rho x_i + b - y_i) = 0.$$

В результате получим систему двух линейных уравнений относительно ρ и b:

$$\begin{cases} \left(\sum x^2\right)\rho + \left(\sum x\right)b = \sum xy \\ \left(\sum x\right)\rho + nb = \sum y \end{cases}$$

Ее решение позволяет найти искомые параметры в виде:

$$\rho_{yx} = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - \left(\sum x\right)^2}; \quad b = \frac{\sum x^2 \sum y - \sum x \sum xy}{n\sum x^2 - \left(\sum x\right)^2}.$$

При этом предполагалось, что все значения X и Y наблюдались по одному разу.

Теперь рассмотрим случай, когда имеется достаточно большая выборка (не менее 50 значений), и данные сгруппированы в виде *корреляционной таблицы*:

	X				
	x_1	x_2		x_k	n_y
1	n_{11}	n_{21}		n_{k1}	$n_{11}+n_{21}+\ldots+n_{k1}$
2	n_{12} … n_{1m}	n_{22} … n_{2m}		n_{k2} … n_{km}	$n_{12}+n_{22}+\ldots+n_{k2}$ $n_{1m}+n_{2m}+\ldots+n_{km}$
m					
x	$n_{11}+n_{12}+\ldots+n_{1m}$	$n_{21}+n_{22}+\ldots+n_{2m}$		$n_{k1}+n_{k2}+\ldots+n_{km}$	$n=\sum n_x=\sum n_y$

Здесь n_{ij} – число появлений в выборке пары чисел (x_i, y_j).

Поскольку $\bar{x}=\frac{\sum x}{n}, \bar{y}=\frac{\sum y}{n}, \overline{x^2}=\frac{\sum x^2}{n}$, заменим в системе (10.3) $\sum x = n\bar{x}$, $\sum y = n\bar{y}$, $\sum x^2 = n\overline{x^2}$, $\sum xy = \sum n_{xy}xy$, где n_{xy} – число появлений пары чисел (x, y). Тогда система примет вид:

$$\begin{cases} (n\overline{x^2})\rho_{yx} + (n\bar{x})b = \sum n_{xy}xy \\ (\bar{x})\rho_{yx} + b = \bar{y} \end{cases}$$

Можно решить эту систему и найти параметры ρ_{yx} и b, определяющие выборочное уравнение прямой линии регрессии:

$$\bar{y}_x = \rho_{yx}\bar{x} + b.$$

Но чаще уравнение регрессии записывают в ином виде, вводя **выборочный коэффициент корреляции**. Выразим b из второго уравнения системы:

$$b = \bar{y} - \rho_{yx}\bar{x}.$$

Подставим это выражение в уравнение регрессии:

$$\bar{y}_x - \bar{y} = \rho_{yx}(x - \bar{x}).$$

$$\rho_{yx} = \frac{\sum n_{xy}xy - n\bar{x}\bar{y}}{n(\overline{x^2} - (\bar{x})^2)} = \frac{\sum n_{xy}xy - n\bar{x}\bar{y}}{n\tilde{\sigma}_x^2},$$

где $\tilde{\sigma}_x^2 = \overline{x^2} - (\bar{x})^2$.

Введем понятие **выборочного коэффициента корреляции**

$$r_B = \frac{\sum n_{xy}xy - n\bar{x}\bar{y}}{n\tilde{\sigma}_x\tilde{\sigma}_y}$$

и умножим равенство на $\frac{\tilde{\sigma}_x}{\tilde{\sigma}_y}$: $\rho_{yx}\frac{\tilde{\sigma}_x}{\tilde{\sigma}_y} = r_B$, откуда $\rho_{yx} = r_B\frac{\tilde{\sigma}_y}{\tilde{\sigma}_x}$.

Используя это соотношение, получим выборочное уравнение прямой линии регрессии Y на X вида

$$\bar{y}_x - \bar{y} = r_B\frac{\tilde{\sigma}_y}{\tilde{\sigma}_x}(x - \bar{x}).$$

Типовая задача 2

Для 10 петушков леггорнов 15-дневного возраста были получены следующие данные о весе их тела $X(г)$ и весе гребня $Y(мг)$:

x_i	83	72	69	90	90	95	95	91	75	70
y_i	56	42	18	84	56	107	90	68	31	48

Требуется:

1) найти коэффициент корреляции и сделать вывод о тесноте и направлении линейной корреляционной связи между признаками;

2) составить уравнение прямой регрессии;

3) нанести на чертеже исходные данные и построить полученную прямую регрессии.

Решение.

1) В малых выборках коэффициент корреляции рассчитывается по формуле:

$$r = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2} \cdot \sqrt{\sum (y_i - \overline{y})^2}}.$$

Промежуточные вычисления удобно проводить в таблице 1, располагая $X(г)$ - вес тела петушка в порядке возрастания.

Таблица 1

№ наблюдения	x_i	y_i	$x_i - \overline{x}$	$(x_i - \overline{x})^2$	$y_i - \overline{y}$	$(y_i - \overline{y})^2$	$(x_i - \overline{x})(y_i - \overline{y})$
1	69	18	-14	196	-42	1764	588
2	70	48	-13	169	-12	144	156
3	72	42	-11	121	-18	324	198
4	75	31	-8	64	-29	841	232
5	83	56	0	0	-4	16	0
6	90	56	7	49	-4	16	-28
7	90	84	7	49	24	576	168
8	91	68	8	64	8	64	64
9	95	90	12	144	30	900	360
10	95	107	12	144	47	2209	564
$\sum$	830	600	0	990	0	6854	2302

Вычисляем средние:

$$\overline{X} = \frac{\sum x_i}{n}; \qquad \overline{X} = \frac{830}{10} = 83, \qquad \overline{Y} = \frac{\sum y_i}{n}; \ \overline{Y} = \frac{600}{10} = 60$$

Заполняем столбцы таблицы. Суммируя элементы в столбцах, находим:

$$\sum_{i=1}^{10} (x_i - \overline{x})^2 = 990, \ \sum_{i=1}^{10} (y_i - \overline{y})^2 = 6854, \ \sum_{i=1}^{10} (x_i - \overline{x})(y_i - \overline{y}) = 2302.$$

Подставляя вычисленные значения в формулу для r, получаем

$$r = \frac{2302}{\sqrt{990} \cdot \sqrt{6854}} = \frac{2302}{2604{,}9} = 0{,}8837 \approx 0{,}88.$$

Вывод: между весом тела X и весом гребня Y у 15 – дневных петушков существует тесная положительная линейная корреляционная связь.

2) Уравнение прямой регрессии имеет вид:

$$y-\bar{y}=b_{y/x}\left(x-\bar{x}\right),$$

где $b_{y/x}$ - коэффициент регрессии, определяется по формуле:

$$b_{y/x}=\frac{\sum\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sum\left(x_i-\bar{x}\right)^2}.$$

Беря данные из таблицы, получим:

$$b_{y/x}=\frac{2302}{990}=2{,}32525\approx 2{,}325\cdot$$

Подставляя теперь в уравнение прямой регрессии $\overline{X}=83$, $\overline{Y}=60$, $b_{y/x}\approx 2{,}32$ будем иметь

$$y-60=2{,}32(x-83).$$

Последнее уравнение преобразуем к виду

$$y=2{,}32x-2{,}32\cdot 83+60;\; y=2{,}32x-132{,}56.$$

3)Нанесем исходные данные на координатную плоскость и построим найденную прямую регрессии (рис. 1).

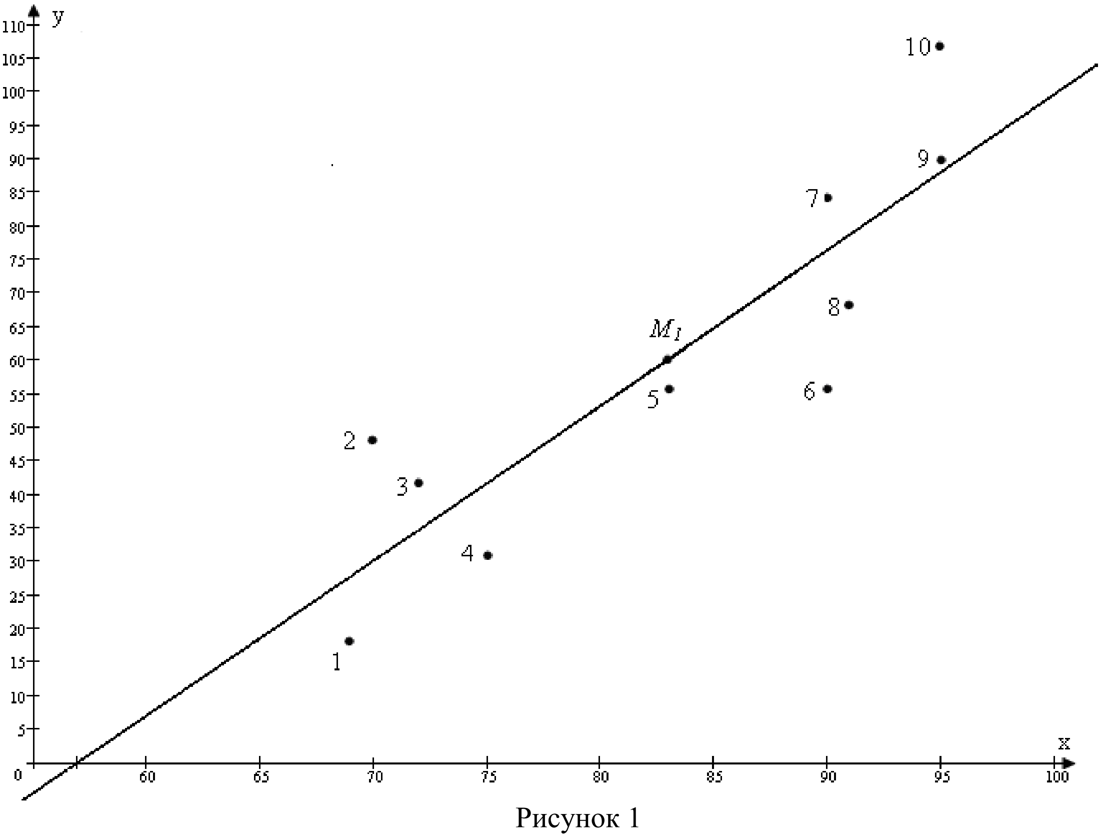


Рисунок 1

Для того чтобы провести прямую в системе координат, достаточно иметь две точки. Одна точка $M_1(83;60)$. Координаты второй точки M_2 определим, подставив в уравнение регрессии $y=0$ и вычислив

$$x = \frac{132,56}{2,32} \approx 57 \cdot$$

Полученная математическая модель (уравнение прямой регрессии) обладает прогнозирующими свойствами лишь при изменении X от 69 до 95. Так, например, можно с достаточной степенью достоверности считать, что при весе петушка 80 $г$ вес его гребня составит $y = 2,32 \cdot 80 - 132,56 \approx 53 мг$.

В задачах 121 - 140 требуется: 1) найти коэффициент корреляции и сделать вывод о тесноте и направлении линейной корреляционной связи между признаками; 2) составить уравнение прямой регрессии Y на X;

3) нанести на чертеже исходные данные и построить прямую регрессии.

3. ВАРИАНТЫ ЗАДАНИЙ

Задача 1

Приведены результаты тестирования студентов по математике (ответы на 50 вопросов программы).

Требуется:

1. Построить интервальные статистические ряды распределения частот и относительных частот (частостей) наблюдаемых значений;
2. Найти размах вариации и разбить его на 9 интервалов;
3. Построить гистограмму и полигон относительных частот, кумуляту. Указать, графикам каких функции в теории вероятностей они соответствуют;
4. Найти эмпирическую функцию распределения и построить ее график;
5. Вычислить числовые характеристики ряда распределения: выборочную среднюю, выборочные моду M_0^* и медиану M_e^*, выборочную дисперсию s^2, выборочное среднее квадратичное отклонение s и выборочный коэффициент вариации V_s^*. Вычислить выборочные начальные и центральные моменты до четвертого порядка включительно, а также выборочные коэффициент асимметрии $А_с^*$ и эксцесса $Е_к^*$;
6. Рассчитать теоретическую нормальную кривую распределения и построить ее на эмпирическом графике;
7. Приняв в качестве нулевой гипотезы H_o (генеральная совокупность, из которой извлечена выборка, имеет нормальное распределение), проверить гипотезу, пользуясь критерием согласия Пирсона (χ^2) при уровне значимости $\alpha = 0,025$;
8. Найти доверительные интервалы для математического ожидания и среднего квадратичного отклонения.

Исходные выборочные данные для типового расчета представлены в таблице. Каждый студент для выполнения данного задания использует n = 50 выборочных данных.

Из таблицы каждый студент выбирает строки с номерами от *к* до *к+9,* где k-номер по списку в групповом журнале:

Таблица данных

	1	**2**	**3**	**4**	**5**
1.	33	42	27	33	23
2.	26	31	43	45	44

3.	16	36	19	19	23
4.	43	49	25	22	23
5.	28	36	36	28	43
6.	45	21	48	49	44
7.	29	25	31	23	33
8.	31	40	35	32	37
9.	18	26	43	33	31
10.	36	25	38	27	35
11.	39	33	26	43	35
12.	32	34	35	35	34
13.	44	21	31	37	32
14.	26	31	43	45	44
15.	16	36	19	19	39
16.	43	49	25	22	23
17.	28	30	36	28	43
I8.	45	21	16	49	44
19.	29	25	31	23	33
20.	31	40	47	32	37
21.	18	24	43	33	31
22.	36	25	38	31	35
23.	33	33	29	18	35
24	37	34	35	35	34
25.	32	16	31	44	25
26	44	31	43	26	33
27.	39	47	19	16	34
28.	23	43	25	43	16
29.	43	33	36	22	31
30.	47	39	26	33	37
31.	36	25	33	33	32
32.	43	31	35	44	48
33.	19	18	25	23	35
34.	25	23	43	25	44
35.	31	43	49	25	32
36.	43	26	43	21	44
37.	36	16	38	32	26
38.	33	26	33	34	35
39.	37	23	43	25	25
40.	32	23	43	18	33

Задача 2

В таблицах 1,2,3,4 представлена выработка товарной продукции X на одного работающего и затраты на один рубль товарной продукции Y по предприятиям Саратовской области.

Требуется:

1. найти коэффициент корреляции и сделать вывод о тесноте и направлении линейной корреляционной связи между признаками;

2. составить уравнение прямой регрессии;

3. нанести на чертеже исходные данные и построить полученную прямую регрессии.

Исходные данные для заданий типового расчета выбирается с номерами задач, которые соответствуют номеру в списке группового журнала

Таблица 4

№ наблюдения	№ задачи									
	1		2		3		4		5	
1	97	35	93	36	104	31	95	36	102	32
2	104	31	101	31	98	35	90	37	95	37
3	103	32	95	34	100	32	103	32	97	35
4	98	34	97	35	102	31	104	31	98	34
5	101	30	102	30	99	32	89	37	94	37
6	102	33	94	35	97	33	97	35	90	38
7	100	31	96	36	95	36	101	34	100	30
8	99	34	100	31	101	32	96	34	101	31
9	96	35	95	36	103	30	99	33	93	36
10	98	32	92	37	98	35	102	32	96	35

Таблица 5

	№ задачи									
	6		7		8		9		10	
1	91	62	82	51	103	79	85	56	97	61
2	86	43	101	59	96	61	94	63	89	48
3	94	60	104	77	92	60	91	59	94	58
4	95	73	96	63	98	68	103	70	105	74
5	104	87	97	73	89	55	101	63	98	62
6	92	64	112	68	97	70	97	60	91	67
7	98	79	106	68	98	66	93	61	85	60
8	83	52	91	62	87	49	94	72	87	53
9	96	65	109	70	105	75	97	58	102	78
10	99	68	91	62	96	61	95	65	96	57

Таблица 6

№ наблюдения	№ задачи

	11		12		13		14		15	
1	96	34	92	36	103	30	94	35	101	30
2	104	31	100	31	97	35	89	37	95	37
3	103	32	95	34	100	32	103	32	97	35
4	98	34	97	35	102	31	104	31	98	34
5	101	30	102	30	99	32	89	37	93	37
6	102	33	94	35	97	33	97	35	90	38
7	100	31	96	36	95	36	101	34	100	30
8	99	34	101	31	101	32	96	34	101	31
9	96	35	95	36	103	30	99	33	93	36
10	98	32	92	37	98	35	102	32	96	35

Таблица 7

№ наблюдения	№ задачи									
	16		17		18		19		20	
1	97	35	93	36	104	31	95	36	102	32
2	104	31	101	31	98	35	90	37	95	37
3	103	32	95	34	100	32	103	32	97	35
4	98	34	97	35	102	31	104	31	98	34
5	101	30	102	30	99	32	89	37	94	37
6	102	33	94	35	97	33	97	35	90	38
7	100	31	96	36	95	36	101	34	100	30
8	99	34	100	31	101	32	96	34	101	31
9	96	35	95	36	103	30	99	33	93	36
10	98	32	92	37	98	35	102	32	96	35

4. КРИТЕРИЙ ОЦЕНКИ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

Качество выполнения внеаудиторной самостоятельной работы студентов оценивается посредством текущего контроля самостоятельной работы студентов. Текущий контроль СРС – это форма планомерного контроля качества и объема приобретаемых студентом компетенций в процессе изучения дисциплины, проводится на практических и семинарских занятиях и во время консультаций преподавателя.

Максимальное количество баллов «отлично» студент получает, если:

- обстоятельно с достаточной полнотой излагает соответствующую тему;
- раскрыл содержание практической части задания на 85%-100%;
- дает правильные формулировки, точные определения, понятия терминов;
- правильно отвечает на дополнительные вопросы преподавателя, имеющие целью выяснить степень понимания студентом данного материала.

Оценку «хорошо» студент получает, если:

- неполно, но правильно решено задание;
- если студент раскрыл содержание практической части задания на 70%-84%;
- при решении были допущены 1-2 несущественные ошибки, которые он исправляет после замечания преподавателя;
- дает правильные формулировки, точные определения, понятия терминов;
- правильно отвечает на дополнительные вопросы преподавателя, имеющие целью выяснить степень понимания студентом данного материала.

Оценку «удовлетворительно» студент получает, если:

- неполно, но правильно изложено задание;
- если студент раскрыл содержание практической части задания на 50%-69%;
- при решении было допущено 2 существенные ошибки;
- излагает выполнение задания недостаточно логично и последовательно;
- затрудняется при ответах на вопросы преподавателя.

Оценка «неудовлетворительно» студент получает, если:

- неполно изложено задание;
- если студент раскрыл содержание практической части задания менее, чем на 49%;
- при изложении были допущены существенные ошибки, т.е. если оно не удовлетворяет требованиям, установленным преподавателем к данному виду работы.

4. СПИСОК РЕКОМЕНДУЕМОЙ ЛИТЕРАТУРЫ

основная литература:

1. **Вентцель, Е. С.** Теория вероятностей и ее инженерные приложения : учебное пособие / Е. С. Вентцель, Л. А. Овчаров. - 4-е изд., стер. - М. : Высш. шк., 2007. - 491 с. - (Учебники для вузов. Математика). - ISBN 978-5-06-005714-0

2. **Шириков, В. Ф.** Теория вероятностей : учебное пособие / В. Ф. Шириков, С. М. Зарбалиев. - М. : КолосС, 2008. - 389 с. - (Учебники и учеб. пособия для студентов высш. учеб. заведений). - ISBN 978-5-9532-0621-1

3. **Прохоров, Ю.В.** Вероятность и математическая статистика : энциклопедический словарь / ред. Ю. В. Прохоров. - Репр. воспроизведение изд. - М. : Большая Российская Энциклопедия, 2003. - 910 с. - (Золотой фонд). - ISBN 5-7107-7433-2

4. **Ермакова, В.И**. Теория вероятностей и математическая статистика : учебное пособие / ред. В. И. Ермакова. - М. : Инфра-М, 2004. - 287 с. - (Высшее образование). - ISBN 5-16 001561-2

5. **Гмурман, В. Е.** Теория вероятностей и математическая статистика : учебное пособие / В. Е. Гмурман. - 12-е изд., перераб. - М. : Высш. образование, 2006. - 479 с. : ил. - (Основы наук).

дополнительная литература:

1. **Шипачев, В. С.** Курс высшей математики : учебник для вузов / Под ред. А. Н. Тихонова. - 3-е изд., испр. - М. : ОНИКС, 2007. - 600 с.
2. **Юрьева, А. А.** Руководство к решению задач по теории вероятностей и математической статистике. : учебное пособие / А. А. Юрьева. - 2-е изд., испр. - Саратов : ФГОУ ВПО "Саратовский ГАУ", 2009. - 208 с.
3. **Кириллова, Т. В., Хучраева Т.С.** Элементы математической статистики / Т. В., Хучраева Т.С. Кириллова. - Саратов : Сарат. гос. агр. ун-т, 2004. - 60 с. - ISBN 5-7011-0394-3

Интернет-ресурсы:

- http://ru.wikipedia.org -Википедия;
- www.newlibrary.ru - новая электронная библиотека;
- www.edu.ru – федеральный портал российского образования;
- www.mathnet.ru – общероссийский математический портал;
- www.elibrary.ru – научная электронная библиотека;
- www.matburo.ru – матбюро: решения задач по высшей математике;
- www.nehudlit.ru - электронная библиотека учебных материалов.